



Spatial and Kinematic Clustering of Stars in the Galactic Disk

Harshil Kamdar¹, Charlie Conroy¹, Yuan-Sen Ting (丁源森)^{2,3,4,5}, and Kareem El-Badry⁶¹ Center for Astrophysics | Harvard & Smithsonian, 60 Garden Street, Cambridge, MA 02138, USA; harshil.kamdar@cfa.harvard.edu² Institute for Advanced Study, Princeton, NJ 08540, USA³ Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA⁴ Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, CA 91101, USA⁵ Research School of Astronomy and Astrophysics, Mount Stromlo Observatory, Cotter Road, Weston Creek, ACT 2611, Canberra, Australia⁶ Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720, USA

Received 2020 August 4; revised 2021 May 4; accepted 2021 May 4; published 2021 November 18

Abstract

The Galactic disk is expected to be spatially and kinematically clustered on many scales due to both star formation and the Galactic potential. In this work we calculate the spatial and kinematic two-point correlation functions (TPCF) using a sample of 1.7×10^6 stars with radial velocities from Gaia DR2. Clustering is detected on spatial scales of 1–300 pc and a velocity scale of 15 km s^{-1} . After removing bound structures, the data have a power-law index of $\gamma \approx -1$ for $1 \text{ pc} < \Delta r < 100 \text{ pc}$ and $\gamma \lesssim -1.5$ for $\Delta r > 100 \text{ pc}$. We interpret these results with the aid of a star-by-star simulation of the Galaxy, in which stars are born in clusters orbiting in a realistic potential that includes spiral arms, a bar, and giant molecular clouds. We find that the simulation largely agrees with the observations at most spatial and kinematic scales. In detail, the TPCF in the simulation is shallower than the data at $\lesssim 20 \text{ pc}$ scales, and steeper than the data at $\gtrsim 30 \text{ pc}$. We also find a persistent clustering signal in the kinematic TPCF for the data at large Δv ($> 5 \text{ km s}^{-1}$) that is not present in the simulations. We speculate that this mismatch between observations and simulations may be due to two processes: hierarchical star formation and transient spiral arms. We also predict that the addition of ages and metallicities measured with a precision of 50% and 0.05 dex, respectively, will enhance the clustering signal beyond current measurements.

Unified Astronomy Thesaurus concepts: Star clusters (1567); Star formation (1569); Milky Way Galaxy (1054); Two-point correlation function (1951)

1. Introduction

Much of our knowledge about how galaxies form and evolve in the universe comes from detailed studies of our own Galaxy. The rich history of the Galaxy is encoded in the distribution of the kinematics and the chemistry of its stars. The unprecedented amount of astrometric (Brown et al. 2018) and spectroscopic (e.g., Kollmeier et al. 2017; Kunder et al. 2017; Ahumada et al. 2020; Buder et al. 2019) data on our Galaxy expected in the coming years will revolutionize our view of the different physical processes in galaxy evolution. To reconstruct the history of the Galaxy we must study the birth, evolution, and death of the building blocks of star formation—star clusters.

Most stars are thought to be born in a spatially and temporally correlated way (see reviews by Bland-Hawthorn et al. 2010; Krumholz et al. 2019); however, most stellar aggregates are quickly disrupted in the Galaxy (Lada & Lada 2003; Gieles et al. 2006). Consequently, much effort has been devoted to the fossil record of star formation by inspecting the chemical makeup of stars to identify those that might have been born in the same birth cloud (i.e., “co-natal” stars; Freeman & Bland-Hawthorn 2002; Bland-Hawthorn et al. 2010; Ting et al. 2015a; Price-Jones et al. 2020). Recent work (e.g., Meingast et al. 2019; Kamdar et al. 2019b; Coronado et al. 2020) has also shown promise in using the kinematic properties of stars to find those that might have been born in the same cluster but have since drifted apart.

Non-axisymmetric features in the Galactic disk⁷ can also have a significant impact on the structure in the disk (e.g.,

Hunt et al. 2018; Sellwood et al. 2019; Trick et al. 2019). Structure on large scales—due to resonances in the kinematics and the enrichment history of the Galaxy in chemistry—has been extensively studied in recent years in both the data (e.g., Kawata et al. 2018; Michtchenko et al. 2018; Bland-Hawthorn et al. 2019; Trick et al. 2019) and simulations (e.g., Fragkoudi et al. 2019; Monari et al. 2019). However, there remain key questions about clustering on intermediate scales, where both correlated star formation and resonances are likely important.

In Kamdar et al. (2019a, hereafter K19a), we presented a star-by-star dynamical model of the Galactic disk that takes into account the clustered nature of star formation and the complexity of the Galactic potential. The highest-resolution cosmological zoom-in simulations and isolated N -body simulations of the Galaxy have stellar particles with masses $\gtrsim 500 M_\odot$, which is the typical mass of a star cluster born today. These simulations are therefore unable to probe the scales relevant for studying the small-scale clustering of individual stars. Our model self-consistently evolves 4 billion stars over the last 5 Gyr in a realistic time-varying potential that includes an axisymmetric component, a bar, spiral arms, and giant molecular clouds (GMCs). All stars are born in clusters with a subgrid model for cluster birth and dissolution (Lada & Lada 2003). As direct N -body calculations for billions of stars are computationally infeasible, we developed a method of initializing star clusters to mimic the effects of direct N -body interactions.

A key, unexpected prediction from K19a was that stars separated spatially by as much as 20–30 pc but moving at similar relative velocities are likely co-natal. We used Gaia DR2 and LAMOST DR4 data in Kamdar et al. (2019b, hereafter K19b) to identify and study these “comoving” pairs of

⁷ For the remainder of this paper, we use “non-axisymmetries” to mean the bar, spiral arms, and GMCs.

stars (e.g., Oh et al. 2017) with both kinematic information and chemical abundances. In K19b, we identified 111 such comoving pairs in the solar neighborhood with reliable astrometric and spectroscopic measurements. These pairs showed a strong preference for having similar metallicities when compared to random field pairs, supporting the idea that they were born together.

The comoving pairs identified in K19b along with wide binaries from other work (e.g., Andrews et al. 2017; El-Badry & Rix 2018; Hawkins et al. 2020) probe (by design) fairly small spatial scales (1–20 pc). Other work studying resonances (e.g., Bovy et al. 2015; Khanna et al. 2019) probes structure at kiloparsec scales. Here we expand upon previous work by measuring the two-point correlation function for the solar neighborhood on physical scales from parsecs to kiloparsecs in order to study the clustered nature of the Galaxy.

The two-point correlation function (TPCF) has been extensively used in cosmology (see references in Peebles 2001) and other areas of physics (e.g., Kagan & Knopoff 1980; Zmolodchikov 1991). The TPCF characterizes the excess probability of two points separated by some r relative to an unclustered distribution. Given its simplicity and relative ease of interpretability, the TPCF has been used to constrain cosmological models (e.g., Eisenstein et al. 2005; Sanchez et al. 2012; Alam et al. 2017), study the galaxy–halo connection (e.g., Conroy et al. 2006; Wechsler et al. 2006; Reddick et al. 2013), probe the epoch of reionization (e.g., McQuinn et al. 2007), and quantify the clustering of young stellar clusters in other galaxies (e.g., Houllahan & Scalo 1990; Elmegreen et al. 2014; Gouliermis et al. 2017; Grasha et al. 2017).

There has been some previous work characterizing the clustering in the Galactic halo (Cooper et al. 2011; Lancaster et al. 2019) to study kinematic substructures and infer the Galactic accretion history. Bovy et al. (2015) and Khanna et al. (2019) calculated the power spectrum (the Fourier transform of the TPCF) of velocity fluctuations in the disk to study the dynamical influence of the bar in the disk. Mao et al. (2015) used the TPCF to probe Galactic disk structure in SEGUE G-dwarf stars. Mao et al. placed strong constraints on the scale heights of the thin disk and the thick disk. However, they also show the strong biases that the selection function and the nonuniform density profile of the Galaxy impart on the TPCF. Consequently, calculating the TPCF in Galactic science has been nontrivial up to now due to the relative dearth of data, the complex density profile of the Galaxy, and the absence of theory or simulations to guide predictions at all scales. The landscape has changed dramatically with the release of Gaia DR2, which provided 6D phase space information for millions of stars. Moreover, the star-by-star simulations presented in K19a enable, for the first time, predictions of stellar clustering on both small and large spatial scales.

In this paper we present the spatial and velocity TPCF for stars in the solar neighborhood using Gaia data and provide predictions from the simulations presented in K19a. The rest of this paper is organized as follows. Section 2 discusses the quality cuts imposed on Gaia data and the simulations in K19a; we also present the mock catalog from a smooth unclustered realization of the Galaxy (Rybizki et al. 2018) as a control. In Section 3 we introduce the TPCF, describe our method for the random catalog, and present several validation tests. The spatial and kinematic TPCFs for the data and the simulations are presented in Section 4.1, and we predict the clustering in the disk when we combine kinematic data with metallicity and age

information in Section 4.2. A summary of our results is provided in Section 5.

2. Data and Simulations

2.1. Observational Data

We focus on stars with radial velocities in Gaia DR2. We start with the 6D Gaia DR2 (Brown et al. 2018) catalog from Marchetti et al. (2019). Stars were selected within a cylinder centered at $(X, Y, Z) = (-8.2, 0.0, 0.025)$ kpc (Bland-Hawthorn & Gerhard 2016) with a radius of 0.5 kpc and a height of 1 kpc (0.5 kpc above and below the solar position). The impact of the volume of the cylinder on the TPCF is discussed in Appendix A, where we show that the chosen volume has minimal impact on our overall results. We choose this volume to ensure that the data are of high quality, and because going out to a larger volume would require a careful treatment of the fluctuations in the local mid-plane of the Galaxy (Beane et al. 2019).

The distance to sources in the catalog with low relative error ($0 < \sigma_{\varpi}/\varpi < 0.1$) in parallax is calculated by simply inverting the parallax. For stars with a larger parallax uncertainty, the distances are calculated using the Bayesian approach outlined in Bailer-Jones et al. (2018). We transform the parallax, proper motions, and radial velocities into a Galactocentric Cartesian coordinate frame. The rotation velocity at the Sun’s position is assumed to be $v_{\text{LSR}} = 238 \text{ km s}^{-1}$, and the Sun’s orbital velocity is assumed to be $(U_{\odot}, V_{\odot}, W_{\odot}) = (14.0, 12.24, 7.25) \text{ km s}^{-1}$ (Schönrich 2012; Bland-Hawthorn & Gerhard 2016). Moreover, we follow K19b and the recommendations in Boubert et al. (2019), and impose the following quality criteria on the Gaia data considered in this analysis: (1) number of visibility periods ≥ 6 , (2) number of radial velocity (RV) transits ≥ 4 , and (3) renormalized unit weight error (RUWE) ≤ 1.6 .⁸

These quality cuts and the geometric selection described earlier result in a catalog of $\sim 1.7 \times 10^6$ stars. The median uncertainties in parallax and the proper motions ($\sigma_{\mu_{\alpha^*}}, \sigma_{\mu_{\delta}}, \sigma_w$) are 0.037 mas, 0.06 mas yr^{−1}, and 0.05 mas yr^{−1}, respectively. The velocity error budget is dominated by the RV measurements; the median RV uncertainty in our selected subsample is 1.15 km s^{−1}.

2.2. Simulations

We use four simulations to interpret the Gaia results, three from K19a and one from Rybizki et al. (2018, hereafter R18).

2.2.1. Kamdar et al. (2019a)

K19a presented three simulations that we will utilize in this work. These simulations are summarized below; we refer the reader to K19a for a more comprehensive overview of the different model ingredients.

1. *A fiducial simulation with both clustered star formation and a realistic gravitational potential.* The fiducial simulation self-consistently evolves 4 billion stars over the last 5 Gyr in a realistic time-varying potential that includes an axisymmetric component, a bar, non-transient

⁸ The traditional square root of the reduced chi-square (unit weight error) has a strong dependence on color and magnitude. These dependences are removed using a renormalization process in Lindegren (2018). The RUWE provides a more robust indicator of the goodness-of-fit for the astrometry.

spiral arms rotating at a constant pattern speed, and actively orbiting GMCs. All stars are initialized in clusters with an observationally motivated range of initial conditions. For stars older than 5 Gyr, we include a smooth, phase-mixed background population of stars. We developed a method of initializing star clusters to mimic the effects of direct N -body interactions, while the actual orbit integrations are treated as test particles within the analytic potential.

2. *A simulation with only small-scale perturbations.* The setup of this simulation is almost identical to the fiducial simulation with one key difference: the potential is axisymmetric with no bar and spiral arms, and no GMCs as perturbers.
3. *A simulation with non-axisymmetric perturbations (with bar and spiral arms) but with no clustered star formation (NCSF simulation hereafter).* Instead of forming stars within clusters, we form them as above but in $N=1$ systems. Since there are no clusters in this simulation, there is little small-scale clustering.

To enable a fair comparison to the Gaia data we create mock catalogs of our simulation in Gaia DR2-like solar cylinders. We perform the same geometry and magnitude selections on the mock catalog from our simulations as the Gaia catalog. We use the MIST stellar evolutionary tracks (Choi et al. 2016) and the C3K stellar library (C. Conroy et al., unpublished) to derive photometry for the simulated stars using a Kroupa initial mass function (Kroupa 2001). We also calculate G_{RVS} of the Gaia radial velocity spectrometer (RVS) using the relations (Equations (2) and (3)) presented in Brown et al. (2018) and apply the same G_{RVS} selection ($G_{\text{RVS}} < 12$).

An accurate error model is essential for comparisons between simulations and observations. The dependence of parallax, proper motion, and radial velocity errors is a complex function of several parameters. We fit a Gaussian mixture model (GMM) with 20 components to the combined $(G, G_{\text{BP}} - G_{\text{RP}}, \sigma_{\omega}, \sigma_{\mu_{\alpha^*}}, \sigma_{\mu_{\delta}})$ and $(G, G_{\text{BP}} - G_{\text{RP}}, \sigma_{\text{RV}})$ spaces, respectively, where $\sigma_{\omega}, \sigma_{\mu_{\alpha^*}}, \sigma_{\mu_{\delta}}, \sigma_{\text{RV}}$ are the uncertainties in the parallax, proper motions, and radial velocities. We build separate mixture models for the astrometric errors and the spectroscopic errors.

To draw realistic error estimates given G and $G_{\text{BP}} - G_{\text{RP}}$ in our simulation, we sample from the conditional distributions for the respective errors given G and $G_{\text{BP}} - G_{\text{RP}}$. However, the errors will also depend on the scanning law, which is not explicitly modeled for this work. The scanning law also has a nontrivial impact on the selection of stars in the solar neighborhood. With recent progress on modeling the scanning law and computing the true RVS selection function (Boubert & Everall 2020), we plan to include a detailed selection function in future work (as opposed to a simple magnitude cut) for both the error model and the selection of stars.

Since the simulations in K19a and R18 do not include binarity, it is important to avoid contamination from bound and disrupting wide binaries in the data. The Jacobi radius, beyond which the Galactic tidal field is stronger than the mutual gravitational attraction of binaries, is ≈ 1.7 pc for $\sim 1 M_{\odot}$ stars in the solar neighborhood (Yoo et al. 2004; Jiang & Tremaine 2010). We employ a simple condition on the projected separation between pairs of stars ($s < 0.5$ pc) to ensure minimal contamination from wide binaries. We discuss our motivations for this selection in Appendix B.1. Moreover,

simulations in Jiang & Tremaine (2010) also predict that there could be a noticeable signature of unbound wide binaries in the phase space density up to $10\text{--}100R_{\text{J}}$ ($\sim 20\text{--}200$ pc), where R_{J} is the Jacobi radius. A discussion on unbound wide binaries is also included in Appendix B.1.

2.2.2. Rybizki et al. (2018)

The smooth, unclustered Gaia DR2 mock catalog presented in R18 is essential to validate the techniques presented in this paper. The mock catalog in R18 was generated using Galaxia (Sharma et al. 2011), sampling stars according to the (spatially and kinematically smooth) Besançon Galactic model (Robin et al. 2003). Moreover, R18 also includes a realistic treatment of 3D dust extinction (Bovy et al. 2016, and references therein), and Gaia DR2-like errors in the astrometry, photometry, and spectroscopy of stars. The R18 mock uses PARSEC isochrones (Marigo et al. 2017) to generate the photometry. Lastly, R18 also includes a model for the Galactic warp (Sharma et al. 2011).

To enable a fair comparison to the Gaia data we create a DR2-like solar cylinder from the R18 mock data. We perform the same geometry and magnitude selections on the mock catalog from R18 as in the Gaia catalog. The R18 mock has no spatial or kinematic clustering on any scales by construction; consequently, we will use the R18 mock as a control to test our technique for measuring the TPCF, and validate our technique to generate random catalogs.

3. Two-point Correlation Function

3.1. Theory and Motivation

The TPCF is a powerful measure of the clustering in data (see Peebles 2001, and references therein). The TPCF measures the excess probability of finding one object within a specified distance of another object against that of a random, unclustered distribution. In the sections that follow, we will calculate the TPCF using a variety of different selection cuts on Δv , Δr , $\Delta[\text{Fe}/\text{H}]$, or Δage . We use the highly optimized, OpenMP-parallelized, publicly available code, CorrFunc⁹ (Sinha & Garrison 2018, 2020), to calculate the TPCFs,

$$dP_{1,2} = n^2 dV_1 dV_2 [1 + \xi(\Delta r_{1,2})]. \quad (1)$$

Here, n is the mean density, and $\xi(\Delta r_{1,2})$ is the excess probability, relative to an unclustered distribution, that two points (1 and 2) are separated by $\Delta r_{1,2}$ (Δr hereafter¹⁰). The two-point correlation is usually estimated using the Landy–Szalay estimator (Landy & Szalay 1993):

$$\xi(\Delta r) = \frac{\text{DD}(\Delta r) - 2\text{DR}(\Delta r) + \text{RR}(\Delta r)}{\text{RR}(\Delta r)}, \quad (2)$$

where DD is the count of data–data pairs, DR is the count of data–random pairs, and RR is the count of random–random pairs. In studies of large-scale structure, a simple uniform distribution is adopted for the random catalog. However, the nonuniform density distribution of the Galactic disk makes the

⁹ <https://github.com/manodeep/CorrFunc>

¹⁰ In other fields, this is sometimes denoted as r but we use Δr to avoid confusion with the overall geometry of the system.

creation of the random catalog highly nontrivial (e.g., Mao et al. 2015).

3.2. Random Catalogs

Random catalogs are an essential ingredient for computing the TPCF. Lancaster et al. (2019) utilized the TPCF to quantify the smoothness of the Galactic halo; the random catalog was created by fitting a parametric form of the density profile to the halo. Mao et al. (2015, M15 hereafter) used the TPCF to probe Galactic disk structure in SEGUE G-dwarf stars. The results presented in M15 show that both the underlying density gradient and the survey geometry can significantly bias the TPCF. M15 chose analytical thin-disk/thick-disk density models, assuming that $\xi(\Delta r)$ should approach 0 at sufficiently large scales if the correct density model is chosen. The key result from M15 was to provide constraints on the scale lengths and the scale heights of both the thin disk and the thick disk.

The combination of a nonuniform density profile and a complex selection function for the data necessitates the use of a very flexible density estimation technique. For this work, we attempted Gaussian mixture models, Dirichlet process Gaussian mixture models (also known as infinite mixture models; Rasmussen 2000), and normalizing flows (e.g., Rezende & Mohamed 2015). In an attempt to balance accuracy and interpretability, we chose to use Dirichlet process Gaussian mixture models (DPGMM hereafter).

Gaussian mixture models have been extensively used in astronomy. GMMs assume that the input data are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. Traditionally, GMMs are trained using the expectation-maximization algorithm and require a choice of hyperparameters. The GMM can be written as

$$p(x|\theta_1, \dots, \theta_K) = \sum_{k=1}^K \pi_k \mathcal{N}(x|\mu_k, \Sigma_k) \quad (3)$$

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right) \quad (4)$$

where $\theta_k = \{\mu_k, \Sigma_k, \pi_k\}$ is the set of parameters for component k , and $\mathcal{N}(x|\mu, \Sigma)$ is the multivariate Gaussian.

The DPGMM is a nonparametric Bayesian extension of GMMs, where each parameter in the model is assigned a prior. The DPGMM is formally written as

$$(\mu_k, \Sigma_k) \sim \mathcal{N}\mathcal{IW}(\mu_0, \lambda_0, S_0, \nu_0) \quad (5)$$

$$\pi_k \sim \mathcal{DP}(\alpha). \quad (6)$$

Here, the means and the covariances of each Gaussian component (μ_k, Σ_k) have the normal inverse Wishart (NIW) distribution as their prior. μ_0, λ_0, S_0 , and ν_0 represent the mean, scale, scale matrix, and degrees of freedom. The weights for each Gaussian component follow a Dirichlet process prior, parameterized by the concentration α . The key advantages that DPGMMs offer over GMMs are twofold: (1) the number of components actively used in the model is automatically inferred using variational inference, and (2) the priors help regularize the model. We use the implementation of DPGMMs presented in scikit-learn (Pedregosa et al. 2011). The model is fit to the entirety of Gaia RVS data because the rigid boundaries of a cylindrical selection created some pathological behavior in the fitting process. The model is fit on half the data

and compared to the other half for validation. We first sample stars from the fit DPGMM, and then make the spatial selection described in Section 2.

The fidelity of the random catalogs generated using the DPGMM method described above is shown in Figure 1. The top left panel shows the distribution of the distance of each star in the Gaia solar cylinder and the random catalog from the Sun. The distance distribution is probing the spatial density distribution of stars in the solar cylinder subjected to the Gaia selection function. The top middle and top right panels of Figure 1 show the distributions of R and Z in the data and the random catalog. The bottom three panels show the distributions of the three velocity components for the data and the random catalog. The data and the random catalog seem to be in excellent agreement for the different spatial and velocity components. Consequently, the random catalogs created here generate an unclustered distribution of stars that adequately reproduce both the phase space distribution and the impact of the selection function of Gaia stars. We use the same method to generate the methods for the R18 mock, the K19a simulations, and the Gaia data.

A much more rigorous test of the random catalog construction is to apply our machinery to the R18 mock catalog, which should have no clustering signal. The resulting TPCF is shown in Figure 2. The top panel shows $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ and the bottom panel shows $\xi(\Delta v; \Delta r < 50 \text{ pc})$. There is no clustering signal at a level exceeding 10^{-2} , indicating that our approach to measuring the TPCF is reliable at this level. We note that the R18 mock contains both a Galactic warp and spatially inhomogeneous dust. Clearly these two physical effects do not have any effect on the measured correlation function.

3.3. Identifying and Removing Bound Structures

Calculating the TPCF is an exercise in pair-counting. Large open clusters that are nearby and well sampled in the data could dominate the TPCF because the pair counts scale as N^2 . There has been a significant amount of effort toward finding and characterizing these large open clusters (e.g., De la Fuente Marcos & De la Fuente Marcos 2009), and many new open clusters are being found with Gaia DR2 (e.g., Cantat-Gaudin et al. 2018; Castro-Ginard et al. 2020). It is easy to see that known open clusters could overwhelm the TPCF signal at small spatial scales.

Consequently, we choose to exclude pairs from these open clusters to isolate the signal of star clusters that are disrupting or have already disrupted. Similar to Oh et al. (2017) and K19b, we form an undirected graph where stars are nodes, and edges between the nodes exist for comoving pairs of stars. For the purposes of this work, we define comoving as having a 3D velocity difference of $< 2 \text{ km s}^{-1}$ and a physical separation of $< 5 \text{ pc}$; these selections are similar to linking lengths in the friends-of-friends algorithm. Consequently, a star could have multiple comoving neighbors, and a pair of stars could be directly or indirectly connected via a sequence of edges. The graph is then split into connected components—a connected component is a subgraph of the original graph in which any two nodes are connected to each other by a path—to calculate the connectivity of each star. A connectivity of 1 means that a star is not a part of any larger structure; a connectivity of 2 means that a star is in a mutually exclusive pair, and so on.

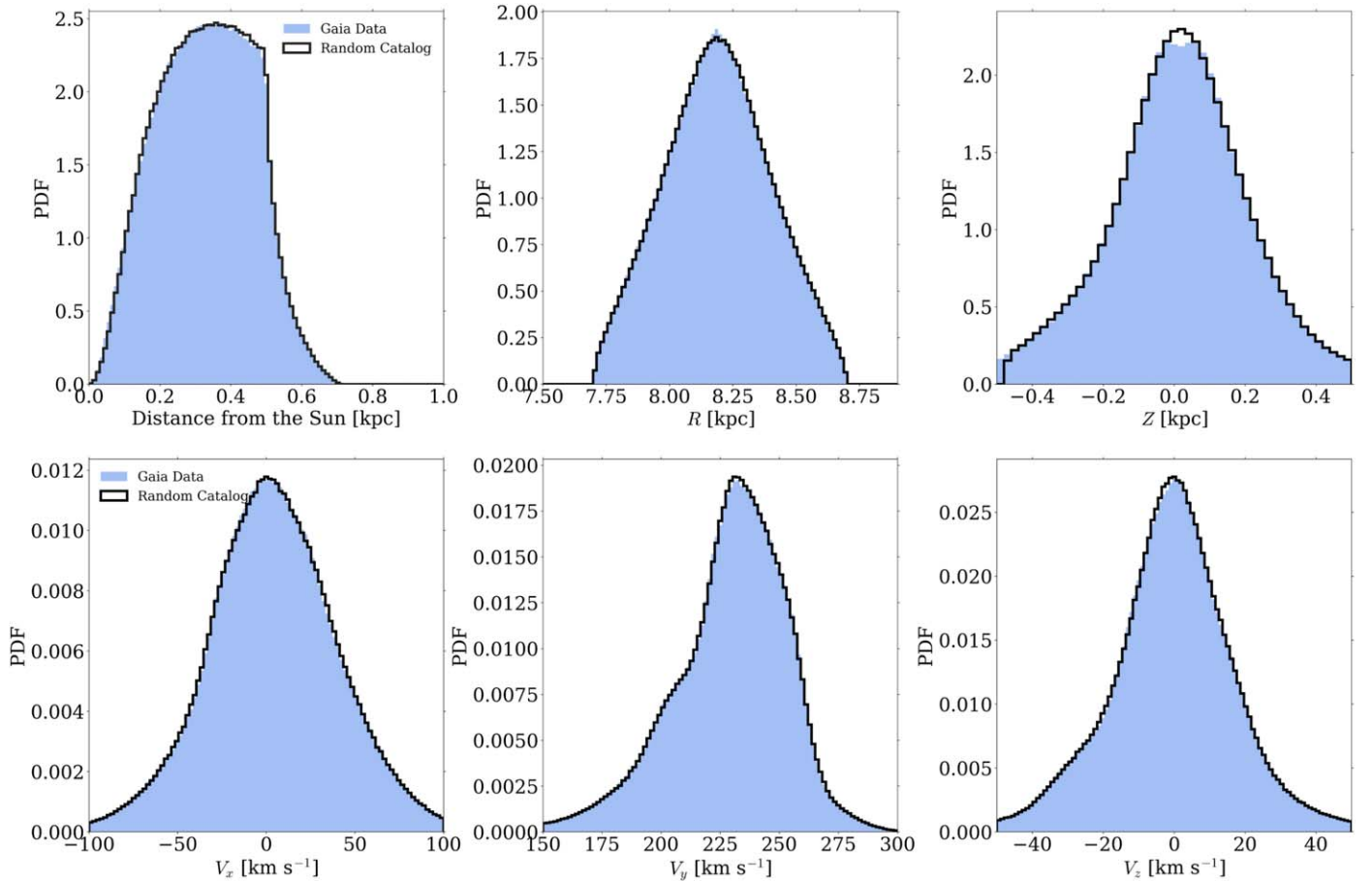


Figure 1. Validation of the method for random catalog construction by comparison to Gaia data. In each panel the distribution of Gaia data (blue) is compared to the random catalog (black). Top panels: comparison in the quantities of heliocentric distance, and Galactocentric R and Z distributions. Bottom panels: comparison in three components of the velocities: V_x , V_y , and V_z . Overall, the random catalog reproduces the distribution of the data in the solar neighborhood.

Bound wide binaries could also impact the TPCF at small spatial scales. The actual separation of most bound wide binaries is likely less than the Jacobi radius in the solar neighborhood (~ 1.7 pc). However, the median parallax uncertainty of ~ 0.04 mas corresponds to an uncertainty of ± 10 pc at 500 pc. Consequently, to minimize contamination from bound wide binaries, we impose the additional condition that all TPCF calculations for the data exclude pairs that have projected separation < 0.5 pc. A discussion of why we choose this criterion is included in Appendix B.1.

The top panel of Figure 3 shows the spatial TPCF calculated for stars with $\Delta v < 2$ km s $^{-1}$ and with different connectivity cuts in the Gaia data. The bottom left panel shows all the stars with a connectivity of > 50 . This selection efficiently identifies nearby open clusters; the clusters picked out above include Melotte 20, Pleiades, NGC 2516, Hyades, and Praesepe. The right panel shows stars with a connectivity between 2 and 50.

The TPCF that includes all stars (including stars from open clusters) shows the strongest clustering in the data. Even with a liberal connectivity cut where stars that are part of connected components with a size of 20 or less are included, the TPCF is notably stable and close to what it is with the very conservative cut of connectivity ≤ 2 , which only selects unconnected stars (connectivity = 1) and mutually exclusive stars (connectivity = 2). The TPCF being largely insensitive to smaller connectivity cuts is reassuring because it indicates that the bound open clusters are being effectively filtered out. Consequently, we

choose to use a selection of connectivity ≤ 5 for the rest of this work.

4. Results

4.1. Comparing Data and Simulations

In this section we present the spatial and kinematic TPCF in the data and compare to three simulations from K19a.

The top panel of Figure 4 shows the TPCF in the data and the fiducial simulation for stars with a 3D velocity difference of less than 2 km s $^{-1}$. The Jacobi radius (R_J) of wide binaries in the solar neighborhood is shown as an arrow on the x -axis; there could be contamination from bound wide binaries at separations smaller than R_J . Poisson uncertainties are shown as shaded bands. The three dashed lines show different curves with power-law index $\xi(\Delta r) \propto \Delta r^\gamma$, where $\gamma = -1.0$, -1.5 , and -2.0 .

There are three spatial regimes to consider: small (< 10 pc), intermediate (10 pc $< \Delta r < 100$ pc), and large (100 pc $< \Delta r < 1000$ pc). The data are slightly more clustered than the fiducial simulation at the smallest scales by a factor of ~ 2 – 3 . The fiducial simulation initializes all stars into star clusters (Lada & Lada 2003), and explicitly undervirializes the stars to mimic the boundedness of stars born together at young ages. The data could include some stars from subsampled bound open clusters at these small spatial scales that would not have an analog in the simulation.

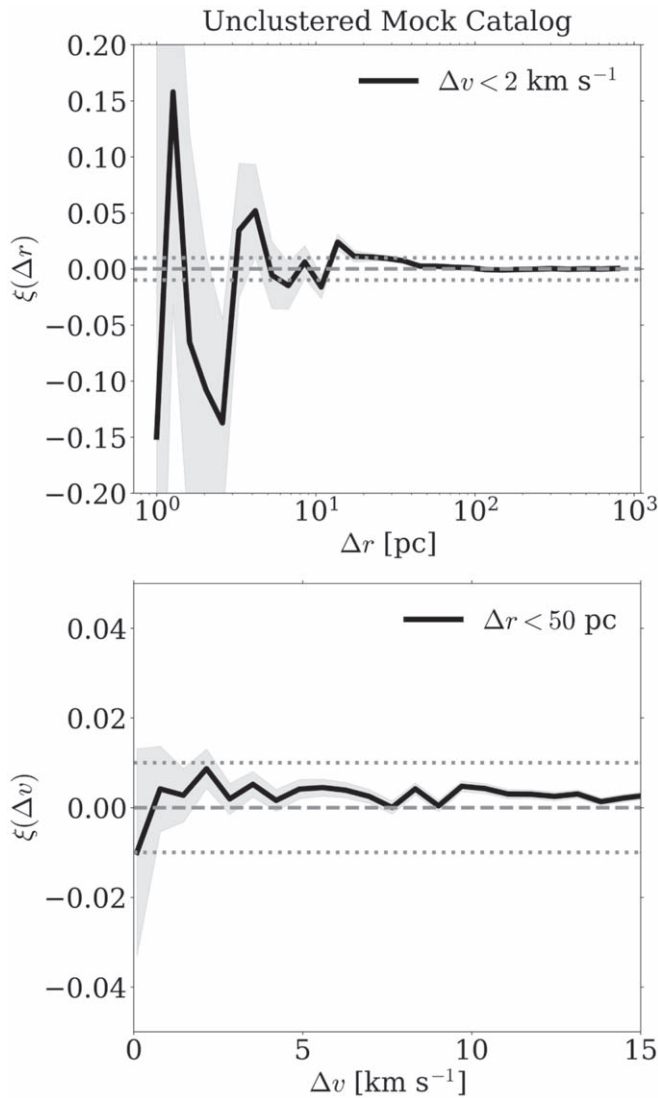


Figure 2. Correlation function measured in a mock data set (R18) that was constructed to be smooth in position and velocity space. Top panel: the spatial TPCF for the R18 mock with $\Delta v < 2 \text{ km s}^{-1}$. The shaded gray region shows the Poisson error. Bottom panel: the kinematic TPCF for the R18 mock with $\Delta r < 50 \text{ pc}$. The shaded gray region shows the Poisson error. In both panels there is no statistically significant clustering above 10^{-2} at any scale. The absence of any signal in the TPCF for the R18 mock indicates the fidelity of the technique used to generate the random catalogs and measure the correlation function.

At intermediate scales, the simulation and the data are in reasonable agreement (see Figure 6 for a discussion on sample variance). The clustering at intermediate scales could be driven by disrupting star clusters, unbound stellar associations, or hierarchical star formation. In our current model, we follow Lada & Lada (2003) and assume an infant mortality rate for clusters of 80%–90% (more details in K19a). Hierarchical star formation, on the other hand, could imply a cluster formation efficiency (CFE) of $<10\%$ for Milky Way–like galaxies (e.g., Kruijssen 2012; Ward et al. 2020). If our assumed CFE in the simulations is too high, that could drive up clustering at small and intermediate scales. In future work we will run a grid of simulations with hierarchical star formation and without it to further test our assumptions.

The data are again more clustered than the simulation at the largest scales. Clustering at these scales could be caused

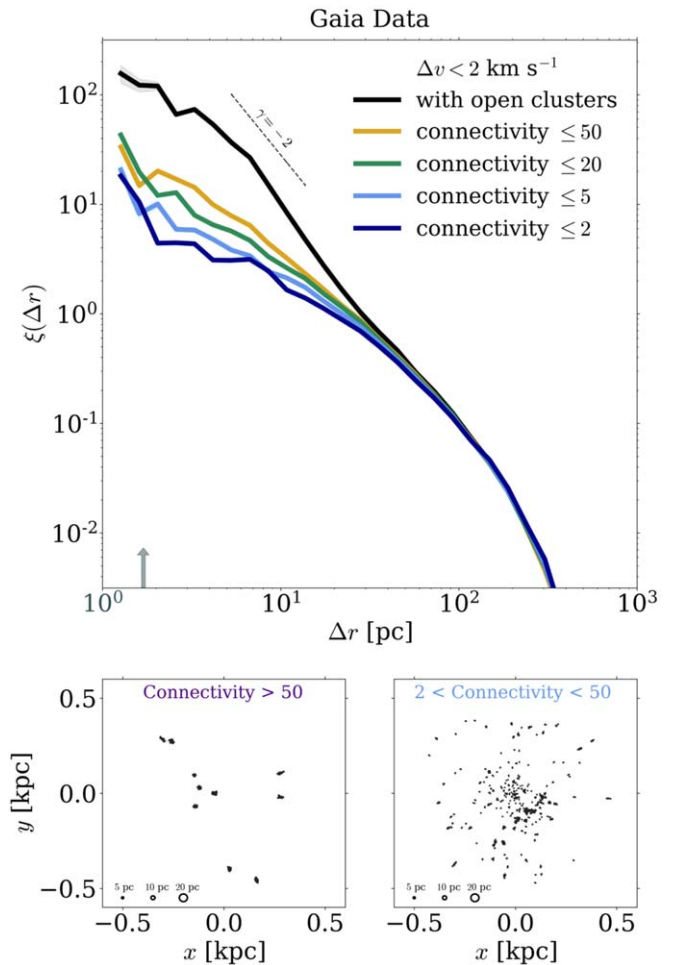


Figure 3. TPCF for the Gaia data with $\Delta v < 2 \text{ km s}^{-1}$ and various connectivity selections, and an illustration of what these different selections look like spatially. Top panel: different colors show the spatial TPCF with open clusters and with varying connectivity cuts ($\leq \{2, 5, 10, 50\}$) to remove bound structures. The Jacobi radius for wide binaries in the solar neighborhood is shown as an arrow on the x-axis. The black line shows the power-law index $\gamma = -2$. The TPCF at scales $< 30 \text{ pc}$ for no connectivity cut is largely dominated by stars from a few bound open clusters. As the connectivity cut gets smaller, there is not a large difference between connectivities of ≤ 20 and ≤ 2 . For the rest of this work, we choose stars with a connectivity of ≤ 5 for both the data and the simulations. Bottom left: stars in the solar neighborhood with a connectivity of > 50 . Most of the stars shown here are part of a bound open cluster. Bottom right: stars with a connectivity between 2 and 50. These stars are may be part of a larger star-forming region, stars that are subsampled from small or more diffuse open clusters, or chance alignments.

either by the few pairs of stars that were born together but drifted apart or by the resonances related to the non-axisymmetries of the Galactic potential. A hierarchical model for star formation could explain the clustering at large Δr since we would expect different star-forming regions to also be spatially correlated (Grasha et al. 2017). The fiducial simulation includes a realistic bar and rotating but fixed spiral arms; consequently, the disagreement at larger Δr could also indicate clustering due to resonances in the transient spiral arms. As shown in previous work (e.g., Hunt et al. 2018; Sellwood et al. 2019), the inclusion of transient spiral arms will likely show richer phase space structure in the solar neighborhood at large scales.

The bottom panel of Figure 4 shows the fraction of pairs in the simulation born together and the fraction of field pairs as a function of their spatial separation for a velocity difference of

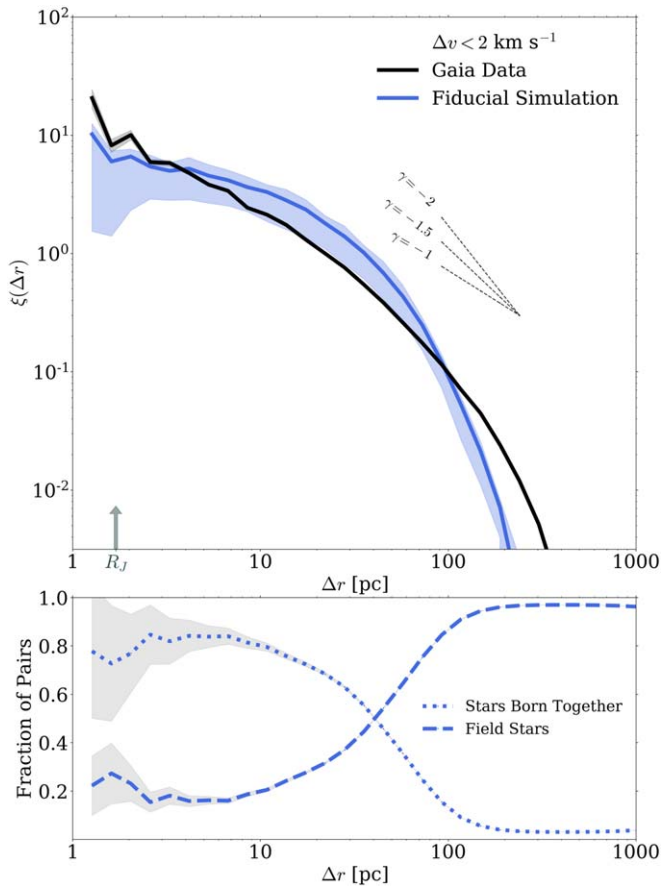


Figure 4. Top panel: TPCF for the Gaia solar cylinder (black) and the fiducial simulation (blue) with a 3D velocity difference of 2 km s^{-1} . Poisson errors for the data are shaded in black. The blue shaded region shows the sample variance for the simulation (calculated by measuring the TPCF for different “solar neighborhoods”) and the Poisson error added in quadrature. An arrow shows the Jacobi radius, R_J , of solar-mass wide binaries in the solar neighborhood. Open clusters are excluded in the calculation of the TPCF (see Section 3.3). Bottom panel: fraction of pairs in the simulation that were born together (blue dotted) and the fraction that are chance field alignments (blue dashed). At low Δr (and $\Delta v < 2 \text{ km s}^{-1}$), the fraction of stars born together dominates the simulation TPCF presented in the upper panel. Beyond $\sim 40 \text{ pc}$ the majority of pairs arise from chance field alignments.

$\Delta v < 2 \text{ km s}^{-1}$. The fraction of stars born together dominates the pair counts up to $\sim 40 \text{ pc}$. Field pairs dominate above for larger spatial scales. Consequently, the small spatial scales ($< 10 \text{ pc}$) probe clustered star formation, the intermediate scales ($10 \text{ pc} < r < 100 \text{ pc}$) probe clustering due to both star formation (SF) and other clustering mechanisms, and the large scales ($100 \text{ pc} < r < 1000 \text{ pc}$) mostly probe non-SF related clustering in the simulation.

Figure 4 simultaneously probes star formation at small scales, the disruption mechanism of star clusters, and the resonances in the Galaxy. However, the chosen velocity difference is only probing clustering for one Δv cut. Figure 5 shows the TPCF for the data, and the three simulations from K19a for the velocity differences: $\Delta v < \{1, 2, 4\} \text{ km s}^{-1}$. The left panel shows the data (solid line) and the fiducial simulation (dashed line) for the three different velocity differences. The shaded regions accompanying each curve show the Poisson errors. The pattern observed in Figure 5 for $\Delta v < 2 \text{ km s}^{-1}$ also holds for $\Delta v < 1$ and 4 km s^{-1} .

The data and simulation largely agree at small scales, the simulation shows slightly more clustering at intermediate

scales, and the data are more clustered at the largest scales. The discrepancy between the data and the simulation at small spatial scales for $\Delta v < 4 \text{ km s}^{-1}$ could suggest that the birth velocity dispersion for stars born together in the data could reach larger values than the model used in the simulation. The birth velocity dispersion for star clusters in the simulation is determined by assuming some potential for the star cluster and drawing from the cluster birth mass–radius relation. The latter is particularly uncertain (e.g., Parmentier & Kroupa 2011), but recent work (Choksi & Kruijssen 2021) has made progress in estimating a physically motivated mass–radius relation. Similar to $\Delta v < 2 \text{ km s}^{-1}$, the difference at intermediate and large scales could be attributed to a hierarchical mode of star formation, resonances due to transient spiral arms, or a combination thereof.

The right panel shows the TPCF for the three different simulations (fiducial, red; axisymmetric, gray; and NCSF, indigo) for $\Delta v < 2 \text{ km s}^{-1}$. The NCSF simulation, as expected, has the lowest clustering amplitude because of the absence of clustered star formation. The NCSF simulation does have non-axisymmetries but these are not detectable in the TPCF at $\Delta v < 2 \text{ km s}^{-1}$. The axisymmetric simulation has the largest clustering amplitude because of the absence of scattering from the bar, spiral arms, and GMCs, and hence star clusters disrupt the slowest. The different red lines near the fiducial simulation show the TPCF for different realizations of the “solar” cylinder from different regions in the solar annulus in an effort to study sample variance. There is considerable scatter in the TPCF for the different solar cylinders—the numbers vary by almost a factor of 3–4 at smaller scales. Such a large sample variance necessitates caution in comparing the data and the simulation, especially when the two are very similar.

In Figure 6 we calculate the power-law slope γ for the different Δv cuts by fitting the relation $\xi(\Delta r) \propto \Delta r^\gamma$ to the Gaia data after making the connectivity selection discussed in Section 3.3 (solid lines) and without a connectivity cut for $\Delta v < 2 \text{ km s}^{-1}$ (dashed). The slope is calculated for five bins on a rolling basis through the different spatial scales. For $\lesssim 50 \text{ pc}$, the index hovers around ~ -1 for all Δv . Regardless of Δv , the power-law slope falls precipitously after 50–70 pc to $\lesssim -2$. With no connectivity cut, the data are well described by $\gamma \sim -2$ from 10–100 pc.

So far, we have focused on the spatial TPCF by selecting samples in a narrow Δv range; we can also consider the converse and measure the velocity-space TPCF in a narrow Δr range. Figure 7 shows such measurements comparing the data and the fiducial simulation. There is clearly more clustering in data for the three different spatial selections for most Δv ; at small Δv , the simulation is slightly more clustered than the data, as shown in Figure 5. For $\Delta r < 5 \text{ pc}$ both the data and simulation are clustered at low Δv (up to $\sim 5 \text{ km s}^{-1}$); however, the data are more clustered at larger Δv (up to 10 km s^{-1}). If the signal at low Δv is due to co-natal stars, the larger clustering in the data could indicate that the birth velocity dispersion of stars born together is larger in data than in the simulation. There is a similar trend, though less drastic, for $\Delta r < 20$ and 50 pc . The data and the model are reasonably close at low Δv but seem to diverge at large Δv . The co-natal fraction in the fiducial simulation at such large Δv is quite low, even for $\Delta r < 5 \text{ pc}$ (K19a). Consequently, the discrepancy could be driven by the rich structure created due to either the non-axisymmetries of the Galaxy and/or non-equilibrium

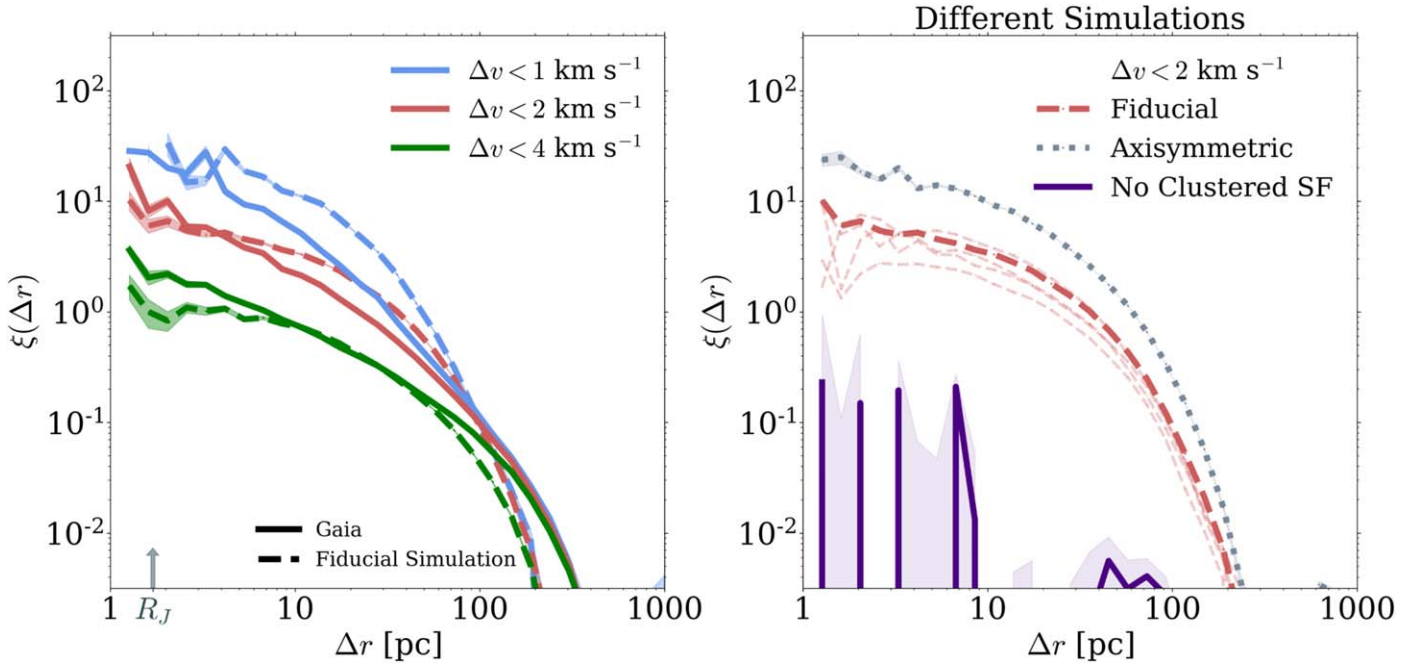


Figure 5. Left panel: TPCF with $\Delta v = \{1, 2, 4\} \text{ km s}^{-1}$ (blue, red, green) for the data (solid lines) and the simulation (dashed lines). Each curve has associated Poisson errors shaded. The features at different spatial scales discussed in Figure 5 for $\Delta v < 2 \text{ km s}^{-1}$ are also apparent at different velocity amplitudes. The data seem to diverge the most from the simulation at $\Delta v < 1 \text{ km s}^{-1}$ at intermediate scales. The discrepancies between the data and the fiducial simulation for different Δv indicate the need for a more complex subgrid model of star formation and the non-axisymmetries in the potential. Right panel: TPCF for the fiducial (red), axisymmetric (gray), and NCSF (purple) simulations. The light red lines near the fiducial TPCF show the TPCF in four different solar cylinders throughout the simulated galaxy. As expected, the axisymmetric simulation is more clustered due to no scattering and the NCSF simulation is less clustered due to the absence of clustered star formation. The sample variance in the different solar neighborhoods contributes a factor of 2–3 variation in the TPCF.

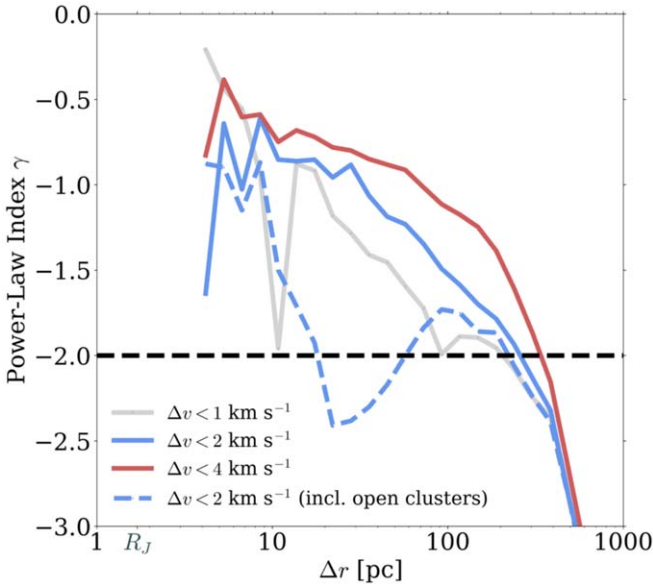


Figure 6. Power-law index ($\xi(\Delta r) \propto \Delta r^\gamma$) for the Gaia spatial TPCF with different Δv cuts. We compute the average power-law index on a rolling basis. For $\lesssim 50 \text{ pc}$, the index is ~ -1 for all Δv . Regardless of Δv , the index falls precipitously beyond $50\text{--}70 \text{ pc}$ to $\lesssim -2$. The power-law index for the data that includes bound open clusters (no connectivity cut) is ~ -2 for $\Delta r > 10 \text{ pc}$.

phenomena in the disk (e.g., Laporte et al. 2018). A thorough test would involve calculating the TPCF for simulations with satellites (either isolated or cosmological), such as Laporte et al. (2018) and Sanderson et al. (2020), and comparing the velocity-space TPCF at large Δv .

The right panel shows the velocity TPCF of the three simulation variants for $\Delta r < 20 \text{ pc}$. The NCSF simulation has

the lowest clustering amplitude, as expected. The NCSF simulation does have non-axisymmetries but this is evidently not detectable in the kinematic TPCF for $\Delta r < 20 \text{ pc}$ at any velocity scale. The thin red lines show the TPCF for different realizations of the “solar cylinder” in the fiducial simulation. As in Figure 4, the sample variance is significant.

We speculate that the discrepancies between the data and the simulations at large Δv scales—where the clustering should be largely driven by field pairs—could be driven by either transient spiral arms or interactions with a satellite. Both simulations include a bar and steady-state spiral arms but do not include transient modes and interactions with external perturbers. Recent work (e.g., Hunt et al. 2018; Sellwood et al. 2019) has shown that transient spiral arms can recreate velocity structure very close to what is observed in the data. Modelling external perturbers is more complex because of the need to run expensive N -body simulations.

4.2. Predicted Clustering as a Function of Gaia Uncertainties, Age, and Metallicity

The analysis so far has focused on the data products from Gaia DR2, released in April 2018. The next decade will see an unprecedented increase in both the volume and precision of observational data on stars in the Galaxy. Gaia, in particular, will deliver the radial velocities of tens of millions of stars, and will decrease the astrometric uncertainties by at least a factor of 1.5.¹¹ The increase in precision and complementary spectroscopic and asteroseismic data prompt the question of

¹¹ The Gaia data improve with time as $t^{-0.5}$ for parallaxes, photometry, and radial velocities, and as $t^{-1.5}$ for proper motions (<http://pc500.astro.lu.se/gaia2017/slides/Brown.pdf>).

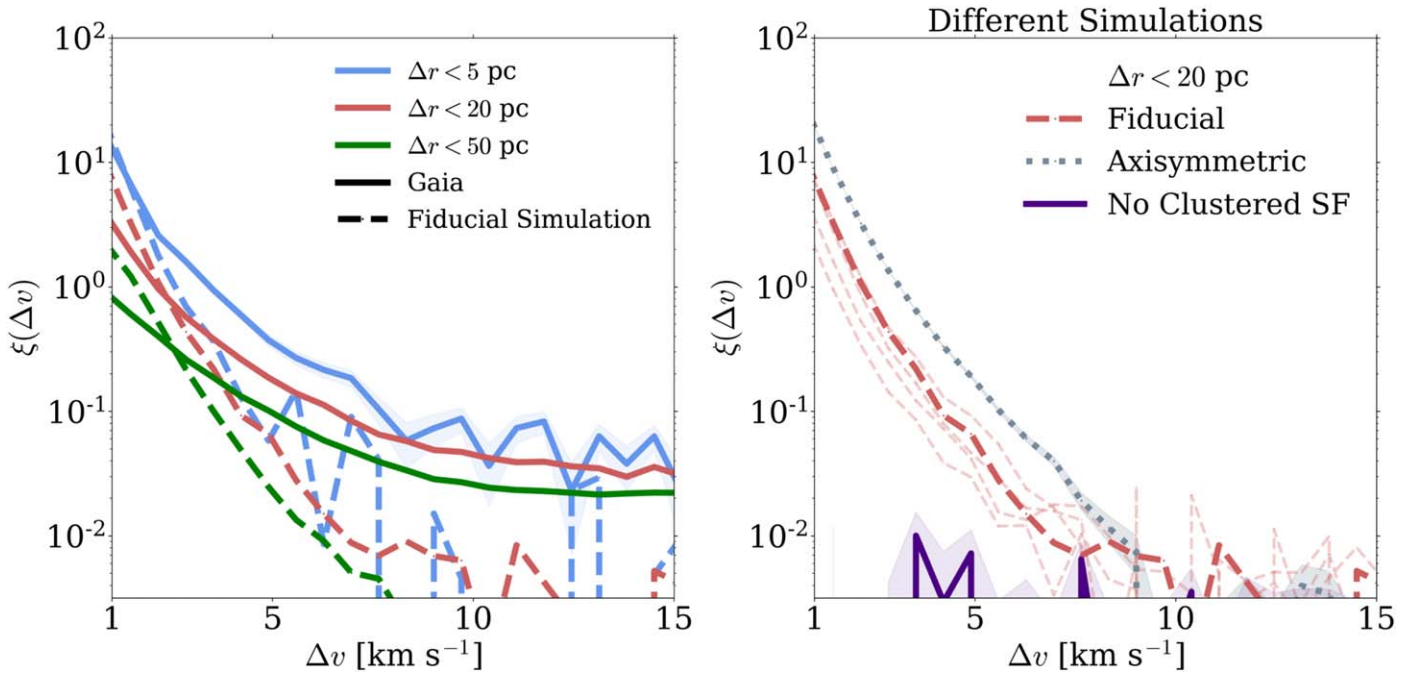


Figure 7. Left panel: velocity TPCF with $\Delta r = \{5, 20, 50\}$ pc (blue, red, green) for the data (solid lines) and the simulation (dashed lines). Each curve has associated Poisson errors shaded. The simulation and the data seem to be reasonably close at low Δv (up to $\sim 3\text{--}4$ km s $^{-1}$); however, for all larger Δv , the data are more clustered than the simulation by a factor of a few regardless of the spatial scale. This rich velocity structure regardless of spatial scale likely points to both a more complex initial velocity profile of cluster dispersion and the impact of resonances in the disk. Right panel: velocity TPCF for the fiducial (red), axisymmetric (gray), and the NCSF (purple) simulations. The light red lines near the fiducial TPCF show the TPCF in different solar cylinders throughout the simulated galaxy. The sample variance for the kinematic TPCF is comparable to that of the spatial TPCF shown in Figure 4.

how much the information content changes as new and more precise data become available.

We begin by exploring the effect of Gaia measurement uncertainties on the TPCF. Figure 8 shows the spatial TPCF for the fiducial simulation for DR2 errors (solid), for DR4 errors (dashed), and for no errors (dashed-dotted). To calculate the DR4 errors, we scale the parallax and the RV errors by $1/1.7$ and the proper motion errors by $1/4.5$. The DR4 $\xi(\Delta r)$ is not too different for $\Delta v < 4$ and 8 km s $^{-1}$, which is not surprising given that the median DR2 uncertainty in our sample is 1.15 km s $^{-1}$. $\xi(\Delta r)$ is expected to be approximately a factor of 1.5 times larger than the DR2 $\xi(\Delta r)$ for small Δv . The dashed-dotted line shows the maximal change in $\xi(\Delta r)$ with perfect phase space information. As with DR4 errors, there is little difference at large Δv . However, the gain in $\xi(\Delta r)$ that can be extracted with perfect phase space information for low Δv is more than a factor of 5 times that that Gaia DR2 currently provides.

Alongside Gaia, many ongoing spectroscopic surveys (e.g., Deng et al. 2012; De Silva et al. 2015; Kunder et al. 2017; Majewski et al. 2017) and upcoming ones (e.g., De Jong et al. 2012; Dalton et al. 2014; Kollmeier et al. 2017) will deliver precise chemical information for millions of stars. The key challenge with modeling the multidimensional chemistry that these surveys will measure is grappling with the inherent dimensionality of the chemical space (e.g., Ting et al. 2012; Price-Jones & Bovy 2018). Moreover, the uncertainties of the derived abundances are difficult to accurately forecast; as shown in Ting et al. (2015b), the effective uncertainty could be much smaller if the covariances are not taken into account, due to the large dimensionality of the chemical space. Moreover, recent work (e.g., Martig et al. 2016; Bovy et al. 2019; Ting &

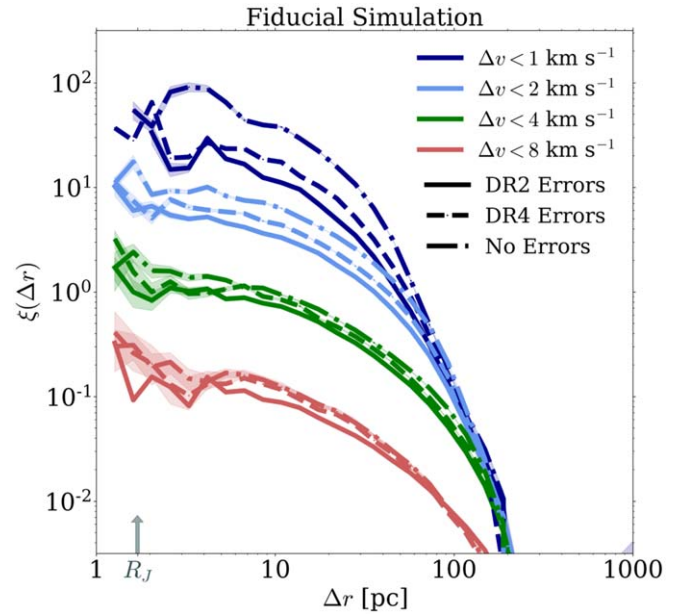


Figure 8. Impact of Gaia DR2 errors, DR4 errors, and no phase space errors on the simulated $\xi(\Delta r; \Delta v < \{1, 2, 4, 8\}$ km s $^{-1}$). The improvement in precision leads to a modest change in $\xi(\Delta r)$ for $\Delta v < 4$ and 8 km s $^{-1}$, and a notable increase in $\xi(\Delta r)$ for $\Delta v < 1$ and 2 km s $^{-1}$.

Rix 2019) has shown the promise of utilizing age information to make inferences about dynamics in the Galactic disk.

Figure 9 shows the potential impact of incorporating metallicity and age information when measuring the TPCF. The top panels show $\xi(\Delta r; \Delta v < 2$ km s $^{-1}$) with various metallicity cuts for $\sigma_{[\text{Fe}/\text{H}]} = 0.01$ and 0.05 dex. For

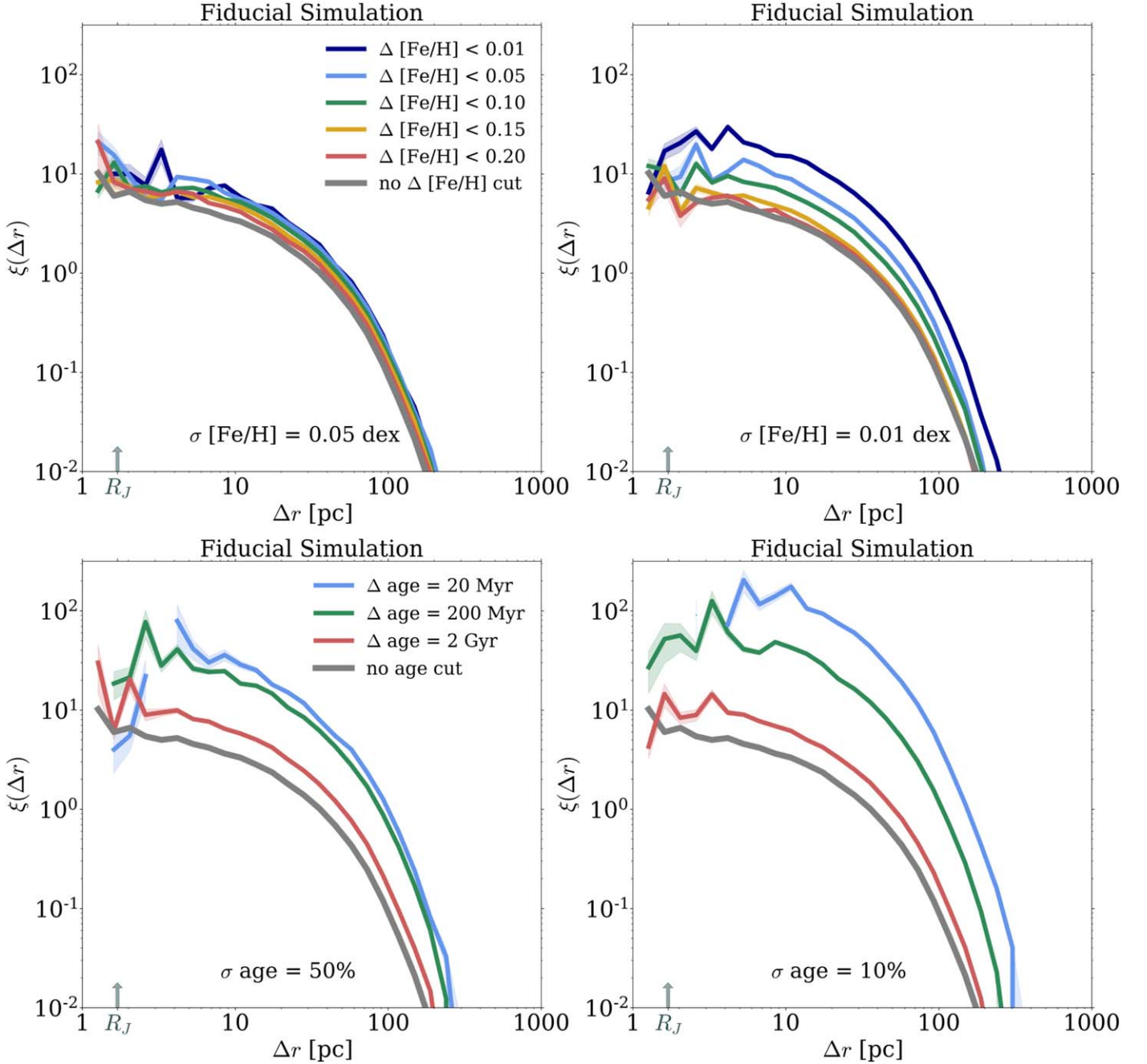


Figure 9. Change in $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ when including metallicities with $\sigma_{[\text{Fe}/\text{H}]} = 0.01$ and 0.05 dex, and ages with $\sigma_{\text{age}} = 10\%$ and 50% . Top left: $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ at various metallicity cuts with $\sigma_{[\text{Fe}/\text{H}]} = 0.05$ dex. The relative change in $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ with the addition of metallicity is about a factor of ~ 1.7 – 2 . Top right: $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ at various metallicity cuts with $\sigma_{[\text{Fe}/\text{H}]} = 0.01$ dex (almost perfect metallicity). $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ with almost perfect metallicity information is an order of magnitude larger for the lowest $\Delta[\text{Fe}/\text{H}]$. Bottom left: $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ with different Δage cuts and an age uncertainty of 50% . Even with an uncertainty of a factor of 2 , the addition of ages leads to a similar change in $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ to the inclusion of almost perfect metallicity information. Bottom right: $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ with different Δage cuts and an age uncertainty of 10% . The change in $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ for the lowest Δage cut is more than an order of magnitude. As expected, precise ages are very effective at identifying co-natal stars.

$\sigma_{[\text{Fe}/\text{H}]} = 0.05$ dex, the relative gain in information for the smallest $\Delta[\text{Fe}/\text{H}]$ is a modest factor of ~ 1.6 – 1.8 . The results for the TPCF line up well with those presented in K19a for the co-natal fraction of stars with additional metallicity information. The right panel shows the same $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ with various metallicity cuts but for an almost perfect $[\text{Fe}/\text{H}]$ measurement with $\sigma_{[\text{Fe}/\text{H}]} = 0.01$ dex. The change in $\xi(\Delta r)$ is almost an order of magnitude because the stringent $\Delta[\text{Fe}/\text{H}]$ selection is very efficient at identifying co-natal stars.

The bottom panels of Figure 9 show $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ with various different age cuts ($\Delta \text{age} = 20, 200$, and 2000 Myr) for age uncertainties $\sigma_{\text{age}} = 10\%$ and 50% . Even with an age uncertainty of 50% , $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ increases by an order of magnitude for the smallest Δage cut of 20 Myr. There are a few possible reasons for this large change with such uncertain ages. First, there is a much larger dynamic range in stellar ages than in metallicities. Second, the uncertainties are relative, and so younger stars have smaller

absolute uncertainties. Last, there is a stronger coupling between age and dynamics than between metallicity and dynamics; for instance, the age–velocity dispersion relation has a smaller scatter than the metallicity–velocity dispersion relation. Consequently, a weak prior on age and a strong prior on metallicity lead to an analogous change in the TPCF. The more precise ages ($\sigma_{\text{age}} = 10\%$) increase $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ by almost two orders of magnitude; precise ages combined with kinematics hold the most information about stars born together in the disk.

5. Summary

Several key physical processes including the clustered nature of star formation, non-axisymmetries of the Galactic potential, and non-equilibrium phenomena determine the structure of the Galaxy in chemodynamical space. The two-point correlation function, a clustering metric widely used in other fields of astronomy and physics, is well suited to the task of disentangling structure in chemodynamical space caused by these distinct physical processes.

In this paper we presented a robust, nonparametric technique to generate realistic random catalogs for a complex density profile and a nontrivial selection function using Dirichlet process Gaussian mixture models. We validated the fidelity of our random catalog by calculating the TPCF of the R18 mock. We calculated the spatial and kinematic TPCF in the data and three simulations from K19a sliced in velocity, separation, metallicity, and age. The resulting structure in these different contexts holds valuable clues about the nature of star formation and the importance of non-axisymmetries in the Galaxy. The spatial and kinematic TPCFs presented above reaffirm the need for a deeper look at the cluster disruption model in K19a, and a thorough treatment of the transient non-axisymmetries of the Galactic potential and a hierarchical model of star formation to study clustering at larger Δr and Δv .

Our key findings are listed below.

1. We calculate the spatial TPCF for stars with velocity differences (Δv) of 1, 2, and 4 km s^{-1} and the kinematic TPCF for stars with spatial (Δr) separations of 5, 20, and 50 pc in the solar neighborhood with data from Gaia DR2. We detect clustering out to large spatial and kinematic scales (up to 300 pc and 15 km s^{-1}). The power-law index of the spatial TPCF that includes bound structures is ~ -2 for $\Delta r > 10$ pc (in line with theoretical predictions), and is ~ -1 up to 50 pc and then drops precipitously to $\lesssim -2$ for larger Δr without bound structures.
2. We analyze a novel star-by-star simulation (K19a) to interpret the observational results. The data and the simulation agree reasonably well at small spatial scales but there is some tension at intermediate and large spatial scales. Since we assume in K19a that all stars are born in clusters (naturally leading to a more clustered population of stars), we suggest that the mismatch at intermediate scales could be explained by hierarchical star formation. For $\Delta v > 5 \text{ km s}^{-1}$ and > 100 pc, the data show rich clustering in the spatial and kinematic TPCF that is absent in the simulations. Since the co-natal fraction in our simulations is small at these scales, we speculate that the low clustering strength in the simulations is due to the lack of transient spiral arms in the simulation.

3. Ongoing Gaia data collection and upcoming spectroscopic surveys of the Galaxy promise to revolutionize the field of Galactic archeology. We make predictions about how future Gaia errors and the inclusion of [Fe/H] and age information will affect measurements of clustering in chemodynamical space. We predict that gains in Gaia precision will increase $\xi(\Delta r)$ by a factor of 1.5, a metallicity uncertainty of $\sigma_{[\text{Fe}/\text{H}]} = 0.05$ dex will increase the TPCF by a factor of two, and even 50% uncertain ages will significantly enhance the TPCF.

We expect that the discrepancies between the simulated Galaxy and Gaia data will lead to new insights regarding the clustered nature of star formation and non-axisymmetric, time-dependent components of the Galactic potential.

We thank Angus Beane, Anthony Brown, Lehman Garrison, Yan-Fei Jiang, Diederik Kruijssen, Hans-Walter Rix, and members of the Conroy group at Harvard for useful discussions and helpful comments. H.M.K. acknowledges support from the DOE CSGF under grant number DE-FG02-97ER25308. C.C. acknowledges support from the Packard Foundation. Y.S.T. is supported by the NASA Hubble Fellowship grant HST-HF2-51425.001 awarded by the Space Telescope Science Institute. The computations in this paper were run on the Odyssey cluster supported by the FAS Division of Science, Research Computing Group at Harvard University.

This work has made use of data from the European Space Agency 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. The Sloan Digital Sky Survey IV is funded by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions and acknowledges support and resources from the Center for High-Performance Computing at the University of Utah.

Software: CorrFunc (Sinha & Garrison 2018, 2020), IPython (Pérez & Granger 2007), Cython (Behnel et al. 2011), Astropy (Astropy Collaboration et al. 2013, 2018), NumPy (Van Der Walt et al. 2011), SciPy (Jones et al. 2001), scikit-Learn (Pedregosa et al. 2011), Matplotlib (Hunter 2007).

Appendix A Volume Effects

In our analyses we found a dramatic decline in the TPCF for both the data and the simulation at around $r \sim 200\text{--}400$ pc. Since the solar cylinders we consider here all have a radius of 0.5 kpc, we test here whether or not the decline beyond $r \sim 200$ pc is due to volume effects. The volume we chose restricted stars to within ~ 0.71 kpc of the Sun (largest possible distance from the Sun with a cylinder radius of 0.5 kpc and height of 0.5 kpc above/below the Sun). The TPCF considers any pair contained within the volume; the largest possible distance between any such pair of stars in this volume is 1.41 kpc. However, there likely are not many pairs this far apart because most stars in this volume are contained within the thin disk (scale height ~ 200 pc).

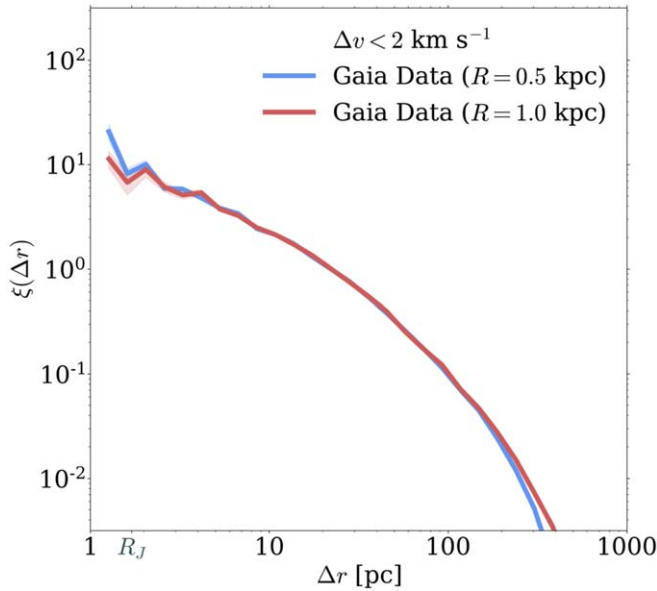


Figure 10. $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ for solar cylinders of radii $R = 0.5 \text{ kpc}$ and 1 kpc . The two TPCFs are nearly identical. We conclude that the rapid decline in the TPCF beyond 100 pc is not a volume effect.

Figure 10 shows the spatial TPCF of Gaia with the fiducial cylinder radius of 0.5 kpc and a larger radius of 1 kpc . The TPCFs for the two volumes are almost identical. There are some discrepancies at the smallest spatial scales, which are driven by Poisson error and the smaller number of pairs. Overall, the excellent agreement confirms that the rapid decline in $\xi(\Delta r)$ at large r is not due to our adopted geometry.

Appendix B

The Impact of Wide Binaries on the TPCF

B.1. Bound Wide Binaries

Wide binaries are ubiquitous in the Galaxy (e.g., Jiang & Tremaine 2010; El-Badry & Rix 2019; El-Badry et al. 2019; Tian et al. 2019). Bound wide binaries are expected to have separations ranging from tens of astronomical units to $\sim 1\text{--}2 \text{ pc}$. The TPCF calculations presented in this work are all for pair separations of $>1 \text{ pc}$. In the absence of observational uncertainties, we would thus not expect a significant fraction of pairs contributing to the TPCF to be bound binaries. However, observational uncertainties—particularly the uncertainty in parallax—make it difficult to measure the true 3D separation of close pairs. Our sample’s median parallax uncertainty of $\sim 0.04 \text{ mas}$ corresponds to a distance uncertainty of $\pm 2 \text{ pc}$ at a distance of 200 pc , and $\pm 10 \text{ pc}$ at a distance of 500 pc . This means that parallax uncertainties can inflate the apparent separations of wide binaries that have true separations of $\ll 1 \text{ pc}$ up to $\sim 20 \text{ pc}$, and could lead to biases in the TPCF at $\Delta r \lesssim 20 \text{ pc}$ if they are not removed from the sample.

“Stretching” of wide binaries along the line of sight due to parallax uncertainties is a configuration space analog of the Fingers-of-God effect in redshift space. For binaries with small true separations, observational uncertainties significantly inflate their separations along the line of sight, but not their projected separations on the plane of the sky. We therefore calculate the projected separation of pairs (angular separation times the mean distance of the two stars) with $\Delta r < 20 \text{ pc}$ in the Gaia data and in the R18 mock (which contains no wide binaries). For true

wide binaries, the projected separation should be much smaller than the calculated 3D separation. For non-binaries, the two separations should be comparable, since it is only for rare geometric alignments that the 3D separation is much greater than the 2D separation. The comparison between the 3D separation and the projected 2D separation is shown in Figure 11. As expected, there is a large population of pairs with projected separation orders of magnitude smaller than the calculated 3D separation in the data, and almost none in the unclustered mock.

The left panel of Figure 12 shows the fraction of stars with projected separation $< 0.5 \text{ pc}$ for the data and the R18 mock. The contamination is $\sim 90\%$ below the Jacobi radius (dashed line), and falls to $\lesssim 10\%$ after 10 pc . The unclustered mock has very few pairs with projected separation $< 0.5 \text{ pc}$ out to 3D separations of 20 pc . Consequently, our procedure of removing pairs with projected separations $< 0.5 \text{ pc}$ effectively selects bound wide binaries in the data, and removes very few non-binary pairs. We therefore use this selection when computing the TPCF in the main text. The right panel of Figure 12 shows $\xi(\Delta r; \Delta v < 2 \text{ km s}^{-1})$ with and without the projected separation cut. The clustering signals for $\Delta r \lesssim 10 \text{ pc}$ differ by an order of magnitude due to the presence of bound wide binaries.

B.2. Unbound Wide Binaries

Jiang & Tremaine (2010, hereafter J10) simulated the orbital evolution and dissolution of wide binaries in the Galactic disk. They argued that unbound wide binaries could remain close in phase space after being disrupted, which would lead to enhanced clustering at spatial separations extending beyond the Jacobi radius. Briefly, the models presented in J10 study the evolution of wide binaries due to gravitational perturbations from passing stars, and the Galactic tidal field. The component stars were tracked even after a wide binary became unbound. The discussion below considers the “Opik 1” model presented in J10, which resembles the solar neighborhood. It is worth noting that the simulations presented in J10 do not include the impact of GMCs, which are likely an additional important scattering mechanism in the Galaxy (e.g., Weinberg et al. 1987).

We argued in K19b that there is a trough in the distribution of unbound wide binary separations for the spatial scales $\sim 2\text{--}20 \text{ pc}$. However, J10 predict that unbound wide binaries that are slowly drifting apart could lead to a peak in the separation distribution out to separations of $\sim 100\text{--}300 \text{ pc}$ for relative velocities $\Delta v \sim 0.1\text{--}0.2 \text{ km s}^{-1}$. Almost all unbound wide binaries in the Opik 1 model have $\Delta v \leq 0.5 \text{ km s}^{-1}$, with the majority at $\Delta v \leq 0.2 \text{ km s}^{-1}$. However, given the uncertainties in the Gaia data (especially the radial velocities), the computed Δv for wide binaries could be a few times that.

The contamination of wide binaries in the data can be estimated by contrasting the spatial TPCF for $\Delta v < 4 \text{ km s}^{-1}$ with the TPCF for $1 \text{ km s}^{-1} < \Delta v < 4 \text{ km s}^{-1}$. If many wide binaries are contributing to the TPCF signal, we would expect a precipitous drop-off in the TPCF when we probe the larger TPCF for $1 \text{ km s}^{-1} < \Delta v < 4 \text{ km s}^{-1}$. The left panel of Figure 13 shows the spatial TPCF with these different velocity slices for the data and the simulation. The simulation is plotted as a control to show what the same relative Δv cut looks like in a mock without wide binaries; this is to motivate how much $\xi(\Delta r)$ changes just due to the change in the co-natal fraction. The right panel shows the fractions $\frac{\xi(\Delta r; 1 < \Delta v < 4)}{\xi(\Delta r; \Delta v < 4)}$ in red.

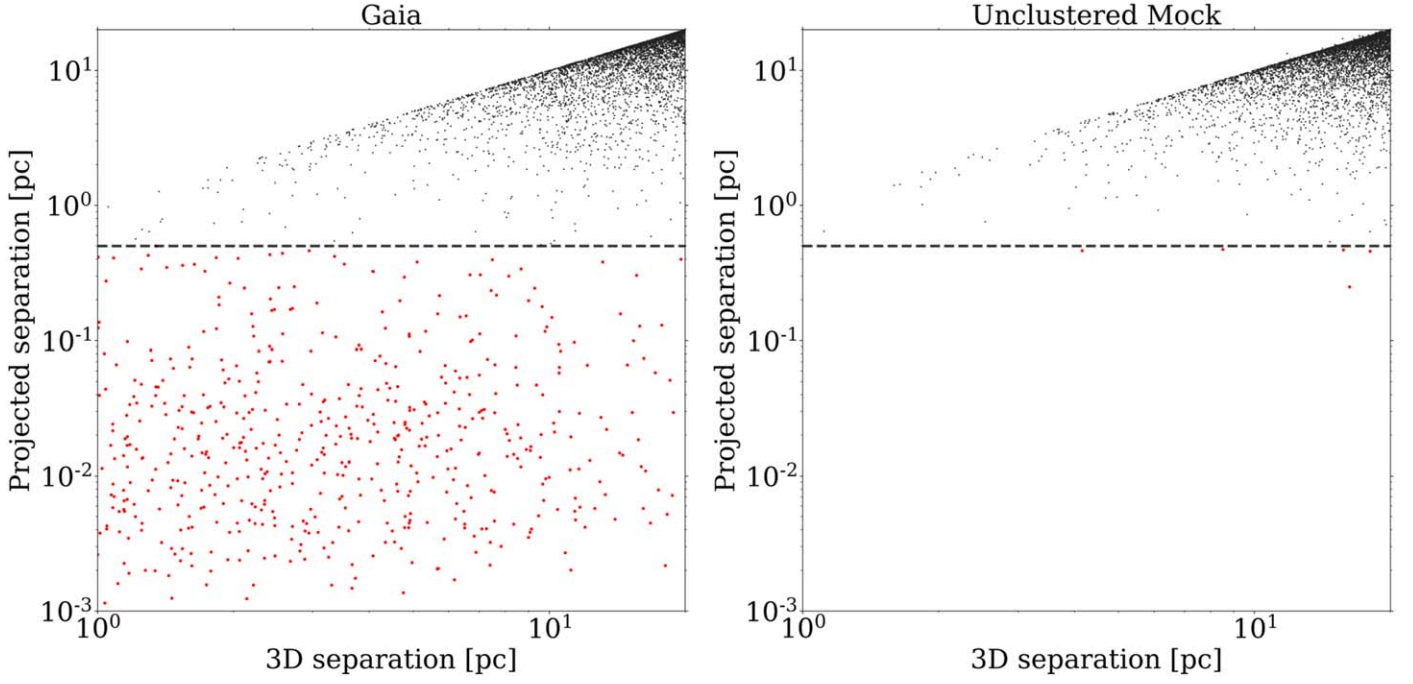


Figure 11. Projected separation and 3D separation of stars in Gaia data (left panel) and an unclustered mock (R18) without wide binaries (right panel). Points with projected separation < 0.5 pc are colored red in both panels. There is a large population of pairs with very low projected separation relative to the 3D separation in the data but not in the unclustered mock. These pairs are likely bound wide binaries with true separations of $\ll 1$ pc, whose calculated 3D separations were significantly inflated by parallax uncertainties. We only count pairs with projected separation > 0.5 pc in our TPCF calculations to minimize contamination from wide binaries.

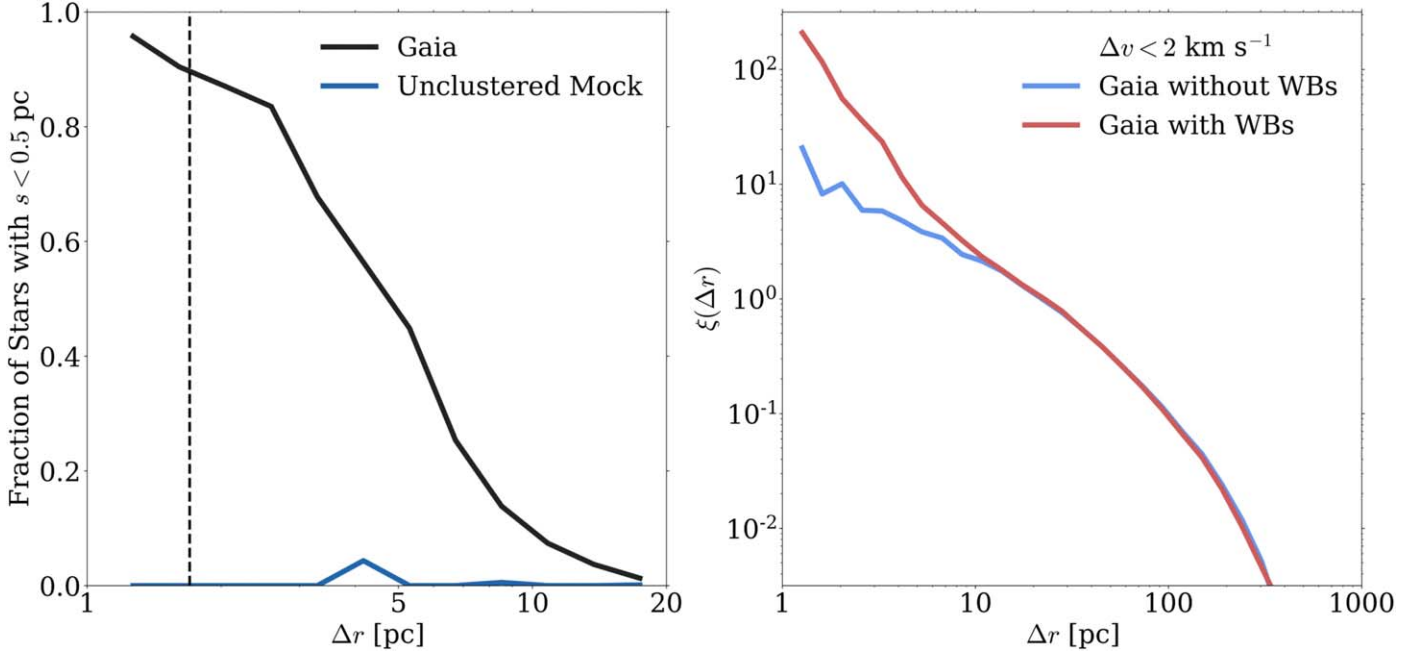


Figure 12. Fraction of pairs with projected separation < 0.5 pc in the Gaia data and in the unclustered mock (R18), and the corresponding change in the TPCF with and without the projected separation cut. Left panel: the large fraction at low Δr in the data, with no analog in the mock catalog, indicates contamination from bound wide binaries. This contamination is generally negligible for 3D separations $\Delta r \gtrsim 10$ pc. Right panel: the spatial TPCF when pairs with projected separation less than 0.5 pc are not counted. There is a large change in the clustering signal at $\Delta r < 20$ pc.

Both the data and the simulation show a fairly small change in $\xi(\Delta r)$ between $\Delta v < 4 \text{ km s}^{-1}$ and $1 \text{ km s}^{-1} < \Delta v < 4 \text{ km s}^{-1}$. $\xi(\Delta r)$ is smaller by a factor of ~ 2 at the smallest scales (perhaps due to bound wide binaries), and ~ 1.1 – 1.3 in the overdense region mentioned in J10 within the range 1 km s^{-1}

$< \Delta v < 4 \text{ km s}^{-1}$ compared to $\Delta v < 4 \text{ km s}^{-1}$. A similar decrease in $\xi(\Delta r)$ is also seen in the fiducial simulation (left panel). These results indicate that wide binaries likely constitute a smaller overdensity in phase space at small Δv and large Δr than predicted in J10.

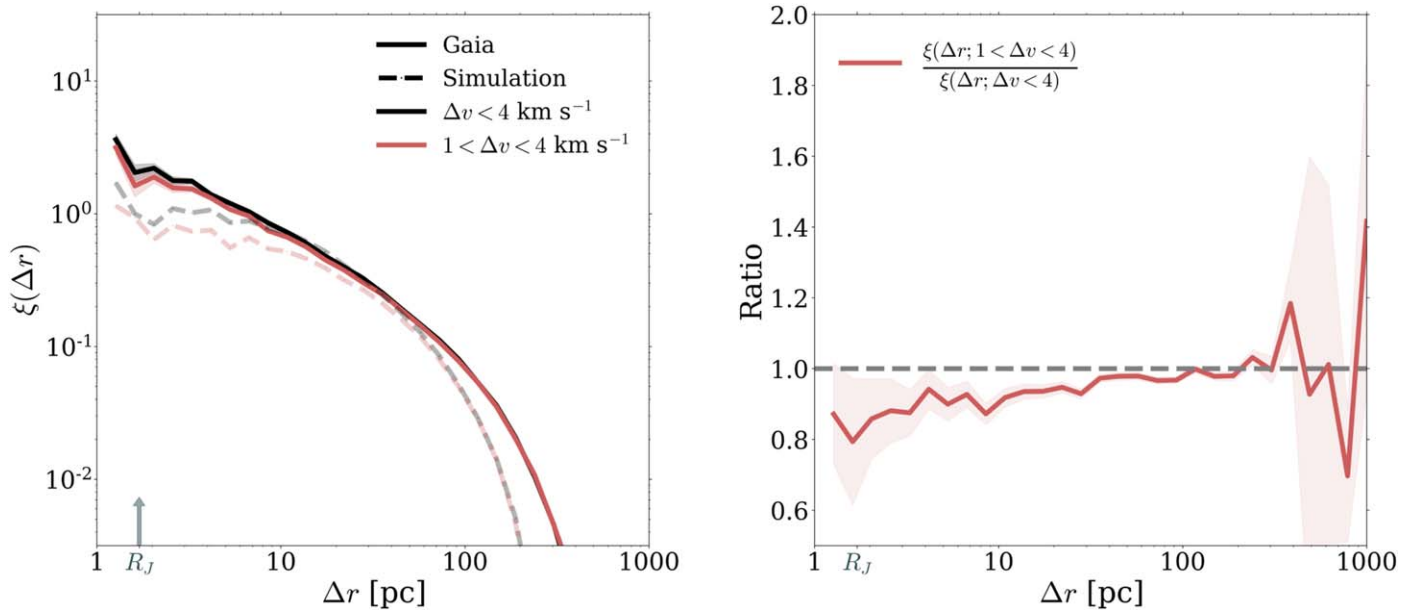


Figure 13. $\xi(\Delta r)$ for two different Δv cuts to assess the contamination from unbound wide binaries. Left panel: the spatial TPCF with these different velocity slices for the data and the simulation ($\Delta v < 4 \text{ km s}^{-1}$, $1 \text{ km s}^{-1} < \Delta v < 4 \text{ km s}^{-1}$). Note that the simulation does not include wide binaries. Right panel: the ratio between $\xi(\Delta r; 1 < \Delta v < 4)$ and $\xi(\Delta r; \Delta v < 4)$ (with Δv in km s^{-1}). If unbound wide binaries were present in the data, we would expect them to be more common at smaller Δv . The fact that the computed ratio is close to one suggests that unbound wide binaries are unlikely to affect the TPCF at a level beyond $\sim 5\%$ – 10% over the scales of interest.

ORCID iDs

Harshil Kamdar <https://orcid.org/0000-0001-5625-5342>
 Charlie Conroy <https://orcid.org/0000-0002-1590-8551>
 Yuan-Sen Ting (丁源森) <https://orcid.org/0000-0001-5082-9536>
 Kareem El-Badry <https://orcid.org/0000-0002-6871-1752>

References

- Ahumada, R., Prieto, C. A., Almeida, A., et al. 2020, *ApJS*, **249**, 3
 Alam, S., Ata, M., Bailey, S., et al. 2017, *MNRAS*, **470**, 2617
 Andrews, J. J., Chanamé, J., & Agüeros, M. A. 2017, *MNRAS*, **473**, 5393
 Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, *AJ*, **156**, 123
 Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, *A&A*, **558**, A33
 Bailer-Jones, C., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, *AJ*, **156**, 58
 Beane, A., Sanderson, R. E., Ness, M. K., et al. 2019, *ApJ*, **883**, 103
 Behnel, S., Bradshaw, R., Citro, C., et al. 2011, *CSE*, **13**, 31
 Bland-Hawthorn, J., & Gerhard, O. 2016, *ARA&A*, **54**, 529
 Bland-Hawthorn, J., Krumholz, M. R., & Freeman, K. 2010, *ApJ*, **713**, 166
 Bland-Hawthorn, J., Sharma, S., Tepper-García, T., et al. 2019, *MNRAS*, **486**, 1167
 Boubert, D., & Everall, A. 2020, *MNRAS*, **497**, 4246
 Boubert, D., Strader, J., Aguado, D., et al. 2019, *MNRAS*, **486**, 2618
 Bovy, J., Bird, J. C., Pérez, A. E. G., et al. 2015, *ApJ*, **800**, 83
 Bovy, J., Leung, H. W., Hunt, J. A., et al. 2019, *MNRAS*, **490**, 4740
 Bovy, J., Rix, H.-W., Green, G. M., Schlafly, E. F., & Finkbeiner, D. P. 2016, *ApJ*, **818**, 130
 Brown, A., Vallenari, A., Prusti, T., et al. 2018, *A&A*, **616**, A1
 Buder, S., Lind, K., Ness, M. K., et al. 2019, *A&A*, **624**, A19
 Cantat-Gaudin, T., Jordi, C., Vallenari, A., et al. 2018, *A&A*, **618**, A93
 Castro-Ginard, A., Jordi, C., Luri, X., et al. 2020, *A&A*, **635**, A45
 Choi, J., Dotter, A., Conroy, C., et al. 2016, *ApJ*, **823**, 102
 Choksi, N., & Kruijssen, J. 2021, *MNRAS*, **507**, 5492
 Conroy, C., Wechsler, R. H., & Kravtsov, A. V. 2006, *ApJ*, **647**, 201
 Cooper, A., Cole, S., Frenk, C., & Helmi, A. 2011, *MNRAS*, **417**, 2206
 Coronado, J., Rix, H.-W., Trick, W. H., et al. 2020, *MNRAS*, **495**, 4098
 Dalton, G., Trager, S., Abrams, D. C., et al. 2014, *Proc. SPIE*, **9147**, 91470L
 De Jong, R. S., Bellido-Tirado, O., Chiappini, C., et al. 2012, *Proc. SPIE*, **8446**, 84460T
 De la Fuente Marcos, R., & De la Fuente Marcos, C. 2009, *ApJ*, **700**, 436
 De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, **449**, 2604
 Deng, L.-C., Newberg, H. J., Liu, C., et al. 2012, *RAA*, **12**, 735
 Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, *ApJ*, **633**, 560
 El-Badry, K., & Rix, H.-W. 2018, *MNRAS*, **480**, 4884
 El-Badry, K., & Rix, H.-W. 2019, *MNRAS*, **482**, L139
 El-Badry, K., Rix, H.-W., Tian, H., Duchêne, G., & Moe, M. 2019, *MNRAS*, **489**, 5822
 Elmegreen, D. M., Elmegreen, B. G., Adamo, A., et al. 2014, *ApJL*, **787**, L15
 Fragkoudi, F., Katz, D., Trick, W., et al. 2019, *MNRAS*, **488**, 3324
 Freeman, K., & Bland-Hawthorn, J. 2002, *ARA&A*, **40**, 487
 Gieles, M., Zwart, S. P., Baumgardt, H., et al. 2006, *MNRAS*, **371**, 793
 Gouliermis, D. A., Elmegreen, B. G., Elmegreen, D. M., et al. 2017, *MNRAS*, **468**, 509
 Grasha, K., Elmegreen, B., Calzetti, D., et al. 2017, *ApJ*, **842**, 25
 Hawkins, K., Lucey, M., Ting, Y.-S., et al. 2020, *MNRAS*, **492**, 1164
 Houllahan, P., & Scalo, J. 1990, *ApJS*, **72**, 133
 Hunt, J. A., Hong, J., Bovy, J., Kawata, D., & Grand, R. J. 2018, *MNRAS*, **481**, 3794
 Hunter, J. D. 2007, *CSE*, **9**, 90
 Jiang, Y.-F., & Tremaine, S. 2010, *MNRAS*, **401**, 977
 Jones, E., Oliphant, T., Peterson, P., et al. 2001, SciPy: Open Source Scientific Tools for Python, <http://www.scipy.org/>
 Kagan, Y., & Knopoff, L. 1980, *GeoJ*, **62**, 303
 Kamdar, H., Conroy, C., Ting, Y.-S., et al. 2019a, *ApJ*, **884**, 173
 Kamdar, H., Conroy, C., Ting, Y.-S., et al. 2019b, *ApJL*, **884**, L42
 Kawata, D., Baba, J., Ciucă, I., et al. 2018, *MNRAS*, **479**, L108
 Khanna, S., Sharma, S., Bland-Hawthorn, J., et al. 2019, *MNRAS*, **482**, 4215
 Kollmeier, J. A., Zasowski, G., Rix, H.-W., et al. 2017, arXiv:1711.03234
 Kroupa, P. 2001, *MNRAS*, **322**, 231
 Kruijssen, J. D. 2012, *MNRAS*, **426**, 3008
 Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, *ARA&A*, **57**, 227
 Kunder, A., Kordopatis, G., Steinmetz, M., et al. 2017, *AJ*, **153**, 75
 Lada, C. J., & Lada, E. A. 2003, *ARA&A*, **41**, 57
 Lancaster, L., Belokurov, V., & Evans, N. W. 2019, *MNRAS*, **484**, 2556
 Landy, S. D., & Szalay, A. S. 1993, *ApJ*, **412**, 64
 Laporte, C.-F., Gómez, F. A., Besla, G., Johnston, K. V., & Garavito-Camargo, N. 2018, *MNRAS*, **473**, 1218
 Lindegren, L. 2018, gAIA-C3-TN-LU-LL-124, http://www.rssd.esa.int/doc_fetch.php?id=3757412

- Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., et al. 2017, *AJ*, **154**, 94
- Mao, Q., Berlind, A. A., Holley-Bockelmann, K., et al. 2015, arXiv:1507.01593
- Marchetti, T., Rossi, E., & Brown, A. 2019, *MNRAS*, **490**, 157
- Marigo, P., Girardi, L., Bressan, A., et al. 2017, *ApJ*, **835**, 77
- Martig, M., Fouesneau, M., Rix, H.-W., et al. 2016, *MNRAS*, **456**, 3655
- McQuinn, M., Hernquist, L., Zaldarriaga, M., & Dutta, S. 2007, *MNRAS*, **381**, 75
- Meingast, S., Alves, J., & Fürnkranz, V. 2019, *A&A*, **622**, L13
- Michtchenko, T. A., Lépine, J. R., Pérez-Villegas, A., Vieira, R. S., & Barros, D. A. 2018, *ApJL*, **863**, L37
- Monari, G., Famaey, B., Siebert, A., Wegg, C., & Gerhard, O. 2019, *A&A*, **626**, A41
- Oh, S., Price-Whelan, A. M., Hogg, D. W., Morton, T. D., & Spergel, D. N. 2017, *AJ*, **153**, 257
- Parmentier, G., & Kroupa, P. 2011, *MNRAS*, **411**, 1258
- Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, *JMLR*, **12**, 2825
- Peebles, P. 2001, arXiv:astro-ph/0103040
- Pérez, F., & Granger, B. E. 2007, *CSE*, **9**, 21
- Price-Jones, N., & Bovy, J. 2018, *MNRAS*, **475**, 1410
- Price-Jones, N., Bovy, J., Webb, J. J., et al. 2020, *MNRAS*, **496**, 5101
- Rasmussen, C. E. 2000, in *Advances in Neural Information Processing Systems 12* (Cambridge, MA: MIT Press), 554
- Reddick, R. M., Wechsler, R. H., Tinker, J. L., & Behroozi, P. S. 2013, *ApJ*, **771**, 30
- Rezende, D. J., & Mohamed, S. 2015, arXiv:1505.05770
- Robin, A. C., Reylé, C., Derriere, S., & Picaud, S. 2003, *A&A*, **409**, 523
- Rybizki, J., Demleitner, M., Fouesneau, M., et al. 2018, *PASP*, **130**, 074101
- Sanchez, A. G., Scóccola, C. G., Ross, A., et al. 2012, *MNRAS*, **425**, 415
- Sanderson, R. E., Wetzel, A., Loebman, S., et al. 2020, *ApJS*, **246**, 6
- Schönrich, R. 2012, *MNRAS*, **427**, 274
- Sellwood, J., Trick, W. H., Carlberg, R., Coronado, J., & Rix, H.-W. 2019, *MNRAS*, **484**, 3154
- Sharma, S., Bland-Hawthorn, J., Johnston, K. V., & Binney, J. 2011, *ApJ*, **730**, 3
- Sinha, M., & Garrison, L. 2018, *Workshop on Software Challenges to Exascale Computing* (Berlin: Springer), 3
- Sinha, M., & Garrison, L. H. 2020, *MNRAS*, **491**, 3022
- Tian, H.-J., El-Badry, K., Rix, H.-W., & Gould, A. 2019, *ApJS*, **246**, 4
- Ting, Y.-S., Conroy, C., & Goodman, A. 2015a, *ApJ*, **807**, 104
- Ting, Y.-S., Conroy, C., & Rix, H.-W. 2015b, *ApJ*, **816**, 10
- Ting, Y.-S., Freeman, K. C., Kobayashi, C., De Silva, G. M., & Bland-Hawthorn, J. 2012, *MNRAS*, **421**, 1231
- Ting, Y.-S., & Rix, H.-W. 2019, *ApJ*, **878**, 21
- Trick, W. H., Coronado, J., & Rix, H.-W. 2019, *MNRAS*, **484**, 3291
- Van Der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, **13**, 22
- Ward, J. L., Kruijssen, J. D., & Rix, H.-W. 2020, *MNRAS*, **495**, 663
- Wechsler, R. H., Zentner, A. R., Bullock, J. S., Kravtsov, A. V., & Allgood, B. 2006, *ApJ*, **652**, 71
- Weinberg, M. D., Shapiro, S. L., & Wasserman, I. 1987, *ApJ*, **312**, 367
- Yoo, J., Chaname, J., & Gould, A. 2004, *ApJ*, **601**, 311
- Zamolodchikov, A. B. 1991, *NuPhB*, **348**, 619