Astronomy & Astrophysics manuscript no. anticentreEDR3
December 2, 2020

Gaia Early Data Release 3: The Galactic anticentre

Gaia Collaboration, T. Antoja[1], P. McMillan[2], G. Kordopatis[3], P. Ramos[1], A. Helmi[4], E. Balbinot[4], T. Cantat-Gaudin[1], L. Chemin[5], F. Figueras[1], C. Jordi[1], S. Khanna[4], M. Romero-Gomez[1], G. Seabroke[6], and et al.

[1] Institut de Ciències del Cosmos, Universitat de Barcelona (IEEC-UB), Martí i Franquès 1, 08028 Barcelona, Spain,
e-mail: `tantoja@fqa.ub.edu`
[2] Lund Observatory, Department of Astronomy and Theoretical Physics, Lund University, Box 43, 22100 Lund, Sweden
[3] Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Bd de l'Observatoire, CS 34229, 06304 Nice Cedex 4, France
[4] Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD Groningen, The Netherlands
[5] Centro de Astronomía - CITEVA, Universidad de Antofagasta, Avenida Angamos 601, Antofagasta 1270300, Chile
[6] Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, United Kingdom

Received September 15, 1996; accepted March 16, 1997

ABSTRACT

Aims. We aim to demonstrate the scientific potential of the *Gaia* Early Data Release 3 (EDR3) for the study of different aspects of the Milky Way structure and evolution, while at the same time we provide a description of several practical aspects of the data and examples of their usage.
Methods. We use astrometric positions, proper motions, parallaxes, and photometry from EDR3 to select different populations and components and to calculate the distances and velocities in the direction of the anticentre. In this direction, the *Gaia* astrometric data alone enables the calculation of the vertical and azimuthal velocities, and the extinction is relatively low compared to other directions in the Galactic plane. We then explore the disturbances of the current disc, the spatial and kinematical distributions of early accreted versus in-situ stars, the structures in the outer parts of the disc, and the orbits of the open clusters Berkeley 29 and Saurer 1.
Results. With the improved astrometry and photometry of EDR3 we find that: i) the dynamics of the Galactic disc are very complex with oscillations in the median rotation and vertical velocities as a function of radius, vertical asymmetries and new correlations, including a bimodality with disc stars with large angular momentum moving vertically upwards from below the plane, and disc stars with slightly lower angular momentum moving preferentially downwards; ii) we resolve kinematic substructure (diagonal ridges) in the outer parts of the disc for the first time; iii) the red sequence that has been associated with the proto-Galactic disc present at the time of the merger with Gaia-Enceladus-Sausage is currently radially concentrated up to around 14 kpc, while the blue sequence that has been associated with debris of the satellite extends beyond that; iv) there are density structures in the outer disc both above and below the plane most probably related to Monoceros, the Anticentre Stream and TriAnd, for which the *Gaia* data allows an exhaustive selection of candidate member stars and dynamical study; v) the open clusters Berkeley 29 and Saurer 1, despite being located at large distances from the Galactic Centre, are on nearly circular disc-like orbits.
Conclusions. Even with our simple preliminary exploration of the *Gaia* EDR3, we demonstrate how, once again, these data from the European Space Agency are crucial for our understanding of the different pieces of our Galaxy and their connection to its global structure and history.

Key words. Galaxy: disc – Galaxy: halo – open clusters and associations: individual – Galaxy: formation – Galaxy: kinematics and dynamics – Stars: distances

1. Introduction

As for previous releases, the Early Third Data Release (EDR3, Gaia Collaboration 2020) of the *Gaia* mission (Gaia Collaboration et al. 2016) of the European Space Agency comes accompanied with a series of performance verification articles that show the quality of the data, the improvements with respect to previous releases, and the scientific potential for multiple research areas in astrophysics (see also Luri 2020; Smart 2020; Klioner 2020).

In the present study we focus on a specific area in the sky that allows us to explore different elements of the Milky Way's structure and history: the Galactic anticentre. This region of the Galaxy has the advantage that from astrometric measurements alone (proper motions and parallaxes), one can calculate the vertical and azimuthal (rotation) motion of the stars with a negligible contribution of the line-of-sight velocity. Also, the anticentre has relatively low extinction compared to other directions of the Galactic disc.

More importantly, the anticentre is a meeting point of several distinct components of the Galaxy (the disc, the halo) and possibly hosts ancient and recently disrupted stellar systems of extragalactic origin. The anticentre is also an excellent window to the dynamics and the past of the Milky Way: due to the lower gravitational potential, any perturbation on the disc would cause more significant deformations than in the inner disc, and, due to the longer dynamical timescales, these could still be observable today (e.g. Binney & Tremaine 2008; Antoja et al. 2018; Laporte et al. 2019b).

In this paper we focus on several aspects of the Galaxy that coexist in the anticentre and that will help us towards answering a single question: "How does the Galaxy appear today and how did it become like this?" Thanks to a combination of models and measurements, in which *Gaia* DR2 (Gaia Collaboration et al. 2018b) played one of the most relevant roles, we have already uncovered part of the Milky Way structure and history. The major accretion event of the so-called Gaia-Enceladus-Sausage around 10 Gyr ago (Helmi et al. 2018; Belokurov et al. 2018), together with the ongoing accretion of the Sagittarius dwarf galaxy (Mateo et al. 1996; Majewski et al. 2003 and a recent detection with *Gaia* data in Antoja et al. 2020), and internal structures such as the bar (de Vaucouleurs 1964; Binney et al. 1991; Blitz & Spergel 1991; Weiland et al. 1994) and the spiral arms (Drimmel & Spergel 2001; Reid et al. 2009) are among the most important phenomena that have shaped our Galaxy throughout its evolution (see also Bland-Hawthorn & Gerhard 2016). The footprints of these phenomena can still be observed today and that is what we investigate here.

First, we look into the kinematic disturbances of the disc that EDR3 allows us to inspect in its outermost parts with more detail. Already known disturbances include vertical asymmetries in the number counts linked to vertical bending and breathing waves (e.g. Widrow et al. 2012; Bennett & Bovy 2019), substructure in the in-plane velocities (e.g. Dehnen 1998; Famaey et al. 2005; Antoja et al. 2008, 2018; Kawata et al. 2018; Khanna et al. 2019a), large scale velocity patterns in the disc (Siebert et al. 2011; Williams et al. 2013; Carlin et al. 2013; Antoja et al. 2017) and other phase space correlations (e.g. Schönrich & Dehnen 2018; Friske & Schönrich 2019; Huang et al. 2018; Cheng et al. 2020). While some of these could be related to the structures such as the bar, the poorly-known spiral arms and the warp, more recently, the *Gaia* vertical phase spiral (Antoja et al. 2018; Bland-Hawthorn et al. 2019) suggests a phase mixing event after the perturbation of Sagittarius (Binney & Schönrich 2018; Laporte et al. 2019b). All of these aspects have proven to be extremely difficult to understand, and also to disentangle or relate. Here we look at the rotation and vertical velocities of the outer disc, showing the power that *Gaia* data can have in our understanding of this complexity and the role that recent and past, internal and external, perturbations have had in the Milky Way.

Second, we go from the current disc of the Galaxy to more ancient components. Two distinct populations were clearly apparent in the Hertzsprung-Russell (HR) diagram of stars with large tangential velocities near the Sun (i.e., representing the stellar halo) by Gaia Collaboration et al. (2018a) using *Gaia* DR2: a blue and a red sequence. It has been proposed that these two populations correspond to an accreted one stemming largely from the galaxy Gaia-Enceladus-Sausage, and an in-situ heated (thick) disc, different from the canonical thick disk, that was present at the time of this merger (Helmi et al. 2018; Belokurov et al. 2018; Di Matteo et al. 2019; Gallart et al. 2019; Belokurov et al. 2020). Here we analyse the spatial distribution and kinematics of stars from each of the HR sequences to investigate out to which distance the debris of Gaia-Enceladus-Sausage may be found, and constrain the extent of the suggested proto-Galactic disc present at that time.

Thirdly, we explore the density structures that appear in the edge of the disc in the anticentre direction. Newberg et al. (2002) using the deep Sloan Digital Sky Survey (SDSS York et al. 2000) discovered the existence of a $\sim 100^\circ$ structure in their A/F star count maps. Now known as Monoceros, later studies have confirmed its existence and large extension on the sky (e.g., Crane et al. 2003; Yanny et al. 2003; Ibata et al. 2003). Together with the Anticentre stream (ACS, Grillmair 2006), both at a distance $\sim$10 kpc from the Sun, and the Triangulum-Andromeda overdensities (TriAnd, Majewski et al. 2004; Rocha-Pinto et al. 2004; Martin et al. 2007), they are part of a complex and substructured outer disc. The initial interpretation that these could be the remains of an accreted dwarf galaxy (e.g, Martin et al. 2004; Peñarrubia et al. 2005) has become less plausible (although not completely ruled out) since: i) there is no known progenitor to the stream (the candidate Canis Major has been discarded – Momany et al. 2004, 2006; Rocha-Pinto et al. 2006), and ii) the kinematics of the structures (e.g., de Boer et al. 2018), their metallicities and their ratio of RR Lyrae to Giants (Li et al. 2012; Sheffield et al. 2014; Price-Whelan et al. 2015; Sheffield et al. 2018; Bergemann et al. 2018; Laporte et al. 2020a) are compatible with that of the disc. Here we explore how these structures look in *Gaia* EDR3 and coexist with other structures such as the Sagittarius stream.

Finally, we explore the open clusters Berkeley 29 (Kaluzny 1994; Lata et al. 2002; Tosi et al. 2004; Bragaglia & Tosi 2006) and Saurer 1 (Frinchaboy & Phelps 2002; Carraro & Baume 2003) in the anticentre direction that, with ages of several Gyr, are among the oldest Galactic clusters known. Their unusual location at Galactocentric distances of $\sim$20 kpc and more than 1 kpc above the Galactic mid-plane is a puzzle that has led several authors to question whether they are associated with the disc, and to propose a possible extragalactic origin (e.g. Frinchaboy et al. 2004). Attempts to characterise the orbits of these two objects have returned widely discrepant results (Carraro et al. 2007; Vande Putte et al. 2010), mainly due to their poorly-constrained proper motions since at such large distances, small proper motion errors translate into large uncertainties in physical velocities. An additional issue has been the uncertain membership status of individual stars. Here, thanks to *Gaia* EDR3, we perform a robust analysis of the membership of these clusters and derive their orbits with high confidence.

To investigate all these aspects, the main *Gaia* data products that we use here are the astrometric measurements. For EDR3 these show a substantial decrease of uncertainties resulting from the use of 34 months of data (12 more than for DR2). Apart from a higher completeness at the faint end, there is a significantly larger number of stars at a given parallax precision. The combination of all these improvements essentially means that we can now explore distant regions of the Galaxy in the direction of the anticentre, even reaching around 16 kpc from the Galactic centre and beyond (see Sect. 2 for details), and thus, the very outskirts of the disc, for a sample with positions and velocities of excellent quality. Moreover, important improvements in the pipelines of the *Gaia* photometry have resulted in photometric bands with significantly less systematic error, from which, combined with the improved parallaxes, cleaner HR diagrams can be built and used to select different populations and components.

The paper illustrates how, once more, the new *Gaia* data are set to revolutionise our knowledge of the Galaxy and its past. Additionally, we describe practical aspects of the data and examples of its use that might be of interest for the community, such as queries in the *Gaia* Archive, quality cuts, derivation of distances (Bayesian inference, considerations on the parallax zero point), etc. We also complement our analysis with the use of simulated data from the Gaia Object Generator (GOG, Luri et al. 2014) & the Gaia Universe Model Snapshot (GUMS, Robin et al. 2012), now available directly in the Gaia Archive, to evaluate the effects of selection, errors and extinction.

We start our paper by explaining the different datasets used and demonstrating the different improvements (but also limitations) of the EDR3 data in the anticentre direction (Sect. 2) We continue by explaining how the distances and phase space coordinates are derived (Sect. 3). The results sections follow, organised into the explorations of the disc dynamics (Sect. 4), halo, thick disc and outer disc structures (Sect. 5), and distant

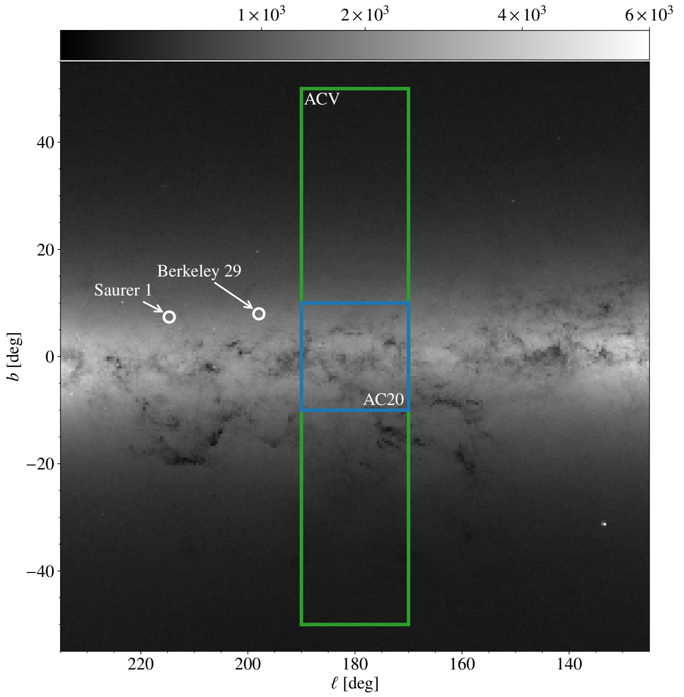

Fig. 1. EDR3 star counts in the anticentre region with the different data selections used. The HEALpix map is obtained by querying the *Gaia* archive the counts of stars within each HEALpix of level 8 (query 2 in Appendix A). The size of the circles to indicate the position of the clusters does not correspond to the size used for the selection which is much smaller. Several other clusters can be seen in the figure, and also the Triangulum Galaxy (M33, bottom right corner).

Table 1. Number of stars in the different samples and comparison with DR2. The numbers are given for the different data samples described in Sect. 2.2 and different sub-selections. The numbers in the first two numerical columns are for samples without the `excess_flux` and `RUWE` selections since these are not equally defined in the different releases. [(a)] See footnote 1.

	DR2	EDR3	EDR3+filters
1. AC20	13 307 312	14 120 029	11 949 642
5p-6p[a]	10 750 864	12 279 076	11 949 642
$\varpi/\sigma_\varpi > 3$	2 645 014	3 518 388	3 369 456
photometry	12 618 364	13 706 954	11 436 625
2. ACV	24 578 296	25 835 286	21 835 927
$\varpi < 0.1$ mas	4 974 104	4 879 087	4 509 263
$\varpi < 0$ mas	3 945 985	3 781 306	3 496 645
4. Clusters			
Berkeley 29	365	370	334
Saurer 1	283	284	263

open clusters (Sect. 6). We present our discussion and conclude in Sect. 7.

2. Data

2.1. Main datasets

In this work we explore the Galactic anticentre region using different data selections obtained from *Gaia* EDR3 (Gaia Collaboration 2020) that can be accessed through the *Gaia* Archive (`https://gea.esac.esa.int/archive/`, Gonzalez-Nunez 2020). More details on the data and validation are given also in Lindegren (2020a), Riello (2020) and Fabricius (2020). The main datasets are shown in Fig. 1 and listed below. The number of stars for these samples are summarised in Table 1, in which a comparison with DR2 is also shown.

(1) AC20: A square on the sky centred at $(\ell, b) = (180, 0)$ deg of 20 deg on a side (blue square in Fig. 1). This sample is used to explore Galactic disc kinematics in Sect. 4. It contains 14 120 029 stars but most of the time we use only the selection with $\varpi/\sigma_\varpi > 3$ (see Sect. 3.2), which comprises 3 518 388 sources (AC20-$\varpi/\sigma_\varpi > 3$). The data are retrieved from the archive using the query 1 in Appendix A. Similar queries have been used for other samples.

(2) ACV: A rectangle on the sky centred at $(\ell, b) = (180, 0)$ deg with a width of 20 deg in ℓ and height 100 deg in b (green rectangle in Fig 1). This sample is used to explore the halo and the structures in the outer disc such as Monoceros or the Sagittarius stream in Sect. 5. For parts of our analysis, we perform a selection of $\varpi < 0.1$ mas to favour distant stars. We note that with this selection there are 2% fewer stars in EDR3 than in DR2 (Table 1). While the total number of stars in that region has increased with respect to DR2, many of the stars added are nearby faint dwarfs (see Sect. 2.3) and the overall quality of the parallaxes has improved significantly as proven by the decrease in the number of sources with a negative parallax (and spurious sources). As a consequence, our parallax cut is now able to reject the nearby sources more efficiently, thus resulting in a slightly smaller sample.

(3) Two clusters in the anticentre: All sources brighter than $G = 19$ within 4 arcmin of the centres of the extremely distant Galactic old open clusters Berkeley 29 and Saurer 1. These data are analysed in Sect. 6.

In all our analysis, our fundamental observables are the astrometric quantities ϖ, $\mu_{\alpha*}$, μ_δ (parallax and proper motions) and the photometric bands G, BP, RP. In order to use the best quality data, we apply several selections. First we apply the following astrometric quality selection on the Renormalised Unit Weight Error (`RUWE`) as recommended in Lindegren (2020a):

$$\texttt{RUWE} < 1.4 \qquad (1)$$

On the other hand, whenever the photometry is used we select good photometry sources with:

$$0.001 + 0.039(BP - RP) < \log_{10} \texttt{excess\_flux} < 0.12 + 0.039(BP - RP). \qquad (2)$$

Sources out of these limits have inconsistent G, $G_{\rm BP}$ and $G_{\rm RP}$ fluxes due to blends (more than one source in the BP/RP windows), contamination by a nearby source (outside the window) or a sign of the extended nature of the source. Additionally, we correct the fluxes in G for 6p sources following Riello (2020) -their Table 5- using directly an ADQL query as suggested in Gaia Collaboration (2020). The last column in Table 1 indicates the number of stars after these selections.

2.2. Complementary datasets

For validation and other purposes, we also use the following complementary data:

(i) 6Dsample: a full sky sample with stars that have DR2 line-of-sight velocity in EDR3 (Seabroke 2020), thus with full 6D phase space information. After filtering, this sample contains 6 156 684 stars and is used mainly in Sect. 4.

(ii) DR2: the same selections as above (AC20, ACV) but for DR2. These are used for comparison with EDR3.

(iii) GOG & GUMS: the same selections as above but for GOG (Luri et al. 2014; GEDR3 documentation Chapter 2) which is a mock *Gaia* catalogue based on the Besançon model (Robin et al. 2003), and for GUMS (Robin et al. 2012) which contains the intrinsic properties of the mock sources before applying the *Gaia* instrument modelling. Here we use the GOG version 20.0.3. with uncertainties that have been scaled to 34 months of data (but see Fig. E.1). These samples are used for the evaluation of completeness and extinction effects and they do not contain any kinematic substructure or asymmetries. Furthermore, GOG and GUMS are used in Appendix C for testing how robust each of the distance estimation methods is. These simulated data have been retrieved through the Gaia Archive querying from the corresponding tables (`gaiaedr3.gaia_source_simulation` and `gaiaedr3.gaia_universe_model`).

(iv) 2MASS: We use the official crossmatch of EDR3 with the *2MASS* point source catalogue (Skrutskie et al. 2006), provided in `gaiaedr3.tmass_psc_xsc_best_neighbour.` For the AC20 sample, this yields about 55% anticentre objects with *2MASS* photometry. These data are used to select Red Clump (RC) stars and compute their photometric distances (Sect. 2.4 and 3.2, Appendix B and C.2.1).

More information can be found in the respective sections in which these samples are used.

2.3. EDR3 data quality and completeness

In this section we examine the quality of the EDR3 data and compare it to DR2. The most relevant improvements in EDR3 for our study include a larger number of sources at the faintest magnitudes and a significant decrease of the astrometric uncertainties and thus a significantly larger number of stars with a certain parallax precision. Below we show these aspects in more detail focusing on the AC20 region as an example unless stated otherwise.

1. General description. Figure 2 shows the AC20 region in Galactic coordinates coloured according to different quantities in bins of 0.1 deg. In panel a we show the number of stars per bin while the rest of the panels show median quantities. As expected, the counts anti-correlate with the patterns seen in the extinction map from Green et al. (2019) (d, see details in Sect. 2.4) combined with the decrease with Galactic latitude $|b|$. The median magnitude (b) and median colour (c) also correlate with extinction (d): higher extinction regions have, on average, more reddened sources that have fainter (more extinguished) apparent magnitudes. Panel d shows that there is higher extinction for $b < 0$. Additionally, there is a horizontal elongated window at $b \sim 2.5$ deg of low extinction with far more counts and brighter magnitudes, which is seen also in other panels where brighter magnitudes essentially translate into smaller astrometric errors (e.g. panel g) and also smaller parallaxes (panel f, stars reaching farther distances). Whether this feature with larger counts reflects more than simply lower extinction (e.g. a flexing of the disk) requires a deeper analysis of the extinction and the selection function. We also note that the thin nearly-horizontal lines in panels a and b are a consequence of the `RUWE` selection.

2. Completeness. The evaluation of the completeness of the *Gaia* data is a difficult task given that there is no deeper survey with a comparable resolution. Distinct methodologies to assess the data completeness can be found in Boubert & Everall (2020), Lindegren (2020a) and Fabricius (2020). Here we examine it in a simpler way. First we note that the AC20 sample (without any cuts) has about one million more stars in EDR3 compared to DR2 (see Table 1). Figure 3 shows histograms of the G magnitude for stars in the AC20 sample in DR2 and EDR3 (blue solid and dashed histograms) showing a great increase of sources at the faint magnitudes with respect to DR2. This was expected given that the detection on board prioritises bright magnitudes and the effect of more months of observations produces new detections mostly at the faintest bins.

Figure 2e shows the map of median `visibility_periods_used`. This panel shows bands at different spatial scales that correspond to regions with higher/lower number of observations and thus higher/lower completeness. The thin, nearly-horizontal, yellowish pattern, separated by roughly 0.7deg, similar to the width (across scan) of *Gaia*'s FOV, corresponds to consecutive scans that did not overlap in across scan. The wider purple bands, indicating areas where the coverage is better, are close to some "nodes" in the scanning law, that is, the positions in the sky that get repeated coverage during some consecutive scans. Figure 4 shows the star counts for different ranges in G for DR2 (top) and EDR3 (bottom) using the same colour scale. The bands of the scanning law appear clearly and correlate with Fig. 2e. Comparing to DR2, we can clearly notice the larger number of stars in EDR3 in these two magnitude ranges as well as the reduction of some of the bands (at scales of $\sim$ 3 deg) imprinted in DR2. In the range of $20.75 < G < 21$ some scanning bands are still present.

3. Completeness of the kinematic samples. Some of the *Gaia* sources have only partial astrometric solutions, from which only sky coordinates are derived (2p solutions) while others have full astrometric solution (positions, parallax and proper motions available) and are dubbed 5p and 6p solutions (Lindegren 2020a) , where the 6th parameter is the colour[1]. In the first rows of Table 2 we give the number of stars with partial (2p) and full (5p, 6p) solutions comparing DR2 and EDR3 for the whole AC20 sample and for different ranges of magnitude. In EDR3, there are two million more stars with full astrometric solution than in DR2. The table also shows the percentage of full solutions relative to all sources in EDR3, which gives an indication of the internal completeness of the kinematic data. Most notably, in the range of $19 < G < 20$ there is now a 98% internal completeness compared to the 82% in DR2, and in the range $20 < G < 20.7$ the percentage is now 90% versus the old 64%, verifying that, as shown also with the orange solid and dashed histograms of Fig. 3, there is an outstanding gain at the faintest magnitudes. These stars have never been used before in kinematic studies with *Gaia* data.

---

[1] 2p partial solutions (only positions) are indicated as `astrometric_params_solved=3` in the *Gaia* Archive. 5p solutions are those for which the *Gaia* colour is used in the astrometric solution, while in the 6p cases, this quantity, more precisely, the pseudocolour, is derived simultaneously in the solution (Lindegren 2020a). The 5p and 6p solutions correspond to `astrometric_params_solved=31` and 95, respectively. In DR2 all full astrometric solutions were included under the `astrometric_params_solved=31` case, even if in some cases a chromaticity different from the photometric colour was used.

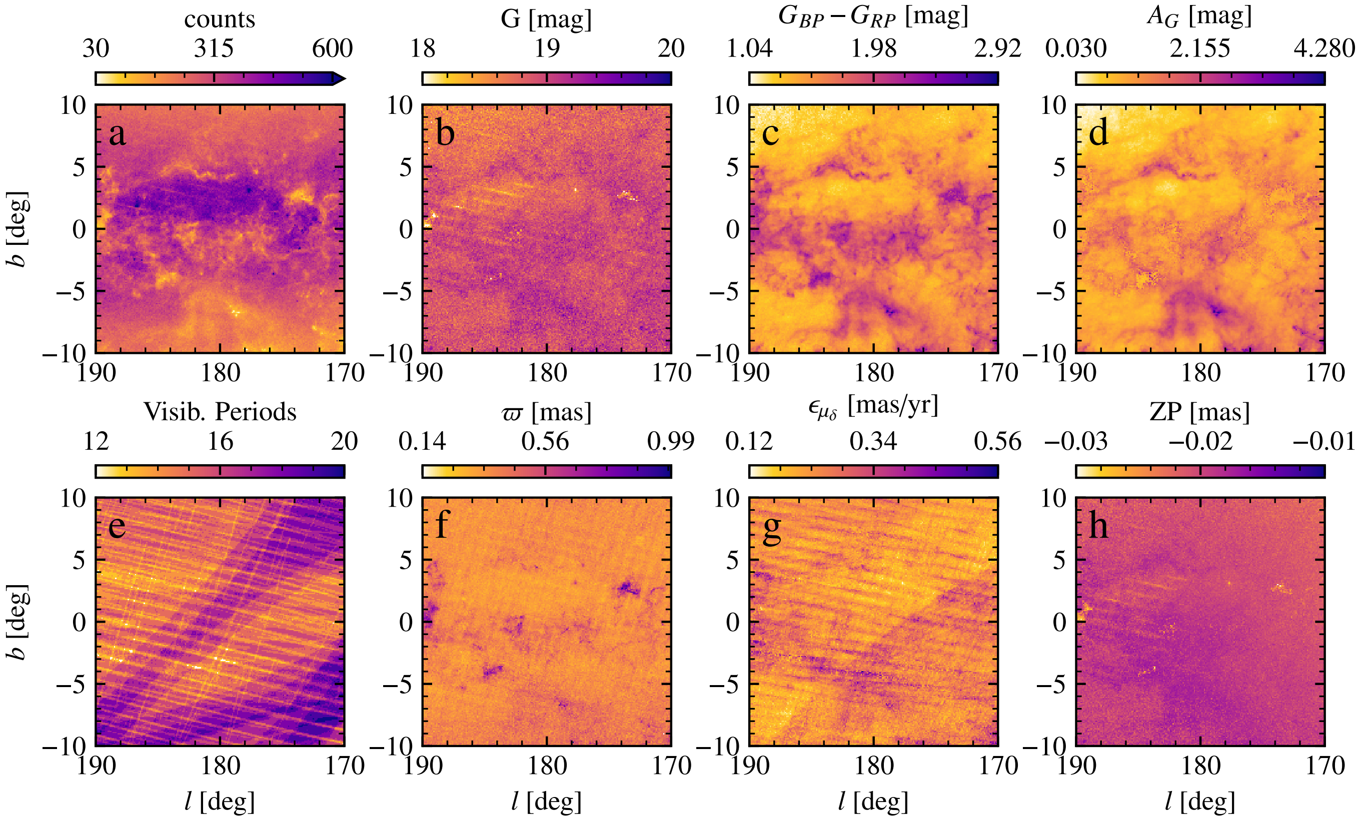

Fig. 2. Characteristics of the anticentre AC20 sky. Panel (a) shows the number of sources and the rest of panels show median quantities for bins of 0.1 deg of the magnitude (b), colour (c), extinction in the G band (d, only for stars with $\varpi/\sigma_\varpi > 3$, see Sect. 2.4), `visibility_periods_used` (e), parallax (f), uncertainty in the proper motion in the δ direction (g), and zero point correction to the parallax ZP_{56} (h, see Sect. 3.1)

Table 2. Indicative completeness of the kinematic samples. Absolute number of stars and fractions for all magnitudes and for distinct magnitude ranges are given for the cases with 2p and 5/6p solutions and for the selection of $\varpi/\sigma_\varpi > 3$. To compute the percentages for DR2, the total number of sources in EDR3 for each case has been used. These numbers are for samples without the `excess_flux` and `RUWE` selections since these are not equally defined in the different releases and the selection in `RUWE` eliminates the 2p solutions.

	DR2			EDR3		
	2p	5p	(5p) /ALL-EDR3	2p	(5p∪6p)	(5p∪6p) /ALL-EDR3
$\forall G$	2 5564 48	10 750 864	76%	1 840 953	12 279 076	87%
$G < 19$	111 638	5 860 281	96%	67 217	6 010 199	99%
$19 < G < 20$	492 849	2 825 129	82%	60 360	3 369 371	98%
$20 < G < 20.7$	940 298	1944 641	64%	312 553	2 705 704	90%
	$\varpi/\sigma_\varpi < 3$	$\varpi/\sigma_\varpi > 3$	$\varpi/\sigma_\varpi > 3$ /ALL-EDR3	$\varpi/\sigma_\varpi < 3$	$\varpi/\sigma_\varpi > 3$	$\varpi/\sigma_\varpi > 3$ /ALL-EDR3
$\forall G$	10 662 298	2 645 014	19%	10 601 641	3 518 388	25%
$G < 15$	30 930	478 565	93%	13 098	500 913	97%
$15 < G < 17$	360 149	1 096 109	74%	170 528	1 301 329	88%
$17 < G < 19$	6 284 445	1 039 701	14%	5 860 892	1 660 388	22%
$19 < G < 21$	3 826 133	30 637	0.7%	4 232 117	55 142	1%

We note that 6p solutions tend to be associated to fainter sources and their astrometric solutions are worse than for 5p ones (they have on average fewer `visibility_periods_used`, i.e. less observations, and larger `ipd_frac_multi_peak`, i.e. relatively large probability of being a double source, either visual or real binary), having larger astrometric errors. While for the AC20 sample the fraction of 6p solutions is comparable to the 5p (42 and 45%, respectively, the remaining 13% being 2p), for the AC20-$\varpi/\sigma_\varpi > 3$ case they represent only a 14% (86% being 5p) since we require good relative parallax errors.

After selecting stars with $\varpi/\sigma_\varpi > 3$ (AC20-$\varpi/\sigma_\varpi > 3$ sample) we find approximately one milion more stars in EDR3 than in DR2 (bottom rows of Table 2), which represents an increase of 33%. Figure 3 (green solid and dashed histograms) shows an improvement of the completeness of the parallax quality selection at magnitudes fainter than $G = 16$, which means better sampling at all distances and probing larger ones. Table 2 also shows that at the relatively bright magnitudes $15 < G < 17$, there were 74% of stars in DR2 satisfying this condition but we have now 88%. It is nevertheless important to remark that the completeness of the sample with good parallaxes is low even at intermediate magnitude ranges both for the DR2 and EDR3 (as low as 14% and 22% in the range $17 < G < 19$, respectively), although we see an overall improvement for the new release.

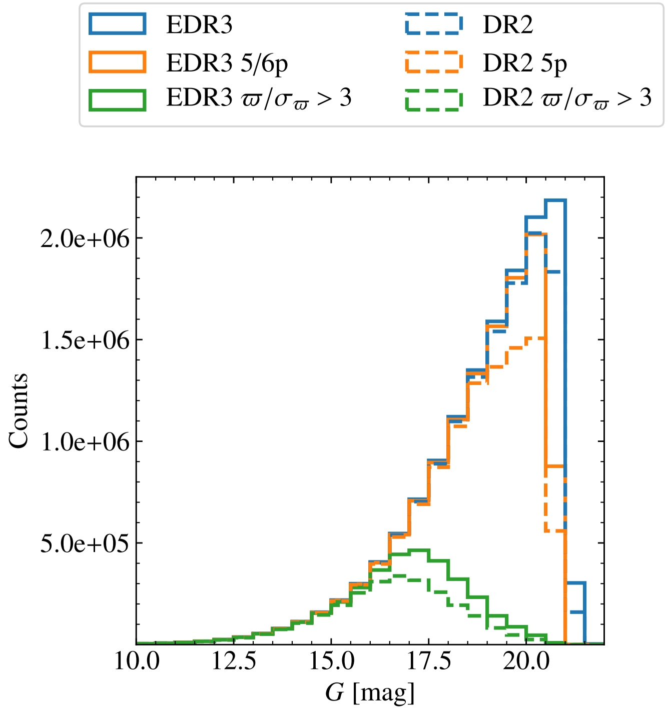

Fig. 3. Distribution of G magnitude in the anticentre. We show the number of stars in bins 0.5 mag for the AC20 case. The gain from DR2 to EDR3 is mostly at the faintest magnitudes where some sources did not have enough observations to appear in the past release.

4. Astrometric quality, systematics and parallax zero point.

The improvement in astrometric quality of EDR3 with respect to DR2 is discussed in Lindegren (2020a) and is reflected in smaller uncertainties and a reduction of the number of negative parallaxes (e.g. for the ACV sample where there are 164679 less sources with negative parallaxes, Table 1). Figure E.1 in Appendix E illustrates the improvement in the uncertainties for the anticentre (similar to figure A.1 by Lindegren (2020a) for all EDR3). There is a reduction by a factor of 0.79 and 0.5 in parallax and proper motion uncertainties, respectively, as expected for the increase in the number of months of observations, and even a larger reduction for sources with $G < 14$. This improvement in the astrometric quality allows us to have now a much larger sample of stars with very good relative parallax errors, and reach farther distances from the Sun. We also note that the uncertainty in μ_δ is smaller than for $\mu_\alpha*$. This is due to a geometrical scaling factor on the uncertainties that appears as a result of the scanning law which in the direction of the anticentre favours μ_δ.

As in previous releases, the astrometric *Gaia* data suffers from some systematics. The median astrometric quantities and their uncertainties show checkered patterns that somehow correlate with the scanning law, as illustrated for the median parallax (Fig. 2f) and median uncertainty in μ_δ (Fig. 2g). The later shows additionally some of the large scale bands mentioned above. The amplitude of these systematics has, however, been reduced in EDR3 (see Lindegren 2020a). Another known systematic is a zero point in parallax Lindegren (2020b) that has also been reduced and which we examine in detail in Sect. 3.1.

Figure 5 shows the differences in all astrometric quantities between DR2 and EDR3 normalised to the errors[2]. The median

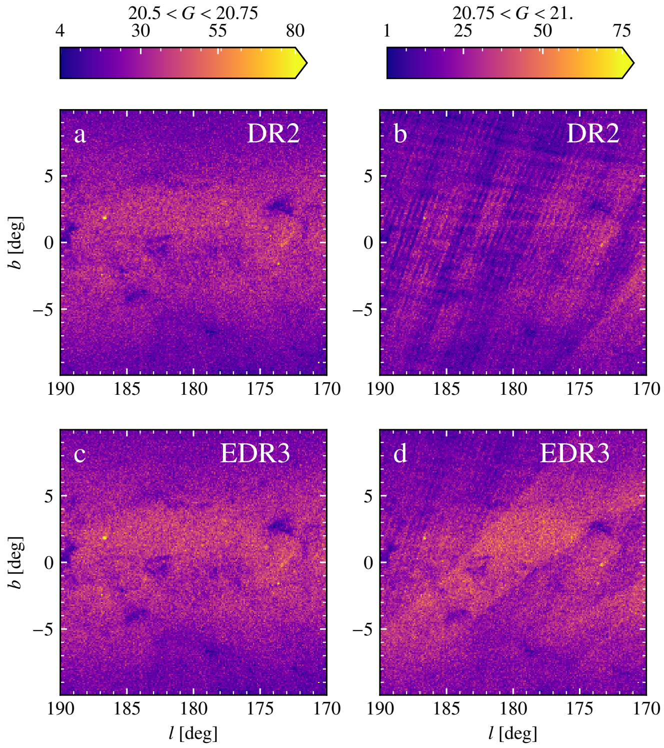

Fig. 4. DR2 and EDR3 counts for different magnitude ranges. The panels show the number of sources in bins of 0.1 deg in two different ranges of magnitude: $20.5 < G < 20.75$ (left) and $20.75 < G < 21.$ (right). To facilitate the comparison, the same colour bars has been used for each vertical pair of panels and the upper limit of the colour scale does not correspond to the maximum number of counts to avoid dominance of bins with clusters. An increase of the number of counts in EDR3 is observed, together with the decrease of some of the small scale patterns, although some bands remain in the faintest magnitude range.

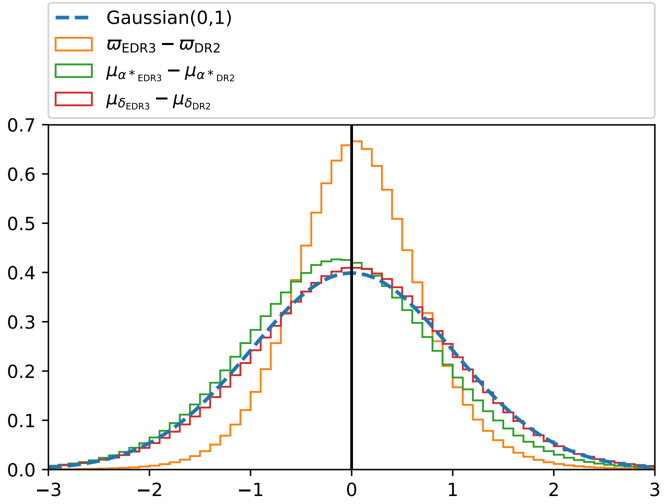

Fig. 5. Consistency between astrometric values in DR2 and EDR3. The histograms show the differences in parallax and proper motion normalised to the errors $(\boldsymbol{x}_{\mathbf{EDR3}} - \boldsymbol{x}_{\mathbf{DR2}})/\sqrt{\sigma^2_{\boldsymbol{x},\mathbf{EDR3}} + \sigma^2_{\boldsymbol{x},\mathbf{DR2}}}$, where x is ϖ, $\mu_\alpha*$ or μ_δ, and compared to a Gaussian distribution with 0 mean and variance of 1. The differences in $\mu_\alpha*$ are due to a systematic in DR2 that has been now corrected. We have corrected the parallaxes (Eq. 3) using the median offset for DR2 (-27 μas) and for EDR3 (-17 μas).

---

[2] We have used sources having the same `source_id` in DR2 and DR3. Even though we know from Gaia Collaboration (2020), Fabricius (2020) and Seabroke (2020) that some sources have changed

absolute differences between EDR3 and DR2 are $15\,\mu$as yr^{-1} in ϖ, $-48\,\mu$as yr^{-1} in $\mu_{\alpha*}$ and $7\,\mu$as yr^{-1} in μ_δ. For comparison, we show a Gaussian distribution with 0 mean and 1 as variance with a blue curve, although the quantities from DR2 and EDR3 are not independent and thus these distributions are not expected necessarily to follow this curve. The large systematic differences in $\mu_\alpha*$ (green histogram) is explained by a correction of the reference frame (spin) for EDR3 that has largely reduced the medium-scale (1-20 deg) inhomogeneities in the median parallax and proper motion of the quasars, which actually were quite large precisely in the direction of the anticentre for $\mu_\alpha*$ (about $0.1\,\mu$as, figures 10 and 11 of Lindegren 2020a). The histogram of parallax differences is narrower than the Gaussian case and is slightly positively biased. In Fig. 5 the zero point has been corrected using the median estimated values for quasars respectively in DR2 and EDR3 (more details are given in Sect. 3.1). We note that the bias was even larger if we neglected the corrections (giving a median of $25\,\mu$as). The persistent bias even after the correction could be attributed to underestimation of the zero point in DR2 (for which there is some evidence, see Lindegren 2020b) or effects of considering a fixed value of the offset (Sect. 3.1).

4. Photometric quality. The improvement of the photometry of EDR3 with respect to DR2 is described in Riello (2020). In summary, the increase of the number of observations and the improvement of several steps of the pipelines (image parameter determination, LSF/PSF calibrations, cross-match and photometry) have led to a significant decrease of the systematics at the bright end ($G < 15$). The effects of blends and contamination by nearby stars are mainly filtered out using Eq. 2.

2.4. Extinction and selection of tracer populations

We use two different approaches to select the tracer populations. The first one uses only *Gaia* data and the populations are selected in the de-reddened HR diagram. In the second approach, external photometric data is used to define a sample of RC sources.

Method I: Using the *Gaia* HR diagram. We obtain the de-reddened HR diagram for the AC20-$\varpi/\sigma_\varpi > 3$ sample using the 3D dust-reddening maps from *Bayestar* (Green et al. 2019). We use the d_{PM} distances (Sect. 3.2 and Appendix C.1) and the galactic (ℓ, b) coordinates to infer the line-of-sight visual extinction A_V for each source. Then we transform A_V to A_G and $E(BP-RP)$ using the *Gaia* extinction factor from Casagrande & VandenBerg (2018). We have compared our results with the ones using the extinction model by Lallement et al. (2019) and the conversion to A_G and $E(BP-RP)$ following Appendix A in Ramos et al. (2020b), finding only mild differences in the classification of stars without consequences on the conclusions from the subsequent sections.

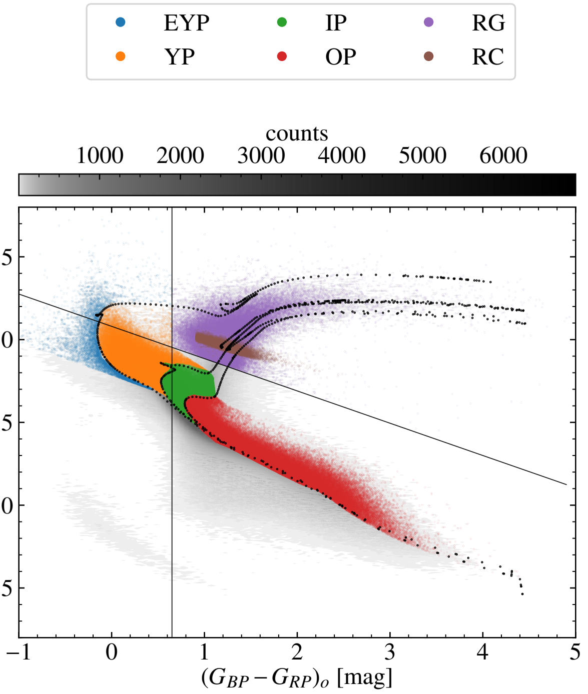

Fig. 6. De-reddened HR diagram of the anticentre region and different selected populations. The diagram is shown for the 3 369 456 sources of the AC20-$\varpi/\sigma_\varpi > 3$ sample with available photometry (G, BP, RP) and extinction data and absolute magnitudes derived considering the d_{PM} distances. We over-plot three `PARSEC` stellar isochrones with $[M/H] = 0$ for the ages of 0.2, 2 and 8 Gyr, a line at $BP-RP = 0.65$ and a diagonal line following the extinction slope used for the selection of populations (Method I) which appear in different colours. The RC have been selected using a different method (Method II).

Table 3. Populations in the AC20 sample. The selections are obtained following the Method I, except for the RC of the last row (Method II).

	Population	Sources
	All	3 369 456
EYP	Extremely Young Massive ($\tau \lesssim 0.2\,Gyr$)	12 792
YP	Young MS with $0.2 \lesssim \tau \lesssim 2\,Gyr$	276 344
IP	Intermediate MS with $2 \lesssim \tau \lesssim 8\,Gyr$	807 910
OP	Old MS	1 032 916
RG	Red Giants	176 193
RC	Red Clump	121 857

Figure 6 shows the Gaia de-reddened HR diagram for the AC20-$\varpi/\sigma_\varpi > 3$ sample. We use the direction of the extinction line in this diagram ($M_G = 1.95(BP-RP)_o - 0.8$, black diagonal line) combined with the vertical cut $(BP-RP)_o > 0.65$ to select (conservatively) giants (purple dots). Then we use the `PARSEC` isochrones (Bressan et al. 2012; Marigo et al. 2017) and respective updates[3] to perform a statistical partition by ages of the main sequence sources into extremely young (EYP, 0.2 Gyr), young (YP, 0.2 - 2 Gyr), intermediate (IP, 2 - 8 Gyr) and old populations (OP), as specified in Table 3 and illustrated in Fig. 6. The massive sources of the EYP are constrained to have $(BP-RP)_o < 0.65$, while the YP, IP and OP are selected between the lower (ZAMS) and the upper (TAMS) luminosity boundary of the main sequence band defined by the `PARSEC` stellar evolutionary tracks at $[M/H] = 0$. We note that while the OP has contribution from young stars, we can claim it is on average older that the IP: we expect an average age of 4-5 Gyr (e.g. figure 13 in Bland-Hawthorn et al. 2019) but with an important contribution of the oldest stars in the disc. In general, we expect these selections to be contaminated by stars of different ages due to several aspects (stars with different metallicities to the ones used in the

---

`source_id`, this occurs for a minority of cases and does not affect the results from Fig. 5

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

isochrones, inaccuracies of the extinction model used, confluence of isochrones around the ZAMS, binarity, etc). Nevertheless, we expect our samples to be dominated by the age ranges desired, which is enough for our basic purposes here.

Method II: Using Gaia & 2MASS. We combine EDR3 parallaxes and G-band photometry with that from *2MASS K*-band for the AC20 sample. The passbands in *2MASS* are narrow and in the infrared, and are thus weakly affected by errors in the extinction estimation. For *2MASS*, the flag (qfl) = 'AAA' indicates the highest photometric quality. However, this would significantly reduce our sample (to only 15% of the entire AC20 sample). Instead, we choose to enforce a quality cut only at the distance estimation stage using the photometric errors, *e_jmag* & *e_kmag*. We first compute the extinction of each source using the 3D dust-reddening maps from *Bayestar* (Green et al. 2019) with the inverse of the parallax as a prior for distance. The RC sources are selected in a Bayesian manner around the literature values for the absolute magnitude of the RC simultaneously for the *G Gaia* band and the *2MASS K* band. We have compared these distances with the ones using Bayesian distances for the prior (d_{PM}, see Sect 3.2) and our results do not change significantly. More details of the procedure and a validation with an external sample are given in Appendix B and C.2.1.

3. Distances and phase space coordinates

In this Section we describe how the distances and phase coordinates are computed in our study. We start by discussing the zero point in the *Gaia* parallaxes (Sect. 3.1), which needs to be corrected in order to estimate first distances (Sect. 3.2), and subsequently Galactic cylindrical positions and velocities (Sect. 3.3).

3.1. Parallax zero-point correction

As for previous releases, the *Gaia* parallaxes have a zero point[4] (ZP) that needs to be considered. In EDR3 the median parallax of the quasars is $-17\,\mu$as (Lindegren 2020b). This negative correction needs to be subtracted from the EDR3 parallaxes (parallaxes need to be increased by $\sim 17\,\mu$as):

$$\varpi^{\text{corrected}} = \varpi - ZP \tag{3}$$

or equivalently, reducing the inferred distance. Here we correct all parallaxes by subtracting $ZP = -17\,\mu$as. Additionally, when relevant, we also compare our results with the more sophisticated approach presented in Lindegren (2020b). In that work, they estimate the parallax zero-point ZP_{56} as a function of magnitude, colour (more precisely, `nu_eff_used_in_astrometry` for the 5p solutions and `pseudocolour` for the 6p solutions, hence the names ZP_5 and ZP_6), and ecliptic latitude, by looking at the parallaxes of quasars, binary stars and sources in the Large Magellanic Cloud for EDR3. Here we compute ZP_{56} using the Python implementation that is available online as part of the *Gaia* EDR3 access facilities[5]. Panel h in Fig. 2 shows the median zero-point ZP_{56} in the AC20 region. We observe a mild dependency of its value on the sky position.The median value for all stars in AC20 region is $ZP_{56} = -20\mu$as, thus similar to that of the quasars, with the 10 and 90 percentiles being -32 and -14 μas, respectively, and $ZP_{56} > 0$ only for 0.02% of the stars. For the AC20-$\varpi/\sigma_\varpi > 3$ sample, which has a significantly different magnitude distribution compared to the case without the parallax quality cut (Fig. 3), we find a median ZP_{56} of $-30\,\mu$as and percentiles of -38 and $-20\,\mu$as, respectively. The ZP_{56} case , thus, yields the largest differences between uncorrected and corrected distances (Fig D.1 in Appendix D). The velocities, which depend linearly on the distances, are consequently scaled as well. All these will be important in order to determine, for example, the exact distances to some kinematic features that we detect but we do not observe any qualitative difference in our results. More details are given in Appendix D and throughout the paper.

3.2. Distances

To convert the astrometric measurements by *Gaia* into phase space coordinates, we require an estimate of the distance to a given star. The complications of estimating distances to stars given their measured parallaxes have been discussed by a number of authors over a long period of time (e.g. Strömberg 1927; Lutz & Kelker 1973; Bailer-Jones 2015; Luri et al. 2018a). The transformation between parallax and distance is non-linear, which leads to a number of issues, including the extreme case of negative measured parallaxes. Simply taking the inverse of the measured parallax gives a biassed estimate of the distance of a star, and this bias grows more serious as the relative uncertainty grows larger. It has therefore become extremely common to apply a Bayesian approach to the problem of providing distance estimates from parallaxes, and/or to use photometric information to produce a better estimate of the distance.

We work primarily with distance estimates from a Bayesian approach (d_{PM}), similar to that applied by Schönrich & Aumer (2017), with a prior that is derived iteratively to be consistent with the data. These distances use a prior $P(d) \propto d^2 P_\rho(r(d))S(d)$, where $r(d)$ is the position at distance d along a given line-of-sight, so $P_\rho(r(d))$ is proportional to the density of a model Galaxy. The term $S(d)$ is the selection function – i.e. the probability that a randomly chosen star at a distance d enters the catalogue. The distance estimate, $\tilde{d}$ is then found as the expectation value of d given this prior and the measured parallax (and its uncertainty). As explained in the previous section, default distances are computed considering a fixed parallax zero point of $-17\,\mu$as. More details can be found in Appendix C.

To check that our results are robust, we compare to results when we estimate the distance as simply the inverse of the parallax, and also with a different Bayesian approach based on that from Bailer-Jones et al. (2018). We have tested each of these approaches on GOG data, and further details are given in Appendix C. From these tests, we conclude that using a parallax quality cut of $\varpi/\sigma_\varpi > 3$ is a good compromise between the performance of the estimate and the number of stars of our samples. However, we emphasise that all the estimators tested here return somehow imperfect distances, which in the Bayesian case depend also on how close the assumed prior on the Galaxy distribution is to the Galaxy model used in GOG (i.e. the Besançon Galaxy Model). We find that the median relative difference between the simulated true distances and the estimated ones can be as large as 20% at 4 kpc and larger than 50% for 25% of the stars even with the $\varpi/\sigma_\varpi > 3$ selection.

On the other hand, for the stars classified as RC (see Sect. 2.4), we infer their distance (d_{RC}) in a Bayesian manner using complementary photometric data from *2MASS* (details are given in Appendix C.2.1). Typical uncertainties are of 1 kpc at a distance of 6 kpc (see Fig. C.5).

---

[4] We use a different notation compared to Lindegren (2020b) to distinguish with the vertical cylindrical coordinate Z.

[5] https://gitlab.com/icc-ub/public/gaiadr3_zeropoint

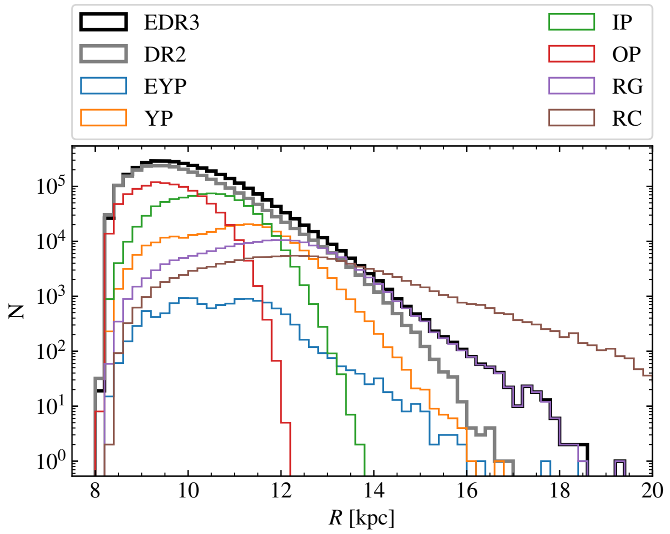

Fig. 7. Distribution of stars in Galactocentric radius. Number of stars per radial bin of 200 pc for the whole AC20-$\varpi/\sigma_\varpi > 3$ sample for EDR3 (black line) and DR2 (grey line), and for each stellar population (colour lines as indicated in the legend). The RC do not have the constraint $\varpi/\sigma_\varpi > 3$ and appears more extended in R than the black distribution.

Table 4. Uncertainties in phase space coordinates for the different *Gaia* releases. In the first rows we show the median uncertainties (first three numerical columns) and upper limit uncertainty for 80% of stars (three columns from the right) for stars in the AC20-$\varpi/\sigma_\varpi > 3$ sample for DR2 (first row), for the stars from EDR3 in common with DR2, and for EDR3. The last rows compare the heliocentric velocity uncertainties in DR2 and EDR3 for the sample with 6D velocities (6dsample) when the error in $v_{\rm los}$ is not (left) and is considered (right).

	AC20-$\varpi/\sigma_\varpi > 3$					
	median			80% of sources		
	ϵ_R	$\epsilon_{V^*_\phi}$	$\epsilon_{V^*_Z}$	ϵ_R	$\epsilon_{V^*_\phi}$	$\epsilon_{V^*_Z}$
DR2	0.30	3.8	2.2	<0.57	<8.4	<4.6
EDR3 ($\cap$ DR2)	0.18	2.4	1.3	<0.44	<5.9	< 3.1
EDR3	0.29	3.1	1.7	<0.58	<7.3	< 3.9
	6dsample					
	median $\epsilon_{v_{\rm los}} = 0$			median		
	ϵ_U	ϵ_V	ϵ_W	ϵ_U	ϵ_V	ϵ_W
DR2	0.09	0.09	0.07	0.43	0.44	0.38
EDR3	0.04	0.04	0.04	0.38	0.39	0.34

3.3. Positions and velocities

From the distances obtained in Sect. 3.2 and the sky positions, we compute the Galactic Cartesian (X, Y, Z) and cylindrical (R, ϕ, Z) positions, assuming that the Sun is located at $d_{\odot-GC} = 8.178$ kpc from the Galactic centre (Gravity Collaboration et al. 2019) and a height above the Galactic plane of $Z_\odot = 0.0208$ kpc (Bennett & Bovy 2019). Figure E.2 in Appendix E shows the spatial distribution of the the AC20 sample. By construction, the vertical Z and azimuthal Y distributions are wider for larger distances from the Sun, with some stars at $R \sim X = 14$ kpc reaching heights of 1 kpc above and below the plane. Figure 7 shows the number of stars as a function of Galactocentric radius. The gain in EDR3 for sources with $\varpi/\sigma_\varpi > 3$ (black line) compared to DR2 (grey line) at large radii is very significant with an increase of one order of magnitude already at 16 kpc and notably more stars at almost all radii.

For the different populations detailed in Sect. 6 (colour lines in Fig. 7), the samples with younger ages have distributions that, as expected, extend to larger radii compared to older populations. The distribution for the whole sample with $\varpi/\sigma_\varpi > 3$ is dominated by dwarfs for $R < 12$ kpc while giant stars take over beyond that. We see some hints of an over-density at around 12 kpc for the EYP and YP that could be the Perseus spiral arm but a good assessment of this requires more investigations of the selection function and the extinction. For the RC whose distances are computed photometrically without the $\varpi/\sigma_\varpi > 3$ constraint (brown line), there is a larger number of stars at large distances compared to the whole sample with $\varpi/\sigma_\varpi > 3$ (black line).

For the velocities, we compute V_ℓ and V_b and correct them for the reflex of the solar motion using the following equations:

$$V_\ell = k d \mu_\ell - U_\odot \sin(\ell) + (v_{c,\odot} + V_\odot)\cos(\ell) \quad (4)$$

$$V_b = k d \mu_b + [-U_\odot \sin(\ell) - (v_{c,\odot} + V_\odot)\cos(\ell)]\sin(b) + W_\odot \cos(b) \quad (5)$$

where $k = 4.7404705$ is the usual factor for units conversion, and we assume $U_\odot = 11.1$, $v_{c,\odot} + V_\odot = 248.5$, $W_\odot = 7.25\,{\rm km\,s^{-1}}$ for the solar motion (Schönrich et al. 2010; Reid & Brunthaler 2020), where $v_{c,\odot} \equiv v_c(R = R_0)$ is the value of the rotation curve at the Sun's position. In the anticentre direction, V_ℓ and V_b are approximately aligned with the usual cylindrical velocities V_ϕ and V_Z and, thus, we use:

$$V_\phi * \equiv -V_\ell \quad (6)$$

$$V_Z * \equiv V_b \quad (7)$$

We note that V^*_ϕ is positive for most of the disc stars with this definition and that V^*_ϕ is not exactly equivalent to V_ϕ, nor is V^*_Z to V_Z, due to a geometric difference in the vector orientation and the contribution of the line-of-sight velocity, but the differences are small in the anticentre. In the Appendix E we have used GOG to quantify this and we find that 80% of the sources with $\varpi/\sigma_\varpi > 3$ have absolute differences smaller than 2.9 and 3.3 km s^{-1} for $V_\phi *$ and $V_Z *$, respectively (Figs. E.4, E.5 and E.3). We see that V^*_ϕ is mainly smaller than V_ϕ with a median of -0.4 km s^{-1}. When examining how these differences are distributed in the ℓ-b projection, we see, as expected, larger differences in V^*_ϕ the farther from the exact anticentre line ($\ell = 180$ deg). The differences in V^*_Z show a quadrupole symmetry, indicating that any kinematic signature following this same shape in the sky would be clearly suspicious but that for most of the cases, since we average over the whole area, the global effect of these differences is null. For stars in the Gaia 6D phase space sample (thus a more realistic case), the differences are similar though slightly larger (80% of the stars with $\varpi/\sigma_\varpi > 3$ have absolute differences smaller than 3.2 and 4.0 km s^{-1} for $V_\phi *$ and $V_Z *$, respectively).

Another reference system for the velocities that we use in Sect. 5.1 is the tangential velocity V_t defined as:

$$V_t \equiv k d \sqrt{\mu_\alpha *^2 + \mu_\delta{}^2} \quad (8)$$

where in particular for that section we use as distances the inverse of the parallax with a more strict selection of $\varpi/\sigma_\varpi > 5$.

We have used the Jacobian matrix to compute the errors in the positions and velocities from the errors (and correlations) of the astrometric quantities. We neglect the errors in the angular positions since they are extremely small. In the case of the Bayesian and photometric distances, no correlation between distance d and proper motions μ has been considered (but see discussion Appendix C). Figure 8 shows the median uncertainty in the radius R (top) and velocities (bottom) as a function of R for EDR3 (solid lines) for the AC20-$\varpi/\sigma_\varpi > 3$ sample, and the area delimited by the 25 and 75 percentiles (shaded regions). The median errors in R (solid blue line) remain lower than 1 kpc for $R < 14$ kpc and the velocity uncertainties (solid orange and green lines) are smaller than 5 and 2 km s^{-1} for V_ϕ^* and V_Z^*, respectively, for most of the radii probed. The slight change of trend in the solid curves at around 12 kpc is due to the contributions of different stellar types, in particular giants stars that are intrinsically brighter at a given R and have, thus, smaller astrometric uncertainties. Table 4 gives a summary of these position and velocity errors: 80% of stars have errors < 0.6 kpc in Galactocentric radius, and < 7 km s^{-1} and < 4 km s^{-1}, respectively for V_ϕ^* and V_Z^*. We note that the errors for V_Z^* are smaller than for V_ϕ^* due to the better alignment of μ_b with $\mu_{\delta}*$ which in turn has smaller errors than μ_α in this sky direction as seen in Sect. 2.3 (Fig. E.1).

In Fig. 8 we also show the equivalent errors in DR2 (dashed lines). However, a fair comparison requires that we compare only the common sources (otherwise the new sources of fainter magnitudes in EDR3 at each bin in R contribute in a negative way to the overall values). The dotted lines obtained for the sources of EDR3 in common with DR2 show a quite significant improvement. For the velocities, the uncertainties are now smaller by about $\lesssim 2$ km s^{-1} at a Galactocentric distance of $R = 12$ kpc, which represents an improvement of 30%.

Figure 9 shows the full error ellipses for a few stars chosen to sample different values of R in the R-V_ϕ^* projection that we explore later. While the black error bars show the errors on the individual quantities, the error ellipses show large correlations between these two variables. This correlation is induced by the coordinate transformations, which in both axes have an approximately linear dependency with the distance error. As expected, the ellipses are all oriented pointing towards the position and velocity of the Sun assumed (indicated with a black star).

Finally, another good illustration of the improvement in the astrometry is the comparison of the uncertainties in the heliocentric velocities U, V and W for the 6dsample in DR2 (Katz et al. 2019) and EDR3 (Seabroke 2020), which is shown in the last three rows of Table 4. Assuming that there are no line-of-sight velocity uncertainties, the median uncertainties (left columns) are reduced by around 50% in EDR3. Including the line-of-sight velocity uncertainties (rightmost columns) does not show such a reduction, highlighting that the line-of-sight velocity uncertainties dominate. This will change in DR3 where these uncertainties are expected to decrease substantially and millions of additional sources will have line-of-sight velocity measurements for the first time.

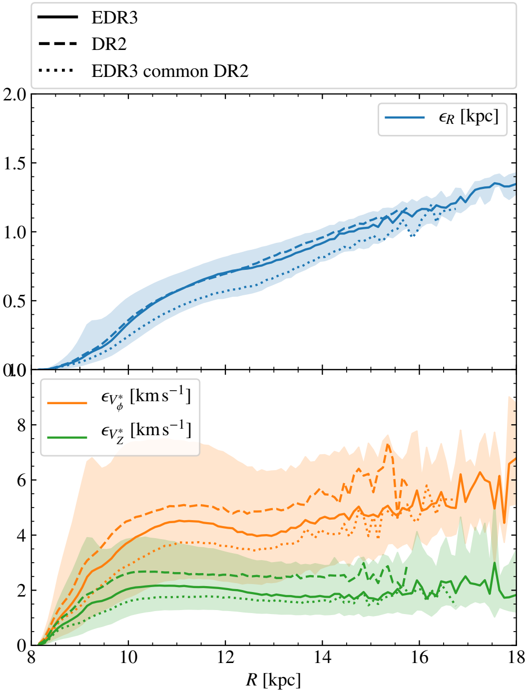

Fig. 8. Errors in phase space coordinates in the anticentre. The curves are for the AC20-$\varpi/\sigma_\varpi > 3$ sample and show the median errors for R (blue) in the top panel, and V_ϕ^* (orange) and V_Z^* (green) in the bottom panel, while the shaded regions show areas enclosing 50% of the stars (that is, limited by the 25 and 75 percentiles). We show the values for EDR3 (solid), DR2 (dashed) and sources in common (dotted).

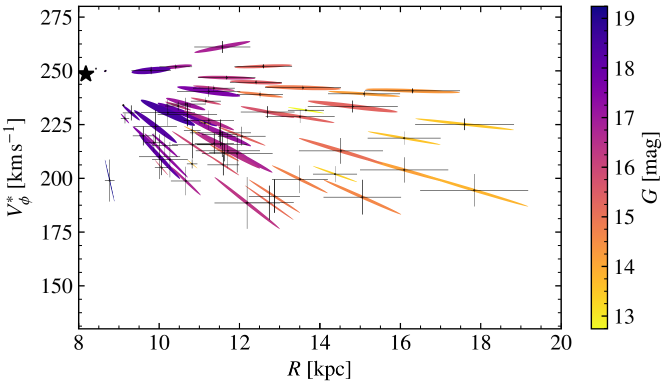

Fig. 9. Error ellipses in the R-V_ϕ^* plane for stars in the anticentre. The ellipses have been drawn for 60 stars from the AC20-$\varpi/\sigma_\varpi > 3$ sample chosen randomly but with weight of $\exp R$ in order to sample better the different R. The ellipses are coloured by magnitude G and the error bars are included as black lines. The error ellipses are oriented pointing towards the R_0-($v_{c,\odot} + V_\odot$) point (black star).

4. Disc kinematics

In this section we explore the dynamics of the Milky Way disc, analysing the velocities as a function of positions. As seen in Sects. 2.3 and 3.3, the improvement in the EDR3 astrometry allows us to probe the disc's outer regions. We start by examining the median velocities and velocity dispersions (Sect. 4.1) as a function of Galactocentric radius. We then look at large scale velocity asymmetries and phase space correlations in Sect. 4.2, to end with the analysis of small scale velocity substructures (Sect. 4.3) that are now resolved for the first time.

4.1. Azimuthal and vertical velocities and dispersions

We measure the median velocity profiles and dispersions of V_ϕ^* and V_Z^* for each stellar population. We use $\sigma * \equiv 1.5 MAD$ (where MAD is the median absolute deviation) as a robust estimate of the standard deviation, to which we subtract the median error in each bin in quadrature. Using a robust estimator, rather than the standard deviation, prevents outliers from producing a noisy dispersion profile. Although the 1.5 factor is strictly valid only for normally distributed data, this approximation puts our values on the same scale as the standard deviation for a more easy

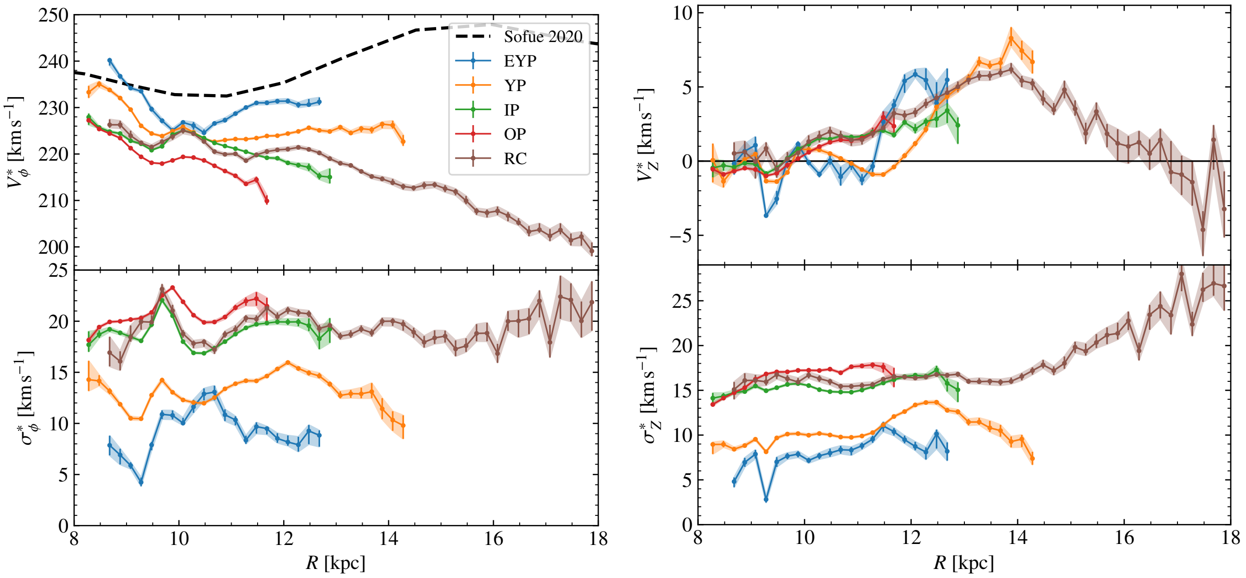

Fig. 10. Rotation and vertical velocity profiles in the anticentre. Top: Median azimuthal and vertical velocities of the populations EYP, YP, IP, OP and RC as indicated in the legends (same as in Fig. 7). Shaded areas represent the uncertainties (see text) but they are very small and barely visible in most of the cases. The rotation curve by Sofue (2020) is over-plotted in the top left panel. Bottom: Same as top but for the velocity dispersions (computed as 1.5*MAD* values). Apart from the expected differences due to the different ages of the populations and the asymmetric drift, we see significant oscillations in all curves.

comparison. We use bins of 200 pc and discard those with less than 100 stars. The uncertainties are then obtained by performing 1000 bootstrap resamplings of these distributions at each radius, choosing the 16th and 84th percentiles, and are indicated as shaded colour bands and error bars in the following panels.

The rotation velocity curves for the different populations are shown in the top left panel of Fig. 10. A difference in the median V_ϕ^* is observed for the different stellar populations with the older stars rotating slower as a result of the asymmetric drift. On average, the EYP stars rotate between 10 and ~20 km s^{-1} faster than the OP or the RC. The curve of the EY stars presents the best agreement with the rotation curve (black dashed lines) derived in Sofue (2020) from a compilation of kinematic data from molecular gas and stars in the infrared. This is consistent with the expectation that younger stars rotate as fast as the cold interstellar gas, thus at velocities closer to the true circular velocity of the Milky Way. Globally, all the rotation curves decline for $R \lesssim 9.5$ kpc and show a bump at around 10 kpc. Beyond this, the curve of YP stars is flat, while those of older stars decrease again.

The effects of the parallax zero point are examined in Appendix D where we show, as an example, the effects of the different adopted values of this offset on the rotation curve of our AC20-$\varpi/\sigma_\varpi > 3$ sample (Fig. D.2). As expected and discussed in Sect. 3.1, we see slight differences in the curves due to a decrease of the distance and scaling of the velocities when the correction is used. However, the general features of the curves remain the same.

Interestingly, we observe stars from the YP rotating as far as 14.5 kpc from the Galactic centre (see also Figs. 7, 12 and 16). In total, we find as many as 1186 stars with $16 < R < 18$ kpc and $V_\phi^* > 200$ km s^{-1} for the $ZP = -17\,\mu$as (with median uncertainties of $\epsilon_R \sim 1.$ kpc), and 275 for the case of ZP_{56}. This establishes a lower limit to the disc size although a more detailed analysis is required, in particular in the context of the biases of the distance estimators, which can be large at these distances (see Fig. C.4).

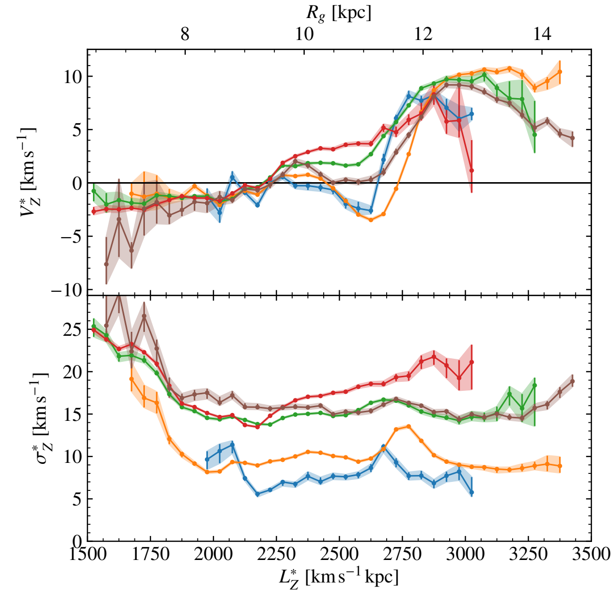

Fig. 11. Vertical velocity profiles as a function of angular momentum. The plot and legend is equivalent to the left panels of Fig. 7 but as a function of L_z^*. To guide the eye, we also show an approximate guiding radius $R_g = L_z^*/(236\,\mathrm{km\,s^{-1}})$ on the top axis.

The top right panel of Fig. 10 shows the median vertical velocities. These velocities appear to have small oscillations of the order of 2 km s^{-1} inside $R \sim 12$ kpc. There are clear dips at $R \sim 9.5$ kpc, coinciding with the dip in V_ϕ^*, and at $R \sim 11$ more prominent for the young stars. Figure 11 (top) shows the same velocities but as a function of L_z^* where the oscillations appear clearer with dips especially at around 2200 and 2600 $\mathrm{km\,s^{-1}}$ kpc. In the later one, we can see a strong age dependence with younger populations showing a deeper valley. We note that at the extremes of L_z^* in this plot, the populations are biased towards high eccentricity orbits that are those that manage to reach the observed volume. Beyond the location of the dips, V_Z^* increases and stars move in median upwards ($V_Z^* > 0$). The profile of the RC stars with R (and of all populations in the L_z^* plot) draw a clear wiggle (with a subsequent decrease), with maxima of $\sim 5\,\mathrm{km\,s^{-1}}$ at $R \sim 14$ kpc ($\sim 10\,\mathrm{km\,s^{-1}}$ at $L_z^* \sim 3000\,\mathrm{km\,s^{-1}}$ kpc). In Fig. 14 of Gaia Collaboration et al. (2018c) only the first part of this positive vertical velocity wiggle was observed and seemed to have certain dependencies on the Galactic azimuth ϕ and vertical position Z of the stars, as we will confirm in Sect. 4.2. The oscillations and the outer increase of the vertical velocities as a function of angular momentum were also observed in Schönrich & Dehnen (2018), Huang et al. (2018) and Cheng et al. (2020).

The bottom panels of Fig. 10 show the diversity in the velocity dispersions σ_ϕ^* and σ_Z^* in the Galactic anticentre direction. Although we expect decreasing dispersions with R (van der Kruit & Freeman 1986, and references therein) supported by observations in external galaxies (Martinsson et al. 2013) including the Large Magellanic Clouds in the *Gaia* data (Luri 2020), the general behaviour here shows bumpy dispersions in all the populations that correlate with the oscillations in the median velocities.

Apart from the oscillations, overall we observe dispersions that are quite flat as a function of R, and even increasing at larger radii for RC stars. For the vertical velocity dispersion as a function of angular momentum (bottom panel of Fig. 11) the oscillations are even clearer. In the inner parts the dispersions decrease with Lz but this could be due to the selection effects explained above, while the profiles are overall flat in the outer parts. We note that the geometry of our AC20-$\varpi/\sigma_\varpi > 3$ samples have larger ranges of Z for increasing R (Fig. E.2). This together with a complex selection function in the more distant regions and the approximation in the velocities of Eqs. 6 and 7 could produce artificial trends in the velocity dispersion. A similar flattening of the vertical velocity dispersion outside the solar radius was observed in Sanders & Das (2018) where the authors also discuss different biases that could explain this behaviour but also the possibility of being related to the flare (see also Mackereth et al. 2019; Sharma et al. 2020).

As for the amplitude of the dispersions, younger stars unsurprisingly present lower velocity dispersions than more evolved stars, most likely because these populations have not had the time to be heated by various internal and external processes, unlike older populations. On average, the azimuthal and vertical velocity dispersions of the EYP stars are around 8 and 5 $\mathrm{km\,s^{-1}}$ lower than those of old stars, respectively for V_ϕ^* and V_Z^*.

The flattening of the velocity ellipsoid σ_Z^*/σ_ϕ^* inside $R \sim 14.5$ kpc is quite homogeneous among the various populations and within the whole sample, all of them showing an azimuthal dispersion larger than the vertical component ($\sigma_Z^*/\sigma_\phi^* = 0.7\text{–}0.8$, on average). The vertical random motion only exceeds the azimuthal component for RC stars beyond 15 kpc, and for EYP stars at $R = 9$ kpc and after 11 kpc. Interestingly, the random motions of the EYP stars with dispersions of values of 9 and 8 $\mathrm{km\,s^{-1}}$ for σ_ϕ^* and σ_Z^* on average, respectively, are comparable to the typical velocity dispersions seen in the gas (~ 9 and 4.5 $\mathrm{km\,s^{-1}}$ respectively for neutral atomic and molecular gas for $R < 8$ kpc, Marasco et al. 2017), for a gas velocity ellipsoid assumed isotropic. Thus, in 200 Myr (the maximum age of the EYP), the youngest stars present already a slight velocity anisotropy as expected since they have oscillated of the order of one-two vertical periods.

4.2. Velocity correlations and asymmetries

We study here kinematic differences as a function of the location with respect to the Galactic mid-plane, and other phase space correlations. First, we compare the kinematics of $Z < 0$ stars with those at $Z \geq 0$ for the whole AC20-$\varpi/\sigma_\varpi > 3$ sample (Fig. 12). There is a notable asymmetry in the median velocities and the velocity dispersions (Fig. 12), starting approximately at 10-11 kpc, thus coinciding with the starting position of the large vertical velocities of Fig. 10. The rotation of $Z < 0$ stars (blue curves) clearly leads that of stars at $Z \geq 0$ (orange) beyond $R \sim 11$ kpc typically by up to $10\,\mathrm{km\,s^{-1}}$. A significant asymmetry is also seen for $R > 10$ kpc in the vertical motion where stars at $Z < 0$ move at larger velocities than $Z \geq 0$ stars, with a difference of up to $\sim 6\,\mathrm{km\,s^{-1}}$ (already noticed in Gaia Collaboration et al. 2018c and Wang et al. 2020b for example). The asymmetries in V_Z^* start close to the Sun, though with opposite trend compared to $R > 10$ kpc. The azimuthal dispersions are comparable at lower radii but asymmetric beyond $R \sim 10.5$ kpc (larger values for $Z < 0$ stars, by up to $5\,\mathrm{km\,s^{-1}}$) and reversing beyond 13 kpc. There is also a vertical velocity dispersion asymmetry but it is weaker ($\lesssim 1\,\mathrm{km\,s^{-1}}$). In any case, the dispersions observed correspond to the typical thin disc velocity dispersions (e.g. Robin et al. 2003; Aumer & Binney 2009).

We now follow up these asymmetries by looking with more detail at the density of stars in the V_ϕ^*-V_Z^* plane. We show the counts in this projection in 1 kpc-wide radial bins for Galactocentric distances ranging from 10 kpc to 15 kpc, and for the north ($Z > 0$, top) and south ($Z < 0$, bottom) Galactic plane (Fig. 13). One of the clearest features in Fig. 13 is the lack of symmetry for stars above and below the plane. Secondly, for the bins at $R > 12$ kpc we observe a bimodality where stars are sitting mainly in two clumps, one with negative V_Z^* at lower V_ϕ^*, which is more prominent in the north, and one with positive V_Z^* at higher V_ϕ^*, more conspicuous in the south. The different proportions of the clumps of the bimodality at different Z seems to be the cause of the vertical asymmetries seen at the top panels of Fig. 12, moving the median velocities to higher/lower V_ϕ^* and higher/lower V_Z^*. However, we emphasise that the bimodality appears on both sides of the disc, just in different ratios.

Figure 14 shows other phase space projections, allowing us to study this phenomenon in a more continuous way: the top panels show V_ϕ^* as a function of R, color-coded by either the Z position (a), the median V_Z^* (b) and the dispersion $\sigma_{V_Z^*}$ (c). At R>11 kpc, the population having large V_ϕ^* ($\sim 30\,\mathrm{km\,s^{-1}}$ larger than the other group) and positive V_Z^* ($\sim 10\,\mathrm{km\,s^{-1}}$, blue colours in panel b) is predominantly at negative Z (red colours in panel a), and vice-versa for the population having smaller V_ϕ^* and negative V_Z^* (of about -2 to -5 $\mathrm{km\,s^{-1}}$), as seen before. We note that the bimodal kinematic behaviour is also present at smaller R but with smaller amplitude in the vertical velocities. In fact some hints of this bimodality were seen in the maps of e.g. Khanna et al. (2019a) and Wang et al. (2020a) but those reach only 12 kpc from the Galactic centre and the bimodality appears marginally at the borders of their distributions. Additionally, we note now a clear spatial evolution, with the region at large V_ϕ^* and positive V_Z^* smoothly diminishing its V_ϕ^* when R increases. A line of constant angular momentum $L_z^* = RV_\phi^* = 2750\,\mathrm{km\,s^{-1}}$ kpc has

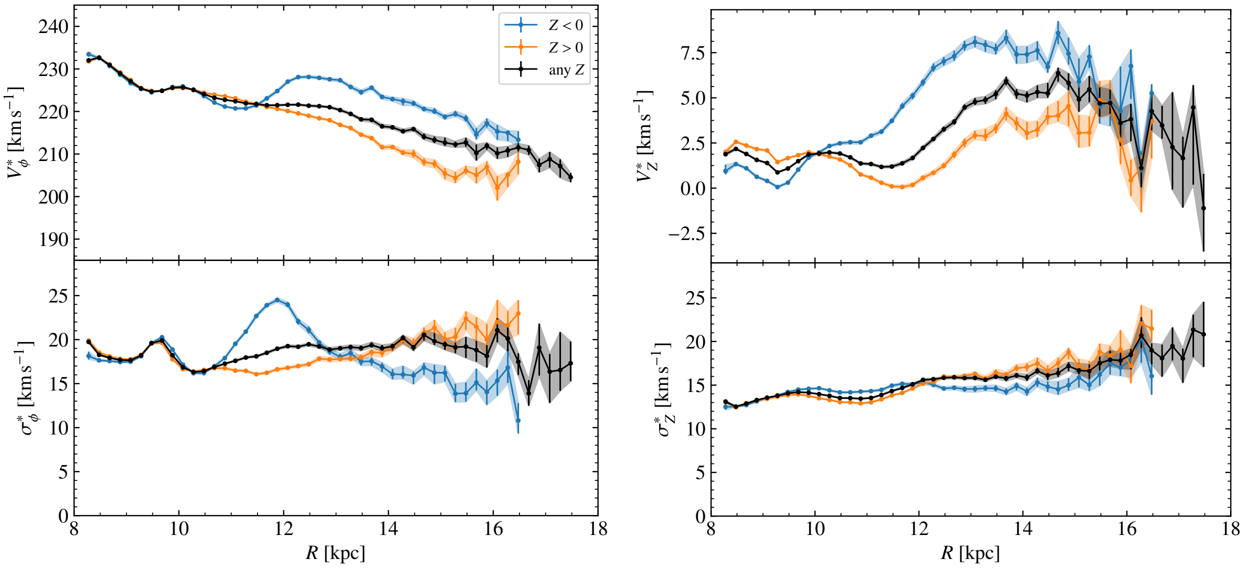

Fig. 12. Comparisons of the velocities above and below the Galactic plane. We plot the median azimuthal and vertical velocities (top) and velocity dispersions (computed as 1.5*MAD* values, bottom) for the whole sample AC20-$\varpi/\sigma_\varpi > 3$ (black lines, not including the sample of RC with photometrically derived distances), and for stars with $Z \geq 0$ (orange lines) and for $Z < 0$ stars (blue lines). Shaded areas represent the uncertainties. We observe notable asymmetries beyond 10-11 kpc.

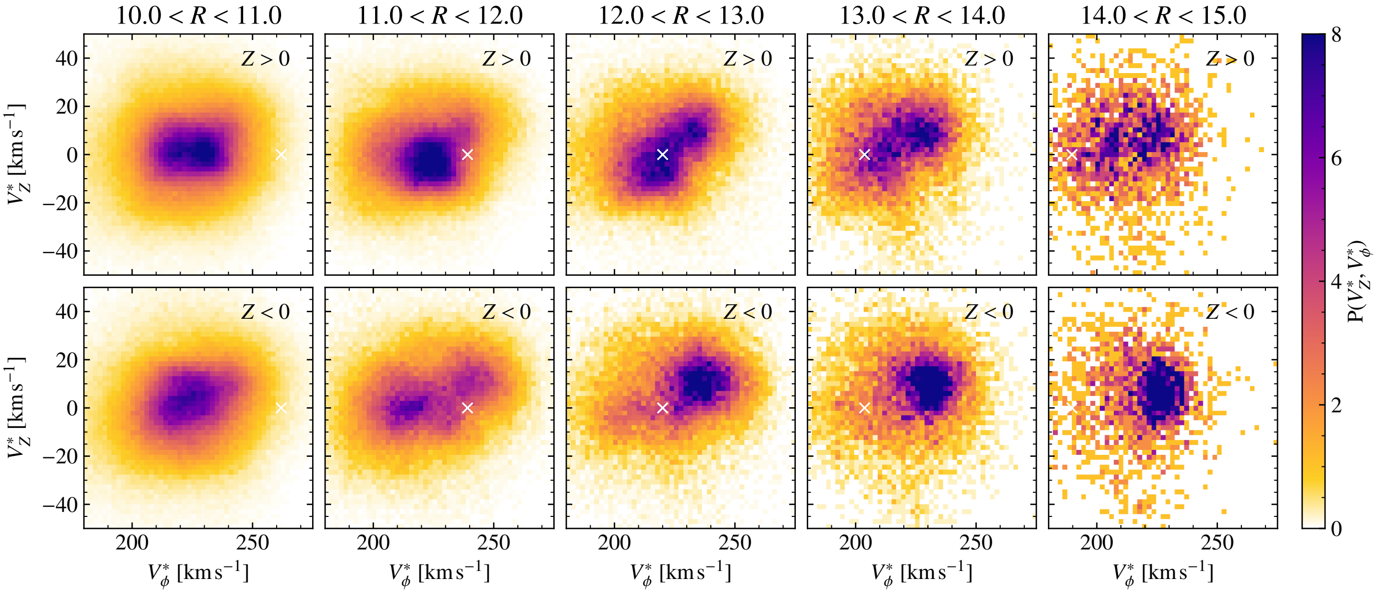

Fig. 13. Density in velocity space at different distances above and below the plane. Stellar density in the $V^*_\phi - V^*_Z$ plane, for bins in R from 10 to 15 kpc for the AC20-$\varpi/\sigma_\varpi > 3$ sample for $Z > 0$ (top) and $Z < 0$ (bottom). We see division into two components in the outer radial bins. To guide the eye a cross has been placed at V^*_ϕ corresponding to $L^*_z = 2750$ km s^{-1} kpc for a point in the centre of the radial bin (see also Fig. 15).

been plotted that roughly marks the transition in the sign of V^*_Z in panel b. There is not an exact match between the transition zone in panels a and b, indicating that the dominance of one clump over the other does not occur exactly at $Z = 0$. We note that the velocity dispersion of both groups of stars is typical of the thin disc ($\sigma_{V^*_Z} \sim 15$ km s^{-1}), as already inferred from the bottom-right plot of Fig. 12. We also see that the dispersion profiles of panels c and f are vey different from the ones for the GOG mock *Gaia* data in Fig. E.6.

The phase spiral identified in Antoja et al. (2018) with DR2 data for stars in the immediate Solar vicinity (within $R_0 \pm 200$ pc), is illustrated in panel d of Fig. 14 now with astrometry from EDR3. The morphological change of this phase spiral (or more precisely a slice of it, centred around $Z \sim 0$ kpc, highlighted with brighter colours in panel d) is traced for radii between 8 and 10 in panel e, by plotting V^*_Z as a function of R color-coded by V^*_ϕ. Up to $R \sim 10$ kpc, one can still see the different arms of the phase spiral at positive and negative V^*_Z, with a diminishing envelope as one moves outwards, due to smaller restoring forces. While this has been observed already in Laporte et al. (2019b) for discrete ranges of R, we see it here in a continuous way. However, farther out than $R \sim 11$ kpc, we see a clump (red colours) of large V^*_ϕ and positive V^*_Z dominating, which corresponds to one

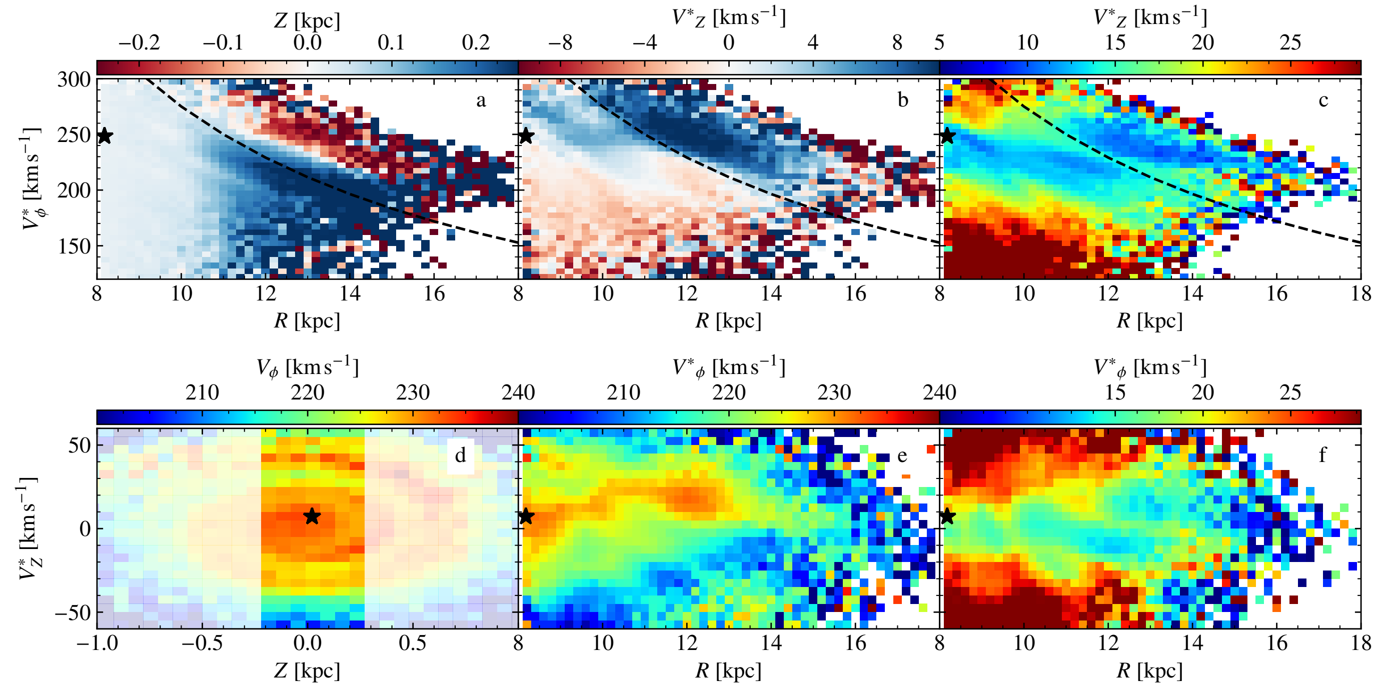

Fig. 14. Phase space projections of the Galaxy disc. The plots show for the AC20-$\varpi/\sigma_\varpi > 3$ sample: a) median Z coordinate in the R-V_ϕ^* plane; b) median vertical velocity V_Z^* in the same projection; c) dispersion in the V_Z^* velocity in the same projection (computed as the 1.5MAD); d) phase spiral in the 6dsample in EDR3 for stars in the Galactic radial range $|R - R_0| < 0.2$ kpc; e) median azimuthal velocity V_ϕ^* in the R-V_Z^* plane; f) dispersion in V_ϕ^* in the same plane. The bimodality appears in the outer parts of the disc in panels a, b and c, with the separation marked with a line of constant angular momentum $L_z^* = 2750\,\mathrm{km\,s^{-1}}$ kpc. In panel e, the evolution of a slice of the phase spiral (marked in brighter colours in panel d) is seen for smaller radii, while a signature related to the above bimodality is seen beyond ~ 12 kpc in panels e and f.

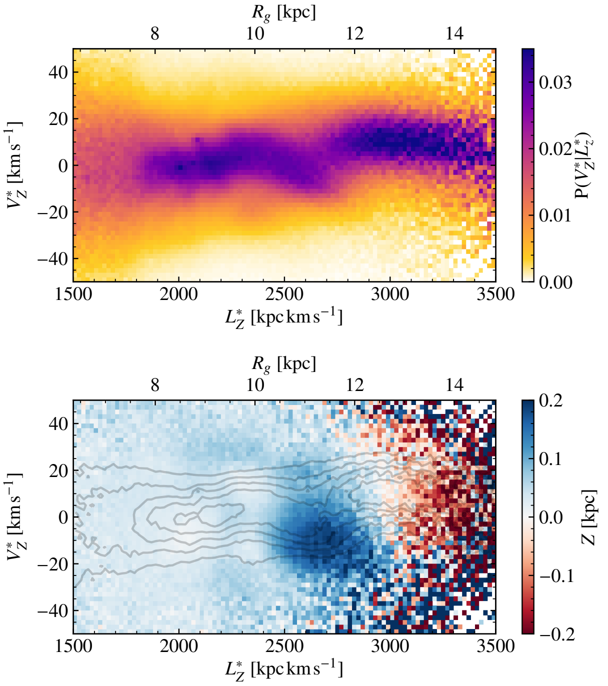

Fig. 15. Structures in vertical velocity and angular momentum. Top: Column normalised histogram of star numbers in the L_z^*-V_Z^* plane for the AC20-$\varpi/\sigma_\varpi > 3$ sample (the colour represents the fraction of stars in a given L_z^* bin that have a certain V_Z^*). Bottom: Average Z of stars in each bin in L_z^*-V_Z^* in our AC20-$\varpi/\sigma_\varpi > 3$ sample. Contours are the same as the colour plot in the top panel. To guide the eye, we also show an approximate guiding radius $R_g = L_z^*/(236\,\mathrm{km\,s^{-1}})$.

of the modes of the bimodality discussed above. Whether this is a manifestation of the same spiral (but blurred since we are now considering a wide range of Z given the cone geometry of the sample and because of the errors), another phase spiral at larger radius or a different phenomenon such as the warp –perhaps with the same origin– is not clear at this point.

In Appendix E we repeat some of the plots presented thus far for the GOG and UM samples (Fig. E.6). From those we conclude that selection effects due to extinction can induce some features in projections such as R-V_ϕ^* coloured as a function of Z. This is because a different extinction below and above the plane favours distinctly the different types of stars (different ages) that have different asymmetric drift (thus different V_ϕ^*) creating correlations between these variables. However, we do not observe any special vertical kinematics for these features in the mock data. We have also checked that the effects of the zero point in parallax does not induce or remove the features observed but merely change the distance scale with the pattern arriving farther or closer, independently whether a constant ZP or ZP_{56} is used (Fig. D.2 in Appendix D). Moreover, these features preferentially occupy positive or negative Galactic latitudes but do not correlate with the smaller scale checkered patterns seen in the astrometry. We note also that the stars participating in this phenomenon are relatively bright stars (Fig. E.7 in the Appendix E), thus with good astrometry in general. Also the difference of $10\,\mathrm{km\,s^{-1}}$ seen in the velocities of the two distinct features mentioned above which are at a typical distance of 4 kpc, correspond to a proper motion difference of around 0.5 $\mathrm{mas\,yr^{-1}}$, which is much bigger than any known systematics.

Finally, Fig. 15 shows the angular momentum-vertical velocity (L_z^*, V_Z^*) space, coloured by density in V_Z^* at each L_z^* (top) and average Z coordinate (bottom). In this plot, we see oscillations in

V_Z^* for the smaller L_z^* (better seen in the top panel of Fig. 11, and seen also in Huang et al. 2018 and Cheng et al. 2020) that likely correspond to the vertical oscillations also seen in the top right panel of Fig. 10 at nearby Galactocentric radii. Most notably, these plots show that the clumpy features seen for $R > 11.5$ kpc in Fig. 13 correspond to a clear break in the (L_z^*, V_Z^*) density at $\sim 2750\,\mathrm{km\,s^{-1}}$ kpc rather than a smooth transition. We note that when we separate our sample into young population (YP+EYP), Main Sequence (IP+OP) or Giants (Fig. E.8 in Appendix E), this trend is seen for all the populations (albeit most clearly in the young one, as it has the lowest velocity dispersion) implying that this break is most likely of dynamical origin. In particular, the change in proportions between the two populations that we see in Fig. 13 as we move outwards is related to the fact that the population with $L_z^* \lesssim 2750\,\mathrm{km\,s^{-1}}$ kpc (and $V_Z^* < 0$) does not reach as large radii as the population with $L_z^* \gtrsim 2750\,\mathrm{km\,s^{-1}}$ kpc (and $V_Z^* > 0\,\mathrm{km\,s^{-1}}$). According to the bottom panel of Fig. 15, and as seen above, the part of the disc at lower angular momentum L_z^* corresponds to stars predominantly at positive Z while the one with higher L_z^* mostly has negative Z, though without perfect one-to-one correlation.

4.3. Small scale velocity structures

Apart from the two clumps discussed in Sect. 4.2, finer substructures in the phase space of the disc can already be seen in the top panels of Fig. 14 for nearby radii. These structures are better visualised in Fig. 16 showing the 2-dimensional histogram of the V_ϕ^*-R projection (panel a). Diagonal ridges, i.e. substructures with decreasing V_ϕ^* as a function of R, can be seen, as already discovered in the *Gaia* DR2 (Antoja et al. 2018; Kawata et al. 2018). To enhance the contrast of these substructures, in panel b we show the density relative to a smoothed density obtained from a Gaussian filter $\frac{N-N_s}{N_s}$, where N are the counts and N_s are the smoothed counts with a $\sigma = 5$ times the bin size (similar to what is done in Laporte et al. 2019b). Panels c to e show this relative density for different stellar types. We do not note any difference between using $ZP = -17\,\mu$as and Z_{56} except for the already mentioned distance scaling.

The location of the main ridges obtained in Ramos et al. (2018) with the DR2 *Gaia* RVS sample are over-plotted with colour lines in Fig. 16b-e. Following their notation, we can identify the ridges associated to Hercules, Hyades, L18 (with a different slope compared to the rest) and one that could be linked to L16 or the so called hat (e.g. Gaia Collaboration et al. 2018c, $V \sim 40\,\mathrm{km\,s^{-1}}$ on their figure 22) - also related to L14 and L17.

Interestingly, for the YP the Sirius ridge appears to have slightly higher V_ϕ^* velocities than the marked ridge (red line), as if following the asymmetric drift relations, and the ridges look thinner than in the RG or RC plots. We estimate the fraction of stars forming the ridges by calculating $\frac{\sum (N-N_s)>0}{\sum N_s}$. These fractions are 30%, 13%, 8%, 8% for the EY, YP, IP and OP, respectively. The fractions are 11%, 14% and 8% for the RG, RC and all AC20-$\varpi/\sigma_\varpi > 3$ stars, respectively. This fraction depends on the σ used to smooth the distribution but the relative trends are the same, from which we see that the younger the population, the higher the fraction of stars in substructures. On the other hand, we do not have enough stars in the lower V_ϕ^* region in any of the populations to notice the low angular momentum ridges suggested in Laporte et al. (2020b).

More importantly, in Fig. 16 the ridges are now seen at much larger distances than before. The Sirius ridge is detected up to $R \sim 12.5$ kpc, while in *Gaia* DR2 a sophisticated method to detect very low contrasts was needed to reach even $R \sim 11$ kpc (Ramos et al. 2018). We can also spot three ridges that reach outer regions of the disc, up to 16 kpc and beyond in the case of the RC. The one at lowest V_ϕ^* could be the extension of L16. The other two were previously unknown and have been marked with arrows in the bottom panel (new anticentre ridges 1 and 2). The new structures do not point towards $V_\phi^* \sim v_{c,\odot} + V_\odot$ and $R = R_0$ (black star in the panels) as expected for structures stretched by errors in distance (see Fig. 9). In addition, we do not see any similar ridge induced by selection effects, uncertainties, or extinction, in the GOG equivalent sample.

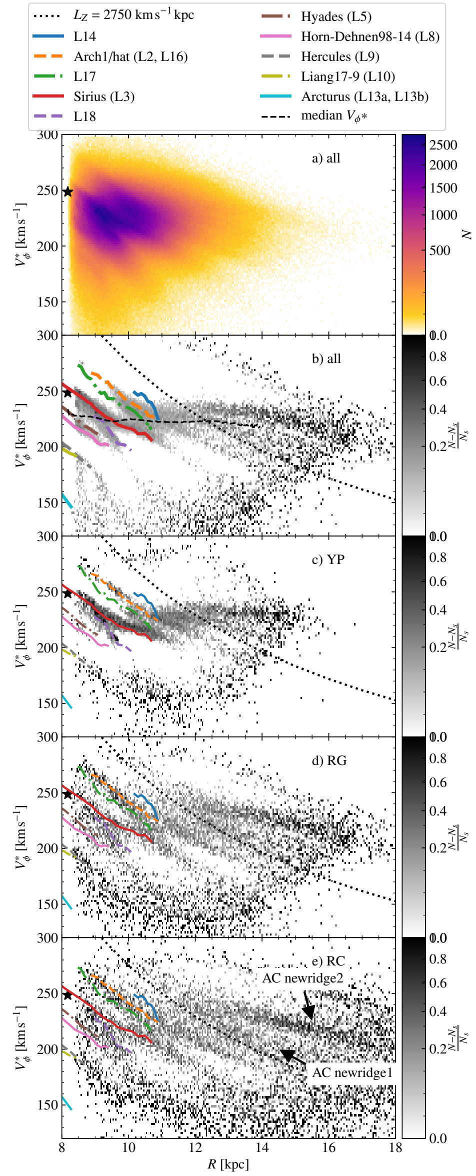

Fig. 16. Substructures in the R-V_ϕ^* projection in the anticentre direction. a) Number counts in the R-V_ϕ^* plane in bins of size $\Delta R = 0.02$ kpc and $\Delta V_\phi^* = 1\,\mathrm{km\,s^{-1}}$ for all stars in the AC20-$\varpi/\sigma_\varpi > 3$ sample. b) Same but applying a substructure mask to highlight the ridges (see text). c-e) Same as b but for different stellar types. We also plot: some ridges from Ramos et al. (2018) with coloured lines, the separation of the bimodality (black dotted line), and the median velocity (black dashed line in panel b). We see the ridges extending beyond their limits in DR2 and new ridges resolved here for the first time.

In the panels b to e of Fig. 16 we also plot the line of angular momentum $L_z^* = 2750\,\mathrm{km\,s^{-1}}$ kpc (dotted black line) which marks the approximate separation of the bimodality described in Sect. 4.2. While this line seems to coincide with the new anticentre ridge 1 (especially in panel e), no dynamical connection is clear at this stage. The median rotation velocities from Fig. 10 are over-plotted as a black dashed line in panel b and we see that the bump at around 10 kpc seems linked to the appearance of the L16 ridge that, with higher V_ϕ^*, moves the median curve slightly upwards. The connection between ridges and bumps in the rotation curve was already suggested by Martinez-Medina et al. (2019, 2020). The bump at 13 kpc could also be connected to the new anticentre ridge 2.

In Fig. 14a, we see some correspondence between the median Z and the density ridges seen in Fig. 16 (e.g. the white ridge in panel a with lower median Z overlaps with the Sirius ridge). Similarly, in Fig. 14b the ridges exhibit a complex pattern of positive and negative vertical velocities, thus indicating coupling between in-plane and off-plane kinematics. These effects were also noticed in *Gaia* DR2 with the RVS sample (Khanna et al. 2019b; Laporte et al. 2019b), where the ridges were stronger at lower Z and had some amplitude in V_Z^* though typically lower than $5\,\mathrm{km\,s^{-1}}$.

5. Halo, thick disc and distant structures

r In this section, we investigate several constituents of the Galaxy through the powerful combination of *Gaia* astrometry and photometry. In Sect. 5.1 we look at the stars of high tangential velocity which are contributed by the halo and the hot thick disc and secondly, in Sect. 5.2 we explore the structures in the outer parts of the Galaxy disc.

5.1. Halo and thick disc

Our goal in this section is to establish the extent and properties of the stellar halo populations beyond the solar vicinity and towards the galactic anticentre.

To enhance the contribution of halo stars and partially mitigate the effects of high-extinction near the disc plane, we use the ACV sample, defined in Sect. 2.2, with an additional selection of $|b| < 40°$. We select on $\varpi/\sigma_\varpi > 5$ and compute distances as the inverse parallax. Since we are interested in precise intrinsic colours and magnitudes, we choose only stars that have G-band extiction $A_G < 1.0$. Here the extinction is computed using the Schlegel et al. (1998) maps (with the correction of Schlafly & Finkbeiner 2011) and a Cardelli et al. (1989) extiction curve with $R_V = 3.1$. Although this extinction correction does not yield intrinsic magnitudes as accurate as in Appendix B, the main goal here is simply to remove high-extinction regions from our analysis, while producing accurate enough colours at large distances.

Following the approach of Gaia Collaboration et al. (2018a), we focus on the HR diagram for stars in the ACV sample that pass the cuts described above. We find that when selecting only stars with high heliocentric tangential velocity V_t (Eq. 8), two sequences arise as shown in Fig. 17, and that at the value of $\sim 150\,\mathrm{km\,s^{-1}}$ both sequences seem to be found in equal numbers around the main-sequence turn-off point. For completeness, see also our animation of how the HR diagram varies as V_t is increased in 5 $\mathrm{km\,s^{-1}}$ slices that will be available online. When V_t is low, there is a significant contribution from the thin and the canonical thick discs, whereas at $V_t \gtrsim 250$ km/s mainly the blue sequence (that locally is dominated by stars from the accreted Gaia-Enceladus-Sausage) is apparent.

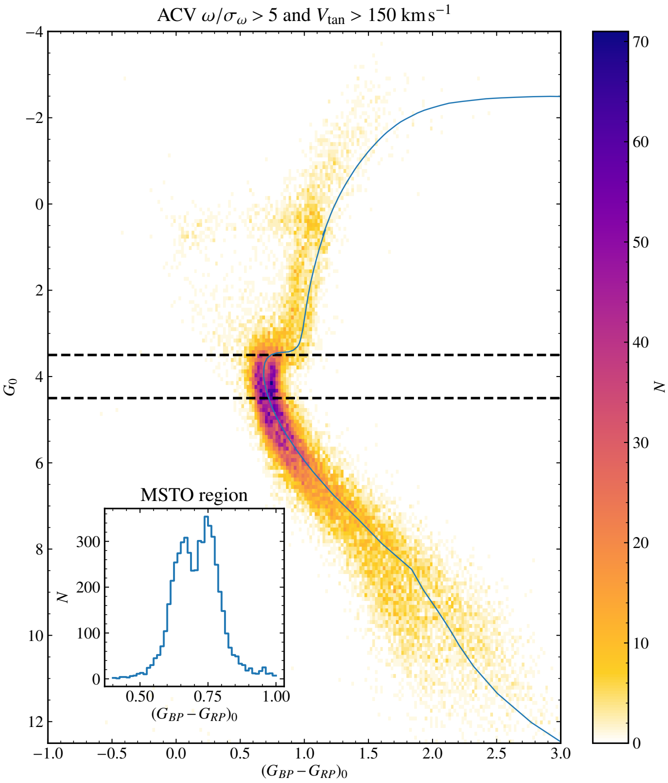

Fig. 17. Red and blue sequences for high tangential velocity stars. $G_{\mathrm{RP}} - G_{\mathrm{BP}}$ vs G Hess diagram for the ACV sample with $|b| < 40°$ and with a $\varpi/\sigma_\varpi > 5$ and $V_t > 150\,\mathrm{km\,s^{-1}}$. A `PARSEC` isochrone with [M/H] = −0.5 and age of 11 Gyr is shown in blue (but shifted by 0.04 in colour and 0.2 in magnitude in order to match the gap between blue and red sequences). The inset histogram shows the colour distribution in the magnitude range of the MSTO (shown as dashed lines in the main figure). A clear separation in two sequences is clearly seen as originally noted in Gaia Collaboration et al. (2018a) with DR2 data. An animated version of this figure for varying V_t limits will be available online.

In Fig. 17 the double sequence extends beyond the turn-off point, but with fewer luminous stars in the red sequence than in the blue one, suggesting that the distance distribution of the two populations is different (since at the largest distances only the brightest stars are apparent, and there are fewer of these on the red sequence). In order to select stars in either sequence we use a `PARSEC` isochrone (Bressan et al. 2012; Marigo et al. 2017) with [M/H] = −0.5 and age of 11 Gyr (blue line). The isochrone was shifted by 0.04 in colour and 0.2 in magnitude in order to match the gap between blue and red sequences. Notice that both the isochrone and extinction coefficients use *Gaia* DR2 transmission curves.

We now explore the dynamical distributions of the stars belonging to these two sequences in more detail. To this end we explore the velocity distribution in V_ℓ and V_b, for three cylindrical galactocentric distance bins for the stars in the blue (left) and red (right) sequences. We note that at the higher latitudes within this sample V_Z^* becomes a poorer approximation to V_z, but that is still a reasonable approximation to the non-radial, non-V_ϕ velocity of stars, and to low latitude stars. On the other hand, V_ℓ is good proxy for V_ϕ given the small range in ℓ. The densest structure at $V_\phi^* \sim 220\,\mathrm{km\,s^{-1}}$ is comprised mainly of disc stars, while the more extended and sparser structures belong to the halo and

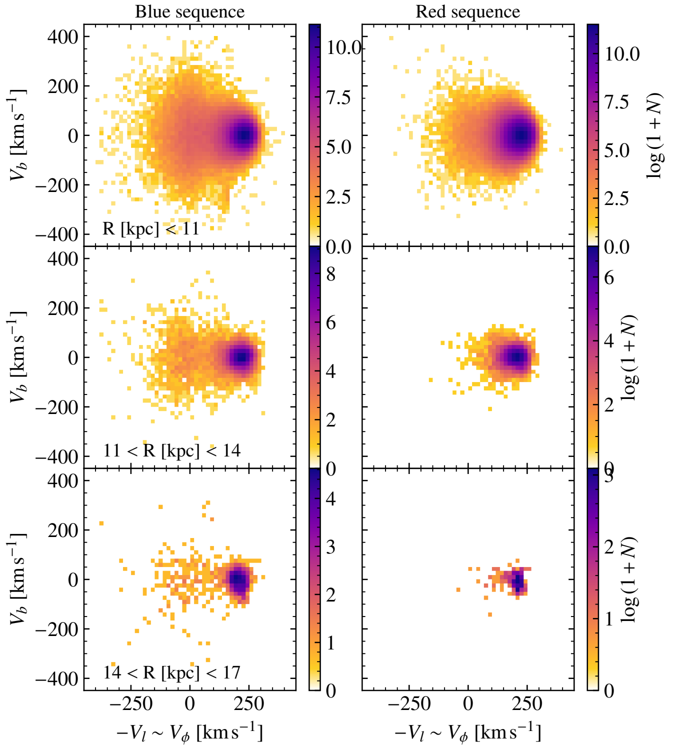

Fig. 18. Velocity distribution of the blue and red sequences. $-V_\ell \sim V_\phi^*$ vs V_b distribution showing in the left (right) column the stars in the blue (red) sequence. Each row shows the distributions for a given distance slice, indicated in the left panels. The stars with low rotation (even the retrogrades ones) are far more prominent in the blue sequence and extend to larger Galactocentric radii.

thick disc. Firstly, we note the presence of the Helmi et al. (1999) streams in the top-left panel at $(-V_\ell, V_b) \sim (150, -250)\,\mathrm{km\,s^{-1}}$ (indicating that these streams are a relatively local feature, in agreement with the results and predictions of Koppelman et al. 2019). We see, however, some hints of structures at similar velocities (and mirrored ones) in the second left panel of more distant stars that could potentially be related to these known streams. For the local sample (top panels) we observe a higher V_b velocity spread for the halo (i.e. at $V_\ell \sim 0$) blue sequence stars compared to the red sequence. In the intermediate distance bin (middle panels) the velocity distribution of red sequence stars barely extends to $V_\ell \sim 0$ and for the most distant stars (bottom panels) only the blue sequence is apparent in the halo population, with the red sequence mostly appearing as a low-dispersion disc-like component.

Similar conclusions can be drawn from Fig. 19 which shows the distribution of V_b velocities for the blue and red sequence for the same distance bins (columns) as in the previous figure, and for five V_t selections (rows). We note again that the red-sequence distribution generally has a lower velocity dispersion than the blue sequence stars (indicated with numbers for cases with at least 50 stars). Note also that for $V_t \gtrsim 100\,\mathrm{km\,s^{-1}}$ the dispersion increases significantly, indicating the transition from the canonical thick disk to a hotter component. For the more distant bins, the contribution of this hot thick disc becomes smaller (bottom right panels), and it is basically absent beyond 14-17 kpc (whereas the canonical thick disc still is apparent in the top panels at these radii). On the other hand, the blue sequence is apparent at all radii, and has a relatively large V_b velocity dispersion.

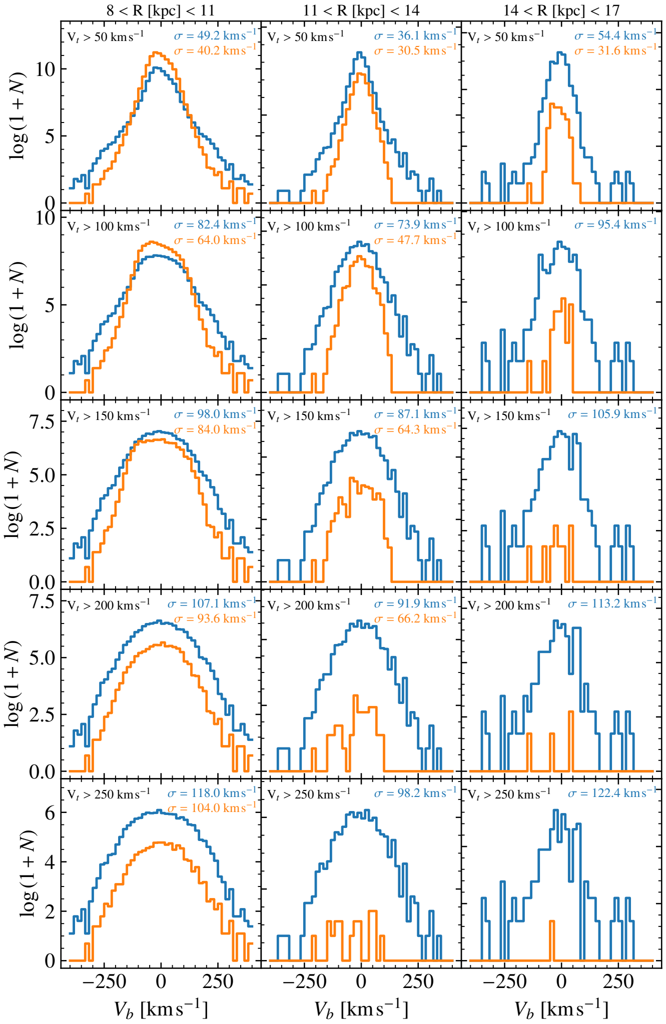

Fig. 19. Vertical tangential velocities V_b for the blue and red sequences. The plots are for different distance slices, indicated in the top of each column. Each row shows a different V_t selection, indicated in each panel. The blue (orange) curve shows the distribution for the blue (red) sequence. The different relative contribution of the sequences in the different panels is indicative of the spatial distribution of the accreted component and the ancient heated disc, and in particular of a shorter extent of the later one.

Therefore, the analyses presented in this section show that the hot thick-disc component, which locally has been associated with the heated disc at the time of the merger with Gaia-Enceladus-Sausage, has a smaller extent presently than the canonical thick disc. This suggests that the disc present at that time was smaller in size, as indeed expected from cosmological models. A more quantitative estimate of its size would require a careful assessment of the density distribution of the older stars in the red sequence, which is beyond the scope of this work. On the other hand, we see that the component locally associated with Gaia-Enceladus-Sausage extends out to large distances from the Sun, as we detect the presence of a retrograde component out to ∼ 17 kpc from the Galactic centre.

5.2. Distant structures

Studying the outskirts of the disc is a difficult task since the anticentre is mostly outshone by the nearby stars which are more numerous due to both the density gradient of the Galaxy and the

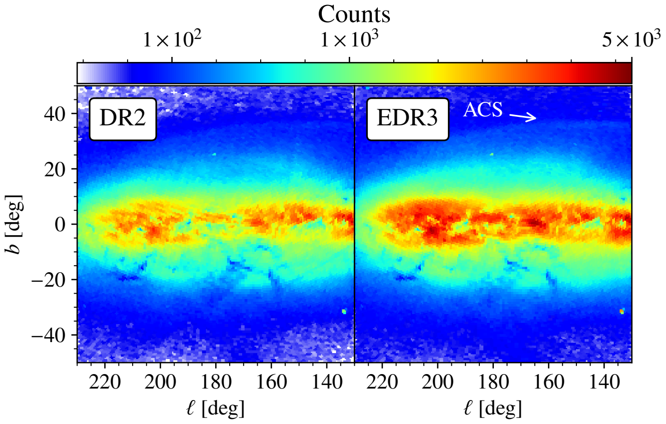

Fig. 20. Counts in the sky for a selection of stars that favours the outer disc structures. The selection is for stars with $\varpi < 0.1$ mas, $-1 < \mu_{\alpha}* < 1$ mas yr^{-1} and $-2 < \mu_{\delta} < 0$ mas yr^{-1}. Left: DR2 (we observe marks of the scanning law). Right: EDR3 (with more stars and better homogeneity) without filters nor parallax zero-point correction. The ACS can be seen more clearly in the right panel.

magnitude limitations inherent to any survey. The majority of the studies of the outer disc detected unexpected overdensities in counts such as Monoceros and ACS and focused on a specific stellar type, generally main sequence turn-off stars or M giants. An alternative way is now possible with *Gaia*, which allows us to detect them by applying the right astrometric selection. First, we can significantly reduce the amount of foreground contamination with a cut in parallax selecting only stars with $\varpi < 0.1$ mas. By doing so, we guarantee that most of the stars closer than 10 kpc are not selected, although the probability of failure is related to the parallax error of the source (the fainter sources being more likely to pass the filter regardless of their true distance). Then, we apply a kinematic selection since the proper motion signatures of these structures, given that they are relatively far from the Sun, are significantly different from the nearby disc and halo stars. The latter tend to have large proper motions due to the large relative velocity with respect to the Local Standard of Rest, while the former also tend to have large proper motions, but in this case due to the small heliocentric distance. Figure 20 is an example of such parallax and kinematic selection ($-1 < \mu_{\alpha}* < 1$ mas yr^{-1} and $-2 < \mu_{\delta} < 0$ mas yr^{-1}) where, in contrast to Fig. 1, we can observe a perfectly defined and thin ACS, as indicated by the arrow. The difference between DR2 (left) and EDR3 (right) is clear: we now have more stars (7 624 697 compared to 5 951 302), mostly due to the higher completeness of stars with proper motions in EDR3, and the sample is less affected by the scanning law and other artefacts.

In the first column of Fig. 21 we show the proper motion 2d histograms for different slices in latitude around the anticentre ($170° < \ell < 190°$) using the sample ACV, now with the astrometric and photometric filters, as well as the parallax zero-point correction. As we move from the north to the south Galactic hemisphere (top to bottom), different structures can be observed. We examine them by selecting stars in the rectangles A to H and plotting their Colour-Magnitude diagrams (CMDs) in the second and third columns using $G - G_{RP}$ instead of $G_{BP} - G_{RP}$ since, as exposed in Riello (2020) –see their Fig. 26–, the flux in the *BP* band can be overestimated for faint sources.

First, we note that the large concentration of sources close to the proper motion origin in the boxes A and G are mostly quasars for several reasons: i) they are faint and too blue, with $G - G_{RP} < 0.5$ mag, which is equivalent, incidentally, to the cut used in Newberg et al. (2002) ($g - r < 0.3$ mag) to remove the SDSS quasars[6], ii) the fraction of primary sources (`astrometric_primary_flag`), a significant fraction of which are quasars (Lindegren et al. 2018), is abnormally high in both A and G, and iii) ~32% of the sources in A and ~75% in G are found in the `agn_cross_id` table.

More interestingly, box A contains other kinematic structures apart from the aforementioned quasars. There is a more extended giant branch formed by the ACS (c.f. Fig. 2 from Laporte et al. 2020a) and, tentatively, two fainter tips of a giant branch that could be related to the Sagittarius stream (similarly to box C, as explained below). The other box (B) at the same latitude corresponds to the distribution of halo stars, their proper motions larger due to the Sun's reflex motion and their CMD compatible with an old isochrone at ~10 kpc or farther, where stars accumulate due to our parallax cut. In the second row, panel C contains parts of both Monoceros, which provides the giant branch, a well defined RC and a very blue turn-off consistent with previous observations (e.g. Newberg et al. 2002; Yanny et al. 2003), and the leading tidal tail of Sagittarius, which is only evident by its AGBs at magnitudes between 17 and 18. Boxes D and F are dominated by the disc which, after the selection in parallax, is expected to have a thick main sequence created by faint dwarfs with large parallax uncertainties, and a few Red Giants[7] above magnitude $G \sim 17$ mag.

On the south, at latitudes $-30\,\mathrm{deg} < b < -20\,\mathrm{deg}$, we observe that the CMD of the small proper motion population is dominated by two RCs (panel E), the densest at magnitude ~17 mag and the other at ~15.5 mag. To confirm their existence, we have obtained the Gaussian kernel of $G^* = G - 1.95(G_{BP} - G_{RP})$, therefore marginalising the apparent magnitudes along the extinction line (see Sect. 2.4). This kernel (shown within panel E) presents two overdensities corresponding to each of the mentioned RCs. By approximately selecting stars in these clumps and computing their distances assuming an absolute magnitude of the RC of $M_G = 0.495$ (Ruiz-Dern et al. 2018) and the extinction by Schlegel et al. (1998) -and thus, upper limits-, we find that they are located at an average heliocentric distance of 9 and 14 kpc with variance of 3 and 2 kpc, for the bright and faint clumps, respectively. These corresponds to Galactocentric cylindrical radii of around 16 and 21 kpc, and heights below the plane of -4 and -6 kpc, respectively. With this it is very likely that the bright RC corresponds to a nearby south extension of Monoceros (Ibata et al. 2003; Li et al. 2017), alternatively called "south middle structure" (Xu et al. 2015), at around ~ 12 to 16 kpc from the Galactic centre. On the other hand, the faint RC could be related to TriAnd (Rocha-Pinto et al. 2004; Xu et al. 2015; Li et al. 2017; Bergemann et al. 2018)), at a Galactocentric radius between 18 and 25. We note however that previous TriAnd detections were located around the range 100-160 deg in ℓ, thus not exactly in the anticentre direction, and our detection would then be a confirmation of the broadness of this structure and their extension up to $\ell = 180$, albeit predicted by models such as that from Sheffield et al. (2018).

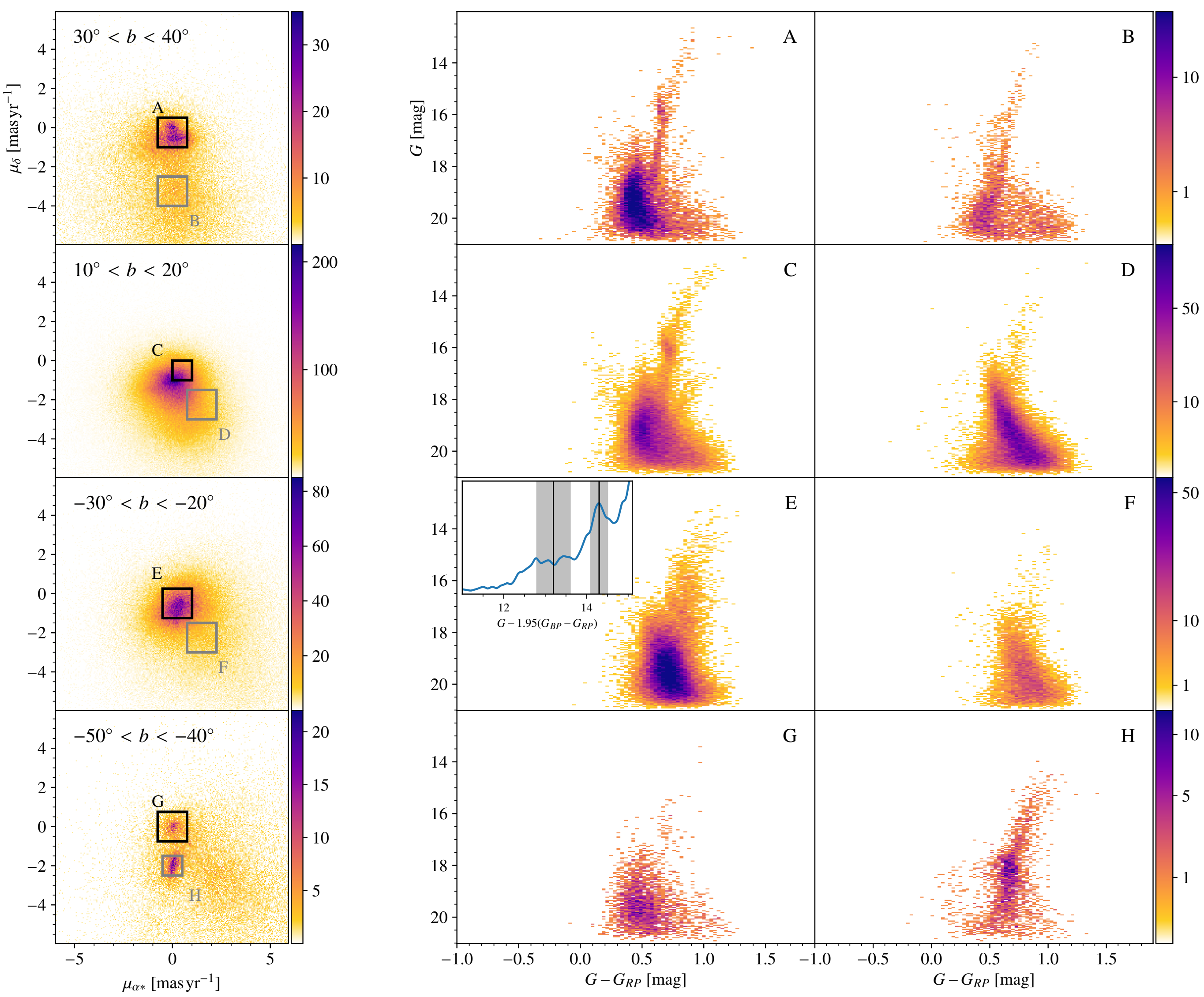

Fig. 21. Colour-magnitude diagrams of different features in the anticentre. The diagrams are for the kinematic groups selected in proper motion in the range $170° < \ell < 190°$. Each row corresponds to a different latitude. The first column shows their distribution in proper motion and the definition of two regions of interested. The stellar population in these selections are shown in the second and third columns. We see structures such as ACS (A), Monoceros (C), the Sagittarius stream (H, C) and other outer disc structures (E).

6. Clusters in the outer disc

In this section, we investigate the peculiar clusters Berkeley 29 and Saurer 1. The *Gaia* EDR3 astrometric data allows us for the first time to perform a reliable member selection of these clusters and to constrain their proper motions in order to determine their orbits. We retained all sources brighter than G=19 within 4 arcmin of the cluster centres. The members were identified from their *Gaia* proper motions and parallaxes with the unsupervised clustering procedure UPMASK (Krone-Martins & Moitinho 2014). Cantat-Gaudin et al. (2018), also using UPMASK, analysed the stars brighter than $G = 18$ mag of *Gaia* DR2 and detected Berkeley 29 but not Saurer 1. The improvement of *Gaia* EDR3 with respect to *Gaia* DR2 allows us to gain one magnitude and reliably detect both clusters. Figure 22 impressively shows how the stars in these clusters appear much more concentrated in proper motion space compared to DR2. The CMDs of the clusters are shown in Fig. 23, highlighting the sources that we consider the most secure members (with membership scores over 50%). We manually fit PARSEC isochrones (Bressan et al. 2012) to the observed CMDs. For Berkeley 29 we employ an isochrone with a metallicity [Fe/H] = −0.5 (Yong et al. 2005; Cantat-Gaudin et al. 2016), an age $\log t$=9.55, and a distance modulus of 15.5 mag with an extinction A_V of 0.2 mag. For Saurer 1 we use an isochrone of metallicity [Fe/H] = −0.4 (Carraro & Baume 2003; Cantat-Gaudin et al. 2016), an age $\log t$=9.6, and a distance modulus of 15.4 mag with an extinction A_V of 0.4 mag.

The mean proper motions of the cluster members are $(\mu_{\alpha}*, \mu_{\delta})$=(0.11, −1.05) mas yr^{-1} for Berkeley 29, and (−0.26, −0.32) mas yr^{-1} for Saurer 1. The mean tangential velocities V_ℓ and V_b (Eq. 4) are represented in the first panel of Fig. 24, along with that of the Sagittarius stream particles from

---

[6] We used the values in Table 5 from Jordi et al. (2010) to convert the SDSS colours to *Gaia* colours.

[7] A RC star fainter than $G \sim 15$ mag at latitudes b>10° is bound to be higher than 2 kpc from the disc, which is unlikely, but stars brighter than that tend to have a reliable parallax and are therefore more likely to be removed with our parallax cut.

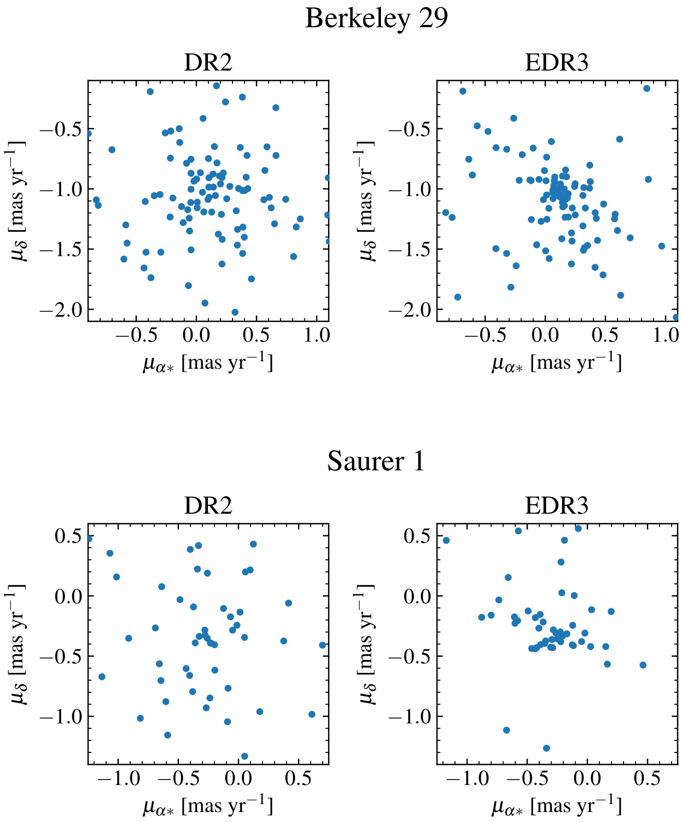

Fig. 22. Proper motions of the stars in Berkeley 29 and Saurer 1. The proper motions are for Berkeley 29 (top) and Saurer 1 (bottom) for *Gaia* DR2 (left) and *Gaia* EDR3 (right), for sources brighter than G=19 in the investigated field of view. The reduced uncertainties in EDR3 make the stars appear much more clumped than in DR2, allowing for a better selection of members and a better determination of the proper motion of the clusters.

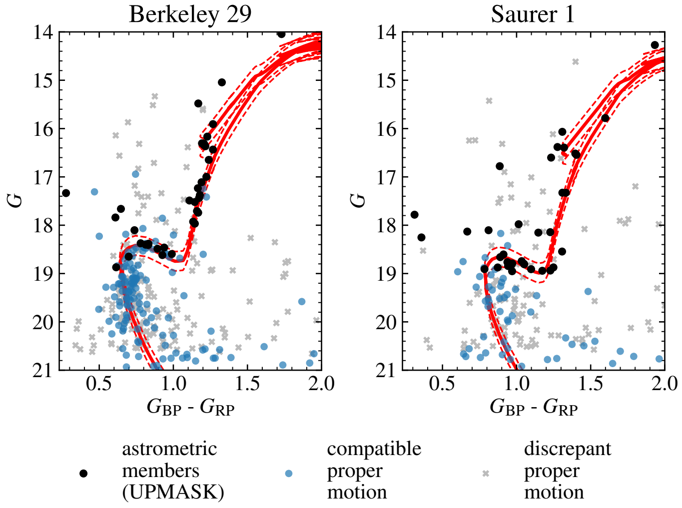

Fig. 23. CMDs for Berkeley 29 and Saurer 1. The secure members identified with UPMASK are indicated and used to compute the mean proper motions. The blue points are sources with similar proper motions but large uncertainties, or magnitudes G>19. The red lines are PARSEC isochrones. The dashed lines correspond to offsets of ±0.2 mag in distance modulus.

the Law & Majewski (2010) model. In this panel, all proper motions were corrected from the effect of the Solar motion. The velocity vector of both clusters is mostly parallel to the Galactic plane, and differs significantly from that of the stream.

We used `galpy` MWPotential2014 model (Bovy 2015) to integrate the orbits of these objects, shown in the left panels of Fig. 24. For this, we supplemented the quantities derived from *Gaia* data with line-of-sight velocities obtained from high-resolution spectra analysed by Cantat-Gaudin et al. (2016). Their mean line-of-sight velocities are 24.8 km s^{-1} for Berkeley 29 (from eight stars), and 98.0 km s^{-1} for Saurer 1 (from two stars). All the stars they used to compute those mean velocities are part of the sample of secure members we obtained in the present study. We estimated the uncertainty on the main orbital parameters by Monte-Carlo sampling of the uncertainties on the distance, line-of-sight velocity, and proper motion. We assume an uncertainty of 0.2 mag on the distance modulus, and 2 km s^{-1} on the line-of-sight velocities of both clusters. The precision on the mean cluster proper motion is limited by systematics, on the level of 11 μas yr^{-1} on each component of the mean cluster proper motion. All sampled orbits correspond to prograde, bound trajectories. The maximum altitude above the Galactic plane is z_{max}=$1.80^{+0.45}_{-0.09}$ for Berkeley 29, and z_{max}=$1.59^{+0.11}_{-0.09}$ for Saurer 1. They also exhibit small eccentricities e=$0.03^{+0.08}_{-0.01}$ and $0.05^{+0.06}_{-0.05}$, respectively. Despite their large Galactocentric distance, the orbits of these clusters are typical of disc objects. We obtained very similar results with the model by McMillan (2017).

7. Discussion and conclusions

7.1. Summary of results

With the combination of photometric and astrometric data from *Gaia* EDR3, we have explored the dynamics of different elements of the MW in the Galactic anticentre. The main results of this study are:

1. There are prominent oscillations in the median rotation and vertical velocities and dispersions of disc stars as a function of radius and angular momentum which depend on the evolutionary state of the stars (Sect. 4.1).
2. There are significant asymmetries in velocity when comparing stars above and below the standard Galactic plane for disc stars that can be as high as 5 km s^{-1} for the vertical velocities and 10 km s^{-1} for the rotation ones (Sect. 4.2).
3. At the outer disc, stars are predominantly following a bimodal distribution, with a group of stars mostly below the plane moving upwards with velocities of ~ 10 km s^{-1} and rotating faster by about ~ 30 km s^{-1} than another group of stars predominantly above the plane moving downwards by 2-5 km s^{-1} (Sect. 4.2).
4. The known R-V_ϕ ridges discovered with *Gaia* DR2, reach larger Galactocentric radius with EDR3 (up to 14 kpc) and there are also new ridges up to about 16-18 kpc, that is much beyond the limits reached in previous studies (Sect. 4.3).
5. Galactic rotation is detected as far as 18 kpc from the Galactic centre, being this a lower limit on the current thin disk size (Sects. 4.1 and 4.2).

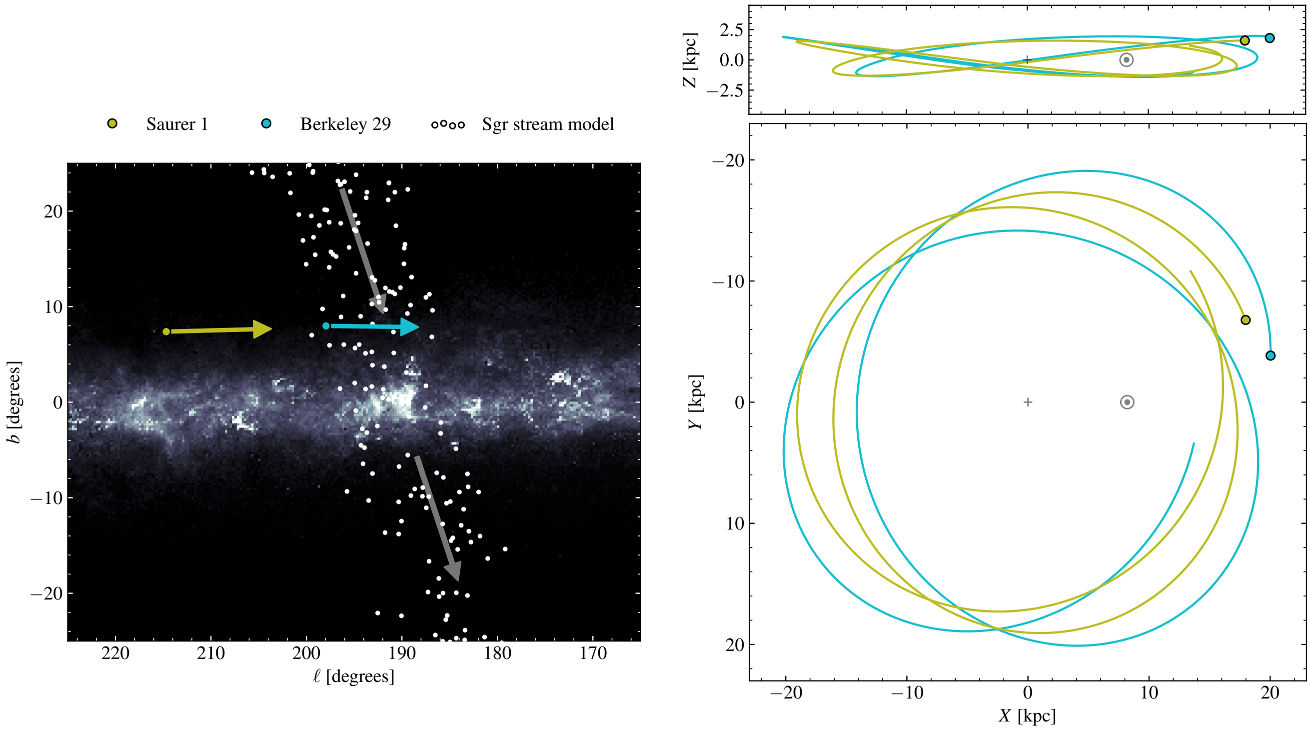

Fig. 24. Orbits if the Berkeley 29 and Saurer 1 clusters from EDR3 data. Left: location of Saurer 1 and Berkeley 29 in Galactic coordinates. The white symbols are the Sagittarius stream particles modelled by Law & Majewski (2010). The arrows indicate mean tangential velocities corrected for the solar motion (V_ℓ, V_b). The background is the integrated extinction model of Green et al. (2015), beyond 2 kpc. Right: integrated orbits of Saurer 1 and Berkeley 29. The Sun's position and Galactic centre are indicated as the usual Sun's symbol and cross, respectively. We find that the orbits of these clusters are very similar and typical of the disc.

6. The red sequence of high tangential velocity stars (suggested to be the ancient disc that was heated after the merger with Gaia-Enceladus-Sausage) now is seen to extend out to ∼ 14 kpc (Sect. 5.1).
7. The blue sequence (assumed to be the debris of the Gaia-Enceladus-Sausage) is much more extended and can be detected at least beyond 17 kpc (Sect. 5.1).
8. The far anticentre shows a intricate superposition of structures in the proper motion and photometry diagrams including the leading (in the north) and trailing (in the south) Sagittarius stream, and known outer disc structures such of Monoceros and ACS in the north (Sect. 5.2).
9. There are two structures at latitudes of $-30 < b < -20$ deg approximately at 9 and 14 kpc from the Sun, tentatively related to the Monoceros in the south and an extension of TriAnd in the anticentre direction, respectively (Sect. 5.2).
10. The clusters Berkeley 29 and Saurer 1, which are among the oldest open clusters known, are found to be on disc-like orbits despite being located at around 20 kpc from the Galactic center (Sect. 6).

7.2. Discussion (I): MW disc dynamics

In line with the discoveries from *Gaia* DR2, the disk is found to be rather complex. Nearby, the rotation velocities are dominated by the known ridges in the (R, V_ϕ^*) plane, which are now detected up to 14 kpc from the Galactic centre, that is 3 kpc farther than for DR2, while two additional ridges are discovered that reach 16-18 kpc. The overlap of distinct ridges in R seems to be the cause of some oscillation seen in the rotation curve, as already suggested by Martinez-Medina et al. (2019), although they could also be related to the physical location of the spiral arms (Sancisi 2004; McGaugh 2019), rather than their resonances (Barros et al. 2013).

The most prominent nearby ridge is Sirius, followed by the hat, L18, Hyades and Hercules. If indeed the Hercules, Sirius and hat ridges are signatures of the corotation, 4:1, and 2:1 Outer Lindblad Resonance of the bar (Monari et al. 2019; Laporte et al. 2020a), respectively, with the latter being the outermost resonance that can exist, the new ridges beyond 12 kpc discovered here require a different explanation, either spiral structure (with a lower pattern speed or transient) or the perturbation from Sagittarius (or the two at the same time since perturbations from satellites inevitably induce density spirals and rings, Purcell et al. 2011).

The *Gaia* EDR3 now allows for a full characterisation of the velocity ellipsoid and the asymmetric drift as a function of age and radius. We see clear oscillations in V_Z^* with radius (and angular momentum) of an amplitude of 1-2 km s^{-1} but increasing for younger stars. As already noticed before (e.g. Schönrich & Dehnen 2018; Huang et al. 2018; Beane et al. 2019; Cheng et al. 2020) but now seen at a higher precision with *Gaia* EDR3, these oscillations could indicate a vertical wave propagating radially and are possibly associated to oscillations in the local mid-plane itself. The vertical velocity dispersions do not show the expected decreasing behaviour with radius but seem to be flat or increase and present very prominent oscillations that appear connected to the oscillations in the median velocities.

In the outer disc ($R > 12$ kpc), the velocity field is dominated by an upwards motion of about 5 km s^{-1}. Already seen in Gaia Collaboration et al. (2018c), Wang et al. (2018), Poggio et al. (2018), Romero-Gómez et al. (2019), Carrillo et al. (2019), López-Corredoira et al. (2020) and Cheng et al. (2020), it has been associated to the warp that in the anticentre is near the line-of nodes, a bending wave due to Sagittarius, or to a disc that never achieved equilibrium. Here, however, we go one step beyond and find, coexisting in R, a bimodal distribution of stars moving vertically with opposite directions and with a different amount of rotation. The feature can be observed also as a vertical velocity oscillation in angular momentum space, which can thus have different phases coexisting at the same R. This bimodality shows similar phase space correlations to those of the phase spiral (Antoja et al. 2018). Each group of the bimodality could be interpreted as different wraps of a phase-mixing feature or a combination of bending waves. Missing data in this study such as line-of-sight velocities and chemistry will help in the understanding of this feature. The WEAVE Galactic Archaeology survey (Dalton et al. 2016; Famaey et al. 2016) has a dedicated science case in the region of the anticentre to obtain line-of-sight velocities in complement to *Gaia*, which will be crucial in this and many other aspects explored in this study. Yet our exploration reveals that simple 2d projections of phase space often do not capture the full complexity of the disc dynamics: when the vertical velocities are explored alone as a function of radius, only the upward motion (as in previous studies) is seen and adding more coordinates is necessary to observe this bimodality.

To interpret the complex patterns observed, a dynamical framework is required that no longer assumes decoupling between the vertical and horizontal movements (D'Onghia et al. 2016) and is capable of linking the small scale features such as the ridges, the global streaming motions, the phase spiral and perhaps structures such as the warp, the flare and the spiral arms. In any case, we probably live in a Galaxy with a highly perturbed outer disc (e.g. Widrow et al. 2012; Wang et al. 2017; Antoja et al. 2018; Beane et al. 2019) as seen in simulations of MW-like galaxies perturbed by a Sagittarius-like galaxy (Purcell et al. 2011; Gómez et al. 2013, 2017; Laporte et al. 2018) having rings with non-null vertical velocities, qualitatively comparable to what we find here.

7.3. Discussion (II): MW constituents

After *Gaia* DR2, our understanding of the Galactic components has changed, in particular recognising that most of the (local) halo is made of debris from a single accretion event (forming a blue sequence in the HR diagram of high transverse velocity stars) and that we here find to be extended beyond the local neighbourhood at least up to distances of 17 kpc from the Galactic centre. This is consistent with expectations from e.g. the orbit integrations of Deason et al. (2018), but also emphasises the global importance of the debris. The redder component of the HR diagram does not extend this far, having very few stars already around 14 kpc. If this redder component is the heated thick disc after the merger (as claimed in Helmi et al. 2018; Di Matteo et al. 2019; Gallart et al. 2019; Belokurov et al. 2020), this would imply that it is more compact than the canonical thick disc, which can be detected up to this radius. It will be interesting to try to constrain its initial properties, particularly through comparison to simulations of mergers and subsequent disc growth.

In line with Carraro et al. (2010) and López-Corredoira et al. (2018) that advocate a disc larger than the previously thought 12-14 kpc (Minniti et al. 2011, and references therein), we find evidence of circular rotation up to about 18 kpc from the Galactic center. A more precise value for such an important measurement needs a detailed analysis of the effects of the adopted zero parallax point, the biases on any distance estimate (see Appendix C and D) and aspects such as the flare (e.g. López-Corredoira & Molgó 2014). We have compared the effects of a constant parallax offset of $-17\,\mu$as (the average offset of the quasars) with that of Lindegren (2020b) - a more sophisticated prescription as a function of magnitude, colour and ecliptic latitude. We find that the latter gives a more compressed distance scale (that propagates to velocities) but at this point it is not straightforward to claim that one prescription is better than the other (Lindegren 2020b). In any case, the features observed remain qualitatively the same regardless of the zero point.

Gaia has also provided us with a window into the structures that dwell at the edge of the disc. We detect the Monoceros and the ACS above the disc plane and other structures in the south. Our southern detections are possibly related to the Monoceros south or south middle structure (e.g. Ibata et al. 2003; Xu et al. 2015) and TriAnd (Rocha-Pinto et al. 2004), which have not been probed in detail at $\ell \sim 180\,$deg so far due to the high extinction (e.g. Slater et al. 2014; Xu et al. 2015). If confirmed, this would be the first TriAnd detection with *Gaia* data (but see Ramos et al. 2020a) and the first time it is observed beyond its previously known longitude limit of $\ell \sim 160\,$deg. Curiously, the part of the disk bimodality of stars below the plane moving upwards strongly at 12 kpc coincides in distance with this nearby southern structure, though the latter is at a lower latitude. The connection between these features certainly needs some attention. Whether these structures are the corresponding northern and southern counterparts of the vertically oscillating disc (bending wave) expected in the scenario proposed by Widrow et al. (2012) and Yanny & Gardner (2013), or they are individual rings/feather structures in the outermost parts of the disc as suggested in Ibata et al. (2003), Kazantzidis et al. (2008), Purcell et al. (2011), Gómez et al. (2013) and Laporte et al. (2019a) also remains undetermined. In any case, future studies can benefit from the *Gaia* data that enables the kinematic selection of members of these features, providing a uniform sample of all the stellar types, and the determination of their proper motion.

Here we have looked at two particular clusters, Berkeley 29 and Saurer 1, that due to their great distances from the Galaxy centre (around 20 kpc, derived photometrically and thus not affected by the parallax offset) and their old age (3-4 Gyr) probe extreme conditions in the Galaxy. Using an improved membership assignation and the better astrometry of *Gaia* EDR3, we ascertain that the two clusters are on disc orbits, unlike what was claimed by previous studies (Frinchaboy 2006; Wu et al. 2009; Carraro et al. 2007; Vande Putte et al. 2010; Carraro & Bensby 2009). Yet, their distant location makes us wonder whether the disc extends to such a distance or whether these clusters were brought there by other means (radial migration, interaction with a satellite, expelled material from the disc). In particular Berkeley 29 has been already associated to Monoceros in Carraro et al. (2004) and in Frinchaboy et al. 2004, though in the latter case advocating for a stream origin. Our distances and proper motion of these clusters are compatible with the ones of Monoceros. Similarly, after examining the literature (Rocha-Pinto et al. 2004; Li et al. 2012; Cantat-Gaudin et al. 2016; Sheffield et al. 2018), we note that their chemistry and line-of-sight velocity are also broadly comparable. These clusters thus can be small but relevant pieces of information on the outer disc unknowns.

7.4. Conclusion

The quality of the EDR3 *Gaia* data together with the advantage of having astrometry and photometry from the same mission have allowed us to extend the horizon for exploration towards the very end of the disc, travel to the past to explore its ancient components and detect its small constituents and phase space features with better resolution. With a simple exploration of the *Gaia* data we find new complex patterns of movement in the outskirts of the Galactic disc, we estimate the extent of the ancient MW disc, show how the anticentre is a crossroad of structures likely both of internal and external origin, and uncover the nature of the orbits of two distant clusters. The anticentre is thus proven to be an excellent testbed region in the quest of deciphering the structure and history of our Galaxy that many astrophysicists are pursuing in the *Gaia* era.

Acknowledgements. We thank the referee for his comments. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (`https://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, `https://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This work presents results from the European Space Agency (ESA) space mission *Gaia* . *Gaia* data are being processed by the *Gaia* Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the *Gaia* MultiLateral Agreement (MLA). The *Gaia* mission website is `https://www.cosmos.esa.int/gaia`. The *Gaia* archive website is `https://archives.esac.esa.int/gaia`. The *Gaia* mission and data processing have financially been supported by, in alphabetical order by country: the Algerian Centre de Recherche en Astronomie, Astrophysique et Géophysique of Bouzareah Observatory; the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (FWF) Hertha Firnberg Programme through grants T359, P20046, and P23737; the BELgian federal Science Policy Office (BELSPO) through various PROgramme de Développement d'Expériences scientifiques (PRODEX) grants and the Polish Academy of Sciences - Fonds Wetenschappelijk Onderzoek through grant VS.091.16N, and the Fonds de la Recherche Scientifique (FNRS); the Brazil-France exchange programmes Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and Coordenação de Aperfeicoamento de Pessoal de Nível Superior (CAPES) - Comité Français d'Evaluation de la Coopération Universitaire et Scientifique avec le Brésil (COFECUB); the National Science Foundation of China (NSFC) through grants 11573054 and 11703065 and the China Scholarship Council through grant 201806040200; the Tenure Track Pilot Programme of the Croatian Science Foundation and the École Polytechnique Fédérale de Lausanne and the project TTP-2018-07-1171 'Mining the Variable Sky', with the funds of the Croatian-Swiss Research Programme; the Czech-Republic Ministry of Education, Youth, and Sports through grant LG 15010 and INTER-EXCELLENCE grant LTAUSA18093, and the Czech Space Office through ESA PECS contract 98058; the Danish Ministry of Science; the Estonian Ministry of Education and Research through grant IUT40-1; the European Commission's Sixth Framework Programme through the European Leadership in Space Astrometry (ELSA) Marie Curie Research Training Network (MRTN-CT-2006-033481), through Marie Curie project PIOF-GA-2009-255267 (Space AsteroSeismology & RR Lyrae stars, SAS-RRL), and through a Marie Curie Transfer-of-Knowledge (ToK) fellowship (MTKD-CT-2004-014188); the European Commission's Seventh Framework Programme through grant FP7-606740 (FP7-SPACE-2013-1) for the *Gaia* European Network for Improved data User Services (GENIUS) and through grant 264895 for the *Gaia* Research for European Astronomy Training (GREAT-ITN) network; the European Research Council (ERC) through grants 320360 and 647208 and through the European Union's Horizon 2020 research and innovation and excellent science programmes through Marie Skłodowska-Curie grant 745617 as well as grants 670519 (Mixing and Angular Momentum tranSport of massIvE stars – MAMSIE), 687378 (Small Bodies: Near and Far), 682115 (Using the Magellanic Clouds to Understand the Interaction of Galaxies), and 695099 (A sub-percent distance scale from binaries and Cepheids – CepBin); the European Science Foundation (ESF), in the framework of the *Gaia* Research for European Astronomy Training Research Network Programme (GREAT-ESF); the European Space Agency (ESA) in the framework of the *Gaia* project, through the Plan for European Cooperating States (PECS) programme through grants for Slovenia, through contracts C98090 and 4000106398/12/NL/KML for Hungary, and through contract 4000115263/15/NL/IB for Germany; the Academy of Finland and the Magnus Ehrnrooth Foundation; the French Centre National d'Etudes Spatiales (CNES), the Agence Nationale de la Recherche (ANR) through grant ANR-10-IDEX-0001-02 for the 'Investissements d'avenir' programme, through grant ANR-15-CE31-0007 for project 'Modelling the Milky Way in the Gaia era' (MOD4Gaia), through grant ANR-14-CE33-0014-01 for project 'The Milky Way disc formation in the Gaia era' (ARCHEOGAL), and through grant ANR-15-CE31-0012-01 for project 'Unlocking the potential of Cepheids as primary distance calibrators' (UnlockCepheids), the Centre National de la Recherche Scientifique (CNRS) and its SNO Gaia of the Institut des Sciences de l'Univers (INSU), the 'Action Fédératrice Gaia' of the Observatoire de Paris, the Région de Franche-Comté, and the Programme National de Gravitation, Références, Astronomie, et Métrologie (GRAM) of CNRS/INSU with the Institut National Polytechnique (INP) and the Institut National de Physique nucléaire et de Physique des Particules (IN2P3) co-funded by CNES; the German Aerospace Agency (Deutsches Zentrum für Luft- und Raumfahrt e.V., DLR) through grants 50QG0501, 50QG0601, 50QG0602, 50QG0701, 50QG0901, 50QG1001, 50QG1101, 50QG1401, 50QG1402, 50QG1403, 50QG1404, and 50QG1904 and the Centre for Information Services and High Performance Computing (ZIH) at the Technische Universität (TU) Dresden for generous allocations of computer time; the Hungarian Academy of Sciences through the Lendület Programme grants LP2014-17 and LP2018-7 and through the Premium Postdoctoral Research Programme (L. Molnár), and the Hungarian National Research, Development, and Innovation Office (NKFIH) through grant KH_18-130405; the Science Foundation Ireland (SFI) through a Royal Society - SFI University Research Fellowship (M. Fraser); the Israel Science Foundation (ISF) through grant 848/16; the Agenzia Spaziale Italiana (ASI) through contracts I/037/08/0, I/058/10/0, 2014-025-R.0, 2014-025-R.1.2015, and 2018-24-HH.0 to the Italian Istituto Nazionale di Astrofisica (INAF), contract 2014-049-R.0/1/2 to INAF for the Space Science Data Centre (SSDC, formerly known as the ASI Science Data Center, ASDC), contracts I/008/10/0, 2013/030/I.0, 2013-030-I.0.1-2015, and 2016-17-I.0 to the Aerospace Logistics Technology Engineering Company (ALTEC S.p.A.), INAF, and the Italian Ministry of Education, University, and Research (Ministero dell'Istruzione, dell'Università e della Ricerca) through the Premiale project 'MIning The Cosmos Big Data and Innovative Italian Technology for Frontier Astrophysics and Cosmology' (MITiC); the Netherlands Organisation for Scientific Research (NWO) through grant NWO-M-614.061.414, through a VICI grant (A. Helmi), and through a Spinoza prize (A. Helmi), and the Netherlands Research School for Astronomy (NOVA); the Polish National Science Centre through HARMONIA grant 2018/06/M/ST9/00311, DAINA grant 2017/27/L/ST9/03221, and PRELUDIUM grant 2017/25/N/ST9/01253, and the Ministry of Science and Higher Education (MNiSW) through grant DIR/WK/2018/12; the Portugese Fundação para a Ciência e a Tecnologia (FCT) through grants SFRH/BPD/74697/2010 and SFRH/BD/128840/2017 and the Strategic Programme UID/FIS/00099/2019 for CENTRA; the Slovenian Research Agency through grant P1-0188; the Spanish Ministry of Economy (MINECO/FEDER, UE) through grants ESP2016-80079-C2-1-R, ESP2016-80079-C2-2-R, RTI2018-095076-B-C21, RTI2018-095076-B-C22, BES-2016-078499, and BES-2017-083126 and the Juan de la Cierva formación 2015 grant FJCI-2015-2671, the Spanish Ministry of Education, Culture, and Sports through grant FPU16/03827, the Spanish Ministry of Science and Innovation (MICINN) through grant AYA2017-89841P for project 'Estudio de las propiedades de los fósiles estelares en el entorno del Grupo Local' and through grant TIN2015-65316-P for project 'Computación de Altas Prestaciones VII', the Severo Ochoa Centre of Excellence Programme of the Spanish Government through grant SEV2015-0493, the Institute of Cosmos Sciences University of Barcelona (ICCUB, Unidad de Excelencia 'María de Maeztu') through grants MDM-2014-0369 and CEX2019-000918-M, the University of Barcelona's official doctoral programme for the development of an R+D+i project through an Ajuts de Personal Investigador en Formació (APIF) grant, the Spanish Virtual Observatory through project AyA2017-84089, the Galician Regional Government, Xunta de Galicia, through grants ED431B-2018/42 and ED481A-2019/155, support received from the Centro de Investigación en Tecnologías de la Información y las Comunicaciones (CITIC) funded by the Xunta de Galicia, the Xunta de Galicia and the Centros Singulares de Investigación de Galicia for the period 2016-2019 through CITIC, the European Union through the European Regional Development Fund (ERDF) / Fondo Europeo de Desenvolvemento Rexional (FEDER) for the Galicia 2014-2020 Programme through grant ED431G-2019/01, the Red Española de Supercomputación (RES) computer resources at MareNostrum, the Barcelona Supercomputing Centre - Centro Nacional de Supercomputación (BSC-CNS) through activities AECT-2016-1-0006, AECT-2016-2-0013, AECT-2016-3-0011, and AECT-2017-1-0020, the Departament d'Innovació, Universitats i Empresa de la Generalitat de Catalunya through grant 2014-SGR-1051 for project 'Models de Programació i Entorns d'Execució Parallels' (MPEXPAR), and Ramon y Cajal Fellowship RYC2018-025968-I; the Swedish National Space Agency (SNSA/Rymdstyrelsen); the Swiss State Secretariat for Education, Research, and Innovation through the ESA PRODEX programme, the Mesures d'Accompagnement, the Swiss Activités Nationales Complémentaires, and the Swiss National Science Foundation; the United Kingdom Particle Physics and Astronomy Research Council (PPARC), the United Kingdom Science and Technology Facilities Council (STFC), and the United Kingdom Space Agency (UKSA) through the following grants to the University of Bristol, the University of Cambridge, the University of Edinburgh, the University of Leicester, the Mullard Space Sciences Laboratory of University College London, and the United Kingdom Rutherford Appleton Laboratory (RAL): PP/D006511/1, PP/D006546/1, PP/D006570/1, ST/I000852/1, ST/J005045/1, ST/K00056X/1, ST/K000209/1, ST/K000756/1, ST/L006561/1, ST/N000595/1,

ST/N000641/1, ST/N000978/1, ST/N001117/1, ST/S000089/1, ST/S000976/1, ST/S001123/1, ST/S001948/1, ST/S002103/1, and ST/V000969/1.

References

Ahumada, R., Allende Prieto, C., Almeida, A., et al. 2020, ApJS, 249, 3
Antoja, T., de Bruijne, J., Figueras, F., et al. 2017, A&A, 602, L13
Antoja, T., Figueras, F., Fernández, D., & Torra, J. 2008, A&A, 490, 135
Antoja, T., Helmi, A., Romero-Gómez, M., et al. 2018, Nature, 561, 360
Antoja, T., Ramos, P., Mateu, C., et al. 2020, A&A, 635, L3
Aumer, M. & Binney, J. J. 2009, MNRAS, 397, 1286
Bailer-Jones, C. A. L. 2015, PASP, 127, 994
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, AJ, 156, 58
Barros, D. A., Lépine, J. R. D., & Junqueira, T. C. 2013, MNRAS, 435, 2299
Beane, A., Sanderson, R. E., Ness, M. K., et al. 2019, ApJ, 883, 103
Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, MNRAS, 478, 611
Belokurov, V., Sanders, J. L., Fattahi, A., et al. 2020, MNRAS, 494, 3880
Bennett, M. & Bovy, J. 2019, MNRAS, 482, 1417
Bergemann, M., Sesar, B., Cohen, J. G., et al. 2018, Nature, 555, 334
Binney, J., Gerhard, O. E., Stark, A. A., Bally, J., & Uchida, K. I. 1991, MNRAS, 252, 210
Binney, J. & Schönrich, R. 2018, MNRAS, 481, 1501
Binney, J. & Tremaine, S. 2008, Galactic Dynamics: Second Edition (Princeton University Press)
Bland-Hawthorn, J. & Gerhard, O. 2016, ARA&A, 54, 529
Bland-Hawthorn, J., Sharma, S., Tepper-Garcia, T., et al. 2019, MNRAS, 486, 1167
Blitz, L. & Spergel, D. N. 1991, ApJ, 379, 631
Boubert, D. & Everall, A. 2020, arXiv e-prints, arXiv:2005.08983
Bovy, J. 2015, ApJS, 216, 29
Bragaglia, A. & Tosi, M. 2006, AJ, 131, 1544
Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127
Cantat-Gaudin, T., Donati, P., Vallenari, A., et al. 2016, A&A, 588, A120
Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, A&A, 618, A93
Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
Carlin, J. L., DeLaunay, J., Newberg, H. J., et al. 2013, ApJ, 777, L5
Carraro, G. & Baume, G. 2003, MNRAS, 346, 18
Carraro, G. & Bensby, T. 2009, MNRAS, 397, L106
Carraro, G., Bresolin, F., Villanova, S., et al. 2004, AJ, 128, 1676
Carraro, G., Geisler, D., Villanova, S., Frinchaboy, P. M., & Majewski, S. R. 2007, A&A, 476, 217
Carraro, G., Vázquez, R. A., Costa, E., Perren, G., & Moitinho, A. 2010, ApJ, 718, 683
Carrillo, I., Minchev, I., Steinmetz, M., et al. 2019, MNRAS, 490, 797
Casagrande, L. & VandenBerg, D. A. 2018, MNRAS, 479, L102
Cheng, X., Anguiano, B., Majewski, S. R., et al. 2020, arXiv e-prints, arXiv:2010.10398
Crane, J. D., Majewski, S. R., Rocha-Pinto, H. J., et al. 2003, ApJ, 594, L119
Dalton, G., Trager, S., Abrams, D. C., et al. 2016, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 9908, Ground-based and Airborne Instrumentation for Astronomy VI, 99081G
de Boer, T. J. L., Belokurov, V., & Koposov, S. E. 2018, MNRAS, 473, 647
de Vaucouleurs, G. 1964, in IAU Symposium, Vol. 20, The Galaxy and the Magellanic Clouds, ed. F. J. Kerr, 195
Deason, A. J., Belokurov, V., Koposov, S. E., & Lancaster, L. 2018, ApJ, 862, L1
Dehnen, W. 1998, AJ, 115, 2384
Di Matteo, P., Haywood, M., Lehnert, M. D., et al. 2019, A&A, 632, A4
D'Onghia, E., Madau, P., Vera-Ciro, C., Quillen, A., & Hernquist, L. 2016, ApJ, 823, 4
Drimmel, R. & Spergel, D. N. 2001, ApJ, 556, 181
Fabricius, C. 2020, A&A
Famaey, B., Antoja, T., Romero-Gomez, M., et al. 2016, in SF2A-2016: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, ed. C. Reylé, J. Richard, L. Cambrésy, M. Deleuil, E. Pécontal, L. Tresse, & I. Vauglin, 281–286
Famaey, B., Jorissen, A., Luri, X., et al. 2005, A&A, 430, 165
Frinchaboy, P. M. 2006, arXiv e-prints, astro
Frinchaboy, P. M., Majewski, S. R., Crane, J. D., et al. 2004, ApJ, 602, L21
Frinchaboy, P. M. & Phelps, R. L. 2002, AJ, 123, 2552
Friske, J. K. S. & Schönrich, R. 2019, MNRAS, 490, 5414
Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018a, A&A, 616, A10
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018b, A&A, 616, A1
Gaia Collaboration, Katz, D., Antoja, T., et al. 2018c, A&A, 616, A11
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gaia Collaboration, Brown, A. 2020, In Prep.
Gallart, C., Bernard, E. J., Brook, C. B., et al. 2019, Nature Astronomy, 407
Gómez, F. A., Minchev, I., O'Shea, B. W., et al. 2013, MNRAS, 429, 159
Gómez, F. A., White, S. D. M., Grand , R. J. J., et al. 2017, MNRAS, 465, 3446
Gonzalez-Nunez, X. 2020, In Prep.
Gravity Collaboration, Abuter, R., Amorim, A., et al. 2019, A&A, 625, L10
Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93
Green, G. M., Schlafly, E. F., Finkbeiner, D. P., et al. 2015, ApJ, 810, 25
Grillmair, C. J. 2006, ApJ, 651, L29
Hawkins, K., Leistedt, B., Bovy, J., & Hogg, D. W. 2017, ArXiv e-prints [arXiv:1705.08988]
Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Nature, 563, 85
Helmi, A., White, S. D. M., de Zeeuw, P. T., & Zhao, H. 1999, Nature, 402, 53
Huang, Y., Schönrich, R., Liu, X. W., et al. 2018, ApJ, 864, 129
Ibata, R. A., Irwin, M. J., Lewis, G. F., Ferguson, A. M. N., & Tanvir, N. 2003, MNRAS, 340, L21
Jordi, C., Gebran, M., Carrasco, J. M., et al. 2010, A&A, 523, A48
Kaluzny, J. 1994, A&AS, 108, 151
Katz, D., Sartoretti, P., Cropper, M., et al. 2019, A&A, 622, A205
Kawata, D., Baba, J., Ciucă, I., et al. 2018, MNRAS, 479, L108
Kazantzidis, S., Bullock, J. S., Zentner, A. R., Kravtsov, A. V., & Moustakas, L. A. 2008, ApJ, 688, 254
Khanna, S., Sharma, S., Bland-Hawthorn, J., et al. 2019a, MNRAS, 482, 4215
Khanna, S., Sharma, S., Tepper-Garcia, T., et al. 2019b, MNRAS, 489, 4962
Klioner, S. 2020, A&A
Koppelman, H. H., Helmi, A., Massari, D., Roelenga, S., & Bastian, U. 2019, A&A, 625, A5
Krone-Martins, A. & Moitinho, A. 2014, A&A, 561, A57
Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, A&A, 625, A135
Laporte, C. F. P., Belokurov, V., Koposov, S. E., Smith, M. C., & Hill, V. 2020a, MNRAS, 492, L61
Laporte, C. F. P., Famaey, B., Monari, G., et al. 2020b, arXiv e-prints, arXiv:2006.13876
Laporte, C. F. P., Johnston, K. V., Gómez, F. A., Garavito-Camargo, N., & Besla, G. 2018, MNRAS, 481, 286
Laporte, C. F. P., Johnston, K. V., & Tzanidakis, A. 2019a, MNRAS, 483, 1427
Laporte, C. F. P., Minchev, I., Johnston, K. V., & Gómez, F. A. 2019b, MNRAS, 485, 3134
Lata, S., Pandey, A. K., Sagar, R., & Mohan, V. 2002, A&A, 388, 158
Law, D. R. & Majewski, S. R. 2010, ApJ, 714, 229
Li, J., Newberg, H. J., Carlin, J. L., et al. 2012, ApJ, 757, 151
Li, T. S., Sheffield, A. A., Johnston, K. V., et al. 2017, ApJ, 844, 74
Lindegren, L. 2020a, A&A
Lindegren, L. 2020b, A&A
Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, A&A, 616, A2
López-Corredoira, M., Allende Prieto, C., Garzón, F., et al. 2018, A&A, 612, L8
López-Corredoira, M., Garzón, F., Wang, H. F., et al. 2020, A&A, 634, A66
López-Corredoira, M. & Molgó, J. 2014, A&A, 567, A106
Lucey, M., Ting, Y.-S., Ramachandra, N. S., & Hawkins, K. 2020, MNRAS, 495, 3087
Luri, X. 2020, In Prep.
Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018a, A&A, 616, A9
Luri, X., Brown, A. G. A., Sarro, L. M., et al. 2018b, A&A, 616, A9
Luri, X., Palmer, M., Arenou, F., et al. 2014, A&A, 566, A119
Lutz, T. E. & Kelker, D. H. 1973, PASP, 85, 573
Mackereth, J. T., Bovy, J., Leung, H. W., et al. 2019, MNRAS, 489, 176
Majewski, S. R., Ostheimer, J. C., Rocha-Pinto, H. J., et al. 2004, ApJ, 615, 738
Majewski, S. R., Skrutskie, M. F., Weinberg, M. D., & Ostheimer, J. C. 2003, ApJ, 599, 1082
Marasco, A., Fraternali, F., van der Hulst, J. M., & Oosterloo, T. 2017, A&A, 607, A106
Marigo, P., Girardi, L., Bressan, A., et al. 2017, ApJ, 835, 77
Martin, N. F., Ibata, R. A., Bellazzini, M., et al. 2004, MNRAS, 348, 12
Martin, N. F., Ibata, R. A., & Irwin, M. 2007, ApJ, 668, L123
Martinez-Medina, L., Pichardo, B., Peimbert, A., & Valenzuela, O. 2019, MNRAS, 485, L104
Martinez-Medina, L. A., Pichardo, B., & Peimbert, A. 2020, MNRAS, 496, 1845
Martinsson, T. P. K., Verheijen, M. A. W., Westfall, K. B., et al. 2013, A&A, 557, A130
Mateo, M., Mirabal, N., Udalski, A., et al. 1996, ApJ, 458, L13
McGaugh, S. S. 2019, ApJ, 885, 87
McMillan, P. J. 2017, MNRAS, 465, 76
McMillan, P. J. 2018, Research Notes of the American Astronomical Society, 2, 51
Minniti, D., Saito, R. K., Alonso-García, J., Lucas, P. W., & Hempel, M. 2011, ApJ, 733, L43
Momany, Y., Zaggia, S., Gilmore, G., et al. 2006, A&A, 451, 515
Momany, Y., Zaggia, S. R., Bonifacio, P., et al. 2004, A&A, 421, L29

Monari, G., Famaey, B., Siebert, A., Wegg, C., & Gerhard, O. 2019, A&A, 626, A41
Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, ApJ, 569, 245
Peñarrubia, J., Martínez-Delgado, D., Rix, H. W., et al. 2005, ApJ, 626, 128
Poggio, E., Drimmel, R., Lattanzi, M. G., et al. 2018, MNRAS, 481, L21
Price-Whelan, A. M., Johnston, K. V., Sheffield, A. A., Laporte, C. F. P., & Sesar, B. 2015, MNRAS, 452, 676
Purcell, C. W., Bullock, J. S., Tollerud, E. J., Rocha, M., & Chakrabarti, S. 2011, Nature, 477, 301
Ramos, P., Antoja, T., & Figueras, F. 2018, A&A, 619, A72
Ramos, P., Antoja, T., Mateu, C., et al. 2020a, arXiv e-prints, arXiv:2011.01241
Ramos, P., Mateu, C., Antoja, T., et al. 2020b, A&A, 638, A104
Reid, M. J. & Brunthaler, A. 2020, ApJ, 892, 39
Reid, M. J., Menten, K. M., Zheng, X. W., et al. 2009, ApJ, 700, 137
Riello, M. 2020, In Prep.
Robin, A. C., Luri, X., Reylé, C., et al. 2012, A&A, 543, A100
Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523
Rocha-Pinto, H. J., Majewski, S. R., Skrutskie, M. F., Crane, J. D., & Patterson, R. J. 2004, ApJ, 615, 732
Rocha-Pinto, H. J., Majewski, S. R., Skrutskie, M. F., et al. 2006, ApJ, 640, L147
Romero-Gómez, M., Mateu, C., Aguilar, L., Figueras, F., & Castro-Ginard, A. 2019, A&A, 627, A150
Ruiz-Dern, L., Babusiaux, C., Arenou, F., Turon, C., & Lallement, R. 2018, A&A, 609, A116
Sancisi, R. 2004, in IAU Symposium, Vol. 220, Dark Matter in Galaxies, ed. S. Ryder, D. Pisano, M. Walker, & K. Freeman, 233
Sanders, J. L. & Das, P. 2018, MNRAS, 481, 4093
Schlafly, E. F. & Finkbeiner, D. P. 2011, ApJ, 737, 103
Schlegel, D. J., Finkbeiner, D. P., & Davis, M. 1998, ApJ, 500, 525
Schönrich, R. & Aumer, M. 2017, MNRAS, 472, 3979
Schönrich, R., Binney, J., & Dehnen, W. 2010, MNRAS, 403, 1829
Schönrich, R. & Dehnen, W. 2018, MNRAS, 478, 3809
Seabroke, G. 2020, In Prep
Sharma, S., Hayden, M. R., Bland-Hawthorn, J., et al. 2020, arXiv e-prints, arXiv:2004.06556
Sheffield, A. A., Johnston, K. V., Majewski, S. R., et al. 2014, ApJ, 793, 62
Sheffield, A. A., Price-Whelan, A. M., Tzanidakis, A., et al. 2018, ApJ, 854, 47
Siebert, A., Famaey, B., Minchev, I., et al. 2011, MNRAS, 412, 2026
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Slater, C. T., Bell, E. F., Schlafly, E. F., et al. 2014, ApJ, 791, 9
Smart, E. 2020, In Prep.
Sofue, Y. 2020, Galaxies, 8, 37
Strömberg, G. 1927, ApJ, 65, 238
Tosi, M., Di Fabrizio, L., Bragaglia, A., Carusillo, P. A., & Marconi, G. 2004, MNRAS, 354, 225
van der Kruit, P. C. & Freeman, K. C. 1986, ApJ, 303, 556
Vande Putte, D., Garnier, T. P., Ferreras, I., Mignani, R. P., & Cropper, M. 2010, MNRAS, 407, 2109
Wang, H., López-Corredoira, M., Carlin, J. L., & Deng, L. 2018, MNRAS, 477, 2858
Wang, H. F., Huang, Y., Zhang, H. W., et al. 2020a, arXiv e-prints, arXiv:2005.14362
Wang, H. F., López-Corredoira, M., Huang, Y., et al. 2020b, MNRAS, 491, 2104
Wang, Q., Wang, Y., Liu, C., Mao, S., & Long, R. J. 2017, MNRAS, 470, 2949
Weiland, J. L., Arendt, R. G., Berriman, G. B., et al. 1994, ApJ, 425, L81
Widrow, L. M., Gardner, S., Yanny, B., Dodelson, S., & Chen, H.-Y. 2012, ApJ, 750, L41
Williams, M. E. K., Steinmetz, M., Binney, J., et al. 2013, MNRAS, 436, 101
Wu, Z.-Y., Zhou, X., Ma, J., & Du, C.-H. 2009, MNRAS, 399, 2146
Xu, Y., Newberg, H. J., Carlin, J. L., et al. 2015, ApJ, 801, 105
Yanny, B. & Gardner, S. 2013, ApJ, 777, 91
Yanny, B., Newberg, H. J., Grebel, E. K., et al. 2003, ApJ, 588, 824
Yong, D., Carney, B. W., & Teixera de Almeida, M. L. 2005, AJ, 130, 597
York, D. G., Adelman, J., Anderson, J. E. J., et al. 2000, AJ, 120, 1579

Appendix A: Queries to the *Gaia* Archive

In this Appendix we show a few examples of queries to the *Gaia* Archive `https://gea.esac.esa.int/archive/` to retrieve the data:

Listing 1. An example of query to retrieve stars in the rectangular sky patch of the AC20 sample.

```
SELECT * from gaiaedr3.gaia_source WHERE l<190 and
    l>170 and b>-10 and b<10 }
```

Listing 2. An example of query to retrieve the number of stars and average quantities in all healpix of level 8 inside a rectangular patch in the sky.

```
SELECT sub.healpix_8,COUNT(*) as
    N,AVG(phot_g_mean_mag) as avg_g,
    AVG(visibility_periods_used) as avg_vp FROM
    (SELECT gaia_healpix_index(8, source_id) AS
    healpix_8,phot_g_mean_mag,visibility_periods_used
    FROM user_edr3int4.gaia_source WHERE l<240 AND
    l>120 AND b<60 and b>-60 AND ruwe < 1.4) AS sub
    GROUP BY sub.healpix_8
```

Appendix B: Selection of Red Clump stars

In this Appendix we describe the selection of the RC subsample.

First, in order to compute the absolute magnitude, we need good estimates of the extinction A$_\lambda$ in band λ. For each star, one could in principle use the 2D (l, b) maps of reddening, $E(B-V)$, from Schlegel et al. (1998) which estimates the extinction at infinity. However, these 2D extinction values will overestimate the reddening. Since, we have parallax information for our sample, we can use this as a prior for distance and estimate the 3D extinction. For this, we make use of the 3D dust-reddening maps from *Bayestar* (Green et al. 2019). These are derived using a Bayesian scheme that combines *Gaia* parallaxes with photometry from the *2MASS* and *Pan-STARRS* surveys, and covers the sky North of declination of −30°. Only 3 stars in our AC20 sample are missing from *Bayestar*. The multiplicative factor (f_λ) between reddening and extinction that we use is listed in Table B.1 for various bands.

For the RC selection, we first apply the following photometric cuts:

$$BP - G > 0.6, BP - RP_0 > 0.91. \tag{B.1}$$

Then, for each star, we compute the absolute magnitude (M_λ) in each of the *2MASS* bands, and in Gaia G:

$$M_\lambda = m_\lambda - A_\lambda - \mathrm{dmod}, \tag{B.2}$$

using dmod = 5 $log_{10}(100/\varpi'[\mathrm{mas}])$. Here, ϖ' is the parallax corrected for the offset of -17 μas. In Table B.1 we list the literature absolute magnitude $(\bar{M}_\lambda)$ and dispersion in various photometric bands for the RC population. Using this, for each star we can write down a likelihood function per bandpass i, and take their product

$$P_{RC}(m_\lambda, A_\lambda, \varpi) = \prod_i \sqrt{2\pi}\sigma_{\bar{M}_\lambda} \mathcal{N}(M_\lambda || \bar{M}_\lambda, \sigma^*_{M_\lambda}), \tag{B.3}$$

where $\sigma^*_{M_\lambda} = \sqrt{\sigma^2_{er,M_\lambda} + \sigma^2_{\bar{M}_\lambda}}$ combines the propagated error in the absolute magnitude from Equation B.2, σ_{er,M_λ}, and the dispersion in the true absolute magnitude, $\sigma_{\bar{M}_\lambda}$.

Table B.1. Median absolute magnitude $\bar{M}_\lambda$, and dispersion in absolute magnitude $\sigma_{\bar{M}_\lambda}$ for RC stars selected from Hawkins et al. (2017). Also listed are the extinction factors (f_λ) for the four passbands used, with the *2MASS* values taken from Green et al. (2019) and *Gaia* from Casagrande & VandenBerg (2018).

Passband (λ)	$\bar{M}_\lambda$	$\sigma_{\bar{M}_\lambda}$	$f_\lambda = \frac{A_\lambda}{E(B-V)}$
J	-0.93 ± 0.01	0.20 ± 0.02	0.7927
H	-1.46 ± 0.01	0.17 ± 0.02	0.469
K	-1.61 ± 0.01	0.17 ± 0.02	0.3026
G	$+0.44 \pm 0.01$	0.20 ± 0.02	2.74
G_{BP}	-	-	3.374
G_{RP}	-	-	2.035

Table B.2. Parameters used for selecting the *RC* population. *ML* is essentially the confidence level used to set a minimum probability threshold (P_{RC} >). Finally, N_{RC} gives the resulting number of stars classed as RC that lie between 170° $< l <$ 190° and $|b| <$ 10°.

ML	N_{RC}	$P_{\mathrm{RC}} >$	band(s)
3	121857	0.01	*2MASS J*, *Gaia G*

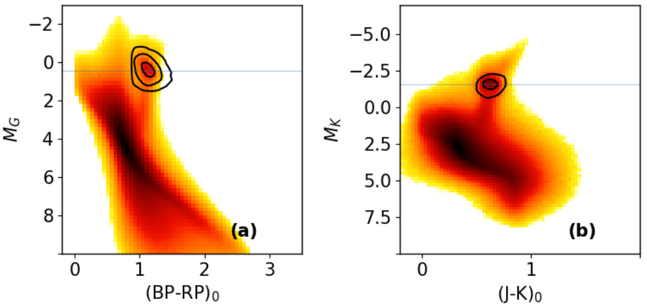

Fig. B.1. HR diagrams for the AC20 sample, with contours marking the RC selection.

For any distribution, the distance between the centroid (x_0) and a point of interest (x_1) can be given in terms of its Mahalanobis distance (ML)

$$ML^2 = (x_1 - x_0)^T \Sigma^{-1} (x_1 - x_0), \tag{B.4}$$

that respects the combined covariance of x_0 and x_1, which we have written as Σ. Essentially, ML is a measure of the distance from the centroid in units of the standard deviation. Then, we can define a p-value, i.e. the probability of finding a value of ML^2 or more extreme under the null-hypothesis of the star not being part of the RC, from a chi-square distribution, and select those stars for which

$$P_{RC} > 1 - P[\chi^2 \leq ML^2] \tag{B.5}$$

i.e., the probability (P_{RC}) is greater than the p-value. In this work we limit our analysis to a maximum of two bands, namely, *Gaia* G and *2MASS* K. So, we use a chi-square distribution with 2 degrees of freedom, and ML is essentially the confidence level used to set a minimum probability threshold. The tolerance parameters used in our selection is shown in Table B.2, and we obtain a high quality RC sample of N_{RC} =121857. The HR diagram with our RC selection is shown in Figure B.1. The parallax quality for the selection is shown in Figure B.2, with the tail of the distribution extending down to $\sigma_\varpi/\varpi \approx 0.8$.

In Figure C.5, we further inspect the RC selection. Panels(a-b) show the absolute magnitude distribution in G and K bands.

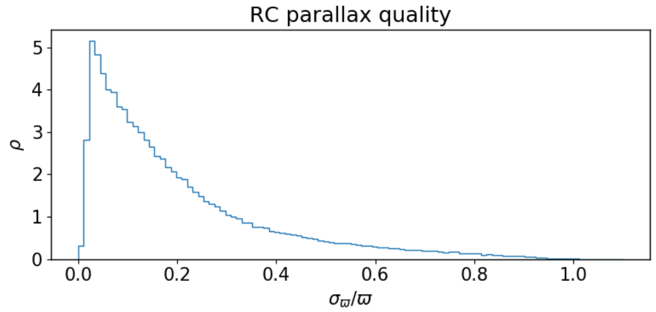

Fig. B.2. Parallax error quality of the RC sample. The tail of the distribution extends down to $\sigma_\varpi/\varpi \approx 0.8$.

We find that the median absolute magnitudes for our sample is offset from their literature values by -0.05 (G) and 0.05 (K) in the two bands. The yellow curves use the *Bayestar* reddening, but we also show the distribution for absolute magnitudes computed with $A_\lambda = 0$, just to illustrate that our extinction correction shifts the distribution in the correct direction. In Panels (c-d), we compare the distances computed as in subsubsection C.2.1, against inverse parallax. It is encouraging to see that the running median for nearby stars lies on the 1:1 line. This is further shown in panels (e-f), where we look at the relative difference between the two distance estimates. Compared to inverse parallax method, our distances are slightly under (over) estimated in G (K) beyond 5 kpc from the Sun. This is likely due to the fact that distance modulus $\propto$ -M_λ. Since the literature absolute magnitudes are slightly offset, this would result in smaller distances, but the effect is minor given the small offset, especially in the K band. Finally, in panels (g-h), is shown the distance error as a function of distance. The errors in the inverse parallax distances are quadratic with d, while the trend is linear for the Red Clump distances. Beyond, d > 5 kpc, the errors in inverse parallax grow significantly, while for RC distances, the prediction is σ_d 1.5 kpc at 10 kpc. The distribution in Galactocentric cylindrical radius R is shown in Figure 7. Our sample extends out to R ~17 kpc, consisting of about 1000 stars at that distance.

Lucey et al. (2020) recently put out a catalogue of 2.6 million RC stars. Their method involves predicting asteroseismic parameters ($\Delta P, \Delta\nu$) and stellar parameters ($\log g$,$T_{\rm eff}$) from spectral energy distributions (SED). They combined photometry from *Pan-STARRS*, *WISE*, *2MASS* and *Gaia*. In their catalogue (hereafter L20) they classified RC stars with contamination rate of ≈ 33% as 'Tier II', and a superior subset with contamination rate of ≈ 20% as 'Tier I'. In Figure B.3 we show the distribution of our sample on a *Kiel* diagram by cross-matching with the L20 catalogue. We notice that their 'Tier I' sample does not have too many cooler stars. Conversely, their less stringent 'Tier II' sample, extends out to $\log g \approx$ 1.8, which is typically the lower limit of the RC range, and thus prone to contamination from regular giants.

Finally, we use *APOGEE*-DR16 (Ahumada et al. 2020), to construct the background *Kiel* diagram. This shows that our RC sample is largely concentrated around the horizontal branch (blue contour), thus missing several common stars with L20, but at the same time is likely a 'purer' sample for the purpose of distance estimation.

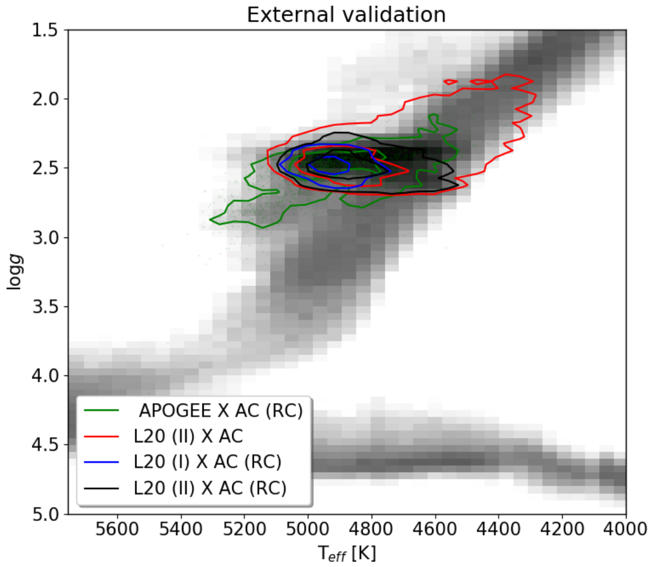

Fig. B.3. External validation comparing the distribution of the AC20 RC stars on a *Kiel* diagram. In grey is the full distribution from *APOGEE*-DR16, and the green contours show common stars between our RC sample and *APOGEE*-DR16. Red contours show common stars between L20 and the entire anticenter sample used here. The blue contours show common stars between L20-'Tier I' (i.e., 20% contamination) and our RC sample. The black contours show common stars between L20-'Tier II' (i.e., 33% contamination) and our RC sample.

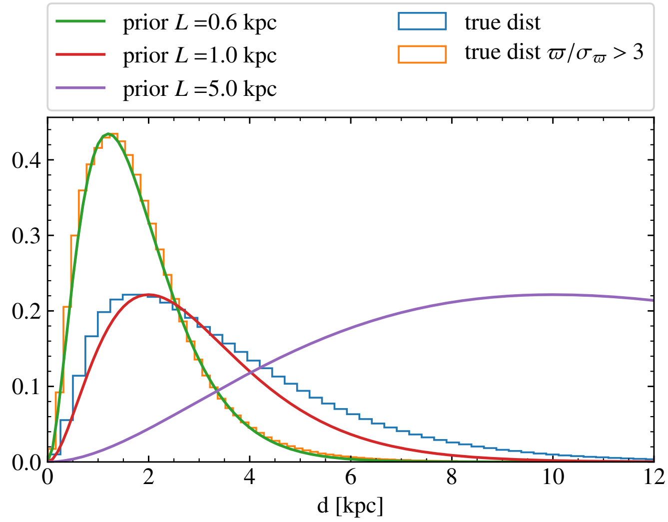

Fig. C.1. Distribution of true distances of GOG in the anticenter. We show all stars in GOG (blue histogram) and stars with $\varpi/\sigma_\varpi > 3$ (orange histogram). The different solid lines are the exponential decreasing prior (Eq. (C.1)) with scalelenght of 0.6, 1 and 5 kpc.

Appendix C: Distances to stars

Appendix C.1: Distance estimates

As discussed in Sect. 3.2 there is no existing perfect recipe for estimating distances from a measured parallax. In this work we approach this problem by testing how robust our conclusions are to the use of different distance estimators. We use three different methods, which we test with the mock *Gaia* data from GOG (described in Sect. 2.2):

1. d_ϖ: simple inversion of parallax $1/\varpi$

2. $d_{\rm PM}$: Bayesian distances with an iterative prior This approach is closely related to that used by Schönrich & Aumer (2017). In general, the Bayesian approach relies on the statement that for an observed parallax, ϖ, and uncertainty, σ_ϖ, the probability of a given distance d is

$$P(d) \propto P(\varpi | d, \sigma_\varpi)\, P(d)$$

where $P(d)$ is the prior on distance. This prior takes three factors into account:

- The volume at distances between d and $d + \delta d$ increases like d^2.
- The true spatial distribution of stars is not uniform.
- Selection effects: the probability of a star at distance d entering the catalogue varies with d because there is a magnitude limit to the survey (because e.g., intrinsically faint stars become too faint to enter the catalogue).

For the distances $d_{\rm PM}$, these considerations lead us to a prior $P(d) \propto d^2 P(r(d)) S(d)$, where $S(d)$ is the selection function, and $r(d)$ is the position in a galaxy of an object at distance d along a given line-of-sight, so $P(r(d))$ is proportional to the density of a model Galaxy. The distance estimate, $\tilde{d}$ and uncertainty σ_d is then found as the expectation value and standard deviation of d given this prior and the measured parallax (with uncertainty).

The model from which we take $P(r(d))$ is taken from McMillan (2018), and has two exponential discs (thin and thick) and a power-law halo. It has no warp. We approximate the selection function as $S(d) \propto \exp(-d/L_s)$ where L_s is a value we determine. Experiments with GOG (see below) and investigation with the *Gaia* data both suggest that this is a reasonable approximation.

Following Schönrich & Aumer (2017), we derive the selection function from the data itself by recognising that $S(d) \propto N(d)/(d^2 \int \rho(d, l, b) \cos b \,{\rm d}l\, {\rm d}b)$ where $N(d)$ is the number of stars in the catalogue at a distance d and the integral over ℓ, b is taken over the field we consider. We don't know $N(d)$, but we can make the approximation that $N(d) \approx N(\tilde{d})$ for some range of distances and subset of the more accurate parallaxes. We use this to find the scale length L_s which enters into $S(d)$. We then iterate this process – using this estimate of the selection function to find new distance estimates, $\tilde{d}$, then using these to make a new estimate of $S(d)$. Experiments with GOG indicate that fitting $S(d)$ for distances $1 < d/{\rm kpc} < 3$ and for stars with $\varpi/\sigma_\varpi > 3$ give a reasonable approximation. The value of L_s we find converges after a few iterations and we find $L_s = 0.963$ for our sample and $L_s = 1.16$ for GOG.

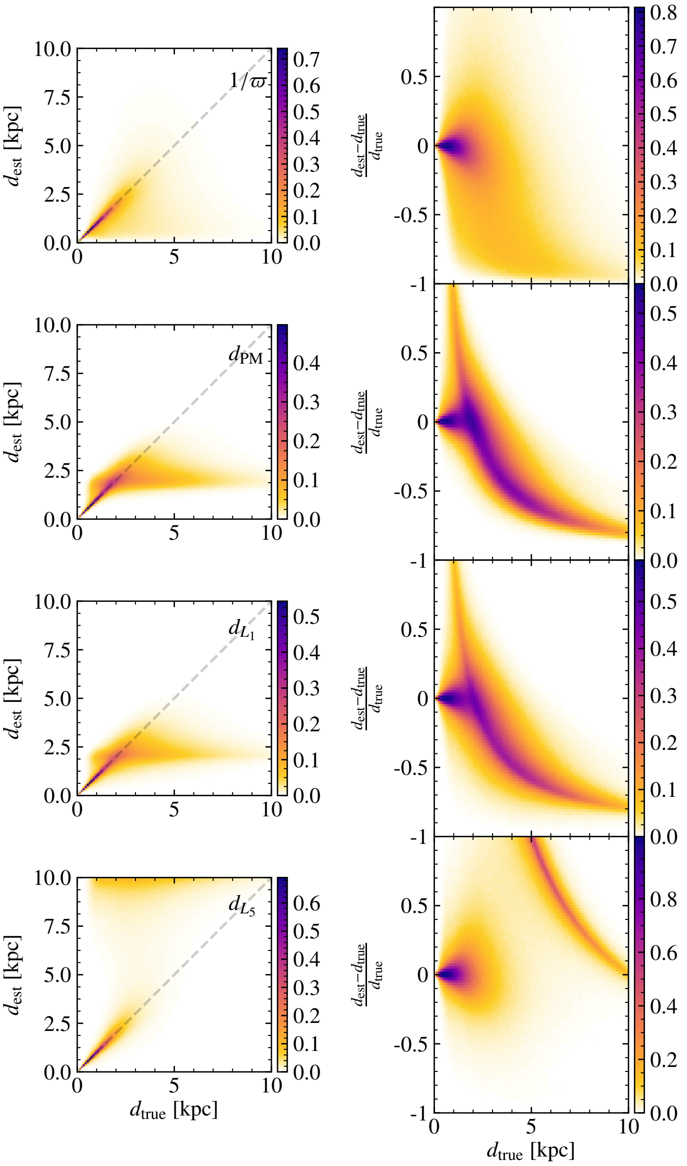

Fig. C.2. True versus estimated distances for different methods. *Left:* Direct comparison between true and estimated distances of GOG for the different distance estimations as indicated in the legends. *Right:* Fractional error in the estimated distance as a function of true distance for the same estimators.

3. d_L: Bayesian distances with exponentially decreasing prior with scale length of L following Bailer-Jones et al. (2018) These distances are computed following Bailer-Jones et al. (2018) with a simpler prior, in particular an exponentially decreasing prior with distance d:

$$P(d|L) = \begin{cases} \dfrac{1}{2L^3}\, d^2 e^{-d/L} & \text{if } d > 0 \\ 0 & \text{otherwise} \end{cases} \tag{C.1}$$

Figure C.1 shows the true distribution of distances of GOG (blue histogram) and the same for a selection of sources with $\varpi/\sigma_\varpi > 3$ (orange histogram). As explained in Bailer-Jones et al. (2018), a good approximation for the maximum likelihood estimate for the scale-length L of the prior Eq. (C.1) of a given distance distribution is $MED(d)/3$, where $MED(d)$ is the median of the distribution. For GOG in the anticenter this turns out to be 0.977 kpc and 0.527 kpc for stars with $\varpi/\sigma_\varpi > 3$. The red and green solid lines show the shape of the prior with $L = 1$ kpc and $L = 0.6$ kpc, respectively, which fairly reproduce the true distribution of distances in each case. We also show the prior for $L = 5$ kpc. Hereafter, we choose two different scale-length L of 1 kpc (that we name d_{L_1}) and 5 kpc (d_{L_5}), motivated by the tests shown below. While Bailer-Jones et al. (2018) uses a scale-length that depends on the sky coordinates, here for simplicity we use a single value for the whole field of 20 × 20 deg.

Appendix C.2: Tests with GOG

Here we test the different distances estimations with GOG. First we note that due to deficiencies in the *Gaia* error model, the uncertainties in the astrometric values in GOG somewhat disagree with the values for EDR3. In particular we see an overestimation of the parallax errors as a function of magnitude G, actually more similar to the DR2 scenario than to EDR3 (Fig. E.1 top). The exercises presented here will thus show a worse case scenario.

Figure C.2 shows the comparison between true distance and estimated distance for the different methods presented above applied to the whole GOG sample, for which the true distances are known. We see in the left panels a good fraction of stars with properly determined distances falling on the 1:1 line (those with small parallax uncertainties). However, we also see large fraction of stars with badly estimated distances corresponding to parallaxes with large uncertainty (including negative parallaxes). For the d_ϖ case (top panels), most of these problematic cases appear scattered in the underestimated region. For Bayesian estimations d_{PM}, d_{L_1} and d_{L_5} (three bottom panels), they appear concentrated at the nearly horizontal line at $d_{\rm est} = 2L$ (coinciding with two times the mode of the prior, that is ~ 2 kpc in the two middle rows, and 10 kpc for the bottom row), completely dominated by the choice of the prior as explained in Bailer-Jones et al. (2018). These numerous uninformative parallaxes forces us to perform a cut in fractional parallax uncertainty, which, unfortunately, may introduce biases in our samples as discussed for instance in Luri et al. (2018b).

Figure C.3 shows the fractional error in the estimated distance as a function of fractional error in parallax with different panels for different distance estimations. All panels show large errors for large parallax uncertainties (smaller ϖ/σ_ϖ) as expected and highlight the need to use a certain criteria to select good parallaxes while finding a proper balance with the final number of sources kept and trying not to bias the sample as a result of eliminating specific populations. Depending on the particular analysis, these considerations might lead to different choices. Here we choose to select sources with $\varpi/\sigma_\varpi > 3$ in the case of the AC20 sample (dash-dotted vertical line) while a more restrictive cut at $\varpi/\sigma_\varpi > 5$ is used in Sect. 5.1 (dotted line).

From Fig. C.3 we also note two important aspects. First, the performance of the 4 different methods is quite similar when one chooses cuts in ϖ/σ_ϖ as the ones mentioned above, with only a slight underestimation of the distances in the case of d_ϖ compared for instance to d_{PM}. Second, we also want to emphasise that, even if the median differences between estimated and true distances are small, at $\varpi/\sigma_\varpi > 5$ (dotted vertical line) 50% of the sources have errors in the derived distances $\gtrsim 20\%$ (sources outside the shaded areas which enclose the other 50%) independently of the method used.

Now focusing on the selection of sources with $\varpi/\sigma_\varpi > 3$, the distance error of these different estimators as a function of true distances is shown in Fig. C.4. We see a slightly better performance of the d_{L_5} at larger distances but a better one for d_{L_1} at nearby distances. The d_ϖ is underestimated in median for all distances while d_{PM} shows overestimated distances at nearby distances, but the contrary beyond 2 kpc. Apart from these little differences, we note that non of the estimators is completely free of bias even with the selection of $\varpi/\sigma_\varpi > 3$, as already mentioned above. We see underestimations of the distance that start to be important (20%) at around 4 kpc and biases larger than ~40% for 25% of the sources at this same distance. Again we emphasise that the parallax errors in GOG are overestimated with respect to *Gaia* EDR3, and therefore the expected biases as a function of distance in EDR3 are possibly smaller than shown here.

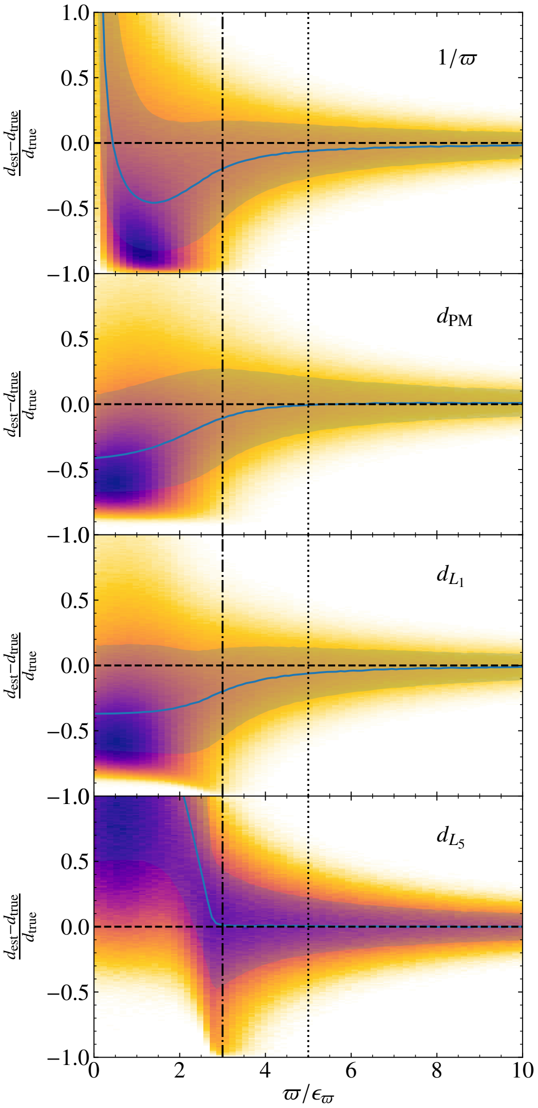

Fig. C.3. Fractional error in the estimated distance as a function of fractional error in parallax. We show the errors for the different distance estimations as indicated in the labels for the GOG sample. Solid lines indicate the median fractional error and shaded areas show the 25 and 75% quartiles.

All these tests show that different priors might work better in different regimes and that there can be multiple criteria to choose which method provides a better estimate (e.g. minimising the median distance error at small versus large distances). We also need to keep in mind that these conclusions are somewhat model dependent, influenced by the particular Milky Way density model and selection function imposed in GOG. Our approach of exploring varied distances estimations wants to mitigate this model-dependency and the appropriateness of different methods and priors in different cases. We highlight that it is necessary to evaluate the impact of this biases and the effects of the parallax quality cut on the different analysis.

So far what we have shown regarded only the estimation of the distance. This estimation and a single value for its uncertainty is then used, together with the proper motions, to calculate

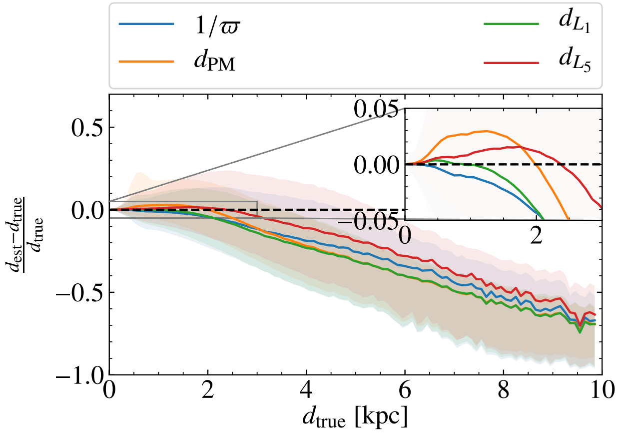

Fig. C.4. Median fractional error in the estimated distance for good quality parallaxes. We show the errors as a function of true distance for different distance estimations as indicated in the labels for stars with $\varpi/\sigma_\varpi > 3$ in the GOG sample. Solid lines indicate the median fractional error and shaded areas show the 25 and 75% quartiles.

velocities and their uncertainties. However, we know this is not strictly correct. On one hand, because the proper motion errors are correlated with the parallax errors and, on the other, because the distribution of uncertainties in the estimated distance in general is not Gaussian and asymmetric. Ideally, then, one would use a method to estimate simultaneously the distance and the tangential velocity of each star. The *Gaia* technical note GAIA-C8-TN-MPIA-CBJ-081 described a way to infer velocities and distances at the same time, from the proper motions and parallax, using a MCMC method. This approach is mathematically more accurate and allows us to deal properly with the correlations between velocities and distances. We have tested it with a random subset of the GOG sample and conclude that: i) the resulting velocities are similar to the ones obtained with the usual and simpler approach, ii) the correlation between velocities and distances is dominated by the transformation rather than by the correlation in the uncertainties, iii) a cut in parallax quality is still necessary and iv) the high computation cost renders it unfeasible to use for even modest-sized samples. For all these reasons, we do not use it here.

Appendix C.2.1: Red Clump distances

For each star classified as RC, we can invert Equation B.2 to calculate the distance modulus. For this we use the literature absolute magnitudes in each band pass, $\bar{M}_\lambda$. The errors in the computed distances using the RC and parallax only is given by,

$$\sigma_{d_{RC,\lambda}} = 0.2\ln(10)\sigma_{M_\lambda} d \tag{C.2}$$

$$\sigma_{d_\varpi} = \sigma_\varpi d^2, \tag{C.3}$$

where σ_{M_λ} is the dispersion in the computed absolute magnitudes of the RC selection, and σ_ϖ is the parallax error. The parameters in Table B.2 are fine tuned in order to maximise the number of RC stars and minimise the dispersion and thus the errors in distances.

As mentioned earlier, we do not apply the 'qfl' quality flag on *2MASS* photometry, but instead use the photometric errors to decide if the distances will be estimated using the K band or G band. In general, the K band suffers from lower extinction than the broader G band, so we prefer to use distances estimated using K. However, if for a given star the photometric errors, (*e_jmag*| *e_kmag*) > 0.025, the typical value above which photometry in *2MASS* becomes unreliable, then we estimate distances for these using the G band. This is illustrated in Figure C.6, where we compare our distance estimates to the external catalogue of L20. Essentially, for stars with poor *2MASS* photometry we overestimate the distances if the K band is used. Replacing these with G band estimates results in a much better agreement with the external catalogue.

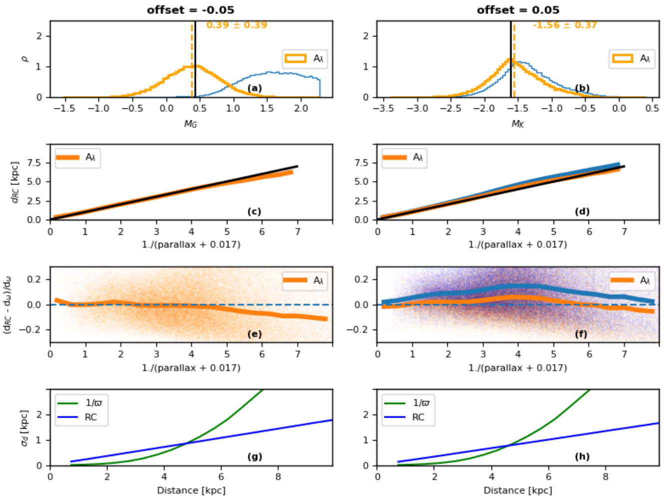

Fig. C.5. Red clump sample inspection. (a-b) Absolute magnitudes in G and K for the selected sample. The yellow curves use the 3D extinctions from *Bayestar*, while the blue curves are for zero extinction shown just for illustration of shift towards the correct literature value upon reddening correction. (c-f) compare the RC distances to inverse parallax, while panels (g-h) show the error in distances as a function of d for the two methods. Beyond 5 kpc, RC distances become more reliable than inverse parallax.

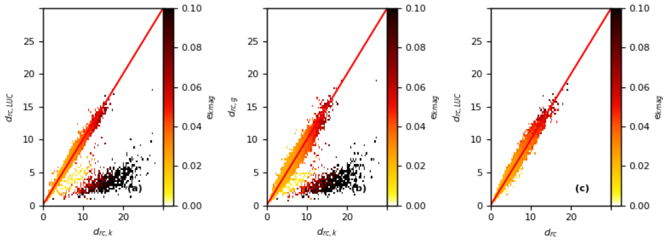

Fig. C.6. Red Clump distance validation with the external catalogue L20. To enhance the illustration we use a larger RC sample here (147° < l < 219° and $|b|$ < 30°). Panel a) Comparison to L20 shows the presence of a population for which distances are overestimated using the K band. This is due to very high photometric errors i.e., (*e_jmag*| *e_kmag*) > 0.025. b) Comparison between G and K band derived distances also highlights the same trend, i.e., K band distances are overestimated for poor photometry stars. c) Replacing K band estimates with G where (*e_jmag*| *e_kmag*) > 0.025 improves agreement with L20.

Appendix C.2.2: Comparison of the different distances for EDR3

Finally, Fig. C.7 compares all sets of distances derived in this work using the d_{PM} case as a baseline (see caption for more details).

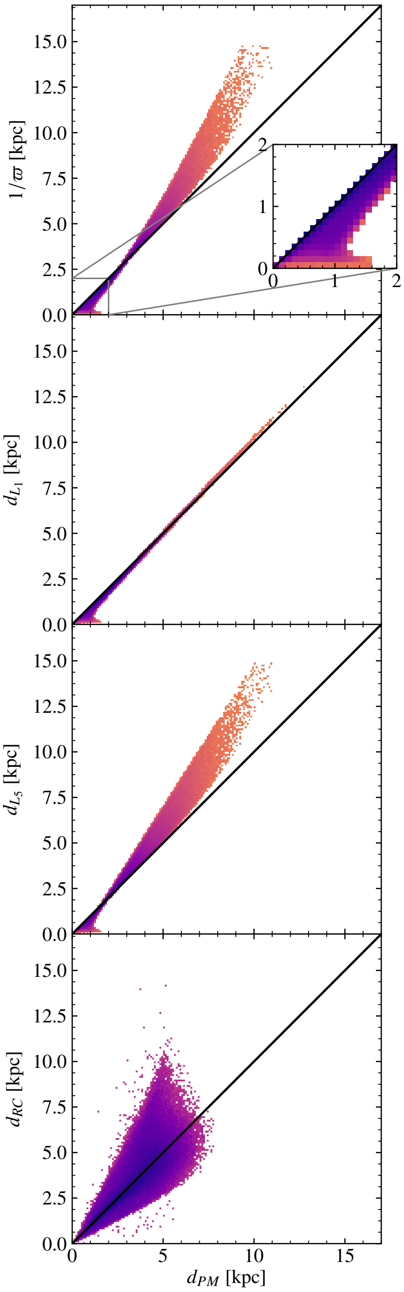

Fig. C.7. Comparison of the different distances used in this study. The comparison is done with respect to the d_{PM} distances. The discrepancies are small for small distances. For the case of the RC, we only compare stars with $\varpi/\sigma_\varpi > 3$, since the rest of RC sources are not included in our set of d_{PM}. This is then misleading since for these stars the parallax retrieves better distances, but the real gain for the RC occurs exactly for the stars missing in this panel, in the regime where the photometric distances might be better than the ones from parallax alone. The peculiar shape shown in the inset of the top panel and present in the three top panels is composed of stars with large parallax error, for which the expectation values used in the d_{PM} estimation are larger than for instance the medians used in d_L.

Appendix D: Parallax zero point

In this Appendix we illustrate the differences in distance and velocities when different parallax zero points are used (Fig. D.1), and we reproduce several figures of the main part done with and without different parallax zero points (Figs. D.2 and Fig. D.3).

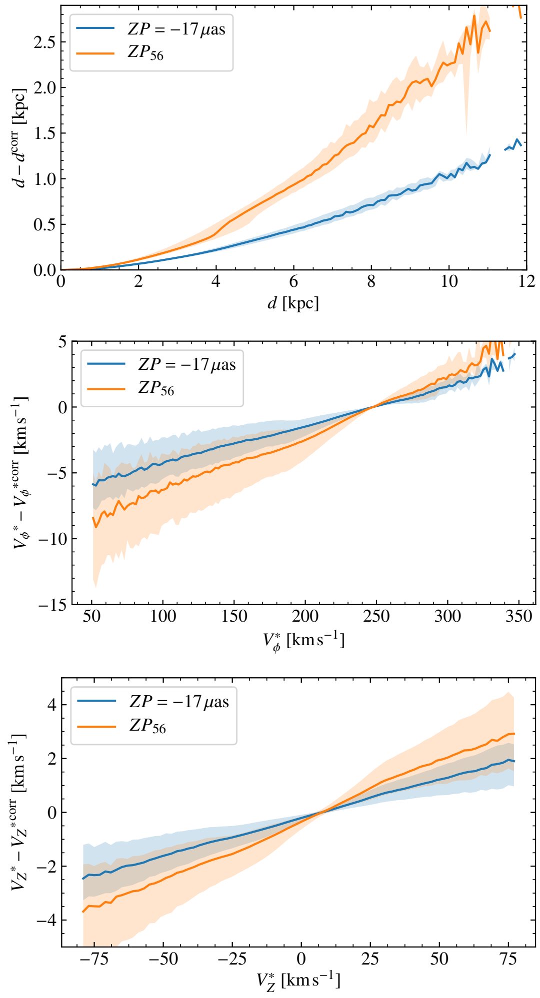

Fig. D.1. Effects of the zero point in parallax on distances and velocities. The comparison is done with respect to the case with the Bayesian distances d_{PM} where the zero point is not considered (x axis) and the shaded areas show the percentiles 10 and 90 (i.e. they enclose 80% of stars). In the top panel, we see how not correcting for the zero point produces overestimated distances. The zero point prescription ZP_{56} reduces even more the distances compared to the case of a fixed zero point $ZP = -17\,\mu$as. The velocities (middle and bottom) scale linearly with the distance and thus we see the absolute magnitude of the velocities being larger when the zero point is not considered. We see null differences in the case of null proper motion, that is when the velocities equal that of the Local Standard of Rest.

Appendix E: Additional material

In this Appendix we present a miscellanies set of plots that serve as supporting material to the rest of the sections. A describing text can be found in each of the figures.

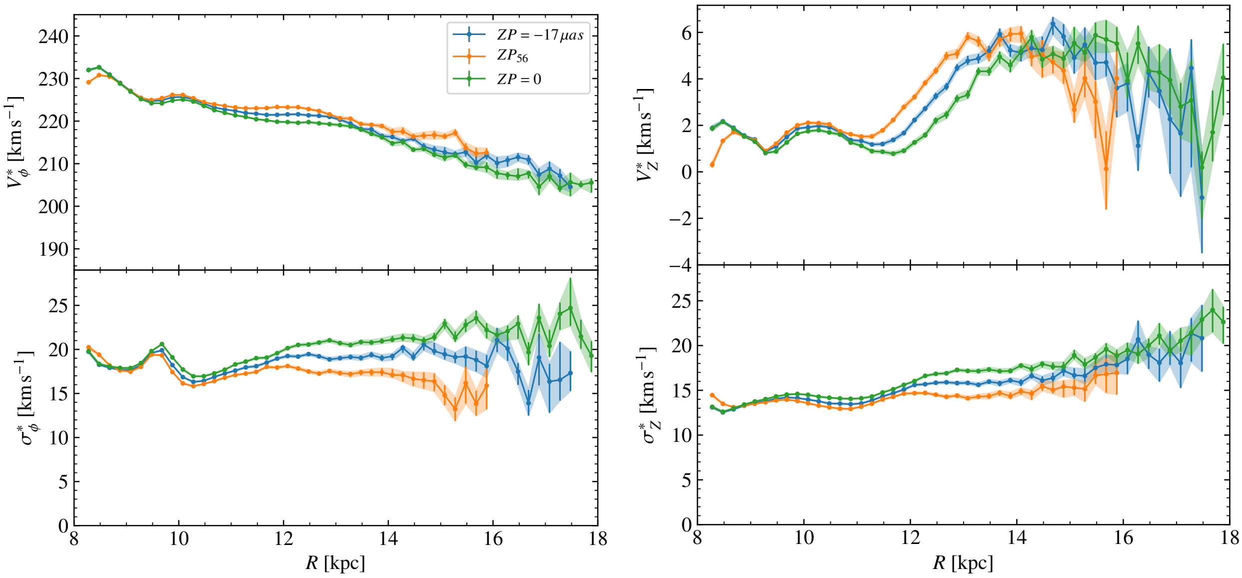

Fig. D.2. Velocity profiles for different parallax zero points. In the rotation curve (top left), as expected, the rotation curve computed using ZP_{56} (orange curve) is slightly shifted to the left (R decreases by about 0.5 kpc at R = 14 kpc) and V^*_ϕ also decreases, but always in amounts smaller than $\sim 2\,\mathrm{km\,s^{-1}}$. In the vertical velocity plot (top right), we observe similar effects, though a notable effect is seen in the first kpc. The velocity dispersions (computed as the *mad* values, bottom) appear also slightly different, with ZP_{56} yielding smaller dispersions but without changing the overall shape.

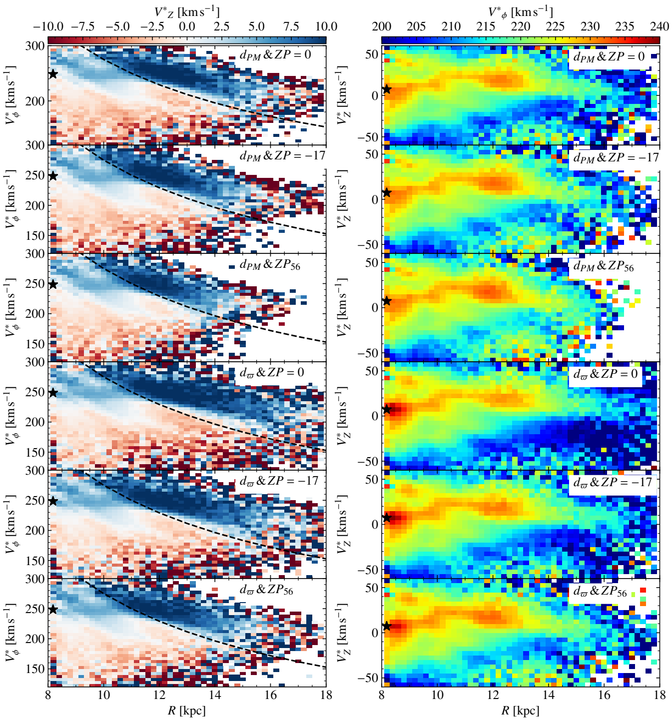

Fig. D.3. Phase space projections for different parallax zero point. The plot reproduces panels b and e of Fig. 14 using different distance estimations and parallax zero point as indicated in the legends. As explained in other parts of the article, the correction of the zero point combined with the different distance estimators used produce a change in the distance scale but in any case induces or removes the phase space substructure such as the one observed in this panels. The smallest distances are found when the Bayesian distances d_{PM} and the zero point ZP_{56} are used.

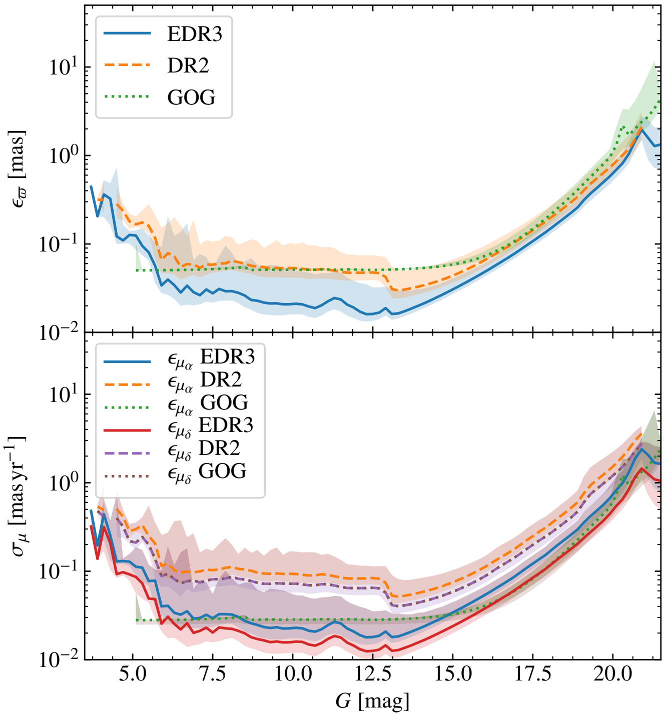

Fig. E.1. Comparison of the astrometric uncertainties for DR2, EDR3 and GOG. Due to deficiencies in the GOG *Gaia* error model, the astrometric uncertainties in GOG do not match perfectly those for EDR3. The error model retrieves unique values of the formal uncertainties as a function of G, while a large range is obtained for the data (shaded areas showing the 10% and 90% percentiles). We also see an overestimation of the parallax errors (top), which actually look more similar to the DR2 scenario than to EDR3. By definition, the errors in μ^*_α and μ_δ are the same for GOG. The errors of the proper motions are closer to the true uncertainties although no distinction between the different components is made for this mock data.

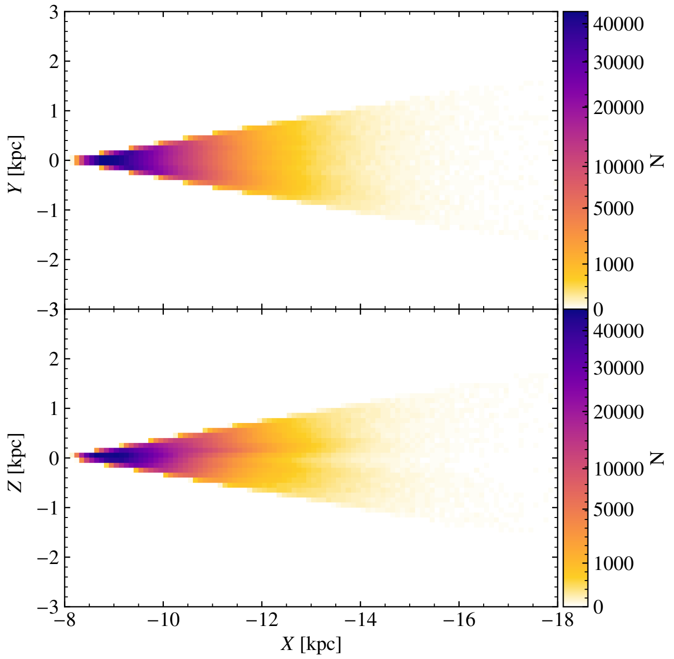

Fig. E.2. Spatial distribution of the AC20-$\varpi/\sigma_\varpi > 3$. a) Counts in bins of 0.1 kpc in the X-Y projection. b) Same but for the X-Z projection.

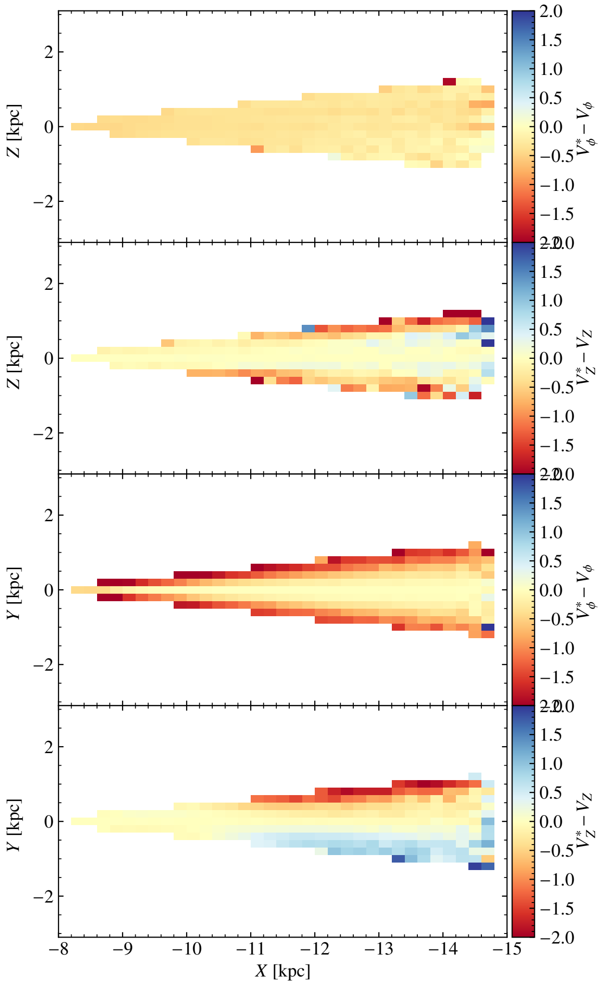

Fig. E.3. Error in the velocities for GOG when using our approximations. Median differences between V^*_ϕ and V_ϕ, and between V^*_Z and V_Z (Eq. 4 and 5) in bins in the X-Y and X-Z projections.

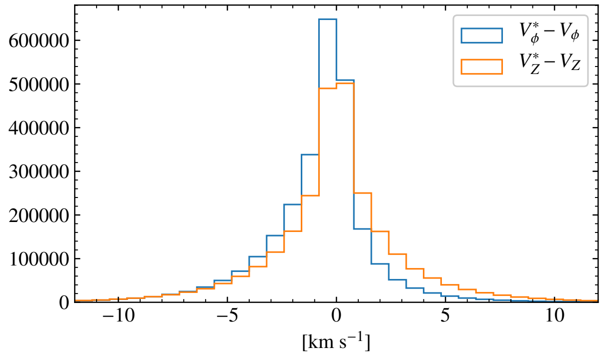

Fig. E.4. Error in the velocities for GOG when using our approximations (part 2). The histogram of the differences between V^*_ϕ and V_ϕ, and between V^*_Z and V_Z is shown. The 10% and 90% percentiles of the differences are -1.5 and 3.9 km s^{-1} and -3.6 and 3.4 for $V_\phi*$ and for V_Z*, respectively.

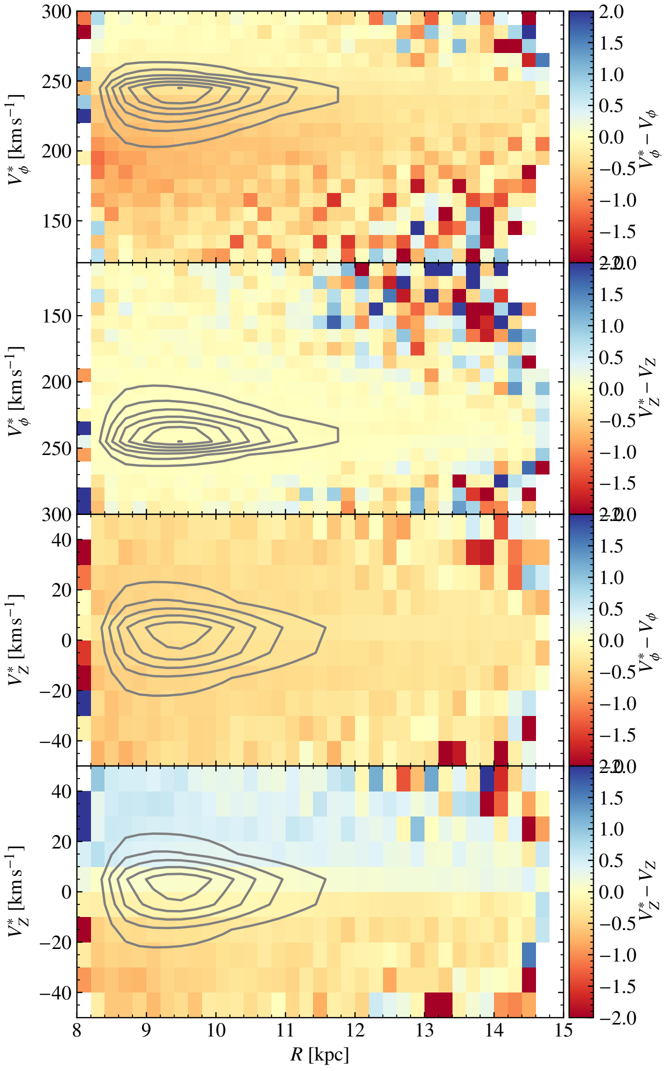

Fig. E.5. Error in the velocities for GOG when using our approximations (part 3). Median differences between V^*_ϕ and V_ϕ, and between V^*_Z and V_Z in bins in the R-V^*_ϕ and R-V^*_Z projections.

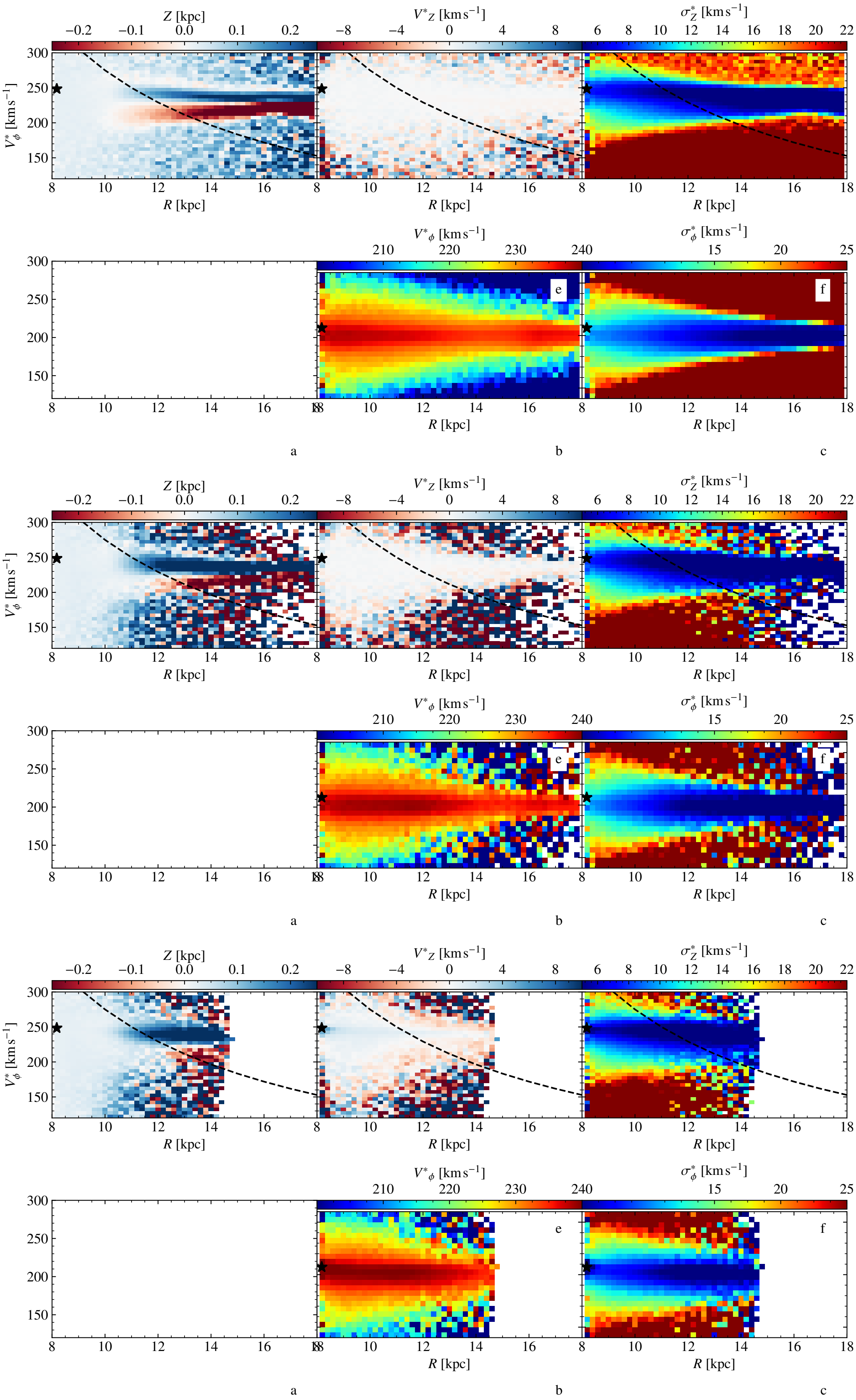

Fig. E.6. Phase space projections for model and mock data. Same as Fig. 14 but for the UM (top), the UM with the sources that in GOG have $\varpi/\sigma_{\varpi} > 3$ (middle), and for GOG with the selection $\varpi/\sigma_{\varpi} > 3$ (bottom). The phase space spiral does not exist in GOG and is not shown in panels d.

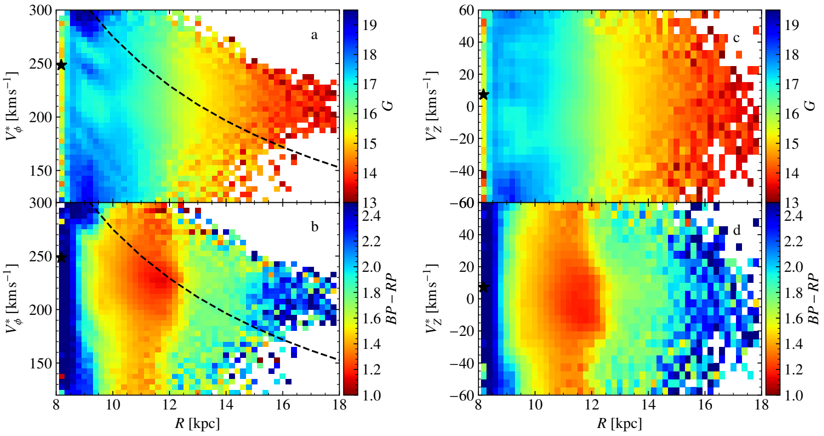

Fig. E.7. Photometry in different phase space projections. Median magnitudes G (top) and colour $BP-RP$ (bottom) in the R-V^*_ϕ plane (left) and R-V^*_Z plane (right) for the AC20-$\varpi/\sigma_\varpi > 3$ sample.

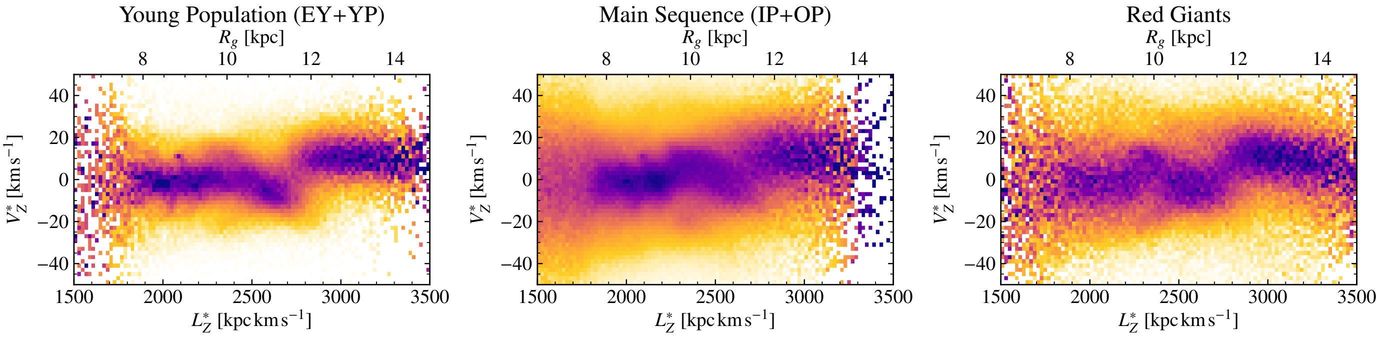

Fig. E.8. Structures in the vertical velocity and angular momentum space for different populations. As in the top panel of Fig. 15, these show a column normalised histogram of star numbers in the L^*_z, V^*_Z plane but for a given population (as in Sect. 2.2). In all cases the feature at $\sim 2750\,\mathrm{km\,s^{-1}}$ kpc is clearly visible. The young population has the lowest velocity dispersion, and therefore shows the feature most cleanly.