The chemical and spatial variations of the bulge's velocity ellipsoids

Luis M. San Martin Fernandez,[1]★ Steven Gough-Kelly [ID],[1] Victor P. Debattista [ID],[1]★ Oscar A. Gonzalez,[2] Ilin Lazar [ID],[3] Alvaro Rojas-Arriagada [ID][4,5,6,7] and Leandro Beraldo e Silva [ID][8]

[1]*Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK*
[2]*UK Astronomy Technology Centre, Royal Observatory, Blackford Hill, Edinburgh EH9 3HJ, UK*
[3]*Centre for Astrophysics Research, School of Physics, Engineering & Computer Science, University of Hertfordshire, Hatfield AL10 9AB, UK*
[4]*Departamento de Física, Universidad de Santiago de Chile, Av. Victor Jara 3659, Santiago 9170124, Chile*
[5]*Millennium Institute of Astrophysics, Avenida Vicuña Mackenna 4860, Macul, Santiago 82-0436, Chile*
[6]*Núcleo Milenio ERIS, Santiago, Chile*
[7]*Center for Interdisciplinary Research in Astrophysics and Space Exploration (CIRAS), Universidad de Santiago de Chile, Santiago, Chile*
[8]*Department of Astronomy & Steward Observatory, University of Arizona, Tucson, AZ 85721, USA*

ABSTRACT

We study the velocity ellipsoids in an N-body+SPH (smooth particle hydrodynamics) simulation of a barred galaxy which forms a bar with a BP bulge. We focus on the 2D kinematics, and quantify the velocity ellipses by the anisotropy, β_{ij}, the correlation, ρ_{ij}, and the vertex deviation, l_v. We explore the variations in these quantities based on stellar age within the bulge and compare these results with the Milky Way's bulge using data from APOGEE (Apache Point Observatory Galactic Evolution Experiment) DR16 and *Gaia* DR3. We first explore the variation of the model's velocity ellipses in galactocentric velocities, v_R and v_ϕ, for two bulge populations, a (relatively) young one and an old one. The bar imprints quadrupoles on the distribution of ellipse properties, which are stronger in the young population, as expected from their stronger bar. The quadrupoles are distorted if we use heliocentric velocities v_r and v_l. We then project these kinematics along the line of sight onto the (l, b)-plane. Along the minor axis β_{rl} changes from positive at low $|b|$ to negative at large $|b|$, crossing over at lower $|b|$ in the young stars. Consequently, the vertex deviation peaks at lower $|b|$ in the young population, but reaches similar peak values in the old. The ρ_{rl} is much stronger in the young stars, and traces the bar strength. The APOGEE stars split by the median [Fe/H] follow the same trends. Lastly, we explore the velocity ellipses across the entire bulge region in (l, b) space, finding good qualitative agreement between the model and observations.

Key words: Galaxy: bulge – Galaxy: kinematics and dynamics – Galaxy: structure.

1 INTRODUCTION

Vertex deviation measures the tilt of the velocity ellipse away from the coordinate axes. In an axisymmetric system, symmetry requires that the vertex deviation is $0°$. When computed for bulge stars along the minor (rotation) axis, a non-zero vertex deviation of the v_r–v_l velocity ellipse is therefore a useful probe of the deviation of the bulge from axisymmetry, and in particular of the bar's triaxiality, even with a relatively small sample of stars. The measurement by Zhao, Spergel & Rich (1994) of a significant vertex deviation for a sample of 62 K giants in Baade's Window provided the first stellar kinematic evidence that the bulge of the Milky Way (MW) is barred. Zhao et al. (1994) found that the vertex deviations of the metal-rich and metal-poor stars were perpendicular to one another, with negative and positive values respectively. Soto, Rich & Kuijken (2007) studied a larger sample of stars and also found a negative vertex deviation for the metal-rich stars, but an untilted velocity

ellipse for the metal-poor ones (see also Soto, Kuijken & Rich 2012). They pointed out that the positive vertex deviation found by Zhao et al. (1994) for metal-poor stars was mainly due to their small sample size, and was not statistically significant. These results were confirmed by Babusiaux et al. (2010), who found a metal-rich population with significant vertex deviation, and a metal-poor population consistent with a spheroid of vanishing vertex deviation. Simion et al. (2021) also found a metallicity dependence of the vertex deviation in Baade's Window, although they did not detect a coherent metallicity dependence in other fields.

The difference in the vertex deviation of metal-rich and metal-poor stars has been interpreted as evidence of the co-existence of two different components in the bulge. In this scenario, the bulge is comprised of both a secularly evolved box/peanut bulge driven by the thickening of the bar (e.g. Sellwood & Gerhard 2020), as well as an older, nearly axisymmetric, accreted 'classical' bulge in the metal-poor stars (e.g. Vásquez et al. 2013). Other evidence that has been invoked in favour of a significant accreted population in the bulge include velocity dispersions that vary with metallicity (Babusiaux 2016; Zhao et al. 1994), the weak bar in RR Lyrae (Dékány et al. 2013), the vertical metallicity gradient (Gonzalez

★ E-mails: vpdebattista@uclan.ac.uk (VPD); luismi98sanmartin@gmail.com (LMSMF)

et al. 2011; Johnson et al. 2011, 2013; Zoccali et al. 2008), the absence of an X-shape in metal-poor stars (Ness et al. 2012; Rojas-Arriagada et al. 2014; Uttenthaler et al. 2012), the bimodal metallicity distribution (Ness et al. 2013; Johnson et al. 2022; Rojas-Arriagada et al. 2020; Schultheis et al. 2017; Williams et al. 2016), and the mostly old bulge stellar ages (Ortolani et al. 1995; Kuijken & Rich 2002; Ferreras, Wyse & Silk 2003; Clarkson et al. 2008; Brown et al. 2010; Clarkson et al. 2011; Calamida et al. 2014; Sahu et al. 2006; Sit & Ness 2020; Surot et al. 2019; Valenti et al. 2013; Zoccali et al. 2003).

However, Debattista et al. (2017) showed that many of the trends with metallicity found in the bulge can be reproduced purely by the secular evolution of the bar. This occurs because the bar exerts a different influence on stellar populations of different initial kinemat-ics, spatially separating them. Cooler (hence relatively young,[1] and metal-rich) stars form a strong bar, with a clear X-shaped distribution, while hotter (hence older and metal-poor) stars form a weaker bar, which is thicker at the centre, and has little or no X-shape. The vertical metallicity gradient then results from the different density distribu-tion of the young and old stars. Since the populations are separated by the bar on the basis of their kinematics, Debattista et al. (2017) referred to this process as kinematic fractionation. In this scenario, the bimodal metallicity distribution requires contribution from a thick disc population (Bekki & Tsujimoto 2011; Di Matteo et al. 2015; Fragkoudi et al. 2017a, b, 2018, 2020; Di Matteo et al. 2019), possibly via the accretion of star-forming clumps (Debattista et al. 2023), which is also an *in-situ* formation mechanism. Observations of high-redshift galaxies also find evidence of an early formation of bulges via the gravitational instability of the gas and subsequent bursty star formation (e.g. Tacchella et al. 2015; ichi Tadaki et al. 2017). Shen et al. (2010) used a set of N-body simulations to show that the kinematics of the bulge can be matched with a bulge that included no more than 8 per cent of the total mass in the form of a slowly rotating classical bulge. Debattista et al. (2017) refined this estimate by considering the kinematics as a function of metallicity, finding that a classical bulge needed to be less massive than ~ 2 per cent of the total stellar mass.

Debattista et al. (2019) studied the variation of the vertex deviation with age and metallicity in a cosmological simulation of an MW-like galaxy from the FIRE project (Hopkins 2015; Hopkins et al. 2018). Rescaling the simulation, they found, for the equivalent of Baade's Window, a vertex deviation trend with metallicity comparable to that of Soto et al. (2007) and Babusiaux et al. (2010): a significant and constant vertex deviation for stars with [Fe/H] $\gtrsim -0.7$ and a vanishing one for metal-poor stars. They showed that the vertex deviation of old stars is low regardless of whether all stars are included or just those that formed *in situ*, suggesting that vertex deviation, as indeed many other kinematic quantities (Gough-Kelly et al. 2022), exhibits the effect of kinematic fractionation.

In this study, we analyse the kinematics of a high-resolution N-body+SPH (smooth particle hydrodynamics) simulation where stars form continuously out of gas, and compare the results with the Rojas-Arriagada et al. (2020) sample of APOGEE-1 (Apache Point Observatory Galactic Evolution Experiment) and APOGEE-2 stars (Majewski et al. 2017), with proper motions from *Gaia* DR3 (Gaia Collaboration 2023). Our model is the same as that used in Debattista et al. (2017), which evolves in isolation purely secularly and exhibits kinematic fractionation. This high-resolution

---

[1]In this context, a young population can be just $1 - 2$ Gyr younger than an ~ 11 Gyr-old population. Youth is very much relative for bulge populations.

simulation allows us to predict trends of the velocity ellipses across the bulge region. A qualitative agreement with the MW's trends would constitute evidence that our Galaxy's bulge formed secularly with stellar populations separated by kinematic fractionation, without the need to invoke a significant accreted bulge component.

We first define the vertex deviation in Section 2. Section 3 describes the simulation from which the model we use is drawn. In Section 4, we map the model's bulge galactocentric and heliocentric kinematics, paying attention not just to the vertex deviation but also to the in-plane anisotropy and correlation. Section 5 describes the observational data we use to compare with the model. Then in Section 6, we explore the vertex deviation along the bulge minor axis, comparing the heliocentric kinematics of the model and of the observations. We explore the kinematics both as a function of latitude, and as a function of age (for the model) or metallicity (for the observations). In Section 8, we extend our study of the ellipses across the entire bulge, for both the model and the observations. In Section 9, we discuss and summarize our results. In the appendices, we rederive the vertex deviation and contrast two differing definitions, briefly present results for two further simulations, and discuss the validity of bootstrapping for our statistics of interest.

2 VERTEX DEVIATION

The spread in stellar velocities at each position $\boldsymbol{x}$ in the Galaxy around the mean velocity is described by the velocity-dispersion tensor,

$$\sigma_{ij}^2(\boldsymbol{x}) \equiv \frac{1}{\rho(\boldsymbol{x})} \int (v_i - \langle v_i \rangle)(v_j - \langle v_j \rangle) f(\boldsymbol{x}, \boldsymbol{v}) \mathrm{d}^3 \boldsymbol{v}$$
$$= \langle v_i v_j \rangle - \langle v_i \rangle \langle v_j \rangle, \tag{1}$$

where $i, j \in \{1, 2, 3\}$ are 3D velocity components, $f(\boldsymbol{x}, \boldsymbol{v})$ is the stellar distribution function at the phase space position $(\boldsymbol{x}, \boldsymbol{v})$, and $\rho(\boldsymbol{x}) = \int f(\boldsymbol{x}, \boldsymbol{v}) \mathrm{d}^3 \boldsymbol{v}$ is the density.

Since the velocity dispersion tensor is symmetric, it can always be diagonalized. The principal axes of this tensor define a 2D projection of the velocity ellipsoid, i.e. a velocity ellipse. The deviation of the velocity ellipse from alignment with the coordinate axes is called the vertex deviation, l_v, and is given by

$$l_v = \frac{1}{2} \arctan \frac{2\sigma_{ij}^2}{|\sigma_{ii}^2 - \sigma_{jj}^2|}, \tag{2}$$

which takes values $-45° \leq l_v \leq 45°$ (see Appendix A). In order to interpret vertex deviation values, it is useful to separately study each of the quantities that appear in equation (2). Instead of the covariance (σ_{ij}^2) and variances (σ_{ii}^2, σ_{jj}^2) of the velocities, we consider their dimensionless equivalents. These are the correlation,

$$\rho_{ij} = \frac{\sigma_{ij}^2}{\sigma_i \sigma_j}, \tag{3}$$

which takes values $-1 \leq \rho_{ij} \leq 1$, and the in-plane anisotropy,

$$\beta_{ij} = 1 - \frac{\sigma_{jj}^2}{\sigma_{ii}^2}, \tag{4}$$

which takes values $-\infty \leq \beta_{ij} \leq 1$. The anisotropy is negative when $\sigma_{ii}^2 < \sigma_{jj}^2$. Then, l_v can be expressed purely in terms of ρ_{ij} and β_{ij} (see Appendix A4, equation A10). Equation (2) measures the degree by which the local direction of highest dispersion deviates from alignment with the coordinate axes with respect to which the velocities are being measured. The direction of highest dispersion in an axisymmetric bulge is everywhere cylindrically radial, i.e. aligned along $\hat{R}$ of galactocentric cylindrical coordinates. The presence of

a bar leads to a deviation from such alignment, because the bar-supporting orbits stream along the bar. The vertex deviation can detect this deviation, making it a useful tool for studying the bar. Thus, for instance, Simion et al. (2021) used the vertex deviation and the simulation of Shen et al. (2010) to show that the MW's bar is tilted by $29° \pm 3°$ with respect to the line joining the Sun and the Galactic Centre.

Vertex deviation quantifies non-axisymmetry when stars are selected in a region where the chosen coordinate system captures the symmetry of an axisymmetric galaxy. Doing so ensures that a deviation of the direction of highest dispersion from the coordinate axes (as measured by vertex deviation) can be interpreted as a deviation from axisymmetry. The Galactocentric cylindrical coordinate system fulfills this everywhere across the Galaxy. However, the Galactocentric cylindrical unit vectors change direction rapidly across the bulge region due to proximity to the Galactic Centre. Given stars selected to build a velocity ellipse have to 'agree' on the directions of the coordinate axes for the vertex deviation value to be meaningful, a vertex deviation analysis using the Galactocentric cylindrical coordinate system requires dense spatial coverage and high spatial resolution (necessitating observations with very precise heliocentric distances). The spherical Galactic coordinate system, on the contrary, has unit vectors whose directions are relatively stable across the bulge region but requires stars to be selected near $l = 0°$, where the unit vectors align most closely with those of the Galactocentric cylindrical coordinate system. Studies of the v_r–v_l velocity ellipse in Baade's Window have been preferred not only because of its low extinction but also because it is close to $l = 0°$. Elsewhere the spherical Galactic unit vectors are a combination of Galactocentric radial and tangential velocities, whose directions vary with heliocentric distance, which complicates the interpretation of the vertex deviation values. Equation (2) shows that the vertex deviation becomes increasingly sensitive to any non-zero correlation as the anisotropy approaches zero. What happens in such cases is that, for an isotropic group of stars, given the lack of preferential direction towards either coordinate axis, any non-zero correlation by symmetry results in an ellipse that is elongated along the diagonal. Mathematically, $l_v \to \text{sign}(\rho_{ij}) \cdot 45°$ as $\beta_{ij} \to 0$ if $\rho_{ij} \neq 0$. The lack of dependence on the magnitude of ρ_{ij} in such situations means that vertex deviation is a blunt probe of non-axisymmetry when the velocity distribution is isotropic. If the correlation is instead negligible, the velocity ellipse of the isotropic population approximates a circle, with poorly defined semimajor axis direction. As a result, its vertex deviation is unstable and exhibits no statistical significance upon bootstrapping (see equation 5). That is the special case of isotropic axisymmetry.

We will reserve the symbol l_v to refer to the vertex deviation of the velocity ellipse constructed using the heliocentric spherical Galactic velocities v_r and v_l. The use of other coordinate systems will be reflected in the notation. For example, $l_v^{R\phi}$ will be computed using the Galactocentric cylindrical velocities v_R and v_ϕ.

3 SIMULATION

We study the dependence of the vertex deviation on stellar ages across the bulge using the same N-body+SPH simulation as was used by Debattista et al. (2017) to demonstrate kinematic fractionation. The simulation is evolved in isolation with GASOLINE (Wadsley, Stadel & Quinn 2004) for 10 Gyr, by which point it has formed $\sim 1.1 \times 10^7$ stellar particles (hereafter stars). We use the 10 Gyr snapshot as our model. This same model was also used by Gough-Kelly et al. (2022) to compare the bulge proper motion rotation curves of old and young

stars; our use of the model is very similar to that of Gough-Kelly et al. (2022). Besides these works, the same model and the associated simulation has also been presented in several other studies (*e.g.* Cole et al. 2014; Gardner et al. 2014; Ness et al. 2014; Gonzalez et al. 2016), therefore we do not describe it at length here, but merely refer the reader to those earlier papers.

The simulation evolves via the cooling of gas off a hot corona in pressure equilibrium within a spherical dark matter halo. The dark matter halo has a virial radius of ~ 200 kpc, a virial mass of 9×10^{11} M$_\odot$, and a concentration $c = 19$; the halo is represented by 5 million particles of unequal mass as described in Cole et al. (2014). The gas follows the same radial profile but has only 11 per cent of the total mass. The gas is imparted with a solid body rotation to give it a spin $\lambda = 0.041$. Initially, the gas corona is also comprised of 5 million particles (for an initial gas particle mass of 2.7×10^4 M$_\odot$). As the gas cools, it settles into a disc and, where it reaches high density (> 100 amu cm^{-3}), it ignites star formation with a probability of 0.1 if the gas is part of a convergent flow and cooler than 1.5×10^4 K. Star formation then leads to supernova (SN) feedback using the blastwave prescription of Stinson et al. (2006).

The stars have an initial mass of 9.4×10^3 M$_\odot$ [up to 46 per cent of the mass of a star is lost via SN feedback and asymptotic giant branch (AGB) stellar winds], leading to a total stellar mass at 10 Gyr of 6.5×10^{10} M$_\odot$. SN explosions couple 0.4×10^{51} erg per SN (Governato et al. 2010) to the gas, and metals pollute the interstellar medium using the yields of Raiteri, Villata & Navarro (1996) for SN II, Thielemann, Nomoto & Yokoi (1986) for SN Ia, and Weidemann (1987) for stellar winds. However, the simulation does not include the diffusion of metals between the gas particles, which results in an excess of low-metallicity stars forming at all ages, and a weakened relation between metallicity and age. As a result, following previous works (Debattista et al. 2017; Gough-Kelly et al. 2022), we define stellar populations on the basis of age rather than metallicity. As the age is more fundamental anyway (albeit harder to measure observationally), this is an acceptable trade-off.

3.1 The model bar

A bar forms in the simulation during the time interval $2 - 4$ Gyr and grows secularly thereafter (Cole et al. 2014). At 2 Gyr around 55 per cent of the total stellar mass in the galaxy has already formed, which is not unusual for galaxies of Sb-type (Tacchella et al. 2015). The result is a bulge with mostly old stars, as shown in the cumulative distribution of ages in Fig. 1, consistent with observations of the MW bulge (e.g. Kuijken & Rich 2002; Clarkson et al. 2008; Valenti et al. 2013; Renzini et al. 2018; Surot et al. 2019).

The model bar has a semimajor axis of ~ 2.9 kpc (Gough-Kelly et al. 2022), whereas in the MW, Wegg, Gerhard & Portail (2015) measured a bar semimajor axis of ~ 5 kpc. We therefore apply a scaling factor of 1.7 to the spatial coordinates of the model. We also apply a kinematic scaling of 0.48 to the model's velocities to scale it to the MW, as in previous studies (Debattista et al. 2017; Gough-Kelly et al. 2022); however this has no effect on the anisotropy, correlation, and vertex deviation since these are based on dimensionless ratios.

Starting with the bar aligned with the x-axis, we rotate it clockwise by $27°$ (Wegg & Gerhard 2013), with the receding near end at positive longitudes. We adopt a right-handed galactocentric rectangular coordinate system, and place the observer on the negative x-axis at $R_0 = 8.1$ kpc from the galactic centre (GRAVITY Collaboration 2018; Qin et al. 2018). In order to match our observational sample (presented in Section 5), we define the bulge as the region enclosed
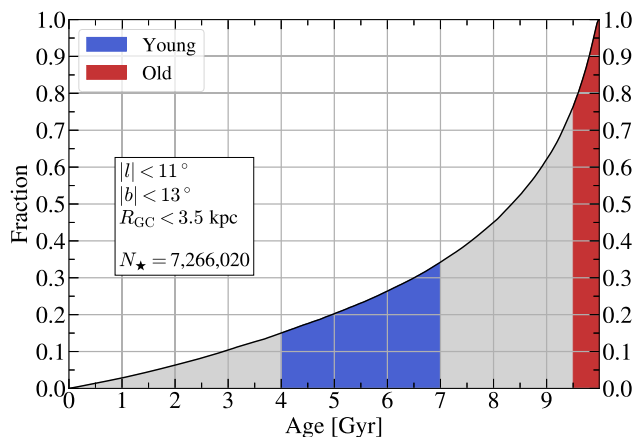
Figure 1. Cumulative distribution of the stellar ages in the model for the bulge region. The total number of bulge stars is $N_\star \sim 7.27 \times 10^6$. We take stars in the range 4–7 Gyr (~19 per cent) and 9.5–10 Gyr (~24 per cent) as representing our young and old bulge populations, respectively.

by $|l| < 11°$, $|b| < 13°$, and $R_{GC} < 3.5$ kpc. This results in a total of 7 266 020 bulge stars.

The face-on view of the stellar populations of different ages in the bulge is shown in Fig. 2, with (a) $|z| < 3$ kpc and (b) $0.5 < |z|/$ kpc < 3. We show stars of ages from 4 to 10 Gyr in intervals of 1 Gyr. The side-on view of the same stellar populations, with $|y| < 5$ kpc, is shown in Fig. 3.

Fig. 2 shows the elongated shape of the distribution of stars older than 4 Gyr, which diminishes progressively as we approach 10 Gyr. Similarly, in Fig. 3, the distribution of younger stars exhibits a prominent X-shape, while that of old stars is boxy. At $|b| \gtrsim 5°$, the arms of the X-shape cause a double peak in the distance distribution (Debattista et al. 2017; Gough-Kelly et al. 2022), consistent with the red clump (RC) bimodality found in the MW (McWilliam & Zoccali 2010; Saito et al. 2011).

We define 'young' and 'old' stellar populations in the model as follows. The young population consists of all the stars in the age range 4−7 Gyr.[2] This population hosts a strong bar and X-shaped bulge. We take stars in the age range 9.5 − 10 Gyr, as the old population, which is spheroidal and boxy. The young and old populations constitute ~19.2 per cent (1 395 563) and ~23.6 per cent (1 713 858) of all bulge stars, respectively (computed over the full volume $|l| < 11°$, $|b| < 13°$, and $R_{GC} < 3.5$ kpc), as shown in Fig. 1.

The model has a prominent nuclear stellar disc, where the youngest stars form (Cole et al. 2014; Debattista et al. 2015). This structure is confined to the galactic plane, with $|z| \lesssim 150$ pc (after the coordinate rescaling). In this study, we are not interested in this structure. We avoid it by excluding stars younger than 4 Gyr, and by only considering stars at $|z| > 0.5$ kpc, which at the galactic centre corresponds to $|b| \gtrsim 3.6°$. The resulting young and old populations are shown in Fig. 4.

4 FACE-ON KINEMATICS

In this section, we explore the differences between the face-on kinematics of the strongly barred young population, and the weakly

---

[2]In this way, we differ from Gough-Kelly et al. (2022) who included all stars younger than 7 Gyr in their young population, and all stars older than 9 in their old population.

barred old population defined for the model in Section 3.1. We bin the stars in $x-y$ space with $0.5 < |z|/$ kpc < 3.

In Section 4.1, we characterize the kinematics of the two populations using the intrinsic galactocentric cylindrical velocities v_R-v_ϕ, while in Section 4.2, we explore how these project into the heliocentric frame by using velocities in galactic coordinates, v_r-v_l. Mean velocity maps of similar populations, for the same model, have already been presented in Gough-Kelly et al. (2022); here we also present mean velocity maps to aid with the interpretation of the different velocity ellipses.

4.1 Galactocentric velocities

Fig. 5 shows maps of the average radial and tangential velocity and velocity dispersion of the galactocentric cylindrical components v_R and v_ϕ. The left block shows $\langle v_R \rangle$ and $\langle v_\phi \rangle$, and the right shows their dispersions, σ_R and σ_ϕ. The left and right columns in each block correspond to the young and old populations, respectively.

The mean velocities for the old population exhibit the expected radial and tangential velocity fields for a relatively slowly rotating, nearly axisymmetric component. The old population has $\langle v_R \rangle \sim 0$ everywhere, and a relatively slow rotation, slightly elongated along the bar major axis, with $\langle v_\phi \rangle$ increasing slowly with R. On the other hand, the quadrupole pattern in $\langle v_R \rangle$ for the young stars reveals the strongly barred nature of this population. We find a good match with the quadrupole found by Bovy et al. (2019, their fig. 2) in the MW. This pattern arises from the radial streaming along the bar. The fast streaming motions of young stars along the bar produce $\langle v_\phi \rangle$ peaks along the bar's minor axis, ~1 kpc away from the centre. The mean velocity then drops in the regions of highest density (in the arms of the X-shape), indicative of stars reaching the apocentre of their orbits. These results are in agreement with those obtained by Gough-Kelly et al. (2022, their fig. 5).

The right block in Fig. 5 shows the galactocentric velocity dispersions of the same populations. The peaks of high σ_R for the young population are caused by the crossing of the high velocity streaming motions along the bar major axis. Weaker peaks are present in σ_ϕ, on the bar's minor axis. The dispersions peak away from the centre of the galaxy. The velocity dispersions of the old stars instead peak at the centre, and then decline slowly radially.

Fig. 6 presents the velocity ellipses for the young (left) and old (right) populations. Even though the ellipses are defined in velocity space, it is useful to overlay them in position space because the orientation of their semimajor axis points in the (local) direction of highest dispersion in the $x-y$ plane. We have centred the ellipses on the centre of each bin (in velocity space they would be offset from the origin by the mean velocity values).

The ellipses are colour-coded by the vertex deviation, $l_v^{R\phi}$. Positive or negative values indicate the semimajor axes are tilted anticlockwise or clockwise, respectively, with respect to the local galactocentric radial vector, $\hat{R}$. An overall radial alignment of the old population is evident, which results in a low vertex deviation everywhere within $R < 1.5$ kpc (ignoring the outer noisy regions with fewer stars), as expected for a nearly axisymmetric system. For the young stars, the ellipses within the bar are aligned along its semimajor axis, which causes radial alignment at the X-shape overdensities but azimuthal alignment along the bar's semiminor axis within $R \lesssim 1$ kpc. In the transition from radial to azimuthal alignment, we find velocity ellipses with very strong tilts, reaching $l_v^{R\phi} \sim \pm45°$, a clear indication of the strong bar in this population.

Fig. 7 quantifies the velocity ellipses across the inner galaxy for the two populations. The first row shows the anisotropy. The dispersion
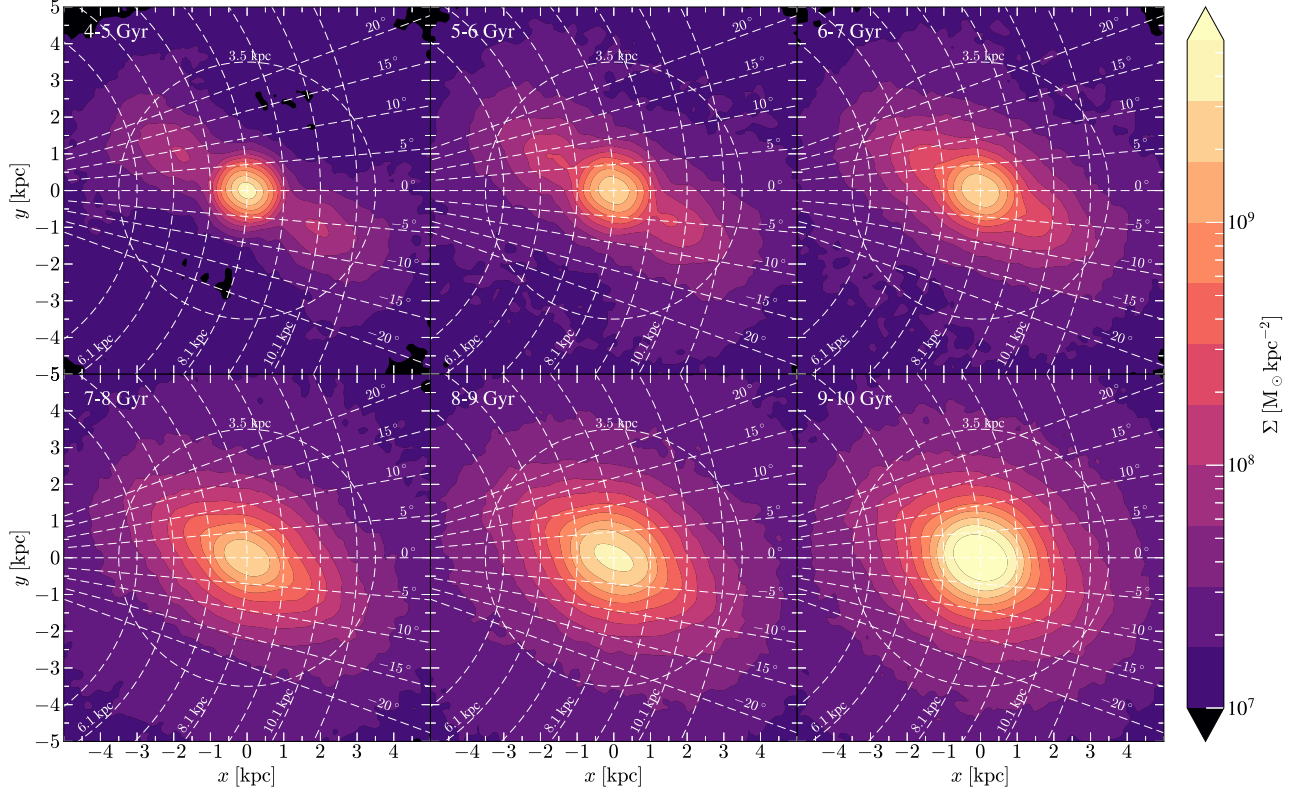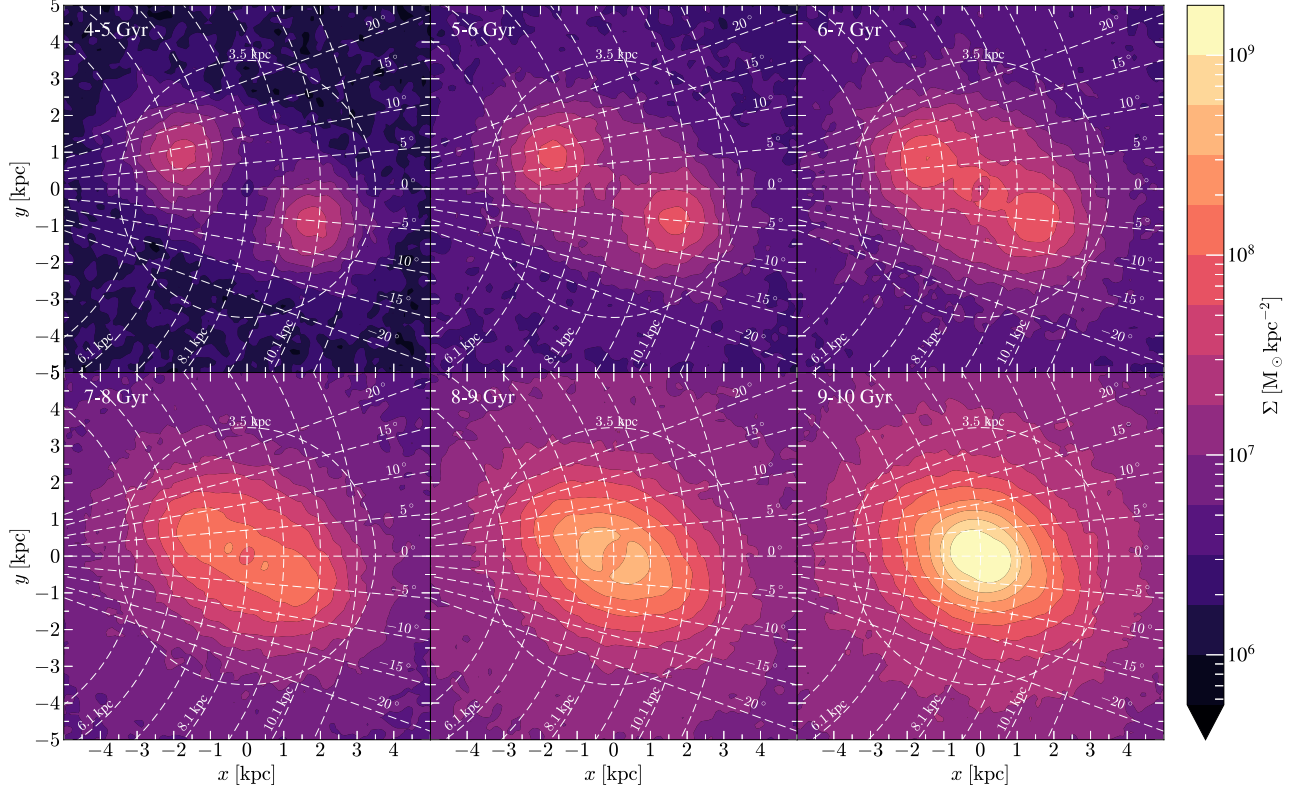
Figure 2. Face-on view of stars of different ages, indicated at top left of each panel, in the bulge of the model, with (a) $|z| < 3$ kpc and (b) $0.5 < |z| < 3$ kpc. The Sun is located at $(x, y) = (-8.1, 0)$ kpc. White dashed lines represent longitudes from $-20°$ to $20°$ in steps of $5°$ (straight lines) and distances from the Sun from 5.1 to 11.1 kpc in steps of 1 kpc (curves). The $R_{GC} = 3.5$ kpc circle is also shown.
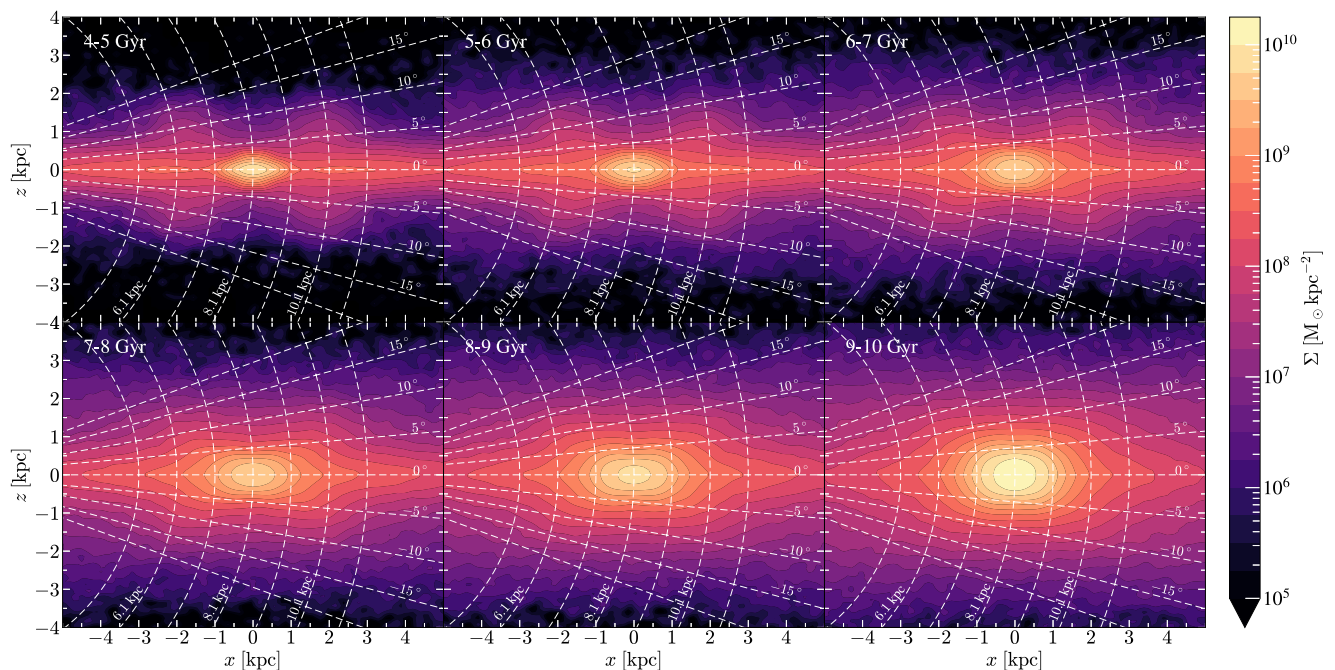
Figure 3. Edge-on view of stars of different ages, as indicated at top left in each panel, in the bulge of the model, with $|y| < 5$ kpc. The Sun is located at $(x, y) = (-8.1, 0)$ kpc. White dashed lines represent latitudes from $-20°$ to $20°$ in steps of $5°$ (straight lines) and distances from the Sun from 5.1 to 11.1 kpc in steps of 1 kpc (curves).

quadrupoles in Fig. 5 imprint quadrupoles in the anisotropy of the young stars. The correlation of the young stars, seen in the middle row, shows a similar quadrupole to that in $\langle v_R \rangle$, although with opposite signs. A quadrupole in $l_v^{R\phi}$ results, which is particularly strong ($\pm 45°$) just outside the centre. The old population instead has mostly positive anisotropy everywhere, as expected for a nearly axisymmetric system. Its weak bar imprints weaker quadrupoles in the anisotropy, correlation and the vertex deviation compared with the young populations.

We estimate uncertainties using the bootstrap method. Given a sample of N stars, after calculating the value of a quantity of interest, x, we obtain B random samples with replacement, of the same size N, and compute the value of the same quantity, x_i^*, for each bootstrap sample i. We then estimate the uncertainty in x as

$$\epsilon(x) = \sqrt{\frac{1}{B} \sum_i^k \left(x_i^* - x \right)^2}. \tag{5}$$

We use $B = 500$ bootstrap iterations.

The uncertainties in Fig. 7 are high where the number of stars is small. The velocity ellipses of old stars have small uncertainties throughout the bulge region because this region is well populated. We also find small errors at the X-shape overdensities of the young stars. The anisotropy of the young stars has a relatively large error (~ 0.1–0.2) along the bar's outer minor axis, approximately at the same location as the negative $\beta_{R\phi}$ lobes. The correlation quadrupole has $|\rho_{R\phi}| > 0.15$, while the errors are $\epsilon(\rho_{R\phi}) < 0.06$. The correlation quadrupole is therefore a robust feature of non-axisymmetry. The regions of $\epsilon(l_v^{R\phi}) \gtrsim 10°$ along the bar's minor axis correspond to ellipses with very small $\beta_{R\phi}$ and $\rho_{R\phi}$. Nearly isotropic, weakly correlated populations have velocity ellipses with ill-defined major axis directions so their vertex deviations are unreliable. The regions

of the vertex deviation quadrupole in Fig. 7 have small associated errors, so this is also a robust feature of non-axisymmetry.

4.2 Velocities in the heliocentric reference frame

We now turn to the bulge as seen from the Sun by considering kinematics in Galactic coordinates. Fig. 8 is the equivalent of Fig. 5 in the heliocentric Galactic coordinates; it shows $\langle v_r \rangle$, $\langle v_l \rangle$, σ_r, and σ_l. The velocities of old stars are nearly symmetric between positive and negative longitudes, which are a result of the near axisymmetry of this population. In contrast, the streaming motions along the bar of the young stars cause the near and far regions at small longitudes to have $\langle v_r \rangle$ of the opposite sign to l. This twist in the velocities with respect to the lines of l gives rise to forbidden velocities (Gough-Kelly et al. 2022) and constitutes part of the observable kinematic imprint of the bar. The strong bar in the young population also causes a slight misalignment between the $\langle v_l \rangle = 0$ line and the line of constant $d = 8.1$ kpc.

The velocity dispersion quadrupoles are no longer aligned along the bar's principal axes when heliocentric velocities are used. Projection results in a weaker quadrupole for young stars in σ_r than in σ_R, while the quadrupole in σ_l is significantly stronger than in σ_ϕ; both quadrupoles are clearly distorted by the effect of perspective.

Fig. 9 shows a strong quadrupole in β_{rl} for the young population. The old stars present a weaker quadrupole which arises largely from projecting their direction of highest dispersion, $\hat{R}$, onto the Galactic coordinates. The same geometric projection gives rise to a quadrupole in the correlation and the vertex deviation of old stars, rotated by $45°$ from that of the anisotropy. In a perfectly axisymmetric population, the black solid contours of $\beta_{rl} = 0$, $\rho_{rl} = 0$, and $l_v = 0°$ would join at the origin. This does not occur due to the weak bar of the old
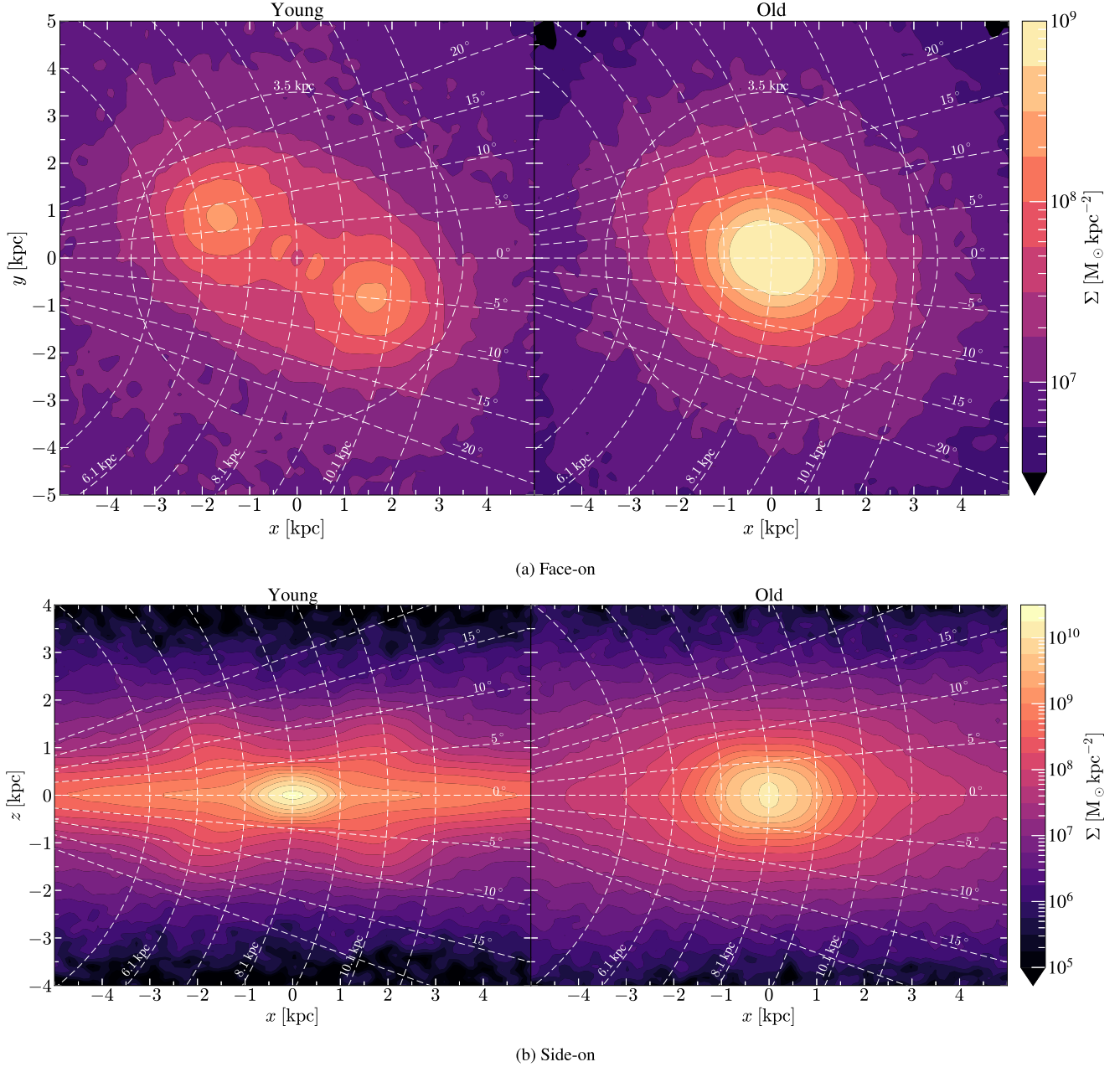
Figure 4. (a) Face-on and (b) side-on views of the surface density of the young (left) and old (right) populations we use. In (a), we use a slice of $0.5 < |z| < 3$ kpc, and in (b), we use $|y| < 5$ kpc. White dashed lines represent longitudes (a) and latitudes (b) from $-20°$ to $20°$ in steps of $5°$ (straight lines) and distances from the Sun from 5.1 to 11.1 kpc in steps of 1 kpc (curves). The $R_{GC} = 3.5$ kpc limit is also shown, as a circle in panel (a).

population. The effect is more extreme for the young stars, with the positive poles of their ρ_{rl} and l_v quadrupoles showing a wider separation. Moreover, the correlation of the young population is significantly stronger.

In the central kpc along $l = 0°$, the vertex deviation is somewhere between $-20°$ and $-30°$. This is not surprising as the bar angle is $-27°$, and the direction of highest dispersion is closely related to the bar direction at $l = 0°$ due to the streaming motions along the bar. This is what enabled Simion et al. (2021) to estimate the bar's angle.

The right block of Fig. 9 shows the bootstrap errors on the heliocentric anisotropy, correlation, and vertex deviation. The old stars have small errors everywhere in the bulge region, just as in

Fig. 7. For the young stars, the anisotropy error is smaller (< 0.1) along the bar's minor axis than for the intrinsic velocities in Fig. 7. The errors on ρ_{rl} and l_v look similar to those in Fig. 7: $\epsilon(\rho_{rl}) < 0.07$ and $\epsilon(l_v) \lesssim 10°$ across most of the bulge region, and larger $\epsilon(l_v)$ peaks located at points of small β_{rl} and ρ_{rl}, along the bar's minor axis. Therefore, the negative correlation and vertex deviation bands along $l = 0°$ in Fig. 9 are stable features of non-axisymmetry.

Thus, the vertex deviation of stars along the $l = 0°$ line is a useful probe of the bar's strength even when the line-of-sight distance of tracer populations is uncertain. The vertex deviation along this direction is robust because $\hat{v}_r$ and $\hat{v}_l$ (and consequently l_v) are well defined and stable for the whole extent of the bulge.
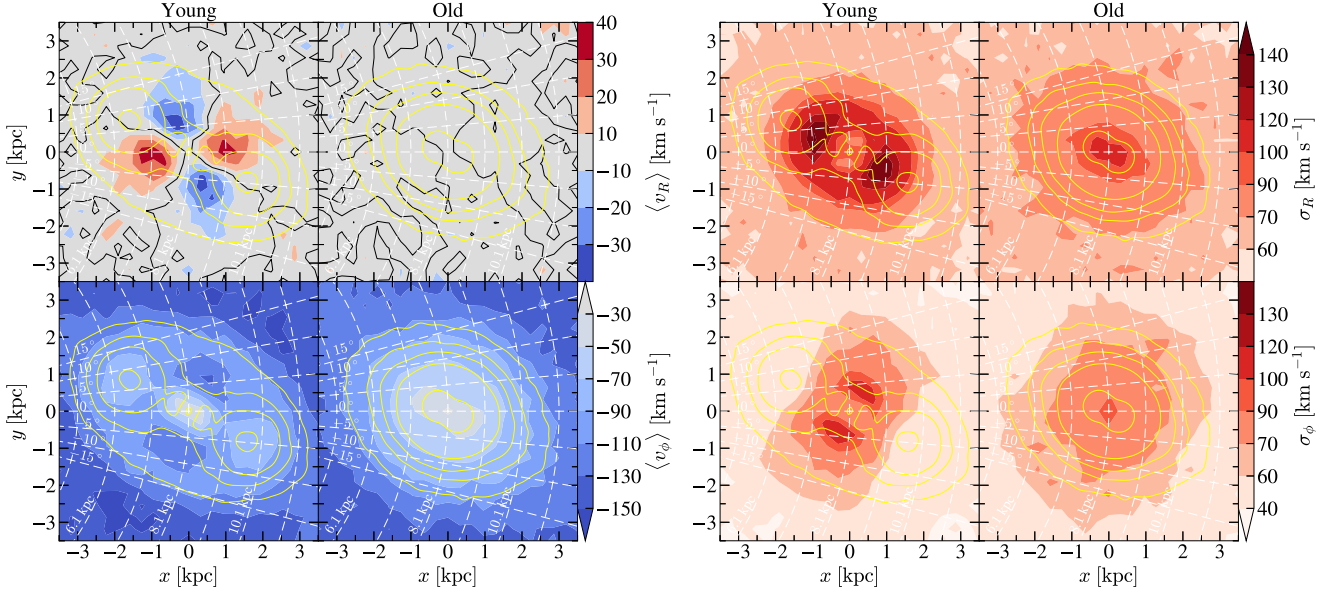
Figure 5. Face-on kinematic maps of the model using galactocentric cylindrical velocities v_R–v_ϕ (see Fig. 8 for v_r–v_l instead). We show two blocks: mean velocities (left) and the corresponding dispersions (right). Each block contains two columns, corresponding to the young (left) and old (right) populations. Black solid contours follow values of $v_R = 0$. Yellow contours outline the density distribution (see Fig. 4a). The Sun is located at $(x, y) = (-8.1, 0)$ kpc. White dashed lines represent longitudes from $-15°$ to $15°$ in steps of $5°$ (straight lines) and distance from the Sun, d, from 6.1 to 10.1 kpc in steps of 1 kpc (curves).
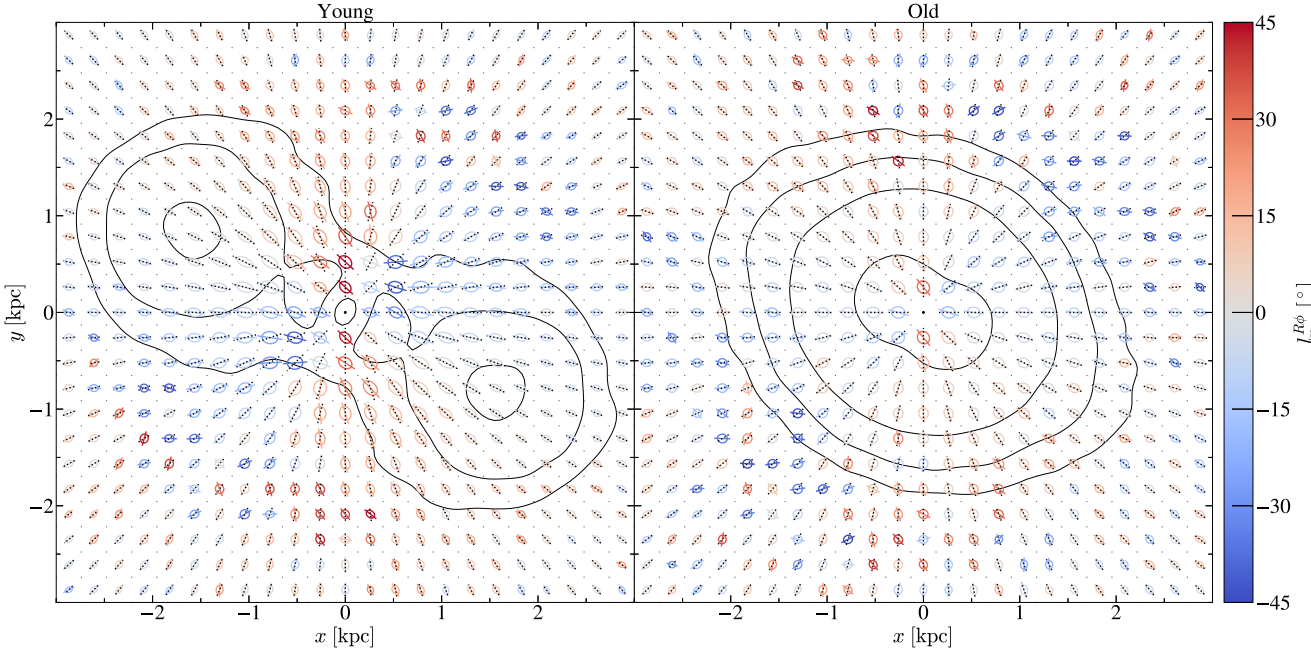
Figure 6. Face-on view of the velocity ellipses of the young (left) and old (right) stars in the model, coloured by the vertex deviation $l_v^{R\phi}$. We have overlaid on each ellipse both its semimajor axis (solid coloured line) and a dotted black line pointing in the radial direction. The black contours outline the density distribution. The grid of black points delimits the different x–y bins. All bins contain at least 50 stars. We omitted the central ellipse due to the instability of the galactocentric unit vectors there.

5 OBSERVATIONAL DATA

In order to test the model's predictions, we define an observational sample of stars in the MW's bulge. In the observations, we use metallicity as a proxy for age, with the expectation that older stars are more metal-poor.

5.1 APOGEE and *Gaia* DR3

We use data from the APOGEE survey (Majewski et al. 2017), part of the Sloan Digital Sky Survey (SDSS). We use APOGEE-1 and APOGEE-2 data, from SDSS-III and -IV, respectively (Eisenstein et al. 2011; Blanton et al. 2017). APOGEE is a high-resolution
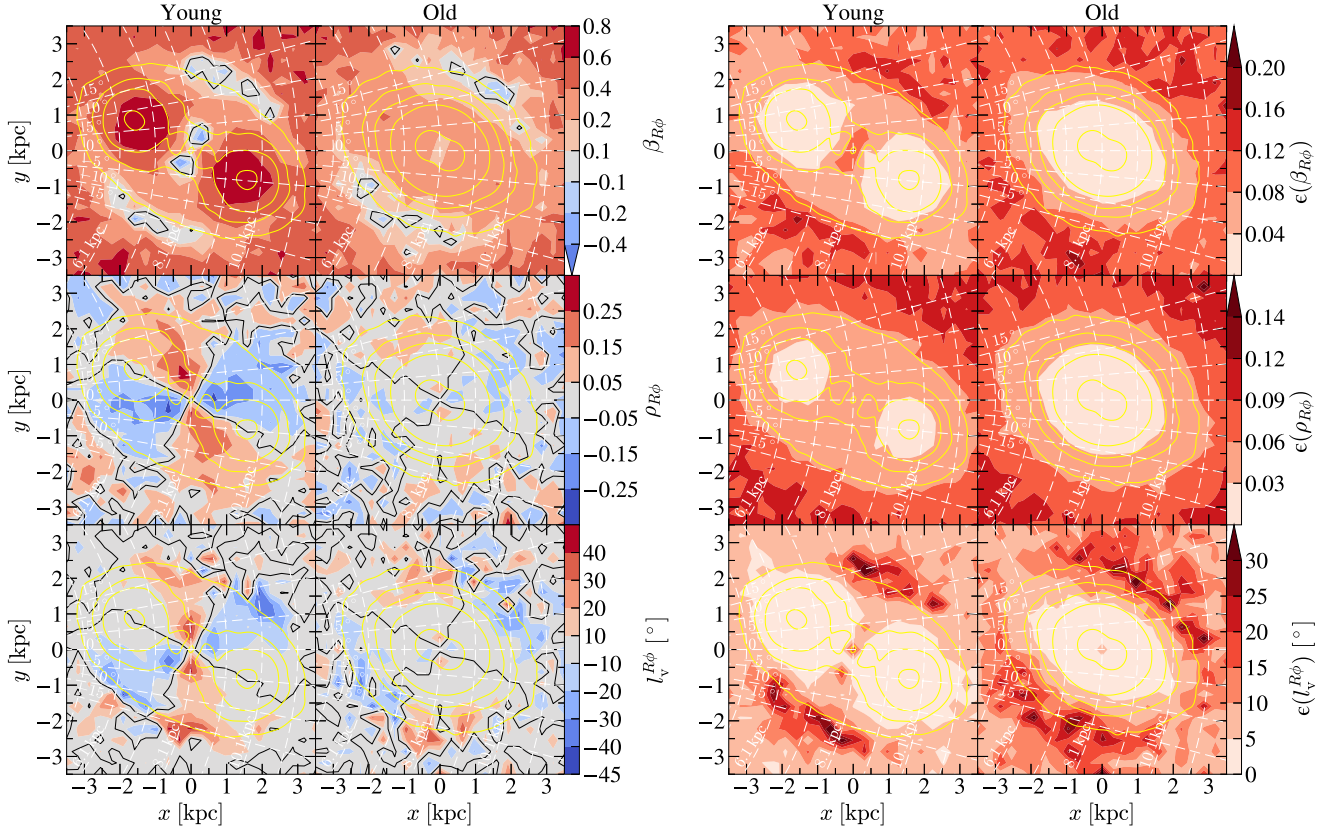
Figure 7. Left block: anisotropy, correlation, and vertex deviation (top to bottom). Right block: the associated errors, computed using bootstrapping with 500 repetitions. Each block contains two columns, corresponding to the young (left) and old (right) populations. Black solid contours follow values of zero. Yellow contours outline the density distribution. The Sun is located at $(x, y) = (-8.1, 0)$ kpc. White dashed lines represent longitudes from $-15°$ to $15°$ in steps of $5°$ (straight lines) and distance from the Sun, d, from 6.1 to 10.1 kpc in steps of 1 kpc (curves).
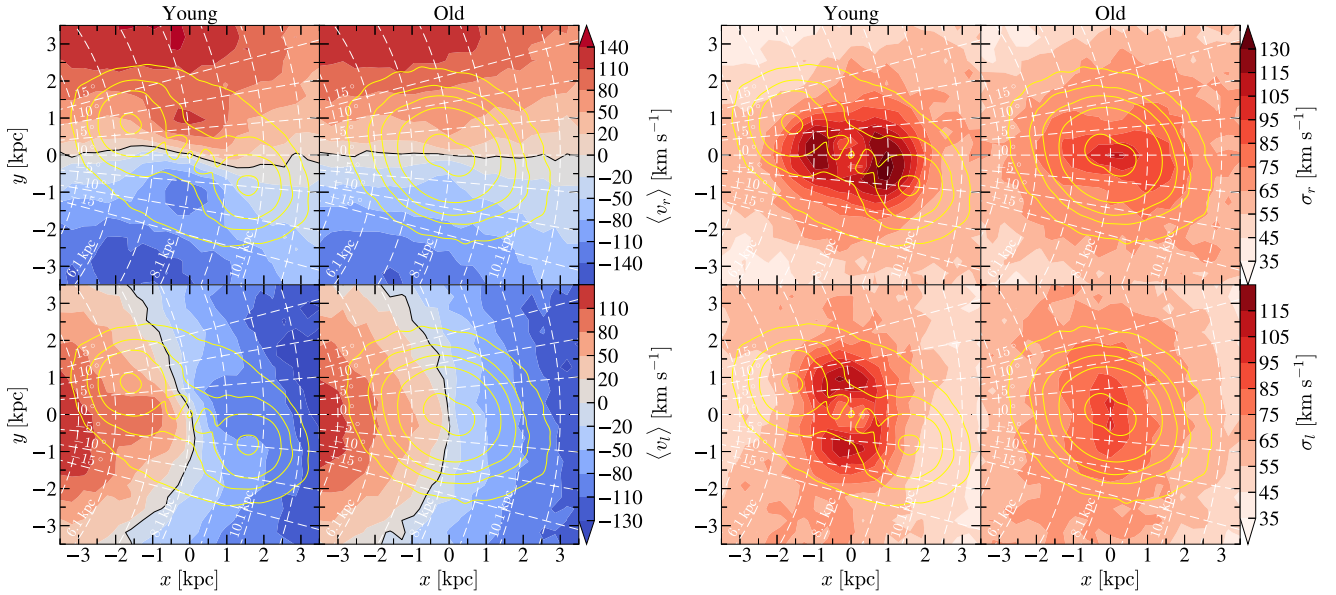

Figure 8. Face-on kinematic maps of the heliocentric galactic velocities, v_r and v_l, for the model (see Fig. 5 for $v_R - v_\phi$ instead). We show two blocks: mean velocities (left) and the corresponding dispersions (right). Each block contains two columns, corresponding to the young (left) and old (right) populations. Black solid contours follow values of zero. Yellow contours outline the density distribution (see Fig. 4a). The Sun is located at $(x, y) = (-8.1, 0)$ kpc. White dashed lines represent longitudes from $-15°$ to $15°$ in steps of $5°$ (straight lines) and distances away from the Sun from 6.1 to 10.1 kpc in steps of 1 kpc (curves).
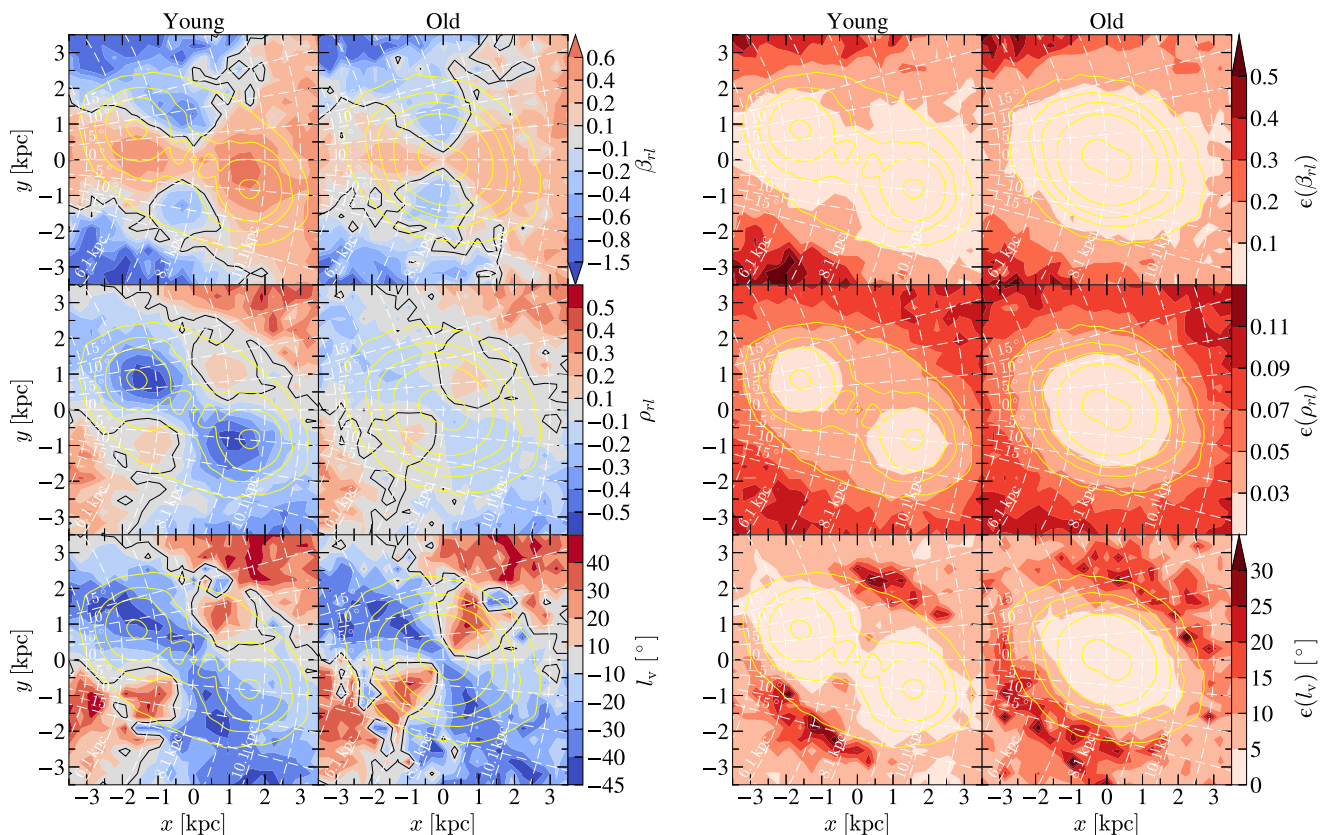
Figure 9. Same as Fig. 7, but now showing, from top to bottom, anisotropy (β_{rl}), correlation (ρ_{rl}), and the vertex deviation, l_v.

spectroscopic survey that observes in the near-infrared H band (1.51–1.70 μm). These wavelengths allow it to collect detailed chemical and kinematic information from the inner Galaxy, piercing through the dust present in the disc. Its main targets are red giant stars from the red giant branch (RGB), the AGB, and the RC, which are luminous tracers present in most stellar populations (Majewski et al. 2013; Zasowski et al. 2017).

We use the cleaned bulge data set from Rojas-Arriagada et al. (2020), comprised of 13 031 APOGEE DR16 stars within $|l| < 11°$, $|b| < 13°$, and $R_{GC} < 3.5$ kpc. We use the radial velocity of the stars computed from the spectroscopic data using the pipeline of Nidever et al. (2015), and the metallicities, [Fe/H], computed using the APOGEE Stellar Parameters and Chemical Abundances Pipeline in García Pérez et al. (2016). We also use the stellar distances computed using the spectrophotometric method in Rojas-Arriagada et al. (2017, 2019) with JHK_s photometric values from Two Micron All-Sky Survey (2MASS; Skrutskie et al. 2006). Their method infers distances and line-of-sight reddening simultaneously through isochrone fitting, using PARSEC isochrones,[3] covering ages from 1 to 13 Gyr and metallicities from -2.2 to $+0.5$ dex. Each star is placed in T_{eff}–log g–metallicity space, and the normalized distance d_{iso} to each model star is computed, accounting for observational errors. Two physical weights are applied: W_{es}, proportional to the mass difference between consecutive model stars along the isochrone, which corrects for oversampling in short-lived phases and assigns greater weight to longer evolutionary stages, and W_{IMF}, proportional to the relative number of stars in each mass interval according to the

initial mass function. The total weight assigned to each model star j is then given by

$$W_j = W_{es} W_{IMF} \exp(-d_{iso}).$$

Weighted estimates of the theoretical absolute magnitudes (M_J, M_H, M_{K_s}) are derived, and by comparing them with the observed 2MASS magnitudes, the distance, and reddening for each star are determined.

We tested our distances against those from the Harris (1996, 2010 edition) catalogue[4] for a sample of globular cluster stars observed by APOGEE, and against *Gaia* DR3 Bayesian distances (Bailer-Jones et al. 2021) for a sample of nearby ($d_{CBJ} < 3$ kpc) APOGEE RGB/AGB stars, selected to be representative of the sample analysed in this work, and whose *Gaia* astrometry have relative parallax errors smaller than 5 per cent. Taking as distance errors the absolute differences between the spectrophotometric and reference distances, we find a median fractional error of 8.3 per cent, with the 25th and 75th percentiles at 3.6 per cent and 16.7 per cent, respectively. Comparing these external errors with the internal errors derived from the isochrone fitting, which indicate the quality of the fit, we find a moderate correlation (Pearson $\rho = 0.52$) and that the internal error underestimates the external error, which indicates that the internal error does not encapsulate all sources of error. This is not an issue because the bootstrap error is the dominant source of error in our results even if we assume a constant fractional distance error of 20 per cent. We analyse the effect of even larger distance uncertainties in Section 7.

---

[3] Available at http://stev.oapd.inaf.it/cgi-bin/cmd

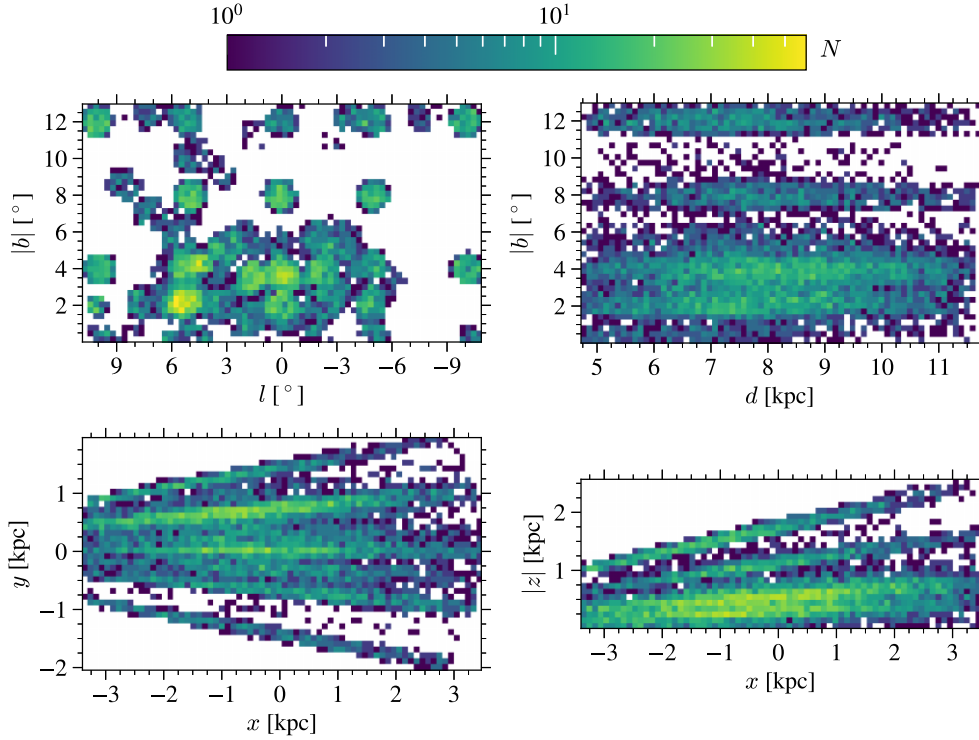[4] Available at https://physics.mcmaster.ca/~harris/mwgc.dat

Figure 10. Distribution of the APOGEE stars in our bulge sample in different spatial representations. The top row shows the distribution in Galactic coordinates, while the bottom row shows Cartesian Galactocentric coordinates. The Sun is located at $(x, y)_\odot = (-8.1, 0)$ kpc. All the heat maps are shown on a log scale, with white indicating bins with no stars.

Rojas-Arriagada et al. (2020) adopted proper motions for the 13 031 star sample from *Gaia* DR2 (Gaia Collaboration 2018). In this study, we update the proper motions with *Gaia* DR3 (Gaia Collaboration 2023). Of the 13 031 stars, 341 do not have a match in *Gaia* DR3. Conversely, 4395 DR2 stars have two or more matches in DR3, and we select as best match the source with the smallest absolute difference in G magnitude. Leaving out 17 problematic sources where the magnitude difference is larger than 1 mag, we are left with 12 673 valid matches. From these, 716 have renormalized unit weight error larger than 1.4, and we removed them. Of the remaining stars, 1422 and 282 are missing [Fe/H] and proper motions, respectively, and we remove them too, leaving us with 10 486 stars.

5.2 The bulge sample

We stick to the $|l| < 11°$, $|b| < 13°$, and $R_{GC} < 3.5$ kpc selections applied by Rojas-Arriagada et al. (2020). We assume a Sun–Galactic Centre distance of $R_0 = 8.1$ kpc (GRAVITY Collaboration 2018; Bobylev & Bajkova 2021), and adopt a Galactocentric Solar velocity $(U, V, W)_\odot = (12.9, 245.6, 7.78)$ km s^{-1} (Drimmel & Poggio 2018; GRAVITY Collaboration 2018; Reid & Brunthaler 2004). We convert all our velocities to the Galactic Standard of Rest using the PYTHON package GALPY (Bovy 2015).

The metallicity values range from -2 to 0.6 dex. We exclude stars with [Fe/H] < -1 dex, which amount to 548 stars (~5.2 per cent), since very metal-poor stars are dominated by the halo (see Lucey et al. 2021), which we are not interested in for this study. After these selections, our final bulge sample contains 9884 stars.

We assume that the bulge of the MW, and that of the model, are symmetric with respect to the mid-plane (Wegg et al. 2015). Thus, we follow Gough-Kelly et al. (2022) and Rojas-Arriagada et al. (2020) in folding the stars across the mid-plane, by projecting $z' = -z$ and $v'_z = -v_z$ for $z < 0$. This allows us to increase the number of stars in each spatial bin, which is particularly useful given the relatively small number of stars in the observational data. Hereafter, we work with $|b|$ and $|z|$. Fig. 10 shows the spatial distribution of the 9884 stars in our bulge sample. The top row shows the view in Galactic coordinates, and the bottom row in Galactocentric Cartesian coordinates. In the top row, we can see that the majority of stars are located at relatively low latitudes, $|b| < 5°$. Most stars were observed by APOGEE in fields of view of $\sim2°$ in diameter, which reach deep into the bulge. We note our APOGEE sample may be biased in distance for different metallicities, as previous studies have suggested that APOGEE star selection is skewed towards metal-poor stars at greater distances, particularly behind the Galactic Centre (Queiroz et al. 2021).

The metallicity distribution of the bulge stars is shown in Fig. 11. There is a clear peak at ~0.4 dex, consistent with component A identified by Ness et al. (2013, their fig. 12), while the rest of the distribution forms a broader cluster around ~ -0.35 dex, consistent with an aggregate of components B and C in Ness et al. (2013). The median metallicity is -0.21 dex. In the kinematic analyses that follow, we define a metal-poor and metal-rich population by splitting the distribution around the median metallicity. This leaves enough stars to perform a broad qualitative comparison with the results from the old and young populations in the model, respectively.
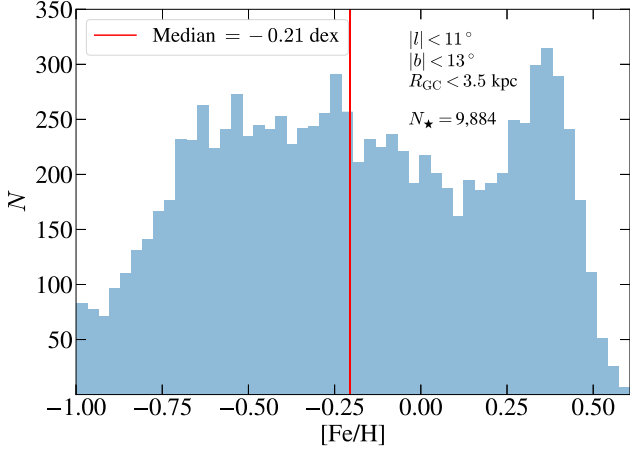
Figure 11. Metallicity distribution function of the APOGEE stars in the bulge region. The vertical line at -0.21 indicates the median metallicity.

We use a Monte Carlo propagation to estimate the effect of individual measurement (distance[5] and proper motions) uncertainties on the velocity ellipse properties for the observational data. We find that the resulting error is negligible relative to the bootstrap error. Therefore, we ignore the measurement uncertainties and, as for the model, we use $B = 500$ bootstrap iterations following equation (5) to estimate the errors on the observational data kinematics.

6 VELOCITY ELLIPSES ALONG THE MINOR AXIS

We now compare the velocity ellipses of the model, and of the APOGEE data, as seen from the Sun (i.e. using heliocentric v_r–v_l velocities) along the bulge minor axis, $|l| < 2°$.

6.1 Variation with latitude

A significant fraction of APOGEE stars is close to $l = 0°$, as can be seen in the top-left panel of Fig. 10. This allows us to explore the variation of the vertex deviation with latitude. We select stars at $|l| < 2°$, $1.5° < |b| < 13°$, and $R_{\rm GC} < 3.5$ kpc. We exclude the $|b| < 1.5°$ region to avoid the potential influence of the nuclear stellar disc. This is a generous cut for the MW (see Sormani et al. 2022), but required by the larger nuclear disc in the model (see Section 3.1). In every panel, we show in a lighter colour the results using a reduced radial cut of $R_{\rm GC} < 2$ kpc.

Given the latitude distribution of the APOGEE stars (see Fig. 10), we divide the $1.5° < |b| < 7°$ region into two bins, each containing an equal number of stars. Additionally, we treat the pencil-beam fields of view centred at $|b| \sim 8°$ and $12°$ as separate bins, resulting in a total of four bins. For the model, instead, we divide the latitude range into eight equally spaced bins. Since the simulation is not a detailed model of the MW, we are only interested in comparing the two data sets qualitatively, so this binning suffices. Fig. 12 shows the results for the anisotropy (top), correlation (middle), and vertex deviation (bottom) as a function of $|b|$. We represent the model by

[5]We propagated the internal distance error which, as discussed in Section 5.1, is only one source of error (isochrone fitting quality). This is the best we can do as it is the only source of errors which we can quantify on a star-by-star basis.
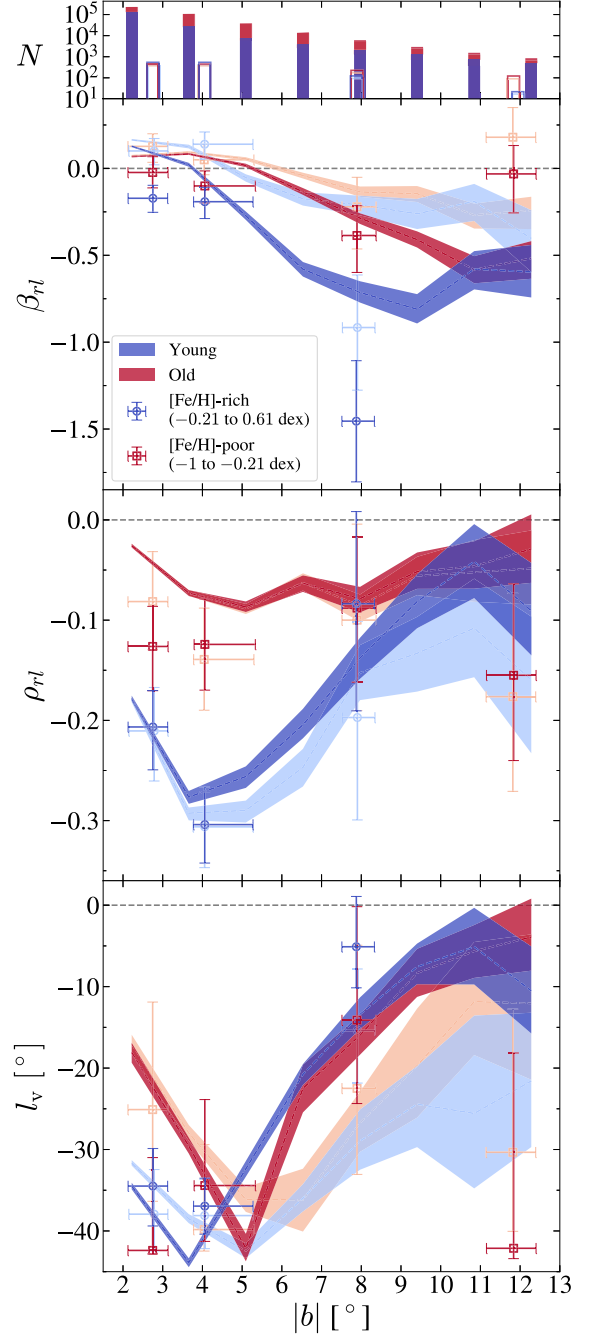

Figure 12. Vertical profiles of anisotropy (top), correlation (middle), and vertex deviation (bottom) along the bulge minor axis, $|l| < 2°$, within $R_{\rm GC} < 3.5$ kpc (darker) and $R_{\rm GC} < 2$ kpc (lighter). The shaded areas show the model, while the data points show the APOGEE data. The number of stars in each bin is shown in the bar plot at the top, with the filled histograms showing the model and the open ones showing the APOGEE data. Only data points with at least 50 stars are shown; there are less than this metal-rich stars in the highest $|b|$ bin, so this bin is excluded.

shaded areas, with the width at each point measuring the 68 per cent bootstrap confidence interval. The observational data are shown as points, placed at the median latitude of the stars in each bin, with kinematic error bars measuring the 68 per cent bootstrap confidence interval, and position error bars measuring half the distance between the point and the bin's edges.

Given the relatively strong positive anisotropy, β_{rl}, of the model in the $x-y$ plane along $l \sim 0°$ across the full depth of the bulge, particularly for the young stars, as seen in the top row of Fig. 9, we might expect to find strong $\beta_{rl} > 0$ in the top panel of Fig. 12. However, the opposite signs of v_l on the near and far sides, seen in Fig. 8, lead to a σ_l almost as large as σ_r at low latitudes, and even larger at higher latitudes. As a result, Fig. 12 shows that the anisotropy of both populations in the model starts small and positive at low latitudes, then crosses the isotropy line (at $|b| \sim 3.7°$ and $\sim 5.3°$ for the young and old populations, respectively), grows stronger and negative beyond that, with a steeper gradient for the young stars, and eventually plateaus at the highest latitudes. The anisotropy of the APOGEE data follows a similar trend, with isotropic values near the plane, growing negative away from it. An exception is the metal-poor data point at the highest latitude, which returns to isotropy. This behaviour might be influenced by contamination from the thick disc (Rojas-Arriagada et al. 2020 found a metallicity distribution consistent with the thick disc at $|b| > 10°$). More data at higher latitudes would help to ascertain the behaviour there. In the model, the effect of changing the radial range to $R_{GC} < 2$ kpc, shown in lighter colour, is a shift of the curves towards $\beta_{rl} = 0$, which is more prominent for the young stars away from the plane. The resulting effect is a flattening of the curves. The anisotropy responds to the change in radius in part because of the broadening caused by the opposite signs of the near and far sides of the galactic rotation, which itself varies radially. For the young stars, the anisotropy changes also due to forbidden velocities (see Section 4.2) becoming more important (as they exist within the inner kpc); the cancellation of the positive and negative sides then leads to a larger σ_r. For the old stars, it is due to σ_l decreasing, given v_l increases radially, so the rotational broadening is reduced. The effect is not as strong close to the plane for both populations because their density is dominated by the hot dense region within $R_{GC} \lesssim 1$ kpc, as shown in Fig. 4(b). The dominance of this inner region reaches further away from the plane for the old stars. The effect of reducing the radial range on the APOGEE data is also an upwards shift of the curve.

The correlation, shown in the middle panel of Fig. 12, is negative for all populations, in both the model and the APOGEE data. The correlation of the young stars in the model and metal-rich ones in the observations exhibit a U-shaped profile at $|b| < 9°$, with the largest (most negative) values in the interval $3° < |b| < 6°$, dropping in magnitude above this. The correlation for the old and metal-poor populations is also negative but smaller in magnitude and flatter, showing less variation with latitude. The effect of the radial change is small to negligible in the correlation, for both the model and the APOGEE stars (except for the metal-rich bin at $|b| \sim 8°$, which, however, has very large error bars). The correlation of the young stars in the model, unlike the old ones, does experience a small ($\lesssim 0.05$) systematic shift towards stronger values except at the lowest latitude.

The bottom panel of Fig. 12 shows the vertical profile of the vertex deviation, l_v. Both the young and old populations in the model exhibit a V-shape below $|b| < 7°$, with magnitude peaks of equal strength (reaching nearly $-45°$). These peaks are located at the points nearest to the $\beta_{rl} = 0$ line (recall that a perfectly isotropic system with negative correlation would have $l_v = -45°$ exactly). Therefore, despite the weaker bar of the old stars, the vertex deviation is robustly capturing its non-axisymmetry, but not the fact that it is weaker than the bar in the young stars. We find similar trends for the metal-rich and metal-poor populations in APOGEE at $|b| < 7°$. The strong correlation and vertex deviation of the metal-rich stars are a clear signature of its non-axisymmetry. As for the model's old population, the near-isotropy of the metal-poor APOGEE stars

leads to their vertex deviation revealing their non-axisymmetry, despite their relatively weak correlation. At the latitude of Baade's Window, $|b| \simeq 4°$, both metal-rich and metal-poor populations have large (negative) l_v values, as is seen in the MW for populations with [Fe/H] > -0.7 dex (Babusiaux et al. 2010; Soto et al. 2007). Beyond $|b| > 7°$, the vertex deviation of all populations diminishes in magnitude, in part because of projection effects. The vertex deviation of the metal-poor stars at the highest latitude is strong but its error is very large. Again a larger sample at higher latitudes would help to elucidate the behaviour there. Using the reduced radial range, $R_{GC} < 2$ kpc, shifts the location of the peak l_v for the young stars, and reduces its amplitude below $|b| = 6°$ in the old stars. In both populations, the amplitude of l_v increases substantially for $|b| > 6°$. There are hints of similar behaviour in the APOGEE data (e.g. $|l_v|$ increases for both metal-rich and metal-poor populations at $|b| = 8°$) but more data are required for the results to be significant.

Overall, we find that, at latitudes $|b| < 7°$, while vertex deviation serves as a blunt tool to detect non-axisymmetry, correlation works best at separating the young and old populations due to their differing bar strengths. This holds for the observations remarkably well too despite their reduced number of stars relative to the model.

6.2 Dependence on age and metallicity

We again select stars on the bulge minor axis, $|l| < 2°$, within $R_{GC} < 3.5$ kpc. We use the same binning in $|b|$ used for the observations in the previous section, both for the observations and for the model. These are $1.5° < |b| < 3.51°$, $3.51° < |b| < 6.6°$, and $7.13° < |b| < 8.85°$. We ignore the highest bin, centred on $|b| \sim 12°$, as it contains too few stars for a meaningful dissection by [Fe/H].

We bin the selected stars by age and metallicity for the model and the APOGEE data, respectively. For the model, we use equal-step bins in age, from 4 to 10 Gyr, with a total of 40, 20, and 10 bins for the lowest, intermediate, and highest latitude cuts, respectively. For the APOGEE data, we use equal-number bins in metallicity, with four bins at the lowest and intermediate latitudes, and three at the highest.

Fig. 13 shows the kinematics of the different metallicity (left) and age (right) bins for the APOGEE data and the model, respectively. We show the age increasing to the left in the panels representing the model, to ease the comparison (which is again purely qualitative). The APOGEE data points are shown at the median of each bin, with kinematic error bars measuring the 68 per cent bootstrap confidence interval, and metallicity error bars indicating half the bin width in each direction. The model values are shown at the mid-point of each bin, and the 68 per cent bootstrap confidence intervals are joined into a shaded region.

From top to bottom, we show β_{rl}, ρ_{rl}, and l_v. The histogram at the very top shows the number of stars in each bin. The star numbers are also written as an inset for the observations. In every panel, we show in a lighter colour the results after reducing the radial cut to $R_{GC} < 2$ kpc. In the following descriptions, we focus on the $R_{GC} < 3.5$ kpc results (saturated colours).

The anisotropy of APOGEE stars and the model follow similar trends. It is negative and strong at the highest latitude (purple), increasing in magnitude towards more metal-rich and younger ages in the MW and the model, respectively. The values at the intermediate (orange) and lowest (green) latitudes are relatively isotropic compared to the highest latitude. At the intermediate latitude, the anisotropy of the model also exhibits a weak gradient with age, starting small and positive for the oldest stars, crossing $\beta_{rl} = 0$ at ~ 9 Gyr, and decreasing towards younger ages. The anisotropy of
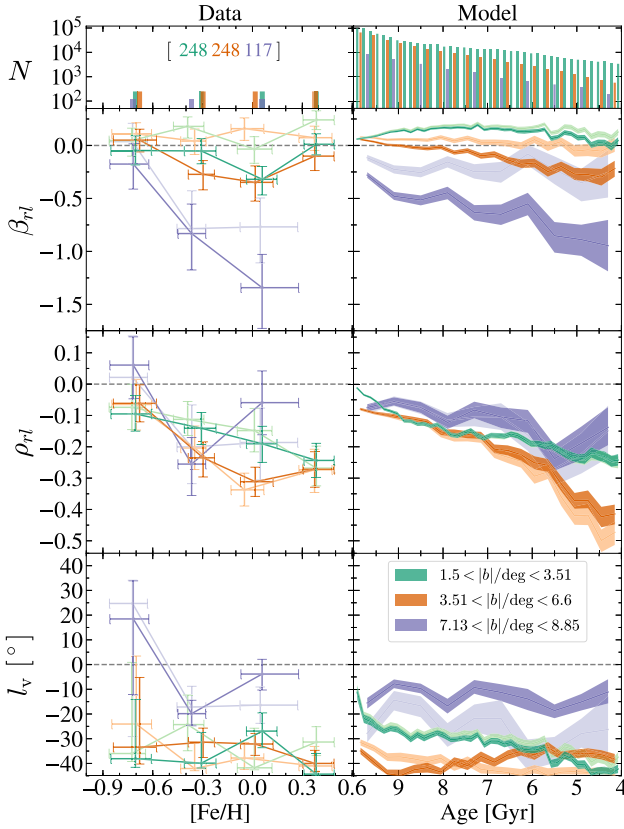
of the observations is more uncertain. The effect of the radial change is relatively small on the ρ_{rl} of both the APOGEE data and the model. In this sense, ρ_{rl} is a robust probe of kinematic fractionation.

The vertex deviation (bottom row) in the model is negative and strong ($\lesssim -30°$) for most ages at the lowest (green) and intermediate (orange) latitudes. As explained in Section 2, vertex deviation is very sensitive to non-zero correlations in isotropic regions. Therefore, even though the oldest stars are significantly less barred than the young ones, their anisotropy is small enough that their vertex deviation robustly reveals their non-axisymmetry. An exception are the stars older than ~9 Gyr at the lowest latitude, whose correlation in this hot region near the plane does drop to small enough values relative to the anisotropy that its vertex deviation eventually drops to ~ $-10°$. The vertex deviation of the model at this lowest latitude is relatively unaffected by a change in the radial range to $R_{GC} < 2$ kpc, resulting in just a slight decrease in magnitude for young stars. At the intermediate latitude, given the radial change affects where the anisotropy crosses the isotropy line, the vertex deviation curve shifts its lowest point from old to young stars. Regardless, the model's vertex deviation is strong and flat at this latitude for all ages. At the highest latitude, the anisotropy of the model is large enough that the vertex deviation is relatively weak, at $l_v = -10°$ to $-20°$ for all populations. Reducing the radial range causes the anisotropy to drop in magnitude for all ages, causing the vertex deviation magnitude to increase to moderate values.

The vertex deviation of the APOGEE stars behaves similar to the model, with negative and relatively strong ($< -25°$) values across all metallicities at the lowest (green) and intermediate (orange) latitudes, for both radial ranges, and more axisymmetric values at the highest (purple) latitude. At the lowest latitudes, even the most metal-poor stars have a very strong vertex deviation despite their small correlation, due to the small β_{rl}. The large error is indicative of a nearly isotropic axisymmetry, especially at the intermediate and highest latitudes.

We have explored the kinematics of the model upon switching from Galactic to Cartesian cuts, and results were relatively unaffected, which means they are not dominated by projection effects.

Additionally, we have recreated the results for the observational data in both Figs 12 and 13 upon applying a 5 per cent cap on the internal (see Section 5.1) fractional distance error, which discards ~19 per cent of the data, and the results remain largely unchanged, with all data points being consistent with the results shown here within a standard deviation.

Following on from the conclusion from the previous section, we find that the correlation is a great tracer of kinematic fractionation, i.e. of varying bar strengths, at low and intermediate latitudes, for both the model and observations. This is in contrast with vertex deviation, which only serves as a blunt test for non-axisymmetry if populations are close to isotropic.

7 EFFECT OF DISTANCE UNCERTAINTIES

We explore the effect of distance errors in the precision and accuracy of the measurements of the velocity ellipses by introducing varying fractional uncertainties in the distance of stars in the model, using Monte Carlo error propagation with 500 repeats, assuming the distance uncertainties are Gaussian.

Fig. 14 shows the variations of the anisotropy (top), correlation (middle), and vertex deviation (bottom) with fractional distance errors from 0.05 to 0.35, for the young (left) and old (right) populations selected within $|l| < 2°$, $3° < |b| < 6°$, and $R_{GC} < 3.5$ kpc.

Figure 13. Kinematics as a function of metallicity (left) and age (right) for the APOGEE data and model, respectively, along the bulge minor axis, $|l| < 2°$. Each panel shows three different latitudes, as indicated in the legend, with saturated and light colours representing results with $R_{GC} < 3.5$ and < 2 kpc, respectively. The number of stars in each bin is shown in the bar plots at the top.

the APOGEE stars at the intermediate latitude also grows slightly negative towards more metal-rich stars, although it then increases back to isotropy at the highest [Fe/H]. At the lowest latitude, β_{rl} is mostly flat across [Fe/H] and age. At this latitude, β_{rl} of the model is mostly positive, crossing the isotropic line at the youngest ages. Switching to $R_{GC} < 2$ kpc leads to an increase in the anisotropy of most populations in the observations and the model, especially at higher latitudes. This is due to an increase in σ_r and a decrease in σ_l. In the model σ_r increases due to the cancellation of the positive and negative sides of the (fast) bar orbits, which dominate the velocity distribution at that reduced radial range (see Fig. 8). As expected, the increase is more prominent for younger stars and at the latitudes where the X-shape is populated (orange and purple). The decrease in σ_l in the model is due to the fact that v_l increases away from $x = 0$, as Fig. 8 shows.

Except for the metal-poor population in the APOGEE stars, the correlation (second row) is negative for all populations in the observations and in the model. The continuum of bar strengths seen in Fig. 2 manifests as a steady increase of the magnitude of the correlation with decreasing age, particularly at the lowest (green) and intermediate (orange) latitudes, with a larger gradient at the latter. The APOGEE data follow a similar trend, with stronger correlation at higher [Fe/H] at these latitudes. The correlation gradient with [Fe/H] is also larger at intermediate latitudes, although it seems to saturate at the highest [Fe/H]. At the highest latitude, the behaviour
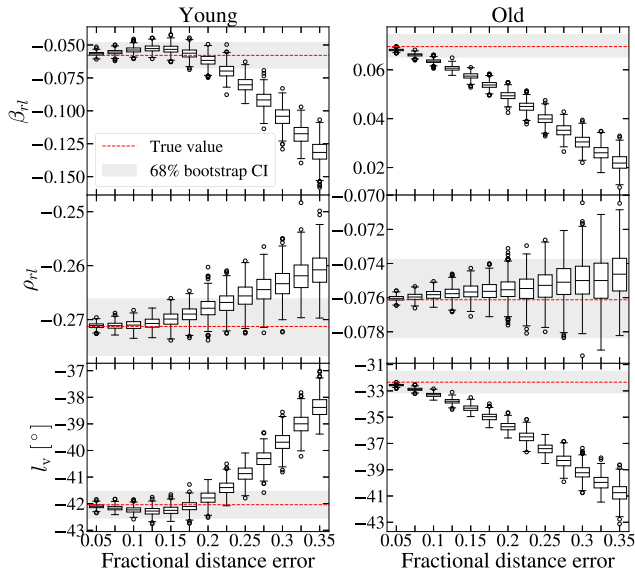
Figure 14. Monte Carlo distributions of anisotropy (top), correlation (middle), and vertex deviation (bottom) for the young (left) and old (right) populations along the bulge minor axis, resulting from increasing levels of fractional distance errors. Each distribution is represented by a boxplot, whose body spans the interquartile range (IQR), containing the central 50 per cent of the data, and is cut through by a solid line at the median. The boxplot whiskers span from $Q1 - 1.5 \cdot IQR$ to $Q3 + 1.5 \cdot IQR$, where $Q1$ and $Q3$ are the 25 per cent and 75 per cent percentile points. Any values beyond the whiskers are shown as individual markers. The (red) dashed lines show the true values of the statistics, with the 68 per cent bootstrap confidence interval shown as a grey shaded region.

The variations of β_{rl} (and consequently l_v) are dominated by relatively small changes in σ_l. After introducing the distance errors, the dependence of v_l on d leads to an increase in σ_l, and consequently a decrease in β_{rl}. The anisotropy magnitude increases for the young population, which already has $\beta_{rl} < 0$, while that of the old, which is initially positive, declines towards zero. This decrease begins immediately for the old population, but for the young once, the 20 per cent error limit is reached. This is due to the differing radial density profiles of these populations. The density of the young population is less peaked at the GC than that of the old (see Fig. 4a), so young stars being lost[6] at the edges of $R_{GC} < 3.5$ kpc after the perturbation represent a more significant fraction of the population. Stars at the edges form part of the tails of the v_l distribution. In addition, more stars with negative v_l than positive are lost because the far and near sides contain mostly negative and positive v_l stars, respectively (see Fig. 8) and further stars experience a stronger distance perturbation. The effect of this asymmetric loss is a decrease in dispersion which is significant enough to counteract the effect of the overall σ_l broadening at the initial fractional errors for the young stars. The correlation magnitude decreases with increasing distance uncertainties, although the relative change is smaller than that in the anisotropy. The increased anisotropy magnitude and weakened correlation leads to a reduction in the magnitude of the vertex deviation of the young stars past the 20 per cent error

---

[6]Stars originally at $R_{GC} > 3.5$ kpc also make it inside the cut after the distance perturbation, but in smaller numbers as the density drops exponentially (see Fig. 4a).

mark. Meanwhile, the anisotropy magnitude drop for the old stars causes its vertex deviation to grow in magnitude, reaching below $-40°$, despite their very small correlation, which again illustrates the sensitivity of vertex deviation to non-zero correlations in isotropic populations.

8 VELOCITY ELLIPSES ACROSS THE BULGE

In this section, we consider the properties of the velocity ellipses for both the model and the APOGEE data across the entire bulge in Galactic coordinates, (l, b). The use of Galactic velocities in fields away from $l = 0°$ introduces projection effects which obscure the interpretation of the vertex deviation values. None the less, provided the line-of-sight distributions of observed stars is fairly representative of the actual ones, it is instructive to compare the model and the APOGEE data. Since the model is not a made-to-measure representation of the MW, we only compare the two qualitatively.

We consider stars within $R_{GC} < 3.5$ kpc. Fig. 15(a) shows the surface density contours of the young (left) and old (right) stars in the model projected onto the (l, b)-plane. The young stars are very dense near the plane and develop an X-shape at $|b| \gtrsim 5°$. The arm of the X-shape at positive longitudes reaches slightly higher than the one at negative longitude, due to perspective. The old stars, instead, are thicker and boxy.

Given the heterogeneous density distribution of stars in the APOGEE data in the (l, b)-plane, as shown in the top-left panel of Fig. 10, we use continuous bins only in the densest region, within $|l|, |b| \lesssim 6°$. For the rest of the stars, we consider each pencil-beam field of view as a single bin. Fig. 15(b) shows the resulting star number in each bin, with individual stars in red. In our analysis we consider only bins containing at least 50 stars.

Given that we are considering the full range of latitudes $|b| < 13°$, down to the Galactic plane, the nuclear disc might affect our results closest to the plane. In the model, we reduce its contribution using our age cuts (the model's nuclear disc is younger than 4 Gyr). Its effect is also small in the observations given most of our stars in our final bins are located at $|b| > 0.5°$ (see top left of Fig. 10), at which point the contamination from it is no larger than 25 per cent according to the modelling of Sormani et al. (2022).

Fig. 16 shows the mean Galactic velocities and their dispersions. For both (a) model and (b) observations, $\langle v_r \rangle$ shows the projection of the clockwise rotation, with a gradient towards decreasing $|b|$ and increasing $|l|$. The contours of the young stars are more pinched than those of the old, exhibiting faster rotation at lower latitudes. The near and far sides of the young stars' forbidden velocities (see Fig. 8) have mostly cancelled out. In the $\langle v_l \rangle$ panels, the rotation of the near and far sides have largely cancelled out, but an asymmetric projection effect remains in both the model and the APOGEE data. This effect is weaker for the old and metal-poor stars than for the young and metal-rich stars, due to projection effects of the near and far ends of the bar.

Both the model and the APOGEE data generally show $\sigma_l > \sigma_r$, and both dispersions grow with decreasing latitudes. The face-on view of the model in Fig. 8 shows that at lines of sight within $|l| \lesssim 5°$, σ_r is larger than σ_l, especially for the young stars, whereas outside σ_l tends to be larger. However, the effect of near- and far-side v_l values cancelling out has led to a larger σ_l than σ_r everywhere except in the very inner region in Fig. 16(a). This is reflected in the isotropic regions near the galactic centre in the first row of Fig. 17(a). Both young and old stars show $\beta_{rl} > 0$ under the isotropic region, as can be seen also in Fig. 13 along the minor axis (green line), and $\beta_{rl} < 0$
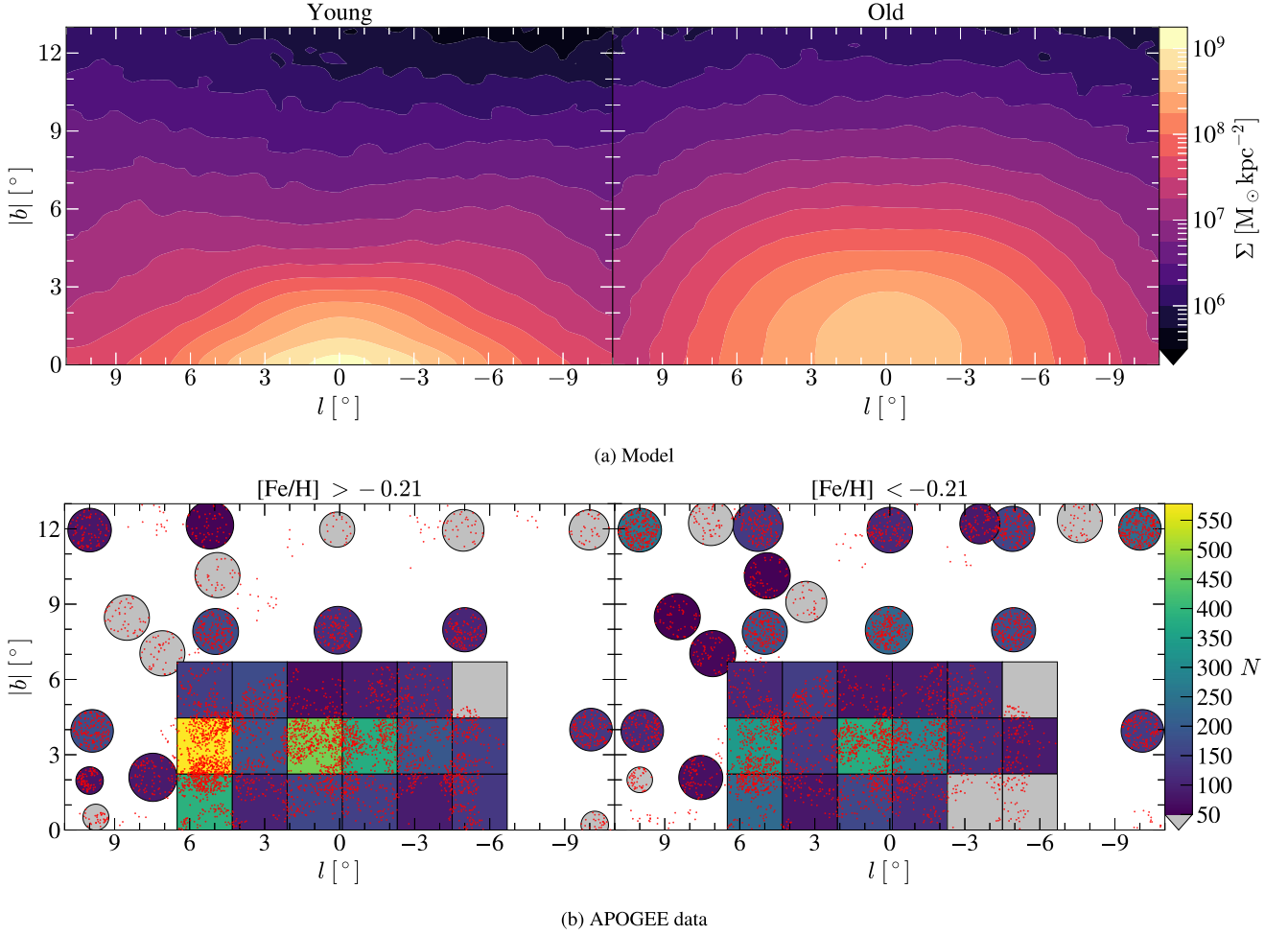
(a) Model

(b) APOGEE data

Figure 15. Spatial distributions of stars in (l, b) space, within $R_{GC} < 3.5$ kpc. Panel (a) shows the surface density in the model for the young (left) and old (right) populations, and panel (b) shows the observed metal-rich (left) and metal-poor (right) star counts in selected bins. Grey bins in panel (b) did not make the $N \geq 50$ cut, so we disregard them in the kinematic maps of Section 8. Individual observed stars are shown as red points.

everywhere else in the (l, b)-plane. There are hints of this behaviour in the APOGEE data, shown in the first row of Fig. 17(b), which are isotropic in the central region near the plane, and have $\beta_{rl} < 0$ everywhere else.

Fig. 9 showed that the $\rho_{rl} < 0$ and $l_v < 0°$ poles of the quadrupoles of the young stars in the model viewed in the (x, y)-plane dominate all lines of sight, and are particularly strong at the X-shape overdensities. The second and third rows of Fig. 17(a) show how this translates to the (l, b)-plane. We have $\rho_{rl} < 0$ and $l_v < 0°$ mostly everywhere. The correlation is particularly strong at intermediate latitudes, and at longitudes where the X-shape resides ($|l| \gtrsim 5°$), particularly on the near side ($l > 0°$) of the bar, which is also where Gough-Kelly et al. (2022) found that the forbidden velocities dominated. We find an agreement with the correlation of the metal-rich population in the observations (second row of Fig. 17b). The vertex deviation of the young population in the model is particularly strong in the inner isotropic regions, particularly at $l < 0°$ where the isotropic region is more extended. The vertex deviation of the metal-rich stars (third row of Fig. 17b) follows a similar trend, with strong negative values in the inner region.

The correlations of the old stars in the model and of the metal-poor ones in the APOGEE data are also negative mostly everywhere,

but the values are smaller compared to the young and metal-rich stars. Despite this, given the isotropy in the inner region, the vertex deviation of the old population in the model is strong at $|b| \sim 4°$ on either side of the bulge minor axis. The vertex deviation of the metal-poor stars in APOGEE also shows some strong values in the inner region. However, some bins also show high vertex deviation errors, indicative of isotropic axisymmetry. The metal-poor population has fewer stars in the $|b| < 6°$ region than the metal-rich (see Fig. 15b); more stars with future surveys would help ascertain the behaviour there.

9 DISCUSSION

We have used the high-resolution N-body $+$ SPH simulation from Debattista et al. (2017) to study age-dependent kinematic trends across the bulge. In this model, all stars form out of cooling gas, which results in correlations between the ages of stars and their kinematics. Once the bar forms, the stellar populations separate through kinematic fractionation, which leads to the young populations being strongly barred and X-shaped, while the older populations are weakly barred and boxy. The model, which has evolved purely in isolation,
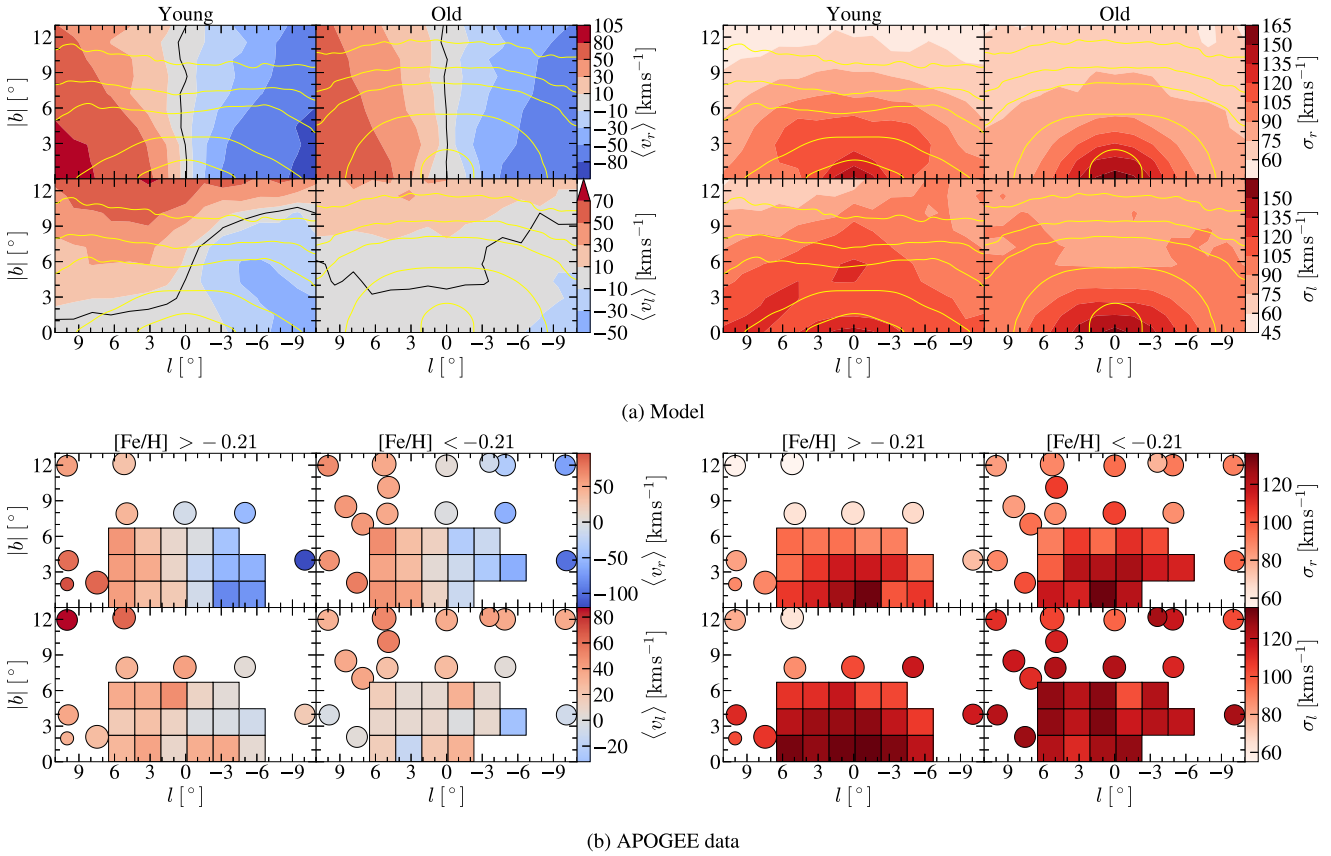
(a) Model

(b) APOGEE data

Figure 16. Kinematic maps in Galactic coordinates of the Galactic velocities for (a) the model (see Fig. 8 for the face-on view) and (b) the APOGEE data, showing the mean velocities (left) and the dispersions (right). In the maps for the model, the black contours follow values of zero, while the yellow contours outline the density distribution. In each block, the top and bottom rows show radial and tangential motions, while the left and right columns correspond to young/metal-rich and old/metal-poor stars.

is able to explain trends observed in the velocity ellipses of the MW's bulge.

9.1 Dependence of bar strength on [Fe/H] in the MW

Fig. 13 shows the variation of the correlation, ρ_{rl}, with [Fe/H] in the APOGEE DR16 + *Gaia* DR3 data. At $3° < |b| < 6°$, which our model suggests is the best location for tracing the bar's amplitude, the overall trend appears to be a rise in $|\rho_{rl}|$ (becoming increasingly negative) from zero at [Fe/H] ~ -0.7 dex until [Fe/H] ~ 0. This is the first indication that the strength of the bar in the MW varies smoothly with [Fe/H], as predicted by the model due to kinematic fractionation. At the highest [Fe/H] bin, ρ_{rl} appears to plateau or even decline in amplitude slightly. However, the number of such metal-rich stars is low, so the uncertainty is larger than the depth of the minimum. Better observations will help establish whether the growing trend extends to the most [Fe/H]-rich stars.

9.2 Future prospects

Future observations, particularly with next-generation surveys like MOONS (Gonzalez et al. 2020), 4MOST (Bensby et al. 2019), and Vera Rubin Observatory Legacy Survey of Space and Time (VRO-LSST), will provide a much more detailed and comprehensive picture of the Galactic bulge. The high multiplexing capabilities of Multi Object Optical and Near-Infrared Spectrograph (MOONS; Cirasuolo

et al. 2020) and 4-metre Multi-Object Spectroscopic Telescope (4MOST) will enable large-scale spectroscopic surveys of the bulge, providing valuable kinematic data for different stellar populations that can be compared with our model predictions. In particular, a Galactic Guaranteed Time Observations survey (Gonzalez et al. 2020) will devote 100 nights to map the central regions of the MW and some of its satellites. The $|l| \leq 10°$ and $|b| \leq 4°$ region (in addition to the adjacent disc at negative longitudes) will be sampled with a very dense and contiguous grid of fields (about 1000 targets across 25 arcmin FoV) obtaining spectra in the *RI*, *YJ*, and *H* bands. The observations in the high-resolution *H* band will pierce through the dust in the Galactic mid-plane and sample this key region with an unprecedented sample size of about 0.5 million RC stars. This data will be complemented with NIR photometry and proper motions from the Vista Variables in the Via Lactea (VVV) survey (Saito et al. 2012). Although the precision of the proper motions will not be as high as future *Gaia* data releases, they will be available for these highly reddened regions where *Gaia* observations are shallower. These data sets combined will provide the unique opportunity to examine in detail, and with statistically significant samples, the kinematic signatures described here. They will be instrumental to settle the debate concerning the origin and evolution of the bulge stellar populations. In addition, VRT will provide photometric data for billions of stars, helping to identify target stars for follow-up spectroscopic observations and enabling detailed structural mapping
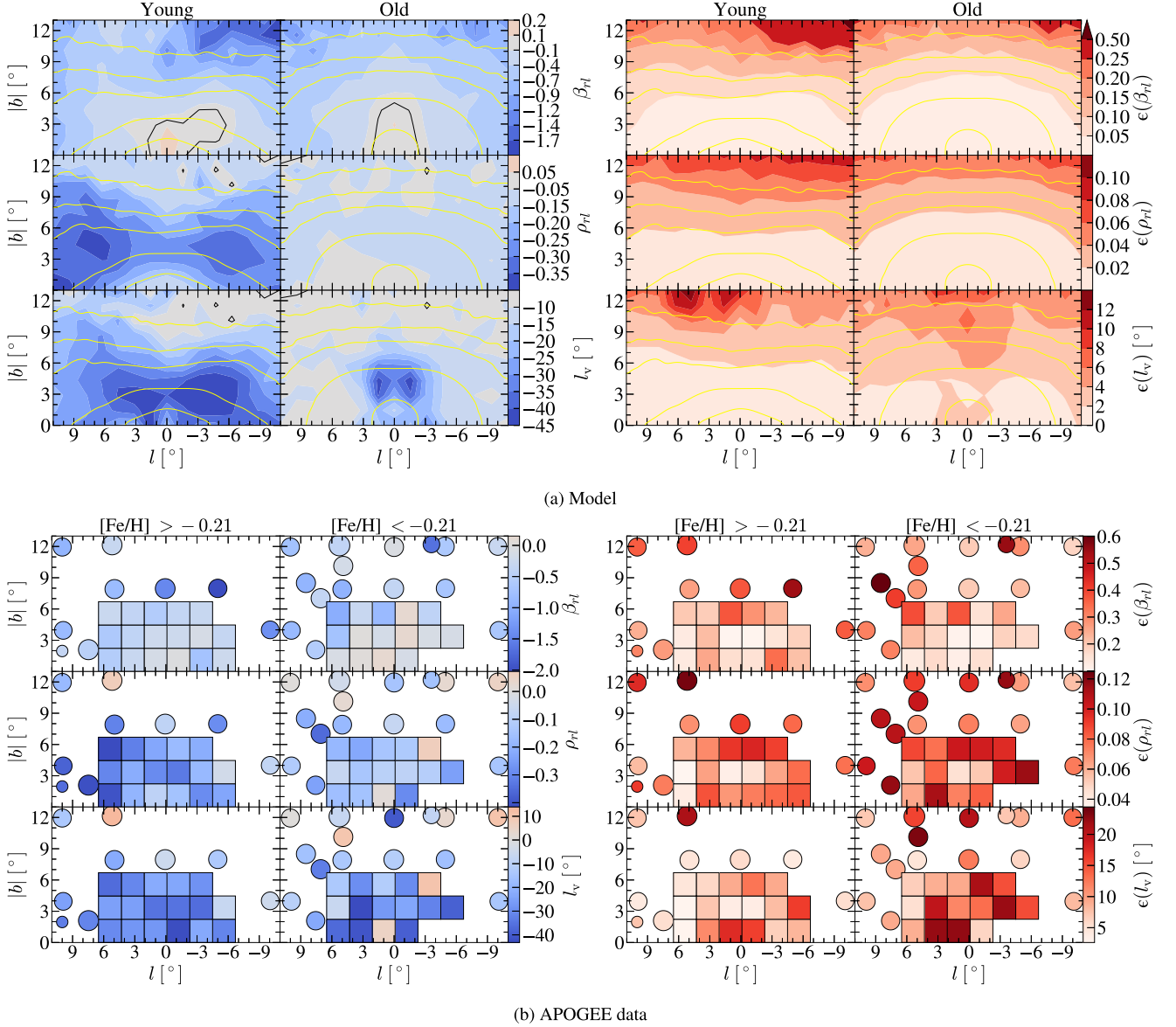
(a) Model

(b) APOGEE data

Figure 17. Same as Fig. 16, but now showing the anisotropy, correlation, and vertex deviation.

of the bulge. This will allow for a more comprehensive distribution of observational data across the bulge. Such data will be invaluable for refining our understanding of the kinematics of the Galactic bulge.

Our results provide guidance for future work using velocity ellipses at the bulge. The velocity ellipses can be parametrized by the dimensionless in-plane anisotropy, β_{rl}, correlation, ρ_{rl}, and vertex deviation, l_v (though these are not independent, see Appendix A4). Non-zero vertex deviation has been used as a tracer of non-axisymmetric structure; however, when $\beta_{rl} \sim 0$, i.e. when $\sigma_l \sim \sigma_r$, then l_v generally takes values near the maximum of $\pm 45°$. The region around $|b| \sim 3°$ to $6°$ (see Figs 12 and 13) in both the model and APOGEE suffers from this effect when all stars are considered. The relative constancy of l_v across [Fe/H], only dropping suddenly to $l_v \sim 0°$ at [Fe/H] ~ -0.7 dex (Soto et al. 2007), has in the past been interpreted as evidence that a metal-poor unbarred, and probably accreted, population is present, alongside a population with a relatively uniformly strong bar. However, the lack of variation in

the model's l_v as a function of age, seen in Fig. 13, does not reflect a constancy in bar strength: Debattista et al. (2017) showed that the bar strength varies with age in the same model as we use here. Likewise, Debattista et al. (2019) using a cosmological simulation from the FIRE-2 suite (Hopkins et al. 2018), showed that even though the bar strength changed quite substantially for stars between 1 to 9 Gyr old, l_v was constant throughout this age range. Additionally, in Appendix B, we compare the model presented in this paper with two additional models, containing a weaker bar and an oval respectively, and find an equally large l_v peak amplitude for the young populations in the strong and weaker bar models. We conclude that l_v is actually a poor tracer of bar strength.

On the other hand, β_{rl} is very sensitive to the distance range employed, and is in general weakly varying with age and [Fe/H] for stars at $|b| \lesssim 4°$. Instead, the correlation, ρ_{rl}, changes monotonically with age and bar strength. The variation of ρ_{rl} in the APOGEE data increases monotonically in amplitude with [Fe/H] at $1.5° < |b| <$

$3.5°$, but at $3.5° < |b| < 6.6°$ it may flatten at the highest [Fe/H], though it is still consistent with increasing amplitude within the errors. Moreover, ρ_{rl} is only very weakly dependent on the distance range used (which is evident in the model and confirmed by the APOGEE data), which makes it robust to distance uncertainties in observational data. We have also verified that selecting stars based on their heliocentric distance, such as $6.1 < d/\mathrm{kpc} < 10.1$, still changes ρ_{rl} mildly. This is also true if we use distance cuts which are asymmetric (e.g. $7.1 < d/\mathrm{kpc} < 10.1$ and $6.1 < d/\mathrm{kpc} < 9.1$). Thus, observational selection functions that are metallicity dependent do not severely affect the comparison of ρ_{rl} for different populations.

In addition, in Appendix C, we test the main assumption of the bootstrap and show that it breaks for vertex deviation for sample sizes smaller than $n \lesssim 300$, and that l_v exhibits biases (which is the average difference between the value of a population and of samples extracted from it) as large as $10°$–$20°$. On the contrary, the correlation is well-behaved, with a standard error decreasing as n^{-2} and no bias even at sample sizes as small as $n = 50$.

Thus, we propose that future studies should devote more attention to measuring the correlation, ρ_{rl}, to explore whether the MW's bar amplitude is a continuous function of [Fe/H]. The region around $|b| \sim 3° - 6°$ along the bulge's minor axis is particularly promising as a location for exploring the variations with [Fe/H]. At larger heights, the mixing between Galactocentric radial and vertical motions reduces the strength of the correlation and are therefore not ideal for probing the variation of the bar's strength.

9.3 Summary

We have computed the kinematics of a young, strongly barred, and peanut-shaped population, and an old, weakly barred, and boxy population in a model of a barred galaxy evolving purely secularly. The old stars have nearly axisymmetric velocities, while large bar streaming motions are present in the young population, as previously shown by Gough-Kelly et al. (2022). Here, we have focused in particular on the in-plane anisotropy, correlation, and vertex deviation of the velocity ellipses of the different populations. We have compared these results with data from APOGEE DR16 cross-matched with *Gaia* DR3. Our main results are as follows:

(i) The dispersions of the galactocentric cylindrical velocities in the model exhibit quadrupoles in the (x, y) plane. The σ_R quadrupole is elongated along the bar direction while that in σ_ϕ is perpendicular to it. The overall velocity ellipse has a quadrupole in the anisotropy, $\beta_{R\phi}$, which is also aligned with the bar, while the quadrupoles in the correlation, $\rho_{R\phi}$, and the vertex deviation, $l_v^{R\phi}$, are rotated by $45°$ ahead of the bar. All of these quadrupoles are stronger in the young population than in the old one. (See Figs 5 and 7.)

(ii) Switching to heliocentric velocities, v_r and v_l, the (x, y) maps of the model's resulting dispersions, σ_r and σ_l, appear as perspective-distorted versions of the intrinsic σ_R and σ_ϕ maps, with comparable peaks and quadrupoles. While the (x, y)-maps of anisotropy, correlation, and vertex deviations retain strong quadrupoles in heliocentric velocities, their orientations are quite different to those of the galactocentric cylindrical velocities. The quadrupoles in the anisotropy and in the correlation in the old population are weaker than in the young one, whereas those in the vertex deviation are comparable. The positive poles of the anisotropy quadrupole and the negative poles of the correlation and vertex deviation quadrupoles are bridged together for the young stars, while the bridge is much shallower in the old stars. This bridging is what gives rise to the negative correlation

and vertex deviation when one looks along $l = 0°$. (See Figs 8 and 9.)

(iii) Along the minor axis ($|l| \lesssim 2°$), the APOGEE data and the model follow similar trends with $|b|$ in β_{rl}, ρ_{rl}, and l_v. In particular, we find isotropic β_{rl} values below $|b| \lesssim 5°$, and $\beta_{rl} < 0$ beyond that both in the [Fe/H]-rich/young strongly barred population, which has the larger β_{rl} amplitudes, and in the [Fe/H]-poor/old weakly barred one. The ρ_{rl} amplitudes are larger in the [Fe/H]-rich/young population than in the [Fe/H]-poor/old one, with a negative U-shaped vertical profile in both. In spite of these differences, l_v reaches a strong peak amplitude, near $-45°$, within $|b| < 6°$ for the two populations. This highlights that l_v is a blunt probe of non-axisymmetry when populations are isotropic, unlike ρ_{rl} which does vary with bar strength. Moreover, changing the distance selection function changes β_{rl} and l_v, but has only a weak effect on ρ_{rl}. (See Fig. 12.)

(iv) Along the minor axis, at fixed $|b|$, β_{rl}, and ρ_{rl} vary with [Fe/H] (APOGEE data) and age (model) but l_v is relatively constant. The increasing amplitude of ρ_{rl} with [Fe/H] in the APOGEE data suggest that in the MW, the bar amplitude varies continuously, as predicted by kinematic fractionation, rather than being constant above some metallicity. (See Fig. 13.)

(v) We simulated line-of-sight distance uncertainties in the model along the bulge minor axis and found that, in the young stars, the effect is negligible in the anisotropy, correlation and vertex deviation for fractional errors less than 20 per cent. For the old stars, the amplitude of the anisotropy decreases and, as a consequence, the vertex deviation reaches values as strong as in the young population. The correlation is the least affected in both populations, barely deviating from the original value, even with distance uncertainties as large as 35 per cent. (See Fig. 14.)

(vi) Projected on to the full (l, b)-plane, the correlation is mostly negative everywhere for the young and old stars. However, while it is uniformly weak for the old stars, the correlation of the young is stronger, particularly on the near side of the bar ($l > 0°$), and exhibits two strong peaks near the arms of the X-shape. The vertex deviation is strong and negative for the young population, with two peaks closer to the minor axis compared to those of the correlation. The old stars have weak vertex deviation except very near the minor axis, where $\beta_{rl} \sim 0$, exhibiting a small-scale but strong double-peak. The maps of the APOGEE data for the [Fe/H]-rich and [Fe/H]-poor stars share some of these properties, including a larger amplitude ρ_{rl} in [Fe/H]-rich stars, with stronger values at $l > 0°$, and relatively isotropic distributions in both populations near the Galactic Centre. However, more data are required to confirm these trends. (See Fig. 17.)

(vii) We have highlighted the promise that measurements of the correlation between v_r and v_l along the bulge minor axis hold. The amplitude of the correlation of a given population increases with bar strength, is robust to variations in the distances probed, and is detectably varying in the MW. We have shown that the vertex deviation is instead a poor tracer of bar strength.

DATA AVAILABILITY

The simulation data underlying this article will be shared on reasonable request. The APOGEE DR16 data are publicly available at https://live-sdss4org-dr16.pantheonsite.io/irspec and *Gaia* DR3 data are publicly available at https://www.cosmos.esa.int/web/gaia/data-release-3.

REFERENCES

Babusiaux C. et al., 2010, A&A, 519, A77
Babusiaux C., 2016, Publ. Astron. Soc. Aust., 33, e026
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, AJ, 161, 147
Bekki K., Tsujimoto T., 2011, MNRAS, 416, L60
Bensby T. et al., 2019, The Messenger, 175, 35
Binney J., Merrifield M., 1998, Galactic Astronomy. Princeton University Press, Princeton, NJ USA
Blanton M. R. et al., 2017, AJ, 154, 28

Bobylev V. V., Bajkova A. T., 2021, Astron. Rep., 65, 498
Bovy J., 2015, ApJS, 216, 29
Bovy J., Leung H. W., Hunt J. A. S., Mackereth J. T., García-Hernández D. A., Roman-Lopes A., 2019, MNRAS, 490, 4740
Brown W. R., Anderson J., Gnedin O. Y., Bond H. E., Geller M. J., Kenyon S. J., Livio M., 2010, ApJ, 719, L23
Büdenbender A., van de Ven G., Watkins L. L., 2015, MNRAS, 452, 956
Calamida A. et al., 2014, ApJ, 790, 7
Cirasuolo M. et al., 2020, The Messenger, 180, 10
Clarkson W. et al., 2008, ApJ, 684, 1110
Clarkson W. I. et al., 2011, ApJ, 735, 37
Cole D. R., Debattista V. P., Erwin P., Earp S. W. F., Roskar R., 2014, MNRAS, 445, 3352
Debattista V. P. et al., 2023, ApJ, 946, 118
Debattista V. P., Gonzalez O. A., Sanderson R. E., El-Badry K., Garrison-Kimmel S., Wetzel A., Faucher-Giguère C.-A., Hopkins P. F., 2019, MNRAS, 485, 5073
Debattista V. P., Ness M., Earp S. W. F., Cole D. R., 2015, ApJ, 812, L16
Debattista V. P., Ness M., Gonzalez O. A., Freeman K., Zoccali M., Minniti D., 2017, MNRAS, 469, 1587
Dehnen W., Binney J., 1998, MNRAS, 294, 429 https://doi.org/10.1046/j.1365-8711.1998.01282.x
Dékány I., Minniti D., Catelan M., Zoccali M., Saito R. K., Hempel M., Gonzalez O. A., 2013, ApJ, 776, L19
Di Matteo P. et al., 2015, A&A, 577, A1
Di Matteo P., Fragkoudi F., Khoperskov S., Ciambur B., Haywood M., Combes F., Gómez A., 2019, A&A, 628, A11
Drimmel R., Poggio E., 2018, Res. Notes AAS, 2, 210
Eisenstein D. J. et al., 2011, AJ, 142, 72
Ferreras I., Wyse R. F. G., Silk J., 2003, MNRAS, 345, 1381
Fragkoudi F. et al., 2020, MNRAS, 494, 5936
Fragkoudi F., Di Matteo P., Haywood M., Gómez A., Combes F., Katz D., Semelin B., 2017a, A&A, 606, A47
Fragkoudi F., Di Matteo P., Haywood M., Khoperskov S., Gomez A., Schultheis M., Combes F., Semelin B., 2017b, A&A, 607, L4
Fragkoudi F., Di Matteo P., Haywood M., Schultheis M., Khoperskov S., Gómez A., Combes F., 2018, A&A, 616, A180
Gaia Collaboration, 2018, A&A, 616, A1
Gaia Collaboration, 2023, A&A, 674, A1
García Pérez A. E. et al., 2016, AJ, 151, 144
Gardner E., Debattista V. P., Robin A. C., Vásquez S., Zoccali M., 2014, MNRAS, 438, 3275
Gnambs T., 2023, Collabra: Psychol., 9, 87615
Gonzalez O. A. et al., 2016, A&A, 591, A7
Gonzalez O. A. et al., 2020, The Messenger, 180, 18
Gonzalez O. A., Rejkuba M., Minniti D., Zoccali M., Valenti E., Saito R. K., 2011, A&A, 534, L14
Gough-Kelly S., Debattista V. P., Clarkson W. I., Gonzalez O. A., Anderson S. R., Gennaro M., Calamida A., Sahu K. C., 2022, MNRAS, 509, 4829
Governato F. et al., 2010, Nature, 463, 203
GRAVITY Collaboration, 2018, A&A, 615, L15
Harris W. E., 1996, AJ, 112, 1487
Hesterberg T. C., 2015, Am. Stat., 69, 371
Hopkins P. F. et al., 2018, MNRAS, 480, 800
Hopkins P. F., 2015, MNRAS, 450, 53
ichi Tadaki K. et al., 2017, ApJ, 834, 135
Johnson C. I. et al., 2022, MNRAS, 515, 1469
Johnson C. I., Rich R. M., Fulbright J. P., Valenti E., McWilliam A., 2011, ApJ, 732, 108
Johnson C. I., Rich R. M., Kobayashi C., Kunder A., Pilachowski C. A., Koch A., de Propris R., 2013, ApJ, 765, 157
Kuijken K., Rich R. M., 2002, AJ, 124, 2054
Lucey M. et al., 2021, MNRAS, 501, 5981
Majewski S. R. et al., 2013, ApJ, 777, L13
Majewski S. R. et al., 2017, AJ, 154, 94
McWilliam A., Zoccali M., 2010, ApJ, 724, 1491
Ness M. et al., 2012, ApJ, 756, 22
Ness M. et al., 2013, MNRAS, 430, 836

Ness M., Debattista V. P., Bensby T., Feltzing S., Roškar R., Cole D. R., Johnson J. A., Freeman K., 2014, ApJ, 787, L19

Nidever D. L. et al., 2015, AJ, 150, 173

Ortolani S., Renzini A., Gilmozzi R., Marconi G., Barbuy B., Bica E., Rich R. M., 1995, Nature, 377, 701

Qin W., Nataf D. M., Zakamska N., Wood P. R., Casagrande L., 2018, ApJ, 865, 47

Queiroz A. B. A. et al., 2021, A&A, 656, A156

Raiteri C. M., Villata M., Navarro J. F., 1996, A&A, 315, 105

Reid M. J., Brunthaler A., 2004, ApJ, 616, 872

Renzini A. et al., 2018, ApJ, 863, 16

Roca-Fàbrega S., Antoja T., Figueras F., Valenzuela O., Romero-Gómez M., Pichardo B., 2014, MNRAS, 440, 1950

Rojas-Arriagada A. et al., 2014, A&A, 569, A103

Rojas-Arriagada A. et al., 2017, A&A, 601, A140

Rojas-Arriagada A. et al., 2020, MNRAS, 499, 1037

Rojas-Arriagada A., Zoccali M., Schultheis M., Recio-Blanco A., Zasowski G., Minniti D., Jönsson H., Cohen R. E., 2019, A&A, 626, A16

Sahu K. C. et al., 2006, Nature, 443, 534

Saito R. K. et al., 2012, A&A, 537, A107

Saito R. K., Zoccali M., McWilliam A., Minniti D., Gonzalez O. A., Hill V., 2011, AJ, 142, 76

Schultheis M. et al., 2017, A&A, 600, A14

Sellwood J. A., Gerhard O., 2020, MNRAS, 495, 3175

Shen J., Rich R. M., Kormendy J., Howard C. D., Propris R. D., Kunder A., 2010, ApJ, 720, L72

Simion I. T., Shen J., Koposov S. E., Ness M., Freeman K., Bland-Hawthorn J., Lewis G. F., 2021, MNRAS, 502, 1740

Sit T., Ness M. K., 2020, ApJ, 900, 4

Skrutskie M. F. et al., 2006, AJ, 131, 1163

Smith M. C., Evans N. W., An J. H., 2009, ApJ, 698, 1110

Sormani M. C. et al., 2022, MNRAS, 512, 1857

Soto M., Kuijken K., Rich R. M., 2012, A&A, 540, A48

Soto M., Rich R. M., Kuijken K., 2007, ApJ, 665, L31

Stinson G., Seth A., Katz N., Wadsley J., Governato F., Quinn T., 2006, MNRAS, 373, 1074

Surot F. et al., 2019, A&A, 629, A1

Tacchella S. et al., 2015, Science, 348, 314

Thielemann F.-K., Nomoto K., Yokoi K., 1986, A&A, 158, 17

Uttenthaler S., Schultheis M., Nataf D. M., Robin A. C., Lebzelter T., Chen B., 2012, A&A, 546, A57

Valenti E., Zoccali M., Renzini A., Brown T. M., Gonzalez O. A., Minniti D., Debattista V. P., Mayer L., 2013, A&A, 559, A98

Vásquez S. et al., 2013, A&A, 555, A91

Wadsley J. W., Stadel J., Quinn T., 2004, New Astron., 9, 137

Wegg C., Gerhard O., 2013, MNRAS, 435, 1874

Wegg C., Gerhard O., Portail M., 2015, MNRAS, 450, 4050

Weidemann V., 1987, A&A, 188, 74

Williams A. A. et al., 2016, ApJ, 824, L29

Zasowski G. et al., 2017, AJ, 154, 198

Zhao H., Spergel D. N., Rich R. M., 1994, AJ, 108, 2154

Zoccali M. et al., 2003, A&A, 399, 931

Zoccali M., Hill V., Lecureur A., Barbuy B., Renzini A., Minniti D., Gómez A., Ortolani S., 2008, A&A, 486, 177

APPENDIX A: VERTEX DEVIATION

Vertex deviation is computed in two different ways in the literature, causing some confusion in its interpretation. For the benefit of future studies, in this appendix we review the two definitions and show the difference between them.

A1 Derivation

The velocity dispersion tensor, equation (1), in 2D can be expressed as the covariance matrix (see appendix in Smith, Evans & An 2009),

C, which contains the ith and jth variances along the diagonal, and the covariance in the off-diagonal terms. Given one of the velocity ellipse's principal axes, $\boldsymbol{v}$, with dispersion σ along its direction, the eigenvalue problem can be set up as $C\boldsymbol{v} = \sigma\boldsymbol{v}$, or

$$\begin{pmatrix} \sigma_{ii}^2 & \sigma_{ij}^2 \\ \sigma_{ij}^2 & \sigma_{jj}^2 \end{pmatrix} \begin{pmatrix} v_i \\ v_j \end{pmatrix} = \sigma \begin{pmatrix} v_i \\ v_j \end{pmatrix}, \tag{A1}$$

which is a system of equations that can be combined to yield

$$\frac{\sigma_{jj}^2 - \sigma_{ii}^2}{\sigma_{ij}^2} = \frac{v_j}{v_i} - \frac{v_i}{v_j}. \tag{A2}$$

Calling l_v the angle that $\boldsymbol{v}$ makes with axis i, and ξ its complementary angle, we can rewrite the equality above as

$$\begin{aligned} \frac{\sigma_{jj}^2 - \sigma_{ii}^2}{\sigma_{ij}^2} &= \tan l_\mathrm{v} - \tan \xi \\ &= (1 + \tan l_\mathrm{v} \tan \xi) \tan(l_\mathrm{v} - \xi) \\ &= 2\tan(2l_\mathrm{v} - \pi/2) \end{aligned} \tag{A3}$$

$$= -2\cot(2l_\mathrm{v}), \tag{A4}$$

having used the trigonometric identity for the tangent of a difference. Equation (A3) follows from writing $\tan l_\mathrm{v} \tan \xi = 1$ and $\xi = \pi/2 - l_\mathrm{v}$.[7] Lastly, we used the trigonometric relations between complementary angles to write equation (A4), which we can recast as

$$\tan(2l_\mathrm{v}) = \frac{2\sigma_{ij}^2}{\sigma_{ii}^2 - \sigma_{jj}^2}. \tag{A5}$$

Our derivation is an alternative to the one provided by Binney & Merrifield (1998, p. 630), which is the one typically referenced, in which they reach the same result but along the way (equations 10.15 and 10.16) introduce a negative sign, which we did not need here.

A2 Vertex deviation definitions

In equation (A2), we absorbed the eigenvalue σ, which means equation (A5) holds for both eigenvectors of C.

It is in the calculation and interpretation of l_v from equation (A5) that authors differ. One way to compute it, as used by several studies (e.g. Dehnen & Binney 1998; Büdenbender, van de Ven & Watkins 2015; Roca-Fàbrega et al. 2014; Smith et al. 2009; Zhao et al. 1994) is

$$\tilde{l}_\mathrm{v} = \frac{1}{2}\mathrm{atan2}\left(2\sigma_{ij}^2, \sigma_{ii}^2 - \sigma_{jj}^2\right), \tag{A6}$$

where

$$\mathrm{atan2}(y, x) = \begin{cases} \arctan(y/x), & \text{if } x \geq 0 \\ \arctan(y/x) + \mathrm{sign}(y)\pi & \text{if } x < 0 \end{cases}. \tag{A7}$$

Equation (A7) is equivalent to writing $\mathrm{atan2}(y, x) \equiv \mathrm{Arg}(x + iy)$, where Arg is the principal argument of the complex number $x + iy$. Selecting the principal branch gives values in the range $(-180, 180]^\circ$. Therefore, equation (A6) limits $\tilde{l}_\mathrm{v}$ to the range $(-90, 90]^\circ$, and measures the angle that the semimajor axis of the velocity ellipse makes with the positive side of the horizontal axis direction, $\hat{\imath}$.

---

[7]To be precise, if $l_\mathrm{v} < 0^\circ$ we should write $\xi = -(\pi/2 + l_\mathrm{v})$. This would give $l_\mathrm{v} - \xi = 2l_\mathrm{v} + \pi/2$, changing equation (A3). However, given $\tan(x) = \tan(x \pm \pi)$, equation (A4) would still hold.
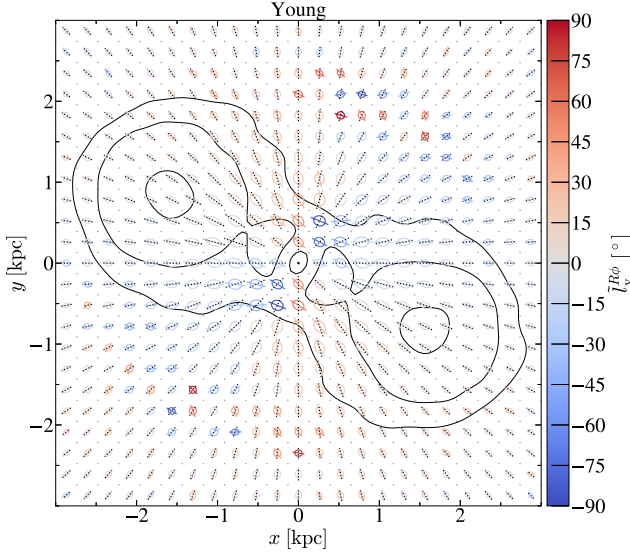
Figure A1. Velocity ellipses of the young population, as previously shown in the left panel of Fig. 6, this time colour-coded by the vertex deviation computed using equation (A6), whose values live in the range $(-90, 90]°$.

We have used a tilde in equation (A6) to differentiate it from the other definition (e.g. Babusiaux et al. 2010; Debattista et al. 2019; Simion et al. 2021; Soto et al. 2007, 2012), which we have used throughout this paper. It is given by equation (2), which involves taking the absolute value of the difference between the dispersions. Taking the absolute value removes the ambiguity in the calculation of l_v from equation (A5), allowing us to apply the standard arctan function. The resulting values live in the range $[-45, 45]°$, and measure the angle that the semimajor axis of the velocity ellipse makes with the coordinate axis it is closest to, with the same sign as the correlation (and as $\tilde{l}_v$).

We can convert between the two definitions using

$$l_v = \begin{cases} \tilde{l}_v, & \text{if } \sigma_{ii}^2 \geq \sigma_{jj}^2 \\ \text{sign}(\tilde{l}_v)(90° - |\tilde{l}_v|) & \text{if } \sigma_{ii}^2 < \sigma_{jj}^2 \end{cases}, \quad (A8)$$

and vice-versa. In words, they are the same except when the anisotropy (equation 4) is negative, in which case $|\tilde{l}_v| > 45°$ and l_v is its complementary angle, with the same sign.

To illustrate the comparison between the two vertex deviation definitions, in Fig. A1 we show the same velocity ellipses as in Fig. 6, this time colour-coded by the values resulting from equation (A6). The inner regions along the bar's semiminor axis, where $\beta_{rl} < 0$, are clearly highlighted, reaching values close to $-90°$ where the velocity ellipses are perpendicular to $\hat{R}$.

Equation (2) should be used when we do not want vertex deviation to encode any information about which velocity dispersion dominates. For example, $\tilde{l}_v$ values of $0°$ and $90°$ returned by equation (A6) both describe a velocity ellipse which is not tilted with respect to the coordinate axes, which means the covariance is negligible relative to the magnitude of the dispersion difference. Equation (2) would treat both of these cases the same by assigning them a value of $l_v = 0°$. The information about which dispersion dominates, i.e. which of the two coordinate axes the ellipses' semimajor axis is closest to, is quantified by the anisotropy.

A3 Caution on error treatment

It is worth noting that, when estimating uncertainties using bootstrapping, equation (5) must be applied with some care when using equation (A6) to compute vertex deviation. This is because of the need to limit the range of atan2 to a particular branch so as to make the values unique. The smallest of the two angles which result from the crossing of two straight lines cannot be greater than $90°$. Using the bootstrap method to compute $\epsilon(\tilde{l}_v)$ can produce bootstrap values which differ from the value computed from the original population by more than $90°$, given the principal branch spans a region of $180°$. This would lead to an overestimation of the error. The solution is to take those bootstrap ellipses and consider the other end of their semimajor axis, reflecting their $\tilde{l}_v^*$ values across the origin. Calling the vertex deviation of the original population $\tilde{l}_v$, the resulting bootstrap distribution belongs to the branch $(\tilde{l}_v - 90, \tilde{l}_v + 90]°$, where for any bootstrap value $\tilde{l}_v^*$ we would have $|\tilde{l}_v^* - \tilde{l}_v| \leq 90°$, as it should be.

A4 Relation to the anisotropy and correlation

Equation (A5) can be expressed in terms of the correlation (ρ_{ij}) and in-plane anisotropy (β_{ij}), defined in equations (3) and (4), respectively, as

$$\tan(2l_v) = 2\rho_{ij} \frac{\sqrt{1 - \beta_{ij}}}{\beta_{ij}}, \quad (A9)$$

so that the equivalent of equation (2) is

$$l_v = \frac{1}{2} \arctan \frac{2\rho_{ij} \sqrt{1 - \beta_{ij}}}{|\beta_{ij}|}. \quad (A10)$$

The vertex deviation using the equivalent of equation (A6) instead is

$$\tilde{l}_v = \frac{1}{2} \text{atan2} \left(2\rho_{ij} \sqrt{1 - \beta_{ij}}, \beta_{ij} \right). \quad (A11)$$

APPENDIX B: ADDITIONAL MODELS

In Fig. B1, we compare the anisotropy, correlation, and vertex deviation of our fiducial model with two additional models. The first model is the weak bar model presented by Gough-Kelly et al. (2022). The second model has not previously been presented; it has only a central weak oval, not a bar. We scale the three models identically, with the bar/oval orientated as in the fiducial model, and using the same age cuts for the young and old stellar populations.

We select stars along the bulge minor axis, namely $|l| < 2°$ and $R_{GC} < 3.5$ kpc. The figure shows that the amplitude of β_{rl} is largest in the young population of the central-oval model, not in the fiducial model, which in turn has a larger amplitude than in the weak bar's young population. In the vertex deviation, the (negative) amplitude of the young population is comparable between the fiducial and weak bars, while in the weak-oval case, the amplitude of l_v in smaller than in the other two models for both populations.

In contrast with β_{rl} and l_v, the correlation has amplitudes that follow the strength of the bar in the $3° < |b| < 6°$ region, with the largest ρ_{rl} in the young population of the fiducial model, and the smallest (near-zero) ρ_{rl} in the old population of the weak-oval model. From these results, we conclude that the correlation is the best tracer of the bar strength of any given population.
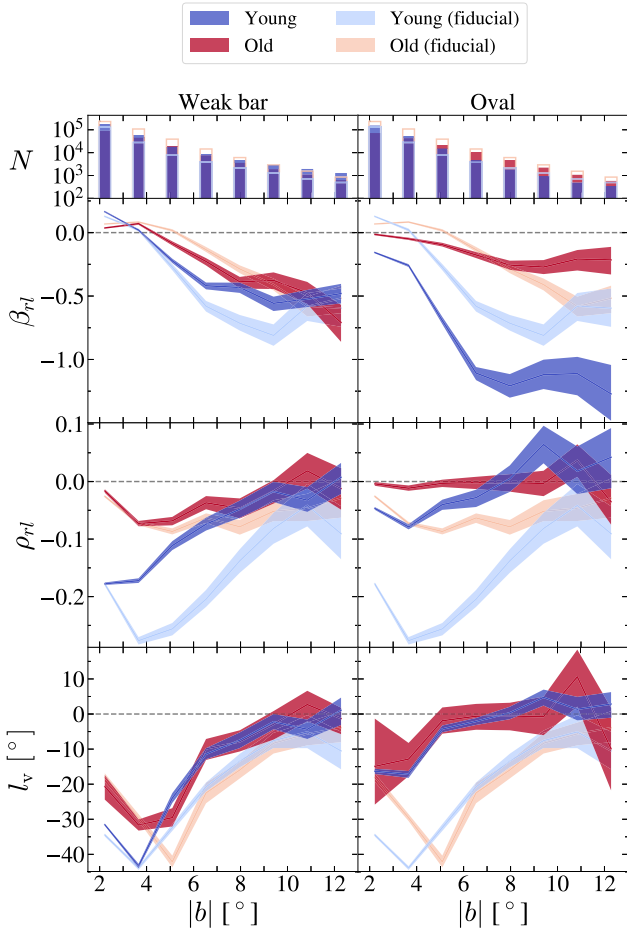
Figure B1. Anisotropy (top), correlation (middle), and vertex deviation (bottom) as a function of latitude along the minor axis of the bulge, $|l| < 2°$, within $R_{GC} < 3.5$ kpc, for models with a weaker bar (left panels) and only a central oval (right panels), compared with the strongly barred fiducial model (light colours). We split the stellar populations, and scale the new models, as described for the fiducial model in the main text. The surfaces represent the 68 per cent percentile bootstrap confidence intervals.

APPENDIX C: TESTING THE ASSUMPTION OF BOOTSTRAPPING

In this section, we use the model to briefly discuss the validity of bootstrapping for the kinematic variables of interest.

Let us suppose we have a sample of size n, extracted from a larger population of size $\gg n$. We are interested in knowing how closely the correlation of the sample, called the sample estimate $\hat{\rho}$, approximates the correlation of the total population, ρ. If we had access to the population, we could investigate this by drawing many random samples of size n with replacement from the population, and computing all their sample estimates $\hat{\rho}$. The histogram resulting from binning these values is called the sampling distribution. Its standard deviation, $\sigma(\hat{\rho})$, is called the standard error, and measures the precision of the sample estimates, quantifying their variability around the mean, $\langle\hat{\rho}\rangle$. The difference between the mean and the value computed from the population, $\langle\hat{\rho}\rangle - \rho$, is called the bias, and it measures the average accuracy of the sample estimates. If the bias is zero, the standard error measures both the precision and accuracy of the sample estimates, as it quantifies their variability around the true value, computed from the population.

In inferential statistics, we do not have access to the population, only to the sample, of size n. Therefore, we cannot compute the precision or accuracy of the sample estimate, $\hat{\rho}$, as described above because we cannot build the sampling distribution. Instead, we can take many bootstrap samples of size n with replacement from the sample itself, and compute all their correlation values, ρ^*, called bootstrap estimates. The histogram resulting from binning these values is called the bootstrap distribution. Its standard deviation, $\sigma(\rho^*)$, is called the bootstrap standard error, and the difference between its mean and the original sample estimate, $\langle\rho^*\rangle - \hat{\rho}$, is called the bootstrap bias. If the bootstrap assumption held true, meaning the original sample was representative of the underlying population, then the bootstrap standard error would approximate the actual standard error, i.e. $\sigma(\rho^*) \approx \sigma(\hat{\rho})$, and the bootstrap bias would approximate the actual bias, i.e. $\langle\rho^*\rangle - \hat{\rho} \approx \langle\hat{\rho}\rangle - \rho$. Therefore, we would be able to use these quantities as a measure of the precision and accuracy of our original sample estimate (Hesterberg 2015).

We use our model to test the bootstrap assumption on our statistics of interest at different sample sizes for the young and old populations. We select the stars on the bulge minor axis, namely $|l| < 2°$, $3° < |b| < 6°$, and $R_{GC} < 3.5$ kpc. From each population we extract 5000 random samples with replacement of sample sizes varying from $n = 50$ to 5000. Samples of size $n = 5000$ represent 14.8 per cent and 3.5 per cent of the young and old populations, respectively. For each sample size, we aggregate the 5000 sample estimates of the statistics of interest ($\hat{\beta}_{rl}$, $\hat{\rho}_{rl}$, and $\hat{l}_v$) into a sampling distribution, and we compute the standard error and bias as described in the first paragraph. Therefore, we obtain a value of standard error and bias for each sample size n, shown as the solid lines in Fig. C1.

From each of the 5000 samples of size n extracted from a population, we then draw, from the sample itself, 500 bootstrap samples of the same size n with replacement, and aggregate their values of the statistics of interest, i.e. the bootstrap estimates (β_{rl}^*, ρ_{rl}^*, and l_v^*), into a bootstrap distribution. We then compute the bootstrap standard error and the bootstrap bias as described above. Doing this for all 5000 samples of size n extracted from the population results in 5000 values of bootstrap error and bootstrap bias. We take the average and show the mean bootstrap standard error and mean bootstrap bias as dashed lines in Fig. C1. The 68 per cent percentile range is also shown as a shaded region.

In Fig. C1(a), we show the standard error as a function of sample size. For the anisotropy and correlation in the first two panels, the standard error (solid line) and mean of the bootstrap standard errors (dashed line) match closely for both populations for all n. Therefore, on average the bootstrap assumption holds for these variables. Given the standard error curves for the young and old populations match quite closely despite their true values of anisotropy and correlation differing (see Fig. 12), these conclusions depend almost solely on sample size. We illustrate this with the dotted lines, which show fits of $f(n) \propto n^{-1/2}$ to the actual standard error curves. For the correlation we show the analytic expression recommended by Gnambs (2023), which depends on the correlation value itself, though this makes little difference in practice at our range of values ($\lesssim 0.3$).

The standard error and mean bootstrap standard error of the vertex deviation of the young stars also roughly match, with maximum deviation of $\sim 2°$. However, for the old stars at sample sizes below $n \lesssim 400$, these two curves separate, with the mean bootstrap standard error underestimating the actual standard error by up to $\sim 6°$. The 68 per cent interval of the old population also takes longer to converge to the actual standard error. Moreover, unlike in the other panels, the
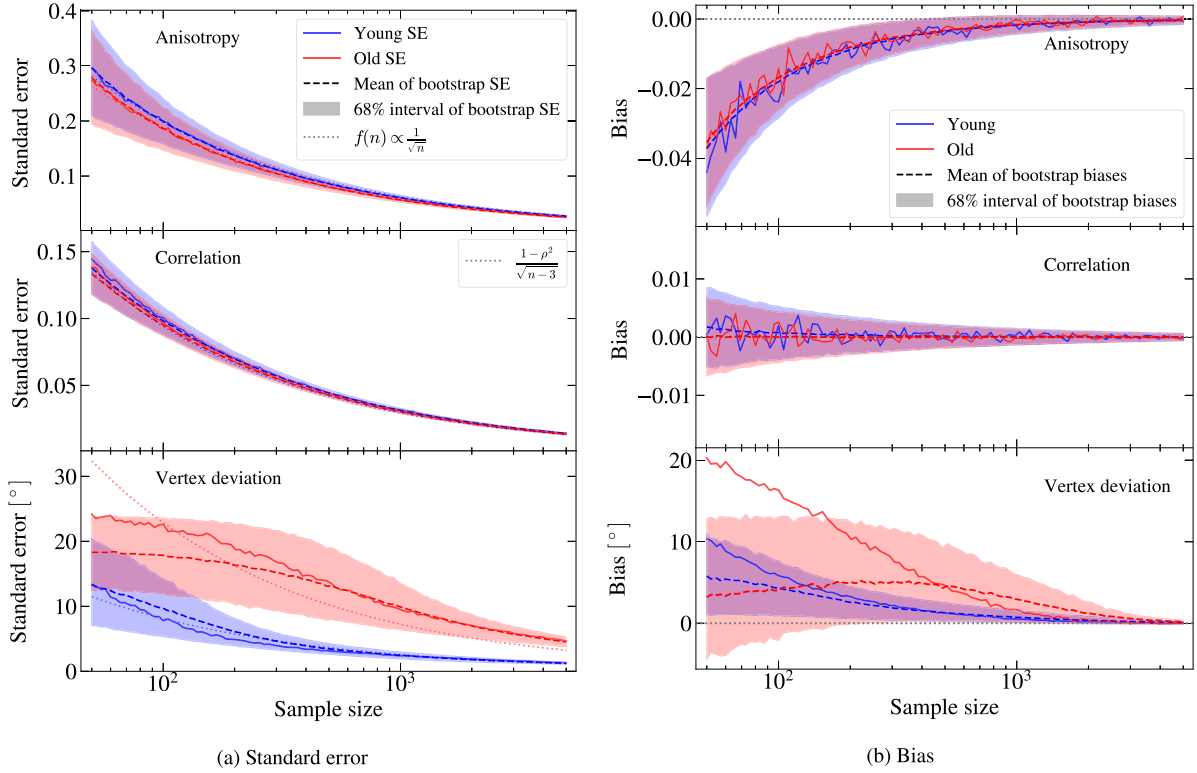
Figure C1. Standard error (a) and bias (b) as a function of sample size, for young (blue) and old (red) stars on the bulge minor axis. The dashed lines show the mean bootstrap standard error (a) and mean bootstrap bias (b). The variability of bootstrap standard errors and bootstrap biases around the mean is illustrated by the 68 per cent percentile range surface. On column (a), the dotted curves show fits of $f(n) \propto n^{-1/2}$ to the standard errors. On column (b), the horizontal dotted lines indicate zero.

standard errors for the young and old populations are different for all sample sizes, and hence do not depend only on n. The vertex deviation of isotropic and weakly correlated populations, like the old one here (see Fig. 12), whose velocity ellipse is closer to a circle with less well-defined semimajor axis direction, have larger standard errors and the bootstrap assumption on average breaks below a certain n, with the bootstrap error on average underestimating the actual standard error.

In Fig. C1(b), we show the bias as a function of sample size. For anisotropy and correlation, the bias (solid line) and the mean bootstrap bias (dashed line) match closely for both populations for all n, which again confirms that the bootstrap assumption holds for these statistics. It is worth noting that the anisotropy develops a negative bias that increases in magnitude as sample sizes drop below $n \approx 10^3$. This means that the sample estimates at those sample sizes on average deliver an anisotropy value lower than the true value computed from the population. However, the effect is small ($\lesssim 0.04$) even for the smallest samples of $n = 50$. Therefore, the correlation and anisotropy are largely unbiased estimators, which means that the standard error of these statistics is a measure of both the precision and accuracy of sample estimates, as it measures their variability around the true value, computed from the population.

The bootstrap assumption again breaks for vertex deviation, this time for both young stars below $n \approx 200$ and old stars below $n \approx 400$, reaching a discrepancy of $\sim 5°$ and $\sim 17°$, respectively, between the actual and the mean bootstrap bias at the smallest sample size of 50. For both populations the mean bootstrap bias underestimates the actual bias. Moreover, unlike the anisotropy and correlation, the vertex deviation develops a significant bias below $n \lesssim 10^3$, reaching

up to $10°$ and $20°$ for the young and old populations, respectively, at $n = 50$. As a result, the standard error of vertex deviation at these sample sizes is a measure of the precision of the sample estimates but not of their accuracy in representing the true value of the underlying population.

We summarize the findings of this appendix below:

(i) The standard error and bias of the anisotropy and correlation depend almost solely on sample size, n. The standard error varies as $\propto n^{-1/2}$, and only small anisotropy biases are introduced at low n. For vertex deviation, the standard error and bias at the same n differ for the young and old populations, with larger standard errors and biases for the old. Vertex deviation develops biases for both the young and old populations at $n \lesssim 10^3$, reaching $\sim 5°$–$15°$ at $n = 100$. This means that at those sample sizes the standard error quantifies the precision of vertex deviation estimates but contains limited information about their accuracy in representing the underlying population.

(ii) The bootstrap assumption holds on average for the anisotropy and correlation at all n, with the mean bootstrap standard error and bias matching the actual standard error and bias, respectively. The same is roughly true for the standard error of the vertex deviation of the young stars, but not for the old, whose mean bootstrap standard error underestimates the actual standard error below $n \approx 400$ by up to $\sim 6°$. Below that n, bootstrapping also underestimates the bias by up to $17°$ for the old stars and $5°$ for the young at $n = 50$.

This paper has been typeset from a T$_{\rm E}$X/L$^{\rm A}$T$_{\rm E}$X file prepared by the author.