

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 [1] [2]. 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) [3] addressed this question by applying Extreme-Deconvolution Gaussian Mixture Modelling (XD-GMM) to the full (v_ϕ, v_R, v_Z) phase-space of Gaia DR3 bright red giant stars, 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 serve as a 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) [4]. 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 the disc-like signatures for the high and low α sequences emerge at different metallicities in line with current understanding.

Methodology

After data cleaning and processing, the velocity distribution (v_R, v_ϕ, v_z) is modelled using Extreme-Deconvolution Gaussian Mixture Modelling (XD-GMM) [5][6], 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 [4], 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 disc-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

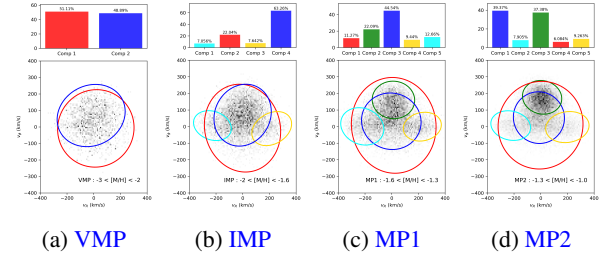


Figure 1: XD-GMM decomposition of red giant kinematics across metallicity bins. Links to the fully interactive 3-D plots for each bin are provided in the subcaptions.

Below $[M/H] \sim -2$, 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 [7]. In the intermediate metallicity bin, two additional highly radial, non-rotating components emerge, linked to the Gaia-Sausage/Enceladus (GS/E) merger [8]. 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 first appears at $[M/H] \gtrsim -1.6$, where a thick-disc component (approximately 22% of stars) rotates at $\langle v_\phi \rangle \approx 160 \text{ km s}^{-1}$ with $\sigma_\phi \approx 57 \text{ km s}^{-1}$ ($V_{\text{rot}}/\sigma_\phi \approx 2.8$, just below the “discy” threshold of ~ 3). As metallicity increases to the MP2 bin, this component grows to 37.4% of the sample and surpasses the discy criterion with $V_{\text{rot}}/\sigma_\phi \approx 3.5$, indicating a well-established, rotation-supported thick disc.

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 $[M/H] \sim -1.6$.

Following Viswanathan et al. [4], we split the sample into high- and low- α sequences—proxies for rapid versus extended star-formation timescales. Figure 2 shows that high- α stars spin up gradually, with $\langle v_\phi \rangle$ rising and σ_ϕ falling steadily from $[M/H] \sim -2$ to -1 , whereas the low- α branch remains dispersion-dominated until a sharp transition at $[M/H] \sim -1.3$, when disc kinematics abruptly appear.

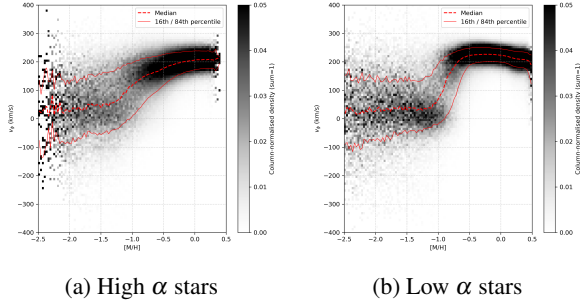


Figure 2: Median v_ϕ and dispersion vs. metallicity, split by α -sequence.

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 $[M/H] \lesssim -2$. A thick-disc Gaussian first appears at $-1.6 \lesssim [M/H] \lesssim -1.3$, with $V_{\text{rot}}/\sigma_\phi \approx 2.4$. By $-1.3 < [M/H] < -1.0$ it grows to 52% and $V_{\text{rot}}/\sigma_\phi \approx 2.8$, indicating a gradual build-up of the thick disc from $[M/H] \sim -1.6$.

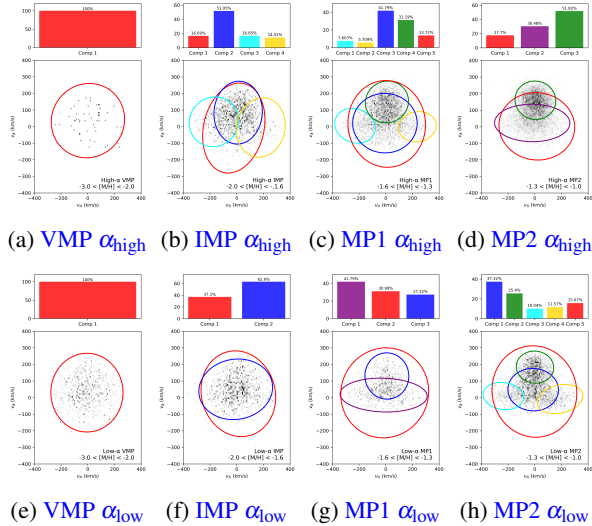


Figure 3: XD-GMM decomposition across α -sequences and metallicity bins. Top row: high- α ; bottom row: low- α .

In the *low- α* branch, a thick-disc component (25% of stars) only appears in the MP2 bin ($-1.3 < [M/H] < -1.0$), but its high $V_{\text{rot}}/\sigma_\phi \approx 3.8$, above the ~ 3 discy threshold, suggests contamination by a colder thin-disc

population. We do not observe a thin disc in the low- α sequence as expected due to sample selection effects.

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 ($[M/H] \gtrsim -1.6$), while the low- α thin disc formed later and swiftly ($[M/H] \gtrsim -1.3$).

Recommendations and Next Steps

- **Richer dynamical models.** Replace Gaussian mixtures with distribution-function or action-based models to capture non-Gaussian and asymmetric structures.
- **Tighter chemistry.** Improve α -abundance precision (or use high-resolution follow-up) to reduce sequence cross-contamination, especially at low metallicity.
- **Explicit selection function.** Model Gaia's magnitude and colour cuts to convert component weights into unbiased population fractions.
- **GS/E contamination.** Re-examine the high- α bins with stricter chemical cuts to confirm whether the apparent GS/E signatures are real or artefacts.

Conclusion and Research Impact

Splitting Gaia DR3 red giants by α abundance reveals a *two-phase* disc build-up: the high- α sequence gains rotational support at $[M/H] \approx -1.6$, forming a thick disc that grows gradually, while the low- α sequence does not reach disc kinematics until $[M/H] \approx -1.3$. This confirms that no metal-poor ($[M/H] < -2$) disc exists and clarifies the distinct evolutionary paths of the thick and thin discs.

Coupling precise chemistry with full 3-D kinematics provides a template for forthcoming surveys (WEAVE, 4MOST, SDSS-V) to isolate, date and map Milky-Way disc components. Pinpointing the metallicity- α thresholds for disc formation tightens constraints on early star-formation, feedback and merger heating in disc galaxies, advancing our reconstruction of the Galaxy's assembly history.

References

- [1] Federico Sestito, Tobias Buck, Else Starkenburg, Nicolas F Martin, Julio F Navarro, Kim A Venn, Aura Obreja, Pascale Jablonka, and Andrea V Macciò. Exploring the origin of low-metallicity stars in milky-way-like galaxies with the nihao-uhd simulations. *Monthly Notices of the Royal Astronomical Society*, 500(3):3750–3762, November 2020.
- [2] Alexander B. Gurvich, Jonathan Stern, Claude-André Faucher-Giguère, Philip F. Hopkins, Andrew Wetzel, Jorge Moreno, Christopher C. Hayward, Alexander J. Richings, and Zachary Hafen. Rapid disc settling and the transition from bursty to steady star formation in milky way-mass galaxies. *Monthly Notices of the Royal Astronomical Society*, 519(2):2598–2614, February 2023.
- [3] Hanyuan Zhang, Anke Ardern-Arentsen, and Vasily Belokurov. On the existence of a very metal-poor disc in the milky way, 2024.
- [4] 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.
- [5] 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.
- [6] Peter Melchior and Andy Goulding. Filling the gaps: Gaussian mixture models from noisy, truncated or incomplete samples. *Astronomy and Computing*, 25, 11 2016.
- [7] 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.
- [8] 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.