

# Robust clustering of the local Milky Way stellar kinematic substructures with *Gaia* eDR3

Xiaowei Ou<sup>1</sup> <sup>1</sup>★ Lina Necib<sup>1,2</sup> and Anna Frebel<sup>1</sup>

<sup>1</sup>Physics Department and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA 02139, USA

<sup>2</sup>The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA 02139, USA

Accepted 2023 February 27. Received 2023 February 7; in original form 2022 October 3

## ABSTRACT

Understanding local stellar kinematic substructures in the solar neighbourhood helps build a complete picture of the formation of the Milky Way, as well as an empirical phase space distribution of dark matter that would inform detection experiments. We apply the clustering algorithm HDBSCAN on the *Gaia* early third data release to identify a list of stable clusters in velocity space and action-angle space by taking into account the measurement uncertainties and studying the stability of the clustering results. We find 1405 (497) stars in 23 (6) robust clusters in velocity space (action-angle space) that are consistently not associated with noise. We discuss the kinematic properties of these structures and study whether many of the small clusters belong to a similar larger cluster based on their chemical abundances. They are attributed to the known structures: the Gaia Sausage-Enceladus, the Helmi Stream, and globular cluster NGC 3201 are found in both spaces, while NGC 104 and the thick disc (Sequoia) are identified in velocity space (action-angle space). Although we do not identify any new structures, we find that the HDBSCAN member selection of already known structures is unstable to input kinematics of the stars when resampled within their uncertainties. We therefore present the stable subset of local kinematic structures, which are consistently identified by the clustering algorithm, and emphasize the need to take into account error propagation during both the manual and automated identification of stellar structures, both for existing ones as well as future discoveries.

**Key words:** methods: data analysis – Galaxy: halo – Galaxy: kinematics and dynamics – Galaxy: structure.

## 1 INTRODUCTION

In the most popular cosmological structure formation model, typical spiral galaxies, like our Milky Way, form hierarchically through many mergers with less massive galaxies (see e.g. White & Rees 1978; Frenk & White 2012; Wechsler & Tinker 2018; Behroozi et al. 2019). Theories and simulations predict that most of the merging stellar populations are tidally disrupted during infall and form various structures in today’s Milky Way stellar halo (Zolotov et al. 2009; Cooper et al. 2010). These accreted stars appear in large all-sky surveys as spatially coherent kinematic structures (‘streams’) (Majewski et al. 2003; Shipp et al. 2018; Yuan et al. 2019; Malhan et al. 2021; Li et al. 2022), as well as mere overdensities in orbital energy and angular momenta after phase mixing [‘tidal debris’; Koppelman et al. 2019b; Naidu et al. 2020; Malhan 2022; for a review see e.g. Helmi (2020)]. Studying these accreted stars allows us to estimate properties of their progenitors (e.g. mass, age, infall time, and kinematics). The stellar chemical signatures of such structures, such as metallicity and  $\alpha$ -element abundances (Freeman & Bland-Hawthorn 2002; Frebel & Norris 2015; Maiolino & Mannucci 2019; Sestito et al. 2019; Fattahi et al. 2020), and kinematic properties (Johnston, Hernquist & Bolte 1996; Johnston 1998; Helmi et al.

1999; Bullock & Johnston 2005; Johnston et al. 2008) thus provide unique information about the formation history of the Milky Way.

Understanding the merger history of the Milky Way and the properties of the accreted and merged progenitors galaxies can help us learn about near-field cosmology and provide constraints on structure formation and galaxy evolution in the late-time universe [see e.g. Mackereth et al. (2019); Myeong et al. (2019); Belokurov et al. (2020); Naidu et al. (2020); Feuillet et al. (2022); Malhan (2022)]. At the same time, Herzog-Arbeitman et al. (2018a), Herzog-Arbeitman, Lisanti & Necib (2018b), and Necib, Lisanti & Belokurov (2019a) found that the stellar component of the accreted systems may be used as a tracer of the accompanying dark matter (DM) accreted along with the system (O’Hare et al. 2020). The DM distribution in the Milky Way, especially near the solar neighbourhood, is one of the key ingredients towards uncovering the particle nature of DM. Direct detection rates of DM particles rely on the proper modelling of the local DM density and velocity distributions, which translates discoveries as well as null results into a DM mass and DM-nucleon scattering cross-section (Goodman & Witten 1985; Drukier, Freese & Spergel 1986; Jungman, Kamionkowski & Griest 1996).

Using accreted stars to answer near-field cosmology questions, as DM tracers relies on efficiently and accurately identifying them in the local solar neighbourhood. Since our Solar system resides in the thin disc of the Milky Way, the local stellar population is dominated by *in situ* stars formed in the thin disc (Soubiran, Bienaymé & Siebert

\* E-mail: [xwou@mit.edu](mailto:xwou@mit.edu)

2003). Thus, while earlier discoveries and studies [see e.g. Ibata, Gilmore & Irwin (1994); Belokurov et al. (2006); Schlaufman et al. (2009)] were carried out with chemical abundance measurements and small kinematic data sets, identifying some of the more elusive accreted stellar populations, such as Thamnos and the metal-weak Atari disc, on top of this *in situ* background relies on obtaining 6D kinematic information for a large data set of local stars. Large kinematics data sets are powerful in identifying intact satellites that are in the early stage of accretion, diffuse satellites that have phase-mixed into the environment, as well as those that are in the process of transitioning from the former to the latter via tidal disruption. This can now be done with large photometric and astrometric surveys such as the *Gaia* mission (Gaia Collaboration 2016, 2018, 2021; Yuan et al. 2019; Ibata et al. 2020; Malhan 2022).

Many studies have already used these data to identify numerous substructures that merged into the Milky Way in the past (Belokurov et al. 2018; Helmi et al. 2018; Myeong et al. 2018; Necib et al. 2019a; Koppelman et al. 2019b; Helmi 2020). Among these local structures, the Gaia-Enceladus/Sausage (GSE) is the most recent major merger event experienced by the Milky Way (Belokurov et al. 2018; Helmi et al. 2018). Additionally, Koppelman et al. (2019b) discovered the Thamnos structure in the *Gaia* second data release (DR2) as part of the local retrograde halo. Sequoia, a high energy retrograde structure originally attributed to being part of the GSE by Helmi et al. (2018), was found to be most likely coming from a separate accretion event in Myeong et al. (2019) with *Gaia* DR2 data. Sharing similar kinematic properties, Arjuna and I'itoi were discovered in Naidu et al. (2020) as distinct accretion debris given their metallicity differences after combining data from the *Gaia* DR2 and the H3 survey (Gaia Collaboration 2018; Conroy et al. 2019). Wukong and Aleph, two prograde structures, were found similarly by Naidu et al. (2020); Aleph is a highly prograde structure with circular orbits and a significant vertical motion, as well as a metal-rich and relatively alpha-poor chemistry. Wukong is also a prograde structure lining the margin of the GSE in the  $E_{\text{tot}}$  versus  $L_z$  plane, shown as two overdensities at  $E_z \sim -1.1$  and  $-1.3 \times 10^5 \text{ km}^2 \text{ s}^{-2}$ . Malhan (2022) discovered Pontus, a merger event that happened  $>5\text{--}6$  Gyr ago and brought in seven globular clusters, by studying groupings of globular clusters in action space. The remnant of this merger is found to have low energy and slightly retrograde motion. Spatially coherent streams such as S1, Icarus, Sagittarius, Centus, and Nyx, to name a few, are also identified or studied with greater details with new members identified (Ibata et al. 1994; Newberg, Yanny & Willett 2009; Myeong et al. 2018; Meingast, Alves & Fürnkranz 2019; Yuan et al. 2019; Ibata et al. 2020; Re Fiorentin et al. 2021). They are associated to either past merger events or disc clusters.

Past discoveries of local over-densities, such as the Helmi stream (Helmi & White 1999), have also been further studied with *Gaia* data (Koppelman et al. 2019a). The long known ‘metal weak thick disc’ (Norris, Bessell & Pickles 1985; Morrison 1990; Chiba & Beers 2000) has recently been postulated to be likely the result of a merger, rather than part of the canonical thick disc. Mardini et al. (2022) confirmed such origin with photometric metallicities from the SkyMapper survey (Keller et al. 2007; Wolf et al. 2018; Chiti et al. 2021) and kinematics from *Gaia*. Using machine learning techniques trained on simulations (Wetzel et al. 2016; Hopkins et al. 2018; Sanderson et al. 2020) and labelling the *Gaia* DR2 data set, Ostdiek et al. (2020) built an accreted star catalogue (Ostdiek et al. 2019) from which multiple structures were identified (Necib et al. 2020b), including Nyx (Necib et al. 2020a), a stream within the thick disc with a highly eccentric orbit.

These studies use the fact that accreted substructures emerge as overdensities in the 6D phase space. The key, thus, is extracting such overdensities effectively and robustly from the data. Current studies mostly rely on two main methods in identifying overdensities and selecting their constituents: manual selection (Helmi et al. 2018; Lee, Beers & Kim 2019; Naidu et al. 2020; Carollo & Chiba 2021) and clustering algorithms (Helmi et al. 2017; Yang et al. 2019; Yuan et al. 2019; Necib et al. 2019a; Necib et al. 2019b; Malhan 2022; Sofie Lövdal et al. 2022; Wu et al. 2022), or some combination of both [Koppelman et al. 2019b; Bird et al. 2021; see Buder et al. (2022) Table A.1 for a review of techniques].

For manual selection, overdensities in phase space are identified based on physical intuition gained from past discoveries (Naidu et al. 2020) or statistical comparisons with some reference random data set (Helmi et al. 2018). They are then characterized based on their phase space properties. Potential background contamination is removed by examining stellar population properties, such as age and metallicity distributions. This method relies on the fact that an overdensity is located in phase space where the background contamination level is low, either because the substructures themselves have peculiar properties, which set them apart from the majority of the *in situ* Milky Way stars, or because astronomers are able to apply well-informed cuts on the data set to remove most of the background. For structures that reside within or close to the Milky Way background, manual selection becomes difficult and tricky.

For machine learning-focused methods, structures are picked out directly from the data sample with minimum pre-processing by various algorithms. The results are, in general, more consistent between works, for example, especially when using the same algorithm. Naturally, this is not frequently true for manual selections. In most cases, however, the final results need to be further examined, combined with additional less complete information such as metallicity and chemical abundances, to remove unphysical structures such as background noise and data artefacts. Brauer et al. (2022) tested kinematic-based clustering algorithms on simulations and showed that they generally lack efficiency and accuracy in identifying structures resulting from now-dissolved ultra faint dwarf galaxies.

Despite the difficulties mentioned above, both methods have yielded many new structures in the solar neighbourhood that have been extensively studied and validated. Yet, it is still often the case that the stellar membership of any of these structures has to be confirmed with archive/follow-up spectroscopic observations if they were initially discovered based on kinematic information only. Even with the best informed physical expectation and/or the most sophisticated algorithms, the kinematic selection alone is typically not enough to robustly determine the constituent stars, even when the overdensities/structures are principally well-established. This is due not only to the background-dominated signal search (i.e. the *in situ* population) but also, and more importantly, due to our inability to measure the stellar kinematics infinitely precisely. Despite the recent advancement in astrometric survey precision, uncertainties still play a large role that affects stellar membership assignments. For practical reasons, uncertainties are often left out in kinematic structure studies, either because the machine learning algorithm inherently does not support uncertainties or because the process becomes too computationally intensive. This lack continues to hamper the robustness of identifying additional structures to eventually identify all progenitors of the Milky Way.

In this paper, we aim to incorporate uncertainties for the first time into an unsupervised machine learning algorithm to search for Milky Way substructures in *Gaia*. The goal is to select a robust sample of star members present in clusters/overdensities purely based on kinematic

inputs. We use the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN; McInnes, Healy & Astels 2017), an unsupervised learning algorithm, to identify clusters in the 6D phase space with data from the *Gaia* eDR3 data set (Gaia Collaboration 2021). We combine Monte Carlo resampling and Jaccard coefficients to estimate the robustness of the HDBSCAN clusters. Doing so currently leads us to identify only the most robust clusters and their member stars. We recover the GSE, Sequoia, the Helmi Stream, and globular clusters NGC 3201 and NGC 104, but do not find Thamnos and Nyx due to an initial cut on the maximum orbital vertical distance  $Z_{\max}$ , or Arjuna and I'itoi, which could be a subset of Sequoia in our analysis, or Aleph and Wukong that we expect to be either unstable, or dominant further away from the disc than our sample. We also do not identify structures consistent with the Sagittarius stream as it is dominant further away and thus have fewer radial velocity measurements from *Gaia* DR2, which is required for the star to enter the sample in the first place. For similar reason, spatially coherent streams in the outer halo, such as Cetus, would also not be retrieved by our method. Substructures originally discovered using groupings of globular clusters, such as Pontus and Kraken, are also not recovered in this work. Our results demonstrate the importance of incorporating uncertainties in machine learning, especially unsupervised clustering, for astronomy studies, and how this approach affects our interpretation of the clustering results and the constituent stars within the clusters.

This paper is structured as follows: We discuss the data sample used for this study in Section 2. The clustering algorithm is discussed in Section 3. We present the clustering results in Section 4 and discuss their interpretations in Section 5.

## 2 DATA SET

### 2.1 Quality cuts

We first apply quality cuts on *Gaia* eDR3 stars, combined with the radial velocity measurements from *Gaia* DR2. Based on recommendations from Lindegren et al. (2021a), we select sources with good astrometric solutions by requiring `parallax_over_error > 5`, `ruwe < 1.4`, `astrometric_excess_noise < 2`, `duplicated_source = 0`,  $G < 19$ , and `visibility_periods_used >= 10. We also perform the cut on colour  $C = \text{phot\_bp\_rp\_excess\_factor}$  with the prescription described by Riello et al. (2021) to remove faint sources in close proximity to bright sources. We then correct for the zero point bias in the parallaxes as described in Lindegren et al. (2021b).`

The data sample size, after applying the quality cuts mentioned above, and removing high-velocity stars ( $|v_i| > 1000 \text{ km s}^{-1}$  where  $i = r, z, \phi$ ), is 5557 670. More cuts are applied to the sample before going through the clustering algorithm, which are described in the following sections.

### 2.2 Cross-matched data sets

For extra chemical information, we cross-match the sample with the spectroscopic surveys LAMOST DR6 (632 223 matches; Cui et al. 2012; Zhao et al. 2012) and APOGEE DR17 (211 798 matches; Majewski et al. 2017; Abdurro'uf et al. 2022). We focus mainly on the metallicity measurements from these two surveys, as LAMOST provides the most matching sources and APOGEE provides the most  $\alpha$ -abundances.

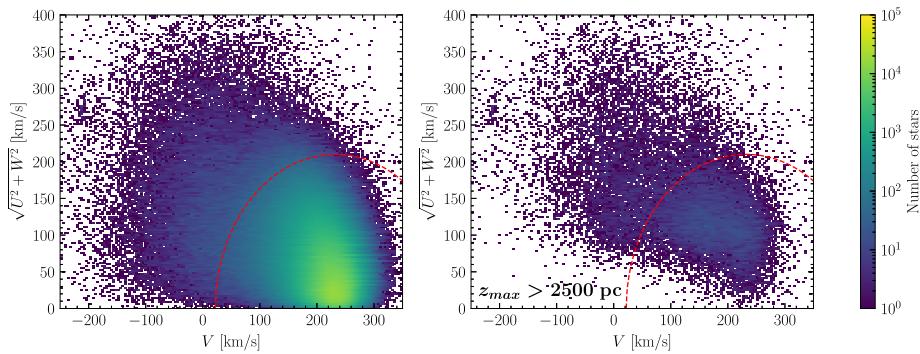
Chemical abundance information is used for the characterization of individual clusters, comparison between clusters, and comparison with literature. To avoid systematic differences across data sets, we limit metallicity characterizations of our identified clusters to one survey at any given time. Unless otherwise noted, we use APOGEE metallicities and  $\alpha$  abundances in summary plots for characterizing individual clusters as APOGEE chemical abundances are derived from high-resolution spectra than those from LAMOST (medium-resolution spectra only). However, for examining relationships between the final clusters in this study and for a comparison with known structures' metallicities from the literature, we instead use metallicities from LAMOST that are available for more of our cluster stars and thus provide better statistics. We discuss more details in Section 5.

### 2.3 $Z_{\max}$ cuts

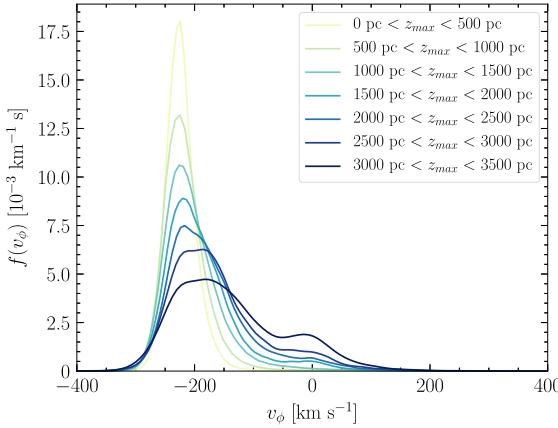
It is extremely difficult, if not impossible, for most clustering algorithms to identify substructure signals over a strong background. The majority of our sample is made up of *in situ* thin disc stars, which, if not removed in some way, will flood all signals and prevent detecting any meaningful clustering. To remove the thin-disc contamination, a standard practice in the field has been to perform a velocity cut in the local standard of rest (LSR) velocity [see e.g. Helmi et al. (2018); Sofie Lövdal et al. (2022)]. Where spectral information is available, velocity cuts are also combined with metallicity cuts to exclude metal-rich disc stars [see e.g. Myeong et al. (2018)]. This method is the most successful in completely removing stars with thin disc-like kinematics, thus leaving behind a sample comprised primarily of halo stars and some hot thick disc stars. However, the disadvantages of this method are that it might bias the overall velocity distribution of the resulting sample, or drop some of the prograde structures [e.g. Nyx (Necib et al. 2020a)]. Since our goal is to identify velocity structures and eventually their impact on the local DM velocity distributions in future work, we here opt for the least kinematically biased sample possible by instead placing a  $Z_{\max}$  cut rather than using a direct velocity cut.

Specifically, we apply a cut at  $Z_{\max} > 2.5 \text{ kpc}$  to remove the thin disc and to keep the possibility of studying potential prograde structures with high vertical motion. The sample size after the cut is 27 921 (out of the initial 5557 670 stars). The left-hand panel of Fig. 1 shows the Toomre plot of our initial sample, where stars in the thin disc ( $V \sim 220 \text{ km s}^{-1}$ ) dominate by orders of magnitude over the rest of the stars, which would also include any prograde kinematic substructures.

Our  $Z_{\max}$  is calculated through orbit integration, as detailed in Section 2.5. We expect most *in situ* stars to have  $Z_{\max}$  no higher than 2.5 kpc, especially with a thin disc scale height approximately at 300 pc (Jurić et al. 2008). Thus, this cut will serve the purpose of removing the thin disc, similar to the LSR velocity cut, but at the same time, it will allow potential prograde structures with high vertical motion to remain in the final sample. As shown in Fig. 2, the  $v_\phi$  distribution of stars at higher  $Z_{\max}$  interval shows a clear transition from the expected rotational velocity of the disc of  $-220 \text{ km s}^{-1}$  towards  $0 \text{ km s}^{-1}$ , the average rotational velocity of the stellar halo. Fig. 1 also indicates a significant portion of the stars that would have been removed by an LSR velocity cut (red dashed line) remaining in the post-cut sample. Our  $Z_{\max}$  cut is not as clean as a LSR velocity cut, by design, as even at vertical distances a few times above the scale height of the thin disc, the exponential tail of the thin disc is still comparable in size to the slower rotating prograde halo, and retrograde halo populations. Jurić et al. (2008) measured the



**Figure 1.** Toomre plots for the full sample (left) and the subsample with  $Z_{\text{max}}$  cut at 2500 pc, as indicated at the bottom left of each panel. The red dashed line indicates a probable LSR velocity cut at 210 km s<sup>-1</sup>, with  $V_{\text{LSR}} = 232 \text{ km s}^{-1}$ .



**Figure 2.** Normalized distribution of  $v_\phi$  for increasing  $Z_{\text{max}}$  intervals, including only stars with available *Gaia* 2 RV. The  $v_\phi$  histogram was constructed from bins between 0 and 3.5 kpc, in 500 pc intervals. There is a gradual shift in the overall distribution away from the LRS at  $\sim -220 \text{ km s}^{-1}$  towards 0 km s<sup>-1</sup>, as the sample evolves from thin-disc dominated to halo dominated.

local halo-to-thin disc stellar number density ratio of 0.5 per cent. In addition to that, the uncertainty in our distance measurements unavoidably creates a systematically puffed-up thin disc, causing a significant amount of prograde stars with kinematics very similar to that of the thin disc even at  $Z_{\text{max}}$  greater than 3 kpc. Yet, doing so gives us the opportunity to study potential overdensities in the high vertical motion tail of the thin disc.

More specifically, in terms of potential velocity biases, while this cut still poses an implicit bias on the distribution of the velocity in the  $z$  direction, it is nevertheless much less restrictive than the LSR cut. Again, this leaves space for prograde accreted stars to be part of the clustering sample. Following this reasoning, we require  $Z_{\text{max}} > 2.5 \text{ kpc}$ , as well as that the mean value of  $Z_{\text{max}}$  has to be larger than two standard deviations above the cut when taking into account the errors in the measurements.

## 2.4 Coordinate transformations

After sample selection, we compute the galactocentric coordinates and velocities assuming the ASTROPY galactocentric frame with default parameters adopted in ASTROPY v4.0 (Astropy Collaboration 2013, 2018). The galactic centre coordinate is at RA = 266.4051° and Dec = −28.936175° (Reid & Brunthaler 2004). The right-handed coordinate system is defined such that the  $x$ -axis points from the Sun

to the galactic centre and the  $y$ -axis points in the direction of local disc rotation. The distance to the galactic centre is assumed to be 8.122 kpc (Gravity Collaboration 2018), with the Sun moving with velocity (12.9, 245.6, and 7.78) km s<sup>-1</sup> and located 20.8 pc above the mid-plane (Drimmel & Poggio 2018; Bennett & Bovy 2019).

Uncertainties on the galactocentric velocities and positions are propagated from uncertainties in parallax, proper motion, and radial velocities. We take into account covariances between *Gaia* measured parallax and proper motions. As a result, there exist non-trivial covariances between the transformed galactocentric velocities, which we will take into account in the orbit integration. Radial velocities are measured from *Gaia* DR2 and have no provided covariances with other *Gaia* EDR3 astrometry measurements.

## 2.5 Orbit integration

We integrate the orbits for our final sample backwards in time for 2.5 Gyr with 1 Myr steps using the Python AGAMA package (Vasiliev 2019). We set up the potential to be the default MilkyWayPotential from GALA v1.4.1, which is a four-component potential model consisting of a bulge and nucleus with Hernquist profile (Hernquist 1990), a disc with Miyamoto–Nagai profile Miyamoto & Nagai (1975), and a halo following the NFW profile Navarro, Frenk & White (1997). The disc model is taken from Bovy (2015) and the rest are fixed by fitting to a compilation of mass measurements of the Milky Way. More details on the parameters can be found in the GALA documentation (Price-Whelan 2017). Orbital parameters including energy, angular momenta, eccentricity, maximum galactic  $Z_{\text{max}}$ , apogalacticon, perigalacticon, and actions are derived from the integrated orbits.

For each star, we estimate the uncertainties for the orbital parameters by resampling the initial positions and velocities 100 times with their respective uncertainties (assuming the uncertainties are inherently Gaussian) and including the covariances between the velocities. Actions are calculated using the ACTIONFINDER routine provided by AGAMA. We note that while their effect is minimal, we include the covariance between the positions and velocities to compute the uncertainties in the orbital parameters. We then record the mean and standard deviation of these parameters as the final values and corresponding uncertainties.

## 3 CLUSTERING ALGORITHM

The main goal of this study is to identify kinematic clusters, from machine learning algorithms, which are stable under resampling within the uncertainties of the stellar kinematics inputs. Our method

thus focuses on ensuring that the members of the extracted clusters are stable against being assigned as noise or as parts of different clusters as can occur when uncertainties are not taken into account. Consequently, stars we assign as cluster members are very stable against any splitting/merging of their apparent host clusters; splitting refers to a cluster breaking into smaller clusters, and merging refers to multiple clusters merging into one single cluster.

Given the complexity of this section, we summarize the main points here. We apply the clustering algorithm HDBSCAN (defined with its parameters in Section 3.1) on two separate spaces: The 3D cylindrical velocity space, and the 3D action space with total energy as an additional axis (see Section 3.2). In order to incorporate the uncertainties, we first define a high-stability sample by only selecting the stars that get clustered more than  $N_{\text{stack}}^{\text{cut}}$  times (defined in Section 3.4) when applying HDBSCAN over 100 realizations, as shown in Section 3.3. We then cluster the high-stability stars, and define the Jaccard coefficient as a measure of the stability of clusters (Section 3.5).

### 3.1 Choice of algorithm

We adopt HDBSCAN<sup>1</sup> as a clustering algorithm for identifying the kinematic substructures in the *Gaia* data (McInnes et al. 2017). The choice of the clustering algorithm depends on the particularity of the data set. Specifically, we do not know a prior total number of clusters; the clusters could be of various shapes, and the data inherently have stars that are not part of a cluster and would therefore need to be treated as noise. Ideally, the required algorithm would address the specifics of the data above, and potentially, it would incorporate measurement uncertainties. Given the difficulty in finding an algorithm that has a standard treatment of measurement errors, a stability study of the resulting clusters is required, as we will discuss in Section 3.5.

HDBSCAN stands out to be the most ideal choice for our study. The algorithm transforms the space according to the density of points and performs single linkage clustering on the transformed space. HDBSCAN assumes dense regions as clusters and thus makes no assumption about the intrinsic distribution or total number of clusters. It is thus capable of finding clusters of generic shape. It also does not require every data point to be assigned to a cluster, allowing the presence of noise. It builds a merger tree, where one can cut at variable heights to decide the smallest sized cluster based on a few hyperparameters, mostly `min_cluster_size`, which determines whether a split of the tree should be treated as two new clusters or one cluster losing noise points. The resulting clusters can thus have varying densities as long as they are stable against further splitting. Additionally, the clusters found are deterministic with the same input data and hyperparameters setting. Brauer et al. (2022) examined the performance of a suite of clustering algorithms on the Caterpillar cosmological simulation (Griffen et al. 2016) while trying to recover substructures left behind by accreted ultra-faint dwarf galaxies. Their result demonstrates that HDBSCAN has the best performance among all algorithms, albeit not powerful enough to identify all subtle, long-accreted clusters.

The parameters of the algorithm are set to default with the exception of `min_cluster_size`, which is set to 20 to capture the smallest substructure, and `cluster_selection_method`, which is set to `leaf` to track the leaf nodes of the clustering tree. For the `cluster_selection_method`, as described in

Section 3.5, the merging of these smaller homogeneous clusters along the hierarchical tree can be altered significantly due to uncertainties. Instead of having HDBSCAN decide how the leaf nodes should be merged, it is more consistent to first examine the stability of the leaf nodes individually (and how they change with uncertainties folded in), as shown in Section 3.5. We then considered separately if and how they may be merged, as shown in Section 4.3. Further, as a result of the `leaf` setting, changing the `min_cluster_size` essentially determines the size of the smallest cluster HDBSCAN returns. Our test shows that the final stable clusters reported in Section 4 do not alter significantly with `min_cluster_size` set ranging from 15 to 25.

### 3.2 Clustering spaces

In this work, we apply HDBSCAN to two separate spaces: The 3D cylindrical velocity space and the 3D action space with total energy as an additional axis. First, the cylindrical velocity clustering space is chosen as it will identify stellar kinematic substructures, which might correlate with DM velocity substructures. As shown in Necib et al. (2019b), the stellar velocity distribution of debris flows would be correlated with the velocity distribution of DM, which is key in studying DM direct detection rates.

Secondly, we also perform clustering in the integrals of motion space for accreted structures that have phase-mixed since their accretion. In particular, we perform clustering on the total energy and the 3D action space. Previous works have discussed the fact that accreted stellar populations remain clustered in such integral of motion space long after the merger, given a slowly varying potential, from both theory (Binney & Spergel 1982) and simulation (Helmi et al. 1999; Knebe et al. 2005; Gómez et al. 2010) perspectives. While these structures may no longer show a coherent velocity distribution, they are none the less important sources of accreted DM and would affect the local DM velocity distribution. With the 3D action fully describing the orbit, the total energy of the orbit is principally a redundant parameter. Our test shows that the clustering results generally perform better with the redundant energy parameter. It is also widely used in previous literature using HDBSCAN or other clustering algorithms [as have been used in Yuan et al. (2019); Koppelman et al. (2019a); Sofie Lövdal et al. (2022)] so for comparison purposes it is included.

For the reasons mentioned above, we do not expect a one-to-one correspondence between the clusters identified in one space and those from the other. None the less, we compare the clustering results from these two spaces in Section 4 as these results complement each other. Moreover, they allow us to study, in future work, the local DM velocity distribution from both coherent and phase-mixed accreted structures.

### 3.3 Uncertainties: random realizations and stacking

HDBSCAN, similar to most clustering algorithms, is not designed for handling data with uncertainties. Previous work (see e.g. Koppelman et al. 2019b) that had used HDBSCAN for cluster identification did not take uncertainties into account when applying the algorithm to kinematic data. Other works have included measurement uncertainties on a much smaller scale, typically on the order of thousands of stars; for example, Gudin et al. (2021), Limberg et al. (2021), and Shank et al. (2022) used HDBSCAN to identify dynamically tagged groups for very metal-poor stars. Limberg et al. (2021) proposed estimating the confidence level of the clusters via feeding Monte Carlo generated perturbed data sets into the HDBSCAN

<sup>1</sup><https://hdbscan.readthedocs.io/>

hierarchical tree generated from the nominal values. The cluster membership assignment is re-evaluated using the prediction method provided by the HDBSCAN package.<sup>2</sup> As pointed out by the authors of HDBSCAN, the algorithm is a transductive method, meaning new/perturbed data points can and should alter the underlying clustering (McInnes et al. 2017). This is especially the case if one attempts to predict the membership assignment of an entire perturbed data set with a size identical to the original nominal data set used to generate the tree. The membership assignments for the perturbed data sets are thus inherently inconsistent with the hierarchical tree.

In this work, we adopt a self-consistent method of simultaneously evaluating the robustness of the clusters themselves and the membership of the stars within the clusters by regenerating the hierarchical clustering tree for each Monte Carlo resampled data set. Specifically, uncertainties are incorporated by repeating the clustering process with input parameters resampled from their respective uncertainties for 100 times. We repeat the clustering procedure with 100 realizations, sampled from the error bars on each of the variables included. The total number of resampled observed phase-space measurements used to estimate orbital parameters uncertainties, described in Section 2.5, is the same as the total number of clustering realizations. We note that, however, we do not directly sample from the 100 calculated orbital parameters for each clustering realization. Inputs to the clustering algorithms are sampled 100 times only using the mean and standard deviation recorded in Section 2.5, which do not contain information on the correlations and covariance of the observed phase-space uncertainties. Our test on a subset of stars shows that the distribution of all orbital parameters (i.e. energy and actions) used for clustering are consistent with a symmetric-Gaussian described by the mean and standard deviation of the 100 calculated orbital parameters from the randomly sampled observed measurements. It would be, none the less, more consistent to sample from the 100 calculated orbital parameters during clustering to account for correlations between parameters. Doing so, however, requires either saving and keeping track of the 100 orbital measurements for four parameters per star, which amounts to several hundred TB of data considering the size of the sample, or calculating the orbital parameters for all stars from sampled phase-space measurements on-the-fly during clustering, which becomes computationally expensive for 100 realizations of HDBSCAN. The computational constraint, together with the goal of studying the effect of uncertainties in the direct input data in HDBSCAN clustering, pushes us to assume no covariance and correlation between orbital parameters.

The clustering results from each random realization are then combined together in order to establish the stability of the clusters. We find that, while the general clustering results may appear similar by eye in various spaces, the exact stellar membership could differ significantly between realizations. Clusters that are in close proximity to other clusters often exchange members due to the existence of noise and background. Such behaviour results in the clusters themselves shifting around in the clustering space and sometimes merging and splitting between different realizations. It is thus extremely difficult to directly quantify the stability of the clusters.

<sup>2</sup>This procedure provides estimates on the probability of a given star being the member of a cluster, considering uncertainties on the kinematic properties, which indirectly evaluates the robustness of the clusters themselves. All probabilities are based on the hierarchical tree generated by HDBSCAN from the nominal (or mean) kinematic properties of the sample stars.

To overcome this challenge, instead of directly evaluating the stability of individual clusters, we first focus our attention on the stars themselves. We stack the clustering results from all 100 realizations, and instead of recording the membership of the stars to specific clusters (or noise), we simply record whether or not the stars are identified to be part of some cluster at all. The result is a distribution of the number of times a given star is clustered to be non-noise. Stars that are consistently clustered can then be selected by making a cut on the number of times the star has been clustered; we call this variable  $N_{\text{stack}}$ . An  $N_{\text{stack}}^{\text{cut}}$  allows us to identify the stable stars first that are more often part of a cluster and not the noise, which can then be used to find the stable clusters. We call the sample of stars after this cut the high-stability sample.

The sample size of the high-stability sample is significantly smaller than the full sample. It does, however, guarantee that the clusters found in this sample are much more robust with respect to any resampling of the kinematic inputs within their uncertainties. The fact that these stars are associated with some cluster in multiple random realizations signifies that these stars stay close to overdensities in the parameter space. More importantly, the overdensities themselves remain significant above the background such that HDBSCAN would identify them as clusters. We can think of this as deepening the contrast of the clustering space. We show these high-stability sample stars (in red) as well as the entire sample (in black) in Fig. 3. Some structures are visible by eye in this high-stability sample.

### 3.4 Selection of $N_{\text{stack}}^{\text{cut}}$

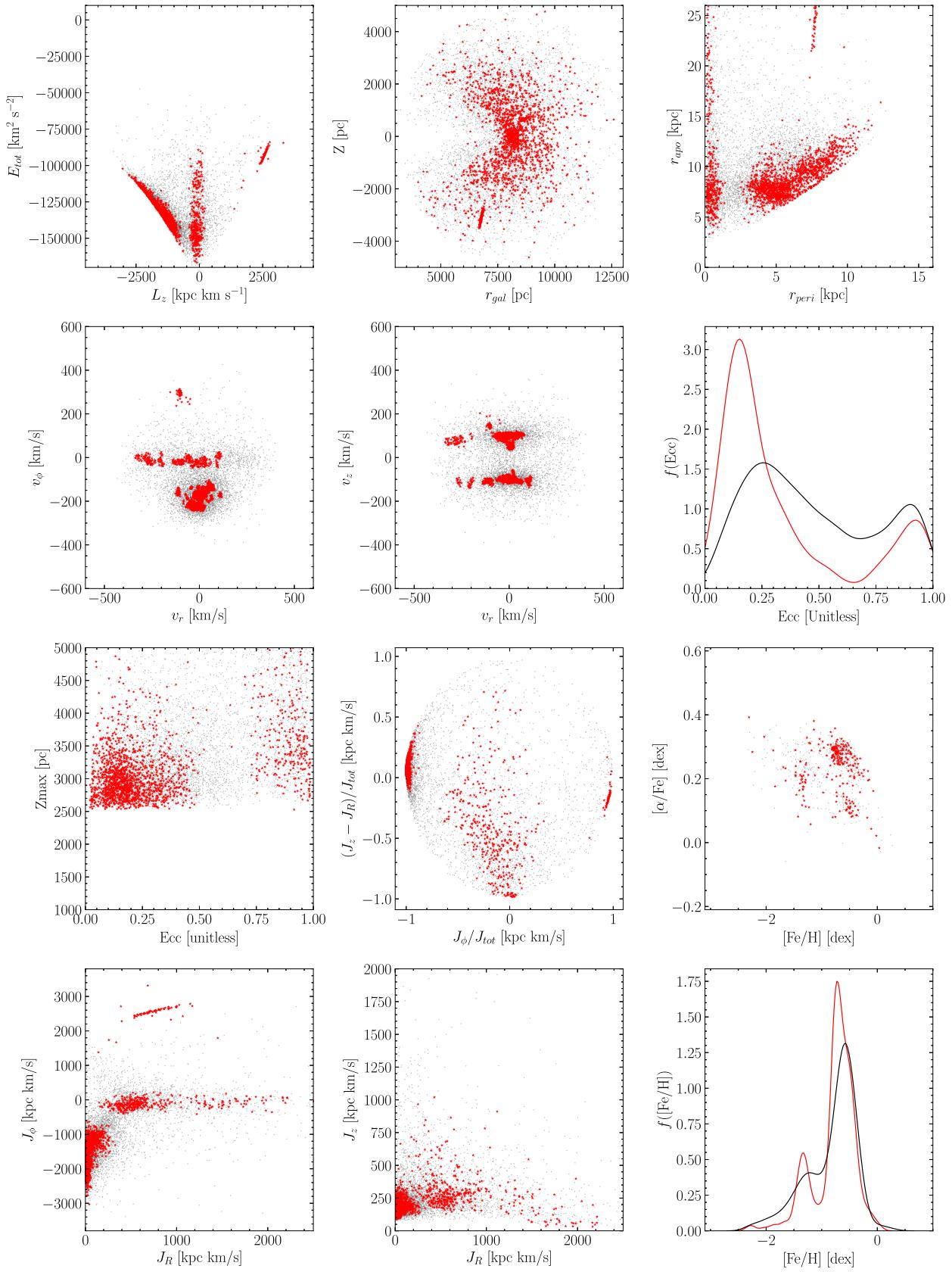
We choose  $N_{\text{stack}}^{\text{cut}} = 20$  to balance sample sizes and overall stabilities of the final clusters. The sample size of the high-stability sample reduces as shown in Fig. A1, while the stabilities of the final clusters improve with increasing  $N_{\text{stack}}^{\text{cut}}$ . We tested the clustering results' stabilities of high-stability samples using different  $N_{\text{stack}}^{\text{cut}}$ , following methods detailed in Section 3.5, and found that  $N_{\text{stack}}^{\text{cut}}$  at 20 yields the largest sample of stable stars with most of the final clusters being stable.

We find no obvious correlation between the  $N_{\text{stack}}^{\text{cut}}$  and the uncertainty distribution of the high-stability sample. A concern of the  $N_{\text{stack}}^{\text{cut}}$  would be that it reduced the data size by essentially performing an uncertainty cut, biasing towards stars that do not move as much in the clustering space. To test this, we examine the uncertainty distribution of the subsample in 3D cylindrical velocity space with and without  $N_{\text{stack}}^{\text{cut}}$ , shown in Fig. A2. The result indicates a minimum effect on the overall distribution of the uncertainties in any of the cylindrical velocities. We are thus confident that the stars with lower  $N_{\text{stack}}$  are not excluded because of high uncertainties but because they are at the edge of the identified clusters in the clustering spaces. In other words, under the same level of uncertainties, they blend in with the background noise more easily given their position in the clustering space, and thus  $N_{\text{stack}}$  is selecting physically more significant stars as part of a cluster than just statistically better measured stellar kinematics. The same test is carried out in the integrals of motion space with a similar result.

After identifying the stable stars, we apply HDBSCAN with the same parameter settings on only the stars, with  $N_{\text{stack}}$  above 20. The resulting clusters on this high-stability sample are taken as the final clusters.

### 3.5 Stability of clusters and stars

We apply the clustering algorithm as well as the 100 random realization process mentioned above to this new high-stability sample. As



**Figure 3.** Summary plot for the full sample (black), downsampled to 20 per cent, with the high-stability ( $N_{\text{stack}} > 20$ ) sample (red) identified from the cylindrical velocity clustering space.

discussed in Sections 3.3 and 3.4, the high-stability sample comprises only stars identified to be part of some clusters in at least 20 of the 100 random realizations. The clusters found this way are thus more robust than those from the full sample. They are also more stable against merging/splitting due to uncertainties because stars consistently identified as noise have been removed. We now discuss in more detail how we evaluate the stability of the final clusters.

We note that the  $N_{\text{stack}}$  distribution of this high-stability sample would be different from that of the original data set, e.g. we will again yield stars that have  $N_{\text{stack}} < 20$ . The goal of the cut, however, is to improve the stability (measured by the Jaccard coefficient as discussed below) of the final clusters from HDBSCAN. Once most clusters obtained from the 100 realizations of HDBSCAN is stabilized, it is not so important where and how many time an  $N_{\text{stack}}$  cut is applied. Suppose the cut on the original data set results in clusters that are still not stable, we could either go back and change the cut, or apply a further cut based on the new  $N_{\text{stack}}$  distribution. Ultimately, the cut is removing stars that are introducing instability in the clustering process.

We call the clusters found from the mean values of the clustering parameters the baseline clusters. We use the Jaccard coefficient  $J$  (Jaccard 1901) between two clusters from two different realizations to quantify the stability of the cluster. The coefficient  $J_s^a$  for a baseline cluster  $s$  is defined as the number of common members between the cluster  $s$  and a cluster  $a$  found in another realization, which we label  $N_{s,a}$ , divided by the total number of members in the initial cluster  $s$  that we label  $N_s$ .

$$J_s^a = \frac{N_{s,a}}{N_s}. \quad (1)$$

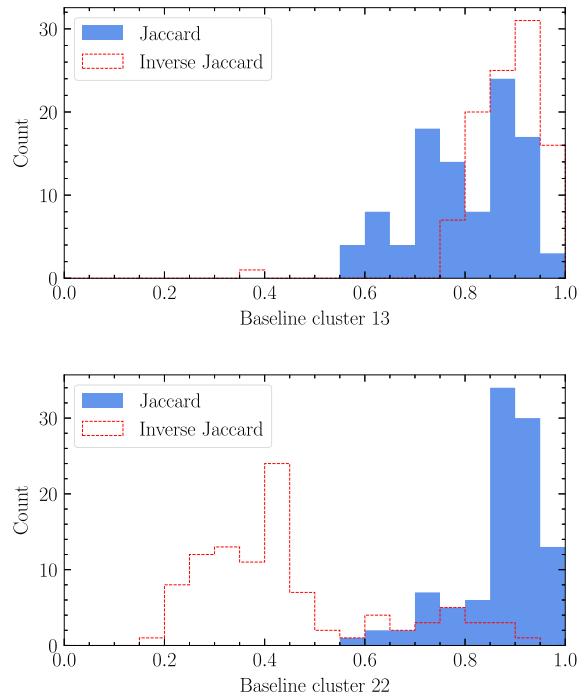
Baseline clusters  $s$  are those from the baseline realization, i.e. clusters found with the mean values, and clusters  $a$  are those found in each of the 100 random realizations.<sup>3</sup>

We also define the inverse Jaccard Coefficient  $J^\dagger$  as the number of common members divided by the total number of members in a random cluster. More specifically, the  $J_s^{a\dagger}$  for a baseline cluster  $s$  and a random realization cluster  $a$  is therefore defined as

$$J_s^{a\dagger} = \frac{N_{s,a}}{N_a}. \quad (2)$$

To study whether a baseline cluster  $s$  is stable, we find its closest cluster in a random realization, i.e. the cluster with which it shares the most members. To do so, we calculate the maximum  $J_s$ , which we call  $J_s^{r,\text{max}}$ , of the baseline cluster  $s$  (with respect to all the clusters) in each of the random realization  $r$ . We thus record at most 100 values of the  $J_s^{r,\text{max}}$  for each baseline cluster  $s$ ; some baseline clusters are not identified at all in some of the random realizations, in which cases the  $J_s^{r,\text{max}}$  is undefined. We also record the corresponding maximum value of the  $J_s^\dagger$  in each realization, called  $J_s^{\dagger r,\text{max}}$ ; this is applying equation (2) on the same pair of clusters (the baseline  $s$  and the random realization  $r$ ). We note that these  $J_s^{\dagger r,\text{max}}$  are therefore

<sup>3</sup>For illustrative purposes, we compute  $J_s^a$  between all baseline clusters and random realization clusters, i.e. for every baseline cluster, we study how much it overlaps with every other cluster in all 100 random realizations. Theoretically, the same thing could be done with any one of the random realizations serving the role of the baseline clusters, because they are ultimately treated as equal. The conclusion about the stability of the cluster membership remains valid, however, regardless of whether we use the baseline clustering as the base, or a random realization as the base. We use the mean values clustering result as the baseline. The only case of clusters that we might miss are the ones that are present in other realizations but not in the baseline.



**Figure 4.** Example Jaccard and inverse Jaccard distributions from clustering results in cylindrical velocity space with  $Z_{\text{max}}$  cut at 2.5 kpc and  $N_{\text{stack}}$  cut at 20. Cluster 13 represents a typical stable cluster, whereas cluster 22 represents a more ambiguous case where the Jaccard distribution seems stable while the inverse Jaccard is not. See text for detailed discussion of this case.

not necessarily the maximum possible values for the given random realization and baseline cluster.<sup>4</sup> We examine the  $J_s^{r,\text{max}}$  and  $J_s^{\dagger r,\text{max}}$  distribution of all the baseline clusters to evaluate their stability.

To distinguish stable from unstable clusters, we first study the distributions of  $J$  and  $J^\dagger$  of each of the baseline clusters. In Fig. 4, we show two examples of the distributions of  $J_s^{r,\text{max}}$  and  $J_s^{\dagger r,\text{max}}$  for two baseline clusters. To define stability, we take 50 per cent as the cut in  $J_s^{r,\text{max}}$  and  $J_s^{\dagger r,\text{max}}$  distributions. When over half of the cluster members in the baseline/random cluster are present in the corresponding random/baseline cluster, we say that the baseline cluster is successfully recovered in that random realization, and thus stable. Although the 50 per cent cut is a choice we make, we verified in the Jaccard distributions of all clusters (similarly to those shown in Fig. 4) that the stability of the cluster is not sensitive to the exact choice of the cut and is in most cases unambiguous.

More specifically, for a baseline cluster  $s$  showing consistently high Jaccard and inverse Jaccard coefficients, specifically  $J_s^{r,\text{max}} > 50$  per cent and  $J_s^{\dagger r,\text{max}} > 50$  per cent for all realizations  $r$ , the baseline cluster is stable against splitting. Similarly, baseline clusters  $s$  with consistently  $J_s^{r,\text{max}} < 50$  per cent and  $J_s^{\dagger r,\text{max}} < 50$  per cent are unstable. The more interesting cases are when the  $J$  and  $J^\dagger$  distributions are inconsistent, as shown in the bottom panel of Fig. 4; the high  $J_s$  but low  $J_s^\dagger$  distributions indicate that the baseline cluster is either absorbing more members, or merging with other clusters in random realizations. We can distinguish the two cases by examining if multiple baseline clusters share the same maximum  $J$  cluster in a given random realization. Lastly, we have baseline clusters with the opposite distributions, low  $J$  but high  $J^\dagger$ ; these clusters split in

<sup>4</sup>For example, in the case of baseline clusters splitting into multiple smaller clusters in random realizations ( $J_s^{\dagger r,\text{max}} \neq J_r^{s,\text{max}}$ ).

**Table 1.** Results from sample with  $N_{\text{stack}}$  greater than 20 in the cylindrical velocity clustering space.

Cluster ID	Colour	$v_R$ km s $^{-1}$	$\sigma_{v_R}$ km s $^{-1}$	$v_\phi$ km s $^{-1}$	$\sigma_{v_\phi}$ km s $^{-1}$	$v_z$ km s $^{-1}$	$\sigma_{v_z}$ km s $^{-1}$	$ecc$	[Fe/H] dex	$N_{[\text{Fe}/\text{H}]}$	$N_{\text{star}}$	Stability
0	■	-100.93	14.21	291.85	15.01	145.05	16.55	0.53		0	63	Stable
1	■	97.72	11.55	-2.75	19.65	-112.26	10.35	0.88	-1.16	9	39	Stable
2	■	110.0	4.39	-171.9	5.72	-95.88	4.74	0.36	-0.61	7	26	Stable
3	■	-111.46	5.48	-159.53	10.44	-106.72	5.86	0.41	-0.7	11	42	Stable
4	■	-109.77	16.78	-21.44	7.12	-110.45	9.18	0.86	-1.07	16	59	Stable
5	■	-281.08	26.82	-5.67	13.93	72.8	25.23	0.96	-1.47	8	64	Stable
6	■	-204.84	12.42	-19.4	11.72	-117.92	13.3	0.91	-0.99	2	24	Stable
7	■	24.75	8.75	-19.96	9.28	93.65	7.35	0.85	-1.11	4	26	Stable
8	■	-41.78	19.17	-11.6	11.1	97.05	8.69	0.89	-0.99	15	56	Stable
9	■	36.39	14.65	-151.04	11.71	-106.22	6.46	0.28	-0.65	26	167	Stable
10	■	-24.14	23.2	-225.72	8.45	88.6	15.14	0.13	-0.54	49	251	Unstable
11	■	-21.08	3.87	-171.0	8.7	-102.21	5.74	0.18	-0.55	11	54	Stable
12	■	-2.5	3.59	-160.82	5.33	-111.6	4.52	0.2	-0.56	10	37	Stable
13	■	52.34	5.4	-121.0	5.82	104.6	3.98	0.42	-0.78	8	34	Stable
14	■	21.54	2.52	-230.59	2.77	-75.43	10.34	0.13	-0.38	5	28	Unstable
15	■	-49.2	4.82	-207.99	5.06	-99.18	5.09	0.15	-0.54	5	26	Unstable
16	■	43.19	3.43	-153.73	5.16	100.18	7.57	0.29	-0.55	9	32	Stable
17	■	-27.67	3.01	-214.13	6.3	-90.66	4.52	0.1	-0.58	3	24	Unstable
18	■	5.53	6.88	-183.34	7.56	51.14	7.83	0.15	-0.3	6	184	Stable
19	■	-15.19	8.52	-172.05	8.28	103.42	7.02	0.17	-0.59	16	63	Stable
20	■	18.78	7.3	-172.49	4.68	104.28	5.6	0.17	-0.59	5	39	Ambiguous
21	■	-7.8	4.65	-233.56	2.34	-94.81	6.03	0.13	-0.31	2	20	Unstable
22	■	4.87	3.06	-208.89	8.21	-98.67	7.19	0.05	-0.48	13	47	Unstable

**Table 2.** Results from sample with  $N_{\text{stack}}$  greater than 20 in the energy and action clustering space.

Cluster ID	Colour	$J_R$ km kpc s $^{-1}$	$\sigma_{J_R}$ km kpc s $^{-1}$	$J_\phi$ km kpc s $^{-1}$	$\sigma_{J_\phi}$ km kpc s $^{-1}$	$J_z$ km kpc s $^{-1}$	$\sigma_{J_z}$ km kpc s $^{-1}$	$E_{\text{tot}}$ km $^2$ s $^{-2}$	$\sigma_{E_{\text{tot}}}$ km $^2$ s $^{-2}$	$ecc$	[Fe/H] dex	$N_{[\text{Fe}/\text{H}]}$	$N_{\text{star}}$	Stability
0	■	349.28	97.08	-1167.43	178.0	1155.14	99.66	-106802.56	2941.69	0.44	-1.56	12	65	Stable
1	■	1713.61	90.77	-27.99	104.25	69.67	17.76	-109012.71	2403.22	0.98	-1.07	7	55	Stable
2	■	757.92	134.15	2486.99	242.59	286.34	25.2	-93541.68	3778.38	0.54		0	59	Stable
3	■	573.94	47.09	-52.75	82.0	217.45	21.46	-147907.82	2284.24	0.93	-0.89	34	155	Stable
4	■	16.38	6.44	-2033.83	23.38	119.05	5.36	-120579.56	549.61	0.11	-0.45	16	57	Stable
5	■	17.11	9.01	-1760.66	86.17	133.7	7.04	-126802.74	2026.59	0.11	-0.51	58	287	Unstable
6	■	31.23	10.13	-1248.4	25.81	180.78	9.29	-139561.44	641.08	0.18	-0.56	14	131	Unstable
7	■	24.86	10.26	-1422.58	41.28	161.64	4.6	-135038.12	1151.29	0.15	-0.56	19	106	Stable

random realizations. Applied to the high-stability sample described earlier in this section, we find that most of the Jaccard distributions across the two clustering spaces show consistently high distributions of  $J$  and  $J^\dagger$ .<sup>5</sup> These clusters are identified as the final stable clusters, and the stars within them as core stable stellar members.

## 4 CLUSTERING RESULTS

We perform HDBSCAN for the high-stability sample (defined in Section 3.5) to obtain the clusters from the stable stars. We repeat the process for both clustering spaces (see Section 3.2). Note that the high-stability samples are different in each of the clustering spaces (and depend on the  $N_{\text{stack}}^{\text{cut}}$ , which we set to 20 as explained in Section 3.4). All clustering summary results shown here are from a single realization with the mean kinematic parameters of the stars. As discussed in Section 3, the clustering procedure is repeated for 100 random realizations, and all clusters are examined

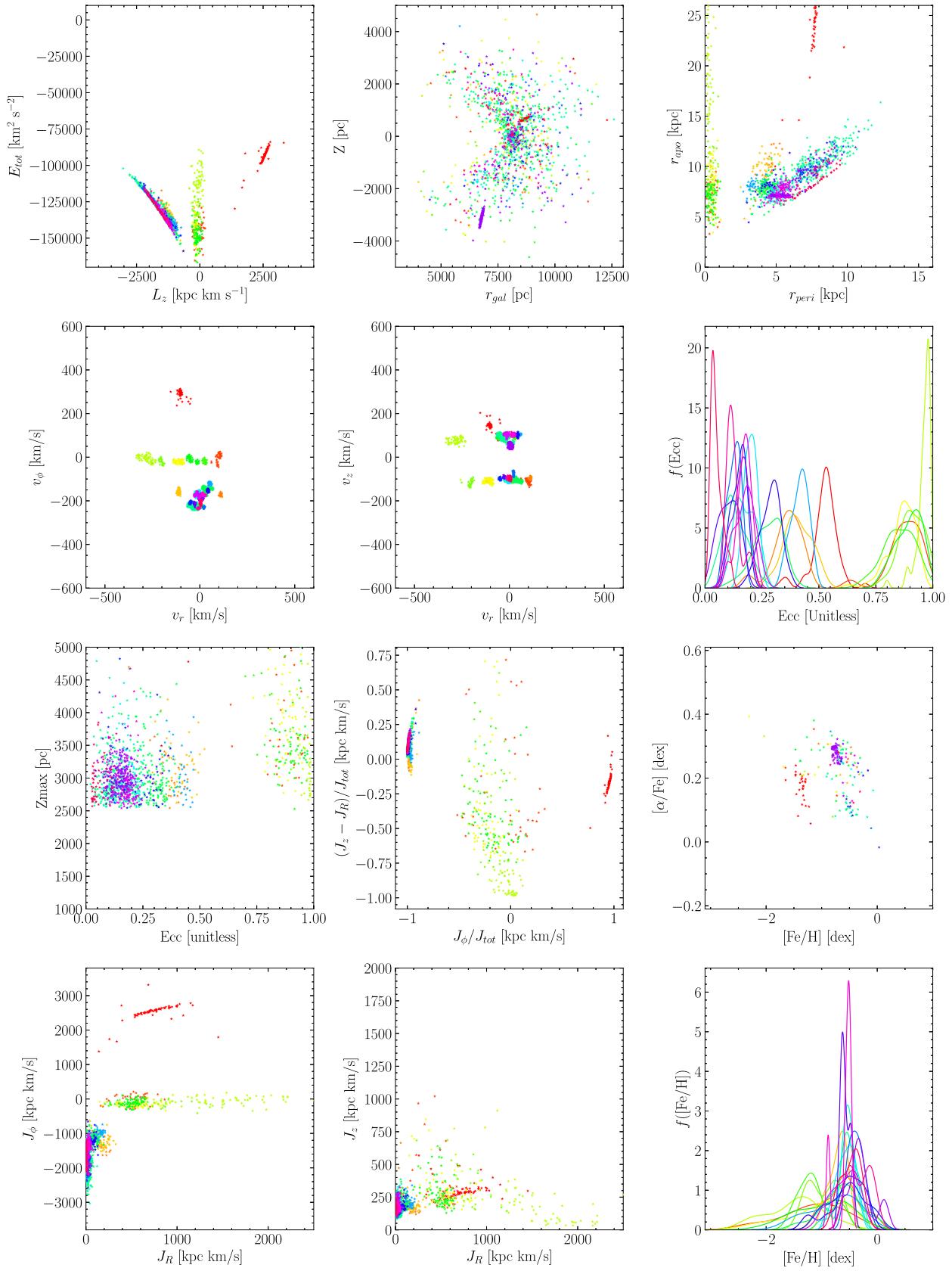
for their stability (see Section 3.5). One could principally pick any of the 100 random realizations as the baseline for the final clusters and the characterization of their stabilities, but because the clusters have stabilized, the result will not in fact change significantly. The reported stable clusters in Tables 1 and 2 are summarized based on the Jaccard stability studies. Therefore, the clusters from the summary figures may not all be classified as stable in the tables. We find 1405 stars in 23 stable clusters in velocity space and 497 stars in six stable clusters in energy and action space.

### 4.1 Velocity clusters

Running the above clustering procedure in velocity space, we show the clustering results in Fig. 5. The stars are colour-coded by the clusters with which they are associated. We note that even with this high-stability sample clustering space, not all stars are classified to be in one of the final clusters. The final stable clusters from 100 random realizations of HDBSCAN are listed in Table 1.

We see prograde structures with velocity distributions similar to that of the thin and thick disc. One noticeable difference is the slight bias towards positive radial velocity, i.e. more of the identified disc-like prograde clusters are found to be moving away from the galactic

<sup>5</sup>This is not the case if we run HDBSCAN over multiple realizations with the original sample, as a large fraction of the clusters merges and splits, making it difficult to identify the robust groups of stars, and their core member stars.



**Figure 5.** Summary plot for clusters that HDBSCAN identified from the high-stability ( $N_{\text{stack}} > 20$ ) sample with  $Z_{\text{max}}$   $2\sigma$  above 2.5 kpc from the cylindrical velocity space.

centre. We discuss each stable cluster identified here in more details in Section 5.

## 4.2 Integrals of motion clusters

The clustering results in the Integrals of Motion (Energy and Action) space applied to the high-stability sample are shown in Fig. 6. As expected, the clustering results are not the same as those identified in the cylindrical velocity space, as seen in Fig. 5. We examine and summarize the properties and stabilities of these clusters in Table 2.

## 4.3 Relationship between clusters

We examine the HDBSCAN merger trees from the clustering results mentioned in Sections 4.1 and 4.2. The trees are extracted by the `condensed_tree_` method from the HDBSCAN package. The merger tree allows us to directly see how clusters are related to each other along the hierarchical tree. This is especially important for our clustering method, `leaf`, which captures the smallest clusters in the sample and is thus designed to subdivide clusters that may be related to each other. It also provides intuition on which axis HDBSCAN primarily uses to distinguish clusters during the clustering process. Fig. 7 shows the merger tree for the clustering in velocity space. Identified clusters are shown with a colour matching that is given in Fig. 5. By examining the figure, we find, for example, that Cluster 0 from Table 1 is the first to be separated from the rest of the clusters, which is sensible given that Cluster 0 is the only retrograde structure found, and it is surrounded by low-density points that HDBSCAN deems to be noise. The stars are generally first separated by  $v_\phi$ , where we see a clear division between clusters with and without prograde kinematics.

To study whether multiple clusters belong to the same object, we perform a KS test (Kolmogorov 1933; Smirnov 1948) on the available metallicity of each cluster, which is obtained through cross-matching from LAMOST DR6 (Cui et al. 2012; Zhao et al. 2012).<sup>6</sup> Note that while we classified the clusters' stabilities based on their  $J$  and  $J^\dagger$  distributions, all clusters from the baseline realization, stable and unstable, are included in the following analysis to avoid inconsistencies caused by removing nodes in the merger tree. Some of the instability may be reflected during the merging process, so it is useful to keep them while inspecting the final results. We note that due to the lack of metallicity measurements for part of the sample, some of the smaller clusters contain too few valid metallicity measurements to provide enough statistics to perform a KS test. In cases where there are fewer than five available [Fe/H] measurements in a given cluster, we skip the comparison, move on to the parent node, and assume the child node can be merged.

We work from the bottom up to compare the metallicities of the clusters on each of the branches. If they are statistically consistent with being drawn from the same group, we merge them together. If not, we keep them as separate groups. We use the `KS_2samp` routine from the SCIPY package (Virtanen et al. 2020) to perform the KS test. The results are discussed further in Section 5.

We additionally note that the merging performed above will merge potential independent substructures with overlapping metallicities. As mentioned above, because the clustering algorithm is returning the smallest clusters above the minimum cluster size, it will artificially

split up substructures (that in fact belongs to one substructure) in the phase space where slight density fluctuations are present. One independent indicator for testing if that is the case is metallicity, which is what we performed here. This method inevitably omits physically independent substructures with similar metallicities distribution. More detailed population characterization and analyses of these individual clusters are needed to provide a conclusive argument on whether they are indeed independent substructures, which is beyond the scope of the current study.

## 5 DISCUSSION AND INTERPRETATION

We compare our final clustering results in the two clustering spaces (see Section 3.2), with parameters set at  $Z_{\max} > 2.5$  kpc, and  $N_{\text{stack}} > 20$ , against known Milky Way substructures from the literature.

### 5.1 Merged clusters

From Fig. 8 (left-hand panel) and Tables 1 and 3, we see that the unstable clusters 14, 15, 17, 21, and 22 are all mergeable (along with the stable clusters 9, 11, and 12). This indicates that the instability of these five clusters against splitting and merging is consistent with the fact that they are compatible with being part of a larger cluster in metallicity space. They are likely all coming from the same larger cluster in velocity space, but only split/merge due to their locations in phase space and the associated uncertainties. Thus, these unstable clusters are fragments of a larger cluster, split in different ways. The member stars in these clusters still remain robust, albeit for a larger cluster.

On the other hand, the unstable cluster 10 has no potentially merging clusters in metallicity space, suggesting that cluster 10 by itself is unlikely a robust cluster and dissolves in velocity space when considering uncertainties. Stellar members in this cluster are thus not robust cluster members.

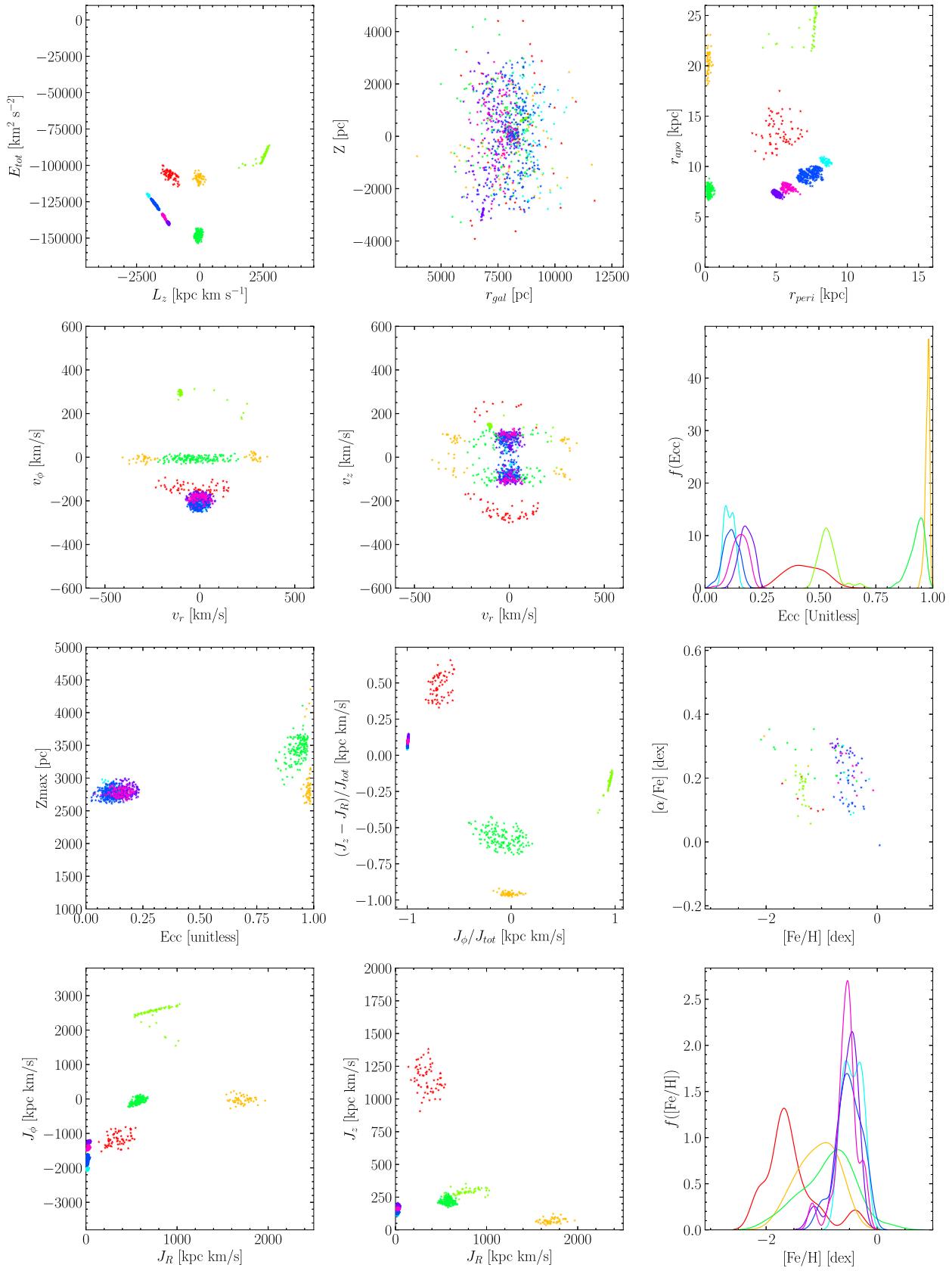
In addition to the previously identified unstable clusters, we also get stable clusters merged, which is the case for clusters 4–9, 11, 12, 19, and 20 from Table 1. Their corresponding parents, after merging, provide hints for the reason. For example, the newly merged clusters VI and X are likely both part of the thick disc, given the prograde rotational velocity of  $v_\phi$  at  $\sim -180$  km s $^{-1}$ . The `leaf` method we used for HDBSCAN divided them into smaller subgroups but their metallicities distributions allow us to confidently say that they can be merged back into a larger group.

We repeat the same process with the clustering result for the integrals of motion space and include the results in Fig. 8 (right-hand panel) and Table 4. Clusters 5 and 6, which were previously identified as unstable in Table 2, are merged together with clusters 4 and 7. These results are shown in Figs 9 and 10 for the velocity and integral of motion spaces, respectively. We discuss in the following subsections how the clusters compare with structures known from previous studies. When applicable, we use references from the literature where the kinematic values are obtained and derived similarly to our study to avoid systematic differences, i.e. original data from *Gaia* and similar galaxy potentials used in orbital integrations.

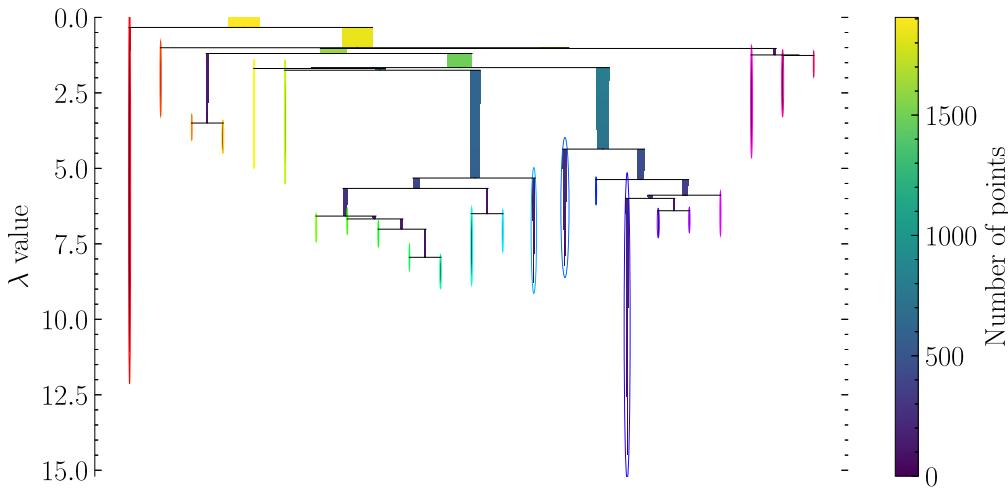
### 5.2 Velocity space

In this section, we associate clusters found in the velocity clustering space with known structures in the literature by discussing their mean kinematics and chemical characteristics in turn. A summary of the all the clusters discussed here can be found in Fig. 9, and Table 3.

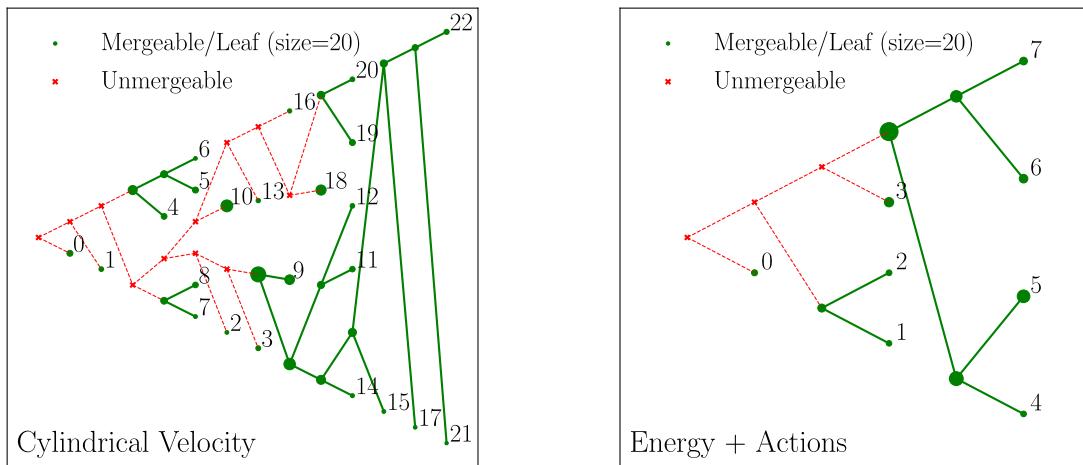
<sup>6</sup>We note that we use the metallicity measurements from LAMOST DR6 for better statistics in the comparison, as they provide more metallicities than APOGEE DR17, which were featured in the summary plots.



**Figure 6.** Summary plot for clusters that HDBSCAN identified from the high-stability ( $N_{\text{stack}} > 20$ ) sample with  $Z_{\text{max}}$   $2\sigma$  above 2.5 kpc from the integral of motion space.



**Figure 7.** Condensed merger tree for the stacked subsample with  $N_{\text{stack}} > 20$  and  $Z_{\text{max}} 2\sigma$  above 2.5 kpc from the cylindrical velocity space. The merger tree for the integral of motion space clustering is shown in Fig. A5. The vertical axis is  $\lambda$ , a measure of distance that HDBSCAN uses to define the edges of the cluster. See McInnes et al. (2017) for more details.



**Figure 8.** Semitree diagrams showing the nodes for the stacked stable subsample with  $Z_{\text{max}} 2\sigma$  above 2.5 kpc from the cylindrical velocity space (left) and the energy and action space (right). Size of the nodes is proportional to the size of cluster after merging children nodes. Green in general means the clusters can be merged together either because one or both of the children nodes have less than 10 member or because they passed the metallicity KS test. Red means the node could not be merged either because the test mentioned above failed or one of the children nodes failed to merge.

**Table 3.** Clusters merged based on metallicity distribution from Table 1.

Cluster ID	Colour	$v_R$ km s $^{-1}$	$\sigma_{v_R}$ km s $^{-1}$	$v_\phi$ km s $^{-1}$	$\sigma_{v_\phi}$ km s $^{-1}$	$v_z$ km s $^{-1}$	$\sigma_{v_z}$ km s $^{-1}$	$ecc$	[Fe/H] dex	$N_{[\text{Fe}/\text{H}]}$	$N_{\text{star}}$	Merged
I	■	-100.93	14.21	291.85	15.01	145.05	16.55	0.53	-	0	63	[0]
II	■	97.72	11.55	-2.75	19.65	-112.26	10.35	0.88	-1.16	9	39	[1]
III	■	-20.69	35.12	-14.25	11.25	95.97	8.44	0.88	-1.02	19	82	[7 8]
IV	■	110.0	4.39	-171.9	5.72	-95.88	4.74	0.36	-0.61	7	26	[2]
V	■	-111.46	5.48	-159.53	10.44	-106.72	5.86	0.41	-0.7	11	42	[3]
VI	■	8.88	29.44	-178.41	31.77	-101.21	10.86	0.19	-0.56	75	403	[9 11 12 14 15 17 21 22]
VII	■	-24.14	23.2	-225.72	8.45	88.6	15.14	0.13	-0.54	49	251	[10]
VIII	■	52.34	5.4	-121.0	5.82	104.6	3.98	0.42	-0.78	8	34	[13]
IX	■	5.53	6.88	-183.34	7.56	51.14	7.83	0.15	-0.3	6	184	[18]
X	■	-2.2	18.38	-172.21	7.13	103.75	6.53	0.17	-0.59	21	102	[19 20]
XI	■	43.19	3.43	-153.73	5.16	100.18	7.57	0.29	-0.55	9	32	[16]
XII	■	-199.88	81.15	-14.24	13.58	-31.89	93.79	0.91	-1.19	26	147	[4 5 6]

**Table 4.** Clusters merged based on metallicity distribution from Table 2.

Cluster ID	Colour	$J_R$ km kpc s $^{-1}$	$\sigma_{J_R}$ km kpc s $^{-1}$	$J_\phi$ km kpc s $^{-1}$	$\sigma_{J_\phi}$ km kpc s $^{-1}$	$J_z$ km kpc s $^{-1}$	$\sigma_{J_z}$ km kpc s $^{-1}$	$E_{tot}$ km $^2$ s $^{-2}$	$\sigma_{E_{tot}}$ km $^2$ s $^{-2}$	$ecc$	[Fe/H] dex	$N_{[Fe/H]}$	$N_{star}$	Merged
I	■	349.28	97.08	-1167.43	178.0	1155.14	99.66	-106802.56	2941.69	0.44	-1.56	12	65	[0]
II	■	1219.0	491.27	1273.62	1270.84	181.81	110.47	-101005.77	8362.99	0.75	-1.07	7	114	[1 2]
III	■	573.94	47.09	-52.75	82.0	217.45	21.46	-147907.82	2284.24	0.93	-0.89	34	155	[3]
IV	■	21.64	11.04	-1610.28	262.5	147.97	22.59	-130571.45	6408.31	0.14	-0.52	107	581	[4 5 6 7]

We start with the two globular clusters identified, NGC 3201 (cluster I shown as dark red in Fig. 9) and NGC 104 (cluster IX, shown as dark blue in Fig. 9). We find these clusters at  $(l, b) = (277.21, 8.63)$  degrees and  $(305.88, -44.90)$  degrees, and heliocentric distances of 4.22 and 4.33 kpc, respectively. These coordinates are consistent with the values of  $(l, b) = (277.23, 8.64)$  degrees and  $(305.89, -44.90)$  degrees from Harris (1996, 2010) considering the half-light radius of these two clusters are at 3.1 and 3.17 arcmin, respectively. The heliocentric distances in Harris (1996, 2010) are 4.9 and 4.5 kpc for NGC 3201 and NGC 104, respectively, where we attribute the disagreement with our values to systematic uncertainties in the different methods used in measuring distances, i.e. photometry versus astrometry. Finding these globular clusters within our framework is a validation of our clustering method. It is, however, worth noting that these two globular clusters appear more extended along the line of sight than they actually are from dedicated globular cluster studies (see Harris 2010, and references therein). This is a common effect in galactic coordinate transformation caused by the uncertainties in the distance to the sources. First, this further emphasizes the importance of incorporating uncertainties into the clustering. Additionally, a similar effect would be present for the entire sample and as a result systematically increase the scale height of the thin disc at high galactic latitude. This plays a role in interpreting the prograde structures we identify, which we discuss below.

Our newly identified clusters II, III, and XII have almost no rotational motion and low metallicity. Their chemical characteristics are overall consistent with being part of GSE, with mean [Fe/H] = -1.16, -1.02, and -1.19, respectively. These values agree with literature estimates of the metallicity of the GSE, which ranges from [Fe/H]  $\sim$  -1.6 to -1.1, depending on the selection criterion (Helmi et al. 2018; Necib et al. 2019a; Naidu et al. 2020). Because of the  $Z_{max}$  cut we applied, the stars identified as part of these clusters most likely correspond to the high vertical motion portion of the GSE. The majority of the GSE would be left out in the initial cut, yet the fact that the overdensity is still identified suggests the GSE has a non-negligible tail in high vertical velocity distribution. Clusters II and XII (shown in orange and red in Fig. 9) have the known high radial velocities of the GSE (Belokurov et al. 2018; Helmi et al. 2018). However, Cluster III (shown in yellow) does not. It could be that this cluster, which has eccentricities and metallicities consistent with clusters II and XII, encompasses the stars that connect the positive and negative radial velocity lobes of the GSE.

We now discuss the possibility of our prograde clusters being the high  $J_z$  tail of the thin disc or thick disc. Knowing that the uncertainties in the distances can systematically increase the scale height of the disc, stars originally on the exponential tail of the disc height distribution could be pushed outward into our selection region,  $Z_{max}$  greater than 2.5 kpc, as was shown earlier in Fig. 2. We thus expect there to be a significant amount of disc stars in the final high-stability sample. Cluster VII (in cyan), despite the significant vertical motion, does otherwise appear to be part of the thin disc. The cluster has rotational velocity  $v_\phi = -225.72$  km s $^{-1}$  and eccentricity of 0.13,

consistent with being part of the thin disc. The mean metallicity of -0.54 and low  $\alpha$ -abundance also support the thin disc origin.

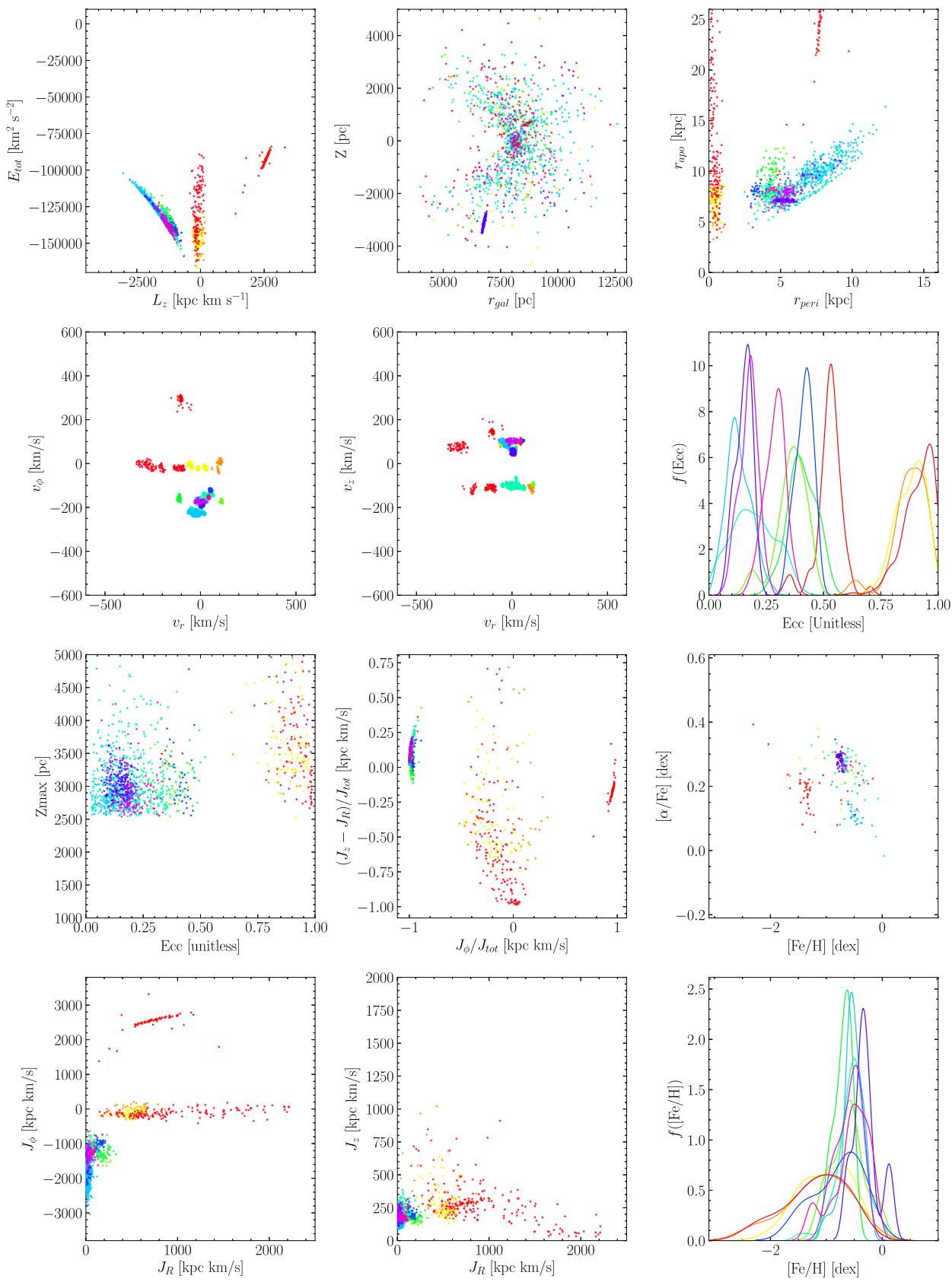
Additionally, clusters IV, V, VI, and X, coloured as fluorescent green, green, turquoise, and purple, respectively, sit in between the two groups mentioned above, with kinematics more consistent with that of the thick disc, and average  $v_\phi$  approximately 150 to 180 km s $^{-1}$ . The current estimate of the canonical thick disc scale height from recent literature study is  $0.9 \pm 0.2$  kpc [see Bland-Hawthorn & Gerhard (2016) for a review on galactic disc parameters]. The metallicity distributions of these clusters peak between -1 and 0 dex, with overall high  $\alpha$ -abundances, which are consistent with these stars being part of the canonical thick disc. We thus associate them as part of the thick disc. Clusters VIII (in dark blue) and XI (in pink), while being prograde, appear distinct from the thin disc with metallicity distributions extending lower than -1. This can be taken as a sign of these clusters containing accreted stellar populations or a portion of the metal-weak Atari disc (Norris et al. 1985; Morrison 1990; Chiba & Beers 2000; Mardini et al. 2022).

### 5.3 Integrals of motion space

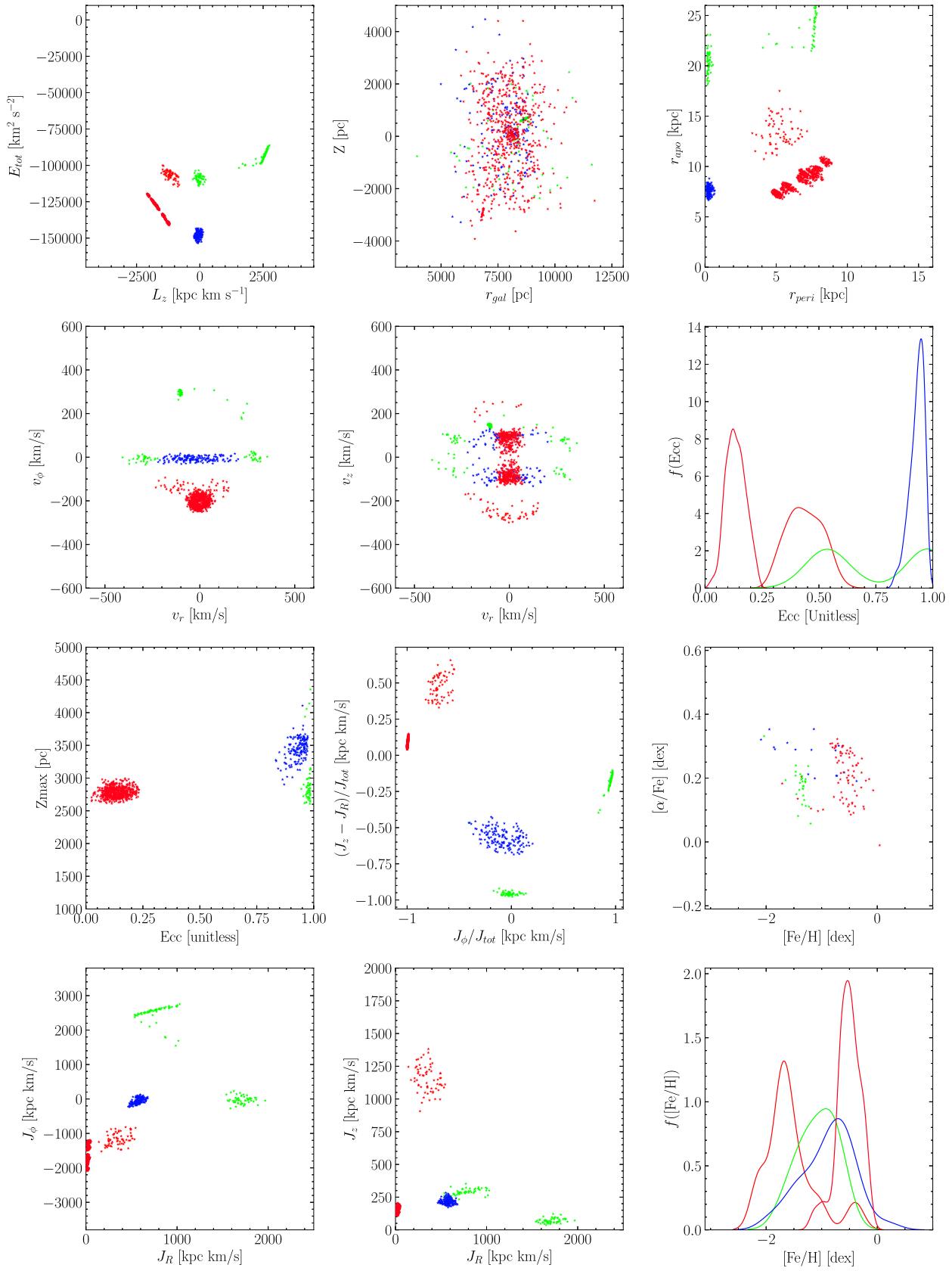
Similarly to the above, we associate the following clusters identified in energy and action space with known Milky Way substructures. We find four clusters that we list in Table 4 and show in Fig. 10. Cluster I, shown in the figure coloured in dark red, closely resembles the Helmi stream (Helmi et al. 1999), which is expected due to its well-known large vertical motion. Thus, most of its members pass the  $Z_{max}$  cut. The mean kinematics and metallicity of the cluster identified in our study are  $J_\phi = -1.17^{+0.10} \times 10^3$  kpc km s $^{-1}$ ,  $E_{tot} = -10.68^{+0.29} \times 10^4$  km $^2$  s $^{-2}$ , and [Fe/H] = -1.56. They agree with literature selection of the Helmi stream in  $L_z$  in  $[-1.7, -0.75] \times 10^3$  kpc km s $^{-1}$  (Koppelman et al. 2019a),  $E_{tot} = -10.3 \times 10^4$  km $^2$  s $^{-2}$ , and peak [Fe/H] ranging from -1.5 to -1.3 (Naidu et al. 2020). As the merger event associated with these streams happened approximately 5–8 Gyr ago (Koppelman et al. 2019a), the stream does not show a significant coherent spatial structure. Its stars distribute spatially similar to the background stars, except being confined by Galactic radial distance (Koppelman et al. 2019a).

Cluster II, shown in green, is a mix of three substructures. Merging based on metallicity alone has caused structures that are obviously not from the same group to be merged. Thus, instead of discussing the merged cluster II, we discuss pre-merging clusters 1 and 2, together with cluster III, i.e. cluster 3.

The radial motion dominated clusters (pre-merging clusters 1 and 3), resembling the GSE, are identified with much fewer clusters than in velocity space, and they are mainly distinguished by energy rather than radial velocity. Both clusters have mean metallicity [Fe/H]  $\sim$  -1, roughly consistent with literature values, but the lower energy cluster 3 shows potential for contamination from the high metallicity *in situ* halo at [Fe/H]  $\sim$  -0.5 and the big splash (stars formed in protodisc and have orbit altered to the highly eccentric orbit by the last major merger event) with metallicity  $>$  -0.7 from Belokurov et al. (2020).



**Figure 9.** Summary plot for merged clusters that HDBSCAN identified in the high-stability ( $N_{stack} > 20$ ) sample with  $Z_{max}$   $2\sigma$  above 2.5 kpc from the cylindrical velocity space. Note the colours do not match those from Fig. 5, but those quoted in Table 3.



**Figure 10.** Summary plot for merged clusters that HDBSCAN identified in the high-stability ( $N_{\text{stack}} > 20$ ) sample with  $Z_{\text{max}}$   $2\sigma$  above 2.5 kpc from the energy and action space. Note the colours do not match those from Fig. 6.

The higher energy cluster, on the other hand, appears to be more in line with the metal-poor nature of an accreted structure like the GSE.

Globular cluster NGC 3201 (our pre-merging cluster 2) is identified as in the cylindrical velocity space. There are, however, a few additional members identified in this cluster, which are most likely not true members of the globular cluster. We suspect they actually belong to the high energy retrograde structure Sequoia Myeong et al. (2019) and are instead mis-classified to be part of this globular cluster due to our HDBSCAN parameter setting. Namely, the minimum cluster size of 20 prohibited these Sequoia stars from forming their own cluster. Repeating HDBSCAN clustering in the full sample space with minimum cluster size reduced to 10, the result indicates Sequoia-like stars are identified separately as their own cluster. We are thus confident that we may safely remove the globular cluster core members and treat the rest of the cluster as a significant overdensity. We can do so given the globular cluster’s well-known coordinates  $[(l, b) = (277.21, 8.63)$  degrees] and angular radius (0.05°; Harris 2010) in the sky (see the details of this overdensity in Fig. A6). Note that the few stars that are left have no metallicities available.

There has been a discussion in the literature about the existence of multiple populations within the currently identified retrograde structure named Sequoia, i.e. Arjuna and I’itoi. The three structures are discussed in more detail in Naidu et al. (2020). The latter two are discovered in chemical space as separate peaks, with  $[Fe/H]$  of  $-1.6, -1.2, <-2$  for Sequoia, Arjuna, and I’itoi, respectively. The lack of metallicity measurements for our cluster prevents us from resolving potential population compositions of this prograde structure. Like GSE, Sequoia is also believed to have most of its stars having a zero mean  $v_z$ . Thus, our initial cut has preferentially removed Sequoia. Despite that, HDBSCAN identifies an overdensity at high  $Z_{\max}$ , signalling that Sequoia may be more kinematically heated and vertically extended than previously thought.

We associate multiple prograde structures (all in merged cluster IV, shown in dark red) with the thin and the thick disc, as they have rotational velocity of  $\sim 200 \text{ km s}^{-1}$  and metallicity  $\sim -0.5 \text{ dex}$ , consistent with the disc of the Milky Way (Bland-Hawthorn & Gerhard 2016). The position of these clusters in the  $Z_{\max}$  versus eccentricity plane also suggest the possibility of these stars being the high vertical motion end of the velocity distribution of the thin disc and thick disc. We evaluated whether these stars belong to Nyx; cross-matching Nyx and Nyx-2 with the high-stability sample in integrals of motion space yields no match. In fact, the majority of the Nyx stars are cut out by our  $Z_{\max}$  cut, so we would not expect this cluster to be associated with the Nyx.

#### 5.4 Comparison with other works

Previous work has been done in identifying local substructures self-consistently with only kinematic data as input. We now compare the results of our clustering algorithm to those of previous studies.

We first discuss the work of Ruiz-Lara et al. (2022) and Sofie Lövdal et al. (2022), which shares similar scientific goals and approaches as this study. Sofie Lövdal et al. (2022) adopted a *Gaia* eDR3 sample with cross-matched radial velocities from other spectroscopic surveys than APOGEE and LAMOST. The halo sample, used for clustering, is isolated by applying a  $v_{\text{LSR}}$  cut. They opted for the single linkage tree as the clustering algorithm, combined with Mahalanobis distances (Mahalanobis 1936) for final cluster and membership selections. While different, HDBSCAN is inherently also a linkage tree algorithm. The clustering space of Sofie Lövdal et al. (2022) is limited to energy and angular momenta, and uncertainties in these quantities have not been taken into account.

While both Sofie Lövdal et al. (2022) and our study identified structures such as the GSE, Helmi Stream, and Sequoia, we do not identify some of the structures (e.g. Thamnos) identified in Sofie Lövdal et al. (2022) mostly because of the different cuts applied. The  $Z_{\max}$  cut has removed the majority of the structures to the point that HDBSCAN can no longer pick up any overdensities in phase space. This is likely the case for Thamnos, which is known to be associated with a low stellar mass accretion event that happened a long time ago, a low metallicity, and low total energy. The mean  $\langle v_z \rangle$  for Thamnos is consistent with being zero from previous studies, and its current identified member count is lower than that of GSE or Sequoia. Moreover, our average cluster size is significantly smaller due to the different approach we used for selecting robust cluster members with uncertainties considered, along with the different algorithm used. While some clusters identified in their study also appear in similar phase space locations in single realizations of this study, the majority of the stars within these overdensities, as discussed in Section 3.3, do not robustly belong to said clusters when all random realizations are taken into account. We see in Fig. A1 that most of the stars ever associated with any cluster by HDBSCAN in the 100 random realizations have  $N_{\text{stack}} < 10$ , i.e. they are flagged as noise 90 per cent of the time. The robust clusters and their constituents are thus a smaller subset of what one would get by applying HDBSCAN to only the full sample with nominal value inputs.

Ruiz-Lara et al. (2022) used a cross-matched chemical information and mass-age relation to characterize the identified clusters in Sofie Lövdal et al. (2022). The membership selection was validated with isochrone fitting and metallicity distribution function study, where there are available matches in spectroscopic surveys. We do not carry out a similar validation because requiring stability of the clusters comes at the expense of statistics. As such, we cannot reliably characterize the progenitors with such few available chemical abundances. While our sample size does allow further study of the structure’s colour–magnitude diagram and potential comparison with synthetic colour–magnitude diagrams, factoring in the observational uncertainties from the data and theoretical uncertainties from the stellar evolution models is not a trivial effort (Cignoni & Tosi 2010). We thus defer characterizing our robust clusters with synthetic colour–magnitude diagram fitting procedures to a future study.

We also compare our results with those from Naidu et al. (2020), where selections are more based on physical intuition. Again, we recover the GSE, Helmi Stream, and Sequoia, but miss Thamnos for the same reasons mentioned above. We also do not detect any structure that resembles Aleph or Wukong, both reported to have high  $J_z$  (Naidu et al. 2020). While Aleph has a relatively high vertical motion, its selection required chemical information that is not available to our study. Wukong is a structure found to be two overdensities near the margin of the GSE in the  $E_{\text{tot}}$  versus  $L_z$  plane. With only kinematic information, we suspect that incorporating uncertainties has caused the cluster to become unstable and thus make it more difficult for HDBSCAN to robustly separate it from the GSE and background. We also point out that Naidu et al. (2020) focused on stars 3 kpc away from the solar neighbourhood, so it is also likely that these two structures are only present/or dominate at high galactic latitude, and thus we do not have large enough statistics within our sample to identify such kinds of structures.

For similar reasons, structures that are known to be dominant in the outer halo are also not identified in this study, for example, the Sagittarius stream (Ibata et al. 1994) and the Cetus stream (Yuan et al. 2019). The distant nature of these structures makes it less likely for their stars to have *Gaia* DR2 radial velocity measurements,

hence preventing from being detected by the algorithm. Likewise, because our method relies on the stellar overdensities in phase-space as tracers for the underlying substructures, Pontus (Malhan 2022), Kraken (Kruijssen et al. 2020) and other substructures identified through grouping of globular clusters are also not identified in this study.

In summary, our method has some limitations. The explorable space is limited to where we have the full 6D kinematic information, as seen in the case of the non-detection of Sagittarius and other substructures that are more dominant at outer halo distances, and to which background cut we apply, as is the case for the non-detections of Nyx or Thamnos. The algorithm also relies solely on kinematic overdensities of stellar tracers and is thus not making use of other accretion event signatures such as accompanying globular clusters. In the absence of easily accessible additional information, such limitations will remain common to studies using stellar kinematics to search for substructures in the MW. They certainly do not invalidate the importance of incorporating measurement uncertainties into any searching method used.

Our clustering results give two clusters associated with the GSE. Donlon Thomas et al. (2022) discussed the GSE being composed of multiple components, namely the VRM, Nereus, and Cronus, using kinematics and metallicities for *Gaia* eDR3 dwarf stars. We compare our pre-merging clusters 1 and 3 in the integrals of motion space with the three radial merger events identified by Donlon et al. (2022). Our pre-merging cluster 1 resembles the VRM and Nereus with its high energy and double-lobed  $v_r$  velocity profile, but the average metallicity of our cluster  $[\text{Fe}/\text{H}] \sim -1.07$  is higher than those of the VRM and Nereus at  $-1.7$  and  $-2.1$ , respectively. Similarly, the lower energy cluster 3 has slightly higher metallicity  $[\text{Fe}/\text{H}] \sim -0.89$  than the Cronus at  $-1.2$ . Further, while the Cronus was shown to be prograde with  $L_z \sim -400 \text{ kpc km s}^{-1}$ , our low energy GSE-like cluster is consistent with being non-rotating. In short, while our clustering results in the integral of motion space separate the GSE into high-energy and low-energy components, the metallicities do not show as significant a difference as shown in Donlon et al. (2022). The discrepancy may be a combined result of systematic differences in metallicity derivation, given that Donlon et al. (2022) used photometric metallicity estimates calibrated using a collection of spectroscopic survey catalogues in the literature [see section 2.3 in Kim & Lépine (2022)] while we only used LAMOST DR6, and/or potential contamination from *in situ* stars in our clusters given that our clustering is based solely on kinematic properties. For a similar reason, we are not able to distinguish potential constituents in our high eccentricity clusters the same way as in Myeong et al. (2022), where chemical abundances from APOGEE DR17 (Abdurro'uf et al. 2022) and GALAH DR3 (Buder et al. 2021) are used to construct the chemo-dynamical space for Gaussian Mixture Modelling.

## 6 INTERPRETATION AND CONCLUSIONS

In this section, we discuss both the challenges in identifying local structures using clustering algorithms as well the potential that new approaches can deliver with current and future data with regards to the detailed exploration of the Galaxy's assembly and accretion history.

In general, while satellite galaxies that have merged or are in the process of merging with the Milky Way can principally be extracted with clustering tools, the observed signal is often too weak compared to the background (e.g. structures within the disc) and identification becomes even more challenging with increasing measurement uncertainties. Accordingly, the current standard practice of

largely mitigating unreliable selections has been focused on using either chemical abundance information or isochrone fitting techniques to validate the stellar membership. However, this approach, when applied to the entirety of the Galaxy, will be limiting going forward; chemical information is severely restricted by the difficulty in obtaining medium or high-resolution spectroscopic metallicities and abundances for quantities of stars that are comparable to those of large-scale astrometric and photometric surveys such as *Gaia*. Isochrone fitting of a collection of stars is dependent on the completeness and selection function of any sample. These methods of *a posteriori* resorting the data also partially undermine the purpose of using an unsupervised machine learning algorithm to handle large data sets and make objective decisions unaffected by human biases.

Another problem on its own is measurement uncertainties associated with input kinematic data. Unfortunately, the current scope of our study is severely limited by detailed and precise 6D kinematics information, especially for fainter stars that probe deep into the galactic halo, when measurement uncertainties become larger or no information is available at all. Membership assignments by hierarchical clustering algorithms, such as HDBSCAN, are generally unreliable when uncertainties are not taken into account. For example, when considering stars that are associated with a random cluster in more than one out of 100 random clustering realizations, the majority of the stars are often assigned to be noise in the other 99 realizations. However, this is not too surprising. Hierarchical clustering algorithms, e.g. single-linkage algorithms such as HDBSCAN, rely on first building some form of a hierarchical merger tree, which can alter significantly when data points are shifted within their respective uncertainties. A changed tree will naturally lead to altered findings and as such will yield unreliable results when it comes to identifying new overdensities and their constituents. Thus, the results from applying such algorithms straight to the nominal mean values of the input parameters cannot be trusted. *We thus stress the need for numerical techniques that produce the most reliable and reproducible clustering results, by taking advantage of incorporating existing observational and astrometric uncertainties in the clustering algorithms. Specifically, we propose a more robust method of utilizing HDBSCAN in identifying local stable stellar overdensities in kinematic space.*

To overcome these challenges and to identify robust and reliable clusters, in this study, we use *Gaia* eDR3 astrometry data together with line-of-sight radial velocity from *Gaia* DR2 to identify clusters in the solar neighbourhood using HDBSCAN. In our novel procedure, we successfully incorporate measurement uncertainties in the clustering algorithm to produce stable and robust clusters and recover previously known structures in the Milky Way, specifically the GSE, Sequoia, the Helmi Stream, and the globular clusters NGC 3201 and NGC 104. We achieved this by resampling the input data, repeating the clustering process, and treating each random realization equally. We then incorporate measurement uncertainties in the clustering algorithm by resampling the input data, repeating the clustering process, and treating each random realization equally. This technique allows us to confidently isolate the most stable and robust cluster members by stacking the clustering results from each realization and removing those that are associated with noise in more than 20 (out of 100) realizations. These cluster members are then re-clustered into our final set of stable clusters.

When properly incorporating measurement uncertainties in a routine fashion, the resulting robust cluster members will naturally be much fewer than those that would be obtained from applying algorithms to only the nominal values. We have thus focused in this study on retrieving stellar core members of structures near us.

Regardless, these results already provide candidate structures that could be targeted for follow-up observations with future spectroscopic surveys, such as 4MOST (de Jong et al. 2019) and WEAVE (Dalton et al. 2012). Since obtaining spectroscopic data requires significant amounts of time, as more clusters as well as cluster candidates are being identified, targeted follow-up will be maximally efficient, ensuring a high discovery rate.

To further increase efficiency and maximal identification of as many cluster member stars as possible, as a next step, we anticipate expanding our technique to include a ‘cluster re-building’ component in which we start with the robust clusters and then carry out an additional member selection with progressively loosened stability criteria to eventually re-build the full clusters step-by-step, still with uncertainties factored in. This approach makes use of the fact that the existence of the cluster is no longer in question, and instead focuses efforts on finding the complete stellar membership.

More broadly, our technique of providing precise identifications of substructure signatures has the potential to go beyond finding just accreted structures. In particular, in the galactic disc, studies of moving groups have long revealed many overdensities of stars moving in common orbits (Antoja et al. 2018), with the most dominant ones being the branches of Sirius, Coma Berenices, Hyades-Pleiades, and Hercules. These local overdensities, proposed to be induced by dynamical resonances, past perturbations, and/or non-axisymmetric features of the Milky Way, are all prime targets of research and provide valuable insights into the formation and evolution of the galactic disc [see e.g. appendix B of Trick, Coronado & Rix (2019) for a short review]. New moving groups are still being discovered thanks to constantly improving methodologies (Ramos, Antoja & Figueras 2018; Bernet et al. 2022; Lucchini et al. 2023). Identifying these overdensities in the disc with high confidence requires careful treatment of the uncertainties, as the structures will most likely be accompanied by strong background noise. Lucchini et al. (2023), for example, incorporated various measurement uncertainties with wavelet transforms, which is another promising way to maximize unsupervised data analysis outputs. While our method provides potentially a way to accomplish this goal, we do note that the galactic disc is a very different environment from the galactic halo. Sufficient testing and cross-validating with other methods are required to establish the applicability of our method to the highly-dense and populated galactic disc.

Apart from the uncertainties on the kinematic measurements of the stars, it would also be both interesting and beneficial to test the stabilities of the clusters as a result of uncertainties in the galactic potential model. Since the integral of motion parameters are computed by assuming some potential model, uncertainties in these models manifest as additional systematic uncertainties in these parameters. By testing our clustering procedure in different galactic potential models, we may examine how robust and stable the clusters are against the assumed potential model of the Milky Way. We aim to explore different galaxy models in future work with particular emphasis on the DM density profiles.

These novel data analysis approaches, and as more and deeper data become available, from both space missions, such as *Gaia*, and ground missions, such as the Vera Rubin Observatory Legacy Survey of Space and Time (Ivezic et al. 2019), showcase the huge discovery potential for stellar (sub)structures in the Milky Way. We thus anticipate that many more delicate, thus far hidden clusters will be discovered. But success will rest solely on our ability to rely on unsupervised machine learning algorithms, such as HDBSCAN, to provide trustworthy outputs. We can only trust algorithms when

we factor in measurement uncertainties that, of course, are always present in observational astronomy.

In closing, understanding in detail how accurate these techniques and algorithms are given the current level of uncertainties in observational measurements is an important step towards building a complete picture of the Milky Way, as well as towards providing improved modelling capabilities of the local DM phase space distribution.

We thus encourage future clustering studies to begin testing and incorporating uncertainties into their new generation of techniques and unsupervised machine learning algorithms to acquire reliable and robust selections of substructures and their member stars based on kinematic information only.

## SOFTWARE

The analysis for this work was coded in PYTHON v. 3.7.6 (Van Rossum & Drake 2009) and includes its packages NUMPY (van der Walt, Colbert & Varoquaux 2011) and SCIPY (Virtanen et al. 2020). We used ASTROPY (Astropy Collaboration 2013, 2018) and AGAMA (Vasiliev 2019) for coordinate transformation and orbit integration. HDBSCAN (McInnes et al. 2017) was used for clustering and figures are generated with MATPLOTLIB (Hunter 2007).

## ACKNOWLEDGEMENTS

We thank M Lisanti, T Nguyen, C Roche, and N Shipp for helpful discussions. We thank the referee for the insightful feedback. XO and AF acknowledge support from NSF grant AST-1716251.

This work presents results from the European Space Agency (ESA) space mission *Gaia*. *Gaia* data are being processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the *Gaia* MultiLateral Agreement (MLA). The *Gaia* mission website is <https://www.cosmos.esa.int/gaia>. The *Gaia* archive website is <https://archives.esac.esa.int/gaia>.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High Performance Computing at the University of Utah. The SDSS website is [www.sdss4.org](http://www.sdss4.org).

Guoshoujing Telescope (the Large Sky Area Multi-Object Fiber Spectroscopic Telescope LAMOST) is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

This work used Stampede-2 under allocation number TG-PHY210118, part of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF grant number ACI-1548562.

This research has made use of NASA’s Astrophysics Data System Bibliographic Services; the arXiv pre-print server operated by Cornell University; the SIMBAD and VizieR data bases hosted by the Strasbourg Astronomical Data Centre.

## DATA AVAILABILITY

Source ID of the clustered *Gaia* stars from this study is available for download as online material accompanying this work. Table A1 is included here to demonstrate its form and content.

## REFERENCES

- Abdurro'uf et al., 2022, *ApJS*, 259, 35
- Antoja T. et al., 2018, *Nature*, 561, 360
- Astropy Collaboration, 2013, *A&A*, 558, A33
- Astropy Collaboration, 2018, *AJ*, 156, 123
- Behroozi P., Wechsler R. H., Hearn A. P., Conroy C., 2019, *MNRAS*, 488, 3143
- 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., Fattah A., Smith M. C., Deason A. J., Evans N. W., Grand R. J. J., 2020, *MNRAS*, 494, 3880
- Bennett M., Bovy J., 2019, *MNRAS*, 482, 1417
- Bernet M., Ramos P., Antoja T., Famaey B., Monari G., Al Kazwini H., Romero-Gómez M., 2022, *A&A*, 667, A116
- Binney J., Spergel D., 1982, *ApJ*, 252, 308
- Bird S. A., Xue X.-X., Liu C., Shen J., Flynn C., Yang C., Zhao G., Tian H.-J., 2021, *ApJ*, 919, 66
- Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529
- Bovy J., 2015, *ApJS*, 216, 29
- Brauer K., Andales H. D., Ji A. P., Frebel A., Mardini M. K., Gomez F. A., O'Shea B. W., 2022, *ApJ*, 937, 14
- Buder S. et al., 2021, *MNRAS*, 506, 150
- Buder S. et al., 2022, *MNRAS*, 510, 2407
- Bullock J. S., Johnston K. V., 2005, *ApJ*, 635, 931
- Carollo D., Chiba M., 2021, *ApJ*, 908, 191
- Chiba M., Beers T. C., 2000, *AJ*, 119, 2843
- Chiti A., Frebel A., Mardini M. K., Daniel T. W., Ou X., Uvarova A. V., 2021, *ApJS*, 254, 31
- Cignoni M., Tosi M., 2010, *Adv. Astron.*, 2010, 158568
- Conroy C. et al., 2019, *ApJ*, 883, 107
- Cooper A. P. et al., 2010, *MNRAS*, 406, 744
- Cui X.-Q. et al., 2012, *Res. Astron. Astrophys.*, 12, 1197
- Dalton G. et al., 2012, in McLean I. S., Ramsay S. K., Takami H. eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460P
- de Jong R. S. et al., 2019, *The Messenger*, 175, 3
- Donlon Thomas I., Newberg H. J., Kim B., Lépine S., 2022, *ApJ*, 932, L16
- Drimmel R., Poggio E., 2018, *Res. Notes Am. Astron. Soc.*, 2, 210
- Drukier A. K., Freese K., Spergel D. N., 1986, *Phys. Rev. D*, 33, 3495
- Fattah A. et al., 2020, *MNRAS*, 497, 4459
- Feuillet D. K., Feltzing S., Sahlholdt C., Bensby T., 2022, *ApJ*, 934, 21
- Frebel A., Norris J. E., 2015, *ARA&A*, 53, 631
- Freeman K., Bland-Hawthorn J., 2002, *ARA&A*, 40, 487
- Frenk C. S., White S. D. M., 2012, *Ann. Phys.*, 524, 507
- Gaia Collaboration, 2016, *A&A*, 595, A1
- Gaia Collaboration, 2018, *A&A*, 616, A1
- Gaia Collaboration, 2021, *A&A*, 649, A1
- Gómez F. A., Helmi A., Brown A. G. A., Li Y.-S., 2010, *MNRAS*, 408, 935
- Goodman M. W., Witten E., 1985, *Phys. Rev. D*, 31, 3059
- Gravity Collaboration, 2018, *A&A*, 615, L15
- Griffen B. F., Ji A. P., Dooley G. A., Gómez F. A., Vogelsberger M., O'Shea B. W., Frebel A., 2016, *ApJ*, 818, 10
- Gudin D. et al., 2021, *ApJ*, 908, 79
- Harris W. E., 1996, *AJ*, 112, 1487
- Harris W. E., 2010, preprint ([arXiv:1012.3224](https://arxiv.org/abs/1012.3224))
- Helmi A., 2020, *ARA&A*, 58, 205
- Helmi A., White S. D. M., 1999, *MNRAS*, 307, 495
- Helmi A., White S. D. M., de Zeeuw P. T., Zhao H., 1999, *Nature*, 402, 53
- Helmi A., Veljanoski J., Breddels M. A., Tian H., Sales L. V., 2017, *A&A*, 598, A58
- Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, *Nature*, 563, 85
- Hernquist L., 1990, *ApJ*, 356, 359
- Herzog-Arbeitman J., Lisanti M., Madau P., Necib L., 2018a, *Phys. Rev. Lett.*, 120, 041102
- Herzog-Arbeitman J., Lisanti M., Necib L., 2018b, *J. Cosmol. Astropart. Phys.*, 2018, 052
- Hopkins P. F. et al., 2018, *MNRAS*, 480, 800
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Ibata R. A., Gilmore G., Irwin M. J., 1994, *Nature*, 370, 194
- Ibata R., Bellazzini M., Thomas G., Malhan K., Martin N., Famaey B., Siebert A., 2020, *ApJ*, 891, L19
- Ivezic Ž. et al., 2019, *ApJ*, 873, 111
- Jaccard P., 1901, *Bull. Soc. Vaudoise Sci. Naturelles*, 37, 241
- Johnston K. V., 1998, *ApJ*, 495, 297
- Johnston K. V., Hernquist L., Bolte M., 1996, *ApJ*, 465, 278
- Johnston K. V., Bullock J. S., Sharma S., Font A., Robertson B. E., Leitner S. N., 2008, *ApJ*, 689, 936
- Jungman G., Kamionkowski M., Griest K., 1996, *Phys. Rep.*, 267, 195
- Jurić M. et al., 2008, *ApJ*, 673, 864
- Keller S. C. et al., 2007, *Publ. Astron. Soc. Aust.*, 24, 1
- Kim B., Lépine S., 2022, *MNRAS*, 510, 4308
- Knebe A., Gill S. P. D., Kawata D., Gibson B. K., 2005, *MNRAS*, 357, L35
- Kolmogorov A. N., 1933, *G. Ist. Ital. Attuari.*, 4, 83
- Koppelman H. H., Helmi A., Massari D., Roelenga S., Bastian U., 2019a, *A&A*, 625, A5
- Koppelman H. H., Helmi A., Massari D., Price-Whelan A. M., Starkenburg T. K., 2019b, *A&A*, 631, L9
- Kruisissen J. M. D. et al., 2020, *MNRAS*, 498, 2472
- Lee Y. S., Beers T. C., Kim Y. K., 2019, *ApJ*, 885, 102
- Li T. S. et al., 2022, *ApJ*, 928, 30
- Limberg G. et al., 2021, *ApJ*, 907, 10
- Lindgren L. et al., 2021a, *A&A*, 649, A2
- Lindgren L. et al., 2021b, *A&A*, 649, A4
- Lucchini S., Pellett E., D'Onghia E., Aguerri J. A. L., 2023, *MNRAS*, 519, 432
- Mackereth J. T. et al., 2019, *MNRAS*, 482, 3426
- Mahalanobis P. C., 1936, in Proc. National Institute of Sciences, Vol. 2, On the Generalized Distance in Statistics. Calcutta, p. 49
- Maiolino R., Mannucci F., 2019, *A&AR*, 27, 3
- Majewski S. R., Skrutskie M. F., Weinberg M. D., Ostheimer J. C., 2003, *ApJ*, 599, 1082
- Majewski S. R. et al., 2017, *AJ*, 154, 94
- Malhan K., 2022, *ApJ*, 930, L9
- Malhan K., Yuan Z., Ibata R. A., Arentsen A., Bellazzini M., Martin N. F., 2021, *ApJ*, 920, 51
- Mardini M. K., Frebel A., Chiti A., Meiron Y., Brauer K. V., Ou X., 2022, *ApJ*, 936, 78
- McInnes L., Healy J., Astels S., 2017, *J. Open Source Softw.*, 2, 205
- Meingast S., Alves J., Fürnkranz V., 2019, *A&A*, 622, L13
- Miyamoto M., Nagai R., 1975, *PASJ*, 27, 533
- Morrison H. L., 1990, *J. R. Astron. Soc. Canada*, 84, 107
- Myeong G. C., Evans N. W., Belokurov V., Amorisco N. C., Koposov S. E., 2018, *MNRAS*, 475, 1537
- 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
- Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, 490, 493
- Necib L., Lisanti M., Belokurov V., 2019a, *ApJ*, 874, 3
- Necib L., Lisanti M., Garrison-Kimmel S., Wetzel A., Sanderson R., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2019b, *ApJ*, 883, 27
- Necib L. et al., 2020a, *Nature Astron.*, 4, 1078
- Necib L., Ostdiek B., Lisanti M., Cohen T., Freytsis M., Garrison-Kimmel S., 2020b, *ApJ*, 903, 25
- Newberg H. J., Yanny B., Willett B. A., 2009, *ApJ*, 700, L61
- Norris J., Bessell M. S., Pickles A. J., 1985, *ApJS*, 58, 463
- O'Hare C. A. J., Evans N. W., McCabe C., Myeong G., Belokurov V., 2020, *Phys. Rev. D*, 101, 023006

- Ostdiek B. et al., 2019, Cataloging Accreted Stars within Gaia DR2 Using Deep Learning, Zenodo, available at: <https://doi.org/10.5281/zenodo.3579379>
- Ostdiek B. et al., 2020, *A&A*, 636, A75
- Price-Whelan A. M., 2017, *J. Open Source Softw.*, 2, 18
- Ramos P., Antoja T., Figueiras F., 2018, *A&A*, 619, A72
- Re Fiorentin P., Spagna A., Lattanzi M. G., Cignoni M., 2021, *ApJ*, 907, 16
- Reid M. J., Brunthaler A., 2004, *ApJ*, 616, 872
- Riello M. et al., 2021, *A&A*, 649, A3
- Ruiz-Lara T., Matsuno T., Sofie Lövdal S., Helmi A., Dodd E., Koppelman H. H., 2022, *A&A*, 665, A58
- Sanderson R. E. et al., 2020, *ApJS*, 246, 6
- Schlaufman K. C. et al., 2009, *ApJ*, 703, 2177
- Sestito F. et al., 2019, *MNRAS*, 484, 2166
- Shank D. et al., 2022, *ApJ*, 926, 26
- Shipp N. et al., 2018, *ApJ*, 862, 114
- Smirnov N., 1948, *Ann. Math. Statist.*, 19, 279
- Sofie Lövdal S., Ruiz-Lara T., Koppelman H. H., Matsuno T., Dodd E., Helmi A., 2022, *A&A*, 665, A57
- Soubiran C., Bienaymé O., Siebert A., 2003, *A&A*, 398, 141
- Trick W. H., Coronado J., Rix H.-W., 2019, *MNRAS*, 484, 3291
- van der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, 13, 22
- Van Rossum G., Drake F. L., 2009, Python 3 Reference Manual. CreateSpace, Scotts Valley, CA
- Vasiliev E., 2019, *MNRAS*, 482, 1525
- Virtanen P. et al., 2020, *Nat. Meth.*, 17, 261
- Wechsler R. H., Tinker J. L., 2018, *ARA&A*, 56, 435
- Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, *ApJ*, 827, L23
- White S. D. M., Rees M. J., 1978, *MNRAS*, 183, 341
- Wolf C. et al., 2018, *Publ. Astron. Soc. Aust.*, 35, e010
- Wu Y., Valluri M., Panithanpaisal N., Sanderson R. E., Freese K., Wetzel A., Sharma S., 2022, *MNRAS*, 509, 5882
- Yang C. et al., 2019, *ApJ*, 880, 65
- Yuan Z., Smith M. C., Xue X.-X., Li J., Liu C., Wang Y., Li L., Chang J., 2019, *ApJ*, 881, 164
- Zhao G., Zhao Y.-H., Chu Y.-Q., Jing Y.-P., Deng L.-C., 2012, *Res. Astron. Astrophys.*, 12, 723
- Zolotov A., Willman B., Brooks A. M., Governato F., Brook C. B., Hogg D. W., Quinn T., Stinson G., 2009, *ApJ*, 702, 1058

## SUPPORTING INFORMATION

Supplementary data are available at *MNRAS* online.

**Table A1.** Source ID of the clustered Gaia stars studied in this work.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors.

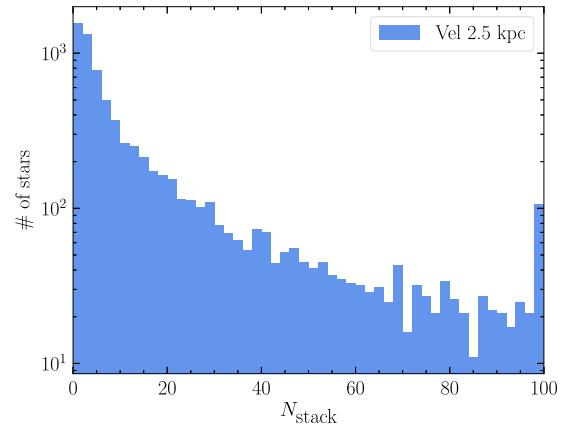
Any queries (other than missing material) should be directed to the corresponding author for the article.

## APPENDIX A: $N_{\text{stack}}$ CUTS AND ADDITIONAL CLUSTERING RESULTS

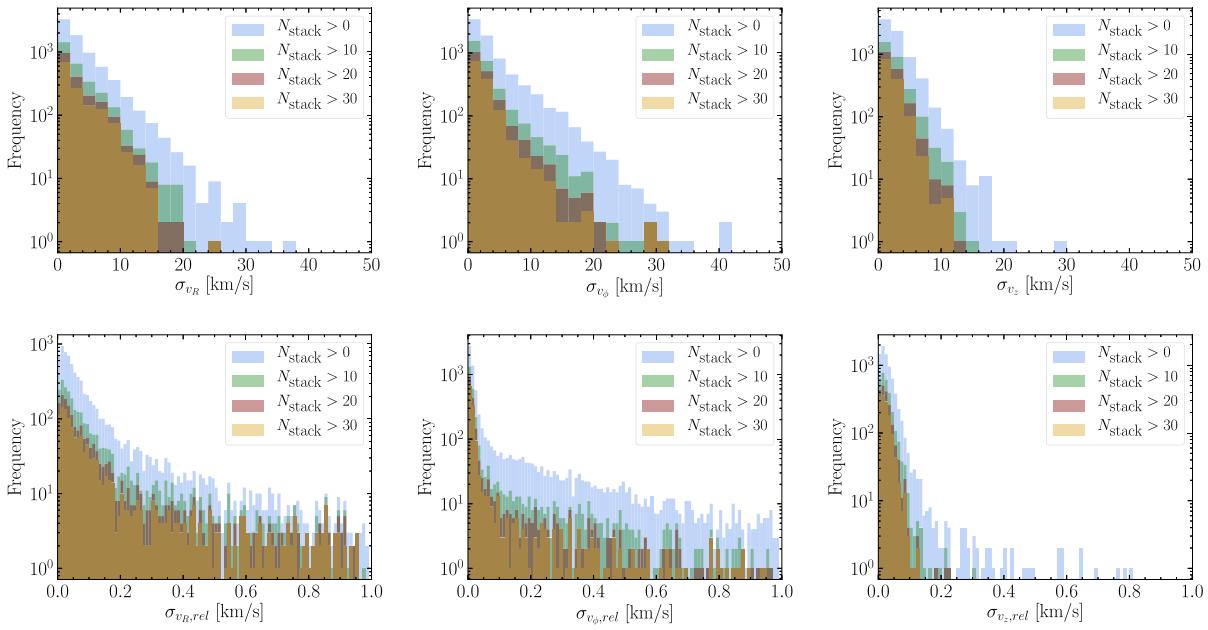
In this appendix, we show the figures associated with the  $N_{\text{stack}}$  cuts. Fig. A1 shows the number of stars associated with any cluster at  $N_{\text{stack}}$  number of realizations. Fig. A2 shows the error distribution of stars with various  $N_{\text{stack}}$  cuts. We also include Figs A3 and A4, showing the clustering result on the velocity and integrals of motion spaces respectively, without placing any  $N_{\text{stack}}$  cut on the stacked sample. As expected, they include a much larger number of clusters. We also include Fig. A5, the equivalent of Fig. 7 in integrals of motion space. Finally, Fig. A6 shows the resulting clustering, having removed cluster NGC 3201 (see related discussion in Section 5.3).

**Table A1.** Source ID of the clustered Gaia stars studied in this work. The complete version of Table A1 is available in the online edition of the journal.

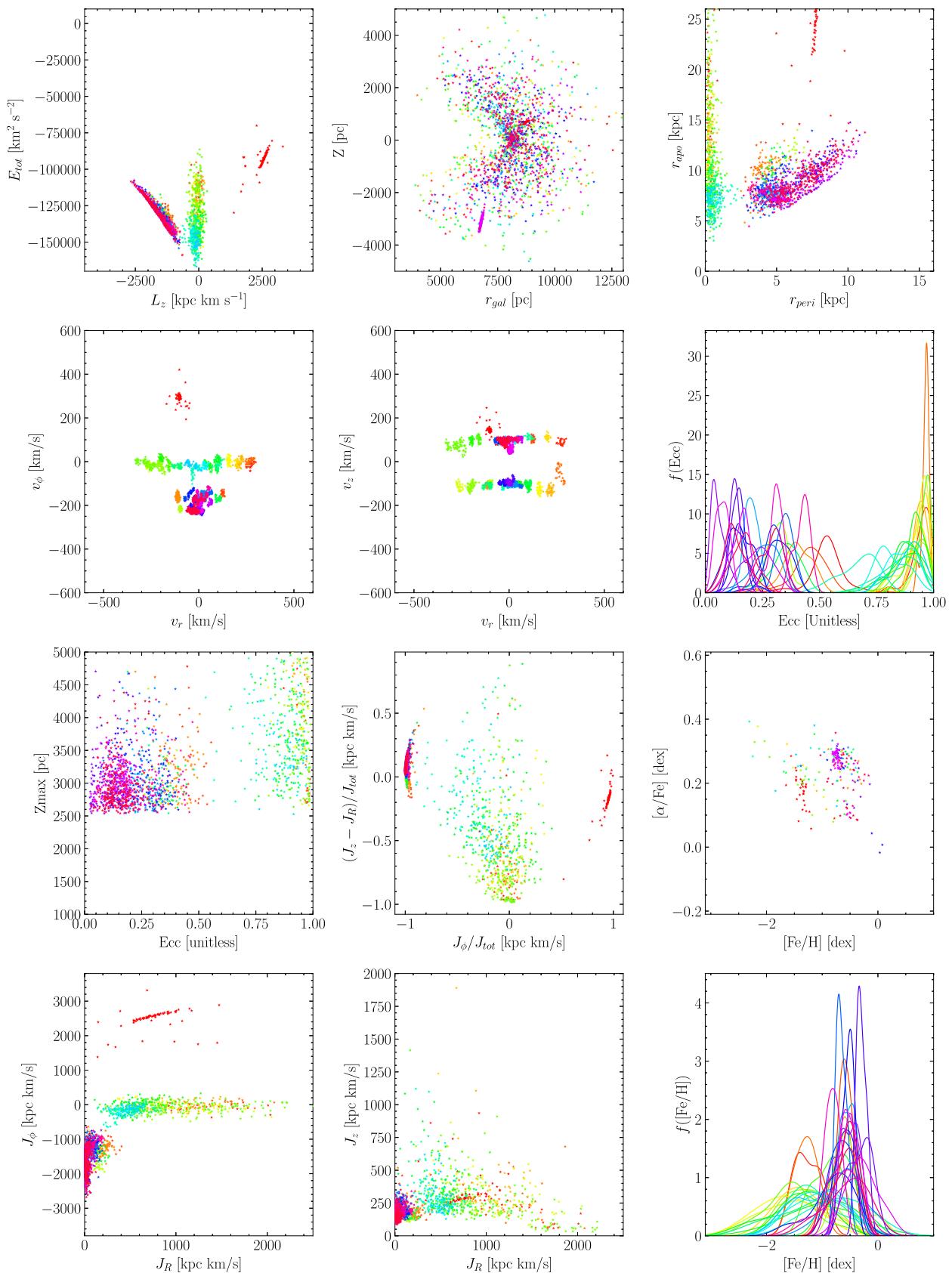
Source ID	Space	Cluster ID	Merged Cluster ID
634469512611911168	velocity	0	I
1277708416434956288	velocity	0	I
2798457105022823296	velocity	0	I
2831494371421184512	velocity	0	I
2910503176753011840	velocity	0	I



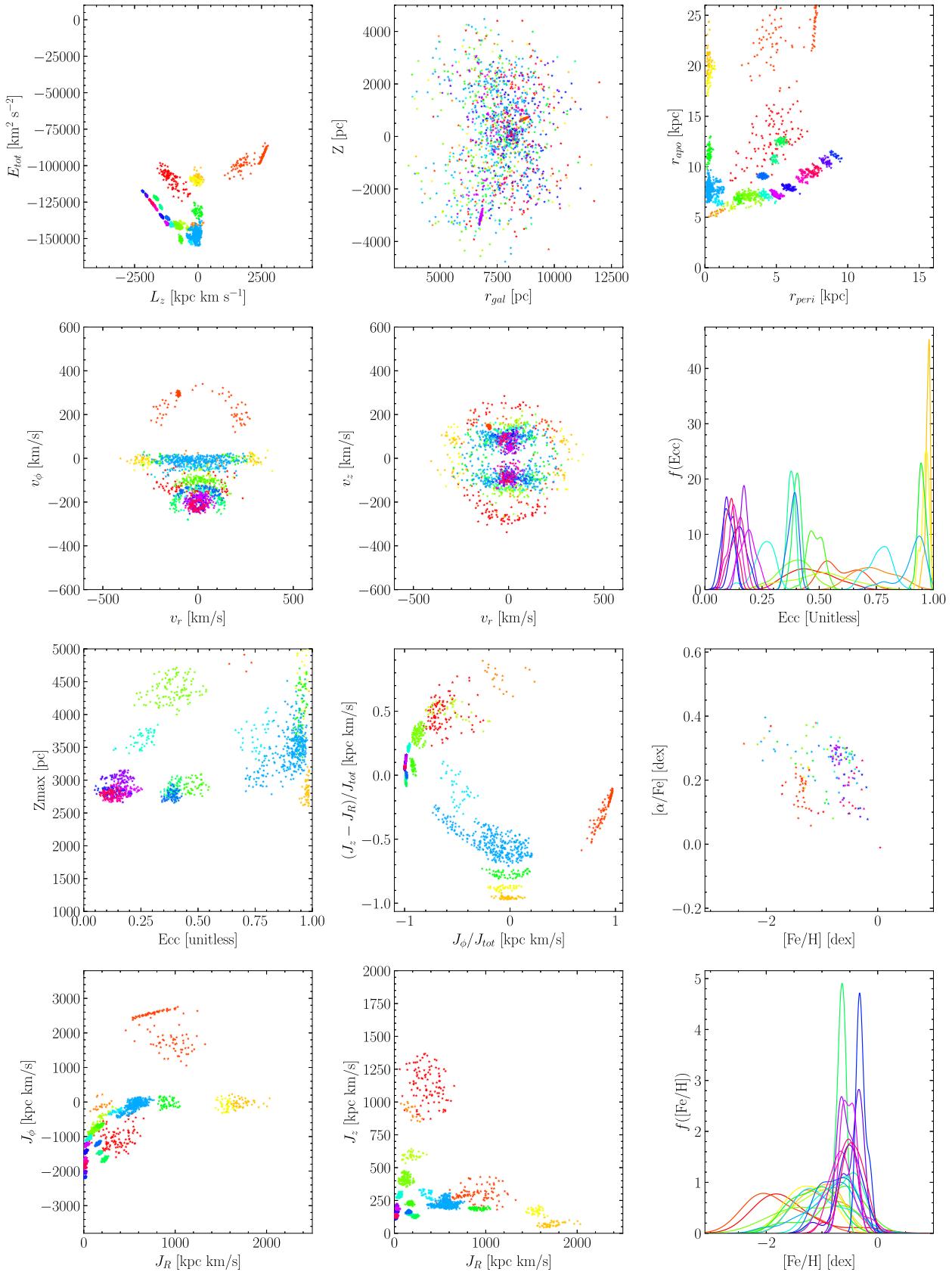
**Figure A1.** Distribution of  $N_{\text{stack}}$  for the stacked full subsample with  $Z_{\text{max}} 2\sigma$  above 2.5 kpc from the cylindrical velocity space.



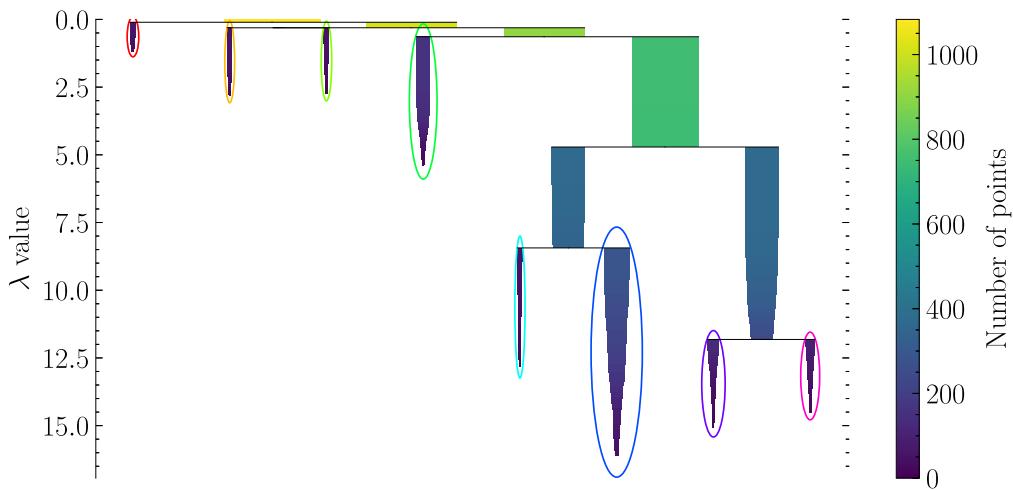
**Figure A2.** Absolute and relative error distribution of cylindrical velocities for the stacked full subsample with  $Z_{\text{max}} 2\sigma$  above 2.5 kpc at different  $N_{\text{stack}}$  cuts. Note that the y-axis is on logarithmic scale.



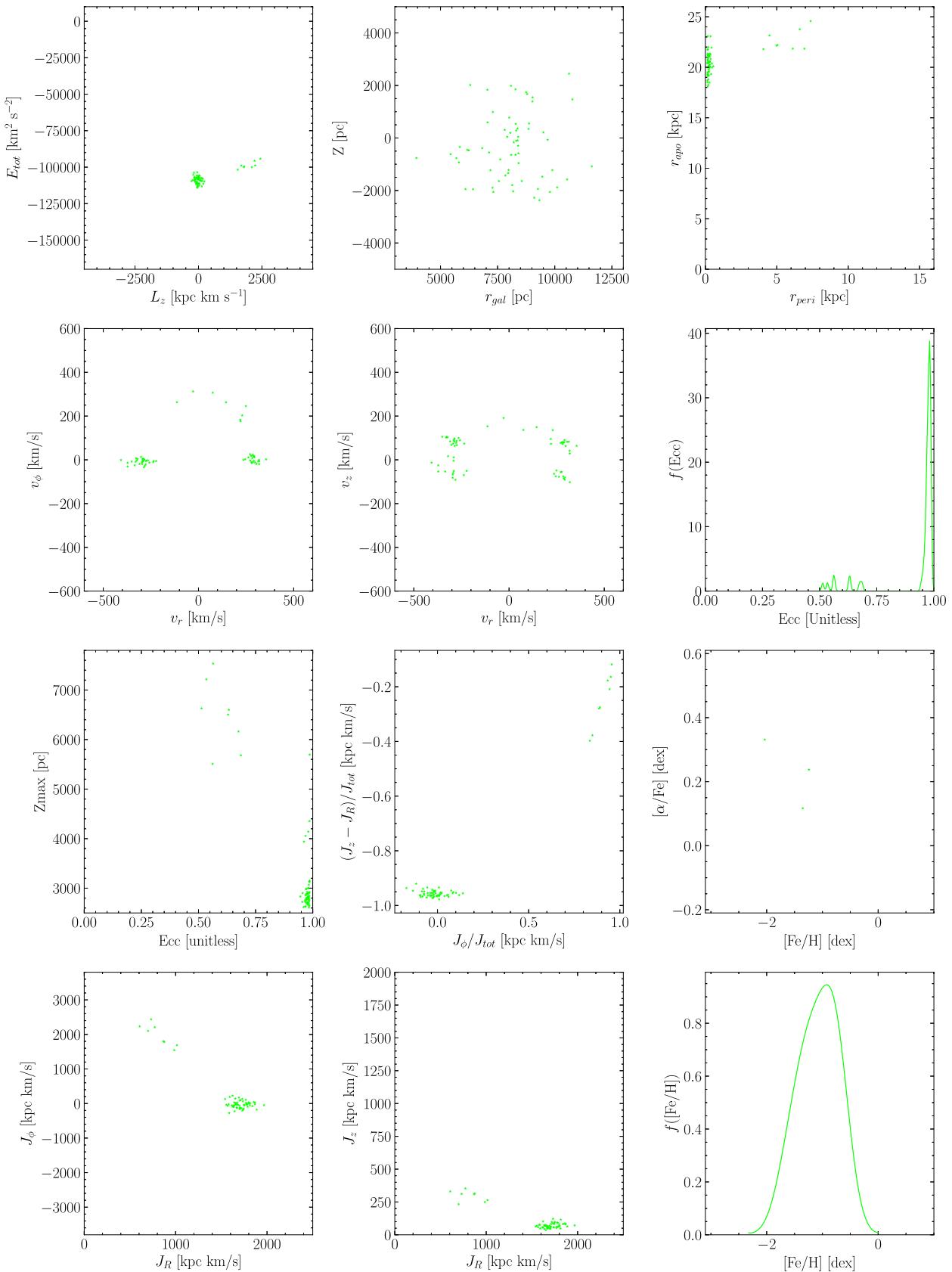
**Figure A3.** Summary plot for clusters that HDBSCAN identified from stacked full subsample with  $Z_{max}$   $2\sigma$  above 2.5 kpc from the cylindrical velocity space.



**Figure A4.** Summary plot for clusters that HDBSCAN identified from stacked full subsample with  $Z_{\max}$   $2\sigma$  above 2.5 kpc from the Energy +3D action space.



**Figure A5.** Condensed merger tree for the stacked subsample with  $N_{\text{stack}} > 20$  and  $Z_{\text{max}}$   $2\sigma$  above 2.5 kpc from the Energy +3D action space.



**Figure A6.** Summary plot for cluster 1 from HDBSCAN identified from stacked stable ( $N_{stack} > 20$ ) subsample with  $Z_{\max}$   $2\sigma$  above 2.5 kpc from the Energy + Action space with NGC 3201 removed.

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