

UNIVERSITY OF CRETE

DOCTORAL THESIS

Reconstructing the Magnetic Field of the Milky Way via Astrophysical Techniques and Numerical Simulations

Author:
Alexandros TSOUROS

Supervisor:
Vasiliki PAVLIDOU

*A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy
in the*

Department of Physics
University of Crete

Declaration of Authorship

I, Alexandros TSOUROS, declare that this thesis titled, "Reconstructing the Magnetic Field of the Milky Way via Astrophysical Techniques and Numerical Simulations" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

“We demand rigidly defined areas of doubt and uncertainty!”

Vroomfondel, The Hitchhiker’s Guide to the Galaxy

UNIVERSITY OF CRETE

Abstract

Department of Physics

Doctor of Philosophy

Reconstructing the Magnetic Field of the Milky Way via Astrophysical Techniques and Numerical Simulations

by Alexandros TSOUROS

Ultrahigh-energy cosmic rays (UHECRs) are the most energetic particles observed in the universe, with energies above 10^{18} eV. Identifying their sources and understanding their production mechanisms could shed light on key questions in astrophysics and high-energy physics. However, as charged particles, UHECRs are deflected by the Galactic magnetic field (GMF), concealing their true sources on the plane of the sky (POS). Tracing UHECRs to their origins is further complicated by the uncertainties in the GMF's three-dimensional structure, as current observations are integrated along the line of sight (LOS). Forthcoming stellar polarisation surveys and Gaia stellar parallax data are expected to enable accurate nonparametric GMF modelling in 3D, with significant implications for UHECR source identification.

This thesis presents a Bayesian approach, based on information field theory, for reconstructing the 3D structure of the GMF in the local Galaxy. We benchmark this method with synthetic data that emulate expected observations, with increasing realism. By sampling the GMF's posterior distribution, we estimate the true arrival directions of UHECRs by backtracking them through sample configurations. Initial tests show that for weakly turbulent fields, UHECR arrival directions can be corrected to within approximately 3° , with substantial improvements even in highly turbulent fields.

We further test the framework on a ground-truth GMF configuration from a dynamo simulation of the Galactic interstellar medium, incorporating LOS-integrated data in addition to sparse local measurements. Our results indicate significant improvements in UHECR arrival direction estimates, particularly for rigidities around 10^{19} eV. The inclusion of LOS-integrated data further reduces maximum errors, especially in complex regions of the sky.

Finally, we assess whether the absence of directional information in polarimetric data can be compensated by prior knowledge from parametric GMF models. For proton primaries, deflections can be corrected with residual biases around a degree, while intermediate-charge nuclei, such as nitrogen, are localised within tens of degrees. The arrival directions of highly charged nuclei, like iron, remain difficult to recover due to extensive diffusion. For weaker GMFs, substantial corrections are achievable across large portions of the sky, with regions of minimal deflection identified for optimal results.

This thesis establishes a rigorous framework for GMF reconstruction from sparse local data, significantly enhancing UHECR source localisation. Future applications to observational data from the PASIPHAE survey and complementary LOS measurements promise to refine our understanding of the GMF's 3D structure and support UHECR source identification efforts.

Acknowledgements

This dissertation has only been possible thanks to the support, guidance, and encouragement I received from many remarkable individuals. I am deeply appreciative of each person who contributed to this journey, helping to make this work a reality.

First and foremost, I extend my heartfelt gratitude to my supervisor and mentor, Professor Vasiliki Pavlidou. Her influence has uniquely shaped my scientific outlook but also my development as an individual. Her steadfast guidance and encouragement have been transformative, and I am profoundly grateful for her belief in me and her unwavering support.

I am also exceptionally grateful to Professor Torsten Enßlin, whose contributions have been instrumental in guiding and advancing this work. His insights and encouragement, particularly during pivotal stages of the project, were invaluable. I extend special thanks to his students, Gordian Edenhofer and Phillip Frank, whose patience and expertise helped me navigate the foundational concepts of this thesis. Their generosity in sharing their knowledge was crucial to this work.

Special thanks go to Vincent Pelgrims, whose technical support and patience during the initial stages of this project provided essential guidance. His presence was a steady and invaluable source of insight that helped lay the groundwork for this thesis.

I owe a particular debt of gratitude to Professor Konstantinos Tassis, whose invaluable insights and mentorship have had a lasting influence on my academic path, from my early undergraduate years to the culmination of my PhD. His support has been instrumental in shaping my approach to science and my career.

I would also like to acknowledge Professor Nikolaos Kylafis, the first professor who recognized my potential as an undergraduate and inspired me to pursue a career in physics. His early mentorship was foundational to my journey, providing me with the confidence to take my first steps into scientific research.

I am fortunate to have collaborated with many exceptional colleagues, past and present, who each brought unique contributions to this work. I am deeply grateful to Paolo Bonfini, Tanio Diaz Santos, Konstantinos Kovlakas, Evangelia Ntormousi, Georgia Panopoulou, and Andreas Zezas. Whether through direct scientific contributions or their invaluable support, each of them played an important role in the progression of this research.

The companionship and support of my office mates have enriched this journey in countless ways: Álvarez-Ortega Diego, Román Fernández Aranda, Anna Konstantinou, Georgios Korkidis, Maria Kopsachili, Elias Kyritsis, Lofaro Cristina Maria, Nikos Mandarakas, Raphael Skalidis, and Michalis Xigkis. Their camaraderie and our stimulating discussions provided both inspiration and respite, offering motivation during the most challenging phases of this PhD—even if usually the temptation of just chatting the day away was too great.

My deepest thanks go to my family and friends, whose love and support have been my foundation. Their belief in me has been a source of strength and inspiration throughout this journey.

In particular, I thank my mother, Louise Maalouf, and my stepfather, Gaudenzio Zonca, for their unending encouragement and understanding over the years. I am also profoundly grateful to my sister, Stella Plyta, and my brother, Giannis Plytas, for their steadfast support and companionship.

I honor the memory of my late father, Pavlos Tsouros, who inspired my love for physics from a young age and supported me wholeheartedly on this path. This thesis is dedicated to his memory.

Finally, I would like to acknowledge support by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “Third Call for H.F.R.I. Scholarships for PhD Candidates” (Project 5332).

Contents

Declaration of Authorship	iii
Abstract	vii
Acknowledgements	ix
1 Ultrahigh Energy Cosmic Rays	1
1.1 Introduction	1
1.2 Historical Overview	2
1.3 Modern Detection Methods and Challenges	3
1.4 UHECR Spectrum	4
1.5 UHECR composition	6
1.6 Observed Anisotropies in UHECR Arrival Directions	8
1.6.1 Dipole anisotropy	9
1.6.2 Small- and Intermediate-scale Anisotropies	9
1.6.3 Implications for UHECR Source Identification	9
1.7 UHECRs and Magnetic fields	10
1.7.1 The Intergalactic Magnetic Field	10
1.7.2 Galactic Magnetic Field	11
1.7.3 Parametric Models of the GMF	12
1.7.4 Prospects for Improved GMF Mapping	13
2 Bayesian Statistics and Information theory	17
2.1 Classical Logic and Its Limitations	17
2.2 From Boolean Algebra to Probability	18
2.2.1 Introduction to Plausibility	18
2.2.2 Marginalization	19
2.3 Example: One-Dimensional Inference	20
2.4 A primer on information and entropy	22
2.5 Example: Multi-Dimensional Inference	24
2.5.1 Prior Distribution	25
2.5.2 Likelihood	25
2.5.3 Posterior Distribution	25
3 Field theoretic approach to the GMF inverse problem	27
3.1 Statement of the Problem	27
3.2 Field Theoretical Approach	28
3.2.1 The Likelihood	30
3.2.2 The Prior	31
Interlude: A Primer on L_2 -Regularization	31
3.2.3 Maximum Entropy Prior	31
3.3 Calculation of the Partition Functional : The Effective Current Density	33

4 Nonparametric Bayesian Reconstruction of the GMF Using IFT for UHECR Backtracking (Part 1)	37
5 Nonparametric Bayesian Reconstruction of the GMF Using IFT for UHECR Backtracking (Part 2)	51
6 The Effect of Heavier Primary Composition	67
7 Conclusions	79
Bibliography	81

List of Abbreviations

CMB	Cosmic Microwave Background
IGMF	InterGalactic Magnetic Field
GMF	Galactic Magnetic Field
EAS	Extensive Air Shower
PAO	Pierre Auger Observatory
TA	Telescope Array
LHC	Large Hadron Collider
FD	Fluorescence Detectors
SD	Surface Detectors
AERA	Auger Enhanced Radio Array
HEAT	High-Elevation Auger Telescope
KASCADE	Karlsruhe Shower Core and Detector Electronics
SKA	Square Kilometre Array
PASIPHAE	Polar-Areas Stellar Imaging in Polarization High-Accuracy Experiment
KL	Kullback-Leibler

*Dedicated to the memory of my father, Pavlos Tsouros, who taught me
to love physics*

Chapter 1

Ultrahigh Energy Cosmic Rays

1.1 Introduction

Cosmic rays are charged particles, primarily protons and atomic nuclei, that travel through space at nearly the speed of light. Among them, *ultrahigh-energy cosmic rays (UHECRs)* represent the most energetic particles known in the universe, with energies exceeding 10^{18} eV (1 EeV). These particles carry energy far beyond what can be achieved in any human-made accelerator, making their study an essential part of high-energy astrophysics (Kotera and Olinto, 2011; Olinto, 2012). However, despite nearly a century of research since their discovery, UHECRs remain one of the most intriguing and enigmatic topics in the field of astrophysics (Cronin, 2005).

The origins of UHECRs, as well as the mechanisms by which they achieve such extreme energies, are still a subject of significant debate (Hillas, 1984; Allard and Protheroe, 2009). The detection of UHECRs presents numerous challenges, as they are rare—arriving at Earth at a rate of about 1 particle per square kilometer per year for energies above 1 EeV, and even more infrequent at higher energies (Nagano and Watson, 2000). Additionally, UHECRs are charged particles, which means their paths are bent by intergalactic and Galactic magnetic fields. This deflection complicates the task of tracing their trajectories back to their sources, unlike neutral messengers like photons or neutrinos, which can travel in straight lines through space (Beck, 2001; Stanev, 1997).

The UHECR energy spectrum provides key insights into the physical processes at play. It shows a steepening at the highest energies, known as the *Greisen-Zatsepin-Kuzmin (GZK) cutoff*, which occurs as UHECRs interact with cosmic microwave background (CMB) photons, losing energy in the process (Greisen, 1966; Zatsepin and Kuz'min, 1966). Understanding whether this steepening is due to interactions with the CMB or the intrinsic energy limits of the sources is central to UHECR research (Berezinsky, Gazizov, and Grigorieva, 2006). The composition of UHECRs at these energies—whether they consist primarily of protons, heavier nuclei, or a mix—holds critical clues about their origins and the conditions within their acceleration environments (Kampert and Unger, 2012).

UHECRs are detected indirectly through the extensive air showers (EAS) they produce when they collide with particles in the Earth's atmosphere. Instruments like the *Pierre Auger Observatory* (The Pierre Auger Collaboration, 2015) (PAO) and the *Telescope Array* (TA) (Kawai et al., 2008a) use a combination of surface detectors and fluorescence detectors to observe these showers and infer key properties of the primary UHECRs, including their energy, arrival direction, and mass composition. However, these measurements are complicated by uncertainties in the underlying physics, such as the interaction models used to simulate the air showers and the limited event statistics at the highest energies (Alves Batista et al., 2019).

The study of UHECRs can provide a window into the most extreme environments in the universe, including regions around black holes, active galactic nuclei (AGN), and gamma-ray bursts (GRBs) (Murase et al., 2012; Kotera and Olinto, 2011). These particles offer a direct

probe of astrophysical conditions at energies far beyond those achievable in particle accelerators like the Large Hadron Collider (LHC), potentially uncovering new physics in the high-energy regime (Waxman, 2005; Romanopoulos, Pavlidou, and Tomaras, 2022a; Romanopoulos, Pavlidou, and Tomaras, 2022b).

Moreover, UHECRs play a crucial role in the broader context of *multimessenger astronomy*, where cosmic rays, photons, neutrinos, and gravitational waves are observed together to provide a more complete picture of energetic cosmic phenomena (Aartsen et al., 2018; Bartos and Kowalski, 2017). Understanding the origin and behavior of UHECRs can complement our knowledge of other messengers and help address fundamental questions about the nature of cosmic accelerators (Ahlers and Halzen, 2018).

In this chapter, we will explore the key features of UHECRs, their detection methods, phenomenological features, and challenges concerning identification of their sources. We will also briefly discuss the magnetic fields that influence their path, and specifically we will discuss how the Galactic magnetic field (GMF) can be probed, especially as part of the effort to correct for its effect in the observed UHECR arrival directions, thus arriving at the main problem that will concern us in this thesis; namely, the nonparametric statistical reconstruction of the local GMF from sparse and local data.

1.2 Historical Overview

The history of cosmic rays began in 1912 when Austrian physicist Victor Hess made a groundbreaking discovery during a series of balloon ascents. As he rose to altitudes over 5,000 meters, Hess measured ionizing radiation levels that significantly increased with height, contradicting the expectation that radiation would decrease with altitude. This observation led to the realization that the radiation was coming from outer space, and Hess named it "cosmic radiation" (Hess, 1912). For his discovery, Hess was awarded the Nobel Prize in Physics in 1936, and his findings laid the foundation for what would become the field of cosmic ray research.

Throughout the 20th century, cosmic rays were studied extensively, leading to the classification of different energy regimes. In the 1930s and 1940s, cosmic rays were identified as primarily high-energy protons and atomic nuclei. Researchers began to measure cosmic rays in the atmosphere using ground-based detectors, discovering that they ranged from a few GeV to much higher energies. The field entered a new era in 1962 when the first UHECR with an energy exceeding 10^{20} eV was detected by the Volcano Ranch experiment (Linsley, 1963). This marked the discovery of a class of cosmic rays that were far more energetic than previously observed, raising fundamental questions about their origin and the astrophysical processes capable of accelerating particles to such extreme energies.

A major milestone in cosmic ray research came in 1991 with the detection of the "Oh-My-God" (OMG) particle, recorded by the Fly's Eye detector in Utah. This particle had an energy of approximately 3×10^{20} eV, the highest energy ever recorded for a cosmic ray (Bird et al., 1995). Its discovery stunned the scientific community, as the energy of this particle was orders of magnitude beyond what could be produced in man-made particle accelerators. The OMG particle highlighted the need to understand the extreme astrophysical environments where such high-energy particles could be produced.

These discoveries led to the establishment of large-scale cosmic ray observatories aimed at understanding UHECRs. In 2004, PAO in Argentina began operations, becoming the world's largest facility dedicated to UHECR detection (The Pierre Auger Collaboration, 2015). TA, located in Utah, followed in 2008, designed to complement Auger's efforts by covering the northern hemisphere (Kawai et al., 2008a). These observatories provided the first statistically significant datasets of UHECR events and led to crucial insights into the energy spectrum, mass composition, and anisotropy of UHECRs.

Despite these advances, the sources of UHECRs remain elusive. Magnetic fields in space deflect the paths of these charged particles, making it difficult to trace them back to their origins. Furthermore, the mechanisms by which UHECRs are accelerated to such extreme energies—whether in AGN, GRBs, or other extreme environments—are still debated (Kotera and Olinto, 2011).

The discovery of UHECRs has not only challenged our understanding of high-energy astrophysics but also pushed the boundaries of particle physics. With energies far exceeding those of terrestrial accelerators, UHECRs provide a unique opportunity to study particle interactions at energies unreachable by current technology.

1.3 Modern Detection Methods and Challenges

UHECRs are detected indirectly through the extensive air showers (EAS) they produce when they interact with the Earth’s atmosphere. When a UHECR strikes an atmospheric particle, it initiates a cascade of secondary particles that spread out over a large area as they travel toward the Earth’s surface. These air showers contain billions of particles, including electrons, muons, and hadrons, which can be detected by ground-based instruments. There are two primary methods for detecting UHECRs: fluorescence detectors (FD) and surface detectors (SD), both of which provide complementary information about the properties of the primary UHECR. Collectively, data from these measurements provide insights into the energy, arrival direction, as well as composition of the primary UHECR (Ulrich, Engel, and Unger, 2011).

- **Fluorescence Detectors (FDs):** Fluorescence is the process by which atoms absorb photons of one energy and emit photons at a smaller energy. FDs detect the ultraviolet (UV) light emitted as charged particles in an extensive air shower (EAS) excite nitrogen molecules in the atmosphere. FD observations are limited by their operational constraints, as they can only operate during clear, moonless nights to avoid background noise from the moon’s light and atmospheric conditions. As a result, FDs have a duty cycle of roughly 10-15%, reducing the already limited UHECR detection rate at the highest energies (Abraham et al., 2010). The fluorescence telescopes function in stereoscopic mode, allowing for the determination of a three-dimensional view of the air shower. When an air shower is observed by two telescopes positioned at different sites, each telescope captures a plane of the shower’s trajectory. The intersection of these two planes enables the precise calculation of the shower’s arrival direction, which can be traced back to the arrival direction of the primary cosmic ray (Tokuno et al., 2012).
- **Surface Detectors (SDs):** SDs complement FD measurements by detecting the secondary particles that reach the ground after the EAS develops. These detectors are spread out over large areas and can operate continuously, regardless of weather or light conditions (The Pierre Auger Collaboration, 2015; Kawai et al., 2008b). SDs measure the lateral distribution of particles from the shower, which can be used to estimate the energy and arrival direction of the primary cosmic ray.

The combination of FD and SD detectors are used in the main two UHECR observatories currently in operation, the PAO and the TA project.

The PAO covers an area of approximately 3,000 square kilometers, and it employs a hybrid detection system, consisting of both SDs and FDs, to observe UHECRs. The observatory is designed to detect cosmic rays with energies above 10^{18} eV, focusing on the southern hemisphere (The Pierre Auger Collaboration, 2015). In the right panel of Figure 1.1 PAO’s layout is shown. Additional components of PAO include the Auger Engineering Radio Array (AERA), designed to detect the radio emission produced by cosmic ray air showers in the Earth’s atmosphere. It

operates in the frequency range of 30 to 80 MHz, providing complementary information about the properties of extensive air showers. The High Elevation Auger Telescopes (HEAT) are three additional fluorescence telescopes located at a higher elevation, designed to extend the observatory's sensitivity to lower-energy cosmic rays (down to approximately 10^{17} eV). HEAT allows for the study of the transition region between Galactic and extragalactic cosmic rays.

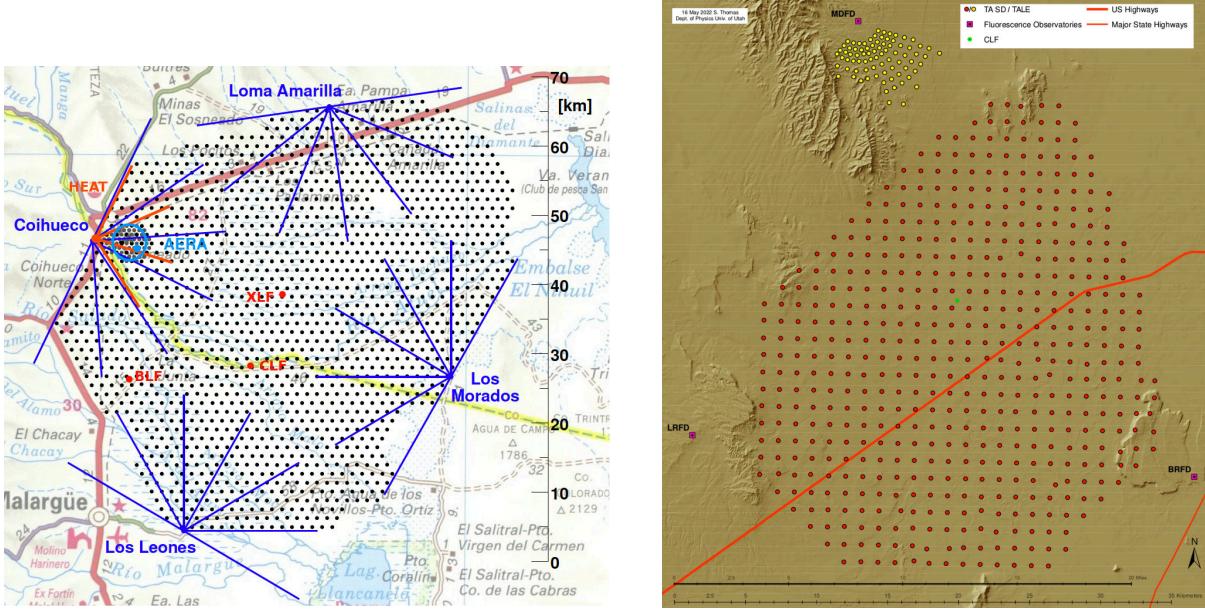


FIGURE 1.1: Left panel: Pierre Auger Observatory layout. Black dots correspond to the SD stations. Blue lines denote the field of view of the four FDs. In addition, there are three high elevation (HEAT) telescopes (red), which provide the low-energy extensions to the setup. The extent of the AERA radio array is also shown. Right panel: Telescope Array layout. The SD stations are shown as black dots while the FDs are shown in magenta. Images taken from PAO and TA websites.

TA, located in Utah, USA, is the largest UHECR observatory in the northern hemisphere, covering an area of about 700 square kilometers. Like PAO, TA employs a hybrid detection method using both Surface Detectors (SD) and Fluorescence Detectors (FD) to measure UHECRs. Its location allows TA to complement the observations made by PAO, providing coverage of the northern hemisphere (Kawai et al., 2008a). In the left panel of Figure 1.1, the basic layout of the TA is shown.

1.4 UHECR Spectrum

In Figure 1.2 the differential flux (multiplied by E^3) of UHECRs is shown as a function of energy, as observed by both PAO and TA.

The spectrum can be seen to adhere to a broken power-law structure, where the breaking points occur roughly at $E \sim 10^{15}$ eV (called the *knee*), at $E \sim 10^{18.5}$ eV (the *ankle*), and finally above $E \sim 10^{19.5}$ eV there is a strong suppression, called the *Greisen-Zatsepin-Kuzmin (GZK) cutoff*.

- **The knee:** First observed in the 1950s by Kulikov and Khristiansen (Kulikov and Khristiansen, 1959), this feature signals a transition in the dominant population of cosmic ray particles, marking a steepening of the spectrum. The knee is generally understood to reflect the upper energy limit for Galactic cosmic ray sources or the maximum energy for which the Galaxy's magnetic field can confine particles (Guido, 2021).

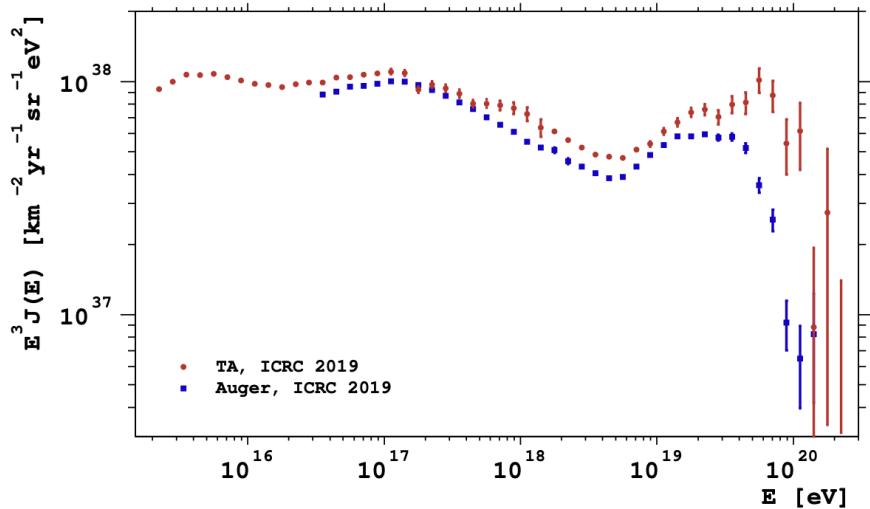


FIGURE 1.2: The differential flux of UHECRs as a function of energy, showing data from the PAO and TA. The steepening at $10^{19.5}$ eV is indicative of the GZK cutoff. Adapted from (Deligny, 2019).

The knee is believed to arise due to the energy-dependent rigidity of cosmic rays, where lighter elements such as protons begin to escape the Galactic magnetic field at lower energies compared to heavier nuclei like iron. This energy dependence is expressed by the relationship:

$$E_{\text{knee}} \sim Z \times 3 \times 10^{15} \text{ eV}$$

where Z is the atomic number of the particle. Hence, protons dominate the flux up to the knee energy, while heavier nuclei contribute more significantly at higher energies (Kotera and Olinto, 2011). Consequently, the spectral steepening observed at the knee is linked to the mass composition of cosmic rays, transitioning from lighter to heavier nuclei.

- **Second knee:** At energies around $E \approx 10^{17}$ eV, a second knee is observed in the spectrum, signaling the further steepening of the cosmic ray flux. This second knee is often interpreted as the energy limit for the heaviest nuclei to be confined by the Galactic magnetic field. Beyond this energy, the contribution from extragalactic sources becomes more prominent, initiating a gradual transition to an extragalactic cosmic ray population (Coleman et al., 2023).

Recent experiments, such as KASCADE-Grande, have provided detailed measurements of cosmic ray composition across this energy range, offering insights into the knee and second knee features. KASCADE-Grande observed a feature close to the expected iron knee at $E \approx 10^{16.92}$ eV, followed by the possible onset of extragalactic protons at $E \approx 10^{17.08}$ eV (Guido, 2021). These observations further support the idea that the knee reflects the end of Galactic cosmic ray confinement for lighter elements, while the second knee marks the corresponding limit for heavier nuclei.

The interpretation of the knee and second knee is crucial for understanding the transition between Galactic and extragalactic cosmic rays. Various models suggest that Galactic sources, such as supernova remnants, are responsible for accelerating particles up to energies around the knee. Beyond this energy, extragalactic sources, potentially AGN or GRBs, are thought to contribute to the flux observed at Earth (Kotera and Olinto, 2011).

- **The ankle:** The hardening of the spectrum at $E \sim 10^{18.5}$ eV feature in the UHECR energy spectrum can be explained by two primary models, each based on different assumptions

about the composition of the extragalactic cosmic rays. In one scenario, the ankle is interpreted as a result of electron-positron pair production during the interactions of protons with photons from the Cosmic Microwave Background (CMB), commonly referred to as the *dip model* (Berezinsky and Grigoreva, 1988; Berezinsky, Gazizov, and Grigorieva, 2006). This model assumes that protons dominate the extragalactic component around energies of approximately 10^{18} eV.

Alternatively, if the extragalactic cosmic rays are primarily composed of heavier nuclei, the ankle can be explained as a transition point from a Galactic to an extragalactic cosmic ray population (Kotera and Olinto, 2011; Guido, 2021). In this case, the ankle feature signals the energy at which cosmic rays from Galactic sources taper off, and cosmic rays from extragalactic sources begin to dominate. The precise nature of this transition, and the mass composition of UHECRs at these energies, is a key factor in distinguishing between these two theoretical models.

Similarly, at higher energies, a flux suppression above $\sim 4 \times 10^{19}$ eV has been observed by both the PAO and the TA (The Pierre Auger Collaboration, 2015; Kawai et al., 2008a). This suppression is still under debate, but it is generally interpreted as either a signature of pion production from the interactions of protons with CMB photons—akin to the GZK process—or as the maximum energy limit of cosmic ray acceleration at extragalactic sources (Guido, 2021). It is also possible that this suppression results from an interplay between both effects (Kotera and Olinto, 2011). As with the ankle, understanding the suppression requires a detailed knowledge of UHECR mass composition, which is essential for discriminating between these different models of origin and propagation.

- **The GZK cutoff:** This feature occurs due to the interaction of UHECRs with the cosmic microwave background (CMB) radiation (Greisen, 1966; Zatsepin and Kuz'min, 1966). For example, at energies above ~ 50 EeV, protons can interact with CMB photons, leading to the production of pions. This interaction results in significant energy loss for the protons as they propagate over cosmological distances, causing a suppression in the observed flux of UHECRs at Earth. This process sets an effective distance limit for UHECR sources, often referred to as the *GZK horizon*, which is defined as the distance above which 10% or less of the original population of a given energy survives. At energies above $E \sim 60$ EeV, only protons and iron nuclei can survive propagation over distances greater than $D \sim 50$ Mpc (see Figure 1.3).

1.5 UHECR composition

The main method to probe the UHECR composition involves inferring the atmospheric depth where the energy deposit profile of secondary particles from extensive air showers reaches its maximum, X_{\max} (see Figure 1.4), which is directly measured by FDs. This is related to the particle charge via an inference of its mass through the cross-section, which is extrapolated from the Standard Model to the observed energies, or calculated from alternative theories at those energies (Gaisser and Hillas, 1977; Mayotte et al., 2023; Abbasi et al., 2019; Arsene, 2021; Salamida, 2023; Mayotte et al., 2023; Romanopoulos, Pavlidou, and Tomaras, 2022a; Romanopoulos, Pavlidou, and Tomaras, 2022b)). Lighter particles, for example protons, penetrate deeper into the atmosphere before initiating a shower, while heavier nuclei initiate showers higher up. The uncertainties in hadronic interaction models at super-LHC energies, which describe how primary cosmic rays interact with the atmosphere, introduce significant ambiguity in inferring X_{\max} from the data measurements (Berezinsky, 2014a; Berezinsky, 2014b). In Figure 1.5 the first two moments of the observed X_{\max} is shown

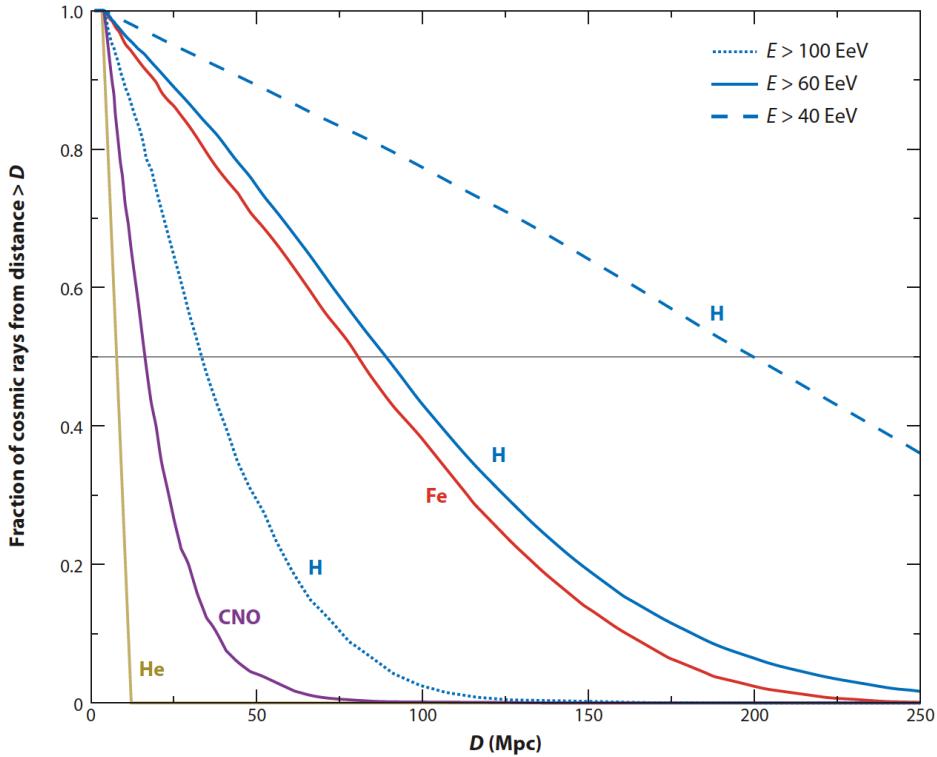


FIGURE 1.3: The fraction of cosmic rays that survive propagation over a distance greater than D is shown for protons with energies above 40, 60, and 100 EeV, as well as for helium, CNO, and iron nuclei with energies above 60 EeV. The gray solid line indicates the distance where 50% of a given species can originate for a specific atomic mass and energy. Figure from (Kotera and Olinto, 2011).

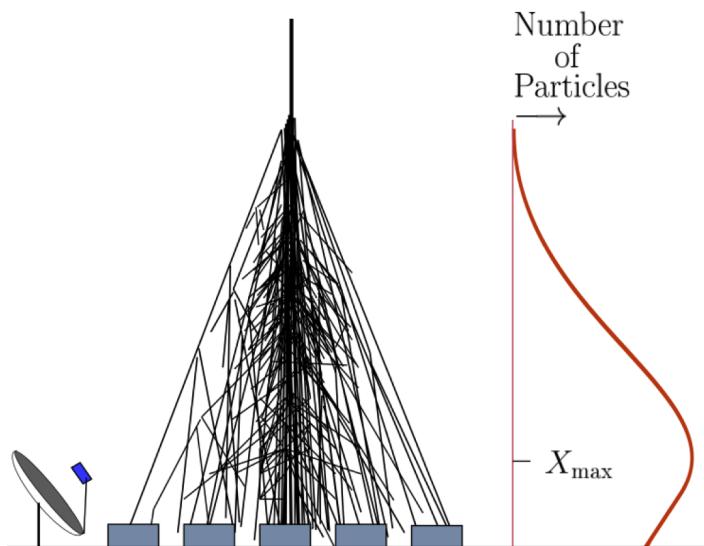


FIGURE 1.4: Definition of X_{\max} . The primary particle interacts with a particle in the Earth's atmosphere, which produces a cascade, called an extensive air shower. The atmospheric depth at which the energy deposited is maximum, is X_{\max} . Figure from Anchordoqui, 2011.

Moreover, the low flux of UHECRs at the highest energies (especially beyond the GZK cutoff) further complicates mass composition measurements. Since only a small number of UHECRs are detected, statistical uncertainties play a significant role in limiting the conclusions that can be drawn about the primary composition (Kampert and Unger, 2012). In Figure 1.5, the first two moments of the PAO measurements of X_{\max} are shown. In addition, the theoretical predictions for a pure proton and pure iron composition are shown, as calculated from Monte Carlo simulations for three different particle physics models (Salamida, 2023). Notice that for smaller energies the X_{\max} measurements agree with the theoretical predictions for a proton composition, and shift towards iron for larger energies.

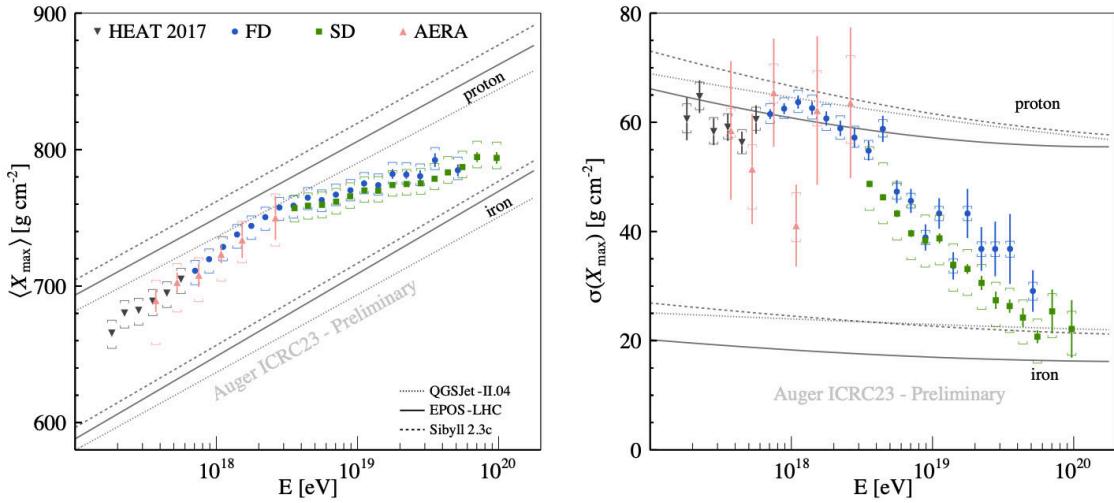


FIGURE 1.5: The first two moments of the X_{\max} posterior distributions as a function of energy, as measured by the PAO. Monte Carlo simulation predictions are shown using different particle physics models for a primary proton and iron nucleus (Figure from Salamida, 2023).

Recent technological advances aim to mitigate these challenges. For instance, space-based UHECR detection experiments like JEM-EUSO (Inoue, Miyazawa, and Kawasaki, 2009) are being developed to capture a significantly larger number of events by observing extensive air showers from space. This approach offers the advantage of wider coverage, which could improve the statistical significance of the highest energy UHECR observations (Guido, 2021). Additionally, innovations in machine learning and artificial intelligence models, applied to UHECR data, are beginning to provide more accurate interpretations of air shower development and X_{\max} measurements (Coleman et al., 2023).

1.6 Observed Anisotropies in UHECR Arrival Directions

The study of anisotropies in UHECR arrival directions is crucial for understanding the sources and propagation of UHECRs. While cosmic rays at lower energies tend to arrive isotropically due to strong deflections by Galactic magnetic fields, UHECRs at higher energies are expected to exhibit anisotropic patterns that can reveal the locations of their astrophysical sources. Both the PAO and TA have made significant contributions to the detection and characterization of UHECR anisotropies.

1.6.1 Dipole anisotropy

One of the key findings in recent years is the observation of large-scale anisotropy dipolar in the arrival directions of UHECRs with energies exceeding 8 EeV. This anisotropy, first detected by PAO, has been observed with a significance larger than 6σ . The data from PAO show that the arrival directions of these cosmic rays exhibit a dipolar pattern, with an excess pointing approximately 125 degrees away from the Galactic center (Kotera and Olinto, 2011; Guido, 2021). This suggests that the majority of UHECRs at these energies are extragalactic in origin, as the anisotropy aligns with regions outside the Milky Way.

TA has also reported large-scale anisotropies in UHECR arrival directions, albeit with some differences compared to the results from PAO. Both observatories contribute to the growing body of evidence supporting the extragalactic nature of UHECRs, particularly at the highest energies. The increase in dipole amplitude with increasing energy suggests that the UHECR flux is dominated by extragalactic sources beyond a certain threshold energy (Coleman et al., 2023). This large-scale anisotropy is a crucial step towards identifying the sources of UHECRs and understanding their propagation through cosmic magnetic fields.

1.6.2 Small- and Intermediate-scale Anisotropies

In addition to large-scale patterns, small- and intermediate-scale anisotropies have been observed in the UHECR data, particularly in the direction of known astrophysical objects. For example, PAO has detected a significant excess of UHECRs in the region surrounding Centaurus A, one of the nearest AGN (Guido, 2021). This excess is particularly pronounced for cosmic rays with energies above 40 EeV, where deflections by the Galactic magnetic field are expected to be smaller, making source identification more feasible. The detection of such intermediate-scale anisotropies provides strong evidence linking UHECRs to specific extragalactic sources.

TA has similarly reported an excess of UHECR events in the northern hemisphere, often referred to as the "TA hot spot". This excess is centered near the supergalactic plane, suggesting a potential association with nearby extragalactic structures (Coleman et al., 2023). These small-scale anisotropies are critical for narrowing down the potential sources of UHECRs, although the exact identification of sources remains challenging due to deflections by both Galactic and intergalactic magnetic fields.

Figure 1.6 shows the Mollweide projection of the arrival directions of UHECRs from both PAO and TA. The cosmic-ray flux is represented in equatorial coordinates, and the map is smoothed using a 45-degree top-hat function to highlight the large-scale anisotropies. The map reveals the dipolar structure observed in the southern hemisphere by PAO, as well as the smaller-scale excesses detected by TA and PAO in the northern and southern hemispheres respectively. These features provide a clear visual representation of the large- and intermediate-scale anisotropies detected in UHECR data.

1.6.3 Implications for UHECR Source Identification

The observed anisotropies in UHECR arrival directions are crucial for constraining the potential sources of these particles, and by extension the mechanisms responsible for their production. Large-scale anisotropies, particularly the dipolar pattern observed by PAO, suggest that UHECRs at the highest energies originate from extragalactic sources. The alignment of the dipole with regions outside the Galactic center further supports this

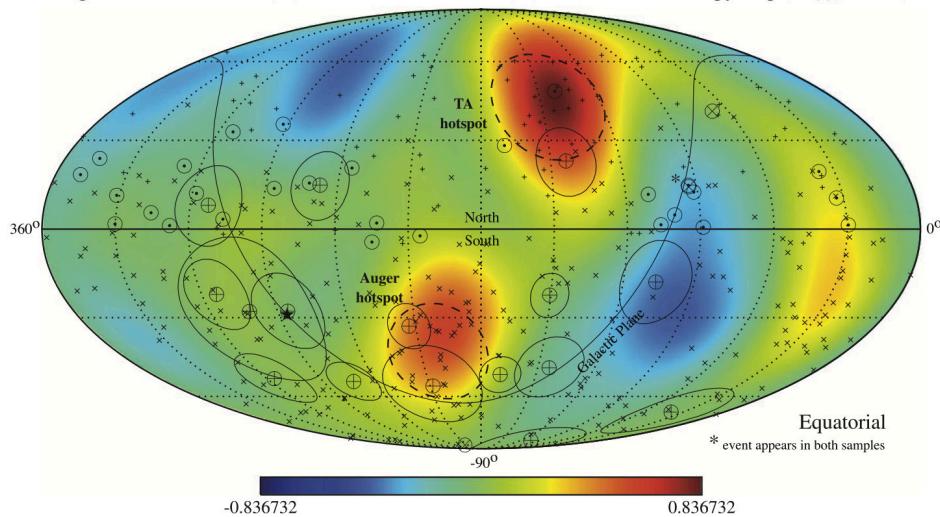


FIGURE 1.6: Mollweide projection of UHECR arrival directions from the PAO and TA. The cosmic-ray flux is shown in equatorial coordinates, smoothed with a 45-degree top-hat function, revealing large-scale anisotropies in the southern and northern hemispheres. The data indicate a dipolar structure for cosmic rays with energies above 8 EeV. Adapted from (The IceCube and collaborations, 2016).

Note that this figure is in equatorial coordinates.

conclusion. Intermediate-scale anisotropies, such as the excess observed near Centaurus A, provide evidence for the association of UHECRs with specific astrophysical objects, though the exact nature of these sources remains uncertain due to magnetic field deflections.

1.7 UHECRs and Magnetic fields

As UHECRs are charged particles, they get deflected by intergalactic and Galactic magnetic fields (IGMF and GMF, respectively). Therefore, accurate models of the intervening magnetic fields are essential for inferring UHECR sources, which can ultimately help in disentangling the contributions of protons versus heavier nuclei at various energies, shedding light into the mechanisms that produced the UHECRs, as well as the unexplained features of UHECR data presented in the previous sections.

The degree to which UHECRs are deflected is energy-dependent, with lower-energy UHECRs generally experiencing more significant deviations, while higher-energy particles are deflected less but still enough to pose challenges in source identification

1.7.1 The Intergalactic Magnetic Field

The IGMF is one of the least well-understood components of the cosmic magnetic field environment. Knowledge of the IGMF is fragmented, particularly in cosmic voids and filaments, which comprise most of the space traversed by cosmic rays. While magnetic fields in galaxy clusters are relatively well-constrained, with strengths ranging from $0.1 - 40 \mu G$ (Ryu, Kang, and Biermann, 1998; Blasi and Colafrancesco, 1999), magnetic fields in filaments and voids are far more uncertain. Observations of synchrotron emission, polarization, and Faraday rotation have provided upper limits on the magnetic field strength,

with the coherence length varying depending on the region (Kronberg, Lesch, and Hopp, 1999; Widrow, 2002; Kotera and Olinto, 2011).

In galaxy clusters, IGMFs can reach up to $\sim 1 \mu\text{G}$ in the central regions (Ryu, Kang, and Biermann, 1998; Dolag et al., 2005). However, in cosmic filaments, the magnetic field strength is typically predicted to be lower, on the order of $0.01 - 100 \text{nG}$ (Sigl, Miniati, and Enßlin, 2004; Takami and Sato, 2008). In cosmic voids, which make up a significant portion of the volume of the Universe (between 20% to 80%), the IGMF is even more poorly constrained. Estimates suggest that the field strength could range from 10^{-17} G to 10^{-15} G (Neronov, Semikoz, and Tkachev, 2009; Alves Batista et al., 2019), as derived from gamma-ray observations and numerical simulations (Bertone et al., 2006; Hackstein et al., 2019). These simulations are essential in determining the range of plausible IGMF configurations, as they model the amplification of magnetic fields during large-scale structure formation, including contributions from Biermann battery effects and astrophysical sources such as galactic winds (Sigl, Miniati, and Enßlin, 2004; Kotera and Lemoine, 2008).

Cosmic rays traversing the IGMF can experience significant deflections, particularly in regions of higher magnetic field strength, such as galaxy clusters. However, the existence of a IGMF within cosmic voids is also expected to yield a considerable impact on UHECR propagation. These fields can cause cosmic rays to undergo diffusive propagation at energies below $E \sim 10^{18} \text{ eV}$, particularly for heavy nuclei, further complicating source identification (Bertone et al., 2006; Kotera and Lemoine, 2008; Arámburo-García et al., 2021). The coherence length of the IGMF is also an important parameter for understanding UHECR propagation, with estimates ranging from a few kiloparsecs in filaments to several tens of megaparsecs in cosmic voids (Takami, Yoshiguchi, and Sato, 2006; Kotera and Lemoine, 2008).

IGMFs play a crucial role not only in UHECR propagation but also in multimessenger studies. By confining cosmic rays and increasing their interaction rates with the surrounding gas and radiation fields, IGMFs can enhance the production of secondary particles such as neutrinos and gamma rays (Aloisio, Berezhinsky, and Gazizov, 2012). This effect has important implications for UHECR composition and energy spectrum observations. The interplay between the IGMF and the GMF also introduces challenges for distinguishing between local and extragalactic UHECR sources. For instance, protons with energies above 50 EeV can still experience deflections of up to 15 degrees if the IGMF is sufficiently strong (Dolag et al., 2005).

While significant progress has been made in understanding the IGMF, much remains unknown about its strength, coherence length, and structure, particularly in cosmic voids. Understanding these properties is essential for advancing our knowledge of UHECR propagation and improving source identification efforts in the coming years. In this thesis, we will neglect the effect of the IGMF, and solely focus on correcting for the effect of the GMF. This is not to be understood as an assumption that the effect of the IGMF is insignificant in UHECR propagation, but rather that our current information is lacking for a comprehensive attempt at its modelling.

1.7.2 Galactic Magnetic Field

The GMF consists of two primary components: a large-scale regular field and a small-scale turbulent field. The large-scale regular component follows an axisymmetric spiral structure with typical field strengths of a few to tens of μG within the disk of the Milky

Way (Jansson and Farrar, 2012; Jaffe, 2019). The coherence length of this regular component is on the order of 1 kpc, and its effects are significant for the propagation of cosmic rays, including protons with energies around 1 EeV (Abbasi et al., 2008).

The small-scale turbulent component has a coherence length of approximately 10–100 pc (Jaffe et al., 2013), which results in significant deflections for cosmic rays with energies below a few 10^{17} eV for protons and even higher for heavier nuclei (Abbasi et al., 2008). This turbulent field contributes to the overall complexity of tracing the paths of UHECRs back to their sources.

Furthermore, the GMF extends beyond the Galactic disk into the halo, potentially affecting cosmic ray propagation over large distances. Observations indicate that the field may extend several kpc above the Galactic plane (Orlando and Strong, 2013), though its structure at such heights is less well understood. Some studies suggest the existence of an "x-shaped" vertical field, which is observed in external galaxies and might be present in the Milky Way as well (Jansson and Farrar, 2012; Beck and Wielebinski, 2013).

The GMF is known to reverse direction at several points in the Galactic midplane on scales of about 1 kpc (Brown et al., 2007). These reversals add further complexity to the deflections experienced by UHECRs as they travel through the Galaxy, making source identification more challenging. The understanding of these reversals and other features of the GMF is expected to improve with upcoming surveys like the Square Kilometer Array (SKA), which will provide more detailed measurements of pulsar rotation measures (RMs) and enable precise 3D mapping of the GMF at kpc scales across much of the Galactic disk (Braun et al., 2015; Jarvis et al., 2016; Heald et al., 2020).

Additionally, the *Polar-Areas Stellar Imaging in Polarization High-Accuracy Experiment* (PASIPHAE) project (Tassis et al., 2018) will contribute significantly to GMF modelling by mapping the orientation of the GMF within 1–2 kpc from the Sun using starlight polarization and distance measurements from the Gaia mission, thus opening the way for the possibility of a 3D map of the magnetic field (Tassis et al., 2018). These advancements are expected to offer critical insights into the role of the GMF in cosmic ray propagation, particularly regarding UHECR deflections and the resulting anisotropies observed in their arrival directions.

1.7.3 Parametric Models of the GMF

To account for the deflections caused by the GMF, parametric models have been developed that break down the GMF of the whole Galaxy into three primary components: a toroidal component, a poloidal component, and a turbulent component. The toroidal and poloidal components represent the large-scale ordered magnetic field, while the turbulent component accounts for small-scale random fluctuations that can further deflect UHECRs, modelled as a Gaussian random field (Sun and Reich, 2010; Jansson and Farrar, 2012). These models form the current state of the art in UHECR arrival directions and estimating how much deflection has occurred during the particles' journey through the Galaxy (Magkos and Pavlidou, 2019).

These GMF models rely heavily on observational data from Faraday rotation measurements and synchrotron radiation, both of which provide information on the line-of-sight component of the magnetic field (Beck, 2001; Sun and Reich, 2010). This data provides GMF information that is integrated along the line of sight (LOS), and typically the parameters of the aforementioned models are inferred by a Bayesian method (i.e. nested sampling), conditional to this data. Such parametric models are invaluable in providing an estimate of the GMF on large scales (as they are able to perform a global estimate for

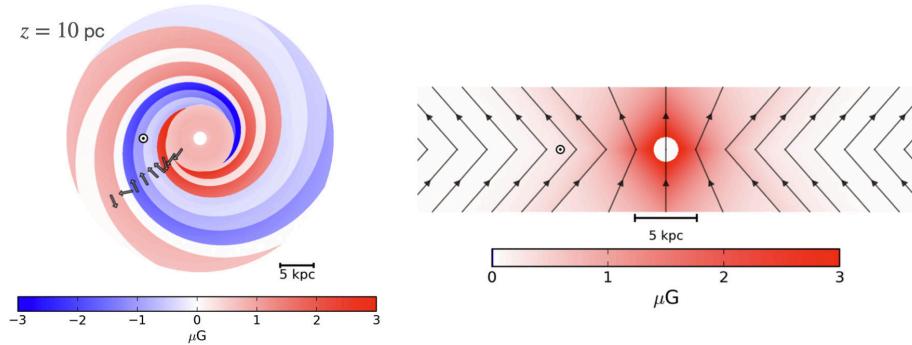


FIGURE 1.7: Left panel: The $z = 10$ pc slice for the best fit parameters of the JF12 model. Right panel: The $x - z$ plane. The arrows show the direction of the GMF, while the location of the Sun is marked with a circle. Adapted from Jansson and Farrar, 2012.

the whole Galaxy), but do sacrifice short-scale information by assuming a specific analytic 3D form which is unable to capture the complexities of the true GMF. Additionally, the turbulent component, which plays a significant role in deflecting UHECRs, is challenging to model accurately, further complicating attempts to backtrack UHECRs to their sources (Guido, 2021).

1.7.4 Prospects for Improved GMF Mapping

A promising avenue for improving the accuracy of GMF models lies in the combination of starlight polarization data with distance measurements, providing 3D tomographic information on the GMF. Starlight passing through interstellar dust grains becomes polarized due to dichroic absorption, a phenomenon where non-spherical dust grains preferentially absorb light with an electric field vector aligned along the grains' longer axes. These grains, in turn, tend to align with the magnetic field, allowing the polarization of starlight to trace the magnetic field's orientation. (Andersson, Lazarian, and Vaillancourt, 2015).

This polarization effect, measured across many stars along a given solid angle, can be combined with distance information from the *Gaia* mission (Collaboration, 2016), which provides precise parallax data for over a billion stars. Through this combination, statistical techniques can be employed in order to infer: a) the location of intervening dust clouds, b) the orientation of the GMF at the location of the dust cloud through the observed induced polarization (Pelgrims et al., 2023), and c) the strength of the GMF at the location of intervening dust clouds (Chandrasekhar and Fermi, 1953; Skalidis et al., 2021). *Therefore, the strength and POS orientation of the GMF can be measured on specific locations in a neighborhood of the Galaxy, up to about 2 kpc from the Earth, which is the distance up to which Gaia has provided measurements.* In Figure 1.8, a cartoon of the this process is shown.

The PASIPHAE survey is set to revolutionize this approach. By obtaining high-precision polarization measurements for millions of stars across both hemispheres, PASIPHAE will allow for direct 3D mapping of the GMF in regions with significant dust presence, especially in high-Galactic latitudes where polarisation data is scarce.

The key advantage of such tomographic reconstructions is that they provide local, yet sparse, information on the magnetic field at specific locations, unlike parametric models

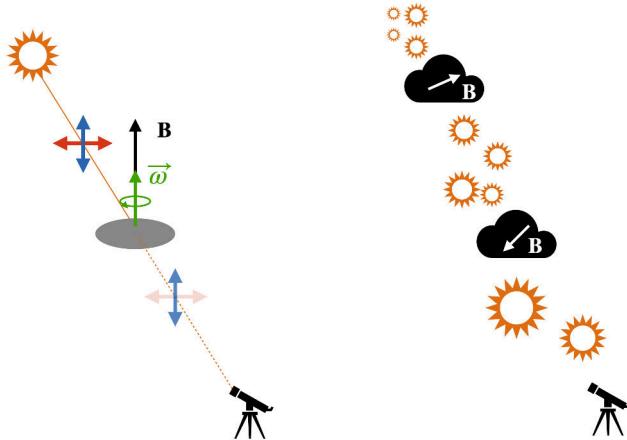


FIGURE 1.8: Left: Prolate rotating dust grains with their major axis aligned with respect to the GMF. As light is emitted unpolarised from its source (star), it is preferentially absorbed along the major axis of the grain. Thus, when observed on Earth, an overall polarisation parallel to that of the GMF at the location of the grain is observed. Right: Starlight-polarization-based tomography: if the polarisation of many stars with known distances is observed along a LOS, the orientation and strength of the GMF at the location of the dust clouds can then be inferred.

that rely solely on LOS-integrated data. By mapping the GMF at the locations of intervening dust clouds and in combination with LOS-integrated data (Faraday and data, synchrotron), it becomes possible to acquire a detailed reconstruction of the magnetic field in for the relevant part of the Galaxy. Such reconstructions can then be used to improve the ability to correct for the ability of the GMF to the observation of arrival directions of UHECRs.

Therefore, a central question that arises is: given the local data provided by stellar polarimetry within a sub-domain of the Galaxy as well as integral data, can we infer the GMF configuration that is compatible with this data in a statistically robust way? If yes, what are the implications for UHECR research, and in particular the removal of the GMF's influence on their paths? The solution of this problem is the central question that we aim to answer in this thesis. Conclusions

In this chapter, we reviewed the role of UHECRs in astrophysics, emphasizing their value as probes of high-energy astrophysical processes. We discussed that despite decades of study, questions about the origins and acceleration mechanisms of UHECRs persist, largely due to their low flux at the highest energies and the deflections they undergo in magnetic fields, which complicate source identification.

We examined the UHECR energy spectrum, noting the features such as the knee, ankle, and GZK cutoff. These spectral transitions help to inform our understanding of UHECR propagation, composition, and interaction with cosmic background photons. We further discussed that identifying the relative proportions of protons versus heavier nuclei remains a central issue, as it has implications for both the nature of UHECR sources and the environments in which these particles are accelerated.

Additionally, we described the observational approaches used by PAO and TA, including their hybrid detection systems that combine fluorescence and surface detectors to measure extensive air showers. These observatories provide crucial data on UHECR properties but face challenges in event interpretation due to uncertainties in particle interaction models and limited event statistics at the highest energies.

Finally, we introduced the role of Galactic and intergalactic magnetic fields in UHECR propagation, with a particular focus on the GMF's influence on arrival direction measurements. The need for accurate GMF models is clear, as these models are essential for minimizing deflection effects and improving source localization. In the next chapter, In the next chapter, we will provide the fundamental mathematical techniques that will allow us to solve this problem, and in the next chapters we will present the applications of this Bayesian approach to solve the problem at hand.

Chapter 2

Bayesian Statistics and Information theory

In the end of the previous chapter, we briefly presented the current status of our knowledge concerning UHECRs, we introduced their phenomenological features, and also discussed the challenges in detecting their sources due to the existence of the IGMF and the GMF. Finally, we showed how local information on the GMF can be acquired through stellar optopolarimetry, and how this gives rise to the main topic of our thesis, namely, the development of a statistical method that constructs local GMF configurations that could have given rise to the observed local data acquired from stellar optopolarimetry, as well as integral data. In this chapter, we will take a detour and concern ourselves with developing the mathematical apparatus that is required to tackle this problem in a robust way. In order to do this, we will develop Bayesian statistics from first principles as an extension of classical binary logic, arrive at Bayes' theorem, and then arrive at the definition of several useful quantities, such as information, entropy, and the Kullback-Liebler divergence. This chapter is by no means comprehensive, and generally follows the development of the unfinished book '*Probability Theory: The Logic of Science*' by E.T. Jaynes (Jaynes, 2003).

2.1 Classical Logic and Its Limitations

Classical logic operates within a binary framework where propositions are either strictly true or strictly false. This system is formalized through Boolean algebra, which governs the manipulation of binary variables using logical operations such as conjunction (AND), disjunction (OR), and negation (NOT) (Jaynes, 2003). In Boolean algebra, each proposition, denoted by a variable such as A , takes a value of either 1 (true) or 0 (false).

The key operations in Boolean algebra are:

- **Conjunction (AND):** The conjunction of two propositions A and B , denoted $A \wedge B$ or $A \cdot B$, is true if both A and B are true. Mathematically,

$$A \wedge B = A \cdot B = \min(A, B). \quad (2.1)$$

- **Disjunction (OR):** The disjunction of two propositions A and B , denoted $A \vee B$ or $A + B$, is true if at least one of A or B is true. Formally,

$$A \vee B = A + B = \max(A, B). \quad (2.2)$$

- **Negation (NOT):** The negation of a proposition A , denoted $\neg A$ or \overline{A} , is true if A is false, and false if A is true:

$$\neg A = \overline{A} = 1 - A. \quad (2.3)$$

In the Boolean calculus propositions A, B, C, \dots are considered only in terms of their truth values, and not their content. Therefore, the equality $A = B$ means that the propositions A and B have the same truth value.

These operations satisfy several important identities, known as the laws of Boolean algebra:

Idempotency:	$A \wedge A = A, \quad A \vee A = A,$
Commutativity:	$A \wedge B = B \wedge A, \quad A \vee B = B \vee A,$
Associativity:	$(A \wedge B) \wedge C = A \wedge (B \wedge C),$ $(A \vee B) \vee C = A \vee (B \vee C),$
Distributivity:	$A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C),$ $A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C).$

Boolean algebra provides a rigorous framework for manipulating logical propositions in a deterministic manner, where truth values are absolute and unaffected by uncertainty.

However, this binary nature of classical logic renders it insufficient for representing uncertainty or degrees of belief. In many applications, especially in fields such as decision-making and scientific inference, propositions are often not strictly true or false but instead have varying levels of plausibility depending on the evidence. This limitation necessitates a more nuanced approach to the formalisation of reasoning under uncertainty.

Furthermore, classical logic does not provide a natural mechanism for updating beliefs in light of new evidence, a requirement that is central to many scientific and statistical processes. This shortfall leads to the development of probability theory as an extension of classical logic, where uncertainty is treated systematically and consistently.

2.2 From Boolean Algebra to Probability

2.2.1 Introduction to Plausibility

Plausibility generalizes binary truth values into a continuous scale, where the degree of belief in a proposition, $p(A|B)$, can take values between 0 and 1. This extension allows for reasoning under uncertainty by quantifying how plausible proposition A is, given some evidence B .

Richard Threlkeld Cox set three postulates that any consistent system of plausible reasoning must satisfy:

1. **Plausibility as Real Numbers:** Plausibility is represented by real numbers, where greater plausibility corresponds to a higher number. Let $p(A|B)$ denote the plausibility of A given B . If $p(A|B) > p(C|B)$, this means A is more plausible than C given B .
2. **Qualitative Correspondence with Common Sense:** Plausibility should qualitatively align with intuitive reasoning. This leads to the **product rule**, which defines the joint plausibility of two propositions A and B given C :
3. **Self-Consistency:** If a conclusion can be derived in multiple ways, the results must agree.

From these requirements, the following rules can be derived (Cox, 1946; Jaynes, 2003). First, the *produce rule*, states that

$$p(A \wedge B|C) = p(A|B \wedge C)p(B|C) = p(B|A \wedge C)p(A|C), \quad (2.4)$$

and the *sum rule* states that

$$p(A|B) = p(A|B) + p(\neg A|B) = 1, \quad (2.5)$$

The proofs of these rules are outside the scope of this introductory chapter, however, the reader should note that they are results of a requirement of the generalisation of classical logic into reasoning under uncertainty, provided the postulates presented above are satisfied. From these rules, one can prove, for instance

$$p(A \vee B|C) = p(A|C) + p(B|C) - p(A \wedge B|C). \quad (2.6)$$

From Eq. 2.4 we can derive Bayes' theorem which is the most central equation in this thesis. Suppose that we want to know how likely A is to be the case, assuming that B is true. Suppressing the background information C , we may write

$$p(A|B) = \frac{P(B|A)p(A)}{p(B)}. \quad (2.7)$$

Therefore the *posterior* probability for A to be the case assuming that B is true, is equal to the *likelihood* that B is the case assuming A is true, times the *prior* probability that A is true regardless of B . The quantity $p(B)$ is a normalisation constant.

2.2.2 Marginalization

Marginalization is a key principle in probability theory, allowing us to account for unknown or irrelevant variables by summing (or integrating) over all possible outcomes of those variables. Suppose we have a set of mutually exclusive and exhaustive possibilities $\{B_i\}_{i=1}^n$ under some background information I . These possibilities satisfy:

$$P(B_i \wedge B_j|I) = 0 \quad \text{for } i \neq j,$$

indicating that the events B_i and B_j are mutually exclusive, and:

$$P(B_1 \vee B_2 \vee \dots \vee B_n|I) = 1,$$

indicating that these events are exhaustive, meaning one of them must occur.

Now, if we are interested in determining the probability of A under the background information I , but A depends on the outcome of some variable B_i , we can marginalize over B_i . The marginalization rule states that the total probability of A is given by summing the joint probabilities of A and B_i over all possible values of B_i :

$$P(A|I) = \sum_{i=1}^n P(A \wedge B_i|I).$$

This expression is known as the B -marginalized probability of A under I because it incorporates the possible values of B_i without requiring specific knowledge of B_i .

By the product rule, we can rewrite each term $P(A \wedge B_i | I)$ as:

$$P(A \wedge B_i | I) = P(A | B_i \wedge I)P(B_i | I),$$

which leads to the following marginalization formula:

$$P(A | I) = \sum_{i=1}^n P(A | B_i \wedge I)P(B_i | I).$$

In this way, marginalization allows us to simplify problems where some variables are not directly observed or are unimportant for the question at hand. It is a fundamental tool in probabilistic reasoning, especially in Bayesian inference, where we often integrate out "nuisance" parameters that we do not need to explicitly account for in the final analysis. Henceforth, we will use the notation A, B in order to signify $A \wedge B$.

2.3 Example: One-Dimensional Inference

In many applications, we are tasked with inferring a random variable X given some observed data Y . In this context, both X and Y are seen some random variables, where x and y denote their specific realizations. As a specific example, let us work with the assumption that X and Y are related through a linear measurement process, which introduces uncertainties both in the measurement itself and in the underlying model.

For simplicity, assume that the measurements process is such that the relationship between the observed data Y and the variable of interest X is linear:

$$Y = aX + N, \quad (2.8)$$

where a is a known real number, and N is a random variable representing measurement noise, which is typically assumed to follow a known distribution. In this case, n is a random variable drawn from a zero-centered Gaussian distribution with σ^2 standard deviation, which is known. Furthermore, let us assume that X is known to be distributed according to

$$P(X) = \frac{1}{\sqrt{2\pi\sigma_X^2}} \exp\left[-\frac{(x - x_0)^2}{2\sigma_X^2}\right], \quad (2.9)$$

that is, X is assumed to be *a priori* distributed according to a Gaussian distribution centered at x_0 with a variance σ_X^2 . From here, we see why the distribution $P(X)$ independent of any measurement Y is called the prior.

Posterior Distribution

Suppose a measurement of a single observable X is made, and our data is observed to take the value $Y = y$. We seek the posterior distribution of X conditioned on Y , denoted $P(X|Y = y)$. This posterior combines our prior knowledge of X , captured by the prior distribution $P(X)$, with the likelihood of observing the data $Y = y$, given X , denoted as

$P(Y = y|X)$. The posterior is

$$\begin{aligned} P(x|y) &\propto P(y|x)P(X) \\ &\propto \int dn \exp\left[-\frac{n^2}{2\sigma^2}\right] \delta(n - y + ax) \exp\left[-\frac{(x - x_0)^2}{2\sigma_X^2}\right] \\ &= \exp\left[-\frac{(y - ax)^2}{2\sigma^2}\right] \exp\left[-\frac{(x - x_0)^2}{2\sigma_X^2}\right] \\ &\propto \exp\left[-\frac{(x - \mu)^2}{2\sigma_{\text{eff}}^2}\right], \end{aligned}$$

where

$$\sigma_{\text{eff}}^2 = \frac{\sigma^2}{a^2 + \sigma^2/\sigma_X^2}, \quad (2.10)$$

and

$$\mu = \frac{\sigma_{\text{eff}}^2}{\sigma^2} \left(ay + x_0 \frac{\sigma^2}{\sigma_X^2} \right) \quad (2.11)$$

In the first line, we marginalise over the noise, and the delta function is used in order to ensure the measurement equation Eq. 2.8 is satisfied. In the third line, we have completed the square in order to rewrite the posterior as a Gaussian distribution, with an updated mean and variance. The posterior as well as prior are shown in Fig. 2.1

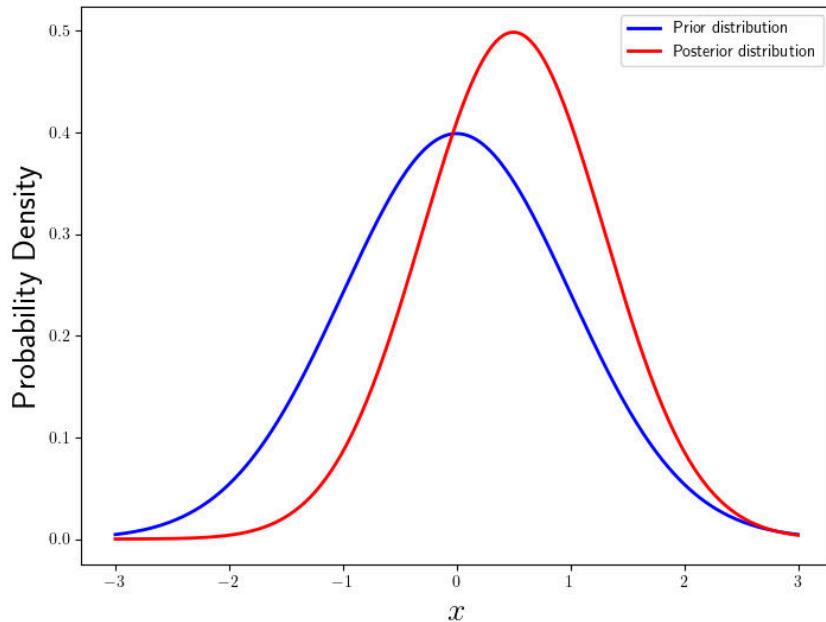


FIGURE 2.1: Inference of a random variable with the forward model of Eq. 2.8. The blue distribution corresponds to the prior knowledge, and the red distribution corresponds to the update after a single measurement.

An immediate question that one might ask is: how much ‘information’ have we gained through our measurement. From Fig. 2.1, it does indeed seem like our measurement has, in some sense, updated our knowledge about the actual value of X , as the prior is less localised than the posterior since it has a larger variance. However, in cases where the

prior and posterior are not Gaussian distributions, and/or we are working in a multi-dimensional setting, this might be so obvious. In the following section we will try to make the notion of information and how informative a distribution is, more general.

2.4 A primer on information and entropy

Information, in its most general sense, quantifies the reduction of uncertainty when learning about an event or outcome. The concept of information can be viewed as a measure of surprise: the more unlikely an event is, the more information is gained upon its occurrence. In the context of inference, information helps quantify how much our knowledge about a certain variable improves when new data is observed.

For example, in observing the outcome of a fair-coin tossing, then the outcome 'Heads' is equally surprising as the outcome of 'Tails', and so these two outcomes contain the same information content.

More formally, the information gained from observing an event x , denoted as $I(x)$, is inversely proportional to the probability of that event:

$$I(x) = -\log P(x), \quad (2.12)$$

where $P(x)$ is the probability of the event $X = x$ occurring, and the base of the logarithm is chosen arbitrarily, with the most common choices being base 2 and natural logarithm. If the former base is used information is measured in *bits*, whereas in the latter case information is measured in *nats*. This formula implies that events with lower probability (i.e., unexpected or surprising events) carry more information. The logarithmic form ensures that the information of independent events can be added.

Therefore, for example, if X a binary variable, like a fair coin, then the outcome $X = \text{Heads}$ has a probability of $P(X = \text{Heads}) = 1/2$, meaning that $I(X = \text{Heads}) = 1$ bit. Specifying a string on N uncorrelated binary variables contains N bits.

A useful measure of the uncertainty of a distribution is the average information content. For a discrete random variable X with a probability distribution $P(X)$, this is given by:

$$H(X) = - \sum_{x \in X} P(x) \log P(x), \quad (2.13)$$

where the sum is over all possible outcomes x of the variable X . This quantity is called the *entropy* of the distribution. Entropy provides a measure of how "spread out" the distribution of X is — the more uncertain we are about X , the larger its entropy.

In the continuous case, where X is a continuous random variable, entropy is defined as:

$$H(X) = - \int_{x \in X} P(x) \log P(x) dx. \quad (2.14)$$

For example, if X follows a Gaussian distribution with mean μ and variance σ^2 , its entropy is:

$$H(X) = \frac{1}{2} \log(2\pi e \sigma^2) = \frac{1}{2} \log(2\pi e) + \log(\sigma) \quad (2.15)$$

Indeed, modulo an additive constant, the entropy of a Gaussian is equal to $\log(\sigma)$, and it increases monotonically with the distribution's variance, according to our intuition from the previous section. Using Eq. 2.10 we can immediately see that measurement of the

variable Y decreases the entropy of the distribution if and only if $a^2 > 0$. If $a = 0$ and our measurement had been pure noise (meaning that Y and X would have been uncorrelated), the measurement of Y would have played no role in our uncertainty regarding the value of the variable X . In different words, measuring pure noise, uncorrelated to the variable X , would add no information.

Of course, the definition of the entropy can be generalised to multiple random variables. For a distribution of two random variables $P(X, Y)$, we define

$$H(X, Y) = - \sum_{x,y} P(x, y) \log P(x, y), \quad (2.16)$$

or, in the continuous case,

$$H(X, Y) = - \int P(x, y) \log P(x, y) dx dy. \quad (2.17)$$

In these expressions, the sums and integrals are always understood to run over all possible values of X and Y . In case our random variables are correlated, we can define the *conditional entropy* of X conditional to Y . This is defined as the entropy of the distribution $P(X|Y)$, averaged over $P(Y)$. That is,

$$H(X|Y) = - \mathbb{E}_Y \sum_x P(x|Y) \log P(x|Y) \quad (2.18)$$

and it is then straightforward to show that

$$H(X|Y) = H(X, Y) - H(Y). \quad (2.19)$$

Now, one can ask whether the entropy of the posterior distribution must always be greater than or equal to the prior. In other words, whether the posterior is more informative than the prior. If that were not the case, then there could be a scenario where additional information could *increase* the uncertainty. If entropy is a candidate for a general determination of a general distribution's information content, then conditioning should *not* increase the distributions entropy. That is, is it true that $H(X) \geq H(X|Y)$? We can work that out by expanding

$$\begin{aligned} H(X) - H(X|Y) &= H(X) + H(Y) - H(X, Y) \\ &= \sum_{x,y} P(x, y) \log P(x, y) - \sum_x P(x) \log P(x) - \sum_y P(y) \log P(y) \\ &= \sum_{x,y} [P(x, y) \log P(x, y) - P(x, y) \log P(x) - P(x, y) \log P(y)] \\ &= \sum_{x,y} P(x, y) \log \frac{P(x, y)}{P(x)P(y)} \equiv D_{\text{KL}}(P(X, Y) \| P(X)P(Y)) \end{aligned} \quad (2.20)$$

In the last line, we defined $D_{\text{KL}}(\cdot, \cdot)$ as a functional whose arguments are the distributions $P(X, Y)$ and $P(X)P(Y)$. It turns out that this function, called the *Kullback-Liebler (KL) divergence* (Kullback and Leibler, 1951) has two key properties (the proof will not be provided here). If P and Q are distributions, then:

- It is non-negative: $D_{\text{KL}}(P||Q) \geq 0$ with equality if and only if $P(X) = Q(X)$ for all X . This means the KL divergence is zero if the two distributions are identical, and positive otherwise.

- It is not symmetric: $D_{\text{KL}}(P||Q) \neq D_{\text{KL}}(Q||P)$. This indicates that it measures the divergence from P to Q specifically, and not the reverse.

It is a non-symmetric ‘distance’ between the distributions P and Q , or rather, it quantifies the extent to which Q is different from P . In the specific case of Eq. 2.20, $D_{\text{KL}}(P(X, Y)||P(X)P(Y))$ is called the *mutual information* between the variables X and Y . Since the KL-divergence is greater than or equal to zero, so is the mutual information, and so

$$H(X) \geq H(X|Y), \quad (2.21)$$

with equality occurring if and only if X and Y are uncorrelated. Therefore, conditioning with respect to correlated variables necessarily *decreases* the entropy (increases the information content). Conversely, conditioning with respect to uncorrelated variables, does not update our knowledge, which is expected: the probability of rain tomorrow doesn’t change if we find out that a cat prefers sleeping in boxes.

Before we move on, an extremely important application of the entropy must be mentioned. Since it is a general characterisation of the uncertainty of a distribution, the distribution that maximizes it under specific conditions is the *least informative choice possible*, while still respecting the desired conditions.

For example, suppose we have good reason to believe that we know the mean and variance of the random variable X , that is, we know x_0 and σ_X . What is the *least informative* distribution we can choose? Rephrasing, we are asking, what is the distribution $P(X)$ such that Eq. 2.14 is maximized, conditional to

$$\int dx P(x) = 1, \quad (2.22)$$

$$\int dx P(x)x = x_0, \quad (2.23)$$

$$\int dx P(x)(x - x_0)^2 = \sigma_X^2. \quad (2.24)$$

Using standard techniques from the calculus of variations, one can show that the desired distribution is a Gaussian, i.e. $\mathcal{N}(x_0, \sigma_X)$. Therefore, if we know the first two moments of the distribution of a random variable, the least informative choice we can make is the Gaussian. This choice is called the *maximum entropy* distribution.

2.5 Example: Multi-Dimensional Inference

Let us now extend the previous one-dimensional case to N variables. Consider a linear measurement process where the observed data Y is related to the variable of interest X via a linear relationship:

$$Y = A \cdot X + C, \quad (2.25)$$

where:

- $Y \in \mathbb{R}^M$ is an M -dimensional observed random variable,
- $A \in \mathbb{R}^{M \times N}$ is a $M \times N$ (generally not invertible) matrix,

- $X \in \mathbb{R}^N$ is an N -dimensional random variable drawn from a Gaussian distribution,
- $C \in \mathbb{R}^M$ is an M -dimensional contamination term, drawn from a zero-centered multivariate Gaussian distribution with diagonal covariance matrix Σ_C .

As before, the goal is to infer the posterior distribution of X , given the observed data Y . To do so, we start by specifying the prior distribution of X and the likelihood of observing Y given X .

2.5.1 Prior Distribution

We assume that X is a Gaussian random variable with mean x_0 and covariance matrix Σ_X :

$$P(X) = \frac{1}{(2\pi)^{N/2}|\Sigma_X|^{1/2}} \exp\left(-\frac{1}{2}(X - x_0)^\top \Sigma_X^{-1}(X - x_0)\right). \quad (2.26)$$

2.5.2 Likelihood

The likelihood of observing the data Y , given X , is determined by marginalising over the measurement contamination C . Since C is Gaussian with covariance matrix Σ_C , the likelihood is:

$$P(Y|X) = \frac{1}{(2\pi)^{M/2}|\Sigma_C|^{1/2}} \exp\left(-\frac{1}{2}(Y - AX)^\top \Sigma_C^{-1}(Y - AX)\right). \quad (2.27)$$

2.5.3 Posterior Distribution

The posterior distribution of X given Y can be obtained by applying Bayes' theorem:

$$P(X|Y) \propto P(Y|X)P(X). \quad (2.28)$$

Substituting the expressions for the prior and the likelihood, we get:

$$P(X|Y) \propto \exp\left(-\frac{1}{2}(Y - AX)^\top \Sigma_C^{-1}(Y - AX)\right) \exp\left(-\frac{1}{2}(X - x_0)^\top \Sigma_X^{-1}(X - x_0)\right) \quad (2.29)$$

$$\propto \exp\left(-\frac{1}{2}\left[(Y - AX)^\top \Sigma_C^{-1}(Y - AX) + (X - x_0)^\top \Sigma_X^{-1}(X - x_0)\right]\right). \quad (2.30)$$

This is a Gaussian distribution for X . By completing the square, we can express the posterior as:

$$P(X|Y) = \mathcal{N}(\mu_{\text{post}}, \Sigma_{\text{post}}), \quad (2.31)$$

where the posterior mean μ_{post} and covariance Σ_{post} are given by:

$$\Sigma_{\text{post}} = \left(A^\top \Sigma_C^{-1} A + \Sigma_X^{-1}\right)^{-1}, \quad (2.32)$$

$$\mu_{\text{post}} = \Sigma_{\text{post}} \left(A^\top \Sigma_C^{-1} Y + \Sigma_X^{-1} x_0\right). \quad (2.33)$$

Thus, the posterior distribution of X is Gaussian, with the updated mean and covariance matrix depending on both the prior distribution and the observed data through the measurement process. A characteristic example of an application might be an image denoising problem: suppose that in that case Y is the observed image and X is the true signal, that we are trying to infer, C , as before, is a contamination with known Gaussian statistics. In this case, we will assume that the prior on the image has a zero mean. Then, the maximum a posteriori choice is straightforward to compute from Eq. 2.32.

This discussion on multidimensional inference can also be generalised to continuous degrees of freedom, by naively taking the continuum limit $N \rightarrow \infty$. The reader should note that, mathematically, this is a non-trivial procedure. For the level of rigour of this thesis, we will assume that in the continuum limit, multi-dimensional random variables are upgraded into fields, and matrices (such as the measurement matrix A) are upgraded into operators (Enßlin, 2019)

So far we have only concerned ourselves with Gaussian problems: the noise is Gaussian with a known covariance, the measurement process is linear, and the prior is also Gaussian with known covariance. These are obviously extremely strict conditions, and most inverse problems do not satisfy all three at the same time. In that case, there are two options. In case the deviation from the above is weak, e.g., the prior on X is weakly non-Gaussian, then one can compute the answer as a perturbation around the Gaussian case, as a power series in the weak nonlinearity. The second option is a non-perturbative, numerical approach where we attempt to directly sample the posterior distribution.

We have now concluded the general introduction into the parts of information theory that are relevant to the solution of the inverse problem that forms the central task of our thesis. In the next chapters we will attempt a solution to the problem, by first attempting a direct approach to the continuous problem, and then

Chapter 3

Field theoretic approach to the GMF inverse problem

In the previous section we introduced the necessary mathematical and statistical apparatus that we will need in order to tackle the inverse problem of inferring the GMF from local measurements. We developed the formalism generally for discrete cases. In this section we will shift gears and go into information theory applied to fields, and turn our attention to the problem at hand, as a first attempt at a solution. The primary purpose of this section is pedagogical, and aims to build up some intuition regarding the complexities that arise even in a setting where significant simplifications are made. In the next chapters we will turn to a numerical approach, but the need for this will be seen by naively attempting to tackle the problem head-on. This chapter is quite mathematical and technical, and the reader could skip it during a first reading. However, its main conclusion should be noted: if we assume a maximum entropy distribution that constrains the magnetic energy of the GMF (and no coupling to plasma), the maximum a posteriori distribution configurations is equal to the Biot-Savart law, where the effective current density that only depends on the observations, normalized by the measurements errors.

3.1 Statement of the Problem

We are given M sparse measurements $\mathbf{d}(\mathbf{x}_m)$ of the Galactic magnetic field, $\mathbf{B}(\mathbf{x})$, at a set of points $\mathbf{x}_m \in \mathcal{M}$, where $\mathcal{M} \subset \mathbb{R}^3$, and $m = 1, \dots, M$. The data at each point \mathbf{x}_m is related to the true signal by the equation¹

$$D_i(\mathbf{x}_m) = \lambda \int d^3\mathbf{x} K_{ij}(\mathbf{x}, \mathbf{x}_m) B_j(\mathbf{x}) + n_i(\mathbf{x}_m), \quad (3.1)$$

where $D_i(\mathbf{x}_m)$ is the i -th component of the vector $\mathbf{d}(\mathbf{x}_m)$. Notice also that, in general, the dimension of the vector $\mathbf{d}(\mathbf{x}_m)$ does not coincide with that of the vector $\mathbf{B}(\mathbf{x})$. The kernel $K_{ij}(\mathbf{x}, \mathbf{x}_m)$ is assumed to be known, and $n_i(\mathbf{x}_m)$ is the i -th component a random 3-vector drawn from a Gaussian distribution with known covariance (called the *noise*):

$$P(\mathbf{n}) = \frac{1}{\sqrt{2\pi \det \Sigma}} \exp \left(-\frac{1}{2} \sum_{m=1}^M \frac{|\mathbf{n}(\mathbf{x}_m)|^2}{\sigma_m^2} \right). \quad (3.2)$$

Finally, λ is a parameter that is typically part of $K_{ij}(\mathbf{x}, \mathbf{x}')$, but is singled out for bookkeeping purposes, and its main function will be to count the powers of $K_{ij}(\mathbf{x}, \mathbf{x}')$.

¹Henceforth, Einstein summation is assumed

By Σ we denote the $M \times M$ covariance matrix of the noise, which is assumed to be diagonal: $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_M^2)$. Eq. 3.1 describes mathematically the act of measurement: the measurement can be seen as a linear operator that acts on the signal space, and maps it into the data space. The noise term corrupts our measurements in a random way.

In general, the operation 3.1 is not invertible. Our task is to reconstruct the original signal $\mathbf{B}(\mathbf{x})$ as best² we can, given $K_{ij}(\mathbf{x}, \mathbf{x}_m)$ and Σ , subject to two *physical conditions*:

- The outcome of the reconstruction, $\mathcal{B}(\mathbf{x})$, must satisfy

$$\nabla \cdot \mathcal{B}(\mathbf{x}) = 0 \quad (3.3)$$

everywhere, and

- It must vanish at infinity.

This problem falls under the general category of inverse problems which are notoriously difficult to tackle because of their ill-posedness, as they do not admit unique solutions. Hence, a statistical approach is required. In the next section we will address this problem in the context of statistical field theory.

3.2 Field Theoretical Approach

Our starting point will be Bayes' theorem:

$$P[\mathbf{B}(\mathbf{x})|\mathbf{d}(\mathbf{x}_m)] = \frac{P[\mathbf{B}(\mathbf{x}), \mathbf{d}(\mathbf{x}_m)]}{P[\mathbf{d}(\mathbf{x}_m)]} = \frac{P[\mathbf{d}(\mathbf{x}_m)|\mathbf{B}(\mathbf{x})]P[\mathbf{B}(\mathbf{x})]}{P[\mathbf{d}(\mathbf{x}_m)]}, \quad (3.4)$$

where, as usual, $P[\mathbf{B}(\mathbf{x})|\mathbf{d}(\mathbf{x}_m)]$ is the posterior probability functional, $P[\mathbf{d}(\mathbf{x}_m)|\mathbf{B}(\mathbf{x})]$ is the likelihood, and $P[\mathbf{B}(\mathbf{x})]$ is the prior probability functional. The marginalised $P[\mathbf{d}(\mathbf{x}_m)]$ is the normalization factor.

Since the notation can get cumbersome we will be introducing a number of different shorthand notations. Unless confusion is possible, we will suppress the dependence by writing

$$\mathbf{B} \equiv \mathbf{B}(\mathbf{x}), \mathbf{D} \equiv \mathbf{d}(\mathbf{x}_m)$$

In order to make contact with statistical physics, we will rewrite Eq. 3.4 as

$$P[\mathbf{B}|\mathbf{D}] = \frac{e^{-\beta H[\mathbf{B}, \mathbf{D}]}}{\mathcal{Z}} = \frac{e^{-\beta H_L[\mathbf{D}|\mathbf{B}] - \beta H_P[\mathbf{D}]}}{\mathcal{Z}}, \quad (3.5)$$

where the redefinitions are

²The term ‘best’ will be defined more rigorously later on.

$$\begin{aligned}
H_L[\mathbf{D}|\mathbf{B}] &\equiv -\frac{1}{\beta} \log(P[\mathbf{D}|\mathbf{B}]), \\
H_P[\mathbf{B}] &\equiv -\frac{1}{\beta} \log(P[\mathbf{B}]) \\
H[\mathbf{B}, \mathbf{D}] &= H_L[\mathbf{D}|\mathbf{B}] + H_P[\mathbf{B}] \\
\mathcal{Z} \equiv P[\mathbf{D}] &= \int \mathcal{D}[\mathbf{B}(\mathbf{x})] P[\mathbf{B}|\mathbf{D}]
\end{aligned}$$

We will call the first three collectively *information Hamiltonians*. The parameter β is a quantity with dimensions Energy $^{-1}$, introduced such that the information Hamiltonians have dimensions of energy. The subscripts L and P refer to the likelihood and prior Hamiltonian respectively. The normalization factor, \mathcal{Z} , is the *partition functional*. Introducing a source term (or singling it out if it already exists in the information Hamiltonian), this reads

$$\mathcal{Z}[\mathbf{J}] = \int \mathcal{D}[\mathbf{B}(\mathbf{x})] \exp \left[-H[\mathbf{B}, \mathbf{D}]|_{\mathbf{J}=0} + \int d^3\mathbf{x} \mathbf{J}(\mathbf{x}) \cdot \mathbf{B}(\mathbf{x}) \right] \quad (3.6)$$

The measure $\mathcal{D}[\mathbf{B}(\mathbf{x})]$ denotes a functional integral (or *path integral*) over *all possible field configurations*. The partition functional will be the central object of our study, since if we are able to compute it in closed form, all the cumulants can be acquired by successive functional differentiation. In particular, the mean field configuration $\mathcal{B}(\mathbf{x}) \equiv \langle \mathbf{B}(\mathbf{x}) \rangle$, is gotten by³

$$\langle B_i(\mathbf{x}) \rangle = \frac{1}{\mathcal{Z}[\mathbf{J}]} \frac{\delta \mathcal{Z}[\mathbf{J}]}{\delta J_i(\mathbf{x})} \Big|_{\mathbf{J}=0}, \quad (3.7)$$

while the variance is

$$\langle B_i(\mathbf{x}) B_j(\mathbf{x}') \rangle = \frac{1}{\mathcal{Z}[\mathbf{J}]} \frac{\delta^2 \mathcal{Z}[\mathbf{J}]}{\delta J_i(\mathbf{x}) \delta J_j(\mathbf{x}')} \Big|_{\mathbf{J}=0}. \quad (3.8)$$

At the end of the calculation, the source is taken to zero if it was introduced ad hoc. If it was part of the information Hamiltonian, then we do not require $\mathbf{J} = 0$ after the differentiations.

Before we proceed to the calculation of the information Hamiltonians, we must address the imposition of the $\nabla \cdot \mathcal{B}(\mathbf{x}) = 0$ requirement. As is natural in electrodynamics, we will require our magnetic field to be written as the curl of a vector potential, $\mathbf{B} = \nabla \times \mathbf{A}$. As it stands, Eq. 3.6 integrates over *all the possible field configurations*, which is too much. We can alleviate this problem by introducing a delta functional that imposes our requirement over the field configurations that are being sampled. Therefore, $\mathcal{Z}[\mathbf{J}]$ will be written as

$$\mathcal{Z}[\mathbf{J}] = \int \mathcal{D}[\mathbf{B}(\mathbf{x})] \exp \left[-H[\mathbf{B}, \mathbf{D}]|_{\mathbf{J}=0} + \int d^3\mathbf{x} \mathbf{J}(\mathbf{x}) \cdot \mathbf{B}(\mathbf{x}) \right] \delta\{\mathbf{B} - \nabla \times \mathbf{A}\}. \quad (3.9)$$

Using the standard identity

³A word of caution: the cumulants $\langle \cdot \rangle$ are themselves functions of space, and *not* averages of their arguments. That is, $\langle \mathbf{B}(\mathbf{x}) \rangle$ is the mean configuration, or the mean field.

$$\delta\{\mathbf{B} - \nabla \times \mathbf{A}\} = \int \mathcal{D}[\mu(\mathbf{x})] e^{i \int d^3\mathbf{x} \mu(\mathbf{x})(\mathbf{B} - \nabla \times \mathbf{A})}, \quad (3.10)$$

we can rewrite 3.6 as

$$\mathcal{Z}[\mathbf{J}] = \int \mathcal{D}[\mu(\mathbf{x})] \mathcal{D}[\mathbf{B}(\mathbf{x})] \exp \left[-H[\mathbf{B}, \mathbf{D}] + i \int d^3\mathbf{x} \mu(\mathbf{x})(\mathbf{B} - \nabla \times \mathbf{A}) \right]. \quad (3.11)$$

The field $\mu(\mathbf{x})$ is called a *Lagrange multiplier*, since it is an auxiliary field introduced for the purpose of fixing a specific condition.

We have therefore recast the inverse problem in a more traditional setting. Colloquially stated, given data on \mathbf{B} , what is the best possible reconstruction of the vector potential $\mathbf{A}(\mathbf{x})$? This way of phrasing the problem elevates the burden of having to impose 3.3, at the cost of another one: we introduce an additional degree of freedom, since even given the original signal $\mathbf{B}(\mathbf{x})$, there are infinitely many $\mathbf{A}(\mathbf{x})$ that could work. This is of course due to the gauge freedom of electrodynamics, and techniques to remedy it are well known, and we will do that in due course.

3.2.1 The Likelihood

Before we proceed, we will be introducing the shorthand notation

$$(KB)_i(\mathbf{x}_m) = (KB)_i \equiv \lambda \int d^3\mathbf{x} K_{ij}(\mathbf{x}, \mathbf{x}_m) B_j(\mathbf{x}).$$

We write the posterior PDF by marginalising the noise, i.e.

$$P(\mathbf{D}|\mathbf{B}) = \int d^3\mathbf{n}_1 \cdots d^3\mathbf{n}_M P(\mathbf{D}, \mathbf{n}_1, \dots, \mathbf{n}_M | \mathbf{B}) \quad (3.12)$$

$$= \int d^3\mathbf{n}_1 \cdots d^3\mathbf{n}_M P(\mathbf{D} | \mathbf{n}_1, \dots, \mathbf{n}_M, \mathbf{B}) P(\mathbf{n}_1, \dots, \mathbf{n}_M, \mathbf{B}) \quad (3.13)$$

Using the fact that the noise is independent of the signal, we have

$$P(\mathbf{n}_1, \dots, \mathbf{n}_M, \mathbf{B}) = P(\mathbf{n}_1, \dots, \mathbf{n}_M), \quad (3.14)$$

which is given by 3.2. Further, due to Eq. 3.1, we have (using index notation)

$$P(D_i, (n_1)_i, \dots, (n_M)_i | B_i) = \delta\{D_i - (KB)_i - n_i\}. \quad (3.15)$$

The integrations over the noise can then easily be done, yielding,

$$P(\mathbf{D}|\mathbf{B}) = \frac{1}{\sqrt{2\pi \det \Sigma}} \exp \left(-\frac{1}{2} \sum_{m=1}^M \frac{[D_m(\mathbf{x}_m) - (KB)_m(\mathbf{x}_m)]^2}{\sigma_m^2} \right) \quad (3.16)$$

3.2.2 The Prior

Interlude: A Primer on L_2 -Regularization

Ill-posed inverse problems typically require some sort of regularization in order to provide stable solutions. In a typical setting, our case being an example, we are interested in recovering a field $\phi(\mathbf{x})$, with $\mathbf{x} \in \Omega$, where $\Omega \subseteq \mathbb{R}^d$. We are given a forward linear operator K , and data on a function $\rho(\mathbf{x})$, satisfying $K(\phi) = \rho$. Given sparse measurements on ρ , the challenge is to find a reasonable ϕ such that the predicted data match the observed data within some tolerance. Discretizing Ω , the field $\phi(\mathbf{x})$ is demoted to a vector taking values at each node of the mesh: $\phi(\mathbf{x}) \rightarrow \phi_i$. The same thing happens to $\rho(\mathbf{x})$. The linear operator K is also demoted to a matrix and in whole, the discretization has the following effect:

$$K(\phi) = \rho \rightarrow K_{ij}\phi_j = \rho_i.$$

The matrix K_{ij} has dimension $m \times n$, where n is the dimension of ϕ_i (the number of nodes of the mesh), while m is the number of data one has for ρ . Of course, $m \leq n$. The nontriviality of the null-space of K_{ij} implies that the inverse does not exist, since the system is underdetermined and has an infinite number of solutions. Typically, then, one writes down a *regularized* problem, by requiring that ϕ satisfies

$$\min_{\phi} (\|K\phi - \rho\|_W^2 + \alpha R(\phi)). \quad (3.17)$$

The function to be minimized is called the *loss function*. Here $\|\cdot\|_W$ denotes the standard weighted norm

$$\|A\|_W \equiv \sqrt{A^T \mathbf{W} A}, \quad (3.18)$$

for some vector A . For our case, \mathbf{W} is the inverse of the covariance matrix of our measurements. The parameter $\alpha \geq 0$ is the *regularization parameter*, and R is a regularization operator. For $\alpha = 0$, this is the ordinary least squares procedure, and the $\alpha R(\phi)$ term adds additional desirable properties. Choosing $R(\phi) \propto \|\phi\|_2^2$, for instance, gives preference to vectors with smaller norms, in order to penalize overfitting⁴. This is known as the L_2 -regularization (or *generalized Tikhonov regularization*). Different choices for R equate to different regularization schemes.

3.2.3 Maximum Entropy Prior

Another choice for the prior could be a maximum entropy distribution that constrains some functional of the field. For instance, one can suppose that the energy fluctuates around a given value that is known. Then, the distribution that is the least informative and still respects this requirement, is the one that maximizes the entropy, as discussed in the previous chapter. If β is the Lagrange multiplier that enforces the condition, the distribution that we acquire is the Gibbs distribution

$$P(\mathbf{B}) \propto e^{-\beta E[\mathbf{B}(\mathbf{x})]}, \quad (3.19)$$

where $E[\mathbf{B}(\mathbf{x})]$ is the energy of the given configuration. For the energy of a magnetic field we know that

⁴This can be seen intuitively: a ‘wiggly’ line has a larger norm than a straighter line.

$$E[\mathbf{B}(\mathbf{x})] = \frac{1}{2\mu_0} \int d^3\mathbf{x} |\mathbf{B}(\mathbf{x})|^2. \quad (3.20)$$

Therefore, setting $H_P[\mathbf{B}(\mathbf{x})] = E[\mathbf{B}(\mathbf{x})]$ in 3.5, where the covariance matrix is just a multiple of the identity matrix, has a very straightforward physical interpretation. Essentially, choosing this expression for the energy (acquired from standard EM theory) implicitly assumes that the GMF is in thermal equilibrium, which is not a valid assumption as the true GMF is not in thermal equilibrium, and moreover it is coupled to other fields, such as the plasma of the ISM. However, in what follows, we will argue that choosing the Boltzmann prior is equivalent to just requiring our interpolation $\mathcal{B}(\mathbf{x})$ to be L_2 -regularized.

Our total information Hamiltonian reads

$$H[\mathbf{B}, \mathbf{D}] = \frac{1}{2\beta} \int d^3\mathbf{x} \sum_{m=1}^M \frac{1}{\sigma_m^2} \delta(\mathbf{x} - \mathbf{x}_m) [D_i(\mathbf{x}) - (KB)_i(\mathbf{x})]^2 + \frac{1}{2\mu_0} \int d^3\mathbf{x} |\mathbf{B}(\mathbf{x})|^2 \quad (3.21)$$

As it stands, due to the sums over the repeated i indices, the information Hamiltonian breaks into a sum over three terms, one for each component of the magnetic field and the data. This means that the three components are statistically independent (since their respective distributions multiply), and we can argue over a single one of them without loss of generality. Let us choose the case $i = 1$, and rewrite $B_1(\mathbf{x}) \rightarrow \phi(\mathbf{x})$, and $D_1(\mathbf{x}) \rightarrow \rho(\mathbf{x})$. Thus, we are allowed to write

$$D_i(\mathbf{x}) - (KB)_i(\mathbf{x}) \rightarrow \rho(\mathbf{x}) - K[\phi(\mathbf{x})]. \quad (3.22)$$

Integrating over the delta function and discretizing our domain thus treating our distributions as multivariable probability density functions, the first term reads

$$H_L[\rho|\phi] = \frac{1}{2\beta} (\rho_i - K_{im}\phi_m) \Sigma_{ij}^{-1} (\rho_i - K_{in}\phi_n) \quad (3.23)$$

as well as

$$H_P[\phi] = \frac{1}{2\mu_0} \|\phi\|_2^2. \quad (3.24)$$

The maximum a posteriori (MAP) estimate, the one that maximizes $P(\phi|\rho)$ and minimizes $H[\rho, \phi]$, is

$$\min_{\phi} \left[(\rho_i - K_{im}\phi_m) \Sigma_{ij}^{-1} (\rho_i - K_{in}\phi_n) + \frac{\beta}{\mu_0} \|\phi\|_2^2 \right]. \quad (3.25)$$

or

$$\min_{\phi} \left[\|\rho - K\phi\|_{\Sigma^{-1}}^2 + \frac{\beta}{\mu_0} \|\phi\|_2^2 \right]. \quad (3.26)$$

Comparing 3.26 with 3.17, and identifying $\Sigma^{-1} \rightarrow \mathbf{W}$ and $\beta/\mu_0 \rightarrow \alpha$ finishes our argument. In fact, this can be seen as physical justification in order to enforce an L_2 regularization.

3.3 Calculation of the Partition Functional : The Effective Current Density

In the following we will assume the maximum entropy prior presented in the previous section, and we will go back to calculating the partition functional. We expand the likelihood Hamiltonian term and keep only terms that depend on $\mathbf{B}(\mathbf{x})$,

$$H_L[\mathbf{D}|\mathbf{B}] \triangleq \frac{1}{2\beta} \int d^3\mathbf{x} \sum_{m=1}^M \frac{1}{\sigma_m^2} \delta(\mathbf{x} - \mathbf{x}_m) [-2D_i(\mathbf{x})(KB)_i(\mathbf{x}) + (KB)_i(\mathbf{x})(KB)_i(\mathbf{x})]. \quad (3.27)$$

We are allowed to dispense with the term quadratic in $\mathbf{d}(\mathbf{x}_m)$, since due to 3.7 and respective expressions for higher-order correlation functions, multiplicative factors do not alter the cumulants. The symbol \triangleq signifies equality modulo irrelevant constants.

We can rewrite 3.27 as

$$H_L[\mathbf{D}|\mathbf{B}] \triangleq \frac{\lambda^2}{2\beta} \int d^3\mathbf{x} d^3\mathbf{x}' B_i(\mathbf{x}) \Lambda_{ij}(\mathbf{x}, \mathbf{x}') B_j(\mathbf{x}') - \int d^3\mathbf{x} \tilde{J}_i(\mathbf{x}) B_i(\mathbf{x}), \quad (3.28)$$

where we have made the definitions

$$\tilde{J}_i(\mathbf{x}) \equiv \lambda \sum_{m=1}^M \frac{1}{\sigma_m^2} K_{ij}(\mathbf{x}, \mathbf{x}_m) D_j(\mathbf{x}_m), \quad (3.29)$$

and

$$\Lambda_{ij}(\mathbf{x}, \mathbf{x}') \equiv \sum_{m=1}^M \frac{1}{\sigma_m^2} K_{ri}(\mathbf{x}, \mathbf{x}_m) K_{rj}(\mathbf{x}', \mathbf{x}_m). \quad (3.30)$$

We can therefore write the full Hamiltonian as

$$H[\mathbf{B}, \mathbf{D}] = \frac{1}{2} \int d^3\mathbf{x} d^3\mathbf{x}' B_i(\mathbf{x}) \left[\frac{\lambda^2}{\beta} \Lambda_{ij}(\mathbf{x}, \mathbf{x}') + \frac{1}{\mu_0} \delta_{ij} \delta(\mathbf{x} - \mathbf{x}') \right] B_j(\mathbf{x}') - \frac{1}{\beta} \int d^3\mathbf{x} \tilde{J}_i(\mathbf{x}) B_i(\mathbf{x}) \quad (3.31)$$

Equipped with the information Hamiltonian we will calculate Eq. 3.11. First, let's do the integral over $\mu(\mathbf{x})$. This will make the change $B_i \rightarrow \varepsilon_{ijk} \partial_j A_k$. Integrating by parts and disregarding boundary terms that are taken to vanish, we end up with

$$H[\mathbf{A}, \mathbf{D}] = \frac{1}{2} \int d^3\mathbf{x} d^3\mathbf{x}' A_m(\mathbf{x}) \left[\frac{\lambda^2}{\beta} \varepsilon_{mrk} \varepsilon_{nls} \partial_r \partial'_l \Lambda_{ks}(\mathbf{x}, \mathbf{x}') + \frac{1}{\mu_0} \delta(\mathbf{x} - \mathbf{x}') (\partial_n \partial_m - \nabla^2 \delta_{mn}) \right] A_n(\mathbf{x}') - \frac{1}{\beta} \int d^3\mathbf{x} A_i(\mathbf{x}) \varepsilon_{ijk} \partial_j \tilde{J}_k(\mathbf{x}). \quad (3.32)$$

where $\partial'_i \equiv \partial / \partial x'_i$.

Defining further,

$$S_{mn}(\mathbf{x}, \mathbf{x}') \equiv \lambda^2 \epsilon_{mrk} \epsilon_{nls} \partial_r \partial'_l \Lambda_{ks}(\mathbf{x}, \mathbf{x}') + \frac{\beta}{\mu_0} \delta(\mathbf{x} - \mathbf{x}') (\partial_n \partial_m - \nabla^2 \delta_{mn}) \quad (3.33)$$

and

$$J_i(\mathbf{x}) \equiv \epsilon_{ijk} \partial_j \tilde{J}_k(\mathbf{x}) \quad (3.34)$$

the Hamiltonian 3.32 is written more compactly as

$$H[\mathbf{A}, \mathbf{D}] = \frac{1}{2\beta} \int d^3 \mathbf{x} d^3 \mathbf{x}' A_m(\mathbf{x}) S_{mn}(\mathbf{x}, \mathbf{x}') A_n(\mathbf{x}') - \frac{1}{\beta} \int d^3 \mathbf{x} A_i(\mathbf{x}) J_i(\mathbf{x}). \quad (3.35)$$

It is relatively simple to check that $S_{mn}(\mathbf{x}, \mathbf{x}') = S_{nm}(\mathbf{x}', \mathbf{x})$. This is sufficient to diagonalize S : in the expansion, note that we can interchange $\mathbf{x} \leftrightarrow \mathbf{x}'$ in the off diagonal terms and see that without loss of generality we can just keep the symmetric part of the matrix $S_{mn}(\mathbf{x}, \mathbf{x}')$. We are only allowed to do this because of the fact that the variables \mathbf{x} and \mathbf{x}' are being integrated over.

We are now ready to perform the integration over $\mathbf{A}(\mathbf{x})$, yielding

$$\mathcal{Z}[J] \propto \exp \left(\int d^3 \mathbf{x} d^3 \mathbf{x}' J_m(\mathbf{x}) \Delta_{mn}(\mathbf{x}, \mathbf{x}') J_n(\mathbf{x}') \right), \quad (3.36)$$

where $\Delta_{mn}(\mathbf{x}, \mathbf{x}')$ is the Green's function, satisfying

$$\int d^3 \mathbf{x}' S_{mn}(\mathbf{x}, \mathbf{x}') \Delta_{nr}(\mathbf{x}', \mathbf{x}'') = \delta_{mr} \delta(\mathbf{x} - \mathbf{x}''). \quad (3.37)$$

In principle, we are done. Given a specific kernel $K_{ij}(\mathbf{x}, \mathbf{x}_m)$, our task is to solve Eq. 3.37, and by using 3.7 calculate the mean field $\langle \mathbf{B}(\mathbf{x}) \rangle$, as well as the variance or any other cumulant we might be interested in.⁵

Before we conclude this section and we move on to specific examples, let us attempt to get some further insight. Assume that the general solution for the Green's function is written as a in λ . i.e.

$$\Delta_{mn}(\mathbf{x}', \mathbf{x}'') = \sum_{k=0}^{\infty} \lambda^k \Delta_{mn}^{(k)}(\mathbf{x}', \mathbf{x}''). \quad (3.38)$$

Inserting 3.38 into 3.37 and equating powers of λ , yields to order λ^0 :

$$(\partial_m \partial_n - \nabla^2 \delta_{mn}) \Delta_{nr}^{(0)}(\mathbf{x}', \mathbf{x}'') = \frac{\mu_0}{\beta} \delta_{mr} \delta(\mathbf{x} - \mathbf{x}''). \quad (3.39)$$

This is just the equation for the standard EM Green's function - or just the expression for all the spacelike components of the photon propagator in QED (modulo factors of i). As it stands, the operator

$$L_{mn} \equiv \partial_n \partial_m - \nabla^2 \delta_{mn} \quad (3.40)$$

is not invertible. This can easily be seen by going to Fourier space and seeing that if p is a momentum mode, then $P_{mn} p_n = 0$. The matrix has (at least one) vanishing eigenvalue,

⁵Since we have decided to single out the parameter λ , the cumulants must also be multiplied by a factor of λ^{-n} , where n is the number of derivatives.

and is thus not invertible. A common trick to getting a solution to 3.39 lies in rewriting 3.40 as

$$L_{mn} \rightarrow L_{mn}(\xi) \equiv \partial_n \partial_m \left(1 - \frac{1}{\xi} \right) - \nabla^2 \delta_{mn}, \quad (3.41)$$

where the parameter ξ fixes the gauge. In Fourier space, the propagator reads

$$\tilde{\Delta}_{mn}^{(0)}(p) = \frac{\mu_0}{\beta} \frac{\delta_{mn} - (1 - \xi) \frac{p_m p_n}{p^2}}{p^2}, \quad (3.42)$$

for an arbitrary gauge, as can be readily checked. For $\xi = 1$ (the Feynman gauge) we simply get the Green's function for the Laplacian, which in configuration space reads

$$\Delta_{mr}^{(0)}(\mathbf{x}', \mathbf{x}'') = \delta_{mr} \frac{1}{\beta} \frac{\mu_0}{4\pi} \frac{1}{|\mathbf{x} - \mathbf{x}'|}. \quad (3.43)$$

We may then proceed to solve for the $\Delta_{mr}^{(n)}(\mathbf{x}', \mathbf{x}'')$ for $n \geq 1$ iteratively, from 3.37. Using 3.7, we get the mean potential field to be

$$\begin{aligned} \langle A_i(\mathbf{x}) \rangle &= \int d^3 \mathbf{x}' \Delta_{ij}(\mathbf{x}, \mathbf{x}') J_j(\mathbf{x}') \\ &= \int d^3 \mathbf{x}' \Delta_{ij}^{(0)}(\mathbf{x}, \mathbf{x}') J_j(\mathbf{x}') + \mathcal{O}(\lambda^2) \\ &= \frac{1}{\beta} \frac{\mu_0}{4\pi} \int d^3 \mathbf{x}' \frac{1}{|\mathbf{x} - \mathbf{x}'|} [J_i(\mathbf{x}') + \mathcal{O}(\lambda^2)] \end{aligned} \quad (3.44)$$

By $\mathbf{J}^{\text{eff}}(\mathbf{x})$ we define the effective current density that appears in the last bracket of equation 3.44 $\langle \mathbf{B}(\mathbf{x}) \rangle$ accurate to all orders in λ , that is, including the full propagator 3.38. Explicitly, to order λ this reads

$$J_i^{\text{eff}}(\mathbf{x}) = \lambda \epsilon_{ijk} \partial_j \sum_{m=0}^M \frac{1}{\sigma_M^2} K_{kr}(\mathbf{x}, \mathbf{x}_m) D_r(\mathbf{x}_m) + \mathcal{O}(\lambda^2). \quad (3.45)$$

The higher order terms are calculated perturbatively, by including higher order terms of the propagator, and solving equation 3.37 iteratively, if a closed form solution of 3.37 is not available.

For the mean magnetic field we just take the curl of expression 3.44, which gives

$$\langle B_i(\mathbf{x}) \rangle = \frac{\mu_0}{4\pi\beta} \int d^3 \mathbf{x}' \frac{\epsilon_{ijk} J_j^{\text{eff}}(\mathbf{x}') (\mathbf{x} - \mathbf{x}')_k}{|\mathbf{x} - \mathbf{x}'|^3}, \quad (3.46)$$

or in vector notation

$$\langle \mathbf{B}(\mathbf{x}) \rangle = \frac{\mu_0}{4\pi\beta} \int d^3 \mathbf{x}' \frac{\mathbf{J}^{\text{eff}}(\mathbf{x}') \times (\mathbf{x} - \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|^3}$$

(3.47)

which is just the Biot-Savart law for a current density given by equation 3.45. If the kernel vanished, our data is just noise and does not provide any information on the actual signal. Hence, our total information Hamiltonian has an $O(3)$ symmetry, and the mean field vanishes in order to respect that symmetry. But our information breaks this symmetry

explicitly by introducing a source. Hence, the mean field acquires a nonzero value for a nonzero kernel.

As an example, let us calculate the MAP field from equation 3.47 for the case where our measurement equation 3.1 is of the form

$$D_i(\mathbf{x}_m) = \lambda B_i(\mathbf{x}_m) + n_i(\mathbf{x}_m), \quad (3.48)$$

for $m = 1, \dots, M$, and $\lambda \ll 1$. This corresponds to the kernel

$$K_{ij}(\mathbf{x}, \mathbf{x}_m) = \delta_{ij}\delta(\mathbf{x} - \mathbf{x}_m). \quad (3.49)$$

Substituting 3.49 into 3.47, we find

$$\langle \mathbf{B}(\mathbf{x}) \rangle = \lambda \frac{\mu_0}{4\pi} \sum_{m=1}^M \frac{1}{\sigma_m^2} \left[\frac{3\mathbf{d}(\mathbf{x}_m) \cdot (\mathbf{x} - \mathbf{x}_m)}{|\mathbf{x} - \mathbf{x}_m|^5} (\mathbf{x} - \mathbf{x}_m) - \frac{\mathbf{d}(\mathbf{x}_m)}{|\mathbf{x} - \mathbf{x}_m|^3} \right] + \mathcal{O}(\lambda^2). \quad (3.50)$$

Thus, to order $\mathcal{O}(\lambda)$ we find that the mean magnetic field is that of a superposition of the fields of individual magnetic dipole moments, each with a magnetic moment given by

$$\boldsymbol{\mu}_m = \frac{\lambda \mathbf{d}(\mathbf{x}_m)}{\sigma_m^2}. \quad (3.51)$$

Notice that if the error of a specific datum is small, then the corresponding magnetic moment is large, and vice versa. Further, we see that in this case, the parameter λ can be taken to scale the variance of the noise, so small λ implies a signal much weaker to the noise component.

Since in this case we assume the prior covariance is known, the complete solution is provided by the generalised Wiener filter, briefly introduced in the previous chapter. However, the choice of prior is the main weak point of this chapter's analysis. In the next chapters, we are going to relax the prior, and instead infer the prior covariance from the data. However, this will increase the complexity of the problem significantly, as the covariance (or the power spectrum, as we will work in Fourier space) can be seen as an independent field coupled to the magnetic field, thus rendering our inference problem non-Gaussian. For this reason, we will need to turn to non-perturbative numerical techniques.

Chapter 4

Nonparametric Bayesian Reconstruction of the GMF Using IFT for UHECR Backtracking (Part 1)

In this chapter, we once again address the challenge of reconstructing the GMF using sparse, localized measurements with the specific aim of enhancing our ability to correct the effect of the GMF on the observed UHECR arrival directions. However, as opposed to the naive field-theoretic approach shown in the previous chapter, we now turn to a more pragmatic, numerical, and non-perturbative approach, utilizing techniques from Information Field Theory (IFT) and the respective computational package, NIFTy. This framework allows us to sample an approximation of the posterior distribution of the GMF's structure within the Galactic region of interest, with an unknown power spectrum. By sampling GMF configurations from this posterior, we can model the deflections experienced by UHECRs as they traverse the Galactic environment. Specifically, given the energy, charge, and observed arrival direction of a UHECR (which are assumed known), we solve the equations of motion across these sampled GMF configurations, deriving a probability distribution for the true arrival directions on the POS. In this work, the true magnetic field is assumed to take the form of a uniform component plus a Gaussian random field component, that attempts to capture the turbulent properties of the GMF, at least up to the scale relevant for our problem. In addition, the distribution of sparse data points is assumed to be uniform within the domain, but different number densities of data points are examined.

Our results demonstrate that for weakly turbulent GMFs, this approach can effectively correct the deflection of UHECRs, achieving a precision within approximately 3° of the true arrival direction. This level of accuracy is maintained even when measurements are relatively sparse, with an average spacing of up to 600 pc, a conservative estimate aligned with the distribution of molecular clouds in the Galaxy. In cases of fully turbulent GMFs (no uniform component), the method remains valuable, offering a significant improvement over simpler models that often fail under such conditions. We compare the results we obtain from directly sampling the (approximate) posterior to more straightforward approaches, such as merely considering the vector mean of the data points, or nearest-neighbor estimates. Generally, our method outperforms such ad hoc methods, and in addition provides a robust uncertainty quantification on the inferred UHECR arrival directions. Overall, this work lays the groundwork for utilizing future GMF measurements to refine UHECR source identification, even in the presence of complex magnetic structures.

Reconstructing Galactic magnetic fields from local measurements for backtracking ultra-high-energy cosmic rays

Alexandros Tsouros^{1,2} , Gordian Edenhofer^{3,4}, Torsten Enßlin^{3,4}, Michalis Mastorakis^{1,2}, and Vasiliki Pavlidou^{1,2}

¹ Department of Physics & ITC, University of Crete, 70013, Heraklion, Greece
e-mail: tsouros@physics.uoc.gr

² Institute of Astrophysics, Foundation for Research and Technology-Hellas, Vasilika Vouton, 70013 Heraklion, Greece

³ Ludwig Maximilian University of Munich, Geschwister-Scholl-Platz 1, 80539 Munich, Germany

⁴ Max Planck Institute for Astrophysics, Karl-Schwarzschild-Straße 1, 85748 Garching, Germany

Received 15 March 2023 / Accepted 14 October 2023

ABSTRACT

Context. Ultra-high-energy cosmic rays (UHECRs) are highly energetic charged particles with energies exceeding 10^{18} eV. These energies are far greater than those achieved in Earth-bound accelerators, and identifying their sources and production mechanism can shed light on many open questions in both astrophysics and high-energy physics. However, due to the presence of the Galactic magnetic field (GMF) they are deflected, and hence the location of their true source on the plane of the sky (PoS) is concealed. The identification of UHECR sources is an open question, exacerbated by the large uncertainties in our current understanding of the three-dimensional structure of the GMF. This difficulty arises from the fact that currently all GMF observations are integrated along the line of sight (LoS). However, thanks to upcoming stellar optopolarimetric surveys as well as *Gaia* data on stellar parallaxes, we expect that local measurements of the GMF in the near future will become available.

Aims. Given such a set of (sparse) local GMF measurements, the question is how to optimally use them in backtracking UHECRs through the Galaxy. In this paper, we evaluate the reconstruction of the GMF, in a limited region of the Galaxy, through Bayesian inference, using principles of information field theory.

Methods. We employed methods of Bayesian statistical inference in order to estimate the posterior distribution of the GMF configuration within a certain region of the Galaxy from a set of sparse simulated local measurements. Given the energy, charge, and arrival direction of a UHECR, we could backtrack it through GMF configurations drawn from the posterior, and hence calculate the probability distribution of the true arrival directions on the PoS, by solving the equations of motion in each case.

Results. We show that, for a weakly turbulent GMF, it is possible to correct for its effect on the observed arrival direction of UHECRs to within $\sim 3^\circ$. For completely turbulent fields, we show that our procedure can still be used to significantly improve our knowledge on the true arrival direction of UHECRs.

Key words. magnetic fields – (ISM:) cosmic rays – turbulence

1. Introduction

The identification of ultra-high-energy cosmic ray (UHECR) sources remains one of the central questions in modern high-energy astrophysics. A possible resolution to this problem would shed light on the astrophysical mechanism that produces UHECRs, as well as their composition (either protons or heavier nuclei), which at high energies still remains a subject of debate. Knowledge of their composition may, in turn, provide insights into high-energy particle phenomenology at energies not yet probed by particle accelerators (Pavlidou & Tomaras 2019; Romanopoulos et al. 2022a; Romanopoulos et al. 2022b).

Despite the fact that several theoretical candidates for UHECR sources have been proposed (Bhattacharjee & Sigl 2000; Torres & Anchordoqui 2004), the identity of their sources remains elusive. The primary reason for this is that UHECRs are charged particles and are hence deflected by the Galactic magnetic field (GMF) as well as the intergalactic magnetic field. Even if several of the detected UHECRs originate from a single bright, nearby cosmic ray source di Matteo et al. 2023, their arrival directions would be very spread out in the sky, and any residual clustering would be centered off-source, due to

magnetic deflections. This is unlike the case of photons or neutrinos, in which positional identifications events with their likely sources can be attempted even with very low-number statistics.

The main difficulty in resolving the GMF stems from the fact that 3D tomographic realizations of the intervening magnetic fields are notoriously difficult to acquire. Specifically for the GMF, most observables that are currently available are integrated along the line of sight (LoS). Due to this limitation, and specifically for the GMF, the principal approach (Takami & Sato 2010) is relying on GMF models that are acquired by parameter-fitting three separate components – a toroidal, a poloidal, and a Gaussian random field (Jansson & Farrar 2012a; Jansson & Farrar 2012b; Sun et al. 2008; Sun & Reich 2010).

Nonetheless, it is possible to acquire direct information regarding the 3D structure of the GMF. *Gaia* data on stellar distances have localized more than a billion stars in the Galaxy by accurately determining stellar parallaxes (Gaia Collaboration 2016, 2021; Bailer-Jones et al. 2021). This data – in addition to other available spectroscopic data – has been used in order to construct 3D tomographic maps of the dust density distribution of certain Galactic regions (Green et al. 2019; Lallement et al. 2019; Leike & Enßlin 2019). These reconstructions, however, do

not constrain the magnetic field that is of primary interest in cosmic ray physics.

Nonetheless, there do exist probes that provide information regarding the structure of the GMF in 3D. For instance, the linear polarization of starlight is of particular interest; while starlight starts off unpolarized from its source, it will usually acquire a linear polarization until it is observed on Earth, due to dichroic absorption by dust particles aligned with the ambient magnetic field (Andersson et al. 2015).

Upcoming optopolarimetric surveys, such as PASIPHAE (Tassis et al. 2018; Maharana et al. 2021; Maharana et al. 2022; Magalhães 2012), are expected to provide a large number of high-quality stellar polarization measurements on more than a million stars. With stellar distances also known from the *Gaia* survey, this data can be used to provide direct tomographic measurements of the plane of the sky (PoS) component of the GMF at the location of the dust clouds (Panopoulou et al. 2017; Pelgrims et al. 2023; Davis 1951; Chandrasekhar & Fermi 1953; Skalidis et al. 2021). Used jointly with available LoS information (see for example Tahani et al. 2022a, Tahani et al. 2022b), we can expect to have local and sparse data of the GMF in the near future, which could be used in order to provide a 3D tomographic map of a particular region of interest. Given such a map, one can then backtrack UHECRs through that region, thus improving the localisation of their source on the sky, modulo the effects of the intergalactic magnetic field. To be more specific, one might be interested in reconstructing the region of the GMF through which the UHECRs “hotspots” (Kawata et al. 2019; Abbasi et al. 2014; Pierre Auger Collaboration et al. 2017) have to travel.

In this paper, we address the problem of reconstructing the posterior density function (PDF) for the true arrival directions of UHECRs, using local, sparse, and (statistically) uniformly distributed GMF measurements. In essence, we are interested in solving the inverse problem in a Bayesian setting, wherein one is given local data and tasked with calculating the posterior distribution of the GMF configurations in the region of interest. For Bayesian field inference, information field theory (IFT) was developed Enßlin et al. 2009, and demonstrated in a number of contexts Enßlin 2022. Here, we use an adapted version of the algorithm used in Leike & Enßlin 2019, after generalising it in order to make it applicable to divergence-free vector fields. Using the reconstructed posterior distribution over GMF realizations, we correct for the effect of the GMF on the observed arrival direction of the UHECR.

The work is structured as follows: in Sect. 2, we describe the Bayesian setting that is employed, and set out the principal components of the algorithm that forms the core of our reconstruction scheme, as well as describe our assumptions. In Sect. 3, we present the main results of this paper, by applying our reconstruction scheme to magnetic fields that are locally and sparsely sampled, for different relative strengths of the turbulent and the uniform component, as well as different sampling intervals. In Sect. 4, we summarize and discuss our conclusions.

2. Methodology

In Bayesian inference for continuous signals, we are in general interested in reconstructing the posterior probability distribution, $\varphi(\mathbf{x})$, defined over a domain, \mathcal{V} , from a given data set, d :

$$P(\varphi|d) \propto P(d|\varphi)P(\varphi). \quad (1)$$

The distribution, $P(d|\varphi)$, is the likelihood, and quantifies how likely acquiring d as our data is, given that the configuration is

$\varphi(\mathbf{x})$. The function, $P(\varphi)$, is the prior probability density, containing our information on $\varphi(\mathbf{x})$ before d is taken into account. The proportionality factor in Eq. (1) is determined by normalising the posterior to unity.

Mathematically, $\varphi(\mathbf{x})$ is a continuous function, and the distributions $P(\varphi)$, $P(\varphi|d)$, and $P(d|\varphi)$ are functionals. For example, an important case is the Gaussian distribution,

$$\mathcal{G}(\varphi - m, \Phi) \equiv \frac{\exp\left[-\frac{1}{2}(\varphi - m)\Phi^{-1}(\varphi - m)\right]}{|2\pi\Phi|^{\frac{1}{2}}} \quad (2)$$

where $m(\mathbf{x}) = \langle\varphi(\mathbf{x})\rangle$ is the mean. The quantity $\Phi(x_1, x_2) = \langle\delta\varphi(x_1)\delta\varphi(x_2)\rangle$ is the covariance of the distribution, where we have defined $\delta\varphi(x) \equiv \varphi(x) - m(x)$. $|\Phi|$ denotes the determinant of Φ . In Eq. (2), there is also an implicit integration, that is,

$$\begin{aligned} (\varphi - m)\Phi^{-1}(\varphi - m) \\ \equiv \int d^3x d^3x' [\varphi(\mathbf{x}) - m(\mathbf{x})]\Phi^{-1}(\mathbf{x}, \mathbf{x}')[\varphi(\mathbf{x}') - m(\mathbf{x}')]. \end{aligned}$$

In order to avoid cluttered notation, this integration will always be written implicitly.

2.1. Prior

The signal that we are interested in reconstructing is the GMF, $\mathbf{B}(\mathbf{x})$, which is a vector quantity. Following Eq. (2), the prior is written as

$$\mathcal{G}(\delta\mathbf{B}(\mathbf{x}), M) \equiv \frac{\exp\left[-\frac{1}{2}\delta B_i M_{ij}^{-1} \delta B_j\right]}{|2\pi M|^{\frac{1}{2}}}, \quad (3)$$

where $\delta B_i(\mathbf{x}) = B_i(\mathbf{x}) - \langle B_i(\mathbf{x})\rangle$, with $\langle B_i(\mathbf{x})\rangle$ denoting the mean of the distribution.

Here the indices correspond to individual components of the vector $\mathbf{B}(\mathbf{x})$, and the Einstein summation convention is assumed, and \mathbf{B}_0 is the mean field. Here, the covariance matrix inherits two indices, since by definition

$$M_{ij}(\mathbf{x}, \mathbf{x}') \equiv \langle\delta B_i(\mathbf{x})\delta B_j^*(\mathbf{x}')\rangle. \quad (4)$$

In the absence of information regarding the geometry and statistics of the GMF in the region of interest, we modeled the prior as being statistically isotropic and homogeneous. In Fourier space, this is equivalent to assuming that the two-point correlation function Eq. (4) takes the form

$$\langle\delta\hat{B}_i(\mathbf{k})\delta\hat{B}_j^*(\mathbf{k}')\rangle = \frac{(2\pi)^3}{2}\mathcal{P}_{ij}(\mathbf{k})\delta^{(3)}(\mathbf{k} - \mathbf{k}')P(k), \quad (5)$$

where

$$\mathcal{P}_{ij}(\mathbf{k}) \equiv \delta_{ij} - \hat{k}_i \hat{k}_j \quad (6)$$

is the transverse projection operator, \hat{k}_i denotes the unit \mathbf{k} -vector in the i -th direction, δ_{ij} is the kronecker delta, $\delta^{(3)}(\mathbf{k} - \mathbf{k}')$ is the three dimensional Dirac delta function, and $\delta\hat{B}_j^*(\mathbf{k})$ denotes the complex conjugate of $\delta\hat{B}_j(\mathbf{k})$. Furthermore, the norm of the \mathbf{k} -vector is henceforth denoted as k . The function, $P(k)$, is the magnetic power spectrum, and needs to be inferred as well. Equation (5) assumes that no prior expectation of the presence

Table 1. Parameters that define the power spectrum prior.

Parameter	Distribution	Mean	Standard deviation
Total offset (\mathbf{B}_0)	Not-applicable	0	Not-applicable
Total offset st. dev.	Log-normal	3 μG	1 μG
Total spectral energy	Log-normal	1 μG	1 μG
Spectral index	Normal	- $\frac{11}{3}$	1
Int. Wiener process amplitude	Log-normal	1.5	1

of the helicity of any sign exists¹. We note that magnetic helicity in the reconstructed magnetic field is not thereby excluded, but just needs to be requested by the data and will not be enforced by the prior.

Instead of inferring $\mathbf{B}(\mathbf{x})$ directly, it is more practical to work with a latent vector field, $\varphi(\mathbf{x})$, that does not satisfy the divergence free condition and that thus contains more degrees of freedom than needed. The Fourier space correlation structure for φ will be assumed to be isotropic and homogeneous, and this will take the form

$$\langle \delta\hat{\varphi}_i(\mathbf{k})\delta\hat{\varphi}_j^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_{ij} \delta^{(3)}(\mathbf{k} - \mathbf{k}') P(k), \quad (7)$$

with the understanding that $\mathbf{B}(\mathbf{x})$ and $\varphi(\mathbf{x})$ are related to each other by an application of $\mathcal{P}_{ij}(\mathbf{k})$ by

$$\hat{B}_i(\mathbf{k}) \equiv \frac{3}{2} \mathcal{P}_{ij}(\mathbf{k}) \hat{\varphi}_j(\mathbf{k}) = \frac{3}{2} \left(\hat{\varphi}_i(\mathbf{k}) - [\hat{\varphi}_j(\mathbf{k}) k_j] \frac{k_i}{k^2} \right), \quad (8)$$

in Fourier space. Are these three words – “in Fourier space” – connected to the previous equation, or are they not supposed to be here? This definition is such that the resulting field, \mathbf{B} , is guaranteed to be divergence-free, since then $k_i \hat{B}_i(\mathbf{k}) = 0$ in harmonic space, implying that $\nabla \cdot \mathbf{B} = 0$ as required. The 3/2 factor compensates for the loss of power that the subtraction of degrees of freedom causes. This is justified by the original a priori assumption of statistical isotropy for φ , which leads to equipartition along the three directions in Fourier space. Equation (8) is in accordance with the correlation structure assumed in Eq. (5) (Jaffe et al. 2012).

The power spectrum, $P(k)$, is unknown and needs to be inferred as well. It is modeled as the sum of a power law component and an integrated Wiener process component (for details regarding the generative model for power spectra, the reader should refer to Arras et al. 2022). The parameters that define it, as well as their respective prior PDFs, are:

1. The total field offset for all field components. This controls the mean value around which the random field fluctuates, \mathbf{B}_0 . We set a value of 0, which means that in the absence of data the mean overall field is assumed to be zero-centered.
2. The standard deviation of the total offset. This is a random variable with a prior log-normal distribution. We set its mean at 3 μG and its own standard deviation at 1 μG . This reflects our current understanding of typical GMF values in the interstellar medium that a (randomly oriented) mean magnetic field of this typical strength within the reconstructed volume is conceivable, but not enforced.

¹ Helicity would add a complex, antisymmetric term of the form $i\varepsilon_{ijl}k_l H(k)$, with $H(k)$ being the real helicity spectrum, which satisfies $-P(k) < H(k) < P(k)$.

3. The total spectral energy². This controls the amplitude of the fluctuations in configuration space. Its prior probability distribution is log-normal, with a mean and standard deviation set to unity. This is essentially the standard deviation of the field values calculated over the complete set of voxels.
4. The spectral index; the exponent of the pure power law component. Its prior distribution is normal, with a mean set as the Kolmogorov index, $-11/3$, and we assumed a unit standard deviation.
5. Amplitude of the integrated Wiener process component, controlling the deviations from a pure power law.

Its prior distribution is log-normal, with a mean and standard deviation equal to 1.5 and 1, respectively. These parameters were chosen so as to allow for deviations from a pure power law without destroying the approximately power-law overall behavior of the power spectrum. “approximately power-law overall behavior” does not sound right. Do you mean power-law-like behavior perhaps? Please rephrase.

The parameters of the amplitude model were assumed to be relative to a unit-less power spectrum; in other words, the parameters were assumed to be agnostic with regard to changes in the volume of the target subdomain, \mathcal{V}^3 . In Table 1, we summarise our choice for the probability distributions of the parameters of the power spectrum, as well as their respective means and standard deviations.

In closing the discussion on the prior, we consider the implementation of anisotropy. In general, the GMF in the region within which we wished to reconstruct it was assumed to admit the two-component structure

$$\mathbf{B}(\mathbf{x}) = \mathbf{B}_0 + \mathbf{b}(\mathbf{x}), \quad (9)$$

where \mathbf{B}_0 is a uniform field parallel to the PoS. Since the observed arrival velocity of the UHECR was chosen to be parallel to the LoS (see Sect. 3), the PoS component of \mathbf{B}_0 dominates the UHECR’s deflection and the LoS component is hardly constrained. Additionally, $\mathbf{b}(\mathbf{x})$ is a fluctuating field with zero mean, which can be physically interpreted as the turbulent field. Our prior is agnostic to the direction and magnitude of \mathbf{B}_0 , and inferring it from the data is a central task. The relative strength of the two components is quantified by the turbulent-to-uniform ratio, λ , defined as

$$\lambda \equiv \frac{b_{\text{rms}}}{B_0}, \quad (10)$$

where B_0 is the norm of \mathbf{B}_0 , and b_{rms} is the root mean square value of the fluctuating component’s magnitude, in the domain \mathcal{V} . For $\lambda \gg 1$ we have strong turbulence, for $\lambda \ll 1$ we have

² Here, the term ‘energy’ is used in the context of signal processing, and it does not refer to physical energy, although the two are related.

weak turbulence, and for $\lambda \simeq 1$ we have intermediate turbulence. As we will demonstrate, λ is the main parameter that controls how well the sources of UHECRs can be localized by the use of local GMF data along, of course, with the sampling rate.

2.2. Likelihood

In order to construct the likelihood, we considered how the data is acquired from the true signal, \mathbf{B} . We assumed that the i -th data point is given by a measurement process of the form

$$\mathbf{d}^{(i)} = \int R(\mathbf{x}, \mathbf{x}_i) \mathbf{B}(\mathbf{x}) d^3x + \mathbf{n}^{(i)}. \quad (11)$$

The operator, R , is a map from the signal space to the data space. For this proof of concept work, we made the simplifying assumption that local measurements of field components are possible, in other words $R(\mathbf{x}, \mathbf{x}_i) = \delta^{(3)}(\mathbf{x} - \mathbf{x}_i)$. In practice, most astrophysical measurements of magnetic fields (Faraday rotation, UHECR deflection, etc.) provide at least a LoS, particle trajectory, or even sub-volume averaged field strength. This only complicates the numerical reconstruction of fields, but does not imply conceptual changes in the formalism.

Thus, since we assumed local measurements, R was here assumed to be a mask operator. In practice, since the mock signal is discretized into voxels, R is simply a very sparse matrix. The full vector, \mathbf{d} , is therefore a concatenation of a number of 3D vectors known on a finite number of points inside the domain wherever we have measurements, and undefined anywhere else. Simply put, then, the operator, R , merely acts a selection operator that picks out the values of $\mathbf{B}(\mathbf{x})$ where we have measurements. Physically, the assumption was that at the positions where GMF measurements can be obtained, all three components of $\mathbf{B}(\mathbf{x})$ can be measured. This can be accomplished, for example, by Zeeman observations for the LoS component of the magnetic field. The PoS component of the magnetic field can be recovered through a combination of stellar optopolarimetry and stellar distance measures. Optopolarimetric measurements of stars of known distances can be decomposed to the contributions of individual clouds along the LoS (Pelgrims et al. 2023). Then, the average direction of polarization angles induced on starlight by dichroic absorption due to a single cloud reveals the direction of the magnetic field in that cloud. The dispersion of these directions can yield a measurement (within a factor of two) of the magnetic field strength in the same cloud (Chandrasekhar & Fermi 1953; Davis 1951; Skalidis et al. 2021; Skalidis & Tassis 2021).

Additionally, the vector, $\mathbf{n}^{(i)}$, was random noise added to the i -th noiseless data point. It was assumed to be drawn from a Gaussian distribution with a standard deviation equal to half the root-mean-square magnitude of the mock signal field. Physically, $\mathbf{n}^{(i)}$ represents the measurement error, assuming an average signal-to-noise ratio S/N of two.

The likelihood was then calculated by marginalizing over the noise:

$$\begin{aligned} P(\mathbf{d}|\mathbf{B}) &= \int \mathcal{D}\mathbf{n} P(\mathbf{d}, \mathbf{n}|\mathbf{B}) = \int \mathcal{D}\mathbf{n} P(\mathbf{d}|\mathbf{n}, \mathbf{B}) P(\mathbf{n}|\mathbf{B}) \\ &= \int \mathcal{D}\mathbf{n} \delta(\mathbf{d} - (R\mathbf{B} + \mathbf{n})) P(\mathbf{n}|\mathbf{B}) \\ &= \mathcal{G}(\mathbf{d} - R\mathbf{B}, N). \end{aligned} \quad (12)$$

where N is the noise covariance, and where we used the short-hand notation $R\mathbf{B} \equiv \int R(\mathbf{x}, \mathbf{x}_i)\mathbf{B}(\mathbf{x})d^3x$. We can write the likelihood in terms of φ by absorbing P_{ij} into R , that is, $R' \equiv R\mathcal{P}_{ij}$. Then, the likelihood becomes

$$P(\mathbf{d}|\varphi) = \mathcal{G}(\mathbf{d} - R'\varphi, N). \quad (13)$$

2.3. Approaching the posterior: Geometric variational inference

In the previous two sections, we described how we modeled the prior and the likelihood for our inference setting. From Eq. (1), the posterior is proportional to the product of the two.

Due to the fact that the magnetic power spectrum, $P(k)$, needs to be inferred along with the configuration of the GMF, this inference problem is nonlinear. A way to see this is that the magnetic field, $\mathbf{B}(\mathbf{x})$, couples to $P(k)$ through Eqs. (3) and (5). Moreover, the prior PDFs for the parameters that determine $P(k)$ are not Gaussian (with the exception of the spectral index), which is an additional source of nonlinearity. Furthermore, there is no small parameter that could be used in a perturbative analysis about a linear inference case. For this reason, a non-perturbative scheme, called geometrical variational inference (geoVI) developed by Frank et al. 2021 was utilized. In this section we motivate the basic premises of geoVI.

The idea is to approximate the true posterior, P , with an approximate one, Q . The approximate posterior, Q , is chosen such that the Kullback-Leibler divergence (Kullback & Leibler 1951),

$$D_{KL}(Q, P) \equiv \int dQ \log\left(\frac{Q}{P}\right), \quad (14)$$

between the actual posterior, P , and an approximate posterior, Q , is minimized. The main idea of geoVI is to achieve this minimization in a new coordinate system, chosen such that P – in the new coordinate system – locally closely resembles a normalized standard distribution. Once this is done, the approximating posterior, Q , is chosen to be of the form (3). Then, the mean and covariance are chosen as the parameters with respect to which the KL divergence is minimized.

A few more details on geoVI can be found in Appendix A and in Frank et al. 2021. Essentially, the algorithm provides approximate posterior samples that can follow the non-Gaussian structure of the posterior to a certain degree. geoVI can be invoked by the numerical information field theory (NIFTy) package in Python (Selig et al. 2013; Steininger et al. 2019; Arras et al. 2019). The input that is required is the likelihood and the prior of the original physical model, as described in Sects. 2.2 and 2.1, respectively.

3. Results

3.1. General procedure

We were now ready to construct a number of representative examples, in which we applied the procedure described in the previous section in order to reconstruct various assumed GMF geometries from a set of local, sparse, noisy, 3D GMF mock observations. All of the examples that we studied were created according to the following scheme:

- We defined the domain: We chose a cube, \mathcal{V} , of side length $L = 3$ kpc, with periodic boundary conditions (topologically a 3-torus), and evenly divided it into N^3 voxels. The number, N , defines our resolution.

- We produced a power spectrum, $P(k)$, by randomly sampling each of its defining parameters from its respective distribution (see Sect. 2.1).
- We produced the latent field, φ , with a correlation structure dictated by Eq. (7), with it taking a constant value over each voxel.
- We created the fluctuating part of the synthetic signal in Fourier space, by acting on φ with the transverse projection operator;

$$b_i(\mathbf{k}) = \mathcal{P}_{ij}(\mathbf{k})\varphi_j(\mathbf{k}).$$

After transforming into configuration space and computing the RMS value of the field, b_{rms} , we added the uniform component that lies along the $x - y$ plane (serving as our PoS) of a magnitude $\lambda^{-1}b_{\text{rms}}$, for a desired value of λ . We rescaled the total field such that the total RMS magnitude is 5 μG .

- We acted on the mock signal with a mask operator, which uniformly chooses a specified small fraction of voxels from the subdivided domain. This was done as follows: if N_d was the (approximate) number of data points we wished to have, then we acted on the original array with an operator that masked each voxel with a probability of $1 - N_d/N$. Roughly, then, the number of voxels that survived was N_d , by construction.

However, it is more useful to refer to the mean sampling ‘rate’; since by construction the distribution of data points is spatially homogeneous, the mean distance between the data points, as a function of N_d , is

$$\ell = \frac{L}{N_d^{1/3}}.$$

The mean sampling rate is thus defined as

$$k_{\text{sample}} \equiv \ell^{-1} = \frac{N_d^{1/3}}{L}. \quad (15)$$

In physical reality, the data points will be located wherever HI clouds exist in the region under study, which are not positioned uniformly within \mathcal{V} , and so the sampling rate will vary with vertical distance from the Galactic plane. Furthermore, there are also nonlocal measurements that, for example, average over a LoS. In future work, we shall also consider such effects.

- We added noise: To each data vector created in the previous step, we added a random vector sampled from a multivariate Gaussian distribution meant to represent observational error. The variance of the distribution is

$$\sigma^2 = \frac{B_{\text{rms}}^2}{4} = \frac{25}{4} (\mu\text{G})^2 \quad (16)$$

where B_{rms} is the RMS value of the total magnetic field’s norm within \mathcal{V} , which is set to 5 μG (see step 4). The denominator on the right-hand side is chosen such that the average S/N is two.

- We applied the geoVI algorithm to the data set (see Sect. 2.3). The output of the algorithm are samples of the posterior distribution.

An illustrative example is shown in Fig. 1, where we choose $k_{\text{sample}} = (600 \text{ pc})^{-1}$ as the mean sampling rate, and a turbulent-to-uniform ratio, $\lambda = 0.2$ (see Eq. (10)). In this case, the posterior

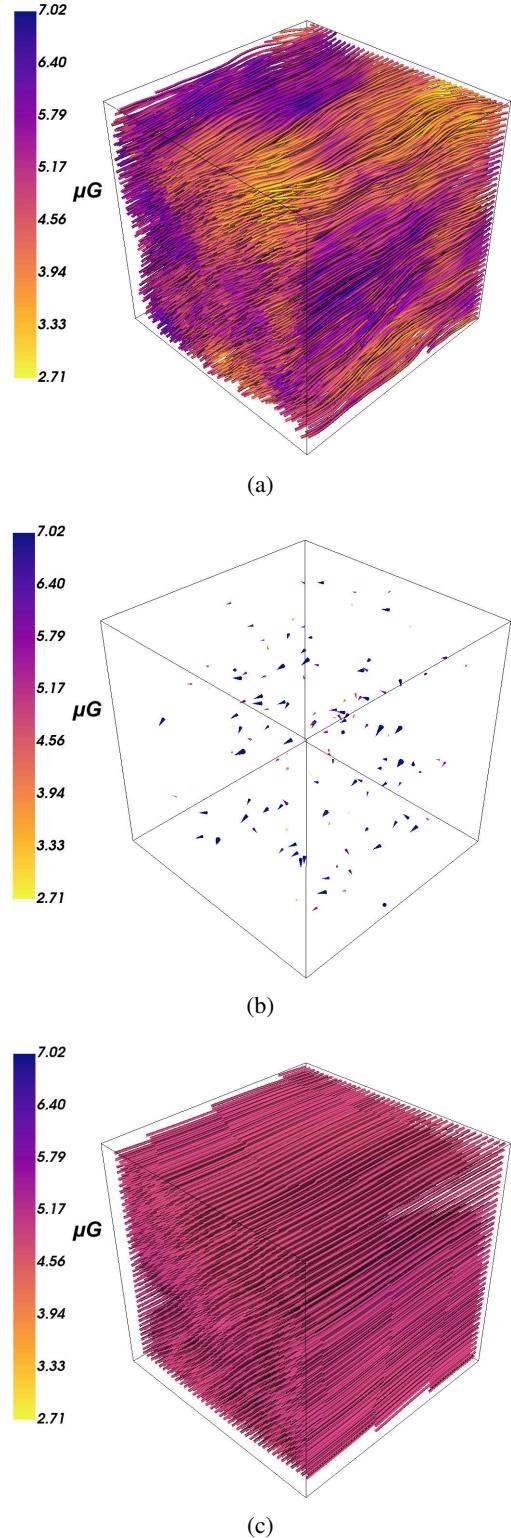


Fig. 1. Reconstruction of a 3D magnetic field within a cube of side $L = 3 = \text{kpc}$. *Top:* the ground truth; a uniform field parallel to the $x - y$ plane with magnitude 5 μG , plus a turbulent, random, field with an RMS magnitude of 1 μG , corresponding to $\lambda = 0.2$ in Eq. (10). *Middle:* local data sampled randomly, with a constant mean sampling rate (600 pc) $^{-1}$. The colormap is saturated at the maximum magnitude that appears in Fig. 1a. *Bottom:* the mean of the approximating posterior distribution attained via the geoVI algorithm based on the data provided in Fig. 1b.

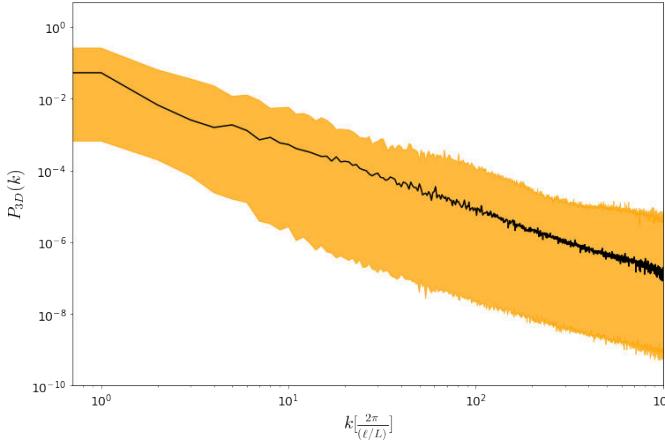


Fig. 2. The three-dimensional power spectrum, $P_{3D}(k)$, corresponding to the example showcased in Fig. 1. The black line is the power spectrum of the mock signal, while the orange envelope encompasses the posterior samples, providing the variance for the posterior mean power spectrum estimate. The wavevector, k , is given in units of $2\pi/N$. The large scatter is due to the fact that we have few and noisy data points, in a signal that is dominated by its uniform component, and thus the modes with $k > 0$ are poorly constrained.

mean essentially identifies the uniform component (Fig. 1c) provided the sparse and local data shown in Fig. 1b. The dominance of the ordered component results in a large scatter in the power spectrum (Fig. 2). To see why this is the case, we note that the zero mode of the field is

$$\hat{\mathbf{B}}(0) = \int \mathbf{B}(\mathbf{x}) d^3\mathbf{x} \propto \mathbf{B}_0. \quad (17)$$

In the scenario of a strong uniform component, the modes $\hat{\mathbf{B}}(\mathbf{k})$ with $k > 0$ are poorly constrained by the data, which mainly inform with respect to the zero mode.

It is important to note, however, that the UHECR's deflection is primarily influenced by the uniform component (or the zero mode), and is affected much less by the small-scale, turbulent fluctuations. Therefore, for the purposes of our inquiry, the recovery of the zero mode is enough to provide the leading order correction on the original arrival direction of UHECRs. In the next section, we apply the geoVI algorithm as exemplified in Fig. 1 specifically to the case of backtracking UHECRs back to their sources by utilising local data.

3.2. Correcting the arrival directions of UHECRs

In the previous section, we demonstrated the results of the geoVI algorithm applied to sparse and local measurements of a magnetic field in the case of a strong uniform component. However, the quality of the reconstruction should be judged in context, as different degrees of resolution are required in different applications. In fact, in this context, the GMF reconstruction is a means to an end, the latter being correcting for the effect of the GMF on the arrival direction of the UHECRs, given their arrival direction and energy as observed on Earth.

Stated explicitly, the question is, given an observed arrival direction, $\hat{\mathbf{v}}^{obs}$, of a charged particle of known lab-frame energy and charge, how to get the posterior distribution for its original, extragalactic direction, assuming we are provided with sparse and local measurements of the GMF within a region, \mathcal{V} , within which it traveled. Since the geoVI algorithm returns as output

sample configurations of the GMF drawn from the posterior, we can reconstruct the UHECR's path through each sample by solving the equations of motion backward (for details on how this is carried out in practice, see Appendix B). In the end, we are left with a distribution for possible paths, all converging to the velocity observed on Earth.

For the situation presented in Fig. 1, we considered a particle of charge e with final position, $\mathbf{r}_i = (L/2, L/2, 0)^T$ (Earth's location inside \mathcal{V}), velocity direction, $\hat{\mathbf{v}}^{obs} = (-90^\circ, 45^\circ)$ in Galactic coordinates, and energy in the lab frame, $E = 5 \times 10^{19}$ eV, that was backtracked through the true GMF, as well as samples of the posterior distribution of the GMF given the local and sparse data shown in Fig. 1b. In Fig. 3, in Galactic coordinates, the observed arrival direction (black star), true arrival direction (red star), and posterior samples (dots in the viridis colormap), are shown. The colormap corresponds to the mean posterior distribution as inferred from our samples using the IFT-based density estimator, DENSe (Guardiani et al. 2022). The advantage of this method is that the parameters of the kernel are inferred instead of assumed, and the result of this algorithm are samples of possible distributions drawn from the posterior distribution of distributions, given the samples³. From the former, the mean is drawn as our estimate for the underlying distribution, shown in the colormap of Fig. 3.

It can be seen that we were able to substantially correct for the effect of the GMF. For the purpose of comparison, we also include the respective result obtained via two different, much simpler, reconstruction methods: 1) merely taking the vector mean of the data points (blue star), and 2) a nearest-neighbor estimate wherein one assumes that at each point in space the value of the GMF is that given by the nearest available data point (pink star). While all three of the methods are able to correct for the effect of the GMF by essentially picking out the zero mode – which in this case predominantly affects the UHECR paths – the IFT-based method is a statistically rigorous way to perform the inference, as it also provides a quantification of the inference's uncertainty, a feature that the simpler methods lack. In addition, the simple methods assume either a low λ , or data points that are populated densely enough – something that cannot be expected from the distribution of dust clouds in real life applications.

In order to quantify the improvement in our knowledge, we used the Mahalanobis distance (Mahalanobis 1936) between a given arrival direction, $\hat{\mathbf{v}}$, and the posterior samples acquired from the geoVI algorithm,

$$d_M[\hat{\mathbf{v}}; P] = \sqrt{(\hat{\mathbf{v}} - \hat{\mu})^T S^{-1} (\hat{\mathbf{v}} - \hat{\mu})}, \quad (18)$$

where S^{-1} and $\hat{\mu}$ are the inverse covariance matrix and mean of $P(\hat{\mathbf{v}}|d)$, respectively. Essentially, d_M measures how many standard deviations $\hat{\mathbf{v}}$ is located away from $\hat{\mu}$. If $\hat{\mathbf{v}}^{true}$ and $\hat{\mathbf{v}}^{obs}$ are the true and observed arrival directions, respectively shown in Figs. 1 and 3, we calculated that $d_M[\hat{\mathbf{v}}^{true}; P] = 1.6\sigma$ and $d_M[\hat{\mathbf{v}}^{obs}; P] = 22.8\sigma$.

We also quantified our results using a different metric. Let $\hat{\mathbf{v}}^{rec}$ denote the arrival direction as obtained by backtracking through a reconstructed magnetic field. If we define the angle

$$\theta' = \cos^{-1}(\hat{\mathbf{v}}^{true} \cdot \hat{\mathbf{v}}^{rec}), \quad (19)$$

we are interested in the posterior mean of θ' , henceforth referred to as $\theta \equiv \langle \theta' \rangle$.

³ In addition to the reference provided, the reader is advised to look at <https://ift.pages.mpcdf.de/public/dense/> for further information.

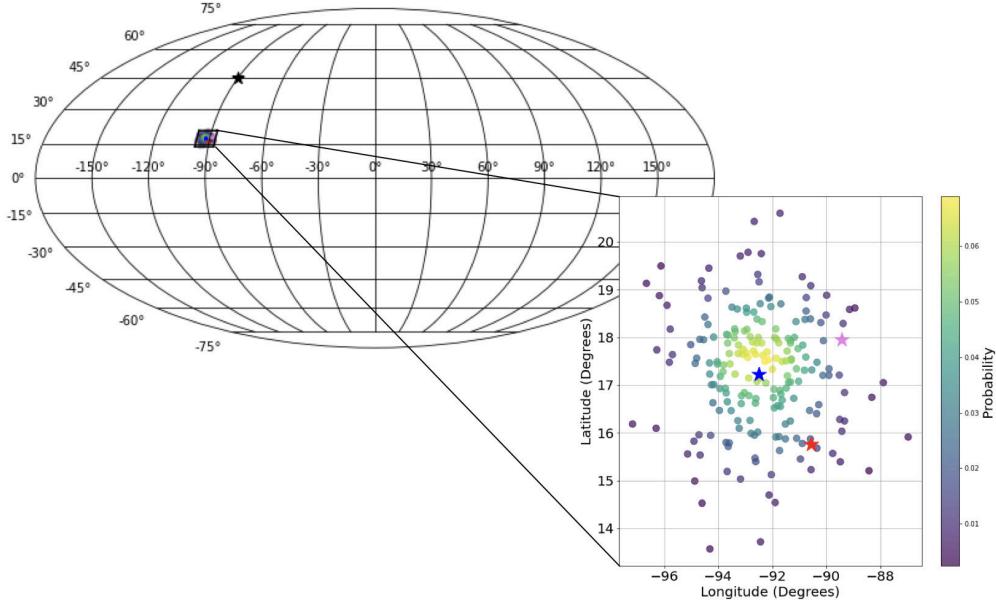


Fig. 3. Plane of the sky projections of the UHECR arrival directions as inferred given local and sparse GMF data based on 100 posterior samples (viridis colormap). The GMF reconstruction problem is the one shown in Fig. 1. The red star denotes the true arrival direction, while the black star denotes the observed arrival direction, and they are found 1.6σ and 22.8σ from the posterior's mean, respectively. For comparison, we also inferred the UHECR arrival direction using a simple vector mean of all the GMF data points (blue star), as well as a nearest-neighbor estimate (pink star), where at each point the GMF was assumed to be that dictated by the nearest available data point. These simple reconstruction methods do not provide an uncertainty quantification. In addition, they tacitly assume either a low λ , or data points that are populated densely enough – something that cannot be expected from the distribution of dust clouds in real life applications.

3.3. Effect of the sampling rate

In order to see how local sampling of the GMF might help us resolve the UHECR initial directions in general, the same backtracking procedure that led to Fig. 3 was carried out for an ensemble of 100 GMF configurations, for three different values of the mean sampling rate; $k_{\text{sample}} = (300 \text{ pc})^{-1}$, $(425 \text{ pc})^{-1}$, and $(600 \text{ pc})^{-1}$, and each time a different field served as ground truth and was reconstructed, via the process described in Sect. 3.1. The angle, θ , defined at the end of the last section, was calculated in each case, and the histograms for each sampling rate considered are shown in blue, in Fig. 4. For reference, we also calculated the angle between $\hat{\mathbf{v}}^{\text{true}}$ and the final velocity direction, $\hat{\mathbf{v}}^{\text{obs}}$,

$$\theta_0 = \cos^{-1}(\hat{\mathbf{v}}^{\text{true}} \cdot \hat{\mathbf{v}}^{\text{obs}}). \quad (20)$$

The histogram for θ_0 is shown in red for each sampling rate considered. At zeroth order, without attempting to reconstruct the GMF at all, the UHECR's path would not be altered, and so in this case $\hat{\mathbf{v}}^{\text{true}} \approx \hat{\mathbf{v}}^{\text{obs}}$ or $\theta \approx \theta_0$, trivially. Therefore, θ_0 is the zeroth-order approximation, without any data points considered. As can be clearly seen, the usage of the local and sparse data, \mathbf{d} , significantly corrects for the effect of the GMF on our knowledge of the UHECR arrival direction. Furthermore, as expected, increasing sampling rates tend to move the PDF for θ , $P(\theta)$, more toward $\theta = 0$. In the extreme case of a perfectly known GMF, $P(\theta)$ would be a delta function centered on zero.

These results are for the set value of $\lambda = 0.2$. As λ becomes larger, or as the fluctuating or turbulent component, $\mathbf{b}(\mathbf{x})$, in Eq. (9) becomes larger, our results should become increasingly worse. The reason for this is that if $\lambda \ll 1$ then the uniform component dominates, and we are required to only infer that, as the fluctuating part will only play a subdominant role in the deflection of the UHECR. As λ becomes larger, then the process will need to discern more irregular structures that play an

increasingly important role in deflecting the UHECR. This task is more difficult to achieve with sparse data. Additionally, the larger λ becomes, the less the UHECR path is expected to be deflected, as for dominating $\mathbf{b}(\mathbf{x})$ (which has zero mean) the UHECR will travel through regions that will partly cancel each other's effect on the path, resulting in a random walk, rather than a systematic deflection, about the true position of the source on the sky. Therefore, in Fig. 4, as λ increases, we expect $P(\theta)$ (blue histogram) to diffuse toward larger θ (owing to the increasing resolution that is required in that case). Conversely, the PDF for the angular distance between the final velocity and the true initial velocity (red histograms) is expected to shift toward smaller θ .

In Fig. 5, we tested this expectation, by performing the same calculation as in Fig. 4, but this time with increasing λ and keeping $k_{\text{sample}} = (600 \text{ pc})^{-1}$ fixed. The results confirm our intuition. We notice, however, that even in the limit where the turbulence is completely isotropic, $\lambda \rightarrow \infty$, the PDF resulting from considering the reconstruction is narrower than the one where only observed directions are used.

To see this, consider raising k_{sample} to $(300 \text{ pc})^{-1}$, which is not an unrealistic expectation for the density of HI clouds relatively close to the Galactic disk. Then, for a strongly turbulent case, $\lambda \rightarrow \infty$, our reconstruction can be seen in Fig. 6. The mean of the posterior as found by the geoVI algorithm, given the data of Fig. 6b, can be seen in Fig. 6c. In this example, the mean is essentially a ‘combed’ version of the true field, in that it does contain all the main structures and larger-scale features, but completely misses out the fluctuations below a certain length scale. It should be made clear, however, that the result of the reconstruction – as is the case with any inference problem – is not just the mean but rather the whole posterior distribution. The mean is drawn as a representative example of the posterior distribution, but in general samples from the whole posterior are utilized in

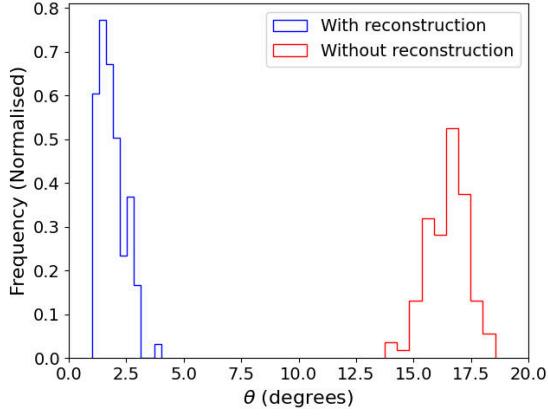
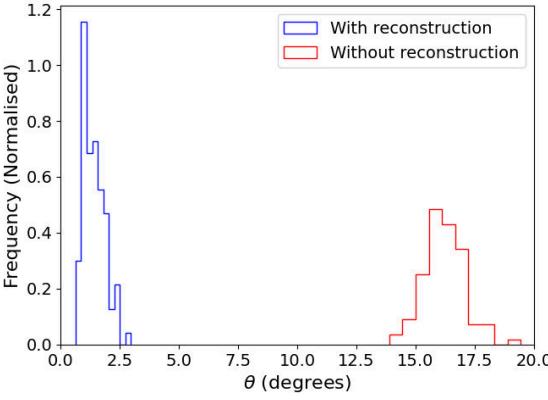
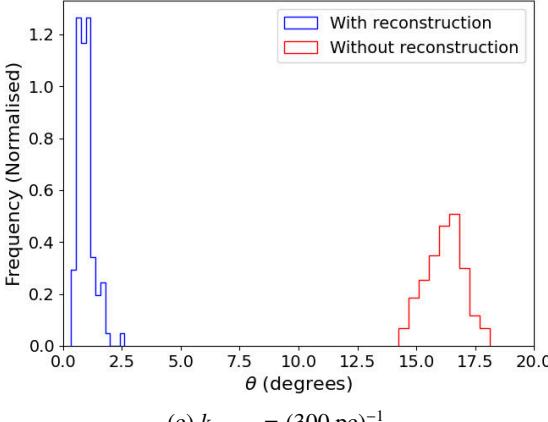
(a) $k_{\text{sample}} = (600 \text{ pc})^{-1}$ (b) $k_{\text{sample}} = (425 \text{ pc})^{-1}$ (c) $k_{\text{sample}} = (300 \text{ pc})^{-1}$

Fig. 4. Histograms of the deviation angle between the true arrival direction of UHECRs and the one obtained by backtracking through the magnetic field, as reconstructed using local data sampled at three different mean rates $k_{\text{sample}} = (600 \text{ pc})^{-1}$, $(425 \text{ pc})^{-1}$, and $(300 \text{ pc})^{-1}$ (blue). For each mean sampling rate, we considered 100 independent GMFs. The observed arrival direction, the charge, and lab-frame energy were assumed to be known and are $(0, 0, -1)^T$, $+e$, and $5 \times 10^{19} \text{ eV}$, respectively. The turbulent-to-uniform magnitude ratio (Eq. (10)) was set at $\lambda = 0.2$. For reference, we also plot the same angle by completely neglecting the effect of the GMF on the particle's trajectory (red). It can be seen that local and homogeneously distributed GMF data significantly improve our estimate of the true arrival directions of UHECRs, and that there is successive improvement as k_{sample} becomes larger, as expected.

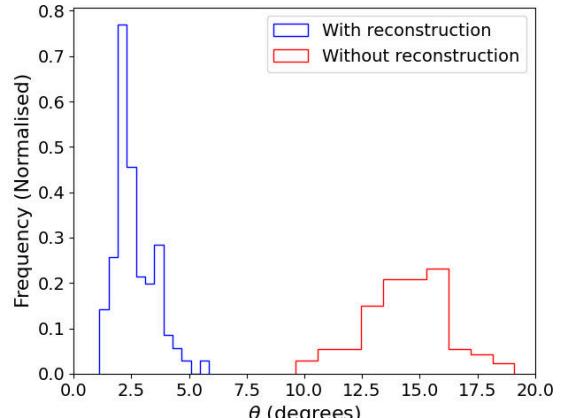
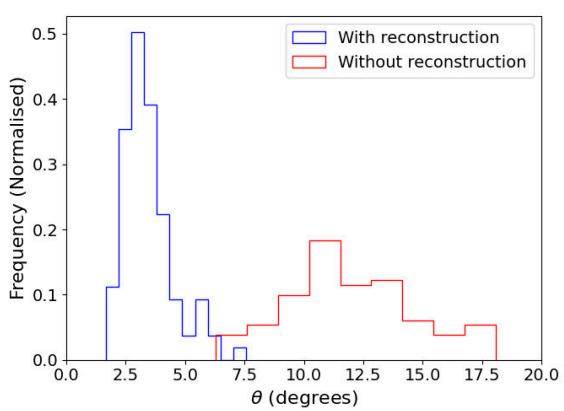
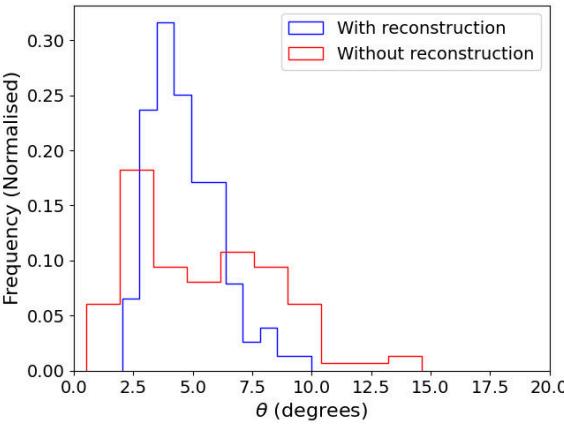
(a) $\lambda = 0.5$ (b) $\lambda = 1$ (c) $\lambda = \infty$

Fig. 5. Histograms of the deviation angle between the true arrival direction of UHECRs and the one obtained by backtracking through the magnetic field, as reconstructed using local data sampled at three turbulent-to-uniform ratios $\lambda = 0.5$, $\lambda = 1$, and $\lambda = \infty$, for λ defined in Eq. (10). For a value of λ , we considered 100 independent GMFs. The observed arrival direction, the charge, and lab-frame energy were assumed to be known and are $(0, 0, -1)^T$, $+e$, and $5 \times 10^{19} \text{ eV}$, respectively. The sampling rate (Eq. (15)) was set at $k_{\text{sample}} = 600 \text{ pc}$. For reference, we also plot the same angle by completely neglecting the effect of the GMF on the particle's trajectory (red). It can be seen that as λ becomes larger, the deviations in the original direction as obtained by backtracking through the reconstructed GMF and the true arrival direction become more dispersed, while the observed discrepancy diffuses toward smaller angles.

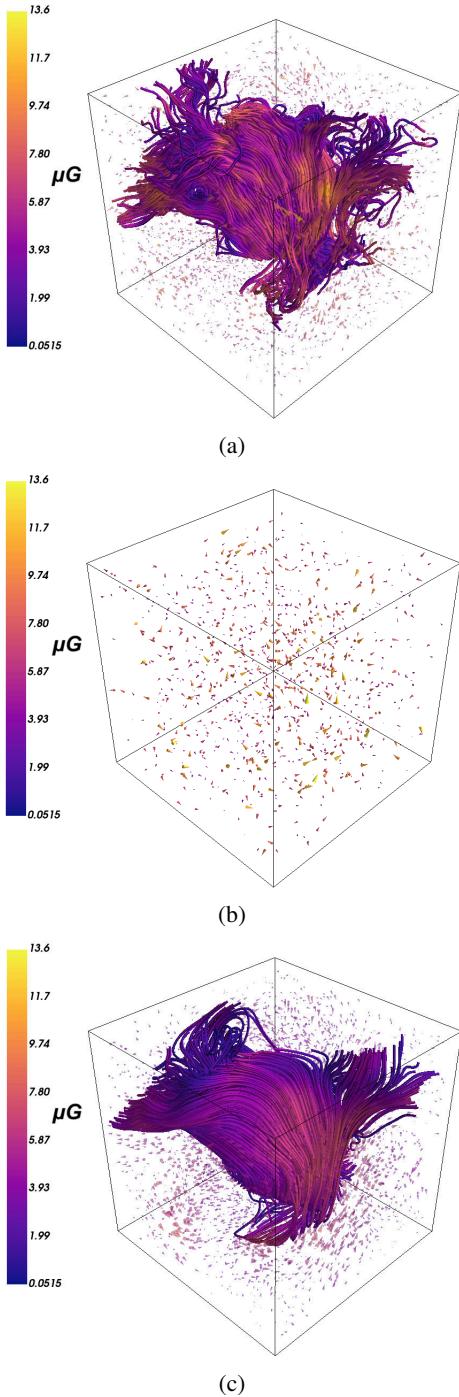


Fig. 6. Reconstruction of an isotropically turbulent 3D magnetic field within a cube of side $L = 3 \text{ kpc}$. *Top*: the ground truth; an isotropically turbulent GMF with an RMS magnitude of $5 \mu\text{G}$. The colormap denotes the vector's magnitude. *Middle*: local data sampled uniformly. The mean distance between each data point is 300 pc. The colormap is saturated at the maximum magnitude that appears in Fig. 6a. *Bottom*: the mean of the approximating Gaussian posterior distribution attained via the geoVI algorithm based on the data provided in Fig. 6b.

backtracking the cosmic rays and thus obtaining the posterior for the initial arrival directions.

Figure 7 shows the power spectrum defined via Eq. (5). We notice that in the case of isotropic turbulence, the modes with a nonzero wavevector are constrained much better from the data compared to the anisotropic case of a dominating zero mode.

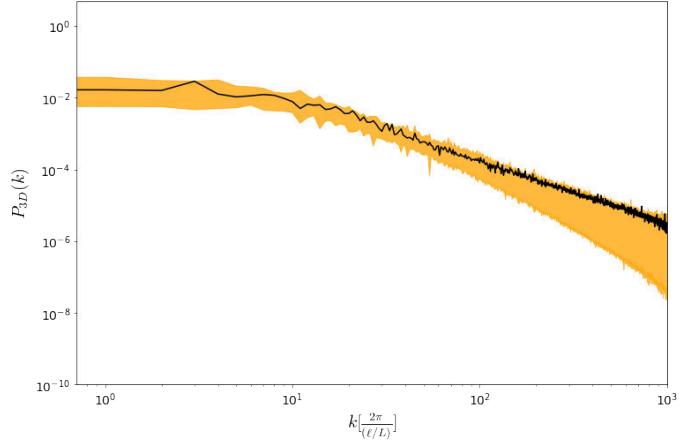


Fig. 7. As with Fig. 2, but for the case displayed in Fig. 6. In this case, all modes are relevant and well constrained, as opposed to the case of Fig. 2. Furthermore, we note that as k becomes larger the posterior of the power spectrum starts deviating from that of the signal, implying a loss of information at smaller lengthscales, as observed in Fig. 6.

As was done with the weakly turbulent example, we considered a UHECR with charge, e , observed arrival direction, $\hat{\mathbf{v}}^{\text{obs}} = (-90^\circ, 45^\circ)$ in Galactic coordinates, and energy in the lab frame, $E = 5 \times 10^{19} \text{ eV}$ through this strongly turbulent field. In Fig. 8, it can clearly be seen that in the extreme case of completely isotropic turbulence there is still a substantial deviation due to the presence of the GMF, and the zeroth-order approximation (that is, assuming $\hat{\mathbf{v}}^{\text{true}} \simeq \hat{\mathbf{v}}^{\text{obs}}$) is further away from the posterior's mean than the true arrival direction. In Fig. 8, we plot the posterior samples of the UHECR arrival directions given the local GMF data shown in Fig. 6b. The viridis colormap is the estimated posterior distribution using DENSe (see Sect. 3.2), based on the 100 posterior samples drawn using geoVI. As with Fig. 3, the red star denotes the true arrival direction, while the black star denotes the observed arrival direction. The Mahalanobis distance between these directions and the arrival directions sampled from the posterior is computed at 1.8σ and 6.7σ , respectively. As before, two simple GMF inference schemes are also employed, for comparison: 1) a simple vector mean of all the data points (blue star) and a nearest-neighbor regression (pink star). Since in this case the GMF's zero mode does not predominantly contribute in the UHECR's deflection, the simpler vector mean inference method fails completely, as it requires a dominant zero mode, or equivalently a low λ – a prior assumption that the geoVI method does not make. The nearest-neighbor inference scheme does perform well, but this also depends on the fact that in this case our data points are dense enough. As noted before, the IFT-based method improves upon the simpler methods by accounting for variation in the magnetic field and we improve upon the second by providing uncertainties (and having physically sound magnetic fields), and so the influence that a possible high λ and/or low sampling rate will have on our confidence in the suggested extragalactic origin is unknown. The IFT-based method presented in the work systematically takes care of this shortcoming.

4. Summary, conclusions, and outlook

4.1. Summary

In this work, we used 3D local and sparse mock observations of the GMF scattered across a cubic domain of 3 kpc sides within

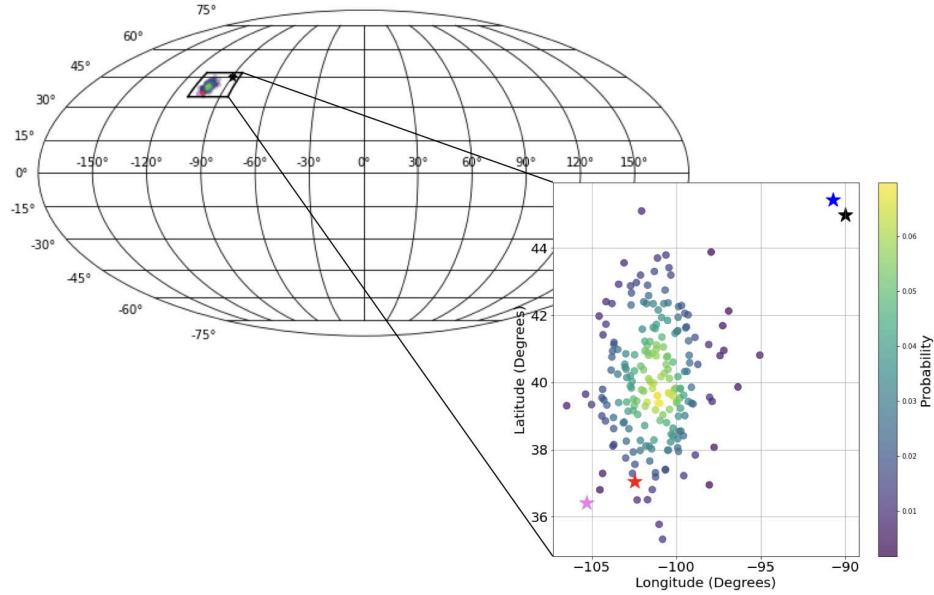


Fig. 8. As with Fig. 3, but for the turbulent magnetic field case displayed in Fig. 6.

the Galaxy, which are statistically uniformly distributed, in order to obtain the posterior distribution of the UHECR arrival directions before they entered the domain of influence of the GMF. We considered a two-component field, which was comprised of a uniform and random (turbulent) part, for various relative strengths between the two components. We used techniques from information field theory and geoVI in order to construct a systematic way of calculating sample configurations from the posterior PDF. We adapted the correlated field model of IFT to include the case divergence-free vector field such as the GMF, so that the divergence-free condition was also taken into account during the computation of the posterior samples.

Once the above framework was established, we used it to test the quality of our reconstruction by comparing it to the ground truth, for various cases of sampling rates and turbulent-to-uniform ratios. By backtracking a UHECR of known arrival direction and a rigidity (energy per unit charge) equal to 5×10^{19} V through sample configurations drawn from the posterior PDF for the GMF given the sparse and local mock data, we calculated their reconstructed direction, thus producing a posterior distribution for the true arrival directions. For an ensemble of GMFs sharing the same turbulence-to-uniform ratios and the same mean sampling rate, we were able to estimate the PDF of the mean angle between the true arrival direction and the ones obtained by backtracking through posterior samples, θ , and we were able to show that, in the case of weak turbulence, very modest sampling rates are needed in order for the predictions to not deviate from the true arrival direction by more than a few degrees in relative angle. In particular, for a mean distance of 600 pc between each measurement, the vast majority of the relative angles between the inferred arrival directions and the true ones are below $\sim 3^\circ$, for the cases considered. For increasing sampling rates, our results become increasingly better, as should be expected. For reference, completely neglecting the information provided by the local and sparse data, and thus also neglecting the effect of the GMF, results in a relative angle between the observed arrival direction and the true one (θ_0) with a mean value, $\theta_0 \sim 15^\circ$.

As soon as the turbulent component starts dominating over the uniform one, our results deteriorate, and in the extreme case

of fully isotropic turbulence there is a significant overlap between the PDF for θ and θ_0 . However, even in this extreme and unrealistic case, the local and sparse GMF data still provide a significant improvement in determining the true arrival directions, and by extension the location UHECR sources.

4.2. Conclusions

In conclusion, our results are briefly summarized as follows:

- We have developed a systematic framework, based on information field theory, that infers the posterior distribution of the true arrival directions of UHECRs, given sparse and local data for the GMF scattered within a region of the Galaxy.
- For weakly turbulent fields, and by using uniformly sampled local data of the GMF with an average S/N of two, we are able to correct for the effect of the GMF to within $\theta \sim 3^\circ$, where θ is the angle between the inferred arrival direction and the true arrival direction. The required mean distance between each measurement can be as large as 600 pc in this case. In applications to real data, the locations of the data points will coincide with the locations of molecular clouds, and a mean distance of ~ 600 pc is a rather conservative estimate.
- In the extreme but unrealistic case of an isotropically turbulent GMF, our results are worse compared to the case of a dominating structured field, but local data can still substantially improve our knowledge on the true arrival directions of UHECRs, and hence on the identity of their sources.
- Other simple inference methods, specifically a vector mean of the data points or a nearest-neighbor estimate of the GMF, require very weak turbulence and/or high sampling rates, and even in this case they do not quantify the uncertainty of the inference in a statistically rigorous way. These shortcomings are addressed by the IFT-based method introduced in this work, as it produces samples of the posterior distribution of arrival directions instead of a single estimate, and no prior assumption about the relative strength of the turbulent component or the sampling rate is required.

4.3. Final comments and outlook

In this work, the data points were chosen to have a constant mean sampling rate throughout the domain of study, with the mean distance between the measurements serving as an adjustable parameter. However, in real applications this parameter is not set by us, but rather by Nature, as our measurements will be localized wherever HI clouds exist inside the domain within which we wish to reconstruct the GMF. In particular, their spacing will depend on the distance from the Galactic plane. It might be the case that the assumption of statistical homogeneity of data introduces biases into our results. In future work, we will consider the case of inferring UHECRs' true arrival directions by using data from MHD simulations of Galactic evolution in Milky Way-like galaxies, where our mock observations will not be chosen at random by a predetermined distribution, but rather from the distribution of clouds as given by the simulation data. It should also be noted that 3D information on each synthetic data point is a rather idealized situation; real data will mainly consist of PoS information alone, often even integrated (with some structured weighting) along the LoS. However, in the small deflection case that is relevant for our problem, the PoS component is expected to dominate the UHECR path's deflection, since the observed velocity is parallel to the LoS. This issue will be addressed in detail in future work. Finally, we like to add that for any UHECR source that could be identified, the difference between the observed and initial travel directions of the corresponding UHECRs will then provide information on the intergalactic magnetic field.

Acknowledgements. A.T. and V.P. acknowledge support from the Foundation of Research and Technology – Hellas Synergy Grants Program through project MagMASim, jointly implemented by the Institute of Astrophysics and the Institute of Applied and Computational Mathematics. A.T. acknowledges support by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “Third Call for H.F.R.I. Scholarships for PhD Candidates” (Project 5332). V.P. acknowledges support by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant” (Project 1552 CIRCE). A.T. would like to thank Vincent Pelgrims, Raphael Skalidis, Georgia V. Panopoulou, and Konstantinos Tassis for helpful tips and stimulating discussions. G.E. acknowledges the support of the German Academic Scholarship Foundation in the form of a PhD scholarship (“Promotionsstipendium der Studienstiftung des Deutschen Volkes”).

References

- Abbasi, R. U., Abe, M., Abu-Zayyad, T., et al. 2014, *ApJ*, 790, L21
- Amari, S.-i. 2016, *Invariant Geometry of Manifold of Probability Distributions* (Tokyo: Springer Japan)
- Andersson, B. G., Lazarian, A., & Vaillancourt, J. E. 2015, *ARA&A*, 53, 501
- Arras, P., Baltac, M., Ensslin, T. A., et al. 2019, *Astrophysics Source Code Library*
- Arras, P., Frank, P., Haim, P., et al. 2022, *Nat. Astron.*, 6, 259
- Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, *AJ*, 161, 147
- Bhattacharjee, P., & Sigl, G. 2000, *Phys. Rep.*, 327, 109
- Chandrasekhar, S., & Fermi, E. 1953, *ApJ*, 118, 113
- Davis, L. 1951, *Phys. Rev.*, 81, 890
- di Matteo, A., Anchordoqui, L., Bister, T., et al. 2023, arXiv e-prints [arXiv:2302.04502]
- Enßlin, T. 2022, *Entropy*, 24, 374
- Enßlin, T. A., Frommert, M., & Kitaura, F. S. 2009, *Phys. Rev. D*, 80, 105005
- Frank, P., Leike, R., & Enßlin, T. A. 2021, *Entropy*, 23
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, 595, A1
- Gaia Collaboration (Brown, A. G. A., et al.) 2021, *A&A*, 649, A1
- Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, *ApJ*, 887, 93
- Guardiani, M., Frank, P., Kostić, A., et al. 2022, *PLoS ONE*, 17, e0275011
- Jaffe, T., Waelkens, A., Reinecke, M., Kitaura, F. S., & Ensslin, T. A. 2012, Hammurabi: Simulating polarized Galactic synchrotron emission, *Astrophysics Source Code Library*, [record ascl:1201.014]
- Jansson, R., & Farrar, G. R. 2012a, *ApJ*, 757, 14
- Jansson, R., & Farrar, G. R. 2012b, *ApJ*, 761, L11
- Kawata, K., di Matteo, A., Fujii, T., et al. 2019, in *International Cosmic Ray Conference*, 36, 36th International Cosmic Ray Conference (ICRC2019), 310
- Kullback, S., & Leibler, R. A. 1951, *Ann. Math. Statist.*, 22, 79
- Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, *A&A*, 625, A135
- Leike, R. H., & Enßlin, T. A. 2019, *A&A*, 631, A32
- Magalhães, A. M. 2012, in *Science from the Next Generation Imaging and Spectroscopic Surveys*, 7
- Mahalanobis, P. C. 1936, *Proc. Natl. Inst. Sci. (India)*, 2, 6
- Maharana, S., Kyriotakis, J. A., Ramaprakash, A. N., et al. 2021, *J. Astron. Telescopes Instrum. Syst.*, 7, 014004
- Maharana, S., Anche, R. M., Ramaprakash, A. N., et al. 2022, *J. Astron. Telescopes Instrum. Syst.*, 8, 038004
- Nocedal, Jorge, P., & Wright, S. J. 2006, *Large-Scale Unconstrained Optimization* (New York, NY: Springer New York), 164
- Panopoulou, G. V., Psaradaki, I., Skalidis, R., Tassis, K., & Andrews, J. J. 2017, *MNRAS*, 466, 2529
- Pavlidou, V., & Tomaras, T. 2019, *Phys. Rev. D*, 99, 123016
- Pelgrims, V., Panopoulou, G. V., Tassis, K., et al. 2023, *A&A*, 670, A164
- Pierre Auger Collaboration (Aab, A., et al.) 2017, *Science*, 357, 1266
- Romanopoulos, S., Pavlidou, V., & Tomaras, T. 2022a, arXiv e-prints [arXiv:2206.14837]
- Romanopoulos, S., Pavlidou, V., & Tomaras, T. 2022b, in *37th International Cosmic Ray Conference*, 12–23 July 2021. Berlin, 475
- Selig, M., Bell, M. R., Junklewitz, H., et al. 2013, *A&A*, 554, A26
- Skalidis, R., & Tassis, K. 2021, *A&A*, 647, A186
- Skalidis, R., Sternberg, J., Beattie, J. R., Pavlidou, V., & Tassis, K. 2021, *A&A*, 656, A118
- Steininger, T., Dixit, J., Frank, P., et al. 2019, *Ann. Physik*, 531, 1800290
- Sun, X.-H., & Reich, W. 2010, *Res. Astron. Astrophys.*, 10, 1287
- Sun, X. H., Reich, W., Waelkens, A., & Enßlin, T. A. 2008, *A&A*, 477, 573
- Tahani, M., Glover, J., Lupyciw, W., et al. 2022a, *A&A*, 660, L7
- Tahani, M., Lupyciw, W., Glover, J., et al. 2022b, *A&A*, 660, A97
- Takami, H., & Sato, K. 2010, *ApJ*, 724, 1456
- Tassis, K., Ramaprakash, A. N., Readhead, A. C. S., et al. 2018, ArXiv e-prints [arXiv:1810.05652]
- Torres, D. F., & Anchordoqui, L. A. 2004, *Rep. Progr. Phys.*, 67, 1663

Appendix A: Geometric variational inference (geoVI)

In this appendix we provide a brief step-by-step overview of the geoVI algorithm, which is the main algorithm used to approximate the posterior distribution of the magnetic field, given sparse and local data. The idea is to approximate the true posterior, P , with an approximate one, Q . The approximate posterior, Q , is chosen such that the Kullback-Leibler divergence (Kullback & Leibler 1951),

$$D_{KL}(Q, P) \equiv \int dQ \log\left(\frac{Q}{P}\right), \quad (\text{A.1})$$

between the actual posterior, P , and an approximate posterior, Q , is minimized. The main idea of geoVI is to achieve this minimization in a new coordinate system, chosen such that P - in the new coordinate system - locally closely resembles a normalized standard distribution. Once this is done, the approximating posterior, Q , is chosen to be of the form (3). Then, the mean and covariance are chosen as the parameters with respect to which the KL divergence is minimized.

1. First, a coordinate transformation, $\varphi = f(\xi)$, is performed, such that the new prior is Gaussian with unit covariance and zero mean, the standardized coordinate system. Henceforth, the measure $D\xi$ signifies integration over all possible configurations of the vector field, ξ .
2. In this new coordinate system, we calculate the Fisher information metric (Amari 2016), $\mathcal{M}(\xi)$, for the likelihood, marginalising the data and joining it with the unit prior metric, \mathbb{M} . Intuitively, this may be regarded as a metric over the statistical manifold associated with the likelihood $P(d|\xi)$.
3. We seek another transformation, g , on top of the original f , that turns $\mathcal{M} + \mathbb{M}$ into the Euclidean metric, locally. The motivation is that in this coordinate system, since the geometry of the statistical manifold is as simple as possible, the original posterior is more likely to be accurately described by a Gaussian. The accuracy of the step depends on the choice of the expansion point for the local transformation. It can be computed locally around the mean of the approximating Gaussian in this new coordinate system.
4. The KL-divergence (A.1) is minimized in the coordinate system to which g maps with respect to the mean referred to in the previous step, using a second-order quasi-Newton method called the Newton conjugate gradient (NewtonCG) (Nocedal & Wright 2006). This is achieved by drawing sample configurations and using them to compute the KL divergence, minimising it with respect to the mean.

Appendix B: Back-propagating the UHECRs through the GMF

The equations of motion for a relativistic charged particle of charge, q , in a static magnetic field in the lab frame, $\mathbf{B} = \mathbf{B}(\mathbf{x})$, are

$$\frac{d(\gamma m \mathbf{v})}{dt} = q \mathbf{v} \times \mathbf{B}, \quad (\text{B.1})$$

and

$$\frac{d(\gamma m c^2)}{dt} = 0, \quad (\text{B.2})$$

where γ is the particle's Lorentz factor, and \mathbf{v} its velocity. Equation (B.2) follows from the absence of an electric field. Substituting equation (B.2) into (B.3), we get

$$\frac{d\mathbf{v}}{dt} = \frac{qc^2}{E} \mathbf{v} \times \mathbf{B}, \quad (\text{B.3})$$

where E is the particle's lab-frame energy.

If δt is a small time interval, then the change in the velocity during the interval, δt , is

$$\delta\mathbf{v} = \mathbf{v}(t) - \mathbf{v}(t - \delta t). \quad (\text{B.4})$$

Using equation (B.3), we may write

$$\frac{\delta\hat{\mathbf{v}}}{\delta t} = \frac{qc^2}{E} \hat{\mathbf{v}} \times \mathbf{B}, \quad (\text{B.5})$$

where we divided both sides by $|\mathbf{v}| \simeq c$, and $\hat{\mathbf{v}} \simeq \mathbf{v}c^{-1}$ is the velocity's direction at any given time. Substituting equation (B.4) into (B.5) and solving for $\mathbf{v}(t - \delta t)$, we obtain

$$\hat{\mathbf{v}}(t - \delta t) = \hat{\mathbf{v}}(t) - \frac{Zec^2}{E} (\hat{\mathbf{v}}(t) \times \mathbf{B}) \delta t, \quad (\text{B.6})$$

where $q = Ze$, with e being the electron charge and Z the atomic number of the UHECR.

If we are also given the position of the UHECR at time t , and we wish to calculate it at time $t - \delta t$, then we may write

$$\mathbf{r}(t - \delta t) = \mathbf{r}(t) - \hat{\mathbf{v}}(t)c\delta t, \quad (\text{B.7})$$

where we once again made the assumption $|\mathbf{v}| \simeq c$ throughout the particle's path.

Therefore, if we are given the position, charge, lab-frame energy, and observed arrival direction of a UHECR, we can use equations (B.6) and (B.7) iteratively in order to solve the equations of motion numerically. We choose δt in the iterative process such that the length, $c\delta t$, is equal to the resolution; the total domain is subdivided into voxels of side length $c\delta t$, within which the GMF is assumed to be constant. For this work, this amounts to setting $c\delta t = 60$ pc, while the side of the total cubic domain is 3 kpc.

Finally, once the initial arrival direction is obtained - this happens when the coordinates of the UHECR's location exceed the boundaries of the domain - it is translated into galactic coordinates via the equations

$$b = \sin^{-1}(\hat{v}_z), \quad (\text{B.8})$$

$$\ell = \text{sign}(\hat{v}_y) \cos^{-1}\left(\frac{\hat{v}_x}{\cos(b)}\right),$$

where if $\ell < 0$, 2π is added to avoid negative angles.

Chapter 5

Nonparametric Bayesian Reconstruction of the GMF Using IFT for UHECR Backtracking (Part 2)

In this chapter, we continue our investigation of a Bayesian inference approach for reconstructing the three-dimensional structure of the GMF using synthetic local measurements. This study will naturally follow the developments presented in the previous chapter, but will extend the analysis already presented by incorporating more realistic assumptions. First, the local data GMF points are now projected on the POS, as stellar polarimetry does not provide access to the LOS dimension. However, in this work we update the likelihood such that idealised integrated measurements of the GMF which can become available via Faraday rotation measurements. In addition, the ground truth field is now provided by dynamo simulation of the Galactic ISM.

Our results demonstrate that, across all tested arrival directions, the algorithm significantly improves the corrects the effect of the GMF on the observed arrival direction of UHECRs with a rigidity of $E/Z = 5 \times 10^{19}$ eV, even when only local data are used. When integrated LOS data are included, the accuracy further increases, reducing the maximum error by a factor of about three in the most challenging regions of the sky that corresponds to the regions where the GMFs ordered component is perpendicular to the LOS.

Additionally, we have extended the analysis from the previous chapter to a more complex setting that includes non-uniform distribution of the local data points, so that the distribution of local GMF data obtained through stellar polarization tomography reflects a more realistic distribution of dust clouds, unlike the earlier work, where data was assumed to be homogeneously distributed. Additionally, using an MHD simulation as the ground-truth GMF enables us to assess the effectiveness of our Gaussian modeling approach in a more realistic scenario, where the statistical properties of the ground truth GMF to be inferred more closely resemble the statistical statistical properties of the true GMF.

Even in the absence of LOS information, our method corrects for GMF effects on UHECR paths with considerable accuracy, though the inclusion of LOS data substantially enhances the results. We are able to identify and analyse the regions where errors are expected to be greatest and demonstrate that, even in these cases, our IFT-based approach reduces angular discrepancies by a factor of three. By reconstructing the large-scale GMF features that predominantly influence UHECR deflection, we provide insights into the regions of the plane of the sky most susceptible to error. Furthermore, we show that we all also able to estimate the correlation length of the GMF. These results underscore the potential of our method to utilize future GMF data for more precise UHECR source identification.

Nonparametric Bayesian reconstruction of Galactic magnetic fields using information field theory

The inclusion of line-of-sight information in ultrahigh-energy cosmic-ray backtracing

Alexandros Tsouros^{1,2,*}, Abhijit B. Bendre^{3,4}, Gordian Edenhofer^{5,6,7}, Torsten Enßlin^{5,6}, Philipp Frank⁵, Michalis Mastorakis^{1,2}, and Vasiliki Pavlidou^{1,2}

¹ Department of Physics & ITCP, University of Crete, 70013 Heraklion, Greece

² Institute of Astrophysics, Foundation for Research and Technology-Hellas, Vasilika Vouton, 70013 Heraklion, Greece

³ Laboratoire d'Astrophysique, EPFL, 1290 Sauverny, Switzerland

⁴ Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, 56126 Pisa, Italy

⁵ Max Planck Institute for Astrophysics, Karl-Schwarzschild-Straße 1, 85748 Garching, Germany

⁶ Ludwig Maximilian University of Munich, Geschwister-Scholl-Platz 1, 80539 Munich, Germany

⁷ University of Vienna, Department of Astrophysics, Türkenschanzstrasse 17, 1180 Vienna, Austria

Received 11 March 2024 / Accepted 20 August 2024

ABSTRACT

Context. Ultrahigh-energy cosmic rays (UHECRs) are charged particles with energies surpassing 10^{18} eV. Their sources remain elusive because they are obscured by deflections caused by the Galactic magnetic field (GMF). This challenge is further complicated by our limited understanding of the 3D structure of the GMF because current GMF observations primarily consist of quantities that are integrated along the line of sight (LOS). Nevertheless, data from upcoming stellar polarization surveys along with Gaia stellar parallax data are expected to yield local GMF measurements.

Aims. This study is the second entry in our exploration of a Bayesian inference approach to the local GMF that uses synthetic local GMF observations that emulate forthcoming local GMF measurements, and attempts to use them to reconstruct its 3D structure. The ultimate aim is to trace back observed UHECRs and thereby update our knowledge about their possible origin.

Methods. In this proof-of-concept work, we assumed as ground truth a magnetic field produced by a dynamo simulation of the Galactic ISM. We employed methods of Bayesian statistical inference in order to sample the posterior distribution of the GMF within part of the Galaxy. By assuming a known rigidity and arrival direction of an UHECR, we traced its trajectory back through various GMF configurations drawn from the posterior distribution. Our objective was to rigorously evaluate the performance of our algorithm in scenarios that closely mirror the setting of expected future applications. In pursuit of this, we conditioned the posterior to synthetically integrated LOS measurements of the GMF, in addition to synthetic local plane of sky-component measurements.

Results. Our results demonstrate that for all locations of the observed arrival direction on the plane of sky, our algorithm is able to substantially update our knowledge on the original arrival direction of UHECRs with a rigidity of $E/Z = 5 \times 10^{19}$ eV, even without any LOS information. When the integrated data are included in the inference, the regions of the celestial sphere in which the maximum error occurs are greatly reduced. The maximum error is diminished by a factor of about 3 even in these regions in the specific setting we studied. Additionally, we are able to identify the regions in which the largest error is expected to occur.

Key words. astroparticle physics – ISM: magnetic fields – local interstellar matter

1. Introduction

Ultrahigh-energy cosmic rays (UHECRs) are cosmic rays with energies surpassing 10^{18} eV. Since they are charged, the Galactic magnetic field (GMF) deflects their paths, and this leads to a mismatch between the true and the observed arrival direction. It is a crucial challenge in the field of high-energy astrophysics to determine their origins. Successfully addressing this challenge could offer insights into astrophysical processes that generate UHECRs, and into their composition. Additionally, knowledge of UHECR sources would be a crucial ingredient in multessenger studies of high-energy systems (e.g., Fang & Murase 2018; Murase 2019).

Although numerous theoretical models have been proposed to explain the sources of UHECRs (e.g., Bhattacharjee & Sigl 2000; Torres & Anchordoqui 2004; Kotera & Olinto 2011), it has proven to be a complicated task to determine these sources concisely. The main challenge arises from the fact that UHECRs are charged particles, and they are deflected by both the GMF and the intergalactic magnetic field. As a result, even if multiple UHECRs were emitted from a single, intense, and proximate cosmic-ray source (di Matteo et al. 2023), their trajectories would be dispersed across the plane of the sky (POS). Consequently, any UHECR hotspot would not align with the source, but would be displaced away from it due to systematic deflections by the ordered component of the GMF, in addition to being spread out due to the random deflections by the turbulent component of the GMF. This situation is different from

* Corresponding author; tsouros@physics.uoc.gr

that of photons or neutrinos, for which it is more straightforward to establish a connection between observed events and their probable sources, even in the limit of low statistics and the poor angular resolution of their detectors. In addition to UHECRs, a better understanding of the 3D structure of the GMF would significantly contribute to our understanding of a multitude of astrophysical problems, such as endeavors to subtract the Galactic foreground from the cosmic microwave background foreground (Tassis & Pavlidou 2015; Rubiño-Martín et al. 2023), as well as the physics of star formation (Pattle et al. 2023; Doi et al. 2024).

The primary challenge in understanding the GMF lies in the difficulty of obtaining a 3D tomographic reconstruction of the intervening GMF because the majority of the currently accessible observations are integrated along the LOS, such as observations of the Faraday rotation induced by the GMF (Pandhi et al. 2022; Hutschenreuter et al. 2024). This limitation has guided the predominant approach in GMF modeling to rely on parametric models. This is typically achieved by fitting parameters to distinct analytic components, for instance, a toroidal component, a poloidal component, and a turbulent component. To model the latter, a Gaussian random field is employed (Sun et al. 2008; Sun & Reich 2010; Takami & Sato 2010; Jansson & Farrar 2012a; Jansson & Farrar 2012b).

However, direct insights into the 3D structure of the interstellar medium of the Milky Way are attainable. The Gaia mission has mapped the positions of over a billion stars in the Galaxy by accurately measuring stellar parallaxes (Gaia Collaboration 2016, 2021; Bailer-Jones et al. 2021). Combined with other spectroscopic data, these parallaxes have enabled the construction of 3D tomographic maps showing the dust density distribution in certain regions of the Galaxy (Lallement et al. 2018, 2019, 2022; Green et al. 2019; Leike & Enßlin 2019; Leike et al. 2020, 2022; Edenhofer et al. 2024b). Nevertheless, these are maps of the dust density and do not directly constrain the magnetic field.

Observational methods that probe the 3D structure of the GMF do exist, however. A notable example is the linear polarization of starlight. Typically, starlight originates from its source as unpolarized light, but it can become linearly polarized due to the dichroic absorption by interstellar dust particles, which align themselves with the surrounding magnetic field (Andersson et al. 2015).

Future optopolarimetric surveys such as the Polar-Areas Stellar Imaging in Polarization High-Accuracy Experiment (PASIPHAE) and SOUTH POL are poised to deliver high-quality stellar polarization measurements for millions of stars (Magalhães 2012; Tassis et al. 2018; Maharana et al. 2021; Maharana et al. 2022). When combined with the stellar distance data obtained from the Gaia survey, these measurements will enable direct tomographic measurements of the GMF POS component in regions with dust clouds (Davis 1951; Chandrasekhar & Fermi 1953; Panopoulou et al. 2017; Skalidis et al. 2021; Skalidis & Tassis 2021; Pelgrims et al. 2023). Notably, optical stellar polarimetry was recently used in order to probe the GMF in 3D in the Sagittarius spiral arm (Doi et al. 2021; Doi et al. 2024). Additionally, local information can be further constrained through the study of HI gas in different velocity bins, which provide GMF information on the number of clouds along a given LOS as well as local information through the use of 3D dust-reddening maps (Green et al. 2018; Tritsis et al. 2018, 2019; Clark & Hensley 2019). In conjunction with available LOS data (see, e.g., Tahani et al. 2022a,b), this information promises to provide localized and sparse GMF data in the future. All of

these techniques will provide crucial information for creating 3D tomographic maps of specific areas of interest. With these maps, the paths of UHECRs can be traced back through these regions and the source localization on the sky can be improved (however, the contribution of the intergalactic magnetic field is still not accounted for). Specifically, there is an intense interest in mapping the GMF in the direction of UHECR hotspots, as well as in parts of the Galaxy that are likely to have been traversed by particles comprising these hotspots (Abbasi et al. 2014; Pierre Auger Collaboration 2017; Kawata et al. 2019).

This study is the second entry in our effort to reconstruct the GMF nonparametrically in 3D in a Bayesian setting. It directly follows Tsouros et al. 2024, hereafter Paper I. Essentially, we address an inverse problem within a Bayesian framework, where the goal is to sample the posterior distribution of GMF configurations in a specific part of the Galaxy using a combination of local and LOS-integrated information. In this work, local measurements only provide information for the POS component of the magnetic field. This corresponds to the information content of tomographic measurements of interstellar magnetized dust through optopolarimetry of starlight. On the other hand, LOS-integrated measurements provide information about the LOS component of the magnetic field as derived from Faraday rotation measurements, for instance. We approach this problem within the context of information field theory, which was developed specifically for Bayesian inference for fields and was successfully applied in various contexts (Enßlin et al. 2009; Enßlin 2019; Enßlin 2022). By reconstructing the posterior distribution of GMF realizations, we aim to accurately recover the true arrival directions of UHECRs given the observed arrival directions. We also account for the influence of the GMF.

In Section 2 we briefly describe the method, the forward models we used, and the sampling of the posterior. In Section 3 we present the main results of the algorithm for the considered scenarios, and in Section 4 we discuss the results further.

2. Method

In general, we are interested in inferring the configuration of the GMF, $\mathbf{B}(\mathbf{x})$ with $\mathbf{x} \in \mathcal{V}$ over a domain $\mathcal{V} \subset \mathbb{R}^3$, given some observed dataset d . In the context of Bayesian inference for continuous signals, the task is to determine the posterior probability distribution of $\mathbf{B}(\mathbf{x})$ for d ,

$$P(\mathbf{B}|d) = \frac{1}{Z} P(d|\mathbf{B})P(\mathbf{B}). \quad (1)$$

Here, $P(d|\mathbf{B})$ is the likelihood, representing the probability of observing magnetic field measurements d given a specific configuration $\mathbf{B}(\mathbf{x})$. The prior, $P(\mathbf{B})$, encapsulates preexisting information about $\mathbf{B}(\mathbf{x})$, and $Z = P(d)$ is the normalization factor.

In this work, the field that serves as a ground truth (the true field) was generated from a dynamo magnetohydrodynamic simulation discussed in Appendix A. The original simulation domain extended to ~ 1 kpc in the $x-y$ direction and to ~ 2 kpc above the Galactic plane. The GMF was rescaled so that its root mean square (RMS) value was $5\mu\text{G}$.

2.1. Likelihood

Tomography of the magnetized ISM from stellar polarization measurements is a highly nontrivial problem, and its full

discussion is beyond the scope of this work (Pelgrims et al. 2023). However, we note that by combining Gaia data and stellar polarization data for stars with a known distance to the Sun, it is possible to acquire information on the Stokes parameters that each intervening dust cloud imposes on the observed starlight if enough stars have polarization measurements and known distances. This can then be translated into local information on the orientation of the POS component of the GMF at that cloud through the connection to grain alignment, as referenced briefly in the previous section and thoroughly used in Tassis et al. (2018). Information on the POS component of GMF in clouds can also be acquired by the use of 21 cm neutral hydrogen (HI) emission measurements (Clark & Hensley 2019). We assumed that the task of determining the locations to which the measurements correspond to has been carried out.

Thus, for the i th datapoint, we assumed a forward model of the form

$$\mathbf{d}_{\text{local}}^{(i)} = \int R_{\text{local}}(\mathbf{x}, \mathbf{x}_i) \mathbf{B}(\mathbf{x}) d^3x + \mathbf{n}_{\text{local}}^{(i)}, \quad (2)$$

$$R_{\text{local}}(\mathbf{x}, \mathbf{x}_i) \equiv \delta^{(3)}(\mathbf{x} - \mathbf{x}_i) P_{\text{POS}}, \quad (3)$$

where $\mathbf{B}(\mathbf{x})$ is the magnetic field, and $\mathbf{n}_{\text{local}}^{(i)}$ are the observational uncertainties that contaminate our measurements. For clarity in the notation, we henceforth suppress the volume integration and write $\mathbf{d}_{\text{local}} = R_{\text{local}} \mathbf{B}$, where the volume integral is implicit. The vector \mathbf{x}_i is the location of the i th cloud where the magnetic field is measured, and P_{POS} signifies a projection operator on the POS, which reflects that (mainly) the POS component of the magnetic field is measured via dust polarization, $P_{\text{POS},ij} = \delta_{ij} - \hat{x}_i \hat{x}_j^T$ with $\hat{x}_i = x_i / \|x_i\|$ (assuming the observer to be at the origin). The Dirac delta function localizes the measurements at specific known locations \mathbf{x}_i .

The option to include the operator P_{POS} into the considered scenario is central to this work because it presents one of the main additions compared to Paper I. A complete projection on the POS is a pessimistic scenario because LOS information can become available by incorporating Zeeman or Faraday rotation data (Tahani et al. 2022a; Tahani et al. 2022b). A complete projection on the POS should therefore be seen as an extreme benchmarking scenario.

We note that this forward model is quite simplistic in that it assumes that accurate 3D locations are measured. Formally, this is captured by the Dirac delta function and that the locations \mathbf{x}_i are to be assumed known up to the resolution length scale. As we show in Section 2.4, the resolution of the grid in which our reconstruction takes place is some tens of parsecs, which is comparable to the uncertainty of the cloud localization (Pelgrims et al. 2023). In this particular application, distance uncertainties are therefore not relevant because they tend to be on scales that are lower than the grid resolution. However, it should be noted that in applications of this method with grids whose resolution is below the scale of the cloud localization uncertainties, the latter should be included because they would indeed have an observable effect.

The vector $\mathbf{n}_{\text{local}}^{(i)}$ is assumed to be a random variable drawn from a Gaussian distribution with a known covariance N_{local} . When specific measurement techniques are identified, other more appropriate error distributions will be chosen. When we marginalize over the noise, the likelihood becomes

$$P(\mathbf{d}|\mathbf{B}) = \mathcal{G}(\mathbf{d}_{\text{local}} - R_{\text{local}} \mathbf{B}, N_{\text{local}}). \quad (4)$$

The covariance N_{local} was chosen to be a multiple of the identity, $(N_{\text{local}})_{ij} = \sigma^2 \delta_{ij}$, where we chose

$$\sigma = \frac{|\mathbf{B}|_{\text{RMS}}}{2}, \quad (5)$$

where $|\mathbf{B}|_{\text{RMS}} = 5 \mu\text{G}$ is the RMS value of the magnitude of the ground truth. This does not imply that the noise is correlated with the GMF covariance. It was merely chosen as such in order to ensure an S/N of about 2. This S/N was chosen as a worst-case scenario for the expected GMF measurement error that arises from the measurement of the magnitude of the POS component of the GMF, which is expected to have an S/N lower than 2 (Skalidis et al. 2021). As the directional error is subdominant to that of the POS magnitude estimate (Pelgrims et al. 2024), we assumed that it was included in our choice of the worst-case scenario S/N.

In addition to local data, we explored the possible use of integrated LOS data as inferred from Faraday measurements, for instance (Hutschenreuter et al. 2024). In this case, the forward model takes the form

$$d_{\text{int}}^{(i)} = (\overline{P_{\text{LOS}} \mathbf{B}})_{L_i} + n_{\text{int}}^{(i)}, \quad (6)$$

$$(\overline{P_{\text{LOS}} \mathbf{B}})_{L_i} \equiv \frac{1}{|L_i|} \int_0^{|L_i|} B_{||}(\mathbf{x}) d\ell, \quad (7)$$

where P_{LOS} projects a vector onto the LOS component ($B_{||}$), and L_i is the specific LOS under consideration. Furthermore, $|L_i|$ denotes the limit up to which we integrate. In this application, $|L_i|$ coincides with the distance between the Earth and the intersection of L_i with the boundary of \mathcal{V} . Essentially, the above is equivalent to assuming that the electron density is roughly constant and known up to $|L_i|$ and then falls to zero. While this is not a valid assumption for low Galactic latitudes, we maintained it in this proof-of-concept work. Finally, the vector $n_{\text{int}}^{(i)}$ corresponds to a random vector on the POS, with a covariance N_{int} .

The likelihood for the combined data is given by

$$P(\mathbf{d}|\mathbf{B}) = \mathcal{G}(\mathbf{d}_{\text{local}} - R_{\text{local}} \mathbf{B}, N_{\text{local}}) \mathcal{G}(d_{\text{int}} - (\overline{P_{\text{LOS}} \mathbf{B}})_{L_i}, N_{\text{int}}). \quad (8)$$

Similarly, we defined the covariance for the noise of the integrated measurements as $(N_{\text{int}})_{ij} = \sigma_{\text{int}}^2 \delta_{ij}$, where¹

$$\sigma_{\text{int}} = \frac{1}{2} \mu\text{G}. \quad (9)$$

Finally, the operator R_{local} , which sparsely samples the GMF, is defined as follows. After discretizing our domain to voxels (see Section 4.1), we applied a Bernoulli trial to each voxel to determine whether it was observed with a probability p and $1 - p$. The probability p was chosen so that the approximate number density of measurements n_{data} was about

$$n_{\text{data}} \sim \begin{cases} 100 (\text{kpc})^{-3}, & \text{if } T \geq 10^4 \text{ K} \\ 1000 (\text{kpc})^{-3}, & \text{if } T < 10^4 \text{ K,} \end{cases} \quad (10)$$

where T is the corresponding gas temperature of that voxel, acquired from the same simulation that produced our ground truth. This choice of n_{data} reflects the decay in the number of dust clouds as a function of distance from the Galactic plane, which

¹ While Faraday data are significantly more accurate than this assumption suggests, we used this pessimistic noise covariance to compensate for the unknown 3D electron density distribution.

directly correlates with the expected number of measurements with respect to the position above the Galactic plane because the local measurements of the GMF will ultimately exist where dust clouds are located after polarized-starlight tomography has been carried out. Moreover, as most of the gas has a temperature above 10^4 K, the number density of local measurements is 262 (kpc)^{-3} . This is about a few hundred parsec per cloud in the cold medium, which is comparable to the expectation for the true ISM (Panopoulou & Lenz 2020).

2.2. Prior

As in Paper I, the only hard constraint that needed to be imposed is that regardless of the candidate magnetic field configuration \mathbf{B} we considered, it must satisfy $\nabla \cdot \mathbf{B} = 0$ in order to be a viable candidate. To ensure that the magnetic field was divergence free, we assumed that it was related to a non-divergence-free random field $\boldsymbol{\varphi}$ by a divergence-cleaning operator \mathcal{P} . This transverse projection operator, defined in Fourier space as

$$\mathcal{P}_{ij}(\mathbf{k}) = \delta_{ij} - \hat{k}_i \hat{k}_j^T, \quad (11)$$

projects out the degrees of freedom of the Gaussian random vector field that violate the divergence-free condition. In other words, it connects a latent field $\boldsymbol{\varphi}(\mathbf{x})$ to the true magnetic field by the harmonic space relation

$$\hat{B}_i(\mathbf{k}) = \frac{3}{2} \mathcal{P}_{ij}(\mathbf{k}) \hat{\varphi}_j(\mathbf{k}), \quad (12)$$

where \mathbf{k} are wavevectors, and the Fourier modes are denoted by hatted fields. Eq. (12) ensures that $\nabla \cdot \mathbf{B} = 0$, while the factor $3/2$ accounts for power loss due to reduced degrees of freedom, aligned with the original assumption of statistical isotropy for $\boldsymbol{\varphi}$ (Jaffe et al. 2012), and the hatted fields denote the respective Fourier transforms. Our aim is to reconstruct the local GMF \mathbf{B} by inferring the latent field $\boldsymbol{\varphi}$, which is related to the latter by Eq. (12). For $\boldsymbol{\varphi}$, we assumed a Gaussian prior of the form

$$\mathcal{P}(\boldsymbol{\varphi}) = \frac{1}{|2\pi\Phi|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} \int d^3x d^3x' \sum_{ij} \varphi_i(\mathbf{x}) \Phi_{ij}^{-1}(\mathbf{x}, \mathbf{x}') \varphi_j(\mathbf{x}') \right]. \quad (13)$$

The quantity Φ_{ij} is the covariance matrix, defined as

$$\Phi_{ij}(\mathbf{x}, \mathbf{x}') = \langle \varphi_i(\mathbf{x}) \varphi_j^*(\mathbf{x}') \rangle, \quad (14)$$

where the symbol $\langle \dots \rangle$ signifies an average over the distribution $P(\boldsymbol{\varphi})$. That is, if $O(\mathbf{x})$ is some quantity of interest, then

$$\langle O(\mathbf{x}) \rangle \equiv \int d\boldsymbol{\varphi} P(\boldsymbol{\varphi}) O(\mathbf{x}).$$

We note that the average was taken over field configurations.

In our analysis, we chose not to integrate any prior knowledge about the GMF geometry and statistics, and we therefore used a prior distribution exhibiting statistical isotropy, homogeneity, and mirror symmetry². This is formally encapsulated by writing the Fourier space covariance in the form

$$\langle \hat{\varphi}_i(\mathbf{k}) \hat{\varphi}_j^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_{ij} \delta^{(3)}(\mathbf{k} - \mathbf{k}') P(k). \quad (15)$$

² A GMF with a broken mirror symmetry would be implemented with the inclusion of a helicity term in the power spectrum. This would be a complex antisymmetric term of the form $i\epsilon_{ijk} k_j H(k)$, where $H(k)$ denotes the real helicity spectrum, satisfying $-P(k) < H(k) < P(k)$ (Jaffe et al. 2012). A broken mirror symmetry can be visualized by twisted flux tubes.

A crucial point is that the 3D prior power spectrum $P(k)$ is not known and is to be inferred as well. It is modeled as a sum of a power law and an integrated Wiener component (Arras et al. 2022). The defining hyperparameters and their prior PDFs (typically called hyperpriors) are summarized in Table 1, and they are also briefly discussed in Paper I.

2.3. Sampling the posterior

Equipped with the likelihood and prior, the posterior in terms of the magnetic field \mathbf{B} is given by Eq. (1). Because the power spectrum $P(k)$ needs to be inferred along with the configuration of the GMF, this inference problem is nonlinear and cannot be solved by a generalized Wiener filter (Pratt 1972). For this reason, a nonperturbative scheme, called geometrical variational inference (geoVI) developed by Frank et al. (2021), was used. A brief exposition on geoVI can be found in Appendix A of Paper I. For the purposes of this work, it suffices to state that we did not sample magnetic field configurations from the true posterior directly, but rather from an approximate posterior, as is usually the strategy in variational methods. For this task, we employed the numerical information field theory (NIFTy³) package in Python (Selig et al. 2013; Steininger et al. 2017; Arras et al. 2019, Edenhofer et al. 2024a). The required input is the likelihood and the prior of the original physical model, as described in Sections 2.1 and 2.2, respectively.

2.4. Procedure

We summarize the specific setting we probed below and describe how we generated the synthetic data on which the method was verified.

- Spatial domain: The modeled space was assumed to be periodic because we implemented the details of the ground truth, where periodic boundary conditions were used in order to emulate the axisymmetry of the Galaxy (see Appendix A, and Bende & Subramanian 2022 for a broader discussion of the simulation used). Furthermore, we padded our space by a factor of two, so that the x and y directions reached an extent of ~ 1.6 kpc. The resulting cube was uniformly partitioned into N_x , N_y , and N_z segments per axis, where $N_x = N_y = 48$, and $N_z = 64$, with padding. In this setting, every voxel had a linear dimension of approximately 30 pc. This accommodated the expected size of the dust clouds, as well as the uncertainty of the measurement positions (at least as an order of magnitude), as shown in Pelgrims et al. 2023.
- Data masking: We applied R_{local} (see Section 2.1) to the ground truth field in order to acquire the noiseless data.
- Adding noise to local data: Gaussian noise with a covariance matrix N_{local} (Eq. (5)) was added to each observed data vector.
- Integrated data: Optionally (see Section 3.3), the likelihood was supplemented by an additional term for the integrated local measurements, as in Eq. (8). In practice, the magnetic field was transformed from a Cartesian coordinate system to a spherical polar coordinate system with the Earth at the origin. Then, the radial component of the GMF, which is equivalent to the LOS component, was integrated along individual LOSs, resulting in a set of 2D integrated measurements that informed the model further.

³ The documentation can be found in <https://ift.pages.mpcdf.de/nifty/index.html>

Table 1. Hyperparameters of the prior.

Parameter	Distribution	Mean	Standard deviation
Total offset (\mathbf{B}_0)	Not-applicable	0	Not-applicable
Total offset st. dev.	Log-normal	3 μG	1 μG
Total spectral energy	Log-normal	1 μG	1 μG
Spectral index	Normal	$-\frac{11}{3}$	1
Int. Wiener process amplitude	Log-normal	1.5	1

- Adding noise to integrated data: Gaussian noise with a covariance N_{int} (Eq. (9)) was added to each pixel on the celestial sphere to contaminate the data acquired from the previous step.
- Sampling the approximated posterior: Finally, the geoVI method was applied to the true posterior distribution, resulting in samples from the approximate distribution. To all the latent fields we sampled, we again applied the projection operator (Eq. (11)) to obtain posterior samples of the divergence-free GMF.
- Application to UHECR back tracing: Through each of the GMF samples drawn from (1) in the previous step, we traced back an UHECR of known observed arrival direction θ_{obs} and rigidity $r_* \equiv E/Z$. A recording of the final velocity of the particles, in particular their original directions θ when they leave \mathcal{V} , essentially provides samples from the distribution $P(\theta|D)$ of the original arrival directions of the particles before they entered the GMF for the data

$$D \equiv \{d, r_*, \theta_{\text{obs}}\}. \quad (16)$$

To keep the discussion simple, we only considered UHECRs with a fixed rigidity $r_* = 5 \times 10^{19}$ eV (equivalently, protons of energy equal to the GZK limit⁴, $E = 5 \times 10^{19}$ eV). As a way to benchmark the quality of our reconstructions in the context of UHECR physics, we compared the angular separation $\delta\theta$ between the true arrival direction θ_{true} and that of the back-propagated UHECR, ending up with a distribution over $\delta\theta$. In this context, the true arrival direction always refers to the direction of the UHECR immediately where it entered \mathcal{V} . In Fig. 1, we provide a visual representation of the quantities defined in this section.

3. Results

In this section, we use NIFTy in order to sample the posterior distribution for three different scenarios: In scenario A, the observed data only consist of local measurements, and at each location, we only probe the components of the GMF that are parallel to the POS. In scenario B, all three components of the GMF (including the LOS) are probed on an equal footing for comparison. Finally, in scenario C, we use the same dataset as in scenario A, but additionally use integrated LOS information over the whole sky.

For each of these scenarios, we benchmarked the success of the reconstruction by using the GMF posterior samples in order to infer the true arrival direction of a UHECR with fiducial rigidity of $r_* = 5 \times 10^{19}$ eV for all possible observed arrival directions on the northern sky, as described in the previous section.

⁴ This is the theoretical upper energy limit of protons that travel over long distances, set by the slowing effect of interactions with the CMB photons (Greisen 1966 ; Zatsepin & Kuzmin 1966).

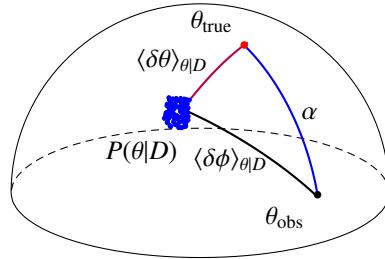


Fig. 1. Illustration and definition of quantities used in the analysis. A UHECR with a known rigidity r_* enters the Galaxy with an arrival direction θ_{true} (red dot). Because of the GMF, it is deflected and is observed on Earth as arriving from θ_{obs} (black dot). The angular distance between θ_{obs} and θ_{true} is α , and it is the error that the GMF induces on the observed arrival direction. We traced back the particle through each GMF configuration sampled using NIFTy, thus ending up with a distribution of arrival directions $P(\theta|D)$, with D defined in Eq. (16). From the posterior samples drawn, we calculated the mean angular distances $\langle\delta\theta\rangle_{\theta|D}$ and $\langle\delta\phi\rangle_{\theta|D}$ to the true and observed arrival directions, respectively, as well as the standard deviations for the former. The scales in this artificial example are larger for visual clarity, and do not correspond to an application of the method.

3.1. Scenario A: Local measurements with information on the plane of the sky alone

The local GMF information that can be acquired through starlight polarization-based tomography alone is confined to the celestial sphere (Panopoulou et al. 2019; Pelgrims et al. 2023). In this section, we therefore sample the posterior Eq. (1) for local GMF data d that are completely blind to the LOS dimension, as is the case for polarization measurements.

To do this, we worked on a spherical polar coordinate system with the Sun at the origin. The magnetic field is expressed as $\mathbf{B}(\mathbf{x}) = (B_r, B_\theta, B_\phi)$ in this coordinate system. In Fig. 2, we reconstruct the simulated GMF described in Appendix A. In Fig. 2a, we show the ground truth. Fig. 2b depicts the synthetic local GMF data obtained from the ground truth for this scenario. The result of the reconstruction algorithm is a set of 100 posterior samples of Eq. (1), given the data of Fig. 2b. In Fig. 2c, we show the mean of the posterior samples.

In Figs. 3a and 4a, we show the mean and standard deviation of the angular distance error ($\langle\delta\theta\rangle_{\theta|D}$ and $\sigma_{\theta|D}$, respectively) obtained through the use of the GMF reconstructions shown in Fig. 2. $\langle\delta\theta\rangle_{\theta|D}$ and $\sigma_{\theta|D}$ vary across the celestial sphere, and the specific structure of these functions depends on the specific chosen ground truth GMF. The greatest error of the reconstruction for this setting is approximately 14° . In order to judge the performance, we depict in Fig. 5a the angular error in the arrival direction assuming the observed ones were true, that is, we ignore the correction using the recovered GMF. By comparing Fig. 5a to Fig. 3a, we observe that reconstructing the local GMF

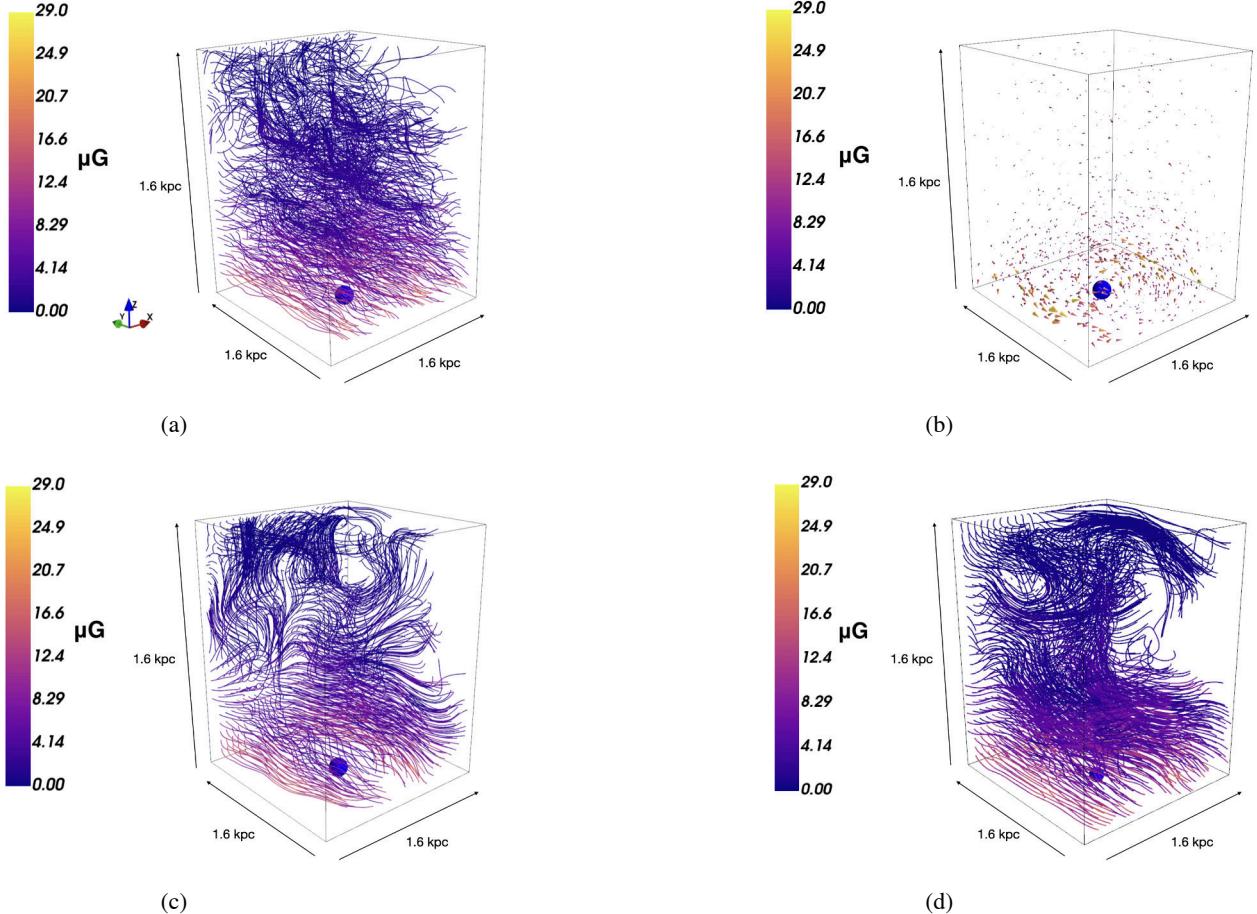


Fig. 2. Reconstruction of the simulated 3D magnetic field with the use of local data that lack LOS field component information. The blue sphere represents the celestial sphere. Panel a: ground truth; the GMF obtained as described in Appendix A. The field is rescaled so that it has an RMS norm of 5 μG . Panel b: synthetic data based on the ground truth of Fig. 2a. The radial component of the magnetic field is not measured. Panel c: the mean of the approximating posterior distribution attained via the geoVI algorithm based on the data provided in Fig. 2b. Panel d: mean of the approximating posterior distribution attained for the local data of Fig. 2b as well as integrated measurements of the radial component (Fig. 6b). The side of the box is about 1.6 kpc, or approximately 20 times the GMF correlation length.

for d yields a significant improvement in our ability to recover UHECR arrival directions. This result suggests that $\langle \delta\theta \rangle_{\theta D}$ is greater for UHECRs that are observed to arrive from directions in which the influence of the GMF is greater (Fig. 5a), in this case, at low longitudes. This correlation is further explored in Section 4.1.

3.2. Scenario B: Local measurements with full 3D information at each measured location

In this section, we examine the impact of a complete lack of observation of the LOS (scenario A) on the UHECR arrival direction reconstruction. For this purpose, we performed the same inference as in Section 3.1, but the LOS component was also probed locally, just like the POS components. In Figs. 3b and 4b, we plot the mean angular error $\langle \delta\theta \rangle_{\theta D}$ and the respective standard deviation for this scenario. Compared to the results of scenario A (see Figs. 3a and 4a), the quality of the reconstruction greatly improves when local LOS information is included. While the maximum mean angular error drops by a few degrees, the improvement is dramatic in general in that the total area of the sky where the maximum bias occurs is substantially reduced. This observation also holds for the variance.

While we considered θ_{obs} over the whole northern hemisphere for benchmarking purposes, in real applications, only sufficiently high Galactic latitudes are relevant. The reasons for this are twofold: First, reconstructions of the GMF will mainly be available at high enough latitudes because a tomographic reconstruction is exceptionally difficult on the Galactic disk due to the high density of HI clouds. Second, as Gaia data extend to approximately two kiloparsecs, this also defines the limits of the Galactic region in which our method can be expected to yield a sufficiently good reconstruction. However, UHECRs observed as arriving at low Galactic latitudes will have traveled through a significant part of the Galaxy that is not reconstructed, and this places a limit on the applicability of our method on low enough latitudes.

We have shown that knowledge of local LOS information would yield a substantial improvement over our ability to reconstruct the GMF, at least as far as UHECR backtracing is concerned. As stellar polarization data alone cannot probe the LOS dimension, this information would have to be supplemented by additional methods (e.g., Zeeman measurements). However, the local measurement of the LOS GMF component is a notoriously difficult task. We therefore attempt to mitigate this below by including integrated LOS information in our likelihood.

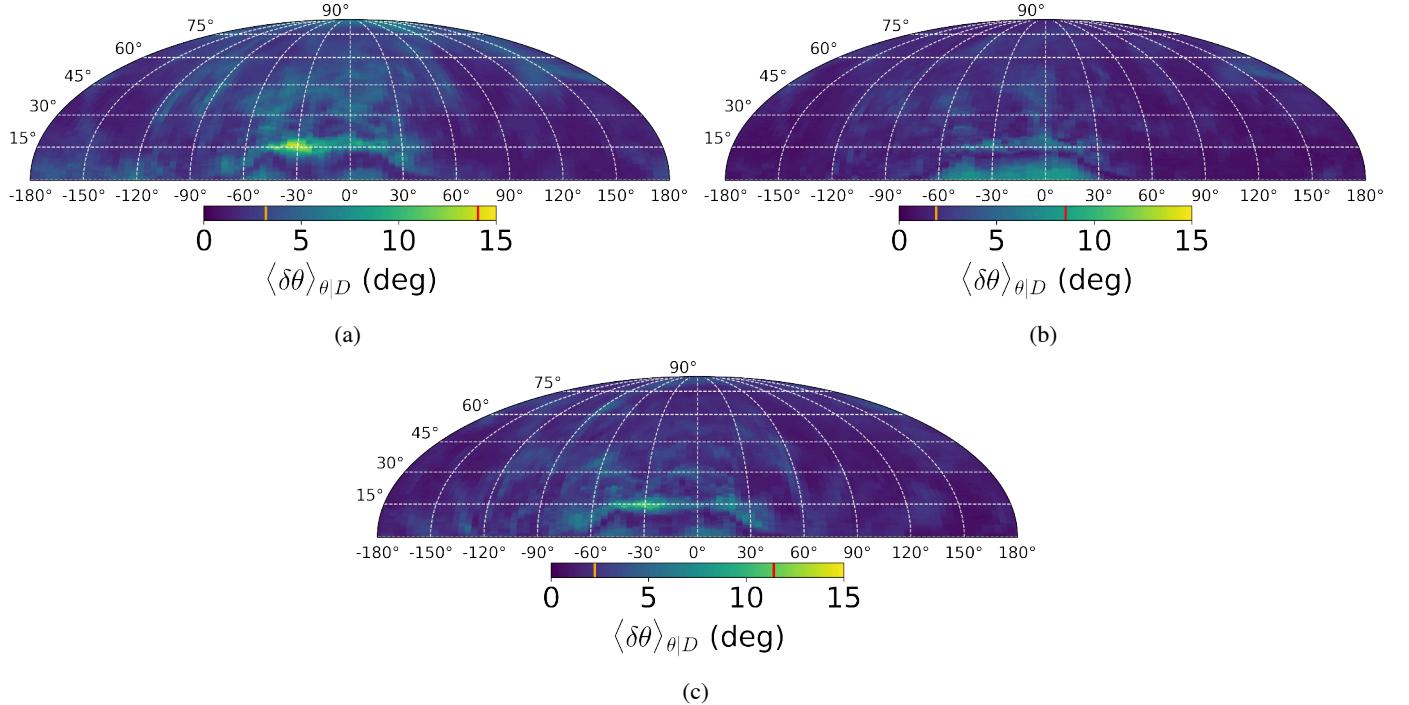


Fig. 3. Mean angular error of the reconstruction (see Fig. 1) as a function of all possible arrival directions on the northern hemisphere for the case of a UHECR with a rigidity $r_* = 5 \times 10^{19}$ eV. Panel a: magnetic field data consisting of local information with the LoS component are projected out (scenario A). Panel b: magnetic field data consisting of local information with the measured LOS component (scenario B) Panel c: as in the top left panel, but the data are supplemented by integrated LOS data (scenario C; see Fig. 6). The red and orange lines on the color bar indicate the maximum and mean values of the map, respectively.

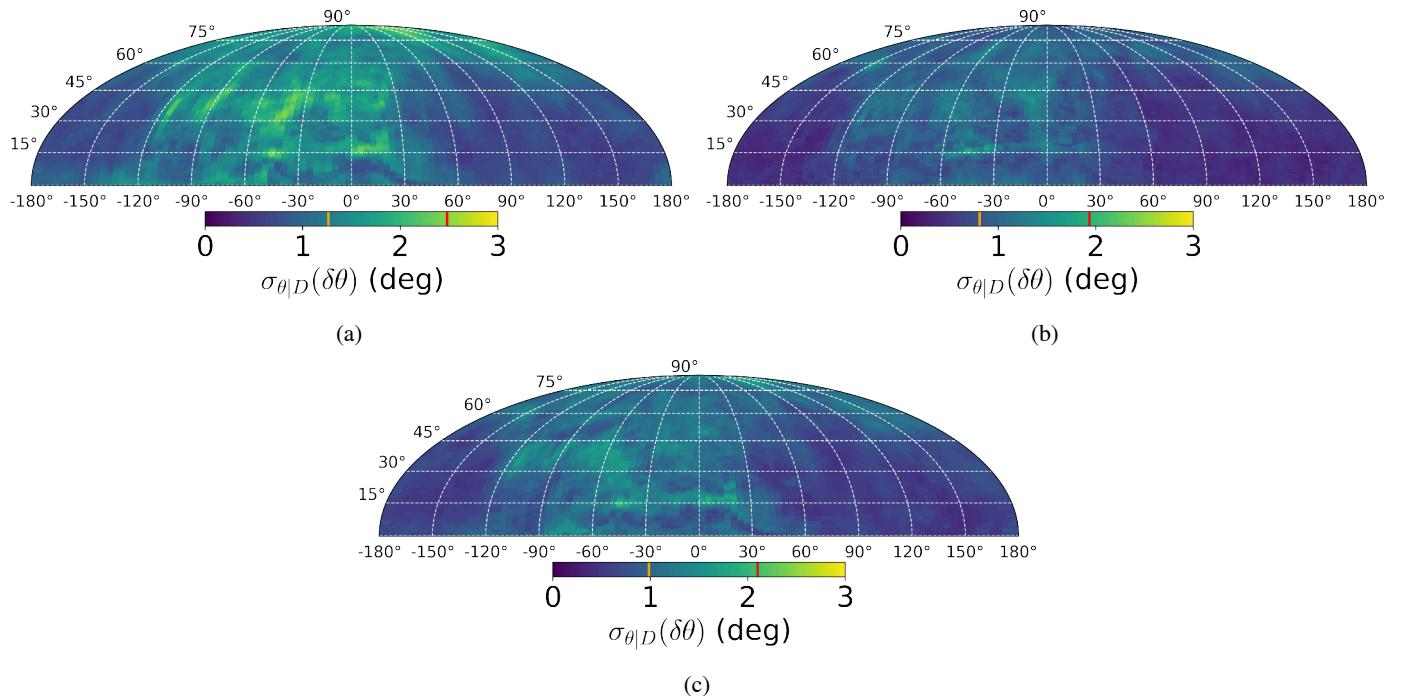


Fig. 4. As in Fig. 3, but for the corresponding angular error standard deviations as a function of the observed arrival direction.

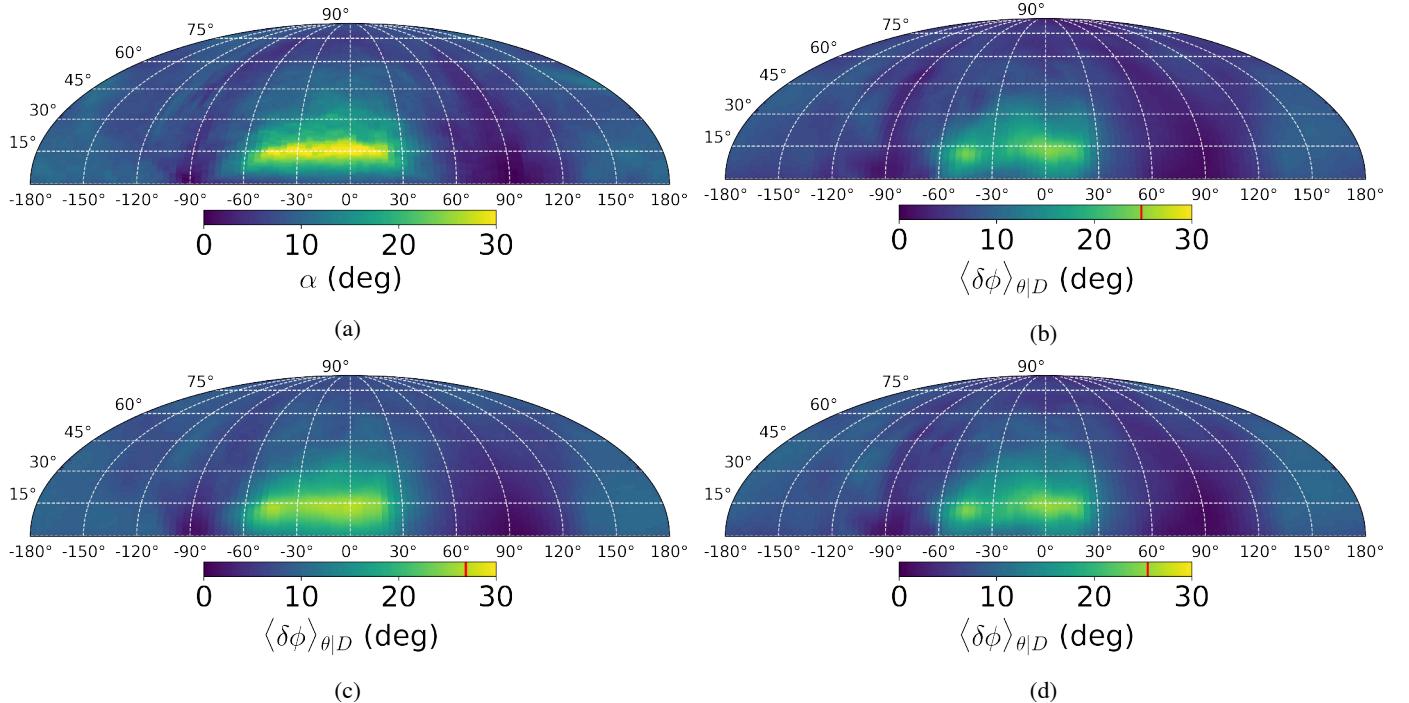


Fig. 5. Amount by which a UHECR of rigidity $r_* = 5 \times 10^{19}$ eV is deflected by different GMF configurations for all possible observed arrival directions on the northern sky, θ_{obs} (the deflection map; see Fig. 1 for the definition of the relevant quantities). Panel a: true deflection map. Panel b: mean deflection over the posterior samples for scenario A. Panel c: as in 5b, but the local measurements of the GMF now contain information on the LOS component as well as the POS component (scenario B). The additional information in this case causes a greater resemblance of the posterior mean to the true field, and so the deflection map is closer to Fig. 5a. Panel d: as in Fig. 5b, but the posterior is additionally constrained by the integrated data shown in Fig. 6b (scenario C). The scale of the color bar is kept up to 30 degrees to aid visual comparison. The red line on the color bar indicates the maximum deflection for each case. The dominant central feature of Fig. 5a is recovered in Figs. 5b–5d, since it is caused by the largest-scale features of the magnetic field, which we are able to infer in every case.

3.3. Scenario C: Local measurements with information on the plane of the sky supplemented by integrated line-of-sight measurements for the whole sky

In this section, we consider the inclusion of integrated constraints on the LOS component of the GMF as shown in Fig. 6b, while the local measurements at the dust clouds, simulating those obtained through polarized starlight tomography, are still projected on the celestial sphere as in Fig. 2b. Therefore, the likelihood we used now had the full form of Eq. (8).

In Figs. 3c and 4c, we show the mean and standard deviation of the angular distance error of the inferred UHECR arrival direction using the samples that were produced through the updated posterior for both local POS data and integrated LOS data. In comparison to scenario A, shown in Figs. 3a and 4a, the improvement in the ability to reconstruct the UHECR arrival direction is substantial in that the maximum mean angular error is reduced by a factor of about 1.5, the part of the POS where the maximum mean angular error occurs is greatly reduced, and the variance of the posterior is diminished by a factor of about 1.2. For the setting we considered, we thus showed that including integrated LOS data of the GMF (which is a much more realistic expectation than full 3D local measurements of scenario B) also leads to significantly better results with regard to recovering the arrival directions of UHECRs with a rigidity r_* .

3.4. Correlation length

In addition to the arrival direction of UHECRs, we can also infer the correlation length of the field for each of the above

scenarios. When we denote the fluctuating component of the GMF as $\delta\mathbf{B}(\mathbf{x})$, the two-point correlation function is

$$\langle \delta B_i(\mathbf{x}) \delta B_i^*(\mathbf{x}') \rangle = \xi(\mathbf{x}, \mathbf{x}'), \quad (17)$$

where a sum over i indices is understood. When we assume homogeneity and isotropy, the function ξ only is a function of the norm of the difference vector $\mathbf{r} \equiv \mathbf{x} - \mathbf{x}'$, that is, $\xi = \xi(r)$, where $r \equiv |\mathbf{r}|$. In this case, the correlation length is defined as

$$L_c \equiv \frac{1}{\xi(0)} \int_0^\infty dr \xi(r). \quad (18)$$

As discussed in Section 2.2, we directly sampled GMF configurations from the posterior, but also inferred a parameterized power spectrum $P(k)$, providing us with two different ways to calculate L_c : directly from the GMF samples, and from the inferred $P(k)$. We recall that the power spectrum was modeled as depending on the magnitude of the wavevectors alone, implying an assumption of homogeneity and isotropy. This is not because we think that the GMF obeys these conditions, but rather due to our a priori ignorance regarding their violation. In truth as well as in our simulated ground truth, the magnetic field is vertically stratified, and homogeneity therefore cannot be assumed to hold along the z direction, but only on constant z slices. The individual GMF samples are informed by the data about violations of homogeneity and isotropy, as shown in Figs. 2 and 7. Therefore, caution should be exercised when calculating the correlation length directly from the samples.

Following Bende & Subramanian (2022), we only computed the correlation length in slices of constant z , thus ending up with

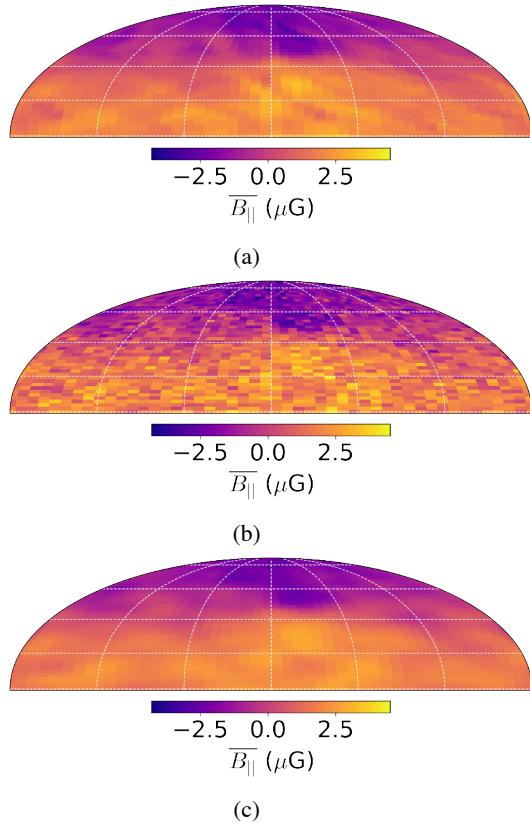


Fig. 6. The result of the inference of the GMF as seen on the northern POS. Panel a: averaged LOS component of the test magnetic field, shown in Fig. 2a. Panel b: noisy integrated data that were used along with the sparse and local data shown in Fig. 2b in order to define the LOS-informed posterior distribution. The noise covariance was set to $0.5 \mu\text{G}^2$, and the density of integrated measurements was 0.1 deg^{-2} . Panel c: averaged LOS component of the mean 3D configuration of the approximating posterior distribution for the data of Figs. 2b and 6b.

2D two-point correlation function of the field, $\xi(x, y) = \xi(\rho)$, for fixed z , where $\rho = \sqrt{(x^2 + y^2)}$. We therefore define

$$L_c(z) \equiv \frac{1}{\xi(0)} \int_0^\infty d\rho \xi(\rho), \quad (19)$$

and

$$\bar{L}_c \equiv \langle L_c(z) \rangle_z, \quad (20)$$

where $\langle \dots \rangle_z$ denotes averaging over the z -axis. For the ground truth, the z -averaged correlation length is calculated to be $L_c = 59 \text{ pc}$. In Fig. 8, we calculate L_c for all the posterior samples of the three scenarios above and compare it with the L_c calculated for the ground truth (dashed vertical line). The correlation length is systematically overestimated. This effect is expected because the reconstruction of noisily and sparsely sampled random fields necessarily loses information at small length scales, which introduces a low-pass filtering effect over the true signal and therefore leads to an increased correlation length. For this reason, the Bayesian inference of the correlation length should be seen as imposing upper bounds. In agreement with the results of the previous sections, knowledge of local LOS information on the location of individual measurements (scenario B) significantly improves our capability to infer the correlation length in this case as well.

The presence of differential shear means that even the 2D correlation function of the ground truth deviates from isotropy, and it therefore admits two distinct correlation lengths. For simplicity, however, we neglected the effect of shear and calculated the correlation length as defined above, with the understanding that the deviation from isotropy is a higher-order effect. The L_c we calculated for the ground truth lies in the range of correlation lengths computed in Breden & Subramanian (2022), who accounted for differential shear.

Finally, for comparison, we show in Fig. 9 the results for L_c calculated directly from Eq. (18), using the inferred power spectrum $P(k)$ of each sample. In this case, while we still obtain a correlation length of $\sim 100 \text{ pc}$, the variance is much greater than the results shown in Fig. 8 for all three scenarios. The tendency to overestimate L_c arises because the stratification and the subsequent large-scale correlation along the z -axis is not taken into account separately in this method, as it was in the previous method, and this manifests itself in the power spectrum by the increased measured correlation length.

4. Discussion

4.1. Identification of a systematic bias

In Fig. 2a, we showed that the ordered component of the field primarily lies (anti)parallel to the $\pm \hat{y}$ direction, which corresponds to a longitude $l = \pm 90^\circ$. In Fig. 5a, this is reflected by the fact that the observed arrival directions parallel to the ordered component, $(l, b) \approx (\pm 90^\circ, 0^\circ)$, are minimally deflected, while the maximum deflection occurs at the arrival directions perpendicular to the ordered component of the field. We call the map of Fig. 5a the deflection map of the GMF for a UHECR with a rigidity r_* . If the deflection map of the GMF for a given of rigidity were available, we would be able to identify the regions of the celestial sphere in which observed UHECRs with this rigidity are deflected most strongly.

A comparison of Fig. 5a with Fig. 3 yields a direct correlation between the regions of the deflection map and the mean angular error of our inferred arrival directions as a function of observed arrival direction for the same rigidity. In qualitative terms, this correlation suggests that for observed arrival directions perpendicular to the GMF zero mode, where the particles must have been deflected most strongly, our inference of their true arrival direction is more prone to a systematic bias. This bias is to be understood as the angular distance of the mean of our posterior distribution with respect to the true value.

Even though we might not be able to correct for this bias using our available data, knowledge of how severely the GMF alters the UHECR trajectories can help us to characterize the regions of the POS where our reconstructions are expected to be affected by it. While the corresponding deflection of the true GMF for a value of the UHECR rigidity will not be known a priori⁵, its structure is largely dictated by the dominating mean value of the field, which is generally well captured by our algorithm, as shown in Paper I. As shown in Figs. 5b–5d, we are able to accurately recover the large-scale features of the deflection map for all three considered scenarios, and in this way, we chart the parts of the POS in which the GMF will affect the UHECR trajectories most strongly, and by extension, the regions in which our arrival direction posterior might be shifted with respect to the true value.

⁵ Its derivation requires knowledge of the full 3D structure of the GMF, which is unknown.

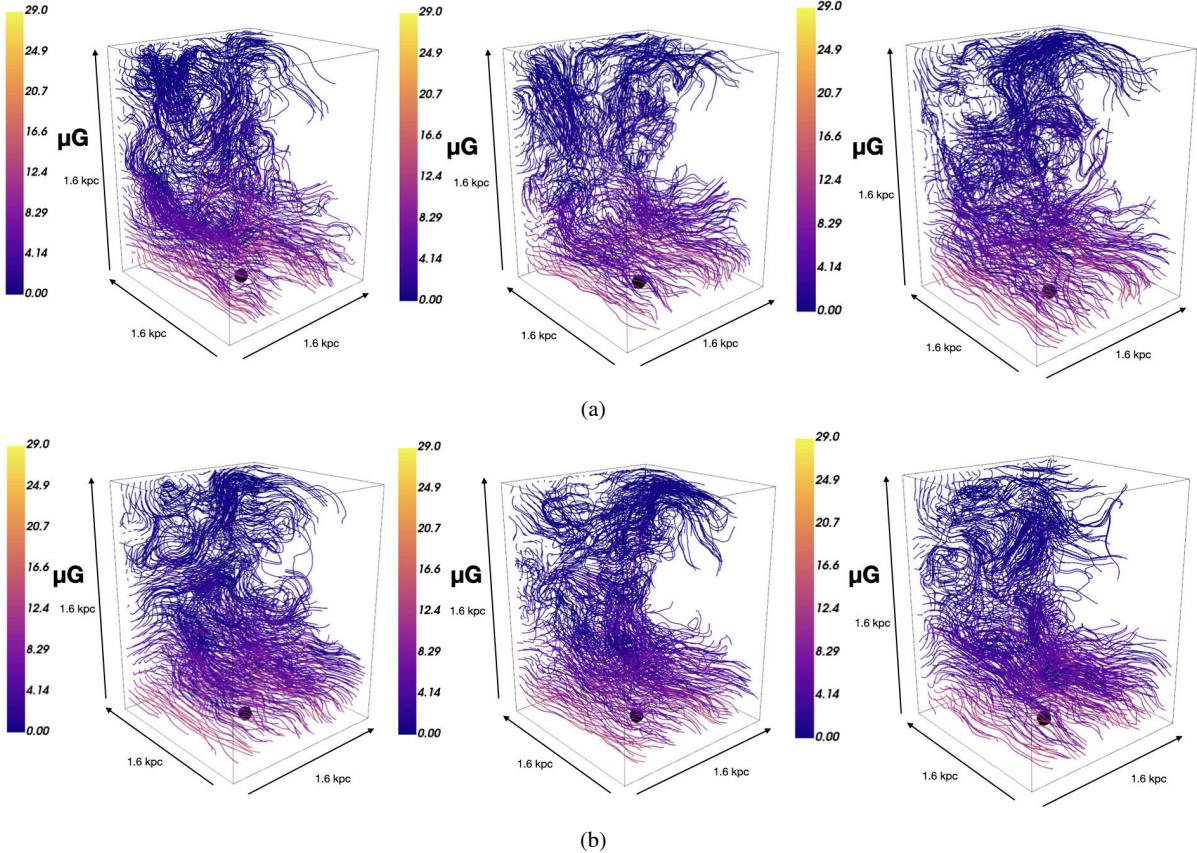


Fig. 7. Individual samples from two of the posterior distributions considered. Panel a: the posterior distribution conditional to the data of Figs. 2b and 6b. Panel b: the posterior distribution is conditional to the data of Fig. 2b alone.

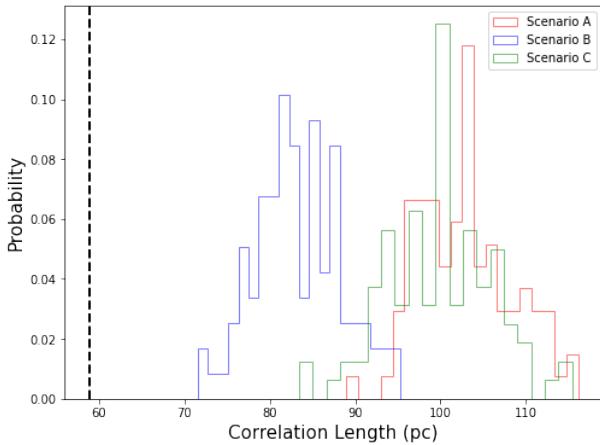


Fig. 8. Inference of the correlation length as defined via Eqs. (19) and (20) computed directly from the posterior samples for the three scenarios. The vertical line corresponds to the correlation length computed from the ground truth.

4.2. Caveats

While tomography using starlight polarization and Gaia data can provide the location of dust clouds in the local Galaxy as well as the POS orientation of the GMF at the location of each cloud, the POS direction of the GMF is generally not known, as this inference makes use of the properties of grain alignment, which cannot infer the POS directionality of the GMF (Tassis et al. 2018).

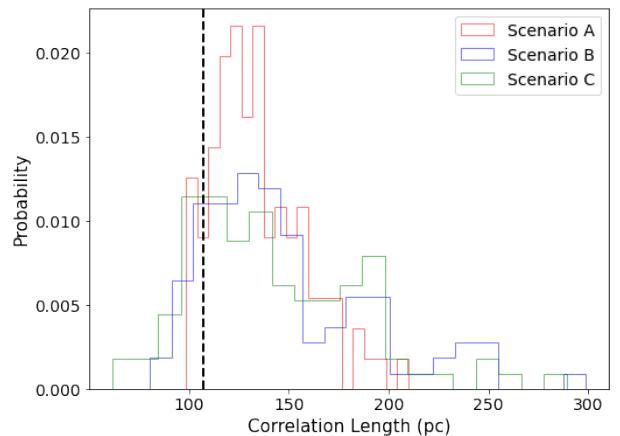


Fig. 9. Inference of the correlation length as defined via Eq. (18), using the inferred power spectrum of each respective field. The vertical line corresponds to the correlation length of the ground truth.

Furthermore, the integrated measurements used here assume that the integrated Galactic LOS component has been measured or inferred. In practice, the observables that need to be measured in order to estimate these integrals is the Faraday rotation measure and the dispersion measure. This means that even if the Galactic component is separated, it will still provide an average weighted over the thermal electron density. Therefore, we practically made the simplifying assumption that the thermal electron density is constant or known. In applications to the real GMF, the electron density will be treated as an additional degree of

freedom to be inferred (Hutschenreuter et al. 2024). However, it must be noted that recent research suggests the possibility that local LOS data can be available, at least in part of the dataset (Tahani et al. 2022a; Tahani et al. 2022b).

In this analysis, we only studied the case of UHECRs with a fixed rigidity of $r_* = 5 \times 10^{19}$ eV. This is equivalent to assuming that the UHECR particles are protons of $E = 5 \times 10^{19}$ eV. In general, the composition of UHECRs is unknown and is most likely mixed, especially if some of the sources have a Galactic origin (Calvez et al. 2010; Kusenko 2011; Jiang et al. 2021). The closer examination of different composition scenarios will be the subject of future work. The robustness of this method must be further confirmed in different simulations (see, e.g., Gent et al. 2024; Korpi-Lagg et al. 2024).

4.3. Conclusions and outlook

We extended the analysis of Paper I to the case of more realistic LOS information and local data distribution. This was motivated by the fact that in real applications, the local GMF data obtained through stellar polarization tomography will not contain LOS information, and the distribution of these measurements will follow the distribution of dust clouds, which is not homogeneous, as was assumed in Paper I.

Additionally, the ground-truth GMF that was used in order to benchmark the performance of our inference algorithm was taken from an MHD simulation with the aim of studying the effect of our Gaussian approach to magnetic field configurations whose statistical properties more closely resemble those of the real GMF. Furthermore, we supplemented the existing framework in order to also include LOS-integrated information.

Our results show that while the complete absence of LOS information in the local data diminishes the accuracy of our inferred UHECR arrival directions, we are able to significantly correct for the effect of the GMF on the observed arrival directions even in this case, at least for the rigidity considered here. The inclusion of integrated LOS data for the GMF, which can realistically be expected to be part of our available information, is enough to provide accurate enough results, however.

Even in directions in which the angular distance between the inferred arrival direction and the true direction are strongest are we still able to correct for the effect of the GMF by a factor of 3 in the setting we considered. Additionally, by our ability to reconstruct the large-scale features of the field that dominate UHECR deflection, we are able to identify the regions of the POS in which our reconstructions are most likely to be affected by the maximum error. Finally, we are able to estimate the correlation length of the GMF with two different methods, one directly from the posterior samples, and the other using the inferred power spectra.

Acknowledgements. A.T. and V.P. acknowledge support from the Foundation of Research and Technology – Hellas Synergy Grants Program through project MagMASim, jointly implemented by the Institute of Astrophysics and the Institute of Applied and Computational Mathematics. A.T. acknowledges support by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “Third Call for H.F.R.I. Scholarships for PhD Candidates” (Project 5332). V.P. acknowledges support by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant” (Project 1552 CIRCE). The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie RISE action, Grant Agreement no. 873089 (ASTROSTAT-II). This work also benefited greatly from discussions during the program “Towards a Comprehensive Model of the Galactic

Magnetic Field” at Nordita in April 2023, which is partly supported by NordForsk and the Royal Astronomical Society. A.T. would like to thank Vincent Pelgrims, Raphael Skalidis, Georgia V. Panopoulou, and Konstantinos Tassis for helpful tips and stimulating discussions. G.E. acknowledges the support of the German Academic Scholarship Foundation in the form of a PhD scholarship (“Promotionsstipendium der Studienstiftung des Deutschen Volkes”). P.F. acknowledges funding through the German Federal Ministry of Education and Research for the project ErUM-IFT: Informationsfeldtheorie fuer Experimente an Großforschungsanlagen (Foerderkennzeichen: 05D23EO1). The authors wish to thank an anonymous referee for insightful comments and suggestions.

References

- Abbasi, R. U., Abe, M., Abu-Zayyad, T., et al. 2014, *ApJ*, **790**, L21
- Andersson, B. G., Lazarin, A., & Vaillancourt, J. E. 2015, *ARA&A*, **53**, 501
- Arras, P., Baltac, M., Ensslin, T. A., et al. 2019, *Astrophysics Source Code Library* [[record ascl:1903.008](#)]
- Arras, P., Frank, P., Haim, P., et al. 2022, *Nat. Astron.*, **6**, 259
- Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, *AJ*, **161**, 147
- Bendre, A. B., & Subramanian, K. 2022, *MNRAS*, **511**, 4454
- Bendre, A., Gressel, O., & Elstner, D. 2015, *Astron. Nachr.*, **336**, 991
- Bhattacharjee, P., & Sigl, G. 2000, *Phys. Rep.*, **327**, 109
- Calvez, A., Kusenko, A., & Nagataki, S. 2010, *Phys. Rev. Lett.*, **105**, 091101
- Chandrasekhar, S., & Fermi, E. 1953, *ApJ*, **118**, 113
- Clark, S. E., & Hensley, B. S. 2019, *ApJ*, **887**, 136
- Davis, L. 1951, *Phys. Rev.*, **81**, 890
- di Matteo, A., Anchordoqui, L., Bister, T., et al. 2023, in *European Physical Journal Web of Conferences*, **283**, 03002
- Doi, Y., Hasegawa, T., Bastien, P., et al. 2021, *ApJ*, **914**, 122
- Doi, Y., Nakamura, K., Kawabata, K. S., et al. 2024, *ApJ*, **961**, 13
- Edenhofer, G., Frank, P., Roth, J., et al. 2024a, *J. Open Source Softw.*, **9**, 6593
- Edenhofer, G., Zucker, C., Frank, P., et al. 2024b, *A&A*, **685**, A82
- Enßlin, T. A. 2019, *Ann. Phys.*, **531**, 1970017
- Enßlin, T. 2022, *Entropy*, **24**, 374
- Enßlin, T. A., Frommert, M., & Kitaura, F. S. 2009, *Phys. Rev. D*, **80**, 105005
- Fang, K., & Murase, K. 2018, *Nat. Phys.*, **14**, 396
- Frank, P., Leike, R., & Enßlin, T. A. 2021, *Entropy*, **23**, 853
- Gaia Collaboration (Prusti, T., et al.) 2016, *A&A*, **595**, A1
- Gaia Collaboration (Brown, A. G. A., et al.) 2021, *A&A*, **649**, A1
- Gent, F. A., Mac Low, M.-M., & Korpi-Lagg, M. J. 2024, *ApJ*, **961**, 7
- Gilmore, G., Wyse, R. F. G., & Kuijken, K. 1989, *ARA&A*, **27**, 555
- Green, G. M., Schlafly, E. F., Finkbeiner, D., et al. 2018, *MNRAS*, **478**, 651
- Green, G. M., Schlafly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, *ApJ*, **887**, 93
- Greisen, K. 1966, *Phys. Rev. Lett.*, **16**, 748
- Hutschenreuter, S., Haeverkorn, M., Frank, P., Raycheva, N. C., & Enßlin, T. A., 2024, *A&A*, in press, <https://doi.org/10.1051/0004-6361/202346740>
- Jaffe, T., Waelkens, A., Reinecke, M., Kitaura, F. S., & Ensslin, T. A. 2012, Hammurabi: Simulating polarized Galactic synchrotron emission, *Astrophysics Source Code Library* [[record ascl:1201.014](#)]
- Jansson, R., & Farrar, G. R. 2012a, *ApJ*, **757**, 14
- Jansson, R., & Farrar, G. R. 2012b, *ApJ*, **761**, L11
- Jiang, Y., Zhang, B. T., & Murase, K. 2021, *Phys. Rev. D*, **104**, 043017
- Kawata, K., di Matteo, A., Fujii, T., et al. 2019, in *International Cosmic Ray Conference*, **36**, 310
- Korpi-Lagg, M. J., Mac Low, M.-M., & Gent, F. A. 2024, *Living Rev. Computat. Astrophys.*, **10**, 3
- Kotera, K., & Olinto, A. V. 2011, *ARA&A*, **49**, 119
- Kusenko, A. 2011, *Nucl. Phys. B Proc. Suppl.*, **212**, 194
- Lallement, R., Capitanio, L., Ruiz-Dern, L., et al. 2018, *A&A*, **616**, A132
- Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, *A&A*, **625**, A135
- Lallement, R., Vergely, J. L., Babusiaux, C., & Cox, N. L. J. 2022, *A&A*, **661**, A147
- Leike, R. H., & Enßlin, T. A. 2019, *A&A*, **631**, A32
- Leike, R. H., Glatzle, M., & Enßlin, T. A. 2020, *A&A*, **639**, A138
- Leike, R. H., Edenhofer, G., Knollmüller, J., et al. 2022, arXiv e-prints, [[arXiv:2204.11715](#)]
- Magalhães, A. M. 2012, in *Science from the Next Generation Imaging and Spectroscopic Surveys*, **7**
- Maharana, S., Kyriotakis, J. A., Ramaprakash, A. N., et al. 2021, *J. Astron. Telescopes Instrum. Syst.*, **7**, 014004
- Maharana, S., Anche, R. M., Ramaprakash, A. N., et al. 2022, *J. Astron. Telescopes Instrum. Syst.*, **8**, 038004
- Murase, K. 2019, in *International Cosmic Ray Conference*, **36**, 965

- Pandhi, A., Hutschenreuter, S., West, J. L., Gaensler, B. M., & Stock, A. 2022, [MNRAS](#), **516**, 4739
- Panopoulou, G. V., & Lenz, D. 2020, [ApJ](#), **902**, 120
- Panopoulou, G. V., Psaradaki, I., Skalidis, R., Tassis, K., & Andrews, J. J. 2017, [MNRAS](#), **466**, 2529
- Panopoulou, G. V., Tassis, K., Skalidis, R., et al. 2019, [ApJ](#), **872**, 56
- Pattle, K., Fissel, L., Tahani, M., Liu, T., & Ntormousi, E. 2023, in [Astronomical Society of the Pacific Conference Series](#), 534, Protostars and Planets VII, eds. S. Inutsuka, Y. Aikawa, T. Muto, K. Tomida, & M. Tamura, 193
- Pelgrims, V., Panopoulou, G. V., Tassis, K., et al. 2023, [A&A](#), **670**, A164
- Pelgrims, V., Mandarakas, N., Skalidis, R., et al. 2024, [A&A](#), **684**, A162
- Pierre Auger Collaboration (Aab, A., et al.) 2017, [Science](#), **357**, 1266
- Pratt, W. 1972, [IEEE Trans. Comput.](#), **C-21**, 636
- Rubiño-Martín, J. A., Guidi, F., Génova-Santos, R. T., et al. 2023, [MNRAS](#), **519**, 3383
- Sánchez-Salcedo, F. J., Vázquez-Semadeni, E., & Gazol, A. 2002, [ApJ](#), **577**, 768
- Selig, M., Bell, M. R., Junklewitz, H., et al. 2013, [A&A](#), **554**, A26
- Skalidis, R., & Tassis, K. 2021, [A&A](#), **647**, A186
- Skalidis, R., Sternberg, J., Beattie, J. R., Pavlidou, V., & Tassis, K. 2021, [A&A](#), **656**, A118
- Steininger, T., Dixit, J., Frank, P., et al. 2017, arXiv e-prints [arXiv:[1708.01073](#)]
- Sun, X.-H., & Reich, W. 2010, [Res. Astron. Astrophys.](#), **10**, 1287
- Sun, X. H., Reich, W., Waelkens, A., & Enßlin, T. A. 2008, [A&A](#), **477**, 573
- Tahani, M., Glover, J., Lupyncew, W., et al. 2022a, [A&A](#), **660**, L7
- Tahani, M., Lupyncew, W., Glover, J., et al. 2022b, [A&A](#), **660**, A97
- Takami, H., & Sato, K. 2010, [ApJ](#), **724**, 1456
- Tassis, K., & Pavlidou, V. 2015, [MNRAS](#), **451**, L90
- Tassis, K., Ramaprasakash, A. N., Readhead, A. C. S., et al. 2018, arXiv e-prints [arXiv:[1810.05652](#)]
- Torres, D. F., & Anchordoqui, L. A. 2004, [Rep. Progr. Phys.](#), **67**, 1663
- Tritsis, A., Federrath, C., Schneider, N., & Tassis, K. 2018, [MNRAS](#), **481**, 5275
- Tritsis, A., Federrath, C., & Pavlidou, V. 2019, [ApJ](#), **873**, 38
- Tsouros, A., Edenhofer, G., Enßlin, T., Mastorakis, M., & Pavlidou, V. 2024, [A&A](#), **681**, A111
- Zatsepin, G. T., & Kuzmin, V. A. 1966, [JETP Lett.](#), **4**, 78
- Ziegler, U. 2004, [Comput. Phys. Commun.](#), **157**, 207

Appendix A: Simulated Magnetic Field

We briefly summarize the setup and results of the Galactic dynamo simulations that have been analyzed here. A detailed description of the numerical setup is presented in [Bendre et al. \(2015\)](#).

These are Magnetohydrodynamic (MHD) simulations of the Galactic interstellar medium (ISM). The simulation domain is an elongated box, located roughly at the solar neighbourhood of the Milky Way. It has dimensions of approximately 1×1 kpc in the radial (x) and azimuth (y) direction and ranges from approximately -2 to $+2$ kpc in z direction, on either side of the Galactic mid-plane. It is split in a uniform Cartesian grid with a resolution of approximately 8.3 pc, and a set of non-ideal MHD equations is solved in this domain using the NIRVANA code ([Ziegler 2004](#)) (see Eq. 1 from [Bendre et al. \(2015\)](#) for the set of equations we have solved). Periodic boundary conditions were used in the y direction to incorporate the axisymmetry of the Galactic disc. The flat rotation curve is incorporated by allowing the angular velocity to scale inversely with the Galactic radius as $\Omega \propto 1/R$, with $\Omega_0 = 100 \text{ km s}^{-1} \text{ kpc}^{-1}$ at the centre of the box. Shearing periodic boundary conditions are used in the radial x direction to accommodate the aforementioned radial dependence of angular velocity. The initial density distribution of the ISM is in hydrostatic balance with the vertical gravity pointing towards the mid-plane, such that the vertical scale-height of the initial density was approximately 300 pc, with its value in the mid-plane of approximately $10^{-24} \text{ g cm}^{-3}$. A vertical profile of gravitational acceleration is adapted from [Gilmore et al. \(1989\)](#). The ISM in this box is stirred by supernovae (SN) explosions, which inject the thermal energy at random locations, at a rate of approximately $7.5 \text{ kpc}^{-2} \text{ Myr}^{-1}$. The vertical distribution of the explosions scale with the mass density. A piece-wise power law, similar to [Sánchez-Salcedo et al. \(2002\)](#), is used to model the temperature-dependent rate of radiative heat transfer, which along with SN explosions, roughly capture the observed multi-phase morphology of the ISM. We started the simulations with negligible initial magnetic fields of strength of the order of nG, and it grew exponentially to the strengths of the order of μG , with an e-folding time of about 200 Myr, such that the final energy density of the magnetic fields reached equipartition with the kinetic energy density of the ISM turbulence (shown in the right-hand panel of Fig. A.1). The exponential amplification of the magnetic energy saturated after about a Gyr, and coherent magnetic fields of scale-height close to 500pc were sustained in the box, consistent with the typical scale-height of GMFs (shown in the left-hand panel of Fig. A.1). The initial amplification and subsequent saturation phases of the magnetic field are termed here respectively, as kinematic and dynamical phases. This refers to the fact that magnetic fields from initial kinematic strengths amplify and reach the strengths that are dynamically significant to the turbulent flow. The growth and saturation of these large-scale fields are understood in terms of a self-consistent large-scale dynamo mechanism, governed by the SN-driven stratified helical turbulence and the Galactic differential rotation ([Bendre et al. 2015](#)). Finally, regarding the correlation length, the magnetic field is anisotropic, in that it is more correlated in the direction of shear. In the direction parallel to the shear the correlation length is calculated at approximately 100 – 150 pc, while in the direction perpendicular to it, it is in the range of 20 – 70 pc (see Fig. A3 of [Bendre & Subramanian 2022](#)). Neglecting shear for simplicity, we calculate the correlation length to be 63 pc.

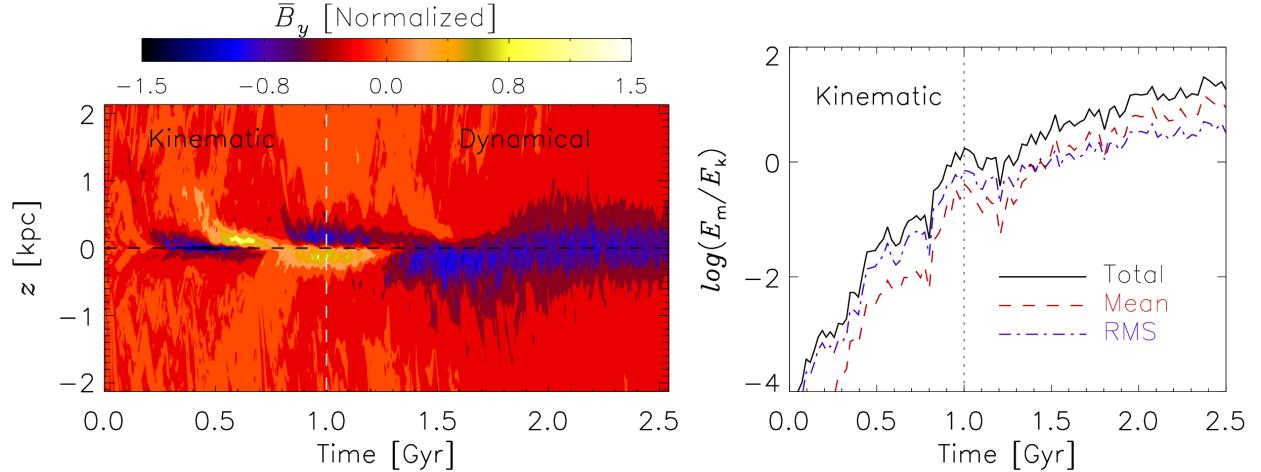


Fig. A.1: Left: Time evolution of the vertical (z) profile of the azimuthal component of the magnetic field averaged over $x - y$ plane. The color code is normalized by an exponential factor to compensate for an exponential growth of magnetic fields. The mean magnetic field eventually grows to a large-scale mode symmetric with respect to the Galactic mid-plane. Right: Time evolution of various contributions to magnetic energy, normalized to the turbulent kinetic energy (which stays roughly constant in time). The black solid line corresponds to the total magnetic energy contribution, the red dashed line corresponds to the magnetic energy of mean magnetic fields (averaged over the horizontal $x - y$ planes) and with the blue dot-dashed line to the magnetic energy in the RMS magnetic fields. The magnetic energy is amplified exponentially for about a Gyr and eventually reaches an equipartition with turbulent kinetic energy. Figure adapted from [Bendre & Subramanian 2022](#).

Chapter 6

The Effect of Heavier Primary Composition

Building on the framework developed in the previous chapters where the GMF was inferred using synthetic data with known directional information on the POS, assuming a known UHECR primary charge. Here, we expand on this approach by introducing probabilistic charge distributions over the UHECR charge and by relaxing the assumption of known GMF directionality of the local data on the POS. Specifically, we investigate the impact of varying the assumed UHECR charges for the same energys—from pure proton composition to heavier nuclei like iron—and introduce an ad hoc adjustment in order to assume a POS directionality to evaluate its efficacy under realistic observational conditions. The ad hoc assumption consists in aligning the local POS magnetic field component with the ordered direction of a parametric GMF model, like JF12 (Jansson and Farrar, 2012), and we explore the extent to which this assumption influences the accuracy of UHECR backtracking.

Our results demonstrate that, for GMF magnitudes around $5\mu\text{G}$, the deflection of protons can be corrected to within approximately 10° . For intermediate nuclei, like nitrogen, the reconstructions provide substantial improvements over existing models, particularly in directions orthogonal to the GMF. However, deflections of the most highly charged UHECRs, such as iron, remain largely irrecoverable due to their significant diffusion, except under lower-magnitude GMFs of around $1\mu\text{G}$, where significant corrections are achievable over broad sky regions, which corresponds to locations of the sky where the GMF is perpendicular to the UHECR arrival direction. By identifying the most uncertain regions in the POS data, priority areas for future observations can be identified.

Can the Galactic Magnetic Field's Influence be Removed from UHECR Arrival Direction Observations?

Alexandros Tsouros^{1,2*}, Abhijit B. Bendre^{3,4}, Gordian Edenhofer^{5,6,7}, Torsten Enßlin^{5,6}, Philipp Frank⁵, Michalis Mastorakis^{1,2}, Vasiliki Pavlidou^{1,2}

¹ Department of Physics & ITC, University of Crete, GR-70013, Heraklion, Greece

² Institute of Astrophysics, Foundation for Research and Technology-Hellas, Vasilika Vouton, GR-70013 Heraklion, Greece

³ Laboratoire d'Astrophysique, EPFL, CH-1290 Sauverny, Switzerland

⁴ Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, 56126 Pisa, Italy

⁵ Max Planck Institute for Astrophysics, Karl-Schwarzschild-Straße 1, 85748 Garching, Germany

⁶ Ludwig Maximilian University of Munich, Geschwister-Scholl-Platz 1, 80539 Munich, Germany

⁷ University of Vienna, Department of Astrophysics, Türkenschanzstrasse 17, 1180 Vienna, Austria

Received ; accepted

ABSTRACT

Context. Ultra-high energy cosmic rays (UHECRs) are the most energetic elementary particles in the Universe, with energies exceeding 10^{18} eV. Their sources are yet unidentified, due to magnetic deflections inside the Galactic magnetic field (GMF). This challenge is exacerbated by our limited knowledge of the 3D structure of the GMF. This is currently mainly confined to analytical models with parameters fitted from line-of-sight observations. Recent studies have shown that the 3D structure of the local GMF can be effectively probed without relying on analytical models, using data from upcoming stellar polarization surveys and stellar parallax measurements from Gaia. This could significantly improve our ability to correct for the GMF's influence on the trajectories and observed arrival directions of UHECRs, if the UHECR charges are known. The issue is further complicated from the fact that UHECR charges are not directly measured, and can only be inferred from other observables.

Aims. In this work, our aim is to: a) test whether the lack of directional information on the GMF measurements derived from stellar polarimetry can be compensated by prior knowledge as provided from analytical models, b) investigate the effect of varying UHECR composition on arrival direction corrections using GMF reconstructions based on optopolarimetry.

Methods. We apply a recently developed Bayesian method for the 3D reconstruction of the local GMF to a simulated field generated from a dynamo MHD simulation. By further simulating mock event datasets and considering different probability distributions over UHECR charges, we investigate how effectively it is possible to recover the original arrival directions.

Results. For a GMF with an approximate magnitude of $5\mu\text{G}$, the deflection of proton UHECRs can be significantly corrected, with a residual bias of about 10° (approximately 2σ) at worst. For UHECRs with intermediate charge (for instance, nitrogen nuclei), our GMF reconstructions still provide substantial improvements over the current knowledge, particularly in regions of the sky corresponding to directions perpendicular to the GMF. However, UHECRs with the largest possible charges, such as iron nuclei, diffuse to the extent that their true arrival directions are largely irrecoverable. For GMFs with a magnitude of the order of $1\mu\text{G}$, a significant correction can be achieved over a large portion of the sky, even for iron nuclei.

Key words. Galactic magnetic field – Ultra high energy cosmic ray sources – Interstellar turbulence

1. Introduction

Even though more than 60 years have passed since the first observation of ultra-high energy cosmic rays (Linsley 1963), their origin remains a matter of debate. Advancements in our understanding of their origins would provide significant information relating to the physical processes that lead to their production. Furthermore, this would also contribute additional observables to be used in multimessenger approaches to the study high-energy astrophysical phenomena (e.g., Fang & Murase 2018; Murase 2019).

The problem of determining their sources stems from the fact that UHECRs, being charged particles, interact with magnetic fields that deflect their paths, masking their origins. Throughout their travelling time from the source to the observer, the ambient magnetic fields will alter their direction. In recent years, a sig-

nificant number of events have been recorded by the two leading experiments, Pierre Auger Observatory and the Telescope Array (Kawai et al. 2008; Sagawa 2020), whose results suggest the existence of UHECR ‘hotspots’ (Pierre Auger Collaboration et al. 2017; Kawata et al. 2019; Abbasi et al. 2014).

The Galactic magnetic field (GMF) is a significant source of UHECR deflections (Stanev 1997; Prouza, M. & Šmíd, R. 2003; Eichmann & Winchen 2020; Higuchi et al. 2023). Accurately correcting for its effect would require its 3D representation to sufficient precision, for the region of the Galaxy through which each UHECR has travelled. As most available observations of the GMF are integrated along the line of sight (LOS), this task is extremely challenging. However, 3D parametric models have been proposed, where the GMF is modeled by fitting parameters of predefined components with an analytic description; a toroidal, poloidal, as well as turbulent fields, with the latter often modeled as a Gaussian random field (Sun et al. 2008; Sun

* tsouros@physics.uoc.gr

& Reich 2010; Takami & Sato 2010; Jansson & Farrar 2012a; Jansson & Farrar 2012b). These models form the state of the art in the field of study that concerns UHECRs and their origins, and have formed the basis in a variety of studies (Kobzar et al. 2019; Globus et al. 2019; Magkos & Pavlidou 2019; Heinze et al. 2019; Watanabe, Keito et al. 2023; Allard, D. et al. 2024; Bister et al. 2024).

Recent developments, however, have made it possible to probe the 3D structure of the local interstellar medium directly. The Gaia mission has provided precise measurements of stellar parallaxes, resulting in a detailed map of over a billion stars in the Galaxy (Gaia Collaboration et al. 2016; Gaia Collaboration et al. 2021; Bailer-Jones et al. 2021). When combined with additional spectroscopic data, this information has been used to create 3D tomographic maps of the dust density in certain regions of the Galaxy (Lallement et al. 2018; Green et al. 2019; Lallement et al. 2019; Leike & Enßlin 2019; Leike et al. 2020; Lallement et al. 2022; Leike et al. 2022; Edenhofer et al. 2023). These maps, however, primarily focus on dust and do not provide direct constraints on the magnetic field itself.

There do exist observational techniques that can provide 3D data on the GMF. One such technique is the measurement of starlight polarization. While light from stars is generally emitted unpolarized, it can become polarized through dichroic absorption by aligned interstellar dust grains that follow the magnetic field lines (Andersson et al. 2015). Taking advantage of this phenomenon, future optopolarimetric surveys, including PASIPHAE and SouthPol, are expected to provide high-precision polarization data for millions of stars (Magalhães 2012; Tassis et al. 2018; Maharana et al. 2021; Maharana et al. 2022), especially at high Galactic latitudes where UHECR hotspots are located. When this data is combined with the stellar distances measured by Gaia, tomographic reconstructions of the GMF's POS component in regions with sufficient dust are made possible (Davis 1951; Chandrasekhar & Fermi 1953; Panopoulou et al. 2017; Skalidis et al. 2021; Skalidis & Tassis 2021; Pelgrims, V. et al. 2023). Indeed, optical stellar polarimetry has already been used in order to probe the GMF in 3D in the Sagittarius spiral arm (Doi et al. 2021; Doi et al. 2024).

Insights into the local GMF structure can also be obtained from HI gas observations across different velocity bins (Tritsis et al. 2018; Tritsis et al. 2019; Clark & Hensley 2019), which, when used alongside LOS-integrated data (e.g., Tahani et al. 2022a; Tahani et al. 2022b), offer the potential for creating localized, sparse GMF maps. These data are crucial for constructing 3D tomographic maps of specific regions, which in turn would enable more accurate UHECR backtracking and source localization. There is thus significant interest in reconstructing the GMF in directions where UHECR hotspots are observed.

An additional challenge is that the UHECR charge is not directly measured. Rather, it is inferred from the atmospheric depth where the energy deposit profile of secondary particles from UHECR-caused extensive air showers reaches its maximum. This is directly tied to the depth at which the first interaction of the UHECR primary particle occurred in the atmosphere, which naturally depends on its cross-section and, therefore, in the case of atomic nuclei, its mass (Gaisser & Hillas 1977; Mayotte et al. 2023). Generally, then, a distribution over UHECR charges is inferred (Abbasi et al. 2019; Arsene 2021; Salamida 2023; Mayotte et al. 2023; Abbasi et al. 2024).

This work follows the development of two previous works, Tsouros et al. 2024b and Tsouros et al. 2024a, hereafter referred to as Paper I and Paper II, respectively. These works introduced a framework for reconstructing the GMF in 3D within a mock

setting, in anticipation of future real optopolarimetric measurements. The inverse problem was addressed in a Bayesian setting, where the objective was to sample the posterior distribution of GMF configurations in a specific Galactic region conditional to a combination of local data on the POS as well as LOS-integrated data. For the former, the synthetic local data was assumed to be inferred from tomographic observations of magnetized interstellar dust via starlight optopolarimetry that provided the POS component of the GMF on the location of dust clouds. In contrast, LOS-integrated observations, such as those obtained from Faraday rotation measurements, offer insights into the LOS component of the magnetic field (Pandhi et al. 2022; Hutschenreuter et al. 2023). The methodology relied on information field theory, a framework developed for Bayesian inference of fields (Enßlin et al. 2009; Enßlin 2019; Enßlin 2022). To evaluate the quality of the reconstructions, UHECR particles with known rigidities and arrival directions were back-traced through the reconstructed GMF in the region of interest. The recovered arrival directions were then compared to the observed ones in order to assess our ability to correct for the bias introduced by the GMF.

In this work, we perform the reconstruction of the GMF using synthetic data as was done in Paper I and Paper II, but in this case our goal is to apply these results to different distributions over the UHECR charge, ranging from pure proton composition to the other extreme, pure iron composition, with the same observed energy, as well as intermediate charges of the UHECR nuclei. In addition, we relax one of the main assumptions made in the two previous works, namely, the that the local GMF data have a known directionality. In real applications, the stellar polarization can provide the orientation of the GMF on the POS as well as the magnitude of its POS component, at specific locations. Importantly, however, the direction of the GMF is not probed, and must somehow be guessed. In this work, we supplement realistic (i.e. non-directional) simulated measurements with an ad hoc assumption, namely that the POS direction aligns with the POS direction of the ordered component of available parametric models. A goal of this work is to test this assumption by applying it on a synthetic scenario.

Section 2 summarizes the Bayesian modelling and sampling procedure. The main results for the charge distributions considered are presented in Section 3, and further discussion of these results is provided in Section 4.

2. Method

Our task is to infer the configuration of the GMF, $\mathbf{B}(\mathbf{x})$ within a domain \mathcal{V} , given some observed data set d . In doing so we are tasked with determining the posterior probability distribution of $\mathbf{B}(\mathbf{x})$ conditional to d :

$$P(\mathbf{B}|d) = \frac{1}{\mathcal{Z}} P(d|\mathbf{B})P(\mathbf{B}). \quad (1)$$

Here, $P(d|\mathbf{B})$ is the likelihood, or the probability of observing magnetic field measurements d given a specific configuration $\mathbf{B}(\mathbf{x})$. The prior, $P(\mathbf{B})$, encapsulates pre-existing information about $\mathbf{B}(\mathbf{x})$ while $\mathcal{Z} = P(d)$ is the normalization factor (see also relevant discussion in Paper I and Paper II).

In this work, the field that serves as a ground truth (the ‘true’ field) is generated from a dynamo MHD simulation, discussed extensively in Bendre & Subramanian 2022 and the Appendix of Paper II. The original simulation domain extended to ≈ 1 kpc in the $x - y$ direction, and ≈ 2 kpc above the Galactic plane. We rescale the simulated GMF so that its root-mean-square (RMS)

value is $5\mu\text{G}$ or $1\mu\text{G}$. We choose two such RMS values in order to study the effect of the GMFs strength on our ability to recover UHECR arrival directions (see section 3.3).

2.1. Prior

As we are dealing with a magnetic field, we need to impose the condition that any candidate of reconstruction must satisfy $\nabla \cdot \mathbf{B} = 0$. This is ensured by working with a random vector field that does not respect the divergence free condition, φ , but is related to the observed field by a divergence cleaning operator \mathcal{P} . In Fourier space, this is defined as

$$\mathcal{P}_{ij}(\mathbf{k}) = \delta_{ij} - \hat{k}_i \hat{k}_j^T, \quad (2)$$

where δ_{ij} is the Kronecker delta and k_i is the i -th component of the wavevector. Thus, this operator is defined so as to project out the degrees of freedom of the Gaussian random vector field that violate the divergence-free condition.

$$\hat{B}_i(\mathbf{k}) = \frac{3}{2} \mathcal{P}_{ij}(\mathbf{k}) \hat{\varphi}_j(\mathbf{k}), \quad (3)$$

where the hatted fields denote Fourier modes. Eq. 3 ensures that $\nabla \cdot \mathbf{B} = 0$.

For φ we assume a Gaussian prior of the form:

$$\mathcal{P}(\varphi) = \frac{1}{|2\pi\Phi|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} \int d^3x d^3x' \sum_{ij} \varphi_i(\mathbf{x}) \Phi_{ij}^{-1}(\mathbf{x}, \mathbf{x}') \varphi_j(\mathbf{x}') \right]. \quad (4)$$

The quantity Φ_{ij} is the covariance matrix, defined as

$$\Phi_{ij}(\mathbf{x}, \mathbf{x}') = \langle \varphi_i(\mathbf{x}) \varphi_j^*(\mathbf{x}') \rangle, \quad (5)$$

where the symbol $\langle \dots \rangle$ signifies an average over the distribution $P(\varphi)$. The Gaussian form of Eq. 5 should not be taken as a physical assumption, but rather as a statement of ignorance: as we do not assume any reliable prior information regarding the breaking of isotropy, homogeneity, and helicity, our prior is symmetric in those respects. The data, however, can inform the posterior with regard to such violations of symmetry, as they are indeed expected to occur.

Thus, in our analysis, we choose not to integrate any prior knowledge about the GMF geometry and statistics, so we use a prior distribution exhibiting statistical isotropy, homogeneity, and mirror symmetry (this is imposed by not incorporating a helicity term into the field correlation tensor Φ). The assumption of statistical homogeneity and isotropy in the prior implies a covariance that is diagonal in Fourier space, through the Wiener-Khinchin theorem. Therefore, the Fourier space covariance in the form

$$\langle \hat{\varphi}_i(\mathbf{k}) \hat{\varphi}_j^*(\mathbf{k}') \rangle = (2\pi)^3 \delta_{ij} \delta^{(3)}(\mathbf{k} - \mathbf{k}') P(k), \quad (6)$$

or, equivalently,

$$\langle \delta \hat{B}_i(\mathbf{k}) \delta \hat{B}_j^*(\mathbf{k}') \rangle = \frac{(2\pi)^3}{2} \mathcal{P}_{ij}(\mathbf{k}) \delta^{(3)}(\mathbf{k} - \mathbf{k}') P(k), \quad (7)$$

where

$$\delta \hat{B}_i \equiv B_i - \langle B_i \rangle. \quad (8)$$

The power spectrum $P(k)$ is modelled as an integrated Wiener process on a logarithmic scale $l = \log(|k|)$ for $k \neq 0$. Specifically, the amplitude spectrum is expressed as:

$$P(k) = ae^{2\gamma(l)}, \quad \frac{d^2\gamma}{dl^2} = \eta \xi_W(l), \quad (9)$$

where $\xi_W(l)$ is distributed according to

$$\mathcal{P}(\xi_W) = \mathcal{G}(\xi_W, \mathbf{I}). \quad (10)$$

The integration of the above process yields:

$$\gamma(l) = ml + \eta \int_{l_0}^l \int_{l_0}^{l''} \xi_W(l'') dl'' dl', \quad (11)$$

where l_0 is the first mode greater than zero, m defines the slope of the amplitude spectrum on the double logarithmic scale, and η is a flexibility parameter controlling the total variance of the integrated Wiener process. The hyperparameters m , η , and a are also inferred from the data along with the GMF configuration, and have their own hyperpriors.

2.2. Likelihood

A complete discussion of tomography of the magnetized ISM using stellar polarization measurements is beyond the scope of this work (but the reader can consult Tassis et al. 2018, Pelgrims, V. et al. 2023, and Pelgrims, V. et al. 2024 for more information). In brief, the central idea is that from stellar polarization data for stars with a measured parallax from Gaia, one can extract information about the Stokes parameters imposed by each intervening dust cloud on the observed starlight. This method is effective provided there is a sufficient density of stars with polarization measurements and known distances. The resulting data can then be used to infer the local orientation of the POS component of the Galactic magnetic field (GMF) within each cloud, which is in turn linked to the alignment of dust grains in the cold neutral medium (Tassis et al. 2018). Additionally, information about the POS component of the GMF in clouds can be obtained through 21 cm neutral hydrogen (HI) emission measurements (Clark & Hensley 2019). Finally, the magnitude of the POS component of the GMF at the location of the dust cloud can be ascertained with standard techniques (Chandrasekhar & Fermi 1953; Skalidis et al. 2021). For the purposes of this work, we assume that the task of associating these measurements with their corresponding locations has already been completed. Therefore, our final measurements will consist of 3-dimensional vectors on the inferred locations of dust clouds.

However, an important issue should be addressed. The local information on the GMF on dust clouds acquired from stellar optopolarimetry provides information on the orientation of the POS component of the GMF, but not its direction (Pelgrims, V. et al. 2023; Pelgrims, V. et al. 2024). In this work, we assume that the direction (and only the direction) of the large scale field is known and is imposed a priori on each local data point. Thus, the direction of the local data will be chosen so that their inner product with known large scale direction is positive, while the orientation on the POS will be the one measured. This imposes a directional prior, but it does so implicitly. In real applications we will take this known direction to align with the direction of the local large scale GMF as derived from parametric models, such as the JF12 field, whose parameters are fitted using LOS integrated rotation measure and synchrotron data (Jansson & Farrar 2012a; Jansson & Farrar 2012b). Of course, the orientation of the

Table 1: Power spectrum hyperparameters and their hyperpriors

Parameter	Distribution	Mean	Standard deviation
m	Normal	$-\frac{11}{3}$	1
η	Log-normal	1.5	1
a	Log-normal	$1 \mu\text{G}$	$1 \mu\text{G}$

GMF at the location of dust clouds is not expected to align completely with the large scale field, even within the small section of the Galaxy that is of interest to us. Comparing with the ground truth GMF, this choice of POS direction for local measurements is wrong 8% of the time.

In light of the above, in constructing our likelihood model we will assume that the POS projection of the GMF is known at specific locations in the local Galaxy¹, scattered around the Sun. Thus for the i -th datapoint, we have a forward model of the form

$$\mathbf{d}_{\text{local}}^{(i)} = \int R_{\text{local}}(\mathbf{x}, \mathbf{x}_i) \mathbf{B}(\mathbf{x}) d^3x + \mathbf{n}_{\text{local}}^{(i)}, \quad (12)$$

$$R_{\text{local}}(\mathbf{x}, \mathbf{x}_i) \equiv \delta^{(3)}(\mathbf{x} - \mathbf{x}_i) P_{\text{POS}}, \quad (13)$$

$\mathbf{B}(\mathbf{x})$ is the magnetic field, and $\mathbf{n}_{\text{local}}^{(i)}$ are the observational uncertainties that contaminate our measurements. The vector \mathbf{x}_i denotes the location of the i -th cloud where the magnetic field is measured, P_{POS} signifies the projection operator on the POS, which reflects that the POS component of the magnetic field is measured via dust polarization, $P_{\text{POS},ij} = \delta_{ij} - \hat{x}_i \hat{x}_j^\top$ with $\hat{x}_i = x_i / \|x_i\|$ (assuming the observer to be at the origin). The Dirac delta function localizes the measurements at specific known locations \mathbf{x}_i .

Note that while the Dirac distribution implies that accurate 3D locations are measured, in reality they will not be known exactly. As also noted in Paper II and discussed in section 3.1 the resolution of our discretized domain is of the order of the expected cloud localization errors, which corresponds to the uncertainty of cloud localization (Pelgrims, V. et al. 2023).

The contaminating vector $\mathbf{n}_{\text{local}}^{(i)}$ is assumed to be a random vector on the POS, drawn from a Gaussian distribution with a known covariance N_{local} . Marginalizing over the noise, the likelihood becomes

$$P(\mathbf{d}|\mathbf{B}) = \mathcal{G}(\mathbf{d}_{\text{local}} - R_{\text{local}} \mathbf{B}, N_{\text{local}}). \quad (14)$$

The covariance N_{local} is chosen to be a multiple of the identity, $(N_{\text{local}})_{ij} = \sigma^2 \delta_{ij}$, where we choose

$$\sigma = \frac{|\mathbf{B}|_{\text{RMS}}}{2}, \quad (15)$$

where $|\mathbf{B}|_{\text{RMS}}$ is the RMS value of the magnitude of the ground truth, which in this work is taken to be $5 \mu\text{G}$ and $1 \mu\text{G}$. It should be noted that this does not imply that the noise is correlated with the GMF covariance, it is merely chosen as such in order to model a realistic error of the measurement of the POS component of the GMF's magnitude (Skalidis et al. 2021).

In addition we consider integrated Faraday LOS data (Hutschenreuter et al. 2023). In this case, the forward model takes the form

$$d_{\text{int}}^{(i)} = (\overline{P_{\text{LOS}} \mathbf{B}})_{L_i} + n_{\text{int}}^{(i)}, \quad (16)$$

$$(\overline{P_{\text{LOS}} \mathbf{B}})_{L_i} \equiv \frac{1}{|L_i|} \int_0^{|L_i|} B_{\parallel}(\mathbf{x}) d\ell, \quad (17)$$

¹ The locations of the inferred dust clouds.

where P_{LOS} projects a vector onto the LOS component (B_{\parallel}), and L_i is the specific LOS under consideration. Further, $|L_i|$ denotes the limit up to which we integrate, which coincides with the distance between the Earth and the intersection of L_i with the boundary of \mathcal{V} , in this application. The vector $n_{\text{int}}^{(i)}$ denotes the contaminating noise of the LOS measurements.

This likelihood implies that the electron density is constant up to a distance $|L_i|$ and then falls to zero. While this assumption is not realistic, we will maintain it in this proof-of-concept work for simplicity, and mitigate it by introducing a noise variance, $N_{\text{int}} = \sigma_{\text{int}}^2 \delta_{ij}$, with $\sigma = 0.5$, which is much larger than expected from Faraday measurements (Hutschenreuter et al. 2023; Hutschenreuter, S. et al. 2022). In applications to real data, the thermal electron density will be an additional degree of freedom that will also be inferred, alongside the GMF.

As the likelihoods of equations 12 and 16 are independent, the total likelihood is their product:

$$P(\mathbf{d}|\mathbf{B}) = \mathcal{G}(\mathbf{d}_{\text{local}} - R_{\text{local}} \mathbf{B}, N_{\text{local}}) \mathcal{G}(d_{\text{int}} - (\overline{P_{\text{LOS}} \mathbf{B}})_{L_i}, N_{\text{int}}). \quad (18)$$

2.3. Posterior of UHECR arrival directions

From the above discussions, we are now equipped with the posterior for the GMF within \mathcal{V} subject to the data d , i.e. the distribution $P(\mathbf{B}|d)$ (here for simplicity we use the symbol d to refer to the totality of data to which the GMF must be constrained).

In addition, and independently of the data on the GMF, d , we have access to data relating to a UHECR:

$$D \equiv \{E, Z, \theta_{\text{obs}}\}, \quad (19)$$

where E is the the UHECR's observed energy, Z is the atomic number of the nucleus, and θ_{obs} is the observed arrival direction on the celestial sphere. What we are ultimately interested in is the posterior

$$P(\theta|d) = \int P(\theta|\mathbf{B}, D) P(\mathbf{B}|d) d\mathbf{B} P(D) dD, \quad (20)$$

where θ is the inferred arrival direction on the celestial sphere, and $P(D)$ are the distributions of the independently measured D . In this work, we will assume for simplicity that E is known with minimal uncertainty, and has a value of $E = 5 \times 10^{19}$ eV, which corresponds to the GZK limit for protons (Greisen 1966; Zatsepin & Kuzmin 1966). We will additionally assume that θ_{obs} is perfectly known, but we will consider all of its possible values, for completeness.

While E and θ_{obs} are directly observable quantities, the mass can only be inferred indirectly. As discussed in the Introduction, what is directly measured is the atmospheric depth where the energy deposit profile of secondary particles from extensive air showers reaches its maximum, X_{max} , which is related to the particle charge via an inference of its mass, through the cross-section and the number of first-collision daughter particles, which are usually extrapolated from the Standard Model to the observed energies, or are calculated from alternative theories at those energies (Gaisser & Hillas 1977; Mayotte et al. 2023; Abbasi et al.

2019; Arsene 2021; Salamida 2023; Mayotte et al. 2023; Romanopoulos et al. 2022a; Romanopoulos et al. 2022b). Thus, regarding charge, one has to use a distribution $P(Z|X_{\max}, \text{Model})$, where the dependence on the particle physics model at the observed energies is made explicit. We will drop the mention to X_{\max} in order to declutter notation, and we will merely refer to the prior information on charges as $P(Z)$.

A realistic modelling of $P(Z)$ is beyond the scope of this analysis. In this work, we will only consider four simple cases that capture the two extremes of the parameter space for Z : a pure proton composition and a pure iron nuclei composition, as well as two cases with moderate primary charge: pure nitrogen nucleus composition, and a mixed proton/nitrogen nucleus composition, in order to test our ability to correct for the effect of the GMF using our method, for each of these scenarios. Therefore, we will consider four priors for the UHECR atomic number:

1. Scenario 1 (Pure proton composition):

$$\begin{cases} P(Z) = 1, & \text{if } Z = 1 \\ P(Z) = 0 & \text{otherwise} \end{cases} \quad (21)$$

2. Scenario 2 (Pure nitrogen composition):

$$\begin{cases} P(Z) = 1, & \text{if } Z = 7 \\ P(Z) = 0 & \text{otherwise} \end{cases} \quad (22)$$

3. Scenario 3 (Mixed proton/nitrogen composition):

$$\begin{cases} P(Z) = 0.8, & \text{if } Z = 1 \\ P(Z) = 0.2, & \text{if } Z = 7 \\ P(Z) = 0 & \text{otherwise} \end{cases} \quad (23)$$

4. Scenario 4 (Pure iron composition):

$$\begin{cases} P(Z) = 1, & \text{if } Z = 26 \\ P(Z) = 0 & \text{otherwise} \end{cases} \quad (24)$$

With the prior $P(Z)$ chosen, and assuming that samples from the posterior $P(\mathbf{B}|d)$ are available (see section 3.2), we are able to sample the distribution of arrival directions 20, and compare our results with the true arrival direction. The metric that we will use is the mean angular distance over the distribution $P(\theta|d)$, which we call the bias, that is, the quantity

$$\langle \delta\theta \rangle_{\theta|d} \equiv \int f(\theta_{\text{true}}, \theta) P(\theta|d) d\theta, \quad (25)$$

where $f(\cdot, \cdot)$ denotes the angular (great circle) distance between the two arguments. In addition to the bias, we are interested in the respective standard deviation. For each observed arrival direction θ_{obs} , the bias and the respective standard deviation are computed with the simple procedure described in Algorithm 1. Of course, due to the finite number of samples that are drawn these answers are only approximate, but are guaranteed to converge to the true values as $N \rightarrow \infty$ from the law of large numbers.

Algorithm 1 Calculate $\langle \delta\theta \rangle_{\theta|d}$ and $\sigma_{\theta|d}$ for a given θ_{obs} , E

Require: Posterior samples $\{\tilde{\mathbf{B}}_i\}_{i=1}^N \sim P(\mathbf{B}|d)$, ground truth GMF configuration \mathbf{B}_{true} , observed arrival direction θ_{obs} , observed energy E , probability distribution $P(Z)$ for UHECR charge

Ensure: Bias $\langle \delta\theta \rangle_{\theta|d}$ and standard deviation $\sigma_{\theta|d}$

- 1: **Compute True Direction:**
 - 2: Sample $Z_{\text{true}} \sim P(Z)$
 - 3: Back-trace a UHECR with parameters $\{\theta_{\text{obs}}, E, Z_{\text{true}}\}$ through \mathbf{B}_{true} , to compute the direction of the velocity the particle had when it entered \mathcal{V} . This is θ_{true} .
 - 4: Initialize list $\delta\theta \leftarrow \{\emptyset\}$
 - 5: **for** each sample $\tilde{\mathbf{B}}_i$ from $\{\tilde{\mathbf{B}}_i\}_{i=1}^N$ **do**
 - 6: Sample $Z_i \sim P(Z)$
 - 7: Back-trace UHECR with parameters $\{\theta_{\text{obs}}, E, Z_i\}$ through $\tilde{\mathbf{B}}_i$, to compute θ_i .
 - 8: Compute angular deviation and append $f(\theta_{\text{true}}, \theta_i)$ to $\delta\theta$
 - 9: **end for**
 - 10: **Compute average angular deviation and its standard deviation:**
 - 11: $\langle \delta\theta \rangle_{\theta|d} \leftarrow \frac{1}{N} \sum_{i=1}^N \delta\theta_i$
 - 12: $\sigma_{\theta|d} \leftarrow \sqrt{\frac{1}{N} \sum_{i=1}^N [\delta\theta_i - \langle \delta\theta \rangle_{\theta|d}]^2}$
 - 13: **return** $\langle \delta\theta \rangle_{\theta|d}, \sigma_{\theta|d}$
-

In Algorithm 1, the initial arrival directions, true or inferred, are defined as the direction of the velocity on the celestial sphere the UHECR had when entering \mathcal{V} . Of course this is true only whenever the boundary of \mathcal{V} coincides with the boundary of the Galaxy. As our local measurements are expected to reach a maximum distance of about 2 kpc - up to which Gaia data is available, this is expected to only be true in high Galactic latitudes. For completeness, in this work we will consider all points of the sky.

3. Results

In what follows we will test the extent to which we can leverage local as well as integrated GMF data in order to correct for the effect of the GMF on the observed arrival direction of each observed UHECR.

3.1. Generation of synthetic data

In order to generate the synthetic data, we will use the dynamo simulation of Bredre & Subramanian 2022, also briefly discussed in Appendix A of Paper II. Briefly, the GMF that will serve as our ground truth was generated by a Magnetohydrodynamic simulation of the Galactic ISM. The simulation domain is located roughly at the solar neighbourhood of the Milky Way. It has dimensions of $\approx 0.8 \times 0.8$ kpc in the radial x and azimuth y direction and ranges from ≈ -1.6 to $\approx +1.6$ kpc in z direction, where $z = 0$ is the Galactic plane. Even though the original simulation was performed on a Cartesian grid with a resolution of 8.3 pc, we average out smaller degrees of freedom and reach a cruder resolution of 33.3 pc. The reason for this is twofold: first, running the reconstruction on a grid of a resolution of 8.3 pc would unnecessarily increase the computation time of the problem. The intended use of the reconstruction is to trace UHECRs back to their original arrival directions. For a μG strength, the Larmor radius of a 10^{19} eV UHECR proton is of kpc order, and so maintaining the original 8.3 pc resolution would not meaningfully improve the backtracking accuracy. The second reason

is that the expected uncertainty on the cloud localization (and by extension, the uncertainty on the local GMF measurement) is of the order of 10 pc (Pelgrims, V. et al. 2023). Therefore, the uncertainty on the localization of the GMF measurements (which was not included in the likelihood) is of the order of the grid resolution, and will therefore not be considered, as it will tend to be smaller or of the order of the grid resolution in case the grid resolution required is smaller. Finally, taking advantage of the simulation's periodic boundary conditions on the $x \times y$ plane, we double the x and y directions in order to achieve a domain of approximately 1.6×1.6 kpc along the Galactic plane. In order to test the sensitivity of our results to the strength of the magnetic field, we rescale it so that it has a strength of $1\mu\text{G}$ or $5\mu\text{G}$. The "ground truth" is shown in Fig. 1a for $5\mu\text{G}$.

Once the ground-truth GMF is available, we apply the same operations described above to the corresponding temperature field produced from the same MHD simulation that produced the ground-truth, in order to end up with a grid of the same dimensions. We then apply R_{local} which applies a Bernoulli trial to choose whether a voxel is observed or not. The probability of the trial is chosen so that the density of measurements is on average 100 per cubic kpc when the temperature exceeds the threshold of 10^4 K, and 1000 otherwise. This ensures that the spacing between the clouds, where measurements are assumed to be located, remains on the order of a few hundred parsecs. The motivation here is that dust clouds are much more frequent in the cold ISM (Panopoulou & Lenz 2020), and our measurements must reflect this inhomogeneity. The resulting synthetic data is shown in Fig. 1b.

For the integrated data, the ground truth from Fig. 1a is first converted to spherical coordinates and then integrated along the radial coordinate. After adding noise, this process yields the data presented in Fig. 2b.

3.2. GMF Reconstruction

The sampling of the posterior distribution, conditional to the data shown in Figs. 1b and 2b, is performed using the Numerical Information Field Theory package in Python, NIFTy² (Selig et al. 2013; Steininger et al. 2017; Arras et al. 2019; Edenhofer et al. 2024), which utilizes the geoVI algorithm (Frank et al. 2021), a variational inference algorithm geared towards high dimensional problems such as field inference. A brief overview of this algorithm is provided in Appendix A of Paper I.

Fig. 1a displays the ground truth with an RMS of $5\mu\text{G}$, while Fig. 1c shows the mean of the 100 posterior samples drawn from the distribution conditional to the data of Figs. 1b and 2b. In Fig. 2c we show the same field as in Fig. 1c, but integrated along the line of sight. Finally note that the correlation length of the GMF is of the order of 100 pc, as shown in Paper II and Bendre & Subramanian 2022.

In Fig. 3 we plot the power spectrum prior (blue-shaded area), as well as the power spectra of the posterior samples (red-shaded area). Comparison of the latter family of power spectra with the true power spectrum (black line). The two vertical dashed lines corresponds to the two lengthscale limits of our problem: the linear scale of \mathcal{V} and the linear scale of the smallest voxel.

² The documentation can be found in ift.pages.mpcdf.de/nifty/index.html.

As a measure of the success of the reconstruction, we define:

$$\chi_B^2 = \frac{\langle |\mathbf{B} - \bar{\mathbf{B}}|^2 \rangle_{\mathcal{V}}}{\langle |\mathbf{B}|^2 \rangle_{\mathcal{V}}}, \quad (26)$$

where \mathbf{B} is the true field, $\bar{\mathbf{B}}$ is the posterior mean, and $\langle \dots \rangle_{\mathcal{V}}$ represents a spatial average over the domain \mathcal{V} . In practice, the accuracy of the reconstruction depends strongly on the specific physical application. For phenomena involving length scales much smaller than the scale resolved by the data, the reconstruction is inherently limited. Conversely, reconstructions are more reliable for length scales larger than those probed by the data. Given that the Larmor radii of UHECRs are comparable to the size of \mathcal{V} , they provide a useful benchmark for evaluating the quality of the reconstruction.

3.3. Correcting UHECR arrival direction from the GMF influence

In Fig 4, we plot the bias of our reconstructions for a ground truth GMF of $B_{\text{RMS}} = 5\mu\text{G}$. In these figures, each pixel corresponds to the observed arrival direction, θ_{obs} . Assuming a pure proton composition (Eq. 21), Figs 4a and 5a, we essentially recover the results presented in Paper II, where the maximum bias over all θ_{obs} on the celestial sphere is 13.3° , and the maximum variance is about 5.2° . For the same composition, the mean bias (standard deviation) is $2^\circ(1^\circ)$. The figures show that the most severe bias occurs for those θ_{obs} that are perpendicular to the GMF mean component, which in this setting corresponds to small longitudes.

In Figs. 4b and 5b we assume pure nitrogen nucleus composition only (Eq. 22). In this case the bias $\langle \delta\theta \rangle_{\theta|D}$ is especially large for small longitudes, reaching a maximum bias (standard deviation) of $157.8^\circ(59.8^\circ)$, whereas the mean bias (standard deviation) are $21.4^\circ(10.2^\circ)$ respectively. As is also the case of protons, for nitrogen nuclei that are observed arriving from directions perpendicular to the GMF, our reconstructions completely fail at recovering the arrival directions, due to the extreme diffusion that the primaries experience until they are observed on Earth. In real applications of reconstructed fields to UHECR data, extreme caution should be paid to this effect, and UHECRs with θ_{obs} that correspond to directions perpendicular to the GMF should not be considered in a backtracking analysis. However, as shown in Paper II, it is possible to infer the directions where the failure of our reconstruction is expected, assuming a particular UHECR charge. For completeness, in Figs 4c and 5c we assume a mixed proton/ nitrogen nucleus composition (Eq. 23), and proceed as in Algorithm 1 for all possible θ_{obs} .

Finally, as a worst case scenario, in Figs. 4d and 5d we assume a pure iron nucleus prior distribution (Eq. 24). In this case, apart from those θ_{obs} that directly correspond to directions parallel to the GMF, we are completely unable to recover the original arrival directions, and $\langle \delta\theta \rangle_{\theta|D}$ fluctuates randomly as a function of θ_{obs} due to the strong diffusion of the iron nucleus primaries.

In Figs. 6 and 7 we perform the same calculations as in 4 and 5, with the only difference being that now the ground truth GMF has a RMS value of $1\mu\text{G}$. In this case, we observe that even in the worst-case scenario of a pure iron composition, the mean bias (standard deviation) fall to $15.9^\circ(7.4^\circ)$.

The numerical results discussed in this section are tabulated in Table 2.

The systematic bias in our inference of the arrival direction is more pronounced for observed arrival directions that are perpendicular to the shearing direction of the GMF, as shown in Figs. 4

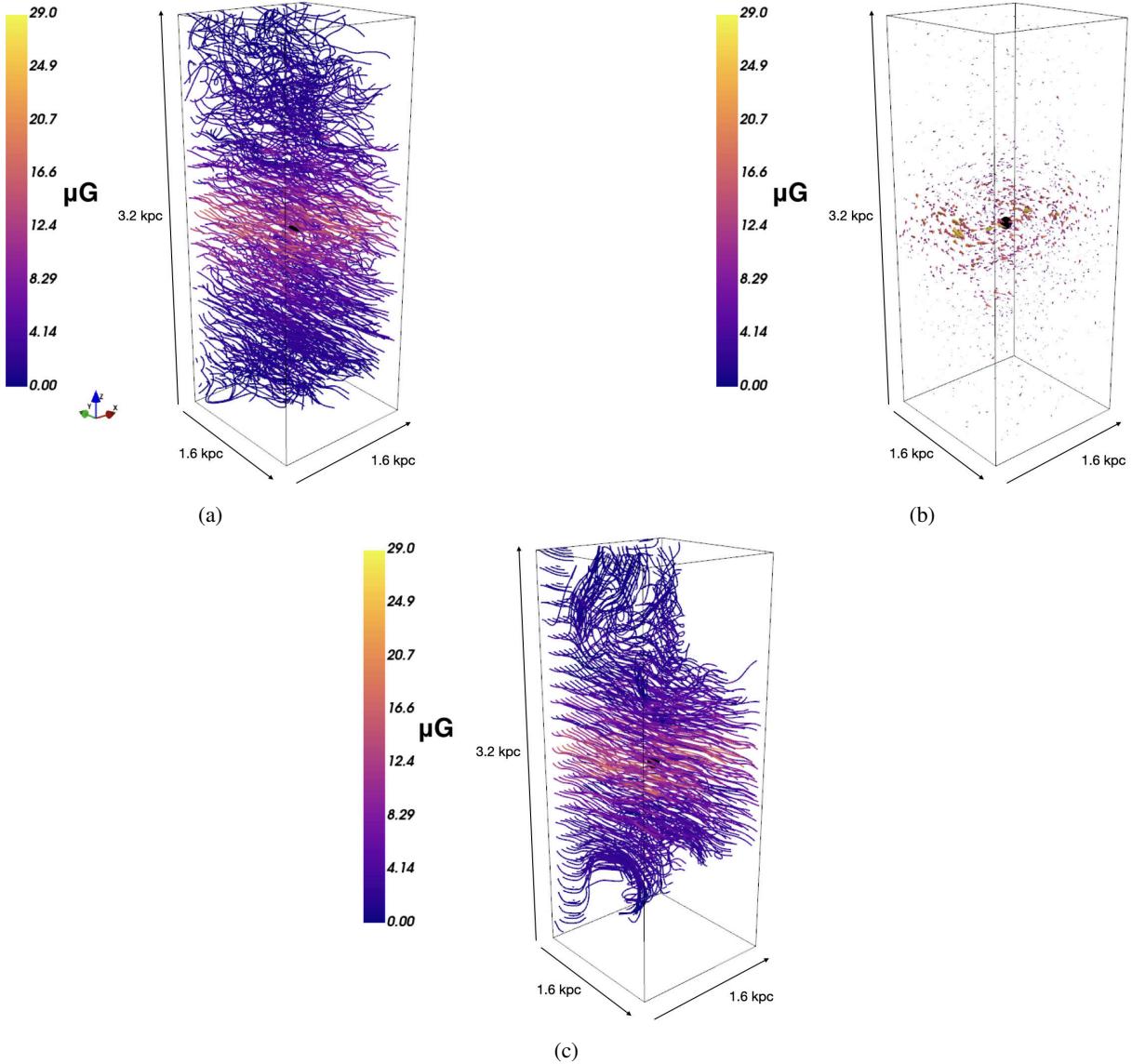


Fig. 1: Panel a: Assumed ground truth GMF. Panel b: The local and sparse data, confined to the POS. Panel c: Mean of the posterior distribution conditional to the data of Fig. 1b and 2b.

Table 2: Bias of the UHECR arrival direction inference from the GMF reconstructions

GMF RMS Magnitude	Scenario	Max Bias (deg)	Max st. dev. (deg)	Mean Bias (deg)	Mean st. dev. (deg)
$5 \mu\text{G}$	1 (Eq. 21)	13.3°	5.2°	2.4°	1.0°
	2 (Eq. 22)	157.8°	59.8°	21.4°	10.2°
	3 (Eq. 23)	160.5°	60.5°	18.8°	17.9°
	4 (Eq. 24)	140.6°	50.8°	71.5°	32.8°
$1 \mu\text{G}$	1 (Eq. 21)	2.6°	0.9°	0.5°	0.2°
	2 (Eq. 22)	24.3°	8.6°	3.6°	1.6°
	3 (Eq. 23)	35.6°	14.4°	4.0°	3.8°
	4 (Eq. 24)	149.7°	54.6°	15.9°	7.4°

and 6. This occurs because unresolved small-scale fluctuations, which are anisotropic due to the presence differential shear, produce a missing component to the inferred force acting on the UHECRs. This missing component is particularly significant for θ_{obs} perpendicular to the shearing direction.

4. Summary and Conclusions

Our results demonstrate that the lack of directional information on the GMF measured by polarization data can be compensated by prior knowledge of its mean direction, without significant error compared to previous studies that had assumed perfect knowledge of the POS directionality of the GMF. Thus, at least

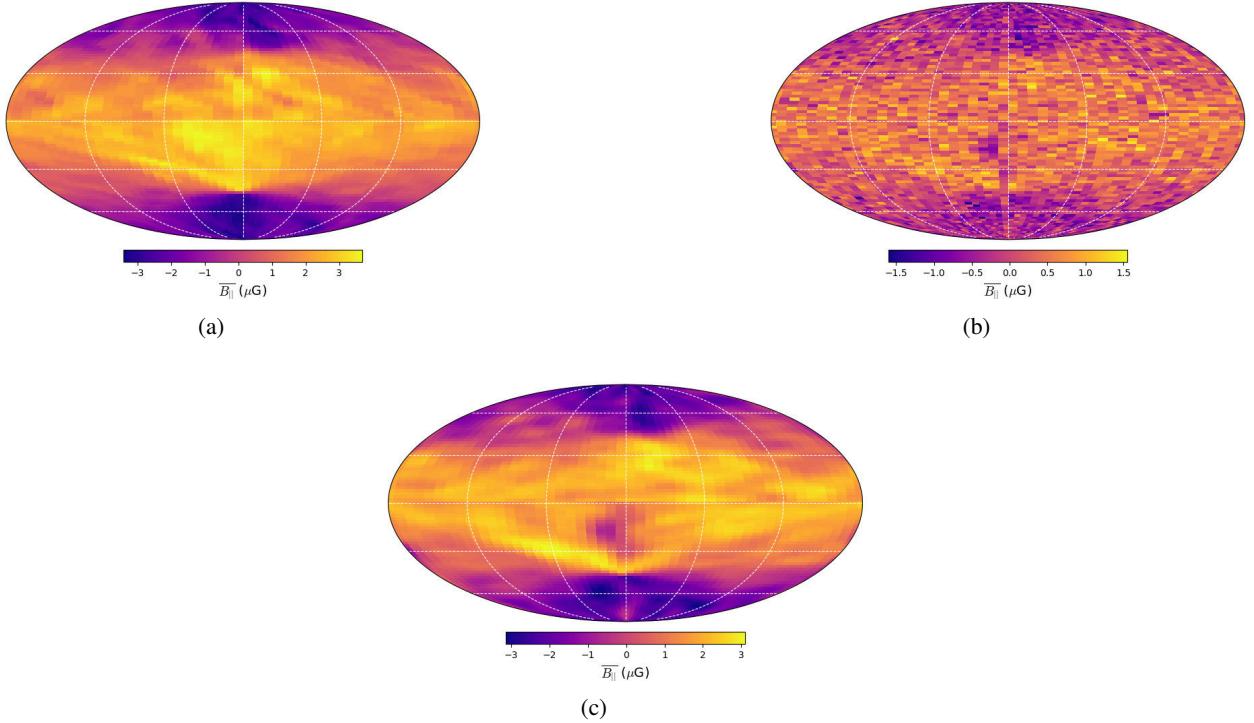


Fig. 2: Panel a: Assumed ground truth, integrated along the LOS. Panel b: The observed LOS-integrated component. Panel c: The mean of the posterior, integrated along the LOS.

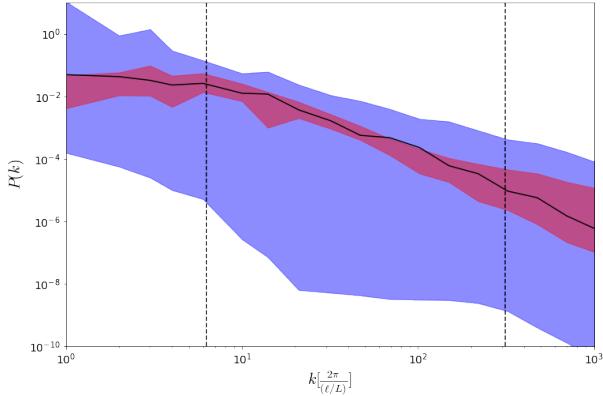


Fig. 3: The 3D power spectrum. The blue-shaded region represents the prior power spectrum, while the red-shaded region corresponds to the power spectra of the posterior samples. The black line depicts the power spectrum of the true field. The left vertical dashed line marks the natural length scale of the domain, \mathcal{V} , while the right vertical dashed line indicates the size of the smallest voxels. The wave vector norm is expressed in units of 2π divided by the linear size of the smallest voxel, normalized to the linear scale of \mathcal{V} .

to the extent that the GMF that we used for analysis is an captures the essential features of the true GMF, the knowledge of the direction of the GMFs mean component may be provided by the currently existing parametric models, which we assume have to capture it accurately.

We have shown that local GMF information derived from stellar polarisation measurements combined with LOS data can indeed be used to correct for the effect of the GMF on the observed UHECR arrival directions. For low Z and a RMS strength of $5\mu\text{G}$, we can recover the arrival directions to within a few degrees, if we assume a proton composition. For moderate Z (for example $Z = 7$ for nitrogen) we can infer the arrival directions with a mean bias of the order of 10° to 20° and a similar mean standard deviation. However, these reconstructions should only be carried out on parts of the sky that do not correspond to arrival directions perpendicular to the GMF mean component. The greatest deviations occur at the directions that coincide with the direction perpendicular to the mean GMF, as expected. These directions are however, independently inferred and thus known, for a given composition. For a weaker local GMF (of the order of $1\mu\text{G}$), this method can indeed yield significant improvements on our understanding of the origin of UHECRs.

Finally, in real life applications this analysis would only be applicable to high Galactic latitudes. Thankfully, upcoming surveys like PASIPHAE will provide accurate stellar polarisation exactly at high Galactic latitudes, where the observed hotspots are also situated.

Acknowledgements. A.T. and V.P. acknowledge support from the Foundation of Research and Technology - Hellas Synergy Grants Program through project MagMASim, jointly implemented by the Institute of Astrophysics and the Institute of Applied and Computational Mathematics. A.T. acknowledges support by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “Third Call for H.F.R.I. Scholarships for PhD Candidates” (Project 5332). V.P. acknowledges support by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the “First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant” (Project 1552 CIRCE). The research leading to these results has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie RISE action, Grant Agree-

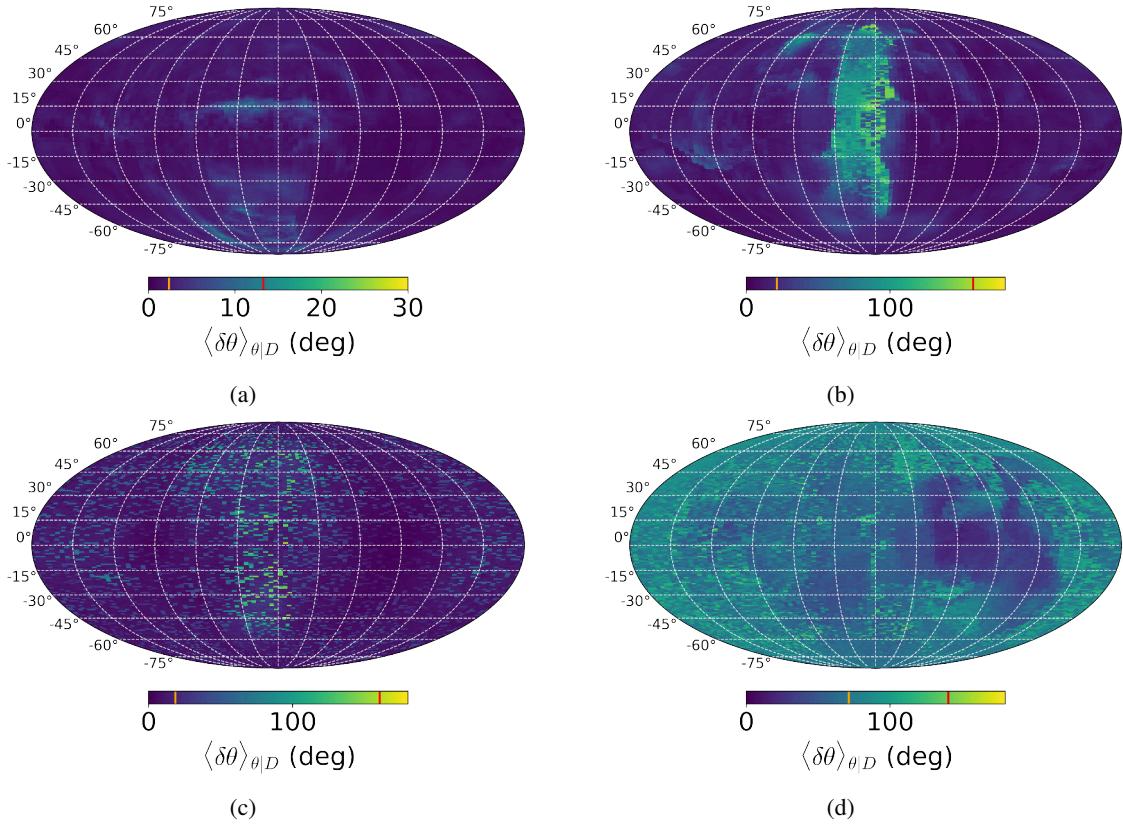


Fig. 4: Mean angular error of the reconstruction for four different composition scenarios. Each pixel corresponds to the observed arrival direction, θ_{obs} . The root-mean-square magnitude of the GMF is $B_{\text{RMS}} = 5 \mu\text{G}$, and the particles are assumed to have known energies of $E = 5 \times 10^{19} \text{ eV}$. The figure is in Galactic coordinates. Panel a: Pure proton composition. Panel b: Pure nitrogen nucleus composition. Panel c: Mixed proton/nitrogen nucleus composition. Panel d: Pure iron nucleus composition.

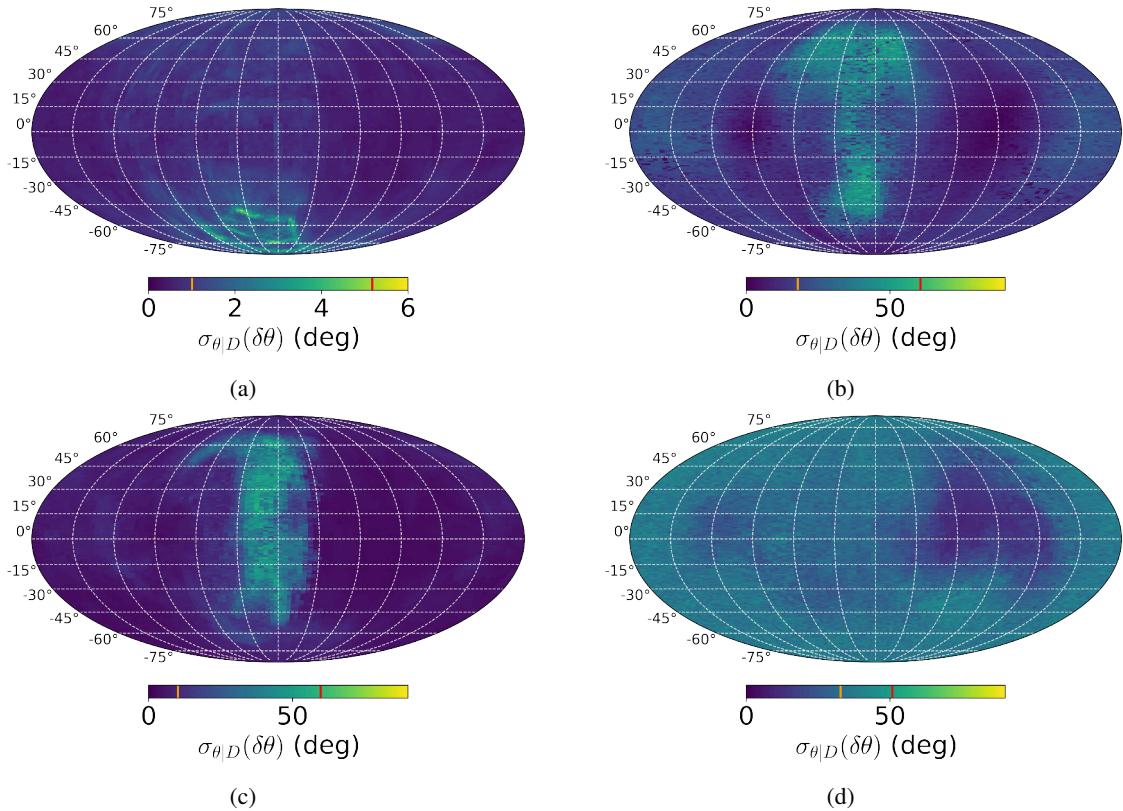


Fig. 5: As in Fig. 4, but the colorbar denotes the corresponding standard deviation.

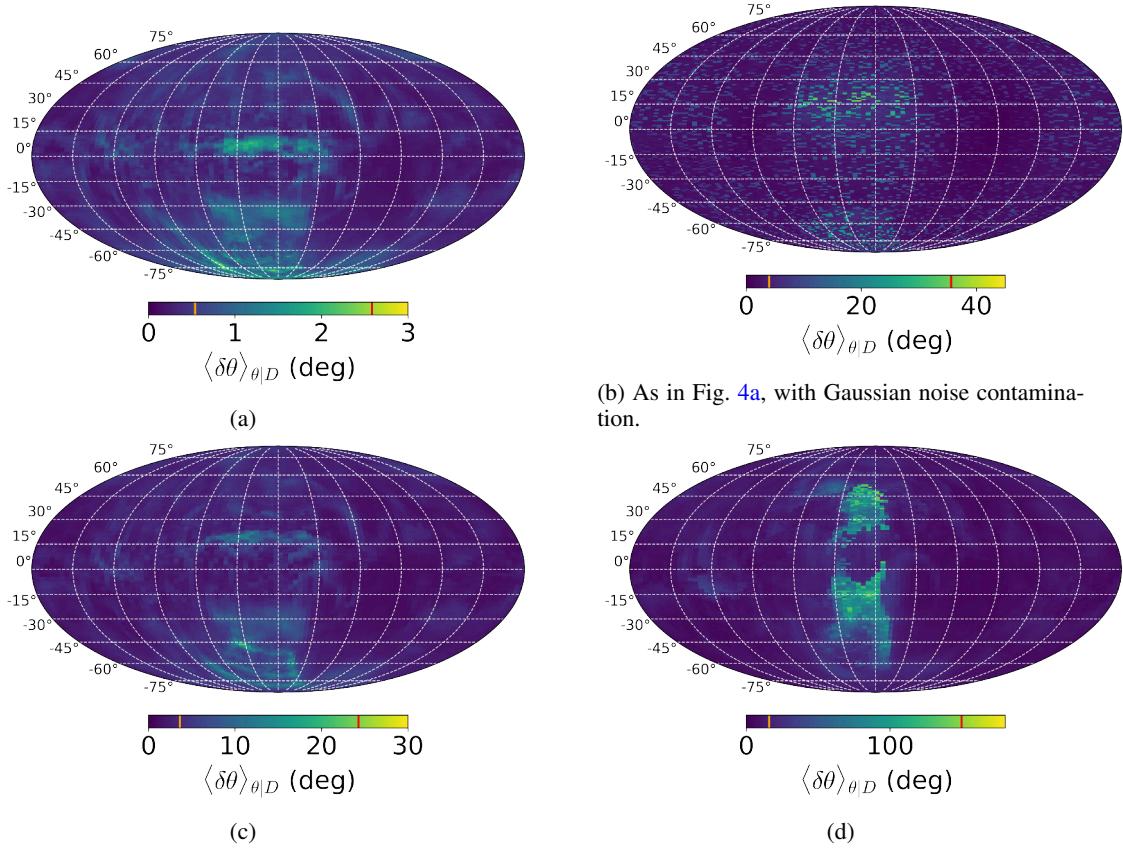


Fig. 6: As in Fig. 4, but the ground truth field that is rescaled so that its RMS magnitude is $1\mu\text{G}$.

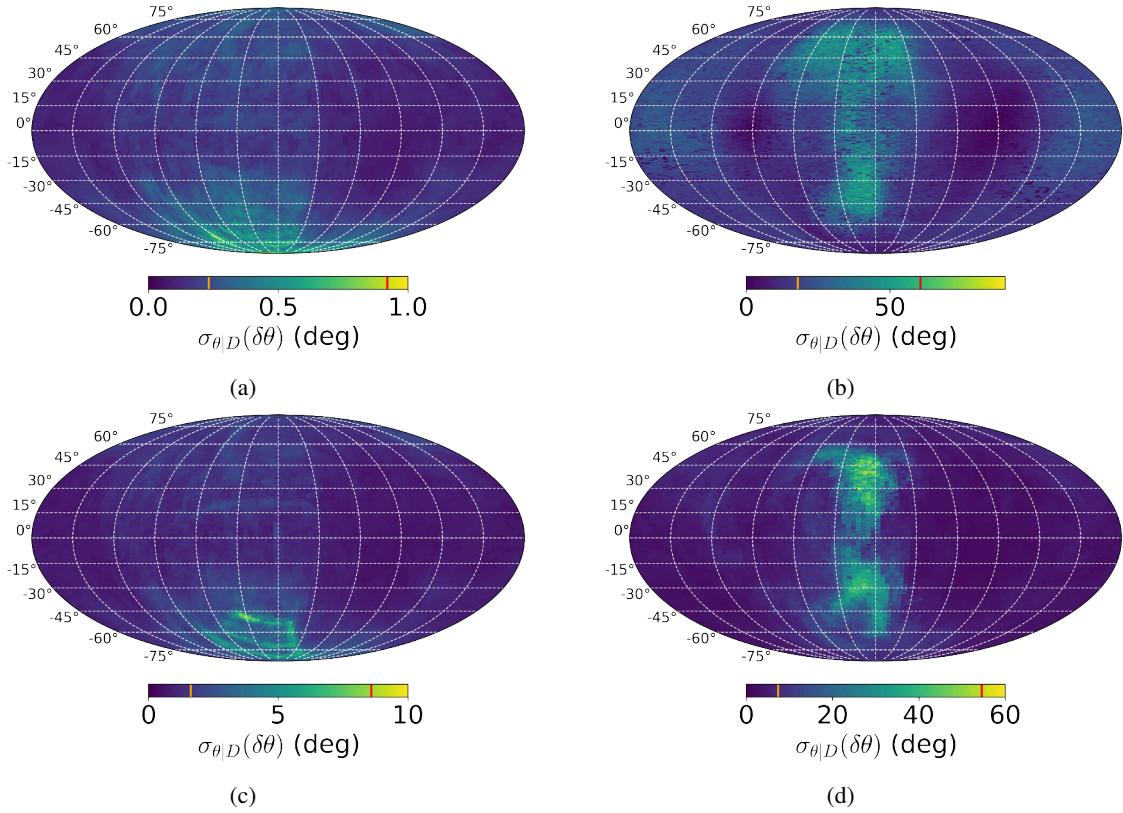


Fig. 7: As in Fig. 5, but with a ground truth field that has an RMS magnitude value of $1\mu\text{G}$.

ment n. 873089 (ASTROSTAT-II). This work also benefited greatly from discussions during the program "Towards a Comprehensive Model of the Galactic Magnetic Field" at Nordita in April 2023, which is partly supported by Nord-Forsk and the Royal Astronomical Society. A.T. would like to thank Vincent Pelgrims, Raphael Skalidis, Georgia V. Panopoulou, and Konstantinos Tassis for helpful tips and stimulating discussions. G.E. acknowledges the support of the German Academic Scholarship Foundation in the form of a PhD scholarship ("Promotionsstipendium der Studienstiftung des Deutschen Volkes"). P.F. acknowledges funding through the German Federal Ministry of Education and Research for the project ErUM-IFT: Informationsfeldtheorie fuer Experimente an Großforschungsanlagen (Foerderkennzeichen: 05D23EO1)

References

- Abbasi, R. U., Abe, M., Abu-Zayyad, T., et al. 2014, *ApJ*, 790, L21
- Abbasi, R. U., Abe, M., Abu-Zayyad, T., et al. 2019, *Phys. Rev. D*, 99, 022002
- Abbasi, R. U., Abe, Y., Abu-Zayyad, T., et al. 2024, *Phys. Rev. D*, 110, 022006
- Allard, D., Aublin, J., Baret, B., & Parizot, E. 2024, *AA*, 686, A292
- Andersson, B. G., Lazarian, A., & Vaillancourt, J. E. 2015, *ARA&A*, 53, 501
- Arras, P., Baltac, M., Ensslin, T. A., et al. 2019, *Astrophysics Source Code Library*
- Arsene, N. 2021, *Universe*, 7, 321
- Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., & Andrae, R. 2021, *AJ*, 161, 147
- Bendre, A. B. & Subramanian, K. 2022, *MNRAS*, 511, 4454
- Bister, T., Farrar, G. R., & Unger, M. 2024, The large-scale anisotropy and flux (de-)magnification of ultra-high-energy cosmic rays in the Galactic magnetic field
- Chandrasekhar, S. & Fermi, E. 1953, *ApJ*, 118, 113
- Clark, S. E. & Hensley, B. S. 2019, *ApJ*, 887, 136
- Davis, L. 1951, *Physical Review*, 81, 890
- Doi, Y., Hasegawa, T., Bastien, P., et al. 2021, *ApJ*, 914, 122
- Doi, Y., Nakamura, K., Kawabata, K. S., et al. 2024, *ApJ*, 961, 13
- Edenhofer, G., Frank, P., Roth, J., et al. 2024, *The Journal of Open Source Software*, 9, 6593
- Edenhofer, G., Zucker, C., Frank, P., et al. 2023, arXiv e-prints, arXiv:2308.01295
- Eichmann, B. & Winchen, T. 2020, *J. Cosmology Astropart. Phys.*, 2020, 047
- Enßlin, T. 2022, *Entropy*, 24, 374
- Enßlin, T. A. 2019, *Annalen der Physik*, 531, 1970017
- Enßlin, T. A., Frommert, M., & Kitaura, F. S. 2009, *Phys. Rev. D*, 80, 105005
- Fang, K. & Murase, K. 2018, *Nature Physics*, 14, 396
- Frank, P., Leike, R., & Enßlin, T. A. 2021, *Entropy*, 23
- Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, *A&A*, 649, A1
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, 595, A1
- Gaisser, T. K. & Hillas, A. M. 1977, in *International Cosmic Ray Conference*, Vol. 8, International Cosmic Ray Conference, 353
- Globus, N., Piran, T., Hoffman, Y., Carlesi, E., & Pomarède, D. 2019, *Monthly Notices of the Royal Astronomical Society*, 484, 4167
- Green, G. M., Schlaufly, E., Zucker, C., Speagle, J. S., & Finkbeiner, D. 2019, *ApJ*, 887, 93
- Greisen, K. 1966, *Phys. Rev. Lett.*, 16, 748
- Heinze, J., Fedynitch, A., Boncioli, D., & Winter, W. 2019, *The Astrophysical Journal*, 873, 88
- Higuchi, R., Sako, T., Fujii, T., Kawata, K., & Kido, E. 2023, *ApJ*, 949, 107
- Hutschenreuter, S., Havercorn, M., Frank, P., Raycheva, N. C., & Enßlin, T. A. 2023, arXiv e-prints, arXiv:2304.12350
- Hutschenreuter, S., Anderson, C. S., Betti, S., et al. 2022, *AA*, 657, A43
- Jansson, R. & Farrar, G. R. 2012a, *ApJ*, 757, 14
- Jansson, R. & Farrar, G. R. 2012b, *ApJ*, 761, L11
- Kawai, H., Yoshida, S., Yoshii, H., et al. 2008, *Nuclear Physics B - Proceedings Supplements*, 175-176, 221, proceedings of the XIV International Symposium on Very High Energy Cosmic Ray Interactions
- Kawata, K., di Matteo, A., Fujii, T., et al. 2019, in *International Cosmic Ray Conference*, Vol. 36, 36th International Cosmic Ray Conference (ICRC2019), 310
- Kobzar, O., Hnatyk, B., Marchenko, V., & Sushchov, O. 2019, *Monthly Notices of the Royal Astronomical Society*, 484, 1790
- Lallement, R., Babusiaux, C., Vergely, J. L., et al. 2019, *A&A*, 625, A135
- Lallement, R., Capitanio, L., Ruiz-Dern, L., et al. 2018, *A&A*, 616, A132
- Lallement, R., Vergely, J. L., Babusiaux, C., & Cox, N. L. J. 2022, *A&A*, 661, A147
- Leike, R. H., Edenhofer, G., Knollmüller, J., et al. 2022, arXiv e-prints, arXiv:2204.11715
- Leike, R. H. & Enßlin, T. A. 2019, *A&A*, 631, A32
- Leike, R. H., Glatzle, M., & Enßlin, T. A. 2020, *A&A*, 639, A138
- Linsley, J. 1963, *Phys. Rev. Lett.*, 10, 146
- Magalhães, A. M. 2012, in *Science from the Next Generation Imaging and Spectroscopic Surveys*, 7
- Magkos, G. & Pavlidou, V. 2019, *J. Cosmology Astropart. Phys.*, 2019, 004
- Maharana, S., Anche, R. M., Ramaprakash, A. N., et al. 2022, *Journal of Astronomical Telescopes, Instruments, and Systems*, 8, 038004
- Maharana, S., Kyriiotakis, J. A., Ramaprakash, A. N., et al. 2021, *Journal of Astronomical Telescopes, Instruments, and Systems*, 7, 014004
- Mayotte, E. W. et al. 2023, PoS, ICRC2023, 365
- Murase, K. 2019, in *International Cosmic Ray Conference*, Vol. 36, 36th International Cosmic Ray Conference (ICRC2019), 965
- Pandhi, A., Hutschenreuter, S., West, J. L., Gaensler, B. M., & Stock, A. 2022, *MNRAS*, 516, 4739
- Panopoulou, G. V. & Lenz, D. 2020, *ApJ*, 902, 120
- Panopoulou, G. V., Psaradaki, I., Skalidis, R., Tassis, K., & Andrews, J. J. 2017, *MNRAS*, 466, 2529
- Pelgrims, V., Mandarakas, N., Skalidis, R., et al. 2024, *AA*, 684, A162
- Pelgrims, V., Panopoulou, G. V., Tassis, K., et al. 2023, *AA*, 670, A164
- Pierre Auger Collaboration, Aab, A., Abreu, P., et al. 2017, *Science*, 357, 1266
- Prouza, M. & Šmilka, R. 2003, *AA*, 410, 1
- Romanopoulos, S., Pavlidou, V., & Tomaras, T. 2022a, arXiv e-prints, arXiv:2206.14837
- Romanopoulos, S., Pavlidou, V., & Tomaras, T. 2022b, in *37th International Cosmic Ray Conference*. 12-23 July 2021. Berlin, 475
- Sagawa, H. 2020, *Journal of Instrumentation*, 15, C09012
- Salamida, F. 2023, PoS, ICRC2023, 016
- Selig, M., Bell, M. R., Junklewitz, H., et al. 2013, *aap*, 554, A26
- Skalidis, R., Sternberg, J., Beattie, J. R., Pavlidou, V., & Tassis, K. 2021, *A&A*, 656, A118
- Skalidis, R. & Tassis, K. 2021, *A&A*, 647, A186
- Stanev, T. 1997, *The Astrophysical Journal*, 479, 290
- Steininger, T., Dixit, J., Frank, P., et al. 2017, ArXiv e-prints [arXiv:1708.01073]
- Sun, X.-H. & Reich, W. 2010, *Research in Astronomy and Astrophysics*, 10, 1287
- Sun, X. H., Reich, W., Waelkens, A., & Enßlin, T. A. 2008, *A&A*, 477, 573
- Tahani, M., Glover, J., Lupciu, W., et al. 2022a, *A&A*, 660, L7
- Tahani, M., Lupciu, W., Glover, J., et al. 2022b, *A&A*, 660, A97
- Takami, H. & Sato, K. 2010, *ApJ*, 724, 1456
- Tassis, K., Ramaprakash, A. N., Readhead, A. C. S., et al. 2018, arXiv e-prints, arXiv:1810.05652
- Tritsis, A., Federrath, C., & Pavlidou, V. 2019, *ApJ*, 873, 38
- Tritsis, A., Federrath, C., Schneider, N., & Tassis, K. 2018, *MNRAS*, 481, 5275
- Tsouros, A., Bendre, A. B., Edenhofer, G., et al. 2024a, arXiv e-prints, arXiv:2403.05531
- Tsouros, A., Edenhofer, G., Enßlin, T., Mastorakis, M., & Pavlidou, V. 2024b, *A&A*, 681, A111
- Watanabe, Keito, Fedynitch, Anatoli, Capel, Francesca, & Sagawa, Hiroyuki. 2023, *EPJ Web Conf.*, 283, 03009
- Zatsepin, G. T. & Kuzmin, V. A. 1966, *JETP Lett.*, 4, 78

Chapter 7

Conclusions

In this thesis, we developed a robust method for reconstructing the GMF with the intent of correcting for its effect on the observed arrival directions of UHECRs, thereby improving our knowledge regarding their sources. Using IFT as the foundation, we were able to efficiently sample the posterior distribution of the GMF using synthetic data, comprising of both sparse and local GMF measurements in anticipation of PASIPHAE optopolarimetric data, as well as integral data, modelling Faraday measurements of the GMF.

The first part of this thesis introduced a systematic IFT-based framework for inferring the true arrival directions of UHECRs by reconstructing the GMF from sparse, local data scattered within a specified region of the Galaxy. We demonstrated that in cases where the GMF is weakly turbulent and data are uniformly sampled with an average signal-to-noise ratio of two, we could correct for the GMF's influence on UHECR arrival directions with an angular accuracy of approximately 3° . This result holds for an average distance between data points of up to 600 pc, which is suitable for real applications where data points will align with molecular cloud locations. Our approach naturally provides posterior samples of UHECR arrival directions without requiring strict assumptions on field turbulence or data density.

Building on this framework, we next addressed more realistic conditions, where the local GMF data derived from stellar polarisation tomography lack line-of-sight (LOS) information and are distributed according to a realistic, non-uniform distribution of local and sparse data points that coincide with the location of dust clouds. Here, we benchmarked our algorithm's performance using a GMF configuration from a dynamo magnetohydrodynamic simulation, enabling us to evaluate the performance of the Gaussian basis of our method when applied to a ground truth that more closely resembles the non-Gaussian statistics of the real GMF. Our findings indicated that while the absence of LOS information in local data reduces accuracy, combining these measurements with LOS-integrated data, such as Faraday measurements, allows us to substantially correct for the effect of the GMF on the observed arrival directions of UHECRs. Even in challenging directions where the effect of the GMF is maximal, our approach provided a threefold improvement over simpler models, reconstructing the large-scale GMF features as well as the power spectrum.

In the final part of this thesis, we examined the case where directional information on the local and sparse measurements of the GMF is lacking, as will be the case with real data. To address this, we assumed that the direction of localised measurements aligns with that of the ordered component of well-known parametric models, like the JF12 model. We demonstrated that such prior knowledge can compensate for the missing data, producing corrections with comparable accuracy to those assuming full directional knowledge of the GMF local and sparse data, always within the context of UHECR arrival direction inference. For lower charge states and field strengths around $5 \mu\text{G}$, arrival directions

could be inferred within a few degrees. Even for moderately charged particles, directional biases could be limited to 10° – 20° , and we verified that the greatest deviations occur in regions perpendicular to the GMF’s mean component. The directions on the POS that are perpendicular to the inferred large scale component of the GMF can be inferred, indicating an area of the sky where our algorithm is expected to have its worst performance. In addition, PASIPHAE aims to provide polarimetric data at high Galactic latitudes, where UHECR hotspots are commonly observed, making stellar polarimetric data from PASIPHAE, in conjunction with our inference method, well-suited for the application of UHECR source localisation up to the effect of the GMF.

Overall, this thesis provides a rigorous statistical framework capable of leveraging both local and integral data to provide a nonparametric reconstruction of the GMF within the local Milky Way, up to the distance where local data can be achieved, approximately within 1.5 kpc from Earth, as localisation depends on Gaia data. The ability to reconstruct the GMF nonparametrically from such sparse and local data, which will become available via the PASIPHAE survey, will mark a significant advancement over simpler, parametric models currently in use by the community.

Bibliography

- Aartsen, M. G. et al. (Nov. 2018). "Astrophysical neutrinos and cosmic rays observed by IceCube". In: *Advances in Space Research* 62.10, pp. 2902–2930. DOI: 10.1016/j.asr.2017.05.030. arXiv: 1701.03731 [astro-ph.HE].
- Abbasi, R. et al. (Feb. 2008). "Air fluorescence measurements in the spectral range 300–420 nm using a 28.5 GeV electron beam". In: *Astroparticle Physics* 29.1, pp. 77–86. DOI: 10.1016/j.astropartphys.2007.11.010. arXiv: 0708.3116 [astro-ph].
- Abbasi, R. U. et al. (2019). "Mass composition of ultrahigh-energy cosmic rays with the Telescope Array Surface Detector data". In: *Phys. Rev. D* 99 (2), p. 022002. DOI: 10.1103/PhysRevD.99.022002. URL: <https://link.aps.org/doi/10.1103/PhysRevD.99.022002>.
- Abraham, J. et al. (Aug. 2010). "The fluorescence detector of the Pierre Auger Observatory". In: *Nuclear Instruments and Methods in Physics Research A* 620.2-3, pp. 227–251. DOI: 10.1016/j.nima.2010.04.023.
- Ahlers, Markus and Francis Halzen (Sept. 2018). "Opening a new window onto the universe with IceCube". In: *Progress in Particle and Nuclear Physics* 102, pp. 73–88. DOI: 10.1016/j.ppnp.2018.05.001. arXiv: 1805.11112 [astro-ph.HE].
- Allard, D. and R. J. Protheroe (Aug. 2009). "Interactions of UHE cosmic ray nuclei with radiation during acceleration: consequences for the spectrum and composition". In: *Astronomy & Astrophysics* 502.3, pp. 803–815. DOI: 10.1051/0004-6361/200911839. arXiv: 0902.4538 [astro-ph.HE].
- Aloisio, R., V. Berezinsky, and A. Gazizov (Dec. 2012). "Transition from galactic to extragalactic cosmic rays". In: *Astroparticle Physics* 39, pp. 129–143. DOI: 10.1016/j.astropartphys.2012.09.007. arXiv: 1211.0494 [astro-ph.HE].
- Alves Batista, Rafael et al. (June 2019). "Open Questions in Cosmic-Ray Research at Ultrahigh Energies". In: *Frontiers in Astronomy and Space Sciences* 6, 23, p. 23. DOI: 10.3389/fspas.2019.00023. arXiv: 1903.06714 [astro-ph.HE].
- Anchordoqui, Luis Alfredo (2011). *Ultrahigh Energy Cosmic Rays: Facts, Myths, and Legends*. arXiv: 1104.0509 [hep-ph]. URL: <https://arxiv.org/abs/1104.0509>.
- Andersson, B. G., A. Lazarian, and John E. Vaillancourt (Aug. 2015). "Interstellar Dust Grain Alignment". In: *Annual Review of Astronomy and Astrophysics* 53, pp. 501–539. DOI: 10.1146/annurev-astro-082214-122414.
- Arámburo-García, Andrés et al. (2021). "Ultrahigh energy cosmic ray deflection by the intergalactic magnetic field". In: *Phys. Rev. D* 104 (8), p. 083017. DOI: 10.1103/PhysRevD.104.083017. URL: <https://link.aps.org/doi/10.1103/PhysRevD.104.083017>.
- Arsene, Nicusor (Aug. 2021). "Mass Composition of UHECRs from X_{max} Distributions Recorded by the Pierre Auger and Telescope Array Observatories". In: *Universe* 7.9, 321, p. 321. DOI: 10.3390/universe7090321. arXiv: 2109.03626 [hep-ph].
- Bartos, Imre and Marek Kowalski (2017). *Multimessenger Astronomy*. DOI: 10.1088/978-0-7503-1369-8.
- Beck, Rainer (Oct. 2001). "Galactic and Extragalactic Magnetic Fields". In: *Space Science Reviews* 99, pp. 243–260. DOI: 10.1023/A:1013805401252. arXiv: astro-ph/0012402 [astro-ph].

- Beck, Rainer and Richard Wielebinski (2013). "Magnetic Fields in Galaxies". In: *Planets, Stars and Stellar Systems. Volume 5: Galactic Structure and Stellar Populations*. Ed. by Terry D. Oswalt and Gerard Gilmore. Vol. 5, p. 641. DOI: 10.1007/978-94-007-5612-0_13.
- Berezinsky, V. (Jan. 2014a). "Extragalactic cosmic rays and their signatures". In: *Astroparticle Physics* 53, pp. 120–129. DOI: 10.1016/j.astropartphys.2013.04.001. arXiv: 1301.0914 [astro-ph.HE].
- (Jan. 2014b). "Extragalactic cosmic rays and their signatures". In: *Astroparticle Physics* 53, pp. 120–129. DOI: 10.1016/j.astropartphys.2013.04.001. arXiv: 1301.0914 [astro-ph.HE].
- Berezinsky, V. S. and S. I. Grigoreva (Jan. 1988). "Spectrum of ultrahigh-energy cosmic rays". In: *Pisma v Astronomicheskii Zhurnal* 14, pp. 3–17.
- Berezinsky, Veniamin, Askhat Gazizov, and Svetlana Grigorieva (Aug. 2006). "On astrophysical solution to ultrahigh energy cosmic rays". In: *Physical Review D* 74.4, 043005, p. 043005. DOI: 10.1103/PhysRevD.74.043005. arXiv: hep-ph/0204357 [hep-ph].
- Bertone, Gianfranco et al. (Apr. 2006). "Gamma-ray bursts and the origin of galactic positrons". In: *Physics Letters B* 636.1, pp. 20–24. DOI: 10.1016/j.physletb.2006.03.022. arXiv: astro-ph/0405005 [astro-ph].
- Bird, D. J. et al. (Mar. 1995). "Detection of a Cosmic Ray with Measured Energy Well beyond the Expected Spectral Cutoff due to Cosmic Microwave Radiation". In: *The Astrophysical Journal* 441, p. 144. DOI: 10.1086/175344. arXiv: astro-ph/9410067 [astro-ph].
- Blasi, Pasquale and Sergio Colafrancesco (Nov. 1999). "Cosmic rays, radio halos and non-thermal X-ray emission in clusters of galaxies". In: *Astroparticle Physics* 12.3, pp. 169–183. DOI: 10.1016/S0927-6505(99)00079-1. arXiv: astro-ph/9905122 [astro-ph].
- Braun, R. et al. (Apr. 2015). "Advancing Astrophysics with the Square Kilometre Array". In: *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, 174, p. 174. DOI: 10.22323/1.215.0174.
- Brown, J. C. et al. (July 2007). "Rotation Measures of Extragalactic Sources behind the Southern Galactic Plane: New Insights into the Large-Scale Magnetic Field of the Inner Milky Way". In: *The Astrophysical Journal* 663.1, pp. 258–266. DOI: 10.1086/518499. arXiv: 0704.0458 [astro-ph].
- Chandrasekhar, S. and E. Fermi (July 1953). "Magnetic Fields in Spiral Arms." In: *The Astrophysical Journal* 118, p. 113. DOI: 10.1086/145731.
- Coleman, A. et al. (2023). "Ultra high energy cosmic rays The intersection of the Cosmic and Energy Frontiers". In: *Astroparticle Physics* 149, p. 102819. ISSN: 0927-6505. DOI: <https://doi.org/10.1016/j.astropartphys.2023.102819>. URL: <https://www.sciencedirect.com/science/article/pii/S0927650523000051>.
- Collaboration, Gaia (2016). "The Gaia mission". In: *A&A* 595, A1. DOI: 10.1051/0004-6361/201629272. URL: <https://doi.org/10.1051/0004-6361/201629272>.
- Cox, R. T. (Jan. 1946). "Probability, Frequency and Reasonable Expectation". In: *American Journal of Physics* 14.1, pp. 1–13. DOI: 10.1119/1.1990764.
- Cronin, James W. (Jan. 2005). "The highest-energy cosmic rays". In: *Nuclear Physics B Proceedings Supplements* 138, pp. 465–491. DOI: 10.1016/j.nuclphysbps.2004.11.107. arXiv: astro-ph/0402487 [astro-ph].
- Deligny, O. (July 2019). "The energy spectrum of ultra-high energy cosmic rays measured at the Pierre Auger Observatory and at the Telescope Array". In: *36th International Cosmic Ray Conference (ICRC2019)*. Vol. 36. International Cosmic Ray Conference, 234, p. 234. DOI: 10.22323/1.358.0234.

- Dolag, K. et al. (Oct. 2005). "The imprints of local superclusters on the Sunyaev-Zel'dovich signals and their detectability with Planck". In: 363.1, pp. 29–39. DOI: 10.1111/j.1365-2966.2005.09452.x. arXiv: astro-ph/0505258 [astro-ph].
- Enßlin, Torsten A. (2019). "Information Theory for Fields". In: *Annalen der Physik* 531.3, p. 1800127. DOI: <https://doi.org/10.1002/andp.201800127>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.201800127>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/andp.201800127>.
- Gaisser, T. K. and A. M. Hillas (Jan. 1977). "Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers". In: *International Cosmic Ray Conference*. Vol. 8. International Cosmic Ray Conference, p. 353.
- Greisen, Kenneth (Apr. 1966). "End to the Cosmic-Ray Spectrum?" In: 16.17, pp. 748–750. DOI: 10.1103/PhysRevLett.16.748.
- Guido, Eleonora (2021). "Constraining ultra-high-energy cosmic ray source scenarios with the data measured by the Pierre Auger Observatory". Defended on February 12, 2021. Committee: Prof. Fiorenza Donato, Prof. Sergio Petrera, Dr. Andrew Taylor. Tutor: Dr. Antonella Castellina, Co-tutor: Prof. Mario Edoardo Bertaina. PhD thesis. Italy: University of Torino.
- Hackstein, Stefan et al. (Feb. 2019). "Propagation of UHECRs in the local Universe and origin of cosmic magnetic fields". In: *arXiv e-prints*, arXiv:1902.04408, arXiv:1902.04408. DOI: 10.48550/arXiv.1902.04408. arXiv: 1902.04408 [astro-ph.HE].
- Heald, George et al. (July 2020). "Magnetism Science with the Square Kilometre Array". In: *Galaxies* 8.3, 53, p. 53. DOI: 10.3390/galaxies8030053. arXiv: 2006.03172 [astro-ph.GA].
- Hess, Victor F. (1912). "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten". In: *Phys. Z.* 13, pp. 1084–1091.
- Hillas, A. M. (Jan. 1984). "The Origin of Ultra-High-Energy Cosmic Rays". In: *Annual Review of Astronomy and Astrophysics* 22, pp. 425–444. DOI: 10.1146/annurev.aa.22.090184.002233.
- Inoue, N., K. Miyazawa, and Y. Kawasaki (2009). "JEM-EUSO mission on the ISS with the aim of performing particle astronomy". In: *Nuclear Physics B - Proceedings Supplements* 196. Proceedings of the XV International Symposium on Very High Energy Cosmic Ray Interactions (ISVHECRI 2008), pp. 135–140. ISSN: 0920-5632. DOI: <https://doi.org/10.1016/j.nuclphysbps.2009.09.024>. URL: <https://www.sciencedirect.com/science/article/pii/S0920563209006604>.
- Jaffe, T. R. et al. (May 2013). "Comparing polarized synchrotron and thermal dust emission in the Galactic plane". In: 431.1, pp. 683–694. DOI: 10.1093/mnras/stt200. arXiv: 1302.0143 [astro-ph.GA].
- Jaffe, Tess R. (Apr. 2019). "Practical Modeling of Large-Scale Galactic Magnetic Fields: Status and Prospects". In: *Galaxies* 7.2, 52, p. 52. DOI: 10.3390/galaxies7020052. arXiv: 1904.12689 [astro-ph.GA].
- Jansson, Ronnie and Glennys R. Farrar (Sept. 2012). "A New Model of the Galactic Magnetic Field". In: *The Astrophysical Journal* 757.1, 14, p. 14. DOI: 10.1088/0004-637X/757/1/14. arXiv: 1204.3662 [astro-ph.GA].
- Jarvis, M. et al. (Jan. 2016). "The MeerKAT International GHz Tiered Extragalactic Exploration (MIGHTEE) Survey". In: *MeerKAT Science: On the Pathway to the SKA*, 6, p. 6. DOI: 10.22323/1.277.0006. arXiv: 1709.01901 [astro-ph.GA].
- Jaynes, E. T. (2003). *Probability Theory: The Logic of Science*. Ed. by G. Larry Bretthorst. Cambridge University Press.
- Kampert, Karl-Heinz and Michael Unger (May 2012). "Measurements of the cosmic ray composition with air shower experiments". In: *Astroparticle Physics* 35.10, pp. 660–678. DOI: 10.1016/j.astropartphys.2012.02.004. arXiv: 1201.0018 [astro-ph.HE].

- Kawai, H. et al. (Jan. 2008a). "Telescope Array Experiment". In: *Nuclear Physics B Proceedings Supplements* 175, pp. 221–226. DOI: 10.1016/j.nuclphysbps.2007.11.002.
- (Jan. 2008b). "Telescope Array Experiment". In: *Nuclear Physics B Proceedings Supplements* 175, pp. 221–226. DOI: 10.1016/j.nuclphysbps.2007.11.002.
- Kotera, Kumiko and Martin Lemoine (Jan. 2008). "Inhomogeneous extragalactic magnetic fields and the second knee in the cosmic ray spectrum". In: 77.2, 023005, p. 023005. DOI: 10.1103/PhysRevD.77.023005. arXiv: 0706.1891 [astro-ph].
- Kotera, Kumiko and Angela V. Olinto (2011). "The Astrophysics of Ultrahigh-Energy Cosmic Rays". In: *Annual Review of Astronomy and Astrophysics* 49. Volume 49, 2011, pp. 119–153. ISSN: 1545-4282. DOI: <https://doi.org/10.1146/annurev-astro-081710-102620>. URL: <https://www.annualreviews.org/content/journals/10.1146/annurev-astro-081710-102620>.
- Kronberg, Philipp P., Harald Lesch, and Ulrich Hopp (Jan. 1999). "Magnetization of the Intergalactic Medium by Primeval Galaxies". In: *The Astrophysical Journal* 511.1, pp. 56–64. DOI: 10.1086/306662.
- Kulikov, G. V. and G. B. Khristiansen (Mar. 1959). "On the Size Spectrum of Extensive Air Showers". In: *Soviet Physics JETP* 8, p. 4.
- Kullback, S. and R. A. Leibler (1951). "On Information and Sufficiency". In: *Ann. Math. Statist.* 22.1, pp. 79–86.
- Linsley, John (Feb. 1963). "Evidence for a Primary Cosmic-Ray Particle with Energy 10^{20} eV". In: 10.4, pp. 146–148. DOI: 10.1103/PhysRevLett.10.146.
- Magkos, G. and V. Pavlidou (Feb. 2019). "Deflections of ultra-high energy cosmic rays by the Milky Way magnetic field: how well can they be corrected?" In: 2019.2, 004, p. 004. DOI: 10.1088/1475-7516/2019/02/004. arXiv: 1802.03409 [astro-ph.HE].
- Mayotte, Eric William et al. (2023). "Measurement of the mass composition of ultra-high-energy cosmic rays at the Pierre Auger Observatory". In: *PoS ICRC2023*, p. 365. DOI: 10.22323/1.444.0365.
- Murase, Kohta et al. (Apr. 2012). "Blazars as Ultra-high-energy Cosmic-ray Sources: Implications for TeV Gamma-Ray Observations". In: *The Astrophysical Journal* 749.1, 63, p. 63. DOI: 10.1088/0004-637X/749/1/63. arXiv: 1107.5576 [astro-ph.HE].
- Nagano, M. and A. A. Watson (July 2000). "Observations and implications of the ultrahigh-energy cosmic rays". In: *Reviews of Modern Physics* 72.3, pp. 689–732. DOI: 10.1103/RevModPhys.72.689.
- Neronov, A. Yu, D. V. Semikoz, and I. I. Tkachev (June 2009). "Ultra-high energy cosmic ray production in the polar cap regions of black hole magnetospheres". In: *New Journal of Physics* 11.6, 065015, p. 065015. DOI: 10.1088/1367-2630/11/6/065015. arXiv: 0712.1737 [astro-ph].
- Olinto, Angela V. (July 2012). "Cosmic rays at the highest energies". In: *Journal of Physics Conference Series*. Vol. 375. *Journal of Physics Conference Series*. IOP, 052001, p. 052001. DOI: 10.1088/1742-6596/375/1/052001. arXiv: 1201.4519 [astro-ph.HE].
- Orlando, Elena and Andrew Strong (Dec. 2013). "Galactic synchrotron emission with cosmic ray propagation models". In: 436.3, pp. 2127–2142. DOI: 10.1093/mnras/stt1718. arXiv: 1309.2947 [astro-ph.GA].
- Pelgrims, V. et al. (Feb. 2023). "Starlight-polarization-based tomography of the magnetized ISM: PASIPHAE's line-of-sight inversion method". In: *Astronomy Astrophysics* 670, A164, A164. DOI: 10.1051/0004-6361/202244625. arXiv: 2208.02278 [astro-ph.GA].
- Romanopoulos, S., V. Pavlidou, and T. Tomaras (June 2022a). "New physics above 50 TeV: probing its phenomenology through UHECR air-shower simulations". In: *arXiv e-prints*, arXiv:2206.14837, arXiv:2206.14837. arXiv: 2206.14837 [astro-ph.HE].
- (Mar. 2022b). "New physics Air-Shower simulations for UHECR above 50 TeV". In: *37th International Cosmic Ray Conference. 12-23 July 2021. Berlin*, 475, p. 475.

- Ryu, Dongsu, Hyesung Kang, and Peter L. Biermann (July 1998). "Cosmic magnetic fields in large scale filaments and sheets". In: 335, pp. 19–25. DOI: 10.48550/arXiv.astro-ph/9803275. arXiv: astro-ph/9803275 [astro-ph].
- Salamida, Francesco (2023). "Highlights from the Pierre Auger Observatory". In: PoS ICRC2023, p. 016. DOI: 10.22323/1.444.0016. arXiv: 2312.14673 [astro-ph.HE].
- Sigl, Günter, Francesco Miniati, and Torsten A. Enßlin (Aug. 2004). "Ultrahigh energy cosmic ray probes of large scale structure and magnetic fields". In: 70.4, 043007, p. 043007. DOI: 10.1103/PhysRevD.70.043007. arXiv: astro-ph/0401084 [astro-ph].
- Skalidis, R. et al. (Dec. 2021). "Why take the square root? An assessment of interstellar magnetic field strength estimation methods". In: *Astronomy Astrophysics* 656, A118, A118. DOI: 10.1051/0004-6361/202142045. arXiv: 2109.10925 [astro-ph.GA].
- Stanev, Todor (Apr. 1997). "Ultra-high-energy Cosmic Rays and the Large-scale Structure of the Galactic Magnetic Field". In: *Astrophysical Journal* 479.1, pp. 290–295. DOI: 10.1086/303866. arXiv: astro-ph/9607086 [astro-ph].
- Sun, Xiao-Hui and Wolfgang Reich (Dec. 2010). "The Galactic halo magnetic field revisited". In: *Research in Astronomy and Astrophysics* 10.12, pp. 1287–1297. DOI: 10.1088/1674-4527/10/12/009. arXiv: 1010.4394 [astro-ph.GA].
- Takami, Hajime and Katsuhiko Sato (July 2008). "Distortion of Ultra-High-Energy Sky by Galactic Magnetic Field". In: *The Astrophysical Journal* 681.2, pp. 1279–1286. DOI: 10.1086/588513. arXiv: 0711.2386 [astro-ph].
- Takami, Hajime, Hiroyuki Yoshiguchi, and Katsuhiko Sato (Mar. 2006). "Propagation of Ultra-High-Energy Cosmic Rays above 10^{19} eV in a Structured Extragalactic Magnetic Field and Galactic Magnetic Field". In: *The Astrophysical Journal* 639.2, pp. 803–815. DOI: 10.1086/499420. arXiv: astro-ph/0506203 [astro-ph].
- Tassis, Konstantinos et al. (Oct. 2018). "PASIPHAE: A high-Galactic-latitude, high-accuracy optopolarimetric survey". In: *arXiv e-prints*, arXiv:1810.05652, arXiv:1810.05652. arXiv: 1810.05652 [astro-ph.IM].
- The IceCube, Pierre Auger and Telescope Array collaborations (2016). "Search for correlations between the arrival directions of IceCube neutrino events and ultrahigh-energy cosmic rays detected by the Pierre Auger Observatory and the Telescope Array". In: *Journal of Cosmology and Astroparticle Physics* 2016.01, p. 037. DOI: 10.1088/1475-7516/2016/01/037. URL: <https://dx.doi.org/10.1088/1475-7516/2016/01/037>.
- The Pierre Auger Collaboration (2015). "The Pierre Auger Cosmic Ray Observatory". In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 798, pp. 172–213. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2015.06.058>. URL: <https://www.sciencedirect.com/science/article/pii/S0168900215008086>.
- Tokuno, H. et al. (June 2012). "New air fluorescence detectors employed in the Telescope Array experiment". In: *Nuclear Instruments and Methods in Physics Research A* 676, pp. 54–65. DOI: 10.1016/j.nima.2012.02.044. arXiv: 1201.0002 [astro-ph.IM].
- Ulrich, Ralf, Ralph Engel, and Michael Unger (Mar. 2011). "Hadronic multiparticle production at ultrahigh energies and extensive air showers". In: 83.5, 054026, p. 054026. DOI: 10.1103/PhysRevD.83.054026. arXiv: 1010.4310 [hep-ph].
- Waxman, Eli (Jan. 2005). "Extra Galactic Sources of High Energy Neutrinos*". In: *Physica Scripta Volume T* 121, pp. 147–152. DOI: 10.1088/0031-8949/2005/T121/022. arXiv: astro-ph/0502159 [astro-ph].
- Widrow, Lawrence M. (Jan. 2002). "Origin of galactic and extragalactic magnetic fields". In: *Reviews of Modern Physics* 74.3, pp. 775–823. DOI: 10.1103/RevModPhys.74.775. arXiv: astro-ph/0207240 [astro-ph].
- Zatsepin, G. T. and V. A. Kuz'min (Aug. 1966). "Upper Limit of the Spectrum of Cosmic Rays". In: *Soviet Journal of Experimental and Theoretical Physics Letters* 4, p. 78.