



# The chemical diversity of the metal-poor Milky Way

Nicole Buckley <sup>1</sup>★, Payel Das <sup>1</sup>, Paula Jofré,<sup>2</sup> Robert M. Yates <sup>3</sup> and Keith Hawkins <sup>4</sup>

<sup>1</sup>Department of Physics, University of Surrey, Stag Hill, Guildford GU2 7XH, UK

<sup>2</sup>Instituto de Estudios Astrofísicos, Facultad de Ingeniería y Ciencias, Universidad Diego Portales, Av. Ejercito 441, 8370191 Santiago, Chile

<sup>3</sup>Centre for Astrophysics Research, University of Hertfordshire, Hatfield AL10 9AB, UK

<sup>4</sup>Department of Astronomy, The University of Texas at Austin, 2515 Speedway Boulevard, Austin, TX 78712, USA

Accepted 2024 September 19. Received 2024 September 17; in original form 2024 July 26

## ABSTRACT

We present a detailed study of the chemical diversity of the metal-poor Milky Way using data from the GALAH DR3 survey. Considering 17 chemical abundances relative to iron ( $[X/Fe]$ ) for 9923 stars, we employ principal component analysis (PCA) and extreme deconvolution (XD) to identify 10 distinct stellar groups. This approach, free from chemical or dynamical cuts, reveals known populations, including the accreted halo, thick disc, thin disc, and *in situ* halo. The thick disc is characterized by multiple substructures, suggesting it comprises stars formed in diverse environments. Our findings highlight the limited discriminatory power of magnesium in separating accreted and disc stars. Elements such as Ba, Al, Cu, and Sc are critical in distinguishing disc from accreted stars, while Ba, Y, Eu, and Zn differentiate disc and accreted stars from the *in situ* halo. This study demonstrates the potential power of combining a latent space representation of the data (PCA) with a clustering algorithm (XD) in Galactic archaeology, in providing new insights into the Galaxy’s assembly and evolutionary history.

**Key words:** methods: data analysis – methods: observational – Stars: abundances – Stars: Population II – Galaxy: abundances – Galaxy: structure.

## 1 INTRODUCTION

The Milky Way’s (MW) stellar components have distinct star formation histories (SFHs) that manifest in unique chemical signatures. Chemical diversity, which measures the spread of elemental abundances, can indicate contributions from both nucleosynthetic processes and accreted systems. Variations in chemical diversity are more significant in the stellar halo than in the disc. This is because the halo has accumulated stars from various smaller galaxies, each with distinct SFHs, whereas the disc has experienced more thorough chemical homogenization due to effective material mixing (Reddy, Lambert & Prieto 2006; Feng & Krumholz 2014).

The thick disc, first described by Gilmore & Reid (1983), consists of prograde stars ( $L_z > 0$ ) with scale heights between  $530 \pm 32$  and  $630 \pm 29$  pc above the Galactic plane (Kumar et al. 2021). These stars are older than those in the thin disc ( $\geq 10$  Gyr), more metal-poor ( $-2.2 \leq [Fe/H]$  (dex)  $\leq -0.5$ ), and show higher enhancements of  $\alpha$  elements relative to iron ( $[\alpha/Fe] > 0.2$  dex) (Freeman & Bland-Hawthorn 2002; Navarro et al. 2011; Hayden et al. 2015). The thick disc likely formed quickly, marked by an early burst of star formation in the MW’s history. Its formation theories include: (i) secular evolution through heating and primordial collapse (Prantzos et al. 2023), (ii) a significant merger, for example, Gaia Enceladus Sausage (GES), influencing the proto-disc before the thin disc’s formation (Belokurov et al. 2018; Helmi et al. 2018; Bignone, Helmi & Tissera

2019; Ciucă et al. 2023), and (iii) direct accretion of stars from satellite galaxies (Reddy et al. 2006).

The thin disc dominates the stellar material in spiral galaxies, with a high and constant star formation rate (SFR) due to abundant gas, leading to many young stars [ $< 8$  Gyr (Yu et al. 2021)]. It is the low- $\alpha$  [ $[\alpha/Fe] < 0.2$  dex (Navarro et al. 2011)] region of the galactic disc and is more metal rich due to later gas accretion. The thin disc extends to heights of  $\approx 230 \pm 20$  to  $330 \pm 11$  pc (Kumar et al. 2021).

The MW stellar halo comprises two main parts: the inner and outer halo (Carollo et al. 2007; Nissen & Schuster 2010; Beers et al. 2012). Within the inner halo ( $R_{gal} < 15$  kpc), there are high- $\alpha$  (Ishigaki, Chiba & Aoki 2012) and low- $\alpha$  components (Bensby, Feltzing & Lundström 2003; Nissen & Schuster 2010, 2012; Sheffield et al. 2012; Hawkins et al. 2014; Jackson-Jones et al. 2014). The high- $\alpha$  component includes *in situ* stars, GES, and the heated high- $\alpha$  disc (Zolotov et al. 2009; Nissen & Schuster 2010; Tissera et al. 2014; Pillepich, Madau & Mayer 2015), formed from bursty star formation triggered by mergers (Ledinauskas & Zubovas 2018; Fernández-Alvar et al. 2019; Liu et al. 2019; Emami et al. 2021). The low- $\alpha$  component is comprised of disrupted satellites with extended, less intense episodes of star formation. The inner halo is primarily influenced by GES, a significant dwarf galaxy accreted  $\approx 8–11$  Gyr ago, marking the MW’s last major minor merger (Belokurov et al. 2020).

Due to the availability of large stellar spectroscopic surveys, for example, GALAH (Buder et al. 2021), APOGEE (Majewski et al. 2017), SDSS/SEGUE (Yanny et al. 2009), LAMOST (Cui et al. 2012), and H3 (Conroy et al. 2019), many studies have used integrals

\* E-mail: [nb00699@surrey.ac.uk](mailto:nb00699@surrey.ac.uk)

of motion and chemical abundances to cluster and disentangle systems with different histories. Dynamical tagging is one method used, in particular, using quasi-conserved quantities such as orbital actions ( $J_z$ ,  $J_R$ , and  $L_z$ ) and energy ( $E$ ). Although these quantities are not conserved over long time-scales, the  $E - L_z$  space has been instrumental in identifying distinct accreted structures such as GES (Helmi & de Zeeuw 2002; Belokurov et al. 2018; Helmi et al. 2018), Sequoia (Myeong et al. 2019), and Kraken (a possible pre-disc *in situ* population) (Forbes 2020; Kruijssen et al. 2020).

Chemical tagging is the process of grouping stars together based on chemical abundance ratios<sup>1</sup> (Freeman & Bland-Hawthorn 2002). It assumes that stars born in the same population are chemically similar to each other but chemically distinct from other stars born at different times and/or locations. This allows us to use their preserved chemical signatures as snapshots of the Galaxy’s accretion history, facilitating a detailed reconstruction of the MWs formation through chemical tagging (Robertson et al. 2005; Font et al. 2006; Bedell et al. 2018). Buder et al. (2022) analyse up to 30 element abundances, using them to successfully distinguish GES from the *in situ* populations and find that the chemical signatures of Mg, Si, Na, Al, Mn, Fe, Ni, and Cu, significantly differ from *in situ* MW stars. Through Gaussian mixture models applied to these chemical abundances, they isolate 1049 stars associated with GES.

In the literature, 2D hyperplanes, such as the Tinsley–Wallerstein hyperplane which features  $[\alpha/\text{Fe}]$  and  $[\text{Fe}/\text{H}]$  (Tinsley 1979) have been used. The early-type dwarf galaxies formed stars early and have predominantly old stellar populations. The later types with ongoing star formation have shallower  $[\alpha/\text{Fe}] - [\text{Fe}/\text{H}]$  slopes, which separates it from the steeper gradients present for MW stars. This is due to a difference in SFR, which is a result of the relative size difference between the smaller satellite systems and the MW (Das, Hawkins & Jofre 2020).

Another informative 2D plane is  $[\text{Mg}/\text{Mn}] - [\text{Al}/\text{Fe}]$ , shown in Das et al. (2020) to effectively distinguish between accreted populations and the metal-poor disc (Hawkins et al. 2015; Mackereth et al. 2019; Price-Jones et al. 2020). Mg and Mn are relatively pristine tracers of core-collapse supernovae (CCSNe) and type 1a supernovae (SNe1a), respectively (Matteucci 2016). Similarly to  $[\alpha/\text{Fe}]$ , the contrasting time-scales of these events make  $[\text{Mg}/\text{Mn}]$  an invaluable tool for decoding the SFHs of systems. As for  $[\text{Al}/\text{Fe}]$ , accreted populations, such as GES (Hawkins et al. 2015) exhibit sub-solar values of this ratio due to the metallicity dependence of the light odd-Z element production (Kobayashi et al. 2006; Ting et al. 2012).

Exploring other nucleosynthesis channels, Manea et al. (2023) emphasized the role of neutron-capture elements in distinguishing between disc stars, significantly reducing the number of chemical doppelgangers. The ratio of s-process to r-process elements, especially  $[\text{Ba}/\text{Eu}]$ , is important here. Analysing  $[\text{Ba}/\text{Eu}]$  ratios within these galaxies reveals sub-solar values at low metallicities ( $[\text{Fe}/\text{H}] < -1.7$  dex), followed by a marked increase as more massive stars dominate ISM enrichment (Lanfranchi, Matteucci & Cescutti 2006). The difference between light s-process (ls) elements, such as yttrium and lanthanum, and heavy s-process (hs) elements, such as Ba is also crucial. These elements are primarily synthesized during the AGB phase of stellar evolution, with the  $[\text{hs}/\text{ls}]$  ratio serving as a gauge for neutron source efficiency and s-process output. Lower metallicity in AGB stars typically results in a higher  $[\text{hs}/\text{ls}]$  ratio, indicating

<sup>1</sup> $[\text{X}/\text{Y}] = \log(N_X/N_Y)_* - \log(N_X/N_Y)_\odot$ , is the logarithmic ratio between two abundances with respect to solar abundances, where  $N_X$  represents the number of ‘X’ atoms per unit volume.

stronger neutron-capture activity and the significant influence of the  $^{13}\text{C}$  pocket (Kappeler, Beer & Wisshak 1989; Busso, Gallino & Wasserburg 1999). Due to the obscurity surrounding the sites and yields of neutron-capture elements, there has been little analysis of these heavier elements in abundance space.

While 2D abundance hyperplanes effectively highlight enrichment differences between the halo and disc, the boundary between the inner halo and metal-poor disc requires more information to break down the degeneracy – implying that this region is more chemically complex and has a greater chemical diversity. To uncover the extent of the chemical diversity in the MW, techniques have been applied to reveal the number of dimensions in chemical space (hereafter referred to as  $\mathcal{C}$ -space, drawing from Freeman & Bland-Hawthorn 2002), and the abundance ratios which are most effective at breaking the chemical degeneracy. Having a high-dimensional  $\mathcal{C}$ -space is heavily recommended in Andrews et al. (2017) for the next generation of multielement stellar abundance surveys.

The choice of appropriate abundances is key when trying to recover signatures of accretion events. In Ting & Weinberg (2022), they determine that for an observational uncertainty of  $\approx 0.01 - 0.015$  dex, 5–7 abundances must be conditioned on. This is shown in Anders et al. (2018), who used t-SNE (t-distributed stochastic neighbour embedding), a dimensionality-reduction technique, to dissect  $\mathcal{C}$ -space revealing insights into the chemical diversity of the disc and suggesting that it is formed from various progenitor systems and is chemically complex. In this paper, we investigate this chemical diversity and assess the chemical degeneracy between the inner halo and the metal-poor MW (defined by  $[\text{Fe}/\text{H}] \leq -0.5$  dex) in the GALAH survey. The focus on the metal-poor MW is due to the high level of overlap in *in situ* disc, halo and accreted populations when viewed in 2D abundance planes and to also find known without the use of dynamical or chemical selection cuts – other than a metallicity cut. We aim to reveal nucleosynthesis contributions in the metal-poor MW and identify reproducible chemical hyperplanes for clustering chemically similar groups. After applying XD to the PCA-transformed  $\mathcal{C}$ -space, we distinguish these groups by their chemodynamical properties and compare them with known structures. By examining the metal-poor MW’s chemical diversity, we seek to determine: (i) the most effective chemical abundances for resolving chemical degeneracy in the disc and disc–halo region; (ii) the effectiveness of unsupervised clustering algorithms in identifying chemically coherent groups in the transformed  $\mathcal{C}$ -space; and (iii) insights into the formation histories of the thick disc and stellar halo.

In Section 2, we discuss the selection of the 17 element enhancements ( $[\text{X}/\text{Fe}]$ ) from GALAH DR3, and the sample cuts. In Section 3, we utilize PCA to transform a high-dimensional (17)  $\mathcal{C}$ -space to a lower one (9), on which extreme deconvolution (XD) is applied to find chemically coherent groups. In Section 4, we evaluate the contributions of the 17  $[\text{X}/\text{Fe}]$  ratios to the variance of the  $\mathcal{C}$ -space, which is examined in Section 5.

## 2 DATA

We first present the GALAH+ (GALactic Archaeology with HERMES) DR3 (Buder et al. 2021) data and then discuss the chemical elements that will be included to break down degeneracy in  $\mathcal{C}$ -space. The GALAH survey aims to explore the chemical and kinematic histories of the MW. Using the HERMES spectrograph at the Australian Astronomical Observatory’s 3.9-m Anglo-Australian Telescope, GALAH is a high resolution ( $R \approx 28\,000$ ), optical survey. One of its core strengths is its large amount of data and determination of chemical abundances for up to 30 different elements for five

different nucleosynthesis channels. Their abundance determination workflow uses spectroscopy made easy (SME), which is a spectral synthesis fitting code (Piskunov & Valenti 2017).

In this work, we select 17  $[X/Fe]$  ratios ( $X = \text{Mg, Si, Ca, Ti, Ba, La, Y, Eu, Mn, Zn, Co, Cr, Cu, Ni, Al, K, and Sc}$ ), which were taken from an initial set of 30 provided by GALAH. These abundance ratios have been chosen for their reliability and ability to capture key nucleosynthesis signatures in metal-poor stars. Abundance ratio measurements that are not measured in the imposed quality cuts are not included. Some elements were deliberately not included in this analysis: O and Na [both are light elements that are affected by internal mixing and depletion (Cohen 2004)]. Additionally  $[Fe/H]$ , which is the iron abundance, was not included as  $[Fe/H]$  is highly correlated with the abundance ratios  $[X/Fe]$  that are considered in this work. We briefly summarize the elements included here as follows:

(i)  $[Mg, Si, Ca, Ti/Fe]$  – The  $[\alpha/Fe]$  ratio is widely used to differentiate between CCSNe, which enrich the ISM with  $\alpha$ -elements on short time-scales, and SNeIa, which contribute iron-peak elements across longer time-scales. The play-off between these nucleosynthesis processes leads to a characteristic ‘knee’ in the  $[Fe/H]$ – $[\alpha/Fe]$  plane, the location of which can indicate the mass of the system, as more massive galaxies retain and accumulate more metals before SNeIa begin to dominate (De Boer et al. 2014; Hendricks et al. 2014). Previous works highlight the chemical dichotomy of  $[\alpha/Fe]$  in both the disc and halo of the MW, acting as key indicators of SFH and the timing of chemical enrichment (Tinsley 1979).

(ii)  $[Ba/Fe]$  – As a *hs*-element, Ba features supersolar values in systems with extended SFHs, indicative of enhancement due to AGB nucleosynthesis around 10 Myr after star formation (De Los Reyes et al. 2022). At low metallicities, Ba is synthesized via the *r*-process, contributing to the chemical diversity observed in ancient stellar populations (Busso et al. 1999; Frebel 2018).

(iii)  $[La, Y/Fe]$  – Observations suggest that *ls* elements exhibit more scatter than their heavy counterparts and are more overabundant at low metallicities (Burriss et al. 2000; Aoki et al. 2005). This scatter could be indicative of varied nucleosynthetic sites and processes contributing to their abundances, reflecting different chemical evolution paths.

(iv)  $[Eu/Fe]$  – Eu is predominantly produced by the *r*-process and serves as an excellent tracer of neutron-capture processes. The sources of the *r*-process are currently thought to be CCSNe, which contributes to the elevated Eu-enhancement observed at low metallicities, as well as neutron star mergers (NSMs). NSMs occur on similar delay time-scales to SNeIa (Wanajo, Hirai & Prantzos 2021), which leads to an overall increase of Eu abundance in time. Although studies indicate that NSMs are rare events, the observed *r*-process abundance trends in dwarf galaxies can predominantly result from stochastic effects (Beniamini, Hotokezaka & Piran 2016; Ji et al. 2016; Carrillo et al. 2022). This pattern is evidenced in dwarf galaxies from the Local Group such as Carina and Draco, which exhibit notable Eu enrichment ( $[Eu/Fe] \approx 0.5$  dex) as observed by Lanfranchi et al. (2006).

(v)  $[Mn/Fe]$  – Mn is a more pristine tracer of SNe Ia compared to iron, which has other formation channels. Studies have shown that  $[Mn/Fe]$  is underabundant in low-metallicity stars and increases towards solar values at higher metallicities (Feltzing, Fohlman & Bensby 2007; North et al. 2012).

(vi)  $[Zn, Co/Fe]$  – Zn and Co are interesting for their roles in hypernovae (HNe) nucleosynthesis, providing insights into the high-energy end of stellar end-of-life processes (Kobayashi, Karakas & Lugaro 2020). High  $[Zn/Fe]$  ratios in very metal-poor stars suggest

CCSNe as a major site for Zn production but the required quantities are more consistent with energetic HNe. In later phases of galactic evolution, the major Zn-production site likely switches from HNe to SNeIa. This dual contribution led to the solar isotope composition of Zn and solar  $[Zn/Fe]$  values in the disc (Tsujiimoto & Nishimura 2018). Low  $[Zn/Fe]$  can imply formation in an external environment with a low star formation efficiency and where the gas has been poorly enriched by massive stars (Minelli et al. 2021).

(vii)  $[Cr, Cu, Ni/Fe]$  – Cr is primarily produced in SNeIa and CCSNe, with its abundance reflecting the integrated yield of these events over time. Cu, on the other hand, has a more complex nucleosynthetic origin, involving both SNeIa and CCSNe, but also significant contributions from AGB stars. Ni is predominantly synthesized in SNeIa, making its abundance a valuable tracer for these events.

(viii)  $[Al, K, Sc/Fe]$  – Al and K are included for their potential to reveal unique nucleosynthetic signatures from massive stars and AGB stars. The *in situ* and accreted halo appear to exhibit different values of  $[Al/Fe]$ , with metal-poor populations having a lower efficiency of Al production due to inefficient C, N, and O in the ISM (see Kobayashi et al. 2006 for further details).

We select stars with a sufficient spectrum signal-to-noise ratio, with no identified problems with stellar parameter determination, no identified problems with abundance determination of any of the elements we consider, and with a low uncertainty in abundance ratios by imposing the following cuts to the GALAH data (as recommended in Buder et al. 2021):

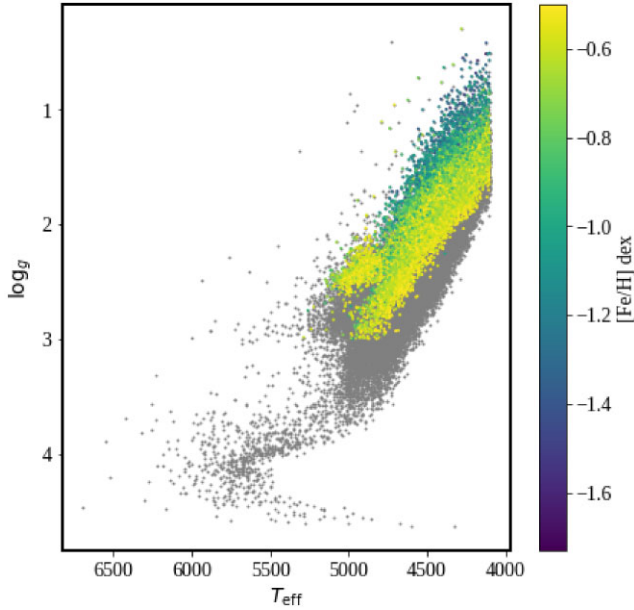
- (i) `snr c3 iraf > 30`
- (ii) `flag sp == 0`
- (iii) `flag Fe H == 0`
- (iv) `flag X Fe == 0`
- (v) `e X Fe < 0.2`

The remaining sample is shown in Fig. 1. Without these cuts, spurious relations can appear and distort the abundance relations by large residual correlations (Jofré, Heiter & Soubiran 2019; Ting & Weinberg 2022). We only select red giant branch (RGB) stars ( $\log g < 3$  and  $\log T_{\text{eff}} < 3.73$ ), due to their high intrinsic luminosity which allows us to probe more distant stars. By conditioning on the stellar parameters  $T_{\text{eff}}$  and  $\log g$ , we also ensure that our sample is not impacted by the different measurement systematics between spectral types (Jofré et al. 2019; de Mijolla et al. 2021). A final additional cut of  $[Fe/H] \leq -0.5$  dex has been made to focus on metal-poor stars. After making these cuts, our sample size is reduced from 678 423 to 9923 stars.

Note that the errors used in the cuts include the error calculated from repeat measurements, while for the rest of the analysis we use only internal SME covariance uncertainties. To provide further clarity on the data quality and the effects of these cuts, in the Appendix, Figs A1 and A2 display the elemental abundances  $[X/Fe]$  and their associated uncertainties as a function of  $T_{\text{eff}}$  and  $[Fe/H]$ . Of all the abundances ratios used,  $[La/Fe]$  has the highest uncertainties.

### 3 METHODS

In the following section, we describe the details of using PCA to find a compact latent space transformation of the high-dimensional  $C$ -space. The transformation allows us to explore the importance of each element in the chemical diversity of the data set and helps the XD clustering method by reducing the dimensionality of the data set (Bovy, Hogg & Roweis 2011; Mukherjee, Doerker & Zhang 2024).



**Figure 1.** The logarithmic surface gravity and effective temperature of GALAH stars. The grey points represent the full sample (without the cuts discussed above), while overlaid points are of our working sample coloured by their metallicity. This limits our sample to RGB stars, which are selected due to their high luminosity.

### 3.1 Principal component analysis

Principal component analysis (PCA) is a dimensionality reduction technique that forms linear combinations of the original variables, ordered by their information content (i.e. contribution to the total variance). By removing those principal components (PCs) with the least information, the data can be simplified in an optimal way, making it easier to visualize and allowing clustering algorithms to find groups more cleanly (Jolliffe & Cadima 2016).

We refer to our analysis as principal component abundance analysis (PCAA, coined by Andrews et al. 2012). Previous studies have shown PCA’s potential to reveal latent correlations between chemical elements, providing insights into nucleosynthetic pathways and the Galaxy’s chemical evolution (Andrews et al. 2012; Ting et al. 2012, 2019). This process yields eigenvectors (PCs) and eigenvalues representing the variance captured by each component. A larger eigenvalue indicates that its corresponding PC captures more variance, holding more ‘information’ about the original data. By projecting on to a subset of the most significant eigenvectors, the data set can be dimensionally reduced.

PCA’s linearity, unlike t-SNE, necessitates that our data reflect this property. The decision to work in  $[X/Fe]$  space is based on two assumptions: (1)  $[X/Fe]$  correlations are linear due to their logarithmic nature; and (2)  $[X/Fe]$  errors are Gaussian distributed.  $[X/H]$  ratios are highly correlated, offering less distinctiveness in the  $\mathcal{C}$ -space. Ting et al. (2012) analysed the dimensionality of  $\mathcal{C}$ -space in various metallicity bands, revealing that neutron-capture abundances significantly contribute to observed variances in low-metallicity ranges ( $-3.5 < [Fe/H] < -1.5$  dex). They demonstrated that the dimensionality of  $\mathcal{C}$ -space, defined by the number of PCs required to encapsulate intrinsic variations, is accurately represented when the cumulative variance of the first  $k$  PCs is around 85 per cent. This threshold is influenced by sample size, observed abundances, and their uncertainties. Unlike 2D chemical hyperplanes, PCA probes

how different elements evolve together, revealing the chemical diversity of the MW.

### 3.2 Extreme deconvolution

In this study, we employ XD, a clustering algorithm, designed to recover an underlying multi-Gaussian distribution, even in the presence of noise. We start with the assumption that the  $N$  observed data points are drawn from a mixture of  $k$  multivariate Gaussian distributions with means  $\mu_k \in \mathbb{R}^d$  and covariances  $\Sigma_k \in \mathbb{R}^{d \times d}$ , where  $d$  is the number of dimensions. For each observed data point  $x_i$ , there is an associated observational error, which can be modelled as a Gaussian. The covariance matrix of this error is represented as  $\chi_i$ . The initial predictions on the fraction of the total  $N$  points that belong to the  $k^{\text{th}}$  Gaussian component are given by  $\pi_k$ , such that  $\sum_{k=1}^K \pi_k = 1$ . The initial estimates for  $\pi_k$ ,  $\mu_k$  and  $\Sigma_k$  are initialized using the `sklearn.mixture.GaussianMixture` package, and are generated using K-Means. The XD algorithm employs the expectation-maximization technique, which consists of two steps: the E-step, where the posterior probabilities  $\omega_{k,i}$  are computed, and the M-step, where the Gaussian parameters are updated.

In the E-step, the algorithm calculates the probability that data point  $x_i$  belongs to the  $k^{\text{th}}$  Gaussian component using Bayes’ theorem:

$$\omega_{k,i} = \frac{p(x_i | \mu_k, \Sigma_k + \chi_i)}{\sum_{j=1}^k \pi_j p(x_i | \mu_j, \Sigma_j + \chi_i)}, \quad (1)$$

where  $p(x_i | \mu_k, \Sigma_k + \chi_i)$  represents the probability density function of the Gaussian distribution with mean  $\mu_k$  and covariance  $\Sigma_k$  evaluated at  $x_i$ . In the M-step, the estimates for  $\mu_k$  and  $\Sigma_k$  are updated. The revised mean  $\mu_k$  for all  $k$  is

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^N \omega_{k,i} x_i, \quad (2)$$

with  $N_k$  being the effective count of data points ascribed to the  $k^{\text{th}}$  Gaussian component. The updated covariance estimate for all  $k$  is

$$\Sigma_k = \frac{1}{N_k} \sum_{i=1}^N \omega_{k,i} (x_i - \mu_k)(x_i - \mu_k)^T - \chi_i. \quad (3)$$

This formula accounts for the observational error, refining the true distribution’s covariance estimate. Lastly, the prior probability  $\pi_k$  is adjusted:

$$\pi_k = \frac{N_k}{N}. \quad (4)$$

The XD algorithm iterates between the E-step and M-step until convergence is reached. XD’s robustness against noise and its ability to discern intricate structures make it particularly valuable in astrophysics. This was shown by Bovy et al. (2011) and Belokurov et al. (2018), to break down the complex velocity distributions of solar neighbourhood stars and the MW stellar halo, respectively.

### 3.3 Group number ( $k$ ) estimation

After employing XD, the next critical task is to determine the optimal number of Gaussian groups, hereafter defined as  $k$ . The Bayesian information criterion (BIC) emerges as a practical tool for this purpose (Fraley & Raftery 2002). It provides a trade-off between the fit of the model to the data and the dimensionality (hereafter defined as  $d$ ) of the model. The formula for BIC is given by

$$\text{BIC} = -2 \ln(L) + m \ln(N), \quad (5)$$

**Table 1.** The compiled results of the XD mock tests, investigating the impact of the dimensionality ( $d$ ), number of groups ( $k$ ), and the errors  $X_{\text{err}}$ .

$d$ -dimensions	$k$ -groups	$X_{\text{err}}$	$k$ -predicted
3	3	0.0, 0.05, 0.10, 0.20	3, 3, 3, 3
3	5	0.0, 0.05, 0.10, 0.20	5, 4, 5, 6
3	10	0.0, 0.05, 0.10, 0.20	5, 5, 5, 4
5	3	0.0, 0.05, 0.10, 0.20	3, 3, 3, 2
5	5	0.0, 0.05, 0.10, 0.20	6, 6, 6, 6
5	10	0.0, 0.05, 0.10, 0.20	11, 11, 11, 8
7	3	0.0, 0.05, 0.10, 0.20	3, 3, 3, 3
7	5	0.0, 0.05, 0.10, 0.20	8, 5, 5, 5
7	10	0.0, 0.05, 0.10, 0.20	11, 6, 9, 9

where  $L$  is the likelihood of the observed data given the model,  $m$  is the number of free parameters in the model and as before,  $N$  is the number of data points.  $L$  is the product of the probabilities of each data point, given the model parameters and is given by

$$L = \prod_{i=1}^N p(\mathbf{x}_i | \theta), \quad (6)$$

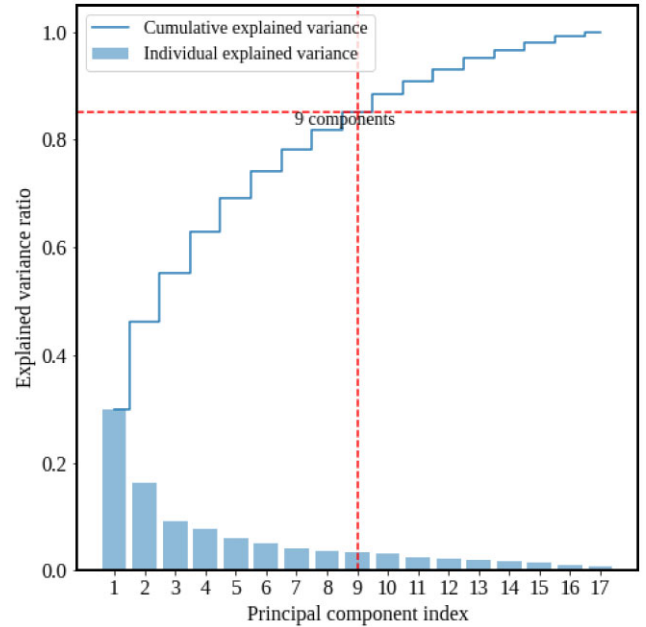
where  $\theta$  denotes the set of parameters of the Gaussian mixture model. The value of  $m$  is given by

$$m = d \frac{d+3}{2} k. \quad (7)$$

The BIC tends to favour models with fewer parameters (simpler models) for smaller data sets and models with more parameters (more complex models) for larger data sets. As a result of this, the BIC penalizes overfitting, discouraging overly complex models. In our study, underestimating the number of groups is more advantageous than overestimating, as during subsequent dynamical analyses these merged groups could be teased apart based on their dynamical properties.

In order to test the accuracy and limitations of XD, we generated a set of mock multivariate distributions, adjusting the number of groups ( $k$ ), means ( $\mu_k$ ), covariances ( $\Sigma_k$ ), weights ( $\pi_k$ ), number of dimensions ( $d = 2, 5, \text{ and } 7$ ) and magnitude of the errors ( $X_{\text{err}} = 0.0, 0.05, 0.10, 0.20$  dex). The summarized results are found in Table 1. These errors were chosen to span the range of errors that are encountered in the GALAH data. To keep these tests applicable to the ranges of our PCA-transformed data, we use the same sample size ( $N = 9923$ ), generating random samples with the following number of artificially included groups: 3, 5, 7, and 10. XD performs well in recovering the correct number of groups,  $k$ , but this performance is influenced by both the dimensionality,  $d$ , of the space and the magnitude of  $k$ . In a 3D space with three groups ( $d = 3, k = 3$ ), the number of groups predicted by the BIC,  $k_{\text{predicted}}$ , is accurate. However, when the number of groups is increased to 10 in the same 3D space the prediction becomes less reliable, resulting in  $k_{\text{predicted}} = 4$ , when the errors are set to  $X_{\text{err}} = 0.2$  dex. As the dimensionality increases, the accuracy of the group recovery seems to improve, for example, in the case of  $d = 7$  and  $k = 10$ , the BIC is able to predict a range of values close to the actual number,  $k_{\text{predicted}} = 6-11$ .

This suggests that higher dimensional spaces provide better separation between groups, allowing for more accurate predictions by the BIC. In summary, when looking at various dimensions and group numbers,  $k_{\text{predicted}}$  does not consistently indicate a tendency toward overfitting or underfitting our mock tests. Nevertheless, in cases where  $k = 10$ , which is the number used in this work as described in Section 4.2, the predicted number of groups falls short across all dimensions, indicating a tendency towards underfitting.



**Figure 2.** The explained variance of a PC is the eigenvalue of that PC. Applying a cut-off at 85 per cent cumulative explained variance shows that nine components are required to reliably capture the patterns present in the data. Thus, a dimensionality reduction can be made from a 17D to a 9D space.

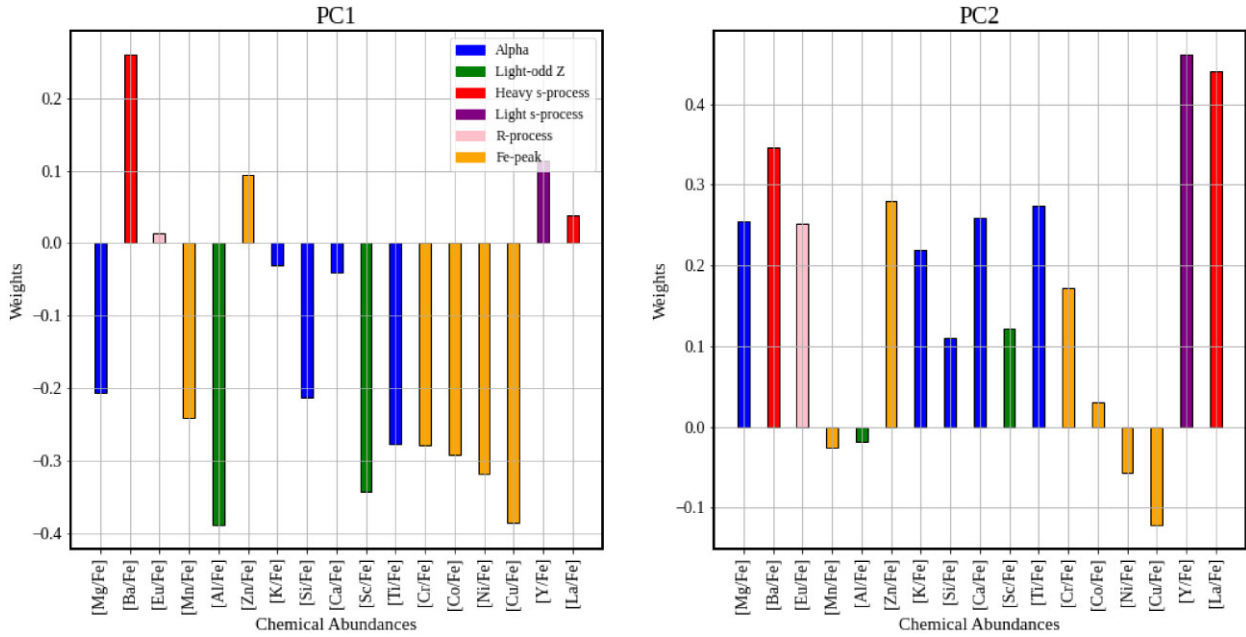
Consequently, we conclude that when applying this method to our data, the BIC score’s ‘knee’ (the point where adding more groups does not significantly improve the model) represents a lower limit of the true number of groups in our data set, so is unlikely to subdivide the  $\mathcal{C}$ -space into too many groups.

## 4 RESULTS

Here, we describe the results of applying PCA and our clustering analysis, followed by a presentation of the chemodynamical properties of the revealed groups- assigning known stellar components to the 10 derived groups as follows: the thick disc (G1, G4, G6, G7, and G8), GES (G2), the *in situ* halo (G3 and G5) and the thin disc (G9). G10 is interpreted as an outlier group.

### 4.1 PCAA

PCA is applied to the 17D  $\mathcal{C}$ -space, with the explained variance shown in Fig. 2. The explained variance is the fraction of the total variance in the original data set that is contained in each PC. Consistent with the approach of Ting et al. (2012), we have implemented a cumulative explained variance cutoff of 85 per cent, thereby excluding those PCs dominated by noise or errors. Notably, the cumulative explained variance increases gradually, requiring a substantial number of PCs to reach the 85 per cent threshold. This pattern could suggest, within the metallicity range of  $-1.7 \leq [\text{Fe}/\text{H}] \leq -0.5$ , the presence of multiple independent nucleosynthetic processes that influence the observed chemical patterns, as well as different SFHs contributing to the enrichment of this region. These aspects are explored further in Section 5. Implementing the 85 per cent cutoff reduces the 17D space to a 9D space, highlighted in Fig. 2. We note that changing the number of PCs we consider by a few does not significantly affect the number of groups that are identified by our subsequent clustering analysis in Section 4.2.

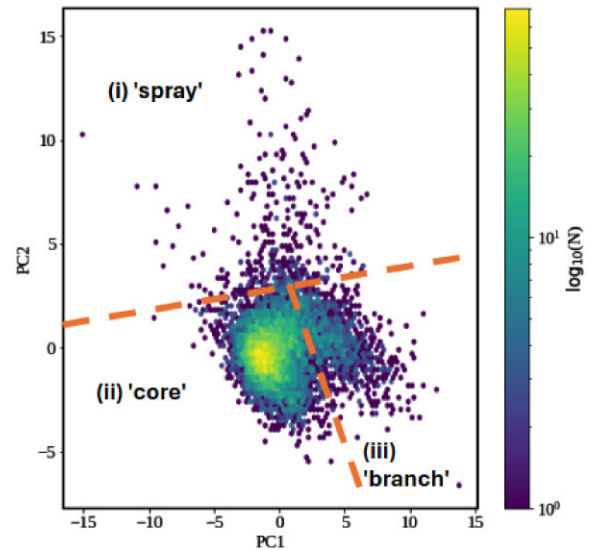


**Figure 3.** The 17  $[X/Fe]$  abundances are coloured approximately by their nucleosynthetic families:  $\alpha$  (blue), light-odd  $Z$  (green), heavy  $s$ -process (red), light  $s$ -process (purple),  $r$ -process (pink), and iron-peak (orange). The first PC (left plot) makes up 29 per cent of the total variance of the  $C$ -space. The second PC (right plot) makes up 18 per cent of the total variance. The weights (positive and negative) show how each  $[X/Fe]$  chemical abundance ratio contributes to the PC.

To further breakdown the elemental contributions to the variance, Fig. 3 shows the weights (i.e. eigenvector coefficients for each element enhancement) for the first two PCs, which contain 50 per cent of the variance of the data. In PC1,  $[Al/Fe]$ ,  $[Cu/Fe]$ , and  $[Ba/Fe]$  contribute the most to the variance (in terms of absolute magnitude). In PC2, there is a greater contribution from the  $ls$ -elements,  $[Y/Fe]$ , and  $[La/Fe]$ . The lack of contribution, positive or negative, from elements like  $[Eu/Fe]$  and  $[K/Fe]$  in PC1, and  $[Mn/Fe]$ ,  $[Al/Fe]$ ,  $[Co/Fe]$ ,  $[Ni/Fe]$ , and  $[Cu/Fe]$  in PC2, is a consequence of the non-linearity of some nucleosynthesis processes, for example, Al, where two trends can cancel out. To see more clearly how these elements contribute to the spread of observed abundances, we show the PC1–PC2 plane in Fig. 4. We use the following terms to define the three general regions that are identifiable by eye in this 2D plane; the ‘core’, the ‘branch’ and the ‘spray’. We argue in the following that the ‘core’ is made up of disc populations, the ‘branch’ is made up from a single accreted population (GES) and the ‘spray’ features supersolar  $s$ -process enhanced stars which are *in situ* in origin. These regions are used for reference when analysing the distributions of each  $[X/Fe]$  abundance ratio in the PC1–PC2 plane in Fig. 5. We can also identify that the positive PC1 direction varies between disc (‘core’) and accreted stars (‘branch’), with  $hs$ -element  $[Ba/Fe]$  having a positive contribution and iron-peak and light odd- $Z$  elements having a negative contribution which points towards metal-poor populations with extended SFHs. The positive PC2 direction varies between disc and *in situ* populations (‘spray’), pointing towards  $s$ -process and  $\alpha$ -enhancement.

Our main findings are as follows:

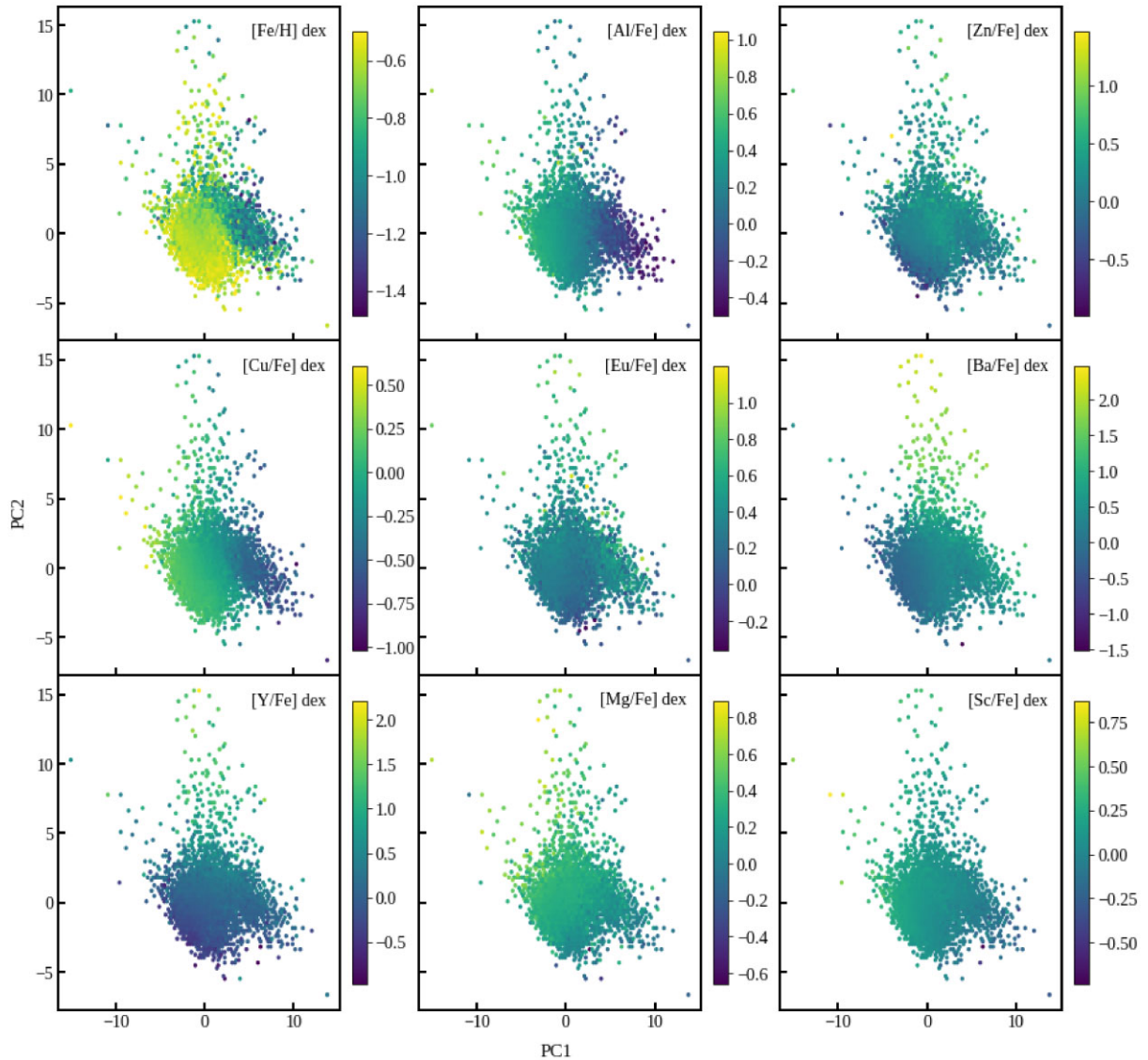
(i) Core: The ‘core’ is characterized by relatively metal-rich stars with  $[Fe/H] \approx -0.6$  dex, noting that the sample is selected to have  $[Fe/H] \leq -0.5$  dex. This region shows significant enhancement in  $[Al/Fe]$ , indicating differences in SFHs compared to other regions with the ‘core’ experiencing recent/ongoing star formation compared with the accreted populations in the ‘branch’. The majority of



**Figure 4.** The ‘core’ is the group centred around  $PC1 \approx 0$  and  $PC2 \approx -2$ . The ‘branch’ is the group centred around  $PC1 \approx 5$  and  $PC2 \approx 0$ . The ‘spray’ is the group centred around  $PC1 \approx 0-1$  and  $PC2 \approx 10$ .

the ‘core’ shows enhanced  $[Cu/Fe]$ , except for the lower region occupying  $PC1 \approx -2$ ,  $PC2 \approx -4$ . The ‘core’ has near-solar values of  $[Ba/Fe]$ , suggesting it is representative of disc populations. There is a lower  $[Mg/Fe]$  value in the lower region of the ‘core’ compared with the upper ‘core’, with a difference of  $\approx 0.3$  dex, which is greater than the average errors of  $[Mg/Fe]$  (0.07 dex). This could indicate the presence of thin disc stars. Additionally, some  $[Sc/Fe]$  enhancement is observed in the left region of the ‘core’ (at  $PC1 \approx -5$  and  $PC2 \approx 0$ ).

(ii) Spray: The ‘spray’ region displays a complex mixture of metallicities. It is notable for  $hs$ -enhancement in  $[Ba/Fe]$  and supersolar  $[Y/Fe]$  stars, indicating a distinct chemical signature from



**Figure 5.** The PC1–PC2 hyperplane coloured by various abundances ratios ([Fe/H], [Al/Fe], [Zn/Fe], [Cu/Fe], [Eu/Fe], [Ba/Fe], [Y/Fe], [Mg/Fe], and [Sc/Fe]). The only abundance ratio here that was not included in the  $C$ -space is [Fe/H]. We evaluate the chemical enrichment in three regions: the ‘core’, the ‘branch’, and the ‘spray’.

both the ‘core’ and ‘branch’. PC2 shows a high positive weighting for  $s$ -process elements towards the ‘spray’, suggesting significant contributions from AGB stars in these populations. This could also indicate the presence of Ba-rich stars resulting from binary processes. Higher AGB contributions imply these populations underwent prolonged or multiple-phase star formation, allowing intermediate-mass stars to evolve into AGB stars, enriching the ISM with  $s$ -process elements. The ‘spray’ has a varied range of abundance ratios for the other elements considered, showing no clear pattern like the ‘core’ or ‘branch’, which makes it a more heterogeneous group in terms of chemical composition.

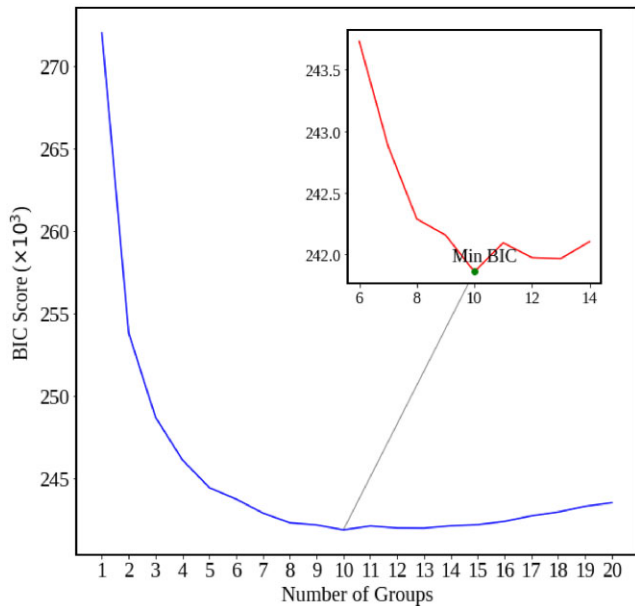
(iii) Branch: The ‘branch’ contains more metal-poor stars, with  $[\text{Fe}/\text{H}] \approx -1.4$  dex. It exhibits significant variance in [Eu/Fe], pointing to stochasticity in  $r$ -process element production in the metal-poor regime. The ‘branch’ exhibits lower [Al/Fe] than the ‘core’ or ‘spray’ regions, indicating significant differences in SFHs, with low [Al/Fe] being a property of accreted populations. Many of the iron-peak elements have negative weights in PC1, which can be seen through the

sub-solar values of elements such as [Cu/Fe] towards the ‘branch’. A similar trend is observed in [Cu/Fe], where the lower [Cu/Fe] agrees with findings for accreted halo stars. The ‘branch’ also has high [Ba/Fe] values, and a slightly sub-solar [Mg/Fe] present in the lower region, potentially indicative of low- $\alpha$  halo stars. In PC1, [Ba/Fe], [Zn/Fe], [Y/Fe], and [La/Fe] have positive weights, indicating these elements increase towards the accreted population in the ‘branch’.

The PC1–PC2 hyperplane is just a 2D representation of the reduced 9D  $C$ -space, which indicates why there are only three components selected by eye (see Fig. 4). This will be explored in more detail in the following Section 4.2, where we cluster the latent space using XD to estimate the underlying number of groups.

## 4.2 Clustering the PCA-transformed space

Our goal is to cluster the metal-poor MW stars into groups with similar chemistry to better understand the chemical diversity and



**Figure 6.** BIC scores for clustering the 9D-transformed  $\mathcal{C}$ -space of metal-poor MW stars. The minimum BIC, the ‘knee’, suggests 10 groups. The inset shows the minimum BIC score at 10 groups. Although 10 to 14 groups yield similar scores, we proceed with 10 groups for simplicity and robustness.

formation history of components such as the thick disc, thin disc, and stellar halo. We apply XD to our 9D-transformed  $\mathcal{C}$ -space for a varying number of groups and calculate the BIC for each case. The BIC scores shown in Fig. 6 suggest the presence of  $\approx 10$  groups, which we label as G1–G10. It is important to note that the optimal number of groups is not very clear from the BIC analysis, as a range of about 10 to 14 groups returns similar BIC scores. For simplicity and to facilitate our analysis, we proceed with the minimum number of groups suggested by the BIC knee, which is 10, noting also that a slight underestimation of the total number of groups is preferable to an overestimation, as discussed in Section 3.3. Additionally, we only assign stars to a group if they have a  $\geq 70$  per cent probability of belonging to that Gaussian component.

### 4.3 Mapping the groups to known populations

In Fig. 7, we show the location of the groups in commonly used chemodynamical hyperplanes,  $[\text{Mg}/\text{Mn}] - [\text{Al}/\text{Fe}]$ ,  $L_z - E$  and  $L_z/L_c - [\text{Fe}/\text{H}]$ , while also showing where the groups lie in the reduced PC1–PC2 space. While  $[\text{Mg}/\text{Fe}]$  has a relatively low weight of  $-0.2$  in PC1 (shown in Fig. 3), indicating it does not vary significantly, we still utilize it to explore its associations with the thick disc, thin disc and halo populations. Combining this with information from the other  $[\text{X}/\text{Fe}]$  distributions for each group, found in the Appendix, we can determine the chemodynamical properties of each of them. We also refer the reader to Table A1 for a complete collection of means and errors of the chemodynamical distributions and Fig. A3 for the  $[\text{X}/\text{Fe}]$  abundance distributions for each group.

#### 4.3.1 G1, G4, G6, G7, and G8 – thick disc groups

G1, G4, G6, G7, and G8 contain 763, 648, 2235, 832, and 610 stars, respectively. In the PC1–PC2 plane, these groups form the ‘core’. They occupy an  $\alpha$ -rich region with  $[\text{Mg}/\text{Fe}]_\mu$  values of  $\approx 0.3$  dex. Their  $[\text{Fe}/\text{H}]$  distributions show the most metal-enhanced

values among the groups of around  $-0.6$  dex. All have prograde orbits ( $L_z \approx 1200$  kpc km s $^{-1}$ ) and  $e$  distributions peaking at  $\approx e = 0.30$ .

Regarding their chemistry, the  $[\text{Ba}/\text{Fe}]$  values differ: G4 and G8 have supersolar values ( $[\text{Ba}/\text{Fe}]_\mu = 0.28$  and  $0.21$  dex respectively), while G6 and G7 have slightly sub-solar values ( $[\text{Ba}/\text{Fe}]_\mu = -0.05$  and  $-0.10$  dex, respectively). The  $[\text{Y}/\text{Fe}]$  trend is similar.  $[\text{Eu}/\text{Fe}]$  abundances are consistently supersolar across these groups ( $\approx 0.3$  dex). G4 shows a high Zn enhancement ( $[\text{Zn}/\text{Fe}]_\mu = 0.42$  dex), while others are closer to solar ( $\approx 0.1$  dex). G7 and G8 show bimodalities in  $[\text{Zn}/\text{Fe}]$  with solar and sub-solar components (e.g.  $G7_{\text{sub-solar}} = -0.3$  and  $G7_{\text{solar}} = 0.1$  dex). The  $[\text{Al}/\text{Fe}]$  distribution is narrow across all groups, with abundances ranging from  $[\text{Al}/\text{Fe}]_\mu = 0.22 - 0.45$  dex. For  $[\text{Cu}/\text{Fe}]$ , all groups have solar values except G4, which has  $[\text{Cu}/\text{Fe}]_\mu = -0.08$  dex. In  $[\text{Ni}/\text{Fe}]$ , all groups have supersolar Ni-abundances except G4 ( $[\text{Ni}/\text{Fe}]_\mu = 0.00$  dex). Finally, the  $[\text{Sc}/\text{Fe}]$  distributions are similar across G4, G6, and G7, with tails extending between  $[\text{Sc}/\text{Fe}] = 0.0 - 0.2$  dex. G1 has an extended distribution with  $[\text{Sc}/\text{Fe}]$  reaching  $\approx 0.4$  dex. G8 is the only group with sub-solar values,  $[\text{Sc}/\text{Fe}]_\mu = 0.04$  dex, and a tail extending down to  $\approx -0.2$  dex. These groups (G1, G4, G6, G7, and G8) lie in the thick disc region of the  $[\text{Mg}/\text{Mn}] - [\text{Al}/\text{Fe}]$  plane.

The properties of these groups, including their  $\alpha$ -rich nature, relatively high metallicity, prograde orbits, and specific chemical abundance patterns, align with the expected SFH of the thick disc. This history involves rapid early star formation from gas enriched by CCSNe, followed by prolonged or multiple-phase star formation episodes (Bland-Hawthorn & Gerhard 2016; Yu et al. 2021). G4 and G8 have some exceptions to expected abundance patterns, in the case of  $[\text{Ba}/\text{Fe}]$  and  $[\text{Zn}/\text{Fe}]$ , this is further discussed in Section 5.1.

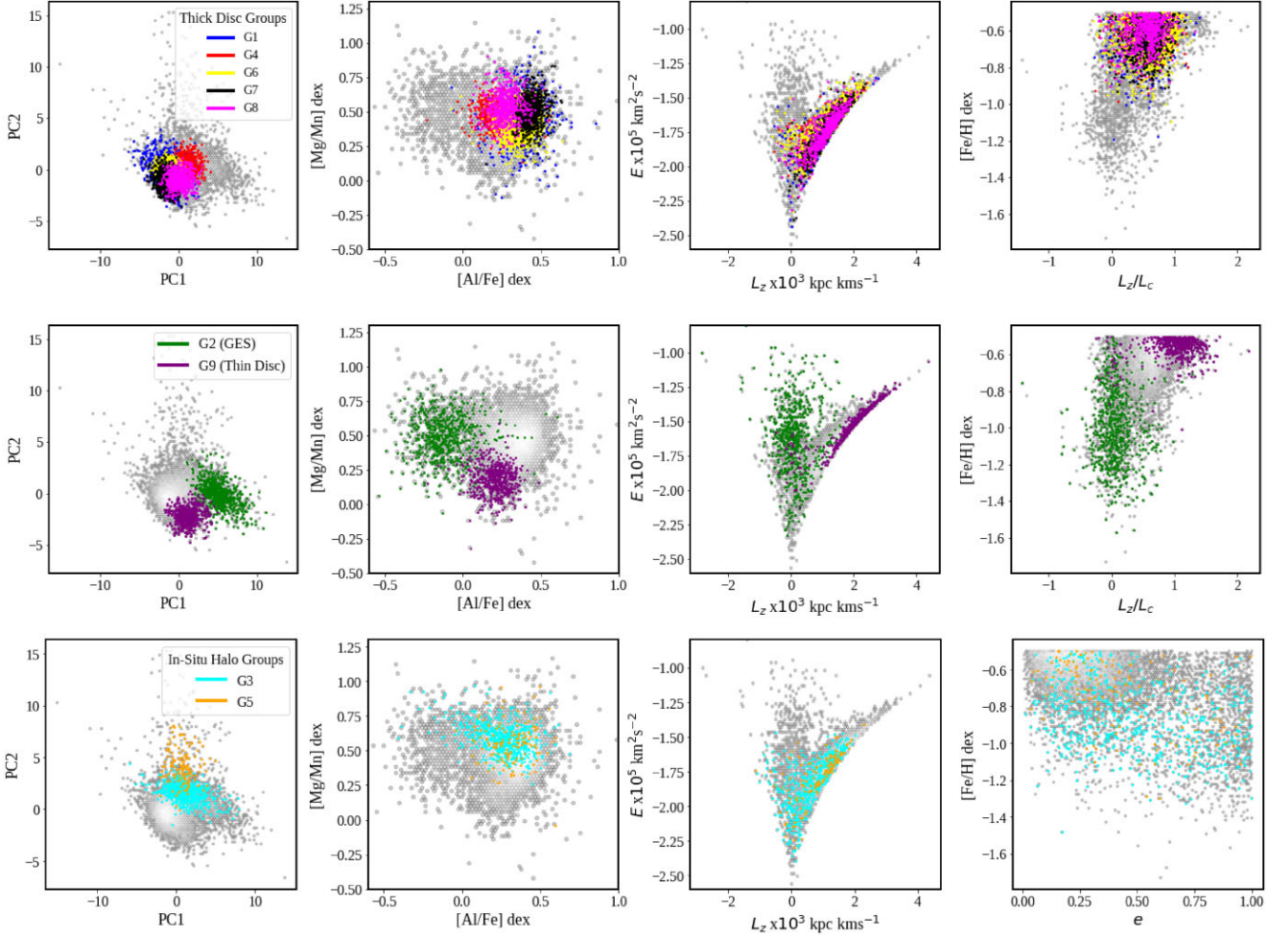
#### 4.3.2 G2–GES

G2 contains 671 stars and dominates the ‘branch’ in the PC2–PC1 plane. G2 has low-metallicity values ( $[\text{Fe}/\text{H}]_\mu = -1.00$  dex) and  $L_{z,\mu} = 145.13 \pm 521.97$  kpc kms $^{-1}$ , with an eccentricity  $e_\mu = 0.80 \pm 0.21$ , extending up to a peak at  $e \approx 1$ , making it the most dynamically hot of all groups.

It is  $\alpha$ -rich, slightly less so than the groups in the ‘core’, with  $[\text{Mg}/\text{Fe}]_\mu = 0.13$  dex, and is the most Mn-depleted group present, with  $[\text{Mn}/\text{Fe}]_\mu = -0.37$  dex. G2 has a high Ba enhancement,  $[\text{Ba}/\text{Fe}]_\mu = 0.41$  dex, though the ls-process element Y enhancement is  $[\text{Y}/\text{Fe}]_\mu = 0.11$  dex, indicating a lower abundance compared to the hs-process. G2 shows supersolar  $[\text{Eu}/\text{Fe}]$  values at  $\mu = 0.44$  dex and a uni-modal, solar  $[\text{Zn}/\text{Fe}]$  distribution with  $[\text{Zn}/\text{Fe}]_\mu = 0.08$  dex. The  $[\text{Al}/\text{Fe}]$  distribution is subsolar, the lowest across all groups, with  $[\text{Al}/\text{Fe}]_\mu = -0.12$  dex. Notably, G2 has a dominant peak at very sub-solar  $[\text{Cu}/\text{Fe}]$  values,  $[\text{Cu}/\text{Fe}]_\mu = -0.46$  dex, making it the least Cu-enriched in the sample. The  $[\text{Ti}/\text{Fe}]$  distribution shows that while G2 is enhanced in Mg (which is also an  $\alpha$ -element), it extends to  $[\text{Ti}/\text{Fe}]$  abundances as low as those of G9. G2 has  $[\text{Ni}/\text{Fe}]_\mu = -0.11$  dex, and  $[\text{Sc}/\text{Fe}]_\mu = -0.01$  dex.

In the  $[\text{Mg}/\text{Mn}] - [\text{Al}/\text{Fe}]$  plane, G2 overlaps with the ‘blob’ identified in Das et al. (2020) and also mirrored by other studies (Myeong et al. 2022), indicating the presence of GES stars. The low metallicity of these stars position G2 within the more metal-rich extent of GES ( $[\text{Fe}/\text{H}]_\mu = -1.05 \pm 0.20$  dex (Ortigoza-Urdaneta et al. 2023), although GES is thought to extend down to  $[\text{Fe}/\text{H}] \approx -1.5$  dex (Naidu et al. 2022), a metal-poor region not fully captured in our sample, which extends down to  $[\text{Fe}/\text{H}] \approx -0.6$  dex. G3 and G10 also show GES-like stars, with G10 reaching lower





**Figure 7.** Plots of all groups in varying chemodynamical and PC planes. Top panel: the thick disc groups, G1, G4, G6, G7, and G8, in order from left to right: PC2 – PC1, [Mg/Mn] – [Al/Fe],  $L_z - E$  and  $L_z/L_c - [Fe/H]$ . Middle panel: GES (G2) and the thin disc (G9), in order from left to right: PC2 – PC1, [Mg/Mn] – [Al/Fe],  $E - L_z$  and  $[Fe/H] - L_z/L_c$ . Bottom panel: the *in situ* halo groups (G3 and G5) in order from left to right: PC2 – PC1, [Mg/Mn] – [Al/Fe],  $E - L_z$ , and  $[Fe/H] - e$ .

metallicities. The high [Ba/Fe] values in G2 present a discrepancy with studies that indicate GES has negligible s-process contribution and subsolar [Ba/Fe] values ( $-0.02$  dex) (Carrillo et al. 2022). However, our  $[Ba/Fe]_\mu = 0.41 \pm 0.19$  dex closely matches Myeong et al. (2022)’s estimate for metal-rich GES ( $[Ba/Fe] = 0.38$  dex,  $[Fe/H] = -0.96$  dex). G2 also shows r-process enhancement with  $[Eu/Fe]_\mu = 0.44$  dex, aligning with the metal-rich GES cited in the literature ( $[Eu/Fe] = 0.47$  dex (Myeong et al. 2022)). The subsolar  $[Al/Fe]_\mu = -0.12$  dex is consistent with an accreted population, matching Myeong et al. (2022)’s metal-rich GES ( $[Al/Fe] = -0.17 \pm 0.09$  dex). G2’s extreme subsolar [Cu/Fe] values define it distinctly.

#### 4.3.3 G9 – thin disc

G9 contains 581 stars and is part of a distinct lower cluster in the segmented ‘core’ shown in Fig. 4. G9 has the highest  $L_z$  among the groups, with  $L_z = 2020.10 \pm 619.82$  kpc km s $^{-1}$ . There is a slight bi-modality in  $L_z$ , with a small group of stars having  $L_z \sim 0$ , and it has the most circular orbits of all the groups, with a distinct peak at  $e \approx 0.1$ . G9 has a tail towards  $L_z \leq 0$ , with 9 stars exhibiting retrograde kinematics, representing  $\approx 1.72$  per cent

contamination from G2. G9 is the most  $\alpha$ -poor group in the sample, with  $[Mg/Fe]_\mu = 0.10$  dex, supported by low [Ti/Fe] values ( $[Ti/Fe]_\mu = 0.09$  dex). It has the most overall Mn-rich distribution of the groups, with  $[Mn/Fe]_\mu = -0.08$  dex, with only G1 and G10 showing extended tails to higher Mn-enhancement. The [Fe/H] distribution is the narrowest and also among the most metal-rich ( $[Fe/H]_\mu = -0.56$  dex). The s-process enhancement shows slight bimodality, more evident in the ls-process element Y, with [Y/Fe] featuring peaks at solar ( $\approx 0.1$  dex) and subsolar ( $\approx -0.2$  dex) values. G9 has the lowest r-process enhancement with  $[Eu/Fe]_\mu = 0.15$  dex). The [Zn/Fe] values are solar,  $[Zn/Fe]_\mu = -0.01$  dex, and feature a tail extending to low [Zn/Fe] values ( $< -0.5$  dex). [Al/Fe] aligns with the ‘core’ groups, and the [Cu/Fe] distribution closely resembles that of G1, G6, and G8. G9 has the lowest [Ti/Fe] of all groups,  $[Ti/Fe]_\mu = 0.09$  dex, with a tail extending to subsolar values,  $\approx -0.2$  dex, similar to G2. The [Ni/Fe] distribution is similar to G1, G4, G6, and G7 ( $[Ni/Fe]_\mu = 0.04$  dex), and [Sc/Fe] values closely match G8, at  $[Sc/Fe]_\mu = 0.06$  dex. In the [Mg/Mn]–[Al/Fe] plane, G9 lies in a region characteristic of the thin disc. This is corroborated by its predominantly prograde and circular orbits. The  $\alpha$ -poor nature and narrow metallicity distribution at  $[Fe/H] > -0.8$  dex are consistent with the thin disc population studied by

Hawkins et al. (2015). The retrograde contamination (1.72 per cent) likely originates from GES, as they possess a similar  $\alpha$  enhancement.

#### 4.3.4 G3 and G5 – *in situ halo*

G3 and G5 contain 651 and 141 stars, respectively. They both occupy the upper region of the ‘core’ (at PC1  $\approx$  0 and PC2  $\approx$  5), with G3 extending into the ‘branch’ and G5 extending into the ‘spray’. It is evident that these two groups overlap across multiple hyperplanes, capturing stars on retrograde and more circular orbits. G3 features a bi-modality at  $L_z \approx 1000$  and  $0 \text{ kpc km s}^{-1}$ . The prograde peak aligns with G1, G4, G6, G7, and G8 (i.e. the thick-disc populations), while the  $L_z \approx 0 \text{ kpc km s}^{-1}$  peak matches G2 (GES). G5 also shows a bimodal distribution in  $L_z$ , with stars at  $L_z = 0$  and retrograde values. The  $e$  distribution in both groups is widely spread between  $e = 0$  and 1, with a bi-modality of thick disc-like and higher eccentricity orbits.

The [Mg/Fe] distribution of G3 is centred at  $[\text{Mg/Fe}]_\mu = 0.32$  dex, similar to the  $\alpha$ -enhanced G1, G4, G6, G7, and G8 (thick-disc populations). G5 has the most  $\alpha$ -rich stars of all groups, with  $[\text{Mg/Fe}]_\mu = 0.36$  dex, and a distribution tail extending to  $\approx 0.75$  dex. Both groups have wide metallicity distributions, with G5 being more metal-rich ( $[\text{Fe/H}]_\mu = -0.75$  dex), and G3 having a lower metallicity ( $[\text{Fe/H}]_\mu = -0.87$  dex). The Ba-enrichment of G3 is similar to G2 ( $[\text{Ba/Fe}]_\mu = 0.44 \pm 0.24$  dex), while G5 has the highest [Ba/Fe] enrichment among all groups ( $[\text{Ba/Fe}]_\mu = 0.97 \pm 0.37$  dex). G3 has a narrow [Y/Fe] distribution ( $[\text{Y/Fe}]_\mu = 0.33 \pm 0.17$  dex), unique among the groups, while G5 has a higher enhancement,  $[\text{Y/Fe}]_\mu = 0.68 \pm 0.26$  dex, with a complex distribution featuring about four peaks. The high s-process enhancements present in G3 are found in the protogalaxy population Aurora (Myeong et al. 2022), and are attributed to the prolonged star formation of *in situ* populations. Similarly, Heracles (Horta et al. 2020; Naidu et al. 2022), an ancient merger event prior to the GES merger has similar s-process properties to G3. The Eu-enhancement of G3 closely resembles G1 and G6 (thick-disc groups), featuring a tail of Eu-enhanced stars trailing up to  $[\text{Eu/Fe}] > 0.7$  dex. G5 has a similar distribution but with a slightly higher mean enhancement ( $[\text{Eu/Fe}]_\mu = 0.39$  dex), and a tail extending up to  $[\text{Eu/Fe}] \approx 0.8$  dex.

The [Al/Fe] distribution are 0.25 and 0.31 dex for G3 and G5, respectively, and are too high to be considered from Heracles, which is limited to  $[\text{Al/Fe}] < 0$  (Horta et al. 2020). This range, however, matches well with the *in situ* halo population Aurora, with  $[\text{Al/Fe}] = 0.10 \pm 0.16$  dex (Myeong et al. 2022). For [Zn/Fe], G3 has  $[\text{Zn/Fe}]_\mu = 0.19$  dex and G5 has  $[\text{Zn/Fe}]_\mu = 0.25$  dex, similar to G1, G6, and G8. Both G3 and G5 extend to the lowest [Cu/Fe] values ( $\approx -0.5$  dex), with G3 having the lowest mean ( $[\text{Cu/Fe}]_\mu = -0.14$  dex). The [Ti/Fe] distributions for G3 and G5 are similar to G1, G6, and G7 (thick-disc groups), with  $[\text{Ti/Fe}]_\mu = 0.26$  dex and  $[\text{Ti/Fe}]_\mu = 0.28$  dex respectively, which is also seen in the case of [Ni/Fe]. The [Sc/Fe] distributions of both G3 and G5 closely match G9 and G8, with  $[\text{Sc/Fe}]_\mu = 0.08$  and  $0.11$  dex, respectively. Retrograde contamination in these groups is 99 stars for G3 (15.2 per cent of the group) and 15 stars for G5 (10.6 per cent of the group). Located in the [Mg/Mn]–[Al/Fe] plane where *in situ* halo stars are found, G3 and G5 are distinguished by their s-process enhancement in [Ba/Fe] and [Y/Fe].

#### 4.3.5 G10 – *outlier group*

G10 contains 365 stars and is an outlier group, overlapping into the ‘branch’, ‘spray’, and ‘core’. G10 appears as a background component in all hyperplanes. In  $L_z$ , G10 has two peaks with a greater spread

towards prograde values ( $L_z > 2000 \text{ kpc km s}^{-1}$ ), and a wide  $e$  distribution. G10 shows a broad range of [Mg/Fe] values from  $-0.75$  to  $0.75$  dex, grouping together some of the most  $\alpha$ -rich and  $\alpha$ -poor stars in the data set. It also has a spread of extreme [Ba/Fe] values, likely captured in the ‘spray’, with hs-process enhancement ( $[\text{Ba/Fe}] > 2.0$  dex). [Y/Fe] shows a similar spread. The [Eu/Fe] distribution has wide dispersion and tails to very high Eu-enhancements ( $[\text{Eu/Fe}] \approx 1.0$  dex). In contrast, [Ni/Fe] and [Sc/Fe] distributions are narrower than other chemical abundances, with  $[\text{Ni/Fe}]_\mu = 0.00 \pm 0.16$  dex and  $[\text{Sc/Fe}]_\mu = 0.12 \pm 0.17$  dex, showing similarities to G5 in both abundances. Kinematically, G10 includes approximately two distinct groups that align with the thick disc kinematics of G1, G4, G6, G7, and G8, as well as the accreted,  $L_z \approx 0$ , G2 group. The wide distribution of chemical and kinematic features contrasts with its tight distributions in both [Ni/Fe] and [Sc/Fe].

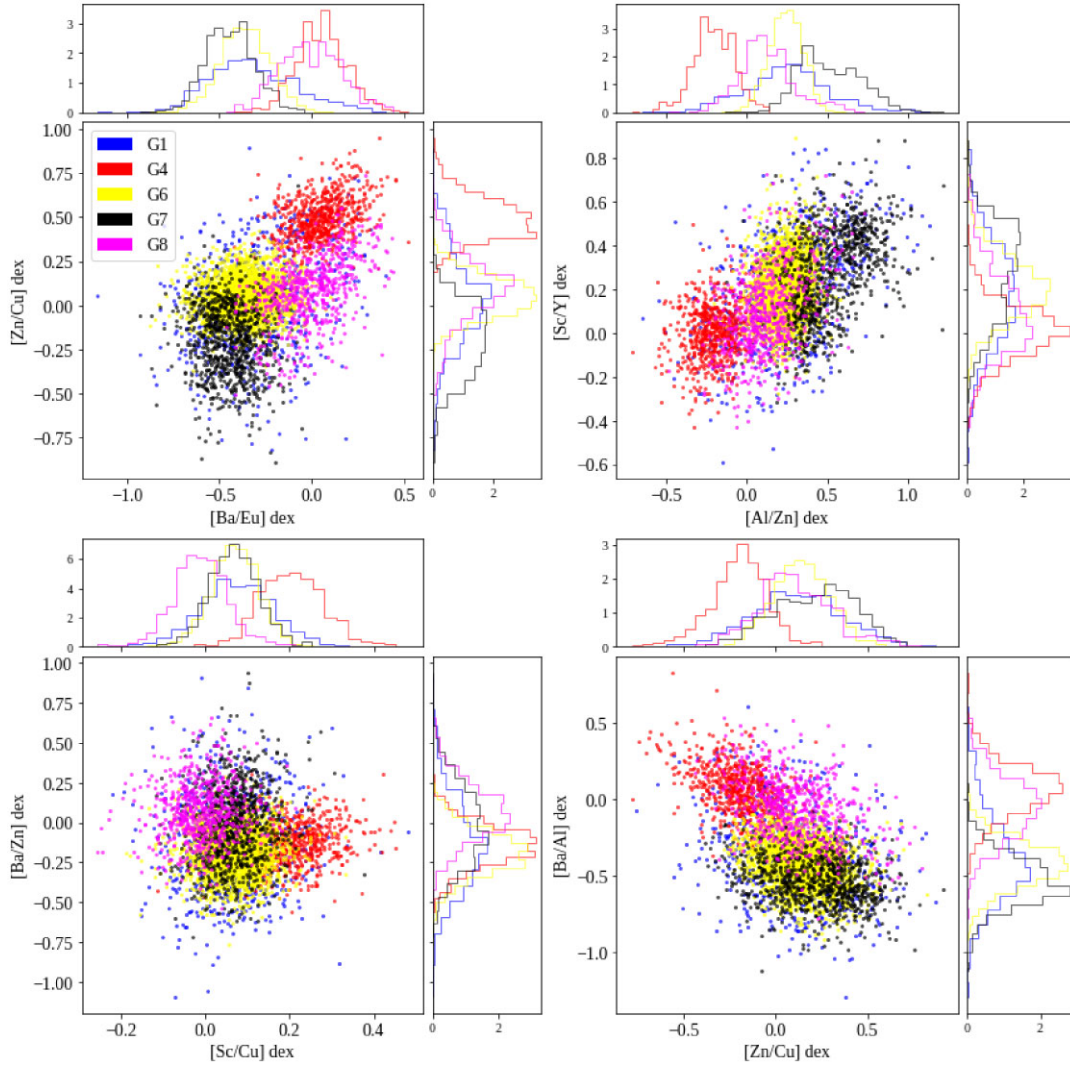
## 5 DISCUSSION

In this section, we assess the robustness of our method and compare our findings with previous studies. We discuss the chemical and kinematic properties of different groups, emphasizing the distinct populations within the thick disc and their implications for the SFH of the metal-poor MW. Additionally, we examine the *in situ* halo groups and GES and explore the chemical diversity present in our sample.

### 5.1 Multiple populations in the thick disc

The magnitude of the weights for Ba (PC1 and PC2), Y (PC2), Zn (PC2), Al (PC1), Cu (PC1), La (PC2), Sc (PC2), and Eu (PC2) are among the highest, and the distinguishing power of these chemical abundances become a useful tool to decompose the thick disc. Fig. 8 shows the chemical composition of groups we associate with the thick disc. We select the following hyperplanes: [Zn/Cu] – [Ba/Eu], [Sc/Y] – [Al/Zn], [Ba/Zn] – [Sc/Cu], and [Ba/Al] – [Cu/Y]. [Zn/Cu] is selected as they both have different time-scales: Cu from SNe Ia and AGB (longer delay) and Zn from CCSNe and HNe (shorter delay). Hence, regions with early star formation might have higher [Zn/Cu], while prolonged star formation shows lower [Zn/Cu]. [Ba/Eu] traces s- versus r-process nucleosynthesis. Higher [Ba/Eu] indicates a stronger AGB star influence and prolonged star formation; lower [Ba/Eu] suggests significant r-process events linked to early, rapid star formation. Thus, in Fig. 8, G4 in the top right (high [Zn/Cu], high [Ba/Eu]) could represent an older population with rapid star formation and early enrichment. We note that G4 lies in the region of the chemical hyperplane, and suspect it could have experienced a high number of HNe. [Sc/Y] indicates the play-off between CCSNe and the ls-process in AGB stars – Y shows a very high weighting in PC2 and this variance is seen in [Sc/Y]. [Al/Zn] is another probe for HNe, as well as for lower mass (low Al) populations. This means G4 and G8 both have low [Al/Zn], indicating lower mass systems with evidence of early intense star formation. The bottom two chemical hyperplanes also demonstrate how the thick disc populations lie, with G4 and G8 being the most distinctly separate.

Overall, G1 and G6 are chemically similar, representing the classical thick disc. G7 differs due to its Mn and Zn enhancement, while G4 is the most chemically distinct among the thick disc groups captured here. G8 alongside G4 has a distinctive hs-element enhancement which is captured in [Ba/Fe]. This enhancement in Ba, but not Y, could indicate that there was an environment with higher neutron density (Bisterzo et al. 2010). Additionally, G4 features a very high [Zn/Fe] enhancement which could be the contribution



**Figure 8.** Four hyperplanes:  $[\text{Zn}/\text{Cu}] - [\text{Ba}/\text{Eu}]$ ,  $[\text{Sc}/\text{Y}] - [\text{Al}/\text{Zn}]$ ,  $[\text{Ba}/\text{Zn}] - [\text{Sc}/\text{Cu}]$ , and  $[\text{Ba}/\text{Al}] - [\text{Cu}/\text{Y}]$  (left to right, top to bottom panels). The five groups are selected to belong to the thick disc as defined by their high  $[\alpha/\text{Fe}]$  and thick disc-like kinematics. In these chemical hyperplanes, the thick-disc populations show distinction in chemistry. G4 is easier to classify separately in each plane, whereas G7 is most clearly distinguishable in  $[\text{Al}/\text{Zn}]$  and  $[\text{Zn}/\text{Cu}]$ .

from HNe (Kobayashi et al. 2020). This suggests while some studies have argued that the thick disc is mostly chemically homogeneous (Haywood et al. 2019; Chandra et al. 2024), we find evidence that the thick disc’s formation could have resulted from three separate events or nuanced nucleosynthesis channel which results in the classical thick disc (G1 + G6 + G7), G4 (high Zn, high Ba) and G8 (low Sc, high Ba) (Navarro et al. 2011; Belokurov & Kravtsov 2022).

## 5.2 *In situ* halo and GES

G3 and G5 are likely the *in situ* halo. We look to Heracles, postulated to be an ancient accreted population. The stars are proposed to have formed prior to the spin-up of the MW disc. It was discovered chemodynamically by Horta et al. (2020) and is thought to have an estimated progenitor stellar mass of  $M_* \approx 5 \times 10^8 M_\odot$ . The  $[\text{Al}/\text{Fe}]$  distribution presented in this paper is on average higher than the expectation for Heracles quoted in Horta et al. (2023). This is because the authors selected Heracles according to the following chemical criteria:  $[\text{Al}/\text{Fe}] < -0.07$  &  $[\text{Mg}/\text{Mn}] \geq 0.25$ , and  $[\text{Al}/\text{Fe}] \geq -0.07$  &  $[\text{Mg}/\text{Mn}] \geq 4.25 \times [\text{Al}/\text{Fe}] + 0.5475$ . Other connections have been

made relating to this population, for example, Kraken (Kruijssen et al. 2020) and Koala (Forbes 2020).

Heracles has been found to have very similar properties to the protogalaxy population, Aurora, as identified in Myeong et al. (2022). Aurora is associated with having a higher  $[\alpha/\text{Fe}]$  ratio compared with Heracles, indicating that Aurora had a higher SFR than Heracles (Horta et al. 2023). Additionally, as G3 occupies a lower orbital energy region,  $E_\mu = -1.82 \times 10^5 \text{ km}^2 \text{ s}^{-2}$ , this appears to kinematically match with Aurora which has orbital energies of  $-1.76 \pm 0.15 \times 10^5 \text{ km}^2 \text{ s}^{-2}$ . Additionally, as noted, G3 is incredibly enhanced in the  $[\text{Ba}/\text{Fe}]$  and  $[\text{Y}/\text{Fe}]$  ratio, indicating the contribution of metallicity-dependent AGB stars. Our values of s-process enhancement,  $[\text{Ba}/\text{Fe}] = 0.44 \pm 0.24 \text{ dex}$ , align with those presented for Aurora at  $[\text{Ba}/\text{Fe}] = 0.48 \pm 0.25 \text{ dex}$  (Myeong et al. 2022). Additionally, Malhan & Rix (2024) has identified two stellar substructures, Shakti and Shiva using Gaia DR3 data, whose chemodynamical properties match G3 and G5. These substructures are characterized by being metal-poor but significantly enhanced in  $[\text{Mg}/\text{Fe}]$  and  $[\text{Al}/\text{Fe}]$  compared to GSE and most known mergers or dwarf galaxies of the MW. This suggests a rapid enrichment inside

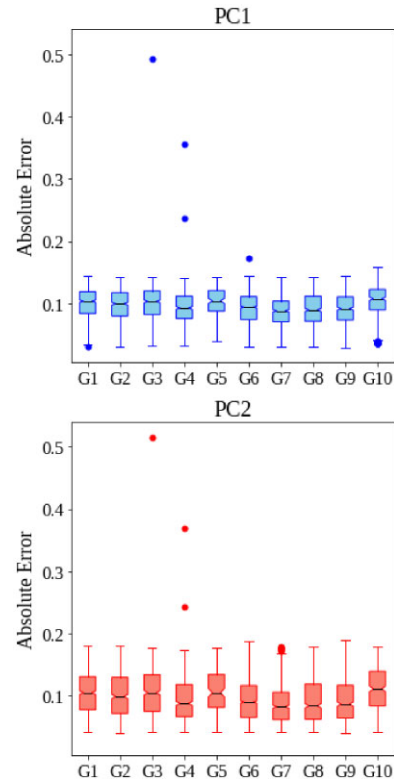
their massive progenitors. They propose possible origins for Shakti and Shiva to be: *in situ*/disc, protogalactic, ‘heated’ disc, ‘spun up’ halo or accreted. Of these origins, they find that the metallicity, chemical abundances and  $[Al/Fe] > 0$  suggest a protogalactic origin, similarly to Aurora. They do note that both Shakti and Shiva have a higher binding energy than Aurora and determine that the unconventional orbital and abundance properties of these two populations makes it difficult to classify their origins. We call G3 and G5 as *in situ* halo populations, as despite their low metallicity, their supersolar  $[Al/Fe]$  does not mark them as accreted populations.

G2’s chemodynamical properties point to it being representative of the accreted population GES. This structure has been extensively studied in terms of its chemodynamical properties. In this work, we find that G2 has lower  $\alpha$ -enrichment than the thick disc, in agreement with previous works (e.g. Haywood et al. 2018; Helmi et al. 2018; Mackereth et al. 2019; Das et al. 2020; Buder et al. 2022; Horta et al. 2023). Additionally, as GES is an accreted population we find that it has a sub-solar  $[Al/Fe]$  ratio which resembles the abundance pattern of stars from satellites of the MW (Hawkins et al. 2015; Horta et al. 2020). Looking to the neutron-capture elements, we see that G2 has a high  $[Eu/Fe]$  ratio of  $0.44 \pm 0.15$  dex, this aligns with many other sources (see e.g. Aguado et al. 2021; Matsuno et al. 2021; Carrillo et al. 2022; Myeong et al. 2022; Naidu et al. 2022). On the other hand, looking to the s-process abundances, such as Ba and Y, we see that our sample of GES candidates have an equivalent enhancement to Eu ( $[Ba/Fe] = 0.41 \pm 0.19$  dex). This is in contradiction with some studies who indicate that GES has solar levels of Ba-enhancement (e.g. Carrillo et al. 2022), while others measure a strong enhancement of up to  $[Ba/Fe] = 0.38 \pm 0.18$  dex (Myeong et al. 2022). GES is thought to have an extended SFH of  $\approx 3.6$  Gyr (Bonaca et al. 2021), meaning theoretically there would be an expected source of enrichment via AGB stars.

### 5.3 Method robustness and comparison

We find in this study of observational data that nine PCs are required to make up 85 per cent of the cumulative variance, reflecting the findings of Ting et al. (2012) and Price-Jones & Bovy (2019), who argued for the necessity of  $\approx 6$ –10 abundance dimensions. The results contrast, however, with those obtained from applying PCA to a set of chemical abundances predicted by a one-zone chemical evolution model, Andrews et al. (2017). Here, the authors found that  $\alpha$ -elements and elements such as Fe, Mn, and Al dominate the encoded information, with two PCs accounting for 99 per cent of the variance in their model. This suggests that chemical evolution models may be underestimating the true diversity of contributions to the high-dimensional  $C$ -space. While we find that Ba, Zn, and Cu are all important elements in chemically distinguishing stars, we do not see a significant variance in  $[Mg/Fe]$ . Ting et al. (2012) found that low-mass AGB stars are significant contributors to the abundance of hs and ls-elements at higher metallicity, which could reflect the Ba-enhancement we see in G4 and G8 among the thick-disc groups. The s-process elements are also shown to be effective for chemical tagging in Blanco-Cuaresma et al. (2015). Weinberg et al. (2022) emphasize that to effectively chemically tag groups, it is recommended to minimize group overlap and reduce residual correlations between abundances to below the level of observational uncertainties.

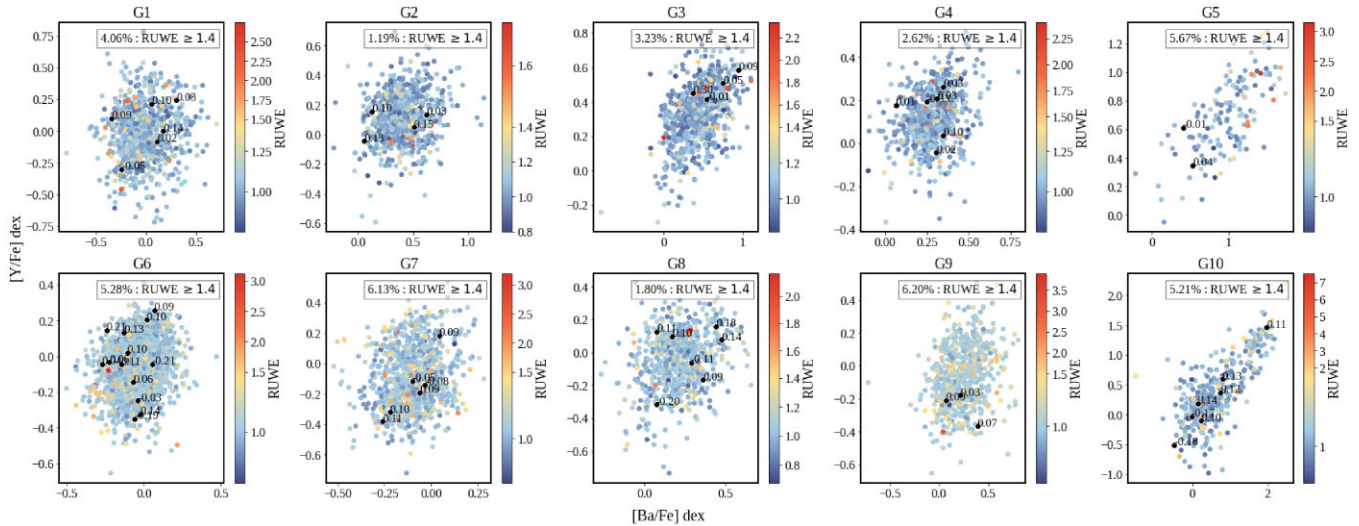
Hogg et al. (2016) assessed the potential of chemical tagging in light of challenges such as noise and incompleteness in chemical abundance measurements and determined that clustering in an abundance space can be effective. We also find that by using the PCA latent space as the clustering target we are able to cleanly recover



**Figure 9.** These box plots show the spread and distribution of absolute errors in PC1 (upper panel) and PC2 (lower panel) for all the groups. The whiskers extend to 1.5 times the interquartile range. Data points beyond the whiskers are considered outliers and are marked as individual dots. It is unlikely that G10’s wide distribution of chemodynamical properties are due to large errors, as shown.

different histories in the metal-poor MW for the majority of groups. In Naidu et al. (2020), they conclude that clustering algorithms (such as DBSCAN, HBDSCAN, and k-means) either fracture the space into too many clusters, or assign all stars to a dominant structure (e.g. GES or the thick disc). Furthermore, Blanco-Cuaresma et al. (2015) indicate that chemical homogeneity across differing groups becomes a challenge, with there being a non-negligible overlap between structures dependent on the  $C$ -space. In light of this, we will discuss the outlier group (G10). The errors of the PC1 and PC2 variables for each group are shown in Fig. 9, indicating that high errors are not the reason for this group capturing an extreme range of chemodynamical properties.

Also, we look at the renormalized unit weight error (RUWE) values of the s-process enhanced *in situ* halo groups (G3 and G5) to evaluate whether they are unresolved AGB binaries which may possess enhanced s-process values due to mass transfer and other effects. After examining the RUWE values across all groups, we quantified the number with  $RUWE \geq 1.4$ , based on the recommended cut-off value from Buder et al. (2021). The number of stars with  $RUWE \geq 1.4$  is very low across all groups, indicating that unresolved binaries are not a significant concern. Additionally, we cross-matched APOGEE (Prieto et al. 2008) and GALAH surveys to obtain  $[C/Fe]$  values for 329 stars, as high C along with Ba is another indicator of binary AGB stars. The data suggest that Ba-rich stars are not a concern in our analysis. This is shown in Fig. 10.



**Figure 10.** G1–G10 plotted in the  $[Y/Fe] - [Ba/Fe]$  chemical space, and colour-mapped by the RUWE values. At  $RUWE \geq 1.4$  (Buder et al. 2021), the single-star astrometric fit becomes poor and this could indicate the presence of an unresolved binary. The number of stars with  $RUWE \geq 1.4$  is very low across all groups. Black points represent stars with  $[C/Fe]$  values obtained from cross-matching APOGEE and GALAH data.

## 6 CONCLUSIONS

In this study, we explored the chemical diversity of the MW’s metal-poor populations by combining PCA and XD on 17 abundances measured for 9923 stars in the GALAH DR3 data set. This methodology allowed for an unbiased and comprehensive examination of the Galaxy’s chemical landscape, leading to the identification of 10 distinct stellar groups. Our approach provides a clear advantage over traditional techniques by avoiding arbitrary cuts based on chemistry or kinematics, offering a less biased representation of the Galaxy’s chemical complexity. Our key findings include:

(i) Identification of known populations: By using chemistry alone, we successfully identified known stellar populations through a combination of latent space representation and a detailed examination of chemical abundances beyond a traditional 2D analysis. This includes GES, the thick disc, the thin disc and *in situ* halo populations.

(ii) Limited discriminatory power of Mg: Mg has less impact in separating accreted and disc stars, with the greatest  $\alpha$ -separation being between the thick and thin discs. Other elements like Ba, Al, Cu, and Sc are important for distinguishing disc from accreted stars, and Ba, Y, Eu, and Zn for distinguishing disc and accreted stars from *in situ* halo.

(iii) Multiple chemically distinct populations in the thick disc: The thick disc is revealed to be a composite of multiple populations with distinct chemical histories, underscoring its complexity and the consideration of it being a diverse assembly rather than a homogeneous entity.

These findings highlight the intricate chemical structure of the thick disc and the broad range of influential elements, reflecting the diverse enrichment environments specific to the metal-poor MW.

## ACKNOWLEDGEMENTS

PD is supported by a UKRI Future Leaders Fellowship (grant reference MR/S032223/1). PJ acknowledges Fondo Nacional de Desarrollo Científico y Tecnológico (Fondecyt) Regular Number 1231057, and the Millennium Nucleus ERIS NCN2021.017. NB would like to thank the GLEAM and ERIS collaborations for their

support with this research, enabling me to attend workshops in Chile and learn from the ERIS team. The research for this publication was coded in python and included its packages: numpy, astropy, pandas, matplotlib and sklearn. This work made use of the Extreme Deconvolution (XD) package (Bovy et al. 2011), available at: <https://github.com/jobovy/extreme-deconvolution>. This work made use of the Third Data Release of the GALAH Survey (Buder et al. 2021). The GALAH Survey is based on data acquired through the Australian Astronomical Observatory, under programs: A/2013B/13 (The GALAH pilot survey); A/2014A/25, A/2015A/19, A/2017A/18 (The GALAH survey phase 1); A/2018A/18 (Open clusters with HERMES); A/2019A/1 (Hierarchical star formation in Ori OB1); A/2019A/15 (The GALAH survey phase 2); A/2015B/19, A/2016A/22, A/2016B/10, A/2017B/16, A/2018B/15 (The HERMES-TESS program); and A/2015A/3, A/2015B/1, A/2015B/19, A/2016A/22, A/2016B/12, A/2017A/14 (The HERMES K2-follow-up program). We acknowledge the traditional owners of the land on which the AAT stands, the Gamilaraay people, and pay our respects to elders past and present. This paper includes data that has been provided by AAO Data Central ([datacentral.org.au](http://datacentral.org.au)). 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.

## DATA AVAILABILITY

The code and data generated to produce the figures will be shared on reasonable request to NB.

## REFERENCES

- Aguado D. S. et al., 2021, *ApJ*, 908, L8  
 Anders F., Chiappini C., Santiago B. X., Matijević G., Queiroz A. B., Steinmetz M., Guiglion G., 2018, *A&A*, 619, A125  
 Andrews B. H., Weinberg D. H., Johnson J. A., Bensby T., Feltzing S., 2012, *Acta Astron.*, 62, 269

- Andrews B. H., Weinberg D. H., Schönrich R., Johnson J. A., 2017, *ApJ*, 835, 224
- Aoki W. et al., 2005, *ApJ*, 632, 611
- Bedell M. et al., 2018, *ApJ*, 865, 68
- Beers T. C. et al., 2012, *ApJ*, 746, 34
- Belokurov V., Kravtsov A., 2022, *MNRAS*, 514, 689
- Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, *MNRAS*, 478, 611
- Belokurov V., Sanders J. L., Fattahi A., Smith M. C., Deason A. J., Evans N. W., Grand R. J. J., 2020, *MNRAS*, 494, 3880
- Beniamini P., Hotokezaka K., Piran T., 2016, *ApJ*, 832, 149
- Bensby T., Feltzing S., Lundström I., 2003, *A&A*, 410, 527
- Bignone L. A., Helmi A., Tissera P. B., 2019, *ApJ*, 883, L5
- Bisterzo S., Gallino R., Straniero O., Cristallo S., Käppeler F., 2010, *MNRAS*, 404, 1529
- Blanco-Cuaresma S. et al., 2015, *A&A*, 577, A47
- Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529
- Bonaca A. et al., 2021, *ApJ*, 909, L26
- Bovy J., Hogg D. W., Roweis S. T., 2011, *Ann. Appl. Stat.*, 5, 1657
- Buder S. et al., 2021, *MNRAS*, 506, 150
- Buder S. et al., 2022, *MNRAS*, 510, 2407
- Burris D. L., Pilachowski C. A., Armandroff T. E., Sneden C., Cowan J. J., Roe H., 2000, *ApJ*, 544, 302
- Busso M., Gallino R., Wasserburg G. J., 1999, *ARA&A*, 37, 239
- Carollo D. et al., 2007, *Nature*, 450, 1020
- Carrillo A., Hawkins K., Jofré P., de Brito Silva D., Das P., Lucey M., 2022, *MNRAS*, 513, 1557
- Chandra V. et al., 2024, *ApJ*, 972, 112
- Ciucă I. et al., 2023, *MNRAS*, 528, L122
- Cohen J. G., 2006, in Randich S., Pasquini L., eds, *Chemical Abundances and Mixing in Stars in the Milky Way and its Satellites. ESO ASTROPHYSICS SYMPOSIA*. Springer, Berlin, Heidelberg, p. 103
- Conroy C. et al., 2019, *ApJ*, 883, 107
- Cui X.-Q. et al., 2012, *Res. Astron. Astrophys.*, 12, 1197
- Das P., Hawkins K., Jofré P., 2020, *MNRAS*, 493, 5195
- De Boer T. J. L., Tolstoy E., Lemasle B., Saha A., Olszewski E. W., Mateo M., Irwin M. J., Battaglia G., 2014, *A&A*, 572, A10
- De Los Reyes M. A. C., Kirby E. N., Ji A. P., Nuñez E. H., 2022, *ApJ*, 925, 66
- Emami N. et al., 2021, *ApJ*, 922, 217
- Feltzing S., Fohlman M., Bensby T., 2007, *A&A*, 467, 665
- Feng Y., Krumholz M. R., 2014, *Nature*, 513, 523
- Fernández-Alvar E., Tissera P. B., Carigi L., Schuster W. J., Beers T. C., Belokurov V. A., 2019, *MNRAS*, 485, 1745
- Font A. S., Johnston K. V., Bullock J. S., Robertson B., 2006, *ApJ*, 638, 585
- Forbes D. A., 2020, *MNRAS*, 493, 847
- Fraley C., Raftery A. E., 2002, *J. Am. Stat. Assoc.*, 97, 611
- Frebel A., 2018, *Annual Rev. Nucl. Part. Sci.*, 68, 237
- Freeman K., Bland-Hawthorn J., 2002, *ARA&A*, 40, 487
- Gilmore G., Reid N., 1983, *MNRAS*, 202, 1025
- Hawkins K., Jofré P., Gilmore G., Masseron T., 2014, *MNRAS*, 445, 2575
- Hawkins K., Jofré P., Masseron T., Gilmore G., 2015, *MNRAS*, 453, 758
- Hayden M. R. et al., 2015, *ApJ*, 808, 132
- Haywood M., Di Matteo P., Lehnert M., Snaith O., Khoperskov S., Gómez A., 2018, *ApJ*, 863, 113
- Haywood M., Snaith O., Lehnert M. D., Di Matteo P., Khoperskov S., 2019, *A&A*, 625, A105
- Helmi A., de Zeeuw P. T., 2002, *MNRAS*, 319, 657
- Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, *Nature*, 563, 85
- Hendricks B., Koch A., Lanfranchi G. A., Boeche C., Walker M., Johnson C. I., Peñarrubia J., Gilmore G., 2014, *ApJ*, 785, 102
- Hogg D. W. et al., 2016, *ApJ*, 833, 262
- Horta D. et al., 2020, *MNRAS*, 500, 1385
- Horta D. et al., 2023, *MNRAS*, 520, 5671
- Ishigaki M. N., Chiba M., Aoki W., 2012, *ApJ*, 753, 64
- Jackson-Jones R. et al., 2014, *A&A*, 571, L5
- Ji A. P., Frebel A., Chiti A., Simon J. D., 2016, *Nature*, 531, 610
- Jofré P., Heiter U., Soubiran C., 2019, *ARA&A*, 57, 571
- Jolliffe I. T., Cadima J., 2016, *Phil. Trans. R. Soc. A*, 374, 20150202
- Käppeler F., Beer H., Wisshak K., 1989, *Rep. Prog. Phys.*, 52, 945
- Kobayashi C., Umeda H., Nomoto K., Tominaga N., Ohkubo T., 2006, *ApJ*, 653, 1145
- Kobayashi C., Karakas A. I., Lugaro M., 2020, *ApJ*, 900, 179
- Krujssens J. M. D. et al., 2020, *MNRAS*, 498, 2472
- Kumar R., Pradhan A. C., Ojha D. K., Piridi S., Baug T., Ghosh S. K., 2021, *JA&A*, 42, 42
- Lanfranchi G. A., Matteucci F., Cescutti G., 2006, *MNRAS*, 365, 477
- Ledinauskas E., Zubovas K., 2018, *A&A*, 615, A64
- Liu F., Asplund M., Yong D., Feltzing S., Dotter A., Meléndez J., Ramírez I., 2019, *A&A*, 627, A117
- Mackereth J. T. et al., 2019, *MNRAS*, 482, 3426
- Majewski S. R. et al., 2017, *AJ*, 154, 94
- Malhan K., Rix H.-W., 2024, *ApJ*, 964, 104
- Manea C., Hawkins K., Ness M. K., Buder S., Martell S. L., Zucker D. B., 2023, *ApJ*, 972, 20
- Matsuno T., Hirai Y., Tarumi Y., Hotokezaka K., Tanaka M., Helmi A., 2021, *A&A*, 650, A110
- Matteucci F., 2016, *J. Phys. Conf. Ser.*, 703, 012004
- de Mijolla D., Ness M., Viti S., Wheeler A., 2021, *ApJ*, 913, 12
- Minelli A., Mucciarelli A., Massari D., Bellazzini M., Romano D., Ferraro F. R., 2021, *ApJ*, 918, L32
- Mukherjee C. S., Doerkar N., Zhang J., 2024, preprint ([arXiv:2204.10888](https://arxiv.org/abs/2204.10888))
- Myeong G. C., Vasiliev E., Iorio G., Evans N. W., Belokurov V., 2019, *MNRAS*, 488, 1235
- Myeong G. C., Belokurov V., Aguado D. S., Evans N. W., Caldwell N., Bradley J., 2022, *ApJ*, 938, 21
- Naidu R. P., Conroy C., Bonaca A., Johnson B. D., Ting Y.-S., Caldwell N., Zaritsky D., Cargile P. A., 2020, *ApJ*, 901, 48
- Naidu R. P. et al., 2022, *ApJ*, 926, L36
- Navarro J. F., Abadi M. G., Venn K. A., Freeman K. C., Anguiano B., 2011, *MNRAS*, 412, 1203
- Nissen P. E., Schuster W. J., 2010, *A&A*, 511, L10
- Nissen P. E., Schuster W. J., 2012, *A&A*, 543, A28
- North P. et al., 2012, *A&A*, 541, A45
- Ortigoza-Urdaneta M. et al., 2023, *A&A*, 676, A140
- Pillepich A., Madau P., Mayer L., 2015, *ApJ*, 799, 184
- Piskunov N., Valenti J. A., 2017, *A&A*, 597, A16
- Prantzos N. et al., 2023, *MNRAS*, 523, 2126
- Price-Jones N., Bovy J., 2019, *MNRAS*, 487, 871
- Price-Jones N. et al., 2020, *MNRAS*, 496, 5101
- Prieto C. A. et al., 2008, *Astron. Nachr.*, 329, 1018
- Reddy B. E., Lambert D. L., Prieto C. A., 2006, *MNRAS*, 367, 1329
- Robertson B., Bullock J. S., Font A. S., Johnston K. V., Hernquist L., 2005, *ApJ*, 632, 872
- Sheffield A. A. et al., 2012, *ApJ*, 761, 161
- Ting Y.-S., Weinberg D. H., 2022, *ApJ*, 927, 209
- Ting Y. S., Freeman K. C., Kobayashi C., De Silva G. M., Bland-Hawthorn J., 2012, *MNRAS*, 421, 1231
- Ting Y.-S., Conroy C., Rix H.-W., Cargile P., 2019, *ApJ*, 879, 69
- Tinsley B. M., 1979, *ApJ*, 229, 1046
- Tissera P., Beers T., Carollo D., Scannapieco C., 2014, *MNRAS*, 439, 3128
- Tsujiimoto T., Nishimura N., 2018, *ApJ*, 863, L27
- Wanajo S., Hirai Y., Prantzos N., 2021, *MNRAS*, 505, 5862
- Weinberg D. H. et al., 2022, *ApJS*, 260, 32
- Yanny B. et al., 2009, *AJ*, 137, 4377
- Yu S. et al., 2021, *MNRAS*, 505, 889
- Zolotov A., Willman B., Brooks A. M., Governato F., Brook C. B., Hogg D. W., Quinn T., Stinson G., 2009, *ApJ*, 702, 1058

## APPENDIX A

In this appendix, Figs A1 and A2 show the  $[X/Fe]$  abundances and their associated SME covariance uncertainties versus  $T_{\text{eff}}$  and coloured by  $[Fe/H]$ . This is referenced in Section 2 and shows that

the shown abundances have no strong systematic trends. Table A1 summarizes the chemodynamical properties of the stellar groups identified in this study – useful for Section 4.3. To support this

further, Fig. A3 shows histograms of the identified stellar groups (the thick disc, *in situ* halo, GES, the thin disc and the outlier group).

**Table A1.** Group characteristics based on sample size, chemistry, kinematics, PC1–PC2 location, and notable information.

Group	Sample size	Chemistry (dex) ( $\mu \pm \sigma$ )	Kinematics ( $L_z$ kpc kms $^{-1}$ ) ( $\mu \pm \sigma$ )	PC1–PC2 location	Notes
G1	763	[Ba/Fe] = $-0.32 \pm 0.21$ [Mg/Fe] = $0.32 \pm 0.11$ [Mn/Fe] = $-0.13 \pm 0.14$ [Y/Fe] = $-0.02 \pm 0.23$ [Eu/Fe] = $0.30 \pm 0.10$ [Zn/Fe] = $0.12 \pm 0.22$ [Al/Fe] = $0.37 \pm 0.11$ [Fe/H] = $-0.62 \pm 0.10$ [Cu/Fe] = $0.09 \pm 0.10$ [Ti/Fe] = $0.28 \pm 0.10$ [Ni/Fe] = $0.07 \pm 0.07$ [Sc/Fe] = $0.17 \pm 0.07$	$L_z = 1175.89 \pm 450.05e =$ $0.29 \pm 0.19$	Core	Thick disc chemistry and kinematics
G2	671	[Ba/Fe] = $0.41 \pm 0.19$ [Mg/Fe] = $0.13 \pm 0.10$ [Mn/Fe] = $-0.37 \pm 0.12$ [Y/Fe] = $0.11 \pm 0.17$ [Eu/Fe] = $0.44 \pm 0.15$ [Zn/Fe] = $0.08 \pm 0.15$ [Al/Fe] = $-0.12 \pm 0.14$ [Fe/H] = $-1.00 \pm 0.20$ [Cu/Fe] = $-0.46 \pm 0.12$ [Ti/Fe] = $0.14 \pm 0.10$ [Ni/Fe] = $-0.11 \pm 0.07$ [Sc/Fe] = $-0.01 \pm 0.07$	$L_z = 145.13 \pm 521.97e =$ $0.80 \pm 0.21$	Branch	GES
G3	651	[Ba/Fe] = $0.44 \pm 0.24$ [Mg/Fe] = $0.32 \pm 0.08$ [Mn/Fe] = $-0.27 \pm 0.10$ [Y/Fe] = $0.33 \pm 0.17$ [Eu/Fe] = $0.30 \pm 0.10$ [Zn/Fe] = $0.19 \pm 0.15$ [Al/Fe] = $0.25 \pm 0.15$ [Fe/H] = $-0.87 \pm 0.18$ [Cu/Fe] = $-0.14 \pm 0.14$ [Ti/Fe] = $0.28 \pm 0.08$ [Ni/Fe] = $0.00 \pm 0.08$ [Sc/Fe] = $0.08 \pm 0.08$	$L_z = 660.54 \pm 599.88e =$ $0.49 \pm 0.26$	Core/branch	Bi-modal $L_z$ , <i>in situ</i> population
G4	648	[Ba/Fe] = $0.28 \pm 0.12$ [Mg/Fe] = $0.31 \pm 0.06$ [Mn/Fe] = $-0.17 \pm 0.06$ [Y/Fe] = $0.12 \pm 0.13$ [Eu/Fe] = $0.24 \pm 0.09$ [Zn/Fe] = $0.42 \pm 0.08$ [Al/Fe] = $0.22 \pm 0.08$ [Fe/H] = $-0.58 \pm 0.07$ [Cu/Fe] = $-0.08 \pm 0.07$ [Ti/Fe] = $0.20 \pm 0.06$ [Ni/Fe] = $0.00 \pm 0.06$ [Sc/Fe] = $0.13 \pm 0.04$	$L_z = 1218.60 \pm 441.24e =$ $0.32 \pm 0.20$	Core	Thick disc kinematics, high [Zn/Fe] distinguished from other thick disc groups

**Table A1** – *continued*

Group	Sample size	Chemistry (dex) ( $\mu \pm \sigma$ )	Kinematics ( $L_z$ kpc kms <sup>-1</sup> ) ( $\mu \pm \sigma$ )	PC1–PC2 location	Notes
G5	141	[Ba/Fe] = $0.97 \pm 0.37$ [Mg/Fe] = $0.34 \pm 0.12$ [Mn/Fe] = $-0.21 \pm 0.11$ [Y/Fe] = $0.68 \pm 0.26$ [Eu/Fe] = $0.39 \pm 0.13$ [Zn/Fe] = $0.25 \pm 0.17$ [Al/Fe] = $0.31 \pm 0.14$ [Fe/H] = $-0.75 \pm 0.18$ [Cu/Fe] = $-0.04 \pm 0.15$ [Ti/Fe] = $0.28 \pm 0.08$ [Ni/Fe] = $0.02 \pm 0.08$ [Sc/Fe] = $0.11 \pm 0.06$	$L_z = 840.21 \pm 609.00e =$ $0.45 \pm 0.25$	Core/spray	High s-process enhancement
G6	2235	[Ba/Fe] = $-0.05 \pm 0.12$ [Mg/Fe] = $0.29 \pm 0.07$ [Mn/Fe] = $-0.14 \pm 0.06$ [Y/Fe] = $-0.06 \pm 0.13$ [Eu/Fe] = $0.31 \pm 0.09$ [Zn/Fe] = $0.13 \pm 0.08$ [Al/Fe] = $0.37 \pm 0.06$ [Fe/H] = $-0.62 \pm 0.08$ [Cu/Fe] = $0.08 \pm 0.06$ [Ti/Fe] = $0.27 \pm 0.06$ [Ni/Fe] = $0.07 \pm 0.05$ [Sc/Fe] = $0.14 \pm 0.04$	$L_z = 1245.04 \pm 430.14e =$ $0.31 \pm 0.19$	Core	Thick disc, chemically similar to G1 and G8
G7	832	[Ba/Fe] = $-0.10 \pm 0.13$ [Mg/Fe] = $0.27 \pm 0.08$ [Mn/Fe] = $-0.23 \pm 0.07$ [Y/Fe] = $0.09 \pm 0.20$ [Eu/Fe] = $0.34 \pm 0.06$ [Zn/Fe] = $-0.08 \pm 0.19$ [Al/Fe] = $0.43 \pm 0.06$ [Fe/H] = $-0.63 \pm 0.09$ [Cu/Fe] = $0.12 \pm 0.06$ [Ti/Fe] = $0.24 \pm 0.06$ [Ni/Fe] = $0.06 \pm 0.05$ [Sc/Fe] = $0.18 \pm 0.05$	$L_z = 1153.59 \pm 416.14e =$ $0.31 \pm 0.18$	Core	Thick disc kinematics with bi-modal [Zn/Fe]
G8	610	[Ba/Fe] = $0.21 \pm 0.15$ [Mg/Fe] = $0.36 \pm 0.08$ [Mn/Fe] = $-0.18 \pm 0.08$ [Y/Fe] = $-0.05 \pm 0.18$ [Eu/Fe] = $0.22 \pm 0.07$ [Zn/Fe] = $0.15 \pm 0.15$ [Al/Fe] = $0.27 \pm 0.07$ [Fe/H] = $-0.61 \pm 0.08$ [Cu/Fe] = $0.06 \pm 0.08$ [Ti/Fe] = $0.19 \pm 0.06$ [Ni/Fe] = $0.10 \pm 0.05$ [Sc/Fe] = $0.04 \pm 0.05$	$L_z = 1073.85 \pm 462.95e =$ $0.35 \pm 0.21$	Core	Thick disc, chemically similar to G1 and G6
G9	581	[Ba/Fe] = $0.24 \pm 0.19$ [Mg/Fe] = $0.10 \pm 0.07$ [Mn/Fe] = $-0.08 \pm 0.08$ [Y/Fe] = $-0.06 \pm 0.18$ [Eu/Fe] = $0.15 \pm 0.09$ [Zn/Fe] = $-0.03 \pm 0.20$ [Al/Fe] = $0.21 \pm 0.09$ [Fe/H] = $-0.56 \pm 0.06$ [Cu/Fe] = $0.05 \pm 0.09$ [Ti/Fe] = $0.09 \pm 0.07$ [Ni/Fe] = $0.04 \pm 0.07$ [Sc/Fe] = $0.06 \pm 0.06$	$L_z = 2020.10 \pm 649.82e =$ $0.19 \pm 0.19$	Core	Most $\alpha$ -poor, thin disc



Table A1 – continued

Group	Sample size	Chemistry (dex) ( $\mu \pm \sigma$ )	Kinematics ( $L_z$ kpc kms $^{-1}$ ) ( $\mu \pm \sigma$ )	PC1–PC2 location	Notes
G10	365	[Ba/Fe] = $0.63 \pm 0.65$ [Mg/Fe] = $0.28 \pm 0.23$ [Mn/Fe] = $-0.19 \pm 0.28$ [Y/Fe] = $0.37 \pm 0.53$ [Eu/Fe] = $0.39 \pm 0.24$ [Zn/Fe] = $0.26 \pm 0.34$ [Al/Fe] = $0.24 \pm 0.28$ [Fe/H] = $-0.78 \pm 0.22$ [Cu/Fe] = $-0.11 \pm 0.27$ [Ti/Fe] = $0.22 \pm 0.21$ [Ni/Fe] = $0.00 \pm 0.16$ [Sc/Fe] = $0.12 \pm 0.17$	$L_z = 982.75 \pm 831.54e =$ $0.44 \pm 0.31$	Core/branch/spray	Significant variance, overlaps with multiple groups

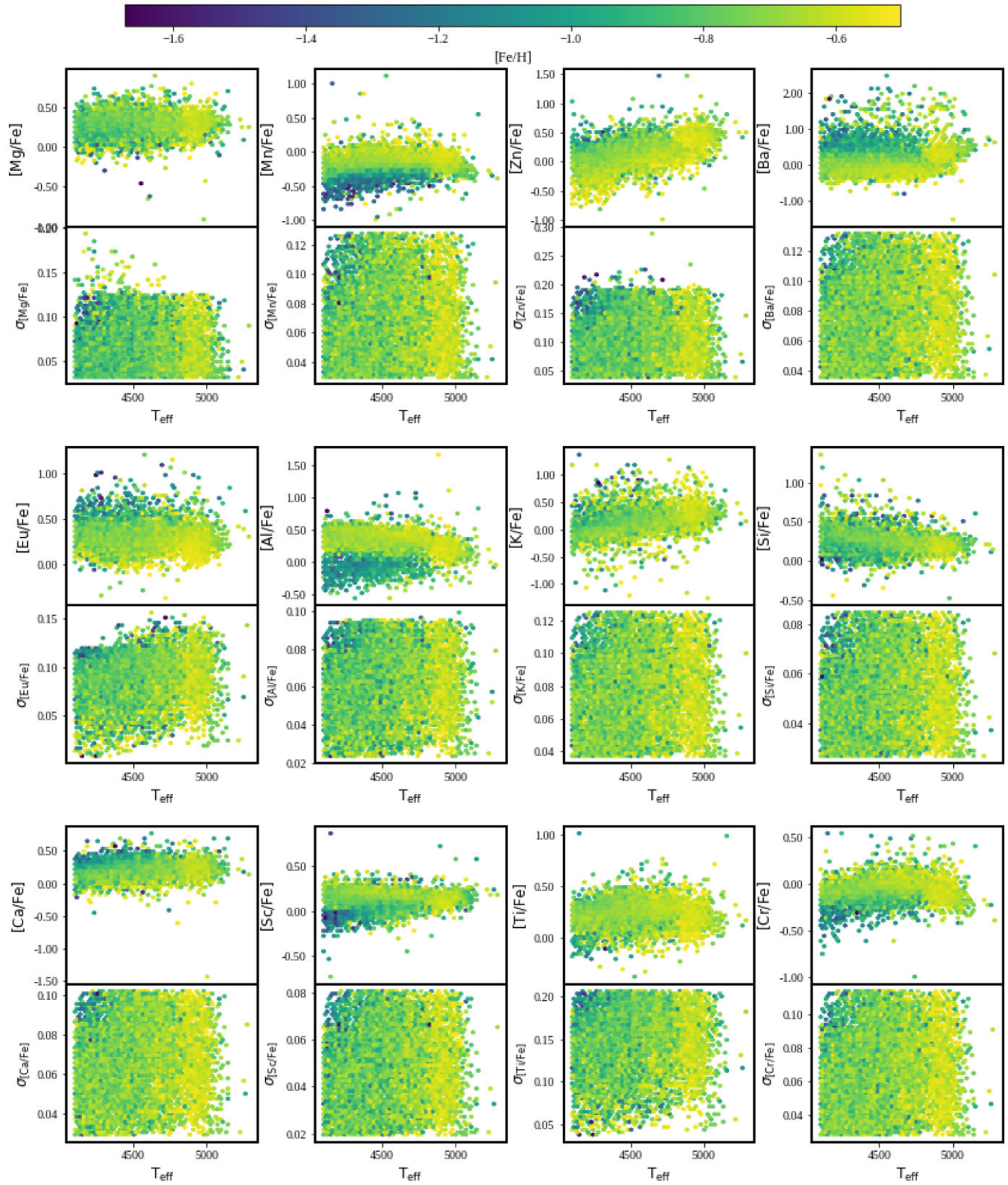
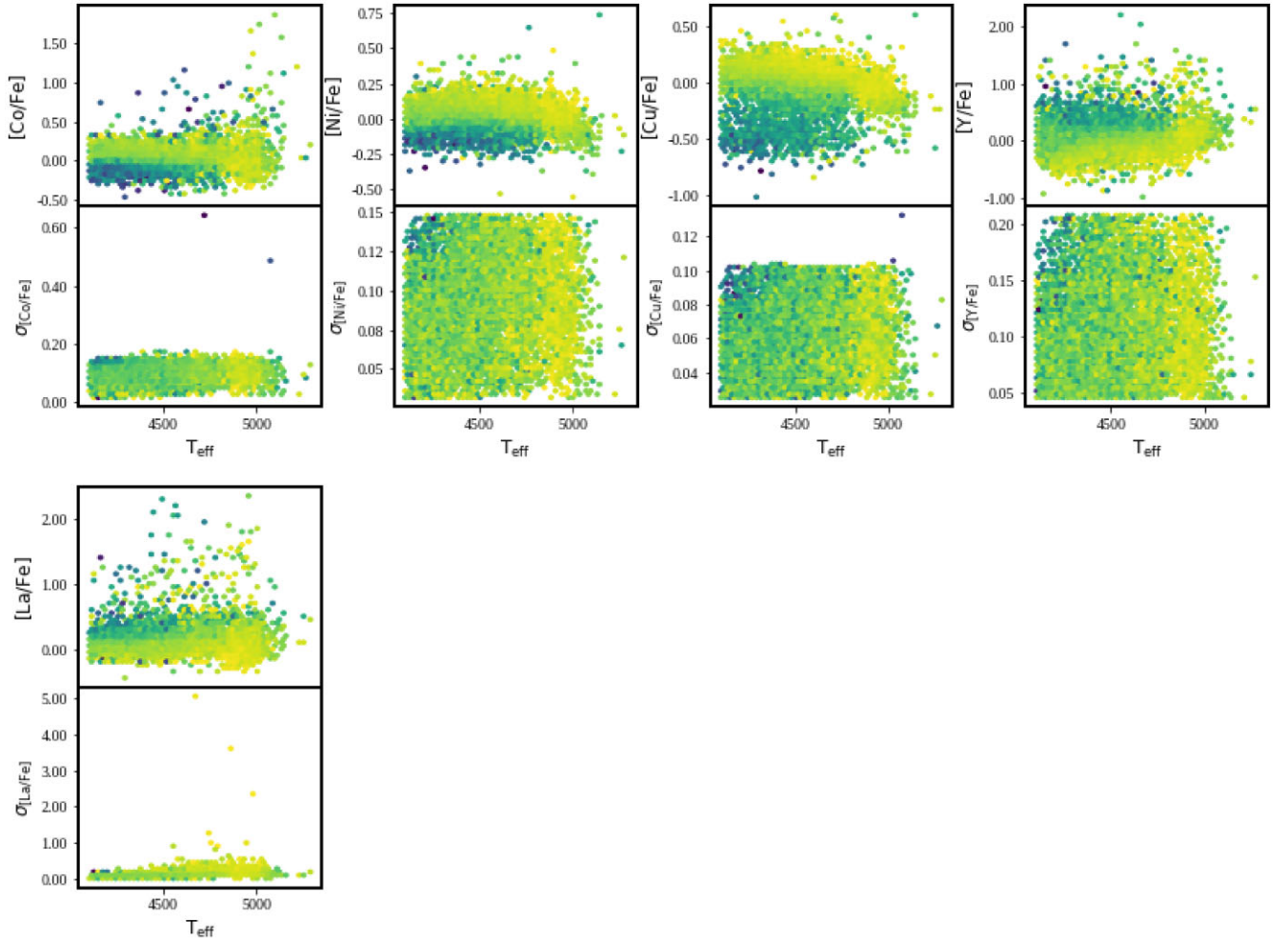
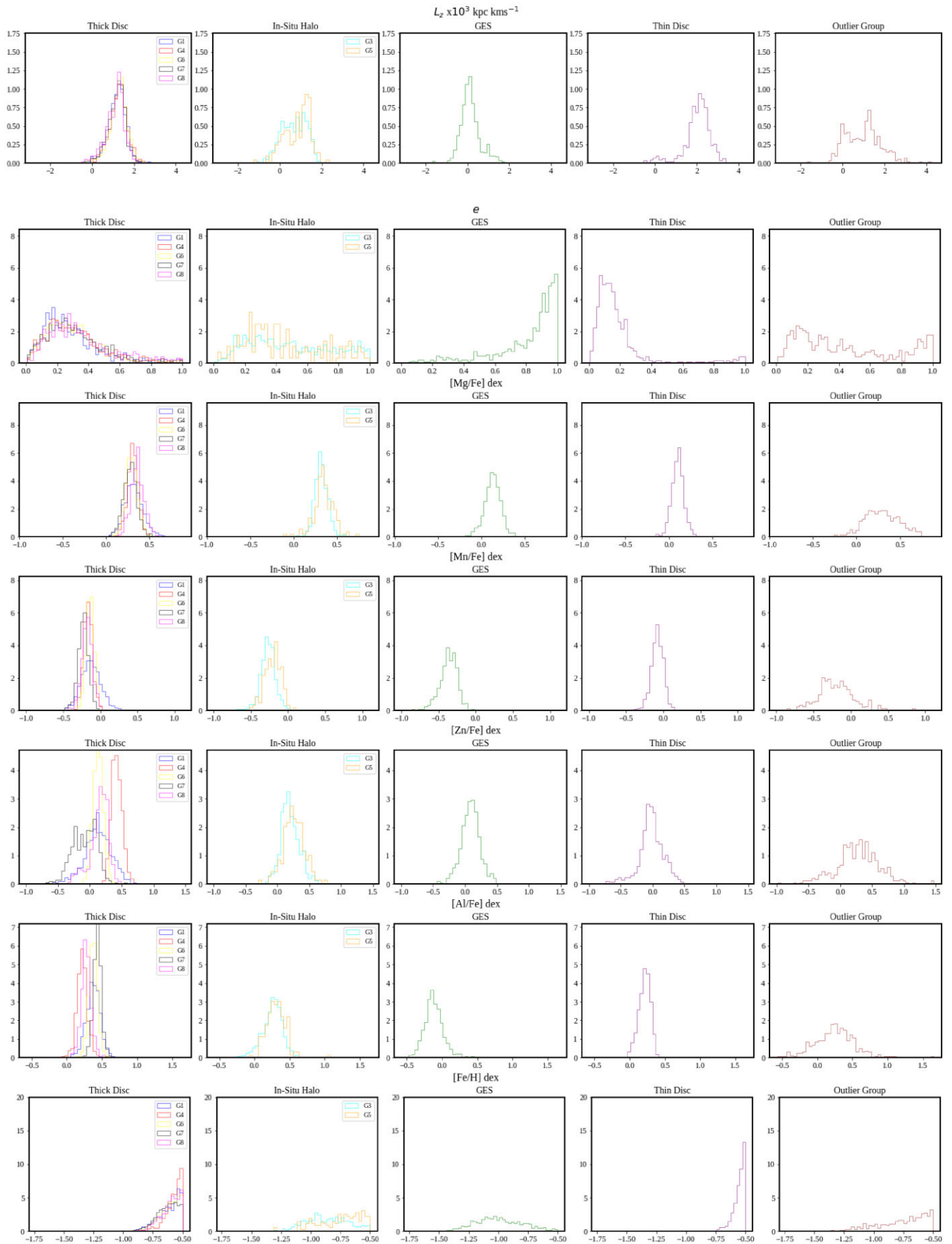


Figure A1. Plots continue on next page.



**Figure A2.** Plots showing  $[X/Fe]$  abundances (top) and associated uncertainties (bottom) versus  $T_{\text{eff}}$ , with a colour bar indicating  $[Fe/H]$ . We note that  $[La/Fe]$  has the greatest errors of all the abundances. We refer to Buder et al. (2021), who state that La is one of the abundances impacted the most by blended spectral lines.



**Figure A3.** Histograms of the chemodynamical properties of each group (thick disc, *in situ* halo, GES, thin disc and the outlier group).

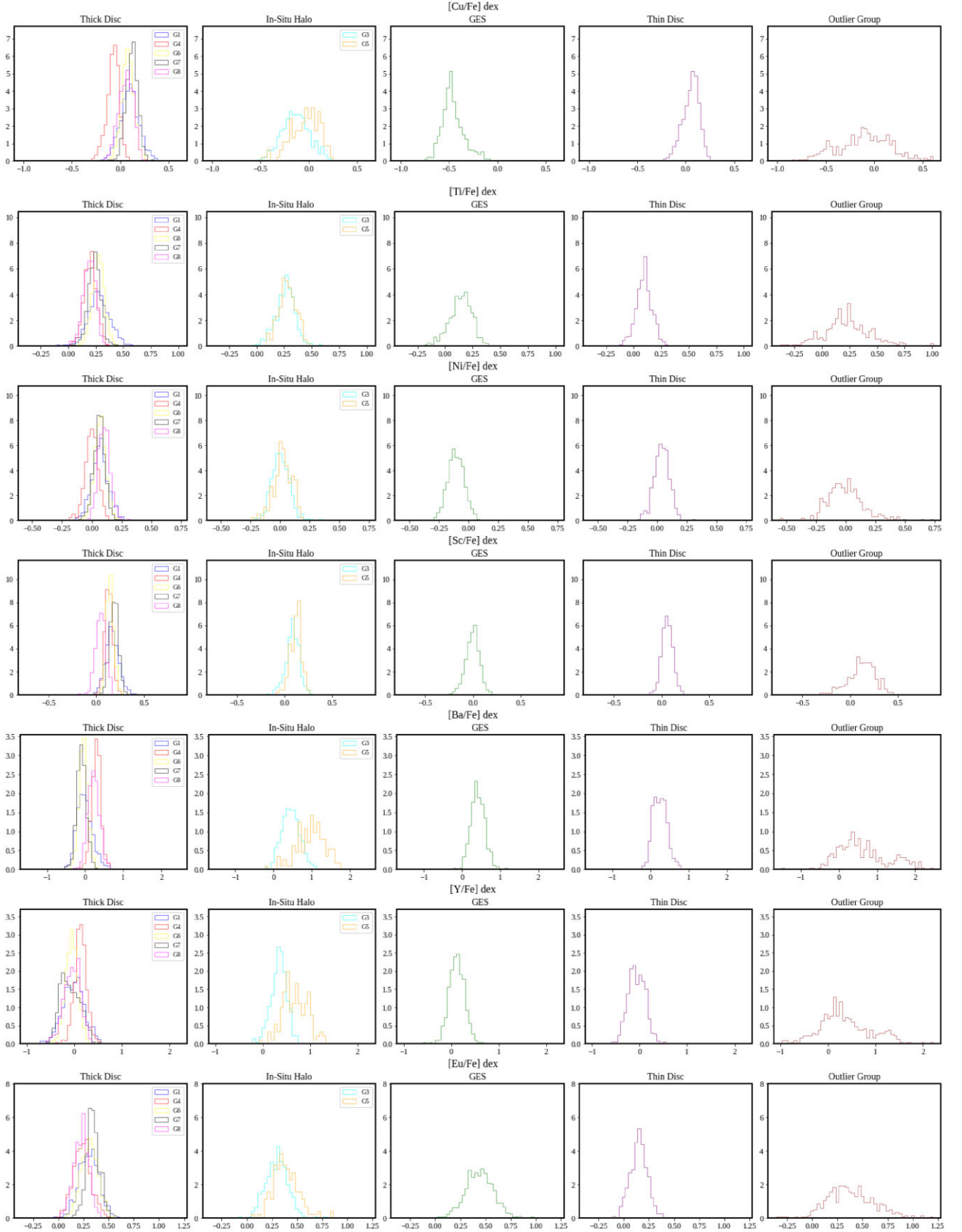


Figure A3. – continued.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.

© 2024 The Author(s).

Published by Oxford University Press on behalf of Royal Astronomical Society. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.