A Gaussian Process Model for the Local Galactic Velocity Field

by

Patrick Nelson

A thesis submitted to the Department of Physics, Engineering Physics, and Astronomy in conformity with the requirements for the degree of Master of Science

Queen’s University
Kingston, Ontario, Canada
December 2022

Copyright © Patrick Nelson, 2022
Abstract

The Milky Way Galaxy has been a source of inspiration and challenge for astronomers who have focused their efforts to understand the motions of stars, the spin of the Galaxy, and the connection beyond to dark matter and our universe. These efforts are being propelled by the Gaia mission, a space observatory dedicated to cataloging more than one billion stars in the Galaxy. Gaia has observed more than 33 million stars with both the positions and velocities in each dimension to give a full six dimensional solution. The next challenge in Galactic astronomy is to analyze and understand how the Galaxy’s motion fits with our knowledge of theory, and the departures within. The size and complexity of the data make it a challenge to analyze the underlying motion. In this thesis we focus on the mean velocity field in the vicinity of the Sun. A number of researchers have constructed views of Galactic motion via binning. While conceptually simple, the connection to theory can be tenuous. Here we use Gaussian process regression, a non-parametric machine learning method that presents an opportunity for modelling the bulk motion of the Galaxy. This method can inform both the current motion and future state of the Galaxy, calculating continuous derivatives of the velocity field. We discuss the use of observational data with Gaussian process regression, modifications to modelling for use with large data, and highlight the observed motions and their connection to our understanding of the Galaxy. Models agree strongly with Galactic motions explored in literature, and provide a measure of the divergence for examining time evolution of regions of the Galaxy.
Statement of Co-authorship

The work presented in this thesis has been done by myself (Patrick Nelson) under the supervision and guidance of Lawrence M. Widrow.

Chapter 5 is a published paper co-authored by myself and Lawrence M. Widrow, for which I am the primary author. I performed the programming, modelling, construction of all figures and writing much of the content. Lawrence M. Widrow and I were actively engaged in discussions throughout the work, with the original goal of the project derived from discussions between Widrow and collaborators. In this work, he wrote content around background, the connection of velocity gradients to theory, and edited the paper.
Acknowledgments

I first thank Larry Widrow for all the work, guidance, and care he has given to supervising me. Prior to working with him I had no formal astronomy background, but combining interests in data science and a love for learning about astrophysics has driven our work projects. I am thankful for the opportunity and freedom to explore topics to study, and to learn from the style and elegance that Larry brings to writing and projects.

To my faculty and graduate course teachers, Judith Irwin, Mark Richardson, Aaron Vincent, Sarah Sadavoy, and Laura Fissel, thank you for the experience and skill I have gained through your courses. Each course explored a different area of physics than my thesis, but the thinking and insights have helped me throughout my degree and I certainly will not forget the mentorship provided.

Lastly, I thank my family, my partner Tiffanie, and my friends in supporting me through the most difficult times of my degree. The wonderful people who have been there to help or distract me with good times are irreplaceable.
Contents

Abstract i

Statement of Co-authorship ii

Acknowledgments iii

Contents iv

List of Tables vii

List of Figures viii

Abbreviations x

Symbols xii

Chapter 1: Introduction 1

Chapter 2: Galactic Dynamics 5
  2.1 The Stellar Distribution Function 6
  2.2 Velocity Field in Equilibrium 8
  2.3 Departures from Equilibrium 9
    2.3.1 The Spiral Arms 10
    2.3.2 The Bar 12
    2.3.3 Random Motions 13
    2.3.4 External Sources 14
  2.4 Galactic Parameters 14

Chapter 3: Gaussian Process Regression 16
  3.1 Background 16
  3.2 Mathematical Basis of Gaussian Processes 20
  3.3 Covariance Functions 24
    3.3.1 Radial Basis Function 24
# Table of Contents

## Chapter 3: Gaussian Process Regression

- 3.3.2 Linear Function ................................................. 25
- 3.3.3 Periodic Function .............................................. 26
- 3.3.4 Functions of Kernels ........................................... 27
- 3.4 Hyperparameter Optimization ................................... 28
- 3.5 Heteroscedastic Noise ............................................ 30
- 3.6 Sparse Gaussian Process Regression .......................... 31
  - 3.6.1 Inducing point method ...................................... 31
  - 3.6.2 Deterministic Training Conditional ....................... 34
  - 3.6.3 Implementation .............................................. 35

## Chapter 4: Application of Gaussian Process Regression to \textit{Gaia} Data

- 4.1 Astrometry ......................................................... 39
  - 4.1.1 Color and Magnitudes ...................................... 40
  - 4.1.2 Coordinate Transformations ................................ 42
  - 4.1.3 Uncertainty in Velocity .................................... 45
- 4.2 Datasets ............................................................ 46
  - 4.2.1 \textit{Gaia} DR2 Radial Velocity Sample .................. 47
  - 4.2.2 \textit{Gaia} DR3 Radial Velocity Sample .................. 49
- 4.3 Application to GP ............................................... 52
  - 4.3.1 Model Setup ................................................. 53

## Chapter 5: Gaussian Process Model for the Local Stellar Velocity Field from Gaia Data Release 2

- 5.1 Introduction ....................................................... 56
- 5.2 Preliminaries ....................................................... 61
  - 5.2.1 Six-Dimensional Phase-space Catalogue .................. 61
- 5.3 Binning ............................................................. 63
- 5.4 Gaussian Process Regression .................................... 64
  - 5.4.1 Overview of Gaussian Processes .......................... 64
  - 5.4.2 Kernel Function .............................................. 66
  - 5.4.3 Sparse GP Regression via Inducing Points ............... 68
  - 5.4.4 Mock Data Tests ............................................. 69
- 5.5 Results ............................................................. 71
  - 5.5.1 Velocity Field in Components ............................. 72
  - 5.5.2 Velocity Vector Field ....................................... 76
- 5.6 Velocity Gradient ................................................ 81
  - 5.6.1 Oort Constants .............................................. 81
  - 5.6.2 Oort Functions .............................................. 84
  - 5.6.3 Divergence of the Velocity Field ......................... 86
- 5.7 Discussion ......................................................... 89
List of Tables

5.1 Optimal hyperparameters of sparse GP analysis for Gaia DR2 data . . . . . 72
5.2 Comparison of literature and measured values of Oort constants . . . . . . . 83
6.1 Optimal hyperparameters of sparse GP analysis for Gaia DR3 data . . . . . 93
6.2 Comparison of Oort constants for DR2 and DR3 with literature . . . . . . . 100
List of Figures

1.1 Artist image of Milky Way ........................................ 2

2.1 The MWPotential2014 Galactic potential model .................. 8

2.2 The Milky Way spiral arms: positions and over-densities .......... 11

3.1 The bi-variate normal distribution .................................. 18

3.2 Identity kernel sample functions ................................... 19

3.3 The radial basis function kernel .................................... 20

3.4 Gaussian process regression toy model ............................ 23

3.5 Samples from various kernel functions ............................. 26

3.6 Comparison between a basic GP and the sparse GP method ........ 32

3.7 Sparse GP model with failure case ................................. 33

4.1 Gaia data distribution over an artist image of the Galaxy .......... 38

4.2 Visualization of sky coordinates .................................... 40

4.3 Example color-magnitude diagram ................................. 42

4.4 2-D histograms of the Schönrich catalog ......................... 50

4.5 2-D histograms of the Gaia DR3 catalog ......................... 51

5.1 Mock data test with Gaussian process regression ................. 70

5.2 Residuals of mock data testing .................................... 71
Abbreviations

DF Distribution function
CBE Collisionless Boltzmann Equation
MVN Multivariate normal
GP Gaussian process(es)
GPR Gaussian process regression
PDF Probability distribution function
RBF Radial basis function
ARD Automatic relevance determination
FITC Fully Independent Training Conditional
DTC Deterministic training conditional
Var-DTC Variational deterministic training conditional
RAM Rapid access memory
SEGUE Sloan Extension for Galactic Understanding and Exploration
RAVE Radial Velocity Experiment
LAMOST Large Sky Area Multi-Object Fiber Spectroscopic Telescope
DR2 Gaia Data Release 2
DR3 Gaia Data Release 3
RA Right ascension
DEC Declination
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMRA</td>
<td>Proper motion in right ascension</td>
</tr>
<tr>
<td>PMDEC</td>
<td>Proper motion in declination</td>
</tr>
<tr>
<td>ICRS</td>
<td>International Celestial Reference Frame</td>
</tr>
<tr>
<td>BP</td>
<td>Blue photometer</td>
</tr>
<tr>
<td>RP</td>
<td>Red photometer</td>
</tr>
<tr>
<td>CMD</td>
<td>Color-magnitude diagram</td>
</tr>
</tbody>
</table>
Symbols

Each quantity is defined explicitly in the context where it is used, with some addition of subscripts beyond those listed here. Some quantities such as $l, \mu, m$, are standard to use in multiple contexts. The context has been made clear for each use. Values are listed below in approximate order of appearance.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>pc</td>
<td>parsec, 1pc $\approx 31 \times 10^{15}$ m</td>
</tr>
<tr>
<td>kpc</td>
<td>kiloparsecs</td>
</tr>
<tr>
<td>Myr</td>
<td>Megayears</td>
</tr>
<tr>
<td>Gyr</td>
<td>Gigayears</td>
</tr>
<tr>
<td>$M_\odot$</td>
<td>Solar masses</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Stellar density</td>
</tr>
<tr>
<td>$x, x$</td>
<td>Position, vector and scalar</td>
</tr>
<tr>
<td>$v, v$</td>
<td>Velocity, vector and scalar</td>
</tr>
<tr>
<td>$t$</td>
<td>Time</td>
</tr>
<tr>
<td>$O$</td>
<td>Set of observables</td>
</tr>
<tr>
<td>$\nabla^2$</td>
<td>Laplacian operator</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>Gravitational potential</td>
</tr>
<tr>
<td>$G$</td>
<td>Gravitational constant</td>
</tr>
<tr>
<td>$\nabla$</td>
<td>Gradient</td>
</tr>
</tbody>
</table>
\( W \) \quad Window function

\( n \) \quad A counted number

\( V \) \quad Mean velocity field

\( A, B, C, K \) \quad Oort’s constants

\( \Omega \) \quad Angular velocity

\( N \) \quad Normal distribution

\( \mu \) \quad Mean vector

\( \Sigma \) \quad Covariance matrix

\( \sigma^2 \) \quad Variance

\( l \) \quad Lengthscale, in context of GP

\( y, y \) \quad Observations/targets, as vector and scalar

\( \epsilon_i \) \quad Gaussian noise

\( m(x) \) \quad Mean function

\( k(x, x') \) \quad Covariance function

\( P \) \quad Probability distribution function

\( X, K \) \quad Input matrix, kernel matrix

\( I \) \quad Identity matrix

\( \theta \) \quad Hyperparameter vector

\( N, M \) \quad Numbers of data and inducing points

\( Q \) \quad Nystrom approximation matrix

\( d, p \) \quad Distance, parallax

\( M, m \) \quad Absolute and apparent magnitude

\( T, A, B, C \) \quad Coordinate transformation matrices

\( \alpha, \delta, l, b \) \quad Coordinate angles
\( \mu_a, \mu_\delta, \mu_l, \mu_b \) Proper motions in angles
\( v_{\text{los}} \) Line of sight velocity (radial relative to the Sun)
\( X, Y, Z, R, \phi \) Spatial coordinates in Cartesian and polar forms
\( V_i, U, V, W \) Velocity components in stated form
\( \mathbf{H} \) 2 \( \times \) 2 matrix of velocity derivatives
\( M^z \) Symmetric and anti-symmetric tensors of the O-M model
Chapter 1

Introduction

The study of galaxies reveals a variety of evolutionary paths and outcomes, ranging from dwarf galaxies, ellipticals, spheriodals, and spiral arm galaxies such as our own Milky Way. Galaxies are usually though of in term of several components: a central bulge, stellar disk, gas disk, and a dark matter halo. Within the disk, the dominant motion is the rotation around the center of the Galaxy, moving about 200 km s$^{-1}$ at the position of the Sun. Stars follow generally circular orbits with velocity changing as a function of radius and distance from the plane of the Galaxy, governed by the gravitational potential. Figure 1.1 shows an artist impression of the Galaxy, presenting the spiral arms, the Sun’s position about 8 kpc from the center, and the dominant clockwise rotation of the disk. Within the dominant rotation, there is a complex set of features and departures from equilibrium driven by internal and external sources as the Galaxy evolves. In this thesis we aim to model the mean velocity field in the vicinity of the Sun.

The Gaia space telescope has the goal of capturing the positions and sky motions of more than one billion stars in the Milky Way (Gaia Collaboration et al., 2016). Accompanying this data is the goal for more than 100 million recorded radial velocities, with ~7 million and ~34 million provided in Data Release 2 (DR2) and Data Release 3 (DR3) (Gaia Collaboration et al., 2018a, 2022a). Now more than halfway through the anticipated mission lifetime, these data releases have allowed great steps to be taken in understanding the state of the
Figure 1.1: Artist depiction of the Milky Way, viewed from above [NASA 2008]. The Sun is positioned about 8 kpc from the center of the Galaxy within the Local Arm (called the Orion Spur in this image). Stars rotate clockwise around the center of the Galaxy.
Galaxy, including identifying evidence of past mergers such as Gaia-Enceladus, and phase space spirals in Galactic disk stars (Belokurov et al., 2018; Helmi et al., 2018; Antoja et al., 2018; Darling & Widrow, 2019). However, a clear picture of the bulk motions of stars near the Sun has yet to emerge. Simple statistics over the data can provide kinematic maps such as those in Gaia Collaboration et al. (2018b), but lack the connection to theory. This work aims to provide that stronger connection by modelling the local stellar velocity field.

The question then becomes how to perform this modelling. Dynamical models of a galaxy describe the orbits of the stars along with the distribution of mass, and thus the gravitational potential via Poisson’s equation. The simplest models assume axisymmetry and time-independence. These simple models are constrained by the rotation curve, which gives the mean azimuthal velocity about the center as a function of radius. More complex models drop the assumption of axisymmetry, and may fit the additional components of the azimuthal and vertical velocities or consider further parameters for describing bars, spiral structure, or the warp (e.g. Spekkens & Sellwood, 2007). Difficulty can arise in selecting the form of equations to fit to the data as variations in the disk can take a shape not easily fit by a set of equations. To sidestep the limitations of a set form and group of parameters, we look to a non-parametric method of modelling.

Considering the connection to theory, as early as 1927, Oort’s constants considered the derivatives of the velocity field at the Sun to give the azimuthal shear and vorticity as a constraint on the potential (Oort, 1927). Such connections between the derivatives of the velocity field and the potential have been expanded by Ogorodnikov and Milne to further explore the disequilibrium in the disk (Ogorodnikoff, 1932; Milne, 1935; Ogorodnikov, 1965). The model we apply must be continuous and differentiable to provide such a connection to the potential and time evolution of the Galaxy.
We apply Gaussian process regression (GPR), a non-parametric modelling method that is continuous and differentiable, allowing the model to provide a strong connection between astrometric data and dynamics. Using machine learning techniques, the model is optimized to capture the underlying bulk Galactic motion.

This thesis consists of seven chapters. We begin by discussing galactic dynamics and the motions within the Galaxy from perturbations in Chapter 2. In Chapter 3, we explore GPR and its mathematical basis, as well as modifications to the method for use with our large dataset. Next, Chapter 4 discusses the treatment of observational data for use in the model. In Chapter 5, we present the published paper containing first results of the model. We extend the model to improved data with Gaia Data Release 3 in Chapter 6, and finally, Chapter 7 concludes and offers further extensions to the work.
Chapter 2

Galactic Dynamics

The Milky Way Galaxy is a gravitationally bound system with $\sim 10^{11}$ stars and an age of about 10 Gyr. The Galaxy is fairly typical of a spiral system and may be broken up into its major structures: a central massive black hole, a central bulge (including a boxy/peanut-shaped bar), the disk, spiral arms, and an extended dark matter halo. The Milky Way’s supermassive black hole is Sagittarius A*, named for its position in the constellation, and with a mass of $4.16 \pm 0.06 \times 10^6 M_\odot$ ([GRAVITY Collaboration et al., 2019](#)). Surrounding the black hole is the stellar bulge and bar with axis ratios $\sim 3:1:1$ containing much of the bulge’s mass. Beyond the bulge, stars and gas orbit the Galaxy in a flat disk extending for more than 15 kpc. Superimposed on the disk are spiral arms that wrap the Galaxy. The overall pattern is that of a density wave with individual stars passing in and out of the spiral arms, giving the wave a different speed than the individual stars themselves. The last component is the dark matter halo. Thought to have a radius on the order of 180 kpc and a mass around $10^{12} M_\odot$, it is composed of some weakly interacting particle(s) that have not been directly detected ([Battaglia et al., 2005](#)). While the nature of such particle is unknown, its influence on the Galaxy can be seen by understanding the motion of visible matter, and additional mass that must be present.
2.1 The Stellar Distribution Function

In this section we describe the standard approach to modelling the stellar component of disk galaxies (Binney & Tremaine, 2008). First we consider the nature of the stellar component as collision-less. A typical galaxy has between $10^6$ and $10^{12}$ stars with the distance between typical stars on the order of lightyears. One may calculate the 2-body relaxation time for perturbations in the Galaxy as

$$t_r \sim \frac{N}{10 \ln N} t_{dyn}$$

where $t_{dyn}$ is the dynamical time of the Galaxy, $\sim 100 – 200$ Myr. This produces a relaxation time greater than the age of the universe, and the effect of random interactions (collisions) is minimal in the Galaxy. This may also be considered as the chance of collisions between stars being small, and so gravity acts as a long range force. A simple thought experiment from Binney & Tremaine (2008) weighs the influence of near stars versus far stars by considering the force from a cone of stars. The force from any one star falls of as $r^{-2}$. Given a uniform density for the region, the number of stars per unit length increases as $r^2$. Each doubling in radius has a length proportional to $r$, so each region attracts with a force $r^{-2} \times r^2 \times r = r$. The force on a star is then dominated by the mean gravitational field of distant stars in the Galaxy. This collisionless nature of the disk is supported observationally by substructure in the velocity distribution of the Galaxy (Dehnen, 1998; De Simone et al., 2004).

We then consider the stellar component to be described by a probability distribution function (DF), $f(x, v, t)$, specifying the stellar density in a 6-D phase space of position $x$ and velocity $v$ at a particular time $t$. This function $f$ obeys the Collisionless Boltzmann Equation (CBE)

$$\frac{\partial f}{\partial t} + v \cdot \frac{\partial f}{\partial x} + \dot{v} \cdot \frac{\partial f}{\partial v} = 0.$$
2.1. THE STELLAR DISTRIBUTION FUNCTION

The DF is coupled to the potential of the Galaxy through Poisson’s equation. Poisson’s equation is given by

\[ \nabla^2 \Phi(x, t) = 4\pi G \rho(x, t), \]  

(2.3)

where \( \Phi \) is the gravitational potential, \( G \) is the gravitational constant and \( \rho(x,t) \) is the mass density. The mass density is found from the stellar DF as an integral over velocity times some mass,

\[ \rho(x, t) = \int d^3v' f(x', v', t) m. \]  

(2.4)

The goal of interpreting the distribution function involves considering a set of observables designed to isolate features of the DF. An observation of the Galaxy is able to find the phase-space positions, velocities, and a few quantities such as magnitude and colour that could serve as an estimate of stellar mass. From this set, some example observables are the stellar number density, the mean velocity, velocity variance, and an estimate to the mass density based on calculations that could be done from colors and magnitudes. These observables may be viewed as a construction from the DF

\[ O = \int d^3x' d^3v' f(x', v', t) F(x', v') \]  

(2.5)

where \( F \) is a function to pick out a specific observable. This work models the first velocity moment of the DF, the mean velocity field. Constructing this observable, we can consider a window function selecting a volume of stars around a point \( W(x - x') \), and a number of stars within the volume \( n \). The mean velocity is then expressed as

\[ V(x) = n^{-1} \int d^3x' d^3v' f(x', v', t) W(x - x') v'. \]  

(2.6)
where lowercase \( v \) is the velocity vector of an individual star, and \( V \) is the mean velocity field as a function of position. To understand and analyze the observed mean velocity field \( V(x) \) we first need to understand the variety of motion that may exist within the Milky Way Galaxy.

## 2.2 Velocity Field in Equilibrium

To begin understanding the motion of the Galaxy we can first consider a simple dynamical model, and then explore departures from this model. A simple model considers a galaxy that is in equilibrium, is axisymmetric, and symmetric about the midplane. In this axisymmetric equilibrium model of the Galaxy, stars move on circular orbits around the Galactic center. Individual stars will possess components in the azimuthal and vertical direction, but no radial motion. The stellar velocities depend on the potential of the Galaxy, which has gradients in radius \( R \) and distance from the disk midplane \( |z| \).

![Figure 2.1: The MWPotential2014 Galactic potential model. \( R_o \) is equal to 8 kpc, with a circular speed of 220 km s\(^{-1}\). Left: Rotation curve produced by galpy (black). Components in dashed lines use the power spherical potential for the bulge (red), Miyamoto-Nagai model for the disk (blue), and the NFW potential for the dark matter halo (green) (Miyamoto & Nagai 1975; Navarro et al. 1996). Right: isopotential curves for the MWPotential2014 model as a function of radius and height from the disk (Bovy 2015).]
2.3. DEPARTURES FROM EQUILIBRIUM

The dominant motion in the disk is the rotation, which varies as a function of \( R \) and \(|z|\) as the potential changes with distance from the center. Modelling the stellar azimuthal speed as a function of \( R \) gives the rotation curve, which may be obtained directly from the potential in simple models by \( v^2(r) = r \frac{\partial \Phi}{\partial r} \). Figure 2.1 presents the MWPotential2014 model produced by Bovy (2015) for GALPY. The Milky Way model includes components for the bulge, disk, and halo in its calculation. The left panel of Figure 2.1 presents the rotation curves of the separate components, and the full model consisting of the components summed in quadrature. The curve is dominated in the region of the Sun (at \( \sim 8 \) kpc) by the disk’s potential, but at increased radius the halo contributes to produce a flatter rotation curve. The right panel of Figure 2.1 shows iso-potential curves for the model. From the gradients in \(|z|\) we can see stars further from the midplane lie in a weaker potential and will experience lower circular velocities as a result of the weakened attraction. The stars will also oscillate in \( z \) as a result of the gradient towards the midplane. While individual stars may have vertical velocities, the bulk motion in an equilibrium model will remain \( V_z = 0 \).

Further complex orbits are studied with Chapter 3 of Binney & Tremaine (2008), but are not discussed further.

2.3 Departures from Equilibrium

We then consider stellar motion that is a departure from this equilibrium field. Observations of the Milky Way through Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al. 2009), the Radial Velocity Experiment (RAVE; Steinmetz et al. 2006), and the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Cui et al. 2012) have shown extensive evidence for bulk motions that signal a Galaxy in disequilibrium. These departures may come from a variety of sources within and beyond
the Milky Way.

2.3.1 The Spiral Arms

The most prominent feature when visualizing a galaxy or observing an artist sketch is a set of spiral arms circling the galaxy. Spiral arms sweep out from the centre of the galaxy in unbarred galaxies, or from the end of the bar in barred galaxies. Spiral arms take the form of density waves that rotate with a radius-independent angular speed through the disk with differentially rotating stars. Stars at the inner section of the arms may move faster than the arms itself, while towards the edge of the galaxy stars lag behind the arms.

The formation and life of spiral arms is a matter of debate. Four basic models of spiral formation and evolution have been suggested. Spirals may be tidally triggered by interaction with companion galaxies, as proposed for the spiral nebula of the Whirlpool Galaxy (or M51) \cite{Kormendy1979,Struck2011,Dobbs2010,Watkins2015}. They may be produced by swing amplification, in which the stellar shear and self gravity change the angle of the arms, transiently amplifying small scale waves into the bulk spiral arms \cite{Fall1981,Sellwood2012,Baba2013}. This mechanism may produce large scale arms disks with low velocity dispersion, but only for short periods before winding washes out the arms. Spiral arms may be driven by bars, extending the gravitational influence beyond the bulge and into the disk \cite{Kormendy1979,Athanassoula1980}. Finally, the classical view is a fixed spiral wave travelling through the disk that will not experience winding or transient behaviour. The number of arms and their duration are shown to be partially dependant on the amount of mass in the disk versus the bulge, with heavier disks tending towards two large long lasting spiral arms over a number of smaller arm segments that are transient \cite{Hart2016,Ann2014}. The number
of arms may also be triggered by the method of formation with gravitational instabilities tending to form multiple arms while interactions are likely to form two arms (Dobbs & Baba, 2014; Hart et al., 2016). Figure 2.2 presents the position of the Galactic spiral arms from the Reid et al. (2019) models (Left), and stellar over-density from Poggio et al. (2021b) (Right). The Milky Way presents four proposed arms that circle the Galaxy, as well as some arm segments such as the local arm in blue in Figure 2.2 (Left).

Figure 2.2: Left: Spiral arms model of Reid et al. (2019), with solid lines tracing the centers and dotted lines tracing the width of 90% of fitting sources. Arms are in order: Scutum (cyan), Sagittarius (magenta), Local Arm (blue), Perseus (black). Right: Stellar over-density with Gaia EDR3 from Poggio et al. (2021a), recreated from available data and code. Stellar overdensity is calculated as the local density over mean density minus 1, applying kernel density estimation with bandwidth 0.3 kpc (local) and 2 kpc (mean). The gray dot represents the position of the Sun.

Motions associated with spiral arms are related to the nature of a density wave and the self gravity. N-body+hydrodynamics simulations by Kawata et al. (2014) and Grand et al. (2014) of spiral arms explore the theoretical motions that may be analyzed. Stars on the leading edge of a spiral arm tend to have higher azimuthal velocities, and interaction with the arms will cause loss of angular momentum. As the arm passes through the section of
stars, this angular momentum will be passed to stars on the trailing edge that are moving at lower azimuthal velocities. The angular momentum change is due to the tangential force of the spiral arm, leading to a radial migration towards the edge of the Galaxy as stars pass through the spiral arm. The difference in angular velocity may be considered as stars in the leading edge being at peri-centre while the trailing side stars are at apo-centre. Stars above and below the plane experience a ‘breathing’ mode of bulk motions either towards or away from the Galactic plane (vertical compression or expansion) (Debattista, 2014). Similar to the in-plane motion, stars on the leading side experience compression, and stars on the trailing side experience expansion.

Theoretical motions from simulation may not be definite for the Milky Way depending on the state of the Galaxy. The possibility that the Milky Way arms are transient has seen some support based on long lived spiral galaxy simulations and spiral arm segments that exist in the Milky Way such as the Local Arm. Baba et al. (2018) present evidence of a disruption phase in the Perseus arm using Gaia DR1 Cepheids. The proposed motion of the Perseus Arm does not agree with spiral arms models, questioning the stability of the arm. Observed motions highlight the need for further research into the motion of galaxies.

2.3.2 The Bar

Like many large spiral galaxies, the Milky Way’s bulge contains a bar. The bar is a density mode with approximate axis proportions 3:1:1 with a diameter of a few kiloparsecs. While the bar has many effects on the movement of gas on and star formation, the focus here is on the effects on the potential and bulk motion of the disk. Models of barred galaxies find potential curves altered into ellipses by the non-axisymmetric mass distribution. Individual star orbits are affected greatly in the bar, however the effect on the outer ranges of the
2.3 DEPARTURES FROM EQUILIBRIUM

Spiral arms are less definite. Simulation results point to a multitude of motion that may be attributed to the bar. Monari et al. (2014) found that the bar amplified vertical breathing modes in the disk, and shifted their position relative to bar-less galaxies. Chakrabarty (2007) attributed phase-space clumping of stars to resonances with a bar and spiral arms. Monari et al. (2016) observed radial motions in the disk to be a superposition of isolated bar and spiral effects, suggesting they may be able to be disentangle the effects when viewed away from resonances.

2.3.3 Random Motions

The previous assumption of collisionless movement for stars in the disk is largely accurate for considering the bulk motions of the Galaxy. However, stars still have a chance for close encounters that alter their trajectory significantly more than the mean potential field. These close encounters have two relevant effects. First, interactions will allow stars to gain significant changes in velocity with measurable changes in the bulk dispersion of the stars. This is the age-velocity dispersion relation and has a few sources including random interactions, and perturbations from other sources such as the spiral arms (Yu & Liu, 2018). The dispersion in the vertical direction is considered as disk heating, and affects the thickness of the disk. The other effect is the rise of patterns and perturbation from random motion. Galactic simulations by Chequers & Widrow (2017) found bending waves arise without provocation from the random distribution noise. These simulated waves are long lived, with power migrating to the edge of the disk and amplitudes in agreement with the stellar warp.
2.3.4 External Sources

External perturbations can cause gravitational disturbances in the disk, especially the outer disk which has a lower surface density and is more fragile to perturbation. Suggested external disturbances include interactions with satellites such as the Magellanic Clouds, infall of material, and a non-uniform or off centre dark matter halo (Gaia Collaboration et al., 2018b). External perturbers are hard to distinguish using observational data, however varying simulation parameters allows an understanding of the effects these interactions may have. Widrow et al. (2014) analyzed on interaction between a satellite and a galactic disk, finding that a ‘slow’ approach will primarily cause a bending mode in the disk, while fast satellites may excite breathing and higher order modes. Angled interactions and satellite accretions have been noted to cause streaming motions in the disk detectable in phase space. Perhaps the most significant perturbation is the warp in the galactic disk. The suggestion of satellite interaction causing the apparent bend in the Galactic disk has been investigated by many (Weinberg & Blitz, 2006; Bailin, 2003, for example). The nature of the warp as long-lived, transient, or repeatedly excited have each been explored through simulation and expected signals produced by the methods of excitation, with the goal that improved observation will provide further evidence in the Galaxy (Poggio et al., 2017).

2.4 Galactic Parameters

Quantifying Galactic motions has been approached using a variety of methods. In order to parameterize the disk, the motion must be understood. A first quantity to consider is the rotation curve $V(r)$. Mentioned previously, the rotation curve considers the azimuthal velocity as a function of radius. While useful for stars, the spiral arms are proposed to travel as density pattern waves. Instead we may also consider the angular velocity $\Omega(r)$ and
2.4. GALACTIC PARAMETERS

search for patterns and signals of the spiral arms.

To understand the complex motions in our Galaxy, we consider not just the snapshot we are able to observe, but how velocities will change across the plane and their connection to the Galactic potential. In 1927, Oort laid the theoretical basis for considering the connection of stars streaming along orbits and the Galactic potential (Oort, 1927). The difference in velocity between two points in the Galaxy may be Taylor expanded to consider the derivatives of the velocities. Oort considered a flat disk with motion in two dimensions - radially and azimuthally. Sums of the derivatives evaluated at the position of the Sun are considered Oort’s constants: A, B, C, and K, measuring the azimuthal shear (A), vorticity (B), radial shear (C), and local divergence (K) of the velocity field. In an axisymmetric model parameters may provide direct constraints on the Galactic potential, for example a harmonic potential produces solid-body rotation with $A = 0$, and $B = \Omega$, a flat rotation curve gives $A = -B$, and a potential with all mass concentrated in the inner Galaxy yields $A = -3B$.

Derivatives of the velocity field may be further expanded to calculate the 3-D, nine component derivative matrix, as considered by the Ogorodnikov-Milne model (Ogorodnikoff, 1932; Milne 1935). Further representations give access to considering the 3-D divergence of the velocity field, considering the vorticity or angular momentum in other planes, and a more accurate representation of the Galactic potential. The derivation and discussion of the Oort constants and velocity field derivatives continues in Section 5.6 accompanying model results.
Chapter 3

Gaussian Process Regression

The Gaussian processes (GP) model is a probabilistic supervised machine learning framework that may be applied to regression and classification problems. A Gaussian process regression (GPR) model incorporates prior knowledge (kernels) to make predictions and provide uncertainty measures over the predictions. Wiener and Kolmogorov first used GPR in the 1940’s for the prediction of time series data (Wiener & Masani, 1976; Brovelli et al., 2003). This method gained popularity in geostatistics in the 1970’s with Krige’s work on gold ore deposits, giving GPR the name kriging in the field (Rasmussen & Williams, 2006). GPR is viewed today as a non-parametric machine learning technique, incorporating supervised iterative learning made more available with increased computing power.

In this chapter, we introduce the theory of Gaussian processes and its application to the problem of regression. This overview covers the methods of GP inference, kernel matrices, parameter optimization, and the sparse GP approximation applied in later chapters. For a more extensive view on the topic, see Rasmussen & Williams (2006).

3.1 Background

A GP is defined as a stochastic process (a collection of random variables indexed by time and space) such that every finite collection of those random variables has a multivariate normal distribution. The distribution of a GP is the joint distribution of the infinitely many random
variables, and as such can be used to represent arbitrary continuous functions. This section explores this definition beginning with a multivariate normal (MVN), the mathematical basis of GP continuing in Section 3.2.

The multivariate normal distribution describes a system that is controlled by more than one variable, a vector \( \mathbf{x} = (x_1, x_2, \ldots, x_D) \) that may have have correlated components. The probability distribution function (PDF) of an MVN with a number of dimensions \( D \) is given by

\[
N(\mathbf{x} | \mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} ((\mathbf{x} - \mu)^T \Sigma^{-1} (\mathbf{x} - \mu)) \right]
\]  

(3.1)

where \( \mathbf{x} \) is our vector of variables, \( \mu \) is the mean vector, and \( \Sigma = \text{cov(\mathbf{x})} \) is the \( D \times D \) covariance matrix. The covariance matrix is a symmetric matrix that stores the variance of each variable along the diagonal, and the pairwise covariance between random variables as \( \Sigma_{ij} = \text{cov}(x_i, x_j) \).

To visualize the MVN we can use a bi-variate normal

\[
\begin{bmatrix}
  x_1 \\
  x_2
\end{bmatrix} \sim N\left( \begin{bmatrix}
  \mu_1 \\
  \mu_2
\end{bmatrix}, \begin{bmatrix}
  \sigma_{11} & \sigma_{12} \\
  \sigma_{21} & \sigma_{22}
\end{bmatrix} \right) \sim N(\mu, \Sigma)
\]  

(3.2)

and apply values to the mean vector and the covariance matrix. Figure 3.1 shows an example of a bi-variate normal with the mean values \( \mu_1 = 0, \mu_2 = 1 \) and the covariance values \( \sigma_{11} = \sigma_{22} = 1, \sigma_{12} = \sigma_{21} = 0.6 \).

To transition from the view of a MVN to that of a Gaussian process we can consider samples from an MVN with a higher number of variables. In the left panel of Figure 3.2 we consider samples from the two parts of a bi-variate normal. The sets of 5 connected points between \( Y = 0 \) and \( Y = 1 \) can be seen as the values of \( x_1 \) and \( x_2 \) in a distribution like Figure 3.1. Extending to 20 dimensions with no covariance we find the right panel of Figure 3.2.
3.1. BACKGROUND

Figure 3.1: The PDF of a bi-variate normal from Equation 3.2 with values \( \sigma_{11} = \sigma_{22} = 1 \), and \( \sigma_{12} = \sigma_{21} = 0.6 \). The 3-D bell curve represents the probability density with height, and its 2-D representation below.

The set of sampled points takes the appearance of a discontinuous function. This is because the relation between each point (each dimension of the MVN) has not been constrained by the covariance matrix.

To smooth the functions in the right panel of Figure 3.2 we can define covariance functions which reflect on prior knowledge about the form of the function we are modelling. Covariance functions (also called kernels) will be covered in a more mathematical basis in Section 3.3. The most widely used kernel function is the radial basis function (RBF),
3.1. BACKGROUND

Figure 3.2: Five samples of an identity kernel (no covariance) MVN with different dimensions. Left: two dimensions. Right: twenty dimensions.

defined as

$$\text{cov}(x_i, x_j) = \sigma_f^2 \exp \left( -\frac{(x_i - x_j)^2}{2l^2} \right)$$  \hspace{1cm} (3.3)

where $\sigma_f^2$ is the variance of the function controlling the peak value of the covariance matrix, and $l$ is the lengthscale, controlling how quickly the relation between points reduces with distance.

Previously, we plotted a set of samples with no covariance (also considered the identity kernel). In Figure 3.3 we plot a set of sample functions using the RBF kernel. The left panel shows the shape of the kernel used. We see that points close together exhibit a stronger covariance than those far apart. This translates into the right panel with slowly varying functions. Samples from the general kernel are called kernelized prior functions, as they have no observed data points yet.

It is natural to then consider increasing the dimension of the MVN. When the dimension of the MVN becomes infinity, it is able to represent a continuous function. In using an MVN with infinite dimensions, we apply infinite parameters to the regression task. In this case,
3.2. MATHEMATICAL BASIS OF GAUSSIAN PROCESSES

In Gaussian processes, we consider the values $y_i$ as observations of the function $f(x)$ at input locations $x_i$. These observations contain some Gaussian noise $\epsilon_i$ that is independent and identically distributed given by

$$y_i = f(x_i) + \epsilon_i \quad \quad \epsilon_i \sim N(0, \sigma_n^2).$$ (3.4)
The inputs are taken as vectors of the input dimension $x_i \in \mathbb{R}^{D \times 1}$ and the outputs as stochastic scalars $y_i \in \mathbb{R}$. The assumption placed on the target function $f(x_i)$ is that any finite number of function values will have a multivariate Gaussian distribution, and as such the function is a GP. A GP is fully defined by a mean function $m(x)$ and a covariance function $k(x, x')$ such that

$$f(x) = \mathcal{N}(m(x), k(x, x')).$$

(3.5)

A Bayesian GP model assumes that the noise-free regression function $f$ comes from a GP that has a prior zero mean function, and a specified covariance function. A zero mean states that the function is equally likely to be positive or negative over the domain, not that the function will have a mean of zero over the domain. This does have the result of pulling the regression towards zero where no data points exist.

The probability distribution function of the noiseless data $f = [f(x_1), ..., f(x_N)]$ is given by

$$P(f|X) = \mathcal{N}(f|\mu, K)$$

(3.6)

where $X = [x_1, ..., x_N] \in \mathbb{R}^{N \times D}$, $\mu = [m(x_1), ..., m(x_N)]$, and $K = k(X, X)$. The aim is then to find a relationship between the input points $X$ and corresponding data with error $y = [y_1, y_2, ..., y_N]^T \in \mathbb{R}^{N \times 1}$, and make predictions for a new set of function outputs $f_*$ with corresponding points $X_*$. The function values are given the assumption that they have a probabilistic joint Gaussian distribution

$$\begin{bmatrix} f \\ f_* \end{bmatrix} \sim \mathcal{N} \left( \begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right)$$

(3.7)

where $f$ is the noiseless data vector, $K = k(X, X)$, $K_* = k(X, X_*)$, and $K_{**} = k(X_*, X_*)$. In reality we consider our noisy observations $y$ with variance $\sigma_n^2$, giving us the prior
cov(y) = K + \sigma_n^2 I. The joint distribution between the observed values and the values at new points becomes

\[
\begin{bmatrix}
  y \\
  f_*
\end{bmatrix} \sim \mathcal{N}
\begin{pmatrix}
  m(X) \\
  m(X_*)
\end{pmatrix},
\begin{bmatrix}
  K + \sigma_n^2 I & K_* \\
  K_+^T & K_{++}
\end{bmatrix}.
\] (3.8)

The joint probability is denoted \( P(y, f_* | X, X_*) \). Regression requires the conditional distribution \( P(f_* | y, X, X_*) \). The derivation applies the marginal and conditional distributions of the MVN theorem in Appendix A.1 (Rasmussen & Williams, 2006). By marginalizing, the posterior distribution is then

\[
P(f_* | y, X, X_*) \sim \mathcal{N}(\mu_*, \Sigma_*)
\] (3.9)

where

\[
\mu_* = m_* + K_+^T (K + \sigma_n^2 I)^{-1} (y - m_*)
\] (3.10)

\[
\Sigma_* = K_{++} - K_+^T (K + \sigma_n^2 I)^{-1} K_+.
\] (3.11)

Equation 3.9 can be used to make predictions of the function value for points \( x_* \) giving both the statistical mean and the model’s confidence in the mean through the variance provided by \( \Sigma_* \). This is the key equation for prediction from the model. In the covariance function, it may be noted that values depend only on inputs \( X, X_* \), and the noise \( \sigma_n^2 \), not the values of \( y \). This is a property provided by the Gaussian distribution (Rasmussen & Williams, 2006).

Figure 3.4 presents an example of a basic GP using the RBF kernel. Previously, in the right panel of Figure 3.3, we saw samples from the kernel that were a prior, showing no constraining data points. The left panel of Figure 3.4 shows the posterior distribution, where the function passes through seven noise-free observations. The 2\( \sigma \) confidence region shows
Figure 3.4: An example of a basic GPR, applying the presented equations. The observed data (black points) are generated from the true function of the red dashed line. The blue dashed line shows the calculated GP mean. 20 samples are drawn from the distribution, and plotted as thin solid lines with the $2\sigma$ confidence region as the light blue shading. The RBF kernel is used with $\sigma = 2$ and $l = 0.2$. The left panel plots a GP with no noise ($\sigma^2_n$) added to the points. The right panel adds a noise variance of 0.1 to the generated points.

most of the sample functions remain close to the mean. The positions of data along the axis shows the effect of data density on the predicted variance. Points close together heavily restrict possible samples that may be drawn from the posterior, meaning the variance also gives information of the data density in that region. This can be understood from Equation 3.11 which is a function of only the location of the input $X$. The right panel shows the effect of added noise creating a larger variance as the confidence in the datapoints is reduced. This allows a larger set of possible sample functions but more closely depicts the effect of real data with observational noise.

The difference between a parametric form of regression and a non-parametric form like GP is made evident by the predictive Equations 3.10 and 3.11. We see the predicted values are a direct function of the datapoints, requiring the dataset for each prediction to occur. Conversely, both simple and complex parametric models such as neural networks only requires data to learn the model parameters, after which it may be discarded. Two
factors are especially important for an accurate GP prediction: the input data which will be discussed in Chapter 4 and the covariance function that must be specified by the user.

3.3 Covariance Functions

The covariance function is perhaps the most crucial part of the GP model. The covariance function defines the statistical relationship between two points in the input space, encapsulating any assumptions about the target function. A valid kernel function \( k \) is required to be positive semi-definite, such that any subset of points will construct a positive semi-definite matrix \( K(X, X) \). A symmetric matrix is positive semi-definite if and only if all of its eigenvalues are non-negative. Many programming implementations of GP additionally require that the kernel function be positive definite eliminating zero as a possible eigenvalue.

3.3.1 Radial Basis Function

The most common choice of kernel is the RBF kernel. By definition, a radial basis function is any function that is only a function of \( r = |x_i - x_j| \). Practically, when discussed as the RBF function or the RBF kernel, the term refers to the squared exponential or Gaussian kernel we introduced in Equation 3.3, which we will now call \( k_{\text{rbf}}(x_i, x_j) \).

The RBF kernel measures the distance \( r \) between datapoints, assigning a high correlation if points are spatially close, and a low correlation if they are further apart. The RBF kernel included two hyperparameters, \( \sigma_f^2 \) giving the variance from the mean values, and the lengthscale \( l \) controlling how quickly correlation reduces with distance. Hyperparameters are often captured jointly in a vector \( \theta = [\sigma_f, \sigma_n, l, ...] \), and changing the value will change the behaviour of the model. Tuning the values of hyperparameters will be discussed in Section 3.4. Figure 3.5(a) shows a set of RBF kernel samples with varying lengthscales. The
positively offset functions present shorter length scales and the negatively offset functions present longer length scales.

The RBF kernel is infinitely differentiable, and will produce functions that are also infinitely differentiable. This fact may be applied as well, given that performing predictions with the derivative of the kernel will return the derivative of the function at the queried point. In practice, applying finite differencing and other simple derivative methods on the queried mean points can give equivalent results.

### 3.3.2 Linear Function

The linear kernel takes the form

$$k_{\text{lin}}(x_i, x_j) = \sigma_b^2 + \sigma_v^2 (x_i - c)(x_j - c)$$

(3.12)

where $\sigma_b$ controls the offset of the line, $\sigma_v$ controls the slope of the line, and $c$ determines a point the posterior will pass through. Using a linear kernel alone is equivalent to performing Bayesian linear regression, for which there are many other tools more easily suited. The linear kernel is also a non-stationary kernel, in that it depends on the absolute position of the points, not the relative different between them. In an implementation of the linear kernel, some programs will separate the two terms in Equation 3.12, considering only the second term to be the linear kernel, and calling the first term ($\sigma_b^2$) a bias kernel that can separately control the offset of the function. These kernels can be combined together again as we will discuss in Section 3.3.4. Figure 3.5(b) presents samples from a linear kernel.
3.3.3 Periodic Function

The periodic kernel function takes the form

\[ k_{\text{per}}(x_i, x_j) = \sigma^2 \exp \left( -\frac{2\sin^2(\pi |x_i - x_j|/p)}{l^2} \right) \] (3.13)

where the period \( p \) determines the distance between repetitions of the function, and \( l \) is the lengthscale. Intuitively, the periodic kernel fits periodic functions, and is often applied to time series data. This data may not be explicitly periodic, exhibiting trends around the periodic variation. We now turn to considering complex kernel functions.

![Figure 3.5: Samples from various kernel functions. Panels (a), (b), and (c) present individual kernel functions. Panels (d), (e), and (f) present functions of kernels. Extension of Roelants (2018).](image)
3.3. COVARIANCE FUNCTIONS

3.3.4 Functions of Kernels

Basic operations over kernel functions will always result in another valid kernel function. The kernel function must yield a symmetric positive semi-definite kernel matrix for any subset of inputs, and any operation that preserves this property may be applied to create a new kernel. The most common operations are:

• sum: \( k(x_i, x_j) = k_1(x_i, x_j) + k_2(x_i, x_j) \)
• product: \( k(x_i, x_j) = k_1(x_i, x_j)k_2(x_i, x_j) \)
• scaling: \( k(x_i, x_j) = ak_1(x_i, x_j) \)
• function multiplication: \( k(x_i, x_j) = f(x_i)k_1(x_i, x_j)f(x_j) \)
• function composition: \( k(x_i, x_j) = k_1(f(x_i), f(x_j)) \)

Combinations of functions may be applied to impose prior knowledge of a function shape. Figure 3.5 (d), (e), and (f) present samples from functions of kernels. The samples show how the space of possible functions changes with alteration of the kernel function.

An important result that comes from performing functions of kernels is the concept of Automatic Relevance Determination (ARD). In this concept, the kernel does not depend on the absolute distance, instead, the distance in each input variable is associated with the hyperparameters of a separable piece of the kernel. In this case, each input dimension will have a different lengthscale, according to the variations seen in that dimension. Use of ARD is crucial to creating accurate GP models with a high input dimension.
3.4 Hyperparameter Optimization

The functional form of the kernel captures the prior knowledge of the function we wish to model. The kernel functions depend on the values of hyperparameters and are the variables of the GPR model that are fit to the data. This section discusses the application of machine learning methods to Gaussian processing for the purpose of optimizing hyperparameters.

Now that we have introduced kernel functions and hyperparameters, we want to consider the posterior probability function $P(f|X, y, \theta, k_i)$. Here, $f$ is a vector of the possible functions, $\theta$ is our vector of hyperparameters, and $k_i$ is selected possible kernel function. The term we are looking for when considering hyperparameter optimization is $P(\theta|X, y, k_i)$, which is the probability that the set of hyperparameters explains the relationship between the data and the kernel function. To get here we first consider Bayes’ rule for inference. Using our posterior, and considering the data, we come to the form

$$P(f|X, y, \theta, k_i) = \frac{P(y|X, f, \theta)P(f|\theta, k_i)}{P(y|X, \theta, k_i)}. \tag{3.14}$$

In this equation, $P(y|X, f, \theta)$ is the likelihood function, or the probability that the observed data is explained by the function, the input points, and the hyperparameters; $P(f|\theta, k_i)$ is the prior, encoding our beliefs; $P(y|X, \theta, k_i)$ is the marginal likelihood. In Bayesian statistics, the marginal likelihood represents the probability of generating the set of data from the prior, and is often referred to as the evidence.

Continuing with the Bayesian formalism, we look to assign our vector of parameters ($\theta$) a probability distribution. To do so, we also consider a prior on our hyperparameter values given the kernel $P(\theta|k_i)$, and our likelihood is the probability that the output $y$ is explained by the data, our hyperparameters, and the kernel. This is also our previous
3.4. HYPERPARAMETER OPTIMIZATION

marginal likelihood, giving the equation

\[ P(\theta | X, y, k_i) = \frac{P(y|X, \theta, k_i)P(\theta|k_i)}{P(y|X, k_i)}. \]  

(3.15)

The denominator is the normalizing constant

\[ P(y|X, k_i) = \int P(y|X, \theta, k_i)P(\theta|k_i)d\theta. \]  

(3.16)

We now look to find the maximum of the probability distribution \( P(\theta|X, y, k_i) \). Considering Equation 3.15, the equation may be simplified in two ways. First, we know our denominator \( P(y|X, k_i) \) is a constant given a fixed kernel function. Second, we may assume the prior \( P(\theta|k_i) \) is flat, giving no prior knowledge of the best values. This gives us the relation

\[ P(\theta|X, y, k_i) \propto P(y|X, \theta, k_i). \]  

(3.17)

The probability distribution of hyperparameters is then proportional to the marginal likelihood, which can be used to perform numerical optimization. For simplicity, the logarithm is taken resulting in a maximization problem. The function to be maximized is given by

\[ \log(P(y|X, \theta)) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |K + \sigma_n^2 I| - \frac{1}{2} y^T (K + \sigma_n^2 I)^{-1} y. \]  

(3.18)

The first term is a normalizing constant, the second contains the trace of the kernel matrix and is considered a penalty term based on the complexity of the fit, and the third is a fitting term finding better values for a good fit of the data.

Within machine learning, the usual convention is to consider minimization problems and apply a negative where applicable. Following that convention, optimization for the
3.5 HETEROSCEDASTIC NOISE

Hyperparameters is done using gradient descent algorithm. This is done by calculating the partial derivatives of the marginal likelihood, at which point any algorithmic scheme such as the Adam optimizer is applied (Kingma & Ba, 2014). Completing optimization results in the best estimate for the hyperparameters and inference gives the GP posterior distribution.

3.5 Heteroscedastic Noise

The GP models we have considered so far concern the covariance matrix given by \( \text{cov}(X) = K + \sigma_n^2 I \) where \( K = k(X, X) \). The value \( \sigma_n^2 \) is the variance of the input points \( X \). The variance is usually assumed to be a constant value across the dataset, which is not true for many datasets including the intended use here for astrometric data. If the variance (or noise) is not constant, it is called heteroscedastic noise.

There are several ways to implement heteroscedastic noise into a model. The method applied here requires changing our data noise \( \epsilon_i \sim N(0, \sigma_n^2) \). We alter \( \sigma_n^2 \) to instead be an indexed value \( \sigma_i^2 \) (Cawley et al., 2003). The noise on the input parameters is then a provided value to the construction of the kernel matrix such that \( \text{cov}(X) = K + \sigma_i^2 I \). The diagonal of the kernel matrix is no longer a definite ridge as each point will have its own individual variance. Providing the value of the noise in this way removes the need to learn the variance of the residual scatter, reducing the number of hyperparameters by one.

Other methods for implementing heteroscedastic noise consider adding a term to the kernel, or an additional unobserved variable that is added to the method. In this way the noise is learned as a function of \( X \), requiring fitting a GP for both the data and the noise (Wang & Neal, 2012; Le et al., 2005; Kersting et al., 2007). While these methods may predict better error ranges for largely anomalous data, performance is similar for most tests, and minimally changes the mean function.
3.6 Sparse Gaussian Process Regression

The biggest challenge with GP is the computational complexity associated with performing
the inverse of the $N \times N$ kernel matrix $K$. The inverse of the matrix scales as $O(N^3)$
operations, requiring $O(N^2)$ storage. This process becomes unfeasible for $N$ much greater
than $10^4$. There are a number of methods to reduce the computational complexity of GP,
including approximating the function, performing interconnected patches of GP, and sparse
GP. In this work we explore the sparse GP framework with the inducing point method.

3.6.1 Inducing point method

GP approximation through inducing points is a popular method to combat the issue of
computational complexity. The concept behind the approach is to allow the estimation of
the marginal likelihood by using a number of points $M$ called the inducing points such that
$M << N$. This gives a dimensional reduction, and prediction can be performed on the
model through the inducing points. Application of the inducing points allows the reduction
of the computational complexity to $O(NM^2)$ and the storage requirement to $O(NM)$.

This can be conceptually understood as reducing the number of comparisons that must
be made. The basic GP model must compare each point to each other point in order to
construct the covariance matrix. The inducing point method designates a set of points that
are able to capture the shape of the posterior well by weighting the model at significant
regions of the input space. The process of prediction is then comparing each datapoint to
each inducing point, a heavily reduced task.

The important decision is selecting the set of inducing points that may best represent the
model. Simple methods such as taking a random subset of the data, or a selection process
to find the best data subset, are constrained to the discrete sample of the space provided by
3.6. SPARSE GAUSSIAN PROCESS REGRESSION

The preferred method is to initialize the inducing points using a data subset, but allow the inducing point locations to vary. This allows each inducing point to represent a region of data more effectively, allowing for both an improved GP fit and a smaller number of required inducing points. Varying the inducing points requires considering each input dimension of the inducing point as a parameter to be optimized. Our set of optimization parameters becomes \( \theta = [\sigma^2, I_i, ..., I_{j,k}, ..., I_{M,D}] \) where \( I \) is the set of inducing points, \( j \) indexes the item in the set, and \( k \) is the index for the input dimension. The set of optimization parameters grows linearly with \( M \), leading quickly to a limit of small \( M \). When \( M \) grows large and sparse GP approaches a full GP, storage and computational costs for each iteration may still be improved, but optimization may require more iterations to approach the maximum likelihood.

A visual example of the inducing point method may be seen in Figure 3.6. The left panel presents a GP model with \( N = 50 \), and the right panel presents a sparse GP model \( N = 50, M = 12 \). This example GP model achieves a computational saving of a factor of

![Figure 3.6: Comparison between a basic GP and the sparse GP method. The mean is in blue, the 50 datapoints as black marks, and the 2\( \sigma \) confidence region as shaded blue. The left panel presents the basic GP. The right panel presents the sparse GP. The 12 inducing points are plotted along the mean function as red dots.](image-url)
12.5 for each iteration of optimization, and a storage saving of 2.5. The difference between the mean and confidence region is negligible for these models. Note the position of the inducing points along the model. The optimized inducing points tend to lie in the center of a cluster of datapoints, or at maxima and minima, with a distribution similar to that of the data.

The model fit depends on the value selected for $M$. Above a threshold, improvement to the model is minimal with increasing $M$ as the model accuracy approaches that of a full GP. Below this threshold, the model is not able to capture the variation in the data. Using our example model, pushing the sparse approximation further to only 5 inducing points would give a computational saving factor of 36 and storage factor of 6, but fails to capture features. This may be seen in Figure 3.7.

Figure 3.7: Sparse GP model of Figure 3.6 with 5 inducing points. The mean is in blue, the 50 datapoints as black marks, and the $2\sigma$ confidence region as shaded blue. The 5 inducing points are plotted along the mean function as red dots.

The fit of the model may be quantified in a few different ways, with the root mean squared error (RMSE) being a basic method. This is given by

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (\text{Predicted}_i - \text{Actual}_i)^2}{N}}.$$  

(3.19)
3.6. SPARSE GAUSSIAN PROCESS REGRESSION

The GP model and sparse GP with 12 inducing points find similar RMSE values with 0.8023 and 0.8024 respectively. The sparse model with 5 inducing points find a score of 0.8602, showing the inability to fit the model.

3.6.2 Deterministic Training Conditional

Applying the sparse GP with inducing points is derived as applying approximations to the GP prior. The same inference and posterior conditioning process of the basic GP is then followed. There are several different approaches to the algorithm, including Subset of Regressors (SoR) (Smola & Bartlett, 2000), Fully Independant Training Conditional (FITC) (Snelson & Ghahramani, 2006), and Deterministic Training Conditional (DTC) (Seeger et al., 2003). Although approached differently, these methods are shown to be derivatives from a unifying framework (Quinonero-Candela & Rasmussen, 2005). This section covers the Variational Deterministic Training Conditional (Var-DTC) of (Titsias, 2009; Bauer et al., 2016).

In the Var-DTC method we consider our \( M \) inducing variables at some inputs \( X_M \). Using the learned inducing points method we need to optimize the set of hyperparameters \( \theta \) which includes our inducing points. The approximation applied is the projected process approximation, which introduces an approximation of the covariance matrix. For notational purposes we add the subscripts \((m, n)\) to denote the shape of each matrix, for example our kernel matrix is now \( K_{nn} \). We replace our exact kernel matrix with

\[
\text{cov}(X) = Q_{nn} + \text{diag}[K_{nn} - Q_{nn}] + \sigma_i^2 I \quad Q_{ab} = K_{am}K_{mn}^{-1}K_{mb} \tag{3.20}
\]

where \( K_{nn} \) is the kernel matrix of the inducing inputs, \( K_{nm} = K_{mn}^T \) is the cross covariance matrix between the data and inducing points, and \( Q_{ab} \) is called the Nystrom approximation
3.6. SPARSE GAUSSIAN PROCESS REGRESSION

(Rasmussen & Williams, 2006). The approximation is corrected in the diagonal by the term \( \text{diag}[K_{nn} - Q_{nn}] \). With this approximation, the marginal likelihood is given by

\[
\log(p(y|X, \theta)) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |Q_{nn} + \sigma_n^2 I| - \frac{1}{2} y^T (Q_{nn} + \sigma_n^2 I)^{-1} y - \frac{1}{2} \sigma_n^2 \text{tr}(K_{nn} - Q_{nn}).
\]  

(3.21)

The additional final term compared to Equation 3.18 is due to the approximation correction \( \text{diag}[K_{nn} - Q_{nn}] \). The predictive equations are then

\[
\mu_* = Q_{*m} (Q_{nn} + \sigma_n^2 I + \text{diag}[K_{nn} - Q_{nn}])^{-1} y 
\]

(3.22)

\[
\Sigma_* = K_{**} - Q_{*m} (Q_{nn} + \sigma_n^2 I + \text{diag}[K_{nn} - Q_{nn}])^{-1} Q_{*n}.
\]

(3.23)

This gives the computational complexity of \( O(NM^2) \) as the matrix inversion is done on the \( M \times M \) matrix \( K_{nm} \) that appears in the Nystrom approximation. The largest matrix required is the component \( K_{nm} \), giving us the storage complexity \( O(NM) \). This selected method often outperforms other sparse GP derivations, given a careful and good quality optimization algorithm (Bauer et al., 2016).

3.6.3 Implementation

Implementation of GP may be done from derived equations (as is done for Figure 3.4), or make use of a variety of publicly available GP packages including Scipy.GP, GPfit, GPy, and Scikit-learn (Virtanen et al., 2020; MacDonald et al., 2013; GPy, 2012; Pedregosa et al., 2011). (Erickson et al., 2017) provides a thorough comparison of quality and computational expense of several available packages, finding that GPy is one of the top publicly available packages in accuracy and computational expense. The calculations done in this work use the package GPy, a Gaussian process package written in Python by the University of Sheffield.
3.6. SPARSE GAUSSIAN PROCESS REGRESSION

machine learning group (GPy 2012). The model is described with the data in Section 4.3.1.
Chapter 4

Application of Gaussian Process Regression to Gaia Data

The Gaia mission has the goal of measuring the positions, distances, space motions, and other physical characteristics of more than 1 billion stars, both in our Galaxy and beyond. The Gaia space telescope has operated from Earth’s Lagrange Point 2 since 2014. Observations have resulted in three large releases, aptly named Gaia Data Release 1, 2 (DR2), and 3 (DR3). This freely available stellar census provides the ability to explore the structure, motion, and evolutionary history of the Milky Way Galaxy.

In Chapter 1 an artist image of the Galaxy depicted the expected structure of the Galaxy and our approximate position within the Local Arm. Figure 4.1 presents how much of our Galaxy has actually been observed, superimposed over this artist image. This view, created from the work of Anders et al. (2019), depicts the distribution of 150 million observed stars from Gaia DR2. This work makes use of both Gaia DR2 and DR3, which we will explore further in Section 4.2.

This chapter covers the understanding and use of astrometric data, conversion between coordinate frames, the data we use, application to our method, and the construction of the input for our GP model.
Figure 4.1: The distribution of 150 million stars from *Gaia* DR2, superimposed on an artist impression of the Galaxy. Yellow-orange indicates more dense, purple indicates less dense, and no color change indicates no data. The Sun is positioned at the most dense region towards the bottom of the image, while the center of the Galaxy sees a large and elongated feature as an indication of the Galactic bar. From [Anders et al.](2019) and A. Khalatyan, using the StarHorse code ([Queiroz et al.](2018)).
4.1 Astrometry

To an observer on Earth, stars’ positions on the sky change over time as a result of the Earth’s spin, the motion around the Sun, and over longer periods, the individual angular motion of each star. This section gives a short overview of calculating velocities in the Galactic coordinate frame from observations.

Observers apply the celestial coordinate system as the projection of the geographic system of longitude and latitude onto the night sky. The equatorial angle is named Right Ascension (RA) which considers 360 degrees to be split into hours of the day, and further divided into minutes and seconds. The angle towards the poles is named Declination (DEC). The zero point of RA is the direction of the Sun at the exact calculated Vernal (spring) Equinox. Declination is measured in the direction of latitude, with the equator at 0°, the North Celestial Pole at 90°, and the South Celestial Sole at −90°. The angular (or proper) motions are named PMRA and PMDEC, corresponding to each dimension. Angular positions are time-variant, leading to use of a standard time, called the epoch to consider positions. The epoch applied with Gaia is J2000.0 in the ICRS system (Hobbs et al., 2022).

Distance to a star and its radial velocity complete the fully defined 3-D positions and velocities. The telescope Gaia determines distance through parallax, measuring the angle \( \alpha \) subtended as it moves over a 2 AU baseline through the year. The distance is then

\[
\tan(p) = p = \frac{1 \text{[AU]}}{d \text{[pc]}}
\]

where \( p \) is measured in arcseconds.

The radial velocity of a star is determined by Gaia using its spectrum. The light from
4.1. ASTROMETRY

Figure 4.2: Visualization of sky coordinates. Positions on the sphere are denoted by RA and DEC. The celestial poles mark the axis or rotation of the Earth. Proper motions are the angular motion in sky coordinates (Li et al., 2007).

A star with velocity relative to the observer will experience the relativistic Doppler shift. Stellar spectral lines with well known positions are used to determine the extent of the Doppler shift, and the velocity required to cause it.

Other factors may affect this determination, including images of the same star not being taken 6 months apart, the relative motion of the Earth and the Sun for radial velocities, and an obscured background giving bad parallax (Gaia Collaboration et al., 2016; Fabricius et al., 2016; Crowley et al., 2016).

4.1.1 Color and Magnitudes

The Gaia telescope observes stars in three main pass bands, given the names G (green), BP (blue photometer), and RP (red photometer). Green is a wide band measuring light from about 400 - 950 nm, while the BP and RP bands cover the region in the two ranges 350 - 675 nm and 625 - 1000 nm. The color bands measure a star’s flux over its region and apply
the *Gaia* calibration to turn this into units of magnitude.

The astronomical magnitude scale is a logarithmic scale of stellar flux. The zero point is selected as the flux from the star Vega, corrected over the observed band to give Vega a magnitude of 0.023 in the Johnson V band \cite{Busso2021}. The magnitude system uses a scale such that a magnitude difference of 5 will be exactly 100 times brighter, or a step is a factor in brightness of $\sqrt[5]{100} \approx 2.512$. The observed brightness of a star is called the apparent magnitude. Observed brightness varies with distance, so knowing the distance to a star, we can find the true or absolute magnitude of a star. The absolute magnitude is defined as the apparent magnitude that an object would have if it were placed at a distance of 10 pc, and the path of light was not affected by extinction due to dust or gas. The apparent magnitude may be converted to absolute magnitude with the equation

$$M = m - 5 \log_{10}(d) + 5$$  \hspace{1cm} (4.2)

where $M$ is the absolute magnitude, $m$ is the apparent magnitude, and $d$ is the distance in parsecs.

The color bands BP and RP may be used to determine the type of star and temperature of the star being observed. Approximating a star to be a black body, the difference in observed light for the two color bands BP $-$ RP is an indicator of the peak of its radiation curve, and its temperature. Low values of BP $-$ RP indicate a redder and hence cooler star. A common figure of stars using their observed light is a color-magnitude diagram (CMD), comparing the absolute magnitude to the difference between two color bands. An example of the CMD is shown in Figure 4.3 depicting the positions of types of stars on the diagram, and true CMDs for the data used are in Section 4.2.1 and 4.2.2. The key features of the CMD are the main sequence, a diagonal band of stars that runs from low luminosity red stars to high
4.1. ASTROMETRY

Figure 4.3: Example color-magnitude diagram, depicting the positions of types of stars on the diagram. Axes show comparisons between stellar spectral class, color and effective temperature on the X axis, and the magnitude scale and solar luminosity on the Y axis. The major regions of stars are the main sequence and the red giant branch at magnitude 0 (Hollow 2022).

luminosity blue stars, and a roughly horizontal band between 0-2 mag and BP – RP between 1-1.7 mag that corresponds to giant stars.

4.1.2 Coordinate Transformations

The study of Galactic motion often uses two systems. The first is the Galactocentric cylindrical coordinate system, using \((R, \phi, Z)\) with the Sun placed at a position \((R_0, \phi_0, Z_0) = (8.27, 0, 0.02)\) kpc (Schönrich et al. 2019, Astropy Collaboration et al. 2013). The second system we often use is a local Cartesian system \((X, Y, Z)\) with velocities
(U, V, W) where U is positive towards the Galactic centre, V is in the direction of rotation, and W is perpendicular to the disk. The changes to these systems requires several transformations from observed values. To begin this transformation we consider unit vectors within the ICRS frame

\[
\begin{bmatrix}
X_{ICRS} \\
Y_{ICRS} \\
Z_{ICRS}
\end{bmatrix} =
\begin{bmatrix}
\cos \alpha \cos \delta \\
\sin \alpha \cos \delta \\
\sin \delta
\end{bmatrix}
\]  \hspace{1cm} (4.3)

where the coordinate variables are (\(\alpha, \delta\)) as RA and DEC in ICRS. We introduce a second observational angular system called the Galactic system (van Altena, 2012). The Galactic system is oriented from the position with the Sun, with the zero point directed to the center of the Galaxy. The Galactic system uses variables (l, b), Galactic longitude (l) increases parallel to the Galactic plane and Galactic latitude (b) varies perpendicular to the plane. We then introduce the applicable unit vector

\[
\begin{bmatrix}
X_{Gal} \\
Y_{Gal} \\
Z_{Gal}
\end{bmatrix} =
\begin{bmatrix}
\cos l \cos b \\
\sin l \cos b \\
\sin b
\end{bmatrix}
\]  \hspace{1cm} (4.4)

These coordinate systems are converted by considering the position of the north Galactic pole, which in the Gaia determination is given by (\(\alpha_G, \delta_G\)) = (192.85948, +27.12825), and the Galactic longitude of the first intersection of the Galactic plane with the equator is given by \(l_\Omega = 32.93192\) (Hobbs et al., 2022). The transformation between the frames is
given by the matrix multiplication

\[ r_{Gal} = T r_{ICRS} \]  \hspace{1cm} (4.5)

where \( T \) is the rotation matrix found from angles \( \alpha_G, \delta_G, \) and \( l_\Omega \), and is given by

\[
T = R_z(-l_\Omega)R_x(90^\circ - \delta_G)R_z(\alpha_G + 90^\circ)
\]

\[
= \begin{bmatrix}
-0.0548755604162154 & +0.4941094278755837 & -0.8676661490190047 \\
-0.8734370902348850 & -0.4448296299600112 & -0.1980763734312015 \\
-0.4838350155487132 & +0.7469822444972189 & +0.4559837761750669
\end{bmatrix}. \hspace{1cm} (4.6)
\]

The transformation of proper motion to a Galactic frame considers the coordinate matrix

\[
A = \begin{bmatrix}
\cos \alpha \cos \delta & -\sin \alpha & -\cos \alpha \sin \delta \\
\sin \alpha \cos \delta & \cos \alpha & -\sin \alpha \sin \delta \\
\sin \delta & 0 & \cos \delta
\end{bmatrix}. \hspace{1cm} (4.7)
\]

The Galactic Cartesian coordinate velocity components are then

\[
\begin{bmatrix}
U \\
V \\
W
\end{bmatrix} = B \cdot \begin{bmatrix}
v_{los} \\
\mu_\alpha k/p \\
\mu_\delta k/p
\end{bmatrix} \hspace{1cm} (4.8)
\]

where \( B = T \cdot A \), \( v_{los} \) is the observed line of sight (radial) velocity relative to the Sun in km s\(^{-1}\), \( (\mu_\alpha, \mu_\delta) \) are the proper motions in the celestial frame, \( k = 4.74057 \) is a transformation factor between the units AU yr\(^{-1}\) and km s\(^{-1}\), and \( p \) is the parallax angle of the star (Johnson & Soderblom, 1987).
4.1. ASTROMETRY

Some astronomers consider a left-handed coordinate system as the preferred method. In this case, the left-handed system may be produced by changing the signs of the element of the top row of $T$, equating to a result of $-U$. This is the case for the dataset in Section 4.2.1.

The Galactic frame considers motion relative to the Sun. When analyzing Galactic dynamics, a non-rotating frame centered on the Galaxy may be advantageous. To consider this Galactocentric frame, stars are transformed using the position and velocity of the Sun as $(X_\odot, Y_\odot, Z_\odot) = (-8.27, 0, 0.02)\text{ kpc}$ and $(U_\odot, V_\odot, W_\odot) = (11.1, 250, 7.24)\text{ km s}^{-1}$ \cite{Schonrich2012}.

We finally consider a Galactocentric polar coordinate system $(R, \phi, Z)$ in which the $\phi$ direction follows the rotation of the galaxy clockwise. In this way the coordinate system may be considered left-handed. The conversion between the Cartesian velocities and the Galactocentric velocities is given by

\[
\begin{bmatrix}
V_R \\
V_\phi \\
V_Z
\end{bmatrix} =
\begin{bmatrix}
\cos \theta & - \sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
U \\
V \\
W
\end{bmatrix}.
\]

4.1.3 Uncertainty in Velocity

The equation for the variance of a function of several variables is given by

\[
\sigma_F^2_{(x,y,z)} = \left(\frac{\partial F}{\partial x}\right)^2 \sigma_x^2 + \left(\frac{\partial F}{\partial y}\right)^2 \sigma_y^2 + \left(\frac{\partial F}{\partial z}\right)^2 \sigma_z^2
\]

with the restriction that the errors must be uncorrelated. In the case of Gaia data, the assumption of uncorrelated errors is adopted. Given that the matrices $T$ and $A$ will produce
4.2. DATASETS

no error, Equation [4.10] may be applied to Equation [4.8], producing the equation

\[
\begin{bmatrix}
\sigma^2_U \\
\sigma^2_V \\
\sigma^2_W
\end{bmatrix}
= C
\begin{bmatrix}
\sigma_{\text{vlos}}^2 \\
(k/p)^2[\sigma^2_{\mu} + \left(\mu_{\alpha}\sigma_{\rho}/p\right)^2] \\
(k/p)^2[\sigma^2_{\mu} + \left(\mu_{\delta}\sigma_{\rho}/p\right)^2]
\end{bmatrix}
+ 2\mu_{\alpha}\mu_{\delta}k^2\sigma^2_{\rho}/p^4
\begin{bmatrix}
B_{12} \cdot B_{13} \\
B_{22} \cdot B_{23} \\
B_{32} \cdot B_{33}
\end{bmatrix}
\]

(4.11)

where \(B_{ij}\) are the elements of \(B\), and \(C\) is a matrix of the squares of individual elements of \(B\), \(C_{ij} = B_{ij}^2\) (Johnson & Soderblom, 1987). The errors may then be converted to Galactocentric coordinates with

\[
\begin{bmatrix}
\sigma^2_{VR} \\
\sigma^2_{V\phi} \\
\sigma^2_{VZ}
\end{bmatrix}
= 
\begin{bmatrix}
\cos^2\theta & \sin^2\theta & 0 \\
\sin^2\theta & \cos^2\theta & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\sigma^2_U \\
\sigma^2_V \\
\sigma^2_W
\end{bmatrix}
\]

(4.12)

These error calculations assume that errors are Gaussian, do not have correlation, and the coordinate transformations do not introduce further error or correlation. The assumptions are not strictly true, as has been explored by Li et al. (2007). For the purposes of this work, these assumptions are taken to be valid based on the relatively small size of these errors compared to the natural variation in stellar velocities, quality cuts performed on the data, and our use of binning which primarily uses the variation between stellar velocities. This is explored further in Sections 4.2.1 and 4.3.

4.2 Datasets

The Gaia mission provides the ability for astronomers to conduct research on a large scale with its comprehensive and precise stellar survey. Application of Gaia data to GP requires
that the dataset be of high quality to ensure that the features are not muddled by excessive noise, outliers, or erroneous data. This section describes two datasets in this work, the Schönrich stellar catalog processed from the *Gaia* DR2 Radial Velocity Sample, and the *Gaia* DR3 Radial Velocity Sample.

### 4.2.1 *Gaia* DR2 Radial Velocity Sample

The first dataset used is the *gaiaRVdelpeqspdelsp43* catalog, constructed by Schönrich et al. (2019) to correct systematic errors in the *Gaia* parallaxes and uncertainties. The catalog provides coordinates in both Galactic ($l, b, p, \mu_l, \mu_b, v_{los}$) and Galactocentric ($X, Y, Z, V_R, V_\phi, V_Z$). The catalog uses a left-handed system, placing the Sun at $X = 8.27$ kpc, and considering $V_R$ to point towards the center of the Galaxy. We transform the coordinates from the left-handed system used in Schönrich et al. (2019) to the right-handed system used in our work, and calculate the errors in Galactocentric coordinates using Equations 4.11 and 4.12 (as these are not provided).

We apply the following recommended quality cuts to the data to ensure minimal kinematic biases (Schönrich, 2012). This is followed by kinematics cuts in the style of Williams et al. (2013) to remove high-velocity outliers.

**Quality cuts:**

- color: $G_{bp} - G_{rp} < 1.5$
- magnitude: $G < 14.5$ and $G_{bp}, G, G_{rp} > 0$
- parallax signal to noise: $p/\sigma_p > 4$
- parallax uncertainty: $\sigma_p < 0.1$ mas with $\sigma_p$ given by the *Gaia* pipeline
- excess $BP - RP$ flux: $1.172 < E_{bp\text{rp}} < 1.3$

**Kinematic cuts:**
4.2. DATASETS

- proper motion: $\mu_\alpha, \mu_\delta < 400$ mas yr$^{-1}$
- proper motion error: $\epsilon_\mu_\alpha, \epsilon_\mu_\delta < 20$ mas yr$^{-1}$
- Galactocentric speed: $|V| < 600$ km s$^{-1}$

Many of the cuts are intuitive such as signal to noise ratio and uncertainties, however some are less intuitive. The color and excess are limited based on biases of underestimated distance that stars outside this range introduce into the catalog. These red stars and stars with excess $BP - RP$ flux likely are contaminated by neighbors or the background. Stars with magnitudes dimmer than 14.5 also find a very strong distance underestimate, affecting their calculated speeds. The kinematic cuts limit proper motions and speeds to exclude stars that are not bound to the disk, including possible stars originating from the halo or high velocity stars.

The set of restrictions reduces the original 6606247 stars to 4579820. In Figure 4.4 we show the spatial distribution of stars in our sample. The catalog presents a usable number of stars to a distance of approximately 4 kpc from the Sun, or about half the distance to the Galactic center, with most stars within 2 kpc of the Sun. Observational restrictions in the midplane, such as the abundance of sight obscuring dust and gas and the high density of stars, mean that most observed stars exist in the bands $Z = \pm[0.1, 1.5]$ kpc. The bottom left panel presents the CMD of the catalog, showing an abundant main sequence and giant branch. This CMD does not take into account any de-reddening that should occur due to line of sight effects such as extinction from dust clouds altering the color (Gaia Collaboration et al., 2018b).

The bottom right panel presents the distribution of stellar absolute magnitudes over distance from the Sun. The lower diagonal edge marks the magnitude limit of the sample at an apparent magnitude of approximately 15. Dimmer stars may be observed at lower distances as they produce the same visual brightness as brighter stars further away. The
upper diagonal edge marks the upper limit of the survey, in which a star’s brightness will saturate the instrument, preventing a measurement. While the upper brightness limit is fixed, the lower limit will move. *Gaia* prioritizes bright stars in its observation and as data grows, more stars with low apparent magnitude will be captured, as shown in Section 4.2.2. The ability to capture brighter stars more easily leads to a difference in the represented populations of stars with distance. For example, we may see that beyond 2 kpc (log 0.3) most of the stars with radial velocities are giants.

The Schönrich catalog of corrected *Gaia* DR2 stars is a great tool to study the kinematics of the Galactic disc. The Galactic motion captured by the catalog is modelled and analyzed with sparse GP regression in the published paper presented in Chapter 5. Further discussion of the selection of stars and possible effects such as velocity varying with stellar population are contained in Chapter 5.

### 4.2.2 *Gaia* DR3 Radial Velocity Sample

The *Gaia* DR3 release builds upon DR2, providing both better measurements of the previous set and expanding the number of stars with radial velocities to more than 34 million. The stellar positions and velocities in the Galactocentric system \((X, Y, Z, V_R, V_\phi, V_Z)\) are calculated from the 6-parameter celestial coordinates using Equations 4.3-4.12. Calculating from directly observed motions easily allows the inclusion of the desired position of the Sun as \((X_\odot, Y_\odot, Z_\odot) = (-8.27, 0, 0.02)\) kpc and \((U_\odot, V_\odot, W_\odot) = (11.1, 250, 7.24)\) km s\(^{-1}\). The series of restrictions from Section 4.2.1 reduces the original 34702191 stars to 23402827. This restriction shows a larger percentage of stars eliminated from the DR3 subset than the previous Schönrich subset. This is not due to the nature of the restrictions, but rather the Schönrich catalog applies a few basic restrictions in their calculations and correction of the
Figure 4.4: 2-D histogram of the Schönrich catalog in several depictions. The colorbar shows the number of stars in each bin. Top left: the X-Y plane, showing most of the stars in the datalog are close to the Sun at (-8.27,0). Top right: the X-Z plane, showing the shape of data with Z, and the reduced number of stars in the midplane. Bottom left: color-magnitude diagram of the catalog. Bottom right: absolute magnitude with distance, showing the distribution across the sample.

set before giving their recommendations for restrictions that have been adopted. Similar corrections to the parallax have not been applied for *Gaia* DR3 as many biases noted in DR2 have been taken into account (Gaia Collaboration et al., 2022b).

The extent of the dataset may be seen in Figure 4.5. The top panels show that a high density of observed stars exists within a distance of 6 kpc in the plane and approximately 3 kpc perpendicular to the plane. Individual stars extend well beyond this distance, but the
Figure 4.5: 2-D histogram of the Gaia DR3 catalog in several depictions. The colorbar shows the number of stars in each bin. Top left: the X-Y plane, showing most of the stars are close to the Sun at (-8.27,0). Top right: the X-Z plane, showing the shape of data with Z, and the reduced number of stars in the midplane. Bottom left: color-magnitude diagram of the catalog. Bottom right: absolute magnitude with distance, showing the distribution across the sample. Black boxes approximate the regions covered by the previous data set with Gaia DR2.

density is lower than 10 stars per $100 \times 100 \times 100$ pc$^3$ bin for most of the extended region, and too low to be an indicator of the disc motion for analysis. The black boxes approximate the range of the Schönrich dataset presented in Figure 4.4 to stress the increase in data available with DR3. The new data release pushes the distance to which radial velocities are available to more than double that of DR2, although sparsely covered. The bottom left panel is the color-magnitude diagram of the dataset, clearly showing the main sequence
stars and the extension to the giant branch and beyond. The dataset also contains a few faint white dwarf stars, although this population is less than 10 stars. The bottom right panel shows the distribution of the stars’ absolute magnitude with radius from the Sun. Similarly to DR2, the shape of the data follows the apparent magnitude limit, tracing a line of minimum observed brightness. The black shape marks the extent of DR2 for a simple comparison. The magnitude limit has moved about 2 magnitudes fainter, allowing both lower main sequence stars and stars at further distances to be captured. Giant stars still comprise most of the observed stars at a radius of more than 3 kpc.

4.3 Application to GP

In Chapter 3, we introduced the computational restrictions of GP regression, and the sparse GPR approximation that allows the computational complexity to be reduced to $O(NM^2)$, and the storage complexity to $O(NM)$. Our computational resources limit us to $N < 10^5$ data points with $M = 3000$ inducing points. Each datapoint is thus derived from a spatial bin that has $O(100)$ stars, with a minimum limit of 20 stars. Mean values and uncertainties for the $V_R$, $V_\phi$, and $V_Z$ components are calculated by a standard least squares algorithm, allowing the bin to be applied as a single point representing the region.

Performing this operation improves the assumption that data may be represented by a Gaussian process. The velocity distribution of stars is distinctly non-Gaussian due to the influence of stars from the thin disc, thick disc, and the stellar halo. The errors on the stellar velocities are also not Gaussian, owing to the larger error in the parallax incorporated into the tangential velocities calculated from proper motions. Applying binning invokes the central limit theorem, stating that a population with a mean and standard deviation will be approximately normally distributed given a sufficiently large sample. The smaller bin size
does not see the non-Gaussian variation over the large scale of the full dataset. The binning method applies a Cartesian grid of cells with size $(\Delta X, \Delta Y, \Delta Z) = (125, 125, 50)$ pc for the region $4 < X(\text{kpc}) < 12$, $-4 < Y(\text{kpc}) < 4$, and $-2 < Z(\text{kpc}) < 2$. The factor of 2.5 difference in $\Delta Z$ is meant to account for the difference between the radial scale length and the vertical scale height of the disk. Bins with less than 20 stars are less likely to be represented well (and less likely to fit the assumption of a Gaussian distribution for velocities) and are excluded from the model. For the DR2 dataset, 27305 cells representing the observations of 3972825 stars extend to about 4 kpc from the Sun. For the DR3 dataset, 86142 cells representing the observations of 20532202 stars extend about 6 kpc from the Sun.

The number of inducing points was selected from test trials. In Section 3.6.1, toy models found minimal improvement to the model fit past a threshold for the value of $M$. Testing with subsets of Gaia data, model improvement stopped at more than 2500 inducing points. To account for some possible differences between testing and the full dataset, 3000 inducing points were selected as the value for $M$.

### 4.3.1 Model Setup

The GP model is constructed using the kernels described in Section 3.3. In the analysis, the velocity components $V_R$, $V_\phi$, and $V_z$ have been modelled as independent scalar functions. This creates three separate GP models, using the same input positions for the binned grid with different mean velocities. The $V_R$ and $V_z$ components are well-represented by the RBF kernel. In an axisymmetric equilibrium model the mean of these components is zero, supporting the basic assumption of a zero mean GP. The assumption of Gaussian variations is then realistic. The $V_\phi$ component does not have a zero mean and must capture both the
rotation curve of the Galaxy, and be quadratic in $z$. Trial and error finds a function of kernels given by $k_{\text{RBF}} + k_{\text{lin}} * k_{\text{lin}}$ is able to capture the motion in the $V_\phi$ component. All kernels apply ARD to allow different lengthscales in each spatial dimension.

Implementation uses the Gaussian process package GPy, written in Python. The sparse GP module is applied with $M = 3000$ inducing points. The heteroscedastic Gaussian likelihood allows the incorporation of the heteroscedastic noise given by the bin uncertainties as discussed in Section 3.5. Optimization is carried out by applying the VarDTC inference method using the GPy optimization module and the L-BFGS-B algorithm from the software package SciPy (Nocedal & Wright, 2006; Virtanen et al., 2020). Optimization is restarted 10 times to reduce the possibility of a local optimum.

The sparse GP model requires about 16 GB of memory for the Gaia DR3 model with ~ 86000 data points and 3000 inducing points. Optimization of the models requires between 6 – 10 hours with a AMD Ryzen 9 5900HS CPU, varying with the number of likelihood calls required for optimization. Results from the model fit are presented in Chapters 5 and 6 for the DR2 and DR3 datasets.
Chapter 5

Gaussian Process Model for the Local Stellar Velocity Field from Gaia Data Release 2

This chapter presents the published manuscript in Monthly Notices of the Royal Astronomical Society, Volume 516, Issue 4, November 2022. It is included with permission of the co-author Lawrence M. Widrow.

Sections 1-4.3 of the paper contain some overlapping information with Chapters 2 - 4 of this thesis. Sections 4.4 - 6 present the results of mock data and applying the Schönrich DR2 catalog, and Sections 7 and 8 cover discussion and conclusions.

Abstract

We model the local stellar velocity field using position and velocity measurements for 4M stars from the second data release of Gaia. We determine the components of the mean or bulk velocity in ~ 27 000 spatially defined bins. Our assumption is that these quantities constitute a Gaussian process where the correlation between the bulk velocity at different locations is described by a simple covariance function or kernel. We use a sparse Gaussian process algorithm based on inducing points to construct a non-parametric, smooth, and differentiable model for the underlying mean stellar velocity field. We estimate the Oort constants $A$, $B$, $C$, and $K$ and find values in excellent agreement with previous results. Maps
of the velocity field within 2 kpc of the Sun reveal complicated substructures, which provide clear evidence that the local disc is in a state of disequilibrium. We present the first three-dimensional map of the divergence of the stellar velocity field and identify regions of the disc that may be undergoing compression and rarefaction.

5.1 Introduction

A common strategy for understanding the dynamics of the Milky Way is to assume that phenomena such as the bar, spiral arms, and the warp can be understood as perturbations from an equilibrium state (see e.g. [Binney 2013, Sellwood 2013]). A further assumption is that stars orbit in the mean gravitational field of gas, dark matter, and the other stars. The stellar components of the Galaxy are then described by a phase-space distribution function (DF), \( f(\mathbf{x}, \mathbf{v}, t) \), which obeys the collisionless Boltzmann equation (CBE) coupled to gravity via Poisson’s equation. The equilibrium/perturbation split then carries over to the DF and gravitational potential \( \Phi \). For a disc galaxy such as the Milky Way, the equilibrium state exhibits symmetry about both its mid-plane and rotation axis. Any departure from these symmetries therefore signals a departure from equilibrium.

The second Gaia data release (Gaia DR2; [Gaia Collaboration et al. 2018a]) includes measurements of positions and velocities for over 7 million stars thereby vastly increasing the number of stars for which all phase space coordinates are known. This data allows one to estimate the present-day stellar DF near the Sun. In general, interpretation of the DF involves the construction of a set of observables

\[
O = \int d^3x' d^3v' f(x', v', t) \mathcal{F}(x', v')
\]  
(5.1)
where $\mathcal{F}$ is designed to pick out particular features of the DF and also account for the selection function of the survey. In addition, one can incorporate observational uncertainties into Equation 5.1. As an example, the number density $n(x)$ is obtained by setting $\mathcal{F} = W(x - x')$ where $W$ is a localized window function:

$$\begin{align*}
n(x) &= \int d^3x' d^3v' f(x', v', t) W(x - x') \, . 
\end{align*}$$

(5.2)

In this paper, we model the mean velocity field for a subset of stars from the Gaia DR2 radial velocity survey as a function of position in the Galaxy. The mean velocity field corresponds to the first velocity moment of the DF. In the language of Equation 5.1, we have

$$\begin{align*}
V(x) &= n^{-1} \int d^3x' d^3v' f(x', v', t) W(x - x') v \, . 
\end{align*}$$

(5.3)

Here and throughout, we use uppercase $V$ to refer to the mean velocity field and lower case $v$ to refer to the velocity vector of individual stars.

For a system in equilibrium, $V$ is a function of Galactocentric radius $R$ and distance from the mid-plane $|z|$. Studies of the bulk velocity in pre-Gaia surveys such as the Sloan Extension for Galactic Understanding and Exploration (SEGUE; Yanny et al., 2009), the Radial Velocity Experiment (RAVE; Steinmetz et al., 2006), and the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Cui et al., 2012) revealed bulk motions that signalled a Galaxy in disequilibrium. For example, Widrow et al. (2012) determined the bulk vertical velocity $V_z$ and its dispersion $\sigma_z$ as functions of $z$ for SEGUE stars and found that they showed asymmetric features about $z = 0$. Williams et al. (2013) determined the full three-dimensional (3D) velocity field from some 70,000 red clump stars in the RAVE survey and found vertical motions that varied with Galactocentric radius $R$.
as well as evidence for radial flows. Similarly, Carlin et al. (2013) and Pearl et al. (2017) used LAMOST spectroscopic velocities with proper motions from the PPMXL catalogue (Roeser et al. 2010) for 400 000 stars to map out both vertical and radial bulk motions. Finally, Schönrich et al. (2019) and Friske & Schönrich (2019) found a wavelike pattern in the mean vertical velocity of Solar neighbourhood stars when plotted as a function of \( L_z \), the angular momentum about the symmetry axis of the disc. Since \( L_z \) can be used a proxy for the orbit-averaged Galactocentric radius, this observation provides further evidence for undulations in the disc.

Unfortunately, a clear picture of bulk motions near the Sun from these and other studies never emerged. The surveys each had their own complicated selection functions, as illustrated in Fig. 1 of Carlin et al. (2013). Furthermore, systematic distance errors could masquerade as velocity-space substructure (Schönrich 2012; Carrillo et al., 2018). These issues were addressed in Gaia Collaboration et al. (2018b) where a series of velocity field maps using data from Gaia DR2 and straightforward binning procedures were presented. For example, face-on maps were constructed by first dividing the sample into vertical slabs of width \( \Delta z = 400 \text{ pc} \) and then computing the mean velocity in \( XY \) cells of 200 pc by 200 pc. Similar binning procedures were used to construct maps of the velocity field in the \( Rz \) and \( R\phi \) planes.

The Gaia kinematic maps are simple and intuitive though the connection with theory, say through the CBE, requires further analysis. In particular, the mean velocities in the Gaia maps differ from the moments that appear in the continuity and Jeans equations since the former are averaged over cells of a finite size that must strike a balance between improved spatial resolution and reduced statistical fluctuations. The situation is only exacerbated when one goes to compute gradients of the velocity field.
In this paper, we take a different approach based on Gaussian process (GP) regression, a Bayesian, (almost) non-parametric approach to fitting data from the field of machine learning (see e.g. [Rasmussen & Williams, 2006]). GP regression has been used extensively in data science and engineering though only recently has it been applied to problems in astrophysics. An example that has some similarities with the problem at hand is the construction of dust and extinction maps in the Galaxy (see e.g. [Sale & Magorrian, 2018]). More recently, [Widmark et al., 2022] used GP regression to model the local number density field of stars using Gaia Data Release 3 (Gaia DR3; [Gaia Collaboration et al., 2022a]). There are, of course, other methods for modelling the mean velocity field. For example, [Bovy et al., 2015] and [Khanna et al., 2022] used Fourier analysis to characterize departures of the mean velocity field or streaming motion from an axisymmetric model.

Our provisional starting point is the assumption that the velocities of stars in the Galaxy constitute, to a good approximation, a GP. Formally, a GP is a collection of random variables such that any subset of these variables is Gaussian. Stellar velocities would seem to be a good candidate for a GP since the velocity distributions for the different stellar components (thin and thick discs, stellar halo) are well-described by anisotropic Maxwellians ([Binney & Tremaine, 2008]).

The fundamental object in GP regression is the covariance matrix, which describes correlations in the output variables for different values of the inputs. In our case, the outputs are the components of the velocity field and the inputs are different positions within the Galaxy. The covariance matrix is constructed from a kernel function, which depends on a set of hyperparameters that control the strength and scale length of correlations in the velocity field. Thus, although the model for the velocity field itself is non-parametric, the covariance function is parametric. The driver in GP regression is the optimization of the likelihood
function, defined below, over the space of hyperparameters. Note that GP regression is different from kernel smoothing or kernel density estimation where one produces a smooth model by convolving the data with a window function. In GP regression, one uses the data to select a model from the space of non-parametric functions in the GP prior.

To evaluate the GP likelihood function, one must invert an $N$ by $N$ covariance matrix where $N$ is the number of observations. This calculation is an $O(N^3)$ operation that requires $O(N^2)$ of rapid-access memory (RAM). Clearly, it is unfeasible to apply GP regression directly to the Gaia data, where $N$ in DR2 already exceeds $10^6$. In this work, we apply a two-fold strategy for handling the large-$N$ bottleneck. First, we bin stars in 3D cells and calculate a mean velocity vector for each cell. Our binning procedure leads to roughly 27 000 cells. We then use a sparse GP algorithm that deploys $M$ inducing points to approximate the likelihood function. This algorithm reduces the computational complexity to $O(NM^2)$ and the RAM requirements to $O(NM)$. Thus, our actual starting point is the assumption that the mean velocity vectors in the set of cells forms a GP.

Our GP analysis leads to a single model from which other properties of the velocity field can be derived. By contrast the [Gaia Collaboration et al. (2018b) used different binning strategies to explore different facets of the velocity field. Moreover, the Gaia maps are somewhere between a projection, where one integrates out one spatial dimension, and a two-dimensional (2D) slice, where one takes a narrow range in one of the dimensions to get the velocity field on a 2D surface. By contrast, our GP model can be queried to give the inferred velocity field with uncertainties at any point in the sample region. Furthermore, since the model is differentiable, we can use it to infer quantities such as the Oort constants at the position of the Sun. In addition, derivatives of $V$ can be used to make statements about the dynamical state of the Galaxy. This feature allows us to map out the velocity
gradient field in the vicinity of the Sun. It also allows us to construct a 3D map of the divergence, which is proportional to the total time derivative of the stellar number density. In so doing, we are able to identify regions near the Sun where the number density of stars may be undergoing compression or rarefaction.

An outline of our paper is as follows. In Section 5.2, we describe the sample used in our analysis as well as our binning strategy. In Section 5.4, we provide a brief introduction to GPs and outline our sparse GP algorithm. We also summarize results of tests performed with mock data. The results of our analysis for Gaia DR2 are discussed in Sections 5.5 and 5.6. We present estimates for the generalized Oort constants and maps of the velocity field as well as a 3D map of the divergence of the velocity field. We discuss possible extensions of this work in Section 5.7 and provide a brief summary of our results and some concluding remarks in Section 5.8.

5.2 Preliminaries

5.2.1 Six-Dimensional Phase-space Catalogue

The Gaia DR2 Radial Velocity survey contains six-dimensional phase-space measurements for over 7 million stars (Gaia Collaboration et al., 2018a,b). In this work, we use the gaiaRVdelpepsdelsp43 catalogue, which was constructed to correct for systematic errors in the Gaia parallaxes and uncertainties (Schönrich et al., 2019).

Following (Gaia Collaboration et al., 2018b) we use both Cartesian coordinates \((X, Y, Z, U, V, W)\) and Galactic cylindrical coordinates \((R, \phi, Z, V_R, V_\phi, V_z)\). The Cartesian coordinate system has the Galactic centre at the origin and the Sun on the \(-X\) axis; \(U, V,\) and \(W\) are the usual velocity components where positive values at the position of the Sun indicate motion towards the Galactic centre, the direction of Galactic rotation, and the
North Galactic Pole, respectively. We use 8.27 kpc for the Sun’s distance to the Galactic centre and 20 pc for its distance to the Galactic mid-plane (see Schönrich et al., 2019 and references therein) so that \((X_\odot, Y_\odot, Z_\odot) = (−8.27, 0, 0.02)\) kpc. We further set \((U_\odot, V_\odot, W_\odot) = (11.1, 250, 7.24)\) km s\(^{-1}\) (Schönrich, 2012). Our cylindrical coordinate system also has the Galactic centre at the origin, the Sun at \(\phi = 0^\circ\), and \(\phi\) increasing in the direction of Galactic rotation. This system is left-handed in the sense that increasing \(\phi\) corresponds to the direction towards the South rather than North Galactic pole. We note that at the position of the Sun \(V_R = −U\) and \(V_\phi = V\); \(V_Z = W\).

We calculate the \(u, v, w\) velocity components and their errors from astrometric observations using the method described in Johnson & Soderblom (1987). We then convert to \(v_R, v_\phi\) and \(v_z\). While the velocity components were provided in the gaiaRVdelpsedlspsp43 catalogue, the uncertainties were not. Note that in that catalogue, the Galactocentric cylindrical velocity components were given as \((U_g, V_g, W_g) = (−v_R, v_\phi, v_z)\). In addition, the \(X\)-coordinate in the catalogue differs from ours by a minus sign.

We apply the following restrictions to the sample:

- color: \(G_{bp} − G_{rp} < 1.5\)
- magnitude: \(G < 14.5\) and \(G_{bp}, G, G_{rp} > 0\)
- parallax signal to noise: \(p/\sigma_p > 4\)
- parallax uncertainty: \(\sigma_p < 0.1\) mas with \(\sigma_p\) given by the Gaia pipeline
- excess \(B−R\) flux: \(1.172 < E_{bprp} < 1.3\)

These quality cuts were recommended in Schönrich et al. (2019) to ensure minimal systematic biases in derived kinematic quantities. They leave 4 584 106 stars from the original catalog of 6 606 247. Finally, we apply the following kinematic cuts to remove high-velocity outliers:

- proper motion: \(\mu_\alpha, \mu_\delta < 400\) mas yr\(^{-1}\)
5.3. BINNING

- proper motion error: $\epsilon_\mu_\alpha, \epsilon_\mu_\delta < 20 \text{ mas yr}^{-1}$
- Galactocentric speed: $|V| < 600 \text{ km s}^{-1}$

These cuts are similar to those implemented by Williams et al. (2013) with the modification that our velocity cut is in terms of the speed in the local frame of rest whereas they apply a cut on the radial velocity. These cuts eliminate only about 400 stars.

The selection criteria adopted from Schönrich et al. (2019) and Williams et al. (2013) are meant to minimize systematic bias in parallax and distance measurements from Gaia and eliminate outliers. However, we have made no attempt to account for selection effects due to sample completeness. In effect, different parts of our sample will contain stars with different properties. Essentially, we are assuming that the mean velocity field is then independent of stellar properties such as intrinsic brightness and colour. We will return to this assumption in Section 5.7.

5.3 Binning

As discussed above, computation time and RAM requirements make it unfeasible to apply GP regression directly to data sets with much more than 104 entries. For this reason, we bin stars so that input data for our GP analysis are the mean velocity components in cells. In a sense, velocity components averaged over stars in a cell are closer to a GP than the stellar velocities themselves. Stars in the region near the Sun can come from the thin disc, the thick disc, or the stellar halo. In addition, the azimuthal velocity distribution of thin and especially thick disc stars is skewed due to asymmetric drift. Thus, the velocity distribution of stars in the Solar neighbourhood is decidedly non-Gaussian. On the other hand, the average velocity of some large number of stars in a cell will be approximately Gaussian due to the central limit theorem. We set up a Cartesian grid of cells with size
(ΔX, ΔY, ΔZ) = (125, 125, 50) pc for the region 4 < X(kpc) < 12, −4 < Y(kpc) < 4, and −2 < Z(kpc) < 2. The factor of 2.5 difference in ΔZ is meant to account for the difference between the radial scale length and the vertical scale height of the disc. Note that the width of the bins in Z is a factor of 8 smaller than the width used in constructing the Gaia Collaboration et al. (2018b) face-on maps. We keep only those cells with more than 20 stars. Mean values and uncertainties for the \( V_R, V_\phi, \) and \( V_Z \) components are calculated by a standard least squares algorithm. In the end, we are left with mean velocities for 27 305 cells representing the observations of 3 972 825 stars.

5.4 Gaussian Process Regression

5.4.1 Overview of Gaussian Processes

We begin with a brief review of GP regression. More thorough discussions can be found in numerous resources such as the excellent book by Rasmussen & Williams (2006). Suppose we have observations of a real scalar process \( f \) such that

\[
y_i = f(x_i) + \epsilon_i \quad i = \{1, 2, \ldots N\}
\]

(5.4)

where \( x_i \) is the input vector for the \( i \)'-th observation, \( y_i \) is the scalar output and \( \epsilon_i \) is additive noise for that observation. For the case at hand, \( x_i = (X_i, Y_i, Z_i) \) is the position vector of the \( i \)'th cell and \( f \) is \( V_R, V_\phi, \) or \( V_Z \). If \( f \) is a Gaussian Process then it is completely specified by a mean function \( m(x) \) and covariance function \( k(x, x') \) such that
5.4. GAUSSIAN PROCESS REGRESSION

\[ m(x) = \mathbb{E} [f(x)] \]  
\[ k(x, x') = \mathbb{E} [(f(x) - m(x))(f(x') - m(x'))] \]

where \( \mathbb{E} \) denotes expectation value. Furthermore, the probability DF of \( f \equiv [f_1, \ldots, f(x_n)] \) is given by

\[ P(f \mid X) = \mathcal{N}(f \mid \mu, K) \]

where \( X, \mu, \) and \( K \) are aggregate vectors of the input vectors \( x_i \), the mean functions \( m_i \equiv m(x_i) \), and the kernel functions \( K_{ij} \equiv k(x_i, x_j) \), respectively. As usual, \( \mathcal{N} \) denotes a multivariate Gaussian. Equation 5.7 constitutes a Gaussian process prior on \( f \). For simplicity, we assume \( m = 0 \).

The goal of GP regression is to infer \( f_* \) at new inputs \( X_* \) given the data \{\( X, y \)\}. If the noise \( \epsilon \) is identical, independent, and Gaussian with dispersion \( \sigma_n \), then the joint probability for \( y \) and \( f_* \) is itself Gaussian

\[ P(y, f_* \mid X, X_*) = \mathcal{N} \left( 0, \begin{bmatrix} K + \sigma_n^2 I & K_* \\ K_*^\top & K_{**} \end{bmatrix} \right) \]

where \( K \equiv K(X, X), K_* \equiv K(X, X_*) \) and \( K_{**} \equiv K(X_*, X_*) \). The quantity of interest is the conditional probability

\[ P(f_* \mid y, X, X_*) = \frac{P(y, f_* \mid X, X_*)}{P(y)} \]
\[ = \mathcal{N} \left( \tilde{f}_*, C \right) \]
where
\[ \tilde{f}_* = K_+^T \left( K + \sigma^2 I \right)^{-1} y \] (5.11)

and
\[ C = K_{++} - K_+^T \left( K + \sigma^2 I \right)^{-1} K_+ \] (5.12)

(see Rasmussen & Williams, 2006, section 2.2 and appendix A).

### 5.4.2 Kernel Function

A key ingredient of GP regression is the kernel or covariance function. Though there is considerable flexibility in choosing a kernel, for most problems it is constructed by taking sums and/or products of standard kernels, which themselves are constructed from elementary functions of the input variables.

In the present analysis, we allow for different kernels for \( V_R, V_\phi, \) and \( V_z \) since they are modeled as independent scalar functions. Each of the kernels uses a 3D radial basis function (RBF) plus a term proportional to the identity matrix that accounts for unknown noise

\[ k(x_i, x_j) = \sigma_f^2 e^{-\tilde{r}_{ij}/2} + \sigma_n^2 \delta_{ij} \] (5.13)

where \( \sigma_f^2 \) is the signal variance and

\[ \tilde{r}_{ij}^2 \equiv \frac{(X_i - X_j)^2}{l_x^2} + \frac{(Y_i - Y_j)^2}{l_y^2} + \frac{(Z_i - Z_j)^2}{l_z^2} \] (5.14)

is a dimensionless pseudo-distance between \( x_i \) and \( x_j \). The RBF kernel is positive definite, differentiable, and maximal for \( x_i = x_j \). These are all features that lead to realistic models for the bulk velocity field of the Galactic disc. Through trial and error, we find that
5.4. GAUSSIAN PROCESS REGRESSION

the fit for $V_\phi$ is significantly improved if we include a term corresponding to the product of two linear kernels. The linear kernel is given by

$$k_{\text{lin}}(x, x') = \sum_{i=x,y,z} \mu_i x_i x'_i .$$

(5.15)

It expands the space of priors on $f$ to include linear functions of the inputs $x$. Unlike the RBF kernel, it is non-stationary in the sense that it depends on the absolute positions of the data points rather than the distance between pairs of data points. Here, we add

$$k_{\text{lin2}} = \left( \sum_{i=x,y,z} \mu_i x_i x'_i \right) \left( \sum_{j=x,y,z} v_j x_j x'_j \right)$$

(5.16)

to the RBF and noise kernels in Equation 5.13.

It is not surprising that $V_\phi$ requires a more complicated kernel than the ones for $V_R$ and $V_z$. In an axisymmetric, equilibrium model, the mean radial and vertical components of the velocity field are zero whereas the azimuthal velocity field depends on both $z$ and $R$. Typically $V_\phi$ for an equilibrium model will be quadratic in $z$ and linear in $R$ in a patch of the Galaxy near the mid-plane. The combination of Equations 5.13 and 5.16 allow one to model these motions along with non-axisymmetric residuals.

The kernels defined above depend on a number of free parameters commonly referred to as hyperparameters since they parameterize the covariance function rather than $f$ itself. The hyperparameters determine characteristics of functions in the prior of $f$. For example, $l_x, l_y,$ and $l_z$ control the correlation lengths along the three coordinate axes. Proper choice of hyperparameters is crucial for obtaining a good model of the data. If the length-scales are too small, then the model will tend to overfit the data, thereby attributing small-scale bumps and wiggles to $f$ rather than noise. Conversely, if the length-scales are too large,
then the model will tend to miss important features represented in the data.

In simple problems, one can often find suitable hyperparameters by trial and error as illustrated in Chapter 2 of [Rasmussen & Williams (2006)]. Here we adopt a principled approach suitable for more complicated problems in which the hyperparameters $\theta$ are determined by maximizing the marginal likelihood function $p(y|X, \theta)$ where

$$
\log(p(y|X, \theta)) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |K + \sigma_n^2 I| \\
- \frac{1}{2} y^T (K + \sigma_n^2)^{-1} y.
$$

(5.17)

### 5.4.3 Sparse GP Regression via Inducing Points

In our optimization scheme, the marginal likelihood function in Equation (5.17) must be evaluated a large number of times for different choices of the hyperparameters $\theta$. Each likelihood call involves inversion of the $N \times N$ matrix $K$, which is an $O(N^3)$ operation that requires $O(N^2)$ RAM. This process becomes unfeasible for $N$ much greater than $10^4$. Fortunately, there are a number of algorithms that allow one to estimate the marginal likelihood using $M < N$ inputs. These algorithms, generally referred to as sparse GP regression, reduce the computational complexity to $O(NM^2)$ and the RAM requirement to $O(NM)$. In this work, we follow the method described in [Bauer et al. (2016)]. We denote the full covariance matrix ($K$ in Equation (5.7)) as $K_{ff}$, the covariance matrix for the inducing points as $K_{uu}$ and the cross covariance matrix as $K_{fu}$. In sparse GP regression, one replaces $K_{ff}$ in Equation (5.17) with $Q_{ff} \equiv K_{fu} K_{uu}^{-1} K_{uf}$ and includes an additional term given by $\text{Tr}(K_{ff} - Q_{ff})/2\sigma_n^2$ ([Titsias 2009; Bauer et al. 2016]). Note that optimization requires that we compute the gradient of the marginal likelihood with respect to the hyperparameters.

We implement our GP algorithm using GPY, a Gaussian process package written in
5.4. GAUSSIAN PROCESS REGRESSION

Python by the University of Sheffield machine learning group (GPy, 2012). We apply the sparse GP module with a heteroscedastic Gaussian likelihood, which allows us to incorporate uncertainties in the mean velocity components for individual cells. Optimization is carried out by applying the VarDTC inference method using the GPy optimization module and the L-BFGS-B algorithm from the software package SciPy (Titsias 2009; Nocedal & Wright 2006; Virtanen et al. 2020).

5.4.4 Mock Data Tests

Before turning to Gaia DR2, we test our algorithm on a mock data sample. This sample is constructed by replacing measured velocities in the gaiaRVdelpepsdelsp43 catalog with artificial velocities drawn from the distribution

\[ \mathbf{v} = \mathbf{V}(X, Y, Z) + \mathbf{N}(0, \sigma) \] (5.18)

where \( \mathbf{N} \) is a vector of normally-distributed variables and \( \mathbf{V} \) is an analytic function that is chosen so that the resulting maps are qualitatively similar to ones found with the real data. For brevity, we present results for a single generic component of the velocity field with \( \sigma = 20 \text{ km s}^{-1} \) and

\[ \mathbf{V}(X, Y, Z) = 2 \left( 11 \cos g_1 - 3 \cos g_2 - 2 \cos g_3 - Z^2 - 10 \right) \] (5.19)

where \( g_1 = Z + (X - 8)/6 \), \( g_2 = 0.52(Y - 2) \), and \( g_3 = X - 1.3Y - 11.3 \) with \( X, Y, \) and \( Z \) in units of kpc. The velocity field for \( Z = 0 \) is shown in the upper left panel of Figure 5.1.

The mock data are analysed using the same procedure that is applied to the real data. We first determine mean values for \( \mathbf{V} \) and errors about the mean for the 27,000 cells. We
Figure 5.1: Results of a mock data test. Top left: the generating field for the mock data. Top right: the confidence region of the model. Bottom left: the GP fit of the mock data. Bottom right: the difference between the mock field and the model, scaled by the dispersion as calculated in the GP regression. Here and throughout, the black dot indicates the position of the Sun and velocities are given in km s$^{-1}$.

then optimize the likelihood function over the hyperparameters using the inducing point algorithm described above with $M = 3000$ and the RBF plus noise kernel function. Armed with optimal hyperparameters, we determine the posterior of the model velocity field $\tilde{V}$ using Equation 5.11. The result for the $XY$-plane is shown in the lower left panel of Figure 5.1. We also determine the covariance matrix of the posterior using Equation 5.12. The square root of its diagonal elements provides an estimate of the dispersion, $\epsilon_V$, which is shown in the upper right panel of Figure 5.1. As expected, the dispersion rises sharply
5.5. Results

As with the mock data, we analyse the Gaia DR2 sample by first determining the mean values and uncertainties for the three velocity components in our 27 000 cells. We then optimize the likelihood function over the prior distribution of hyperparameters. The results
Table 5.1: Optimal hyperparameters from our sparse GP analysis of *Gaia* data. Units are kpc for $l_i$, km s$^{-1}$ for $\sigma_f$ and km s$^{-1}$kpc$^{-2}$ for $\mu_i$ and $v_i$.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>$V_R$</th>
<th>$V_\phi$</th>
<th>$V_Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$l_x$</td>
<td>1.06</td>
<td>1.13</td>
<td>1.89</td>
</tr>
<tr>
<td>$l_y$</td>
<td>2.72</td>
<td>1.75</td>
<td>4.18</td>
</tr>
<tr>
<td>$l_z$</td>
<td>0.536</td>
<td>0.242</td>
<td>0.484</td>
</tr>
<tr>
<td>$\sigma_f$</td>
<td>12.4</td>
<td>10.4</td>
<td>3.31</td>
</tr>
<tr>
<td>$\mu_x$, $\nu_x$</td>
<td>0.236, 0.228</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_y$, $\nu_y$</td>
<td>0.207, 0.209</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_z$, $\nu_z$</td>
<td>3.86, 3.71</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

are given in Table 5.1. We see that $l_y/l_x \sim 1.5 - 2.5$ and $l_z/l_x \sim 2 - 5$. The hierarchy of scales is expected given that the radial length scale of the disk is a factor of 5 - 10 times larger than the scale height and that variations in the disk tend to be stronger in the radial direction than the azimuthal direction.

### 5.5.1 Velocity Field in Components

Figures 5.3 and 5.4 show model predictions for the velocity field in the $z = 0$ and $\phi = 0^\circ$ planes. These figures are produced by querying the model at the appropriate points using Equation 5.11. By and large, the maps are in agreement with those found in *Gaia Collaboration et al.* (2018b), and in particular, their Figures 10 and 11. For example, $V_R$ in the midplane shows a stronger dependence on $R$ than $\phi$. In particular, $V_R$ is positive inside the Solar circle and negative just beyond the Solar circle, with hints of another sign reversal beyond $R = 9$ kpc. The azimuthal streaming motion seems to be faster inside the Solar circle than just beyond it. On the other hand, the gradient in the $V_z$, as seen in both *Gaia Collaboration et al.* (2018b) and our maps, is in the direction of increasing $R$ and $\phi$. The GP analysis tends to produce smoother maps where the model prediction for the velocity
Figure 5.3: Left: Components of the velocity field $V_R$, $V_\phi$, and $V_z$ in the $z = 0$ plane. Right: One-sigma confidence regions of the GP models. Colormap units are in km s$^{-1}$. 
Figure 5.4: Left: Components of the velocity field in the $\phi = 0$ plane. Right: One-sigma confidence regions of the GP models. Colormap units are in $\text{km s}^{-1}$. 
at any point is informed by the data within a volume characterized by $l_x$, $l_y$, and $l_z$. By contrast, the Gaia maps simply record a mean velocity field as calculated for individual voxels centered on $z = 0$ across the $XY$ plane.

Figures 5.3 and 5.4 also show the statistical uncertainties in the mean for the three components, $\epsilon_i$, as calculated from Equation 5.12. In general the uncertainties are less than 1 km s$^{-1}$ within the sample volume, but rise rapidly as one approaches the edge of the volume. Our maps can also be compared to those in Khanna et al. (2022) who combine data from Gaia EDR3 with radial velocity measurements from a number of spectroscopic surveys. They construct parametric models of the velocity field in heliocentric coordinates so a direct comparison is difficult.

The upper left panels of Figures 5.3 and 5.4 point to several regions of radial bulk flows. In particular, there is inward bulk motion just beyond the Solar circle that is approximately independent of $\phi$ across the sample volume. This motion is primarily in the region $|z| < 500$ pc and is thus likely associated with the thin disk. On the other hand, the outward radial flow just inside the Solar circle is mainly found at positive $\phi$ but across the full range in $z$ of the survey. There also appears to be another outward flow at $R > 10$ kpc, $\phi \approx -10^\circ$ and positive $z$.

The middle-left panels of Figures 5.3 and 5.4 show $V_\phi$. The dominant flow here, and indeed for $V$ in general, is the motion about the Galactic centre. We see that contours of constant $V_\phi$ roughly follow contours of constant $R$ in the panel of Figure 5.3. As discussed below, the average of $V_\phi$ over $\phi$ at $z = 0$ yields the rotation curve near the Solar circle. The main feature in the panel of Figure 5.4 is a decrease of $V_\phi$ as one moves away from the mid-plane of the Galaxy. This trend is easily explained by an increase in asymmetric drift due to a larger contribution from dynamically warmer stars in the thick disc. As with
5.5. RESULTS

\( \dot{V}_R \), we see that \( \dot{V}_\phi \) is not perfectly symmetric about the midplane. Consider, for example, the ridge in \( \dot{V}_\phi \) between \( X = -6 \) and \( X = -8 \) in Figure 5.3, which coincides with the peak seen in Figure 5.4 at \( R = 7 \) kpc. The peak and its outer slopes are shifted by a small, but non-negligible, amount above the mid-plane.

The bottom left panels of Figures 5.3 and 5.4 show our results for the vertical bulk motion, \( \dot{V}_z \). In the mid-plane we see a clear trend of increasing \( \dot{V}_z \) as one moves across the local patch of the Galaxy in the direction of increasing \( R \) and \( \phi \). The view in the \( Rz \) plane shows what may be described as a mix of bending and breathing motions. The former refers to motion in the same direction above and below the midplane and can be thought of as the coherent motion of a patch of the disk in the \( z \) direction. The latter refers to motion either toward or away from the midplane on either side of \( z = 0 \), i.e. compression or rarefaction of the disk. Inside the Solar circle, the dominant motion is in the negative \( z \)-direction. Outside the Solar circle \( \dot{V}_z \) is a combination of motion in the positive \( z \)-direction and motion away from the midplane. The generation of bending and breathing waves through internal disk dynamics and interactions with satellites such as the Sagittarius dwarf has been studied in N-body simulations by numerous authors including [Gómez et al. (2013); Chequers et al. (2018); Laporte et al. (2019); Poggio et al. (2021a); Bennett et al. (2022); Thulasidharan et al. (2022)]. Observational studies using data from various surveys in [Widrow et al. (2012); Williams et al. (2013); Carlin et al. (2013); Carrillo et al. (2018); Gaia Collaboration et al. (2018b); Wang et al. (2019); López-Corredoira et al. (2020)].

5.5.2 Velocity Vector Field

To aid our understanding of the velocity field we present velocity vector maps in three projections. Similar vector field maps were presented in [Pearl et al. (2017)] using data from
5.5. RESULTS

Figure 5.5: Map of the velocity field in $\phi Z$ plane along a cylindrical surface at $R = 8.27$. Arrows represent the components $V_\phi$ and $V_Z$ where, for clarity, we subtract 220 km s$^{-1}$ from $V_\phi$. The background color map shows $V_R$ while black contour indicates the curve where $V_R = 0$.

LAMOST and, most recently, by Fedorov et al. (2022) using data from Gaia DR2. We begin with Figure 5.5, which presents the vector field in the $\phi Z$ plane at $R = R_0$, that is, along a curved cylindrical surface that includes a part of the Solar circle. To help visualize the velocity field, we have subtracted off the vector 220 km s$^{-1}\hat{\phi}$. Radial velocities are shown as a color map. Thus, all three components of the velocity field are represented in the figure.

As already noted in the $V_\phi$ panel of Figure 5.6, the dominant feature of the map is the decrease in $V_\phi$ due to asymmetric drift as one moves away from the midplane. Velocities
5.5. RESULTS

Figure 5.6: Velocity vector map in the $RZ$-plane for $\phi = 0^\circ$ with arrows representing $V_R$ and $V_Z$. The background color map shows $V_\phi$ with the black contour indicating the curve where $V_\phi = 220$ km s$^{-1}$.

in the midplane are about 10 km s$^{-1}$ higher than they are at $|Z| = 500$ pc. Though the dominant flow is in the azimuthal direction, we do see a clear downward motion for $\phi < 0^\circ$, in agreement with Figure 5.7.

In Figure 5.6 we show the velocity field in the $RZ$-plane at $\phi = 0^\circ$, which passes through the position of the Sun. We see that there are three distinct regions defined primarily by the sign of $V_R$. Specifically, we find an inward flow centered on $(R, z) = (8.5, -0.2)$ kpc, an outward and downward flow inside the Solar circle, and an outward and upward flow for $R > 10$ kpc and $z \sim 500$ pc. The results are in good agreement with Figure 10 of Pearl et al. (2017) for regions where the samples overlap.

In Figure 5.7 we show the velocity field in the midplane while the background color
5.5. RESULTS

Figure 5.7: Velocity vector map in midplane. Arrows represent the $V_X$ and $V_Y$ components with the vector $230 \text{ km s}^{-1} \hat{\phi}$ subtracted for clarity. The background colormap shows $V_Z$ and the black contour indicates the curve where $V_Z = 0$. Arcs show the model for nearby spiral arms from Reid et al. (2019). In order from the inner to the outer disc they are: Scutum (cyan), Sagittarius (magenta), Local Arm (blue), and Perseus (black).

The map shows vertical bulk motions. Note that in this figure the in-plane vectors are shown relative to $230 \text{ km s}^{-1} \hat{\phi}$ rather than $220 \text{ km s}^{-1} \hat{\phi}$. The latter value was used in Figures 5.5 and 5.6 where we wanted to highlight the variation in $V_\phi$ as one moved from the midplane to $|z| \sim 1 \text{ kpc}$. Here, we want to highlight non-axisymmetric motions in the $z = 0$ plane, where the mean velocity is close to $230 \text{ km s}^{-1} \hat{\phi}$. The map paints an image of several flows in the vicinity of the Sun, which can be describe by a combination of expansion or compression and shear. For example, roughly along the Solar circle, we find radial compression and azimuthal shear — the dominant motion inside the Solar circle is outward and toward
positive $\phi$ while the motion just beyond the Solar circle is inward and toward negative $\phi$.

As discussed in Gaia Collaboration et al. (2018b) and references therein, departures in the stellar velocity field from an axisymmetric model can arise from internal mechanisms such as the bar and spiral structure as well as external ones such as the interaction of the disk with a satellite galaxy. Siebert et al. (2012) constructed a model for the radial motion of stars in the vicinity of a spiral arm and found that the stars along the inner or trailing edge should move radially outward while stars along the outer or leading edge should move inward. They attributed the change in sign of $V_R$ around $R \approx 8$ kpc in their analysis of RAVE data (their Figure 3) and also seen in our $V_R$ map to this effect. Using results from various N-body+hydrodynamics simulations, Kawata et al. (2014) and Grand et al. (2015) found that stars along the trailing edge of a spiral arm should rotate slower than average in addition to moving outward while stars along the leading edge should rotate faster than average. Hunt et al. (2015) also carried out simulations of spiral galaxies and used the SNAPDRAGON code (Hunt et al., 2019) to tag stars according to stellar populations. The results presented a more complicated picture where motions varied significantly among the different stellar populations.

Following Gaia Collaboration et al. (2018b) we overlay the Reid et al. (2019) model for the four closest spiral arms on our velocity field map in Figure 5.7. Our first observation is that the spatial scale of the features picked out in our model is comparable to the inter-arm spacing. Thus, the current model and data do not appear to have the resolution to connect velocity flows to particular spiral arms let alone pick out differences in motion between leading and trailing sides of an arm. For example, the radially outward flow on the trailing side of the Local Arm discussed above appears to originate inside the Sagittarius arm.

Thus, the connection between streaming motions and current models for spiral arms
appears tenuous at best. Nevertheless, there should be a connection between the velocity field and structure within the disk through the continuity and Euler or momentum equations. With this in mind, we turn to our next topic, gradients of the velocity field.

### 5.6 Velocity Gradient

In the previous section, we showed that our GP model could be queried to give the bulk velocity field \( \mathbf{V} \) at any position in the sample region. Since the GP solution is continuous and differentiable, we can also use it to infer the gradient of the velocity field, a \( 3 \times 3 \)-tensor whose components are \( \partial V_i / \partial X_j \). These components can be computed either by replacing the kernel functions in Equations 5.11 and 5.12 with their gradients or by a finite difference scheme. We have confirmed that the two methods agree for the RBF kernel. In what follows we use finite differencing. Similar results are presented in Fedorov et al. (2021).

#### 5.6.1 Oort Constants

Gradients of the velocity field played a central role in Oort’s seminal work on the rotation of the Galaxy (Oort [1927]). The basic idea was to expand the velocity field near the Sun as a Taylor series in position. Though Oort considered only axisymmetric flows in the \( X - Y \) plane, the method was extended to general three dimensional flows by Ogrodnikoff (1932); Milne (1935); Chandrasekhar (1942) and Ogorodnikov (1965).

To linear order in position, the velocity field near the Sun can be written as a Taylor series

\[
\mathbf{V} = \mathbf{V}_\odot + \mathbf{H} \cdot \mathbf{x} + O(\mathbf{x}^2). \tag{5.20}
\]

If we restrict ourselves to the projection of the velocity field onto the midplane of the Galaxy,
5.6. VELOCITY GRADIENT

then $H$ reduces to the $2 \times 2$ matrix

$$
H = \begin{pmatrix}
\frac{\partial V_x}{\partial x} & \frac{\partial V_y}{\partial x} \\
\frac{\partial V_x}{\partial y} & \frac{\partial V_y}{\partial y}
\end{pmatrix}_{x=0} = \begin{pmatrix}
K + C & A - B \\
A + B & K - C
\end{pmatrix}.
$$

(5.21)

where the second equality defines the Oort constants $A$, $B$, $C$, and $K$. These constants measure, respectively, the azimuthal shear, vorticity, radial shear, and divergence of the velocity field in the midplane of the disk. In Galactocentric polar coordinates $(R, \varphi)$ the Oort constants are given by

$$
2A = \frac{V_\varphi}{R} - \frac{\partial V_\varphi}{\partial R} - \frac{1}{R} \frac{\partial V_R}{\partial \varphi},
$$

(5.22)

$$
2B = -\frac{V_\varphi}{R} - \frac{\partial V_\varphi}{\partial R} + \frac{1}{R} \frac{\partial V_R}{\partial \varphi},
$$

(5.23)

$$
2C = -\frac{V_R}{R} + \frac{\partial V_R}{\partial R} - \frac{1}{R} \frac{\partial V_\varphi}{\partial \varphi},
$$

(5.24)

$$
2K = \frac{V_R}{R} + \frac{\partial V_R}{\partial R} + \frac{1}{R} \frac{\partial V_\varphi}{\partial \varphi}.
$$

(5.25)

Chandrasekhar (1942).

The usual method for determining the Oort constants from astrometric data is based on the observation that the proper motion of a star in the direction of Galactic longitude, $\mu_l$, and the line-of-sight velocity, $v_{\text{los}}$, can be expressed in terms of sine and cosine functions of $l$. The Oort constants appear as coefficients in this truncated Fourier series and can therefore be determined by standard statistical methods. (See, for example, Feast & Whitelock (1997); Olling & Dehnen (2003); Bovy (2017); Vityazev et al. (2018); Wang et al. (2021)).

In this paper, we compute the Oort constants directly from our GP model for the velocity field using the above equations. Our results along with a selection of values from the literature are given in Table 5.2. We find excellent agreement with recent measurements.
Table 5.2: Comparison of measured values for the Oort constants from the literature and from this work in units of km s$^{-1}$ kpc$^{-1}$. For the GP model we provide the inferred values at the exact position of the Sun and average values over a spherical volume of radius 500 pc.

<table>
<thead>
<tr>
<th>Origin</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oort (1927)</td>
<td>~19</td>
<td>~-24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feast (1997)</td>
<td>14.8 ± 0.8</td>
<td>-12.4 ± 0.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Olling (2003)</td>
<td>15.9 ± 2</td>
<td>-16.9 ± 2</td>
<td>-9.8 ± 2</td>
<td></td>
</tr>
<tr>
<td>Bovy (2017)</td>
<td>15.3 ± 0.4</td>
<td>-11.9 ± 0.4</td>
<td>-3.2 ± 0.4</td>
<td>-3.3 ± 0.6</td>
</tr>
<tr>
<td>Vityazev (2018)</td>
<td>16.3 ± 0.1</td>
<td>-11.9 ± 0.1</td>
<td>-3.0 ± 0.1</td>
<td>-4.0 ± 0.2</td>
</tr>
<tr>
<td>Li (2019)</td>
<td>15.1 ± 0.1</td>
<td>-13.4 ± 0.1</td>
<td>-2.7 ± 0.1</td>
<td>-1.7 ± 0.2</td>
</tr>
<tr>
<td>Wang (2021)</td>
<td>16.3 ± 0.9</td>
<td>-12.0 ± 0.8</td>
<td>-3.1 ± 0.5</td>
<td>-1.3 ± 1.0</td>
</tr>
<tr>
<td>This work (p)</td>
<td>16.2 ± 0.2</td>
<td>-11.7 ± 0.2</td>
<td>-3.1 ± 0.2</td>
<td>-3.0 ± 0.2</td>
</tr>
<tr>
<td>This work (v)</td>
<td>15.2 ± 0.8</td>
<td>-12.4 ± 0.9</td>
<td>-2.9 ± 0.8</td>
<td>-2.3 ± 0.5</td>
</tr>
</tbody>
</table>

As noted above, the values from the literature all determine the Oort constants by fitting $\mu_l$ and $v_{\text{los}}$ to a low-order Fourier series in $l$. They do, however, use data from different surveys and with different geometric selection functions and sample sizes. For example, both Bovy (2017) and Li et al. (2019) consider large samples of stars within 500 pc of the Sun. Bovy (2017) considers main sequence stars from the Tycho–Gaia Astrometric (TGAS) catalog (Michalik et al., 2015) while Li et al. (2019) considers all stars from Gaia DR2. Wang et al. (2021) uses a relatively small sample of A-stars from LAMOST with a similar range in distance from the Sun. Finally, Vityazev et al. (2018) uses stars from TGAS but with the larger reach of 1.5 kpc.

Our model allows us to predict the Oort constants at a single point, namely the position of the Sun. These values are given in the second to last line in Table 5.2. To allow for a closer comparison to literature values, we also include values for the Oort constants averaged over a spherical volume of radius 500 pc.
5.6. VELOCITY GRADIENT

Figure 5.8: Galactic parameters as a function of radius. Top: Oort parameters $A$ (left) and $B$ (right). Bottom left: $V_\phi$. Bottom right: Angular velocity $\Omega = A - B$. The model is queried along the line passing from the Galactic centre through the Sun. The red dashed lines show the 1 $\sigma$ uncertainties.

5.6.2 Oort Functions

In Oort's [1927] original work, the Galaxy is assumed to be axisymmetric, and stars follow circular orbits in the midplane of the Galaxy. Under this assumption, the circular speed can be expressed in terms of the potential: $V_c = R \partial \Phi / \partial R$. The constants $C$ and $K$ are then identically zero and $A$ and $B$ can be written in terms of circular speed and its gradient at the position of the Sun. They therefore provide a probe of the gravitational potential and hence matter distribution in the Galaxy. Olling & Merrifield (1998) extended the idea of Oort constants to Oort functions by writing
Figure 5.9: Oort constants $A$, $B$, $C$, and $K$ calculated from Equations 5.22 - 5.25. Colormap units are in km s$^{-1}$ kpc$^{-1}$. The spiral arms model of Reid et al. (2019) is also over-plotted for the local divergence ($K$). In order of inner to the outer disc the arms are: Scutum (cyan), Sagittarius (magenta), Local Arm (blue), and Perseus (black).

\begin{align}
2\tilde{A}(R) &= \frac{V_c}{R} - \frac{\partial V_c}{\partial R} \\
2\tilde{B}(R) &= -\frac{V_c}{R} - \frac{\partial V_c}{\partial R}
\end{align}
\( \tilde{A} \) and \( \tilde{B} \) differ from \( A \) and \( B \) in Equations 5.22 and 5.23 in three important ways. First, \( V_\phi \), which includes asymmetric drift, is replaced by \( V_c \). Second, \( A \) and \( B \) include a contribution from the azimuthal gradient of \( V_R \). And finally, \( \tilde{A} \) and \( \tilde{B} \) are functions of \( R \).

In what follows, we generalize the work of Olling & Merrifield (1998) by treating the elements of \( H \) in Equation 5.21 as functions of \( X \) and \( Y \) rather than as constants of a Taylor series about the position of the Sun. We begin in Figure 5.8 by showing \( A \) and \( B \) as functions of \( R \) for \( \phi = 0 \). The functions are evaluated using Equations 5.22 and 5.23. We observe the same broad trends as seen in Olling & Merrifield (1998) and more recently, Fedorov et al. (2021). \( A(R) \) is a decreasing function of \( R \) with a "bump" just inside the Solar circle. \( B(R) \) is an increasing function of \( R \). In the bottom panels of Figure 5.8 we plot the rotation curve \( V_\phi \) and angular velocity curve \( \Omega = |A - B| \), which show departures from a flat rotation curve at the few km s\(^{-1}\) level.

The expressions for the remaining Oort constants in Equations 5.24-5.25 can similarly be extended to function of both \( R \) and \( \phi \) to yield maps of Oort functions in the Galactic midplane. These are shown in Figure 5.9. Were the Galaxy in an axisymmetric steady state, we would find that \( C = K = 0 \) and \( A \) and \( B \) depend only on \( R \). Though \( C \) and \( K \) are about a factor of five smaller than \( A \) and \( B \) they are clearly nonzero. Moreover, all four functions show variations in \( \phi \) though the dominant gradients are in the \( R \)-direction. The prominent regions of shear (lower left panel showing \( C \)) and compression (lower right panel showing \( K \)) provide another way of visualizing the in-plane flows already seen in Figure 5.7.

### 5.6.3 Divergence of the Velocity Field

As mentioned above, the Oort constants are constructed from \( X - Y \) derivatives of the in-plane components of the velocity field. More generally, \( H \) in Equation 5.20 is a \( 3 \times 3 \)
Figure 5.10: Divergence of the velocity field, calculated using cylindrical coordinates. Regions with $|\nabla \cdot \mathbf{V}| > 5 \text{ km s}^{-1} \text{ kpc}^{-1}$ are plotted, positive in red and negative in blue. Coloring and opacity is scaled with amplitude of the divergence.
5.6. VELOCITY GRADIENT

tensor. This tensor can be written as the sum of a six-component symmetric tensor \( M^+ \) and a three-component anti-symmetric tensor \( M^- \)

\[
M^\pm = \frac{1}{2} \left( \frac{\partial V_i}{\partial x_k} \pm \frac{\partial V_k}{\partial x_i} \right)
\]

where as before, the components are evaluated at the position of the Sun \cite{Ogrodnikoff1932, Milne1935, Ogorodnikov1965, Tsvetkov2019, Fedorov2021}. Note that the coefficients of \( M^\pm \) include the Oort constants. For example, \( A = M^+_{12} \) and \( B = M^-_{21} \). Recently, the nine parameters of the velocity gradient method have been estimated using data from \textit{Gaia} DR2 \cite{Vityazev2018, Bobylev2021}.

Here, we focus on the divergence of the velocity field \( \nabla \cdot \mathbf{V} \), which corresponds to the trace of \( M^+ \). Recall that the zeroth moment of the collisionless Boltzmann equation yields the continuity equation

\[
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{V}) = 0
\]

where \( \rho \) is the stellar density. This equation may be written as

\[
\frac{1}{\rho} \frac{d\rho}{dt} = -\nabla \cdot \mathbf{V}
\]

where \( d/dt = \partial/\partial t + \nabla \cdot \mathbf{V} \) is the total derivative. Thus, a region of positive divergence indicates that either the stellar distribution is expanding, or stars are being transported into a region of lower density. Either way, \( \nabla \cdot \mathbf{V} \) is a clear signal of disequilibrium in the disk.

In Figure 5.10 we show the divergence in three dimensions from six different viewing angles. We find a prominent region of negative divergence corresponding to compression in the velocity field running through the Solar Neighborhood and roughly aligned with Galactic azimuth. This is in good agreement with the results for \( K \) in the Galactic plane as seen in the
lower-right panel of Figure 5.9. This region is bracketed by regions of positive divergence, one below the midplane and inside the Solar circle and the other above the midplane and outside the Solar circle. The amplitude of the divergence is $5 - 10 \text{ km s}^{-1}\text{kpc}^{-1}$, which implies that the time-scale for the logarithm of the density to change is of order $100 - 200 \text{ Myr}$.

5.7 Discussion

Our discussion of the divergence of the stellar velocity field illustrates how a GP model can provide a link between theory and data. In particular, the divergence gives us a direct indication of regions in the disk where the total time derivative of $\ln \rho$ is nonzero. Clearly, if we have $\rho$ itself (see Widmark et al., 2022), then we can estimate the convective derivative term $\mathbf{V} \cdot \nabla \rho$ and hence infer the partial time derivative of $\rho$. In this way, we gain information about time-dependent phenomena in the disk from a kinematic snapshot. Models of $\rho$ for particular tracer populations require detailed knowledge of the selection function but are certainly accessible given the wealth of astrometric data available.

In a similar fashion, GP models for components of the velocity dispersion tensor would allow one to study the first moments of the CBE, namely the Jeans equations. In contrast with the continuity equation, these equations involve gradients of the gravitational potential. Thus, we will typically have two unknown quantities, one involving time-derivatives of the moments and the other involving the potential. Nevertheless, one should be able to infer something about the dynamics of the stellar disk, and in particular, departures from equilibrium by modelling moments of the DF via GP regression.

These considerations suggest an extension of the present work where we treat the output as a single three-components vector rather than three independent scalars. Doing so would
allow for correlations between the different components of the velocity field. The framework for extending GP regression to vector outputs is already well-established (Alvarez et al., 2011).

Finally, we return to the comment made at the end of Section 5.2 regarding sample selection. As noted, we select stars without requiring that the sample be complete. Thus, different regions in our analysis use different populations of stars. Since later data releases of \textit{Gaia} will have many more stars, we should be able to refine our analysis by choosing narrower regions of the color-magnitude diagram where completeness can be assured. We would then be able to compare the velocity field for different subpopulations of stars as was done for the vertical number density profile in Bennett & Bovy (2019). An alternative would be to extend GP regression to a higher dimensional space that included one or more variables to distinguish among different stellar populations.

5.8 Conclusions

In this paper, we present a GP model for the mean or bulk stellar velocity field in the vicinity of the Sun using astrometric measurements from the \textit{Gaia DR2} radial velocity survey. The model is nonparametric in the sense that there are no prior constraints on the functional form of the velocity field. Instead, one specifies the functional form of the kernel function, which, through a set of hyperparameters, controls properties of the prior on the velocity field such as its coherence length.

The main challenge in applying GP regression to \textit{Gaia} data comes from the large-$N$ requirements in both computing time and RAM. Fortunately, there is a large effort within the machine learning community on addressing these problems (See e.g. Hensman et al., 2013). In this work, we have just started to exploit methods developed in that field.
Our model provides an alternative to the velocity field maps found in Gaia Collaboration et al. (2018b) and elsewhere. The *Gaia* maps used different binning schemes to compute different projections of the velocity field. In our case, we compute a single GP model (or more precisely, three independent models for each of the cylindrical velocity field components) from which properties of the velocity field could be derived. As a check of the model, we confirm that the values for the Oort constants derived by differentiating the velocity field agree extremely well with recent determinations in the literature.

Perhaps the most important property of our GP model is that it is smooth and differentiable and can therefore be used to connect astrometric data with theory. Already, our results for the divergence of the velocity field have allowed us to identify, via the continuity equation, regions of the Solar neighbourhood that may be undergoing compression or expansion. With the availability of significantly more data we should be able to build GP models for higher moments of the DF and therefore perform similar analyses using the Jeans equations. Thus GP modelling represents a promising new avenue for understanding the complex dynamics of the Milky Way.

**Data Availability**

Data used in this paper is available through Zenodo (https://doi.org/10.5281/zenodo.2557803)
Chapter 6

Gaussian Process Model with *Gaia* Data Release 3

*Gaia* Data Release 3 provides a further opportunity for modelling the velocity field of the Galaxy by extending available 6-D phase space data from 6.6 million to 34.7 million stars. With the increase in the number of data points, we can extend the model range from 4 kpc to 6 kpc, more than doubling the area covered by the model. This allows us to confirm the previously observed bulk motions within the model with reduced uncertainties. This chapter presents the results of the GP model applied to the *Gaia* DR3 dataset. Section 1 presents the model velocity component fields, Section 2 presents constructed vectorfields, and Section 3 explores improved maps of Oort constants and divergence in the model.

We construct the velocity field with *Gaia* DR3 following the same process as presented previously in Chapters 4 and 5. Binning was applied in a Cartesian grid with size \((\Delta X, \Delta Y, \Delta Z) = (125, 125, 50)\) pc for the region \(2 < X(\text{kpc}) < 14, -6 < Y(\text{kpc}) < 6,\) and \(-2.5 < Z(\text{kpc}) < 2.5\). Bins with more than 20 stars were retained for the model. For the DR3 dataset, 86142 cells representing the observations of 20532202 stars extend about 6 kpc from the Sun. The likelihood function was optimized to produce the set of hyperparameters in Table 6.1. Proportions between the hyperparameter directions are similar to the models of *Gaia* DR2 data, with the \(Z\) direction finding the smallest length-scale, and the \(Y\) direction (approximately azimuthal) finding the largest length-scale. Compared to the model with DR2, the addition of the squared linear kernel (lin2: Equation 5.16) has a
Table 6.1: Optimal hyperparameters from our sparse GP analysis of Gaia DR3 data. Units are kpc for $l_i$, km s$^{-1}$ for $\sigma_f$ and km s$^{-1}$ kpc$^{-2}$ for $\mu_i$ and $\nu_i$.

<table>
<thead>
<tr>
<th>Kernel</th>
<th>$V_R$</th>
<th>$V_\phi$</th>
<th>$V_Z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>RBF</td>
<td>0.977</td>
<td>1.44</td>
<td>1.52</td>
</tr>
<tr>
<td>RBF + lin2</td>
<td>1.66</td>
<td>1.83</td>
<td>4.29</td>
</tr>
<tr>
<td>RBF</td>
<td>0.398</td>
<td>0.349</td>
<td>0.586</td>
</tr>
<tr>
<td>$\mu_x$, $\nu_x$</td>
<td>3.64</td>
<td>13.8</td>
<td>11.4</td>
</tr>
<tr>
<td>$\mu_y$, $\nu_y$</td>
<td>0.245, 0.228</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_z$, $\nu_z$</td>
<td>0.232, 0.213</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\mu_z$, $\nu_z$</td>
<td>1.41, 0.81</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

reduced impact. The model fit values for $\mu$ and $\nu$ are smaller, and a model fit without the added linear kernels finds similar output results. For consistency, the models presented in this chapter apply the same kernel structure as in Chapter 5.

6.1 Velocity Field in Components

Figures 6.1 and 6.2 show model predictions for the velocity field in the $z = 0$ and $\phi = 0^\circ$ planes. These maps are in good agreement with our previous models, as can be seen by comparing Figure 5.3 with 6.1 and 5.4 with 6.2. Uncertainty values for the three models are similar within the overlapping region of the models, while beyond, DR3 models see a continued rise in uncertainty. The highest uncertainty is seen at the model edge, towards the center of the Galaxy. The Galactic bulge is predicted to extend to a radius of about 2.25 kpc, and the edge may be covered by the model (Amòres et al., 2013). While some bulge stars may be captured by the model, the model assumes a single underlying flow in the disk, and would not fit the bugle if it were extended further.

The radial bulk flows in the upper left panels of Figure 6.1 show a consistent general pattern of three regions: two with outwards bulk motion and one with inwards. This view
6.1. VELOCITY FIELD IN COMPONENTS

Figure 6.1: Left: Components of the velocity field $V_R$, $V_\phi$, and $V_z$ in the $z = 0$ plane. Contours are placed at 0 km s$^{-1}$ for $V_R$ and $V_z$, and 230 km s$^{-1}$ for $V_\phi$. Right: One-sigma confidence regions of the GP models. Colormap units are in km s$^{-1}$. The position of the Sun is placed as a grey dot at (-8.27,0).
is improved by showing a larger range in $\phi$, and revealing larger changes with $\phi$ compared to strongly radial dependence with the DR2 model. The region of inward motion just beyond the solar circle is resolved as being split by a small region of stars with little radial motion, but with an angle similar to that of the spiral arms. The region of outwards motion at $X = 6\text{ kpc}$ varies quickly towards $\phi \approx -20^\circ$, between positive and negative with low amplitude. At higher azimuthal angle, the bulk motion is larger in size, spanning several kpc. The inner disk behaviour is in contrast to the larger regions with high amplitudes in the outer disk. Figure 6.2 presents the model view in $RZ$ coordinates. In this view, we may see the model is entirely consistent with DR2 in the overlapping regions, and provides further modelling towards the Galactic center. The region within $R = 6\text{ kpc}$ shows a variety of motions varying with both $R$ and $Z$, though the uncertainty in this region is high and a link between modelled motions and theory cannot be made.

The middle panels of Figures 6.1 and 6.2 present the azimuthal field, $V_\phi$. This model shows the largest difference between the two datasets. In Figure 5.3, we saw the region around $R = 11\text{ kpc}$ present an increase in the azimuthal velocity. The DR3 model of this region shows a slowly decreasing azimuthal velocity which is more closely aligned with the expected behaviour of a galactic rotation curve. The azimuthal speed has a higher dependence on $\phi$, presenting several local peaks along the region of highest $V_\phi$. The region follows radial curves reasonably well, with one notable difference at $Y = 0$ showing a thinner band than the rest of the model. Away from the midplane, the model shows the expected gradients towards lower azimuthal velocities.

The bottom panel of Figures 6.1 and 6.2 present $V_Z$. The vertical velocity shows an even pattern in the midplane with slowly varying velocity increasing with Galactic radius. The view in the $RZ$ plane presents a pattern of motion above and below the plane consistent
Figure 6.2: Left: Components of the velocity field in the $\phi = 0$ plane. Contours are placed at 0 km s$^{-1}$ for $V_R$ and $V_z$, and 200 km s$^{-1}$ for $V_\phi$. Right: One-sigma confidence regions of the GP models. Colormap units are in km s$^{-1}$. The position of the Sun is placed as a grey dot at (-8.27,0).
with DR2. The extension to a larger $Z$ confirms bending motions apparent in the disk and some breathing motions, though limited. The trends are supported by the outer disk analysis of McMillan et al. (2022), focusing on $V_Z$. They note an increasing trend in $V_Z$ with radius for young blue stars, additionally finding a discontinuity near $R = 11$ kpc when considering specific planes, in which the velocities show an increase of about 10 km s$^{-1}$. Such a discontinuity could not be explicitly found in a GP model. Instead, the model finds a large gradient over a short length that may be seen in the bottom panel of Figure 6.1 near $(0, -11)$ kpc, showing agreement with the found evidence of the warp. The edge of the $V_Z$ model in Figure 6.2 finds a change at $Z = -1.5$ kpc from slowly varying $V_Z$ to a sharp gradient. Five binned points at this edge show a large $V_Z$ that is inconsistent with the negative values of neighboring points. This influences the model to consider a larger mean at this edge and a correspondingly large uncertainty. More data is needed in this region as this feature is likely not a strong representation of the mean.

6.2 DR3 Velocity Vectors

We construct velocity vectorfields to visualize the complex motions modelled by GPR. In Figure 6.3 we present the velocity field in the midplane with vertical bulk motions shown as the background map. Velocity components are shown relative to a vector of 230 km s$^{-1}\hat{\phi}$ for clarity. Vectors paint a picture of several flows of radial and azimuthal motion within the bulk rotation of the disk. Beyond the solar circle, vectors follow two gradients, from larger to smaller azimuthal velocity as azimuth and radius increase, and from negative to positive radial velocity as radius increases. The gradient in azimuthal velocity with radius is the signal of the rotation curve, while the radial component and underlaid vertical component colormap point to disequilibrium in the disk. Within the solar circle, motions
6.2. DR3 VELOCITY VECTORS

Figure 6.3: Velocity vector map in midplane. Arrows represent the $V_X$ and $V_Y$ components with the vector 230 km s$^{-1}$ $\hat{\phi}$ subtracted for clarity. The background colormap shows $V_Z$ and the black contour indicates the curve where $V_Z = 0$. Arcs show the model for nearby spiral arms from Reid et al. (2019). In order from the inner to the outer disc they are: Scutum (cyan), Sagittarius (magenta), Local Arm (blue), and Perseus (black).

Vary significantly with azimuthal angle. Low azimuth finds motion away from the Galactic center within $R = 6$ kpc. This motion is the opposite at higher azimuth creating regions of converging and diverging mean fields. This will be explored further in Section 6.3.1.

In Figures 6.4 and 6.5 we present vectorfields in the $Y = 0$ kpc and $R = 8.27$ kpc planes. In these planes, the models show motions strongly consistent with those presented by DR2 data. In the $RZ$-plane, data extends beyond the DR2 model in radius, showing further confidence in bending modes found within the disk. The dominant feature in Figure 6.5 is the $V_\phi$ gradient with disk height that has been previously explored. The confidence in the velocities near the solar circle is high, leading to a very similar figure for the $\phi Z$-plane of both datasets.
6.2. DR3 VELOCITY VECTORS

Figure 6.4: Velocity vector map in the $RZ$-plane for $\phi = 0^\circ$ with arrows representing $V_R$ and $V_Z$. The background color map shows $V_\phi$ with the black contour indicating the curve where $V_\phi = 220 \text{ km s}^{-1}$.

Figure 6.5: Map of the velocity field in $\phi Z$-plane along a cylindrical surface at $R = 8.27 \text{ kpc}$. Arrows represent the components $V_\phi$ and $V_Z$ where, for clarity, we subtract $220 \text{ km s}^{-1}$ from $V_\phi$. The background colormap shows $V_R$ while black contour indicates the curve where $V_R = 0$. 
6.3. VELOCITY GRADIENT

Table 6.2: Comparison of measured values for the Oort constants from DR2 data, DR3 data, and literature in units of km s\(^{-1}\) kpc\(^{-1}\). For the GP model inferred values are provided at the position of the Sun (p) and average values over a spherical volume of radius 500 pc (v).

<table>
<thead>
<tr>
<th>Origin</th>
<th>(A)</th>
<th>(B)</th>
<th>(C)</th>
<th>(K)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vityazev (2018)</td>
<td>16.3 ± 0.1</td>
<td>-11.9 ± 0.1</td>
<td>-3.0 ± 0.1</td>
<td>-4.0 ± 0.2</td>
</tr>
<tr>
<td>Li (2019)</td>
<td>15.1 ± 0.1</td>
<td>-13.4 ± 0.1</td>
<td>-2.7 ± 0.1</td>
<td>-1.7 ± 0.2</td>
</tr>
<tr>
<td>Wang (2021)</td>
<td>16.3 ± 0.9</td>
<td>-12.0 ± 0.8</td>
<td>-3.1 ± 0.5</td>
<td>-1.3 ± 1.0</td>
</tr>
<tr>
<td>DR2 model (p)</td>
<td>16.2 ± 0.2</td>
<td>-11.7 ± 0.2</td>
<td>-3.1 ± 0.2</td>
<td>-3.0 ± 0.2</td>
</tr>
<tr>
<td>DR2 model (v)</td>
<td>15.2 ± 0.8</td>
<td>-12.4 ± 0.9</td>
<td>-2.9 ± 0.8</td>
<td>-2.3 ± 0.5</td>
</tr>
<tr>
<td>DR3 model (p)</td>
<td>15.2 ± 0.2</td>
<td>-12.2 ± 0.2</td>
<td>-4.0 ± 0.2</td>
<td>-4.6 ± 0.2</td>
</tr>
<tr>
<td>DR3 model (v)</td>
<td>15.3 ± 0.4</td>
<td>-12.3 ± 0.2</td>
<td>-3.5 ± 1.0</td>
<td>-3.9 ± 0.6</td>
</tr>
</tbody>
</table>

6.3 Velocity Gradient

6.3.1 Oort Constants and Functions

The continuous and differentiable GP solution allows inference of the gradient of the velocity field, the \(3 \times 3\)-tensor whose components are \(\partial V_i / \partial X_j\). Following Equations 5.21 - 5.26, Oort constants, Oort functions, the rotation curve, and angular velocity curve are calculated for the model.

Table 6.2 presents the Oort constants calculated from the DR3 model. For comparison to literature, both the prediction at a single point and the averaged value over a region are provided. The DR3 model constants agree strongly for \(A\) and \(B\) with both literature and with the DR2 model. Constants \(C\) and \(K\) from the DR3 model show stronger negative radial shear and negative divergence than their DR2 model counterparts, while \(K\) is in agreement with Vityazev et al. (2018). From Equations 5.24 - 5.25 we can see the sum of \(C\) and \(K\) is the derivative of the radial velocity with respect to radius, \(\partial V_R / \partial R\). The DR3 model observes a steeper gradient in the region of the Sun as it contains higher amplitude peaks and valleys in the radial velocity. The individual components may be viewed in Appendix...
6.3. VELOCITY GRADIENT

Figure 6.6: Galactic parameters as a function of radius. Top: Oort parameters $A$ (left) and $B$ (right). Bottom left: $V_\phi$. Bottom right: Angular velocity $\Omega = A - B$. The model is queried along the line passing from the Galactic centre through the Sun. The red dashed lines show the $1\sigma$ uncertainties.

B.1

The Oort constants are calculated along the length from the center of the Galaxy to the Sun and are shown in Figure 6.6. The top panels show constants $A$ and $B$ as a function of radius, with overall similar trends between DR2 and DR3. There is some variation, especially with constant $A$ finding a local peak near 9 kpc that is not present in the DR2 data. This difference may be a result of mapping a slightly different set of stellar populations due to observation restrictions of Gaia. The bottom panels present the rotation and angular velocity curves given by $V_\phi$ and $\Omega = |A - B|$. The extension of the model to larger distances
Figure 6.7: Oort constants $A$, $B$, $C$, and $K$ calculated from Equations 5.22 - 5.25. Colormap units are in km s$^{-1}$ kpc$^{-1}$. The spiral arms model of Reid et al. (2019) is also over-plotted for the local divergence ($K$). In order of inner to the outer disc the arms are: Scutum (cyan), Sagittarius (magenta), Local Arm (blue), and Perseus (black).

from the Sun gives a stronger model for the velocity curve compared to the larger errors found by the DR2 model.

The Oort constants may be evaluated over the region of the model as shown in Figure 6.7. From $A$ and $B$, we see lower amplitudes of azimuthal shear and vorticity at higher radius, owing to the gently varying rotation curve in this region. There is some azimuthal dependence, but this change is minimal. $C$ and $K$ find a general overlap in being positive and negative towards the outer disk. This is in contrast to the inner disk presenting largely opposite signs. Further investigation shows $\partial V_\phi/\partial \phi$ is much stronger in the inner disk,
while $\partial V_R/\partial R$ has larger amplitude in the outer disk. Beyond $R = 6$ kpc, features in $C$ and $K$ strongly follow the tilt of the overlaid spiral arms model of [Reid et al. (2019)], owing to that $\partial V_R/\partial R$ contribution. While this may be indicative of the spiral arms, the simulated motions of [Kawata et al. (2014)] and [Grand et al. (2015)] indicate these trends could be seen on a lengthscale of half the inter-arm spacing, while the lengthscale here is doubled.

### 6.3.2 Divergence

The full $3 \times 3$ set of gradients may be calculated to construct the divergence of stars in the disk, following the process in Section 5.6.3. The divergence is presented in Figure 6.8 in six viewing angles ranging from above and below the plane. The divergence with DR2 in Figure 5.10 found a volume with three main regions: a large central region around the Sun with negative divergence and two straddling smaller volumes of positive divergence. The DR3 model reveals this picture to be generally true, but with increased substructure in the disk. The region of negative divergence near the Sun extends further from the plane of the Galaxy towards $-\phi$ and beyond the solar circle. The regions of positive divergence are more numerous, but smaller and straddle the negative divergence region both above and below the plane. The divergence has similar amplitudes of $5 - 15 \text{ km s}^{-1} \text{ kpc}^{-1}$ compared to the $5 - 10 \text{ km s}^{-1} \text{ kpc}^{-1}$ discussed in Section 5.6.3. This applies the same $100 - 200$ Myr time-scale for the logarithm of the density to change.

In Figure 6.9 we present the divergence as a function of Galactocentric radius for 200 pc bands. The major features of Figure 6.8 present as $\pm 5 \text{ km s}^{-1} \text{ kpc}^{-1}$ variations. The regions of positive and negative divergence closely align with the radius of the spiral arms in this region, though the angle of the divergence regions following the spiral arms leads to some obscuring of the peaks in this azimuthal average. Considering divergence over the
6.3. VELOCITY GRADIENT

Figure 6.8: Divergence of the velocity field, calculated using cylindrical coordinates. Regions with $|\nabla \cdot \mathbf{V}| > 5 \, \text{km s}^{-1} \, \text{kpc}^{-1}$ are plotted, positive in red and negative in blue. Coloring and opacity is scaled with amplitude of the divergence, with a maximum around $15 \, \text{km s}^{-1} \, \text{kpc}^{-1}$.
6.4 DISCUSSION

Figure 6.9: Divergence of the velocity field with Galactocentric radius. The blue line presents the mean divergence over a 200 pc band in radius, and the shaded region is the standard distribution of the band. Azimuthal range of the bands varies with model coverage. The Reid et al. (2019) spiral arm radius between $\phi = \pm 15^\circ$ is plotted for each arm, in the same color convention.

whole model, a mean and standard deviation of $-0.20 \pm 5.86$ is found. Given a large area of the Galaxy, a total divergence of zero is expected. Smaller scale motion creates local regions of divergence, but the disk as a whole would obey continuity. A true measure with the continuity equation would require the density of the regions, but the divergence of the mean velocity field is a strong indicator.

6.4 Discussion

The extension to Gaia DR3 increases both the confidence in the model and provides improved predictions for the velocity field over a larger volume of the Galactic disk. The continuous GP model provides a link between the DF and Boltzmann equation to the observed data. Departures from equilibrium may be leveraged to infer the state of the stellar
The definite link between the velocity field and the potential requires the mass density. The presented GP models do not consider the stellar or mass density in modelling the bulk flows, though modelling $\rho$ for a tracer population and a specific selection function is certainly possible with available astrometric data. The DR2 and DR3 models present some differences. While showing improved velocity and Oort constant measurements, some of the difference may be due to slight changes in the represented population. Section 4.2.2 highlights the difference between the selected data and the magnitude limits that allow the DR3 data to include more stars. The application of a variety of selection functions based on HR diagram position, tracer populations, or age may extend the link between data and theory.
Chapter 7

Conclusion

The goal of understanding the motion of our Galaxy has made great progress. The Milky Way is a complex environment, and while in the past it has been modelled by simplistic axisymmetric or equilibrium potentials, stars experience many perturbing forces. Motions disrupt the disk from both within the Galaxy, such as the spiral arms, the influence of the bar, and close interaction with other stars, and beyond the plane of the Galaxy from Milky Way satellites, in-falling material, and the dark matter halo. Understanding the motion of our local patch of the Galaxy furthers the goal of understanding the Galactic potential and how it will evolve in the future.

In this thesis, we apply astrometric observations from the Gaia DR2 and DR3 surveys to model the mean velocity field of the local patch of the Galaxy. Modelling with Gaussian Process Regression creates a continuous and differentiable mean velocity field of the Galaxy’s motion, allowing for understanding the bulk motions of stars and derivatives of the velocity field. The models find a complex variety of disequilibrium motions including the warp, breathing modes, and radial patterns with a tilt similar to the spiral arms.

The nine derivative components allow comparisons with further kinematic parameters including the Oort constants. Kinematic parameters are in good agreement for the position of the Sun, and full 6-D astrometry allows the extension beyond previous models to map the full region. For the first time, the 3D divergence for the local Galaxy has been modelled,
identifying the negative divergence of the Solar region angled with the Local Arm. The divergence within the region provides a link to the time derivative of the density and evolution of the region.

7.1 Future Work

There are two main routes for expanding upon this work. Either the method of modelling can be altered to try to obtain better measures, or the data analyzed by the method can be changed for different scientific outcomes.

In this modelling method, the GPR treated the velocity vector to be uncorrelated between the \((V_R, V_\phi, V_z)\) components, allowing for a separation into a model field for each component. An alteration to GPR to consider direct fitting of 3-D vectors either through correlations between components or angle rotation of vectors could prove interesting. A restriction of the model is the limitation of model hyper-parameters being oriented with the input dimensions. A choice of reference frame is required, while certain Galactic structures may be better suited for modelling in a different frame. Allowing for variable hyper-parameter values through an altered method, such as layered GPR, could improve structure capture.

In this work, the first velocity moment of the distribution function was considered for the mean velocity field. GPR modelling may be applied to further tasks such as velocity variance over the disk for spiral arms signals, or the stellar density and mass could accompany the velocity field to give a measure of momentum in the disk. On a different avenue, applying further restrictions to the data selection can allow for populations of stars to be isolated. This may be done either by stellar age or by stellar type to analyze the dynamical relations within or between stellar age groups.

Lastly, the Gaussian Process model may be scaled up to larger computer systems. This
work was performed utilizing 16 GB of memory, and 6 – 10 hours of CPU time per model fit. Extending to 64 GB or 128 GB would allow the inclusion of more binned datapoints, further reducing any impact of binning on the model and allowing for an improved fit. These discussed extensions may answer further questions about the motion of the Galaxy, the gravitational potential, and the evolution we may see in the future.
Bibliography


Bennett M., Bovy J., 2019, *MNRAS* 482, 1417


Crowley C., et al., 2016, Astronomy & Astrophysics, 595, A6

Cui X.-Q., et al., 2012, Research in Astronomy and Astrophysics, 12, 1197


Fall S., ed. 1981, The structure and evolution of normal galaxies


Gaia Collaboration et al., 2016, Astronomy & Astrophysics 595, A1

Gaia Collaboration et al., 2018a, A&A 616, A1


Hobbs D., et al., 2022, Gaia DR3 documentation Chapter 4: Astrometric data, European Space Agency; Gaia Data Processing and Analysis Consortium


Li H., Fu C.-W., Li Y., Hanson A., 2007, IEEE transactions on visualization and computer graphics 13, 1640

Li C., Zhao G., Yang C., 2019, ApJ 872, 205


Miyamoto M., Nagai R., 1975, PASJ 27, 533


Ogorodnikoff K., 1932, Z. Astrophys., 4, 190


Sale S. E., Magorrian J., 2018, *MNRAS*, 481, 494


Tsvetkov A. S., Amosov F. A., 2019, Astronomy Letters, 45, 462

Vityazev V. V., Popov A. V., Tsvetkov A. S., Petrov S. D., Trofimov D. A., Kiyaev V. I., 2018, Astronomy Letters, 44, 236


Appendix A

Mathematical Formulae

A.1 The Marginal and Conditional Distributions of the MVN theorem

Suppose $X = (x_1, x_2)$ is a joint Gaussian with parameters

$$
\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \Lambda = \Sigma^{-1} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}, \quad (A.1)
$$

then the marginal distributions are given by

$$
p(x_1) = \mathcal{N}(x_1 | \mu_1, \Sigma_{11}) ,
$$

$$
p(x_2) = \mathcal{N}(x_2 | \mu_2, \Sigma_{22}) , \quad (A.2)
$$

and the posterior conditional is given by

$$
p(x_1|x_2) = \mathcal{N}(x_1 | \mu_{1|2}, \Sigma_{1|2})
$$

$$
\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2) = \mu_1 - \Lambda_{11}^{-1} \Lambda_{12} (x_2 - \mu_2) \quad (A.3)
$$

$$
\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \Lambda_{11}^{-1}
$$
Appendix B

Additional Figures
B.1 Velocity Derivatives

Figure B.1: Maps of the inplane partial derivatives for $V_R$ and $V_\phi$ with DR3 data. Used in calculating the Oort constants and Oort constant maps in Figures 6.6 and 6.7. Units are in km s$^{-1}$ kpc$^{-1}$. 
Figure B.2: $Z$ partial derivative of $V_Z$ for DR3 data. Presents the in-plane data for calculating the divergence in Figure 6.8. Units are in km s$^{-1}$ kpc$^{-1}$.