3D map of the dust distribution in the Milky Way

SARA REZAEIKHOSHBAKHT

Dissertation
submitted to the
Combined Faculties of the Natural Sciences and Mathematics
of the Ruperto-Carola-University of Heidelberg. Germany
for the degree of
Doctor of Natural Sciences

Put forward by
Sara Rezaeikhoshbakht (Rezaei Kh.)
born in: Tehran, Iran
Oral examination: 9 November 2018

3D map of the dust distribution in the Milky Way

Referees: Dr. Coryn Bailer-Jones

Prof. Dr. Hans-Walter Rix

Abstract

In this thesis, I present a new non-parametric model for inferring the three-dimensional (3D) distribution of dust density in the Milky Way. Our approach uses the extinction measured towards stars at different locations in the Galaxy at known distances. Each extinction measurement is proportional to the integrated dust density along its line of sight (l.o.s). Making simple assumptions about the spatial correlation of the dust density, we infer the most probable 3D distribution of dust across the entire observed region, including along sight lines which were not observed. This is possible because our model employs a Gaussian process to connect all l.o.s. The result is a smooth, 3D map of the dust density, which is the local property of the interstellar medium (ISM) rather than an integrated quantity. Owing to our smoothness constraint and its isotropy, the method provides one of the first maps without "fingers of God" artefact.

I then present the first continuous map of the dust distribution in the Galactic disk out to 7 kpc within 100 pc of the Galactic midplane, using red giant stars from SDSS APOGEE DR14. The resulting map traces some features of the local Galactic spiral arms, even though the model contains no prior suggestion of spiral arms, nor any underlying model for the Galactic structure. This is the first time that such evident arm structures have been captured by a dust density map in the Milky Way. Our resulting map also traces some of the known giant molecular clouds in the Galaxy and puts some constraints on their distances, some of which were hitherto relatively uncertain.

I also demonstrate a map of the 3D distribution of dust in the Orion complex. Orion is the closest site of high-mass star formation, making it an excellent laboratory for studying the ISM and star formation. We use data from the Gaia-TGAS catalogue combined with photometry from 2MASS and WISE to get the distances and extinctions of individual stars in the vicinity of the Orion complex. We find that the distance and depth of the cloud are compatible with other recent works, which show that the method can be applied to local molecular clouds to map their 3D dust distribution. We also use data from the recent second Gaia data release (GDR2) to update the map that shows complex dust clouds in the Orion region.

I finally show a 3D map of hydrogen density in the local ISM. The hydrogen equivalent column densities were obtained from the Exploring the X-ray Transient and variable Sky project (EXTRAS), which provides equivalent N_{H} values from X-ray spectral fits of observations within the XMM-Newton Data Release. A cross-correlation between the EXTRAS catalogue and the first Gaia Data Release was performed in order to obtain accurate parallax and distance measurements. The resulting map shows small-scale density structures which can not be modelled using analytic density profiles.

Zusammenfassung

Ich präsentiere ein neues nicht-parametrisches Modell zur Ableitung der dreidimensionalen (3D) Verteilung der Staubdichte in der Milchstraße. Wir verwenden die gemessene Extinktion zu Sternen an verschiedenen Orten der Galaxis in bekannten Entfernungen. Jede Extinktionsmessung ist proportional zur integrierten Staubdichte entlang ihrer Sichtlinie (l.o.s). Ausgehend von einfachen Annahmen über die räumliche Korrelation der Staubdichte unter Verwendung eines Gauß-Prozesses, schließen wir die wahrscheinlichste 3D-Verteilung von Staub über die gesamte beobachtete Region einschließlich entlang nicht sichtbarer Sichtlinien ab. Die resultierende Karte bietet eine glatte 3D-Karte der Staubdichte, die die lokale Eigenschaft des ISM ist und keine integrierte Größe darstellt. Aufgrund unserer Glättebeschränkung und ihrer Isotropie bietet die Methode eine der ersten Karten ohne "Finger Gottes" Artefakt.

Ich stelle dann die erste durchgehende Karte der Staubverteilung in der galaktischen Scheibe bis zu 7 kpc innerhalb von 100 pc der galaktischen Mittelebene vor und verwende rote Riesensterne von APOGEE DR14. Die resultierende Karte zeichnet einige Merkmale der lokalen galaktischen Spiralarme nach, obwohl das Modell keine vorherigen Hinweise auf Spiralarme enthä lt, noch ein zugrunde liegendes Modell für die galaktische Struktur. Dies ist das erste Mal, dass solche offensichtlichen Armstrukturen von einer Staubdichtekarte in der Milchstraße erfasst wurden. Unsere resultierende Karte zeichnet auch einige der bekannten riesigen Molekülwolken in der Galaxis nach und stellt einige Einschränkungen für ihre Entfernungen auf, von denen einige bisher relativ unsicher waren.

Ich zeige auch eine Karte der 3D-Verteilung von Staub im Orion-Komplex. Orion ist der nächstgelegene Ort der Sternentstehung mit hoher Masse, was es zu einem ausgezeichneten Labor für die Untersuchung der interstellaren Medium- und Sternentstehung macht. Wir verwenden Daten aus dem Gaia-TGAS-Katalog, kombiniert mit Photometrie von 2MASS und WISE, um die Entfernungen und Aussterben einzelner Sterne zu erhalten. Wir finden, dass die Entfernung und die Tiefe der Wolke mit anderen neueren Arbeiten kompatibel sind, was zeigt, dass die Methode auf lokale Molekülwolken anwendbar ist, um ihre Staubverteilung abzubilden. Wir verwenden auch Daten aus der jüngsten zweiten Gaia-Datenfreigabe, um die Karte zu aktualisieren, die komplexe Staubwolken in der Orion-Region zeigt.

Ich zeige schließlich eine 3D-Karte der Wasserstoffdichte im lokalen interstellaren Medium. Die wasserstoffäquivalenten Säulendichten wurden aus dem EXTRAS-Projekt erhalten, das äquivalente N_H -Werte aus Räntgenspektralanpassungen von Beobachtungen innerhalb der XMM-Newton-Datenfreigabe liefert. Eine Kreuzkorrelation zwischen dem EXTRAS-Katalog und der ersten Gaia Data Release gab uns genaue Parallaxen- und Abstandsmessungen. Die resultierende Karte zeigt kleinskalige Dichtestrukturen, die nicht mit analytischen Dichteprofilen modelliert werden können.

To my parents

for all your sacrifices

تقدیم به پدر و مادر عزیزم

بخاطر تمام از خود گذشتگی ها

Acknowledgments

This thesis represents not only my (one-handed) work on the keyboard, it is the result of four years of a great time at MPIA and Heidelberg. I hereby would like to express my gratitude to all the people who, one way or the other, supported me over the last four years of my life.

I would like to express my deepest appreciation to my supervisor, Coryn Bailer-Jones, who gave me the opportunity to experience an amazing PhD time with him, who guided me through every step from the beginning to the end; yet gave me enough freedom to redirect the project according to my interests. Words can not describe how grateful I am for your continued support and guidance.

I would also like to extend my deepest gratitude to Hans-Walter Rix, who has always been very supportive to make sure everything is in order for us to get our scientific work done. It has been a pleasure to be a member of the GC department at MPIA. I would also like to thank you for agreeing to referee my thesis.

I would like to extend my sincere thanks to David Hogg for his very fruitful comments and discussions. In particular, the improvement of the method was not possible without your help and ideas.

The coordination of the IMPRS students would not have been possible without the support and nurturing of Christian Fendt. I am extremely grateful to all your support and kindness; from organising all exciting social events to personal guidance through bureaucratic procedures in Heidelberg.

I am extremely grateful to Arjen van der Wel, for all his help and support over the past four years; as a great colleague, teammate, and friend who helped me in many aspects of my PhD. A wonderful friend who, together with his lovely family; Sharon, Thijs, and Kasper, made my life more joyful in Heidelberg. I owe you a world of gratitude.

I would like to extend my sincere thanks to all friends and colleagues at MPIA, in particular, Morgan Fouesneau, Rene Andrae, Gaia and Milky Way group members for

all the science we shared and useful comments and discussions. I would also like to thank my fellow IMPRS-14 generation with whom I shared a lot of memorable moments.

Many thanks to all my friends who supported me during the past four years. I am extremely grateful, in particular, to Reza Moetazedian and Bahar Bidaran for their very kind help during the especially hard last month I had due to the shoulder surgery. I could not pass through this stage without you guys; I owe you big time. My special gratitude also goes to María Jesús Jiménez Donaire, my beloved flatmate and friend with whom I spent most of my time (before she abandoned me ;)); life-lasting memories I will never forget, Golam. Another special thank goes to Chiara Mazzucchelli, who has always been a supportive friend, officemate and teammate. I will definitely miss a good friend when you move to Chile, but I wish you a great time. I am also extremely grateful to Anna Sippel for her unwavering help and support. Thank you for being my good friend / big sister. I would also like to thank Christina Eilers, Richard Teague, Gesa Bertrang, Iskren Georgiev, Michael Rugel, Robert Harris, Roberto Decarli, Faezeh Shabani, and many others who I can not fit in this little section but have always been great friends. My special gratitude also goes to the E's headquarters ("Saraane E-Neshast"); Ehsan Kourkchi, Maryam Saberi, and Farhang Habibi, for the very joyful and useful weekly Skype meetings ("E-Neshasts"): you are the best.

I am also grateful to Arjen van der Wel, Reza Moetazedian, Chiara Mazzucchelli, and María Jesús Jiménez Donaire for proofreading my thesis.

This journey would not have been possible without the great support of my family. I am deeply indebted to my parents and my beloved brothers, Saeed and Mohammad for the continued moral support and love they gave me. Being away from you has been the hardest part of my PhD life. Thanks for all your encouragement, kindness and sacrifices.

This project is partially funded by the Sonderforschungsbereich SFB 881 "The Milky Way System" of the German Research Foundation (DFG).
This work has made use of data from SDSS-IV. SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU) / University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatories of China, New Mexico State University, New York University, University of Notre Dame, Observatário Nacional / MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional

Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University. This work has made use of data from the European Space Agency (ESA) mission *Gaia* (https://www.cosmos.esa.int/gaia), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement.

Publication list

Refereed Journals:

- **Rezaei Kh., S.**, Bailer-Jones, C. A. L., Hanson, R. J., and Fouesneau, M. Inferring the three-dimensional distribution of dust in the Galaxy with a non-parametric method . Preparing for Gaia. A&A, 598:A125, February 2017. doi: 10.1051/ 0004-6361/201628885.

- **Rezaei, Kh. S.**, Bailer-Jones, C. A. L., Schlafly, E. F., and Fouesneau, M. Threedimensional dust mapping in the Orion complex, combining Gaia-TGAS, 2MASS, and WISE. A&A, 616:A44, August 2018. doi: 10.1051/0004-6361/201732503.

- **Rezaei Kh., S.**, Bailer-Jones, C. A. L., Hogg, D. W., and Schultheis, M. Detection of the Milky Way spiral arms in dust from 3D mapping. A&A in press. doi: 10.1051/0004-6361/201833284, ArXiv e-prints: 1808.00015, July 2018.

- Gatuzz, E., **Rezaei Kh., S.**, Kallman, T. R., Kreikenbohm, A., Oertel, M., Wilms, J., and Garc a, J. A. 3D mapping of the neutral X-ray absorption in the local interstellar medium: the Gaia and XMM-Newton synergy. MNRAS, 479:3715- 3725, September 2018. doi: 10.1093/MNRAS/sty1738.

Conference Proceedings:

- **Rezaei Kh., S.**, Bailer-Jones, C. A. L., Hanson, R. J. Getting ready for Gaia: three-dimensional modeling of dust in the Milky Way. Memorie della Societa Astronomica Italiana, v.86, p.642 (2015).

- **Rezaei Kh., S.**, Bailer-Jones, C. A. L., Fouesneau, M., Hanson, R. Can we detect Galactic spiral arms? 3D dust distribution in the Milky Way. Astrometry and Astrophysics in the Gaia sky, Proceedings of the International Astronomical Union, IAU Symposium, Volume 330, pp. 189-192 (2018).

Contents

Listing of Figures

Listing of tables

1

Introduction

Attempts to map our Milky Way date back to the 18th century. One of the most important works was by William Herschel, who constructed a map of the Milky Way by counting stars in more than 600 different lines of sight. He concluded that the Milky Way is a flattened disk and the Sun is located very close to the centre (Herschel, 1785). Jacobus Kapteyn improved Herschel's map using photometric star counts, parallaxes and proper motions of stars and estimated the size and shape of the Milky Way (Kapteyn, 1922). However, neither Kapteyn nor Herschel were aware of the importance of the extinction of light by interstellar dust, which resulted in erroneous estimates of the size and shape of the Galaxy. Robert Trumpler found the first evidence of interstellar reddening and demonstrated how significant the effect is on the brightness of distant objects (Trumpler, 1930). Even until now, almost 90 years later, the detailed structure of the Galaxy has remained literally obscured due to our location within the dusty disk.

In this chapter, I give a general introduction to dust and extinction. I start with the first observational pieces of evidence for the Galactic reddening, continue with the physical properties and formulations of dust extinction, what causes this obscuration and what we currently know about it. I finish with the current state-of-the-art Milky Way dust/extinction maps and how I have contributed to the field.

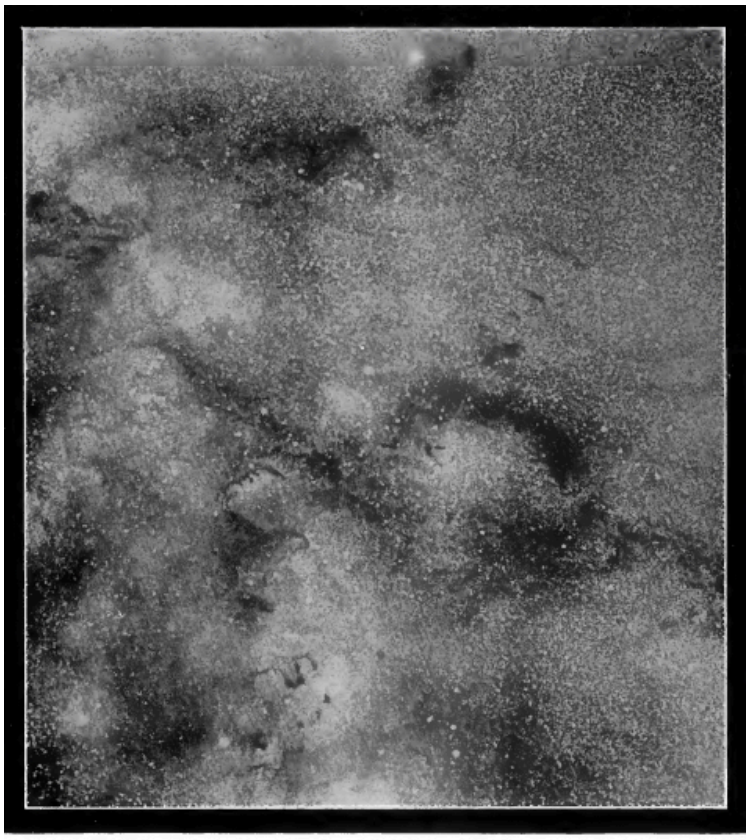
1.1 A walk through history

When William Herschel mapped the Milky Way in 1785, he noticed the existence of dark regions in the Galaxy. He mentions that while he was approaching the Milky Way [disk], the number of stars started to increase but all of a sudden they dropped down to nothing (Herschel, 1785). These regions were considered as "holes in the heavens". In 1847, Wilhelm Struve realised that the number of stars per unit volume decreases in all directions from the sun. This could be possible if the sun was located at the centre of the Galaxy, or if there exist some observational obscuration effects. He estimated that this obscuration has an effect of about 1 mag/kpc (Li and Greenberg, 2003). A similar effect was found by Jacobus C. Kapteyn who obtained a roughly spherical distribution of stars around the sun (Kapteyn, 1909).

By the early 20th century, Edward Barnard (Barnard, 1919) reported that there are actually obscuring bodies rather than holes in the sky, by taking amazing pictures of the clouds of various forms and structures (see Fig. 1.1). In 1921, some dark lanes were seen in the Milky Way that were believed by Heber D. Curtis to be caused by obscuring material (in a famous debate with Shapley, Shapley and Curtis, 1921). This obscuring dark region was then called "the zone of avoidance" which is our today's Galactic plane.

It was Robert J. Trumpler in 1930 who for the first time demonstrated the effects of interstellar reddening and the existence of the general and selective absorptions in the interstellar medium (ISM) (Trumpler, 1930). He studied open clusters in the Milky Way and compared their distances determined from their photometry and spectral types with the distances estimated from their angular diameters. He found that the estimated distances deviate systematically from one another as the distance increases. Since he assumed that the dimensions of open clusters should not depend on their distances from the sun, he concluded that "a general absorption is taking place within our stellar system" (Trumpler, 1930). He estimated the value of 0.7 mag/kpc for the general absorption and mentioned that it mainly affects stars in the Galactic plane. He also demonstrated that the absorption varies with wavelength and defined the colour excess as the difference between the "normal colour index" and the observed one and estimated the "selective absorption" of 0.3 mag/kpc between the photographic ($\lambda = 4300$) and visual ($\lambda = 5500$) bands.

Shortly after, Rudnick (1936) measured for the first time the wavelength dependency of the extinction in optical bands, via comparing pairs of obscured and non-obscured stellar spectra of the same spectral type. Further observations by Hall (1937) and



REGION NORTH OF THETA OPHIUCHI
$\alpha = 17^h\ 13^m$, $\delta = -21^\circ\ 0'$

Figure 1.1: Figure adopted from Barnard (1919); substructures that showed the presence of dark obscuring clouds (rather than holes in the sky) are clearly visible.

Stebbins et al. (1939) revealed the $1/\lambda$ reddening law using only $\lambda = 0.3 - 1\,\mu m$ wavelength range. In addition, in 1934, Paul W. Merrill discovered "undefined interstellar lines" (Merrill, 1934); widened absorption features now known as the "diffuse interstellar bands".

Having all the observational pieces of evidence, the subsequent question concerned the kind of material that can cause the extinction of about 1 mag/kpc, together with other observational effects that had been detected.

Since small metallic particles were perfect fits to the $1/\lambda$ extinction law, It was first thought that meteors are responsible for the interstellar reddening (schalén, 1929). It was believed that the metallic grains are produced by fragmentation of the small meteors or micrometeorites (e.g. Greenstein, 1937). However, it became clear later that meteors do not have an interstellar origin. In 1935, based on the hypothesis of particle growth by Arthur Eddington, Lindblad proposed that the dust clouds could form by the condensation of gas (Lindblad, 1935). Later on, Oort and van de Hulst expanded Lindblad's theory and proposed the "dirty ice" model for the interstellar dust. They used already existing information about the interstellar medium and surface chemistry and suggested that the dirty ice molecules are made out of existing atoms (H, O, C and H by then) in space, leading to the H2O, CH4, NH3 ices (Oort and van de Hulst, 1946).

In 1949, two astronomers, Hall (1949); Hiltner (1949), independently discovered the interstellar polarisation based on the prediction in 1946 by Chandrasekhar (Chandrasekhar, 1946). The discovery revealed the non-spherical nature of dust grains, and that they had to be globally aligned, likely by large-scale magnetic fields (Draine, 2003). By the discovery of the interstellar polarisation, the dirty ice model seemed to fail to explain it by that time which resulted in the reconsideration of metallic grains. Cayrel and Schatzman (1954) and Hoyle and Wickramasinghe (1962) proposed graphite as a dust component that can also explain the polarisation of the light, as a result of its anisotropic optical properties.

The development of Infrared astronomy made it finally possible to observe silicate particles in the atmosphere of M stars emitting at their characteristic 10 μm wavelengths (Woolf and Ney, 1969; Knacke et al., 1969). Nowadays we know that silicate exists everywhere in space; from interstellar clouds and circumstellar disks to different types of stars and cometary dust (Li and Greenberg, 2003).

1.2 Dust Extinction

One of the most well-studied properties of dust is its ability to extinguish the starlight passing through it; interstellar extinction. It is defined as the difference between the extinguished ($m(\lambda)$) and unextinguished ($m_0(\lambda)$) apparent magnitudes of stars at a given wavelength (λ):

$$A(\lambda) = m(\lambda) - m_0(\lambda)\,. \tag{1.1}$$

This can be written in terms of the apparent magnitude (m(λ)), absolute magnitude (M(λ)), and distance modulus (μ) as

$$A(\lambda) = m(\lambda) - M(\lambda) - \mu\,. \tag{1.2}$$

Defining the extinction in this way requires knowing the distance and absolute magnitude of stars. Instead, one can measure the distance-independent selective extinction or "colour excess" as

$$E(B-V) = A(B) - A(V) = (B-V) - (B-V)_0\,, \tag{1.3}$$

which is the excessive colour of stars due to the extinction ($(B-V)$ is the observed and $(B-V)_0$ is the intrinsic colour of a star).
Cardelli et al. (1989) defined the "normalised extinction" as

$$A(\lambda)/A(I) = a(\lambda) + b(\lambda)/R(V)\,, \tag{1.4}$$

and demonstrated that the normalised extinction as a function of wavelength can be parametrised as what is called an "interstellar extinction curve". The two parameters, a(λ) and b(λ), are fixed polynomials that determine the slope and curvature of the continuous extinction and the strength and shape of the 2175 Å bump. The interstellar extinction curve is an empirical relation measured classically by comparing the spectra of pairs of stars with the same spectral type, one in a dust-free region of the sky and another one behind a dusty region. Cardelli et al. (1989) found that the optical-UV extinction curves can be approximated by only one parameter, the total-to-selective extinction ratio:

$$R(V) \equiv A(V)/E(B-V). \tag{1.5}$$

The parametrisation of the interstellar extinction curve was developed by Fitzpatrick (1999) with a particular attention to the finite-width photometric passbands and is

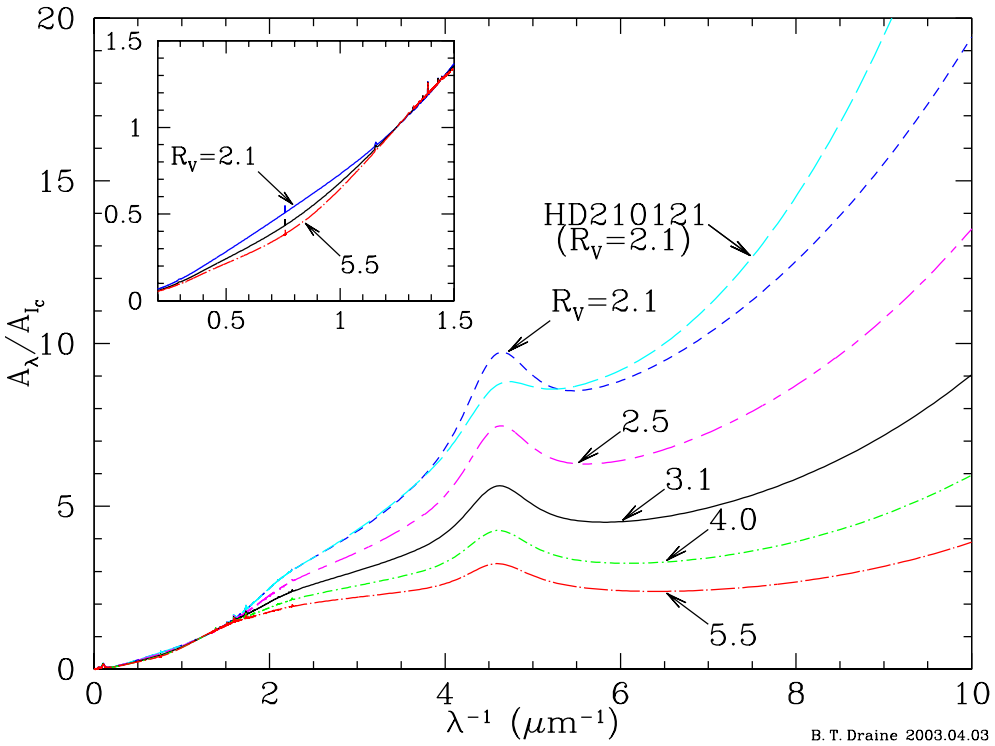


Figure 1.2: Interstellar extinction curves for different R(V) values (figure adopted from Draine, 2003). The inset is zoomed in at infrared wavelengths.

provided as a general recipe for the extinction curve determination. Figure 1.2 shows the extinction curves for different R(V) values using Fitzpatrick (1999) fitting formula (adopted from Draine, 2003). The interstellar extinction curves are flatter for higher R(V) values. Also, two different prescriptions for the extinction curves, Fitzpatrick (1999) vs. Cardelli et al. (1989), are compared for the same R(V) values, which shows significant variation in the UV extinction curve.

The value of R(V), consequently the shape of the extinction curve, depends on the environment and the properties of the dust grains: the larger dust grains provide larger R(V) values and flat, grey extinction curves in optical, a weaker 2175 Å bump and a flatter far-UV rise, relative to the smaller grains in more diffuse regions (Li and Greenberg, 2003; Schlafly et al., 2016, see figure 1.2). Mathis (1990) reported single values of R(V) for each of the diffuse and dense interstellar medium of 3.1 and 5 respectively. A recent study by Schlafly et al. (2016) on the optical-near infrared extinction curve confirmed the universality of the extinction curve, showing small deviation for measured R(V) for different sight lines. On the other hand, it showed

that variations in the R(V) in the Galactic plane occur in scales much larger than individual dust clouds, indicating that R(V) variations trace much more than just grain growth in the dense molecular clouds (Schlafly et al., 2016).

In general, from UV to infrared wavelengths, the extinction law follows $A(\lambda) \propto 1/\lambda$, although especially in the UV there are significant deviations. Apart from this general trend, there are remarkable features in the extinction curve, each of which determines the different physical and chemical properties of the dust grains. The most dominant feature is the 2175 Å UV bump that is thought to be driven by small particles rich in carbon, such as graphite and polycyclic aromatic hydrocarbon (PAH) molecules (organic molecules containing only carbon and hydrogen), as possible carriers of the 2175 Å bump (e.g. Li and Draine, 2001; Draine and Li, 2007; Mishra and Li, 2015). It is notable that there is a strong correlation between the 2175 Å bump and the metallicity, with the bump appearing weaker in the Large Magellanic Cloud (LMC) with lower metallicity compared to the Milky Way, and almost vanishing in case of the Small Magellanic Cloud (SMC) with much lower metallicity (Draine, 2003).

Other features in the extinction curve are the diffuse interstellar bands (DIBs), very weak, broad absorption features seen from UV to near infrared. Despite their very early detection in 1934 (Merrill, 1934), it was only in 2015 when $C_{60}{}^{+}$ was confirmed in the laboratory as the carrier of two of these bands (Campbell et al., 2015). The strength of the DIBs are correlated with dust extinction and can be used to infer the interstellar reddening in different sight lines (e.g. Capitanio et al., 2017). In the mid-infrared wavelength, from 3 to 19 μm, the extinction curve is dominated by the relatively broad PAH emission features. These features are present everywhere and are seen in the spectra of almost all objects, from HII regions to young stellar objects, planetary nebulae, AGB stars, and nuclei of galaxies (Tielens, 2008). There are also distinct silicate absorption features at 9.7 μm and 18 μm. At longer wavelengths, the extinction curve contains mainly emission from small, heated dust grains in the star-forming regions (Draine, 2003).

1.2.1 Correlation between extinction and hydrogen column density

The amount of extinction along a line of sight is strongly correlated with the hydrogen column density (Bohlin et al., 1978):

$$A(V)/N_{\rm H} \approx 5.35 \times 10^{-22}\,{\rm mag\,cm^2} \quad \text{for } R(V) = 3.1 \quad . \tag{1.6}$$

Fitzpatrick (1999) found

$$A(I_C)/N_{\rm H} \approx 2.96 \times 10^{-22}\,{\rm mag\,cm^2} \tag{1.7}$$

for the same R(V) value but computing the extinction in the Cousins I band ($\lambda = 0.802\mu m$, $A(I_C)/A(V) = 0.554$). Rachford et al. (2002) discovered that this ratio is not universal and varies with R(V) variations: the extinction per unit hydrogen column density increases with larger R(V) values. This correlation can be very helpful to get the total hydrogen column density using the measured extinction or vice versa. Hydrogen column density can be measured, for example, using HI or H_2 bands in the UV (e.g. Bohlin et al., 1978), or through X-ray absorption lines from particular Galactic sources (e.g. using XMM-Newton, Gatuzz and Churazov, 2018; Gatuzz, Rezaei Kh. et al., 2018, as will be seen in chapter 5).

1.3 Interstellar dust

The dust-to-gas ratio in the Milky Way is about 1% (e.g. Draine, 2003). The ISM itself contains about 10% of the baryonic mass of the Galaxy. Therefore dust makes up only a tiny fraction of the baryonic matter of the Galaxy, yet plays an important role. It absorbs around 30 - 50 % of the light emitted by the Galaxy at short wavelengths, scatters and re-emits it at long wavelengths, thereby only 0.1% of the baryonic matter is responsible for up to 50% of the Galactic radiation field (Li and Greenberg, 2003).

Apart from the observational effects, mentioned in the previous sections, we now realise that dust truly matters for the stellar and galactic formation and evolution. Intermediate-mass stars dredge up metals from their nuclei to the outer layers near the end of their lives. Dust particles are then formed as the atmosphere of the red giant and Asymptotic Giant Branch (AGB) stars are expelled by stellar winds and mass loss during the thermal pulse AGB (TP-AGB) phase. These dust particles grow further in the enriched ISM via grain-grain collisions in the denser environments or by accreting atoms and molecules onto their surfaces (e.g. Draine, 2009). They can then shield molecular hydrogen from destruction and catalyse its formation, allowing molecular clouds to form. Dust grains also cool down the hot molecular clouds via radiation, enabling them to collapse and form the next generation of stars (Tielens, 2010). Therefore, studying the three-dimensional dust distribution in the Galaxy provides valuable information about past and future star formation on large scales. Dust also plays a major role in planet formation: the remaining dust around the

new star forms a protoplanetary disk, then larger dust grains stick together, grow and eventually form planets. Not only dust grows in the ISM to form stars and planets, it can be destructed by high-energy UV radiations or due to supernovae blast waves (Draine, 2003). The timescale in which dust grains are being destroyed by supernovae shockwaves are orders of magnitude shorter than they are being formed in the atmosphere of cool stars and grow in the ISM; that is to admit that our current models for dust formation and growth are relatively incomplete (Draine, 2003).

Having access to observations in the UV and IR wavelengths, important facts about the nature of the dust particles revealed. Interstellar dust comprises a wide range of particle shapes, sizes and compositions: from large grains (about 10- 500 nm) of silicate and refractory mantles to the very small (about 1-10 nm) carbonaceous nano-grains, as well as a population of Polycyclic Aromatic Hydrocarbons (PAH) macromolecules. The effects of the dust grains of various shapes and sizes are evident through their interaction with the starlight. They extinguish starlight passing through them by the combination of absorption and scattering. The extinction efficiency of a dust grain is therefore divided into scattering and absorption terms:

$$Q_{\rm ext} = Q_{\rm scat} + Q_{\rm abs} \tag{1.8}$$

where $Q_{\rm ext}$ is the extinction efficiency, $Q_{\rm scat}$ the scattering efficiency, and $Q_{\rm abs}$ the absorption efficiency. The grain "albedo" is then defined as $\omega = Q_{\rm scat}/Q_{\rm ext}$, which is $\omega = 1$ for an idealised pure scattering grain, and $\omega = 0$ for a pure absorbing grain. For icy particles, scattering is the dominant factor, yet absorption is not exactly zero that heats up the grains making them emit some thermal emission. Metallic grains, on the other hand, are effective absorbers (van de Hulst, 1957).

Assuming only the scattering dominated regime (e.g. icy grains), there are two interesting cases:
The first one is regarding the long-wavelength regime where the sizes of the dust grains are much smaller compared to the wavelength of the incident light. This is the classical Rayleigh scattering for which

$$Q_{\rm scat} \propto \lambda^{-4}. \tag{1.9}$$

This means that shorter wavelengths are scattered much stronger than longer wavelengths (Rayleigh, 1919).
The second case is the short-wavelength regime where dust grains are larger in size

than the wavelength of radiating light, the so-called Mie scattering (Mie, 1908). A Mie scattering curve for a specific grain size rises steeply at long wavelengths and flattens out at short wavelengths.

If we now consider also the absorption term, at long wavelengths (i.e. small particles) absorption is more efficient than scattering and we have (by solving Maxwell's equations, Mie, 1908; van de Hulst, 1957)

$$Q_{\rm abs} \approx \lambda^{-1} \,. \tag{1.10}$$

Looking back at the interstellar extinction curve (section 1.2), the broadness of the curve illustrates a wide range of particle sizes, and the fact that the curve flattens out at short wavelengths suggests that there are more small grains than large grains in the ISM. The empirical interstellar extinction curve follows λ^{-1} relation, that, in the presence of small particles, is proven by the Mie theorem (Mie, 1908; Li and Greenberg, 2003).

In addition to absorption and scattering, the light of the bright stars is also reflected by the dust clouds located right behind them (reflection nebulae, Slipher, 1912; Hertzsprung, 1913; Hubble, 1922). Moreover, scattering off non-spherical particles polarises light. The grains are presumably aligned globally by the magnetic field (Li and Greenberg, 2003; Draine, 2003). Dust grains can also emit the electromagnetic radiation: they can be heated up in the ISM by absorption of a photon of starlight, or collisions with other atoms, dust grains, or cosmic rays. Once heated, dust grains need to cool to retain equilibrium. Radiative cooling is the main cooling process in the ISM (Draine, 2003). Their radiation is a strongly wavelength-dependent modified blackbody radiation. The normal size dust grains ($\sim$ 10 nm) have temperatures of around 30 – 50 K and radiate at far-infrared wavelengths up to about 100 μm. The tiny dust grains ($\sim$ 0.5 – 5 nm), on the other hand, can be heated up to temperatures of about 500 – 1000 K or more by absorbing only one UV photon, thus, they never achieve equilibrium with the radiation field (Draine, 2003). The tiny heated up grains in non-equilibrium emit at near- to mid-infrared wavelength regimes (1 – 30 μm).

1.4 Milky Way maps

Recognising the effects of interstellar dust on observations and ISM evolution, many attempts have been made by astronomers to map the extinction in the Milky Way. One of the most significant studies in this regard is the work by Schlegel, Finkbeiner,

and Davis (1998) who mapped the dust column density using far-infrared dust emission from the IRAS and COBE satellites. A more sensitive 2D map with higher resolution was made by Planck Collaboration et al. (2014) using a similar method to Schlegel, Finkbeiner, and Davis (1998). However, for many Galactic studies, 2D measurements do not suffice; we often need an estimate of the three-dimensional location of the emitting and/or extinguishing sources in the Galaxy. Moreover, knowing the local distribution of dust provides valuable information about the Galactic structure and the probable sites of star formation.

This opened a new area of studies in which various groups have been trying to map the Galactic dust extinction in 3D using different data sets and techniques. Marshall et al. (2006) presented a 3D extinction model in the Galactic plane using a Galactic model and the near-infrared colour excess to estimate distances and extinctions. Schlafly et al. (2010) used the blue tip of the distribution of stellar colours to measure the colour of the main sequence turnoff stars and measured the reddening of stars in the SDSS-III footprint. Sale (2012) developed a hierarchical Bayesian model to simultaneously infer extinction and stellar parameters from multi-band photometry, and Sale et al. (2014) used this method to build a 3D extinction map of the northern Galactic plane using IPHAS photometry. A similar probabilistic method was developed by Hanson and Bailer-Jones (2014) to estimate the effective temperature and extinction based on a method previously introduced by Bailer-Jones (2011). They used a Bayesian framework to account for the degeneracy between extinction and stellar effective temperature to produce a 3D extinction map of the Galactic high latitudes ($b > \sim 30°$) using Sloan Digital Sky Survey (SDSS) and The UKIRT Infrared Deep Sky Survey (UKIDSS). Hanson et al. (2016) then used photometry from the Panoramic Survey Telescope and Rapid Response System 1 (Pan-STARRS1) and Spitzer Glimpse surveys to map the dust extinction in the Galactic plane. Green et al. (2014) introduced a method similar to Sale (2012) to determine dust reddening from stellar photometry which was then used by Schlafly et al. (2014) to map the dust reddening of the entire sky north of declination $-30°$. Later on, this was used by Green et al. (2015) to build a 3D map of dust reddening for three-quarters of the sky using Pan-STARRS1 and the Two Micron All-Sky Survey (2MASS). Green et al. (2018) recently introduced an updated version of the map, taking advantage of a more accurate extinction law and additional new data from Pan-STARRS1.

The main drawback of these methods is that they treat each line of sight (l.o.s) independently from one another. This creates artefacts and discontinuities in their results (See Fig. 1.3, adopted from Green et al., 2014). Vergely et al. (2010) used

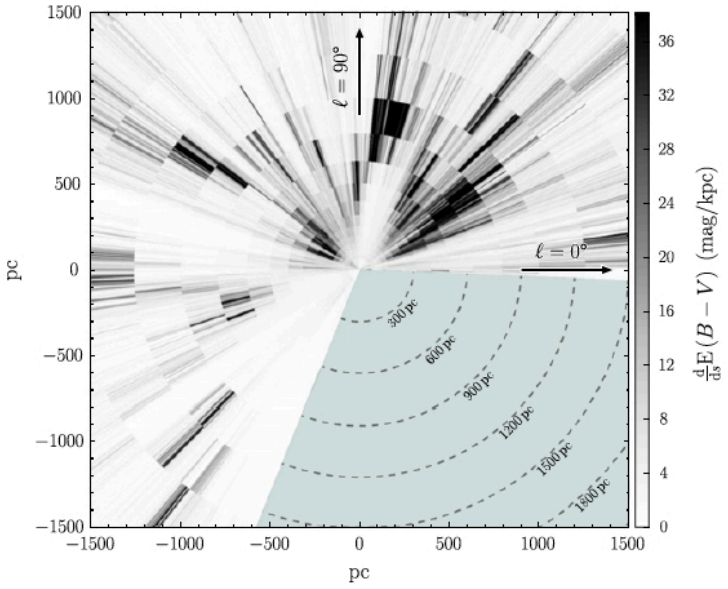


Figure 1.3: An example of the artefacts seen in the dust extinction maps due to the underlying assumptions. (Figure adopted from Green et al., 2014)

a method with a smoothing kernel to account for gaps in the data, and mapped the dust opacity (mag/pc) in the Sun's vicinity. A similar approach was taken by Lallement et al. (2014) who presented a 3D map of the local opacity. They later updated this map in Capitanio et al. (2017) using distance information from Gaia TGAS and colour excess estimates from diffuse interstellar bands (DIBs) from SDSS/APOGEE spectra, adopting a low-resolution map based on Pan-STARRS1 reddening measurements as a prior. Sale and Magorrian (2014) introduced a new method to map the Galactic extinction and dust in which the logarithm of the extinction (log A) is modelled as a Gaussian random field. Its covariance function has a Kolmogorov-like power spectrum which is motivated by a physical model of the interstellar medium. Dust is also modelled as a semi-stationary random field which produces a log A distribution that is very close to Gaussian.

Despite several attempts at 3D dust mapping, none of the aforementioned dust maps reveals the Galactic spiral arm structure. This is in contrast with the fact that spiral arms are rich in gas and dust where many stars are formed (as seen in external galaxies: Kennicutt et al., 2011; Schinnerer et al., 2017). Evidence for spiral structure in the Milky Way dates back to 1951 when W.W. Morgan and collaborators determined the distances towards emission regions (Oort and Muller,

1952; Morgan et al., 1953). This was confirmed shortly after by the discovery of 21 cm radio observations (van de Hulst et al., 1954; Morgan, 1955). The reason for this failure of extinction maps is mainly due to the lack of precise distance measurements as well as the assumptions behind dust mapping techniques. Most of these maps illustrate dust extinction or reddening which is an integrated property, and so cannot trace local properties of the Galaxy. In addition, many of these methods treat each l.o.s separately such that no information is propagated from neighbouring points, resulting in discontinuities between neighbouring lines of sight in the resulting maps.

1.5 Thesis overview

This thesis summarises the work I have done as part of my Doctoral degree on the topic of 3D map of the dust distribution in the Milky Way. In this thesis, I address the shortcomings of previous 3D dust extinction maps, mentioned in section 1.4, by using a non-parametric method to capture complex structures present in the observed data. Furthermore, the method directly maps dust density - which represents the local properties of the Galaxy - rather than the integrated extinction. The correlation between dust points in space is taken into account using an isotropic Gaussian process that provides a continuous map without l.o.s artefacts. I provide the first continuous 3D dust maps of the Milky Way, for the first time with the potential to detect the Galaxy's spiral arms.
I have developed a new method for mapping the dust distribution in the Milky Way using 3D positions of stars and their l.o.s extinctions. Technical details of the method are explained in chapter 2 and have been published in Rezaei Kh. et al. (2017) and Rezaei Kh. et al. (2018, b). In chapter 3, I illustrate the capability of the method in capturing arbitrary dust variations using both simulated and real datasets. This has been also part of the Rezaei Kh. et al. (2017) publication. In chapter 4, I summarise the work that I have done on the Galactic disk resulting in the first detection of the Milky Way spiral arms in the dust. This has been published in Rezaei Kh. et al. (2018, b). Chapter 5 represents the dust distribution in our local ($< 1kpc$) Galaxy. In the first section, 5.1, I concentrate on the Orion complex, mainly taking advantage of the Gaia data. Part of this, section 5.1.2, has been published in Rezaei et al. (2018, a). In the second section, 5.2, I present the 3D map of the local ($< 1kpc$) hydrogen density using X-ray data. This has been published in Gatuzz, Rezaei Kh. et al. (2018). The analysis of the X-ray spectra and derivation of the hydrogen column has been done by the co-authors and I have led the rest of the analysis.

Finally, in chapter 6, I summarise the main achievements during my PhD studies and how I have contributed to the field, and provide an outlook on how this can be developed and used in the future.

2

3D dust mapping technique

In this chapter, I explain in detail the new method of mapping the dust distribution in the Galaxy in 3D. The content of this chapter is adopted from Rezaei Kh. et al. (2017) and Rezaei Kh. et al. (2018, b).

The chapter is organised as follows. At the beginning, I introduce the problem we are about to tackle and describe the principles of our model. Then, I introduce our Gaussian process prior and show that the posterior distribution has an analytic solution. Finally, I explain how the dimensionality of the method can be decreased, resulting in a faster and more feasible method.

2.1 Problem setup

Our goal is to determine the 3D spatial distribution of dust given measurements of the l.o.s extinction caused by this dust toward a number of stars. Specifically, given these extinction measurements, we would like to find the probability distribution over the dust density at *any* point in space, and not necessarily a point along the l.o.s to one of these stars.

Let $\rho(\mathbf{r})$ be the dust density at vector position $\mathbf{r}$ measured from the observer. A model for the attenuation of starlight caused by this dust for a star at position $\mathbf{r}_n$

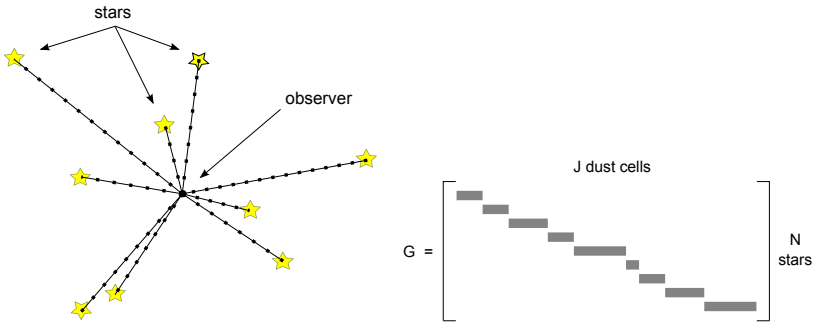


Figure 2.1: Dust cell geometry. Left: The l.o.s towards each of the N stars is divided up into a number of dust cells. The centre of each cell is shown as a small black square. (Here we show cells of constant size.) The total number of dust cells (towards all stars) is J, and the length of dust cell j towards star n is denoted $g_{n,j}$. Right: These cells are represented by a sparse matrix G of size $N \times J$. Each row has non-zero elements just for the cells along the l.o.s to that star.

is

$$f_n \propto \int_0^{r_n} \rho(\mathbf{r})\, dr \tag{2.1}$$

where $r = |\mathbf{r}|$. The principle of our method is to invert the above to get $\rho(\mathbf{r})$ for an arbitrary point in space given measurements of the attenuation towards multiple stars. If we adopt a parametric form for $\rho(\mathbf{r})$ then this is straightforward, but the result would be highly limited by the form adopted. Here we use a non-parametric model by dividing the pencil beam along the l.o.s toward each star into several dust cells, as shown in figure 2.1 (left). Let the (unknown) average dust density in cell j towards star n be $\rho_{n,j}$. The integral in equation 2.1 can then be replaced by a sum

$$f_n = \sum_j g_{n,j} \rho_{n,j} \ . \tag{2.2}$$

The attenuation is unitless. $g_{n,j}$ (the "geometric factor") is the length of the cell along the l.o.s. Thus $\rho_{n,j}$, which we think of as the "dust density", has units of attenuation per unit length. As we will use parsecs for distances, $\rho_{n,j}$ has units pc^{-1}. Let a_n be a measurement of the attenuation towards star n. Adopting a Gaussian noise model with standard deviation σ_n means that the probability of the measurements is

$$P(a_n|\{\rho_{n,j}\}) = \frac{1}{\sqrt{2\pi}\sigma_n} \exp\left[-\frac{1}{2\sigma_n^2}(a_n - f_n)^2\right] \tag{2.3}$$

where a_n is the measured attenuation and f_n is the model prediction of that, and $\{\rho_{n,j}\}$ denotes the mean densities of those cells on the l.o.s towards star n. If the non-extincted source intensity is I_0, then we assume that the observed intensity due to an attenuation a_n is

$$I \;=\; I_\circ \, e^{-a_n}. \tag{2.4}$$

This is related to the measured extinction A_n (in magnitudes) via the usual expression

$$A_n \;=\; -2.5 \log_{10} \left(\frac{I}{I_\circ} \right) \tag{2.5}$$

which gives $A_n \simeq 1.0857 a_n$.

Let us suppose that we measure the extinction toward N stars and use a total of J dust cells, for all stars. We define G as the $N \times J$ matrix with elements $g_{n,j}$, such that the n^{th} row of G contains the geometric factors just for star n. With this particular geometry this matrix is very sparse (see figure 2.1), because most row elements are zero (corresponding to the cells for all other stars), and each column has just one non-zero element (stars do not share l.o.s). Writing the set of dust densities in all cells (for all l.o.s) as the J-dimensional vector $\boldsymbol{\rho}_J$, and the model predictions for the attenuation towards the N stars as $\mathbf{f}_N$, we can write equation 2.2 as

$$\mathbf{f}_N \;=\; \mathrm{G} \boldsymbol{\rho}_J \; . \tag{2.6}$$

Writing the N attenuation measurements as the vector $\mathbf{a}_N$ with covariance V_N, then we can generalise equation 2.3 to be an N-dimensional Gaussian

$$\begin{aligned} P(\mathbf{a}_N | \boldsymbol{\rho}_J) \;=\; & \frac{1}{(2\pi)^{N/2} |\mathrm{V}_N|^{1/2}} \\ & \exp \left[-\frac{1}{2} (\mathbf{a}_N - \mathrm{G}\boldsymbol{\rho}_J)^{\mathsf{T}} \mathrm{V}_N^{-1} (\mathbf{a}_N - \mathrm{G}\boldsymbol{\rho}_J) \right] \; . \end{aligned} \tag{2.7}$$

The above equation is the likelihood: the probability of the data given the model parameters.

Our goal is to estimate the dust density ρ_{new} at an arbitrary point $\mathbf{r}_{J+1}$ in 3D space, given N measurements of the attenuation, $\mathbf{a}_N$, at known positions. Put probabilistically, we want to find $P(\rho_{new} | \mathbf{a}_N)$.

It should be noted that each element of $\boldsymbol{\rho}_J$ refers to the average dust density in the corresponding cell, although we can consider it to be the dust density at the centre

of the cell. ρ_{new}, in contrast, is the density at the *point* $\mathbf{r}_{J+1}$. There is no concept of a cell for points where we want to predict the dust density.
We have $J \geq N$ and $N \gg 1$. A typical problem may involve $N = 10^4$ and $J = 10^5$ (of the order of ten cells per star on average). In order to infer the dust densities, we need to introduce some connection between the l.o.s, otherwise we just have N independent equations like equation 2.2 with J unknowns, which would be insoluble.

2.2 Gaussian process prior

We connect the l.o.s using a Gaussian process (e.g. Gibbs and MacKay, 1997; Rasmussen and Williams, 2006). This states that the joint probability distribution of the dust density at any J different points (or cell centres) is a J-dimensional Gaussian, with a covariance matrix, C_J, which depends on the distance between the points (or cell centres), i.e.

$$P(\boldsymbol{\rho}_J) \;=\; \frac{1}{(2\pi)^{J/2}|\mathrm{C}_J|^{1/2}} \exp\left[-\frac{1}{2}\boldsymbol{\rho}_J^{\mathsf{T}}\mathrm{C}_J^{-1}\boldsymbol{\rho}_J\right] \;. \tag{2.8}$$

We assume a zero mean Gaussian in order to have zero values for the dust density in regions where we do not have any constraints from the data. A useful property of Gaussian processes is that the conditional distribution, $P(\rho_{new}|\boldsymbol{\rho}_J)$, is also Gaussian.
A Gaussian process is just a way of specifying a prior on the covariances between points, as opposed to specifying the functional form of the dust variation in physical space (which is what a parametric model would usually do). This permits a much wider form of functional variations than a parametric model.
An important aspect of the Gaussian processes is to choose an appropriate covariance function. This determines the elements, $c_{i,j}$, of the covariance matrix between two points (or cells) i and j, with position vectors $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively. Here we use a covariance function from Gneiting (2002)

$$c_{i,j} = \begin{cases} \theta\,(1+t^{\alpha})^{-3}\left[(1-t)\cos(\pi t) + \frac{1}{\pi}\sin(\pi t)\right] & \text{if} \quad 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases} \tag{2.9}$$

where

$$t \;=\; \frac{|\mathbf{r}_i - \mathbf{r}_j|}{\lambda} \quad \text{and} \quad \theta > 0\,,\, \lambda > 0\;.$$

which we illustrate in figure 2.2. We use $\alpha = 1$ (the solid line). A larger value of α (e.g. $\alpha = 2$, which has zero gradient at zero separation) produces functions with a

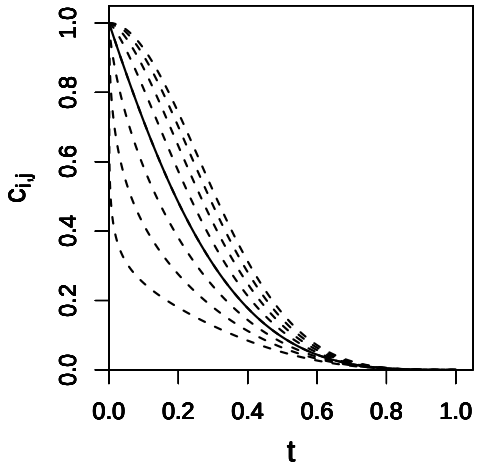


Figure 2.2: Covariance function in equation 2.9. The solid line is the case of $\alpha = 1$ (used in this work) and the dashed lines, from bottom to top, are for different values of α from 0.25, 0.5, 0.75, 1.25, 1.5, 1.75, to 2.

smoother spatial variation (plots for $\alpha = 1$ are shown later). The covariance drops monotonically as the separation between the points increases, to a value of zero once the points are separated by more than λ (the scale length). Note, however, that the covariance already drops to half its maximum at $t = 0.2$, so the effective correlation distance is much less than λ. The hyperparameter, θ, in the covariance function determines the overall scale of variations of the dust density. Here we consider the two hyperparameters, λ and θ, to be fixed (see chapter 3), although they can be inferred from the data.

This specific form of the covariance function does not have a particular physical motivation, apart from the fact that closer points have higher correlation, but it has compact support, which means it goes exactly to zero beyond some distance: two points separated by more than λ will not influence each other. While other covariance functions such as an exponential will give negligible covariance beyond many scale lengths, a truncated covariance function has the advantage of giving rise to sparse covariance matrices, which reduces memory use and accelerates computations. It is important to note that although one can easily write down any function which is truncated, this is not sufficient for it to be a *covariance* function (see Rasmussen and Williams, 2006).

We can now use this Gaussian process prior together with the likelihood (equation

2.7) to determine $P(\rho_{new}|\mathbf{a}_N)$. This will result in estimating $J+1$ parameters from N measurements. As $J \geq N$, this means that the resulting density estimates will not be independent. This is of course the whole point: to introduce correlations between the dust cells to make the problem tractable and – more significantly – to allow us to infer a PDF over the dust density at unobserved points.

Figure 2.3 shows samples drawn from the prior for different values of the hyper-parameters. To make this we define a 1D grid of 1000 equally-spaced points from $r = 0$ to $r = 5000pc$. One draw from this 1000-dimensional Gaussian gives us 1000 points which are plotted at their respective positions in space, and then connected with a line. Here, for each fixed pair of λ and θ, we draw two samples from the prior (red and black lines). The sharpness of the function is due to the shape of our covariance function (Fig. 2.2) which allows for sharper fluctuations. As mentioned earlier, a larger value of α in equation 2.9 produces smoother variations. It is clear that higher values of θ result in sharper and larger amplitude variations. Larger values of λ produce smoother variations.

2.3 Analytic solution

Using the law of marginalization over each l.o.s towards observed stars and then applying Bayes theorem, we can write the posterior PDF of the dust density at a given point given the data as

$$\begin{aligned} P(\rho_{new}|\mathbf{a}_N) &= \int P(\rho_{new}, \boldsymbol{\rho}_J|\mathbf{a}_N)\, d\boldsymbol{\rho}_J \\ &= \int \frac{P(\rho_{new}, \boldsymbol{\rho}_J)P(\mathbf{a}_N|\rho_{new}, \boldsymbol{\rho}_J)}{P(\mathbf{a}_N)}\, d\boldsymbol{\rho}_J \\ &= \frac{1}{P(\mathbf{a}_N)} \int P(\rho_{new}, \boldsymbol{\rho}_J)P(\mathbf{a}_N|\boldsymbol{\rho}_J)\, d\boldsymbol{\rho}_J \end{aligned} \tag{2.10}$$

where in the last line we use the fact that $\mathbf{a}_N$ is independent of ρ_{new} once conditioned on $\boldsymbol{\rho}_J$. This is a J-dimensional integral evaluated over all values of each component of $\boldsymbol{\rho}_J$. The term outside the integral is independent of ρ_{new} so is just part of the normalization constant. The first term under the integral is the Gaussian process prior (equation 2.8), now in $J+1$ dimensions. The second term is the likelihood (equation 2.7). Both are Gaussians, but not in ρ_{new}. Yet because their arguments

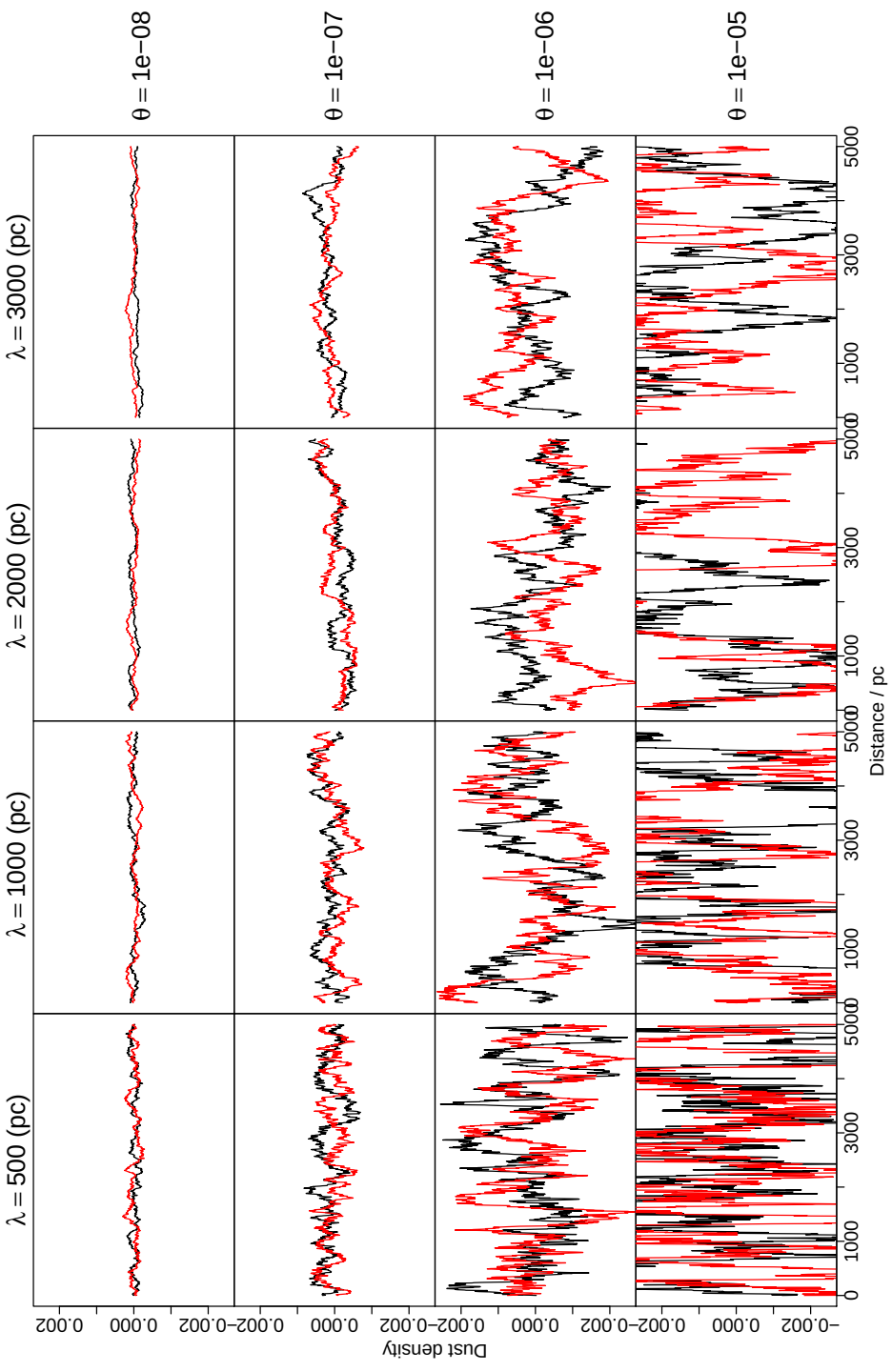


Figure 2.3: Dust density (attenuation per parsec) drawn from the prior with different hyperparameters of the covariance function (equation 2.9). We define a 1D grid of 1000 equally-spaced points from $r = 0$ to $r = 5000pc$ and each red and black line shows one draw from a 1000-dimensional Gaussian for a fixed λ and θ.

are linear functions of ρ_{new}, the integral will have an analytic solution. Let

$$\boldsymbol{x}_{N+1} = \begin{bmatrix} \boldsymbol{\rho}_J \\ \rho_{new} \end{bmatrix} \tag{2.11}$$

be the concatenation of the J dust densities with the dust density at the new point. Denote its covariance matrix as Ω_{N+1}. The distribution of $\boldsymbol{x}_{N+1}$ follows equation 2.8 with $J \to J+1$. We partition the inverse of this covariance matrix by writing it as

$$\Omega_{N+1}^{-1} = \begin{bmatrix} \mathrm{Q}_N & \mathbf{q}_N \\ \mathbf{q}_N^\mathsf{T} & \mu \end{bmatrix} \tag{2.12}$$

where Q_N is a $J \times J$ matrix, $\mathbf{q}_N$ is a $J \times 1$ vector, and μ is a scalar. We show in appendix A that the result of the integration is a Gaussian with mean $-\beta/\alpha$ and variance $1/\alpha$

$$P(\rho_{new}|\mathbf{a}_N) = \sqrt{\frac{\alpha}{2\pi}} \exp\left[-\frac{\alpha}{2}\left(\rho_{new} + \frac{\beta}{\alpha}\right)^2\right] \tag{2.13}$$

where

$$\begin{aligned} \alpha &= \mu - \mathbf{q}_N^\mathsf{T} \mathrm{R}_N^{-1} \mathbf{q}_N \\ \beta &= \mathbf{a}_N^\mathsf{T} \mathrm{V}_N^{-1} \mathrm{G} \mathrm{R}_N^{-1} \mathbf{q}_N \\ \mathrm{R}_N &= \mathrm{Q}_N + \mathrm{G}^\mathsf{T} \mathrm{V}_N^{-1} \mathrm{G} \ . \end{aligned} \tag{2.14}$$

In order to calculate the one-dimensional PDF over any single ρ_{new} at position $\mathbf{r}_{new}$, we must invert several matrices of size J, and this takes time $\mathcal{O}(J^n)$, where $n \lesssim 3$ for exact matrix inversion. The calculations can, however, be accelerated using certain matrix identities, as described in appendix A.

2.3.1 Decreasing the dimensionality

As discussed earlier (section 2.3), inferring the dust densities involves a J×J matrix inversion, followed by a set of J-dimensional matrix manipulations, which makes the calculations computationally expensive, even if we use accelerating techniques. The time-consuming part is the (one-off) inversion of the $J \times J$ covariance matrix C_J, which takes time $\mathcal{O}(J^n)$ to compute, where n is typically $\lesssim 3$ but can be reduced to around 2.3 (Demmel et al., 2007), as well as various matrix inversions and multiplications taking time $\mathcal{O}(NJ^2)$, which must be done for every prediction. For a problem with $N = 230$ and $J = 3203$, inverting C_J took two minutes (using a

single core on a modest AMD Opteron 6380 CPU). Making predictions at multiple points can then be done in parallel: 200 predictions took 4 minutes with 40 cores, or 1.2 seconds per point. The computation time for more points is proportional to the number of points. For a problem with $N = 1000$ and $J = 8185$, it took 40 minutes to invert C_J, and 34 seconds per point to make new predictions for 1000 new points (again with 40 cores). This is 28 times longer than the previous case, which agrees reasonably well with the $\mathcal{O}(NJ^2)$ scaling suggested above (which gives $(1000 \times 8185^2)/(200 \times 3203^2) = 33$).

The limiting factor when scaling this up to larger applications is the memory rather than the run-time. For the case of $J = 12\,000$, we needed 8 GB of RAM per core. This number is determined primarily by the number of cells, J, because the largest matrix has size $J \times J$. However, as we use sparse matrix methods and a truncated covariance function, the RAM required will not continue to grow as J^2. It is rather the density of cells in space, rather than the number of cells, which will ultimately drive the memory requirements. Using the run-time numbers from above, and ignoring memory limitations, then with N=10 000 and J=100 000, the C_J inversion takes around 30 days (with just one core; this could be accelerated if C_J inversion is parallelized too). Then even with 10 000 cores running for 30 days we could only make predictions at 30 000 points. This (and N) is too small to build up a useful dust density map over a large volume of space.

In order to accelerate the computations, we need to decrease the dimensionality of the problem. We introduce a new prior as

$$P(\mathrm{G}\boldsymbol{\rho}_J) \;=\; \frac{1}{(2\pi)^{J/2}|\mathrm{G}\mathrm{C}_J\mathrm{G}^\mathsf{T}|^{1/2}} \exp\left[X\right] \;,$$
$$X \;=\; -\frac{1}{2}(\mathrm{G}\boldsymbol{\rho}_J - \mathrm{G}\boldsymbol{\rho}_\mu)^\mathsf{T}(\mathrm{G}\mathrm{C}_J\mathrm{G}^\mathsf{T})^{-1}(\mathrm{G}\boldsymbol{\rho}_J - \mathrm{G}\boldsymbol{\rho}_\mu) \;. \quad (2.15)$$

This is still a Gaussian but now in $\mathrm{G}\boldsymbol{\rho}_J$. This has the advantage of a dramatic drop in its dimensionality from J to N, where N is the total number of stars in the sample and G is an $N \times J$ matrix containing geometric factors of the cells. The $J \times J$ matrix C_J needs to be calculated once, and $\mathrm{G}\mathrm{C}_J\mathrm{G}^\mathsf{T}$ is an $N \times N$ matrix which needs to be inverted also once. Note that the matrix C_J that is calculated first is built based on the distances between J dust cells, then it is multiplied by the matrix G. We also allow for a non-zero mean in the Gaussian process prior ($\mathrm{G}\boldsymbol{\rho}_\mu$). As described later, this is determined from a global property of the input data.

The likelihood function needs to be changed accordingly too:

$$P(\mathbf{a}_N|\mathrm{G}\boldsymbol{\rho}_J) = \frac{1}{(2\pi)^{N/2}|\mathrm{V}_N|^{1/2}} \times \exp\left[-\frac{1}{2}(\mathbf{a}_N - \mathrm{G}\boldsymbol{\rho}_J)^{\mathsf{T}}\mathrm{V}_N^{-1}(\mathbf{a}_N - \mathrm{G}\boldsymbol{\rho}_J)\right] , \tag{2.16}$$

where $\mathbf{a}_N$ is the vector of attenuation measurements with covariance V_N. Note that the right hand side of the new likelihood is identical to the previous likelihood (equation 2.7), since we assume the positions of input stars, consequently the geometric matrix, G, are fixed.

The posterior PDF of the dust density at a new point is calculated by multiplying the Gaussian process prior (equation 2.15) by the likelihood (equation 2.16) and marginalising over the $\boldsymbol{\rho}_J$ to give

$$P(\rho_{new}|\mathbf{a}_N) = \sqrt{\frac{\alpha}{2\pi}}\exp\left[-\frac{\alpha}{2}\left(\rho_{new} + \frac{\beta}{\alpha} - \rho_\mu\right)^2\right] \tag{2.17}$$

which is a Gaussian with mean $-\beta/\alpha + \rho_\mu$ and variance $1/\alpha$ where

$$\begin{aligned} \alpha &= q - \mathbf{q}_N^{\mathsf{T}}\mathrm{R}_N^{-1}\mathbf{q}_N \\ \beta &= \mathbf{a}_N^{\mathsf{T}}\mathrm{V}_N^{-1}\mathrm{R}_N^{-1}\mathbf{q}_N + \mathbf{q}_N^{\mathsf{T}}\mathrm{R}_N^{-1}\mathrm{Q}_N^{\mathsf{T}}\mathrm{G}\boldsymbol{\rho}_\mu - \mathbf{q}_N^{\mathsf{T}}\mathrm{G}\boldsymbol{\rho}_\mu \\ \mathrm{R}_N &= \mathrm{Q}_N + \mathrm{V}_N^{-1} \end{aligned} \tag{2.18}$$

and Q_N ($N\times N$ matrix), $\mathbf{q}_N$ ($N\times 1$ vector) and q (scalar) are the partitioned elements of matrix $(\mathrm{GC}_J\mathrm{G}^{\mathsf{T}})^{-1}$ (see appendix A for the derivations).

This leads to a dramatic computational gain, both in terms of speed and the maximum J which can be handled, while providing results identical to using the previous J-dimensional Gaussian process (equation 2.8). For instance, adopting the same J in both cases, the improved method runs about 1000 times faster.

It should be noted that the Gaussian model allows the dust density, ρ, to be negative. This is indeed non-physical but is due to the fact that the data is by nature noisy. Both extinctions and distances are measurements which contain uncertainties. Having inconsistent distance-extinction estimates and underestimated uncertainties would cause more/larger negative predictions. We will later see that in the presence of good data our posterior is mostly determined by the likelihood rather than the prior which results in mostly positive predictions. We will explore this later with real datasets (see chapters 4 and 5).

2.3.2 Distance uncertainty

While extinction uncertainties are taken into account in the likelihood (covariance V_N), distance uncertainties are not considered as we assume fixed positions for input stars and corresponding dust cells. Yet we can propagate the distance uncertainty into the extinction uncertainty. To get a first order approximation, we assume a constant dust density along the line of sight to each star. This way we have

$$\begin{aligned} a &= \rho r \\ \sigma_{a_d} &= \frac{a}{r}\sigma_r \end{aligned} \tag{2.19}$$

where a is the attenuation, ρ is a constant dust density along the l.o.s, σ_{a_d} is the uncertainty in attenuation propagated from the distance uncertainty, and σ_r is the distance uncertainty. Then the total input attenuation uncertainty in the model ($\sigma_{a_{tot}}$) is

$$\sigma_{a_{tot}} = ({\sigma_{a_d}}^2 + {\sigma_a}^2)^{1/2} \tag{2.20}$$

where σ_a is the measured attenuation uncertainty.
This way, both uncertainties in the input data, i.e. extinction uncertainty and distance uncertainty, are taken into account to infer the 3D distribution of the dust density.

3

Demonstration of the metod

In this chapter, I illustrate how the method works and show the capability of the method in capturing arbitrary dust variation using both simulated and real data. This chapter is based on the work published in Rezaei Kh. et al. (2017).
The chapter is organised as follows. First, using simple mock data, I show how the method works. Afterwards, I represent a 3D dust map using a more complex simulated data from the Gaia Universe Model Snapshot (GUMS). I then finish with showing the results on a real dataset.

3.1 Simple mock data

We first investigate the ability of our model to capture structures in the dust distribution and to infer it in unobserved regions. We further investigate the influence of the two model hyperparameters, θ and λ, on the dust density models produced. Figure 3.1 illustrates the simulation set up, which specifies the distribution of dust and observations of stars in the Galactic plane. There is a general distribution of dust, the density of which decreases exponentially from the Galactic Center in all directions with a length scale of 1 kpc. Observations are made in region 1 (a 12° wedge between longitudes 354° and 6°) and region 3 (a 2° wedge between

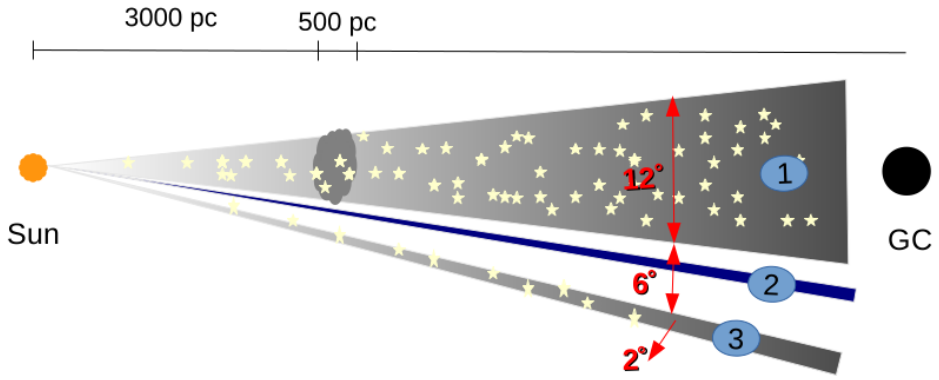


Figure 3.1: Illustration of the simple mock data in the Galactic plane. The true dust density decreases exponentially from the Galactic center (GC) in all directions. The observer is at the Sun. Region 1: in addition to the main dust variations there is a cloud between 3 and 3.5 kpc. 200 stars are observed here. Region 2: no stars are observed. Region 3: 100 stars are observed here (there is no cloud). Stars are drawn from throughout regions 1 and 3, their attenuations calculated, and observational noise added.

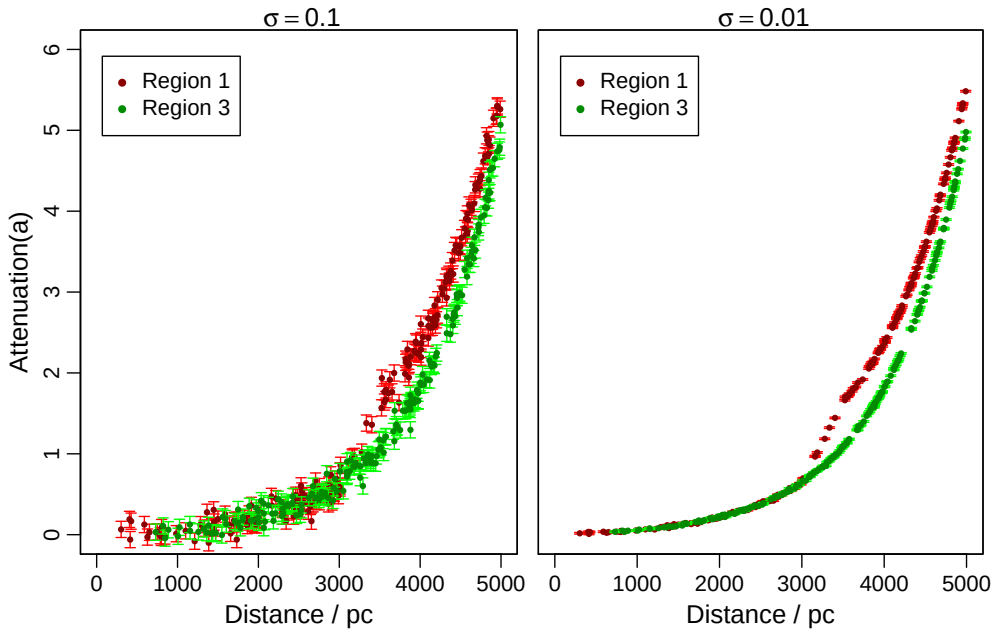


Figure 3.2: Attenuation in the simple mock data set as a function of distance for two different attenuation uncertainties (σ_n in equation 2.3) of 0.1 (left) and 0.01 (right) (no units). An increase in attenuation where the cloud is located (3 kpc to 3.5 kpc) is evident at the smaller noise level (right panel). It should be noted that attenuation is unitless (see equation 2.2).

longitudes 12° and 14°), but not in region 2, a narrow wedge ($l = 7° - 8°$) which lies between them. In region 1 there is an additional dust cloud of 500 pc depth centered on a distance of 3.25 kpc from the Sun. This does not extend into region 3 (whether it extends into region 2 in "reality" is immaterial as we have no observations there). We want to infer the distribution of the dust density over all three regions using measurements of the attenuations towards 200 stars spread uniformly over region 1, and of 100 stars spread uniformly over region 3. Two different sets of noisy measurements are considered: standard deviations (σ_n in equation 2.3) of 0.1 and 0.01 on the attenuations. Figure 3.2 shows the noisy input data for these two situations. In region 3, the dust density increases roughly exponentially with increasing distance from the Sun. The attenuation is the integral of this, so is more or less exponential too. Region 1 is similar, but has the dust cloud in addition. This can be made out reasonably well in the data at higher signal-to-noise (right panel), but is barely noticeable at the lower signal-to-noise (left panel).

We use our model to estimate the dust density at 100 points distributed at random at each of the three regions. We set the hyperparameters to $\theta = 10^{-7}$ and $\lambda = 2\,\text{kpc}$ (equation 2.9) and use uniform cell sizes of size $g = 250\,\text{pc}$. The choice of these values will be discussed later in this section. Figure 3.3 shows (as blue points) the estimated dust densities as a function of distance, as well as the uncertainty on this estimate (as error bars): these are the mean and standard deviation of the Gaussian posterior in equation 2.17. These predictions can be compared to the true values for regions 1 and 3, which are shown as red and green crosses respectively. We see that the inference of the overall exponentially-varying dust is good in all three regions, including in region 2, where there were no observations. This shows that our model performs sensible, plausible interpolations across unobserved regions. This is possible because of the smoothness prior imposed by the Gaussian prior. For region 1 (top row), the model predicts the location and density of the dust cloud well, even at the lower signal-to-noise ratio. In that case the estimated uncertainties (error bars) are also larger, which is what we want from a model. In region 2 (middle row), we ask the model to predict dust densities along a very narrow wedge between $l = 7°$ and $l = 8°$, located between regions 1 and 3 but closer to region 1 (with cloud) than region 3 (without cloud). With $\sigma = 0.01$ (right), an increase is obvious at 3 to 3.5 kpc corresponding to the distance of the dust cloud in region 1. A smaller increase is visible for the larger noise case (left). This too is a sensible interpolation of the available data: the cloud must stop or peter out somewhere between region 1 and region 3 because it is no longer observed in region 3. We have no information

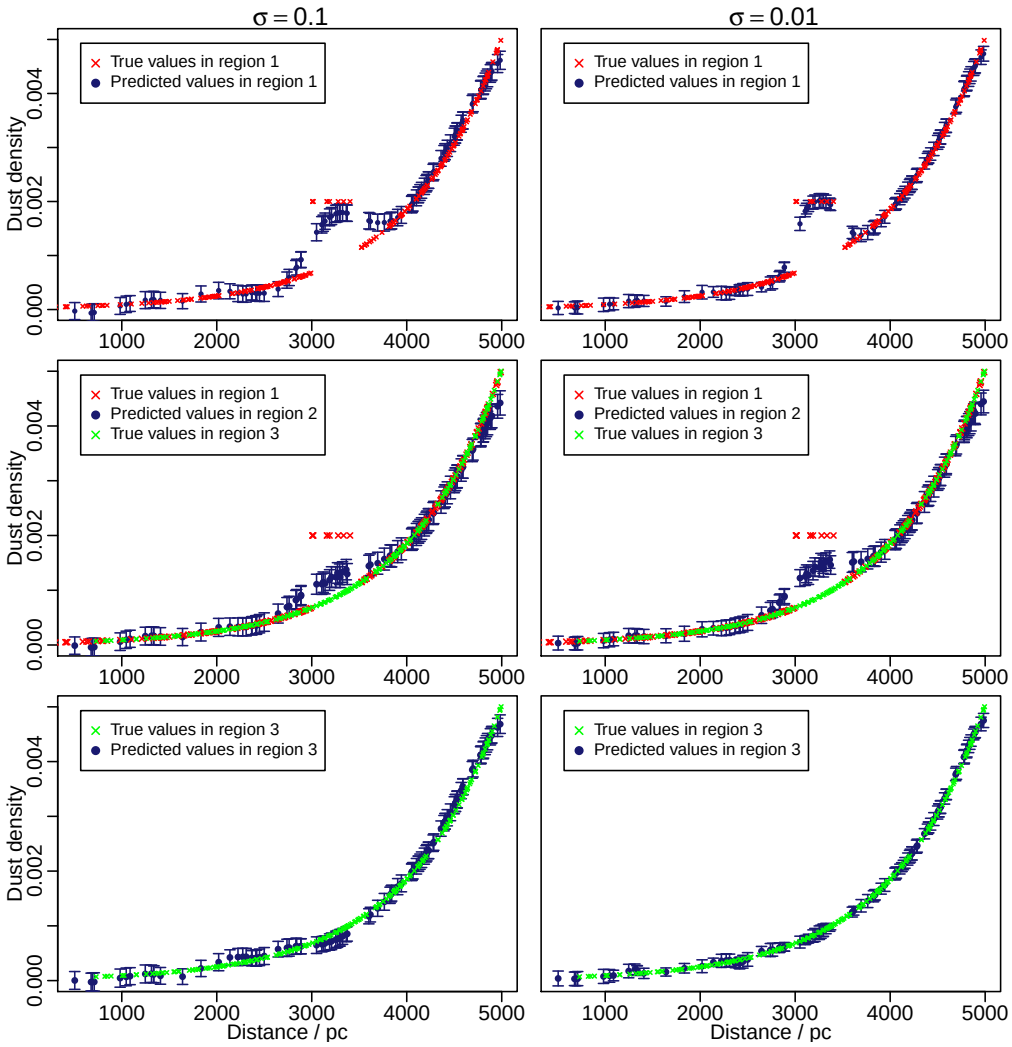


Figure 3.3: Predicted dust densities (attenuation per parsec) in the different regions of figure 3.1 and for two different attenuation errors: 0.1 (left) and 0.01 (right). Red and green crosses show true values for regions 1 and 3 respectively, and blue points show predicted values for different regions (1, 2 and 3 from top to bottom). The error bars on the latter points are also predicted by our model. Although there are no observations in region 2 the model can still predict dust densities there.

on where, but the covariance prior tells is that the closer we are to region 1, the more likely we are to still encounter the cloud. (It should be noted that the physical transverse extent from 6° to 12° at 3 kpc is much less than the length scale, λ, we have adopted.) In region 3, the model predictions show no indication of a cloud: they are influenced primarily by the nearer, cloud-free attenuation estimates.

The length scale, λ, sets the maximum distance over which dust cells are correlated. Note, however, that the correlation is only significant for values considerably smaller than λ (see figure 2.2). Choosing a value of λ that is too small will result in too many cells being disconnected (or having very low correlations), with the outcome that the information in the data is propagated less well. A value of λ that is too large will make even quite distant cells relatively highly correlated, potentially blurring out the local variance determined by the data.

The cell size (here taken as constant) is the radial length over which we assume the dust density to be constant when setting up the model. It is only used by equation 2.2 (generally equation 2.6) to discretize the dust density for representing the dust attenuation towards observed stars. It is used neither in the calculation of the covariance nor in the computation of dust density at new points, so contrary to possible expectations it does not represent the minimum scale over which we can compute density variations. It does, however, set some kind of minimum length scale over which we are sensitive to dust variations. Ideally we would use very small cells, but the computation time grows as the third power of the number of cells, so in practice we are limited by computational considerations.

The hyperparameter, θ, sets the scale of the covariance and thus the amplitude of variations in the dust. For a given λ and cell size, a larger value of θ means we can capture larger variations in the dust. We see from equation 2.9 with $t = 0$ that θ is the expected variance in the dust at any point. An estimate for the value of θ is therefore the variance in the expected distribution of the dust density over all cells. We can get an order of magnitude estimate of this before applying the model, by using the simplifying assumption that towards a given star, n, every cell has the same dust density, ρ_n, and same variance therein, $\mathrm{Var}(\rho_n)$. Adopting a constant cell size g for a given star, it follows from equation 2.2, using a_n as our estimate of f, that

$$\rho_n = \frac{a_n}{j_n g} \tag{3.1}$$

where $j_n g$ is just the distance to the star. Let μ_ρ be the average of the $\{\rho_n\}$ across

all N stars. The weighted variance of this distribution is

$$\theta = \frac{\sum_{n=1}^{N} w_n(\rho_n - \mu_{\rho_n})^2}{\sum_{n=1}^{N} w_n} \tag{3.2}$$

where each weight, w_n, can be set equal to the inverse of the variance, $\mathrm{Var}(\rho_n)$, in the corresponding value of ρ_n. These can be found by taking variance equation 2.2 (with a_n as our estimate of f) in which $\rho_{n,j} = \rho_n$ is consistent with our above assumptions. This gives

$$\mathrm{Var}(\rho_n) = \frac{\mathrm{Var}(a_n)}{j_n\, g_n^2} \tag{3.3}$$

where j_n is the number of cells towards the star n. Equations 3.1–3.3 allow us to estimate θ using the measured attenuations, a_n, and their uncertainties, $\sigma = \sqrt{\mathrm{Var}(a_n)}$ as well as the distances and adopted cell sizes. We will see that the larger the value of θ, the larger the error bars on our dust density predictions will be, as we expect. Using the simple mock data, we get $\theta = 1.9 \times 10^{-7}$ for $\sigma = 0.1$ and $\theta = 1.8 \times 10^{-7}$ for $\sigma = 0.01$.

It is important to realize that the minimum length scale over which we can probe dust variations is *not* set by λ. We see in figure 3.4 and we will see later that we can probe significant variations on length scales much less than λ. The finite cells sizes (g) aside, the minimum length variation is actually set by a combination of λ and θ, plus, most importantly, by the data themselves.

Figure 3.4 shows the effects of varying the values of the hyperparameters on the model predictions. Here we compare the true and estimated dust density for different values of λ (columns) and θ (rows) for region 1 with $\sigma = 0.1$. As before, red crosses show true values and blue points show the predictions. Values of λ and θ increase from left to right and top to bottom, respectively. Larger values of λ produce smoother variations in the dust. This is largely because this connects more cells, thereby increasing the amount of data used to estimate the dust densities. This also decreases the uncertainties on the estimates (the error bars). Smaller values of θ prevent the model from following steep changes in the dust density: for a given λ we see smoother variations in the dust density at smaller θ.

The very nature of our covariance model is that points separated by less than λ have correlated posterior distributions. Thus not only are the predicted dust estimates correlated, but so are their predicted uncertainties. Thus the error bars of adjacent points in figure 3.4 are highly correlated. They are simple point estimates of a continuous function (we could make estimate on a grid twice as dense; the error

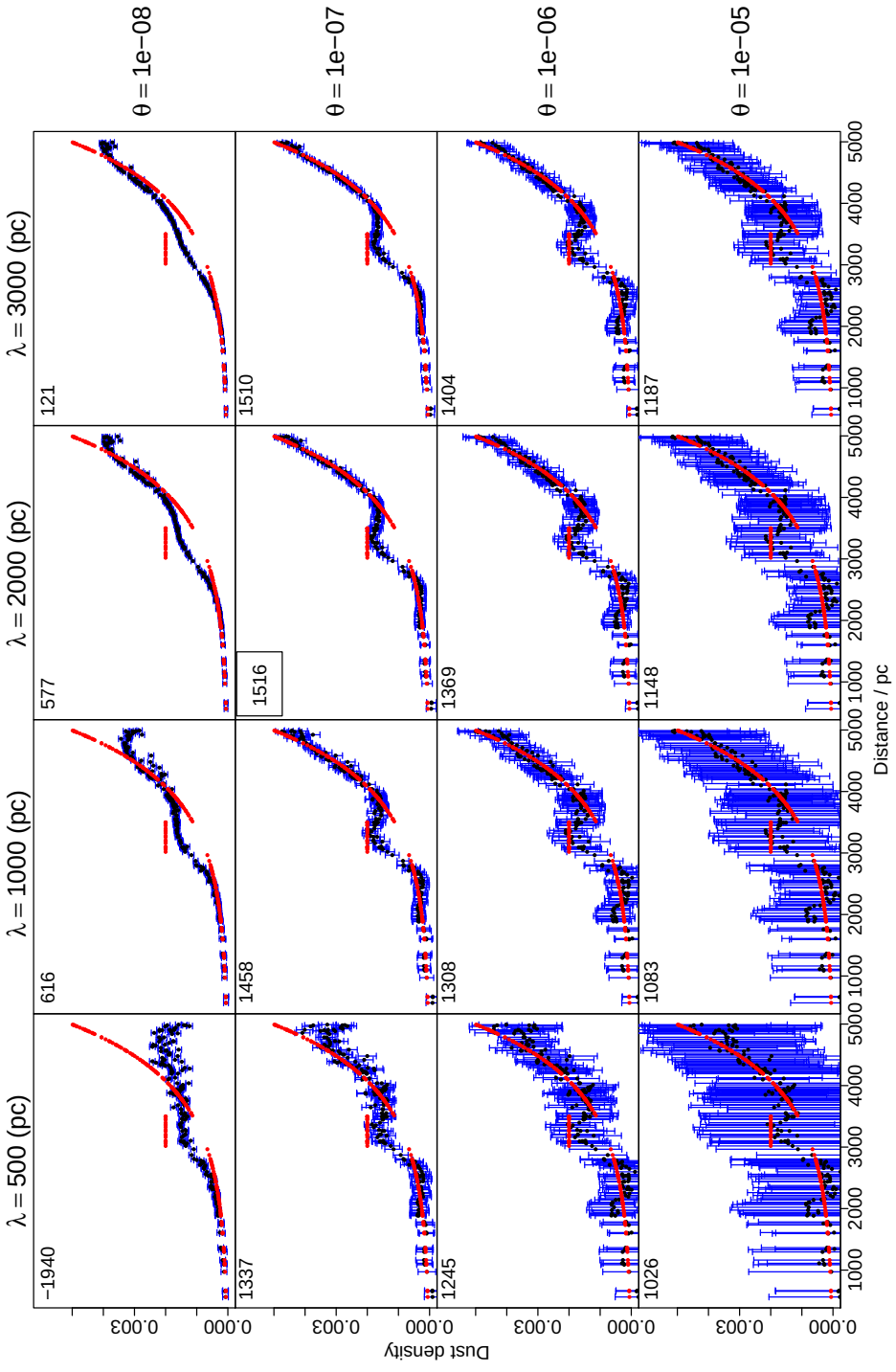


Figure 3.4: Effects of the hyperparameters on the model predictions (attenuation per parsec) in region 1 with $\sigma = 0.1$. Red points are the true values and black points with blue error bars are predictions by the model. It is important to note that the predicted uncertainties are highly correlated. Each column is for different values of λ (length scale), and each row is for different values of θ. The number in the top left corner is the natural logarithm of the reconstruction probability, defined by equation 3.4

bars would not change). Significant variations of the dust density can be (and are) obtained which are far smaller than these error bars.

In a real application – where we do not know the true dust densities – we could compute the likelihood of the data at the predicted values of the dust densities from equation 2.7. We could do this for a range of hyperparameters λ and θ and find the model value which gives the highest likelihood. This is a form of Bayesian model selection, because the regularizing prior has been used to estimate the individual dust densities. The different models correspond to different values of the hyperparameters.[1] The one practical disadvantage of this is that we would have to compute the dust densities along the l.o.s to every star and for all values of hyperparameters, which could become very time consuming.

For the sake of this simulation – where we do know the true dust densities – we compute instead the probability of the inferred dust densities given the true dust densities. By construction, the inferred dust densities have a joint Gaussian distribution, and the true dust densities have no variance. The log probability is therefore given by

$$\ln P = -\frac{1}{2}\triangle\boldsymbol{\rho}^{\mathsf{T}}\mathrm{C}\triangle\boldsymbol{\rho} - \frac{1}{2}\ln((2\pi)^n|\mathrm{C}|) \tag{3.4}$$

which we will call the *reconstruction probability*. It is analogous to the likelihood (or negative log of the sum-of-squared residuals, if the covariance were unity) but with model predicted values replaced by their true ones. n is the number of predicted points, C is the $n \times n$ covariance matrix or these points with elements given by equation equation 2.9, and $\triangle\boldsymbol{\rho}$ is the $n \times 1$ vector of the differences between true and predicted dust densities. The numbers in the corners of the panels in figure 3.4 show the values of this metric. The highest value is at $\lambda = 2000\,\mathrm{pc}$ and $\theta = 1\times10^{-7}$, which are the values we used for our predictions in figure 3.3. Recall that our pre-modelling order of magnitude estimate of what to use for θ gave a value of $\theta = 1.9\times10^{-7}$ for $\sigma = 0.1$, which is very similar. This procedure could not specify λ, because we are free to specify this according to the flexibility of the fitting we wish to achieve (and/or our knowledge of the true scale of the variations). For a fixed θ of 10^{-7} and different values of λ, we see from figure 3.4 that as long as λ is large

[1] What we call the likelihood here is not the likelihood in the sense of parametric models, which is the probability of the data for a given model *and* a given set of parameters. While we could maximize this likelihood to find the best parameters for a given model, we cannot use it to choose among models, because it contains no regularization and so will just identify the most complex model (we can eventually fit the data perfectly, noise and all). The likelihood we are talking about in this Gaussian process context is at a higher level. The "parameters", if you will, have effectively been marginalized over by the Gaussian process to produce the probability of the data for given hyperparameters.

enough (a few times cell sizes) to connect many cells in 3D space ($\lambda = 1000$, 2000 and 3000 pc in this example), we can achieve good results.
From figure 3.4, it is clear that the model predicts the dust densities with smaller uncertainties when using larger values of λ. This is because it then uses more points to predict the dust densities for every new point, although the closer points of course still have more influence (correlation) than more distant points. As the dependence on λ is not strong (once θ is set), we choose to fix λ to a relatively large value (a few times the cell size).

3.2 Gaia Universe Model Snapshot (GUMS)

We now look at a more realistic set of simulations taken from the Gaia Universe Model Snapshot (GUMS) (Robin et al., 2012).
GUMS is a simulation, generated by the DPAC prior to the Gaia launch, of what the Gaia catalogue can be expected to contain. It contains both the intrinsic properties of the objects as well as simulations of the noise-free Gaia observations (or more precisely, the corresponding catalogue products). It comprises around 1.6×10^9 stars (in single or multiple systems) with G-band magnitudes brighter than 20. We select the coordinates, distance, G-band magnitude, absolute V-band magnitude, $(V-I)$ colour, extinction A_V, and effective temperature T_{eff} for these stars. GUMS uses a dust model to generate its extinction values, but this is not part of our simulated catalogue (we have no knowledge of the true dust densities).
To map dust extinction we need not – and should not – use all stars. Due to computational limitations, the model, as it stands, cannot cope with anything nearly as large as the number of objects in the catalogue (see section 5.1.2 for possible improvements). We therefore first select just those stars with $G < 15$, parallax errors less than 5%, and T_{eff} between 5 000 and 10 000 K. Using cooler stars, e.g. down to 3 000 K, we found that it does not make a significant difference in the results, but as their parallax and/or extinction estimates would often be less precise, we would probably omit them in practice. The magnitude selection is imposed to ensure that the Gaia spectrophotometry have high signal-to-noise ratio so that A_V is determined to better than about 0.05 mag by the DPAC processing (Bailer-Jones et al., 2013) and we calculate the expected end-of-mission parallax errors from the simplified formula in de Bruijne et al. (2014). Although a selection on apparent magnitude biases our sample towards less extinct stars, this is hard to avoid in practice (as almost all surveys have a magnitude limit). As Gaia provides parallaxes from which we can infer distances, we could instead attempt to select all stars within a given

volume. But this selection (magnitude and temperature limit) is better because the distance uncertainties are asymmetric and are themselves a strong function of magnitude, plus the interstellar extinction would significantly limit the size of a complete volume at low Galactic latitudes. A better approach might be to limit the selection to intrinsically bright stars, over a narrow $T_{\rm eff}$ range (and therefore absolute magnitude range) for which we can estimate accurate extinctions. This will be explored in more detail in subsequent work.

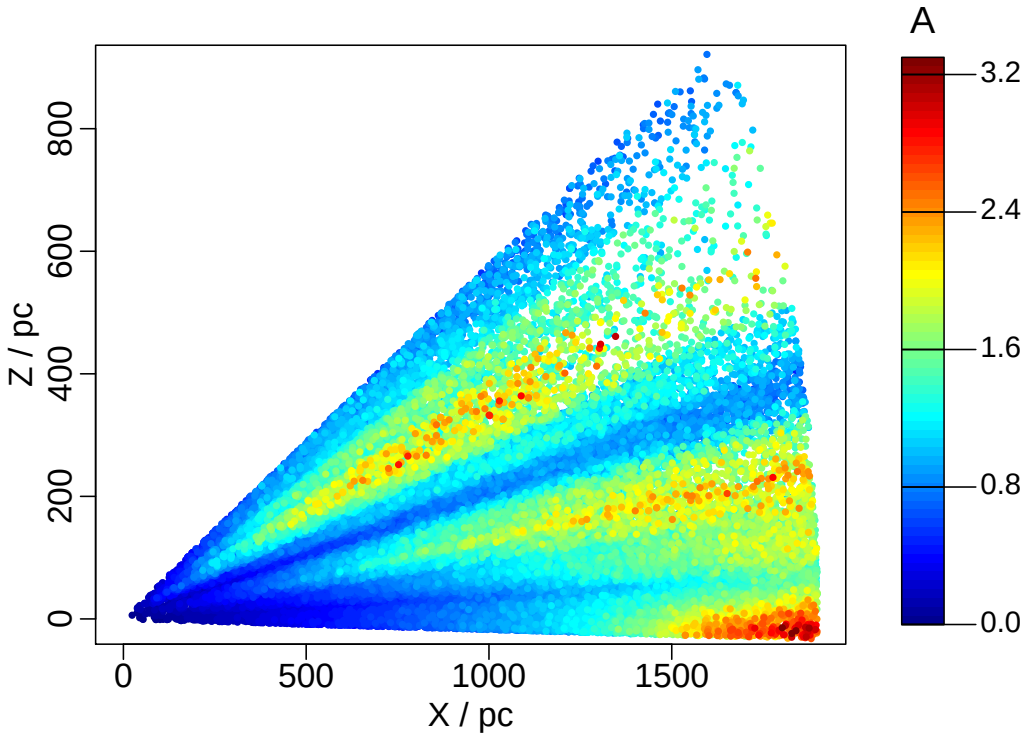


Figure 3.5: Two-dimensional positions of the 52 000 stars meeting our selection criteria over the region $5° < l < 7°$ and $-1° < b < 30°$ out to 2 kpc from the GUMS catalogue, colour-coded by the true (GUMS) extinctions in magnitudes.

Applying the above selections gives us 30 million stars distributed throughout the simulated Galaxy. To ease the interpretation of the results for the sake of this demonstration, we select stars in a wedge narrow in longitude (5° to 7°) but broad in latitude (-1° to 30°) within 2 kpc containing around 52 000 stars. Their positions and extinctions are shown in figure 3.5. The drop off at high latitudes is due to the decline in density of the Galactic disk population away from the plane. The presence of small, nearby regions of higher dust density are apparent from the higher

attenuations along some l.o.s.

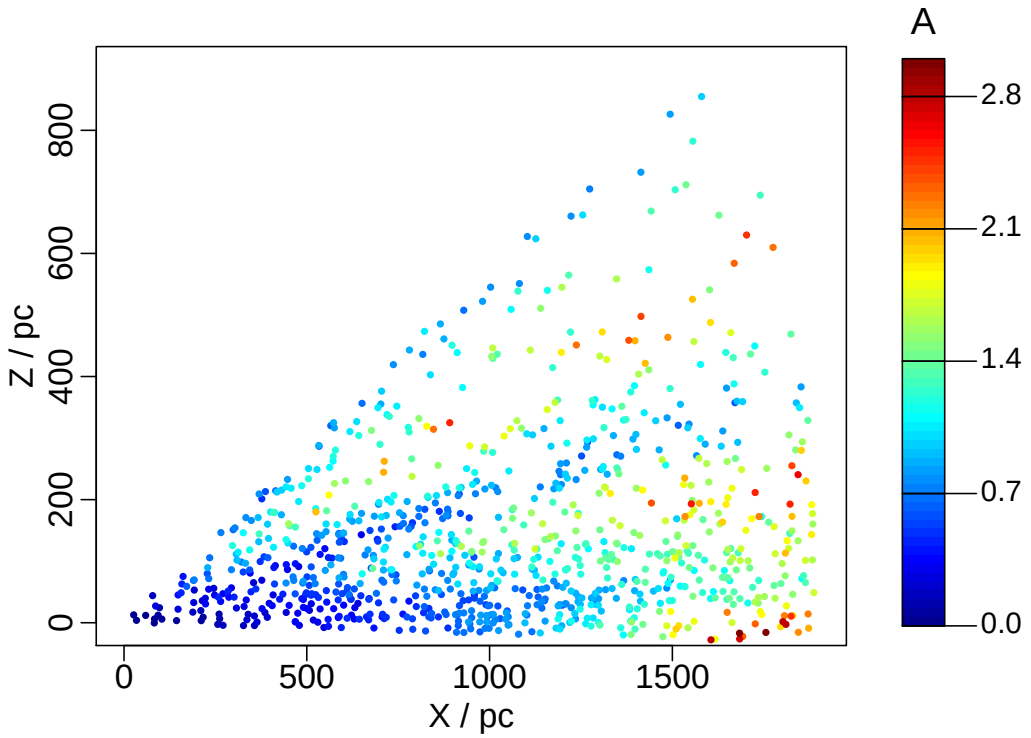


Figure 3.6: Random set of 1000 stars selected from those shown in figure 3.5. See section 3.2.

A useful feature of our method is that we do not have to use all stars within a region. Although the stars are independent probes of their l.o.s extinctions, stars in close proximity to one another probe much of the same dust. Provided the stars retained have a high enough spatial density to map out the minimum scale of the spatial variations we want to probe, randomly removing additional stars will not qualitatively change the resulting map. It may reduce the precision, because we then have fewer measurements to determine the dust density, but often other factors dominate the uncertainties (we will examine this in section 5.1.2). However, as reducing the number of stars can lead to great computational savings and increased numerical stability, this is a useful strategy to pursue. Figure 3.6 shows the positions of 1000 stars randomly selected from the sample of 52 000. Although there are far fewer stars, we can still see most of the structures from the full sample.

Figure 3.7 shows the predicted dust densities for 2000 random new points in the

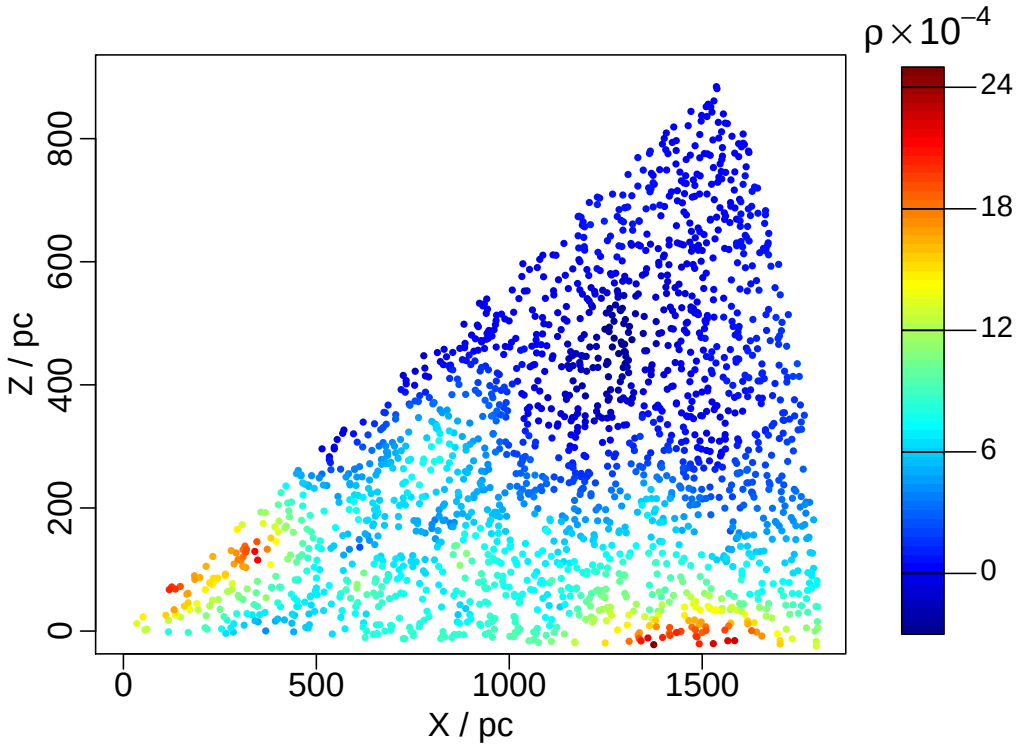


Figure 3.7: Predictions of the dust densities for the selected area from GUMS catalogue ($5°<l<7°$ and $-1°<b<30°$) for 2000 new points in two dimension coloured coded by the dust density values. A high value dust cloud is seen at the large distances and low latitudes as well as some lower-value dust clouds in other locations.

selected region. A localized region of high dust density ("dust cloud") is apparent at around $(X, Z) = (400, 150)$ which is responsible for the high extinctions beyond this point visible in figure 3.6, namely the upper diagonal wedge. Likewise, the very high extinction region in the plane beyond about 1400 pc is assigned a cloud at the same distance. Less apparent is a dense dust cloud responsible for the intermediate diagonal wedge in figure 3.6; the model has instead attributed this to a more diffuse region of higher dust density, perhaps because there is a larger region of higher extinctions at intermediate latitudes. It should be noted that the position of these new points have been chosen at random; we are free to select them and predict dust density at any point in 3D space. For this predictions we used a dust correlation length scale, λ, of 2 kpc, dust cell lengths, g, of 250 pc and a dust variation scale, θ, of 1×10^{-7}.

In the above we selected just 1000 stars from a possible 52 000 which met our selection

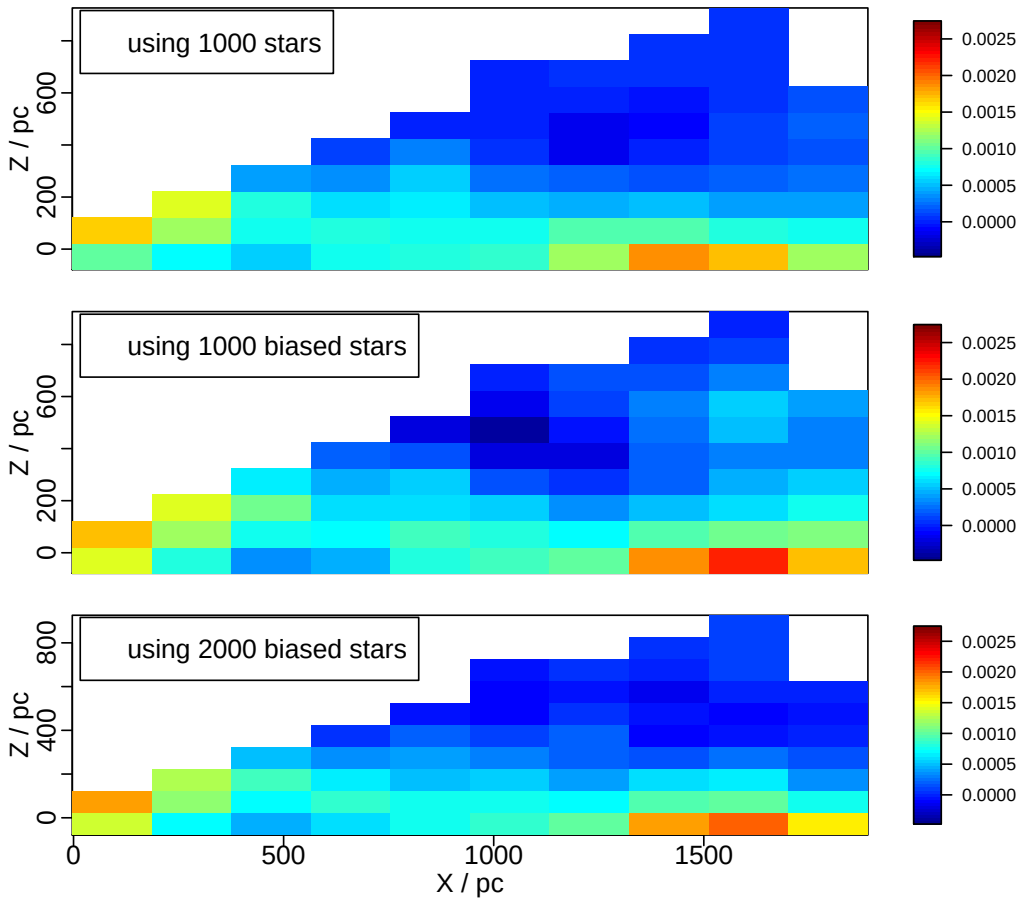


Figure 3.8: Effect of different sampling on the dust density predictions (attenuation per parsec, shown as the colour scale) for the region of $5°<l<7°$ and $-1°<b<30°$ using GUMS catalogue (section 3.2). Top panel shows the predictions using 1000 randomly sampled input data (from the entire 52 000), the middle panel is the results using 1000 input stars which sampled in a biased way towards high value attenuations, and the lower panel shows predictions in case of the same biased sampling but using 2000 stars. The results are pretty much consistent, capturing similar trends.

criteria. How might this selection affect our inference?

The top panel of Figure 3.8 shows the same results as in Figure 3.7 but now giving the average dust density in the rectangular region. The middle panel shows the results when we instead select 1000 stars but biased towards selecting higher attenuations. For the biased selection, we sample from the data with the probability of being selected proportional to the rank of the sorted attenuation values. The bottom panel is for 2000 stars selected in this same biased manner. Overall we see a high level of consistency, although not surprisingly, these latter two data sets reveal larger densities in the higher extinction areas. There is essentially no difference between using 1000 and 2000 stars, however. This is good news for our method, because it has a poor computational time scaling with the number of data points.

3.3 Application to APOKASC data

Having demonstrated the basic features of our model on simulated data, we now apply it to a set of real data to construct a 3D dust map in a small region of the Galaxy.

We use extinctions and positions of nearly 2000 stars provided by Rodrigues et al. (2014). They use spectroscopic and asteroseismic data of giants observed by APOGEE and the Kepler satellite (APOKASC catalogue), together with photometry from SDSS, 2MASS, and WISE and apply a Bayesian method to determine their extinctions and distances. The sample covers 17 degrees in longitude (from 68° to 85°) and 14 degrees in latitude (from 6° to 20°). We select stars with distance uncertainties of less than 10% within 3 kpc and remove stars with negative extinctions. This leaves around 1900 stars with observed extinctions shown in figure 3.9. The spatial distribution arises from the Kepler's CCD placement in the focal plane. Many of the extinctions are small, but we do see patches of higher extinction on different scales, especially in the lower right of the plot.

We apply our model to estimate the dust density (specifically: the mean and standard deviation of a Gaussian distribution at each point) at 2000 points distributed at random throughout the volume occupied by the stars. This has the shape of a pyramid with its apex at the Sun. We use $\lambda = 2$ kpc, $g = 250$ pc, and $\theta = 3{\times}10^{-7}$ (set as described in section 3.1). Figure 3.10 shows the mean estimated dust density of each of these points projected onto the sky. Comparing the coverage of this dust map with the input data makes it apparent that we can predict dust densities for points outside of the measured field. The model predicts an extended

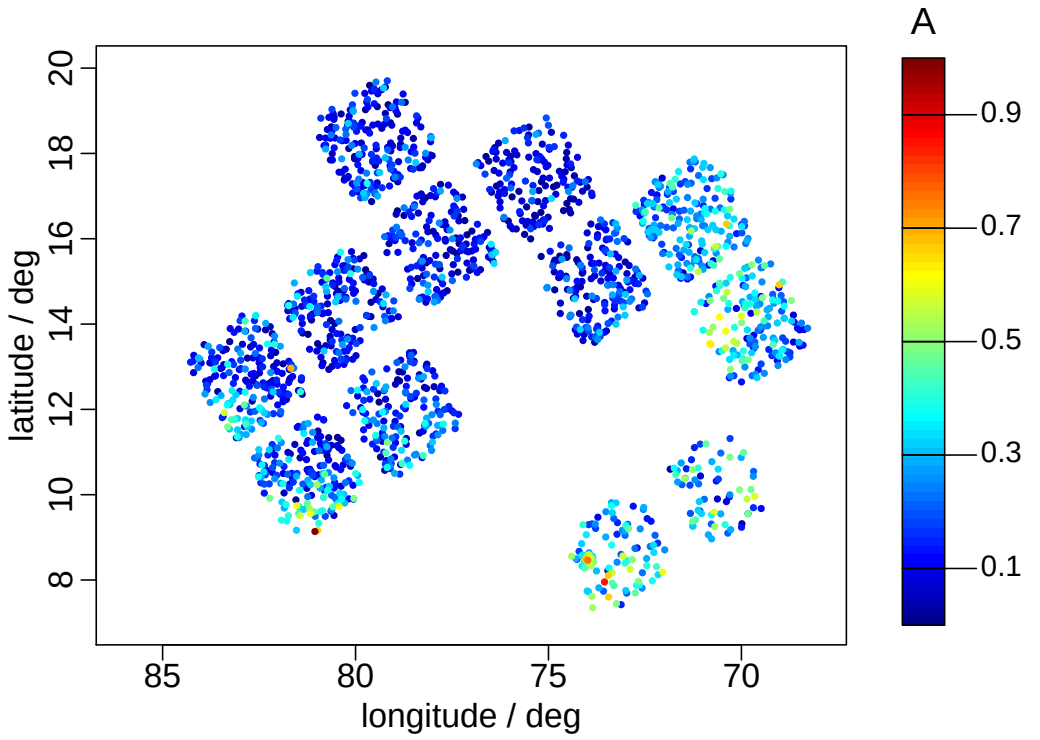


Figure 3.9: Extinction values of 1900 stars from Rodrigues et al. (2014) lying within 3 kpc. The colour indicates the extinction in magnitudes. The discrete squares are due to the Kepler satellite fields.

region of higher density at the bottom of the region, presumably driven by the higher extinctions for some stars in the lower fields in figure 3.9.

As the inferred dust distribution covers a range of distances, a sky projection like Figure 3.10 does not give the full picture. Figure 3.11 attempts to show the three-dimensional distribution. As expected from the input data, regions with higher dust densities are located at the lower latitudes, as is visible at the bottom of the east and west triangles. There is also a relatively high density region at about 1.5 kpc from the Sun which, as it is most visible in the south triangle, must be located predominantly in the southern part of the region. It is important to note that there is no l.o.s ("fingers of god") effect in the dust reconstruction. This is a direct consequence of the non-parametric nature of our model plus its minimal assumption on a smoothness prior. This is one of the advantages that our approach brings over more traditional mapping techniques, which are generally based on independent estimates of the dust

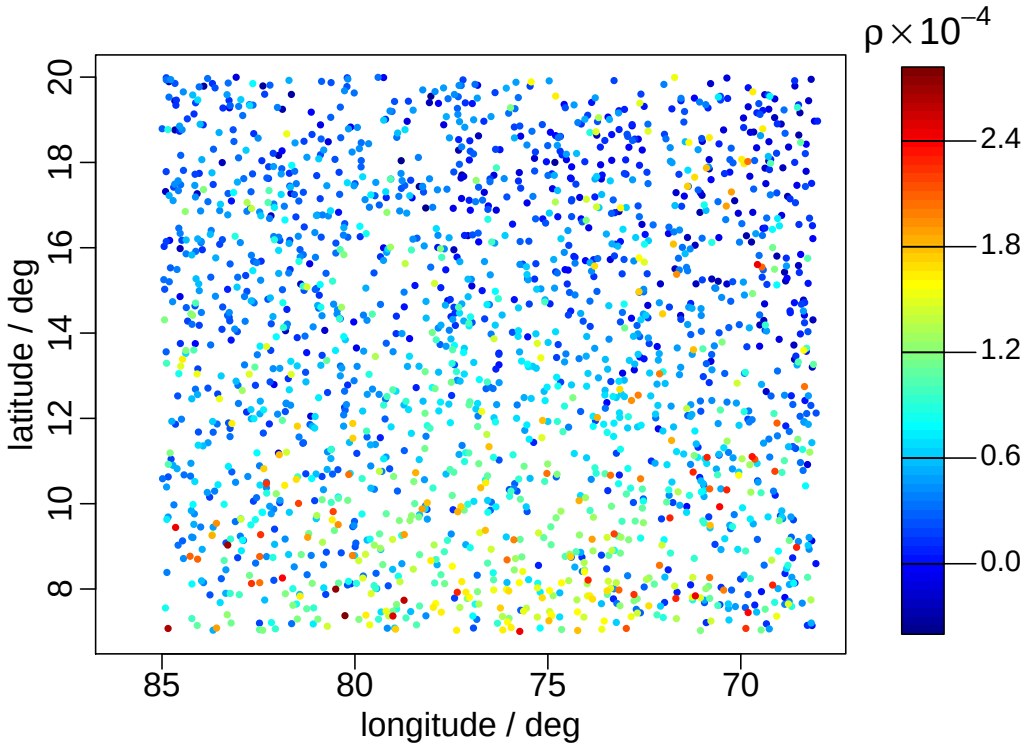


Figure 3.10: Inferred dust density (attenuation per parsec) for the area of $68°<l<85°$ and $6°<b<20°$ out to 3 kpc using data from Rodrigues et al. (2014) shown in 2D with the same frame as figure 3.9 (the input data). It should be noted that dust density is local in 3D and these points lie at a range of distances as seen in figure 3.12.

density along different l.o.s, which lead necessarily to the fingers of god.

One way to assess the performance of our model is to estimate the dust density at several points along the l.o.s towards to a star, and then to use equation 2.2 to predict the attenuation, f_n. The standard deviation in this prediction is found by taking the variance of equation 2.2,

$$\mathrm{Var}(f_n) = \mathbf{g}^\mathsf{T}\,\mathrm{C}\,\mathbf{g} \tag{3.5}$$

where $\mathbf{g}$ is the the vector of cell sizes along that l.o.s, and C is the covariance matrix (with elements given by equation 2.9) of the dust densities in these cells. It should be noted that this expression takes into account the (often large) covariance between cells along a l.o.s. (This is essential, because our model by its very nature assumes spatial correlations in the dust.) We compute attenuation estimates and uncertainties in this way for 200 stars in the two lower right Kepler fields in figure

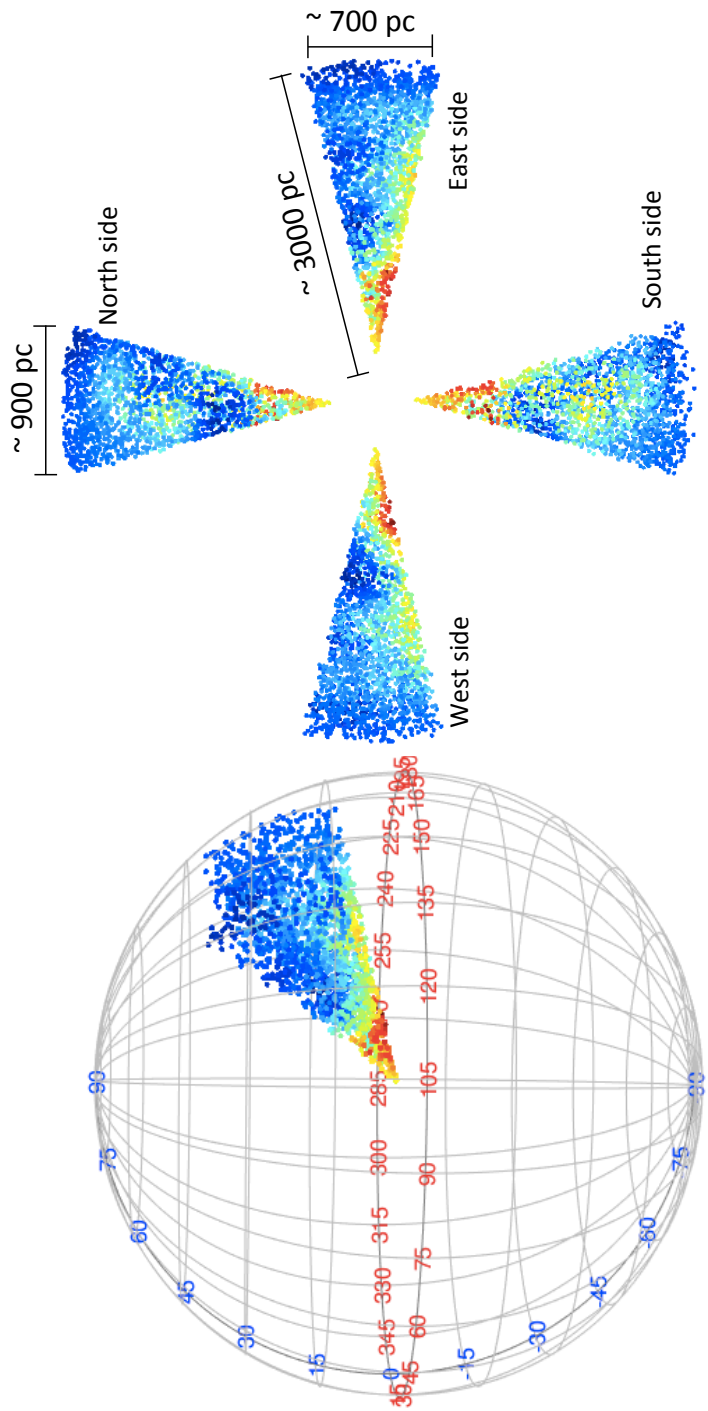


Figure 3.11: Dust densities estimated by our model over the pyramid-shaped region $68°<l<85°$, $6°<b<20°$, $d < 3$ kpc, computed using the extinction and distance data from Rodrigues et al. (2014). The left panel figure shows the data in three dimensions. The right panel shows the dust density in projection as viewed from the outside sides of the pyramid. 'north' means looking from top to bottom, perpendicular to the upper side of the pyramid, and similarly for 'south' (looking from the bottom). The 'east' and 'west' views are looking perpendicular to the sides of the pyramid, where 'east' means looking from higher longitudes to lower ones, and 'west' from lower to higher. Colour scale is as in figure 3.10. It should be noted that these triangles are plotted with a larger scale parallel to their base (transverse to the l.o.s from the Sun) – a larger opening angle than in reality – in order to better resolve the details.

3.9. For each star we use 15 cells with constant cells sizes per l.o.s. As the stars are at a side range of distances, this corresponds to cells sizes between 39 and 334 pc.

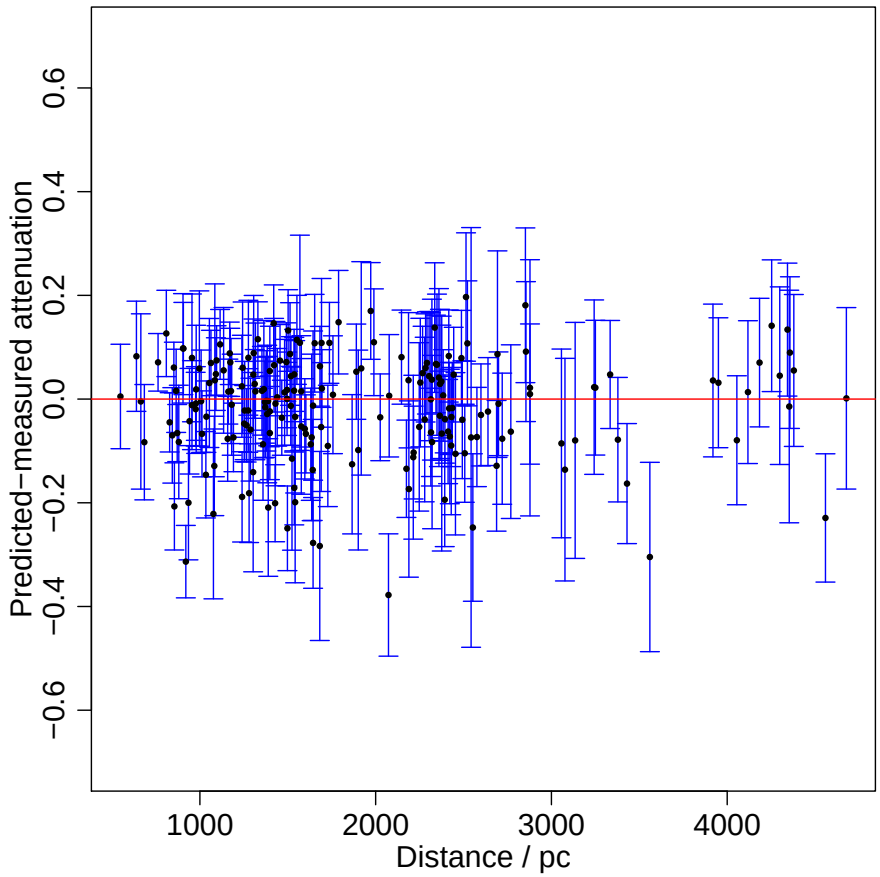


Figure 3.12: Residuals between reconstructed and measured attenuations as a function of distance for the 200 stars in the APOKASC data (from the bottom right two fields shown in Figure 3.9).

Figure 3.12 shows the residuals (predicted minus measured) of the attenuations to these 200 stars, along with the corresponding error bars. These error bars are computed from the sum of the variance in the measurement, σ_a^2, and the variance from equation 3.5. We see no particular trend in the magnitude of either the residuals or their uncertainties with distance. However, it looks as though there is a small negative bias, in the sense that our model slightly under estimates the measured attenuations. This is better seen in Figure 3.13, which shows the residuals scaled by

their uncertainty estimates. A linear model with ideal Gaussian residuals would show a Gaussian distribution with zero mean and unit standard deviation. A negative bias is apparent. This is not necessarily a sign of a bad model, because the whole point of the model is to make an inference from the data subject to the smoothness constraint. Our model uses a zero mean prior for the dust density; thus, in the absence of data the model gives zero density. We can always add an offset to the covariance function of the Gaussian process to have a non-zero mean (e.g. if we want to examine the high density regions).

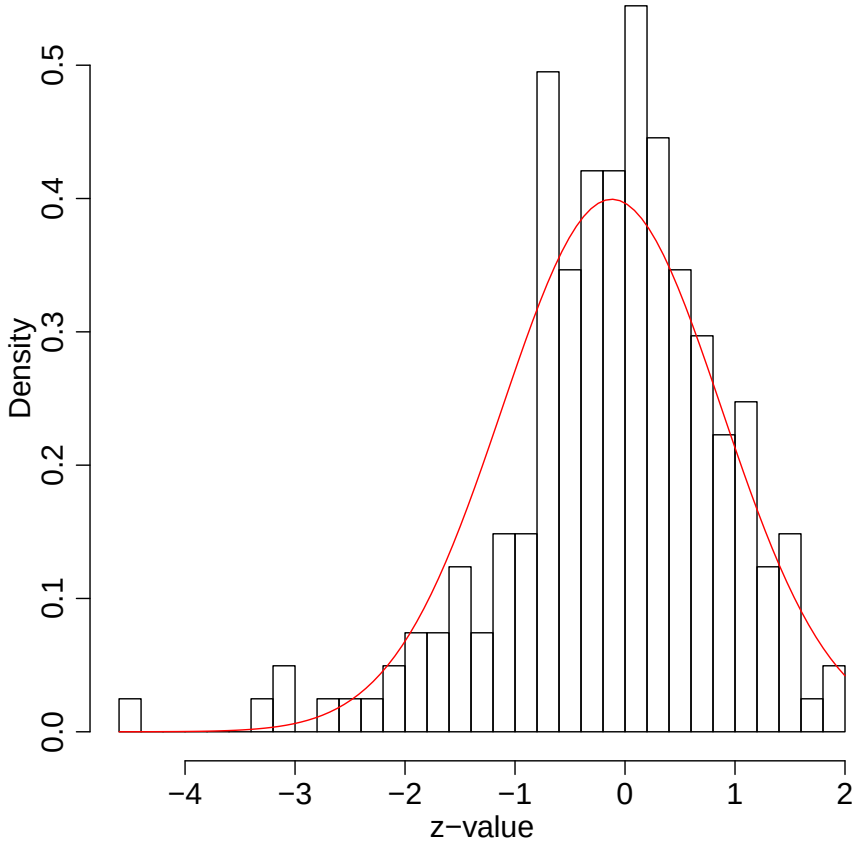


Figure 3.13: Distribution of the scaled residuals, $(a_n - f_n)/\sigma$, where $\sigma^2 = \sigma_a^2 + \mathrm{Var}(f_n)$, for the set of 200 stars in the APOKASC data. The mean is -0.12 and the standard deviation is 0.99. The curve shows a Gaussian with these values for comparison.

In contrast, the standard deviation is almost exactly unity, indicating that our Gaussian process is very good at estimating the uncertainty in its predictions. Figure 3.14 shows the predicted values together with the measured values as a function of distance. Recall that the error bars on the predictions are correlated. They increase with distance because the more distant stars have larger dust cells. There are a few measurements which lie well outside of the main envelope of the data, but given the number of points (and the size of their error bars in some cases) this is entirely consistent. We see how the model has smoothed out these "outliers", on account of its built-in assumptions. We further see that our model's underestimation of high extinctions is a consequence of its smooth variation with distance.

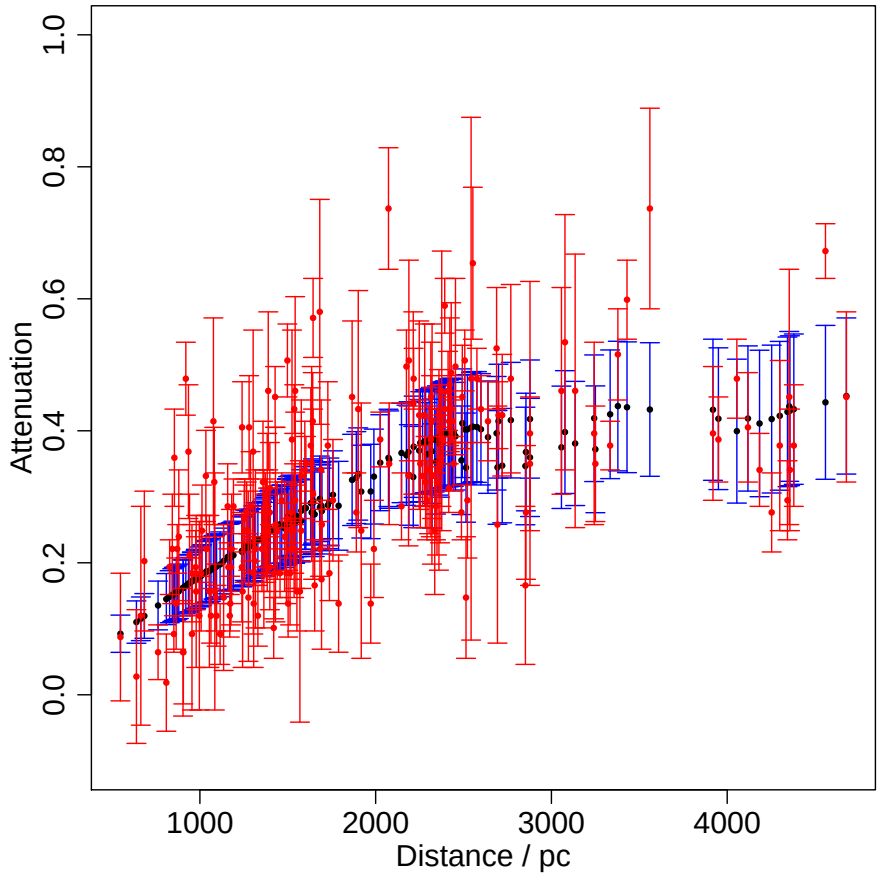


Figure 3.14: Reconstructed attenuations as a function of distance (black dots with blue error bars) on top of the measured values (in red) for the set of 200 stars in the APOKASC data. It is important to note that the error bars in the predictions are highly correlated.

Table 3.3.1: Logarithm of the Bayes factors for APOKASC data for different ranges of λ and θ.

$\log_{10}$ (Bayes factor)				
$\theta \setminus \lambda$ (pc)	500 pc	1000 pc	2000 pc	3000 pc
1×10^{-8}	-1.41	-0.87	-0.43	-0.21
1×10^{-7}	0.05	0.14	0.15	0.13
1×10^{-6}	0.05	-0.06	-0.18	-0.26
1×10^{-5}	-0.40	-0.48	-0.65	-0.69

We could try to fix the hyperparameters by calculating the Bayesian evidence. The evidence (or "marginal likelihood") is the probability of observing the data, for fixed λ and θ, averaged over all possible instantiations of the model. We compute this by drawing one sample from the J-dimensional Gaussian process prior (which gives us J values for the dust density), calculating the likelihood for these model dust densities, repeating it for a large number of times (e.g. $K = 10^5$), and then averaging these likelihoods

$$P(\mathbf{a}_N|\lambda,\theta) = \frac{1}{K}\sum_{k=1}^{K} P_k(\mathbf{a}_N|\{f_i\},V_N) \tag{3.6}$$

where $\{f_i\}$ are calculated attenuations (equation 2.6) using dust densities drawn from the prior. Having done this for various λ and θ, we then calculate the Bayes factors, which are the ratio of these evidences (for different λ and θ) to the one with the specific values of λ and θ used for our APOKASC data ($\lambda = 2$ kpc and $\theta = 4\times10^{-7}$). We report these in table 3.3.1.

We get values for the Bayes factors in the case of APOKASC data which agree broadly with what we calculated for λ and θ. But in the case of the simulated data (section 3.1), the Bayesian evidence does not give us a useful discrimination between models. Most values are very close to zero for a range of θ and λ because our simulated data have high extinctions, which are not well represented by a Gaussian process prior with zero mean. The APOKASC data, in contrast, have smaller extinctions. This shows that using a non-zero mean in the Gaussian process prior will better construct the dust density in regions with higher extinctions, such as the disk of the galaxy and the spiral arms. A different covariance function could also be used in the Gaussian process. We tested various forms of the covariance function, such as truncated exponential forms, but they did not make a significant difference to our results. The covariance function that we are using has the advantage that we

can get different variation slopes by changing α (see fig. 2.2).

4

Galactic disk structure

In this chapter, I represent the 3D map of the dust distribution in the Galactic disk out to about 7 kpc from the sun. This chapter is adopted from our recent paper Rezaei Kh. et al. (2018, b).

The chapter is organised as follows. First, I explain the APOGEE data used as the input. Then I discuss the results, and finally illustrate the detection of the spiral arms from our map.

4.1 APOGEE data

To map the dust distribution in the Galactic disk, we use data from APOGEE-2 (Blanton et al., 2017; Majewski et al., 2017; Abolfathi et al., 2017), a near-infrared high-resolution spectroscopic survey targeting bright stars (Eisenstein et al., 2011; Zasowski et al., 2013). As the survey operates in the near-infrared, the effects of extinction are about an order of magnitude lower than at optical wavelength, enabling it to observe stars in the highly obscured regions of the Galactic disk and towards the Galactic centre.

We select giants from APOGEE DR14 (Majewski et al., 2017; Abolfathi et al., 2017) using their $\log g$ information ($0.5 < \log g < 3.5$). We estimate distances and colour

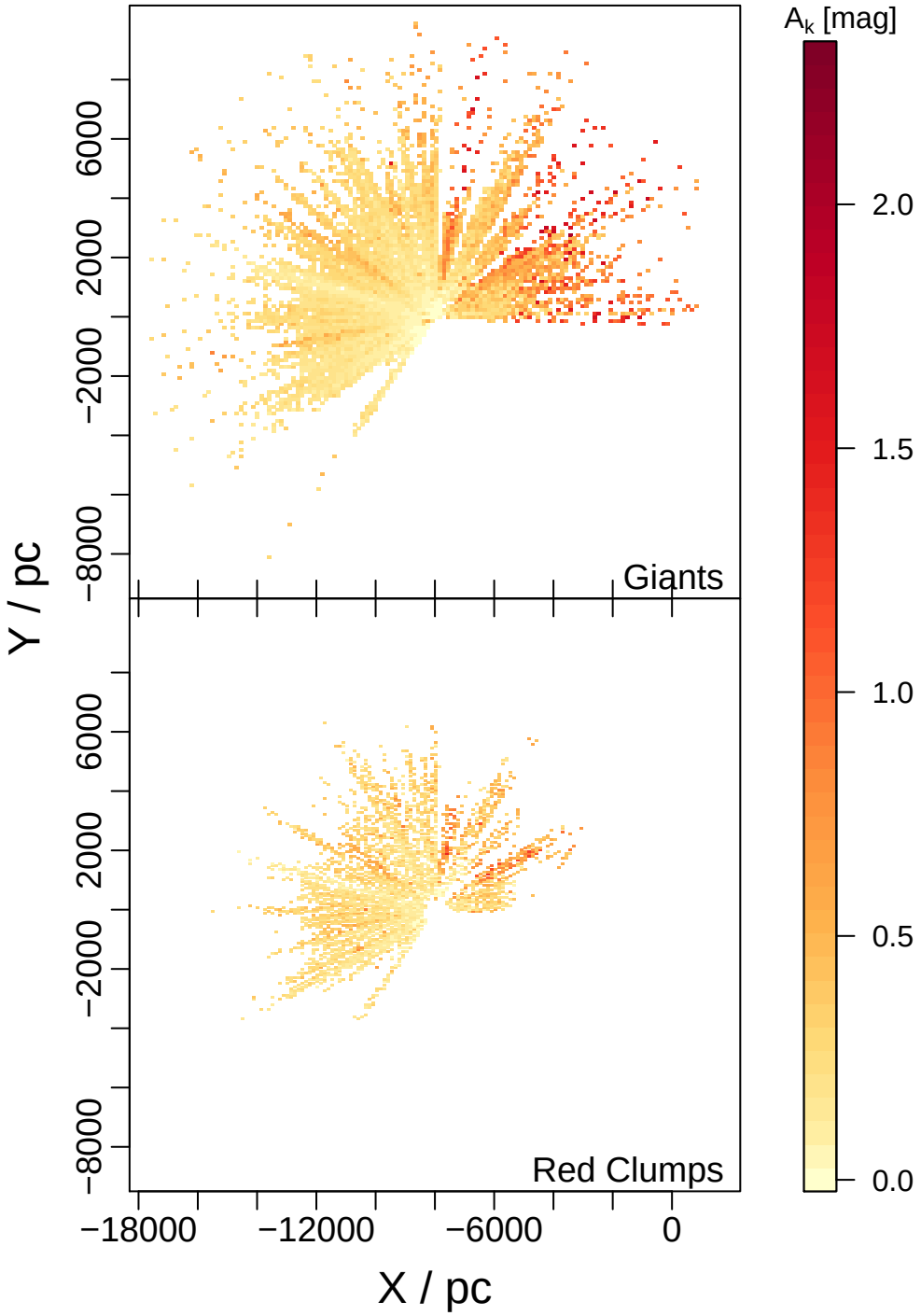


Figure 4.1: Distribution of input stars in the X-Y plane (±100 pc in Z). The sun is at (-8000, 0) and the Galactic centre is at (0,0). The top panel shows giants stars and the bottom panel shows red clump stars. The colour represents their l.o.s K-band extinction. Most giants can probe out to about 7 kpc, while red clump stars probe to about 5 kpc on average.

excess (E(J – K), as described below) using the stellar parameters (T_{eff}, log g and [M/H]) together with isochrones from stellar evolution models [1] (PARSEC, Tang et al., 2014; Chen et al., 2015). The stellar parameters were determined by the APOGEE Stellar Parameters and Chemical Abundances Pipeline (ASPCAP, García Pérez et al., 2016).

For each star, we look for the closest point in the T_{eff} vs. log g plane for the corresponding isochrone at a given metallicity, taking into account the uncertainties in the stellar parameters. We use the calibrated stellar parameters (PARAM) which were calibrated using a sample of well-studied field and cluster stars as well as stars with asteroseismic stellar parameters (Holtzman et al., 2015). Each star in the APOGEE sample has 2MASS J, H, and K_s magnitudes. In the isochrone grid we find the absolute magnitudes M_J , M_H, and M_{K_s} corresponding to the stellar parameters. The colour excess E(J – K_s) is calculated as E(J – K_s) = J – K_s – (M_J – M_{K_s}). To convert E(J – K) to A_{K_s} we use the extinction law of Nishiyama et al. (2009) with $A_{K_s} = 0.528E(J - K_s)$.

We then calculate the distances as

$$d = 10^{0.2(K_s - M_{K_s})+5-A_{K_s}} \tag{4.1}$$

For a more detailed description see Schultheis et al. (2014). The distance from each star to the nearest point in the isochrone grid in the T_{eff} and log g dimension, together with the individual uncertainties σ_{Teff} and $\sigma_{\log g}$ (from the ASPCAP pipeline), gives the distance uncertainty for each star. As shown in Schultheis et al. (2014) the median uncertainty is approximately 30%.

Apart from giants, APOGEE also targets red clump (RC) stars. Bovy et al. (2014) introduced a new method to select these RC stars from APOGEE data based on their position in the colour–metallicity–surface-gravity–effective-temperature space. Because of the narrowness of the RC locus, distances to these stars can be estimated with an accuracy of 5% – 10% (Bovy et al., 2014). The extinctions for this sample are calculated by Bovy et al. (2014) using Rayleigh Jeans Colour Excess method (RJCE; Majewski et al., 2011) which relies on the fact that near- to mid- infrared colours are almost constant for all stars. Therefore the change in the colour of a star is due to interstellar extinction. The catalogue of RC stars is available from APOGEE DR14 containing around 30 000 stars with accurate distance and extinction estimates (Bovy et al., 2014), which we use as inputs for our model to map the

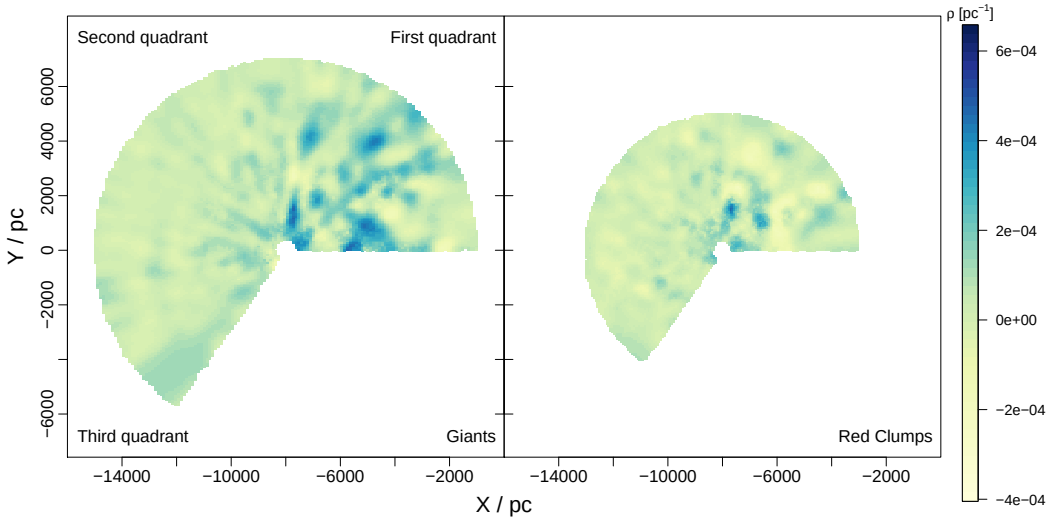


Figure 4.2: 2D image of the 3D map of the dust distribution in the Galactic disk (±100 pc in the Z direction) using giants (left panel) and RC stars (right panel). The sun is at (-8000, 0) and the Galactic centre at (0,0). The colour shows the mean of the dust density predictions over the column through the disk (pc^{-1}).

dust distribution.

From both the giant and RC catalogues we select targets within Z=100 pc above and below the Galactic mid-plane. The RC catalogue provides the Galactic Z values assuming the Sun is 8 kpc from the Galactic centre and 25 pc above the Galactic mid-plane (Bovy et al., 2014). We make the same assumption and calculate the Galactic Z for the giants to select stars in the disk of ±100 pc. Also, to be consistent with RC distance precision, from our giant sample we only select those with fractional distance uncertainties less than 0.05. This leaves us with about 5 000 stars from the RC catalogue and about 16 000 giants. Figure 4.1 shows both samples: the giant sample probes larger distances and gives higher extinction measurements, in particular in the first quadrant and towards the Galactic centre.

We use each of these samples separately as the input data in our model to infer the dust distribution in the Galactic disk. The typical separation between stars in both RC and giant samples is of the order of 200 pc. We use this as the cell size in the model, and adopt a correlation length of $\lambda = 1000$ pc. The corresponding θ for giant and RC samples is $9{\times}10^{-9}$ and $6.5{\times}10^{-9}$ pc^{-2} respectively (see chapter 2). The common mean dust density of the Gaussian process prior (ρ_μ) is computed

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

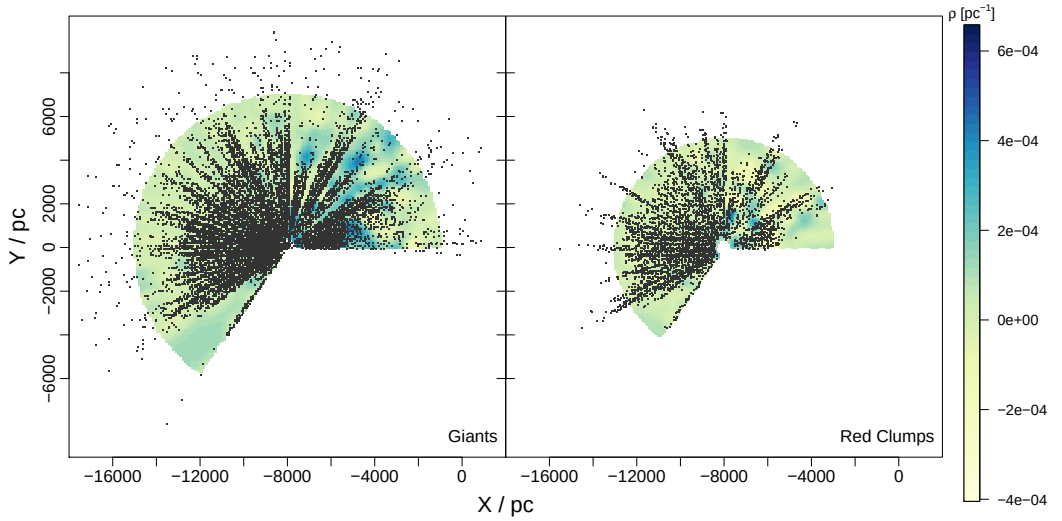


Figure 4.3: As Fig. 4.2 but now overplotting as black points the locations of the stars used to derive the underlying dust map. The giant sample extends to larger distances than the RC stars.

to be 1.2×10^{-4} pc^{-1} for both the giants and the RC samples (see chapter 2). It is important to note that the correlation length scale, θ, is *not* the resolution limit of our maps: we can probe much smaller scales depending on the distance between stars (Rezaei Kh. et al., 2017; Rezaei et al., 2018). The correlation length is in fact the upper limit on the distance between two points which can still interact, since we use a covariance function with finite support.

4.2 Galactic dust map

Figure 4.2 shows our dust maps for both the giants and the RC sample. There are many high dust density clouds and structures in both maps. They trace similar structures in the second and third quadrants out to distances where they overlap: there are many dust clouds with relatively high densities (higher than the average mean density of 1.2×10^{-4} pc^{-1}) spread around the area. However, the maps show dramatic differences in the first quadrant. The reason for this can be seen from Fig. 4.3 that shows the input data over-plotted on the predicted dust densities: it is clear that the RC sample poorly covers the first quadrant, especially towards the Galactic centre. In this case the posterior will be dominated by the prior. The giant sample, on the other hand, covers much greater distances in general, in particular they cover most of the area in the first quadrant, resulting in better constrained predictions in the first quadrant compared to that of the RC sample.

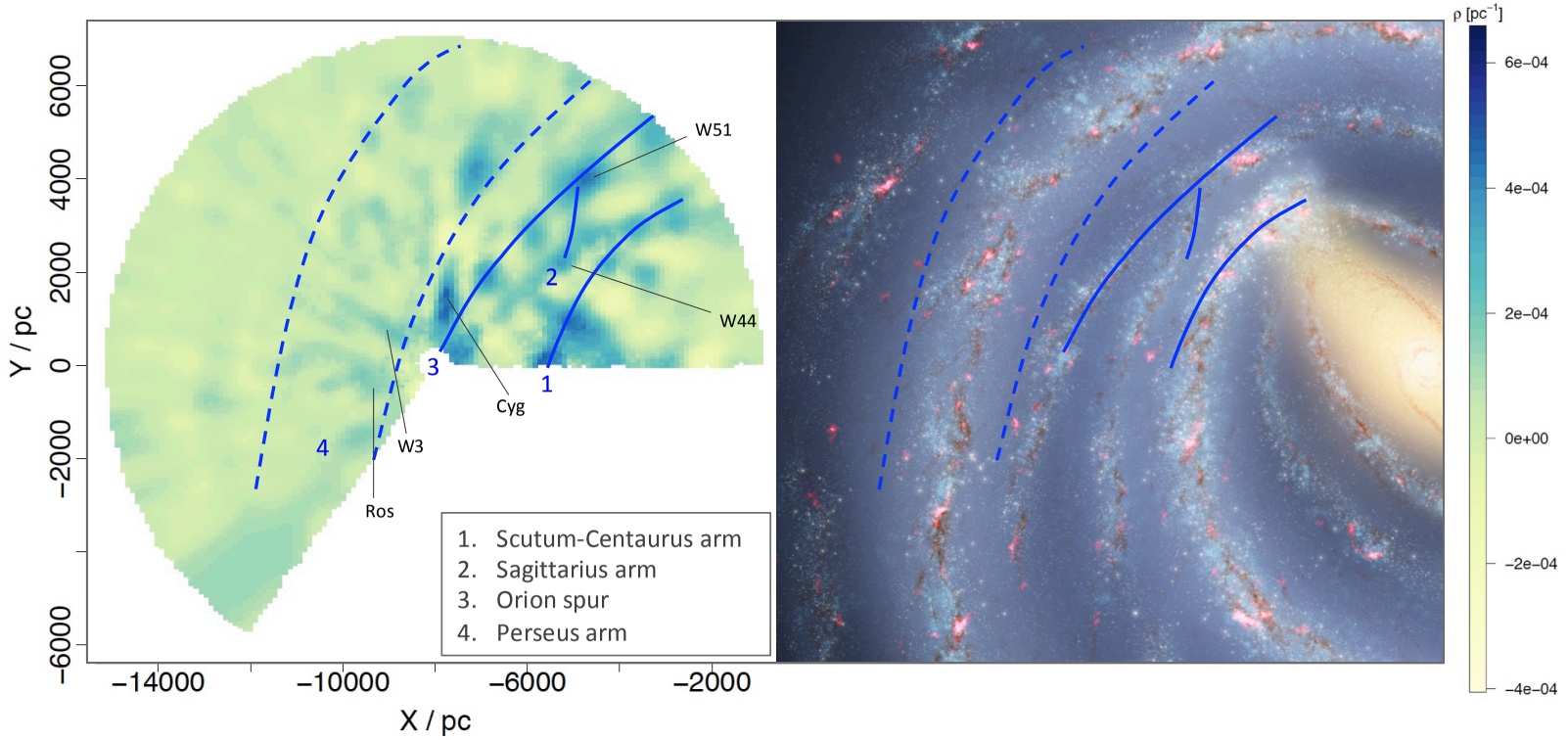


Figure 4.4: Left panel: our dust density predictions as in Fig. 4.2 (left panel), but now over-plotting with blue lines the approximate locations of the arms as we deduce them from this dust map. The dashed lines show an area in which relatively high dust density clouds are seen, but which do not lead to as such a distinct pattern as seen for the other three lines. The known giant molecular clouds detected in the map are also labelled. Right panel: our estimated location of the arms (blue lines) from the left panel plotted on top of the Spitzer sketch of the Galactic arms (by Robert Hurt, courtesy of NASA/JPL-Caltech/ESO).

Another area of interest in Fig. 4.3 is the lower left corner of the maps where there is a gap in the input stars in both the RC and giant samples. For close distances where the stars on both sides of the gap are within the 1000 pc correlation length, predictions are affected by the neighbouring stars. For more distant points, however, the correlation drops to zero and the predictions would be close to the input mean of the Gaussian process prior. The same criterion applies to the outer regions of the first quadrant in the map produced using RC stars.

We now consider only the map based on the giants, since it extends to greater distances. The high density dust clouds in the first quadrant of this map seem to be lined up to form arc-shaped structures along the expected locations of the spiral arms of the Galaxy. To explore this, we connect the lined-up high density clouds to predict the locations of the arms. As seen in Fig. 4.4, left, in the first quadrant, we connect three sets of lined-up clouds with blue lines; however, in the second quadrant, the high density clouds seem to be spread widely and do not shape clear lines; therefore, we show borders (as dashed lines) of the area where most of high density clouds are located. Afterwards, we over-plot our predictions on the Spitzer sketch of the Galaxy[2] (Fig. 4.4 right). It is worthwhile to mention that this artist's impression of the Milky Way is based on the Spitzer observations; the spiral structure in the impression is more qualitatively related to observational data. This is the the reason we choose to compare our spiral arm predictions with this particular map. Other present Galactic spiral arm models are heavily model-dependant, consequently quite different from one another, while the Spitzer map is mainly based on the stellar observations and does not have strong Galactic model assumptions, which makes it similar to our approach.

The locations of our predicted arms in the first quadrant are in relatively good agreement with the Spitzer arm structures, especially at the position of the Orion spur (line 3 in Fig. 4.4). Parts of the Scutum-Centaurus and Sagittarius arms (arms 1 and 2 in Fig. 4.4) also match nicely with our inferred dust clouds. The location of the Perseus arm coincides with high density clouds in between our predicted two dashed lines. This is in agreement with the recent finding of Baba et al. (2018) which uses data from Gaia DR1 concluding that the Perseus arm is being disrupted.

In addition to spiral arm structures, our map detects some of the known giant molecular clouds (GMCs) that are labelled in Fig. 4.4. Literature distances to the W51 GMC ranges from 5 to 8 kpc depending on the method used (Parsons et al., 2012). From our map, the distance to W51 GMC is about 5.5 kpc, which is in

[2]https://www.eso.org/public/images/eso1339e/

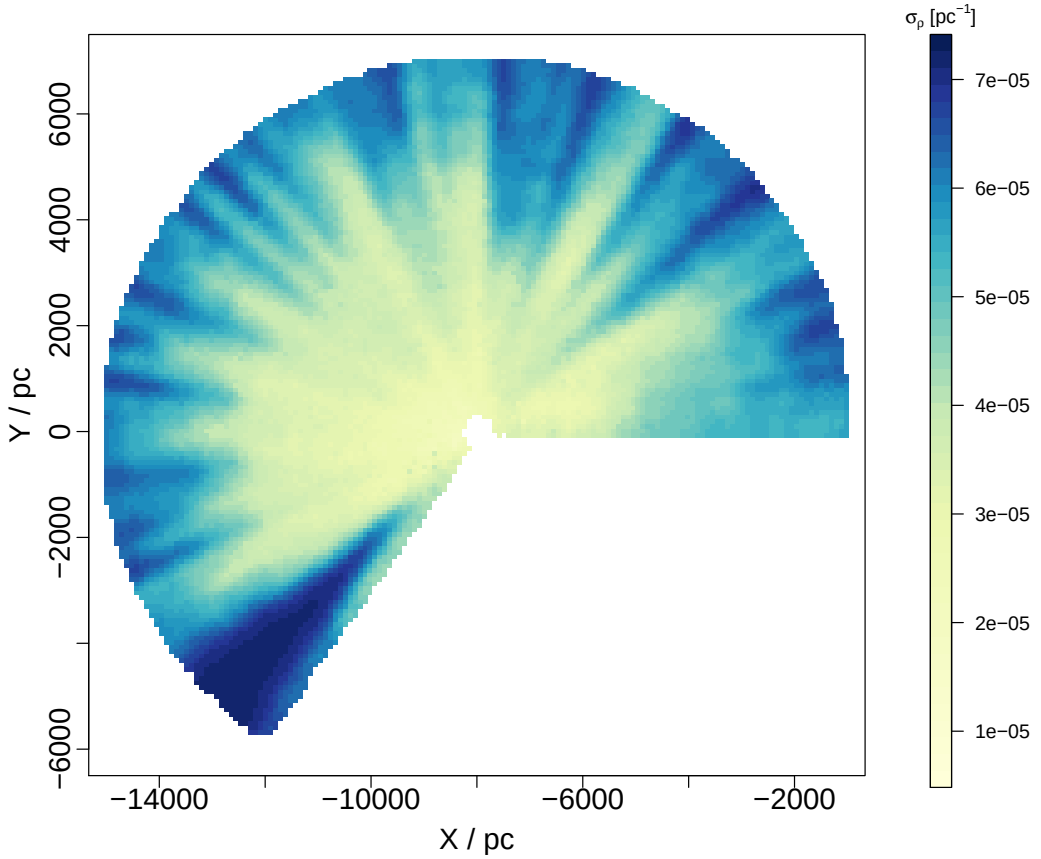


Figure 4.5: Standard deviation of the dust density predictions. The larger uncertainties (dark blue colour) appear at the places not well-populated by the stars.

agreement with the distances obtained by Sato et al. (2010) and Russeil (2003). We estimate the distance to the W44 GMC, the dust cloud that seems to be associated with the Sagittarius arm, to be about 3 kpc. There are two nearby clumps of high-density dust at the location of the Cygnus GMC at distances of about 1.5 and 2 kpc. The distance to the main OB associations in Cygnus X has been reported to be about 1.7 kpc (Schneider et al., 2006). We also see some moderate density of dust at the expected locations of the W3 and Rosette GMCs at about 2 and 1.5 kpc respectively.

Figure 4.5 shows the model predicted uncertainties. Generally the uncertainties increase as going to larger distances due to the drop in the density of stars in the

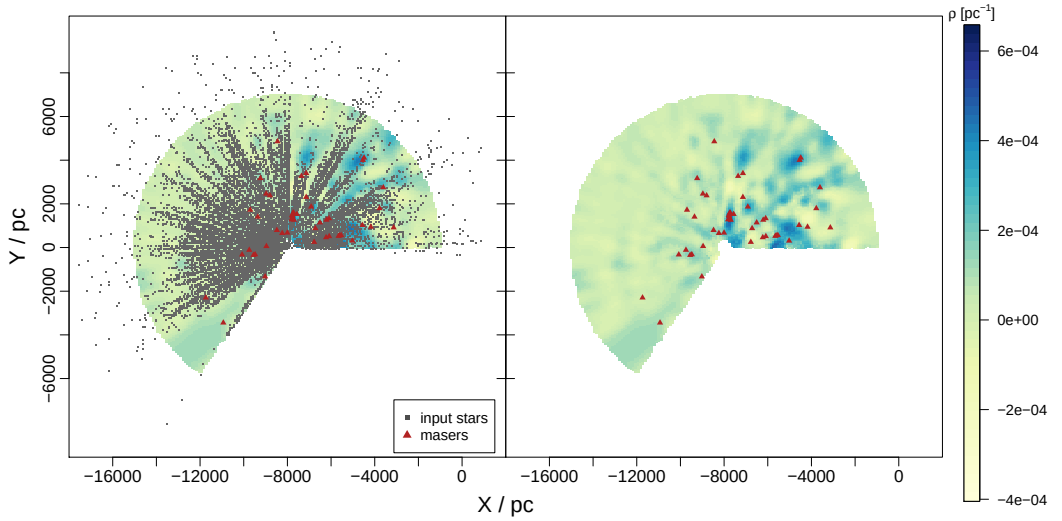


Figure 4.6: The left panel is the same as the left panel of Fig. 4.3, but now overplotting as triangles the positions of masers from Reid et al. (2014) which are within ±100 pc of the disk midplane. Some of the masers show correlation with high density predictions. The right panel is the same, but with the stars removed.

sample. Comparing this to the input data (Fig. 4.3, left) reveals that regions of larger uncertainty occur where the gaps between the input stars are larger, resulting in the dust density being less well constrained. As can be seen from Fig. 4.4, we are not able to constrain the location of the Sagittarius arm where it comes close to the sun. This is due to the lack of precise measurements in the direction towards the Galactic centre and lacking data in the southern hemisphere. There is a relatively high density line-shaped area parallel to the Orion spur that seems to be part of the Sagittarius arm. But Fig. 4.5 suggests the predictions in this area have relatively high uncertainties and are inferred primarily from their surrounding points. We therefore decided to not include it as part of the arms.

4.3 Galactic spiral arms

Our model does not assume any functional form or prior assumption in favour of the spiral arm or Galactic disk structure in our model. Any feature in our map is the outcome of the input data used for inferring the underlying dust densities coupled with the smoothness assumption from the Gaussian process. In order to draw a stronger conclusion for the arm structure of our Galaxy, velocity information from stars could be of great help. Future data from Gaia, 4MOST and WEAVE

will provide precise distances, radial velocities and proper motions for millions of stars in the Milky Way which allow us to trace the motion of stars in the Galaxy in order to reveal better the position of the spiral arms. It will be possible then to see whether the arm predictions by the stellar velocities and over-densities coincide with what the dust density probes.

Beyond stellar kinematics, HI and CO observations of gas provide longitude-velocity information which can trace the spiral arms of the Galaxy (e.g. Dame, Hartmann, and Thaddeus, 2001). Since the molecular gas and dust are mixed in the ISM (e.g. Tielens, 2005; Corbelli et al., 2012), they are expected to trace the same arm structure. However, the main challenge of using gas velocity information is that their distance estimates (kinematic distances) have large uncertainties , so the sources cannot be precisely located. There have been various studies trying to overcome this issue and determine distances to the star forming regions (Wienen et al., 2015; Whitaker et al., 2017) thought to be associated with the spiral arms. One of the major works in this aspect is that of Reid et al. (2014) who use trigonometric parallaxes and proper motion of masers associated with high-mass star forming regions that trace the spiral arms in the Milky Way (Reid et al., 2014). From their sample we select those masers that are within our probed volume (out to 7 kpc in distance and ±100 pc in the Z direction) and over-plot them on our dust map (Fig. 4.6). Some of the maser locations match inferred dust density clouds we find, especially those located in regions well-populated by the input stars. Some others, on the other hand, appear where little dust is inferred. This may be due to the fact that some regions are poorly covered as a result of APOGEE's target selection (Zasowski et al., 2013). A more likely reason is the nature of the masers: they cover much smaller physical scales (less than 1000 AU) than the resolution of our map (200 pc scale here, which is related to the typical separation of stars in the sample: see section4.1). The supposed high dust density they trace only extends over a small volume, so makes only a small contribution to the average over a larger volume.

Figure 4.7 is the same as the left panel of Fig. 4.4 but additionally shows the spiral arms from the Reid et al. (2014) model. Our predicted locations for the Scutum-Centaurus arm and the Perseus arm match that of Reid et al. (2014) quite well. In contrast, our predictions for the Sagittarius arm and the Orion spur (local arm) do not match, except for the part of the Sagittarius arm that merges with the Orion spur. These differences could be due to the observational limitations in both cases, i.e. a limited number of lines-of-sights/regions have been observed. It is also important to note that our results are determined without recourse to any

prior information on the nature or even the existence of spiral arms or any other Galactic structures. The spiral arm model of Reid et al. (2014), in contrast, contains strong assumptions for the Galactic and arm rotations and structures. Moreover, the assignment of the masers to the spiral arms based on their longitude-velocity measurements is relatively undetermined, especially for the inner arms. Only based on the masers locations in Reid et al. (2014), which represent the locations of the high-density clouds as in our map, it is extremely hard to reach a conclusion about the spiral arms.

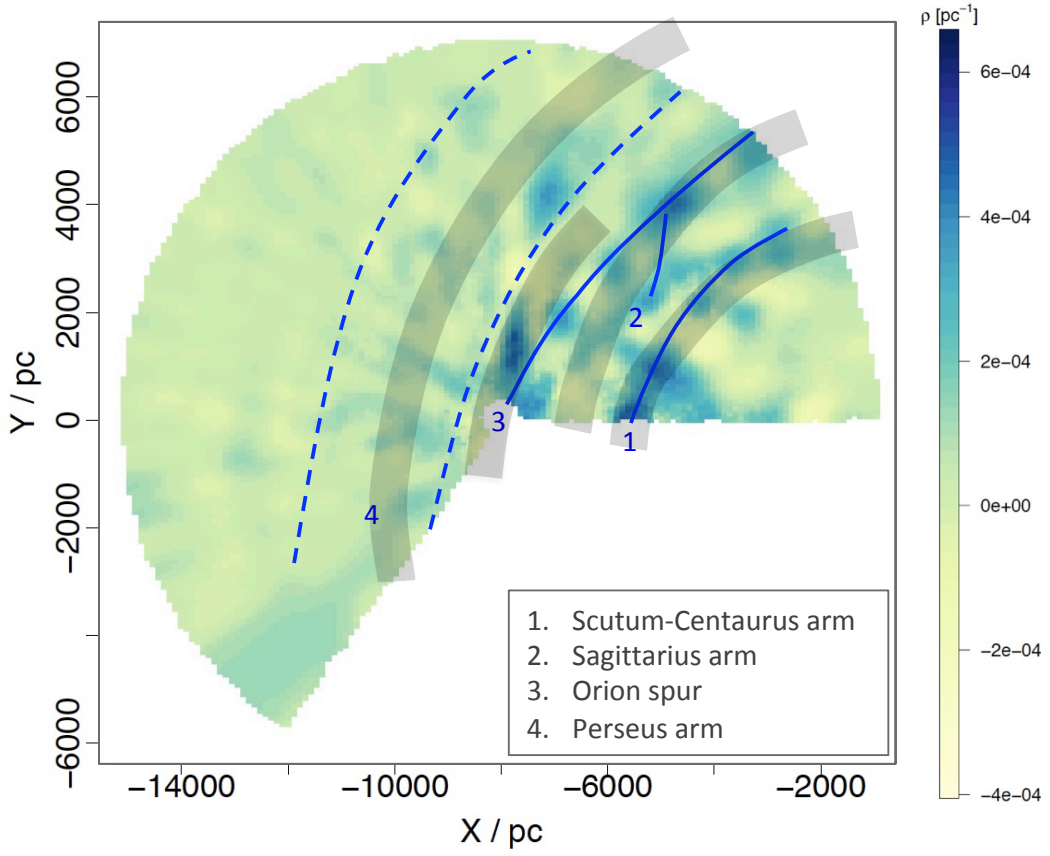


Figure 4.7: As Fig. 4.4 left, but overplotting spiral arms model from Reid et al. (2014) as grey shaded curves. The width of the shaded curves is equal to 2σ uncertainty in the arm fitting (1σ on each side of the central curve, Reid et al., 2014).

Apart from limitations in the precision of the data, there are limitations in our results due to the assumptions of the model. We assume a constant dust density within each cell to relate the measured extinction to the model dust density along

observed lines of sight (chapter 2). This, together with the separation between stars, limits the resolution of the map and can cause the map to miss smaller structures in the ISM, as discussed above with the masers.

As mentioned in chapter 2, we do not constrain our dust density predictions or the input extinctions to be positive; as a result, our dust density predictions contain negative values. This is indeed non-physical but is due to the fact that the data is by nature noisy. Both extinctions and distances are measurements which contain uncertainties. Having inconsistent distance-extinction estimates and underestimated uncertainties would cause more/larger negative predictions. In fact, in the presence of the "good" data, the predictions are positive (as seen in the results of the simulated data in chapter 3).

5

Dust distribution in the local molecular clouds

In this chapter, I present the 3D maps of our local ISM using different datasets and targets. Most of the first section is adopted from Rezaei et al. (2018, a). The second section is also part of the Gatuzz, Rezaei Kh. et al. (2018) publication.
The chapter is organised as follows. The first section concentrates on the Orion complex, first using the combination of the Gaia DR1, 2MASS and WISE, and then using Gaia DR2. The second section illustrates the 3D map of the local ($< 600pc$) hydrogen density using X-Ray data.

5.1 Orion complex

The Orion molecular complex is one of the most studied regions in the Galaxy. Containing M42, it is the closest H II region and closest site of massive star formation (O'dell, 2001). Being an on hand laboratory, Orion has been studied in various aspects; from distance estimates to different parts of the cloud to the star formation processes and individual stellar populations (e.g. Brown et al., 1994; Menten et al., 2007; Jeffries, 2007; Bally, 2008; Alves and Bouy, 2012; Bouy et al., 2014; Schlafly et al., 2015; Zari et al., 2017).
Estimates of the distance to the Orion nebula cluster, the active star-forming regions in Orion, which extend over 200 pc both in radial direction and in the plane of the sky

(e.g. Brown et al., 1994; Bally, 2008), range from 347 pc to 483 pc using optical and near-infrared photometry and colour-magnitude diagrams (CMD) (Jeffries, 2007). Other methods have found different distances to the cloud of 440 pc and 392 pc (Jeffries, 2007), 480 pc (Genzel et al., 1981), 434 pc and 387 pc (Kraus et al., 2007), and 414 pc (Menten et al., 2007). Recently, Schlafly et al. (2015) mapped the dust in Orion using Pan-STARRS1 photometry, revealing a 150 pc depth dust ring extending from around 415 pc to 550 pc in distance from the Sun.

Stellar activities in the Orion region is known also to be responsible for the creation of the Orion-Eridanus superbubble, a large cavity in the vicinity of the Orion that extends to the constellation of Eridanus in the sky (e.g. Bally, 2008; Pon, 2015). Bubble structures are very common in the ISM. They are results of the presence and evolution of young, massive stars that influence their surrounding ISM through radiations, stellar winds and supernovae explosions (e.g. Heiles, 1979; Mac Low et al., 1989). Although the projected picture of the Orion-Eridanus superbubble in the plane of the sky has been determined from various observations of the gas/dust emission (e.g. Ochsendorf et al., 2015; Soler et al., 2018), the distance to the front and back edges of the bubble is still debated.

5.1.1 Gaia Data

Most of the work I present in this chapter makes use of data from the recent European Space Agency (ESA) mission: Gaia. The Gaia astrometric satellite was launched in December 2013 aiming for providing the most accurate astrometric survey to date (Gaia Collaboration et al., 2016b). It will determine accurate positions, parallaxes and proper motions for more than a billion sources, and radial velocities for brighter stars ($G < 17mag$, Gaia Collaboration et al., 2016a). Measurements with ultimate accuracy will only be provided at the end of the post-operational phase of the mission, currently planned for 2023, but the Gaia collaboration decided to provide intermediate data releases to the community (Gaia Collaboration et al., 2016a).

The first Gaia data release (GDR1) is one of the intermediate data releases based only on 14 months of observation. GDR1 includes astrometry for about 2 million bright stars in common with the HIPPARCOS and Tycho-2 catalogue (Høg et al., 2000) to calibrate the parallaxes (Gaia Collaboration et al., 2016a; Michalik et al., 2015). The typical uncertainty in the so called Tycho-Gaia Astrometric Solution (TGAS) catalogue is about 0.3 mas for the positions, about 1 mas / yr for the proper motions, and 0.3 mas for the parallaxes (Gaia Collaboration et al., 2016a).

I use data from TGAS catalogue in section 5.1.2.

The second Gaia data release (GDR2) was provided to the community in April 2018 with precise parallaxes and proper motions for about 1.3 billion sources (Gaia Collaboration et al., 2018). Apart from astrometry, the GDR2 contains astrophysical parameters, including stellar effective temperature, extinction, reddening, radius, and luminosity, for about a hundred millions of sources (Gaia Collaboration et al., 2018; Andrae et al., 2018). A typical uncertainty for the positions and parallaxes of bright stars ($G < 15mag$) can reach down to 0.02 mas, and for the 17th magnitude is about 0.1 mas. A proper motion uncertainty for a star at $G = 17mag$ is 0.2 mas. A typical extinction uncertainty in G band is about 0.5 mag (Gaia Collaboration et al., 2018). I use the positions and extinctions from GDR2 in section 5.1.3.

5.1.2 Combining Gaia-TGAS, 2MASS, and WISE

As explained in chapter 2, we need the 3D positions of the stars, i.e. their longitudes, latitudes, and distances (l, b, d), together with their l.o.s extinction measurements to derive dust densities. Positions and parallaxes of around two million stars were published as the Tycho-Gaia astrometric solution (TGAS; Michalik et al., 2015; Lindegren et al., 2016) as part of the first Gaia data release (Gaia DR1; Gaia Collaboration et al., 2016a). Astraatmadja and Bailer-Jones (2016) used this survey to infer distances from noisy parallaxes. We use their published distances, those inferred with the Milky Way prior without the additional systematic errors, together with the Gaia positions to determine the 3D positions of stars.

We then estimate extinctions for these stars using the Rayleigh-Jeans Colour Excess method (RJCE) (Majewski et al., 2011), which uses near- and mid-infrared (NIR and MIR, respectively) data to estimate the l.o.s K_s-band extinction. Since the dereddened colour (H - $[4.5\mu]$)$_\circ$ is almost constant for different types of stars, extinction values can be calculated as

$$A_{k_s} = 0.918(H - [4.5\mu] - 0.08) \tag{5.1}$$

with an approximate extinction uncertainty of less than 0.1 mag, where $H - [4.5\mu]$ is the measured colour (Majewski et al., 2011). We use NIR photometry (J, H, and K_s) from the Two Micron All-Sky Survey (2MASS; Skrutskie et al., 2006) and the MIR photometry (W1 and W2) from WISE (Wright et al., 2010). Both catalogues are cross-matched with TGAS in the Gaia archive[1], making estimation of the extinctions

[1] https://gea.esac.esa.int/archive/

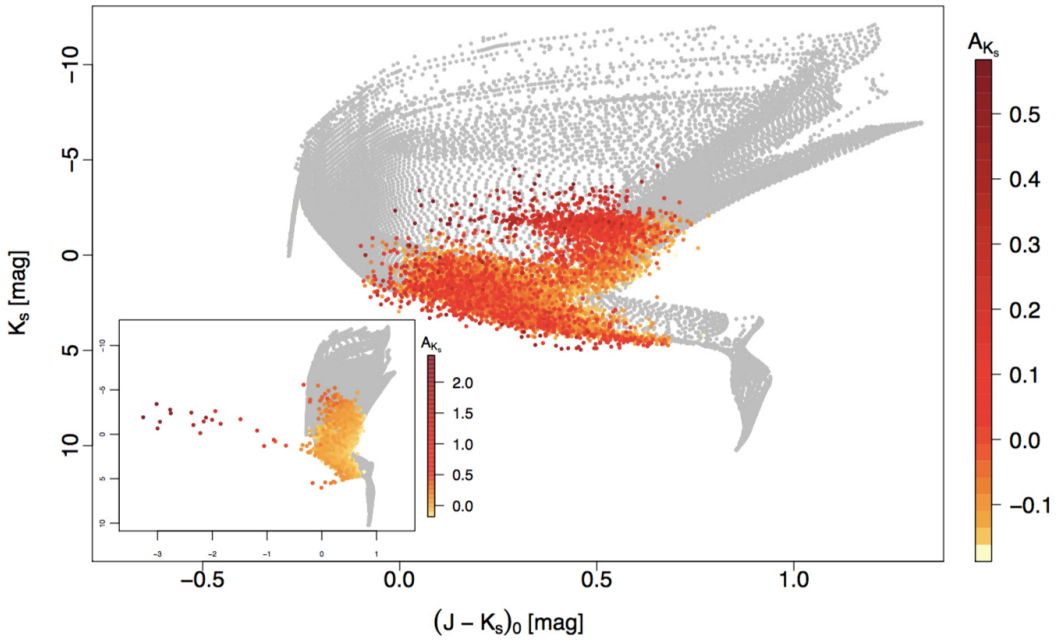


Figure 5.1: Colour-magnitude diagram (absolute K_s band magnitude vs. dereddened colour $(J-K_s)_0$) for the theoretical models (grey points) and our data (in colour). The colour indicates the derived extinctions from the RJCE method. We cut some stars that are not compatible with the theoretical models, as can be seen in the inset plot (which is the same, but extending to bluer colours). The colour scales are different in the two plots and the rare stars have spuriously high extinctions of up to about 2 mag in the K_s band.

for our sample rather trivial.

Input sample

We only select stars with fractional distance uncertainties below 0.15 to have a precise sample. This gives around 650 000 stars out to 700 pc.

From this sample, we select stars towards the Orion complex, which we defined as the region $187.5° < l < 218°$ and $-25° < b < -4°$. This leaves around 12 000 stars within distance of 100–650 pc that we use as the input data. To check the derived extinctions we can look at the distribution of stars in the CMD. Figure 5.1 shows the absolute K_s band magnitude versus dereddened colour $(J-K_s)_0$ for the input data with RJCE extinction values (coloured points) on top of the theoretical isochrones with no extinction (grey points), for which we have used PARSEC 1.2S[2] (Tang et al.,

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

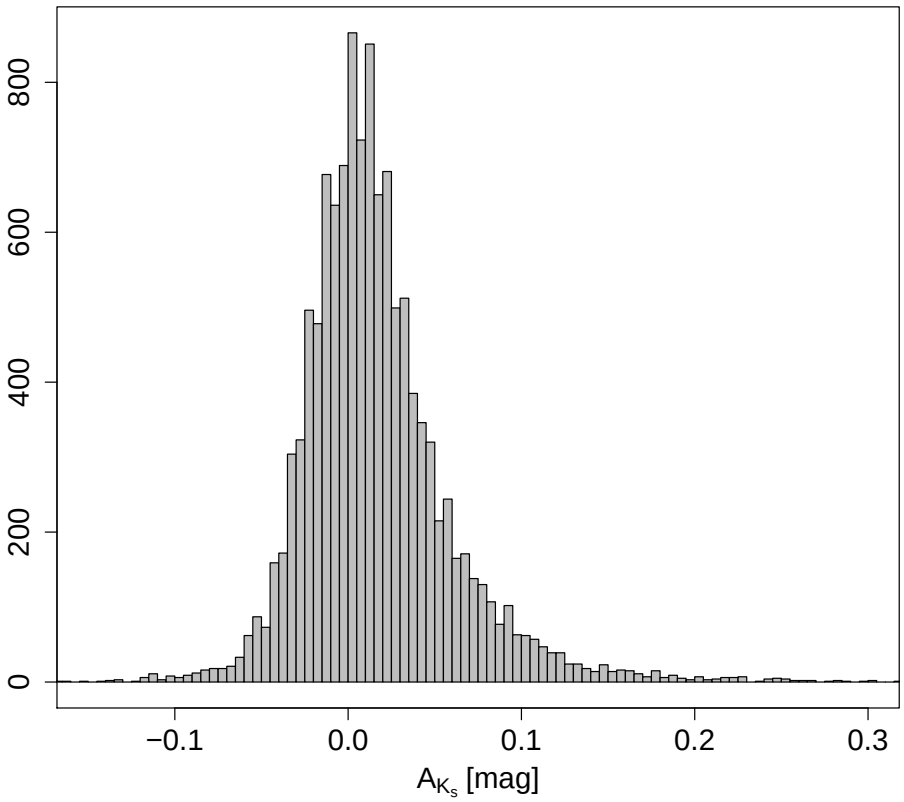


Figure 5.2: Histogram of extinction of stars. For illustrative purposes, the main plot is limited to the range -0.2 to 0.3 mag while very few stars have extinctions up to about 0.6 mag.

2014; Chen et al., 2015) to compute 2MASS JHK_s dereddened photometry for solar metallicity stars ($Z_\odot$ = 0.0152; Bressan et al., 2012), using the extinction law of Cardelli et al. (1989) and O'Donnell (1994) with R_V=3.1 (Girardi et al., 2008). As can be seen in the subplot at the lower left corner of figure 5.1, a small number of points have spuriously large A_{K_s} and accordingly extremely blue inferred intrinsic $J - K_s$ colours. Examining these stars in detail reveals that almost all of them are variable stars and young stellar objects. We choose to eliminate stars by discarding those with $J - K_s <$ - 0.3 mag (this removes 19 stars). We discuss these stars and their effects on the predictions in section 5.1.2.

As the measured extinctions are noisy, they can be negative, and in fact about

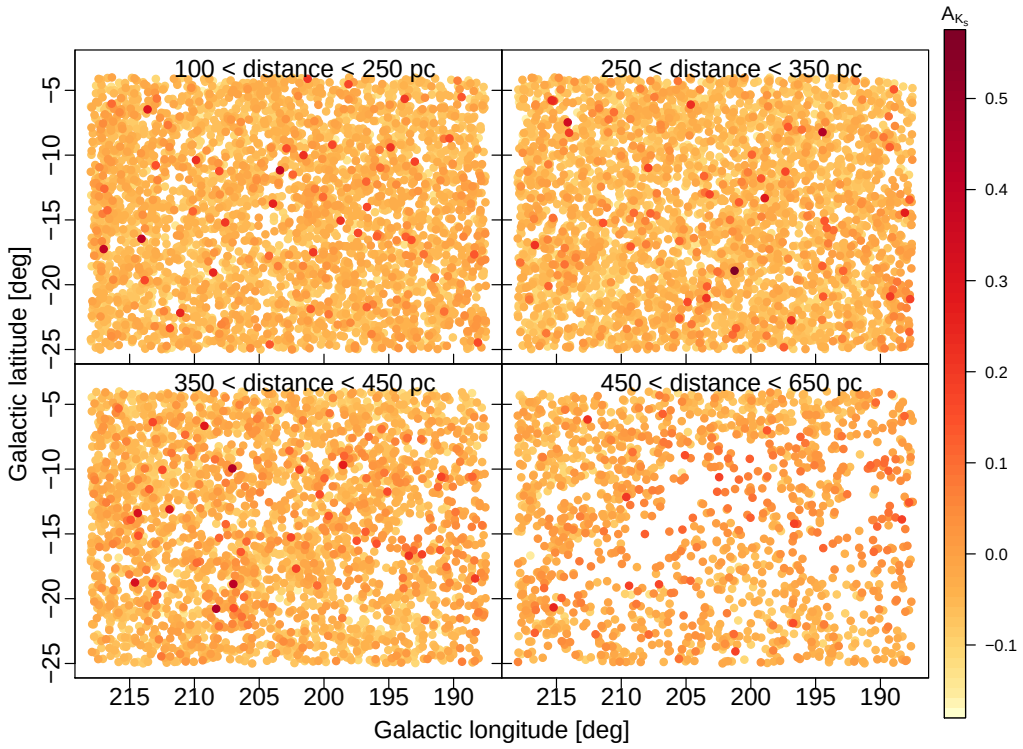


Figure 5.3: Galactic latitude vs. longitude of input stars towards Orion complex ($187.5° < l < 218°$ and $-25° < b < -4°$) for four different distance slices. Colour shows the extinctions. Most of stars have extinction values lower than about 0.3 magnitudes but there are some higher extinction values up to about 0.6 magnitudes. The Orion cloud can be identified in the lower right panel (high-extinction stars surrounding a slight void), but it is hard to see any trend in the other three panels.

40% of our sample have negative extinctions. Figure 5.2 shows the histogram of extinction for all input stars. The A_{K_s} values go up to about 0.5 magnitudes but most of the stars have extinction values lower than about 0.2 magnitudes (around 2 magnitudes in V band). Unlike in our previous work (Rezaei Kh. et al., 2017), we use all these data (including negative extinctions) in our model to predict the underlying dust density. Subsequently, in section 5.1.2, we discuss the effects of input stars with negative extinctions on the predictions by running the model only for positive extinctions. The RJCE method assumes three sources of uncertainty in the predicted extinctions: intrinsic scatter in the stellar colour, photometric uncertainty, and uncertainties in the extinction law that they used. Combining all these sources, (Majewski et al., 2011) estimate extinction uncertainty of less than 0.11 mag in K_s band for a typical individual star, while being more precise for red clump stars and

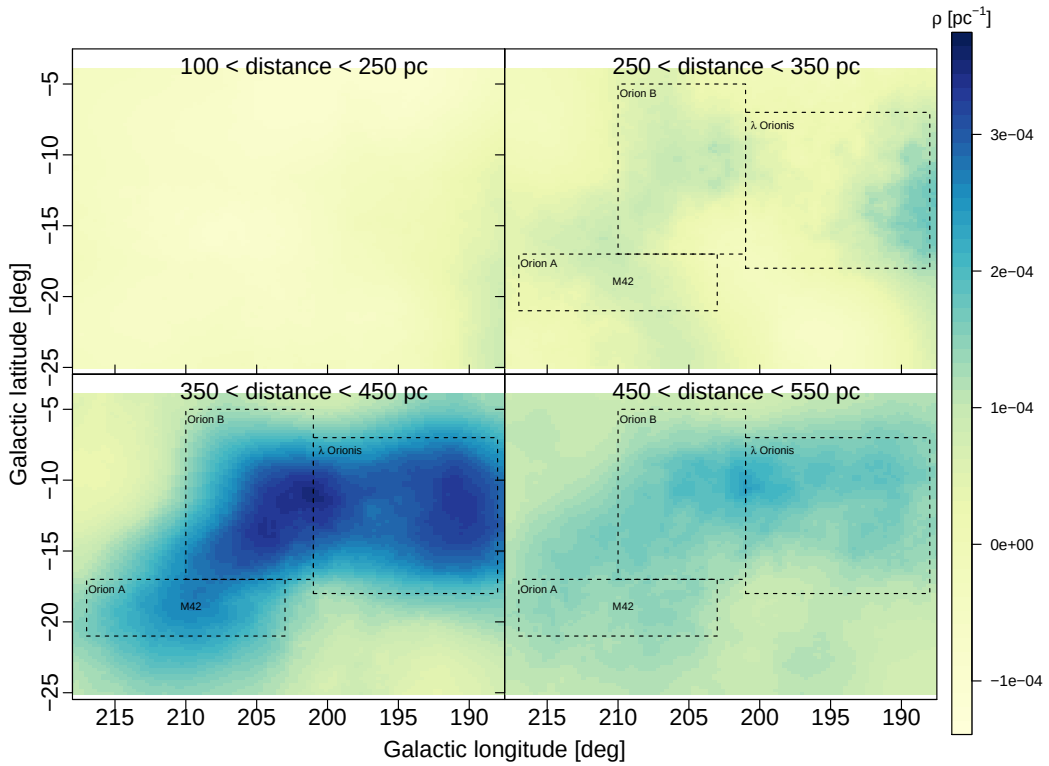


Figure 5.4: Dust density predictions for 50 000 points randomly distributed towards the Orion complex (as in figure 5.3), where we plot by taking the mean of each $0.3^\circ \times 0.3^\circ$ grid, then smoothing the plot with a Gaussian function with a scale parameter of 0.4 to cover empty pixels. Each panel shows a different distance range. Green and blue colours show the points with higher dust densities. Orion A, Orion B, and λ Orionis are denoted by dashed lines Lombardi et al. (2011).

giants. Since we do not have any selection on types of stars, we assume the upper limit of the extinction uncertainty for all stars in our sample. We discuss the effects of different extinction uncertainties on our predictions in section 5.1.2.

Figure 5.3 shows the spatial distribution of the input data (latitude versus longitude) for four different distance slices (less that 250 pc, then 100-pc slices up to 450 pc and 450–650 pc slice), colour-coded by their extinction values. We use these values as the input data for our model, assuming an extinction uncertainty of 0.1 mag for all stars. Using this we calculate the underlying dust densities for 50 000 points randomly distributed in space towards the Orion complex with distances between 100 pc and 600 pc.

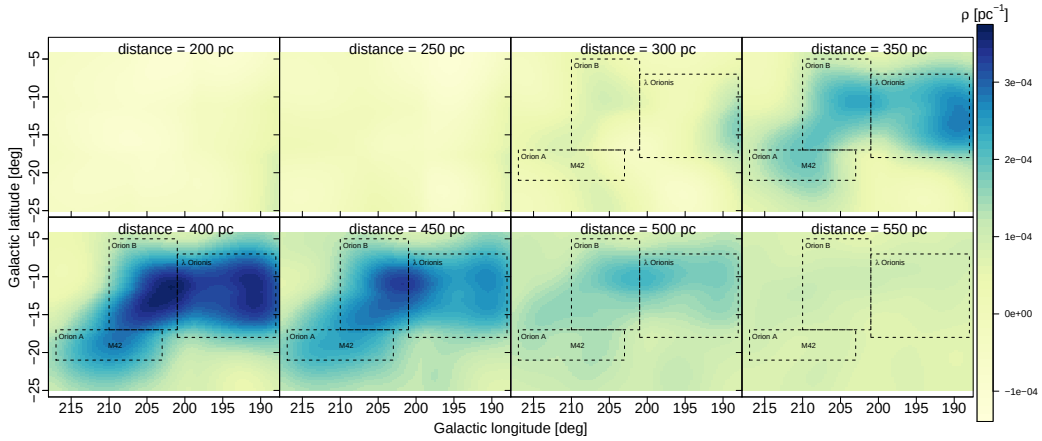


Figure 5.5: Dust density predictions for 10 000 points at fixed distance distributed on a regular grid. There is no smoothness or interpolation for plotting and the features are the direct outcome of our predictions. Orion A, Orion B, and λ Orionis are indicated by dashed lines Lombardi et al. (2011).

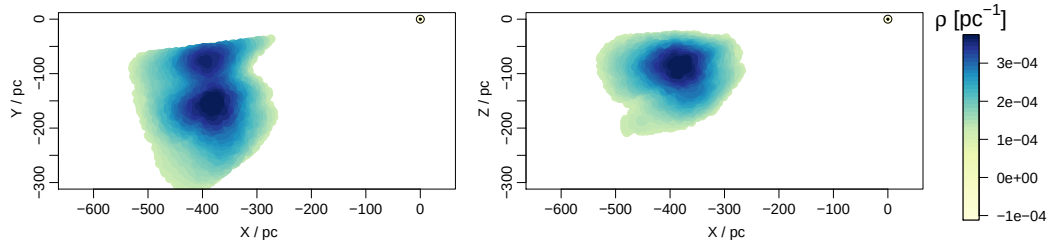


Figure 5.6: Two Cartesian projections of the 3D dust distributions in Orion. Only dust densities higher than the mean density of the predictions (1.2 $\times 10^{-4}$ pc^{-1}) are shown. The Sun is at (X, Y, Z) = (0 , 0, 0), with X increasing towards the Galactic centre and Z point to the north Galactic pole, perpendicular to the Galactic disk.

Dust density predictions

Figure 5.4 shows the dust density predictions for the same panels as figure 5.3. Each panel represents the mean of the dust density predictions for the corresponding distance range. There is not much dust seen out to 250 pc (upper left panel). The dust density starts to increase from the second (250–350 pc) up to the last (450 – 550 pc) distance range, suggesting that the Orion cloud in some parts is extended for about 200–300 pc. In the lower left panel, the clear detection of Orion A and Orion B towards λ Orionis is evident (as indicated by dashed lines from Lombardi et al., 2011), where Orion B and λ Orionis seem to extend to distances of larger than 450 pc. Figure 5.5, on the other hand, shows the exact value of the predictions at fixed distances (in 50 pc steps). The first traces of over-densities starts appearing

at 300 pc and extends up to at least 500 pc. It is worth mentioning that we restrict our data to the fractional distance uncertainty of < 0.15, which on average is about 50 pc at 400 pc distance.

In order to see the structures and locations of different parts, we need to look at their spatial distributions. Figure 5.6 shows 2D projections of the dust from two directions: a top-down view (i.e. looking from above the Galactic plane) and a side view. For better visualisation, the plot only shows points with dust predictions higher than the mean density of the predictions (1.2×10^{-4} pc^{-1}). We see that the front part of the higher density region starts as close as 300 pc and extends to 550 pc at the direction of the Orion A, Orion B, and λ Orionis.

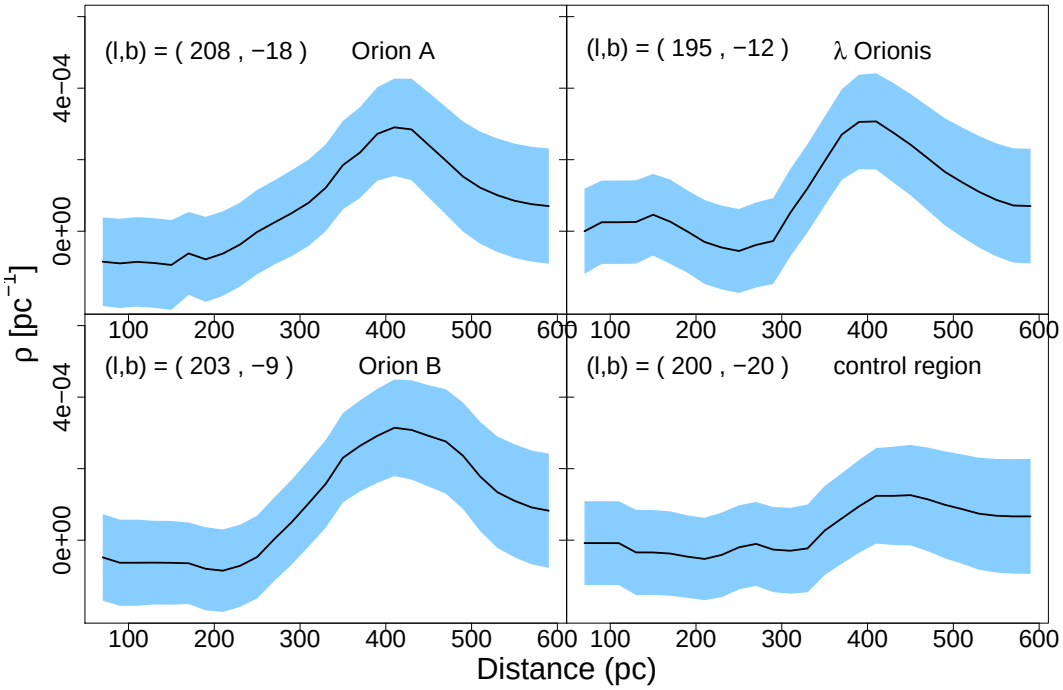


Figure 5.7: Dust density vs. distance for four different l.o.s.: three towards known regions of Orion and one towards a control region where we do not expect a cloud. The black line shows the mean and the blue shades represent one standard deviation (also computed by the Gaussian process model).

Our model predicts a Gaussian distribution for our knowledge of the dust density at every point in space, from which we can extract an estimate (the mean, which has been plotted so far) and its uncertainty (standard deviation). Figure 5.7 shows the variation of both of these quantities along different l.o.s. Each panel of the figure

shows one l.o.s towards Orion A, Orion B, and λ Orionis, plus one l.o.s outside these regions where we do not expect to see a prominent peak in the dust density. The dust density in Orion A and B starts increasing slightly after 200 pc, while dust towards λ Orionis seems to be more concentrated, increasing only after 300 pc. The dust density in Orion A decreases to the input mean value by a distance of 500 pc; however, in Orion B the density remains high to larger distances, which demonstrates a more extended dense part at the direction of Orion B.

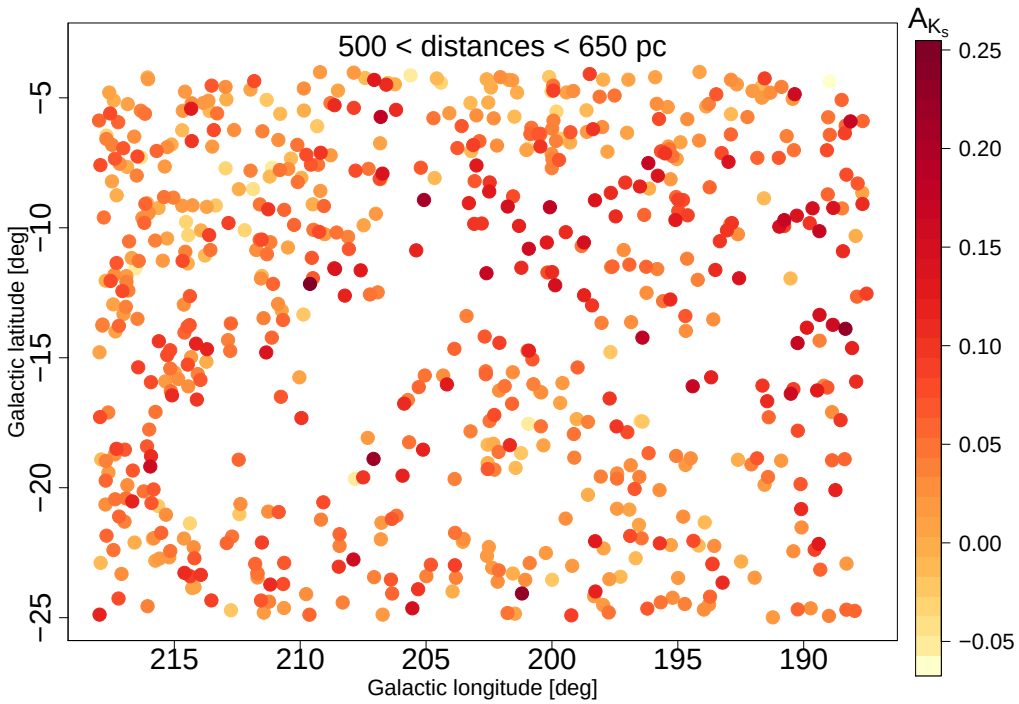


Figure 5.8: Latitude vs. longitude of input stars towards Orion complex further than 500 pc colour coded by their input extinctions. A region extending down and left of centre and down and right of centre is devoid of stars due to the high foreground extinction. The extinctions of the stars at the edge of this region are increasing towards the centre of the region.

Figure 5.8 shows the input stars that lie behind the Orion cloud (further than 500 pc), colour coded by their extinctions. The first feature from this plot is the gap – missing stars – extending from the centre of the plot to the lower left and right. This is a consequence of stars behind the dense regions of the Orion complex being highly extinct and thus too faint for our data sample, on account of the magnitude limit of Gaia-TGAS ($G \simeq 13.5$, although incompleteness sets it at brighter magnitudes). The second feature is the increase of the extinction towards the edges of these missing parts, from which we can better identify the location of the foreground obscuring

cloud. However, this makes our model underestimate the predictions in these highly extinct regions where stars are missing, and makes us uncertain of the back edges of the clouds.

Discussion

Our results indicate that the Orion molecular complex includes a higher density dust region starting as close as about 300 pc and extending up to around 500 pc, suggesting a depth of 200 pc for the complex. We compare our results with the recent work of Schlafly et al. (2015) in figure 5.9, which shows the same distance slices as in the first two panels of figure 2 of Schlafly et al. (2015). The colour shows the integrated dust densities through the cloud (A_{K_s} in mag. which is approximately 0.1 times A_V). There is a good overall agreement. Most of the dust we infer is at distances between 300 pc and 600 pc and the three main parts related to Orion A, Orion B, and λ Orionis are obvious in the right panel. We also over-plot contours of E(B - V) = 0.5 mag from the Schlegel, Finkbeiner, and Davis (1998) (SFD) 2D reddening map. This matches our dust density predictions fairly well, except for the regions closer to the Galactic disk and larger longitudes where the dust probed by SFD is further away and beyond our probed distance range.

If we compare these reconstructed extinctions towards the Orion complex with the input extinctions (figure 5.3), we see the maximum A_{K_s} value in figure 5.9 is only about 0.1 mag while the input extinctions extend to 0.6 mag. This is the expected outcome of our isotropic Gaussian process prior with a particular scale length (100 pc here), which considers correlations in all directions. This means that each point is affected by the surrounding points within the 100 pc correlation length. Therefore, to compare the predicted and input extinctions, we need to estimate the average extinctions within the 100 pc radius in figure 5.3. The average extinction over 100 pc radius for most of the regions in figure 5.3 is around 0.05 ± 0.05 mag, which is in agreement with the reconstructed extinctions.

As explained in previous chapters, we notice that some dust density predictions are negative. This does not have a physical meaning, but it is a consequence of the Gaussian process assumption and the noisy input data. We could put some stronger constraints (priors) on the predictions (e.g. with truncated functions) to get only positive values. But we chose not to do this, as these negative values are informative and show inconsistencies in the input extinctions and/or distances. This is because the integral over dust density predictions along the l.o.s towards each star needs to be equal to the input extinction for the likelihood model. Therefore,

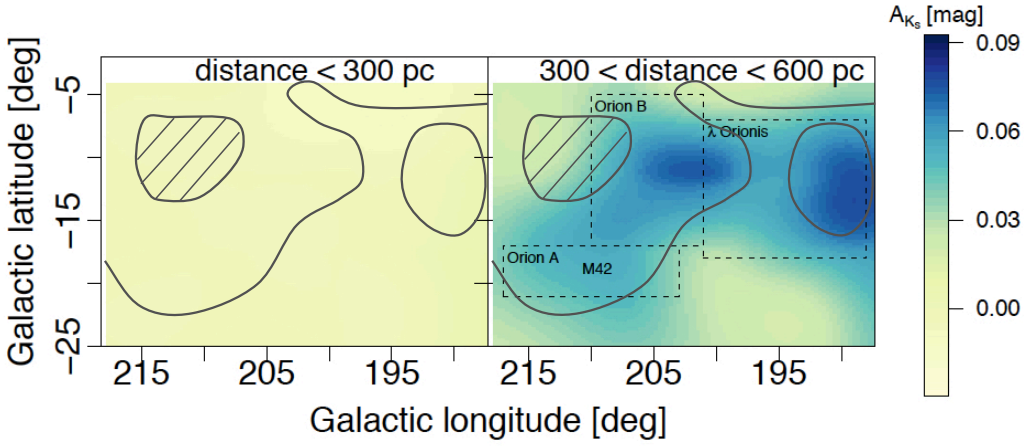


Figure 5.9: Integrated dust density predictions (i.e. extinctions A_{K_s} in mag.) for the Orion complex for the same two distance slices as shown in figure 2 of Schlafly et al. (2015). Most of the extinction is at distances more than 300 pc. Orion A, Orion B, and λ Orionis are indicated by dashed lines (Lombardi et al., 2011) and the solid contours show the E(B-V) = 0.5 mag from the 2D dust map of Schlegel, Finkbeiner, and Davis (1998) (SFD), where the hatched area indicates lower reddening.

if there are some extinctions that cause the model to predict higher dust densities than what is expected for more distant stars (along or near to the l.os.), then the model has to predict some negative dust densities to stay consistent with other input extinctions within their error bars. This is evident from figure 5.10, where not only the impact of the very high input extinction values appears as an increase in the maximum predicted extinction, but also they produce maps with much larger negative dust densities than before, which is another indication that the some of the input extinctions are unrealistically high.

Another point to consider is that by using the Gaia-TGAS catalogue, we underestimate the amount of dust in the Orion complex. This is because even a modest amount of extinction means stars do not appear in Gaia-TGAS because of its relatively bright magnitude limit ($G \simeq 13.5$). Thus we can only see the cloud to a limited depth. This will improve with the use of Gaia DR2, which will reach around seven magnitudes deeper than the TGAS catalogue. With Gaia DR2 we will be able to make a deeper and more precise 3D view of the Orion cloud and infer more accurate distances towards its different parts. Moreover, the extinction estimates in Gaia DR2 have uncertainties of order 0.2 mag in the G band (Andrae et al. 2018 A&A submitted), which is equivalent to about 0.02 mag in the K band; this is five

times smaller as used in this study, although there are additional issues (such as systematics) with using the Gaia DR2 extinctions.

Discussion 1 – Spurious high-extinction sources

As mentioned briefly in section 5.1.2, input sample, using the observed colours of stars and RJCE method results in having extremely high extinction values for 19 out of 13 000 stars, mostly located at the lower centre of the second and third distance slices of figure 5.3. We decided to discard these stars based on the CMD and the theoretical models. To study these stars we include them in the sample and predict the dust densities one more time to see their effects on the results. Figure 5.10 shows the resulting map.

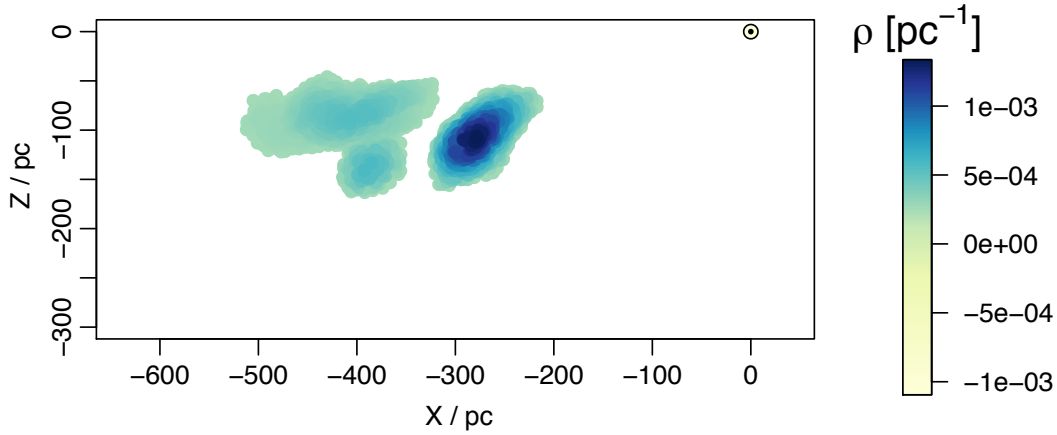


Figure 5.10: Effects of the spuriously blue stars with extremely high extinction values on the predictions; there is a foreground cloud in front of Orion. The axis and view angle are the same as in figure 5.6, right panel.

The main change (cf. 5.6, right panel) is the appearance of a distinct cloud in front of Orion A. Looking up these stars on SIMBAD suggests that many of these supposedly high-extinction sources are young stellar objects in the foreground of Orion and are associated with bursts of star formation in Orion over the last several million years (e.g. Bouy et al., 2014; Bally, 2008; Brown et al., 1994, 1995). These stars have MIR excesses, which are due to their circumstellar dust rather than interstellar dust, leading to spurious extinction predictions from RJCE, for which stars are assumed to be older. Another reason for having such extreme, unlikely extinction values is our approach to deriving colours for these stars. Since we obtain NIR and MIR

photometry from different surveys (2MASS and WISE), and as these surveys did not collect their data at the same time, stellar variability could result in erroneous colours ($H - K_s$ colour of more than 2 magnitudes). We therefore used the location of stars in the CMD to help assess the validity of their dereddened colours. This experiment demonstrates that using only colours of stars without other constraints such as the CMD can cause unexpected results because the colours might have been affected by different factors which could, in our case, noticeably affect the predictions.

Discussion 2 – Negative input extinction effects

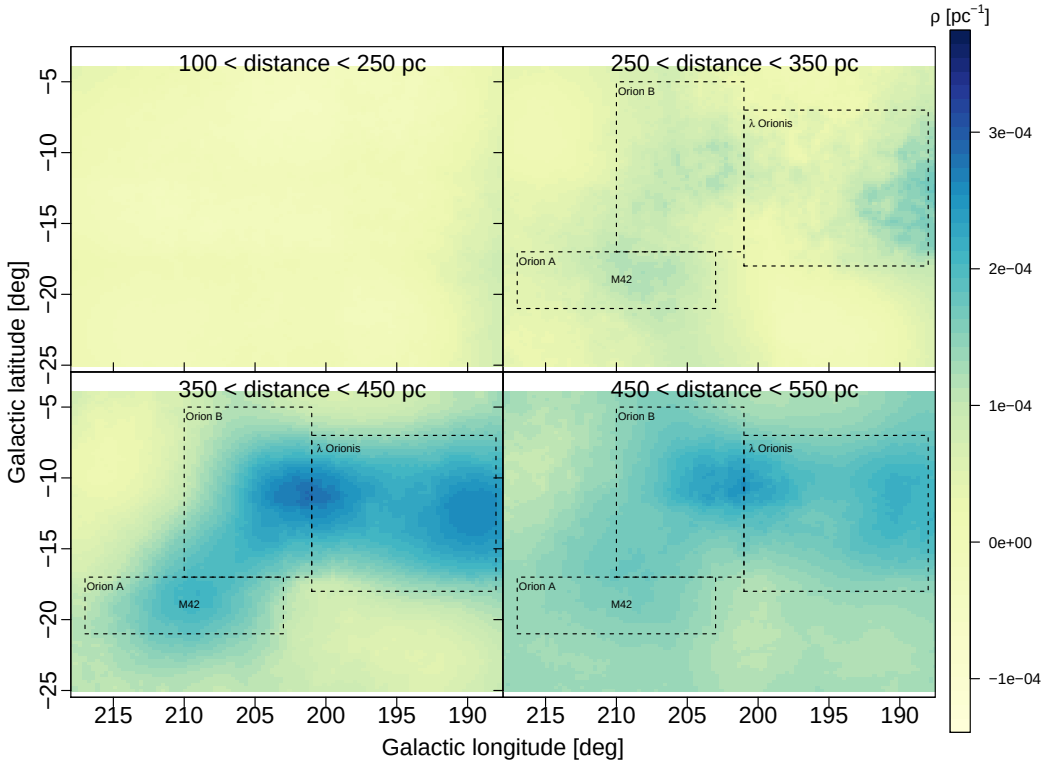


Figure 5.11: Dust density predictions as in figure 5.4 but using only stars with positive extinctions as the input. The colour bar is the same as in figure 5.4 enabling direct comparison. Orion A, Orion B, and λ Orionis are indicated by dashed lines Lombardi et al. (2011).

As mentioned in section 5.1.2, input sample, around 40% of stars in our sample have negative extinctions (see figure 5.2); we have included these stars in our analysis so far. To see the effects of these negative extinctions, we remove them from the sample

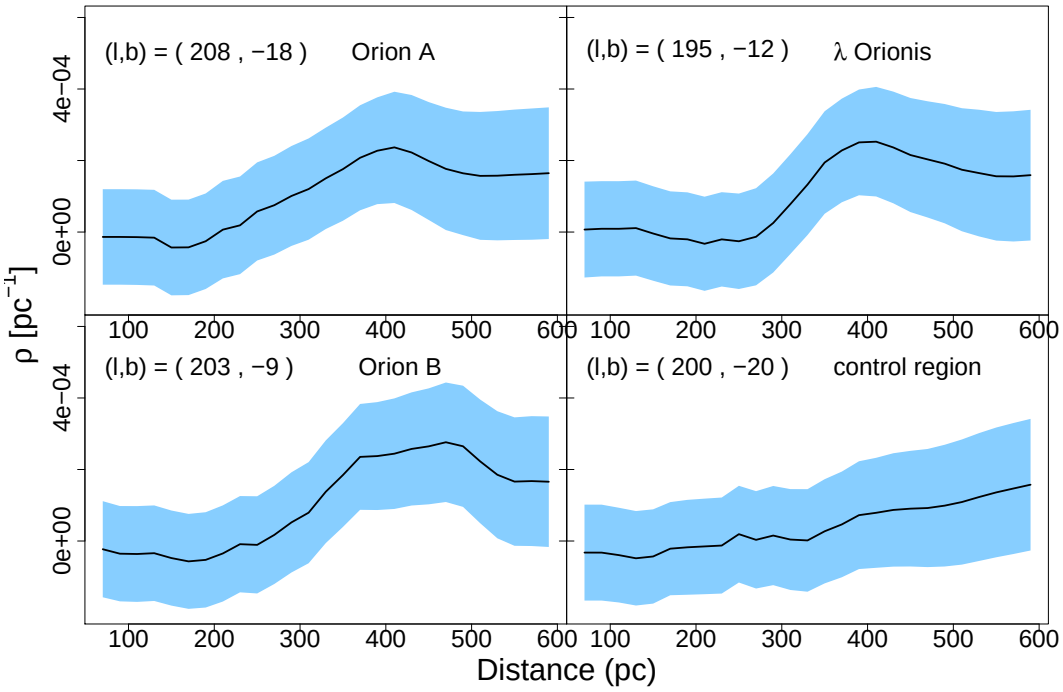


Figure 5.12: As Fig. 5.7 but using the positive extinctions to infer the dust densities.

and redo the analysis using only positive extinctions. This makes the input mean density almost twice as high as when including negative extinctions. Figure 5.11 shows the predictions using only positive extinctions with the same colour range as in figure 5.4. The results look rather similar. They trace similar structures in the dust densities, although the range of the dust density is now narrower. This can be seen better by comparing figures 5.7 and 5.12, where figure 5.12 shows dust density predictions per l.o.s for the case of using only positive extinctions. This decrease in dust amplitude arises because discarding negative extinctions decreases the contrast between high and low density regions. Excluding negative extinctions produces less negative dust densities overall.

Discussion 3 – Input extinction uncertainties

As mentioned in section 5.1.2, input sample, we have assumed for all input stars a constant extinction uncertainty of 0.1 mag, which is the upper limit of the uncertainties expected from the RJCE method. Here we investigate the effects of input extinction uncertainties on our density predictions. Figure 5.13 shows l.o.s predic-

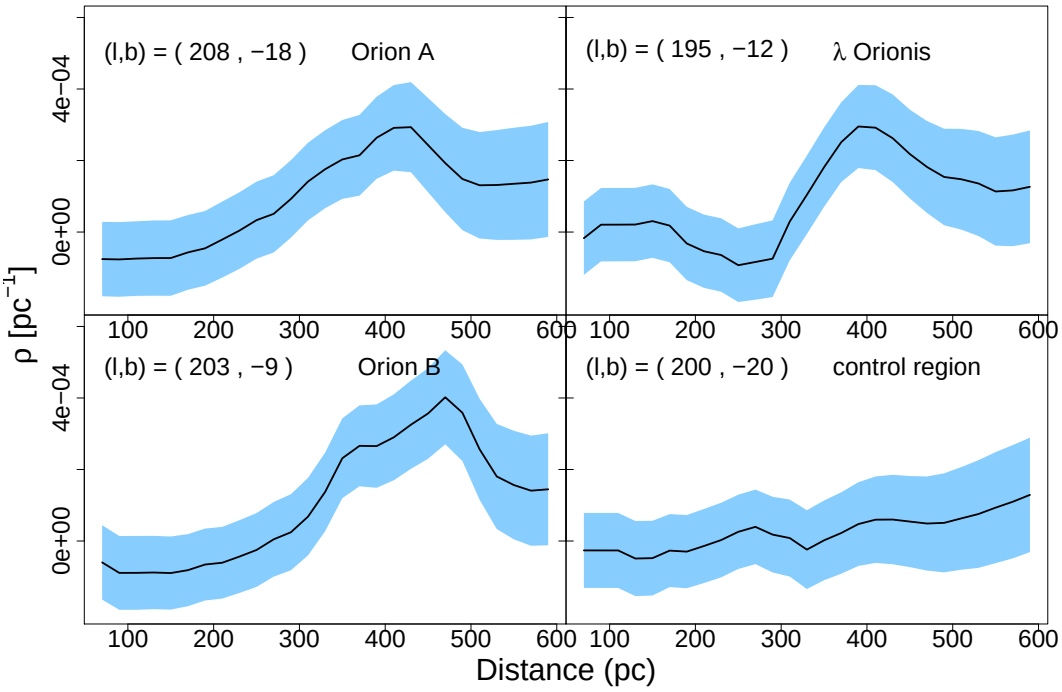


Figure 5.13: As figure 5.12 but from using uncertainties on the extinctions of 0.05 mag as opposed to 0.10 mag.

tions as in figure 5.12 using only positive extinctions with a smaller uncertainty of 0.05 mag. The predictions are less smooth and the error bars are smaller. This is as expected. Having smaller uncertainties in the extinction values indicates a sharper likelihood, which therefore has more impact on the posterior than does our Gaussian process prior. There are also more negative density predictions at lower distances, which is because dropping the uncertainties make input extinctions less consistent within their error bars, therefore predicting negative values at lower density regimes.

5.1.3 Using Gaia DR2

The Gaia DR2 contains G-band extinctions that are derived using BP - RP colours and parallaxes, together with the HR diagram constraints (Andrae et al., 2018). However, since there are strong degeneracy between the extinctions and temperatures of stars, the catalogue contains extinctions for only about 88 million sources. The validation of the Gaia DR2 extinction estimates for the Red Clump (RC) stars

demonstrates the reliability of those measurements (Andrae et al., 2018).

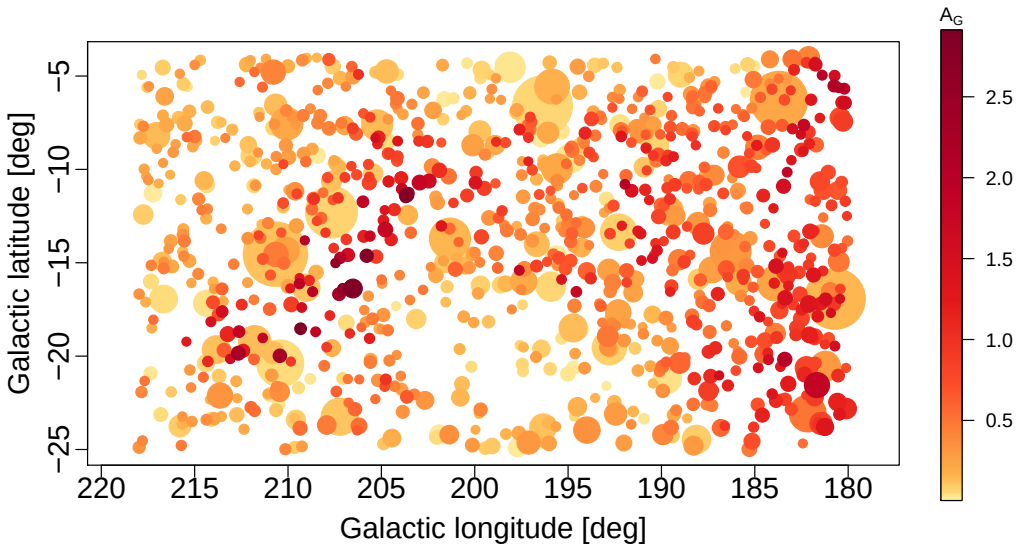


Figure 5.14: G-band extinction estimates from Gaia DR2 (Andrae et al., 2018; Gaia Collaboration et al., 2018) in the plane of the sky. Colour shows the extinction value and the size of the points represents their distances: close points are larger and further away ones are smaller.

We, hence, select only RC stars towards the Orion region ($180° < l < 218°$ and $-25° < b < -4°$) from the Gaia DR2 catalogue using the colour, absolute magnitude and colour cut as explained in the section 7.3 of (Andrae et al., 2018). We join the Gaia DR2 catalogue of RC stars with the distance estimates from Bailer-Jones et al. (2018) and limit the distances to 700 pc to focus on the Orion complex. This selection criteria leaves us with about 1000 stars spread all over the area. The G-band extinctions for the sample ranges from 0 to 3 mag as seen from figure 5.14. We use this sample as our input data to derive the corresponding dust distribution in the area.

Dust density predictions

Figure 5.15 shows the predicted dust densities towards the Orion for distance slices at every 25 pc. There is a dust cloud centred at (l , b) = (208 , -14) that appears to be at closer distances than the main Orion complex, peaking at about 270 pc. It seems to be connected to Orion B which peaks at around 400 pc, while Orion A is located at larger distances of about 450 pc. The distance to λ Orionis looks to be

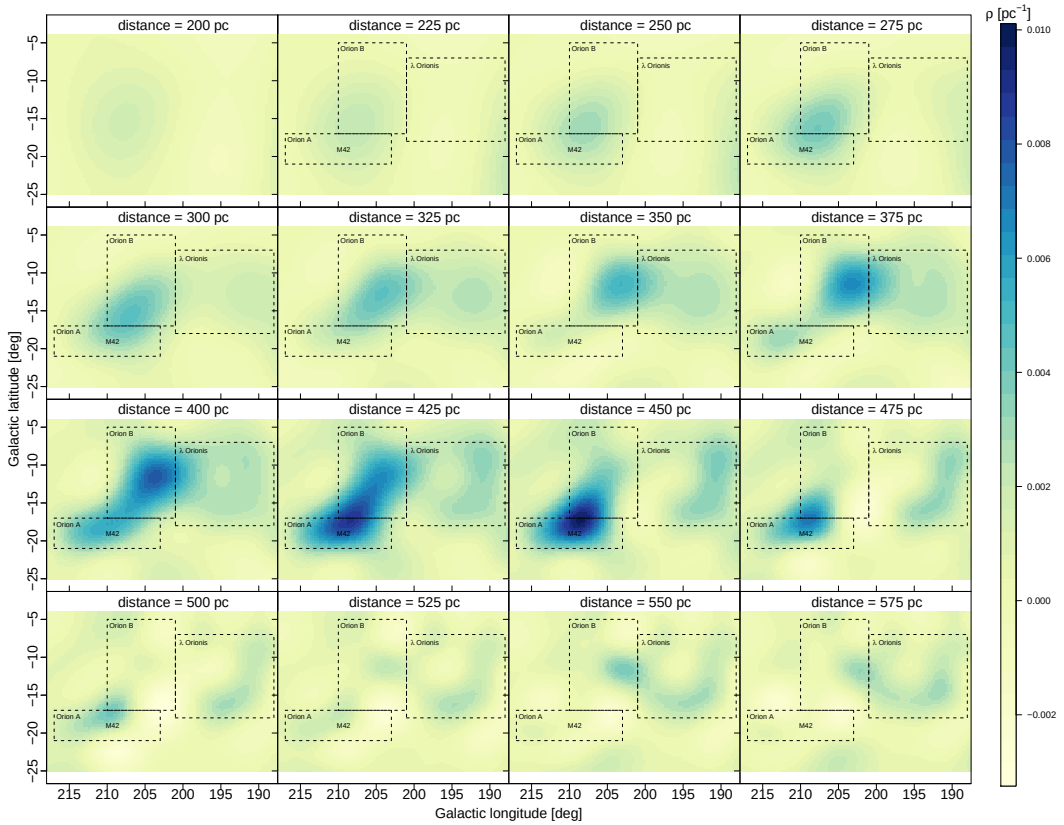


Figure 5.15: Dust density predictions at fixed distances for every 25 pc using GDR2. The boundaries of the Orion A, Orion B, and λ Orionis are marked by dashed lines (Lombardi et al., 2011). The presence of a foreground cloud to Orion A, and a background cloud to Orion B is evident in the first and last rows.

somewhat in between Orion B and A distances. These interpretations can be seen better from figure 5.16 which shows the 2D projections of the 3D distribution from two different view angles. The foreground cloud and its connection to Orion B can be seen clearly in this figure. Moreover, this results suggest that the dust in/around the Orion region consists of multiple dust clumps with different size and density that can be related to the young OB stellar associations in the area (cite).

In the previous figures, we only plotted the mean of our density predictions. To investigate the uncertainties of the predictions and the significance of the predicted dust clouds we look at the certain lines-of-sight through the Orion region and plot the predicted dust densities with their uncertainties as a function of distance, as illustrated in 5.17. The presence of the foreground cloud in front of the Orion A (in

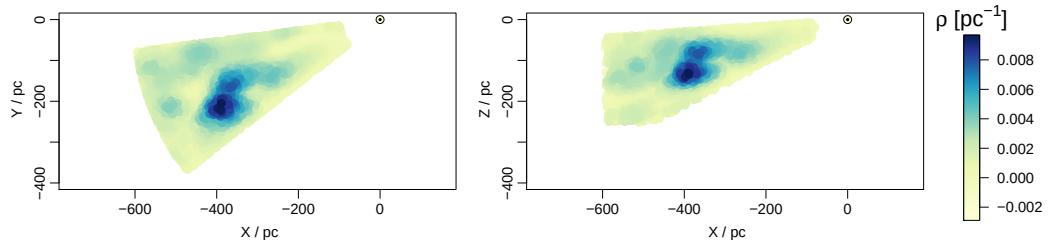


Figure 5.16: Two Cartesian projections of the 3D dust distributions in Orion using GDR2. The Sun is at (X, Y, Z) = (0 , 0, 0), with X increasing towards the Galactic centre and Z point to the North Galactic pole, perpendicular to the Galactic disk. Clumps of dust are seen in the Orion region from different angles.

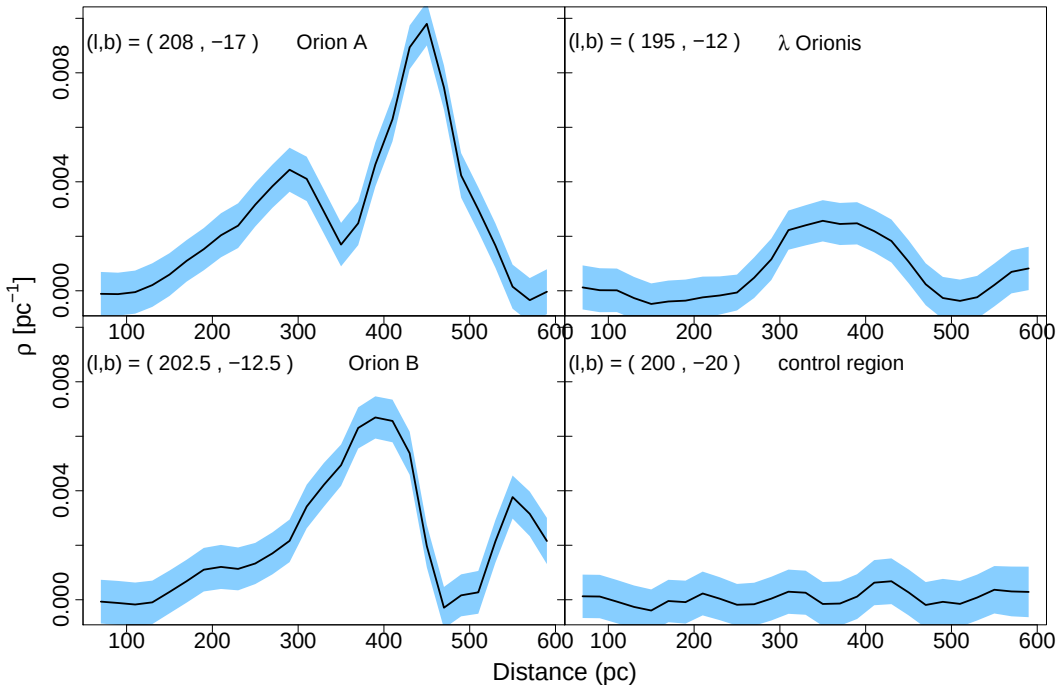


Figure 5.17: Dust density vs. distance for four different l.o.s. using GDR2: three are towards known regions of Orion and one towards a control region where we do not expect a cloud. The black line shows the mean and the blue shades represent one standard deviation (also computed by the Gaussian Process model). The foreground cloud to Orion A as well as the background cloud of the Orion B are clearly detected here.

l.o.s projection) is evident.

Discussion

These results shows multiple dust clumps spread around the area and suggest that there is a dust cloud as close as 275 pc in the direction of the Orion complex which can put a constraint on the distance to the back edge of the Orion-Eridanus superbubble. However, looking back at figure 5.14, there is only one star deriving that foreground cloud. In another words, selecting only red clump stars towards the Orion region with precise parallax measurements provides only around 1000 stars in this area, resulting in low statistics. We can not claim the foreground cloud based only on these few stars. The extinction measurements for the red clump stars are validated, yet expanding the sample to include other stellar types at this point is not possible. This is due to the fact that extinction measurements from GDR2 contain some systematic artefacts and is not suggested to be used for star-by-star base analysis (Andrae et al., 2018).

Future work on deriving better extinction measurements for Gaia sources, e.g. using multi-band photometry from other surveys, will provide a more reliable sample to find a precise dust distribution towards the Orion-Eridanus superbubble. Since the OB associations are thought to be responsible for the production of the superbubble (e.g. Schlafly et al., 2015; Pon et al., 2014), we look at the known OB associations towards the Orion complex. There are distance estimates to various groups of OB associations ranging from about 330 pc to 500 pc (e.g. Brown et al., 1994; Jeffries, 2007; Menten et al., 2007; Zari et al., 2017). However, if the discovered foreground dust cloud would be confirmed, it is in line with the foreground population of stars reported by Alves and Bouy (2012); Bouy et al. (2014). Further studies on the distance of the young stellar populations towards the Orion region are needed in order to find the possible link between our predicted dust clumps, in particular the foreground and background clouds, and the young stellar associations.

5.2 3D local hydrogen density map using X-Ray spectra

In this section, I present a study of the 3D hydrogen density, $n_{\rm H}$ [cm^{-3}], distribution in the Milky Way, using equivalent hydrogen column density, N(H) [cm^{-2}], obtained from the X-ray Multi-Mirror Mission (XMM-Newton) spectral fitting, using the parallax measurements obtained by Gaia. The angular resolution of the XMM-Newton is about 6-arcsec FWHM, that, compare to that of the 21-cm all-sky surveys ($\sim$ 36 arcmin), allows a study of the small-scale ISM structures.

The Exploring the X-ray Transient and variable Sky project (EXTRAS, De Luca

et al., 2016) provides a catalogue of X-ray spectral fitting parameters for observations within the XMM-Newton catalogue. The current version of EXTRAS includes 137212 observations from the XMM-Newton Data Release (3XMM-DR6, Rosen et al., 2016). A phenomenological classification scheme, developed by the EXTRAS team, allows the identification of a given source according to their spectral properties. The scheme is based on the random forests probabilistic method developed by Breiman (2001) which has been successfully applied to classify X-ray sources (Lo et al., 2014; Farrell et al., 2015). This provides us with hydrogen column density, $N(\mathrm{H})$, towards individual sources.

The first Gaia data release Gaia Collaboration et al. (GDR1[3] 2016a) contains parallax measurements for 2.05 million sources. It is important to note that the estimation of distances needs to be done in a proper way in the sense that inverting the parallax is only valid in the absence of noise and therefore the distance calculation should be treated as an inference problem. Therefore, we used the distances and uncertainties inferred by Astraatmadja and Bailer-Jones (2016, hereafter AST16) using the Milky Way Prior.[4].

We performed a cross-matching between the EXTRAS and AST16 catalogues by computing the angular distance between sources. For each source in our sample we assigned the distance obtained from the closest AST16 source. The upper limit in the angular distance is determined from the XMM-Newton angular resolution ($<$ 12.5 arcsecond). We found a mean angular distance in the cross-matching of 2.1 arcsecond. At this step we noted that (i) numerous sources have large column density errors and (ii) for sources with number of counts $< 10^2$ we are not able to recover low column densities ($<10^{21}$ cm^{-2}). In order to build a final sample, we decide to exclude sources with number of counts $< 10^2$ and those with multiple observations and $\Delta N(\mathrm{H}) > 50\%$. This final sample consists of 2128 sources and is used in the following analysis.

Figure 5.18 shows the Galactic distribution of the sample in Aitoff projection for different distance ranges. The colours indicate the hydrogen equivalent column density for each source (obtained from the X-ray fitting procedure described above), in units of 1×10^{22} cm^{-2}. For illustrative purposes, all sources with equivalent column densities larger than 1×10^{22} cm^{-2} show the same colour. Although sources are concentrated near the Galactic plane, the sample also includes high latitude sources.

[3]This is a proof-of-concept work that used DR1. We are in the process of analysing the DR2 data.

[4]http://www.mpia.de/homes/calj/tgas_distances/main.html

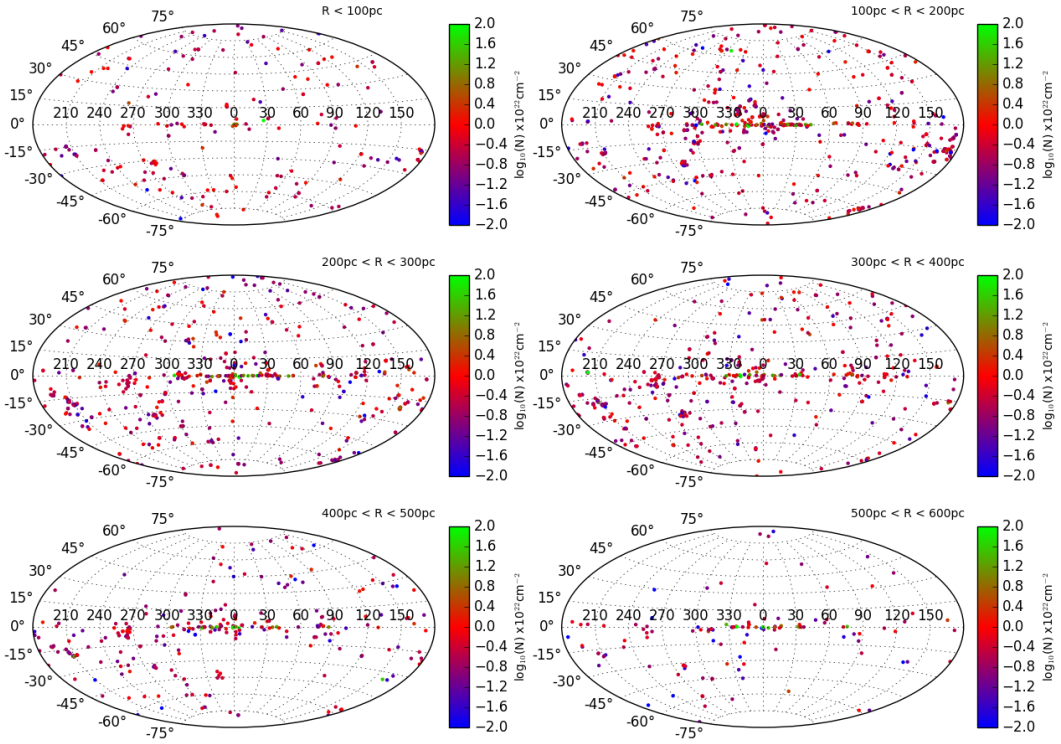


Figure 5.18: Galactic distribution of the X-ray sources in Galactic coordinates.

5.2.1 3D MAPPING OF THE ISM NEUTRAL ABSORPTION

Two regimes can be distinguished when studying the H I distribution in the Milky Way: large-scale and small-scale structures. Large-scale structure refers to the global distribution of the gas assuming hydrostatic equilibrium conditions. Using the equivalent $N(\mathrm{H})$ values obtained from the EXTRAS catalogue the gas distribution can be modeled according to the equation

$$N_i = \int_{\text{observer}}^{\text{source}} n_i(r) dr \tag{5.2}$$

where r is the distance along the i l.o.s and $n_i(r)$ is the density profile. Because our sample does not include sources near the Galactic centre or near the outer regions, we cannot constrain the parameters that define the analytic density profiles defined above (e.g. core density n_0). We use instead a method explained in chapter 2 (Rezaei Kh. et al., 2017, 2018, b) to infer hydrogen densities from the hydrogen column density.

To begin with, the size of the 1D cells towards each source needs to be set to the

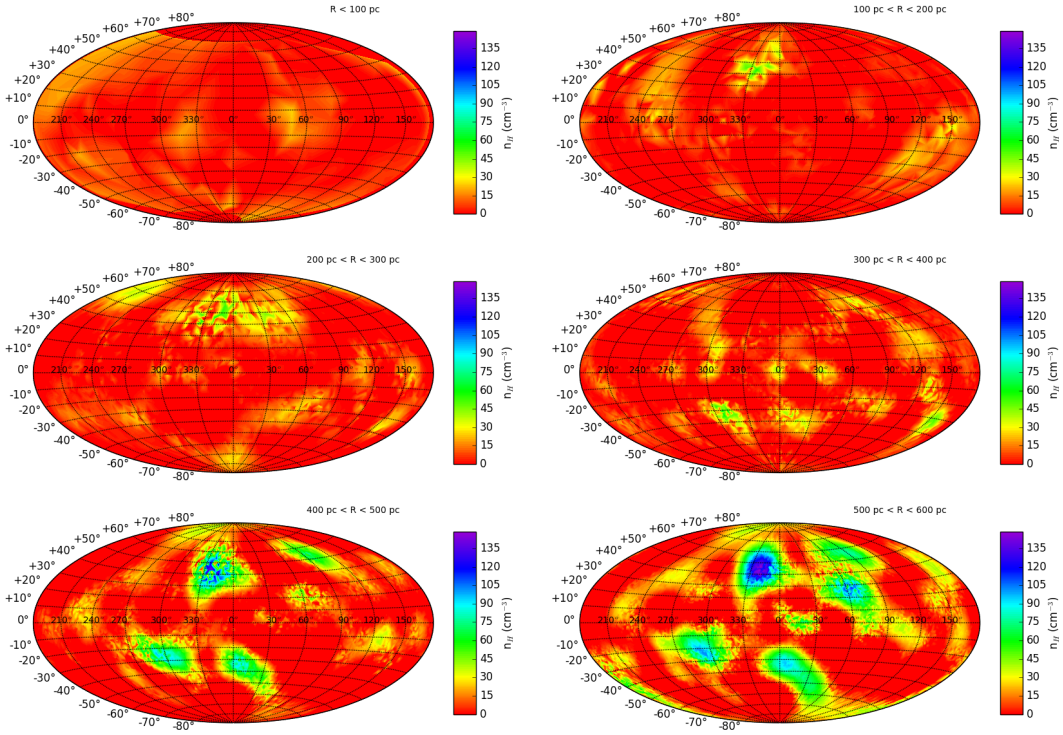


Figure 5.19: Full-sky 2D map of the density distribution. The map is an Aitoff projection covering distances from 0 to 600 pc.

typical separation between the input sources, which in our case is 70 *pc*. Afterwards, there are three hyper-parameters of the Gaussian process which need to be fixed: first is λ which is the correlation length. It needs to be a few times the cell sizes so that it can connect the nearby cells. Here we use the correlation length of 500 *pc*. Second is the variation scale, θ which sets the maximum fluctuation the model can have to capture the variations. It is calculated based on the variance in the input distribution. We fix this parameter at 2×10^{-4}. Finally, we can set a non-zero mean for the Gaussian process based on the input data, with $n_{\rm H} = 7 \times 10^{19}$ cm^{-2}pc^{-1} (≈ 23 cm^{-3}). Having set these hyper-parameters we predict the probability distribution function over densities, which is characterised by the mean and the standard deviation, for 50,000 points randomly distributed in the 3D data space. The results contain negative values mainly due to very noisy equivalent N(H); indeed approximately half of the predicted results are negative. These negative values have been set to zero for illustrative purposes (we come back to this point in section 5.2.2).

Figure 5.19 shows a 2D full-sky map of the density distribution obtained for different

distance ranges. The map is in Aitoff projection. Since the predictions were made for points randomly distributed in the area, we have performed a linear interpolation in order to included those regions for which no points were computed with the method described above. The maximum distance value in our final sample corresponds to $r = 600$ pc. In this sense, we are doing an analysis of the very local ISM density distribution, surrounding the solar system. Clouds of different sizes are observed along all l.o.s, an evidence of small-scale structures. In this regime the ISM is found outside equilibrium mainly because of the presence of shock fronts associated with supernovae explosions, although the contribution from magnetic fields and cosmic-rays to the physical conditions of the gas are significant (Kalberla and Kerp, 2009; Kalberla et al., 2016). HI intermediate-velocity clouds at high latitudes have been identified by Röhser et al. (2016), which can explain the high-density region located near (360°,+40°) for large distances. However, we noted that density uncertainties increase as we move to the boundaries of the data region.

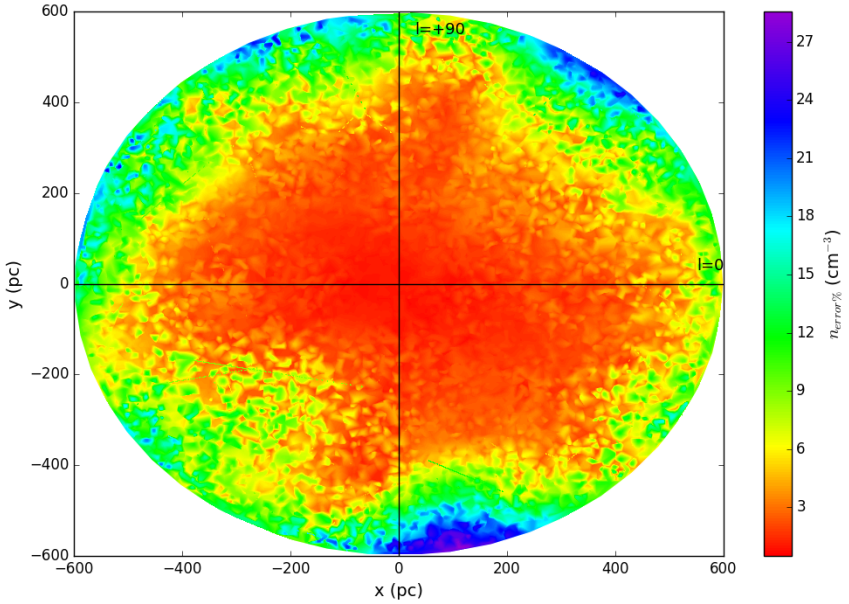


Figure 5.20: (x,y) density uncertainty distribution map for z = 0. The Sun is located at coordinates $(x, y) = (0, 0)$ and the Galactic centre direction to the right.

Figure 5.20 shows a (x,y) percentage density uncertainties distribution for z = 0, normalised to the maximum density value obtained. The uncertainties for the density predictions tend to increase as the predictions get closer to the boundaries of

the data volume. This is due to the fact that as we get closer to the boundary, the number of input sources within the correlation volume decreases which makes the predictions more uncertain in these areas. By including more sources from the next Gaia data releases the uncertainties for the density prediction will decrease.

5.2.2 Caveats and limitations

The main caveat in our analysis comes from the equivalent column density estimation by X-ray spectral fitting, which depends on the number of counts. We estimate the expected uncertainties in the column densities for sources with 500 counts in the complete energy range (0.5-10 keV) to be of the order of $\sim 25\% - 40\%$, depending on the complexity of the model (Gatuzz, Rezaei Kh. et al., 2018). For sources with lower number of counts, the uncertainties can be >100%. However, based on our results, that contained around 50% negative predictions, we think that we underestimate the uncertainty per source. Having performed more sophisticated X-Ray spectral analysis and modelling, which can result in realistic uncertainty estimate, will certainly provide more accurate results.
Moreover, the angular separation in the cross match between EXTRAS and GDR1 has a considerable effect on the final sample, thus the predictions. So far we used a maximum separation of 12.5 arcsecond, which corresponds to the point-spread function (PSF) of the EPIC PN instrument on board of XMM-Newton[5]. However, if we use 4 arcsecond, a value used in similar cross-matching surveys (see for example Caccianiga et al., 2008; Pineau et al., 2011; Lansbury et al., 2017), a lot of sources will be excluded, especially those with higher hydrogen column density located near the Galactic plane. The lower angular separation corresponds to a more accurate identification, nevertheless, the resulting sample covers a small column density range. This is the first time such a map has been created using X-ray spectral fits. Although, because of systematic uncertainties due to the N(H) dependency on the continuum fitting model and the source identifications, the present maps should be considered qualitatively at this point. Future work using GDR2, together with a more sophisticated X-Ray modelling is currently being performed. Observational constraints, such as the densities derived from our analysis, are necessary to compare with the high-resolution 3D hydrodynamical simulations that have been performed in the last decades (de Avillez, 2000; de Avillez and Berry, 2001; Lagos et al., 2013; Gent et al., 2013).

[5] https://heasarc.nasa.gov/docs/xmm/uhb/onaxisxraypsf.html

6

Concluding remarks

In this thesis, I have introduced a new non-parametric method for building a smooth, three-dimensional map of dust density which avoids l.o.s artefacts. It uses a Gaussian process prior to constrain the variation of the dust density in 3D space, but without assuming a specific functional form for the spatial dependence. It instead uses a covariance function which varies with the separation of points. This allows the model to infer the dust density in unobserved regions as well. Our model uses the 3D positions of stars together with their l.o.s extinctions as its input data and infers the posterior probability density function (PDF) over the dust at selected points. This PDF is a Gaussian, and we showed that its mean and standard deviation have analytic solutions. While the l.o.s to the observed stars are divided into discrete cells, predictions are made at arbitrary points without any discretisation being necessary. The result is a continuous, smooth map of the dust distribution, which can trace the local properties of the Galaxy, as opposed to extinction maps.

With this method, we have presented a map of the Galactic disk using giants from APOGEE DR14. This is the first time that such a continuous map of the dust in the disk is presented out to 7 kpc from the Sun. We showed that some of the dust features in our map are possibly associated with spiral arms in our Galaxy. However, our result is limited by the spatial coverage of the input data and observational artefacts due to the APOGEE target selection, plus a limited distance precision of 5%. Future

data from APOGEE south and SDSS-V will be great compliments to the current data by covering the southern hemisphere and providing continuous observation. In addition, the Gaia DR2 data can be of help as it covers some of the missing l.o.s in our current APOGEE sample, although the fact that it is an optical survey limits the depth it can probe in dusty regions.

We have also shown that our method can be used to map the 3D distribution of the dust in the local molecular clouds. Our results in the Orion region shows multiple dust clumps in different distances. Further studies on the Orion complex with the new GDR2 data can help to put new constraints on the shape and distance of the Orion-Eridanus superbubble and the connections of the dust clouds and young stellar OB associations.

We also presented a proof of concept of the kind of science possible with the synergy of the X-ray catalogues, e.g. XMM-Neuton. They can provide measurements of the hydrogen column density $N_{\rm H}$, using which we derived the 3D local hydrogen density map. However, given the systematic uncertainties connected to the source identification and to the dependence of $N_{\rm H}$ on the spectral model, the present maps should be considered qualitatively at this point.

The method has limitations that can be improved in the future. We assume a constant for the mean of the dust in the Gaussian process prior everywhere in our map. This has the advantage of not imposing any spatial dust density prior to the data, therefore the features appear in the map are mainly derived by the input data (i.e. there is no spatial preference for the appearance of the high-density clouds in the prior), but has the disadvantage of not being very informative. Using a more informative prior, like the mean density derived from an actual map, will better constrain the predictions in the area not well-populated by the data. Another limitation of the method is its computation time and memory. We have developed this already by decreasing the dimensionality (section 2.3.1), yet needs further improvement in order to be applicable to larger, global datasets and/or provide higher resolution maps. We are currently working on this by trying to divide the input data in overlapping chunks, but how to combine those overlapping area needs further investigations.

Once these developments have been made and tested, the model will be ready to provide global dust density maps. In particular, using future near-infrared data, we will be able to build a 3D dust map covering more obscured regions of the Galactic disk and provide a more accurate picture of the spiral structure of the Milky Way.

A

Analytic solution of the integral

As explained before, our likelihood function is:

$$P(\mathbf{a}_N|\mathrm{G}\boldsymbol{\rho}_J) = \frac{1}{(2\pi)^{N/2}|\mathrm{V}_N|^{1/2}} \times \exp\left[-\frac{1}{2}(\mathbf{a}_N - \mathrm{G}\boldsymbol{\rho}_J)^{\mathsf{T}}\mathrm{V}_N^{-1}(\mathbf{a}_N - \mathrm{G}\boldsymbol{\rho}_J)\right] , \tag{A.1}$$

where $\mathbf{a}_N$ is the vector of attenuation measurements with covariance V_N. Using our new Gaussian process prior, the law of marginalization and then applying Bayes theorem, we can write the posterior as

$$\begin{aligned} P(\rho_{new}|\mathbf{a}_N) &= \int_{\mathrm{G}\boldsymbol{\rho}_J} P(\rho_{new}, \mathrm{G}\boldsymbol{\rho}_J|\mathbf{a}_N)\, d(\mathrm{G}\boldsymbol{\rho}_J) \\ &= \int_{\mathrm{G}\boldsymbol{\rho}_J} \frac{P(\rho_{new}, \mathrm{G}\boldsymbol{\rho}_J)P(\mathbf{a}_N|\rho_{new}, \mathrm{G}\boldsymbol{\rho}_J)}{P(\mathbf{a}_N)}\, d(\mathrm{G}\boldsymbol{\rho}_J) \\ &= \frac{1}{P(\mathbf{a}_N)} \int_{\mathrm{G}\boldsymbol{\rho}_J} P(\rho_{new}, \mathrm{G}\boldsymbol{\rho}_J)P(\mathbf{a}_N|\mathrm{G}\boldsymbol{\rho}_J)\, d(\mathrm{G}\boldsymbol{\rho}_J) \ . \end{aligned} \tag{A.2}$$

Both the first (the new Gaussian process prior) and the second (the likelihood) terms are linear functions of ρ_{new} which suggests an analytic solution for the integral. For brevity we write equation A.2 as

$$P(\rho_{new}|\mathbf{a}_N) = \frac{1}{Z}\int_{\mathrm{G}\boldsymbol{\rho}_J} e^{-\psi/2}\, d(\mathrm{G}\boldsymbol{\rho}_J) \tag{A.3}$$

where Z is a normalisation constant, and

$$\begin{aligned}\psi &= \boldsymbol{x}_{N+1}^{\mathsf{T}}\Omega_{N+1}^{-1}\boldsymbol{x}_{N+1} + (\mathbf{a}_N - \mathrm{G}\boldsymbol{\rho}_J)^{\mathsf{T}}\mathrm{V}_N^{-1}(\mathbf{a}_N - \mathrm{G}\boldsymbol{\rho}_J) \\ &= \boldsymbol{x}_{N+1}^{\mathsf{T}}\Omega_{N+1}^{-1}\boldsymbol{x}_{N+1} + \mathbf{a}_N^{\mathsf{T}}\mathrm{V}_N^{-1}\mathbf{a}_N - \boldsymbol{\rho}_J^{\mathsf{T}}\mathrm{G}^{\mathsf{T}}\mathrm{V}_N^{-1}\mathbf{a}_N + \boldsymbol{\rho}_J^{\mathsf{T}}\mathrm{G}^{\mathsf{T}}\mathrm{V}_N^{-1}\mathrm{G}\boldsymbol{\rho}_J \\ &\quad - \mathbf{a}_N^{\mathsf{T}}\mathrm{V}_N^{-1}\mathrm{G}\boldsymbol{\rho}_J \end{aligned} \tag{A.4}$$

where the third and fifth terms are identical as each term is a scalar and

$$\boldsymbol{x}_{N+1} = \begin{bmatrix} \mathrm{G}\boldsymbol{\rho}_J - \mathrm{G}\vec{\rho}_\mu \\ \rho_{new} - \rho_\mu \end{bmatrix} \tag{A.5}$$

$$\Omega_{N+1}^{-1} = \begin{bmatrix} \mathrm{Q}_N & \mathbf{q}_N \\ \mathbf{q}_N^{\mathsf{T}} & q \end{bmatrix} \tag{A.6}$$

where Q_N ($N \times N$ matrix), $\mathbf{q}_N$ ($N \times 1$ vector), and q (scalar) are components of the inverted matrix, and ρ_μ is the mean of the dust densities for the Gaussian process prior and is calculated from the input attenuation (see chapter 2). We can then write

$$\begin{aligned}&\boldsymbol{x}_{N+1}^{\mathsf{T}}\Omega_{N+1}^{-1}\boldsymbol{x}_{N+1} \\ &= \begin{bmatrix} (\mathrm{G}\boldsymbol{\rho}_J - \mathrm{G}\vec{\rho}_\mu)^{\mathsf{T}} & (\rho_{new} - \rho_\mu) \end{bmatrix} \begin{bmatrix} \mathrm{Q}_N & \mathbf{q}_N \\ \mathbf{q}_N^{\mathsf{T}} & q \end{bmatrix} \begin{bmatrix} \mathrm{G}\boldsymbol{\rho}_J - \mathrm{G}\vec{\rho}_\mu \\ (\rho_{new} - \rho_\mu) \end{bmatrix} \\ &= (\mathrm{G}\boldsymbol{\rho}_J - \mathrm{G}\vec{\rho}_\mu)^{\mathsf{T}}\mathrm{Q}_N(\mathrm{G}\boldsymbol{\rho}_J - \mathrm{G}\vec{\rho}_\mu) + 2(\rho_{new} - \rho_\mu)\mathbf{q}_N^{\mathsf{T}}(\mathrm{G}\boldsymbol{\rho}_J - \mathrm{G}\vec{\rho}_\mu) \\ &\quad + q(\rho_{new} - \rho_\mu)^2 \,. \end{aligned} \tag{A.7}$$

Substituting this into equation A.4 and gathering together terms gives

$$\begin{aligned}\psi &= (\mathrm{G}\boldsymbol{\rho}_J)^{\mathsf{T}}(\mathrm{Q}_N + \mathrm{V}_N^{-1})\mathrm{G}\boldsymbol{\rho}_J + 2((\rho_{new} - \rho_\mu)\mathbf{q}_N^{\mathsf{T}} - \mathbf{a}_N^{\mathsf{T}}\mathrm{V}_N^{-1} \\ &\quad - (\mathrm{G}\vec{\rho}_\mu)^{\mathsf{T}}\mathrm{Q}_N)\mathrm{G}\boldsymbol{\rho}_J + ((\mathrm{G}\vec{\rho}_\mu)^{\mathsf{T}}\mathrm{Q}_N(\mathrm{G}\vec{\rho}_\mu) \\ &\quad - 2(\rho_{new} - \rho_\mu)\mathbf{q}_N^{\mathsf{T}}(\mathrm{G}\vec{\rho}_\mu) + \mathbf{a}_N^{\mathsf{T}}\mathrm{V}_N^{-1}\mathbf{a}_N + q(\rho_{new} - \rho_\mu)^2) \end{aligned} \tag{A.8}$$

This is a quadratic expression in $\mathrm{G}\boldsymbol{\rho}_J$. The last term is independent of $\boldsymbol{\rho}_J$ so can be taken out of the integral, allowing us to write equation A.3 as

$$\begin{aligned}P(\rho_{new}|\mathbf{a}_N) &= \frac{1}{Z}\exp\left[-\frac{1}{2}(\mathrm{G}\vec{\rho}_\mu)^{\mathsf{T}}\mathrm{Q}_N(\mathrm{G}\vec{\rho}_\mu) - \frac{1}{2}\mathbf{a}_N^{\mathsf{T}}\mathrm{V}_N^{-1}\mathbf{a}_N\right] \\ &\quad \times \exp\left[-\frac{1}{2}q(\rho_{new} - \rho_\mu)^2 + \mathbf{q}_N^{\mathsf{T}}(\mathrm{G}\vec{\rho}_\mu)(\rho_{new} - \rho_\mu)\right] \\ &\quad \times \int_{\mathrm{G}\boldsymbol{\rho}_J} \exp\left[-\frac{1}{2}(\mathrm{G}\boldsymbol{\rho}_J)^{\mathsf{T}}\mathrm{R}_N\mathrm{G}\boldsymbol{\rho}_J + \mathbf{b}_N^{\mathsf{T}}\mathrm{G}\boldsymbol{\rho}_J\right] d\mathrm{G}\boldsymbol{\rho}_J \end{aligned} \tag{A.9}$$

where

$$\mathrm{R}_N = \mathrm{Q}_N + \mathrm{V}_N^{-1} \tag{A.10}$$
$$\mathbf{b}_N = \mathrm{Q}_N^{\mathsf{T}}(\mathrm{G}\vec{\rho}_\mu) + \mathrm{V}_N^{-1}\mathbf{a}_N - (\rho_{new} - \rho_\mu)\mathbf{q}_N \tag{A.11}$$

The integral is a standard one allowing us to write equation A.9 as

$$P(\rho_{new}|\mathbf{a}_N) = \frac{1}{Z}e^{-\phi/2} \quad \text{where}$$
$$\phi = q(\rho_{new} - \rho_\mu)^2 - 2\mathbf{q}_N^{\mathsf{T}}(\mathrm{G}\vec{\rho}_\mu)(\rho_{new} - \rho_\mu) - \mathbf{b}_N^{\mathsf{T}}\mathrm{R}_N^{-1}\mathbf{b}_N \tag{A.12}$$

and we have absorbed all factors which do not depend on ρ_{new} into the normalization constant. Substituting for $\mathbf{b}_N$ this becomes

$$\phi = (\text{terms free of } \rho_{new}) + (q - \mathbf{q}_N^{\mathsf{T}}\mathrm{R}_N^{-1}\mathbf{q}_N)(\rho_{new} - \rho_\mu)^2 + 2(\mathbf{a}_N^{\mathsf{T}}\mathrm{V}_N^{-1}\mathrm{R}_N^{-1}\mathbf{q}_N + \mathbf{q}_N^{\mathsf{T}}\mathrm{R}_N^{-1}\mathrm{Q}_N^{\mathsf{T}}\mathrm{G}\vec{\rho}_\mu - \mathbf{q}_N^{\mathsf{T}}\mathrm{G}\vec{\rho}_\mu)(\rho_{new} - \rho_\mu)\ . \tag{A.13}$$

The first parenthesis contains terms which do not depend on ρ_{new} so can be absorbed into the normalization constant. Putting this into equation A.12 gives us

$$P(\rho_{new}|\mathbf{a}_N) = \frac{1}{Z}\exp\left[-\frac{1}{2}\alpha(\rho_{new} - \rho_\mu)^2 - \beta(\rho_{new} - \rho_\mu)\right] \tag{A.14}$$

where

$$\alpha = q - \mathbf{q}_N^{\mathsf{T}}\mathrm{R}_N^{-1}\mathbf{q}_N$$
$$\beta = \mathbf{a}_N^{\mathsf{T}}\mathrm{V}_N^{-1}\mathrm{R}_N^{-1}\mathbf{q}_N + \mathbf{q}_N^{\mathsf{T}}\mathrm{R}_N^{-1}\mathrm{Q}_N^{\mathsf{T}}\mathrm{G}\vec{\rho}_\mu - \mathbf{q}_N^{\mathsf{T}}\mathrm{G}\vec{\rho}_\mu\ . \tag{A.15}$$

By completing the square in the exponent we see that

$$P(\rho_{new}|\mathbf{a}_N) = \sqrt{\frac{\alpha}{2\pi}}\exp\left[-\frac{\alpha}{2}\left(\rho_{new} + \frac{\beta}{\alpha} - \rho_\mu\right)^2\right]\ . \tag{A.16}$$

which is a Gaussian with mean $-\beta/\alpha + \rho_\mu$ and variance $1/\alpha$.

References

Abolfathi, B., Aguado, D. S., Aguilar, G., Allende Prieto, C., Almeida, A., Tasnim Ananna, T., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., and et al. The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the extended Baryon Oscillation Spectroscopic Survey and from the second phase of the Apache Point Observatory Galactic Evolution Experiment. *ArXiv e-prints*, July 2017.

Alves, J. and Bouy, H. Orion revisited. I. The massive cluster in front of the Orion nebula cluster. *A&A*, 547:A97, November 2012. doi: 10.1051/0004-6361/201220119.

Andrae, R., Fouesneau, M., Creevey, O., Ordenovic, C., Mary, N., Burlacu, A., Chaoul, L., Jean-Antoine-Piccolo, A., Kordopatis, G., Korn, A., Lebreton, Y., Panem, C., Pichon, B., Thevenin, F., Walmsley, G., and Bailer-Jones, C. A. L. Gaia Data Release 2: first stellar parameters from Apsis. *ArXiv e-prints*, April 2018.

Astraatmadja, T. L. and Bailer-Jones, C. A. L. Estimating Distances from Parallaxes. III. Distances of Two Million Stars in the Gaia DR1 Catalogue. *ApJ*, 833: 119, December 2016. doi: 10.3847/1538-4357/833/1/119.

Baba, J., Kawata, D., Matsunaga, N., Grand, R. J. J., and Hunt, J. A. S. Gaia DR1 Evidence of Disrupting the Perseus Arm. *ApJL*, 853:L23, February 2018. doi: 10.3847/2041-8213/aaa839.

Bailer-Jones, C. A. L. Bayesian inference of stellar parameters and interstellar extinction using parallaxes and multiband photometry. *MNRAS*, 411:435–452, February 2011. doi: 10.1111/j.1365-2966.2010.17699.x.

Bailer-Jones, C. A. L., Andrae, R., Arcay, B., Astraatmadja, T., Bellas-Velidis, I., Berihuete, A., Bijaoui, A., Carrión, C., Dafonte, C., Damerdji, Y., Dapergolas, A., de Laverny, P., Delchambre, L., Drazinos, P., Drimmel, R., Frémat, Y., Fustes, D., García-Torres, M., Guédé, C., Heiter, U., Janotto, A.-M., Karampelas, A., Kim, D.-W., Knude, J., Kolka, I., Kontizas, E., Kontizas, M., Korn, A. J., Lanzafame, A. C., Lebreton, Y., Lindstrøm, H., Liu, C., Livanou, E., Lobel, A., Manteiga, M., Martayan, C., Ordenovic, C., Pichon, B., Recio-Blanco, A., Rocca-Volmerange, B., Sarro, L. M., Smith, K., Sordo, R., Soubiran, C., Surdej, J., Thévenin, F., Tsalmantza, P., Vallenari, A., and Zorec, J. The Gaia astrophysical parameters

inference system (Apsis). Pre-launch description. *A&A*, 559:A74, November 2013. doi: 10.1051/0004-6361/201322344.

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., and Andrae, R. Estimating distances from parallaxes IV: Distances to 1.33 billion stars in Gaia Data Release 2. *ArXiv e-prints*, April 2018.

Bally, J. *Overview of the Orion Complex*, page 459. The Northern Sky ASP Monograph Publications, Vol. 4. Edited by Bo Reipurth, December 2008.

Barnard, E. E. On the dark markings of the sky, with a catalogue of 182 such objects. *ApJ*, 49, January 1919. doi: 10.1086/142439.

Blanton, M. R., Bershady, M. A., Abolfathi, B., Albareti, F. D., Allende Prieto, C., Almeida, A., Alonso-García, J., Anders, F., Anderson, S. F., Andrews, B., and et al. Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. *AJ*, 154:28, July 2017. doi: 10.3847/1538-3881/aa7567.

Bohlin, R. C., Savage, B. D., and Drake, J. F. A survey of interstellar H I from L-alpha absorption measurements. II. *ApJ*, 224:132–142, August 1978. doi: 10.1086/156357.

Bouy, H., Alves, J., Bertin, E., Sarro, L. M., and Barrado, D. Orion revisited. II. The foreground population to Orion A. *A&A*, 564:A29, April 2014. doi: 10.1051/0004-6361/201323191.

Bovy, J., Nidever, D. L., Rix, H.-W., Girardi, L., Zasowski, G., Chojnowski, S. D., Holtzman, J., Epstein, C., Frinchaboy, P. M., Hayden, M. R., Rodrigues, T. S., Majewski, S. R., Johnson, J. A., Pinsonneault, M. H., Stello, D., Allende Prieto, C., Andrews, B., Basu, S., Beers, T. C., Bizyaev, D., Burton, A., Chaplin, W. J., Cunha, K., Elsworth, Y., García, R. A., García-Herńandez, D. A., García Pérez, A. E., Hearty, F. R., Hekker, S., Kallinger, T., Kinemuchi, K., Koesterke, L., Mészáros, S., Mosser, B., O'Connell, R. W., Oravetz, D., Pan, K., Robin, A. C., Schiavon, R. P., Schneider, D. P., Schultheis, M., Serenelli, A., Shetrone, M., Silva Aguirre, V., Simmons, A., Skrutskie, M., Smith, V. V., Stassun, K., Weinberg, D. H., Wilson, J. C., and Zamora, O. The APOGEE Red-clump Catalog: Precise Distances, Velocities, and High-resolution Elemental Abundances over a Large Area of the Milky Way's Disk. *ApJ*, 790:127, August 2014. doi: 10.1088/0004-637X/790/2/127.

Breiman, L. Random forests. *Machine Learning*, 45(1):5–32, Oct 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

Bressan, A., Marigo, P., Girardi, L., Salasnich, B., Dal Cero, C., Rubele, S., and Nanni, A. PARSEC: stellar tracks and isochrones with the PAdova and TRieste Stellar Evolution Code. *MNRAS*, 427:127–145, November 2012. doi: 10.1111/j.1365-2966.2012.21948.x.

Brown, A. G. A., de Geus, E. J., and de Zeeuw, P. T. The Orion OB1 association. 1: Stellar content. *A&A*, 289:101–120, September 1994.

Brown, A. G. A., Hartmann, D., and Burton, W. B. The Orion OB1 association. II. The Orion-Eridanus Bubble. *A&A*, 300:903, August 1995.

Caccianiga, A., Severgnini, P., Della Ceca, R., Maccacaro, T., Cocchia, F., Barcons, X., Carrera, F. J., Matute, I., McMahon, R. G., Page, M. J., Pietsch, W., Sbarufatti, B., Schwope, A., Tedds, J. A., and Watson, M. G. The XMM-Newton bright serendipitous survey. Identification and optical spectral properties. *A&A*, 477:735–746, January 2008. doi: 10.1051/0004-6361:20078568.

Campbell, E. K., Holz, M., Gerlich, D., and Maier, J. P. Laboratory confirmation of C_{60}^+ as the carrier of two diffuse interstellar bands. *Nature*, 523:322–323, July 2015. doi: 10.1038/nature14566.

Capitanio, L., Lallement, R., Vergely, J. L., Elyajouri, M., and Monreal-Ibero, A. Three-dimensional mapping of the local interstellar medium with composite data. *A&A*, 606:A65, October 2017. doi: 10.1051/0004-6361/201730831.

Cardelli, J. A., Clayton, G. C., and Mathis, J. S. The relationship between infrared, optical, and ultraviolet extinction. *ApJ*, 345:245–256, October 1989. doi: 10.1086/167900.

Cayrel, R. and Schatzman, E. Sur la polarisation interstellaire par des particules de graphite. *Annales d'Astrophysique*, 17:555, January 1954.

Chandrasekhar, S. On the Radiative Equilibrium of a Stellar Atmosphere. X. *ApJ*, 103:351, May 1946. doi: 10.1086/144816.

Chen, Y., Bressan, A., Girardi, L., Marigo, P., Kong, X., and Lanza, A. PARSEC evolutionary tracks of massive stars up to 350 M at metallicities $0.0001 \leq Z \leq 0.04$. *MNRAS*, 452:1068–1080, September 2015. doi: 10.1093/mnras/stv1281.

Corbelli, E., Bianchi, S., Cortese, L., Giovanardi, C., Magrini, L., Pappalardo, C., Boselli, A., Bendo, G. J., Davies, J., Grossi, M., Madden, S. C., Smith, M. W. L., Vlahakis, C., Auld, R., Baes, M., De Looze, I., Fritz, J., Pohlen, M., and Verstappen, J. The Herschel Virgo Cluster Survey. X. The relationship between cold dust and molecular gas content in Virgo spirals. *A&A*, 542:A32, June 2012. doi: 10.1051/0004-6361/201117329.

Dame, T. M., Hartmann, D., and Thaddeus, P. The Milky Way in Molecular Clouds: A New Complete CO Survey. *ApJ*, 547:792–813, February 2001. doi: 10.1086/318388.

de Avillez, M. A. Disc-halo interaction - I. Three-dimensional evolution of the Galactic disc. *MNRAS*, 315:479–497, July 2000. doi: 10.1046/j.1365-8711.2000.03464.x.

de Avillez, M. A. and Berry, D. L. Three-dimensional evolution of worms and chimneys in the Galactic disc. *MNRAS*, 328:708–718, December 2001. doi: 10.1046/j.1365-8711.2001.04865.x.

de Bruijne, J. H. J., Rygl, K. L. J., and Antoja, T. Gaia Astrometric Science Performance - Post-Launch Predictions. In *EAS Publications Series*, volume 67 of *EAS Publications Series*, pages 23–29, July 2014. doi: 10.1051/eas/1567004.

De Luca, A., Salvaterra, R., Tiengo, A., D'Agostino, D., Watson, M. G., Haberl, F., and Wilms, J. Science with the EXTraS Project: Exploring the X-Ray Transient and Variable Sky. *The Universe of Digital Sky Surveys*, 42:291, 2016. doi: 10.1007/978-3-319-19330-4_46.

Demmel, J., Dumitriu, I., and Holtz, O. Fast linear algebra is stable. 2007.

Draine, B. T. Interstellar Dust Grains. *ARA&A*, 41:241–289, 2003. doi: 10.1146/annurev.astro.41.011802.094840.

Draine, B. T. Interstellar Dust Models and Evolutionary Implications. In Henning, T., Grün, E., and Steinacker, J., editors, *Cosmic Dust - Near and Far*, volume 414 of *Astronomical Society of the Pacific Conference Series*, page 453, December 2009.

Draine, B. T. and Li, A. Infrared Emission from Interstellar Dust. IV. The Silicate-Graphite-PAH Model in the Post-Spitzer Era. *ApJ*, 657:810–837, March 2007. doi: 10.1086/511055.

Eisenstein, D. J., Weinberg, D. H., Agol, E., Aihara, H., Allende Prieto, C., Anderson, S. F., Arns, J. A., Aubourg, É., Bailey, S., Balbinot, E., and et al. SDSS-III: Massive Spectroscopic Surveys of the Distant Universe, the Milky Way, and Extra-Solar Planetary Systems. *AJ*, 142:72, September 2011. doi: 10.1088/0004-6256/142/3/72.

Farrell, S. A., Murphy, T., and Lo, K. K. Autoclassification of the variable 3xmm sources using the random forest machine learning algorithm. *The Astrophysical Journal*, 813(1):28, 2015. URL http://stacks.iop.org/0004-637X/813/i=1/a=28.

Fitzpatrick, E. L. Correcting for the Effects of Interstellar Extinction. *PASP*, 111:63–75, January 1999. doi: 10.1086/316293.

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Mignard, F., Drimmel, R., Babusiaux, C., Bailer-Jones, C. A. L., Bastian, U., and et al. Gaia Data Release 1. Summary of the astrometric, photometric, and survey properties. *A&A*, 595:A2, November 2016a. doi: 10.1051/0004-6361/201629512.

Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., Brown, A. G. A., Vallenari, A., Babusiaux, C., Bailer-Jones, C. A. L., Bastian, U., Biermann, M., Evans,

D. W., and et al. The Gaia mission. *A&A*, 595:A1, November 2016b. doi: 10.1051/0004-6361/201629272.

Gaia Collaboration, Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., and Bailer-Jones, C. A. L. Gaia Data Release 2. Summary of the contents and survey properties. *ArXiv e-prints*, April 2018.

García Pérez, A. E., Allende Prieto, C., Holtzman, J. A., Shetrone, M., Mészáros, S., Bizyaev, D., Carrera, R., Cunha, K., García-Hernández, D. A., Johnson, J. A., Majewski, S. R., Nidever, D. L., Schiavon, R. P., Shane, N., Smith, V. V., Sobeck, J., Troup, N., Zamora, O., Weinberg, D. H., Bovy, J., Eisenstein, D. J., Feuillet, D., Frinchaboy, P. M., Hayden, M. R., Hearty, F. R., Nguyen, D. C., O'Connell, R. W., Pinsonneault, M. H., Wilson, J. C., and Zasowski, G. ASPCAP: The APOGEE Stellar Parameter and Chemical Abundances Pipeline. *AJ*, 151:144, June 2016. doi: 10.3847/0004-6256/151/6/144.

Gatuzz, E. and Churazov, E. Probing the structure of the gas in the milky way through x-ray high-resolution spectroscopy. *Monthly Notices of the Royal Astronomical Society*, 474(1):696–711, 2018. doi: 10.1093/mnras/stx2776. URL http://dx.doi.org/10.1093/mnras/stx2776.

Gatuzz, Rezaei Kh., S., Kallman, T. R., Kreikenbohm, A., Oertel, M., Wilms, J., and García, J. A. 3D mapping of the neutral X-ray absorption in the local interstellar medium: the Gaia and XMM-Newton synergy. *MNRAS*, 479:3715–3725, September 2018. doi: 10.1093/mnras/sty1738.

Gent, F. A., Shukurov, A., Fletcher, A., Sarson, G. R., and Mantere, M. J. The supernova-regulated ISM - I. The multiphase structure. *MNRAS*, 432:1396–1423, June 2013. doi: 10.1093/mnras/stt560.

Genzel, R., Reid, M. J., Moran, J. M., and Downes, D. Proper motions and distances of H2O maser sources. I - The outflow in Orion-KL. *ApJ*, 244:884–902, March 1981. doi: 10.1086/158764.

Gibbs, M. and MacKay, D. J. Efficient implementation of gaussian processes. 1997.

Girardi, L., Dalcanton, J., Williams, B., de Jong, R., Gallart, C., Monelli, M., Groenewegen, M. A. T., Holtzman, J. A., Olsen, K. A. G., Seth, A. C., Weisz, D. R., and the ANGST/ANGRRR Collaboration. Revised Bolometric Corrections and Interstellar Extinction Coefficients for the ACS and WFPC2 Photometric Systems. *PASP*, 120:583, May 2008. doi: 10.1086/588526.

Gneiting, T. Compactly supported correlation functions. *Journal of Multivariate Analysis*, 83(2):493–508, 2002.

Green, G. M., Schlafly, E. F., Finkbeiner, D. P., Jurić, M., Rix, H.-W., Burgett, W., Chambers, K. C., Draper, P. W., Flewelling, H., Kudritzki, R. P., Magnier, E., Martin, N., Metcalfe, N., Tonry, J., Wainscoat, R., and Waters, C. Measuring

Distances and Reddenings for a Billion Stars: Toward a 3D Dust Map from Pan-STARRS 1. *ApJ*, 783:114, March 2014. doi: 10.1088/0004-637X/783/2/114.

Green, G. M., Schlafly, E. F., Finkbeiner, D. P., Rix, H.-W., Martin, N., Burgett, W., Draper, P. W., Flewelling, H., Hodapp, K., Kaiser, N., Kudritzki, R. P., Magnier, E., Metcalfe, N., Price, P., Tonry, J., and Wainscoat, R. A Three-dimensional Map of Milky Way Dust. *ApJ*, 810:25, September 2015. doi: 10.1088/0004-637X/810/1/25.

Green, G. M., Schlafly, E. F., Finkbeiner, D., Rix, H.-W., Martin, N., Burgett, W., Draper, P. W., Flewelling, H., Hodapp, K., Kaiser, N., Kudritzki, R.-P., Magnier, E. A., Metcalfe, N., Tonry, J. L., Wainscoat, R., and Waters, C. Galactic reddening in 3D from stellar photometry - an improved map. *MNRAS*, 478:651–666, July 2018. doi: 10.1093/mnras/sty1008.

Greenstein, J. L. The Theory of Interstellar Absorption. *Harvard College Observatory Circular*, 422:1–36, July 1937.

Hall, J. S. Observations of B-Type Stars in the Red and Infra-Red Regions of the Spectrum. *ApJ*, 85:145, April 1937. doi: 10.1086/143810.

Hall, J. S. Observations of the Polarized Light from Stars. *Science*, 109:166–167, February 1949. doi: 10.1126/science.109.2825.166.

Hanson, R. J. and Bailer-Jones, C. A. L. 3D Galactic dust extinction mapping with multiband photometry. *MNRAS*, 438:2938–2953, March 2014. doi: 10.1093/mnras/stt2401.

Hanson, R. J., Bailer-Jones, C. A. L., Burgett, W. S., Chambers, K. C., Hodapp, K. W., Kaiser, N., Tonry, J. L., Wainscoat, R. J., and Waters, C. Dust in three dimensions in the Galactic plane. *MNRAS*, 463:3604–3615, December 2016. doi: 10.1093/mnras/stw2240.

Heiles, C. H I shells and supershells. *ApJ*, 229:533–537, April 1979. doi: 10.1086/156986.

Herschel, W. On the Construction of the Heavens. *Philosophical Transactions of the Royal Society of London Series I*, 75:213–266, 1785.

Hertzsprung, E. Über die Helligkeit der Plejadennebel. *Astronomische Nachrichten*, 195:449, September 1913. doi: 10.1002/asna.19131952302.

Hiltner, W. A. Polarization of Light from Distant Stars by Interstellar Medium. *Science*, 109:165, February 1949. doi: 10.1126/science.109.2825.165.

Høg, E., Fabricius, C., Makarov, V. V., Urban, S., Corbin, T., Wycoff, G., Bastian, U., Schwekendiek, P., and Wicenec, A. The Tycho-2 catalogue of the 2.5 million brightest stars. *A&A*, 355:L27–L30, March 2000.

Holtzman, J. A., Shetrone, M., Johnson, J. A., Allende Prieto, C., Anders, F., Andrews, B., Beers, T. C., Bizyaev, D., Blanton, M. R., Bovy, J., Carrera, R., Chojnowski, S. D., Cunha, K., Eisenstein, D. J., Feuillet, D., Frinchaboy, P. M., Galbraith-Frew, J., García Pérez, A. E., García-Hernández, D. A., Hasselquist, S., Hayden, M. R., Hearty, F. R., Ivans, I., Majewski, S. R., Martell, S., Meszaros, S., Muna, D., Nidever, D., Nguyen, D. C., O'Connell, R. W., Pan, K., Pinsonneault, M., Robin, A. C., Schiavon, R. P., Shane, N., Sobeck, J., Smith, V. V., Troup, N., Weinberg, D. H., Wilson, J. C., Wood-Vasey, W. M., Zamora, O., and Zasowski, G. Abundances, Stellar Parameters, and Spectra from the SDSS-III/APOGEE Survey. *AJ*, 150:148, November 2015. doi: 10.1088/0004-6256/150/5/148.

Hoyle, F. and Wickramasinghe, N. C. On graphite particles as interstellar grains. *MNRAS*, 124:417, 1962. doi: 10.1093/mnras/124.5.417.

Hubble, E. P. The source of luminosity in galactic nebulae. *ApJ*, 56, December 1922. doi: 10.1086/142713.

Jeffries, R. D. The distance to the Orion Nebula cluster. *MNRAS*, 376:1109–1119, April 2007. doi: 10.1111/j.1365-2966.2007.11471.x.

Kalberla, P. M. W. and Kerp, J. The Hi Distribution of the Milky Way. *ARA&A*, 47:27–61, September 2009. doi: 10.1146/annurev-astro-082708-101823.

Kalberla, P. M. W., Kerp, J., Haud, U., Winkel, B., Ben Bekhti, N., Flöer, L., and Lenz, D. Cold Milky Way HI Gas in Filaments. *ApJ*, 821:117, April 2016. doi: 10.3847/0004-637X/821/2/117.

Kapteyn, J. C. On the Absorption of Light in Space. *ApJ*, 29:46, January 1909. doi: 10.1086/141618.

Kapteyn, J. C. First Attempt at a Theory of the Arrangement and Motion of the Sidereal System. *ApJ*, 55:302, May 1922. doi: 10.1086/142670.

Kennicutt, R. C., Calzetti, D., Aniano, G., Appleton, P., Armus, L., Beirão, P., Bolatto, A. D., Brandl, B., Crocker, A., Croxall, K., Dale, D. A., Donovan Meyer, J., Draine, B. T., Engelbracht, C. W., Galametz, M., Gordon, K. D., Groves, B., Hao, C.-N., Helou, G., Hinz, J., Hunt, L. K., Johnson, B., Koda, J., Krause, O., Leroy, A. K., Li, Y., Meidt, S., Montiel, E., Murphy, E. J., Rahman, N., Rix, H.-W., Roussel, H., Sandstrom, K., Sauvage, M., Schinnerer, E., Skibba, R., Smith, J. D. T., Srinivasan, S., Vigroux, L., Walter, F., Wilson, C. D., Wolfire, M., and Zibetti, S. KINGFISH – Key Insights on Nearby Galaxies: A Far-Infrared Survey with Herschel: Survey Description and Image Atlas. *PASP*, 123:1347, December 2011. doi: 10.1086/663818.

Knacke, R. F., Gaustad, J. E., Gillett, F. C., and Stein, W. A. A Possible Identification of Interstellar Silicate Absorption in the Infrared Spectrum of 119 Tauri. *ApJL*, 155:L189, March 1969. doi: 10.1086/180333.

Kraus, S., Balega, Y. Y., Berger, J.-P., Hofmann, K.-H., Millan-Gabet, R., Monnier, J. D., Ohnaka, K., Pedretti, E., Preibisch, T., Schertl, D., Schloerb, F. P., Traub, W. A., and Weigelt, G. Visual/infrared interferometry of Orion Trapezium stars: preliminary dynamical orbit and aperture synthesis imaging of the theta Orionis C system. *A&A*, 466:649–659, May 2007. doi: 10.1051/0004-6361:20066965.

Lagos, C. d. P., Lacey, C. G., and Baugh, C. M. A dynamical model of supernova feedback: gas outflows from the interstellar medium. *MNRAS*, 436:1787–1817, December 2013. doi: 10.1093/mnras/stt1696.

Lallement, R., Vergely, J.-L., Valette, B., Puspitarini, L., Eyer, L., and Casagrande, L. 3D maps of the local ISM from inversion of individual color excess measurements. *A&A*, 561:A91, January 2014. doi: 10.1051/0004-6361/201322032.

Lansbury, G. B., Stern, D., Aird, J., Alexander, D. M., Fuentes, C., Harrison, F. A., Treister, E., Bauer, F. E., Tomsick, J. A., Baloković, M., Del Moro, A., Gandhi, P., Ajello, M., Annuar, A., Ballantyne, D. R., Boggs, S. E., Brandt, W. N., Brightman, M., Chen, C.-T. J., Christensen, F. E., Civano, F., Comastri, A., Craig, W. W., Forster, K., Grefenstette, B. W., Hailey, C. J., Hickox, R. C., Jiang, B., Jun, H. D., Koss, M., Marchesi, S., Melo, A. D., Mullaney, J. R., Noirot, G., Schulze, S., Walton, D. J., Zappacosta, L., and Zhang, W. W. The NuSTAR Serendipitous Survey: The 40-month Catalog and the Properties of the Distant High-energy X-Ray Source Population. *ApJ*, 836:99, February 2017. doi: 10.3847/1538-4357/836/1/99.

Li, A. and Draine, B. T. Infrared Emission from Interstellar Dust. II. The Diffuse Interstellar Medium. *ApJ*, 554:778–802, June 2001. doi: 10.1086/323147.

Li, A. and Greenberg, J. M. In dust we trust: an overview of observations and theories of interstellar dust. In Pirronello, V., Krelowski, J., and Manicò, G., editors, *Solid State Astrochemistry*, volume 120, pages 37–84, 2003.

Lindblad, B. A Condensation Theory of Meteoric Matter and its Cosmological Significance. *Nature*, 135:133–135, January 1935. doi: 10.1038/135133a0.

Lindegren, L., Lammers, U., Bastian, U., Hernández, J., Klioner, S., Hobbs, D., Bombrun, A., Michalik, D., Ramos-Lerate, M., Butkevich, A., Comoretto, G., Joliet, E., Holl, B., Hutton, A., Parsons, P., Steidelmüller, H., Abbas, U., Altmann, M., Andrei, A., Anton, S., Bach, N., Barache, C., Becciani, U., Berthier, J., Bianchi, L., Biermann, M., Bouquillon, S., Bourda, G., Brüsemeister, T., Bucciarelli, B., Busonero, D., Carlucci, T., Castañeda, J., Charlot, P., Clotet, M., Crosta, M., Davidson, M., de Felice, F., Drimmel, R., Fabricius, C., Fienga, A., Figueras, F., Fraile, E., Gai, M., Garralda, N., Geyer, R., González-Vidal, J. J., Guerra, R., Hambly, N. C., Hauser, M., Jordan, S., Lattanzi, M. G., Lenhardt, H., Liao, S., Löffler, W., McMillan, P. J., Mignard, F., Mora, A., Morbidelli, R., Portell, J., Riva, A., Sarasso, M., Serraller, I., Siddiqui, H., Smart, R., Spagna, A., Stampa, U., Steele, I., Taris, F., Torra, J., van Reeven, W., Vecchiato, A.,

Zschocke, S., de Bruijne, J., Gracia, G., Raison, F., Lister, T., Marchant, J., Messineo, R., Soffel, M., Osorio, J., de Torres, A., and O'Mullane, W. Gaia Data Release 1. Astrometry: one billion positions, two million proper motions and parallaxes. *A&A*, 595:A4, November 2016. doi: 10.1051/0004-6361/201628714.

Lo, K. K., Farrell, S., Murphy, T., and Gaensler, B. M. Automatic classification of time-variable x-ray sources. *The Astrophysical Journal*, 786(1):20, 2014. URL http://stacks.iop.org/0004-637X/786/i=1/a=20.

Lombardi, M., Alves, J., and Lada, C. J. 2MASS wide field extinction maps. IV. The Orion, Monoceros R2, Rosette, and Canis Major star forming regions. *A&A*, 535:A16, November 2011. doi: 10.1051/0004-6361/201116915.

Mac Low, M.-M., McCray, R., and Norman, M. L. Superbubble blowout dynamics. *ApJ*, 337:141–154, February 1989. doi: 10.1086/167094.

Majewski, S. R., Zasowski, G., and Nidever, D. L. Lifting the Dusty Veil with Near- and Mid-infrared Photometry. I. Description and Applications of the Rayleigh-Jeans Color Excess Method. *ApJ*, 739:25, September 2011. doi: 10.1088/0004-637X/739/1/25.

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., Allende Prieto, C., Barkhouser, R., Bizyaev, D., Blank, B., Brunner, S., Burton, A., Carrera, R., Chojnowski, S. D., Cunha, K., Epstein, C., Fitzgerald, G., García Pérez, A. E., Hearty, F. R., Henderson, C., Holtzman, J. A., Johnson, J. A., Lam, C. R., Lawler, J. E., Maseman, P., Mészáros, S., Nelson, M., Nguyen, D. C., Nidever, D. L., Pinsonneault, M., Shetrone, M., Smee, S., Smith, V. V., Stolberg, T., Skrutskie, M. F., Walker, E., Wilson, J. C., Zasowski, G., Anders, F., Basu, S., Beland, S., Blanton, M. R., Bovy, J., Brownstein, J. R., Carlberg, J., Chaplin, W., Chiappini, C., Eisenstein, D. J., Elsworth, Y., Feuillet, D., Fleming, S. W., Galbraith-Frew, J., García, R. A., García-Hernández, D. A., Gillespie, B. A., Girardi, L., Gunn, J. E., Hasselquist, S., Hayden, M. R., Hekker, S., Ivans, I., Kinemuchi, K., Klaene, M., Mahadevan, S., Mathur, S., Mosser, B., Muna, D., Munn, J. A., Nichol, R. C., O'Connell, R. W., Parejko, J. K., Robin, A. C., Rocha-Pinto, H., Schultheis, M., Serenelli, A. M., Shane, N., Silva Aguirre, V., Sobeck, J. S., Thompson, B., Troup, N. W., Weinberg, D. H., and Zamora, O. The Apache Point Observatory Galactic Evolution Experiment (APOGEE). *AJ*, 154:94, September 2017. doi: 10.3847/1538-3881/aa784d.

Marshall, D. J., Robin, A. C., Reylé, C., Schultheis, M., and Picaud, S. Modelling the Galactic interstellar extinction distribution in three dimensions. *A&A*, 453: 635–651, July 2006. doi: 10.1051/0004-6361:20053842.

Mathis, J. S. Interstellar dust and extinction. *ARA&A*, 28:37–70, 1990. doi: 10.1146/annurev.aa.28.090190.000345.

Menten, K. M., Reid, M. J., Forbrich, J., and Brunthaler, A. The distance to the Orion Nebula. *A&A*, 474:515–520, November 2007. doi: 10.1051/0004-6361:20078247.

Merrill, P. W. Unidentified Interstellar Lines. *PASP*, 46:206–207, August 1934. doi: 10.1086/124460.

Michalik, D., Lindegren, L., Hobbs, D., and Butkevich, A. G. Gaia astrometry for stars with too few observations. A Bayesian approach. *A&A*, 583:A68, November 2015. doi: 10.1051/0004-6361/201526936.

Mie, G. Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen. *Annalen der Physik*, 330:377–445, 1908. doi: 10.1002/andp.19083300302.

Mishra, A. and Li, A. Probing the Role of Carbon in the Interstellar Ultraviolet Extinction. *ApJ*, 809:120, August 2015. doi: 10.1088/0004-637X/809/2/120.

Morgan, W. W. The Spiral Structure of the Galaxy. *Scientific American*, 192:42–47, May 1955. doi: 10.1038/scientificamerican0555-42.

Morgan, W. W., Whitford, A. E., and Code, A. D. Studies in Galactic Structure. I. a Preliminary Determination of the Space Distribution of the Blue Giants. *ApJ*, 118:318, September 1953. doi: 10.1086/145754.

Nishiyama, S., Tamura, M., Hatano, H., Kato, D., Tanabé, T., Sugitani, K., and Nagata, T. Interstellar Extinction Law Toward the Galactic Center III: J, H, K_S Bands in the 2MASS and the MKO Systems, and 3.6, 4.5, 5.8, 8.0 μm in the Spitzer/IRAC System. *ApJ*, 696:1407–1417, May 2009. doi: 10.1088/0004-637X/696/2/1407.

Ochsendorf, B. B., Brown, A. G. A., Bally, J., and Tielens, A. G. G. M. Nested Shells Reveal the Rejuvenation of the Orion-Eridanus Superbubble. *ApJ*, 808:111, August 2015. doi: 10.1088/0004-637X/808/2/111.

O'dell, C. R. The Orion Nebula and its Associated Population. *ARA&A*, 39:99–136, 2001. doi: 10.1146/annurev.astro.39.1.99.

O'Donnell, J. E. R_{nu}-dependent optical and near-ultraviolet extinction. *ApJ*, 422:158–163, February 1994. doi: 10.1086/173713.

Oort, J. H. and Muller, C. A. Spiral Structure and Interstellar Emission. *Monthly Notes of the Astronomical Society of South Africa*, 11:65, 1952.

Oort, J. H. and van de Hulst, H. C. Gas and smoke in interstellar space. *BAN*, 10:187, November 1946.

Parsons, H., Thompson, M. A., Clark, J. S., and Chrysostomou, A. Molecular clumps in the W51 giant molecular cloud. *MNRAS*, 424:1658–1671, August 2012. doi: 10.1111/j.1365-2966.2012.21106.x.

Pineau, F.-X., Motch, C., Carrera, F., Della Ceca, R., Derrière, S., Michel, L., Schwope, A., and Watson, M. G. Cross-correlation of the 2XMMi catalogue with Data Release 7 of the Sloan Digital Sky Survey. *A&A*, 527:A126, March 2011. doi: 10.1051/0004-6361/201015141.

Planck Collaboration, Abergel, A., Ade, P. A. R., Aghanim, N., Alves, M. I. R., Aniano, G., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., and et al. Planck 2013 results. XI. All-sky model of thermal dust emission. *A&A*, 571:A11, November 2014. doi: 10.1051/0004-6361/201323195.

Pon, A. Blowing Bubbles in Space: The Orion-Eridanus Superbubble. *JRASC*, 109: 7, February 2015.

Pon, A., Johnstone, D., Bally, J., and Heiles, C. Kompaneets model fitting of the Orion-Eridanus superbubble. *MNRAS*, 444:3657–3669, November 2014. doi: 10.1093/mnras/stu1704.

Rachford, B. L., Snow, T. P., Tumlinson, J., Shull, J. M., Blair, W. P., Ferlet, R., Friedman, S. D., Gry, C., Jenkins, E. B., Morton, D. C., Savage, B. D., Sonnentrucker, P., Vidal-Madjar, A., Welty, D. E., and York, D. G. A Far Ultraviolet Spectroscopic Explorer Survey of Interstellar Molecular Hydrogen in Translucent Clouds. *ApJ*, 577:221–244, September 2002. doi: 10.1086/342146.

Rasmussen, C. E. and Williams, C. Gaussian processes for machine learning. 2006.

Rayleigh. Scattering of Light by Resonating Molecules. *Nature*, 104:276, November 1919. doi: 10.1038/104276c0.

Reid, M. J., Menten, K. M., Brunthaler, A., Zheng, X. W., Dame, T. M., Xu, Y., Wu, Y., Zhang, B., Sanna, A., Sato, M., Hachisuka, K., Choi, Y. K., Immer, K., Moscadelli, L., Rygl, K. L. J., and Bartkiewicz, A. Trigonometric Parallaxes of High Mass Star Forming Regions: The Structure and Kinematics of the Milky Way. *ApJ*, 783:130, March 2014. doi: 10.1088/0004-637X/783/2/130.

Rezaei, K. S., Bailer-Jones, C. A. L., Schlafly, E. F., and Fouesneau, M. Three-dimensional dust mapping in the Orion complex, combining Gaia-TGAS, 2MASS, and WISE. *A&A*, 616:A44, August 2018. doi: 10.1051/0004-6361/201732503.

Rezaei Kh., S., Bailer-Jones, C. A. L., Hanson, R. J., and Fouesneau, M. Inferring the three-dimensional distribution of dust in the Galaxy with a non-parametric method . Preparing for Gaia. *A&A*, 598:A125, February 2017. doi: 10.1051/0004-6361/201628885.

Rezaei Kh., S., Bailer-Jones, C. A. L., Hogg, D. W., and Schultheis, M. Detection of the Milky Way spiral arms in dust from 3D mapping. *ArXiv e-prints*, July 2018.

Robin, A. C., Luri, X., Reylé, C., Isasi, Y., Grux, E., Blanco-Cuaresma, S., Arenou, F., Babusiaux, C., Belcheva, M., Drimmel, R., Jordi, C., Krone-Martins, A.,

Masana, E., Mauduit, J. C., Mignard, F., Mowlavi, N., Rocca-Volmerange, B., Sartoretti, P., Slezak, E., and Sozzetti, A. Gaia Universe model snapshot. A statistical analysis of the expected contents of the Gaia catalogue. *A&A*, 543: A100, July 2012. doi: 10.1051/0004-6361/201118646.

Rodrigues, T. S., Girardi, L., Miglio, A., Bossini, D., Bovy, J., Epstein, C., Pinsonneault, M. H., Stello, D., Zasowski, G., Prieto, C. A., Chaplin, W. J., Hekker, S., Johnson, J. A., Mészáros, S., Mosser, B., Anders, F., Basu, S., Beers, T. C., Chiappini, C., da Costa, L. A. N., Elsworth, Y., García, R. A., Pérez, A. E. G., Hearty, F. R., Maia, M. A. G., Majewski, S. R., Mathur, S., Montalbán, J., Nidever, D. L., Santiago, B., Schultheis, M., Serenelli, A., and Shetrone, M. Bayesian distances and extinctions for giants observed by Kepler and APOGEE. *MNRAS*, 445:2758–2776, December 2014. doi: 10.1093/mnras/stu1907.

Röhser, T., Kerp, J., Lenz, D., and Winkel, B. All-sky census of Galactic high-latitude molecular intermediate-velocity clouds. *A&A*, 596:A94, December 2016. doi: 10.1051/0004-6361/201629141.

Rosen, S. R., Webb, N. A., Watson, M. G., Ballet, J., Barret, D., Braito, V., Carrera, F. J., Ceballos, M. T., Coriat, M., Della Ceca, R., Denkinson, G., Esquej, P., Farrell, S. A., Freyberg, M., Grise, F., Guillout, P., Heil, L., Law-Green, D., Lamer, G., Lin, D., Martino, R., Michel, L., Motch, C., Nebot Gomez-Moran, A., Page, C. G., Page, K., Page, M., Pakull, M. W., Pye, J., Read, A., Rodriguez, P., Sakano, M., Saxton, R., Schwope, A., Scott, A. E., Sturm, R., Traulsen, I., Yershov, V., and Zolotukhin, I. VizieR Online Data Catalog: XMM-Newton Serendipitous Source Catalogue 3XMM-DR6 (XMM-SSC, 2016). *VizieR Online Data Catalog*, 9050, September 2016.

Rudnick, J. On the Reddening in B-Type Stars. *ApJ*, 83:394, May 1936. doi: 10.1086/143733.

Russeil, D. Star-forming complexes and the spiral structure of our Galaxy. *A&A*, 397:133–146, January 2003. doi: 10.1051/0004-6361:20021504.

Sale, S. E. 3D extinction mapping using hierarchical Bayesian models. *MNRAS*, 427:2119–2131, December 2012. doi: 10.1111/j.1365-2966.2012.21662.x.

Sale, S. E. and Magorrian, J. Three-dimensional extinction mapping using Gaussian random fields. *MNRAS*, 445:256–269, November 2014. doi: 10.1093/mnras/stu1728.

Sale, S. E., Drew, J. E., Barentsen, G., Farnhill, H. J., Raddi, R., Barlow, M. J., Eislöffel, J., Vink, J. S., Rodríguez-Gil, P., and Wright, N. J. A 3D extinction map of the northern Galactic plane based on IPHAS photometry. *MNRAS*, 443: 2907–2922, October 2014. doi: 10.1093/mnras/stu1090.

Sato, M., Reid, M. J., Brunthaler, A., and Menten, K. M. Trigonometric Parallax of W51 Main/South. *ApJ*, 720:1055–1065, September 2010. doi: 10.1088/0004-637X/720/2/1055.

schalén, C. . *Astron. nachr.*, 236:249–267, August 1929.

Schinnerer, E., Meidt, S. E., Colombo, D., Chandar, R., Dobbs, C. L., García-Burillo, S., Hughes, A., Leroy, A. K., Pety, J., Querejeta, M., Kramer, C., and Schuster, K. F. The PdBI Arcsecond Whirlpool Survey (PAWS): The Role of Spiral Arms in Cloud and Star Formation. *ApJ*, 836:62, February 2017. doi: 10.3847/1538-4357/836/1/62.

Schlafly, E. F., Finkbeiner, D. P., Schlegel, D. J., Jurić, M., Ivezić, Ž., Gibson, R. R., Knapp, G. R., and Weaver, B. A. The Blue Tip of the Stellar Locus: Measuring Reddening with the Sloan Digital Sky Survey. *ApJ*, 725:1175–1191, December 2010. doi: 10.1088/0004-637X/725/1/1175.

Schlafly, E. F., Green, G., Finkbeiner, D. P., Jurić, M., Rix, H.-W., Martin, N. F., Burgett, W. S., Chambers, K. C., Draper, P. W., Hodapp, K. W., Kaiser, N., Kudritzki, R.-P., Magnier, E. A., Metcalfe, N., Morgan, J. S., Price, P. A., Stubbs, C. W., Tonry, J. L., Wainscoat, R. J., and Waters, C. A Map of Dust Reddening to 4.5 kpc from Pan-STARRS1. *ApJ*, 789:15, July 2014. doi: 10.1088/0004-637X/789/1/15.

Schlafly, E. F., Green, G., Finkbeiner, D. P., Rix, H.-W., Burgett, W. S., Chambers, K. C., Draper, P. W., Kaiser, N., Martin, N. F., Metcalfe, N., Morgan, J. S., Price, P. A., Tonry, J. L., Wainscoat, R. J., and Waters, C. Three-dimensional Dust Mapping Reveals that Orion Forms Part of a Large Ring of Dust. *ApJ*, 799:116, February 2015. doi: 10.1088/0004-637X/799/2/116.

Schlafly, E. F., Meisner, A. M., Stutz, A. M., Kainulainen, J., Peek, J. E. G., Tchernyshyov, K., Rix, H.-W., Finkbeiner, D. P., Covey, K. R., Green, G. M., Bell, E. F., Burgett, W. S., Chambers, K. C., Draper, P. W., Flewelling, H., Hodapp, K. W., Kaiser, N., Magnier, E. A., Martin, N. F., Metcalfe, N., Wainscoat, R. J., and Waters, C. The Optical-infrared Extinction Curve and Its Variation in the Milky Way. *ApJ*, 821:78, April 2016. doi: 10.3847/0004-637X/821/2/78.

Schlegel, D. J., Finkbeiner, D. P., and Davis, M. Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds. *ApJ*, 500:525–553, June 1998. doi: 10.1086/305772.

Schneider, N., Bontemps, S., Simon, R., Jakob, H., Motte, F., Miller, M., Kramer, C., and Stutzki, J. A new view of the Cygnus X region. KOSMA ^{13}CO 2 to 1, 3 to 2, and ^{12}CO 3 to 2 imaging. *A&A*, 458:855–871, November 2006. doi: 10.1051/0004-6361:20065088.

Schultheis, M., Zasowski, G., Allende Prieto, C., Anders, F., Beaton, R. L., Beers, T. C., Bizyaev, D., Chiappini, C., Frinchaboy, P. M., García Pérez, A. E., Ge,

J., Hearty, F., Holtzman, J., Majewski, S. R., Muna, D., Nidever, D., Shetrone, M., and Schneider, D. P. Extinction Maps toward the Milky Way Bulge: Two-dimensional and Three-dimensional Tests with APOGEE. *AJ*, 148:24, July 2014. doi: 10.1088/0004-6256/148/1/24.

Shapley, H. and Curtis, H. D. The Scale of the Universe. *Bulletin of the National Research Council, Vol. 2, Part 3, No. 11, p. 171-217*, 2:171–217, May 1921.

Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam, R., Van Dyk, S., and Wheelock, S. The Two Micron All Sky Survey (2MASS). *AJ*, 131:1163–1183, February 2006. doi: 10.1086/498708.

Slipher, V. M. On the spectrum of the nebula in the Pleiades. *Lowell Observatory Bulletin*, 2:26–27, 1912.

Soler, J. D., Bracco, A., and Pon, A. The magnetic environment of the Orion-Eridanus superbubble as revealed by Planck. *A&A*, 609:L3, February 2018. doi: 10.1051/0004-6361/201732203.

Stebbins, J., Huffer, C. M., and Whitford, A. E. Space Reddening in the Galaxy. *ApJ*, 90:209, September 1939. doi: 10.1086/144099.

Tang, J., Bressan, A., Rosenfield, P., Slemer, A., Marigo, P., Girardi, L., and Bianchi, L. New PARSEC evolutionary tracks of massive stars at low metallicity: testing canonical stellar evolution in nearby star-forming dwarf galaxies. *MNRAS*, 445:4287–4305, December 2014. doi: 10.1093/mnras/stu2029.

Tielens, A. G. G. M. *The Physics and Chemistry of the Interstellar Medium*. Cambridge University Press, August 2005.

Tielens, A. G. G. M. Interstellar Polycyclic Aromatic Hydrocarbon Molecules. *ARA&A*, 46:289–337, September 2008. doi: 10.1146/annurev.astro.46.060407.145211.

Tielens, A. G. G. M. *The Physics and Chemistry of the Interstellar Medium*. August 2010.

Trumpler, R. J. Absorption of Light in the Galactic System. *PASP*, 42:214, August 1930. doi: 10.1086/124039.

van de Hulst, H. C. *Light Scattering by Small Particles*. 1957.

van de Hulst, H. C., Muller, C. A., and Oort, J. H. The spiral structure of the outer part of the Galactic System derived from the hydrogen emission at 21 cm wavelength. *BAN*, 12:117, May 1954.

Vergely, J.-L., Valette, B., Lallement, R., and Raimond, S. Spatial distribution of interstellar dust in the Sun's vicinity. Comparison with neutral sodium-bearing gas. *A&A*, 518:A31, July 2010. doi: 10.1051/0004-6361/200913962.

Whitaker, J. S., Jackson, J. M., Rathborne, J. M., Foster, J. B., Contreras, Y., Sanhueza, P., Stephens, I. W., and Longmore, S. N. MALT90 Kinematic Distances to Dense Molecular Clumps. *AJ*, 154:140, October 2017. doi: 10.3847/1538-3881/aa86ad.

Wienen, M., Wyrowski, F., Menten, K. M., Urquhart, J. S., Csengeri, T., Walmsley, C. M., Bontemps, S., Russeil, D., Bronfman, L., Koribalski, B. S., and Schuller, F. ATLASGAL - Kinematic distances and the dense gas mass distribution of the inner Galaxy. *A&A*, 579:A91, July 2015. doi: 10.1051/0004-6361/201424802.

Woolf, N. J. and Ney, E. P. Circumstellar Infrared Emission from Cool Stars. *ApJL*, 155:L181, March 1969. doi: 10.1086/180331.

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., Ressler, M. E., Cutri, R. M., Jarrett, T., Kirkpatrick, J. D., Padgett, D., McMillan, R. S., Skrutskie, M., Stanford, S. A., Cohen, M., Walker, R. G., Mather, J. C., Leisawitz, D., Gautier, T. N., III, McLean, I., Benford, D., Lonsdale, C. J., Blain, A., Mendez, B., Irace, W. R., Duval, V., Liu, F., Royer, D., Heinrichsen, I., Howard, J., Shannon, M., Kendall, M., Walsh, A. L., Larsen, M., Cardon, J. G., Schick, S., Schwalm, M., Abid, M., Fabinsky, B., Naes, L., and Tsai, C.-W. The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. *AJ*, 140: 1868-1881, December 2010. doi: 10.1088/0004-6256/140/6/1868.

Zari, E., Brown, A. G. A., de Bruijne, J., Manara, C. F., and de Zeeuw, P. T. Mapping young stellar populations toward Orion with Gaia DR1. *A&A*, 608: A148, December 2017. doi: 10.1051/0004-6361/201731309.

Zasowski, G., Johnson, J. A., Frinchaboy, P. M., Majewski, S. R., Nidever, D. L., Rocha Pinto, H. J., Girardi, L., Andrews, B., Chojnowski, S. D., Cudworth, K. M., Jackson, K., Munn, J., Skrutskie, M. F., Beaton, R. L., Blake, C. H., Covey, K., Deshpande, R., Epstein, C., Fabbian, D., Fleming, S. W., Garcia Hernandez, D. A., Herrero, A., Mahadevan, S., Mészáros, S., Schultheis, M., Sellgren, K., Terrien, R., van Saders, J., Allende Prieto, C., Bizyaev, D., Burton, A., Cunha, K., da Costa, L. N., Hasselquist, S., Hearty, F., Holtzman, J., García Pérez, A. E., Maia, M. A. G., O'Connell, R. W., O'Donnell, C., Pinsonneault, M., Santiago, B. X., Schiavon, R. P., Shetrone, M., Smith, V., and Wilson, J. C. Target Selection for the Apache Point Observatory Galactic Evolution Experiment (APOGEE). *AJ*, 146:81, October 2013. doi: 10.1088/0004-6256/146/4/81.