

The Inside-out Growth of the Galactic Disk

NEIGE FRANKEL,¹ JASON SANDERS,² HANS-WALTER RIX,¹ YUAN-SEN TING(丁源森),^{3,4,5,*} AND MELISSA NESS^{6,7}

¹*Max Planck Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*

²*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA*

³*Institute for Advanced Study, Princeton, NJ 08540, USA*

⁴*Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA*

⁵*Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA*

⁶*Department of Astronomy, Columbia University, 550 W 120th St, New York, NY 10027, USA*

⁷*Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA*

(Accepted in ApJ)

ABSTRACT

We quantify the inside-out growth of the Milky Way’s low- α stellar disk, modelling the ages, metallicities and Galactocentric radii of APOGEE red clump stars with $6 < R < 13$ kpc. The current stellar distribution differs significantly from that expected from the star formation history due to the redistribution of stars through radial orbit mixing. We propose and fit a global model for the Milky Way disk, specified by an inside-out star formation history, radial orbit mixing, and an empirical, parametric model for its chemical evolution. We account for the spatially complex survey selection function, and find that the model fits all data well. We find distinct inside-out growth of the Milky Way disk; the best fit model implies that the half-mass radius of the Milky Way disk has grown by 43% over the last 7 Gyr. Yet, such inside-out growth still results in present-day age gradient weaker than 0.1 Gyr kpc^{-1} . Our model predicts the half-mass and half-light sizes of the Galactic disk at earlier epochs, which can be compared to the observed redshift -size relations of disk galaxies. We show that radial orbit migration can reconcile the distinct disk-size evolution with redshift, also expected from cosmological simulations, with the modest present-day age gradients seen in the Milky Way and other galaxies.

Keywords: Galaxy: abundances — Galaxy: disk — Galaxy: evolution — Galaxy: formation — ISM: abundances — stars: abundances

1. INTRODUCTION

The star formation history is a key formation and evolution aspect for any disk galaxy, as it governs its resulting stellar structure. Subsequent orbit mixing processes can redistribute the stars and remove dynamical memory of their birth conditions. Constraining galaxy evolution requires knowledge of both the initial formation distribution and the importance of orbit mixing.

Galaxy disks are thought to grow from inside-out. Peebles (1969) first postulated that gas disks acquire their angular momenta from tidal torques; with the low angular momentum gas cooling, settling and forming

stars on shorter time-scales than the high angular momentum gas (Larson 1976). (semi-)Analytic hierarchical models for disk formation in a cosmological context (e.g., Fall & Efstathiou 1980; Mo et al. 1998; Somerville et al. 2008; Dutton et al. 2011) have predicted that galaxy disks must grow from inside-out to reproduce the observed size-luminosity-velocity relations (Tully & Fisher 1977; Courteau et al. 2007). Similarly, cosmological and zoom-in simulations show disks with star formation time-scales that increase with the distance to the galactic center or gas accretion of higher angular momentum at later times (Avila-Reese et al. 2018; Grand et al. 2017; Aumer et al. 2014; Brook et al. 2006, 2012; Pilkington et al. 2012; Agertz et al. 2011). Measuring the rate at which stellar disks grow can give insight to constrain the interplay between the physical processes

frankel@mpia.de

* Hubble fellow

involved in gas accretion, cooling and forming stars, as well as the global evolution of disk galaxies.

Observationally, a number of independent measurements suggests that disk galaxies grow from inside-out. Several studies indicate that disk galaxies of a given stellar mass are smaller at higher redshift (Ferguson et al. 2004; Barden et al. 2005; Franx et al. 2008; Buitrago et al. 2008; van Dokkum et al. 2013; van der Wel et al. 2014; Rodríguez-Puebla et al. 2017). Likewise, in massive disk galaxies of the local Universe, integrated light from older stellar populations, as traced by color, is more centrally concentrated than that of younger populations, and star formation is more spatially extended than the overall stellar distribution (e.g. MacArthur et al. 2004; Muñoz-Mateos et al. 2007; Boissier et al. 2008; Muñoz-Mateos et al. 2011; Wang et al. 2011; Pezzulli et al. 2015). Similarly, resolved stellar observations of Local Group galaxies (e.g. NGC300, M33 and more recently NGC7793) using the *Hubble Space Telescope* showed that old populations are often more centrally concentrated than young populations (Gogarten et al. 2010; Williams et al. 2009; Sacchi et al. 2019). However, this does not appear to be the case in M31 (Bernard et al. 2015). Finally, integral field spectroscopy surveys, such as MaNGA (Bundy et al. 2015) and CALIFA (Sánchez et al. 2012) have led to the detection of mostly weak age radial gradients in massive disk galaxies (Goddard et al. 2017; García-Benito et al. 2017; González Delgado et al. 2014; Pérez et al. 2013).

Most of the described galaxy observations are consistent with inside-out growth, but are restricted to studying present-day galactocentric radii of stellar populations, instead of considering their (unknown) birth sites. Dynamical processes in disk galaxies rearrange the stars, often weakening or erasing any formation gradients (e.g. Minchev et al. 2014). Therefore, to state anything quantitative about inside-out growth from these observations requires knowledge of the strength of these dynamical processes. The importance of dynamical heating from non-axisymmetric perturbations such as spiral arms and molecular clouds has long been recognised (e.g. Lacey 1984; Carlberg & Sellwood 1985; Sellwood 2014; Aumer et al. 2016). Transient spiral patterns produce changes in angular momentum and heating around the Lindblad resonances, restructuring the disc (Lynden-Bell & Kalnajs 1972). Sellwood & Binney (2002) recognized that at corotation, a star could be scattered in angular momentum without associated increase in random motion. This process is known as ‘radial migration’ or ‘churning’ and is generally distinguished from ‘blurring’ which describes changes in angular momentum with associated heating. A truly radial-migrated population

will appear dynamically identical to a population that has not experienced any dynamical mixing (Sellwood 2014), making their separation hard. In this work, we ignore kinematics of the stars so cannot distinguish between the different dynamical processes that scatter stars radially, and consider the combined effect as ‘radial orbit migration’. If radial mixing is strong, then the age gradients expected from inside-out growth can be severely weakened. An old star at a large Galactocentric radius can have formed at a smaller radius, erasing evidence for inside-out growth. All these aspects (inside-out star formation, radial migration) must be accounted for simultaneously in any modelling as demonstrated in Schönrich & McMillan (2017a); Minchev et al. (2019); Frankel et al. (2018).

For the Milky Way, we have an opportunity to disentangle the effect of inside-out formation from radial orbit migration as we have access to positions, chemical compositions and ages of individual stars from spectroscopic observations. Although dynamical processes cause stars to lose dynamical memory of their birth sites, they are thought to retain chemical memory. With assumptions on the past history of the Galaxy, these observations allow the linking of stars to their birth locations from their chemistry or ages through ‘weak chemical tagging’, (different from the classical ‘chemical tagging’ as described in Freeman & Bland-Hawthorn (2002) and Ting et al. (2015) for example) such that the dynamical processes can be ‘rewound’ and the formation properties measured (e.g., Schönrich & Binney 2009a; Sanders & Binney 2015; Hayden et al. 2015; Schönrich & McMillan 2017b; Frankel et al. 2018; Minchev et al. 2018).

The recent advance of wide spectroscopic surveys of the Milky Way disk has produced pioneering work in Galactic archaeology enabled by unprecedented samples of $\sim 10^5 - 10^6$ stars well beyond the solar neighbourhood. Earlier work considered the distribution of abundances at each Galactic location $p([\alpha/\text{Fe}], [\text{Fe}/\text{H}] | R)$ (e.g. Hayden et al. 2015) or $p(R | [\alpha/\text{Fe}], [\text{Fe}/\text{H}])$ (e.g. Bovy et al. 2012), finding that stars with high $[\alpha/\text{Fe}]$ were more centrally concentrated. But $[\alpha/\text{Fe}]$ was used as a chemical clock proxy for stellar age, and the focus was on the basic differences between the high- α and the low- α disks. More recently, stellar ages have become available for many stars, making it feasible to study the age distribution at different locations, and the spatial distributions of mono age populations $p(R | \tau, [\text{Fe}/\text{H}])$ – a more explicit measure of the history of the Galaxy (e.g. Bergemann et al. 2014; Bensby et al. 2010; Mackereth et al. 2017). However, as the Galaxy evolves it is non-trivial to relate these distributions to the star formation history of

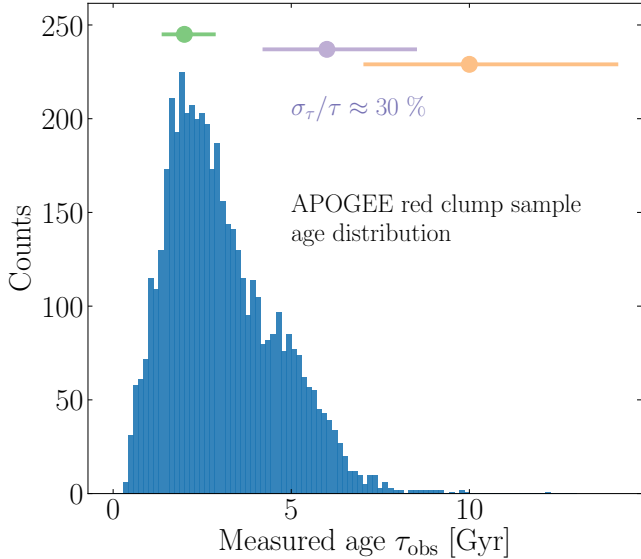


Figure 1. Age distribution of the APOGEE red clump stars from the low- α Galactic disk, used in the analysis presented here. These ages were determined in Ting & Rix (2018), tied to asteroseismic mass estimates. Age uncertainties are $\sigma_{\log_{10} \tau} \approx 0.15$ dex, resulting in about 30% age uncertainties, as illustrated for three ages: 2 Gyr (green), 6 Gyr (purple), 10 Gyr (orange). The peak in the distribution at 2 Gyr does not reflect the age distribution of the “underlying” stellar population, but is expected from theory and can be modelled quantitatively: it reflects the mass dependence – and thereby age dependence – of the life time of the core helium burning evolutionary stage that defines red clump stars.

the disk. Here, we set out to build a model for where and when stars were born over a large range of Galactocentric radii $p(R_0, \tau)$, with direct use of stellar ages, and accounting for radial orbit migration. We focus on the low- α disk only (the last 8 Gyr of the Milky Way evolution), propose an evolution scenario through parametrized equations and fit the parameters using APOGEE data.

This paper is the second in a series, developing and applying a framework for a global evolutionary Milky Way disk model introduced in Frankel et al. (2018), with emphasis on inside-out growth. We present the data we model in Section 2. In Section 3, we describe aspects of the model itself: (1) the survey selection function and (2) the Galactic disk; this model is an extension of Frankel et al. (2018), where it is described in some detail. We then present the results of the model fit to the data in Section 4. Finally, we interpret these in a more global context of galaxy disk formation and evolution and compare them to disk galaxies observed at different redshifts in Section 5, and discuss the limitations in Section 6.

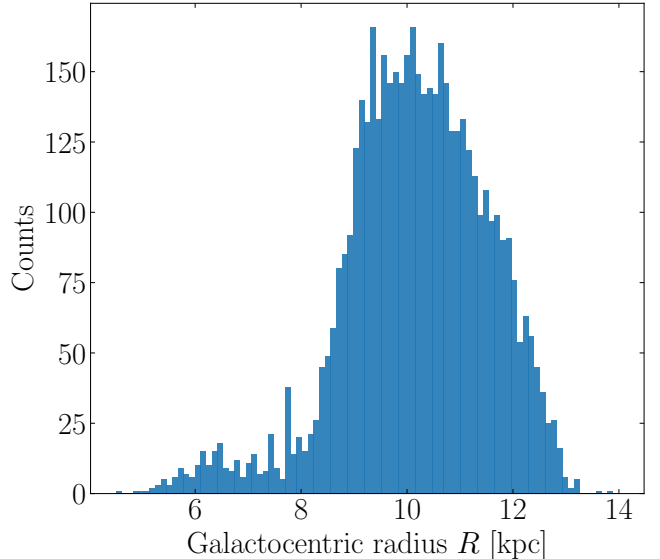


Figure 2. Galactocentric radius distribution of the red clump stars (as in Fig. 1), used in this analysis. They span a range of 6 kpc to 13 kpc, with the vast majority of sample stars beyond the Solar radius; this latter property traces back to the Galactocentric radius R distribution of the initial Bovy et al. (2014) parent sample and the subsequent selection cuts we have imposed (Section 2).

2. DATA: APOGEE RED CLUMP GIANTS

We use asteroseismically calibrated ages, $[\text{Fe}/\text{H}]$ and 3D positions of red clump stars from the 12th data release of the APOGEE near-infrared spectroscopic survey (Apache Point Observatory for Galactic Evolution Experiment, Alam et al. 2015; Majewski et al. 2017). Red clump stars are low mass core helium burning stars that have gone through the helium flash. Since they have similar core masses, they have similar luminosities, which make them good standard candles and suitable for Galactic archaeology studies that require precise distances. The core helium burning stage life-time is longer for initially more massive stars. Therefore, the overall red clump population is fairly young, with an age distribution that peaks around 2 Gyr (Girardi 2016), as can be seen in Fig. 1.

Bovy et al. (2014) describes the selection of the red clump population from the APOGEE data. It is based on cuts in stellar parameters $\log g$, T_{eff} and infra-red photometry $(J - K)_0$ and H . We cross-match this red clump catalog with that of Ting et al. (2018) to remove possible contaminants. Ting et al. (2018) used a data-driven approach trained on APOKASC2 (Pinsonneault et al. 2018) to predict asteroseismic parameters (and hence the evolutionary stage) from stellar spectra, and evaluate their contamination fraction to 3%.

We further restrict our sample to the low alpha stars as in Frankel et al. (2018), tracing the most recent evolution of the Milky Way disk, in order to cut down undesired information from older stars. In addition, we consider only the ‘short cohort’ fields of APOGEE, as defined in Zasowski et al. (2013), which contain the brightest and closest objects (with approximately $7 < H < 12$). Stars further away, in longer ‘cohorts’ because they are fainter, may have higher extinctions, at a level where modelling extinction may be challenging. We restrict our sample to the APOGEE fields where the median extinction of APOGEE stars is less than ≈ 0.6 in H band. After these cuts, our sample consists of ~ 5381 stars in 142 APOGEE fields.

We use metallicities $[\text{Fe}/\text{H}]$ known to about ± 0.05 dex and 3D positions (l, b, D) with (l, b) standing for Galactic longitude and latitude and with photometric distances D , known to about 7% for such standard candles, from Ness et al. (2016). The ages τ_{obs} known to about 30% are taken from Ting & Rix (2018). APOGEE has delivered measurements of 15 chemical abundances, which we could, in principle, use. We choose to restrict to $[\text{Fe}/\text{H}]$ and age, because they are sufficient for our purpose: at given $[\text{Fe}/\text{H}]$ and τ , other abundances $[\text{X}/\text{Fe}]$ of the low- α disk can be predicted with good precision (Ness et al. 2019).

The ages were determined using neural networks trained on stars that have asteroseismic age estimates from APOKASC2 (Pinsonneault et al. 2018). The neural networks were trained to predict ages from stellar spectra in Ting & Rix (2018). This age determination method applies the same philosophy as the work described in Ness et al. (2016), who used a quadratic model to map from stellar spectra to ages rather than non-linear functions. It was shown, in a separate works using respectively asteroseismic ages (Martig et al. 2016; Silva Aguirre et al. 2018) and Bayesian isochrone fitting (Feuillet et al. 2016, 2018), that most of the spectral information on the stellar mass (and hence stellar age) comes from surface abundances of the CNO cycle elements brought up during the mass dependent dredge-up process.

Since the largest source of uncertainties comes from the ages (i.e. the age uncertainties are much larger than metallicity and distance uncertainties), we will account for them through a noise model, and treat the distances and metallicities as noise-free variables. The observed age distribution of the resulting sample is illustrated in Figure 1, and the Galactocentric radius distribution is shown in Figure 2. Most of the stars of our sample are young and located in the outer 8-12 kpc of the Milky Way disk. Therefore, our modelling

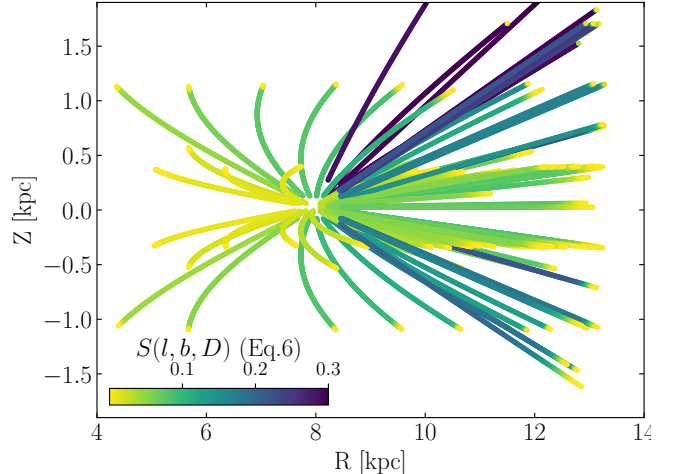


Figure 3. Distance-dependent fraction of observed red clump stars in the APOGEE pointings used here. This is the product of the fraction of photometrically eligible stars in each of the 142 fields that were spectroscopically targetted (Eq. 4), and the distance-dependent probability that a star of red clump luminosity does not have prohibitive dust extinction, as calculated from the Green et al. (2018) extinction map (Eq. 5).

will describe mainly the recent evolution of the outer disk. Our final data set consists of these 5381 stars with $\mathcal{D} = \{l, b, D, [\text{Fe}/\text{H}], \tau_{\text{obs}}\}$.

3. MODELLING THE DATA SET

We set out a global model for the data described above: $\mathcal{D} = \{l, b, D, [\text{Fe}/\text{H}], \tau_{\text{obs}}\}$ and their uncertainties. We make a clear difference between the measured age τ_{obs} and the true age τ that we use for the modelling, to account rigorously for age uncertainties (see Section 3.3). For clarity, all the model variables and parameters are summarized in Table 1. We start by splitting this model in three main terms: (1) the model for the selection of the disk stars in the APOGEE survey, (2) a global forward model for the Galactic disk with model parameters in an array \mathbf{p}_m and (3) the normalization constant over the observable space limited by the selection as in Rix & Bovy (2013).

$$p(\mathcal{D} | \mathbf{p}_m, \text{selection}) = \frac{S(l, b, D) f_{\text{RC}}(\tau) p(\mathcal{D} | \mathbf{p}_m)}{V_s(\mathbf{p}_m)} \quad (1)$$

where $p(\mathcal{D} | \mathbf{p}_m, \text{selection})$ is the normalized model of the data set, $V_s(\mathbf{p}_m)$ is a normalization constant, $p(\mathcal{D} | \mathbf{p}_m)$ is the physical model for the Galactic disk, and $S(l, b, D) = p(\text{selection} | l, b, D)$ is the survey selection function: the probability that a star ends up in the catalog, given its properties: position (l, b) , magnitude H or distance D in our particular case for a standard candle, dereddened color $(J - K_s)_0$, and integrated extinction $A_H(D)$ along the line of sight (Section 3.1). The

Symbol	Name	Type and appearance in the text
l, b, D	Galactic longitude, latitude, distance from the Sun	Assumed noise-free observables
X, Y	Cartesian coordinates for stellar positions in the plane	Disk model variables
R, Z	Galactocentric radius, height above the Galactic plane	Disk model variables
R_0	Galactocentric radius of a star at its birth	Disk model variable
$[\text{Fe}/\text{H}]$	Metallicity of stars	Assumed noise-free observable
τ_{obs}	Measured stellar age	Observable that has errors
τ	Modelled true age	Disk model variable
$\sigma_{\log_{10} \tau}$	Uncertainties in measured log age	Noise model parameter, fixed
R_d	Integrated scale-length of the star-forming disk	Element of \mathbf{p}_m in Eq. 10, to fit
$\tau_{\text{SFR}}, \tau_m, x$	Star formation time-scale, onset time, inside-out degree	Elements of \mathbf{p}_m in Eq. 11, to fit
$h_z(a_z)$	Disk scale-height as a function of the scaling parameter a_z	Nuisance parameter in Eq. 16, to fit
R_{old}	Scale-length of the disk for $\tau > \tau_m$	Nuisance parameter in Section 3.2.5, to fit
σ_{RM7}	Strength of radial orbit mixing	Element of \mathbf{p}_m in Section 3.2.2, to fit
$\gamma_{[\text{Fe}/\text{H}]}$	chemical evolution parameter time dependency	Nuisance parameter in Eq. 18, to fit
$R_{[\text{Fe}/\text{H}]=0}^{\text{now}}$	Radius of solar metallicity in the star-forming gas	Nuisance parameter in Eq. 17, to fit
$\nabla[\text{Fe}/\text{H}]$	$[\text{Fe}/\text{H}]$ radial gradient in the star-forming gas	Nuisance parameter in Eq. 17, to fit
Ω_i	Solid angle of APOGEE pointing i	Selection function parameter, fixed
H_{RC}	Red Clump absolute magnitude in H band	Selection function parameter, fixed
A_H, A_K	Extinction in H and K bands	Selection function variable
$(J - K)_0$	Dereddened color of a star	Selection function variable

Table 1. Main variables and parameters used in the model.

term $f_{\text{RC}}(\tau)$ is the fractional stellar mass contained in the red clump population at given age. It can be determined assuming an initial mass function and using stellar isochrones. Generally, selecting stellar populations through cuts in observables implies biasing the sample in age and metallicity. For the red clump population, Bovy et al. (2014) showed that f_{RC} is as strong function of age, and a weak function of metallicity. Therefore, we neglect the weak metallicity dependence and adopt the fit of f_{RC} as a function of age from Bovy et al. (2014).

The normalization constant (or survey volume) can be computed by integrating the unnormalized model over the observable space:

$$V_S(\mathbf{p}_m) = \int_{\mathcal{D}} p(\mathcal{D} | \mathbf{p}_m) S(l, b, D) f_{\text{RC}}(\tau) d\mathcal{D}, \quad (2)$$

which is a 5 dimensional integral over all the physical properties of the data. The integrals over (l, b) are obtainable analytically and can be transformed into a sum over APOGEE fields, assuming that the properties of stars do not vary in (l, b) over a single APOGEE pointing (of about 1.5 degree radius). We evaluate this integral in subsection 3.4. Expanding out the data and writing the spatial distributions in Cartesian coordinates (X, Y, Z) , we have

$$\begin{aligned} p(l, b, D, \tau_{\text{obs}}, [\text{Fe}/\text{H}] | \mathbf{p}_m) \\ = D^2 \cos(b) p(X, Y, Z, \tau_{\text{obs}}, [\text{Fe}/\text{H}] | \mathbf{p}_m), \end{aligned} \quad (3)$$

where $p(X, Y, Z, \tau, [\text{Fe}/\text{H}] | \mathbf{p}_m)$ is our model for the Galactic disk described in Section 3.2, and its relation to the similar term in Eq. 3 is the convolution over age uncertainties as described in Section 3.3.

3.1. Modelling APOGEE Selection Function

The APOGEE survey targeted stars in different lines of sights, or pointings, centered on directions (l, b) . Therefore, a spatial histogram of the stars contained in the APOGEE catalog does not reflect the spatial density of stars in the Milky Way disk. Using these stars to infer the spatial structure of the Milky Way disk requires forward modelling of this selection process.

The details of the APOGEE-1 selection function are described in Zasowski et al. (2013). We apply and summarize here the general method laid-out in Rix & Bovy (2013); Bovy et al. (2014) and refer the reader to these references for details¹. For most stars of the APOGEE sample, we can assign a probability that this star was observed given its location, magnitude and color (selection function), which is then combined with a probability for this star to be at this location (the density model). In most APOGEE disk fields, stars were selected according to simple magnitude cuts in the H band and cuts in

¹ we have additionally made the selection function for fields used here available as a fits table at https://github.com/NeigeF/apogee_selection_function with a tutorial

dereddened $(J-K)_0$ color from the photometric sample 2MASS (Skrutskie et al. 2006) which is assumed complete within these cuts. The de-reddened colors were obtained in Zasowski et al. (2013) using the Rayleigh Jeans Color Excess method (Majewski et al. 2011), based on combinations of photometry in different near- and mid-infrared bands and on the assumption that most stars have a similar intrinsic color in the Rayleigh-Jeans part of their spectrum. When too many stars in a pointing satisfy these criteria, the subset of stars to observe was drawn roughly randomly from the 2MASS. This subset contains a fraction S_i of all available stars, \star ,

$$S_i = \frac{\# \star \text{ in APOGEE field } i}{\# \star \text{ in 2MASS matching selec. criteria in } i}. \quad (4)$$

We focus on the main disk fields of APOGEE that are not dominated by ancillary programs (Eq. 4 is approximately valid only for targets drawn randomly from 2MASS, not for targets chosen for a specific purpose). We also remove the fields for which the selection criteria were complex to model (see a detailed description of the numerous particular cases in Zasowski et al. (2013)) or irrelevant to the scope of the present work, e.g., halo fields which contain fewer disk stars.

In addition, a fraction of stars in each field in the Milky Way disk is too extinguished by interstellar dust to be detectable within the magnitude limits of APOGEE (Bovy et al. 2016). This is a function of distance, and can be modelled, if we have a model for the extinction spatial distribution in H band. We use the Bayestar17 3D extinction map (Green et al. 2018), which predicts the extinction distribution in different bands at any (l, b, D) . We convert the map’s output values to H band using the extinction coefficient 0.468 and assume the extinction law of Indebetouw et al. (2005) ($A_H/A_K = 1.55$) to convert the K band extinction values stated by APOGEE, to the H band. Assuming red clump stars are standard candles of magnitude $H_{RC} = -1.49$ (Laney et al. 2012), the probability that a red clump star can be seen at a given distance D within APOGEE magnitude limits H_{\min} and H_{\max} is the fractional area of the field at this given distance that is not hidden by dust extinction A_H . The map’s highest resolution (3.4’) allows computation of this ratio by counting the distance-dependent fraction of pixels that satisfy the inequality

$$f_i(D | \text{field } i) = \frac{\# \text{ pix}(H_{\min} < H(D, H_{RC}, A_H) < H_{\max})}{\# \text{ pix in field } i}. \quad (5)$$

We can finally recast this into the overall selection function, dubbed ‘*effective selection function*’ when it accounts for both the probability to select a star from the

photometric sample and for the probability to see a star given dust extinction, plotted in Figure 3,

$$S(l_i, b_i, D) = S_i \times f_i(D | \text{field } i). \quad (6)$$

This is one of the three terms to cast in Eq. 1, and we are left with the two others: global model for the Galactic disk, and normalizing integral.

3.2. Global Model for the Galactic Disk Evolution

We aim to model the Galactic disk in terms of the distributions $p(\tau, [\text{Fe}/\text{H}], R, Z | \mathbf{p}_m)$ of ages, metallicities, Galactocentric radii and heights above the plane. To fit the large data set efficiently, we build the different model aspects (star formation history, enrichment, radial mixing) from parametrized functional families. These are not designed to describe the exact physics in a self-consistent manner, but are physically plausible, fast to compute and have parameters that can be physically interpreted. This methodology permits us to disentangle different effects at play and measure the scales on which they happen. Applying the probabilistic chain rule and marginalizing over the birth Galactocentric radius R_0 , we introduce the different model aspects:

$$\begin{aligned} p(\tau, [\text{Fe}/\text{H}], X, Y, Z | \mathbf{p}_m) \\ = \int p(R_0, \tau | \mathbf{p}_m) p([\text{Fe}/\text{H}] | R_0, \tau, \mathbf{p}_m) \\ \times p(X, Y, Z | R_0, \tau, \mathbf{p}_m) dR_0. \end{aligned} \quad (7)$$

The first term in the integral is the birth radius-age distribution of stars and is closely related to the star formation history, (subsection 3.2.1). The second term is the distributions of metallicity in the star forming gas as a function of time, which we adopt from Frankel et al. (2018) and re-fit. The last term is the spatial density resulting from radial orbit migration and vertical heating. It can be split into a radial component (the main focus here) and a vertical component, which depends on the distance from the Galactic mid-plane Z :

$$\begin{aligned} p(X, Y, Z | R_0, \tau, \mathbf{p}_m) = p(X, Y | R_0, \tau, \mathbf{p}_m) \\ \times p(Z | R_0, R, \tau, \mathbf{p}_m). \end{aligned} \quad (8)$$

The first term on the right hand side is the present-day surface density of stars of true age τ and birth radius R_0 (subsection 3.2.2), and the second term is the present-day vertical profile of the disk resulting from the vertical heating of stellar orbits between their birth radius and their present day radius over time (subsection 3.2.3).

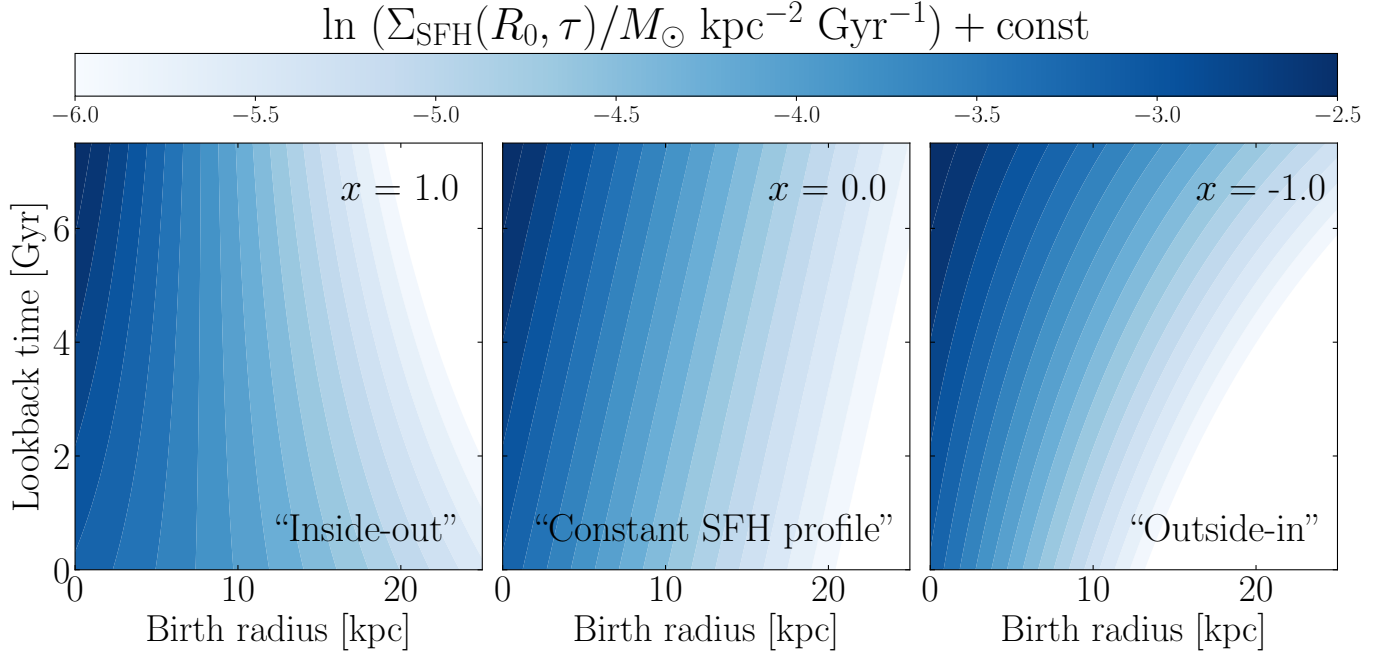


Figure 4. Model star formation rate as a function of time and birth Galactocentric radius for different scenarios considered here. We show the star formation rate surface density $\Sigma_{\text{SFR}}(R_0, \tau | \tau_{\text{SFR}}, \tau_m, x)$, normalized to yield unit total stellar mass, as a function of birth radius R_0 , lookback time τ and inside-out growth degree x as defined in Eq. 11. The panels reflect star formation histories for different ‘inside-out growth parameters’ x , from left to right: strong inside-out growth ($x = 1$) that would predict a flat star formation history at 8 kpc, uniform star formation history ($x = 0$), and strong outside-in growth, where the disk formed stars initially on a larger scale-length than at present.

3.2.1. Radial Dependent Star Formation History

We model the distribution of stars at birth as

$$\Sigma_{\text{SFR}}(R_0, \tau | \mathbf{p}_m) = \Sigma(R_0 | \mathbf{p}_m) \text{SFH}(\tau | R_0, \mathbf{p}_m), \quad (9)$$

where the time integrated surface density profile at birth is an exponential in birth radius R_0

$$\Sigma(R_0 | \mathbf{p}_m) \propto \exp\left(-\frac{R_0}{R_d}\right), \quad (10)$$

and the radially-dependent normalized star formation history SFH is a function of age τ

$$\begin{aligned} \text{SFH}(\tau | R_0, \mathbf{p}_m) &= c(R_0, \mathbf{p}_m) \\ &\times \exp\left[\frac{1}{\tau_{\text{SFR}}}\left((1 - x \frac{R_0}{8\text{kpc}})\tau - \tau_m\right)\right]. \end{aligned} \quad (11)$$

The inside-out growth is encoded in the dimensionless parameter x . If $x = 0$, the star formation time-scale is constant across the disk: no inside-out growth. If $x > 0$, stars form on a shorter time-scale in the inner disk than in the outer disk: there is inside-out growth. If $x < 0$, the star formation time-scale decreases to the outer disk: this would be outside-in growth. We fit for x by maximizing the likelihood, without imposing any prior. The values taken by $\Sigma_{\text{SFR}}(R_0, \tau | \mathbf{p}_m)$ for three different values of x (1, 0 and -1) are illustrated in Figure 4. More

generally, x sets the strength of the linear dependency of the star formation history with Galactocentric radius. For positive x and τ_{SFR} , there is always a Galactocentric radius R_{SFR0} such that the star formation rate is a decreasing function of time inside R_{SFR0} , and is an increasing function of time outside R_{SFR0} . In the particular case where $x = 1$, the star formation history is constant at 8 kpc, decreasing in the inner 8 kpc, and increasing in the outer 8 kpc. This is illustrated in the left panel of Figure 4, where at the radius $R_0 = 8$ kpc, the contour lines are vertical. Finally, the normalization constant $c(R_0, \mathbf{p}_m)$ is such that the SFH in Eq. 11 is normalized between 0 and τ_m . The model parameters to fit for are $\{\tau_m, \tau_{\text{SFR}}, x, R_d\}$. This equation is related to the first term of Eq. 7 through

$$p(R_0, \tau | \mathbf{p}_m) = 2\pi R_0 \Sigma_{\text{SFR}}(R_0, \tau | \mathbf{p}_m). \quad (12)$$

3.2.2. Radial Orbit Migration

Since the present-day Galactocentric radii of stars may be different from their birth positions, we model the effect of radial orbit migration in order to ‘rewind’ back to the birth properties of stars. We use the radial orbit migration prescription from Frankel et al. (2018) for the probability of a star moving from birth radius R_0 to current radius R in a time τ . This is modelled as a

global diffusion process with a diffusion strength $\sigma(\tau) = \sigma_{\text{RM7}}\sqrt{\tau/7}$ Gyr,

$$p(R | R_0, \tau, \mathbf{p}_m) = N(\mathbf{p}_m, \tau, R_0) \exp\left(-\frac{(R - R_0)^2}{2\sigma^2(\tau)}\right) \quad (13)$$

with $N(\mathbf{p}_m, \tau, R_0)$ a normalization constant. We re-fit for σ_{RM7} .

The surface density profile of migrated stars of age τ and from birth radius R_0 is

$$p(X, Y | R_0, \tau, \mathbf{p}_m) = \frac{1}{2\pi R} p(R | R_0, \tau, \mathbf{p}_m), \quad (14)$$

and is the first term in Eq. 8.

3.2.3. Vertical Distribution of Stars

Since the present work focuses on the radial structure of the Milky Way disk, we are essentially not interested in its vertical structure. However, we must not ignore it as the survey selection function is three dimensional. We therefore adopt a description for the vertical profile that is good enough to characterize the vertical heating history of the disk, and for which we only fit a single parameter.

The Milky Way disk scale-height depends on the ages of stellar populations, and the vertical profile of populations of given age flares (e.g., Mackereth et al. 2017; Bovy et al. 2016). We model the vertical structure of the Milky Way disk as a function of age and Galactocentric radius, using the approximation of an isothermal disk and the harmonic limit (e.g., Binney & Tremaine 2008):

$$p(Z | R, R_0, \tau, \mathbf{p}_m) = \frac{1}{2h_z(R, R_0, \tau)} \text{sech}^2\left(\frac{Z}{h_z(R, R_0, \tau)}\right), \quad (15)$$

where the scale-height is modelled as

$$h_z(R, R_0, \tau) = a_z \sqrt{\frac{2\bar{J}_z(R, R_0, \tau)}{\nu(R)}}. \quad (16)$$

We fit for the dimensionless parameter a_z and adopt the terms in the square root from the literature: $\bar{J}_z(R, R_0, \tau)$ is the mean vertical action of stars of age τ born at R_0 and now at Galactocentric radius R . Ting & Rix (2018) have studied the vertical heating history of the Galactic disk using an APOGEE red clump data set and Gaia data (Gaia Collaboration et al. 2018; Lindgren et al. 2018), and provided a functional form for $\bar{J}_z(R, R_0, \tau)$, which we adopt here. We calculate the vertical frequency ν such that $\nu^2 = \frac{\partial^2 \Phi}{\partial z^2}$ using the MWPotential2014 potential of the Galpy python package (Bovy 2014), which is the same gravitational potential Φ used by Ting & Rix (2018) to infer the orbital actions of their data.

3.2.4. Chemical Evolution

We model the chemical evolution of the gas as in Frankel et al. (2018): the metallicity of a star at birth is modelled as a simple function of its birth Galactocentric radius R_0 and time after birth τ :

$$[\text{Fe}/\text{H}] = F_m - (F_m + \nabla[\text{Fe}/\text{H}]R_{[\text{Fe}/\text{H}]=0}^{\text{now}})f(\tau) + \nabla[\text{Fe}/\text{H}]R \quad (17)$$

with the time dependency

$$f(\tau) = \left(1 - \frac{\tau}{12 \text{ Gyr}}\right)^{\gamma_{[\text{Fe}/\text{H}]}}. \quad (18)$$

We fit for the model parameters $\{\nabla[\text{Fe}/\text{H}], R_{[\text{Fe}/\text{H}]=0}^{\text{now}}, \gamma_{[\text{Fe}/\text{H}]}\}$ and keep $F_m = -1$ dex fixed. This chemical evolution model makes the following assumptions:

- there is a tight, unique relation between the birth location and time and the metallicity of a star,
- there is always a negative radial metallicity gradient in the ISM (modelled through $\nabla[\text{Fe}/\text{H}]$)
- at any Galactocentric radius, the $[\text{Fe}/\text{H}]$ of the ISM only increases with time (modelled through the exponent $\gamma_{[\text{Fe}/\text{H}]}$).

These assumptions are supported by several chemical evolution models for the late evolution of the Milky Way disk (the past 8 Gyr) after the last major merger (e.g. Schönrich & Binney 2009b; Grisoni et al. 2018). Since this model is purely parametric and fitted to the data, it has the important advantage of bypassing the large uncertainties currently present in chemical evolution models, for example the coupling between the star formation history, possible gas inflows, outflows, radial flows, enrichment sources, supernovae progenitors, supernovae models and nucleosynthesis yields. However, the shortcoming of this model is that it does not contain a self-consistent link between the chemical evolution description and the star formation history. Our approach comes closer to ‘weak chemical tagging’, where $[\text{Fe}/\text{H}](R_0, \tau)$ is used to tag stars to their possible birth radius (e.g. Minchev et al. 2018; Schönrich & Binney 2009b; Sanders & Binney 2015).

3.2.5. Accounting for an “Old” Disk Component

The evolution model presented above may not be valid at early times in the evolution of the Milky Way; yet, the size-able age uncertainties mean that we have to incorporate the existence of an “old” (still low- α) disk component. We aim to introduce an uninformative model for this old low- α disk by presuming there is an old star

fraction ϵ in our data set with an uniform age distribution between τ_m and 12 Gyr, where τ_m remains part of the formal model fit. After some experimentation, we have adopted empirically $\epsilon = 0.05$, as this leads to astrophysically sensible τ_m .

We model the old star metallicity distribution as a Gaussian centered on solar metallicity with 0.2 dex spread (inspired from Frankel et al. (2018)), the radial distribution as an exponential of scale-length R_{old} for which we fit, and the vertical distribution as a *sech2* function of scale-height $h_{\text{old}} = 0.85$ kpc (which is roughly consistent with the local scale-height of old stars (Mackereth et al. 2017)). We do not fit for the vertical scale-height of the older stars, but instead use prior knowledge that old stars have generally dynamically hotter orbits (so greater scale-heights) than younger stars. This makes our separation between ‘young’ and ‘old’ better informed in the presence of large age uncertainties at large ages.

3.3. Model for Age Uncertainties

To account for age uncertainties, we convolve the evolution model with a noise model:

$$p(\tau_{\text{obs}}, \mathcal{D}' | \mathbf{p}_m) = \int_0^{\tau_m} p(\tau, \mathcal{D}' | \mathbf{p}_m) p_{\text{obs}}(\tau_{\text{obs}} | \tau, \sigma_{\log_{10} \tau}) d\tau \quad (19)$$

where here \mathcal{D}' is all assumed noise-free observables (position and metallicity, see Table 1), and $\sigma_{\log_{10} \tau}$ is our noise parameter. We assume that age uncertainties are Gaussian in log age with a spread of $\sigma_{\log_{10} \tau} = 0.15$ dex, which we calculated by measuring the standard deviation of the difference in inferred log age and APOKASC2 test set. $p_{\text{obs}}(\log_{10}(\tau_{\text{obs}}) | \tau, \sigma_{\log_{10} \tau})$ represents the probability of measuring an age τ_{obs} given the true age τ :

$$p_{\text{obs}}(\log_{10}(\tau_{\text{obs}}) | \tau, \sigma_{\log_{10} \tau}) \sim \mathcal{N}(\log_{10}(\tau), \sigma_{\log_{10} \tau}). \quad (20)$$

The integral in Eq. 19 must be computed numerically for each of the 5381 stars, which makes the fitting procedure computationally expensive.

3.4. Normalization of the Probability Density Function: Survey Volume

The probability density function must be normalized over the observables. The survey volume, as defined in Equation 2, is

$$V_S(\mathbf{p}_m) = \int_{\mathcal{D}''} p(\mathcal{D}'' | \mathbf{p}_m) S(l, b, D) d\mathcal{D}'' = \sum_i^{\text{fields}} \iiint p(\tau, R_0, X, Y, Z | \mathbf{p}_m) S_i(D) \Omega_i D^2 dD dR_0 d\tau, \quad (21)$$

with $\mathcal{D}'' = \{l, b, D, R_0, \tau\}$ and $\Omega_i = \iint_{\text{field } i} \cos(b) dl db$ the solid angle of plate i . In the second line of Eq. 21, we sum over fields instead of integrating over the entire sky (S is zero outside the fields), and we have performed the integral over (l, b) in each field assuming the density varies slowly across the (small) angular size of the field. We have also implicitly integrated over metallicity as $\int p([\text{Fe}/\text{H}] | R_0, \tau) d[\text{Fe}/\text{H}] = 1$ and the selection function is assumed independent of metallicity. We compute the sum of the remaining 3D integrals using trapezoidal integration on a regular grid of $[D, R_0, \tau]$ with 42, 38 and 36 points in each dimension respectively. $V_S(\mathbf{p}_m)$ is a function of the model parameters only, so is only evaluated once per optimization step and not for each star.

3.5. Constructing the Likelihood

We can now cast the model aspects back into Eq. 7, and build Eq. ?? from Equations 3 and 6. In practice, our model is a mixture of the evolution model and the old, uninformative model:

$$p_{\text{tot}}(\mathcal{D}_i | \mathbf{p}_m, \text{selection}) = (1 - \epsilon) \cdot p(\mathcal{D}_i | \mathbf{p}_m, \text{selection}) + \epsilon \cdot p_{\text{old}}(\mathcal{D}_i | \mathbf{p}_m, \text{selection}), \quad (22)$$

with an ‘old fraction’ $\epsilon = 5\%$. Assuming all measurements are independent, we construct the log likelihood \mathcal{L} that the model parameters \mathbf{p}_m generated the data \mathcal{D} from our model

$$\mathcal{L}(\mathbf{p}_m, \{\mathcal{D}\}) = \sum_i \ln p_{\text{tot}}(\mathcal{D}_i | \mathbf{p}_m, \text{selection}). \quad (23)$$

To optimize calculations, we compute the survey volume once per optimization step as it is not a function of the observables. We perform a maximum likelihood estimate of the parameters \mathbf{p}_m of the global model for all the 5381 stars. We are mainly interested in three parameters: the star formation time-scale, the inside-out growth parameter x and the scale-length of stars at birth R_d . However, we fit for all ten parameters in the model (summarized in Table 5) because the global model needs a good description of the Milky Way disk in all aspects in order to describe inside-out growth correctly.

4. RESULTS

The parameter values that maximize the likelihood are presented in Table 5, where the uncertainties quoted are determined from 10 samples of bootstrapped data. We comment below the direct implications and describe the main tests carried to verify the results.

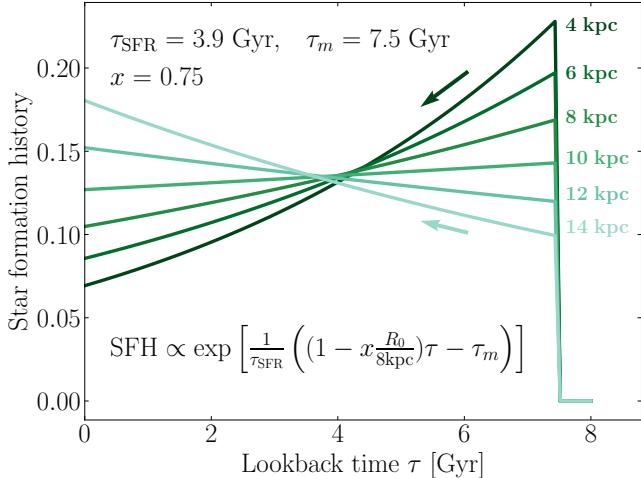


Figure 5. Star formation history $\text{SFH}(R_0 | \tau)$ implied by the best-fit model at different Galactocentric birth radii R_0 (from dark to light: 4, 6, 8, 10, 12 and 14 kpc). The best fit model has parameters $x = 0.75$, $\tau_m = 7.5$ Gyr and $\tau_{\text{SFR}} = 3.9$ Gyr. This yields a star formation rate decreasing with time in the inner disk (dark green arrow), flat at 10.5 kpc, and increasing with time in the outer disk (light green arrow).

Table 2. Maximum Likelihood Estimates of \mathbf{p}_m

Parameter	\mathbf{p}_m	Best fit
Inside-out growth	x	0.75 ± 0.17 (0.69) ^a
Star formation time-scale	$\frac{\tau_{\text{SFR}}}{\text{Gyr}}$	3.9 ± 1.5 (1.4)
Star formation onset	$\frac{\tau_m}{\text{Gyr}}$	7.5 ± 0.5 (7.6)
Disk scale-length	$\frac{R_d}{\text{kpc}}$	3.1 ± 0.4 (2.9)
Radial orbit migration	$\frac{\sigma_{\text{RM7}}}{\text{kpc}}$	3.9 ± 0.2 (3.7)
Radius of solar [Fe/H]	$\frac{R_{[\text{Fe}/\text{H}]=0}^{\text{low}}}{\text{kpc}}$	8.3 ± 0.3 (8.9)
Enrichment exponent	$\gamma_{[\text{Fe}/\text{H}]}$	0.19 ± 0.03 (0.212)
Metallicity gradient	$\frac{\nabla[\text{Fe}/\text{H}]}{\text{dex kpc}^{-1}}$	-0.073 ± 0.002 (-0.078)
Old disk scale-length	$\frac{R_{\text{old}}}{\text{kpc}}$	1.3 ± 0.9 (2.2)
Vertical adaptive scale	a_z	0.85 ± 0.03 (0.79)

^aNumbers in parentheses are fits using the Bayestar19 extinction map in 240 fields and 7600 stars as a check described in subsection 4.3.2.

4.1. Inside-out Growth

We quantify the radial dependence of the star formation history with $x = 0.75$ for the Milky Way’s low- α disk. The resulting star formation history is illustrated in Figure 5. According to this fit, the star formation processes that generated our data set started about $\tau_m = 7.5$ Gyr ago. This maximum age differs from the measured age of the oldest star in the data set because (1) a fraction of old stars constrain the old component of the model more than the young one, and (2) this pa-

rameter is in units of the dummy variable ‘true age’ of the model, which is convolved over age uncertainties. Therefore, τ_m is different from a best fit that would not account for age uncertainties, where it would take the value of the oldest star of the ‘young sample’ (where the ‘young sample’ would be defined as the 95% youngest stars of the APOGEE sample because of our split in young and old components, see Section 3.2.5).

The best fit value of x implies that the star formation rate has been constant at ~ 10.5 kpc, a decreasing function of time in the inner ~ 10.5 kpc, and an increasing function if time in the outer ~ 10.5 kpc. This means that the Milky Way disk is still forming stars, with a slower decay in the outer disk than in the inner disk. The star formation history decayed slowly in the Solar neighbourhood.

4.2. Other Parameters

Even if the present work focuses on the radius-dependent star formation history of the Milky Way disk, the processes determining the evolution of the Milky Way are multiple and complex. We therefore constructed a global model to account for the effect of several of them, in particular radial orbit migration and chemical evolution.

We find a radial orbit migration strength of about $\sigma_{\text{RM}} = 3.9 \text{ kpc} \sqrt{\tau/7 \text{ Gyr}}$. This value is greater than, but consistent with, the amount of radial migration in Frankel et al. (2018) who found $\sigma_{\text{RM}} \approx 3.4 \text{ kpc} \sqrt{\tau/7 \text{ Gyr}}$. The difference results from the different age sets used between that work and the present study. The current age catalog is, on average, younger. So the stars have had less apparent time to migrate the same distance, which pushes the radial orbit migration strength up accordingly. The effects of different age catalogs is discussed more extensively in Sections 4.3.4 and 6.

4.3. Tests

We test the robustness of the results to various numerical approximations and data uncertainties, and discuss the most important aspects below.

4.3.1. Basic Tests

We perform a series of simple tests to ensure that (1) model prediction in data space compare well to our data set, (2) a density model recovers similar work in the literature, (3) integrals are computed with the necessary level of accuracy, (4) the optimization scheme recovers parameters correctly.

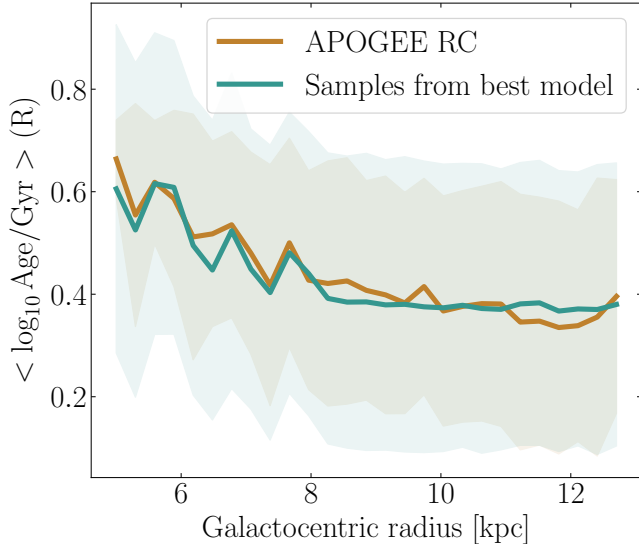


Figure 6. Radial profile of the stellar age distribution in the Galactic disk. The observed data are shown brown, and samples from the best fit model, after applying the APOGEE selection function, are shown in green; the thick lines reflect the mean age and the filled areas are the age dispersion. The jaggedness of the median age is a consequence of the spatially complex selection function, not an indication of spatially discontinuous star formation. This is why the model (green), combined with the APOGEE selection function, is able to reproduce these features even with a smooth star formation history. This Figure illustrates that the observed, present-day age distribution as a function of radius, can be modelled well by our model.

1. To verify that the fitting procedure has worked, we produce a mock data set from the best fit parameters using the APOGEE selection function and compare it to the real data. The mass-weighted age profile is well recovered, but the predicted age scatter is greater than observed (see Fig. 6 and Section 6.3). Overall, the best fit evolution model reproduces well the observed scatter in age-metallicity at given Galactocentric radius, as shown in Fig. 7. The mean metallicity profile is well reproduced, indicating that our parametric prescription for the gradual enrichment of the gas was well fitted. In Figure 7, there is an apparent flattening of the metallicity profile at $R < 7$ kpc, both in the model and in the data. In the present work, these are purely due to selection effects, since our chemical evolution model has a constant metallicity gradient in space and in time. The inner disk APOGEE fields tend to point to high Galactic latitudes away from the mid-plane (Fig. 3). Since there are vertical metallicity gradients rising from vertical heating of older, metal-

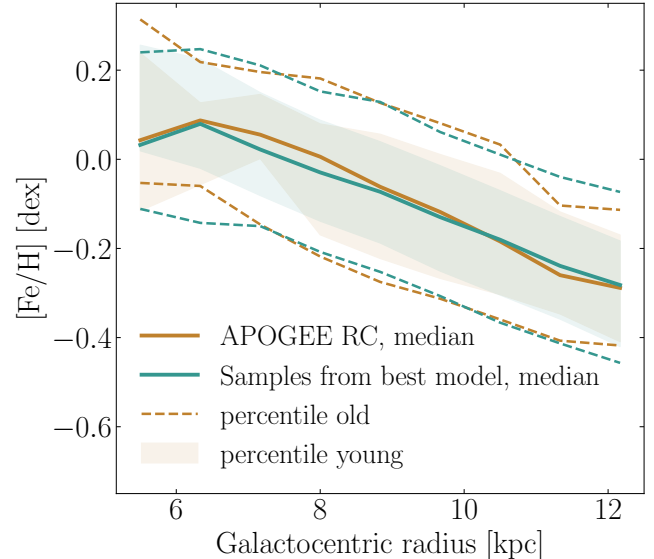


Figure 7. Radial metallicity profile of the observed data (brown) and samples from the best fit model (green), analogous to Fig. 6. The solid lines are the median metallicity profiles using all stars, and using a best fit metallicity gradient of $-0.073 \text{ dex kpc}^{-1}$ (see Table 5). The 16% and 84% percentiles of two age bins are also plotted: the shaded area for stars younger than 1.5 Gyr, and the dashed lines for stars with ages $2.5 < \tau_{\text{obs}} < 5$ Gyr. The metallicity scatter increases with age due to radial orbit migration, and seems well reproduced by the model. The slight differences in the inner disk are expected because there are few data points at radii less than 7 kpc. Our model reproduces well both the metallicity gradient and the age-dependent metallicity scatter.

poorer stars, the data set lacks the metal rich stars from the inner mid-plane, underrepresented due to the spatial selection. Of course, this is not an issue in our case because these effects are fully accounted for in our model. But it highlights the importance of accounting carefully for selection effects when using data from surveys to draw conclusions on the evolution of the Galaxy.

2. We set out a model for the 3D density of the Galactic disk $p(X, Y, Z | [\text{Fe}/\text{H}], \tau)$, as in Bovy et al. (2016); Mackereth et al. (2017). Qualitatively, fitting for the scale-lengths and heights in age-metallicity bins gave consistent results with Mackereth et al. (2017), where at given age, the distribution of metal-rich stars peak in the inner disk and those of metal-poor stars peak in the outer disk. And at given metallicity, the distributions of old stars are in broader ‘donut-like’ structures than those of young stars.

3. We verify that the normalizing integrals of our global model were evaluated with enough accuracy. Varying the ranges and regularity of the integration grids, we find that increasing the accuracy further does not influence the best estimate results presented here.
4. To ensure that the optimization scheme (Nelder-Mead algorithm) does not get stuck in local maxima, we optimize the likelihood several times, with different initial parameters. Additionally, we check that the optimization scheme was able to recover the true parameters. We sample 5400 mock data points with various parameter values (as well as the best fit) and add noise to their ages from our noise model, Eq. 20. We fit these data, and recover the parameters with good precision. The enrichment parameters used to tag birth radii, and the radial orbit migration strength are recovered to 4%, τ_m is recovered to 0.5%, x to 7% and τ_{SFR} to 15% (depending how much noise is added to the ages, here for $\sigma_{\log_{10} \tau} = 0.15$ dex). The scale-length of the disk is recovered to 3% and a_z to 0.9%.

4.3.2. Modelling Extinction

We verify the consistency of the effective selection function of this work (Eq. 6), using the Bayestar17 extinction map (Green et al. 2018), with the more recent extinction map Bayestar19 (Green et al. 2019). Since we have excluded all the APOGEE fields where the median extinction of the APOGEE sample is greater than $A_H = 0.6$, the Bayestar17 map is largely sufficient for our description of the disk in the 142 remaining disk fields and yields the same effective selection function as the Bayestar19 map. However, the effective selection function differs significantly in the Galactic mid-plane where there are more extinguished fields, which we have excluded. Assuming we can trust the Bayestar19 dust map for such fields, we have fitted the parameters again using the 7600 stars present in the 213 disk fields including those high extinction fields. The best fit parameters are listed in parentheses in Table 5. We find a similar estimate of the disk scale-length R_0 , which is the parameter that would have been most affected by variations in the effective selection function (direct density modelling). The star formation time-scale τ_{SFR} changes significantly after including more stars from the inner disk (due to co-variances with the scale-length of the old disk component and the fact that the data set has changed but not our outlier fraction ϵ), but the inside-out growth parameter x remains strongly positive.

4.3.3. Modelling Age Uncertainties

Stellar ages are challenging to determine. They are modelled quantities rather than direct observables, so using stellar ages relies on assumptions in the underlying stellar evolution models. Ages are heteroscedastic, their uncertainties not well quantified, and their systematic behaviors unknown. In the present work, we assume that asteroseismic ages are the ground truth, and that age errors arise when mapping from stellar spectra to asteroseismic ages through data-driven methods. For the data set we have used, these errors are approximately constant in log age, and the standard deviation of log age on a test set (APOKASC2) is roughly 0.15 dex in decimal logarithm, (Ting & Rix 2018). We used a noise model based on this value, where the measured log age is normally distributed around the true value with a 0.15 dex spread (Eq. 20). This assumes that the neural network used to map stellar spectra to log age produces Gaussian errors of 0.15 dex.

To benefit from the direct use of stellar ages, we need to test and understand the impact of the (unknown) uncertainties on our conclusions. We test the robustness of the results to the assumed noise model by varying the value of this scatter between 0.12 and 0.17 dex, and optimizing the likelihood with these new values. We find noticeable changes in the total age of the low- α disk (parameter τ_m) and a (weak) dependency on the inside-out growth parameter x that takes values between 0.6 (underestimate errors) and 0.82 (overestimate errors). This leaves our conclusions on inside-out growth unchanged (with a strongly positive x), but it affects the comparisons between model predictions and the data set for old stars. So the best fit parameter τ_m should not be seen as the age of the Milky Way disk, but rather as the maximum age at which we trust our model - based on the age scale and assumed uncertainties of our data.

4.3.4. Different Age Determination Methods

In the era of large surveys and data-driven methods, there are at present multiple catalogs of parameters derived for the same stars in large homogeneous data sets. The values inferred from different methods have systematic differences, which are sometimes significant. It is central to realize how these systematics influence our understanding of Galaxy evolution. We test a total of five age catalogs (described in details in Appendix A), on the exact stars that we presented in Section 2 (with some minor loss during the cross-match), keeping metallicity and position the same and changing only the age data column in the likelihood optimization. As expected, the best fit values changed between age sets, see Table 3 where the best fit parameter x is summarized, as well

Age set	x	R_{const}
Ting18	0.75	11 kpc
Ness16	0.62	12 kpc
Ness18	0.50	16 kpc
Ness19	1.1	8 kpc
Sanders18	-0.3	–

Table 3. Inside-out results using different ages

as the Galactocentric radius at which the star formation rate is a constant in time R_{const} . The first data (which we will refer to as ‘Ting18’) set was used for the analysis described above. The ages of the other sets were derived either using *The Cannon* (Ness et al. 2015, 2016, 2019) or a combination of data-driven mapping from abundances to ages and stellar evolution models (Das & Sanders 2019; Sanders & Das 2018). and are named from these references (see Table 3).

These five data sets, composed of the same stars but with different age estimates, lead to five different best-fit values for inside-out growth x due to systematic differences between methods. These differences in inside-out growth best fit arise naturally from the different radial age gradients present in each data set: the data set containing the strongest age radial gradient (Ness19) leads to the strongest inside-out growth best value ($x = 1.1$), and the data that show the weakest age gradient (or the least negative, Sanders18) also has the least positive inside-out growth ($x = -0.3$). Apart from the Sanders18 age set, all data are consistent with an inside-out growth formation scenario for the Milky Way disk, with some slight difference on the decay of star formation with Galactocentric radius. We found that the Sanders18 age set might over estimate stellar ages at large distance: this is the only method making direct use of Gaia parallaxes, and comparing these ages to the four other spectroscopic ages (with no known bias with distance) shows systematic differences as a function of distance. The systematic zeropoint issues with the Gaia parallaxes are now well studied (Leung & Bovy 2019) which bias the ages via the the strong mass-luminosity degeneracy for giant stars. Folding this distance bias with the APOGEE spatial selection function can produce an inversion of the overall age gradient in the data set.

4.3.5. Comparisons with Literature Results

The test presented above consists in comparing inference results from the same stars, of the same population (red clump), but using ages derived from different methods. In complement, we now use the best fit parameters from these five age sets to make predictions that can be compared with literature. We compare the model

predictions to the age distributions from Xiang et al. (2018), who measured ages from turn off and sub-giant LAMOST stars (Cui et al. 2012) with a 0.12 dex precision on log age. They corrected for selection effects, and derived selection-corrected age histograms in Galactocentric radius bins. To compare, we simulate this procedure. We sample stars of all populations (not just red clump) from our best fit evolutionary model and select them in Galactocentric radius bins as in Xiang et al. (2018). We then add 0.12 dex of noise to the ages, to emulate their age histograms. We find that the best fit parameters obtained from the Ting18 sample (the ages we have focused on during the analysis) are more consistent with the trends in Xiang et al. (2018) than those from the other age sets, at young ages (where it matters and where the data are constraining). The model predictions and those of Xiang et al. (2018) differ significantly at large ages: we under predict the number of old stars. This could come from that our model is mostly constrained by young stars, and that we only model the low- α disk, whereas Xiang et al. (2018) derived these age histograms considering all stars.

4.3.6. Inside-out Model Variants

We test several models for the distribution of $\Sigma(R_0, \tau)$ in order to (1) see whether the general result ‘the Milky Way disk grew from inside-out’ is robust to model variations and to our definitions, and (2) set out to understand what the best description for inside-out growth is. We describe two of them below.

(1) *Time-varying Disk Scale-length*—We first build a model of a birth surface density profile with a time varying scale-length, where

$$\Sigma(R_0, \tau) = \text{SFH}(\tau)\Sigma(R_0 | \tau), \quad (24)$$

inspired by that in Frankel et al. (2018). The star formation history is simply a universally-decreasing exponential, where stars started forming 12 Gyr ago on a time-scale τ_{SFR} , and $\Sigma(R_0 | \tau)$ is an exponential profile of the disk with a varying scale-length $R_d(\tau) = R_{d,0}(1 - \alpha_{\text{Rexp}} \frac{\tau}{8 \text{ Gyr}})$. The fitted parameter quantifying inside-out growth is α_{Rexp} . The fit results in overly strong inside-out growth, α_{Rexp} going to 1 (which confirms and amplifies the inside-out growth results found here). But this model is conceptually deceiving: integrated over time, it predicts an overall disk with a profile that deviates significantly from the expected exponential for disk galaxies. Additionally, this model is restrictive: by construction, it is (1) unable to describe outside-in growth in a physically plausible way, (2) unable to describe very strong inside-out growth, which would result in stars born with a negative scale-length.

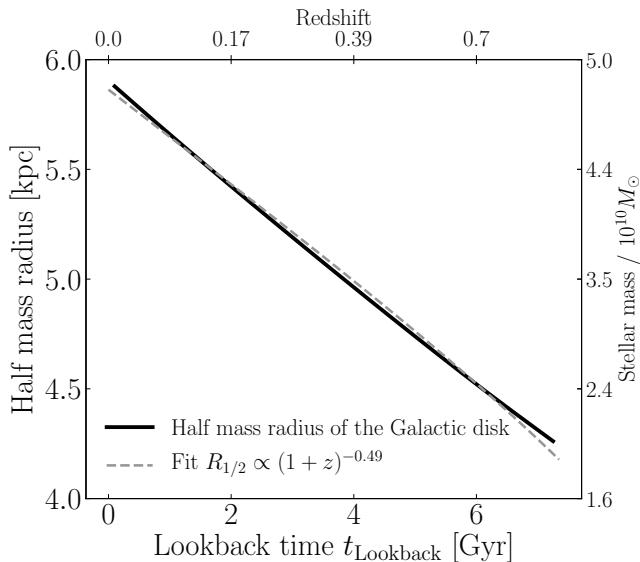


Figure 8. Size evolution of the Milky Way’s low- α stellar disk over the past 7 Gyr, according to the best fit model (as computed in Eq. 26). The disk has grown by 43% its size 7.5 Gyr ago. The dashed grey line is an approximation of this evolution by $R_{1/2} \propto (1+z)^\alpha$ with $\alpha = -0.49$. The stellar mass as a function of lookback time (as computed in Eq. 27) is illustrated on the right hand side y axis.

(2) *Radius-varying Star Formation Peak*—We test another model, where inside-out growth is not described by a radially-varying star formation time-scale as in equation 11, but rather with a radially-dependent star formation peak. This raises the question of what inside-out growth is: do stars form on a longer time-scale at large radii, or does star formation begin later at large radii? Unfortunately, the second question, ‘when do stars begin to form?’ requires data that are constraining at large ages, which are currently not available due to the red clump selection. The best fit parameters are not in contradiction with inside-out growth and samples from that model reproduce the observed data better than the model we chose to show in this analysis. But the fitting procedure is unfortunately not more conclusive in a quantifiable way. This is because age errors at large ages are so large, and old data are so sparse, that our likelihood function is almost flat in the parameter quantifying the radial dependency in the star formation peak.

5. ASTROPHYSICAL IMPLICATIONS

We have constructed and constrained a model that describes how the Milky Way’s low- α disk may have built up and grown over time. We discuss below the implications in the more general context of the formation and evolution of galactic disks, and show that under the as-

sumptions of this work, the Milky Way is typical in this aspect.

For external galaxies, there exist two families of observational constraints for inside-out growth. Redshift-size relations are obtained by measuring the size of the visible disks (half-light radius, or effective radius), assuming exponential or de Vaucouleur (de Vaucouleurs 1948) profiles. These observations reflect the overall size of galaxies of fixed stellar mass (derived from their observed light) at given lookback times, and show that on average, disks were smaller in the past. Since these observations are made on populations of galaxies of a given stellar mass at all redshifts, they do not reflect the time evolution of individual galaxies (especially as, due to star formation, the stellar mass of a galaxy should in principle increase with time). In this picture, a star-forming disk galaxy’s stellar mass and size should grow with time. Interpreting the evolution of scaling relations with time requires modelling (e.g., Mo et al. 1998).

On the other hand, galactic archaeology approaches have studied the present-day positions of different populations of stars in individual galaxies, including the Milky Way. But stellar ages have large uncertainties, and present-day stars Galactocentric radii are not their birth radii. If stellar ages indicated the lookback time at which stars formed, and if present-day stellar positions reflected well their birth positions, this approach would provide a direct knowledge of the formation of galactic disks. But under extensive radial orbit migration and diffusion, information on stellar birth sites, and therefore local star formation histories, is lost.

Measured mass weighted age gradients are small. For instance, Goddard et al. (2017) found mass weighted age gradient on average consistent with zero. This is in apparent contradiction with the picture provided by redshift-size relations, implying that old stars should be more centrally concentrated if stars were not radially redistributed in some way. If these differences are physical (and not due to measurement systematics), two scenarios can reconcile these observations: (a) either star formation is constant across disks, and disks grow in size due to redistribution of stars or mergers for non isolated galaxies, or (b) star formation moves outwards with time but radial orbit migration erases the traces of it, such that present-day studies only see the weak remnants of formation gradients. Our present model for the Milky Way disk allowed both pictures to exist through the parameter x , that would be zero in (a) and positive in (b). The data preferentially constrained it towards (b) since the best fit $x = 0.75$.

Below, we use the best fit model to predict (1) the evolution of the half-mass radius of the Milky Way disk

with time, (2) the evolution of its surface brightness profile and its half-light radius, to compare to redshift-size relations, and (3) the present-day positions of stars of different ages.

5.1. Evolution of the Disk Half-mass Radius

We use the best fit model to predict the size of the Milky Way disk at different times in the past. Accounting for inside-out growth and radial redistribution of stars, we compute the radius containing half of the total disk mass,

$$M(< R_{1/2}) = M_{tot}/2, \quad (25)$$

as a function of lookback time $R_{1/2}(t_{\text{lback}})$ by solving numerically for the radius $R_{1/2}(t_{\text{lback}})$ enclosing half of the total disk stellar mass M_{tot} .

The stellar mass $M(< R_{1/2})$ contained within $R_{1/2}$ is defined as

$$M(< R_{1/2}, t_{\text{lback}}) = M_{tot} \int_0^{R_{1/2}} \int_0^\infty \int_{t_{\text{lback}}}^{\tau_m} p(R_0) \times p(\tau | R_0) p(R | R_0, \tau - t_{\text{lback}}) d\tau dR_0 dR, \quad (26)$$

where t_{lback} is lookback time, τ is time from now (dummy lookback time in the integral, marking the birth of stars) and the quantity $\tau - t_{\text{lback}}$ is the age of stars at lookback time t_{lback} and is used to compute the radial migration term $p(R | R_0, \tau - t_{\text{lback}})$. The evolution of the half-mass radius for the best fit model is illustrated by the black solid line in Figure 8. The size of the disk has evolved almost linearly in time from about 4.2 kpc 7.5 Gyr ago to about 5.9 kpc today. This implies an almost constant growth rate of $\dot{R}_{1/2} = 0.2$ kpc Gyr⁻¹. The associated time-scale of radial disk growth is $\tau_R = R_{1/2}/\dot{R}_{1/2} = 30$ Gyr. This compares well with the findings of Pezzulli et al. (2015), who measured $\dot{R}_{1/2}$ using star formation and $R_{1/2}$ assuming an exponential surface density profile of about 30 disk galaxies and quote the same growth time-scale. There may be differences by factors of a few due to systematic uncertainties or the definition of a disk size used (half-mass radius, half-light radius, or exponential scale-length).

Using $\Omega_\Lambda = 0.7$, $\Omega_m = 0.3$, $H_0 = 70$ km s⁻¹ Mpc⁻¹, the half-mass radius obtained from Eq. 26 is well approximated by a function of redshift z as $R_{1/2} \propto (1+z)^\alpha$, with $\alpha = -0.49$. This seems globally consistent with redshift-size measurements of external disk galaxies. Franx et al. (2008) find an average relation for galaxies of stellar mass $M_\star > 2.5 \times 10^{10} M_\odot$ with $\alpha = -0.6 \pm 0.1$ and Trujillo et al. (2006) find $\alpha = -0.40 \pm 0.06$ for late type galaxies of stellar mass $M_\star > 3 \times 10^{10} M_\odot$. This

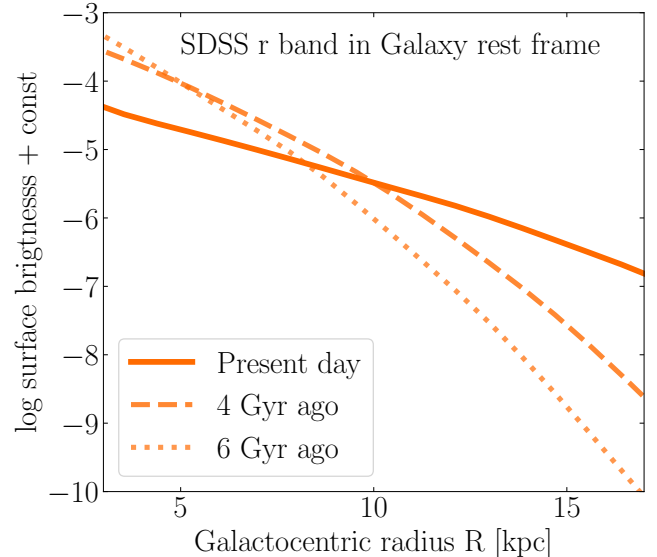


Figure 9. Predicted SDSS r band surface brightness profile for the Galactic disk at three different lookback times. This illustrates the inside-out growth of the stellar disk implied by our model fit.

approximation of the half-mass radius is illustrated by the grey dashed line in Figure 8. However, this direct comparison is only approximate: the literature relations were fitted for galaxy populations of given stellar mass, whereas this work predicts the evolution of the half-mass radius of the Milky Way at different times of its evolution, including its growth in stellar mass. We propose to compensate for these differences in the next subsection.

5.2. Evolution of the Half-light Radius: Comparing to Redshift-Size Relations

In external galaxies, the physical half-mass radius is not a direct observable. It is often assumed that stellar light at different wavelengths traces stellar mass of different populations, and the half-light radius is used as a proxy for the disk size. We predict the surface brightness profile and the half-light radius that would be measured at different lookback times and in different photometric bands, roughly emulating observations of the Milky Way at different redshifts. For this purpose, we use grids based on the single stellar population synthesis code E-MILES SEDs (Vazdekis et al. 2010, 2012; Ricciardelli et al. 2012) to predict the mass-to-light ratio of a single stellar population of given age and metallicity. Assuming a Kroupa initial mass function (Kroupa 2001), we use the best fit model to predict the distributions of $p(\tau, [\text{Fe}/\text{H}], R | \mathbf{p}_m, t_{\text{lback}})$ at given lookback time to estimate the stellar mass density at radius R (up to a normalization constant). The resulting surface brightness profiles in the SDSS-r band are plotted

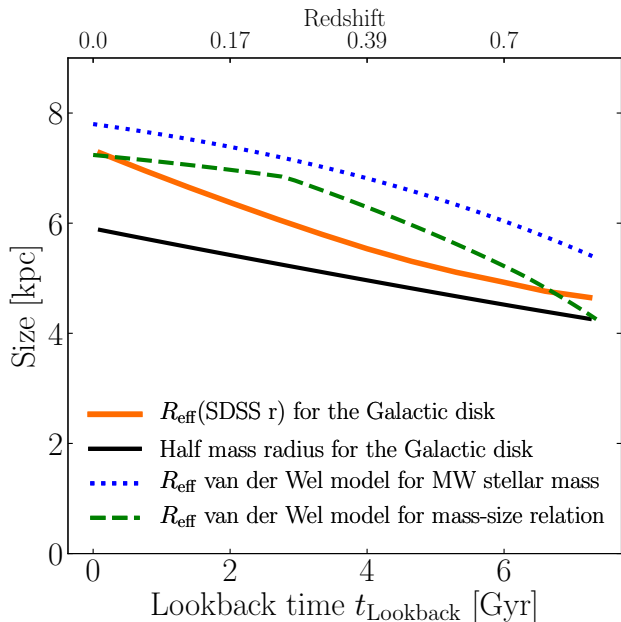


Figure 10. Predicted half-light radius as a function of lookback time for the Milky Way disk (thick orange), compared with that of other galaxies (green and blue). The blue dotted line is a prediction from the van der Wel et al. (2014) model for Milky Way mass redshift-size relations, as an element of comparison. It is different from the evolutionary track of a single galaxy, because galaxies grow in stellar mass and size (whereas the van der Wel et al. (2014) blue line is at fixed stellar mass). The green dashed line shows the the same prediction of the van der Wel et al. (2014), but including the evolution of the stellar mass of the Galaxy as in Eq. 27. It leads to a redshift-size evolution close to our inference for the Milky Way.

We include the half-mass radius (from Fig. 8) in black as a reference.

as an example in Figure 9. Qualitatively, the trends are similar in other bands. Generally, blue bands are spatially more extended and red more centrally concentrated. This is expected because shorter wavelengths are better tracers of young populations.

From these surface brightness profiles, we solve a similar equation to Eq. 26 for the half-light radius, illustrated by the thick orange line in Fig. 10. The half-light radius is greater than the half-mass radius, in particular at late times, when the young stars dominating the light are spatially more extended than the overall stellar population. This compares well with the model of van der Wel et al. (2014), who fitted a model to galaxies at different redshifts (illustrated in dotted blue in Fig. 10 for galaxies of stellar mass $M_{\star} \approx 5 \times 10^{10} M_{\odot}$). However, this only serves as an element of qualitative comparison. This model fitted the size of galaxies of \sim Milky Way stellar mass at different redshifts. But the stellar mass

of the Milky Way has grown with time. Therefore, the comparison holds best at $z = 0$. At higher redshifts, one should compare to the sizes of galaxies of smaller stellar mass, which should also be smaller than the blue dotted line that Fig. 10 predicts.

van der Wel et al. (2014) have also fitted the evolution of the total stellar mass M_{\star} - size R_{eff} relation as a function of redshift z . With our best fit model of the evolution of the stellar mass of the Milky Way, we can use the van der Wel et al. (2014) fits to predict the corresponding size. Assuming that the total stellar mass of the Milky Way is the sum of the bulge mass M_b and that of the disk, and assuming that the bulge mass has been constant over the past 8 Gyr (because the bulge is relatively old compared to the disk (Bland-Hawthorn & Gerhard 2016)), we consider the total Milky Way stellar mass as

$$M_{\star}(t) = M_b + M_d(t). \quad (27)$$

We assume a bulge stellar mass of about $1.5 \times 10^{10} M_{\odot}$ (Bland-Hawthorn & Gerhard 2016) and the fraction of mass in the bulge about 30% at present. The evolution of the stellar mass is illustrated in Fig. 8.

Following van der Wel et al. (2014), we assume a stellar mass-size relation of the form

$$R_{\text{eff}}(z) = A(z) \left(\frac{M_{\star}(z)}{5 \times 10^{10} M_{\odot}} \right)^{\alpha(z)}, \quad (28)$$

where $A(z)$ and $\alpha(z)$ were measured by van der Wel et al. (2014) at different redshifts. We interpolate these values linearly as a function of redshift and compute the corresponding value of $R_{\text{eff}}(z)$ and our predicted Milky Way's stellar mass as a function of time. The resulting size as a function of lookback time (or redshift) is illustrated with the dashed green line in Figure 10, and compares well with our predicted evolution of the half-light radius of the Milky Way.

5.3. Present-day Scale-lengths of Stellar Populations

In nearby galaxies, it is common to determine the spatial distribution of stars of different populations (e.g., Gogarten et al. 2010). When old stellar populations are found more centrally concentrated than the young ones, this is interpreted as the result of inside-out growth.

We compute these quantities in two cases: (a) for our Milky Way best fit, and (b) for our best fit excluding radial migration (i.e. finding the half-mass birth radii). (b) does not lead to a scale-length that is possibly observable today if stars migrate, but decouples the contributions of the star formation history and radial orbit migration to the growth of the disk. To compute these, we perform a Monte Carlo simulation sampling ages, present-day Galactocentric radii and

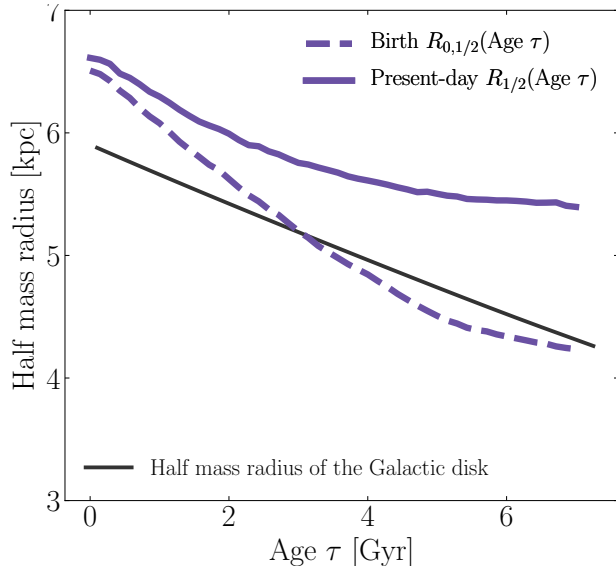


Figure 11. The present-day half-mass radii of stars with ages τ , $R_{1/2}(\tau)$, compared to the star formation scale-length a time τ ago. We show $R_{1/2}(\tau)$ from the best fit model with solid purple and compare it to the analogous, hypothetical radii expected in the absence of radial orbit migration (dashed purple). As a reference, we overplot the overall size evolution of the disk at different lookback times in solid black. Radial mixing leads to a moderate present-day dependence of $R_{1/2}$ on age, even in the presence of distinct inside-out growth in the star formation history.

Galactocentric radii at birth for stars from the best fit model. For each sample, we compute the present-day radius $R_{1/2}(\tau \approx t_{\text{lkback}})$ containing half of the mass of stars at given age, and similarly for the birth radius $R_{0,1/2}(\tau \approx t_{\text{lkback}})$. These two quantities are plotted in purple in Figure 11.

We find a significant difference in the half-mass radius of different stellar populations that have migrated (solid purple in Fig. 11) and those that have not (dashed purple). In the limits of our modelling, we can argue that if we were to observe the Milky Way present-day stellar populations in order to constrain inside-out growth, and assume its stars do not redistribute radially, we would overestimate the scale-length at birth of old stars by about 1 kpc, and therefore underestimate the amount of inside-out growth. This is a caveat for galactic archaeology approaches that cannot be neglected for disk galaxies with significant radial orbit migration. For the Milky Way, this can be overcome by detailed modelling, because one can measure a physical scatter in the age-metallicity distributions, and interpret it as an effect of radial orbit migration. However, for external galaxies, the question is more challenging as only integrated properties are derived (the mean, but not the scatter). Therefore, the present-day distributions of stellar popu-

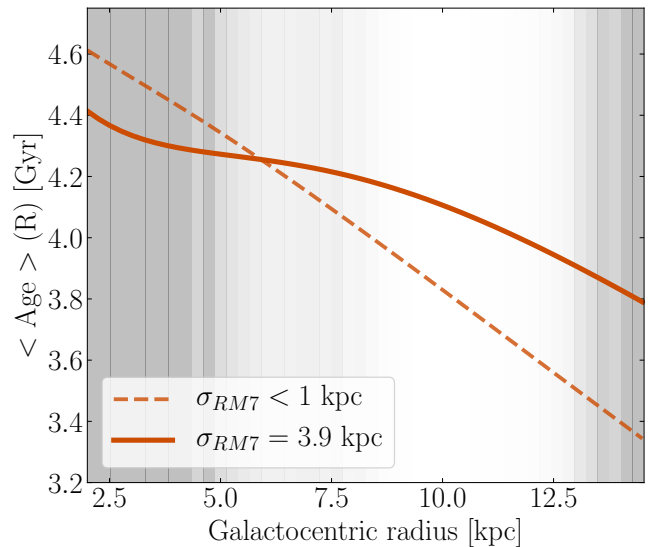


Figure 12. Radial profile of the mean stellar age in the Galactic disk, determined through our model fit (solid line, for $\sigma_{RM7} = 3.9$ kpc computed using Equation 29). This is compared to the hypothetical case of ‘no radial orbit migration’ (dashed line), where stars stay at their birth radii (also determined from our model fit). Age gradients resulting from the distinct inside-out growth that the best model implies are severely weakened by radial orbit migration. Areas containing few data to constrain our model at their present-day Galactocentric radii are shaded in proportions of the Galactocentric radius histogram in Fig. 2 on a logarithmic scale.

lations can be a weak diagnostic for inside-out growth, giving only lower limits.

5.4. Weakening of Age Radial Gradients

Age gradients are expected to result from inside-out growth: if stars formed on shorter time-scales in the inner disk than in the outer disk, then the proportions of old stars should be larger in the inner disk than in the outer disk. But in external galaxies, the mass weighted age gradients seen in MaNGA seem weak for a large number of disk galaxies (Goddard et al. 2017; Bundy et al. 2015).

We compute the expected mass weighted age profile from our model and the best fit parameters with

$$\langle \text{Age} \rangle (R) = \iint \tau \cdot p(R_0, \tau | R) dR_0 d\tau, \quad (29)$$

and illustrate it in Figure 12 for two cases: no radial migration (where we take $\sigma_{RM7} < 1$ kpc), and our best fit radial migration. We find that, in the ‘no migration’ case, the best fit star formation history produces an age gradient that is already weak, with only about $-0.1 \text{ Gyr kpc}^{-1}$. But in addition, radial orbit migration weakens severely this gradient. Therefore, it is pos-

sible for a galaxy to have grown from inside-out without exhibiting strong age radial gradients, making present-day age gradients a poor diagnostic of inside-out growth.

6. LIMITATIONS OF OUR METHODOLOGY

Even though the present method is promising for later studies of Galaxy evolution, we state below important shortcomings that should be addressed in the future.

6.1. *Inevitable Limitations: Direct Use of Ages*

The direct use of ages in global models for the Galactic archaeology field is new and can only be as powerful as the ages are good. Since our method relies *directly* on stellar age measurements, the results are directly affected by any systematics in age determination methods (see Section 4). These systematics are numerous and unknown; and age uncertainties are large and heteroscedastic. Age determination methods differ by their complexity. Data-driven methods depend strongly on the reliability of their training sets, those using luminosity and distances may induce large scale spatial biases, affecting our understanding of Galaxy evolution. To estimate the extent of these limitations, we have considered several data-driven age catalogs, and presented the results for the fit on the Ting18 sample (Ting & Rix 2018). Even though this data set has its shortcomings, such as an unexpectedly small fraction of old stars, it showed more reliability than the three others in the several aspects:

- the precision of the determined ages on a test set on APOKASC 2 red clump stars was the highest with 0.15 dex scatter in log age;
- model predictions using the best fit on these data compared better than those the three other data sets to independent work on selection-corrected age histograms using LAMOST turn off stars;
- the age determination method (spectroscopic ages) has no known dependencies on distance measurement, so there should not be age biases with distance, which is of fundamental importance to study inside-out growth through the spatial structure of ages;
- the training of the neural networks that determined these ages was done only on red clump stars, so the neural networks were not learning possibly undesired information from other stellar evolutionary stages.

Inevitably, the present results are bound to all assumptions that were made by using this data set and by modelling its age uncertainties.

6.2. *Limitations from the Data Selection*

This work made exclusive use of red clump stars, because they are excellent standard candles, they are bright and they are abundant in the Galactic disk. However, this stellar population is younger than the underlying mean population, with an age distribution that peaks at 2 Gyr and that has few stars older than 5 Gyr (see Fig. 1). This introduces two shortcomings. First, most of the information constraining our model lies between 1 - 4 Gyr. Therefore, our model describes best the most recent evolution of the Galactic disk, but not earlier times when the physical processes of disk formation were supposedly faster. Secondly, we corrected for population selection biases by modelling their age distribution (or, more accurately, the relative mass of stars on the red clump evolutionary stage, at given age and assuming an initial mass function). This was based on theoretical expectations of a star's lifetime on the core helium burning stage, which is only constrained by theoretical models of stellar evolution.

In order to constrain the evolution of the disk at earlier times, and reduce stellar population effects on ages, we should turn to a different stellar population with more old stars, for example RGB stars for which distances are known precisely (Hogg et al. 2018; Leung & Bovy 2019). But this would introduce more technical difficulties that are not straightforward to overcome: using standard candles limits the number of integrals to evaluate in magnitude-limited selection functions, bringing computational expenses down. However, the lack of old stars in our red clump data set, and the large age uncertainties at large ages, made us unable to build and constrain a more adequate model for the early behavior of star formation. This question is left to be addressed in the future.

Additionally, due to the spatial selection of stars in our survey, most of the stars we have used lie in the outer disk (5-14 kpc, but with few stars between 5 and 7 kpc, see Fig. 2). Since these are the data constraining our model, our results describe best the outer disk of the Milky Way. But the inner low- α disk should contain additional information on inside-out growth in its spatial-age-metallicity distributions, unless it had a different formation scenario (as argued in e.g. Haywood et al. 2019). As shown in the test of subsection 4.3.2, including more APOGEE fields with data from the inner disk mid-plane affects some parameters, for example the star formation time-scale τ_{SFR} and the inferred scale-length of our outlier model (old component). We have kept consistency with the data set used in Frankel et al. (2018), but we plan to turn to APOGEE DR16 and the inner Galaxy in future work.

6.3. Model Shortcomings

To construct a parametric model for the Galactic disk, we have made several assumptions. The core of this model, a radially-dependent star formation history, was built to satisfy the following properties: (1) be qualitatively plausible; (2) be sufficiently flexible to allow the ‘no inside-out growth’, ‘inside-out growth’ and ‘outside-in growth’ scenarios to exist, and to be fit; (3) be a global decreasing exponential in the case of ‘no inside-out growth’, and (4) be such that our data can constrain it.

The functional form presented in Eq. 11 satisfies the properties (2)-(4), and is good with (1) on average. However, it proposes a rather simplistic description of the star formation history: star formation begins at the same time over the entire disk and the only parameter quantifying inside-out growth is the time-scale regulating the star formation. By definition, such a functional form produces an age distribution that has a large variance, and our model does not reproduce exactly the age distribution of the data at given Galactocentric radius. This can be seen in Fig. 6, where our model predicts the mean age profile correctly, but overestimates the standard deviation, and is skewed at large ages. We have experimented with different functional forms for the star formation history, where star formation rises at different times at different Galactocentric radii (see Section 4). Samples from such a model compared much better to the observed data than those from the model we have chosen to present. But due to the lack of old stars in our data set, and large age uncertainties at large ages, we were not able to fit this model reliably (and condition (4) was violated).

In order to improve the description of the data in a physically sensible way, it will be important in the future to introduce realistic parametric models inspired by simulations.

7. SUMMARY AND FUTURE PROSPECTS

We have built a model to constrain the degree of inside-out growth of the Milky Way disk. Applied to APOGEE red clump stars, the model fit leads to significant inside-out growth and has several implications. In this modelling context, we find that

- the model fit implies a growth rate of the stellar disk of 0.2 kpc Gyr^{-1} and a present-day half-mass radius of 5.9 kpc , resulting from a 43% growth over the past 7.5 Gyr;
- comparing the half-mass and half-light radii with redshift-size relations and with the evolution of the mass-size relations of other galaxies shows that the

Milky Way is a rather typical disk galaxy in terms of present-day mass and size, but also in terms of global mass-size evolution. We fitted the evolution of the Milky Way’s half-mass radius $R_{1/2} \propto (1+z)^{-0.49}$;

- important radial redistribution of stars erases present-day evidence for the past history of the disk, including the spatial variations of the star formation history and therefore inside-out growth. This implies that a local age histogram differs from a local star formation history as local stars may have been born at different Galactocentric radii and stars formed locally may have migrated to different radii;
- present-day age gradients can be a poor diagnostic for inside-out growth if radial orbit migration is strong and makes population studies in external galaxies challenging as assessing the strength of radial orbit migration is more difficult (the physical scatter in age-metallicity resulting from radial orbit migration is not accessible through integrated light). Indirect possibilities to assess radial mixing strength in external galaxies and their impact on the stellar age structure are, for example, dynamical studies as in [Gogarten et al. \(2010\)](#).

This methodology builds on a large data set, with extended spatial coverage and for which stellar ages are known, combined with a forward model with parameters physically understandable. This allows to quantify the amount of inside-out growth in the Milky Way disk. However, there are clear limitations to outcome: the data set we have used is overall young (whereas the evolution of the disk was possibly faster at early times), which allows us to fit the recent evolution of the disk well but not earlier times. In addition, the results are bound to all assumptions made while constructing this model.

To confirm and improve these results, reliable age estimates of stellar populations that cover wide age and Galactocentric radius ranges are needed. Red clump stars are excellent distance indicators, but have poor age estimates and are younger than the underlying population. We showed that five different age sets for our red clump stars lead to five different estimates of the degree of inside-out growth, which is a general problem for Galactic archaeology.

With a data set with precise distances (e.g., [Hogg et al. 2018](#); [Leung & Bovy 2019](#)) and less prone to population selection effects, it should be feasible to improve the inside-out growth model for a better description of the build up of the Milky Way disk and inspire the model

construction from numerical simulations, to benefit from direct physical insights. But these requires more computational issues to overcome.

ACKNOWLEDGEMENTS

It is a pleasure to thank Maosheng Xiang and Ted Mackereth for providing us insights on their results for our safety checks, and Gregory Green for practical help on using the Bayestar17 extinction map. We thank Diane Feuillet and Andy Gould for useful discussions on APOGEE selection function, and Alina Boecker and Ryan Leaman for advice on stellar populations. We are grateful to Sofia Feltzing and Paul McMillan for helping to improve the clarity of the writing. This project was developed in part at the 2018 NYC Gaia Sprint, hosted

by the Center for Computational Astrophysics of the Flatiron Institute in New York City and at the 2018 Gaia-LAMOST hack-a-thon, held in Shanghai. N.F. acknowledges support from the International Max Planck Research School for Astronomy and Cosmic Physics at the University of Heidelberg (IMPRS-HD). H.-W.R. received support from the European Research Council under the European Union’s Seventh Framework Programme (FP 7) ERC Grant Agreement n. [321035]. YST is grateful to be supported by the NASA Hubble Fellowship grant HST-HF2-51425 awarded by the Space Telescope Science Institute.

The following softwares were used during this research: Astropy (Astropy Collaboration et al. 2013), Matplotlib (Hunter 2007).

APPENDIX

A. INSIDE-OUT GROWTH INFERENCE USING DIFFERENT AGE SETS

As part of the analysis, we explore the effect of age determination methods on the results. We repeat the analysis keeping the same stars, the same [Fe/H] and position columns, but changing the age column using five different age sets described below. We find that the overall inside-out growth scenario is robust to age systematics, but the details can change quantitatively. We describe the detailed procedure and results for each age set below.

Ness16—We use ages derived with *The Cannon* (Ness et al. 2015, 2016) trained on spectra and masses of APOKASC1 stars (Pinsonneault et al. 2014), and derived log age with a precision of 0.2 dex. Adapting our noise model from Eq. 20 with $\sigma_{\log\tau} = 0.2$, and optimizing the likelihood, we find a value for the parameter x of 0.62, which is roughly consistent with our results with the Ting age set: the star formation time-scale decreases slowly towards the outer disk and flattens at $R_0 = 12$ kpc. Additionally, the best fit parameters of the other model aspects are consistent with the work in Frankel et al. (2018), who used this data set. In particular, the value of the radial orbit migration term was exactly the same: a mixing scale of 3.6 kpc after 8 Gyr. This confirms the robustness of this result, and suggests that our new model aspects and APOGEE selection function are incorporated correctly.

Ness18—This data set also used *The Cannon*, trained on APOKASC2 spectra and masses. We apply the same noise model as for the Ness16 data set. The best fit inside-out parameter for this data set is $x = 0.50$, which means that the star formation rate depends more weakly on Galactocentric radius, and goes as $1 - 0.5R_0/8$ kpc. According to this fit, the star formation rate is as slowly decreasing function of time throughout the disk until $R_0 = 16$ kpc.

Ness19—Ness et al. (2019) used empirical tight age-abundance relations, and trained *The Cannon* to map from precise abundances measurements for 17 elements to APOKASC2 age estimates. This procedure was applied to the low- α red clump sample of Ting et al. (2018), and yields ages with a precision $\sigma_\tau = 1.6$ Gyr. The resulting distribution of observed ages resembles most that of the Ting18 sample compared to the other age sets. We modified our noise model accordingly with $p(\tau_{\text{obs}} | \tau, \sigma_\tau) \sim \mathcal{N}(\tau, \sigma_\tau)$. The best fit parameter for the inside-out growth on this data set is $x = 1.1$, which implies that the star formation history at 8 kpc has been constant over the past 7.5 Gyr.

Sanders18—We also used the ages presented in Sanders & Das (2018), using the method outlined in Das & Sanders (2019). This method consists of an initial data-driven stage to measure stellar mass from APOGEE spectroscopy as in Ting & Rix (2018) and Ness et al. (2016), but using the information contained in the abundances directly rather than the full spectrum. A second stage compares the derived mass, spectroscopic parameters, photometry and Gaia parallaxes to a set of isochrones folded with a Galactic prior. We first re-run the age determination procedure, removing the priors on the ages. The fit to the data leads to $x = -0.3 < 0$, favouring weak outside-in growth. A key difference between this age determination method and the previous ones is that the luminosity and parallaxes of the stars were explicitly used to derive the stellar masses. This could introduce a bias with distance or extinction, in particular at large distances where the distance information from the parallax is modest and the extinction may

be significant. When comparing these ages with all of the previous ones (Ting18, Ness16, Ness18, Ness19), which are not distance-dependent, we find systematic differences that are function of distance: compared to the other data sets, these ages tend to be overestimated at large distances leading to a contaminating population of old stars at large distances. In our case, due to selection effects, distant stars are mainly located in the outer disk. This produced a general trend with older stars at larger distances and larger Galactocentric radii, weakening (and inverting, in some parts) the already quite weak radial age gradient that contains information about inside-out growth in the data set (see Fig. 6).

REFERENCES

- Agertz, O., Teyssier, R., & Moore, B. 2011, *MNRAS*, 410, 1391
- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, 219, 12
- Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, 558, A33
- Aumer, M., Binney, J., & Schönrich, R. 2016, *MNRAS*, 462, 1697
- Aumer, M., White, S. D. M., & Naab, T. 2014, *MNRAS*, 441, 3679
- Avila-Reese, V., González-Samaniego, A., Colín, P., Ibarra-Medel, H., & Rodríguez-Puebla, A. 2018, *ApJ*, 854, 152
- Barden, M., Rix, H.-W., Somerville, R. S., et al. 2005, *ApJ*, 635, 959
- Bensby, T., Alves-Brito, A., Oey, M. S., Yong, D., & Meléndez, J. 2010, *A&A*, 516, L13
- Bergemann, M., Ruchti, G. R., Serenelli, A., et al. 2014, *A&A*, 565, A89
- Bernard, E. J., Ferguson, A. M. N., Chapman, S. C., et al. 2015, *MNRAS*, 453, L113
- Binney, J., & Tremaine, S. 2008, *Galactic Dynamics: Second Edition* (Princeton University Press)
- Bland-Hawthorn, J., & Gerhard, O. 2016, *ARA&A*, 54, 529
- Boissier, S., Gil de Paz, A., Boselli, A., et al. 2008, *ApJ*, 681, 244
- Bovy, J. 2014, *galpy: Galactic dynamics package*, *Astrophysics Source Code Library*, , ascl:1411.008
- Bovy, J., Rix, H.-W., Green, G. M., Schlafly, E. F., & Finkbeiner, D. P. 2016, *ApJ*, 818, 130
- Bovy, J., Rix, H.-W., Liu, C., et al. 2012, *ApJ*, 753, 148
- Bovy, J., Nidever, D. L., Rix, H.-W., et al. 2014, *ApJ*, 790, 127
- Brook, C. B., Kawata, D., Martel, H., Gibson, B. K., & Bailin, J. 2006, *ApJ*, 639, 126
- Brook, C. B., Stinson, G. S., Gibson, B. K., et al. 2012, *MNRAS*, 426, 690
- Buitrago, F., Trujillo, I., Conselice, C. J., et al. 2008, *ApJL*, 687, L61
- Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, *ApJ*, 798, 7
- Carlberg, R. G., & Sellwood, J. A. 1985, *ApJ*, 292, 79
- Courteau, S., Dutton, A. A., van den Bosch, F. C., et al. 2007, *ApJ*, 671, 203
- Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *Research in Astronomy and Astrophysics*, 12, 1197
- Das, P., & Sanders, J. L. 2019, *MNRAS*, 484, 294
- de Vaucouleurs, G. 1948, *Annales d’Astrophysique*, 11, 247
- Dutton, A. A., van den Bosch, F. C., Faber, S. M., et al. 2011, *MNRAS*, 410, 1660
- Fall, S. M., & Efstathiou, G. 1980, *MNRAS*, 193, 189
- Ferguson, H. C., Dickinson, M., Giavalisco, M., et al. 2004, *ApJL*, 600, L107
- Feuillet, D. K., Bovy, J., Holtzman, J., et al. 2016, *ApJ*, 817, 40
- . 2018, *MNRAS*, 477, 2326
- Frankel, N., Rix, H.-W., Ting, Y.-S., Ness, M., & Hogg, D. W. 2018, *ApJ*, 865, 96
- Franx, M., van Dokkum, P. G., Förster Schreiber, N. M., et al. 2008, *ApJ*, 688, 770
- Freeman, K., & Bland-Hawthorn, J. 2002, *ARA&A*, 40, 487
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, *A&A*, 616, A1
- García-Benito, R., González Delgado, R. M., Pérez, E., et al. 2017, *A&A*, 608, A27
- Girardi, L. 2016, *ARA&A*, 54, 95
- Goddard, D., Thomas, D., Maraston, C., et al. 2017, *MNRAS*, 466, 4731
- Gogarten, S. M., Dalcanton, J. J., Williams, B. F., et al. 2010, *ApJ*, 712, 858
- González Delgado, R. M., Pérez, E., Cid Fernandes, R., et al. 2014, *A&A*, 562, A47
- Grand, R. J. J., Gómez, F. A., Marinacci, F., et al. 2017, *MNRAS*, 467, 179
- Green, G. M., Schlafly, E. F., Zucker, C., Speagle, J. S., & Finkbeiner, D. P. 2019, *arXiv e-prints*, arXiv:1905.02734
- Green, G. M., Schlafly, E. F., Finkbeiner, D., et al. 2018, *MNRAS*, 478, 651

- Grisoni, V., Spitoni, E., & Matteucci, F. 2018, *MNRAS*, 481, 2570
- Hayden, M. R., Bovy, J., Holtzman, J. A., et al. 2015, *ApJ*, 808, 132
- Haywood, M., Snaith, O., Lehnert, M. D., Di Matteo, P., & Khoperskov, S. 2019, *A&A*, 625, A105
- Hogg, D. W., Eilers, A.-C., & Rix, H.-W. 2018, arXiv e-prints, arXiv:1810.09468
- Hunter, J. D. 2007, *Computing In Science & Engineering*, 9, 90
- Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2005, *ApJ*, 619, 931
- Kroupa, P. 2001, *MNRAS*, 322, 231
- Lacey, C. G. 1984, *MNRAS*, 208, 687
- Laney, C. D., Joner, M. D., & Pietrzyński, G. 2012, *MNRAS*, 419, 1637
- Larson, R. B. 1976, *MNRAS*, 176, 31
- Leung, H. W., & Bovy, J. 2019, arXiv e-prints, arXiv:1902.08634
- Lindgren, L., Hernández, J., Bombrun, A., et al. 2018, *A&A*, 616, A2
- Lynden-Bell, D., & Kalnajs, A. J. 1972, *MNRAS*, 157, 1
- MacArthur, L. A., Courteau, S., Bell, E., & Holtzman, J. A. 2004, *ApJS*, 152, 175
- Mackereth, J. T., Bovy, J., Schiavon, R. P., et al. 2017, *MNRAS*, 471, 3057
- Majewski, S. R., Zasowski, G., & Nidever, D. L. 2011, *ApJ*, 739, 25
- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, 154, 94
- Martig, M., Fouesneau, M., Rix, H.-W., et al. 2016, *MNRAS*, 456, 3655
- Minchev, I., Chiappini, C., & Martig, M. 2014, *A&A*, 572, A92
- Minchev, I., Anders, F., Recio-Blanco, A., et al. 2018, *MNRAS*, 481, 1645
- Minchev, I., Matijevic, G., Hogg, D. W., et al. 2019, *MNRAS*, 487, 3946
- Mo, H. J., Mao, S., & White, S. D. M. 1998, *MNRAS*, 295, 319
- Muñoz-Mateos, J. C., Boissier, S., Gil de Paz, A., et al. 2011, *ApJ*, 731, 10
- Muñoz-Mateos, J. C., Gil de Paz, A., Boissier, S., et al. 2007, *ApJ*, 658, 1006
- Ness, M., Hogg, D. W., Rix, H.-W., Ho, A. Y. Q., & Zasowski, G. 2015, *ApJ*, 808, 16
- Ness, M., Hogg, D. W., Rix, H.-W., et al. 2016, *ApJ*, 823, 114
- Ness, M. K., Johnston, K. V., Blancato, K., et al. 2019, arXiv e-prints, arXiv:1907.10606
- Peebles, P. J. E. 1969, *ApJ*, 155, 393
- Pérez, E., Cid Fernandes, R., González Delgado, R. M., et al. 2013, *ApJL*, 764, L1
- Pezzulli, G., Fraternali, F., Boissier, S., & Muñoz-Mateos, J. C. 2015, *MNRAS*, 451, 2324
- Pilkington, K., Few, C. G., Gibson, B. K., et al. 2012, *A&A*, 540, A56
- Pinsonneault, M. H., Elsworth, Y., Epstein, C., et al. 2014, *ApJS*, 215, 19
- Pinsonneault, M. H., Elsworth, Y. P., Tayar, J., et al. 2018, *ApJS*, 239, 32
- Ricciardelli, E., Vazdekis, A., Cenarro, A. J., & Falcón-Barroso, J. 2012, *MNRAS*, 424, 172
- Rix, H.-W., & Bovy, J. 2013, *A&A Rv*, 21, 61
- Rodríguez-Puebla, A., Primack, J. R., Avila-Reese, V., & Faber, S. M. 2017, *MNRAS*, 470, 651
- Sacchi, E., Cignoni, M., Aloisi, A., et al. 2019, arXiv e-prints, arXiv:arXiv:1905.00020
- Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, *A&A*, 538, A8
- Sanders, J. L., & Binney, J. 2015, *MNRAS*, 449, 3479
- Sanders, J. L., & Das, P. 2018, *MNRAS*, 481, 4093
- Schönrich, R., & Binney, J. 2009a, *MNRAS*, 396, 203
- . 2009b, *MNRAS*, 399, 1145
- Schönrich, R., & McMillan, P. J. 2017a, *MNRAS*, 467, 1154
- . 2017b, *MNRAS*, 467, 1154
- Sellwood, J. A. 2014, *Reviews of Modern Physics*, 86, 1
- Sellwood, J. A., & Binney, J. J. 2002, *MNRAS*, 336, 785
- Silva Aguirre, V., Bojsen-Hansen, M., Slumstrup, D., et al. 2018, *MNRAS*, 475, 5487
- Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, *AJ*, 131, 1163
- Somerville, R. S., Barden, M., Rix, H.-W., et al. 2008, *ApJ*, 672, 776
- Ting, Y.-S., Conroy, C., & Goodman, A. 2015, *ApJ*, 807, 104
- Ting, Y.-S., Hawkins, K., & Rix, H.-W. 2018, *ApJL*, 858, L7
- Ting, Y.-S., & Rix, H.-W. 2018, arXiv e-prints, arXiv:1808.03278
- Trujillo, I., Förster Schreiber, N. M., Rudnick, G., et al. 2006, *ApJ*, 650, 18
- Tully, R. B., & Fisher, J. R. 1977, *A&A*, 54, 661
- van der Wel, A., Franx, M., van Dokkum, P. G., et al. 2014, *ApJ*, 788, 28
- van Dokkum, P. G., Leja, J., Nelson, E. J., et al. 2013, *ApJL*, 771, L35
- Vazdekis, A., Ricciardelli, E., Cenarro, A. J., et al. 2012, *MNRAS*, 424, 157

Vazdekis, A., Sánchez-Blázquez, P., Falcón-Barroso, J.,
et al. 2010, MNRAS, 404, 1639

Wang, J., Kauffmann, G., Overzier, R., et al. 2011,
MNRAS, 412, 1081

Williams, B. F., Dalcanton, J. J., Dolphin, A. E.,
Holtzman, J., & Sarajedini, A. 2009, ApJL, 695, L15

Xiang, M., Shi, J., Liu, X., et al. 2018, ApJS, 237, 33

Zasowski, G., Johnson, J. A., Frinchaboy, P. M., et al.
2013, AJ, 146, 81