RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Testing the Robustness of Via Machinae Stellar Stream Candidates

Rafael Porto, David Shih

Phenomenology 2026 Symposium

What are Stellar Streams?

- Groups of stars with a common orbit around a galaxy
- Formed by tidal forces that stretch out orbiting globular clusters or dwarf galaxies
- Around 100 known streams in the Milky Way

Why do we care?

- Learn the history of mergers that formed the Milky Way
- Measure the Galactic potential
- **Gaps in streams linked to dark matter substructure**
  - Sensitive to dark matter subhalos down to 10^6 $M_\odot$
  - Subhalo mass function puts constraints on dark matter models

Searching for Streams in *Gaia* Data Release 2

- *Gaia* DR2 contains kinematic data for roughly 1.5 billion Milky Way stars
- Traditional (non-neural net) algorithm example:
  - *Streamfinder* looks for 6D position+velocity “hypertubes” with high occupation
- Neural networks provide a more model-agnostic alternative

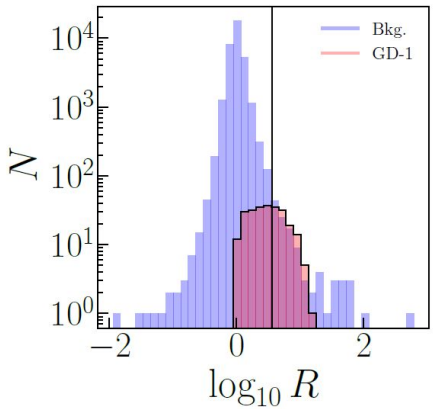
Anomaly Detection with ANODE

Define a signal region (SR) and sidebands (SB)

Train density estimator in both regions

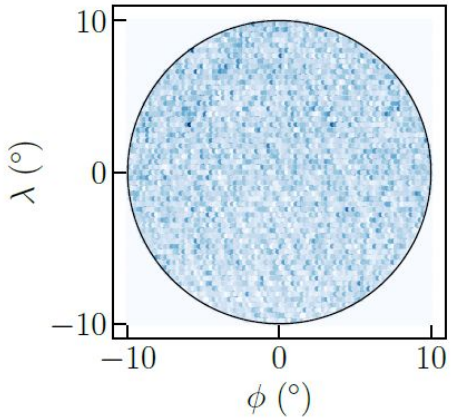
Interpolate from SB into SR

Calculate $R = \frac{p_{data}(x|x_i)}{p_{background}(x|x_i)}$ for each x in SR



Known stream stars have higher R than background stars

Shih, Buckley, Necib, Tamanas 2021



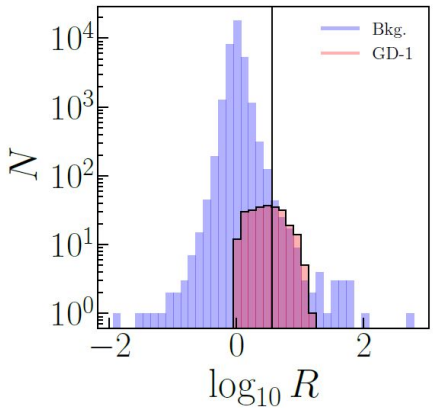
Anomaly Detection with ANODE

Define a signal region (SR) and sidebands (SB)

Train density estimator in both regions

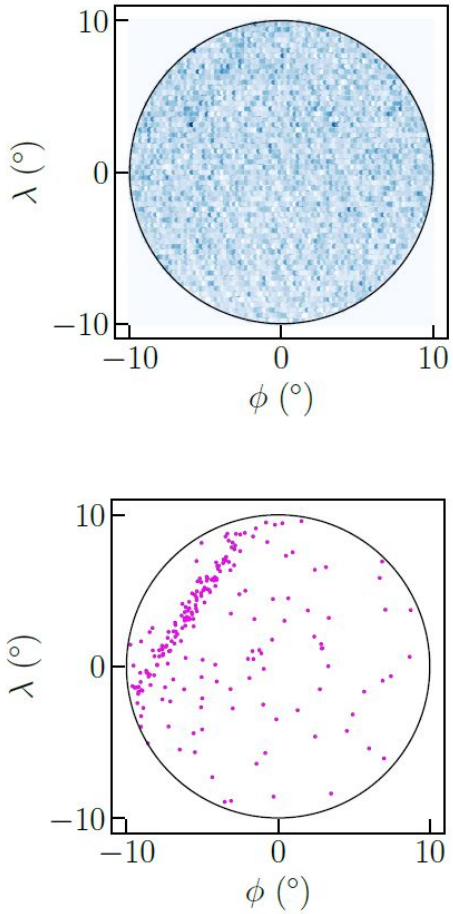
Interpolate from SB into SR

Calculate $R = \frac{p_{data}(x|x_i)}{p_{background}(x|x_i)}$ for each x in SR



Known stream stars have higher *R* than background stars

Shih, Buckley, Necib, Tamanas 202



An (abridged) Overview of *Via Machinae*

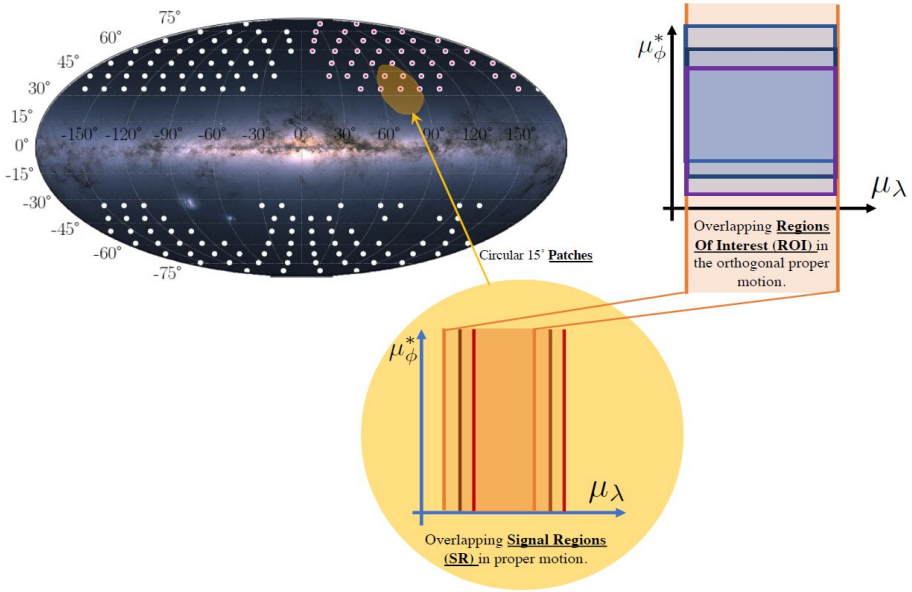
Cut up the sky into SRs

Run ANODE on each SR - get R value for each star

Bin SRs into ROIs

Run line-finding algorithm on high-*R*-stars

Cluster ROIs into protoclusters, protostreams, and then streams



Shih, Buckley, Necib 2023

Estimating the False Positive Rate

Many candidates correspond to known objects - reassuring

False positives: background mismodelling, random alignment

Ran VM on a mock dataset containing no stream-like structures (made using *Galaxia* code)

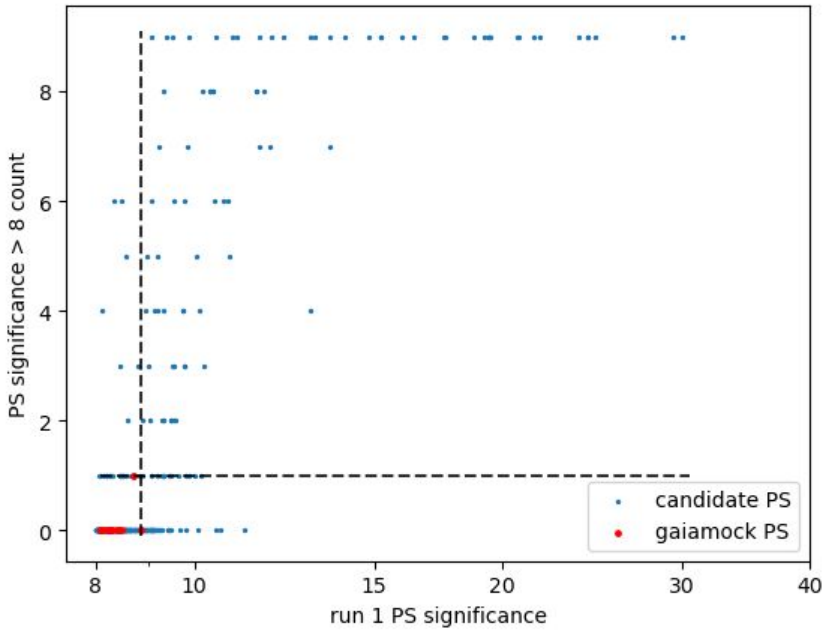
- Stream significance from line finding step is used to classify mock vs *Gaia* DR2 candidates

Robustness of *Via Machinae* Candidates

Retraining density estimators reproduces real streams more often than false positives

Reran ANODE 9 additional times

We count the number of times each PS has significance above detection threshold



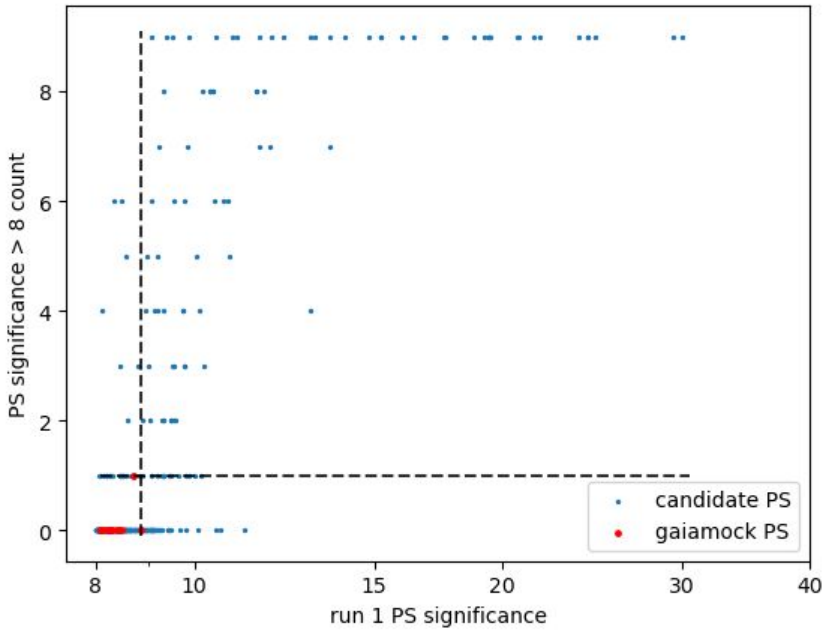
Robustness of *Via Machinae* Candidates

Cut only on significance: 87 candidates, FPR = 12.8%

Cut on both consistency and significance: 106 candidates, FPR = 10.5%

Streams that pass *both* cuts are our top candidates

- 44 candidates have at least one PS in top right of plot



Orbit Fitting

Stream candidates that track an orbit around the Galactic center are more likely to be real

An independent test since VM does not take orbits into consideration

Gaia DR2 provides 4D data

Choose a seed star, optimize 2 missing coordinates such that integrated orbit minimizes chi2 distance to other stars

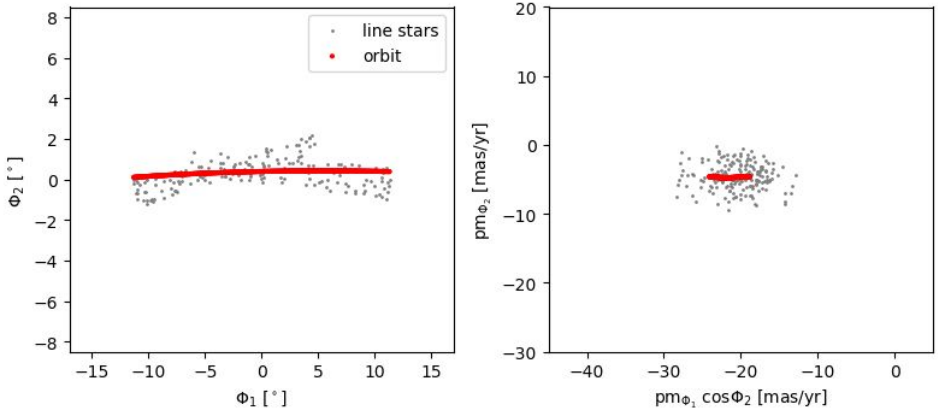
Orbit Fitting

Some but not all candidates produce good fits

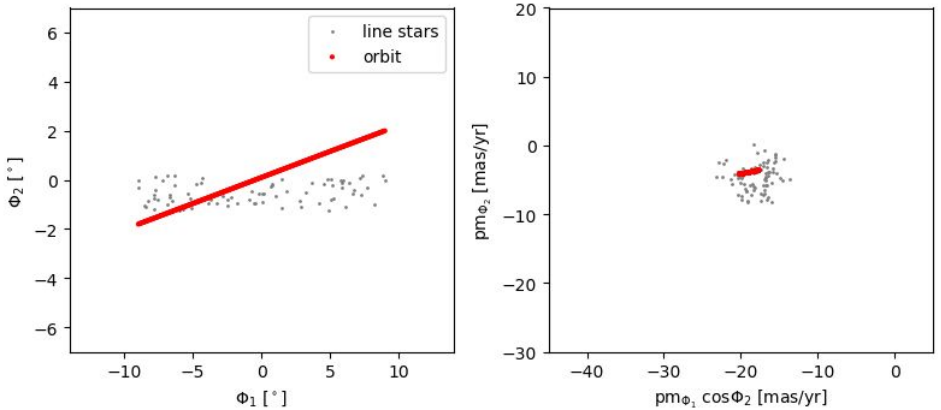
Orbits are evaluated by eye (in position and proper motion), and values of free parameters must be reasonable

Gaiamock candidates do not produce good fits

Optimal initial conditions: d = 5.36 kpc, rvel = 26.99 km/s



Optimal initial conditions: d = 34.17 kpc, rvel = 1000 km/s

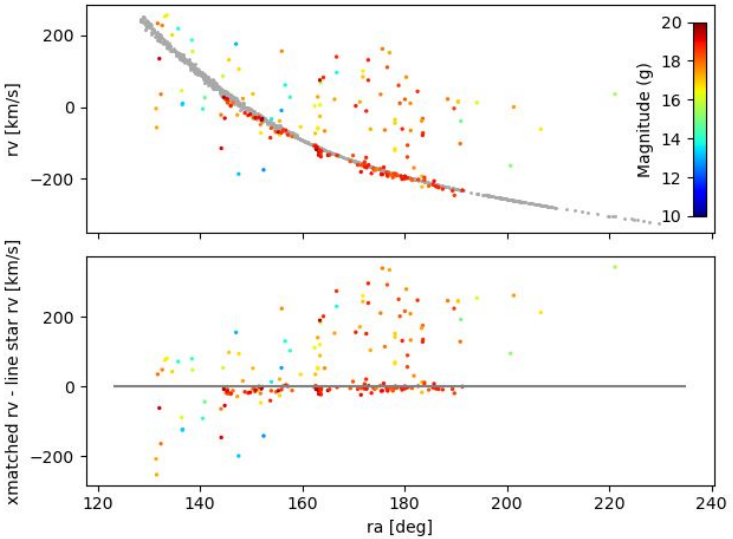


RUTGERS rafael.porto@rutgers.edu

Orbit Fitting

We can predict the distances and radial velocities of candidate stars

GD-1 radial velocity predictions largely agree with measurements from other surveys



Crossmatching GD-1 stars to other surveys allows us to compare our radial velocity predictions to real measurements. (mostly from DESI edr)

Results and Future Work

Of 44 candidates with high significance *and* consistency, 22 have reasonable best-fit orbits

9/22 appear to match previously known streams (in galstreams database)

This leaves us with 13 top candidates to investigate in further detail

- High significance
- High consistency
- Good orbit fit

Results and Future Work

Investigate individual stars that appear repeatedly in the 9+1 VM runs

Fit candidates to isochrones

Perform further observations (metallicity, radial velocity, distance) of stars in top candidates

- Stream stars should have similar metallicities (same progenitor)
- Velocities and distances should match predictions

Summary

- Via Machinae produced 202 stream candidates, of which we've selected the most promising
  - High significance
  - High consistency across retrainings
  - Good best-fit orbit
- Careful not to discard the other candidates (some correspond to known streams!)
- Closer observations of these candidates would confirm or reject their discoveries