

Thick-to-thin disc transition and gas disc shrinking induced by the Gaia-Sausage-Enceladus merger

Natsuki Funakoshi,^{1*} Daisuke Kawata,¹ Jason L. Sanders,² Ioana Ciucă,³ Robert J. J. Grand,⁴ HanYuan Zhāng⁵

¹Space and Climate Physics, Mullard Space Science Laboratory, University College London Holmbury St. Mary, Dorking, Surrey, RH5 6NT, UK

²Department of Physics and Astronomy, University College London, London, WC1E 6BT, UK

³Kavli Institute for Particle Astrophysics and Cosmology, Stanford University, 452 Lomita Mall, Stanford, CA 94305, UK

⁴Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK

⁵Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Understanding the Milky Way disc formation requires characterising its structural and kinematic properties as functions of stellar age. Using red giant stars from APOGEE DR17 and *Gaia* DR3, we model the age-dependent stellar kinematics with a quasi-isothermal distribution function and fit disc parameters as a function of age using non-parametric splines. We identify a transition from thick to thin disc populations around 10 Gyr ago. Stars older than this have short scale lengths (~ 1.7 kpc), typical of the thick disc, while younger stars exhibit increasing scale length with decreasing age, consistent with inside-out formation of the thin disc. This transition possibly coincides with the end of the starburst triggered by the Gaia-Sausage-Enceladus (GSE) merger. Stars formed around 10 Gyr ago exhibit a dip in scale length, even shorter than that of the thick disc. Comparison with an Auriga simulation suggests that this scale-length dip reflects gas disc shrinking caused by the transition from a cold to hot gas accretion mode. We propose the following disc formation scenario: (1) the thick disc formed under cold-mode accretion; (2) the GSE merger triggered a starburst and increased the total mass of the Galaxy, causing the transition to hot-mode accretion; (3) rapid gas consumption led to temporary shrinking of the star-forming gas disc; and then (4) thin disc grows in an inside-out fashion, as the size of the star-forming gas disc grows via hot-mode smooth gas accretion.

Key words: Galaxy: disc – Galaxy: evolution – Galaxy: kinematics and dynamics

1 INTRODUCTION

Gaia-Sausage-Enceladus (GSE) is believed to be the last significant merger event experienced by the Milky Way. The GSE remnants are identified through the kinematics of metal-poor halo stars typically with low- $[\alpha/\text{Fe}]$ abundances (Belokurov et al. 2018; Helmi et al. 2018; Haywood et al. 2018). The merger is believed to have occurred between 8 and 11 Gyr ago (e.g., Helmi 2020; Mackereth et al. 2019; Di Matteo et al. 2019; Gallart et al. 2019; Belokurov et al. 2020; Montalbán et al. 2021) and had a significant impact on the Galactic disc (e.g., Belokurov et al. 2020; Grand et al. 2020; Renaud et al. 2021).

Observations show that the Galactic disc consists of two distinct components (Yoshii 1982; Gilmore & Reid 1983), although there is ongoing debate as to whether the thin and thick discs should be regarded as truly distinct components (e.g., Bensby et al. 2007; Bovy et al. 2012a; Kawata & Chiappini 2016; Hayden et al. 2017). The first is the high- $[\alpha/\text{Fe}]$ disc, which is older and geometrically thicker—hence referred to as the “thick disc”. The second is the younger low- $[\alpha/\text{Fe}]$ disc, which is typically geometrically thin and

known as the “thin disc”¹ (e.g., Fuhrmann 1998; Prochaska et al. 2000; Feltzing et al. 2003; Bensby et al. 2014; Hayden et al. 2015). Recent years have seen theories develop that connect together the GSE merger with the configuration of the Galactic discs. By analysing a series of high-resolution cosmological magnetohydrodynamical simulations (Auriga: Grand et al. 2017, 2024), Grand et al. (2020) showed that a gas-rich merger can trigger a central starburst, leading to a transition from thick to thin disc formation (see also Brook et al. 2004, 2012b). Motivated by this result, it has been proposed that the GSE merger may have caused a similar transition in the Milky Way.

This theoretical scenario finds support in observational results. Ciucă et al. (2024) analysed red giant stars from APOGEE (The Apache Point Observatory Galactic Evolution Experiment), whose ages were estimated using BINGO (Bayesian INFerence for Galactic archaeOlogy), a Bayesian machine learning framework developed by Ciucă et al. (2021). Their analysis revealed a distinct phase of rapid

¹ Flaring in the outer disc means the low- $[\alpha/\text{Fe}]$ population can be geometrically thick in some parts of the Galaxy (e.g., Kawata & Chiappini 2016; Kawata et al. 2024, for reviews). However, for convenience, we refer to the chemically defined older high- $[\alpha/\text{Fe}]$ disc and younger low- $[\alpha/\text{Fe}]$ disc as the thick and thin discs, respectively.

* E-mail: n.funakoshi@ucl.ac.uk

increase of $[\text{Fe}/\text{H}]$ and decrease of $[\alpha/\text{Fe}]$ in stars aged between 12 and 10 Gyr (on the *relative* age scale used in Ciucă et al. 2021). They interpreted this as evidence for a major starburst episode, which they named the Great Galactic Starburst (GGS). This event was likely driven by a rapid gas inflow that began around 12 Gyr ago and continued until roughly 10 Gyr ago. Comparisons with Auriga simulations suggest that the GGS was likely driven by the gas-rich GSE merger event. Ciucă et al. (2024) indicated that after this gas-rich merger event, the high- $[\text{Fe}/\text{H}]$ and low- $[\alpha/\text{Fe}]$ thin disc population began to form. This suggests that the transition in the disc formation phase—from thick to thin—occurred after the gas-rich GSE merger.

A key question is whether *structural and kinematical* properties of the disc also exhibit a clear transition from the thick to the thin disc around the same epoch as GSE merger. Many studies have analysed the age-dependent structure of the Galactic disc to understand how the disc has evolved. For example, Frankel et al. (2019) modelled the ages, metallicities and radial distribution of APOGEE red clump stars and found clear evidence for inside-out growth of the low- $[\alpha/\text{Fe}]$ thin disc. Lian et al. (2023) estimated the surface brightness profile of the Milky Way as a function of stellar age. They revealed a broken radial structure, indicating a more complex and extended disc than previously assumed based on single exponential models (e.g., see also Mackereth et al. 2017; Lian et al. 2022). Khoperskov et al. (2025) used APOGEE DR17 data and an orbit superposition modelling. They identified two distinct age-metallicity sequences in the Milky Way disc, which they associated with the formation histories of an older inner disc and a younger outer disc. Imig et al. (2025) demonstrated that the disc scale length, scale height and flaring are strongly correlated with stellar age, highlighting the importance of time-dependent processes in the formation and evolution of the Milky Way disc.

Zhang & Sanders (2023) analysed the velocity distribution of the stellar disc as a function of stellar age to calibrate the period-age relation for O-rich Mira variable stars. They estimated the structural properties of the disc by fitting the velocity distributions of individual stars with an action-based distribution function (hereafter, DF) after dividing their sample into different age bins. This method does not require knowledge of the stellar density distribution and is less affected by spatial selection bias in observational data, since the kinematics of stars are unlikely to depend on the observational selection function.

Following the approach of Zhang & Sanders (2023), we investigate the kinematical distribution of red giant stars in APOGEE DR17 as a function of age. We use stellar ages measured by Ciucă et al. (2024) and stellar kinematics from APOGEE and *Gaia* DR3. By fitting action-based DFs as a function of age, for the first time we successfully trace the age dependence of the structural and kinematic properties of the Galactic disc seamlessly over the full stellar age range. Particular attention is paid in the age dependence of the scale length of the radial surface density profile of the Galactic disc. Grand et al. (2018) suggested that the transition from thick to thin disc formation phase involved a shrinking of the star-forming gas disc analysing an Auriga cosmological simulation. They showed that the transition of the gas accretion mode contributes to the thick-to-thin disc formation phase transition and this mode transition causes a temporary gas supply gap, which leads to the gas shortage and shrinking of the star-forming gas disc. If this is the case, the scale length of the stars formed in this transition phase would be smaller in the scale length, compared to the other period.

In Section 2, we describe the dataset of red giant stars used in this study. Section 3 outlines our modelling method based on an action-based quasi-isothermal DF. Section 4 presents our results of

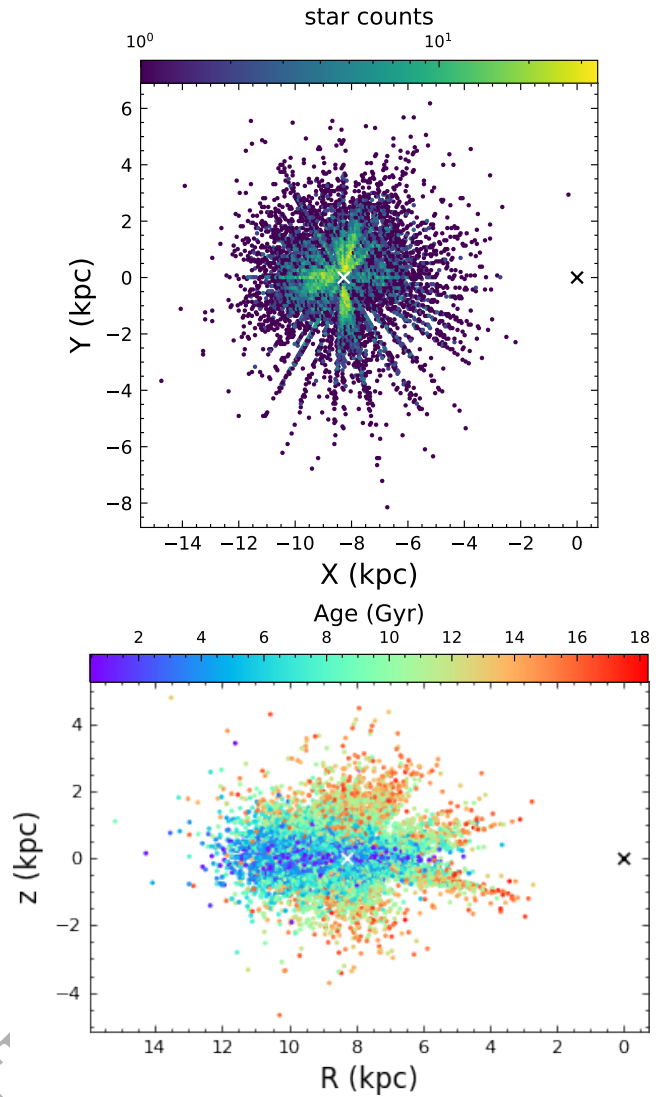


Figure 1. Spatial distribution of the red giant stars used in our study. The upper panel shows the face-on view of the stellar density distribution coloured with the logarithmic density. The lower panel presents the distribution in Galactocentric radius and vertical height, where colour denotes stellar age. The position of the Sun is indicated by a white cross in both panels, located at $(X, Y) = (-8.275, 0)$ kpc in the upper panel and $(R, z) = (8.275, 0.0208)$ kpc in the lower panel. The Galactic centre is marked by a black cross.

fitting the observational data with the DFs as a function of age. Our discussion in Section 5 is divided into three parts: in Section 5.1, we analyse which physical properties in the data drive the trends found in the DF fitting results; in Section 5.2, we apply the same DF fitting method to Auriga simulation data and compare the resulting trends with those observed in the Milky Way; in Section 5.3, we discuss the implications of our results in the context of the evolution of the Galaxy. Section 6 provides a summary of this study.

2 DATA

The red giant sample was taken from Ciucă et al. (2024). Ciucă et al. (2021) developed BINGO, a supervised Bayesian Neural Network to estimate stellar ages and their uncertainties of red giant stars

using APOGEE-2 DR17 (Nidever et al. 2015; Majewski et al. 2017; Wilson et al. 2019). The input features are stellar parameters T_{eff} , $\log g$ and $[\text{Fe}/\text{H}]$, and chemical abundance ratios, $[\text{C}/\text{Fe}]$, $[\text{N}/\text{Fe}]$ and $[\text{Mg}/\text{Fe}]$, which were derived with APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP; García Pérez et al. 2016) using the line lists described in Shetrone et al. (2015) and Smith et al. (2021). Its training data were sourced from the APOKASC-2 stars with asteroseismic ages calculated by Miglio et al. (2021). Since asteroseismic ages are more reliable for stars on the Red Giant Branch (RGB) and high-mass ($> 1.8 M_{\odot}$) Red Clump (RC) stars, the training data were restricted to these populations. In addition, Ciucă et al. (2024) applied quality cuts of ASPCAPFLAG=0, signal-to-noise ratio > 100 , $1 < \log g < 3.5$ and $4000 \text{ K} < T_{\text{eff}} < 5500 \text{ K}$. Further, they developed a classifier to select only the RGB and the high-mass RC stars, and applied it to the APOGEE data to select the stars similar to the training data. After applying the trained BINGO model to these stars, they estimated ages for 89,591 stars with their statistical uncertainties. These estimated ages are independently validated by their consistency with spectroscopic ages of APOGEE red giant stars derived using XGBoost (Chen & Guestrin 2016) model trained on asteroseismic data (see Fig. 6 of Anders et al. 2023). It should be noted that some stars in the sample have estimated ages older than the age of the Universe. This arises from the non-informative age prior used by Miglio et al. (2021), which has a high upper limit of 40 Gyr. Nevertheless, Ciucă et al. (2021) discussed that the age estimates are reliable in terms of relative age. Hence, throughout this paper, we adopt the specific age scale used in BINGO, which may not correspond to the absolute age scale of the Universe.

We cross-matched the APOGEE sample with the *Gaia* DR3 sources (Gaia Collaboration et al. 2023), to obtain proper motions and parallaxes. V_{helio} from the APOGEE catalogue was adopted as the radial velocity. We applied quality cuts on the *Gaia* data by selecting the data with $\text{parallax}/\text{parallax_error} > 5$ and $\text{RUWE} < 1.4$ (Lindgren et al. 2021a). We only included stars with an uncertainty of less than 0.02 dex in the base-10 logarithmic stellar age estimate and a probability higher than 95% of being an RGB or high-mass RC star. Note that the age uncertainty of BINGO represents aleatoric uncertainty of the neural network model predictions. This uncertainty reflects how well the Neural Network model can replicate the training data, and can be smaller than the measured uncertainties of the training data (Ciucă et al. 2021). The age uncertainty of BINGO can still be used to select the stars with high-confidence age inference.

In addition, we removed stars with vertical distances greater than 5 times the dispersion of the vertical distribution of stars with similar ages. This process removed three stars with stellar ages of less than 1 Gyr. In total, 16,617 star samples were obtained. Most of the stars in our sample are disc stars with $[\text{Fe}/\text{H}] > -1.0$. This is because stars with lower metallicity ($[\text{Fe}/\text{H}] < -1.0$) have larger age uncertainties of BINGO. The training data include very few such low-metallicity stars due to the *Kepler* asteroseismic survey's narrow field of view, which is restricted to the Galactic disc. With so few examples in training data, the neural network cannot reliably constrain the ages for the low-metallicity stars. As a result, such stars rarely meet our age-uncertainty requirement (< 0.02 dex). This selection effect is, however, appropriate for our purposes, because it naturally reflects the training data, which mainly consist of Galactic disc stars, and keeps the sample focused on the disc population.

The spatial distribution of our selected data is shown in Fig. 1. Observational data are sparse within 5 kpc of the Galactic centre. Since the effects of the Galactic bar and bulge are likely to be prominent within this inner region, it is reasonable to exclude the bar-dominated

area from the analysis when using an axisymmetric model, as adopted in this study (see Section 3).

3 MODELLING

We aimed to fit the stellar kinematics data with a probability DF of stars, $p(\mu, v_{\parallel} | \ell, b, \varpi, \log_{10} \tau)$, where μ and v_{\parallel} are the proper motion vector and the line-of-sight velocity, respectively, ℓ and b are the Galactic longitude and latitude and τ is the stellar age. ϖ is the *Gaia* DR3 parallax, which was corrected for the zero-point offset based on Lindgren et al. (2021b) using the Python package `gaiadr3_zeropoint`². Since stellar density measurements are highly sensitive to observational selection functions and are difficult to obtain accurately, we focused on fitting (action-based) DFs using only stellar velocities, following Zhang & Sanders (2023) and Sanders et al. (2024).

We adopted $R_0 = 8.275 \text{ kpc}$ (GRAVITY Collaboration et al. 2021) as the Galactocentric radius of the Sun, and $z_0 = 20.8 \text{ pc}$ (Bennett & Bovy 2019) as its vertical height above the Galactic mid-plane. For the solar motion, we used $(u_0, v_0, w_0) = (9.65, 14, 8.59) \text{ km s}^{-1}$ (Almannaei et al. 2024), where u_0 , v_0 and w_0 are the components of the solar proper motion with respect to the Local Standard of the Rest in the direction of the Galactic centre, the Galactic rotation and the north Galactic pole, respectively. The circular velocity at the solar radius, $V_c(R_0)$, was taken from a fixed axisymmetric gravitational potential of the Milky Way described in Price-Whelan (2017) and was set to 231.21 km s^{-1} .

We begin by writing the probability DF as follows:

$$p(\mu, v_{\parallel} | \ell, b, \varpi, \log_{10} \tau) = \frac{p(\ell, b, \varpi, \mu, v_{\parallel}, \log_{10} \tau)}{p(\ell, b, \varpi, \log_{10} \tau)}, \quad (1)$$

where both the numerator and denominator have a cancelling contribution from the selection function. We accounted for uncertainties in the proper motion, line-of-sight velocity, parallax and stellar age by marginalizing over them. The numerator is evaluated as follows;

$$\begin{aligned} & p(\ell, b, \varpi, \mu, v_{\parallel}, \log_{10} \tau) \\ &= \int d^2 \mu' dv'_{\parallel} d\varpi' d\log_{10} \tau' \mathcal{N}(\mu' | \mu, \Sigma_{\mu}) \mathcal{N}(\varpi' | \varpi, \sigma_{\varpi}^2) \\ & \quad \mathcal{N}(v'_{\parallel} | v_{\parallel}, \sigma_{v_{\parallel}}^2) \mathcal{N}(\log_{10} \tau' | \log_{10} \tau, \sigma_{\log_{10} \tau}^2) p(\ell, b, \varpi', \mu', v'_{\parallel}, \log_{10} \tau'), \end{aligned} \quad (2)$$

where $\mathcal{N}(x | \mu, \sigma^2)$ denotes a Gaussian with mean μ and variance σ^2 . The sky position of stars, (ℓ, b) , was assumed to be accurately measured and was not included in the marginalisation. Then, we related the DF of the observable coordinates to a DF of action as follows:

$$p(\ell, b, \varpi', \mu', v'_{\parallel}, \log_{10} \tau) = \left| \frac{\partial(\mathbf{J}, \theta)}{\partial(\ell, b, \varpi, \mu, v_{\parallel})} \right| f_{\tau}(\mathbf{J}) \propto s^6 \cos b f_{\tau}(\mathbf{J}), \quad (3)$$

where $\mathbf{J} = (J_R, J_{\phi}, J_z)$ is the vector of actions corresponding to the observable coordinates in six dimensions (with the corresponding angle coordinates, θ) and s is the distance corresponding to the parallax. The factor $s^6 \cos b$ is derived from the Jacobian between the Galactic

² <https://pypi.org/project/gaiadr3-zeropoint/>

coordinates and Cartesian coordinates, (\mathbf{x}, \mathbf{v}) . The Jacobian between (\mathbf{x}, \mathbf{v}) and (\mathbf{J}, θ) is unity due to the canonical transformation. The factor $s^4 \cos b$ appears in the transformation of a volume element in three-dimensional position space, which includes $s = 1/\varpi$, while s^2 appears in the transformation of an area element in proper motion velocity space. Here, the subscript τ in $f_\tau(\mathbf{J})$ means that the parameters of the action-based DF are functions of stellar age τ , so that the DF varies with stellar age τ .

Adopting a quasi-isothermal DF, $f(\mathbf{J})$, from Binney (2010), we used an implementation provided in AGAMA (Vasiliev 2018), which takes the following form:

$$f_\tau(\mathbf{J}) = \frac{\tilde{\Sigma}\Omega}{2\pi^2\kappa^2} \times \frac{\kappa}{\tilde{\sigma}_r^2} \exp\left(-\frac{\kappa J_R}{\tilde{\sigma}_r^2}\right) \times \frac{\nu}{\tilde{\sigma}_z^2} \exp\left(-\frac{\nu J_z}{\tilde{\sigma}_z^2}\right) \times B(J_\phi),$$

$$B(J_\phi) = \begin{cases} 1 & \text{if } J_\phi \geq 0, \\ \exp\left(\frac{2\Omega J_\phi}{\tilde{\sigma}_r^2}\right) & \text{if } J_\phi < 0, \end{cases}$$

$$\tilde{\Sigma}(R_c) \equiv \Sigma_0 \exp(-R_c/R_{\text{disc}}(\tau)),$$

$$\tilde{\sigma}_r^2(R_c) \equiv \sigma_{R,0}^2(\tau) \exp(-2(R_c - R_0)/R_{\sigma,R}(\tau)),$$

$$\tilde{\sigma}_z^2(R_c) \equiv \sigma_{z,0}^2(\tau) \exp(-2(R_c - R_0)/R_{\sigma,z}(\tau)),$$
(4)

where R_c is the guiding radius, defined as the radius of a circular orbit with angular momentum, $J_\phi \equiv L_z = R_c \times V_c(R_c)$. Note that L_z is a conserved quantity in axisymmetric potentials and can also be expressed as $L_z = R \times V_\phi$ using the star's current Galactocentric radius and azimuthal velocity. The quantities κ, Ω and ν denote the epicyclic, angular and vertical frequencies of the orbit, respectively. The DF is described with five parameters: $R_{\text{disc}}, \sigma_{R,0}, \sigma_{z,0}, R_{\sigma,R}$ and $R_{\sigma,z}$, where R_{disc} is the radial scale length, $\sigma_{R,0}$ and $\sigma_{z,0}$ are the normalisation parameters for the radial and vertical velocity dispersions of the stars at $R_c = R_0$ and $R_{\sigma,R}$ and $R_{\sigma,z}$ represent the radial scale lengths of the radial and vertical velocity dispersion profiles, respectively. Table 1 summarises these parameters along with the priors described in the following subsections. It is worth noting that these parameters describe the shape of the DF and approximate the behaviour of physical quantities, but they are not strictly equivalent to them (see also Binney & McMillan 2011; Bovy & Rix 2013). For example, the radial scale length parameter, R_{disc} , controls the exponential fall-off of the DF in guiding radius, but the actual stellar density profile is also influenced by the shape of the gravitational potential. Nevertheless, the resulting stellar density profile is still approximately exponential, with a scale length similar to R_{disc} (e.g., Fig. 6 of Binney 2010). Although previous studies have shown that the radial surface density profile for mono-age disc population is better described by a broken exponential, especially for the low- $[\alpha/\text{Fe}]$ disc (e.g., Mackereth et al. 2017; Lian et al. 2024), here we model the disc with a single exponential scale length. This enables us to trace the stellar age dependence of the overall structure of the disc with fewer parameters.

We adopted a fixed axisymmetric gravitational potential of the Milky Way from Price-Whelan (2017) and computed the actions, $\mathbf{J} = \mathbf{J}(\mathbf{x}, \mathbf{v})$, of each star using the Stäckel fudge approximation as implemented in AGAMA.

3.1 Importance sampling

The denominator in equation (1), $p(\ell, b, \varpi, \log_{10} \tau)$, involves the Monte Carlo integrals over the full three-dimensional velocity space, making it computationally expensive. To compute the denominator

efficiently, we applied importance sampling, as described by Zhang & Sanders (2023) and Sanders et al. (2024). Here, a sample of the three-dimensional velocities, \mathbf{v} , used for integration was drawn from a probability DF, $G_\tau(\mathbf{v} | \ell, b, \varpi)$. This function is proportional to an assumed quasi-isothermal DF, $f'_\tau(\mathbf{J})$, with fixed parameters. Importance sampling integration is more accurate when the sampling function, $f'(\mathbf{J})$, closely resembles that of the function being integrated, $f(\mathbf{J})$. As explained in Section 4, we found that the parameters of our DF that best describe the observational data vary with stellar age and are expressed as an age-dependent function $f_\tau(\mathbf{J})$. Therefore, we used an age-dependent sampling function of $f'_\tau(\mathbf{J})$. Given a stellar age τ and a 3D position (ℓ, b, ϖ) , the probability DF can be drawn as follows:

$$G_\tau(\mathbf{v} | \ell, b, \varpi) = \frac{p(\ell, b, \varpi, \mathbf{v}, \log_{10} \tau)}{\int d^3\mathbf{v} p(\ell, b, \varpi, \mathbf{v}, \log_{10} \tau)} = \frac{f'_\tau(\mathbf{J})}{\int d^3\mathbf{v} f'_\tau(\mathbf{J})},$$
(5)

To minimise the bias in the Monte Carlo integration, we adjusted the parameters of $f'_\tau(\mathbf{J})$ so that it approximates the true distribution $f_\tau(\mathbf{J})$. This adjustment was done iteratively by repeating the entire Markov Chain Monte Carlo (MCMC) fitting process, each time updating the parameters of $f'_\tau(\mathbf{J})$ based on the results of the previous run.

The denominator on the right side of equation (5) can be computed using AGAMA. We generate a set of $N = 1280$ samples for each star using the MCMC performed with emcee (Foreman-Mackey et al. 2013), sampling velocities $\mathbf{v} = (v_x, v_y, v_z)$ based on $G_{\tau_k}(\mathbf{v} | \ell_k, b_k, \varpi_k)$ as the likelihood for the k -th star at (ℓ_k, b_k, ϖ_k) and τ_k . Also, $N = 1280$ random parallaxes and stellar ages are drawn from Gaussian distributions, $\mathcal{N}(\varpi_k, \sigma_{\varpi_k}^2)$ and $\mathcal{N}(\log_{10} \tau_k, \sigma_{\log_{10} \tau_k}^2)$, respectively, to account for the uncertainty in parallax and stellar age. The integral of the denominator in equation (1) for each star is then computed as follows,

$$p(\ell_k, b_k, \varpi_k, \log_{10} \tau_k) \approx \frac{A}{N} \times$$

$$\frac{\varpi_i \text{ from } \mathcal{N}(\varpi_k, \sigma_{\varpi_k}^2)}{\mathbf{v}_i \text{ from } G(\mathbf{v} | \dots)} \frac{\tau_i \text{ from } \mathcal{N}(\log_{10} \tau_k, \sigma_{\log_{10} \tau_k}^2)}{\sum_i s_i^4 \cos b \frac{f_{\tau_i}(\mathbf{J}(\ell_k, b_k, \varpi_i, \mathbf{v}_i))}{f'_{\tau_k}(\mathbf{J}(\ell_k, b_k, \varpi_k, \mathbf{v}_i))}},$$
(6)

where

$$A = \int d^3\mathbf{v} f'_{\tau_k}(\mathbf{J}(\ell_k, b_k, \varpi_k, \mathbf{v}))$$

$$= \sum_i \mathbf{v}_i \text{ from } G(\mathbf{v} | \dots) f'_{\tau_k}(\mathbf{J}(\ell_k, b_k, \varpi_k, \mathbf{v}_i)).$$
(7)

Here, k indexes stars and i indexes samples per star. The term $s^4 \cos b$ represents the Jacobian for transforming a Cartesian spatial volume into Galactic coordinates. The denominator of equation (6), $f'_{\tau_k}(\mathbf{J})$ is precomputed once since the parameters of f'_{τ_k} is fixed during the MCMC fitting process. The action variables \mathbf{J} used here are computed for each star using the fixed spatial coordinates (ℓ_k, b_k, ϖ_k) and a velocity component \mathbf{v}_i drawn from $G_{\tau_k}(\mathbf{v} | \ell_k, b_k, \varpi_k)$. On the other hand, for the numerator of equation (6), since the age dependence of the parameters for f_{τ_i} keeps being updated at each step of MCMC, $f_{\tau_i}(\mathbf{J})$ needs to be computed at every step of MCMC fitting. \mathbf{J} for each star is computed using sampled ϖ_i, \mathbf{v}_i and τ_i . The constant factor A defined in equation (7) is precomputed once for each star in the same manner as the denominator of equation (6).

Nevertheless, thanks to importance sampling, the computational cost of three-dimensional integration is effectively reduced to a simple summation, as shown in equation (6).

3.2 Likelihood and priors

For fitting the data, we maximise the following log-likelihood using MCMC,

$$\ln L = \sum_k^{\text{stars}} \ln p \left(\mu_k, v_{\parallel k} \mid \ell_k, b_k, \varpi_k, \log_{10} \tau_k \right), \quad (8)$$

for the five parameters: R_{disc} , $\sigma_{R,0}$, $\sigma_{z,0}$, $R_{\sigma,R}$, $R_{\sigma,z}$. Since our primary interest is to investigate how these parameters vary with stellar age, we model them as functions of age, τ (Gyr), i.e., $R_{\text{disc}}(\tau)$, $\sigma_{R,0}(\tau)$, $\sigma_{z,0}(\tau)$, $R_{\sigma,R}(\tau)$ and $R_{\sigma,z}(\tau)$. Their age dependence is modelled using flexible cubic spline functions, following Sanders et al. (2024). We found that 10 knots provides a good balance between capturing the age variation of these parameters and avoiding overfitting. These knots were evenly spaced across the age range between 0.46 to 18.26 Gyr, which are the minimum and maximum stellar ages in our observational data. The five parameters for the DFs are adjusted at these 10 knots to fit the data, resulting in a total of 50 free parameters. The following uniform prior distributions are assigned to all the knots,

$$\begin{aligned} \ln R_{\text{disc}}/\text{kpc} &\sim \mathcal{U}(\ln(0.01), \ln(8)), \\ \ln \sigma_{R,0}/\text{km s}^{-1} &\sim \mathcal{U}(\ln(10), \ln(100)), \\ \ln \sigma_{z,0}/\text{km s}^{-1} &\sim \mathcal{U}(\ln(10), \ln(100)), \\ \ln R_{\sigma,R}/\text{kpc} &\sim \mathcal{U}(\ln(1), \ln(100)), \\ \ln R_{\sigma,z}/\text{kpc} &\sim \mathcal{U}(\ln(1), \ln(100)). \end{aligned} \quad (9)$$

To suppress oscillations in values between spline knots, we introduced a smoothing prior P_{smooth} ,

$$\ln P_{\text{smooth}} = - \sum_{\theta} (R_{\text{disc}}, \sigma_{R,0}, \sigma_{z,0}, R_{\sigma,R}, R_{\sigma,z}) N_{\text{knots}}^{-1} \sum_{n=0}^{N_{\text{knots}}-1} \frac{(\ln \theta_{n+1} - \ln \theta_n)^2}{2(\sigma_{\theta})^2}, \quad (10)$$

where σ_{θ} is the scaling parameter, set to 0.3 for all the parameters, $\theta = (R_{\text{disc}}, \sigma_{R,0}, \sigma_{z,0}, R_{\sigma,R}, R_{\sigma,z})$, with θ_{n+1} and θ_n indicating the parameter values at the $(n+1)$ -th and n -th knots, respectively. N_{knots} is the number of knots, which was set to 10 as mentioned above. Adopting the smoothing prior distribution helps reduce the risk of over-fitting. For MCMC fitting, we used the codes developed by Sanders et al. (2024) and modified them for this application. This code uses Hamiltonian Monte Carlo with the NUTS sampler (Hoffman & Gelman 2011), implemented in NumPyro (Phan et al. 2019), which utilises JAX (Bradbury et al. 2018) to speed up the computation. Appendix A presents the verification of this model using the mock observational data, demonstrating that it recovers the ground truth within approximately 1σ across most of the age range. This validation indicates that the uncertainties in the parameter estimates tend to increase in the age range where the number of stars is small (i.e., $\tau < 2$ Gyr and $\tau > 14$ Gyr). Moreover, while $\sigma_{R,0}$ and $\sigma_{z,0}$ are relatively well constrained across the full age range, $R_{\sigma,R}$ and $R_{\sigma,z}$ are more difficult to constrain especially where their values are large. This is because once a scale length exceeds the radial range spanned by the data, the radial profile becomes effectively flat within the observed region, so that the scale length cannot be tightly determined.

4 RESULT

Panels (a)–(e) of Fig. 2 present the fitting results for our DF model parameters as a function of stellar age, described with a cubic spline with 10 knots. Note that these fitted parameters represent present-day values for different stellar age populations in the disc, rather than the conditions at the time of their formation. The original actions at birth may not be preserved if the gravitational potential of the system has undergone rapid changes, such as a major merger. Also, radial migration (Sellwood & Binney 2002) due to the bar and spiral arms leads to a change of the angular momentum of the stars. The solid blue line shows the median of the sampled spline curves, and the blue-shaded region shows the $\pm 1\sigma$ uncertainty range. The uncertainties in each parameter are mainly driven by the sample size. The dashed grey line shows the age-dependent functions adopted for f'_{τ} , which is used for the importance sampling in equation (5). We iteratively adjusted f'_{τ} to be similar to our derived f_{τ} , and our final parameters for f'_{τ} are described in equation (A1). Panel (f) shows the age distribution of APOGEE stars used in our analysis. The large number of stars younger than about 1.5 Gyr is due to the cut-off of low-mass RC stars. Table 2 shows the best-fit values and uncertainties for each parameter at individual spline knots. These are derived from the median and 16th and 84th percentiles of the posterior distribution function, respectively. Appendix B shows the distribution of actions for the observational data and those for the best-fit model shown with solid lines in Fig. 2, divided into four age groups. Overall, the observed distributions are reproduced by the best-fit model reasonably well.

The scale length of the disc, R_{disc} , remains small around 1.7 kpc for stellar populations older than 10 Gyr. For younger populations, the disc scale length increases with decreasing stellar age. This suggests that the age dependence of the disc scale length undergoes a transition around $\tau = 10$ Gyr. For stars older than 10 Gyr, the small scale length is consistent with that of the high- $[\alpha/\text{Fe}]$ thick disc previously reported (e.g., Bovy et al. 2012b). The increase in R_{disc} for the younger stars demonstrates the inside-out disc growth after 10 Gyr ago. This is consistent with the growth of the low- $[\alpha/\text{Fe}]$ thin disc indicated by previous studies (e.g., Bovy et al. 2012b; Frankel et al. 2019). Our modelling robustly identifies a clear transition from the thick to the thin disc traced by stars with ages around $\tau = 10$ Gyr, marking a key epoch in the Milky Way's disc evolution.

Both the radial and vertical velocity dispersions at the solar radius, $\sigma_{R,0}$ and $\sigma_{z,0}$, show similar trends. The velocity dispersion declines as the age decreases, consistent with previous studies (e.g., Casagrande et al. 2011; Silva Aguirre et al. 2018). We find that the vertical velocity dispersion, $\sigma_{z,0}$, declines steeply in stars with ages $12 < \tau < 10$ Gyr. This also indicates the transition from the thick disc, characterised by higher vertical velocity dispersion, to the thin disc, characterised by lower vertical velocity dispersion.

Both $R_{\sigma,R}$ and $R_{\sigma,z}$ also show a significant transition traced by stars with ages around 10 Gyr. Both parameters remain around 5–10 kpc for older stellar populations, and start to increase rapidly for stars aged around 10 Gyr. $R_{\sigma,R}$ reaches a peak of ~ 35 kpc at around 4 Gyr and $R_{\sigma,z}$ reaches a high value and stays high for $\tau \lesssim 4$ Gyr. These trends indicate that, for the thick disc population older than $\tau = 10$ Gyr, radial and vertical velocity dispersions show strong radial dependence, i.e., higher velocity dispersions in the inner disc. In the younger thin disc population, the radial dependence of the velocity dispersions becomes weaker, and their profiles flatten. This may be linked to the flaring of young stars in the outer disc (see the bottom panel of Fig. 1, e.g., Rahimi et al. 2014; Minchev et al. 2017), and may also be consistent with a scenario where the thin disc formed

Table 1. Summary of quasi-isothermal DF parameters described in equation (4) used in the fitting.

parameter θ	units	description	prior range	smoothing parameter σ_θ	figure panel
R_{disc}	kpc	radial scale length	(0.01, 8)	0.3	(a)
$\sigma_{R,0}$	km s^{-1}	radial velocity dispersion at $R_c = R_0$	(10, 100)	0.3	(b)
$\sigma_{z,0}$	km s^{-1}	vertical velocity dispersion at $R_c = R_0$	(10, 100)	0.3	(c)
$R_{\sigma,R}$	kpc	radial scale length of the radial velocity dispersion profiles	(1, 100)	0.3	(d)
$R_{\sigma,z}$	kpc	radial scale length of the vertical velocity dispersion profiles	(1, 100)	0.3	(e)

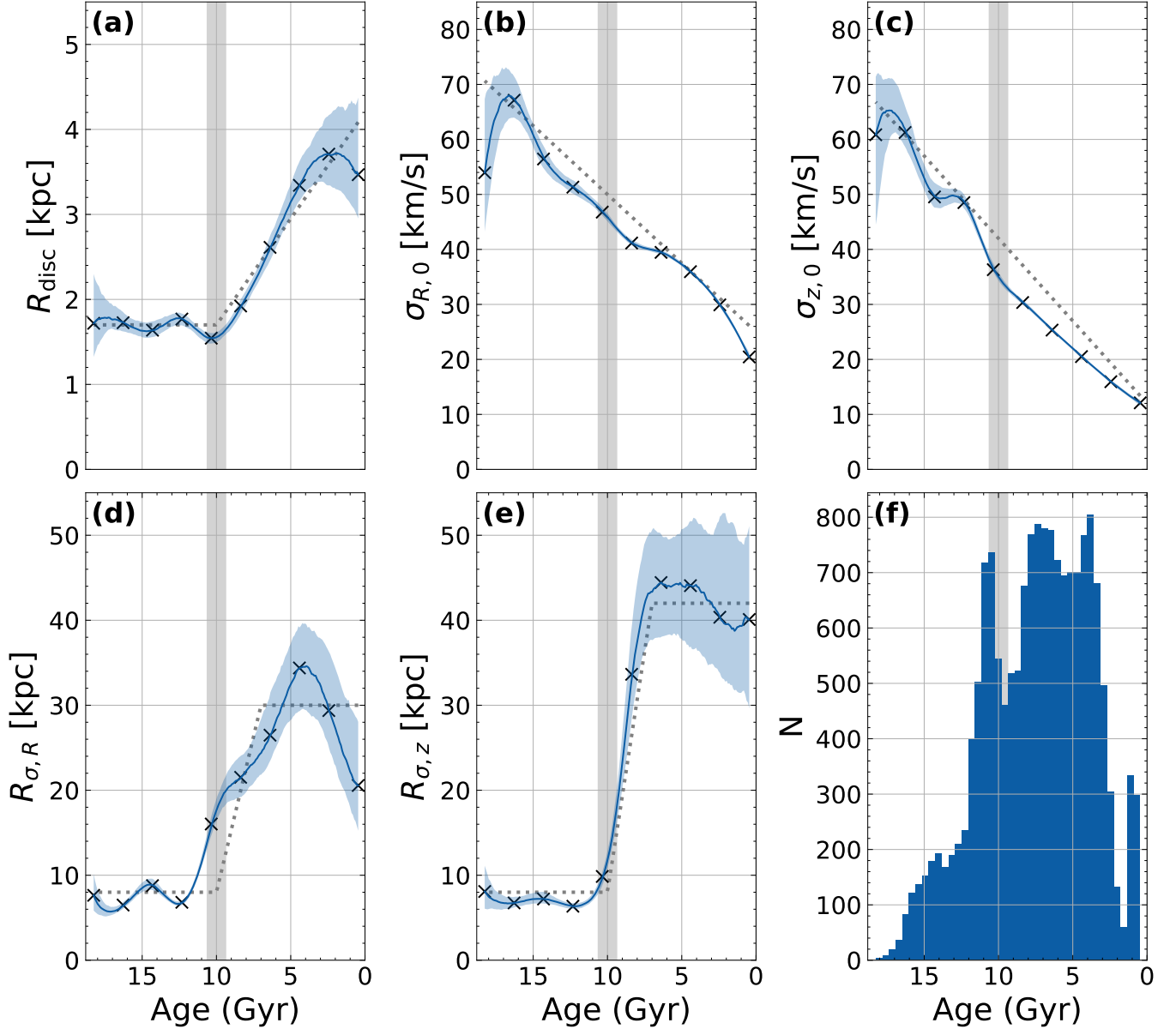
**Figure 2.** Posterior distributions of the splines that describe the age-dependent parameters of the quasi-isothermal DF, which are obtained after MCMC fitting. Panels (a), (b), (c), (d) and (e) correspond to the age dependence of the DF parameters R_{disc} , $\sigma_{R,0}$, $\sigma_{z,0}$, $R_{\sigma,R}$ and $R_{\sigma,z}$, respectively, and panel (f) shows the stellar age distribution. The blue solid lines indicate the median of the fitting results, while the shaded blue regions represent the 1σ confidence intervals. The grey dotted lines show the parameters used in f'_r in equation (5). Crosses denote the position of knots for the fitted spline function. The grey-shaded region in all the panels highlights the age around the dip in the scale length R_{disc} seen in panel (a), which coincides with the transition from the small older, thick disc, to the growing younger, thin disc.

Table 2. Our best fit values (medians of the posterior distributions) for the DF parameters at the knots shown in Fig. 2 together with their uncertainties, which correspond to the 16th and 84th percentiles of the posterior distributions.

τ [Gyr]	R_{disc} [kpc]	$\sigma_{R,0}$ [km s ⁻¹]	$\sigma_{z,0}$ [km s ⁻¹]	$R_{\sigma,R}$ [kpc]	$R_{\sigma,z}$ [kpc]
0.46	3.49 ^{+0.96} _{-0.56}	20.46 ^{+0.76} _{-0.63}	12.06 ^{+0.62} _{-0.59}	20.62 ^{+7.29} _{-5.15}	39.46 ^{+10.86} _{-10.25}
2.44	3.72 ^{+0.48} _{-0.37}	29.94 ^{+0.49} _{-0.41}	15.98 ^{+0.32} _{-0.32}	29.98 ^{+6.80} _{-4.83}	39.26 ^{+10.72} _{-6.32}
4.41	3.36 ^{+0.26} _{-0.19}	35.94 ^{+0.34} _{-0.34}	20.56 ^{+0.28} _{-0.30}	34.60 ^{+5.18} _{-4.99}	44.16 ^{+7.64} _{-7.30}
6.39	2.60 ^{+0.12} _{-0.15}	39.51 ^{+0.39} _{-0.43}	25.43 ^{+0.29} _{-0.34}	26.69 ^{+4.45} _{-3.19}	44.29 ^{+6.39} _{-6.36}
8.37	1.91 ^{+0.09} _{-0.09}	41.20 ^{+0.56} _{-0.47}	30.39 ^{+0.48} _{-0.46}	21.62 ^{+2.45} _{-2.32}	33.17 ^{+5.62} _{-4.76}
10.35	1.54 ^{+0.06} _{-0.06}	46.92 ^{+0.70} _{-0.80}	36.61 ^{+0.78} _{-0.71}	15.75 ^{+1.33} _{-1.04}	9.63 ^{+0.79} _{-0.36}
12.33	1.77 ^{+0.07} _{-0.08}	51.40 ^{+1.17} _{-1.25}	48.80 ^{+1.29} _{-1.24}	6.75 ^{+0.26} _{-0.26}	6.37 ^{+0.35} _{-0.35}
14.31	1.64 ^{+0.10} _{-0.09}	56.78 ^{+2.06} _{-1.91}	49.76 ^{+2.59} _{-2.04}	8.83 ^{+0.67} _{-0.59}	7.24 ^{+0.82} _{-0.57}
16.28	1.74 ^{+0.12} _{-0.11}	67.17 ^{+4.14} _{-3.12}	62.02 ^{+4.14} _{-3.52}	6.38 ^{+0.44} _{-0.36}	6.82 ^{+0.69} _{-0.61}
18.26	1.72 ^{+0.58} _{-0.46}	55.81 ^{+11.71} _{-11.83}	60.41 ^{+10.66} _{-13.82}	7.40 ^{+2.18} _{-1.65}	8.23 ^{+3.02} _{-1.99}

from a globally cold molecular disc. We find for the first time that $R_{\sigma,R}$ and $R_{\sigma,z}$ have an age dependence, showing trend transition with stars aged around $\tau = 10$ Gyr.

Our results show that the transition from the thick to the thin disc underlies in stars with ages around $\tau = 10$ Gyr, as indicated by a clear change in their kinematical properties. The transitional age is indicated by vertical grey-shaded regions in Fig. 2. Notably, stars at this transitional age exhibit a temporary dip in scale length reaching 1.5 kpc in panel (a) of Fig. 2. Appendix C compares the action distributions of stars with $10.5 > \tau > 9.5$ Gyr generated from two DF models. The first is the best-fit DF model that includes the dip in R_{disc} . The second is a comparison model without the dip, created by setting R_{disc} to 1.72 kpc, corresponding to 3σ above the best-fit value of the knot at $\tau = 10.35$ Gyr, while keeping all other parameters the same as in the best-fit DF (see Table. 2). We found that the model without the dip in R_{disc} shows a slight mismatch with the observed data in the J_R distribution (see details in Appendix. C). Hence, we consider that this dip is meaningfully inferred from the data. Interestingly, this is consistent with what is predicted by Grand et al. (2018) as introduced in Section 1. In the next section, we discuss the implications of this dip.

5 DISCUSSION

This section begins by examining how our model constrains the stellar density scale length, R_{disc} , without density information. We demonstrate that the dip in R_{disc} seen in Fig. 2 reflects a genuine signature that can be inferred from kinematic characteristics. Next, we analyse data from an Auriga cosmological simulation, which exhibits gas disc shrinking during the transition from the thick to the thin disc formation phase (Grand et al. 2018). We then discuss what the R_{disc} dip reveals about the Milky Way's formation history.

5.1 R_{disc} recovery from stellar kinematics

The scale length of a galactic disc is linked to stellar kinematics through the axisymmetric Jeans equation,

$$V_c - \langle V_\phi \rangle = \frac{\sigma_R^2}{2V_c} \left[\frac{\sigma_\phi^2}{\sigma_R^2} - 1 + R \left(\frac{1}{R_d} + \frac{2}{R_{\sigma,R}} \right) \right]$$

$$\Leftrightarrow \frac{1}{R_d} = \frac{1}{R} \left[\frac{2V_c}{\sigma_R^2} (V_c - \langle V_\phi \rangle) - \frac{\sigma_\phi^2}{\sigma_R^2} + 1 \right] - \frac{2}{R_{\sigma,R}}, \quad (11)$$

where the stellar surface density profile is described by an exponential law as $\Sigma(R) = \Sigma_0 \exp(-R/R_d)$, and velocity dispersion profiles as $\sigma_R(R) = \sigma_{R,0} \exp(-(R - R_0)/R_{\sigma,R})$. R_d is the scale length of the surface density profile, and $R_{\sigma,R}$ is that of the radial velocity dispersion profile. V_c , $\langle V_\phi \rangle$, σ_R and σ_ϕ denote circular velocity, mean rotation velocity, radial and rotational velocity dispersions at radius R , respectively. Here, R refers to the current Galactocentric radius of a star. This is different from the guiding radius, R_c , which appears in equation (4), i.e., R_d in equation (11) is defined differently from R_{disc} in equation (4).

To compute R_d from equation (11), we focused on stars around the Galactocentric radius of the Sun, i.e. $R \sim R_0 = 8.275$ kpc. We selected stars within $|R - R_0| < 0.1$ kpc and $6.5 < \tau < 13.5$ Gyr, and analysed them in different age bins. Note that this binning is used only for illustrative purposes; the full DF modelling in Section 4 does not rely on binning and can account for age uncertainties. The top three panels of Fig. 3 show (a) the mean rotational velocity, $\langle V_\phi \rangle$, (b) the radial velocity dispersion, σ_R and (c) the ratio of the azimuthal velocity dispersion to the radial velocity dispersion, σ_ϕ^2/σ_R^2 , at different age bins. Using these values and equation (11), we computed R_d for each age bin, shown in panel (d). Because $R_{\sigma,R}$ cannot be reliably estimated within each age bin due to the limited sample sizes and restricted radial coverage, we adopt values of $R_{\sigma,R}$ from the fitting result presented in Fig. 2 (Section 4). Since we restricted our sample to a narrow range in Galactocentric radius, we fix $R = R_0$ and $V_c = V_c(R_0)$ in equation (11) for simplicity ignoring the radial dependence of them. The error bars in each panel were evaluated using the Monte Carlo method, i.e., computing these values after re-sampling the observational data using the observational uncertainties. The number of selected stars for each bin is shown in panel (e) of Fig. 3.

Panel (a) shows that $\langle V_\phi \rangle$ increases from approximately 195 km s⁻¹ at around 11 Gyr to about 220 km s⁻¹ at 7 Gyr, with a transition occurring around 10 Gyr. As shown in equation (11), an increase of $\langle V_\phi \rangle$ makes R_d larger. Similarly, in panel (b), σ_R decreases from around 60 km s⁻¹ at 13 Gyr to about 40 km s⁻¹ after 9 Gyr. This behaviour is consistent with what we found in our cubic spline fitting in panel (b) of Fig. 2. A decrease of σ_R is connected to an increase of R_d . The effect of the change in σ_R on R_d is stronger the closer the value of $\langle V_\phi \rangle$ is to V_c , as seen from equation (11). In panel (c), a slight increase is seen in σ_ϕ^2/σ_R^2 at 10 Gyr. This leads to a larger R_d following equation (11), but the contribution of this slight increase is smaller than the terms mentioned above. Panel (d) shows the resultant R_d computed with equation (11) at different age bins. R_d remains small for older stars and increases for younger stars, con-

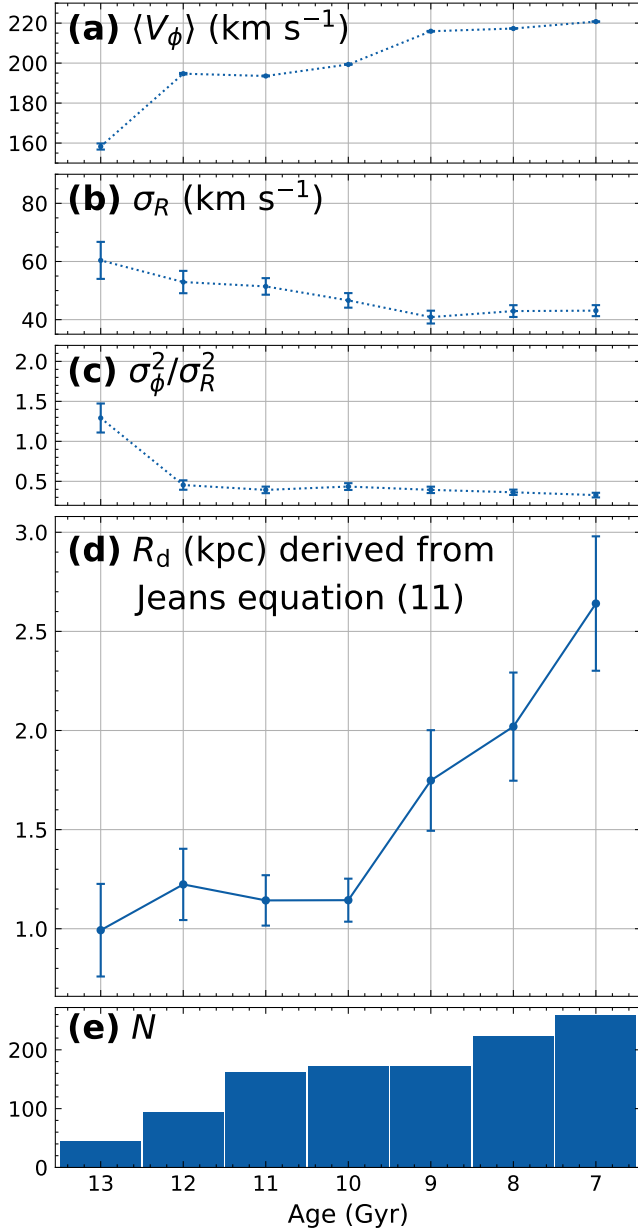


Figure 3. Measured kinematical properties (panel a-c), the derived scale radius, R_d , (panel d) and the number of stars (panel e) for the stars around the solar radius in different age bins. Solid circles with error bars connected with the dotted lines represent (a) the mean azimuthal velocity, $\langle V_\phi \rangle$, (b) the radial velocity dispersion, σ_R , and (c) the squared ratio of radial to azimuthal velocity dispersions, σ_ϕ^2/σ_R^2 , respectively. Error bars indicate the 1σ uncertainties. Solid circles connected with the solid lines in panel (d) show the scale length computed from these kinematic quantities via equation (11), with 1σ uncertainties estimated using a Monte Carlo method shown as error bars.

sistent with the result shown in Fig. 2. Panel (d) also shows a slight decrease in R_d at $\tau = 10$ Gyr, where the dip of R_{disc} is observed in panel (a) of Fig. 2. Note that, as mentioned above, R_d and R_{disc} are defined differently, and hence it is not surprising to see that R_d is not quantitatively the same as R_{disc} . Fig. 3 shows that there is no single parameter driving the dip of R_d . This suggests that the dip in R_{disc}

at $\tau \sim 10$ Gyr in Fig. 2 is obtained due to the combined trends of the mean rotation velocity and velocity dispersion properties.

The depth of the dip shown in panel (d) in Fig. 3 is less than 0.2 kpc, which is comparable to 1σ uncertainty, and the dip is not very clear in this figure. However, this is simply because we took a narrow radial range of the data to compute equation (11), which makes the error larger and the result less reliable. Note that the main aim of this section is to demonstrate that kinematical information alone is sufficient to constrain the scale length of the density distribution, not to assess the depth of the dip feature.

5.2 Shrinking disc in Auriga cosmological simulation

In this section, we apply our kinematic DF fitting method to a simulated galaxy from the Auriga cosmological simulation suite. The Auriga project (Grand et al. 2017) is a set of cosmological magnetohydrodynamical zoom simulations of the formation and evolution of isolated Milky Way-mass galaxies, (mass range from 10^{12} to $2 \times 10^{12} M_\odot$). The Auriga simulations use cosmological parameters of $\Omega_m = 0.307$, $\Omega_b = 0.048$, $\Omega_\Lambda = 0.693$ and Hubble constant of $H_0 = 67.77 \text{ km s}^{-1} \text{ Mpc}^{-1}$ (Planck Collaboration et al. 2014).

We selected the Au23 simulation for our analysis because it closely resembles the Milky Way in terms of stellar mass and disc structure. Its stellar mass is $9.02 \times 10^{10} M_\odot$ and the radial scale length of its disc is 4.99 kpc. In addition, it is known that it exhibits a temporary shrinking of the star-forming gas disc around 6 Gyr ago, which coincides with a transition from thick to thin disc formation (Grand et al. 2018). Therefore, it is a useful test case to see whether traces of past gas disc shrinking remain in such a present-day ($z = 0$) simulated galaxy as an age-dependent scale length.

For our DF fitting we selected star particles with ages between 4 and 10 Gyr. We limited the sample to stars with $|z| < 0.5$ kpc. Then, we randomly chose 16,000 stars within 5 kpc from the assumed observer position at $(x, y) = (-R_0, 0)$ in the simulation, where $R_0 = 8.275$ kpc corresponds to the Sun’s Galactocentric radius in the Milky Way. The gravitational potential of the galaxy was computed with AGAMA’s multipole potential model assuming axisymmetry, based on the mass distributions of stars, gas and dark matter. From this potential, the circular velocity at the observer’s radius was determined to be $V_c(R_0) = 244.09 \text{ km s}^{-1}$.

Panels (a)-(e) of Fig. 4 show fitting results for the simulation data using cubic spline functions with nine knots indicated by crosses. As noted in Section 4, the parameters shown here do not reflect the values at the time of each star’s birth (i.e., lookback time), but rather represent the present-day properties of different stellar age populations in the disc, while the three snapshots in panel (a) illustrate the spatial gas distribution at the respective past epoch, $t_{\text{lookback}} = 4.25, 6.18$ and 8.20 Gyr ago. The solid blue lines represent the median of the sampled spline curves from the posterior distribution function of each parameter, while the blue shaded regions indicate the 1σ uncertainty range. The dashed grey lines show the age-dependent function for the adopted parameters of f'_τ . The shape of this function was determined by iteratively adjusting f'_τ to be similar to the best-fitting f_τ . Panel (f) shows the age distribution of the selected stars.

Fig. 4 shows generally similar results to those obtained from the observational data of the Milky Way disc, and shows a clear transition of the disc structures at $\tau \sim 6$ Gyr. According to Grand et al. (2018), this coincides with the epoch of the transition from the high- $[\alpha/\text{Fe}]$ thick disc formation phase to the low- $[\alpha/\text{Fe}]$ thin disc formation phase in Au 23. Panel (a) shows that R_{disc} for stars older than the transitional age is relatively small, around 3 kpc, and the younger disc ($\tau < 6$ Gyr) shows larger R_{disc} . The radial and vertical velocity

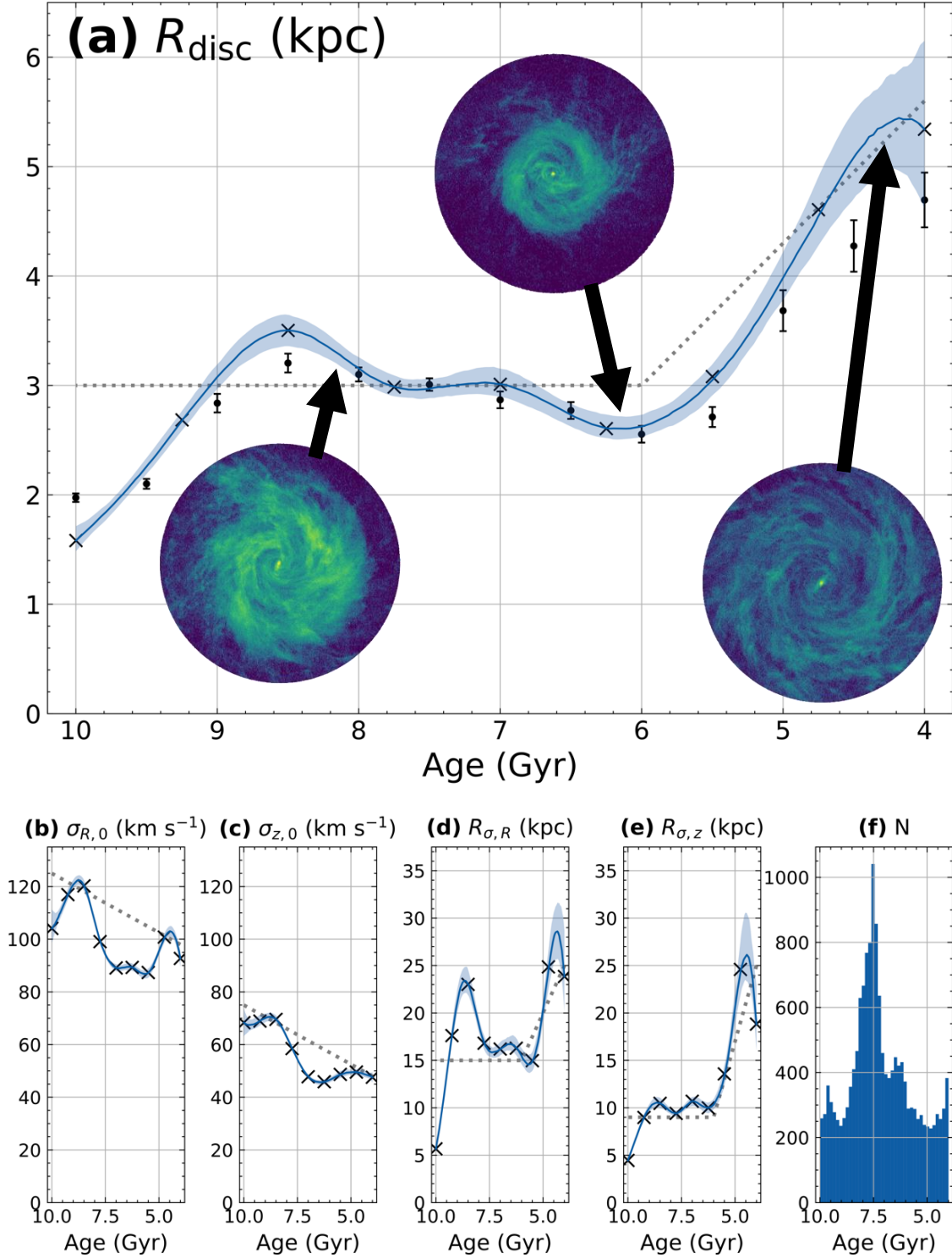


Figure 4. (Galactic disc parameters as a function of stellar age as obtained from our MCMC DF fitting for the Au 23 simulation data. Panels (a), (b), (c), (d) and (e) present the model fitting results for parameters of R_{disc} , $\sigma_{R,0}$, $\sigma_{z,0}$, $R_{\sigma,R}$ and $R_{\sigma,z}$, respectively and panel (f) shows stellar age distribution of the selected star particles for this fitting. The blue solid lines indicate the median of the posterior distribution function for each parameter, while the shaded blue regions represent the 1σ confidence intervals. The thick orange solid lines show f'_τ in equation (5). Crosses denote the position of nine knots for the cubic spline fitting. Black dots with error bar in panel (a) show R_d of the simulated galaxy in each stellar age bin measured from its surface density profile. The three snapshots shown in panel (a) represent face-on maps of the gas disc at 4.25, 6.18, and 8.10 Gyr ago, respectively. Each image is cropped to a circular region within a radius of 20 kpc. Note that the snapshots are taken from these three different past epochs, while the parameter values shown in all panels are obtained from fitting the present-day ($z = 0$) simulation data.

dispersions (panels b and c) show the overall trend of decreasing velocity dispersion with decreasing age, though there is a large fluctuation at some ages. Panels (d) and (e) show that $R_{\sigma,R}$ and $R_{\sigma,z}$ also undergo a transition with stars aged around 6 Gyr, showing a smaller scale length for older stars, and an increasing scale length for younger stars. These are also similar trends to those seen in Fig. 2.

The black dots with error bars in panel (a) of Fig. 4 show the scale lengths measured by fitting a single exponential profile to the radial surface density of star particles in each age bin using a least-squares method. In this calculation, stars within $5 < R < 11$ kpc were used without applying any vertical selection. It should be noted that the value of R_{disc} used in the DF (eq. 4) is not strictly identical to the scale length measured from the surface density profiles particularly when the discs are relatively warm. Although R_{disc} is formally defined as a scale length parameter in the DF, the radial density profile of stars as a function of R or R_c slightly deviates from an exponential profile (e.g., Binney 2010). This is because the DF is described with actions, J , which implicitly encode the radial dependence of the gravitational potential. This difference in the definition likely explains the difference between the scale length R_{disc} obtained from our model fitting and that derived from the surface density profiles. For example, at $\tau = 7$ Gyr, our fitting result gives $R_{\text{disc}} = 3.00^{+0.14}_{-0.13}$ kpc, while the value derived from the surface density profile is $R_d = 2.87 \pm 0.08$ kpc. At $\tau = 5.5$ Gyr, the corresponding values are $R_{\text{disc}} = 3.08^{+0.15}_{-0.17}$ kpc and $R_d = 2.71 \pm 0.09$ kpc, respectively. Despite this small difference, the overall trend of R_{disc} remains consistent with the scale length measured from the surface density profiles. This demonstrates that our quasi-isothermal DF fitting can successfully capture the age dependence of R_{disc} , even when applied to a cosmological simulation where the galaxy does not strictly follow a quasi-isothermal DF.

Notably, panel (a) of Fig. 4 shows a dip in R_{disc} around $\tau \sim 6$ Gyr, which coincides with the transition from thick to thin disc formation in Au 23. This dip is also seen in the scale length directly measured from the radial surface density profile at the different age bins (black dots with the error bars in panel (a) of Fig. 4). As reported in Grand et al. (2018), Au 23 has a shrinking of the star-forming gas disc at $t_{\text{lookback}} \sim 6$ Gyr ago. Face-on views of the gas disc at three different look-back times (8.10, 6.18, and 4.25 Gyr from left to right) are presented in panel (a) of Fig. 4, showing the inner 20 kpc of the galaxy. It is evident that the gas disc temporarily shrinks at 6.18 Gyr ago. Our results demonstrate that stars formed during the gas disc's shrinking at $t_{\text{lookback}} \sim 6$ Gyr retain a small radial scale length even at the present epoch, i.e., $z = 0$. Hence, drawing on these similarities between the fitting results of the Auriga simulation and observational data, the dip in R_{disc} seen in Fig. 2 could indicate that the Milky Way's star-forming gas disc shrank around 10 Gyr ago.

5.3 Star-forming gas disc shrinking at the thick-to-thin disc transition

Our analysis suggests that the Milky Way's star-forming gas disc shrank around 10 Gyr ago, coinciding with a structural and dynamical phase transition from the thick to thin disc. In Fig. 2, stellar populations older than 10 Gyr show disc scale lengths of $R_{\text{disc}} \sim 1.7$ kpc. This is consistent with previous measurements of thick disc scale length, most of which are based on star counts, e.g. 2.0 kpc by Bensby et al. (2011), $1.8^{+2.1}_{-0.5}$ kpc by Cheng et al. (2012), 2.2 ± 0.2 kpc by Bovy et al. (2016), 1.9 ± 0.1 kpc by Mackereth et al. (2017) and 1.9 kpc by Yu et al. (2021). Importantly, our constraint is derived purely from stellar kinematics, making it complementary to star-count studies. The close agreement indicates that non-equilibrium

effects are unlikely to dominate and that the gravitational potential adopted in this paper is a reasonable description of the Galaxy.

For populations younger than 10 Gyr, R_{disc} increases steadily as age decreases. This trend suggests that the thin disc grew in size smoothly. Such a pattern is consistent with the inside-out formation scenario (e.g., Chiappini et al. 2001), which is proposed based on observed chemical abundance gradients. It is also supported by mono-abundance population analyses (e.g., Bovy et al. 2012b). Furthermore, Frankel et al. (2019) quantified this growth, showing that the half-mass radius of the Galactic disc has expanded by approximately 43% over the past 7 Gyr. The inside-out formation of the thin disc is consistent with observations showing that progenitors of Milky Way-sized galaxies had smaller disc sizes at higher redshifts (e.g., van Dokkum et al. 2013; Tan et al. 2024; Tsukui et al. 2025).

Grand et al. (2018) analysed the Auriga simulations and demonstrated that the shrinking of the star-forming gas disc is related to the thick-to-thin disc formation transition and the chemical bimodality in the solar neighbourhood and the outer disc of the Milky Way. They proposed that the star-forming disc shrinkage is attributed to a temporary decline in gas inflow, leading to an insufficient supply to replenish gas consumed by star formation. In other words, when gas consumption exceeds inflow, the star-forming gas disc shrinks.

Fig. 5 shows the age-[Fe/H] relation in the APOGEE data used in our study, colour coded with [Mg/Fe]. Stars older than 12 Gyr are predominantly metal-poor ([Fe/H] < -0.4) with high [Mg/Fe], representing the thick disc population. In contrast, stars younger than 10 Gyr tend to have higher [Fe/H] with lower [Mg/Fe], corresponding to the thin disc. Ciucă et al. (2024) referred to the period $12 > \tau > 10$ Gyr as the GGS phase, where the metallicity rapidly increases with decreasing age, and [Mg/Fe] decreases at the same period. By comparing APOGEE data with Auriga simulations, Ciucă et al. (2024) suggested that this chemical transition was driven by the gas-rich GSE merger. The gas inflow from the GSE merger likely induced a starburst in the Galactic disc, initially boosting [Mg/Fe], followed by a rise in [Fe/H] and a fall in [Mg/Fe] (e.g., Brook et al. 2007). The grey-shaded region in Fig. 5 marks the same epoch as in Fig. 2, highlighting the phase transition epoch with the temporary disc shrinking in the Milky Way. This period coincides with the final stage of the GGS phase. One possible scenario is that the intense starbursts caused by the GSE could have exceeded the gas supply and lead to the shrinking of the star-forming gas disc.

Another interesting explanation is the transition from cold-mode to hot-mode gas accretion. Theoretical work by Noguchi (2018) proposed that high- $[\alpha/\text{Fe}]$ stars formed rapidly through cold-mode gas accretion, followed by the formation of low- $[\alpha/\text{Fe}]$ stars under hot-mode accretion. In the cold mode, which dominates in lower-mass galaxies, gas flows efficiently along cosmic filaments into the cold star-forming gas disc without experiencing shock heating, since their virial temperatures are low. In contrast, hot-mode accretion dominates in more massive galaxies. In the hot mode dominant galaxies, inflowing gas is first heated to the temperature of the hot halo. The present-day Milky Way is thought to be massive enough for hot-mode accretion, which may be essential for the thin disc formation (e.g., Kereš et al. 2005; Grand et al. 2018; Hafen et al. 2022). Thus, the Milky Way would have undergone this transition at some point in the past, eventually reaching the hot mode today. Grand et al. (2018) showed that the shrinking of the star-forming gas disc at the onset of thin disc formation occurred when the Milky Way became massive enough for hot-mode gas accretion to dominate, reducing the gas supply to the central cold gas disc.

Building on the insights from the simulation study in Grand et al. (2020) and the timing coincidence between the end of the GGS

phase and the scale length dip, our results may suggest the following potential scenario to explain the shrinking of the gas-disc at the transition from the thick to thin disc formation phase in the Milky Way: (1) before the GSE merger, the Milky Way was still relatively low in mass, allowing the formation of the thick disc through intense star formation and occasional gas-rich mergers under cold-mode gas accretion, which maintained the formation of high- $[\alpha/\text{Fe}]$ thick disc stars (Brook et al. 2004, 2005, 2012b). (2) The gas-rich nature of the GSE merger (Grand et al. 2020; Ciucă et al. 2024) likely triggered an intense starburst, leading to rapid consumption of the available gas. At the same time, the GSE merger may have been substantial enough to increase the total mass of the Milky Way and triggered a transition from cold-mode to hot-mode gas accretion. Alternatively, this transition may have resulted from a combination of factors, with the GSE being only one of them. The total mass of GSE is subject to considerable debate, with estimates ranging from $\sim 10^8$ to $10^{11} M_\odot$, corresponding to $\sim 10 - 50\%$ of the Milky Way's progenitor mass at the time, depending on the study (e.g., Helmi et al. 2018; Naidu et al. 2021; Buck et al. 2023; Lane et al. 2023). The critical halo mass threshold between hot- and cold-modes is estimated approximately $10^{11.4}$ to $10^{12} M_\odot$ (Kereš et al. 2005; Dekel & Birnboim 2006). (3) The transition to hot-mode accretion reduced the inflow of cold gas, just after the rapid consumption of the gas due to the GSE merger, limiting the Galaxy's replenishment of its star-forming gas and thereby accelerating the depletion of the gas reservoir. This process may have led to the shrinking of the cold gas disc, leaving an imprint that is now observed as the dip in R_{disc} traced by stars aged around $\tau \sim 10$ Gyr as in Fig. 2. (4) The subsequent formation of the thin disc proceeded in an inside-out manner, as indicated by the steady increase of R_{disc} for younger stellar populations shown in Fig. 2.

A star formation quenching in the solar-neighbourhood of the Galactic disc has been proposed in the literature (e.g., Gratton et al. 2000). This can explain the observed bimodality in the $[\alpha/\text{Fe}]$ – $[\text{Fe}/\text{H}]$ distribution of solar-neighbourhood stars (e.g., Haywood et al. 2016; Snaith et al. 2014, 2015). To account for this feature, chemical evolution models such as the two-infall model have been proposed, in which a temporary suppression in star formation is caused by a delay between two episodes of gas accretion (e.g., Chiappini et al. 1997, 2015). The halt of star formation is thought to have occurred during the transitional epoch between the formation of the thick and thin discs (e.g., Spitoni et al. 2024), which corresponds to around 10 Gyr ago on the stellar age scale used in this study. In our study, we identify a dip in the disc scale length R_{disc} around $\tau \sim 10$ Gyr, which aligns with this thick-to-thin disc transition period. If the gas disc during the thick disc formation phase was initially extended beyond the solar radius and shrank to within the solar radius (R_0) at the transitional epoch, star formation would have been temporarily halted in the outer disc—including at R_0 , while continuing in the inner disc, as suggested in Grand et al. (2018). This "outer-disc quenching" scenario may provide an explanation for the observed chemical bimodality: star formation near the solar radius was paused when the gas disc shrank, and resumed once fresh gas was steadily accreted and the disc regrew in an inside-out manner, eventually reaching R_0 . In this scenario, star formation at the solar radius naturally ceased due to the temporarily reduced size of the star-forming disc.

While the scenario above offers a more physically motivated explanation for the pause in star formation inferred from solar-neighbourhood stars, it is not the only possible interpretation. To test this scenario, further modelling will be required, including the effects of radial migration and the kinematical heating due to the merger, which may have played a key role in reshaping the stellar

disc's scale length. In fact, the dip of scale length we found in Fig. 2 may not look drastic to infer such outer-disc quenching. However, our analysis is the present-day scale length of the stars formed during this transition period, whose angular momentum distribution must be significantly modified by radial migration and radial mixing due to the heating. For example, thick disc stars may have formed initially at smaller radii and then moved outward as the thin disc grew through inside-out formation, receiving angular momentum from the thin disc or being dynamically heated by minor mergers and/or the bar (e.g., Brook et al. 2012b). The upper panel of Fig. 4 demonstrates that the gas disc experienced a significant shrinking of about one-third of its size between $t_{\text{lookback}} = 8.10$ and 6.18 Gyr ago. In contrast, the difference in the stellar scale length at $z = 0$ is only about one tenth. This suggests that the small dip observed in the present-day scale length may reflect a larger difference in the star-forming gas disc at the birth epoch of those stars.

Zhang et al. (2025) compared the age-metallicity relation from *Gaia* data with simulations including a slowing bar, and proposed that the age-metallicity relation shows two sequences: a higher- $[\text{Fe}/\text{H}]$ sequence formed by stars that migrated radially due to the slowing bar, and a lower- $[\text{Fe}/\text{H}]$ sequence reflecting the chemical evolution of stars formed locally. The sequence of local chemical evolution for stars with guiding radii around $R_g \sim 8$ kpc appears to begin approximately 8 Gyr ago, at a metallicity of $[\text{Fe}/\text{H}] \sim -0.75$. Note that their age scale is different from our age scale. This trend is consistent with our proposed scenario of inside-out thin disc formation, after the star forming gas disc shrinking. Star formation at $R \sim 8$ kpc likely began when the gas disc grew outward and reached that radius, fuelled by the accretion of low- $[\text{Fe}/\text{H}]$ gas from the hot halo gas. Interestingly, in the scenario proposed by Zhang et al. (2025), stars formed in the inner disc can migrate outward to $R \sim 8$ kpc due to the growth of the bar. This implies that the star-forming disc during the thick disc formation phase did not need to extend as far as $R \sim 8$ kpc, and that the star formation at this radius began only when the thin disc had grown large enough to reach it.

Clumpy early disc formation has also been suggested to explain the early thick disc formation (Noguchi 1998; Bournaud et al. 2009; Inoue et al. 2016). In high-redshift galaxies, gas-rich discs become gravitationally unstable, leading to the formation of massive, high-density gas clumps. Within these clumps, star formation proceeds rapidly and can drive the production of high- $[\alpha/\text{Fe}]$ populations. Previous studies comparing Milky Way observations with simulations have suggested that clump formation can reproduce key features of the disc such as the $[\alpha/\text{Fe}]$ bimodality (e.g., Clarke et al. 2019; Beraldo e Silva et al. 2020). Clumpy disc scenario generally predicts that the high- and low- $[\alpha/\text{Fe}]$ discs evolve in parallel rather than sequentially (e.g., Beraldo e Silva et al. 2021), which could be a distinctive feature from the scenario of the thick to thin disc transition due to the GSE merger highlighted in this paper. However, our model focuses on the age dependence of the kinematical structure of the Galactic disc, and cannot distinguish these scenarios. Fitting chrono-chemodynamical data with the model including the parametric descriptions of the impacts of GSE-like merger, clumpy disc formation, the bar and spiral arms would be interesting to further disentangle the formation history of the Milky Way, though it would be challenging to construct such model.

It is important to keep in mind the uncertainties in stellar age estimates when considering any scenario or assessing causal links between different events. Fig. 5 shows typical age uncertainties for the stars used in this study, indicated by error bars at several stellar ages. Around $\tau = 10$ Gyr, where the transition from the thick to thin disc occurs and a dip in the scale length is observed, the typical age

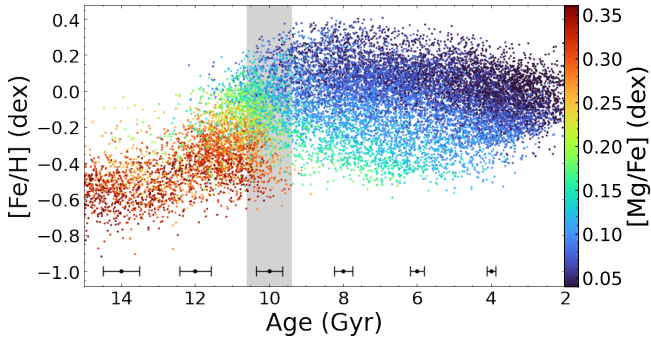


Figure 5. The age-metallicity relation coloured by $[Mg/Fe]$. The region shaded in grey shows the same period shown in Fig. 2. Black dots with error bars represent typical 1σ uncertainties of stars in each age.

uncertainty is about 0.35 Gyr. Additionally, the spline fitting in Fig. 2 uses knot intervals of approximately 2 Gyr (see Table 2). Therefore, the effective age resolution of our analysis is around 2 Gyr, and trends on finer age scales (i.e., <2 Gyr) cannot be reliably discussed. Moreover, the stellar ages used in this study are based on the BINGO age trained with asteroseismic age (Miglio et al. 2021). Its absolute scale differs from other methods, and the relation between different age estimates is not necessarily linear. A key challenge for future work is to calibrate the age scales across different estimation techniques.

6 SUMMARY

In this study, we analysed the stellar age-kinematics relation of 16,617 red giant stars from APOGEE DR17, using age estimates from Ciucă et al. (2024), cross-matched with *Gaia* DR3. Following the approaches of Zhang & Sanders (2023) and Sanders et al. (2024), we modelled the stellar kinematic distribution using a dynamical action-based distribution function parametrized by cubic spline curves to describe the age dependence of the disc parameters.

We provide the first DF model that smoothly tracks structural and kinematic properties across the entire stellar age range in the Milky Way disc. Our results reveal a clear kinematic transition at age ~ 10 Gyr, marking the shift from the thick disc to the thin disc population. This transition is accompanied by a decrease in the disc scale length, followed by inside-out growth of the thin disc. This is the first observational indication of such a temporary shrinking at the transition period between high- $[\alpha/Fe]$ thick disc and low- $[\alpha/Fe]$ thin disc. Importantly, we also find for the first time that $R_{\sigma,R}$ and $R_{\sigma,z}$ exhibit a clear age dependence, with a transition in their trends occurring at age ~ 10 Gyr. How the transition of these parameters is connected to the formation history of the Galactic disc remains to be clarified. These trends may reflect processes such as flaring and dynamical heating in the outer disc. This should be further tested with numerical simulations in future work.

We applied the same fitting approach to an Auriga cosmological simulation, Au 23, which shows the temporary shrinking of the star-forming cold gas disc at the period of the transition from the high- $[\alpha/Fe]$ thick to the low- $[\alpha/Fe]$ thin disc formation (Grand et al. 2018). Our method recovers the short radial scale length of the stars formed during the cold gas-disc shrinking period. This demonstrates that past gas disc shrinking can leave imprints on present-day stellar kinematics, which can be recovered through our DF fitting.

Hence, we suggest that the small disc found for stars aged around

$\tau = 10$ Gyr reflects a temporary shrinking of the Milky Way’s star-forming gas disc at that time. This event coincided with the end of the GGS phase, which was likely triggered by the gas-rich GSE merger (Ciucă et al. 2024). Notably, this timing also corresponds to the onset of the low- $[\alpha/Fe]$ thin disc formation (Fig. 5). Drawing on our results and previous simulation studies (e.g., Grand et al. 2018, 2020), we outline a possible scenario that may explain our findings. Before the GSE merger, the Milky Way was low enough mass to maintain predominantly cold-mode gas accretion, driving high star formation rates and fed by frequent gas-rich mergers, as expected at high redshift in a Λ CDM Universe (e.g., Brook et al. 2004). The gas-rich GSE merger then triggered an intense starburst (GGS phase; Ciucă et al. 2024) and rapid gas shrinking of the disc. At the same time, a transition from cold- to hot-mode gas accretion may have occurred, possibly triggered by the increase in the Milky Way’s mass, due to the significant mass added by GSE merger. The transition to hot-mode accretion slowed the supply of gas to the cold gas disc, just as the cold gas was being rapidly consumed by the starburst. This combination could have caused the temporary shrinking of the star-forming gas disc. As a result, the thin disc formation started from a gas disc smaller than the thick disc. Subsequently, the thin disc grew inside-out, fuelled by smooth and gradual gas accretion from the hot halo. Over time, gas with progressively higher angular momentum accreted, driving the disc’s outward expansion (e.g., Brook et al. 2012a).

We caution that stellar age estimates carry uncertainties that affect the interpretation of features such as the dip in scale length observed in stars with ages around $\tau \sim 10$ Gyr. The BINGO inferred age used here may differ in scale from ages derived by other methods. Calibrating age scales across different techniques is essential for comparing the timing of key events and building a coherent picture of the Milky Way’s formation history.

Recently, Merrow et al. (2024) suggested that the GSE merger can trigger bar formation. Hence, comparing key epochs in the Milky Way’s evolution—including the GGS period triggered by the GSE merger (Ciucă et al. 2024), the thick-thin disc transition epoch revealed in our study and the bar formation period suggested by Sanders et al. (2024)—is essential for understanding the comprehensive picture of the Milky Way’s history. However, the age scale of our study based on Miglio et al. (2021) is different from the age scale used in Sanders et al. (2024) which relies on the calibrated Mira variable age-period relation in Zhang & Sanders (2023). We will carefully calibrate these age scales and investigate how bar formation influences the transition between thick and thin disc formation phases in future work.

ACKNOWLEDGEMENTS

We thank the anonymous referee for the careful review and helpful comments. This work is a part of MWGaiaDN, a Horizon Europe Marie Skłodowska-Curie Actions Doctoral Network funded under grant agreement no. 101072454 and also funded by UK Research and Innovation (EP/X031756/1). This work was also partly supported by the UK’s Science & Technology Facilities Council (STFC grant ST/S000216/1, ST/W001136/1). JLS acknowledges the support of the Royal Society (URF/R1191555; URF/R241030).

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national

institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

This work has made use of data from the Apache Point Observatory Galactic Evolution Experiment (APOGEE), which is part of the Sloan Digital Sky Survey (SDSS). Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss4.org.

SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, Center for Astrophysics | Harvard & Smithsonian (CfA), the Chilean Participation Group, the French Participation Group, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, the Korean Participation Group, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatório Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author. We have used simulations that are part of the Auriga Project public data release (Grand et al. 2024) available at <https://wwwmpa.mpa-garching.mpg.de/auriga/data>.

REFERENCES

- Almannaei A. S., Kawata D., Baba J., Hunt J. A. S., Seabroke G., Yan Z., 2024, *MNRAS*, **529**, 1035
- Anders F., et al., 2023, *A&A*, **678**, A158
- Belokurov V., Erkal D., Evans N. W., Koposov S. E., Deason A. J., 2018, *MNRAS*, **478**, 611
- Belokurov V., et al., 2020, *MNRAS*, **496**, 1922
- Bennett M., Bovy J., 2019, *MNRAS*, **482**, 1417
- Bensby T., Zenn A. R., Oey M. S., Feltzing S., 2007, *ApJ*, **663**, L13
- Bensby T., Alves-Brito A., Oey M. S., Yong D., Meléndez J., 2011, *ApJ*, **735**, L46
- Bensby T., Feltzing S., Oey M. S., 2014, *A&A*, **562**, A71
- Beraldo e Silva L., Debattista V. P., Khachaturyants T., Nidever D., 2020, *MNRAS*, **492**, 4716
- Beraldo e Silva L., Debattista V. P., Nidever D., Amarante J. A. S., Garver B., 2021, *MNRAS*, **502**, 260
- Binney J., 2010, *MNRAS*, **401**, 2318
- Binney J., McMillan P., 2011, *MNRAS*, **413**, 1889
- Bournaud F., Elmegreen B. G., Martig M., 2009, *ApJ*, **707**, L1
- Bovy J., Rix H.-W., 2013, *ApJ*, **779**, 115
- Bovy J., Rix H.-W., Hogg D. W., 2012a, *ApJ*, **751**, 131
- Bovy J., Rix H.-W., Liu C., Hogg D. W., Beers T. C., Lee Y. S., 2012b, *ApJ*, **753**, 148
- Bovy J., Rix H.-W., Schlafly E. F., Nidever D. L., Holtzman J. A., Shetrone M., Beers T. C., 2016, *ApJ*, **823**, 30
- Bradbury J., et al., 2018, JAX: composable transformations of Python+NumPy programs, <http://github.com/jax-ml/jax>
- Brook C. B., Kawata D., Gibson B. K., Freeman K. C., 2004, *ApJ*, **612**, 894
- Brook C. B., Gibson B. K., Martel H., Kawata D., 2005, *ApJ*, **630**, 298
- Brook C., Richard S., Kawata D., Martel H., Gibson B. K., 2007, *ApJ*, **658**, 60
- Brook C. B., Stinson G., Gibson B. K., Roškar R., Wadsley J., Quinn T., 2012a, *MNRAS*, **419**, 771
- Brook C. B., et al., 2012b, *MNRAS*, **426**, 690
- Buck T., Obreja A., Ratcliffe B., Lu Y., Minchev I., Macciò A. V., 2023, *MNRAS*, **523**, 1565
- Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, *A&A*, **530**, A138
- Chen T., Guestrin C., 2016, *arXiv e-prints*, p. arXiv:1603.02754
- Cheng J. Y., et al., 2012, *ApJ*, **752**, 51
- Chiappini C., Matteucci F., Gratton R., 1997, *ApJ*, **477**, 765
- Chiappini C., Matteucci F., Romano D., 2001, *ApJ*, **554**, 1044
- Chiappini C., Minchev I., Martig M., 2015, in IAU General Assembly. p. 2255326
- Ciucă I., Kawata D., Miglio A., Davies G. R., Grand R. J. J., 2021, *MNRAS*, **503**, 2814
- Ciucă I., et al., 2024, *MNRAS*, **528**, L122
- Clarke A. J., et al., 2019, *MNRAS*, **484**, 3476
- Dekel A., Birnboim Y., 2006, *MNRAS*, **368**, 2
- Di Matteo P., Haywood M., Lehnert M. D., Katz D., Khoperskov S., Snaith O. N., Gómez A., Robichon N., 2019, *A&A*, **632**, A4
- Feltzing S., Bensby T., Lundström I., 2003, *A&A*, **397**, L1
- Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, **125**, 306
- Frankel N., Sanders J., Rix H.-W., Ting Y.-S., Ness M., 2019, *ApJ*, **884**, 99
- Fuhrmann K., 1998, *A&A*, **338**, 161
- GRAVITY Collaboration et al., 2021, *A&A*, **647**, A59
- Gaia Collaboration et al., 2023, *A&A*, **674**, A37
- Gallart C., Bernard E. J., Brook C. B., Ruiz-Lara T., Cassisi S., Hill V., Monelli M., 2019, *Nature Astronomy*, **3**, 932
- García Pérez A. E., et al., 2016, *AJ*, **151**, 144
- Gilmore G., Reid N., 1983, *MNRAS*, **202**, 1025
- Grand R. J. J., et al., 2017, *MNRAS*, **467**, 179
- Grand R. J. J., et al., 2018, *MNRAS*, **474**, 3629
- Grand R. J. J., et al., 2020, *MNRAS*, **497**, 1603
- Grand R. J. J., Fragkoudi F., Gómez F. A., Jenkins A., Marinacci F., Pakmor R., Springel V., 2024, *MNRAS*, **532**, 1814
- Gratton R. G., Carretta E., Matteucci F., Snenen C., 2000, *A&A*, **358**, 671
- Hafen Z., et al., 2022, *MNRAS*, **514**, 5056
- Hayden M. R., et al., 2015, *ApJ*, **808**, 132
- Hayden M. R., Recio-Blanco A., de Laverny P., Mikolaitis S., Worley C. C., 2017, *A&A*, **608**, L1
- Haywood M., Lehnert M. D., Di Matteo P., Snaith O., Schultheis M., Katz D., Gómez A., 2016, *A&A*, **589**, A66
- Haywood M., Di Matteo P., Lehnert M. D., Snaith O., Khoperskov S., Gómez A., 2018, *ApJ*, **863**, 113
- Helmi A., 2020, *ARA&A*, **58**, 205
- Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, *Nature*, **563**, 85
- Hoffman M. D., Gelman A., 2011, *arXiv e-prints*, p. arXiv:1111.4246
- Imig J., et al., 2025, *ApJ*, **990**, 203
- Inoue S., Dekel A., Mandelker N., Ceverino D., Bournaud F., Primack J., 2016, *MNRAS*, **456**, 2052
- Kawata D., Chiappini C., 2016, *Astronomische Nachrichten*, **337**, 976
- Kawata D., Grand R. J. J., Hunt J. A. S., Ciucă I., 2024, *arXiv e-prints*, p. arXiv:2412.12252
- Kereš D., Katz N., Weinberg D. H., Davé R., 2005, *MNRAS*, **363**, 2
- Khoperskov S., et al., 2025, *A&A*, **700**, A89
- Lane J. M. M., Bovy J., Mackereth J. T., 2023, *MNRAS*, **526**, 1209

- Lian J., et al., 2022, *MNRAS*, **513**, 4130
- Lian J., Bergemann M., Pillepich A., Zasowski G., Lane R. R., 2023, *Nature Astronomy*, **7**, 951
- Lian J., Zasowski G., Chen B., Imig J., Wang T., Boardman N., Liu X., 2024, *Nature Astronomy*, **8**, 1302
- Lindgren L., et al., 2021a, *A&A*, **649**, A2
- Lindgren L., et al., 2021b, *A&A*, **649**, A4
- Mackereth J. T., et al., 2017, *MNRAS*, **471**, 3057
- Mackereth J. T., et al., 2019, *MNRAS*, **482**, 3426
- Majewski S. R., et al., 2017, *AJ*, **154**, 94
- Morrow A., Grand R. J. J., Fragkoudi F., Martig M., 2024, *MNRAS*, **531**, 1520
- Miglio A., et al., 2021, *A&A*, **645**, A85
- Minchev I., Steinmetz M., Chiappini C., Martig M., Anders F., Matijevic G., de Jong R. S., 2017, *ApJ*, **834**, 27
- Montalbán J., et al., 2021, *Nature Astronomy*, **5**, 640
- Naidu R. P., et al., 2021, *ApJ*, **923**, 92
- Nidever D. L., et al., 2015, *AJ*, **150**, 173
- Noguchi M., 1998, *Nature*, **392**, 253
- Noguchi M., 2018, *Nature*, **559**, 585
- Phan D., Pradhan N., Jankowiak M., 2019, *arXiv e-prints*, p. [arXiv:1912.11554](https://arxiv.org/abs/1912.11554)
- Planck Collaboration et al., 2014, *A&A*, **571**, A16
- Price-Whelan A. M., 2017, *The Journal of Open Source Software*, **2**, 388
- Prochaska J. X., Naumov S. O., Carney B. W., McWilliam A., Wolfe A. M., 2000, *AJ*, **120**, 2513
- Rahimi A., Carrell K., Kawata D., 2014, *Research in Astronomy and Astrophysics*, **14**, 1406
- Renaud F., Agertz O., Read J. I., Ryde N., Andersson E. P., Bensby T., Rey M. P., Feuillet D. K., 2021, *MNRAS*, **503**, 5846
- Sanders J. L., Kawata D., Matsunaga N., Sormani M. C., Smith L. C., Minniti D., Gerhard O., 2024, *MNRAS*, **530**, 2972
- Sellwood J. A., Binney J. J., 2002, *MNRAS*, **336**, 785
- Shetrone M., et al., 2015, *ApJS*, **221**, 24
- Silva Aguirre V., et al., 2018, *MNRAS*, **475**, 5487
- Smith V. V., et al., 2021, *AJ*, **161**, 254
- Snaith O. N., Haywood M., Di Matteo P., Lehnert M. D., Combes F., Katz D., Gómez A., 2014, *ApJ*, **781**, L31
- Snaith O., Haywood M., Di Matteo P., Lehnert M. D., Combes F., Katz D., Gómez A., 2015, *A&A*, **578**, A87
- Spitoni E., Matteucci F., Gratton R., Ratcliffe B., Minchev I., Cescutti G., 2024, *A&A*, **690**, A208
- Tan V. Y. Y., et al., 2024, *arXiv e-prints*, p. [arXiv:2412.07829](https://arxiv.org/abs/2412.07829)
- Tsukui T., Wisnioski E., Bland-Hawthorn J., Freeman K., 2025, *MNRAS*, **540**, 3493
- Vasiliev E., 2018, *arXiv e-prints*, p. [arXiv:1802.08255](https://arxiv.org/abs/1802.08255)
- Wilson J. C., et al., 2019, *PASP*, **131**, 055001
- Yoshii Y., 1982, *PASJ*, **34**, 365
- Yu Z., et al., 2021, *ApJ*, **912**, 106
- Zhang H., Sanders J. L., 2023, *MNRAS*, **521**, 1462
- Zhang H., et al., 2025, *ApJ*, **983**, L10
- van Dokkum P. G., et al., 2013, *ApJ*, **771**, L35

APPENDIX A: MOCK DATA VALIDATION

We here demonstrate that our model-fitting method described in Section 3 can recover the known parameters of mock galaxy data. We used AGAMA to generate mock data with the same spatial and age distribution as our observational data, but assigning kinematics based on a quasi-isothermal DF (eq. 4) with the assumed parameter values as a function of age. Then, we fitted the mock data using a quasi-isothermal DF, modelling the age dependence of the parameters with cubic splines. To generate mock data we used the same galactic potential and solar position as mentioned in Section 3.

The age-dependent parameters for the DF used to generate mock

data are set as:

$$\begin{aligned}
 R_{\text{disc}} &= \begin{cases} -0.25 \times \tau + 4.2 & \text{if } \tau \leq 10, \\ 1.7 & \text{otherwise,} \end{cases} \\
 \sigma_{R,0} &= 2.5 \times \tau + 25, \\
 \sigma_{z,0} &= 3 \times \tau + 12, \\
 R_{\sigma,R} &= \begin{cases} 30 & \text{if } \tau \leq 7, \\ -\frac{22}{3} \times \tau + \frac{244}{3} & \text{if } 7 < \tau \leq 10, \\ 8 & \text{if } \tau > 10, \end{cases} \\
 R_{\sigma,z} &= \begin{cases} 42 & \text{if } \tau \leq 7, \\ -\frac{34}{3} \times \tau + \frac{364}{3} & \text{if } 7 < \tau \leq 10, \\ 8 & \text{if } \tau > 10, \end{cases}
 \end{aligned} \tag{A1}$$

where τ is in Gyr, R_{disc} , $R_{\sigma,R}$ and $R_{\sigma,z}$ are in kpc and $\sigma_{R,0}$ and $\sigma_{z,0}$ are in km s^{-1} . Based on the stellar age of each star in the observational dataset, we determined five parameters for the DF using equation (A1). We generated 10,000 sample stars for each observed star using its corresponding DF defined by the parameters described above. Among these samples, we selected the one whose (R, z) position was closest to that of the corresponding observed star. Next, we replaced the azimuthal angle of the selected data point with that of the observed star, keeping its R and z coordinates fixed. The stellar age of the observed star was then assigned to the selected point. This procedure allows us to incorporate observational uncertainties in a manner consistent with the real data. We note that the observed (R, z) positions include measurement errors, while the matched samples are drawn from error-free distributions. By repeating this process for all stars in the observational data, we generated a mock dataset containing the same number of stars as the observational data. The mock data have a similar spatial distribution and an identical stellar age distribution to the observational data. Thanks to the sampling method above, the velocities of the generated mock data follow the DF whose parameters vary according to equation (A1).

The mock data were converted into observable coordinates, $(\ell, b, \varpi, \mu, v_{\parallel})$. We then assigned the observational errors for each star, including the uncertainty in stellar age, to the corresponding mock data points. We used the log-uniform priors described in equation (9) and the smoothing priors given in equation (10). For the importance sampling, we assumed that f'_τ in equation (5) shares the same age dependence as that of the target mock data described in equation (A1).

Panels (a)-(e) in Fig. A1 show the results of our fitting with 10 knots indicated by crosses. The solid blue line shows the median of the sampled spline curves from the posterior distribution of the parameters, while the blue-shaded region shows the 1σ dispersion range. The thick orange lines show the true parameters used to generate the target mock data, as described in equation (A1). Panel (f) shows the age distribution of the mock data. The stellar ages in the mock sample are taken from the observational data so that the age distribution of the mock data is identical to that of the observational data.

Fig. A1 shows that our fitting successfully recovers the true parameter values within approximately 1σ over most of the age range. Panel (a) of Fig. A1 shows that R_{disc} fluctuates at $\tau > 10$ Gyr even though there are no corresponding features in the mock data. At 10 Gyr, R_{disc} is smaller than at 12 Gyr and shows a dip similar to that seen in the observed data. However, the depth of this dip is less than 1σ , while the dip in the real data reaches approximately 4σ (Fig. 2). The uncertainties in the estimates of each parameter are

mainly dominated by the small number of data points. The number of stars is lower, particularly at $\tau \lesssim 2$ Gyr and $\tau \gtrsim 14$ Gyr, as shown in panel (f). In this age range, all parameters exhibit relatively large uncertainties. The error ranges for $\sigma_{R,0}$ and $\sigma_{z,0}$ appear small across all ages simply because these parameters are inherently easier to constrain with relatively high accuracy. Although the uncertainties remain small even in the age range $\tau \lesssim 2$ Gyr, where the number of stars is low, this is a result of both the smaller estimated parameter values and their tighter constraints. The uncertainties in $R_{\sigma,R}$ and $R_{\sigma,z}$ increase sharply for stars younger than 7 Gyr, despite the large number of stars around that age range. This is likely because the larger values of these parameters reduce the radial contrasts in velocity dispersions, making the fitting less stable.

It should be noted that the results exhibit some variation due to the stochastic nature of the sampling process to generate the mock data from the limited number of observed stars. The validation result shown in Fig. A1 represents the results from one realisation of the mock data. We have also generated several mock data using different random sampling from the above-mentioned methodology, keeping the DF same. We confirm that the results are consistent with each other roughly within 1σ . Nevertheless, overall, the parameter recoveries achieved by our fitting are good, and the associated uncertainties are statistically meaningful. This demonstrates that the model can robustly capture age-dependent trends in Galactic parameters.

APPENDIX B: VERIFICATION OF OUR BEST-FIT MODEL

To verify the fitting results presented in Section 4, we constructed mock samples based on the best-fitting age-dependent DF parameters (shown in Fig. 2). For each observed star, we used AGAMA to get the velocity DF $f(\mathbf{v}|\mathbf{x}_i)$ at its position \mathbf{x}_i , and generated 100 velocity samples from this DF. Random shifts were then applied to the parallax, proper motions, and line-of-sight velocity of each sample, based on the corresponding observational uncertainties. Finally, we computed the actions for all generated mock stars.

The comparison of the action distribution between the best-fit model and the observational data are shown in Fig. B1. For clarity, the samples are divided into four groups by age with almost equal sample sizes. The action distributions exhibit distinct structures in different age groups, which is consistent with the result shown in Section 4. We find generally good agreement between our best-fit DF model and the observational data across all age ranges. The histogram of angular momentum, L_z appears to be less consistent than the other two physical quantities. These may be due to resonances caused by sub-structures, such as a bar, that are not incorporated in the models used in this study. For stars younger than 10 Gyr, there is an excess around $L_z = 1850 \sim 2000$ kpc km s⁻¹ and a deficit around $L_z = 2200$ kpc km s⁻¹ that are not reproduced by the model. For stars older than 10 Gyr, there is an excess around $L_z \sim 1650$ kpc km s⁻¹ and a deficit around $L_z = 1300 \sim 1500$ kpc km s⁻¹.

APPENDIX C: VERIFICATION OF THE DIP IN SCALE LENGTH

We test the robustness of the dip in scale length around $\tau = 10$ Gyr found in the fitting results for the observational data presented in Section 4. Here, using the same method as in Appendix B, we compared the action distribution of 1,295 stars with ages between $\tau = 9.5$ and 10.5 Gyr in the observational data and in two mock data sets generated from two different DF models. Model 1 follows the best-fit age

dependence of each parameter obtained from the fitting result shown in Fig. 2, same as the model used in Appendix B. Model 2 is identical to Model 1 except that it assumes a scale length $R_{\text{disc}} = 1.72$ kpc, which is 3σ larger than the R_{disc} value at the dip found in our results and similar to the scale length of the old thick disc (see Table. 2).

The left and right panels in Fig. C1 compare the action distributions of the observational data with those of Model 1 and Model 2, respectively. The only difference between the models is a shift of approximately 0.2 kpc in R_{disc} , yet both provide similarly good fits to the data. However, upon closer inspection, the J_R histogram in Model 2 (right-hand panel) is slightly more strongly concentrated around zero than in the observational data or Model 1, indicating a mild mismatch with the observed distribution. There appears to be a subtle difference between the two models around $J_R = 100$ kpc km s⁻¹, where the data shows several fluctuations. Model 1 seems to go through the middle of this wobble, while Model 2 goes through the lower side of this wobble. J_z shows little difference between the models. This may suggest that differences in R_{disc} do not have a significant effect on J_z . For L_z , there is a bimodality around $L_z \sim 1600$ kpc km s⁻¹ in the observation, but this is not reproduced in either model. This may be due to a resonance from the Galactic bar, which is not considered in our model. Still, Model 1 reproduces a distribution slightly closer to the observation than Model 2, as the sharpness around the peak is slightly weaker and has a wider distribution. Hence, we consider that this dip in R_{disc} at $\tau \sim 10$ Gyr is meaningfully inferred from the data.

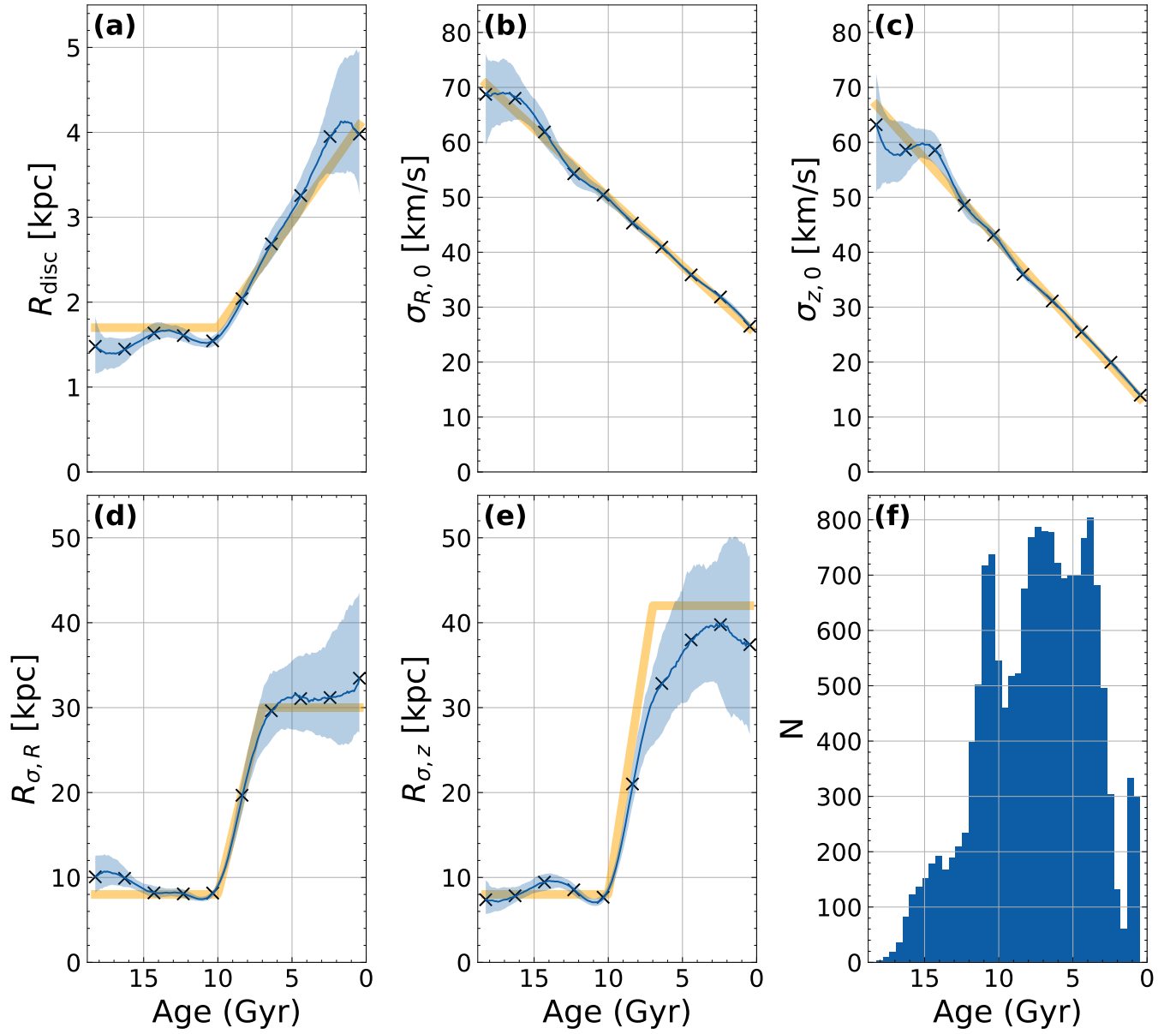


Figure A1. Results of model validation with a mock observational data. Panels (a), (b), (c), (d) and (e) present the model fitting results for parameters of R_{disc} , $\sigma_{R,0}$, $\sigma_{z,0}$, $R_{\sigma,R}$ and $R_{\sigma,z}$ as a function of age, respectively, and panel (f) shows stellar age distribution of the mock data. The blue solid lines indicate the median probability of the fitting results for each parameter, while the shaded blue regions represent the 1σ confidence intervals. The thick orange solid lines show the age-dependent functional forms for each parameter, used to generate the mock data, i.e. true parameter values. Crosses denote the position of 10 evenly spaced knots set for the cubic spline fitting of the age dependence of each parameter.

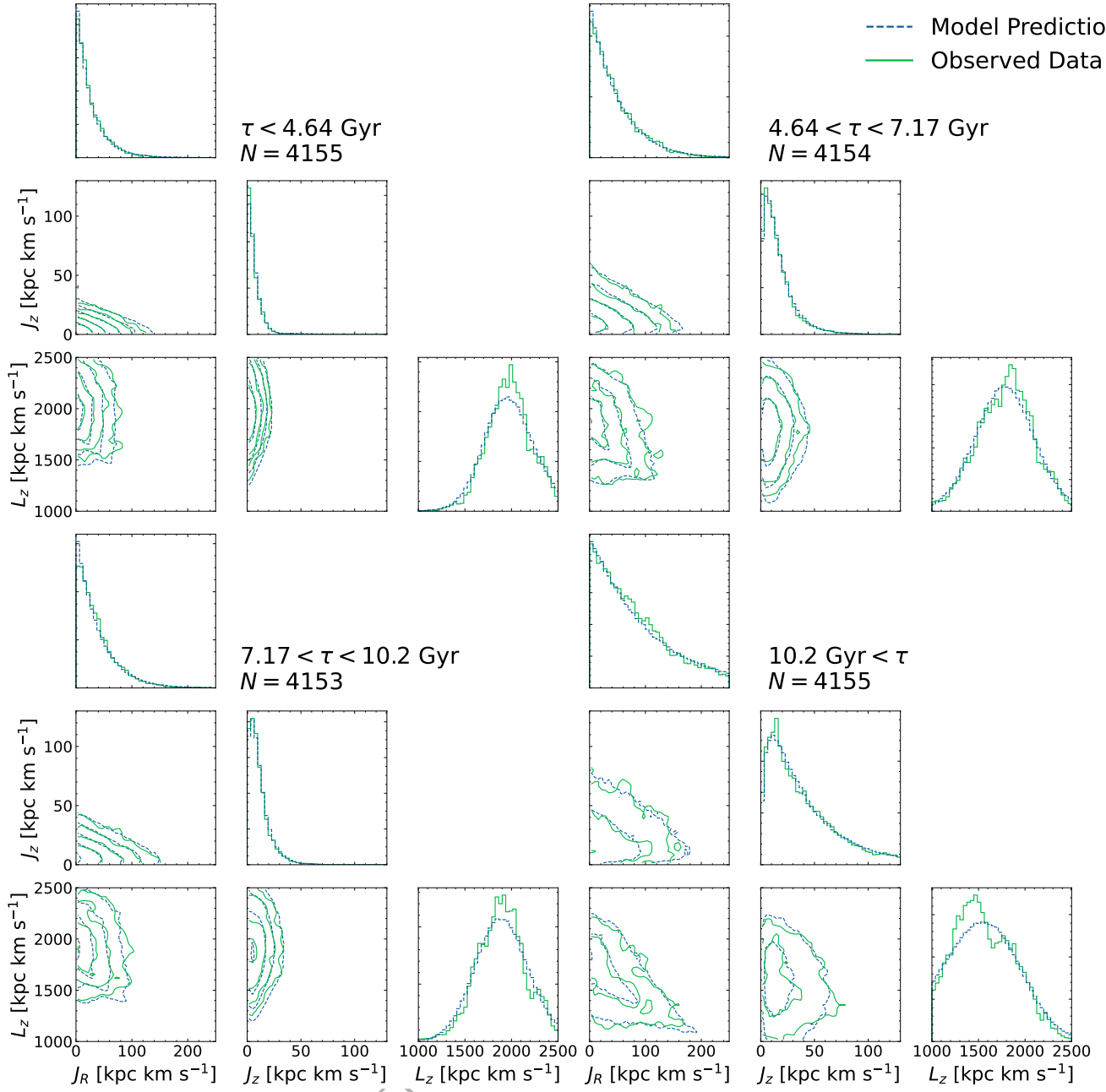


Figure B1. Comparison of the action-space distributions between our best-fit DF model (blue dashed lines), shown in Fig. 2, and the observed data (green solid lines). For clarity, the sample is divided into four different stellar age groups. Overall, the model reproduces the observed distributions well.

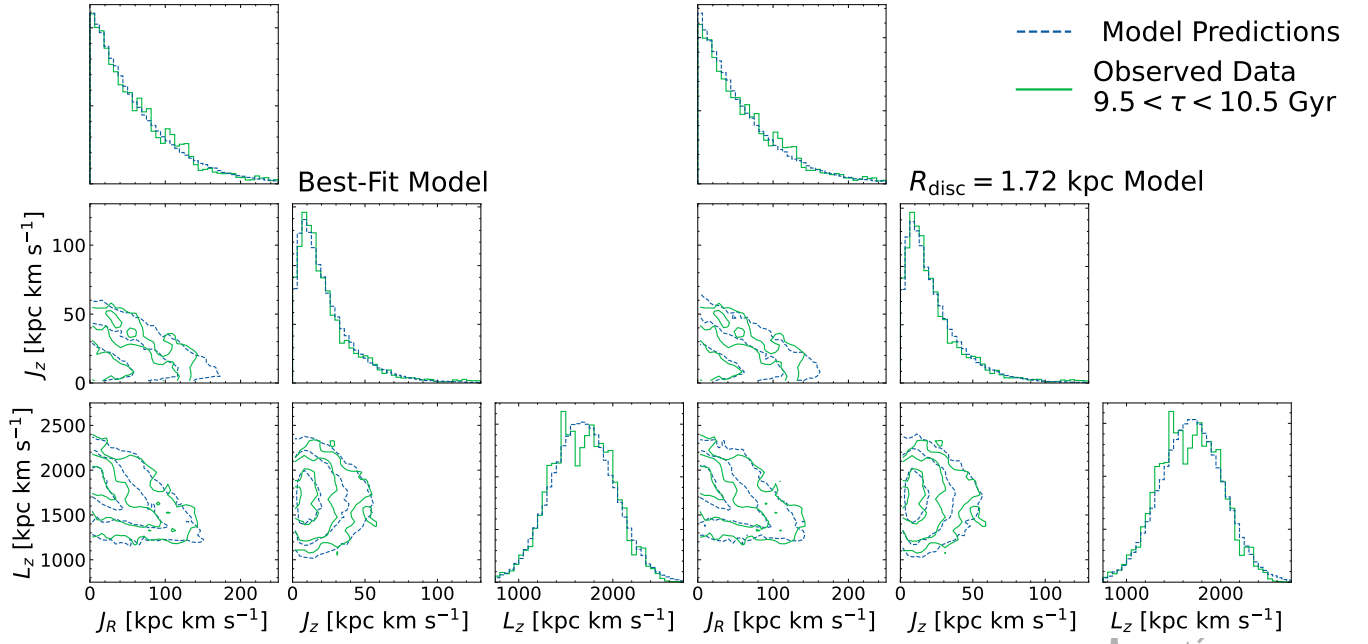


Figure C1. Comparison of action-space distributions for stars with ages in the range $9.5 < \tau < 10.5$ Gyr, corresponding to the epoch where a dip in R_{disc} is observed in Fig. 2. The left panel compares the observed data (green solid lines) with predictions from the best-fit model (blue dashed lines), using the parameter estimates shown in Fig. 2. The right panel shows a comparison with an alternative model in which R_{disc} is fixed at 1.72 kpc, simulating a scenario without the observed dip—while keeping all other parameters identical to those of the best-fit model. Both models appear to be in good agreement with the observed data, but when focusing on the shape of the distribution of the J_R and L_z models, the best-fit model fits the observed data slightly better. This comparison highlights the impact of the R_{disc} dip on the resulting action distributions.