Galactic dust and dynamics

Reimar Leike

München 2020

Galactic dust and dynamics

Reimar Leike

Dissertation
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Reimar Leike
aus Berlin

München, den 03.06.2020

Contents

List of Figures

List of Tables

Zusammenfassung

Physik befasst sich mit der Modellierung der Welt. Ein Modell zu bauen kann zwei Bedeutungen haben: Einerseits kann man damit die Konstruktion eines Modells bezeichnen, das das Verhalten eines Systems imitiert, eine Simulation. Andererseits kann ein Modell etwas bezeichnen, das Aspekte des Originals zeigt, nur nicht so groß ist, z.B. eine Karte. Diese Dissertation beschäftigt sich mit der Modellierung der Welt in beiderlei Bedeutungen, und verbindet diese auch.

Wir betrachten eine Karte als bedingte Wahrscheinlichkeit, denn die Karte hat Freiheitsgrade, die Aussagen über das System ermöglichen. Kartografiert man ein zeitveränderliches System, so muss man Karten erneuern wenn das System sich verändert. Kennt man die Zeitevolution des Systems, so kann man Aussagen einer Karte in die Zukunft extrapolieren. Die Freiheitsgrade einer erneuerten Karte sollte man dann so wählen, dass man möglichst wenig Informationen über das System verliert. Folgt man diesem Paradigma wiederholt, so erhält man eine Simulation des Systems, abgebildet durch die Serie an Karten. Auf diese Art und Weise leiten wir Simulationen eines einfachen fluiddynamischen Systems von Grund auf her. Dabei ist die durch die Karte induzierte bedingte Wahrscheinlichkeit entscheidend, da sie die einzige Stellschraube für das resultierende Simulationsschema ist.

Der zweite Teil dieser Arbeit behandelt das Erstellen von dreidimensionalen Karten von galaktischem Staub. Auch hierbei spielt die Wahl der bedingten Wahrscheinlichkeit, die von der Karte induziert wird, eine zentrale Rolle. Wir modellieren Staub als ein korreliertes Feld, wobei der Grad der Korrelation ein zusätzlicher Parameter der Karte ist. Um die Parameter der Karte zu inferieren werden Daten über Staub in drei Dimensionen benötigt. Diese beziehen wir aus Sternenkatalogen, die Informationen über die Staubdichte durch Abdunklungswerte von Sternen enthält; Sternen von welchen auch die Positionen zu gewissem Grad bekannt sind. Drei Staubkarten werden hier präsentiert. Die erste Staubkarte verwendete synthetische Daten und dient der Validierung unseres Ansatzes. Der zweiten Staubkarte liegt der neuste und präziseste Katalog von Sternen, durchgeführt von dem Gaia Satelliten, zu Grunde. Die finale Staubkarte benutzt Daten von allen größeren öffentlichen Katalogen von Sternen zusammen. Diese Karte zeigt die Abdunklung durch Staub bis zu einer Distanz von 1000 Lichtjahren in drei Dimensionen. Sie ist sowohl für Beobachter zur Korrektur von Staubabsorption relevant, als auch für Astrophysiker, die sich für die Zusammensetzung des interstellaren Mediums interessieren. Auch Parameter von Simulationen des interstellaren Mediums können durch die hergeleiteten statistischen Eigenschaften eingeschränkt werden.

Zusammenfassend demonstriert diese Arbeit die Wichtigkeit von Modellen und deren Aussagen über die Realität, sowie die Bedeutung statistischer Analysen, die von Grund auf hergeleitet werden.

Abstract

Physics is about building a model of the world. Building a model can have two different interpretations. On the one hand, it can refer to the construction of a model that mimics the behavior of a system, i.e. in the form of a simulation. On the other hand it can denote the process of building something that has properties of the original, i.e. a map. This dissertation contributes to modeling the world in both meanings of the word, and also connects them.

We regard a map as a conditional probability, the map has degrees of freedom that constrain the mapped system. Maps of time variable systems have to be updated as the system evolves. Given only the information that a map contains about a system at a previous point in time, and the time evolution of the system, the degrees of freedom of an updated map should be selected such that the least amount of information about the system is lost. Iterating this procedure, one obtains a simulation scheme, as the time evolution of the system is imprinted in the sequence of maps. In this thesis, simulation schemes for a simple fluid dynamic equation are constructed this way from first principles. Of paramount importance is the conditional probability of the system given the map data, as it is the only way to influence the resulting simulation scheme.

The second part of this thesis focuses on constructing three dimensional maps of the Galactic dust. In this application one has to specify as well, which statements the map degrees of freedom make about the actual distribution of Galactic dust. We choose to model dust as a correlated field, where the degree of correlation is an additional parameter of the map. To infer the parameters of the map, data about dust in three dimensions is needed. To this end, data from stellar surveys are used, which reflects dust density through the extinction towards millions of sources; sources of which also the distance is known to a limited precision. Three dust maps are presented, one using simulated data through which we verify the validity of our approach, one using data from the most recent and precise stellar survey obtained by the Gaia satellite, and a final map using data from a combination of many larger stellar surveys that are available. Our final result is a map showing the extinction due to Galactic dust up to a distance of about 1000 light years in three dimensions. The map is of importance for observers, to whom dust extinction comprises a foreground to observations, as well as for astrophysicists interested in the composition and structure of the interstellar medium. Also parameters of simulations of the interstellar medium can be constrained using our derived statistical properties.

In conclusion, this thesis demonstrates the importance of models and how they constrain reality, as well as the impact of statistical analyses that are derived from first principles.

1 Introduction

Humans have long dreamt of predicting the future. This longing is reflected by its omnipresence in the popular literature: From the ancient Greeks to modern Hollywood films, Oracles and prophecies play fundamental roles; and their tales have fascinated people for centuries.

In some way, physics can realize the dream of predicting the future. For example the motion of planetary bodies in the solar system can be predicted to remarkable precision up to thousands of years in the future. Physics is about modeling the world to such a degree of fidelity, that accurate predictions can be made about the future, or more generally, about the behavior of systems.

However, to achieve this goal one needs a model that is predictive for the system at hand, and one needs to be able to predict the behavior of the model. Furthermore, one has to infer the model parameters, such that the model accurately matches the reality one observes. For the motion of planets, it is sufficient to abstract the system to only few parameters, namely the mass, position, and velocity of the massive planets. For other systems, such as the movement of water, a simple abstraction can not be achieved, and models involve many or infinite degrees of freedom. These infinite degrees of freedom are represented by fields, which have degrees of freedom for every point in space and/or time. Often, the behavior of models, which follow reality to a sufficient degree, is too complicated to be explored using analytical calculations.

When constraining models with infinitely degrees of freedom with the finitely many parameters a computer gives access to, statements can in general only be probabilistic. To deal with these probabilistic statements over fields, we employ the formalism of information field theory [28, 65], to which there is a brief introduction in the next section.

1.1 A Formal Introduction to Information Field Theory

In information theory, knowledge I about a quantity s is expressed in terms of a conditional probability $P(s|I)$. This conditional probability can be updated in light of new information d by the use of Bayes theorem:

$$P(s|d, I) = \frac{P(d|s, I)P(s|I)}{P(d|I)} \tag{1.1}$$

To make an example, we will consider the case of a multivariate Gaussian distribution

$$P(s|I) = \mathscr{G}(s|S(I)) \tag{1.2}$$

$$= |2\pi S|^{-1/2} \exp\left(-\frac{1}{2}(s)^T S^{-1}(s)\right) \tag{1.3}$$

where $|X|$ denotes the determinant of X, S is the positive definite symmetric covariance matrix of the Gaussian distribution. We now assert that the new information d is given in form of a linear measurement with Gaussian additive zero-centered noise:

$$d = R(s) + n \tag{1.4}$$

$$\text{with } P(n|I,s) = \mathscr{G}(n|N) , \tag{1.5}$$

with the noise covariance N and the measurement response matrix R. This yields a likelihood

$$P(d|s,I) = \int \mathrm{d}n\, P(d,n|s,I) \tag{1.6}$$

$$= \int \mathrm{d}n\, P(d|n,s,I)\, P(n|s,I) \tag{1.7}$$

$$= \int \mathrm{d}n\, \delta(d - R(s) - n)\, \mathscr{G}(n,N) \tag{1.8}$$

$$= \mathscr{G}(d|R(s),N) , \tag{1.9}$$

where we denote with δ the Dirac delta distribution. To derive the posterior density $P(s|d,I)$, it is useful to introduce the concept of information Hamiltonians. We define

$$H(a|b) \equiv -\ln(P(a|b)) . \tag{1.10}$$

Using this definition, we obtain

$$H(d,s) = H(d|s) + H(s) \tag{1.11}$$

$$\widehat{=} \frac{1}{2}\left(R(s) - d\right)^T N^{-1}\left(R(s) - d\right) + \frac{1}{2}s^T S^{-1}s + \frac{1}{2}\mathrm{tr}\ln(N) + \frac{1}{2}\mathrm{tr}\ln(S) \tag{1.12}$$

$$= \frac{1}{2}\left(s^T\left(R^T N^{-1}R + S^{-1}\right)s + d^T N^{-1}Rs + sR^T N^{-1}d + \mathrm{tr}\ln(N) + \mathrm{tr}\ln(S)\right) , \tag{1.13}$$

where "$\widehat{=}$" denotes equality up to irrelevant constants, in this case constants in both s and d. We define $j = R^T N^{-1}d$ and $D^{-1} = \left(R^T N^{-1}R + S^{-1}\right)$ to obtain

$$H(d,s) = \frac{1}{2}(s - Dj)^T D^{-1}(s - Dj) + H_0(d) , \tag{1.14}$$

where we accumulated all constants independent of s into $H_0(d)$. From this result it is easy to see that the posterior in s is again Gaussian

$$P(s|d) = \mathscr{G}(s|m, D) \tag{1.15}$$

$$\text{with mean } m = Dj \tag{1.16}$$

$$\text{and covariance } D = \left(R^T N^{-1} R + S^{-1}\right)^{-1} . \tag{1.17}$$

The solution to this multivariate Gaussian conditioning is also known as the generalized Wiener Filter.

To go from information theory to information field theory, we have to take the continuum limit, for which we get

$$s_i \to s^x , \tag{1.18}$$

with $x \in \mathbb{R}^n$ denoting the spacial coordinate. We denote the field s with an upper index. This is to distinguish between fields that are element of the vector space, and elements in the dual space more clearly. Distinguishing between space and dual space is especially important in infinite dimensions, as they might have different properties and are not isomorphic in general. To make this consideration more explicit, consider the continuum limit of a Riemann sum:

$$r_n = \sum_{n=1}^{n} w_i s^i , \tag{1.19}$$

where s_i are the field values at e.g. pixels i, and w_i is the volume of the pixels. When we take the limit $n \to \infty$ by refining the discretization of s, the values s_i will stay more or less the same, while the values of w_i will shrink as the number of pixels grow. As a consequence, the field limit of w_i is not sensible, as its values tend to zero, however one can identify w through its action on fields, making it naturally an element of the dual space:

$$r_n \to r_\infty \tag{1.20}$$

$$s_i \to s^x \tag{1.21}$$

$$w_i \to w_x \tag{1.22}$$

$$r_\infty = w_x s^x \tag{1.23}$$

To be able to assign values to w_x, we need to choose a basis of the dual space. One convenient way of doing this is by introducing a metric g_{xy}, such that the space and the dual space can be identified. We choose the standard scalar product

$$g_{xy} s^x t^y = \int \mathrm{d}x \, s^x t^x . \tag{1.24}$$

Using this definition, we can write the action of w as

$$w(s) = w_x s^x = w^x g_{xy} s^y = \int \mathrm{d}x \, w^x s^x = \int \mathrm{d}x \, s^x = g_{xy} 1^x s^y . \tag{1.25}$$

Thus we get $w^x = 1^x$, a field of constant value 1, as expected from a field representing a Riemann integral.

The matrices S and R become operators in the continuum limit, whose action can usually be written down in terms of the standard scalar product as

$$S_{ij} \rightarrow S^{xy} \tag{1.26}$$

$$R_{ij} \rightarrow R^i{}_y \tag{1.27}$$

$$\left(S^{-1}\right)(s) = \left(S^{-1}\right)_{xy} s^y = g_{xz} \int \mathrm{d}y \, \left(S^{-1}\right)^{zy} s^y \tag{1.28}$$

$$R(s) = R^i{}_y s^y = \int \mathrm{d}y \, (R)^{iy} s^y \ . \tag{1.29}$$

Note that we imposed that only s is a field, the data d are still finite, reflected by the index notation of R. Although the metric g_{xy} helps to represent operators, it should be noted that it does not actually influence the result of calculations, as all relevant quantities like

$$H(s) = \frac{1}{2} s^x (S^{-1})_{xy} s^y + H_0 \tag{1.30}$$

$$\text{and } m^x = D^{xy} R^i_x N^{-1}_{ij} d^j \ , \tag{1.31}$$

are naturally arranged such that indices occur in opposing pairs. This has theoretical and practical implications. The theoretical implication is that the result of inference is independent of the coordinate system, a very reassuring observation. The practical consequence is that one does not need to take the metric g_{xy} into account when writing computational code, as it ultimately always cancels out in the end.

For the work carried out in this thesis, the continuum usually has to be discretized again, in order to handle computations on a computer. Nevertheless, it is advantageous to consider this continuum limit. If all computer operations are discretized from the continuum, it is easy to make them consistent to each other. If one in contrast considers the discretization first, then different operations could be discretized inconsistently to each other, leading to numerical problems. We will see an example of this discretization consistency in Sec. 2, where the discretization of space dictates how derivatives have to be discretized.

A question that is often raised, is how to choose an appropriate prior. A good prior should constrain the signal s to sensible ranges while being as agnostic as possible about a-priori indistinguishable outcomes. For example if no point in space is a priori special, it can be argued that the prior covariance matrix should be invariant under shifts:

$$S^{xy} = C(x - y) \ , \tag{1.32}$$

where C is the correlation kernel. Such covariances are diagonal in Fourier space, a property that will be used throughout this thesis. By change into the Fourier basis, one gets:

$$S^{xy} = \left(\mathbb{F}^{-1}\right)^x{}_k S^{kl} \left(\mathbb{F}^{-1}\right)_l{}^y \tag{1.33}$$

$$\text{with } S^{kl} = \begin{cases} 0 & \text{for } k \neq l \\ V P_s(k) & \text{otherwise.} \end{cases} \tag{1.34}$$

where $(\mathbb{F}^{-1})^x{}_k$ denotes the inverse Fourier transform, $(\mathbb{F}^{-1})_l{}^y$ denotes the adjoint inverse Fourier transform, V denotes the total volume of the space s is defined on, and $P_s(k)$ is the power spectrum of s. Noteworthy about this definition is the absence of delta distributions. This absence is achieved by the coordinate free interpretation, if one were to introduce a metric in order to write S^{kl} as the operator $S^k{}_l$, then the delta distribution would reappear. We will use this Fourier representation of the signal covariance matrix S in the upcoming chapters, as it provides the basis for a flexible and sufficiently agnostic prior. However, we will often adopt the standard definition in order to cause less confusion.

1.2 Overview of my scientific work

1.2.1 This Thesis

Numerical simulations are a core pillar of physics since the dawn of the computer age. Models which are defined on the continuum are hereby discretized to allow their exploration on a computer. While simulation on a computer drastically increases the wealth of models that can be investigated, the necessity of discretizing models to arrive at simulation schemes introduces further complications. In Chapter 2 of this thesis, we explore a fundamentally different approach to deriving simulation schemes. Instead of imposing that the differential equation underlying a model has to be discretized, we assume to have discrete knowledge about a continuous system.

This knowledge imposes a probability distribution of plausible continuous systems at each time step of a simulation, parameterized by our finitely many parameters. We then derive how to change the finite number of parameters representing our knowledge, such that minimal information about the system is lost while it evolves in time. We hereby revisit the fundamentals of the approach taken by Enßlin [29] and extent the applicability to the regime of nonlinear partial differential equations. We furthermore present two different parametrizations of the continuous probability distribution, and discuss their performance relative to the commonly used finite difference method. Both these parametrizations are build from an analogy to the inference of fields [27], where it is common to constrain fields that have infinitely many degrees of freedom by the use of finitely many data points.

In the remaining chapters we will use the same statistical language to model the interstellar dust in the vicinity of about one thousand light years of the Sun. We do so by the use of different stellar catalogs, which provide a wealth of data on the stellar parallax, temperature, radius, proper motion, and source intensity at various wavebands. To reconstruct the dust, we are mostly interested in the G-band extinction of sources, which can be calculated by assuming stars to be black-body radiators as well as a model for the wavelength dependence of dust extinction. The parallax of the source supplements the necessary spatial information to enable determining the distance to the reconstructed dust extinction. We then combine the information provided by the individual sources by imposing an underlying coherent three dimensional dust distribution. This dust density is assumed to be correlated, which allows for interpolation. The degree and form of corre-

lation is thereby reconstructed from the data, so that we only interpolate to an realistic extent.

We present a first proof of concept reconstruction in Chapter 3. We verify the validity of our approach by performing a reconstruction on a synthetic dust density that resulted from a simulation of the interstellar medium. For our first reconstruction on real data, we only use data of the Gaia survey, hoping to avoid systematic errors that can occur when merging different datasets. The resulting map shows nearby dust clouds with a distance resolution that surpasses other methods in the field substantially. However, the map also shows some limitations. Smaller dust clouds are not well resolved or might be entirely absent in the reconstruction.

In Chapter 4 we improve upon our first attempt. By using data that was obtained through a combination of four available surveys and by improving on our method in terms of statistical modeling and computational optimization, we are able to achieve significantly higher resolution while at the same time slightly extending the reconstructed volume. The final reconstruction is able to compete with other methods in angular resolution, while far surpassing them in radial resolution. The detailed view we provide enables seeing dust clouds in unprecedented resolution, provides the ground to estimate distances to molecular clouds and allows for insights into the dynamic of the Galaxy itself. Our reconstruction might serve as a benchmark for simulations of the interstellar medium on small scales, as it is the most resolved three dimensional map of dust density distribution so far.

We shed light on the underlying numerical aspects of dust reconstruction in Chapter 5, where we reflect upon properties of and decision about the used minimization routine. The minimization is crucial for performing the dust reconstruction, and the reconstruction performed in this thesis comprises the largest application of information field theory to date.

A summary of our findings can be found in Chapter 6.

1.2.2 Additional Work

Additional to the work presented in this thesis, I have contributed to various other projects. The reconstructions presented in chapter 3 and 4 were performed using the statistical software package NIFTy. I have made contributions to NIFTy 3,4,5 and 6 [7, 55, 88] with ideas and source code. This software package combines statistical functions, Fourier analysis, implicit operators, and optimization algorithms with tools, that allow the discretization of fields. Since version 5, it supports automatic differentiation.

In the domain of radio interferometry, telescopes probe the electromagnetic field at different locations on the Earth. Data from individual telescopes are being correlated to increase the effective telescope size. However, systematic errors arise from the different conditions at the telescope sites, especially from the differing electron density in the ionosphere, which acts as a lens to the incoming radio waves. These systematic errors need to be corrected for. The process of calculating these telescope dependent corrections is called calibration. To infer the calibration solutions, the radio interferometer alternates between observing the science target and a calibration target, which is usually a known

point source. Typically, the initial guess for the calibration is then inferred using the calibration source, and later refined by using the science target. A more elaborate method is proposed by Arras et al. [8], where the calibration solution is reconstructed in parallel to the science target. Hereby, the reconstruction of the science target is informed of the uncertainty of the calibration solution, and vice versa, leading to an overall more consistent reconstruction. My contribution to the work is in various discussions, by providing minor code contributions and by helping to develop the model for correlation structure of the science target, and the calibration solution. I also made contributions to the text of the publication.

A very extreme case of radio interferometer is the Event Horizon Telescope (EHT) [32, 33, 34, 35, 36, 37]. This interferometer reaches an effective size comparable to the size of the Earth by joining and correlating data of many radio telescopes all over the world. Because of the larger distances of the individual telescopes, and because the individual telescopes are build differently, calibration is especially hard for the EHT. One can circumvent this problem by combining the data to calibration invariant quantities, so called closure quantities [11], and using these for inference. I have contributed to a time resolved reconstruction of the black hole shadow of M87* [9] using data from the EHT. In this publication, we present a holistic approach to time domain imaging, where correlations of the science target in space and time are taken into account. My contribution involves development and implementation of the algorithm as well as contribution to the text of the publication.

For some time variable science targets, assuming correlations in time might not be sufficient. One can go one step further and not only reconstruct a time resolved signal, but also its dynamic. This is a highly degenerate problem. As such, inference is problematic, but can be stabilized by imposing additional constraints. In this context, I have contributed to a proof of concept study, where the dynamics is constrained to be local or causal [38]. My contribution includes many scientific discussions as well as helping in reviewing the manuscript.

2 Towards information optimal simulation of partial differential equations

This chapter is published as an article in the journal Physical Review E [63] with me as the first author. I performed the necessary theoretical calculations, programmed and ran the simulation, and wrote the text of the publication. Torsten Enßlin has contributed through many valuable discussions and by reviewing and correcting the text several times. All authors read, commented, and approved the final manuscript.

2.1 Abstract

Most simulation schemes for partial differential equations (PDEs) focus on minimizing a simple error norm of a discretized version of a field. This paper takes a fundamentally different approach; the discretized field is interpreted as data providing information about a real physical field that is unknown. This information is sought to be conserved by the scheme as the field evolves in time. Such an information theoretic approach to simulation was pursued before by information field dynamics (IFD). In this paper we work out the theory of IFD for nonlinear PDEs in a noiseless Gaussian approximation. The result is an action that can be minimized to obtain an informationally optimal simulation scheme. It can be brought into a closed form using field operators to calculate the appearing Gaussian integrals. The resulting simulation schemes are tested numerically in two instances for the Burgers equation. Their accuracy surpasses finite-difference schemes on the same resolution. The IFD scheme, however, has to be correctly informed on the subgrid correlation structure. In certain limiting cases we recover well-known simulation schemes like spectral Fourier Galerkin methods. We discuss implications of the approximations made.

2.2 Introduction

Simulation of partial differential equations (PDEs) is a wide field with countless applications. This stems from the fact that there is no general known analytic solution for most of the interesting, in practice occurring PDEs. Thus one has to resort to simulation in order to make predictions about the behavior of the solutions. PDEs are differential equations for fields, which have infinite degrees of freedom. However, on a computer one is not able

to store the whole field for any point in time. Furthermore the time evolution has to be discretized as well, because time is a continuous variable.

Information field dynamics (IFD) [29] takes an approach that differs slightly on a fundamental level from conventional field discretization. Instead of simulating a discretized field, finite information about the real continuous field is stored in a data vector, as if it were obtained from a measurement. The time evolution of the data is then derived from the evolution of the real field. This interpretation enables the application of information theory, specifically information field theory [28, 65] which is information theory for the reconstruction of fields.

The application of information theory to get an improvement or better understanding of existing numerical methods is not new. One of the early prominent examples is Ref. [24], where Bayesian inference is used to compute integrals. There are also examples of groups working on applying information theory to simulations. Some historic examples are [16, 26, 48, 50] who regard the problem as a hidden Markov model which is then treated in a Bayesian fashion through a filtering approach. See e.g. [15] for an overview, [79] for a more generic review of sequential Monte Carlo methods. There is still ongoing research for the filtering approach, see e.g. [12, 20, 52]. These methods can also be applied to neural networks, see e.g. [23]. In some cases, for example for linear differential equations, one can infer the solution directly [73]. Other approaches focus on parametrizing the posterior as a Gaussian and learning the dynamics in a way motivated by machine learning [74].

The approach that is probably the closest to the one in this paper is described in [5], where stochastic differential equations are approximately solved using a variational approach. The differences to our approach lie in the way the probability density is parametrized and how the KL divergence is used.

All in all, Bayesian simulation is an active and growing field of research.

In our approach, we do not rely on sampling, neither are we restricted to linear PDEs. Instead, we approximate the true evolved probability distribution of solutions by a parametrized one in each time step; and choose the parameters so that the loss of information is minimized. Hereby we parameterize the probability distribution such that it mimics a physical measurement instrument.

In Sec. 2.3 the reader is introduced to the fundamental concepts of IFD and general formula for discretizing PDEs. This formula is then tested in Sec. 2.4 for its numerical performance. Advantages and disadvantages of the proposed scheme are discussed in Sec. 2.5. We conclude in Sec. 2.6.

2.3 General formalism

IFD is a formalism for simulating differential equations for fields $s = s(x, t)$ of the form

$$\frac{\mathrm{d}s}{\mathrm{d}t} = f(s) \tag{2.1}$$

using only the finite resources that are available on computers. This implies that from the infinite degrees of freedom of a field s, only finitely many can be taken into account. IFD

is a specific kind of Bayesian forward simulation scheme.

A forward simulation scheme takes a data vector d_0 and returns a new data vector $d_1 = d_1(d_0)$. Here a data vector is referring to an array of numbers on a computer. The data d_0 are supposed to contain information about the real physical field s_0 at time t_0. What kind of information d contains is also specified by the simulation scheme. The new data vector d_1 is supposed to contain information about the field s_1 at the time $t_1 = t_0 + \mathrm{d}t$. One then iterates the application of this scheme until one arrives at a target time. On this abstract level IFD yields a forward simulation scheme. The difference to the construction of most other simulation schemes is its information theoretic foundation and very restrictive formalism. The formalism is restrictive in the sense that once it is defined what information the data d contains about the field s, the time evolution of the scheme $d_1(d_0)$ is completely specified.

IFD attempts to mimic the optimal Bayesian simulation. In an optimal Bayesian simulation we take the knowledge about the initial conditions $P(s_0|\text{"init"})$ and then compute the time evolved probability density using the exact analytic time evolution

$$P(s_n|\text{"init"}) = P(s_0(s_n)|\text{"init"}) \left\|\left\| \frac{\partial s_0(s_n)}{\partial s_n} \right\|\right\| . \tag{2.2}$$

Here we have assumed that there exists an exact solution for the PDE to be simulated (at least up to a zero set of $P(s_0|\text{"init"})$). Thus there is a one to one mapping between fields s_0 at time t_0 and fields s_n at time t_n, such that we can write the initial field $s_0(s_n)$ as function of the later field s_n, or vice versa. We denote with $\left\|\left\| \frac{\partial s_0(s_n)}{\partial s_n} \right\|\right\|$ the absolute value of the Jacobi determinant that arises from transforming the probability density. Note that no information is lost, since time forward and backward evolution is a one to one mapping of the phase space of the field s. This optimal Bayesian simulation scheme is practically not accessible in most interesting cases because it requires the exact (backward) time evolution to be known, the Jacobi determinant to be computable and the storage of whole probability densities over fields. In this paper, we propose a scheme that aims to overcome these limitations at the cost of losing some information in the process of simulation. It does so by parameterizing the probability density and then evolving these parameters such that the least amount of information is lost in each of the small time steps. We proceed by describing in detail how the probability density is parametrized, then we describe how we approximate the time evolution.

In IFD we store a finite amount of data d on the actual continuous field s, *as if* measured by an instrument whose action is described by a measurement equation of the following form:

$$d = R(s)(t) + n \tag{2.3}$$

Here n is accumulated numerical noise and R is some response function. As an example for R one could choose a matrix of Dirac δ-distributions for R to mimic point measurements of the field at certain locations. This measurement equation does not imply that there is an actual measurement, it just defines the probability theoretic connection between the data

d on our computer and the actual physical field s that we want to simulate. The initial conditions of the PDE will determine the first data d_0, future data will then be determined by the scheme to mimic the time evolution of the field s. To recover the full field s_0 from the data d_0 at time t_0 one can use Bayes theorem:

$$P(s_0|d_0) = \frac{P(d_0|s_0)P(s_0)}{P(d_0)} \ . \tag{2.4}$$

For this, a prior $P(s_0)$ is necessary. It reflects the knowledge about s_0 when no data d_0 are available. A simulation scheme also has to discretize the time evolution such that in a time step from t_0 to $t_1 = t_0 + dt$ the data gets updated from d_0 to d_1. In the language just introduced, the purpose of a simulation scheme is to choose a proper time discretization $d_1(d_0)$ that is as close to the real evolution of $U(s_0) = s_1$ as possible (or even equal if feasible) and a proper discretization R of space such that the features of the field are represented well.

In IFD the time evolution of the data d is defined indirectly, that is we assign d_1 such that the posterior $P(s_1|d_1)$ using our new data d_1 matches the time evolved probability density $P(s_1|d_0)$ using our old data d_0 as well as possible. Note that $P(s_1|d_0)$ is what we defined to be the optimal Bayesian simulation, but simulated only for a small time step dt, where a linearization of the time evolution might still be justified. Because we cannot store the whole density $P(s_1|d_0)$ we store an approximation of it that uses the same parametrization as the probability density $P(s_0|d_0)$ but with new values d_1 assigned to the parameters such that it approximates the time evolved probability density $P(s_1|d_0)$ as well as possible. For probability densities corresponding to a Bayesian belief, there is only one consistent notion of "approximating as well as possible", given the two requirements that the optimal approximation is no approximation and that an approximation can be judged by what it predicts for actual outcomes. We refer to Ref. [62] for a practice-oriented discussion why this uniquely determines the "approximation" KL distance as the appropriate loss function to be used here, see Ref. [10] for the original proof on probability densities. This proposed loss is different than that in the originally proposed IFD scheme [29] and leads to matching the two distributions via

$$D_{\mathrm{KL}}(d_0, d_1) = \int ds_1\, P(s_1|d_0) \ln \frac{P(s_1|d_0)}{P(s_1|d_1)} \tag{2.5}$$

In this matching, d_0 is given and d_1 is searched for, such that the KL divergence serves as an action that is minimized to obtain the discretized time evolution $d_1(d_0)$.

It was also proposed in [62] that for information preserving dynamics, that is for non-stochastic time evolution, one has $P(s_1|d_0) = P(s_0|d_0)\left|\left|\frac{\partial s_0(s_1)}{\partial s_1}\right|\right|$ and therefore this Kullback-Leibler distance is equal to the Kullback-Leibler distance with both probability densities time evolved to the past (note the changed indices):

$$D_{\mathrm{KL}}(d_0, d_1) = \int ds_0\, P(s_0|d_0) \ln \frac{P(s_0|d_0)}{P(s_0|d_1)} \tag{2.6}$$

Note that the equality between Eqs. (2.5) and (2.6) is nothing else than the invariance of the KL under invertible transformations. In this case the transformation is the time evolution of the field s. The latter KL can be calculated once one made a suitable choice for R. For this note that there is a degeneracy between R and s. That means, if R is altered by an invertible operator T

$$R' = RT \tag{2.7}$$

then this is equivalent to instead simulating the differential equation for Ts

$$\frac{\mathrm{d}}{\mathrm{d}t}(Ts) = \frac{\mathrm{d}T}{\mathrm{d}s}\left(f(T^{-1}(Ts))\right) \tag{2.8}$$

and using the unaltered response R. This is because

$$(R')s(t) = (RT)s(t) = R(Ts)(t) \ . \tag{2.9}$$

This provides some freedom to simplify R, thus R can be chosen such that it is linear at the cost of possibly making the time evolution more complicated. If the prior $P(s_0) = \mathscr{G}(s_0, S_0)$ and the noise $P(n_0) = \mathscr{G}(n_0, N_0)$ are zero-centered Gaussian distributions with covariance matrices S_0 and N_0, respectively, then the inverse problem can be solved by a (generalized) Wiener Filter [94] and a Gaussian posterior distribution is obtained:

$$\begin{aligned} P(s_0|d_0) &= \mathscr{G}(s_0 - m_0, D_0) \\ &= |2\pi D_0|^{-\frac{1}{2}} e^{-\frac{1}{2}(s_0-m_0)^\dagger D_0^{-1}(s_0-m_0)} \end{aligned} \tag{2.10}$$

Here $m_0 = D_0 R^\dagger N_0^{-1} d_0$ and $D_0^{-1} = S_0^{-1} + R^\dagger N_0^{-1} R$. One also gets a Gaussian posterior distribution for s_1:

$$P(s_1|d_1) = \mathscr{G}(s_1 - m_1, D_1), \tag{2.11}$$
$$\text{with } D_1^{-1} = S_1^{-1} + R^\dagger N_1^{-1} R \tag{2.12}$$
$$\text{and } m_1 = D_1 R^\dagger N_1^{-1} d_1 \tag{2.13}$$
$$= SR^\dagger \left(RS_1 R^\dagger + N_1\right)^{-1} d_1 \ . \tag{2.14}$$

Note that Eqs. (2.13) and (2.14) are two equivalent ways to obtain a reconstruction m_1. In our paper we will mostly use Eq. (2.14) as the matrix inversion only needs to be computed for $RS_1 R^\dagger + N_1$, which is a finite dimensional operator.

To compute the necessary quantities for our action as given by Eq. (2.6) we have to compute the distribution for s_0 given d_1. It is obtained from the backward time evolution of $P(s_1|d_1)$:

$$P(s_0|d_1) = \mathscr{G}(U(s_0) - m_1, D_1)\left\|\frac{\mathrm{d}U(s_0)}{\mathrm{d}s_0}\right\| \tag{2.15}$$

Here $U(s(t_0)) = s(t_1)$ is the exact analytical time evolution. Using this, the Kullback-Leibler divergence that needs to be minimized so that d_1 is obtained is

$$D_{\mathrm{KL}}(d_0, d_1) = \int \mathrm{d}s_0 \, P(s_0|d_0) \ln \frac{P(s_0|d_0)}{P(s_0|d_1)} \tag{2.16}$$

$$= \int \mathrm{d}s_0 \, \mathscr{G}(s_0 - m_0, D_0)$$

$$\ln \frac{\mathscr{G}(s_0 - m_0, D_0)}{\mathscr{G}(U(s_0) - m_1, D_1) \left\| \frac{\mathrm{d}U(s_0)}{\mathrm{d}s_0} \right\|} \quad .$$

We only minimize for parameters of $\mathscr{G}(U(s_0) - m_1, D_1)$, so we can ignore any additive terms that do not depend on d_1. Thus

$$D_{\mathrm{KL}}(d_0, d_1) \,\widehat{=}$$
$$\int \mathrm{d}s_0 \, \mathscr{G}(s_0 - m_0, D_0) \ln \frac{1}{\mathscr{G}(U(s_0) - m_1, D_1)} \quad . \tag{2.17}$$

Here "$\widehat{=}$" denotes equality up to irrelevant constants, which in this case are constants that are not a function of d_1. These will drop out when the expression is minimized with respect to d_1 later on. Note that the absolute value of the Jacobian $\left\| \frac{\mathrm{d}U(s_0)}{\mathrm{d}s_0} \right\|$ can be ignored because it only depends on s_0. The integral above can be quite difficult to evaluate in general. For integrals involving Gaussian distributions there is however a general method [60] to write down a closed expression for the result. Replacing every instance of s_0 with the field operator

$$O_{m_0} = m_0 + D_0 \frac{\mathrm{d}}{\mathrm{d}m_0} \tag{2.18}$$

allows us to evaluate the integral at the cost of having to evaluate operator expressions. The integral is rewritten as

$$D_{\mathrm{KL}}(d_0, d_1) \,\widehat{=}\, \ln \frac{1}{\mathscr{G}(U(O_{m_0}) - m_1, D_1)}$$

$$\widehat{=} \frac{1}{2}(U(O_{m_0}) - m_1)^\dagger D_1^{-1}(U(O_{m_0}) - m_1)$$

$$+ \frac{1}{2}\mathrm{tr}\left(\ln\left(2\pi D_1\right)\right) \quad . \tag{2.19}$$

We now minimize this Kullback-Leibler divergence with respect to d_1. For this we compute the derivative

$$\frac{\mathrm{d}D_{\mathrm{KL}}(d_0, d_1)}{\mathrm{d}d_1} = \left(\frac{\mathrm{d}m_1}{\mathrm{d}d_1}\right)^\dagger D_1^{-1}(m_1 - U(O_{m_0}))$$

$$= N_1^{-1} R(m_1 - U(O_{m_0})) \tag{2.20}$$

with respect to d_1. We now assume that we leave the noise matrix, the response, and the prior invariant, thus omitting the indices on these operators. At the minimum this derivative is 0, so we can solve it for d_1:

$$0 = N^{-1}R(m_1 - U(O_{m_0}))$$
$$0 = RSR^\dagger \left(N + RSR^\dagger\right)^{-1} d_1 - RU(O_{m_0})$$
$$d_1 = \left(N + RSR^\dagger\right)\left(RSR^\dagger\right)^{-1} RU(O_{m_0}) \tag{2.21}$$

One way to use IFD is to reformulate a PDE like Eq. (2.1) to an ordinary differential equation (ODE), for which potent solvers already exist. For this we expand Eq. (2.21) to first order in dt:

$$d_1 = \left(N + RSR^\dagger\right)\left(RSR^\dagger\right)^{-1} R\left(O_{m_0} + \mathrm{d}t f(O_{m_0})\right)$$
$$d_1 = \left(N + RSR^\dagger\right)\left(RSR^\dagger\right)^{-1} R$$
$$\left(SR^\dagger \left(N + RSR^\dagger\right)^{-1} d_0 + \mathrm{d}t f(O_{m_0})\right)$$
$$d_1 = d_0 + \left(N + RSR^\dagger\right)\left(RSR^\dagger\right)^{-1} R\mathrm{d}t f(O_{m_0}) \tag{2.22}$$

Inserting $d_1 = d_0 + \mathrm{d}t \frac{\mathrm{d}d}{\mathrm{d}t}$ we arrive at an ODE for d:

$$\frac{\mathrm{d}d}{\mathrm{d}t} = \left(N + RSR^\dagger\right)\left(RSR^\dagger\right)^{-1} Rf(O_m) \tag{2.23}$$

Using this in the limit of no-noise $N \to 0$ we get the following compact expression for the updating rule:

$$\frac{\mathrm{d}d}{\mathrm{d}t} = Rf(O_m) = Rf(O_{SR^\dagger\left(RSR^\dagger\right)^{-1}d}) \tag{2.24}$$

Eqs. (2.23) and (2.24) are the central equations of this paper, allowing us to discretize any differential equation. They were derived through minimizing the action given by Eq (2.6) and thus mimic the Bayes optimal simulation up to a minimized information loss. Using Eqs. (2.23) and (2.24) and an appropriate choice of the response R of the virtual measurement connecting field and data, IFD tells us how the differential operators need to be discretized.

2.4 Numerical tests

As a benchmark we simulate the Burgers equation

$$\frac{\partial s}{\partial t} = f(s) = \eta \frac{\partial^2 s}{\partial x^2} - s\frac{\partial s}{\partial x} \; . \tag{2.25}$$

This equation is numerically challenging as it develops shock waves for small diffusion constants η. First, we have to specify our choice of R. We demonstrate the formalism for two different choices of R.

2.4.1 Box grid

We choose

$$R_{ix} = 1_{x_i, x_{i+1}} = \begin{cases} 1 & x_i < x \le x_{i+1} \\ 0 & \text{otherwise.} \end{cases} \tag{2.26}$$

This type of grid is commonly used in simulations. Starting from equation (2.24), we compute:

$$
\begin{aligned}
\frac{\mathrm{d}d}{\mathrm{d}t} &= Rf(O_m) \\
&= R(\Delta O_m - O_m \nabla O_m) \\
&= R(\Delta m - m \nabla m) \\
&\quad - \int \mathrm{d}x\mathrm{d}y\mathrm{d}z \, R_{.x} D_{xy} \frac{\mathrm{d}}{\mathrm{d}m_y}(\nabla_{xz} m_z) \\
&= R\Delta SR^\dagger (RSR^\dagger)^{-1} d \\
&\quad - R\left(SR^\dagger (RSR^\dagger)^{-1} d(\nabla SR^\dagger (RSR^\dagger)^{-1} d)\right) \\
&\quad - \int \mathrm{d}x\mathrm{d}y \, R_{.x} D_{xy} \nabla_{xy}
\end{aligned}
\tag{2.27}
$$

We introduce the short hand notation

$$d' = (RSR^\dagger)^{-1} d \tag{2.28}$$

so that the IFD Burgers scheme simplifies to

$$
\begin{aligned}
\frac{\mathrm{d}d}{\mathrm{d}t} &= \eta R\Delta SR^\dagger d' \\
&\quad - R\left(SR^\dagger d'(\nabla SR^\dagger d')\right) \\
&\quad - \int \mathrm{d}x\mathrm{d}y \, R_{.x} D_{xy} \nabla_{xy}
\end{aligned}
\tag{2.29}
$$

Assuming that our a priori knowledge favors no certain points in space or certain directions, according to the Wiener-Khintchin theorem [95] the covariance operator S has to be diagonal in Fourier space. This is equivalent to a convolution with a convolution kernel C_x in configuration (x-) space, such that

$$(SR^\dagger)_{xi} = C_x * R_{xi}^\dagger = \int \mathrm{d}y \, C_{x-y} R_{yi}^\dagger \ . \tag{2.30}$$

We now compute the three terms of Eq. (2.29) all separately, starting with the term
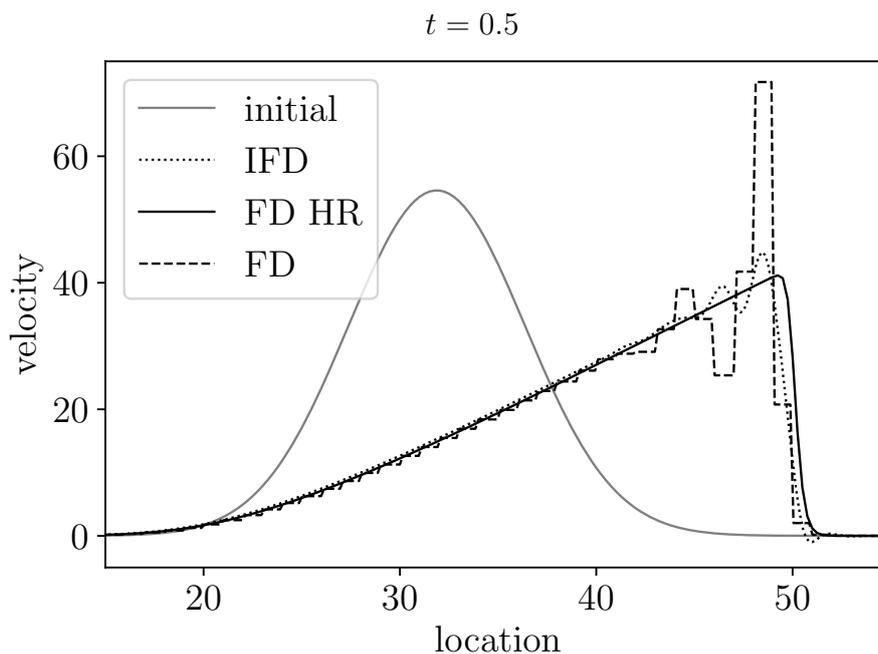
Figure 2.1: Simulation of the Burgers equation using a Gaussian velocity profile as initial condition as represented by the gray line. The dotted line shows the reconstruction as it is obtained from the IFD formalism, the dashed line shows a finite-difference simulation with the same resolution. The solid line is a more exact simulation obtained by simulating with a finite-difference scheme in 4-times higher spatial resolution.
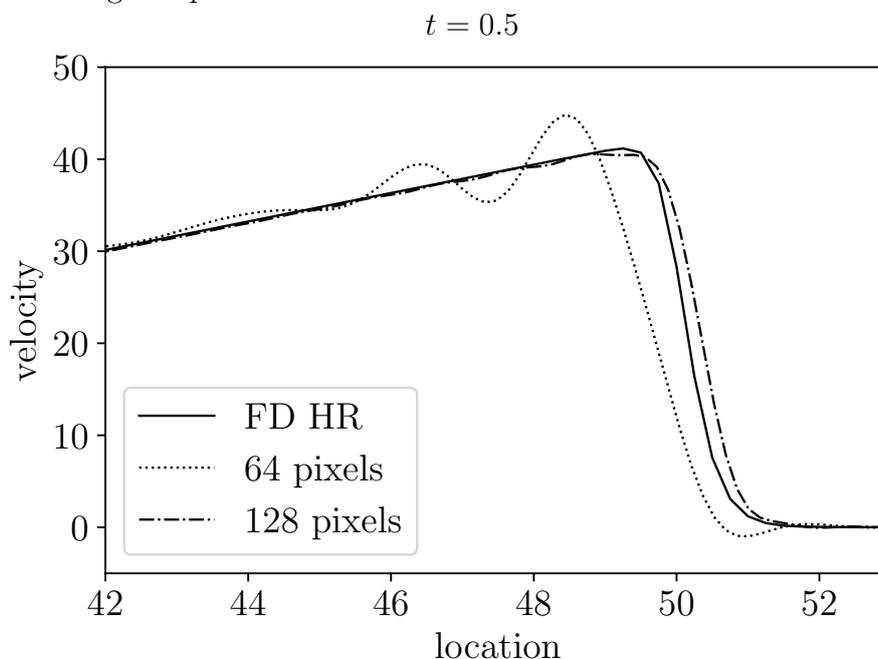
Figure 2.2: Zoom in into the simulation of the Burgers equation shown in Fig. 2.1. The dash-dotted line shows the reconstruction as it is obtained from the IFD formalism, but with double the spatial resolution. The other lines are the same as in Fig. 2.1.

involving the Laplace operator:

$$
\begin{aligned}
\left(R\Delta SR^\dagger d'\right)_i &= \int_{x_i}^{x_{i+1}} \mathrm{d}x \sum_j \Delta \int_{x_j}^{x_{i+j}} \mathrm{d}y \, S_{xy} d'_j \\
&= \int_{x_i}^{x_{i+1}} \mathrm{d}x \sum_j \Delta C_x * 1_{x_j, x_{j+1}} d'_j \\
&= \int \mathrm{d}x \, 1_{x_i, x_{i+1}} \sum_j \Delta C_x * 1_{x_j, x_{j+1}} d'_j \\
&= \int_{x_i}^{x_{i+1}} \mathrm{d}x \sum_j \nabla \left(C_x * \left(\delta(x - x_j) - \delta(x - x_{j+1})\right)\right) d'_j \\
&= \int \mathrm{d}x \left(\delta(x - x_{i+1}) - \delta(x - x_i)\right) \\
&\qquad \sum_j \left(C_{x - x_j} - C_{x - x_{j+1}}\right) d'_j \\
&= \sum_j \left(C_{x_{i+1} - x_j} - C_{x_{i+1} - x_{j+1}}\right) \\
&\qquad - \sum_j \left(C_{x_i - x_j} - C_{x_i - x_{j+1}}\right) d'_j \\
&= \sum_j \left(C_{l(i-j+1)} - 2C_{l(i-j)} + C_{l(i-j-1)}\right) d'_j
\end{aligned}
\tag{2.31}
$$

Here we assumed the x_i to be equally spaced with distance l. Note that this version of the discretized Laplace operator has similarities with the normal finite-difference [21] Laplace operator, but accounts for the field correlation structure. We continue by computing the second term

$$
\begin{aligned}
\left(R\left(SR^\dagger d'(\nabla SR^\dagger d')\right)\right)_i &= \\
&= \int \mathrm{d}x \, R_{ix}(SR^\dagger d')_x(\nabla SR^\dagger d')_x \\
&= -\int \mathrm{d}x \left[\nabla\left(R_{ix}\left(SR^\dagger d'\right)\right)_x\right] (SR^\dagger d')_x \\
&= -\int \mathrm{d}x \left(\nabla R_{ix}\right)\left(SR^\dagger d'\right)\left(SR^\dagger d'\right) \\
&\qquad - \int \mathrm{d}x \, R_{ix}\left(\nabla SR^\dagger d'\right)\left(SR^\dagger d'\right).
\end{aligned}
\tag{2.32}
$$

The last summand in Eq. (2.32) is the same term we started with, only with a negative

sign. Thus we can bring both to the same side of the equation and get

$$
R\left(SR^\dagger d'(\nabla SR^\dagger d')\right) =
$$
$$
= -\frac{1}{2} \int \mathrm{d}x\, (\nabla R_{ix}) \left(SR^\dagger d'\right) \left(SR^\dagger d'\right)
$$
$$
= \frac{1}{2} \left[\left(C_x * \sum_j R_{xj} d'_j\right) \left(C_x * \sum_k R_{xk} d'_k\right) \right]_{x=x_i}^{x_{i+1}} . \tag{2.33}
$$

The third term is

$$
\int \mathrm{d}x\mathrm{d}y\, R_{ix} D_{xy} \nabla_{xy} . \tag{2.34}
$$

This term vanishes in the case of periodic boundary conditions. One can see this by rewriting $\nabla_{xy} = \epsilon^{-1} \left(\delta(x - y + \epsilon) - \delta(x - y - \epsilon)\right)$ for a sufficiently small ϵ to obtain

$$
\int \mathrm{d}x\mathrm{d}y\, R_{ix} D_{xy} \epsilon^{-1} \left(\delta(x - y + \epsilon) - \delta(x - y - \epsilon)\right)
$$
$$
= \int \mathrm{d}x\, \epsilon^{-1} R_{ix} \left(D_{x(x+\epsilon)} - D_{x(x-\epsilon)}\right) . \tag{2.35}
$$

Because S and R have no favored direction, $D_{x(x+\epsilon)} = D_{x(x-\epsilon)}$ and thus the third term vanishes. Finally we have to compute

$$
d' = \left(RSR^\dagger\right)^{-1} d
$$
$$
= \left(\int_{x_i}^{x_{i+1}} \int_{x_j}^{x_{j+1}} \mathrm{d}x\mathrm{d}y\, C_{x-y}\right)^{-1} d_j . \tag{2.36}
$$

Now that one has all the terms of Eq. (2.27) one can choose a prior and obtain a simulation scheme as a result. Normally, one would choose the prior according to physical properties of the system, such that it meaningfully encodes our knowledge in the absence of data. The matter of choosing priors will be further addressed in Sec. 2.5.2. We just want to demonstrate the formalism, so we simply choose the analytic form of C_x such that we can easily compute the three integrals given by Eqs. (2.31, 2.33, 2.36). One convenient choice of C_x is a Gaussian[1], for which we know all the above terms analytically. One might equally well choose any correlation function and do these integrations numerically. Because these integrals only have to be done once, this does not significantly increase the computation time of the resulting simulation scheme.

Note that all computed operators are local, meaning that they fall off as C_x falls off. Thus, they can be truncated at a certain distance and the whole simulation scales only linearly with the number of data points.

---

[1]or a mixture of Gaussians

Fig. 2.1 shows an example of a simulation that was performed using the scheme that was worked out in this section. As a comparison, the figure also shows a simulation using the finite-difference method with the same spatial resolution. This simulation uses 64 data values and the function

$$s(x) = e^{4-(x/64-0.5)^2} \tag{2.37}$$

as initial condition. The simulation space is an interval of length 64 with periodic boundary conditions. The prior covariance was chosen to be a convolution with a zero-centered Gaussian that has a standard deviation of 0.5. The diffusion constant η was chosen to be 5. The result of a simulation using the finite-difference method with a four times higher resolution is displayed as well. This high resolution simulation should capture all features produced by the Burgers dynamics. To investigate the resolution dependence of the IFD scheme, we compare the IFD simulation of Fig. 2.1 with a simulation on a two times more resolved grid in Fig. 2.2. For comparison one can again see the fine resolved finite-difference method. One can clearly see an increased performance as the resolution increases.

2.4.2 Fourier grid

Now we switch to a different response. We choose a Fourier space grid

$$R_{ik} = \sum_i \delta(k - k_i) \tag{2.38}$$

with Fourier space grid points k_i. There is a choice whether to view the Fourier transform

$$F_{kx} = e^{ikx} \tag{2.39}$$

as part of the measurement R or as transformation of the field $s_k = F_{kx}s_x$, see Eqs. (2.7, 2.8, 2.9). We choose the latter and obtain as transformed time evolution:

$$\begin{aligned}
\frac{\mathrm{d}s_k}{\mathrm{d}t} &= F_{kx}\frac{\mathrm{d}s_x}{\mathrm{d}t} \\
&= F_{kx}\left(\Delta s_x - (s_x)(\nabla s_x)\right) \\
&= -k^2 s_k + (s_k) * (iks_k)
\end{aligned} \tag{2.40}$$

We insert this into Eq. (2.24) and obtain

$$\begin{aligned}
\frac{\mathrm{d}d}{\mathrm{d}t} &= Rf(O_m) = \\
&- Rk^2 SR^\dagger \left(RSR^\dagger\right)^{-1} d + Rk^2 SR^\dagger \left(RSR^\dagger\right)^{-1} d \ .
\end{aligned} \tag{2.41}$$

$$t = 0.5$$
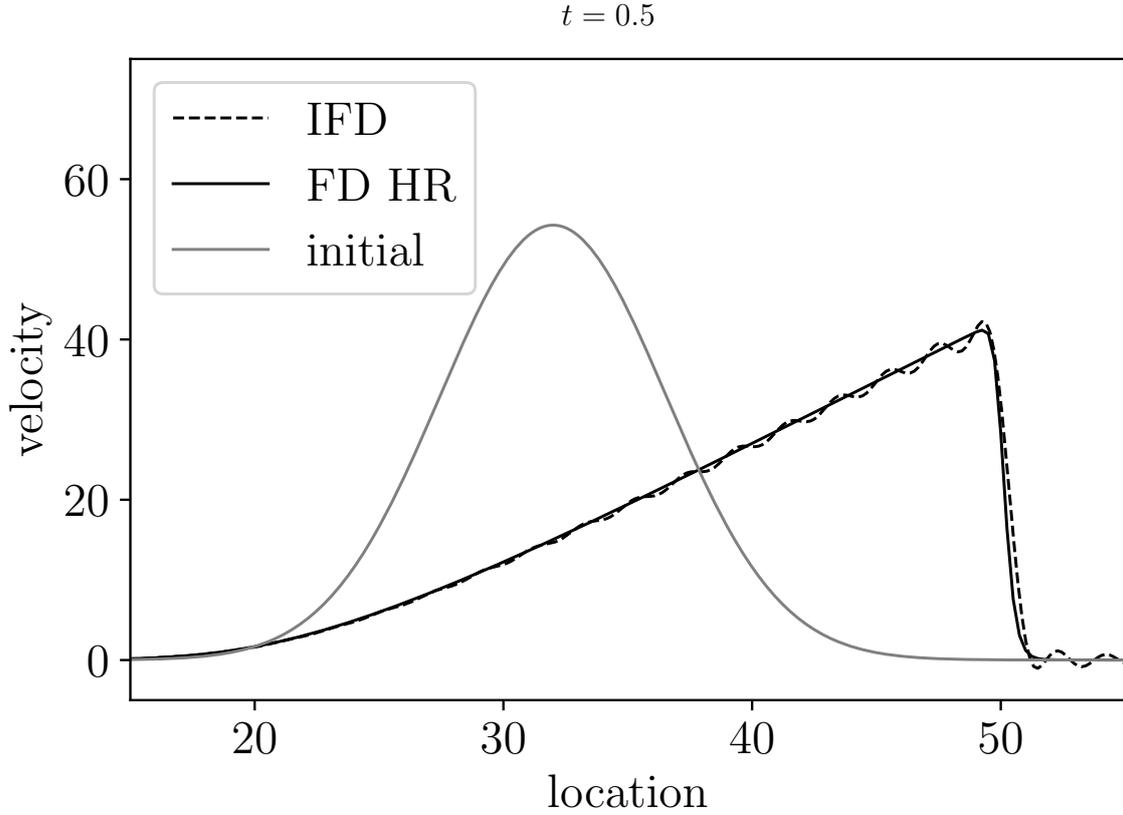
Figure 2.3: Simulation of the Burgers equation using a Gaussian velocity profile as initial condition as represented by the grey line and a Fourier space response. The dotted line is a reconstruction that is obtained through simulating with the IFD scheme. This simulation scheme is equivalent to a Fourier-Galerkin scheme. The solid line is a high resolution simulation using a finite-difference scheme.

If the a priori knowledge does not favor any specific direction or point in space, then S is diagonal in Fourier space. Thus

$$\left(SR^\dagger(RSR^\dagger)^{-1}\right)_{ki} = \sum_{lrjp} S_{kl}\delta_{lm_j}(\delta_{m_ip}S_{pr}\delta_{rm_j})^{-1}$$

$$= \sum_j S_{km_j}(S_{m_im_j})^{-1}$$

$$= \delta_{km_i} = R^\dagger_{ki} \tag{2.42}$$

and the prior covariance drops out, making time evolution on this grid invariant under Fourier space priors. For the final time evolution we arrive at

$$
\begin{aligned}
\frac{\mathrm{d}d_i}{\mathrm{d}t} &= -k_i^2 d_i \\
&\quad + \sum_j \int \mathrm{d}k \mathrm{d}k' \, \delta(k - k_i - k') d_i i k' \delta(k' - k_j) d_j \\
&= -k_i^2 d_i + \sum_j d_{j-i} i k_j d_j
\end{aligned}
\tag{2.43}
$$

where the continuous convolution was translated to a finite convolution on a grid. This resulting time evolution equation can be implemented efficiently and the results can be seen in Fig. 2.3. The simulation constraints, initial conditions, and degrees of freedom were chosen to be the same as in Sec. 2.4.1.

The developed Fourier space IFD method is equivalent to a Fourier Galerkin method [40]. In the Fourier Galerkin method, the error to the correct solution is minimized for a vector in a subspace. This subset is a linear subspace of the real solution space, with selected basis functions that are often chosen to be polynomials or, as in the case of the Fourier Galerkin method, Fourier basis functions. In the case of IFD this subspace is the co-image of the response R. Thus Galerkin schemes can be regarded as IFD schemes with R being specified by the Galerkin basis. The prerequisites for this equivalence are that the prior S commutes with the discretization $R^\dagger R$ and that we work in the no-noise regime $N \to 0$.

2.5 Current advantages and disadvantages

In the current development status of the IFD method, there are some caveats as well as some advantages over classical approaches. Some of them stem from the theoretical side, where approximations had to be done in order to arrive at a computable scheme. In this section we discuss all the observed problems and benefits.

2.5.1 Subgrid structure

Information field dynamics employs the field evolution of the real physical field, and thus automatically takes into account subgrid structure. However, this leads to problems when the time evolution is truncated after the first order in $\mathrm{d}t$. In most physical systems, the time evolution is faster for the smaller scales, thus if we take into account all scales, no $\mathrm{d}t$ is small enough to justify this approximation. In other schemes, the discretized field automatically yields a cutoff at high frequencies. However for IFD, in the derivation we truncate the whole time evolution of the real system at first order, which is not justified.

As an example consider the diffusion equation

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \Delta s \tag{2.44}$$

$$\frac{\mathrm{d}s_k}{\mathrm{d}t} = -k^2 s_k \ . \tag{2.45}$$

For higher modes k the change of s scales with k^2. Thus, if we truncate the Taylor expansion of the time evolution at order of $\mathrm{d}t$, then for $k > \sqrt{\frac{2}{\mathrm{d}t}}$ we get

$$s_k^{t+\mathrm{d}t} = s_k^t - \mathrm{d}t\, k^2 s_k^t = -\left(1 - \mathrm{d}t\, k^2\right) s_k^t \ . \tag{2.46}$$

where $|1 - \mathrm{d}t\, k^2| > 1$, thus the scheme is boosting small frequencies instead of damping them.

This is, however, a general problem of any simulation scheme.

2.5.2 How to choose a prior

The goal of a prior is to incorporate as much information as one has about the system and nothing more. If the system has no special directions or singled out locations, one should choose a prior that is homogeneous and isotropic. These two requirements force the covariance matrix $S_{xy} = \left\langle s_x s_y^\dagger \right\rangle_{P(s)}$ to be diagonal in Fourier space. If we restrict ourselves to Gaussian priors, the prior is fully characterized by its power spectrum $P(k) \propto \left\langle |s_k|^2 \right\rangle_{P(s)}$. Thus, the only a priori information that enters the simulation in the Gaussian case is how smooth the physical field is. But this is also a significant restriction. For example for infinitely sharp shocks, as they occur in the Burgers equation with $\eta = 0$, no smoothness at all is justified at the location of the shock, whereas at other points the solution might be perfectly smooth. To capture this kind of behavior one would need to either use a prior that allows for higher order statistics, use a dynamical prior which evolves in time, or to introduce data that capture the discontinuities. In our simulation we observed that the scheme diverges quickly if an unjustified prior was chosen, that is for example a prior that enforces significantly more smoothness than is present in the solution.

To choose the right power spectrum of the prior one could use a fine grid simulation and take the occurring power spectrum as input for a coarser simulation.

2.5.3 Static prior

In the current formulation we assume that the prior does not evolve in time. However, ideally the prior should evolve with the system. This is because some a priori assumptions that were made for time step t_0 might not be true at a later time (for example initial smoothness might be violated by the formation of a shock). However, from an agnostic point of view, if one has no knowledge about points in time, then the prior should be invariant under time translation. In principle, IFD provides guidance on how to evolve any

kind of degree of information on the field, like its covariance structure, and not only some measurement data. By minimizing Eq. 2.19 with respect to any such piece of information, we obtain an evolution equation for it that loses the minimal amount of information. This way, when the prior is parametrized, an update rule for it is automatically obtained.

2.5.4 No-noise approximation

To derive our algorithms, we take the limit $N \to 0$. While this significantly simplifies the derivation of the schemes, it also deprives the resulting schemes of the advantages that an information theoretic treatment has in general. In the no-noise approximation field configurations s with $Rs \neq d$ are assigned zero probability, thus the information loss in the presence of numerical rounding errors and finite time steps, where the evolution of the data d cannot satisfy Eq. (2.24) exactly, is infinite. Further development in the field of IFD will have to investigate into approaches incorperating noise.

2.6 Conclusion

The requirement of minimal information loss per time-step defines a unique simulation framework. This concrete simulation scheme requires the specification of the field measurement (response) and the incorporation of prior known correlation structure. It exhibits similarities to the finite-difference scheme when the response is a grid of boxes and becomes a spectral scheme in the case of Fourier space response. This yields a new interpretation for linear and in some cases even nonlinear Galerkin schemes. These are information optimal up to the approximations made in this paper if no spatial a priori knowledge about the field is available.

IFD can thus be regarded as a general theory for simulations that explains what assumptions about the simulated field enter a given simulation scheme, if one is able to reproduce that scheme in IFD. For some schemes one can enhance the performance by using a prior that is correctly informed on the field correlation structure. When the prior is chosen incorrectly, for example if it is chosen such that the simulation produces features that are regarded very unlikely by the prior, the scheme tends to diverge quickly. In principle, IFD can provide a guideline how to evolve any degree of freedom; by minimizing Eq. 2.6 one gets a unique simulation scheme. An interesting route, at least for the Burgers and other hydrodynamic equations, would be to automatically infer the position of the virtual measurements, allowing the scheme to sample the field where it is most informative. The investigation in that direction is however beyond the scope of this paper and might be the target of future research.

All in all information theory provides a powerful language to talk about simulation tasks. Even though the series of approximations made in this paper permitted the resulting simulation schemes to only outperform finite differences by a small amount, further advancements in the field could yield substantial enhancements.

3 Charting nearby dust clouds using Gaia data only

This chapter is published as an article in the journal Astronomy and Astrophysics [64] with me as the first author. My contribution is in doing the theoretical calculations, writing and running the underlying code, and writing the text of the publication. Torsten Enßlin has contributed through many valuable discussions and by reviewing and correcting the text several times. All authors read, commented, and approved the final manuscript.

3.1 Abstract

Highly resolved maps of the local Galactic dust are an important ingredient for sky emission models. In nearly the whole electromagnetic spectrum one can see imprints of dust, many of which originate from dust clouds within 300pc. Having a detailed 3D reconstruction of these local dust clouds enables detailed studies, helps to quantify the impact on other observables and is a milestone necessary to enable larger reconstructions, as every sightline for more distant objects will pass through the local dust. To infer the dust density we use parallax and extinction estimates published by the Gaia collaboration in their second data release. We model the dust as a log-normal process using a hierarchical Bayesian model. We also infer non-parametrically the kernel of the log-normal process, which corresponds to the physical spatial correlation power spectrum of the log-density. Using only Gaia data of the second Gaia data release, we reconstruct the 3D dust density and its spatial correlation spectrum in a 600pc cube centered on the Sun. We report a spectral index of the logarithmic dust density of 3.1 on Fourier scales with wavelengths between 2pc and 125pc. The resulting 3D dust map as well as the power spectrum and posterior samples are publicly available for download.

3.2 Introduction

Emission and extinction by Galactic dust is a prominent astronomical foreground at many wavelengths. Therefore, knowing its distribution on the 2D sky and in 3D is essential for many astronomical observations. However, dust is also interesting to be studied on its own, as it provides information about the physical conditions in the interstellar medium and informs us about star forming regions. Dust has been mapped out by surveys for a long time, the first notable contribution being [17]. Their dust reconstruction, as most

reconstructions of the dust distribution so far, was focused on mapping the dust in 2D on the sky. This can be done by looking at the sky in infra-red wavelengths, where it is dominated by dust emission. However, when mapping out dust using infra-red emission one is biased by the radiation field, as dust emission is not only proportional to the dust density but also to the amount of starlight absorbed by the dust. Furthermore, dust maps that were produced by mapping infrared light might contain extended infrared sources that are not within our galaxy, as was shown by [19]. On the other hand, a hypothetical cold dust cloud cannot be seen in infra-red, leading to systematic errors in the analysis of distant targets, for example quasars or the Cosmic Microwave Background (CMB).

For accurate analyses of objects in our galactic vicinity, it is vital to have a 3D dust map as a foreground model, which informs us about regions that cannot be observed, or only be observed with less fidelity, due to dust obscuration. The first non-parametric reconstruction of galactic dust in 3D published is [6]. Since then there have been many attempts to chart the dust density in 3D in increasing resolution, accuracy and for an ever greater part of our galaxy [18, 42, 43, 76, 77, 78, 81, 82, 91, 92]. A large driving force for 3D dust reconstruction and astronomy in general are large surveys like 2MASS [86], Pan-STARRS [51] and SDSS/APOGEE [1] and WISE [96]. These surveys provide photometric measurements and some of them spectra for thousands of stars, from which the calculation of photometric distances is possible. There are two 3D dust reconstructions based on these data sets that are closely related to the approach taken in this paper.

In [58] reddening data from 71 000 sources has been used to perform a 3D reconstruction using Gaussian process regression. The resulting dust map covers a 4kpc square of the galactic plane and 600pc in perpendicular direction with a voxel size of $(5\,\mathrm{pc})^3$.

In [44] a 3D dust map is produced by combining the star data of Pan-STARRS and 2MASS, binning it in angular and distance bins, and performing independent Bayesian reconstructions per angular bin. The result is a dust map that covers three quarters of the sky to a distance up to 2kpc. This reconstruction shows artificial radial structures called the "fingers of God effect" in analogy to the well known phenomenon in cosmology [47, 90]. One way to mitigate this effect is to use more accurate parallax information.

A prominent new survey is performed by the Gaia collaboration [14]. In its second data release (DR2, [13]) accurate parallaxes for roughly 2 billion stars are published. The provided catalog also contains estimates of extinction coefficients for a subset of about 88 million stars, using spectral information of the Gaia satellite's three energy bands. Due to the limited spectral information, the accuracy of the extinction coefficients estimated for individual sightlines is quite low. For this reason it is recommended by [4] to not use the information of individual sightlines but only the joint information of several sightlines. Even though the data quality of individual sightlines is rather low, the sheer amount of data points and the accuracy of the parallaxes outweigh this limitation as our work shows.

So far, 3D dust reconstructions have never been performed using solely Gaia data, instead Gaia data has been used for its accurate parallax measurements only and the more accurate spectral information of other surveys was used [58, for example].

In this paper we present a 3D dust reconstruction using Gaia DR2 data only. The results of the reconstruction are provided online on `https://wwwmpa.mpa-garching.mpg.`

`de/~ensslin/research/data/dust.html` or by its doi:10.5281/zenodo.2577337, and can
be used under the terms of the ODC-By 1.0 license. The inference of the unknown dust den-
sity from the extinction data is performed by a critical filter, a method for Gaussian process
regression with non-parametric kernel learning, first published in [30]. While the statistical
model used here is up to minor details equivalent to the model introduced in that paper,
the algorithm to arrive at an approximate posterior summary statistics is quite different.
The relevant numerical method used in this paper is outlined in [54], which describes
a general method to derive posterior summary statistics for high dimensional Bayesian
inference problems. For a theoretical discussion of the underlying inference framework of
information field theory we refer to [27]. The algorithm was implemented using the Python
package NIFTy5, which is the newest version of the software package NIFTy [7, 83, 88][1].
Even though mathematical theory, statistical motivations, and numerical details are dis-
tributed over the aforementioned papers, this paper is entirely self-contained by describing
the whole method.

In Sec. 3.3 we discuss which part of the Gaia data we used. We introduce our statistical
model of the interstellar dust density as well as of the measurement in Sec. 3.4. In Sec. 3.5
we present a test application of the algorithm using synthetic data, verifying the predictive
power of the algorithm. The main results of the dust density reconstruction using Gaia data
are presented in Sec. 3.6. This section also contains a brief recommendation on how to use
our results. Our dust reconstruction is compared to other 3D dust density reconstructions
in Sec. 3.7. In Sec. 3.8 we summarize the findings of this paper.

3.3 Data

We used the data from the Gaia DR2 catalog by [13], to reconstruct the galactic dust in
the nearby interstellar medium. From the Gaia data archive we extract the parallaxes, the
G-band extinction, the latitude and longitude as well as their respective uncertainties. A
plot of the full Gaia extinction data set can be seen in Fig. 3.1.

We select sources according to the following criteria:

1. the above mentioned data are available for the source

2. the parallax $\widetilde{\omega}$ is inside a 600pc cube around the Sun

3. the relative parallax error is sufficiently low, $\widetilde{\omega}/\sigma_{\widetilde{\omega}} > 5$

4. Priam flag 0100001 or 0100002

The last two criteria are suggested by [4]. There are about 3.7 million stars selected by
these criteria. Fig. 3.1 shows a sky average of the data points used in the reconstruction.
In this data plot one can observe structures present also in other dust maps, for example
the Planck dust map (Fig 3.1).

---

[1]The version of NIFTy used for this reconstruction is available on `https://gitlab.mpcdf.mpg.de/ift/`
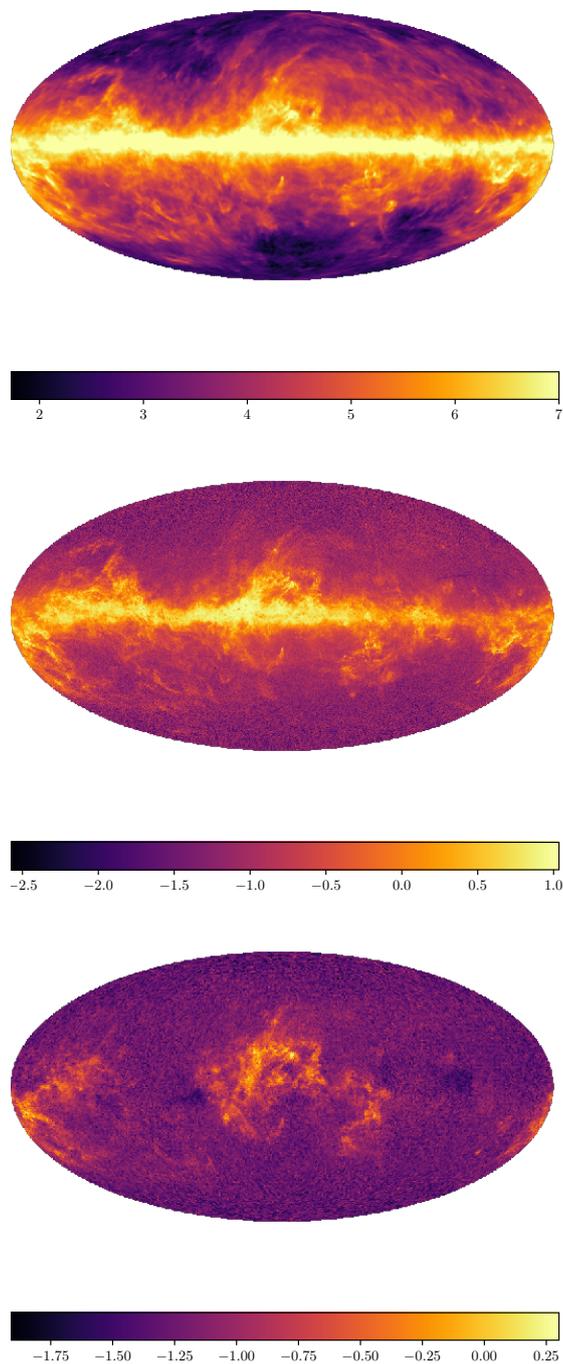`NIFTy` .

Figure 3.1: The top panel shows the natural logarithm of the thermal dust emission map
produced by the Planck collaboration [89]. The color-scheme was saturated
to visually match that of the Gaia dust extinction data shown in the middle
panel. The bottom panel shows the subset of Gaia extinction data within a
600pc cube centered on the Sun; the data used in this paper. The scale is
natural logarithm of the extinction data in magnitudes. Data points in the
same direction were averaged.

3.4 Model

3.4.1 Algorithm

The algorithm is derived from Bayesian reasoning. In Bayesian reasoning information some data d provides about a quantity of interest s is calculated according to Bayes theorem:

$$P(s|d) = \frac{P(d|s)P(s)}{P(d)} \tag{3.1}$$

Note that the quantity of interest can be a (possibly high-dimensional) vector, in our case it is the dust density for every point in space (256^3 degrees of freedom after discretization). There are three main ingredients necessary for the inference of the quantity of interest s:

1. The likelihood $P(d|s)$ of the data d given a realization of the quantity of interest s. We describe our likelihood in Subsec. 3.4.2.

2. The prior $P(s)$ describing the best available knowledge about the quantity of interest s in absence of data. We describe our prior in Subsec. 3.4.3.

3. An inference algorithm that yields a statistical summary of $P(s|d)$ given the joint distribution $P(d, s) = P(d|s)P(s)$ of d and s. We use the inference algorithm described in [54].

The main quantity of interest s is the logarithmic G-band dust extinction cross-section density $s = \ln(\alpha\rho/\mathrm{pc})$, henceforth called the logarithmic dust density. Hereby ρ denotes the actual dust mass density and α the average G-band dust cross section per mass. The value of α is uncertain, which is why we report extinction densities, also called dust pseudo-densities, instead.

We approximate the posterior with a Gaussian

$$Q(s) = \mathscr{G}(s - m, D) \tag{3.2}$$

$$= \frac{\exp\left(-\frac{1}{2}(s - m)^\dagger D^{-1}(s - m)\right)}{|2\pi D|^{\frac{1}{2}}} \,, \tag{3.3}$$

by adopting a suitable mean m and uncertainty dispersion D. The approximation is obtained by minimizing the Kullback-Leibler divergence [57]

$$\mathrm{KL}(Q, P) = \int \mathrm{d}Q \ln\frac{Q}{P} \tag{3.4}$$

with respect to the parameters of Q. This approach is known as variational Bayes [67] or Gibbs free energy approach [31]. The approach we take in finding the unknown approximate posterior mean m and covariance D of Eq. 3.3 is described in detail in [54]. It can be summarized as follows:

1. Calculate the negative log-probability $-\log(P(s, d))$ for the problem, disregarding normalization terms like the evidence $P(d)$.

2. Perform coordinate transformations of the unknown quantities until those are a-priori Gaussian distributed with unit covariance [53, 56].

3. Choose the class of approximating distributions to be Gaussian with variable mean m and covariance $D = (\mathbb{1} + M_m)^{-1}$, where M_m is the Fisher information metric at the current m. Here $\mathbb{1}$ is the contribution of the prior which was transformed in step 2 to have unit covariance. This uncertainty dispersion is a lower bound to the uncertainty [22, 75] and has been shown to be an efficient technique to take cross-correlations between all degrees of freedom into account without having to parameterize them explicitly [54]

4. Minimize Eq. (3.4) with respect to m using Newton Conjugate gradient as second order scheme with the covariance D of Q as curvature. The expectation value with respect to Q is hereby approximated through a set of samples drawn from the approximating distribution Q. Second order minimization by preconditioning with the inverse Fisher metric is also called natural gradient descent [2] in the literature.

A description of the used likelihood and the prior follows.

3.4.2 Likelihood

The likelihood $P(d|s)$ can be split into two parts, one part states how the true extinction depends on the dust density and one part that states how the actual data is distributed given the true extinction on that line of sight. The first part, which we call the response R, states how the unknown dust extinction density ρ imprints itself on the data. The extinction of light on the i-th line of sight L_i is given by the line integral

$$(A_G)_i = [R(\rho)]_i = \int_{L_i} \mathrm{d}l \, \alpha \, \rho(l) \, . \tag{3.5}$$

Here α is the average dust cross section per unit of mass and the line of sight $L_i = L_i(\omega)$ is dependent on the true parallax ω. As noted in section 3.4.1, the value of α is uncertain and we reconstruct the dust extinction density $s = \ln(\alpha\rho/\mathrm{pc})$ instead.

The extinction is additive because the extinction data are given in the magnitudes scale, which is logarithmic. The true parallax ω of the star is uncertain. We assume the true parallax ω to be Gaussian distributed around the published parallax $\widetilde{\omega}$ with a standard deviation equal to the published parallax error σ_ω:

$$P(\omega|\widetilde{\omega}, \sigma_\omega) = \mathscr{G}(\omega - \widetilde{\omega}, \sigma_\omega^2) \tag{3.6}$$

The parallaxes of Gaia DR2 were shown to be Gaussian distributed with incredible reliability by [66]. However, it was also noted by the same authors that there can be outliers. By

restricting ourselves to close-by sources for which G-band extinction values are published, we expect to have cut out most of the outliers.

We do not reconstruct the actual positions of the stars in our reconstruction, thus we have to marginalize them out to obtain the response:

$$P((A_G)_i|\widetilde{\omega}_i, \sigma_{\omega_1}, \rho) =$$
$$= \int d\omega_i \, P((A_G)_i, \omega_i|\widetilde{\omega}_i, \sigma_{\omega_i}, \rho)$$
$$= \int d\omega_i \, P((A_G)_i|\omega_i, \widetilde{\omega}_i, \sigma_{\omega_i}, \rho) P(\omega_i|\widetilde{\omega}_i, \sigma_{\omega_i}, \rho)$$
$$= \int d\omega_i \, P((A_G)_i|\omega_i, \rho) P(\omega_i|\widetilde{\omega}_i, \sigma_{\omega_i})$$
$$= \int d\omega_i \, \delta\left((A_G)_i - \int_{L_i(\omega_i)} dl_i \, \alpha\rho(l_i)\right) P(\omega_i|\widetilde{\omega}_i, \sigma_{\omega_i})$$
$$\approx \delta\left((A_G)_i - \int d\omega_i \int_{L_i(\omega_i)} dl_i \, \alpha\rho(l_i) \, P(\omega_i|\widetilde{\omega}_i, \sigma_{\omega_i})\right)$$
$$= \delta\left((A_G)_i - \int d\omega_i \int_{L_i(0)} dl_i \, \alpha\mathbb{1}_{[0,\frac{1}{\omega}]}(l_i) \, \rho(l_i) \mathscr{G}\left(\omega_i - \widetilde{\omega}_i, \sigma_{\omega_i}^2\right)\right)$$
$$= \delta\left((A_G)_i - \int_{L_i(0)} dl_i \, \alpha\rho(l_i) \, \mathrm{sf}_{\mathscr{G}}\left(\frac{\frac{1}{l_i} - \widetilde{\omega}_i}{\sigma_{\omega_i}}\right)\right) . \tag{3.7}$$

Here

$$\mathrm{sf}_{\mathscr{G}}(x) = 1 - \int_{-\infty}^x dt \, \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}t^2\right) \tag{3.8}$$

denotes the survival function of a standard normal distribution and

$$\mathbb{1}_{[a,b]}(x) = \begin{cases} 1 & \text{for } x \in [a,b] \\ 0 & \text{otherwise} \end{cases} \tag{3.9}$$

denotes the indicator function for the closed interval $[a,b]$. Note that we did an approximation where we replace the true extinction by the expected extinction. As a consequence the lines of sight are smeared out by the parallax uncertainty in our approximation. This smoothing can be regarded as a first order correction for the uncertainty of the parallax and was already used by [92]. A fully Bayesian analysis would treat the true parallax as unknown and infer these along the other unknowns, but this is beyond the scope of this work.

For the algorithm, the integral in Eq. (3.7) is discretized into a weighted sum, such that each voxel contributes to the line integral over L_i exactly equal to the length of the line segment of L_i within that voxel while being discounted by the probability $P(l|\widetilde{\omega}, \sigma_\omega)$ of that voxel being on the line of sight. Applying the response R thus takes $\mathscr{O}(N_{\mathrm{data}}N_{\mathrm{side}})$

operations, where $N_{\text{data}} = 3\,661\,286$ is the number of data points used in the reconstruction and $N_{\text{side}} = 256$ is the number of voxels per axis. Due to the large number of data points, evaluating the response on a computer turns out to be numerically expensive. The inference algorithm (see Sec. 3.4.1) is a minimization for which this response has to be evaluated many times. To make the dust inference feasible in a reasonable amount of time we restricted our reconstruction to a 600pc cube centered on the Sun.

The second part of the likelihood states how the published data $\widetilde{A_G}$ is distributed given the true extinctions $(A_G)_i$. We use the data likelihood recommended by the Gaia collaboration in [4]. This likelihood assumes the data $\widetilde{A_G}$ to be distributed according to a truncated Gaussian with a global variance $N = (0.46\,\text{mag})^2\,\mathbb{1}$. This leads to the likelihood

$$P(\widetilde{A_G}|\rho) = \prod_i \frac{\mathscr{G}(\widetilde{A_G}_i - R(\rho)_i, N_{ii})}{\text{cdf}_{\mathscr{G}(R(\rho)_i, N_{ii})}(A_G^{\max}) - \text{cdf}_{\mathscr{G}(R(\rho)_i, N_{ii})}(A_G^{\min})} \tag{3.10}$$

$$\text{for } d \in [A_G^{\min}, A_G^{\max}] \,. \tag{3.11}$$

Here $\text{cdf}_{\mathscr{G}(R(\rho)_i, N_{ii})}$ denotes the cumulative density function of a normal distribution with mean $R(s)_i$ and variance N_{ii}. We took the boundaries of the truncated Gaussian to be $A_G^{\min} = 0$ and $A_G^{\max} = 3.609\,\text{mag}$ as recommended in [4].

3.4.3 Prior

We assume the dust density to be a positive quantity that can vary over orders of magnitude. The dust is assumed to be spatially correlated and statistically homogeneous and isotropic. The statistical model is constructed to be as general as possible with these two properties in mind.

To reflect the positivity and to allow variations of the dust density by orders of magnitude we assume the dust density ρ to be a-priori log-normal distributed with

$$\alpha\rho = \rho_0 \exp(s) \,, \tag{3.12}$$

$$\text{where } s \curvearrowleft \mathscr{G}(s, S) \tag{3.13}$$

is assumed to be Gaussian distributed with Gaussian process kernel S. Here $\rho_0 = 1/1000\,\text{pc}$ is a constant introduced to give ρ the correct unit and to bring it to roughly the right order of magnitude. By using an exponentiated Gaussian process we allow the dust density to vary by orders of magnitude while simultaneously ensuring that it is a positive quantity. In Eq. (3.13) S is the prior covariance. If we assume no point or direction to be special a-priori, then according to the Wiener-Khinchin theorem S can be fully characterized by its spatial power spectrum $S_{kk'} = 2\pi\delta(k - k')P_s(k)$. We non-parametrically infer this power spectrum $P_s(k)$ as well. There are two main motivations to reconstruct the power spectrum. From a physical perspective the power spectrum provides valuable insights into the underlying processes. From a signal processing point of view, many linear filters can be identified with a Bayesian filter that assumes a certain prior power spectrum. The optimal linear filter is obtained when the power spectrum used for the filter is exactly

$$\boxed{S_{kk}} \longrightarrow \boxed{\log(\rho)|S_{kk}} \longrightarrow \boxed{A_G|\rho} \longrightarrow \boxed{\widetilde{A_G}|A_G}$$
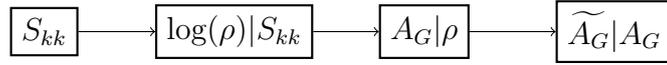
Figure 3.2: Graphical representation of the data model for our reconstruction. The logarithmic dust density $\log(\rho)$ is a Gaussian process with a smooth Gaussian process power spectrum $S_{kk'} = 2\pi\delta(k - k')P_s(k)$. The true extinctions A_G are directly dependent on the dust density ρ on each line of sight. The measured extinctions $\widetilde{A_G}$ are assumed to be distributed around the true extinctions A_G following a truncated Gaussian distribution as described in section 3.4.2.

equal to the power spectrum of the unknown quantity [30]. However, the power spectrum of the unknown quantity is usually also unknown, thus one has to reconstruct it as well. While this argumentation holds for linear filters, certainly many aspects of it carry over to nonlinear filters such as the reconstruction performed in this paper.

Fig. 3.2 depicts the hierarchical Bayesian model for the extinction data $\widetilde{A_G}$ resulting from the logarithmic dust density $\ln(\rho)$, which itself is shaped by the power spectrum $P_s(k)$.

Our statistical model for the power spectrum $P_s(k)$ is a falling power law with Gaussian distributed slope and offset modified by differentiable non-parametric deviations. It is up to minor details[2] an integrated Wiener process [25] on log-log-scale.

This is realized by the following formula:

$$\sqrt{P_s(k)} =$$
$$\exp\left((\phi_m\sigma_m + \bar{m})\log(k) + \phi_y\sigma_y + \bar{y} + \mathbb{F}^{-1}_{\log(k)t}\left(\frac{a}{1 + t^2/t_0^2}\tau(t)\right) \right) \quad (3.14)$$

Here ϕ_m, ϕ_y and $\tau(t)$ are the parameters to be reconstructed, $\sigma_m = 1, \bar{m} = -4, \sigma_y = 2., \bar{y} = -16, \sigma_y = 3., a = 11, t_0 = 0.2$ are fixed hyperparameters, $\mathbb{F}^{-1}_{\log(k)t}$ denotes the inverse Fourier transform on log-scale, and $V = (600\,\mathrm{pc})^3$ is the total volume of the reconstruction. These hyperparameters settings were determined by trial and error such that data measured from a prior sample has roughly the same order of magnitude as the actual data and such that the dust density varies by more than one order of magnitude in prior samples.

In our reconstruction the parameter τ for the smooth deviations of the log-log power spectrum was discretized using 128 pixels. The mathematical motivation to take Eq. (3.14) as a generative prior for power spectra is discussed in [8]. As a rule of thumb, k-modes for which the data constrains the power spectrum very well will be recovered in great detail due to the non-parametric nature of the model. For k-modes on which the data provide little information, the power spectrum will be complemented by the prior which forces it into a falling power law whenever the data is not informative. If the actual physical process deviates strongly from a falling power law for the unobserved k-modes, the prior might artificially suppress or amplify the posterior uncertainty of the result, possibly biasing

---

[2]The amplitude model given by Eq. 3.14 is not exactly equivalent to an integrated Wiener process, but shown by [8] to be equivalent to it in a certain limit while still allowing a numerically stable transformation of the prior to a white Gaussian.
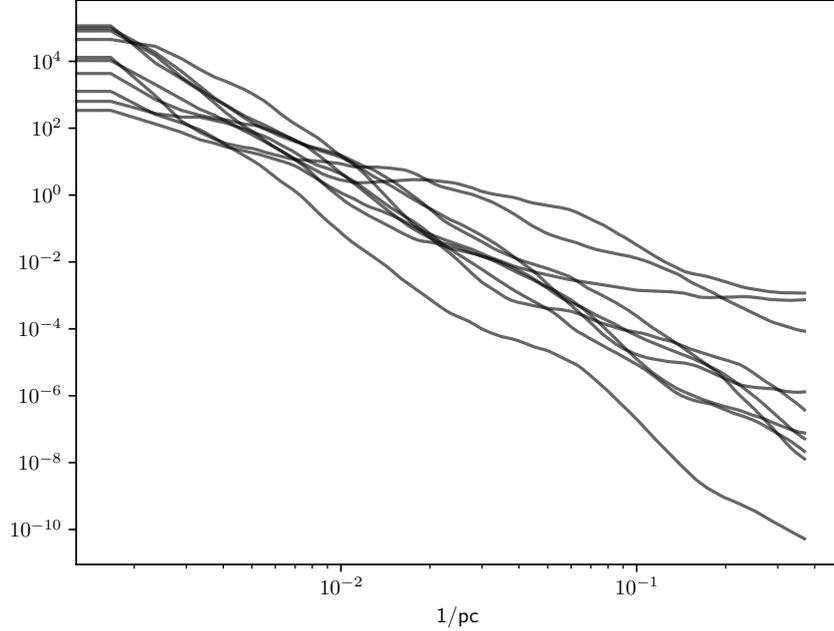
Figure 3.3: Several prior samples of the logarithmic spatial correlation power spectrum in units of pc^3.

the uncertainty quantification. Fig. 3.3 shows a few examples of prior samples of power spectra using our choice of hyperparameters. While the individual samples might not look too different qualitatively, it should be noted on the one hand that any kind of power spectrum is representable with our model given enough data and on the other hand that the figure depicts the power spectrum of the log-density on log-log scale. A small deviation in this figure can have a huge impact on the actual statistics. Reconstructing the power spectrum is equivalent to reconstructing the correlation kernel. We show our reconstructed normalized kernel as well as the one assumed by [58] in Fig. 3.4. Certain biases can appear when using a fixed kernel, for example introducing a characteristic length scale of the order of the FWHM of the kernel.

Putting together likelihood and prior, the overall joint information Hamiltonian for our parameters ξ, τ, and ϕ is

$$P(d, \xi, \tau, \phi) = \mathrm{TG}_{A_G^{\min},\ A_G^{\min},\ 0.46,\ d}(R(\alpha\rho))$$
$$\mathscr{G}((\xi, \phi, \tau)^T, \mathbb{1}) \tag{3.15}$$

$$\text{where } [R(\alpha\rho)]_i = \int_{L_i} \mathrm{d}l\ \alpha\,\rho(l) \tag{3.16}$$

$$\alpha\rho = \frac{1}{1000}\exp\left(\mathbb{F}_{xk}^{-1}\sqrt{P_s(k)}V(\phi, \tau)\,\xi_k\right) \tag{3.17}$$
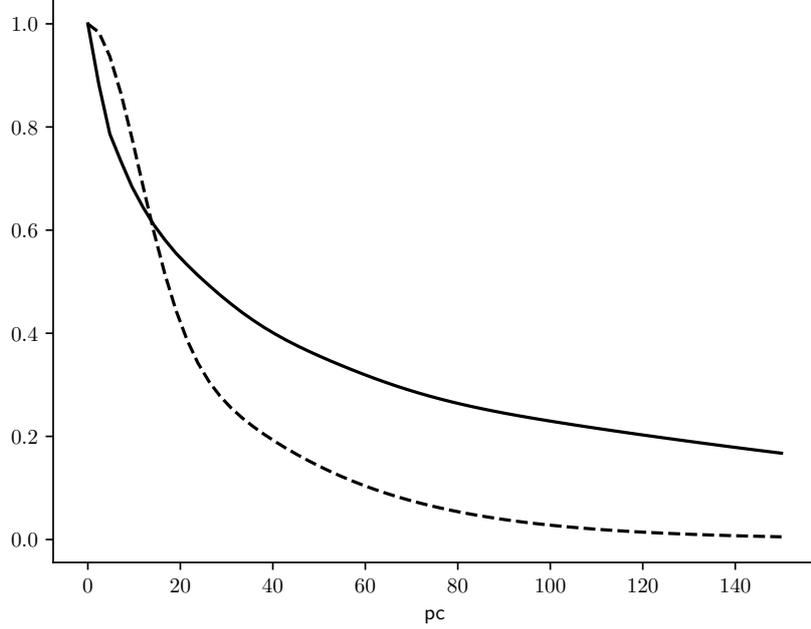
Figure 3.4: The log-normal process normalized 2-point correlation reconstructed by our method (solid line) and imposed in the reconstruction by [58] (dashed line). One can see that the dust is assumed to be strongly correlated at a distance scale of up to about 30pc. This plot shows normalized one dimensional cuts through the three dimensional Fourier transform of the log-normal spatial correlation power spectrum shown in Fig. 3.7. []

$$\sqrt{P_s(k)}(\phi, \tau) =$$
$$\exp\left((\phi_m \sigma_m + \bar{m})\log(k) + \phi_y \sigma_y + \bar{y} + \mathbb{F}^{-1}_{\log(k)t}\left(\frac{a}{1 + t^2/t_0^2}\tau_t\right)\right) \qquad (3.18)$$

$$\mathrm{TG}_{x_{\min},\ x_{\max},\ \sigma,\ \bar{x}}(x) = \prod_i \frac{\mathscr{G}(\bar{x}_i - x_i, \sigma^2)}{\mathrm{cdf}_{\mathscr{G}(x_i, \sigma^2)}(x_{\max}) - \mathrm{cdf}_{\mathscr{G}(x_i, \sigma^2)}(x_{\min})} \qquad (3.19)$$
$$\text{for } x \in [x_{\min}, x_{\max}],$$

is a truncated Gaussian, and $V = (600\,\mathrm{pc})^3$. The application, calculation of the gradient, and the application of the Fisher metric of Eq. (3.15) scales almost linearly with the number of voxels N_{side}^3, more specifically it takes $\mathscr{O}(N_{\mathrm{data}}N_{\mathrm{side}} + N_{\mathrm{side}}^3\log N_{\mathrm{side}})$ operations to evaluate Eq. 3.15, where $N_{\mathrm{data}} = 3\,661\,286$ is the number of data points used in the reconstruction and $N_{\mathrm{side}} = 256$ is the number of voxels per axis.

3.5 Simulated Data Test

3.5.1 Data generation

In this section, a test on simulated data is presented. This test enables comparing the results of the reconstruction to a known ground truth. As ground truth dust density public data from the SILCC collaboration [93] was used, more specifically from the magneto-hydrodynamic simulation of the interstellar medium *B6-1*pc at 50 Myr published by [41]. This simulation result spans a cube with size $(512\,\text{pc})^3$. We computed our synthetic ground truth differential absorption ρ_{mock} from the gas density of the simulation ρ_{sim} via

$$\rho_{\text{mock}}\left(x - (150, 150, 0)^T\right) = \sqrt[3]{\rho_{\text{sim}}\left(\frac{512x}{600}\right)10^{17}\frac{\text{cm}^3}{\text{g}}\frac{1}{\text{pc}}} \ . \tag{3.20}$$

Thus we stretch the 512 pc simulated cube to the 600 pc of our reconstruction, scale it with a constant factor, and shift it by 150 pc. The shift is performed in order to have an underdense region at the center. We also take the third root of the gas density in Eq.(3.20). There are two reasons for this.

A practical motivation for taking the third root is that it reduces the dynamic range. If one does not do this, the sky will be dominated by one very small, but very strongly absorbing blob.

A more physical motivation is that very dense regions lead to star formation. These forming stars again reduce the density by blowing the material out of these regions. This feedback mechanism was not included into the simulation by [41] but was shown to have a strong impact on the gas density in a followup simulation by [46]. The third root can be seen as a very crude way of reducing the density in these overdense regions.

To obtain the synthetic data from the ground truth differential extinction cube ρ_{mock}, the following operations were performed:

1. Sampling ground truth parallaxes $\omega_i \curvearrowleft \mathscr{G}(\omega_i - \widetilde{\omega}_i, \sigma_i^2)$ according to the parallax likelihood published by the Gaia collaboration.

2. Integrating the dust density from the center of the cube to the location of the sampled star location $^1/_\omega$ using the full resolution of 512^3 voxels[3].

3. Sampling an observed extinction magnitude according to the truncated Gaussian likelihood described in section 3.4.2.

3.5.2 Results

We were able to recover a slightly smeared out version of the original synthetic extinction cube. In Fig. 3.5 integrations with respect to the x-, y-, and z-axis of the synthetic ex-

---

[3]Note that the simulation of which the data is used was performed on an adaptive grid. The full resolution of 512^3 is only realized in the high density regions.

tinction cube and the reconstructed extinction cube are shown. This visually confirms the reliability of the reconstruction.

For a more quantitative analysis, we compared the reconstructed differential extinctions with the ground truth voxel-wise. More specifically, we computed the uncertainty weighted residual r

$$r = \frac{\rho_{\text{reconstruction}} - \rho_{\text{ground truth}}}{\sigma_\rho}, \qquad (3.21)$$

where $\rho_{\text{reconstruction}}$ and σ_ρ are the posterior mean and standard deviation computed from the approximate posterior samples. The ground truth differential extinction was recovered well within the recovered approximate posterior uncertainty, apart from outliers which make up about 0.15% of the voxels. See Fig. 3.6 for a histogram of the uncertainty weighted residual.

Overall, the reconstruction seems very reliable on a qualitative and quantitatively level within the uncertainty for most of the voxels.

3.6 Results from Gaia Data

We reconstruct the dust density in a 600pc cube using 256^3 voxels, resulting in a resolution of $(2.34\,\text{pc})^3$ per voxel. For our reconstructed volume we also infer the spatial correlation power spectrum of the log-density, see Fig. 3.7.

In Fig. 3.8b one can see a projection of the reconstructed dust onto the sky in galactic coordinates. Fig. 3.9b shows the corresponding expected logarithmic dust density.

In Fig. 3.8a the projection of the dust reconstruction on the galactic plane is shown. This view is especially interesting to study the dust morphology as this projection introduces no perspective-based distortion. It is especially suited to spot underdense regions such as the local bubble in high resolution. A logarithmic plot of the projection on the galactic plane can be seen in Fig. 3.9a. We show integrated dust density for sightlines parallel to the x-, y-, and z-axis in Fig. 3.13.

We provide posterior uncertainty estimation via samples. One should note that these uncertainties might be underestimated due to the variational approach taken in this paper. One can see a map of the expected posterior variance of the sky projection in Fig. 3.10a and in the plane projection in Fig. 3.11a.

3.6.1 Using the reconstruction

The results of the reconstruction are provided online on `https://wwwmpa.mpa-garching.mpg.de/~ensslin/research/data/dust.html`, or by its doi:10.5281/zenodo.2577337, and can be used under the terms of the ODC-By 1.0 license. Proper attribution should be given to this paper as well as to the Gaia collaboration [39].

We give an overview of known systematic effects and advice on how to use the provided dust map.
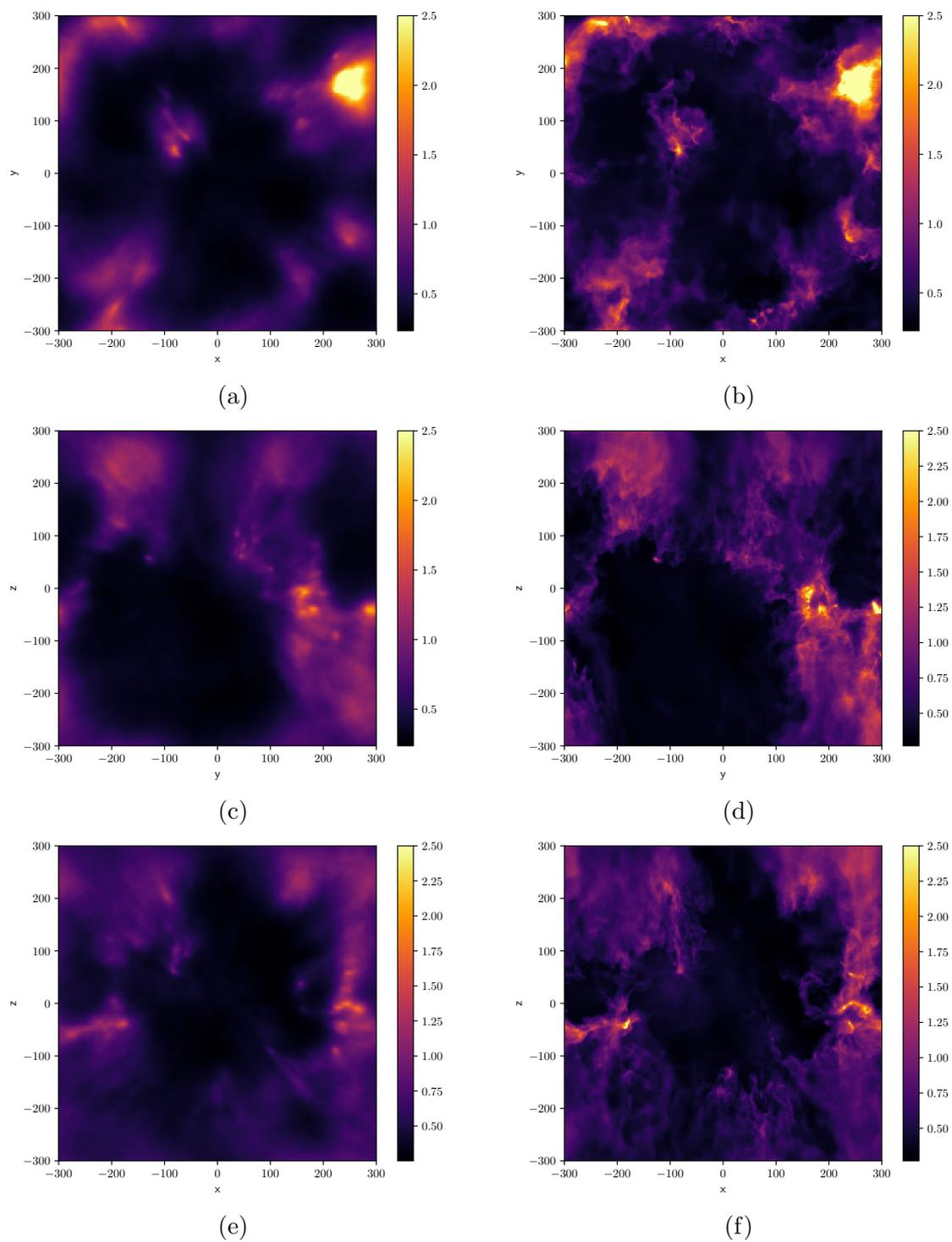
Figure 3.5: Results of our test using simulated data. The rows show integrated dust extinction for sightlines parallel to the z- x- and y- axis respectively. The first column corresponds to the test reconstruction, the second column is the ground truth synthetic extinction.
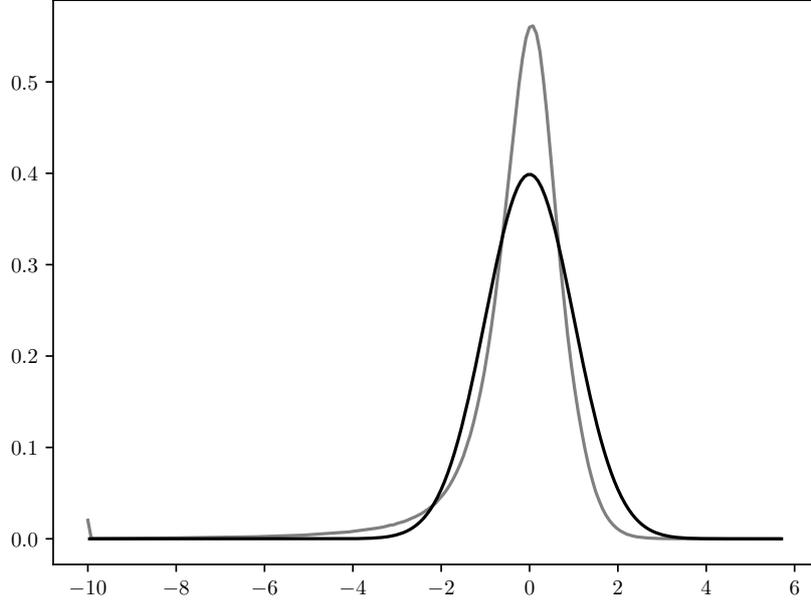
Figure 3.6: The gray curve shows a normalized histogram of the deviation of the reconstruction from the true solution, in sigmas. The black curve is the probability density function of a standard normal distribution, which is plotted as a reference. Note that the values were clipped to the range from -10 to 10, i.e. the bump in the gray curve at -10 corresponds to outliers that can be up to 250 sigmas. These outliers correspond to about 0.15% of all voxels.

- We do not recommend to use the outer 15pc of the reconstruction. Periodic boundary conditions were assumed for algorithmic reasons, which leads to correlations leaking around the border of the cube. The inferred prior correlation kernel (Fig. 3.4) suggests that correlations are vanishing after 30pc.

- We provide posterior samples. When doing further analysis of our reconstruction we recommend doing so for every sample in order to propagate errors.

- It was observed in Sec.3.5.2 that there can be a small number of outliers, that is differential dust extinction values that are much larger than the reconstructed value, by amounts that cannot be explained by the reconstructed uncertainty.

- We anticipate a perception threshold that leads to the absence of extremely low density dust clouds. The two main reasons for this are that the truncated Gaussian likelihood provides less evidence in the regime where the extinction is close to zero and that variational Bayesian schemes are known to underestimate errors. Studying a larger volume will shed further light on this subject, as sightlines for more distant
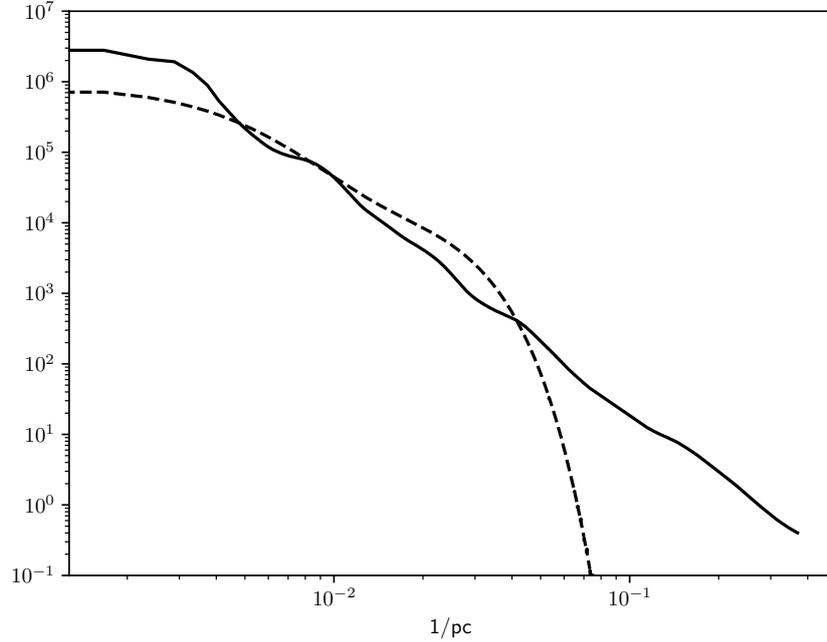
Figure 3.7: The log-normal process spatial correlation power spectrum inferred in our re-
construction (solid line) as well as the imposed power spectrum of [58] (dashed
line). The shaded area around the solid line indicates 1σ error bounds. The
unit of the y-axis is pc^3. The functions can be interpreted as the a-priori ex-
pected value of $|\mathbb{F}\ln(\rho)|^2/V$, where V is the volume the density ρ is defined on
and $\mathbb{F}$ is the Fourier transform. The region between $0.0008/pc$ and $0.426/pc$ is
almost power-law like with a slope of 3.1, the spectral index of the power law.

stars still provide information about nearby dust clouds. One should note that the
overall Gaia extinction data provides 20 times more sightlines than were used in this
reconstruction.

3.7 Discussion

Here we discuss qualitative, quantitative, and methodological differences to other dust
mapping efforts. In table 3.1 a detailed break down of methodological differences to other
papers are shown. There are three notable differences of our method to other methods
that we would like to stress.

1. The here used dataset is one of largest one used so far.

2. We use a high amount of data while still taking 3D correlations into account.
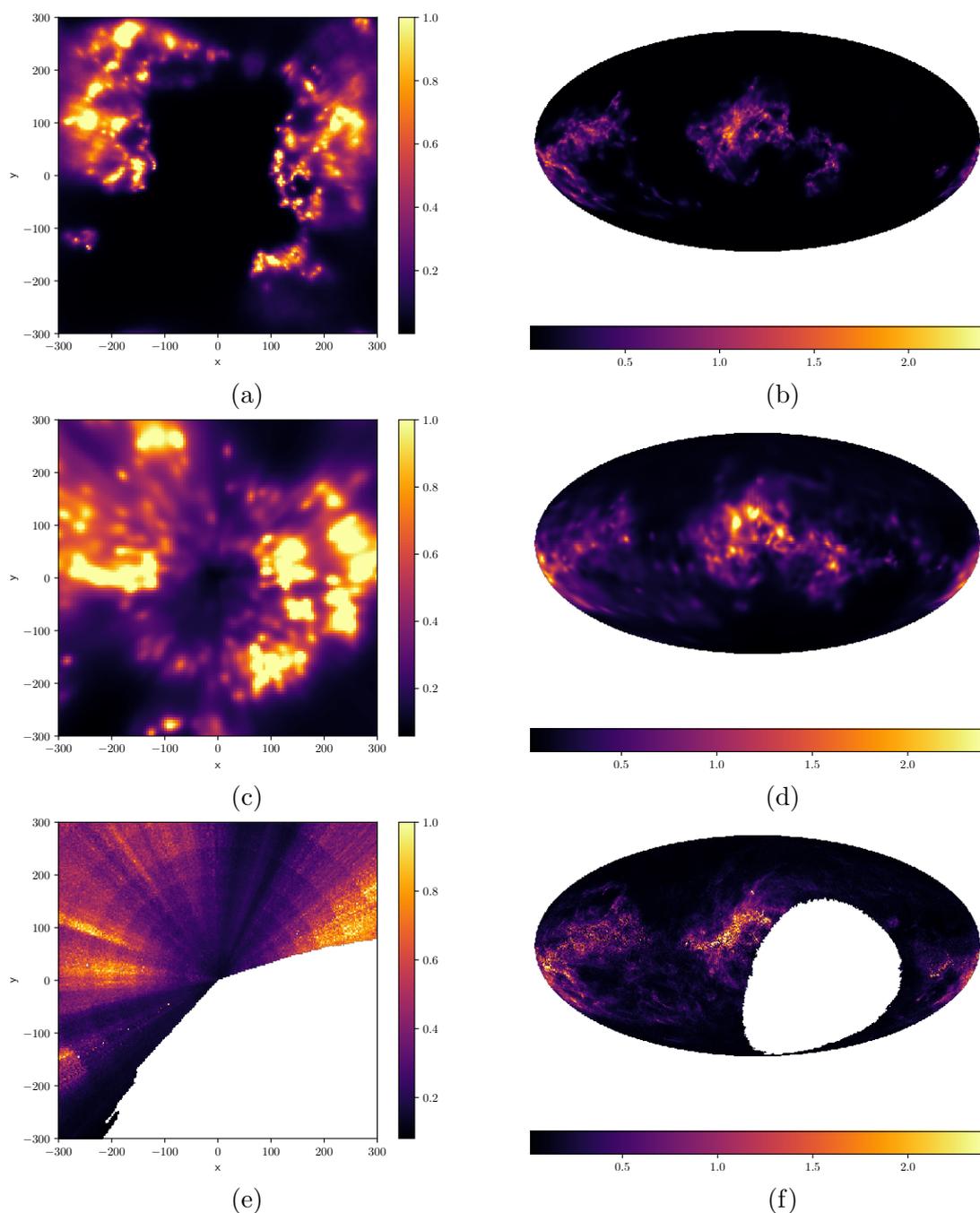
Figure 3.8: The left column shows integrated dust extinction from $-300\,\mathrm{pc}$ to $300\,\mathrm{pc}$ for sightlines perpendicular to the galactic plane. The image covers a 600pc cube centered around the Sun. The units are e-folds of extinction. The coordinates are galactice cartesian coordinates. The Sun is at coordinate $(0,0)$, the galactic center is located towards the left of the plot, and the galactic West is located towards the top. The right column shows all-sky integrated dust extinction maps of the same region, but for sightlines towards the location of the Sun. The first row is the result of the reconstruction discussed in this paper, the second row is the reconstruction performed by [58], the last row shows the reconstruction by [44].
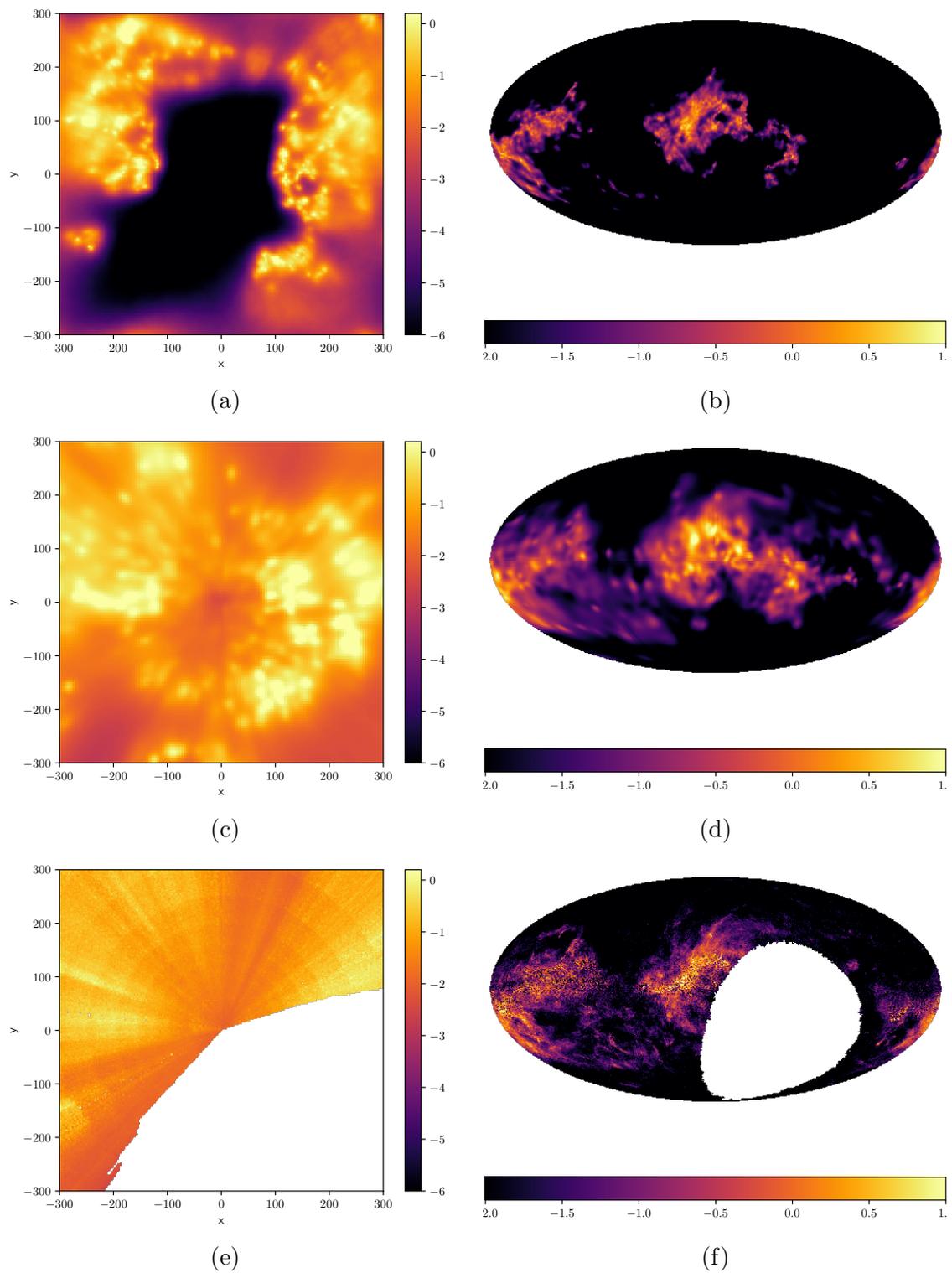
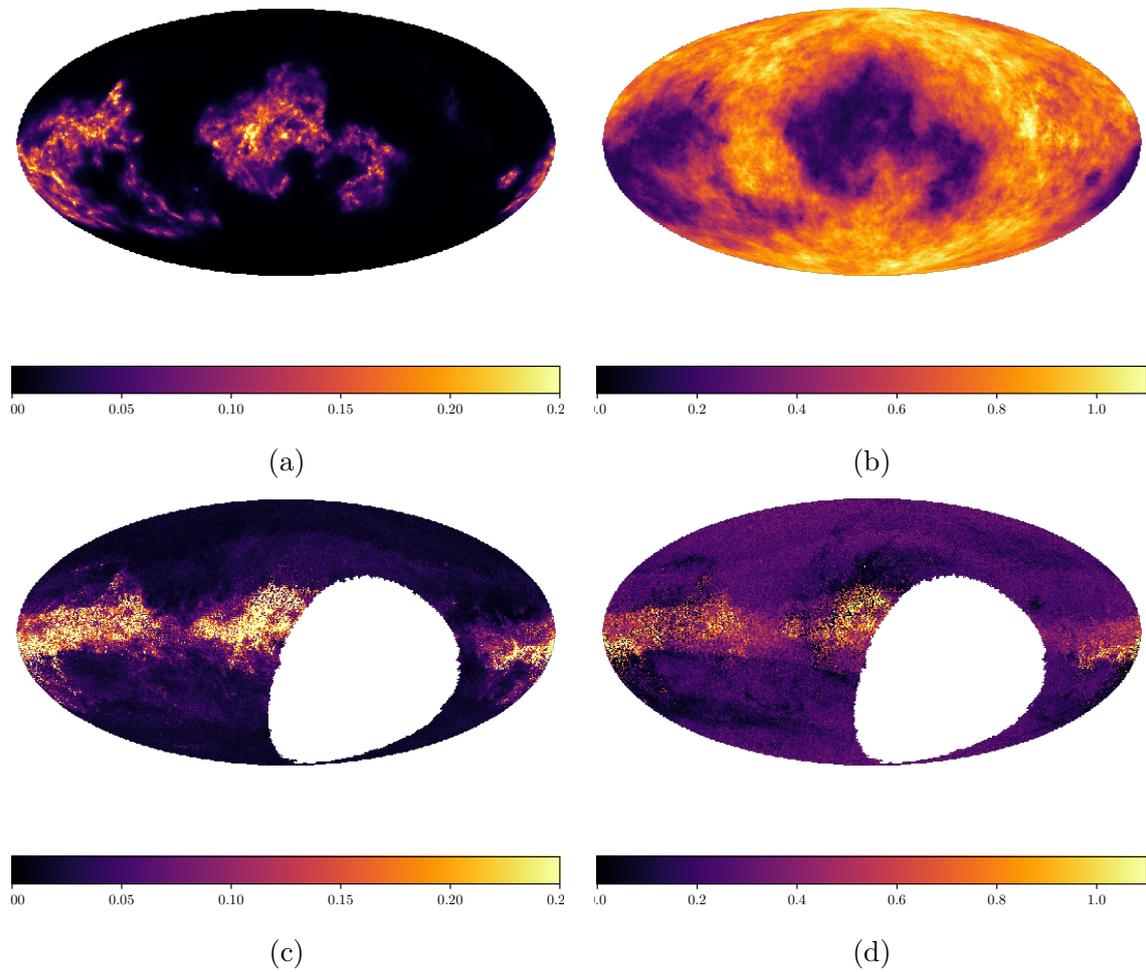Figure 3.9: A natural logarithmic version of Fig. 3.8.

Figure 3.10: Uncertainty of the reconstruction of this paper derived from posterior samples (first row) and of the reconstruction of [44] (second row), both in the sky projection. The uncertainties are in the same unit as the corresponding maps in Fig. 3.8, or dimensionless for logarithmic uncertainties. The first column shows the variance for the dust extinction and the second column shows the variance of the logarithmic projected dust density on natural log-scale, which can be interpreted as a relative error.
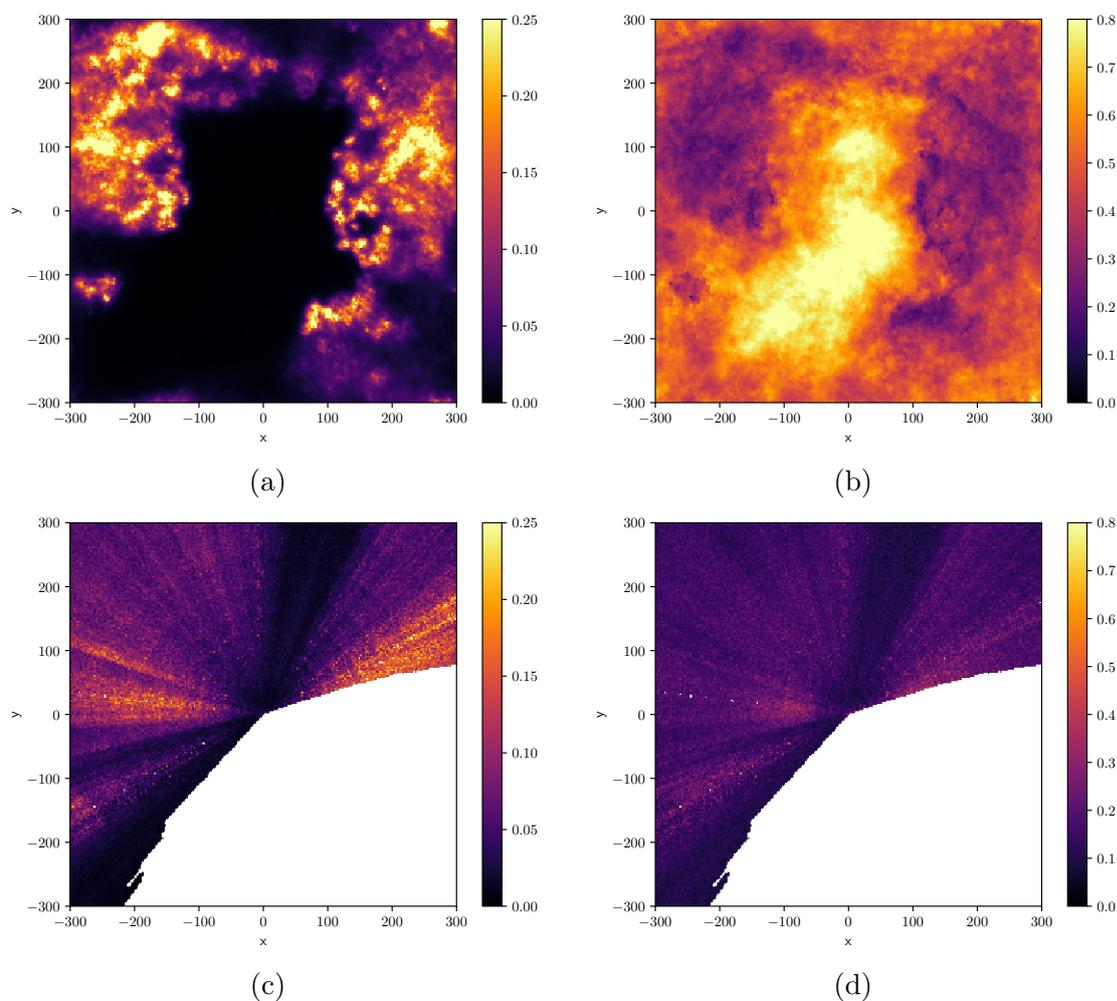
Figure 3.11: Posterior uncertainty of the reconstruction of this paper derived from samples (first row) and of the reconstruction of [44] (second row) in the plane projection. The uncertainties are in the same unit as the corresponding maps in Fig. 3.8, or dimensionless for logarithmic uncertainties. The first column shows the variance for the dust extinction and the second column shows the variance of the logarithmic projected dust density on natural log-scale which can be interpreted as a relative error.

3. We reconstruct the spatial correlation power spectrum. The motivation and impact of this already briefly discussed in Sec. 3.4.3

	this paper	[81]	[78]	[58]	[44]
parallax uncertainty	smoothing only	marginalization by sampling	neglected	neglected	proper uncertainty handling
max distance	$300\sqrt{3}\,\mathrm{pc}$	$5\,\mathrm{kpc}$	$6\,\mathrm{kpc}$	$\approx 2\sqrt{2}\,\mathrm{kpc}$	$3\,\mathrm{kpc}$
max voxel resolution	$2.3\,\mathrm{pc}$	not applicable	about $200\,\mathrm{pc}$	$5\,\mathrm{pc}$	$16.4\,\mathrm{pc}/0.063\,\mathrm{pc}$
number of datapoints	3.7 million	6 349	21 000	71 357	806 million
power spectrum inference	yes	no	no	no	no
correlations	3D	3D	2D map only	3D	1D correlations only
positiveness	yes	only of reddening	no	yes	yes
statistical method	Variational Bayes	Expectation Propagation	analytic	maximum posterior	Hamiltonian Monte Carlo
data sets	Gaia DR2	synthetic Gaia data	APOGEE	Gaia DR1 + APOGEE + 2MASS	Pan-STARRS + 2MASS

Table 3.1: A table comparing different dust inference methods with the one performed in this paper. The first row indicates how the parallax uncertainty of the stars was treated. Hereby smoothing refers to weighting a voxel in the line of sight by the survival function of the star radial distance, as is described in Eq. (3.7). The distance of the furthest point in the reconstruction is given in the second row. The dimensions of the smallest voxel are given in the third row. For the reconstruction of [81] the concept of voxel resolution is not readily applicable; [81] use 140 inducing points spanning a region for which one could evaluate the posterior mean at any point. The resolution for [44] contains two values because the resolution is different in radial/angular direction. The fourth row provides the number of used data points. The fifth row indicates whether the power spectrum is inferred. The sixth row states which kind of correlations are assumed for the reconstruction. Whether positivity of dust density is enforced can be read in the seventh row. The second to last row states the method, with which the posterior summary statistics was calculated from the unnormalized log posterior. In the last row the data sets used for the reconstruction are listed.

We compare our dust map to other maps. Comparisons to 2D dust maps are only possible on a qualitative level, since it is not clear whether structures visible in the 2D maps that are not present in the 3D map are simply further away or are too noisy in the data for the algorithm to pick them up. On a qualitative level it is possible to see several morphological similarities of our reconstruction in Fig. 3.8b to the Planck dust map [89] in Fig. 3.1. These figures also show that many dust structures that are not inside the galactic plane are local features.

The two 3D dust maps mentioned in Sec. 3.2 permit a more thorough analysis. Fig. 3.8 shows a compilation of projected dust densities for our reconstruction as well as the reconstruction of [58] and [44][4], restricted to the same volume as the reconstruction discussed in this paper. A logarithmic version of this figure is provided by Fig. 3.9.

While our map seems to agree on large scales with the other maps, there seems to be a pronounced tension in the predictions of the position of some dust clouds compared to the reconstruction of [58]. Compared to the map of [44] we recover the small scales significantly better and suffer far less from radial smearing. It should be noted that [44] mapped a significantly larger part of our galaxy, and that the region that overlaps with our map was declared to be not that reliable by the authors themselves. The differences are probably due to the different nature of the used datasets. The Gaia DR2 data used in our reconstruction has a vastly higher amount of data points than those used for the other reconstructions. These data points, taken from Gaia DR2, have a very small parallax error. Additionally our reconstruction takes the full 3D correlation structure into account.

Our reconstruction as well as the reconstruction of [44] permit quantifying uncertainties using samples. A plot of uncertainties of the dust density reconstructions projected into the galactic plane can be seen in Fig. 3.11. Uncertainties of the dust density reconstructions in the sky projection can be seen in Fig. 3.10.

To quantify the dynamic range of the reconstruction and as a prediction on the variability of the logarithmic dust density we calculated histograms of dust density which show how many voxels have which dust density. These histograms can be seen in Fig. 3.12a. One can see that the histogram of our reconstruction extends slightly more towards high dust densities and substantially towards low dust densities. This is possibly because our reconstruction is more sharply resolved, thus regions of high dust density get captured better and bleed less into the regions where dust is absent.

We characterize how much pairs of those reconstructions agree by the heatmaps of their voxel-wise value pairs. These heatmaps can be seen in Fig. 3.12. For two perfectly agreeing reconstructions the heatmap would show a line with slope 1. Again it can be seen that the dust density in our reconstruction varies significantly more than in the two other reconstruction. While all maps agree more or less for high dust densities, our dust map exhibits vastly more volume with low dust density.

The reconstruction of [58] is performed using Gaussian process regression, as is ours. Thus one can compute the prior Gaussian process correlation power spectrum used in their reconstruction. Fig. 3.7 shows both our inferred power spectrum as well as their

---

[4]It should be noted that there is a new version [45] that appeared during the revision of this paper.
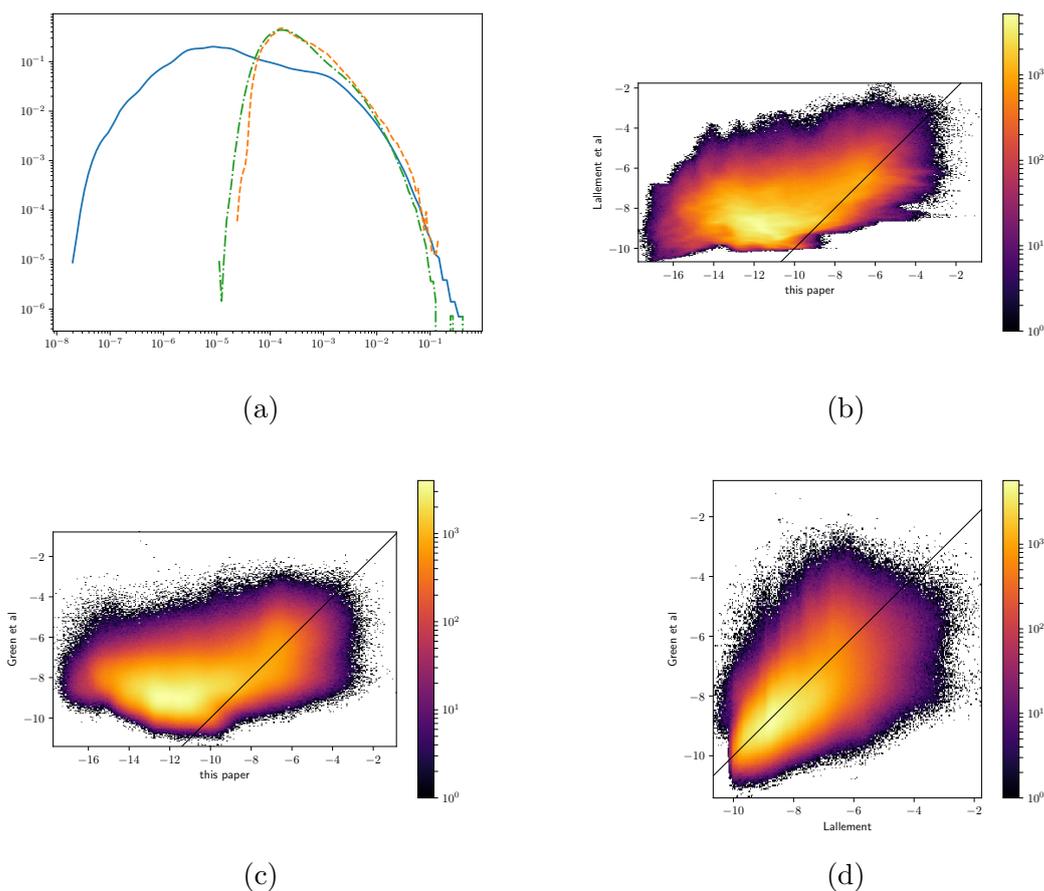
(a)　　　　　　　　　　　　　　　　　　(b)

(c)　　　　　　　　　　　　　　　　　　(d)

Figure 3.12: Panel a shows normalized histograms of dust densities. The solid line corresponds to our reconstruction, the dashed line is the reconstruction of [58] and the dash-dotted line is the reconstruction of [44]. The other three plots are heatmaps of voxel-wise correlations between reconstructed logarithmic dust densities, where the color shows bin counts. The black line in the heatmaps is the identity function, corresponding to perfect correlation.
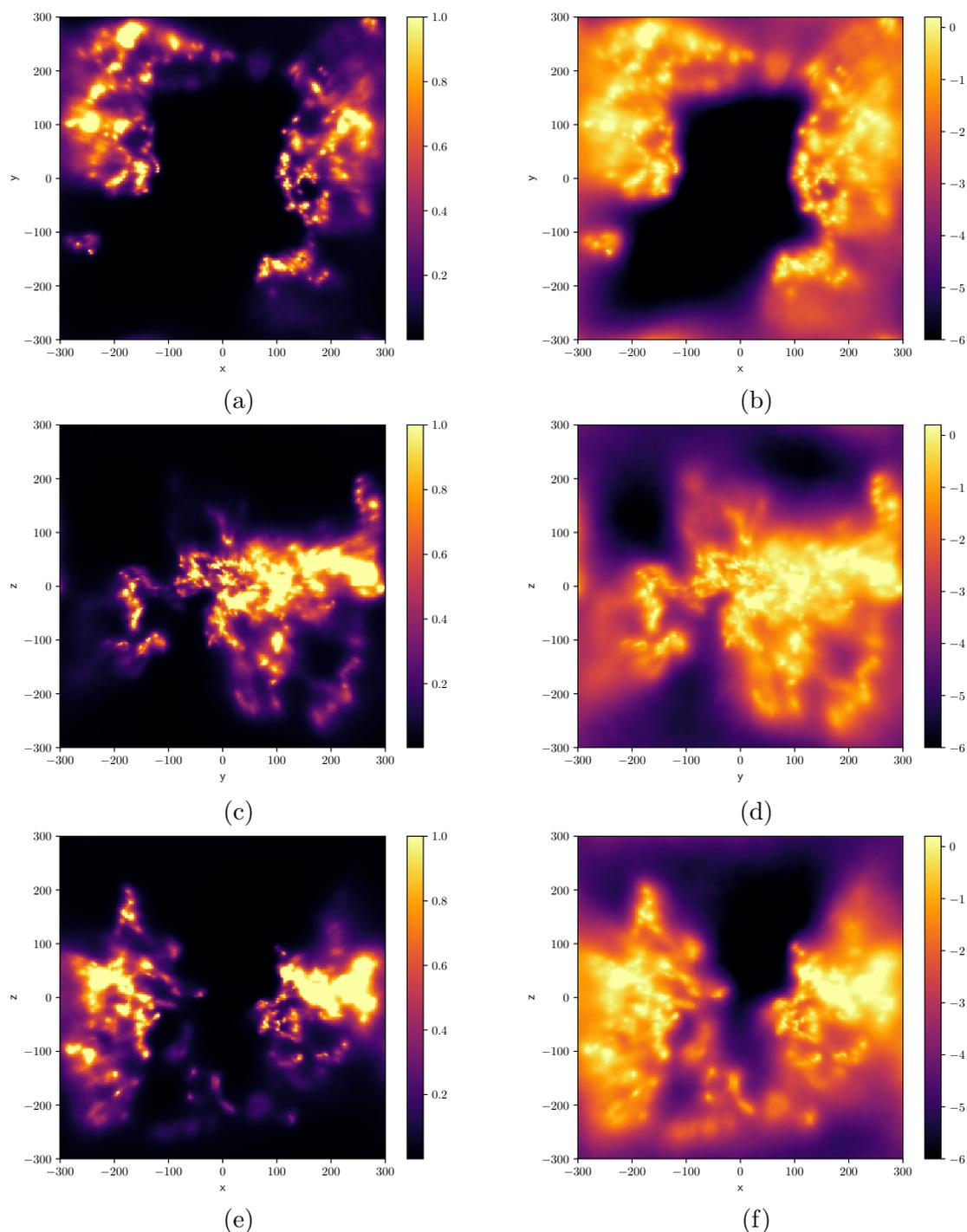
Figure 3.13: The reconstructed dust density in different projections. The rows show integrated dust extinction for sightlines parallel to the z- x- and y- axis respectively. In the first row, the galactic center is located towards the left of the plot, in the other two rows the galactic North is located towards the top of the plot. The cube is in galactic coordinates, thus the x-axis is oriented towards the galactic center and the z-axis is perpendicular to the galactic plane. The first column shows the integrated G-band extinction in e-folds of extinction, the second column is a logarithmic version of the first column.
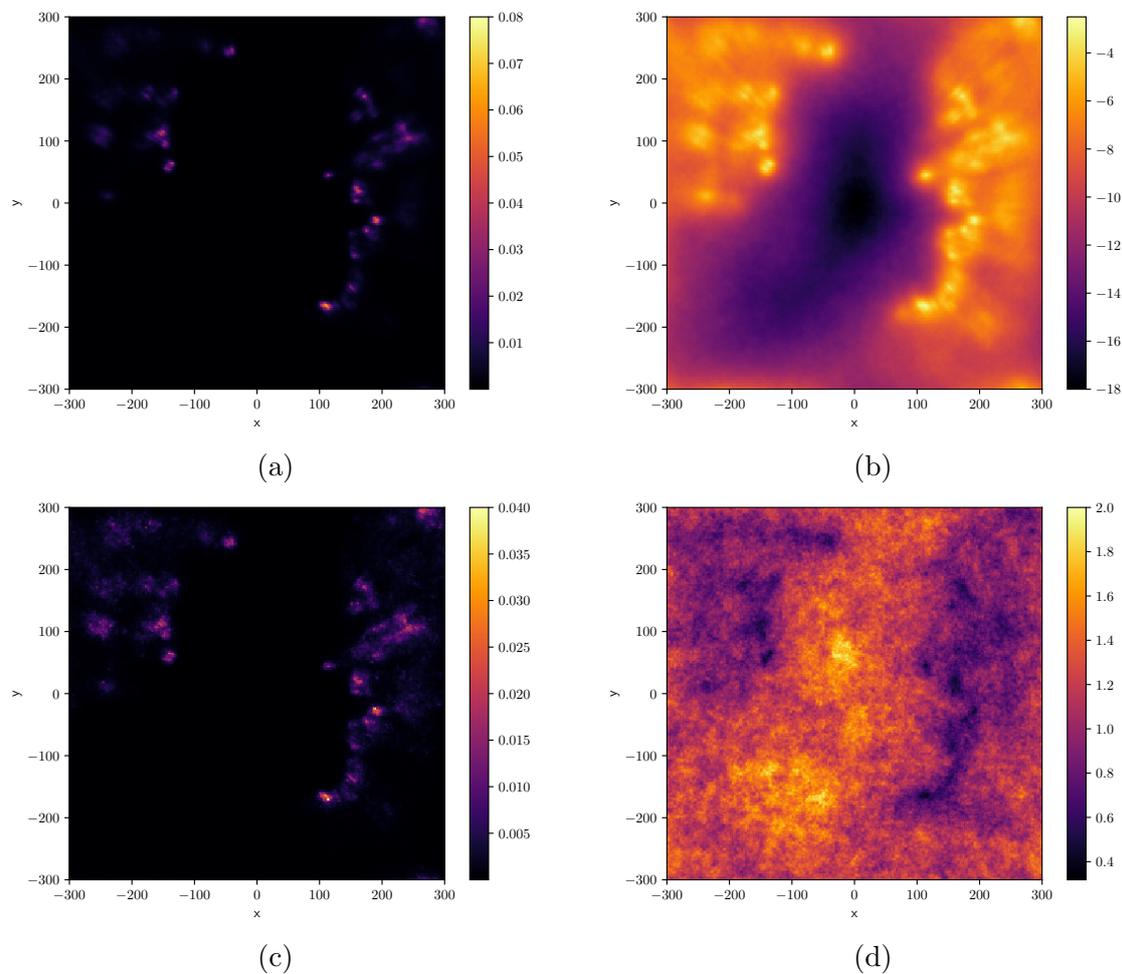
(a)

(b)

(c)

(d)

Figure 3.14: The reconstructed dust pseudo-density in a slice of the galactic plane. The first plot shows differential dust extinction in the plane containing the Sun. The second plot of the first row is a logarithmic version of the first plot. The second row shows the corresponding uncertainty maps. The unit of the dust is G-band extinction in *e*-folds per parsec. The coordinates are galactic cartesian coordinates.
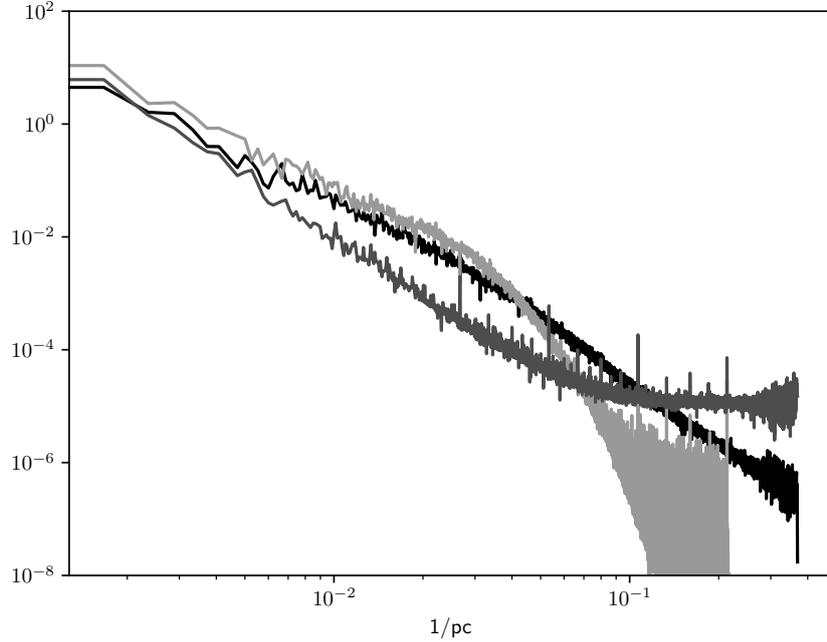
Figure 3.15: Empirical spatial correlation power spectra of the reconstructed mean dust density in units of pc. The black line was computed from our reconstruction, the dark-grey line is computed from the reconstruction of [58] and the light-grey line is computed from the reconstruction of [44]. For the reconstruction of [44] unspecified voxel values on sightlines that lacked data were replaced with 0.

assumed power spectrum. These two power spectra agree more or less for the larger modes (low k), where the data is very constraining.

One can empirically compute power spectra of the dust density using a Fourier transformation. A comparison plot with all the three mentioned reconstruction can be found in Fig. 3.15. This shows a white noise floor in the reconstruction of [44], which can visually also be seen as small scale structures in the plane projections shown in Figs. 3.8 and 3.9.

3.8 Conclusions

1. We provide a highly resolved map of the local dust density using only Gaia data. This map agrees on large scales with previously published maps of [58] and [44], but also shows significant differences on small scales. These differences might to a large degree stem from the different data used. Our map shows many structures visible in the Planck dust map [89].

2. In comparison to previous maps, we were able to improve on 3D resolution while still being mostly consistent on the large scales. A comparison to 2D maps like the Planck dust map seems to confirm the features present in our map.

3. We find that the logarithmic density of dust exhibits a power-law power spectrum with a 3D spectral index of 3.1, corresponding to a 1D index of 1.1. This is a significantly harder spectrum as that expected for a passive tracer in Kolmogorov turbulence, which would be a 1D index of 5/3. The harder spectrum is probably caused by the sharp edges of the local bubble and other ionization or dust evaporation fronts.

4. In contrast to other dust reconstructions, we predict very low dust densities inside the local bubble. This discrepancy is possibly an artifact of our reconstruction as there are known dust clouds in our vicinity, for example the northern high latitude shells [72] and the local Leo cold cloud [70]. The Leo cold cloud is however considerably smaller than a voxel of our simulation. The possibility that Gaia extinction estimates are biased for small distances can also not be excluded.

5. We hope that by providing accurate reconstructions of the nearby dust clouds, further studies of dust morphology will be possible as well as the construction of more accurate extinction models for photon observations in a large range of frequency bands.

4 Resolving nearby dust clouds

This chapter is accepted as an article in the journal Astronomy and Astrophysics [61]. My contribution to the work is the development of the method, implementing and testing the code, running the reconstruction, and writing of the text. Martin Glatzle contributed the code used for the computation of the line of sight integrals and has written the first two paragraphs of the introduction. Torsten Enßlin has contributed through many valuable discussions and by reviewing and correcting the text several times. All authors read, commented, and approved the final manuscript. You can find larger versions of sub-figures of figure 4.2 in the appendix.

4.1 Abstract

Mapping the interstellar medium in 3D provides a wealth of insights into its inner working. The Milky Way is the only galaxy, for which a detailed 3D mapping can be achieved in principle. In this paper, we reconstruct the dust density in and around the local super-bubble. The combined data from surveys such as Gaia, 2MASS, PANSTARRS, and ALLWISE provide the necessary information to make detailed maps of the interstellar medium in our surrounding. To this end, we use variational inference and Gaussian processes to model the dust extinction density, exploiting its intrinsic correlations. We reconstruct a highly resolved dust map, showing the nearest dust clouds at a distance of up to $400\,\mathrm{pc}$ with a resolution of $1\,\mathrm{pc}$. Our reconstruction provides insights into the structure of the interstellar medium. We compute summary statistics of the spectral index and the 1-point function of the logarithmic dust extinction density, which may constrain simulations of the interstellar medium that achieve similar resolution.

4.2 Introduction

Although dust contributes only a small fraction in terms of mass, it is an important constituent of the interstellar medium (ISM) that is observable in many wavebands of the electromagnetic spectrum. Dust efficiently absorbs and scatters ultra-violet and visible range photons, obscuring large parts of the Galaxy and hiding star forming regions at these wavelengths. The dust absorbed energy is re-emitted in the infrared to microwave bands, offering a diagnostic for physical conditions of the ISM. The microwave emission of dust is a significant foreground to the Cosmic Microwave Background (CMB).

Dust plays a role in many processes that drive galactic evolution. Grain surfaces can adsorb material from interstellar gas and act as catalytic sites for chemical reactions.

Stars, including the most massive ones, are observed to form from dusty molecular clouds. Thermal emission from dust grains can be an important cooling channel for these clouds and grains can drive their chemistry, suggesting that dust plays an important role in regulating the star formation process. Photons absorbed by dust can convey radiation pressure to interstellar matter or, if they are energetic enough, eject electrons, contributing to the heating of interstellar gas.

Finally, the distribution of dust can be used as a tracer of other quantities. A significant portion of the observed Galactic gamma rays in the GeV-range originates in dense clouds, where it is produced by hadronic interactions of cosmic rays with gas. This can be seen e.g. in the morphology of cosmic rays with hadronic spectrum from FERMI [84]. Dust can be used to trace these dense clouds and identify gamma ray production sites. Another example is the magnetic field structure of the Galaxy, which is imprinted in the dust density, as dust filaments tend to be aligned to the line of sight magnetic field [69]. Dust also reveals the large scale dynamics and structure of the Galaxy, as the gravitational and differential rotation imprints on the filaments of dust.

Studying how dust is distributed in the Galaxy can not only provide understanding of its contents and structure but also into its inner workings and aid in the interpretation of observations in dust affected wavebands. Most 3D mapping efforts so far have aimed at reconstructing the distribution of dust in our galaxy on large scales. This is interesting as it reveals the structure of our galaxy such as spiral arms. Some notable recent contribution in this direction was provided by Green et al. [45], who map three quarters of the sky using Gaia, 2MASS and PANSTARRS data using importance sampling on a gridded parameter space and by assuming a Gaussian process prior. Lallement et al. [59] reconstructed a map extending out to $3\,\mathrm{kpc}$ with a $25\,\mathrm{pc}$ resolution based on Gaia and 2MASS data with Gaussian process regression. Chen et al. [18] reconstructed a map extending out to $6\,\mathrm{kpc}$ with a $0.2\,\mathrm{kpc}$ radial resolution based on Gaia, 2MASS and WISE data with random forest regression.

This paper can be regarded as a follow-up to Leike and Enßlin [64]. Some derivations are kept short here, and we advise Leike and Enßlin [64] as a co-read for the statistically inclined reader. We focus on reconstructing only the nearby dust clouds, within $\sim 400\,\mathrm{pc}$. While this prohibits revealing spiral arms, it enables us to achieve higher resolution. This way, we hope to be able to constrain simulations of the ISM, which achieve similar resolution. The map might also prove relevant for foreground corrections to the CMB, especially for CMB polarization studies. It was shown that most of the Galactic infrared polarization at high latitudes ($|b| > 60$) comes from close by regions around 200-300 pc [85]. Corrections maps so far were based on infrared observations, and could be biased through different starlight illumination or differing dust temperatures.

4.3 Data

For our 3D reconstruction, we use combined observational data of Gaia DR2, ALLWISE, PANSTARRS and 2MASS. These data-sets were combined and processed to yield one

consistent catalogue with stellar parameters by Anders et al. [3]. We use these high level preprocessed data for our reconstruction. Table 4.1 contains a summary of the columns we extract from this data set. We further only select sources that are inside an $800\,\mathrm{pc} \times 800\,\mathrm{pc} \times 600\,\mathrm{pc}$ cube centered on the Sun. To determine whether a source is inside this cube, we use their 84% distance quantile dist_{84}. We assume a Gaussian error on the parallax, with mean m_ω and standard deviation computed from the distance quantiles as

$$m_\omega = \frac{1}{2}\left(1/\mathrm{dist}_{16} + 1/\mathrm{dist}_{84}\right) \tag{4.1}$$

$$\sigma_\omega = \frac{1}{2}\left(1/\mathrm{dist}_{16} - 1/\mathrm{dist}_{84}\right). \tag{4.2}$$

Furthermore, we apply the following selection criteria:

$$\mathrm{SH\_OUTFLAG} = 00000 \tag{4.3}$$
$$\mathrm{SH\_GAIAFLAG} = 000 \tag{4.4}$$
$$\mathrm{ph} \in \mathrm{Table}\ 4.2 \tag{4.5}$$
$$\sigma_\omega/m_\omega < 0.3 \tag{4.6}$$
$$\mathrm{av}_{05} \neq \mathrm{av}_{16}. \tag{4.7}$$

In words, we selected only stars, which do have clean starhorse pipeline flags, a clean Gaia flag, a specific photo-flag, and sufficiently small parallax error. We require the constraint on the photo-flag, because we only derived the noise statistic for stars with this flag. For details see Sec. 4.4.2. Additionally, we excluded stars for which the 5% V-band extinction quantile is equal to the 16% quantile, as this suggests that the pipeline had difficulties for these sources.

These criteria result in the selection of a total of 5 096 642 sources. Fig. 4.1 shows an inverse-noise weighted average of our data projected onto the sky. To consistently combine the information of many data points, it is crucial to know the likelihood of a data point given the true amount of extinction for that source. We call this likelihood of one data point given its true extinction the *noise statistic* to distinguish it from the likelihood of the whole data set given the true 3-dimensional dust extinction distribution, which contains additional operations (see Sec. 4.4 for details). Unfortunately, [3] did not publish a noise statistic for their data-set, and a noise statistic is not readily derivable from posterior quantiles. This is because posterior quantiles a give very limited information on the distribution $P(a^*|a)$ of the true extinction a^*, while a full noise statistic would be given by $P(a|a^*)$. In particular, there is no natural way to derive an analytic form of $P(a^*|a)$, inhibiting the calculation of $P(a|a^*)$. We will thus choose a different approach to infer the noise statistic which we describe in Sec 4.4.2.
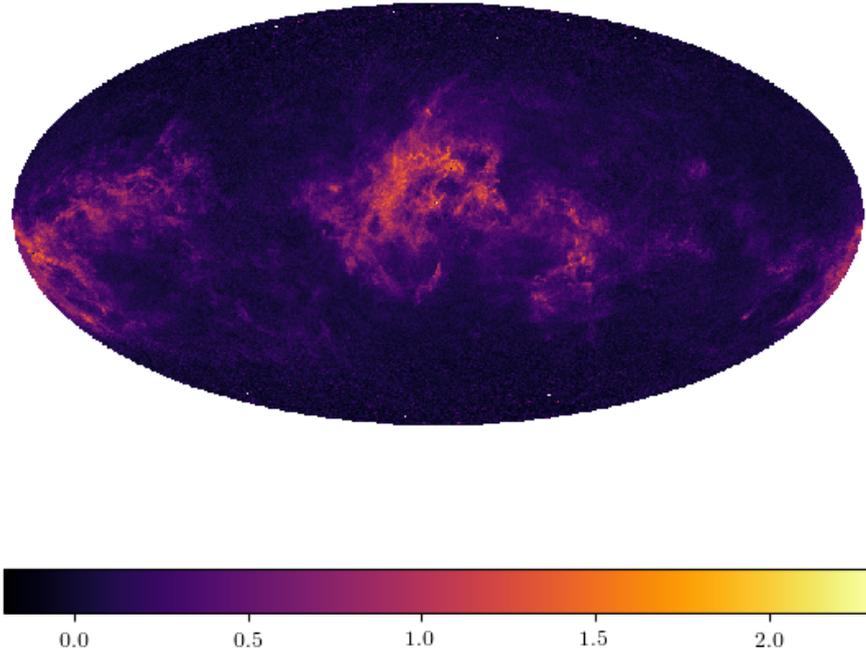
Figure 4.1: A Mollweide projection of the G-band extinction optical depth a to all sources
in the used dataset. For this healpix nside 128 plot, we average the data sources
that are in the same pixel, using the inverse noise dispersion as weights. Pixels
with no data appear in white.

name in Anders et al. [3]	our notation	explanation
dist16	dist_{16}	16% distance quantile
dist50	dist_{50}	50% distance quantile
dist84	dist_{84}	84% distance quantile
ag50	a	50% G-band extinction quantile
SH_PHOTOFLAG	ph	photo-bands used for data point
SH_GAIAFLAG	SH_GAIAFLAG	output flag of Gaia
SH_OUTFLAG	SH_OUTFLAG	output flag of the starhorse pipeline

Table 4.1: data columns extracted from Anders et al. [3]

4.4 Likelihood

4.4.1 Response

If given the true 3D extinction density $s(x)$, we can compute the extinction a_i^* for each source i by computing the line integral R_i

$$a_i^* = R_i^{\omega*}(s) = \int_0^{\frac{1}{\omega_i^*}} s(r\theta_i)\mathrm{d}r \,, \tag{4.8}$$

where θ_i is the position of the i-th source projected onto the unit sphere and ω_i^* is the true parallax of the source. The true parallax ω_i^* is assumed to be Gaussian distributed with error and mean computed from the 16% and 84% percentiles of the starhorse dataset according to Eqs. (4.1), (4.2):

$$\omega_i^* \curvearrowright \mathscr{G}(\omega_i^* | \frac{1}{2}(\omega_{84,i} + \omega_{16,i}), \frac{1}{4}(\omega_{84,i} - \omega_{16,i})^2) \,. \tag{4.9}$$

Given this uncertainty of the true source distance, we can compute the expected extinction density for the source i as a weighted line integral R_i

$$\langle a_i^* \rangle_{P(\omega_i^*|\omega_{16,i},\omega_{84,i})} = R_i(s) = \left\langle R_i^{\omega^*}(s) \right\rangle_{P(\omega_i^*|\omega_{16,i},\omega_{84,i})}$$

$$= \int_0^{\frac{1}{\omega_i^*}} s(r\theta_i)(1 - \mathrm{cdf}(r|\omega_{16,i}, \omega_{84,i}))\mathrm{d}r \,, \tag{4.10}$$

where cdf denotes the cumulative density function of Eq. 4.9 with $r = (\omega^*)^{-1}$. We compute the line integral of Eq. (4.10) on the fly for every step, using a parallelized fortran code[1].

The uncertainty of the true position of the source introduces a source dependent supplementary noise contribution $\widehat{\sigma}_i^2$. This uncertainty arises due to the uncertainty of the true source distance, which introduces uncertainty on the line of sight extinction even when given the true extinction density s. The standard deviation of this supplementary noise contribution can be computed as

$$\widehat{\sigma}_i^2 = \mathrm{Var}\left[P\left(a_i^* | \omega_{16,i}, \omega_{84,i}, s\right) \right]$$

$$= \mathrm{Var}\left[\int \mathrm{d}\omega_i^* \, P\left(a_i^*, \omega_i^* | \omega_{16,i}, \omega_{84,i}, s\right) \right]$$

$$= \mathrm{Var}\left[\int \mathrm{d}\omega_i^* \, P\left(a_i^* | s, \omega_i^*\right) P\left(\omega_i^* | \omega_{16,i}, \omega_{84,i}, s\right) \right]$$

$$\leq \mathrm{Var}\left[\int \mathrm{d}\omega_i^* \, P\left(a_i^* | s, \omega_i^*\right) P\left(\omega_i^* | \omega_{16,i}, \omega_{84,i}\right) \right]$$

$$= \mathrm{Var}\left[\int \mathrm{d}\omega_i^* \, \delta\left(a_i^* - R_i^{\omega^*}(s)\right) P\left(\omega_i^* | \omega_{16,i}, \omega_{84,i}\right) \right] \,. \tag{4.11}$$

---

[1]https://gitlab.mpcdf.mpg.de/mglatzle/gda_futils

The last inequality holds as $P(\omega_i^*|\omega_{16,i}, \omega_{84,i})$ has strictly more variance than $P(\omega_i^*|\omega_{16,i}, \omega_{84,i}, s)$. We sample this additional noise contribution before every step of our algorithm. We do this by drawing $M = 20$ samples j of parallaxes ω_i^j according to the statistic given by Eq. (4.9). We then compute

$$\widehat{\sigma}_i^2 = \frac{1}{M} \sum_j R^{\omega_i^j}(s) \tag{4.12}$$

as the sample variance of the extinction estimate using the samples j and the current reconstructed dust extinction density s. This error correction was not done in Leike and Enßlin [64]. However, for this paper the smaller data uncertainty and slightly higher parallax error of the sources raises the importance of computing this error correction, while the use of the new code for the response enables its calculation.

4.4.2 Noise Statistic

The noise statistic specifies how probable an observed G-band extinction value is, given one would know the true amount of G-band extinction for that source. Since there is no detailed noise statistic published for the dataset we use, we have to construct it. To do this, we look at regions of the sky where there is no significant amount of dust expected. These regions were identified by using the Planck dust map [89], more specifically the dust map from the COMMANDER pipeline of the 2014 Planck data release. Here, regions with less than $\exp(2)^{\mu K}/_{rJ}$ were taken to be dustless. This criterion selects 606 pixels of the healpix nside 256 dust map, corresponding to 0.077% of the sky. For every SH_PHOTOFLAG for which we have more than 100 values in these dustless regions, we calculate the mean m_{ph} and standard deviation σ_{ph} of all G-band extinctions. Using these values, we define the probability to measure an extinction a given the true extinction a^* as

$$P(a|a^*, \mathrm{SH\_PHOTOFLAG} = \mathrm{ph}) = \mathscr{G}(a|a^* + m_{\mathrm{ph}}, \sigma_{\mathrm{ph}}^2) \ . \tag{4.13}$$

Table 4.2 show our used means m_{ph} and standard deviations σ_{ph} for all used photoflags ph. As can be seen by investigating Table 4.2, the mean values deviate strongly from zero, and correcting the zero-point is vital to our reconstruction. Note that because we fix the noise statistic for an actual extinction value of zero, the reconstruction might be biased for high extinction values. We discuss some biases that could be attributed to this effect in Sec. 4.7.2.

4.5 Prior

We fold our physical knowledge into the prior of the dust extinction density. We choose the exact same prior model as in [64]. We assume the extinction density s to be positive and spatially correlated. This can be enforced by assuming a log-normal Gaussian process

ph=SH_PHOTOFLAG	mean m_{ph}	standard deviation σ_{ph}
GBPRP	0.493	0.439
GBPRPJHKs	0.131	0.259
GBPRPJHKs#W1W2	0.315	0.538
GBPRPJHKsW1W2	0.116	0.232
GBPRPgrizyJHKs	0.223	0.209
GBPRPgrizyJHKsW1W2	0.156	0.172
GBPRPiJHKsW1W2	0.101	0.219
GBPRPiyJHKsW1W2	0.165	0.234

Table 4.2: SH_PHOTOFLAG values and the corresponding mean and standard deviations for sources in dustless regions. Regions are considered as dustless if the Planck dust map shows weaker emission than $\exp(2)^{\mu K}/_{rJ}$.

prior

$$s_x = \rho_0 \exp(\tau_x), \tag{4.14}$$

$$\tau \curvearrowleft \mathscr{G}(\tau|0, T), \tag{4.15}$$

where ρ_0 is the prior median extinction density and T is the correlation kernel of the Gaussian process τ. The prior median extinction density is a hyper-parameter of our model and we choose $\rho_0 = {}^1\!/_{1000}\mathrm{pc}^{-1}$. We infer the kernel T during our reconstruction. This can be achieved by rewriting s in terms of a generative model

$$s_x = \rho_0 \exp(\mathbb{F}\sqrt{T_k(\xi_T)}\xi_k), \tag{4.16}$$

where all ξ are a-priori standard normal distributed and $T_k(\xi_T)$ is a non-parametric model for the Fourier transformed correlation kernel T_k, also called the spatial correlation power spectrum. One should note that this model is degenerate, any change in T_k can be absorbed into ξ_k instead, as only the product of these two fields enters the overall dust extinction density s. Because of this property, the reconstructed power spectrum T_k does not have to be the empirical power spectrum of s_x, that can be calculated by Fourier transforming and binning. To avoid misunderstandings and artifacts from the degenerate model, we mainly report the empirical power spectrum in this paper, which is computed from posterior samples of s_x. We now focus on our model for the power spectrum T_k. This model assumes the spatial correlation power spectrum to be a preferentially falling power-law, but allows for arbitrary deviations. It can be written as

$$\sqrt{T_k(\xi_T)} = \mathrm{Exp}^*\mathrm{Exp}\big[(m_s + \sigma_s\xi_s)\ln(k) + m_0 + \sigma_0\xi_0$$
$$+ \mathbb{F}_{\ln(k)t}\mathrm{sym}({}^A\!/_{(1+(t/t_0)^2)}\xi_\phi(t))\big], \tag{4.17}$$

where the first part describes a linear function on log-log scale, i.e. a power law; and the second part describes the non-parametric deviations which are assumed to be differentiable on log-log-scale. The operation Exp^* denotes the exponentiation of the coordinate system. More explicitly, $\mathbb{F}_{\ln(k)t}$ is Fourier transformation on log-log scale, and the function sym is

defined as

$$f : [0, 2b] \to \mathbb{R} \tag{4.18}$$

$$\mathrm{sym}(f)(x) = \left. (f(x) - f(2b - x)) \right|_{[0,b]} , \tag{4.19}$$

where $\left. f \right|_M$ denotes the restriction of the domain of the function f to M. The function sym is required to deal with the periodic boundary conditions introduced by the Fourier transform. Details can be found in the appendix of Arras et al. [8]. The hyper-parameters of the model are $(A, t_0, m_s, \sigma_s, m_0, \sigma_0)$ which we chose to be $(11, 0.2, -4, 1, -14, 3)$ in complete analogy to Leike and Enßlin [64].

4.6 Algorithm

We combine the prior and the likelihood into one generative model of the data. We compute approximate posterior samples using Metric Gaussian Variational Inference (MGVI) [54]. This variational approach alternates between drawing samples around the current estimate for the latent parameters and optimizing the current estimate using the average gradient of the samples. The final set of samples is used to derive an uncertainty estimate on all our maps as well as on all derived quantities.

For further parallelization, we split the problem into the 8 octants. Each octant has size 410pc × 410pc × 310pc, such that they overlap for 20pc.

We hereby use a threefold parallelization scheme, parallelizing by octants, parallelizing by samples and a parallelized response. The latter two parallelizations are enabled by our new fortran implementation, which computes the arising line integrals (Eq. (4.10)) on the fly. This is in contrast to our previous paper [64], where we computed the line integral using sparse matrices. Computing the response on the fly takes approximately the same time, but does not have any additional memory requirements and therefore allows for parallelization and a larger reconstruction.

The total number of degrees of freedom is ≈ 417 Million, exceeding those of our previous map by a factor of 30. The total computation time was about 2 weeks of wall clock time, or about 0.5 Million CPUh on 1920 cores.

The final samples of the independently reconstructed octants are combined into the full reconstruction using a differentiable variance-preserving interpolation scheme. The details are described in appendix A.

One noteworthy point is that we cut away the outer 30pc due to artifacts from periodic boundary conditions, resulting in a final map volume of $740\,\mathrm{pc} \times 740\,\mathrm{pc} \times 540\,\mathrm{pc}$.

4.7 Results and Discussion

4.7.1 Results

We were able to reconstruct the nearby dust clouds. Fig. 4.2 shows various maps produced from our result and their relative uncertainty. The maps show tendrils and filaments of dust on scales as small as 2pc up to scales of several hundred parsecs, at which they become disconnected.

All octants inferred similar logarithmic convolution kernels, as can be seen in Fig.4.3. These correlation kernels were computed by taking a slice out of the reconstructed Fourier transformed square root power spectrum.

A comparison of the empirical power spectra of the 8 different octants can be found in Fig. 4.4. Most octants have very similar power spectra, only octant 3 deviates strongly. This octant, located at $180 < l \leq 270$ and $b > 0$ (disregarding the overlap), is strongly devoid of dust, explaining the significantly lower power spectrum.

For the power of the full-volume extinction density, we find a power law with spectral index of 2.52 ± 0.015 at scales from 2pc to 100pc. For the logarithmic power, we report a spectral index of 2.82 ± 0.022 at scales from 2.3pc to 125pc.

Using our reconstruction, we can determine distances to nearby dust clouds. We derive two distance maps. Fig. 4.5 shows the distance to the nearest dust clouds in all directions, as well as an uncertainty on that distance estimate. Note that we compute the distance by checking for the first voxel that exceeds a the threshold of 0.005 e-folds per pc of extinction density. Some of our samples do not exclude the existence of nearby dense clouds, which raises the uncertainty in the corresponding directions tremendously. Fig. 4.6 shows the distance to the densest dust clouds in all directions, as well as an uncertainty on that distance estimate. Note that the uncertainty estimate is quite high on the boundaries of dust clouds, as the reconstruction is uncertain which voxel is densest along these lines of sight.

4.7.2 Comparison

An implicit assumption of the algorithm is that the voxels are smaller than the achievable resolution. Phrased in physical terms, an increase in pixel resolution can be regarded as a renormalization and we need to reach the continuous limit, i.e the limit of negligible discretization effects, for the algorithm to work. This is a byproduct of the inference of the power spectrum, if the achieved posterior resolution is of the order of the imposed voxel resolution, then the reconstruction changes drastically from one voxel to another and the extinction of sources behind an affected voxel also changes dramatically at the boundary. This sudden change in extinction is not compatible with a falling spatial correlation power law in Fourier space, thus the reconstructed power will fall less steep than the real one. This would significantly hamper the ability of the algorithm to extrapolate between measurements. We avoid this behavior by significantly increasing resolution compared to our previous reconstruction [64]. However, it is conceivable that the reconstruction would
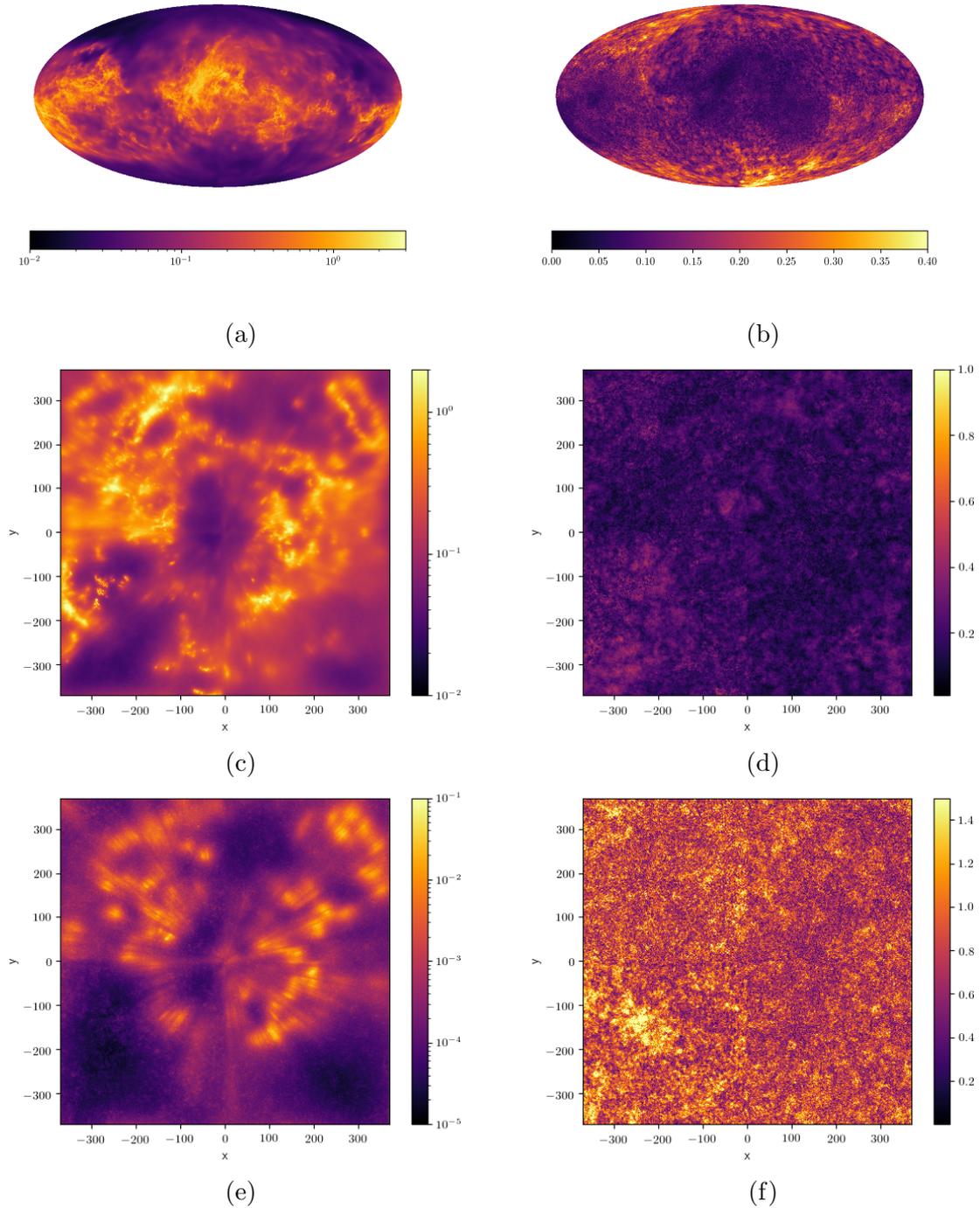
Figure 4.2: Result of our 3D dust reconstruction. The first column shows dust extinction, the second shows the relative error. The first row shows the integrated extinction in e-folds in a Mollweide projection of the whole reconstructed box of $740\,\mathrm{pc} \times 740\,\mathrm{pc} \times 540\,\mathrm{pc}$. The second row also shows integrated extinction in e-folds in the same box, but integrated normal to the Galactic plane instead of radially. The third row shows differential extinction in e-folds per parsec in a slice along the Galactic plane.
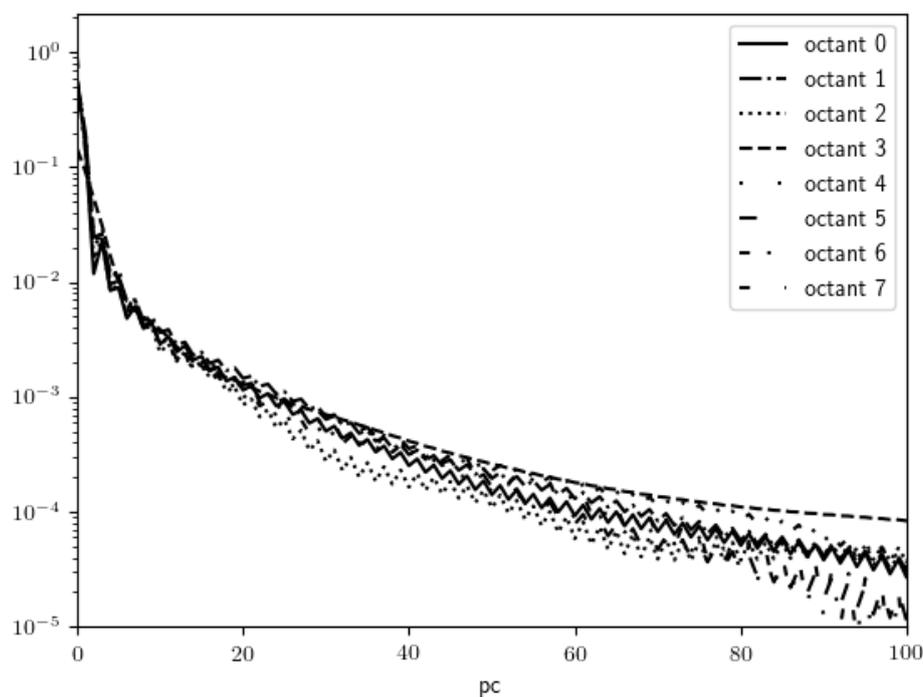
Figure 4.3: Reconstructed correlation kernels for the different octants. Note that the logarithmic dust extinction in our model is the result of an a-priori normal distributed field that is folded with these kernels, dependent on the octant. The octants are arranged such that octant $i = 4b_2 + 2b_1 + b_0$ (for $b_i \in \{0, 1\}$) extends in positive x-direction if and only if $b_0 = 0$, in positive y-direction if and only if $b_1 = 0$ and in positive z-direction if and only if $b_2 = 0$. Note that all kernels fall to about 10% in the first $2\,\mathrm{pc}$.
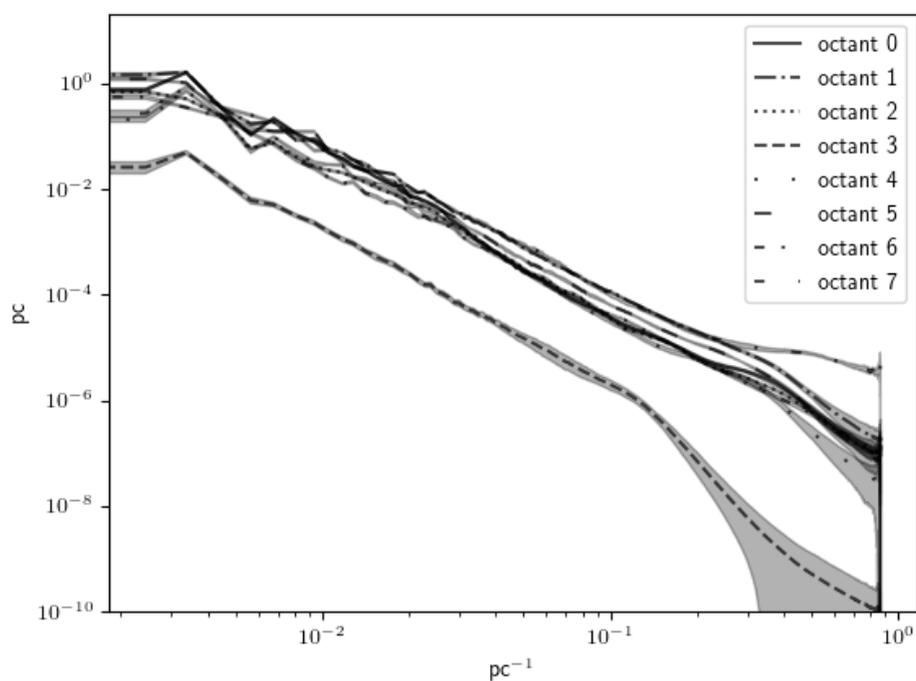
Figure 4.4: Empirical power spectra of the dust extinction density of the eight octants. The octants are arranged such that octant $i = 4b_2 + 2b_1 + b_0$ (for $b_i \in \{0, 1\}$) extends in positive x-direction if and only if $b_0 = 0$, in positive y-direction if and only if $b_1 = 0$ and in positive z-direction if and only if $b_2 = 0$.
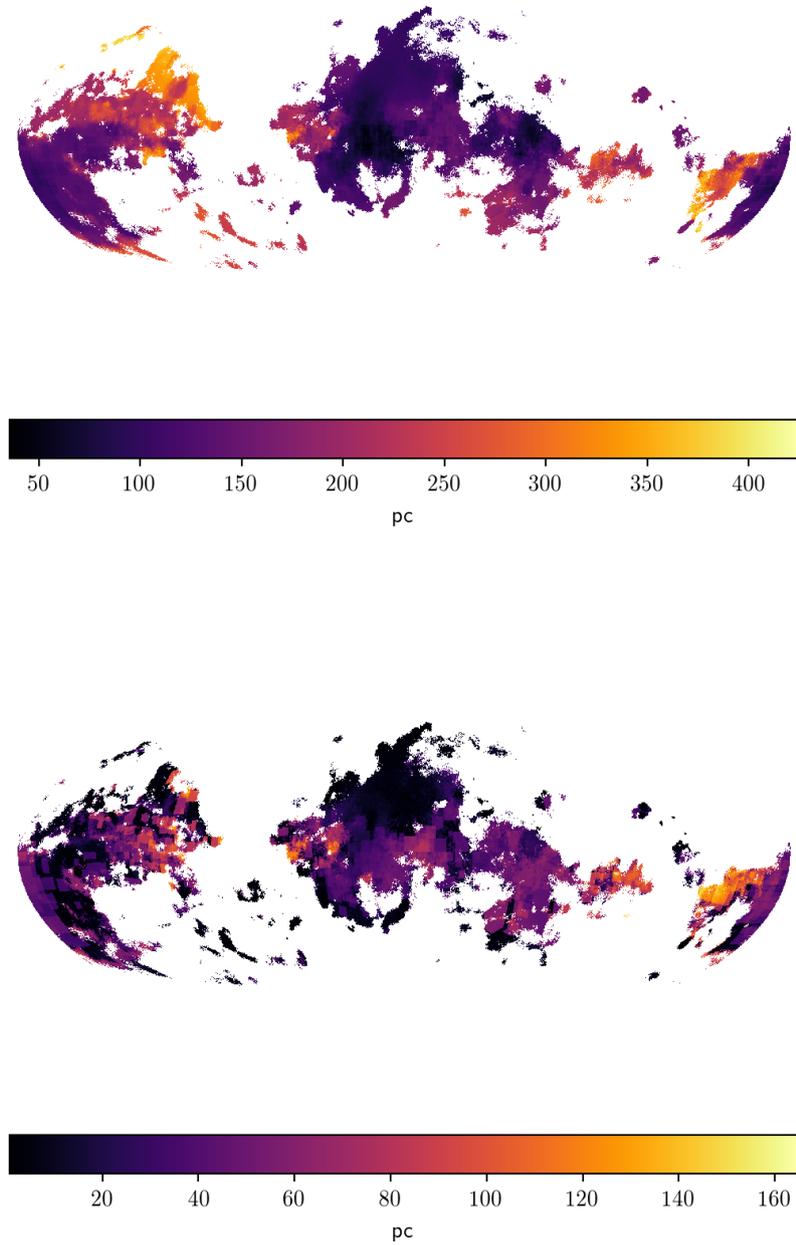
Figure 4.5: A Mollweide projection showing the distance to the first voxel of our reconstruction that exceeds an extinction estimate of 0.005 e-folds per parsec (top panel) and corresponding uncertainty map (bottom panel). Directions for which the threshold is never reached appear in white.
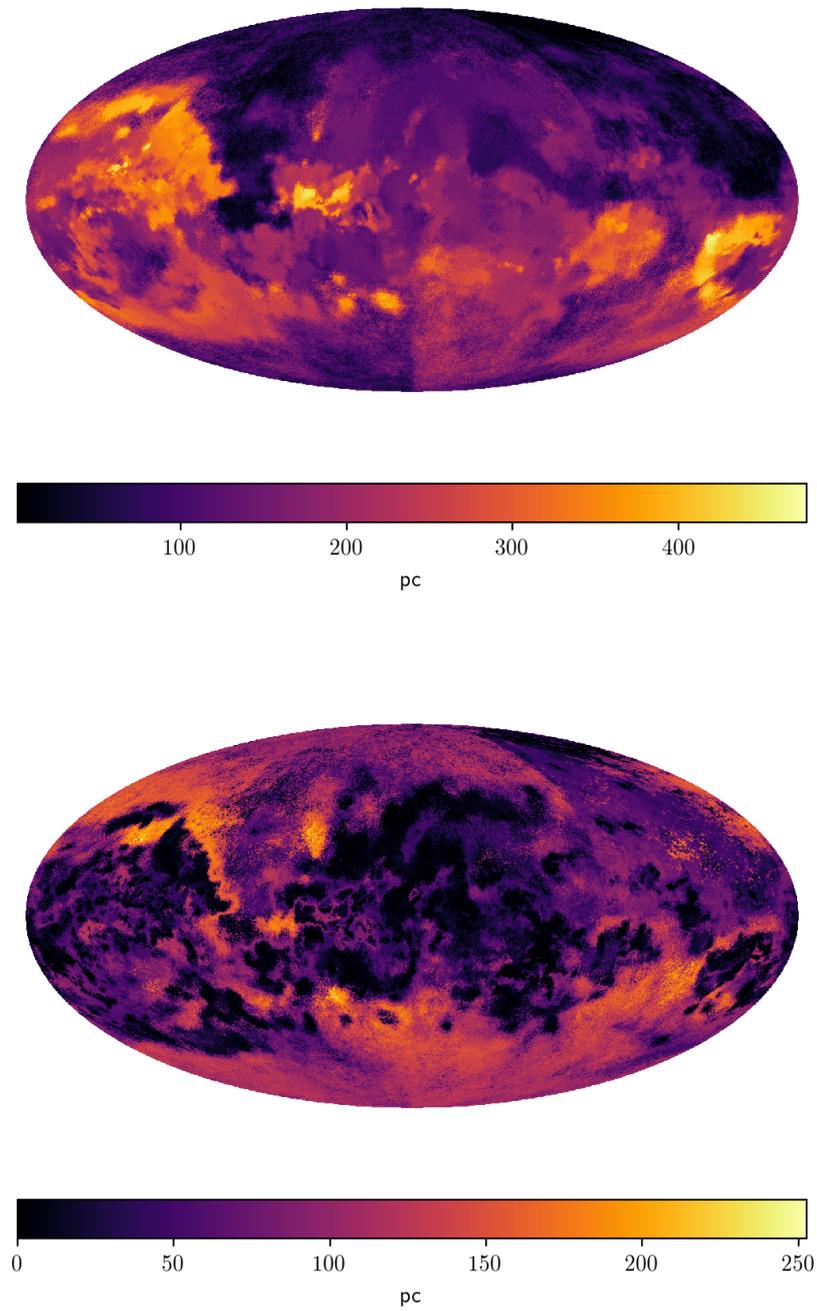
Figure 4.6: A Mollweide projection showing the distance to the voxel with the highest extinction estimate in that direction (top panel) and corresponding uncertainty map (bottom panel).

still benefit from increasing the amount of voxels. We recommend distrusting the smallest scales of our reconstruction, only at scales of 2 pc or larger can the result be considered to be stable. This resolution limit was deduced from the reconstructed logarithmic correlation kernels as seen in Fig. 4.3. At this limit, the reconstructed correlation kernels of the logarithmic dust extinction density have fallen to 10%.

A comparison of our results to Leike and Enßlin [64] can be found in Fig. 4.7; see Fig. 4.8 for a logarithmic version.

We compare our results to the map of Green et al. [45]. Fig. 4.9 shows column density comparisons of the two reconstructions. Fig. 4.10 shows the same column densities, but on a logarithmic scale. A more detailed comparison to Green et al. [45] in angular coordinates can be seen in Fig. 4.11.

In contrast to our old map, we use the dataset of [3], which provides more sources and tighter constraints on the parallax and G-band extinction than the previously used Gaia data. The new reconstruction has a volume of $800\,\text{pc} \times 800\,\text{pc} \times 600\text{pc}$, compared to the $(600\,\text{pc})^3$ cube in Leike and Enßlin [64]. Furthermore, using a designated fortran routine for the computation of the line of sight integrals lead to the necessary speedup to handle the additional data constraints and massively more degrees of freedom. Finally, in the new reconstruction the parallax error is propagated into the measurement error, causing extinction values with stars of high parallax error to be less informative. In Fig. 4.7 one can see dust column densities along Galactic x, y, and z coordinates. Both dust maps agree on the morphology of large dust clouds on large scales. However, the current dust map contains significantly more dust. Part of the reason is that the data we use in the reconstruction of this paper has higher resolution and lower noise, allowing more dust to be reconstructed. We also believe the data used in Leike and Enßlin [64] to be slightly biased to underestimating the amount of dust, an effect that accumulates in a reconstruction that uses many data points. In contrast, the data used in this reconstruction might have a tendency to overestimate the amount of dust, despite our effort to calibrate the zero point (see Sec. 4.4.2).

We furthermore reconstruct our correlation kernel nonparametrically, which should lead to an unbiased estimate of the power spectrum. Fig. 4.12 shows power spectra of Leike and Enßlin [64], the reconstruction of this paper, and of the reconstruction of Green et al. [45]. Our new reconstruction and Leike and Enßlin [64] seem to have quite consistent power spectra. The general tendency of the falling power law is also remarkably consistent with Green et al. [45], however at scales of a few parsec the power spectrum of Green et al. [45] flattens which we believe to be an artifact of how we put their reconstruction on a cartesian grid, i.e. the boundaries of the reconstructions intrinsic voxels introduce steep cuts which flatten the resulting power spectrum. However, none of the power spectra are consistent within the uncertainties estimates. While this seems problematic, one has to bear in mind that all reconstructions focus on dust in differing regions, potentially explaining the difference in the power spectrum. In Fig. 4.13 we show the power spectra of the logarithmic reconstructions. These seem to be less consistent in general, however one has to bear in mind that the logarithmic power spectrum is dominated by regions of low dust content, as these occur more frequently. Using Gaia data our method was found to

underestimate low dust regions, and we anticipate that with the starhorse data we tend to overestimate low-dust regions. Nonetheless we find that the spectral index of 2.82 ± 0.022 at scales from $2.3\,\mathrm{pc}$ to $125\,\mathrm{pc}$ is compatible with the empirical spectral index of Leike and Enßlin [64] within a 2σ joint uncertainty margin. The spectral index of the empirical power spectrum of Leike and Enßlin [64] is 3.2 ± 0.14.[2]The logarithmic power spectrum of Green et al. [45] seems to be inconsistent with our measurements. However, this effect is probably due to how we treat the missing values in that map, where a quarter of the sky was not measured. We have to set these values and every possible choice will impact the derived power spectra. We chose to set them to 10^{-7}, which has minimal impact on the power spectrum on linear scale, but biases the power spectrum of the logarithmic dust extinction density and could potentially explain the difference.

Fig. 4.14 shows a histogram of dust extinction density per voxel. One can see a good agreement between the histogram of our old and our current reconstruction in the region between $10^{-3}\,\mathrm{pc}^{-1}$ and $10^{-1}\,\mathrm{pc}^{-1}$. A dust extinction density of $10^{-4}\,\mathrm{pc}^{-1}$ integrated to the boundary of our simulation cube yields an integrated extinction of 0.046, which is below our noise level even when pooling the information of many stars. For this reason, we do not show the histogram below $10^{-4}\,\mathrm{pc}^{-1}$ as its shape is mostly dependent on how the reconstruction extrapolates into dustless region. From the histogram it can be seen that the dust density is well described by a log-normal distribution. Note that since we show only the part of the histogram that has high signal to noise, this result should be relatively unbiased by our choice of prior. The fitted log-normal model has a standard deviation of $\sigma = 1.906 \pm 0.009$ and a mean of $m = -9.79 \pm 0.04$.

4.7.3 Using the reconstruction

One should note that the reconstruction shows a non-negligible amount of dust in the local bubble. We believe that the level found is an artifact of our noise statistic. As described in Sec. 4.4, our data model involves some heuristics which might systematically affect the reconstruction. This causes estimates made with this data to be biased and we were not able to fully correct for this bias. When integrating the reconstructed dust density to $70\,\mathrm{pc}$, we find that the nearby dust looks like a smeared out version of farther dust clouds, indicating that it is indeed an artifact related to systematic data biases.

The posterior samples of the extinction density are available for download under `https://doi.org/10.5281/zenodo.3750926` or by its DOI 10.5281/zenodo.3750926 . When using the reconstruction we advise to beware of systematic overestimations of dust, especially in the local bubble. When deriving numeric quantities, we advise to do so for every sample and then estimate the mean and standard deviation of the results in order to get an error estimation.

---

[2]Note that [64] reports a spectral index of 3.1 for the reconstructed power spectrum. For this paper, we instead chose to analyze the power spectrum of the resulting maps, which yields slightly different values but enables us to derive uncertainty estimates for all compared maps in the same way.
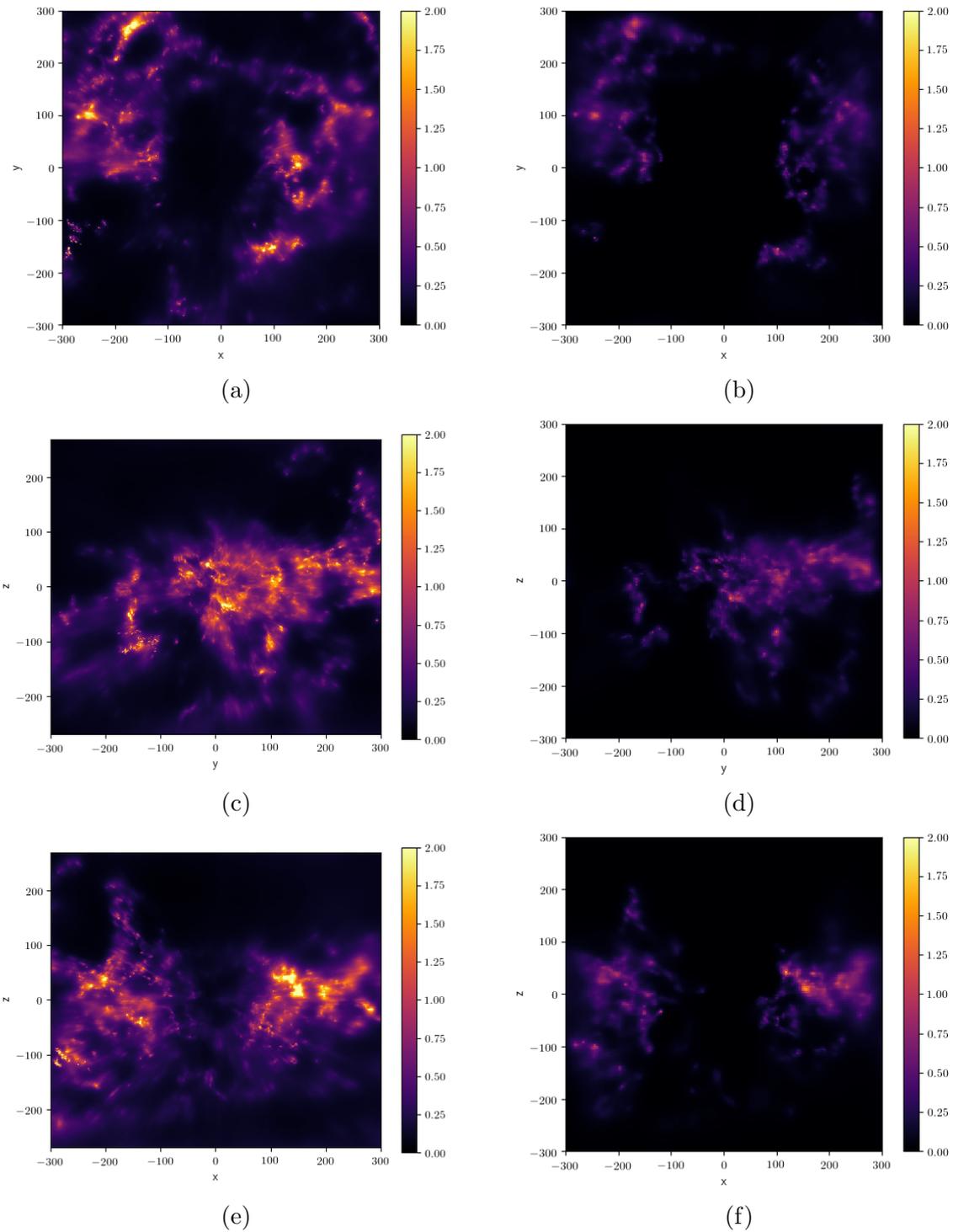
Figure 4.7: Comparison of column densities of our current reconstruction (left column) and Leike and Enßlin [64] (right column). The rows show integrated dust extinction for sightlines parallel to the *z*- *x*- and *y*-axis respectively.
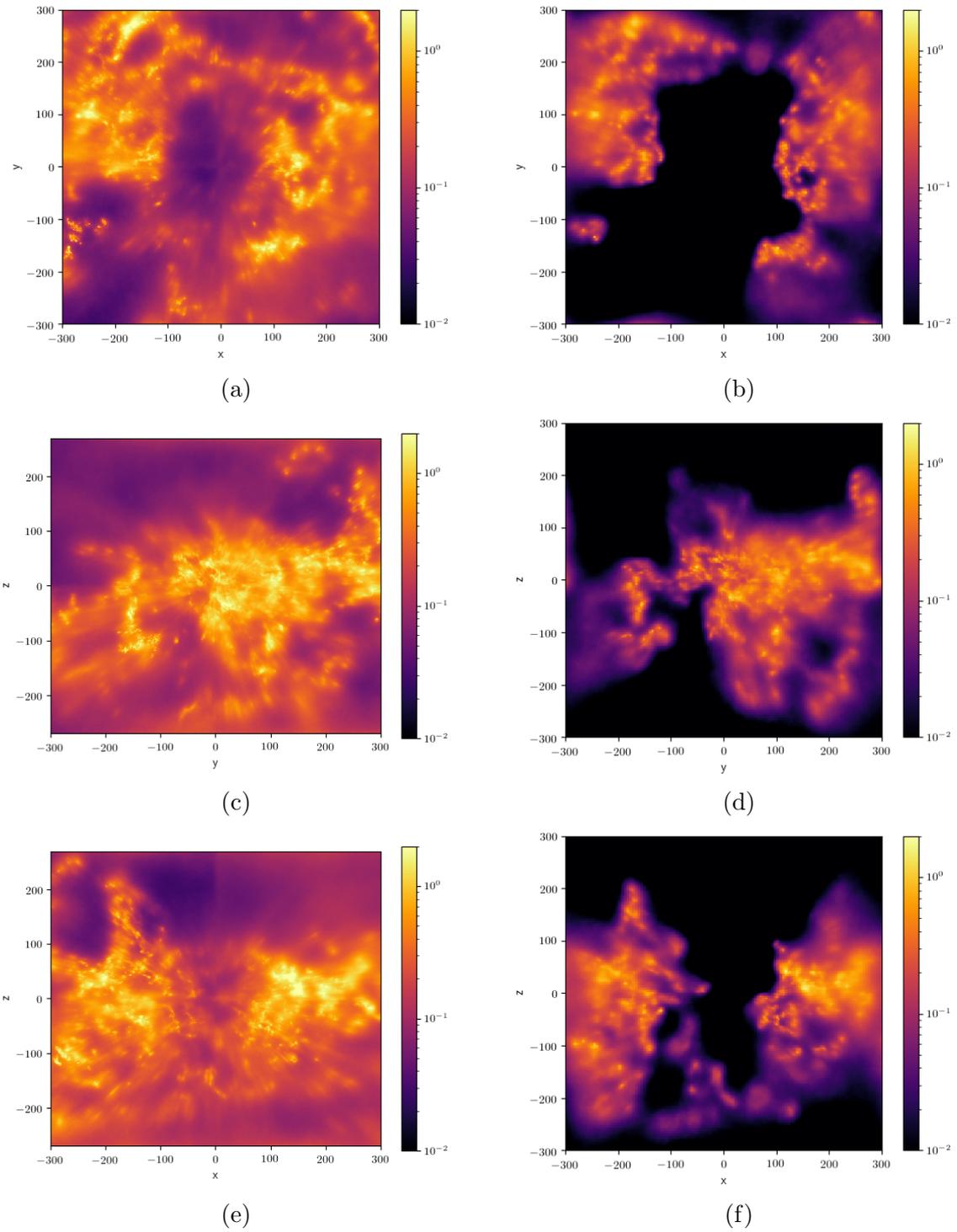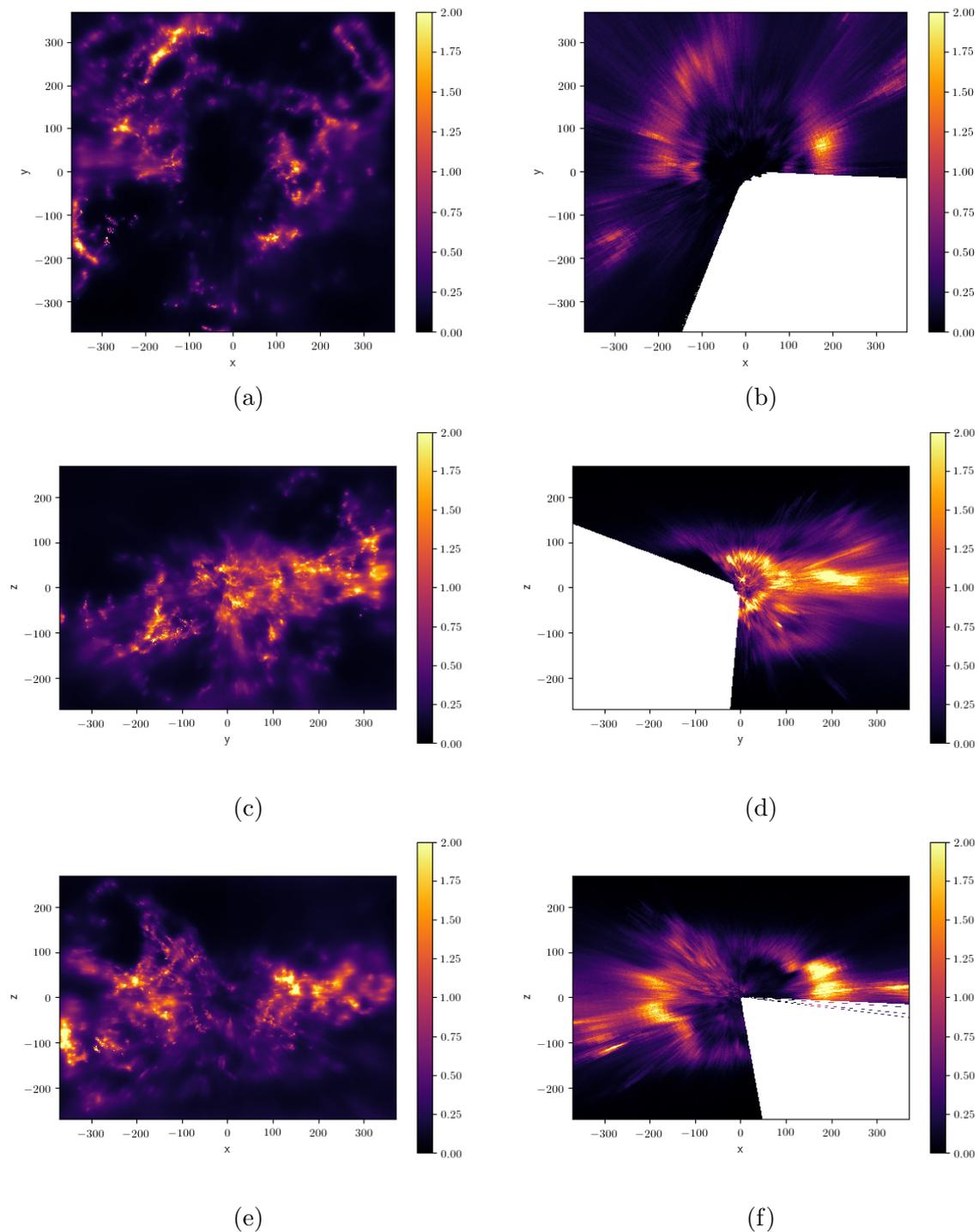
Figure 4.8: As Fig. 4.7 but on logarithmic scale.

Figure 4.9: Column density comparison of our current reconstruction (left column) and that of Green et al. [45] (right column). The rows show integrated dust extinction for sightlines parallel to the z- x- and y-axis respectively. Note that for Green et al. [45] we show the integrated extinction only if more than 50% of the projected voxels exist in the reconstruction.

(a)

(b)

(c)

(d)

(e)

(f)
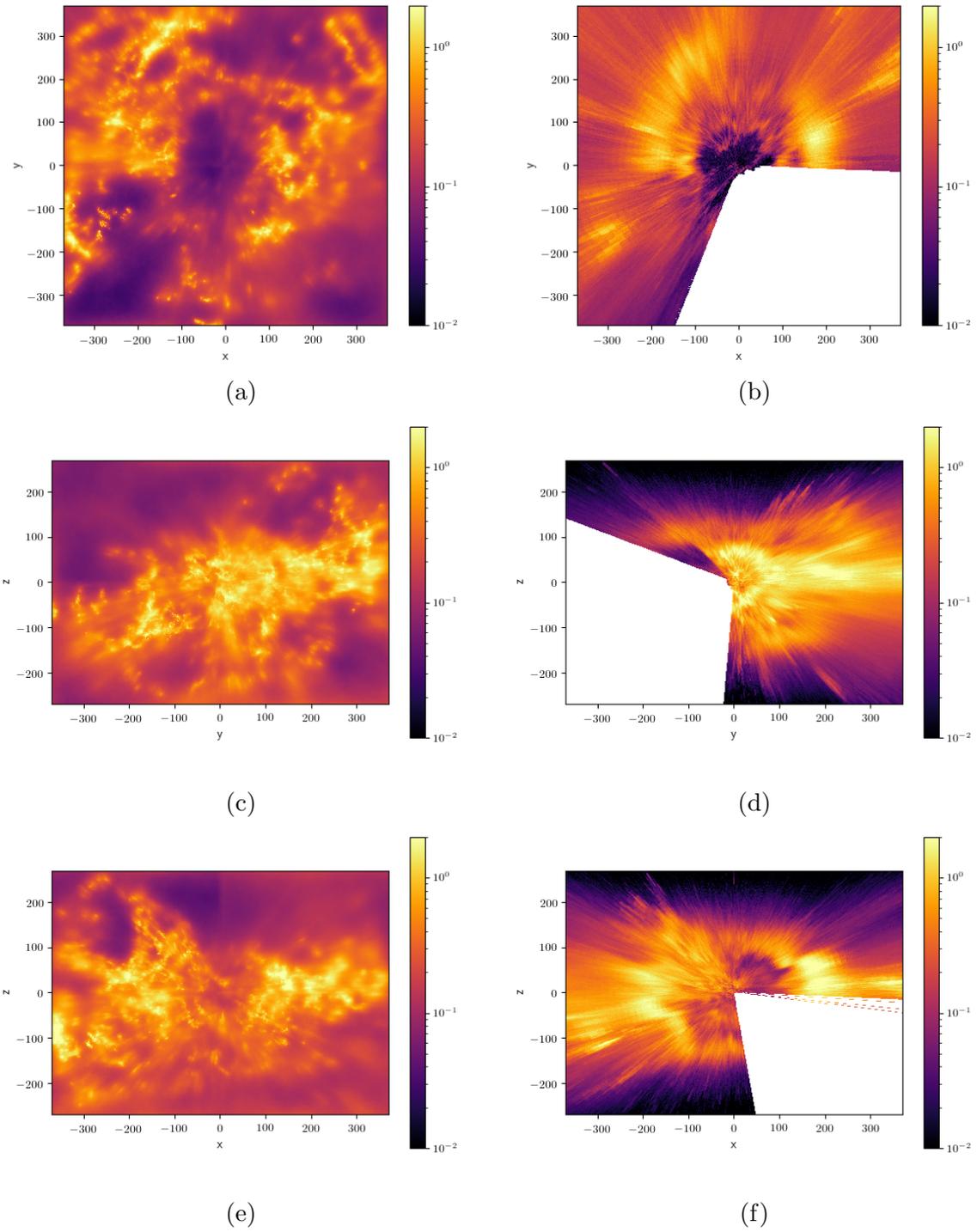
Figure 4.10: As Fig. 4.9 but on logarithmic scale.

(a)   (b)

(c)   (d)

(e)   (f)
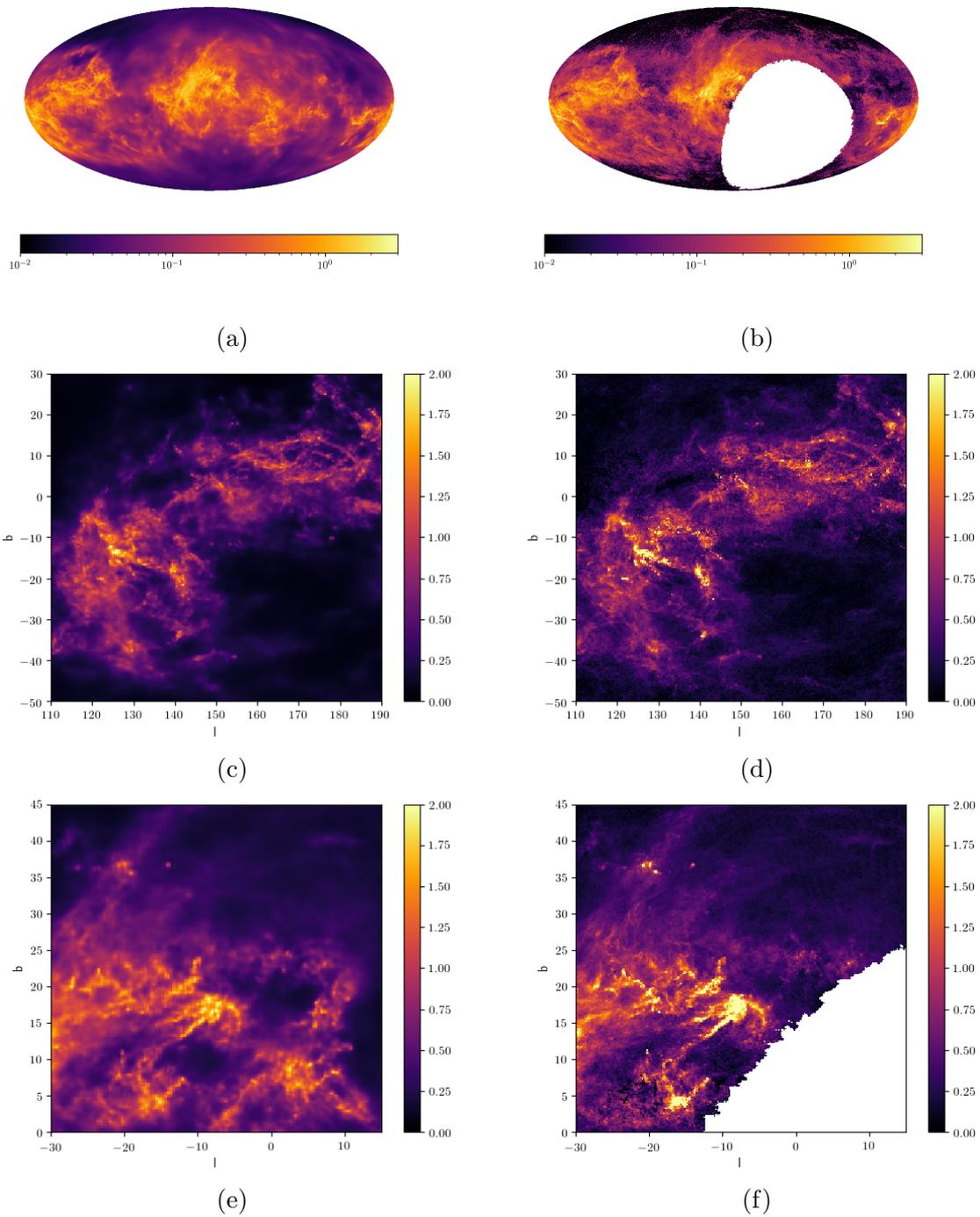
Figure 4.11: Comparison of integrated extinction of our reconstruction (left column) and that of Green et al. [45] (right column) in sky projection. The rows show integrated dust extinction out to the boundary of our $740\,\text{pc} \times 740\,\text{pc} \times 540\,\text{pc}$ box in an all sky view (first row), as well as two selected directions towards the Galactic anticenter (middle row) and center (last row).
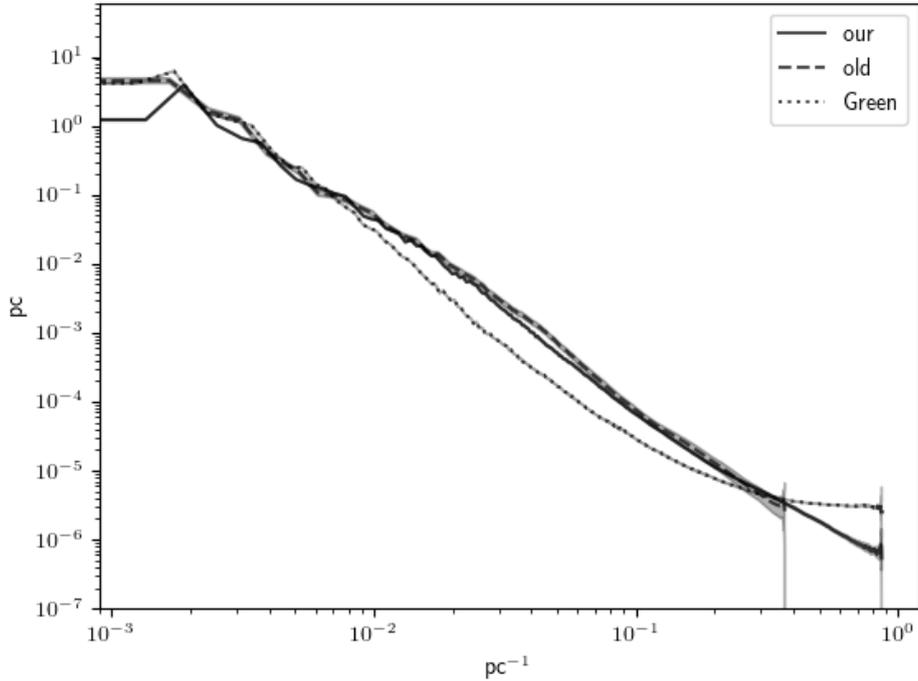
Figure 4.12: Empirical power spectra of the dust extinction density of this paper
             (solid line), Leike and Enßlin [64] (dashed line) and the reconstruction of
             Green et al. [45] (dotted line).

4.7.4 Implications

Our map can be used to constrain simulations of the ISM. For example, in simulations of
radiatively cooling dust clouds in hot winds, it has been shown that dust density power
spectra are flatter than was previously thought [87]. Our maps show power spectra com-
patible with these simulations, and morphologically similar structures. Our reconstructed
spectral index of 2.82 ± 0.022 at scales from $2.3\,\mathrm{pc}$ to $125\,\mathrm{pc}$ could be used to constrain
parameters of sub-grid models of simulations of the ISM. Furthermore we find the density
histogram of the logarithm of the G-band dust extinction density in e-folds per parsec
shown in Fig. 4.14 is well described by a log-normal distribution with standard deviation
$\sigma = 1.906 \pm 0.009$ and mean $m = -9.79 \pm 0.04$ on extinction density scales from $10^{-4}\,\mathrm{pc}^{-1}$
to $1\,\mathrm{pc}^{-1}$.

4.8 Conclusion

We were able to reconstruct the dust clouds within $\sim 400\,\mathrm{pc}$ of the Sun down to a resolution
of $2\,\mathrm{pc}$, improving in resolution and volume on our previous reconstruction [64]. The
resulting map is public and can be downloaded; see Sec. 4.7.3 for details. Distances to
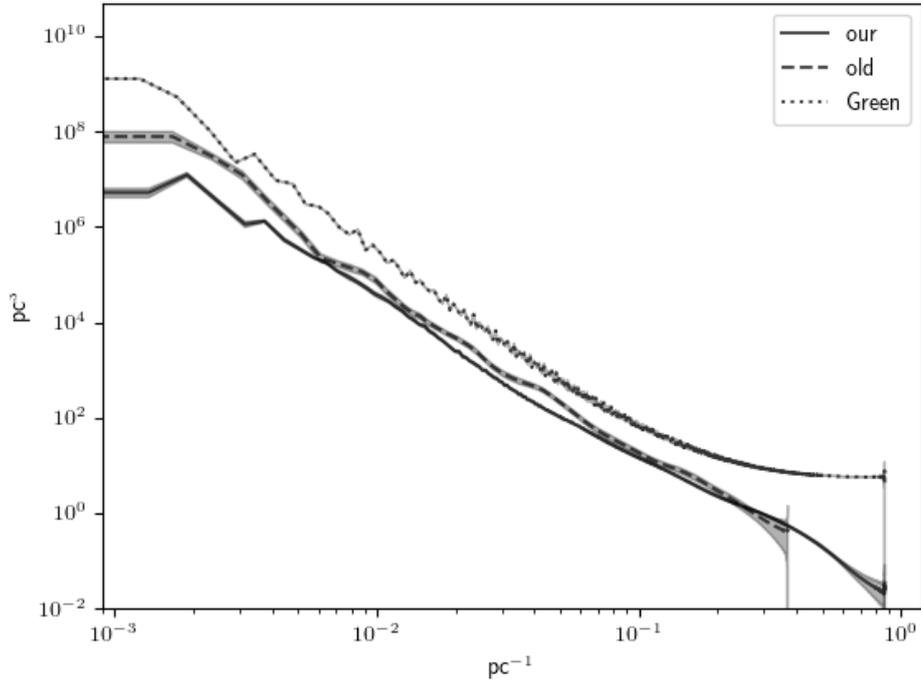
Figure 4.13: Empirical power spectra of the logarithmic dust extinction density of this paper (solid line), Leike and Enßlin [64] (dashed line) and the reconstruction of Green et al. [45] (dotted line).

and densities of all dust clouds larger than $2\,\mathrm{pc}$ are expected to be well constrained by the reconstruction. We report our estimate on the power spectrum of the dust extinction density as well as the logarithmic density. Furthermore, we provide a histogram of dust densities in the interstellar medium and find them to be well described by a log-normal model. We hope that our diverse summary statistics allow simulations of the ISM to be constrained.
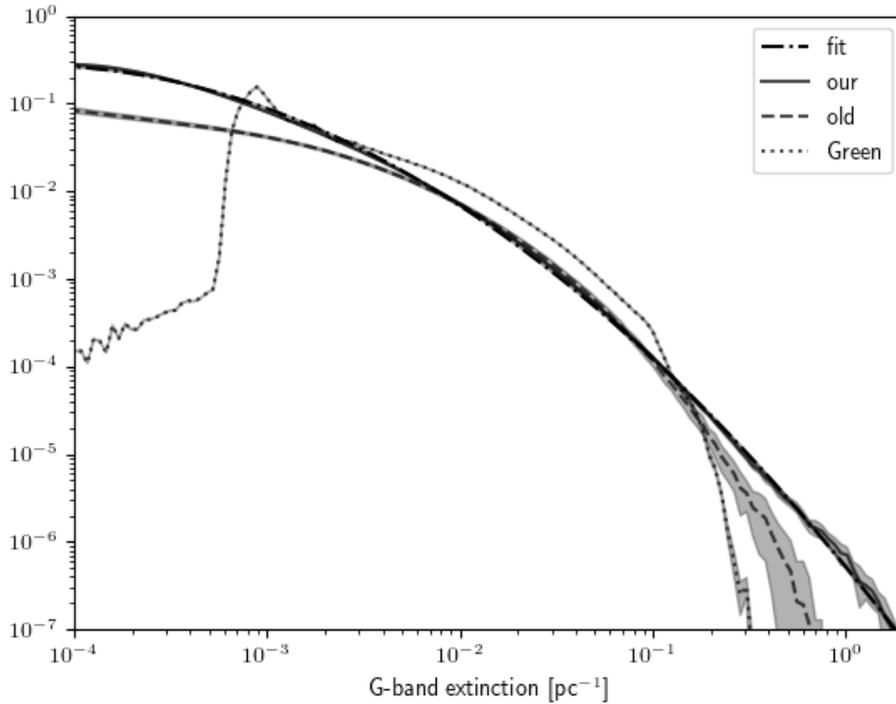
4.9 Acknowledgements

Figure 4.14: Histogram of dust extinction density per voxel of this paper (solid line), Leike and Enßlin [64] (dashed line) and the reconstruction of Green et al. [45] (dotted line). We overplot a log-normal model that was fitted to our reconstructed logarithmic extinction density (dash-dotted line). The curve of the fit is described by $f(x) \propto \exp\left(-0.5\sigma^{-2}\left(\ln(x) - m\right)\right)$ with $\sigma = 1.906$ and $m = -9.79$ and follows our empirical distribution function closely.

4.10 Appendix A: Interpolation scheme

To parallelize the reconstruction, we reconstructed the eight octants of the coordinate system independently, with a 20 pc overlap region. To get one final reconstruction, we have to glue these reconstructions together and specify how we deal with the overlap region. We do so using a differentiable, variance preserving interpolation scheme, i.e. if the octants are differentiable then the result is differentiable; and the final samples have at least the variance that the individual reconstructions imposed. We compute the uncorrected interpolated logarithmic extinction samples $\tau'(x)_j$ from the samples of the

eight octants $o^i(x)_j$ as

$$\tau'(x)_j = \sum_i w_i(x) o^i(x)_j \ .$$
(4.20)

Hereby the weights $w_i(x)$ can be computed as

$$w_i(x) = \prod_{k=0}^{2} \left| b_k(i) - f\left(\frac{x_k - 9\,\mathrm{pc}}{18\,\mathrm{pc}}\right)\right| \ ,$$
(4.21)

$$\text{where } f(x) = \begin{cases} 0 & x \in (-\infty, 0] \\ x^2(3 - 2x) & x \in (0, 1) \\ 1 & \text{otherwise,} \end{cases}$$
(4.22)

and $b_k(i)$ denotes the k-th digit of i in binary format. Noteworthy properties of this scheme are that the weights sum to one

$$\forall x \sum_i w_i(x) = 1,$$
(4.23)

and the polynomial f is the unique polynomial of degree 3 such that

$$f(0) = 0 \ ,$$
(4.24)

$$f(1) = 1 \ ,$$
(4.25)

$$\frac{\partial f}{\partial x}(0) = 0 \ ,$$
(4.26)

$$\frac{\partial f}{\partial x}(1) = 0 \ .$$
(4.27)

By using this interpolation scheme only voxels that have a coordinate x_k of which the absolute value is at most $8\,\mathrm{pc}$ get non-zero contributions from more than one octant, or put in a different way: We cut away the outermost $2\,\mathrm{pc}$ of all reconstructions, which mitigates artifacts from periodic boundary conditions. From the preliminary logarithmic extinction density $\tau'(x)_j$ we can compute the logarithmic sample mean $\bar{\tau}(x)$ as

$$\bar{\tau} = \frac{1}{N} \sum_j \tau'_j \ .$$
(4.28)

Here N denotes the number of samples. The variance of τ'_j is artificially low at overlapping regions, as independent samples were averaged. We correct for this effect and compute the overall logarithmic extinction density samples $\tau(x)_j$ as

$$\tau(x)_j = \frac{\tau'(x)_j - \bar{\tau}(x)}{\sqrt{\sum_i w_i(x)^2}} + \bar{\tau}(x) \ .$$
(4.29)

4.11 Supplementary Material

This chapter is not part of the publication and was supplemented in this thesis to provide a more detailed view of the main result of this thesis, the three dimensional dust reconstruction.
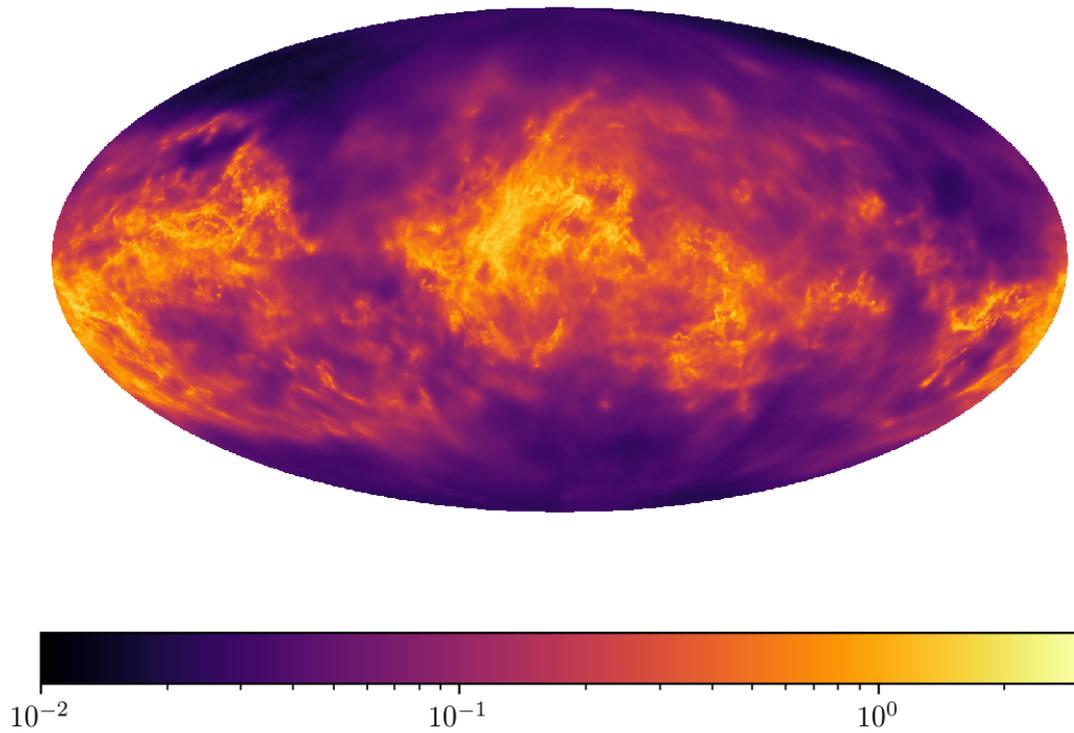

Figure 4.15: Integrated G-band extinction to the boundary of our reconstruction volume, viewed from the location of the Sun.
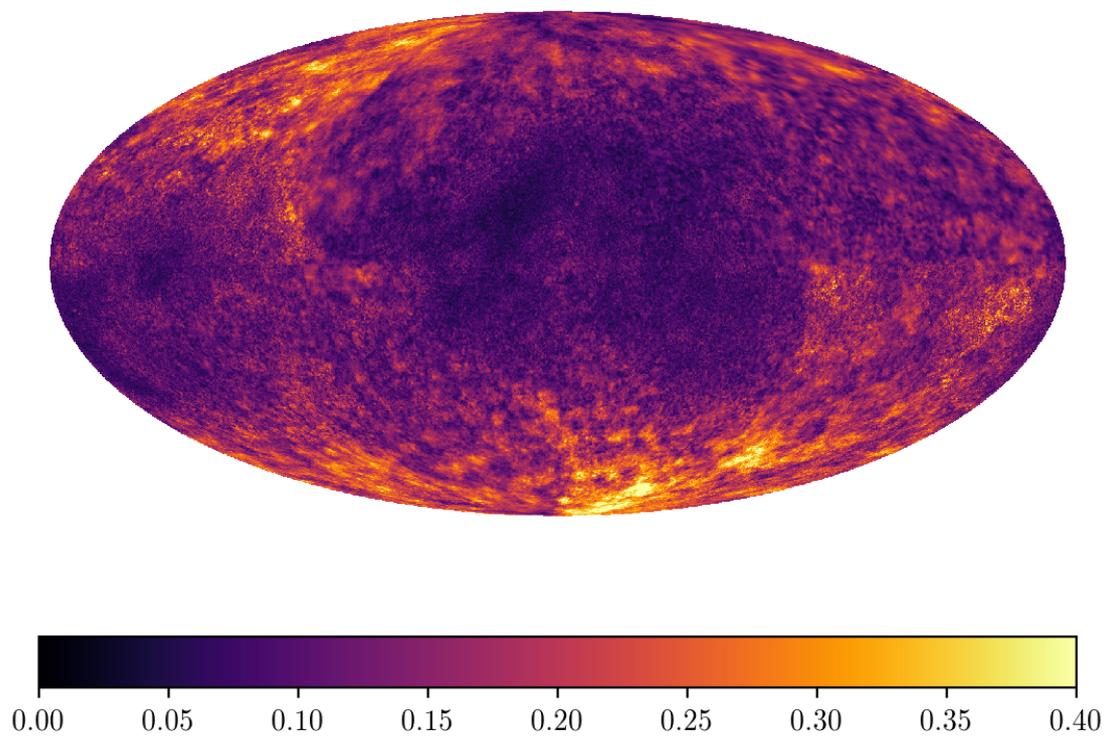
Figure 4.16:
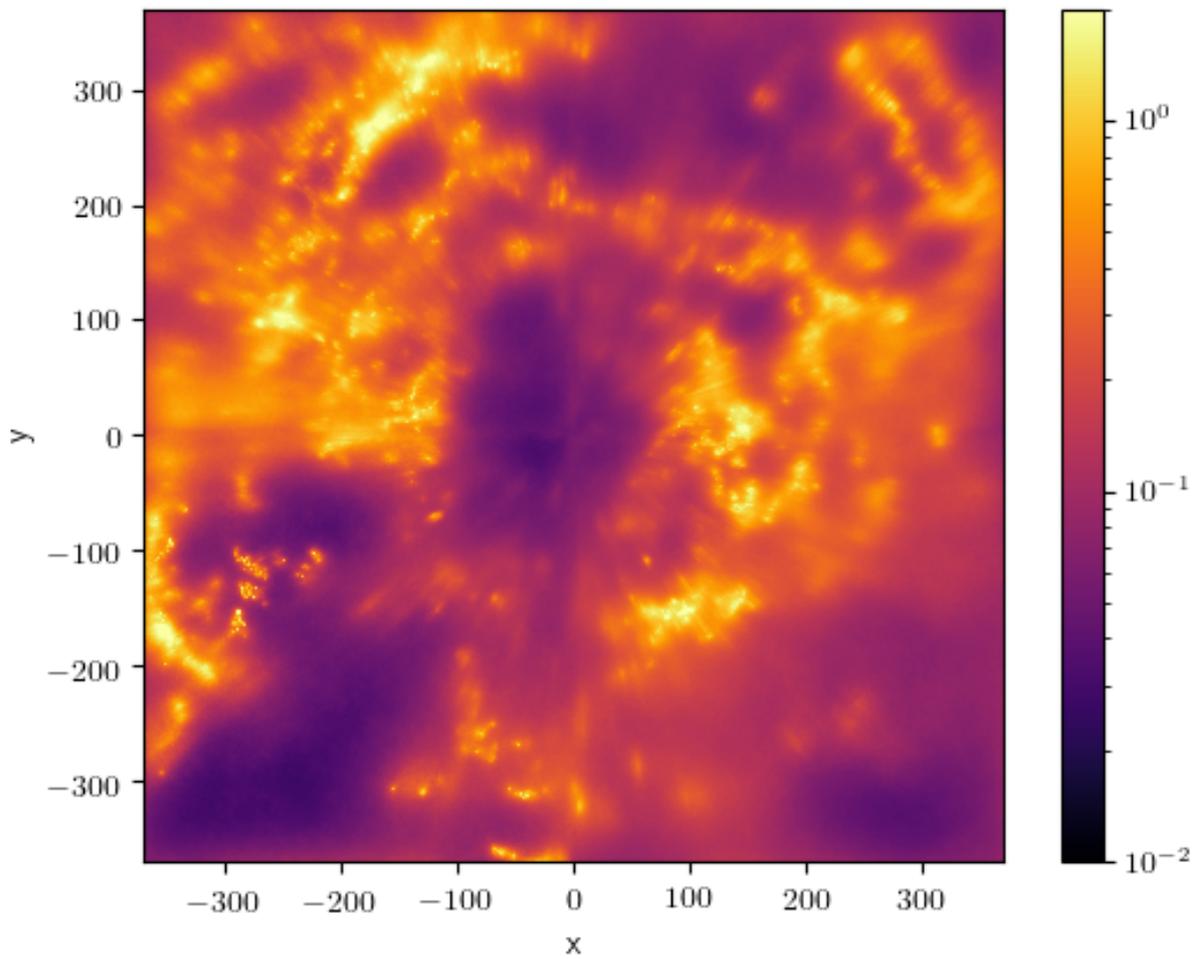Relative uncertainty map to Fig. 4.15.

Figure 4.17: Column density of G-band extinction in the Galactic plane around the Sun.
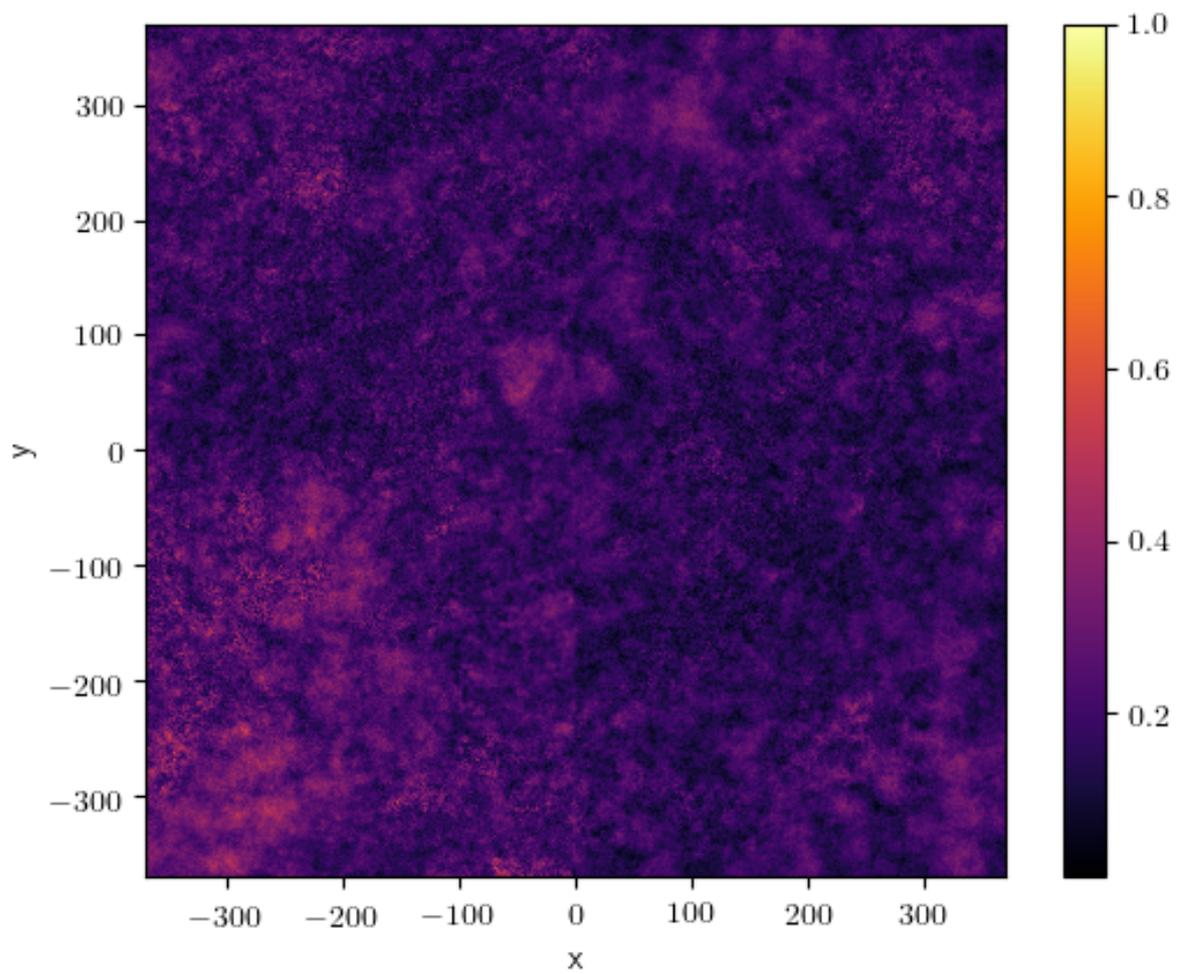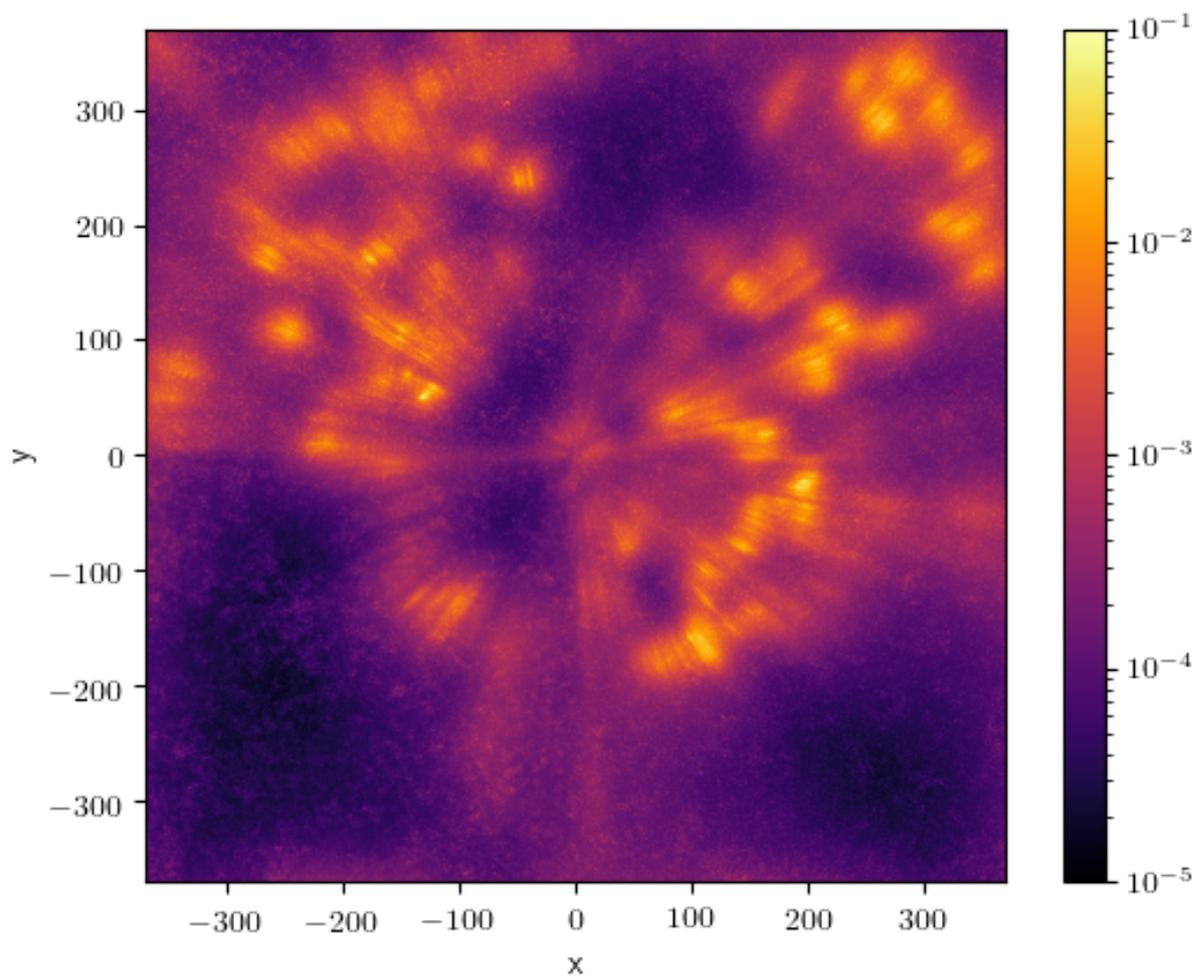
Figure 4.18: Relative uncertainty map to Fig. 4.17

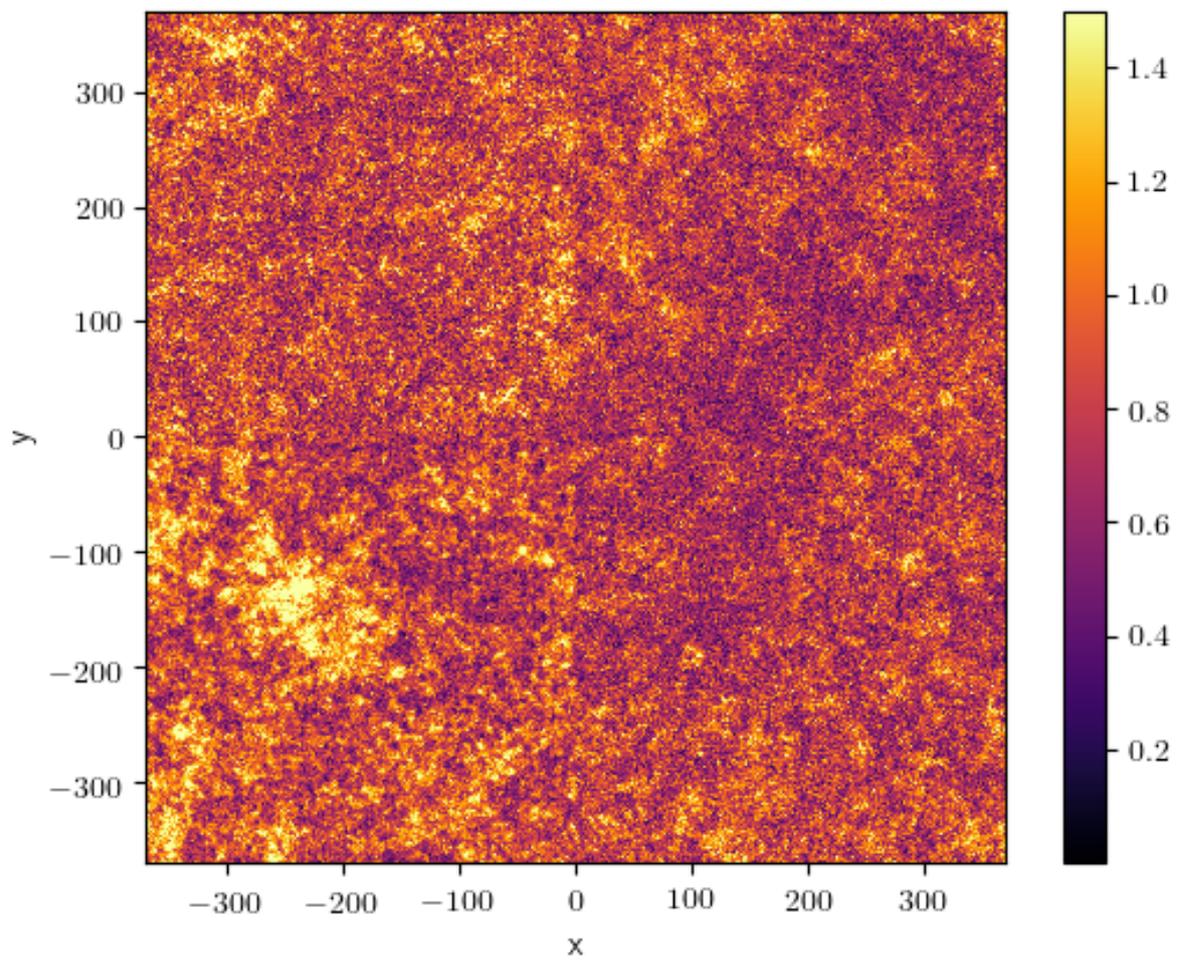Figure 4.19: Differential G-band extinction within 2 pc of the Galactic plane.

Figure 4.20
Relative uncertainty map to Fig. 4.19.

5 Computational aspects

This chapter discusses some numerical aspects of the employed algorithms. We presented some of the first applications of the MGVI algorithm [54], and as such we would like to discuss some numerical considerations and practices that are relevant for the understanding, application, and derivation of MGVI. Here, we especially focus on the minimization, discussing why we use the Newton conjugate gradient method and test the performance of two different variants of the method. The discussed aspects might generalize to other high dimensional minimization problems.

Many of the numerical considerations are not new and can be found in books about numerical optimization such as Nocedal and Wright [68]. We reproduce some of these considerations here for clarity and educational purposes, and in order to make this thesis self-contained. Though some aspects of this chapter are introductory, a familiarity of the reader with the Newton method as well as the conjugate gradient algorithm is assumed.

5.1 Discussing the heuristic of the Newton conjugate gradient algorithm for minimizations in a high dimensional setting

5.1.1 Theory

The goal of minimization is to find the location of a local minimum $\overline{x}$ of an objective function f. Minimizers do so by proposing a series of trial points x_i, where the function f, its gradient, and metric M can be evaluated. By metric M we mean a position dependent positive definite operator that is equal to the second derivative of f near its minima[1].

In this chapter we will mainly regard the Newton conjugate gradient minimizer and the gradient descent minimizer and explain their properties and limitations in high dimensional minimization.

One way to classify minimizers that is often considered is their convergence order. A minimizer is said to have convergence order λ if and only if the series

$$\ln\left(|x_i - \overline{x}|\right) \in \mathscr{O}(i^\lambda), \tag{5.1}$$

i.e. it is said to have linear convergence order if the difference to the true value reduces by a factor in every iteration of the minimization. In theory, if it takes a certain time to

---

[1]In practice, the metric will not be exactly equal to the second derivative at the minima, but many considerations apply nonetheless.

evaluate the first digit of the minimum's location, one just has to invest 16 times more time to get to double precision floating point accuracy. Thus, looking at the definition from a practitioner's perspective, it is not clear why anyone would need faster than linear convergence. We will see that the problem is not the linear convergence, but the factor by which one shrinks the distance to the minimum being too close to 1.

To investigate this, it is instructive to consider the performance of a steepest descent scheme, which always follows the gradient downwards. Let us assume the objective function f given by the simple quadratic form

$$f(x) = \frac{1}{2}x^\dagger M x = \frac{1}{2}x^\dagger \begin{pmatrix} 1 & 0 \\ 0 & 100^2 \end{pmatrix} x \,. \tag{5.2}$$

Consider the initial value for the minimization to be $x_0 = (1, 1/100^2)^\dagger$. Then the gradient of f is

$$\frac{\partial f}{\partial x}(x_0) = M x_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \,. \tag{5.3}$$

A gradient descent scheme then descents in negative gradient direction, i.e. $x_1 = x_0 - \alpha(M x_0)$, where α is the step size. Choosing

$$\alpha = \frac{(M x_0)^\dagger (M x_0)}{(M x_0)^\dagger M (M x_0)} = \frac{2}{1 + 100^2} \,, \tag{5.4}$$

minimizes the objective function along the descent direction and yields the new position

$$\begin{aligned} x_1 &= \left(\mathbb{1} - \frac{x_0^\dagger M^2 x_0}{x_0^\dagger M^3 x_0} M \right) x_0 \\ &= \begin{pmatrix} 1 - \frac{2}{100^2+1} & 0 \\ 0 & 1 - 100^2 \frac{2}{100^2+1} \end{pmatrix} x_0 \\ &= \begin{pmatrix} 1 - \frac{2}{100^2+1} & 0 \\ 0 & -(1 - \frac{2}{100^2+1}) \end{pmatrix} x_0 \,. \end{aligned} \tag{5.5}$$

Thus, x_1 is closer to the minimum of f by a factor z

$$z = \frac{\|x_1\|}{\|x_0\|} \approx 1 - \frac{2}{100^2} \,. \tag{5.6}$$

In other words: To decrease the distance to the minimum by a factor of e it takes at least 5000 evaluations of the objective function and its derivative. One might expect that this bad convergence rate is an artifact of the specific numbers and initial conditions of our example. However, one can show that in general the convergence rate of a steepest descent approaches (Theorem 3.4 of Nocedal and Wright [68]) is

$$z = \frac{1}{1 + 1/C} - \frac{1}{C + 1} \approx 1 - \frac{2}{C} \,, \tag{5.7}$$

where C is the condition number of the matrix M, i.e. the quotient of its largest and smallest eigenvalue.

One can generalize this consideration to all minimization problems, as near the minimum $\overline{x}$, the objective function f is well approximated by the quadratic form

$$f(\overline{x} + \mathrm{d}x) \in f(\overline{x}) + \mathrm{d}x^\dagger M \mathrm{d}x + \mathscr{O}(\mathrm{d}x^3) \ . \tag{5.8}$$

Typically, one has no control over the condition number of M and it tends to be very large in high dimensions. The more parameters a minimization problem has, the less likely it is that all eigenvalues of M end up at the same order of magnitude, and the less possible it is to somehow tune the problem by hand in order to reduce the condition number. To draw a conclusion out of the considerations so far, the linear convergence order of steepest descent is problematic for minimizations, not because linear convergence is bad in general, but because one has no control over the factor z by which one shrinks the distance to the minimum in each step, and z tends to be very close to 1. To give some rough order of magnitude, for our minimization problem in chapter 3 the condition number of the metric is about[2] 10^6, and in chapter 4 we had $C \approx 10^7$.

One way to overcome the complexity induced by a large condition number in minimization is by using a second order minimizer, such as the Newton conjugate gradient. We also applied the Newton conjugate gradient algorithm for our Bayesian inverse problem presented in Chapter 3 and 4. The Newton conjugate gradient iterates between determining the descent direction and performing a line search in the descent direction. The descent direction d_i in the i-th step of the iteration is found by numerically solving the linear inverse problem

$$M_i d_i = \frac{\partial f}{\partial x}(x_i) \, , \tag{5.9}$$

with M the metric at location x_i for the descent direction d_i. This linear inverse problem can also be regarded as a minimization problem with objective function

$$f_{\mathrm{CG}} = \frac{1}{2} d_i^\dagger M_i d_i - d_i^\dagger \frac{\partial f}{\partial x}(x_i) \, , \tag{5.10}$$

The numerical inversion of the linear problem given by Eq. (5.9) is not exact, and the resulting descent direction has a residual

$$r_i^j = \frac{\partial f}{\partial x}(x_i) - M_i d_i^j \, , \tag{5.11}$$

which quantifies how good Eq. (5.9) was solved. Here the upper index j denotes that j iterations of the conjugate gradient algorithm were performed. Once the descent direction d_i is calculated, $x_{i+1}' = x_i - d_i$ is proposed as the new iteration step. To increase the stability of the minimization, a line search is performed in this descent direction, and the

---

[2]The exact value can vary during minimization.

new location is the result of that line search. However, the involved criteria and algorithms for the line search are tedious and ultimately irrelevant for this discussion. For the coming section we will assume $x_{i+1} = x_i - d_i$, which is true when x_i is near the minimum $\overline{x}$.

Choosing a step $x_{i+1} = x_i - d_i$ minimizes the quadratic part of Eq. (5.8). The residuals to the actual objective function are thus of third order in the difference to the minimum's location. In the Newton conjugate gradient algorithm, one needs to specify how many conjugate gradient steps are taken to solve the linear inverse problem given by Eq. (5.9). This choice involves a tradeoff. On the one hand, taking more steps for solving the linear problem given by Eq.(5.9), one might end up with a descent direction that ideally directly leads towards the minimum. On the other hand, computing steps for the conjugate gradient algorithm is almost as expensive as proposing a new location x_{i+1}. For the new location, the approximation given by Eq. (5.8) is more accurate. This makes conjugate gradient steps at the new location more efficient than at the old location x_i.

According to Nocedal and Wright [68], the point at which the conjugate gradient is usually terminated is determined by imposing a factor η_i by which the norm of the residual r_i^j has to have decreased:

$$\|r_i^j\| \le \eta_i \left\|\frac{\partial f}{\partial x}(x_i)\right\| , \tag{5.12}$$

For example the Python package Scipy [49], which bases its Newton conjugate gradient method on Nocedal and Wright [68], chooses

$$\eta_i = \min\left(0.5, \sqrt{\left\|\frac{\partial f}{\partial x}(x_i)\right\|}\right) , \tag{5.13}$$

with the norm $\|.\|$ being the 1-norm. In words, the conjugate gradient algorithm is run until the 1-norm of the residual is half its initial value, unless its 1-norm was smaller than 0.5 to begin with. This dependence of η on the 1-norm is troublesome from the perspective of dimensional analysis, as the gradient is in general not a unitless quantity, but 0.5 is. To be more precise, the magnitude of the norm of the gradient $\frac{\partial f}{\partial x}$ could be anywhere, it can vary wildly for different objective functions. Thus one would expect that for one minimization problem one is either always in the regime where $\eta_i = 0.5$, or always in the regime where $\eta_i = \left\|\frac{\partial f}{\partial x}(x_i)\right\|$. In fact for MGVI, one tends to be in the branch were $\eta_i = 0.5$, as the 1-norm scales with the number of dimensions, and through our standardization [53] the entries of the gradient usually are at least of order 1.

In the software package NIFTy 6, the conjugate gradient is instead terminated if

$$|f_{\mathrm{CG}}(d_i^j) - f_{\mathrm{CG}}(d_i^{j-1})| \le \eta' \left(f(x_i) - f(x_{i-1})\right) , \tag{5.14}$$

with $\eta' = 0.1$ and f_{CG} as in Eq. (5.10). This satisfies the requirements of dimensional analysis, as now both sides of the equation have the same units. We will evaluate the performance of this heuristic rule in several examples.

5.1.2 Tests

We will test the performance of the Newton conjugate gradient with heuristic I given by Eq. (5.12) and the here proposed heuristic II given by Eq. (5.14) by applying them to the minimization of different objective functions. We use increasingly practice relevant examples, designed to mimic the optimization tasks that are being solved with NIFTy.

Rosenbrock function

The Rosenbrock function [80] and its multidimensional analogs are popular minimization benchmarks. It is given by

$$f_{\text{Rosenbrock}} : \mathbb{R}^N \to \mathbb{R} \,, \tag{5.15}$$

$$f_{\text{Rosenbrock}}(x) = \sum_{i=0}^{N-1} b \left(x_{i+1} - x_i^2\right)^2 + \left(1 - x_i\right)^2 \,, \tag{5.16}$$

with the parameter b usually taken to be $b = 100$. We define a metric for it by noticing it can be interpreted as a sum of two $N - 1$ dimensional Gaussian likelihoods

$$f_{\text{Rosenbrock}} \,\widehat{=}\, - \ln\left(\mathscr{G}(0|x_{i+1} - x_i^2, 2b^{-1}\mathbb{1})\right) - \ln\left(\mathscr{G}(1|x_i, 2\mathbb{1})\right) \,. \tag{5.17}$$

with "$\widehat{=}$ denoting equality up to irrelevant constants. By writing the Rosenbrock function in the form of Eq. (5.17) we can identify its metric with the Fisher metric of the likelihood.

Poisson log-normal

The objective function defined in this subsection is a variant of demo 2 of NIFTy5 [7]. The objective function is the posterior Hamiltonian of a log-normal process regression task using a Poisson likelihood with homogeneous exposure. The objective function can be defined as

$$f_{\text{Poisson}}(\xi_k) = \frac{1}{2}\xi_k^\dagger \xi_k + \lambda(\xi_k) - d^\dagger \ln\lambda(\xi_k) \tag{5.18}$$

$$\text{with } \lambda(\xi_k) = \mathbb{F}^{-1} \frac{1}{20 + k^2} \xi_k \,, \tag{5.19}$$

where $\mathbb{F}$ is the Fourier transform, $1/{20+k^2}$ is a falling square root power spectrum enforcing a differentiable field, and ξ_k are the latent degrees of freedom. The target of the Fourier transform is a one dimensional grid of length 1 with 1024 pixels. The data d is generated by sampling normal distributed ξ_k from which λ can be calculated. The data d is then drawn to be Poisson distributed with rate λ.

Radial tomography

The objective function defined in this subsection is a variant of demo 3 of NIFTy5 [7]. It can also be regarded as a simple variant of the models introduced in Chapters 3 and 4.

The objective function is the Kullback Leibler divergence to the posterior Hamiltonian of a Gaussian process regression task. This Gaussian regression task involves a Gaussian likelihood for radial integrations of positive field along several lines of sight. The Hamiltonian can be defined as

$$H_{\text{tomography}}(\xi) = \frac{1}{2}\xi^\dagger\xi + \frac{1}{2}(d - \theta(Rs))^\dagger N^{-1}(d - \theta(Rs)) \tag{5.20}$$

$$\text{with } s = \mathbb{F}^{-1}A(\xi_l)\xi_k \tag{5.21}$$

$$\text{and } \theta(x) = \frac{1}{\exp(-x) + 1} \ , \tag{5.22}$$

where $\mathbb{F}$ is the Fourier transform, $\xi = (\xi_k, \xi_l)^T$ are the latent degrees of freedom, R is the response which integrates radial lines of sight, $N = 0.001\,\mathbb{1}$ is the diagonal noise covariance, and $A(\xi_l)$ is a smooth square root power spectrum parameterized by the latent parameters ξ_l. The target of the Fourier transform is a $N \times N$ grid. The radial lines of sight are chosen with random target inside the reconstructed square and there are M lines. The data d is generated by sampling normal distributed ξ from which s can be calculated. The data d is then drawn to be independently normal distributed with mean $\theta(Rs)$ and variance 0.001. The overall objective function is then calculated by

$$f_{\text{tomography}}(m) = \frac{1}{10}\sum_{i=0}^{5}\left[H_{\text{tomography}}(m + \xi_i) + H_{\text{tomography}}(m - \xi_i)\right] \tag{5.23}$$

with ξ_i being independent Gaussian samples using the Bayesian Fisher metric of $H_{\text{tomography}}$ as covariance. The overall setup performs one iteration of the MGVI algorithm [54] with five mirrored samples on a two dimensional tomography inverse problem.

5.1.3 Results

We minimized all objective functions using both termination criteria for the conjugate gradient defined in Sec. 5.1.1. We terminated the minimization when

$$\left(\frac{\partial f}{\partial x}\right)^\dagger M^{-1}\left(\frac{\partial f}{\partial x}\right) < C \ , \tag{5.24}$$

with $C = 10^{-6}$ for $f_{\text{Rosenbrock}}$, $C = 10^{-4}$ for f_{Poisson} and $C = 10^{-2}$ for $f_{\text{tomography}}$. Table 5.1 shows a summary of the results. Overall, both termination criteria for the conjugate gradient perform similarly for the chosen problems.

Heuristic II is overall more stable under coordinate transformations. In all examples, both heuristics are within a factor 2 of each other in terms of number of model applications. Based on the here performed tests, no heuristic outperforms the other significantly, however heuristic II can be preferred due to its invariance under coordinate transformations.

	$f_{\text{Rosenbrock}}(x)$	$f_{\text{Rosenbrock}}(10^{-4}x)$	f_{Poisson}	$f^{(1)}_{\text{tomography}}$	$f^{(2)}_{\text{tomography}}$
heuristic I	5340	6896	916	2692	194866
heuristic II	6056	6056	626	3582	118606

Table 5.1: Calls to the objective function, its gradient, or metric during minimization. Calls to the metric are counted twice, they involve both the Jacobian of the model as well as its adjoint. The rows show the two tested heuristics for terminating the conjugate gradient. The columns are the different objective functions defined in Sec. 5.1.2. Hereby $f^{(1)}_{\text{tomography}}$ is defined to have $N = 128$ pixels per side and $M = 100$ lines of sight, whereas $f^{(2)}_{\text{tomography}}$ has $N = 512$ and $M = 10000$.

6 Conclusion

In this dissertation, two different applications of Bayesian reasoning on continuous fields are discussed. In the first part, we investigate simulation schemes constructed by a principle of minimal information loss. Contrary to common simulation schemes which require a discretization of the partial differential operators, discretization in information field dynamics requires the specification of what statements the simulation parameters make about a continuous field to be simulated. Based on this parameterized probability distribution, the simulation scheme that causes the least information loss per time step is uniquely defined. We investigate two different parametrizations and their resulting simulation schemes in simulating the Burger's equation. When comparing the found solutions to the results of finite difference schemes, we find an overall better performance when using the same amount of degrees of freedom. While the general direction of inferring solutions to PDEs using Bayesian reasoning seems promising, the approach that was investigated in this thesis has limited applicability. On the one hand, the increase in numerical performance comes at the cost of additional computations that have to be performed. Additionally, the choice of parametrization seems to play a fundamental role for the performance. The performance for an IFD method derived for one application might not translate to a different setup using the same differential equation, if the prior on the continuous field is not flexible enough. Choosing a different prior that might cope with the new situation enforces the user to re-calculate the updating rule for the new parameters. This derivation involves the analytical calculation of several integrals, a hurdle that makes the method not readily applicable in other contexts or for other differential equations. All these drawbacks could potentially be overcome by future work. The use of explicit matrices in the simulation is the largest drawback, as it enforces a quadratic scaling of the computational cost with the number of degrees of freedom. However, the cost of explicit matrices can often be overcome by a clever choice of basis, or by the use of sparse or implicit operators.

In the second part of this thesis, we present an example where such quadratic scaling was overcome by the use of the MGVI algorithm [54] in the context of inference. Hereby, a complex posterior distribution is approximated by a Gaussian with full covariance, that is represented through an implicit operator. This enables a fast inference with more accurate uncertainty estimates compared to methods that only involve diagonal covariance matrices [54]. We apply this algorithm to infer the dust extinction density in a vicinity of about $300\,\mathrm{pc}$ of the sun. A large scale validation example shows that the uncertainties that the algorithm imposes indeed represent the true uncertainty quite well, with the exception of a few outliers. Using data from the Gaia catalog, we reconstruct a three dimensional dust extinction map showing the nearby dust clouds. Our work contributes to the modeling of the magnetized Local Bubble [71]. However, our reconstruction seems to be artificially

void of dust. We attribute this effect to the likelihood, which is not very informing in regions with low but non-zero dust extinction. Furthermore, it only included Gaia data although further datasets were available.

We improve upon this first proof of concept in chapter 4, where we present an improved version of our algorithm and reconstruction, using a richer dataset. The main algorithmic improvements involve optimizations in runtime and parallelization. However, also improvements in the treatments of parallax uncertainty were achieved. The dataset used in this second reconstruction incorporates a larger amount of sources, and up to 3 times more precise dust extinction measurements. The resulting map covers a slightly larger area in vastly higher resolution, revealing nearby dust features in unprecedented resolution. By reconstructing the dust density in such detail, we pose a challenge to simulations of the interstellar medium to reproduce our results, or a possibility to tune the simulation hyper parameters such that they produce the dust statistics we observe for the Galaxy.

Finally, chapter 5 provides insights into the inner workings of the employed minimization procedure, discusses the Newton conjugate gradient algorithm and reflects upon its implementation in NIFTy. We discuss the tradeoff between conjugate gradient and Newton steps, and analyze two possible heuristics in several numerical examples. Our numerical results show that both heuristics show comparable performance for standard minimization benchmarks.

Bibliography

[1] Franco D Albareti, Carlos Allende Prieto, Andres Almeida, Friedrich Anders, Scott Anderson, Brett H Andrews, Alfonso Aragón-Salamanca, Maria Argudo-Fernández, Eric Armengaud, Eric Aubourg, et al. The 13th data release of the sloan digital sky survey: first spectroscopic data from the sdss-iv survey mapping nearby galaxies at apache point observatory. *The Astrophysical Journal Supplement Series*, 233(2):25, 2017.

[2] Shun-ichi Amari. Neural learning in structured parameter spaces-natural riemannian gradient. In *Advances in neural information processing systems*, pages 127–133, 1997.

[3] Friedrich Anders, Arman Khalatyan, C Chiappini, AB Queiroz, BX Santiago, C Jordi, L Girardi, AGA Brown, G Matijevič, G Monari, et al. Photo-astrometric distances, extinctions, and astrophysical parameters for gaia dr2 stars brighter than g= 18. *Astronomy & Astrophysics*, 628:A94, 2019.

[4] René Andrae, Morgan Fouesneau, Orlagh Creevey, Christophe Ordenovic, Nicolas Mary, Alexandru Burlacu, Laurence Chaoul, Anne Jean-Antoine-Piccolo, Georges Kordopatis, Andreas Korn, et al. Gaia data release 2-first stellar parameters from apsis. *Astronomy & Astrophysics*, 616:A8, 2018.

[5] Cedric Archambeau, Dan Cornford, Manfred Opper, and John Shawe-Taylor. Gaussian process approximations of stochastic differential equations. In *Gaussian Processes in Practice*, pages 1–16, 2007.

[6] Frédéric Arenou, Michel Grenon, and A Gomez. A tridimensional model of the galactic interstellar extinction. *Astronomy and Astrophysics*, 258:104–111, 1992.

[7] Philipp Arras, Mihai Baltac, Torsten A Ensslin, Philipp Frank, Sebastian Hutschenreuter, Jakob Knollmueller, Reimar Leike, Max-Niklas Newrzella, Lukas Platz, Martin Reinecke, et al. Nifty5: Numerical information field theory v5. *Astrophysics Source Code Library*, 2019.

[8] Philipp Arras, Philipp Frank, Reimar Leike, Rüdiger Westermann, and Torsten Enßlin. Unified radio interferometric calibration and imaging with joint uncertainty quantification. *arXiv preprint arXiv:1903.11169*, 2019.

[9] Philipp Arras, Philipp Frank, Philipp Haim, Jakob Knollmüller, Reimar Leike, Martin Reinecke, and Torsten Enßlin. The variable shadow of m87. *arXiv preprint arXiv:2002.05218*, 2020.

[10] José M Bernardo. Expected information as expected utility. *The Annals of Statistics*, pages 686–690, 1979.

[11] Lindy Blackburn, Dominic W Pesce, Michael D Johnson, Maciek Wielgus, Andrew A Chael, Pierre Christian, and Sheperd S Doeleman. Closure statistics in radio interferometric data. *arXiv preprint arXiv:1910.02062*, 2019.

[12] M Branicki and AJ Majda. Quantifying bayesian filter performance for turbulent dynamical systems through information theory. *Commun. Math. Sci*, 12(5):901–978, 2014.

[13] AGA Brown, A Vallenari, T Prusti, JHJ de Bruijne, C Babusiaux, CAL Bailer-Jones, Gaia Collaboration, et al. Gaia data release 2. summary of the contents and survey properties. *arXiv preprint arXiv:1804.09365*, 2018.

[14] Anthony GA Brown, A Vallenari, T Prusti, JHJ De Bruijne, F Mignard, R Drimmel, C Babusiaux, CAL Bailer-Jones, U Bastian, M Biermann, et al. Gaia data release 1-summary of the astrometric, photometric, and survey properties. *Astronomy & Astrophysics*, 595:A2, 2016.

[15] Marcelo GS Bruno. Sequential monte carlo methods for nonlinear discrete-time filtering. *Synthesis Lectures on Signal Processing*, 6(1):1–99, 2013.

[16] Richard S Bucy and Peter D Joseph. *Filtering for stochastic processes with applications to guidance*, volume 326. American Mathematical Soc., 1987.

[17] Dp Burstein and C Heiles. Hi, galaxy counts, and reddening-variation in the gas-to-dust ratio, the extinction at high galactic latitudes, and a new method for determining galactic reddening. *The Astrophysical Journal*, 225:40–55, 1978.

[18] BQ Chen, Y Huang, HB Yuan, C Wang, DW Fan, MS Xiang, HW Zhang, ZJ Tian, and XW Liu. Three-dimensional interstellar dust reddening maps of the galactic plane. *Monthly Notices of the Royal Astronomical Society*, 483(4):4277–4289, 2018.

[19] Yi-Kuan Chiang and Brice Ménard. Extragalactic imprints in galactic dust maps. *arXiv preprint arXiv:1808.03294*, 2018.

[20] J. Cockayne, C. Oates, T. Sullivan, and M. Girolami. Bayesian Probabilistic Numerical Methods. *ArXiv e-prints*, February 2017.

[21] Richard Courant and Peter Lax. On nonlinear partial differential equations with two independent variables. *Communications on pure and applied mathematics*, 2(2-3): 255–273, 1949.

[22] Harald Cramér. Mathematical methods of statistics. 1946. *Department of Mathematical SU*, 1946.

[23] Nando De Freitas, C Andrieu, Pedro Højen-Sørensen, M Niranjan, and A Gee. Sequential monte carlo methods for neural networks. In *Sequential Monte Carlo methods in practice*, pages 359–379. Springer, 2001.

[24] Persi Diaconis. Bayesian numerical analysis. *Statistical decision theory and related topics IV*, 1:163–175, 1988.

[25] Joseph Leo Doob. *Stochastic processes*, volume 101. New York Wiley, 1953.

[26] Arnaud Doucet. On sequential simulation-based methods for bayesian filtering. 1998.

[27] T. A. Enßlin. Information theory for fields. *ArXiv e-prints*, April 2018.

[28] T. A. Enßlin, M. Frommert, and F. S. Kitaura. Information field theory for cosmological perturbation reconstruction and nonlinear signal analysis. *Phys Rev D*, 80(10): 105005, November 2009. doi: 10.1103/PhysRevD.80.105005.

[29] Torsten A Enßlin. Information field dynamics for simulation scheme construction. *Physical Review E*, 87(1):013308, 2013.

[30] Torsten A Enßlin and Mona Frommert. Reconstruction of signals with unknown spectra in information field theory with parameter uncertainty. *Physical Review D*, 83(10):105014, 2011.

[31] Torsten A Enßlin and Cornelius Weig. Inference with minimal gibbs free energy in information field theory. *Physical Review E*, 82(5):051112, 2010.

[32] The EHT Collaboration et al. First m87 event horizon telescope results. i. the shadow of the supermassive black hole. *ApJL*, 875:1, 2019. URL `https://iopscience.iop.org/article/10.3847/2041-8213/ab0ec7`.

[33] The EHT Collaboration et al. First m87 event horizon telescope results. ii. array and instrumentation. *ApJL*, 875:2, 2019. URL `https://iopscience.iop.org/article/10.3847/2041-8213/ab0c96`.

[34] The EHT Collaboration et al. First m87 event horizon telescope results. iii. data processing and calibration. *ApJL*, 875:3, 2019. URL `https://iopscience.iop.org/article/10.3847/2041-8213/ab0c57`.

[35] The EHT Collaboration et al. First m87 event horizon telescope results. iv. imaging the central supermassive black hole. *ApJL*, 875:4, 2019. URL `https://iopscience.iop.org/article/10.3847/2041-8213/ab0e85`.

[36] The EHT Collaboration et al. First m87 event horizon telescope results. v. physical origin of the asymmetric ring. *ApJL*, 875:5, 2019. URL `https://iopscience.iop.org/article/10.3847/2041-8213/ab0f43`.

[37] The EHT Collaboration et al. First m87 event horizon telescope results. vi. the shadow and mass of the central black hole. *ApJL*, 875:6, 2019. URL `https://iopscience.iop.org/article/10.3847/2041-8213/ab1141`.

[38] Philipp Frank, Reimar Leike, and Torsten A Enßlin. Field dynamics inference for local and causal interactions. *arXiv preprint arXiv:1902.02624*, 2019.

[39] Gaia Collaboration et al. Description of the gaia mission (spacecraft, instruments, survey and measurement principles, and operations). *Gaia Collaboration et al.(2016a): Summary description of Gaia DR1*, 2016.

[40] B.G. Galerkin. On electrical circuits for the approximate solution of the laplace equation. *Vestnik Inzh.*, 19:897–908, 1915.

[41] Philipp Girichidis, Daniel Seifried, Thorsten Naab, Thomas Peters, Stefanie Walch, Richard Wünsch, Simon CO Glover, and Ralf S Klessen. The silcc project–v. the impact of magnetic fields on the chemistry and the formation of molecular clouds. *Monthly Notices of the Royal Astronomical Society*, 480(3):3511–3540, 2018.

[42] GA Gontcharov. 3d interstellar extinction map within the nearest kiloparsec. *Astronomy letters*, 38(2):87–100, 2012.

[43] George A Gontcharov. 3d stellar reddening map from 2mass photometry: An improved version. *Astronomy Letters*, 43(7):472–488, 2017.

[44] Gregory M Green, Edward F Schlafly, Douglas Finkbeiner, Hans-Walter Rix, Nicolas Martin, William Burgett, Peter W Draper, Heather Flewelling, Klaus Hodapp, Nicholas Kaiser, et al. Galactic reddening in 3d from stellar photometry–an improved map. *Monthly Notices of the Royal Astronomical Society*, 478(1):651–666, 2018.

[45] Gregory M. Green, Edward F. Schlafly, Catherine Zucker, Joshua S. Speagle, and Douglas P. Finkbeiner. A 3D Dust Map Based on Gaia, Pan-STARRS 1 and 2MASS. *arXiv e-prints*, art. arXiv:1905.02734, May 2019.

[46] Sebastian Haid, Stefanie Walch, Daniel Seifried, Richard Wuensch, Frantisek Dinnbier, and Thorsten Naab. Silcc-zoom: The early impact of ionizing radiation on forming molecular clouds. *Monthly Notices of the Royal Astronomical Society*, 482(3):4062–4083, 2018.

[47] JC Jackson. A critique of rees's theory of primordial gravitational radiation. *Monthly Notices of the Royal Astronomical Society*, 156(1):1P–5P, 1972.

[48] Andrew H Jazwinski. Mathematics in science and engineering. *Stochastic Processes and Filtering Theory*, 64, 1970.

[49] Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–. URL `http://www.scipy.org/`. [Online; accessed].

[50] Thomas Kailath. A general likelihood-ratio formula for random signals in gaussian noise. *IEEE Transactions on Information Theory*, 15(3):350–361, 1969.

[51] Nicholas Kaiser, Herve Aussel, Barry E Burke, Hans Boesgaard, Ken Chambers, Mark Richard Chun, James N Heasley, Klaus-Werner Hodapp, Bobby Hunt, Robert Jedicke, et al. Pan-starrs: a large synoptic survey telescope array. In *Survey and Other Telescope Technologies and Discoveries*, volume 4836, pages 154–165. International Society for Optics and Photonics, 2002.

[52] Hans Kersting and Philipp Hennig. Active uncertainty calibration in bayesian ode solvers. *arXiv preprint arXiv:1605.03364*, 2016.

[53] Jakob Knollmüller and Torsten A Enßlin. Encoding prior knowledge in the structure of the likelihood. *arXiv preprint arXiv:1812.04403*, 2018.

[54] Jakob Knollmüller and Torsten A Enßlin. Metric gaussian variational inference. *arXiv preprint arXiv:1901.11033*, 2019.

[55] A. Kostic, G. Edenhofer, L. Platz, M. Reinecke, P. Arras, P. Frank, P. Haim, R. H. Leike, R. Lemmerz, T. Enßlin, and V. Eberle. NIFTy 6: Open source scientific tools for Python, 2020. URL `https://gitlab.mpcdf.mpg.de/ift/nifty/-/tree/NIFTy_6`. [Online; accessed 22.05.2020].

[56] Alp Kucukelbir, Dustin Tran, Rajesh Ranganath, Andrew Gelman, and David M Blei. Automatic differentiation variational inference. *The Journal of Machine Learning Research*, 18(1):430–474, 2017.

[57] S. Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22 (1):79–86, 1951. doi: doi:10.1214/aoms/1177729694.

[58] R Lallement, L Capitanio, L Ruiz-Dern, C Danielski, C Babusiaux, JL Vergely, M Elyajouri, F Arenou, and N Leclerc. 3d maps of interstellar dust in the local arm: using *gaia*, 2mass and apogee-dr14. *arXiv preprint arXiv:1804.06060*, 2018.

[59] R Lallement, C Babusiaux, JL Vergely, D Katz, Frédéric Arenou, Bernard Valette, C Hottier, and L Capitanio. Gaia-2mass 3d maps of galactic interstellar dust within 3 kpc. *Astronomy & Astrophysics*, 625:A135, 2019.

[60] R. H. Leike and T. A. Enßlin. Operator calculus for information field theory. *Phys Rev E*, 94(5):053306, November 2016. doi: 10.1103/PhysRevE.94.053306.

[61] R. H. Leike, M. Glatzle, and T. A. Enßlin. Resolving nearby dust clouds. *ArXiv e-prints*, April 2020. URL `https://arxiv.org/abs/2004.06732`.

[62] Reimar H Leike and Torsten A Enßlin. Optimal belief approximation. *Entropy*, 19 (8):402, 2017.

[63] Reimar H Leike and Torsten A Enßlin. Towards information-optimal simulation of partial differential equations. *Physical Review E*, 97(3):033314, 2018.

[64] RH Leike and TA Enßlin. Charting nearby dust clouds using gaia data only. *Astronomy & Astrophysics*, 631:A32, 2019.

[65] J. C. Lemm. Bayesian Field Theory: Nonparametric Approaches to Density Estimation, Regression, Classification, and Inverse Quantum Problems. *ArXiv Physics e-prints*, December 1999.

[66] Xavier Luri, AGA Brown, LM Sarro, F Arenou, CAL Bailer-Jones, A Castro-Ginard, J de Bruijne, T Prusti, C Babusiaux, and HE Delgado. Gaia data release 2: using gaia parallaxes. *arXiv preprint arXiv:1804.09376*, 2018.

[67] Nasser M Nasrabadi. Pattern recognition and machine learning. *Journal of electronic imaging*, 16(4):049901, 2007.

[68] Jorge Nocedal and Stephen Wright. *Numerical optimization.* Springer Science & Business Media, 2006.

[69] GV Panopoulou, I Psaradaki, and K Tassis. The magnetic field and dust filaments in the polaris flare. *Monthly Notices of the Royal Astronomical Society*, 462(2):1517–1529, 2016.

[70] JEG Peek, Carl Heiles, Kathryn MG Peek, David M Meyer, and JT Lauroesch. The local leo cold cloud and new limits on a local hot bubble. *The Astrophysical Journal*, 735(2):129, 2011.

[71] V Pelgrims, K Ferrière, F Boulanger, R Lallement, and L Montier. Modeling the magnetized local bubble from dust data. *Astronomy & Astrophysics*, 636:A17, 2020.

[72] Lucky Puspitarini and Rosine Lallement. Distance to northern high-latitude hi shells. *Astronomy & Astrophysics*, 545:A21, 2012.

[73] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Inferring solutions of differential equations using noisy multi-fidelity data. *Journal of Computational Physics*, 335:736–746, 2017.

[74] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Numerical gaussian processes for time-dependent and non-linear partial differential equations. *arXiv preprint arXiv:1703.10230*, 2017.

[75] C Radhakrishna Rao. Minimum variance and the estimation of several parameters. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 43, 2, pages 280–283. Cambridge University Press, 1947.

[76] S Rezaei Kh, CAL Bailer-Jones, EF Schlafly, and M Fouesneau. Three-dimensional dust mapping in the orion complex, combining gaia-tgas, 2mass, and wise. *Astronomy & Astrophysics*, 616:A44, 2018.

[77] Sara Rezaei Kh, Coryn AL Bailer-Jones, Morgan Fouesneau, and Richard Hanson. Can we detect galactic spiral arms? 3d dust distribution in the milky way. *Proceedings of the International Astronomical Union*, 12(S330):189–192, 2017.

[78] Sara Rezaei Kh, Coryn AL Bailer-Jones, David W Hogg, and Mathias Schultheis. Detection of the milky way spiral arms in dust from 3d mapping. *Astronomy & Astrophysics*, 618:A168, 2018.

[79] Greg Ridgeway and David Madigan. A sequential monte carlo method for bayesian analysis of massive datasets. *Data Mining and Knowledge Discovery*, 7(3):301–319, 2003.

[80] HoHo Rosenbrock. An automatic method for finding the greatest or least value of a function. *The Computer Journal*, 3(3):175–184, 1960.

[81] SE Sale and J Magorrian. Large-scale three-dimensional gaussian process extinction mapping. *Monthly Notices of the Royal Astronomical Society*, 481(1):494–508, 2018.

[82] SE Sale, JE Drew, G Barentsen, HJ Farnhill, R Raddi, MJ Barlow, J Eislöffel, JS Vink, P Rodríguez-Gil, and NJ Wright. A 3d extinction map of the northern galactic plane based on iphas photometry. *Monthly Notices of the Royal Astronomical Society*, 443 (4):2907–2922, 2014.

[83] M. Selig. The NIFTY way of bayesian signal inference. In *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, volume these proceedings, 2013.

[84] Marco Selig, Valentina Vacca, Niels Oppermann, and Torsten A Enßlin. The denoised, deconvolved, and decomposed fermi γ-ray sky-an application of the d3po algorithm. *Astronomy & Astrophysics*, 581:A126, 2015.

[85] R Skalidis and V Pelgrims. Local bubble contribution to the 353-ghz dust polarized emission. *Astronomy & Astrophysics*, 631:L11, 2019.

[86] MF Skrutskie, RM Cutri, R Stiening, MD Weinberg, S Schneider, JM Carpenter, C Beichman, R Capps, T Chester, J Elias, et al. The two micron all sky survey (2mass). *The Astronomical Journal*, 131(2):1163, 2006.

[87] Martin Sparre, Christoph Pfrommer, and Mark Vogelsberger. The physics of multiphase gas flows: fragmentation of a radiatively cooling gas cloud in a hot wind. *Monthly Notices of the Royal Astronomical Society*, 482(4):5401–5421, 2019.

[88] Theo Steininger, Jait Dixit, Philipp Frank, Maksim Greiner, Sebastian Hutschenreuter, Jakob Knollmüller, Reimar Leike, Natalia Porqueres, Daniel Pumpe, Martin Reinecke, et al. Nifty 3-numerical information field theory-a python framework for multicomponent signal inference on hpc clusters. *arXiv preprint arXiv:1708.01073*, 2017.

[89] The Planck Collaboration et al. Planck 2018 results. i. overview and the cosmological legacy of planck. *arXiv preprint arXiv:1807.06205*, 2018.

[90] R Brent Tully and J Richard Fisher. Nearby small groups of galaxies. In *Symposium-International Astronomical Union*, volume 79, pages 31–47. Cambridge University Press, 1978.

[91] J-L Vergely, D Egret, R Freire Ferrero, B Valette, and J Koeppen. The extinction in the solar neighbourhood from the hipparcos data. In *Hipparcos-Venice'97*, volume 402, pages 603–606, 1997.

[92] J-L Vergely, R Freire Ferrero, A Siebert, and B Valette. Nai and hi 3-d density distribution in the solar neighbourhood. *Astronomy & Astrophysics*, 366(3):1016–1034, 2001.

[93] S Walch, P Girichidis, T Naab, A Gatto, SCO Glover, R Wünsch, RS Klessen, PC Clark, T Peters, D Derigs, et al. The silcc (simulating the lifecycle of molecular clouds) project–i. chemical evolution of the supernova-driven ism. *Monthly Notices of the Royal Astronomical Society*, 454(1):238–268, 2015.

[94] N. Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. New York: Wiley, 1949.

[95] Norbert Wiener. Generalized harmonic analysis. *Acta mathematica*, 55(1):117–258, 1930.

[96] Edward L Wright, Peter RM Eisenhardt, Amy K Mainzer, Michael E Ressler, Roc M Cutri, Thomas Jarrett, J Davy Kirkpatrick, Deborah Padgett, Robert S McMillan, Michael Skrutskie, et al. The wide-field infrared survey explorer (wise): mission description and initial on-orbit performance. *The Astronomical Journal*, 140(6):1868, 2010.

Acknowledgments