Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **527,** 9767–9781 (2024) https://doi.org/10.1093/mnras/stad3817
Advance Access publication 2023 December 11

Exploring the *ex-situ* components within *Gaia* DR3

Zhuohan Li [1,2] Gang Zhao [1,2]★ Ruizhi Zhang [1,2] Xiang-Xiang Xue [1,3] Yuqin Chen [1,2] and João A. S. Amarante [4,5]†

[1]*CAS Key Laboratory of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, 100101 Beijing, PR China*
[2]*School of Astronomy and Space Science, University of Chinese Academy of Sciences, 100049 Beijing, PR China*
[3]*Institute for Frontiers in Astronomy and Astrophysics, Beijing Normal University, 102206 Beijing, PR China*
[4]*Institut de Ciencies del Cosmos (ICCUB), Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, E-08028 Barcelona, Spain*
[5]*Jeremiah Horrocks Institute, University of Central Lancashire, Preston, PR1 2HE, UK*

Accepted 2023 December 6. Received 2023 November 28; in original form 2023 September 29

ABSTRACT

The presence of *Gaia* DR3 provides a large sample of stars with complete 6D information, offering a fertile ground for the exploration of stellar objects that were accreted to the Milky Way through ancient merger events. In this study, we developed a deep learning methodology to identify *ex-situ* stars within the *Gaia* DR3 catalogue. After two phases of training, our neural network (NN) model was capable of performing binary classification of stars based on input data consisting of 3D position and velocity, as well as actions. From the target sample of 27 085 748 stars, our NN model managed to identify 160 146 *ex-situ* stars. The metallicity distribution suggests that this *ex-situ* sample comprises multiple components but appears to be predominated by the Gaia-Sausage-Enceladus (GSE). We identified member stars of the Magellanic Clouds, Sagittarius, and 20 globular clusters throughout our examination. Furthermore, an extensive group of member stars from GSE, Thamnos, Sequoia, Helmi streams, Wukong, and Pontus were meticulously selected, constituting an ideal sample for the comprehensive study of substructures. Finally, we conducted a preliminary estimation to determine the proportions of *ex-situ* stars in the thin disc, thick disc, and halo, which resulted in percentages of 0.1 per cent, 1.6 per cent, and 63.2 per cent, respectively. As the vertical height from the Galactic disc and distance from the Galactic centre increased, there was a corresponding upward trend in the *ex-situ* fraction of the target sample.

Key words: Galaxy: evolution – Galaxy: general – Galaxy: kinematics and dynamics – Galaxy: structure.

1 INTRODUCTION

The hierarchical merger hypothesis (White & Rees 1978) has long been a prominent theory in explaining the formation of galaxies. Within the framework of the Lambda cold dark matter (ΛCDM) cosmology and in light of current observations, a picture has been constructed in which large galaxies are formed through a process of hierarchical mergers with smaller satellite galaxies. For instance, the ongoing merger between the Milky Way (MW) and the Sagittarius dwarf spheroidal galaxy (Sgr; Ibata, Gilmore & Irwin 1994) provides direct observational evidence of the anticipated accretion events. Furthermore, remnants of ancient accretion events have also been observed in the form of kinematic substructures and stellar streams (e.g. Zhao, Zhao & Chen 2009; Xue et al. 2011; Zhao et al. 2015; Helmi 2020).

Facilitated by the availability of high-quality astrometric data for a large number of stars, the *Gaia* mission (Gaia Collaboration 2016a) has initiated a new era in the investigation of Galactic formation. One of the most notable substructures confirmed through studies based on *Gaia* data (Gaia Collaboration 2016b, 2018) is the Gaia-Sausage-Enceladus (GSE; Belokurov et al. 2018; Haywood et al. 2018; Helmi et al. 2018; Myeong et al. 2018). It is believed to be the remnant of the last major merger event experienced by the MW, which was completed approximately 10 Gyr ago (Gallart et al. 2019; Helmi 2020). Kinematic studies of other substructures, including Sequoia (Myeong et al. 2019), Thamnos (Koppelman et al. 2019b), and the Helmi streams (Helmi et al. 1999; Koppelman et al. 2019a) were also conducted contemporaneously.

On the other hand, spectroscopic surveys such as the RAdial Velocity Experiment (RAVE; Steinmetz et al. 2006), Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009), LAMOST survey (Zhao et al. 2006, 2012; Deng et al. 2012; Liu et al. 2013; Luo et al. 2015; Yan et al. 2022), Galactic Archaeology with HERMES (GALAH; De Silva et al. 2015), Apache Point Observatory Galactic Evolution Experiment (APOGEE; Majewski et al. 2017), and H3 survey (Conroy et al. 2019), provide us with insights into chemical abundances. The formation and chemical enrichment history of dwarf galaxies (DG) differs from that of the MW, and DGs of different masses exhibit distinct chemical evolution patterns. Based on chemical abundance data obtained from the H3 survey (Conroy et al. 2019), Naidu et al. (2020) identified both previously known and newly discovered structures, such as Aleph,

★ E-mail: gzhao@nao.cas.cn
† Visiting Fellow at UCLan.

Arjuna, I'itoi, and Wukong. Another chemo-dynamical study that employed data from the APOGEE DR16 (Ahumada et al. 2020) revealed evidence for the existence of a metal-poor stellar structure in the inner Galaxy (Horta et al. 2021), which was previously identified as the Kraken (Kruijssen et al. 2020) and may be a signature of the proto-MW (Belokurov & Kravtsov 2022; Rix et al. 2022). With the support of APOGEE DR17 (Accetta et al. 2022), Horta et al. (2023) further conducted a comprehensive investigation into the chemical patterns exhibited by various previously known substructures and provided an in-depth analysis of their origins. Moreover, chemically peculiar stars were examined in APOGEE data (e.g. Fernández-Trincado et al. 2019, 2022), offering further insights into the merger and evolutionary history of the MW.

The incorporation of chemical abundance significantly enhances the analysis of substructures. However, there is a substantial discrepancy in sample sizes between spectroscopic and astrometric surveys. The third data release of *Gaia* (Gaia Collaboration 2023) provided us with astrometric data for over 1.8 billion stars, of which more than 30 million have 6D phase space information. Concurrently, the number of sources observed by a single high-resolution spectroscopic survey is limited to the order of one million (Buder et al. 2021; Accetta et al. 2022). The LAMOST survey has achieved tens of millions of spectra, leading to numerous significant findings (e.g. Li et al. 2015; Li, Tan & Zhao 2018; Yan et al. 2018, 2021; Xing et al. 2019, 2023; Zhao & Chen 2021). However, as a median-to-low resolution survey, the preponderance of its data is confined to information on metal abundance and a limited number of elements. Given the current circumstances, it would be highly beneficial for the community to devise a data-driven methodology that solely relies on kinematic data to attain comparable selection outcomes as those obtained through the utilization of multi-element abundance information.

Data-driven approaches have already been utilized in researches endeavor to detect merger debris within the MW. In the pursuit of uncovering the aforementioned substructures, several classic machine learning algorithms have been employed, including KNN (Fix & Hodges 1951), GMM (Dempster, Laird & Rubin 1977), and HDBSCAN (Campello, Moulavi & Sander 2013). Recent efforts to identify stellar streams and kinematic substructures have also incorporated the use of machine learning algorithms. By performing clustering using the DBSCAN algorithm (Ester et al. 1996) within a 6D phase space, Borsato, Martell & Simpson (2020) were able to detect five high-confidence streams, including one that had not been previously discovered. Yuan et al. (2020) employed a self-organizing map algorithm (Kohonen 2001; Yuan et al. 2018) within a 4D space of orbital energy and angular momentum to identify 57 dynamically tagged groups, most of which belong to previously known substructures such as the GSE. Building upon the ANODE algorithm (Nachman & Shih 2020), Shih et al. (2022) developed a novel methodology, termed the VIA MACHINE, for the identification of cold stellar streams within data acquired from the *Gaia*. Through the implementation of this approach, they were able to successfully identify the presence of the GD-1 stream (Grillmair 2006). By conducting t-Distributed Stochastic Neighbor Embedding (t-SNE; Van der Maaten & Hinton 2008) analysis in chemical space, Ortigoza-Urdaneta et al. (2023) identified seven structures, including the Splash, GSE, the high-α heated-disc population, N-C-O peculiar stars, inner disc-like stars, and two previously unreported structures. Unsupervised learning algorithms do not necessitate prior labeling of data. Instead, the algorithms make decisions based on the inherent distribution of the data. Nevertheless, the absence of a labeling process makes it challenging to artificially control the behavior of unsupervised algorithms and to quantify their performance. Consequently, it cannot be guaranteed that structures with clustering properties in phase space are related to ancient merger events. These structures may also originate from the disruption of local globular clusters (GCs) in the MW. Therefore, it is important to ensure a pure *ex-situ* sample before using clustering algorithms for detailed analysis.

Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction (LeCun, Bengio & Hinton 2015), it holds considerable promise as a tool in the search for *ex-situ* components. Through the implementation of supervised deep learning techniques on a known data set, an optimal model can be derived for the purpose of classification. Ostdiek et al. (2020) implemented an neural network (NN) trained to identify accreted stars using only 5D kinematics information and construct a catalogue comprising 767 000 accreted stars. Subsequent companion works revealed the existence of a vast prograde stellar stream in the solar vicinity, which they named Nyx (Necib et al. 2020a, b). The methodology presented by Ostdiek et al. (2020) provides an instructive paradigm for the application of supervised learning to the identification of merger debris. By pre-training the model on simulated data, it is able to classify accreted stars labeled at truth level and capture their major kinematic features. The employment of transfer learning further improved the model's ability to adapt to real data of the MW. However, the absence of radial velocity limits the classification to kinematic level, as the derivation of integrals of motion requires full 6D phase space information. With the availability of *Gaia* DR3 data and the inclusion of photo-astrometric distances derived by the `StarHorse` code (Queiroz et al. 2018; Anders et al. 2022), the sample size of stars with accurately determined full 6D phase space parameters is no longer a constraining factor. In addition, recent data releases of spectroscopic surveys have facilitated the tagging of accreted stars in a more efficient manner, thereby enabling the improvement of the second phase of network training based on previous research.

In this study, we present a deep learning methodology for the identification of *ex-situ* stars through the utilization of 6D phase space parameters and the actions. Following the completion of two training phases, our NN model demonstrated the ability to effectively identify *ex-situ* components comprised of DGs, GCs, and merger debris within the target sample of *Gaia* DR3. We conducted a preliminary analysis of the NN classified sample and dedicate it to the community in the hope that it could provide novel insights and inspiration. Further analysis, including clustering and chemical follow-up, may be performed to fully explore the potential of this sample and deepen our understanding of the MW's merger history.

This paper is organized as follows: Section 2 introduces the NN model, the training process, and the data utilized in this study. In Section 3, we present the classification results of the NN and provide an initial analysis of the *ex-situ* sample. Section 4 discusses particular elements involved in our algorithm and analysis. Finally, we summarize the results of this study in Section 5.

2 METHOD

2.1 Base model

As in Ostdiek et al. (2020), we constructed an NN model that was trained on synthetic data to learn fundamental kinematic features. The NN model was developed and trained using the `Keras` (Chollet et al. 2015) library, with `TensorFlow` (Abadi et al. 2015) serving as the backend framework. The input to the NN consists of 3D position and velocity data in the Cartesian coordinate system. The model

comprises five fully connected hidden layers, with the intermediate three layers containing 128 nodes and the remaining two layers containing 64 nodes. We selected the Rectified Linear Unit (ReLU; Hahnloser et al. 2000) as the activation function, and applied batch normalization prior to each activation. The output layer of the NN is composed of a single node that uses a sigmoid activation function. This function scales the output to fall within the range of 0–1, where values closer to 1 indicate a higher probability of the star being *ex-situ*. We refer to this NN model as `NN_FIRE` in subsequent discussions.

2.1.1 Synthetic data

As a supervised learning algorithm, the performance of the NN is greatly influenced by the labels of the training set. To obtain truth level *ex-situ* labels, we train our NN with simulated data, where the accretion history of each star is traceable. The state-of-the-art cosmological hydrodynamic simulations, specifically FIRE-2 (Hopkins et al. 2018; Wetzel et al. 2023), provide us a robust platform for our training endeavors. Rather than directly employing the raw data of FIRE-2 simulations, we opted to utilize a mock *Gaia* catalogue constructed by Sanderson et al. (2020) to facilitate the alignment between the patterns of the simulated data and the observational data from *Gaia* satellite.

Detailed elaboration of the synthetic catalogue is presented in Sanderson et al. (2020), and reviewed by Ostdiek et al. (2020). The mock catalogues are derived from three MW-mass galaxies from the *Latte* suite of FIRE-2 simulations, namely `m12i`, `m12f`, and `m12m`. Assuming each star particle of mass $\approx 7000\,M_{\odot}$ represents a single stellar population, Sanderson et al. (2020) sampled synthetic stars following the algorithm described in Sharma et al. (2011). Three different local standard of rest (LSR) were deployed in each simulated galaxy as 'solar viewpoints' to construct the synthetic catalogue. The data structure of the synthetic catalogue resembles that of the *Gaia* catalogue, which supplies astrometric and photometric data. As in Ostdiek et al. (2020), we utilized data from all three LSRs in the `m12i`, encompassing a total of 47 673 267 synthetic stars. We only selected stars with available error-convolved radial velocity, which covers a distance range of 0–3 kpc. The data utilized included error-convolved right ascension (`ra`), declination (`dec`), proper motion in right ascension (`pmra`), proper motion in declination (`pmdec`), and radial velocity (`radial_velocity`). In addition, the true LSR-centric distance (`dhel_true`) was used instead of the error-convolved parallax (`parallax`) to ensure compatibility with the final model. The astrometric data were then transformed into 3D position and velocity in the Galactocentric Cartesian coordinate system using the `astropy` package (Price-Whelan et al. 2018), as the input of the NN.

Although the synthetic catalogue is build to mock *Gaia* DR2, we note that it is not necessary to reproduce the exact selection function of the latest data release of *Gaia*. The primary objective of `NN_FIRE` is to acquire fundamental knowledge of stellar kinematics. The inconsistency between the distribution of training samples and target samples will be resolved in the subsequent training process, as described in Section 2.2.

To identify the *ex-situ* stars in the synthetic catalogue, we made use of a text file named `star_exsitu_flag_600.txt` in the `m12i` directory (Wetzel et al. 2023), which provides a binary flag for each star particle at redshift equals to 0. A flag value of 1 indicates that the star particle is *ex-situ*, while a value of 0 indicates that it is *in-situ*. Star particles are classified as '*ex-situ*' if they originated at a spherical distance greater than 30 kpc comoving from the centre of the primary galaxy (Bellardini et al. 2022). Synthetic stars generated from the same star particle are assigned the same `parentid`, which is the array index of the star particle. We labeled each star with the binary flag corresponding to the index in the text file, resulting in a total of 380 326 *ex-situ* stars.

2.1.2 Training and evaluation

Our training process adheres to the standard workflow within the `Keras` framework. We employed the `Adam` optimizer (Kingma & Ba 2014), with an initial learning rate of 10^{-3}. A learning rate reduction mechanism was implemented, whereby if the validation loss failed to decrease for five consecutive epochs, the learning rate would be automatically reduced to half of its original value. The minimum learning rate threshold was set at 10^{-6}. Furthermore, to avoid overfitting, the training process would be terminated if the validation loss failed to decrease for 10 epochs. However, this early stopping mechanism was not triggered during a total of 50 epochs of training.

The synthetic stars in our data set were partitioned into training, validation, and test sets at a ratio of 6:2:2. The 6D phase space data in these three data sets were normalized using a `StandardScaler` from the `sklearn.preprocessing` module (Pedregosa et al. 2011), which was fit to the training set. The partitioning process was accomplished using the `sklearn.train_test_split` function, during which we employed stratified sampling on the label to ensure that the proportion of *ex-situ* stars remained consistent across all data sets. However, *ex-situ* stars comprise a very small fraction of our synthetic sample, which presents a classification problem with an extreme class imbalance. One common method to address class imbalance is sample-weighted training, but determining the appropriate weight can be challenging. If the weight is set too high, it can skew the distribution of the prediction and result in poor classification performance. Conversely, if the weight is set too low, it may not effectively address the issue of imbalanced sample size. Therefore, we aim to solve this problem by choosing a proper loss function. The `Focal Loss` is a loss function introduced by Lin et al. (2017) to mitigate the issue of class imbalance during training in tasks such as binary classification. This function applies a modulating term to the cross-entropy loss to focus on the samples that are challenging to classify. The function is mathematically defined as

$$FL(p) = \begin{cases} -(1-p)^{\gamma}\log(p) & \text{if y = 1} \\ -p^{\gamma}\log(1-p) & \text{if y = 0} \end{cases} \quad (1)$$

where p denotes the output of the classifier and γ represents a tunable hyper parameter. In our training process, we adopted the default setting of $\gamma = 2$. As the confidence in the correct class increases, the scaling factor decays to zero, automatically down-weighting the contribution of samples that are easier to classify during training and rapidly focusing the model on more challenging samples.

Due to the extreme class imbalance in our classification problem, the commonly used metric of accuracy is not a suitable measure of NN classification performance. Accuracy reflects the proportion of samples that are correctly classified within the entire sample. However, in our sample, less than 1 per cent of the stars are labeled as *ex-situ*. As a result, a classifier that indiscriminately assigns all stars as *in-situ* could easily achieve an accuracy rate exceeding 99 per cent, even in the absence of any classification activity.

In order to adopt a more appropriate metric, we initially consider *ex-situ* stars as positive samples while *in-situ* stars the negative

samples. With respect to the classification results, *ex-situ* samples that are accurately identified by the NN are defined as true positive samples (TP). In contrast, *ex-situ* samples that are erroneously classified as *in-situ* by the NN are defined as false negative samples (FN). Correspondingly, *in-situ* samples that are accurately identified by the NN are regarded as true negative samples (TN), while *in-situ* samples that are erroneously classified as *ex-situ* by the NN are regarded as false positive samples (FP). To evaluate the purity of *ex-situ* samples selected by the NN, we employed the metric of precision,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \tag{2}$$

Furthermore, the true positive rate (TPR), commonly referred to as recall, illustrates the completeness of the selection

$$TPR = \frac{\text{TP}}{\text{TP} + \text{FN}}. \tag{3}$$

Additionally, we calculated the false positive rate (FPR) to quantify the proportion of *in-situ* samples that were incorrectly classified as *ex-situ*,

$$FPR = \frac{\text{FP}}{\text{TN} + \text{FP}}. \tag{4}$$

Precision and recall are often considered to be conflicting measures, making it challenging for a classifier to achieve high values for both simultaneously. In the task of identifying *ex-situ* stars, our goal is to ensure that the *ex-situ* samples selected by the NN have a high degree of purity, which corresponds to a higher precision. Similarly, we also strive to minimize the FPR. However, this may result in sacrificing the completeness of the *ex-situ* sample, leading to a relatively lower recall.

The output of an NN consists of a series of continuous values. In order to map these values to discrete labels, it is necessary to establish a threshold. The final classification result can be expressed as

$$\text{Prediction} = \begin{cases} 1 & \text{if } p > \text{threshold} \\ 0 & \text{if } p \leq \text{threshold} \end{cases} \tag{5}$$

where p represents the raw output of the NN. The choice of threshold can have a significant impact on the performance of the classifier, as different thresholds will result in varying levels of TPR and FPR. By selecting a series of thresholds, a corresponding series of TPRs and FPRs can be obtained, allowing for the construction of a receiver operating characteristic (ROC) curve. According to the performance of NN_FIRE on the test set, we present the ROC curve as shown in Fig. 1. The area under the curve (AUC) is a quantitative measure of classifier performance, with values ranging from 0 to 1. An AUC value closer to 1 indicates superior classifier performance. In this case, NN_FIRE obtained an AUC value exceeding 0.98, denoting an excellent result. As represented by the gold star, an ideal classifier would achieve a TPR of 1 and a FPR of 0, corresponding to a precision of 100 per cent. However, as previously discussed, this is not achievable in practice and a trade-off must be made between the purity and completeness of the *ex-situ* sample. When the threshold was established at 0.5, NN_FIRE had a precision of 80.1 per cent and a recall of 34.5 per cent. Elevating the threshold to 0.75 yielded a precision of 98.3 per cent, but diminished the recall to 13.4 per cent. As NN_FIRE is not the final model, we will refrain from discussing the threshold setting in detail at present. Based on the analysis of the ROC curve, we consider that the initial training phase has imbued NN_FIRE with the ability to identify *ex-situ* stars from a kinematic perspective.

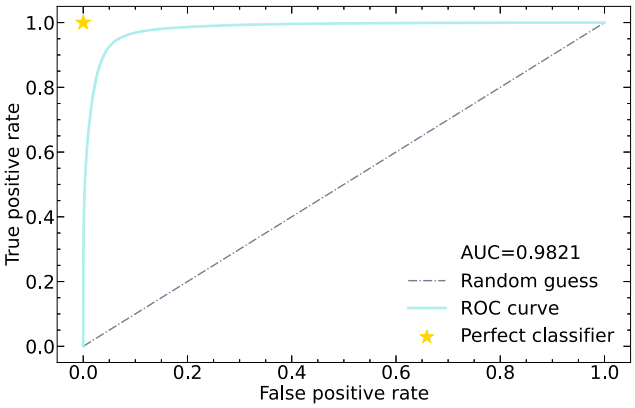

Figure 1. As demonstrated by the ROC curve, the NN_FIRE achieves an AUC of 0.98 on the test set. The performance of a perfect classifier, indicated by a star, has a TPR of 1 and an FPR of 0. The dashed line represents the performance of a random guess.

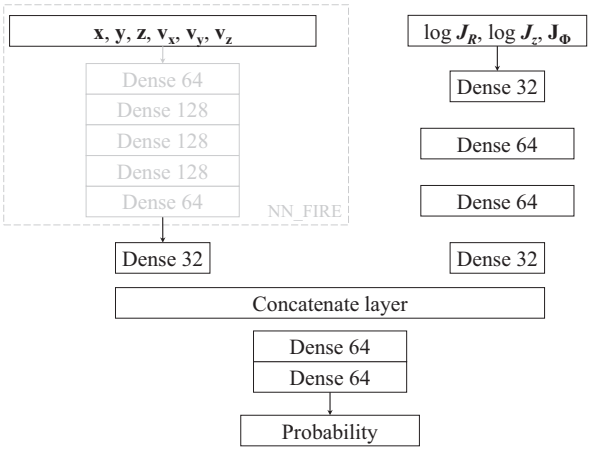

Figure 2. The parallel model comprises two components: the NN_FIRE network and a sub-network that processes actions. The outputs of these two components are concatenated and subsequently processed by a series of dense layers. The weighted layers of the NN_FIRE network are frozen, as indicated by their grey coloration.

2.2 Parallel model

In the context of NN regression tasks, the distribution of labels within the training set warrants substantial attention. Predictions outside this range or in sparse parts of the distribution are regarded as extrapolations of the model. Unlike regression tasks where label values are continuous, our classification task has only two categories, corresponding to label values 0 and 1. In this case, caution should still be exercised regarding extrapolation, but the focus should be on the distribution of NN input parameters. During the training phase of NN_FIRE, only sources with distances less than 3 kpc are considered. However, the target sample may encompass distances reaching tens of kiloparsecs. Moreover, even if the simulated data is derived from a synthetic *Gaia* survey catalogue, it may not precisely align with observational data. Therefore, we need to design a training framework to improve our model's performance on the target sample.

Instead of performing the standard workflow of transfer learning as in Ostdiek et al. (2020), we opted for a more innovative approach. Based on NN_FIRE, a parallel model was constructed as shown in Fig. 2. The training methodology employed for this model

was similar to that of the base model, but with a shorter training period. Furthermore, the `StandardScaler` was supplanted by a `RobustScaler` to facilitate improved normalization of the realistic data. All the weights of the base model are frozen during the training process and its output is concatenated with the output of a sub-network that takes J_R, J_z and J_ϕ as input, which denotes the radial, vertical, and azimuthal actions, respectively (e.g. Binney 2012; Sanders & Binney 2016). To enhance the normalization process, the logarithm of J_R and J_z was taken. The concatenated output will be further processed by the rest two dense layers, finally giving the classification result. The inputs of the sub-network introduce the potential of the MW into the parallel model and transform the model from a purely kinematic-based classifier to one that is based on dynamics. In this way, the parallel structure not only solves the extrapolation problem to a certain extent, but also maximizes the utilization of all available information of the 6D parameters. We denote this model as `NN_parallel` in the following text.

2.2.1 *Observational data*

The third data release of *Gaia* (Gaia Collaboration 2023) provided radial velocity measurement for over 30 million stars, thereby presenting an ideal target sample for our study. To overcome the limitation of parallax measurement, we employed the photo-astrometric distance estimated by Anders et al. (2022). We adopted stars with available radial velocity and photo-astrometric distance within *Gaia* DR3, and imposed certain restrictions to ensure the quality of the data, which include a renormalized unit weight error $<$ 1.4 (e.g. Lindegren et al. 2021), a radial velocity error $<$ 20 km s^{-1}, and a distance error $<$ 30 per cent.

In order to derive robust dynamic parameters, we employ a Monte Carlo methodology. For each observational parameter, we perform 100 times of random sampling, which follows a Gaussian distribution with the observed quantity serving as the mean and the error of the observed quantity as the standard deviation. In this way, each star is represented by 100 sets of Monte Carlo observations, which are then transformed into Cartesian positions and velocities. We apply a right-hand Galactocentric Cartesian coordinate system (x, y, z), where the x-axis is oriented from the Sun towards the Galactic centre, the y-axis aligns with the direction of the MW's rotation, and the z-axis points towards the North Galactic Pole. We assume the circular speed at the Sun as 232.8 km s^{-1} and the Solar Galactocentric distance as 8.2 kpc (McMillan 2017). The distance from the Sun to the Galactic plane is set at 20.8 pc (Bennett & Bovy 2019). For the solar peculiar motion, we adopt the value reported by Schönrich, Binney & Dehnen (2010), yielding (U, V, W) = (11.1, 12.24, 7.25) km s^{-1}. Subsequently, we calculate the corresponding actions and orbital parameters using `AGAMA` (Vasiliev 2019), with the assumption of McMillan (2017) potential and the application of `Stäckel` approximation as described by Binney (2012). The median values of the dynamic parameters are adopted as the final estimation.

It is important to note that in the region behind the Galactic centre and close to the Galactic plane, the distances of a small number of stars may be overestimated. Since Anders et al. (2022) utilized the extinction curve provided by Schlafly et al. (2016), whose fitting quality for the inner Galaxy is relatively inferior (see fig. 10 in Schlafly et al. 2016), we consider that this overestimation could potentially be attributed to extinction issues. Consequently, we remove stars in the region defined by: $x > 0$ kpc, $z >$ -3 kpc, $r_{gc} >$ 6 kpc, -50° $< l <$ 50° and -10° $< b <$ 15°. Here, r_{gc} denotes the Galactocentric distance, while l and b represent the Galactic longitude and Galactic latitude, respectively. Given the asymmetric distribution of the stars whose distances may be overestimated, we adopt an asymmetric criterion. Furthermore, we exclude stars with positive total energy, leaving a target sample of 27 085 748 stars, which is comparable in magnitude to the sample size of stars with 5D phase space parameters in Ostdiek et al. (2020).

2.2.2 *Chemical tagging of the training samples*

For the purpose of training the `NN_parallel` model to be applicable to the target sample, it is necessary to construct a new data set based on observational data. The primary challenge in this process is determining an appropriate method for labeling the observational data. Ostdiek et al. (2020) implemented a selection criterion, referred to as the ZM method (e.g. Herzog-Arbeitman et al. 2018), where stars with [Fe/H] <-1.5 dex and $|z| > 1.5$ kpc were labeled as *ex-situ*. Nevertheless, given that the input of our NN consists solely of dynamic parameters, it is desirable to label *ex-situ* and *in-situ* stars using a criterion that is independent of dynamics, which ensures that the sample remains dynamically unbiased. Thus the measurement of chemical abundance is considered to be the ideal criterion for the tagging of *ex-situ* components in observations, as it offers an independent perspective, and serves as a reliable indicator of stellar origin.

In pursuit of a large sample of stars with detailed chemical abundances, we cross-matched the *Gaia* DR3 target sample with APOGEE DR17 catalogue (Accetta et al. 2022). Adhering to the online documentation of APOGEE DR17 and Holtzman et al. (2015), we only adopted stars that satisfy the following criteria:

(i) No BAD bit in `STARFLAG` or `ASPCAPFLAG`,
(ii) 3500 K $< T_{\rm eff} <$ 6000 K,
(iii) S/N $\geq$ 70,
(iv) VSCATTER $<$ 1 km s^{-1},
(v) `X_Fe_FLAG` == 0 and `X_FE_ERR` $<$ 0.15 dex.

Upon applying these selection parameters, a total of 254 046 stars remained. Additionally, we utilized the value-added catalogue (VAC) of LAMOST DR8 presented by Li et al. (2022), which contains abundance information for 10 elements. The elemental abundances were estimated by an NN trained on APOGEE DR17 data and show high consistency with APOGEE measurements. Therefore, we further adopted 562 626 stars with `X_FE_ERR` $<$ 0.10 dex in the VAC that are in common with our target sample as a supplement to the data set, resulting in a final sample size of 816 672 stars.

As demonstrated in Hawkins et al. (2015), the Galactic components can be labeled using [Al/Fe] and [Mg/Mn]. Magnesium is the first element to be affected by Type II supernovae, while manganese is produced in higher fractions compared to iron in Type Ia supernovae. Aluminium, on the other hand, is produced by Type II supernovae and is sensitive to the initial abundance of carbon and nitrogen, which originate from helium burning or asymptotic giant branch evolution. As a result, the [Mg/Mn] and [Al/Fe] distributions of *ex-situ* stars are expected to differ from those of stars formed locally in the MW due to differences in their star formation rate and chemical evolution history. Based on APOGEE DR14 (Abolfathi et al. 2018), Das, Hawkins & Jofré (2020) introduced the [Mg/Mn]-[Al/Fe] diagram and discovered a'blob' of accreted stars located in the region of high [Mg/Mn] and low [Al/Fe], which is distinct from the stellar disc. Horta et al. (2023) and Ortigoza-Urdaneta et al. (2023) also characterized numerous substructures distributed within this region, including the GSE, Sequoia, and Heracles. Furthermore, recent studies have attempted to extract the *ex-situ* components in

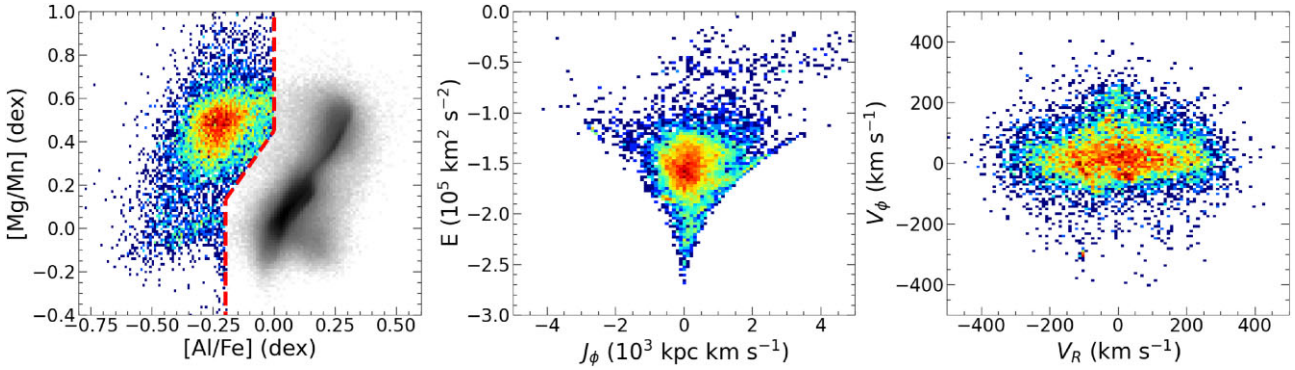

Figure 3. As depicted by the red dashed line in the [Mg/Mn]-[Al/Fe] plane, our criterion divides the plane into two distinct regions. The region to the left of the line is predominantly occupied by *ex-situ* stars, while the region to the right is primarily composed of *in-situ* stars. The middle and right subplots illustrate the distribution of stars labeled as *ex-situ* in $E - J_\Phi$ space and velocity space, respectively. Each subplot is colour-coded to represent the number density of stars, with warmer colours indicating higher densities for *ex-situ* stars and darker colours indicating higher densities for *in-situ* stars.

the [Mg/Mn]-[Al/Fe] diagram (e.g. Horta et al. 2021; Carrillo et al. 2024; Feltzing & Feuillet 2023). However, most of these efforts have focused primarily on the GSE region. As noted by Hasselquist et al. (2021) and Fernandes et al. (2023), satellite galaxies of the MW such as the Large Magellanic Cloud (LMC), the Small Magellanic Cloud (SMC), and the Sgr exhibit relatively low [Mg/Mn] values. With the aim of uniformly distinguishing between the *ex-situ* and *in-situ* components, it may be more effective to select a region that encompasses as many DGs and substructures as possible, rather than focusing on a specific accretion component such as the GSE.

We performed a segmental analysis of our training samples, as depicted in Fig. 3. Stars were labeled as *ex-situ* if they satisfy the following selection criteria:

$$\begin{cases} [\mathrm{Al/Fe}] < 0.00 & \text{if } [\mathrm{Mg/Mn}] \geq 0.45 \\ [\mathrm{Mg/Mn}] > 1.6[\mathrm{Al/Fe}] + 0.45 & \text{if } 0.13 \leq [\mathrm{Mg/Mn}] < 0.45 \\ [\mathrm{Al/Fe}] < -0.20 & \text{if } [\mathrm{Mg/Mn}] < 0.13 \end{cases} \quad (6)$$

According to these criteria, 11 478 *ex-situ* stars were selected in our data set. This selection includes member stars of the MW's satellite galaxies and GC debris (e.g. Fernández-Trincado et al. 2022). Inevitably, it also encompasses a minor fraction of stars associated with the Galactic disc. As shown in Fig. 3, there is a group of stars with a v_Φ of approximately 230 km s^{-1}, whose distribution on the $E - J_\Phi$ plane closely resembles that of the disc. Fernandes et al. (2023) showed that these disc stars occupy the region of low [Al/Fe] and low [Mg/Mn], which overlaps with the *ex-situ* region as defined by equation (6). However, a criterion stringent enough to exclude these disc stars would also fail to capture the region where the majority of DGs reside. Upon careful evaluation, we determined that this classification method is reasonable and the pollution of disc stars will not have a significant impact on the training of `NN_parallel`.

2.2.3 *Model performance*

Since the weights of the base model are frozen, the chemical criterion turns out to be sufficient and robust enough to support the subsequent training of the `NN_parallel`. During the training phase, the training samples were divided into training set, validation set, and test set as in Section 2.1.2. On the test set, `NN_parallel` achieved an AUC of 0.98, which is comparable to that of the `NN_FIRE`. Employing a default threshold value of 0.5, the NN attained a precision of 81.0 per cent and a recall of 64.6 per cent, which already signified an equilibrium between purity and completeness. This threshold will be utilized in subsequent analysis. Nevertheless, it is imperative to note that the metric values should be regarded as intuitive references only, as they are not predicated on truth-level labels as explicated in Section 2.1.2.

Independent of metrics, we conducted a more comprehensive analysis of the classification results given by `NN_parallel`. As depicted in Fig. 4, the selection made by the parallel model does not precisely mirror the pattern of the label. The `NN_parallel` avoids selecting the disc stars that was incorrectly labeled as *ex-situ* through chemical tagging. Conversely, stars selected by the NN but labeled as *in-situ* through chemical tagging exhibit chemical patterns and dynamic characteristics consistent with those of the thick disc and the GSE. Since the catalogue in Sanderson et al. (2020) did not contain information about [Al/Fe] or [Mn/Fe], we were not able to reproduce the workflow of the second training phase on the simulation. However, based on the results depicted in Fig. 4, it appears that `NN_parallel` may offer an even more appropriate classification result than chemical tagging. A similar situation arose in Ostdiek et al. (2020), who randomly selected 200 000 synthetic stars from m12f and labeled them using the traditional ZM selection method. They then performed transfer learning on an NN pre-trained on m12i using this data set. After transfer learning, the NN achieved a precision of 41 per cent and a recall of 47 per cent when the threshold was set at 0.75. When the threshold was raised to 0.95, the precision increased to 59 per cent while the recall decreased to 13 per cent. In comparison, the ZM selection method only achieved a precision of 50.9 per cent and a recall of 2.4 per cent. Our second training phase, like transfer learning, is designed to retain the results of the first training phase while adapting the model to observational data. The training process enabled the parallel model to overcome the limitations of both the base model and chemical tagging, resulting in improved performance. So far, the `NN_parallel` model is well-prepared for the classification task of the target sample.

3 RESULT

After two phases of training, the `NN_parallel` model was applied to the target sample from *Gaia* DR3, resulting in the identification of 160 146 *ex-situ* stars. As illustrated in Fig. 5, the region of the stellar disc was virtually eliminated, with the majority of the *ex-situ* stars being diffusely distributed in the rest of the plane. Specifically,

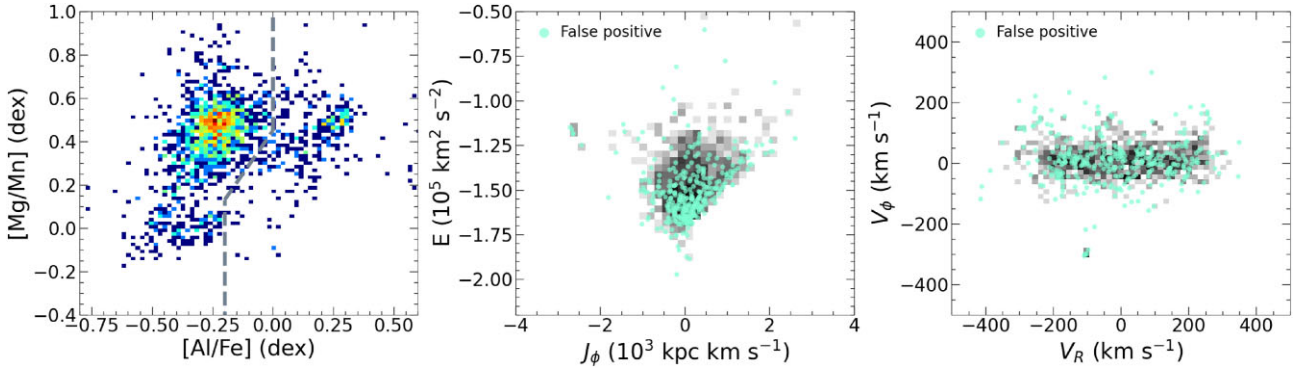

Figure 4. *Ex-situ* stars selected by the `NN_parallel` model in the test set. The left panel displays the distribution in the [Mg/Mn]-[Al/Fe] diagram, with the selection criterion indicated by a grey dashed line. Samples located in the right part are considered false positive samples with respect to the chemical tagged label. The middle and right subplots show the distribution in $E - J_\Phi$ space and velocity space, respectively. The background of these two subplots represents the number density of *ex-situ* stars selected by the NN, with darker colours indicating higher densities. False positive samples are highlighted for emphasis, constituting 0.2 per cent of the samples in the test set.

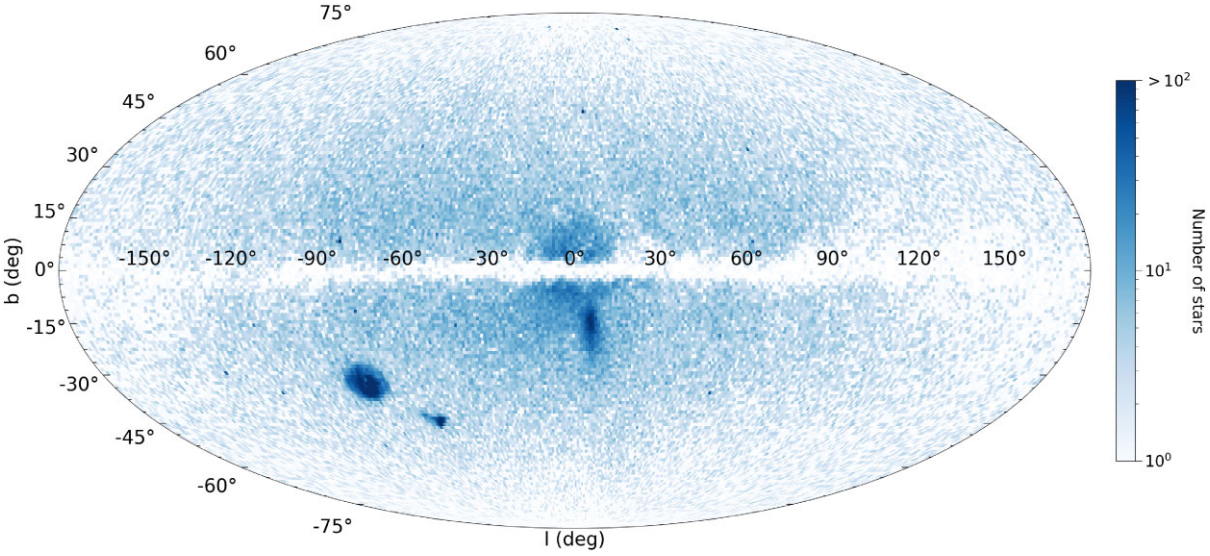

Figure 5. Spatial distribution of 160 146 *ex-situ* stars, as selected by the `NN_parallel` model, is presented within the Galactic coordinate system. Three prominent overdensities are identified, corresponding to the Large Magellanic Cloud (LMC), Small Magellanic Cloud (SMC), and Sgr, respectively. Furthermore, GCs can be found as smaller overdensities. A total of 20 GCs are present in our *ex-situ* sample.

three prominent spacial overdensities can be intuitively identified, corresponding to the LMC, SMC, and Sgr, respectively. Although not as conspicuous as DGs, smaller overdensities in the plot reveal the presence of GCs, with a total of 20 being found. The further selection for DGs and GCs is detailed in Section 3.1. Substructures, which are considered to be disrupted DGs that merged into the MW, cannot be directly identified in the coordinate space but are crucial for understanding the MW's merger history. The selected member stars from substructures including the GSE, Thamnos, Sequoia, Helmi streams, Wukong, and Pontus are presented in Section 3.2. As exhibited in Table 1, we summarize the sample size of the DGs, GCs, and substructures in the subsequent discussion.

Furthermore, we examined the metallicity distribution of the *ex-situ* sample utilizing the metallicity derived from SkyMapper Southern Survey (SMSS; Wolf et al. 2018; Onken et al. 2019; Huang et al. 2022) and the Stellar Abundance and Galactic Evolution Survey (SAGES; Fan et al. 2023; Huang et al. 2023). The SMSS is a photometric survey of the southern sky, while the SAGES covers the

Table 1. Sample size of the three DGs, 20 GCs, and six substructures included in Section 3.

Structures	N stars
LMC	9822
SMC	1899
Sgr core	3319
Sgr stream	1269
GCs	831
GSE	47 521
Thamnos	5291
Sequoia	1714
Helmi streams	1655
Wukong	3140
Pontus	668
Total	77 129

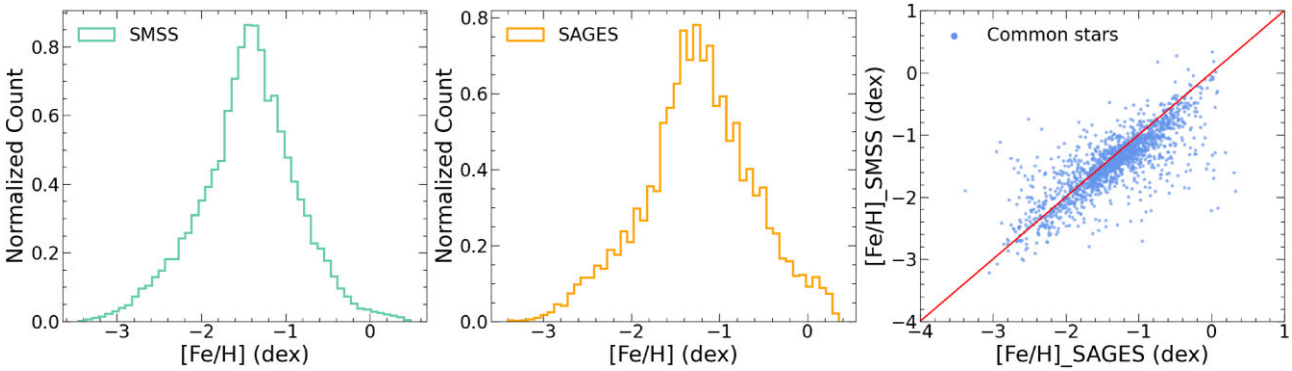

Figure 6. The left subplot presents the metallicity distribution of 62 025 *ex-situ* stars derived from SMSS, while the middle subplot exhibits the metallicity distribution of 18 413 *ex-situ* stars derived from SAGES. On the right, the subplot illustrates the correlation of iron abundance measurements for 1581 common stars between the SMSS and SAGES samples, with the diagonal indicating a one-to-one correlation. Despite slight system differences, the result demonstrate a good consistency.

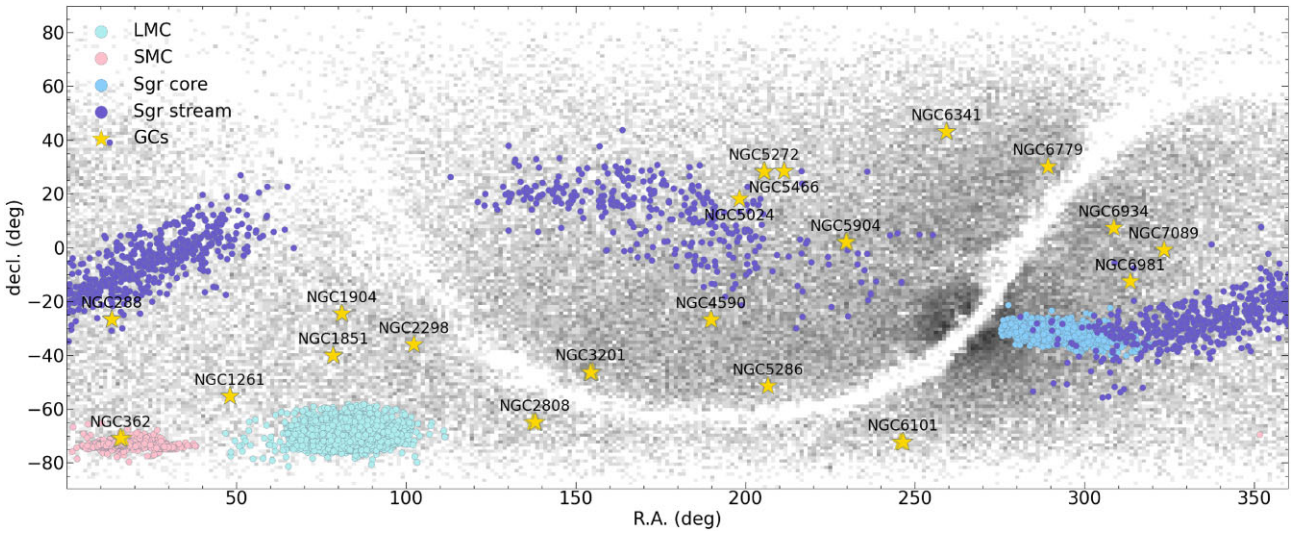

Figure 7. Distribution of the LMC, SMC, Sgr core, Sgr stream and GCs on the celestial sphere. DGs are depicted in distinct colours, and the names of GCs are annotated near their corresponding positions. The background illustrates the number density of the *ex-situ* sample.

northern sky. By cross-matching our *ex-situ* sample with these two catalogues, we obtained 62 025 and 18 413 common stars, respectively. As shown in the left column of Fig. 6, the *ex-situ* stars exhibits a peak metal abundance around −1.4 dex. The peak metallicity of the SAGES sample is slightly higher at approximately −1.3 dex. We checked the agreement of metal abundance measurements for the two samples using 1581 common stars between them, resulting in a good consistency. However, the measurements of the SAGES samples were slightly higher than those of the SMSS samples, possibly due to differences in the filters utilized by the two surveys. Consequently, the peak metallicity of the *ex-situ* sample is likely to lie between −1.3 and −1.4 dex, which is consistent with the typical metallicity of GSE (e.g. Myeong et al. 2019; Amarante, Smith & Boeche 2020; Wu et al. 2022). Taking into account both metallicity analysis and star counts as shown in Table 1, our *ex-situ* sample appears to be composed of multiple components but predominated by the GSE.

3.1 DGs and GCs

In our *ex-situ* sample, DGs and GCs are the most prominent structures, exhibiting a distinct clustering in coordinate space. As shown in Table A1, we selected three DGs combining the position and velocity. Moreover, we selected the disrupted portion of the Sgr by cross-matching with Sgr stream catalogues compiled by Yang et al. (2019) and Ramos et al. (2022). They are shown here as a comparison of the Sgr core. The member stars of these four systems are plotted on the celestial sphere in Fig. 7, where the LMC, SMC, and Sgr core exhibit distinct spatial clustering, while the Sgr stream displays a more extended distribution. Thereafter, we removed the selected member stars of the DGs from our sample, which made the overdensity regions of the GCs more prominent, facilitating the subsequent selection.

For each GC, we primarily select its member stars using the centre coordinate and angular size (θ) from references listed in Table A2. We retain stars within a projected distance ($d_{\rm proj}$) to the GC centre of $d_{\rm proj} < 1.5 \times \theta$, where $1.5 \times \theta$ corresponds to three times the angular radii. Subsequently, we verify whether the selected GC candidates exhibit similar motion properties by examining their proper motion and radial velocity. Outliers in proper motion phase space are manually eliminated, and stars with radial velocity outside the range of $\mu_{\rm rv} \pm 3\sigma_{\rm rv}$ are further removed. Here, $\mu_{\rm rv}$ represents the median radial velocity of a GC, and $\sigma_{\rm rv}$ is half the difference

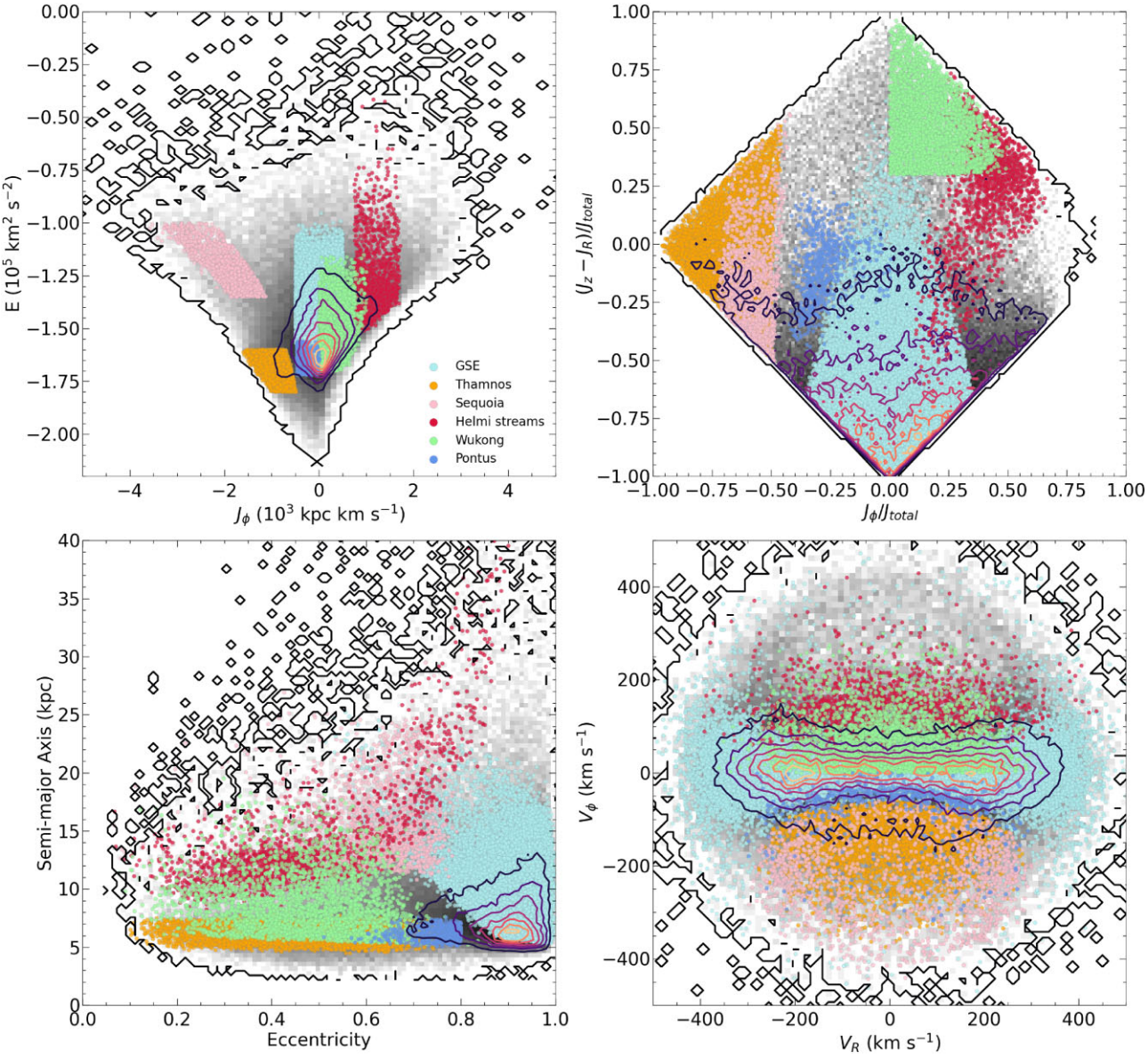

Figure 8. The member stars of the six substructures, as selected using the criteria outlined in Table A3, are depicted in four distinct phase spaces. Each substructure is represented by a unique colour. The background and the contours illustrate the number density of the remaining 143 006 *ex-situ* stars after the removal of DGs and GCs.

between the corresponding value of the 84th percentile and 16th percentile. We only selected GCs with more than 10 member stars. This resulted in a total of 20 GCs, all of which are classified as *ex-situ* in Belokurov & Kravtsov (2023).

The metric values, such as precision and recall, provide an indication of the average classification ability of the overall sample. While the NN demonstrates a high degree of accuracy in identifying certain portions of the sample, its performance may be somewhat diminished when classifying more challenging portions. As DGs and GCs exhibit distinct spatial clustering, their identification by the `NN_parallel` represents a relatively minor test of its capabilities. The true potential of the NN is further demonstrated by its ability to accurately identify member stars of substructures that no longer cluster in coordinate space.

3.2 Substructures

Prior to the selection of substructures, we excluded the member stars of DGs and GCs as identified in Section 3.1, which was necessary to prevent any overlap with the regions where substructures reside in the phase space. During the selection process, we adopted an approach similar to that of Naidu et al. (2020), where each time member stars of a substructure was selected, we removed them from the sample to prevent their influence on the subsequent selection of other substructures. In total, 143 006 *ex-situ* stars were included in our analysis. We then selected 59 989 member stars from six substructures according to the order and criteria outlined in Table A3.

Our selection focuses on substructures including the GSE, Thamnos, Sequoia, Helmi streams, Wukong (also known as LMS-1), and Pontus. Fig. 8 displays the distribution of their member stars in four

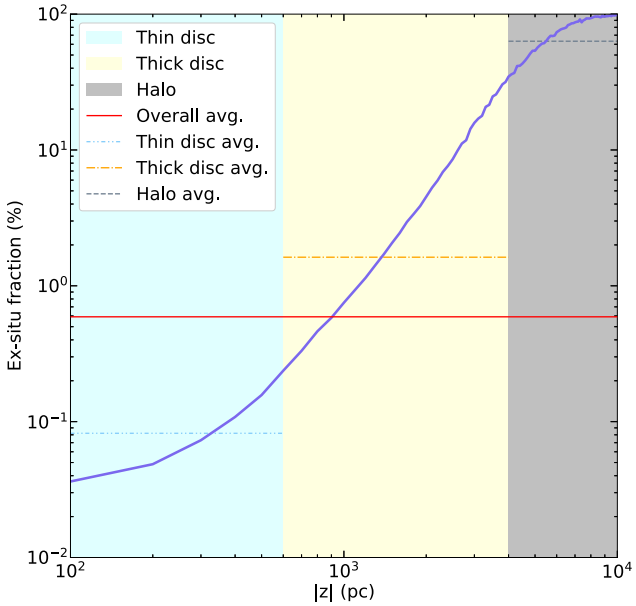

Figure 9. *Ex-situ* fraction as a function of $|z|$. The region with $|z|$ below 0.6 kpc is considered as the thin disc, while the region with $|z|$ between 0.6 and 4 kpc is considered as the thick disc. The region above 4 kpc and below 10 kpc is occupied by the halo. Different regions are represented in distinct colours, and the average *ex-situ* fraction is indicated by horizontal dashed lines in the corresponding region. The red horizontal lines indicate the average *ex-situ* fraction of the target sample.

phase spaces. The second subplot presents an action diamond, with $J_{\rm total}$ defined as the sum of J_R, J_z, and $|J_\Phi|$ (e.g. Myeong et al. 2019). In the third subplot, we characterize the semi-major axis as $(r_{\rm apo} + r_{\rm peri})/2$, and the eccentricity as $(r_{\rm apo} - r_{\rm peri})/(r_{\rm apo} + r_{\rm peri})$, where $r_{\rm apo}$ and $r_{\rm peri}$ represent the distances to the apocentre and pericentre, respectively. The selection criteria, as listed in Table A3, involve multidimensional linear cuts. Fig. 8 illustrates the projection of samples with high-dimensional features in four 2D spaces. As a result, the substructures may display clear linear boundaries in the corresponding plane.

3.3 *Ex-situ* percentage

In the presence of a large sample of high purity *ex-situ* stars, we are able to conduct a statistical analysis of the *ex-situ* fraction in the MW. To ensure the robustness of our analysis, we focus exclusively on target samples located in the region where $|z| \leq 10$ kpc. In Fig. B1, we present a 2D *ex-situ* fraction map that exhibits significant layering along the $|z|$-direction. Based on the layering pattern, we divide our sample into three distinct regions, corresponding to the thin disc ($|z| < 0.6$ kpc), the thick disc (0.6 kpc $\leq |z| < 4$ kpc), and the halo (4 kpc $\leq |z| \leq 10$ kpc). The *ex-situ* samples, as identified by the `NN_parallel` model, were binned according to a bin size of $|z| = 100$ pc. We then plotted the *ex-situ* fraction as a function of $|z|$ in Fig. 9, where the x-axis represents the right boundary of each bin and the y-axis stands for the corresponding *ex-situ* fraction. Fig. 9 provides a clear illustration of how the proportion of *ex-situ* stars varies with the vertical height from the Galactic disc. It also shows the proportion of *ex-situ* stars in the three components of the MW. The average *ex-situ* percentages of the thin disc, the thick disc, and the halo are 0.1 per cent, 1.6 per cent, and 63.2 per cent,[1] respectively. Consistent with our previous knowledge, the thin disc is predominantly composed of *in-situ* stars. In the region of the thick disc, there is a substantial increase in the proportion of *ex-situ* stars, while on average, the thick disc is still primarily composed of *in-situ* stars. In contrast, the stellar halo is dominated by *ex-situ* stars, with this dominance increasing with $|z|$.

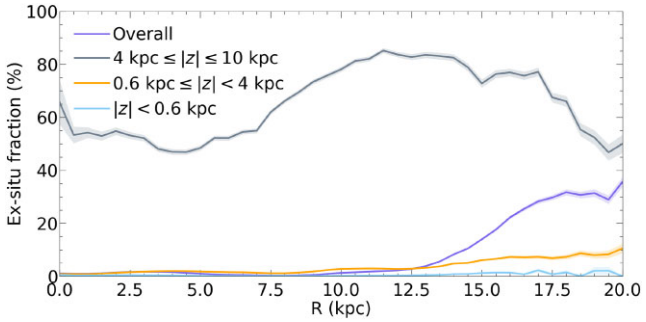

Figure 10. *Ex-situ* fraction as a function of R is illustrated by the purple line, which shows the overall trend of the sample. Distributions for the thin disc, thick disc, and halo as defined in Fig. 9 are depicted in distinct colours. The standard error of each bin is indicated by the shaded area.

Fig. 10 presents an alternative perspective in the Galactocentric cylindrical coordinate system, illustrating the variation in the proportion of *ex-situ* stars with respect to their distance from the Galactic centre (R). The data is binned using a bin size of R = 500 pc, and the x-axis represents the left boundary of each bin. As illustrated by the purple line in Fig. 10, the fraction of *ex-situ* stars is less than 2 per cent up until R ≈12 kpc, and it significantly increases towards the outskirts of the MW. A general growth in the *ex-situ* fraction with increasing R can also be observed in both the thick disc and thin disc, with the trend being more pronounced in the thick disc. The *ex-situ* fraction in the thick disc is higher than that in the thin disc over the entire coverage of R, which is also indicated by the average values in Fig. 9. Within the halo, *ex-situ* stars hold a dominant position and the curve is mainly influenced by the spatial distribution of member stars from distinct substructures. When R is less than 5 kpc, the *ex-situ* fraction is relatively stable. As R extends beyond 5 kpc, the trend of the curve begins to be dominated by the GSE (e.g An & Beers 2021; Wu et al. 2022), exhibiting an upward trajectory until R approximates 12 kpc. Subsequently, the curve commences its decline, aligning with the spatial distribution trend of GSE member stars (e.g. Naidu et al. 2021).

4 DISCUSSION

4.1 Threshold setting

In Section 3, all the *ex-situ* stars are selected using a threshold value of 0.5. This threshold was chosen as it maintains an optimal balance between precision and recall, as evidenced by the results in the test set. However, as the metrics are not based on truth-level label, our choice can only be taken as a reference. In practice, the setting of threshold may vary depending on the scientific goals. If the goal is to obtain a purer *ex-situ* sample, the threshold can be increased. On the

[1]The estimations for both the thin disc and the thick disc yield standard errors less than 0.01 per cent, while for the halo, the standard error is 0.02 per cent. All estimated values are rounded to one decimal places for consistency.

other hand, if a purer *in-situ* sample is desired, stars with small raw prediction values from the NN should be selected. Additionally, the adjustment to threshold value can reduce the impact of samples with inaccurate observations to a certain extent. By combining a relatively stringent threshold with a restriction on the error of observations, the quality of the required samples can be effectively ensured.

If we take one step further, it becomes apparent that the ultimate goal of both traditional linear segmentation methods (e.g. Nissen & Schuster 2010) and deep learning methods is to obtain an *ex-situ* sample that conforms to experience and previous knowledge. The primary difference between the two methods lies in their performance. As demonstrated in Ostdiek et al. (2020), traditional methods struggle to balance precision and recall, with high recall often accompanied by low precision. Although deep learning methods are not currently capable of breaking this deadlock, they are able to achieve higher recall under the same precision level compared with traditional methods, and vice versa. Under certain threshold values, an NN model can even surpass traditional methods in both metrics. The dynamic classification of stars involves many dimensions of information, and NN excels at processing high-dimensional data. Traditional selection methods directly restrict dynamic parameters, providing clear physical meaning. However, this approach may overlook certain dimensions and fail to account for potential connections between parameters of different dimensions. The NN method, on the other hand, directly outputs a continuous value between 0 and 1. All we need to do is determine a threshold to divide these continuous values into two categories. While adjusting the threshold may appear less physically meaningful than adjusting limits on dynamic parameters, it is important to note that the physical meaning is already reflected in the mapping established by the NN from high-dimensional dynamic parameters to output values. Adjusting the threshold is simply an intuitive and efficient process to obtain the final classification result. Moreover, the threshold is a hyperparameter that is independent of the model structure, enabling the attainment of different classification outcomes with out necessitating any modifications to the NN itself. By fine-tuning the threshold, classification results can be tailored to achieve various scientific objectives and meet specific requirements.

4.2 Result assessment

In contrast to simulations, it is extremely difficult to perfectly distinguish between *in-situ* and *ex-situ* stars in observations. Consequently, validating the *ex-situ* sample identified by `NN_parallel` presents a significant challenge, and we have refrained from quantitatively evaluating the performance of the NN. Nevertheless, as depicted in Fig. 5, the NN has effectively eliminated the majority of the disc stars. This is further evidenced by Fig. 8, which clearly shows that the area where the disc is located has been accurately deducted. In Section 3, we present a selection of member stars from DGs, GCs, and substructures within our *ex-situ* sample. The member stars are selected based on the criteria established in the literature, without the use of any clustering algorithm. This approach was chosen as a quick and convenient means to demonstrate the potential of our algorithm and to verify the reliability of our *ex-situ* samples. It should be noted that most of the selection methods employed are relatively conservative. As such, the selected member stars are not exhaustive, as evidenced by the remaining overdensities in Fig. 8. In forthcoming studies, clustering algorithms will be employed to facilitate more comprehensive investigations and detailed examination of each substructure. Furthermore, the origin of the substructures and potential correlations between various substructures will be investigated through simulations (e.g. Amarante et al. 2022).

In Section 3.3, we provided a comprehensive illustration of the variation in the proportion of *ex-situ* stars with respect to the vertical height from the Galactic disc and the distance from the Galactic centre. However, it is important to note that our results are subject to the selection effect of *Gaia* and the performance of the NN. As such, the result we present only reflect the analysis of our current sample. We divide our sample into three distinct parts according to the layering pattern in Fig. B1, which do not precisely align with the three components of the MW as defined canonically (Bland-Hawthorn & Gerhard 2016). The estimation of the proportion of *ex-situ* stars in the thin disc, thick disc, and halo should be considered as preliminary. What we present is far from a definitive conclusion but rather a possibility for statistical exploration of the *ex-situ* components within the MW. Despite the current limitations, we believe that our study provides novel insights and serves as a catalyst for future research endeavors.

5 CONCLUSION

In this work, we developed a deep learning methodology to identify the *ex-situ* stars within *Gaia* DR3. Our base model, referred to as `NN_FIRE`, was built to learn fundamental kinematic features. The training of `NN_FIRE` was conducted on a synthetic *Gaia* catalogue derived from a MW-mass galaxy from the *Latte* suite of FIRE-2 simulations, specifically the m12i. The input for `NN_FIRE` consists of 3D position and velocity data, which are transformed from `ra`, `dec`, `pmra`, `pmdec`, `radial_velocity`, and `dhel_true`. The output of the NN is a series of continuous values ranging from 0 to 1, which can be interpreted as the probability of a star being *ex-situ*. According to the truth-level label, the `NN_FIRE` model achieved an AUC exceeding 0.98 on the test set. At a threshold of 0.5, we obtained a precision of 80.1 per cent and a recall of 34.5 per cent. By increasing the threshold to 0.75, the precision escalated to as high as 98.3 per cent.

Based on `NN_FIRE`, a senior model noted as `NN_parallel` was built, which processes dynamic data derived from real observations of *Gaia*. During the training phase of `NN_parallel`, the weights of the base model were frozen. The output of the base model was concatenated with the output of a sub-network that processes J_R, J_z, and J_Φ. We adopted stars with available radial velocity and photo-astrometric distance estimated by Anders et al. (2022) within *Gaia* DR3 as our target sample and cross-matched it with the APOGEE DR17 catalogue and LAMOST DR8 VAC to build a data set with elemental abundance information for training `NN_parallel`. The stars were labeled using a chemical tagging method, as outlined in Section 2.2.2, which performed a segmental selection in the [Mg/Mn]-[Al/Fe] plane. Although the chemical tagging method did mix some disc stars into the *ex-situ* sample, with the aid of the first training phase, `NN_parallel` overcame this problem and produced better results than selecting solely through chemical criteria.

Out of the 27 085 748 stars in the target sample, the `NN_parallel` model successfully identified 160 146 *ex-situ* stars. According to the metallicity distribution, the *ex-situ* sample is composed of multiple components, with the GSE being the most prominent. A significant number of member stars from the LMC, SMC, Sgr, as well as 20 GCs were found in the *ex-situ* sample. Additionally, member stars from substructures including the GSE, Thamnos, Sequoia, Helmi streams, Wukong, and Pontus were also selected. The identification of these member stars not only verifies the reliability of our *ex-situ* sample and our algorithm but also provides a substantial sample for future research on DGs, GCs, and substructures. Moreover, we provided an intuitive illustration

demonstrating an increase in the *ex-situ* fraction as the vertical height from the Galactic disc and the distance from the Galactic centre increase. Finally, We conducted a preliminary estimation of the proportion of *ex-situ* stars in the thin disc, thick disc, and halo, yielding percentages of 0.1 per cent, 1.6 per cent, and 63.2 per cent, respectively.

This study serves as a prior work, with the primary objective of introducing our proposed algorithm and providing a reliable large sample of *ex-situ* stars. The findings delineated herein are the product of a preliminary analysis of our *ex-situ* sample. While these findings do not encompass the full scope of potential results, we posit that they effectively underscore the viability and promise of our proposed methodology. It is our aspiration that this research will bring the community fresh insights and inspiration regarding the evolution and merger history of the MW. Looking ahead, we plan to augment our research by incorporating simulations, conducting chemical abundance analyses, and employing clustering algorithms to delve deeper into our findings. We eagerly anticipate unearthing new revelations in this exciting field of study.

ACKNOWLEDGEMENTS

This study is supported by the National Natural Science Foundation of China under grant Nos. 11988101, 11890694, 12373020, National Key R&D Program of China Nos. 2019YFA0405500, 2023YFE0107800 and CAS Project for Young Scientists in Basic Research grant No. YSBR-062. ZL thanks Drs Haining Li and Wenbo Wu for their helpful discussions and insightful suggestions. JA acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 852839). We have made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

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.

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.sdss.org.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics| Harvard & Smithsonian, the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

DATA AVAILABILITY

The simulated data utilized during the training phase is accessible at http://ananke.hub.yt. Astrometric and chemical abundance data can be obtained from the websites of *Gaia* (https://gea.esac.esa.int/archive/), APOGEE (https://www.sdss4.org/dr17/irspec/dr_synopsis) and LAMOST (http://www.lamost.org/dr8/v1.1/doc/vac), respectively. The result of this study will be shared upon reasonable request to the corresponding author.

REFERENCES

Abadi M. et al., 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. https://www.tensorflow.org/ (Access Dec 2023)
Abolfathi B. et al., 2018, ApJS, 235, 42
Accetta K. et al., 2022, ApJS, 259, 35
Ahumada R. et al., 2020, ApJS, 249, 3
Amarante J. A. S., Smith M. C., Boeche C., 2020, MNRAS, 492, 3816
Amarante J. A., Debattista V. P., Silva L. B. E., Laporte C. F., Deg N., 2022, ApJ, 937, 12
An D., Beers T. C., 2021, ApJ, 918, 74
Anders F. et al., 2022, A&A, 658, A91
Bellardini M. A., Wetzel A., Loebman S. R., Bailin J., 2022, MNRAS, 514, 4270
Belokurov V., Kravtsov A., 2022, MNRAS, 514, 689
Belokurov V., Kravtsov A., 2023, preprint (arXiv:2309.15902)
Belokurov V., Erkal D., Evans N., Koposov S., Deason A., 2018, MNRAS, 478, 611
Bennett M., Bovy J., 2019, MNRAS, 482, 1417
Binney J., 2012, MNRAS, 426, 1324
Bland-Hawthorn J., Gerhard O., 2016, ARA&A, 54, 529
Borsato N. W., Martell S. L., Simpson J. D., 2020, MNRAS, 492, 1370
Buder S. et al., 2021, MNRAS, 506, 150
Campello R. J., Moulavi D., Sander J., 2013, in Pei J., Tseng V. S., Cao L., Motoda H., Xu G., eds, Proc. 17th Pacific-Asia Conf., Advances in Knowledge Discovery and Data Mining. Springer, Berlin, p. 160
Carrillo A., Deason A. J., Fattahi A., Callingham T. M., Grand R. J., 2024, MNRAS, 527, 2165
Chollet F. et al., 2015, Keras. https://github.com/fchollet/keras (Access Dec 2023)
Conroy C. et al., 2019, ApJ, 883, 107
Cunningham E. C., Hunt J. A. S., Price-Whelan A. M., Johnston K. V., Ness M. K., Lu Y., Escala I., Stelea I. A., 2023, preprint (arXiv:2307.08730)
Das P., Hawkins K., Jofré P., 2020, MNRAS, 493, 5195
De Silva G. M. et al., 2015, MNRAS, 449, 2604
Dempster A. P., Laird N. M., Rubin D. B., 1977, J R Stat. Soc. Series B Stat. Methodol., 39, 1
Deng L.-C. et al., 2012, Res. Astron. Astrophys., 12, 735
Di Criscienzo M., Caputo F., Marconi M., Musella I., 2006, MNRAS, 365, 1357
Ester M., Kriegel H.-P., Sander J., Xu X., 1996, SIGKDD Explorations, 96, 226
Fan Z. et al., 2023, ApJS, 268, 9
Feltzing S., Feuillet D., 2023, ApJ, 953, 143
Fernandes L. et al., 2023, MNRAS, 519, 3611
Fernández-Trincado J. G., Beers T. C., Tang B., Moreno E., Pérez-Villegas A., Ortigoza-Urdaneta M., 2019, MNRAS, 488, 2864

Fernández-Trincado J. G. et al., 2022, A&A, 663, A126
Feuillet D. K., Feltzing S., Sahlholdt C. L., Casagrande L., 2020, MNRAS, 497, 109
Fix E., Hodges J. L., 1951, Discriminatory Analysis: Nonparametric Discrimination: Consistency Properties. USAF School of Aviation Medicine, Randolph Field, TX, Project Number: 21-49-004, Report Number: 4
Gaia Collaboration, 2016a, A&A, 595, A1
Gaia Collaboration, 2016b, A&A, 595, A2
Gaia Collaboration, 2018, A&A, 616, A1
Gaia Collaboration, 2023, A&A, 674, A1
Gallart C., Bernard E. J., Brook C. B., Ruiz-Lara T., Cassisi S., Hill V., Monelli M., 2019, Nat. Astron., 3, 932
Goldsbury R., Richer H. B., Anderson J., Dotter A., Sarajedini A., Woodley K., 2010, AJ, 140, 1830
Grillmair C. J., 2006, ApJ, 645, L37
Hahnloser R. H., Sarpeshkar R., Mahowald M. A., Douglas R. J., Seung H. S., 2000, Nature, 405, 947
Hasselquist S. et al., 2021, ApJ, 923, 172
Hawkins K., Jofre P., Masseron T., Gilmore G., 2015, MNRAS, 453, 758
Haywood M., Di Matteo P., Lehnert M., Snaith O., Khoperskov S., Gómez A., 2018, ApJ, 863, 113
Helmi A., 2020, ARA&A, 58, 205
Helmi A., White S. D., De Zeeuw P. T., Zhao H., 1999, Nature, 402, 53
Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G., 2018, Nature, 563, 85
Herzog-Arbeitman J., Lisanti M., Madau P., Necib L., 2018, Phys. Rev. Lett., 120, 41102
Holtzman J. A. et al., 2015, AJ, 150, 148
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Horta D. et al., 2021, MNRAS, 500, 1385
Horta D. et al., 2023, MNRAS, 520, 5671
Huang Y. et al., 2022, ApJ, 925, 164
Huang Y. et al., 2023, ApJ, 957, 65
Ibata R. A., Gilmore G., Irwin M., 1994, Nature, 370, 194
Kingma D. P., Ba J., 2014, preprint (arXiv:1412.6980)
Kohonen T., 2001, Self-organizing maps, 3rd edn. Springer Series in Information Sciences, Vol. 30. Springer, Berlin
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
Kruijssen J. D. et al., 2020, MNRAS, 498, 2472
LeCun Y., Bengio Y., Hinton G., 2015, Nature, 521, 436
Li H.-N., Zhao G., Christlieb N., Wang L., Wang W., Zhang Y., Hou Y., Yuan H., 2015, ApJ, 798, 110
Li H., Tan K., Zhao G., 2018, ApJS, 238, 16
Li Z., Zhao G., Chen Y., Liang X., Zhao J., 2022, MNRAS, 517, 4875
Limberg G. et al., 2023, preprint (arXiv:2308.13702)
Lin T.-Y., Goyal P., Girshick R., He K., Dollár P., 2017, in Proc. IEEE International Conference on Computer Vision. IEEE, New York, p. 2980
Lindegren L. et al., 2021, A&A, 649, A4
Liu X.-W. et al., 2013, Proc. Int. Astron. Union, 9, 310
Luo A.-L. et al., 2015, Res. Astron. Astrophys., 15, 1095
McMillan P. J., 2017, MNRAS, 465, 76
Majewski S. R. et al., 2017, AJ, 154, 94
Malhan K., 2022, ApJ, 930, L9
Myeong G., Evans N., Belokurov V., Sanders J., Koposov S., 2018, ApJ, 863, L28
Myeong G., Vasiliev E., Iorio G., Evans N., Belokurov V., 2019, MNRAS, 488, 1235
Nachman B., Shih D., 2020, Phys. Rev. D, 101, 75042
Naidu R. P., Conroy C., Bonaca A., Johnson B. D., Ting Y.-S., Caldwell N., Zaritsky D., Cargile P. A., 2020, ApJ, 901, 48
Naidu R. P. et al., 2021, ApJ, 923, 92
Necib L. et al., 2020a, Nat. Astron., 4, 1078
Necib L., Ostdiek B., Lisanti M., Cohen T., Freytsis M., Garrison-Kimmel S., 2020b, ApJ, 903, 25
Nidever D. L. et al., 2020, ApJ, 895, 88
Nissen P. E., Schuster W. J., 2010, A&A, 511, L10
Onken C. A. et al., 2019, PASA, 36, e033
Ortigoza-Urdaneta M. et al., 2023, A&A, 676, A140
Ostdiek B. et al., 2020, A&A, 636, A75
Pedregosa F. et al., 2011, J. Mach. Learn. Res., 12, 2825
Price-Whelan A. M. et al., 2018, AJ, 156, 123
Queiroz A. B. d. A. et al., 2018, MNRAS, 476, 2556
Ramos P. et al., 2022, A&A, 666, A64
Rix H.-W. et al., 2022, ApJ, 941, 45
Sanders J. L., Binney J., 2016, MNRAS, 457, 2107
Sanderson R. E. et al., 2020, ApJS, 246, 6
Schlafly E. et al., 2016, ApJ, 821, 78
Schönrich R., Binney J., Dehnen W., 2010, MNRAS, 403, 1829
Shao Z., Li L., 2019, MNRAS, 489, 3093
Sharma S., Bland-Hawthorn J., Johnston K. V., Binney J., 2011, ApJ, 730, 3
Shih D., Buckley M. R., Necib L., Tamanas J., 2022, MNRAS, 509, 5992
Sinnott R. W., 1988, NGC 2000.0: The Complete New General Catalogue and Index Catalogues of Nebulae and Star Clusters by J. L. E. Dreyer. Cambridge Univ. Press and Sky Publishing Corporation, Cambridge
Steinmetz M. et al., 2006, AJ, 132, 1645
Van der Maaten L., Hinton G., 2008, J. Mach. Learn. Res., 9, 2579
Vasiliev E., 2019, MNRAS, 482, 1525
Vasiliev E., Belokurov V., Erkal D., 2021, MNRAS, 501, 2279
Wetzel A. et al., 2023, ApJS, 265, 44
White S. D., Rees M. J., 1978, MNRAS, 183, 341
Wolf C. et al., 2018, PASA, 35, e010
Wu W., Zhao G., Xue X.-X., Bird S. A., Yang C., 2022, ApJ, 924, 23
Xing Q.-F., Zhao G., Aoki W., Honda S., Li H.-N., Ishigaki M. N., Matsuno T., 2019, Nat. Astron., 3, 631
Xing Q.-F. et al., 2023, Nature, 620, 323
Xue X.-X. et al., 2011, ApJ, 738, 79
Yan H.-L. et al., 2018, Nat. Astron., 2, 790
Yan H.-L. et al., 2021, Nat. Astron., 5, 86
Yan H. et al., 2022, Innovation, 3, 100224
Yang C. et al., 2019, ApJ, 886, 154
Yanny B. et al., 2009, AJ, 137, 4377
Yuan Z., Chang J., Banerjee P., Han J., Kang X., Smith M., 2018, ApJ, 863, 26
Yuan Z. et al., 2020, ApJ, 891, 39
Zhao G., Chen Y., 2021, Sci. China Phys. Mech. Astron., 64, 239562
Zhao G., Chen Y.-Q., Shi J.-R., Liang Y.-C., Hou J.-L., Chen L., Zhang H.-W., Li A.-G., 2006, Chin. J. Astron. Astrophys., 6, 265
Zhao J., Zhao G., Chen Y., 2009, ApJ, 692, L113
Zhao G., Zhao Y.-H., Chu Y.-Q., Jing Y.-P., Deng L.-C., 2012, Res. Astron. Astrophys., 12, 723
Zhao J.-K., Zhao G., Chen Y.-Q., Tan K.-F., Gao M.-T., Yang M., Zhang Y., Hou Y.-H., 2015, Res. Astron. Astrophys., 15, 1378

APPENDIX A: SELECTION CRITERIA

Table A1. Summary of the selection criteria we adopted to select DGs and the literature we consulted.

Structures	Position	Criteria	References
LMC	$\alpha = 80°.893\,860, \delta = -69°.756\,126$	$d_{\rm proj} \leqq 12°$; $160 \leqq {\rm RV} \leqq 348$ (km s^{-1}); $1.2 \leqq {\rm PM}_\alpha \leqq 2.5$ (mas yr^{-1}); $-0.7 \leqq {\rm PM}_\delta \leqq 1.5$ (mas yr^{-1})	(1)(2)
SMC	$\alpha = 13°.186\,87, \delta = -72°.8286$	$d_{\rm proj} \leqq 8°$; $71 \leqq {\rm RV} \leqq 220$ (km s^{-1}); $0.2 \leqq {\rm PM}_\alpha \leqq 1.8$ (mas yr^{-1}); $-1.8 \leqq {\rm PM}_\delta \leqq -0.8$ (mas yr^{-1})	(1)(2)
Sgr core	$\Lambda \approx 0°$, $B \approx 1°.5$	$\lvert B\rvert < 10°$; $b < 0°$; $90 \leqq {\rm RV} \leqq 220$ (km s^{-1}); $-3.5 \leqq {\rm PM}_\alpha \leqq -2$ (mas yr^{-1}); $-2.5 \leqq {\rm PM}_\delta \leqq -0.5$ (mas yr^{-1})	(3)(4)
Sgr stream		Cross-match with Sgr stream catalogues	(5)(6)

(Λ, B) are coordinates in the right-handed coordinate system with respect to the Sgr stream as in Cunningham et al. (2023), $d_{\rm proj}$ represents the projected distance on the celestial sphere.
References: (1) Nidever et al. (2020); (2) Fernandes et al. (2023); (3) Vasiliev, Belokurov & Erkal (2021); (4) Cunningham et al. (2023); (5) Yang et al. (2019); (6) Ramos et al. (2022).

Table A2. The table exhibits the properties of 20 GCs we selected, and the literature we referred to during the selection.

Name	N stars	Centre (α, δ)	Angular size (arcmin)	PM$_\alpha$ (mas yr^{-1})	PM$_\delta$ (mas yr^{-1})	RV (km s^{-1})	References
NGC 3201	143	$10^{\rm h}17^{\rm m}36.82^{\rm s}$, $-46°24'44.9''$	18.2	$7.870 \sim 8.771$	$-2.363 \sim -1.438$	$468.14 \sim 506.36$	(1)(2)
NGC 5904	127	$15^{\rm h}18^{\rm m}33.22^{\rm s}$, $+02°04'51.7''$	17.4	$3.786 \sim 4.483$	$-10.284 \sim -9.503$	$36.01 \sim 69.77$	(1)(2)
NGC 5272	86	$13^{\rm h}42^{\rm m}11.62^{\rm s}$, $+28°22'38.2''$	16.2	$-0.458 \sim 0.176$	$-2.894 \sim -2.362$	$-161.77 \sim -131.13$	(1)(2)
NGC 6341	47	$17^{\rm h}17^{\rm m}07.39^{\rm s}$, $+43°08'09.4''$	11.2	$-5.144 \sim -4.646$	$-0.914 \sim -0.445$	$-133.81 \sim -112.20$	(1)(2)
NGC 2808	40	$09^{\rm h}12^{\rm m}03.10^{\rm s}$, $-64°51'48.6''$	13.8	$0.787 \sim 1.271$	$-0.172 \sim 0.618$	$89.55 \sim 125.43$	(1)(2)
NGC 7089	40	$21^{\rm h}33^{\rm m}27.02^{\rm s}$, $-00°49'23.7''$	12.9	$3.141 \sim 3.708$	$-2.480 \sim -1.906$	$-17.50 \sim 10.02$	(1)(2)
NGC 288	39	$00^{\rm h}52^{\rm m}45.24^{\rm s}$, $-26°34'57.4''$	19.2	$4.033 \sim 4.328$	$-5.877 \sim -5.592$	$-51.54 \sim -36.89$	(1)(4)
NGC 362	37	$01^{\rm h}03^{\rm m}14.26^{\rm s}$, $-70°50'55.6''$	12.9	$6.426 \sim 6.963$	$-2.944 \sim -2.373$	$209.43 \sim 234.15$	(1)(2)
NGC 1904	35	$05^{\rm h}24^{\rm m}10.59^{\rm s}$, $-24°31'27.3''$	8.7	$2.292 \sim 2.623$	$-1.704 \sim -1.473$	$200.14 \sim 215.42$	(2)(3)
NGC 6101	32	$16^{\rm h}25^{\rm m}48.12^{\rm s}$, $-72°12'07.9''$	10.7	$1.655 \sim 1.867$	$-0.360 \sim -0.159$	$358.13 \sim 373.06$	(1)(2)
NGC 1851	28	$05^{\rm h}14^{\rm m}06.76^{\rm s}$, $-40°02'47.6''$	11.0	$2.053 \sim 2.273$	$-0.886 \sim -0.498$	$309.92 \sim 326.88$	(1)(2)
NGC 6779	27	$19^{\rm h}16^{\rm m}35.57^{\rm s}$, $+30°11'00.5''$	7.1	$-2.174 \sim -1.854$	$1.431 \sim 1.759$	$-148.02 \sim -125.18$	(1)(2)
NGC 4590	25	$12^{\rm h}39^{\rm m}27.98^{\rm s}$, $-26°44'38.6''$	12.0	$-2.802 \sim -2.582$	$1.592 \sim 1.888$	$-100.39 \sim -87.51$	(1)(2)
NGC 5024	24	$13^{\rm h}12^{\rm m}55.25^{\rm s}$, $+18°10'05.4''$	12.6	$-0.274 \sim -0.029$	$-1.468 \sim -1.234$	$-71.21 \sim -52.03$	(1)(2)
NGC 5466	20	$14^{\rm h}05^{\rm m}27.29^{\rm s}$, $+28°32'04.0''$	11.0	$-5.415 \sim -5.227$	$-0.866 \sim -0.736$	$104.73 \sim 112.14$	(1)(2)
NGC 5286	18	$13^{\rm h}46^{\rm m}26.81^{\rm s}$, $-51°22'27.3''$	9.1	$-0.074 \sim 0.520$	$-0.390 \sim 0.049$	$51.69 \sim 72.63$	(1)(2)
NGC 6934	18	$20^{\rm h}34^{\rm m}11.37^{\rm s}$, $+07°24'16.1''$	5.9	$-2.743 \sim -2.547$	$-4.880 \sim -4.559$	$-419.38 \sim -399.28$	(1)(2)
NGC 1261	16	$03^{\rm h}12^{\rm m}16.21^{\rm s}$, $-55°12'58.4''$	6.9	$1.538 \sim 1.721$	$-2.210 \sim -2.010$	$60.74 \sim 81.34$	(1)(2)
NGC 2298	16	$06^{\rm h}48^{\rm m}59.41^{\rm s}$, $-36°00'19.1''$	9.6	$3.184 \sim 3.406$	$-2.253 \sim -2.065$	$141.01 \sim 158.16$	(1)(4)
NGC 6981	13	$20^{\rm h}53^{\rm m}27.70^{\rm s}$, $-12°32'14.3''$	5.9	$-1.325 \sim -1.175$	$-3.427 \sim -3.294$	$-334.19 \sim -322.34$	(1)(2)

In particular, the PM$_\alpha$, PM$_\delta$, and RV columns respectively indicate the minimum and maximum values of the proper motion and radial velocity of the samples obtained by the selection method introduced in Section 3.1. *References:* (1) Goldsbury et al. (2010); (2) Sinnott (1988); (3) Di Criscienzo et al. (2006); (4) Shao & Li (2019).

Table A3. Summary of the selection criteria we employed to identify the substructures and the literature we utilized for reference.

Substructures	Criteria	References
GSE	$\lvert L_z\rvert < 0.5(\times 10^3$ kpc km s$^{-1})$; $30 < \sqrt{J_r} < 50$ (kpc$^{1/2}$ km$^{1/2}$ s$^{-1/2}$); $r_{\rm apo} < 40$ (kpc)	(1)(2)
Thamnos	$\eta < -0.4$; $-1.8 < E < -1.6$ ($\times$ 10^5 km^2 s^{-2})	(3)(4)
Sequoia	$-0.65 < \eta < -0.4$; $-1.35 < E < -1.0$ ($\times$ 10^5 km^2 s^{-2})	(3)
Helmi streams[a]	$0.75 < J_\phi < 1.7(\times 10^3$ kpc km s$^{-1})$; $1.6 < L_\perp < 3.2(\times 10^3$ kpc km s$^{-1})$	(4)(5)
Wukong[b]	$0 < J_\phi < 1.0(\times 10^3$ kpc km s$^{-1})$; $-1.7 < E < -1.15$ ($\times$ 10^5 km^2 s^{-2}); $L_y > 2.0(\times 10^3$ kpc km s$^{-1})$; $(J_z - J_r)/J_{\rm tot} > 0.3$; $90° < \arccos(L_z/L) < 120°$	(4)(6)
Pontus	$-1.72 < E < -1.56$ ($\times$ 10^5 km^2 s^{-2}); $-470 < J_\phi < 5($ kpc km s$^{-1})$; $245 < J_r < 725($ kpc km s$^{-1})$; $115 < J_z < 545($ kpc km s$^{-1})$; $390 < L_\perp < 865($ kpc km s$^{-1})$; $0.5 < e < 0.8$; $1 < r_{\rm peri} < 3$ (kpc); $8 < r_{\rm apo} < 13$ (kpc)	(7)(8)

[a] To avoid the inclusion of Sgr, we impose a restriction that $\lvert B\rvert > 10°$.
[b] We further exclude the sky areas where the distance may be overestimated as mentioned in Section 2.2.1, specifically, the region where $-50° < l < 50°$ and $-10° < b < 15°$.
Circularity is represented with η, $L_\perp = \sqrt{L_x^2 + L_y^2}$, and e stands for eccentricity.
References: (1) Feuillet et al. (2020); (2) Wu et al. (2022); (3) Koppelman et al. (2019b); (4) Naidu et al. (2020); (5) Koppelman et al. (2019a); (6) Limberg et al. (2023); (7) Horta et al. (2023); (8) Malhan (2022).

APPENDIX B: 2D *EX-SITU* FRACTION MAP

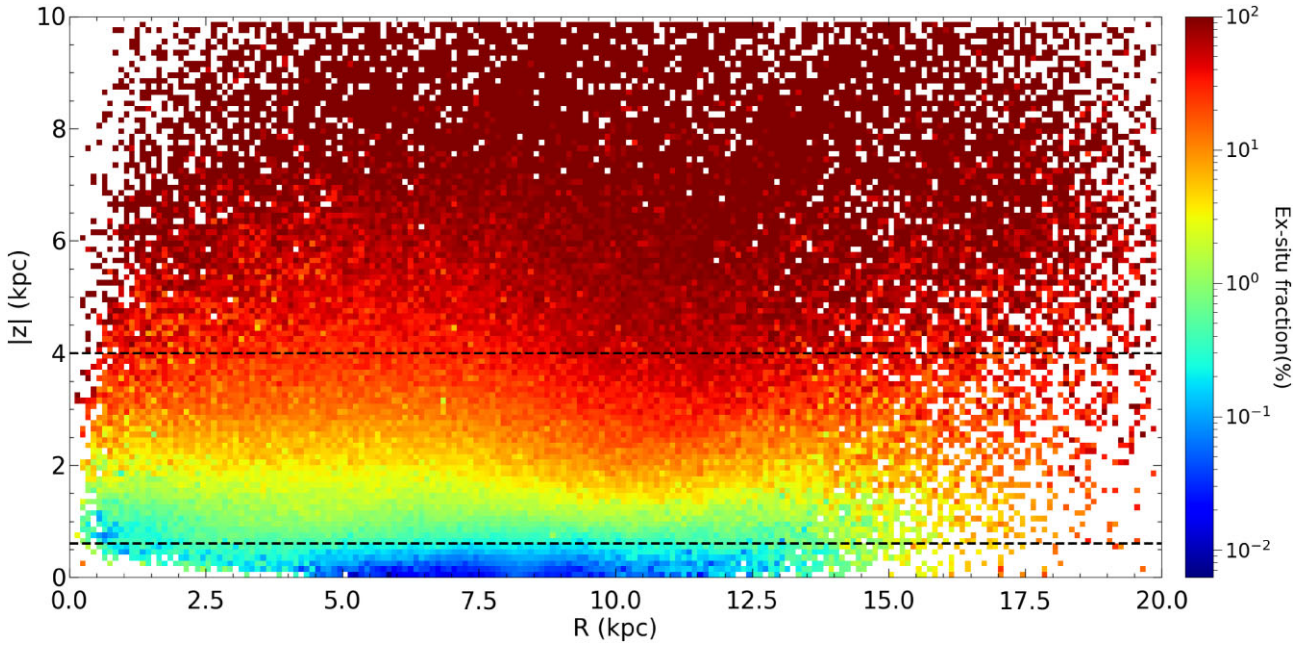

Figure B1. We illustrate the distribution of *ex-situ* fraction within the Galactocentric cylindrical coordinate system. The map is binned with a step size of 100 pc in both R and $|z|$ axis. The *ex-situ* fraction in each grid is visually represented through colour coding, which distinctly exhibits layering along the $|z|$-direction. According to the layering pattern, we divide our sample into three distinct regions: (1) a region characterized by an almost negligible presence of *ex-situ* stars ($|z| < 0.6$ kpc); (2) a region where *ex-situ* stars are slightly interspersed (0.6 kpc $\leq |z| < 4$ kpc); (3) a region predominantly occupied by *ex-situ* stars (4 kpc $\leq |z| \leq 10$ kpc).

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