



# Topography of the Young Galactic Disk: Spatial and Kinematic Patterns of Clustered Star Formation in the Solar Neighborhood

Emilio J. Alfaro<sup>1</sup> , Manuel Jiménez<sup>1</sup> , M. Carmen Sánchez-Gil<sup>2</sup> , Néstor Sánchez<sup>3</sup> , Marta González<sup>3</sup> , and Jesús Maíz Apellániz<sup>4</sup> 

<sup>1</sup> Instituto de Astrofísica de Andalucía (CSIC), Glorieta de la Astronomía s/n, E-18 008 Granada, Spain; [emilio@iaa.es](mailto:emilio@iaa.es), [mjimenez@iaa.es](mailto:mjimenez@iaa.es)

<sup>2</sup> Depto. de Estadística e Investigación Operativa, Universidad de Cádiz, Campus Universitario Río San Pedro s/n, E-11 510 Puerto Real, Spain  
[mcarmen.sanchez@uca.es](mailto:mcarmen.sanchez@uca.es)

<sup>3</sup> Universidad Internacional de Valencia (VIU), C/Pintor Sorolla 21, E-46 002 Valencia, Spain

<sup>4</sup> Centro de Astrobiología (CAB), CSIC-INTA, campus ESAC, Camino bajo del castillo s.n., E-28 692 Madrid, Spain

Received 2022 June 13; revised 2022 July 25; accepted 2022 August 17; published 2022 October 4

## Abstract

The accuracy in determining the spatial-kinematical parameters of open clusters makes them ideal tracers of the Galactic structure. Young open clusters (YOCs) are the main representatives of the clustered star formation mode, which identifies how most of the stars in the Galaxy form. We apply the Kriging technique to a sample of Gaia YOCs within a 3.5 kpc radius around the Sun and  $\log(\text{age}) \leq 7.5$ , as the age in years, to obtain  $Z(X, Y)$  and  $V_z(X, Y)$  maps. Previous work by Alfaro et al. has shown that Kriging can provide reliable results even with small data samples ( $N \sim 100$ ). We approach the 3D spatial and vertical velocity field structure of the Galactic disk defined by YOCs and analyze the hierarchy of the stellar cluster formation, which shows a rich hierarchical structure, displaying complexes embedded within each other. We discuss the fundamental characteristics of the methodology used to perform the mapping and point out the main results obtained in phenomenological terms. Both the 3D spatial distribution and the vertical velocity field reveal a complex disk structure with a high degree of substructures. Their analysis provides clues about the main physical mechanisms that shape the phase space of the clustered star formation in this Galactic area. Warp, corrugations, and high local deviations in  $Z$  and  $V_z$  appear to be intimately connected, in a single but intricate scenario.

*Unified Astronomy Thesaurus concepts:* [Open star clusters \(1160\)](#); [Star formation \(1569\)](#); [Catalogs \(205\)](#); [Surveys \(1671\)](#); [Milky Way disk \(1050\)](#); [Milky Way dynamics \(1051\)](#)

## 1. Introduction

The Milky Way disk is highly structured. In 2D space, the spiral morphology is the most conspicuous feature (Becker 1964, Kerr 1970, and Schmidt-Kaler 1975 are examples of earlier views), although its origin and even the number of grand design arms are still a matter of debate (see Turner 2014; Poggio et al. 2021; Martínez-Medina et al. 2022, and references therein).

Since the late 1950s, we well know that various Galactic disk constituents have spatial distributions far away from planarity. The first evidence came from the study of the HI distribution carried out to redefine the fundamentals of the Galactic coordinate system (Kerr 1958; Gum et al. 1960). Atomic hydrogen exhibited nonplanar structures at different spatial scales and Galactic locations. The outermost edge of the disk showed a warped shape that could reach amplitudes of several hundred parsecs (Gum et al. 1960; Henderson et al. 1982). In the central region, a tilted HI distribution with respect to the formal plane could be observed (Liszt & Burton 1980), and in the internal Galactic regions, some ripples in  $Z$  (the vertical distance from the formal plane) of low amplitude ( $\approx 50$  pc), with variable spatial scales of the order of a few kpc, were the most outstanding features (Dixon 1967; Henderson 1967; Kerr 1970; Varsavsky & Quiroga 1970; Quiroga 1974; Lockman 1977; Quiroga & Schlosser 1977; Spicker &

Feitzinger 1986). Similar structures also appeared to be outlined by different molecular species (see Cohen & Dent 1983; Spicker & Feitzinger 1986; Mead & Kutner 1988; Wouterloot et al. 1990; Combes 1991; Malhotra 1994, among others). The young and massive stars that delineate the spiral structure in the solar neighborhood provided the first evidence of a stellar population with a vertical distribution far from a planar symmetry. For reasons of proximity, the local arm (LA) has been the main feature that has revealed the complicated morphology shown by the young stellar components of the disk. Dixon (1967), using a sample of early B stars, defined the locus of the LA in the solar neighborhood and drew, for the first time, its vertical profile (his Figure 4), concluding that: “Spiral arms are ribbon-shaped.”

Later studies kept their focus on the LA, increasing the length of the analyzed segment and including new spiral-arm tracers. In particular, the innovative HI analysis performed by Quiroga and collaborators (Varsavsky & Quiroga 1970; Quiroga 1974; Quiroga & Schlosser 1977) not only provided a very detailed description of the wave-like vertical structure of the arms, but also pointed out some possible formation scenarios. Other star formation tracers were soon incorporated into this picture, such as fertile gases (HI, CO, molecular clouds, etc.) and young stellar objects with different masses and evolutionary states, as well as the typical signposts of feedback between the newborn objects and the leftover gas (H II regions, reflection nebulae, etc.; see Spicker & Feitzinger 1986 for a compilation of different works). Due to all these efforts, by the end of the 1980s, the spatial corrugations associated with the three spiral arms of the solar neighborhood had been analyzed



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

in great detail. Then, the question arose as to whether this phenomenology was only associated with the spiral structure, or if, on the contrary, it was a phenomenon that encompassed the entire Galactic disk. The only well-known extended structure in the solar neighborhood at that time, the Gould Belt (GB), had to be incorporated into this debate.

If we consider a Galactic area within a radius of 1 kpc around the Sun, the typical star formation tracers (originally, the brightest stars in the sky) seem to split into two stellar systems: one, the Milky Way disk; and the other, tilted around  $20^\circ$  with respect to the latter, the GB (Herschel 1847; Gould 1879; van den Bergh 1966; Taylor et al. 1987; Lindblad et al. 1997; Poppel 1997; Elias et al. 2006b, among many others). Although both the Galactic corrugations and the GB rest on, and are described by, the same set of young Population I-type objects, no study has jointly analyzed both phenomenologies. The GB is not considered today to be representative of a single and extended region of coherent star formation (Moreno et al. 1999; Sánchez et al. 2007a; Elias et al. 2009). All the properties and structural parameters estimated for the GB can be obtained considering only two star-forming regions, Orion and Sco-Cen, with different star formation modes (clustered star formation in Orion and a looser star formation pattern in Sco-Cen; Elias et al. 2009). On the other hand, Sco-Cen’s kinematics is not compatible with those expected for a coherent star-forming ring (or disk) surviving longer than a few Ma (Moreno et al. 1999). Thus, since 2009, we know that the GB is not a single star-forming complex with a coherent phase-space structure. Quoting Elias et al. (2009), “perhaps we should definitely drop the traditional hypothesis of a single, common origin for all the features of the GB, and begin to look at it as a hazardous alignment -from our point of view- of at least two of the many clumps in the hierarchical structure of the Local Arm, with different densities and star formation history. In this sense, the GB would be simply the projection over the sky of the recent star formation in the clouds close to the Sun but located far away from the fundamental Galactic plane.” This point of view has lately been adopted by Bouy & Alves (2015). Nevertheless, even considering that we are not dealing with a coherent star-forming system, its apparent morphology must also be explained by the same mechanisms that would generate the observed wobbly disk.

In 1991, a 3D map of the Galactic disk in the solar neighborhood was obtained for the first time, based on the distribution of young open clusters (YOCs; Alfaro et al. 1991, hereafter, ACCD). The map was tailored using the Kriging technique (Krige 1951; Matheron 1963), which allowed the authors to obtain a detailed 3D view of the Galactic disk with only 82 sample points. It was also the first time that the mining prospecting toolkit, named Kriging, or *Krigeage* in its French designation, was incorporated into astronomy (Lombardi 2002). Two years later, the idea of a whole corrugated disk, rather than a bunch of undulating spiral arms, was supported by the 3D analysis of the Galactic Cepheids performed by Berdnikov & Efremov (1993).

Open clusters have been widely used in the study of the structure of the Galactic disk. Prior to Gaia, there have been several catalogs and compilations of these objects, whose best-known representatives are Lynga (1982), Janes & Adler (1982), Dias et al. (2002), Kharchenko et al. (2005), and their updates. The determination of the parallaxes and proper motions of 20 stellar clusters obtained with the final Hipparcos calibration

(van Leeuwen 2009) deserves special mention and has played an important role as an observational constraint on stellar evolution theories. Few of the papers derived from these catalogs deal with the 3D structure of the young Galactic disk. The vertical structure of the III Galactic quadrant has been one of the best-studied issues (Moitinho et al. 2006; Vázquez et al. 2008). Three observational features have been reported in this quadrant: a deep and extensive depression called “Big Dent” (ACCD), hints of a southern stellar warp (Freudenreich et al. 1994), and a stellar overdensity in Canis Major (Newberg et al. 2002).

The vertical velocity field in the Milky Way has had a less rich history, mainly due to the small number of good quality kinematic data. It was after the publication of the systematic stellar radial velocity catalogs (RAVE, Steinmetz et al. 2006; APOGEE, Allende Prieto et al. 2008; LAMOST, Cui et al. 2012; GES, Gilmore et al. 2012; GALAH, De Silva et al. 2015; among others) and the Gaia releases (Gaia Collaboration et al. 2016, 2018, 2021) that the kinematic analysis could be extended to significant areas of the Galactic disk.

Henderson (1968), analyzing the velocity variations of the 21 cm H I line, claimed the detection of motions departing from circular rotation, which could also be interpreted as variations of the velocity vertical component. These results were subsequently confirmed and refined by other authors (e.g., Yuan & Wallace 1973; Lockman & Bania 1981; Feitzinger & Spicker 1985). Comerón (1999) found a systematic variation of the vertical component of the velocity in the region associated with the GB, pointing to an oscillatory motion around an axis nearly parallel to the solar Galactocentric radius. This study was mainly based on OB stars from the Hipparcos catalog (Perryman et al. 1997).

The corrugations of the Galactic disk, together with the warping of the outermost regions, appear to be universal phenomena that are observed in most edge-on spiral galaxies (e.g., Battaner et al. 1990; Sánchez-Saavedra et al. 1990; Florido et al. 1991, 1992; Matthews & Uson 2008, among others). There is also evidence of the undulating behavior of the vertical velocity component in some external disks (Alfaro et al. 2001; Sanchez-Gil et al. 2015). It seems clear that both structures represent projections on different subspaces of the same phenomenon, affecting the entire phase space. However, this connection between spatial structure and velocity field has not been easy to establish, due to the lack of quality data for an adequate sample of tracers, either in distance or velocity. Consequently, studies were limited to a few local Galactic regions—the GB, essentially—before the advent of Gaia.

A detailed analysis of the first topography of the young Galactic disk (ACCD, Alfaro et al. 1992) led the authors to propose that: “The three-dimensional spatial distribution of the Young Population I-type objects can help determine more clearly the structure and size of the young star-gas complexes and give some clues as to the origin and development of these large clouds and to the possible mechanisms driving the large-scale star formation in our Galaxy.” This is the main objective of our work, to determine the spatial structure and the vertical velocity field in a wide region of the young Galactic disk around the Sun. That is, to establish observational constraints in phase-space subspaces that enable a link between the formation of stellar clusters and the generation of a wobbly disk. We can now do this, because of Gaia.

The Gaia astrometric mission has revolutionized most astronomical fields and, in particular, our view of the Milky Way. Several teams have focused on the study of the Galactic open cluster system. They have cataloged the actual clusters and measured their main physical variables, including, in some cases, the complete kinematic information (Soubiran et al. 2018; Cantat-Gaudin et al. 2020; Dias et al. 2021; Tarricq et al. 2021). YOCs are the main representatives of the clustered star formation mode, which identifies the way in which most of the stars in the Galaxy form (Lada & Lada 2003; Portegies Zwart et al. 2010). Nevertheless, both the proportion of the stars born in bounded clusters, and how they migrate to the disk, are still matters of debate (Dinnbier et al. 2022; Maíz Apellániz et al. 2022; Pang et al. 2022, and references therein).

We choose Kriging (Matheron 1963) to obtain the functions  $Z(X, Y)$  and  $V_Z(X, Y)$ , as the mathematical method best suited to the characteristics of the problem. If we consider the star clusters as probes of the 3D disk structure, we can infer that the observational sample is more similar to what a gold deposit would reveal than to another ore that is more reproducible by continuous and derivable functions. In other words, the distribution of the YOCs appears to be a realization of a stochastic process, analysis of which requires the right mathematical tools. As has been shown in previous works (ACCD), Kriging is able to provide reliable results, even when working with not-too-large samples (Oliver & Webster 2014;  $N \approx 100$ ).

With this approach, we analyze the following items:

1. The hierarchy of the stellar cluster formation.
2. The 3D spatial distribution of the star formation in clusters and comparisons with other young stellar objects.
3. The vertical velocity field of the YOCs ( $\log t \leq 7.5$ ;  $t$  in years) in an area of  $\sim 40 \text{ kpc}^2$  around the Sun.
4. The role of the inner solar neighborhood within the larger spatial area analyzed in this work.
5. The warp as drawn by the YOCs.
6. The physical mechanisms shaping the 3D Galactic disk structure.

The paper is organized into six sections, the first being this introduction. Section 2 describes the data sets and the physical variables that we use in this work. The Kriging technique—in particular, the algorithm employed in our estimations—is described and discussed in Section 3. Section 4 is devoted to obtaining a view of the hierarchical clustering structure and the 3D spatial pattern. In addition, we obtain the vertical velocity field drawn by the YOC population. In Section 5, we discuss our results in relation to the scientific objectives listed in this introduction; finally, Section 6 summarizes our conclusions.

## 2. Data

### 2.1. Input Data

Stellar open clusters exhibit several advantages as tracers of the Galactic disk structure. Their properties can generally be estimated with greater precision compared to isolated stars (Becker 1964). The outstanding accuracy in parallax and proper-motion measurements achieved by the Gaia mission has enabled the elaboration of catalogs of such objects in terms of distance and age at an unprecedented quality. We rely on the catalogs published by Cantat-Gaudin et al. (2020) and Tarricq et al. (2021), which primarily leverage Gaia data release 2

(DR2), to assemble a set of clusters within a radius  $r \leq 3.5 \text{ kpc}$  around the Sun and age  $\leq 10^{7.5} \text{ yr}$ . Thus, these selection criteria result in a sample of YOCs very representative of the places where they were born.

The catalog presented in Cantat-Gaudin et al. (2020) represents the most complete collection of stellar clusters detected and analyzed from the Gaia DR2 individual stars, following a homogeneous procedure in the selection of the cluster members.<sup>5</sup> Both the five astrometric variables and the three Gaia photometric bands are employed to list the probable members and determine the physical properties of the cluster. Their methodology consists of a data-driven approach, based on the traditional isochrone fitting of the observational color–magnitude diagram (CMD), by means of an artificial neural network (ANN). This ANN inputs a flattened representation of the cluster’s CMD, along with two additional quantities estimated from it and the median parallax, and learns the correlation between these inputs and the targets: the cluster’s age, distance modulus, and extinction. The network is trained on a set of 347 well-known clusters that are considered as reference points (Bossini et al. 2019). Then, it is applied to a set of 2017 clusters identified in previous studies (Liu & Pang 2019; Cantat-Gaudin & Anders 2020), using the UPMASK algorithm (Krone-Martins & Moitinho 2014) and involving photometric and astrometric data. At the end of the process, they compile a catalog<sup>6</sup> of 1867 clusters with reliable parameters, including their position in Galactic Cartesian coordinates, age, parallax, and extinction. However, data about their proper motions are not supplied.

In contrast, the catalog produced by Tarricq et al. (2021) constitutes the largest collection of open clusters including radial velocity (RV) measurements, as a result of a study that was carried out to investigate the kinematic behavior of the open cluster population. These RV data are primarily gathered from Gaia DR2, although the number of stars and measures per star are increased by considering nine additional RV surveys (Tarricq et al. 2021). To compute the RVs, a weighted mean of the RV of each star is calculated, in relation to the different measures in the catalogs considered and their errors (Soubiran et al. 2013, 2018). The cluster membership is taken directly from the list of 2017 clusters by Cantat-Gaudin et al. (2020), referred to above, from which they also get the mean positions, distances, proper motions, and ages. As a result, the final catalog<sup>7</sup> includes complete kinematic information for 1382 open clusters, signifying a great improvement with respect to the previous largest compilations (Kharchenko et al. 2013; Dias et al. 2002).

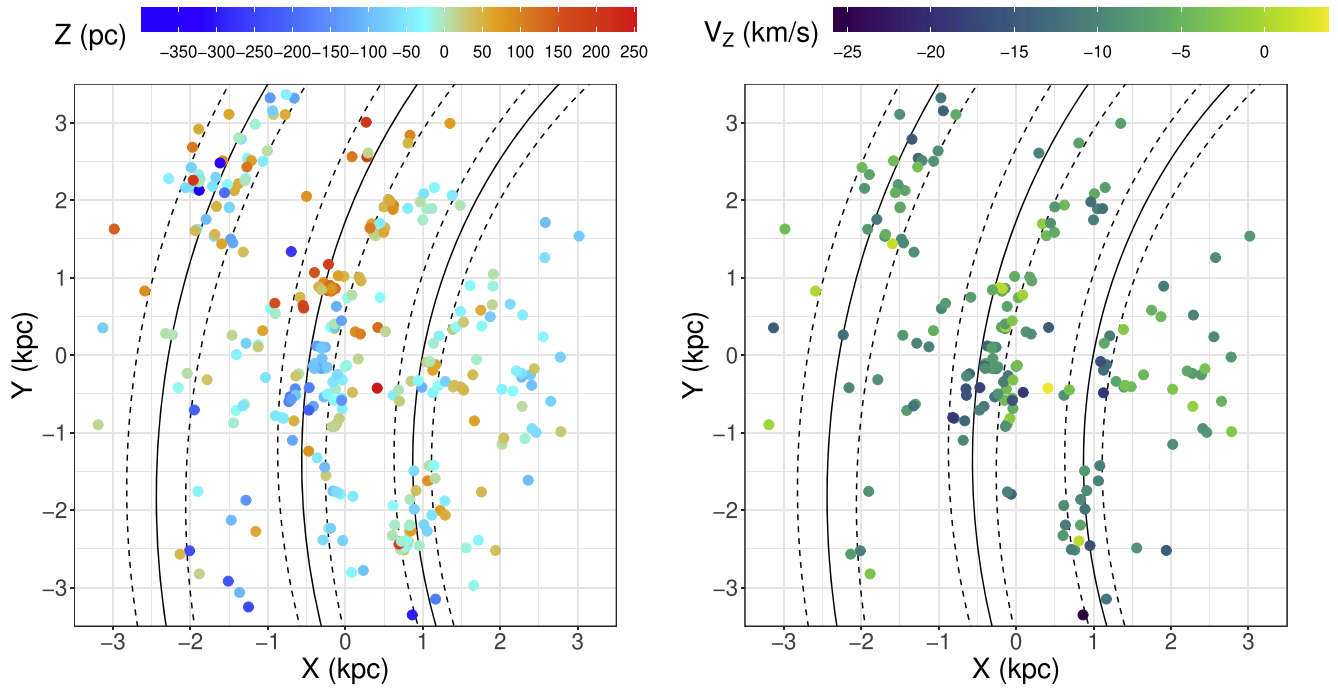
We perform several steps, beginning with these two collections, to compose the final sample of YOCs for our study. First, we apply the criteria detailed above about distance and age separately, yielding a total of 301 and 166 YOCs from the Cantat-Gaudin and Tarricq catalogs, respectively. Then, we crossmatch these two samples, incorporating into the Tarricq sample a set of 143 clusters from Cantat-Gaudin that were not previously present in the Tarricq sample. Of these, 108 have

<sup>5</sup> The compilation of Dias et al. (2021) lists more clusters than that of Tarricq et al. (2021). We choose the latter as the basis of our study because the cluster members are selected using the same method, in a homogeneous way.

<sup>6</sup> The cluster catalog with a complete list of members is available at <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/640/A1>.

<sup>7</sup> The cluster catalog with a complete list of members is available at <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/647/A19>.





**Figure 1.** The G-YOC sample from the merging of the Cantat-Gaudin et al. (2020) and Tarricq et al. (2021) catalogs, in Galactic Cartesian coordinates. The selection criteria are  $|r| \leq 3.5$  kpc and  $\log(\text{Age}) \leq 7.5$ , resulting in 309 YOCs (left), of which 183 have their vertical velocity component measured (right). The color codes for  $Z$  (pc) and  $V_z$  ( $\text{km s}^{-1}$ ) are shown at the top. The  $X$ - and  $Y$ -axes take positive values toward the Galactic center and the Galactic rotation directions, respectively. The Sun is at  $(0, 0)$ . The shown spiral arms are, from left to right: Perseus, LA, and Sagittarius. The spiral-arm parameters are taken from Castro-Ginard et al. (2021), with the width of the fixed arms based on Reid et al. (2014).

velocity measures that are obtained from the catalog<sup>8</sup> provided in Soubiran et al. (2018). This results in a final sample of 309 YOCs with positions and ages, of which 183 also have a vertical velocity component.<sup>9</sup> Figure 1 displays our final YOC sample on the Galactic plane. From now on, we will refer to this sample as the Gaia-YOC (G-YOC) sample.

## 2.2. Our Reference System

We choose the formal Galactic plane as the main reference plane. The Cartesian space and velocity coordinates are referenced to the Sun, such that  $\mathbf{r}_\odot \equiv (X_\odot, Y_\odot, Z_\odot) = (0, 0, 0)$  pc and  $\mathbf{V}_\odot \equiv (V_{\odot X}, V_{\odot Y}, V_{\odot Z}) = (0, 0, 0)$   $\text{km s}^{-1}$ . We do not include any additional information to determine the maps  $Z(X, Y)$  and  $V_z(X, Y)$ . The reason for this decision is that we do not need to estimate additional constants, such as  $R_\odot$ ,  $Z_\odot$ , or the solar motion, whose determination would depend in each case on the chosen age group, the area of the Galactic plane studied, and the quality of the data used. This is an inescapable fact if, as has already been shown in different studies, the Galactic disk is a nonstationary system in phase space (e.g., Widrow et al. 2012; Antoja et al. 2018; Bennett & Bovy 2019). Therefore, all the information gathered from other published works and catalogs is transformed to Galactic coordinates, with the Sun as the origin of the system, if it formerly referred to other Galactic reference centers or other sets of coordinates.

We also select other reference lines in the  $(X, Y)$  plane that represent typical features of the Galactic disk, such as spiral arms, and other peculiar loci that have previously been detected

<sup>8</sup> This catalog is available at <http://cdsarc.u-strasbg.fr/viz-bin/qcat?J/A+A/619/A155>.

<sup>9</sup> There exist overlaps of 158 and 91 clusters in the two crossmatches performed, concerning positions and vertical velocities, respectively.

**Table 1**

Parameters of the Spiral Arms, According to Castro-Ginard et al. (2021), and the Cepheus Spur Linear Fit, Based on the Study Presented in Pantaleoni Gonzalez et al. (2021)

Spiral Arms				
Arm	$R_{ref}$ (kpc)	$\Phi_{ref}$ (deg)	$\psi$ (deg)	Width (kpc)
Perseus	10.88	-13.0	9.8	0.38
LA	8.69	-2.3	8.9	0.33
Sagittarius	7.10	3.5	10.6	0.26
Cepheus Spur				
$X_{min}$ (pc)	$X_{max}$ (pc)	Equation (pc)		
-2157.2	202.0	$Y = 0.6528X + 995.1418$		

in the region, such as the Cepheus spur, introduced in Pantaleoni Gonzalez et al. (2021). There are a great variety of definitions (different formula parameters and even different mathematical functions) for the three main spiral arms within the analyzed area. These differences are due to the variety of spiral tracers and the different distance estimation methods used (e.g., Hou 2021; Poggio et al. 2021; Xu et al. 2021; Vallée 2022, and references therein). In this work, we choose the spiral arms defined by Castro-Ginard et al. (2021), from the joint population of YOCs and maser tracers. The Cepheus spur has been fitted by us as a linear segment, according to the previous information available (Figure 7 in Pantaleoni Gonzalez et al. 2021). The equation for the spiral arms, as well as the range of values in which they and the Cepheus spur are defined, are shown in Equation (1) and Table 1, respectively. In this logarithmic model,  $R$  and  $\Phi$  represent the Galactocentric radius and azimuth along the arm, respectively,

$(R_{\text{ref}}, \Phi_{\text{ref}})$  are the coordinates of a reference point, and  $\psi$  is the pitch angle of the arm. We consider a constant width of the spiral arms for their graphical representation, although there is evidence that the width varies along the arm (e.g., Vallée 2021). The equation for the spiral arms is:

$$\ln \left[ \frac{R}{R_{\text{ref}}} \right] = -(\Phi - \Phi_{\text{ref}}) \tan \psi . \quad (1)$$

These 2D lines are mere references, to enable us to compare our results with those of previous works, but we are far from considering that the spatial and kinematic structure of the young Galactic disk is limited to a few linear features on the plane. We are faced with a 3D structure far from planarity that affects the entire disk, regardless of the stellar population that we are analyzing (Widrow et al. 2012). On the other hand, the defining features of Galactic spiral arms, including their nature and origin, are not yet fully understood (Turner 2014; Lepine 2018; Poggio et al. 2021; Martinez-Medina et al. 2022; Vallée 2022). The characterization of spiral arms in the vicinity of the Sun does not fall within the scope of this paper. However, we will use the main 2D features provided by other authors to highlight some aspects of the 3D pattern of the young Galactic disk and to analyze its possible origins.

### 3. Methods

#### 3.1. Kriging Fundamentals

The genesis of Kriging goes back to the mid-twentieth century, when Danie G. Krige aimed to predict the location of ore grades by considering a spatially correlated data sample in the gold mines of South Africa (Krige 1951). Later in the 1960s, George Matheron developed the theoretical foundations for the method (Matheron 1963), and since then it has been widely implemented in many fields, such as engineering, mining, geology, meteorology, and remote sensing, to name just a few (Chilès & Delfiner 2000; Simpson et al. 2001). Now Kriging is recognized as an extensive family of estimation methods for multidimensional stochastic processes (Webster & Oliver 2007), although it has barely been exploited in astronomy (Pastorello et al. 2014). It provides the best linear unbiased estimator (BLUE), making optimal use of existing knowledge by exploring the way that a target variable varies in space through a variogram model (Oliver & Webster 2014). The Kriging technique is closely connected with Gaussian process regression (GP), at least in its final objectives, the latter being more extended in astronomy (e.g., Rasmussen & Williams 2005; Nelson & Widrow 2022; Widmark et al. 2022). Kriging and GP are families of interpolation methods that have similarities and subtle differences. Mathematically, GP can be considered as a generalization of Kriging into higher-dimensional space, and, in some cases, both methods can provide quite similar mathematical solutions (see Feigelson & Babu 2012 and Cui et al. 2021 for deeper discussions of both techniques).

Let  $z(\mathbf{r})$  be a real function defined in some region of the plane  $\mathbb{R}^2$ , with known values at a certain number of points  $\mathbf{r}_i(x_i, y_i)$  of a sample set  $S$ , which is regarded as a single realization of a random process  $Z(\mathbf{r})$ . The basic problem is to achieve a valid approximation  $z(\mathbf{r})$  of the function  $Z(\mathbf{r})$  for the points not belonging to  $S$ , assuming that some sort of spatial correlation underlies the points in  $S$ . In its simpler form, a Kriging estimate

is a weighted average of the sample data in their neighborhood. However, more elaborated versions have been devised, to take into account general trends in  $S$  or several limitations on its knowledge. In addition, a distinctive characteristic of Kriging is that it enables the quantification of the goodness of the prediction, providing a variance  $\sigma_z^2$  associated with the estimated function  $z(\mathbf{r})$ . For our purposes here,  $S$  will be the G-YOC sample introduced above, whose  $(X, Y, Z)$  and  $(V_Z, X, Y)$  coordinates are well determined. The goal is to obtain an estimation of  $Z(X, Y)$  and  $V_Z(X, Y)$  on a rectangular grid that covers the area investigated, in order to outline the 3D structure of the solar neighborhood and the vertical velocity field.

Under the assumption of stationarity, we can decompose  $Z(\mathbf{r}) = m + e(\mathbf{r})$ , where  $m$  is the mean of the process,  $m = E[Z(\mathbf{r})]$ , and  $e(\mathbf{r})$  is a zero-mean random process, whose covariance  $C(\mathbf{h}) = E[e(\mathbf{r})e(\mathbf{r} + \mathbf{h})]$  is known, with  $E$  being the expectation and  $\mathbf{h}(\mathbf{r}_i, \mathbf{r}_j)$  the vector separation between pairs of sample points. Due to the aforementioned stationarity,  $m$  does not depend on  $\mathbf{r}$  and  $C(\mathbf{h}) = C(h)$  only depends on the distance (isotropy). However, a weaker hypothesis originally introduced by Matheron is needed in practical situations, known as the intrinsic model (Matheron 1965). Using this,  $E[Z(\mathbf{r} + \mathbf{h}) - Z(\mathbf{r})] = 0$  and the variance yields  $\text{Var}[Z(\mathbf{r} + \mathbf{h}) - Z(\mathbf{r})] = 2\gamma(h)$ . The resulting  $\gamma(h)$  is the variogram of the random process, which is assumed to measure the spatial correlation in the actual realization of  $Z(\mathbf{r})$ . Generally, it is estimated using the method of moments (Matheron 1965) as:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2p(\mathbf{h})} \sum_{j=1}^{p(\mathbf{h})} \{Z(\mathbf{r}_j) - Z(\mathbf{r}_j + \mathbf{h})\}^2, \quad (2)$$

with  $p(\mathbf{h})$  being the number of point pairs separated by a distance  $\mathbf{h} \in \mathbb{R}^3$ , or  $h \in \mathbb{R}$  in the particular case under isotropy. When  $S$  shows a general trend—that is, a gradual variation in space—the method can be generalized, considering that  $m = m(\mathbf{r})$  is responsible for the trend. This is the so-called universal Kriging (UK). In this case,  $m = E[Z(\mathbf{r})]$  is a functional drift and  $e(\mathbf{r})$  remains as an intrinsic zero-mean random process. That is, in the UK model, such a trend  $Z(\mathbf{r}) = m(\mathbf{r}) + e(\mathbf{r})$  is modeled as a linear function,

$$Z(\mathbf{r}) = \sum_{j=0}^p f_j(\mathbf{r})\beta_j + e(\mathbf{r}) = F\beta + e(\mathbf{r}), \quad (3)$$

based on  $p$  known predictors or spatial regressors  $f_j(\mathbf{r})$ , evaluated at each observation  $\mathbf{r}_i$  (also known as the design matrix  $F = (f_1(\mathbf{r}), \dots, f_p(\mathbf{r}))$ , with  $f_j(\mathbf{r}) = (f_j(\mathbf{r}_1), \dots, f_j(\mathbf{r}_n))$ , and  $p$  unknown regression coefficients  $\beta = (\beta_1, \dots, \beta_p)^T$ ). The most common strategy for dealing with it is to first estimate the drift, by fitting a polynomial curve, and then obtaining the residuals  $e(\mathbf{r}')$ , by subtracting the curve values from the sample data. Given that  $e(\mathbf{r}')$  is expected to be an intrinsic process, the usual Kriging estimator is calculated for the residuals.

#### 3.2. The Variogram Estimation

The variogram analysis provides a useful tool for summarizing the spatial data and measuring the spatial dependence. Moreover, its main contribution is related to the statistical inference of spatially correlated random variables to estimate the value of the spatial variable at an unsampled location. The Kriging approach differs from classical regression, in that local

features can affect the solution. For that reason, a Kriging estimate considers a weighted mean, bearing in mind that some measurements in the vicinity of the location investigated are more closely related to the actual unknown value than farther ones. Variogram estimation plays a decisive role in the Kriging approach, and is commonly used to find the optimal values of the weights.

The variogram estimation is split into two stages: (i) the computation of the experimental variogram from the sample set  $S$  (Oliver & Webster 2014), the so-called “empirical variogram estimation,” with subsequent fitting with some theoretical model  $\gamma(h)$ ; and (ii) the application of the previous variogram estimator to the data to make predictions.

Under the intrinsic model (see Equation (2)), the distance  $h$  is split into a certain number of bins, resulting in an ordered set of  $(h, \gamma(h))$  points. However, the underlying model is assumed to be continuous in all  $h$ . We then get a variogram estimator by fitting the empirical variogram with some theoretical model. We select the theoretical smooth curve that best fits the experimental variogram at this stage. This function must belong to a family of valid variograms and capture the available data’s underlying spatial dependence. More details of the variogram models and variogram fitting, together with the theoretical model selection, by means of the cross-validation comparison and the WSSE statistics, or the weighted sum of the squared errors of the fitted model, can be found in Appendix appendix.

Figure 15 shows the variogram fitting of the spatial  $Z$  component for the G-YOC sample. The black dots represent the experimental variogram. We test four theoretical variogram models  $\gamma(h)$ : the Spherical (A4) and Exponential (A5) variogram models, which are the most widely used, along with the Gaussian (A6) and Wave (A7) models. The latter was also considered because of its atypical irregular behavior, taking into account the expected corrugated nature of the  $Z$  and  $V_Z$  variables.

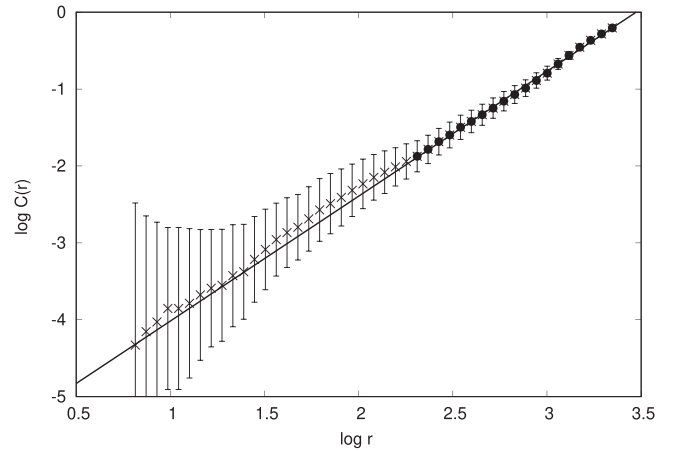
Once we get the variogram estimator, which determines the spatial dependency or autocorrelation of our data, we can estimate the  $Z(X, Y)$  and  $V_Z(X, Y)$  functions using the fitted model. Therefore, the last stage of the Kriging approach is to use the data to make predictions and create a continuous surface of the phenomenon. Predictions are made at each location of a spatial grid around the studied area, based on the fitted semivariogram model and the spatial arrangement of the closer input data.

## 4. Results

In this section, we present the results from the Kriging analysis of the 4D phase-space subspace, inferring both the 3D spatial distribution and the vertical velocity field of the young Galactic disk. We have previously performed a study of the hierarchical spatial distribution of the G-YOC sample.

### 4.1. Clustering of Clusters

Star formation tends to be distributed in space, forming groups that in turn contain other smaller groupings, in a cascade of sizes and ages (Efremov & Elmegreen 1998; Elmegreen et al. 2000; Elmegreen 2011). In this section, we analyze the spatial hierarchy of the distribution of star clusters. To do so, we obtain the fractal dimension in  $(X, Y)$  2D subspace and perform a clustering analysis on the 3D spatial coordinates.



**Figure 2.** The calculated correlation integral  $C(r)$  for the data used in this work. The error bars represent the standard deviations  $\sigma_C$ . The solid line is the best linear fit in the range of reliable values indicated by the solid circles, i.e., the range for which  $\sigma_C \lesssim C(r)$ .

#### 4.1.1. Fractal Dimension of the Distribution of the G-YOC Sample

One way to characterize the spatial distribution of the YOCs and their associations is through the fractal dimension, which gives a simple and objective measurement of the degree of clumpiness of the distribution. We calculated the so-called correlation dimension (Grassberger & Procaccia 1983) using a previously developed algorithm (Sánchez et al. 2007a), which avoids boundary effects and finite data problems at small scales, and, consequently, yields unbiased results even for relatively small sample sizes. The correlation dimension is derived by calculating the correlation integral  $C(r)$ , that is, the probability of finding a point within a circle of radius  $r$  centered on another point. If the points are distributed obeying a fractal geometry of dimension  $D_c$ , then the relation  $C(r) \sim r^{D_c}$  holds, and  $D_c$  can be estimated as the slope of the best linear fit in a log–log plot. Figure 2 shows the obtained result. The 2D correlation dimension for the sample gives  $D_c = 1.62 \pm 0.05$ , where the uncertainties are estimated from bootstrapping. This fractal behavior extends over a range of scales from 200 to 2500 pc, and the slope changes that are observed below  $r \sim 200$  pc can be attributed to random fluctuations in  $C(r)$ , and not to a real transition between two different physical regimes, such as has been reported for the distribution of young stars, YOCs, and H II regions in other galaxies (Odekon 2008; Sánchez et al. 2010; Menon et al. 2021).

The degree of clustering of the distribution of star-forming regions may vary among galaxies. Fractal dimensions in the range  $1.5 \lesssim D_c \lesssim 2.0$  have been obtained for the distribution of H II regions in disk galaxies (Sánchez & Alfaro 2008), depending on properties such as brightness, size, or age. These distributions are the consequence of physical processes acting at large scales, that is, Galactic motions on the disk. At smaller scales, where our analysis is relevant, turbulence is probably the main mechanism controlling the underlying structure where coherent star formation occurs. Our result is consistent with the recent detailed analysis of the distribution of  $\sim 2500$  young stellar structures in the Large Magellanic Cloud by Miller et al. (2022), who derived a 2D fractal dimension of around 1.5–1.6 in the range of spatial scales from 10 pc to 700 pc. The value  $D_c = 1.62 \pm 0.05$  implies that the corresponding 3D fractal dimension should be  $D_f = 2.5 - 2.6$  (see Table 1 in Sánchez & Alfaro 2008), which is similar to the range of values



$D_f \sim 2.5\text{--}2.7$  obtained from emission maps of several spectral lines for different molecular clouds in the Milky Way (Sánchez et al. 2005, 2007b) and for early-type stars in the GB (Sánchez et al. 2007a). In addition, this value is in agreement with the simulations of compressively driven supersonic turbulence reported by Federrath et al. (2009). Overall, the obtained result supports the idea that young newborn stars and clusters should, on average, reflect the same spatial structure of the interstellar medium from which they were formed.

The fractal structure can be observed over a range of spatial scales between 200 and 2500 pc. We analyze below the hierarchical groupings of the 3D G-YOC sample.

#### 4.1.2. Clustered Star Formation Hierarchy: Complexes and Supercomplexes

We use the density-based tool OPTICS (Ordering Points To Identify the Clustering Structure; Ankerst et al. 1999) to analyze the hierarchical spatial substructure of the 3D Galactic Cartesian coordinates ( $X, Y, Z$ ) of our sample. Despite the fact that it can be used to retrieve clustering, and has typically been used that way in astrophysics (Costado et al. 2017; Cánovas et al. 2019; van Terwisga et al. 2022), we want to highlight that, strictly speaking, OPTICS is not a clustering tool, but a density-based reordering algorithm. OPTICS can be considered as an extension of DBSCAN (Density-Based Spatial Clustering of Applications with Noise; Ester et al. 1996). We briefly explain the main characteristics of both algorithms, and refer the reader to González et al. (2021) for a more detailed, but intuitive, explanation of how DBSCAN works; to Cánovas et al. (2019) for a global comparison of density-based clustering algorithms in an astrophysical context; and to the original papers by Ester et al. (1996) and Ankerst et al. (1999) for complete and rigorous descriptions of both algorithms. In essence, DBSCAN decides whether a point belongs to a cluster or to noise, based on whether a minimum number  $N_{\min}$  of neighbors is within a radius of size  $\epsilon$  around that point. The parameters  $N_{\min}$  and  $\epsilon$  define the typical scale and density of the clusters retrieved by the method, which limits the analysis when considering samples where structure is present at several scales and densities. OPTICS overcomes this limitation by defining the ordering of the points according to the concept of the reachability distance between the points. For a fixed  $N_{\min}$ , this distance can be intuitively understood as the  $\epsilon_{N_{\min}}$  needed for both points to be in the same DBSCAN cluster. The main product of the algorithm is the reachability plot, where the reachability distance of each point is shown against the ordering index provided by OPTICS. The reachability plot gives a complete and intuitive visual description of the clustering structure of a sample, since clusters and dense subsets are seen as dents or valleys in the plot.

This view of clusters as valleys in the reachability plot allows for the retrieval of the clusters in the sample: if a specific reachability distance  $\epsilon$  is fixed, the DBSCAN clusters with  $N_{\min}$  and  $\epsilon$  can be directly obtained from the plot as the valleys deeper than  $\epsilon$ , i.e., they are characterized by reachability distances smaller than  $\epsilon$ . The original paper by Ankerst et al. (1999) proposes a different possibility for cluster retrieval, whereby the clusters are separated based on the steepness of the slopes of the valleys formed in the reachability plot, instead of their depth, fixing a parameter  $\xi \in (0, 1)$  that is associated with the difference in density between adjacent points. As explained in the original paper, using this procedure, clusters are features

in the reachability plot that start with a downward slope, end with an upward slope, and fulfill the requirements for both a minimum number of points and a minimum density increment, with respect to their environment.

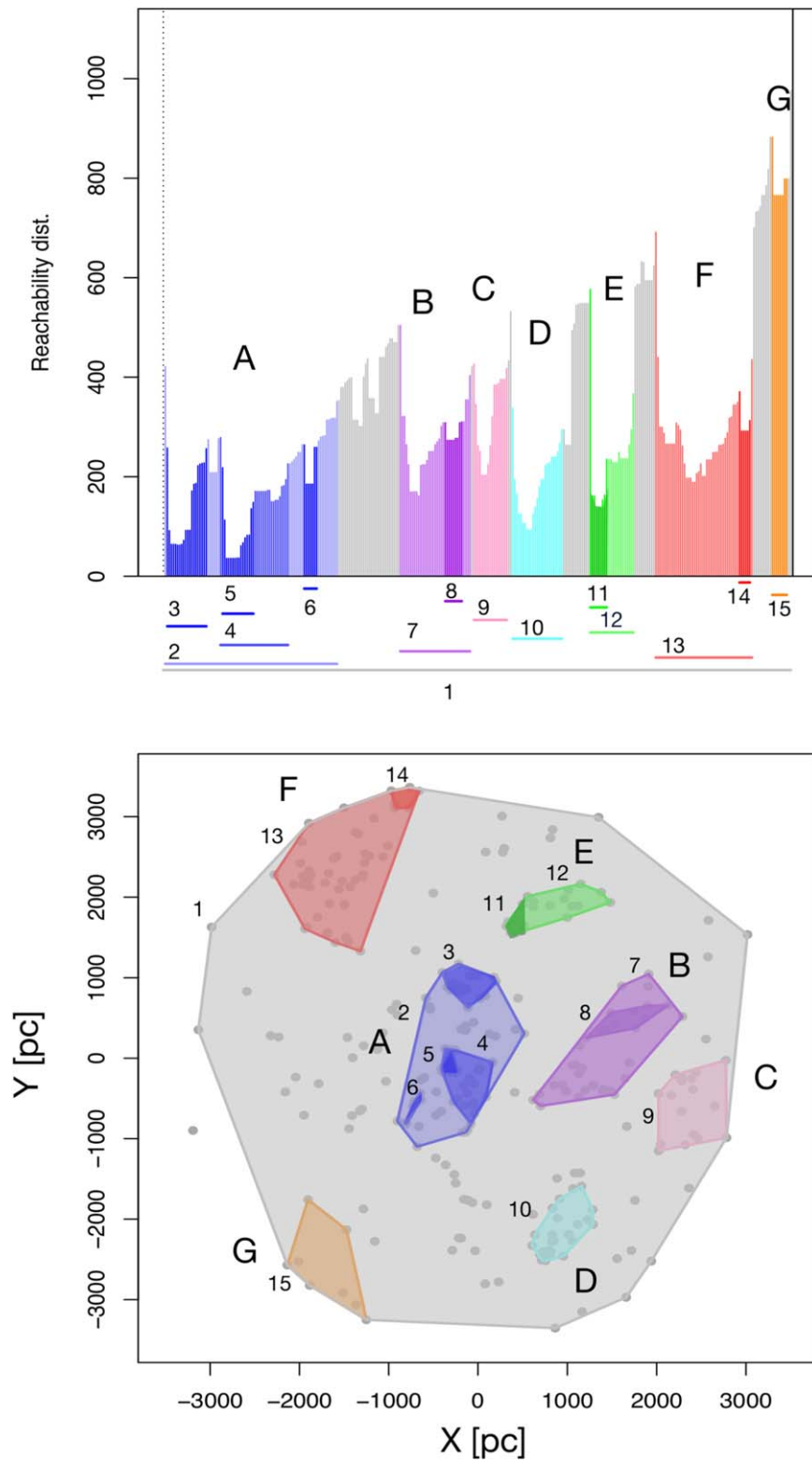
In this work, we use the  $\xi$  strategy, since it takes advantage of the multiscale approach of OPTICS, allowing for hierarchical structures to emerge, with groupings potentially embedded within each other. The main results of our analysis are shown in Figure 3 and Table 2, and reveal the clustering hierarchy of the YOCs. If we define the hierarchical level of a cluster recursively, with level 0 clusters not containing any other cluster, and clusters of level  $l$  containing clusters of level  $l-1$ , the clustering obtained in this work reaches level 3, corresponding to the entire distribution. We have fixed  $N_{\min} = 6$  as a reasonable balance between noise and the detection of small structures, and we have explored the parameter  $\xi$ , choosing  $\xi = 0.09$ , which prioritizes the significance of the structures detected instead of the richness of hierarchy. Additionally, for further confirmation of the robustness of the retrieval, we have also run OPTICS with  $N_{\min} = 8$  and confirmed that all the main families in Table 2 are also present in the corresponding reachability plot, while other dents (like those between families A and B in the upper plot of Figure 3) are smoothed out.

The top panel of Figure 3 shows the reachability plot for the G-YOC sample. The plot shows numerous valleys, some embedded within each other, revealing the presence of a hierarchical structure. The bottom panel of Figure 3 shows the hierarchy of the groupings found in the Galactic plane, plotting only the ( $X, Y$ ) variables, although the hierarchical distribution is based on three spatial Cartesian coordinates. The results from OPTICS show that the G-YOC population is, as already suggested by the fractal analysis in the previous section, structured at various densities and scales, showing both global and local hierarchical clustering all over the region sampled.

In Figure 4, we explore some of the characteristics of the groupings, in relation to their hierarchical level. The top panel shows the size (measured as its 3D diameter) of each grouping against its level. All the structures below level 2 are well within the fractal spatial scale of 200–2500 pc. We remind the reader that the only structure at level 3 is almost the complete sample, and thus it has a diameter corresponding to the 3.5 kpc initial radius. We observe a clear increasing trend in the median size with the hierarchical level. A similar trend can be found in the population of the groupings, shown in the lower panel of Figure 4.

Some aspects of interest should be noted: (a) the main groupings coincide with the overdensities of massive OB stars (sample M; Pantaleoni Gonzalez et al. 2021) and with those detected from upper main-sequence stars, young Galactic clusters, and Cepheid stars younger than 100 Ma (Poggio et al. 2021) in the common area analyzed; and (b) there is full agreement between the hierarchical stellar clumps classified in this work and the stellar complexes and supercomplexes detected from the 2D distribution of YOCs, OB associations, and H II regions by Efremov & Sitnik (1988, hereafter, ES88).

If we compare the position and size of the supercomplexes obtained from our analysis with the distribution of the young stars' filaments from Kounkel et al. (2020; top panel of their Figure 1, corresponding to  $7.0 < \log(\text{age}) < 7.7$ ), we see many similarities and some differences. The supercomplex A is practically the LA central grouping and its spine, defined in Kounkel et al. (2020) as the locus perpendicular to adjacent



**Figure 3.** Top: reachability plots obtained by OPTICS. Bottom: convex hulls of the structures obtained projected onto the Galactic plane. Each separate complex/supercomplex is shown in a different color, and tagged according to Table 2.

filaments, which connects the lower-hierarchy complexes within A. YOCs, and looser and extended concentrations of young stars, show a similar pattern in the LA. The same happens in the Sagittarius region, and in the area associated with the Scutum arm, although the groupings (B and C) of the G-YOC sample appear to be better separated than the structures

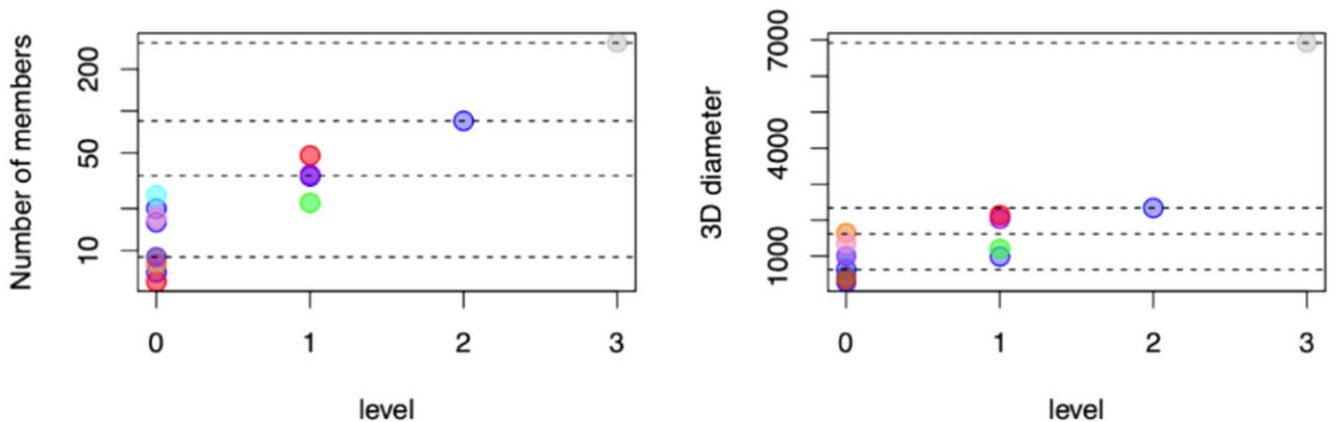
of Kounkel et al. (2020) in the same area. In contrast, the lower segment of the Perseus arm, in the III Galactic quadrant, appears to be richer in filaments and other looser OB groupings than in YOCs. The Cygnus region, on the other hand, defined by complex E, shows a high concentration of clusters, as does Carina, too (complex D). Figure 1 (top panel) in Kounkel et al. (2020)



**Table 2**  
Central Cartesian Coordinates and 3D Diameters of the Hierarchical Clumps Retrieved by OPTICS

Family	Id	$N$	Level	$X$ (pc)	$Y$ (pc)	$Z$ (pc)	3D Diam. (pc)	ES88 Name
A	2	85	2	-250.6	42.7	-22.2	2338.5	Local, 4 (10,8)
	3	20	0	-157.5	899.4	72.1	626.0	
	4	34	1	-201.9	-242.2	-70.8	993.0	4
	5	16	0	-319.5	-109.7	-98.9	294.6	4
	6	7	0	-720.6	-611.4	-120.7	425.0	
B	7	35	1	1394.7	40.7	-5.4	2038.2	1-I, 2-I
	8	9	0	1673.4	484.4	-20.5	1006.2	2-I
C	9	17	0	2349.4	-585.8	-36.8	1357.5	17
D	10	25	0	895.3	-2207.3	1.4	1019.	15-IV, 16
E	11	9	0	415.4	1651.8	36.2	390.2	3
	12	22	1	730.5	1833.2	17.6	1205.4	3-3*
F	13	48	1	-1510.7	2326.8	-30.8	2142.3	7-II, 6-II, 5
	14	6	0	-839.8	3228.0	-52.5	373.8	
G	15	8	0	-1691.7	-2630.3	-112.6	1642.3	13*

**Note.** Cluster 1, of level 3, which encompasses almost the entire sample, is not shown. The correspondences between the stellar complexes and supercomplexes listed by ES88 and those detected in this work are detailed in the last column. Each separate hierarchical family is identified with a capital letter, in accordance with Figures 3 and 5.  $N$  is the number of YOCs in each group. The rest of the column names are self-explanatory.



**Figure 4.** Left: the number of YOCs in each structure, as a function of their hierarchical level. Right: 3D diameters (pc) of the structures found in this work as a function of their hierarchical level. In both panels, the color code indicates the hierarchical family of the structure, and the horizontal dashed lines correspond to the median values for each level.

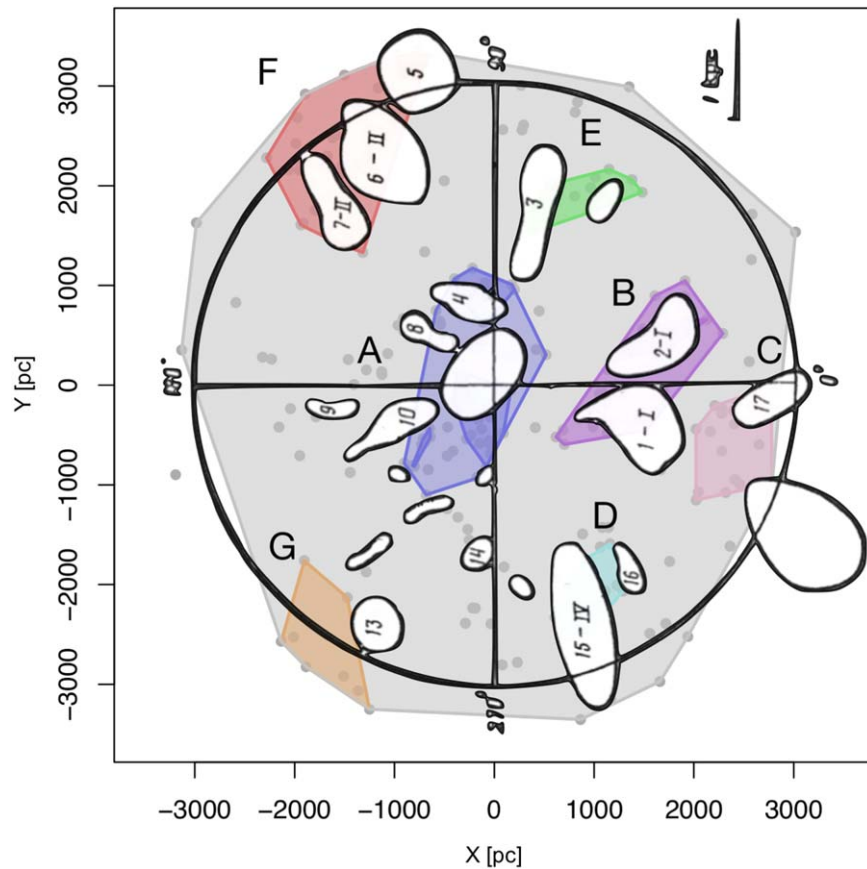
encompasses a slightly wider range of ages than our Figure 3 (bottom panel). In addition, our sample is representative of the current clustered star formation. This last issue could be the cause of the main differences found between both distributions. The star formation in Cygnus and Carina appears to occur mainly in clusters, while few young and massive stars are isolated or located in looser groups.

To explore the distribution of the G-YOC groupings in more detail, we compare our results with those of the seminal work by ES88. Figure 5 shows the results of this comparison, where the colored polygons represent the convex hulls of the structures found in this work and the overlapping white blobs are the complexes and supercomplexes from ES88. The results from ES88 show a striking resemblance to those obtained in this work, despite the obvious differences in the detection techniques, and nature and quality of the data. In particular, most of the groupings found in this work can be linked to corresponding complexes from ES88. All supercomplexes (marked with roman numerals in ES88) correspond to

structures of hierarchical level 1, with the exception of structure D in the Sagittarius arm, which corresponds to the Carina OB1 association. Structure E in the Cygnus area, corresponding to complex 3, also reaches hierarchical level 1.

The 2 kpc sideways box centered on the Sun shows the richest pattern, where the GB is included. A Copernican education leads us to doubt the singularities appearing at positions close to the observer's location. However, we will analyze the  $Z(X, Y)$  and  $V_z(X, Y)$  maps in this region in more detail later.

Numerically, level 0 clusters ( $n = 9$ ; the smallest scale of the hierarchy) have a size central value (median) of 630 pc, although within a wide range of diameters. Level 1 ( $n = 4$ ) has a median of 1600 pc, and the only level 2 cluster, located in the solar vicinity, shows a 3D diagonal of 2300 pc. Level 0 and 1 clusters present similar sizes to those determined for Milky Way stellar complexes and supercomplexes in previous works, suggesting that these hierarchical levels correspond to these structures. In particular, ES88 estimated a range between 150



**Figure 5.** Map of the structures retrieved in this work, as shown in the lower panel of Figure 3. Figure 2 from ES88 is overlaid for comparison, with their complexes and supercomplexes shown in white.

and 700 pc for stellar complexes within a radius of 3 kpc around the Sun, which agrees with our measurements. They also provided a typical value of 1.5 kpc for supercomplexes, in perfect agreement with the central value for the level 1 clustering that we consider to be associated with these structures. Structure A (Table 2 and Figure 3), with the largest size and highest degree of complexity, deserves special mention. Its diameter of 2.3 kpc seems to relate it to those star-forming supercomplexes that are associated with sheared spiral arms (Elmegreen & Efremov 1996). These star-forming regions can reach sizes three times the width of the spiral arm, and their origin could be associated with different gravitational resonances in the Galactic disk (Elmegreen et al. 1992). These supercomplexes also show a high degree of internal spatial hierarchy, as well as an intricate kinematic structure, with the presence of several moving groups (Lepine 2018). Structure A forms the core of the LA, the nature and origin of which is still a matter of debate. However, the size of A, its high degree of structural hierarchy, and the corrugations in  $Z$  and  $V_z$ , which show the largest amplitude and spatial scale, suggest an origin associated with gravitational traps (Lepine 2018; Michtchenko et al. 2019), plus a transient event (e.g., interaction with high-velocity clouds, satellite galaxies, tidal tails, or dark matter blobs), which would generate the higher local vertical and vertical velocity amplitudes. The distribution of the clusters obtained with OPTICS on the plane indicates that the Perseus arm only contains the F supercomplex located in the II Galactic quadrant, between Galactic longitudes  $100^\circ \leq l \leq 140^\circ$ . Young clusters and massive stars appear to be more scarce in the arm's

prolongation into the III Galactic quadrant (Negueruela & Marco 2003; Cantat-Gaudin et al. 2020). Clusters A and E seem to define the Orion–Cygnus (or LA) arm. Its natural continuation through the III Galactic quadrant seems to be the connection with cluster G, between the directions  $240^\circ \leq l \leq 250^\circ$ , as proposed by Moitinho et al. (2006) and later confirmed by Vázquez et al. (2008). The Carina–Sagittarius arm is defined by structures B and D, which are concentrations similar to those of massive stars and OB associations (Wright 2020; Pantaleoni Gonzalez et al. 2021). Concerning the spiral structure in the solar neighborhood, even if we only consider the seven larger structures in Figure 3 (bottom), it is difficult to draw a feasible scheme of such a structure. Thus, perhaps we should think of a hierarchy of groupings, rather than three spiral arms, to analyze the phase space of this Galactic region. The outstanding consistency of the global picture found in this work using state-of-the-art detection and analysis techniques with the results from ES88 highlights both the human potential for pattern detection and the reliability of the parameter estimates for clusters and associations, supporting their effectiveness in studying the large-scale structure of the Galaxy in depth (e.g., Elmegreen & Efremov 1997).

#### 4.2. Kriging Results: Structure of the Galactic Disk

We finally show and describe the results obtained from the Kriging inference method presented in Section 3.1. This interpolation approach provides the  $Z(X, Y)$  map and the  $V_z(X, Y)$  vertical velocity field of the Galactic disk defined by

the G-YOC sample. As the YOCs are young objects that remain close to their birthplace, and, in most cases, are associated with massive star-forming regions, these vertical structures can be related to the large-scale clustered star-forming activity in the Galactic disk. Figure 6 shows the  $Z(X, Y)$  Kriging spatial map, which provides an estimation of the 3D Cartesian structure of the disk drawn by the YOCs. Analogously, Figure 7 shows the corresponding one for  $V_Z(X, Y)$ . These maps correspond to a region within a 3.5 kpc radius around the Sun. We take a grid with steps at each 10 pc to perform the Kriging estimation, based on the  $Z$  and  $V_Z$  measurements from the G-YOC sample, represented by the filled circles in the latter figures. We also plot the position of the Cepheus spur and the spiral arms from Castro-Ginard et al. (2021), with fixed spiral widths based on Reid et al. (2014), to explore the phase-space behavior along the location of the spiral arms, as is discussed in Section 5. Furthermore, the Kriging estimate at each grid point is mainly affected by the density of the input data in its proximity. We can see this effect in the bottom plots of Figures 6 and 7. The regions with lower variance correspond to the locations closer to the higher surface density of the observed clusters.<sup>10</sup>

The  $Z(X, Y)$  map shows a complex structure at different scales. A gradient of the vertical coordinate  $Z$  is seen along the entire length of the  $Y$ -axis. Superimposed on this large-scale gradient is a series of peaks and valleys, some of which show larger vertical displacements. The Galactic disk is corrugated with different amplitudes and spatial scales. One of the more striking substructures is located in the solar neighborhood between  $-1.0$  and  $1.2$  kpc in the  $Y$  coordinate, where the  $Z$  versus  $Y$  gradient becomes deeper, reaching a  $Z$  interval close to 200 pc. The averaged  $Z$  shows a negative value of  $-20.9$  pc within a 50 pc radius around the Sun, and of  $-17.5$  pc for the whole region under study. These heights are within the ranges of values obtained by other authors for different tracers with different ages and Galactic locations (see Elias et al. 2006b; Maíz Apellániz et al. 2008; Everall et al. 2022, and references therein). The  $\sigma_Z$  map also provides a rough estimation of the Galactic plane height scale. This dispersion, as already pointed out in ACCD, accounts for the intrinsic Galactic  $Z$  dispersion, the precision of individual  $Z$  measurements, and the inherent errors in the Kriging interpolation. We obtained an average estimate of  $\sigma_Z = 77.4$  pc for the G-YOC sample, similar to those found by other authors for young stellar populations in this region (Maíz Apellániz 2001; Piskunov et al. 2006; Zari et al. 2018; Kounkel et al. 2020).

Pantaleoni Gonzalez et al. (2021) state that the main features observed in the 3D map of their sample of OB stars are similar to those of the 3D map of YOCs derived 30 yr earlier (ACCD). The same can be said of the comparison between the map obtained in this work and that of 1991: (a) there is a global  $Z$  gradient along the axis parallel to Galactic rotation; (b) the III Galactic quadrant appears to be located, in large extension, below the Galactic plane; (c) the steeper spatial gradient in  $Z$  appears close to the Sun; and (d) the overall appearance of the disk is that of a corrugated surface, like an egg crate that has undergone huge deformation, giving rise to a not necessarily periodic distribution of peaks and valleys of different amplitudes.

<sup>10</sup> These Kriging matrices are available at CDS via anonymous ftp to [cdsarc.u-strasbg.fr](ftp://cdsarc.u-strasbg.fr) (130.79.128.5) or via <https://cdsarc.unistra.fr/viz-bin/cat/J/ApJ>.

$V_Z(X, Y)$  shows some large- and intermediate-scale similarities with the spatial map, but a clumpier kinematic structure at shorter scales. The velocity gradient observed in  $V_Z$  is remarkable along the total length of the  $Y$ -axis. Overimposed on this general trend, a much steeper vertical velocity gradient is observed in the central region, limited by a 2.0 kpc side box. These steep gradients, at both the  $Z$  and  $V_Z$  variables, and their spatial coincidence with the location of the classical GB, are features that need to be inserted into any global scenario for explaining the phase structure of the young Galactic disk.

Finally, we show some cuts of these maps along some selected directions on the Galactic plane. Figure 8 shows the profiles of  $Z$  and  $V_Z$  throughout the Perseus, LA, and Sagittarius spiral arms, the Cepheus spur, and the main diagonal ( $Y = X$ ).

The top panels of Figure 8(a) display the corresponding  $Z$  profile, as thick black lines, along the three spiral arms. The shaded area accounts for the associated Kriging dispersion  $\sigma_Z$ . In contrast to the Perseus and Sagittarius arms, LA shows a sharper  $Z$  gradient in the central region. The expected corrugations observed in other works (e.g., Spicker & Feitzinger 1986; Alfaro et al. 1992) are also visible in these profiles.

For the vertical velocity, we observe that the global trend is a  $V_Z$  gradient along the arms, being less pronounced in the case of the Perseus arm. The  $V_Z$  profiles show their maximum slopes in the regions close to the Sun, as with the  $Z$  spatial coordinate. Other special features, such as the Cepheus spur (Figure 8(b)), and the main diagonal ( $Y = X$ ; Figure 8(c)), also host similar behaviors in  $Z$  and  $V_Z$ , with undulations in both variables. These results are discussed in more detail in the following section 5.

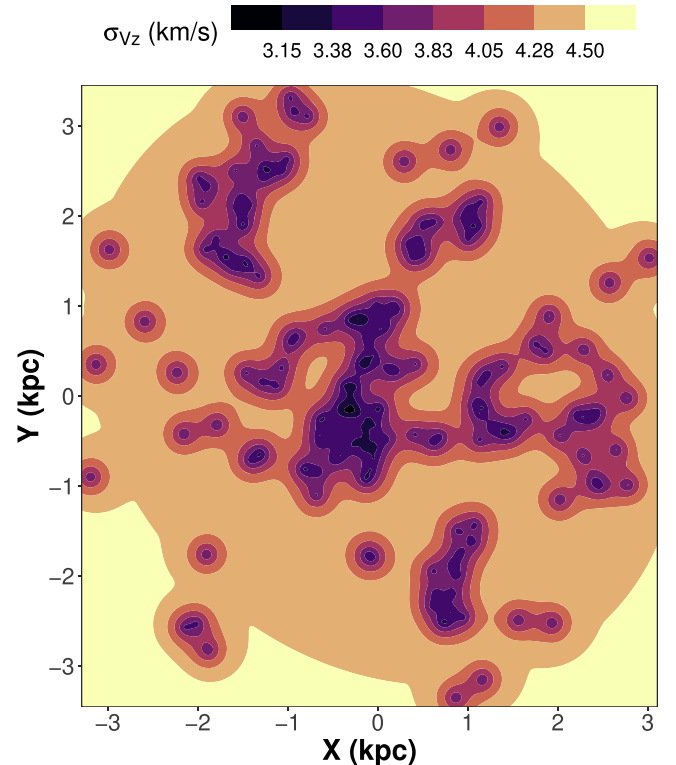
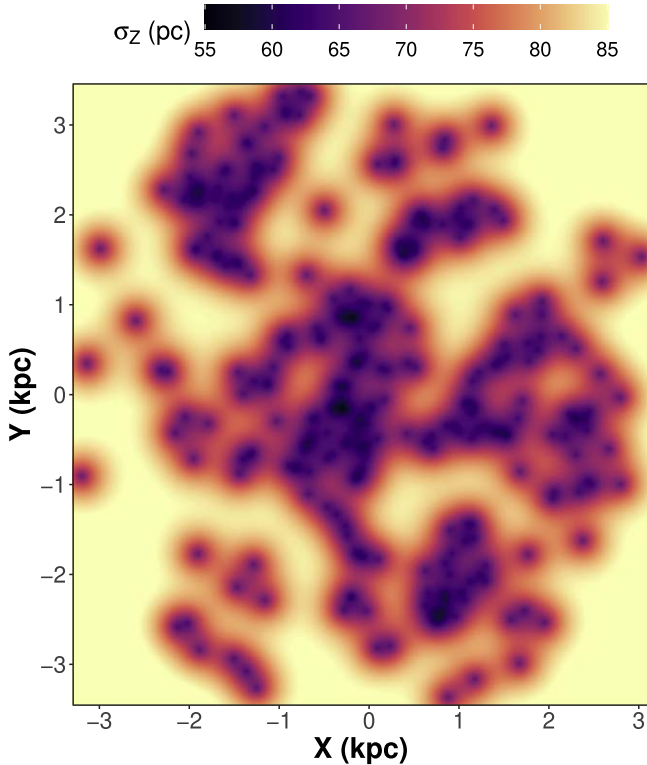
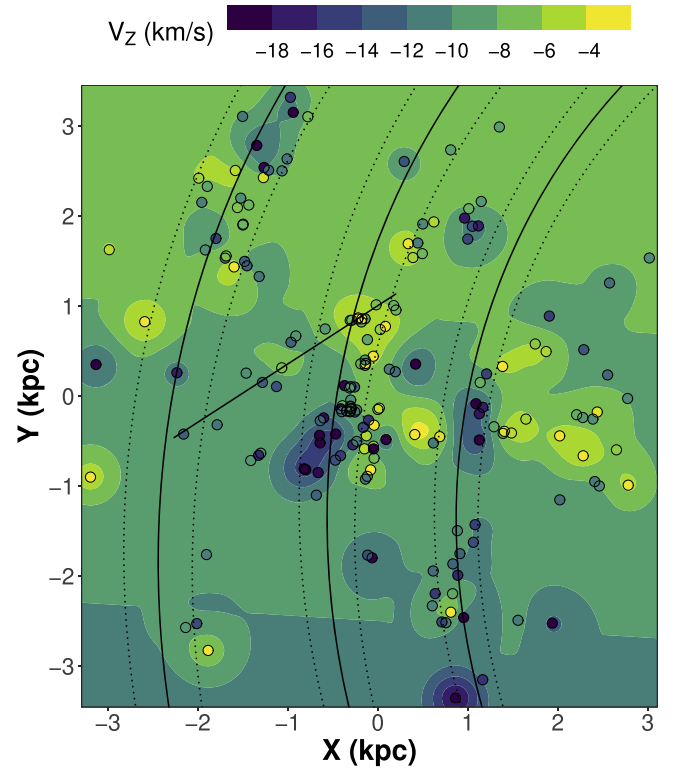
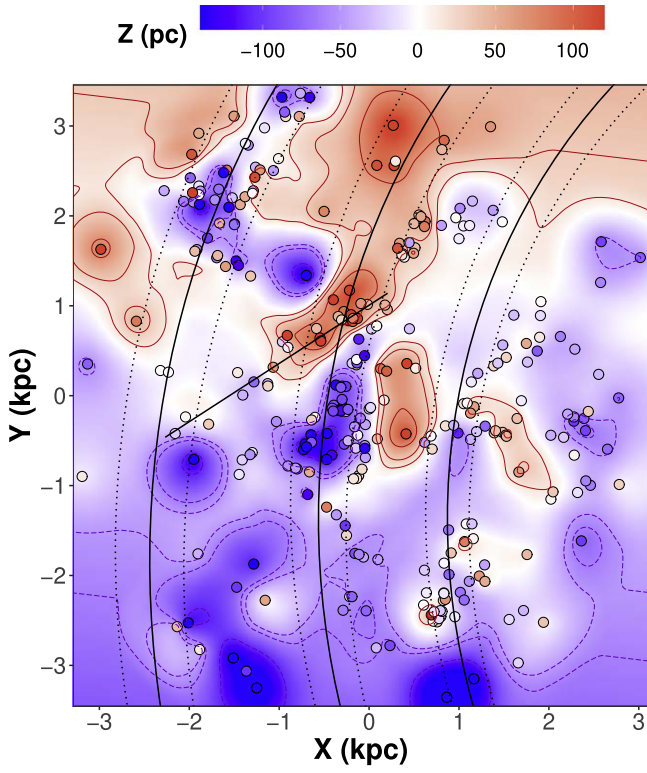
## 5. Discussion

Both the 3D spatial distribution and the vertical velocity field have three common features:

1. A large-scale spatial variation spanning the entire area under study, which has been historically associated with the Galactic warp (Levine et al. 2006; Romero-Gómez et al. 2019; Cheng et al. 2020, among others).
2. A central region, 2.0 kpc on a side, with a high degree of 2D, 3D, and  $V_Z$  structure, which includes the traditional GB (Elias et al. 2006b; Comerón 1999; Romero-Gómez et al. 2019) and dominates the LA vertical profile (Dixon 1967).
3. Superimposed on the large-scale spatial variation, smaller-amplitude ripples are present, for both  $Z$  and  $V_Z$ , in any direction we observe. These corrugations affect the entire disk and cannot be considered specific to particular directions, although some gas and young stellar object alignments, such as the Cepheus spur (Pantaleoni Gonzalez et al. 2021), merit a more detailed analysis.

The profiles of  $Z(X, Y)$  and  $V_Z(X, Y)$  along the main diagonal  $Y = X$  clearly show the three features highlighted above. In Figure 8(c), a well-defined gradient with distance in  $Z$  is observed along the main diagonal, which can be estimated as  $14.78 \pm 1.56$  pc kpc<sup>-1</sup> for the whole linear segment. The same pattern is observed in the variation of  $V_Z$  from the I to the III Galactic quadrants, with a global slope of  $0.50 \pm 0.36$  km s<sup>-1</sup> kpc<sup>-1</sup>. The central region shows steeper slopes in  $Z$  and  $V_Z$



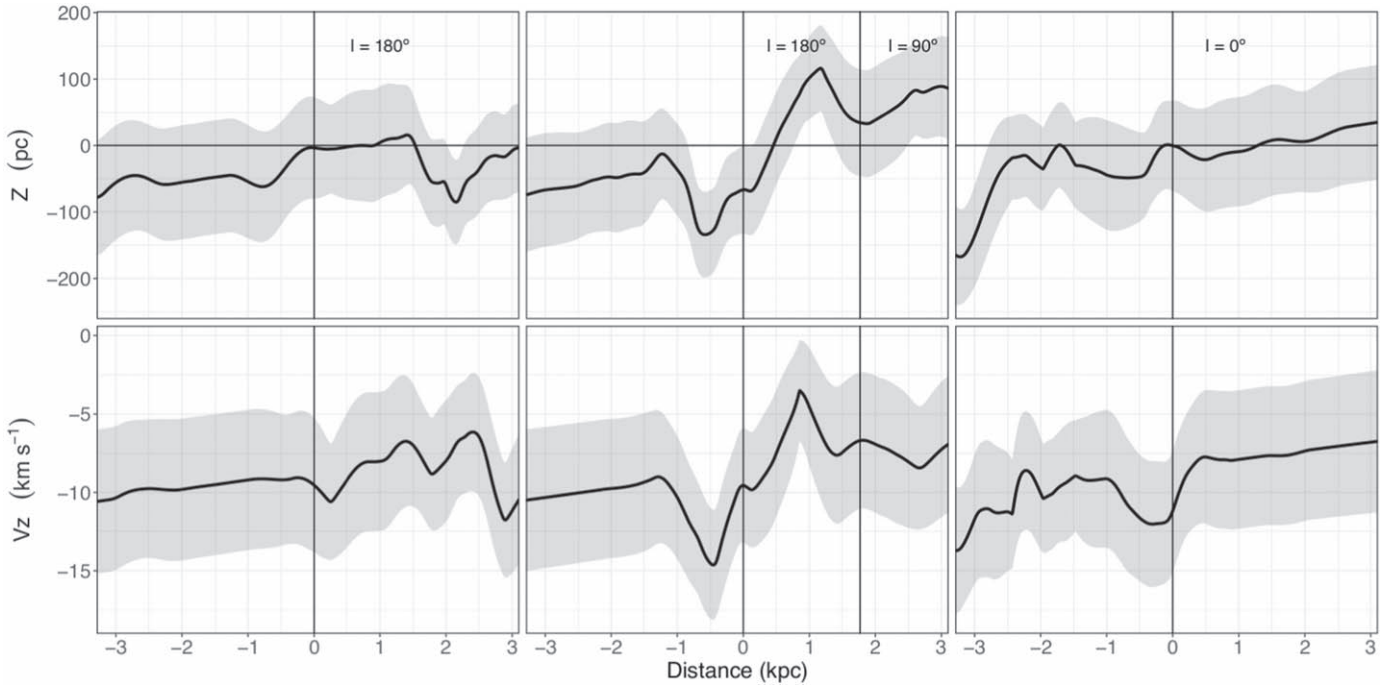


**Figure 6.** Top:  $Z(X, Y)$  map or estimated  $Z$  young Galactic disk structure, inferred from the Kriging method. The colored dots represent the input data. The spiral arms are the same as in Figure 1, and the linear segment represents the Cepheus spur (see Table 1). Bottom: the Kriging prediction standard deviation  $\sigma_Z$  map (Equation (A3)), allowing for a better assessment of the significance of the detected structures.

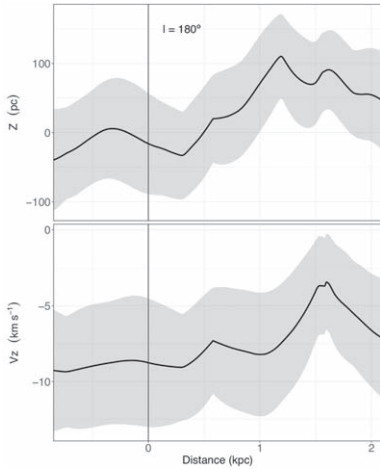
$(170 \pm 9.7 \text{ pc kpc}^{-1}$  and  $0.92 \pm 0.26 \text{ km s}^{-1} \text{ kpc}^{-1}$ , respectively), within a radius of 1 kpc around the Sun.

**Figure 7.** Analogous to Figure 6. Top:  $V_Z(X, Y)$  map. In this case, we have used a discrete scale, as it better suits the minor deviation or range of this variable. Bottom: the Kriging prediction standard deviation  $\sigma_{V_Z}$  map (Equation (A3)).

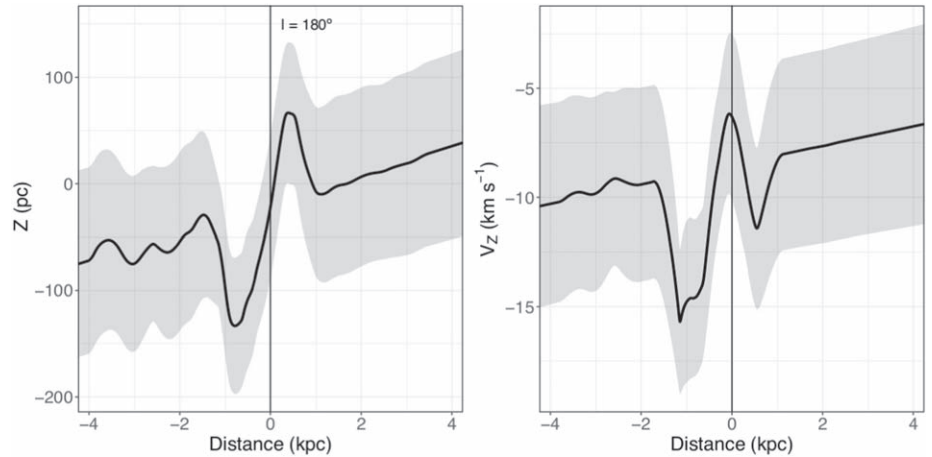
Finally, the existence of  $Z$  and  $V_Z$  ripples with amplitudes of  $\sim 50 \text{ pc}$  and  $\sim 3 \text{ km s}^{-1}$  and spatial scales between 1 and 3 kpc, respectively, dominate the whole disk. We are using the  $Z$  and  $V_Z$  main diagonal profiles to illustrate the three different



(a)  $Z$  and  $V_Z$  profiles along the Perseus, Local, and Sagittarius spiral arms, from left to right.



(b) Profiles along the Cepheus spur.



(c) Profiles along the Main diagonal.

**Figure 8.**  $Z$  and  $V_Z$  profiles along different spatial structures or features. The shaded area accounts for the corresponding dispersion  $\sigma_Z$  and  $\sigma_{V_Z}$ . The distances are measured along the cut, with the origin located at  $Y = 0$  in heliocentric spatial coordinates.

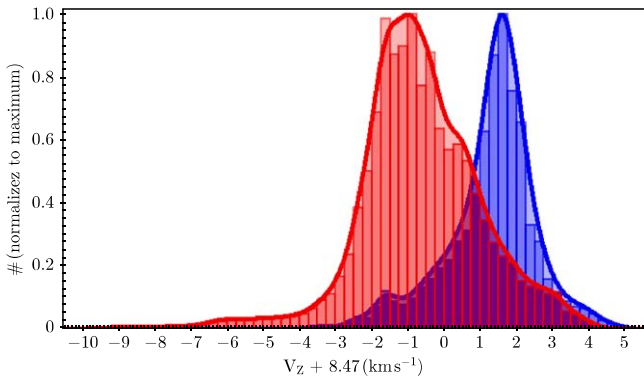
phenomenologies observed in this Galactic region with their different scales and amplitudes.

*5.1. Large-scale Vertical Spatial and Velocity Variations: Can the Warp Be Detected in this Galactic Region?*

Romero-Gómez et al. (2019) have already reported asymmetry in the median  $Z$  about the solar Galactocentric line ( $\phi = 180^\circ$ , counterclockwise), for both OB (age  $< 1$  Ga) and older red giant branch (RGB) stars, where the warp-down amplitude (at  $\phi \geq 180^\circ$ ) is larger than the warp-up amplitude (at  $\phi \leq 180^\circ$ ). This asymmetry can be interpreted as a signature of the Galactic warp lopsidedness. They also found an age dependence with the amplitude of the stellar warp, increasing with the age of the tracer population. In the case of the warp kinematics, the median proper motion in Galactic latitude also

shows a stripe in the warp-down amplitude. Our maps are limited to the central 3.5 kpc around the Sun. This fact, together with the age differences of the samples, could yield different views. However, we see a good agreement with our kinematic results. Overall, if we focus on the clear gradient of  $V_Z$  and  $\mu_b$  along the  $Y$ -axis, restricting the analysis to the same location (see their Figure 8, left panel), the  $Z(X, Y)$  and  $V_Z(X, Y)$  maps show characteristics that are compatible with the detection of a warp defined by YOCs, according to the features presumed to be associated with stellar warp in previous works:

1. Galactic north–south asymmetry in the vertical velocity  $V_Z$  distribution.
2. Asymmetry in the tracer density and  $Z$  between I and II and between III and IV Galactic quadrants, being more evident between the II and III quadrants.



**Figure 9.** Histogram of  $V_Z + 8.47 \text{ km s}^{-1}$ . The constant is the median of  $V_Z$  in the area under study. We have shifted  $V_Z$  for a better visualization of the vertical velocity behavior in both hemispheres. The red shows  $V_Z$  for  $Z \leq 0$  and the blue that for  $Z > 0$ .

3. Undulating behavior of  $Z$  and  $V_Z$  with the Galactocentric radius  $R$ .
4. Vertical velocity gradient  $V_Z$  with the spatial component  $Y$ .
5. Warp amplitude dependence on the age of the tracers, decreasing for the younger ones.

This list includes and summarizes a wide variety of works carried out with different observational samples, which, in many cases, imply different age ranges and different Galactic volumes. Now, all these observational features can be seen, for the first time, in the G-YOC distribution described by the  $Z(X, Y)$  and  $V_Z(X, Y)$  maps.

When analyzing the  $Z(X, Y)$  and  $V_Z(X, Y)$  maps to study their relationship with the Galactic warp, we must take into account two fundamental variables: the age of the tracers and the area under study. The latter puts constraints on the range of Galactic radii analyzed. Recent studies of the Galactic warp (e.g., Romero-Gómez et al. 2019; Cheng et al. 2020), based on the Gaia DRs, concluded that both the  $Z(X, Y)$  and  $V_Z(X, Y)$  maps present a high degree of substructure that cannot be solely explained by classical Galactic warping models.

These previous results highlight the importance of accurately knowing the sample properties and comparing the results obtained in previous works based on tracers of a similar nature.

In Figure 9, we show the histogram of the vertical component of the velocity,  $V_Z$ , for the two Galactic hemispheres ( $b > 0$ —north and  $b \leq 0$ —south) within the area under study. We have added a constant of 8.47 km (median  $V_Z$ ) to the  $V_Z$  values, to better visualize the vertical kinematics of the YOCs of the northern Galactic hemisphere being dominated by positive values. In contrast, those in the southern hemisphere show a negative distribution mode.

Figure 10 shows the  $Z(X, Y)$  and  $V_Z(X, Y)$  maps projected onto the  $Y$ -axis. The projections of the Kriging maps again show a similar scheme to the one observed in the main diagonal: a well-defined large-scale gradient with a superposition of peaks and valleys of different amplitudes, densities, and spatial scales, finding the steepest slopes in locations close to the Sun.

The same fact is observed in Figure 11, where the moving average of  $V_Z$  versus  $R$  has been plotted for the northern and southern hemispheres, as well as for the entire map. The separation between  $Z$  and  $V_Z$  for the I and II Galactic quadrants and for the III and IV ones is clearly observed. Figure 11 shows

the weighted average of  $Z(R)$  and  $V_Z(R)$  profiles on the left, and  $Z(\Phi)$ ,  $V_Z(\Phi)$  on the right. We took bins of 20 pc and bins of  $0.5^\circ$  for  $R$  and  $\Phi$ , respectively, to compute the weighted average of the  $Z$  and  $V_Z$  maps (Figures 6 and 7). The overall weighted mean, encompassing the whole region, is plotted for all cases with a black line, together with the weighted mean for the different cases for each variable. The weights are the normalized inverse of the Kriging standard deviation  $w_i = \frac{\sigma_i^{-1}}{\sum_{i=1}^n \sigma_i^{-1}}$ , and the global uncertainty or variance of the weighted average is computed as  $\sigma_{\text{avg}}^2 = w \Sigma w^T$ , where the covariance matrix  $\Sigma$  is estimated by the sample cross covariogram provided by Gstat. At larger  $R$ , the uncertainty increases, due to the scarcity of the observations at those Galactocentric radii.

If we look at the right plot of Figure 11, where the variations of the vertical coordinate with the Galactocentric azimuth ( $\Phi$ ) for different intervals of  $R$  are shown, we can observe that there is a gradient of  $Z$  and  $V_Z$  with  $\Phi$  that becomes more marked as  $R$  increases. These results are in general agreement with those of previous studies (Levine et al. 2006; Romero-Gómez et al. 2019; Wang et al. 2019; Cheng et al. 2020), but with a caveat—here, they have been obtained for a sample of young stellar clusters with ages less than 50 Ma, which is a useful observational constraint for narrowing down the possible origins of these structures.

Previous studies have pointed out that the Galactic warp, defined by the young stellar population, seems to show its larger southern deviations between Galactic longitudes  $230^\circ$  and  $260^\circ$  (Moitinho et al. 2006; Vázquez et al. 2008; Carballo-Bello et al. 2021).

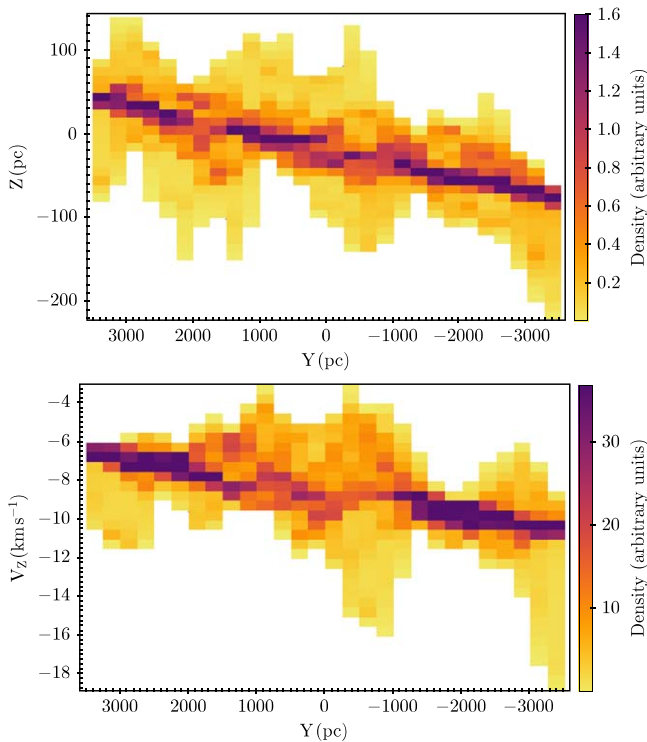
To answer the question that titles this subsection: yes, the spatial and kinematic distributions of the YOCs show average values along different variables that can be associated with warping. However, the map drawn by  $Z$  and  $V_Z$ , if we look at short spatial scales, is very far from resembling what the models predict (see Romero-Gómez et al. 2019 for different mock models), both in shape and magnitude.

## 5.2. The Solar Neighborhood

The central region of the area under study requires a more detailed analysis. We define this region as a  $2 \times 2$  kpc square centered on the Sun. The square box contains the highest amplitudes in  $Z$  and  $V_Z$  (Figure 8(c)) found in our study, which already singles it out as a peculiar region. Regarding the distribution of YOCs analyzed in Section 4.1, this box is preferentially associated with the supercomplex A, which presents a 3D diameter of 2.3 kpc. The supercomplex A shows the highest spatial hierarchy (level 2), except for the one corresponding to the total area (level 3), and also presents the highest density of YOCs, especially in the Orion region (as already noted by other authors: Piskunov et al. 2006; Bally 2008; Elias et al. 2009).

Within the complicated phenomenology of phase space in the solar neighborhood, the central box includes the most populated LA segment and the classical GB. Figures 6 and 7 show the  $Z(X, Y)$  and  $V_Z(X, Y)$  maps, respectively, for this region, where we can clearly appreciate the presence of three substructures, at least, in  $Z(X, Y)$ , whose relative position on the plane explains and defines both the  $Z$  undulation along the LA segment (e.g., Dixon 1967; Quiroga 1974; Spicker & Feitzinger 1986; Alves et al. 2020; Pantaleoni Gonzalez et al. 2021, among many





**Figure 10.** In the top panel, a  $Z(X, Y)$  map projected onto the  $Y$ -axis, both in pc units. Negative  $Y$  values correspond to the III and IV Galactic quadrants. A gradient  $Z(Y)$  is clearly visible, overlaid with a wave-like structure. In the bottom panel, a  $V_Z(X, Y)$  map in  $\text{km s}^{-1}$  projected onto the  $Y$ -axis in pc. Negative  $Y$  values correspond to the III and IV Galactic quadrants. A gradient  $V_Z(Y)$  is the main feature of the plot.

others), and the classical inclination of the GB with a line of nodes at  $l \approx 280^\circ$  (Comerón et al. 1994; Elias et al. 2006b).

The vertical velocity field,  $V_Z$ , also presents a similar pattern, with four different clumps, which draws a clear variation of  $V_Z$  along the LA central segment (see Figure 8(a) for a longitudinal cut along the arm). In turn, we observe a  $V_Z$  gradient between the two interarm substructures, which had previously been associated with an oscillation of the GB around the solar Galactocentric axis (Comerón 1999).

Without entering into other considerations about the nature and origin of the spiral arms, it is evident that the amplitudes of the spatial Cartesian coordinate,  $Z$ , and of the vertical velocity component,  $V_Z$ , as well as the high concentration of YOCs in the region, suggest that large amounts of energy and momentum have come into play in a relatively small spatial volume.

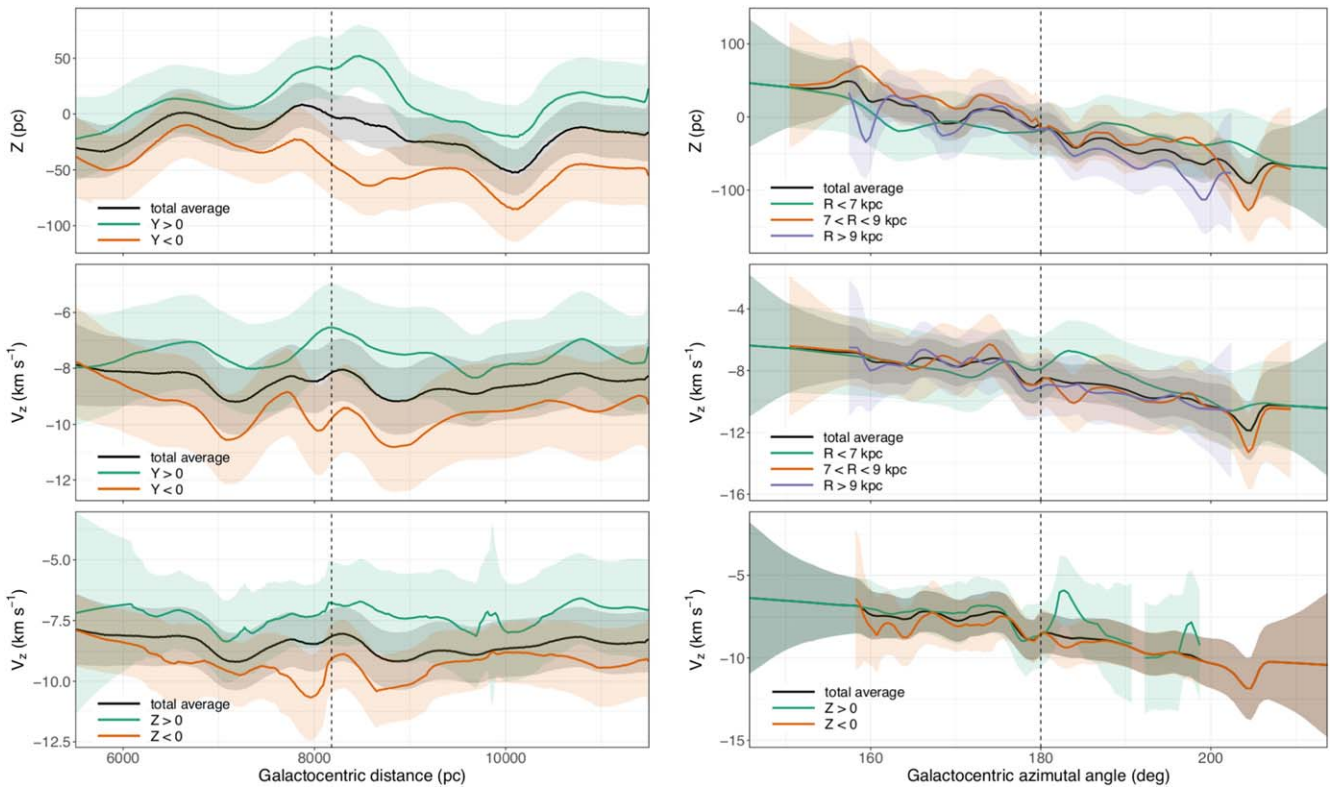
The physical mechanisms proposed in the literature for injecting these enormous amounts of momentum and energy into the solar neighborhood have changed over time (see, e.g., Alfaro & Efremov 1996 for a review of the generators of the possible corrugations at that time). New sources of momentum and energy have been incorporated in recent years, which, together with the classical ones, could be summarized as: supernova explosions (Olano 1982), collisions with high-velocity clouds (Franco et al. 1988; Comerón & Torra 1994), encounters with dark matter blobs (Bekki 2009), and interactions with tidal streams (Laporte et al. 2018; Ruiz-Lara et al. 2020; Tepper-García et al. 2022, and references therein), among others. This wide phenomenology can inject, into the interstellar medium, energies higher than  $10^{51}$  erg in a single event. Note, for example, the case of NGC 6946, with

high velocity cloud (HVC) kinetic energies over  $10^{53}$  erg (Kamphuis 1993), able of displacing giant molecular clouds at heights of 150 pc from the formal Galactic plane. The existence of empty bubbles in the solar neighborhood, bordered by gas clouds and young stellar groups, and generated by any of these mechanisms has been detected, and described in several papers (see e.g., Lallement et al. 2003; Gontcharov & Mosenkov 2019; Zucker et al. 2022, and references therein). The subsequent dynamical evolution of the stellar population would explain the formation of the observed tilted plane associated with GB.

Otherwise, there is kinematic information that shows the solar neighborhood as a collection of moving groups, rather than as a single structure (disk or ring), with a coherent motion able to survive longer than a few Ma (Elias et al. 2006a, 2009; Lepine 2018). Two OB associations (or, better, two groups of star-forming regions: Orion and Scorpius–Centaurus) located at the extremes of the  $Z$  interval in the local spatial distribution of OB stars, show very different properties. While Orion has a high concentration of star clusters, Sco–Cen shows more dispersed star formation, where it is difficult to find any substantial cluster in the optical range (Elias et al. 2009). Likewise, the kinematics of Sco–Cen cannot be explained only as an expanding star-forming region subjected to a stationary Galactic gravitational field (Moreno et al. 1999; Antoja et al. 2008; Elias et al. 2009; Lepine 2018). Sco–Cen OB associations appear to be associated with the Pleiades moving group, which includes the star cluster of the same name (Chen et al. 1997; Asiain et al. 1999; Montes et al. 2001; Makarov 2007). The cluster and the associations are spatially separated by more than 300 pc and show ages differing by a factor of 10. The formation of these moving groups in the solar neighborhood, with an age range exceeding, in some cases, several hundred Ma, can be explained by dynamical resonances, generating gravitational traps where several moving groups can orbit for some Ga in a bounded region of the solar neighborhood (Fux 2001; Antoja et al. 2008; Minchev et al. 2010; Lepine 2018; Michtchenko et al. 2018; Barros et al. 2020).

On the other hand, the main phase-space feature in this region is a spiral-shaped structure in the  $Z$ – $V_Z$  plane, whose contours are even sharper when modulated by the in-plane velocity components (Antoja et al. 2018; Tian et al. 2018; Bland-Hawthorn et al. 2019; Michtchenko et al. 2019; Li 2021). The stellar samples that draw these spirals are mainly formed by stars with parallax errors lower than 20% (avoiding negative parallaxes) and with radial velocities included in Gaia DR2 and/or in any of the other extensive RV surveys listed in Section 2.1. Although there are differences in the selection criteria of the distinct samples, they show some common properties (e.g., see Michtchenko et al. 2019 for details). The samples are representative of a stellar population with a large proportion of late-type stars over a very wide age interval, spanning the age of the Galactic disk. The wide range of ages drawing this structure would suggest that the initial phase-space pattern (if any) is not completely erased over a time period shorter than the typical age of the sample, or, on the contrary, that the Galactic driver needs this long time to generate the peculiar snail shells in the vertical phase space.

Other works have demonstrated that the formation of vertical phase spirals is independent of the Galactocentric radius (Li 2021 and references therein) and of the chemical properties of the sample (Bland-Hawthorn et al. 2019). In other words, the



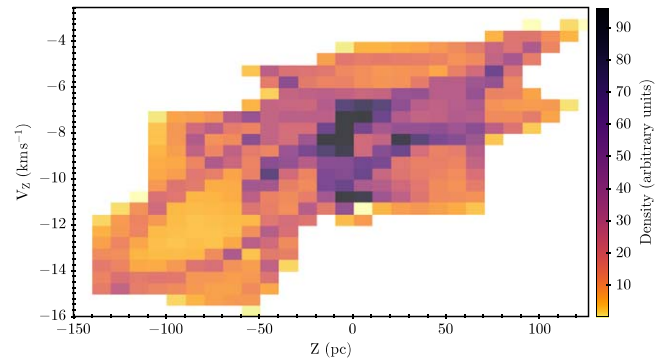
**Figure 11.** Weighted average Kriging estimations for  $Z$  (pc) and  $V_z$  ( $\text{km s}^{-1}$ ) as a function of the Galactocentric radius ( $R$ ; bins of 20 pc) and the Galactocentric azimuthal angle ( $\Phi$ ; bins of  $0.5^\circ$ ), respectively.  $\sigma_i^{-1}$  is taken as the weight. The shaded areas correspond to the standard deviations of the weighted averages, assuming correlated measurements. The vertical dashed line accounts for the Sun's location.

mechanism generating these structures seems to be acting over a long time period ( $\sim$  the disk's age) and over the entire Galactic disk.

The  $Z(X, Y)$  and  $V_z(X, Y)$  fields defined by the YOCs show a clear correlation between  $V_z$  and  $Z$  for the central box in the solar neighborhood (Figure 12). If we compare this plot with the spirals observed in much older stellar samples (Antoja et al. 2018; Tian et al. 2018; Khanna et al. 2019; Michtchenko et al. 2019), we see that the linear structure,  $Z$  versus  $V_z$ , depicted by the YOCs appears to be within the central region of the snails detected in these works.

Tian et al. (2018) have used Gaia DR2 astrometric data, together with LAMOST radial velocities, to analyze the vertical phase in different age groups. Their youngest sample (age  $< 0.5$  Ga) is 10 times older than G-YOC, and the corresponding vertical phase does not seem to show any well-defined structure, contrary to the roughly linear relationship obtained in this work.

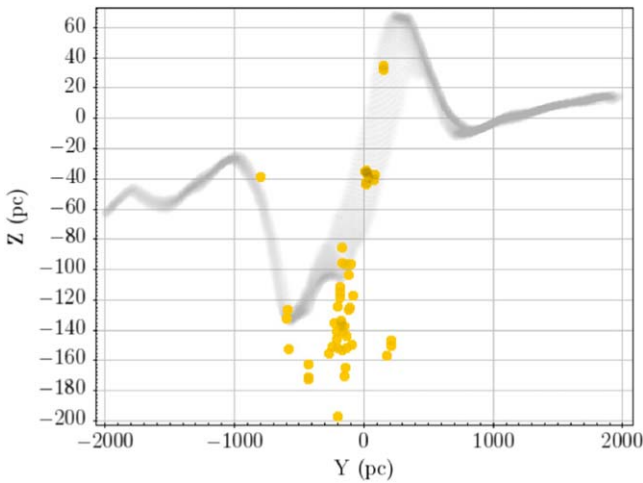
Thus, we have three well-tested observational features associated with the central box of the solar neighborhood: (1) a highly hierarchical stellar supercomplex of 2.3 kpc in diameter, associated with a high density of YOCs; (2) the largest  $Z$  and  $V_z$  deviations within the entire region under study; and (3) a well-defined linear relationship between  $Z$  and  $V_z$ . We can add a fourth observational constraint, including gas. Molecular gas clouds, cataloged by Zucker et al. (2019, 2020), show a vertical profile along the main diagonal ( $Y = X$ ), similar to that found for the YOCs in this work (see Figure 13). Although the  $Z$  values of the molecular clouds seem to show a systematic shift with respect to the main diagonal profile, the agreement between the two distributions is



**Figure 12.**  $V_z$  vs.  $Z$  for the central box, 2.0 kpc on a side, containing stellar supercomplex A. This area includes the most populated LA segment, as well as the classical GB. The color code indicates the density of the map points in arbitrary units. A clear correlation between both phase-space coordinates is observed.

remarkable. On the other hand, we are not able to discern whether this  $Z$  difference between the molecular gas and the YOCs is real or due to any observational bias, especially considering the different distance estimation methods for both types of objects (see Zucker et al. 2019 for a more detailed explanation of how molecular cloud distances are calculated). These four observational issues establish important physical constraints on any proposed scenario to explain the clustered formation and the phase-space structure of the young Galactic disk.

The age of the G-YOC sample is less than 50 Ma, which is too short for the Galactic potential to have substantially modified, or erased, the imprint of the stellar cluster formation



**Figure 13.** Vertical profile,  $Z$  vs.  $Y$ , for the main diagonal of the  $Z(X, Y)$  map (gray band). Molecular gas clouds (Zucker et al. 2020) placed along the main diagonal band are overimposed as yellow circles. The YOCs’ Kriging map and the molecular clouds show a similar trend, although the latter seems to show larger displacements with respect to the formal disk in the III Galactic quadrant. This connection between the gas and the YOCs establishes an important observational constraint for the origin of the phase-space structures.

on the vertical phase space. The vertical perturbations of the disk are typically thought to have originated from different sources injecting momentum and energy into the Galactic disk, as mentioned above (e.g., Tepper-Garcia et al. 2022 and references therein). However, the gravitational resonances due to spiral arms, or rotating bars, can also give rise to vertical deviations of the disk, and to ridges and groups in the in-plane velocity components (e.g., Fux 2001; Michtchenko et al. 2019), especially if we take into account the Galactic magnetic field (Santillán et al. 1999; Cox & Gómez 2002; Franco et al. 2002; Gómez & Cox 2002).

### 5.3. Corrugations

In this section, we refer to corrugations in the broadest and most phenomenological sense of the term. That is, as any of the wavy undulations of the disk components (Nandakumar et al. 2022). The origin of the term “vertical corrugation” is from Quiroga (1974), and it names the vertical variations of the different disk components (atomic and molecular gas, dust, and stars in different evolutionary states and with different age ranges) in the Milky Way and external galaxies. This includes both the azimuthal and radial  $Z$  variations in our Galaxy (Spicker & Feitzinger 1986), as well as the wavy profiles observed in gas (Matthews & Uson 2008), dust (Narayan et al. 2020), and other wavelengths representative of different stellar populations in edge-on galaxies (Matthews & Uson 2008 and references therein). Corrugations are a global disk phenomenon (wobbly or corrugated disks), but their detection and study has been mainly carried out through work on 2D profiles along privileged directions (disk profiles in edge-on galaxies, spiral arms of the Milky Way, vertical profiles along Galactocentric radii, etc.) However, we should not forget that both observational (Widrow et al. 2012; Romero-Gómez et al. 2019; Pantaleoni Gonzalez et al. 2021) and theoretical (Schonrich & Dehnen 2018; Laporte et al. 2019) studies converge in considering corrugations as a global disk phenomenon. This is our first approach to this concept.

Our main results are the 3D spatial and vertical velocity maps of the Galactic disk, restricted to a radius of 3.5 kpc around the Sun. Figures 6 and 7 display a remarkably complex and high-degree structure. We have included 3D versions of them in Figure 14, for better visualizations.

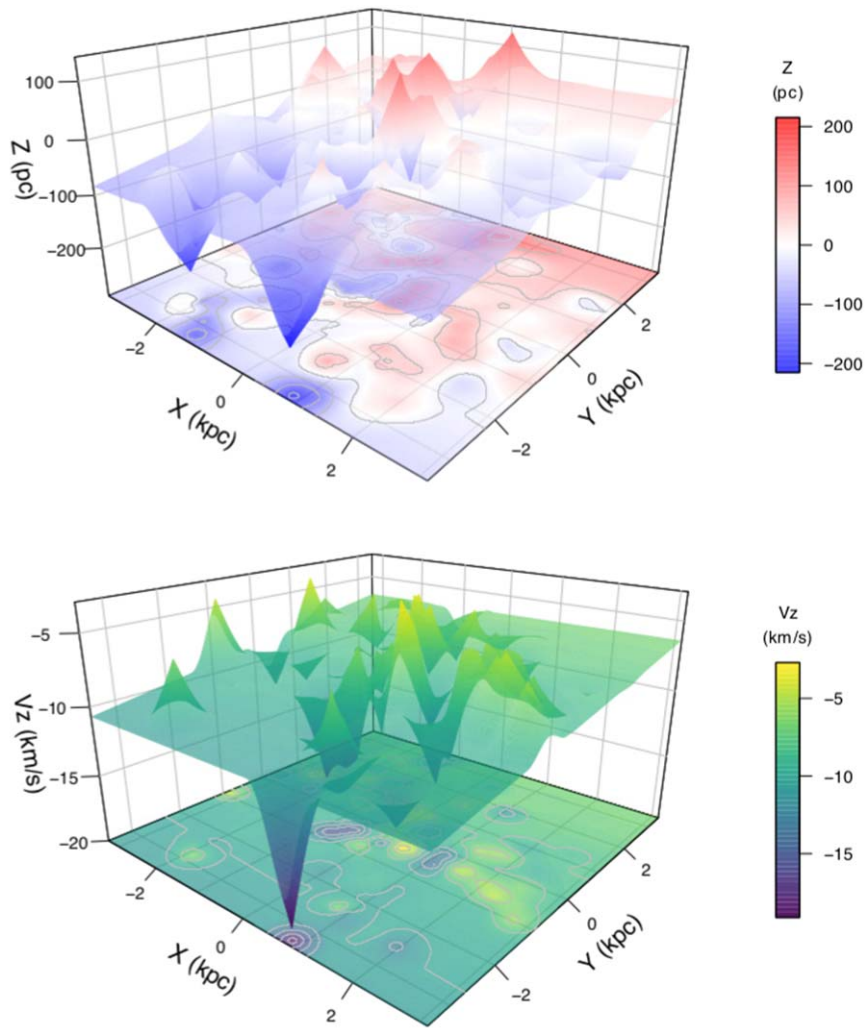
Spiral galaxy disks display many distinct features, such as spiral arms, central bars, resonant structures or rings, outer warps (Tepper-Garcia et al. 2022 and references therein), inner tilted gas to the Galactic disk (Gum et al. 1960), and wave-like disk corrugations (Spicker & Feitzinger 1986; Alfaro & Efremov 1996). In particular, spatial corrugations have been known for more than 60 yr, but nevertheless their systematic study has been intermittent over time. Gaia has brought this issue back to the table. Gas vertical velocity variations have also been detected in some face-on galaxies (Alfaro et al. 2001; Sanchez-Gil et al. 2015; Nandakumar et al. 2022).

Figure 8 has already displayed these expected wavy-like spatial and kinematic corrugations, throughout different linear features, such as some spiral arms and other selected directions. Corrugations are also evident along the Galactocentric radius and azimuthal angle (Figure 11), being the  $Z$  Galactocentric radius profile very similar to those found in other edge-on galaxies (e.g., Matthews & Uson 2008 and references therein). The radial vertical velocity variations strongly resemble the corrugated profiles of  $V_z$  throughout the spiral arms in nearby spiral galaxies (Sanchez-Gil et al. 2015). However, these latter amplitudes showed higher values, ranging from 10–15 up to 30–40  $\text{km s}^{-1}$  in some cases. Regardless, they were obtained from  $H\alpha$  data with a completely different observational technique and methodology.

The maximum amplitude of the corrugations in  $Z$  is 31 pc, and  $\approx 1 \text{ km s}^{-1}$  in  $V_z$ . The vertical velocity shows a periodogram with wavelengths roughly corresponding to those found spatially, that is, between 1.3 and 2.5 kpc. A more exhaustive analysis of the amplitudes and spatial scales of the young Galactic disk corrugations is beyond the scope of this article. It should be said that the values of the spatial scales found in this work are within the range of the values determined in the literature, being amplitudes slightly lower than those found in other pre-Gaia works (Feitzinger & Spicker 1985; Spicker & Feitzinger 1986; Alfaro et al. 1992). Moreover, Figure 14 clearly shows a kinematically and spatially corrugated disk, and not, necessarily, a collection of wavy-like profiles along some privileged directions. These plots offer an overview of the 3D topography of the YOCs’ spatial distribution and their perpendicular velocity field.

The current observations tend to suggest that warps and corrugations are universal phenomena associated with disk galaxies (e.g., Quiroga 1977; Spicker & Feitzinger 1986; ACCD, Florido et al. 1991; Matthews & Uson 2008; Sanchez-Gil et al. 2015; Tepper-Garcia et al. 2022, among others). Despite this universality, what triggers these perturbations has been barely studied and still poses open questions. Different scenarios have been proposed to explain this phenomenon (see Alfaro & Efremov 1996; Alfaro et al. 2001, 2004; Sanchez-Gil et al. 2015; Nandakumar et al. 2022), in some way linked to the mechanisms involved in large-scale star formation processes, such as density waves, tidal interactions, collisions with high-velocity clouds, or galactic bore generated from the interaction of a spiral density wave with a thick magnetized gaseous disk, among others (Gómez & Cox 2002; Franco et al. 2002).





**Figure 14.** 3D topography of the Galactic disk inferred from the G-YOC sample.

The hypothesis of an external perturber interacting with the Galaxy, where the main role is attributable to the passage of the Sagittarius dwarf galaxy through the Galactic disk, is gaining more and more strength for explaining the different vertical phase features. One recent example is found in Tepper-Garcia et al. (2022), where they study the effects of a single satellite passing through the Galactic disk. They reproduce stellar and gas corrugations, initially in phase, but disentangling after a few rotation periods (500–700 Myr). The gas and stellar components evolve in a different way, and their correlation degree could be used to date this encounter. On the other hand, a damping with time of the corrugation amplitudes is predicted, although the vertical energy of the stars remains almost constant throughout the galaxy’s evolution. However, a direct comparison of the complete phase space of the gas and the stellar component is not possible at this time. We have approached this comparison using the spatial information for molecular clouds from Zucker et al. (2019, 2020), shown in Figure 13.

Systematic studies of the gas and the stellar population in different age ranges would help us to understand these structures, their possible origins, and whether these phenomena arise from external actions rather than from internal processes.

#### 5.4. Scenario

The 4D view, ( $X, Y, Z, V_z$ ), of the central box ( $2 \times 2$  kpc) obtained in this work (Figures 6 and 7) introduces important constraints on proposing a global scenario to explain the intricate pattern of the vertical phase space. The main condition stems from the narrow age range of the G-YOC sample ( $\leq 10^{7.5}$  a). Otherwise, the almost linear relationship observed between  $V_z$  and  $Z$  is different from the spiral shapes found for older stellar populations covering a wider age range (Antoja et al. 2018; Bland-Hawthorn et al. 2019; Li 2021). The steep local gradients in both  $Z$  and  $V_z$ , together with the youth of the tracers, would suggest a transient phenomenon producing large local deformations of the vertical phase. The relative positions of the molecular gas and the YOCs in the solar neighborhood (Figure 13) suggest that this encounter is recent. Although the gas and stars could have been in phase for much longer periods ( $\sim 500$  Ma), it is more difficult to explain that the maximum deviation from the plane, with amplitude far in excess of the secondary corrugations, was not damped over that time period (Tepper-Garcia et al. 2022). The correlation between  $Z$  and  $V_z$  (Figure 12), the larger amplitudes at Figure 8(c), and the evident spatial connection between the molecular gas and young stars (Figure 13) suggest that the encounter is taking place here and now. There are indications

that the leading arm of the Sagittarius dwarf galaxy may be entering the Galactic disk in the solar neighborhood (Martínez-Delgado et al. 2007). The quasilinear relationship between  $Z$  and  $V_Z$  for a perturbed gaseous disk is inferred from the collision between HVCs and a magnetized disk (Santillán et al. 1999).

The behavior of the older stellar population is beyond our study, and we can say nothing about it, so far. However, the perturbation of the Galactic disk by encounters with momentum and energy injectors of different natures could account for the observed gas and YOC features, as well as the generation of vertical phase spirals when the older disk population is taken into account (Antoja et al. 2018). The high degree of hierarchy in the distribution of the YOCs and the existence of several moving groups in the central region suggest that a dynamic trap associated with gravitational resonances could explain this phenomenology (Lepine 2018; Michtchenko et al. 2019; Barros et al. 2020).

If we move to larger spatial scales, we observe vertical behaviors that have traditionally been associated with, and have supported, the presence of Galactic warping. However, the main feature that we observe is a gradient in  $V_Z$  with  $Y$  (see Figures 7 and 10) that takes an average value of  $0.5\text{--}0.6\text{ km s}^{-1}\text{ kpc}^{-1}$ . Comparing the  $V_Z(X, Y)$  map of the YOCs with that obtained for the OB sample’s proper motions in latitude from Romero-Gómez et al. (2019) for the same region, we observe a similar pattern. This structure is no longer detected in the RGB sample of these authors. On the other hand, different simulations of the behavior of  $\mu_b$  for different warping shapes (Romero-Gómez et al. 2019) foresee a range of variations four times larger than those observed in that work, with predicted spatial distributions that do not fit the observed vertical velocity gradient in the solar neighborhood. All this leads us to consider that for the young stellar population, at least, transient phenomena dominate the structure of the velocity map, and classical warping approaches do not fit detailed observational kinematic maps of the younger stellar population in the solar neighborhood.

Thus, gravitational resonances and frequent encounters with passing perturbers appear to be the main drivers of the complex structure of the phase subspace analyzed in this paper. The large vertical deviations in both  $Z$  and  $V_Z$  in the solar neighborhood, the quasilinear relationship between both variables, and the spatial connection between molecular gas and YOCs lead us to conclude that such an encounter has taken place not longer than 50 Ma ago.

## 6. Conclusions

In this work, we have applied the Kriging inference technique to a sample of G-YOCs in order to study the 3D spatial structure and the vertical velocity field of the Galactic disk around the Sun. In general, the G-YOC sample exhibits a clumpy, hierarchical distribution, with a 3D fractal dimension in the range  $D_f = 2.5\text{--}2.6$ , similar to those values obtained for different molecular clouds in the Milky Way. This supports the idea that young stars and clusters are distributed following the spatial structure of the interstellar medium from which they were formed. A detailed analysis of this hierarchical structure yields a set of groupings embedded within other groupings that can be identified with the so-called stellar complexes or supercomplexes. Our results reveal a complex structure that is far from a simple planar geometry, in which warp,

corrugations, and large spatial and kinematic deviations at the solar neighborhood are part of a single but complex picture. There is a large-scale spatial variation in both the 3D spatial distribution and the vertical velocity field associated with the known Galactic warp. There is also an intermediate-scale ( $2 \times 2$  kpc) region with a high degree of spatial and kinematic structure, including the so-called GB, delimited in  $Z$  by the positions of the Ori OB and Sco-Cen OB associations. Superimposed throughout the entire structure, small-scale corrugations are observed both in the spatial distribution and in the vertical velocity field of the clusters, with peaks and valleys of different amplitudes and scales. These corrugations affect the entire disk and, in general, can be observed in any direction on the plane. The most pronounced of these fluctuations is located close to the Sun’s position, where the  $V_Z(X, Y)$  and  $Z(X, Y)$  fields show a clear correlation, which we associate with a transient phenomenon produced by a relatively fast transfer of energy and momentum into a magnetized gas disk. We have discussed the physical mechanisms behind this complex structure, but, in any case, our results provide a global 4D ( $X, Y, Z, V_Z$ ) picture that imposes important constraints on models of Galactic disk formation and evolution.

From this analysis we conclude that:

1. A hierarchy of star formation with at least three levels exists, corresponding in general to the complexes and supercomplexes detected by ES88.
2. Supercomplex A, the highest-hierarchy level, is located in the immediate solar vicinity, with a maximum diameter of 2.3 kpc. It contains most of the region’s moving groups, and its origin and existence could be explained by dynamic traps associated with Galactic resonances.
3. We observe three different behaviors in the structures of the maps  $Z(X, Y)$  and  $V_Z(X, Y)$ : (i) at large scales, we notice a gradient of both variables with the  $Y$  coordinate; (ii) the central region shows the largest deviations of  $Z$  and  $V_Z$ , and a linear relationship between both variables; and (iii) superimposed on these features, we see peaks and valleys in both maps, with varying (but low) amplitude and spatial scales between 1.3 and 2.5 kpc.
4. The combined actions of encounters with passing perturbers and the resonances generated by an asymmetric and nonstationary Galactic potential can explain, at least qualitatively, these observations.
5. We consider that, on a short spatial scale, we are observing the effects of the last passage of the Sagittarius dwarf galaxy, no more than 50 Ma ago.

We are grateful to Bruce Elmegreen for comments on the manuscript. This work was motivated in memoriam to our friend and colleague Yuri N. Efremov. Bruce Elmegreen and Emilio J. Alfaro are forever grateful to him for his inspiring insight into the hierarchical structure of star formation through his cataloging of star complexes and supercomplexes in the Milky Way and neighboring galaxies. He brought us together in 1992 on the island of Elba and we were friends and collaborators until his untimely passing in 2019. We acknowledge the referee for a careful reading of the manuscript and very useful comments, which helped us to improve the paper. E.J.A. and M.J. acknowledge financial support from the State Agency for Research of the Spanish MCIU through the “Center of Excellence Severo Ochoa” award to the Instituto de Astrofísica de Andalucía (SEV-2017-0709). E.J.A. and

C.S.-G. have been supported by grant No. PY20-00753 from Junta de Andalucía, Spain (Autonomic Government of Andalusia). N.S. and J.M.-A. acknowledge financial support from grant Nos. PGC2018-095049-B-C21 and PGC2018-095049-B-C22 (MCIU), respectively. In this work, we have made extensive use of TopCat (Taylor 2005). We thank its author, and the subsequent contributors, for the creation and development of this tool.

## Appendix A Variogram Model Selection

### A.1. Variogram Estimation and Model Fitting

Spatial prediction refers to the prediction of the unknown quantity  $Z(r_0)$ , based on sample data  $\{Z(r_i)\}_{i=1,\dots,n}$ , and assumptions regarding the form of the trend of  $Z$  and its variance and spatial correlation. For the UK model, we assume  $Z(r) = m(r) + e(r)$ , with  $E(e(r)) = 0$ , and the mean, which varies spatially, is modeled as a linear function of  $p$  known predictors  $F = (f_1(r), \dots, f_p(r))$ , with  $f_j(r) = (f_j(r_1), \dots, f_j(r_n))^T$ . This is

$$Z(r) = m(r) + e(r) = F\beta + e(r), \quad (\text{A1})$$

with unknown regression coefficients  $\beta = (\beta_1, \dots, \beta_p)^T$  (see Equation (3)). Given this model, the best linear unbiased predictor, or the Kriging predictor, of  $Z(r)$  consists of an estimated mean value for the location  $r_0$ , plus a weighted mean of the residuals from the mean function, with weights  $v^T V^{-1}$ , known as the simple Kriging weights,

$$\hat{Z}(r_0) = f(r_0)\hat{\beta} + v^T V^{-1}(z(r) - F\hat{\beta}), \quad (\text{A2})$$

where  $f(r_0) = (f_1(r_0), \dots, f_p(r_0))$ ,  $z(r) = (Z(r_1), \dots, Z(r_n))^T$ , and  $v = (Cov(e(r_1), e(r_0)), \dots, Cov(e(r_1), e(r_n)))$  are the covariance vectors of  $Z(r_0)$  and  $z(r)$ ,  $V = Cov(e(r))$  is the known covariance matrix of  $Z(r)$ , and  $\hat{\beta} = (F^T V^{-1} F)^{-1} F^T V^{-1} z(r)$  is the generalized least squares estimate of the trend coefficients  $\beta$ . The Kriging prediction has a prediction variance, or Kriging

variance, given as

$$\begin{aligned} \sigma_Z^2(r_0) &= \text{Var}(Z(r_0) - \hat{Z}(r_0)) \\ &= \text{Var}(Z(r_0)) - v^T V^{-1} v \\ &\quad + (f(r_0) - v^T V^{-1} F)(F^T V^{-1} F)^{-1} \\ &\quad \times (f(r_0) - v^T V^{-1} F)^T. \end{aligned} \quad (\text{A3})$$

The variogram estimation is a significant issue for the statistical inference of spatially correlated variables. Typically, its estimation is split into two stages: empirical variogram estimation  $\hat{\gamma}(h)$ , and model fitting with a valid theoretical model  $\gamma(h)$ , to capture the actual spatial dependence of the data. Variogram analysis provides a helpful tool for summarizing and measuring spatial dependence in spatial data. However, its main contribution is to estimate the value of the spatial variable at an unsampled location within some inference procedures, such as the Kriging method. The latter approach differs from classical regression, in that local features can affect the solution, and so it considers a weighted mean. Therefore, estimating a variogram plays a decisive role, as it is commonly used to find the optimal values of the weights.

We considered the widely used exponential and spherical semivariogram models, as well as the Gaussian and wave models. The latter was also considered because of its atypical irregular behavior:

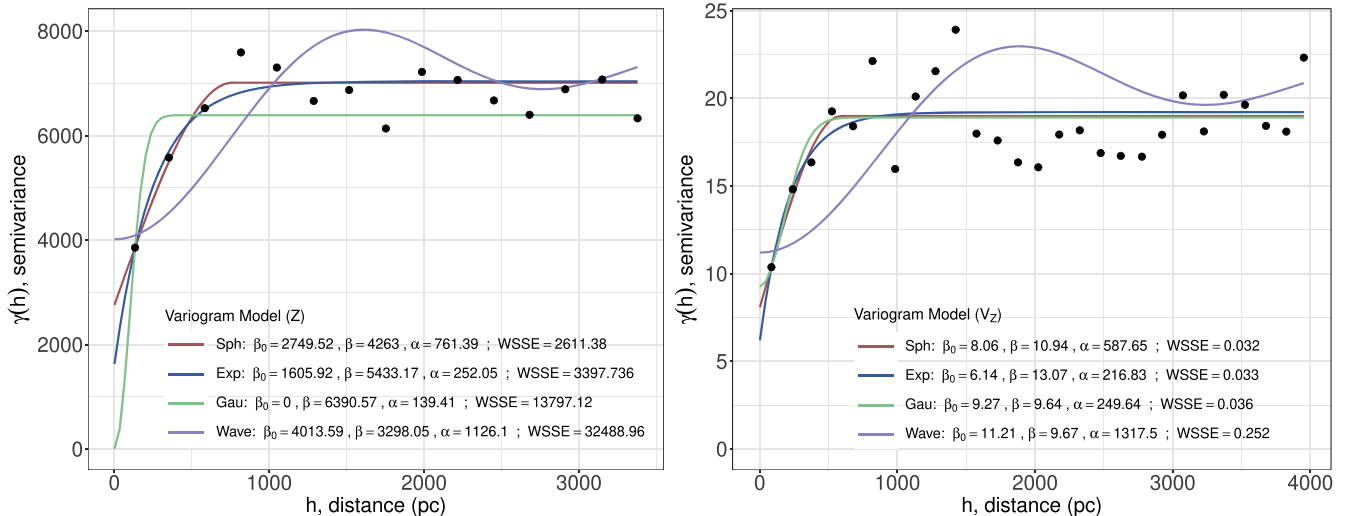
$$\gamma_{\text{sph}}(h) = \begin{cases} \beta_0 + \beta \left( \frac{3h}{2\alpha} - \frac{1}{2} \left( \frac{h}{\alpha} \right)^3 \right), & 0 < h \leq \alpha, \\ \beta_0 + \beta & h > \alpha \end{cases}, \quad (\text{A4})$$

$$\gamma_{\text{exp}}(h) = \beta_0 + \beta(1 - e^{-h/\alpha}), \quad h > 0, \quad (\text{A5})$$

$$\gamma_{\text{gau}}(h) = \beta_0 + \beta(1 - e^{-h^2/\alpha^2}), \quad h > 0, \quad (\text{A6})$$

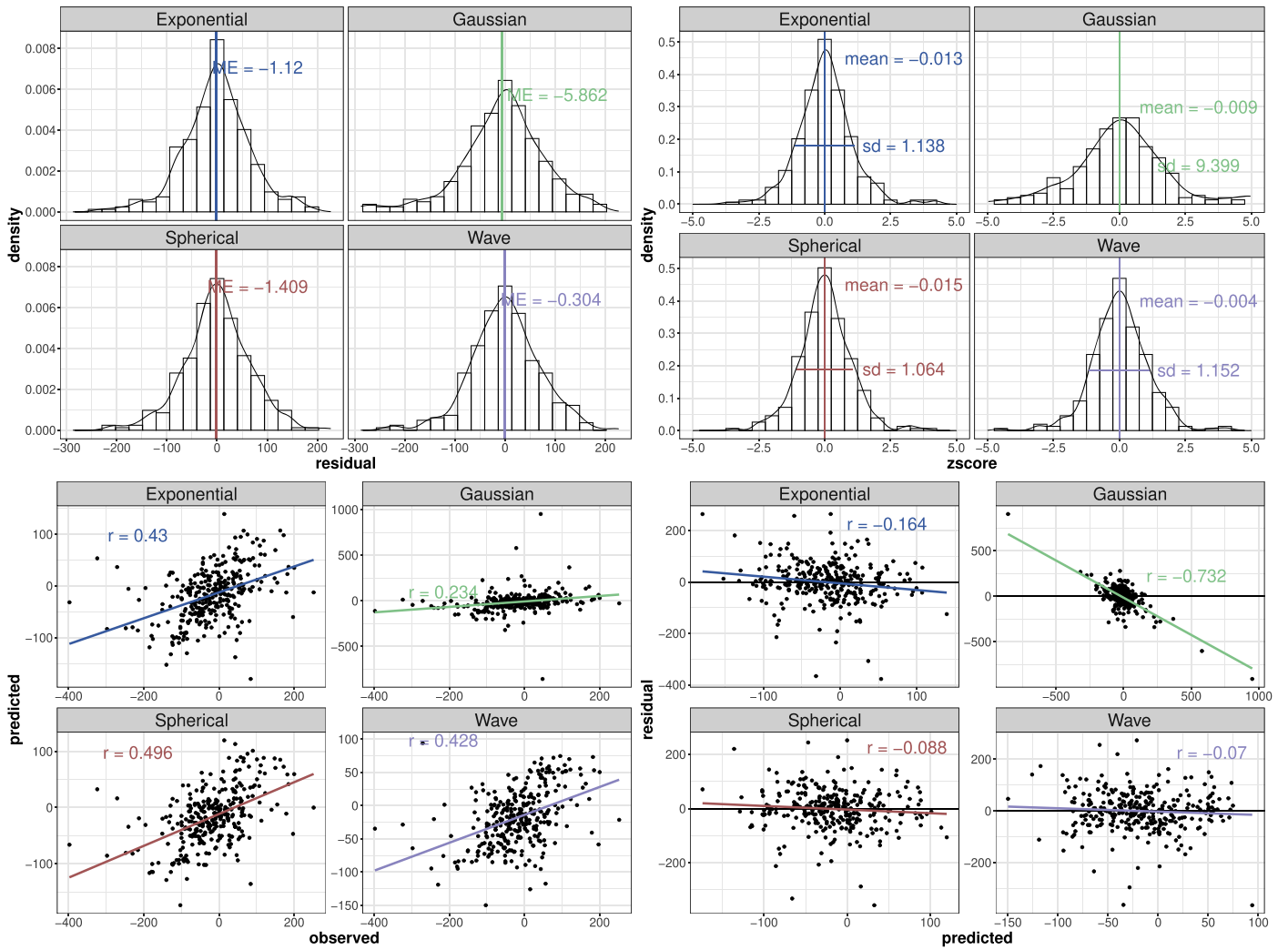
$$\gamma_{\text{wave}}(h) = \beta_0 + \beta \left( 1 - \alpha \frac{\sin(-h/\alpha)}{h} \right), \quad h > 0. \quad (\text{A7})$$

The main characteristics, or parameters, of many semivariogram models are as follows. (i) The nugget effect,  $\beta_0$ . In theory, the semivariogram should be zero for a lag distance  $h$  of zero, but in



**Figure 15.** Variogram fit model selection, corresponding to the  $Z$  Galactic coordinate on the left and  $V_z$  on the right. Four different models for the variogram fit have been tested: spherical “Sph,” exponential “Exp,” Gaussian “Gau,” and wave. The spherical model appears to be the best choice, having the lowest WSSE statistics.





**Figure 16.** Cross-validation analysis for the  $Z$  variogram model selection. The bottom left panels show the correlation between the observed and predicted data, which is ideally close to 1. The bottom right panels show the correlation between the predictions and the residuals, which is ideally close to zero. In both cases, except for the Gaussian model, the results are quite similar.

**Table 3**  
Variogram Fit Results

Model	$Z$			$V_Z$		
	WSSE	AIC	$R^2$	WSSE	AIC	$R^2$
Exponential	3397.74	-257.89	99.98	0.03	-151.36	93.54
Gaussian	13797.12	-257.86	98.34	0.04	-116.22	93.58
Spherical	2611.38	-257.89	99.98	0.03	-150.99	93.35
Wave	32488.96	-257.98	99.98	0.25	-152.53	92.5

practice it can be significantly different from zero. This nonzero value  $\beta_0$  reflects local effects or sampling errors. (ii) Another important characteristic is that the semivariance function  $\gamma(h)$  can approach, or asymptotically converge to, a constant value known as the *sill*,  $\beta_0 + \beta$ . When there is a nugget  $\beta_0 > 0$ ,  $\beta$  is called the partial sill. (iii) The distance at which the semivariogram reaches the sill is called the range,  $\alpha$ . This means that there is a distance beyond which the correlation between variables is zero. Sample locations separated by distances shorter than the range are spatially autocorrelated, while locations that are further apart than the range are not.

Figure 15 shows the different variogram models described above, for both the  $Z$  and  $V_Z$  variables, fitted to the sample or

empirical semivariogram  $\hat{\gamma}(h)$ , represented with the black dots. The sample variograms and covariograms are calculated from the predicted residuals  $\hat{e}(r_i) = z(r_i) - \hat{m}(r_i)$ , with  $\hat{m}(r_i) = F\hat{\beta}$  being the ordinary least squares estimates of the mean, as (equivalent to Equation (2))

$$\hat{\gamma}(\bar{h}_j) = \frac{1}{2N_j} \sum_{i=1}^{N_j} (\hat{e}(r_i) - \hat{e}(r_i + h))^2, \quad \forall(r_i, r_i + h): h \in [h_j, h_j + \delta], \quad (\text{A8})$$

for a number  $N_j$  of pairs within a regular distance interval  $[h_j, h_j + \delta]$ , with  $\bar{h}_j$  being the average of such an interval.

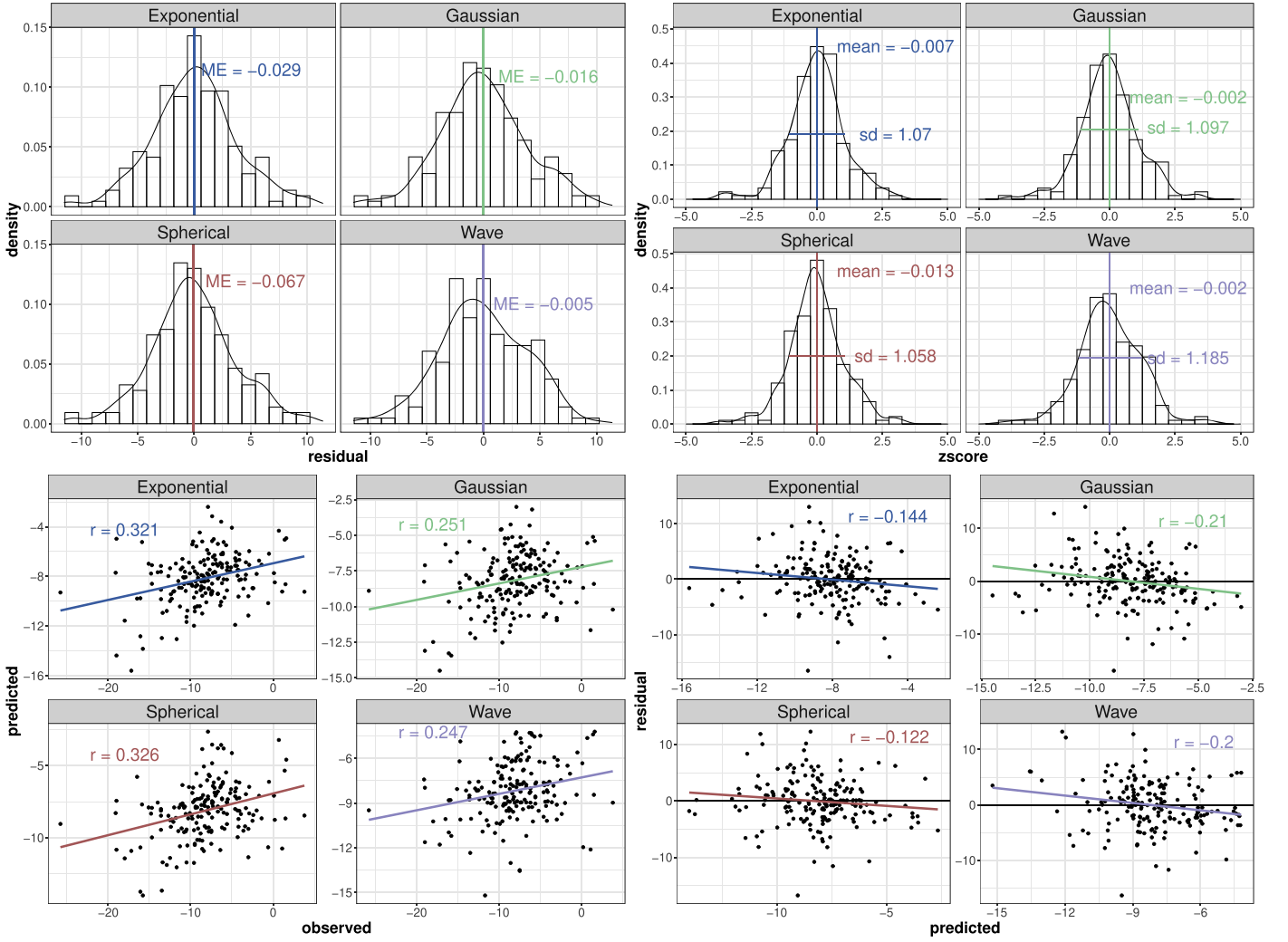


Figure 17. Analogous to Figure 16, cross-validation analysis for the  $V_Z$  variogram model selection.

### A.2. Cross Validation and Model Selection

The fitting method at the model variogram fitting stage uses nonlinear regression to fit the coefficients or parameters of the model. For this, the following WSSE is minimized,

$$\text{WSSE} = \sum_{j=1}^n \omega_j (\hat{\gamma}(\bar{h}_j) - \gamma(\bar{h}_j))^2, \quad (\text{A9})$$

where  $\gamma(h)$  is according to the chosen variogram model (from (A4) to (A7)) and  $\hat{\gamma}(h)$  is the empirical variogram estimation (A8). We have chosen the weights to be the ratio between the number of data point pairs grouped into the corresponding distance intervals and the distance itself, i.e.,  $\omega_j = N_j/h_j$ .

We use this statistic to evaluate how well the different models fit the data and to determine which one is the best fit for the data. We also compute the Akaike information criterion (AIC) to compare the different possible models; however, the AIC results are pretty similar in all cases (bearing in mind that all these models have the same number of parameters). The WSSE results are more discriminant for the case of the  $Z$  component, as can be checked in Figure 15, where the spherical model seems to better fit the data. For the case of the  $V_Z$  component, the results are also similar for both WSSE and AIC between the different models. Therefore, we choose the

spherical model for variogram fitting, based on the discriminant result in  $Z$ .

Besides the above statistics criterion, we also carry out cross validation, another means of model diagnostics for testing the quality of the variogram fitting itself and for model comparison. The data set is split into two sets: a modeling set and a validation one (we took one-fifth of the data or an  $n=5$ -fold partition). The former is used for variogram modeling and Kriging predictions on the locations of the validation set. Then the validation measurements can be compared to their predictions. This process is repeated a hundred times, with the modeling and validation sets being randomly chosen at each run.





We can then compute the sum of the squared errors from the mean or  $R^2$ , another goodness-of-fit measurement, which indicates by how much the Kriging prediction is a better predictor than the mean. The average  $R^2$  is around 99% for all models in the  $Z$  component (slightly lower for the Gaussian model), and around 93% for the  $V_Z$  variable. So  $R^2$  is not a determinant for model selection in this case; see Table 3.

We also inspect the errors from the cross-validation analysis, such as the residuals  $Z(r_i) - \hat{Z}(r_i)$  and the z-scores  $z_i = (Z(r_i) - \hat{Z}(r_i)) / \sigma_{\text{krig}}(r_i)$  (the standardized residual, but taking into account the corresponding Kriging standard error).

Figures 16 and 17 display these cross-validation errors in the top panels, for one of the random runs. If the variogram fitting is correct, the cross-validation residuals are small, have zero mean, and have no apparent structure. This appears to be the case for the four models, for both  $Z$  and  $V_Z$ . In contrast to standard residuals, the  $z$ -score takes the Kriging variance into account. Whereas we expect small residuals with no trend, if the variogram model is correct, the  $z$ -score should have mean and variance values close to 0 and 1, respectively. For the  $Z$  variable, we can discard the Gaussian model, as it has a dispersion more than 1, which may imply some misfitting. The Spherical model would head the model rankings, but with little difference from the Exponential one. For the  $V_Z$  variable, we cannot see meaningful differences between the models.

Finally, we check these errors with respect to the predictions  $\hat{Z}(r_i)$  and the observed  $Z(r_i)$ , or the test data set at each  $n$ -fold, in the bottom panels of Figures 16 and 17. Visual verification that the variogram model fit is sensible is obtained through the observations versus predictions scatter plots (bottom left panels) and the predictions versus residuals plots (bottom right panels). We expect a high correlation (the higher the correlation, the better the model) for the former and a low correlation for the latter (the residuals should be random, without any trend). We can observe that the Spherical and Exponential models are slightly better than the other two. The different models yield similar results for the  $V_Z$  variable, where the Spherical and Exponential models are somewhat better, analogous to the  $Z$  case. The only exception is the Gaussian model for the  $Z$  variable, where the predictions versus residuals plot shows a clear trend, likely attributed to a misfitting. We then choose the Spherical model, based on this final model diagnosis, together with a moderately better WSSE statistic. We use the same model for homogeneity for both the  $Z$  and  $V_Z$  variables.

### ORCID iDs

Emilio J. Alfaro  <https://orcid.org/0000-0002-2234-7035>  
 Manuel Jiménez  <https://orcid.org/0000-0002-8651-9377>  
 M. Carmen Sánchez-Gil  <https://orcid.org/0000-0003-0829-4157>  
 Néstor Sánchez  <https://orcid.org/0000-0002-0042-3180>  
 Marta González  <https://orcid.org/0000-0002-6823-3869>  
 Jesús Maíz Apellániz  <https://orcid.org/0000-0003-0825-3443>

### References

- Alfaro, E. J., Cabrera-Cano, J., & Delgado, A. J. 1991, *ApJ*, **378**, 106  
 Alfaro, E. J., Cabrera-Cano, J., & Delgado, A. J. 1992, *ApJ*, **399**, 576  
 Alfaro, E. J., & Efremov, Y. N. 1996, *RMxAC*, **4**, 1  
 Alfaro, E. J., Pérez, E., & Franco, J. 2004, *How Does the Galaxy Work?*, Astrophysics and Space Science Library, Vol. 315 (Dordrecht: Kluwer)  
 Alfaro, E. J., Pérez, E., González Delgado, R. M., Martos, M. A., & Franco, J. 2001, *ApJ*, **550**, 253  
 Allende Prieto, C., Majewski, S. R., Schiavon, R., et al. 2008, *AN*, **329**, 1018  
 Alves, J., Zucker, C., Goodman, A. A., et al. 2020, *Natur*, **578**, 237  
 Ankerst, M., Breunig, M. M., Kriegel, H.-P., & Sander, J. 1999, *SIGMOD Rec.*, **28**, 49  
 Antoja, T., Figueras, F., Fernández, D., & Torra, J. 2008, *A&A*, **490**, 135  
 Antoja, T., Helmi, A., Romero-Gómez, M., et al. 2018, *Natur*, **561**, 360  
 Asaiin, R., Figueras, F., & Torra, J. 1999, *A&A*, **350**, 434  
 Bally, J. 2008, in *Handbook of Star Forming Regions, Volume I: The Northern Sky*, ed. B. Reipurth, Vol. 4 (San Francisco, CA: ASP), 459  
 Barros, D. A., Pérez-Villegas, A., Lépine, J. R. D., Michtchenko, T. A., & Vieira, R. S. S. 2020, *ApJ*, **888**, 75  
 Battaner, E., Florido, E., & Sanchez-Saavedra, M. L. 1990, *A&A*, **236**, 1  
 Becker, W. 1964, *ZA*, **58**, 202  
 Bekki, K. 2009, *MNRAS Lett.*, **398**, L36  
 Bennett, M., & Bovy, J. 2019, *MNRAS*, **482**, 1417  
 Berdnikov, L. N., & Efremov, Y. N. 1993, *AstL*, **19**, 389  
 Bland-Hawthorn, J., Sharma, S., Tepper-García, T., et al. 2019, *MNRAS*, **486**, 1167  
 Bossini, D., Vallenari, A., Bragaglia, A., et al. 2019, *A&A*, **623**, A108  
 Bouy, H., & Alves, J. 2015, *A&A*, **584**, A26  
 Cánovas, H., Cantero, C., Cieza, L., et al. 2019, *A&A*, **626**, A80  
 Cantat-Gaudin, T., & Anders, F. 2020, *A&A*, **633**, A99  
 Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., et al. 2020, *A&A*, **640**, A1  
 Carballo-Bello, J. A., Martínez-Delgado, D., Corral-Santana, J. M., et al. 2021, *MNRAS*, **501**, 1690  
 Castro-Ginard, A., McMillan, P. J., Luri, X., et al. 2021, *A&A*, **652**, A162  
 Chen, B., Asiain, R., Figueras, F., & Torra, J. 1997, *A&A*, **318**, 29  
 Cheng, X., Anguiano, B., Majewski, S. R., et al. 2020, *ApJ*, **905**, 49  
 Chilès, J.-P., & Delfiner, P. 2000, *Geostatistics: Modeling Spatial Uncertainty* (New York: Wiley), 695  
 Cohen, R. J., & Dent, W. R. F. 1983, in *Surveys of the Southern Galaxy*, Astrophysics and Space Science Library, ed. W. B. Burton & F. P. Israel, Vol. 105 (Dordrecht: Springer), 159  
 Combes, F. 1991, *ARA&A*, **29**, 195  
 Comerón, F. 1999, *A&A*, **351**, 506  
 Comerón, F., & Torra, J. 1994, *A&A*, **281**, 35  
 Comerón, F., Torra, J., & Gomez, A. E. 1994, *A&A*, **286**, 789  
 Costado, M. T., Alfaro, E. J., González, M., & Sampedro, L. 2017, *MNRAS*, **465**, 3879  
 Cox, D. P., & Gómez, G. C. 2002, *ApJS*, **142**, 261  
 Cui, T., Pagendam, D., & Gilfedder, M. 2021, *Environ. Modell. & Software*, **144**, 105170  
 Cui, X.-Q., Zhao, Y.-H., Chu, Y.-Q., et al. 2012, *RAA*, **12**, 1197  
 De Silva, G. M., Freeman, K. C., Bland-Hawthorn, J., et al. 2015, *MNRAS*, **449**, 2604  
 Dias, W. S., Alessi, B. S., Moitinho, A., & Lépine, J. R. D. 2002, *A&A*, **389**, 871  
 Dias, W. S., Monteiro, H., Moitinho, A., et al. 2021, *MNRAS*, **504**, 356  
 Dinnbier, F., Kroupa, P., & Anderson, R. I. 2022, *A&A*, **660**, A61  
 Dixon, M. E. 1967, *MNRAS*, **137**, 337  
 Efremov, Y. N., & Elmeegreen, B. G. 1998, *MNRAS*, **299**, 588  
 Efremov, Y. N., & Sitnik, T. G. 1988, *SvAL*, **14**, 347  
 Elias, F., Alfaro, E. J., & Cabrera-Caño, J. 2006a, *AJ*, **132**, 1052  
 Elias, F., Alfaro, E. J., & Cabrera-Cano, J. 2009, *MNRAS*, **397**, 2  
 Elias, F., Cabrera-Caño, J., & Alfaro, E. J. 2006b, *AJ*, **131**, 2700  
 Elmeegreen, B. G. 2011, *EAS Publ. Ser.*, **51**, 31  
 Elmeegreen, B. G., Efremov, Y., Pudritz, R. E., & Zinnecker, H. 2000, in *Protostars and Planets IV*, ed. V. Mannings, A. P. Boss, & S. S. Russell (Tucson, AZ: Univ. Arizona Press), 179  
 Elmeegreen, B. G., & Efremov, Y. N. 1996, *ApJ*, **466**, 802  
 Elmeegreen, B. G., & Efremov, Y. N. 1997, *ApJ*, **480**, 235  
 Elmeegreen, B. G., Elmeegreen, D. M., & Montenegro, L. 1992, *ApJS*, **79**, 37  
 Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. 1996, in *Proc. Second Int. Conf. Knowledge Discovery and Data Mining, KDD'96* (Palo Alto, CA: AAAI Press), 226, <http://dl.acm.org/citation.cfm?id=3001460.3001507>  
 Everall, A., Belokurov, V., Evans, N. W., Boubert, D., & Grand, R. J. J. 2022, *MNRAS*, **511**, 3863  
 Federath, C., Klessen, R. S., & Schmidt, W. 2009, *ApJ*, **692**, 364  
 Feigelson, E. D., & Babu, G. J. 2012, *Modern Statistical Methods for Astronomy: With R Applications* (Cambridge: Cambridge Univ. Press)  
 Feitzinger, J. V., & Spicker, J. 1985, *MNRAS*, **214**, 539  
 Florido, E., Battaner, E., Gros, A., Prieto, M., & Mediavilla, E. 1992, *Ap&SS*, **190**, 293  
 Florido, E., Battaner, E., Prieto, M., Mediavilla, E., & Sanchez-Saavedra, M. L. 1991, *MNRAS*, **251**, 193  
 Franco, J., Kim, J., Alfaro, E. J., & Hong, S. S. 2002, *ApJ*, **570**, 647  
 Franco, J., Tenorio-Tagle, G., Bodenheimer, P., Rozyczka, M., & Mirabel, I. F. 1988, *ApJ*, **333**, 826  
 Freudenreich, H. T., Berriman, G. B., Dwek, E., et al. 1994, *ApJL*, **429**, L69  
 Fux, R. 2001, *A&A*, **373**, 511  
 Gaia Collaboration, Babusiaux, C., van Leeuwen, F., et al. 2018, *A&A*, **616**, A10  
 Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, *A&A*, **649**, A1  
 Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, **595**, A1  
 Gilmore, G., Randich, S., Asplund, M., et al. 2012, *Msngr*, **147**, 25  
 Gómez, G. C., & Cox, D. P. 2002, *ApJ*, **580**, 235  
 Gontcharov, G. A., & Mosenkov, A. V. 2019, *MNRAS*, **483**, 299  
 González, M., Joncour, I., Buckner, A. S. M., et al. 2021, *A&A*, **647**, A14



- Gould, B. A. 1879, *RNAO*, **1**, **1**
- Grassberger, P., & Procaccia, I. 1983, *PhRvL*, **50**, 346
- Gum, C. S., Kerr, F. J., & Westerhout, G. 1960, *MNRAS*, **121**, 132
- Henderson, A. P. 1967, *AJ*, **72**, 803
- Henderson, A. P. 1968, *AJS*, **73**, 182
- Henderson, A. P., Jackson, P. D., & Kerr, F. J. 1982, *ApJ*, **263**, 116
- Herschel, J. F. W. S. 1847, Results of astronomical observations made during the years 1834, 5, 6, 7, 8, at the Cape of Good Hope; being the completion of a telescopic survey of the whole surface of the visible heavens, commenced in 1825 (London: Smith, Elder and Co.)
- Hou, L. G. 2021, *FrASS*, **8**, 103
- Janes, K., & Adler, D. 1982, *ApJS*, **49**, 425
- Kamphuis, J. J. 1993, PhD Thesis, Univ. of Groningen, Netherlands
- Kerr, F. J. 1958, *RvMP*, **30**, 924
- Kerr, F. J. 1970, in IAU Symp. 38, The Spiral Structure of our Galaxy, ed. W. Becker & G. I. Kontopoulos (Cambridge: Cambridge Univ. Press), 95
- Khanna, S., Sharma, S., Tepper-García, T., et al. 2019, *MNRAS*, **489**, 4962
- Kharchenko, N. V., Piskunov, A. E., Röser, S., Schilbach, E., & Scholz, R. D. 2005, *A&A*, **438**, 1163
- Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., & Scholz, R. D. 2013, *A&A*, **558**, A53
- Koukkel, M., Covey, K., & Stassun, K. G. 2020, *AJ*, **160**, 279
- Krige, D. G. 1951, *J. South. Afr. Inst. Min. Metall.*, **52**, 119
- Krone-Martins, A., & Moitinho, A. 2014, *A&A*, **561**, A57
- Lada, C. J., & Lada, E. A. 2003, *ARA&A*, **41**, 57
- Lallement, R., Welsh, B. Y., Vergely, J. L., Crifo, F., & Sfeir, D. 2003, *A&A*, **411**, 447
- Laporte, C. F. P., Johnston, K. V., Gomez, F. A., Garavito-Camargo, N., & Besla, G. 2018, *MNRAS*, **481**, 286
- Laporte, C. F. P., Minchev, I., Johnston, K. V., & Gomez, F. A. 2019, *MNRAS*, **485**, 3134
- Lepine, S. 2018, in 20th Cambridge Workshop on Cool Stars, Stellar Systems and the Sun (Zenodo), 84
- Levine, E. S., Blitz, L., & Heiles, C. 2006, *ApJ*, **643**, 881
- Li, Z.-Y. 2021, *ApJ*, **911**, 107
- Lindblad, P. O., Palous, J., Loden, K., & Lindegren, L. 1997, *ESA Spec. Publ.*, **402**, 507
- Liszt, H. S., & Burton, W. B. 1980, *ApJ*, **236**, 779
- Liu, L., & Pang, X. 2019, *ApJS*, **245**, 32
- Lockman, F. J. 1977, *AJ*, **82**, 408
- Lockman, F. J., & Bania, T. M. 1981, *BAAS*, **13**, 538
- Lombardi, M. 2002, *A&A*, **395**, 733
- Lynga, G. 1982, *A&A*, **109**, 213
- Maíz Apellániz, J. 2001, *AJ*, **121**, 2737
- Maíz Apellániz, J., Alfaro, E. J., & Sota, A. 2008, arXiv:0804.2553
- Maíz Apellániz, J., Pantaleoni González, M., Barbá, R. H., & Weiler, M. 2022, *A&A*, **657**, A72
- Makarov, V. V. 2007, *ApJS*, **169**, 105
- Malhotra, S. 1994, *ApJ*, **433**, 687
- Martínez-Delgado, D., Peñarrubia, J., Jurić, M., Alfaro, E. J., & Ivezić, Z. 2007, *ApJ*, **660**, 1264
- Martínez-Medina, L., Pérez-Villegas, A., & Peimbert, A. 2022, *MNRAS*, **512**, 1574
- Matheron, G. 1963, *Econ. Geol.*, **58**, 1246
- Matheron, G. 1965, Les variables régionalisées et leur estimation: une application de la théorie de fonctions aléatoires aux sciences de la nature, Vol. 4597 (Paris: Masson et CIE)
- Matthews, L. D., & Uson, J. M. 2008, *ApJ*, **688**, 237
- Mead, K. N., & Kutner, M. L. 1988, *ApJ*, **330**, 399
- Menon, S. H., Grasha, K., Elmegreen, B. G., et al. 2021, *MNRAS*, **507**, 5542
- Michtchenko, T. A., Barros, D. A., Pérez-Villegas, A., & Lépine, J. R. D. 2019, *ApJ*, **876**, 36
- Michtchenko, T. A., Lépine, J. R. D., Pérez-Villegas, A., Vieira, R. S. S., & Barros, D. A. 2018, *ApJL*, **863**, L37
- Miller, A. E., Cioni, M.-R. L., de Grijs, R., et al. 2022, *MNRAS*, **512**, 1196
- Minchev, I., Boily, C., Siebert, A., & Bienayme, O. 2010, *MNRAS*, **407**, 2122
- Moitinho, A., Vázquez, R. A., Carraro, G., et al. 2006, *MNRAS Lett.*, **368**, L77
- Montes, D., Lopez-Santiago, J., Galvez, M. C., et al. 2001, *MNRAS*, **328**, 45
- Moreno, E., Alfaro, E. J., & Franco, J. 1999, *ApJ*, **522**, 276
- Nandakumar, M., Narayan, C., & Dutta, P. 2022, *MNRAS*, **513**, 3065
- Narayan, C. A., Dettmar, R.-J., & Saha, K. 2020, *MNRAS*, **495**, 3705
- Negueruela, I., & Marco, A. 2003, *A&A*, **406**, 119
- Nelson, P., & Widrow, L. M. 2022, arXiv:2206.04627
- Newberg, H. J., Yanny, B., Rockosi, C., et al. 2002, *ApJ*, **569**, 245
- Odekon, M. C. 2008, *ApJ*, **681**, 1248
- Olano, C. A. 1982, *A&A*, **112**, 195
- Oliver, M. A., & Webster, R. 2014, *Catena*, **113**, 56
- Pang, X., Tang, S.-Y., Li, Y., et al. 2022, *ApJ*, **931**, 156
- Pantaleoni González, M., Maíz Apellániz, J., Barbá, R. H., & Reed, B. C. 2021, *MNRAS*, **504**, 2968
- Pastorello, N., Forbes, D. A., Foster, C., et al. 2014, *MNRAS*, **442**, 1003
- Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, *A&A*, **500**, 501
- Piskunov, A. E., Kharchenko, N. V., Röser, S., Schilbach, E., & Scholz, R. D. 2006, *A&A*, **445**, 545
- Poggio, E., Drimmel, R., Cantat-Gaudin, T., et al. 2021, *A&A*, **651**, A104
- Poppel, W. 1997, *FCPh*, **18**, 1
- Portegies Zwart, S. F., McMillan, S. L. W., & Gieles, M. 2010, *ARA&A*, **48**, 431
- Quiroga, R. J. 1974, *Ap&SS*, **27**, 323
- Quiroga, R. J. 1977, *Ap&SS*, **50**, 281
- Quiroga, R. J., & Schlosser, W. 1977, *A&A*, **57**, 455
- Rasmussen, C. E., & Williams, C. K. I. 2005, *Gaussian Processes for Machine Learning* (Cambridge, MA: MIT Press)
- Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, *ApJ*, **783**, 130
- Romero-Gómez, M., Mateu, C., Aguilar, L., Figueras, F., & Castro-Ginard, A. 2019, *A&A*, **627**, A150
- Ruiz-Lara, T., Gallart, C., Bernard, E. J., & Cassisi, S. 2020, *NatAs*, **4**, 965
- Sánchez, N., Añez, N., Alfaro, E. J., & Crone Odekon, M. 2010, *ApJ*, **720**, 541
- Sánchez, N., & Alfaro, E. J. 2008, *ApJS*, **178**, 1
- Sánchez, N., Alfaro, E. J., Elias, F., Delgado, A. J., & Cabrera-Caño, J. 2007a, *ApJ*, **667**, 213
- Sánchez, N., Alfaro, E. J., & Pérez, E. 2005, *ApJ*, **625**, 849
- Sánchez, N., Alfaro, E. J., & Pérez, E. 2007b, *ApJ*, **656**, 222
- Sanchez-Gil, M. C., Alfaro, E. J., & Perez, E. 2015, *MNRAS*, **454**, 3376
- Sánchez-Saavedra, M. L., Battaner, E., & Florido, E. 1990, *MNRAS*, **246**, 458
- Santillán, A., Franco, J., Martos, M., & Kim, J. 1999, *ApJ*, **515**, 657
- Schmidt-Kaler, T. 1975, *VA*, **19**, 69
- Schonrich, R., & Dehnen, W. 2018, *MNRAS*, **478**, 3809
- Simpson, T. W., Poplinski, J. D., Koch, P. N., & Allen, J. K. 2001, *Engineering with Computers*, **17**, 129
- Soubiran, C., Jasniewicz, G., Chemin, L., et al. 2013, *A&A*, **552**, A64
- Soubiran, C., Cantat-Gaudin, T., Romero-Gómez, M., et al. 2018, *A&A*, **619**, A155
- Spicker, J., & Feitzinger, J. V. 1986, *A&A*, **163**, 43
- Steinmetz, M., Zwitter, T., Siebert, A., et al. 2006, *AJ*, **132**, 1645
- Tarricq, Y., Soubiran, C., Casamiquela, L., et al. 2021, *A&A*, **647**, A19
- Taylor, D. K., Dickman, R. L., & Scoville, N. Z. 1987, *ApJ*, **315**, 104
- Taylor, M. B. 2005, in ASP Conf. Ser. 347, *Astronomical Data Analysis Software and Systems XIV*, ed. P. Shopbell, M. Britton, & R. Ebert (San Francisco, CA: ASP), 29
- Tepper-García, T., Bland-Hawthorn, J., & Freeman, K. 2022, *MNRAS*, **515**, 5951
- Tian, H.-J., Liu, C., Wu, Y., Xiang, M.-S., & Zhang, Y. 2018, *ApJL*, **865**, L19
- Turner, D. G. 2014, *CaJPh*, **92**, 959
- Vallée, J. P. 2021, *MNRAS*, **506**, 523
- Vallée, J. P. 2022, *Ap&SS*, **367**, 26
- van den Bergh, S. 1966, *AJ*, **71**, 990
- van Leeuwen, F. 2009, *A&A*, **497**, 209
- van Terwisga, S. E., Hacar, A., van Dishoeck, E. F., Oonk, R., & Portegies Zwart, S. 2022, *A&A*, **661**, A53
- Varsavsky, C. M., & Quiroga, R. J. 1970, in IAU Symp. 38, The Spiral Structure of our Galaxy, ed. W. Becker & G. I. Kontopoulos (Cambridge: Cambridge Univ. Press), 147
- Vázquez, R. A., May, J., Carraro, G., et al. 2008, *ApJ*, **672**, 930
- Wang, H.-F., Carlin, J. L., Huang, Y., et al. 2019, *ApJ*, **884**, 135
- Webster, R., & Oliver, M. A. 2007, *Geostatistics for Environmental Scientists* (Second Edition; Chichester: John Wiley & Sons)
- Widmark, A., Widrow, L. M., & Naik, A. 2022, arXiv:2207.03492
- Widrow, L. M., Gardner, S., Yanny, B., Dodelson, S., & Chen, H.-Y. 2012, *ApJL*, **750**, L41
- Wouterloot, J. G. A., Brand, J., Burton, W. B., & Kwee, K. K. 1990, *A&A*, **230**, 21
- Wright, N. J. 2020, *NewAR*, **90**, 101549
- Xu, Y., Hou, L. G., Bian, S. B., et al. 2021, *A&A*, **645**, L8
- Yuan, L., & Wallace, C. 1973, *ApJ*, **185**, 453
- Zari, E., Hashemi, H., Brown, A. G. A., Jardine, K., & de Zeeuw, P. T. 2018, *A&A*, **620**, A172
- Zucker, C., Speagle, J. S., Schlafly, E. F., et al. 2020, *A&A*, **633**, A51
- Zucker, C., Speagle, J. S., Schlafly, E. F., et al. 2019, *ApJ*, **879**, 125
- Zucker, C., Goodman, A. A., Alves, J., et al. 2022, *Natur*, **601**, 334