The Astrophysical Journal, 966:108 (14pp), 2024 May 1 https://doi.org/10.3847/1538-4357/ad35ba

Synthetic Gaia DR3 Surveys from the FIRE Cosmological Simulations of Milky Way-mass Galaxies

Tri Nguyen[1,2,7], Xiaowei Ou[1,7], Nondh Panithanpaisal[3], Nora Shipp[1,6], Lina Necib[1,2], Robyn Sanderson[3,4], and Andrew Wetzel[5]

[1] Department of Physics and MIT Kavli Institute for Astrophysics and Space Research, 77 Massachusetts Avenue, Cambridge, MA 02139, USA; tnguy@mit.edu, xwou@mit.edu

[2] The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

[3] Department of Physics & Astronomy, University of Pennsylvania, 209 S 33rd St., Philadelphia, PA 19104, USA

[4] Center for Computational Astrophysics, Flatiron Institute, 162 5th Ave., New York, NY 10010, USA

[5] Department of Physics & Astronomy, University of California, Davis, CA 95616, USA

[6] McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA

Received 2023 July 17; revised 2024 March 18; accepted 2024 March 18; published 2024 April 29

Abstract

The third data release (DR3) of Gaia has provided a fivefold increase in the number of radial velocity measurements of stars, as well as a stark improvement in parallax and proper motion measurements. To help with studies that seek to test models and interpret Gaia DR3, we present nine Gaia synthetic surveys, based on three solar positions in three Milky Way-mass galaxies of the *Latte* suite of the FIRE-2 cosmological simulations. These synthetic surveys match the selection function, radial velocity measurements, and photometry of Gaia DR3, adapting the code base `Ananke`, previously used to match the Gaia DR2 release by Sanderson et al. The synthetic surveys are publicly available and can be found at http://ananke.hub.yt/. Similarly to the previous release of `Ananke`, these surveys are based on cosmological simulations and thus are able to model nonequilibrium dynamical effects, making them a useful tool in testing and interpreting Gaia DR3.

Unified Astronomy Thesaurus concepts: Hydrodynamical simulations (767); Gaia (2360); the Milky Way (1054); Milky Way dynamics (1051)

1. Introduction

The Gaia mission (Gaia Collaboration et al. 2016) has revolutionized the study of our Galaxy, the Milky Way (MW). The second data release (DR2; Gaia Collaboration et al. 2018) provided positions, proper motions, and parallaxes for over one billion stars, including the first kinematic measurements of many stars across the Galaxy. In addition, DR2 included radial velocities for ∼7 million stars, making it the largest six-dimensional kinematic catalog at the time. These data have enabled the discovery of new merging events, such as the Gaia Sausage Enceladus (Belokurov et al. 2018; Helmi et al. 2018) and the Kraken (Kruijssen et al. 2019) (see Helmi 2020 for a review), the construction of a new 3D dust map of the MW (Green et al. 2019), a detailed study of open clusters to unveil the history of the Galactic disk (Cantat-Gaudin et al. 2018), an accurate measurement of the circular velocity of the Galaxy (Eilers et al. 2019), and detailed studies of the fine resonances of the MW disk (see, e.g., Antoja et al. 2018). The third data release (DR3; Gaia Collaboration et al. 2021, 2023) builds upon DR2, incorporating 12 months of additional observations, and significantly increasing the catalog of stars with 6D phase-space measurements, including radial velocities, to ∼33 million stars, as well as reducing uncertainties on and increasing the size of the sample of stars with full astrometry. These data have further enabled a deeper understanding of the dynamics of the Galaxy, for example by extending measurements of the circular velocity of the MW to larger distances (Ou et al. 2024).

Synthetic catalogs and mock observations generated from cosmological simulations provide a valuable comparison to these rich observations of our own Galaxy. They enable tests of analytical tools and of our ability to recover true properties of our Galaxy from observations. Sanderson et al. (2020, hereafter S20) produced nine Gaia DR2 synthetic surveys of the *Latte* suite of simulations (Wetzel & Hopkins 2016; Hopkins et al. 2018), using the code `Ananke`. Such synthetic surveys have been used in many studies involving the dynamics of the MW, for example to estimate the detectability of simulated stellar streams (Shipp et al. 2023), as a training set for a neural network that built the first catalog of accreted stars in the MW (Ostdiek et al. 2020), leading to the discovery of the prograde local structure Nyx (Necib et al. 2020), as a framework to test the ability of unsupervised machine learning techniques to reproduce the stellar phase-space density (Buckley et al. 2023), and as a link to connect the formation history and the components of the MW (Belokurov et al. 2020). In this work, we present synthetic Gaia DR3 surveys based on the same suite of *Latte* simulations.

The *Latte* simulations first introduced in Wetzel & Hopkins (2016) are baryonic zoom-in simulations of MW analogs from the Feedback in Realistic Environments (FIRE) project (Hopkins 2015; Hopkins et al. 2018). With an initial stellar particle mass resolution of 7070 $M_\odot$, the *Latte* simulations resolve stellar populations down to the masses of individual star clusters. They self-consistently model baryonic processes, including star formation and the metal enrichment of gas, which is essential for accurately calculating the extinction of

---

[7] Authors Nguyen and Ou have contributed equally to this work, and therefore this paper should be cited as Nguyen, Ou, et al. (2023).

Table 1
The Coordinates of each LSR as Shown in Table 4 of S20

Label	$x_{\rm LSR}$ (kpc)	$y_{\rm LSR}$ (kpc)	$z_{\rm LSR}$ (kpc)	$v_{x,{\rm LSR}}$ (km s^{-1})	$v_{y,{\rm LSR}}$ (km s^{-1})	$v_{z,{\rm LSR}}$ (km s^{-1})	$v_{R,{\rm LSR}}$ (km s^{-1})	$v_{Z,{\rm LSR}}$ (km s^{-1})	$v_{\phi,{\rm LSR}}$ (km s^{-1})
`m12i-lsr-0`	0.0	8.2	0.0	224.7092	−20.3801	3.8954	−17.8	−3.9	224.4
`m12i-lsr-1`	−7.1014	−4.1	0.0	−80.4269	191.7240	1.5039	−24.4	−1.5	210.9
`m12i-lsr-2`	7.1014	−4.1	0.0	−87.2735	−186.8567	−9.4608	22.1	9.5	206.5
`m12f-lsr-0`	0.0	8.2	0.0	226.1849	14.3773	−4.8906	14.9	4.9	227.9
`m12f-lsr-1`	−7.1014	−4.1	0.0	−114.0351	208.7267	5.0635	−3.4	−5.1	244.3
`m12f-lsr-2`	7.1014	−4.1	0.0	−118.1430	−187.7631	−3.8905	−11.4	3.9	227.4
`m12m-lsr-0`	0.0	8.2	0.0	254.9187	16.7901	1.9648	16.2	−2.0	254.7
`m12m-lsr-1`	−7.1014	−4.1	0.0	−128.2480	221.1489	5.8506	2.4	−5.9	252.7
`m12m-lsr-2`	7.1014	−4.1	0.0	−106.6203	−232.2056	−6.4185	15.4	6.4	265.3

observed stars. At the same time, they incorporate the effects of galaxy formation in a cosmological context, including a realistic history of mergers and accretion events.

`Ananke` is a framework for producing synthetic surveys based on the FIRE simulations, first presented by S20. Such work is based on Galaxia (Sharma et al. 2011), which generated synthetic surveys of the MW based on kinematic distributions and *N*-body simulations. The framework entails sampling a population of individual stars from simulated star particles, assigning them realistic physical properties, and applying a simple error model to produce mock observations. `Ananke` has been applied to produce mock observations of other surveys from a range of simulated data sets, such as APOGEE (Beaton et al. 2022), Dark Energy Survey (DES Collaboration 2005, 2016), and the Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al. 2019) in Shipp et al. (2023).

In this paper, we use `Ananke` to produce synthetic Gaia DR3 surveys of three MW analogs from the *Latte* simulation suite, focusing on the updates to the surveys compared to S20. This paper is organized as follows: in Section 2 we review the simulations and mock catalogs used in this work, in Section 3 we discuss the synthetic Gaia DR3 observations, and in Section 4 we present the resulting synthetic surveys, comparing them with those from Gaia DR2. We list the columns of the public release and their definitions in Section 4.3, and discuss the use cases and limitations of these synthetic surveys in Section 5.

2. Simulations and Mock Catalogs

In this section, we outline the different steps to build both a mock catalog and a synthetic survey. We first seek to define these two terms. Generating a *mock catalog* consists of spawning stars from star particles in the initial simulations. This process is independent of the target survey. The star particles in simulations typically have masses orders of magnitude greater than a single solar mass, depending on the initial resolution of the simulation. For the FIRE simulations, the star particles have a mass $\sim 7000\,M_\odot$. Generating a *synthetic survey* involves incorporating the specifics of a particular survey into the catalog of simulated stars, including, for example, photometric passbands, measurement errors, dust extinction, and the observer's location.

In order to build the new synthetic `Ananke` DR3 survey, we use the same three zoom-in simulations of MW-mass galaxies from the *Latte* suite of FIRE-2 simulations as in S20 (`m12i`, `m12f`, `m12m`).[8] The choice of these specific simulations is motivated by Wetzel & Hopkins (2016) and Sanderson et al. (2018), which have shown that these simulated galaxies reasonably reproduce a realistic galaxy of MW mass. Specifically, studies have shown that these simulated galaxies have galactic bar morphology (Debattista et al. 2019; Ansar et al. 2023), stellar thin and thick disk morphology (Ma et al. 2017), gas kinematics (El-Badry et al. 2018), etc., that are broadly similar to the MW. It is impossible, however, for any simulated galaxy to fully reproduce all properties of the MW, and thus we caution those using the synthetic surveys to be aware of the differences and refer to the literature mentioned above for details.

2.1. Locations of the Sun

To build a synthetic survey, we must assume the location of the observer, which we define as the solar position in the simulation. The procedure we adopt here for the coordinate transformation and the definition of the local standards of rest (LSRs) remain unchanged from S20, and we briefly summarize them here. We assume that the Sun is at $R_\odot = 8.2$ kpc (Bland-Hawthorn & Gerhard 2016) in the three simulations,[9] and define the principal axes based on the moment of inertia tensor of the youngest stars (with ages <1 Gyr) located within $R_\odot$. We note that it is possible to scale the solar position differently for each simulation based on disk scale radii, local density, or the local circular velocity. As pointed out in S20, however, doing so introduces extra complexity for users to accommodate different solar radii for different simulations, and the variation in the results is not significant compared to changing the azimuthal position of the Sun.

The three positions of the Sun in each galaxy are chosen to be evenly distributed in azimuthal angle to allow different view angles of axisymmetric features such as the bar and spiral arms, and at vertical distance $Z_\odot = 0$ kpc. We define the velocity of the LSR as the median velocity of the star particles within 200 pc of the solar position. We summarize the positions and velocities of the LSRs in Table 1, which matches Table 4 of S20.

2.2. Building a Mock Catalog

In this section, we discuss the procedure to build a *mock catalog* by converting the star *particles* from the FIRE-2 simulations into synthetic stars, leaving the construction of

---

[8] For a brief overview of the simulation and the simulated galaxies used, we refer the reader to Section 2 of S20.

[9] This is an appropriate approximation given that these simulations have comparable scale heights and radii to the MW.

the *synthetic survey* in which we add the correct properties to these synthetic stars as drawn by the survey to Section 3.

Each star particle, with a mass $M_* \sim 7070\,M_\odot$, is assumed to represent a population of synthetic stars with a single age and metallicity. To generate such mock catalogs, S20 used the GALAXIA code (Sharma et al. 2011). In this work, we adopt the same mock catalogs as in S20, modifying the stellar isochrones used in the generation of stars to the updated Gaia DR3 isochrones. This modification is performed at Step 2 below, while keeping the masses and the phase-space positions the same as in S20 in Steps 1 and 3.

To build a mock catalog, we perform the following three steps. We will leave a detailed description of the DR3 isochrones to Section 3.1.

1. First, we sample the stellar masses of synthetic stars using the initial mass function (IMF) in Kroupa (2001) until the total mass equals the mass of the particle.[10]
2. Select the isochrone model that is closest in age and metallicity to the parent particle and obtain stellar properties and Gaia DR3 passband magnitudes by interpolating across initial stellar mass. Only stars with estimated unextincted apparent Gaia DR2/3 magnitudes of $3 < G < 21$ are kept in the catalog, before applying the full selection function in Section 3.4.
3. Assign the positions in phase space to each star by sampling from a parabolic or Epanechikov density kernel (Epanechnikov 1969) centered on the parent particle. The density kernel generates smooth phase-space distributions from the discrete individual parent particles. We then sample the position and velocities of the synthetic stars from the smooth distribution. The smoothing kernel is computed from the six-dimensional phase-space coordinates using the ENLINK code (Sharma & Steinmetz 2006; Sharma & Johnston 2009). Similarly to S20, we use the nearest eight neighboring star particles to compute the kernel size. The kernel size is taken with respect to two independent smoothing lengths, one for the distances and one for the velocities. The size is proportional to the geometric mean of the smoothing lengths along each of the three dimensions. To preserve the dynamic ranges of the different stellar populations and avoid the oversmoothing of structures from different stellar populations, a kernel is computed for in situ stars, which are defined as those formed within 30 physical kpc of the main galaxy, while a separate kernel is computed for stars formed outside this radius. In addition, we subdivide in situ stars into eight age bins corresponding to the populations of the Besançon Milky Way model (Table 2.1 of Robin et al. 2012) and compute a different kernel for each of them. For full details on the phase-space smoothing procedure, we refer readers to Section 4.3 of S20. This kernel is not optimized for small-scale structures, and in some cases may introduce unphysical features into substructures such as low-mass satellite galaxies and stellar streams. For example, Shipp et al. (2023) adopted a different kernel (albeit also based on the Epanechikov kernel) but with the 16 neighboring star particles and a kernel size that is inversely proportional to the cube root of the local density around each parent particle, to properly smooth out stellar streams.

3. Synthetic Surveys

We describe the procedure used to produce the `Ananke` DR3 *synthetic surveys*. As mentioned in Section 2, we use the mock catalogs presented in S20, and apply updated DR3 isochrones (Section 3.1), extinction modeling (Section 3.2), observational uncertainty modeling (Section 3.3), and selection function (Section 3.4).

3.1. Isochrones

We use updated Gaia DR3 passbands and isochrones from Padova CMD v3.6[11] to generate updated intrinsic Gaia DR3 magnitudes for stars in the mock catalogs in the G, $G_{\rm BP}$, and $G_{\rm RP}$ bands. The photometric system follows the revised and expanded library described in Chen et al. (2019), adopting a revised spectral energy distribution for Vega from Bohlin et al. (2020). Two assumptions are made while adopting the isochrones. First, circumstellar dust is ignored as it mostly affects the bright end of the isochrones, where the grid is the sparsest. Therefore, linear interpolation with the circumstellar dust included creates unphysical features when stars fall between these sparse grid points. Second, we remove the isochrone grid points representing white dwarfs, as the transition from the tip of the red giant branch to the white dwarf is not modeled by GALAXIA. Since GALAXIA takes the edge value for magnitudes when a star is outside the isochrone grid, stars beyond the last non-white dwarf grid point are all assigned the same magnitudes, creating artificial overdensities at the tip of the giant branches in the final sample. We expect these stars to be potential white dwarfs and flag all affected stars in the final synthetic survey as `flag_wd` and recommend removing stars with `flag_wd` set to 1 before conducting analysis. We expect a minimal effect on the overall completeness of the sample as a result of this treatment. As shown by Gentile Fusillo et al. (2021), 359,073 white dwarfs are confidently identified in Gaia DR3, comprising less than 0.05% of the full Gaia sample. Even accounting for the fact that the white dwarf catalog presented by Gentile Fusillo et al. (2021) is less complete in crowded regions near the Galactic plane, the total white dwarf count in the actual Gaia catalog is expected to be a tiny fraction of the full sample. Thus, for the synthetic survey, the overall loss in stellar count and impact on sample completeness as a result of this cut are expected to be minimal.

3.2. Extinction Modeling

We adopt a self-consistent extinction model similar to Section 5.1 of S20, which we briefly describe below. The FIRE-2 simulations do not resolve the creation and destruction of dust grains, so we assume the line-of-sight extinction by dust traces the metal-enriched gas in the simulations. We calculate the reddening $B - V$ of each star using the metal-weighted column density of hydrogen along the line of sight between the star and the solar position. The extinction is therefore calculated self-consistently, using the gas and metal distributions of each individual simulated

---

[10] The number of stars sampled is required to be an integer, while the fraction of the IMF within a subrange of mass is not. Some rounding is assumed, and given that the highest possible stellar mass is still two orders of magnitude lower than the mass of the star particle, in this work, as in S20, we assume that this is a valid approach with a small fractional error.

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

galaxy, and thus accurately captures the spatial structures of the galaxy (see Figure 12 of S20 for the dust map of the simulated galaxy). The extinction at 550 nm, A_0, is calculated using the standard relation, $A_0 = 3.1E(B - V)$ (Johnson 1965; Schultz & Wiemer 1975; Whittet & van Breda 1980), and then converted into extinction in the Gaia DR3 passbands.

Using the coefficient A_0 from the `Ananke` DR2 mock catalogs, as described in S20, we recalculate the extinction coefficients $A_{G,\mathrm{BP,RP}}$ in the Gaia DR3 passbands. We adopt the extinction conversion relation provided by the Gaia collaboration as part of the auxiliary data for eDR3 to compute $A_{G,\mathrm{BP,RP}}$ as functions of A_0 and the unextincted color $(G_{\mathrm{BP}} - G_{\mathrm{RP}})$.[12] Specifically, we compute $A_{G,\mathrm{BP,RP}} = k_{G,\mathrm{BP,RP}} A_0$, where $k_{G,\mathrm{BP,RP}}$ is a function of A_0 and $(G_{\mathrm{BP}} - G_{\mathrm{RP}})$. Using the extinction coefficients $(A_{G,\mathrm{BP,RP}})$, we convert the intrinsic magnitudes interpolated from the isochrones into the extincted intrinsic magnitudes. These extincted intrinsic magnitudes are combined with the distance modulus to calculate the true extincted apparent magnitudes.

Following the recommendation from the Gaia collaboration, we do not directly apply the extinction law outside the applicable color range, $-0.06 < G_{\mathrm{BP}} - G_{\mathrm{RP}} < 2.5$. However, excluding stars outside this range introduces an unnatural cut on the $G_{\mathrm{BP}} - G_{\mathrm{RP}}$ distribution. We therefore extrapolate the extinction coefficients of their Gaia passbands using the nearest extreme value of $G_{\mathrm{BP}} - G_{\mathrm{RP}}$ (i.e., −0.06 or 2.5).

The extinction law is also limited to extinction coefficients (A_0) in the range from 0.01 to 20. At the low end, since the extinction law returns finite positive values for $k_{G,\mathrm{BP,RP}}$, the resulting $A_{G,\mathrm{BP,RP}}$ always converge to 0 as A_0 goes to 0, and thus the extinction law naturally extrapolates to $A_0 = 0$. At the high end, unlike with the color, we do not adopt the extreme value (i.e., 20) for A_0 or attempt to approximate the extinction law outside the applicable A_0 range. Stars with $A_0 > 20$ are not included in the final synthetic survey for two reasons. First, A_0 is implicitly related to the distance of the star because the extinction arises from the dust between the star and the observer. If we were to adopt the extreme value for the extinction coefficient for a given star, the reported extincted photometry in the final synthetic survey would be inconsistent with the reported parallax and the underlying dust map. Second, the extincted apparent magnitudes for the majority of the stars with $A_0 > 20$ are expected to be significantly fainter than the observational limit of the synthetic survey ($G_{\mathrm{obs}} < 21$, as described in Section 3.4). We therefore do not expect the cut on $A_0 > 20$ to have a significant impact on the completeness of the final synthetic survey.

3.3. Error Modeling

We construct the photometric error model from the fit Gaia DR3 photometric uncertainties tool provided by the Gaia Data Processing and Analysis Consortium (DPAC),[13] based on data originally described in Riello et al. (2021). We adopt the astrometric measurement error models from the `PyGaia` package.[14] The spectroscopic error model is obtained from private communication with the Gaia collaboration as a function of T_{eff} and G_{RVS}. We calculate the errors and the error-convolved quantities by randomly sampling from a one-dimensional Gaussian centered on the truth values. In the final catalog, we report both the truth values and the error-convolved values.

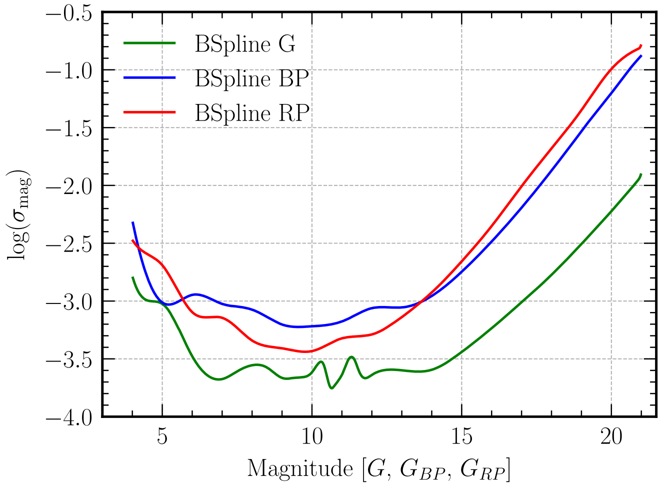

Figure 1. Logarithm of the expected uncertainties for sources with 200 (20) observations in G $(G_{\mathrm{BP}}/G_{\mathrm{RP}})$.

Table 2
Coefficients of the Astrometric errors in Equation (1)

ϖ	$\alpha_\star$	δ	$\mu_{\alpha\star}$	μ_δ
1.0	0.80	0.70	1.03	0.89

Table 3
Coefficients for Color Transformation from $G - G_{\mathrm{RP}}$ to $G_{\mathrm{RVS}} - G_{\mathrm{RP}}$

a_0	a_1	a_2	a_3	$G - G_{\mathrm{RP}}$ range
−0.0397	−0.2852	−0.0330	−0.0867	[−0.15, 1.2]
−4.0618	10.0187	−9.0532	2.6089	[1.2, 1.7]

Table 4
Coefficients for σ_{RV} as a Function of G_{RVS}

σ_{floor}	a	b	$G_{\mathrm{RVS},0}$	Applicable Range
0.12	0.9	6.0	14.0	$T_{\mathrm{eff}} < 6500$ K
0.4	0.8	20.0	12.75	$T_{\mathrm{eff}} > 7000$ K

3.3.1. Photometric Error

As mentioned, we adopt the photometric uncertainties tool from Gaia DPAC to calculate the errors in $(G, G_{\mathrm{BP}}, G_{\mathrm{RP}})$. The tool models the median behavior of the real Gaia (e)DR3 photometric uncertainties in the three Gaia passbands via cubic B-spline fitting. The errors in each photometric band are calculated as a function of the band *extincted* magnitudes. Because the B-spline is restricted to a range [4, 21] in all three bands, we extrapolate the photometric uncertainties of each band using the nearest extreme values (i.e., 4 or 21). In addition, the tool is capable of scaling the fit B-splines with different numbers of observations. We take, for simplicity, the default number of observations (i.e., 200 for G and 20 for $G_{\mathrm{BP}}/G_{\mathrm{RP}}$) for all stars in our catalogs. We show in Figure 1 the errors as a function of extincted magnitude and reproduce

---

[12] The relationship and coefficients can be downloaded from https://www.cosmos.esa.int/web/gaia/edr3-extinction-law.

[13] https://www.cosmos.esa.int/web/gaia/fitted-dr3-photometric-uncertainties-tool

[14] https://github.com/agabrown/PyGaia

Figure 14 of Riello et al. (2021). We note that this error modeling does not take into account systematic effects originating from the properties of the source, e.g., position and color.

3.3.2. Astrometric Error

`PyGaia` models the astrometric errors (i.e., parallax, position, and proper motion) as solely dependent on the apparent G magnitude. The position and proper motion errors are returned in the International Celestial Reference System (ICRS) frame, i.e., in R.A. and decl. To obtain the error-convolved positions and proper motions in Galactic coordinates (ℓ, b), we calculate the error-convolved ICRS coordinates and apply a coordinate transformation. Similarly to Gaia, we do not report the error in the Galactic coordinates. The astrometric errors, σ_X, for $X \in (\alpha_\star, \delta, \mu_{\alpha\star}, \mu_\delta)$ can be summarized as follows:

$$\sigma_X = c_X \sigma_\varpi, \quad \sigma_\varpi = \sqrt{40 + 800z + 30z^2} \tag{1}$$

$$\log_{10} z = 0.4(\max[G, G_{\rm bright}] - 15.0), \tag{2}$$

where σ_ϖ is the parallax error and $G_{\rm bright} = 13$. The coefficients c_X are reported in Table 2. Because `PyGaia` returns the error in R.A. cos(decl.) $\sigma_{\alpha\star}$, we convert $\sigma_{\alpha\star}$ to the R.A. error σ_α via analytical error propagation.

3.3.3. Spectroscopic Error

For spectroscopic measurements, Gaia DR3 provides radial velocity spectra (RVS, with magnitude $G_{\rm RVS}$), object classifications, and measured stellar parameters, such as effective temperature, surface gravity, extinction coefficient, and metallicity, in addition to radial velocities. Our synthetic survey only provides error-convolved radial velocity measurements. For DR3 radial velocities, we first use relationships provided by Sartoretti et al. (2023) to obtain true Gaia RVS magnitude, $G_{\rm RVS}$, from G and $G_{\rm RP}$. To do so, we use

$$G_{\rm RVS} - G_{\rm RP} = a_0 + a_1(G - G_{\rm RP}) + a_2(G - G_{\rm RP})^2 + a_3(G - G_{\rm RP})^3, \tag{3}$$

where the coefficients are provided in Table 3. As for the extrapolation of the extinction law, we approximate the conversion for stars outside the applicable range ($-0.15 < G - G_{\rm RP} < 1.7$), using the coefficients corresponding to the nearest extreme value of $G - G_{\rm RP}$ (i.e., −0.15 or 1.7).

The radial velocity uncertainty is fit as a function of $G_{\rm RVS}$,

$$\sigma_{\rm RV} = \sigma_{\rm floor} + b\exp(a(G_{\rm RVS} - G_{\rm RVS,0})). \tag{4}$$

The coefficients a and b are fit independently for cooler ($T_{\rm eff} < 6750$ K) and warmer ($T_{\rm eff} > 6750$ K) stars as in Table 4, which was obtained from private communication with Gaia Collaboration et al. (2022) prior to the official release of the third data release. Warmer stars generally have a larger error in radial velocities. While the error modeling for warm stars, as shown in Figure 2, appears to greatly exceed $10\,{\rm km\,s^{-1}}$ at the very faint end ($G_{\rm RVS} \sim 14$), we note that only warm stars with $G_{\rm RVS} < 12$ are selected to have a measured radial velocity in the final catalog, as described in more detail in Section 3.4. The maximum uncertainties on radial velocity measurements are thus $\sim 6\,{\rm km\,s^{-1}}$ for cool stars and $\sim 11\,{\rm km\,s^{-1}}$ for warm stars.

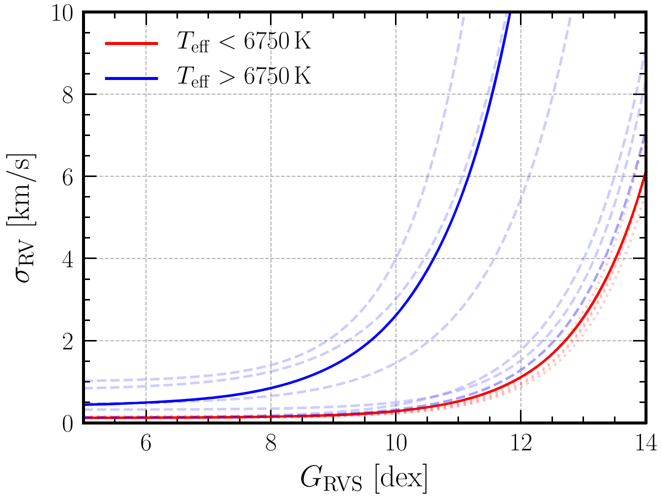

Figure 2. Comparison between the radial velocity error models adopted in this study (solid lines) and those provided by the Gaia collaboration with the official release of DR3 (dashed/dotted lines). Blue dashed lines represent error estimates for dwarfs, whereas red dotted lines represent error estimates for giants. Our estimate for cooler stars ($T_{\rm eff} < 6750$ K) is largely consistent with the estimates for giants and, similarly, our estimate for warmer stars ($T_{\rm eff} > 6750$ K) with the estimates for dwarfs.

Table 5
Coefficients for f as a Function of $G_{\rm RVS}$

a	b	c	Applicable Range
0.318	0.3884	−0.02778	$G_{\rm RVS} < 12$ mag
16.554	−2.4899	0.09933	$G_{\rm RVS} > 12$ mag

During the preparation of this manuscript, we were made aware of a more detailed Gaia DR3 radial velocity error model based on the derived stellar population.[15] Since our synthetic survey does not include the stellar evolutionary stage, we opt for the simple recipe that assigns errors based on the effective temperature of the stars. The coefficients provided are not identical to those adopted here, but a comparison between the two indicates that our adopted error modeling is roughly consistent with that from the more detailed model. For the most part, our error modeling falls on the conservative side of the latest model.

Katz et al. (2023) noted that during scientific validation of the published DR3 radial velocities, the above uncertainties were underestimated and thus require an additional multiplicative correction factor f. This multiplicative factor (f) is a function of $G_{\rm RVS}$,

$$f = a + b\,G_{\rm RVS} + c\,G_{\rm RVS}^2, \tag{5}$$

with coefficients given in Table 5. The velocity uncertainties should therefore be $f \times \sigma_{V_{\rm R}}$. We note that the relation is only valid for $G_{\rm RVS} > 8$. For $G_{\rm RVS} < 8$, we still apply the correction function but assume $G_{\rm RVS} = 8$. The correction factor is not applied directly to the uncertainties in the final Gaia DR3 data set. Following that practice, we calculate this correction factor and provide it separately in the final synthetic survey.

---

[15] https://www.cosmos.esa.int/web/gaia/science-performance

Table 6
Number of Stars in the `Ananke` DR3 Surveys of the *Latte* MW-mass Suite of FIRE Simulations

File Information			Number of Stars					
			`m12i`			`m12i` with radial velocity		
Index	d_{min} (kpc)	d_{max} (kpc)	lsr-0	lsr-1	lsr-2	lsr-0	lsr-1	lsr-2
0	0	3	316,095,707	357,883,822	392,451,308	15,646,316	18,615,906	20,116,160
1	3	4.25	290,904,524	299,505,191	302,616,084	5,468,516	5,571,554	5,555,894
2	4.25	5.5	401,479,587	420,118,792	371,841,356	7,015,408	7,442,827	6,477,804
3	5.5	6.5	400,845,878	451,022,716	362,386,254	6,624,552	7,557,280	6,134,119
4	6.5	7.25	365,130,175	436,233,406	320,076,487	6,257,625	7,157,334	5,952,305
5	7.25	8	418,818,886	526,227,703	362,084,404	8,324,411	9,089,852	8,126,577
6	8	9	507,799,164	823,576,930	435,396,583	11,999,243	14,032,870	11,840,268
7	9	10	320,749,442	382,157,501	269,564,468	8,105,168	8,450,961	7,978,302
8	10	15	510,906,338	604,188,984	455,152,086	15,677,739	15,913,215	15,528,607
9	15	300	149,436,821	137,907,134	133,845,033	11,557,428	9,579,237	10,301,140
		Total	3,682,166,522	4,438,822,179	3,405,414,063	96,676,406	103,411,036	98,011,176
		DR2 Total	3,215,565,725	3,754,501,977	2,932,162,112	38,183,839	44,583,007	39,191,496
	d_{min}	d_{max}	`m12f`			`m12f` with radial velocity		
Index	(kpc)	(kpc)	lsr-0	lsr-1	lsr-2	lsr-0	lsr-1	lsr-2
0	0	3	424,006,559	608,837,889	547,265,888	20,533,729	31,377,347	26,192,624
1	3	4.25	399,657,145	422,500,160	434,830,068	7,581,674	7,891,448	8,082,358
2	4.25	5.5	577,804,699	526,740,558	506,589,056	11,046,401	9,879,024	9,407,981
3	5.5	6.5	606,129,500	519,988,527	484,524,451	11,670,711	9,207,476	8,932,140
4	6.5	7.25	590,389,532	475,891,126	438,928,174	11,559,183	9,064,200	8,673,940
5	7.25	8	719,850,219	590,388,390	532,925,594	14,469,396	13,004,977	11,436,403
6	8	9	883,408,757	747,272,883	671,353,993	18,926,746	19,388,311	15,961,569
7	9	10	554,034,772	416,802,559	401,635,820	11,909,320	11,913,256	10,193,449
8	10	15	1,099,412,587	719,518,336	772,769,865	21,025,371	20,362,945	17,498,773
9	15	300	536, 944, 872	389,292,686	500,404,364	18,697,669	16,215,180	14,172,354
		Total	6,391,638,642	5,417,233,114	5,291,227,273	147,420,200	148,304,164	130,551,591
		DR2 Total	5,851,407,276	4,706,540,756	4,678,842,172	62,673,864	61,393,185	57,808,862
	d_{min}	d_{max}	`m12m`			`m12m` with radial velocity		
Index	(kpc)	(kpc)	lsr-0	lsr-1	lsr-2	lsr-0	lsr-1	lsr-2
0	0	3	984,809,951	1,073,978,992	910,734,608	47,393,328	54,240,600	43,119,764
1	3	4.25	728,265,777	798,150,011	686,462,276	12,171,491	13,592,305	11,429,030
2	4.25	5.5	814,806,044	863,540,944	796,780,191	12,767,460	13,727,680	12,465,994
3	5.5	6.5	685,954,361	723,050,215	689,642,062	10,229,572	10,706,710	9,751,780
4	6.5	7.25	528,436,556	558,951,415	531,816,221	7,774,165	7,855,599	7,300,150
5	7.25	8	523,399,484	551,230,847	532,527,598	7,611,444	7,589,807	7,471,248
6	8	9	2,003,093,353	639,727,826	617,194,075	17,707,039	9,881,911	10,009,474
7	9	10	422,716,282	458,827,031	432,726,049	8,073,088	8,469,156	8,438,569
8	10	15	835,507,954	1,267,343,926	1,192,679,933	23,167,408	26,128,219	26,126,666
9	15	300	261,056,409	268,075,320	244,124,886	20,074,951	20,492,819	19,063,333
		Total	7,788,046,171	7,202,876,527	6,634,687,899	166,969,946	172,684,806	155,176,008
		DR2 Total	5,701,759,381	6,415,674,623	5,516,835,110	84,931,532	108,808,464	78,520,886

3.4. Selection Function and Data Release

With error-convolved values computed, we next apply the selection function to produce the final synthetic surveys. We apply two selection functions, one for selecting stars that are detectable in all three photometric bands and another for selecting stars with reported radial velocity.

We apply a G-band magnitude cut to select stars with reported photometry in each catalog. We note that the cuts are applied on the error-convolved observed magnitudes. We select the sample of stars with reported photometry via a cut on the observed G magnitude, $3 < G_{\mathrm{obs}} < 21$. This is the same selection cut as applied in S20.

To select the sample of stars with reported radial velocities, we make a cut on effective temperature, T_{eff}, and G_{RVS}. S20 reported radial velocity measurements for bright stars with $G_{\mathrm{RVS}} < 14$ and effective temperature of $3550\ \mathrm{K} < T_{\mathrm{eff}} < 6900\ \mathrm{K}$. We extend the radial velocity selection to $3600\ \mathrm{K} < T_{\mathrm{eff}} < 14{,}500\ \mathrm{K}$ for bright stars ($G_{\mathrm{RVS}} \leqslant 12$) and $3100\ \mathrm{K} < T_{\mathrm{eff}} < 6750\ \mathrm{K}$ for fainter stars ($12 < G_{\mathrm{RVS}} < 14$), in order to match the temperature range reported in Katz et al. (2023), reflecting the improvements from Gaia DR2 to DR3.

We bin the stars in each catalog by their LSR-centric distance into 10 radial slices. Table 6 shows the total number of stars, as well as the number of stars with radial velocity measurements, in each radial slice and catalog.

4. Results

4.1. Comparison with Ananke DR2

We compare our final synthetic survey for Gaia DR3 using FIRE with the synthetic Gaia DR2 survey from S20. We updated the photometry to be consistent with Gaia DR3, using isochrones and extinction laws corresponding to the Gaia DR3 photometric system. We also updated the error modeling for photometric measurements and radial velocity measurements.

The detailed numbers of sources in each radial bin of each galaxy are given in Table 6. In general, there is a small increase in the total number of observed stars in the DR3 catalogs as compared to the DR2 catalogs. The number of stars with radial velocity measurements in each catalog has increased by ∼2–3 times, as expected from the wider range of effective temperature $T_{\rm eff}$ in the selection cut (see Section 3.4). However, this is a more moderate increase than the factor of 5 between the two Gaia data releases (from ∼7 million stars in Gaia DR2 to ∼33 million in Gaia DR3; Katz et al. 2023). This is due to the radial velocity selection cut in S20 being overly optimistic, already at $G_{\rm RVS} < 14$ when considering the actual performance of Gaia DR2 at $G_{\rm RVS} < 12.5$ (Katz et al. 2019). In all catalogs, the overall fraction of stars with radial velocity measurement compared to the total sample, which can be calculated from Table 6, is about 2%–3%, which is indeed comparable to that of Gaia DR3, which was about ∼2% (Katz et al. 2023). For reference, the fractions of stars with radial velocity measurements in the `Ananke` DR2 catalogs are about $1\% - 1.5\%$.

4.2. Synthetic Surveys

In Figures 3 and 4, we compare the distributions of radial velocity errors and proper motion errors between DR2 and DR3 for all stars in `m12i-lsr0-rslice0`.[16] Figure 3 shows the distributions of the radial velocity errors for DR2 (blue) and DR3 (solid black) for all stars with radial velocities. As expected, the radial velocity errors in `Ananke` DR3 are significantly lower than those in DR2. The radial velocity errors for DR3 are composed of two stellar populations: one with low $T_{\rm eff}$ and one with high $T_{\rm eff}$, while the radial velocity errors in DR2 are modeled by a single exponential (Sanderson et al. 2020). In `Ananke` DR3, the low-$T_{\rm eff}$ population makes up most of the distribution below $\sigma_{V_{\rm R}} \lesssim 6$ km s^{-1}, while the high-$T_{\rm eff}$ population is responsible for the tail at high $\sigma_{V_{\rm R}} \gtrsim 6$ km s^{-1}. The sharp cut at the lower end of the DR2 error is the systematic noise floor at 0.11 km s^{-1} mentioned in S20. Figure 4 shows the distributions of errors in the proper motions $\mu_{\alpha,\star}$ and μ_δ for DR2 (blue) and DR3 (solid black). Similarly, as with the radial velocity errors, the proper motion errors in `Ananke` DR3, shown in Figure 4, are typically much lower than those in DR2. The DR2 proper motion errors have a cutoff of 0.0861852 mas yr^{-1} at the low end, as described in Equation (16) and Table 5 of S20.

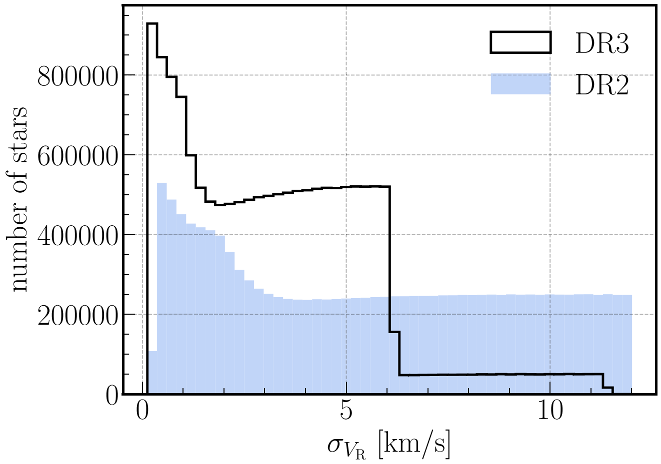

Figure 3. Distributions of the radial velocity errors for DR3 (solid black) and DR2 (blue) for all stars with radial velocities in `m12i-lsr0-rslice0`.

We examine the Hertzsprung–Russell diagram of one synthetic survey (`m12i` at lsr-0 rslice 0) as shown in Figure 5. We plot only stars with an estimated parallax error of less than 10%. Figure 5 shows that our results are qualitatively similar to what was shown in `Ananke` DR2 from S20. We see minor differences at the low-mass end of the main sequence and at the tip of the red giant branch, likely arising from the updated isochrones. Some echoes of the underlying grid of isochrones are still visible at the brightest magnitudes, where the model grid is sparsest, and potential artifacts from linear isochrone interpolation near the tip of the red giant branch are present (see Section 3). The diagrams for the other eight synthetic surveys presented in this study are shown in the Appendix.

We additionally compare our results with an actual Gaia DR3 color–magnitude diagram (CMD) from Fouesneau et al. (2023). In Figure 6, we partially reproduce Figure 1 in Fouesneau et al. (2023). Our synthetic survey generates CMDs qualitatively similar to the Gaia DR3 data. When we only consider the subsample with radial velocity measurements (`has_rvs`), the synthetic survey distributions qualitatively resemble that of the real Gaia DR3 survey.

4.3. List of Parameters in the Synthetic Surveys

In Table 7, we present the column names of the parameters used in the synthetic surveys, as well as their definitions, data types, and units. These column names are categorized by those matching Gaia DR2/DR3 (as well as `Ananke` DR2), those that are relating the properties of the simulations (for example, the true non-error-convolved values), and the properties of the stars (for example, their FIRE-2 chemical abundances).

5. Use Cases and Limitations

Synthetic surveys can be extremely powerful in testing modeling procedures, calculating false-positive rates, and validating methods. This is largely due to the fact that cosmological simulations in general track nonequilibrium dynamics self-consistently, and are therefore powerful tools for exploring dynamical inferences. The need for synthetic surveys is becoming even more critical with the large swaths of data being collected by current and upcoming surveys such as Gaia and LSST. There are, however, limitations to these studies, based on the nature of the construction of the synthetic surveys. In particular, as was highlighted in Section 2, the original simulations have a stellar mass resolution limit of ∼7000 $M_\odot$. We spawn individual stars from these simulated

---

[16] The choice of the `m12i` and lsr-0 synthetic survey is just an example that we use to illustrate different properties. Similar treatment can be done with any of the other synthetic surveys.

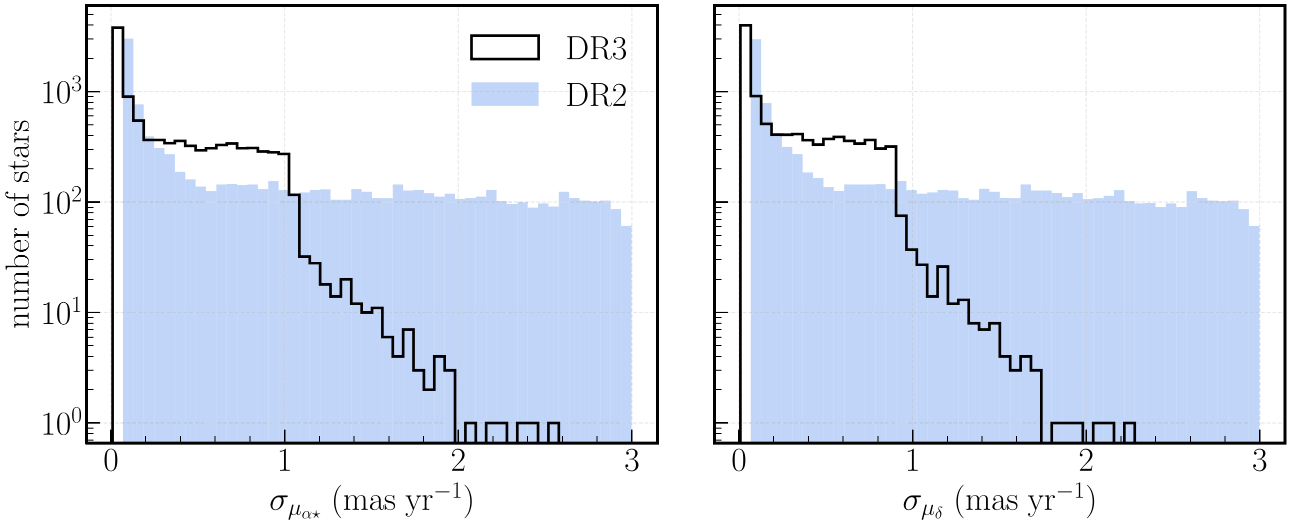

Figure 4. Distributions of the proper motion errors for R.A. cos(decl.) $\alpha^\star$ (left) and decl. δ (right) for DR3 (solid black) and DR2 (blue) for all stars in `m12i-lsr0-rslice0`.

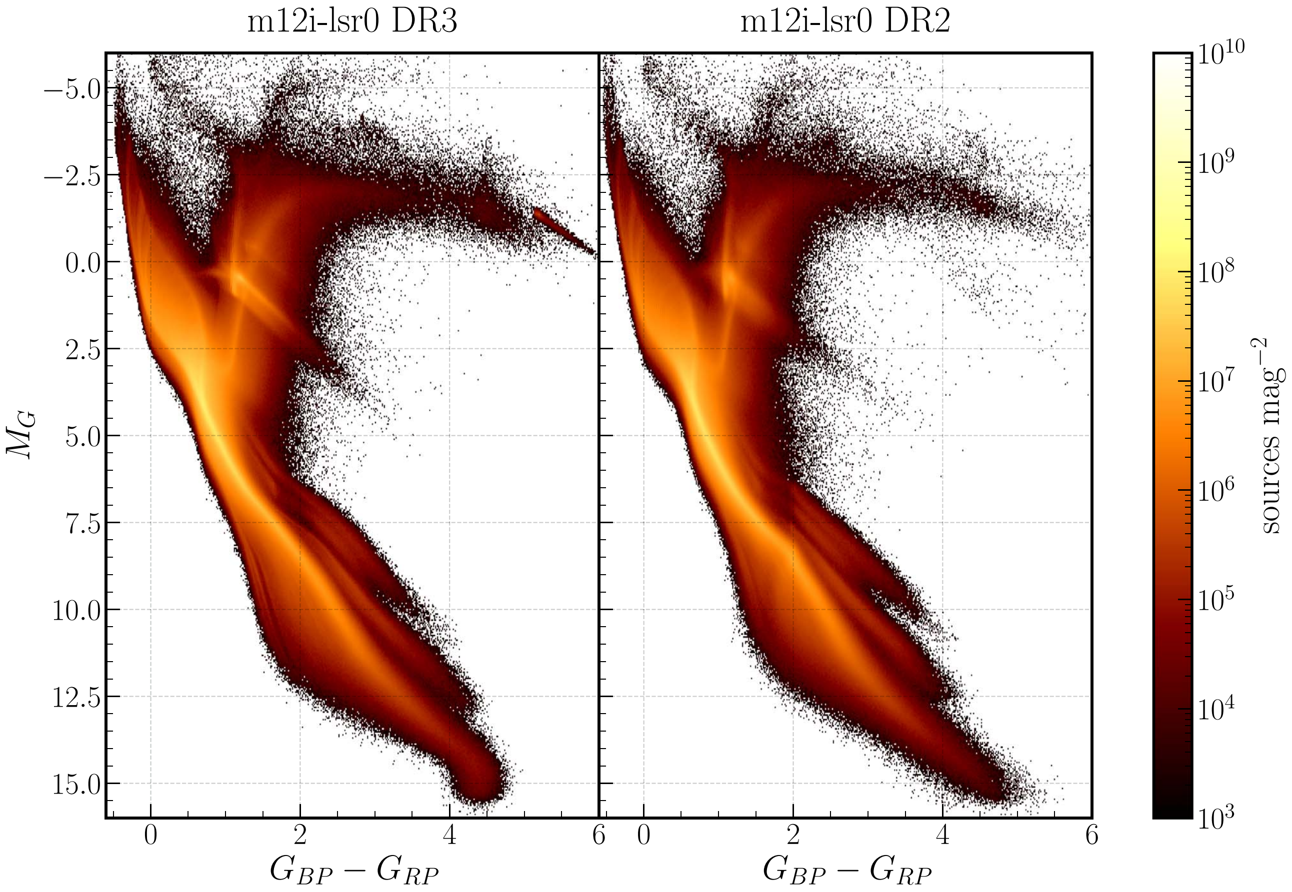

Figure 5. The Hertzsprung–Russell diagram for `m12i-lsr0-rslice0` for stars satisfying the parallax cut $\sigma_\varpi/\varpi > 10$.

star particles to generate the mock catalogs. The resulting positions and velocities of the synthetic stars depend on the kernel of choice, limiting the usage of such synthetic surveys. In particular, the internal dynamics of small-scale structures such as satellite galaxies and stellar streams are sensitive to the choice of kernel. Studies of small-scale MW structures therefore required careful kernel selection based on the science question at hand; Shipp et al. (2023), for example, changed the kernel to be able to perform detectability studies of stellar streams with LSST.

More generally, studies of large stellar structures (larger than a few star particles in the original simulation), velocity anisotropies of the Galaxy, and the MW potential should be robust to the choices of construction of the synthetic surveys, while studies of smaller structures should be treated with care, and potentially a more adequate choice of kernels.

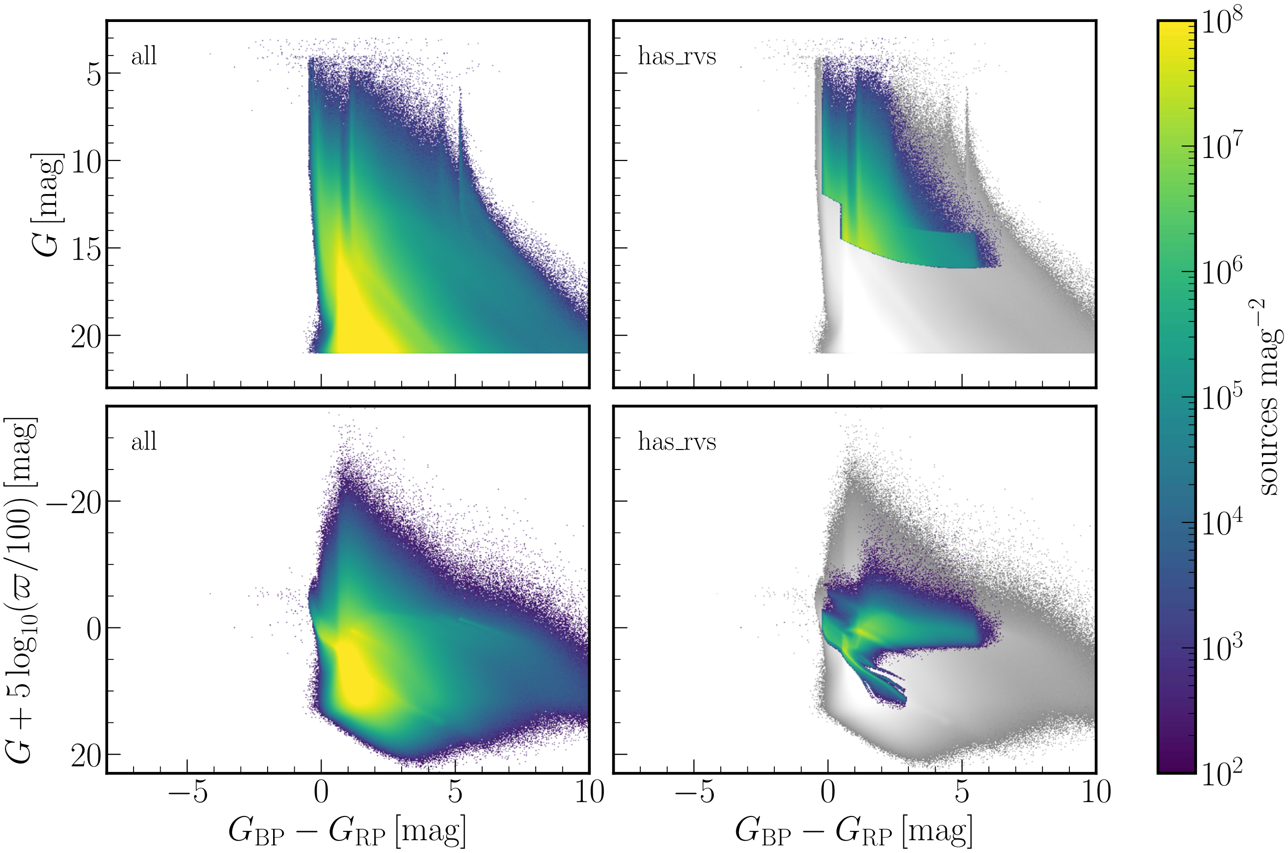

Figure 6. The color–magnitude diagram for `m12i-lsr0-rslice0` to `m12i-lsr0-rslice9` for all stars (left) and those with radial velocity measurements (right) satisfying the positive parallax cut $\varpi > 0$. The top panels shows the observed color–magnitude diagram, whereas the bottom panels shows the absolute G magnitude computed from the measured parallax.

Additionally, the original simulations from which we built these synthetic surveys are not meant to reproduce the MW itself. Indeed, these are cosmological simulations with varying initial conditions, and therefore varying histories; some contain a late merger like `m12f`, while `m12i` has a quieter merger history, for example (see, e.g., Necib et al. 2019, for details of the merger histories of these two galaxies). Therefore, it is critical to treat these galaxies as examples of galaxies with the same mass as our own, but with their own individual properties. A corollary to this is that the dust model adapted for these synthetic surveys is self-consistent with that of the simulations themselves, and therefore is different from that of the MW.

Lastly, synthetic surveys, by definition, are tailor-made to reproduce the observing properties of a particular survey—in the case of Gaia, of a particular data release. As new theories, instruments, surveys, and data become available, many key ingredients of synthetic surveys (isochrones, selection functions, error models, and so on) will become obsolete and inaccurate. To that end, Thob et al. (2023) developed `py-ananke`, a universal pipeline designed to generate synthetic surveys from cosmological simulations, adaptable to various instruments. The package builds upon the method employed in S20 and this study, with the crucial feature that allows users to change the ingredients mentioned above. This tool will greatly reduce the time and effort required to update synthetic surveys, and will ultimately provide a bridge between simulations and observations.

6. Conclusions

In this paper, we presented a new set of synthetic surveys that match Gaia DR3, based on the *Latte* suite of the FIRE-2 simulations. This is an update to the synthetic surveys released by Sanderson et al. (2020) that matched the previous data release Gaia DR2. These synthetic surveys include three different solar positions for three galaxies. The major changes compared to S20 are an updated set of isochrones matching the latest release, a different treatment of the radial velocity errors that increased the precision of the radial velocity measurements, an update to the treatment of proper motion, which decreased the measurement proper motion errors in the synthetic surveys, and an increase in the total number of stars with radial velocity measurements through the update of the selection cuts.

These synthetic surveys are made available to the community on http://ananke.hub.yt/, where they can be used to test any model/analysis pipeline on simulations prior to application to Gaia DR3. In particular, these synthetic surveys are the best tool for studies involving the dynamic properties of the MW, especially given that the "true" star particles from the original simulations are also provided. For example, recent studies on the Milky Way rotation curve (e.g., Ou et al. 2024) using Gaia data have found tentative tension on the dark matter distribution in the Milky Way. Synthetic surveys in this case may be used to quantify the effect of Gaia selection function on the accuracy of the measured rotation curve.

Table 7
Data Model for Synthetic Surveys

Quantity	Explanation	Data Type	Unit
Fields with names identical to those in DR2			
Astrometry			
`ra`	Right ascension	double	Angle (deg)
`ra_error`	Standard error of R.A.	double	Angle (deg)
`dec`	Decl.	double	Angle (deg)
`dec_error`	Standard error of decl.	double	Angle (deg)
`parallax`	Parallax	double	Angle (mas)
`parallax_error`	Standard error of parallax	double	Angle (mas)
`parallax_over_error`	Parallax divided by its error	float	⋯
`pmra`	Proper motion in R.A. direction	double	Angular velocity (mas yr^{-1})
`pmra_error`	Standard error of proper motion in R.A. direction	double	Angular velocity (mas yr^{-1})
`pmdec`	Proper motion in decl. direction	double	Angular velocity (mas yr^{-1})
`pmdec_error`	Standard error of proper motion in decl. direction	double	Angular velocity (mas yr^{-1})
`l`	Galactic longitude (converted from R.A., decl.)	double	Angle (deg)
`b`	Galactic latitude (converted from R.A., decl.)	double	Angle (deg)
Photometry			
`phot_g_mean_mag`	Extincted apparent G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag`	Extincted apparent $G_{\rm BP}$-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag`	Extincted apparent $G_{\rm RP}$-band mean magnitude	float	Magnitude (mag)
`bp_rp`	Reddened $G_{\rm BP} - G_{\rm RP}$ color	float	Magnitude (mag)
`bp_g`	Reddened $G_{\rm BP} - G$ color	float	Magnitude (mag)
`g_rp`	Reddened $G - G_{\rm RP}$ color	float	Magnitude (mag)
Spectroscopy			
`radial_velocity`	Radial velocity	double	Velocity (km s^{-1})
`radial_velocity_error`	Standard error of radial velocity	double	Velocity (km s^{-1})
Other fields not in the Gaia *DR2 data model*			
Indices			
`starid`	Array index of the star (per mock catalog)	long	⋯
`parentid`	Array index of the generating star particle in the snapshot file	long	⋯
`partid`	0 if phase-space coordinates are identical to the generating star particle, 1 otherwise	short	⋯
Phase Space			
`ra_true`	True R.A.	double	Angle (deg)
`dec_true`	True decl.	double	Angle (deg)
`dmod_true`	True distance modulus	double	Magnitude (mag)
`ra_cosdec_error`	Standard error in R.A.cos(decl.)	double	Magnitude (deg)
`parallax_true`	True parallax	double	Angle (mas)
`pmra_true`	True pm in R.A. direction	double	Angular velocity (mas yr^{-1})
`pmdec_true`	True pm in decl. direction	double	Angular velocity (mas yr^{-1})
`radial_velocity_true`	True radial velocity	double	Velocity (km s^{-1})
`l_true`	True Galactic long.	double	Angle (deg)
`b_true`	True Galactic lat.	double	Angle (deg)
`pml`	pm in Galactic long. direction	double	Angular velocity (mas yr^{-1})
`pmb`	pm in Galactic lat. direction	double	Angular velocity (mas yr^{-1})
`pml_true`	True pm in Galactic long. direction	double	Angular velocity (mas yr^{-1})
`pmb_true`	True pm in Galactic lat. direction	double	Angular velocity (mas yr^{-1})
`px_true`, `py_true`, `pz_true`	True position relative to LSR	double	Distance (kpc)
`vx_true`, `vy_true`, `vz_true`	True velocity relative to LSR	double	Velocity (km s^{-1})
Photometry			

Table 7
(Continued)

Quantity	Explanation	Data Type	Unit
	Fields with names identical to those in DR2		
`phot_g_mean_mag_true`	True (i.e., after extinction, but before error convolution) apparent G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag_true`	True apparent $G_{\rm BP}$-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag_true`	True apparent $G_{\rm RP}$-band mean magnitude	float	Magnitude (mag)
`phot_g_mean_mag_int`	Intrinsic (i.e., before extinction or error convolution) apparent G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag_int`	Intrinsic apparent $G_{\rm BP}$-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag_int`	Intrinsic apparent $G_{\rm RP}$-band mean magnitude	float	Magnitude (mag)
`phot_g_mean_mag_abs`	Absolute G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag_abs`	Absolute $G_{\rm BP}$-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag_abs`	Absolute $G_{\rm RP}$-band mean magnitude	float	Magnitude (mag)
`phot_g_mean_mag_error`	Standard error of G-band mean magnitude	float	Magnitude (mag)
`phot_bp_mean_mag_error`	Standard error of $G_{\rm BP}$-band mean magnitude	float	Magnitude (mag)
`phot_rp_mean_mag_error`	Standard error of $G_{\rm RP}$-band mean magnitude	float	Magnitude (mag)
`bp_rp_true`	True $G_{\rm BP} - G_{\rm RP}$ color	float	Magnitude (mag)
`bp_g_true`	True $G_{\rm BP} - G$ color	float	Magnitude (mag)
`g_rp_true`	True $G - G_{\rm RP}$ color	float	Magnitude (mag)
`vmini_true`	True $V - I$ color used for error modeling	float	Magnitude (mag)
Extinction			
`lognh`	$\log_{10}$ equivalent H column density along line of sight to star	float	Surface number density (cm^{-2})
`ebv`	$E(B - V)$ reddening, calculated from $N_{\rm H}^{\rm eff}$	float	Magnitude (mag)
`A0`	A_0, extinction at 550 nm, assuming $R_V = 3.1$	float	Magnitude (mag)
`a_g_val`	True line-of-sight extinction in the G band, A_G	float	Magnitude (mag)
`e_bp_min_rp_val`	True line-of-sight reddening $G_{\rm BP} - G_{\rm RP}$	float	Magnitude (mag)
Spectroscopy			
`radial_velocity_error_corr_factor`	Correction factor for `radial_velocity_error`	double	Velocity (km s^{-1})
Stellar Parameter			
`mact`	Current stellar mass	float	Mass (solar mass)
`mtip`	Mass of a star at tip of giant branch for given age, metallicity	float	Mass (solar mass)
`mini`	Stellar mass on zero-age main sequence	float	Mass (solar mass)
`age`	$\log_{10}$ of stellar age; identical for all stars generated from the same particle	float	Time (log yr)
`teff`	Stellar effective temperature	float	Temperature (K)
`logg`	Surface gravity	float	Surface gravity (log cgs)
`lum`	$\log_{10}$ of stellar luminosity	float	Luminosity (log solar luminosity)
Abundances			
`feh`	[Fe/H]	float	Magnitude (mag)
`alpha`	[Mg/Fe]	float	Magnitude (mag)
`carbon`	[C/H]	float	Magnitude (mag)
`helium`	[He/H]	float	Magnitude (mag)
`nitrogen`	[N/H]	float	Magnitude (mag)
`sulfur`	[S/H]	float	Magnitude (mag)
`oxygen`	[O/H]	float	Magnitude (mag)
`silicon`	[Si/H]	float	Magnitude (mag)
`calcium`	[Ca/H]	float	Magnitude (mag)
`magnesium`	[Mg/H]	float	Magnitude (mag)
`neon`	[Ne/H]	float	Magnitude (mag)
Quality Control			
`flag_wd`	Flag for potential white dwarfs; see Section 3.1	int	…

More generally, the adoption of synthetic surveys is applicable not only in the study of stars, but also in the dynamics of dark matter and properties of gas particles. Wetzel et al. (2023) have made the simulations used in this work publicly available,[17] including the formation coordinates of all star particles, as well as catalogs of all satellite galaxies/halos.

---

[17] http://flathub.flatironinstitute.org/fire

Therefore, the community can use such information to answer more general questions about what the field can learn through Gaia about the Galaxy as a whole, from its stellar components as measured through the Gaia lens, to the inner workings of the dark matter in the Galaxy that governs the dynamics of the stars.

Acknowledgments

We thank Kacper Kowalik and Matthew Turk for their tremendous help on data storage. We also thank Sarah Loebman for her work on the original `Ananke` paper.

T.N. and L.N. are supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, http://iaifi.org/).

N.S. is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-2303841.

A.W. received support from: NSF via CAREER award AST-2045928 and grant AST-2107772; NASA ATP grant 80NSSC20K0513; HST grants AR-15809, GO-15902, GO-16273 from STScI.

N.P. and R.E.S. acknowledge support from NASA grant 19-ATP19-0068. N.P. was supported in part by a Zacheus Daniel fellowship from the University of Pennsylvania. R.E.S. additionally acknowledges support from NSF grant AST-2007232, from the Research Corporation through the Scialog Fellows program on Time Domain Astronomy, and from HST-AR-15809 from the Space Telescope Science Institute (STScI), which is operated by AURA, Inc., under NASA contract NAS5-26555.

This work made use of Stampede-2, a large computing cluster funded by the National Science Foundation (NSF) through award ACI-1540931. The analysis was conducted using the allocation number PHY210118. We generated simulations using Pleiades, via the NASA HEC program through the NAS Division at Ames Research Center. The computations in this work were, in part, run at facilities supported by the Scientific Computing Core at the Flatiron Institute, a division of the Simons Foundation. The data used in this work were, in part, hosted on equipment supported by the Scientific Computing Core at the Flatiron Institute, a division of the Simons Foundation.

This research or product makes use of public auxiliary data provided by ESA/Gaia/DPAC/CU5 and prepared by Carine Babusiaux. 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.

Appendix

Comparison of Hertzsprung–Russell Diagrams for All Synthetic Surveys

In this Appendix, we present the Hertzsprung–Russell diagrams for the rest of synthetic surveys, including `lsr1` and `lsr2` of `m12i` in Figure 7, and `m12f` and `m12m` in Figure 8.

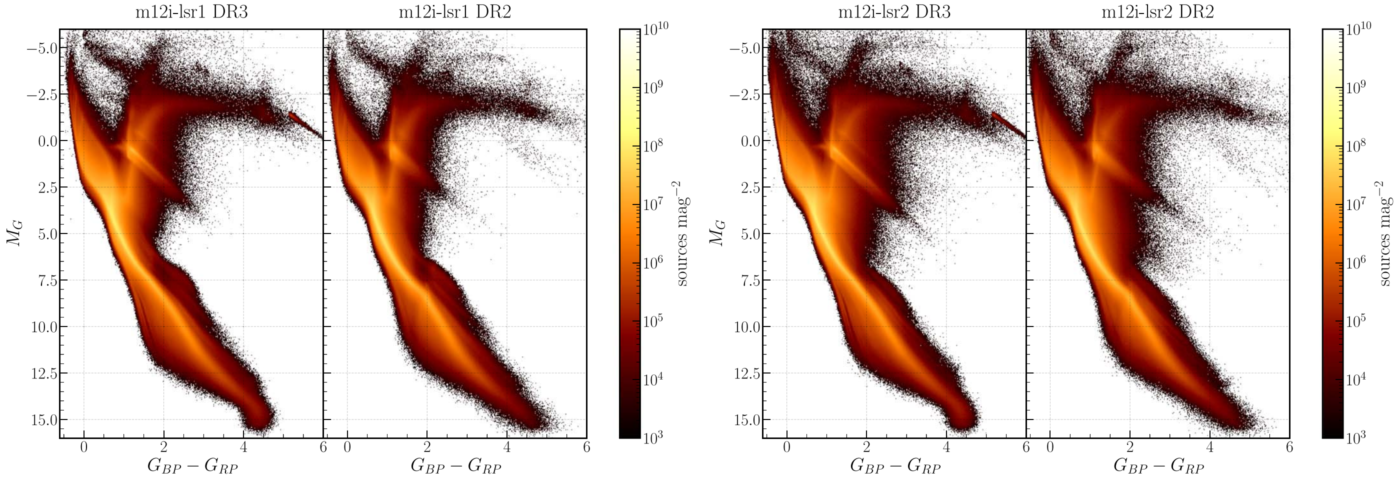

Figure 7. Same as Figure 5 but for `rslice-0` of `m12i-lsr1` and `m12i-lsr2`.

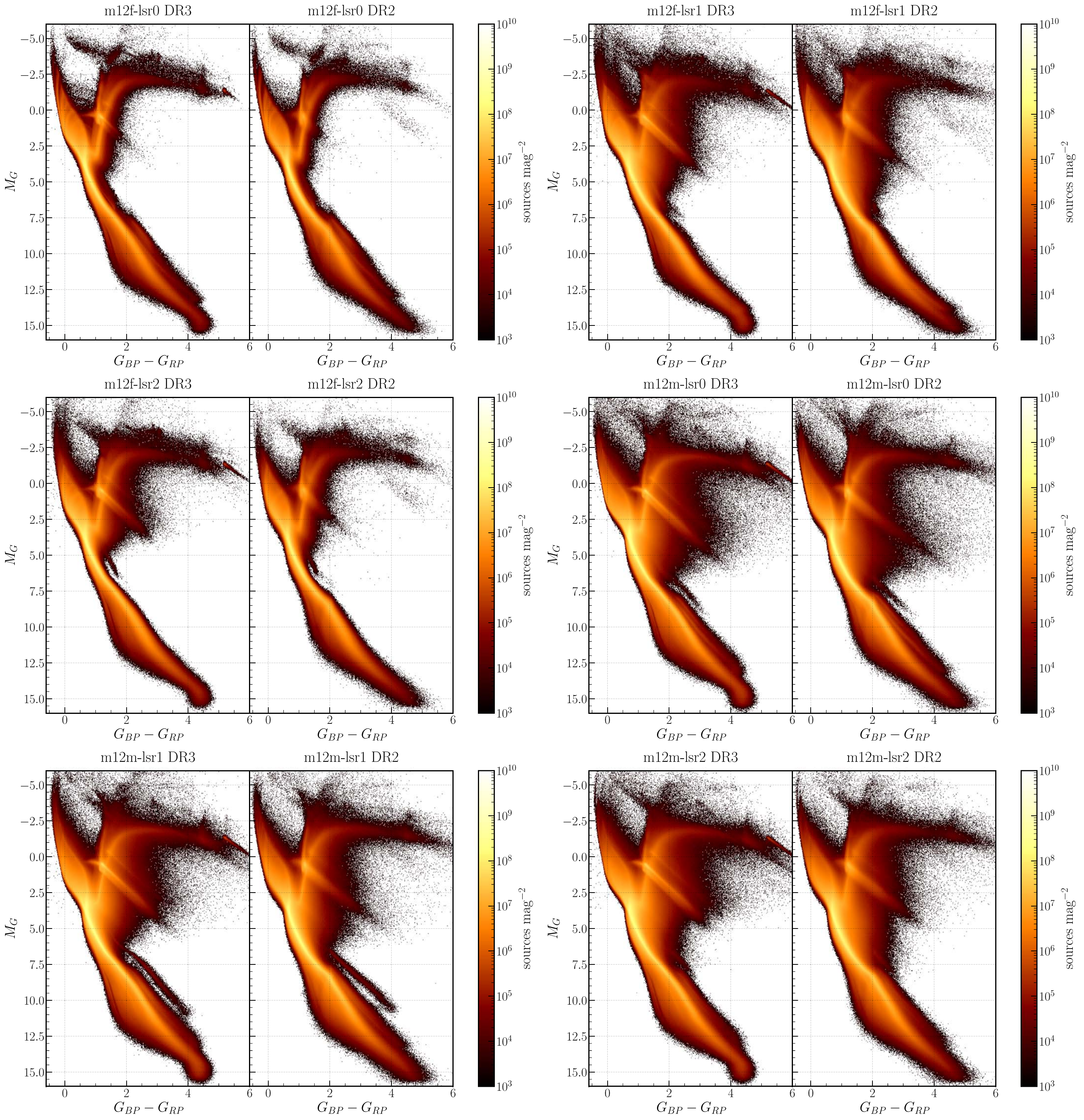

Figure 8. Same as Figure 5 but for RSLICE-0 of `m12f` and `m12m` for all three LSRs.

ORCID iDs

Tri Nguyen https://orcid.org/0000-0001-6189-8457
Xiaowei Ou https://orcid.org/0000-0002-4669-9967
Nondh Panithanpaisal https://orcid.org/0000-0001-5214-8822
Nora Shipp https://orcid.org/0000-0003-2497-091X
Lina Necib https://orcid.org/0000-0003-2806-1414
Robyn Sanderson https://orcid.org/0000-0003-3939-3297
Andrew Wetzel https://orcid.org/0000-0003-0603-8942

References

Ansar, S., Pearson, S., Sanderson, R. E., et al. 2023, arXiv:2309.16811
Antoja, T., Helmi, A., Romero-Gómez, M., et al. 2018, Natur, 561, 360
Beaton, R. L., Werner, S., Mitschang, A. W., et al. 2022, RNAAS, 6, 125
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
Bland-Hawthorn, J., & Gerhard, O. 2016, ARA&A, 54, 529
Bohlin, R. C., Hubeny, I., & Rauch, T. 2020, AJ, 160, 21
Buckley, M. R., Lim, S. H., Putney, E., & Shih, D. 2023, MNRAS, 521, 5100

Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, A&A, 618, A93
Chen, Y., Girardi, L., Fu, X., et al. 2019, A&A, 632, A105
Debattista, V. P., Gonzalez, O. A., Sanderson, R. E., et al. 2019, MNRAS, 485, 5073
DES Collaboration 2005, arXiv:astro-ph/0510346
DES Collaboration 2016, MNRAS, 460, 1270
Eilers, A.-C., Hogg, D. W., Rix, H.-W., & Ness, M. K. 2019, ApJ, 871, 120
El-Badry, K., Quataert, E., Wetzel, A., et al. 2018, MNRAS, 473, 1930
Epanechnikov, V. A. 1969, Theory of Probability & Its Applications, 14, 153
Fouesneau, M., Frémat, Y., Andrae, R., et al. 2023, A&A, 674, A28
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gaia Collaboration, Brown, A. G. A., & Vallenari, A. 2021, A&A, 649, A1
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, A&A, 674, A1
Gentile Fusillo, N. P., Tremblay, P. E., Cukanovaite, E., et al. 2021, MNRAS, 508, 3877
Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, ApJ, 887, 93
Helmi, A. 2020, ARA&A, 58, 205
Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Natur, 563, 85
Hopkins, P. F. 2015, MNRAS, 450, 53
Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, MNRAS, 480, 800
Johnson, H. L. 1965, ApJ, 141, 923
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Katz, D., Sartoretti, P., Cropper, M., et al. 2019, A&A, 622, A205
Katz, D., Sartoretti, P., Guerrier, A., et al. 2023, A&A, 674, A5
Kroupa, P. 2001, MNRAS, 322, 231
Kruijssen, J. M. D., Pfeffer, J. L., Reina-Campos, M., Crain, R. A., & Bastian, N. 2019, MNRAS, 486, 3180
Ma, X., Hopkins, P. F., Wetzel, A. R., et al. 2017, MNRAS, 467, 2430
Necib, L., Lisanti, M., Garrison-Kimmel, S., et al. 2019, ApJ, 883, 27
Necib, L., Ostdiek, B., Lisanti, M., et al. 2020, NatAs, 4, 1078
Ostdiek, B., Necib, L., Cohen, T., et al. 2020, A&A, 636, A75
Ou, X., Eilers, A.-C., Necib, L., & Frebel, A. 2024, MNRAS, 528, 693
Riello, M., De Angeli, F., Evans, D. W., et al. 2021, A&A, 649, A3
Robin, A. C., Luri, X., Reylé, C., et al. 2012, A&A, 543, A100
Sanderson, R. E., Garrison-Kimmel, S., Wetzel, A., et al. 2018, ApJ, 869, 12
Sanderson, R. E., Wetzel, A., Loebman, S., et al. 2020, ApJS, 246, 6
Sartoretti, P., Marchal, O., Babusiaux, C., et al. 2023, A&A, 674, A6
Schultz, G. V., & Wiemer, W. 1975, A&A, 43, 133
Sharma, S., Bland-Hawthorn, J., Johnston, K. V., & Binney, J. 2011, ApJ, 730, 3
Sharma, S., & Johnston, K. V. 2009, ApJ, 703, 1061
Sharma, S., & Steinmetz, M. 2006, MNRAS, 373, 1293
Shipp, N., Panithanpaisal, N., Necib, L., et al. 2023, ApJ, 949, 44
Thob, A. C. R., Sanderson, R. E., Eden, A. P., et al. 2023, arXiv:2312.02268
Wetzel, A., Hayward, C. C., Sanderson, R. E., et al. 2023, ApJS, 265, 44
Wetzel, A. R., Hopkins, P. F., Kim, J.-H., et al. 2016, ApJL, 827, L23
Whittet, D. C. B., & van Breda, I. G. 1980, MNRAS, 192, 467