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

When did our Galaxy stop behaving like a chaotic proto-galaxy and settle into the orderly, rotating disc we see today? An important question in Galactic Archaeology is *when* the Milky Way's disc first settled. Standard models place this event relatively late, after the interstellar medium was enriched by multiple generations of stars [1, 2], implying a dearth of disc-like stars at very low metallicity. Because metallicity rises over cosmic time, it serves as a rough stellar clock: metal-poor stars are generally older. Finding a coherent, metal-poor disc would therefore overturn the *late-disc* paradigm and force a rethink of in-situ versus accreted growth.

A practical hurdle is that every Gaia DR3 velocity carries correlated uncertainties; if these "error ellipsoids" are ignored, genuine kinematic sub-structure can be blurred into (or out of) existence, producing components with no physical meaning.

Following the methodology of Zhang et al. 2024 [3], we revisit the same Gaia DR3 sample, now splitting the stars into high- and low-α sequences [4] and using an uncertainty-aware modelling approach to ask: *at what metallicity does disc-like rotation truly appear in each sequence?*

Methodology – Solution Path

To pinpoint *when* disc-like rotation truly emerges, we must see the **intrinsic** velocity field, stripped of Gaia's measurement blur and interpreted in its chemical context. Our five-step pipeline does exactly that, each step mapping onto a question posed in the Motivation.

1. **Uncertainty-aware modelling.** Every star's velocity (v_R, v_ϕ, v_z) carries a 3×3 covariance matrix—its error ellipsoid. We employ *extreme-deconvolution* Gaussian mixtures [5, 6] to find the set of noise-free Gaussians which, once convolved with those ellipsoids, reproduce the observed cloud. *Why:* de-blurring prevents spurious artefacts and ensures any recovered disc component is physically real.

2. **Metallicity slicing with parsimony.** Because [M/H] is an age proxy, we bin the sample in metallicity and let the Bayesian Information Criterion select the minimal number of Gaussians per bin. *Why:* isolates the precise metallicity—hence epoch—at which rotation support first appears, directly testing the late-disc paradigm.

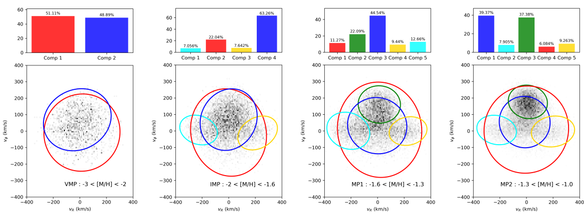
3. **α-sequence split.** We repeat the analysis separately for high- and low-α populations [4], tracers of rapid and prolonged star-formation histories. *Why:* reveals whether the thick and thin discs followed distinct evolutionary timelines.

4. **Disc diagnostic.** For each recovered Gaussian we compute $V_{\rm rot}/\sigma_\phi$; a ratio $\gtrsim 3$ flags a rotation-supported disc. *Why:* gives an objective, quantitative "disc/no-disc" verdict instead of subjective eyeballing.

5. **Residual stress-test.** We Monte-Carlo sample mock data from the best-fit model, convolve them with Gaia errors, and compare to the real sky. Any significant leftover overdensity signals structure the model missed. *Why:* confirms that no low-metallicity disc is hiding below our detection threshold, tightening the upper limit to $< 1\%$ of stars.

Together, these steps yield an unbiased, chemistry-aware map of the Milky Way's phase-space—precisely what we need to answer the core questions: *Is a metal-poor disc present at all, and if not, which stellar population spun up first?*

Key Findings

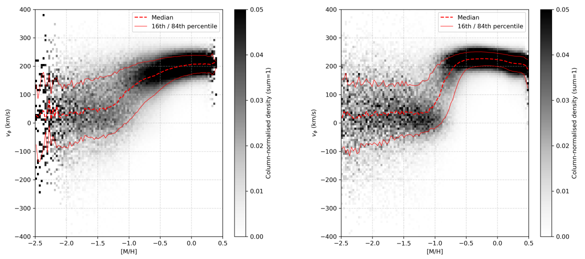


(a) VMP (b) IMP (c) MP1 (d) MP2

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 $[\mathrm{M/H}] \lesssim -2$ we detect **no rotation-supported disc**: the velocity field is fully described by two halo Gaussians—one almost stationary and one mildly prograde, the latter matching the Aurora structure [7]. At intermediate metallicities, a pair of highly radial, non-rotating components emerges, tracing debris from the Gaia–Sausage/Enceladus merger [8]. Disc-like rotation appears only for $[\mathrm{M/H}] \gtrsim -1.6$, where a thick-disc component (approximately 22% of stars) rotates at $\langle v_\phi \rangle \approx 160\,\mathrm{km\,s^{-1}}$ with $V_{\rm rot}/\sigma_\phi \approx 2.8$; by $[\mathrm{M/H}] \approx -1.3$ it grows to approximately 37% of the sample and surpasses the "discy" threshold with $V_{\rm rot}/\sigma_\phi \approx 3.5$. Monte Carlo residual tests confine any hidden low-metallicity disc to less than 1% of stars, confirming **the absence of a disc below** $[\mathrm{M/H}] \approx -1.6$.

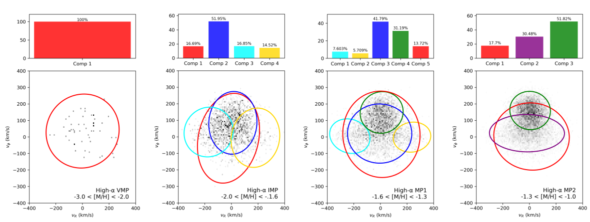
Following the method set by Viswanathan et al. [4], we split the sample into high- and low-α sequences which are proxies for rapid versus extended star-formation timescales. Figure 2 shows that high-α stars spin up gradually, with $\langle v_\phi \rangle$ rising and σ_ϕ falling steadily from $[\mathrm{M/H}]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



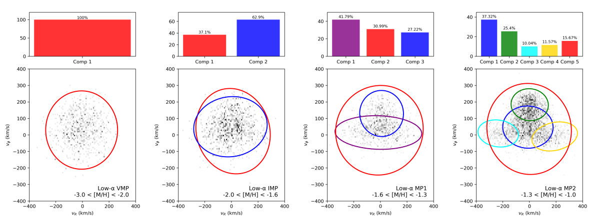
(a) High α stars (b) Low α stars

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



(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.

In the *low-α* branch, a thick-disc component (25% of stars) only appears in the MP2 bin ($-1.3 < [\mathrm{M/H}] < -1.0$), but its high $V_{\rm rot}/\sigma_\phi \approx 3.8$, 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.

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.
- **Chemical-abundance validation at low metallicity.** Gaia XP α-element measurements become unreliable below $[\mathrm{M/H}] \approx -1.5$, so cross-match our sample with high-resolution spectroscopic surveys (e.g., APOGEE, GALAH) to secure precise abundances and verify whether the apparent GS/E signatures persist.

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 $[\mathrm{M/H}] \approx -1.6$, forming a thick disc that grows gradually, while the low-α sequence does not reach disc kinematics until $[\mathrm{M/H}] \approx -1.3$. This confirms that no metal-poor ($[\mathrm{M/H}] < -2$) 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.

965 / 1000 words

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.