Identifying ATLAS stream members with a Bayesian mixture model

Andrew Li — 1006675625 — andrewp.li@mail.utoronto.ca

May 15, 2023

1 Introduction

This report aims to determine stellar members of the ATLAS-Aliqa Uma stream, henceforth referred to as ATLAS. Data from the Southern Stellar Stream Spectroscopic Survey (S^5) (Li et al., 2019), and Gaia DR2 (Gaia Collaboration et al., 2018) is used to construct a Bayesian mixture model, from which we can determine the membership probabilities of each individual star. We will also compare our results to the identified parameters and members in (Li et al., 2021).

First, we select our data and perform cuts to ensure good quality measurements. We then convert the celestial equatorial (α, δ) to spatial stellar stream coordinates (ϕ_1, ϕ_2) defined by (Shipp et al., 2019). This conversion is given by

$$\begin{pmatrix} \cos(\phi_1)\cos(\phi_2) \\ \sin(\phi_1)\cos(\phi_2) \\ \sin(\phi 2) \end{pmatrix} = \begin{pmatrix} 0.83697865 & 0.29481904 & -0.4610298 \\ 0.51616778 & -0.70514011 & 0.4861566 \\ 0.18176238 & 0.64487142 & 0.74236331 \end{pmatrix} \times \begin{pmatrix} \cos(\alpha)\cos(\delta) \\ \sin(\alpha)\cos(\delta) \\ \sin(\delta) \end{pmatrix}. \tag{1}$$

The velocity measurements, given as heliocentric (v_{hel}) are also converted to be with respect to the Galactic Standard of Rest (v_{GSR}), which will be completed using a function described in Price-Whelan (2023).

We now construct a mixture model, consisting of one stream component and one background component. The model is based on three measurement quantities: v_{GSR}, the proper motion in right ascension (μ_α), and the proper motion in declination (μ_δ). Each quantity has a mean and scattering, which will be modelled into a Gaussian. Additionally, the means of these quantities will be a quadratic function of ϕ_1 for the stream component. We use a Markov Chain Monte Carlo (MCMC) sampler to optimize these parameters. Once they are determined, we can find the membership probabilities of each star.

This report is structured as follows. In Section 2, we present the data from S^5, Gaia DR2, and Li et al. (2021), along with the quality cuts made. In Section 3, we present the mixture model and perform MCMC to find the best-fit parameters. The membership probabilities are subsequently determined. In Section 4, we discuss our results and compare to Li et al. (2021), and conclude in Section 5. The methods used in this report will be completed through Python using Jupyter Notebook. The Python libraries used are `numpy` (Harris et al., 2020), `matplotlib` (Hunter, 2007), `astropy` (Astropy Collaboration et al., 2022), `emcee` (Foreman-Mackey et al., 2013), `corner` (Foreman-Mackey et al., 2023), `scipy` (Virtanen et al., 2020), and `schwimmbad`.

2 Data

The data used in this report is collected from S^5 and Gaia DR2. This data is provided as a FITS file and the stars in spatial proximity to the ATLAS stream can be extracted by the key `'object_name'=='ATLAS'`. v_{GSR} is also determined using the function described in Price-Whelan (2023), and the stream coordinates using Equation 1.

Quality cuts are also performed to exclude both poor measurements and stars far away from the known parameters. We select stars with `'good_star'==1`, [Fe/H] < -1.5, $|v_{\mathrm{GSR}}| < 250$ km s^{-1}, $|\mu_\alpha \cos\delta| <$

4 mas yr^{-1}, and $|\mu_\delta| < 4$ mas yr^{-1}. After these cuts, we are left with 1044 stars.

Data of known members of ATLAS is also collected from Li et al. (2021). This data is provided as a FITS file and these stars are selected by the key `'regionname'=='AAU'`. The $v_{\rm GSR}$, ϕ_1, ϕ_2 transformations are also applied similarly. All of this data can be seen in Figure 1.

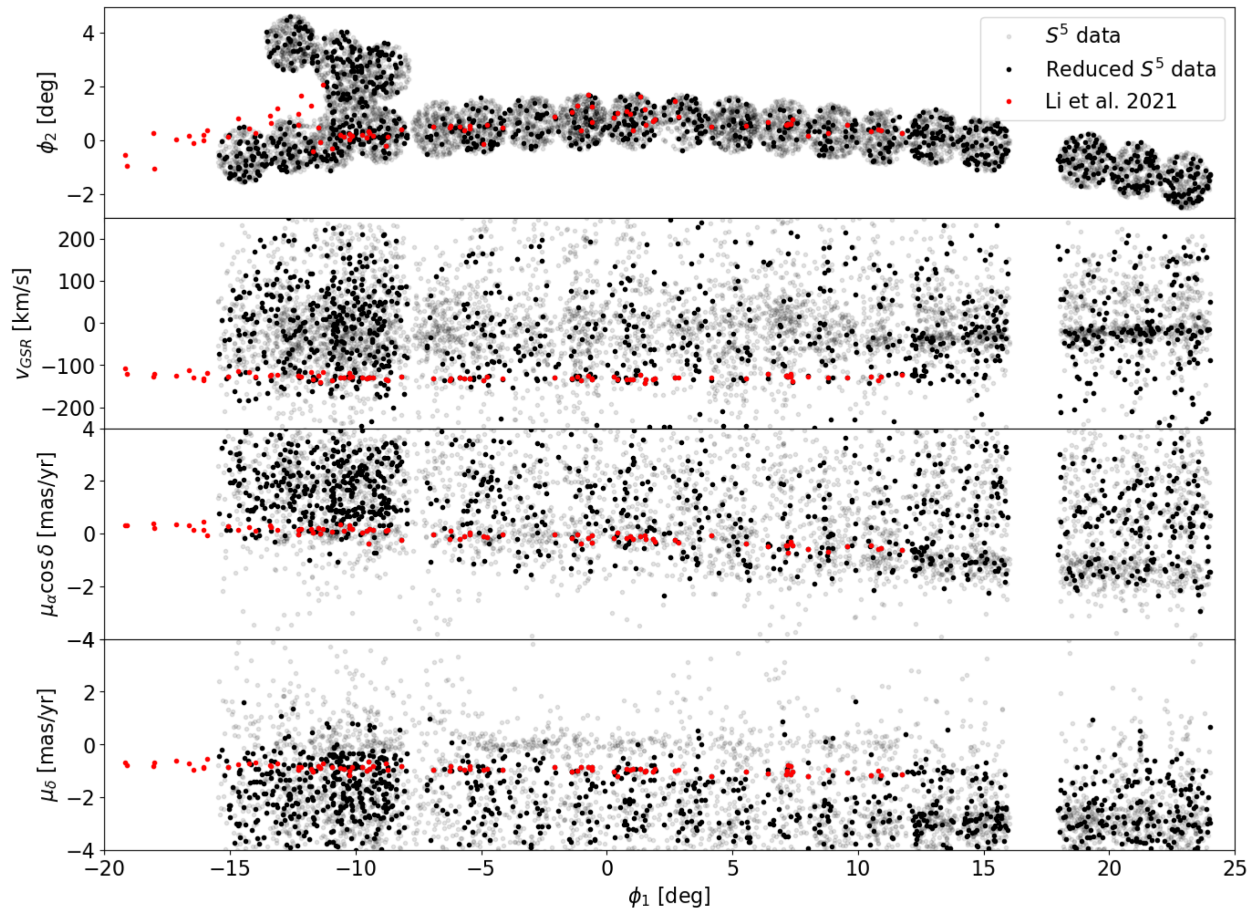

Figure 1: The data members used in this report. For all panels, grey markers represent S^5 members that did not pass the quality cuts described in Section 2, and black markers represent those that did. The red markers indicate the members of ATLAS identified by Li et al. (2021). All panels all share an axis of the stellar stream coordinate ϕ_1, in units of degrees. The panels (top to bottom) show ϕ_2 in degrees, $v_{\rm GSR}$ in km s^{-1}, $\mu_\alpha \cos\delta$ in mas yr^{-1}, and μ_δ in mas yr^{-1}.

3 Data analysis

3.1 Modelling

The mixture model used will consist of two components, which we will call 'stream' and 'background'. The first parameter $p_{\rm stream}$, will be the fraction of stars in the stream component. For the stream component, three measurement qualities will be used: $v_{\rm GSR}$, $\mu_\alpha \cos\delta$, and μ_δ. Each of these will be represented by two parameters, mean and variance, which compose a Gaussian distribution. Additionally, the means for each of these quantities will be a quadratic function of ϕ_1. For example, $v_{\rm GSR}$, the mean will be represented as

$$v_{\rm GSR} = v_{\rm GSR,1} + v_{\rm GSR,2}x + v_{\rm GSR,3}x^2, \tag{2}$$

where $x = \phi_1/10^\circ$ and ϕ_1 is in units of degrees and $v_{\rm GSR}$ is in km s^{-1}. The quadratic coefficients for the proper motion components are defined similarly. This gives the stream component 12 parameters. For the background component, we will also use a Gaussian for each quantity, but this time there will be no dependence on ϕ_1 for the means. This gives another 6 parameters, for a total of 19 parameters. These parameters can be seen in Table 1.

We will now construct the probability function to be used by the MCMC sampler. The probability function is the product of the prior and the likelihood. The prior will simply be hard cutoffs for each parameter, determined by inspection of Figure 1 and results from Li et al. (2021). The selected bounds are also shown in Table 1. The likelihood functions for the stream and background components are products of each of the three Gaussians for each measurement as stated above. They are then scaled by their respective contributions ($p_{\text{stream}}, 1 - p_{\text{stream}}$) and multiplied together.

Our initial guesses for the parameters are formulated both visually and using values from Li et al. (2021). Using `minimize` from `scipy.optimize`, we minimize the negative probability function to get a better estimate for our parameters. Finally, we maximize the likelihood function using `EnsembleSampler` from `emcee`, run with 64 walkers and 2000 iterations. The initial guesses, `scipy` results and `emcee` results are all shown in Table 1. The corner plot of the MCMC sampling result is also shown in Figure 2.

Table 1: Summary of the modelling and analysis completed in Section 3. In the leftmost column, the 19 parameters are shown, along with their representation in the code and their units. The subscript b indicates the parameter is part of the background component. The priors set for these parameters are shown next, followed by their initial guesses. The results of both `scipy.optimize.minimize` and `emcee` are shown in the last two columns.

Parameter (var in code) [units]	Prior	Initial guess	`scipy` results	`emcee` results
p_{stream} (`pstream`) [unitless]	$[0, 1]$	0.20	0.08	0.10 ± 0.01
$v_{\text{GSR},1}$ (`v1`) [km s^{-1}]	$[-160, 100]$	-120	-132	-131 ± 1
$v_{\text{GSR},2}$ (`v2`) [km s^{-1}]	$[-1, 4]$	0.07	-0.05	2.29 ± 0.57
$v_{\text{GSR},3}$ (`v3`) [km s^{-1}]	$[0, 10]$	5.68	2.88	2.30 ± 0.60
$\log_{10}[\sigma_{v_{\text{GSR}}}]$ (`lsigv`) [km s^{-1}]	$[-1, 4]$	1.40	0.72	0.67 ± 0.04
$(\mu_\alpha \cos\delta)_1$ (`pmra1`) [mas yr^{-1}]	$[-1, 1]$	-0.10	-0.19	-0.17 ± 0.02
$(\mu_\alpha \cos\delta)_2$ (`pmra2`) [mas yr^{-1}]	$[-1, 1]$	0.24	-0.31	-0.32 ± 0.02
$(\mu_\alpha \cos\delta)_3$ (`pmra3`) [mas yr^{-1}]	$[-1, 1]$	0.09	-0.08	-0.08 ± 0.02
$\log_{10}[\sigma_{\mu_\alpha \cos\delta}]$ (`lsigpmra`) [mas yr^{-1}]	$[-1, 3]$	-0.20	-0.86	-0.84 ± 0.04
$\mu_{\delta,1}$ (`pmdec1`) [mas yr^{-1}]	$[-2, 0]$	-0.96	-0.99	-0.98 ± 0.02
$\mu_{\delta,2}$ (`pmdec2`) [mas yr^{-1}]	$[-1, 1]$	0.07	-1.22	-0.07 ± 0.01
$\mu_{\delta,3}$ (`pmdec3`) [mas yr^{-1}]	$[-1, 1]$	0.07	0.05	0.01 ± 0.01
$\log_{10}[\sigma_{\mu_\delta}]$ (`lsigpmdec`) [mas yr^{-1}]	$[-2, 3]$	-0.50	-0.85	-0.98 ± 0.04
$v_{\text{GSR,b}}$ (`bv`) [km s^{-1}]	$[-50, 50]$	20.0	10.9	-3.03 ± 3.25
$\log_{10}[\sigma_{v_{\text{GSR,b}}}]$ (`lsigbv`) [km s^{-1}]	$[-1, 4]$	2.00	1.99	2.00 ± 0.01
$(\mu_\alpha \cos\delta)_b$ (`bpmra`) [mas yr^{-1}]	$[0, 2]$	1.40	1.10	1.06 ± 0.05
$\log_{10}[\sigma_{(\mu_\alpha \cos\delta)}]$ (`lsigbpmra`) [mas yr^{-1}]	$[-1, 3]$	0.20	0.18	0.17 ± 0.01
$\mu_{\delta,b}$ (`bpmdec`) [mas yr^{-1}]	$[-4, 0]$	-2.50	-2.08	-2.09 ± 0.04
$\log_{10}[\sigma_{\mu_{\delta,b}}]$ (`lsigbpmdec`) [mas yr^{-1}]	$[-2, 3]$	0.00	0.01	0.04 ± 0.01

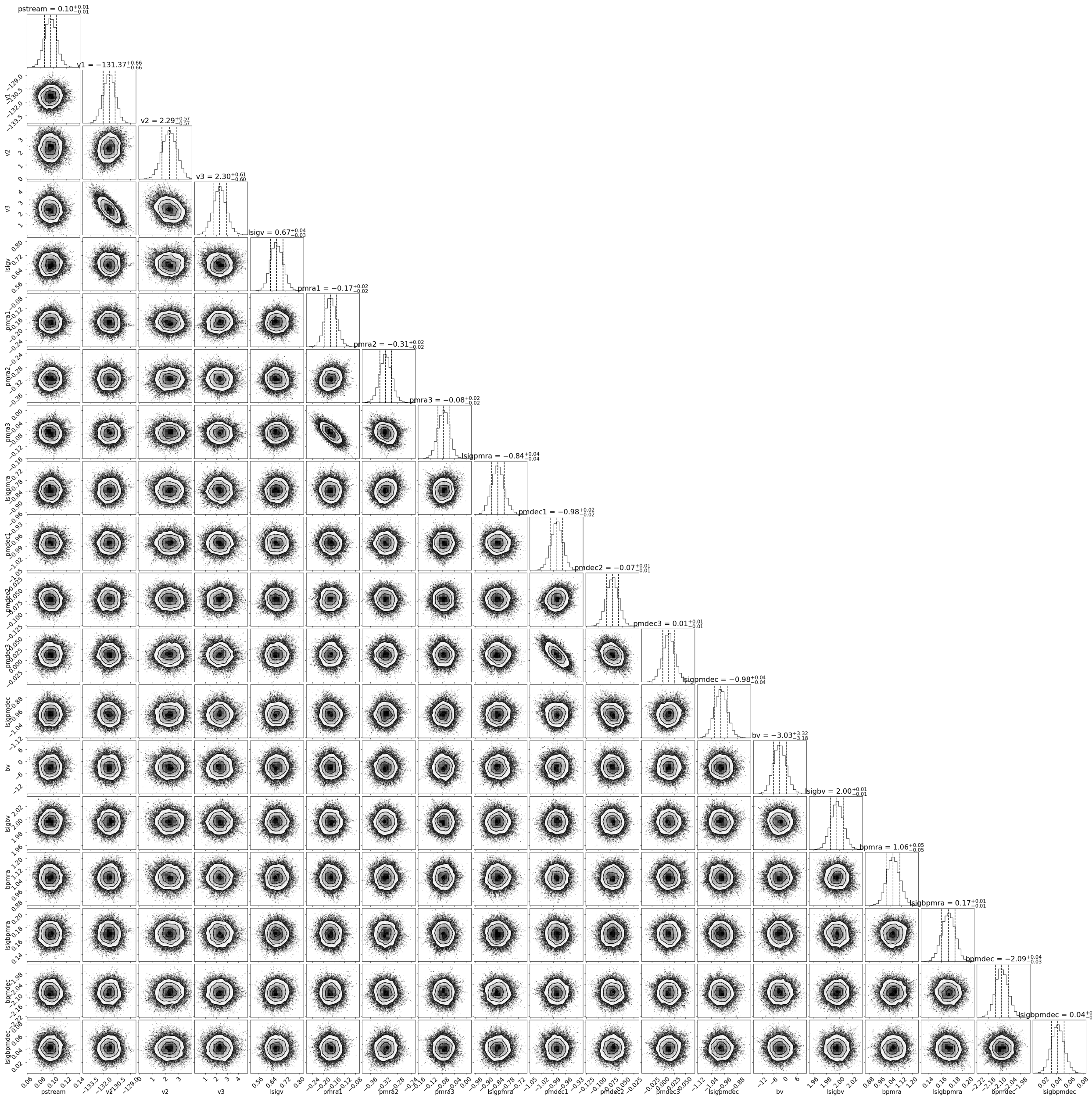

Figure 2: The corner plot of the results of the MCMC sampler. The probability function and procedure used is described in Section 3.

3.2 Membership probability

To find the membership probability, we can use the likelihood functions of the components $\mathcal{L}_{\text{stream}}, \mathcal{L}_{\text{background}}$ as defined in Section 3.1. The membership probability p is given by

$$p = \frac{p_{\text{stream}} \cdot \mathcal{L}_{\text{stream}}}{p_{\text{stream}} \cdot \mathcal{L}_{\text{stream}} + (1 - p_{\text{stream}}) \cdot \mathcal{L}_{\text{background}}}. \tag{3}$$

The membership probabilities of each star can be seen in Figure 3.

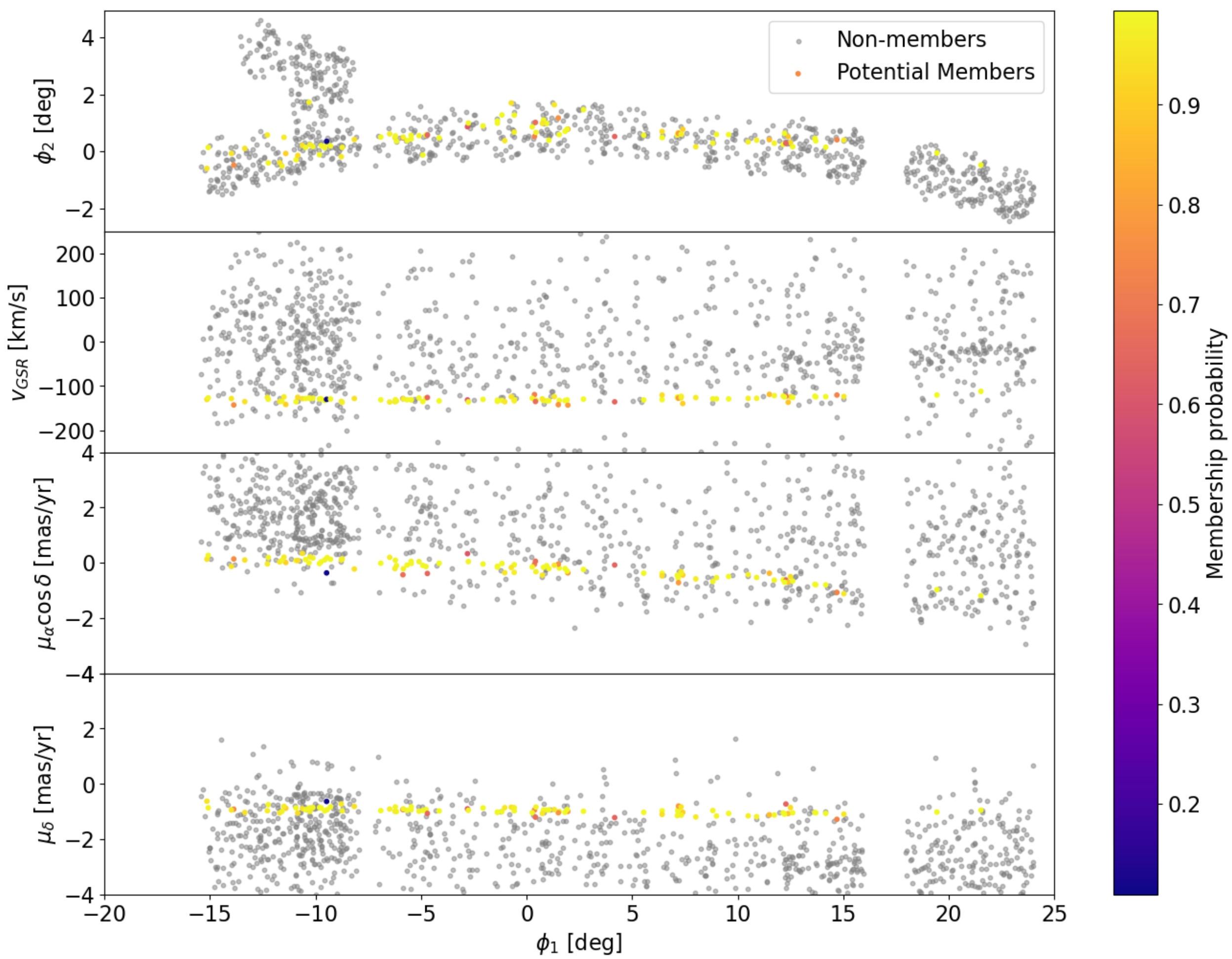

Figure 3: The membership probabilities of stars in the S^5 data set, displaying all $p > 0.1$ in accordance with the colorbar on the right. Displayed in grey are all stars where $p \leq 0.1$. The axes are identical to those in Figure 1.

4 Discussion

Overall the processes in this report ran smoothly. 92 stars were identified with $p > 0.7$, which can be seen in Figure 4. It can also be seen in this figure that there is considerable overlap between the identified stars and those determined in Li et al. (2021), which adds support to our findings. One potential issue can be seen in Figure 3, where there is a noticeable lack of members with $0.1 < p < 0.8$. The reasons for this are unclear and are worth mentioning.

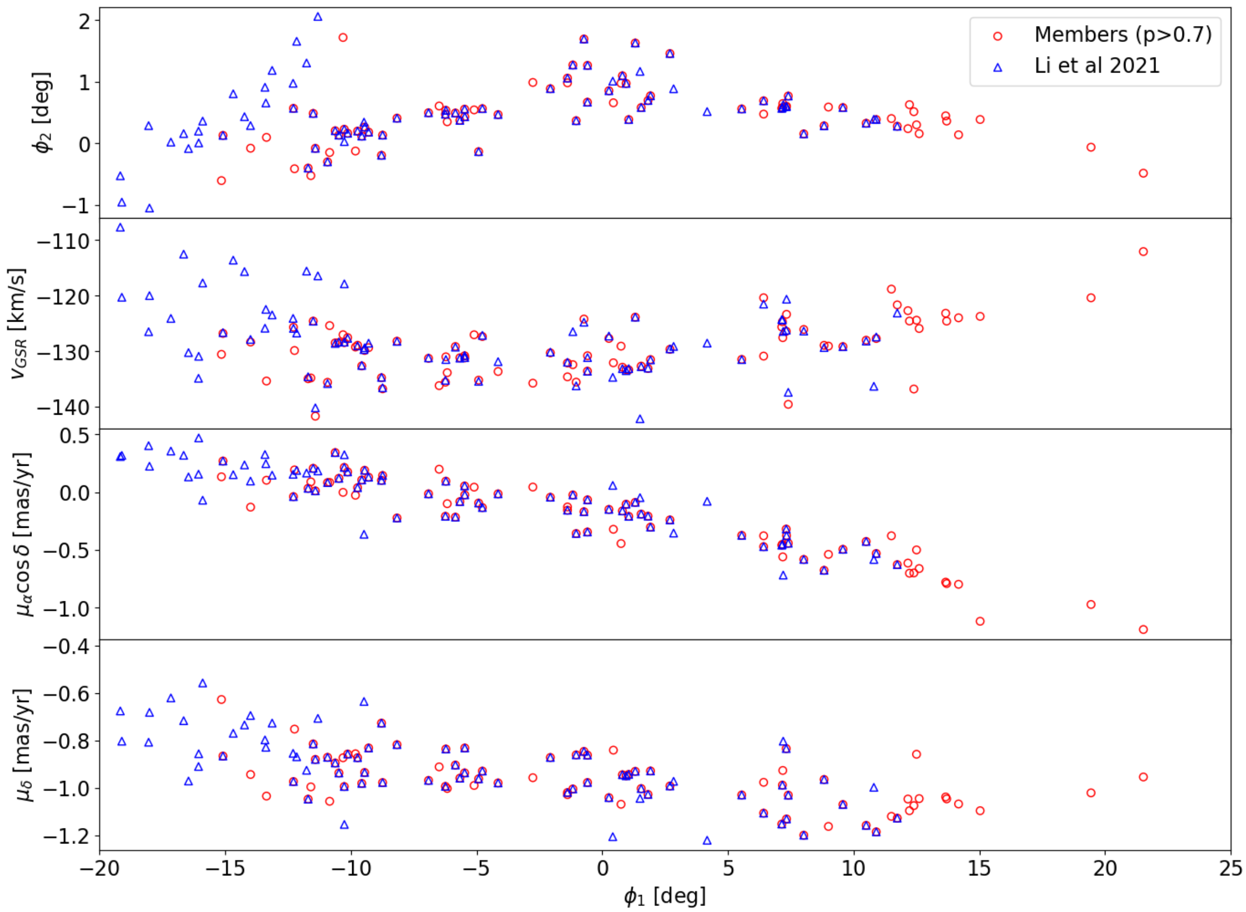

Figure 4: Members of stars in the S^5 data set ($p > 0.7$) compared to the identified members in Li et al. (2021), displayed as red circles and blue triangles respectively.

Some areas for improvement can be considered. Firstly, our data cuts eliminated around 87% of the sample population. The data cuts seem reasonable, as there is unlikely to be stream members at such high velocities or proper motions, but additional data could have been discarded nonetheless. Secondly, it is possible that a Gaussian fit for the background components may not be ideal. This can be seen in Figure 1, which shows distributions for proper motion that seem more complicated. Similarly, the quadratic functions for the means of the stream component may find more success and accordance with Li et al. (2021) if models with more degrees of freedom were considered. Lastly, the covariance between $\mu_\alpha \cos\delta$ and μ_δ is not considered, although the effect of this on our results should not be large.

5 Conclusion

This report was able to determine 92 members of the ATLAS stream using a Bayesian mixture model. Stars close to the ATLAS stream were selected from S^5 and quality cuts were made. The stream coordinates of each star was determined, and the heliocentric velocities were converted with respect to the Galactic Standard of Rest. A 19 parameter mixture model was constructed and sampled using MCMC methods to find the best fit parameters. This model was used to determine the membership probability of each star, and those with $p > 0.7$ were selected. Our results were in accordance with Li et al. (2021). Overall, this report was completed with relative success.

References

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, , 935, 167, doi: 10.3847/1538-4357/ac7c74

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, Publications of the Astronomical Society of the Pacific, 125, 306, doi: 10.1086/670067

Foreman-Mackey, D., Price-Whelan, A., Vousden, W., et al. 2023, dfm/corner.py: corner v2.2.2rc3, v2.2.2rc3, Zenodo, doi: 10.5281/zenodo.7808805

Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1, doi: 10.1051/0004-6361/201833051

Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, Nature, 585, 357, doi: 10.1038/s41586-020-2649-2

Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90

Li, T. S., Koposov, S. E., Zucker, D. B., et al. 2019, Monthly Notices of the Royal Astronomical Society, 490, 3508, doi: 10.1093/mnras/stz2731

Li, T. S., Koposov, S. E., Erkal, D., et al. 2021, , 911, 149, doi: 10.3847/1538-4357/abeb18

Price-Whelan, A. 2023, Convert a radial velocity to the Galactic Standard Of Rest (GSR), BSD. https://docs.astropy.org/en/stable/generated/examples/coordinates/rv-to-gsr.html

Shipp, N., Li, T. S., Pace, A. B., et al. 2019, The Astrophysical Journal, 885, 3, doi: 10.3847/1538-4357/ab44bf

Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nature Methods, 17, 261, doi: 10.1038/s41592-019-0686-2