

VIA MACHINAE 2.0: Full-Sky, Model-Agnostic Search for Stellar Streams in *Gaia* DR2

David Shih,¹★ Matthew R. Buckley,¹ and Lina Necib^{2,3}

¹NHETC, Dept. of Physics and Astronomy, Rutgers, Piscataway, NJ 08854, USA

²Department of Physics and Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA 02139, USA

³The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge MA 02139, USA

6 March 2023

ABSTRACT

We present an update to VIA MACHINAE, an automated stellar stream-finding algorithm based on the deep learning anomaly detector ANODE. VIA MACHINAE identifies stellar streams within *Gaia*, using only angular positions, proper motions, and photometry, without reference to a model of the Milky Way potential for orbit integration or stellar distances. This new version, VIA MACHINAE 2.0, includes many improvements and refinements to nearly every step of the algorithm, that altogether result in more robust and visually distinct stream candidates than our original formulation. In this work, we also provide a quantitative estimate of the false positive rate of VIA MACHINAE 2.0 by applying it to a simulated *Gaia*-mock catalog based on *Galaxia*, a smooth model of the Milky Way that does not contain substructure or stellar streams. Finally, we perform the first full-sky search for stellar streams with VIA MACHINAE 2.0, identifying 102 streams at high significance within the *Gaia* Data Release 2, of which only 10 have been previously identified. While follow-up observations for further confirmation are required, taking into account the false positive rate presented in this work, we expect approximately 90 of these stream candidates to correspond to real stellar structures.

Key words: Galaxy: Stellar Content – Galaxy: Structure – Stars: Kinematics and Dynamics

1 INTRODUCTION

As dwarf galaxies and globular clusters are tidally stripped by the Milky Way’s gravitational field, they form extended stellar structures called streams. The composition (see e.g. Yuan et al. 2022; Li et al. 2022), orbits (see e.g. Shipp et al. 2021; Li et al. 2022; Shipp et al. 2022), and internal structure (see e.g. Carlberg et al. 2012; Bonaca et al. 2019a) of these streams are valuable probes into the merger history (Johnston 1998; Helmi & White 1999; Belokurov et al. 2006; Shipp et al. 2018; Belokurov et al. 2018; Malhan et al. 2021, 2022), structure (Johnston et al. 1999; Ibata et al. 2001; Koposov et al. 2010; Newberg 2010; Varghese et al. 2011; Purcell et al. 2012; Sanders & Binney 2013; Küpper et al. 2015; Malhan & Ibata 2019; Necib et al. 2019; Reino et al. 2020; Tronrud et al. 2022), and substructure of the Galaxy (Carlberg et al. 2012; Sanders et al. 2016; Erkal et al. 2017; Bonaca et al. 2019a; Banik & Bovy 2019; Bonaca et al. 2020). As the number of wide field and all-sky surveys has grown, the number of known stellar streams has likewise increased (see Mateu 2022, for a summary). Most recently and notably, the *Gaia* Space Telescope has opened a new frontier in the discovery and study of stellar streams, providing both all-sky coverage and proper motion measurements for 1.5 billion stars (Gaia Collaboration et al. 2016a,b, 2018, 2021, 2022).

A number of automated techniques have been developed to search for stellar streams within the *Gaia* data (see e.g. Malhan & Ibata 2018; Malhan et al. 2018a; Yuan et al. 2018; Meingast & Alves 2019; Borsato et al. 2020; Meingast et al. 2019; Ibata et al. 2020). Generally speaking, existing approaches require some model-dependent assumptions about the dynamics and astrophysics of the Milky Way. For example, the STREAMFINDER algorithm (Malhan & Ibata 2018) – by far the most comprehensive automated approach to stream finding to date – identifies stellar streams by searching the Milky Way for stars that occupy the same position/velocity “hypertube” in their projected orbits through the Galaxy; this requires an assumption for the Galactic potential, something that is still uncertain and a topic of current research (see e.g. Koposov et al. 2022; Lilleengen et al. 2023).

In a prior paper (Shih et al. 2022, hereafter referred to as Paper I), we developed VIA MACHINAE: a new, fully-automated, machine-learning-assisted algorithm to discover stellar streams within the *Gaia* data. Unlike other stream finding algorithms, VIA MACHINAE is *model-agnostic*, in that it does not assume anything about the form the Galactic potential, orbits, or isochrones.

VIA MACHINAE is based on the ANODE algorithm (Nachman & Shih 2020), a simulation-independent, data-driven, unsupervised machine learning method for anomaly detection originally developed for the Large Hadron Collider. ANODE uses normalizing flows (for reviews and original references, see Kobayev et al. (2019); Papa-

★ E-mail: shih@physics.rutgers.edu

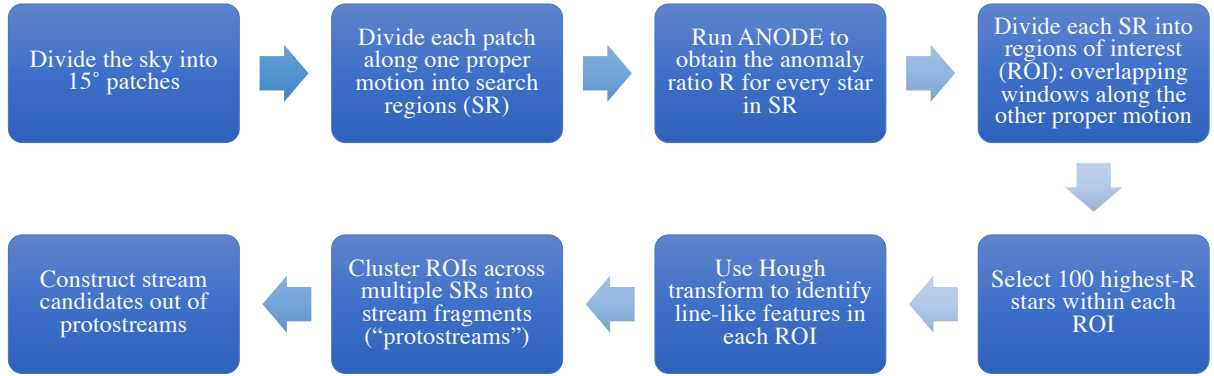


Figure 1. Flowchart of VM2. In this updated version, numerous steps, including the line-finding algorithm, and the stages whereby the ROIs are clustered into full stream candidates, have been improved in order to produce higher quality stream candidates, as well as to minimize the false positivity rate. Also, in VM2 we have run ANODE on both orthogonal proper motions instead of just a single one; this increases the chances and robustness of stream detections. See Sections 2 and 3 for details.

makarios et al. (2019)) to estimate the probability density of data in a “search region,” as well as a background probability density from sidebands. (See Nachman & Shih (2020) and Paper I for more details.) Applied to the angular positions, proper motions, color and magnitude of the stars in the *Gaia* catalog, VIA MACHINAE first uses ANODE to identify stars that are anomalous (overdense) with respect to the background. Subsequent steps of VIA MACHINAE aim to specialize to overdensities that are broadly consistent with stellar streams: localized in both proper motion coordinates, with locally line-like spatial extent on the sky, and color and magnitudes consistent with old, metal-poor stars, but not necessarily following a specific isochrone.

In Paper I, we developed the steps of VIA MACHINAE, and demonstrated their efficacy in rediscovering GD-1 (Grillmair & Dionatos 2006), a particularly extended, narrow, and dense stellar stream. While the basic outline of VIA MACHINAE has not changed from Paper I, in this work we present many refinements and improvements to the algorithm in order to improve the quality of the discovered stream candidates and its robustness against false positives. We will refer to the new-and-improved version of VIA MACHINAE as VM2 throughout this work.

Figure 1 shows the schematic of the steps of the VM2 algorithm, which are:

(i) We divide the sky into overlapping patches, 15° in radius. Patches too close to the Galactic disk are discarded, as are patches containing the Large Magellanic Cloud or significant dust occultations. Each patch has its own centered angular coordinates λ and ϕ , obtained from rotating the ICRS right ascension (α) and declination (δ) coordinates. The rest of the analysis is done in these patch-centered coordinates.

(ii) The stars in each patch are then divided into overlapping *search regions* according to one of their two proper motion coordinates μ_λ and μ_{ϕ^*} . These regions have a set width of 6 mas/yr, and the centers of the overlapping regions are separated by 1 mas/yr. The complement of each search region defines its paired *control region*. In VM2, we perform separate scans with search regions defined in μ_λ and again in μ_{ϕ^*} . Only the first proper motion coordinate was used in the analysis of Paper I.

(iii) The ANODE algorithm is then run on each search region, using the associated control region to determine a background probability density. The result is an anomaly score for each star in the

search region:

$$R(x) = \frac{p_{\text{SR}}(x)}{p_{\text{bg}}(x)}, \quad (1)$$

where $p_{\text{SR}}(x)$ is the probability density of the star in the signal region, and $p_{\text{bg}}(x)$ is the background probability density of the star, obtained by interpolating from the control region. Here x are all the features of the star used in training ANODE, which for VIA MACHINAE consists of $x = (\lambda, \phi, \mu', g, b-r)$, with μ' being the orthogonal proper motion coordinate that was not used to define the search region, g the *Gaia* magnitude, and $b-r$ the *Gaia* color.

We note that each star will typically appear in multiple search regions and patches, since these are highly overlapping. Therefore, each star will typically have multiple $R(x)$ values. As the anomaly score depends on the exact selection of stars within the search and control regions (as well as stochastic dependence on the training process), each $R(x)$ constitutes approximately-independent tests for the anomalous properties of the star. Below, we will describe how these multiple quasi-independent $R(x)$ values are combined to build more robust, higher-confidence stream candidate detections.

(iv) We apply a set of *fiducial selection cuts*, selecting stars whose angular coordinates are within the inner 10° of the patch, $g < 20.2$, and $0.5 < b-r < 1$. To these cuts – which were implemented in Paper I – we add a new cut, designed to remove disk stars, which act as foreground in the search of stellar streams located in the stellar halo. For stars with well-measured parallaxes, we require their Galactocentric altitude $|z|$ to be greater than 2 kpc at 95% CL (the full details of this new criteria are described below). These fiducial cuts are imposed *after* the ANODE training, as hard boundaries within the data set would result in errors in the density estimation near the edges.

(v) We further subdivide each search region into overlapping regions of interest (ROIs) by selecting windows of width 6 mas/yr in the orthogonal proper motion coordinate that was not used to define the search region (e.g., for search regions defined in μ_λ , the ROIs are constructed by slices in μ_{ϕ^*}).¹ Within each ROI, we select the 100 stars with the highest $R(x)$ values.

(vi) Using a Hough transform (Hough 1959; Duda & Hart 1972a),

¹ Note this results in two sets of ROIs for each patch that have the same boundaries in μ_λ and μ_{ϕ^*} . However, as the training of ANODE occurs before the division into ROIs, the R values of these two sets are distinct.

we perform an automated line-finding on the angular positions of these 100 stars for each ROI. This line-finding process returns, in addition to the location and orientation of the best-fit line, a statistical significance for the line. Compared to Paper I, this step has been modified to better estimate the background and statistical significance of the most-probable line of stars within any ROI.

(vii) We expect a real stream to appear as an anomaly within multiple overlapping ROIs in a given patch. We therefore cluster together the lines from ROIs whose line location, orientation, and proper motions are concordant. We refer to these combined ROIs as “protoclusters.” The combination process also results in a single protocluster significance parameter obtained from re-running the line-finding on the concatenation of all the ROIs in the protocluster. Compared to Paper I, the algorithm for this line combination step has been greatly improved in VM2, with more flexible and accurate measures of concordance between ROIs, resulting in fewer spurious combinations. We also now combine lines from the ROIs trained on μ_λ and μ_ϕ^* within the same patch at this step, which was not possible in Paper I given that the analysis there was conducted in a single proper motion parameter.

(viii) Finally, we construct stream candidates by clustering the protoclusters from different patches into a single set of stars on the sky, again requiring a consistent set of proper motion values and line directions across the stream candidate. From the protocluster significances, a single stream significance parameter is calculated, which can be used to reject false positives. Again, as compared to Paper I, the algorithm used in this step is improved to more accurately combine the fragments of a stream located in different patches, taking into account the fact that the proper motion and apparent direction of the stream on the sky can change along its path.

In addition to the improvements to VIA MACHINAE described above, this paper contains a quantitative estimate the false positive rate (fpr) of VIA MACHINAE for stream finding, using the set of *Gaia*-like observations described in Rybizki et al. (2018). These synthetic stars were drawn from a completely smooth model of the Milky Way with no substructure, thus any stream candidates our algorithm detects in this data set are spurious.

Finally, in this paper we present the first results from an all-sky scan of the *Gaia* Data Release 2 (DR2) dataset using VIA MACHINAE. Altogether, we find 102 stream candidates, and with the fpr derived for this working point, we may expect ~ 90 of the stream candidates to be real stellar streams. VM2 is able to rediscover six previously known streams, as well as fragments of Sagittarius. The rest of the stream candidates found by VM2 do not correspond to known streams. We briefly describe 15 of the most-significant new stream candidates identified using VM2, saving a more detailed study for an upcoming follow-up work.

The remainder of the paper is organized as follows. In Section 2, we describe the *Gaia* data used in this analysis, as well as our methods to detect and remove regions which are too contaminated with dust, globular clusters, or thick disk stars for our stream-finding algorithm to return sensible results. Section 3 explains in detail the changes and improvements to the VIA MACHINAE algorithm from Version 1 to the current Version 2. In Section 4, we estimate the false-positive rate of VIA MACHINAE using the *Gaia* mock-catalog. Lastly, we apply VM2 to the entire *Gaia* DR2 sky and describe our results in Section 5.

We note that the technical details of our data processing in Section 2 and of our algorithm in Section 3 are rather dense and may not be of interest to all readers. The high-level description of the VM2 algorithm provided here in the Introduction is intended to be sufficient to understand the algorithm for most readers, and provide

the necessary context for the fpr calculations in Section 4 and the stellar streams identified in Section 5.

2 DATA SET

2.1 Overview

In Paper I, we demonstrated the VIA MACHINAE method using a data set derived from *Gaia* DR2. While the current work was already underway, *Gaia* early Data Release 3 (eDR3) and Data Release 3 (DR3) was published. As we do not use parallax, radial velocity, or spectroscopic information from individual stars in the machine learning stage of our algorithm, the primary advantage of (e)DR3 data over DR2 for our purposes is the greatly improved and reduced measurement errors, which likely reduces false positive stream detections and improves the purity of our stream candidates with respect to contamination by non-stream stars. However, given that significant computational resources are required for the machine learning algorithm, we continue to present results from DR2 in this paper. In addition, the results of a VIA MACHINAE scan on DR3 would not completely supersede our DR2 results; rather, given the stochastic nature of the training of the machine learning algorithm, evidence for streams provided by these scans could be considered quasi-independent: we expect that real streams would robustly show up in both DR2 and DR3 scans, while spurious false positives in DR2 would be less robust and might not be confirmed by DR3. An analysis of the DR3 data set is forthcoming in a future work.

The data we use in the VIA MACHINAE method (both here and in Paper I) is organized into three nested levels: patches, signal regions (SRs), and regions-of-interest (ROIs). We now review their definitions in more detail (see Figure 2 for a graphical illustration):

- *Patches*: Starting with *Gaia* DR2, we select stars with measured parallax less than 1 mas. We do not correct for the zero-point parallax offset as we do not use the detailed parallax information, and therefore we do not expect this to have any significant effect on our analysis. We construct overlapping circular patches on the sky of radius 15° , with patch centers in Galactic coordinates selected using HEALPY (Górski et al. 2005; Zonca et al. 2019) (with `nside=5`). Patches near the disk contain a large number of stars, and data can be highly affected by the presence of dust. The former issue results in prohibitively long training at the ANODE step, and the latter causes spurious anomalous features. For these reasons, we remove all patches with centers that have $|b| < 30^\circ$. We further exclude those that overlap with the Large or Small Magellanic Clouds, as well as those patches away from the disk that contain significant dust features (these are primarily located near the Galactic bulge). We are left with a total of 163 patches overall, which form the basis for our all-sky scan. (This should be contrasted with Paper I, which focused on just 22 patches containing GD-1.)

We select the angular position, proper motion, photometric magnitude g , and photometric $b - r$ color as features of each star, to be used in the ANODE training. In each patch, we recenter the ICRS coordinate system to a set of local angular coordinates ϕ and λ using the built-in centering function of the `SkyOffsetFrame` object in the `ASTROPY` package (Astropy Collaboration et al. 2013, 2018, 2022). The associated proper motions are likewise rotated from the ICRS coordinate system to the patch-centered coordinates: μ_λ and $\mu_\phi^* \equiv \mu_\phi \cos \lambda$. These steps are all the same as in Paper I.

- *Signal regions (SRs)*: Within each patch, we divide stars into overlapping search SRs in proper motion space. In Paper I, we used only the μ_λ coordinate to define the SRs; here, we construct two sets:

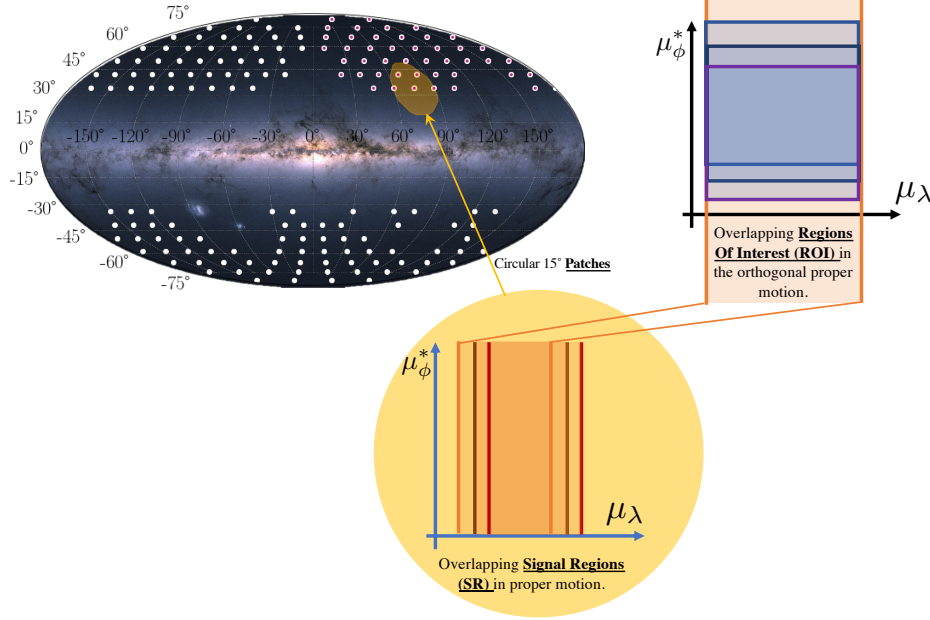


Figure 2. Definition of patches, signal regions (SRs), and regions of interest (ROIs). See Section 2.1 for details. The dots on the sky map indicate the centers of the patches considered in the analysis. The analysis was run on the full sky with *Gaia* (see Sec 5), and on the quarter of sky shown with red dots with *Gaia-Mock* (see Sec 4).

first using μ_λ , and then a separate set using μ_{ϕ^*} . In each case, the SRs are defined as those stars within the given proper motion's range of 6 mas/yr, with the center of each SR separated by 1 mas/yr from its neighbors (see Fig. 2):

$$[\mu_{\min}, \mu_{\max}] = \dots, [-10, -4], [-9, -3], \dots, [3, 9], [4, 10], \dots \quad (2)$$

SRs with fewer than 2×10^4 stars or more than 10^6 stars are excluded from further analysis. At the low end, this number of stars is not sufficient for accurate ANODE training; at the high end the training times become prohibitively long. Each SR implicitly defines its control region (CR), the complement of stars outside the selected range of proper motions. In summary, there are 5,152 SRs in μ_λ and 5,465 SRs in μ_{ϕ^*} . That corresponds to an average of 33 SRs in each patch and each proper motion coordinate.

The SRs and their CR counterparts are fed into the ANODE training step, described in detail in Paper I and briefly reviewed in section 3.1. ANODE produces an overdensity score $R(x)$ for every star in the SR; by cutting on this score, we can discover stellar streams hidden in the data. This is the core of the VIA MACHINAE method.

- **Regions of Interest (ROIs):** After running the ANODE method on the SRs, and before running the subsequent steps of VIA MACHINAE, we further divide up the SRs into ROIs by cutting on the other, orthogonal proper motion coordinate. In other words, if the SR is defined using μ_λ (μ_{ϕ^*}) $\in [n, n+6]$ mas/yr then the ROI is defined by additionally restricting μ_{ϕ^*} (μ_λ) $\in [m, m+6]$ mas/yr. Like SRs, the cuts on the orthogonal proper motion coordinate are also defined with width 6 mas/yr and stepsize 1 mas/yr, so the ROIs are overlapping square regions in proper motion space (see Fig. 2). The purpose of further dividing up the data into ROIs is that streams are supposed to be localized in both proper motion coordinates, so by restricting to the stars in an ROI, we hope to enhance the signal to noise ratio of any stream that might be present. All in all, there are 126,727 ROIs

corresponding to the μ_λ SRs and 129,830 ROIs corresponding to the μ_{ϕ^*} SRs.

Finally, after dividing the data into ROIs, two more selection criteria are required before running the subsequent steps of VM2: identifying and excluding ROIs that contain globular clusters or dwarf galaxies, and defining a fiducial region in which we perform the rest of the analysis; this region is chosen to maximize the signal-to-noise of stream candidates. Since both of these steps are considerably changed from Paper I, we will describe them in more detail in the following two subsections.

2.2 New Globular Cluster/Dwarf Galaxy finding algorithm

Like stellar streams, globular clusters (GCs) and dwarf galaxies (DGs) are collections of stars in compact regions of kinematic phase space, and are therefore anomalous compared to the background. The difference lies in the clustering of these objects in different spaces: Streams are extended in one angular direction and localized in the orthogonal direction as well as the proper motions. GCs and DGs, however, are compact in all angular and proper motion coordinates. Due to the large number of stars within a GC/DG and their anomalous properties as identified by ANODE, we find that the presence of a GC or DG within a search region can cause complete failure of our later steps in identifying stream candidates. This issue was identified in Paper I, but our initial algorithm to remove GCs and DGs was not sufficiently robust for the full-sky analysis. We provide the revamped analysis here.

To locate the GC/DG overdensities, we make a two-dimensional histogram of all the stars within a SR in the angular coordinates ϕ and λ , using 120×120 bins (enumerated by i and j in the ϕ and λ coordinates respectively) to cover the full 15° patch. The goal is to identify significant overdensities in the number of stars within each

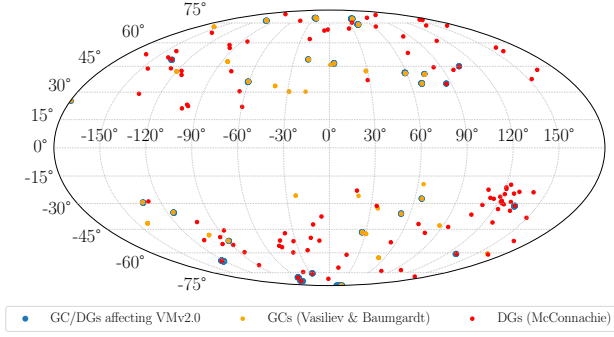


Figure 3. The GC/Dwarf galaxy candidates identified and removed in our analysis (blue), overlaid with the known GCs from Vasiliev & Baumgardt (2021) and DGs from McConnachie (2012). We only show the GCs and DGs contained in the analyzed region of this work (163 patches of the *Gaia* DR2 scan, as defined in Sec. 2.1). This figure confirms that every GC/DG identified by our algorithm as affecting VM2 is indeed a known object. The converse is of course not true – the vast majority of GCs and DGs are too dim and/or far away to even show up in the *Gaia* DR2 catalog.

pixel, N_{ij} , as compared to neighboring pixels. Given that this set up is similar to identifying line-like structures in Hough space, we develop a common algorithm that can be adapted for both problems. This algorithm is described in detail in Appendix A; briefly, the overdensity in each pixel is quantified as the difference between N_{ij} and the average counts in an annulus centered on the i, j pixel, normalized to the statistical and systematic errors. For the GC/DG-finding, we use an annulus with a width and height of 11 pixels, minus the inner 3×3 pixels. The normalized significance assigned to each pixel by this procedure is labeled S_{ij} .

We find that pixels with $S_{ij} \lesssim 9$ are indistinguishable from noise by eye, whereas pixels with $S_{ij} \gtrsim 50$ are clearly recognizable as localized objects. In between, we find that not every bright pixel with $9 < S_{ij} < 50$ causes ANODE to fail. We therefore construct a two-step cut as follows. Given a GC/DG candidate pixel with $S_{ij} > 9$, we consider every ROI r in the SR. Let $N_{r,ij}$ be the number of the 100 highest- R stars in the ROI localized within a 0.5° circle around the GC/DG candidate. We then exclude SRs containing pixels that satisfy

$$S_{ij} > 50 \quad \text{OR} \quad (S_{ij} > 9 \quad \text{AND} \quad \max_r N_{r,ij} > 10). \quad (3)$$

In other words, we exclude SRs containing the highest significance pixels, as well as the pixels with somewhat lower significance but that contain a high number of anomalous stars. In all, 1,174 (1,098) SRs are excluded from the analysis, leaving 3,978 (4,367) SRs corresponding to the μ_λ (μ_{ϕ^*}) directions.

A map of all the GC/DG candidates identified and removed by our analysis is shown in Figure 3, overlaid with all known GCs (Vasiliev & Baumgardt 2021) and DGs (McConnachie 2012) that are contained within the 163 patches in our *Gaia* DR2 scan. We see that every object identified and removed by our method corresponds to a known GC or DG, but there are many more existing GCs and DGs not removed by our analysis. This is due in part to the fact that our tests have shown that not every bright localized GC/DG-like object caused ANODE to fail, and we opted therefore not to remove these regions from our analysis. Furthermore, many GCs and DGs are simply too faint or distant to be detectable in the *Gaia* DR2 data set.

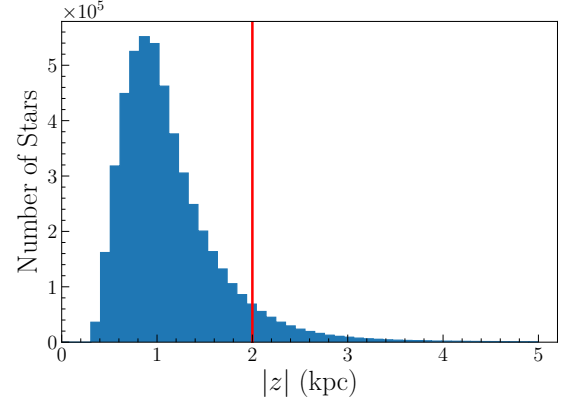


Figure 4. Absolute value of Galactocentric z , for all stars with well-measured parallax (Eq. 4) in our data set.

2.3 New fiducial region: excluding thick disk stars

In Paper I, we imposed several fiducial cuts post-ANODE training based on the quality of the density estimation (avoiding sharp edges in probability distributions), namely $g < 20.2$ and $r < 10^\circ$. We also removed stars with both proper motions too close to zero ($|\mu_\lambda| < 2$ mas/yr and $|\mu_{\phi^*}| < 2$ mas/yr), to remove a population of very distant stars whose presence would overwhelm the ANODE anomaly finder. Finally, we required $(b - r) \in [0.5, 1]$ after ANODE training, in order to target cold stellar streams containing old, metal-poor stars.

In this work, we do not target close streams within the disk (for stream searches that do target this region of the Milky Way, see for example Helmi & White 1999; Myeong et al. 2018a,b; Necib et al. 2020). To avoid foreground contamination, particularly that of the Galactic disk, we therefore exclude nearby stars. Significant contamination from bright foreground stars can indeed be seen in the initial formulation of the VIA MACHINAE algorithm: Figure 13 (lower right panel) of Paper I shows a population of bright ($G \lesssim 16$) stars superimposed on our detection of GD-1. This contamination is not unique to the GD-1 stream – disk stars are spatially and kinematically coherent, which makes it likely that ANODE will identify them as anomalous.

Fortunately, these stars can be removed using the following new fiducial cut that we impose. This cut focuses only on stars that have “well-measured parallaxes” ϖ , which we define as

$$\varpi > 0 \quad \text{and} \quad \frac{\sigma_\varpi}{\varpi} < 0.2, \quad (4)$$

where σ_ϖ is the measured parallax error. Approximately 16% of the stars across our 163 patches of *Gaia* DR2 satisfy these requirements. For these stars, we have enough information to transform their coordinates to cartesian Galactic coordinates. In Figure 4, we plot the vertical distance $|z|$ (altitude relative to the Galactic midplane) of these stars. As expected, the vast majority of these stars have low values of z , implying that they are predominantly disk stars.² To remove these stars from our stream search, we impose the new fiducial cut

$$|z_{95}| > 2 \text{ kpc} \quad (5)$$

on all stars with measured parallaxes. Here, $|z_{95}|$ is the z coordinate obtained by transforming the 95% upper limit on the parallax, $\varpi +$

² The absence of stars with $|z|$ close to zero is due to our initial cuts on parallax $\varpi < 1$ mas and Galactic latitude $|b| > 30^\circ$.

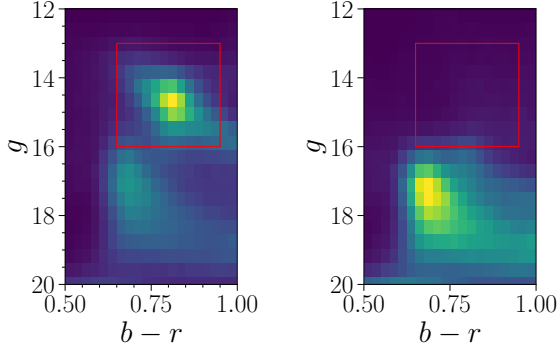


Figure 5. Left: Observed magnitude g versus $b - r$ color for the collection of 100 highest- R stars from all ROIs across the entire *Gaia* data set. Red rectangle indicates bright stars that are contamination from the Galactic disk. Right: The same but after the selection on the vertical distance z given by Eq. (5).

$2\sigma_w$, to Cartesian coordinates.³ The specific value of 2 kpc was chosen to correspond to approximately twice the scale height of the Milky Way thick disk (Li & Zhao 2017). This selection is imposed after running ANODE, along with our other existing fiducial cuts described above.

In Figure 5, we illustrate the effectiveness of this cut with a two-dimensional histogram of the color and magnitude of the 100 highest- R stars in every ROI across our entire *Gaia* data set. The left panel shows all the stars, while the right panel only shows the stars remaining after the selection of Eq. (5). The red rectangle is added to visually emphasize the location of the bright disk stars removed by this fiducial cut on $|z_{95}|$.

3 VIA MACHINAE 2.0

3.1 Initial steps

We now describe the VIA MACHINAE algorithm in detail. An overview has been provided in the Introduction (Sec. 1), and we recapitulate some of that information here. We begin by describing the initial steps of VM2. These steps are mostly unchanged from Paper I where we refer the reader for more details.

For each SR in each circular 15° patch defined in Section 2.1, we run the less-than-supervised anomaly detection algorithm known as ANODE (Nachman & Shih 2020). The input features to ANODE are $\vec{x} = (\lambda, \phi, \mu', g, b - r)$, where μ' is the orthogonal proper motion to the one used to define the SR (for more details, see Section 2.1). ANODE is based on a technique for density estimation known as normalizing flows (for recent reviews, see Kobyzev et al. (2021); Papamakarios et al. (2021)), in particular the architecture of Masked Autoregressive Flows (MAF) (Papamakarios et al. 2017). For full details of the MAF architecture, see Paper I.

By training ANODE on the stars in the SR and then interpolating the sideband density estimate from the control region into the SR, we obtain two estimates of the phase space density of stars in the SR. Taking the ratio of these densities gives us an anomaly score $R(x)$ for each star x in the SR. This R value is an estimate of how overdense that star is in the phase space defined by the input features, relative

to what would be expected given the distribution of stars in control region (which presumably does not contain a stellar stream, if one exists in the range of proper motions contained in the SR).

After the ANODE training, we impose fiducial cuts on the data (Section 2.3) and further divide up the SRs into ROIs (Section 2.1). We select the 100 highest- R stars in each ROI, which will be used to search for and construct the stellar streams.

The subsequent steps of VM2 (line finding, protoclustering, streamclustering) are quite different from Paper I, so we will now describe each subsequent step in more detail.

3.2 Line-finding

The next step is to look within the 100 most anomalous stars of each ROI for line-like features that would be the signature of a stellar stream. As in Paper I, we use the Hough transform (Hough 1959; Duda & Hart 1972b) to identify lines within the angular positions of these stars, and to provide a parameter that defines the significance of the line. As part of our improvements to VM2, we modify our definition of this significance parameter from Paper I.

The Hough transform parametrizes all possible lines through a point in (ϕ, λ) space in terms of a *Hough curve* in *Hough space* (ρ, θ) :

$$\rho = \phi \sin \theta - \lambda \cos \theta. \quad (6)$$

Here, ρ is the distance of closest approach between the center of the patch $(\phi, \lambda) = (0, 0)$ and the line inclined at angle θ to the $\hat{\phi}$ axis that passes through the point (ϕ, λ) . If many stars lie on the same line, then many Hough curves will pass through the same (ρ, θ) point. Thus, the Hough transform converts the problem of line-finding into the simpler task of identifying high-density points in the Hough space of ρ and θ .

For each ROI, we bin the Hough space in a 100×100 grid from $|\rho| \leq 10^\circ$ and $0 \leq \theta \leq \pi$. We index the bins with $i, j = 0, \dots, 99$, such that the map between bin number (i, j) and the left edge of the (ρ, θ) bin is given by $\rho = -10 + i/5$ and $\theta = \pi \times j/100$.

Let N_{ij} denote the number of stars whose Hough curves pass through a box in Hough space with corners (i, j) and $(i + \Delta i, j + \Delta j)$. Different lines cover varying amounts of the patch, due to their trajectory across the circle. In order to fairly compare the number of stars in each possible line, we normalize N_{ij} by the length of the line across the patch.

The hyperparameters Δi and Δj define the thickness of the streams we are sensitive to. In this work, we have picked relatively small values $\Delta i = 5$ and $\Delta j = 3$ (corresponding to $\Delta \rho = 1^\circ$ and $\Delta \theta = 0.09$ rad), so we are explicitly searching for narrow streams (akin to GD-1) as well as somewhat broader streams. Streams covering more of an individual patch would require a larger choice for Δi and/or Δj .

To find the most-significant line parameters in the binned Hough space, we use the same overdensity algorithm as in the GC-identification in Section 2.2, and described in detail in Appendix A. Here, we use an annulus of width and height of 7 pixels, strided by $\Delta i = 5$ and $\Delta j = 3$ and masking the inner 3×3 pixels. The stride is so as to not double-count stars in overlapping boxes. The significance of the pixel i, j using the overdensity algorithm is $\sigma_{\text{line}, ij}$ (see Appendix A for details of the significance calculation). For each ROI, we select the bin with the highest-significance σ_{line} as the single “best” line candidate.

In Figure 6, we show an example of the Hough transform applied to high- R stars of an ROI containing the known stream Jhelum (Shipp et al. 2018). The large overdensity in the binned Hough space (left panel) indicates that many of the stars in this ROI lie along a single

³ Using the upper limit instead of the mean value of the z coordinate is a tighter fiducial cut as $|z_{95}| < |z|$.

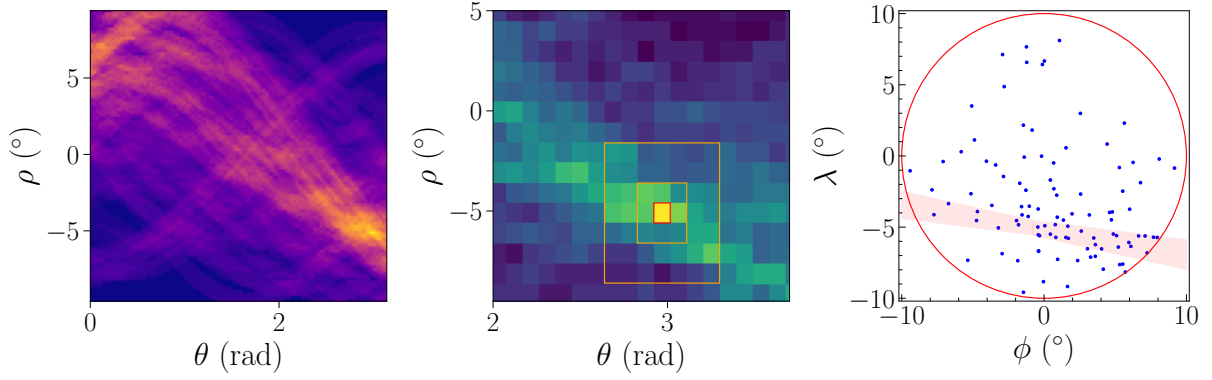


Figure 6. Left: all the counts in binned Hough space, for an ROI that belongs to Jhelum. The intensity of each pixel corresponds to the number of stars that lie within a corresponding box in Hough space, defined by width $\Delta i = 5$ and $\Delta j = 3$, normalized to the length of the corresponding line segment in the patch. Middle: the same 2d histogram as left, but zoomed in around the highest significance pixel, and taking only every pixel strided by $\Delta i = 5$ and $\Delta j = 3$ in the left 2d histogram. The red box shows the highest significance pixel and the orange boxes indicate the 7×7 strided annulus (with the central 3×3 pixels removed) that is used to estimate the background level. Right: the lines that bound the highest significance bin in Hough space.

line. The middle panel of Figure 6 is a zoomed-in look at the highest-significance Hough-space bin and its neighbors, indicating also the annulus in Hough-space used to compute the significance. Finally, the right-hand panel shows the best-fit line converted back to angular position-space, with the range of ρ and θ values spanned by the highest-significance bin shown.

3.3 New protoclustering algorithm

The line-finding step described in Sec. 3.2 identifies high- R stars within an ROI that lie along a relatively narrow line in angular position, producing a significance parameter σ_{line} . As ROIs are overlapping in proper motion, real stellar streams are expected to be identified by the line-finder in multiple “nearby” ROIs within the same patch. In order to obtain the full stream, we must combine the high-significance lines from different ROIs, requiring that they have consistent orientations in position-space as well as having similar values in proper motion-space. This combination of multiple self-consistent ROIs within a *single patch* is referred to as a “proto-cluster.”

Given that there are $O(10^5)$ ROIs, we need an efficient, automated clustering algorithm to group together ROIs into protoclusters. The algorithm from Paper I combined nearby ROIs through *single-linkage clustering*; an ROI was joined to a protocluster if the distance (in line parameters and proper motion space) between the ROI and any ROI within the protocluster was less than some tunable threshold. This resulted in protoclusters composed of a chain of ROIs, each of which is sufficiently close to a neighbor to complete the linkage. However, the net result was a very diffuse assembly in proper motion, something we do not expect for real streams.

In VM2, we shift from single-linkage clustering to an approach based on hierarchical, iterative clustering. At each step of the new clustering algorithm, we have groups of ROIs

$$C_i = \{ROI_{a_1(i)}, ROI_{a_2(i)}, ROI_{a_3(i)}, \dots\} \quad (7)$$

The basic idea of the new algorithm is that two groups, C_{i_1} and C_{i_2} , are joined together if their aggregate proper motions are “close enough” (to be defined below) and if the line finder run on their concatenation returns a higher significance than on each one separately. This new algorithm results in high-significance protoclusters that by-eye are

more similar to properties of known streams (e.g. more compact in proper motion space).

We now describe our new protoclustering algorithm in more detail. We begin with some definitions.

- We call a group of ROIs $C = \{ROI_1, ROI_2, ROI_3, \dots\}$ *valid* if the ROIs are fully pairwise independent, in the sense that they came from pairwise distinct SRs. ROIs from different SRs have their stars chosen by different runs of ANODE, so given the stochastic nature of the neural network training, we expect their anomaly scores to be quasi-independent. Therefore, a stream-like structure that appears in multiple independent ROIs is more likely to be real, compared to a stream-like structure that appears in only a single ROI. Conversely, ROIs that come from the same SR are highly correlated, since their stars were chosen by the same run of ANODE. Thus any structure that appears in multiple ROIs derived from the same SR is not necessarily more likely to be real.

- We define the *line significance* of a group of ROIs to be the significance of the line-finder *re-run* on the concatenation of the highest- R stars from each ROI:

$$\sigma(C) = \sigma_{\text{line}}(ROI_1 \oplus ROI_2 \oplus ROI_3 \dots). \quad (8)$$

Note that we do *not* delete duplicate stars in this concatenation, we deliberately double count them, as there is information in the result that a particular star has high- R values from independent runs of ANODE.

- We define the *proper motion distance* between C_1 and C_2 to be

$$\chi^2_{\mu}(C_1, C_2) = \frac{(\langle \mu_{\lambda} \rangle_1 - \langle \mu_{\lambda} \rangle_2)^2}{\sigma_1^2(\mu_{\lambda}) + \sigma_2^2(\mu_{\lambda})} + \frac{(\langle \mu_{\phi^*} \rangle_1 - \langle \mu_{\phi^*} \rangle_2)^2}{\sigma_1^2(\mu_{\phi^*}) + \sigma_2^2(\mu_{\phi^*})}. \quad (9)$$

Here $\langle \rangle_{1,2}$ denotes the mean taken over the stars in $C_{1,2}$ and $\sigma_{1,2}^2$ the variances.

- Two valid ROI groups C_1 and C_2 are *mergeable* if their concatenation is valid, and if first, their line significance grows upon concatenation

$$\sigma(C_{i_1} \oplus C_{i_2}) > \sigma(C_{i_1}), \sigma(C_{i_2}) \quad (10)$$

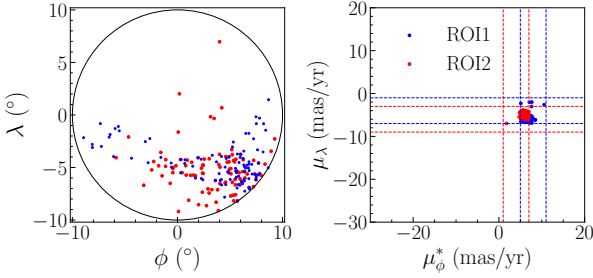


Figure 7. Two ROIs from Jhelum that are clustered together into a protocluster, shown in angular position space (left) and proper motion space (right). The proper motion limits defining each ROI are shown in dashed lines with color matching that of the ROI stars (red for ROI1, blue for ROI2)

and second they are “close enough” in proper motion:⁴

$$\chi_\mu^2(C_{i_1}, C_{i_2}) < 1. \quad (11)$$

Now, with all these definitions in hand, we are ready to describe our new protoclustering algorithm. Given a set of valid ROI groups C_1, C_2, \dots in a patch, consider all mergeable pairs. Among all mergeable pairs, take the one C_{i_1}, C_{i_2} with the highest significance after merging. Replace this pair in the list of valid ROI groups with a new ROI group $C_{i_1 i_2} = C_{i_1} \oplus C_{i_2}$. Repeat until there are no more mergeable pairs of valid ROI groups remaining.

The algorithm is initialized by defining each individual ROI as its own valid ROI group, i.e. $C_1 = \{ROI_1\}$, etc.. After it finishes, one obtains a collection of protoclusters that can be ordered by their significance.

Figure 7 shows an example of the above procedure: two ROIs that are clustered together for the real stream Jhelum. ROI_1 (blue) is defined by $(\mu_\lambda)_{\min} = -7$ mas/yr, $(\mu_\phi^*)_{\min} = 5$ mas/yr with the R values derived from ANODE scan over μ_ϕ^* . ROI_2 (red) is defined by $(\mu_\lambda)_{\min} = -9$, $(\mu_\phi^*)_{\min} = 1$ with the R values derived from ANODE scan over μ_λ . Thus these two ROIs are independent as required. Their line significances are $\sigma(ROI_1) = 6.5$, and $\sigma(ROI_2) = 6.0$, and their concatenation $C = \{ROI_1, ROI_2\}$ has increased line significance $\sigma(C) = 8.1$. Finally, the proper motion distance between the two ROIs is $\chi_\mu^2(ROI_1, ROI_2) = 0.7$.

Through our characterization of the fpr using simulated *Gaia*-like observations, to be described in detail in Section 4, we are motivated to impose a cut of $\sigma(C) > 8$ on the protocluster significance going forward. That is, we will only consider protoclusters with significance greater than this threshold for the subsequent stream-clustering steps of VM2.0, discussed in Sec 3.5. Below $\sigma(C) = 8$, all of the *Gaia* protoclusters are consistent with being false positives, according to both visual inspection and our study of *Gaia-Mock* protoclusters (see especially Fig. 11 in Sec 4).

3.4 Merging Duplicate Protoclusters

The requirement that protoclusters be made up of ROIs that are fully pairwise independent means that our protoclustering algorithm can result in multiple duplicate protoclusters corresponding to the same stellar stream within a given patch. We must therefore develop some set of criteria to decide if two protoclusters C_1 and C_2 within the same patch are “duplicates” of the same object, and if so, merge

them. Since the constituent ROIs of the merged protoclusters are not fully independent of each other, the merging algorithm differs from that used to combine ROIs into independent protoclusters.

Consider two protoclusters C_1 and C_2 in a patch. Each protocluster has a best-fit line associated with it; let the line stars of each be \mathcal{L}_1 and \mathcal{L}_2 respectively. C_1 and C_2 are considered duplicates if \mathcal{L}_1 and \mathcal{L}_2 overlap above a certain threshold. To quantify this overlap we define

$$f_{\text{line}} = \max \left(\frac{|\{s \in \mathcal{L}_2 | s \in \mathcal{L}_1\}|}{|\mathcal{L}_2|}, \frac{|\{s \in \mathcal{L}_1 | s \in \mathcal{L}_2\}|}{|\mathcal{L}_1|} \right). \quad (12)$$

Here, $|\{s \in \mathcal{L}_1 | s \in \mathcal{L}_2\}|$ denotes the number of stars in Line 1 which are also in Line 2 (with the obvious extension to $|\{s \in \mathcal{L}_2 | s \in \mathcal{L}_1\}|$), and $|\mathcal{L}_{1,2}|$ are the number of stars in Lines 1 and 2. Thus our parameter f_{line} is the fraction of one protocluster’s line stars that are found in the other protocluster’s set of line stars.⁵

Two protoclusters are considered duplicates if

$$f_{\text{line}} > 0.4, \quad (13)$$

i.e., more than 40% of the line stars of one protocluster are found in the line stars of the other protocluster. The threshold of 40% has been found through testing on high significance protoclusters that are clearly duplicates by eye, especially those corresponding to previously known streams.

Occasionally, two protoclusters are clearly duplicates, by visual inspection of their highest- R stars, but the line finding algorithm results in different best-fit lines that do not satisfy Eq. (13). To account for this, apart from the overlap in total number of stars explicitly described in Eq. (12), we also measure the overlap fraction of the highest- R stars, via

$$f_{\text{highest-}R} = \max \left(\frac{|\{s \in \mathcal{C}_2 | s \in \mathcal{C}_1\}|}{|\mathcal{C}_2|}, \frac{|\{s \in \mathcal{C}_1 | s \in \mathcal{C}_2\}|}{|\mathcal{C}_1|} \right), \quad (14)$$

and we consider two protoclusters as duplicates if

$$f_{\text{highest-}R} > 0.65, \quad (15)$$

where again, this threshold was set after extensive tests on known streams and other high-significance protoclusters that are clearly duplicates by eye.

Using these two criteria, $f_{\text{line}} > 0.4$ OR $f_{\text{highest-}R} > 0.65$, we merge all the protoclusters in a patch into *protostreams*. A protostream \mathcal{P} is the union of a group of duplicate protoclusters: $\mathcal{P} = \{C_1, C_2, \dots\}$ which satisfy the two overlap-fraction criteria using simple single-linkage clustering, i.e. a protocluster C_i is merged into an existing protostream if it passes the overlap-fraction criteria with *any single* protocluster within the protostream.

We now add one last criterion to our merging algorithm: Our overlap-fraction criteria (especially that of Eq. (15)) can sometimes result in constituent protoclusters in a protostream whose best-fit lines disagree visually; the protoclusters are not well-aligned. If we denote the highest significance protocluster in a protostream by C_1 , we then require all other protoclusters $C_i \in \mathcal{P}$ to satisfy:

$$|\theta(C_1) - \theta(C_i)| < 9^\circ, \text{ and } |\rho(C_1) - \rho(C_i)| < 1^\circ, \quad (16)$$

where θ and ρ are the Hough transform parameters found, discussed in Section 3.2. Any protocluster in \mathcal{P} that does not satisfy the requirements given in Eq (16) is *deleted* from the protostream. It is justified to remove these lower significance protoclusters as these are

⁴ The threshold for χ_μ^2 was chosen after inspection of known streams, the detection of false streams in the *Gaia-Mock* data (see Section 4), and the by-eye quality of high-significance streams from the *Gaia* data.

⁵ We note that $|\{s \in \mathcal{L}_1 | s \in \mathcal{L}_2\}|$ and $|\{s \in \mathcal{L}_2 | s \in \mathcal{L}_1\}|$ are not necessarily the same, because the line stars can be duplicated through the concatenation of the ROIs.

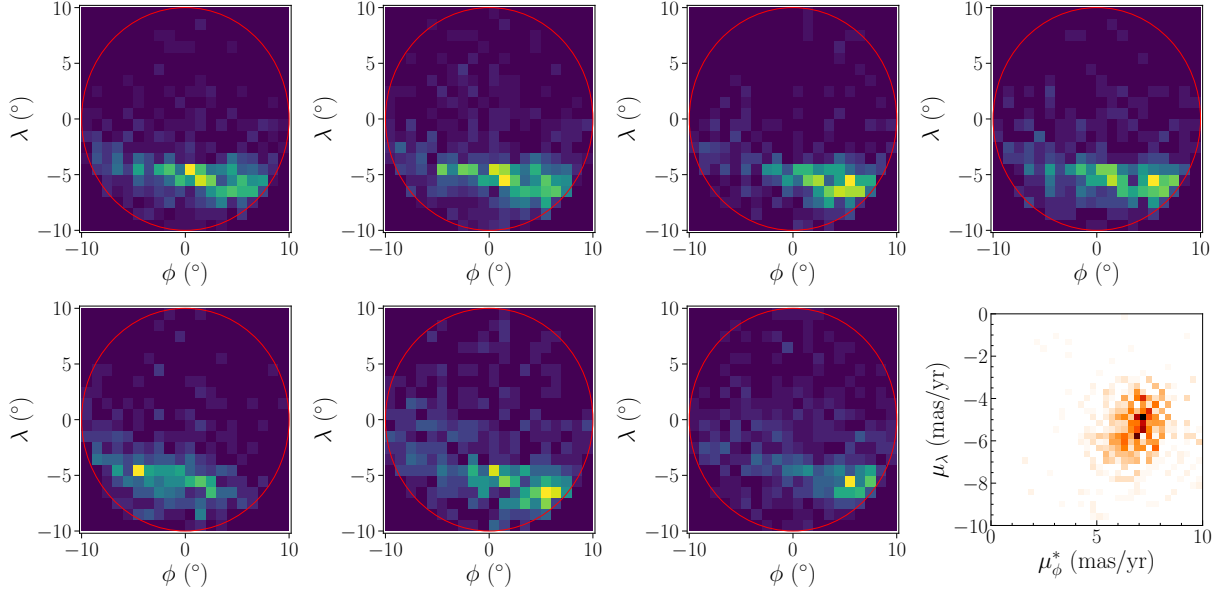


Figure 8. An example protostream belonging to Jhelum. This protostream contains 7 duplicate protoclusters. Shown in the last panel is a 2D histogram of the proper motions of all the protocluster line stars (color indicates number density in linear scaling).

two protoclusters with significant overlap at the level of individual stars, but their best-fit lines are discrepant, so they are not likely to be both correct.

After this last step, the resulting set of protoclusters in a protostream all have best fit lines that agree visually with one another, and we can take the spread of these lines as an approximate measure of “uncertainty” on the protostream itself. We define the significance of a protostream to be the significance of its highest-significance protocluster,

$$\sigma(\mathcal{P}) = \max\{\sigma(C_1), \sigma(C_2), \dots\}. \quad (17)$$

As an example, we consider a protostream that contains Jhelum (the same ROIs and protocluster considered in the previous subsection). This protostream contains seven duplicate protoclusters with max significance $\sigma(C) = 15.2$. These are shown in Figure 8 as individual 2D histograms, as well as a single 2D histogram of the proper motions of all their line stars. As can be seen, our merging algorithm correctly identifies these seven duplicate protoclusters as corresponding to the same underlying object (Jhelum). The seven duplicate protoclusters have an average proper motion dispersion of $\bar{\sigma}_{\mu_{\alpha^*}} = 1.04$ mas/yr and $\bar{\sigma}_{\mu_{\delta}} = 1.05$ mas/yr, which is in good agreement with the proper motion dispersions (0.7 – 1.2 mas/yr) reported in Bonaca et al. (2019b).

3.5 New stream clustering algorithm

The final step of VM2, after obtaining the unique protostreams in each patch, is to link the protostreams of *different* patches in the sky to produce connected stream candidates. We have made many modifications to the stream clustering algorithm relative to Paper I to improve the quality of the stream candidates (according to a by-eye test) and to reduce the fpr as measured by a scan using synthetic *Gaia* observations of a smooth simulated Milky Way (described in detail in Section 4). We now describe the newly adapted stream clustering algorithm.

To merge protostreams, we must determine whether they agree in

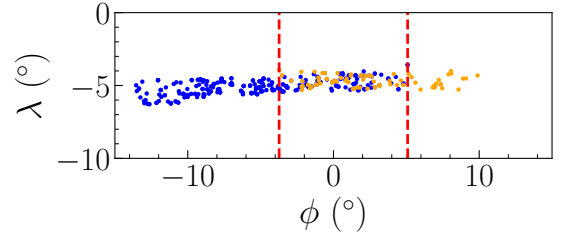


Figure 9. An example illustrating the protostream merging criteria, again for Jhelum. Shown in blue and orange are protocluster line stars from two protostreams belonging to two separate patches. The stars are shown in local coordinates centered on the midpoint between the patch centers and rotated so the line stars are aligned along the x -direction. The overlap region between the protocluster stars is indicated by the red dashed vertical lines.

both proper motion and line direction, similarly to the metrics previously developed for the merging of protoclusters. Given a pair of protostreams $\mathcal{P}^{(1)} = \{C_1^{(1)}, C_2^{(1)}, \dots\}$ and $\mathcal{P}^{(2)} = \{C_1^{(2)}, C_2^{(2)}, \dots\}$, we consider every pair of constituent protoclusters $C_i^{(1)}, C_j^{(2)}$. We first transform them to local patch coordinates centered on their average sky position. We also rotate the coordinate frame so that their line stars are aligned along the x direction. Finally, we define the overlap region of $C_i^{(1)}$ and $C_j^{(2)}$ to be the common extent on the x -axis where both protoclusters have stars. See Figure 9 for an explicit example of this setup, for two high-significance protostreams from Jhelum that are merged together by our criteria.

If the overlap region is at least 3° and the line stars of $C_i^{(1)}$ and $C_j^{(2)}$ in the overlap region satisfy $f_{\text{line}} > 0.4$, then we compute the difference in θ and ρ between the line stars, call this $\Delta\theta_{ij}$ and $\Delta\rho_{ij}$. The two protostreams are merged into the same stream candidate, if $\Delta\theta_{ij}$ and $\Delta\rho_{ij}$, averaged over all such pairs of protoclusters, satisfy the requirements $|\langle\Delta\theta_{ij}\rangle| < 9^\circ$ and $|\langle\Delta\rho_{ij}\rangle| < 1^\circ$. (This is the

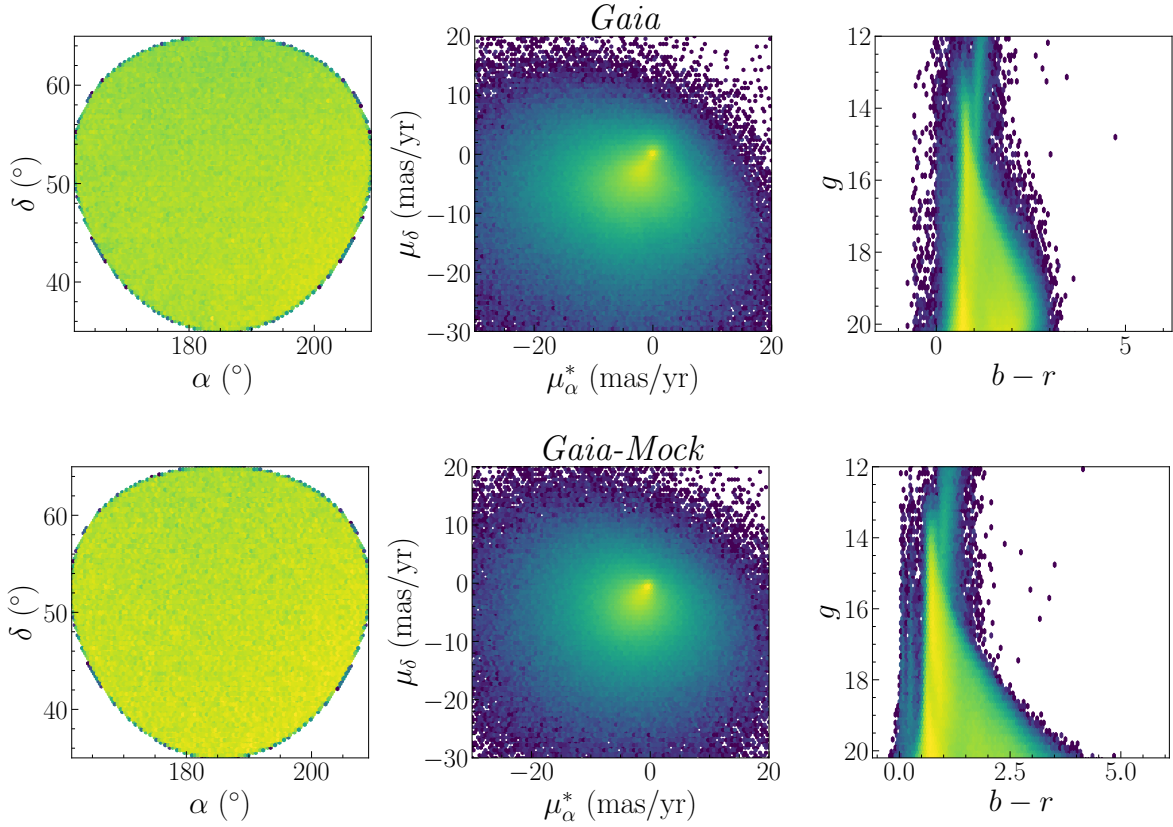


Figure 10. Comparison of the position (left column), proper motion (center column) and color-magnitude (right column) of *Gaia* DR2 (top row) and *Gaia-Mock* (bottom row) patches centered on $(\alpha, \delta) = (185.4^\circ, 50.0^\circ)$. No fiducial cuts have been applied to the data.

same requirement for visual agreement that we used to clean up protostreams, see Eq. (16).)

The stream clustering then proceeds via a simple single-linkage clustering, i.e. stream candidates $\mathcal{S}_1 = \{\mathcal{P}_{11}, \mathcal{P}_{12}, \dots\}$ and $\mathcal{S}_2 = \{\mathcal{P}_{21}, \mathcal{P}_{22}, \dots\}$ are clustered together into a single stream candidate if any pair of protostreams in them satisfy the above requirements.

Finally, we define the significance of a stream to be the sum-in-quadrature of its constituent protostream significances, which were given by Eq. (17)

$$\sigma(\mathcal{S})^2 = \sum_{\mathcal{P} \in \mathcal{S}} \sigma(\mathcal{P})^2. \quad (18)$$

The final result of this stream clustering algorithm is to smoothly connect stream fragments in a way that allows for the direction of the stream on the sky and for proper motion values to gradually change as one moves along the stream.

4 GALAXIA ANALYSIS: ESTIMATING THE FALSE POSITIVE RATE

As the preceding sections indicate, the VM2 algorithm is a multi-step process, with a large number of “hyperparameters” selected to target specific classes of cold stellar streams that are spatially narrow. (For example, the narrowness condition is enforced by our choices for the size of the inner region and outer annulus used to identify the most-significant line in the Hough space, see Section 3.2 for details.) Given this complexity, it is important to investigate the behavior of

VM2, in particular with regards to the fpr, i.e., how often does VM2 identify a stellar stream candidate when none exists?

To provide a benchmark for studies of false positives, we turn to the *Gaia* DR2 mock catalog described in (Rybizki et al. 2018).⁶ This set of mock observations was generated using the GALAXIA code (Sharma et al. 2011), and we will refer to it as *Gaia-Mock* throughout. Simulated stars within the *Gaia-Mock* are drawn from a semi-analytic Besançon model (Robin et al. 2003) of the Milky Way, a combination of smooth distributions of stars representing the Milky Way’s thin and thick disks, halo, and bulge. The stellar population models are tuned to approximate the known properties of Galaxy’s disks, bulge, and halo components. The mock observations of these stars are passed through a 3D dust extinction model based on the Milky Way (Bovy et al. 2016), and smeared by measurement errors derived from the nominal *Gaia* DR2 error model (de Bruijne et al. 2005).

In Figure 10, we show the distribution of stars in a randomly-selected example patch (centered on $\alpha = 185.4^\circ, \delta = 50.0^\circ$) from *Gaia* DR2 (top) and the same patch in *Gaia-Mock*. There are 896,912 stars in the *Gaia* patch and 667,426 in the *Gaia-Mock* equivalent. While differences in the color-magnitude diagrams are visible by eye, the position and proper motion plots are broadly similar.

The smoothness of the underlying model ensures that *Gaia-Mock* does not have any substructure on the scale of stellar streams. Any

⁶ The simulated data is available through the German Astrophysical Virtual Observatory website <https://dc.zah.uni-heidelberg.de/browse/gdr2mock/q>.

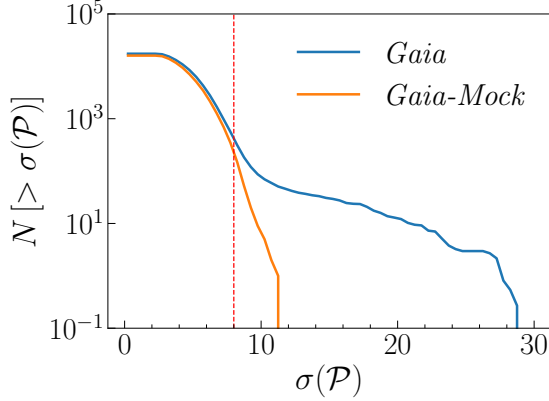


Figure 11. The cumulative number of protoclusters with significance greater than the threshold indicated on the x -axis, in *Gaia* and *Gaia-Mock*. We see that they start to diverge strongly after $\sigma(\mathcal{P}) \sim 8$, which is indicated by the vertical red dashed line.

streams detected by VM2 in the *Gaia-Mock* catalog are therefore false positives. Assuming that the *Gaia-Mock* is modeling the larger-scale features of the *Gaia* data sufficiently well, we should expect similar numbers of spurious stream candidates created by dust occlusion, motion of disk stars, and large-scale correlations of stellar motion within a patch in both *Gaia* and *Gaia-Mock*. We therefore propose to use the *Gaia-Mock* to estimate the fpr of stream detections in *Gaia*.⁷

To construct the false positive streams in *Gaia-Mock*, we ran the entire VM2 algorithm on a quarter of the *Gaia-Mock* sky, as indicated by the red dots in Figure 2. (Computational limitations prevented us from running VM2 on a larger portion of the *Gaia-Mock*.) Other than sky coverage, the VM2 analysis steps for the *Gaia-Mock* data are identical in all respects to those for *Gaia*. By comparing the number of stream candidates above a given significance in *Gaia-Mock* and *Gaia* (properly rescaling for the difference in sky coverage), we obtain an estimate of the fpr in *Gaia*.

4.1 Initial comparisons – no additional cuts

Initially, there are 65,220 *Gaia-Mock* ROIs within the 44 patches, which should be compared with the 256,557 *Gaia* ROIs across all 163 patches. After rescaling the former by 163/44, these values agree to within $\approx 5\%$, showing that *Gaia-Mock* is an excellent match for the kinematic properties of the real sky.⁸ In the following, we will quote the counts of ROIs, protoclusters, protostreams and streams from *Gaia-Mock*, rescaled by 163/44 to match the number of *Gaia* patches.

⁷ Synthetic *Gaia* catalogs based on state-of-the-art N -body+hydrodynamical simulations also exist, see e.g. Ananke (Sanderson et al. 2020), which is based on FIRE (Hopkins 2015; Wetzel et al. 2016; Hopkins et al. 2018) and Auriga (Grand et al. 2018) based on Auriga (Grand et al. 2017). However, the way these mock catalogs “upsampled” the initial simulated star particles into synthetic stars led to residual correlations that caused VM2 to fail already at the ANODE overdensity-finding step. We note that normalizing flows can generate upsampled populations without these unphysical clumping effects (Lim et al. 2022), which may allow future stream-finding studies to test methodology using fully cosmological simulations of galaxies.

⁸ The numbers are not exactly the same, because (as in Paper I), ROIs are required to have at least 200 stars in order to be considered in our analysis. So this cut has slightly different efficiency in *Gaia* and *Gaia-Mock*.

Given that some of the *Gaia* ROIs are deleted due to GC’s, we choose to delete the exact same ROIs from *Gaia-Mock* for an even comparison (even though *Gaia-Mock* does not contain any GCs). This is crucial for an accurate estimate of the fpr. After the GC removal, there are 195,344 valid *Gaia-Mock* ROIs rescaled to 163 patches, compared to 213,874 *Gaia* ROIs. The numbers of ROIs across both data sets are within 10% of one another, indicating continued good agreement between *Gaia* and *Gaia-Mock*.

After the protoclustering step, there are 59,373 rescaled number of protoclusters in the *Gaia-Mock* patches, compared to 64,495 protoclusters in all the *Gaia* patches. These are again within 10% of one another.

Shown in Figure 11 are the number of protoclusters vs. protocluster significances for *Gaia* and *Gaia-Mock*. We see that the bulk of the *Gaia* distributions are reproduced well by the *Gaia-Mock* scan. However, above $\sigma_{\text{protocluster}} \sim 8$, the *Gaia* and *Gaia-Mock* distributions appear to diverge in Figure 11. As discussed above in Section 3.3, protoclusters with $\sigma_{\text{protocluster}} < 8$ are not visible by eye, and so – following our philosophy that each protocluster should be a strong detection of a stream fragment – we cut out such protoclusters. It is reassuring that our comparison with *Gaia-Mock* confirms that these protoclusters are all likely to be false positives, while above this threshold the proportion of false positives starts to decrease rapidly.

After keeping only protoclusters with $\sigma_{\text{protocluster}} > 8$ and merging duplicate protoclusters, we find 848 rescaled number of protostreams in *Gaia-Mock*. This should be compared to the 1,063 protostreams found in *Gaia*. While this indicates there could be around 200 protostreams in *Gaia* corresponding to real streams, our goal was to bring down the fpr. In the next subsection, we will propose one additional cut that will greatly reduce this fpr and leave us with a much more robust sample of candidate streams in *Gaia*.

4.2 Cutting on fraction of dim stars

Given the fpr for protostreams described in the previous subsection, we are motivated to develop additional criteria that we can impose on the protostreams that would improve the fpr. It is important that any cut we devise does not rely on mismodeling differences between *Gaia-Mock* and the real Milky Way galaxy; i.e., the variables used should be in good agreement between *Gaia-Mock* and *Gaia*.

From inspection, we observe that many of the *Gaia-Mock* protostreams are built from stars towards the brighter end of the magnitude range. This is illustrated in Figure 12, which shows the color/magnitude of the stars in the two highest-significance *Gaia-Mock* protostreams. Such protostreams also occur amongst the *Gaia* sample; shown in the third panel of Figure 12 is a representative *Gaia* protostream with a similar predominance of bright stars.

To better identify such protostreams that are mostly composed of brighter stars, we introduce the quantity f_{dim} , defined to be the fraction of stars in each protostream with magnitude $g > 18.4$. In Figure 13, we show a histogram of the f_{dim} distribution for *Gaia* and *Gaia-Mock* protostreams (the latter reweighted to 163 patches, as above). Both *Gaia* and *Gaia-Mock* have a similar peak around $f_{\text{dim}} \sim 0.2$, indicating that the protostreams dominated by bright stars are a feature of *Gaia* which is fully reproduced by the *Gaia-Mock*. This strongly suggests that protostreams with large fractions of bright stars (low f_{dim}) are more likely to be false positives.⁹ Meanwhile above $f_{\text{dim}} \sim 0.5$, the number of protostreams in *Gaia-Mock* is

⁹ Whether this is due to particular choices of hyperparameters of our stream-finding algorithm or from ANODE itself is unclear. It is possible that these

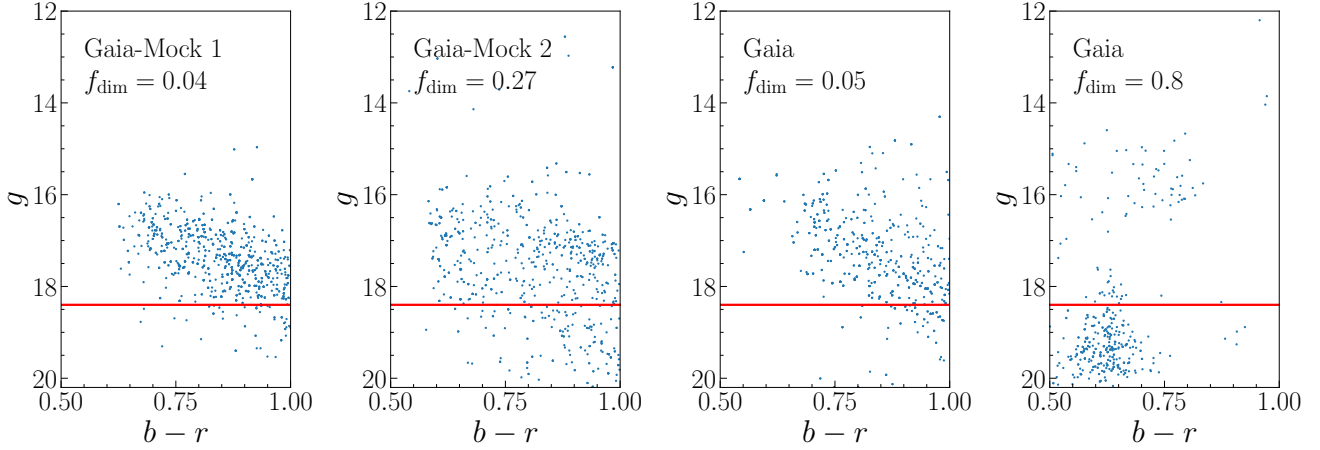


Figure 12. From left to right: the color and magnitude of the stars in the two highest-significance *Galaxia* protostreams, in a representative *Gaia* protostream with low f_{dim} , and in a representative *Gaia* protostream with high f_{dim} . We see that the first three all have a predominance of bright stars, delineated by the red line at magnitude 18.4.

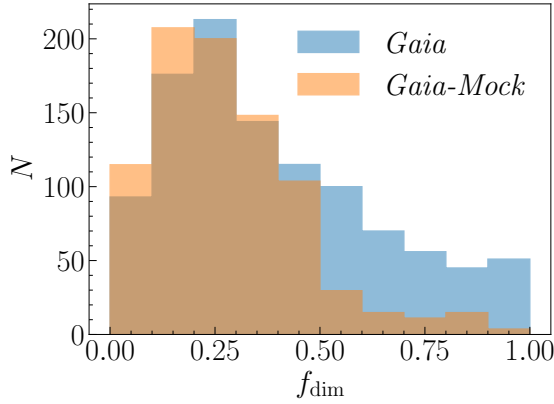


Figure 13. The fraction f_{dim} of stars in each protostream with magnitude $g > 18.4$. Both *Gaia* and *Gaia-Mock* have a similar peak around $f_{\text{dim}} \sim 0.2$, but above $f_{\text{dim}} \sim 0.5$ the number of protostreams (all of which are false positives) in *Gaia-Mock* is significantly reduced.

significantly reduced, but a sizeable number of *Gaia* protostreams remain.

To reduce the rate of false positives, we will include a quality cut of

$$f_{\text{dim}} > 0.5 \quad (19)$$

on the protostreams. After imposing this cut, there are 317 *Gaia* protostreams but only 70 rescaled number of *Gaia-Mock* protostreams. In the rightmost panel of Figure 12 is shown a representative *Gaia* protostream that survives Eq. (19). With these protostreams in hand, we are ready to compare the final step of our algorithm – stream clustering – within *Gaia* and *Gaia-Mock*.

are all thick-disk stars with poorly-measured parallaxes, such that they were not removed by the fiducial cut described in Section 2.3.

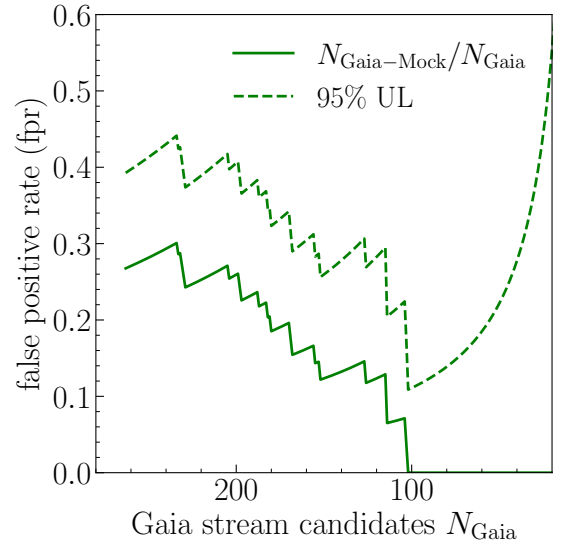


Figure 14. The fpr plot for complete stream candidates (i.e. after the final stream clustering step). Solid and dashed denote the nominal fpr and the 95% UL on the fpr assuming Poisson statistics, respectively.

4.3 Streams in *Gaia-Mock*

After clustering the *Gaia-Mock* protostreams with the $f_{\text{dim}} > 0.5$ cut in place, we find 70 rescaled number of *Gaia-Mock* streams, with max significance $\sigma_{\text{stream}} = 8.86$. Similar clustering on the *Gaia* data results in 262 streams over the full sky. Placing the same cut on significance for *Gaia* and *Gaia-Mock* streams, we obtain an estimate of the fpr for the streams above that significance.

The fpr vs. number of *Gaia* streams above a given significance is shown in Figure 14. Given the reduced statistics of the *Gaia-Mock*, we take a conservative approach and adopt the 95% upper limit (UL) on the fpr, derived from Poisson statistics on the actual (integer) number of *Gaia-Mock* streams. The 95% UL on the fpr is minimized at 11% with a requirement that $\sigma_{\text{stream}} > 8.86$ – when considering the 95% UL, it is not beneficial to cut harder on significance after the

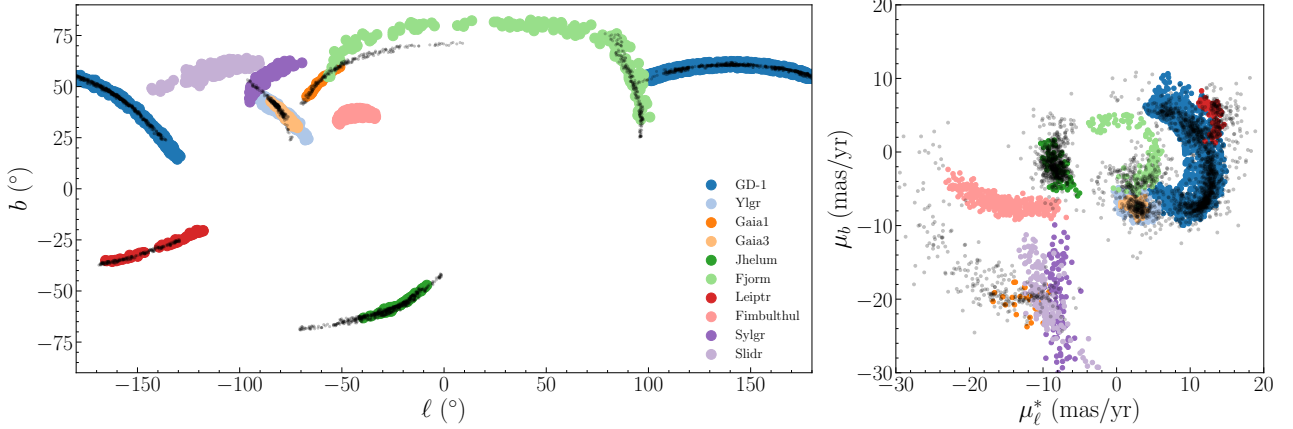


Figure 15. Streams discovered or re-discovered in *Gaia* DR2 data by STREAMFINDER, with position and proper motions digitized from STREAMFINDER papers (Malhan & Ibata 2018; Malhan et al. 2018a). (GD-1 coordinates from Price-Whelan & Bonaca (2018).) Overlaid in black are the VIA MACHINAE stream candidates corresponding to a subset of these known streams. Fimbulthul, Sylgr, and Slidr are not identified in VIA MACHINAE, see text for details. Left: positions in Galactic coordinates. Right: proper motions in Galactic coordinates.

last *Gaia*-Mock stream. Applying this threshold to the real Milky Way data results in 102 *Gaia* stream candidates in the full-sky scan (163 patches), of which we expect at least ~ 90 of these to correspond to real streams, assuming the *Gaia*-Mock-derived fpr holds in the *Gaia* data.

5 STREAMS IN GAIA DR2

Having built and tested VM2, and quantified its fpr, we now proceed to apply it to *Gaia* DR2. In Section 5.1, we focus on VM2’s re-identification of the well-known stream GD-1. We treat this stream separately as it has been identified in numerous other analyses, as well as being the focus of the first implementation of VIA MACHINAE in Paper I. Section 5.2 contains a brief discussion of VM2 stream candidates that may correspond to fragments of the Sagittarius stream. The present work, being optimized for narrow streams, is not expected to recover all of the Sagittarius stream, given its rather large width. Still, Sagittarius is a very distinctive overdensity in the sky and so it is not surprising that VM2 picks up some fragments of it. In Section 5.3, we consider the stellar streams that have been identified by STREAMFINDER, another automated method for finding stellar streams within *Gaia* data that has also been applied over the whole sky (Malhan & Ibata 2018; Malhan et al. 2018a). Both the STREAMFINDER stellar streams that have and have not been re-identified in VM2 are of interest, as the intersection and complementarity of the two methods can better our understanding of the methods and their limitations. Finally, the newly found stream candidates that do not have a previously discovered analog in the literature are briefly presented in Section 5.4. Given that this is still primarily a methodology paper, we leave a more detailed study of the new stream candidates and their broader astrophysical implications to future work.

5.1 GD-1

The GD-1 stellar stream served as the test-bed for the first iteration of the VIA MACHINAE algorithm, due to its extent, stellar density, and existing catalogs of likely stellar members. Originally discovered in

SDSS data, the stellar membership list has been refined and extended using *Gaia* data. For a study of the member GD-1 stars produced by VM2, we use as a benchmark Price-Whelan & Bonaca (2018), which identifies 1,985 stars as likely members of GD-1, based on proper motion, color, magnitude, and position within a defined footprint. In Shih et al. (2022), the first iteration of VIA MACHINAE identified 1,688 likely members of GD-1, of which 738 were likewise identified as stream members by Price-Whelan & Bonaca (2018).

Due to its distinctiveness in all its kinematic, spatial, and photometric features, GD-1 is the highest-significance stream candidate in this updated VM2 analysis, with $\sigma_{\text{stream}} = 83$. We identify 1,252 stars as likely members using VM2, of which 820 (65%) are likewise identified as stream members by Price-Whelan & Bonaca (2018). As shown in Fig. 15, the reconstructed stream overlays the known path of GD-1 in both angular position and proper motion space.

In Figure 16, we show the comparison between the VM2 stars and the existing GD-1 catalog (Price-Whelan & Bonaca 2018) in the stream-aligned coordinate system of Koposov et al. (2010).¹⁰ As can be seen, the VM2 algorithm reproduces well-known features of GD-1, including the progenitor, the spur, and the two known gaps in the stream. Note that below $\phi_1 \sim -80^\circ$ (indicated by the gray shaded region in Fig. 16), GD-1 is located near the Galactic disk, and as a result the patches containing this part of the stream were not part of the VM2 analysis.

Compared to the analysis of GD-1 from Paper I, the new algorithm no longer identifies a secondary stream candidate perpendicular to GD-1. This is a result of our stricter criteria for combining high-significance ROIs into protoclusters and then into protostreams. Our algorithm now identifies more of the stream than in Paper I: in particular, we now capture a section between $\phi_1 \sim [-10^\circ, 10^\circ]$, which was missed in our original algorithm due to the single proper motion analysis (μ_λ used in the ANODE step was too close to zero). As described in Section 2.1, in VM2, both proper motion coordinates are used in separate ANODE trainings, allowing for this set of stream

¹⁰ The location of the GD-1 stars in proper motion and color/magnitude are shown in Figures 15 and Figures 18, respectively.

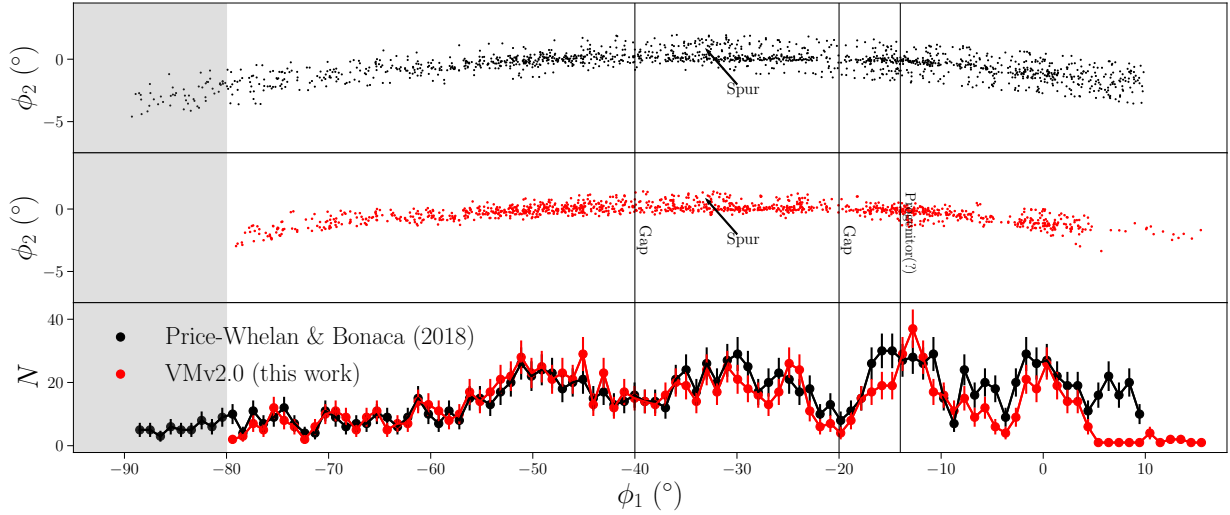


Figure 16. Top: positions of stars identified as likely members of GD-1 by Price-Whelan & Bonaca (2018), using cuts in position, proper motion, and color-magnitude space. Middle: positions of stars identified as likely members of GD-1 by the VM2 algorithm. Bottom: histogram of both catalogs of GD-1 stars in 1° -wide bins along the stream coordinate ϕ_1 . Stream coordinate system $\phi_1 - \phi_2$ defined in Koposov et al. (2010). The region below $\phi_1 \approx -80$ is shaded gray to indicate that it was not included in the 163 patches of the VM2 scan, due to being too close to the Galactic disk.

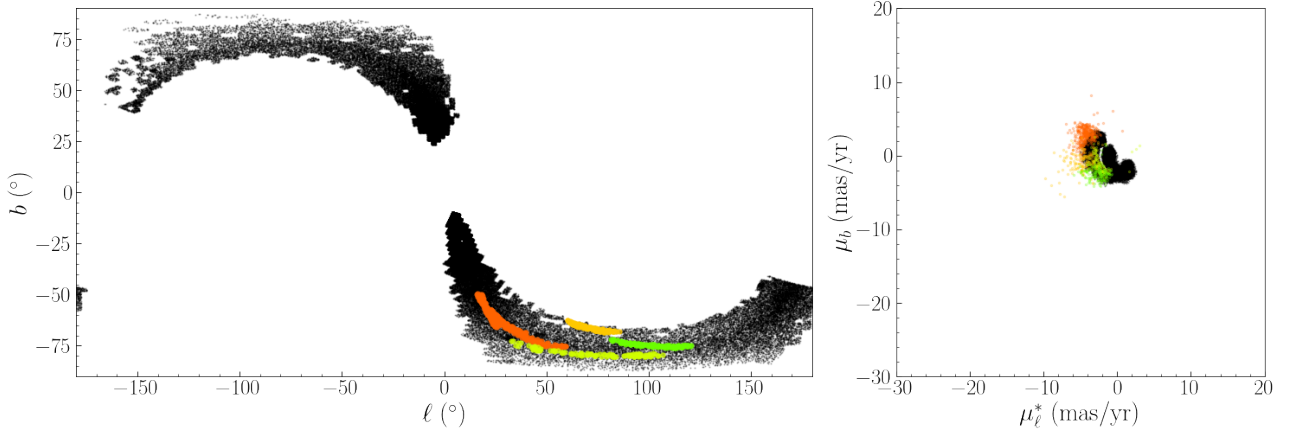


Figure 17. The position (left) and proper motion (right) of *Gaia* DR2 Sagittarius stream stars (black points) overlaid with the four stream candidates from VM2 which are possible members of Sagittarius. Sagittarius stream stars obtained from Antoja et al. (2020).

stars to be tagged as anomalous using the μ_{ϕ^*} search regions. Despite this improvement, VM2 continues to miss stars in GD-1 around $\phi_1 \sim 15^\circ$, likely due to both proper motions of these stars being close to zero.

Overall, we re-identify a higher percentage of likely stream members compared to Paper I. The fewer number of stars in our stream compared to Paper I (along with the tighter clustering around the stream-track) suggests that the new algorithm is successful in combining fragments of the stream cohesively, and rejecting high-significance ROIs that are incompatible with the stream’s path across the sky.

5.2 The Sagittarius Stream

Based on the stellar membership, at least four of the 102 stream candidates identified by VM2 may be subcomponents of the Sagittarius stream (Newberg et al. 2002; Majewski et al. 2003).

To identify potential components of Sagittarius, we compare the stars within all our stream candidates with the 294,344 *Gaia* DR2 stars identified as belonging to the Sagittarius stream in Antoja et al. (2020). Nine stream candidates have at least one star in the Sagittarius catalog. However five of these appear to not overlap significantly in proper motion space with Sagittarius. We show the four likely subcomponents of Sagittarius in Figure 17.

Sagittarius itself is far too wide to be reconstructed by the post-ANODE stream-finding steps. However, the member stars of Sagittarius do often have high- R value, and it is therefore not surprising that some subcomponents are identified as stream candidates by the V_{IA}

Stream Name	References	N_{SF}	N_{VM2}	Status
Gaia1	M2018b	31	308	Found
Gaia2	M2018b	-	-	SRs excluded (GCs)
Gaia3	M2018b	57	276	Same as Ylgr
Gaia4	M2018b	8	-	Not found (too few stars)
Gaia5	M2018b	37	-	Not found (too few stars)
Indus	M2018b, S2019	150	-	Not found (too dim)
Jhelum	M2018b, S2019	63	665	Found
Orphan	M2018b, F2019	-	-	SRs excluded (pms too small)
Gjoll	I2019	-	-	SRs excluded (too few stars)
Fjorm	I2019	148	219	Found
Leiptr	I2019	67	233	Found
Svol	I2019	45	-	Not found (too few stars)
Fimbulthul	I2019	309	-	Found by ANODE but too wide
Ylgr	I2019	349	276	Found
Sylgr	I2019	103	-	Found by ANODE but too wide
Slidr	I2019	156	-	Found by ANODE but too wide
Phlegethon	I2018	-	-	SRs excluded (too few stars)
GD1	GD2006, PWB2018	1985	1252	Found

Table 1. Streams discovered or re-discovered in *Gaia* DR2 using the STREAMFINDER algorithm. The references in the second column correspond to the following: M2018b (Malhan et al. 2018b), S2019 (Shipp et al. 2019), F2019 (Fardal et al. 2019), I2019 (Ibata et al. 2019), I2018 (Ibata et al. 2018), GD2006 (Grillmair & Dionatos 2006), PWB2018 (Price-Whelan & Bonaca 2018). The columns N_{SF} and N_{VM2} correspond to the number of stars found by STREAMFINDER and VM2 respectively.

MACHINAE algorithm. It is perhaps notable that these possible Sagittarius stream segments mostly have significances between $\sim 9 - 10$ (with one having $\sigma_{\text{stream}} = 16$). These are at the lower-end of the stream significances we consider in this work (recall the cutoff from the Galaxia fpr study was $\sigma_{\text{stream}} > 8.86$). The fact that these less-significant stream candidates still correspond to real objects gives us more confidence that the other 102 streams above the cutoff are also real (and not only the highest-significance ones).

5.3 Comparison with STREAMFINDER

Next, we turn to an evaluation of the performance of VM2 on streams previously identified by the STREAMFINDER algorithm. STREAMFINDER is the only other automated search for stellar streams using the *Gaia* data set, and a comparison enables us to better understand the strengths and limitations of VM2. Unlike VM2, however, STREAMFINDER assumes a Milky Way potential and searches for stellar streams within defined orbital cones, with photometry following an isochrone. VM2 remains agnostic to such choices.

In STREAMFINDER survey I, Malhan et al. (2018b) discovered five new stream candidates – named Gaia-1 through Gaia-5 – and demonstrated that their method could find four previously-discovered streams (GD-1, originally found in Grillmair & Dionatos (2006), Indus and Jhelum discovered in Shipp et al. (2018), and the Orphan stream found in Belokurov et al. (2007)). The second sur-

vey, STREAMFINDER II (Ibata et al. 2019), discovered eight new stream candidates (Gjoll, Fjorm, Leiptr, Svol, Fimbulthul, Ylgr, Sylgr, Slidr). Finally, Ibata et al. (2018) describes a new stream candidate, Phlegethon.¹¹

First, we point out that Gaia-3 and Ylgr appear to be the same stream based on their positions and proper motions, which we show in Figure 15. A detailed study of the chemical abundances of members of both streams would be required to confirm this identification. Counting these as the same stream yields a total of 17 streams and stream candidates that were discovered or confirmed with STREAMFINDER in *Gaia* DR2. These stream candidates, their proper motion ranges, and their status in VM2 are summarized in Table 1. Having already discussed GD-1 in the previous subsection, here we focus on the rest.

Three of the streams, Gjoll, Phlegethon, and Orphan, were excluded from our search because their SRs contained too few stars (Gjoll, Phlegethon) or occurred at proper motions close to zero (Orphan). Additionally, Gaia-2 was excluded from our search because all of its ROIs contained a GC candidate (see the selection requirements for VM2 in Section 2).

¹¹ In this work, we restrict our comparison to the STREAMFINDER surveys that used *Gaia* DR2, for a fair comparison. STREAMFINDER did subsequently apply their method to (e)DR3 (Ibata et al. 2020), but we save a comparison to these results for forthcoming work.

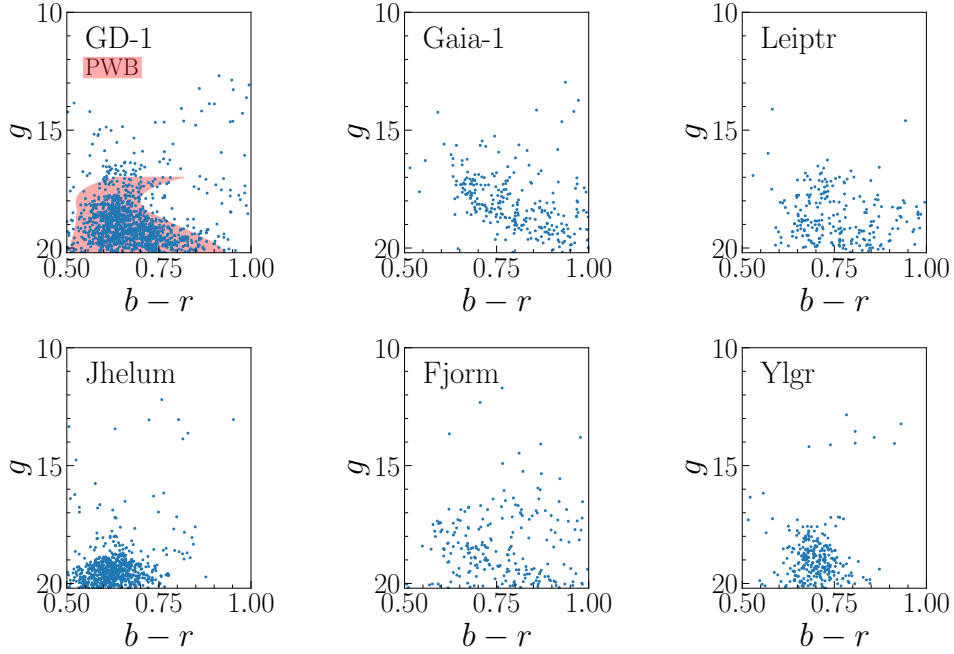


Figure 18. Stars in GD-1 and the five streams re-identified by VM2, in color $b - r$ versus magnitude g space. The shaded red region shown for GD-1 indicates a (smoothed) envelope enclosing the subset of *Gaia* stars in our GD-1 catalogue which were identified as likely stream members by Price-Whelan & Bonaca (2018).

Based on the cuts that we use in VM2 (and excluding the previously-discussed GD-1), we are left with 12 previously known stream candidates that fall into regions of the sky that were analyzed by our algorithm. Of these, we rediscover five using VM2: Gaia-1, Gaia-3/Ylgr, Jhelum, Fjorm and Leiptr. Our method did not identify seven candidates: Gaia-4, Gaia-5, Indus, Svol, Fimbulthul, Sylgr, and Slidr. We now discuss both sets of streams in turn, beginning with the streams that are found in *Gaia* data by both STREAMFINDER and VIA MACHINAE.

A comparison of the positions and proper motions as identified by STREAMFINDER and VIA MACHINAE is provided in Figure 15. The magnitudes and colors of the stars which VM2 identifies as members of these known streams are shown in Figure 18. For GD-1, we also show (with a shaded red region) the range of *Gaia* DR2 stellar colors and magnitudes for the likely stream members identified by Price-Whelan & Bonaca (2018). This provides an estimate for the dispersion of color and magnitudes that might be observed for stellar streams in *Gaia*. It indicates that *Gaia* DR2 photometry is insufficiently precise to resolve a clean isochrone in GD-1, and thus we should probably not expect to see one in the other VM2 stream candidates as well.

In Figure 19, we plot the binned number of stars as a function of the stream coordinate ϕ_1 ,¹² comparing our results to those from STREAMFINDER. From this figure, we see that some of our re-identified streams appear to be significantly extended by VIA MACHINAE in various directions. However, we note that these extensions are sensitive to the hyperparameters of VIA MACHINAE, in particular the accept-

able angle between protostreams at the combination step. Thus more robust confirmation of these potential stream extension awaits further study. In future work, we will study the orbits of these extensions as well as the chemical abundances of these members to check whether or not these stars are part of the streams.

From Figure 19, we see that VM2 does not identify the full extent of Leiptr or Fjorm as found by STREAMFINDER. For Leiptr we have traced this back to the fact that the missing segment of the stream is fully contained in a patch that was deemed too close to the disk and was cut out of our VM2 scan (see Section 2.1 for details). For Fjorm, we found that the ANODE step of our algorithm actually correctly identified the remainder of Fjorm as anomalous stars, but the missing parts are too broad for the narrow line-finder settings used in this work (which are applied in the Hough coordinates after the ANODE training is completed). Repeating the analysis for a wider stream setting will be an interesting direction to explore in future work, potentially highlighting a new set of wider stellar streams.

We also note from Figure 19 that VM2 tends to find a much higher density of stars than STREAMFINDER where the two methods do overlap. The increase in membership stars will enable more spectroscopic follow-ups, which will improve the characterization of these streams.

Finally, with Ylgr, there seems to be a mismatch in the direction of the stream between VIA MACHINAE and STREAMFINDER—overall, STREAMFINDER’s Ylgr seems to be at an angle relative to VIA MACHINAE’s, more clearly seen in Fig. 15. This will be interesting to study further and to check whether the two streams are indeed Ylgr, or they are independent structures that happen to overlap.

Next we turn to the stream candidates found by STREAMFINDER that were not re-identified by VM2. For these streams, we define two categories: The first category – Indus, Fimbulthul, Sylgr, and Slidr – have a large number of stars (150, 309, 103, and 156, respectively), while the second category – consisting of Gaia-4, Gaia-5, and Svol –

¹² The stream-aligned coordinates are obtained by rotating the angular positions of the stars to new coordinate system (ϕ_1, ϕ_2) , which minimizes the sum of $|\phi_2|$ over all the VM2 stars. The rotation step uses the GALA Python package (Price-Whelan 2017; Price-Whelan et al. 2020).

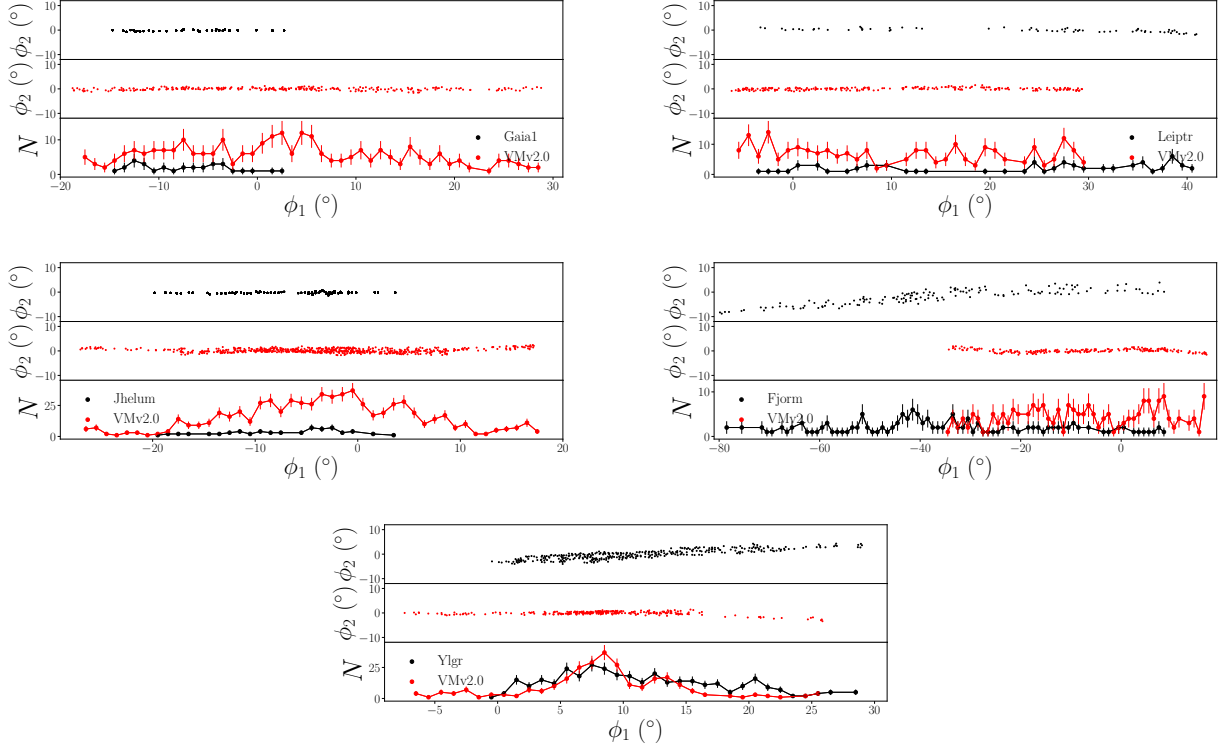


Figure 19. Known stellar streams (black, top row) and VM2-identified stellar stream candidates (red, middle row) in approximate stream-aligned coordinates (ϕ_1, ϕ_2) (stream-aligned coordinates are defined separately for each stream). Star counts in 1° -wide bins of ϕ_1 are shown in the bottom row.

have relatively few stars per the STREAMFINDER membership catalog (8, 37, and 45, respectively).

Investigating the first category, we discover that actually Fimbulthul, Sylgr and Slidr are tagged at the ANODE step, i.e., the 100 highest R stars in their respective ROIs do coincide with the position and proper motions of these STREAMFINDER streams. However, these streams, like the previously-mentioned segment of Fjorm, are again too wide in position space, so they fail to be picked up by the post-ANODE line-finder on the narrow stream setting. This will be explored in more detail in a future work.

For the second category of missing streams, we find no trace of Gaia-4, Gaia-5 and Svöl in any step of VM2, not even after running ANODE. Given the low number of stars within each stream, it is likely these are simply below the detection threshold of ANODE.

Finally, we also find no trace of Indus, despite its robust star count and relative narrowness. It is possible that this failure mode is not attributable to any single property of the stream. However, we note that Indus is composed primarily of dim stars (see Figure 4 of Malhan et al. (2018b)), near the magnitude limit we adopt for *Gaia* ($g = 20.2$). It is possible that the increased number of background stars in this region of phase space is the reason ANODE did not flag the Indus stars as anomalous.

It is notable that the six streams (GD-1, Gaia-1, Leiptr, Jhelum, Fjorm, Gaia-3/Ylgr) identified by both STREAMFINDER and VM2 are also the highest significance stream candidates according to VM2. (Their significances are $\sigma_{\text{stream}} = 83.0, 46.9, 29.2, 29.1, 22.6$ and 22.5 , respectively.) Given the differences in the methodologies of STREAMFINDER and VIA MACHINAE and their agreement on the highest significance streams, we conjecture that these six streams are the

most robust thin cold streams in the volume of *Gaia* data analyzed in this work.

5.4 New stream candidates

Finally, we turn to stellar stream candidates that have not been previously identified (referencing against the GALSTREAMS database (Manteu 2022)). There are a total of 102 VM2 stream candidates by VM2 with significance $\sigma_{\text{stream}} \geq 8.86$, the cutoff we adopt based on our comparison with *Gaia-Mock* (see Sec 4). This number includes GD-1, the five other previously identified streams, and the four possible Sagittarius stream components.

As to why the majority of our stream candidates have not been previously discovered by STREAMFINDER, we can offer a few possible explanations:

- One possibility is the fpr is being underestimated due to some systematic mis-modeling by the *Gaia-Mock*. We deem this unlikely, given how well the *Gaia-Mock* seems to be reproducing the bulk properties of the ROIs and protoclusters found by *Gaia* DR2. Nevertheless, it would be important to perform more cross-checks and estimates of the fpr. For some ideas on how to do this in the future, see Section 6.
- A second possibility is that these stream candidates tend to be shorter on average than the typical STREAMFINDER stream, which may impact their detectability using the STREAMFINDER algorithm. This is indicated in Figure 20, where we plot the density of the stream candidates as a function of their length, and single out the streams found by STREAMFINDER in red.
- Also, STREAMFINDER only reported on a subset of their stream

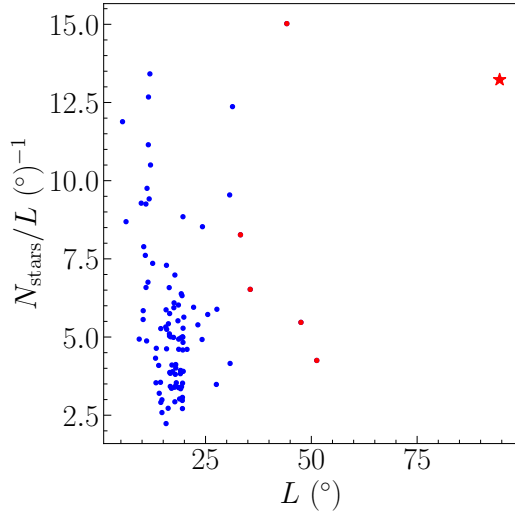


Figure 20. Length L of stellar stream candidates as identified by VM2 versus number of VIA MACHINAE-identified member stars per unit length N_{stars}/L for all 102 VM2 stream candidates. Stream candidates that correspond to streams previously identified by STREAMFINDER are shown in red. GD-1 is denoted with a red star.

candidates. It is possible that our new stream candidates were also found by STREAMFINDER, but just were not made public.

- A final possible explanation is that the Galactic potential modeling assumptions of STREAMFINDER are being violated for these streams. We expect this behavior to be affecting mostly the southern hemisphere streams, due to the proximity to the Large Magellanic Cloud LMC (see e.g. Shipp et al. (2021); Koposov et al. (2022); Lilleengen et al. (2023) for studies on the effect of the LMC on the Milky Way potential). To this point, we note that STREAMFINDER has a clear bias towards the Northern Galactic hemisphere, which could be related to the influence of the LMC/SMC on the Galactic potential model, whereas VM2 is fairly symmetric between Northern and Southern hemispheres.

In Figure 21, we show the locations and proper motions of the 15 highest-significance VIA MACHINAE new stream candidates that do not overlap with Sagittarius (only one possible Sagittarius stream component also has a significance as high as these 15). These top candidates have σ_{stream} values between 11.6 and 19.9, and include all candidates composed of protoclusters drawn from more than one patch on the sky. Only two of these 15 candidates are composed of stars from a single patch (the 9th and 10th most significant stream candidates). The photometric properties of the constituent stars are shown in Figure 22. Stream candidate indexing goes in decreasing stream significance. We chose to highlight these candidates as – due to their length, high σ_{stream} values, and presence in multiple patches – we judge them to be among the most robust candidates in our sample. These, as well as the other stream candidates, require additional cross-matched observations to confirm or reject their existence as stellar streams.

One can see in Figure 21 that three of the 15 highest-significance stream candidates (VM-3, 5, and 9) are overlapping in angular position and proper motion. Upon closer examination (see Fig. 23), we discover that all three streams appear to be built around the same stellar overdensity. Though a stream-like extended structure can be seen extending from this object (most clearly in VM-9), differences in the highest- R stars among each of the streams resulted in three

different stream candidates with different paths across the sky. As a result of these path differences, these three stream candidates were not merged together. Further investigation of the stars around which these three stream candidates are built is necessary; it does not seem to correspond to a known object.

6 CONCLUSIONS

In this work, we have described VM2, an update to the model-agnostic, fully-automated stream finding algorithm VIA MACHINAE. VIA MACHINAE is based on the weakly-supervised anomaly detection protocol ANODE (Nachman & Shih 2020), originally developed for resonant anomaly detection at the Large Hadron Collider (LHC). VIA MACHINAE takes as inputs only the sky positions, proper motions, colors, and magnitudes of stars in the *Gaia* catalog, and searches for stream-like overdensities in this space in a model-agnostic way.

Our work builds upon a previous version of VIA MACHINAE described in Shih et al. (2022), which focused on a more limited demonstration of finding the well-known stream GD-1 in a model-agnostic way. Here, we generalize the application of VIA MACHINAE to the entire *Gaia* DR2 data set. This required a number of improvements to the algorithm. The core step remained the same: training ANODE on *signal regions* and *sideband regions* consisting of windows in a single proper motion coordinate and their complements, in order to derive an “overdensity score” or “anomaly score” for each star in the signal region. However, the subsequent steps of combining multiple detections of stream fragments into a coherent and consistent high-significance stream candidate were redesigned in order to make the algorithm more robust to the different streams. We have also improved on the version of VIA MACHINAE presented in Shih et al. (2022) by incorporating a second ANODE scan over the orthogonal proper motion coordinate, and improving the data quality with a more sensitive globular cluster / dwarf galaxy finding algorithm and a new fiducial cut that excludes foreground contamination from bright stars originating in the thick disk of the Milky Way.

We have validated VIA MACHINAE on a combination of *Gaia* DR2 data and the *Gaia-Mock* simulation, an idealized simulation of the Milky Way that, by design, does not include any substructure. In particular, using *Gaia-Mock*, we estimated the fpr of stream detections in VIA MACHINAE. After showing that this fpr is at an acceptable level ($\sim 10\%$ for 102 detected stream candidates), we applied VIA MACHINAE to a full-sky scan of *Gaia* DR2 data. We showed how VIA MACHINAE could rediscover the well-known GD-1 stream, as well as five other streams previously detected in *Gaia* DR2 by the STREAMFINDER method (Gaia-1, Gaia-3/Ylgr, Fjorm, Leiptr and Jhelum). Finally, we described and characterized the 90+ additional stream candidates that VIA MACHINAE finds, focusing in particular on 15 of the highest-significance new stream candidates.

Although this current paper focused on *Gaia* DR2, a major priority is to re-run it on the *Gaia* Data Release 3 (DR3) data. We expect, with the improved measurements of DR3, all the stream detections will become even more significant and robust, and perhaps discoveries of even more new stream candidates await.

Our work presents many directions for further study. Foremost among these are confirming our new stream candidates with in-depth studies of orbital parameters as well as their chemical abundances, which will help identifying their origins (similar to what Martin et al. 2022, has done for some of the STREAMFINDER streams). If some of these structures are remnants of dwarf galaxy mergers, it will lead to a better understanding of the Milky Way’s stream population (Li et al. 2022), which is crucial for understanding the merger history of

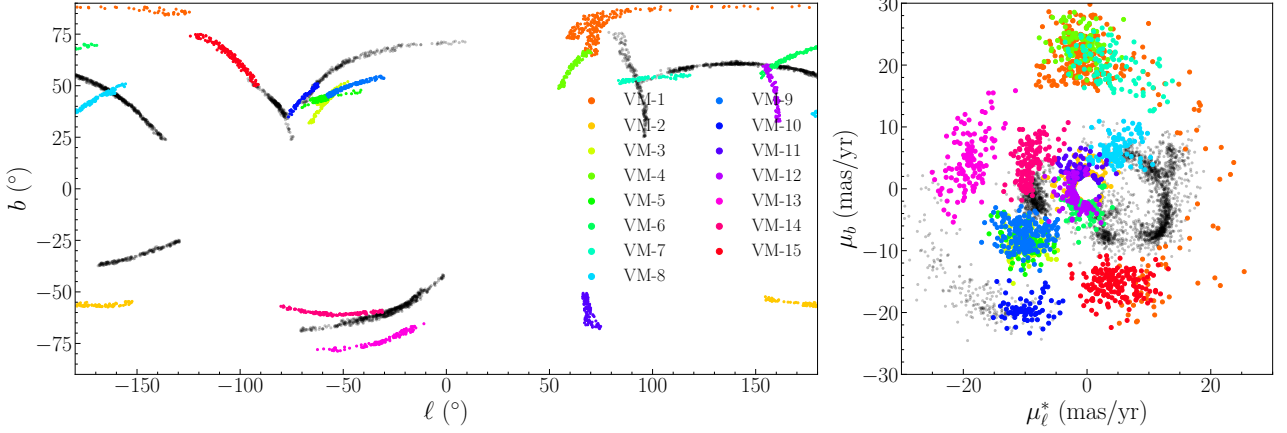


Figure 21. The 15 stream candidates identified by VM2 which are composed of protoclusters from more than one patch in Galactic position (left) and proper motion (right). Stars belonging to the new stream candidates are identified by colored points; previously-identified streams which have been re-identified in VM2 are shown in black (see Sections 5.1 and 5, and Figure 15).

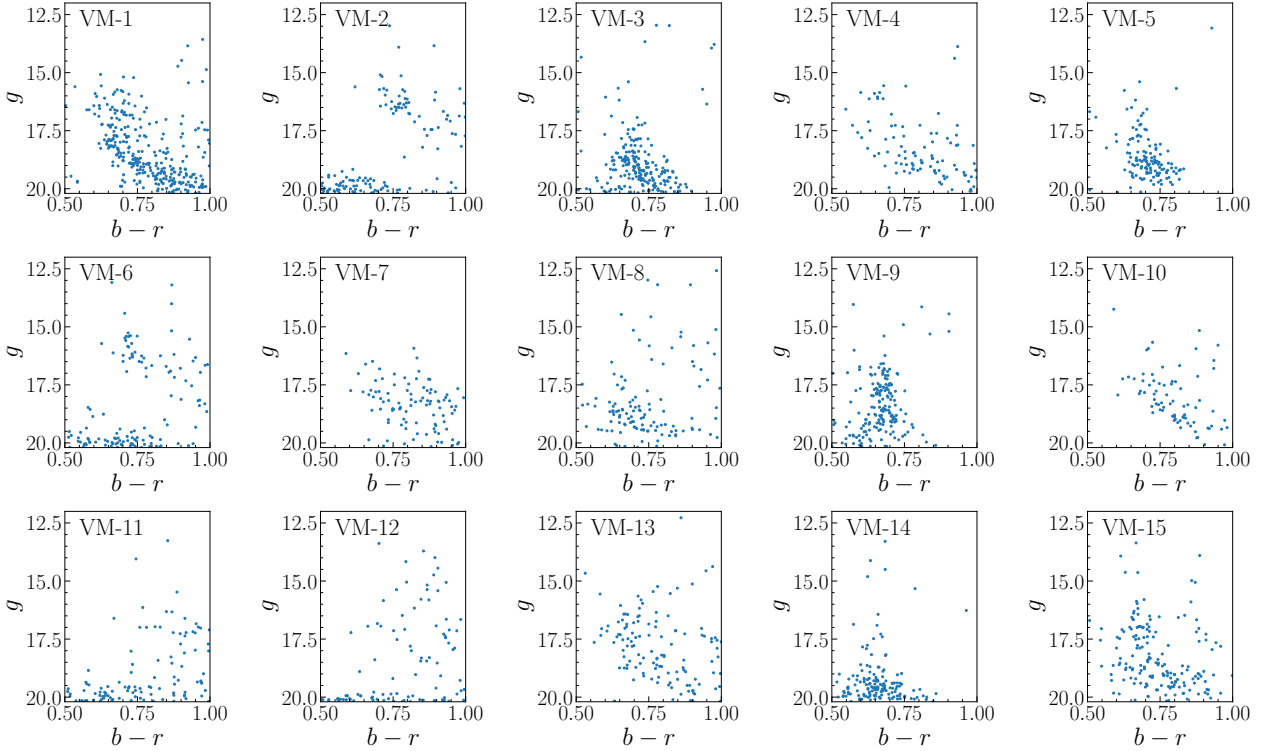


Figure 22. The 15 highest-significance stream candidates identified by VM2 which are not identified with previously known streams, in color $b - r$ versus magnitude g .

the Milky Way, the rates of disruptions, and the feedback models in numerical simulations (as illustrated for example in the comparative work of Shipp et al. 2022).

Besides re-running VM2 on *Gaia* DR3 and performing more in-depth astrophysical studies of our stream candidates, it is important to perform additional checks on the stream candidates and the fpr estimated in this work. While re-running ANODE on the entire DR2 sky multiple times is computationally unfeasible, it should be possi-

ble to re-run on just those patches or SRs that contain the new stream candidates. If the stream candidate is found with repeated ANODE scans, this makes it more likely to be real. Finally, repeating the fpr study with a different simulation would provide a very important cross check. In the future, *Gaia* mock catalogs based on state-of-the-art, kinematically consistent hydrodynamical simulations could become viable candidates for such a study (Lim et al. 2022).

Our findings also suggest that several of the previously discov-

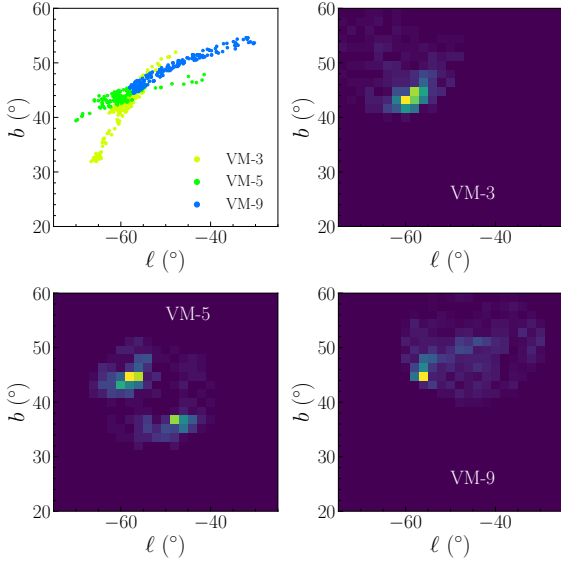


Figure 23. Upper left: the stars in the highest-significance protoclusters of the three stream candidates (VM-3, 5, and 9) which cross in angular position. Clockwise from upper left: two-dimensional histograms of all the stars in all the protoclusters in each stream, clearly showing the shared clump of stars around which each stream candidate is built.

ered streams may be longer than previously known. We will address the extended streams in upcoming work by comparing the chemical abundances of the newly found members with the previously confirmed members. These extensions will be helpful in better studying the orbital properties of these streams, which will then lead to updated measurements of the properties of the Milky Way halo, such as its total mass and shape, as well as the effect of the LMC.

Looking further beyond the present work, potentially one of the most interesting findings are the many stream candidates discovered by our algorithm and not reported by STREAMFINDER. If these stream candidates are indeed real and not false positives (which we expect given our estimated low fpr presented in this paper), this could indicate that the model-dependent methodology of STREAMFINDER is causing it to miss many streams. This would be the ultimate benefit of a model-agnostic technique such as VIA MACHINAE.

Beyond this work, a few generalizations of VIA MACHINAE will reveal more of the interesting merger history of Milky Way. In particular, we want to highlight the wide stream search, where by inspection, some of the streams from the literature are being missed in our analysis due to the settings on the width of the stream. Improving the machine-learned anomaly detection step currently performed by AN-ODE may also reduce the false positive rate and/or be more sensitive to other aspects of the stellar streams. For example, the CATHODE anomaly detection algorithm (Hallin et al. 2022) has been shown to improve anomaly detection within LHC data, and may do the same here. More light-weight anomaly detection methods, such as CWoLa-Hunting (Collins et al. 2018, 2019), could be used to improve refine and improve stream membership post-discovery. Finally, this method will be of upmost importance to future surveys, in particular the Vera Rubin Observatory Legacy Survey of Space and Time (LSST) (LSST Science Collaboration et al. 2009; Ivezić et al. 2019), which will not only improve measurements of proper motions, but will unveil a deeper side of the sky.

ACKNOWLEDGEMENTS

We would like to acknowledge to John Tamanas for collaboration in the early stages of this work. We would also like to thank Anna Hallin, Claudius Krause, Ben Nachman, Mariel Pettee, Nora Shipp and Sowmya Thanvantri for helpful discussions. Finally, we are grateful to Anna Hallin and Claudius Krause for feedback on the draft. MB and DS are supported by the DOE under Award Number DOE-SC0010008. LN is grateful for the generous support and hospitality of the Rutgers NHETC Visitor Program, where this work was initiated.

This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231.

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

DATA AVAILABILITY

This paper made use of the publicly available *Gaia* DR2 data. For more detailed information about the data (e.g. patch center locations) please contact the corresponding author. Membership stars of the discovered streams will be provided following journal acceptance.

APPENDIX A: BINNED TWO-DIMENSIONAL OVERDENSITY DETECTION

In this work, we find it necessary in two contexts to identify local overdensities in two-dimensional coordinates: first when locating GC/DGs in angular coordinate-space, and second in finding line-parameters in the (ρ, θ) Hough space. In both cases, the problem at hand is locating the coordinates of the highest-density of points in a two-dimensional space, over a background which is not at all uniform and cannot be predicted *a priori*. We use a single flexible algorithm for both situations, modifying the hyperparameters to address each of these two cases.

This algorithm operates on a 2d histogram consisting of a set of pixels indexed by i and j , each with a number of counts N_{ij} . We are interested in whether one of the pixels is elevated (overdense) relative to its neighbors. The main “hyperparameter” of this algorithm is how the neighbors are chosen in order to estimate the background level. Given some list $(i_1, j_1), (i_2, j_2), \dots$ of neighbors of the (i, j) pixel, let \mathcal{A}_{ij} be their counts:

$$\mathcal{A}_{ij} = \{N_{i_1 j_1}, N_{i_2 j_2}, \dots\}. \quad (\text{A1})$$

The position-dependent background estimate is then the average of the counts in this neighbor list:

$$N_{\text{bg}, ij} = \bar{\mathcal{A}}_{ij}. \quad (\text{A2})$$

We also define a “systematic error” in our background estimate in terms of the standard deviation of the neighbor counts $\sigma(\mathcal{A}_{ij})$ and the number of pixels in the neighbor list $|\mathcal{A}_{ij}|$:

$$\delta N_{\text{bg}, ij} = \sigma(\mathcal{A}_{ij}) / \sqrt{|\mathcal{A}_{ij}|}. \quad (\text{A3})$$

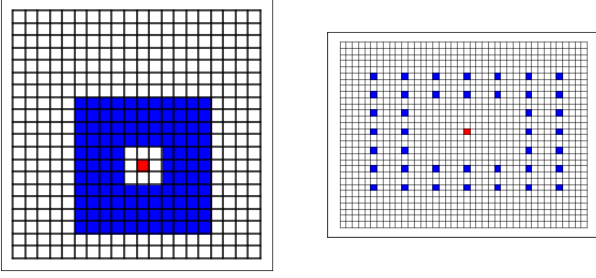


Figure A1. Left: the 11×11 annulus used for estimating the background level in the GC-finding algorithm. Right: the 7×7 annulus with stride 5×3 used for estimating the background level in the line-finding algorithm.

This allows us to define a significance for each pixel i, j :

$$\sigma_{ij} = \frac{N_{ij} - N_{bg,ij}}{\sqrt{N_{bg,ij} + (\delta N_{bg,ij})^2}}. \quad (A4)$$

where the Poisson statistical error and the systematic error are added in quadrature.

For the GC-finding algorithm, the neighbor list is an 11×11 pixel annulus centered at i, j with the central 3×3 pixels removed.

For the line-finding algorithm, the neighbor list is more complicated. Here each pixel (i, j) counts the number of curves in Hough space passing through a box of width $\Delta i = 5$ and $\Delta j = 3$, centered at $\rho = -10 + i/5$ and $\theta = \pi \times j/100$. Thus adjacent pixels correspond to overlapping boxes and are not independent of one another. To derive the background counts with independent boxes, we use a 7×7 annulus of pixels centered at (i, j) , spaced apart by Δi and Δj , with the middle 3×3 pixels removed.

These neighbor lists for the GC-finding and the line-finding are shown in Figure A1.

REFERENCES

- Antoja T., Ramos P., Mateu C., Helmi A., Anders F., Jordi C., Carballo-Bello J. A., 2020, *A&A*, **635**, L3
- Astropy Collaboration et al., 2013, *A&A*, **558**, A33
- Astropy Collaboration et al., 2018, *AJ*, **156**, 123
- Astropy Collaboration et al., 2022, *apj*, **935**, 167
- Banik N., Bovy J., 2019, *MNRAS*, **484**, 2009
- Belokurov V., et al., 2006, *Astrophys.~J.-Lett.*, **642**, L137
- Belokurov V., et al., 2007, *ApJ*, **658**, 337
- Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, *MNRAS*, **478**, 611
- Bonaca A., Hogg D. W., Price-Whelan A. M., Conroy C., 2019a, *Astrophys.~J.*, **880**, 38
- Bonaca A., Conroy C., Price-Whelan A. M., Hogg D. W., 2019b, *ApJ*, **881**, L37
- Bonaca A., et al., 2020, *Astrophys.~J.-Lett.*, **892**, L37
- Borsato N. W., Martell S. L., Simpson J. D., 2020, *MNRAS*, **492**, 1370
- Bovy J., Rix H.-W., Green G. M., Schlafly E. F., Finkbeiner D. P., 2016, *ApJ*, **818**, 130
- Carlberg R. G., Grillmair C. J., Hetherington N., 2012, *Astrophys.~J.*, **760**, 75
- Collins J. H., Howe K., Nachman B., 2018, *Phys. Rev. Lett.*, **121**, 241803
- Collins J. H., Howe K., Nachman B., 2019, *Phys. Rev. D*, **99**, 014038
- Duda R. O., Hart P. E., 1972a, *Commun. ACM*, **15**, 11
- Duda R. O., Hart P. E., 1972b, *Commun. ACM*, **15**, 11
- Erkal D., Koposov S. E., Belokurov V., 2017, *MNRAS*, **470**, 60
- Fardal M. A., van der Marel R. P., Sohn S. T., del Pino Molina A., 2019, *MNRAS*, **486**, 936
- Gaia Collaboration et al., 2016a, *A&A*, **595**, A1
- Gaia Collaboration et al., 2016b, *A&A*, **595**, A2
- Gaia Collaboration et al., 2018, *Astron. & Astrophys.*, **616**, A1
- Gaia Collaboration et al., 2021, *A&A*, **649**, A1
- Gaia Collaboration et al., 2022, arXiv e-prints, p. arXiv:2208.00211
- Górski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, *Astrophys.~J.*, **622**, 759
- Grand R. J. J., et al., 2017, *MNRAS*, **467**, 179
- Grand R. J. J., et al., 2018, *MNRAS*, **481**, 1726
- Grillmair C. J., Dionatos O., 2006, *Astrophys.~J.*, **643**, L17
- Hallin A., et al., 2022, *Phys. Rev. D*, **106**, 055006
- Helmi A., White S. D. M., 1999, *MNRAS*, **307**, 495
- Hopkins P. F., 2015, *MNRAS*, **450**, 53
- Hopkins P. F., et al., 2018, *MNRAS*, **480**, 800
- Hough P. V. C., 1959, Conf. Proc. C, 590914, 554
- Ibata R., Lewis G. F., Irwin M., Totten E., Quinn T., 2001, *Astrophys.~J.*, **551**, 294
- Ibata R. A., Malhan K., Martin N. F., Starkenburg E., 2018, *ApJ*, **865**, 85
- Ibata R. A., Malhan K., Martin N. F., 2019, *Astrophys.~J.*, **872**, 152
- Ibata R., et al., 2020, arXiv e-prints, p. arXiv:2012.05245
- Ivezić Ž., et al., 2019, *ApJ*, **873**, 111
- Johnston K. V., 1998, *Astrophys.~J.*, **495**, 297
- Johnston K. V., Zhao H., Spergel D. N., Hernquist L., 1999, *Astrophys.~J.-Lett.*, **512**, L109
- Kobyzev I., Prince S. J. D., Brubaker M. A., 2019, arXiv e-prints, p. arXiv:1908.09257
- Kobyzev I., Prince S. J., Brubaker M. A., 2021, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **43**, 3964
- Koposov S. E., Rix H.-W., Hogg D. W., 2010, *ApJ*, **712**, 260
- Koposov S. E., et al., 2022, arXiv e-prints, p. arXiv:2211.04495
- Küpper A. H. W., Balbinot E., Bonaca A., Johnston K. V., Hogg D. W., Kroupa P., Santiago B. X., 2015, *Astrophys.~J.*, **803**, 80
- LSST Science Collaboration et al., 2009, arXiv e-prints, p. arXiv:0912.0201
- Li C., Zhao G., 2017, *ApJ*, **850**, 25
- Li T. S., et al., 2022, *ApJ*, **928**, 30
- Lilleengen S., et al., 2023, *MNRAS*, **518**, 774
- Lim S. H., Raman K. A., Buckley M. R., Shih D., 2022, arXiv e-prints, p. arXiv:2211.11765
- Majewski S. R., Skrutskie M. F., Weinberg M. D., Ostheimer J. C., 2003, *ApJ*, **599**, 1082
- Malhan K., Ibata R. A., 2018, *MNRAS*, **477**, 4063
- Malhan K., Ibata R. A., 2019, *MNRAS*, **486**, 2995
- Malhan K., Ibata R. A., Goldman B., Martin N. F., Magnier E., Chambers K., 2018a, *MNRAS*, **478**, 3862
- Malhan K., Ibata R. A., Martin N. F., 2018b, *MNRAS*, **481**, 3442
- Malhan K., Yuan Z., Ibata R., Arentsen A., Bellazzini M., Martin N. F., 2021, arXiv e-prints, p. arXiv:2104.09523
- Malhan K., et al., 2022, *ApJ*, **926**, 107
- Martin N. F., et al., 2022, *MNRAS*, **516**, 5331
- Mateu C., 2022, arXiv e-prints, p. arXiv:2204.10326
- McConnachie A. W., 2012, *AJ*, **144**, 4
- Meingast S., Alves J., 2019, *Astron. & Astrophys.*, **621**, L3
- Meingast S., Alves J., Fürnkranz V., 2019, *Astron. & Astrophys.*, **622**, L13
- Myeong G. C., Evans N. W., Belokurov V., Amorisco N. C., Koposov S. E., 2018a, *MNRAS*, **475**, 1537
- Myeong G. C., Evans N. W., Belokurov V., Sanders J. L., Koposov S. E., 2018b, *MNRAS*, **478**, 5449
- Nachman B., Shih D., 2020, *Phys. Rev. D*, **101**, 075042
- Necib L., Lisanti M., Garrison-Kimmel S., Wetzel A., Sanderson R., Hopkins P. F., Faucher-Giguère C.-A., Kereš D., 2019, *ApJ*, **883**, 27
- Necib L., et al., 2020, *Nature Astronomy*, **4**, 1078
- Newberg H. J., 2010, in AAS/Division of Dynamical Astronomy Meeting #41. p. 5.01
- Newberg H. J., et al., 2002, *ApJ*, **569**, 245
- Papamakarios G., Pavlakou T., Murray I., 2017, in Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R., eds, Vol. 30, Advances in Neural Information Processing Systems. Curran

- Associates, Inc., <https://proceedings.neurips.cc/paper/2017/file/6c1da886822c67822bcf3679d04369fa-Paper.pdf>
- Papamakarios G., Nalisnick E., Jimenez Rezende D., Mohamed S., Lakshminarayanan B., 2019, arXiv e-prints, [p. arXiv:1912.02762](https://arxiv.org/abs/1912.02762)
- Papamakarios G., Nalisnick E. T., Rezende D. J., Mohamed S., Lakshminarayanan B., 2021, *J. Mach. Learn. Res.*, 22, 1
- Price-Whelan A. M., 2017, *The Journal of Open Source Software*, 2
- Price-Whelan A. M., Bonaca A., 2018, *ApJ*, 863, L20
- Price-Whelan A., Sipőcz B., Lenz D., Greco J., Major S., Koposov S., Oh S., Lim P. L., 2020, adrn/gala: v1.1.1, [doi:10.5281/zenodo.3700827](https://doi.org/10.5281/zenodo.3700827), <https://doi.org/10.5281/zenodo.3700827>
- Purcell C. W., Zentner A. R., Wang M.-Y., 2012, *J. Cosmology Astropart. Phys.*, 2012, 027
- Reino S., Rossi E. M., Sanderson R. E., Sellentin E., Helmi A., Koppelman H. H., Sharma S., 2020, arXiv e-prints, [p. arXiv:2007.00356](https://arxiv.org/abs/2007.00356)
- Robin A. C., Reylé C., Derrière S., Picaud S., 2003, *A&A*, 409, 523
- Rybicki J., Demleitner M., Fouesneau M., Bailer-Jones C., Rix H.-W., Andrae R., 2018, *PASP*, 130, 074101
- Sanders J. L., Binney J., 2013, *MNRAS*, 433, 1813
- Sanders J. L., Bovy J., Erkal D., 2016, *MNRAS*, 457, 3817
- Sanderson R. E., et al., 2020, *ApJS*, 246, 6
- Sharma S., Bland-Hawthorn J., Johnston K. V., Binney J., 2011, *ApJ*, 730, 3
- Shih D., Buckley M. R., Necib L., Tamasan J., 2022, *MNRAS*, 509, 5992
- Shipp N., et al., 2018, *ApJ*, 862, 114
- Shipp N., et al., 2019, *ApJ*, 885, 3
- Shipp N., et al., 2021, *ApJ*, 923, 149
- Shipp N., et al., 2022, arXiv e-prints, [p. arXiv:2208.02255](https://arxiv.org/abs/2208.02255)
- Tronrud T., Tissera P. B., Gómez F. A., Grand R. J. J., Pakmor R., Marinacci F., Simpson C. M., 2022, *MNRAS*, 515, 3818
- Varghese A., Ibata R., Lewis G. F., 2011, *MNRAS*, 417, 198
- Vasiliev E., Baumgardt H., 2021, *MNRAS*, 505, 5978
- Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, *ApJ*, 827, L23
- Yuan Z., Chang J., Banerjee P., Han J., Kang X., Smith M. C., 2018, *ApJ*, 863, 26
- Yuan Z., et al., 2022, *ApJ*, 930, 103
- Zonca A., Singer L., Lenz D., Reinecke M., Rosset C., Hivon E., Gorski K., 2019, *Journal of Open Source Software*, 4, 1298
- de Bruijne J. H. J., Lammers U., Perryman M. A. C., 2005, in Turon C., O’Flaherty K. S., Perryman M. A. C., eds, *ESA Special Publication Vol. 576, The Three-Dimensional Universe with Gaia*. p. 67

This paper has been typeset from a $\mathrm{T}_{\mathrm{E}}\mathrm{X}/\mathrm{L}^{\mathrm{A}}\mathrm{T}_{\mathrm{E}}\mathrm{X}$ file prepared by the author.