Can we really pick and choose? Benchmarking various selections of *Gaia* Enceladus/Sausage stars in observations with simulations

Andreia Carrillo [ORCID],[1,2]★ Alis J. Deason [ORCID],[1,2] Azadeh Fattahi [ORCID],[1] Thomas M. Callingham [ORCID][3] and Robert J. J. Grand [ORCID][4,5,6]

[1]*Institute for Computational Cosmology, Department of Physics, Durham University, Durham DH1 3LE, UK*
[2]*Centre for Extragalactic Astronomy, Department of Physics, University of Durham, South Road, Durham DH1 3LE, UK*
[3]*Kapteyn Astronomical Institute, University of Groningen, Landleven 12, NL-9747 AD Groningen, the Netherlands*
[4]*Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK*
[5]*Instituto de Astrofisica de Canarias, Calle Via Lactea s/n, E-38205 La Laguna, Tenerife, Spain*
[6]*Departamento de Astrofisica, Universidad de La Laguna, Av. del Astrofisico Francisco Sanchez s/n, E-38206 La Laguna, Tenerife, Spain*

ABSTRACT

Large spectroscopic surveys plus *Gaia* astrometry have shown us that the inner stellar halo of the Galaxy is dominated by the debris of *Gaia* Enceladus/Sausage (GES). With the richness of data at hand, there are a myriad of ways these accreted stars have been selected. We investigate these GES selections and their effects on the inferred progenitor properties using data constructed from APOGEE and *Gaia*. We explore selections made in eccentricity, energy-angular momentum (E-Lz), radial action-angular momentum (Jr-Lz), action diamond, and [Mg/Mn]-[Al/Fe] in the observations, selecting between 144 and 1279 GES stars with varying contamination from *in-situ* and other accreted stars. We also use the Auriga cosmological hydrodynamic simulations to benchmark the different GES dynamical selections. Applying the same observational GES cuts to nine Auriga galaxies with a GES, we find that the Jr-Lz method is best for sample purity and the eccentricity method for completeness. Given the average metallicity of GES ($-1.28 <$ [Fe/H] < -1.18), we use the $z = 0$ mass–metallicity relationship to find an average $M_\star$ of $\sim 4 \times 10^8 \, M_\odot$. We adopt a similar procedure and derive $M_\star$ for the GES-like systems in Auriga and find that the eccentricity method overestimates the true $M_\star$ by $\sim 2.6 \times$ while E-Lz underestimates by $\sim 0.7 \times$. Lastly, we estimate the total mass of GES to be $10^{10.5-11.1} \, M_\odot$ using the relationship between the metallicity gradient and the GES-to-*in-situ* energy ratio. In the end, we cannot just 'pick and choose' how we select GES stars, and instead should be motivated by the science question.

Key words: Galaxy: formation – Galaxy: halo – galaxies: dwarf.

1 INTRODUCTION

Galaxies grow from the accretion of other smaller galaxies and the Milky Way is no stranger to this process. Particularly, our Galaxy has had a relatively quiet recent merger history (Gilmore, Wyse & Norris 2002), save for the current interactions with Sagittarius (Ibata, Gilmore & Irwin 1994; Majewski et al. 2017) and the Magellanic Clouds (Gómez et al. 2015; Laporte et al. 2018), in addition to dwarf galaxies and globular clusters that are tidally disrupted into streams in the stellar halo (e.g. Belokurov et al. 2006; Shipp et al. 2018). However, a seemingly significant merger in hiding has been hinted at in the past (Nissen & Schuster 1997; Chiba & Beers 2000; Meza et al. 2005; Nissen & Schuster 2010) because of the distinct kinematics and chemistry imprinted in some halo stars in the local Solar neighbourhood. With data from the *Gaia* mission providing astrometry for $>10^9$ stars, in tandem with large spectroscopic surveys providing chemistry for $>10^6$ stars, we have indeed confirmed the existence of this merger we now call *Gaia*-Enceladus/Sausage (GES;

Belokurov et al. 2018; Helmi et al. 2018). With virial mass, $M_{vir} > 10^{10} \, M_\odot$ (Belokurov et al. 2018), GES has undoubtedly changed the course of the evolution and the resulting picture of the Milky Way – from dominating the inner stellar halo and causing a break in the stellar halo density profile (Carollo et al. 2010; Deason et al. 2018; Han et al. 2022), to heating up pre-existing disc stars to halo kinematics (Belokurov et al. 2020; Bonaca et al. 2020), to potentially diluting the metallicity in the interstellar medium (ISM) causing the distinct chemistry of the Milky Way thin and thick discs (Grand et al. 2020; Ciucǎ et al. 2022). The role of GES in understanding the formation and evolution of the Milky Way is therefore unmistakably important.

An obvious next step is putting GES into context with other intact and disrupted satellites (e.g. Fattahi et al. 2020; Hasselquist et al. 2021; Naidu et al. 2022). Many spectroscopic surveys such as the Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017) and the Hectochelle in the Halo at High Resolution Survey (H3 Survey; Conroy et al. 2019) have been extremely powerful for these purposes. A glaring theme from these works however is that the stellar halo is comprised of substructures. The extended nature of GES in its kinematic,

★ E-mail: andreia.carrillo@durham.ac.uk

dynamical, and chemical properties, especially makes it more prone to overlap with accreted stars from other progenitors or even *in-situ* populations. On the other hand, distinct phase-space substructures in the halo such as Wukong, Arjuna, I'itoi, and Sequoia have been suggested to be actually part of GES (e.g. Naidu et al. 2020; Horta et al. 2023). It is becoming increasingly apparent and important to first understand which stars belong to GES and which do not, as this would grossly affect the properties that we infer for its progenitor.

This is further complicated by the fact that different surveys with different selection functions *also* have different ways of selecting the GES population. This, therefore, raises the question *'can we really pick and choose the way we select GES stars'?* One of the main ways GES stars have been selected is in E-Lz space as originally done in Helmi et al. (2018) with *Gaia* DR2 and APOGEE DR14. Here they find that GES forms a separate sequence in the colour–magnitude diagram and has slightly retrograde kinematics. Another widely used selection is in eccentricity space. A high eccentricity cut i.e. $e > 0.7$, is made to select GES stars that have a large dispersion in V_r compared to V_ϕ (e.g. Mackereth et al. 2019 with APOGEE; Naidu et al. 2020 with H3), therefore falling on the *'sausage'* region in the V_ϕ-V_r diagram (Belokurov et al. 2018). A selection in Jr-Lz was also introduced by Feuillet et al. (2020) using data from the SkyMapper Southern Sky Survey (Casagrande et al. 2019) and *Gaia* DR2. Specifically, they selected stars with high radial action and found that this is a cleaner selection as showcased by a normal, Gaussian metallicity distribution function (MDF) for the resulting GES sample. Multiple works have also used the *action diamond* to select GES stars (e.g. Myeong et al. 2019; Lane, Bovy & Mackereth 2022) which uses the actions J_R, J_z, and J_ϕ (or L_z), and delineates stars that have prograde versus retrograde and polar versus radial orbits. Others have used APOGEE data and selected purely on chemistry specifically in [Mg/Mn] versus [Al/Fe], as this distinctly separates the accreted from the *in-situ* stars (Hawkins et al. 2016; Das, Hawkins & Jofré 2020; Carrillo et al. 2022). Specifically, this selection benefits from chemical abundances being an intrinsic and inherited property of stars which could then be used to tag them to different populations (Freeman & Bland-Hawthorn 2002).

The variety of selection methods for GES stars necessitates comparisons of these different selections and the resulting GES samples. Buder et al. (2022) used abundances from the Galactic Archaeology HERMES (GALAH) DR3 and found that the GES chemical selection made in [Mg/Mn] versus [Na/Fe] space overlaps with 29 per cent of their dynamical selection made in Jr-Lz. Lane, Bovy & Mackereth (2022) explored six kinematic selections for GES in an idealized simulation where they found that the scaled action space is best in separating GES from the isotropic halo and achieves sample purity of 82 per cent. Limberg et al. (2022) found a similar level of sample purity for their GES sample using the Jr-Lz selection in observations, estimating the contamination from APOGEE abundances.

Selecting GES stars is non-trivial as the performance of these selections are usually evaluated based on cross-validations between kinematics and chemistry. While this is the best that we can do in the observations, one should be cognizant that this is far from perfect. For example, the samples are overlapping but still different for the Milky Way thin and thick discs depending on how we define them – chemically, kinematically, or spatially.

This is where simulations come in, as they can help us understand the biases and contamination from the different GES selection methods since we *know* which stars truly come from the GES-like system. Indeed, Milky Way-like galaxies that have accreted GES-like systems have been found in cosmological simulations. Mackereth et al. (2019) used EAGLE simulations and found that the GES in the observations are well-reproduced by progenitors with $10^{8.5} \lesssim M_\star \lesssim 10^9$ M$_\odot$. Fattahi et al. (2019) also explored the Auriga hydrodynamical cosmological zoom-in simulations (Grand et al. 2017) and found haloes with radially anisotropic stellar halo populations at higher metallicities, similar to GES. These are produced by progenitors with $10^9 \lesssim M_\star \lesssim 10^{10}$ M$_\odot$ that merged with the host galaxy $6-10$ Gyr ago. The properties of these haloes have been followed up in detail by Orkney et al. (2023) where they found a diversity in the GES progenitors in Auriga, for example in their diskiness, or in their associated satellite population. These works show that the simulations are extremely valuable in understanding the nature of GES.

In this paper, we aim to understand how the different GES selections compare to each other by systematically selecting from a sample constructed from APOGEE DR17 and *Gaia* DR3, and applying the same selections to Milky Way-GES systems in the Auriga simulations. In particular, we investigate the E-Lz, eccentricity, and Jr-Lz selections in both observations and simulations, and the [Mg/Mn] versus [Al/Fe] selection in the observations. We specifically aim to answer the following questions: (1) What is the overlap between each selection and what contamination do we have from other accreted and in-situ components? (2) What progenitor properties do we infer from these different selections? (3) What selection is the purest and what selection is the most complete? In trying to address these questions, we want to impart more careful deliberation on how we select GES stars, as this would affect the sample data and ultimately the nature and picture of the GES progenitor that we paint from this.

This work is organized as follows: In Section 2 we describe our observational data and the selections made within. In Section 3 we compare the different GES samples and their inferred progenitor properties. Section 4 focuses on the Milky Way-GES systems in the Auriga simulations, applying the selections we use in the observations, and investigating the performance of these selections. In Section 5 we discuss and explore other avenues of validating the GES selections, and lastly, we summarize the results from this work in Section 6.

2 ACCRETED STARS FROM AN OBSERVATIONAL PERSPECTIVE

In this section, we construct a chemodynamical data set from the *Gaia* Data (e)DR3 (Gaia Collaboration 2021, 2023a) and APOGEE DR17 (Abdurro'uf et al. 2022). We take the positions, radial velocity, and proper motions from *Gaia* DR3 as well as the *Gaia* Early DR3-derived distances from Bailer-Jones et al. (2021) for calculating dynamical properties of the sample.

The APOGEE data, taken in the *H*-band (1.5–1.7 μm) with $R \sim$ 22 500, contains individual element abundances for up to 20 species for 733 901 stars. Of the elements available, we specifically use the Mg, Mn, and Al which are useful for selecting accreted stars purely from a chemical standpoint (e.g. Hawkins et al. 2015; Das, Hawkins & Jofré 2020; Carrillo et al. 2022).

To ensure the quality of our sample, we make the following cuts in this cross-matched data set:

(i) parallax error/parallax < 0.20
(ii) radial velocity error < 2.0 km s^{-1}
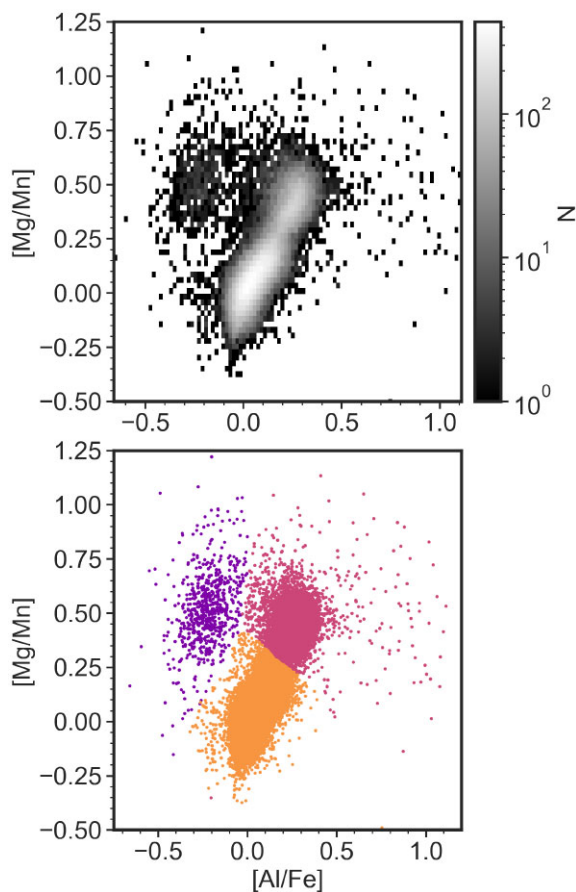
Figure 1. [Mg/Mn] versus [Al/Fe] from APOGEE DR17 used for the chemical selection of accreted stars. Top panel: 2D histogram showing the distinct group of stars at [Al/Fe] <0 and and [Mg/Mn] > 0.25 that corresponds to accreted stars. Bottom panel: Three separate regions in the diagram that roughly correspond to accreted stars (Region 1, purple), thick disc (Region 2, pink), and thin disc (Region 3, orange).

(iii) STARFLAG = 0 and SNR > 70.[1]
(iv) ASPCAPFLAG = 0 and ASPCAP_CHI2 < 25.[2]
(v) [Mg/Fe], [Mn/Fe], [Al/Fe], and [Fe/H] are not null

This selection criteria naturally produces a sample of stars with effective temperatures that are well within the suggested values for APOGEE (e.g. $4000 < T_{eff}/K < 6500$). To reduce the contamination from *in-situ* material, we also avoid the disc and apply a cut of $|z| > 1$ kpc. With these cuts, we have a resulting sample of 41 534 stars. We use the Cautun et al. (2020) potential within the galpy package (Bovy 2015) to derive the orbital properties of this *Gaia*-APOGEE cross-matched sample. We discuss later in detail in Section 2.2 the effects of observational uncertainties on the derived properties. From the parent sample of 41 534 stars, we select GES through various ways, as will be discussed next.

2.1 Chemical selection of accreted stars in observations

The [Mg/Mn] versus [Al/Fe] plane with APOGEE data has extensively been used to select the accreted population of the Milky Way.
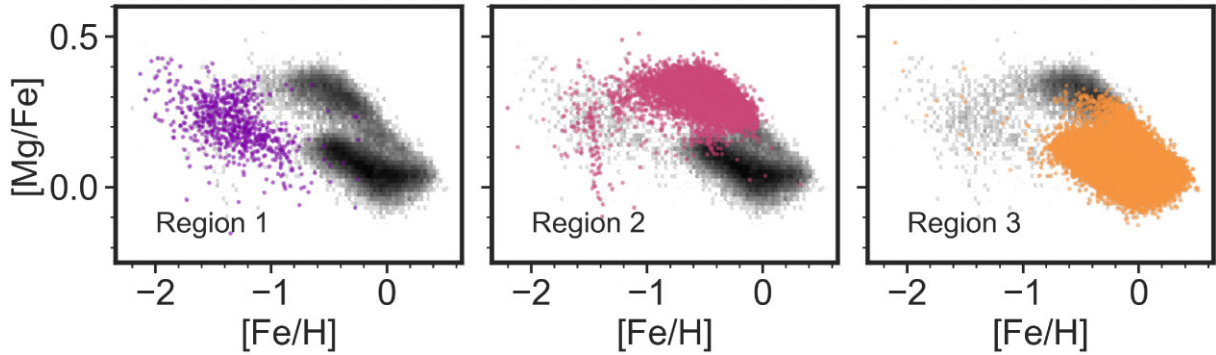
[1]for good spectra
[2]for converged stellar parameters
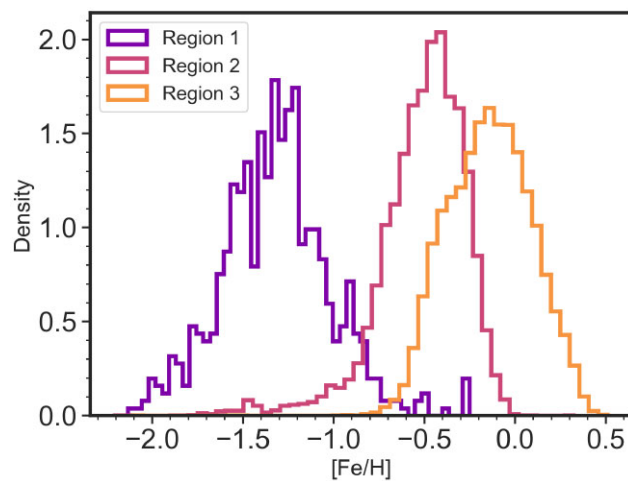
Figure 2. MDFs of the three regions in [Mg/Mn] versus [Al/Fe] (see Fig. 1) that correspond to the accreted stars (Region 1), thick disc (Region 2), and thin disc (Region 3).

This is an effective way to distinguish accreted versus *in-situ* material because Mg is an indicator of core-collapse supernovae, Mn is an indicator of Type Ia supernovae, and Al has been shown to be lower for dwarf galaxies compared to *in-situ* stars (e.g. Hasselquist et al. 2021). The accreted stars separate nicely as a concentration of stars at [Mg/Mn] > 0.25 and [Al/Fe] < 0, away from the bulk of the *in-situ* material as shown in the top panel of Fig. 1. It is especially effective in selecting GES stars in the Solar neighbourhood and inner halo, showing agreement with previous studies in terms of their ages and orbital properties (Bonaca et al. 2020; Das, Hawkins & Jofré 2020), as well as in their detailed chemical abundances, especially in the neutron-capture elements (Matsuno et al. 2020; Aguado et al. 2021; Carrillo et al. 2022).

To illustrate how well this plane is able to separate different stellar populations in the Milky Way, we show the diagram split into three different regions as shown in the bottom panel of Fig. 1. Region 1 (purple) represents the accreted population, region 2 (pink) the thick disc, and region 3 (orange) the thin disc.

Fig. 2 illustrates the MDF and Fig. 3 shows the [Mg/Fe] versus [Fe/H] for these three regions. The Region 1 MDF, which is associated with the accreted halo population, has a peak at the lowest metallicity of all three regions at [Fe/H] ~ -1.2. This is in line with the typical [Fe/H] peak for GES from previous works (Das, Hawkins & Jofré 2020; Feuillet et al. 2021; Buder et al. 2022). There is a smaller secondary peak at [Fe/H] ~ -0.2 which is due to some contribution at lower [Mg/Mn] values, at the boundary with the pink region that corresponds to the thin disc. As shown in Fig. 3, these stars follow the lower track in the [Mg/Fe] versus [Fe/H] plane compared to the higher [Mg/Fe] *in-situ* population. At [Fe/H] < -1.5 however, the track changes slope, similar to what has been previously found for GES in the APOGEE data (e.g. Myeong et al. 2022). Recently, Feltzing & Feuillet (2023) looked into this diagram in greater detail and found that though the majority of stars in this region are accreted, some stars show kinematics more akin to the Milky Way disc. Accreted material from other systems can and do exist in this region (see Horta et al. 2023 for a detailed exploration of the halo substructures in APOGEE). None the less, the majority of our sample in this region comes from GES given that it dominates the accreted population for $R_{GC} < 20$ kpc (Naidu et al. 2020) and we are probing a local region given our cuts in the observations.

Figure 3. [Mg/Fe] versus [Fe/H] of the three regions in [Mg/Mn] versus [Al/Fe] as defined in Fig. 1 that correspond to the accreted stars (Region 1), thick disc (Region 2), and thin disc (Region 3).

The Region 2 MDF has a peak at [Fe/H] ~ -0.4, in line with values for the thick disc (Kordopatis et al. 2013; Gaia Collaboration 2018). There is a long tail towards lower metallicity, which is due to contamination from both *in-situ* halo (Gallart et al. 2019) and accreted halo populations. The middle panel of Fig. 3 shows that the Region 2 stars mainly track the [Mg/Fe] versus [Fe/H] trend for the thick disc, but also show some contamination from stars that are associated with the accreted halo.

Lastly, the Region 3 MDF has a peak at solar metallicity, much like what is expected for the thin disc of the Galaxy. The boundary between the thin and thick discs are not as distinct in the [Mg/Mn] versus [Al/Fe] plane compared to the boundary of the *in-situ* population and that of the accreted population (see top panel of Fig. 1). Thus, there is a bump – not just a tail – for the lower metallicity half of the Region 3 MDF due to contributions from the thick disc. This contamination is especially apparent in the [Mg/Fe] versus [Fe/H] plane showing a considerable portion of the thick disc track being occupied by Region 3 stars in Fig. 3. None the less, based on the MDF and the [Mg/Fe] versus [Fe/H], splitting the different concentrations in the [Mg/Mn] versus [Al/Fe] is a robust way of distinguishing accreted versus *in-situ* populations from each other.

In addition to the chemistry, the kinematics for these three regions in the [Mg/Mn] versus [Al/Fe] plane are consistent with the accreted halo, thick disc, and thin disc stellar populations for regions 1, 2, and 3, respectively. In Fig. 4, we show the Galactocentric spherical velocity components for these regions. These are defined as follows: V_r is the radial component, V_ϕ is the azimuthal component, and V_θ is the polar component. We derived these quantities following section 2 in Bird et al. (2019) and folded in the errors from the distance, proper motions, radial velocities from *Gaia* through resampling 100 times and drawing a new velocity each time. The distributions in Fig. 4 are created from the median spherical velocity component for each star. The uncertainties in the three velocities are $\sim 15 \, \mathrm{km \, s^{-1}}$ for Region 1, $\sim 5 \, \mathrm{km \, s^{-1}}$ for Region 2, and $\sim 3 \, \mathrm{km \, s^{-1}}$ for Region 3.

Region 3, which we associate with the thin disc, shows the lowest dispersion in V_r, V_ϕ, and V_θ with $V_\phi \approx 220 \, \mathrm{km \, s^{-1}}$ therefore showing rotation with the disc, and V_r and V_θ centred at zero. The region associated with the thick disc (Region 2) shows a larger dispersion in all three velocity components compared to the thin disc, as is expected for a more kinematically hot component. In addition, the V_ϕ for the thick disc is lower than the thin disc's at $\sim 170 \, \mathrm{km \, s^{-1}}$. Lastly, the accreted halo component (Region 1) shows the largest dispersion in all three spherical velocity components and show no rotation with V_ϕ centred at $0 \, \mathrm{km \, s^{-1}}$.

Having shown the discerning power of this combination of elements, we therefore use the [Mg/Mn] versus [Al/Fe] to chemically select accreted halo stars in the APOGEE-*Gaia* cross-match without any additional cuts in kinematics. To do this less arbitrarily, we broke down the different regions using the Gaussian Mixture Model (GMM) from `scikit-learn`. We use Bayesian Information Criterion (BIC) to determine the best number of components to describe our data. We ran GMM using 3 to 20 components, and using BIC, we determine the best number of components to be 11. This appears to be too many components to explain the different stellar populations in this chemical plane, with the majority of them in the *in-situ* region. However, for our purposes, this is less important as we are focusing on the accreted halo population. In fact, there is only one associated component in the accreted halo region as the concentration of these stars is easily separated from the rest as shown in the top panel of Fig. 1. We selected the stars with ≥ 70 per cent probability of belonging to the accreted component in the [Mg/Mn] versus [Al/Fe] diagram as given by the `predict_proba()` function in GMM and consider this the purely chemically selected GES sample. Das, Hawkins & Jofré (2020) and Carrillo et al. (2022) make a further cut, i.e. [Fe/H] ≤ -0.7 to remove contamination from kinematic thick disc stars. However, we do not perform this as we want to compare the different levels and sources of contamination among the various ways GES is selected in their very simplest forms. *With this selection, we have a sample of 351 GES stars.*

2.2 Dynamical selection of accreted stars in observations

The more common way for selecting GES stars is through their phase-space information and dynamics. We apply such selections in the APOGEE-*Gaia* cross-match, specifically in energy-angular momentum (E-Lz; Helmi et al. 2018; Horta et al. 2023), eccentricity (*e*; Naidu et al. 2020; Myeong et al. 2022), radial action-angular momentum (Jr-Lz; Feuillet et al. 2021; Buder et al. 2022; Limberg et al. 2022), and action diamond (Myeong et al. 2019; Lane, Bovy & Mackereth 2022) spaces. We note that we *do not* make any other additional cuts, such as in chemistry, in creating these GES samples. We also explored the effects of observational uncertainties on these dynamical selections. Due to computational time, we folded in the uncertainties from the distance, proper motions, and radial velocities for only 0.5 per cent of the parent sample to understand their effects. For each star, we drew a new orbital property 100 times given the input parameters and their uncertainties, and inspected the median and standard deviation from this exercise. With this subsample, we note (1) the median difference in the orbital property between the
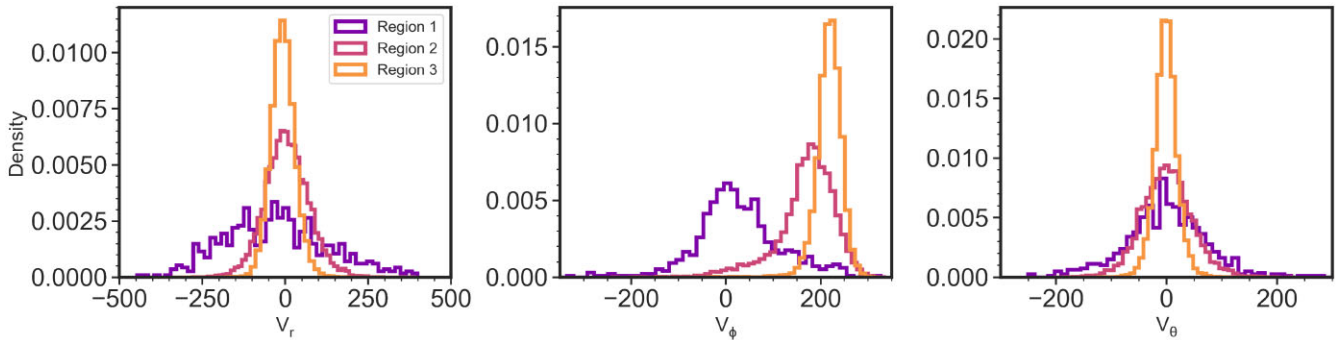
Figure 4. Distributions of kinematics in Galactocentric spherical coordinates showing the radial (V_r), azimuthal (V_ϕ), and polar (V_θ) velocity components of the three regions in [Mg/Mn] versus [Al/Fe] (see Fig. 1) that correspond to the accreted stars (Region 1), thick disc (Region 2), and thin disc (Region 3). Positive V_ϕ corresponds to rotation with the disc.

run with and without resampling and (2) the median error from the resampling as follows: energy = (192, 2325) km^2 s^{-2}, Lz = (6, 78) kpc km s^{-1}, Jr = (6, 50) kpc km s^{-1}, Jz = (2, 22) kpc km s^{-1}, eccentricity = (0.01, 0.031), apocentre = (0.07, 0.72) kpc, pericentre = (0.03, 0.22) kpc, and z$_{max}$ = (0.11, 0.92) kpc. From this test, we can see that the differences and errors in the orbital properties would have a negligible effect on the dynamical selections because they are significantly smaller in magnitude than the cuts applied.

For the E-Lz method, we applied the same cut in Lz as originally used in Helmi et al. (2018) i.e. $-1500 < L_z < 150$ kpc km s^{-1} although more recent works using *Gaia* DR3 have further shown the efficacy of the E-Lz selection space for GES (see Gaia Collaboration 2023b). Note that the positive (and therefore prograde) part of this selection does not cover as large of a range in Lz but this was originally done to avoid *in-situ* material. We modify the bound in energy from -1.8×10^5 km^2 s^{-2} to -1.6×10^5 km^2 s^{-2}. This cut was determined from visual inspection to also avoid the region largely dominated by *in-situ* material and that associated with Heracles in the deeper part of the potential (Horta et al. 2021). The difference in the appropriate energy bound is likely due to the difference in the Milky Way potential that we used. *With the E-Lz selection, we have a sample of 619 GES stars.* Similar to our chemical selection, we do not employ additional cuts outside of E and Lz in order to compare these selection methods in a more straightforward way.

Next, we explore the GES selection in eccentricity. From the APOGEE-*Gaia* cross-match, we apply a cut in eccentricity using $e > 0.7$ as similarly done by Naidu et al. (2020) with the H3 survey. This selection reproduces the highly radial, 'sausage'-like component of the halo in V_ϕ versus V_r plane, which is one of the initial ways GES was discovered with *Gaia* data (Belokurov et al. 2018). A massive satellite like the GES would have been affected significantly by dynamical friction, and the orbits of its most bound stars would have been more radialized with time (Amorisco 2017; Vasiliev, Belokurov & Evans 2022); GES, therefore, has a large portion of stars at highly eccentric orbits. As this selection is quite simplistic, it is (1) prone to contamination of high-eccentricity, *in-situ* stars and (2) misses the low-eccentricity tail of GES. We keep these caveats in mind throughout our comparisons and again, similar to the two other previously mentioned selections, we do not make any further cuts outside of eccentricity. *With the eccentricity selection, we have a sample of 1253 GES stars.*

We also use the radial action, Jr, and angular momentum, Lz of the sample to select highly radial stars in the APOGEE-*Gaia* cross-match. This selection, made in the $\sqrt{(J_R)}$-Lz space, was originally introduced by Feuillet et al. (2020) using the SkyMapper and *Gaia*

surveys as a purer selection for GES stars determined from its MDF. The cut is made with $30 \leq \sqrt{(J_R)} \leq 50$ (kpc km s^{-1})$^{1/2}$ and $-500 \leq$ Lz ≤ 500 kpc km s^{-1}. In the original paper, this cut produces a sample of stars with a narrow MDF (e.g. dispersion of 0.34 dex) centred at [Fe/H] = -1.17. *With the same Jr-Lz selection, we have a sample of 144 GES stars.* Similar to the other selection methods we have already introduced, we do not make any further cuts in chemistry or kinematics in order to understand the effects of purely selecting in this frame.

Lastly, we use the action diamond to select highly radial stars. One axis of the action diamond is $(J_z - J_R)/J_{tot}$ and the other is J_ϕ/J_{tot} where $J_{tot} = |J_z| + |J_R| + |J_\phi|$ and $J_\phi = L_z$. Along the $(J_z - J_R)/J_{tot}$ axis, values closer to -1 correspond to stars with radial orbits while those closer to 1 correspond to stars with polar orbits. On the other hand, along the J_ϕ/J_{tot} axis, values closer to -1 correspond to stars with retrograde orbits while those closer to 1 are stars with prograde orbits. We employ a similar selection as Lane, Bovy & Mackereth (2022) i.e. $|L_z/J_{tot}| < 0.07$ and $(J_z - J_R)/J_{tot} < -0.3$ for the GES stars, as they found that this action diamond selection gives the purest sample of GES stars among six other kinematic methods. *With this action diamond selection, we have a sample of 157 GES stars.*

3 CHEMICAL VERSUS KINEMATIC SELECTION IN OBSERVATIONS

3.1 GES samples from the different methods

We now look at the performance of each selection method by showing the selected GES stars in the other spaces we considered for selection. This is seen in Fig. 5 – from left to right shows the GES stars (orange markers) selected from [Mg/Mn]-[Al/Fe], E-Lz, eccentricity, Jr-Lz, and action diamond, respectively, projected onto the other selection spaces from top to bottom. We also include V_ϕ–V_r, even though we do not use this space for our selection, because this shows the elongated feature along V_r that is attributed to GES. The whole *Gaia*-APOGEE data set from which we culled out the different GES samples is shown as the greyed out background. In rows 4 and 5 of Fig. 5, we mark the selection box for the Jr-Lz and action diamond methods, respectively, with a purple box.

3.1.1 [Mg/Mn]-[Al/Fe]

This chemical selection by construction only occupies the accreted area (i.e. Region 1 from previous discussion) in the [Mg/Mn] versus [Al/Fe] diagram. This GES sample is largely at Lz≈0 kpc km s^{-1}
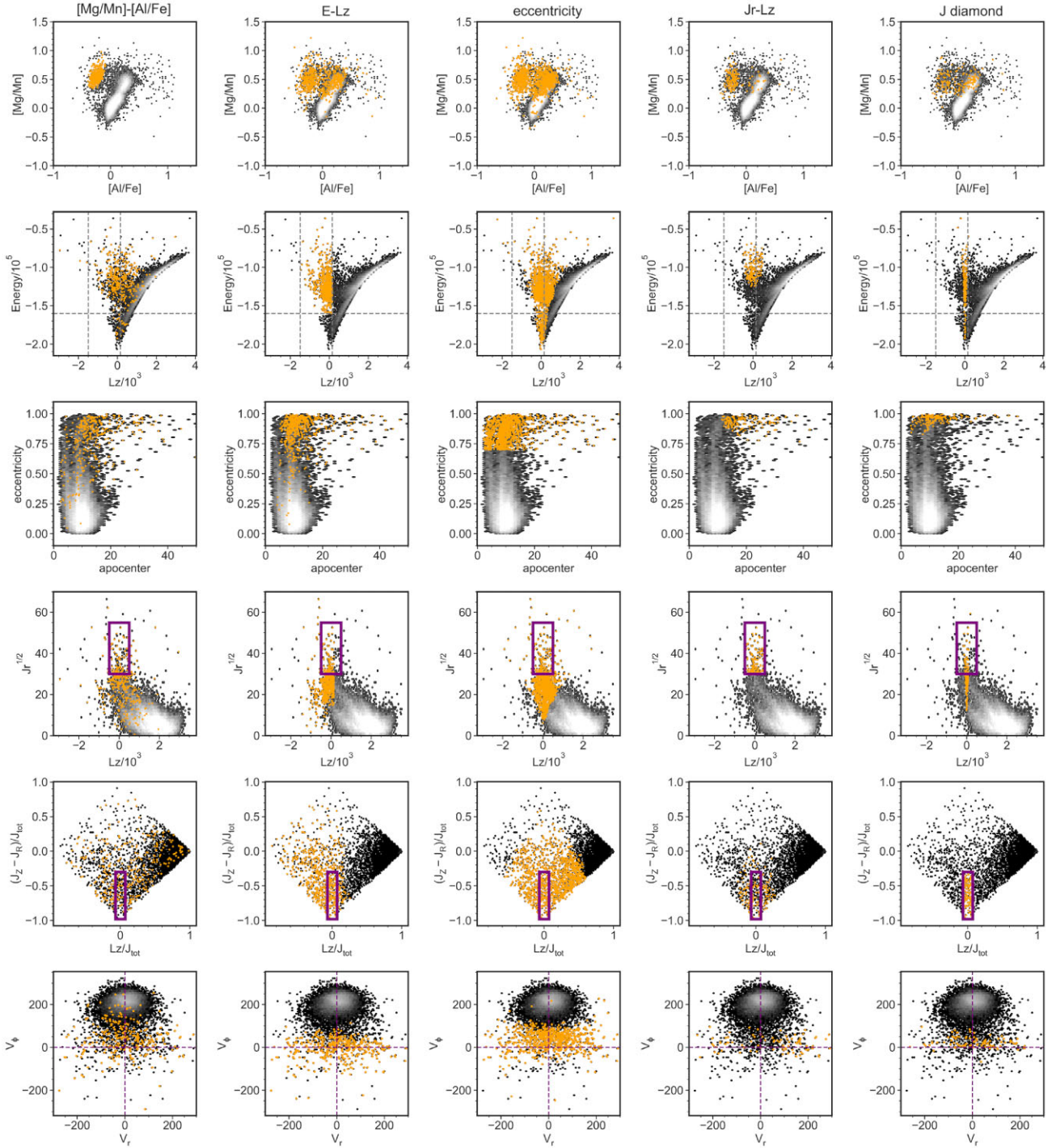
Figure 5. *Different GES samples projected onto the various chemo(dynamical) diagrams.* Columns from left to right: selections in [Mg/Mn]-[Al/Fe], E-Lz, eccentricity, Jr-Lz, and action diamond, respectively. Rows from top to bottom: projections of each GES sample onto [Mg/Mn]-[Al/Fe], E-Lz, eccentricity, Jr-Lz, action diamond, and V_ϕ-V_r. The GES samples are shown in orange and the greyed out background is the parent Gaia-APOGEE sample. The selection box for the Jr-Lz and action diamond methods are marked with a purple box in rows 4 and 5, respectively.

and with energy $> -1.6 \times 10^5$ km^2 s^{-2}. Some GES stars go deeper into the potential with even lower energies, and some stars show retrograde motion, reminiscent of another accreted population, Sequoia (Myeong et al. 2019). Around 66 per cent (236) of these GES stars have high eccentricities i.e. $e > 0.75$, and this sample has a median apocentre of 13.9 kpc but even reaches values >50 kpc. The lower eccentricity stars ($e < 0.75$) within this selection have a

lower median apocentre of 9.5 kpc. We also project these chemically selected GES stars onto the Jr-Lz space. We have marked the Jr-Lz box as defined in Feuillet et al. (2021) to guide the eye. The majority of the stars have high Jr, and interestingly, the floor for our sample lies at a lower $\sqrt{\text{Jr}}$ bound at $\sim$22 (kpc km/s)$^{1/2}$ instead of 30 (kpc km/s)$^{1/2}$ as defined for the box. This could be due to the difference in the potential the stars were integrated in i.e.

we use Cautun et al. (2020) while Feuillet et al. (2020, 2021) use MWPotential14 as included in `galpy`. Buder et al. (2022) similarly find that their chemical selection reach lower $\sqrt{\text{Jr}}$ than the Feuillet et al. (2020) selection using the McMillan (2017) potential. There are 129 stars (36 per cent) below ∼22 (kpc km/s)$^{1/2}$ and these lower Jr stars show higher dispersion in Lz compared to the higher Jr stars. We show these chemically selected GES stars in the action diamond where we have also marked the GES selection box as used by Lane, Bovy & Mackereth (2022) and Myeong et al. (2019). This GES selection occupies a large region of this diagram but for the most part have $(J_z - J_R)/J_{\text{tot}} < 0$ and therefore move more radially. Only 35 stars within the chemical selection overlap with action diamond box. Lastly, the V_ϕ-V_r of these GES stars are indeed elongated, showing stars with high dispersion in V_r at 112 km s^{-1} but also a high dispersion in V_ϕ at 78 km s^{-1}. A handful of stars (i.e. 54 stars, 15 per cent) seem to show kinematics more akin to the *in-situ* population with $V_\phi > 100$ km s^{-1} and V_r centred at 0 km s^{-1}. This is in line with Feltzing & Feuillet (2023) that found Milky Way disc stars overlapping with this GES selection.

3.1.2 *E-Lz*

These GES stars occupy a larger space in the [Mg/Mn]-[Al/Fe] plane, covering both the accreted and thick disc regions. Indeed, 309 stars (49 per cent) are in the region associated with the *in-situ* material in this chemical space. This sample of GES stars are more retrograde by construction as shown in the E-Lz diagram. Most of the stars are concentrated towards Lz∼0 kpc km s^{-1} but there are also regions of stars with more retrograde motions and at higher energies. The majority of this GES sample (82 per cent) have eccentricities greater than 0.75, showing an increasing range in apocentre with increasing eccentricity. In the Jr-Lz space, these stars have a larger range in Jr and extends to the downward retrograde branch that Feuillet et al. (2021) associates with Sequoia. Similar to the chemical selection, the majority of the stars do not lie within the Jr-Lz selection box. We note, however, that the Jr-Lz box was constructed to select more purely and is therefore a tighter constraint. In the action diamond space, it is clear that the majority of these stars move more radially with $(J_z - J_R)/J_{\text{tot}} < 0$. This GES selection spans a large range in $(J_z - J_R)/J_{\text{tot}}$ and L_z/J_{tot}, with 141 stars overlapping with the action diamond selection box (Lane, Bovy & Mackereth 2022), which is 88 per cent of the GES stars selected from the action diamond method. These GES stars also show the classic 'Sausage' feature in V_ϕ-V_r with a noticeably smaller dispersion compared to the [Mg/Mn]-[Al/Fe] selection in the first column.

3.1.3 *Eccentricity*

Similar to the E-Lz selection, the eccentricity method covers the regions in the [Mg/Mn]-[Al/Fe] diagram corresponding to both the accreted and *in-situ* populations, with 824 stars (64 per cent) that overlap with the *in-situ* population. This sample of GES stars are also centred around Lz∼0 kpc km s^{-1}, although this is by construction as a consequence of using the eccentricity. This selection includes stars that are even deeper in the potential, crossing the energy bound that has been put forth as the region corresponding to Heracles, Kraken, and Aurora (Kruijssen et al. 2020; Horta et al. 2021; Belokurov & Kravtsov 2022; Myeong et al. 2022) that have similar chemistry, different posited origins, but all part of the inner Galaxy (Rix et al. 2022). Within this sample, there are 272 stars in this lower energy (i.e. $< -1.5 \times 10^5$ km^2 s^{-2}) substructure. Projecting these stars onto

the [Mg/Mn]-[Al/Fe] diagram, only nine of them are in the accreted region and the rest (263 stars) are in the *in-situ* region.

By construction, the stars in this sample have eccentricities larger than 0.75. Among these stars however, there are still some substructures; there is a concentration of stars with apocentres at 6 kpc or less (237 stars, 19 per cent) and are therefore contained to the centre of the Galaxy. Interestingly, these stars also span both the accreted (26 stars) and *in-situ* (237 stars) regions in the [Mg/Mn]-[Al/Fe] diagram. In the Jr-Lz space for the eccentricity-selected sample, the majority of the stars lie below the selection box. This diagram also shows more clearly the overlap with the *in-situ* prograde material, which is apparent as a shell starting at $\sqrt{\text{Jr}} \sim 5$ (kpc km/s)$^{1/2}$ that curves up and to the right to higher Lz values. It is also quite remarkable that although the E-Lz and eccentricity selections seem the most similar of the methods we have explored, the populations they are probing are quite different in Jr-Lz. The eccentricity-selected GES stars barely go down the low Jr and retrograde branch while the E-Lz selected stars do populate this region. In the action diamond space, these stars are dominated by radial action while constrained in Lz, such that they appear to occupy a smaller diamond in the diagram. By construction, this eccentricity cut selects the elongated feature in V_ϕ-V_r, though interestingly, these stars are not centred at $V_\phi = 0$ km s^{-1}. This is mostly driven by the *in-situ* contamination, as previously noted.

3.1.4 *Jr-Lz*

The Jr-Lz selection has the least number of stars but it successfully selects mostly accreted stars in the [Mg/Mn] versus [Al/Fe] diagram. This dynamical selection results in 116 out of the 144 stars being in the accreted halo region (Region 1) without any additional cuts, with the remaining 28 stars being in the *in-situ* regions. This is, in fact, very similar to the contamination found by Limberg et al. (2022) in their exploration of the Jr-Lz selection for GES stars, i.e. at ∼18 per cent. In the E-Lz plane, these stars are the most restricted in Lz, but are still centred around Lz∼0 kpc km s^{-1}. They are also at higher energies compared to the other methods with E $> -1.3 \times 10^5$ km^2 s^{-2}. These stars exhibit high eccentricities (i.e. >0.75) and apocentres greater than 10 kpc, and they show the elongated feature in V_ϕ-V_r. These stars are bound within the selection box in Jr-Lz since they were selected in this plane, and the density of the stars decrease with increasing Jr. As we noted in the other selection methods, the lower bound in Jr for this selection is higher compared to the others. These stars overlap largely with the GES selection in the action diamond space, perhaps unsurprisingly as both methods select stars with high J_R. Feuillet et al. (2021) similarly used APOGEE (DR16) data, applied the same selection method, and found 299 stars associated with GES. Their GES stars occupy high-V_r lobes similar to our Jr-Lz GES sample, although this signature in our work is not as strong due to the smaller sample. Interestingly, the chemical cartography of the Milky Way from Gaia Collaboration (2023b) also showed that stars within this Jr-Lz selection show higher z_{max} and an average [Fe/H] ≈ -1.2 (see their fig. 31).

3.1.5 *Action diamond*

Finally, we look at the GES stars selected from the action diamond. In the [Mg/Mn] versus [Al/Fe] diagram, 69 stars (43 per cent) lie in the accreted region in [Mg/Mn] versus [Al/Fe] while the rest i.e. 91 stars, seem more *in-situ*-like in their chemistry. Contamination from *in-situ* material in this selection has been similarly found
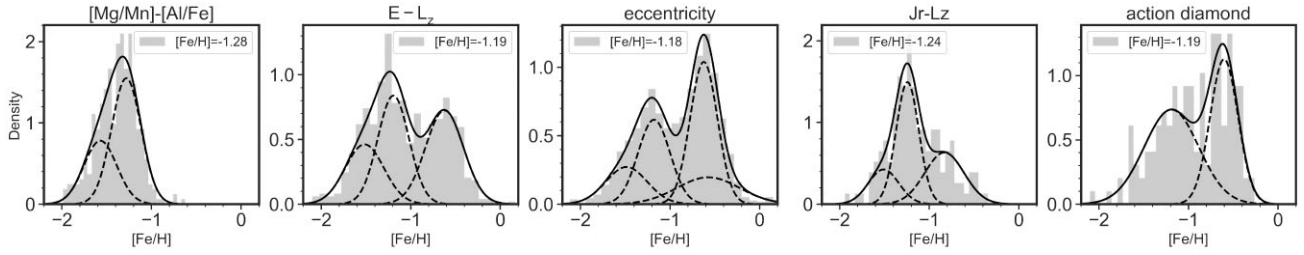
Figure 6. MDF of GES from the different selection methods as indicated in the title of each panel. The black line indicates the total best fit from the Gaussian mixture modelling while the solid dashed lines are the individual Gaussians. The associated GES [Fe/H] peak is noted in the legend of each subpanel.

Table 1. [Fe/H] and mass estimates for GES from the different selection methods. The columns are labelled as follows: (1) Selection method, (2) Number of stars in the GES sample from the selection, (3) GES [Fe/H] from the GMM, (4) Weight associated with the GES component in the GMM given between 0 to 1 where 1 is the highest, (5) Stellar mass, $M_\star$, derived from the mass–metallicity relationship (MZR) in Kirby et al. (2013) using the [Fe/H] in column 3 (6 & 7), $M_\star$ derived from Ma et al. (2016) MZR at $z = 0$ and 2, respectively, (8) $M_\star$ derived from the observationally motivated redshift-evolved MZR from Naidu et al. (2022), essentially shifting the $z = 0$ Kirby et al. (2013) MZR down by ∼0.3 dex, (9) halo mass, $M_{\rm halo}$, of GES using the stellar mass–halo mass relations from Behroozi et al. (2019) at $z = 0$ and 2, with the $M_\star$ from column 5 as input.

(1) Method	(2) N	(3) [Fe/H]	(4) GMM weight	(5) $M_{\star, z=0}$, K13 $10^8 M_\odot$	(6) $M_{\star, z=0}$, M15 $10^8 M_\odot$	(7) $M_{\star, z=2}$, M15 $10^8 M_\odot$	(8) $M_{\star, z-\mathrm{evol}}$, N22 $10^8 M_\odot$	(9) $M_{\mathrm{halo}, z=0, 2}$, B19 $10^{11} M_\odot$
[Mg/Mn]−[Al/Fe]	356	−1.28 ± 0.03	0.64	2.37 ± 1.74	1.72 ± 0.32	19.15 ± 3.24	24.51 ± 20.27	0.69, 1.15
E-Lz	630	−1.19 ± 0.03	0.37	4.53 ± 1.92	2.72 ± 0.43	31.19 ± 5.18	43.66 ± 26.56	0.96, 1.61
eccentricity	1279	−1.18 ± 0.03	0.28	5.26 ± 2.96	3.02 ± 0.50	34.72 ± 5.62	52.96 ± 51.17	1.04, 1.74
Jr-Lz	144	−1.24 ± 0.02	0.47	3.13 ± 2.02	2.11 ± 0.34	23.80 ± 3.73	33.90 ± 23.59	0.80, 1.33
Action diamond	160	−1.19 ± 0.09	0.56	4.75 ± 2.91	2.79 ± 0.47	32.37 ± 5.52	46.57 ± 37.96	0.99, 1.65

by Feltzing & Feuillet (2023) in comparing this method to other selections. This GES sample is centred at Lz∼0 kpc km s^{-1} by construction and though most of the stars are at high energies, 19 of them are below -1.5×10^5 km^2 s^{-2}. This action diamond-selected GES sample has stars with highly eccentric orbits with $e > 0.8$, therefore also occupying the 'sausage' region in V_ϕ-V_r. Though this range in eccentricity is quite similar to that of the Jr-Lz selection, the action diamond method also contains stars with lower apocentres (i.e. <6 kpc). Compared to the Jr-Lz selection which was similarly found to select a purer GES sample, the action diamond also includes stars with lower J_R, delving into the region that overlaps more with the *in-situ* material.

3.2 Metallicity distribution functions

We next explore the MDF of the resulting GES samples from each selection method, and these are shown in Fig. 6. As discussed in the previous section, each method has varying levels of contamination from non-GES stellar populations, both *in-situ* and accreted (i.e. Sequoia). The contamination is also more prominent when looking at the MDFs, especially for the E-Lz, eccentricity, and action diamond selections that clearly show multiple peaks.

We next broke down each MDF into a mixture of Gaussian distributions to further ascertain the level of contamination and to find the associated distribution and the [Fe/H] peak of GES. We first determined the best number of components by minimizing the BIC and performed Gaussian Mixture Modelling (GMM) using the `sklearn` package. The individual Gaussian distributions are shown in Fig. 6 as the dashed lines with the total, best-fitting mixture model shown as the solid lines. Now with the different peaks identified, we assigned the GES peak informed by previous works with ranges between −1.4 and −1.1 dex (e.g. Das, Hawkins & Jofré 2020; Feuillet et al. 2020; Naidu et al. 2020; Bonifacio et al. 2021; Buder

et al. 2022) and list these values in Table 1. Specifically, the MDF peak for the GES are at −1.28, −1.19, −1.18, −1.24, and −1.19 for the chemical, E-Lz, eccentricity, Jr-Lz, and action diamond selections, respectively. It is interesting how the chemical selection provides the lowest peak in the MDF. This difference between the chemical and dynamical selections has indeed been previously noted for the APOGEE data (see discussion in Buder et al. 2022) and is likely a result of such selections methods probing different stellar populations of GES.

We make sense of the other sources of contamination by investigating the origin of the individual metallicity distributions. The negatively skewed MDF from the chemical selection (first panel) necessitates two Gaussians to describe the data best. On one hand, it is known that massive Milky Way satellites have negatively skewed MDFs (Kirby et al. 2013) and in that case, the lower [Fe/H] peak is most likely associated with the same system i.e. GES. Indeed, Amarante et al. (2022) found in their simulations of a GES-like galaxy merging with a Milky Way-like galaxy that there is an associated low-[Fe/H] tail with GES; these stars end up with lower eccentricities and lower Lz (i.e. more retrograde). On the other hand, we do find stars that are typically associated with Sequoia on the basis of their dynamics in this GES selection, and in that case, they would be contributing to the lower metallicity tail of the distribution. The lower metallicity Gaussian has [Fe/H] at −1.57 dex, and is weighted at 36 per cent from the mixture modelling. Interestingly, Sequoia is also seen to peak at [Fe/H] ≈–1.6 (Myeong et al. 2019; Naidu et al. 2020). However, from Fig. 5, the highly retrograde Sequoia component contributes minimally to this selection. Therefore, even if Sequoia is part of the negative metallicity tail, the majority of the tail is likely part of GES.

A similar lower [Fe/H] peak is also observed for both the E-Lz, eccentricity, and Jr-Lz selections. These stars are largely associated with the GES based on their dynamics. The E-Lz, eccentricity,

and action diamond selections all have clear higher [Fe/H] peaks corresponding to the *in-situ* material. The prominence of the higher [Fe/H] peak roughly tracks the contamination measured for the E-Lz and eccentricity selections: their weights from the GMM are 39 per cent and 55 per cent,[3] respectively while the *in-situ* contamination as determined from Sections 3.1.2 and 3.1.3 are 47 per cent and 64 per cent. On the other hand, though the high [Fe/H] Gaussian component in the action diamond selection has higher density in Fig. 6, the GES component contributes a larger weight at 56 per cent due to the large dispersion associated with it. Interestingly, the Jr-Lz selection is best fit with three Gaussians. This is surprising as the motivation for this selection comes from producing a well-behaved, normal MDF as determined by Feuillet et al. (2020). However, this seems to be due to the spatial cut in z, causing a larger portion of the higher metallicity population to be missed. Lastly, the action diamond method is best fit with two Gaussians, with a higher metallicity component corresponding to the *in-situ* material as previously noted in Section 3.1.5 and seen in Fig. 5.

On the basis of the number of components, the [Mg/Mn]-[Al/Fe] and the action diamond selections seem the purest in selecting GES stars. These selections also have the highest weights for the Gaussian component associated with GES, at 64 per cent and 56 per cent, respectively. The MDFs of massive Milky Way satellites show that the negative skew such as that from the [Mg/Mn]-[Al/Fe] selection *is* what is to be expected for GES. The action diamond has previously been shown to purely select GES stars as well, but we find that this still produces a clear multimodal MDF without any further cuts. Another pure selection is the Jr-Lz method that although needed to be fitted with three Gaussian components, shows significant purity in terms of [Mg/Mn] versus [Al/Fe] (see Section 3.1.4). In addition, this selection weighs the GES as the dominant component at 47 per cent. For the E-Lz, eccentricity, and action diamond methods that all show multimodal MDFs, the GMM tends to give a higher [Fe/H] for GES. In fact, we find that with increasing contribution from the higher [Fe/H] Gaussian component – i.e. going from the [Mg/Mn] versus [Al/Fe], to Jr-Lz, to E-Lz, to action diamond, to the eccentricity method – the GES component goes to higher [Fe/H] as well.

3.3 Stellar masses from mass–metallicity relationship

We now use the [Fe/H] from the GMM of the different selection methods to estimate the stellar mass of GES. We do this by taking advantage of the mass–metallicity relationship (MZR). We use the relation defined from observations of satellites (Kirby et al. 2013) in the local Universe, as well as the MZR at $z = 0$ and 2 determined from the FIRE simulations which Ma et al. (2016) showed agrees with observations for $z = 0$ (Tremonti et al. 2004) up to $z \sim 3$ (Mannucci et al. 2009). We take the $z = 2$ relation to roughly illustrate the MZR at the time that GES was accreted onto the Galaxy. We also calculate the stellar masses from the redshift-evolved MZR as determined from disrupted versus intact satellites (Naidu et al. 2022) – this is essentially shifted by ~ 0.3 dex from the $z = 0$ relation from Kirby et al. (2013). These stellar mass estimates are listed in Table 1.

The [Mg/Mn]-[Al/Fe] method gives the lowest [Fe/H] for GES and therefore gives the lowest stellar mass estimate across the different MZRs. Conversely, the eccentricity method has the highest [Fe/H] and resulting stellar mass for GES. From the Kirby et al. (2013) MZR, the mass for GES has a range of $2.37–5.26 \times 10^8 \, M_\odot$, a factor

of ~ 2.2 difference depending on the selection method. From the Ma et al. (2016) redshift-evolving MZR, the $z = 0$ stellar masses are an order of magnitude smaller than the $z = 2$ stellar masses, which has a range of $1.91–3.47 \times 10^9 \, M_\odot$. Using the redshift-evolved MZR determined from Naidu et al. (2022) gives the highest stellar mass range of $2.45–5.30 \times 10^9 \, M_\odot$.

With the many ways that the GES stars are defined, there are as many (or even more) ways that its stellar mass has been derived through stellar density (Mackereth & Bovy 2020), N-body simulations (Naidu et al. 2021), matching to cosmological hydrodynamical simulations (Mackereth et al. 2019), globular clusters (Kruijssen et al. 2020; Callingham et al. 2022), and chemical evolution modelling (Fernández-Alvar et al. 2018; Vincenzo et al. 2019; Hasselquist et al. 2021) that give a wide range of total stellar masses spanning an order of magnitude, from 3×10^8 to $7 \times 10^9 \, M_\odot$. We similarly find a wide range in stellar masses based on the MZR which is modulated by the accretion redshift. More recently, Lane, Bovy & Mackereth (2023) found a lower stellar mass estimate for GES of $1.45 \times 10^8 \, M_\odot$, but our mass estimates agree with their findings within the errors.

Next, we determine the halo mass of GES. We adopted the redshift-evolving stellar mass-halo mass (SMHM) relation from Behroozi et al. (2019) and calculated the halo mass at $z = 0$ and $z = 2$ which are listed in Table 1. Because we want to be as observationally motivated as possible, we use the stellar masses derived from the Kirby et al. (2013) MZR at $z = 0$ in calculating the halo mass of GES. The GES halo mass has a range of $0.69–1.04 \times 10^{11} \, M_\odot$ with the $z = 0$ relation and a range of $1.15–1.74 \times 10^{11} \, M_\odot$ with the $z = 2$ relation. This puts the GES-Milky Way total mass merger ratio at $11–17$ per cent using the total Milky Way mass from Deason et al. (2021) pre-LMC infall and the $z = 2$ GES halo mass. On average, however, the Milky Way halo mass at $z = 2$ would have been ~ 20 per cent of its present day mass, thus increasing the merger ratio to $56–85$ per cent. On the other hand, the stellar mass merger ratio for the GES-Milky Way collision has a range of $2–5$ per cent at $z = 0$ and $19–53$ per cent at $z \sim 2$ if we similarly assumed as Helmi et al. (2018) that the stellar mass of the Milky Way was $\sim 10^{10} \, M_\odot$ at the time of the merger. In comparison, Grand et al. (2020) showed that the GES-Milky Way stellar mass merger ratio could be as low as 5 per cent at infall with the Auriga simulations, while Helmi et al. (2018) derived a total mass merger ratio of 25 per cent to produce the Toomre diagram in the observations. Lane, Bovy & Mackereth (2023) also estimated a minor merger ratio of 1:8 at the time of accretion. There is visibly a wide range of merger ratios for the Milky Way-GES collision, with the redshift evolution of the GES and Milky Way masses being a main factor. With that said, we similarly find the merger ratios are larger for the total mass than the stellar mass, as previous studies have noted.

A huge caveat in deriving the stellar and total mass of GES is the redshift dependence of the MZR and the SMHM, both pushing our estimates to higher values. We therefore aim to be conservative and report the mass estimates from the $z = 0$ relation from observations as lower limits to the true GES mass. Based on our exploration of the MDFs (Section 3.2), we deem the [Mg/Mn] versus [Al/Fe] method to be the best in selecting GES stars in the observations, with $M_\star = 2.37 \pm 1.74 \times 10^8 \, M_\odot$ and $M_{halo} = 0.69–1.15 \times 10^{11} \, M_\odot$. Though this is a lower limit, we can also place an upper limit to the total stellar mass of GES such that it does not exceed the total stellar halo mass i.e. $\sim 1.4 \times 10^9 \, M_\odot$ (Deason, Belokurov & Sanders 2019).

We have so far explored the different ways we select GES stars in observations, their respective sources and level of contamination, and the resulting stellar and halo mass estimates of the GES progenitor. Depending on the selection method, the stellar mass estimate could

---

[3]Here, we combine the weights from the two Gaussians at [Fe/H] > -1

Table 2. Properties of Auriga haloes with GES-like systems. The columns are labelled as follows: (1) Auriga halo that contains a GES progenitor, (2) GES peak $M_\star^{true}$ given in $10^9\,M_\odot$, (3) GES peak total mass M_{total} (i.e. M_{200}) given in $10^{11}\,M_\odot$, (4) Milky Way-like host's peak $M_\star$ in $10^{11}\,M_\odot$, (5) Milky Way-like host's peak M_{total} in $10^{12}\,M_\odot$, (6) Average [Fe/H] of GES stars, (7) Percentage of GES stars in the observational window as described in Section 4.

(1)	(2)	(3)	(4)	(5)	(6)	(7)
						Per cent GES in
Halo	GES $M_\star^{true}$ $10^9 M_\odot$	GES M_{total} $10^{11} M_\odot$	host $M_\star$ $10^{11} M_\odot$	host M_{total} $10^{12} M_\odot$	$\langle[Fe/H]\rangle$	window
Au-5	3.83	1.26	0.71	1.19	-0.39	12.78
Au-9	1.88	1.76	0.63	1.16	-0.85	6.47
Au-10	0.97	0.39	0.62	1.02	-0.65	1.29
Au-15	2.53	1.26	0.43	1.04	-0.50	8.22
Au-17	0.38	0.33	0.79	1.02	-0.98	0.83
Au-18	1.44	0.75	0.84	1.39	-0.70	1.62
Au-24	2.56	1.09	0.77	1.57	-0.60	4.16
Au-26	10.54	3.33	1.14	1.72	-0.39	19.18
Au-27	4.08	1.72	1.03	1.85	-0.53	9.40

differ by a factor of 2, but depending on the adopted redshift, it could differ by a factor of 10. In the next section, we benchmark our different selection methods with simulations where we actually know where the stars come from – i.e. from GES or not.

4 ACCRETED STARS FROM A SIMULATION PERSPECTIVE

In the first part of this work, we investigated the different selections for GES stars in observational data. Now we check which of these methods is best in selecting GES stars using simulations where we have absolute knowledge of which stars belong to the GES-like progenitor. Through this, we hope to answer the question: *Can we really pick and choose?*

We use the Auriga hydrodynamical simulations (Grand et al. 2017), which contain 30 high-resolution, cosmological zoom-in simulations of Milky Way-mass haloes (i.e. with virial mass[4] $1 - 2 \times 10^{12}\,M_\odot$) selected from the dark matter-only 100^3Mpc^3 periodic box of the EAGLE project (Crain et al. 2015; Schaye et al. 2015). The cosmological parameters were adopted from the Planck Collaboration (Planck Collaboration XXIII 2014). These selected haloes were resimulated with the AREPO code (Springel 2010) at higher resolution. In this work, we use the resolution level named Level 4 in Grand et al. (2017) with mass resolution of $\sim 3 \times 10^5\,M_\odot$ for the dark matter particles and $\sim 5 \times 10^4\,M_\odot$ for the gas cells.

The simulation has a comprehensive prescription for galaxy formation physics that includes primordial and metal-line cooling, star formation and stellar feedback, chemical enrichment (from core-collapse supernovae, Type Ia supernovae, and winds from asymptotic giant branch stars), a sub-grid model for the interstellar medium, black hole formation and feedback, uniform photoionizing UV background, and magnetic fields (see Grand et al. 2017 for details). Notably, we use Auriga because it has been shown to contain Milky Way systems with GES-like mergers in the past, as explored in detail by Fattahi et al. (2019). We use the Fattahi et al. (2019) data set and briefly describe it below, but we refer the reader to the original paper for further details.

In Fattahi et al. (2019), star particles were considered '*in-situ*' if they were bound, according to the SUBFIND algorithm (Springel, Yoshida & White 2001), to the main progenitor of the Milky Way analogue at their formation time.[5] If the formation time is at $z > 3$, $z = 3$ association is adopted. Star particles bound to the main halo at $z = 0$ but that were previously born in a different halo (in the snapshot after the time of formation) were considered 'accreted'.[6]

From within the Auriga sample, Fattahi et al. (2019) determined a subset of ten haloes that contain accreted stars exhibiting high orbital anisotropy, $\beta > 0.8$ and high metallicity, [Fe/H] ~ -1, reminiscent of GES in the observations. They find that the stars contributing to the 'sausage' feature in the simulations come from a single progenitor in most cases, and in fact the most massive progenitor to the halo with stellar mass of $10^9 - 10^{10}\,M_\odot$ accreted 6–10 Gyr ago. We use this subsample of Auriga haloes in exploring the different observational selection methods when applied to the simulations. The properties of these haloes are listed in Table 2. We do not include Auriga 22 although it was in the GES sample from Fattahi et al. (2019) as our observational cuts result in too few star particles for our analysis.

For the dynamical properties, we use the orbital energy and actions (specifically Lz and Jr), as well as the apocentre and pericentre distances for star particles, determined by Callingham et al. (2022) for the Auriga haloes using AGAMA (Vasiliev 2019). This is especially important in selecting GES stars in similar ways compared to the observations.

To approximately recreate the same observable area, we limit our sample to ~ 10 kpc from an arbitrary solar viewpoint at 8.3 kpc and with Galactocentric radii larger than 3 kpc to avoid the bulge region. We similarly avoid the disc and apply a cut in $|z| > 1$ kpc. The GES stars contribute between 1 and 20 per cent to the stellar population in the observational window and we include these values in Table 2. With the sample of stars in this region for the nine MW-GES haloes, we applied the different GES selections which we discuss in the next section.

---

[4]Virial mass in the simulations is the mass within a sphere where the mean matter density is 200 times the critical density.

[5]In practice this is determined at the snapshot immediately following the birth time.

[6]From this definition, stars that formed from the gas stripped from the satellite are considered *in-situ*.

Table 3. Properties of the next four most massive accreted satellites (Column 1) of Au-18 after GES. The rest of the columns are defined as follows: (2) peak $M_\star^{\mathrm{true}}$ given in $10^9\,M_\odot$, (3) peak M_{total} given in $10^{11}\,M_\odot$, (4) Average [Fe/H] of stars associated with each progenitor, (5) Percentage of stars in the observational window associated with each progenitor.

Halo	$M_\star$ $10^9\,M_\odot$	M_{total} $10^{11}\,M_\odot$	$\langle[\mathrm{Fe/H}]\rangle$	Per cent in window
m1	0.55	0.25	-0.82	0.43
m2	0.25	0.21	-0.80	0.69
m3	0.07	0.11	-0.93	0.02
m4	0.06	0.09	-1.27	0.03

4.1 Selection of accreted stars in simulations

We highlight Auriga 18 (Au-18) to illustrate the different methods of selecting GES stars but note that we apply these selections to all of the haloes in Table 2. Following Callingham et al. (2022), we also look at the next four most massive contributors (labelled m1-m4) to the stellar halo of Au-18 listed in Table 3, with the most massive progenitor corresponding to GES. This is to give an idea of the other accretion events, in addition to the *in-situ* material, that can affect the purity and completeness of each selection method.

We used the same selection in the simulations as in the observations for the E-Lz, eccentricity, Jr-Lz, and action diamond methods (see Section 2). Unfortunately, the simulations do not contain element abundances for Mn and Al, making the exact comparison with the chemical selection unavailable. However, Tronrud et al. (2022) have shown that the chemistry in Auriga, specifically [α/Fe] and [Fe/H], are very promising in distinguishing accreted versus *in-situ* stars in the disc using neural network models. We note that the different assembly histories and orbital evolution of these haloes (by virtue of being taken from a cosmological simulation) pose differences in where exactly the GES stars lie in these diagrams, as well as where stars from other progenitors lie. However, the selections in observations are generally good approximations for the simulations. In the end, we aim to compare the performance of the different selections with respect to each other within the same halo, so adopting the cuts from the observations is a reasonable approach.

Fig. 7 shows the different GES samples (orange) in the observable window selected in E-Lz, eccentricity, Jr-Lz, and action diamond methods, from left to right. The 'true' GES stars in the defined observable window are also shown in blue. The contours indicate the density of stars from the top five massive merger events that contributed to the stellar halo and are assigned as follows: GES-black, m1-blue, m2-red, m3-green, and m4-magenta. The background grey histogram shows the density of all the stars in the observable window, which is dominated by the *in-situ* stars. Lastly, all these stellar populations are projected onto the E-Lz, eccentricity-apocentre, $\sqrt{\mathrm{Jr}}$-Lz, action diamond, and V_ϕ-V_r diagrams from top to bottom.

It is reassuring that for Au-18, the different selection methods *do* select the GES stars although to varying degrees. For all methods, the majority of the *in-situ* material in the observable window is avoided by the GES selection i.e. the selected GES stars overlap less with the bright grey regions of the background 2D histogram. This overlap is the greatest for the eccentricity selection, followed by E-Lz, then the action diamond, then lastly by Jr-Lz. We can also compare these GES selections to other accreted material. For example, m3 (green) and m4 (magenta) are both at higher energy and high eccentricity, so there are contaminants from these accretion events in the eccentricity selection without making further cuts in other properties. In the E-Lz selection method, the contamination from m2 (red) is greatly reduced

because they are at lower energies, while these stars remain in the GES selection if we only make a cut in eccentricity. The m2 (red), m3 (green), and m4 (magenta) accretion events also overlap with the action diamond box, though we can see in the other selection diagrams that making additional cuts in energy and apocentre will remove these contaminants. The Jr-Lz method seems to be the most selective, having the least overlap with the other accretion events. With that said, the largest contaminant in each GES selection *is* still the *in-situ* material, which will be discussed further in Section 4.2.2.

Again, we note that the overlap among the GES, *in-situ*, and other accreted stars is dependent on the halo and its assembly history, and we are merely showing the case for Au-18. However, this does give an idea of what contributes to our calculated purity and completeness values for each selection method. These values are calculated with respect to the GES stars that are within the observable window, instead of the total GES population. We do these for all nine haloes, and list these values in Table 4. There is variety in the purity and completeness from halo to halo, as some haloes are more GES-like than others. However, in general, the eccentricity method performs best in terms of completeness (62 per cent) and the Jr-Lz method is best in terms of purity (33 per cent). Conversely, the eccentricity method obtains the least pure sample (15 per cent) while the action diamond method is the least complete (6 per cent).

Now that we have a sample of GES stars from the different methods in the simulations, we can look into the MDFs from these selections versus the real MDF of GES, both for the total population and the observable window.

4.2 Metallicity distribution function

4.2.1 Total versus observed GES population

One way of distinguishing accreted versus *in-situ* material is by looking at their chemistry. We have shown earlier in the observations that regardless of the selection method, GES has a distinct MDF from the *in-situ* material centred at lower [Fe/H]. We similarly investigate this in the simulations. In Fig. 8, we show the normalized MDF of the corresponding GES in Au-18 as well as the *in-situ* material both for the total population (top panel) and the population within the observed window (bottom panel). We will refer to the metallicity from the total stellar populations as [Fe/H]$_{\mathrm{sim}}$ while those from the stellar populations in the observational window as [Fe/H]$^*_{\mathrm{sim}}$. We also include a lower mass accreted system with $M_\star = 1.4 \times 10^7\,M_\odot$ for additional comparison. Lastly, we mark the median [Fe/H] from the total population (dotted line) and those only in the observational region (dot-dash line) for the *in-situ* stars (purple) and GES (orange).

For the total population (top panel), there is a distinct progression of higher metallicity for higher stellar mass systems (see also Fattahi et al. 2020). That is, the GES MDF is centred at lower [Fe/H]$_{\mathrm{sim}}$ compared to the *in-situ* material whilst at higher [Fe/H]$_{\mathrm{sim}}$ compared to the lower mass accreted system. This stellar mass–metallicity relationship in the simulations is encouraging as we can potentially use the metallicity of the selected GES stars to estimate their progenitor's stellar mass as we have done in the observations.

In addition, the shapes of the MDF seem to be quite telling as well. The MDF of GES is negatively skewed, similar though to a lesser degree to that of the *in-situ* material. On the other hand, the MDF of the lower mass system is wider, and has more of a platykurtic distribution. These are reminiscent of the different MDFs of more massive versus less massive dwarfs around the Milky Way (Kirby et al. 2013).
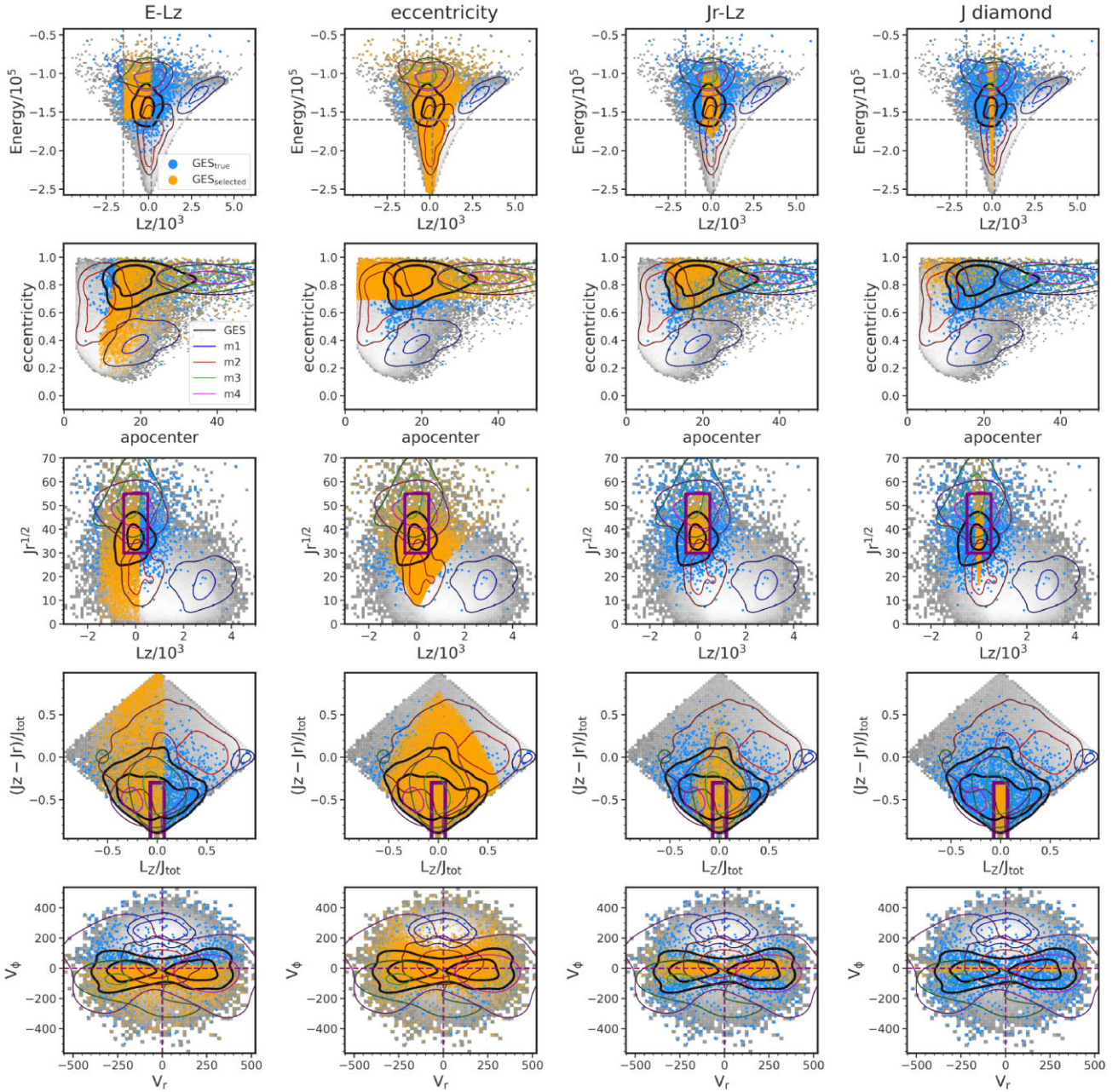
Figure 7. Different GES samples in Auriga 18 projected onto the various selection diagrams. Columns from left to right: selections in E-Lz, eccentricity, $\sqrt{\text{Jr}}$-Lz, and action diamond, respectively. Rows from top to bottom: projections of each GES sample onto E-Lz, eccentricity, $\sqrt{\text{Jr}}$-Lz, action diamond, and V_ϕ-V_r. The *selected* GES stars in the observational window are orange while the *true* GES stars in the same galactic region are light blue. All the stars in the observational window (which is dominated by the *in-situ* material) is shown as the grey background. The five most massive contributors to the accreted halo in the observable region are shown as contours (33rd and 66th percentile, for the inner and outer contours, respectively) with GES-black, m1-blue, m2-red, m3-green, and m4-magenta.

The MDFs for the observed window (bottom panel) however are quite different, due to multiple factors. The biggest factor is our spatial cut in z. This preferentially removes the highest metallicity component of the *in-situ* material i.e. the star-forming disc. Therefore, the GES and *in-situ* MDFs seem much closer to each other. The spatial cut in galactocentric radius ($\lesssim 20$ kpc) also preferentially selects the higher metallicity component of the GES population. This is because in the Auriga simulations, GES-like systems have been shown to have a negative metallicity gradient before merging with the Milky Way-like galaxy, wherein its central

regions have higher [Fe/H] compared to its outskirts (Orkney et al. 2023). The more centrally located stars pre-merger are more bound and stripped later, ending up at lower galactocentric radii in the host galaxy, post-merger. Therefore, with our spatial cut, we tend to select these higher metallicity stars from the progenitor. On the flip side, for some of the haloes where the GES sinks deeper into the centre i.e. $\lesssim 3$ kpc, our galactocentric radius cut removes the most-metal rich part of the GES population. Interestingly, previous works with the Auriga simulations (e.g. Ciucă et al. 2022; Orkney et al. 2022; Pinna, Walo-Martín & Grand 2023), found that a starburst is induced during

Table 4. Purity and completeness in terms of percentage from the different selection methods for the nine Auriga haloes investigated in this work.

halo	E-Lz		Eccentricity		Jr-Lz		Action diamond	
	Purity	Completeness	Purity	Completeness	Purity	Completeness	Purity	Completeness
Au-5	45	52	41	78	66	24	56	10
Au-9	16	41	15	58	23	13	20	6
Au-10	3	39	6	83	12	23	11	10
Au-15	14	43	21	52	29	8	23	5
Au-17	9	36	5	71	15	23	8	6
Au-18	29	49	13	83	33	33	23	11
Au-24	27	45	14	62	35	24	21	7
Au-26	46	4	38	46	40	10	41	4
Au-27	24	3	28	54	33	9	32	5
Median	**24**	**41**	**15**	**62**	**33**	**23**	**23**	**6**
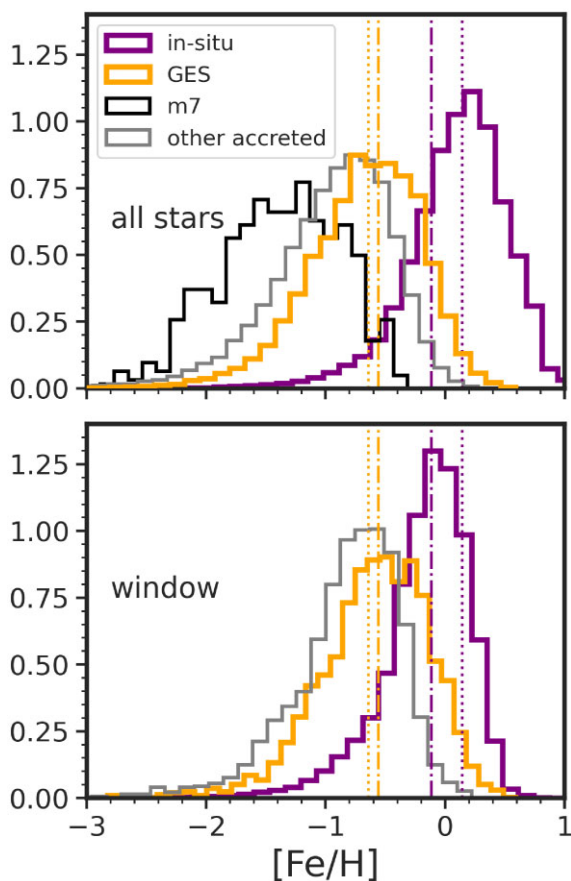
Figure 8. MDFs for the total population and in the observational window. Top: MDF of the total populations for GES (orange), in-situ stars (purple), a lower mass system with $M_\star = 1.4 \times 10^7\,M_\odot$ (m7, black), and the rest of the accreted material (grey) for Au-18. The MDFs peak at higher [Fe/H] for higher mass systems. Bottom: MDF for the GES (orange), *in-situ* (purple), and the rest of the accreted (grey) stars in our defined observational window for Au-18. Both panels include the median [Fe/H] from the total population (dotted line) and in the observational region (dot-dash line) for the *in-situ* stars and GES. The MDFs of the *in-situ* stars and GES seem much closer to each other due to the removal of the higher metallicity disc stars and some lower metallicity GES stars. The MDFs are normalized such that the area under the curve equals one.
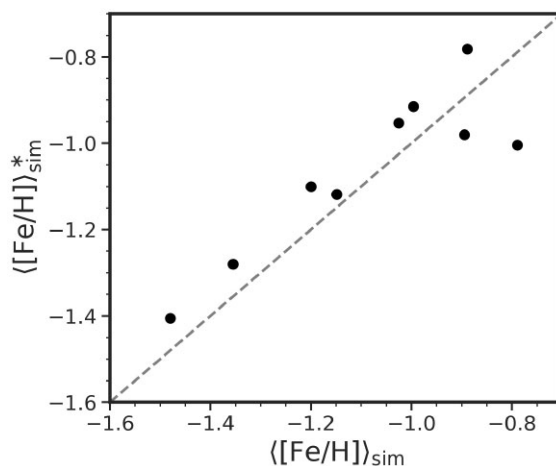
Figure 9. Comparison of the metallicity for the total GES population (x-axis, [Fe/H]$_{\rm sim}$) versus the GES stars in the observable window (y-axis, [Fe/H]$_{\rm sim}^*$). Dashed grey line shows 1:1 correspondence. The majority of the GES material in the observable window has higher [Fe/H] than the total population. Due to the progenitor's negative metallicity gradient, i.e. its central region tends to have higher metallicity stars than the outskirts, and our spatial selection of $\lesssim 20\,{\rm kpc}$, we miss the lower metallicity tail of the progenitor that are at greater distances.

the MW-GES merger, increasing the overall stellar metallicity.[7] We show the relation between the average metallicity in the observable window, [Fe/H]$_{\rm sim}^*$, versus the total population, [Fe/H]$_{\rm sim}$, in Fig. 9. Table 5 also lists these [Fe/H] values for the observable window and the total population (in addition to their associated stellar mass estimates, discussed later Section 4.3). Although the $\langle$[Fe/H]$\rangle$ of the GES population is different from the observable window compared to that from the total population, it is reassuring that there is still a positive trend – i.e. the haloes with higher $\langle$[Fe/H]$\rangle$ for the total GES population generally have higher $\langle$[Fe/H]$\rangle$ from the observable window as well, with an offset of $\sim 0.1-0.2$ dex. This difference therefore affects the estimated $M_\star$ from the [Fe/H] of GES stars in total versus those in the observable window, as later discussed in Section 4.3. In general, the $\langle$[Fe/H]$\rangle$ in the observable window tends to be higher compared to the $\langle$[Fe/H]$\rangle$ of the overall population. There are two haloes where this is not the case, due to these progenitors being more massive and their higher metallicity stars sinking deeper

---

[7]By definition, these stars would be considered *in-situ* in this work.

Table 5. Mean [Fe/H] of the *true* GES population within the observational window and the whole progenitor, and the resulting $M_{\star}$ derived from the mass–metallicity relationship of disrupted satellites shown later in Fig. 11.

Halo	$[Fe/H]^{*\downarrow}_{sim}$ observed	$[Fe/H]^{*\downarrow}_{sim}$ all	$\log_{10}(M^{est}_{\star})$ observed	$\log_{10}(M^{est}_{\star})$ all
Au-5	-0.78	-0.89	10.08 ± 0.38	9.75 ± 0.38
Au-9	-1.28	-1.36	8.68 ± 0.38	8.48 ± 0.38
Au-10	-1.12	-1.15	9.12 ± 0.38	9.05 ± 0.38
Au-15	-0.91	-1.00	9.71 ± 0.38	9.46 ± 0.37
Au-17	-1.41	-1.48	8.37 ± 0.38	8.19 ± 0.37
Au-18	-1.10	-1.20	9.19 ± 0.39	8.92 ± 0.37
Au-24	-1.00	-0.79	9.48 ± 0.37	10.03 ± 0.38
Au-26	-0.98	-0.89	9.49 ± 0.37	9.75 ± 0.37
Au-27	-0.95	-1.03	9.60 ± 0.38	9.40 ± 0.38

into the centre of the Milky Way-like galaxy, which we exclude in our selection.

4.2.2 MDF from different selections

We now investigate the MDFs of GES from the observable window. Before this, however, we reiterate that the cuts we applied generally made the GES MDF closer to that of the *in-situ* population. And in fact, with the different selection methods explored, the MDFs have essentially been indistinguishable from each other. Therefore, separating the GES from the *in-situ* material based on their total MDF from each selection is non-sensical. Modelling this MDF to fit multiple components would produce artificial distributions that do not have physically motivated different stellar populations.

This effect is largely due to the high-mass satellites in Auriga being too metal-rich (by ~ 0.5 dex) compared to the observations, as has been shown in Fig. 13 from Grand et al. (2021). To alleviate this, we shifted the [Fe/H] of all tagged accreted material down by 0.5 dex, which we now call $[Fe/H]^{*\downarrow}_{sim}$. In Grand et al. (2021), the mismatch in the satellite MZR between the observations and simulations is stronger for GES-mass systems, which lessens and shows better agreement at the lower mass end. However, we apply the 0.5 dex shift in [Fe/H] to all accreted population because the lower mass systems contribute very minimally within the observable window, and even more so in the selection methods explored. We use $[Fe/H]^{*\downarrow}_{sim}$ in identifying the MDF of the GES versus *in-situ* stellar populations.

With the $[Fe/H]^{*\downarrow}_{sim}$, we show the normalized (un-normalized) MDFs for the different GES selections in Au-18 as the grey solid histograms in the top row (bottom row) of Fig. 10. We perform GMM on these MDFs to determine the number of different components that contribute to the distribution. The optimal number of components was determined through BIC, and these Gaussians are shown as the dashed lines in the first row of Fig. 10 (the total best-fitting GMM is shown with the solid line). In determining the peak that is associated with the GES, we are informed by our exploration in Section 4.1, specifically pertaining to the purity of the selection summarized in Table 4. This tells us that in the simulations, the majority of the selected GES stars from any of the methods *are not* associated with GES. However, of all the accreted material, the GES merger is typically the most dominant contributor in the halo at these distances (see also Fattahi et al. 2019 regarding how this sample of Auriga haloes was selected).

With this in mind, we ascertained that it is the second most dominant component (which turns out to also be the second highest metallicity component) that is associated with GES. For example

and as shown in Fig. 10, the GMM for the MDF of Au-18 gives a GES that peaks at $[Fe/H]^{*\downarrow}_{sim} = -1.28, -0.81, -1.19,$ and -1.15 for the E-Lz, eccentricity, Jr-Lz, and action diamond selections, respectively. On the other hand, the actual, most dominant component for each selection method in the simulations corresponds to the *in-situ* material which peaks at a higher [Fe/H]. We tabulate the different mean $[Fe/H]^{*\downarrow}_{sim}$ for GES from each selection in Table 6 for all nine haloes investigated. We note again that we are quoting $[Fe/H]^{*\downarrow}_{sim}$ in the table, which are shifted lower by 0.5 dex from the true value in the observational window in the simulations, $[Fe/H]^{*}_{sim}$. Comparing the MDFs in the observations (Fig. 6) to these ones in the simulations (Fig. 10) highlights that the latter also look different in that they are not multimodal. The *in-situ* stars in the Milky Way have a more widely separable MDF from the GES stars in the observations, whereas they are less distinguishable in the simulations, as illustrated in Fig. 8. We note that these two effects i.e. the *in-situ* material dominating the selection instead of GES stars, and the GES and *in-situ* MDFs being less distinguishable from each other, are due to the following factors: (1) the larger stellar masses for the GES-like systems in Auriga especially with respect to the Milky Way host enable the galaxies to be more chemically enriched than is expected in the observations, (2) at a similar mass scale, the satellite galaxies in Auriga generally have higher [Fe/H] than seen in the observations, (3) the *in-situ* discs in Auriga are different and thicker than the real Milky Way's, therefore the spatial cut we applied is not getting rid of a lot of *in-situ* stars as we do in the observations, and (4) none of these are 'exact' GES in terms of their mass and orbit so the ratio of *in-situ* to GES stars can vary and will not be exactly like what we see in the Milky Way. Therefore, we expect the contamination in the simulations to be different and in fact much larger compared to the true contamination in the observations. None the less, exploring these haloes in the simulations is informative as they still correspond to a massive merger that happened at earlier epochs, and contribute significantly to the radially anisotropic stellar population in the halo, much like the GES in the observations.

We also show the MDFs for the true *in-situ* and GES populations from each selection in the bottom row of Fig. 10. This confirms our assignment of which Gaussian component corresponds to GES, aside from the eccentricity selection which was best fit to have three components, because of (1) the lower [Fe/H] peak of GES compared to the *in-situ* material and (2) the larger spread in the MDF of GES stars. One obvious drawback in using GMM to separate the *in-situ* versus GES MDFs is also highlighted here: the true *in-situ* and GES MDFs do not necessarily follow a Gaussian distribution, though we have applied this assumption. None the less, it is reassuring that the GES (orange histogram) and *in-situ* (purple histogram) populations make up the bulk of the sample in each selection method, as adding those two histograms together would give the total MDF (grey histogram).

4.3 Stellar mass determination

In the previous section, we have thoroughly explored the MDF of the GES population – as a whole (e.g. $[Fe/H]_{sim}$), in the observable window (e.g. $[Fe/H]^{*}_{sim}$ and $[Fe/H]^{*\downarrow}_{sim}$ which is shifted down by 0.5 dex), and with the different selection methods. Now we use the [Fe/H] estimates from the different GES selections to convert to a stellar mass estimate as we have done in the observations. In Fig. 8, there appears to be a mass–metallicity relationship in the satellites around the Milky Way systems in Auriga with the MDFs arranged from lowest to highest metallicity going from the least to the most massive system.
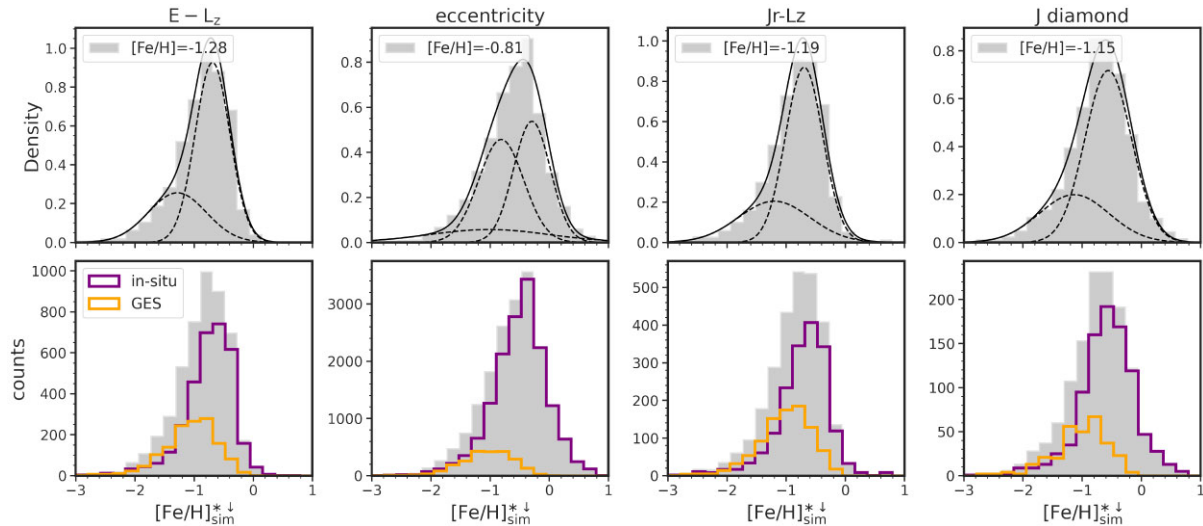
Figure 10. Metallicity distribution functions from different selection methods in the simulations. Top: Normalized MDF of the stellar populations from the different selection methods in the simulations. From left to right, we show the samples from the E-Lz, eccentricity, Jr-Lz, and action diamond selections. The black line indicates the total best fit from the GMM while the solid dashed lines are the individual Gaussians. The associated GES [Fe/H] determined from the GMM is noted in the legend of each subpanel. Bottom: Un-normalized MDF of the stellar populations from the selection methods showing the true *in-situ* (purple) and GES (orange) populations within each sample. The metallicity shown here is $[Fe/H]_{sim}^{*\downarrow}$i.e. the [Fe/H] in the observable window, with the accreted material shifted down by -0.5 dex. See Sections 4.2.1 and 4.2.2 for this discussion.

Table 6. Mean [Fe/H] of the GES component from the GMM in the observable window, and the associated $M_\star$ estimate for GES for the different selection methods in the simulations. The [Fe/H] listed in columns 2–5 are $[Fe/H]_{sim}^{*\downarrow}$, i.e. shifted down by 0.5 dex which was applied during the GMM (see Section 4.2.2). The stellar masses in columns 6–9 are derived from shifting back up the listed metallicities by 0.5 dex and using the MZR for the satellites in the simulations.

| (1) Halo	(2) $[Fe/H]_{sim}^{*\downarrow}$ E-Lz	(3) $[Fe/H]_{sim}^{*\downarrow}$ eccentricity	(4) $[Fe/H]_{sim}^{*\downarrow}$ Jr-Lz	(5) $[Fe/H]_{sim}^{*\downarrow}$ J diamond	(6) $\log_{10}(M_\star^{est})$ E-Lz	(7) $\log_{10}(M_\star^{est})$ eccentricity	(8) $\log_{10}(M_\star^{est})$ Jr-Lz	(9) $\log_{10}(M_\star^{est})$ J diamond
Au-5	-1.00 ± 0.11	-1.09 ± 0.08	-0.89 ± 0.05	-0.99 ± 0.04	9.45 ± 0.50	9.27 ± 0.43	9.78 ± 0.42	9.48 ± 0.40
Au-9	-1.01 ± 0.24	-1.04 ± 0.39	-0.93 ± 0.07	-1.17 ± 0.28	9.43 ± 0.76	9.33 ± 1.13	9.64 ± 0.41	9.07 ± 0.87
Au-10	-0.90 ± 0.24	-1.09 ± 0.35	-1.10 ± 0.26	-0.97 ± 0.12	9.71 ± 0.76	9.24 ± 0.99	9.18 ± 0.84	9.52 ± 0.49
Au-15	-0.96 ± 0.34	-0.63 ± 0.08	-0.75 ± 0.07	-0.89 ± 0.20	9.55 ± 0.99	10.45 ± 0.43	10.15 ± 0.44	9.79 ± 0.67
Au-17	-1.42 ± 0.31	-1.19 ± 0.31	-1.04 ± 0.13	-0.84 ± 0.16	8.26 ± 0.93	8.97 ± 0.90	9.39 ± 0.54	9.92 ± 0.55
Au-18	-1.28 ± 0.24	-0.81 ± 0.14	-1.19 ± 0.36	-1.15 ± 0.35	8.71 ± 0.75	9.98 ± 0.54	8.93 ± 1.09	9.05 ± 1.03
Au-24	-1.09 ± 0.19	-0.81 ± 0.07	-1.22 ± 0.20	-0.90 ± 0.09	9.21 ± 0.64	9.98 ± 0.44	8.87 ± 0.66	9.73 ± 0.46
Au-26	-1.15 ± 0.07	-0.87 ± 0.07	-1.11 ± 0.18	-0.90 ± 0.10	9.01 ± 0.41	9.81 ± 0.41	9.18 ± 0.63	9.72 ± 0.45
Au-27	-1.22 ± 0.25	-0.59 ± 0.03	-1.09 ± 0.14	-0.96 ± 0.11	8.86 ± 0.79	10.56 ± 0.40	9.24 ± 0.53	9.63 ± 0.49

We further explore this relationship and show the mean $[Fe/H]_{sim}$ versus peak stellar mass, $M_\star$, for the destroyed satellites around the Auriga haloes with GES-like mergers in Fig. 11. It is apparent that the MZR in the simulations holds for a large range in $M_\star$ and [Fe/H], specifically over the range where the GES progenitors lie. This has been similarly noted in previous studies on the Auriga satellites (e.g. Grand et al. 2021; Deason et al. 2023). We adopted a simple linear relationship to describe the MZR, given by $\langle[Fe/H]\rangle = 0.37 \times \log(M_\star/M_\odot) - 3.99$ with a scatter of 0.13 dex/$\log(M_\star/M_\odot)$.

We adopt this MZR to get the stellar mass estimates, $M_\star^{est}$, for the GES-like systems in Auriga listed in Tables 5 and 6. We shifted back up the $[Fe/H]_{sim}^{*\downarrow}$ in Tables 5 and 6 by 0.5 dex to $[Fe/H]_{sim}^{*}$ and used this metallicity to derive $M_\star^{est}$ such that the GES metallicities are back on the same scale as the MZR in the simulations. We illustrate the distribution of the different mass estimates for the nine Auriga halos in Fig. 12. Specifically, we show a violin plot of the $\log_{10}(M_\star^{est}/M_\star^{true})$ from the different selection spaces to more easily see their distribution with respect to the $M_\star^{true}$ (marked as dashed line). The white dot shows the median, the grey bar shows the first and third

quartiles, and the thin grey line shows the rest of the distribution, barring outliers. In general, all mass estimates are able to reproduce the $M_\star^{true}$ but to varying precision and accuracy. We note that again, a lot of the scatter here is mostly due to the different assembly histories of the haloes that we investigated. However, this is also why comparisons to the real GES stars in the observational window (pink) and its total population (brown) are elucidating.

Interestingly, the $M_\star^{est}$ derived from the GES stars in the observational window seems to be closer to the true value compared to if we take *all* of the GES stars. The former has the median value right at the true value with $\log_{10}(M_\star^{est}/M_\star^{true}) = 0.02$ while the latter is underestimated with $\log_{10}(M_\star^{est}/M_\star^{true}) = -0.24$. However, based on Fig. 11, more than half (5/9) of the GES in the simulations lie below the MZR. The offset between the MZR and the progenitors below this relationship is also larger compared to the offset with the progenitors that lie above it. The underestimated mass from the total population is therefore expected, while the 'correction' to the true value from the observational window is due to the bias towards the higher metallicity stars of the GES progenitor. We note however that
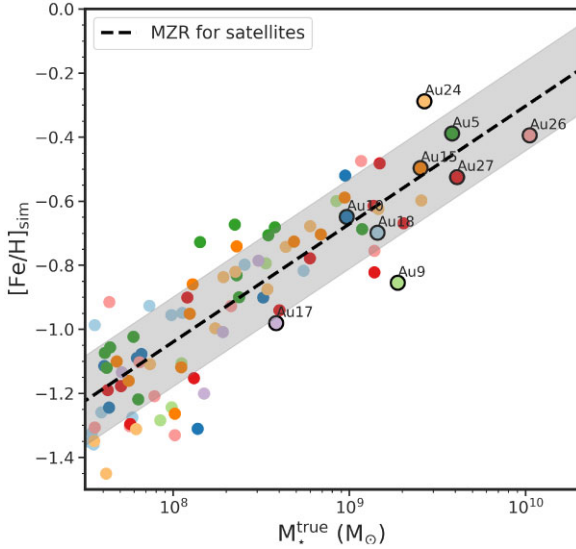
Figure 11. Mass–metallicity relationship (MZR) in the satellites around the Auriga haloes that have GES-like mergers. Every colour corresponds to a different satellite population around the same central galaxy. The GES in each halo is highlighted with a black circle. The MZR in the simulations is shown as a dashed black line with one sigma above and below this relation filled in grey.
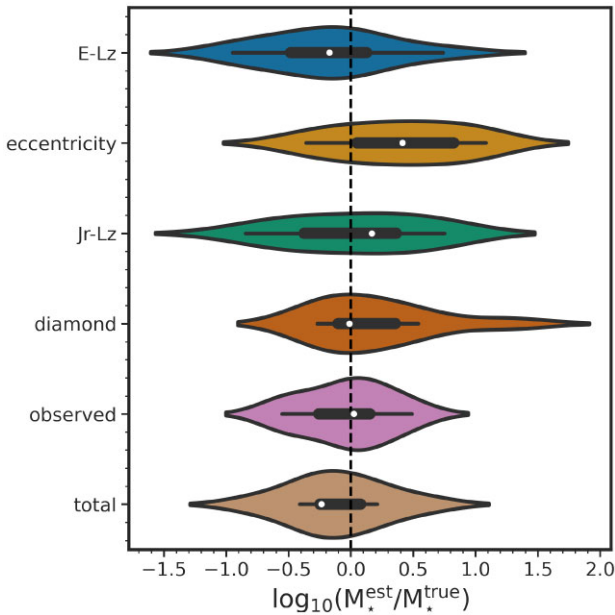
Figure 12. Violin plot of mass estimate ratios derived from [Fe/H] for the E-Lz, eccentricity, Jr-Lz, and action diamond methods as well as GES population in the observable window and the GES total population, ordered from top to bottom.

this distribution of the $M_\star^{est}$ from the total population happened by chance, purely based on where they lie on the MZR.

In terms of accuracy on $M_\star^{true}$, the action diamond (orange) method is the most accurate with median $\log_{10}(M_\star^{est}/M_\star^{true}) = -0.01$. The $M_\star^{est}$ distributions from the E-Lz (blue) and Jr-Lz (green) methods are also generally accurate, i.e. crossing the 1-to-1 line, though in bulk the E-Lz method underestimates the true mass with $\log_{10}(M_\star^{est}/M_\star^{true}) = -0.17$ (factor of 0.7 lower) and Jr-Lz overestimating with

$\log_{10}(M_\star^{est}/M_\star^{true}) = 0.17$ (factor of 1.5 higher). The eccentricity method is the least accurate with a median $\log_{10}(M_\star^{est}/M_\star^{true}) = 0.41$, which overestimates the $M_\star^{true}$ by a factor of 2.6. Although the eccentricity selection is the most complete, it also has the lowest purity. The fact that it is the least accurate in estimating the true $M_\star$ is in line with this. The converse is not true however; the purest selection, Jr-Lz, does not necessarily give the most accurate $M_\star^{est}$ but the action diamond does. This is also interesting as the action diamond by far has the lowest completeness of all dynamical selections and it has a comparable purity to the E-Lz selection. The efficacy of the action diamond method is likely due to the contaminating *in-situ* stars in the same selection region having distinctly higher metallicity (as they come from the innermost regions of the galaxy), which makes their MDF more easily separable from those of the GES stars. We do want to emphasize that these distributions are based only on nine GES-like systems and are therefore prone to small number statistics. None the less, it seems that a smaller GES sample size and a purer selection given by the Jr-Lz and action diamond methods, lend to more accurate stellar mass estimates. Meanwhile, a larger GES sample that is more complete but less pure such as that from the eccentricity method is the least accurate and overestimates the stellar mass.

This exploration of nine GES-like progenitors in the Auriga simulations has shown that each selection method is beneficial in their own right. The selection in Jr-Lz wins in terms of purity (33 per cent), eccentricity in completeness (62 per cent), and the E-Lz, Jr-Lz, but most of all the action diamond, in inferring $M_\star^{true}$. Interestingly, though unsurprisingly, the ⟨[Fe/H]⟩ for the total GES population is lower compared to the ⟨[Fe/H]⟩ for the GES population in the observational window because we are probing populations that were more centrally located in the progenitor pre-merger, which are also higher in [Fe/H]. Although the $M_\star$ estimate from the MZR for the total population is generally underestimated because the majority of the haloes happened to be below the MZR, the bias towards higher [Fe/H] in the observational window brings the $M_\star$ estimate closer to the true value.

5 DISCUSSION

In the first part of this work focusing on the observations (Section 2), we have determined that the estimated stellar mass of GES differs by a factor of $\sim$2 depending on the selection method. The [Mg/Mn]-[Al/Fe] selection gives the lowest estimate while the eccentricity selection gives the highest estimate based on their MDFs. The assumption that largely changes the $M_\star$ however is the adopted MZR which ranges from 2.37 to 5.26×10^8 $M_\odot$ assuming the $z = 0$ Kirby et al. (2013) relation to $2.45-5.30 \times 10^9$ $M_\odot$ assuming the disrupted dwarfs ($z > 0$) relation from Naidu et al. (2022). This is an order of magnitude difference! Suffice to say, the real $M_\star$ of GES is likely in between these values.

In the second part focusing on the simulations, we then test our method of identifying GES stars and similarly derive their $M_\star$ using the MZR for the disrupted satellites in Auriga. We are then able to compare this to the true value in the simulations shown in Fig. 12. There is not one method that gives the highest (or lowest) estimate for the $M_\star$ across *all* the Auriga haloes listed in Table 6 because of their different assembly histories. However, in general, the action diamond method gives the most accurate $M_\star^{est}$, while the eccentricity method is the least accurate and on average overestimates the $M_\star$ by a factor of $\sim$2.6. In the observations, the eccentricity method similarly gives the highest $M_\star^{est}$ for GES, followed by the action diamond, E-Lz, Jr-Lz, and lastly [Mg/Mn]-[Al/Fe]. The GES system in Au-24 follows a similar trend in the $M_\star^{est}$ as in the observations, which has a peak
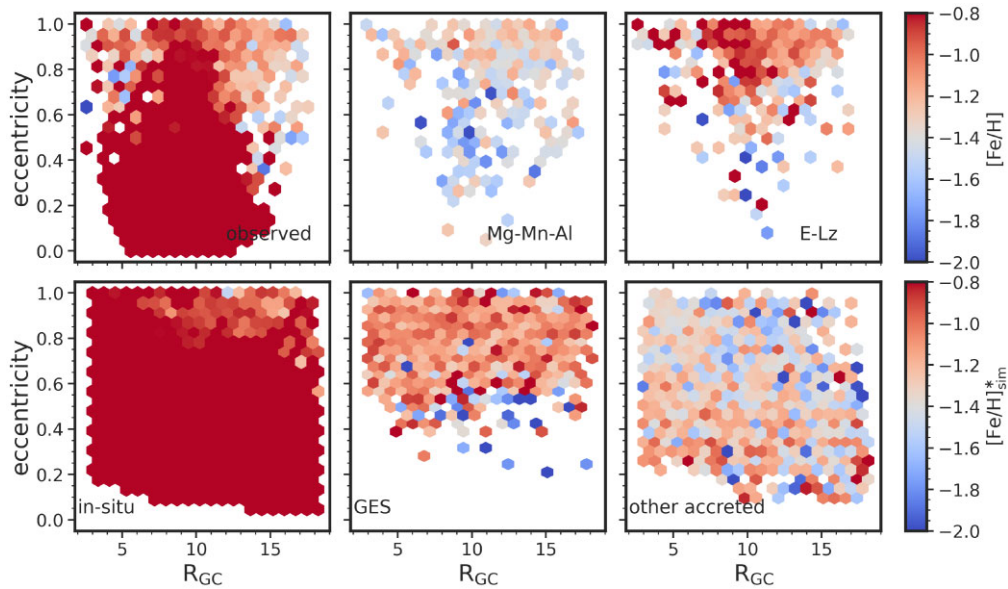
Figure 13. Eccentricity versus current Galactocentric radius, R_{GC}, coloured by metallicity. The top row shows this for all the observed data as well as the GES samples through the [Mg/Mn]-[Al/Fe] and E-Lz selection methods (from left to right, respectively). The bottom row shows the *in-situ*, GES, and all other accreted material in the observational window for Auriga 18. It is noteworthy that the *in-situ* stars have decreasing metallicity with higher eccentricity while the GES shows the opposite trend. The two selection methods in the observations qualitatively match our expectations from the simulations.

$M_\star^{\text{true}} = 2.56 \times 10^9 \, M_\odot$. From Table 6, the variation in $M_\star^{\text{est}}$ *within* each Auriga halo is larger in the simulations compared to that from the observations. We conjecture that this is driven by the GES being more massive and having higher metallicity in the simulations, making it less distinguishable from the *in-situ* population compared to the observed data and therefore affecting the [Fe/H] from which we derive a stellar mass.

For the majority of this work, we have gone through the details of deriving the $M_\star$ of GES. But for understanding whether or not we can pick and choose a GES selection method, it is worth looking into other progenitor properties as well.

5.1 GES eccentricity

We further check the validity of our selections by looking at the eccentricity of stars versus their current Galactocentric radius, R_{GC}, as shown in Fig. 13. The top row shows these properties as seen in the observations for all the stars (left), and the GES stars as selected from [Mg/Mn] versus [Al/Fe] (centre), and from E-Lz (right). We only show results from these two selection methods because they are the only ones that span a large range in eccentricity (see Fig. 5). The bottom row shows the same plot from the observable window for stars that are *in-situ* (left), from GES (centre), and from all other accreted material (right) for Au-18 in the simulations. In addition, both the top and bottom rows are coloured by [Fe/H] but we note that the ranges are different between the observations and the simulations.

It is apparent that the way *in-situ* stars occupy this space is different compared to the GES stars. First, the parent sample in the observations is clearly dominated by higher metallicity *in-situ* stars that in general show progressively lower metallicities with higher R_{GC}. This is similarly observed for the *in-situ* stars in the simulations as well. One difference however, is that the lower R_{GC} sample in the observations show lower metallicities at high eccentricity whereas this does not exist in the simulations. This is likely due to the differences in the Auriga discs versus the Milky Way in addition to

our simple selection function in the simulations. In contrast, the true GES stars in the simulations do not show such a progression which at first glance seems to counter our intuition about the presence of a metallicity gradient. However, this is because we are mainly looking at their current R_{GC} (over a small range in R_{GC}) which is not a preserved quantity. Of the two selection methods investigated here, the chemistry selection resembles GES better as suggested by the simulations. The E-Lz selection shows a metallicity gradient with higher [Fe/H] at lower R_{GC} which is due to contamination from *in-situ* stars as shown in Figs 5 and 6.

Another interesting distinction between the *in-situ* and GES stars is the opposite [Fe/H] gradient from low to high eccentricity. For the *in-situ* stars, the lower eccentricity stars have higher [Fe/H] which progressively goes to lower [Fe/H] at higher eccentricity, especially at higher R_{GC}. Indeed we see this both in the observations and the simulations. On the other hand, it is the opposite story for the GES stars – at lower eccentricity, though there are fewer stars in this region, the GES stars have lower [Fe/H].[8] A larger percentage of the higher [Fe/H] sample are in fact at higher eccentricities as well. This is reasonable if we consider a pre-existing negative metallicity gradient for the GES progenitor and that the stars that are stripped later are more centrally located and highly radialized due to dynamical friction (Amorisco 2017; Vasiliev 2019; Amarante et al. 2022). Indeed, we similarly see this in the observations for both selection methods. However, due to the contamination from *in-situ* stars in the E-Lz method, this [Fe/H] gradient along the eccentricity axis seems to be stronger compared to that from the [Mg/Mn] versus [Al/Fe] method and from the true GES population in Au-18.

We also show this eccentricity versus R_{GC} diagram for all other accreted material in the observable window in Au-18 to highlight that this looks different from the GES stars. For all other accreted material,

[8]We have investigated this diagram for the other Auriga haloes in this work and they in fact show stronger trends. However, we choose to show Au18 for consistency.
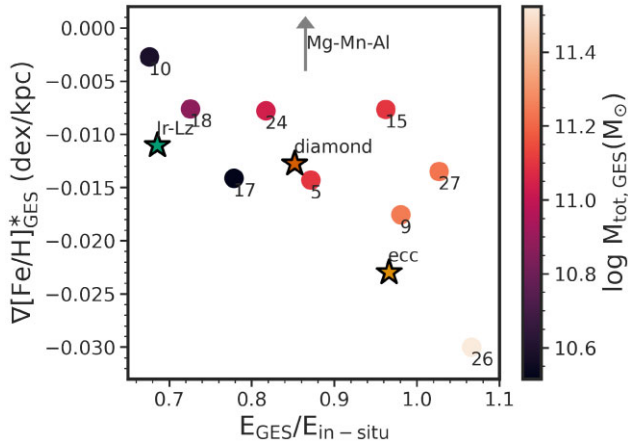
Figure 14. Metallicity gradient versus energy of Auriga GES systems in the observational window. The metallicity gradient, $\nabla[\text{Fe/H}]^*_{\text{GES}}$, is obtained using the mean orbital radius i.e. the average between the apocentre and pericentre of the star. $E_{\text{GES}}/E_{\text{in-situ}}$ is defined as the ratio between the 10th percentile of the GES stars and the 10th percentile of the *in-situ* stars in energy space. The points are also coloured by the GES total mass. In addition, we show the GES from the Jr-Lz (green), action diamond (orange), and eccentricity (yellow) selections in the observations as stars. The data point for the [Mg/Mn]-[Al/Fe] selection is outside this diagram and marked with an arrow.

the [Fe/H] gradient is opposite to the trend for GES as a function of eccentricity. That is, all other accreted stars have decreasing [Fe/H] with higher eccentricity. In addition, at the high eccentricity region, the [Fe/H] for the other accreted material are lower than the GES.

With this exploration, we find that the chemical and E-Lz selections qualitatively reproduce the trends in eccentricity, R_{GC}, and [Fe/H] as the true GES population. However, due to the larger contamination of *in-situ* stars in the E-Lz selection, the chemical selection is the better method for encompassing the nature of the GES along these axes.

5.2 The total mass of GES

So far, we have been looking at the stellar mass of GES – now we will focus on obtaining the total mass using only observational quantities. In Fig. 14, we show the metallicity gradient for the GES stars versus how deep the GES sinks into the potential for the nine haloes we looked at in Auriga. Here, we use only the stars that are within the observational window in the simulations. In addition, we coloured these circles by the total mass of GES in Auriga. The metallicity gradient, $\nabla[\text{Fe/H}]^*_{\text{GES}}$, is obtained using the mean orbital radius i.e. the average between the apocentre and pericentre of the star, which should retain more pre-merger information compared to the current R_{GC}. To quantify how deep the GES sinks into the potential, we use the *in-situ* stars as a reference point. We took the ratio of the 10th percentile of the GES stars and the 10th percentile of the *in-situ* stars in energy space, $E_{\text{GES}}/E_{\text{in-situ}}$. The relative trend is more important than the absolute values here – GES systems with lower ratios (to the left) lie at higher energies while those with larger ratios are at lower energies and are more bound.

By comparing these two observables in the simulations we already see a clear trend: systems with a steeper metallicity gradient (i.e. more negative) sink deeper into the potential (i.e. higher $E_{\text{GES}}/E_{\text{in-situ}}$). This is made even clearer when looking at the total mass of GES. Indeed, the systems that sink deeper and have steeper metallicity

gradients correspond to haloes that are more massive. This is expected as massive systems would have stars that are more bound to the progenitor, which would therefore be stripped much later when GES is deeper into the host's potential. It also makes sense that these same systems have steeper gradients as Monachesi et al. (2019) found that the Auriga stellar haloes that have steeper metallicity gradients tend to be dominated by a few significantly large mergers.

With this promising result, we can potentially use the metallicity gradient and $E_{\text{GES}}/E_{\text{in-situ}}$ of our selected GES stars in the observations to obtain a total mass of the GES progenitor. We test this out for the GES stars selected through the chemistry, eccentricity, Jr-Lz, and action diamond methods shown as the non-circular markers in Fig. 14. Because our selection function in the simulations is very simple, we are not mimicking the fact that we should observe fewer stars with lower R_{GC} (that are at larger heliocentric distances) because they are fainter. These stars are more bound and have lower energies which increase the absolute value in the denominator of $E_{\text{GES}}/E_{\text{in-situ}}$. This is *not* what we see for the parent sample in the observations (dominated by *in-situ* stars) because we are biased by the higher energies and larger Lz of the brighter disc stars. Therefore, we only look at stars with [Fe/H] < -0.8 in calculating $E_{\text{GES}}/E_{\text{in-situ}}$ to mitigate the effect of the higher energy stars from the disc. On the other hand, we note that we are potentially losing the most bound and highest metallicity stars from GES by imposing this. In addition, although we apply this cut in calculating how deep the GES sinks into the Milky Way potential, we do not apply the same cut for calculating the metallicity gradient as this would introduce a sharp and artificial boundary.

The Jr-Lz, action diamond, and eccentricity selected GES samples have similarly negative metallicity gradients as the GES in the simulations. They in fact lie on the same metallicity gradient–energy relation which is reassuring. In contrast, the chemically selected GES sample has a positive metallicity gradient. Upon inspection, we find that this is largely due to a metallicity bias in that this sample has the lowest value in [Fe/H] in the innermost mean orbital radius i.e. [Fe/H]~-1.0 compared to the other selections that reach [Fe/H]~-0.5 at the same radius. This had the net effect of inverting and making the $\nabla[\text{Fe/H}]^*_{\text{GES}}$ positive for the chemical selection. We are more likely to believe the metallicity gradient obtained from the Jr-Lz method because (1) it is less contaminated by *in-situ* material in the observations as shown in Figs 5 and 6, and (2) this is backed by the simulations that show it selects the purest GES sample (see Table 4). The action diamond selection also seems promising and gives $\nabla[\text{Fe/H}]^*_{\text{GES}}$ and $E_{\text{GES}}/E_{\text{in-situ}}$ values in between the Jr-Lz and eccentricity methods. The recent work by Khoperskov et al. (2023) also estimate a metallicity gradient of -0.14 dex kpc^{-1} with respect to the guiding radius, in agreement with our results.

From where the GES lies based on the Jr-Lz and action diamond methods, we estimate that *GES has a total mass of $10^{10.5-11.1}$ $M_\odot$* in the same range as that of Au5, Au10, Au17, Au18, and Au24. This independent estimate is lower but well within the range of halo masses obtained using the SMHM relation from Behroozi et al. (2019) in Table 1. This also agrees with previous estimates of the GES total mass using tailored N-body simulations (e.g. Naidu et al. 2021) and globular clusters (e.g. Callingham et al. 2022). This method is indeed promising because the total mass estimate bypasses any assumptions on the form of the stellar mass–metallicity relationship and the SMHM relation. In addition, it does not rely on the absolute values of GES properties but instead deals with the *relative trends* in [Fe/H] and energies. The main assumption here is that the real GES resembles the GES-like systems in the Auriga simulations.

6 CONCLUSIONS

Arguably the most significant merger that happened in our Galaxy's history, the GES event, has been studied in great detail utilizing a variety of selection methods and surveys. In this work, we try to understand if we can really pick and choose the way that we select these accreted halo stars and how this affects the inferred progenitor properties. We do this in two parts – first in the observations then in the simulations where we know the *true* GES-like population. The main takeaways from this work are as follows:

(i) In the observations, given the same APOGEE + *Gaia* parent sample, we select between 144 (from Jr-Lz) and 1279 (from eccentricity) GES stars depending on the selection method. These GES samples when projected onto the other selection diagrams show varying levels of agreement and contamination (Fig. 5).

(ii) The level of contamination is further supported by the MDFs of the GES samples: the E-Lz, eccentricity, and action diamond methods all show multimodal distributions, with a higher [Fe/H] peak corresponding to *in-situ* stars. On the other hand the [Mg/Mn]-[Al/Fe] and Jr-Lz methods have more well-behaved, normal distributions, indicating purer selections. From Gaussian mixture modelling, the GES MDF peaks somewhere between $-1.28 < $ [Fe/H] $ < -1.18$ dex (Fig. 6).

(iii) The range in [Fe/H] naturally gives a range in the $M_\star$ estimate from the MZR that are different by a factor of 2 between the different selections. However, more importantly than the selection method, the adopted MZR is the more sensitive factor that changes the estimated $M_\star$ by an order of magnitude (Table 1). Based on the [Mg/Mn]-[Al/Fe] MDF in the observations, GES has a lower limit on the $M_\star$ of 2.37×10^8 $M_\odot$ and an upper limit set by the total stellar halo mass (e.g. $\sim 1.4 \times 10^9$ $M_\odot$; Deason, Belokurov & Sanders 2019). The $M_\star$ based on the average of all five methods in the observations using the $z = 0$ relation is $\sim 4 \times 10^8$ $M_\odot$.

(iv) In the simulations, we found that the Jr-Lz method is best in selecting the purest GES sample at 33 per cent purity and the eccentricity method is best in selecting the most complete GES sample at 62 per cent completeness (Fig. 7 and Table 4).

(v) The observed $\langle$[Fe/H]$\rangle$ for GES is generally higher compared to the $\langle$[Fe/H]$\rangle$ from the total GES population due to the presence of a negative metallicity gradient in the progenitor (Figs 8 and 9).

(vi) Using the MZR from the destroyed satellites in the Auriga to estimate $M_\star$ (Fig. 11), we found that the eccentricity method tends to overestimate the most by a factor of 2.6, followed by the Jr-Lz method by a factor of 1.5. The action diamond gives the most accurate $M_\star^{est}$ while the E-Lz method underestimates the true $M_\star$ of GES by a factor of 0.7. None the less, they all are able to get close to the true $M_\star$ depending on the Auriga halo (Fig. 12).

(vii) Bringing together the observations and simulations, we found that the GES and *in-situ* stars in the simulations have distinct [Fe/H] trends in the eccentricity versus R_{GC} space i.e. the GES stars tend to have higher [Fe/H] at higher eccentricities while the opposite trend exists for the *in-situ* material. This is qualitatively reproduced by the [Mg/Mn]-[Al/Fe] and E-Lz methods in the observations (Fig. 13).

(viii) Lastly, we are able to estimate a total mass for GES ($10^{10.5-11.1}$ $M_\odot$) using only the relationship between the metallicity gradient and energy of the GES remnant, and the underlying distribution of GES total masses that drive this relation in the Auriga simulations (Fig. 14). This is very encouraging as it does not rely on the MZR and stellar mass–halo mass relation.

Although we have ventured to find the one selection method to rule them all, we have found that each method is uniquely good for estimating some progenitor properties, but not *all* across the board. Perhaps unsurprisingly, the selection method should instead be driven by the science question. In this work, we found that the chemical selection is best in estimating the $M_\star$ based on the MDF in the observations while the action diamond method is best for an $M_\star$ estimate in the simulations.[9] The Jr-Lz method gives the purest sample, while the eccentricity method is the most complete. In the future, especially with dedicated surveys aimed at observing fainter targets than APOGEE (e.g. DESI, WEAVE), we can extend our questions into the fuller picture and nature of GES as we will have access to its more distant and presumably least bound stars.

DATA AVAILABILITY

The APOGEE DR17 data used in this article are available at https://www.sdss.org/dr17. The *Gaia* data used in this article are available at https://gea.esac.esa.int/archive/. Other data used in this article can be made available upon reasonable request to the corresponding authors.

REFERENCES

Abdurro'uf et al., 2022, ApJS, 259, 35
Aguado D. S. et al., 2021, ApJ, 908, L8
Amarante J. A. S., Debattista V. P., Beraldo e Silva L., Laporte C. F. P., Deg N., 2022, ApJ, 937, 12
Amorisco N. C., 2017, MNRAS, 464, 2882
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, AJ, 161, 147
Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, MNRAS, 488, 3143
Belokurov V., Kravtsov A., 2022, MNRAS, 514, 689
Belokurov V. et al., 2006, ApJ, 642, L137
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
Bird S. A., Xue X.-X., Liu C., Shen J., Flynn C., Yang C., 2019, AJ, 157, 104
Bonaca A. et al., 2020, ApJ, 897, L18

[9] We note here again that the same elements use in the observations are not present in the simulations so we are not able to directly compare.

Bonifacio P. et al., 2021, A&A, 651, A79
Bovy J., 2015, ApJS, 216, 29
Buder S. et al., 2022, MNRAS, 510, 2407
Callingham T. M., Cautun M., Deason A. J., Frenk C. S., Grand R. J. J., Marinacci F., 2022, MNRAS, 513, 4107
Carollo D. et al., 2010, ApJ, 712, 692
Carrillo A., Hawkins K., Jofré P., de Brito Silva D., Das P., Lucey M., 2022, MNRAS, 513, 1557
Casagrande L., Wolf C., Mackey A. D., Nordlander T., Yong D., Bessell M., 2019, MNRAS, 482, 2770
Cautun M. et al., 2020, MNRAS, 494, 4291
Chiba M., Beers T. C., 2000, AJ, 119, 2843
Ciucă I. et al., 2022, preprint (arXiv:2211.01006)
Conroy C. et al., 2019, ApJ, 883, 107
Crain R. A. et al., 2015, MNRAS, 450, 1937
Das P., Hawkins K., Jofré P., 2020, MNRAS, 493, 5195
Deason A. J., Belokurov V., Koposov S. E., Lancaster L., 2018, ApJ, 862, L1
Deason A. J., Belokurov V., Sanders J. L., 2019, MNRAS, 490, 3426
Deason A. J. et al., 2021, MNRAS, 501, 5964
Deason A. J., Koposov S. E., Fattahi A., Grand R. J. J., 2023, MNRAS, 520, 6091
Fattahi A. et al., 2019, MNRAS, 484, 4471
Fattahi A. et al., 2020, MNRAS, 497, 4459
Feltzing S., Feuillet D., 2023, ApJ, 953, 143
Fernández-Alvar E. et al., 2018, ApJ, 852, 50
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
Freeman K., Bland-Hawthorn J., 2002, ARA&A, 40, 487
Gaia Collaboration, 2018, A&A, 616, A1
Gaia Collaboration, 2021, A&A, 649, A1
Gaia Collaboration, 2023a, A&A, 674, A1
Gaia Collaboration, 2023b, A&A, 674, A38
Gallart C., Bernard E. J., Brook C. B., Ruiz-Lara T., Cassisi S., Hill V., Monelli M., 2019, Nat. Astron., 3, 932
Gilmore G., Wyse R. F. G., Norris J. E., 2002, ApJ, 574, L39
Gómez F. A., Besla G., Carpintero D. D., Villalobos Á., O'Shea B. W., Bell E. F., 2015, ApJ, 802, 128
Grand R. J. J. et al., 2017, MNRAS, 467, 179
Grand R. J. J. et al., 2020, MNRAS, 497, 1603
Grand R. J. J. et al., 2021, MNRAS, 507, 4953
Han J. J. et al., 2022, AJ, 164, 249
Hasselquist S. et al., 2021, ApJ, 923, 172
Hawkins K., Jofré P., Masseron T., Gilmore G., 2015, MNRAS, 453, 758
Hawkins K. et al., 2016, A&A, 592, A70
Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, Nature, 563, 85
Horta D. et al., 2021, MNRAS, 500, 1385
Horta D. et al., 2023, MNRAS, 520, 5671
Ibata R. A., Gilmore G., Irwin M. J., 1994, Nature, 370, 194

Khoperskov S., Minchev I., Steinmetz M., Marabotto J., Kordopatis G., Delgado Gomez J., Libeskind N., 2023, preprint (arXiv:2310.05287)
Kirby E. N., Cohen J. G., Guhathakurta P., Cheng L., Bullock J. S., Gallazzi A., 2013, ApJ, 779, 102
Kordopatis G. et al., 2013, MNRAS, 436, 3231
Kruijssen J. M. D. et al., 2020, MNRAS, 498, 2472
Lane J. M. M., Bovy J., Mackereth J. T., 2022, MNRAS, 510, 5119
Lane J. M. M., Bovy J., Mackereth J. T., 2023, MNRAS, 526, 1209
Laporte C. F. P., Gómez F. A., Besla G., Johnston K. V., Garavito-Camargo N., 2018, MNRAS, 473, 1218
Limberg G., Souza S. O., Pérez-Villegas A., Rossi S., Perottoni H. D., Santucci R. M., 2022, ApJ, 935, 109
Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, MNRAS, 456, 2140
Mackereth J. T., Bovy J., 2020, MNRAS, 492, 3631
Mackereth J. T. et al., 2019, MNRAS, 482, 3426
Majewski S. R. et al., 2017, AJ, 154, 94
Mannucci F. et al., 2009, MNRAS, 398, 1915
Matsuno T., Hirai Y., Tarumi Y., Hotokezaka K., Tanaka M., Helmi A., 2021, A&A, 650, A110
McMillan P. J., 2017, MNRAS, 465, 76
Meza A., Navarro J. F., Abadi M. G., Steinmetz M., 2005, MNRAS, 359, 93
Monachesi A. et al., 2019, MNRAS, 485, 2589
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., 2021, ApJ, 923, 92
Naidu R. P. et al., 2022, preprint (arXiv:2204.09057)
Nissen P. E., Schuster W. J., 1997, A&A, 326, 751
Nissen P. E., Schuster W. J., 2010, A&A, 511, L10
Orkney M. D. A. et al., 2022, MNRAS, 517, L138
Orkney M. D. A. et al., 2023, MNRAS, 525, 683
Pinna F., Walo-Martín D., Grand R. J. J., 2023, preprint (arXiv:2301.10666)
Planck Collaboration XXII, 2014, A&A, 571, A11
Rix H.-W. et al., 2022, ApJ, 941, 45
Schaye J. et al., 2015, MNRAS, 446, 521
Shipp N. et al., 2018, ApJ, 862, 114
Springel V., 2010, MNRAS, 401, 791
Springel V., Yoshida N., White S. D. M., 2001, New Astron., 6, 79
Tremonti C. A. et al., 2004, ApJ, 613, 898
Tronrud T., Tissera P. B., Gómez F. A., Grand R. J. J., Pakmor R., Marinacci F., Simpson C. M., 2022, MNRAS, 515, 3818
Vasiliev E., 2019, MNRAS, 482, 1525
Vasiliev E., Belokurov V., Evans N. W., 2022, ApJ, 926, 203
Vincenzo F., Spitoni E., Calura F., Matteucci F., Silva Aguirre V., Miglio A., Cescutti G., 2019, MNRAS, 487, L47

This paper has been typeset from a TEX/LATEX file prepared by the author.