DISENTANGLING THE COMPONENTS OF THE MILKY WAY

INFERRING THE STRUCTURE OF THE MILKY WAY IN PHASE-SPACE USING GAUSSIAN MIXTURE MODELLING WITH EXTREME DECONVOLUTION

Raunaq Singh Rai   —   MPhil Data Intensive Science   —   University of Cambridge

Motivation and Scientific Justification

An important question in Galactic Archaeology is *when* the Milky Way's disc first settled. Standard models suggest that the disc formed relatively late, after the interstellar medium had been enriched by multiple generations of stars. This implies a lack of stars with disc-like kinematics at very low metallicities. Since metallicity generally increases over cosmic time, it can be used as a rough proxy for stellar age: lower metallicity stars are typically older, having formed when the Galaxy was less chemically evolved. Thus, identifying a coherent, metal-poor disc population would challenge the "late–disc" paradigm and force a reassessment of the balance between in-situ star formation and accreted growth.

Zhang et al. (2024) [1] addressed this question by applying Extreme-Deconvolution Gaussian Mixture Modelling (XD-GMM) to Gaia DR3 red giants, and found no evidence for a cold, rotation-supported disc at very low metallicity. However, their analysis did not account for α abundance, which can be used as tracer of star formation timescales.

We reproduce their metallicity-binned XD-GMM using the same bright RGB catalogue, and extend the work by dividing stars into high and low α sequences following Viswanathan et al. (2024) [2]. Within each α branch, we perform XD-GMM fits in successive metallicity bins, selecting the optimal number of components via the Bayesian Information Criterion. By inspecting the weights and kinematic properties of these components, we assess whether disc-like signatures at low metallicity are genuine or attributable to contamination, noise, or accreted substructure.

Methodology

We use a cleaned sample of red giant branch (RGB) stars from Gaia DR3, with metallicities and α abundances from Andrae et al. Catalogue [3] and the Li et al. Catalogue [4], and distances from the Bailer-Jones et al. Catalogue [5]. As analysis largely depends on the accuracies of metallicities regions of high extinction, where XP spectra is known to be bias, are excluded with the sacrifice of losing a large proportion of RGB stars from subsequent anaysis..

The velocity distribution (v_R, v_ϕ, v_z) is modelled using Extreme-Deconvolution Gaussian Mixture Modelling (XD-GMM) [6][7], which accounts for observational uncertainties. We bin stars by metallicity and use the Bayesian Information Criterion (BIC) to determine the number of Gaussian components per bin. This allows us to identify structure without over fitting and introducing too many gaussians.

We extend the original method by splitting the sample into high and low α sequences [2], fitting separate XD-GMMs to each. By inspecting the component weights and kinematic properties, we quantify the emergence of rotational support and assess the presence of disc-like populations across both chemical tracks.

Following this, to determine the presence of disc-like kinematics, we compute the ratio of azimuthal velocity v_ϕ and its dispersion σ_ϕ for each component: a ratio approximately above 3 is considered to be disk-like. We then move to perform residual analysis to test whether disc-like kinematics are present but are too weak to be detected by XD-GMM.
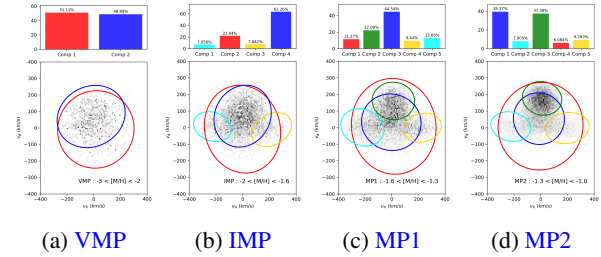
Key Findings

(a) VMP   (b) IMP   (c) MP1   (d) MP2

Figure 1: XD–GMM decomposition of red giant kinematics across metallicity bins.

Below $[\mathrm{M/H}] \sim -2$, no disc-like component is detected. The kinematics are fully explained by two broad halo Gaussians: one stationary and one mildly prograde likely associated with the Aurora population [8]. In the intermediate metallicity bin, two additional highly radial, non-rotating components emerge, linked to the Gaia–Sausage/Enceladus (GS/E) merger [9]. These represent remnants of a major head-on collision between the Milky Way and a massive dwarf galaxy, which deposited stars on radial orbits and contributed significantly to the inner halo's anisotropic structure.

Disc-like rotation only emerges above $[\mathrm{M/H}] \sim -1.6$: a distinct prograde component with relatively low dispersion, consistent with an early thick disc is observable. This component strengthens with increasing metallicity, supporting a gradual buildup of ordered rotation as the Galaxy chemically enriched.

In the MP1 bin (around $[\mathrm{M/H}] \sim -1.5$), a thick disc

begins to appear, rotating at $\langle v_\phi \rangle \approx 160$ km s^{-1} with $\sigma_\phi \approx 57$ km s^{-1}, yielding $V_{\rm rot}/\sigma_\phi \approx 2.8$ (just below the conventional "discy" threshold of ~ 3) accounting for $\sim 22\%$ of the stars. By the MP2 bin (higher metallicity), the disc becomes the dominant component (37.4% by weight) and surpasses the discy criterion with $V_{\rm rot}/\sigma_\phi \approx 3.5$, suggesting a well-established, rotation-supported structure.

Monte Carlo residual analysis in the VMP and IMP bins detects only minor excesses, constraining any hidden low-metallicity disc to at most $\sim 1\%$ of the sample. Together, these diagnostics suggest the absence of a disc below $[\rm M/H] \sim -1.6$.

Extending our analysis, we follow the method proposed by Viswanathan et al. [2] to separate our sample into high and low α sequences. This α abundance reflects star formation timescales: high α stars typically formed rapidly in intense, early starbursts, while low α stars formed more gradually over extended periods, often in lower density environments.

Figure 2 shows how these two populations differ in their kinematic evolution. In the high α branch, we observe a smooth increase in mean azimuthal velocity and a concurrent decrease in dispersion with increasing metallicity, indicating a gradual spin-up from a dynamically hot halo to a thick disc. In contrast, the low α population undergoes a sharper transition: rotation and cooling only emerge above $[\rm M/H] \sim -1.3$, consistent with a younger, thin disc origin.

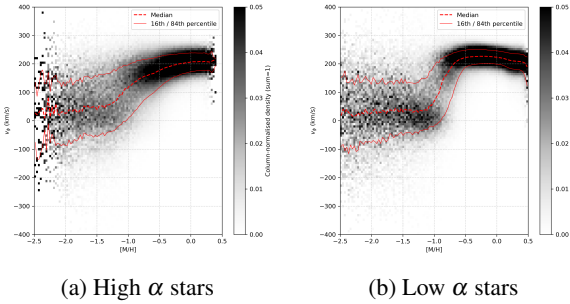(a) High α stars   (b) Low α stars

Figure 2: Median v_ϕ and dispersion vs. metallicity, split by α-sequence.

We applied XD–GMMs to the high- and low-α sequences separately (Figure 3). In the *high-α* track, only a non-rotating halo is present at $[\rm M/H] \lesssim -2$. A thick-disc Gaussian first appears at $-1.6 \lesssim [\rm M/H] \lesssim -1.3$ (31 % by weight), rotating at $\langle v_\phi \rangle \approx 148$ km s^{-1} with $\sigma_\phi \approx 63$ km s^{-1} ($V_{\rm rot}/\sigma_\phi \approx 2.4$). By $-1.3 < [\rm M/H] < -1.0$ it grows to 52 % and $V_{\rm rot}/\sigma_\phi \approx 2.8$, charting the gradual build-up of the thick disc from $[\rm M/H] \sim -1.6$.

In the *low-α* branch no disc is seen until $-1.3 < [\rm M/H] < -1.0$, where a thin, cold component (25 %) emerges abruptly with $\langle v_\phi \rangle \approx 183$ km s^{-1}, $\sigma_\phi \approx 48$ km s^{-1} and $V_{\rm rot}/\sigma_\phi \approx 3.8$—well above the discy

threshold of 3—implying a later, rapid thin-disc formation.

GS/E-like radial Gaussians appear in both sequences, probably due to misclassified α-poor debris, underscoring the need for robust chemical labelling. Overall, the high-α thick disc assembled earlier and gradually ($[\rm M/H] \gtrsim -1.6$), while the low-α thin disc formed later and swiftly ($[\rm M/H] \gtrsim -1.3$).
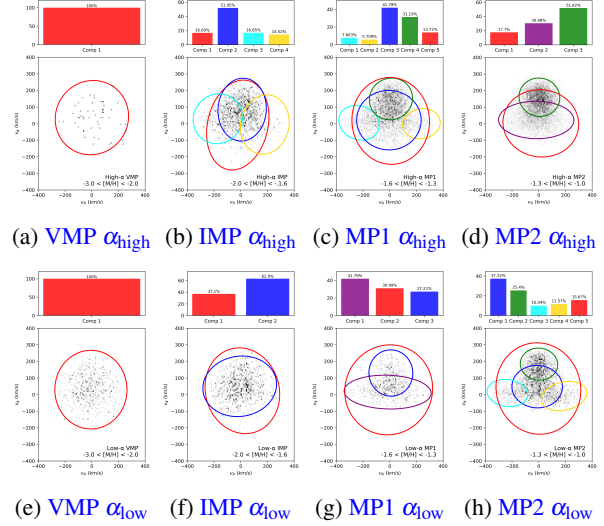

(a) VMP $\alpha_{\rm high}$ (b) IMP $\alpha_{\rm high}$ (c) MP1 $\alpha_{\rm high}$ (d) MP2 $\alpha_{\rm high}$

(e) VMP $\alpha_{\rm low}$ (f) IMP $\alpha_{\rm low}$ (g) MP1 $\alpha_{\rm low}$ (h) MP2 $\alpha_{\rm low}$

Figure 3: XD–GMM decomposition across α-sequences and metallicity bins. Top row: high-α; bottom row: low-α.

Recommendations and Next Steps

This project confirms the absence of disc-like kinematics below $[\rm M/H] \sim -2$ and finds earlier rotational support in the high α sequence. Future work should move beyond Gaussian mixtures, which cannot capture non-Gaussian or asymmetric structures, and consider models based on distribution functions. Addressing measurement uncertainties, particularly in α, will improve sequence classification, especially at low metallicities. Incorporating a formal selection function will also allow for global population estimates. The unexpected appearance of GS/E-like components in the high α sequence requires further investigation, likely reflecting contamination or substructure overlap.

Conclusion and Research Impact

We reproduce the key results of Zhang et al. [1] and extend them by separating stars by α abundance. Our findings support a two-phase disc formation model, with the high α sequence developing rotation at lower metallicity than the low α population. This demonstrates the power of combining chemistry and kinematics to trace the Milky Way's evolution and provides a framework for analysing future survey data.

References

[1] Hanyuan Zhang, Anke Ardern-Arentsen, and Vasily Belokurov. On the existence of a very metal-poor disc in the milky way, 2024.

[2] Akshara Viswanathan, Danny Horta, Adrian M. Price-Whelan, and Else Starkenburg. A slow spin to win – the gradual evolution of the proto-galaxy to the old disc, 2024.

[3] René Andrae, Hans-Walter Rix, and Vedant Chandra. Robust data-driven metallicities for 175 million stars from gaia xp spectra. *The Astrophysical Journal Supplement Series*, 267(1):8, jul 2023.

[4] Jiadong Li, Kaze W. K. Wong, David W. Hogg, Hans-Walter Rix, and Vedant Chandra. Aspgap: Augmented stellar parameters and abundances for 23 million rgb stars from gaia xp low-resolution spectra, 2023.

[5] C. A. L. Bailer-Jones, J. Rybizki, M. Fouesneau, M. Demleitner, and R. Andrae. VizieR Online Data Catalog: Distances to 1.47 billion stars in Gaia EDR3 (Bailer-Jones+, 2021). February 2021.

[6] Jo Bovy, David W. Hogg, and Sam T. Roweis. Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *The Annals of Applied Statistics*, 5(2B), June 2011.

[7] Peter Melchior and Andy Goulding. Filling the gaps: Gaussian mixture models from noisy, truncated or incomplete samples. *Astronomy and Computing*, 25, 11 2016.

[8] Vasily Belokurov and Andrey Kravtsov. From dawn till disc: Milky way's turbulent youth revealed by the apogee+gaia data. *Monthly Notices of the Royal Astronomical Society*, 514(1):689–714, May 2022.

[9] V Belokurov, D Erkal, N W Evans, S E Koposov, and A J Deason. Co-formation of the disc and the stellar halo. *Monthly Notices of the Royal Astronomical Society*, 478(1):611–619, June 2018.