SkyCURTAINs: Model Agnostic Search for Stellar Streams with Gaia Data

Debajyoti Sengupta[1], **Stephen Mulligan[1]**, David Shih[2], John Andrew Raine[1], Tobias Golling[1]

Arxiv: [2405.12131](2405.12131)

1. *Département de physique nucléaire et corpusculaire, University of Geneva, Switzerland*
2. *NHETC, Dept. of Physics and Astronomy, Rutgers, Piscataway, NJ 08854, USA*

UNIVERSITÉ DE GENÈVE

Stellar Streams

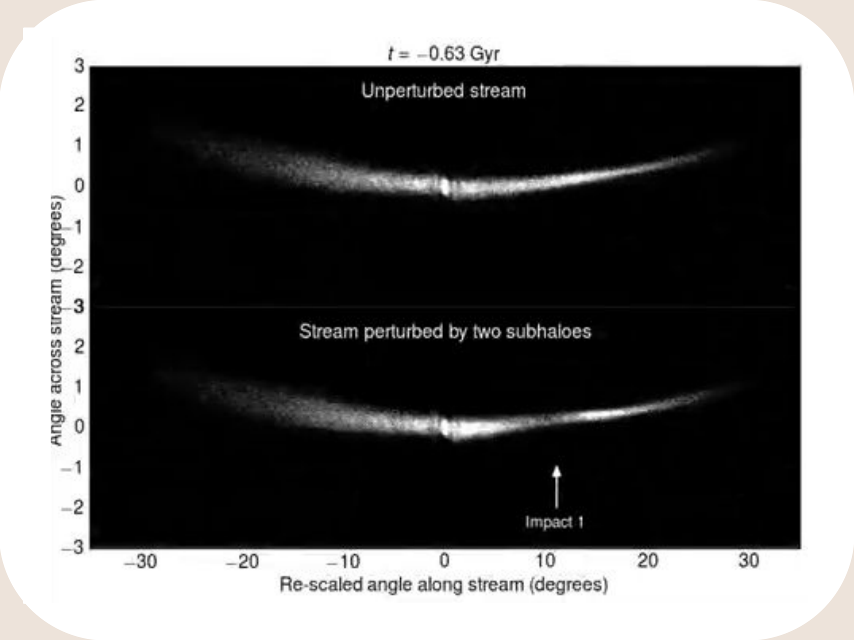
- Formerly globular clusters/ dwarf galaxies in orbit around larger host galaxy
- Disrupted and elongated by tidal forces
- Spread along former orbit
- Due to shared origin, differing from general Milky Way background, expect stellar stream stars to appear as ***overdensity*** in certain features
    - Proper motion

Credit: Gabriel Pérez Díaz, SMM (IAC)

Stellar Streams

Formation occurs over very large timescales - provide information on

- Merger history of systems
- Mass distribution of host galaxies
- Dark matter distribution of host galaxy via density perturbations



Unperturbed (top) vs perturbed stream (bottom)
Credit: Dr. Denis Erkal, University of Surrey

Current Approaches for finding Stellar Streams

“Traditional”
Streamfinder - 1804.11338
Match-filter techniques - 1801.03097

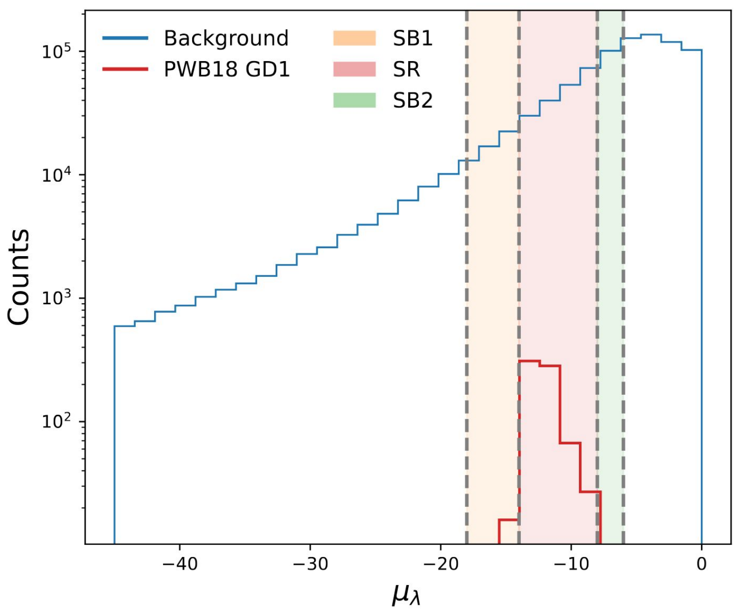
HEP Inspired ML
Via Machinae - 2303.01529
CWoLa - 2305.03761

SkyCURTAINs

- Data driven and model agnostic method to search for stellar streams
- Application of CURTAINsF4F technique developed for template based searches in HEP

Template based searches outside of HEP

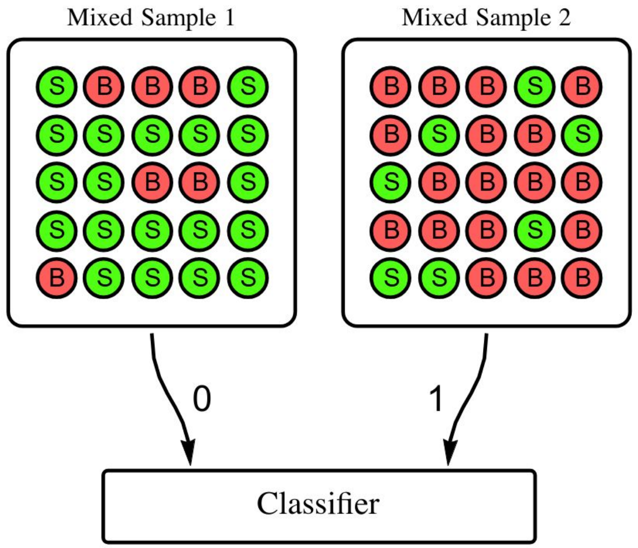
- Resonance search assumes some smooth background distribution and that signals (if they exist) are localised as a resonance about some mean value in the resonant variable
- Does not only occur in HEP cases
- Stellar streams appear as “resonances” in proper motion variables



Latitudinal proper motion of GD1 stream and background stars

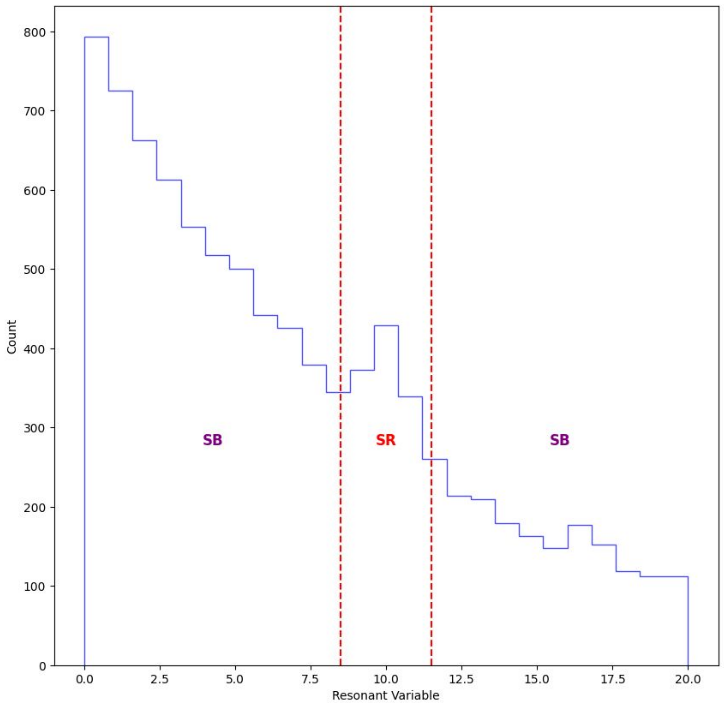
Classification Without Labels (CWoLa)

- Classifier trained on two mixed samples *M1* and *M2*
- CWoLa Theorem: The optimal classifier trained to distinguish M1 and M2 is also optimal for distinguishing S and B



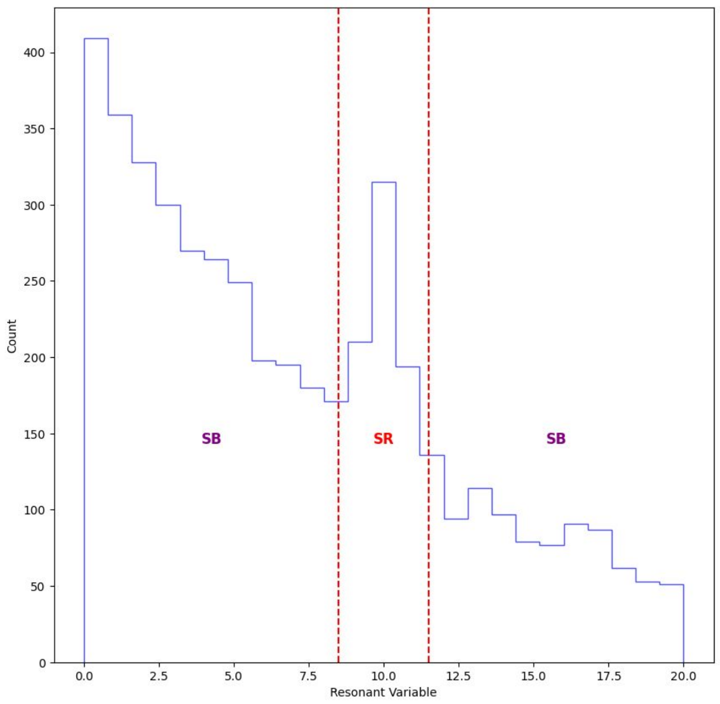
Resonance Searches with CWoLa

- Define Sideband (SB) and Signal Region (SR) where each should have differing Signal/Background ratios
- Train a classifier CWoLa style to distinguish between SB and SR
- Perform a search on some resonant variable after cutting with classifier at some threshold
- Ideally, signal strength is enhanced after classifier cut



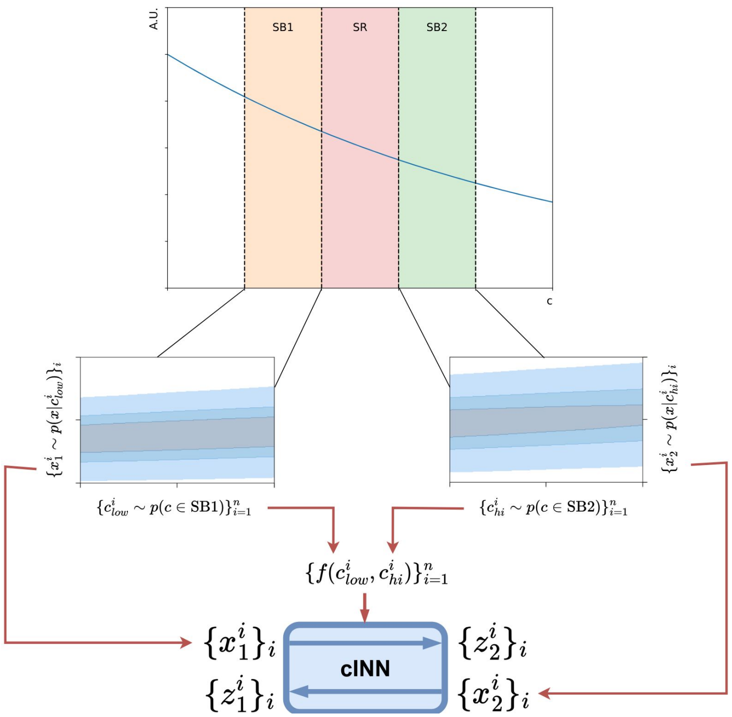
Resonant Searches with CWoLa

- Define Sideband (SB) and Signal Region (SR) where each should have differing Signal/Background ratios
- Train a classifier CWoLa style to distinguish between SB and SR
- Perform a search on some resonant variable after cutting with classifier at some threshold
- Ideally, signal strength is enhanced after classifier cut



CURTAINsF4F

- Flow based method for constructing background templates
- Baseflow - maps from normal to SBs
- Topflow - maps from SB to SB
- Final mapping from SB to SB, conditioned on the resonant variable
  - interpolates the background data into the SR
- Training a CWoLa classifier between template and real SR data



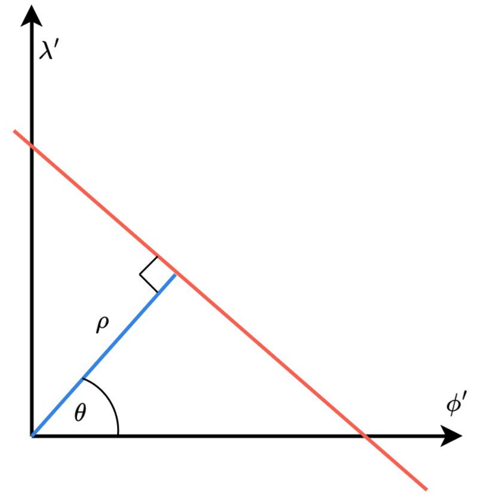
2305.04646

SkyCURTAINs

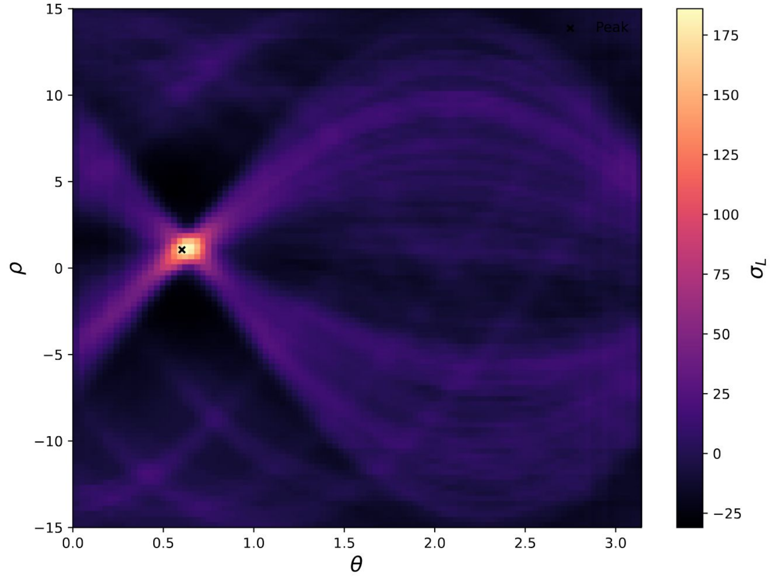
- Train CWoLa classifier on SR data versus a background template generated by CURTAINsF4F
- Perform cut on classifier output to enhance signal - identifies all anomalous stars relative to background distribution
- Specifically interested in *line-like* anomalies
- Cut is followed by ***Hough transform*** and line fitting of stellar data to identify potential stellar streams

Hough Transform

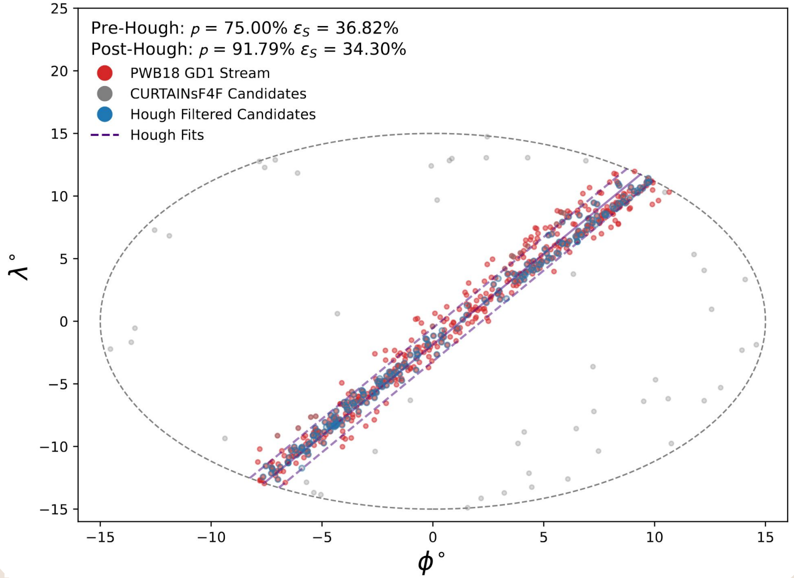
- Introduced for stellar streams in VM
- For a star at $(\phi', \lambda') \rightarrow (\varrho, \theta)$ by $\varrho = \phi'\sin\theta - \lambda'\cos\theta$



Hough Transform and Line Fitting

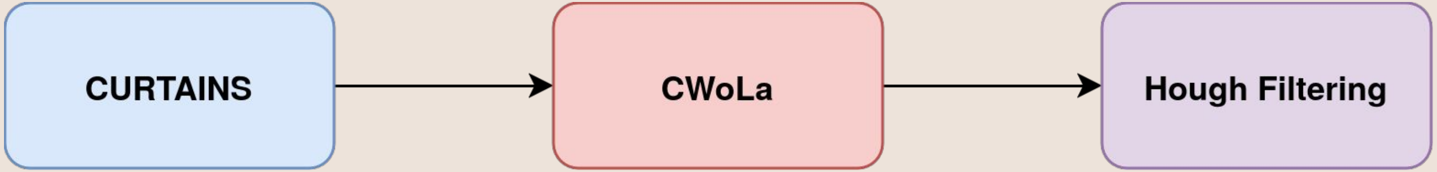


Significance map in Hough space for candidate signal stars



Filtering of candidate stars using line parameters found for maximum significance in Hough space

Pipeline Overview



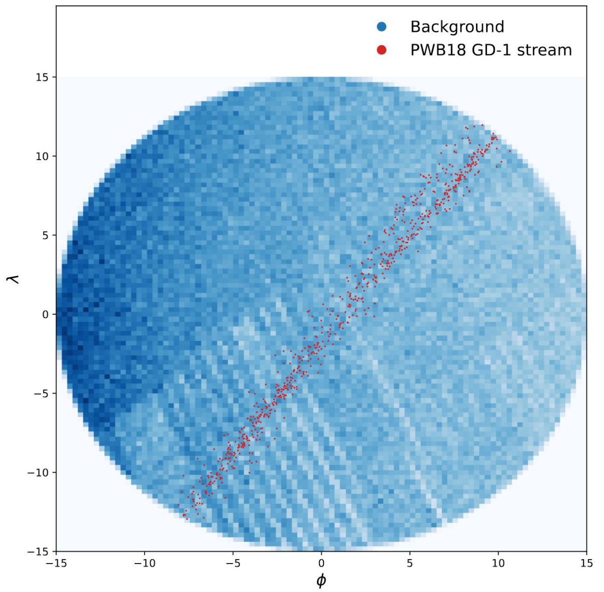
1. Train CURTAINs to generate template of background in SR

2. Train CWoLa classifier between real SR data and background template

3. Cut on classifier output and filter stars using Hough filter to find line like anomalies

Rediscovering GD-1 - Proof of Concept

- GD-1 - long dense stream in the Milky Way
- Discovered in 2006 - Grillmair and Dionatos
- Use *Gaia Data Release 2*(GDR2) for training data
- Stream membership taken labels taken from a 2018 study by *Price-Whelan and Bonaca* (PWB18)
- Train on 21 patches
  - Kinematic features - $(\mu*_{\phi} = \mu_{\phi} \cos \lambda, \mu_{\lambda})$
  - Spatial - (ϕ, λ)
  - Photometric - $(G, G_{BP} - G_{RP})$



Example section of GD-1 stream from patch in GDR2 data

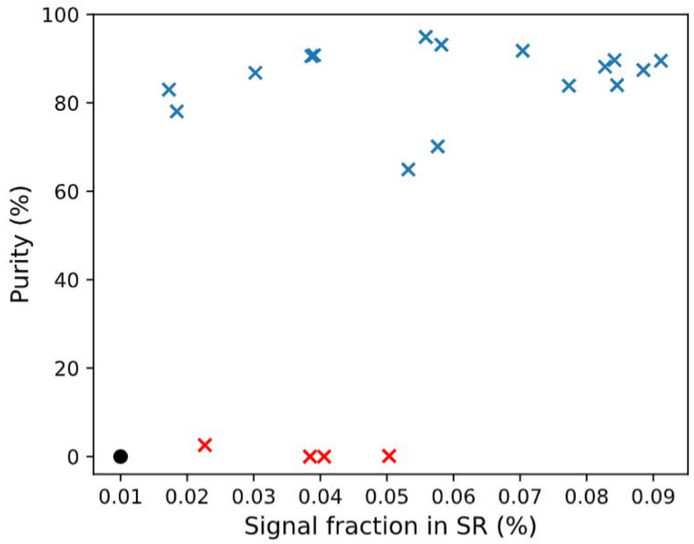
Performance - Purity

SkyCURTAINs largely enhances CWoLa performance, with few exceptions

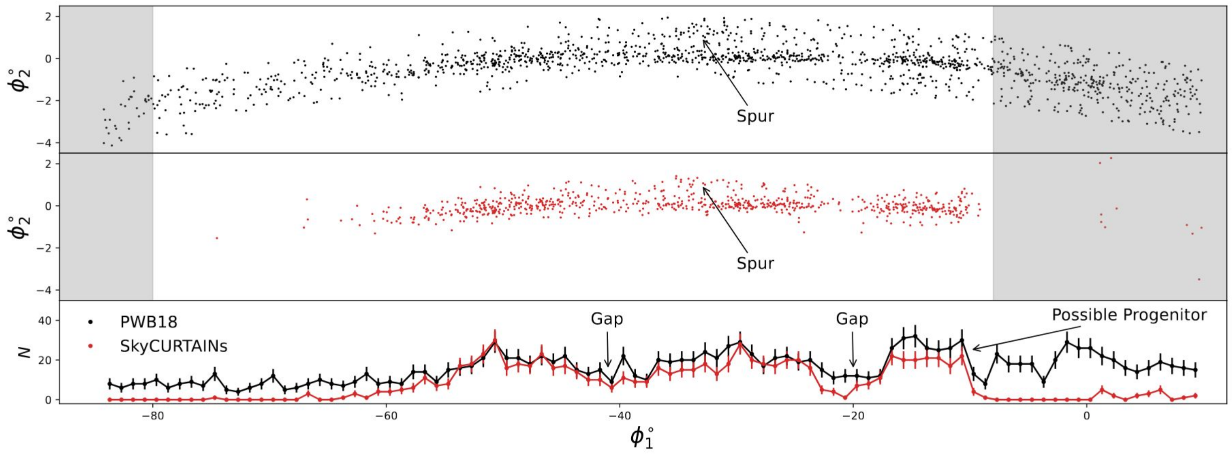
Patch (α, δ)	p SkyCURTAINs	CWoLa
(128.4°, 28.8°)	**82.99**	77.0
(132.6°, 16.9°)	**78.05**	62.0
(136.5°, 36.1°)	**90.56**	86.0
(138.8°, 25.1°)	**90.79**	84.0
(142.7°, 14.5°)	**86.79**	65.0
(146.9°, 35.6°)	**91.79**	90.0
(148.6°, 24.2°)	**94.9**	87.0
(148.6°, 47.0°)	**93.15**	78.0
(156.2°, 57.5°)	**70.14**	54.0
(156.9°, 34.1°)	**88.17**	86.0
(160.5°, 45.5°)	**87.43**	73.0
(171.4°, 43.0°)	**89.52**	72.0
(171.8°, 54.7°)	**89.66**	53.0
(174.3°, 65.1°)	**64.94**	47.0
(185.4°, 50.0°)	**84.0**	57.0
(192.0°, 58.7°)	**83.87**	66.0
(138.1°, 5.7°)	0.0	0.0
(203.7°, 49.1°)	**0.13**	0.0
(212.7°, 55.2°)	0.0	0.0
(224.7°, 60.6°)	**2.58**	0.0
(202.4°, 66.5°)	0.0	**50.0**

Performance - Purity

- Red
  - SR centred around low proper motion
  - Dominated by distant stars - reduces sensitivity
- Black
  - Low signal to background ratio



Recovery of GD-1



Conclusions and Outlook

- Model agnostic data driven methods developed for one domain *can* be applied in others

- SkyCURTAINs recovers the GD-1 stream with minimal model assumptions and high purity in nearly all patches

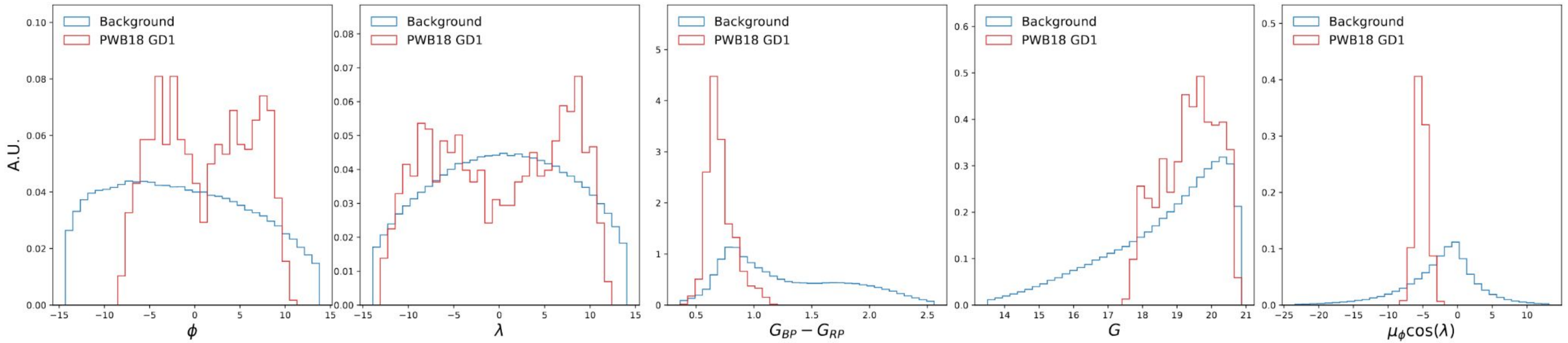
- Full sky scan

- Training two normalising flows for multiple SRs in multiple patches is expensive
  - Can train single baseflow per patch, multiple topflows?

- Modular design allows for use of different template generation approaches e.g. TRANSIT
  - TRANSIT presented by Ivan Oleksiyuk - 11:30am tomorrow, anomaly detection track

Thanks for listening! Questions?

Features



- Marginal feature distributions of patch $(\alpha,\delta) = (146.9^{\circ},35.6^{\circ})$
- Normalised to 1 to highlight qualitative differences

Feature Processing

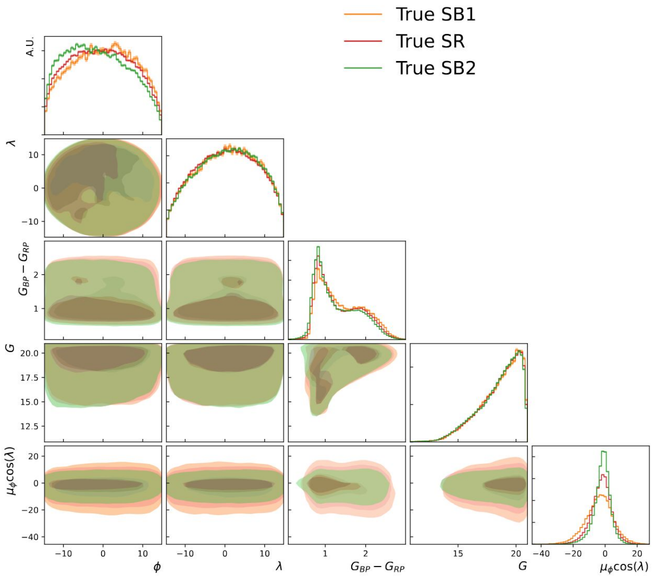
- Restrict analysis to distant stars with parallax < 1mas
- Patches too close to galactic plane are dropped to avoid foreground contamination
  - Also excluded patches overlapping with Large and Small Magellanic clouds
- Trained on
  - Kinematic features - $(\mu*_\phi = \mu_\phi \cos\lambda, \mu_\lambda)$
  - Spatial - (ϕ, λ)
  - Photometric - $(G, G_{BP} - G_{RP})$
- Cuts after CWoLa step before downstream
  - Kinematic cut - $|\mu_\phi^*| > 2mas/yr$ OR $|\mu_\lambda| > 2mas/yr$
    - To reject distant stars
  - Photometric cut - $0.5 \leq G_{BP} - G_{RP} \leq 1$ and $G < 20.2$
    - Isolates older, low metallicity stars

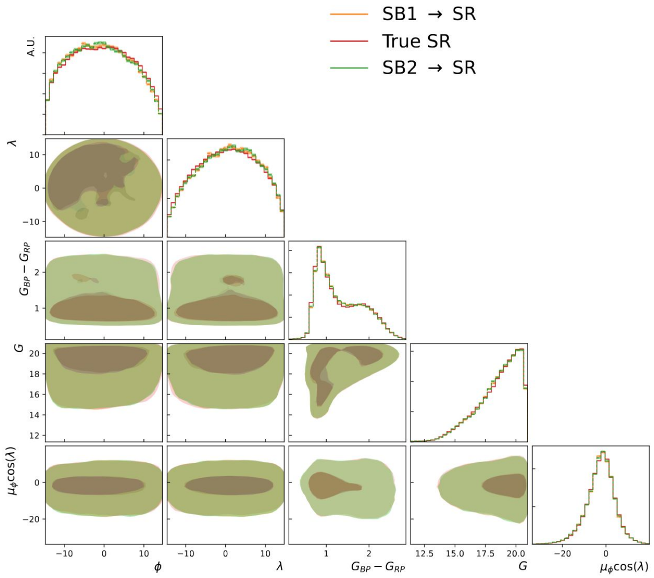
Hough Significance

- N_{ij} - number of curves passing through a pixel
- $N_{bg,ij}$ - position-dependent background estimate
- Delta $N_{bg,ij}$ - “systematic error” on this estimate

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

Template Correlations





Stellar Streams

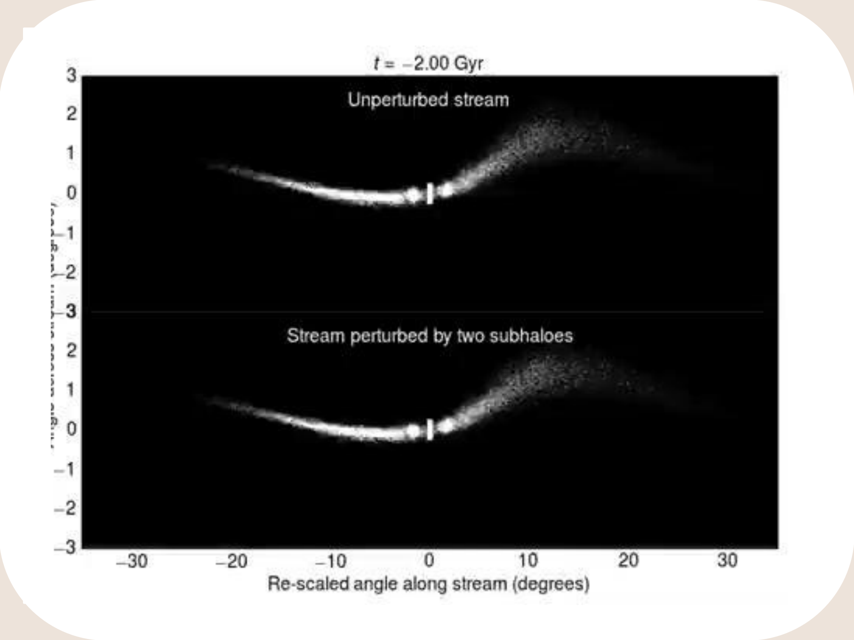
- Formerly globular clusters/ dwarf galaxies in orbit around larger host galaxy
- Disrupted and elongated by tidal forces
- Spread along former orbit
- Due to shared origin, differing from general Milky Way background, expect stellar stream stars to appear as ***overdensity*** in certain features
  - Proper motion

Credit: Gabriel Pérez Díaz, SMM (IAC)

Stellar Streams

Formation occurs over very large timescales - provide information on

- Merger history of systems
- Mass distribution of host galaxies
- Dark matter distribution of host galaxy via density perturbations



Unperturbed (top) vs perturbed stream (bottom)
Credit: Dr. Denis Erkal, University of Surrey