



Two-point Correlation Function Studies for the Milky Way: Discovery of Spatial Clustering from Disk Excitations and Substructure

Austin Hinkel^{1,2} , Susan Gardner¹ , and Brian Yanny³ ¹ Department of Physics and Astronomy, University of Kentucky, Lexington, KY 40506-0055, USA; ahinkel@coloradocollege.edu² Department of Physics, Colorado College, Colorado Springs, CO 80903, USA³ Fermi National Accelerator Laboratory, Batavia, IL 60510, USA

Received 2022 August 12; revised 2022 October 19; accepted 2022 October 20; published 2023 January 6

Abstract

We introduce a two-particle correlation function (2PCF) for the Milky Way, constructed to probe spatial correlations in the orthogonal directions of the stellar disk in the Galactic cylindrical coordinates of R , ϕ , and z . We use this new tool to probe the structure and dynamics of the Galaxy using the carefully selected set of solar neighborhood stars ($d \lesssim 3$ kpc) from Gaia Data Release 2 that we previously employed for studies of axial symmetry breaking in stellar number counts. We make additional, extensive tests, comparing to reference numerical simulations, to ensure our control over possibly confounding systematic effects. Supposing either axial or north–south symmetry, we divide this data set into two nominally symmetric sectors and construct the 2PCF, in the manner of the Landy–Szalay estimator, from the Gaia data. In so doing, working well away from the midplane region in which the spiral arms appear, we have discovered distinct symmetry-breaking patterns in the 2PCF in its orthogonal directions, thus establishing the existence of correlations in stellar number counts alone at subkiloparsec length scales for the very first time. In particular, we observe extensive wavelike structures of amplitude greatly in excess of what we would estimate if the system were in a steady state. We study the variations in these patterns across the Galactic disk, and with increasing $|z|$, and we show how our results complement other observations of non-steady-state effects near the Sun, such as vertical asymmetries in stellar number counts and the Gaia snail.

Unified Astronomy Thesaurus concepts: [Milky Way Galaxy \(1054\)](#); [Galaxy structure \(622\)](#)

1. Introduction

The long-range nature of the gravitational force distinguishes the statistical description of stars in the Galaxy from the terrestrial systems commonly studied. Indeed the gravitational “charge” cannot be shielded, so the stars accelerate, smoothly, through the force field dominated by the Galaxy’s most distant stars. Thus they can be modeled as a collisionless fluid with distribution functions that are assumed to be separable, making the stars uncorrelated, even if correlations can be expected to exist (Binney & Tremaine 2008). For example, stars can be formed in spatially and temporally correlated ways, as discussed in the case of the “solar family” (Bland-Hawthorn & Freeman 2004; Bland-Hawthorn et al. 2010), where we note the review of Krumholz et al. (2019). It is thought, though, that these correlations cannot survive for long in the Galactic environment (Gieles et al. 2006; Lada & Lada 2003). However, with the accumulating evidence for apparent wavelike or perturbed features in the Galactic disk (Antoja et al. 2018; Bennett & Bovy 2018; Ferguson et al. 2017; Widrow et al. 2012; Williams et al. 2013; Yanny & Gardner 2013), and with the observed pattern of axial- and vertical symmetry breaking in the stellar number counts speaking to their origin in non-steady-state effects (Gardner et al. 2020; Hinkel et al. 2020), we believe the time is now ripe for the development of sensitive probes of the structure and dynamics of our Galaxy in the region within a few kiloparsecs of the Sun. In this paper, we show how the two-point correlation function (2PCF) can be computed in a fully data-driven way using symmetry-breaking

effects in the stellar number counts and then proceed to make its direct determination using Gaia Data Release 2 (DR2) data (Gaia Collaboration et al. 2016; Lindegren et al. 2018). We do not employ Gaia Data Release 3 (DR3) because we estimate the improvement in the average precision of the parallax measurements for our particular subset of DR2 stars (Hinkel et al. 2020) to be less than a few percent, so we think an update is not warranted, even if other regions of the Galaxy see marked improvements in parallax precision and measurements of 3D velocity data (Gaia Collaboration et al. 2022). We refer to Section 5 for an overview of the features of our data set. We also note Nelson & Widrow (2022) for a study of local stellar velocity correlations within a Gaussian process model.

The 2PCF, or the pair correlation function, has been used broadly and extensively as a sensitive probe of the structure of matter, with applications, e.g., to terrestrial studies in condensed matter (Goodstein 2014) and nuclear (Blatt & Weisskopf 2010) physics and to cosmology (Peebles 1980). In the last case it has been used as a probe of the large-scale structure of the universe, through studies of galaxy–galaxy clustering. The supposed isotropic nature of the cosmos implies that the 2PCF in this case appears in terms of the scalar separation d of any two galaxies in three-dimensional space, with the 2PCF capturing the likelihood, given one galaxy, that another will be found a distance d away (Hauser & Peebles 1973; Peebles 1973, 1993).

In order to develop a suitable 2PCF for studies in our Galaxy, additional considerations enter: certainly the Galaxy is not isotropic, and the 2PCF, studied here in its stars, can depend on the independent components of the displacement vector between any two of them. We develop this line of thinking in this paper, starting in Section 2. We note that studies of the 2PCF in the Galaxy do exist (Cooper et al. 2011;



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Kamdar et al. 2021; Lancaster et al. 2019; Mao et al. 2015), though they are focused on somewhat different questions and are also limited in different ways. An assumption of spherical symmetry has carried over to those studies. In Cooper et al. (2011) and Lancaster et al. (2019) the 2PCF is used to study substructure in the stellar halo of the Milky Way, with the later, larger study made as a probe of the Galaxy’s accretion history. There, observations of RR Lyrae stars from the CRTS (Drake et al. 2009) survey, with 31,301 objects over Galactocentric radii from ≈ 2 to 90 kpc, and from the PanSTARRS1 (Chambers et al. 2016; Flewelling et al. 2020) survey, with 44,208 objects over Galactocentric radii from ≈ 0.5 to 150 kpc, are used to determine the 2PCF via the Landy–Szalay (LS; Landy & Szalay 1993) correlator. The appearance of substructure is inferred through comparison of the data to a reference theoretical background distribution. Extensive study has shown the LS method to be a superior choice (Keihanen et al. 2019; Lancaster et al. 2019), which we and Kamdar et al. (2021) also employ, though we note Wall & Jenkins (2012) for discussion of the broader possibilities. In Kamdar et al. (2021) spatial and kinematic clustering of the stars in the Galactic disk is studied using a sample of 1.7×10^6 stars with 6D phase-space information within 1 kpc of the Sun from Gaia DR2 data. The construction of a suitable reference distribution is essential to the determination of the 2PCF, and in the Galaxy its construction is challenging (Kamdar et al. 2021; Mao et al. 2015). Kamdar et al. (2021) employ Dirichlet process Gaussian mixture models to that end. With that in place, they find evidence of clustering of approximately comoving stars up to large spatial and kinematic scales, i.e., up to 300 pc and 15 km s $^{-1}$. These results agree well at small scales with their simulations built to model star formation (Kamdar et al. 2019a), and their co-natal correlations (Kamdar et al. 2019b), in the Galaxy.

Although the relatively bright radial velocity data set employed by Kamdar et al. (2021) is appropriate to their study of young stars close to the Galactic midplane, the stars in Gaia DR2 (and DR3) with radial velocity information within the Galaxy have sampling biases that could impact the outcomes of our study (Katz et al. 2019, 2022), and thus a larger-scale 2PCF analysis requires a different approach. Instead, we exploit the exceptionally (>99%) complete Gaia DR2 sample of Hinkel et al. (2020) to effect a data-driven 2PCF analysis of up to some 11.7 million stars, employing spatial information only. Here we split that data set into subsamples that are related by symmetry, either axial, i.e., in the plane of the Galactic disk, or north/south, for our 2PCF analysis, allowing us to focus on the spatial correlations in our sample consequent to the existence of spatial symmetry breaking effects. For reference, we note that the discovery of a vertical wavelike asymmetry in the stellar number counts using Sloan Digital Sky Survey (SDSS; York et al. 2000) data (Widrow et al. 2012; Yanny & Gardner 2013), and in the vertical velocity distribution from RAVE data (Williams et al. 2013), with the observed effect changing in different regions of the Galactic disk (Ferguson et al. 2017), hints at a complex vertical landscape in the local Galactic potential. Moreover, in addition to this picture of planar vertical waves in the Milky Way, Antoja et al. (2018) have shown that a “phase-space spiral” pattern in z - v_z phase space exists, with its very visibility suggesting it is a fairly recent development in the Galaxy’s past (Antoja et al. 2018). Antoja et al. (2018) interpret this snail-shaped pattern using a heuristic, anharmonic

oscillator model in order to derive an approximate date for the perturbation that is thought to have caused it. They find a timescale of approximately 300–900 Myr, which appears to be consistent with the Sagittarius Dwarf’s last passage through the disk (Purcell et al. 2011), which may be a driver of the vertical asymmetries seen (Gómez et al. 2012; Widrow et al. 2012). We note, moreover, that corrugations across the disk have been observed (Bland-Hawthorn & Tepper-García 2021; Xu et al. 2015). Such effects may be due to the last impact of the Sagittarius Dwarf Galaxy, modulated by the influence from the Large Magellanic Cloud (LMC; Laporte et al. 2018). The Galaxy also has a warp, both in its gas (Burke 1957; Kerr et al. 1957; Levine et al. 2006a) and in its stars (Drimmel & Spergel 2001; Freudenreich et al. 1994; Poggio et al. 2018). The LMC may also be warping the disk (Gardner et al. 2020; Kerr 1957; Weinberg & Blitz 2006). Corrugations of a similar nature are observed in Milky Way-like galaxies (Gómez et al. 2021), and those in the Milky Way may arise from the superposition of different wavelike effects (Bland-Hawthorn & Tepper-García 2021); we refer to Gardner et al. (2021) for further discussion.

We view the observed wavelike patterns in stellar number counts with position and spirals in position and velocity space as likely having shared origins, even if the particular observational data sets employed in the two sorts of studies are quite different—and we will use “wavelike patterns” to refer to position-space structures henceforth. The corrugations, or radial waves, noted by Xu et al. (2015), become ring- or shell-like structures still farther from the Sun and Galactic midplane. These stellar overdensities could have an accretion origin, via a tidally disrupted satellite galaxy (Searle & Zinn 1978), but they could also have come from stars ejected from the disk (Xu et al. 2015), with further observational studies supporting that latter, novel interpretation (Li et al. 2017; Price-Whelan et al. 2015; Sheffield et al. 2018). This, in turn, has set the stage for broader studies tying the global response of the Galactic disk to the local disturbances we have noted. For example, stellar kinematics measurements from Gaia have been used to map asymmetric features of the Galactic disk (Drimmel et al. 2022; Katz et al. 2018). We also note evidence that its asymmetries may be tied to its large-scale spiral structure (Eilers et al. 2020; Levine et al. 2006b; Poggio et al. 2021b), as well as various theoretical developments: namely, that of a systematic theoretical framework for the study of dynamical phase spirals (Banik et al. 2022) and of numerical simulations of the collision of a Sagittarius-like dwarf galaxy with the Milky Way. The latter body of work acts to discern its local and global kinematics signatures (Hunt et al. 2021), the vertical response of the disk (Poggio et al. 2021a), and the existence of various snail-like features (Gandhi et al. 2022) that can emerge in such a context.

Recent studies have relied heavily on the use of stellar velocity information to tease out important dynamical effects, and our neglect of such information might therefore seem a limitation. Rather, we emphasize that our studies are complementary, in that we gain in sensitivity not only through the sheer size and quality, in terms of completeness and precision parallax information, of the data set we have chosen, but also through our reliance on symmetry-breaking effects to boost the visibility of subtle non-steady-state effects. Here we develop a 2PCF analysis of the stars near the Sun to give sharpened insights into the structure and nature of the

perturbations on the stars within 3 kpc of the Sun. As observations with Gaia continue, we expect that studying the 2PCF for stars with particular velocity selections away from the midplane region, but loosely in the manner of Kamdar et al. (2021), will yield even more discriminating insights.

We conclude our introduction with a brief sketch of the sections to follow. In Section 2 we develop the theory of the 2PCF for galactic dynamics, computing it under steady-state conditions with and without spherical symmetry, to establish that visible effects necessarily come from non-steady-state effects. In Section 3 we describe the extensive control studies we have made to ensure that the selections we make of the Gaia DR2 data can be robustly interpreted in terms of physical, rather than systematic, effects. Because we are appreciative that studies of the nature of dark matter spur interest in structure at the shortest distance scales (Buckley & Peter 2018; Gardner et al. 2021), we carefully delineate our systematic limitations in resolving small-scale structures in Section 4. In Section 5 we discuss our data selection, based on Hinkel et al. (2020), and note the control we have over observational systematic errors. Finally, we report all of our 2PCF results in Section 6, consider their possible origin in Section 7, and offer a final summary and outlook in Section 8.

2. Theory

2.1. The 2PCF in Steady State

An isolated galaxy is described by a distribution function (DF) in its stars in six-dimensional phase space: $f(\mathbf{v}, \mathbf{x}, t)$, with different DFs for different stellar populations possible. A DF can be self-consistently determined by the solution of the collisionless Boltzmann, or Vlasov, and Poisson equations, where we emphasize that the Vlasov equation itself emerges only if correlations between the stars are neglected (Binney & Tremaine 2008). In steady state, such a galaxy with a stellar disk is expected to be axially symmetric with respect to rotations about an axis, through its center of mass, perpendicular to the plane of the disk, and thus is also reflection symmetric about the galactic midplane (An et al. 2017; Schutz et al. 2018). We have determined that in our own Galaxy, however, that even if axial symmetry is very nearly conserved, reflection symmetry can be markedly broken (Gardner et al. 2020; Hinkel et al. 2020), implying that the Galaxy is not isolated and/or not in steady state. We interpret the small axial symmetry breaking we have found in our carefully selected sample of Gaia DR2 stars as arising, in part, from the net torque exerted on our sample within the Galaxy by the massive LMC/SMC system, yet the differences in axial symmetry breaking we find, comparing north with south, are much larger still. Thus we think our results are particularly indicative of the presence of non-steady-state effects (Gardner et al. 2020; Hinkel et al. 2020). This and the appearance of striking wavelike features in stellar number counts north and south of the Galactic plane (Bennett & Bovy 2018; Widrow et al. 2012; Yanny & Gardner 2013) suggest that the stars are likely correlated as well, possibly on many different length scales (Kamdar et al. 2021). To explore this concretely, we revisit the derivation of the Vlasov equation itself: we return to the Bogoliubov, Born, Green, Kirkwood, and Yvon (BBGKY) hierarchy, which comes from the analysis of Liouville’s equation in the presence of pairwise forces, relating the time-evolution of the s -particle distribution function f_s , to the $(s+1)$ -particle distribution

function f_{s+1} (Gardner et al. 2021). Consequently, the s -particle distribution function is not simply proportional to $(f)^s$; rather, we introduce (Thorne & Blandford 2017)

$$f_2(\mathbf{v}_1, \mathbf{x}_1, \mathbf{v}_2, \mathbf{x}_2, t) = f_1(\mathbf{v}_1, \mathbf{x}_1, t)f_1(\mathbf{v}_2, \mathbf{x}_2, t)(1 + \xi_{12}), \quad (1)$$

where ξ_{12} is the 2PCF, with the same arguments as that of the two-particle distribution function. In general, the 2PCF can either enhance or suppress the joint probability of finding a particle in a phase-space volume $d^3\mathbf{x}_1 d^3\mathbf{v}_1$ and another in $d^3\mathbf{x}_2 d^3\mathbf{v}_2$: the probability of finding one is now no longer independent of the probability of finding the other. The BBGKY hierarchy links f_2 to f_3 , where

$$\begin{aligned} f_3(\mathbf{v}_1, \mathbf{x}_1, \mathbf{v}_2, \mathbf{x}_2, \mathbf{v}_3, \mathbf{x}_3, t) \\ = f_1(\mathbf{v}_1, \mathbf{x}_1, t)f_1(\mathbf{v}_2, \mathbf{x}_2, t)f_1(\mathbf{v}_3, \mathbf{x}_3, t) \\ \times (1 + \xi_{12} + \xi_{13} + \xi_{23} + \xi_{123}), \end{aligned} \quad (2)$$

with $\xi_{ij} \equiv \xi_{ij}(\mathbf{x}_i, \mathbf{v}_i, \mathbf{x}_j, \mathbf{v}_j, t)$. If we neglect the possibility of the *three-particle* correlation function ξ_{123} , we can extract a single differential equation for the two-particle correlation functions. Exact expressions for ξ_{12} exist in simplified contexts, as in, e.g., Kirkwood et al. (1950). Of particular interest to us is the case of electrons in an unmagnetized, thermalized plasma, as long-range forces are present. There the protons can be treated as a stationary background, so that the distribution functions are in electron degrees of freedom only, with correlations arising from electron–electron interactions—and external forces can be neglected (Thorne & Blandford 2017). We adapt this analysis to the particular case of stars in steady state within a few kiloparsecs of the Sun, although the two problems differ in fundamental aspects, as we shall see. Since the gravitational interaction between any two stars does not depend on their velocities, we might also expect the 2PCF in this case to be velocity-independent and depend only on the interparticle separation: $x_{ij} \equiv |\mathbf{x}_i - \mathbf{x}_j|$; we explore this possibility through explicit calculation. To our knowledge, this is the first time such an extension has been explored, and we lay out the connection to statistical physics carefully. Returning to the BBGKY hierarchy, we note the lowest two equations for an N -particle system are of the form (Kardar 2007)

$$\begin{aligned} \frac{\partial f_1}{\partial t} - \nabla_{\mathbf{x}_1} \Phi_{\text{ext}} \cdot \nabla_{\mathbf{v}_1} f_1 + \mathbf{v}_1 \cdot \nabla_{\mathbf{x}_1} f_1 \\ = \int d^3\mathbf{x}_2 d^3\mathbf{v}_2 \nabla_{\mathbf{x}_1} \Phi_{12} \cdot \nabla_{\mathbf{v}_1} f_2, \end{aligned} \quad (3)$$

and

$$\begin{aligned} \frac{\partial f_2}{\partial t} - \nabla_{\mathbf{x}_1} \Phi_{\text{ext}} \cdot \nabla_{\mathbf{v}_1} f_2 - \nabla_{\mathbf{x}_2} \Phi_{\text{ext}} \cdot \nabla_{\mathbf{v}_2} f_2 \\ + \mathbf{v}_1 \cdot \nabla_{\mathbf{x}_1} f_2 + \mathbf{v}_2 \cdot \nabla_{\mathbf{x}_2} f_2 \\ - [\nabla_{\mathbf{v}_1} f_2 \cdot \nabla_{\mathbf{x}_1} \Phi_{12} + \nabla_{\mathbf{v}_2} f_2 \cdot \nabla_{\mathbf{x}_2} \Phi_{12}] \\ = \int d^3\mathbf{x}_3 d^3\mathbf{v}_3 [\nabla_{\mathbf{v}_1} f_3 \cdot \nabla_{\mathbf{x}_1} \Phi_{13} + \nabla_{\mathbf{v}_2} f_3 \cdot \nabla_{\mathbf{x}_2} \Phi_{23}], \end{aligned} \quad (4)$$

where $\Phi_{\text{ext}}(\mathbf{x}_i)$ is the external gravitational potential and $\Phi_{ij}(\mathbf{x}_{ij})$ is the two-body gravitational potential, where we employ dimensions of energy per mass throughout (Binney & Tremaine 2008). In this context it is conventional to normalize $f_1(\mathbf{x}, \mathbf{v}, t)$ so that $\int d^3\mathbf{x} d^3\mathbf{v} f_1 = \mathcal{M}$, the total mass of the system,

rather than unity. We thus adjust the right-hand side of Equations (3) and (4) accordingly in what follows. The collisionless Boltzmann equation follows by neglecting correlations, replacing f_2 with $(f_1)^2$ in Equation (3), yielding

$$\frac{\partial f_1}{\partial t} - \nabla_{\mathbf{x}_1} \Phi_0 \cdot \nabla_{\mathbf{v}_1} f_1 + \mathbf{v}_1 \cdot \nabla_{\mathbf{x}_1} f_1 = 0 \quad (5)$$

with

$$\begin{aligned} \Phi_0 &= \Phi_{\text{ext}} + \frac{1}{\mathcal{M}} \int d^3\mathbf{x}_2 d^3\mathbf{v}_2 f_1(\mathbf{x}_2, \mathbf{v}_2, t) \Phi_{12} \\ &= \Phi_{\text{ext}} + \frac{1}{\mathcal{M}} \int d^3\mathbf{x}_2 \rho(\mathbf{x}_2, t) \Phi_{12}. \end{aligned} \quad (6)$$

The mass density ρ and the effective potential Φ_0 , which are both regarded as smooth distributions (Binney & Tremaine 2008) because we assume $N \gg 1$, can be determined self-consistently by solving Equation (5) and the Poisson equation

$$\nabla^2 \Phi_0 = 4\pi G \rho, \quad (7)$$

with the equilibrium solution satisfying $\partial f_1 / \partial t = 0$. Turning to the analysis of the f_2 equation, Equation (4), and assuming f_1 satisfies Equations (5) and (7), with f_1 and ξ_{ij} in steady state, we find

$$\begin{aligned} &f_1(\mathbf{x}_1, \mathbf{v}_1) f_1(\mathbf{x}_2, \mathbf{v}_2) [-\nabla_{\mathbf{x}_1} \Phi_0 \cdot \nabla_{\mathbf{v}_1} - \nabla_{\mathbf{x}_2} \Phi_0 \\ &\quad \cdot \nabla_{\mathbf{v}_2} + \mathbf{v}_1 \cdot \nabla_{\mathbf{x}_1} + \mathbf{v}_2 \cdot \nabla_{\mathbf{x}_2}] \xi_{12} \\ &- [\nabla_{\mathbf{x}_1} \Phi_{12} \cdot \nabla_{\mathbf{v}_1} + \nabla_{\mathbf{x}_2} \Phi_{12} \cdot \nabla_{\mathbf{v}_2}] f_1(\mathbf{x}_1, \mathbf{v}_1) f_1(\mathbf{x}_2, \mathbf{v}_2) (1 + \xi_{12}) \\ &= \frac{1}{\mathcal{M}} \int d^3\mathbf{x}_3 d^3\mathbf{v}_3 [\nabla_{\mathbf{x}_1} \Phi_{13} \cdot \nabla_{\mathbf{v}_1} + \nabla_{\mathbf{x}_2} \Phi_{23} \cdot \nabla_{\mathbf{v}_2}] \\ &\quad \times f_1(\mathbf{x}_1, \mathbf{v}_1) f_1(\mathbf{x}_2, \mathbf{v}_2) f_1(\mathbf{x}_3, \mathbf{v}_3) (\xi_{12} + \xi_{13} + \xi_{23}) \end{aligned} \quad (8)$$

— noting we have set $\xi_{123} = 0$. In what follows we analyze two simple cases. Although neither one offers a realistic description of the Milky Way galaxy, we find our analysis of value in that it shows clearly that the parametric behavior of the 2PCF in steady state is grossly different from what we observe in the data. In the first we suppose that f_1 is that of an infinite, isotropic, spherically symmetric system, for which

$$f_1(\mathbf{v}) = \frac{\rho_0}{(2\pi\sigma^2)^{3/2}} \exp\left(-\frac{v^2}{2\sigma^2}\right), \quad (9)$$

where ρ_0 is a constant. This is not a self-consistent model because only Equation (5) is satisfied; here we commit the “Jeans swindle” and set $\Phi_0 = 0$ (Binney & Tremaine 2008). In the second case we suppose that f depends on the vertical coordinate only, giving a slab geometry. For concreteness we employ the self-consistent thin-disk model of Spitzer (1942) in this latter case. In both cases we assume that ξ_{ij} is independent of velocity, and we justify this assertion a posteriori, from the form of the solutions we find. Since $\Phi_{ij} = -GM/|\mathbf{x}_i - \mathbf{x}_j|$, where M is the mass of an ϵ -sphere centered on $\mathbf{x}_j$, $M = \int_{|\mathbf{x}'| \leq |\mathbf{x}_j| + \epsilon} d^3\mathbf{x}' \rho_0(\mathbf{x}')$, nominally the mass of a star. Thus

$\nabla_{\mathbf{x}_i} \Phi_{ij} = -\nabla_{\mathbf{x}_j} \Phi_{ij}$ for any $i \neq j$, and Equation (8) becomes

$$\begin{aligned} &(\mathbf{v}_1 \cdot \nabla_{\mathbf{x}_1}) \xi_{12} + (\mathbf{v}_2 \cdot \nabla_{\mathbf{x}_2}) \xi_{12} + \frac{1}{\sigma^2} (1 + \xi_{12}) (\mathbf{v}_1 - \mathbf{v}_2) \cdot \nabla_{\mathbf{x}_1} \Phi_{12} \\ &= -\frac{\rho_0}{\mathcal{M}\sigma^2} \int d^3\mathbf{x}_3 (\xi_{13} + \xi_{23}) [(\mathbf{v}_1 \cdot \nabla_{\mathbf{x}_1}) \Phi_{13} + (\mathbf{v}_2 \cdot \nabla_{\mathbf{x}_2}) \Phi_{23}]. \end{aligned} \quad (10)$$

The velocity dependence is explicit, and the equation must hold for any choice of $\mathbf{v}_1$ and $\mathbf{v}_2$. Setting $\mathbf{v}_2 = 0$ for simplicity, and assuming $\xi_{12} \ll 1$, we determine

$$\nabla_{\mathbf{x}_1} (\sigma^2 \xi_{12} + \Phi_{12}) = -\frac{\rho_0}{\mathcal{M}} \int d^3\mathbf{x}_3 (\xi_{13} + \xi_{23}) \nabla_{\mathbf{x}_1} \Phi_{13}. \quad (11)$$

Since $\nabla_{\mathbf{x}_1} \Phi_{13}$ changes sign if $\mathbf{x}_1 \gtrless \mathbf{x}_3$, we see that if ξ_{13} depends only on x_{13} its contribution to the integral vanishes. Assuming this and with $\nabla_{\mathbf{x}_1}^2 \Phi_{12} = 4\pi GM \delta^{(3)}(\mathbf{x}_{12})$, we finally have

$$\nabla_{\mathbf{x}_1}^2 \xi_{12} + \frac{4\pi\rho_0 GM}{\mathcal{M}\sigma^2} \xi_{12} = -\frac{4\pi GM}{\sigma^2} \delta^{(3)}(\mathbf{x}_{12}). \quad (12)$$

This reveals that ξ_{12} is the response of the system to the interparticle interaction sourced at $\mathbf{x}_2$. Introducing $\lambda_G \equiv \sqrt{\mathcal{M}\sigma^2/4\pi GM\rho_0}$, we first solve this equation for $x_{12} \neq 0$ to yield

$$\xi_{12}(x_{12}) = A \frac{\cos(x_{12}/\lambda_G)}{x_{12}} + B \frac{\sin(x_{12}/\lambda_G)}{x_{12}}, \quad (13)$$

where A and B are arbitrary constants. Through consideration of the divergence theorem in the $\rho_0 \rightarrow 0$ limit we find

$$\xi_{12}(x_{12}) = \left(\frac{GM}{\sigma^2}\right) \frac{\cos(x_{12}/\lambda_G)}{x_{12}}, \quad (14)$$

The attractive nature of the gravitational interaction dictates the form of our solution. In contrast, the unmagnetized, thermal electron gas has a 2PCF of form $\xi_{12} \propto \exp(-x_{12}/\lambda_D)/x_{12}$, where λ_D is the Debye length (Thorne & Blandford 2017). It is useful to contrast our result with the encounter operator (Binney & Tremaine 2008), namely,

$$\Gamma[f(\mathbf{x}_1, \mathbf{v}_1, t)] \equiv \frac{\mathcal{M}}{M} \int d^3\mathbf{x}_2 d^3\mathbf{v}_2 \nabla_{\mathbf{x}_1} \Phi_{12} \cdot \nabla_{\mathbf{v}_1} \xi_{12}, \quad (15)$$

which drives the change of the (one-body) DF,

$$\frac{df}{dt} = \Gamma[f]. \quad (16)$$

Employing our result in Equation (14) we see that the right-hand side does indeed evaluate to zero, so that our 2PCF is compatible with a steady-state limit. It is worthwhile to study the explicit length scales associated with our solution. We see that Equation (14) can attain a value of $\mathcal{O}(1)$ or larger if and only if $x_{12} < GM/\sigma^2$. Supposing σ to be given by the vertical velocity dispersion of the disk, 25 km s⁻¹ (Minchev et al. 2014), $M \approx M_\odot = 1.988 \times 10^{30}$ kg, and $G = 6.674 \times 10^{-11}$ m³ kg⁻¹ s⁻² (Zyla et al. 2020), we determine that $GM/\sigma^2 \approx 2.1 \times 10^{11}$ m = 6.9×10^{-6} pc. Thus for the length scales of interest to us, the BBGKY hierarchy should give a reasonable estimate of the size of the correlation effects in steady state, and ξ_{12} is also extremely small. We note that λ_G can be written as $\lambda_G = \sqrt{\sigma^2 V_{\text{eff}}/4\pi GM}$, where V_{eff} is the effective volume per star. Referring to the extremely complete Gaia DR2 sample of

Hinkel et al. (2020), within roughly 3 kpc of the Sun, we note that 11.7 million stars occupy a volume of at least 16 kpc^3 to yield $\lambda_G \gtrsim \sqrt{f_{\text{col/mag}}} (4.0 \text{ kpc})$, where $f_{\text{col/mag}}$ is the fraction of the total number of stars that fall within the chosen color and magnitude cuts. Querying the Gaia DR2 database, we determine $f_{\text{col/mag}} \approx 0.3$, so that $\lambda_G \gtrsim 2.2 \text{ kpc}$. With $x_{12} = 0.20 \text{ kpc}$, e.g., we find that Equation (14) evaluates to $\xi_{12} = 3.4 \times 10^{-8}$. Thus it would appear that any nonzero values of ξ_{12} that we would be able to observe cannot come from steady-state effects. The spherical symmetry of f_1 dictates that ξ_{12} in Equation (14) can only depend on the scalar x_{12} . To consider how our results might change with the symmetries of the problem, we now turn to our second example, that of a slab geometry, with f_1 depending on the vertical energy E_z only, i.e.,

$$f(z, v) = \frac{\rho_0}{\sqrt{2\pi\sigma^2}} \exp(-E_z/\sigma^2), \quad (17)$$

where $E_z = v^2/2 + \Phi_0(z)$, $\Phi_0(z) = 2\sigma^2 \ln(\cosh(z/2z_d))$, and $\rho_0 = \sigma^2/8\pi G z_d^2$ (Spitzer 1942). In this case, Φ_{ij} represents the potential at z_i due to an infinitely thin and uniform sheet of surface mass density Σ at z_j , so that $\Phi_{ij} = 2\pi G \Sigma |z_i - z_j|$. Returning to Equation (8), we note that if ξ_{12} is independent of velocity then the terms in Φ_0 do not contribute, and we find that ξ_{12} satisfies

$$\begin{aligned} & (v_1 \partial_{z_1} + v_2 \partial_{z_2}) \xi_{12} + \frac{1}{\sigma^2} (1 + \xi_{12}) (v_1 - v_2) \partial_{z_1} \Phi_{12} \\ &= -\frac{1}{\Sigma \sigma^2} \int dz_3 \rho(z_3) (\xi_{12} + \xi_{13} + \xi_{23}) [v_1 \partial_{z_1} \Phi_{13} + v_2 \partial_{z_2} \Phi_{23}], \end{aligned} \quad (18)$$

where $\rho(z) = \rho_0 \text{sech}^2(z/2z_d)$ and the surface mass density is $\Sigma = \int_{-\infty}^{\infty} dz \rho(z) = 4\rho_0 z_d$. Supposing that ξ_{ij} depends in some manner on $|z_i - z_j|$ as well, we see that

$$\int dz_3 \rho(z_3) \xi_{i3} \partial_{z_3} \Phi_{i3} \approx 0; \quad \int dz_3 \rho(z_3) \partial_{z_3} \Phi_{i3} \approx 0 \quad (19)$$

for $i = 1, 2$ if z_1, z_2 are in the vicinity of the Galactic midplane. Thus

$$\begin{aligned} & (v_1 \partial_{z_1} + v_2 \partial_{z_2}) \xi_{12} + \frac{1}{\sigma^2} (1 + \xi_{12}) (v_1 - v_2) \partial_{z_1} \Phi_{12} \\ &= -\frac{1}{\Sigma \sigma^2} \int dz_3 \rho(z_3) [v_1 \xi_{23} \partial_{z_1} \Phi_{13} + v_2 \xi_{13} \partial_{z_2} \Phi_{23}]. \end{aligned} \quad (20)$$

Proceeding as in the previous case, we find

$$\partial_{z_1} \left(\xi_{12} + \frac{1}{\sigma^2} \Phi_{12} \right) = -\frac{1}{\Sigma \sigma^2} \int dz_3 \rho(z_3) \xi_{23} \partial_{z_1} \Phi_{13}, \quad (21)$$

and since

$$\partial_{z_1}^2 \Phi_{12} = 4\pi G \Sigma_{z_2} \delta(z_1 - z_2), \quad (22)$$

where we introduce Σ_{z_2} as the surface mass density of a thin and uniform sheet at z_2 , with $\Sigma_{z_2} = \Delta \rho(z_2)$ and a parameter Δ of $\mathcal{O}(z_d)$, we find

$$\partial_{z_1}^2 \xi_{12} + \frac{4\pi G \rho(z_1) \Sigma_{z_1}}{\Sigma \sigma^2} \xi_{12} = -\frac{4\pi G \Sigma_{z_1}}{\sigma^2} \delta(z_1 - z_2) \quad (23)$$

for z_1, z_2 close to the Galactic midplane, noting $\Sigma_{z_1} \ll \Sigma$. Since explicit z_1 dependence appears in the second term of

Equation (23), which might be expected because the matter distribution is not isotropic, we see that ξ_{12} cannot depend on $|z_1 - z_2|$ alone. Thus we introduce $z_{12} = z_1 - z_2$ and $Z = (z_1 + z_2)/2$ to find

$$\begin{aligned} & (\partial_{z_{12}}^2 + \partial_{z_{12}} \partial_Z + \frac{1}{4} \partial_Z^2) \xi_{12} + \frac{4\pi G \rho(z_{12}/2 + Z) \Sigma_{z_{12}/2+Z}}{\Sigma \sigma^2} \xi_{12} \\ &= -\frac{4\pi G \Sigma_{z_{12}/2+Z}}{\sigma^2} \delta(z_{12}), \end{aligned} \quad (24)$$

and finally

$$\partial_{z_{12}}^2 \xi_{12} + \frac{4\pi G \Delta \rho_0^2}{\Sigma \sigma^2} \xi_{12} = -\frac{4\pi G \rho_0 \Delta}{\sigma^2} \delta(z_{12}), \quad (25)$$

near the Galactic midplane. With $\lambda_{\tilde{G}} \equiv \sqrt{\Sigma \sigma^2 / 4\pi G \Delta \rho_0^2}$, we solve this equation for $z_{12} \neq 0$ to yield

$$\xi_{12}(z_{12}) = A \cos(z_{12}/\lambda_{\tilde{G}}) + B \sin(z_{12}/\lambda_{\tilde{G}}), \quad (26)$$

where A and B are arbitrary constants. Considering the $z_d \rightarrow \infty$ limit with Δ fixed, we see that only the term with A survives and thus we estimate

$$\xi_{12}(z_{12}) \approx A \left(1 - \frac{z_{12}^2}{2\lambda_{\tilde{G}}^2} \right) \quad (27)$$

in the midplane region—and we note that ξ_{12} is symmetric under $z_1 \leftrightarrow z_2$ exchange. In this case we are unable to fix the strength of the homogeneous solution through consideration of the source term and a suitable Gaussian surface, but we observe that variations in ξ_{12} are determined by $\lambda_{\tilde{G}}$, which we evaluate to be $\lambda_{\tilde{G}} \gtrsim \sqrt{\sigma^2 / \pi G \rho_0} = \sqrt{\sigma^2 4z_d / \pi G \Sigma}$ for $\Delta \lesssim z_d$. Using the surface mass density in stars at $|z| = 1.1 \text{ kpc}$ from Bovy & Rix (2013), namely $38 \pm 4 M_{\odot} \text{ pc}^{-2}$, and a scale height of 280 pc (Bovy 2015) we determine $\lambda_{\tilde{G}} \gtrsim 1.1 \text{ kpc}$, which is within a factor of 2 of our estimate in the purely isotropic case. We thus conclude that the spatial variations associated with ξ_{12} in the steady-state case are roughly comparable to the physical dimensions of our stellar sample, though in the regions in which Equation (27) would be valid the spatial correlations we would be able to observe cannot come from steady-state effects. In this regard we emphasize that our analytic solutions for ξ_{12} , namely, Equations (14) and, (27), do not break the underlying symmetries present in each case. This also supports our notion that symmetry breaking speaks to the appearance of non-steady-state effects. In the next section we determine how the 2PCF can be determined from observations.

2.2. Evaluation of the 2PCF: Connecting Theory to Observations

In this paper, we wish to access the 2PCF of the Milky Way in an entirely data-driven way. We do so by adapting the 2PCF analysis in galactic number counts familiar from the study of cosmic large-scale structure (Peebles 1993) and thus begin by reviewing that setting, referring to Peebles (1993) for all details. The joint probability of finding two galaxies, which is assumed to be stationary, at separation r centered within volume elements dV_1 and dV_2 , respectively, is $dP_2 = n^2(1 + \xi(r/r_0))dV_1 dV_2$, where the probability of finding one

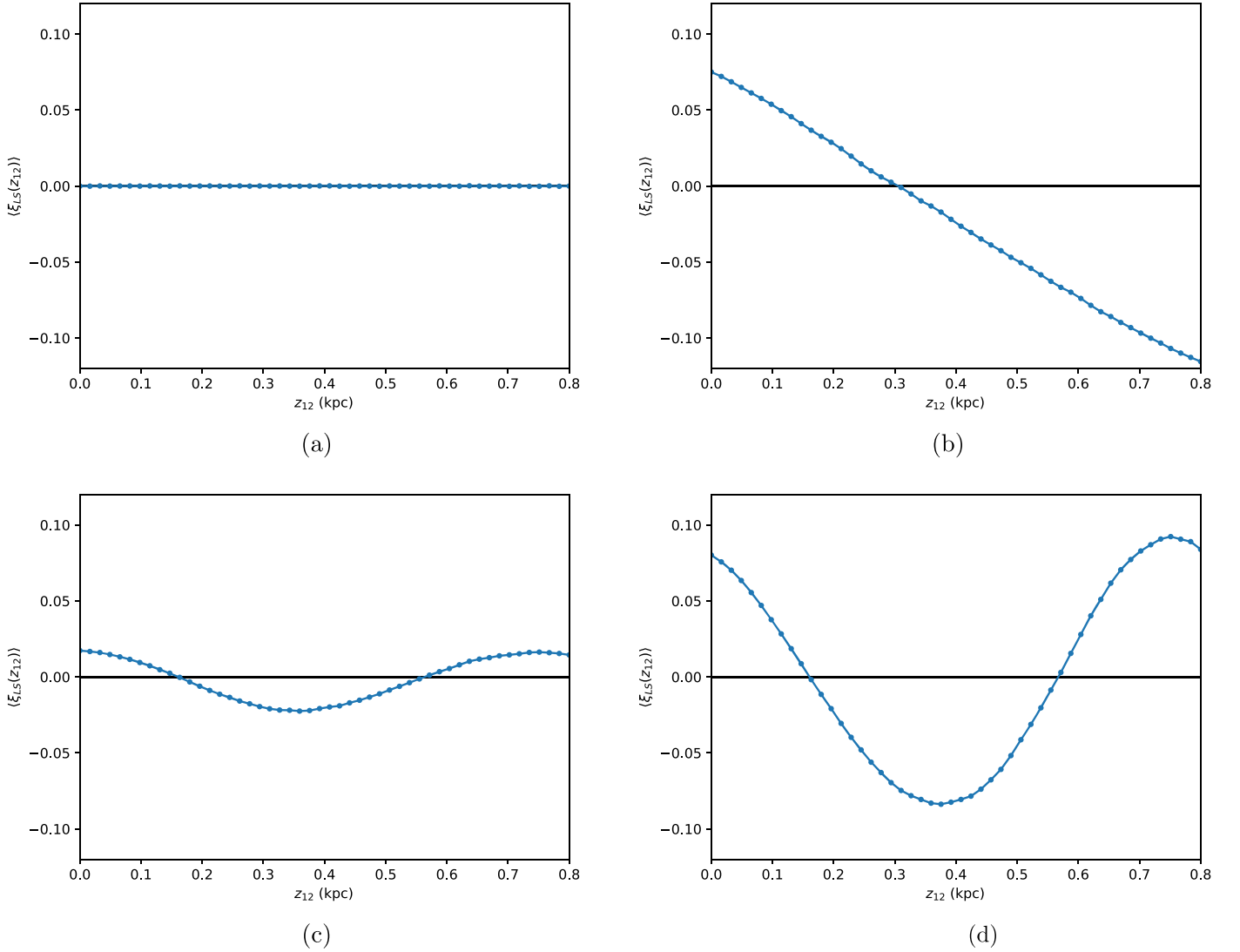


Figure 1. (a) The north vs. south z2PCF comparing two models drawn from identical distribution functions. (b) The north vs. south z2PCF comparing two models drawn from distribution functions with different scale heights. In this case, one model has a scale height of $z_s = 280$ pc, while the other has a scale height of $z_s = 420$ pc. (c) The z2PCF comparing a model with vertical density waves against a smooth model. (d) The north vs. south z2PCF comparing a model with vertical density waves in the north against the same model with vertical density waves in the south. The wave is modeled as in Equation (34) and thus the antisymmetric nature of the wave results in more significant correlations in the north vs. south analysis than in the comparison of the density wave model against a smooth model. All panels implement the following cuts: $7.6 \text{ kpc} < R < 8.2 \text{ kpc}$, $176^\circ < \phi < 184^\circ$, and $0.2 \text{ kpc} < |z| < 2.0 \text{ kpc}$.

galaxy is $dP_1 = ndV$. In this context $\xi(r/r_0)$, the two-point correlation function, depends on r_0 , a characteristic clustering length, which is determined from observations. If the universe is homogeneous, we can convert this quantity to an angular correlation function that can be directly determined from the data by including a selection function S_i that determines the likelihood that a galaxy i at some distance is bright enough to be detected. With this, the joint probability becomes

$$dP_2 = n^2 d\Omega_1 d\Omega_2 \int r_1^2 dr_1 r_2^2 dr_2 (1 + \xi(r_{12}/r_0)) S_1 S_2, \quad (28)$$

and thus we have

$$dP_2 = N^2 d\Omega_1 d\Omega_2 (1 + w(\theta)), \quad (29)$$

with

$$w(\theta) = \frac{\int r_1^2 dr_1 r_2^2 dr_2 \xi(r_{12}/r_0) S_1 S_2}{\left(\int r^2 dr dS\right)^2}, \quad (30)$$

where $r_{12} = (r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta)^{1/2}$ and N is the mean number of galaxies per steradian. In order to assess $w(\theta)$ from the observational data, the LS estimator (Landy & Szalay 1993) is employed, though other choices are possible (Wall & Jenkins 2012). In this method the data D with d points are compared to a reference model R with r points, which is composed of randomly distributed galaxies, and three separate histograms are constructed: RR , DD , and DR . Each histogram counts the number of pairs of stars at separations of θ to $\theta + d\theta$, and DD counts these pairs using the data, RR counts them within the reference model, and DR counts the number of cross-correlation pairs, to yield

$$w_{\text{LS}}(\theta) = \frac{RR(\theta) - 2DR(\theta) + DD(\theta)}{RR(\theta)}, \quad (31)$$

where the histograms must be suitably normalized.

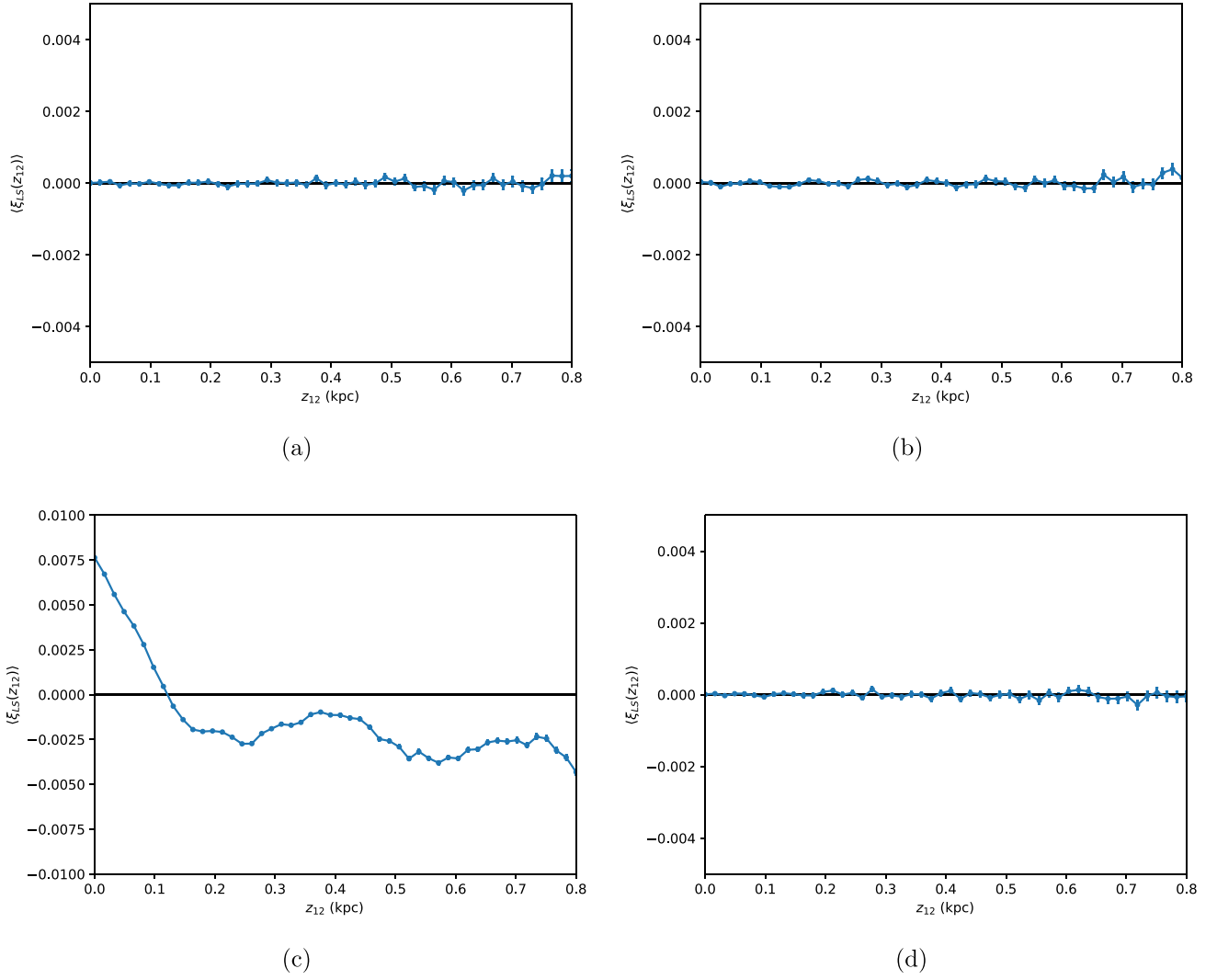


Figure 2. (a) The north vs. south z -separation 2PCF comparing two models drawn from identical distribution functions in the region $7.6 \text{ kpc} < R < 8.2 \text{ kpc}$, $176^\circ < \phi < 184^\circ$, and $0.2 \text{ kpc} < |z| < 2.0 \text{ kpc}$, including $|b| > 30^\circ$ cuts and the set of LMC/SMC cuts mentioned in the text. (b) The same scenario as panel (a), but with a 20% mismatch in the number of stars between the two models. (c) The north vs. south $z2$ PCF comparing two models drawn from a distribution function that accounts for $z_\odot$, resulting in a geometry mismatch, north and south, because the heliocentric cuts no longer emanate from midplane. (d) The north vs. south $z2$ PCF comparing two models drawn from a distribution function that accounts for $z_\odot$, but with Galactocentric geometry chosen in such a way that the l and b cuts are avoided. In this particular example, $7.9 \text{ kpc} < R < 8.3 \text{ kpc}$, $179^\circ < \phi < 181^\circ$, and $0.3 \text{ kpc} < |z| < 2.0 \text{ kpc}$, and thus the $|b| > 30^\circ$ cuts and the LMC/SMC cuts do not impact the geometry of the sample, which results in no geometry mismatch from heliocentric cuts.

Segueing to our Milky Way studies, we first note that our selected Gaia DR2 set is exceptionally complete in our selected color and magnitude windows (Hinkel et al. 2020), and we have not applied a selection function as a result. We note that complete velocity information is only available for stars that are brighter than those in our sample, and the gravitational interaction does not depend on velocity, so we consider the density-density correlation function and do not restrict the relative velocities of the stars in any way. The matter distribution in the Milky Way is not spherically symmetric, so we expect the joint probability of finding two stars at separation $\mathbf{x}_{12} \equiv \mathbf{x}_1 - \mathbf{x}_2$ to depend on the projection onto the Cartesian vectors $\hat{e}_x$, $\hat{e}_y$, and $\hat{e}_z$ as well as on $\mathbf{x}_1$, or $\mathbf{x}_2$, itself. For the stars of our sample within a fixed region of the sky, we expect the sample-averaged 2PCF to be determined by

$$\langle \xi_{12}(|\hat{e}_Q \cdot \mathbf{x}_{12}|) \rangle \equiv \langle \rho(\mathbf{x}_i) \rho(\mathbf{x}_i + \mathbf{x}_{12}) \rangle_{\text{fixed } |\hat{e}_Q \cdot \mathbf{x}_{12}|}, \quad (32)$$

for a fixed choice of $Q \in x, y, z$, where the average is determined by summing over the coordinate $\mathbf{x}_i$ of each of the stars in our selected sample. We employ standard practice, so that $\mathbf{x}$ points along the anti-center line toward the Galactic center, and $\mathbf{y}$ points in the direction of $\phi < 180^\circ$. In concrete terms we use the LS estimator (Landy & Szalay 1993), with histograms formed by the histograms that count the number of pairs of stars separated by some distance $q = |\hat{e}_Q \cdot \mathbf{x}_{12}|$ to $q + dq$ with $dq = d|\hat{e}_Q \cdot \mathbf{x}_{12}|$ (and henceforth $z_{12} > 0$, for example). In usual practice, DD would count pairs within our Gaia data set, RR would count pairs within a reference theoretical model, sampled statistically, and DR would count the cross-correlation pairs between the two data sets to yield

$$\langle \xi_{LS}(q_i) \rangle = \frac{RR(q_i) - 2DR(q_i) + DD(q_i)}{RR(q_i)}, \quad (33)$$

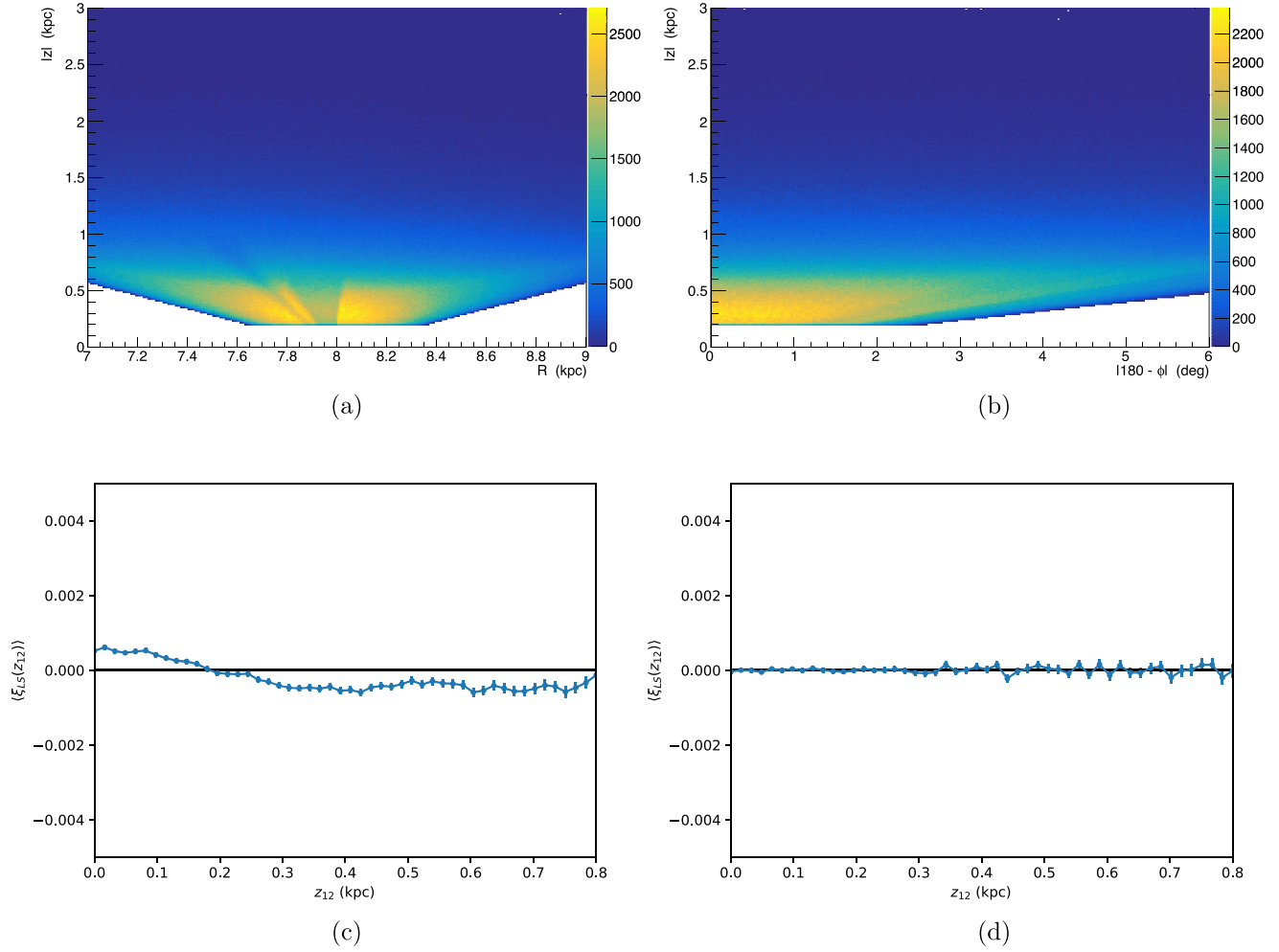


Figure 3. (a) The R - z projection of the data set, illustrating the regions free of l , b cut effects. (b) The ϕ - z projection of the data set, illustrating the regions free of l , b cut effects. (c) A heliocentric analysis of the 2PCF, so that it avoids the geometry mismatch problem of the Galactocentric analysis shown in Figure 2(c). Here, $7.6 \text{ kpc} < R < 8.2 \text{ kpc}$, $176^\circ < \phi < 184^\circ$, $0.2 \text{ kpc} < |z| < 2.0 \text{ kpc}$. In this case the small, nonzero correlation arises due to the sampling of slightly different regions of the Galactic distribution function because the $z_\odot$ offset is *not* included. (d) A heliocentric analysis similar to panel (c), but for higher z : $1.2 \text{ kpc} < |z| < 3.0 \text{ kpc}$. Here the smallest correlations one can probe are significantly smaller for regions well above the midplane, as explained in the text.

where we have averaged over the coordinates not fixed by q_i . As noted previously, the RR , DD , and DR histograms must be normalized so that they have unit areas (Wall & Jenkins 2012). This normalization accounts for the potentially different numbers of stars in real and mock samples, as well as the fact that there are more cross-correlation pairs than pairs within one data set.

Here we differ from usual practice, as we wish to exploit the near symmetry of our chosen data set to tease out the 2PCF arising from non-steady-state effects. That is, we choose “ D ” and “ R ” to be two distinct selections of the Gaia data sample of Hinkel et al. (2020), which are related by either axial or north-south reflection symmetry, rather than comparing the observational data with a model-dependent mock catalog. For example, one can examine the difference in structure between the northern hemisphere of the Galaxy and (a reflection of) the southern hemisphere, so long as the geometries of the two regions are identical. This is enabled through the approximate reflection symmetry of the Galaxy, to yield a new and sensitive probe of the symmetries the Milky Way.

3. Control Studies: Interpreting the 2PCF

Regardless of whether the LS estimator is used in a traditional manner (i.e., comparing model versus data) or to probe symmetry-breaking effects as in this work (i.e., comparing data versus data), it is critical to understand how the LS estimator reveals structure in order to (i) avoid being tricked by the effects of the various geometric cuts, the $z_\odot$ offset, etc., and (ii) to understand how effects, such as north-south symmetry breaking, observed in the data vis-à-vis asymmetries in the one-body density across the Galactic plane, are manifested in a two-point correlation study. In this section we demonstrate that comparing two samples drawn from suitably chosen models, with explicit symmetries in place, suffices as probes of both of these issues, because an explicitly symmetric model can serve as either a model to compare with data *or* a simulation of reflected data. To this end, we have simulated a number of different scenarios and detail them in this section.

In Figure 1(a), the LS estimator is shown for z -separation distances, and it compares two data sets drawn from an identical distribution function, of form $\text{sech}^2(z/2z_s)$. This

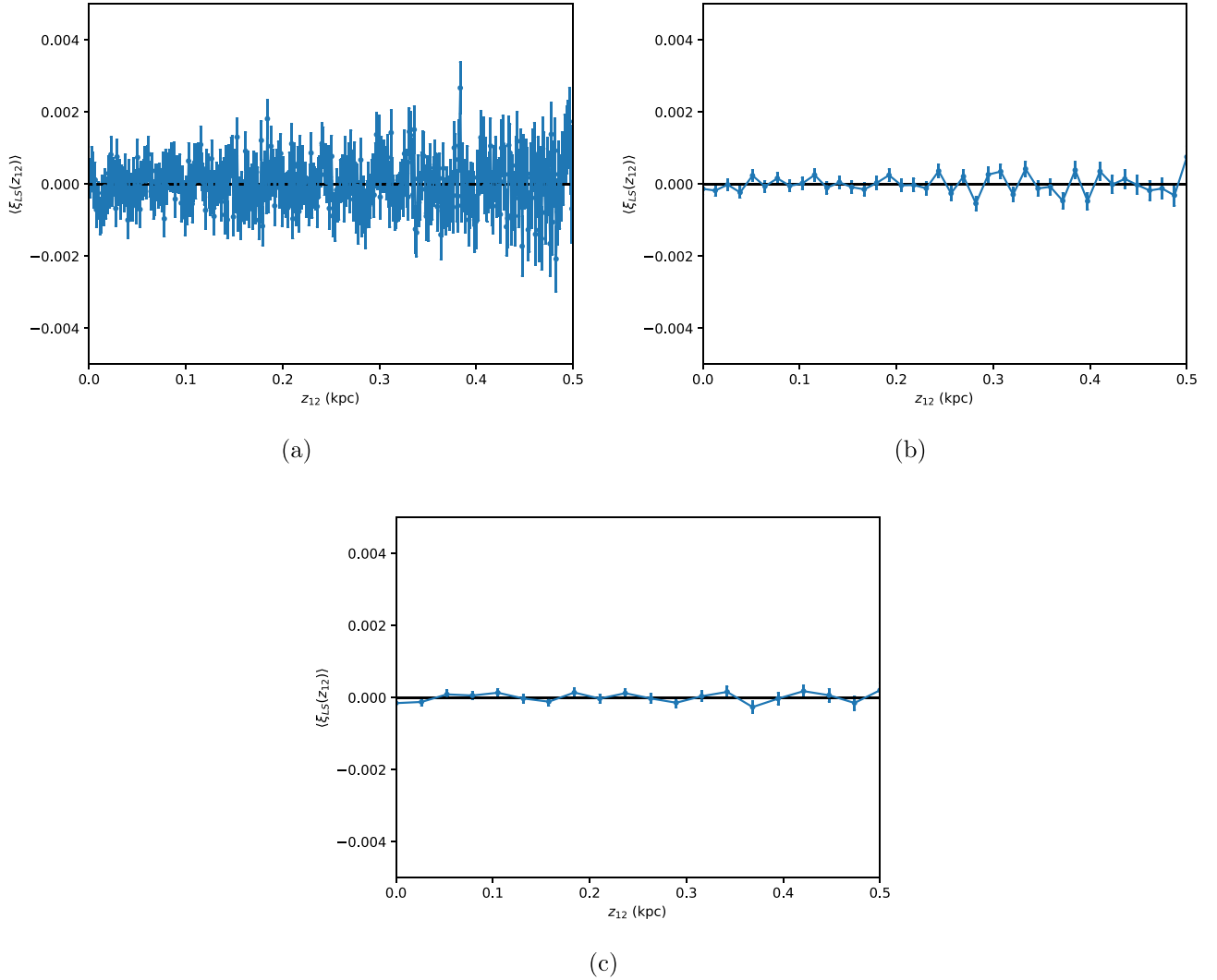


Figure 4. A comparison of two model Galaxies drawn from an identical distribution, as noted in the text, with bin widths of (a) 2 pc, (b) 20 pc, and (c) 40 pc. As the models were drawn from the same distribution, there should be no excess structure, yet binning too finely exposes the study to the finite density limitations described in the text. The approximate limiting length scale for this example simulation is $\lambda_{\text{lim}} \approx 20$ pc, and thus the apparent structure in panel (a) is not significant, while both panels (b) and (c) correctly indicate a lack of significant structure.

control test illustrates the case where no structure exists, because the two samples are drawn from identical distributions, and indeed the LS estimator is consistent with zero. A result with structure is shown in Figure 1(b), wherein two data sets are compared with markedly different scale heights. The first data set (D) has a scale height of $z_s = 280$ pc (Bovy 2015), while the other (R) has an intentionally inflated scale height of $z_s = 420$ pc, and thus an excess of structure is found at small scales and a dearth of structure at larger scales, resulting in the slanted estimator shown.

If instead of a smooth, hyperbolic-secant-squared distribution function, we use a DF with structure embedded in it, the LS estimator will then pick out the characteristic scales of that particular structure. For example, by introducing a toy-model of a vertical wave into one data set (D) as in the equation

$$n(R, z) = e^{-R/R_s} \text{sech}^2\left(\frac{z}{2z_s}\right) (1 + 0.2 \times \sin(8z)) \quad (34)$$

and comparing this against smoother data drawn from a hyperbolic-secant-squared distribution function (R), both with

the same $z_s = 280$ pc, the LS estimator will indicate an excess of structure corresponding to the maxima of the density waves, and a dearth of structure near the trough of the density waves, as in Figure 1(c). An interesting feature of an antisymmetric structure like that of the vertical waves found in Widrow et al. (2012), Yanny & Gardner (2013), and Bennett & Bovy (2018) is that any north–south comparison will result in increased significance in a structure search. In other words, when data in the north (DD) are compared against a reflection of the data in the south (RR), the antisymmetric nature of the vertical waves results in the peaks in the north lining up with the troughs in the south, and thus the LS estimator strongly highlights this structural difference, as depicted in Figure 1(d).

While helpful for illustrative purposes, the toy models studied in Figure 1 are missing a key consideration. In selecting a reliable data set as free from observational artifacts as possible, we have implemented various cuts on *heliocentric* longitude and latitude. However, because the Sun is not truly situated on the Galactic midplane, any analysis in Galactocentric coordinates will necessarily run into problems caused

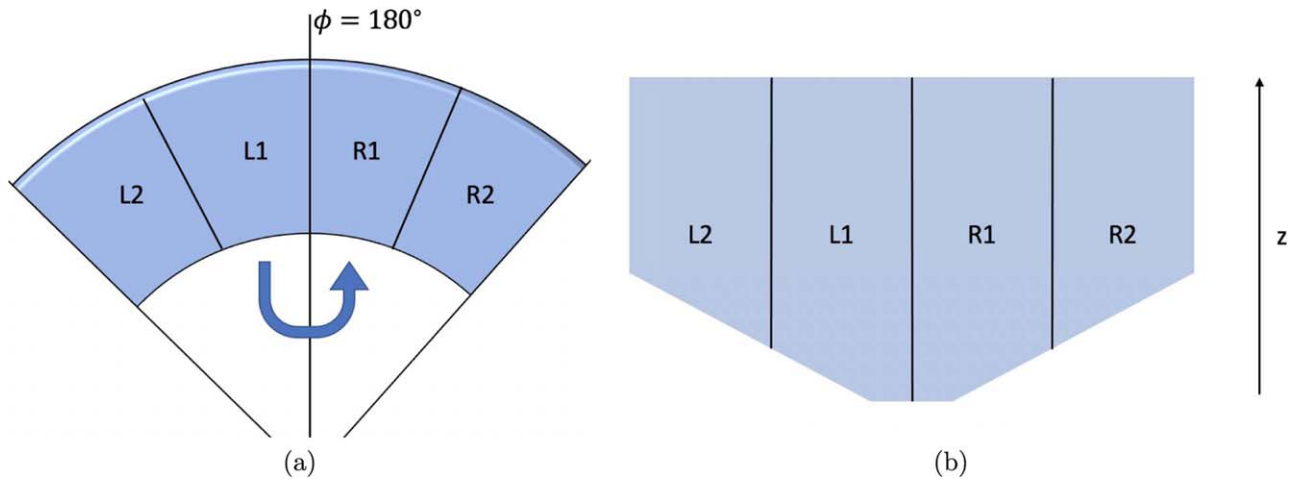


Figure 5. An illustration of the axial symmetry we exploit in order to build a data vs. data LS correlator. (a) A schematic of our data set projected onto the midplane as viewed from the north of the Galaxy. Using the $\phi = 180^\circ$ ray as the mirror line, the reflection of the data from L1 is compared with R1, while the data from L2 are compared with R2. (b) The northern half of the same model projected onto the y - z plane at $x \approx -8$ kpc, the Sun’s location, as viewed from the Galactic center. In this view of the rough schematic, the vertical structure of a reflection of L1 across the $\phi = 180^\circ$ ray can be compared to the vertical structure of R1, as the structure examined in z could change with ϕ . An analogous method is used for the reflection symmetry about $z = 0$. Here we have chosen regions in R and ϕ such that $z_\odot$ can be neglected with little consequence, as we have noted in Figure 3(c).

by a geometry mismatch. To better illustrate this concept, let us consider Figure 2. In panel (a), we repeat the control test of Figure 1(a), but now include cuts on latitude ($|b| > 30^\circ$) as well as the LMC and SMC cuts of Gardner et al. (2020) and Hinkel et al. (2020). In this case, a $z_\odot$ shift has not been applied, and thus these toy models implicitly assume $z_\odot = 0$ pc. It is clear from panel (a) that l and b cuts alone do not bias the LS estimator if $z_\odot = 0$ pc. Indeed, even if the samples have an egregiously large mismatch in the number of stars, as in Figure 2(b), the LS estimator still takes into account both the post-cut geometry and normalization considerations, correctly resulting in no indication of structure. However, if $z_\odot \neq 0$, where we convert heliocentric to Galactocentric coordinates via $z \rightarrow z + z_\odot$, a Galactocentric analysis will incur substantial geometric effects in the LS estimator, as seen in Figure 2(c). In this case, an offset of $z_\odot = 20$ pc (Bennett & Bovy 2018) has been included in the models, such that heliocentric cuts on l and b effectively emanate from a region that is not coincident with the Galactic midplane. This difference in geometries, north and south, is falsely registered in the LS estimator as structure, even though the models are identical in all other regards. Thus, we must be extremely careful to avoid geometric differences caused by a combination of nonzero $z_\odot$ and cuts on l and b .

There are two potential remedies for this issue. First, it is possible to select data such that the l and b cuts are avoided entirely. In these cases, an analysis in Galactocentric coordinates works without trouble, as the heliocentric cuts simply do not enter the geometry in question. To illustrate this point, Figure 2(d) shows how samples drawn from identical distributions that avoid the l and b cuts result in an LS estimator that is consistent with zero. In this particular example, we have raised the minimum value of z to which we probe, and have restricted the region of R and ϕ as well. While this is certainly a viable workaround for the geometry mismatch issue stemming from heliocentric cuts in a Galactocentric coordinate system, the procedure limits the regions that we can explore. To illustrate this, the regions of the data set impacted by the l and b cuts are shown via their projections in the R - z and z - ϕ planes in Figures 3(a) and 3(b), respectively. As the latter shows, increasing the minimum z to which we probe is an effective

way of avoiding the l and b cuts, at the cost of cutting out the region with the highest number of stars. Regions at low z are still available to explore with this method in limited regions of R , as shown in the former panel.

An alternative remedy for the heliocentric cut mismatch issue is to conduct the analysis in heliocentric coordinates. That is, we assume $z_\odot = 0$ pc. While not strictly true, we show in Figure 3(c) that any false correlations due to this incorrect choice of $z_\odot$ are small. In particular, the correlations we find in this case are $\xi_{LS} < 0.001$, and they appear because we are sampling slightly different regions of the Galaxy’s distribution function in the north and in the south. To wit, the 20 pc $z_\odot$ “shift” is small compared to the scale height of ~ 280 pc, and so the resulting effect is small, but it is nonetheless important vis-à-vis the smallest significant correlation we can probe.

Any structure will need to exceed this background correlation of about 0.0005, as shown in Figure 3(c), to be physically significant. Moreover, for high $|z|$, the disk’s density profile falls off approximately as an exponential decay function. Because a shift in an exponential function is equivalent to an overall normalization factor, and because the LS estimator takes into account differences in normalization, a $z_\odot$ shift will not matter in this case. In general, though, other distributions would retain some effects from a $z_\odot$ shift. We note for regions high above the midplane that neglecting $z_\odot$ results in negligible correlations for an otherwise identical model-model comparison, indicating that the stars in our simple model do fall off approximately exponentially at high $|z|$, as shown in Figure 3(d). Turning to the Gaia data, in regions not impacted by our heliocentric cuts, we have explicitly verified that the numerical choice of $z_\odot$ does not affect the outcomes of our study; we regard this check as complementary to the one of Figure 3(c).

4. Systematic Limitations in Resolving Small-scale Structures

Although the Gaia data in the solar neighborhood are remarkably complete, boasting an extraordinary number ($>10^9$) of stars, it is nevertheless the case that the distribution

of stars is finite, so that the stars, on average, are separated by some typical length scale λ_{lim} . This density-derived limitation fundamentally limits the smallest length scale to which our analysis can reliably probe.

As an illustrative example of this limitation, let us examine two mock data sets drawn from the same distribution, of Gaussian form, centered on $z=0$ with standard deviation $\sigma_z = 280$ pc, each with 100,000 stars. These simulations were made in a small volume with a height of 800 pc and the zPCF was computed to assess structural difference between the two mock galaxies. As the galaxies were drawn from identical distributions, any structure must be due to systematic effects. The data are binned in three different ways with 400, 40, and 20 bins as depicted in Figures 4(a)–(c), respectively. With a vertical extent of 800 pc, these histograms of the LS estimator have bin widths of 2 pc, 20 pc, and 40 pc, respectively. Turning to Figure 4(a), we see that binning the data too finely results in false structure, while binning more coarsely as in panels (b) and (c) correctly indicates a lack of structure, matching expectations for two mock galaxies drawn from identical distributions. Again, because the limiting scale length increases with decreasing density of stars, we expect some kind of volume-per-particle dependence. Indeed, Figure 4 suggests the limiting length scale is approximately the cube root of the volume per particle, or more conveniently,

$$\lambda_{\text{lim}} \approx \frac{L_i}{N^{1/3}}, \quad (35)$$

where L_i is the length of the sample in the direction in which we analyze the one-dimensional LS estimator and N is the total number of stars. For example, the illustrative simulation in Figure 4 had a height of 800 pc and 100,000 stars, yielding a limiting scale of about 20 pc. The ultrafine binning in panel (a) is much smaller than this limiting length scale, and we observe false structure from overbinning the data. As panels (b) and (c) are each plotted with a bin size of at least λ_{lim} , we see no significant structure.

The finite density effect we have considered is critical to interpreting the results of our LS estimator analysis. While bin widths will sometimes be finer than the limiting scale in what follows, only structures with scales exceeding λ_{lim} can be significant. Future data releases may possibly enable the study of still smaller scales, possibly even at the sub-solar system level through consideration of planetary-scale objects.

In addition to the finite density effects we have noted, the precision of our 2PCF study is limited by uncertainties in Gaia’s parallax measurements and thus by their distance determinations. Errors in distance assessments will likely “smear” or “blur” structure, and thus they are unlikely to result in false structure. No evidence for appreciable direction-dependent parallax errors has been observed in the Gaia data, and the sample of nearby stars we have employed has, on average, relative parallax errors of 8.6% (Hinkel et al. 2020).

Finally, the uncertainties in the LS estimator are calculated from the individual uncertainties in its three component histograms. The RR , DD , and DR histograms count pairs of stars, and thus each obeys Poisson statistics. These individual uncertainties, assumed to be uncorrelated, are then propagated forward to quantify the uncertainty of the LS estimator.

Altogether, when the value of the LS estimator exceeds 1σ from zero over a scale in excess of λ_{lim} , structure is deemed

significant. Additionally, horizontal error bars are not included, as distance errors are uncorrelated with location on the sky and thus not expected to create false structure given the large number of stars in the analysis (see Section 5). With a firmer picture of the correlations we can probe now in place, we turn to a description of our methods.

5. Methodology and Data Selection

In this paper we consider the Gaia versus Gaia 2PCF to probe the structure associated with the broken symmetries observed in Widrow et al. (2012), Yanny & Gardner (2013), Ferguson et al. (2017), Bennett & Bovy (2018), and Gardner et al. (2020), though we suppose that diffuse structures such as dissolved or dissolving clusters or streams, if sufficiently large and massive, could potentially contribute to the 2PCF as well. To this end, we compute the correlations in the vertical direction by comparing data from the north with a reflection (across $z=0$) of the data from the south and comparing data from the right ($\phi < 180^\circ$) with a reflection (across $\phi = 180^\circ$) of data from the left ($\phi > 180^\circ$). These two comparisons test reflection and axial symmetry, respectively. Our selections in R and ϕ in this study correspond to the case described in Figure 3(c), and we set $z_\odot = 0$ in this section. We illustrate the geometry of these selections in Figure 5.

Moreover, we are able to probe structure in the x and y directions as well, and we do so by computing the correlations in the x and y directions from comparing data on the right of the $\phi = 180^\circ$ line with a reflection of the data on the left of this line. As our analysis is close to the $\phi = 180^\circ$ line, the x and y directions are proxies for the radial and azimuthal directions, respectively—and we pick rectilinear coordinates so that we can combine the computed correlations across the sample. We compute the separation between two stars as $|\hat{e}_Q \cdot \mathbf{x}_{12}|$ for $Q \in x, y$, or z , as we define after Equation (32). These separations are logged into the RR , DD , and DR histograms in order to construct the LS correlator for several small regions of the Galaxy. This “Gaia versus Gaia” method allows for the examination of structure in each coordinate direction independently, and it is completely independent of any theoretical model, though we do assume that our data set is free from any artificial symmetry-breaking effects, which have been carefully assessed in Section 2.4 of Hinkel et al. (2020). This is in contrast to the model-dependent analysis of Kamdar et al. (2021), which also utilized the full three-dimensional distance between two stars. Since our analysis is focused on the identification of symmetry-breaking effects, we also believe it to be particularly sensitive to non-steady-state effects.

Overall, our method does require a large number of stars free of significant biases. As such, we employ the 11.7 million stars in the data set of Hinkel et al. (2020), which has been selected to minimize the impact of faint-end incompleteness, crowded fields, extinction from dust, and artificial bias from the Gaia scan law. The selection satisfies $7 \text{ kpc} < R < 9 \text{ kpc}$, $174^\circ < \phi < 186^\circ$, $0.2 \text{ kpc} < |z| < 3.0 \text{ kpc}$, $|b| > 30^\circ$, $14 \text{ mag} < G < 18 \text{ mag}$, and $0.5 \text{ mag} < G_{\text{BP}} - G_{\text{RP}} < 2.5 \text{ mag}$, to yield an average uncertainty parallax of some 8.6%. With this data set, we are able to subdivide the 11.7 million stars into smaller wedges, which still possess $\sim 10^5$ stars in a typical, small region close to the plane and $\sim 50,000$ stars in the most limiting, high- $|z|$ cases, with an average completeness in excess of 99% (Hinkel et al. 2020) when compared to number counts from the

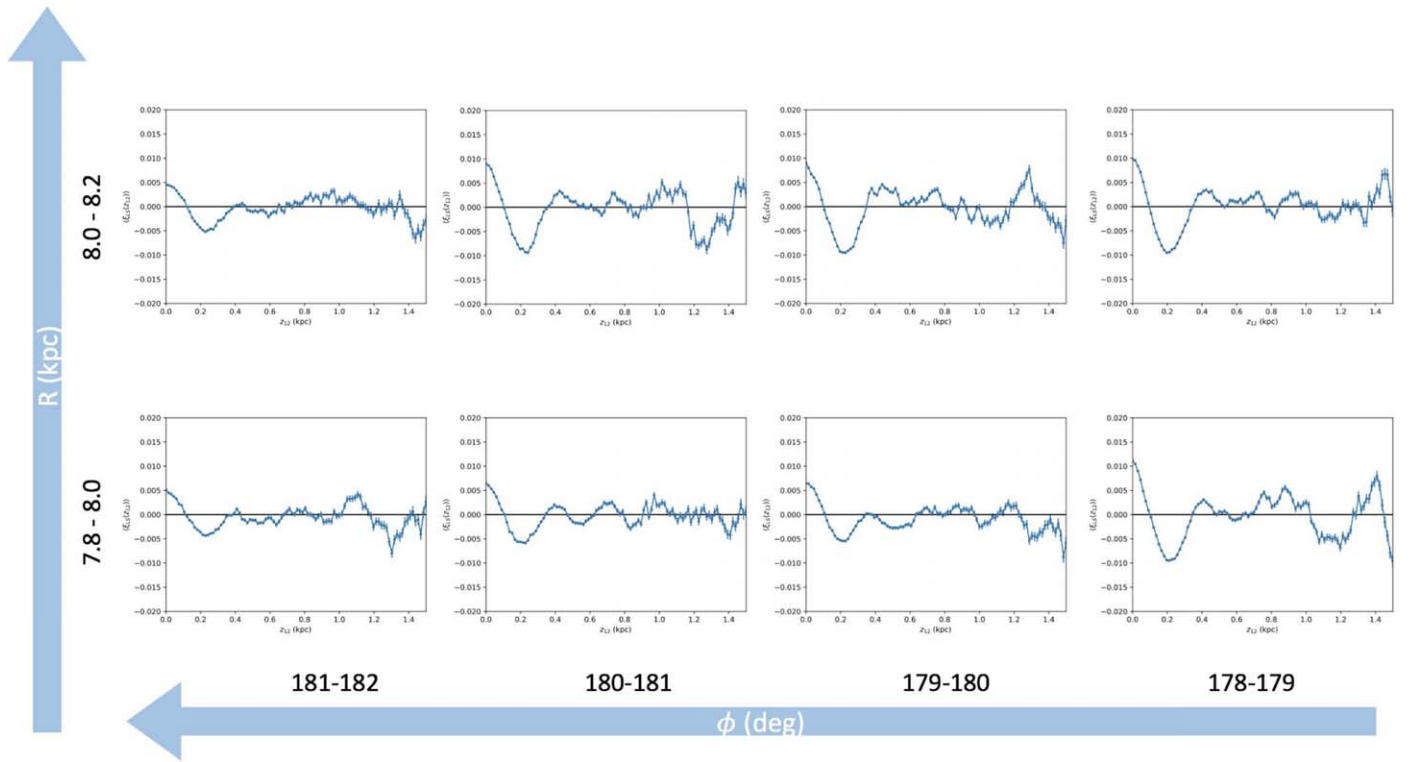


Figure 6. Vertical structure probes: several Gaia vs. Gaia, z2PCF computations for various R , ϕ wedges, comparing the north to the south. The region pictured exhibits highly correlated stars, revealing vertical waves in a new light. As the smallest, limiting length scale for this figure is about 40 pc, all of the longer wavelength features seen in the z2PCF would appear to be significant, while the shorter wavelength features are more likely not. Compare with Figure 1.

Hubble Space Telescope, as both dim stars and crowded fields are avoided (Arenou et al. 2018). This selection intersects the regions that contain the Gaia snail shell pattern (Antoja et al. 2018) and vertical waves (Widrow et al. 2012), as well as the corrugation patterns noted in the simulation of Bland-Hawthorn & Tepper-García (2021). This also holds for the more limited choices in R and ϕ we make in our current study. It should also be noted that our study relies on comparisons between nominally symmetric portions of the Galaxy. Therefore, we do not anticipate significant differences (say, north and south) in stellar crowding, parallax error, and other systematics between the portions of data being compared, in light of the elimination of scan-law patterns in our data set (Hinkel et al. 2020). Indeed, Hinkel et al. (2020) explicitly quantifies these systematic limitations.

Finally, we emphasize that our interest in small-scale structure and symmetry breaking has prompted us to examine the Gaia data in smaller regions. This subdivision of the data aided the computational efficiency of the study, as cross-correlations between various wedges of data were not computed. The analysis was made quicker still by computing only one component of the pairwise displacement. Nonetheless, each small wedge of the Galaxy in our analysis possesses $\mathcal{O}(10^5)$ stars (and thus $\mathcal{O}(10^{10})$ pairs), and so a C++ program, optimized for speed, was written to analyze the large number of pairs.

6. Analysis

We now turn to our Gaia versus Gaia 2PCF analysis, employing the LS correlator of Equation (33), considering first its vertical structure for different selections in R , ϕ and then its

radial and azimuthal structure, studying the north and south separately. Since our LS correlator is not explicitly $D \leftrightarrow R$ symmetric, the choice of D and R can impact the final result, in principle. Thus in building the Gaia versus Gaia 2PCF we assign the first data set (north) to “ D ” and the second data set (south reflection) to “ R .” We ignore the effect of $z_\odot$ throughout, so that we set $z_\odot = 0$. As a result we consider selections within the region for which $R \in [7.6, 8.4]$ kpc, $\phi \in [178^\circ, 182^\circ]$, and $|z| \in [0.2, 3]$ kpc.

6.1. Vertical Structure

For the vertical 2PCF analysis (z2PCF hereafter) we examine data selections that form annular wedges in R , ϕ , choosing a range of 200 pc in R and 1° in ϕ , with z satisfying $0.2 \text{ kpc} < |z| < 2.0 \text{ kpc}$. The separation distances are computed up to 1.5 kpc in separation, as the geometry fundamentally limits the number of pairs near the maximal 1.8 kpc of separation possible in the wedges, corresponding to stars at the maximal $|z| = 2.0 \text{ kpc}$ and minimal $|z| = 0.2 \text{ kpc}$. Each bin for the north versus south analysis has a width of 15 pc, for a rough sampling size of 10^8 pairs per bin, given the large number of pairs possible in a sample of 10^5 stars distributed over 100 bins, for example,

As noted in earlier studies (Bennett & Bovy 2018; Ferguson et al. 2017; Gardner et al. 2020; Widrow et al. 2012; Yanny & Gardner 2013) of the one-body density, i.e., the stellar number counts, north and south, significant vertical structure exists in the Galaxy near the Sun. As we demonstrate in Figure 6, the z2PCF is highly correlated, with an array of wavelike structures across the selected regions. That is, the particular wavelike pattern observed differs from wedge to wedge. Interestingly,

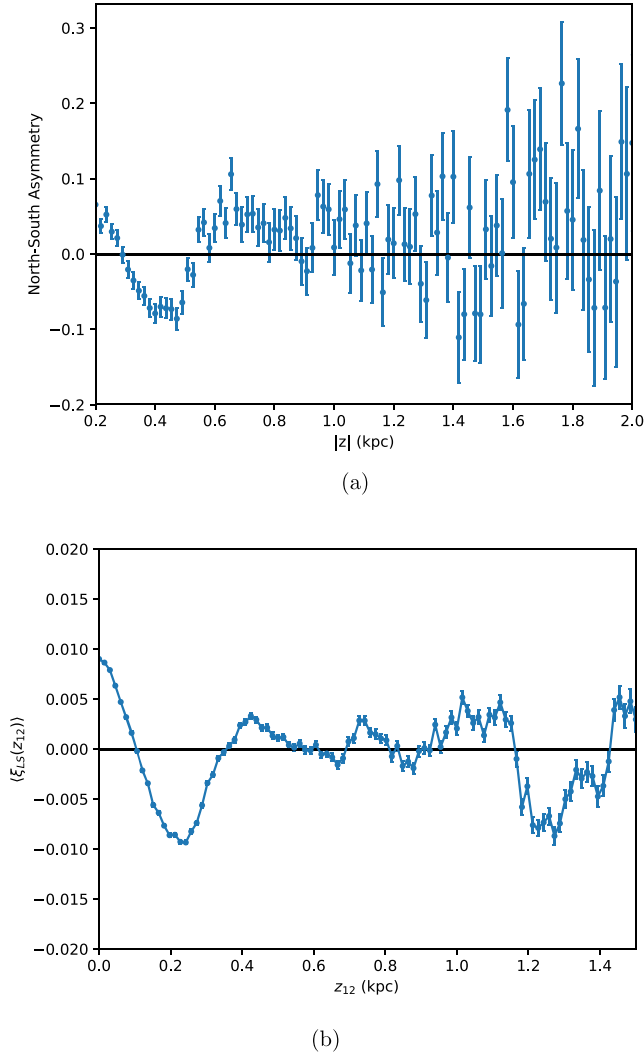


Figure 7. A comparison of (a) the north–south asymmetry and (b) the Gaia–Gaia z2PCF, north vs. south, in the region for which $R \in [8.0, 8.2]$ kpc, $\phi \in [180^\circ, 181^\circ]$, and $z \in [0.2, 2]$ kpc. The peaks and troughs in the asymmetry can be linked to some—but seemingly not all—of the peaks in the z2PCF. Namely, the crests of the wave as seen in the asymmetry at $|z| \approx 0.2$ and $|z| \approx 0.6$ kpc result in the peak near a vertical separation distance of about $z_{12} = 0.4$ kpc in the 2PCF. The peak-to-trough distances between either of the first two peaks and the first trough are around 0.2 kpc, and this registers as a cross-correlation in the LS estimator and thus a trough in the 2PCF near $z_{12} = 0.2$ kpc. The vertical asymmetry plotted in panel (a) has been centered at an asymmetry of zero by subtracting the error-weighted mean from the raw asymmetry to aid in the comparison of the two plots, as a gross offset toward the north or south will not affect the 2PCF. The vertical structure comes into focus much more clearly in the 2PCF due to the $\mathcal{O}(N^2)$ statistics when compared to the asymmetry analysis for the same region, as the latter runs out of statistics very quickly at high $|z|$.

the most significant differences appear at the largest separation distances, i.e., at high z_{12} , whereas the peak-to-trough feature at low z_{12} appears to be universal, even if some variation in its strength appears across the sample, with it attaining greatest significance for $\phi < 180^\circ$. The wave pattern seen bears comparison to the vertical wave features observed in the one-body density we have previously noted but benefits from the $\mathcal{O}(N^2)$ statistics afforded by pair counting statistics, as opposed to the $\mathcal{O}(N)$ statistics of star counts. We compare the north–

south asymmetry, defining it for $z > 0$ as

$$\mathcal{A}(|z|) = \frac{n(z) - n(-z)}{n(z) + n(-z)}, \quad (36)$$

where $n(z)$ is the number of stars at z within our selected R , ϕ sample and the z2PCF in Figure 7. We note that the sign of the north–south asymmetry indicates whether excess counts exist in the north or south, but the sign of the z2PCF—and the 2PCF in general—does not have that interpretation. Rather, as per Equation (33), the sign of the 2PCF is determined by whether direct ($DD + RR$) or cross (DR) correlations dominate. The $\mathcal{O}(N^2)$ statistics of the 2PCF give a more sensitive view of the correlations within the sample than the north–south asymmetry. Indeed, the 2PCF appears to have more finely resolved structures than the corresponding asymmetry results. These differences—especially at higher z_{12} —may speak to differences in the vertical waves across the plane, or perhaps to additional, as yet unappreciated effects.

In addition to north–south differences, differences in the vertical structure from axial symmetry breaking can be assessed via a left–right comparison. That is, a wedge of data on the left (say, $180^\circ < \phi < 181^\circ$) can be reflected across the $\phi = 180^\circ$ ray and compared against a wedge on the right (say, $179^\circ < \phi < 180^\circ$). Nonzero correlations from such a study speak to axial symmetry breaking in the vertical waves. Figures 8 and 9 illustrate these effects in the north and in the south, respectively. Different features are apparent. First, structural variations appear for the waves in the northern hemisphere, but hardly at all in the southern hemisphere. Additionally, the azimuthally adjacent wedges (Figures 8(a) and (b)) exhibit much less structure than wedges that are not azimuthally adjacent. This may just speak to larger effects in the $\phi < 179^\circ$ versus $\phi > 181^\circ$ region, as also suggested in Figure 6. Critically, however, the effect causing substantial differences in the wave structures with ϕ appears to be only significant in the north. Again, the largest differences between the various waves occur at higher values of $|z|$.

6.2. Radial and Azimuthal Structure

We have seen how the vertical structure of the Galaxy can change substantially with Galactic radius and azimuthal angle, so it is natural to ask whether in-plane correlations exist as well. Indeed, the Gaia phase-space spiral (Antoja et al. 2018), the observed axial symmetry breaking in stars at heights well away from the Galactic midplane (Gardner et al. 2020; Hinkel et al. 2020), the axial differences in vertical structure noted in Ferguson et al. (2017) and in this paper, as well as the complex corrugation patterns suggested by Bland-Hawthorn & Tepper-García (2021) all point to the tangible possibility of nonzero radial and azimuthal correlations. Thus, we examine radial and azimuthal structure in the 2PCF in x_{12} and y_{12} , which we denote as x2PCF and y2PCF, respectively. We note that x and y act as proxies for R and ϕ near the anti-center line, as we consider here. For the x2PCF and y2PCF analyses we once again examine annular wedges of data, this time with $7.6 \text{ kpc} < R < 8.4 \text{ kpc}$ and azimuthal widths of 2° in ϕ , and for various selections in z . The separation distances are computed up to 0.4 kpc in x_{12} and to 0.25 kpc in y_{12} , as the geometry limits the number of pairs beyond these scales. Each bin for the x and y analyses has a width of 8 pc.

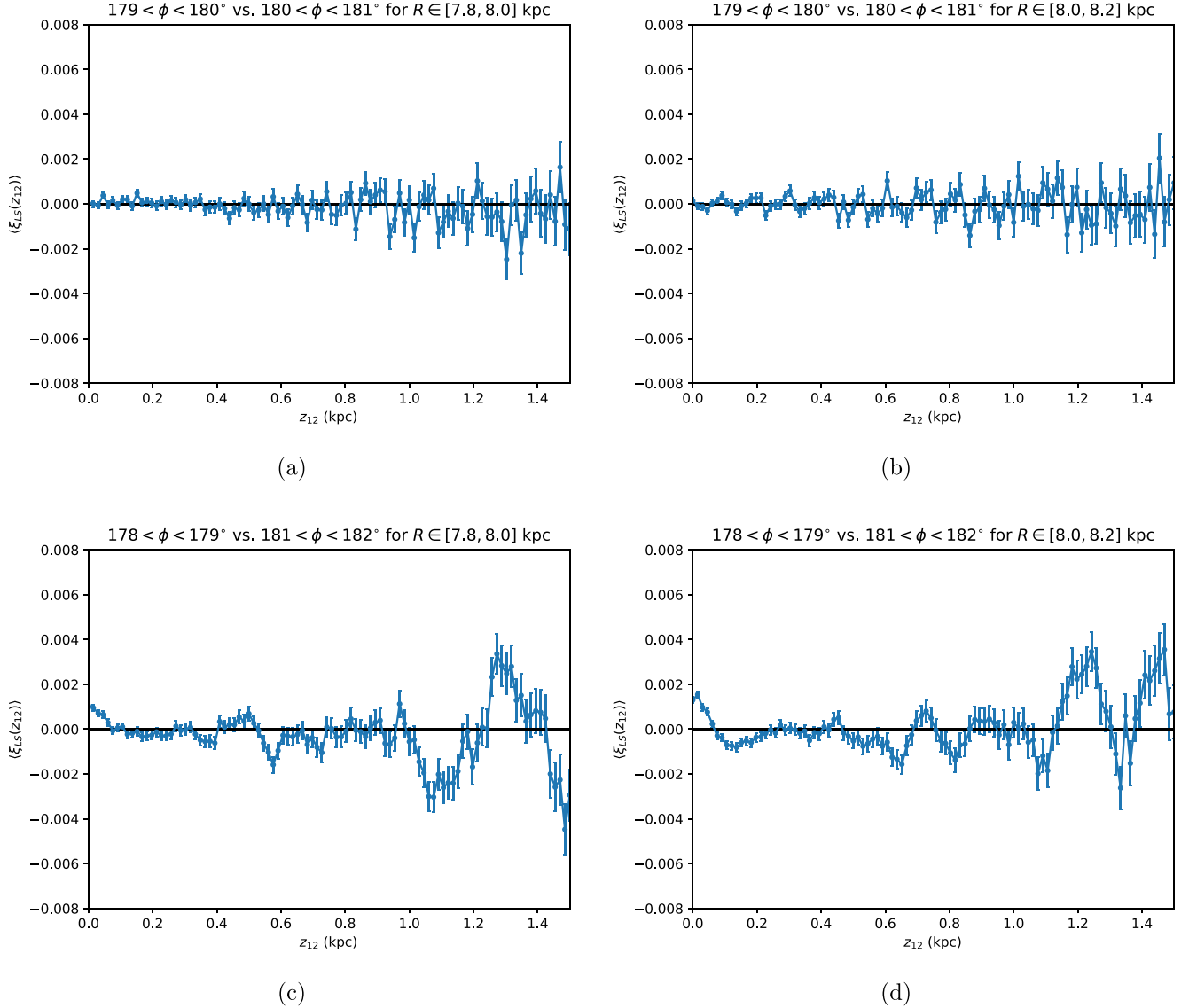


Figure 8. Northern hemisphere azimuthal structure probes. (a) The left–right z2PCF for $z > 0$ kpc, $7.8 \text{ kpc} < R < 8.0 \text{ kpc}$, and $179^\circ < \phi < 180^\circ$ vs. $180^\circ < \phi < 181^\circ$. (b) The left–right z2PCF for $z > 0$ kpc, $8.0 \text{ kpc} < R < 8.2 \text{ kpc}$, and $179^\circ < \phi < 180^\circ$ vs. $180^\circ < \phi < 181^\circ$. (c) The left–right z2PCF for $z > 0$ kpc, $7.8 \text{ kpc} < R < 8.0 \text{ kpc}$, and $178^\circ < \phi < 179^\circ$ vs. $181^\circ < \phi < 182^\circ$. (d) The left–right z2PCF for $z > 0$ kpc, $8.0 \text{ kpc} < R < 8.2 \text{ kpc}$, and $178^\circ < \phi < 179^\circ$ vs. $181^\circ < \phi < 182^\circ$. The smallest length scale to which we can probe is about 40 pc, indicating significant structure at high values of z_{12} in panels (c) and (d) and a dearth of structure in panels (a) and (b).

Left–right structural differences in the x -direction are shown in Figure 10 for various z slices in the northern hemisphere, and the waves appear to become shorter in wavelength as z increases from panel (a) through panel (e). Interestingly, the stars in the north are correlated in the x -direction, especially for $|z| < 1.2$ kpc. The same is not true for the south, however, as depicted in Figure 11. While both the north and the south share some structural similarities at the lowest $|z|$ values, the south lacks the structure seen in the north in the region with $0.3 \text{ kpc} < |z| < 1.2 \text{ kpc}$, as shown in panels (b)–(d). The preponderance of x -direction structure in the north and the lack thereof in the south is also consistent with our vertical studies. Indeed, the vertical waves we observe exhibit marked left–right structural differences in the north (Figure 8) but not the south (Figure 9). It appears that axial symmetry breaking combines with the vertical waves to create a corrugation pattern of

some kind, even beyond the region of the Gaia snail (Antoja et al. 2018), though this appears to be restricted to the north.

Additionally, slight hints of y -direction structure exist in the north (Figure 12) but not the south (Figure 13). While not as significant as the x -direction structures we have just noted, the stars nonetheless appear to be correlated. Moreover, a tantalizing, wavelike structure appears at the highest $|z|$ values in the north, perhaps hinting at a new structure: an azimuthal wave of some sort. The wavelength of this wavelike correlation pattern is about 40–50 pc.

We now turn to the consideration of the possible origins of the effects we have observed.

7. Theoretical Origins

The intricate phase-space correlations presented by the observed vertical asymmetries (Bennett & Bovy 2018;

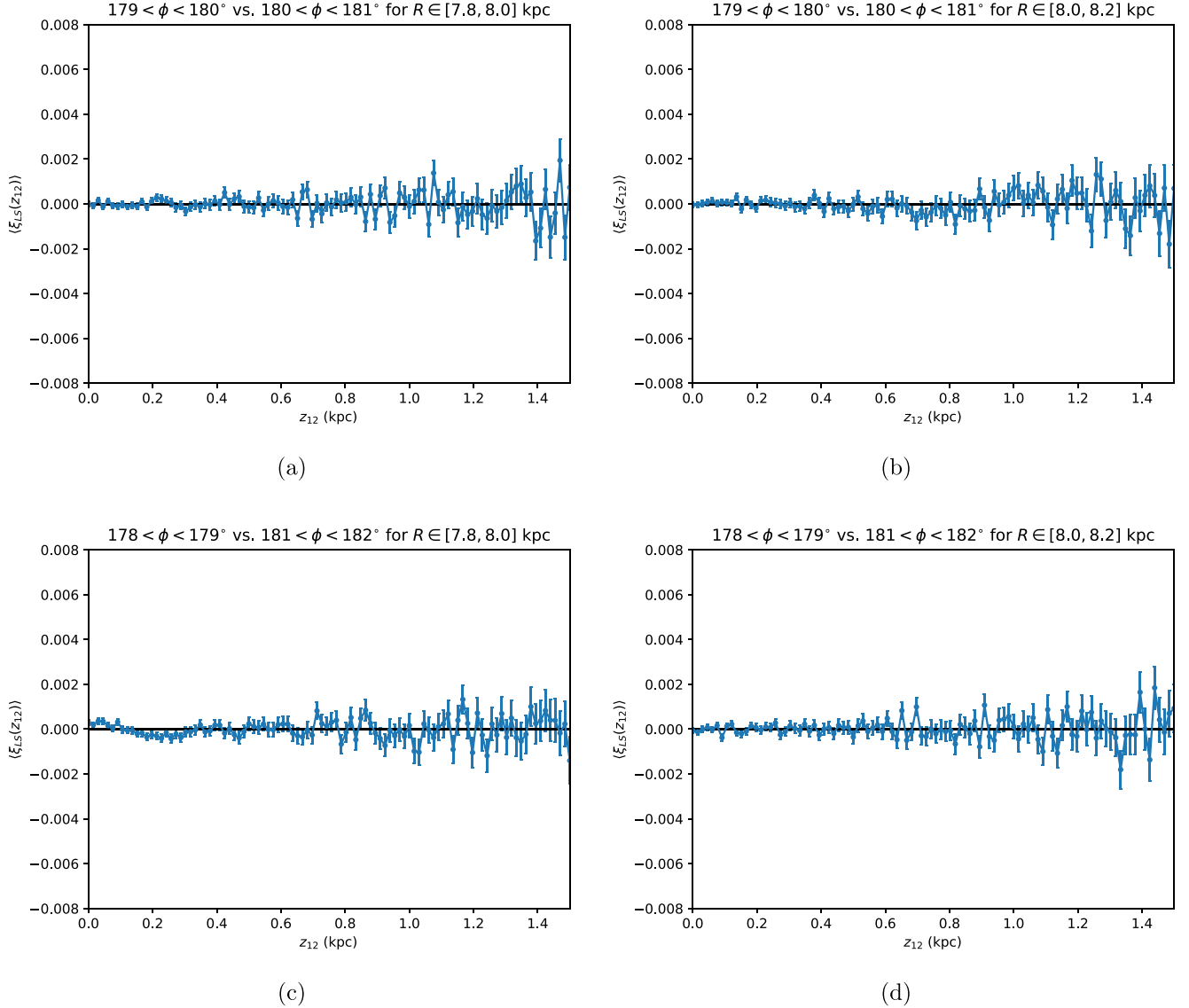


Figure 9. Southern hemisphere azimuthal structure probes. (a) The left–right z2PCF for $z < 0$ kpc, $7.8 \text{ kpc} < R < 8.0 \text{ kpc}$, and $179^\circ < \phi < 180^\circ$ vs. $180^\circ < \phi < 181^\circ$. (b) The left–right z2PCF for $z < 0$ kpc, $8.0 \text{ kpc} < R < 8.2 \text{ kpc}$, and $179^\circ < \phi < 180^\circ$ vs. $180^\circ < \phi < 181^\circ$. (c) The left–right z2PCF for $z < 0$ kpc, $7.8 \text{ kpc} < R < 8.0 \text{ kpc}$, and $178^\circ < \phi < 179^\circ$ vs. $181^\circ < \phi < 182^\circ$. (d) The left–right z2PCF for $z < 0$ kpc, $8.0 \text{ kpc} < R < 8.2 \text{ kpc}$, and $178^\circ < \phi < 179^\circ$ vs. $181^\circ < \phi < 182^\circ$. The smallest length scale to which we can probe is 40 pc, indicating a dearth of differences with ϕ in the vertical structure in the south.

Thulasidharan et al. 2022; Widrow et al. 2012; Yanny & Gardner 2013) and the existence of various snail-like correlations in position and velocity space (Bland-Hawthorn & Tepper-García 2021; Antoja et al. 2018; Gandhi et al. 2022) in the stars within roughly 1–2 kpc of the Sun have already prompted much discussion. In this section we revisit these suggestions and consider whether any of them can explain the particular features we have found. The symmetry-based 2PCF framework we have developed in this paper allows us to study the intricate three-dimensional nature of stellar correlations in position space, and a striking outcome of our study is that the R and ϕ correlations are much richer in the north than in the south. We think that the various phase-space correlations we have found and which have been noted shed light on the merger history of the Milky Way—and note, for context, that mergers have been invoked, variously, to explain the origin of the spiral arms (Purcell et al. 2011) and of the thick disk (Helmi et al. 2018). In regard to specific effects, it has been argued that

the vertical asymmetries and the Gaia snail could be an aftermath of a collision of the Sagittarius (Sgr) dwarf galaxy with the Galactic disk (Gandhi et al. 2022; Binney & Schönrich 2018; Bland-Hawthorn & Tepper-García 2021; Widrow et al. 2012), with others noting that the last Sgr impact could not have been the only source of perturbation (Bennett & Bovy 2021; Bennett et al. 2022). We note, too, that numerical simulations with vertical asymmetries (García-Conde et al. 2022) or with bending and breathing features (Bland-Hawthorn & Tepper-García 2021; Ghosh et al. 2022; Hunt et al. 2021, 2022) can give rise to phase spirals, suggesting that the different phenomena are intertwined.

The interpretations we have noted have been made in a one-body picture. Thus further simulation and analysis are needed to compare more easily with the 2PCF results we have found. While we leave the task of explicit simulations to future work, we can nonetheless comment on the possible theoretical origins for the structures we find. First, as motivated in Section 3 and

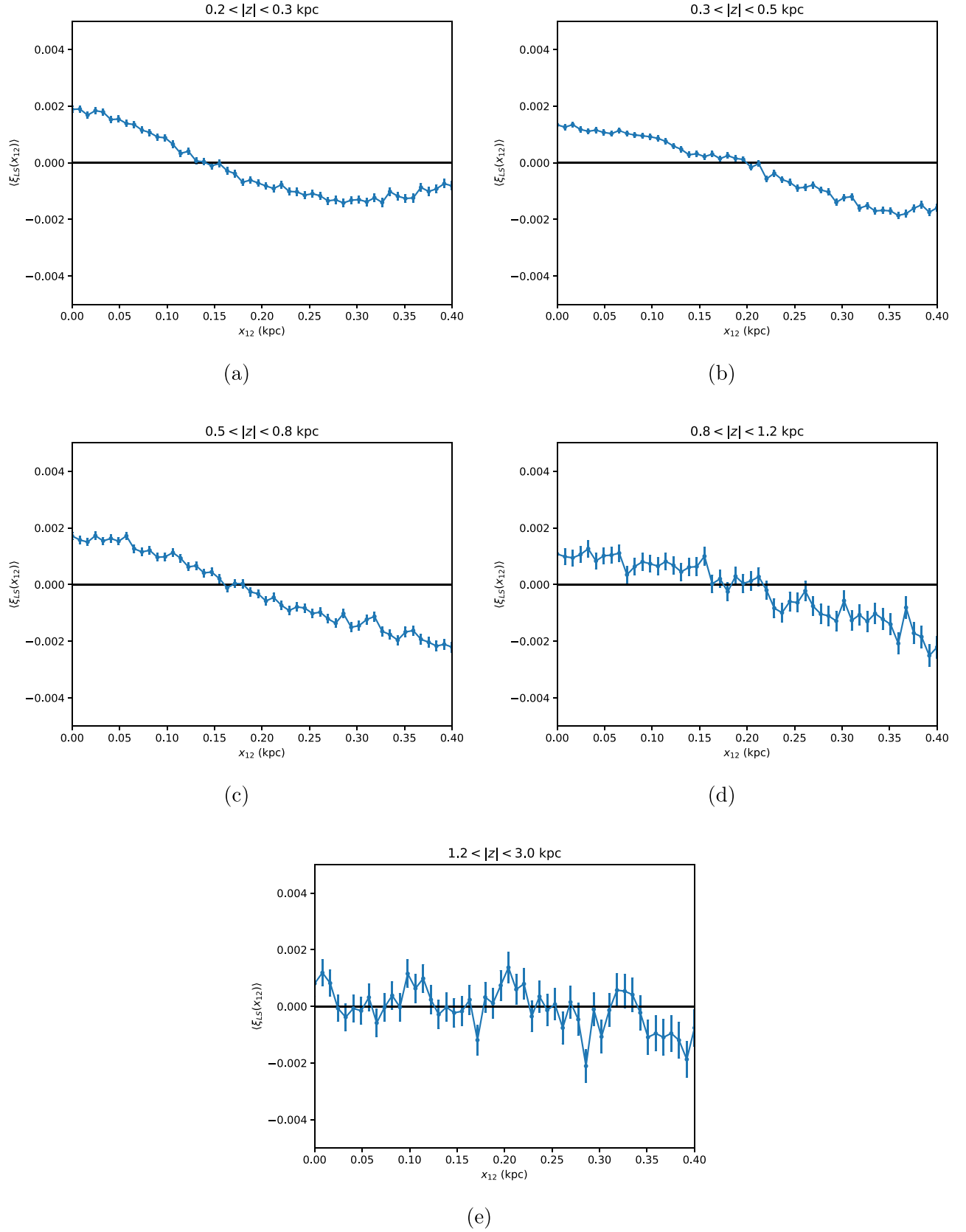


Figure 10. Northern hemisphere radial structure probes. Left ($180^\circ < \phi < 182^\circ$) vs. right ($178^\circ < \phi < 180^\circ$) comparisons of structure in the x -direction for the northern hemisphere, with $7.6 \text{ kpc} < R < 8.4 \text{ kpc}$, and for various slices of $|z|$. (a) $0.2 \text{ kpc} < |z| < 0.3 \text{ kpc}$, (b) $0.3 \text{ kpc} < |z| < 0.5 \text{ kpc}$, (c) $0.5 \text{ kpc} < |z| < 0.8 \text{ kpc}$, (d) $0.8 \text{ kpc} < |z| < 1.2 \text{ kpc}$, (e) $1.2 \text{ kpc} < |z| < 3.0 \text{ kpc}$. The smallest length scale to which we can probe is 20 pc at lower values of z and 30 pc at the highest range of z .

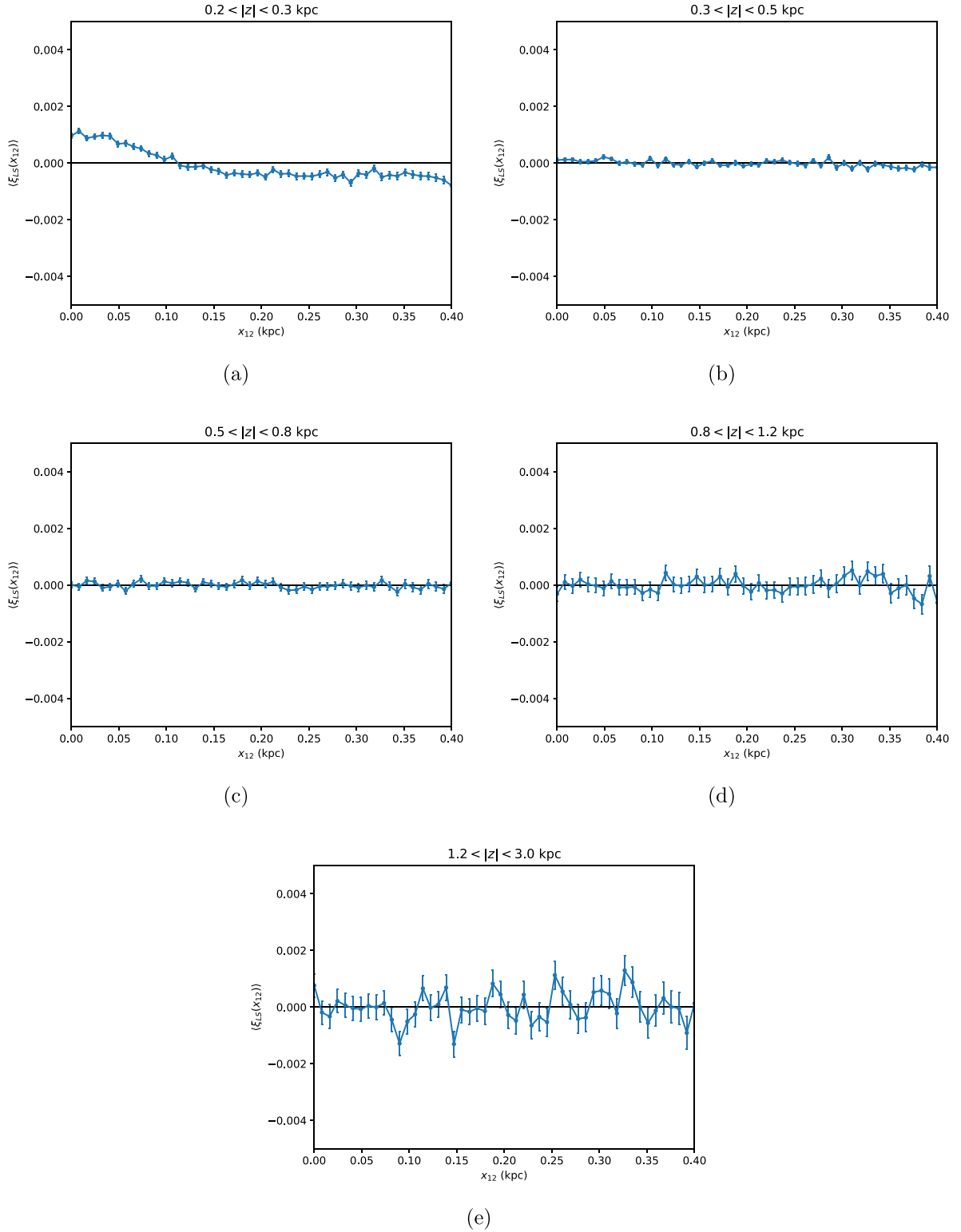


Figure 11. Southern hemisphere radial structure probes. Left ($180^\circ < \phi < 182^\circ$) vs. right ($178^\circ < \phi < 180^\circ$) comparisons of structure in the x -direction for the southern hemisphere, with $7.6 \text{ kpc} < R < 8.4 \text{ kpc}$, and for various slices of $|z|$. (a) $0.2 \text{ kpc} < |z| < 0.3 \text{ kpc}$, (b) $0.3 \text{ kpc} < |z| < 0.5 \text{ kpc}$, (c) $0.5 \text{ kpc} < |z| < 0.8 \text{ kpc}$, (d) $0.8 \text{ kpc} < |z| < 1.2 \text{ kpc}$, (e) $1.2 \text{ kpc} < |z| < 3.0 \text{ kpc}$. The smallest length scale to which we can probe is 20 pc at lower values of z and 30 pc at the highest range of z .

depicted in Figure 7, the over- and underdensities of the vertical waves (Bennett & Bovy 2018; Widrow et al. 2012; Yanny & Gardner 2013) appear to map directly onto our vertical structure (z2PCF) results. Indeed, the z2PCF analysis brings the wavelike features into sharper focus, particularly at larger relative separations z_{12} . Additionally, that this vertical

structure appears to vary across the disk agrees broadly with the findings of Ferguson et al. (2017) and the corrugations suggested by Bland-Hawthorn & Tepper-García (2021). While Bland-Hawthorn & Tepper-García (2021) detail via simulation how the Gaia phase-space spiral (Antoja et al. 2018) may have come about due to a superposition of density waves and

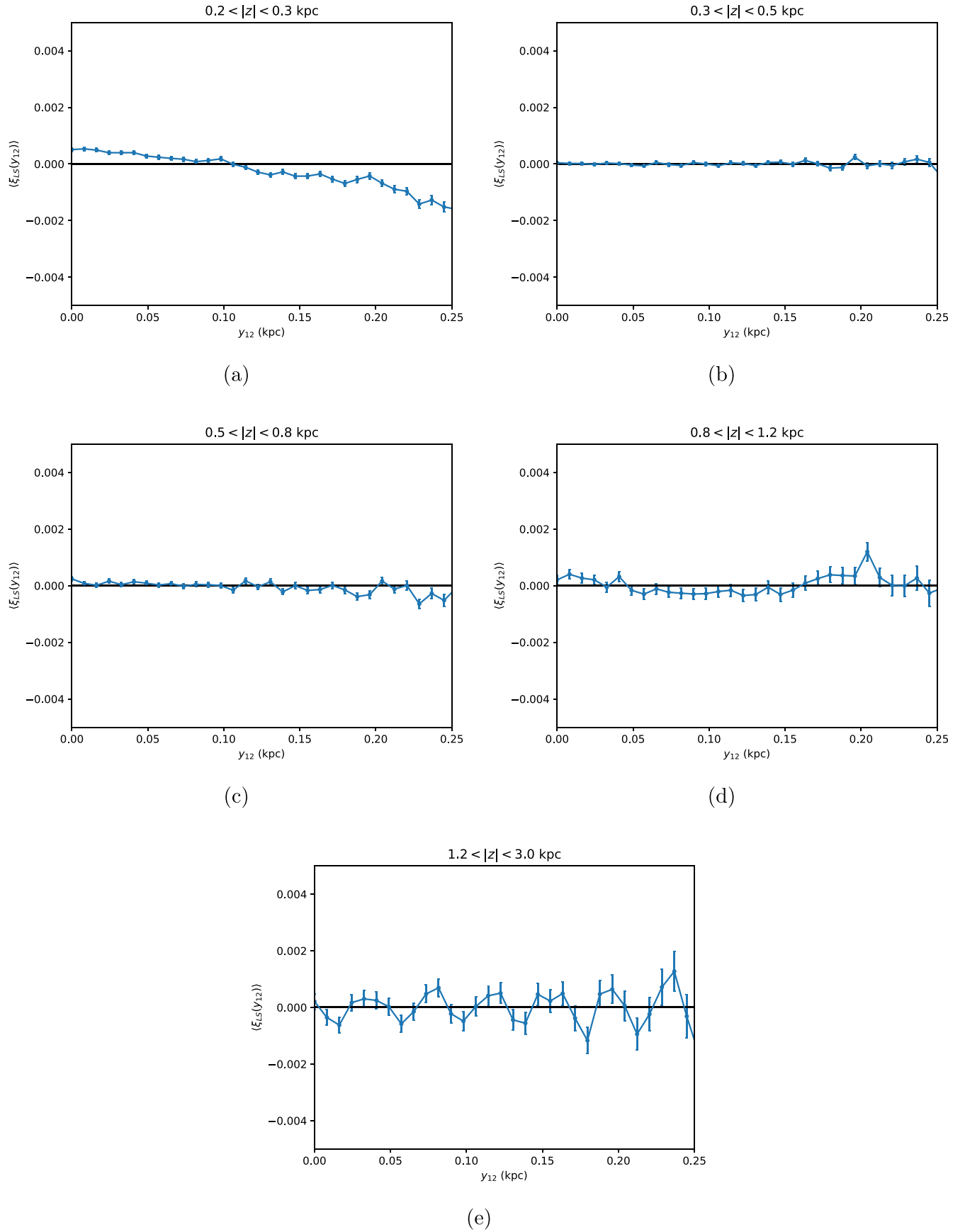


Figure 12. Left ($180^\circ < \phi < 182^\circ$) vs. right ($178^\circ < \phi < 180^\circ$) comparisons of structure in the y -direction for the northern hemisphere, with $7.6 \text{ kpc} < R < 8.4 \text{ kpc}$, and for various slices of $|z|$. (a) $0.2 \text{ kpc} < |z| < 0.3 \text{ kpc}$, (b) $0.3 \text{ kpc} < |z| < 0.5 \text{ kpc}$, (c) $0.5 \text{ kpc} < |z| < 0.8 \text{ kpc}$, (d) $0.8 \text{ kpc} < |z| < 1.2 \text{ kpc}$, (e) $1.2 \text{ kpc} < |z| < 3.0 \text{ kpc}$. The smallest significant length scale is 10 pc for this figure.

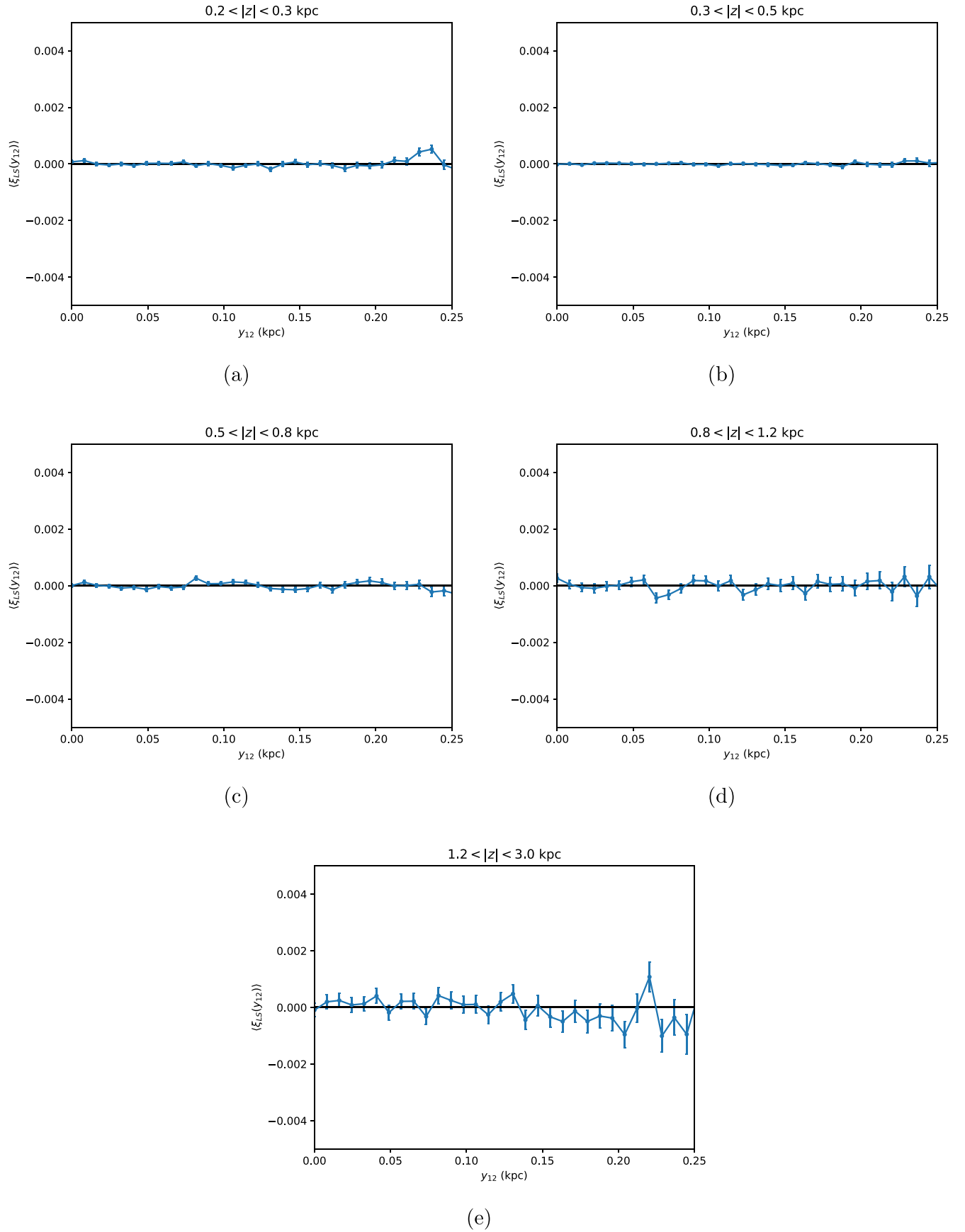


Figure 13. Left ($180^\circ < \phi < 182^\circ$) vs. right ($178^\circ < \phi < 180^\circ$) comparisons of structure in the y -direction for the southern hemisphere, with $7.6 \text{ kpc} < R < 8.4 \text{ kpc}$, and for various slices of $|z|$. (a) $0.2 \text{ kpc} < |z| < 0.3 \text{ kpc}$, (b) $0.3 \text{ kpc} < |z| < 0.5 \text{ kpc}$, (c) $0.5 \text{ kpc} < |z| < 0.8 \text{ kpc}$, (d) $0.8 \text{ kpc} < |z| < 1.2 \text{ kpc}$, (e) $1.2 \text{ kpc} < |z| < 3.0 \text{ kpc}$. The smallest significant length scale is 10 pc for this figure.

bending waves caused by a collision with the Milky Way, assessing a full picture of the comparison between phase-space structures and a purely spatial 2PCF requires a detailed study. Nevertheless, the suggestion of Bland-Hawthorn & Tepper-García (2021) that the complex superposition of these two distinct waves is responsible for the phase-space spiral(s) is compatible with the complex vertical landscape we find. Certainly, our 2PCF analysis disfavors purely planar vertical waves. Indeed, the breathing modes suggested by Ghosh et al. (2022) and Hunt et al. (2022), the vertical waves (Widrow et al. 2012), and the bending of the Galaxy (Chen et al. 2019; Levine et al. 2006a; Poggio et al. 2018; Skowron et al. 2019) could all plausibly contribute to the vertical structures we find, and they could act in varying combinations.

In fact, Ghosh et al. (2022) suggest that Galactic breathing modes are excited by spiral structure, and the particular phase of a star’s breathing mode motion would create a z - ϕ coupling. As our sample is represented by a very small slice of the simulation in Figure 3 of Ghosh et al. (2022), we expect this z - ϕ coupling to be small in our context. This matches the results in Figure 8 where vertical structure in adjacent azimuthal bins appears to be nearly identical, but azimuthal bins separated by some distance begin to show differences in vertical structure at high $|z|$. This also matches the suggestion of Ghosh et al. (2022) that the breathing mode amplitude is directly proportional to the height above the plane, $|z|$, and so the small effect only becomes noticeable in our data at high $|z|$.

Similarly, Widmark et al. (2022) find evidence for breathing modes in the solar neighborhood and suggest that the pattern speed of the local spiral arm is slower than the rotation of stars in the solar neighborhood. Widmark et al. also claim that their Gaussian process fit method indicates that the Galactic warp affects only the stars in the thick disk, while the thin disk is largely unaffected. While it is unclear how such a configuration might occur given the observed warp in both groups of stars (Poggio et al. 2018; Skowron et al. 2019) and H I gas (Burke 1957; Kerr 1957; Levine et al. 2006a), this behavior does potentially explain the results in Figure 8. That is, the gradual warping introduces a small z - ϕ coupling only for the thick disk, and thus only the high- $|z|$ peaks separated by some azimuthal distance exhibit a measurable breaking of axial symmetry in their vertical structure. Nonetheless, the Gaia data used by Widmark et al. (2022) may be more difficult to interpret because of dust and stellar crowding effects in the midplane. Suggesting that the thick and thin disks respond in observably different ways to perturbations may also suggest that they acted in the relatively recent past, which is intriguing.

We emphasize that the structural differences we find appear more marked in the north, and it is likely that this effect cannot be explained solely by the mechanisms we have noted in previous paragraphs. For context, we recall that Bennett et al. (2022) and Bennett & Bovy (2021) contend that the Sagittarius impact picture alone cannot reproduce either the Gaia snail or the vertical waves in simulations. Moreover, Gandhi et al. (2022) show that the current passage of the Sagittarius dwarf has already had its greatest effect while below the disk, and is likely to have made a much smaller effect when it crossed into the north, due to a larger average Galactocentric radius.

Regardless of whether the observed perturbations come from a superposition of waves (Bland-Hawthorn & Tepper-García 2021) or from a Sagittarius impact modulated by the Large Magellanic Cloud (Laporte et al. 2018), it seems increasingly

clear that a multitude of effects are likely recorded in the structure of the Milky Way—and this is made all the more likely by the growing census of past mergers (Lövdal et al. 2022; Malhan et al. 2022). If the Milky Way did grow hierarchically over cosmic time, as expected in the cold dark matter paradigm (Peebles 1993), the aggregate of its entire merger history and its extremely long relaxation time (Binney & Tremaine 2008) does lend credence to a picture of multiple effects. Swathes of stars perturbed by past impacts with the disk would then retain some information about these long-ago mergers, resulting in compounding alterations to the Galaxy’s structure. Presumably, the 2PCF analyses effected within this paper measure the integrated properties of these myriad mergers, along with other effects, with the various probes of symmetry breaking hinting at particular effects. Comparing future simulation with future 2PCF analyses could yield sharpened constraints on the various scenarios, as the discriminating ability of the 2PCF surpasses that of studies of stellar number counts thanks to their $\mathcal{O}(N^2)$ statistics.

8. Conclusions

We have introduced a new realization of the 2PCF and have derived useful functional forms for spherical and slab geometries in the steady-state limit that show that the 2PCF is vanishingly small at the length scales probed in our study. This provides a setting for our observational analysis in which, by exploiting reflection and axial symmetry, we have compared the structural differences of various regions of the Galaxy against one another. In particular, we have examined the 2PCF as a function of the separation in x , y , or z only, for different selections of R , ϕ , and z , fingerprinting the effects of the various perturbations that have acted over the Galaxy’s history. As we have developed in Section 2, these observed effects attest to the existence of time-dependent perturbations. Ultimately, it is clear from this analysis that the stars in our Galaxy are not perfectly uncorrelated as commonly assumed (Binney & Tremaine 2008). Rather, we have discovered that the stars are highly correlated in the vertical direction—confirming and sharpening previous discoveries of vertical waves in the Milky Way disk (Bennett & Bovy 2018; Widrow et al. 2012; Yanny & Gardner 2013). These wavelike, vertical structures exhibit small differences in phase and amplitude across the Galactic disk, especially at higher $|z|$, and non-adjacent wedges of data show marked azimuthal differences in the waves, also at higher $|z|$.

In addition to the aforementioned vertical structures, we find substantial evidence for radial and azimuthal structures in the 2PCF. To summarize:

1. Substantial radial structure appears at lower $|z|$, though it is much more apparent in the northern hemisphere. Radial structure in the north extends all the way to 1.2 kpc above the plane, while radial structure in the south is mostly confined to $0.2 \text{ kpc} < |z| < 0.3 \text{ kpc}$. This north-heavy structure trend is consistent with the azimuthal differences in vertical structure seen in the north, and with the relative lack of structure seen in the south.
2. Further, some hints of azimuthal structure exist—again predominantly in the north. Some slight azimuthal correlations exist at low $|z|$ in the north, but otherwise the thin disk appears to be devoid of azimuthal structure at the z we consider, which ought be well away from the

spiral arms. Very interestingly, an azimuthal wave structure with a wavelength around 40 or 50 pc is apparent at high $|z|$ in the north, perhaps speaking to previously undiscovered dynamical effects.

3. Additionally, we find evidence of substantial structural variations across R and ϕ in the vertical direction. Not only have we resolved the vertical waves discovered by Widrow et al. (2012) in our z2PCF analysis, but a left-right comparison shows significant differences at high $|z|$, suggesting that the waves are being disrupted (i.e., not perfectly planar waves, agreeing with the findings of Ferguson et al. (2017)) or superimposed on an entirely different effect, such as the corrugations suggested by Bland-Hawthorn & Tepper-García (2021) or a tilt in the midplane location (Eilers et al. 2020; Katz et al. 2018).

More study is required to determine the precise origin(s) of the correlations we find. We do think that warping or tilting of the disk is not a sole contributing cause to the azimuthal structures that we see in Figures 8–13. Since this analysis is effected in heliocentric coordinates, the true Galactic midplane is really below our $z = 0$ plane. This in itself would seem to disfavor disk warping or tilting as a sole cause because these effects would presumably be visible in both the north and south, which is not at all what we observe.

S.G. and A.H. acknowledge partial support from the University Research Professor (S.G.) fund of the University of Kentucky and from the U.S. Department of Energy under contract DE-FG02-96ER40989. A.H. acknowledges support from the Universities Research Association and the University of Kentucky College of Arts & Sciences' Dean's Competitive Fellowship.

We thank Joss Bland-Hawthorn for early remarks that we considered in framing our systematic error analysis and Scott Tremaine for comments on the accepted paper that we have addressed in the proofs. We also thank the anonymous referee for helpful comments that have improved the presentation of our paper.

This document was prepared in part using the resources of Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by the Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359.

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.

ORCID iDs

Austin Hinkel  <https://orcid.org/0000-0002-9785-914X>
 Susan Gardner  <https://orcid.org/0000-0002-6166-5546>
 Brian Yanny  <https://orcid.org/0000-0002-9541-2678>

References

An, J., Evans, N., & Sanders, J. 2017, *MNRAS*, **467**, 1281
 Antoja, T., Helmi, A., Romero-Gomez, M., et al. 2018, *Natur*, **561**, 360
 Arenou, F., Luri, X., Babusiaux, C., et al. 2018, *A&A*, **616**, A17
 Banik, U., Weinberg, M. D., & van den Bosch, F. C. 2022, *ApJ*, **935**, 135

Bennett, M., & Bovy, J. 2018, *MNRAS*, **482**, 1417
 Bennett, M., & Bovy, J. 2021, *MNRAS*, **503**, 376
 Bennett, M., Bovy, J., & Hunt, J. A. S. 2022, *ApJ*, **927**, 131
 Binney, J., & Schönrich, R. 2018, *MNRAS*, **481**, 1501
 Binney, J., & Tremaine, S. 2008, *Galactic Dynamics* (2nd edn.; Princeton, NJ: Princeton Univ. Press)
 Bland-Hawthorn, J., & Freeman, K. C. 2004, *PASA*, **21**, 110
 Bland-Hawthorn, J., Krumholz, M. R., & Freeman, K. 2010, *ApJ*, **713**, 166
 Bland-Hawthorn, J., & Tepper-García, T. 2021, *MNRAS*, **504**, 3168
 Blatt, J. M., & Weisskopf, V. F. 2010, *Theoretical Nuclear Physics* (1st ed.; New York: Dover)
 Bovy, J. 2015, *ApJS*, **216**, 29
 Bovy, J., & Rix, H.-W. 2013, *ApJ*, **779**, 115
 Buckley, M. R., & Peter, A. H. G. 2018, *PhR*, **761**, 1
 Burke, B. F. 1957, *AJ*, **62**, 90
 Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
 Chen, X., Wang, S., Deng, L., et al. 2019, *NatAs*, **3**, 320
 Cooper, A. P., Cole, S., Frenk, C. S., & Helmi, A. 2011, *MNRAS*, **417**, 2206
 Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, *ApJ*, **696**, 870
 Drimmel, R., & Spergel, D. N. 2001, *ApJ*, **556**, 181
 Drimmel, R., Romero-Gomez, M., Chemin, L., et al. 2022, arXiv:2206.06207
 Eilers, A.-C., Hogg, D. W., Rix, H.-W., et al. 2020, *ApJ*, **900**, 186
 Ferguson, D., Gardner, S., & Yanny, B. 2017, *ApJ*, **843**, 141
 Flewelling, H. A., Magnier, E. A., Chambers, K. C., et al. 2020, *ApJS*, **251**, 7
 Freudreich, H. T., Berriman, G. B., Dwek, E., et al. 1994, *ApJL*, **429**, L69
 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. 2022, *A&A*, arXiv:2208.00211
 Gandhi, S. S., Johnston, K. V., Hunt, J. A. S., et al. 2022, *ApJ*, **928**, 80
 García-Conde, B., Roca-Fàbrega, S., Antoja, T., Ramos, P., & Valenzuela, O. 2022, *MNRAS*, **510**, 154
 Gardner, S., Hinkel, A., & Yanny, B. 2020, *ApJ*, **890**, 110
 Gardner, S., McDermott, S. D., & Yanny, B. 2021, *PrPNP*, **121**, 103904
 Ghosh, S., Debatista, V. P., & Khachatryan, T. 2022, *MNRAS*, **511**, 784
 Gieles, M., Zwart, S. F. P., Baumgardt, H., et al. 2006, *MNRAS*, **371**, 793
 Gómez, F. A., Minchev, I., O'Shea, B. W., et al. 2012, *MNRAS*, **429**, 159
 Gómez, F. A., Torres-Flores, S., Mora-Urrejola, C., et al. 2021, *ApJ*, **908**, 27
 Goodstein, D. L. 2014, *States of Matter* (1st ed.; New York: Dover)
 Hauser, M. G., & Peebles, P. J. E. 1973, *ApJ*, **185**, 757
 Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, *Natur*, **563**, 85
 Hinkel, A., Gardner, S., & Yanny, B. 2020, *ApJ*, **893**, 105
 Hunt, J. A. S., Stelea, I. A., Johnston, K. V., et al. 2021, *MNRAS*, **508**, 1459
 Hunt, J. A. S., Price-Whelan, A. M., Johnston, K. V., & Darragh-Ford, E. 2022, *MNRAS*, **516**, L7
 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
 Kamdar, H., Conroy, C., Ting, Y.-S., & El-Badry, K. 2021, *ApJ*, **922**, 49
 Kardar, M. 2007, *Statistical Physics of Particles* (1st ed.; Cambridge: Cambridge Univ. Press)
 Katz, D., Antoja, T., Romero-Gómez, M., et al. 2018, *A&A*, **616**, A11
 Katz, D., Sartoretti, P., Cropper, M., et al. 2019, *A&A*, **622**, A205
 Katz, D., Sartoretti, P., Guerrier, A., et al. 2022, arXiv:2206.05902
 Keihänen, E., Kurki-Suonio, H., Lindholm, V., et al. 2019, *A&A*, **631**, A73
 Kerr, F. J. 1957, *AJ*, **62**, 93
 Kerr, F. J., Hindman, J. V., & Carpenter, M. S. 1957, *Natur*, **180**, 677
 Kirkwood, J. G., Maun, E. K., & Alder, B. J. 1950, *JChPh*, **18**, 1040
 Krumholz, M. R., McKee, C. F., & Bland-Hawthorn, J. 2019, *ARA&A*, **57**, 227
 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. P., Johnston, K. V., Gómez, F. A., Garavito-Camargo, N., & Besla, G. 2018, *MNRAS*, **481**, 286
 Levine, E. S., Blitz, L., & Heiles, C. 2006a, *ApJ*, **643**, 881
 Levine, E. S., Blitz, L., & Heiles, C. 2006b, *Sci*, **312**, 1773
 Li, T. S., Sheffield, A. A., Johnston, K. V., et al. 2017, *ApJ*, **844**, 74
 Lindegren, L., Hernández, J., Bombrun, A., et al. 2018, *A&A*, **616**, A2
 Löfdal, S. S., Ruiz-Lara, T., Koppelman, H. H., et al. 2022, *A&A*, **665**, A57
 Malhan, K., Ibata, R. A., Sharma, S., et al. 2022, *ApJ*, **926**, 107
 Mao, Q., Berlind, A. A., Holley-Bockelmann, K., et al. 2015, arXiv:1507.01593
 Minchev, I., Chiappini, C., Martig, M., et al. 2014, *ApJL*, **781**, L20
 Nelson, P., & Widrow, L. M. 2022, *MNRAS*, **516**, 5429
 Peebles, P. J. E. 1973, *ApJ*, **185**, 413
 Peebles, P. J. E. 1980, *The Large-scale Structure of the Universe* (Princeton, NJ: Princeton Univ. Press)

- Peebles, P. J. E. 1993, *Principles of Physical Cosmology* (Princeton, NJ: Princeton Univ. Press)
- Poggio, E., Laporte, C. F., Johnston, K. V., et al. 2021a, *MNRAS*, **508**, 541
- Poggio, E., Drimmel, R., Cantat-Gaudin, T., et al. 2021b, *A&A*, **651**, A104
- Poggio, E., Drimmel, R., Lattanzi, M. G., et al. 2018, *MNRAS*, **481**, L21
- Price-Whelan, A. M., Johnston, K. V., Sheffield, A. A., Laporte, C. F. P., & Sesar, B. 2015, *MNRAS*, **452**, 676
- Purcell, C. W., Bullock, J. S., Tollerud, E. J., Rocha, M., & Chakrabarti, S. 2011, *Natur*, **477**, 301
- Schutz, K., Lin, T., Safdi, B. R., & Wu, C.-L. 2018, *PhRv*, **121**, 081101
- Searle, L., & Zinn, R. 1978, *ApJ*, **225**, 357
- Sheffield, A. A., Price-Whelan, A. M., Tzanidakis, A., et al. 2018, *ApJ*, **854**, 47
- Skowron, D. M., Skowron, J., Mroz, P., et al. 2019, *Sci*, **365**, 478
- Spitzer, L., Jr. 1942, *ApJ*, **95**, 329
- Thorne, K. S., & Blandford, R. D. 2017, *Modern Classical Physics* (Princeton, NJ: Princeton Univ. Press)
- Thulasidharan, L., D’Onghia, E., Poggio, E., et al. 2022, *A&A*, **660**, L12
- Wall, J. V., & Jenkins, C. R. 2012, *Practical Statistics for Astronomers* (Cambridge: Cambridge Univ. Press)
- Weinberg, M. D., & Blitz, L. 2006, *ApJL*, **641**, L33
- Widmark, A., Widrow, L. M., & Naik, A. 2022, *A&A*, **668**, A95
- Widrow, L. M., Gardner, S., Yanny, B., Dodelson, S., & Chen, H.-Y. 2012, *ApJL*, **750**, L41
- Williams, M. E. K., Steinmetz, M., Binney, J., et al. 2013, *MNRAS*, **436**, 101
- Xu, Y., Newberg, H. J., Carlin, J. L., et al. 2015, *ApJ*, **801**, 105
- Yanny, B., & Gardner, S. 2013, *ApJ*, **777**, 91
- York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, *AJ*, **120**, 1579
- Zyla, P. A., Barnett, R. M., Beringer, J., et al. 2020, *PTEP*, **2020**, 083C01