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

A central question in Galactic Archaeology is *when* the Milky Way's disc first settled. Standard models place disc formation after the interstellar medium had already been enriched, implying few (if any) stars on disc-like orbits below $[\mathrm{Fe/H}] \simeq -1.5$. Finding even a small, coherent very-metal-poor (VMP) disc would therefore overturn the "late–disc" paradigm and force a rethink of in-situ versus accreted growth.

Zhang et al. (2024) [1] tackled this problem with Extreme-Deconvolution Gaussian Mixture Modelling (XD-GMM) of Gaia DR3 red giants and reported *no* cold VMP disc. Their analysis, however, did not separate stars by α-abundance, a key tracer of formation timescale.

We reproduce their metallicity-binned XD-GMM on the same bright-RGB catalogue and extend it by splitting the sample into high- and low-α sequences (Viswanathan et al. 2024 [2]). For each chemical branch we fit XD-GMMs in successive metallicity bins, letting the Bayesian Information Criterion choose the minimum number of Gaussian components. Comparing the weights and kinematics of these components lets us decide whether any disc-like signal at low metallicity is genuine or an artefact of noise, misclassification, or accreted debris.

By tightening the chemical and kinematic tests in this way, we provide a more robust verdict on early disc formation and thereby refine constraints on the Milky Way's assembly history.

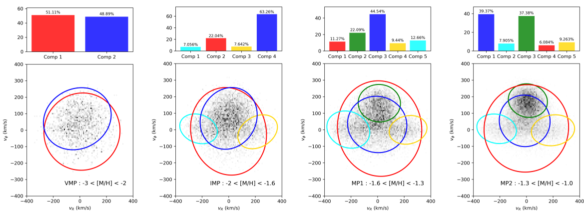
Methodology

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.

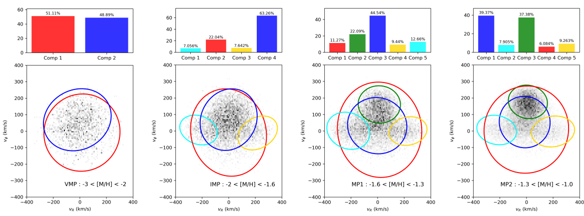
Key Findings



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

Figure 1: XD–GMM decomposition of red giant kinematics in v_R–v_ϕ across increasing metallicity bins: very metal-poor (VMP), intermediate (IMP), and two mildly metal-poor bins (MP1, MP2).

Key Findings



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

Figure 2: XD–GMM decomposition of red giant kinematics across metallicity bins.

Below $[\mathrm{M/H}] \sim -1.6$, no disc-like component is detected. The kinematics are fully explained by two broad halo Gaussians—one stationary and one mildly prograde. In the intermediate metallicity bin (IMP), two additional radial components appear, associated with the Gaia–Sausage/Enceladus merger. These are kinematically hot and non-rotating, and contribute to halo anisotropy.

Disc-like rotation only emerges above $[\mathrm{M/H}] \sim -1.3$, where we identify a cold, prograde component consistent with an early thick disc. This component strengthens with increasing metallicity, supporting a gradual build-up of ordered rotation as the Galaxy chemically enriched.

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). VizieR Online Data Catalog: I/352. Originally published in: 2021AJ....161..147B, 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.