

# Characterisation of local halo building blocks: Thamnos and Sequoia

Emma Dodd<sup>1</sup>, Tomás Ruiz-Lara<sup>2,3</sup>, Amina Helmi<sup>1</sup>, Carme Gallart<sup>4,5</sup>, Thomas M. Callingham<sup>1</sup>, Santi Cassisi<sup>6,7</sup>,  
Emma Fernández-Alvar<sup>4,5</sup>, and Francisco Surot<sup>4</sup>

<sup>1</sup> Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands

<sup>2</sup> Universidad de Granada, Departamento de Física Teórica y del Cosmos, Campus Fuente Nueva, Edificio Mecenas, E-18071, Granada, Spain

<sup>3</sup> Instituto Carlos I de Física Teórica y computacional, Universidad de Granada, E-18071 Granada, Spain

<sup>4</sup> Instituto de Astrofísica de Canarias, E-38200 La Laguna, Tenerife, Spain

<sup>5</sup> Departamento de Astrofísica, Universidad de La Laguna, E-38205 La Laguna, Tenerife, Spain

<sup>6</sup> INAF – Astronomical Observatory of Abruzzo, via M. Maggini, sn, 64100 Teramo, Italy

<sup>7</sup> INFN, Sezione di Pisa, Largo Pontecorvo 3, 56127 Pisa, Italy

## ABSTRACT

**Context.** A crucial aspect of galaxy evolution is the pace at which galaxies build up their mass. We can investigate this hierarchical assembly by uncovering and timing accretion events experienced by our Galaxy.

**Aims.** In the Milky Way, accreted debris has been previously identified in the local halo, thanks to the advent of *Gaia* data. We aim to couple this dataset with advancements in colour-magnitude diagram (CMD) fitting techniques to characterise the building blocks of the Galaxy, based on their age and metallicity distributions. Here, we focus on the retrograde halo, specifically those of Thamnos and Sequoia.

**Methods.** We conducted this study as part of the ChronoGal project by fitting the absolute CMDs (using CMDfit.*Gaia*) of samples of stars associated with these sub-structures, extracted from a local 5D *Gaia* DR3 dataset. Comparing their derived age and metallicity distributions with those of the expected contamination, from the dominant *Gaia* Enceladus (GE) and low-energy (LE) in situ populations, we can unveil the stellar population signatures of the progenitors of Sequoia and Thamnos.

**Results.** We show that both Thamnos and Sequoia have a metal-poor population ( $[Fe/H] \sim -2.5$  to  $-1.5$  dex) that is distinct from the expected contamination. Their age distributions offer us the ability to see the pace of the build-up of their progenitors. Half of the stars in Sequoia were formed by the look-back time of  $\sim 12_{-0.3}^{+0.3}$  Gyr. Thamnos appears slightly older, on average, and declines quickly, having formed half its stars at  $\sim 12.3_{-0.3}^{+0.3}$  Gyr. Compared to GE and the LE in situ populations, they formed half of their stars by  $12.1_{-0.1}^{+0.1}$  Gyr and  $12.9_{-0.1}^{+0.1}$  Gyr, respectively. Caution should be taken when interpreting the age distributions, especially that of Sequoia, due to the low number of stars, which can cause shifts to younger ages of up to  $\sim 1$  Gyr. However, considering these potential shifts and the underlying contamination that is inherently difficult to remove completely, our results allow us to safely conclude that Thamnos, *Gaia* Enceladus, and Sequoia are all predominantly old and accreted at similar epochs, within  $\sim 1$ -2 Gyr of each other.

**Conclusions.** We present, for the first time, the age distributions of the retrograde halo sub-structures: Sequoia and Thamnos. These have been derived from purely photometric data using CMD fitting techniques, which also provide metallicity distributions that successfully reproduce the results from spectroscopy, highlighting the applicability of CMDfit.*Gaia*.

## 1. Introduction

According to  $\Lambda$ CDM cosmology, galaxies assemble hierarchically over time, growing their halos through mergers and the accretion of smaller galaxies (Springel et al. 2005). Identifying and characterising the debris in our own Galaxy’s halo offers insights into the significance of accretion in this process, along with overall processes of galaxy formation and evolution.

Thanks to the advent of *Gaia* data (Prusti et al. 2016; Gaia Collaboration et al. 2022), we are now able to identify this type of merger debris in our local stellar halo as an over-density in the integrals of motion (IoM) space (Helmi & Tim de Zeeuw 2000). This has provided insights into the assembly history, as well as on how mergers have played a role in shaping the Galaxy (for recent reviews see Helmi 2020; Deason & Belokurov 2024). *Gaia* data have revealed that the local stellar halo is dominated by debris from a single merger, *Gaia* Enceladus/Sausage (Helmi et al. 2018; Belokurov et al. 2018). Since its discovery, extensive work has been done to characterise the debris of this merger, providing stellar mass estimates in the range of  $\sim 10^8$ - $10^{10} M_{\odot}$  (e.g. Helmi

et al. 2018; Vincenzo et al. 2019; Feuillet et al. 2020; Mackereth & Bovy 2020; Lane et al. 2023, with more recent studies predicting toward the lower mass end than initial estimates) and an estimated accretion time of  $\sim 10$  Gyr ago, equivalently  $z \sim 2$  (e.g. Helmi et al. 2018; Gallart et al. 2019; Bonaca et al. 2020; Montalbán et al. 2021). Additionally, several other dynamical sub-structures have been uncovered (e.g. Koppelman et al. 2019; Naidu et al. 2020; Yuan et al. 2020; Horta et al. 2021; Lövdal et al. 2022; Ruiz-Lara et al. 2022; Dodd et al. 2023), although their nature (accreted or in situ) and characterisation are rather limited or even non-existent.

The retrograde part of the halo has been revealed to be especially rich in small sub-structures that are either chemically or kinematically distinct from one another (e.g. Dodd et al. 2023; Oria et al. 2022; Ceccarelli et al. 2024). The dominant debris appears to come from three main accretion events; *Gaia* Enceladus, Sequoia, and Thamnos, the latter of which have not yet been characterised (in detail).

Sequoia is a loosely bound retrograde sub-structure identified first by Myeong et al. (2019) as a chemically distinct accre-

tion event in the halo (see also [Matsuno et al. 2019, 2022](#)). An estimate of the stellar mass of  $\sim 5 \times 10^7 M_{\odot}$  has been proposed, along with an accretion time of around 9 Gyr ago ([Myeong et al. 2019](#)). However, it is becoming more clear that the retrograde halo is more complex than we first thought. Simulations have shown that debris from a massive merger such as *Gaia* Enceladus (GE, hereafter) can populate this region, depending on its initial orbital configuration and morphology (e.g. [Koppelman et al. 2020; Naidu et al. 2021; Amarante et al. 2022](#)). Observations have revealed that even though Sequoia stars at low metallicity are chemically distinct from GE ([Myeong et al. 2019; Matsuno et al. 2022](#)), above  $[\text{Fe}/\text{H}] \sim -1.5$ , the chemical abundances are indistinguishable from those of GE debris ([Ceccarelli et al. 2024](#)). The level of contamination by GE in that region of IoM is suggested to be  $\sim 10\text{-}20\%$  ([Matsuno et al. 2022](#)).

Initially, Sequoia was defined in [Myeong et al. \(2019\)](#) to extend in energy down to low orbital energies (high binding energy). Since then, [Koppelman et al. \(2019\)](#) discovered the Thamnos sub-structure, separating the retrograde halo into high  $E_n$  (Sequoia) and low  $E_n$  (Thamnos) for the first time, owing to differences in chemistry between the two populations. Given the mean metallicities of the Sequoia stars, it seems unlikely that its progenitor could have been massive enough (because of the mass-metallicity relationship; e.g. [Kirby et al. 2013; Ma et al. 2016](#)) to create such a large spread in IoM, as that first proposed by [Myeong et al. \(2019\)](#).

Thamnos itself was first defined by [Koppelman et al. \(2019\)](#) as two independent sub-structures: Thamnos 1 and 2. Given their similar abundances and stellar populations, it is possible that the two sub-structures share the same progenitor (see also [Ruiz-Lara et al. 2022](#)) and from the extent in IoM space, [Koppelman et al. \(2019\)](#) placed an upper limit on the stellar mass of  $5 \times 10^6 M_{\odot}$ . It is worth noting that Thamnos 2, which is the less retrograde and more bound sub-structure, contains stars that have  $[\text{Fe}/\text{H}]$  of  $\sim -1.3$  dex that are Al-enhanced; a population that is likely to be contamination of in situ origin (see Fig. 4 in [Koppelman et al. 2019](#), and also [Horta et al. 2021](#)). Subsequently, [Ruiz-Lara et al. \(2022\)](#) using *Gaia* EDR3 supplemented by LAMOST ([Cui et al. 2012](#)), also confirmed that Thamnos 2 is more metal-rich than Thamnos 1, again possibly due to increasing contamination as  $L_z$  becomes less negative. We note, however, that this is based on a limited number of stars with chemical information. More recently, thanks to *Gaia* DR3 data, Thamnos has been identified with clustering in IoM space as one single sub-structure ([Dodd et al. 2023](#)) and its chemistry still displays a mix of stellar populations, including contamination from in situ components and probably from *Gaia* Enceladus as well.

In addition to accreted sub-structures, the local halo is known to host structures of an in situ origin. The colour-magnitude diagram (CMD) of local halo stars was shown with *Gaia* DR2 to consist of two sequences ([Babusiaux et al. 2018](#)). The blue sequence has been attributed to the accreted halo and dominated by *Gaia* Enceladus debris and the red sequence is said to be heated thick-disc stars (e.g. [Haywood et al. 2018](#)). It is now believed that thick-disc stars already present during the merger with *Gaia* Enceladus were heated onto halo-like orbits, producing what we now call the hot thick disc or ‘the Splash’ (e.g. see [Helmi et al. 2018; Gallart et al. 2019; Belokurov et al. 2020](#)).

A more ancient in situ halo component has been uncovered using  $[\text{Al}/\text{Fe}]$  abundances, referred to as the Aurora population ([Belokurov & Kravtsov 2022](#)), and is hypothesised to have formed during a period of chaotic (pre-disc) evolution. This population exhibits a rapid evolution of  $[\text{Al}/\text{Fe}]$  to anomalously high  $[\text{Al}/\text{Fe}]$ , which can be explained by a rapidly star-forming and

self-enriching galaxy before supernovae of type Ia (SN Ia) start to contribute significantly and dilute the  $[\text{Al}/\text{Fe}]$ . As star formation happens on a longer timescale in dwarf galaxies, they do not reach as high  $[\text{Al}/\text{Fe}]$  before the SN Ia kick in ([Hawkins et al. 2015; Belokurov & Kravtsov 2022](#)). The chemical abundances of the Aurora population also suggest that the proto-Milky Way formed with a large contribution from massive star clusters ([Belokurov & Kravtsov 2022; Myeong et al. 2022](#)). However, [Horta et al. \(2021, 2024a\)](#) show that this proto-Milky Way population can be chemically indistinguishable from an early (massive) merger event.

Our understanding of the accretion events that contribute to the local Milky Way halo has, to date, mostly focussed on a dynamical perspective, made possible thanks to the *Gaia* mission. For sub-sets of the stars associated to sub-structures, this has been coupled with chemical information from ground-based spectroscopic surveys (e.g. LAMOST, [Zhao et al. 2012](#); GALAH, [De Silva et al. 2015](#); APOGEE, [Majewski et al. 2017](#)) or targeted follow-up (e.g. [Aguado et al. 2021; Matsuno et al. 2022; Ceccarelli et al. 2024](#)) to obtain a chemo-dynamical view. The future of the characterisation of halo sub-structures relies upon our access to the chemistry and age information for a larger sample of stars, for which we already have the dynamical information from *Gaia*. The chemical characterisation will improve greatly with the upcoming wide-area spectroscopic surveys, such as WEAVE ([Jin et al. 2023](#)) and 4MOST ([De Jong et al. 2019](#)).

An important dimension of the assembly history that is still (for the most part) missing is the chronology of the individual accretion events. Specifically, to be able to assess how quickly galaxies build up their stellar populations, as well as to determine what the timelines for galaxy assembly and the duration of merger events are. Unfortunately, age estimates for stars are notoriously difficult to obtain for large samples, since the ages of stars cannot be directly measured and have to be inferred by comparing observed properties with stellar evolution models. Currently, there are two main methods for obtaining age estimates of individual stars: isochrone fitting and asteroseismology (e.g. for a review see [Soderblom 2010](#)); both of which have their own limitations in terms of precision and sample sizes.

Alternatively, stellar age distributions can be obtained from a CMD fitting of composite stellar populations ([Dolphin 2002; Cignoni & Tosi 2010](#)). This technique has already been well established for deriving the star formation histories of dwarf satellites in the Local Group (e.g. [Gallart et al. 1999; Noël et al. 2009; Tolstoy et al. 2009; Bernard et al. 2012; Weisz et al. 2012; Cignoni et al. 2013; Cole et al. 2014; Skillman et al. 2017; McQuinn et al. 2024](#)). Thanks to the combination of *Gaia* data and developments in the CMD fitting technique, it is now feasible to do this for homogeneous samples of stars in the Milky Way (e.g. [Ruiz-Lara et al. 2020; Dal Tio et al. 2021; Mazzi et al. 2024; Gallart et al. 2024](#)), including the building blocks identified as accreted sub-structures (e.g. see [Gallart et al. 2019](#) and [Ruiz-Lara et al. 2022](#) for the characterisation of the stellar content of *Gaia* Enceladus and the Helmi streams, respectively). We note that this method does not require any assumptions on the age-metallicity relation, the metallicity distributions, or the functional form of the star formation rate over time (for more details, see [Gallart et al. 2024](#)). It provides us with a detailed age distribution for a given sample of stars. Applying this to the building blocks of the Milky Way allows us to construct a timeline of the mergers our Galaxy has experienced and examining when a system had been quenched by accretion.

In this paper, we focus on Thamnos and Sequoia. By fitting their absolute CMDs, we have been able to obtain age and metal-

licity distributions for these sub-structures. Section 2 presents the datasets used in this study and the selections of our samples. The methodology is presented in Sect. 3, including the CMD fitting procedure (Sect. 3.2), the treatment of contamination (Sect. 3.3), and the performance of mock tests to validate the results (Sect. 3.4). The results are presented in Sect. 4 were derived using only photometry, demonstrating that the metallicity distributions of the sub-structures resemble the spectroscopic distributions, while also providing the added age dimension that enables us to see more clearly differences in their formation histories. A discussion of these results is presented in Sect. 5, highlighting the importance of taking into account contamination from the dominant *Gaia* Enceladus and the in situ halo populations (Sect. 5.1) and the complexity of the Thamnos region of IoM space (Sect. 5.4). Section 6 summarises our results and conclusions.

## 2. Data

In a previous work, we developed a clustering algorithm that is able to identify several overdensities, including Thamnos and Sequoia, when applied to the 6D local halo sample in a statistically robust manner (Lövdal et al. 2022; Ruiz-Lara et al. 2022; Dodd et al. 2023). To perform the CMD fitting, we must be able to model the selection functions of our samples; so, for our 6D samples this means the *Gaia* RVS selection function (see Sect. 3.1). For smaller sub-structures (e.g. Sequoia), the number of member stars with radial velocities remains limited, making CMD fitting of these 6D samples difficult. For these reasons, we turned to 5D datasets. We first used the 6D samples to select the sub-structures (Sec 2.1 & 2.2) and to then motivate our selections from a local 5D sample (Sec 2.3; similarly to what was done in Ruiz-Lara et al. 2022). Selecting our samples in 5D allows us to avoid having to model complicated selection functions. Even more importantly, it increases the number of bright stars, allowing us to obtain a complete sample down to  $M_G < 5$  within 2.5 kpc.

### 2.1. 6D Halo Sample

Our 6D halo sample was defined from *Gaia* DR3 data according to Dodd et al. (2023). We selected stars with (total) relative parallax uncertainty less than 20% as:  $(\varpi - \Delta_\varpi) / \sqrt{\sigma_{\text{parallax}}^2 + \sigma_{\text{sys}}^2} \geq 5$ , where  $\Delta_\varpi$  represents the individual zero-point offsets determined following Lindegren et al. (2021),  $\sigma_{\text{parallax}}$  is the `parallax_error`, and  $\sigma_{\text{sys}}$  is the systematic uncertainty on the zero-point, which we take to be 0.015 mas (Lindegren et al. 2021). To obtain a distance, we inverted the corrected parallax and we selected stars within 2.5 kpc of the Sun. One difference with respect to the study of Dodd et al. (2023) is that we here we have extended our sample from stars with a radial velocity in *Gaia* DR3 to also include stars with a radial velocity from ground-based spectroscopic surveys (as done also for EDR3 in Lövdal et al. 2022) such as LAMOST DR7 (Cui et al. 2012), APOGEE DR17 (Acetia et al. 2022), GALAH DR3 (Buder et al. 2021), and RAVE DR6 (Steinmetz et al. 2020). The sub-structures are defined still as those in Dodd et al. (2023) from clustering on the RVS sample, which is more robust. Then we added extra stars with radial velocities from ground-based spectroscopic surveys to our predefined clusters (see Sect. 2.2 for more details).

We cleaned the sample with  $\text{RUWE} < 1.4$ , line of sight velocity uncertainty,  $\epsilon(V_{\text{los}}) < 20$  km/s,  $(G_{\text{RVS}} - G) > -3^2$ , and for stars with  $|b| < 7.5^\circ$  we require<sup>3</sup> `rv_expected_sig_to_noise`  $> 5$ . After correcting for the solar motion using  $(U, V, W)_\odot = (11.1, 12.24, 7.25)$  km/s (Schönrich et al. 2010) and for the motion of the local standard of rest (LSR) using a  $|\mathbf{V}_{\text{LSR}}|$  of 232.8 km/s (McMillan 2017), we selected halo stars as  $|\mathbf{V} - \mathbf{V}_{\text{LSR}}| > 210$  km/s. This resulted in a 6D halo sample of 101,419 stars, 71,849 of which have a *Gaia* radial velocity. We calculate the integrals of motion (IoM) for stars (energy;  $E$ , and angular momenta;  $L_z$ ,  $L_\perp$ ), computing  $E$  using the same potential as in Dodd et al. (2023). This potential consists of a Miyamoto-Nagai disc with parameters  $(a_d, b_d) = (6.5, 0.26)$  kpc,  $M_d = 9.3 \times 10^{10} M_\odot$ , a Hernquist bulge with  $c_b = 0.7$  kpc,  $M_b = 3.0 \times 10^{10} M_\odot$ , and an NFW halo with  $r_s = 21.5$  kpc,  $c_h = 12$ , and  $M_{\text{halo}} = 10^{12} M_\odot$ . We defined  $L_z$  to be positive for prograde stars. While we limited our sample to those with  $\epsilon(V_{\text{los}}) < 20$  km/s, the majority of the stars have a significantly smaller  $V_{\text{los}}$  uncertainty. The mean  $V_{\text{los}}$  uncertainty for the full sample is 5.2 km/s and for the *Gaia* RVS sub-sample is 4.5 km/s. By propagating the errors in  $V_{\text{los}}$  into errors in IoM space  $(E, L_z, L_\perp)$ , we checked that these uncertainties had limited effect on the assigning of stars to sub-structures.

### 2.2. Sub-structures in 6D

We followed the definitions for the local halo sub-structures identified in Dodd et al. (2023), obtained by clustering on the *Gaia* DR3 RVS sample that has lower uncertainties. Then we added extra stars with radial velocities from ground-based spectroscopic surveys to our predefined sub-structures, by selecting those that fall within an ellipsoid which contains 80% of the original sub-structure members. This is equivalent to cutting in Mahalanobis distance<sup>4</sup> at 2.13, which was found previously to minimise adding noise to sub-structures (see Ruiz-Lara et al. 2022, for more details).

Figure 1 shows this 6D halo sample within 2.5 kpc, with *Gaia* Enceladus (GE), Thamnos, and Sequoia selections shown in coloured markers. We also included another sub-set that is seen to be below GE in terms of energy, which we refer to as the ‘low energy’ (LE) sample. This samples was selected to mimic the contamination we expect in the Thamnos region of IoM space. This selection is defined using stars in a narrow range of  $v_y$  between  $-230$  and  $-250$  km/s, and within  $|v_x| < 180$  km/s, combined with a cut of  $E_n < -1.35 \times 10^5$  km<sup>2</sup>/s<sup>2</sup> so that we are not overlapping with our pure GE selection in energy. This selects a general halo sample on highly bound orbits and appears to be a mix of populations, including *Gaia* Enceladus debris at low energy and with an in situ component.

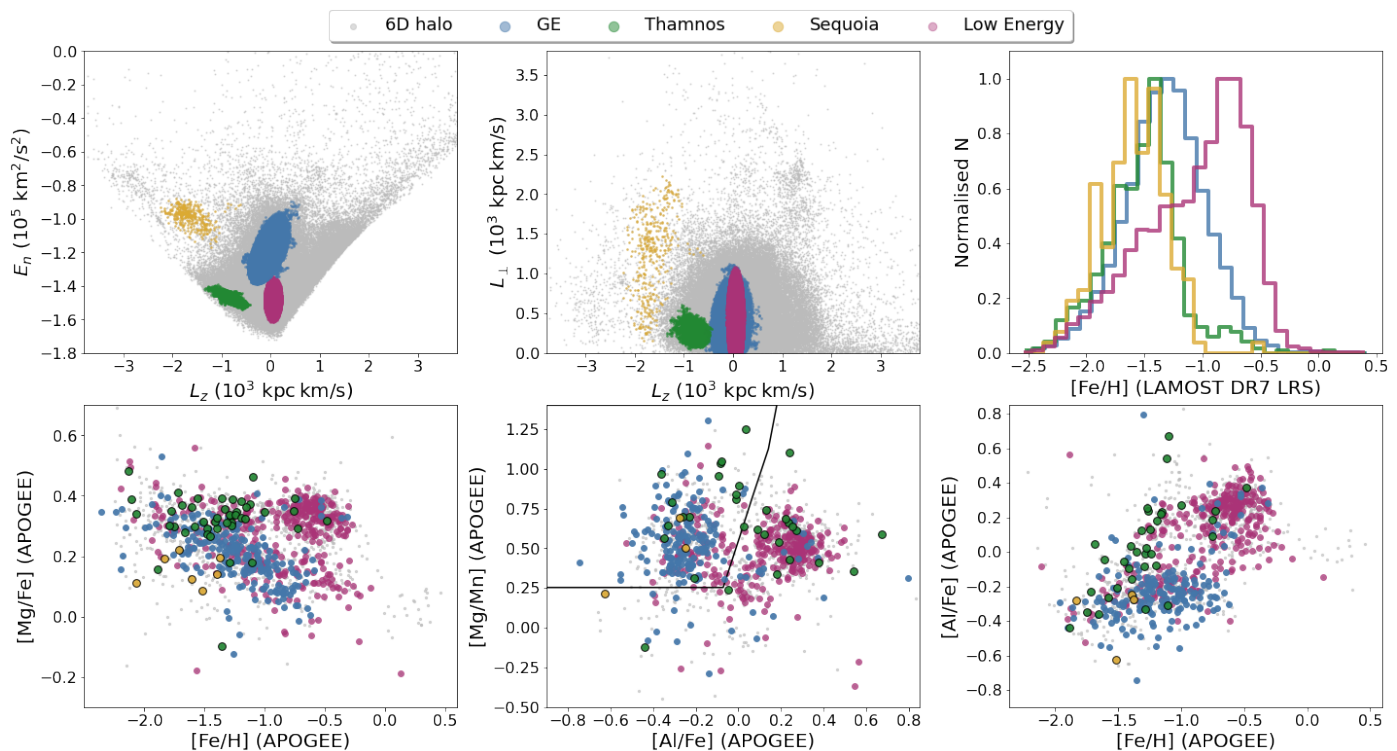
Figure 2 shows the CMDs of these populations, demonstrating that all of these sub-structures are composed mainly of old stars with metallicities that are different (as shown in Fig. 1) among the structures (as seen by the different colours of the RGB

<sup>1</sup> after applying the correction to `radial_velocity_error` recommended by Babusiaux et al. (2022)

<sup>2</sup> to remove problematic radial velocities as recommended by Babusiaux et al. (2022)

<sup>3</sup> to avoid highly contaminated spectra and spurious velocities, following Katz et al. (2023, Sect. 9)

<sup>4</sup> The Mahalanobis distance between a star from a ground-based spectroscopic survey and a sub-structure in IoM space is given as:  $D' = \sqrt{(\boldsymbol{\mu}_{\text{sub}} - \boldsymbol{\mu}_*)^T \boldsymbol{\Sigma}_{\text{sub}}^{-1} (\boldsymbol{\mu}_{\text{sub}} - \boldsymbol{\mu}_*)}$  where  $\boldsymbol{\mu}_{\text{sub}}$  and  $\boldsymbol{\Sigma}_{\text{sub}}$  correspond to the mean and covariance matrix of the sub-structure stars, and  $\boldsymbol{\mu}_*$  is the position of the star in IoM space.



**Fig. 1.** 6D samples relevant to this work. We selected *Gaia* Enceladus, Thamnos, and Sequoia, according to Dodd et al. (2023) and defined another sample referred to as the low-energy (LE) sample, shown in magenta. This sample mimics the contamination we expect at low energy for Thamnos, as we can see more clearly in the APOGEE DR17 abundances (bottom row). The top row shows integrals of motion (IoM) space and LAMOST DR7 LRS metallicity distributions normalised and binned each with a relative shift of 0.01 dex to aid in the comparison. The bottom row shows APOGEE DR17 abundances for the different sub-structures, where the Thamnos and Sequoia stars are shown with a black outline.

and position of the old main sequence turn-off; oMSTO). The CMD fitting will help us further unveil these differences in age and metallicity. We note that both Sequoia and Thamnos exhibit an extended horizontal branch with higher relative contributions of horizontal branch (HB) stars towards the blue end (especially in Thamnos; see Fig. 2).

Figure 1 shows that Sequoia is a sub-structure lying in a region of IoM space that is relatively well separated from other sub-structures. We define Sequoia as a smaller region in IoM space than the typical size assumed in past studies (e.g. Myeong et al. 2019; Koppelman et al. 2019; Yuan et al. 2020; Naidu et al. 2020, for comparison see Figure 13 in Lövdal et al. 2022). This is because Ruiz-Lara et al. (2022) showed that the larger retrograde selection includes sub-structures that have distinct kinematics and that given Sequoia’s estimated mass (from its mean metallicity), it would be inconsistent with a very broad extent in the IoM space. As previously outlined, it is evident from both simulations (e.g. Koppelman et al. 2020; Naidu et al. 2021; Amante et al. 2022) and observations using chemical abundances (e.g. Matsuno et al. 2022; Ceccarelli et al. 2024) that this region of IoM space is contaminated by debris from the *Gaia* Enceladus merger. This is something that we consider in this work.

Thamnos is a low (orbital) energy retrograde structure (see Fig. 1) and its location at low energy means that any selection of Thamnos stars will be contaminated from *Gaia* Enceladus (Helmi et al. 2018) and other populations, such as possibly an in situ halo or Aurora (i.e. the proto-Galaxy), as both dominate at low energy (Belokurov & Kravtsov 2022). These structures are much larger than what we believe the true Thamnos population to be and, hence, on average more metal-rich. This is why we introduced the LE sample, which is a combination of the in situ

halo and likely some *Gaia* Enceladus debris, used to decontaminate Thamnos and characterise the true signal in our analyses.

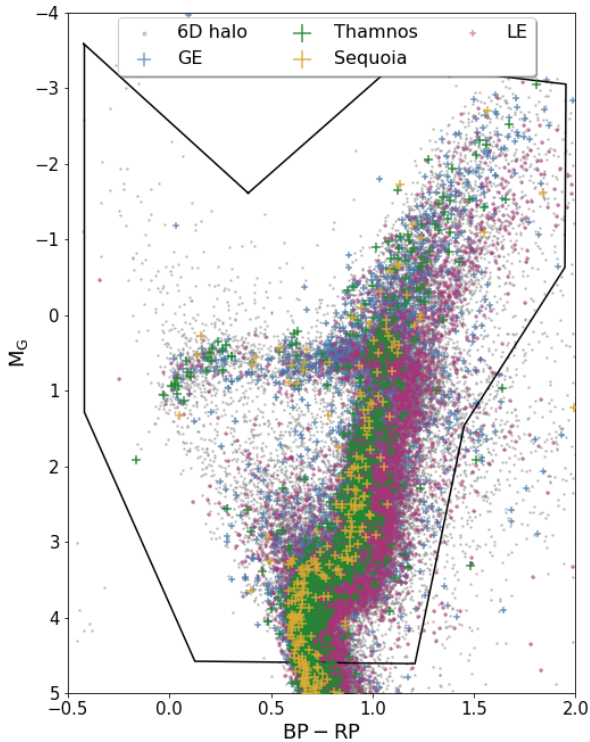
### 2.2.1. Metallicity distributions

To quantitatively establish the differences in the LAMOST DR7 LRS metallicity distributions of Sequoia with GE and Thamnos with the LE sample, we compared the MDFs using the two sample Kolmogorov–Smirnov (KS) test (Kolmogorov 1933; Smirnov 1939), where the KS statistic gives the maximum difference of the cumulative metallicity distributions.

For Sequoia and *Gaia* Enceladus, we found a KS test statistic of 0.38 (with a significance of  $\sim 9\sigma$ ) and we can confidently say that the two samples are not drawn from the same distribution. This maximum difference in the cumulative metallicity distributions occurs at  $[\text{Fe}/\text{H}] = -1.31$  dex, however, the distributions are already statistically different at the 95% level<sup>5</sup> from  $[\text{Fe}/\text{H}] \sim -1.8$ .

For Thamnos (looking at Fig. 1), we can see from the metallicity distributions that Thamnos (green) has a more pronounced metal-poor tail around  $[\text{Fe}/\text{H}] \lesssim -2$  than GE and the LE sample, which we believe is the signature of the small dwarf galaxy. On the other hand, the more metal-rich stars in the Thamnos region are a signature of this contamination. In fact, Fig. 1 shows that there is a large peak in the metallicity distribution around  $[\text{Fe}/\text{H}] \sim -1.4$  dex, which is a combination of GE contamination and an Al-enhanced population of similar metallicity that can be seen in the bottom-left and right-most panels. This is the signature of

<sup>5</sup> using that the critical value for the two sample KS test of sizes  $n_1, n_2$  can be approximated as  $c(\alpha) \sqrt{(n_1 + n_2)/n_1 n_2}$  with  $c(0.05) = 1.36$



**Fig. 2.** CMD of bright ( $M_G < 5$ ) 6D selected Thamnos, Sequoia, *Gaia* Enceladus (GE), and low energy (LE) samples. The region outlined in black is the bundle used for fitting the CMDs, as described in Sect. 3.2, and is chosen to extend  $\approx 1$  magnitude below the main sequence turn-off. The background sample in grey is our full 6D halo sample.

the ancient population of Aurora (Belokurov & Kravtsov 2022) and likely the result of intense rapid star formation in the early Milky Way.

By statistically comparing the metallicity distributions for Thamnos and our LE sample, we can confidently say that the samples are drawn from different distributions, with a KS test statistic of 0.54 (at a significance of  $> 10\sigma$ ). This maximum difference of the cumulative metallicity distributions occurs at  $[\text{Fe}/\text{H}] = -1.13$  dex; however, the distributions are already statistically different at the 95% level from  $[\text{Fe}/\text{H}] \sim -1.9$  dex. This supports that there is a statistically significant metal-poor tail in our Thamnos selection corresponding to something other than the overall low-energy halo.

### 2.2.2. Chemical abundances

In the bottom row of Fig. 1 we show APOGEE DR17 abundances of the 6D sub-structures relevant to this work, which supports our definitions of contamination. We show in the first panel the  $[\alpha/\text{Fe}]-[\text{Fe}/\text{H}]$  plane with Mg as our  $\alpha$ . Accreted galaxies are expected to have lower  $[\alpha/\text{Fe}]$  at fixed  $[\text{Fe}/\text{H}]$  than in situ stars since this is indicative of a slower enrichment (e.g. Horta et al. 2023). Individual progenitors are expected to follow distinct tracks in this space; for example, Sequoia which has been shown to present lower  $[\alpha/\text{Fe}]$  than GE at fixed metallicity (Matsuno et al. 2022). The middle panel shows the  $[\text{Al}/\text{Fe}]-[\text{Mg}/\text{Mn}]$  plane, which is commonly used to distinguish accreted (chemically unevolved) and in-situ (chemically evolved) populations (as proposed by Das et al. 2020). This is shown by the black line

where accreted and unevolved stars exhibit lower  $[\text{Al}/\text{Fe}]$  at similar  $[\text{Mg}/\text{Mn}]$  to in situ stars. The last panel shows an alternative projection of  $[\text{Fe}/\text{H}]$  versus  $[\text{Al}/\text{Fe}]$ , for which we see the plume of stars reaching high  $[\text{Al}/\text{Fe}]$  at  $[\text{Fe}/\text{H}] \sim -1.3$  dex for Thamnos (present also in the LE sample) and is likely in situ stars associated to the Aurora population (Belokurov & Kravtsov 2022). The LE sample is dominated by in situ stars, but does contain a significant accreted or unevolved population. *Gaia* Enceladus dominates the accreted halo regime and presents a distinct track in  $[\alpha/\text{Fe}]-[\text{Fe}/\text{H}]$  space along which some Thamnos and Sequoia stars lie.

### 2.3. 5D Sample

We now move onto the 5D sample. We used this sample to draw samples with important contributions from Thamnos and Sequoia to characterise their stellar content (see Sect. 3.2). For our 5D sample, we calculate absolute  $G$  magnitude and colour ( $G_{\text{BP}} - G_{\text{RP}}$ ) correcting for extinction using the Lallement et al. (2022) dust maps and using the transformation from E(B-V) to *Gaia* extinction from Fitzpatrick et al. (2019).

We then applied several quality cuts on the photometry to obtain clean samples for fitting the CMDs. We removed stars affected by high extinction,  $A_G > 0.5$  and only kept stars with  $M_G < 5$  (the photometric regime for which *Gaia* is complete in our volume of 2.5 kpc) and a good `phot_bp_rp_excess_factor` using:

$$0.001 + 0.039 \times \text{bp\_rp} < \log(\text{phot\_bp\_rp\_excess\_factor})$$

and

$$\log(\text{phot\_bp\_rp\_excess\_factor}) < 0.12 + 0.039 \times \text{bp\_rp}.$$

Here, `bp_rp` is the apparent ( $G_{\text{BP}} - G_{\text{RP}}$ ) colour. This leaves 30.6 million bright 5D stars within 2.5 kpc, where the distances are calculated by inverting the corrected parallax, as described in Sect. 2.1 for the 6D halo sample. Within this sample, 12.5 million stars have a radial velocity (from *Gaia* or ground-based spectroscopic surveys as described in Sect. 2.1) but the majority of these are thin-disc and thick-disc stars. The CMD of this 5D sample presents stars of all evolutionary stages, including young, bright main sequence stars (as expected from a sample dominated by disc stars e.g. Babusiaux et al. 2018). There are 71,977 of the 12.5 million stars with a radial velocity that would be considered halo stars according to  $|\mathbf{V} - \mathbf{V}_{\text{LSR}}| > 210$  km/s. We note that for this sample no quality cuts have been applied to the radial velocity from *Gaia* and no cut in RUWE has been applied since this removes binaries which are included in our synthetic mother CMD (see also Sect. 3.1).

Although the sample is dominated by disc stars we can quite confidently separate the (accreted) halo stars using 5D information, specifically using the pseudo-Cartesian velocity space,  $\tilde{v}_x$ ,  $\tilde{v}_y$ , calculated assuming the radial velocity is 0 km/s such that

$$\tilde{v}_x = -v_{l,*} \sin(l) - v_{b,*} \cos(l) \sin(b), \quad (1)$$

$$\tilde{v}_y = v_{l,*} \cos(l) - v_{b,*} \sin(l) \sin(b), \quad (2)$$

$$\tilde{v}_z = v_{b,*} \cos(b), \quad (3)$$

where  $v_{i,*} = v_i + v_{i,\odot}$  with  $i = (l, b)$  and

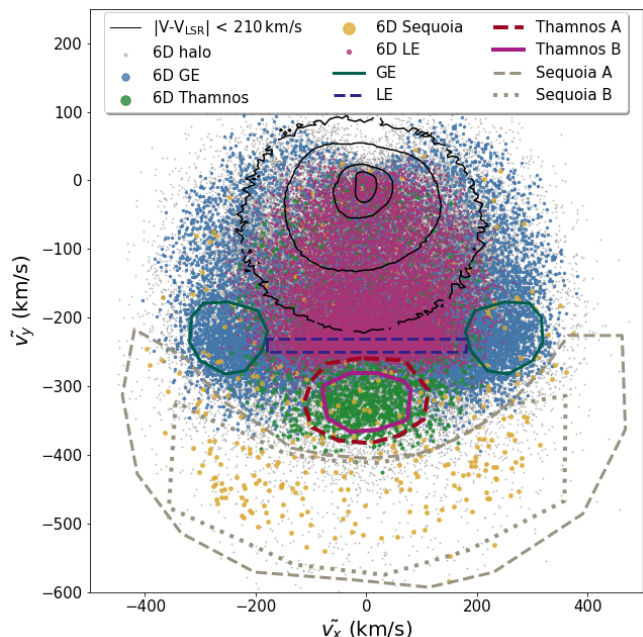
$$v_i = 4.74057 \text{ km s}^{-1} \left( \frac{\mu_i}{\text{mas yr}^{-1}} \right) \left( \frac{d}{\text{kpc}} \right) \quad (4)$$

and

$$v_{l,\odot} = -U_{\odot} \sin(l) + (V_{\odot} + V_{\text{LSR}}) \cos(l), \quad (5)$$

$$v_{b,\odot} = W_{\odot} \cos(b) - \sin(b)[U_{\odot} \cos(l) + (V_{\odot} + V_{\text{LSR}}) \sin(l)], \quad (6)$$

using  $(U, V, W)_\odot$  given in Sect. 2.1. This is demonstrated in Fig. 3 which shows where the 6D stars would fall in this pseudo-Cartesian velocity space if we did not know their radial velocity and assumed it to be 0 km/s. The majority of stars with  $|\mathbf{V} - \mathbf{V}_{\text{LSR}}| < 210$  km/s (disc stars) have  $\tilde{v}_y \gtrsim -100$  km/s, which does not overlap with the regions of interest in the halo.



**Fig. 3.** Selections of Thamnos, Sequoia, *Gaia* Enceladus (GE), and low-energy (LE) samples in pseudo-Cartesian velocity space, with  $\tilde{v}_x$  and  $\tilde{v}_y$ . Stars with 6D information that belong to the aforementioned structures are shown as coloured markers. The polygons are chosen to select each structure in this space, where the 6D stars pile up, whilst reducing contamination from other structures. The background sample in grey is our full 6D halo sample. Those stars with  $|\mathbf{V} - \mathbf{V}_{\text{LSR}}| < 210$  km/s (i.e. the disc and hot thick disc) are shown with the black contours at levels of 20%, 68%, 95%, and 99%. The LE sample was defined in  $v_x - v_y$  space using the same box selection as shown here and used to select stars that are offset in  $v_y$  from Thamnos and in  $v_x$  from *Gaia* Enceladus.

To characterise the Sequoia and Thamnos’ stellar populations, we need to be able to select sub-sets of these stars from our 5D sample. As already discussed, we chose to progress to 5D because it is complete (in the range of distance and absolute magnitude relevant to this work) and also allows us to avoid having to model complicated selection functions and, most importantly, it increases the number of stars. However, it is important to note that going to 5D does add additional contamination, which could make the results more difficult to interpret.

### 2.3.1. 5D selection of sub-structures

We followed a two-step process to select the relevant sub-structures in the 5D sample. First, we addressed the pseudo-Cartesian velocity space,  $\tilde{v}_x$ ,  $\tilde{v}_y$ , calculated assuming the radial velocity is 0 km/s as shown in Eqs. 3, 4, and 6. In Fig. 3 the 6D Thamnos, Sequoia, GE, and LE samples are also shown in this pseudo-Cartesian velocity space. The regions where each specific sub-structure dominates and separates from the other structures are outlined by polygons that we used to make a first selection of the structures. For Thamnos and Sequoia, we had two selections for each, one of which (B) is more conservative (than A) to minimise contamination at the expense of the number of stars.

We note that the LE sample selection was defined in the  $v_x - v_y$  space using the same box selection as used in pseudo-Cartesian velocity space ( $\tilde{v}_x$ ,  $\tilde{v}_y$ ), which selects a general LE halo sample that is offset in  $v_y$  from Thamnos and in  $v_x$  from *Gaia* Enceladus.

The second step involved cleaning each sample further. For this, we followed Ruiz-Lara et al. (2022) in assigning each star in the pseudo velocity selections a radial velocity of the closest 6D star (in  $l$ ,  $b$ ,  $v_l$ ,  $v_b$ ) in the same sub-structure; namely, for cleaning the Sequoia sample, we assign to each 5D Sequoia star the radial velocity of the closest true 6D Sequoia star. To select the closest 6D star in  $l$ ,  $b$ , we used measurement of the distance between two points  $(l_1, b_1)$  and  $(l_2, b_2)$  as  $\cos(\delta) = \sin(b_1) \sin(b_2) + \cos(b_1) \cos(b_2) \cos(l_1 - l_2)$  and for small  $\delta$  values, we then assumed  $\delta^2 \approx 2(1 - \cos(\delta))$ , scaling this between 0 and 1 our metric for the angular distance on the sky is  $d_S^2 = (1 - \cos(\delta))/2$ . For the closest star in  $v_l$ ,  $v_b$ , we used the distance metric of  $d_V^2 = (v_{l1} - v_{l2})^2/\delta_{v_l}^2 + (v_{b1} - v_{b2})^2/\delta_{v_b}^2$ , where  $\delta_{v_{l/b}}$  are  $\max(v_{l/b}) - \min(v_{l/b})$  of the 6D sample, such that  $d_V^2$  is also scaled between 0 and 1 to give equal importance to the distance in velocity space. The total distance to each star is given by  $d_{\text{tot}}^2 = d_S^2 + d_V^2$  and minimising  $d_{\text{tot}}$  gives the nearest neighbour 6D star. We then assigned the radial velocity of this nearest 6D star and use this to estimate energy,  $E$ , and angular momenta ( $L_z$ ,  $L_\perp$ ) using the same potential described in Sect. 2.1.

We selected only those stars whose integrals of motion are within the ellipsoidal selection of the sub-structure already defined in 6D. By definition, any 6D sub-structure stars in the initial pseudo selection will remain in the sample after cleaning in this way but the majority of stars belonging to other sub-structures which overlap in pseudo-velocity space are removed. This method was used before in Ruiz-Lara et al. (2022) for the case of the Helmi streams.

### 2.3.2. Validation of 5D selected samples

**Table 1.** Sizes, completeness, and purity of the samples.

Sample	N in 5D	With $v_{\text{los}}$	Completeness	Purity
Thamnos A	1773	705	32%	53%
Thamnos B	1107	437	20%	53%
Sequoia A	498	165	41%	57%
Sequoia B	406	131	34%	59%

**Notes.** Samples are selected in 5D and the sizes are given within the fitting bundle (see Sect. 3.2). The purity is only a lower limit, see the text for details.

Table 1 presents the number of stars in the different sub-sets after first selecting in pseudo velocity space (as shown in Fig. 3), followed by cleaning in IoM space as described above and then selecting the stars that are within the CMD fitting bundle (described in Sect. 3.2 and shown in Fig. 2). Here, we see the benefit of going to 5D for the smaller sub-structures such as Sequoia, as we have at least three times more stars than if we took the bright 6D samples.

Along with the contamination already present in the 6D samples, due to the samples not being pure (described in Sect. 2.2), when we selected our samples in 5D, we had to take into account the fact that this also adds an extra level of contamination due to the missing line-of-sight velocities. In Table 1, we present lower limits on the completeness and purity for each sample when compared to the 6D halo in Dodd et al. (2023, extended with spectroscopic surveys; described previously in Sect. 2.2)

which we consider to be the ‘truth’. The completeness for a sub-structure (identified in 6D) was taken to be the ratio of the number of member stars of the sub-structure that fall in the pseudo velocity selection to its total number of (bright) member stars; this is equivalent to the percentage of (bright) member stars of the 6D sub-structure that fall within the pseudo velocity selections in Fig. 3. For example, for Sequoia A, the completeness is  $N_{\text{SeqA \& 6D Seq}}/N_{\text{6D Seq \& } M_G < 5}$ . The purity was calculated as the ratio of the number of member stars of the sub-structure that fall in the pseudo velocity selection to the total number of stars (with 6D information) that enter the pseudo velocity selection (e.g. for Sequoia A, the purity is  $N_{\text{SeqA \& 6D Seq}}/N_{\text{SeqA \& 6D total}}$ ).

The purity estimates can be interpreted as lower limits because of our restrictive cut in Mahalanobis distance of 2.13 to assign stars to sub-structures, which was chosen to reduce contamination. In addition, our 6D samples of the sub-structures, that we are defining as the ‘truth’, are not fully complete themselves. These sub-structures have been defined based on clustering in integrals of motion space and while we were not able to capture the full spread of the merger debris, it can be assumed that this represents the main over-density of the debris. In the local halo sample, only  $\sim 10\%$  of the stars are in a significant sub-structure, the remaining will contain accreted stars, including likely stars of GE, Thamnos and Sequoia.

For any given sub-structure, the number of 6D stars that belong to another sub-structure but have still been selected by our criteria using 5D is less than 1%. However, the 6D stars classified as general halo make up the remaining  $\sim 40\text{--}50\%$ . This explains why our selection makes the purity appear quite low. Most of these general halo stars that are added by our 5D method are slightly more extended in IoM space than the cut we used to define our 6D samples (e.g. see Fig. A.2). However, after an examination of the APOGEE DR17 (Accetta et al. 2022) abundances of the stars and the metallicity distributions with LAMOST DR7 LRS (Zhao et al. 2012), we can conclude that these stars are likely to be true members of the sub-structure (see Appendix A for more details) and our 5D selection only adds a small amount of extra contamination.

### 3. Methods

By fitting the absolute CMDs of our 5D selected samples, we obtained age and metallicity distributions of their stellar populations. We used an updated method of CMD fitting that has been tailored for use on *Gaia* data, ‘CMDft.*Gaia*’, the details of which are given in Gallart et al. (2024, with an overview given in their Fig. 2). This is a three-step process including 1) a synthetic CMD computation (using ChronoSynth, see Sect. 3.1), 2) then the error and completeness simulation (DisPar-*Gaia*, see Ruiz-Lara et al. 2022, Fernández-Alvar et al. 2025), and, finally, a search for the best-fit combination of simple stellar populations (SSPs) with DirSFH (a sophisticated update of previous CMD fitting software namely IAC-pop Aparicio & Hidalgo 2009 and TheStorm Bernard et al. 2018). This methodology has also been used to characterise the stellar content of the Helmi streams in Ruiz-Lara et al. (2022).

#### 3.1. CMD fitting procedure

We generated a synthetic mother CMD containing 101 million bright stars ( $M_G < 5$ ) and a flat distribution in age and metallicity ( $Z$ ) from 0.02 to 13.5 Gyr and 0.0001 to 0.032 (equivalently: [Fe/H] of  $-2.5$  to  $0.5$  dex) respectively. This CMD was computed

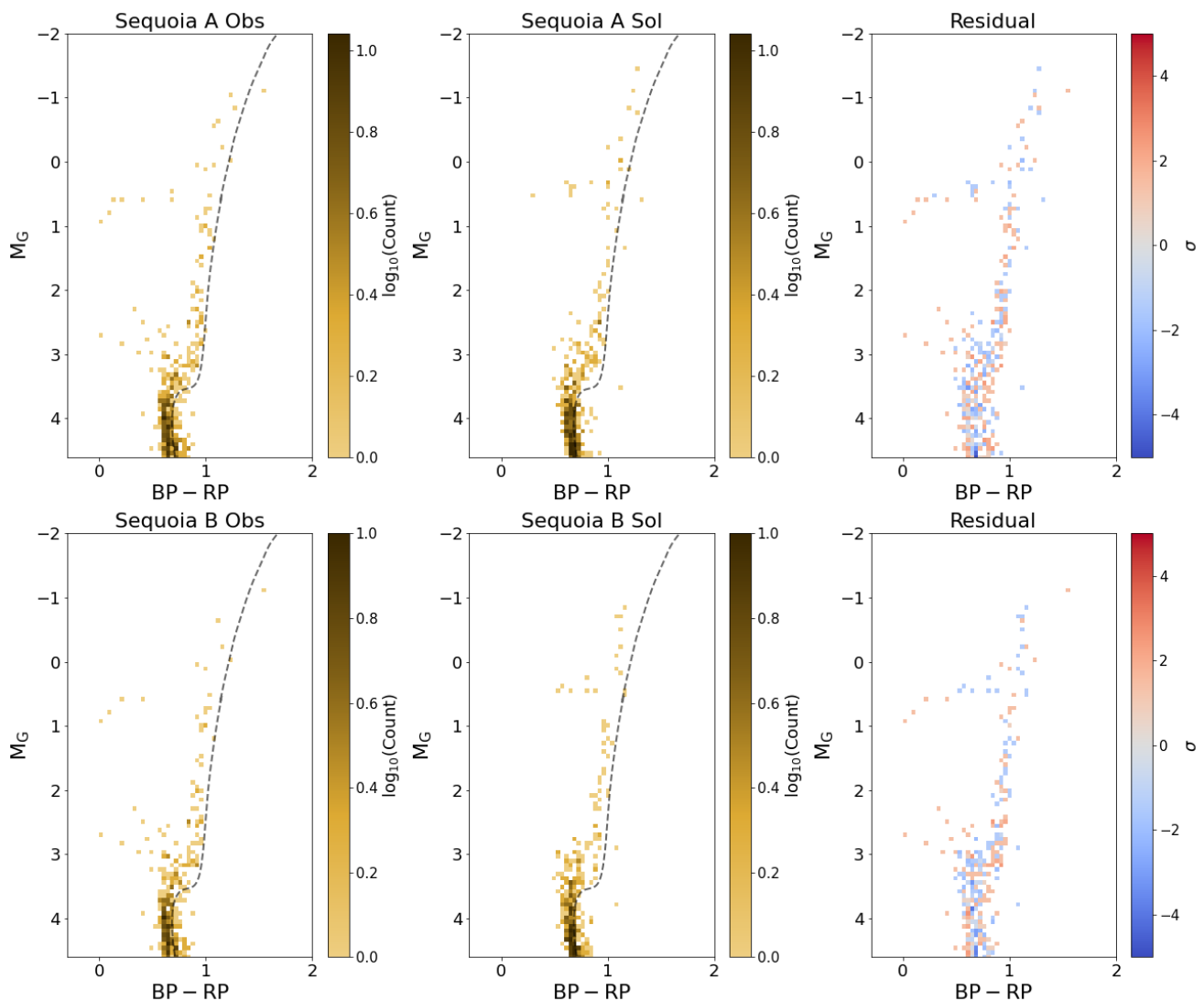
using the updated BaSTI-IAC alpha-enhanced models (Pietrinferni et al. 2021) with a Reimers mass loss parameter,  $\eta$ , of 0.3, a Kroupa et al. (1993) initial mass fraction, a minimum mass ratio for binaries of 0.1 and a fraction of unresolved binaries of 30%.

Our samples are defined within 2.5 kpc of the Sun and contain only stars brighter than  $M_G = 5$ , the magnitude limit for which *Gaia* is  $>99\%$  complete in this volume (Everall & Boubert 2022). However, in creating our samples we make various quality cuts (e.g. `parallax_over_error`  $\geq 5$ ) which can affect the distribution of stars in the CMD. We therefore must mimic all the same selections in our mother CMD by simulating the quality cuts. To make comparisons between the mother CMD and observations we must also simulate the observational errors. This is done with the step we call DisPar-*Gaia*. For more details of how this is done we refer to Sect. 3.1 of Ruiz-Lara et al. (2022) and the comprehensive description presented in Appendix A of Fernández-Alvar et al. (2025). This produces an error-convolved mother CMD that is affected by *Gaia* observational errors and selection effects in the same way as the observed sub-samples, allowing for direct comparisons once the possible systematic shift between data and models is also considered. The colour and magnitude shift applied to the mother CMD to match the observations is  $(\Delta_c, \Delta_m)$  where  $\Delta_c$ , the shift in  $(G_{\text{BP}} - G_{\text{RP}})$ , is taken to be  $-0.035$  and  $\Delta_m$ , the shift in  $M_G$ , is taken to be  $0.04$ . The details of how this shift is determined are given in Gallart et al. (2024) using the high-quality *Gaia* Catalogue of Nearby stars (Smart et al. 2021).

For fitting the observed CMD of our samples using the error-convolved mother CMD we then use DirSFH (Gallart et al. 2024). This extracts a series of SSPs from the mother CMD with typical size defined by an array of age and metallicity ( $Z$ ) seed points and finds the linear combinations of SSPs which best fit the observed CMD. The spacing of the seed points has been chosen to reflect the varying age and metallicity resolutions towards older ages and more metal-poor populations. We perform many realisations of the fit using a Dirichlet tessellation. In each realisation, the seed points are shifted in age-metallicity space. The final solution is the average best fit of all the realisations.

For this work, we employ the ‘weighted’ strategy within DirSFH which uses the logarithm of the inverse of the variance of ages across the mother CMD to weight the fitting. This allows for more importance to be given to regions of the CMD that have more age resolution (see Sect. 3.3.2 in Gallart et al. 2024). Through extensive testing, it was shown by Ruiz-Lara et al. (2022) and Gallart et al. (2024) that the solutions remain robust to the choice of weighting, age-metallicity seed points and differences in the mother CMD (see also Gallart et al. 2024, as well as Sect. 3.4 of this work for an extensive testing of the method).

The fitting of the CMD is done using the stars within a single defined bundle, which can be seen in Fig. 2. It was defined to encompass all of the stars brighter than  $M_G \sim 4.5$ ; namely, chosen to extend  $\sim 1$  magnitude below the old main sequence turn-off. At magnitudes fainter than this, most of the age information is lost. Since the number of stars increases for fainter magnitudes, they will also dominate in number and, consequently, would be given a lot of importance for fitting, adding more noise than signal, and making the fit less reliable (see Sect. 4.3.2 in Gallart et al. 2024, for more details). The vertices of this bundle are given in Table B.1 in Appendix B.



**Fig. 4.** Absolute CMDs of the observed 5D Sequoia samples (first column). The top row shows Sequoia A and the bottom row shows the stricter selection Sequoia B. The middle column shows the best fit solution CMD and the final column shows the residual in terms of  $\sigma$ . The black dashed line in the first two columns corresponds to an isochrone (BaSTI-IAC alpha-enhanced with an age of 11.6 Gyr and  $[\text{Fe}/\text{H}]$  of  $-0.8$  dex) that separates the blue and red halo sequences.

### 3.2. Fitting the observed CMDs of Sequoia and Thamnos

For each observed sample, we computed 100 solutions of the CMD fitting, such that each time the grid of age-metallicity seeds (see Table B.2 in Appendix B for starting seeds) are slightly modified. DirSFH provides the star formation history (SFH) of the population and the solution CMD (obtained by sampling from the mother CMD according to the best fit SFH, matching the number of stars in the observed CMD within the bundle). Each star in the solution CMD therefore has an associated age and metallicity. Here, we used these best-fit CMDs and the associated age and metallicity distributions to infer differences between the halo sub-structures. We combined the 100 solutions in a weighted average using the goodness of fit as weights. The weighted variance gives our uncertainty on the best fit CMD and on the derived ages, metallicities, and SFHs. In our fits, we find that the weights are very similar across all 100 realisations, resulting in the weighted mean and weighted variance being almost identical to the standard mean and variance.

Fitting the CMDs of our samples allows us to obtain age and metallicity distributions for our 5D selections of Sequoia and

Thamnos. The CMDs of Sequoia A and B contain 498 and 406 stars within the bundle respectively. For Thamnos A and B, there are 1773 and 1107 stars within the fitting bundle respectively. The solution CMDs for each sample therefore also contain approximately this number of stars. It was shown in Gallart et al. (2024) that this method can achieve very precise ages ( $\leq 10\%$ ) for single stellar populations, even with  $\sim 350$  stars, a similar number to our smallest sample. For our samples, we are considering populations with an extended age and metallicity distribution and not a single stellar population so we also check with mock tests that the age and metallicity recovery is precise with this number of stars (see Sect. 3.4 for more details).

In Fig. 4 for Sequoia and Fig. 5 for Thamnos, the observed and best fit CMDs for each sample are shown, in the left and middle panels respectively. The residuals are shown on the right panels in terms of  $\sigma = (\text{obs} - \text{sol}) / \sqrt{(\text{obs} + \text{sol})/2}$  where obs and sol are the observed and best fit solution CMD respectively. The residuals are small ( $< \pm 2\sigma$ ) demonstrating that we obtain a good fit for each observed CMD.

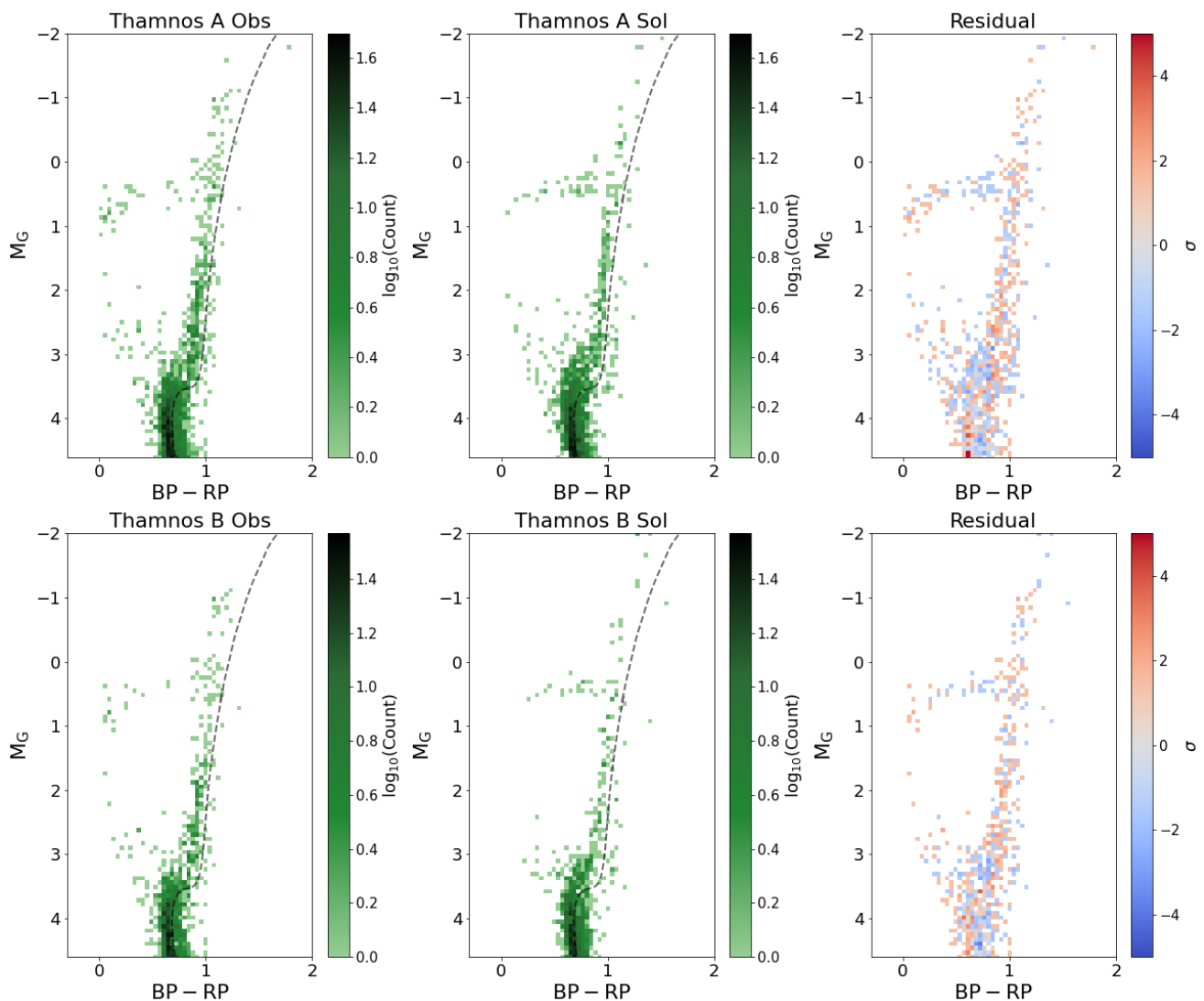


Fig. 5. Same as Fig. 4, but for Thamnos.

We can also validate our 5D samples by comparing their observed CMDs (first column of Figs 4 & 5) to the 6D samples shown in Fig. 2, confirming that the 5D are representative of the 6D samples. We see that like in 6D, our 5D samples show that Sequoia and Thamnos exhibit an extended horizontal branch with higher relative contributions of blue HB stars (first column of Figs 4 & 5) than seen in GE and LE samples (Fig. 6). However, this signature is not as well reproduced in our solution CMDs (middle column of Figs 4 & 5) likely due to low numbers of stars.

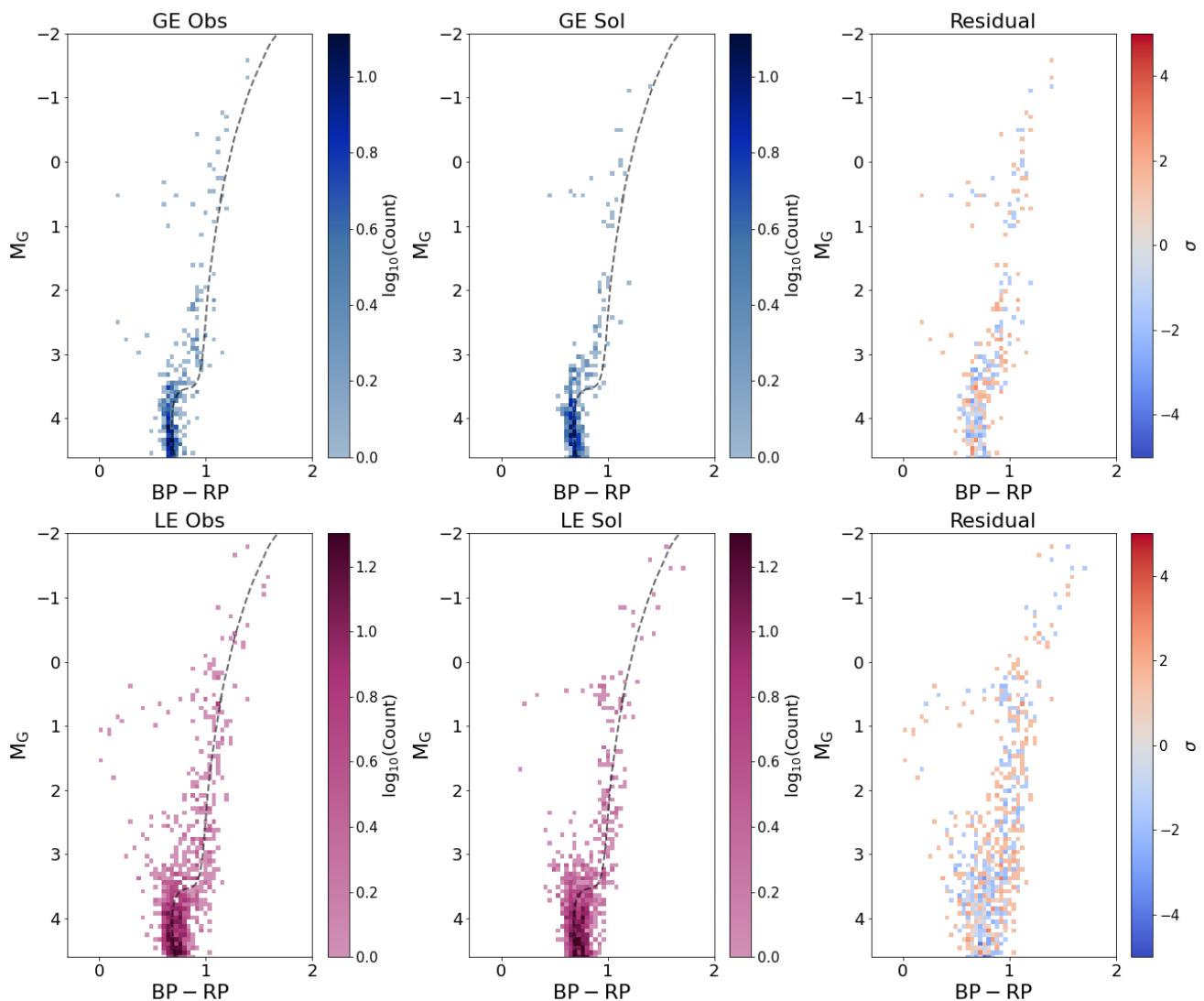
HB stars are low mass ( $\sim 0.5 - 0.8 M_{\odot}$ ) core helium-burning stars and their presence is indicative of an old stellar population. The parameters that determine the morphology of a stellar population's HB have been a long-standing question, with studies of globular clusters (excellent laboratories for stellar evolution) providing evidence (observational and theoretical) suggesting that the main parameter is metallicity, the second main parameter is age and then there is also a strong dependence on the He content (e.g. Gratton et al. 2010). The more metal-poor a stellar population the bluer the HB, the older the population the bluer the HB and the more He enhanced a population the bluer the HB, with the bluest stars forming a tail to lower luminosities. There is not a one-to-one correlation between the HB

morphology and the  $[\text{Fe}/\text{H}]$  or age of the population, however, it is still interesting to note that these populations have extended blue HBs (see Fig. 2).

### 3.3. Fitting the CMDs of contaminants

Both Sequoia and Thamnos samples present contamination from *Gaia* Enceladus debris, and Thamnos also has the additional contamination coming from possibly the in situ halo. As discussed earlier we have represented all of the contamination in Thamnos with our LE sample. Fitting the CMDs of *Gaia* Enceladus and the LE samples will allow us to compare how their contamination presents itself in age-metallicity space.

The 5D selected GE and LE samples contain 4602 (1903 6D) and 6814 (2438 6D) stars respectively. Since our contaminant populations are significantly larger in the number of stars than Thamnos and Sequoia samples, to be able to compare directly the age-metallicity distributions obtained from CMD fitting, we take random sub-samples matching the number of stars of the relevant sample, within the bundle used for fitting. We choose to match the number of stars within the bundle for the Sequoia B and Thamnos B samples, our more restrictive selections with  $N_S = 406$  and  $N_T = 1107$  stars respectively. That is, we take 20



**Fig. 6.** Same as Fig. 4, but for a random sub-sample of the GE and LE samples of sizes  $N_S$  and  $N_T$ , respectively.

random sub-samples of *Gaia* Enceladus each of size  $N_S$  stars and 20 random sub-samples of the LE group each of size  $N_T$  stars. As well as having a comparable number of stars for fitting this also allows us to assess any variations in the solutions from sub-sampling. The resulting age and metallicity space for our Sequoia and Thamnos samples can now be compared to our *Gaia* Enceladus and LE sub-samples given their similar number of stars and can be used to identify which signatures in age-metallicity space are associated with true Thamnos/Sequoia populations and which are coming from contamination.

Fig. 6 shows the best fit CMDs with the black dashed line, in the first two columns, corresponding to an isochrone (BaSTI-IAC alpha-enhanced with an age of 11.6 Gyr and  $[\text{Fe}/\text{H}]$  of  $-0.8$  dex) that separates the blue and red halo sequences visually. Here we can see that *Gaia* Enceladus populates the blue halo sequence and the LE population contains stars across both sequences but predominantly the red halo sequence, in agreement that this sample is dominated by in situ stars (Babusiaux et al. 2018).

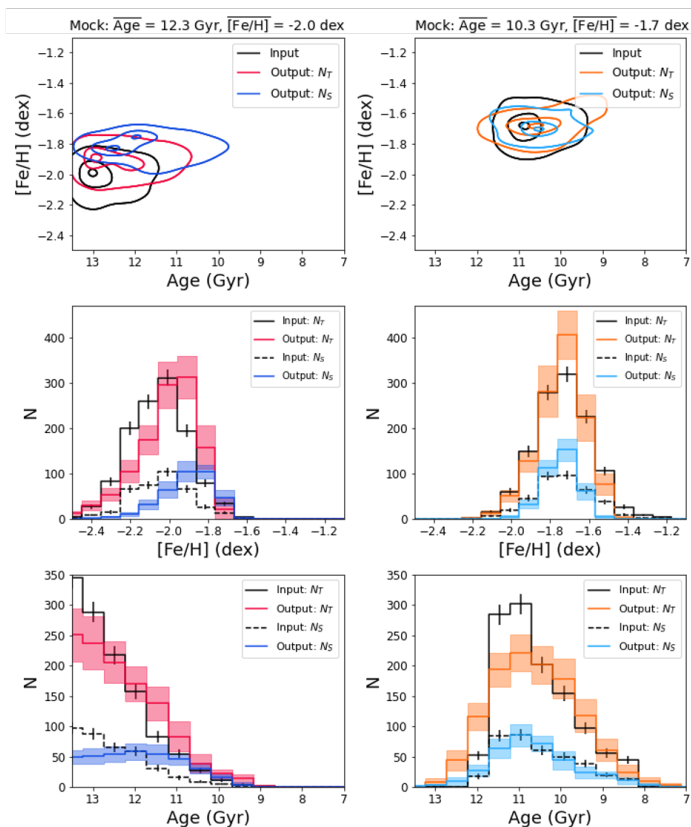
### 3.4. Mock tests

To interpret the outputs of the CMD fitting technique, we test the recovery of the age-metallicity distributions of old and metal-

poor samples. To create suitable mock populations, we use a mother CMD similar to that used for the CMD fitting as described in Sect. 3.1. However, since we want to focus on the metal-poor and old regime, we generate more stars here - this mother CMD has an age range of 8 – 13.5 Gyr, and a metallicity ( $[\text{Fe}/\text{H}]$ ) range of  $-0.5$  to  $-2.5$  dex and contains around 30 million stars.

We took sub-samples of this mother CMD following a specific age-metallicity distribution that represents old metal-poor populations like the Sequoia and Thamnos signal. We do this for samples of the same size as Thamnos B and Sequoia B to see how the number of stars affects the recovery of the age and metallicity information. These tests complement those on single stellar populations by Gallart et al. (2024) as here we focus on the recovery for extended age and metallicity distributions, concentrating on old ages and low metallicities which is the region of interest for this study.

We selected stars according to a skewed bi-variate normal distribution (eq 7), to be able to represent the distributions of ages and metallicities realistically. The age distributions are modelled with a high negative skew such that we reproduce the theoretical age distribution for an old accreted population, con-



**Fig. 7.** Age and metallicity distributions of the input mocks (black) compared to the output of DirSFH. The first column shows the old and metal-poor mock peaking at 12.3 Gyr and -2.0 dex, with blue being the smaller sample ( $N_S$  stars) than red ( $N_T$  stars). The second column shows the younger, slightly more metal-rich mock (mean age of 10.3 Gyr and  $[\text{Fe}/\text{H}]$  of -1.7 dex) with blue corresponding to the smaller sample of size  $N_S$  and orange the larger sample of  $N_T$  stars. The top row shows the contours of the 2D distributions containing the percentage of stars within 1, 2 and 3  $\sigma$  levels in 2D. The middle row shows the 1D metallicity distributions and the bottom row the age distributions, where the errors come from averaging over the 100 solutions produced by DirSFH. In the bottom two rows, the input for the samples of size  $N_S$  are shown with dashed lines to distinguish between distributions. Errors on the input distributions, plotted as black error bars, are taken as  $\sqrt{N}$  per bin.

taining old stars before a decline or cut-off at accretion. The metallicity distributions are not given any skew.

The form of this distribution is given by:

$$f(\mathbf{x} | \boldsymbol{\mu}, \boldsymbol{\Sigma}, \alpha_{\text{age}}) = \frac{2}{2\pi \sqrt{|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \cdot \Phi\left(\alpha_{\text{age}} \frac{\text{age} - \mu_{\text{age}}}{\sqrt{\sigma_{\text{age}}^2}}\right), \quad (7)$$

where  $\mathbf{x} = \begin{pmatrix} \text{age} \\ [\text{Fe}/\text{H}] \end{pmatrix}$ ,  $\boldsymbol{\mu} = \begin{pmatrix} \overline{\text{age}} \\ \overline{[\text{Fe}/\text{H}]} \end{pmatrix}$ ,  $\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{\text{age}}^2 & 0 \\ 0 & \sigma_{[\text{Fe}/\text{H}]}^2 \end{pmatrix}$ ,  $\Phi(\cdot)$  is the CDF (cumulative distribution function) of the standard normal distribution and  $\alpha_{\text{age}}$  represents the skewness in the age distribution.

We make two mocks, one with a  $\boldsymbol{\mu}$  of (13.5 Gyr, -2.0 dex), and a younger more metal-rich mock with a  $\boldsymbol{\mu}$  of (11.5 Gyr, -1.7 dex). For both we use a age spread ( $\sigma_{\text{age}}$ ) of 2.5 Gyr and metallicity spread ( $\sigma_{[\text{Fe}/\text{H}]}$ ) of 0.3 dex. With an age skew  $\alpha_{\text{age}}$  of -10, this corresponds to age distributions with mean ages of 12.3 Gyr and 10.3 Gyr, respectively. We then perform the DisPar-Gaia step to

simulate observational errors in the CMDs of our toy mocks and run DirSFH to obtain age and metallicity distributions.

We can see a comparison of the input and outputs in age and metallicity space in Fig. 7. We see that for the younger, more metal-rich mock (right column of Fig. 7) the recovery is better for both sample sizes. For the old and metal-poor mock (left column of Fig. 7) the peaks of the distributions are shifted to slightly higher metallicities and younger ages. This is especially apparent for the smaller sample size (blue in Fig. 7), where the shift is more significant.

These experiments (as well as similar tests not quoted here) are reassuring, although we find that a small sample ( $N_S = 406$ ) is more affected by shifts and an old distribution near the edge of the grid (13.5 Gyr in age) is more affected, making the age distribution less steep and shifting the median to younger ages. In all cases, any shift in age and metallicity is always in the direction of younger ages and higher metallicities, with maximum shifts of the mean age on the order of 1 Gyr and of the mean  $[\text{Fe}/\text{H}]$  up to  $\sim 0.15$  dex. Any effect from being near the edge of the grid that results in a shift to younger ages in the output is always accompanied by a shift to high metallicities as a compensation.

## 4. Results

### 4.1. Age-metallicity space of best fit CMDs

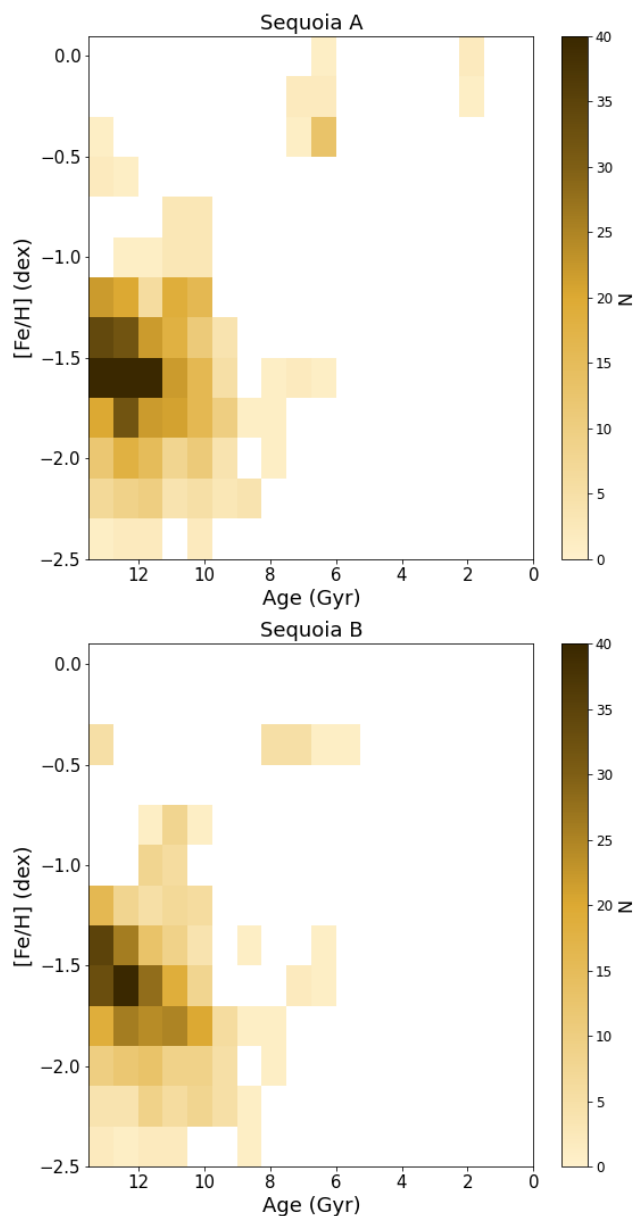
We go on to examine the age-metallicity space for our best fit CMD stars of the different samples of Sequoia in equally spaced age bins of 0.75 Gyr and 0.2 dex in metallicity. For these analyses, we used larger bins than the age-metallicity seeds used for fitting the observed CMD, chosen to have enough signal per bin.

Fig. 8 shows the age-metallicity distribution for Sequoia A and B. We can see a dominant population of stars around  $[\text{Fe}/\text{H}] = -1.5$  with ages  $> 10$  Gyr, and a tail to lower metallicities and younger ages. Sequoia A and B (the more restrictive selection) present very similar distributions.

Fig. 9 shows the age-metallicity space for Thamnos A and B, where Thamnos B is the more conservative selection. In both Thamnos A and B, there is a significant fraction of stars at very old ages ( $> 12$  Gyr) from  $[\text{Fe}/\text{H}] \sim -1.4$  dex, extending down to  $[\text{Fe}/\text{H}] = -2.0$  dex. We can already see differences when comparing to the results of Sequoia in Fig. 8; for instance, in our Thamnos selections, there is a higher fraction of stars that are older than 12 Gyr.

However, due to contamination, it is difficult to interpret these differences. This is why we now also characterise in age-metallicity space the contamination that could be present. In Fig. 10 we show the mean age-metallicity distributions from fitting the CMDs of the 20 GE sub-samples (left panel) and that of the 20 LE sub-samples in the right panel. Here, the total number of stars in the LE sub-sample is  $N_T$ , the same as the number of stars in Thamnos B and in the GE sub-sample is  $N_S$ , the same as the number of stars in Sequoia B. *Gaia* Enceladus presents a well-defined age-metallicity relation extending from  $[\text{Fe}/\text{H}] \sim -2$  dex to  $-1$  dex across 13.5 – 10 Gyr in age (see the left panel of Fig. 10). For the LE sample, shown in the right panel of Fig. 10, we can see that it is dominated by very old stars  $> 12$  Gyr, across a wide range of metallicities from  $[\text{Fe}/\text{H}] \sim -1.8$  to  $-0.5$  dex, and peaking around  $[\text{Fe}/\text{H}] \sim -0.8$ . There is a significant tail of stars going down to  $[\text{Fe}/\text{H}] = -2$  dex with ages  $> 12$  Gyr old, which could be contributing to the signal in Thamnos.

A visual comparison of the age-met distributions for Sequoia and Thamnos with those of the expected contamination is al-

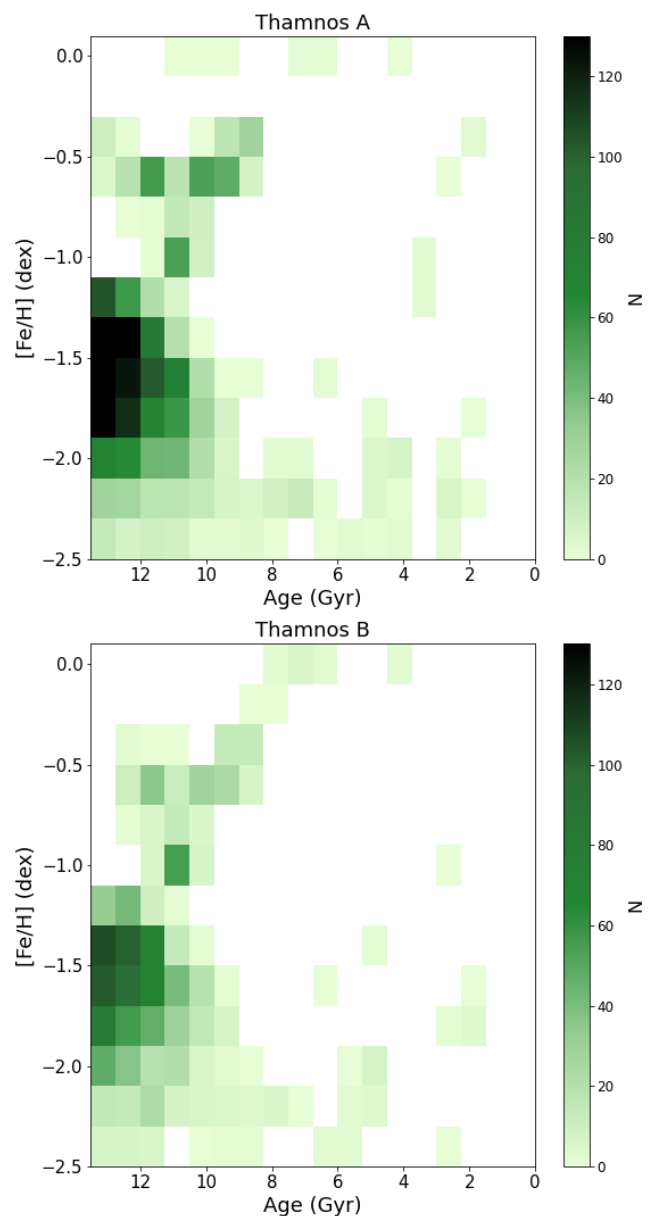


**Fig. 8.** Age-metallicity space of best fit CMD stars for Sequoia A and B (with B being the more conservative selection).

ready informative. We can see that especially Sequoia (and to some extent Thamnos) contains a significant fraction of stars that seem to follow the *Gaia* Enceladus age-metallicity sequence. We can also see that Thamnos has a higher contribution of old stars  $> 12$  Gyr, similar to the LE population, but only at lower metallicities. Nonetheless, both Sequoia and Thamnos show a larger fraction of stars at lower metallicities than in GE and the LE sample. This is the signal we are interested in.

#### 4.2. Comparison of the age-metallicity distributions

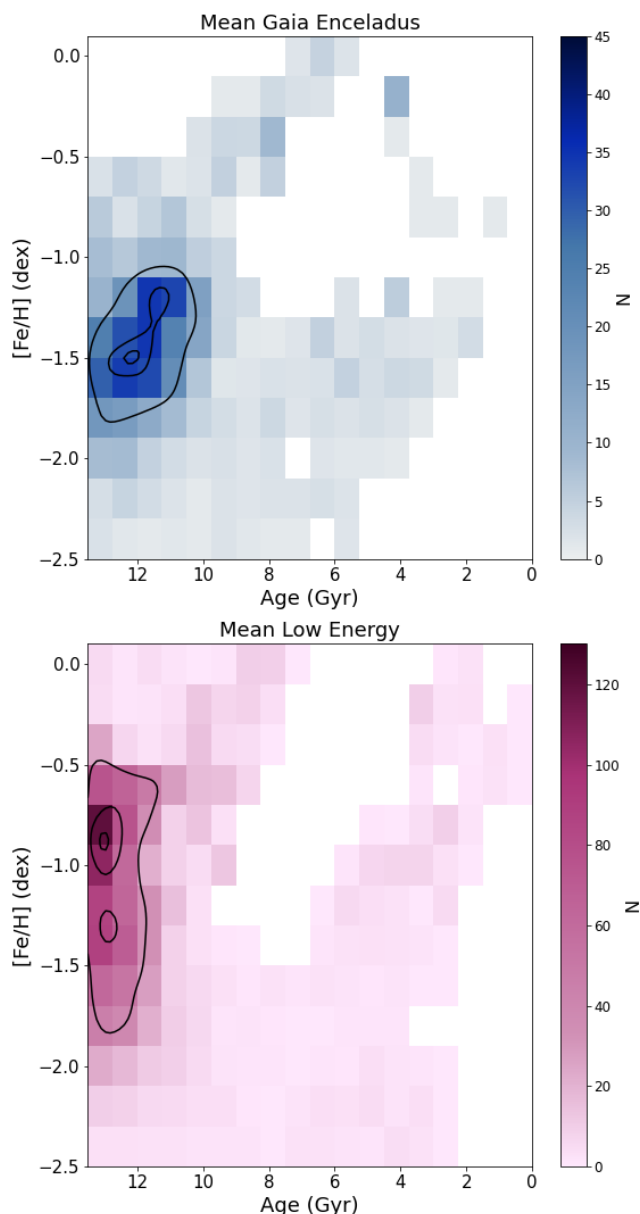
To differentiate the signatures coming from contamination and the signal of Sequoia and Thamnos we subtract the mean counts in the age-metallicity space of GE from Sequoia (Fig. 11) and of the LE sample from Thamnos (Fig. 12). We can subtract our mean counts of LE from Thamnos B directly since the samples contain the same number of stars by definition,  $N_T$ . The same is true for subtracting the mean counts of GE from Sequoia B,



**Fig. 9.** Age-metallicity space of best fit CMD stars for Thamnos A and B (B being the more conservative selection).

both have  $N_S$  stars by definition. For comparison to Sequoia A or Thamnos A which have a different number of stars, we first introduce a normalisation. In this way, we can directly compare sub-structures and see what fraction of the stars in the samples dominate in which region of the age-metallicity space. We scaled the difference in counts with the combined errors to assess the impact of variance per bin. For GE and LE, we took the errors,  $\sigma_{GE}$  and  $\sigma_{LE}$ , to be the standard deviation of counts in the age-metallicity bins across all 20 sub-samples. This measures the variation of the solutions and is larger than the Poisson noise of the mean GE or LE of each sub-sample.

For Sequoia, the residual age-metallicity space after subtracting GE and dividing by the error ( $\sigma$  is taken to be  $\sqrt{\sigma_{GE}^2 + \sigma_S^2}$  where  $\sigma_S$  is the Poisson noise on the counts for Sequoia) is shown in Fig. 11. The colour bar gives a measure of the excess of stars in the sample that are not present in the contamination. We see in this figure that there is an excess of stars



**Fig. 10.** Age-metallicity space showing the mean of the best fit CMDs for 20 GE (left) and 20 LE (right) sub-samples. The colour bar shows the count. Contours are shown containing 39%, 87% and 99% of the counts (corresponding to 1, 2, and 3  $\sigma$  in 2D, respectively).

in Sequoia at low metallicity and up to metallicities of  $-1.5$  dex, with ages greater than  $\sim 9$ – $10$  Gyr.

Fig. 12 shows the residual age-metallicity space after subtracting the mean count of the LE sub-samples from Thamnos in each bin and dividing by the error. Here the error  $\sigma$  is taken to be  $\sqrt{\sigma_{LE}^2 + \sigma_T^2}$  where  $\sigma_T$  is the Poisson noise on the counts for Thamnos. Now we can see that the Thamnos region has an excess of metal-poor stars  $[Fe/H] < -1.5$  dex at old ages ( $> 10$  Gyr) when compared with the contamination.

These figures reveal more clearly that there are subtle differences between Sequoia and Thamnos in their age-metallicity distributions, providing insights into their formation histories and hence possibly on the timing of these mergers. Sequoia reaches similar metallicities as Thamnos ( $[Fe/H]$  of  $-1.5$  dex) but over a slightly more extended period of time. We will see this more

clearly in Sect. 4.4. However, we should bear in mind that this slightly more extended age distribution for Sequoia could be an effect of having a low number of stars, which acts in the direction of spreading towards higher metallicities and younger ages, as discussed in Sect. 3.4. Our findings suggest that Sequoia and Thamnos formed their stars over a similar duration to *Gaia* Enceladus, without enriching to as high metallicities. This can be explained by Sequoia being a smaller progenitor galaxy and Thamnos even smaller still (considering also that Thamnos has a smaller extent in IoM space than Sequoia).

### 4.3. Comparing MDFs of the best-fit CMDs

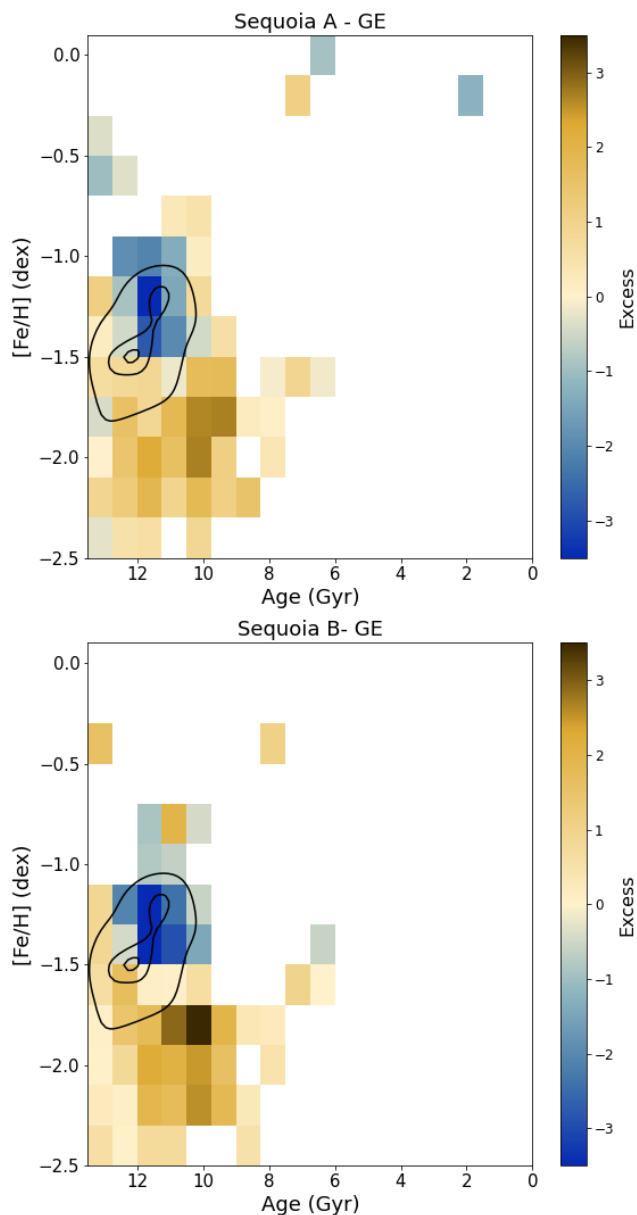
In Fig. 13, we compare the metallicity distributions of the solution CMD stars for the different sub-samples. This figure shows the metallicity distributions derived from fitting the CMDs of our samples for Thamnos, Sequoia, GE and LE. The MDFs of Sequoia and Thamnos are obtained by projecting Figs. 8 & 9 onto their y-axis; namely, they still contain the contamination from GE and LE. The MDFs presented in Fig. 13 are binned according to the metallicity seeds used for fitting the CMDs, see Sect. 3.2 and Table B.2.

We see that the shape of the MDFs derived from fitting the CMDs of our samples closely matches the spectroscopic (LAMOST LRS DR7) MDFs of the samples, which are shown in red in Fig. 13 for comparison. The photometric MDFs are derived with metallicities coming from BaSTI-IAC models and appear to have an offset compared to the spectroscopic metallicities shifting the peaks of the distributions to slightly lower metallicities. We can see visually that there is a shift in the  $[Fe/H]$  scale of our photometrically derived metallicities of  $\sim 0.1$  –  $0.2$  dex to lower metallicities. For the Sequoia sample, this is less clear due to the lower number of stars. Despite the offset, the MDFs derived from CMD fitting do recover the same shape and relative distributions as the spectroscopic metallicities.

Sequoia presents an MDF whose peak is shifted to lower metallicities than *Gaia* Enceladus. The same can be said for Thamnos, with some hints at a more prominent metal-poor tail below  $[Fe/H] \sim -2$ . The LE group presents two main peaks, a large contribution from (hot) thick disc stars with high metallicity ( $[Fe/H] \sim -1.0$  to  $-0.5$ ) and a second more metal-poor peak similar to that of GE debris. It also presents a significant metal-poor tail, as seen in the spectroscopic metallicities, which is the superposition of the metal-poor in situ halo component and metal-poor tail of GE.

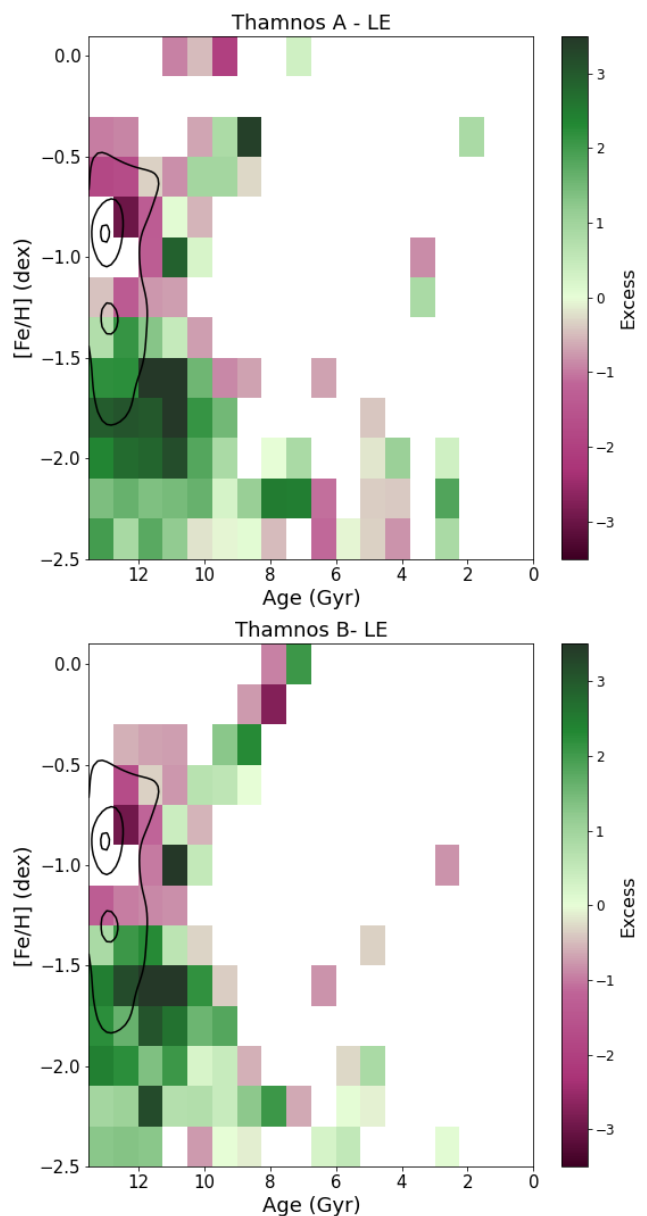
To establish the differences between the MDFs and where they occur in a quantitative way, we proceeded as follows; we compared the MDFs obtained from the best fit CMD of GE, including the 20 sub-samples, with one another and with Sequoia. We did the same for Thamnos compared with the LE samples. To compare two MDFs, we used the two-sample Kolmogorov–Smirnov (KS) test. Figure 14 (top panel) shows the KS statistic (including the sign of the difference in the distributions) and the  $[Fe/H]$  value at which the maximum difference occurs. Here, the yellow points represent Sequoia A (triangle markers) and B (circle markers) sample MDFs compared with GE and the 20 GE sub-samples. Green represents Thamnos A (triangle markers) and B (circle markers) compared with LE and the 20 LE sub-samples. Blue points show the GE MDFs compared with one another. Pink points show the LE MDFs compared with one another. This allows us to quantify how much variation comes from random sub-samples of GE and LE, and to check that the differences we measure for Sequoia and Thamnos with the GE and LE samples are significant. The fact that the difference in cu-

mulative metallicity distributions of Sequoia and Thamnos with GE and LE respectively, is always positive tells us that there is an excess in these sub-structures at these metallicities, while the sub-samples compared with themselves have smaller variations with positive and negative sign, which is consistent with being driven by noise.



**Fig. 11.** Residual age-metallicity space of the GE CMD stars subtracted from those of Sequoia A and B samples (B being the more conservative selection). The colour bar is the difference in the number of stars per bin divided by the combined errors on the count. The contours of the contamination (GE) age-metallicity distribution are shown for visual comparison.

Looking at the results of the KS tests (shown in Fig. 14) we can see that the largest differences between GE and Sequoia are present at low metallicity. We see a clear difference in the MDFs around  $[\text{Fe}/\text{H}]$  of  $-1.5$  to  $-1.6$  and this is slightly more significant in Sequoia B which is the stricter sample with lower contamination. The mean difference for Sequoia (yellow dashed line in Fig. 14) compared with GE is higher than the differences in the MDFs produced by sub-sampling from *Gaia* Enceladus, mean-

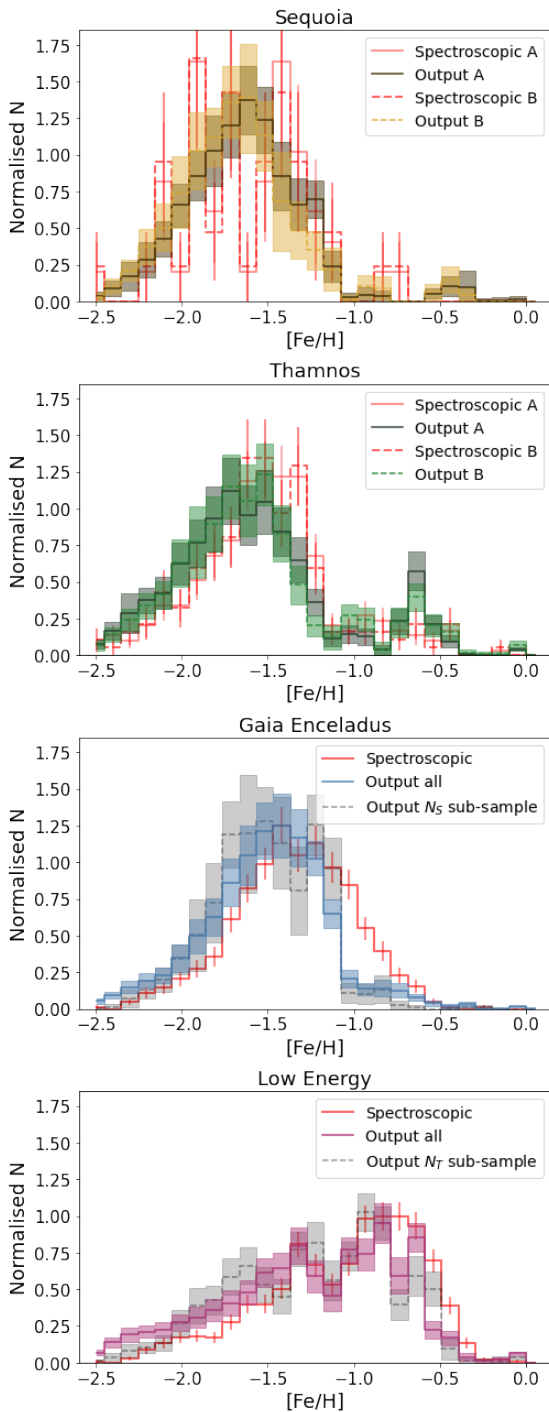


**Fig. 12.** Same as Fig. 11, but for Thamnos. The residual here is calculated by subtracting the LE sample from Thamnos and the contours are that of the LE age-metallicity distribution, shown for visual comparison.

ing that we cannot explain the Sequoia MDF as just a random sub-sampling of *Gaia* Enceladus. For Thamnos, in comparison with the LE sub-samples, the maximum difference is of a larger amplitude, since the metallicity distributions become very different as we go to higher metallicities and this maximum difference occurs at  $[\text{Fe}/\text{H}] \sim -1.4$  dex.

Since the  $[\text{Fe}/\text{H}]$  of the maximum difference in the distributions is not the most informative for detecting the differences in the metal-poor tails, the bottom panel of Fig. 14 instead shows the  $[\text{Fe}/\text{H}]$  at which the sample's cumulative metallicity distributions become significantly different at the 95% level. Since the value of the difference in the distributions,  $\Delta(\text{CDFs})$ , equal to 95% significance, depends on the sizes of the samples<sup>6</sup>; then,

<sup>6</sup> The critical value for the two sample KS test of sizes  $n_1, n_2$  can be approximated as  $c(\alpha) \sqrt{(n_1 + n_2)/n_1 n_2}$  with 95% significance corresponding to  $c(0.05) = 1.36$



**Fig. 13.** Metallicity distributions of the solution CMD stars for the different selections, binned according to the metallicity seeds used for fitting. The solid line shows the weighted mean count per metallicity bin and the shading indicates  $1\sigma$  across 100 iterations. For GE (LE) the output of fitting all stars in the parent sample is shown and one random  $N_5$  ( $N_7$ ) sub-sample. The spectroscopic metallicity distributions (using LAMOST LRS DR7), of the 5D selected samples used for CMD fitting, are also shown here in red. Here, the error bar indicates the Poisson noise ( $\sqrt{N}$ ) in each bin.

in the bottom panel of Fig. 14, samples of the same sizes are at a fixed  $y$  value. For Sequoia, the statistical difference occurs already at  $[\text{Fe}/\text{H}] \sim -2.0$  dex, and again this signal is always a positive excess in Sequoia. For Thamnos, the metal-poor tail of the distribution is even more important, as Thamnos does not

contain the metal-rich stars that are present in the LE sample. From Fig. 14, we see that the Thamnos and LE samples are already statistically different at  $[\text{Fe}/\text{H}] \lesssim -2.1$  dex and this signal is always an excess in Thamnos, confirming the distinct metal-poor tail. The GE/LE samples compared with one another can produce a difference that is significant; however, the sign of the difference is random (consistent with noise) and is typically at higher metallicities.

We had already seen (in Fig. 1 and Sect. 2.2) similar differences for the spectroscopic metallicity distributions and now they have been derived for the 5D samples and using our CMD fitting techniques. This provides confidence in the methodology and the soundness of the results obtained.

#### 4.4. Comparing age distributions of best fit CMDs

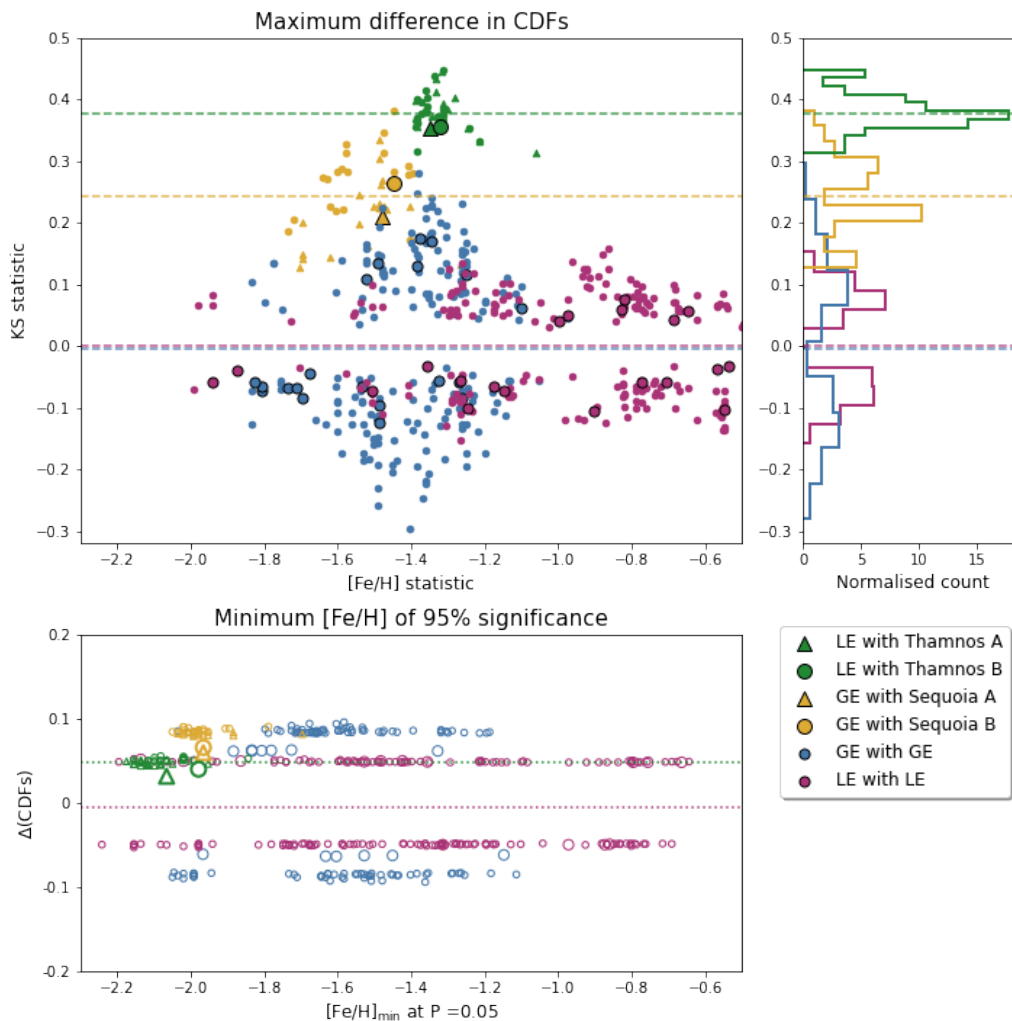
Figure 15 shows the age distributions of the best fit samples. For Sequoia and Thamnos in the first two rows of Fig. 15, the left panel shows the age distribution for all stars in the sample and the right panel shows stars below a metallicity threshold. This metallicity threshold is chosen to investigate the age distributions of stars where the signal from Thamnos and Sequoia is dominant over contamination; namely, where there is a positive excess in Figs. 11 and 12, and also motivated by where the KS tests show maximum differences with GE and LE (see Fig. 14).

From the full age distributions (see left column of Fig. 15) we can already see that both Sequoia and Thamnos are made up of very old stars (mostly  $> 10$  Gyr). Looking at the age distribution below  $[\text{Fe}/\text{H}] = -1.5$  dex (right column of Fig. 15) for these sub-structures, we see that for both Sequoia and Thamnos the metallicity cut has little effect on the age distributions. For Thamnos the age distribution is steep with a faster decline than Sequoia.

We also show the age distribution of all stars in the contaminating samples; GE and LE (bottom row of Fig. 15). The first thing to note is that all of Thamnos, Sequoia, and GE have more extended age distributions than the LE population which is very old, containing mostly stars older than 12 Gyr (possibly even older given the edge effects, see Sect. 3.4). The GE sample shows an approximately flat age distribution up to  $\sim 10$  Gyr with a sharp truncation. This truncation is different from that seen in Sequoia, which presents a more gradual decline to slightly younger ages than GE.

For GE, the age-metallicity distribution in 2D (left panel of Fig. 10) shows two peaks, one at  $\sim 12.5$  Gyr and  $[\text{Fe}/\text{H}]$  of  $-1.5$ , and the other at  $\sim 11$  Gyr and  $[\text{Fe}/\text{H}] \sim -1.3$  dex. These two peaks possibly indicate two starbursts in the GE progenitor and could be related to the accretion process (e.g. as the progenitor reaches the first pericentric passage, a burst of star formation could be triggered). We do not see these two bursts in our 1D age distribution, however, since these are the projected ages across all metallicities.

In a more quantitative sense, looking at the cumulative age distributions, we can estimate the lookback time at which the system formed half of its stars. Sequoia A (B) formed half of its stars by a lookback time of  $11.9^{+0.2}_{-0.3}$  Gyr ( $12.0^{+0.3}_{-0.3}$  Gyr). Thamnos A (B) is on average older, does not display a flat age distribution but declines faster, having formed half its stars at  $12.4^{+0.3}_{-0.3}$  Gyr ( $12.3^{+0.3}_{-0.3}$ ). For both Sequoia and Thamnos samples, we quote here the time at which the cumulative age distributions, with the  $[\text{Fe}/\text{H}]$  cut of  $-1.5$  dex applied, reach 0.5, but we find consistent values without the  $[\text{Fe}/\text{H}]$  cut. For the GE and LE samples, they formed half of their stars by a lookback time  $12.1^{+0.1}_{-0.1}$  Gyr and



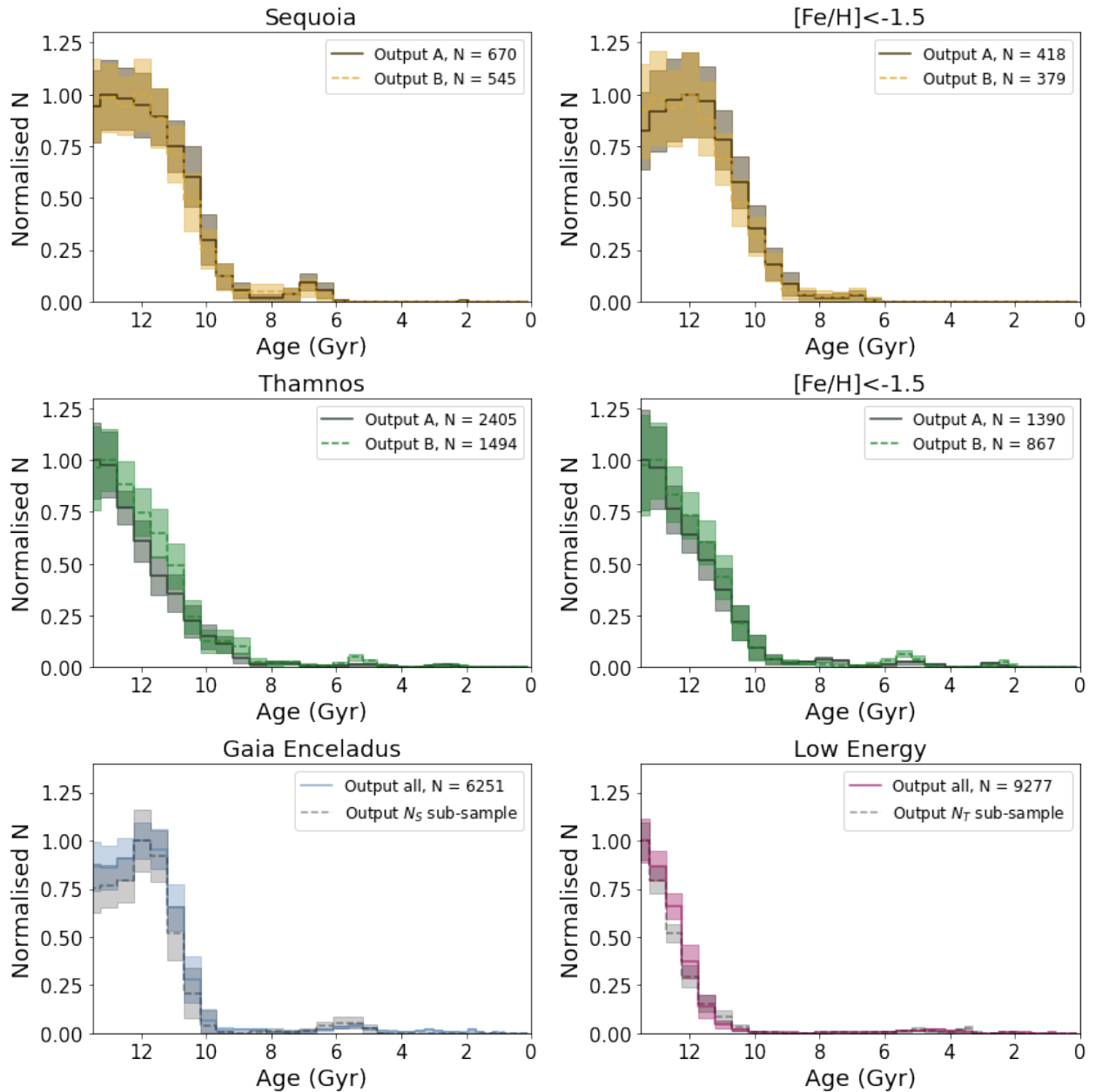
**Fig. 14.** Top-left panel: KS statistic (maximum difference) and metallicity of the maximum difference for the comparison of best fit CMD metallicity distributions of various samples. Yellow and green points represent Sequoia and Thamnos samples compared with GE and LE samples, respectively. Large markers represent the comparison of Thamnos/Sequoia A (triangle) with LE and GE and Thamnos and Sequoia B (circle) with LE/GE. Small markers represent the same but with the sub-samples. Blue points show the GE metallicities compared with one another. Pink points show the LE metallicities compared with one another. Large circle markers represent the 20 sub-samples compared with the parent GE/LE and small markers the 20 sub-samples compared with one another. The dashed green (yellow) line represents the mean maximum difference for Thamnos (Sequoia) samples with LE (GE). Top-right panel: Projected distribution of the KS statistics, highlighting the differences between sub-samples and true differences in the case of Sequoia (yellow) and Thamnos (green). Bottom panel:  $[\text{Fe}/\text{H}]$  at which the difference in the cumulative metallicity distributions of various sub-samples becomes significant to the 95% level. The same colours and markers are used for the different samples.

$12.9^{+0.1}_{-0.1}$  Gyr, respectively. Here, the uncertainties are purely statistical, coming from the 100 different realisations of the CMD fitting and do not take into account possible systematics.

The LE sample age distribution confirms that this population is indeed made up of very ancient stars, in line with the theory that this population contains a significant contribution from the very ancient proto-Milky Way component dubbed Aurora (Belokurov & Kravtsov 2022). Our age distribution for GE is also consistent with previous studies (including earlier CMD fitting works; Gallart et al. 2019, spectra and photometry fitting to MSTO stars; Bonaca et al. 2020 and asteroseismology derived ages; Montalbán et al. 2021) with a truncation at 10 Gyr. Previous studies have found a tail to younger ages (7 – 9 Gyr, e.g. Horta et al. 2024b) in GE samples; however, we do not see this here, above the noise.

## 5. Discussion

We have presented age and metallicity distributions for the retrograde halo sub-structures Sequoia and Thamnos, solely derived from photometry. This is the first time that age distributions have been provided for these sub-structures (although see Feuillet et al. 2021, for age distributions of stars in the retrograde halo that they call ‘Sequoia’ which likely contains Sequoia and Thamnos stars with contamination from GE and Aurora). Age information adds an important dimension for understanding the build-up of the stellar halo. Still, we caution that proper handling of contamination is needed for interpreting these small halo sub-structures.



**Fig. 15.** Age distributions of the different samples. The line shows the weighted mean count per age bin and the shading indicates  $1\sigma$  across 100 iterations. For the first two rows, Sequoia A (B) and Thamnos A (B) are shown with dashed (solid) lines. The left panels of these first two rows show all stars in the Sequoia and Thamnos best fit CMDs, and the right panels show only stars below a  $[\text{Fe}/\text{H}]$  threshold of  $-1.5$  dex. The bottom row shows all stars in the GE and LE samples for comparison and one random  $N_T$  sub-sample.

### 5.1. Contamination

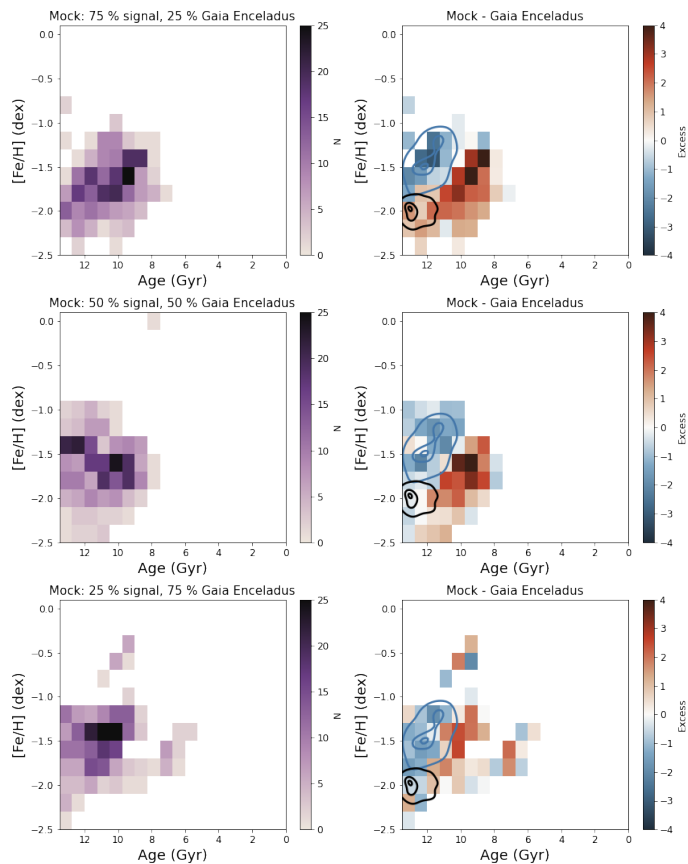
Throughout, we have highlighted the importance of taking into account contamination from the dominant populations in our local halo (i.e. *Gaia* Enceladus and an in situ halo component) when inferring properties of the progenitors. If contamination is not taken into account we can infer very different progenitor properties (for example, if using the mean metallicity to derive the mass of the progenitor). A more thorough understanding of how *Gaia* Enceladus and the in situ halo are distributed in IoM space will help in interpreting the signal from these smaller accretion events. For example, it has been shown in the past that debris from *Gaia* Enceladus are expected (e.g. Koppelman et al. 2020; Naidu et al. 2021; Amarante et al. 2022) and found through chemical abundance studies (e.g. Matsuno et al. 2022;

Ceccarelli et al. 2024) to populate the regions of IoM space where Sequoia and Thamnos are found. The Thamnos region is very complex owing to the low-energy orbit and added contamination of an in situ component, dubbed Aurora (Belokurov & Kravtsov 2022), which is another dominant component at low energy identified with  $[\text{Al}/\text{Fe}]$  abundances. While potential stellar population gradients are expected in massive accretion events such as GE (e.g. Amarante et al. 2022) and present themselves in IoM space, we note that the contamination observed in Sequoia from GE contains stars at  $[\text{Fe}/\text{H}] \sim -1.5$  dex (Matsuno et al. 2022). Thus, we do not expect the issue of a metallicity gradient in the progenitor of GE to affect our results, which are based on the low metallicity signal,  $[\text{Fe}/\text{H}] < -1.5$  dex, of Sequoia.

By selecting samples of *Gaia* Enceladus and the low energy in situ halo that are the contaminating populations, we have

been able to compare their age-metallicity distributions to that of Thamnos and Sequoia to unveil the differences. We have shown that there is a clear excess of stars at low metallicity in both the Sequoia and Thamnos sub-structures when compared with the respective contamination (see Figs. 11 and 12).

## 5.2. Mocks with contamination



**Fig. 16.** Mocks of an old, metal-poor signal with contamination from *Gaia* Enceladus. The top, middle, and bottom rows correspond to the mock with 25%, 50%, and 75% of the stars coming from the old, metal-poor signal and the rest from contamination. In all cases, the total number of stars is  $N_S$ . The first column shows the outputted age-metallicity distribution from DirSFH and the second column shows this minus the observed *Gaia* Enceladus age-metallicity distribution. The contours show the observed GE distribution (blue) and the black shows the input age-metallicity distribution for the old-metal poor signal.

We have previously shown (Sect. 3.4) that with the number of stars in our samples, we have the ability to recover an old and metal-poor population. Our resulting [Fe/H] and age distributions resemble the input, but with some shift in the mean to slightly higher metallicities and younger ages. This shift is less than 0.15 dex and 1 Gyr – always pointing in the direction of slightly younger and more metal-rich stars, where the shift has a slight dependence on how old the population is. However, these mock tests were made with a pure old and metal-poor signal and since our observed Sequoia and Thamnos samples also include contamination, we have constructed mocks that have contamination to check that we can recover an old, metal-poor signal.

For these new experiments, we took the mocks described in Sect. 3.4 and combined them with sub-samples of the observed contamination samples. These mocks consist of an old, metal-

poor signal (mean age of 12.3 Gyr and mean [Fe/H] of -2.0 dex) with contamination from *Gaia* Enceladus (Fig. 16) or the LE sample (Fig. C.1). The mocks are created with varying amounts of the stars (25%, 50%, and 75%) coming from the old, metal-poor signal and the rest from contamination. In all cases, the total number of stars is the same ( $N_S$  in Fig. 16,  $N_T$  Fig. C.1). The first column shows the outputted age-metallicity distribution from DirSFH and the second column shows this minus the contamination age-metallicity distribution (mimicking what we show in Figs 11 and 12).

We see that even for a highly contaminated sample, the old-metal poor signal is recovered in the residual age-metallicity space, but it is always shifted to younger ages and higher metallicities. When the contamination is highest (75% of the stars, the bottom row of Figs. 16 and C.1), then the residual shows an excess population that is distinct from the contamination (at lower metallicities and younger ages); however, this excess does not overlap with the old and metal-poor input sample. The recovered excess is shifted to higher metallicities and younger ages than the input. When contamination is reduced to 50% and 25% of the total number of stars, then we start to recover signal at the right ages and metallicities as the input signal but the mean is still shifted to younger ages and higher metallicities. This shift is especially prominent for the smaller sample size of  $N_S$ , as expected from the outcome of toy mocks without contamination in Sect. 3.4.

Our mock tests do show that for a sample of stars of the same sizes as our observed samples and with realistic contamination, we are still able to recover an excess signal from an extended old and metal-poor population. Even though in the residual age and metallicity space, there are shifts compared to the input, we are still able to tell that there is something other than the contamination in these samples. The question of how much this signal is shifted from its true age and metallicity distribution depends on the size of this population and also how it is distributed in age-metallicity space (e.g. if the age-metallicity is correlated, if the stars are predominantly old and at the edge of the grid, how extended the distribution is) and therefore our mocks do not provide an exact quantitative estimate of how the signal will have shifted in our fits of the observed Thamnos and Sequoia samples. However, we note that the residual signal is always shifted to younger ages and higher metallicities, so the signal we find from our observed Thamnos and Sequoia could be older or more metal-poor than what we are finding. Furthermore, the CMD fitting never creates a signal at older ages and lower metallicities than the input population (even when taking into account contamination) and so the observed residuals of Thamnos and Sequoia at lower metallicity (Figs 11 and 12) cannot be created artificially if there is no true old and metal-poor signal present in these samples.

It has been estimated that the level of contamination from *Gaia* Enceladus in the Sequoia region of IoM space is  $\leq 20\%$  (Matsuno et al. 2022). Our closest mock is shown in the top row of Fig. 16, consisting of an old and metal-poor sample with contamination from GE making up 25% of the number of stars. The residual signal (top-right panel in Fig. 16) does qualitatively resemble the signal we see in the observed Sequoia sample in Fig. 11 with the peak of the signal at ages of  $\sim 9$ –12 Gyr and extending from the lowest metallicities to [Fe/H]  $\sim -1.5$  dex (although the observed Sequoia does appear to extend to larger metallicities). This could indicate that the true Sequoia signal is older and more metal-poor than what we observe in Fig. 11. However, more tailored tests are needed to have a clearer picture of the true Sequoia signal. The fraction of contamination from

GE expected in the Sequoia region of IoM space depends on the configuration of the merger and could be a useful constraint in modelling efforts.

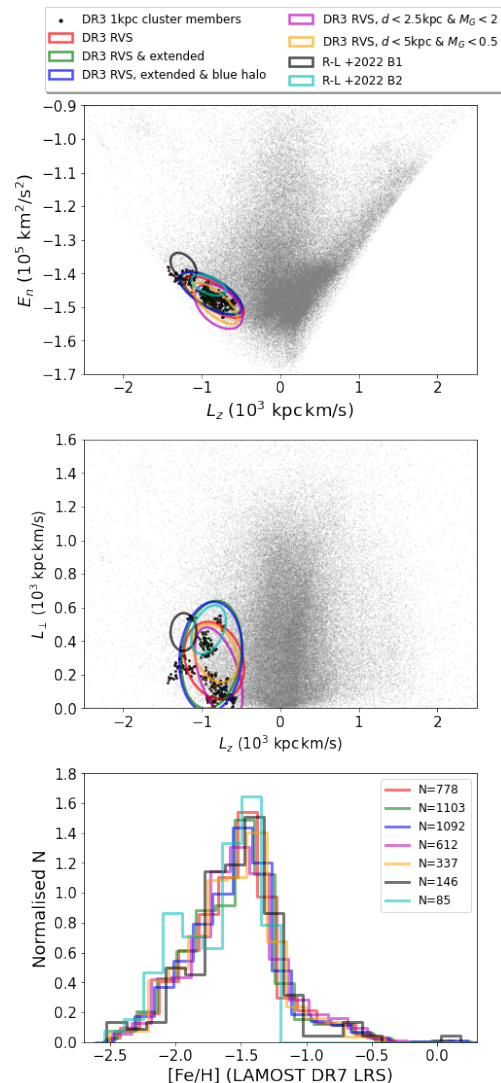
For Thamnos, the highly bound orbit already suggests that contamination is more significant. Looking at the mock tests in a qualitative sense with contamination from the LE sample (mimicking that expected in Thamnos) as shown in Fig. C.1, the case in which the contamination is 75% is most similar to the observed signal (Fig. 12). If the contamination is lower than 75%, then the signal extends to slightly younger ages than those seen in Fig. 12, with more stars with ages of <11 Gyr than what is seen in Thamnos. In these cases, the residual excess shows a strong age-metallicity relationship extending from old ages of  $\sim 13$  Gyr and  $[\text{Fe}/\text{H}]$  of  $-2.5$  dex to ages of  $\sim 9$  Gyr and a  $[\text{Fe}/\text{H}]$  of  $-1.5$  dex. This is not what is seen in what has been observed for Thamnos. The observed strength of the excess is qualitatively more similar to the 75% contamination sample. For lower contamination, the excess signal is stronger and not as representative of Fig. 12. This suggests that the levels of contamination from general LE populations in the region of Thamnos is of the order of  $\sim 75\%$  or more, but more tailored tests would be needed.

### 5.3. Contamination from 5D selections

We have presented results from CMD fitting of our 5D-selected samples of halo sub-structures. In an ideal case, we would like to make selections in 6D, but we are currently limited by the number of bright stars (oMSTO and above) with radial velocity measurements for the smaller halo sub-structures. We explain in Sect. 2.3.2 and Appendix A that these 5D selected samples resemble the 6D samples in terms of metallicity distributions and chemical abundances. In comparing the CMD of our 5D selected samples (first panel of Figs 4, 5, and 6) and 6D samples (Fig. 2), we have checked that the two agree and confirmed we are not adding different populations to the samples by our 5D selection, although subtle differences are hard to determine. We provided estimates on the completeness and purity of our 5D sample in Sect. 2.3.2, while both estimates appear quite low, we discuss the fact that our purity is a lower limit and, in fact, the stars that are being added by our 5D selection are generally halo stars that are slightly more extended in IoM space than the sub-structure in question (see Fig. A.2). These stars that are slightly more extended in IoM space are very likely members of the same population; however, they are not labelled as such due to our conservative selections when defining sub-structures in IoM space (Dodd et al. 2023), since this method is designed to find the main over-densities of debris and sub-structures.

Both the purity and completeness could still be improved upon by selecting samples in 6D, something that will become possible with *Gaia* DR4 thanks to the increase in radial velocities. However, we reiterate that the 6D selections are not themselves pure and still contain significant amounts of contamination from the general halo (given it is dominated by *Gaia* Enceladus and in situ components). It is not possible to completely remove contamination from the dominant halo components by selecting these small sub-structures purely dynamically, whether in 6D or 5D. The only way to obtain pure selections is to combine chemical selections with dynamic selections of the sub-structures. However, for our CMD fitting, this will not lead to improvements unless the selection function and completeness effects are very well understood and can be properly modelled.

### 5.4. Selection of Thamnos



**Fig. 17.** IoM space of the local halo sample described in Sect. 2.1 with ellipsoids showing over-density definitions of Thamnos from running a clustering algorithm (Lövdal et al. 2022) on various samples. The *Gaia* RVS sample definition used in this work is shown by a red ellipsoid. Other selections include extending the RVS sample (addition of ground-based RVs before clustering instead of adding them to the pre-defined clusters), a blue halo selection (those that are left of the isochrone separating the two sequences shown in e.g. Fig. 4), and complete samples (taking the  $M_G$  at the level that all stars down to that magnitude, in that volume, have a  $g_{RVS}$  above the RVS limit). The definitions by Ruiz-Lara et al. (2022) of Thamnos 1 and 2 are shown as black and cyan ellipsoids here (see R-L+2022 in the legend). Black points show the individual clusters in the Thamnos region when clustering on a 1kpc halo sample, highlighting the sub-structure we observe. The bottom panel shows the LAMOST DR7 LRS metallicity distributions for selecting from our 6D halo sample (Sect. 2.1) according to the different ellipsoids. These are normalised so that the area of each is equal to 1 and binned each with a relative shift of 0.01 dex to aid the comparison.

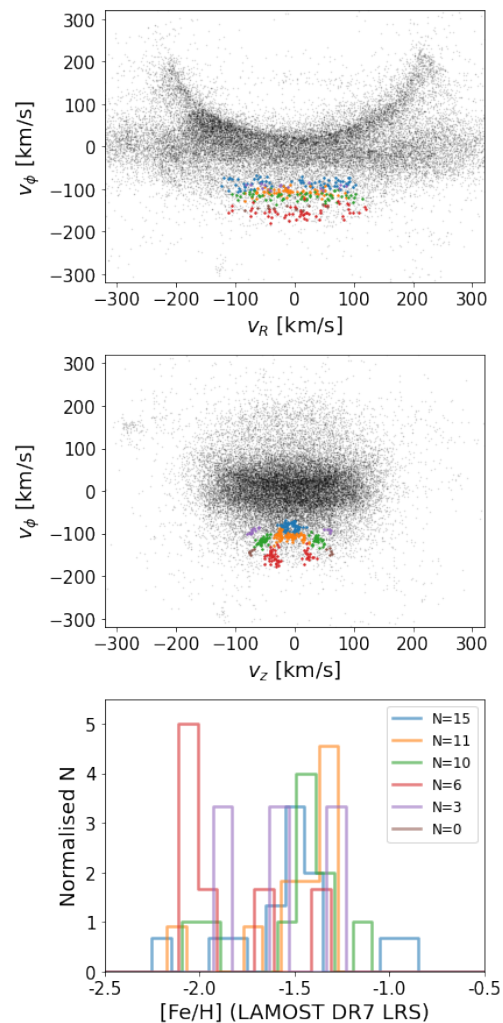
Our selection of sub-structures was motivated by the results of clustering in IoM space on *Gaia*DR3 data presented in Dodd et al. (2023). Prior to this, the same algorithm was applied to *Gaia* EDR3 data (the main difference between the two data releases is a factor of 4.7 less radial velocities in *Gaia* EDR3) supplemented with ground-based spectroscopic survey radial veloc-

ities (Lövdal et al. 2022; Ruiz-Lara et al. 2022). We found limited differences between the two datasets with regard to the definitions of sub-structures that are relevant to this work. However, one difference worth noting is the definition of Thamnos. The definition from Ruiz-Lara et al. (2022) of Thamnos is shown in Fig. 17 as black and cyan ellipsoids in IoM space. Here Thamnos is split into two sub-structures; Thamnos 1 at higher energy and Thamnos 2 at slightly lower  $L_z$ . Compared to our current DR3 definition (red ellipsoid in Fig. 17 and green points in Fig. 1) we can see that the ellipsoidal contours of Ruiz-Lara et al. (2022) are smaller and at the higher  $L_\perp$  end of our distribution, with Thamnos 1 being at higher energy than our definition. We find that in our clustering run on DR3, if we lower the significance threshold for detecting a cluster, we start to have sub-structures overlapping with Thamnos 1 in  $E_n$ .

The other ellipsoids shown in Fig. 17 correspond to running the clustering algorithm on different samples and are shown here to highlight the robustness of the Thamnos selection in this IoM region. These samples include extending the *Gaia* RVS sample with ground-based spectroscopic survey RVs before clustering (green ellipsoid in Fig. 17) and combining this with a blue halo selection (blue ellipsoid in Fig. 17) by selecting stars bluewards of an isochrone that separates the two sequences (BaSTI-IAC alpha-enhanced with an age of 11.6 Gyr and  $[\text{Fe}/\text{H}]$  of  $-0.8$  dex; see dashed line in the first two columns of Figs 4, 5, and 6). Since a radially incomplete sample can produce artificial overdensities in IoM space (and specifically along parabolic lines in  $L_z - E_n$  space, see Lane et al. 2022) we also run our clustering algorithm on three other samples: 1 kpc in which we are complete in *Gaia* RVS, 2.5 kpc with a cut of  $M_G < 2$  such that all stars down to that magnitude will have a *Gaia* RV in that volume and 5 kpc with a cut of  $M_G < 0.5$ , such that all stars in this volume will have a *Gaia* RV. The individual clusters in the Thamnos region within a 1 kpc sample are shown in Fig. 17, highlighting the individual sub-structure seen in this region. Despite these different samples, we can see that our definition of Thamnos remains very similar and we see that also in the metallicity distributions of the selected stars (see bottom panel of Fig. 17), there is always a peak around  $[\text{Fe}/\text{H}] \sim -1.5$  dex and then a significant metal-poor tail.

One reason we highlight the difference between Ruiz-Lara et al. (2022) and our Thamnos definition here is that Thamnos 1 in Ruiz-Lara et al. (2022) has a larger contribution of metal-poor stars and a more significant metal-poor tail (cyan histogram in the bottom panel of Fig. 17), likely due to lower contamination from the in situ halo at higher  $E_n$  and  $L_\perp$  and it being more retrograde. This was already seen in the discovery of Thamnos 1 and 2 in Koppelman et al. (2019); namely, in their Fig. 2 with  $L_z - E$  colour coded by  $[\text{Fe}/\text{H}]$  we see an excess of very metal-poor stars ( $[\text{Fe}/\text{H}] \sim -2$  dex) at the higher  $E_n$  region of Thamnos. The number of stars here is very low and likely similar to the number of metal-poor stars we are seeing at slightly lower  $E_n$  in our sample, but the relative contamination is greatly reduced so it appears to be a purer selection. If we were to select a Thamnos 1 sample from our *Gaia*DR3 halo sample (supplemented with ground-based RVs; see Sect. 2.1), according to the definition in Ruiz-Lara et al. (2022), then we would only have 209 stars (172 with  $M_G < 5$ ). This was not a sufficient number of stars for the CMD fitting, which is why we stuck to our DR3 definition and modelled the contamination carefully.

Thamnos occupies a region of IoM space that is very difficult to disentangle due to the low-energy orbit and the expected overlap of debris in this region, including with the tails of other larger sub-structures (*Gaia* Enceladus and Aurora) as discussed throughout this paper. However, the complexity becomes even



**Fig. 18.** Velocity space of local 1kpc halo sample with significant clusters in the Thamnos region highlighted in different colours. We can see the sub-structure presents itself as tight clumps in  $v_\phi$  and  $v_z$ . The bottom panel shows the LAMOST DR7 LRS metallicities (normalised such that the area sums to 1 and binned with a relative shift of 0.01 dex to aid comparison), indicating that some of these velocity sub-structures could be tight in terms of metallicity, but they always present some contamination from the  $[\text{Fe}/\text{H}] \sim -1.5$  dex peak.

greater when we look at the Thamnos region in more detail. There is a large amount of sub-structure within this region that can be picked up by clustering in IoM space. In Fig. 17, we show the results of clustering on the *Gaia* RVS halo sample confined to a 1 kpc volume around the Sun, the significant ( $> 3\sigma$ ) clusters are shown in black points. Looking at the velocity sub-structure in this small volume of 1kpc around the Sun (Fig. 18), we can see these clusters correspond to distinct and tight velocity sub-structures. An accreted sub-structure within a small volume is expected to display clumping in velocity space corresponding to individual streams on different orbital phases. Likely we are seeing the individual streams within the Thamnos debris. Some of these individual streams are very tight in velocity and with limited metallicity information (from LAMOST DR7 LRS: see bottom panel of Fig. 18) show a small spread (e.g. red points in Fig. 18 show a small spread with four stars around  $[\text{Fe}/\text{H}]$  of  $-2.0$  dex, one at  $[\text{Fe}/\text{H}] -1.6$  and  $-1.3$  dex). It would be interesting to follow these sub-clumps up further.

In summary, the accreted debris, which we refer to as Thamnos, appears to be present across quite an extended range of  $L_z$ ,  $L_\perp$  &  $E_n$  (see Fig. 17), as we see an excess in metal-poor stars. A more detailed look at this region with high-precision chemical abundances is needed to provide more insights into Thamnos' debris. It is also important to note that if we make the in situ versus accreted selections suggested by Belokurov & Kravtsov (2022) we lose the Thamnos sub-structure, which we have shown here is distinct from the general low-energy stellar populations.

## 6. Conclusions

In this paper, we present, for the first time, age distributions of the retrograde halo sub-structures: Sequoia and Thamnos. These data have been derived purely photometrically using CMD fitting techniques (i.e. CMDft.Gaia: Gallart et al. 2024) applied to samples of stars extracted from Gaia 5D data (missing radial velocities). CMD fitting also provides us with metallicity distributions that successfully reproduce the spectroscopic metallicity distributions of these samples (with a mean offset of  $\sim 0.1$ – $0.2$  dex to lower metallicities; see Fig. 13) testing the capability of CMDft.Gaia on old, metal-poor, and extended populations. We have shown that these age and metallicity distributions are well recovered, even with small sample sizes on the order of  $\sim 400$  stars and with extended age-metallicity distributions as expected for old accreted debris.

Our analysis has shown that Sequoia and Thamnos present metal-poor populations that are statistically different from the contamination expected in those regions of IoM space. From Figs 11 and 12, we can see that both present an excess of stars below  $[\text{Fe}/\text{H}] \sim -1.5$  dex. For Thamnos, these stars are mostly older than 11 Gyr; however, for Sequoia, this extends to slightly younger ages of  $\sim 9$  Gyr at  $[\text{Fe}/\text{H}]$  of  $-1.5$  dex. However, we caution that with the lower number of stars in our Sequoia samples, we could expect some shifts in the distribution to younger ages and higher metallicities. This is shown by our mock tests (see Sects. 3.4 & 5.2), which could explain this signal and the shape of the age-metallicity distribution seen in Fig. 11.

We have looked at the projected age distributions of Thamnos and Sequoia stars below  $[\text{Fe}/\text{H}]$  of  $-1.5$  dex to compare the differences with the full population (dominated by contamination) in Fig. 15. The metallicity cut shows a distribution that remains close to constant up to  $\sim 11$  Gyr for Sequoia, before a gradual decline to  $\sim 8$  Gyr. This is different from what is seen for Gaia Enceladus, where the decline (signifying the accretion) is a lot faster. We are not able to say whether this slower decline in Sequoia is significant or a limitation of the number of stars in our CMD fitting, but it could be telling us about the process and duration of accretion for the Sequoia progenitor. It would be interesting to look into this further with a larger sample of stars.

Thamnos is strikingly different in its age distribution showing a decline already from  $>13$  Gyr ago. This could mean a very early accretion, earlier than Gaia Enceladus, which would be in line with the low-energy orbit. Since there is still an important contribution from the contamination at low energy (see e.g. the LE contours in Fig. 12 compared to residual Thamnos signal), we have checked that below  $[\text{Fe}/\text{H}] < -2.0$  dex, we still see this steep decline in the Thamnos age distribution. Thus, the result appears to be robust.

The in situ low-energy (LE) population is dominated by the oldest stars in the local halo, telling us that these stars had already formed before the accretion of Gaia Enceladus, Sequoia, and Thamnos. For Gaia-Enceladus, we find that CMDft.Gaia is

now able to uncover an age-metallicity relation (Fig. 10) and this can be investigated in more detail with 6D selections due to the large number of stars in this sub-structure.

The fact that the age distributions of Gaia Enceladus match predictions by other studies (e.g. Bonaca et al. 2020; Montalbán et al. 2021, using different methods and data) and that the findings for the timeline for accretion of Sequoia and Thamnos match predictions based on their orbits (earlier accretion events sink to lower energy orbits) highlights the power of CMD fitting for uncovering the chronology of the assembly of our halo. This finding, combined with the fact that the spectroscopic metallicity distributions match our photometric metallicities very well, allows us to have substantial confidence in the results provided by CMDft.Gaia (Gallart et al. 2024). This will serve as an important tool for providing the age information that we have been so desperately missing. We show that this tool can already begin to be used, in combination with the rich Gaia dataset, to map out the local halo providing the temporal axis that is enabling us to interpret halo sub-structures and piece together the build-up of our halo.

With the upcoming Gaia DR4, we will have access to more stars with radial velocity information, which will allow us to increase our sample sizes for more reliable determinations of the age and metallicity distributions. We can combine this with detailed chemical abundance information to obtain a multi-dimensional understanding of our local halo. An interesting avenue would be to map the low-energy regime in terms of ages and understand how these ambiguous populations change over their evolution. Another interesting comparison is to look at the globular clusters (GCs) associated with these different sub-structures and see where they lie in age-metallicity space to confirm their association. This is the main aim of the CARMA project (Massari et al. 2023), namely, it has been designed to provide a very accurate chronological map of the Galactic GCs system that is fully consistent with that provided by CMDft.Gaia and based on the same stellar evolutionary framework (i.e. BaSTI-IAC).

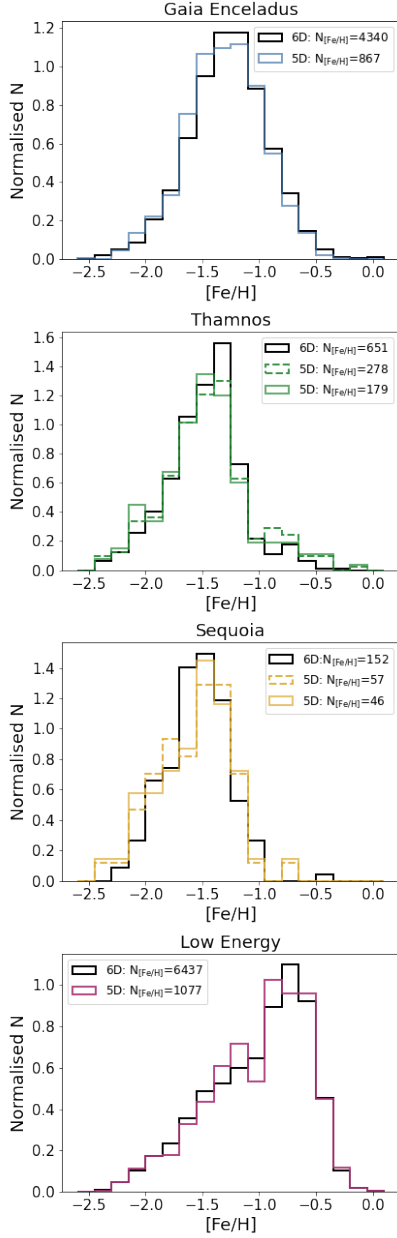
*Acknowledgements.* The authors thank the anonymous referee for their helpful and constructive feedback. This research has been partially funded by a Spinoza award by NWO (SPI 78-411) and support from a MW-Gaia COST (European Cooperation in Science and Technology) action (CA18104) STSM (Short-Term Scientific Mission) grant. TRL acknowledges support from Juan de la Cierva fellowship (IJC2020-043742-I) as well as research projects PID2020-113689GB-I00, and PID2020-114414GB-I00, financed by MCIN/AEI/10.13039/501100011033, the project A-FQM-510-UGR20 financed from FEDER/Junta de Andalucía-Consejería de Transformación Económica, Industria, Conocimiento y Universidades and by grants P20-00334 and FQM108, financed by the Junta de Andalucía (Spain). EFA acknowledges support from HORIZON TMA MSCA Postdoctoral Fellowships Project TEMPOS, number 101066193, call HORIZON-MSCA-2021-PF-01, by the European Research Executive Agency. CG, TLR, SC and EFA acknowledges support from the Agencia Estatal de Investigación del Ministerio de Ciencia e Innovación (AEI-MCINN) under grant “At the forefront of Galactic Archaeology: evolution of the luminous and dark matter components of the Milky Way and Local Group dwarf galaxies in the Gaiaera” with reference PID2020-118778GB-I00/10.13039/501100011033. CG also acknowledges financial support from the Spanish Ministry of Science, Innovation and University (MICIN) through the Spanish State Research Agency, under Severo Ochoa Centres of Excellence Programme 2020-2023 (CEX2019-000920-S). SC acknowledges financial support from PRIN-MIUR-22: CHRONOS: adjusting the clock(s) to unveil the CHRONO-chemo-dynamical Structure of the Galaxy (PI: S. Cassisi) finanziato dall’Unione Europea – Next Generation EU, and Theory grant INAF 2023 (PI: S. Cassisi). This work has made use of data from the European Space Agency (ESA) mission Gaia (<https://www.cosmos.esa.int/gaia>), processed by the Gaia Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular, the institutions participating in the Gaia Multilateral Agreement. The analysis has benefited from the use of the following packages: vaex (Breddels & Veljanoski 2018), numpy (Van

Der Walt et al. 2011), matplotlib (Hunter 2007), scipy (Virtanen et al. 2020), seaborn (Waskom 2021), swiftdatasets (Borrow 2021) and jupyter notebooks (Kluyver et al. 2016).

## References

- Accetta, K., Aerts, C., Aguirre, V. S., et al. 2022, *ApJS*, 259, 35
- Aguado, D. S., Belokurov, V., Myeong, G., et al. 2021, *ApJ*, 908, L8
- Amarante, J. A., Debattista, V. P., Silva, L. B. E., Laporte, C. F., & Deg, N. 2022, *ApJ*, 937, 12
- Aparicio, A. & Hidalgo, S. L. 2009, *ApJ*, 138, 558
- Babusiaux, Fabrice, C., Khanna, S., et al. 2022, *A&A*
- Babusiaux, C., van Leeuwen, F., Barstow, M., et al. 2018, *A&A*, 616, A10
- Belokurov, V., Erkal, D., Evans, N., Koposov, S., & Deason, A. 2018, *MNRAS*, 478, 611
- Belokurov, V. & Kravtsov, A. 2022, *MNRAS*, 514, 689
- Belokurov, V., Sanders, J. L., Fattahi, A., et al. 2020, *MNRAS*, 494, 3880
- Bernard, E. J., Ferguson, A. M., Barker, M. K., et al. 2012, *MNRAS*, 420, 2625
- Bernard, E. J., Schultheis, M., Di Matteo, P., et al. 2018, *MNRAS*, 477, 3507
- Bonaca, A., Conroy, C., Cargile, P. A., et al. 2020, *ApJ*, 897, L18
- Borrow, J. 2021, *swiftdatasets* 1.2.1
- Breddels, M. A. & Veljanoski, J. 2018, *A&A*, 618, A13
- Buder, S., Sharma, S., Kos, J., et al. 2021, *MNRAS*, 506, 150
- Ceccarelli, E., Massari, D., Mucciarelli, A., et al. 2024, *A&A*, 684, A37
- Cignoni, M., Cole, A. A., Tosi, M., et al. 2013, *ApJ*, 775, 83
- Cignoni, M. & Tosi, M. 2010, *Advances in Astronomy*, 2010, 158568
- Cole, A. A., Weisz, D. R., Dolphin, A. E., et al. 2014, *ApJ*, 795, 54
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *Research in A&A*, 12, 1197
- Dal Tio, P., Mazzi, A., Girardi, L., et al. 2021, *MNRAS*, 506, 5681
- Das, P., Hawkins, K., & Jofré, P. 2020, *MNRAS*, 493, 5195
- De Jong, R., Agertz, O., Berbel, A., et al. 2019, *The Messenger*, 175, 3
- De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, 449, 2604
- Deason, A. J. & Belokurov, V. 2024, *New A Rev.*, 99, 101706
- Dodd, E., Callingham, T. M., Helmi, A., et al. 2023, *A&A*, 670, L2
- Dolphin, A. E. 2002, *MNRAS*, 332, 91
- Everall, A. & Boubert, D. 2022, *MNRAS*, 509, 6205
- Fernández-Alvar, E., Ruiz-Lara, T., Gallart, C., et al. 2025, *A&A*, arXiv:2503.19536
- Feuillet, D. K., Feltzing, S., Sahlholdt, C. L., & Casagrande, L. 2020, *MNRAS*, 497, 109
- Feuillet, D. K., Sahlholdt, C. L., Feltzing, S., & Casagrande, L. 2021, *MNRAS*, 508, 1489
- Fitzpatrick, E. L., Massa, D., Gordon, K. D., Bohlin, R., & Clayton, G. C. 2019, *ApJ*, 886, 108
- Gaia Collaboration, Vallenari, A., Brown, A., Prusti, T., et al. 2022, *A&A*, 649, A1
- Gallart, C., Bernard, E. J., Brook, C. B., et al. 2019, *Nature Astronomy*, 3, 932
- Gallart, C., Freedman, W. L., Aparicio, A., Bertelli, G., & Chiosi, C. 1999, *ApJ*, 118, 2245
- Gallart, C., Surot, F., Cassisi, S., et al. 2024, *A&A*, 687, A168
- Gratton, R. G., Carretta, E., Bragaglia, A., Lucatello, S., & D’Orazi, V. 2010, *A&A*, 517, A81
- Hawkins, K., Jofre, P., Masseron, T., & Gilmore, G. 2015, *MNRAS*, 453, 758
- Haywood, M., Di Matteo, P., Lehnert, M., et al. 2018, *ApJ*, 863, 113
- Helmi, A. 2020, *ARA&A*, 58, 205
- Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, *Nature*, 563, 85
- Helmi, A. & Tim de Zeeuw, P. 2000, *MNRAS*, 319, 657
- Horta, D., Cunningham, E. C., Sanderson, R., et al. 2024a, *MNRAS*, 527, 9810
- Horta, D., Lu, Y. L., Ness, M. K., Lisanti, M., & Price-Whelan, A. M. 2024b, *ApJ*, 971, 170
- Horta, D., Schiavon, R. P., Mackereth, J. T., et al. 2021, *MNRAS*, 500, 1385
- Horta, D., Schiavon, R. P., Mackereth, J. T., et al. 2023, *MNRAS*, 520, 5671
- Hunter, J. D. 2007, *IEEE Annals of the History of Computing*, 9, 90
- Jin, S., Trager, S. C., Dalton, G. B., et al. 2023, *MNRAS*, stad557
- Katz, D., Sartoretti, P., Guerrier, A., et al. 2023, *A&A*, 674, A5
- Kirby, E. N., Cohen, J. G., Guhathakurta, P., et al. 2013, *ApJ*, 779, 102
- Kluyver, T., Ragan-Kelley, B., Pérez, F., et al. 2016, *Jupyter Notebooks-a publishing format for reproducible computational workflows.*, Vol. 2016 (IOS Press), 87
- Kolmogorov, A. N. 1933, *Giorn Dell’inst Ital Degli Att*, 4, 89
- Koppelman, H. H., Bos, R. O., & Helmi, A. 2020, *A&A*, 642, L18
- Koppelman, H. H., Helmi, A., Massari, D., Price-Whelan, A. M., & Starkenburg, T. K. 2019, *A&A*, 631, L9
- Kroupa, P., Tout, C. A., & Gilmore, G. 1993, *MNRAS*, 262, 545
- Lallement, R., Vergely, J. L., Babusiaux, C., & Cox, N. L. J. 2022, *A&A*, 661, A147
- Lane, J. M., Bovy, J., & Mackereth, J. T. 2022, *MNRAS*, 510, 5119
- Lane, J. M., Bovy, J., & Mackereth, J. T. 2023, *MNRAS*, 526, 1209
- Lindgren, L., Bastian, U., Biermann, M., et al. 2021, *A&A*, 649, A4
- Lövdal, S. S., Ruiz-Lara, T., Koppelman, H. H., et al. 2022, *A&A*, 665, A57
- Ma, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2016, *MNRAS*, 456, 2140
- Mackereth, J. T. & Bovy, J. 2020, *MNRAS*, 492, 3631
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *ApJ*, 154, 94
- Massari, D., Aguado-Agelet, F., Monelli, M., et al. 2023, *A&A*, 680, A20
- Matsuno, T., Aoki, W., & Suda, T. 2019, *ApJ*, 874, L35
- Matsuno, T., Koppelman, H. H., Helmi, A., et al. 2022, *A&A*, 661, A103
- Mazzi, A., Girardi, L., Trabucchi, M., et al. 2024, *MNRAS*, 527, 583
- McMillan, P. J. 2017, *MNRAS*, 465, 76
- McQuinn, K. B., Newman, M. J., Savino, A., et al. 2024, *ApJ*, 961, 16
- Montalbán, J., Mackereth, J. T., Miglio, A., et al. 2021, *Nature Astronomy*, 5, 640
- Myeong, G., Belokurov, V., Aguado, D. S., et al. 2022, *ApJ*, 938, 21
- Myeong, G., Vasiliev, E., Iorio, G., Evans, N., & Belokurov, V. 2019, *MNRAS*, 488, 1235
- Naidu, R. P., Conroy, C., Bonaca, A., et al. 2020, *ApJ*, 901, 48
- Naidu, R. P., Conroy, C., Bonaca, A., et al. 2021, *ApJ*, 923, 92
- Noël, N. E., Aparicio, A., Gallart, C., et al. 2009, *ApJ*, 705, 1260
- Oria, P.-A., Tenachii, W., Ibata, R., et al. 2022, *ApJ*, 936, L3
- Pietrinferni, A., Hidalgo, S., Cassisi, S., et al. 2021, *ApJ*, 908, 102
- Prusti, T., De Bruijne, J., Brown, A. G., et al. 2016, *A&A*, 595, A1
- Ruiz-Lara, T., Gallart, C., Bernard, E. J., & Cassisi, S. 2020, *Nature Astronomy*, 4, 965
- Ruiz-Lara, T., Helmi, A., Gallart, C., Surot, F., & Cassisi, S. 2022, *A&A*, 668, L10
- Ruiz-Lara, T., Matsuno, T., Lövdal, S. S., et al. 2022, *A&A*, 665, A58
- Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, 403, 1829
- Skillman, E. D., Monelli, M., Weisz, D. R., et al. 2017, *ApJ*, 837, 102
- Smart, R. L., Sarro, L., Rybizki, J., et al. 2021, *A&A*, 649, A6
- Smirnov, N. V. 1939, *Bulletin Moscow University*, 2, 3
- Soderblom, D. R. 2010, *ARA&A*, 48, 581
- Springel, V., White, S. D., Jenkins, A., et al. 2005, *Nature*, 435, 629
- Steinmetz, M., Matijević, G., Enke, H., et al. 2020, *ApJ*, 160, 82
- Tolstoy, E., Hill, V., & Tosi, M. 2009, *ARA&A*, 47, 371
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in science & engineering*, 13, 22
- Vincenzo, F., Spitoni, E., Calura, F., et al. 2019, *MNRAS Letters*, 487, L47
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, *Nature Methods*, 17, 261
- Waskom, M. L. 2021, *Journal of Open Source Software*, 6, 3021
- Weisz, D. R., Zucker, D. B., Dolphin, A. E., et al. 2012, *ApJ*, 748, 88
- Yuan, Z., Myeong, G., Beers, T. C., et al. 2020, *ApJ*, 891, 39
- Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, *Research in A&A*, 12, 723

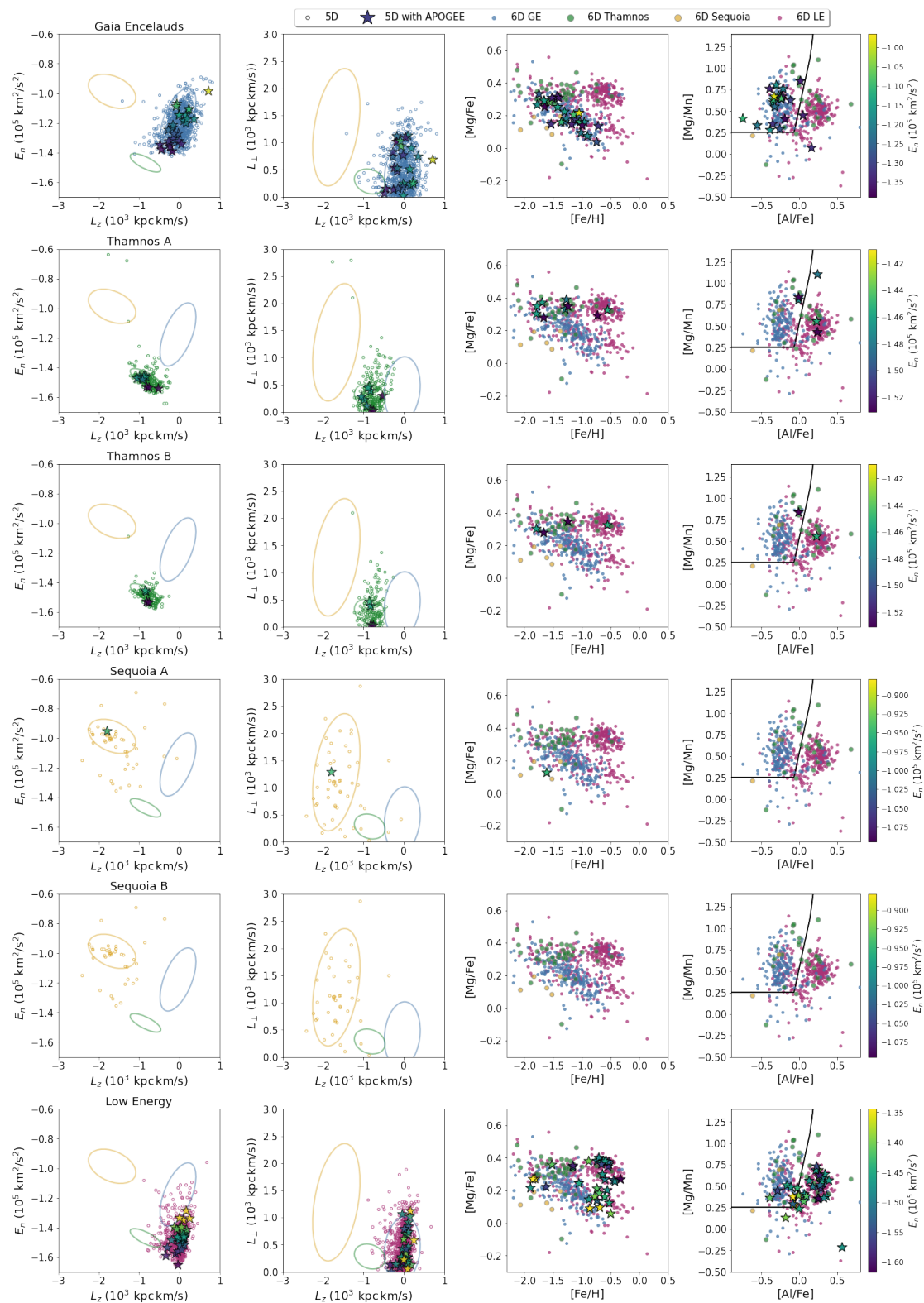
## Appendix A: Validity of our 5D selections



**Fig. A.1.** LAMOST DR7 LRS metallicity distributions of the 6D sub-structures and the 5D selected samples. Since the 5D method removes some 6D stars (those outside of the pseudo velocity selection) and adds stars, we can check the metallicity distributions of the stars that have a LAMOST DR7 LRS [Fe/H] that have been added to check if we are adding noise or if the population resembles the original 6D defined population.

Our 5D selection method could be adding more contamination to our samples. We can use the 6D stars that end up in our 5D selected samples to check that we are not adding noise to these samples. In Sect. 2.3.2 we have shown that the 6D stars we are adding by our selection do not belong to other sub-structures but general halo stars. Here we show that these general halo stars are consistent with the 6D sample we are considering, in terms of their metallicity distribution. Fig. A.1 shows the LAMOST DR7 LRS metallicity distributions of stars in our 5D samples compared to our 6D samples of the same sub-structure. Any changes in our MDFs are due to adding contaminating stars but we can

see that our MDFs resemble the 6D MDFs, with some difference in the shape for Sequoia that is due to the low number of stars. Stars which have a LAMOST DR7 LRS metallicity or APOGEE DR17 abundances must also have a radial velocity from the corresponding survey so we can also check where these added stars are in IoM space. We show in Fig. A.2 the distribution, in IoM space, of all 6D stars that enter our 5D selections in the first two columns. Star markers correspond to those with APOGEE DR17 abundances coloured by their  $E_n$  and these are shown in the last two columns compared to the true 6D samples in the background. This shows us that the added stars are slightly more extended in IoM space than our true 6D sub-structure but chemically they are following the same sequences as the 6D selections.



**Fig. A.2.** Left panels: Stars with radial velocities from the 5D samples (selected assuming  $V_{\text{los}}=0 \text{ km/s}$ ) and where they are distributed in IoM space using their RVs. We see that the 5D selected samples are more extended than the original definitions from [Dodd et al. \(2023\)](#) shown for GE, Thamnos, and Sequoia by the ellipsoids in the corresponding colours. Typically selecting stars to higher  $L_{\perp}$ . Stars with a star marker are highlighted since they have APOGEE abundances, and they are shown in the last two panels compared to the true 6D definitions. We see that the 5D selected samples follow the same distributions as the parent 6D sample in these spaces.

**Appendix B: Parameters for CMD fitting routine**

Table B.1 can be used for reproducing the fitting bundle that has been used for selecting the region in which we perform the CMD fitting. Table B.2 shows the seed points used for the CMD fitting.

**Table B.1.** Vertices of the bundle used for defining the region within which we performed the CMD fitting.

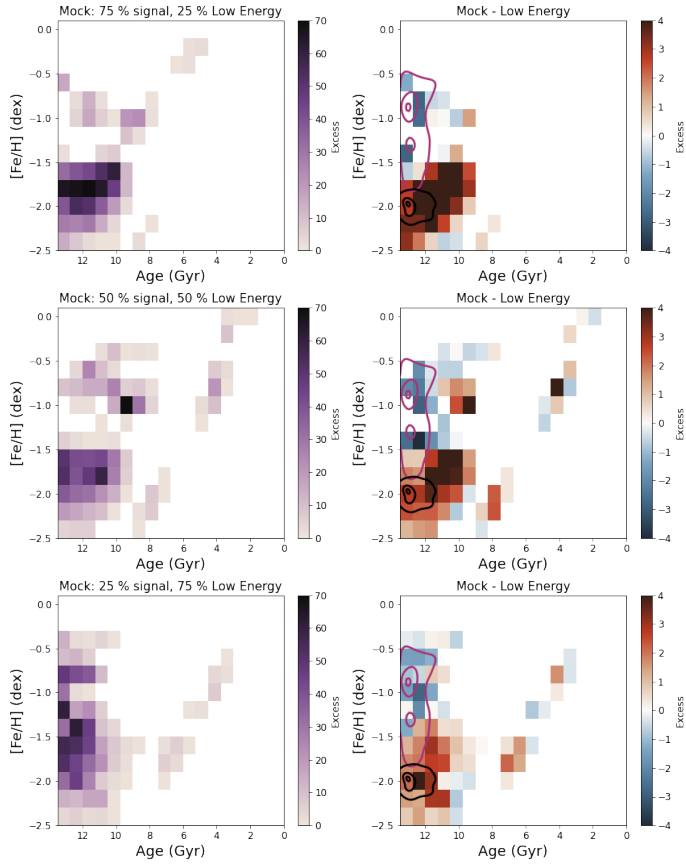
BP-RP Colour	$M_G$
0.388	-1.615
1.112	-3.326
1.953	-3.056
1.948	-0.633
1.452	1.458
1.209	4.611
0.126	4.579
-0.419	1.284
-0.419	-3.563
-0.424	-3.595
0.388	-1.615

**Table B.2.** Age (Z) seed points used for CMD fitting.

Age (Gyr)	Z	[Fe/H] (dex)
0.08	0.0001	-2.498
0.126	0.00013	-2.4
0.192	0.00015	-2.302
0.262	0.00020	-2.204
0.334	0.00025	-2.106
0.404	0.00031	-2.008
0.469	0.00039	-1.91
0.532	0.00048	-1.812
0.596	0.00061	-1.714
0.656	0.00076	-1.616
0.718	0.00095	-1.518
0.784	0.00119	-1.42
0.857	0.00149	-1.322
0.938	0.00187	-1.224
1.028	0.00234	-1.126
1.128	0.00293	-1.028
1.244	0.00366	-0.93
1.391	0.00457	-0.832
1.576	0.00571	-0.734
1.81	0.00712	-0.636
2.066	0.00887	-0.538
2.337	0.01104	-0.44
2.609	0.01372	-0.342
2.882	0.01702	-0.244
3.156	0.02105	-0.146
3.427	0.02595	-0.048
3.695	0.03187	0.05
3.978	0.039	0.148
4.272		
4.581		
4.946		
5.389		
5.858		
6.351		
6.861		
7.372		
7.883		
8.393		
8.904		
9.415		
9.925		
10.436		
10.947		
11.457		
11.968		
12.479		
12.989		
13.5		

## Appendix C: Mocks with contamination

Here, we show the output age-metallicity distributions of our mock with added contamination from the low-energy (LE) sample described in Sect. 5.2.



**Fig. C.1.** Same as Fig. 16, but for the case of the contamination from LE and the total number of stars being  $N_T$ .