# Galactic substructures as tracers of dark matter and stellar evolution

Reino, S.

# *Chapter 4*

## *Constraints on the Galactic potential from action-space clustering of halo stars from the H3 survey*

The stellar halo of the Galaxy, built up through accretion of satellite galaxies, stores a treasure trove of information in its orbits which can be used to extract the Galactic mass profile. We place new constraints on the mass profile by considering the halo stars contained in the Hectochelle in the Halo at High Resolution (H3) Survey with the action-clustering method, where the degree of clustering is measured by Kullback-Leibler divergence (KLD). To study the influence of phase space measurement errors and contamination in the data, we apply the same analysis to a set of accreted structures formed in the FIRE cosmological-hydrodynamical simulations. We uncover a limitation of the KLD-based method: while it delivers accurate results irrespective of the presence of interlopers and/or measurement errors, both effects deteriorate the precision of the constraints. Leave-one-out cross-validation (LOOCV) shows that the uncertainties returned using the KLD between neighboring distributions are greatly overestimated in such cases, and that LOOCV should be used when possible. The KLD-based uncertainty derivation yields $v_{\rm circ}(15.4\ {\rm kpc}) = 240.4^{+79.3}_{-72.2}$ km s$^{-1}$, while LOOCV returns $v_{\rm circ}(15.4\ {\rm kpc}) = 240.4^{+0}_{-21.9}$ km s$^{-1}$.

## 4.1 Introduction

In the current paradigm of structure formation galaxies grow in a hierarchical manner: larger galaxies accrete dwarf galaxies some of which assimilate into the host by undergoing tidal disruption. This process embeds the Galactic halo with a considerable amount of structure of different morphologies: as the dwarf galaxies undergo tidal stripping they stretch out in long stellar streams which eventually evolve to a phase-mixed state and lose their spatial coherence. However, the stars from accreted structures bear memory of their origin, from their progenitor's orbit - encoded into actions and/or energy - to chemistry.

The potential to harness this stored information makes accreted structures a considerable source of interest for multiple fields of astronomy, from Galactic archaeology, which aims to reconstruct the assembly history of the Milky Way by understanding each of its accreted building blocks (Eggen et al. 1962; Freeman & Bland-Hawthorn 2002; Helmi 2020), to the studies of dark matter that attempt to locate artifacts created by interactions with dark subhalos in spatially coherent streams (Johnston et al. 2002; Ibata et al. 2002; Yoon et al. 2011; Carlberg et al. 2012; Bonaca et al. 2019, 2020a).

These structures also hold a key to determining the gravitational field and mass distribution at the outer reaches of the Galaxy (Johnston et al. 1999). The trajectories of a group of stars belonging to the same tidally disrupted structure can be used to place constraints on the gravitational field in which they move. This property has been taken advantage of by several authors such as Koposov et al. (2010); Law & Majewski (2010); Küpper et al. (2015); Malhan & Ibata (2019); Erkal et al. (2019) and Vasiliev et al. (2021), mostly using individual structures in the form of spatially coherent streams. Thus far, however, the general investigation has neglected a major source of information: the phase-mixed streams. Since stars retain memory of the progenitor's orbit even after phase-mixing, these types of accreted structures are in principle well-suited for constraining the potential, although in practice they are harder to detect than their spatially-coherent counterparts and their membership is more difficult to discern. More pressing, however, is the need to transition from fitting single tidal structures to fitting several simultaneously. As individual structures are limited in their phase space coverage and therefore pose an increased tendency to lead to biased predictions (Bonaca et al. 2014; Reino et al. 2022), robust constraints can only be reached by analysis of multiple accreted structures simultaneously.

Progress in these directions has, until recently, been hindered mainly by the lack of precise full 6-dimensional phase space maps of the stellar halo and

by the challenge of analysing multiple tidally disrupted structures simultaneously without excessively inflating the considered parameter space because most methods require modeling the progenitor of each stream in the sample individually. Many of the most widely used methods, such as the orbit-fitting (e.g. Koposov et al. 2010) or particle ejection methods (e.g. Küpper et al. 2012; Gibbons et al. 2014; Fardal et al. 2015), are also unsuitable for fitting phase-mixed streams because they usually compare a model to the stream's track in position and/or velocity space to obtain a figure of merit, while phase-mixed streams by definition are not coherent in these spaces.

The last few years have seen a revolutionary increase in the number of Milky Way stars with precise position and velocity information due to the *Gaia* satellite (Gaia Collaboration et al. 2016a). However, the full 6-dimensional phase space map from *Gaia* data alone is restricted to a relatively nearby region: while proper motion measurements can reach out to 100 kpc, radial velocities and precise distances ($\sim$ 20% precision) are constrained to within 10 kpc from the Sun by the end of mission. Motivated by this disparity, the H3 (Hectochelle in the Halo at High Resolution) Survey (Conroy et al. 2019b) aims to measure the radial velocities and spectrophotometric distances to $\sim$ 300 000 stars in the halo to complement the *Gaia* proper motions beyond 10 kpc. The survey is ongoing, but intermediate data has already been shown to exhibit a high amount of structure by Naidu et al. (2020), who performed a detailed categorization of stars in the Galactic halo in chemodynamical space.

The aim of this work is to place new constraints on the Galactic mass profile by maximizing the global clustering in action space of stars observed by the H3 survey and *Gaia* DR2 (Gaia Collaboration et al. 2018), without reference to their membership in specific streams. The action-clustering method (Sanderson et al. 2015, 2017; Reino et al. 2021, 2022) comes with several advantages: it can be used on multiple tidally disrupted structures simultaneously without adding complexity to the explored parameter space, it is applicable to both spatially coherent and phase-mixed accreted structures, and it does not require information about the progenitor(s) or membership of certain stars in specific streams. Since the stellar halo as viewed by H3 is thought to be mainly composed of accreted structures (Naidu et al. 2020), we apply the action-clustering algorithm to the full set of halo stars observed by H3. While the already detected substructure can help us constrain the Galactic potential, this process can be reversed to detect further substructures in action space by utilizing the robustly measured Galactic potential.

We support our findings with analysis conducted on an ensemble of accreted structures formed in the FIRE cosmological-hydrodynamical simula-

tions (Hopkins et al. 2018). Specifically, we use the simulations to investigate the ways in which measurement errors and interlopers, such as in situ halo stars born in the disc and heated to halo-like orbits by mergers (Bonaca et al. 2017) or formed in outflows (Yu et al. 2020), influence our results.

This paper is organised as follows. In Section 2, we present the method of action-clustering and the potential model. In Section 3 we introduce the H3 survey and discuss the data set we use in our analysis. The results with the H3 survey data are presented in Section 4. In Section 5, we repeat our analysis on simulated streams which, in Section 6, helps us put our H3 results in context. Finally, Section 7 is dedicated to making conclusions and discussing the implications of our results.

## 4.2 Method

We constrain the Galactic potential by employing the action-clustering method following Sanderson et al. (2015), Reino et al. (2021) and Reino et al. (2022). This method exploits the principle that stars belonging to the same accreted structure collapse to a single cluster in action space if the transformation between position-velocity space and action space is performed under the assumption of the true Galactic potential. We determine the potential that best fits our set of stellar data by maximizing the clustering present in action space across all trial potentials.

### 4.2.1 Measuring action-space clustering

We measure the degree of clustering with the Kullback-Leibler divergence (KLD, Kullback 1959), which quantifies the divergence between two probability distributions $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$. The KLD between two identical probability distributions is 0, while an increase in their difference correspondingly increases the KLD value.

In our approach, $p(\boldsymbol{x})$ is set to $p(\boldsymbol{J} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})$, the distribution of actions $J$ transformed from the phase space coordinates of the sample stars, $\boldsymbol{\omega}$, under the assumption of a trial potential parameterized by $\boldsymbol{\zeta}$. This distribution of actions is then compared to a uniform distribution representing a totally featureless action space, in other words, $q(\boldsymbol{x}) = u(\boldsymbol{J})$.

Our first step is to calculate the difference between these two distributions as we vary the trial potential parameters and to select the potential with the highest KLD value as the best-fit potential, with parameters $\boldsymbol{\zeta}_0$, for that particular data set, $\boldsymbol{\omega}$.

In this work, we use two different versions of the KLD depending on the availability of the membership information for each star in our data set. We calculate the standard KLD whenever the membership is undetermined following

$$\mathrm{KLD1}(\boldsymbol{\zeta}) = \frac{1}{N}\sum_{i}^{N} \log \frac{p(\boldsymbol{J} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})}{u(\boldsymbol{J})}\bigg|_{\boldsymbol{J}=\boldsymbol{J}^{i}_{\zeta}} , \tag{4.1}$$

where $N$ is the total number of stars in the sample, $\boldsymbol{J}^{i}_{\zeta} = \boldsymbol{J}(\boldsymbol{\zeta}, \boldsymbol{\omega}_i)$ and $\boldsymbol{\omega}_i$ are the phase space coordinates for star i.

The function $p(\boldsymbol{J} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})$ is constructed from the observed points $\boldsymbol{J}$ via a density estimator, for which purpose we employ the Enlink algorithm developed by Sharma & Johnston (2009). Meanwhile, $u(\boldsymbol{J})$, being constant across all trial potentials, can be set to any preferred value.

The standard KLD gives equal weight to each star in the sample. This arrangement is not always desirable: for example, when analysing multiple structures simultaneously, it would lead to structures with higher star counts influencing the results to a higher degree. In a halo assembled hierarchically from a steep CDM-like mass function, the star counts in different satellites, and hence their influence on the fit, will vary by several orders of magnitude. Therefore, whenever membership information for the stars is available, we can make adjustments to the KLD computation to take full advantage of all available data.[1] First, we modify the equation so that each *structure*, rather than each star, is weighted equally. Second, since the action-clustering method works best without significant overlap between the different structures in action-space, we calculate the probability distributions $p_j(\boldsymbol{J}_j \mid \boldsymbol{\zeta}, \boldsymbol{\omega}_j)$ of each structure (here labeled by $j$) independently. This *weighted* KLD is then calculated as

$$\mathrm{wKLD1}(\boldsymbol{\zeta}) = \sum_{j}^{N_\mathrm{s}}\sum_{i}^{N_j} \frac{1}{N_\mathrm{s}}\frac{1}{N_j} \log \frac{N_j}{N}\frac{p_j(\boldsymbol{J}_j \mid \boldsymbol{\zeta}, \boldsymbol{\omega}_j)}{u(\boldsymbol{J})}\bigg|_{\boldsymbol{J}=\boldsymbol{J}^{ij}_{\zeta}} , \tag{4.2}$$

where $N_\mathrm{s}$ is the number of structures, $N_j$ the number of stars in structure $j$, $\boldsymbol{J}^{ij}_{\zeta} = \boldsymbol{J}(\boldsymbol{\zeta}, \boldsymbol{\omega}_{ij})$ and $\boldsymbol{\omega}_{ij}$ are the phase space coordinates for star i in stream j.

Our second objective is to calculate the confidence intervals on the best-fit potential parameters $\boldsymbol{\zeta}_0$. In this step, we use the KLD to quantify the difference between the action distribution of the best-fit potential, $p(\boldsymbol{J} \mid \boldsymbol{\zeta}_0\,, \boldsymbol{\omega})$ and the action distributions of each other trial potential, $p(\boldsymbol{J} \mid \boldsymbol{\zeta}_{\mathrm{trial}}\,, \boldsymbol{\omega})$. In

[1] However, these adjustments are not necessary for the method to work as we have demonstrated in Sanderson et al. 2015 and Reino et al. 2021.

other words, this set-up allows us to establish how much the action distribution changes as we move further from the best-fit parameters. In Reino et al. (2021) we showed that trial potential parameters that produce an action space with a KLD of or lower than 0.5 from the action space in the best fit potential correspond to $1\sigma$ confidence interval around the best-fit parameter values. This boundary was derived exploiting the interpretation of KLD as average log-likelihood ratio and assuming Gaussian posterior probability distributions. We will report the confidence intervals using this boundary throughout this paper unless otherwise stated.

As with KLD1, we construct both a standard and a weighted version for our calculations, each to be used with the corresponding KLD1. The standard version is defined as

$$\mathrm{KLD2}(\boldsymbol{\zeta}) = \frac{1}{N}\sum_{i}^{N} \log \frac{p(\boldsymbol{J} \mid \boldsymbol{\zeta}_0\,, \boldsymbol{\omega})}{p(\boldsymbol{J} \mid \boldsymbol{\zeta}_{\mathrm{trial}}\,, \boldsymbol{\omega})}\bigg|_{\boldsymbol{J}=\boldsymbol{J}_0^i}\,, \tag{4.3}$$

where, as before, both functions are constructed using the observed actions $\boldsymbol{J}(\boldsymbol{\zeta}_0\,, \boldsymbol{\omega})$ and $\boldsymbol{J}(\boldsymbol{\zeta}_{\mathrm{trial}}\,, \boldsymbol{\omega})$ via Enlink and evaluated at $\boldsymbol{J}_0^i = \boldsymbol{J}(\boldsymbol{\zeta}_0\,, \boldsymbol{\omega}_i)$. The weighted version of KLD2 is computed following

$$\mathrm{wKLD2}(\boldsymbol{\zeta}) = \sum_{j}^{N_s}\sum_{i}^{N_j} \frac{1}{N_\mathrm{s}}\frac{1}{N_j} \log \frac{N_j}{N}\frac{p_j(\boldsymbol{J}_j \mid \boldsymbol{\zeta}_0, \boldsymbol{\omega}_j)}{p_j(\boldsymbol{J}_j \mid \boldsymbol{\zeta}_{\mathrm{trial}}, \boldsymbol{\omega}_j)}\bigg|_{\boldsymbol{J}_i=\boldsymbol{J}_0^{ij}}\,. \tag{4.4}$$

### 4.2.2 Trial potentials

In this work, as in Reino et al. (2021) and Reino et al. (2022), our trial potentials are drawn from a set of oblate axisymmetric two-component Stäckel potentials described by spheroidal coordinates $(\lambda, \nu, \phi)$. $\lambda$ and $\nu$ can be derived from cylindrical coordinates $(R, z, \phi)$ as the roots for $\tau$ in

$$\frac{R^2}{\tau - a^2} + \frac{z^2}{\tau - c^2} = 1\,, \tag{4.5}$$

where $a$ and $c$ are constants, with $a > c$ for oblate density contours, determining the shape of the coordinate system and defining the axis ratio of the coordinate surfaces, $e \equiv \frac{a}{c}$. For more information about the coordinate system we refer the reader to de Zeeuw (1985) and Dejonghe & de Zeeuw (1988).

We construct the full potential by combining two individual Stäckel potentials, $\Phi_{\mathrm{outer}}$ and $\Phi_{\mathrm{inner}}$, each with their own parameters $a_{\mathrm{outer}}$, $c_{\mathrm{outer}}$ and

$a_{\rm inner}$, $c_{\rm inner}$ which are related by the constant $q > 0$ (Batsleer & Dejonghe 1994; Reino et al. 2021, 2022) as

$$\begin{aligned} a^2_{\rm outer} - a^2_{\rm inner} = c^2_{\rm outer} - c^2_{\rm inner} = q \ , \\ \lambda_{\rm outer} - \lambda_{\rm inner} = \nu_{\rm outer} - \nu_{\rm inner} = q \ . \end{aligned} \tag{4.6}$$

The full potential is defined as

$$\begin{aligned} &\Phi(\lambda_{\rm outer}, \nu_{\rm outer}, q) = \Phi_{\rm outer} + \Phi_{\rm inner} = \\ &- GM_{\rm tot}\left[\frac{1-k}{\sqrt{\lambda_{\rm outer}} + \sqrt{\nu_{\rm outer}}} + \frac{k}{\sqrt{\lambda_{\rm outer} - q} + \sqrt{\nu_{\rm outer} - q}}\right] \end{aligned} \tag{4.7}$$

where $M_{\rm tot}$ is the total mass, $k$ is the ratio between the inner component mass and the total mass and G the gravitational constant.

Our decision to adopt the Stäckel potential is motivated by the fact that it is the most realistic potential model that allows for calculation of exact actions. This property is a great advantage when repeated action calculations for large data sets need to be performed as it supports very efficient exploration of the parameter space. Although often considered to be insufficient to describe real galaxies, in Reino et al. (2022) we showed with a galaxy from the FIRE simulations that the two-component Stäckel model recovers the galactic potential precisely and without bias[2].

The first two actions for the Stäckel potential, $J_\lambda$ and $J_\nu$, are defined as

$$J_\tau = \frac{1}{2\pi} \oint p_\tau d\tau \ , \tag{4.8}$$

where $p_\tau$ is the conjugate momentum to the coordinate $\tau = \lambda, \nu$ and the integration is over the full oscillation in either coordinate. The third action $J_\phi$ is a constant in our axisymmetric potential and equal to $L_z$, the z-component of the angular momentum.

We select our two-component trial Stäckel potentials on a grid of five parameters $\boldsymbol{\zeta} = (M_{\rm tot}, a_{\rm outer}, a_{\rm inner}, e_{\rm inner}, k)$. We draw 21 points for each

[2]A least squares fit of the Stäckel model to the true velocity curve of the simulated galaxy recovers the true velocity curve within 1.2% over the range of galactocentric distances where we had stellar data (10-176 kpc). Thus, the Stäckel model can approximate the galactic potential well and the bias originating from our choice of potential is likely negligible. When applying the action-clustering method, with the assumption of a Stäckel model, on two long streams we recovered the true rotation curve within 6.5% between the 5 to 95-percentile stellar distance range (23 - 109 kpc). These results show that only small discrepancies between the best-fit potential and the true potential, all of which well within the associated uncertainties, could be attributed to the choice of Stäckel model.

$M_{\rm tot}$ and $a_{\rm outer}$, and 10 points for each of $a_{\rm inner}$, $e_{\rm inner}$ and $k$ from uniform distributions over the following ranges: $[6 \times 10^{11}, 2 \times 10^{12}]$ in $M/M_{\odot}$, [15, 55] in $a_{\rm outer}$/kpc, [1, 5] in $a_{\rm inner}$/kpc, [1.1, 50] in $e_{\rm inner}$ and [0.05, 0.3] in $k$. In total, our grid contains 441 000 trial potentials. During the course of our analysis, we however discard any potentials that cause more than 10% of the stars in the analysed sample to be unbound from the Galaxy.

Figure 4.3.1 illustrates the set-up of our action-clustering method. Each of the three top panels shows the action space of two simulated streams calculated under the assumption of three different Stäckel potentials, while the lower panels show the circular velocity of each of the trial potentials as compared to the true circular velocity of the host galaxy. The left and right panels both assume trial potentials that are not good representations of the true potential resulting in diffuse action space clusters. Instead, the middle panel adopts a trial potential that well represents the true potential and thus gathers the stars into clearly defined clusters in action-space. The KLD1 values in the top right corner of the top panels quantifies and reinforces this visual inspection: the higher the KLD1 value, the more clustered the distribution is.

## 4.3 Data

The Hectochelle in the Halo at High Resolution (H3) Survey (Conroy et al. 2019b) is a ground-based spectroscopic survey intended to complement *Gaia* data (Gaia Collaboration et al. 2016a) for the stellar halo. Specifically, it will target $\sim 300\,000$ halo stars measuring their radial velocities with a precision of $\lesssim 1$ km s$^{-1}$, spectrophotometric distances with a precision of $\lesssim 10\%$, and metallicities with a precision of $\lesssim 0.1$ dex (Conroy et al. 2019c). Combined with *Gaia* proper motions, the output of the H3 Survey is the full 6D phase space and chemical composition information for the stellar halo.

The selection function of the primary sample consists of the following criteria: (i) $15 < r < 18$, where the photometric magnitude r refers to the Pan-STARRS r-band (Chambers et al. 2016), (ii) $\pi - 2\sigma_{\pi} < 0.5$ (later updated to $\pi < 0.4$) based on *Gaia* DR2 parallaxes (Gaia Collaboration et al. 2018), (iii) $|b| > 30^{\circ}$ to avoid the Galactic plane (although thus far all measured fields are $|b| > 40^{\circ}$), (iv) Dec. $> -20^{\circ}$ to account for the visibility from the MMT in Arizona, USA.

Observations started in 2017 and are ongoing till present day. In this paper, we work with the data collected up to July 2020 which contains more than 130 000 stars. We concentrate on giant stars ($\log g < 3.5$) with a

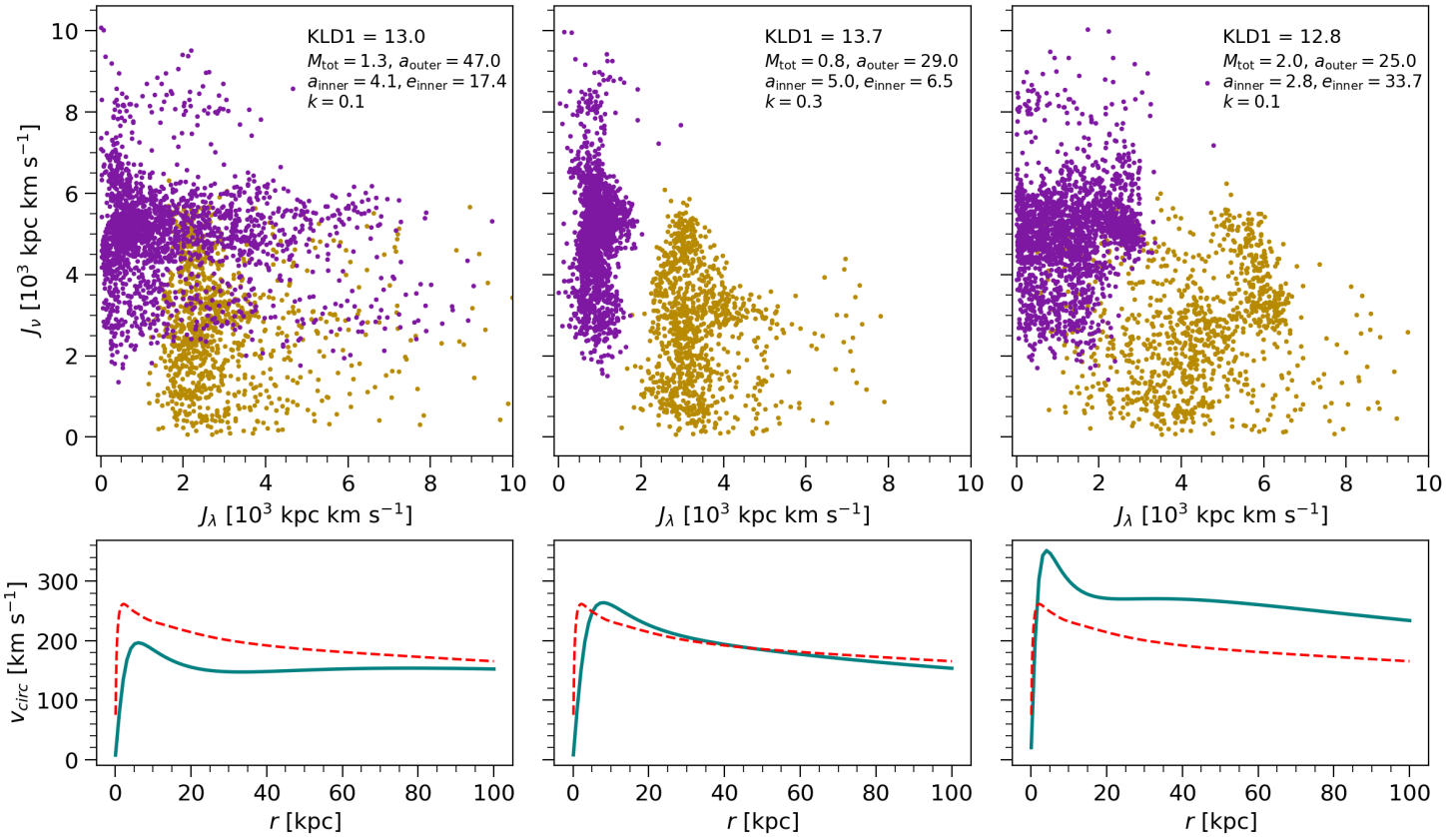


**Figure 4.3.1:** Illustration of the application of the action-clustering method with the degree of clustering measured by the Kullback-Leibler Divergence. The top panel shows the action space of two simulated streams in three different two-component Stäckel trial potentials, and each lower panel compares the circular velocity of the trial potential (teal line) to the true potential (red dashed line) of the simulated host galaxy. The trial model parameters alongside the KLD1 value corresponding to each model is given on the top right corner in the top panels. This figure shows that the highest KLD value (central panel) corresponds to the potential that causes the most clustered distribution of stars in action-space and better reproduces the true potential.

| Structure | $N_*$ | $r$ [kpc] |
|---|---|---|
| Gaia-Sausage-Enceladus | 2684 | 14.4 |
| Sagittarius | 612 | 26.9 |
| Helmi streams | 91 | 15.7 |
| Sequoia | 72 | 14.6 |
| Thamnos | 32 | 8.9 |
| Aleph | 122 | 11.5 |
| I'itoi | 65 | 14.0 |
| Arjuna | 139 | 16.7 |
| Wukong | 111 | 12.9 |

**Table 4.1:** List of accreted structures as determined by Naidu et al. (2020). The columns give for each structure their name, number of stars $N_*$ and median Galactocentric distance $r$.

high signal-to-noise ratio and high-quality stellar parameters (SNR $> 3$ and quality flag $= 0$). We further refine our sample to stars that belong to the kinematic halo by requiring that $|\boldsymbol{V} - \boldsymbol{V}_{\rm LSR}| > 220$ km s$^{-1}$. These conditions restrict our sample to 5162 giants.

Alongside this sample of halo giants, we make use of the selection of stars outlined in Naidu et al. (2020). These authors identify both known and new structures in the H3 halo data and find that about 75% of the stellar halo has been built up by accreted dwarf galaxies. Here, we select the stars assigned by Naidu et al. (2020) to be likely members of Gaia-Sausage-Enceladus (GSE), Sagittarius stream, the Helmi streams, Sequoia, Thamnos, Aleph, I'itoi, Arjuna and Wukong to make up our accreted structures sample, consisting of 3928 stars. A summary of this data can be found in Table 4.1 and a visual representation in energy - angular momentum in Figure 4.4.2. The essential advantage of this selection is that, due to the accompanying membership information, it allows us to sort stars into different structures during our analysis and employ the weighted KLD scheme.

The majority of the stars (2941) are common between the halo giants and the accreted substructures samples. Although the basis of Naidu et al. (2020) inventory of structures is a sample of halo stars selected in much the same way as our halo giants sample - giants with high signal-to-noise ratio and high-quality stellar parameters - there are some key differences. First, Naidu et al. (2020) analysed the H3 survey data that was collected as of March 2020, while our sample extends to July 2020. Second, the kinematic selection we perform to remove stars on disc-like orbits to form our halo giants sample

was not performed by Naidu et al. (2020), who identified stars belonging to the disc following more detailed chemodynamical arguments. Third, Naidu et al. (2020) included in their basis sample a small number of special targets categorized as K giants, BHB or RR Lyrae, while we restrict our analysis to the main sample only. And, finally, we have removed from our accreted structures list those stars which Naidu et al. (2020) designated as belonging to the high-$\alpha$ disk and in situ halo, the metal-weak thick disk or unclassified debris, as our main interest was in the substructure found in the halo region.

We intend to analyse both the halo giants and the accreted structures samples in parallel, the former with the standard and the latter with the weighted KLD version.

To convert from the Heliocentric observables to the Galactocentric frame, we adopt the following parameters: $R_\odot = 8.3$ kpc in the direction of the negative x-axis, $Z_\odot = 27$ pc and $\boldsymbol{v}_\odot = [v_{x,\odot}, v_{y,,\odot} v_{z,\odot}] = [11.1, 232.24, 7.25]$ km s$^{-1}$.

## 4.4 Results

We present the results of our action-clustering analysis on Figure 4.4.1, where we show the circular velocity profiles of the best-fit Stäckel potentials for both the halo giants and the accreted structures samples. These can be compared to other known potentials of Bovy (2015) (dashed black line) and McMillan (2017) (dashed cyan line) and to data from Eilers et al. (2019) (grey dots). Although we show the best-fit curves over the full Galactocentric distance range of 50 kpc, our stellar data mostly covers Galactocentric distances from about 8 to 32 kpc (the thicker line sections on the best-fit rotation curves indicate the 5 to 95-percentile distance range for the stars in each sample). This distance range therefore constitutes the true scope of our results. Furthermore, we demonstrated in Reino et al. (2022) that the potential is most reliably recovered at the median Galactocentric distance of the stars in the sample, shown on Figure 4.4.1 with a cross.

The best-fit potentials of the two data sets are in agreement at their median distances: the halo giants sample predicts $v_{\rm circ}(15.7\ {\rm kpc}) = 238.6$ km s$^{-1}$ while the accreted structures sample prefers a marginally higher $v_{\rm circ}(15.4\ {\rm kpc}) = 240.4$ km s$^{-1}$. There is some variation in the recovered potentials throughout the whole distance range covered, especially at lower Galactocentric radii, the maximum difference of $\sim 45$ km s$^{-1}$ occurring at a Galactocentric distance of $\sim 8$ kpc. However, both estimates appear to be somewhat higher that the measurements from (Bovy 2015), McMillan (2017) and Eilers et al. (2019). The angular momentum - energy diagrams corresponding to these best-fit

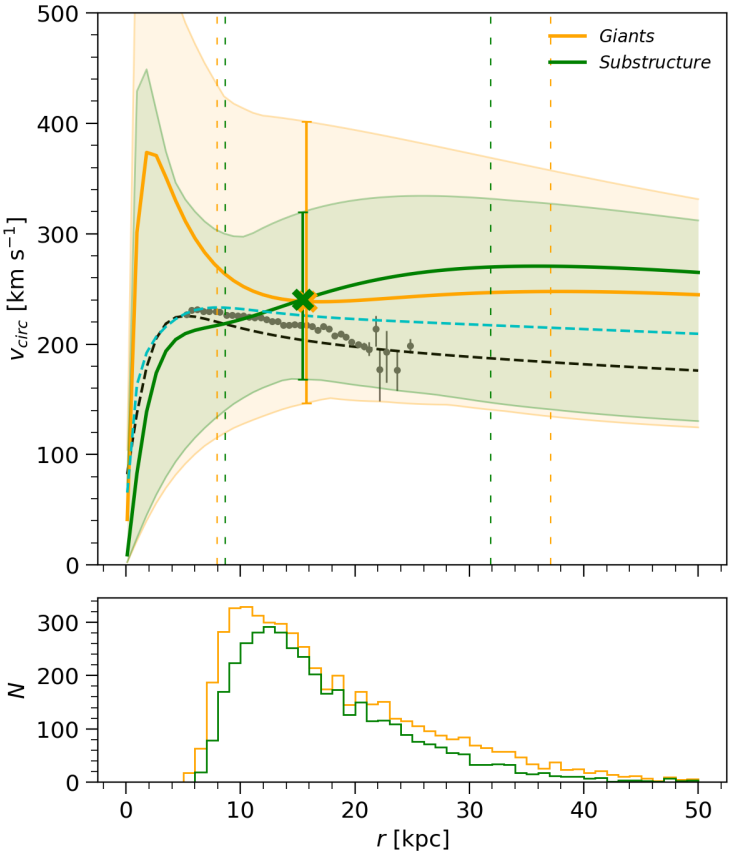


**Figure 4.4.1:** Results for the halo giants (yellow) and accreted substructure (green) samples in circular velocity space. The top panel shows the best-fit Stäckel potential from each data set with the solid coloured lines and the corresponding $1\sigma$ uncertainty regions with the shaded area. The median Galactocentric distance of the stars in each data set is marked with a cross and the vertical dashed lines show the 5 to 95-percentile distance range covered by stars in each sample. We also show the rotation curves from the `galpy MWPotential2014` (Bovy 2015) (dashed black line) and McMillan (2017) (dashed cyan line) and data from Eilers et al. (2019) (grey dots), for comparison. The bottom panel shows the histogram of star counts over this distance range.

potentials are shown in Figure 4.4.2.

The $1\sigma$ confidence regions on the best-fit potentials are shown as the coloured shaded regions on Figure 4.4.1. Both the halo giants and accreted structures samples exhibit strikingly large uncertainty regions relative to the deviation between the best fit and those from other work, with $v_{\rm circ}(15.7\,{\rm kpc}) = 238.6^{+163.0}_{-92.1}$ km s$^{-1}$ for the halo giants sample and $v_{\rm circ}(15.4\,{\rm kpc}) =$

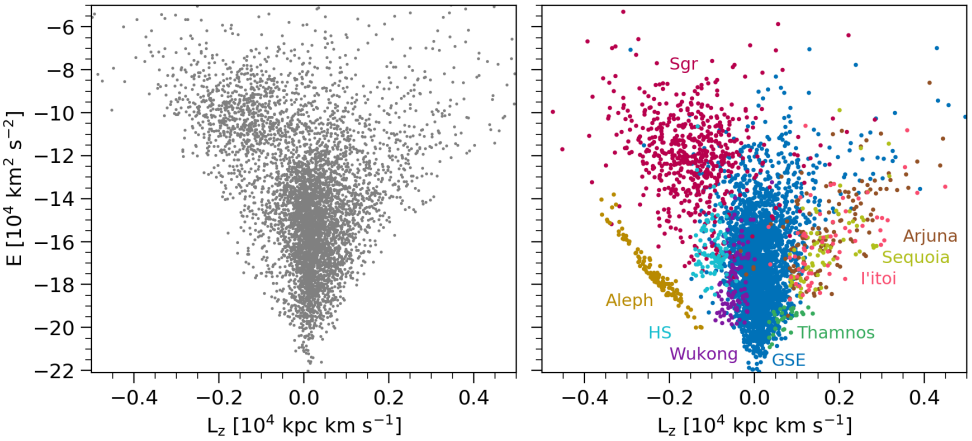


**Figure 4.4.2:** Energy - angular momentum diagrams for the halo giants sample (left) and accreted structures sample (right) in their corresponding best-fit potentials.

$240.4^{+79.3}_{-72.2}$ km s$^{-1}$ for the accreted structures sample. These extensive uncertainties could be induced by several possible causes that we explore in the next section: measurement errors, contamination (from remaining in situ component or other debris) in the sample, or misclassified members.

Intuitively, measurement errors on phase space coordinates should have the effect of diluting the action space clusters over a larger volume in action space for any trial potentials. This does not necessarily affect the determination of the best-fit potential, but rather lowers the maximum clustering, or KLD1, that could be achieved with the data set. However, the increased volume of the action-space clusters could mean that more trial potentials now produce an action space which is considered similar to that of the best-fit potential. In other words, this could induce an increase in the number of trial potentials with KLD2 $< 0.5$.

The presence of interlopers adds a randomly scattered background of stars to the action space, thus also innately lowering the maximum possible KLD1. Since interlopers are unlikely to condense into tight clusters irrespective of the trial potential, their inclusion to the data should be analogous to a overall decrease in signal to noise ratio. Thus, although the best-fit potential is not likely to be affected by interlopers, the noisy background they create does decrease the differences between the action distributions of different trial potentials leading, once again, to an increase in the number of potentials with KLD2 $< 0.5$.

We will explore these two effects as possible causes of these extended

uncertainty regions, in addition to discussing and validating an alternative approach to determining our confidence in our best-fit results, in detail in the next section.

## 4.5 Simulations

This section is dedicated to considering the possible sources for the large uncertainty regions we encountered in the previous section. Our intention here is not to deliver a detailed solution to the H3 samples described above, but rather to provide a proof of concept by analogy. For this purpose we use structures formed from the tidal disruption of dwarf galaxies in cosmological-baryonic simulations of Milky Way-like galaxies from the Latte suite (Wetzel et al. 2016) of the Feedback In Realistic Environments (FIRE-2) project[3] (Hopkins et al. 2018), which was recently made public Wetzel et al. (2022). Panithanpaisal et al. (2021) conducted a search for stellar-stream-like accreted structures across these simulated galaxies and catalogued not only the structures classified as coherent stellar streams, but, as a by-product, also those determined to be either phase-mixed structures or dwarf galaxies yet to be strongly influenced by tidal disruption.

Here, we focus on the structures identified in the halo of the isolated[4] galaxy simulation m12i. We chose this galaxy because it has a relatively large number of coherent streams at present day, giving a good variety of different orbits and masses to test, and because it has a relatively quiet recent merger history, avoiding (for now) the complications and systematic biases in modeling the potential introduced by interactions (e.g. Vasiliev et al. 2021). The rotation curve and disk structure of this simulated galaxy are close to those of the Milky Way (Sanderson et al. 2020). Out of the 17 catalogued structures, we discard all four dwarf galaxies[5] and any structures that contain fewer than 200 star particles after the removal of any gravitationally-bound remnant. The median star particle mass in the simulation is about 5000 solar masses, so this cutoff corresponds to a minimum stellar stream mass of about $10^6$ $M_\odot$. After applying this selection, we retain seven structures: five spatially coherent stellar streams and two phase-mixed streams. To keep all structures approximately the same size, we subsample 4500 star particles from each of the two largest structures. The details of this data set are summarized in Table 4.2.

[3] `http://fire.northwestern.edu`

[4] no massive companion within $\sim 5$ Mpc

[5] The dwarf sample is incomplete, containing only dwarf galaxies with $M < 10^6 M_\odot$ accreted more than 2.7 Gyr ago.

| Structure | Structure type | $N_*$ | $r$ [kpc] |
|---|---|---|---|
| Str1 | Phase-mixed stream | 1067 | 56.0 |
| Str2 | Phase-mixed stream | 9689 (4500) | 69.4 |
| Str3 | Coherent stream | 79774 (4500) | 57.2 |
| Str4 | Coherent stream | 4456 | 112.1 |
| Str5 | Coherent stream | 2063 | 37.7 |
| Str6 | Coherent stream | 4038 | 78.8 |
| Str7 | Coherent stream | 522 | 117.8 |

**Table 4.2:** List of structures from the FIRE simulation m12i used in this work. The columns give for each structure their name, type, number of stars $N_*$ and median galactocentric distance $r$.

This sample represents an idealised version of the situation: a halo entirely comprised of accreted structures with star particles for which we know the true error-free present day positions and velocities. In what follows, we start adding complexity to this pure sample to determine if and how this affects our results.

### 4.5.1 Interlopers

First, we construct a sample of interlopers in the form of "in situ halo" stars: stars formed in the inner galaxy that now follow halo-like orbits. As they are sourced from a roughly equilibrium distribution, we expect that such stars should have no discernible phase-space clustering. To do this, we select all stars that formed at a distance of less than 10 kpc from the centre of the m12i galaxy across all times, which ensures the stars' in situ origin. In addition, we select stars within 200 kpc of the main galaxy at the present day, and implement a cut in total energy which allows us to remove the stars that remain in the disc and bulge at present day. This procedure results in a set of $\sim 30\,000$ "in situ halo" stars, a collection from which we can randomly sample stars to add as interlopers to our pure structures data set.

The top panel in Figure 4.5.1 depicts the effect of adding more and more interlopers to a single stream: a pure sample of Str5 stars. A comparison between results from the clean sample (blue line - covered partly by the golden line at low galactocentric radii and by the pink line at high galactocentric radii - and shaded regions) and the true potential of the host galaxy (red dashed line), derived from the spherically binned total mass profile of the simulation snapshot, reveals that at the median distance (shown with the blue cross) the Str5 sample recovers the true potential within 10.5 km $s^{-1}$. The full

range of the uncertainty region at this distance is 35.2 km $s^{-1}$. Next, we add randomly selected interlopers from our in situ halo sample to this clean data, in the ratios of 1:1 (teal), 1:3 (golden) and 1:5 (pink) true member stars to interlopers. Despite the ever increasing number of included interlopers, all three groups recover the true potential with ease: the best-fit potentials predict velocities that within 9.3, 7.3 and 9.3 km $s^{-1}$, respectively, from that of the true potential at the median distance of Str5. The recovery of the true potential is thus largely unaffected by the inclusion of interlopers. However, their influence is substantial on the associated uncertainty regions: at the median distance of Str5, the coverage of the uncertainty regions extends to 61.3, 111.6 and 129.9 km $s^{-1}$, respectively.

### 4.5.2 Measurement errors

Second, we explore the effect of phase space measurement errors by modifying the positions and velocities of the stars in our clean sample. We run this test on the entire set of 7 streams. The expected size for these measurement uncertainties is determined based on the H3 accreted structures sample by calculating the median relative error across the whole sample for each Galactocentric Cartesian position and velocity component.[6] The new positions and velocities for the simulation star particles are then drawn from a normal distribution centred on their true values with a standard deviation determined by the product of the true values and the corresponding median relative errors from H3.

The middle panel of Figure 4.5.1 presents the comparison between error-free and error-convolved samples, where we show the result obtained with our full error-free pure structures sample, i.e. a sample that includes all seven structures in Table 4.2, (blue line) and the result obtained with the error-convolved version of the same sample (pink line). The inclusion of errors has only a marginal effect on our best-fit results: the best-fit curve of the error-free data is within 5.6 km $s^{-1}$ of the true potential at the median distance while the best-fit curve of the error-convolved data lies within 14.2 km $s^{-1}$. Measurement errors are therefore unlikely to significantly affect the accuracy of potential inference. The confidence regions, however, are strongly affected. The full uncertainty range at median distance increases from 28.2 km $s^{-1}$ of the error-free sample to 107.7 km $s^{-1}$ for the error-convolved sample. To illustrate the relative impact of interlopers and measurement errors, we also

[6]Since the distance uncertainties dominate, this over-estimates the measurement error. However, given that most lines of sight in the survey are nearly parallel to the vertical direction, the over-estimate is relatively small.

show the result of the error-free interloper-filled all-structures sample, where in situ halo stars where added to reach the ratio of 1:1 true members to interlopers (golden line). H3-like measurement errors thus affect the confidence regions to a larger extent than the presence of interlopers does, at least as long as the amount of interlopers does not exceed that of the true accreted structure members.

We also confirm that the inclusion of interlopers or measurement errors has the effect of lowering the amount of clustering, or KLD1, across all trial potentials. The maximum KLD1 of the pure error-free data set is 13.7, while the maximum KLD1 for the contaminated and error free samples is 12.8 and 12.4, respectively. In addition, we see a reduction in the full range of KLD1 values across trial potentials for the contaminated and error-convolved samples: when $\Delta$KLD1 is 1.5 for the pure error-free data, it is 1.0 and 0.8, respectively, for the contaminated and the error-convolved samples. This means that there is less variation in the amount of clustering present in the action space between different trial potentials and, therefore, that it is more difficult to distinguish between them. The impact of the measurement errors is, however, greater, as is also reflected in the KLD2-based confidence regions.

### 4.5.3 Leave-one-out cross-validation

We conclude that neither the presence of a considerable amount of interlopers nor the inclusion of measurement errors individually impact the accuracy with which the action-clustering method recovers the galactic potential. The impact on the associated confidence regions is, however, significant. In the ensuing section we show the influence of the combination of these features and use a leave-one-out cross-validation (LOOCV) to determine whether the extended confidence regions are justified by the data.

We first run the action-clustering method on a set of stars where both of the two effects are present: we add interlopers to our sample of accreted structures in the ratio of 1:1 members to interlopers, as previously, and subsequently draw new positions and velocities as prescribed above, to this full sample. The result of our analysis is shown in the bottom panel in Figure 4.5.1 with the pink line, which can be compared to the error and interloper-free result shown in blue (identical to the result shown in blue in the middle panel). While the best-fit potential shows little variation, the KLD-based confidence regions clearly reflect the combined power of the two effects having, in this instance, expanded even further. The comparison of the energy - angular momentum diagrams of these two data sets in the best-fit potential of the clean sample (blue line in Figure 4.5.1) is shown in Fig-

ure 4.5.2 and illustrates the effect that both interlopers (yellow points on the right panel) and measurement errors have on the structure cluster sizes.

We now create seven new samples from this interloper-filled error-convolved sample, by leaving out one of the seven structures per sample (and a corresponding number of interlopers to retain the 1:1 ratio), and apply our action-clustering method to each. The results are presented with the grey lines in the bottom panel of Figure 4.5.1 and show that the variation in the leave-one-out fits is significantly lower than the full KLD-driven confidence region for the same sample would dictate. Instead, this variation appears to be of the order of magnitude of the confidence region obtained with the error and interloper-free data set (blue shaded region). At median distance, we obtain $172.8^{+12.8}_{-15.3}$ km s$^{-1}$ with the error and interloper-free data using the KLD-based confidence regions and $189.3^{+4.0}_{-19.5}$ km s$^{-1}$ with the interloper-filled error-convolved data using LOOCV.

We conclude that while the KLD-based uncertainties are a good estimate of the confidence in the best-fit results when the analysed data set consists solely of accreted structures with negligible measurement errors, this framework no longer suffices in the presence of a significant number of interlopers and/or larger measurement errors, in which case the KLD-based confidence regions are greatly overestimated relative to the variation in the best-fit potential among the different accreted structures comprising the data set. In such situations LOOCV gives a more appropriate gauge of the uncertainty on the best-fit potential. When membership information is available, this procedure is straightforward; however when it is unavailable, progress could be made by dividing the best-fit action space into contiguous volumes each containing equal numbers of stars, and leaving out the contents of each volume successively.

All results in this section were obtained with the standard KLD version that does not rely on membership information. This was made necessary by the inclusion of interlopers which were not assigned membership to any particular structure. The constraints here are therefore more conservative than if the weighted version had been applied.

## 4.6 Leave-one-out cross-validation in the H3 sample

In the previous section, we showed with simulated data that while contamination and measurement uncertainties in the analysed samples do not have a marked effect on the accuracy with which we can recover the true potential,

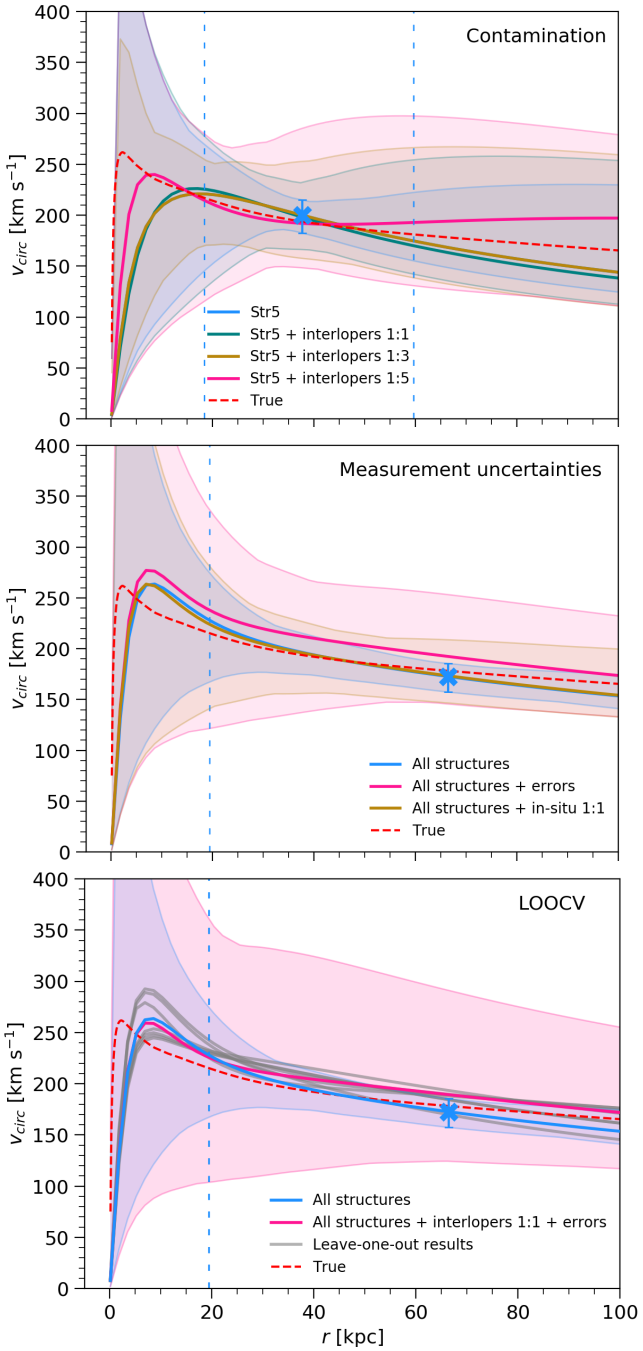


**Figure 4.5.1:** Results of accreted structures from FIRE simulations. The solid lines show the best-fit potentials and the shaded areas the KLD-based $1\sigma$ confidence regions. The red dashed line gives the true rotational velocity. The cross marks the median galactocentric distance of the error- and interloper-free data set, while the vertical dashed lines show the 5 to 95-percentile distance range covered by stars in that sample (in the middle and bottom panels, the higher bound is outside the figure range). Top panel: analysis of different levels of interlopers in the sample data. Middle panel: analysis of data with measurement errors. Bottom panel: LOO analysis of the sample with measurement errors and interlopers.

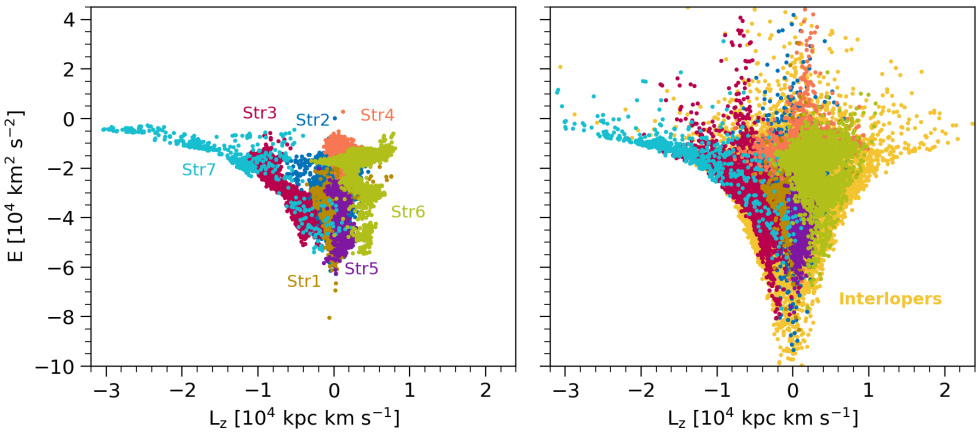


**Figure 4.5.2:** Energy - angular momentum diagrams for accreted structures in FIRE simulations: the pure error-free sample (left) and the error-convolved sample with added interlopers in 1:1 ratio (right), both in the best-fit potential of the unmodified, clean sample corresponding to the blue line in the middle and bottom panels of Figure 4.5.1.

the confidence regions associated with the inference are greatly affected.

The level of contamination we assumed when conducting these tests was much more extreme than we would reasonably expect to find in either of our H3 samples. Nevertheless, contaminants likely make up not an insignificant portion of our halo sample. Naidu et al. (2020) have argued that $\sim 75\%$ of the stellar halo is made up of stars from accreted substructures which—although they have used a slightly different data set to reach this conclusion as discussed in Section 4.3—leads us to estimate that approximately 1 in every 4 stars in our halo giants sample could be an interloper. The source of this contamination could be stars that really belong to other dynamical groups such as the in situ halo and the thick disk, but also stars that were indeed fed to the Galaxy by accretion events but which, possibly due to selection effects, do not appear in sufficient numbers per structure to contribute to the clustering.

Although the substructure sample was more carefully selected by Naidu et al. (2020) following chemodynamical arguments, which allowed stars belonging to groups such as the thick disk and in situ halo to be discarded, the membership of each accreted structure was generally determined using a hard-edged selection in just two to three parameters. For example, the parameters most often used for selection were total energy, angular momentum in the z-direction and metallicity. This leaves room for *misclassification*: the erroneous assignment of stars with a subset of similar parameters to a

particular structure. The effect of misclassified stars is distinct from that of interlopers from a broad and relatively smooth distribution. Structures that have been identified and selected based on their position in energy and angular momentum space have a higher chance of yielding poor constraints on the potential. This is because misclassified stars are likely to overlap with the true members in action space thus inflating the minimum cluster size.

Furthermore, as shown in the previous section, the measurement uncertainties that are typical to the H3 data have an effect of significantly inflating the confidence regions around the best-fit potential, even without any contamination in the sample. Therefore, as an alternative view on the confidence that we should have in our results, we apply LOOCV on the H3 substructure sample. Since LOOCV requires knowledge of structure membership, we do not apply the same analysis to the general halo giants sample.

The result of applying LOOCV to the H3 accreted structures sample is presented in Figure 4.6.1, where the best-fit potential from the full sample is shown with a green line and the results from LOOCV with grey lines. The KLD-based confidence region (green shaded area) is significantly larger than the variation indicated by LOOCV across all Galactocentric radii. At median distance, the KLD-based confidence gives $v_{\rm circ}(15.4\,{\rm kpc}) = 240.4^{+79.3}_{-72.2}$ km s$^{-1}$, while the LOOCV indicates $v_{\rm circ}(15.4\,{\rm kpc}) = 240.4^{+0.0}_{-21.9}$ km s$^{-1}$.

Two further features of these results require comment. First of these is the apparent overestimation of mass at higher Galactocentric radii. This feature persists across all the leave-one-out results and thus does not arise due to the influence of a single structure. There are several possible reasons for this. Due to our viewpoint of the Galaxy it is likely that we observe a portion of each accreted substructure that is closest to the Galactic centre, i.e. the pericentre portion of each group. In Reino et al. (2022) we showed that pericentre sections of coherent stellar streams lead to a systematic bias in the results, specifically, such data overestimates the mass of the Galaxy at all Galactocentric radii. The nature of the H3 data is therefore likely to trigger just such a bias in the present results. An additional bias could arise due to the tendency of measurement errors to be larger for more distant stars. This would create a gradient of minimal cluster size across each accreted substructure: the closer stars with smaller measurement uncertainties would form a tighter action-space knot while the further stars with higher measurement errors would be spread out with larger gaps between them. Therefore, a trial potential that manages to concentrate the further stars to a stronger degree than the closer stars, i.e. a potential that overestimates the mass at higher Galactocentric radii, is preferred. Finally, as mentioned above, selection of structure members in the space of energies and/or angular momenta

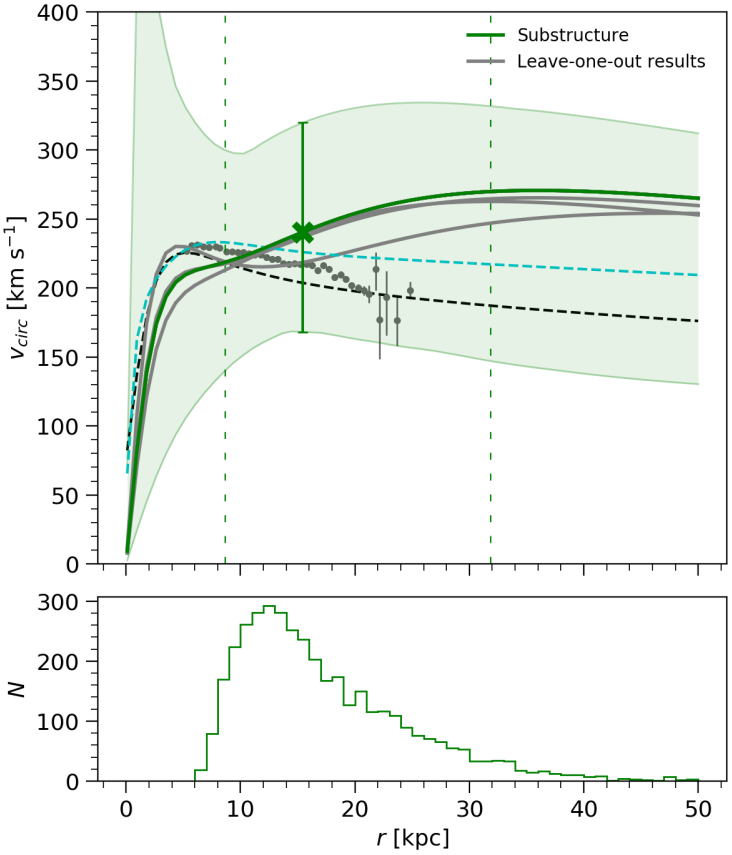


**Figure 4.6.1:** Leave-one-out analysis of the H3 accreted structures sample. The full sample result is shown with the green line and its associated KLD-based uncertainty with the green shaded region (identical to those in Figure 4.4.1), while the leave-one-out results are shown with the grey lines. As in Figure 4.4.1, we show, for comparison, `galpy` `MWPotential2014` (Bovy 2015) (dashed black line) and McMillan (2017) (dashed cyan line) and data from Eilers et al. (2019) (grey dots). The bottom panel shows the histogram of star counts over this distance range, while the cross and the dashed vertical lines in the top panel illustrate the median distance and the 5 to 95-percentile distance range, respectively.

is likely to reduce both the accuracy and precision of the fit: the precision by including interlopers on similar orbits to the stream stars, and the accuracy by potentially focusing on a biased subsample of stream orbits.

The second is the number of leave-one-out samples which do not divert from the full sample fit. Specifically, we find that the omission of each of 6 substructures (GSE, Sagittarius, Sequoia, Aleph, Wukong, Arjuna) will have

no impact on the results: we obtain the same best-fit potential with each of these data sets as we would with the full sample. Exclusion of any of the rest of the three structures (I'itoi, Thamnos, Helmi streams), only induces a small change in the best-fit potential. This behaviour indicates that there is a high level of agreement between the data of different structures but also that the fit is likely driven by the smaller structures (in terms of action-space volume) rather than the larger structures, such as GSE and Sagittarius. This is confirmed when comparing the maximum KLD1 values (signifying the average degree of clustering) of leave-one-out results with the KLD1 of the best-fit potential of the full sample: the maximum KLD1 increases the most when either the Sagittarius or, especially, GSE are removed from the sample, while the maximum KLD1 decreases when Aleph or Thamnos are omitted. GSE and Sagittarius are therefore the least informative structures for this method in our data set.

## 4.7 Conclusions and Discussion

In this work, we measured the mass profile of the Milky Way by applying the action-clustering method – implementing the Kullback Leibler Divergence to quantify the the degree of clustering – to the halo stars from the H3 survey. In particular, we analysed two data sets: a general sample of halo giants and a sample made up of accreted structures as identified by Naidu et al. (2020). The halo giants sample yields $v_{\rm circ}(15.7\ {\rm kpc}) = 238.6^{+163.0}_{-92.1}$ km s$^{-1}$ while the accreted structures sample yields $v_{\rm circ}(15.4\ {\rm kpc}) = 240.4^{+79.3}_{-72.2}$ km s$^{-1}$, the uncertainties of both estimates being strikingly large.

To help us interpret these results, we further analysed accreted structures in a simulated MW analog from the FIRE suite. We compared the results obtained with a clean error-free sample of purely accreted structures to those obtained after adding contamination in the form of in situ halo stars and/or H3-like measurement errors. We found that neither the presence of interlopers nor inclusion of measurement errors significantly reduces the *accuracy* with which we can recover the true potential of the simulated host galaxy. However, both components have a profound effect on the confidence interval around the best fit, inflating the error bars and therefore causing the inferred *precision* of the constraints to suffer.

As an alternative method of error estimation, we propose using leave-one-out cross-validation (LOOCV), whereby the action-clustering method is reapplied to samples from which a single accreted structure has been removed in each iteration and the variation in the best-fit results across all samples

is assigned as the uncertainty in the full sample result. Applying LOOCV to the simulated sample reveals that the KLD-based uncertainties significantly overestimate the errors with respect to the results of LOOCV. We instead find that the uncertainties inferred from LOOCV of error-convolved, interloper-filled data approach the KLD-based uncertainties of the error-free clean sample. KLD-based error estimation is therefore only applicable to data sets with few interlopers and small measurement uncertainties, while in other circumstances, as long as the membership information, albeit imperfect, is available, LOOCV is preferred.

Applying LOOCV thus allowed us to reassess the uncertainty in our H3 accreted structures sample, which we found to be $v_{\rm circ}(15.4 \text{ kpc}) = 240.4^{+0.0}_{-21.9}$ km s$^{-1}$. Although we consider this estimate the more reliable one, there are a few possible sources of bias which we have discussed in Section 4.6.

Here we summarise and reflect on the main elements that influence our potential inference and their implications for the future:

- **Interlopers**. We have found that even a large amount of contamination in the data set does not have a detrimental effect on the accuracy of potential recovery. However, as the ratio of interlopers to accreted structure stars increases, so do the KLD-based uncertainties associated with the potential measurement, i.e. the precision of the constraints deteriorates. Thus, in a situation where the contamination is expected to be significant and the membership information is available, one should instead apply LOOCV to get a sense of the uncertainty in the best-fit potential. This suggests that the current KLD-based method of deriving uncertainties is mostly suitable for use with well-defined coherent stellar streams and not with the stellar halo as a whole. To use the action-clustering method with such a data set, a different approach to error estimation is needed.

- **Measurement errors**. The impact of phase space measurement errors is similar to but even more pronounced than the impact of the interlopers: the potential recovery remains reliable while the associated uncertainties become unjustifiably inflated. This indicates even more resolutely that KLD-based errors are not universally appropriate: although reliable uncertainty is, in principal, obtained when observational uncertainties in the phase space data are small, in practice our objects of interest, stars from accreted structures, reside at large distances out in the halo where measurement uncertainties are significant. Therefore, once again, as long as even partial membership information is available, it is best to utilize LOOCV.

- **Misclassification of member stars**. Coherent stellar streams have the simple yet powerful benefit of being identifiable as clear structures among others in position and/or velocity space. Phase-mixed streams have lost this coherence and therefore their detection and identification relies on more approximate methods, where the selection is determined by a cut in two or three parameters, such as energy, angular momentum and metallicity. These approximate membership selection methods can, however, lead to field stars being misclassified as structure members or assigned to incorrect structures, thus complicating parameter estimation. An additional problem arises when the selection is performed specifically in energy or angular momentum space. Stars that are misclassified as members in this way are likely to overlap with the true members also in action space, causing the minimal cluster size to increase which, in turn, leads to poor potential inference.

- **Pericentre bias**. Our position in the Galaxy makes it inevitable that, as we observe the stellar halo, we predominately observe the stars that are closest to the Galactic centre and, therefore, likely at their pericentre phase. In Reino et al. (2022) we demonstrated with simulations that coherent stellar streams at their pericentre phase are less informative than their counterparts on intermediate or apocentre phase and showed how the quality of the constraints suffers when the analysed data set is fully made out pericentre phase stars. We estimate this bias to be about $\sim 10\%$ of the measured velocity curve at median distance. However, since this is based on simulated coherent streams without any measurement errors, it should be considered an indicative estimate. For accurate high confidence constraints we need to probe deeper into the halo so that stars on intermediate and apocentre phases make up a more prominent fraction of our observations.

We therefore conclude that the robustness of the constraints on the Galactic potential can be increased either by improvements over the current KLD-based method for error estimation or by the growing number of precision observations of the stellar halo.

Figure 4.7.1 compares our measurement of the Galactic potential to those from other recent works in the enclosed mass space. We show our measurement at median distance, $M(15.4\,\mathrm{kpc}) = 1.99^{+0}_{-0.37} \times 10^{11} M_{\odot}$ with the associated LOOCV uncertainty, with the green cross marker and, with a green line, the best-fit potential over the 5 to 95-percentile distance range of the stars in the sample. The majority of the literature results included in the figure as point estimates are measurements based on different stellar streams

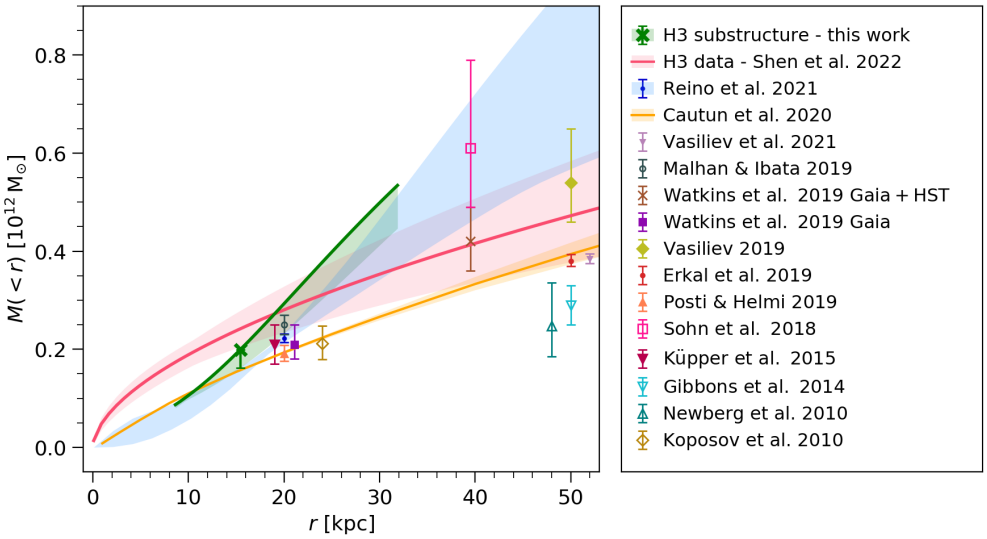


**Figure 4.7.1:** Comparison of our results with previous measurements of the enclosed mass of the Milky Way at different radii. Our results with the H3 accreted substructures sample is shown in green: the measurement at the median distance (with associated LOOCV uncertainty) is illustrated with the cross marker, while the line shows the best-fit measurement across the entire 5 to 95-percentile distance range. The coloured markers show measurements of the enclosed mass by other authors. We have offset the markers showing the results of Koposov et al. (2010) from 20 kpc, and the results of Newberg et al. (2010) and Vasiliev et al. (2021) from 50 kpc, for clarity.

using a variety of methods. With the exception of Reino et al. (2021), they are all acquired from single stream fits: Vasiliev et al. (2021) and Gibbons et al. (2014) analyse the Sagittarius stream, Erkal et al. (2019) and Newberg et al. (2010) analyse the Orphan stream, Malhan & Ibata (2019) and Koposov et al. (2010) the GD-1 stream, and Küpper et al. (2015) the Pal 5 stream.

We also show three measurements that span a larger Galactocentric range which we will now discuss in more detail. In Reino et al. (2021) we applied the action-clustering method on a combination of three coherent streams: GD-1, Pal 5 and Orphan. We obtained $M(20\,\mathrm{kpc}) = 2.22^{+0.10}_{-0.08} \times 10^{11} M_\odot$, shown on Figure 4.7.1 with a blue marker. The KLD-based $1\sigma$ uncertainties over a larger range of radii is shown with the shaded blue regions. In this work we find the enclosed mass at the corresponding distance to be $2.93^{+0}_{-0.60} \times 10^{11} M_\odot$, which is $\sim 30\%$ higher than the results from Reino et al. (2021). However, adjusting for a possible pericentre bias in our H3 accreted structures results

will bring the two measurements more in line. The results of Reino et al. (2021) likely avoid this pitfall, since Pal 5 and Orphan streams are at their apocentre and intermediate phases, respectively.

Shen et al. (2022) infer the Galactic mass profile through the phase-space distribution function of H3 halo stars. Specifically, they analysed the kinematics of halo stars between 50 and 142 kpc, using H3 Survey and *Gaia* EDR3 observations. Their result, extrapolated to lower radii, is shown with the red line, with the 68% confidence intervals shows with the red shaded area. Similarly to our results, they report a mass profile that is generally higher than most other measurements. At radii below $\sim$ 20 kpc, where the majority of our data lies, we measure a lower mass, while at radii larger than $\sim$ 20 kpc, our measurements yield a significantly higher mass. However, fewer than 30% of the stars in our sample are at distances greater than 20 kpc, so we can expect lower accuracy over this distance range compared to the mass profile reported below 20 kpc.

Finally, Cautun et al. (2020) use Milky Way rotation curve data from *Gaia* DR2 to fit the Galactic mass profile with a model that includes six components to describe the baryon distribution and a contracted NFW model for the dark matter halo. Their results, shown with the yellow line including the yellow shaded regions to illustrate the 68% confidence regions, tend to be on the lower side when compared to the rest of the measurements. Our accreted substructure results agree with the Cautun et al. (2020) results only at low radii, $\sim$ 10 kpc, thereafter rising much higher.

Overall our results are higher than most other measurements, possibly owning to the four elements summarized earlier in the section. As the considerable variation between the estimates from different tracers illustrates, unaccounted for systematics dominate as sources of error, emphasizing the necessity of gaining deep insight into the applied methods and the value of considering a variety of methods concurrently.