Monthly Notices of the ROYAL ASTRONOMICAL SOCIETY

MNRAS **509,** 5365–5381 (2022) https://doi.org/10.1093/mnras/stab3176
Advance Access publication 2021 November 3

Orbital phase-driven biases in galactic mass constraints from stellar streams

Stella Reino [1]★ Robyn E. Sanderson [2,3] Nondh Panithanpaisal,[2] Elena M. Rossi[1] and Konrad Kuijken[1]

[1] *Leiden Observatory, Leiden University, Niels Bohrweg 2, NL-2333 CA Leiden, the Netherlands*
[2] *Department of Physics and Astronomy, University of Pennsylvania, 209 S 33rd St, Philadelphia, PA 19104, USA*
[3] *Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave., New York, NY 10010, USA*

Accepted 2021 October 23. Received 2021 October 20; in original form 2021 July 2

ABSTRACT

One of the most promising tracers of the Galactic potential in the halo region is stellar streams. However, individual stream fits can be limited by systematic biases. To study these individual stream systematics, we fit streams in Milky Way-like galaxies from Feedback In Realistic Environments cosmological galaxy formation simulations with an analytical gravitational potential by maximizing the clustering of stream stars in action space. We show that for coherent streams the quality of the constraints depends on the orbital phase of the observed stream stars, despite the fact that the phase information is discarded in action-clustering methods. Streams on intermediate phases give the most accurate results, whereas pericentre streams can be highly biased. This behaviour is tied to the amount of correlation present between positions and momenta in each stream's data: weak correlation in pericentre streams prohibits efficient differentiation between potentials, while strong correlation in intermediate streams promotes it. Although simultaneous fitting of multiple streams is generally prescribed as the remedy to combat individual stream biases, we find that combining multiple pericentric streams is not enough to yield a bias-free result. We finally show that adopting the two-component Stäckel model does not fundamentally induce a biased mass estimate. With our full data set of two multiwrap streams, we recovered the true rotation curve of the simulated galaxy within 12 per cent over the entire range of radii covered by our set of stars (10–176 kpc) and within 6.5 per cent between the 5 and 95 percentile distance range (23–109 kpc).

Key words: methods: numerical – Galaxy: halo – Galaxy: kinematics and dynamics – Galaxy: structure – dark matter.

1 INTRODUCTION

Stellar streams, relics of tidally disrupted globular clusters and dwarf galaxies, are one of the most promising probes of the gravitational field of the Milky Way at large Galactocentric distances. Since the first detections of streams in the Milky Way (Ibata, Gilmore & Irwin 1994; Grillmair et al. 1995; Helmi et al. 1999), many methods have been developed to constrain the Galactic mass profile using streams. Some of these make comparisons between predictions and data directly in position and velocity space such as the orbit-fitting technique (e.g. Koposov, Rix & Hogg 2010), particle ejection methods (e.g. the 'streakline' method of Küpper, Lane & Heggie 2012, the modified Lagrange Cloud Stripping method of Gibbons, Belokurov & Evans 2014, the 'particle-spray' method of Fardal, Huang & Weinberg 2015), or full *N*-body simulations (e.g. Law & Majewski 2010), while others utilize the action-angle coordinates, such as the angle-frequency slope method (Sanders & Binney 2013) and action-clustering method (Sanderson, Helmi & Hogg 2015). These methods have thus far been applied to only a handful of streams, most commonly GD-1 (Koposov et al. 2010; Malhan & Ibata 2019), Pal 5 (Küpper et al. 2015), Sagittarius (Law & Majewski 2010; Dierickx & Loeb 2017; Vasiliev, Belokurov & Erkal 2021), and Orphan (Newberg et al. 2010; Erkal et al. 2019).

To date, most of these studies have focused on measuring the potential with single streams (although see Bovy et al. 2016). However, using a sample of streams evolved in the Via Lactea II simulation Bonaca et al. (2014) showed that constraints from individual streams can be highly biased. Only simultaneous fitting of a collection of streams would result in accurate potential recovery. They estimated that GD-1 and Pal 5-like streams could individually return up to 50 per cent biased mass estimates for the Milky Way. In Reino et al. (2021), we showed that this also holds true when using the action-clustering method. In particular, we saw a significant bias arise from the analysis of GD-1 compared to the constraints derived with a collection of streams. The strategy of simultaneous fitting of multiple streams to avoid the pitfalls of a single stream fit was also advocated by Lux et al. (2013), Sanders & Binney (2013), and Sanderson, Hartke & Helmi (2017).

Today, more than 60 streams have been discovered in the Milky Way (Newberg & Carlin 2016; Mateu, Read & Kawata 2018; Ibata, Malhan & Martin 2019; Koppelman et al. 2019; Myeong et al. 2019; Naidu et al. 2020; Malhan et al. 2021) and ideally we should seek a consensus fit. However, the observational data required for Galactic potential inference is only available for a subset of these streams. Attaining the full 6D phase space information tends to be difficult and necessitates a cross-matching of information from different surveys, e.g. proper motions from *Gaia* (Gaia Collaboration 2016), distances of RR Lyrae from PanSTARRS1 survey (Sesar et al. 2017) or Gaia's Specific Object Study catalogue (Clementini et al. 2019), and radial velocities from RAVE (Kunder et al. 2017), WEAVE (Dalton et al.

★ E-mail: reino@strw.leidenuniv.nl

2012), 4MOST de Jong et al. (2019), or DESI (Levi et al. 2019). Since steams are typically distant and faint, targeted follow-up surveys are often needed, such as the H3 survey (Conroy et al. 2019) targeting the stellar halo and the S5 survey (Li et al. 2019) targeting the stellar streams in the Southern Hemisphere. With this in mind, knowledge of which streams are the most useful for putting accurate constraints on the Galactic potential would be valuable for selecting which streams to focus both our modelling and observing efforts on.

To gain this insight, Bonaca & Hogg (2018) explored the intrinsic information content in the tracks of 11 mock globular cluster stellar streams as a function of their properties using the Fisher-matrix approach. They found that angular length of the stream was the best predictor of the tightness of their parameter constraints and that streams on more eccentric orbits were the most sensitive to the halo shape. However, while the Fisher-matrix approach allowed them to investigate the precision of the constraints the different streams were capable of reaching, they did not address the question of accuracy.

In this paper, we aim instead to gauge the accuracy of the constraints that arise from different streams using the action-clustering method. In particular, we explore the systematics of stream-driven constraints as a function of their orbital phase. To this end, we select two long spatially coherent streams from Feedback In Realistic Environments (FIRE) cosmological-hydrodynamical simulations (Hopkins et al. 2018) for our analysis. First, we set the expectation for the accuracy obtainable with two full-length streams and then divide the streams into smaller sections based on their current orbital phase and analyse these segments independently. This approach is motivated, first, by the fact that in reality it is more likely to detect (or have the full 6D phase space information for) only a small nearby segment of the whole longer stream, or that associations between stream sections observed in different parts of the sky are uncertain. Second, the streams most commonly used for characterizing the potential of the Milky Way are from globular cluster origin, and therefore much shorter than the dwarf galaxy streams found in FIRE simulations. Since globular cluster streams do not intrinsically form in the FIRE simulations, we can instead approximate their length by treating each section as an independent stream. Third, this approach allows us to keep constant some properties of the streams we are working with (for instance eccentricity, inclination, apocentre distance), and to direct our focus on phase differences. Finally, with this framework we can easily apply our method to the full stream data to verify that our potential model does not induce severe systematic biases on these sectional results. As a by-product, this setup will give us an indication of whether, and how much, the constraints vary over the length of the full long stream. Another goal of our study is to investigate how accurately we can recover the true potential of the simulated galaxy when modelling the streams with a Stäckel potential. Although Stäckel potentials are generally considered inadequate for describing realistic galaxies, they have the great benefit of exact actions. This property was our incentive for adopting the Stäckel potential when analysing real Milky Way stellar streams in Reino et al. (2021). In this work, we will test whether this assumption can introduce any significant additional bias into our results.

This paper is organized as follows. We discuss the elements of our method in Section 2, introducing the Stäckel potential (2.1), the action coordinates (2.2), and our action-space clustering measure (2.3). We give an overview of the FIRE suite of simulations and our stream sample in Section 3. In Section 4, we present our results for the full streams, while the results for stream sections are shown in Section 5. Next, in Section 6 we explore the stream section results as a function of orbital phase. In Section 7, we explore the dependence of our result on other stream section properties and provide a reasoning for the orbital-phase effects. Finally, we discuss our results and make conclusions in Section 8.

2 METHOD

We follow the action-clustering method outlined in Sanderson et al. (2015) and Reino et al. (2021) which aims to constrain the galactic potential by maximizing the clustering of stream stars in action space. We vary the potential used for converting the stars' positions and velocities into action coordinates and adopt as the best-fitting potential the one that gives rise to the most clustered distribution of actions.

2.1 Stäckel potential

Analytical transformation of phase space coordinates $(\boldsymbol{x}, \boldsymbol{v})$ to action coordinates $\boldsymbol{J}$ is possible only for a small set of potentials for which Hamilton–Jacobi equations can be solved by separation of variables. Action estimation for general potentials requires the use of numerical approximation algorithms (see Sanders & Binney 2016), although repeated calculation of actions can be computationally costly in these cases. Analytical calculation, in contrast, is less expensive and as such allows both for a larger number of stream stars to be included in the fit and more potentials to be considered.

The Stäckel potential, which separates in ellipsoidal coordinates, is the most general of the small group of separable potentials to describe a real galaxy, as it allows for arbitrarily flattened density profiles and construction of flat rotation curves with the two-component Stäckel model (Batsleer & Dejonghe 1994).

However, the applicability of Stäckel potentials is limited by the fact that all orbits are defined by the same foci. This restriction has been shown to be incorrect for real galaxies (Kuijken & Gilmore 1989; Binney 2012) and therefore a perfect global fit is not possible. Many action-calculation methods still utilize its advantages and, for example, fit each orbit with a local Stäckel potential instead (Sanders 2012) or apply the Stäckel formulae to another more complex potential of interest (Binney 2012).

One of the aims of this work is to test the practicality of adopting a Stäckel potential to describe a realistic galaxy and to see whether, despite its known limits, it could still be utilized as a valuable tool. Discussion on the expected size of errors originating from our choice of the Stäckel potential model is included in Section 4.

In this work, we consider oblate axisymmetric two-component Stäckel potentials described by spheroidal coordinates, a limiting case of ellipsoidal coordinates. Spheroidal coordinates (λ, ν, ϕ) are related to cylindrical coordinates (R, z, ϕ) by the following quadratic equation:

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

where $\tau = \lambda, \nu$ are the roots.

Parameters a and c are constants that determine the location of the foci $\Delta = \sqrt{a^2 - c^2}$ and define the axial ratio of the coordinate surfaces, $e \equiv \frac{a}{c}$ and therefore establish the shape of the coordinate system. For an oblate density distribution, we have $a > c$. Further details about this coordinate system can be found in de Zeeuw (1985) and Dejonghe & de Zeeuw (1988).

A Stäckel potential, Φ, in spheroidal coordinates can be written as

$$\Phi(\lambda, \nu) = -\frac{f(\lambda) - f(\nu)}{\lambda - \nu},$$
$$f(\tau) = (\tau - c^2)\mathcal{G}(\tau), \tag{2}$$

where we set $\mathcal{G}(\tau)$ to be the Kuzmin–Kutuzov potential

$$\mathcal{G}(\tau) = \frac{GM_{\rm tot}}{\sqrt{\tau} + c}, \tag{3}$$

where $M_{\rm tot}$ is the total mass and G is the gravitational constant.

To construct a two-component Stäckel model, we combine two individual Stäckel potentials, $\Phi_{\rm outer}$ and $\Phi_{\rm inner}$. The motivation for this is to add more flexibility to our potential model thereby allowing for a more realistic model of the galaxy (Batsleer & Dejonghe 1994; Reino et al. 2021). The two components are defined by different parameters $a_{\rm outer}$, $c_{\rm outer}$ and $a_{\rm inner}$, $c_{\rm inner}$, and therefore each has a different scale and axial ratio but, crucially, they must have the same foci for the total potential to retain the Stäckel form (as defined by equation 2). It then follows that

$$\begin{aligned} a^2_{\rm outer} - a^2_{\rm inner} &= c^2_{\rm outer} - c^2_{\rm inner} = q, \\ \lambda_{\rm outer} - \lambda_{\rm inner} &= \nu_{\rm outer} - \nu_{\rm inner} = q, \end{aligned} \tag{4}$$

where $q > 0$ is a constant. The total $\mathcal{G}(\tau)$ is now a combination of two parts, $\mathcal{G}_{\rm outer}(\tau)$ and $\mathcal{G}_{\rm inner}(\tau)$

$$\mathcal{G}(\tau) = \frac{GM_{\rm tot}(1-k)}{\sqrt{\tau_{\rm outer}} + c_{\rm outer}} + \frac{GM_{\rm tot}k}{\sqrt{\tau_{\rm outer} - q} + c_{\rm inner}}, \tag{5}$$

and the total two-component potential is

$$\begin{aligned} &\Phi(\lambda_{\rm outer}, \nu_{\rm outer}, q) \\ &\quad = -GM_{\rm tot}\left[\frac{1-k}{\sqrt{\lambda_{\rm outer}} + \sqrt{\nu_{\rm outer}}} + \frac{k}{\sqrt{\lambda_{\rm outer} - q} + \sqrt{\nu_{\rm outer}}}\right], \end{aligned} \tag{6}$$

where k is the ratio between the inner component and the outer component masses and $M_{\rm tot}$ is the sum of the two-component masses.

We define our two-component Stäckel potentials on a grid of five parameters $\boldsymbol{\zeta} = (M_{\rm tot}, a_{\rm outer}, e_{\rm outer}, a_{\rm inner}, k)$. We select the trial potentials by drawing 50 points for each of the shape parameters, from uniform distributions in log space, over the ranges: [0.7, 1.8] in $\log_{10}(a_{\rm outer}/{\rm kpc})$, $[\log_{10}(1.0), \log_{10}(2.0)]$ in $\log_{10}(e_{\rm outer})$, and $[\log_{10}(0.), \log_{10}(0.7)]$ in $\log_{10}(a_{\rm inner}/{\rm kpc})$. However, we only use a subset of these parameter combinations (∼8000) that constructs a mathematically valid potential according to equation (4); i.e. the parameter combinations which adhere to $c^2_{\rm outer} = \frac{e^2_{\rm outer}}{a^2_{\rm outer}} > c^2_{\rm inner}$. We also draw 20 points for each mass parameter over the ranges: [11.5, 12.5] in $\log_{10}(M/M_\odot)$ and $[\log_{10}(0.01), \log_{10}(0.3)]$ in $\log_{10}(k)$. In total, our grid contains 3253 600 trial potentials.

We further discard the potentials that cause any of the star particles in our sample to be unbound from the host galaxy. In Reino et al. (2021), we showed that our results did not change appreciably if, instead of this strict condition, we allowed a small percentage of the stars to become unbound. Furthermore, in that work we were dealing with real stream data with measurement errors while here we know the true position and velocity of all our particles and do not need to worry about measurement errors causing unbound stars.

2.2 Actions

The action-angle coordinates are a set of canonical coordinates which considerably simplify the equations of motion of a bound star in a static or adiabatically time-evolving potential: the actions, J_i, are integrals of motion that uniquely define the stellar orbit and the angles, θ_i, are periodic coordinates that express the phase of the orbit. For the Stäckel potential, we define the actions J_λ and J_ν as

$$J_\tau = \frac{1}{2\pi}\oint p_\tau d\tau\ , \tag{7}$$

where p_τ is the conjugate momentum to the coordinate τ and the integral is over the full oscillation in τ. The third action J_ϕ is equal to the z-component of the angular momentum, L_z, and hence is constant in our axisymmetric potentials.

The conjugate momenta, p_τ, can be found by solving the Hamilton–Jacobi equation by separation of variables. In addition to the momenta, the separation of variables introduces three isolating integrals: I_2, I_3, and the total energy E. The integrals I_2 and I_3 are defined as (Dejonghe & de Zeeuw 1988)

$$\begin{aligned} I_2 &= \frac{L_z^2}{2}, \\ I_3 &= \frac{1}{2}\left(L_x^2 + L_y^2\right) + (a^2 - c^2)\left[\frac{1}{2}v_z^2 - z^2\frac{\mathcal{G}(\lambda) - \mathcal{G}(\nu)}{\lambda - \nu}\right]. \end{aligned} \tag{8}$$

The solution to the Hamilton–Jacobi equation then allows the momenta, p_τ, to be expressed as a function of the τ coordinate and the three isolating integrals:

$$p_\tau^2 = \frac{1}{2(\tau - a^2)}\left[\mathcal{G}(\tau) - \frac{I_2}{\tau - a^2} - \frac{I_3}{\tau - c^2} + E\right], \tag{9}$$

which can then be used in equation (7) and integrated numerically.

2.3 Clustering measurement

An intrinsic element of the action-clustering method is the procedure of quantifying and comparing the degree of clustering present in the action space of different potentials. Following Sanderson et al. (2015) and Reino et al. (2021), we measure this degree of clustering with the Kullback–Leibler divergence (KLD). The KLD is a measure of the divergence between two probability distributions $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$. For a discrete sample $[\boldsymbol{x}_i]$ with $i = 1, \ldots, N$ drawn from a distribution $p(\boldsymbol{x})$, the KLD can be calculated as

$$\mathrm{KLD}(p \,||\, q) \approx \frac{1}{N}\sum_i^N \log\frac{p(\boldsymbol{x}_i)}{q(\boldsymbol{x}_i)}, \quad \text{if } q(\boldsymbol{x}_i) \neq 0\ \forall i. \tag{10}$$

The value of KLD increases with increasing difference between $p(\boldsymbol{x})$ and $q(\boldsymbol{x})$ and is equal to 0 when $p(\boldsymbol{x}) = q(\boldsymbol{x})$.

Since we are looking to measure the amount of clustering in action space, irrespective of the cluster locations, we proceed with the idea of comparing the distribution of actions in a particular trial potential to a completely unclustered, featureless distribution. In other words, we set $q(\boldsymbol{x})$ to a uniform distribution in the actions, $u(\boldsymbol{J})$, and $p(\boldsymbol{x})$ to a probability distribution of actions, $p(\boldsymbol{J} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})$, found by mapping the phase space coordinates $\boldsymbol{\omega}$ to action space $\boldsymbol{J}$ with a trial potential parametrized by $\boldsymbol{\zeta}$. The difference between the two distributions is greater the more clustered the action distribution.

Our goal is to maximize the difference, and therefore the KLD value, between these probability distributions as we vary the trial potential and explore the parameter space. The trial potential with the highest KLD value is adopted as the best-fitting potential, with parameters $\boldsymbol{\zeta}_0$, for that particular data set, $\boldsymbol{\omega}$.

The standard KLD gives equal weight to each star particle in the sample and as a consequence, when multiple streams (or stream sections) are analysed simultaneously, streams (or stream sections) containing more star particles have stronger influence over the results. Since stream membership in this case is known, we can make the best use of the data by giving equal weight to all *streams* instead, by weighting the contribution of each star particle with

$$w_j = \frac{1}{N_{\rm s}} \times \frac{1}{N_j}, \tag{11}$$

where N_s is the number of streams and N_j is the number of star particles in stream j.

This *weighted* KLD is thus calculated as follows:

$$\mathrm{KLD1}(\boldsymbol{\zeta}) = \sum_{j}^{N_s} \sum_{i}^{N_j} w_j \log \frac{p(\boldsymbol{J} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})}{u(\boldsymbol{J})}\bigg|_{\boldsymbol{J}=\boldsymbol{J}_{\zeta}^{ij}}, \tag{12}$$

where $\boldsymbol{J}_{\zeta}^{ij} = \boldsymbol{J}(\boldsymbol{\zeta}, \boldsymbol{\omega}_{ij})$ and $\boldsymbol{\omega}_{ij}$ are the phase space coordinates for star i in stream j.

Although the function $p(\boldsymbol{J} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})$ is not known a priori, it can be constructed using the observed points $\boldsymbol{J}$ via a density estimator. Here, we obtain $p(\boldsymbol{J} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})$ using the Enlink algorithm developed by Sharma & Johnston (2009). $u(\boldsymbol{J})$, on the other hand, is constant across all trial potentials and does not have an impact on the results. It can therefore be set to any preferred value.

Finally, the method will work best if the different streams do not overlap with each other in action space. Since in this case, as in Reino et al. (2021), we know which stars belong to which stream, we can ensure this absence of overlap by calculating the probability distributions for each stream independently. So, instead of estimating $p(\boldsymbol{J} \mid \boldsymbol{\zeta}, \boldsymbol{\omega})$ with the full set of sample points $\boldsymbol{J}$, we construct a probability density function $p_j(\boldsymbol{J}_j \mid \boldsymbol{\zeta}, \boldsymbol{\omega}_j)$ for each stream j individually from points $\boldsymbol{J}_j$. To keep the p_j at the correct relative size between the different streams, we normalize each p_j with $\frac{N_j}{N}$. In practice, the KLD equation that we use is therefore

$$\mathrm{KLD1}(\boldsymbol{\zeta}) = \sum_{j}^{N_s} \sum_{i}^{N_j} w_j \log \frac{N_j}{N} \frac{p_j(\boldsymbol{J}_j \mid \boldsymbol{\zeta}, \boldsymbol{\omega}_j)}{u(\boldsymbol{J})}\bigg|_{\boldsymbol{J}=\boldsymbol{J}_{\zeta}^{ij}}. \tag{13}$$

It is important to note that neither the weighing nor the separation of streams in action space is critical for the action-clustering method to work, as we have shown in previous works (Sanderson et al. 2015; Reino et al. 2021).

The goal of this procedure, as already stated above, is to determine the set of Stäckel potential parameters that maximize the KLD1 value. As a second step in our procedure, we calculate the confidence intervals on these best-fitting parameters, by comparing the action distribution of the best-fitting potential, $p(\boldsymbol{J} \mid \boldsymbol{\zeta}_0, \boldsymbol{\omega})$, to the action distributions of the other trial potentials, $p(\boldsymbol{J} \mid \boldsymbol{\zeta}_{\mathrm{trial}}, \boldsymbol{\omega})$ using again the Kullback–Leibler divergence.

The KLD defined in such a way can be interpreted as the relative probability of the potential parameters $\boldsymbol{\zeta}_{\mathrm{trial}}$ to the best-fitting potential parameters $\boldsymbol{\zeta}_0$ (Kullback & Leibler 1951; Kullback 1959). As we move further from the best-fitting parameter values, the difference between the two action distributions grows and so does the KLD value. In other words, we can measure how far from the best-fitting parameters we can move before the action distribution starts to significantly differ from that of the best-fitting action distribution. The confidence intervals can then be drawn based on the value of KLD that we deem to correspond to significant difference. The interpretation of KLD as the expectation value of the difference in the log of two posterior probabilities allows us to determine the value of KLD that corresponds to any preferred level of significance. For example, the significance of 1σ corresponds to KLD $= 0.5$. An in-depth derivation of the relation between the KLD values and the confidence levels can be found in Reino et al. (2021) and a full discussion of this interpretation of KLD in Sanderson et al. (2015).

As with KLD1($\boldsymbol{\zeta}$) in equation (13), this version of KLD will incorporate weights and a separate density estimation for different streams. It is defined as follows:

$$\mathrm{KLD2}(\boldsymbol{\zeta}) = \sum_{j}^{N_s} \sum_{i}^{N_j} w_j \log \frac{N_j}{N} \frac{p_i(\boldsymbol{J}_i \mid \boldsymbol{\zeta}_0, \boldsymbol{\omega}_i)}{p_i(\boldsymbol{J}_i \mid \boldsymbol{\zeta}_{\mathrm{trial}}, \boldsymbol{\omega}_i)}\bigg|_{\boldsymbol{J}_i=\boldsymbol{J}_0^{ij}}. \tag{14}$$

Once again, we use Enlink to estimate the probability density functions $p_i(\boldsymbol{J}_i \mid \boldsymbol{\zeta}_0, \boldsymbol{\omega}_i)$ and $p_i(\boldsymbol{J}_i \mid \boldsymbol{\zeta}_{trial}, \boldsymbol{\omega}_i)$ using the two sets of actions. Both functions are then evaluated at $\boldsymbol{J}_0 = \boldsymbol{J}(\boldsymbol{\zeta}_0, \boldsymbol{\omega})$, the actions computed with the best-fitting potential parameters $\boldsymbol{\zeta_0}$.

Finally, throughout the rest of the paper we discuss the uncertainties of our measurements as 1σ confidence intervals. This corresponds to the subset of trial potentials with KLD2($\boldsymbol{\zeta}$) ≤ 0.5. The individual parameter confidence intervals are determined as the full range of parameter values in this subset of potentials.

3 SIMULATION DATA

In this work, we make use of streams 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) and ELVIS on FIRE suite (Garrison-Kimmel et al. 2019) of the FIRE project (Hopkins et al. 2018). All haloes were simulated in ΛCDM cosmology at particle mass resolution of 3500–7100 $M_\odot$ and spatial resolution of 1–4 pc for star/gas particles; 18 000–35 000 $M_\odot$ and 40 pc for DM particles. The resolution of this suite of simulations allows both luminous and dark subhaloes to be resolved well even near each Milky Way-like galaxy, and follows the formation of tidal streams from dwarf galaxies down to slightly below the mass of the Milky Way's 'classical' dSphs: around 10^8 $M_\odot$ in total mass or 10^6 $M_\odot$ in stellar mass (at $z = 0$).

Panithanpaisal et al. (2021) search these simulated galaxies for accreted structures that are spatially coherent and stream-like at present day. They identify 100 such streams across 13 simulations (see their table 1) and confirm that the progenitor galaxies of these coherent streams are consistent with the mass–size–velocity dispersion relationship of observed present-day Milky Way satellites. This implies that the streams' phase-space volumes, and therefore their sizes and densities in action space, are representative of real streams from satellite galaxies.

In this work, we focus on two of the nine coherent tidal streams found in the halo of the isolated galaxy simulation `m12i`. This simulated galaxy has had a quiet recent accretion history involving mostly quite low-mass galaxies, as the Milky Way's is expected to have been since the Gaia–Enceladus merger (e.g. Bonaca et al. 2017, 2020; Belokurov et al. 2018, 2020; Haywood et al. 2018, Di Matteo et al. 2019, Naidu et al. 2020). Specifically, `m12i` experiences no mergers with mass ratios more similar than 1:3 after $z = 1.7$ (about 9.5 Gyr ago; Santistevan et al. 2020). Its thin disc stabilizes in its current configuration more than 5 Gyr before present day (Garrison-Kimmel et al. 2018), and has a stellar mass and surface density at the Solar circle comparable to the Milky Way (Sanderson et al. 2020).

We select two long streams identified in `m12i`, which each have multiple wraps around their host galaxy and contain between 2000 and 4000 star particles. Although they are far longer than nearly every known Milky Way stream, this length is ideal for our purposes as it allows us to divide each stream into many sections that are each comparable to most known stream lengths, and to select several instances of the same orbital phase from each stream. Throughout the paper, we use the error-free present-day positions and velocities of the star particles that belong to these streams and assume complete and contamination-free knowledge of stream membership for each star particle.

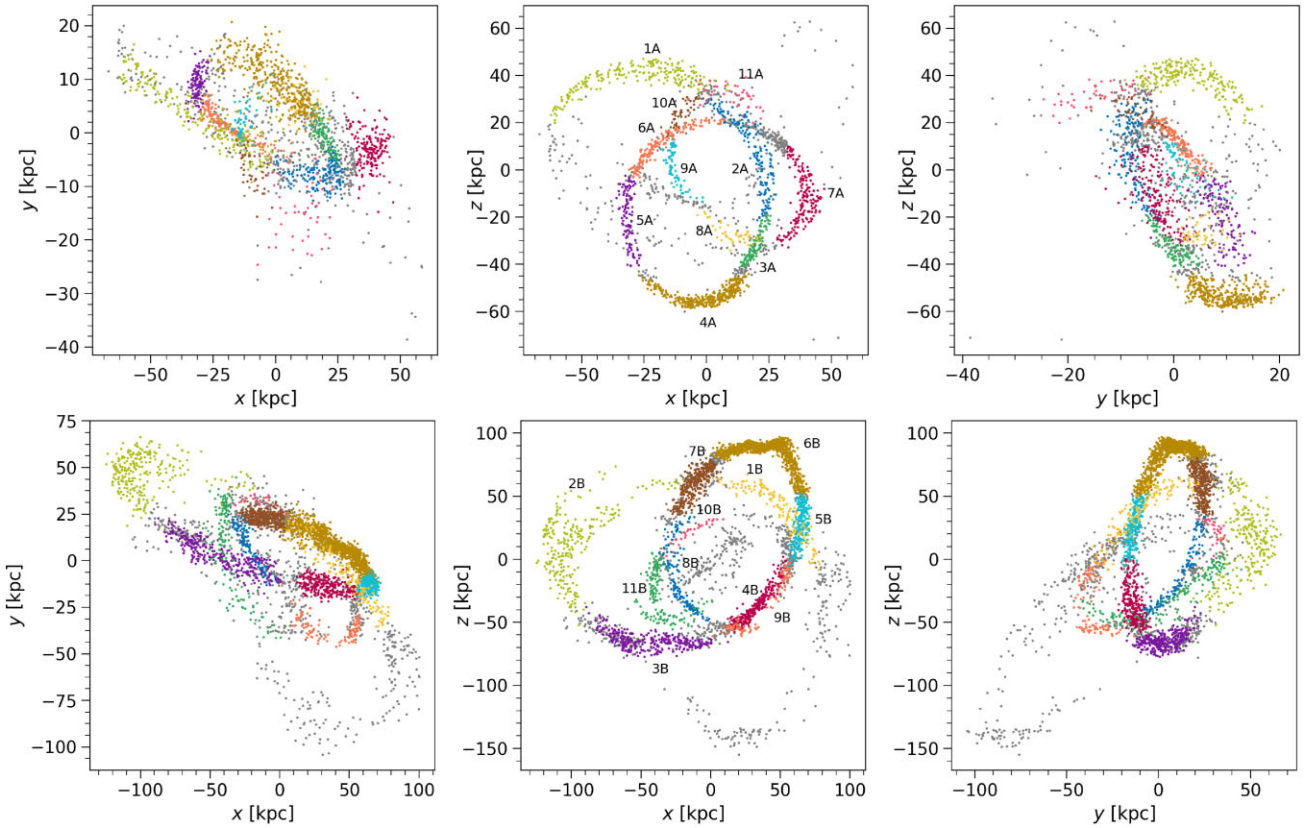

Figure 1. Streams A (top) and B (bottom) in galactocentric coordinates. Each stream section is identified by a given colour throughout the whole paper (see Table 1 for a summary). Star particles not belonging to any section are shown as grey.

We define 11 sections in each stream by eye using a combination of the position, velocity, and orbital phase information. Fig. 1 shows the two streams in the galactocentric reference frame with each section highlighted using a different colour (the velocities of the streams are shown in Fig. A1). Throughout this paper, we consistently use the same colour to represent a particular section of each stream, the legend for this section-specific colour scheme is given in Table 1. Gold-coloured points represent the progenitor-containing section in both streams. The star particles that do not belong to any section are shown in light grey. In Fig. 2, we show the streams after we have unwound them using angle coordinates. The angle coordinates were computed using the AGAMA library (Vasiliev 2019) in a low-order multipole (dark matter and hot gas distribution) and cylindrical spline (stellar and cold gas distribution) model fit to the potential of `m12i` (Arora et al., in preparation). To unwind each stream, we utilized the Hough transform (Duda & Hart 1972), a line detection algorithm, to identify line overdensities in the θ_r versus θ_ϕ projection (Pearson et al. 2021; Shih et al. 2021). All the lines identified were then connected by exploiting the periodic boundary condition in the angle projection (i.e. if a line terminates at $(\theta_r, \theta_\phi) = (2\pi, \tilde{\theta}_\phi)$, it will reappear at $(\theta_r, \theta_\phi) = (0, \tilde{\theta}_\phi)$). Starting from the line with the most members as the stem, we progressively connected more lines to both sides, shifting (θ_r, θ_ϕ) of the members of the newly attached lines with suitable offsets, until all the lines are used. A more detailed explanation of the unwinding process will be presented in Panithanpaisal et al. (in preparation).

Fig. 2 thus allows for clear identification of the phase that each section is on, which we have also summarized in Table 1. Although we define for both streams the section which contains the remaining progenitor (4A and 6B), we do this purely for visualization purposes and do not make use of these sections individually. In total, discounting the progenitor-containing sections, we identify six apocentre sections, seven pericentre sections, and seven intermediate sections between the two streams. The sizes of the sections vary from 30 to 318 star particles, with a total of 973 star particles within apocentre sections, 984 within pericentre sections, and 1176 within intermediate sections.

4 FULL STREAM RESULTS

In this section, we discuss the results of our action-clustering method when applied to the two streams as a whole, both independently and in combination. Fig. 3 compares our recovered Stäckel potential with the true potential in two projections: the circular velocity curve (left) and enclosed mass profile (right). The true potential which is derived from the spherically binned total mass profile of the simulation snapshot is shown with the black dash–dotted line. The solid lines show the best-fitting Stäckel potential of a particular data set. Both stream A (solid green line) and stream B (solid blue line) recover the true potential relatively accurately, especially at the galactocentric distances where most of their star particles are. Stream A, whose star particles lie between ∼10 and ∼96 kpc, recovers the true potential more accurately at lower radii than Stream B, whose median distance is twice as large. The best-fitting velocity curve lies within 32 km s^{-1} of the true velocity curve across all the distances of its star particles, with ∼10 km s^{-1} at the median distance (∼38 kpc). Stream B, covering the range of distances from ∼20 to ∼176 kpc, on the other hand recovers the true potential better at

Table 1. Sections defined in our streams. The upper portion lists the sections from stream A and the lower portion contains those of stream B, each with their respective full stream details on top. The columns give for each stream their signifier, number of stars N_*, median galactocentric distance r, approximate orbital phase, the colour scheme (which is used to mark the corresponding sections throughout the paper), length, width, and velocity dispersion. Discussion on how these properties were computed is included in Appendix B.

Stream	N_*	r [kpc]	Phase	Colour	Length [kpc]	Width [kpc]	σ_v [km s^{-1}]
A	2063	37.7	-	-	-	-	-
1A	215	50.8	apocentre		74.7	3.63	31.9
2A	165	26.4	pericentre		61.6	2.60	20.8
3A	109	38.6	intermediate		38.6	1.84	18.6
4A	362	55.6	apocentre (progenitor)		50.8	3.12	12.1
5A	117	38.4	intermediate		33.5	2.69	14.5
6A	167	23.3	pericentre		47.0	1.67	23.2
7A	183	41.5	apocentre		45.7	3.49	36.8
8A	43	31.1	intermediate		25.8	3.77	25.1
9A	70	15.0	pericentre		32.1	2.37	61.3
10A	41	27.8	intermediate		23.3	2.92	34.6
11A	45	37.2	apocentre		34.1	6.28	40.6
B	4038	78.8	-	-	-	-	-
1B	149	68.2	pericentre		115.5	5.46	20.6
2B	263	115.2	apocentre		179.7	13.0	21.7
3B	305	80.8	intermediate		110.9	6.48	16.4
4B	259	53.5	pericentre		69.0	4.39	16.1
5B	243	72.4	intermediate		58.4	3.66	12.0
6B	1152	96.8	apocentre (progenitor)		93.8	4.75	11.1
7B	318	67.9	intermediate		47.3	5.86	19.0
8B	144	38.8	pericentre		103.1	3.45	22.6
9B	109	66.9	apocentre		108.2	4.71	20.3
10B	30	42.9	pericentre		55.8	2.42	19.0
11B	159	53.3	apocentre		118.1	6.44	24.7

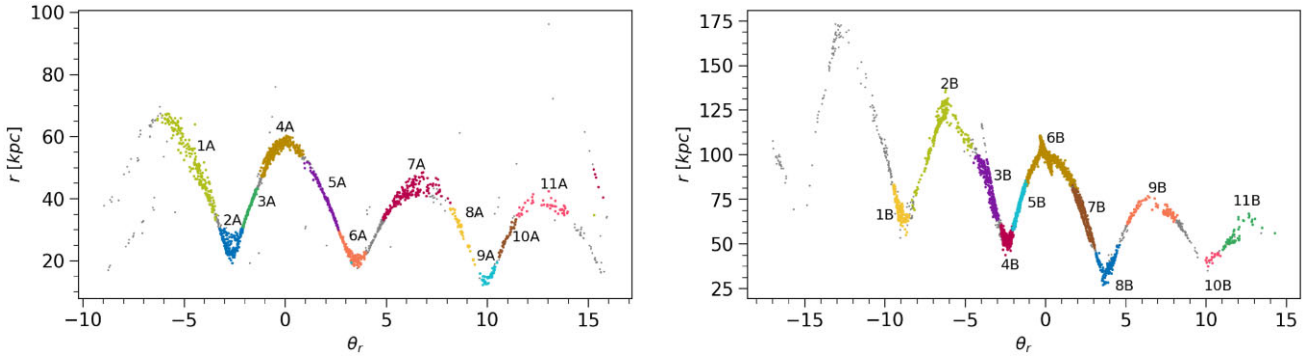

Figure 2. Phase information of streams A (left) and B (right). The progenitor-containing section (in gold) is at $\theta_r = 0$. Colour scheme as in Fig. 1 (see also Table 1).

larger radii. Its best-fitting velocity curve is within 20 km s^{-1} of the true velocity curve across its radial range, with 2.6 km s^{-1} at the median distance (∼79 kpc). The largest differences between the predicted and true velocity curves occur at the shortest and the largest radii for both streams. The true potential at galactocentric distances less than ∼20 kpc cannot be well reproduced by either stream likely due to the small number of star particles within that range (only ∼2 per cent of the stars in our full sample). A further discrepancy might arise also due to the lack of flexibility in the Stäckel potential but this possibility cannot be explored with the current data.

Unsurprisingly, the uncertainties of these best-fitting measurements (shows as the shaded regions) are also the tightest where each stream contains the most data: the median distance of star particles in each stream (marked with a cross) correlates with the distance at which the uncertainty region has reduced to its minimum extent over all distances.

The results from combined data (solid green line) show improved accuracy and precision along the whole range of galactocentric distances probed by our streams. The uncertainty region no longer has a specific galactocentric distance where it is the tightest, instead we are able to recover the true velocity curve within 12 per cent over the range of radii covered by all the star particles (10–176 kpc). As is the case with the individual stream results, the best-fitting rotation curve most deviates from the true rotation curve at very short and very large distances: between 23 and 109 kpc, the 5–95 percentile galactocentric distance range of the star particles, we recover the true rotation curve within 6.5 per cent.

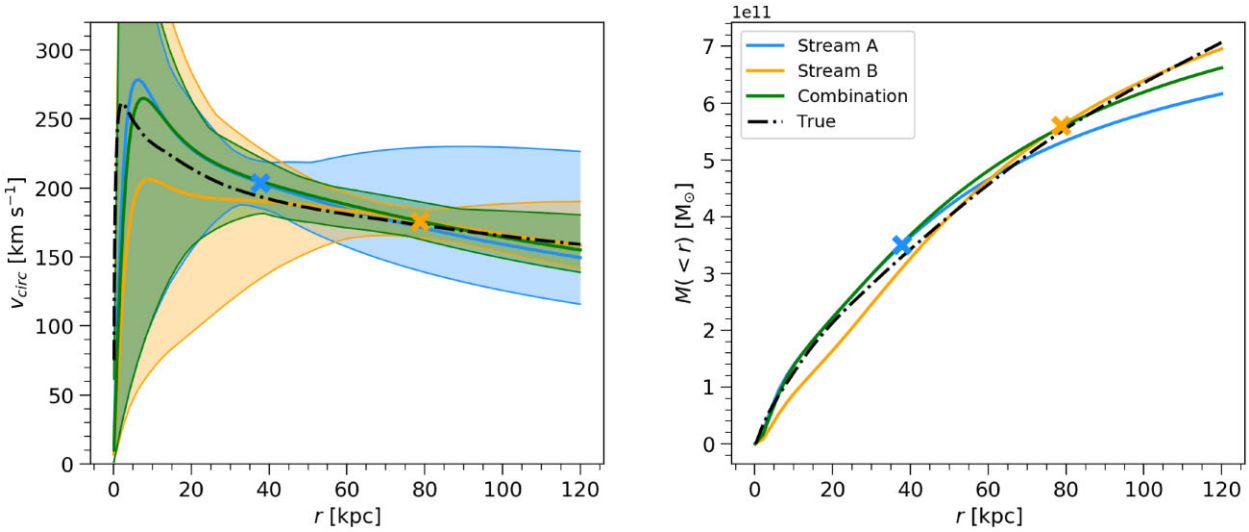

Figure 3. Results for stream A, stream B, and their combined data. We compare the results from these data sets to the true potential (black dash–dotted line) in circular velocity and enclosed mass. The solid coloured lines show the best-fitting Stäckel potential of each data set and the shaded regions show the corresponding 1σ uncertainty regions. The median distance of the star particles in the stream A and stream B data sets is marked with a cross (see Table 1).

To confirm that this accuracy is typical of all streams in the simulation, we randomly selected three of the other nine coherent streams present in the simulation and applied the action-clustering method to each whole stream separately. We find that all three streams recover the true potential very well – within 13 per cent, 3 per cent, and 10 per cent between the 5 and 95 percentile galactocentric distance range covered by their respective star particles.

These full stream results, for which we used all the star particles of the two streams, are virtually unaffected if we remove the sections that contains the progenitors (4A and 6B). Stream B identifies the exact same potential as best fit with or without the progenitor, while stream A finds a best-fitting potential with a somewhat lower mass across all radii in the without-progenitor-section case compared to the full stream case. However, this lower mass profile is still well within the uncertainty region of both stream A and the combined streams (shown in Fig. 3). The *extent* of the uncertainty regions between the with-progenitor and no-progenitor results only show very minor differences for both streams.

To address the question whether the difference between the best-fitting Stäckel potentials and the true potential might largely be attributed to our adoption of the Stäckel model, we performed a least-squares fit to the true velocity curve over the range of galactocentric distances where we have stellar data (10–176 kpc) using only our set of trial potentials. The potential that minimizes the least squares can be considered the best possible Stäckel potential approximation to the true potential in terms of circular velocity. We find that this Stäckel potential approximates the true potential extremely well: it recovers the true velocity curve within 2.8 km s^{-1} or 1.2 per cent over this distance range. Therefore, we expect the bias due to adopting the Stäckel model to be negligible even for the innermost stream sections in our sample, and well within our uncertainties.

Fig. 4 shows the comparison of the vertical acceleration (i.e. acceleration in z-direction) field of the simulated galaxy (left-hand panel) and that of our best-fitting potential model of the combined stream data set (middle panel). The relative residuals are shown in the right-hand panel. In general, the two acceleration fields are consistent with each other. It is not surprising that the greatest difference between the true and model z-accelerations occur near the plane of the galaxy. Part of the reason lies in the fact that the simulated galaxy is dynamic and the accelerations are not perfectly symmetric around the mid-plane (see e.g. Beane et al. 2019). In addition to this m12i has a warp in its outer disc, that our model, or any other axisymmetric model, is not designed to reproduce. This warping is likely the cause for such pronounced differences just above the z-axis on the positive x-direction and just below the z-axis on the negative x-direction. Another reason for the mismatch lies in the fact that, like most parametrized, axisymmetric potentials, the Stäckel model has limited flexibility when used as a global model of the potential of a realistic galaxy. This is likely because the disc is flatter than our Stäckel model can account for, causing it to underestimate the maximum density in the plane. In particular, as is clear from Fig. 3, it does not deliver a good match to the galaxy at small radii. However, this is also the region where the uncertainties in our best-fitting model are the largest, and where we have no data. The green circle in the right-hand panel of Fig. 4 shows the minimum radius of the star particles in our sample.

We compare the alignment between the streams and the predicted orbits in Fig. 5. The figure shows the orbits of a single representative star particle per each stream section, selected to lie near the centre of its respective section. These stars have then been integrated backwards and forwards in time in the best-fitting potential of their respective full stream. We do not expect to see perfect alignment between the star particles and the integrated orbits for several reasons: there is a natural small misalignment between streams and orbits due to the range of energies in the stream stars, but more importantly in this situation, these streams have evolved is lumpy, time-evolving potentials in stark contrast to our smooth and static Stäckel model. The disruption of a dwarf galaxy in a realistic halo is a complex process often taking several pericentre passages, which each produce their own trailing and leading arms based on the evolving energy and angular momentum distribution of the stars being stripped. This naturally leads to different segments of the stream being on slightly different orbits.

Despite this caveat, the orbits of stream A (top panels) align with its star particles markedly well. The orbits of stream B (bottom panels),

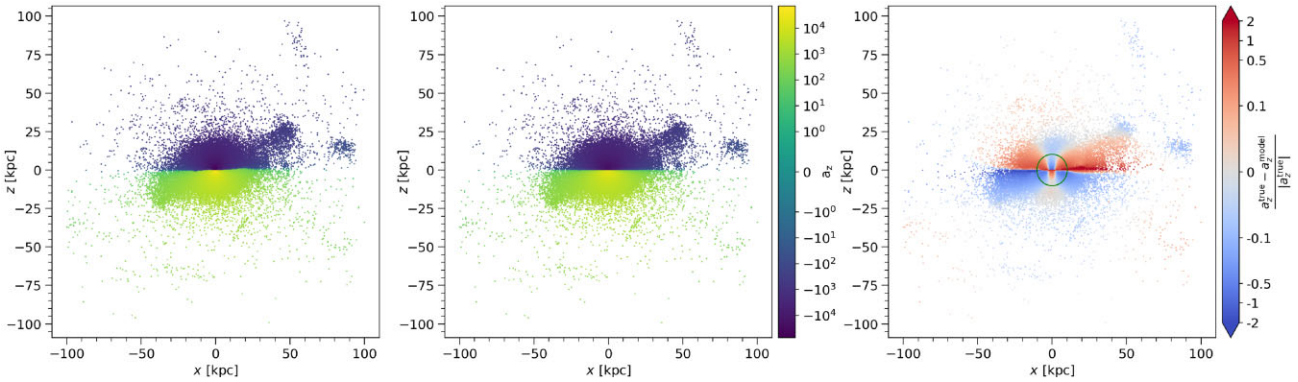

Figure 4. The binned z-accelerations of a slice of the simulated galaxy. The left-hand panel shows the z-accelerations of the real star particles, the middle panel shows the predicted z-accelerations for the star particles based on our best-fitting Stäckel model with the combined stream data set, and the right-hand panel shows their relative residuals. The slice has a thickness of 2 kpc and is centred on galactocentric y-axis. The green circle in the right-hand panel marks the minimum distance of the star particles in our two streams.

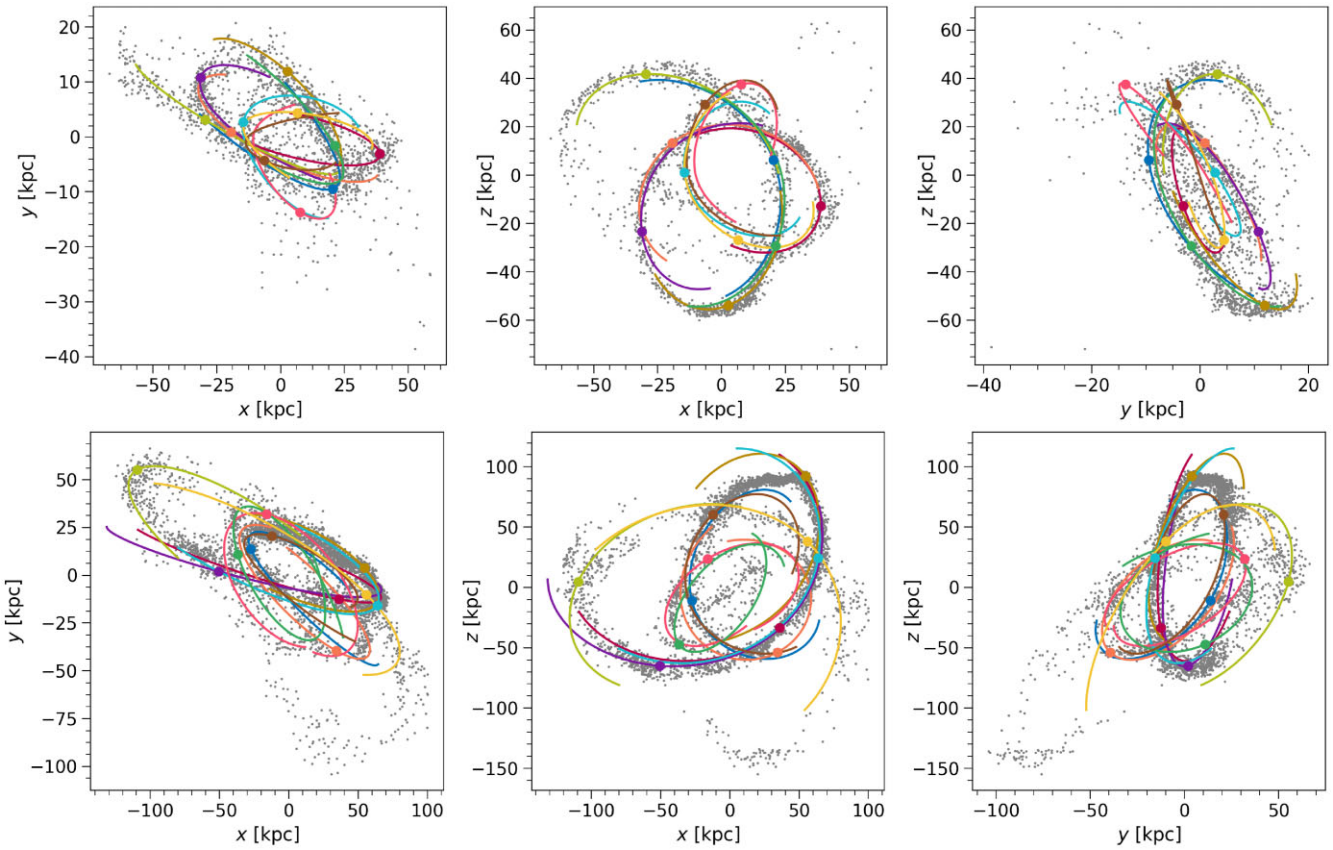

Figure 5. The orbits of streams A (top) and B (bottom) in their respective best-fitting potentials. The star particles whose orbits are shown (marked with a coloured dot) were picked by eye to be near the centre of their section. The orbits are coloured based on the colour scheme of the section they belong to (see Table 1).

however, show some inconsistencies with the data. One of the main mismatches is near the location of the progenitor, which is to be expected, as leading and trailing tails by definition have a mismatch in their orbits. The second inconsistency is between the orbits of 2B and 3B. The rest of the orbits, especially those in the inner parts of the stream, align well with the data.

The action space of each stream in their respective best-fitting potentials is shown in Fig. 6. The colours of the star particles once again correspond to the sections they belong to and reveal the structure of these streams in action space. As the stars get stripped from the progenitor they settle either into leading or trailing tails, which should form two slightly separated clusters in action space. Here, we see this behaviour clearly: the section containing the progenitor (gold-coloured dots) is near the centre of the action space and the sections comprising the trailing tail are above it, while the sections in the leading trail are below it. It is also noticeable that the sections that have evolved further from the progenitor along the tails, tend to be further from the progenitor also in action space. This

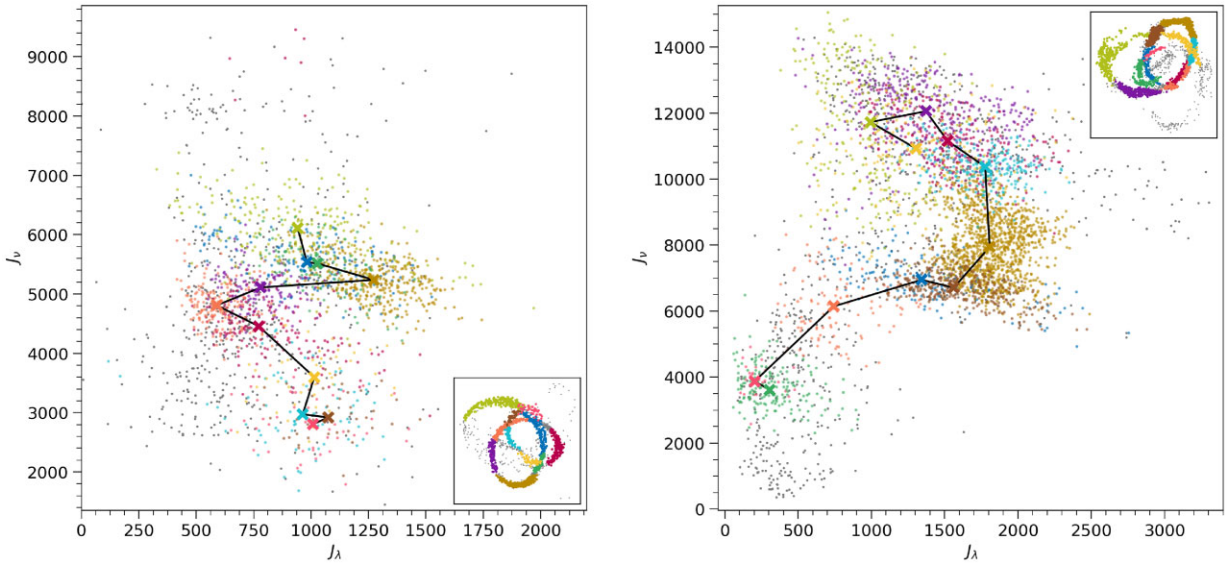

Figure 6. The action space of stream A (left) and stream B (right) in their respective best-fitting potentials. The insets show the stream in x–z frame for reference. The star particles are coloured based on their sections (see Table 1) with star particles not belonging to any section shown as grey. The centre of each section is marked with an 'X' in its respective colour, and the black line connects them in the order that they appear in phase space.

is because the least-bound stars, with the largest energy and action difference from the progenitor, escape the progenitor first.

On the whole, we conclude that the action-clustering method and the two-component Stäckel potential can reproduce the true global potential of the simulated galaxy fairly well and without obvious biases.

5 SECTIONS OF STREAMS

With simulated streams it is convenient to utilize the full stream in our analysis. However, we typically observe only the close-by segments of whole streams or, even if multiple sections are observed, we may not realize that they belong to the same larger structure (e.g. Bonaca et al. 2021).

To investigate the possible consequences, we split our two streams into sections (as specified in Table 1 and Fig. 1) and use them in our analysis individually as if they were independent streams.

Fig. 7 shows the best-fitting results for all sections of stream A (top panels) and stream B (bottom panels). There is considerable variation in the global fit of the different sections, further evidence that individual streams that span only a small region of the position-velocity space can lead to biased estimates of the global galactic potential. The majority of the sections do, however, give a good local prediction of the potential. With two exceptions, the crosses that signify the median distance of the star particles of that section, lie within 25 km s^{-1} of the true velocity curve as shown with the black dash–dotted line.

The galactocentric radii at which each section gives the tightest uncertainties correlates linearly with their median galactocentric radius. The same relationship was found by Bonaca & Hogg (2018) who further discovered that they could tighten this correlation by adding flexibility to their potential model. In other words, the more flexible a model, the more localized the best constraints became. This demonstrates that streams (or stream sections) do not contain information about the entire extent of their orbit but rather are sensitive to the underlying potential at their current location. Furthermore, Peñarrubia et al. (2006) showed that the past history of an evolving gravitational potential cannot be constrained using present-day observables: the properties of stellar streams only reflect the present-day galactic potential.

As streams recover best the current potential at their current location, the variation we see in Fig. 7 is unsurprising – the sections behave as if they were completely separate streams.

Finally, single streams have been shown to produce biased estimates of their host's potential (see e.g. Lux et al. 2013; Bonaca et al. 2014). This serves to add even more complexity to the differences we see in Fig. 7. In the next section, we will explore the underlying causes for the variations we see in both the global and local results between different streams.

6 ORBITAL PHASE

We now organize the sections of both streams into three groups based on their orbital phase: pericentre, intermediate, and apocentre sections. This allows us to, first, explore any systematic differences in the section results based on the phase and, second, perform a joint analysis of all the sections that belong to a particular phase group using our action-clustering method. The interplay between the section results and other stream properties is explored in Section 7. Our findings are summarized in Fig. 8, where the top, middle, and bottom panels present the results for pericentre, intermediate, and apocentre sections, respectively. The right-hand panels show again the individual stream section best-fitting Stäckel potentials while the left-hand panels show the confidence regions of these measurements overlaid in semitransparent grey colour: the brightness of the grey colour tells us where most of the confidence regions overlap. The error bars show each section's measurement at their median distance. The left-hand panels also contain a red line which signifies the best-fitting Stäckel potential of the joint data set of each respective phase. The confidence region for that is shown with the red shaded region in the right-hand panel.

The pericentre sections generally overestimate the mass of their host galaxy both locally and globally, as evidenced by their best-fitting potentials and the associated uncertainties. The lower edge of

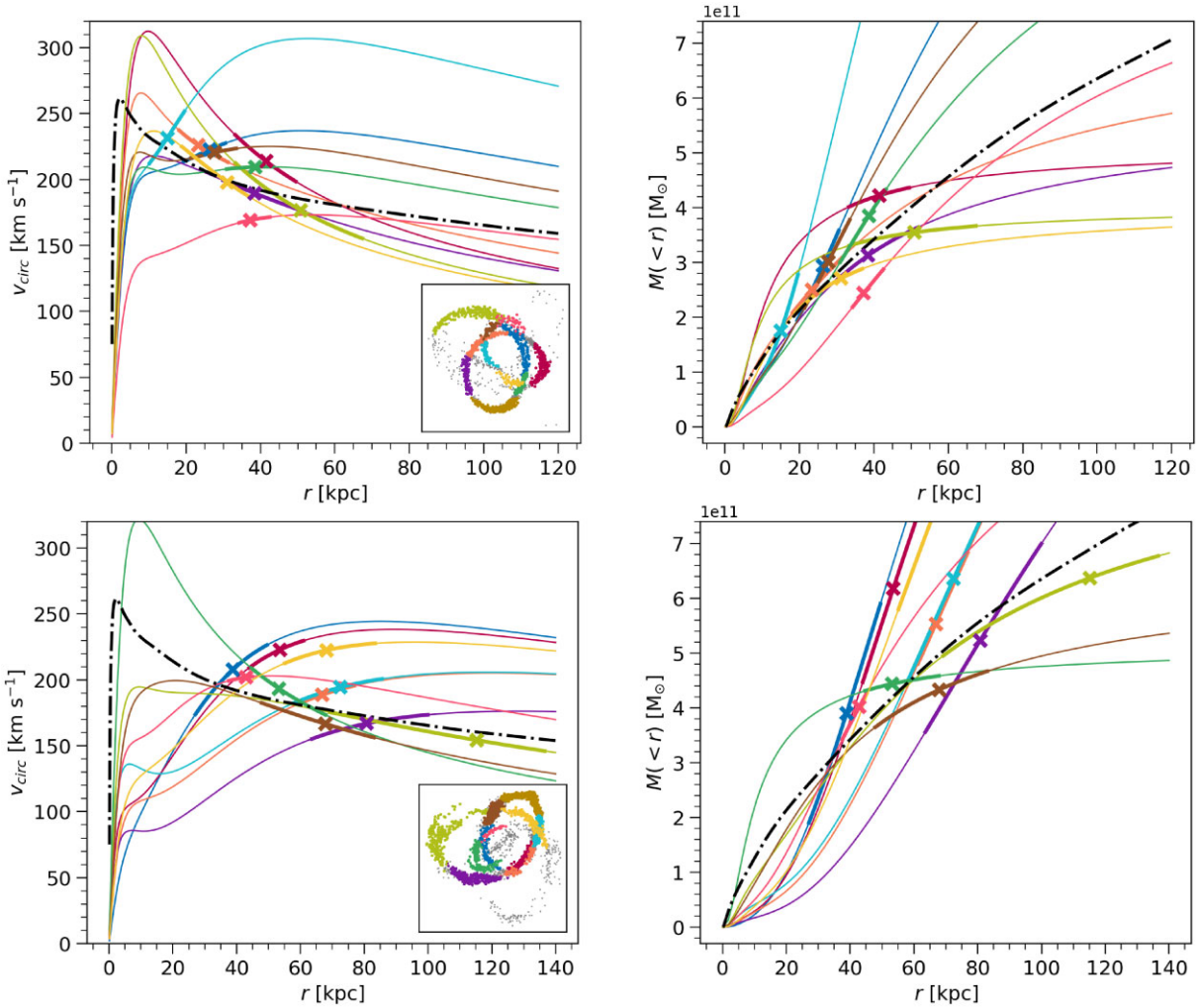

Figure 7. The best-fitting Stäckel potentials for different sections of stream A (top panels) and stream B (bottom panels) shown in circular velocity and enclosed mass profiles. The crosses mark the median distance of each section, while the thicker part of the lines corresponds to the full range of distances of star particles in that section. The insets show the stream in x–z frame for reference.

their confidence regions typically just about covers the true circular velocity curve at high radii, while only a few do so closer in: there is a visible gap in the coverage of the true curve between about 40 and 60 kpc. This is confirmed by the joint phase results: the best-fitting potential is consistently above that of the true potential and the confidence region only barely reaches the true potential at high radii, while being somewhat above it from about 30–100 kpc.

The intermediate sections, on the other hand, show a much better agreement with the true velocity curve. Although there are still large variations in the individual best-fitting potentials in a global sense, the local measurements are better matched. The grey shaded regions now also clearly envelope the true potential across all distance scales. The joint best-fitting potential shows good agreement with the true potential nearly everywhere: we recover the true velocity curve within 6 per cent over the range of radii covered by the data (19–100 kpc).

Finally, the apocentre sections again generally show good agreement with the true velocity curve. The local measurements mostly agree with the true velocities, without showing a preferred bias, while the predicted velocity curves at high radii do mostly prefer lower masses when compared to the true potential. This is also clear when looking at the uncertainty regions: while at low distances the agreement between the stream sections looks fairly chaotic, at high distances, most of the confidence regions overlap slightly below the true velocity curve. The joint best-fitting curve echos these individual results: at high distances we have a fairly good fit, while at low distances the difference is quite large.

7 BIAS DEPENDENCE ON OTHER STREAM PROPERTIES

In Fig. 9, we explore the dependence of the accuracy of our potential fit on several stream section properties (see also Fig. B1 where further stream properties are considered and Appendix B for the description of how these properties were calculated). In the top left and bottom left panels, we show the points with the section-specific colours as per Table 1 while in the rest of the panels the points are coloured based on their orbital phase. The top left panel shows for each section the residuals between the best fit and the true velocity curve at the distance of minimum uncertainty as a function of orbital phase (solid data points). For comparison we also show, with empty markers and lighter colours, the residuals at the median distance of each section (this corresponds to the vertical difference between each of the crosses and the dash–dotted line in Fig. 7). In the bottom left panel, we show the residuals relative to their 1σ uncertainty.

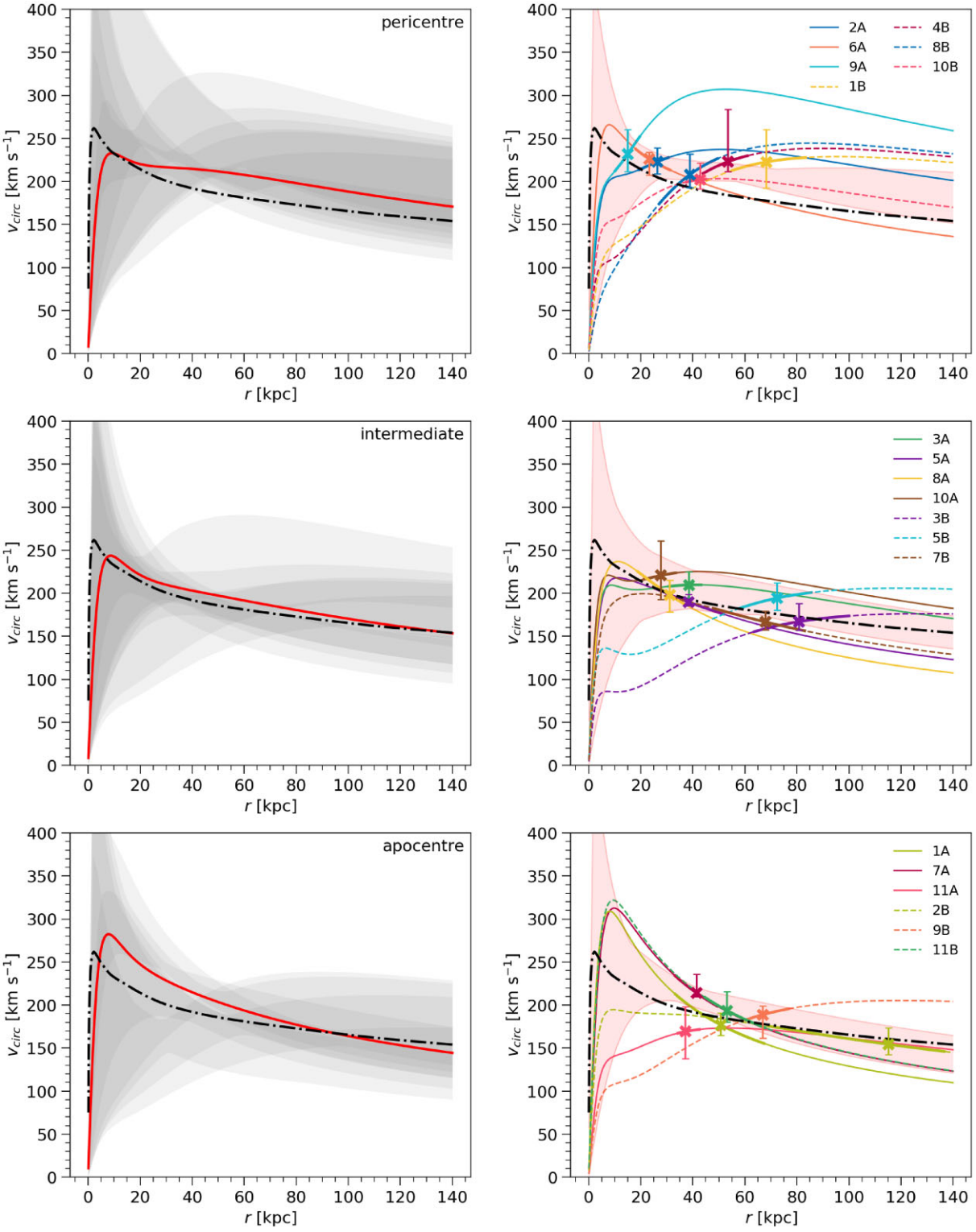

Figure 8. The results of all stream A and stream B sections organized by orbital phase. The top panels show the pericentre sections, while the middle and bottom panels show the intermediate and apocentre sections, respectively. The coloured lines in the right-hand panels show the best-fitting Stäckel potential of the individual sections, with the error bar indicating the uncertainty at their median distance. These are coloured based on Table 1, with stream A sections shown with solid lines and stream B sections shown with dashed lines. Their full confidence regions are shown as a function of radius in the left-hand panel as grey semi-opaque regions. The red line in the left-hand panel shows the best-fitting Stäckel potential of the joint phase data set, and the red shaded area in the right-hand panel is its associated uncertainty. The true potential is shown with a black dash–dotted line in every panel.

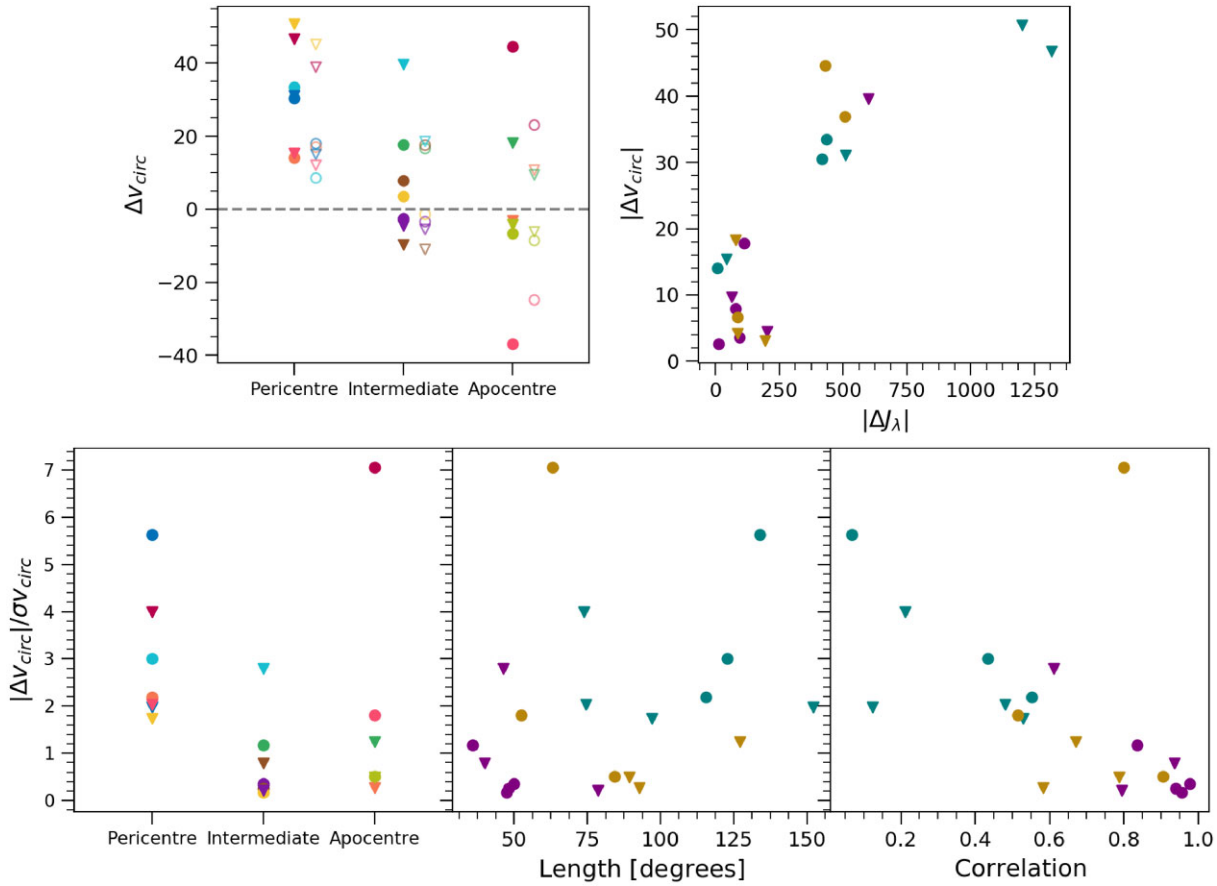

Figure 9. The residuals of the circular velocity curve at the distance of the smallest 1σ uncertainty as a function of the sections' properties. In the top left panel, we also show the residuals at the sections' median distance with lighter empty markers, for comparison. In the left-hand panels, we adopt our section-specific colour scheme, while on the rest of the panels we indicate the pericentre, intermediate, and apocentre sections with teal, purple, and yellow colours, respectively. A description of how the stream properties were computed is available in Appendix B.

We find that there is a clear correlation between the accuracy of the fit and the orbital phase of the stream. The best-fitting results of the intermediate sections show the least amount of scatter around the true velocity curve, and do not appear to have a preferred bias. The apocentre sections, although not exhibiting an obvious bias either, have a lot more scatter around the true potential. The pericentre sections, in contrast, consistently overestimate the mass. Moreover, their uncertainties are small compared to the residuals, indicating a systematic bias.

We do not see any clear trends with either the angular length (bottom middle panel) or physical length (shown in Fig. B1) of the stream sections. Indeed, the intermediate sections in our sample tend to be quite short and nevertheless give better estimates than the often longer pericentre sections. We also see no correlations with the number of stars in each stream section, the median galactocentric distance, the median distance from the galactic plane, width or velocity dispersion in each section (all shown in Fig. B1) nor with the angular momentum vector or the galactocentric distance range covered by a certain stream section (not shown). All of this gives us confidence that it is truly the effect of the orbital phase that causes the variation in the quality of our constraints.

To explain this effect, we investigated the correlation between λ and p_λ in each stream section. The bottom right panel in Fig. 9 shows that stream stars exhibit strong correlations between their motions and positions during the intermediate phase. The pericentre streams meanwhile have the weakest correlations. In general, the stronger the correlations between motions and positions of stream stars the better the constraints from that stream tend to be. Fig. 10 illustrates this concept. On the upper panel, we show a cartoon of a dwarf galaxy stream near pericentre passage for a more radial orbit (top left panel) and a more circular orbit (top right panel). The black points represent stream stars, which have been created by selecting points along the pericentre of a single orbit (the blue line) and adding scatter in both λ and p_λ. The light blue shaded regions approximate the 'scatter' in these orbits. The yellow line shows a fit to the black points and is in both cases almost perfectly vertical, signifying no correlation between λ and p_λ. This in turn indicates lack of statistical power for significant differentiation between different potential models.

The origin of the systematic error that we see arising with pericentre streams therefore lies in this lack of correlation between the positions and momenta of stream stars during pericentre passage. As a consequence this data cannot differentiate between potentials as successfully as intermediate and apocentre streams. In the case of the action-clustering method, this property of the pericentric motion manifests as overestimation of mass. Due to the lack of correlations between motions and positions, it is possible to find a high-mass potential that confines all the stream stars on to orbits with very little radial motion. This potential is then selected as the best fit because it forms a dense cluster near $J_\lambda \sim 0$. However, to accommodate this configuration the stars have to be placed on a variety of different

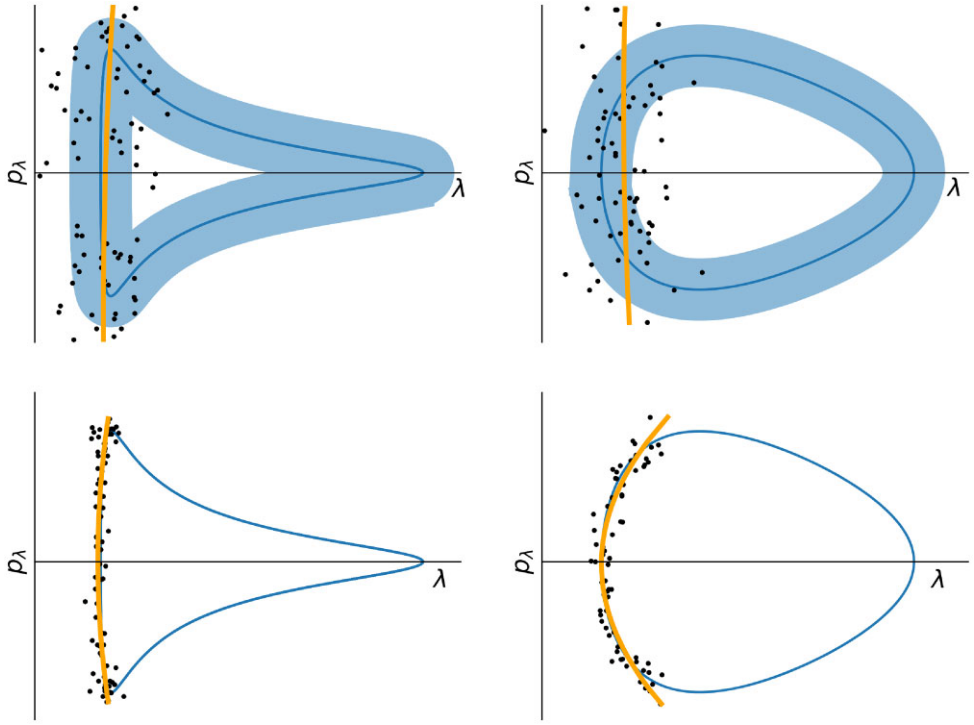

Figure 10. Schematic phase diagram of dwarf galaxy streams (top panels) and cold stellar streams (bottom panels) near pericentre passage for radial orbits (left-hand panels) and circular orbits (right-hand panel). The black points represent stream stars created by selecting points along the pericentre of the orbit shown with the blue line and adding scatter in both λ and p_λ. This scatter is five times larger in both positions and momenta for the dwarf streams compared to the cold stellar streams. The light blue shaded regions on top panels approximate the much larger 'scatter' in the orbits of dwarf galaxy streams. The yellow lines show a fit to the black points on each panel.

phases on their respective orbits. This means that the ordering of the stars along the stream, and in fact the spatial coherence of the stream itself, breaks down in this potential (see also Buist & Helmi 2015). We call this spurious reordering of the stream stars in the incorrect potential 'phase scrambling'.

Although the pericentre streams are likely to yield potentials that are highly biased, the confidence regions can nevertheless be small. This is due to the fact that to calculate the uncertainty, we compare the action-space distribution of the best-fitting potential to that of all other trial potentials, and draw the error contours so that they mark the boundary where the action-distributions begin to significantly differ. The weakness of this technique is that even if there were other potentials that produced a highly clustered action-space (i.e. had a high KLD1) there is no guarantee that they would be included in the uncertainty region if their action-space looks significantly different from that of the best fit, e.g. when clusters simply form at a sufficiently different locations in action-space. So our set-up is reliant on having managed to determine the 'correct' potential as the best fit, while the uncertainty just measures the variation around it.

Finally, in the top right panel, we show the goodness of fit as a function of the difference in median J_λ between the best-fitting potentials of individual sections and that of the combined full streams (green line in Fig. 3). The further the individual results are from the true potential, the bigger this change in J_λ. Although it is expected that a greater difference between two potentials results in a greater change in the action-space, we do not see such a trend with J_ν.

To confirm that the quality of our results is indeed determined by the phase of the stream, we looked at the two stream sections with the highest difference between the true and estimated velocity curves – pericentre sections 1B and 4B – in a simulation snapshot corresponding to redshift $z = 0.038$, when both of these stream sections occupied the intermediate phase. We reapplied the action-clustering algorithm for the stars in each section but now using their past positions and velocities from snapshot $z = 0.038$. The results of this test are presented in Fig. 11, where with the blue line we show the result of the analysis of the current day (pericentre phase) positions of the stream stars, and with the yellow line the past (intermediate phase) positions of the same stars. In both cases, the results originating from the past positions, when the stars were at intermediate phase, perform markedly better.

8 CONCLUSIONS AND DISCUSSION

In this work, we have shown that with two *whole* dwarf galaxy streams we recovered the true rotation curve of the simulated galaxy within 12 per cent over the entire range of radii covered by our set of star particles (10–176 kpc) when adopting the two-component Stäckel model. However, this accuracy is much improved over the distances where we have the most data, recovering the true rotation curve within 6.5 per cent between the 5 and 95 percentile distance range (23–109 kpc). This leads us to conclude that using the Stäckel potential does not introduce a significant bias into our results, at least not more so than any other axisymmetric, parametrized potential model would.

Bonaca et al. (2014) explored the accuracy of smooth analytical potentials in representing realistic – lumpy and time-evolving – dark matter haloes. Using a collection of streams evolved (using the streakline method) in the Via Lactea II simulation, they estimated the galactic mass by comparing these 'observed' streams to models generated in trial analytical potentials and showed that just

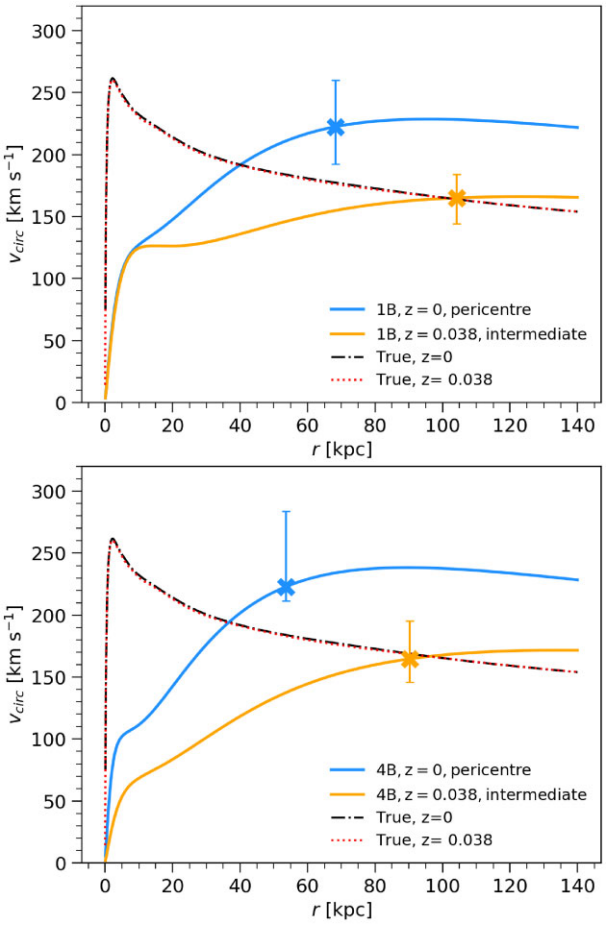

Figure 11. Comparison of the results from analysis of the present day and past (z = 0.038) positions of the stars in sections 1B (top panel) and 4B (bottom panel). The solid coloured lines represent the best-fitting Stäckel potential and the error bar shows the 1σ uncertainty at the median galactocentric distance for each data set. The black and red dash–dotted lines represents the true galactic potential at z = 0 and z = 0.038, respectively.

assuming an analytical potential limits the measurement accuracy to 5−20 per cent. This limit was reached only with the full collection of 256 streams in their sample, while individually the streams were much less accurate (only 40−60 per cent of the individual streams could recover the true parameters within 10 per cent). A similar exploration was performed by Sanderson et al. (2017) who fitted analytical potentials to the streams occurring natively in the Aquarius A simulation using the action-clustering method.[1] They found that with simultaneous fitting of 15 streams they could recover M_{200} within 10 per cent.

The tight constraints we achieve here with just two streams are remarkable considering that, in contrast to Via Lactea II and Aquarius simulations, both of which are dark matter-only simulations, we model a galaxy from a fully cosmological-baryonic simulation which contains a stellar and gas disc shaped by star formation in addition to a time-evolving dark matter halo. Yet we obtain this precision using only a global, two-component Stäckel potential to represent the entire complexity of this galaxy. This is likely due to the good orbital phase coverage of these two streams: both streams have several wraps around the host galaxy and as such cover each orbital phase multiple times.

Both Bonaca et al. (2014, using streakline) and Sanderson et al. (2017, using action clustering) demonstrated that an oversimplification of the potential model does not intrinsically produce a biased mass profile when fitting a collection of streams. Our findings agree with this: our results with the full streams show no presence of systematic bias (Section 4).

We next split each of the two streams into 11 smaller sections based on their orbital phase (Section 5) and analysed them independently. We find that the quality of the constraints on the mass profile depends on the orbital phase of the stream (Section 7). There is a clear systematic bias when using only the pericentre streams in our analysis: this data overestimates the mass of the host galaxy at all galactocentric radii (see Fig. 8, top panel and Fig. 9, lower left panel). This systematic error stems from the fact that during pericentre passage the positions and momenta of stream stars are not correlated (see Fig. 9, bottom right panel). Although a joint fit of multiple streams is usually recommended to get a better fix on the potential, this bias remains even when all pericentre sections are analysed jointly. We find that streams on the intermediate phase are the most likely to give bias-free local mass estimates individually, and a bias-free and accurate global mass profile in combination (see Fig. 8, middle panel and Fig. 9, left-hand panels).

In Reino et al. (2021), we showed that when analysed with the action-clustering method GD-1 data produced a mass estimate that was considerably larger than those from Pal 5, Orphan, and the combination of all three streams. We explored the range of orbital phases the GD-1 stars were on with the best-fitting GD-1 potential and found that the stars were all placed on very different orbital phases on their respective orbits. We briefly discussed that the cause for this, and therefore the high mass that GD-1 recovers, is likely due to the natural energy gradient along the stream not being reproduced. This is another symptom of the phase scrambling we discussed above and, since GD-1 is believed to be a pericentre stream, it aligns with our results here.

This inability of pericentre streams to distinguish robustly between potential models can manifest in other ways for different methods. Previously, Sanders & Binney (2013) remarked on having more difficulty constraining the potential with streams observed at their pericentre. They analysed a mock stream with their angle-frequency slope method both during its apocentre and the subsequent pericentric passage and found that they could not recover the true potential parameters as successfully in the pericentric case. However, despite several local minima in their likelihood surface, they do not calculate a large systematic bias. The cause for such a behaviour could be that streams cover a smaller angle space near their pericentres making the determination of the slope more prone to errors.

Similarly, Koposov et al. (2010) found that they could not constrain all the parameters in their three-component potential model with six-dimensional GD-1 data using the orbit-fitting method and noted that, due to being near its pericentre, GD-1 might not have sufficient phase coverage to differentiate between orbits produced in different potentials, resulting in a poor fit.

Pericentre streams therefore lead to either biased results or weak constraining power irrespective of the applied method. Although our position in the Galaxy and the increased stellar density of streams near pericentre makes detecting streams near their pericentres the easiest, their relative accessibility does not lead to an appreciable

[1]Although mostly similar to the action-clustering method described here, Sanderson et al. (2017) used the product of the marginal distributions of p instead of the uniform distribution as the comparison distribution q, gave equal weight to all stars, and did not separate different streams during the process of density estimation.

improvement in our understanding of the Galactic potential or the field of near-field cosmology.

Although in this work we have made use of dwarf galaxy type streams, the conclusions drawn here are also applicable to most cold stellar streams of globular cluster origin. This is because the cause for the weak constraining power in pericentre streams is not unique to dwarf type streams. In the bottom panels of Fig. 10, we show the cartoon versions of two cold stellar streams to provide comparison with the dwarf galaxy streams in the top panels. Both cartoon streams in the left-hand panels were created from the same original orbit (the blue line), however, the cold stellar stream stars are five times less scattered both in λ and in p_λ. The same holds true for the streams in the right-hand panels: the stars were created from the same original orbit but with five times more scatter added to the dwarf galaxy stream stars. Although with considerably less scatter, the positions and momenta of stars in cold stellar streams on more radial orbits (bottom left panel) would still be uncorrelated during pericentre passage and therefore result in poor constraining power. However, in contrast with dwarf galaxy streams, a cold stellar stream on a more circular rather than radial orbit (bottom right panel) can potentially have sufficient curvature in the pericentre part of the phase diagram to constrain a model potential.

Many studies have shown that in general longer streams have more constraining power. When investigating the information content in the tracks of stellar streams Bonaca & Hogg (2018) found that longer streams (in degrees) achieve the highest precision in recovering the potential parameters. We do not find any correlation between either angular or physical distance with the accuracy of the constraints or the precision of our confidence regions. In fact, our intermediate phase streams often tend to be the shortest and the pericentre streams the longest. However, our shortest streams are around ∼50 deg while only 1 of the 11 streams Bonaca & Hogg (2018) studied reaches this length, the rest being considerably shorter. It could be that the trend with length is no longer as relevant as other factors when it comes to longer streams. Alternatively, the effect could be related to the difference in our methods. Bonaca & Hogg (2018) made comparisons between the tracks of stream data and models in position and velocity space, so longer streams will allow the comparisons to be made over a larger extent and thus enhance the results. Conversely, the length of the stream has no direct impact on the constraints derived with the action-clustering method as we are only measuring the density of the stars in action space.

We summarize our findings as follows:

(i) Although individual streams are likely to deliver accurate estimations of the local galactic profile, they should not be relied on for yielding good global fits.

(ii) We have shown that the pericentre streams can lead to significant systematic errors when used to constrain the potential of their host galaxy.

(iii) Meanwhile apocentre and, especially, intermediate phase streams lead to accurate inference.

For accurate high-confidence constraints on the Galactic potential, we therefore advocate targeting streams that are likely at intermediate or apocentre phases.

ACKNOWLEDGEMENTS

SR would like to thank the Faculty of Science of Leiden University for their support through the Leiden/Huygens fellowship. RES and NP gratefully acknowledge support from National Aeronautics and Space Administration (NASA) grant 19-ATP19-0068. RES is additionally supported by *HST* grant AR-15809 from Space Telescope Science Institute, National Science Foundation grant AST-2007232, and a Scialog-Time Domain Astronomy Award from the Research Corporation for Science Advancement. This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. Simulations used in this work were run using XSEDE supported by National Science Foundation grant ACI-1548562, Blue Waters via allocation PRAC NSF.1713353 supported by the National Science Foundation, and NASA HEC Program through the NASA Advanced Supercomputing (NAS) Division at Ames Research Center. The authors also would like to thank the Flatiron Institute Scientific Computing Core for providing computing resources that made this research possible, and especially for their hard work facilitating remote access during the pandemic. Analysis for this paper was carried out on the Flatiron Institute's 'Iron' computing cluster, which is supported by the Simons Foundation.

DATA AVAILABILITY

The data of the two streams used in this manuscript are available as part of the SAPFIRE: Streams on FIRE catalogue at https://flathub.flatironinstitute.org/sapfire.

REFERENCES

Batsleer P., Dejonghe H., 1994, A&A, 287, 43
Beane A. et al., 2019, ApJ, 883, 103
Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, MNRAS, 478, 611
Belokurov V., Sanders J. L., Fattahi A., Smith M. C., Deason A. J., Evans N. W., Grand R. J. J., 2020, MNRAS, 494, 3880
Binney J., 2012, MNRAS, 426, 1324
Bonaca A., Hogg D. W., 2018, ApJ, 867, 101
Bonaca A., Geha M., Küpper A. H. W., Diemand J., Johnston K. V., Hogg D. W., 2014, ApJ, 795, 94
Bonaca A., Conroy C., Wetzel A., Hopkins P. F., Kereš D., 2017, ApJ, 845, 101
Bonaca A. et al., 2020, ApJ, 897, L18
Bonaca A. et al., 2021, ApJ, 909, L26
Bovy J., Bahmanyar A., Fritz T. K., Kallivayalil N., 2016, ApJ, 833, 31
Buist H. J. T., Helmi A., 2015, A&A, 584, A120
Clementini G. et al., 2019, A&A, 622, A60
Conroy C. et al., 2019, ApJ, 883, 107
Dalton G. et al., 2012, in McLean I. S., Ramsay S. K., Takami H., eds, Proc. SPIE Conf. Ser. Vol. 8446, Ground-Based and Airborne Instrumentation for Astronomy IV. SPIE, Bellingham, p. 84460P
de Jong R. S. et al., 2019, Messenger, 175, 3
de Zeeuw T., 1985, MNRAS, 216, 273
Dejonghe H., de Zeeuw T., 1988, ApJ, 333, 90
Di Matteo P., Haywood M., Lehnert M. D., Katz D., Khoperskov S., Snaith O. N., Gómez A., Robichon N., 2019, A&A, 632, A4
Dierickx M. I. P., Loeb A., 2017, ApJ, 847, 42
Duda R. O., Hart P. E., 1972, Commun. ACM, 15, 11
Erkal D. et al., 2019, MNRAS, 487, 2685
Fardal M. A., Huang S., Weinberg M. D., 2015, MNRAS, 452, 301
Gaia Collaboration, 2016, A&A, 595, A1
Garrison-Kimmel S. et al., 2018, MNRAS, 481, 4133
Garrison-Kimmel S. et al., 2019, MNRAS, 487, 1380
Gibbons S. L. J., Belokurov V., Evans N. W., 2014, MNRAS, 445, 3788
Grillmair C. J., Freeman K. C., Irwin M., Quinn P. J., 1995, AJ, 109, 2553
Haywood M., Di Matteo P., Lehnert M. D., Snaith O., Khoperskov S., Gómez A., 2018, ApJ, 863, 113

Helmi A., White S. D. M., de Zeeuw P. T., Zhao H., 1999, Nature, 402, 53
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Ibata R. A., Gilmore G., Irwin M. J., 1994, Nature, 370, 194
Ibata R. A., Malhan K., Martin N. F., 2019, ApJ, 872, 152
Koposov S. E., Rix H.-W., Hogg D. W., 2010, ApJ, 712, 260
Koppelman H. H., Helmi A., Massari D., Price-Whelan A. M., Starkenburg T. K., 2019, A&A, 631, L9
Kuijken K., Gilmore G., 1989, MNRAS, 239, 571
Kullback S., 1959, Information Theory and Statistics. Wiley, New York
Kullback S., Leibler R. A., 1951, Ann. Math. Stat., 22, 79
Kunder A. et al., 2017, AJ, 153, 75
Küpper A. H. W., Lane R. R., Heggie D. C., 2012, MNRAS, 420, 2700
Küpper A. H. W., Balbinot E., Bonaca A., Johnston K. V., Hogg D. W., Kroupa P., Santiago B. X., 2015, ApJ, 803, 80
Law D. R., Majewski S. R., 2010, ApJ, 714, 229
Levi M. et al., 2019, BAAS, 51, 57
Li T. S. et al., 2019, MNRAS, 490, 3508
Lux H., Read J. I., Lake G., Johnston K. V., 2013, MNRAS, 436, 2386
Malhan K., Ibata R. A., 2019, MNRAS, 486, 2995
Malhan K., Yuan Z., Ibata R. A., Arentsen A., Bellazzini M., Martin N. F., 2021, ApJ, 920, 51
Mateu C., Read J. I., Kawata D., 2018, MNRAS, 474, 4112
Myeong G. C., Vasiliev E., Iorio G., Evans N. W., Belokurov V., 2019, MNRAS, 488, 1235
Naidu R. P., Conroy C., Bonaca A., Johnson B. D., Ting Y.-S., Caldwell N., Zaritsky D., Cargile P. A., 2020, ApJ, 901, 48
Newberg H. J., Carlin J. L., 2016, Astrophysics and Space Science Library, Vol. 420, Tidal Streams in the Local Group and Beyond. Springer, Cham
Newberg H. J., Willett B. A., Yanny B., Xu Y., 2010, ApJ, 711, 32
Panithanpaisal N., Sanderson R. E., Wetzel A., Cunningham E. C., Bailin J., Faucher-Giguère C.-A., 2021, ApJ, 920, 10
Pearson S., Clark S. E., Demirjian A. J., Johnston K. V., Ness M. K., Starkenburg T. K., Williams B. F., Ibata R. A., 2021, preprint (arXiv:2107.00017)
Peñarrubia J., Benson A. J., Martínez-Delgado D., Rix H. W., 2006, ApJ, 645, 240
Reino S., Rossi E. M., Sanderson R. E., Sellentin E., Helmi A., Koppelman H. H., Sharma S., 2021, MNRAS, 502, 4170
Sanders J., 2012, MNRAS, 426, 128
Sanderson R. E., Helmi A., Hogg D. W., 2015, ApJ, 801, 98
Sanderson R. E., Hartke J., Helmi A., 2017, ApJ, 836, 234
Sanderson R. E. et al., 2020, ApJS, 246, 6
Sanders J. L., Binney J., 2013, MNRAS, 433, 1826
Sanders J. L., Binney J., 2016, MNRAS, 457, 2107
Santistevan I. B., Wetzel A., El-Badry K., Bland-Hawthorn J., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., Benincasa S., 2020, MNRAS, 497, 747
Sesar B. et al., 2017, AJ, 153, 204
Sharma S., Johnston K. V., 2009, ApJ, 703, 1061
Shih D., Buckley M. R., Necib L., Tamanas J., 2021, MNRAS, in press
Vasiliev E., 2019, MNRAS, 482, 1525
Vasiliev E., Belokurov V., Erkal D., 2021, MNRAS, 501, 2279
Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, ApJ, 827, L23

APPENDIX A: STREAM VELOCITIES

In Fig. A1, we show the galactocentric velocities for our two streams with each section highlighted in the colour as specified by Table 1.

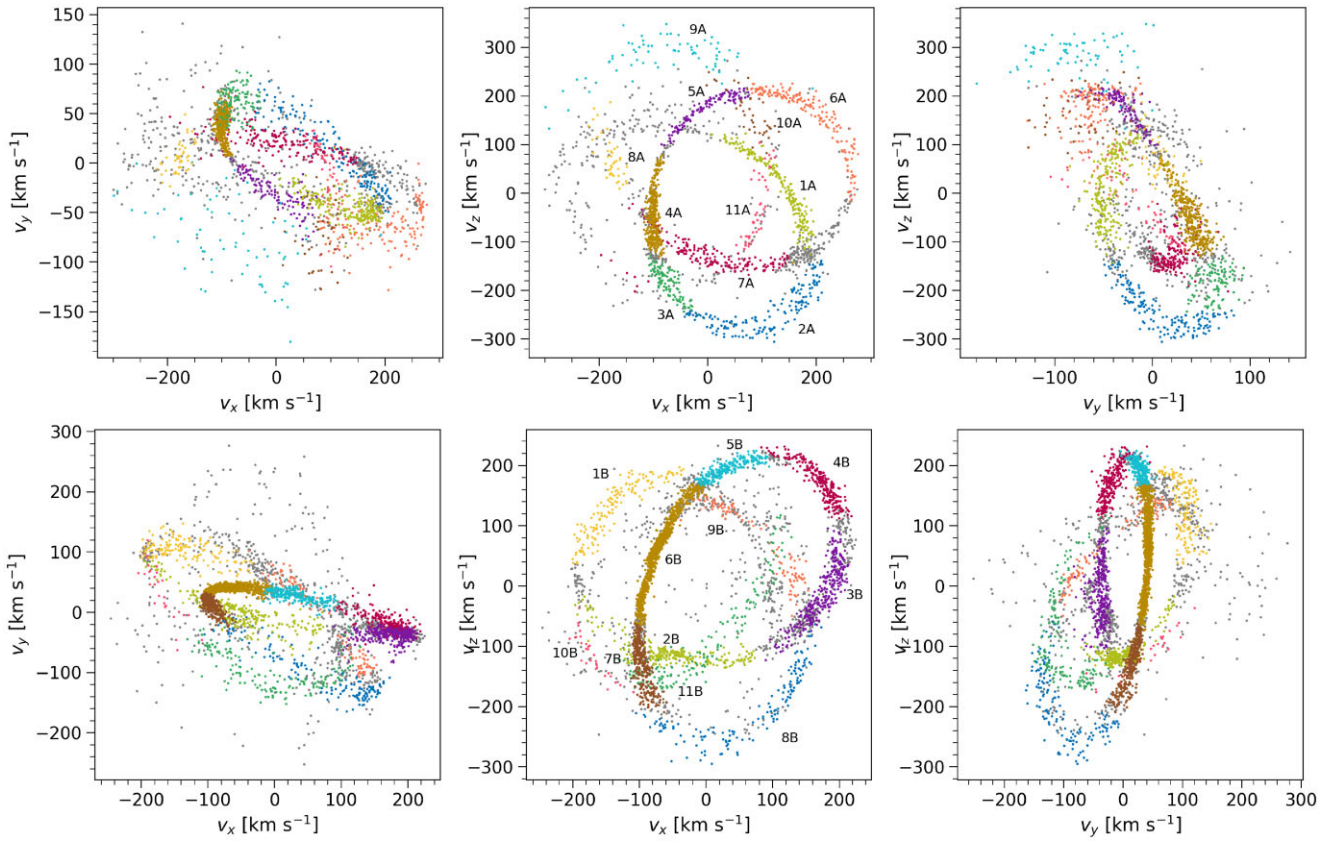

Figure A1. Galactocentric velocities of streams A (top) and B (bottom). Each stream section is identified by the colour given in Table 1 and a label. Star particles not belonging to any section are shown as grey.

APPENDIX B: BIAS AS A FUNCTION OF FURTHER STREAM PROPERTIES

In this section, we show in Fig. B1 the accuracy of our potential fit as a function of some further stream section properties. As already remarked in Section 7, none of these stream section characteristics are sufficient to explain the variation that we see in the quality of our potential constraints.

To measure the stream length, width, and velocity dispersion, we first convert to stream-aligned coordinates. This is a spherical coordinate system (r, ξ, η) where the equator, $\eta = 0$, is defined by a great circle best fitting the stream data centred on the galactic centre and the radius, r is the median galactocentric radius of the stream stars. The angular length of the stream is then defined as $\Delta\xi$. The physical length is found by computing the arc length of the circle subtended by the stream, i.e. $r \times \Delta\xi$. To estimate the width of the stream, we fit the galactocentric x, y, and z coordinates of stream stars as a function of the angle along the stream, ξ, with a quadratic polynomial. The distance of each star from this stream 'axis' can then be calculated at their respective ξ, i.e. $d_i = \sqrt{(x_i - x(\xi_i)^2 + (y_i - y(\xi_i)^2 + (z_i - z(\xi_i)^2}$. We then define the width as the root mean square of these distances. An analogous technique is employed to calculate the velocity dispersion, σ_v, except now the polynomial is fitted to the galactocentric v_x, v_y, v_z as a function of ξ. Correlation between λ and p_λ for each stream is defined as the absolute value of the Pearson correlation coefficient between these coordinates.

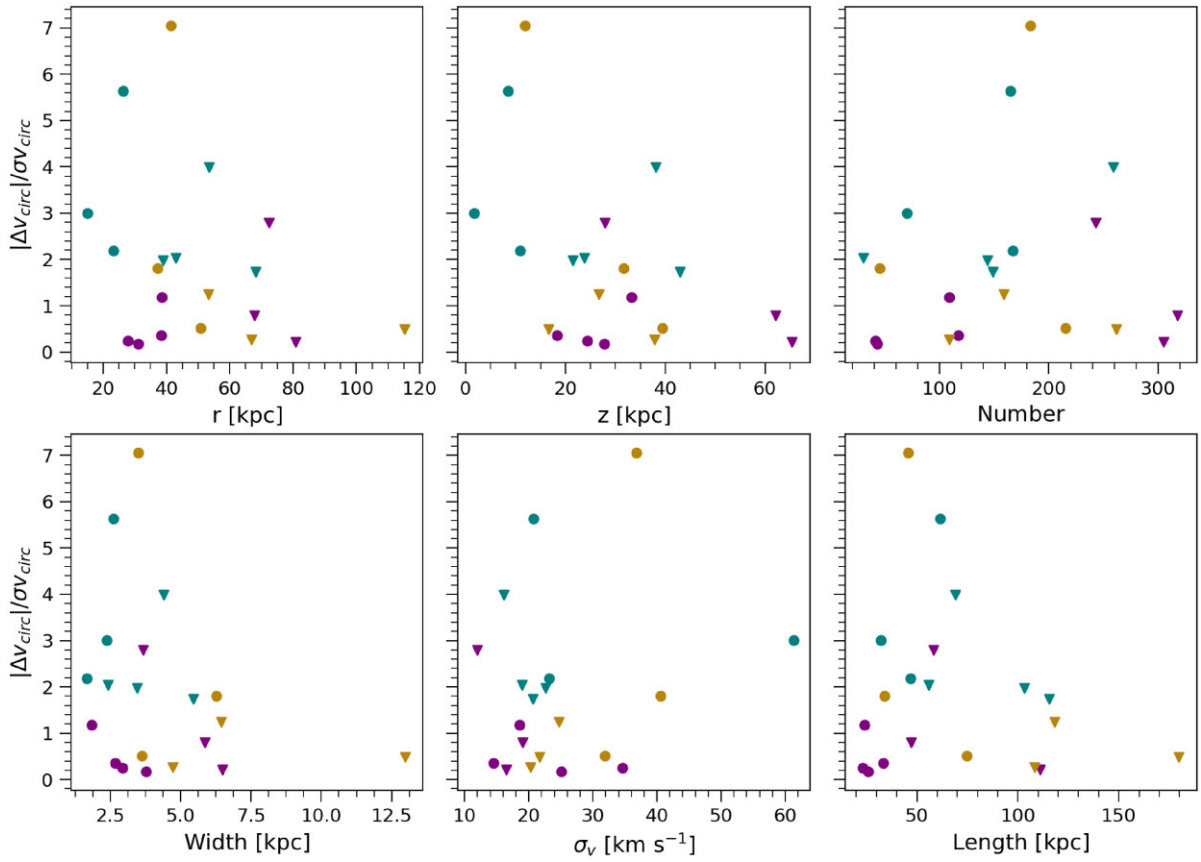

Figure B1. The residuals of the circular velocity curve at the distance of the smallest 1σ uncertainty as a function of the sections' properties. We indicate the pericentre, intermediate, and apocentre sections with teal, purple, and yellow colours, respectively.

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