

Toward the fabric of the Milky Way

I. The density of disk streams from a local 250^3 pc³ volume

Sebastian Ratzenböck^{1,2,3,*}, João Alves^{1,2}, Emily L. Hunt⁴, Núria Miret-Roig¹,
Stefan Meingast¹, and Torsten Möller^{2,3}

¹ University of Vienna, Department of Astrophysics, Türkenschanzstraße 17, 1180 Vienna, Austria

² University of Vienna, Research Network Data Science at Uni Vienna, Kolingasse 14-16, 1090 Vienna, Austria

³ University of Vienna, Faculty of Computer Science, Währinger Straße 29/S6, 1090 Vienna, Austria

⁴ Landessternwarte, Zentrum für Astronomie der Universität Heidelberg, Königstuhl 12, 69117 Heidelberg, Germany

Received 12 August 2024 / Accepted 6 January 2025

ABSTRACT

We studied 12 disk streams found in a 250^3 pc³ volume in the solar neighborhood, which we define as coeval and comoving stellar structures with aspect ratios greater than 3:1. Using *Gaia* Data Release 3 data and the advanced clustering algorithms SigMA and Uncover, we identified and characterized these streams beyond the search volume, doubling, on average, their known populations. We estimate the number density of disk streams to be ≈ 820 objects/kpc³ (for $|Z| < 100$ pc), or surface densities of ≈ 160 objects/kpc². These estimates surpass N-body estimates by one to two orders of magnitude and challenge the prevailing understanding of their destruction mechanisms. Our analysis reveals that these 12 disk streams are dynamically cold with 3D velocity dispersions between 2 and 5 km s⁻¹, exhibit narrow sequences in the Hertzsprung-Russell diagram, and are highly elongated with average aspect ratios of 7:1, extending up to several hundred parsecs. We find evidence suggesting that one of the disk streams, currently embedded in the Scorpius-Centaurus association, is experiencing disruption, likely due to the primordial gas mass of the association.

Key words. methods: data analysis – methods: statistical – open clusters and associations: general – solar neighborhood – Galaxy: structure

1. Introduction

Disk streams are coeval and comoving elongated stellar structures within the Milky Way's disk. They are the disk counterparts of halo streams, which are traditionally studied in the context of the Milky Way halo (e.g., Grillmair et al. 1995; Odenkirchen et al. 2001; Malhan et al. 2018; Bovy 2016; Ibata et al. 2016; Price-Whelan & Bonaca 2018). Unlike halo streams, which are primarily associated with tidally disrupted globular clusters and dwarf galaxies accreted by the Milky Way, disk streams are believed to originate from the disruption of bound and unbound clusters (associations) born in the Milky Way (e.g., Eggen 1996; Meingast et al. 2019; Kamdar et al. 2019; Meingast et al. 2021; Kamdar et al. 2021). Disk streams, in particular nearby ones, constitute excellent laboratories for studying planet formation (e.g., Curtis et al. 2019; Newton et al. 2021) and can provide insights into processes such as the dissolution of star clusters, the influence of Galactic dynamics, and the interaction between stellar structures and giant molecular clouds (GMCs).

The remarkable precision of ESA's *Gaia* data within the local Milky Way has revolutionized our ability to uncover disk streams, an endeavor that previously seemed near impossible due to their subtle presence against the densely populated backdrop of the Milky Way disk. The first bona fide disk stream, Meingast-1 (Pisces-Eridanus; Meingast et al. 2019), is a coeval and comoving unbound structure with an age of about 120 Myr (Curtis et al. 2019), a length of at least 400 pc, and a vertical extent of only about 50 pc, covering about 120° of the sky.

* Corresponding author; sebastian.ratzenboeck@univie.ac.at

This stellar structure was the first chemically homogeneous disk stream identified on the Milky Way disk (Ratzenböck et al. 2020; Hawkins et al. 2020).

While there is currently no working definition of a disk stream, in this work we define it as a coeval and comoving stellar structure with aspect ratios (between the first and third principal components in *XYZ*) greater than 3:1. This broad definition includes cluster tidal tails (e.g., Röser et al. 2019; Meingast & Alves 2019; Jerabkova et al. 2021; Kroupa et al. 2022), unbound clusters (associations), young moving groups, young local associations (e.g., Gagné & Faherty 2018; Beccari et al. 2020; Miret-Roig et al. 2020; Tian 2020), and cluster coronas (Meingast et al. 2021; Moranta et al. 2022). This working definition is appropriate at this stage, as it keeps us from over-classifying the outcome of different complex processes (e.g., Galactic tidal forces, differential rotation, and initial star-forming gas configurations). We leave a discussion on the relative role of the different formation processes, and a better disk stream definition, to be tackled whenever a statistically significant sample of disk streams becomes available.

Notwithstanding recent advancements in the detection of initial disk streams, their potential utility as observational laboratories for the processes governing planet formation and the genesis of the Galactic field population remains largely unexamined. This is due to the lack of a census of disk streams in the Milky Way. While we know of about 100 halo streams (Mateu 2023), only a handful of disk streams have been identified. This limits our understanding of the processes transforming clustered young stellar populations into the Galaxy's field population. To

address this, we need advanced tools to identify coeval and comoving populations down to densities well below the average volume density of stars in the Milky Way.

In this work we focus on establishing the volume and surface density of disk streams in the local Milky Way. We studied 12 disk streams within a fully sampled 250^3 pc^3 local volume. These 12 disk streams were identified as interloper structures in the Scorpius-Centaurus (Sco-Cen) study [Ratzenböck et al. \(2023a, hereafter Paper I\)](#), which employed the SigMA algorithm. Here, we search for additional members of these disk streams outside the initially defined search box of 250^3 pc^3 , using the Uncover algorithm from [Ratzenböck et al. \(2023c\)](#) to characterize their basic physical properties, such as mass and age. Our volume-complete sample will allow us to provide a first estimate of the abundance and lifetime of disk streams in the local Milky Way disk.

2. Data

This study uses position and velocity information from the *Gaia* Data Release 3 (DR3; [Gaia Collaboration 2023](#)). We selected all sources within 500 pc that pass the following quality criteria, as defined in Paper I:

$$\begin{aligned} \varpi/\sigma_{\varpi} &> 4.5, \\ \varpi &> 0. \end{aligned} \quad (1)$$

We determined the distance by inverting the parallax measurement. Due to the large parallax signal-to-noise criterion and relatively small distances ($<500 \text{ pc}$), inverted parallaxes are in good agreement with distance predictions from, for example, [Bailer-Jones et al. \(2021\)](#) but also avoid introducing a space distribution prior that may smooth out real substructure that we wish to find. Combined with *Gaia* DR3 right ascension (ra, deg) and declination (dec, deg), we determined 3D space positions in the heliocentric Galactic Cartesian coordinate frame *XYZ* (in pc). After applying the quality criteria in Eq. (1), our final dataset contains 25 475 384 sources, referred to as the 500 pc search data. We also selected a subsample of the 500 pc search data (hereafter, 6D search data) with an absolute radial velocity error of less than two: $\sigma_{v_r} < 2 \text{ km s}^{-1}$. This sample contains 2 078 715 sources within 500 pc for which we computed precise 3D space motions in heliocentric Galactic Cartesian coordinates *UVW* (in km s^{-1}). We find no systematic differences in the resulting streams' morphology (length, location, velocity dispersion) when clustering in the Galactocentric cylindrical velocity coordinates (v_R, v_ϕ, v_z) instead and only percent level differences in the identified stream sizes. Since the impact of these two reference frames on our pipeline is negligible, we opted to use the heliocentric Galactic Cartesian coordinate system.

This study focuses on 12 disk streams identified as interlopers by the SigMA clustering pipeline in Paper I. Paper I aimed to investigate the sub-structuring of the Sco-Cen association, identifying distinct Sco-Cen subpopulations as well as 48 additional stellar clusters that were not kinematically related to Sco-Cen and thus excluded from further discussion. Among these 48 unrelated clusters, this study follows up on 12 populations that exhibit significant prolate morphologies with pronounced elongations, meeting our aspect ratio criterion of $>3:1$. The remaining 36 clusters, classified as established open clusters or moving groups, do not display the morphological or kinematic characteristics consistent with disk streams and are therefore excluded from this analysis.

These 12 disk streams are located within the search box defined in Paper I with dimensions *X*: $[-50, 250]$, *Y*: $[-200, 50]$, and *Z*: $[-95, 100]$ pc. This corresponds to a box size of $300 \text{ pc} \times 250 \text{ pc} \times 195 \text{ pc}$, encompassing a total volume of $14\,625\,000 \text{ pc}^3$. For simplicity and clarity, we approximate and refer in this manuscript to this volume as a cube with 250 pc sides (i.e., a $250^3 \text{ pc}^3 \text{ box}^1$). Most of the 12 streams appear to terminate abruptly at the edges of this box, suggesting its boundaries truncate them. To recover the potential missing members of these truncated clusters, we expanded the search to the full 500 pc dataset, utilizing a combination of precise phase-space information (*XYZ + UVW*) and its projected 5D counterpart using *XYZ* and tangential velocities v_α and v_δ (in km s^{-1}) for cases where radial velocity data are unavailable.

3. Method

We used a combination of two automated search tools, SigMA and Uncover ([Ratzenböck et al. 2020](#)), for the stream membership analysis. In this study, we employed SigMA to recover the full extent of each progenitor stream using 6D phase space data. Given a 6D selection, Uncover was used to identify members without (precise) radial velocity measurements.

Due to its size, our search strategies (see Sects. 3.1 and 3.2) cannot directly apply to the 500 pc search data. Consequently, we adopted a targeted approach that individually examines the 12 progenitor streams within a more manageable local selection of the abovementioned datasets. We discarded sources in the position and velocity space that are likely unrelated with progenitor streams. To do so, we first computed the three principal axes of the progenitor stream by applying principal component analysis to their *XYZ* coordinates. Then, we determine the extent Δx_i of each progenitor stream along its principal axes by identifying the minimum and maximum positions of sources projected onto each axis. The box within which we search for additional stream members is chosen to extend $3 \times \Delta x_i$ out from the progenitor stream's median *XYZ* position along each principal axis, with $i \in \{1, 2, 3\}$. Thus, we limit the streams' sizes to expand by a factor of 6. The extent of additional members we recover across all 12 streams is well constrained within their respective search boxes and does not come close to any border.

To improve the contrast of the populations against the background field in *XYZ*, we remove unlikely members through a kinematic selection. To this end, we compute the median Galactic Cartesian 3D velocity v for each stream and retain sources in (local versions of) the 6D search data if a source's velocity difference to v is less than 20 km s^{-1} . Sources beyond this velocity cut, concerning, for example, members with large radial velocity uncertainties, can still enter the final selection through our two-step selection pipeline detailed in Sects. 3.1 and 3.2. This velocity selection is still useful, though, as the blind search (see Sect. 3.1) becomes empirically more sensitive to the low-contrast tails of the streams.

3.1. Expanding the progenitor stream's extent

We search for overdensities in 6D phase space using the SigMA pipeline to identify each stream in its respective local 6D search

¹ A more accurate equivalent is a cube with sides of $\sim 244 \text{ pc}$. For ease of notation, we have rounded this value to the nearest 50 pc. Naturally, all volume and surface density calculations are done using the true box volume of $14\,625\,000 \text{ pc}^3$.

data. **SigMA** is an unsupervised hierarchical density-based learning method that identifies clusters as statistical overdensities in input space (here, 6D phase space) separated by regions of significantly lower density. For a detailed description of the **SigMA** pipeline, we refer to Paper I.

Due to their size and relative proximity, the 12 progenitor streams cover huge areas on the sky, with angular extents of about 90° on average and as large as 170° . Even with distinguished compact overdensities in *UVW*, projection effects can cause highly dispersed and non-convex distributions in tangential velocity space, significantly hindering detection capabilities with density-based clustering tools. Thus, in contrast to previous applications of **SigMA**, we ran the pipeline directly on the 6D search data to tailor the pipeline to these large structures. We adopt the parameter choices discussed in Paper I, except for the appropriate scaling factor values, which are affected by the change of the input space from the 5D to the 6D phase space. We update the kinematic scale factors (see Sect. 3.3.3 in Paper I) using the selection of progenitor stream members (see Appendix A for a detailed description).

Due to the large box sizes, running **SigMA** in each of the 12 local 6D datasets results in multiple recovered overdensities. The population of interest is determined as the cluster maximizing the crossmatch rate with the progenitor stream. Except in one case, which exhibits a possible overlap with the Coma-Ber cluster, we find a clean and unique match between resulting **SigMA** groups and progenitor streams. We discuss this overlap in more detail in Sect. 4.

3.2. Identifying unknown members

The resulting 6D **SigMA** selection discussed in Sect. 3.1 is subsequently fed into the **Uncover** pipeline to identify unknown stream members for which no (precise) radial velocities are available. **Uncover** is an extended membership analysis technique that integrates known members of stellar populations to identify undetected members.

Uncover merges a powerful black-box algorithm, one-class support vector machines (OCSVM; Schölkopf et al. 1999), into a statistical framework to provide meaningful parameter selection tools and improve membership accuracy. For a detailed description of the **Uncover** pipeline, we refer to Ratzenböck et al. (2020), Grasser et al. (2021), and Ratzenböck et al. (2023c). Instead of manually choosing OCSVM model parameters, **Uncover** employs ranges of interpretable summary statistics the trained model must adhere to, such as estimates on the number or the maximally allowed velocity dispersion of yet unseen members. We adopted the parameter selection approach presented in Ratzenböck et al. (2020) due to the similarity in the presented use case to ours and refer the reader to this manuscript for an in-depth discussion on parameter choices.

We applied **Uncover** to the local 5D search box (using the features $\{X, Y, Z, v_\alpha, v_\delta\}$) where the candidate members identified in the previous clustering step were used as the training set. The final inferred member selection remains remarkably stream-like, with all 12 populations appearing highly elongated along their bulk velocity direction. Inside the original search box, encompassing a volume of 250^3 pc^3 , these 12 disk streams are tightly packed and appear to envelop Sco-Cen.

4. Results

Table 1 summarizes the 12 recovered disk streams. We assigned each object to an increasing stream identifier (SID) from S1 to

S12 and report the name assigned by crossmatching our sources to the literature. Except for disk stream S1, we find a sufficiently good match between the two population morphologies in the literature (see Sect. 4.6). Since object S1 is unknown, we name it “Ratzenboeck 1”. In the following, we discuss the properties of the identified streams in more detail.

4.1. Spatial distribution

Figure 1 shows the positional extent of the 12 identified disk streams in Galactic Cartesian coordinates (*XYZ*). Disk streams are displayed via simplified colored shapes along with Sco-Cen’s subcluster population, represented by gray shapes (see Ratzenböck et al. 2023b). Figure 1 highlights the scale of the recovered disk streams whose size exceeds that of Sco-Cen by up to a factor of 3. Although the entire Sco-Cen population extends across 150 pc in the *X-Y* plane, the length of the identified disk streams ranges between 120 and 430 pc with an average length of 280 pc. Furthermore, all clusters are highly prolate, meaning the median absolute deviation (MAD) along one principal axis is significantly larger than the MAD along the other two principal axes. The MAD is used here to provide a reliable and robust measure of the spread along each principal axis. All 12 disk streams have aspect ratio measurements (ratio between largest to smallest principal component axis) ranging from 3.3 to 10.1 with an average aspect ratio of 7.1. Additionally, Table 1 provides MAD ratios among the three principal axes in *XYZ* normalized to the smallest component, showcasing the prolate nature of the identified populations. However, we note that the provided aspect ratio values likely constitute lower limits, and we expect these structures to grow in length with improved precision of upcoming *Gaia* data releases.

Figure 2 displays the individual recovered sources color-coded by disk stream membership. The boxes are chosen to retain an equal aspect ratio, highlighting the streams’ extents of up to a few hundred parsecs in the *X-Y* plane compared to their compact size along the *Z* dimension. In the *X-Y* plane, the disk streams cover different inclination angles between the *X* and *Y* axis ranging from 45° to 90° . Figure 3 shows the on-sky distribution of the 12 recovered disk streams, further highlighting the large extent of the identified populations.

4.2. Velocity distribution

The populations are dynamically cold, with 3D velocity dispersions between 2.1 and 5.1 km s^{-1} . These values were determined using a robust deconvolution method that simultaneously addresses large radial velocity measurement errors and accounts for outliers (see Appendix D for details). This approach reduces the estimated 3D velocity dispersion by a factor of ~ 5 when compared to traditional empirical covariance estimates. Figures E.2–E.4 illustrate the distribution of sources in Cartesian velocity space (*UVW*). To mitigate the risk of underestimating 3D velocity dispersions, we also calculated them using the MAD. While this robust measure effectively accounts for potential outliers, it does not (inherently) correct for large radial velocity measurement errors. Consequently, it yields slightly higher velocity dispersion estimates, ranging from 2.7 to 7.9 km s^{-1} . Both estimates are presented in Table 1 where the MAD estimate is indicated in parentheses. (For a discussion on the velocity dispersion estimation, see Sect. 5.1 and Appendix D.)

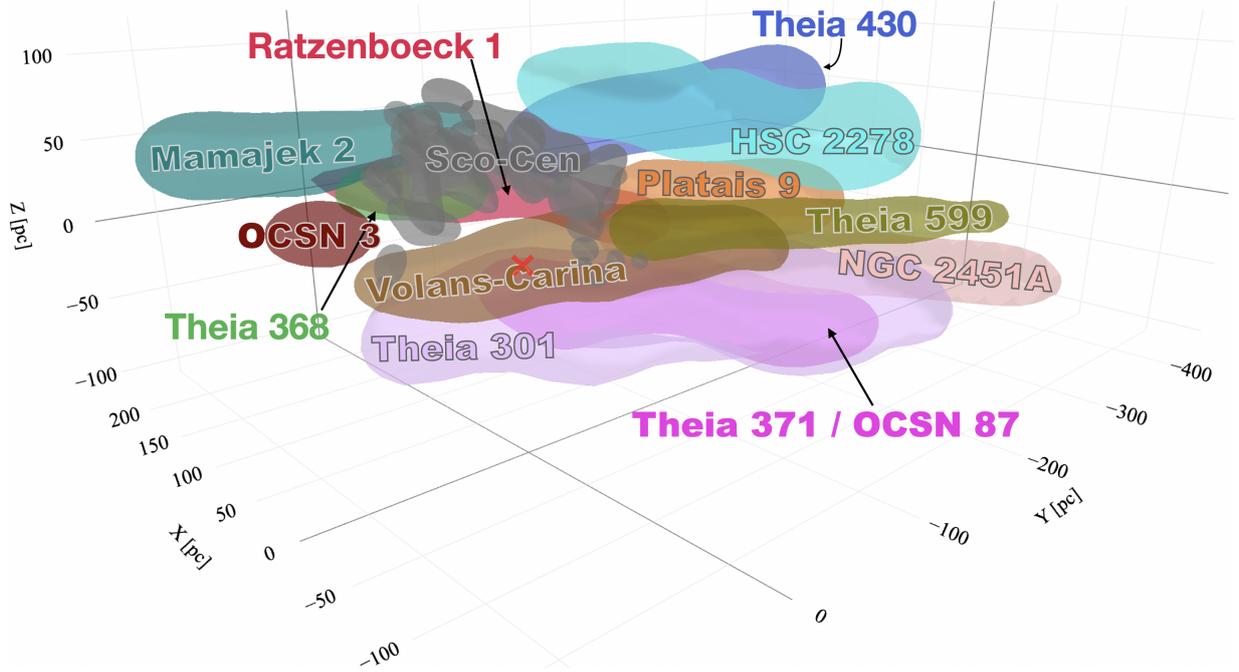


Fig. 1. 3D distribution of 12 disk streams (in color) alongside Sco-Cen from Ratzenböck et al. (2023b). The Sun is at (0,0,0) and is represented by the red “x”. For better visualization, see the link to the [interactive 3D version](#) of this figure, which allows the user to toggle on and off individual sources and the initial search box of 250^3 pc^3 .

Table 1. Overview of the computed cluster parameters and statistics of the 12 identified disk streams.

SID	Literature name ^(a)	Size	Age ^(g) (Myr)	Literature age ^(b) (Myr)	Reference ^(c)	Length ^(d) (pc)	Aspect ratio	M_{tot} (M_{\odot})	M_{sys} ^(d) (M_{\odot})	σ_{3D} ^(h) (km s^{-1})	V (UVW) (km s^{-1})	S/N
1	Ratzenboeck 1	273	103^{+19}_{-7}	–	–	320 ± 10	7.5 : 1.3 : 1	140	369 ± 37	5.1 (7.9)	(-6.6, -26.9, -11.7)	10 ± 2
2	Theia 368	172	96^{+17}_{-11}	140, 93, 126, 158	2, 3, 4, 7	120 ± 20	7.8 : 2.2 : 1	96	198 ± 34	2.8 (3.8)	(-1.0, -28.9, -14.0)	76 ± 16
3	Theia 430	248	288^{+35}_{-20}	80, 190	3, 4	250 ± 20	7.6 : 1.9 : 1	144	310 ± 23	2.5 (4.4)	(-26.4, -7.5, -10.7)	10 ± 3
4	Platais 9	454	66^{+4}_{-2}	78, 50, 37, 54	1, 2, 3, 4	280 ± 30	6.9 : 1.7 : 1	248	524 ± 41	2.8 (4.2)	(-23.6, -15.4, -6.1)	21 ± 11
5	HSC 2278	203	976^{+251}_{-185}	571	3	280 ± 30	8.2 : 2.6 : 1	80	168 ± 20	2.7 (4.3)	(-4.9, -4.4, 0.1)	5 ± 1
6	Theia 371 / OCSN 87	172	158^{+26}_{-34}	355, 144, 146, 631	2, 3, 4, 7	300 ± 30	5.6 : 1.9 : 1	105	262 ± 15	2.1 (2.7)	(-0.6, 1.3, -1.1)	12 ± 3
7	NGC 2451A	845	49^{+5}_{-2}	44, 56, 26, 50	1, 2, 3, 4	330 ± 10	5.3 : 1.6 : 1	435	812 ± 77	3.5 (5.7)	(-27.3, -14.4, -12.7)	28 ± 9
8	Mamajek 2	324	128^{+13}_{-18}	126, 99, 178, 125	2, 3, 4, 5	220 ± 10	8.4 : 2.3 : 1	169	380 ± 25	3.4 (5.9)	(-10.3, -25.8, -4.4)	21 ± 8
9	Theia 301 ^(e)	534	101^{+11}_{-6}	107	4	430 ± 10	7.1 : 2.3 : 1	279	652 ± 68	2.1 (3.6)	(-8.9, -27.3, -11.8)	6 ± 2
10	Volans-Carina	566	106^{+23}_{-5}	355, 212, 189, 89	2, 3, 4, 6	280 ± 10	6.6 : 1.6 : 1	255	570 ± 44	3.8 (5.6)	(-16.2, -27.0, -0.46)	9 ± 4
11	OCSN 3	196	411^{+75}_{-109}	282, 705	2, 3	160 ± 40	3.3 : 1.0 : 1	120	328 ± 26	2.1 (3.3)	(-7.9, -17.2, -13.4)	114 ± 21
12	Theia 599 ^(f)	418	461^{+106}_{-26}	264	4	350 ± 10	10.1 : 2.3 : 1	233	527 ± 53	3.1 (4.5)	(-35.5, -17.0, -0.6)	22 ± 9

Notes. ^(a)We aim to report the name assigned to a cluster based on its first literature appearance, given a one-to-one relationship and enough similarity between our and the literature selection. For clusters that appear to be contaminated or describe a smaller (noncentral) portion of the disk stream, we also report the name of a more recent study whose extraction aligns better with our findings. This is particularly the case for disk streams S1 and S6, where we have either established a new name (Ratzenboeck 1) or attributed two names to describe the population. ^(b)Literature ages are compiled from the discovery paper and/or follow-up studies with refined membership lists. In cases of very contaminated extractions, we omit the age altogether (see Table C.1). For the open clusters Platais 9 and NGC 2451A, we report the age determined by Bossini et al. (2019); see Meingast et al. (2021) for a more thorough discussion on their fundamental physical parameters across the literature. Here, we report the mean population age rounded to the nearest integer and omit reported uncertainties for easier readability. We refer the reader to the respective source material referenced in the respective “Reference” column for more information. ^(c)Ages listed in the “Literature age” column correspond to the following references: (1) Bossini et al. (2019), (2) Qin et al. (2023), (3) Hunt & Reffert (2023), (4) Kounkel & Covey (2019), (5) Mamajek (2006), (6) Gagné et al. (2018b), (7) Fürnkranz et al. (2024). ^(d)Due to possible contamination, we determine the disk stream lengths not directly from the data but as average peak-to-peak lengths from 100 bootstrap samples, which we report alongside respective standard deviations. ^(e)The disk stream S9 overlaps with the group Theia 301 identified by Kounkel & Covey (2019) who relate it to the AB Doradus moving group. We note here that our selection of S9 is significantly different from Theia 301. While Theia 301 consists of just over 1300 sources, only 188 of them overlap with S9, which itself has over 500 members. Regardless of these differences, we refer to S9 as “Theia 301” throughout this manuscript. ^(f)The disk stream S12 overlaps with the group Theia 599 identified by Kounkel & Covey (2019). Their morphologies and membership lists are similar enough to document a match here. However, we note that the observational HRD and XYZ distribution of Theia 599 indicates substantial field contamination. ^(g)We find significantly nonzero ($A_V \geq 0.1 \text{ mag}$) extinction toward disk streams S1 ($A_V = 0.2 \text{ mag}$), S3 ($A_V = 0.1 \text{ mag}$), S4 ($A_V = 0.1 \text{ mag}$), S8 ($A_V = 0.5 \text{ mag}$), S11 ($A_V = 0.3 \text{ mag}$), S12 ($A_V = 0.1 \text{ mag}$). ^(h)We show two estimations of the 3D velocity dispersion. First, the dispersion estimate is derived from the XD procedure. Second, the value in parentheses is determined via the MAD. For more details, see Appendix D.

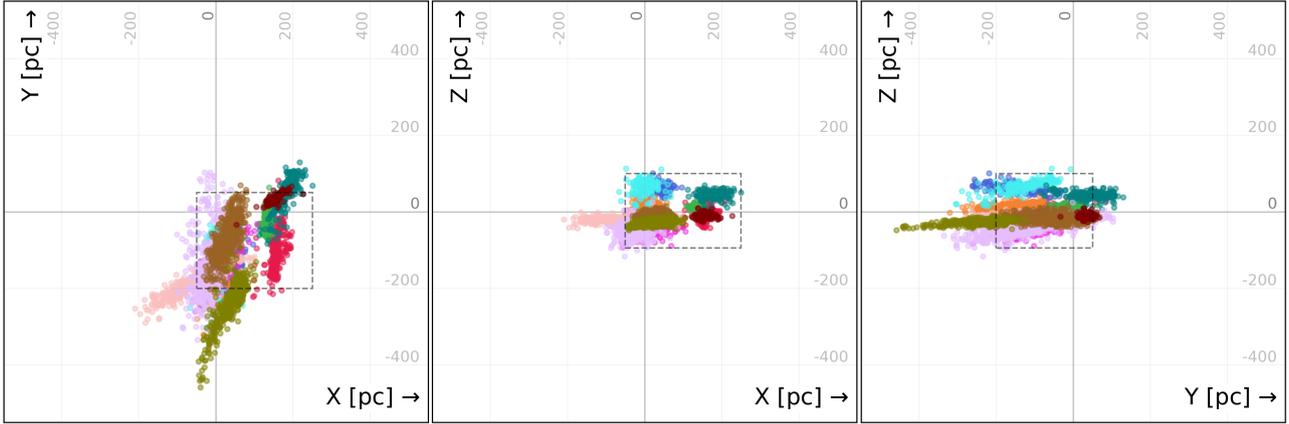


Fig. 2. Spatial distribution of our selection for 12 disk streams in heliocentric Galactic coordinates. Colors have the same meaning as in Fig. 1. Because of sensitivity, the elongations of these disk streams are lower limits to their true elongation. The dashed rectangle indicates the search volume of approximately 250^3 pc^3 within which we aim to identify the local disk stream population. Many of the progenitor streams identified extend far beyond the initial search box.

4.3. Result validation

Since the recovered disk streams cover large volumes with many co-spatial stars, the chance of random field contaminants is much higher than, for example, in more compact open cluster configurations. To validate our clustering pipeline, we applied three different contamination estimation procedures.

First, our deconvolution approach enables us to quantify the contamination rate in each sample by incorporating a dedicated outlier component into a two-component Gaussian mixture model (GMM). This component accounts for both “true” outliers and artificial kinematic outliers caused by binaries or large radial velocity uncertainties. While its primary purpose is to prevent the artificial inflation of the signal component, the inferred mixture weight of the outlier component can also be directly interpreted as the fraction of random outliers in the sample. Using this procedure, we find an average contamination rate across the 12 disk streams of 9%, with individual values ranging between 5 and 17% (see Appendix D).

Second, we compared a background population of co-spatial sources that share the same volume (in XYZ) with our selected stream members in the observational Hertzsprung-Russell diagram (HRD). We find that the distribution of identified stream members represents a significantly narrower configuration around selected model isochrones than a sample from the background population. This provides substantial evidence that the disk stream members we identified constitute coeval populations (see Appendix E).

Third, we compared the velocity distribution of recovered disk streams to that of the corresponding background sources. By determining the number density of stream members and contrasting it with the expected number density of background sources moving with the stream’s bulk motion, we derived a signal-to-noise ratio (S/N) and contamination estimate for each stream. We derive a mean S/N of 28, with values ranging from 5 to 114, and estimate a contamination level of $7 \pm 4\%$. We provide the estimated S/N in Table 1 (see Appendix E).

Although kinematically distinct, several disk streams exhibit partial spatial overlap with their respective members, which are intermixed to varying degrees. Notably, we find three groups: (1) Theia 430 and HSC 2278, (2) Theia 371/OCSN 87, NGC 2451A, Theia 301, and OCSN 3, and (3) a group made up of disk streams Ratzenboeck 1, Theia 368, and Mamajek 2. Members of the third

group are found to also spatially coexist, at varying degrees, with members of the Sco-Cen association, as discussed in Sect. 5.4. These spatial overlaps raise the possibility of shared origins among these structures. However, a detailed kinematic analysis reveals that in all but one case, these co-spatial streams show statistically significant differences in their velocities and ages. This analysis is further detailed in Appendix D.3 and Sect. 5.1, where we discuss a case that warrants further scrutiny, Ratzenboeck 1 and Theia 386.

Compared to previous unsupervised studies (not including targeted searches such as Meingast et al. 2021), our pipeline identifies, on average, more than twice as many candidate members (see Sect. 4.6). Our pipeline allows us to detect median stream densities (number of sources divided by the entire population volume) of 1 star per 10^3 pc^3 (or 0.001 stars/ pc^3), which is about 50 times lower than the surrounding field density. At both extremes, average stream densities are an order of magnitude apart. Whereas the densest structures, such as NGC 2451A and Platais 9, have average densities (across the entire population) of about 2–3 sources per 10^3 pc^3 , we reach average densities as low as 0.2–0.5 sources per 10^3 pc^3 (min 0.0002 stars/ pc^3) across the entire disk stream in HSC 2278, Theia 371/OCSN 87, Theia 301, Theia 430, and Ratzenboeck 1, effectively resolving structures 250 times below the field density in XYZ coordinates.

4.4. Age and mass

To estimate the masses and ages of the streams, we adopted the isochrone fitting procedure from Ratzenböck et al. (2023b) (using PARSEC; Marigo et al. 2017), where noise contributions around isochronal curves are modeled via skewed Cauchy distributions. The skewed Cauchy distribution naturally accounts for nonsymmetric noise sources, such as unresolved binaries and differential reddening effects inside the cluster. At the same time, its heavy tails are known for their robustness to outliers (Hampel et al. 2011). We derived the total stellar mass (M_{tot}) using the inferred isochronal ages and the sources’ relative positions to the best-fitting isochrone (assuming solar metallicity $Z_{\odot} = 0.0152$) in the *Gaia* color–absolute magnitude diagram (using $G_{\text{BP}} - G_{\text{RP}}$ as color).

Combined cluster masses of the entire system (M_{sys}) are estimated by taking into account *Gaia*’s incompleteness toward very bright and faint objects. Assuming a Kroupa initial mass

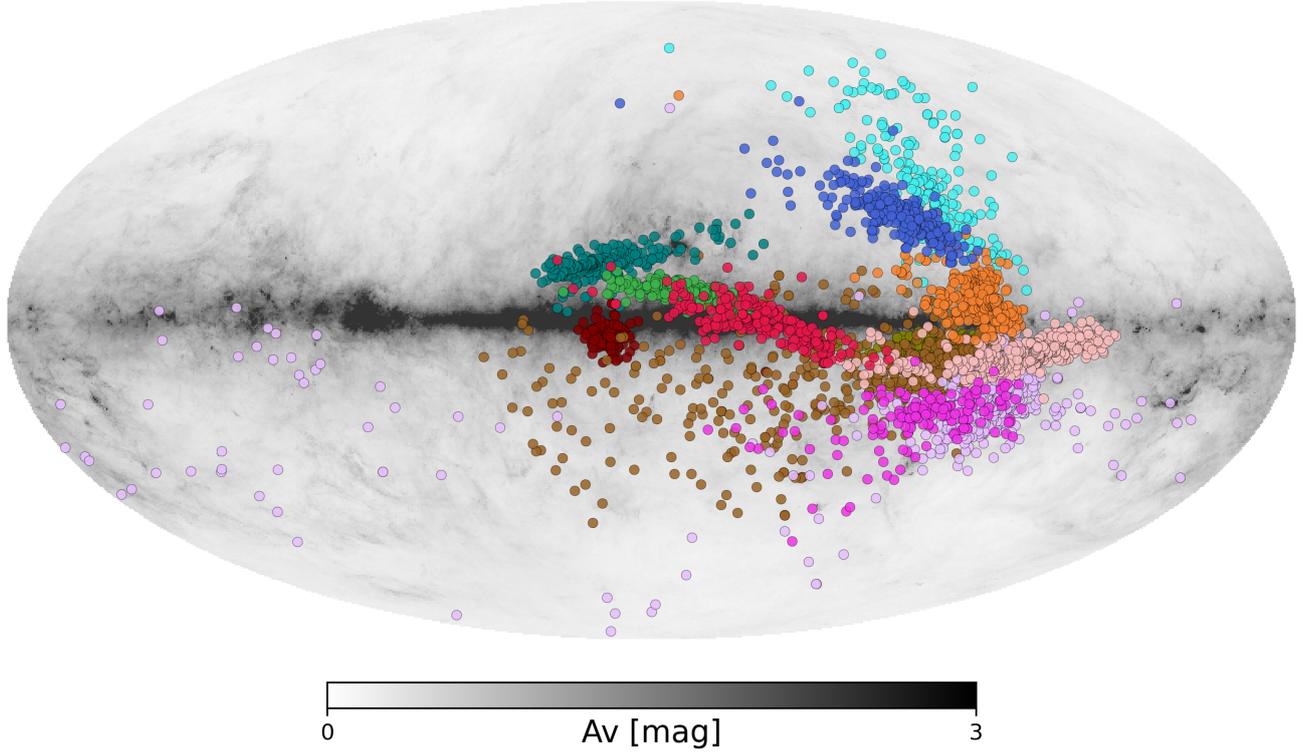


Fig. 3. On-sky distribution of our selection for 12 disk streams on top of the *Planck* dust map (Planck Collaboration XI 2014). All the streams were identified inside a fully sampled 250^3 pc^3 in the local Milky Way. Colors have the same meaning as in Fig. 1.

function (IMF; here Kroupa 2001), we determined M_{sys} by fitting an IMF to the observed mass distribution as a function of variable cluster mass. Figure B.1 shows the observed mass functions for the 12 disk streams along with the best-fitting IMF. In practice, for each cluster, we binned ($N=10$) the mass range where *Gaia* is approximately complete and minimized the χ^2 statistic (i.e., the normalized sum between the squared difference of observed and estimated counts across all mass bins). This mass range is determined from the *G*-band range between magnitudes 12 and 17 where *Gaia* DR3 is expected to be complete (Riello et al. 2021), which we subsequently translated to a mass range using the distances to cluster members to obtain absolute magnitudes and corresponding isochrones. An as yet under-explored field concerns the systematic effects of clustering algorithms on derived physical parameters. For the mass function, we suspect that the cluster selection function is a second-order effect that does not drastically affect our results. However, these selection biases should be considered when studying the detailed shape of the observational mass function.

Figure 4 shows the observational HRDs for the 12 identified disk stream along with the best fitting isochronal curve. Table 1 summarizes the main physical parameters of the 12 identified disk streams and compares their ages to those reported in the literature. We find that our age estimates, in general, agree with other references. The ages of the disk streams range from about 50 Myr to around 1 Gyr, with a median age of 117 Myr. Most streams (8 of 12) are relatively young, with ages less than 200 Myr.

We find no apparent relationship between a population’s age and length, as measured by the sample Pearson correlation coefficient ($r = -0.06$). This observation contradicts expectations from tidal disruption processes, where stars gradually migrate into a system’s tidal tails. However, as explained above, the

lengths derived in this work are lower limits, and this result will likely change with better *Gaia* DR4 data. In contrast, we observe a minimal correlation between a cluster’s volume and age ($r = 0.1$), although with considerable scatter. On average, our findings suggest that (when ignoring outliers) the volume of disk streams increases by approximately 500–1000 pc^3 per Myr. Figure B.2 provides an overview of all pairwise correlations mentioned in this section.

We find that the disk stream mass can moderately predict the length of a stream ($r = 0.55$), which is reasonable since a larger system mass provides a higher signal contrast over the background for the entire stream size. A related statistic, the population density (stars per cubic parsec) over time, shows a moderate negative correlation ($r = -0.56$). This anticorrelation suggests that populations dissolve into the surrounding field over the lifetime of a cluster.

4.5. Boundedness estimation

Similarly to Meingast et al. (2021, hereafter MAR21), our objective was to investigate the “boundedness” and dissolution process of recovered populations. Compared to their study, we did not start from a set of prominent open clusters but rather aimed to analyze all stream-like structures inside a given search volume. Therefore, we hypothesize that most identified disk streams do not have a bound core. To assess the boundness of a cluster’s core, we analyzed the Jacobi radius, r_J (i.e., the dynamical tidal radius) of each cluster population as outlined in MAR21 and Hunt & Reffert (2024). To do so, we computed the cumulative total and completeness-corrected mass enclosed at varying distances, $M_{\text{obs}}(r)$, from the density mode of each population, estimated with a kernel density estimate (KDE); for a detailed description, we refer the reader to Hunt & Reffert (2024), whose

Table 2. Overview of the Jacobi radii, r_J , and masses, M_J , alongside the bound mass fraction of the four clusters where we find a bound core; see Sect. 4.4 for more information.

SID	Literature name	r_J (pc)	$M_J^{(a)}$ (M_\odot)	Bound mass fraction $^{(b)}$ (%)
4	Platais 9	5.9	79.5	32.1
7	NGC 2451A	7.7	170.2	39.1
8	Mamajek 2	5.1	56.0	33.2
11	OCSN 3	4.6	44.2	36.9

Notes. ^(a)The Jacobi mass is determined using the observed masses (see M_{tot}) and not the system masses that are derived from fitting a Kroupa IMF to the observed mass histograms. ^(b)The boundedness fraction is determined by dividing the Jacobi masses by the observed system masses, i.e., M_{tot} as provided in Table 1; see also footnote (a).

implementation we have adopted. By intersecting the observed radial mass profile with the theoretical Jacobi mass for a cluster of a given size $M_J(r)$, denoting the distance from a cluster center to the Lagrange points L_1 and L_2 of a bound system (King 1962; Ernst et al. 2010; Portegies Zwart et al. 2010), we obtained a system’s Jacobi radius, r_J , and mass, M_J . We note that although the Jacobi radius procedure as implemented Hunt & Reffert (2024) (and here) has some differences from its application in MAR21, the Jacobi radius results for NGC 2451A and Platais 9 (both groups appear in MAR21 and this study) agree within 0.5 pc and 2.3 pc, respectively.

We employed the definition of Hunt & Reffert (2024) to determine whether a system has a bound core. A partially bound cluster has to have a valid Jacobi radius alongside a minimum enclosed mass of $M_J \geq 40 M_\odot$. If a cluster has $M_{\text{obs}}(r) < M_J(r)$ for all radii r , then no valid Jacobi radius can be determined and the disk stream is classified as entirely unbound. Four of the twelve disk streams meet this boundedness criterion: Platais 9, NGC 2451A, Mamajek 2, and OCSN 3. Table 2 lists their respective Jacobi radii and masses alongside the fraction of bound mass inside the disk stream. As first reported by MAR21, we also find that most mass, on average around 65%, of these four groups does not reside in their bound core but rather in the tails or coronae of these disk streams. The remaining disk streams – Ratzenboeck 1, Theia 386, Theia 430, HSC 2278, Theia 371/OCSN 87, Theia 301, Volans-Carina, and Theia 599 – appear to be fully unbound.

Lastly, we observe a small negative correlation between the age of the stream and its 3D velocity dispersion. This result is unexpected, as we anticipated an increase in velocity dispersion with age due to processes such as disk heating and GMC encounters. Similarly, the lack of correlation between a stream’s age and its length challenges the assumption that tidal tails grow uniformly over time. These findings are presumably explained by the diversity in the initial properties of the clusters, such as stellar density and mass, alongside the time since disruption and the specifics of the disruption process itself, all of which strongly influence the streams’ present day phase space distribution and survival times. Furthermore, the observed lack of correlation might also suggest limitations in our pipeline’s sensitivity to distinguish the low-density, highly dissolved tails from the surrounding field. These sources could potentially be recovered more effectively by clustering in a different feature space, such as action-angle coordinates (see, e.g., Fürnkranz et al. 2024), or through higher-resolution measurements from upcoming *Gaia* data releases.

4.6. Comparison with established cluster catalogs

Most stellar structures we identify can be crossmatched with literature samples. Table C.1 lists the 12 disk streams and provides an overview of the literature matches we find. Our search box also contains two prominent young open star clusters, NGC 2451A and Platais 9, around which we find substantial coronae, previously studied in detail by MAR21. Our pipeline reproduces their results in cluster morphology and approximate size.

An in-depth comparison is provided in Appendix C, where we also compare our results across multiple literature catalogs². Here in the main text, we focus on the cluster catalog of Hunt & Reffert (2023) (hereafter, HR23), which represents an unsupervised and homogeneous state-of-the-art reference that exhibits a high level of similarity in cluster morphology compared to our selection. Out of the 12 disk streams identified, 9 groups have clear counterparts in HR23. On average, we identify twice as many sources compared to these counterparts.

When comparing our results across cluster catalogs in the literature, it appears that disk stream S1 has not yet been identified as such in previous research. Our selection of S1 shows minimal overlap with HR23, who detected two small fragments, each constituting approximately 5 and 7% of S1 (see Fig. 5, top panel). Hence (as briefly mentioned in Sect. 3.2), we have named the population “Ratzenboeck 1”.

Comparisons with other catalogs are more challenging due to sometimes differing cluster morphologies in the literature, which complicates direct comparisons. Figure 5 illustrates these challenges. The middle and bottom panels show comparisons with Kounkel & Covey (2019) and Fürnkranz et al. (2024), respectively. While some cluster sources align, the overall cluster distributions can appear significantly different.

Lastly, we briefly highlight the potential relationship between several well-known clusters and associations and our disk stream. First, our analysis indicates a possible overlap between the Coma-Berenices cluster (also known as Melotte 111; using the selection by Fürnkranz et al. (2019) who identify its tidal tails) and the group HSC 2278, suggesting that HSC 2278 is the trailing (tidal) tail of the Coma-Ber cluster.

Second, we find a potential relationship between our selection of Theia 301 (S9) and the AB Doradus moving group (AB Dor; Zuckerman et al. 2004) when comparing it to the AB Dor selection of Gagné et al. (2018a). As proposed by Gagné et al. (2021), AB Dor and Theia 301 may be parts of the same system (alongside the Pleiades and other Theia members). In both cases, the analysis of both groups’ space motion and observational HRDs reveals neither definitive support nor a contradiction to the claim of a single-coeval population, and further data are required to generate a definitive answer.

Third, we find 47 crossmatches between the X-ray-selected Sco-Cen members of Schmitt et al. (2022) and our disk streams. We investigated these candidates in the observational HRD and find that these sources are clearly separated from the young pre-main sequence stars of Sco-Cen, pointing toward a potential source of contamination in X-ray-only selections of young sources.

² Cluster papers not mentioned in this section or Table C.1 do not cross-match to the identified disk streams or are already comprehensively discussed by HR23; we refer the reader to this work for further in-depth matches with catalogs not covered here.

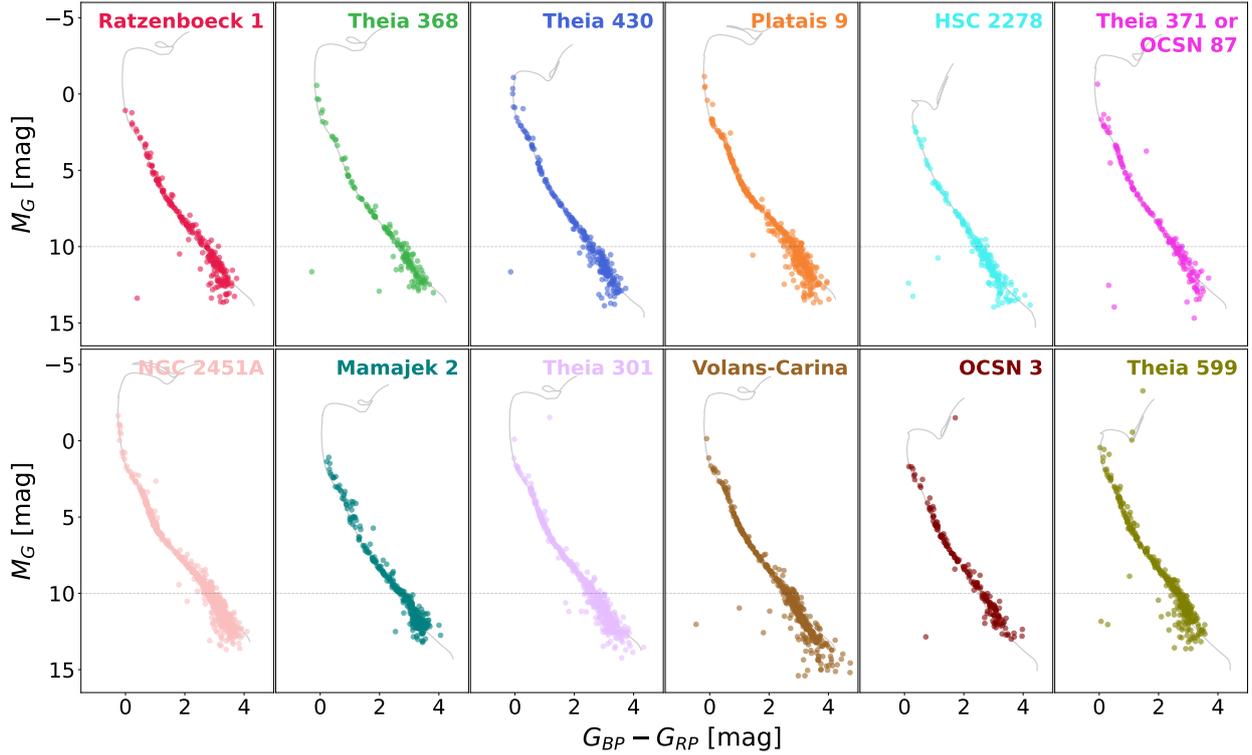


Fig. 4. Observational HRDs for the 12 identified disk streams. Sources that do not satisfy the astrometric quality criterion $\text{RUWE} < 1.4$ (Lindgren et al. 2021) or have large photometric uncertainties ($G_{\text{err}} < 0.007$ mag; $G_{\text{RP, err}} < 0.03$ mag; $G_{\text{BP, err}} < 0.15$ mag) have been removed to reduce large random scatters due to bad measurements. Colored points represent each cluster’s members. The corresponding best-fitting isochrones are shown as gray lines; their respective ages can be found in Table 1. The dashed horizontal line indicates the fitting range limit. Sources fainter than absolute G magnitudes of 10 are removed from the fit due to empirical discrepancies between isochronal curves and data points (see Ratzenböck et al. 2023b).

5. Discussion

Recent *Gaia* data releases have enabled researchers to discover several stream-like structures in the local Milky Way. Beyond the fiducial disk stream Meingast-1, tidal tails have been detected around open clusters (see Meingast & Alves 2019; Röser et al. 2019) and are now a ubiquitous feature around open clusters, reaching almost 100 detections (Tarricq et al. 2022). More recently, cluster coronae (see Meingast et al. 2021; Moranta et al. 2022) have been detected as a loose coeval ensemble of stars surrounding dense cluster cores that are likely also produced in tidal disruption processes. However, it is not clear at the moment the impact of the original gas distribution or the relative roles of residual gas expulsion and violent relaxation (see MAR21). Beyond that, further low-density and elongated structures have been identified, such as stellar snakes (Tian 2020), filamentary structures (Beccari et al. 2019), and the stellar disk stream Theia 456 (Kounkel & Covey 2019; Andrews et al. 2022).

We note that the physical difference between these structures is not clear (yet), and various names might be used to refer to similar dissolution processes. This study does not focus on the origins of disk streams or similar structures, and we call for a more comprehensive sample for an in-depth analysis of their origin. Nevertheless, our preliminary findings warrant a brief discussion. Akin to Meingast-1 and cluster coronae (see Meingast et al. 2019 and MAR21), the populations identified in this work show similar patterns, such as large elongations > 100 pc, and inclination angles in the X - Y plane reminiscent of Galactic tidal interactions. Tidal disruption processes are a plausible explanation for most populations that appear to have a

symmetrical leading and trailing arm oriented toward and away from the Galactic center.

5.1. Peculiar phase space signatures

Notably, velocity dispersions in Galactic Cartesian velocity space below $\sim 5 \text{ km s}^{-1}$ are exceptionally low for structures extending over several hundred parsecs. In this space, Galactic rotation has a contribution to the total velocity dispersion for such large structures, which ranges from $1\text{--}3 \text{ km s}^{-1}$. Appendix D discusses this issue in detail, while we summarize two major points in the following. One potential explanation for these low values is the nature of the deconvolution process. To provide a more robust estimate and minimize the potential bias introduced by the deconvolution method, we also computed velocity dispersions using the MAD (see Sect. 4.2). This robust measure yields slightly higher values, with velocity dispersions ranging from $2.7\text{--}7.9 \text{ km s}^{-1}$ (see Table 1). Another likely source of underestimation arises from selection effects inherent in density-based clustering methods. Given the large extent of the identified structures, the recovered members are embedded within a significant background population. Members in the tails of the velocity distribution, just a few km s^{-1} from the central overdensity (in any coordinate system), lack sufficient contrast relative to the dominant background (i.e., a S/N of ~ 1). As a result, these members often go undetected by clustering algorithms. Addressing this limitation by identifying and recovering the “missing” members with higher velocity dispersions is outside the scope of this work but represents a key direction for future work.

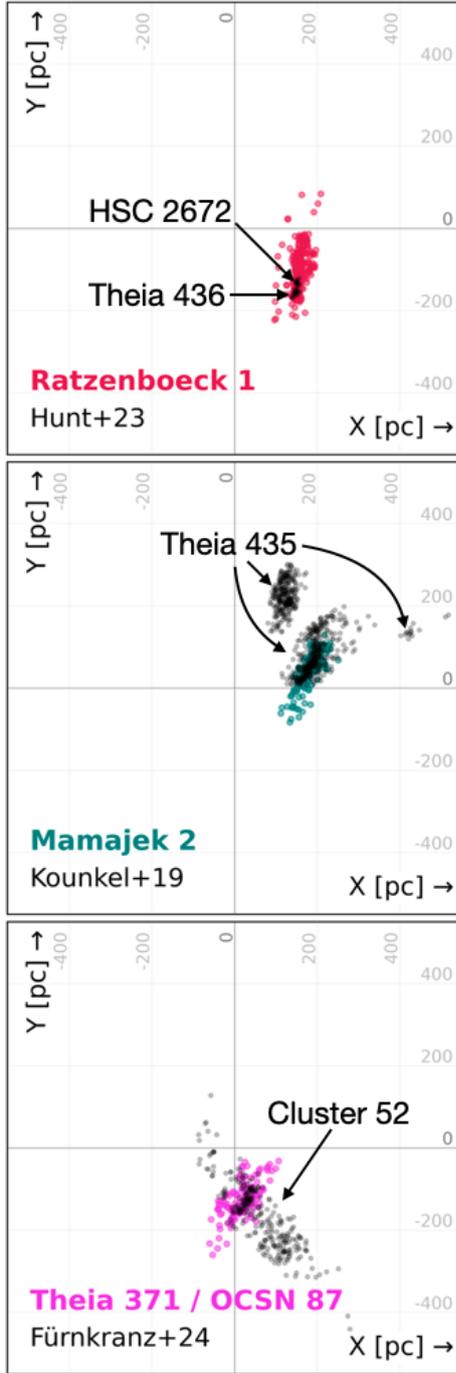


Fig. 5. Examples of challenging cluster comparisons highlighted in the X - Y plane. The colored scatter points show the identified disk streams Ratzenboeck 1, Mamajek 2, and Theia 371/OCSN 87 (from top to bottom). The black scatter points show the crossmatch to a literature cluster in the HR23, KC19, and Fürnkranz et al. (2024) surveys (from top to bottom). See Sect. 4.6, Appendix C, and Table C.1 for further details.

Beyond velocity space, the positional distribution raises an important question regarding the shared origins of several structures. As discussed in Sect. 4, several disk streams exhibit significant overlaps in position space. However, a detailed kinematic analysis reveals that, in all but one case, these co-spatial streams show statistically significant differences in their 3D velocities and ages. In the case of Ratzenboeck 1 and Theia 386, we find similar 3D velocities and ages, which suggests

they represent fragments of a larger structure or a joint formation scenario akin to substructures identified in various OB associations (e.g., Damiani et al. 2019; Chen et al. 2020; Kerr et al. 2021; Ratzenböck et al. 2023a). A traceback analysis of their Galactic orbits in the past and future 20 Myr alongside a distinct bi-modal signal in their joint phase space distribution (see Appendix D.3) supports their classification as separate clusters. For a comprehensive discussion of the pairwise kinematic comparisons, including the Mahalanobis distances quantifying these differences and the implications for stream independence (see Appendix D.3). We also refer readers to Fürnkranz et al. (2019) for an earlier analysis of co-spatial populations and their potential connections.

5.2. Estimating the disk stream number density

Using detailed N -body simulations (see Kamdar et al. 2019 and Kamdar et al. 2021³, hereafter HCT21) recently estimated the number of disk streams we can find considering the capabilities of *Gaia* DR2. They considered two major factors that restrict their detection. First, limited accuracy of astrometric measurements, which restricts the identification of populations with low-density contrast (against the background). And second, destructive encounters with GMCs. The authors tested the impact of different initial conditions in which the progenitor star cluster is born, such as initial cluster mass and dynamical state. They conclude that with the astrometric precision of *Gaia* DR2, 1 to 10 disk streams are likely yet to be found in the solar neighborhood (defined by the authors as a 500 pc radius sphere centered on the Sun). This number, the authors argue, will improve by a factor of 5–10 with *Gaia* 10-year end-of-mission data where high-fidelity parallaxes allow an increase in the effective search volume out to a radius of 1.5 kpc. These N -body estimates suggest that *Gaia* data are and will continue to be sensitive enough to reach volume densities of around 2 to 20 disk streams per kpc^3 .

As discussed above, many elongated, stream-like structures have been identified in the Milky Way disk. In this work, we aim to provide the first empirical estimate of the abundance and, more specifically, the number density of dynamically cold, coeval stream-like structures in our Galaxy. Taking a volume-controlled sample, we observe an abundance of stream-like structures within our local test volume of 250^3pc^3 . For this volume, which samples $|Z| < 100$, we derive a density of approximately 820 disk streams per kpc^3 . This estimate is one to two orders of magnitude higher than predicted by N -body simulations. The equivalent surface density estimate on the Galactic X - Y plane is around 160 objects per kpc^2 . The co-occurrence of 12 stream-like structures within such a confined region creates tension with the conventional understanding of these structures, which are believed to be heavily suppressed by destructive interactions with one or a few GMCs (e.g., Gieles et al. 2006; Kamdar et al. 2021).

To determine whether the over-classification of disk streams has artificially inflated the estimated disk stream density, we evaluated the impact of refining the classification criteria on the density estimates. Specifically, we excluded known open clusters with distinct coronae, such as Platais 9, NGC 2451A, Mamajek 2,

³ We note that this particular research article has not yet been officially published. However, its results build on disk streams arising from N -body simulations of the Galaxy (see Kamdar et al. 2019) that are independent of this particular research and have been successfully peer-reviewed and published.

and OCSN3. These clusters feature bound cores, as indicated by our computations (see Sect. 4.4). However, even after these exclusions, the volume density of the remaining disk streams remains an order of magnitude higher than the most optimistic predictions from N -body simulations. This significant discrepancy warrants further discussion on its origin and points either toward a more efficient production mechanism of low-density stream-like disk populations or a less efficient destruction mechanism with a lower disruption time via, for example, GMC interactions (see, e.g., Kruijssen et al. 2011; Krumholz et al. 2019; Kamdar et al. 2021). Whichever mechanism dominates, leading to this increased disk stream density, is beyond the scope of this discovery work but will be the subject of future study.

We acknowledge that the proximity of our search box also plays a key role in ensuring data quality, as the precision of astrometric measurements and the resulting phase-space resolution decrease significantly at larger distances. The selected box lies within a region of the Milky Way disk that we believe is representative of typical stellar populations, as supported by its relatively large age spread that approximately matches results from all-sky cluster searches (e.g., Hunt & Reffert 2023, 2024). While this region provides high-quality data, we anticipate that similar structures and, thus, similar disk stream densities will be detected in other regions of the Milky Way disk.

In future work, we plan to extend this analysis by repositioning the box to various regions of the Galactic disk. This will increase the currently limited statistical sample of disk streams, allowing us to assess whether our local measurements are typical and to better understand the processes underlying the formation and survival of these structures. Having access to a larger sample of disk streams will also open the door to deriving more detailed constraints on cluster disruption times along GMC properties, such as their mass function, number density, and overall formation and evolution in the Milky Way.

5.3. Cluster dissolution

HCT21 find four main predictors for star clusters to eventually become detectable streams. Those are (a) large initial cluster masses, (b) “young” ages below 1 Gyr, (c) the number and severity of GMC interactions, and (d) an initial dynamical state that is preferentially bound, having initial virial ratios of less than one half. While a higher initial cluster mass, a younger age, and less disruption from GMCs seem like straightforward predictors, the authors also argue about the importance of initial boundedness, which enhances a stream’s chance of being detected in the present day.

Here, we aim to investigate the role of (partial) boundedness in our sample of 12 disk streams in the context of their age. Using the dichotomy of “fully unbound” and “has a bound core” as introduced in Sect. 4.4, we searched for any differences in the age distribution between these classes. Figure 6 presents the results of this analysis, using KDE⁴ to visualize the age distribution of two distinct subpopulations: clusters with evidence of a bound core (blue) and those that are fully unbound (orange). The individual data points, marked along the x -axis, further illustrate the age distribution within each subpopulation.

Our analysis reveals that fully unbound clusters tend to be, on average, approximately 100 Myr older than their counterparts with a bound core. Additionally, all but one disk streams beyond 200 Myr are unbound, except OCSN 3. This observation

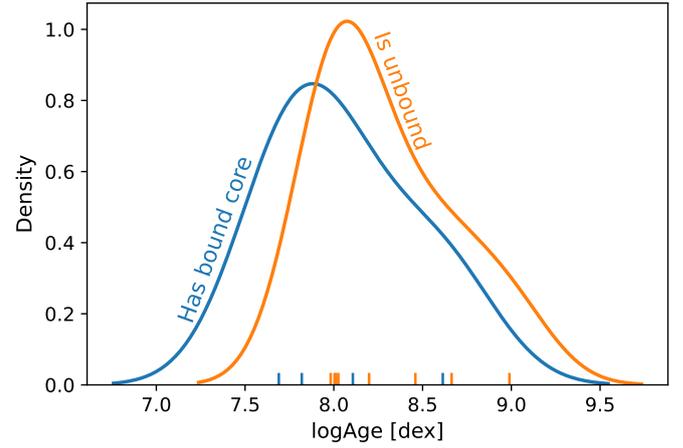


Fig. 6. KDE (colored lines) of the age distribution along individual data points (vertical marks on the x -axis), stratified by cluster boundedness: those with evidence of a bound core (blue) and those that are fully unbound (orange). The unbound clusters have a slightly older average age of approximately 100 Myr. Despite this observed difference, the limited sample size prevents us from drawing statistically significant conclusions regarding the age distributions of the two subpopulations.

points to a possible link between cluster age and the likelihood of core dissolution, suggesting that some older (unbound) disk streams we see in our sample initially come from stellar systems that previously had a bound core. This bound core also brings higher survivability of disk streams into old age, as found in N -body simulations, as it makes them more resilient to disruption processes, such as GMC collisions and other Galactic potential variations. At some point, these disruption processes critically accumulate, resulting in fully unbound systems, which are then rapidly dissolved by future GMC interactions and the Galactic tidal field, explaining the low density of disk streams at ages beyond 200 Myr.

However, it is important to note that the limited sample size prevents us from making statistically significant conclusions regarding the differences in age distributions between the two subpopulations. While the trend is intriguing, more data are needed to robustly determine whether a significant age difference exists between stream-like stellar systems with a bound core and fully unbound disk streams. Another caveat is our simple model of cluster dissolution, ignoring different evolutionary phases and subclasses of stream-like structures such as tidal tails, young moving groups, young local associations, and cluster coronae. In future papers of this series, we aim to work toward a statistically significant sample of disk streams to facilitate the analysis of their different and complex formation and dissolution processes, from Galactic tidal forces, differential rotation, initial star-forming gas configurations, and GMC interactions.

5.4. Relationship with Sco-Cen

The initial sample selection (see Sect. 2) entails that identified disk streams are close to the Sco-Cen OB association. Spatially, the disk streams are tightly packed and can be divided into two groups based on their relative position to Sco-Cen. The disk streams Ratzenboeck 1, Theia 368, Mamajek 2, and OSCN 3 are situated on the far side of Sco-Cen, while the remaining disk streams fill the space between the Sun and Sco-Cen. Remarkably, Ratzenboeck 1, Theia 368, and Mamajek 2 share the same volume to various degrees, with stars associated with the Sco-Cen association. Ratzenboeck 1 appears to “flow” through

⁴ We estimate the bandwidth of the Gaussian kernel using Scott’s rule (Scott 1979).

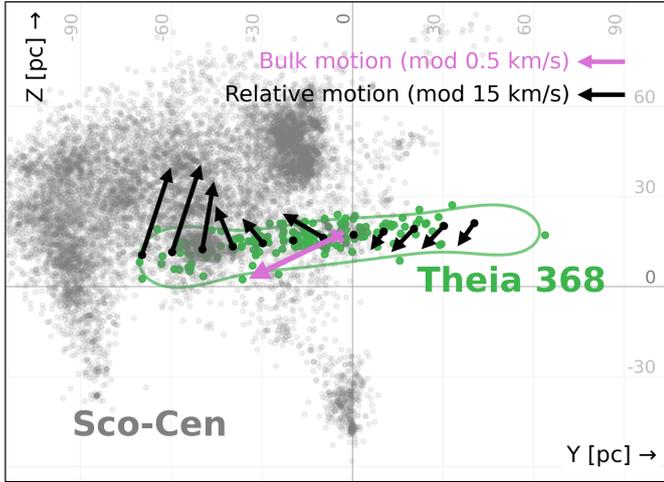


Fig. 7. Evidence suggesting an interaction between disk stream Theia 368 (in green) and Sco-Cen (in gray). The black arrows show a running median of the stream’s motion relative to its bulk motion, depicted by the purple arrow. The relative motion is scaled up by a factor of 30 compared to the bulk motion to highlight their distribution properly; see the legend on the top left for a size comparison. The relative motions across Theia 368 highly correlate with the stream’s interaction time with Sco-Cen; i.e., the further inside Theia 368’s members are found in Sco-Cen, the more drastically their motions have been altered.

the Sco-Cen subgroups V1062 Sco, and μ Sco while partially overlapping with the Cen-Far group (see Paper I for Sco-Cen subcluster names). The disk stream Theia 368 lies almost entirely inside the Sco-Cen subgroups of η Lup, Sco Body, and θ Oph. Lastly, although the bulk of Mamajek 2 is outside Sco-Cen, several sources extend into the classical Blaauw subgroup Upper Scorpius.

Sco-Cen may have exerted measurable forces on the disk stream Theia 368, which is fully embedded within Sco-Cen. Given that Sco-Cen is located at the edge of the Local Bubble (Zucker et al. 2022a), which contributes additional gas mass to Sco-Cen’s progenitor cloud, there has been sufficient primordial gas mass present to directly and cumulatively influence the disk stream over the past few million years. To determine if this interaction has left measurable effects in present-day observables, we investigated potential imprints in the velocity distribution of Theia 368 members along the length of the stream.

Specifically, we aim to compare the velocity distribution of sources currently embedded within the Sco-Cen association to those currently outside. Figure 7 depicts the potential interaction with Sco-Cen (its members are shown as gray points in the background) that affects the relative 3D velocities observed. The black arrows indicate the motion relative to the stream’s bulk motion (purple arrow; see also Table 1) across the entire stream (represented in green). We examined the Y - Z velocity space to detect this interaction, as it is less influenced by large radial velocities that primarily align with the X -axis. To further limit the influence from outliers in UVW on the relative velocity signal, we removed sources with radial velocity uncertainties larger than 5 km s^{-1} and sources flagged as potential outliers by the extreme deconvolution (XD) process (see Appendix D). This quality cut results in a total of 33 sources with “good” radial velocity measurements. Lastly, we computed a running median (across five neighbors) at 12 grid points along the Y -axis spaced

10 pc apart (black scatter points in Fig. 7) to reduce the influence of the remaining random scatter⁵.

Sources located outside Sco-Cen (with $Y > 0$ pc) exhibit nearly constant relative motions that are approximately parallel to the bulk motion (purple arrow in Fig. 7). In contrast, sources “entering” Sco-Cen show relative velocity vectors that appear to be deflected or scattered by their interaction with the OB association. Moreover, this deflection appears to correlate with the “depth” of Theia 368 within Sco-Cen, which gradually increases. These observations provide qualitative and tentative evidence of a scattering process and/or exerted force on parts of the disk stream Theia 368. To substantiate these tentative results, we aim to further investigate this interaction in future work by taking into account a traceback analysis of the unaffected and (apparent) deflected portion of Theia 368 in relation to Sco-Cen, considering the available gas mass and performing a detailed momentum analysis, and contrasting these claims with simpler models that involve only the Galactic potential.

In future work, we aim to investigate similar effects on a larger sample of disk streams, in particular, where our analysis has not yet revealed a clear signal.

6. Conclusion

We refined the initial selection of a sample comprising 12 stream-like stellar populations, previously identified as interlopers in Paper I, by using the established machine learning pipelines SigMA and Uncover to improve their census. One disk stream had not been identified in the previous literature, and we have named it Ratzenboeck 1. Compared to previous unsupervised studies, we find, on average, twice as many candidate members per stream. Our pipeline is sensitive to median stream densities (stream densities are averages throughout the volume of the population) of one star per 10^3 pc^3 (0.001 stars/pc^3), which is about 50 times lower than the surrounding field. At the very extreme, we can recover streams with average densities as low as 0.2 sources per 10^3 pc^3 ($0.0002 \text{ stars/pc}^3$) across the entire population (i.e., resolving structures 250 times below the field density, in XYZ).

1. The 12 disk streams found within the 250^3 pc^3 volume yield an estimated volume density of approximately $820 \text{ objects/kpc}^3$ and a surface density of roughly $160 \text{ objects/kpc}^2$. These estimates exceed previous number density calculations by one to two orders of magnitude, as documented by HCT21.
2. We find evidence that the disk stream Theia 368 (S2), predominantly embedded within Sco-Cen, has recently undergone disruption, likely due to interactions with the primordial gas of the OB association.
3. These 12 disk streams are highly prolate and have lengths between 120 and 430 pc, with large aspect ratios (longest principal axis over shortest axis) of 3–11.
4. The identified disk streams are dynamically cold, having low 3D velocity dispersions ranging from $2 - 5 \text{ km s}^{-1}$, and show clearly defined and narrow sequences in the HRD, strongly suggesting a coeval nature. Stream ages range from 50 Myr to 1 Gyr, with a median age of around 100 Myr.
5. We find that disk streams with bound cores are typically younger than fully unbound ones, with fully unbound systems being, on average, about 100 Myr older. Beyond

⁵ We find that binning the data along the Y -axis produces a similar pattern. However, some bins contain only one or two members, leading to a significantly greater scatter than a running median.

200 Myr, most disk streams are fully unbound, likely reflecting the cumulative effects of disruptive processes such as GMC interactions and Galactic tidal forces.

- We have identified an approximately linear relationship between stream ages and their respective volumes, which increase by around 500–1000 pc³/Myr.

Much like their halo counterparts, disk streams can serve as critical probes for understanding the mass distribution of the Galaxy on both large and small scales. The prevalence of these stream-like features within such a confined region represents a significant departure from conventional understanding, calling for a revision of the formation and dissolution scenarios and for a larger systematic census of disk streams.

Data availability

The full source catalog described in Tables 1 and F.1 is available at the CDS via anonymous ftp to cdsarc.cds.unistra.fr (130.79.128.5). Figure E.5 and additional figures are available online via Zenodo: <https://doi.org/10.5281/zenodo.14278685>

Acknowledgements. S.R. acknowledges funding by the Federal Ministry Republic of Austria for Climate Action, Environment, Energy, Mobility, Innovation and Technology (BMK, <https://www.bmk.gv.at/>) and the Austrian Research Promotion Agency (FFG, <https://www.ffg.at/>) under project number FO999892674. Co-funded by the European Union (ERC, ISM-FLOW, 101055318). Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. 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. This research has used Python, <https://www.python.org>; *Astropy*, a community-developed core Python package for Astronomy (*Astropy Collaboration* 2013, 2018); NumPy (van der Walt et al. 2011); Matplotlib (Hunter 2007); and Plotly (Plotly Technologies Inc. 2015). This research has made use of the SIMBAD database operated at CDS, Strasbourg, France (Wenger et al. 2000); of the Vizier catalog access tool, CDS, Strasbourg, France (Ochsenbein et al. 2000); and of “Aladin sky atlas” developed at CDS, Strasbourg Observatory, France (Bonnarel et al. 2000; Boch & Fernique 2014). This research has used TOPCAT, an interactive graphical viewer and editor for tabular data (Taylor 2005).

References

- Andrews, J. J., Curtis, J. L., Chanamé, J., et al. 2022, *AJ*, 163, 275
 Astropy Collaboration (Robitaille, T. P., et al.) 2013, *A&A*, 558, A33
 Astropy Collaboration (Price-Whelan, A. M., et al.) 2018, *AJ*, 156, 123
 Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, *AJ*, 161, 147
 Beccari, G., Boffin, H. M. J., & Jerabkova, T. 2019, *MNRAS*, 491, 2205
 Beccari, G., Boffin, H. M. J., & Jerabkova, T. 2020, *MNRAS*, 491, 2205
 Bennett, M., & Bovy, J. 2019, *MNRAS*, 482, 1417
 Boch, T., & Fernique, P. 2014, *ASP Conf. Ser.*, 485, 277
 Bonnarel, F., Fernique, P., Bienaymé, O., et al. 2000, *A&AS*, 143, 33
 Bossini, D., Vallenari, A., Bragaglia, A., et al. 2019, *A&A*, 623, A108
 Bovy, J. 2015, *ApJS*, 216, 29
 Bovy, J. 2016, *Phys. Rev. Lett.*, 116, 121301
 Bovy, J., Hogg, D. W., & Roweis, S. T. 2011, *Annal. Appl. Stat.*, 5, 1657
 Cantat-Gaudin, T., & Anders, F. 2020, *A&A*, 633, A99
 Chen, B., D’Onghia, E., Alves, J., & Adamo, A. 2020, *A&A*, 643, A114
 Curtis, J. L., Agüeros, M. A., Mamajek, E. E., Wright, J. T., & Cummings, J. D. 2019, *AJ*, 158, 77
 Damiani, F., Prisinzano, L., Pillitteri, I., Micela, G., & Sciortino, S. 2019, *A&A*, 623, A112
 Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 39, 1
 Eggen, O. J. 1996, *AJ*, 112, 1595
 Ernst, A., Just, A., Berczik, P., & Petrov, M. I. 2010, *A&A*, 524, A62

- Fisher, R. A. 1934, in *Breakthroughs in Statistics* (Berlin: Springer), 66
 Fürnkranz, V., Meingast, S., & Alves, J. 2019, *A&A*, 624, L11
 Fürnkranz, V., Rix, H.-W., Coronado, J., & Seeburger, R. 2024, *ApJ*, 961, 113
 Gagné, J., & Faherty, J. K. 2018, *ApJ*, 862, 138
 Gagné, J., Mamajek, E. E., Malo, L., et al. 2018a, *ApJ*, 856, 23
 Gagné, J., Faherty, J. K., & Mamajek, E. E. 2018b, *ApJ*, 865, 136
 Gagné, J., Faherty, J. K., Moranta, L., & Popinchalk, M. 2021, *ApJ*, 915, L29
 Gaia Collaboration (Vallenari, A., et al.) 2023, *A&A*, 674, A1
 Gieles, M., Zwart, S. F. P., Baumgardt, H., et al. 2006, *MNRAS*, 371, 793
 Golovin, A., Reffert, S., Vani, A., et al. 2024, *A&A*, 683, A33
 Grasser, N., Ratzenböck, S., Alves, J., et al. 2021, *A&A*, 652, A2
 GRAVITY Collaboration (Abuter, R., et al.) 2018, *A&A*, 615, L15
 Grillmair, C. J., Freeman, K. C., Irwin, M., & Quinn, P. J. 1995, *AJ*, 109, 2553
 Hampel, F., Ronchetti, E., Rousseeuw, P., & Stahel, W. 2011, *Wiley Series in Probability and Statistics* (Hoboken: John Wiley & Sons, Inc.), 503
 Hawkins, K., Lucey, M., & Curtis, J. 2020, *MNRAS*, 496, 2422
 Hunt, E. L., & Reffert, S. 2023, *A&A*, 673, A114
 Hunt, E. L., & Reffert, S. 2024, *A&A*, 686, A42
 Hunter, J. D. 2007, *Comput. Sci. Eng.*, 9, 90
 Ibatova, R. A., Lewis, G. F., & Martin, N. F. 2016, *ApJ*, 819, 1
 Jerabkova, T., Boffin, H. M. J., Beccari, G., et al. 2021, *A&A*, 647, A137
 Kamdar, H., Conroy, C., Ting, Y.-S., et al. 2019, *ApJ*, 884, 173
 Kamdar, H., Conroy, C., & Ting, Y.-S. 2021, arXiv e-prints [arXiv:2106.02050]
 Kerr, R. M. P., Rizzuto, A. C., Kraus, A. L., & Offner, S. S. R. 2021, *ApJ*, 917, 23
 King, I. 1962, *AJ*, 67, 471
 Kounkel, M., & Covey, K. 2019, *AJ*, 158, 122
 Kroupa, P. 2001, *MNRAS*, 322, 231
 Kroupa, P., Jerabkova, T., Thies, I., et al. 2022, *MNRAS*, 517, 3613
 Kruijssen, J. M. D., Pelupessy, F. I., Lamers, H. J. G. L. M., Portegies Zwart, S. F., & Icke, V. 2011, *MNRAS*, 414, 1339
 Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, *ARA&A*, 57, 227
 Levene, H. 1960, *Robust Tests for Equality of Variances* (Palo Alto: Stanford University Press)
 Lindegren, L., Klioner, S. A., Hernández, J., et al. 2021, *A&A*, 649, A2
 Malhan, K., Ibatova, R. A., & Martin, N. F. 2018, *MNRAS*, 481, 3442
 Mamajek, E. E. 2006, *AJ*, 132, 2198
 Marigo, P., Girardi, L., Bressan, A., et al. 2017, *ApJ*, 835, 77
 Mateu, C. 2023, *MNRAS*, 520, 5225
 Meingast, S., & Alves, J. 2019, *A&A*, 621, L3
 Meingast, S., Alves, J., & Fürnkranz, V. 2019, *A&A*, 622, L13
 Meingast, S., Alves, J., & Rottensteiner, A. 2021, *A&A*, 645, A84
 Miret-Roig, N., Galli, P. A. B., Brandner, W., et al. 2020, *A&A*, 642, A179
 Moranta, L., Gagné, J., Couture, D., & Faherty, J. K. 2022, *ApJ*, 939, 94
 Newton, E. R., Mann, A. W., Kraus, A. L., et al. 2021, *AJS*, 161, 65
 Ochsenbein, F., Bauer, P., & Marcout, J. 2000, *A&AS*, 143, 23
 Odenkirchen, M., Grebel, E. K., Rockosi, C. M., et al. 2001, *ApJ*, 548, L165
 Planck Collaboration XI. 2014, *A&A*, 571, A11
 Plotly Technologies Inc. 2015, Collaborative data science, Montreal, QC, <https://plot.ly>
 Portegies Zwart, S. F., McMillan, S. L., & Gieles, M. 2010, *Ann. Rev. Astron. Astrophys.*, 48, 431
 Price-Whelan, A. M., & Bonaca, A. 2018, *ApJ*, 863, L20
 Qin, S., Zhong, J., Tang, T., & Chen, L. 2023, *ApJS*, 265, 12
 Ratzenböck, S., Meingast, S., Alves, J., Möller, T., & Bomze, I. 2020, *A&A*, 639, A64
 Ratzenböck, S., Großschedl, J. E., Möller, T., et al. 2023a, *A&A*, 677, A59
 Ratzenböck, S., Großschedl, J. E., Alves, J., et al. 2023b, *A&A*, 678, A71
 Ratzenböck, S., Obermüller, V., Möller, T., Alves, J. a., & Bomze, I. M. 2023c, *IEEE Trans. Visualiz. Comp. Graph.*, 29, 3855
 Riello, M., De Angeli, F., Evans, D. W., et al. 2021, *A&A*, 649, A3
 Röser, S., Schilbach, E., & Goldman, B. 2019, *A&A*, 621, L2
 Schmitt, J. H. M. M., Czesla, S., Freund, S., Robrade, J., & Schneider, P. C. 2022, *A&A*, 661, A40
 Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., & Platt, J. 1999, in *Proceedings of the 12th International Conference on Neural Information Processing Systems, NIPS’99* (Cambridge, MA, USA: MIT Press), 582
 Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, 403, 1829
 Scott, D. W. 1979, *Biometrika*, 66, 605
 Tarricq, Y., Soubiran, C., Casamiquela, L., et al. 2022, *A&A*, 659, A59
 Taylor, M. B. 2005, *ASP Conf. Ser.*, 347, 29
 Tian, H.-J. 2020, *Astrophys. J.*, 904, 196
 van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Comp. Sci. Eng.*, 13, 22
 Wenger, M., Ochsenbein, F., Egret, D., et al. 2000, *A&AS*, 143, 9
 Wilson, D. J. 2019, *Proc. Natl. Acad. Sci.*, 116, 1195
 Zucker, C., Goodman, A. A., Alves, J., et al. 2022a, *Nature*, 601, 334
 Zucker, C., Peek, J. E. G., & Loebman, S. 2022b, *ApJ*, 936, 160
 Zuckerman, B., Song, I., & Bessell, M. S. 2004, *ApJ*, 613, L65

Appendix A: SigMA parameter selection

The SigMA pipeline requires the selection of multiple parameters, each contributing to a unique output for a given set of input parameters. Generally, we adopted the parameter choices outlined in Paper I, except for the scaling factor values, which need to be adjusted due to the change in input space from 5D to 6D phase space. As explained in Paper I (see Sect. 3.3.3), the purpose of the scaling factors is to normalize the different data ranges across various subspaces. In our study, the input data space comprises three positional axes (*XYZ*) and three velocity axes (*UVW*). Within each subspace, the Euclidean norm effectively represents the similarity between sources, such as the distance between stars in parsecs or the velocity difference in km s^{-1} . Generally, our goal would be to normalize one subspace relative to another (see Paper I) based on the characteristic dispersion of the objects we intend to cluster within each subspace. However, stellar streams are notably elongated and exhibit significantly different extents along their three principal axes in *XYZ* space (see Table 1).

Instead of using a single value or a small range of values for the scale factor, we aim to explore a broad space of theoretically meaningful scaling factor values and track a single cluster through this series of SigMA runs. We used the progenitor streams as a reference to determine this range. For each progenitor stream, we calculated the three eigenvalues of its covariance matrix by performing principal component analysis in both positional and kinematic subspaces. Using these six dispersion coefficients (three for *XYZ* and three for *UVW*), we generated nine scaling factors by considering all possible pairs of these coefficients. We repeated this procedure for all 12 disk streams, generating a distribution of possible scale factors. To exclude extreme outliers, we took this distribution's 5th and 95th percentiles and divided the resulting range into 20 equally spaced scaling factors⁶. Running SigMA 20 times yields an ensemble of clustering results in which a given progenitor stream is recovered slightly differently in each run. On average, a stream is identified as such in approximately 35% of the runs, predominantly for runs with lower scale factors below 10.

After automatically identifying the stream by crossmatching with the progenitor stream in each run, we retained a source as a stream member if it appears in at least 50% of the runs where the stream was successfully recovered. This clustering result is then used to train further the membership pipeline Uncover, as described in Sect. 3.2.

Appendix B: Result supplements

This section provides auxiliary information to our result section in Sect. 4. Figure B.1 shows the mass histograms for the 12 disk streams characterized in this work alongside the best fit Kroupa IMFs (Kroupa 2001) and the corresponding uncertainties determined via bootstrap samples 100 times (see Table 1 for an overview of all fit values).

Figure B.2 is a scatter plot matrix that highlights pairwise correlations among various physical parameters determined for the 12 disk streams. Specifically, Fig. B.2 shows the relationships between age, length, volume, mass, density, and 3D velocity dispersion for all disk streams (see Table 1 for an overview of the physical parameters).

⁶ This results in the following scale factors (rounded to the first decimal) with which we multiplied the velocity subspace axes: {1.5, 3.3, 5.0, 6.8, 8.6, 10.4, 12.1, 13.9, 15.7, 17.5, 19.2, 21.0, 22.8, 24.6, 26.3, 28.1, 29.9, 31.7, 33.4, 35.2}.

Appendix C: Literature comparison supplements

This section provides auxiliary information to our literature comparison presented in Sect. 4.6. Table C.1 overviews the 12 identified disk streams and their matches in the literature.

Our comparison of the literature reveals that we find the best agreement in stream morphology and approximate size compared to the work of Hunt & Reffert (2023, hereafter HR23). Seven of the 12 identified disk streams have a similar counterpart in HR23 (S2, S3, S4, S6, S7, S8, and S10). In addition, groups S5 and S11 also have a clear counterpart in HR23, albeit with a fairly smaller extent and source count. The streams S6 and S11, named OCSN 87 and OCSN 3, respectively, were first identified by Qin et al. (2023, hereafter QZTC23). QZTC23 also identify (and claim the discovery of) S2 (OCSN 99) and S10 (OCSN 88), which we found in earlier works as Theia 368 by Kounkel & Covey (2019) and Volans-Carina by Gagné et al. (2018b), respectively. Finally, HR23 has identified fragments of Ratzenboeck 1 (S1; see Fig. 5, top panel) and S12 but does not connect them to a larger structure. Our analysis of the sources' 3D velocity space suggests that these fragments likely correspond to the same stellar structure and should thus not be separated (see Ratzenböck et al. in prep). Disk stream S9 is not recovered by HR23.

The comparison to Kounkel & Covey (2019, hereafter KC19) also reveals many overlaps, although we find that many cross-matches appear to describe (to some degree) different stellar aggregates. We find a good alignment between disk streams S2 and Theia 368 (OCSN 99 in QZTC23), S3 and Theia 430 (HSC 2303 in HR23), and S10 and Theia 424 (Volans-Carina in Gagné et al. (2018b)). Comparisons between other disk streams and Theia groups do not yield a precise alignment; for example, Theia 371 (OCSN 87 in QZTC23 and HSC 2407 in HR23) corresponds to the core of S6; however, Theia 371 and S6 strongly disagree on the remaining extent of respective populations. Except for Theia 134 and Theia 508, which we find represent Platais 9 (S4), we find even more extreme mismatches (compared to Theia 371 and S6) where crossmatched groups correspond to a (sometimes vastly) different stellar population or exhibit (high) contamination (e.g., apparent in the color-magnitude diagram), already indicated by other authors (see, e.g., Zucker et al. 2022b). Figure 5 (middle panel) exemplifies one of these difficult comparisons, in this case, between Theia 435 and Mamajek 2 (S8). Disk streams Ratzenboeck 1 and HSC 2278 (S5) are not recovered by Paper KC19.

We find further crossmatches with the following literature catalogs: Fürnkranz et al. (2024) recover S2 (Theia 368 in KC19) and S6 (OCSN 87 in QZTC23), although with substantial contamination (see Fig. 5, bottom panel). Cantat-Gaudin & Anders (2020) and MAR21 both contain Platais 9 (our S4) and NGC 2451A (our S7). Compared to the census of Cantat-Gaudin & Anders (2020), we recover approximately three times the number of sources for Platais 9 and NGC 2451A. The comparison to MAR21 yields similar cluster morphologies but an average improvement in the number of identified sources between 30 and 50%. However, the improvement in cluster size is likely due to the improved astrometry of Gaia DR3 over DR2 and less stringent error cuts in our work. Finally, the disk stream S10 (OCSN 88 in QZTC23) corresponds to the Volans-Carina Association, first discovered by Gagné et al. (2018b), whose source population we have increased approximately tenfold (and roughly doubled against HR23). Moranta et al. (2022) also identify Volans-Carina alongside its corona, where they uncover a total of 141 high-fidelity members (here we find 566 potential members).

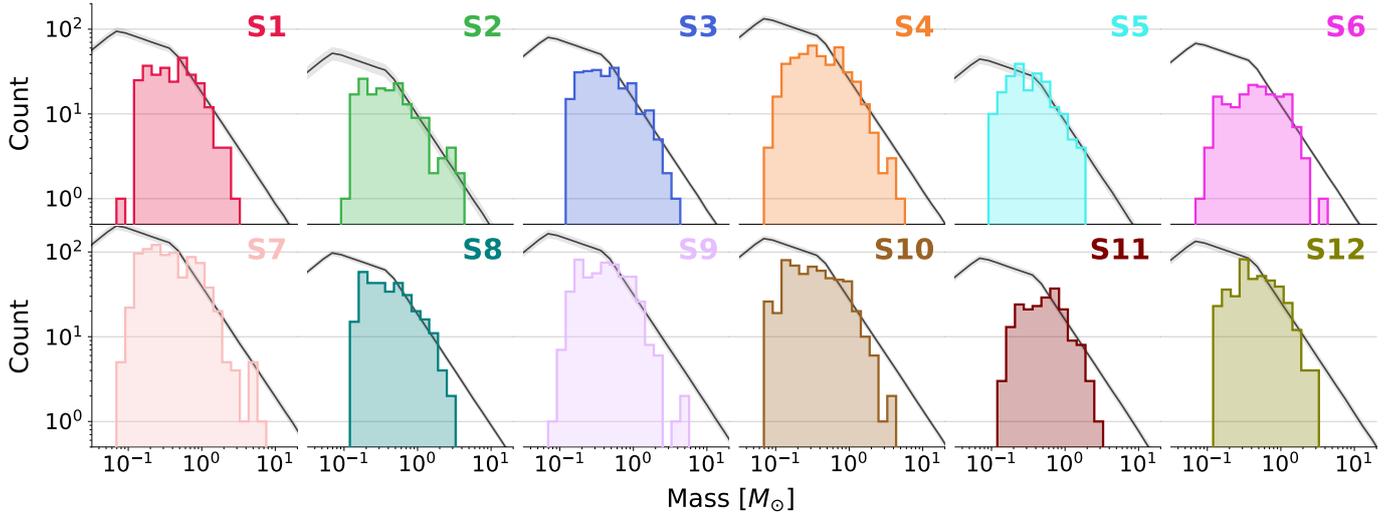


Fig. B.1. Mass functions for the 12 disk streams used to derive system masses for each population. The best-fitting Kroupa IMFs (solid black line; Kroupa 2001) and 1σ uncertainties (gray shaded area) are plotted on top of each histogram.

Appendix D: Velocity dispersion estimation

In this section we provide some additional details on the velocity dispersion estimation alongside a discussion on the size of the derived dispersion estimates.

D.1. Modified extreme deconvolution

We employed the XD method from Bovy et al. (2011) to approximate the noise-free distribution of sources in the Galactic Cartesian velocity space (UVW). Using XD, we aim to minimize the impact of large radial velocity measurement errors on the resulting 3D velocity dispersion.

The term “extreme” in XD highlights its ability to reconstruct the underlying density function even when each source has a unique Gaussian noise covariance matrix. XD utilizes an expectation-maximization (EM) algorithm (Dempster et al. 1977) that iteratively maximizes the likelihood of a GMM representing the noise-free distribution, convolved with the individual error covariances of all measurements.

We aim to use XD to infer a deconvolved density in the 3D Galactic Cartesian velocity space for each of the 12 disk streams. To achieve this, we transform the observations (proper motions and radial velocities) and their corresponding error covariances into UVW space. This involves computing the Jacobian J of the transformation between on-sky velocities ($\mu_\alpha, \mu_\delta, v_r$) and space velocities (U, V, W). The Jacobian was then used to transform the observed covariance matrix, C_{ICRS} , into the error covariance matrix in Galactic Cartesian velocity space, C_{Gal} , as follows⁷:

$$C_{\text{Gal}} = J C_{\text{ICRS}} J^T. \quad (\text{D.1})$$

Assuming approximately Gaussian-distributed 3D velocities, we represent the signal as a single Gaussian distribution. Given the expected nonzero contamination, we explicitly model the contaminating sources. Contamination in this context refers to sources whose space motions and corresponding uncertainties

are incompatible with the signal distribution. Adding a second Gaussian component to the mixture distribution effectively describes the background contamination. To prevent one Gaussian component from collapsing, we ensure that the background component accounts for at least 5% of the observations, discouraging a single mixture component from modeling the entire distribution.

We initialized the mean and covariance matrix of the signal component with values computed from the progenitor stream. The parameters of the background component were initialized as the mean velocity and empirical covariance estimate of the entire local 6D dataset (see Sect. 2), which serves as the basis for each stream search. To ensure that the background component only models uncorrelated and long-range structures, we constrained it to have a diagonal covariance matrix with each diagonal element exceeding a velocity dispersion of 10 km s^{-1} . This constraint prevents the background component from “collapsing” and modeling parts of the signal distribution.

Using these constraints, we empirically find that the signal component in the GMM effectively captures the space velocity distribution of each disk stream. Figures E.2 - E.4 show the distribution of sources in Cartesian velocity space alongside 1- and 2- σ covariance ellipses of the normal signal component inferred by the XD procedure. Table 1 presents the corresponding mean space motion and the 3D velocity deviation, σ_{3D} . The 3D velocity deviation is defined as $\sigma_{3D}^2 = \sigma_U^2 + \sigma_V^2 + \sigma_W^2$, which corresponds to the trace of the covariance matrix.

Finally, we derived an estimate of contamination using this procedure. Our GMM explicitly models the density as a mixture of signal and background components, allowing us to obtain the contamination estimate directly from the mixture weight assigned to the background component. This estimate inevitably includes biases introduced by the discussed model choices, notably that the contamination estimate cannot fall below 5%. Nevertheless, we consider this estimate a reasonable first-order approximation of the contamination fraction within our sample. We find the contamination ranges from 5% to 17%, with a mean contamination of 9%.

⁷ Both the covariances C_{ICRS} and C_{Gal} and the Jacobian depend on a source’s on sky position (ra, dec) and its parallax. Thus, we operate Eq.(D.1) for each source in the catalog.

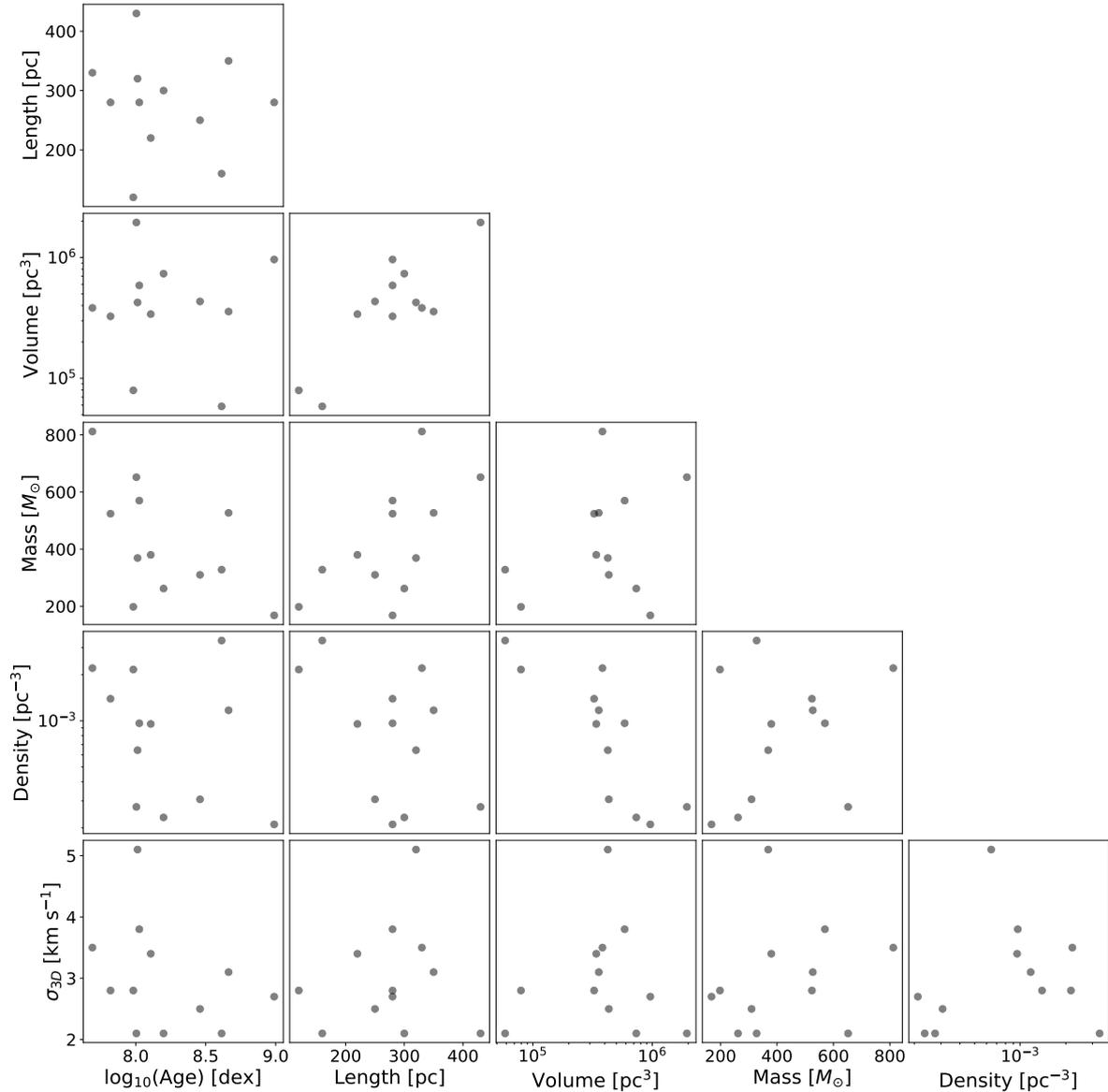


Fig. B.2. Scatter plot matrix showing pairwise correlations between disk stream ages, lengths, volumes, masses, densities, and 3D velocity dispersions.

D.2. Discussion

Our XD analysis reveals that the disk stream populations are dynamically cold, with 3D velocity dispersions between 2.1 and 5.1 km s^{-1} . Velocity dispersions below $\sim 5 \text{ km s}^{-1}$ are notably low for structures spanning several hundred parsecs. Especially since, in Galactic heliocentric velocities, the contribution from Galactic rotation for these large structures is not negligible, amounting between 1 and 3 km s^{-1} .

Changing to Galactocentric cylindrical or action-angle coordinates can provide a more natural way to estimate the remaining intrinsic dispersion as it removes the contribution of Galactic rotation. Our tests conclude that the clustering pipeline is robust to the change of velocity coordinate system (especially for velocities in Galactocentric and Galactic coordinates), with results finding no significant systematic differences in the resulting streams' morphology (length, location, velocity dispersion). Ultimately, we opted to use and report our findings in the heliocentric Galactic Cartesian coordinate system primarily because

it facilitates direct comparisons with other cluster studies that mainly use the same coordinate system.

Comparison to literature values, for example, the fiducial disk stream Meingast-1 (Pisces Eridanus) with a length of approximately 400 pc , has a 3D velocity dispersion in Galactic heliocentric velocities of $\sim 3 \text{ km s}^{-1}$ (Meingast et al. 2019). We also empirically find velocity dispersions below 5 km s^{-1} (in Galactic heliocentric velocities) when analyzing other elongated stellar structures with extents of $100 - 200 \text{ pc}$ like cluster coroneae (Meingast et al. 2021; Moranta et al. 2022) and tidal tails (Meingast & Alves 2019; Röser et al. 2019; Tarricq et al. 2022). As discussed in Sect. 5.3, we believe the identified disk stream sample has a similar origin to coroneae and tidal tails – namely Galactic tidal forces and differential rotation – which might be responsible for such a small velocity dispersion. In future work, when a more comprehensive sample is available, we aim to investigate the kinematic profile of disk streams and in more detail.

Another major factor contributing to small velocity dispersion measurements (also across similar literature examples) is selection bias. The large size of identified structures means that the recovered members are embedded in a large background population. Members from the tail of the velocity distribution that are a few km s^{-1} away from the central overdensity (in whichever velocity coordinate system) will not have enough contrast over the dominating background (i.e., a S/N of ~ 1) to be detected by density-based clustering methods that are typically employed to search for these structures.

Finally, the XD algorithm may bias the results by favoring a more “compact” signal component over a broader or more dispersed one, potentially leading to underestimating the true velocity dispersion. Since we employed a two-component mixture model with restrictions placed on one component, the model may be prone to overfitting, potentially leading to an underestimation of the signal component’s size. Evidence supporting this hypothesis includes contamination estimates from the XD procedure, which are, on average, approximately 2% higher than those derived from the 3D velocity histogram method (see Appendix E).

To address the potential underestimation of the velocity dispersion, we provided a reference estimate using a robust measure, the MAD, to calculate the 3D velocity dispersion. The estimation was done including only sources with radial velocity errors below 1 km s^{-1} to minimize biases from large measurement uncertainties. This approach increased the average 3D velocity dispersion from 2.9 km s^{-1} to 4.6 km s^{-1} , as shown in Table 1.

D.3. Kinematic independence of disk streams

This study’s derivation of disk stream densities assumes that the 12 identified streams are independent structures. However, the spatial overlap (see Fig. 1) between several of these streams raises the question of whether some might represent fragments of larger structures instead. Concretely, we find three groups of disk streams that are, at least partially, co-spatial: (1) Theia 430 and HSC 2278, (2) the four disk streams Theia 371/OCSN 87, NGC 2451A, Theia 301, and OCSN 3, and (3) Ratzenboeck 1, Theia 368, and Mamajek 2.

To investigate this, we examined the kinematic and spatial properties of the identified disk streams. Pairwise comparisons between disk streams reveal that most streams exhibit significantly distinct kinematics, except for one notable case (see below). Using the estimated mean and covariance matrix of each stream’s 3D velocity distribution obtained through the XD procedure, we computed pairwise Mahalanobis distances between the streams’ 3D velocities. The Mahalanobis distance quantifies the separation between two distributions in units of standard deviations, accounting for the shape of their covariance ellipsoids. Distances above 3 typically indicate statistically significant differences in kinematics. As the Mahalanobis distance depends on the covariance matrix of each stream, it is inherently asymmetric.

The pairwise Mahalanobis distances, rounded for clarity, are shown in Fig. D.1. The rows and columns of the pairwise distance matrix are ordered such that disk streams that are spatially close are next to each other for ease of comparison. Orange squares in the figure highlight streams that (at least partially) overlap in 3D spatial volume. This analysis reveals significant kinematic differences among all but one pairwise comparison, pointing toward co-spatial but inherently independent coeval structures. This claim is substantiated by large age differences among individual clusters in these three groups.

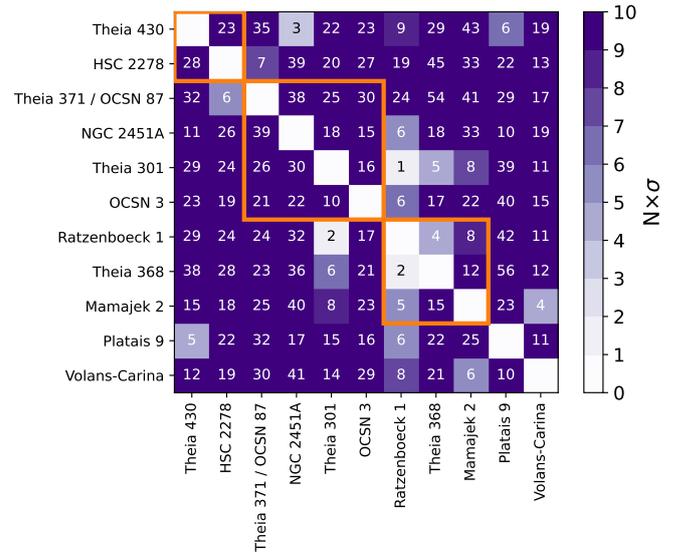


Fig. D.1. Pairwise Mahalanobis distances between the 3D velocities of all identified disk streams. The Mahalanobis distance quantifies the separation between two streams in units of standard deviations, accounting for the covariance structure of their velocity distributions. Distances greater than 3 (highlighted in darker shades) indicate statistically significant kinematic differences. Orange squares denote stream pairs that partially overlap in 3D spatial volume. The disk streams Ratzenboeck 1 and Theia 386, despite their spatial and kinematic proximity, show evidence of being distinct structures based on their 3D morphology and orbital histories, as discussed in the text.

We find one case that needs further investigation, which is Ratzenboeck 1 and Theia 386. These two disk streams are kinematically similar, hinting at fragments of a larger structure. However, closer examination of their 3D morphology and velocity properties provides a more nuanced picture. The two streams show distinct overdensities in XYZ space, and their absolute velocity values point toward separate entities. Specifically, Theia 386, located further along in Galactic rotation, exhibits a relative velocity that points “backward” toward Ratzenboeck 1. This velocity signature indicates a contractive motion, which goes against the expected effects of Galactic tidal forces and differential rotation. To substantiate this claim, we performed a traceback analysis, integrating their orbits 20 Myr into the past and future⁸. This analysis reveals that both streams diverge when their orbits are integrated into the past, providing further evidence that they are two separate, coeval populations.

Despite these differences, the streams share similar, though not identical, 3D velocities and ages. This suggests a potential joint formation scenario analogous to the substructure observed in OB associations such as Sco-Cen or Orion. Their similarities may reflect a common origin or a shared dynamical history within a larger parent structure.

⁸ We used the `galpy` Python library (Bovy 2015) with the default parameters. This parametrization uses the `MWPotential2014` as the Galactic potential, the solar motion relative to the local standard of rest from Schönrich et al. (2010), $(U, V, W)_{\odot, LSR} = (-11.1, 12.24, 7.25) \text{ km s}^{-1}$, and the solar position relative to the Galactic center of $(XYZ_G) = (8122.0, 0.0, 20.8) \text{ pc}$ (GRAVITY Collaboration 2018; Bennett & Bovy 2019).

Appendix E: Contamination estimates

Identifying stellar streams and clusters in a local volume of the Milky Way is inherently challenging due to contamination from the dominant background of field stars in the same phase-space volume. This issue is particularly pronounced for low-density structures encompassing large spatial regions, as even with advanced clustering algorithms, false positives can arise when background stars exhibit positions and velocities that coincidentally match those of stream members.

We addressed these concerns by estimating contamination rates through an approach independent of the clustering and XD procedures. Specifically, we estimated the contamination fraction by comparing the identified stream members with co-spatial *Gaia* sources serving as a background population from the same volume. In contrast to the contamination estimate from the XD procedure, which constitutes rather an outlier approximation (see Appendix D), we explicitly considered the background distribution of field stars and their phase-space density.

Identifying the background population for each stream is a task of finding all co-spatial *Gaia* sources that share the same XYZ volume as respective stream members. These sources serve as a reference sample, which should be distributed differently in phase space and the observational HRD if the stream members are genuine.

We used OCSVM to estimate the support of the positional distribution of each disk stream (i.e., its extent) in Galactic heliocentric coordinates using the identified members. These contours, shown in Fig. 1, were used to define the 3D spatial boundaries of each stream. The background population for each stream was subsequently identified as *Gaia* sources from the 500 pc search data (see Sect. 2) that lie within these 3D volumes but exclude the respective stream members. The median background population size of each stream is approximately 20 000 sources, with a minimum background size of 3 000 and a maximum background size of 100 000 sources. This sample of background stars allows us to quantify contamination and S/N levels by comparison with the identified stream candidate members. The subsequent sections provide detailed results from these analyses and discuss their implications for the reliability of the identified disk streams.

E.1. CMD test

Here, we aim to assess whether the stream populations (see Fig. 4) are significantly narrower in the color-magnitude diagram (CMD) compared to a random sample from the respective background population. We tested the hypothesis that the distribution of sources in the CMD closely follows a single isochronal curve. If the recovered stream members show a significantly narrower pattern around an isochrone curve than the background population, the contamination likely does not dominate the selected sample.

To test whether these two distributions differ drastically in the CMD, we drew ten bootstrap samples (with replacement) from the disk stream members and the corresponding background populations. The bootstrap sample size is chosen to represent the size of respective streams. For each sample, we computed the sources' closest (and signed) distances to an isochronal curve. For the disk streams, these distances were computed relative to the best fitting isochrone determined in Sect. 4 (see Table 1 and Fig. 4). For the background populations, a new best-fitting isochrone was computed for each bootstrap sample to avoid biases in the comparison. To mitigate the biasing effects

of outliers, we removed likely binaries via a cut in the RUWE parameter and white dwarf candidates via a cut in the CMD following Golovin et al. (2024):

$$\begin{aligned} M_G &> 10 + 2.5 \times (G_{BP} - G_{RP}) \\ G_{BP} - G_{RP} &< 1.9 \\ \text{RUWE} &> 1.4 \end{aligned}, \quad (\text{E.1})$$

Each bootstrap draw resulted in two distributions of (signed) distances: one from the stream members and one from the background population. Across all draws, the signed distance distributions of disk streams have smaller variances than respective background samples. To determine whether the stream variances are significantly smaller and, thus, its members are more compactly distributed around the best fitting isochronal curve, we employed Levene's test (Levene 1960). Levene's test evaluates the equality of variances between two distributions. It was chosen due to its robustness against deviations from normality. If the p-value from Levene's test falls below a significance level (e.g., 5%), indicating that the differences in variance are unlikely a result of random sampling from populations with equal variances, thereby supporting our hypothesis.

From the ten bootstrap samples, we obtained ten p-values for each comparison. Using these ten tests, we performed a combined test of the null hypothesis that no p-value is significant. We applied the harmonic mean p-value method (Wilson 2019), which is more robust to dependences among p-values than, for example, the Fisher (1934) method, enhancing its reliability in this context.

Our analysis shows that for all but one disk stream (Theia 599), the hypothesis of equal variances can be rejected at a 2σ level. For most streams, this hypothesis is rejected at a 3σ level, and for some, even at a 5σ level, strongly supporting the conclusion that the stream members are more tightly clustered around the isochrone than their respective background populations.

E.2. 3D velocity test

To explicitly quantify the contamination of each identified disk stream, we performed a velocity-based analysis leveraging the significant differences in velocity dispersion between stream members and the background population. While approximately 10 000 – 100 000 background sources share the same volume alongside selected stream members, the background exhibits a much larger velocity dispersion than the stream members themselves, characterized by velocity dispersions of approximately 5 km s^{-1} or less. Hence, the stream's probability density function is densely concentrated in a small region of the UVW space, while the background population's probability density function is expected to be relatively diffuse. In the following, we expand on this idea to quantify each stream's S/N and contamination rate.

Using the Galactic heliocentric velocity space components U , V , and W within the range of $(-100, 100) \text{ km s}^{-1}$, we divided this space into bins to construct a 3D histogram. We chose a bin size of 5 km s^{-1} along each axis, which guarantees that most stream members fall into one or very few central voxels due to a comparable velocity dispersion for each stream. In contrast, the background population, with its likely more extensive velocity spread, covers a significantly larger volume. For a total of 64 000 voxels in the defined UVW range, even the largest background population ($N \sim 100\,000$) has an average density of only a few stars per voxel, assuming a uniform distribution.

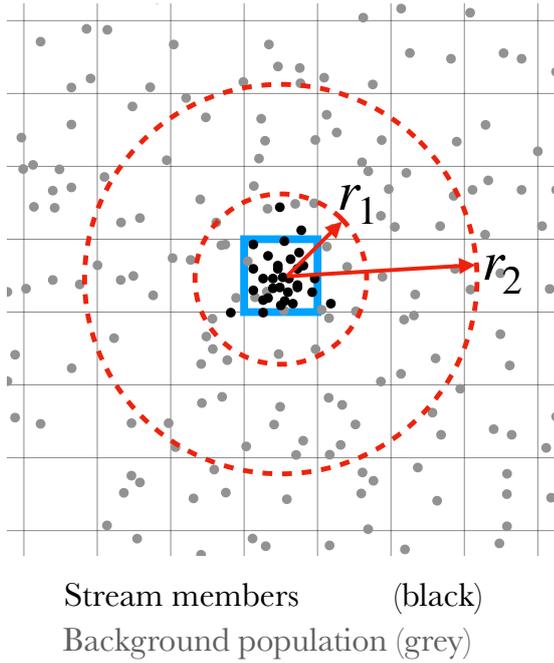


Fig. E.1. Schematic representation of the procedure used to estimate contamination in velocity space. The figure highlights the central voxel (blue) centered on the bulk velocity of stream members, where their density (black dots) is maximal. Background sources (gray points) are distributed more broadly across velocity space. This spatial separation in UVW velocity space enables the estimation of the expected background density at the stream signal location, which is used to calculate each stream’s S/N. The red concentric circles illustrate the neighborhoods across which the expected background number density is estimated.

Although the background is certainly not uniformly distributed, its broader dispersion assumes a likely lower probability mass in the specific regions (i.e., voxels) occupied by the stream members. Figure E.1 shows a schematic overview of this procedure, where stream members (in black) are tightly distributed and, thus, predominantly lie in a small volume in velocity space (the blue pixel). In contrast, background sources (gray points) cover the space more uniformly. Therefore, the expected number density of background sources at the signal location is likely significantly lower than the signal density.

We estimated the expected background density at the stream’s bulk (i.e., mean) motion location to quantify the separation between the stream members and the background. Combining this with the signal density, we calculated each stream’s S/N. Using the background sources, the expected background at the mean stream velocity was determined via the following procedure. First, we determined the number density in the “central voxel” centered on the stream’s bulk velocity, denoted as the signal number density. Figure E.1 shows the central voxel highlighted in blue in which the number density of stream members (black points) is maximal. Second, the expected number density of background sources at the signal location is determined by averaging the voxel count across neighboring cells. Since the background source count varies slightly from voxel to voxel, we averaged the bin counts across multiple cells to obtain a more robust estimate. Specifically, we selected voxels around the central voxel in a growing concentric sphere and computed an expected source count for all voxels whose center location is inside this sphere. We schematically depict this procedure in

Fig. E.1 via the red circles. We selected a minimum radius of 5 km s^{-1} that includes the central voxel and its immediate neighbors and a maximum radius of 30 km s^{-1} . This results in six unique estimates of the expected number density of background sources at the signal location. We obtain an S/N estimate by dividing the signal number density by the expected background number density for each radius value.

Figure E.5 displays the 3D histogram as three marginalized 2D histograms, showing U - V , U - W , and V - W combinations of the disk stream Ratzenboeck 1 and its respective background population. The top row shows only the stream members in red, while the middle and bottom rows show the distribution of background samples and the combined sample of signal and background, respectively. The color map represents the number density where dark gray regions symbolize high and light gray regions low number densities. Each histogram is normalized such that the total area integrates to unity. The horizontal and vertical dashed red lines indicate the bulk (i.e., the mean) velocity of the disk stream. The voxel, which contains most of the disk stream members, has a significant number density increase when adding the identified stream candidate members, as shown in the bottom row. This results in an approximate S/N of 10 ± 2 for Ratzenboeck 1. The remaining 3D histogram plots are provided online via Zenodo, using the following [link](#).

This analysis revealed an average S/N of 27 across all disk streams, with individual values ranging from 5 to 114. These S/N values translate into average contamination estimates of $7 \pm 4\%$. These estimates align well with contamination fractions inferred by XD (see Appendix D). Table 1 provides each stream’s mean S/N estimates alongside its standard deviation from different radius values. A mean S/N of 27 highlights the robustness of the stream selection process, further supporting the result from our CMD test (see Appendix E.1) that contamination does not dominate the identified disk stream catalog.

Appendix F: Auxiliary tables and figures

In Table F.1 we give an overview of the contents of the source-level catalog containing all identified disk stream members as selected in this work alongside membership labels. The full version of the table is available online as a machine-readable version. The phase space coordinates ($XYZUVW$) are defined within the heliocentric Galactic Cartesian coordinate system where the X -axis grows positive toward the Galactic center, the Y -axis grows positive in direction of Galactic rotation, and the Z -axis grows positive toward the Galactic north pole.

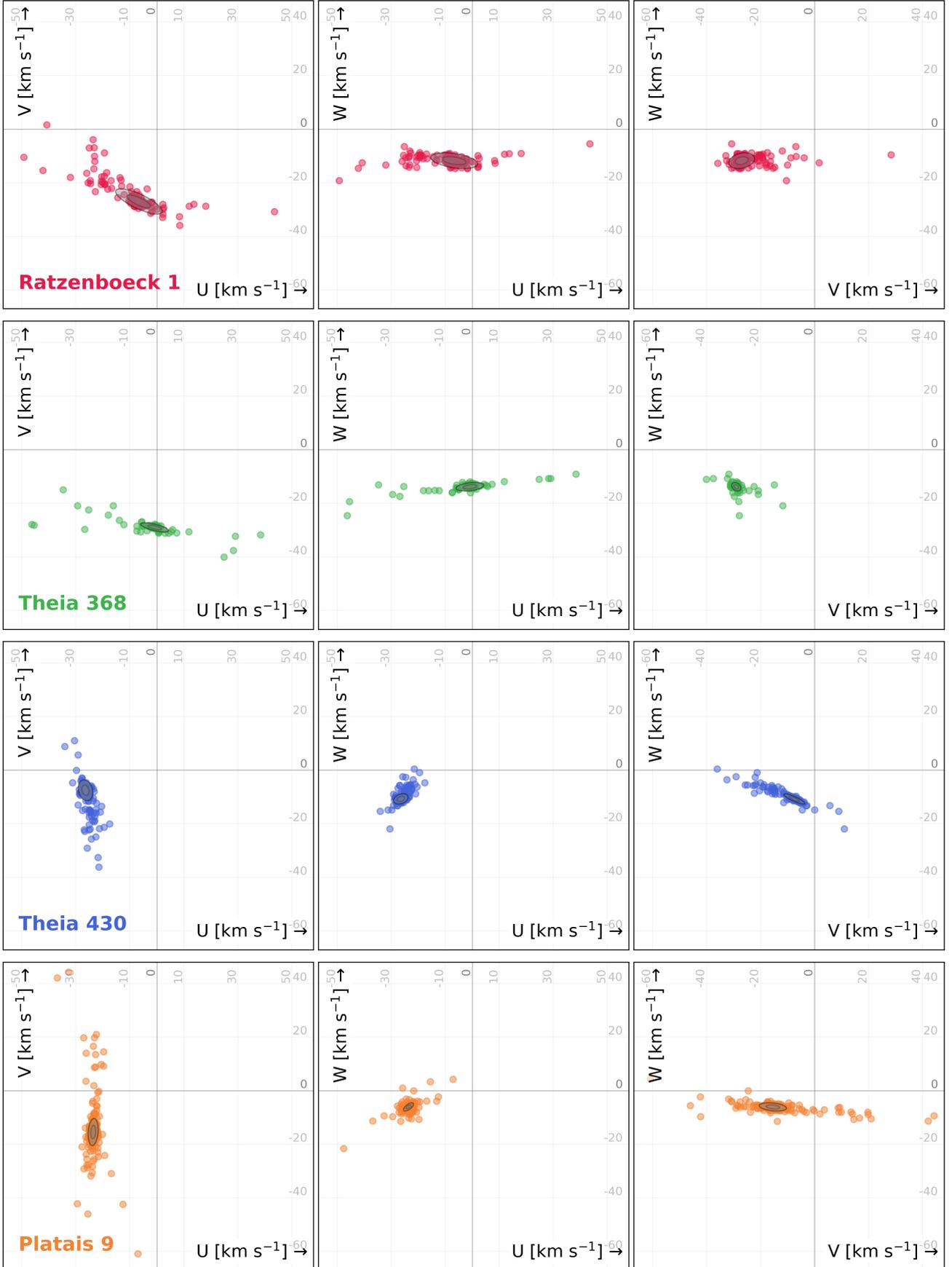


Fig. E.2. *Gaia* Cartesian velocity distributions (UVW) of the 12 recovered disk streams. The black ellipses show the 1σ and 2σ confidence regions inferred by the XD procedure (see Appendix D for more details). The XD procedure can effectively ignore the large line-of-sight velocity errors, which produce the pronounced elongation feature seen in most scatter plots. This figure shows the UVW distribution of disk streams S1 - S4.

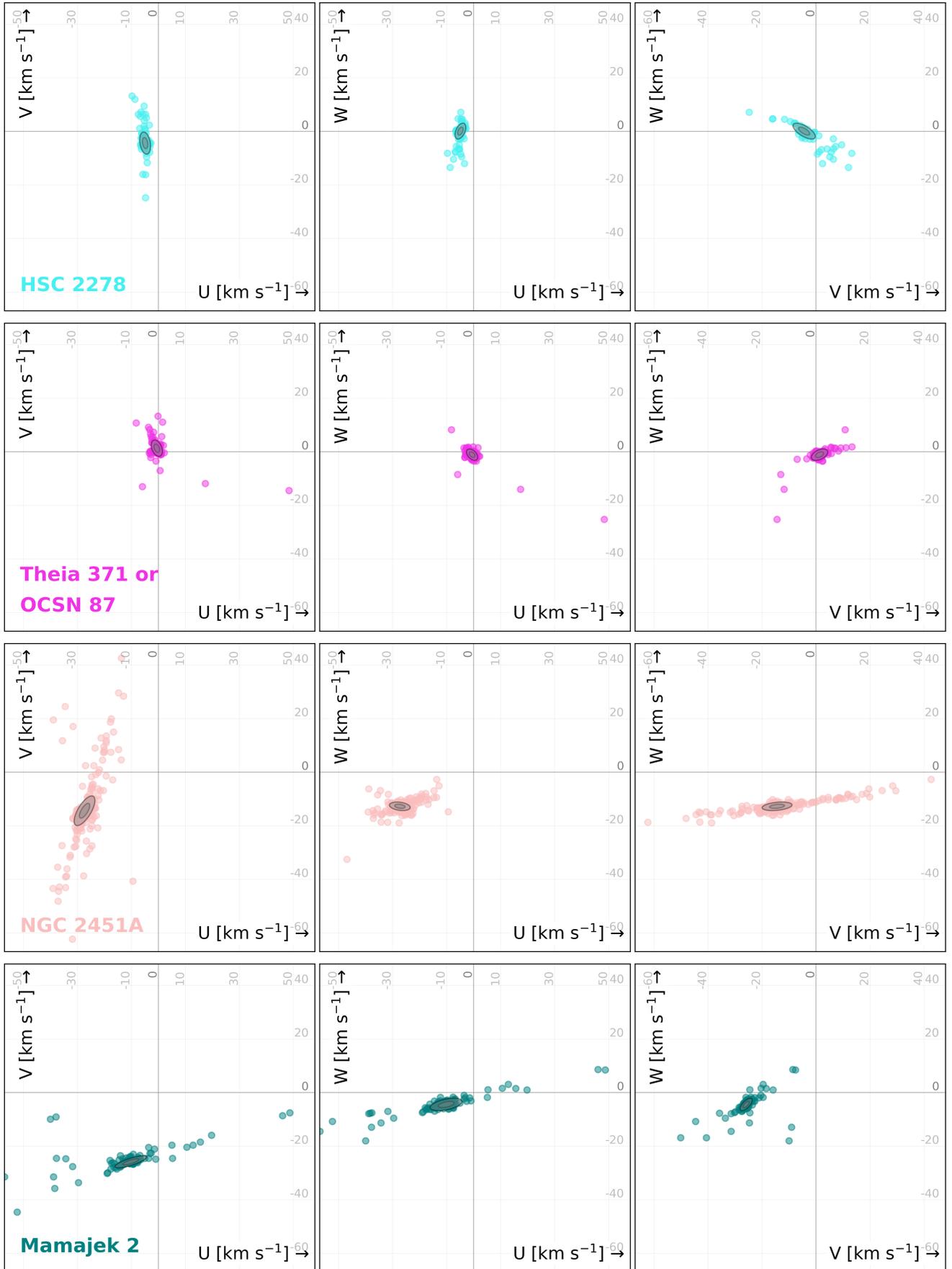


Fig. E.3. Same as Fig. E.2, but for disk streams S5 - S8.

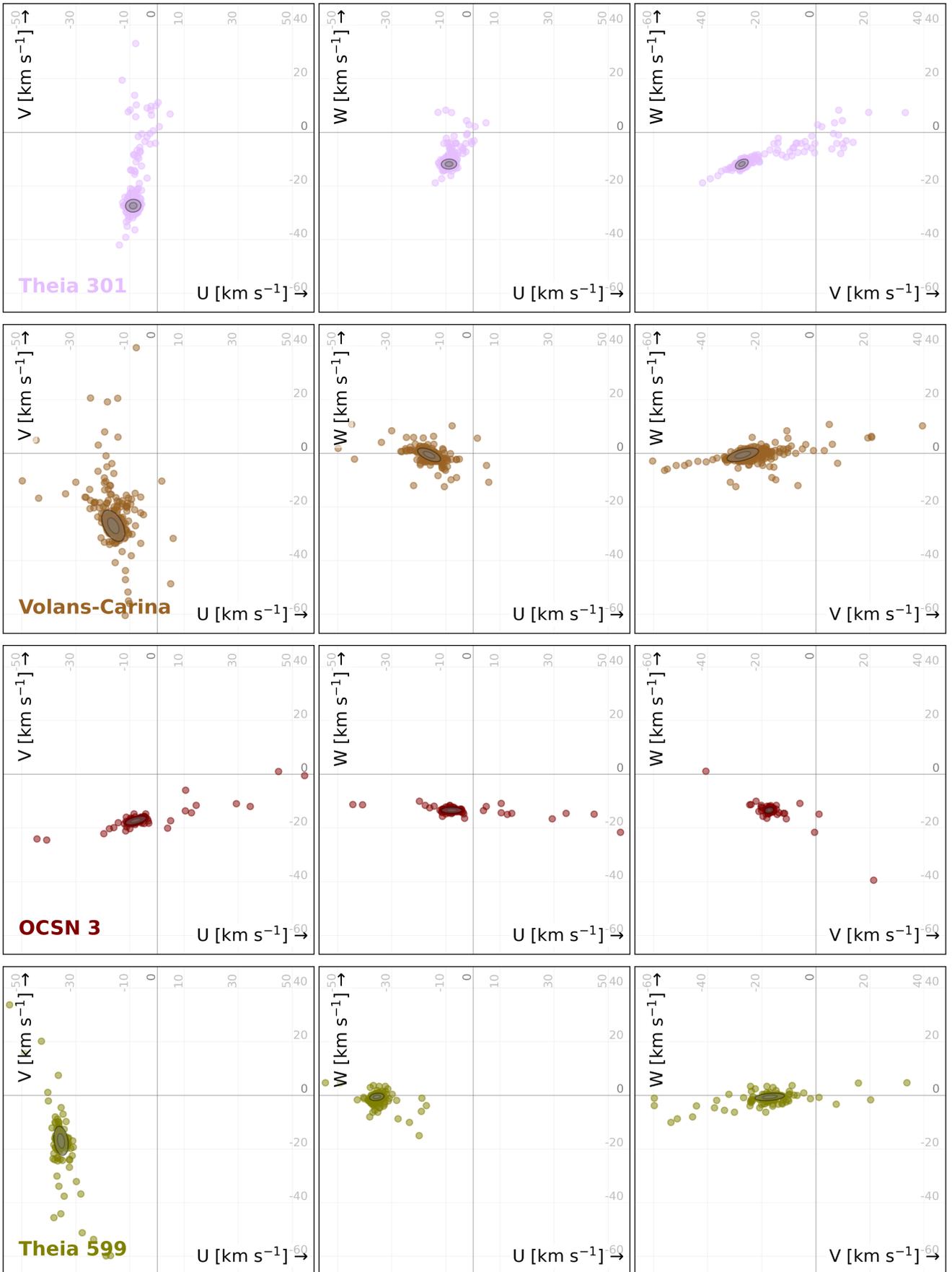


Fig. E.4. Same as Fig. E.2, but for disk streams S9 - S12.

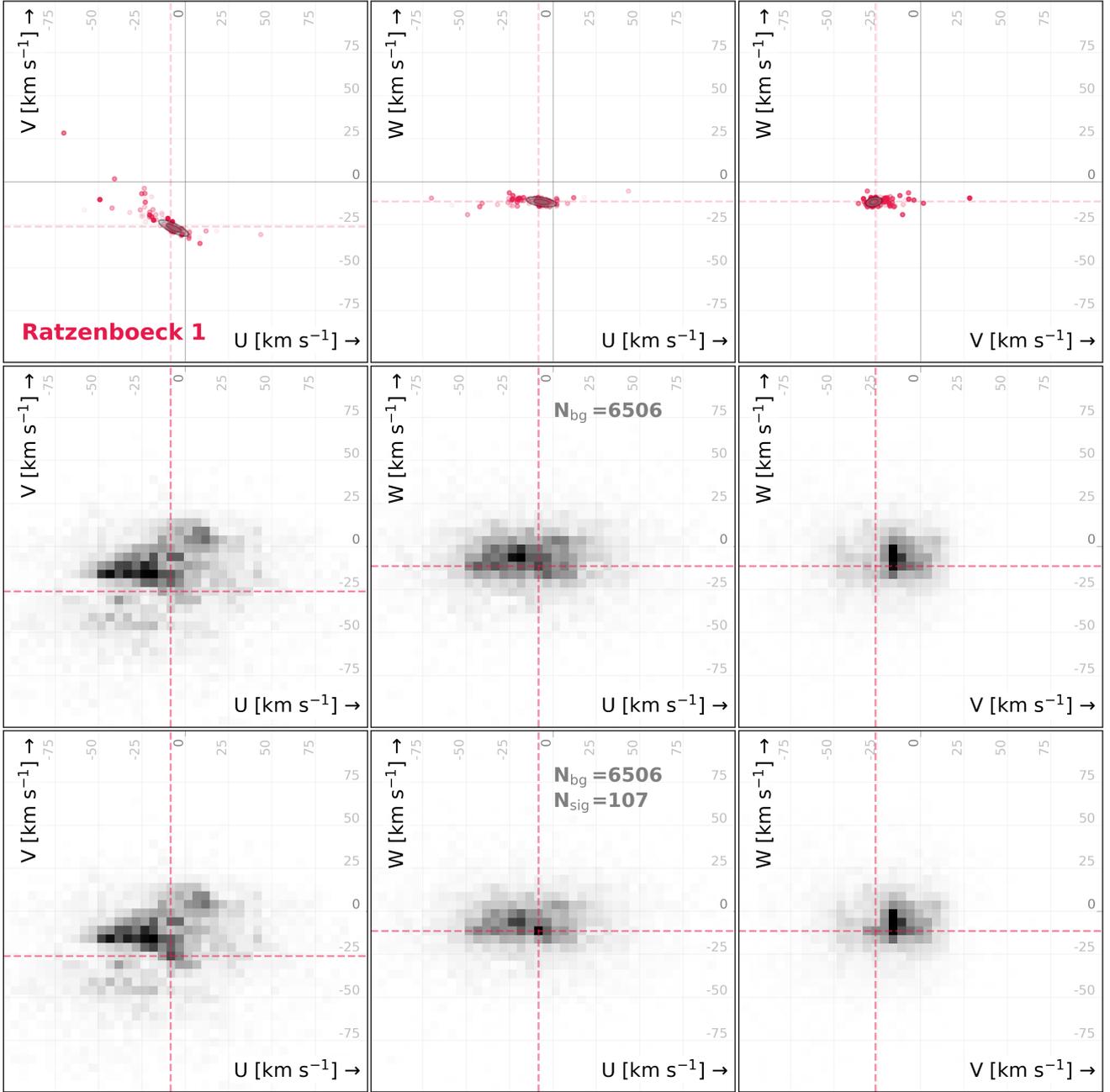


Fig. E.5. S/N estimation procedure applied to the Ratzenboeck 1 stream. The 3D histogram in the middle and bottom row is shown as three marginalized 2D histograms, showing U - V , U - W , and V - W combinations of the disk stream Ratzenboeck 1 and its respective background population. The top row displays the individual stream members in red (akin to Fig. E.2). The middle and bottom rows show the distribution of background sources and the combined sample of signal and background, respectively. The color map represents the number density. Dark regions symbolize high and light gray regions low number densities. Each histogram is normalized and integrates to unity. The horizontal and vertical dashed line indicates the bulk (i.e., the median) velocity of the disk stream. The voxel, which contains most of the disk stream, has a significant number density increase when adding the identified stream candidate members, as shown in the bottom row. The 3D histogram plots showing the other 11 disk streams are provided online via Zenodo, using the following [link](#).

Table C.1. Disk stream comparison against literature catalogs.

SID	This work Size	Blind search										Targeted search						
		Koumkel & Covey (2019)					Fürnkranz et al. (2024)					Meingast et al. (2021)						
		Name	Size ^(a)	Shared ^(b)	Gain ^(c)	f_c [%] ^(d)	Name	Size	Shared	Gain	f_c	Name	Size	Shared	Gain	f_c		
1	273	HSC 2672 ^(f) Theia 436 ^(f)	14 19	14 19	19.5 14.4	100 100	-	-	-	-	-	-	-	-	-	-		
2	172	OCSN 99	196	131	1.3	73	Theia 368 Theia 1569 ^(e)	126 124	79 31	2.2 5.5	79 65	2	128	52	3.3	69		
3	248	HSC 2303	153	133	1.9	93	Theia 430	214	128	1.9	74	-	-	-	-	-		
4	454	Platais 9	251	232	2.0	96	Theia 134 ^(f) Theia 508 ^(f)	339 200	233 77	1.9 5.9	81 79	-	-	-	-	286	1.6	93
5	203	HSC 2278	46	40	5.1	97	-	-	-	-	-	-	-	-	-	-		
6	172	HSC 2407	81	61	2.8	90	Theia 371 ^(e)	86	37	4.6	78	52 ^(e)	295	70	2.5	43		
7	845	NGC 2451A	407	375	2.3	90	Theia 118 ^(e)	1857	628	1.3	41	-	-	-	-	546	1.5	90
8	324	Mamajek 2	226	185	1.8	89	Theia 435 ^(e)	682	152	2.1	38	-	-	-	-	-		
9	534	-	-	-	-	-	Theia 301 ^(e)	1313	188	2.8	32	-	-	-	-	-		
10	566	OCSN 88	333	303	1.9	95	Theia 424	150	90	6.3	90	-	-	-	-	-		
11	196	OCSN 3	78	71	2.8	97	Theia 431 ^(e)	297	137	1.4	55	-	-	-	-	-		
12	418	HSC 2327 ^(f) HSC 2351 ^(f)	91 34	86 32	4.9 13.1	99 100	Theia 599 ^(e)	636	258	1.6	53	-	-	-	-	-		

Notes. Clusters from literature catalogs are listed in this table if their overlap is at least 10% of the respective smaller cluster. For example, if the disk stream size is 300 and a given literature cluster has 150 members, their combined overlap must equal or exceed 15. We additionally impose a minimum overlap size of 10 to exclude chance alignments between unrelated clusters. ^(a)The “size” reports the absolute number of sources in respective literature searches. ^(b)The “shared” column reports the size of the intersection/overlap between sources from our selection and the corresponding literature catalog cluster. ^(c)The “gain” column shows how much our survey increases source counts compared to a specific literature catalog. It compares our group size to the intersection size, representing confirmed samples in the literature. The gain is always greater than one and adjusts for potential contamination in the literature sample (see Fig. 5). This statistic becomes meaningful under the assumption of low contamination in our sample. ^(d)The column “ f_c ” reports an estimation of the completeness fraction (in %) of our sample based on a given reference sample from the literature. We compute the completeness f_c by first considering the incompleteness fraction ($1 - f_c$). We define incompleteness as the number of literature sources we are unable to find divided by the union of our sample and the literature sample. The union provides an estimate of the total number of potential cluster members relative to a given literature selection. Note, however, that the incompleteness fraction is only meaningful as a comparison metric as it depends heavily on the quality of the literature selection and does not take into account the reference’s contamination and completeness. ^(e)(contaminated) The corresponding literature cluster partially overlaps with our selection but appears to be (sometimes significantly) contaminated or describes a (sometimes vastly) different stellar population. See examples in Fig. 5. ^(f)(merge proposed) Several literature clusters suggest that our selection of two disk streams (S1, S4, and S12) should actually be split into multiple clusters. Analyzing the 3D velocity and HRD distribution of these stream fragments (see Ratzenböck et al. in prep), we conclude that these fragments likely correspond to the same stream and should thus not be separated.

Table F.1. Catalog of the identified disk stream members, labeled by membership.

Column name	Column description
source_id	<i>Gaia</i> DR3 source identification number
SID	Membership label; see Table 1
name	Literature name; see Table 1
ra	RA (deg)
dec	Declination (deg)
X	<i>X</i> position (pc)
Y	<i>Y</i> position (pc)
Z	<i>Z</i> position (pc)
U	<i>U</i> velocity component (km s ⁻¹)
V	<i>V</i> velocity component (km s ⁻¹)
W	<i>W</i> velocity component (km s ⁻¹)

Notes. The full machine-readable version of the catalog is available online, while a column overview is given here. We list all relevant derived parameters. Original *Gaia* DR3 parameters can be queried from the *Gaia* Archive by using the `source_id`.