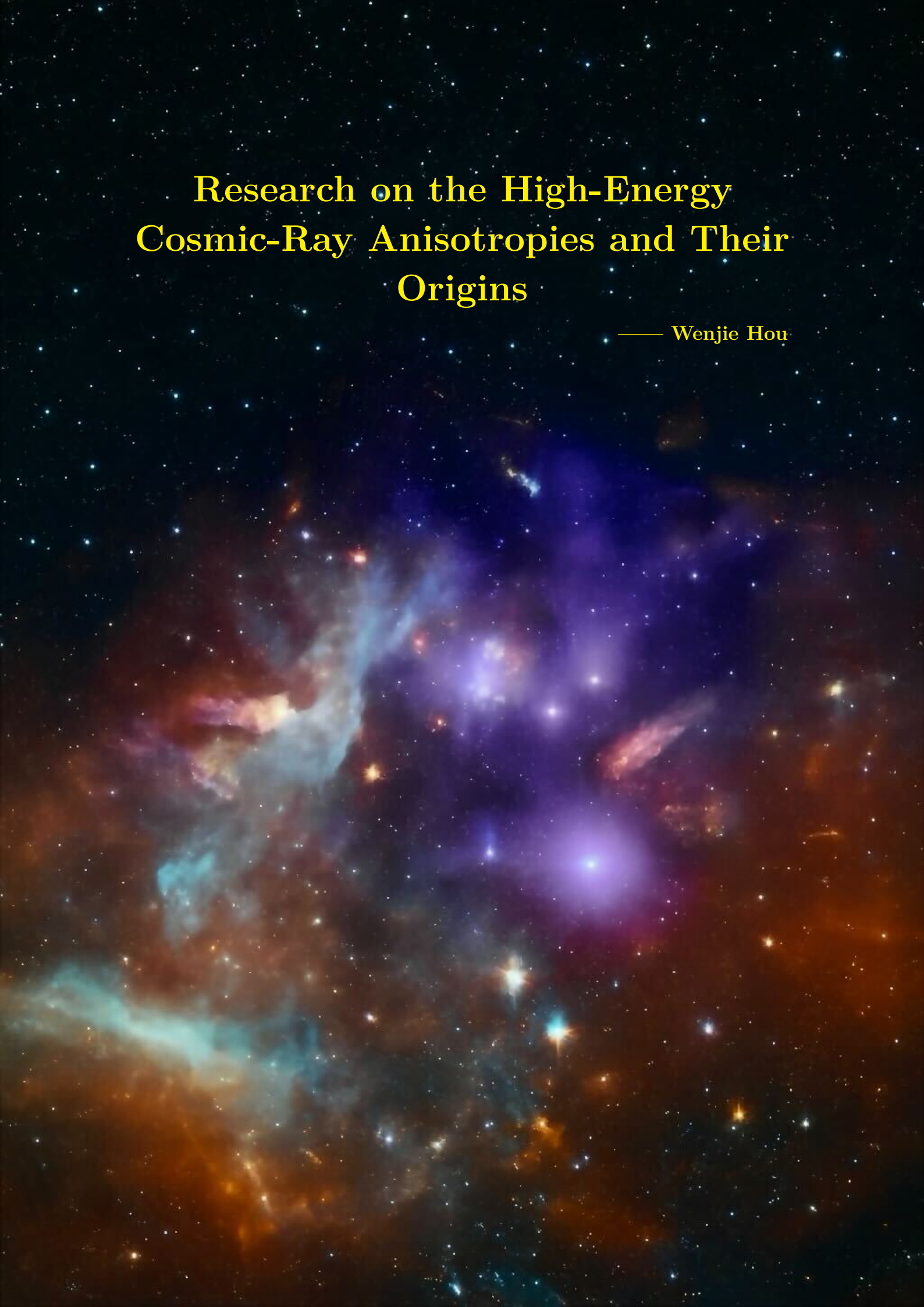


Research on the High-Energy Cosmic-Ray Anisotropies and Their Origins

—— Wenjie Hou

Photo credit: NASA

Research on the High-Energy Cosmic-Ray Anisotropies and Their Origins

For the attainment of the academic degree of
Doctor of Natural Science
from the Faculty of Physics of the
Karlsruhe Institute of Technology (KIT)

accepted
Dissertation
of

Wenjie Hou (M.Sc.)
from Jiangsu, China

Karlsruhe
December 2025

First Reviewer: Prof. Dr. Ralph Engel
Second Reviewer: Prof. Dr. Thomas Schwetz-Mangold
Supervisor: Dr. Andreas Haungs
Day of the oral examination: 19. 12. 2025

Research on the High-Energy Cosmic-Ray Anisotropies and Their Origins

Karlsruhe Institute of Technology

Zur Erlangung des akademischen Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)
von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

genehmigte
Dissertation
von

Wenjie Hou (M.Sc.)
aus Jiangsu, China

Karlsruhe
Dezember 2025

Referent:	Prof. Dr. Ralph Engel
Korreferent:	Prof. Dr. Thomas Schwetz-Mangold
Betreuer:	Dr. Andreas Haungs

Tag der mündlichen Prüfung: 19. 12. 2025

Abstract

Research on the High-Energy Cosmic-Ray Anisotropies and Their Origins

In recent decades, cosmic-ray physics has encountered several unresolved questions, particularly about the origin and acceleration mechanisms of high-energy cosmic rays. Despite improvements in experimental methods, the exact sources of cosmic rays with energies above the PeV scale remain unclear. The transition between Galactic and extragalactic cosmic rays, especially around the "knee" and "ankle" regions of the energy spectrum, continues to be debated. Additionally, the influence of Galactic magnetic fields on cosmic-ray propagation and anisotropies is not fully understood. Solving these issues is important for understanding the basic processes behind cosmic-ray acceleration and propagation throughout the universe.

Cosmic-ray anisotropy is another important aspect of research for understanding cosmic-ray origins and their movement through space. Anisotropies, observed on both small and large scales across different energy ranges, provide useful information about cosmic-ray sources and the magnetic fields they travel through. At lower energies (below a few TeV), large-scale anisotropies are likely affected by the distribution of nearby sources and the structure of the local interstellar magnetic field. However, at higher energies, particularly near the "knee", the anisotropy becomes more noticeable and harder to explain, suggesting more complex transport processes and possible nearby sources of Galactic cosmic rays. Beyond the ankle, anisotropy decreases, hinting at a possible extragalactic origin for cosmic rays in this range, though this transition is still under study. Additionally, small-scale anisotropies observed by experiments such as Tibet ASγ and IceCube may help identify localized cosmic-ray sources and shed light on the role of magnetic field turbulence in their propagation.

In this thesis, the sensitivity of the IceCube-Gen2 surface array to cosmic-ray anisotropy reconstruction is investigated, and it is demonstrated that IceCube-Gen2 will contribute important additional data points to the overall picture. In parallel, I then optimize analytical methods for reconstructing cosmic-ray anisotropies and identifying significant large- and medium-scale anisotropies above 10 PeV in the Milky Way. The sensitivity of the data to anisotropy reconstruction is evaluated using traditional methods, after which several optimized reconstruction techniques are introduced to overcome statistical limitations and reduce uncertainties in the recovered dipole parameters. A unified picture of cosmic-ray anisotropy is obtained by combining measurements from major experiments spanning the TeV domain to beyond EeV energies, ultimately yielding a continuous and global view of the evolution of dipole anisotropies from Galactic to extragalactic origins, and the global fit of the dipole amplitude exhibits a characteristic "W"–shaped structure. However, limited statistics leave the dipole phase between 10^{15} eV and 10^{19} eV uncertain. In particular, the 1–100 PeV range shows complex phase variations, especially in the KASCADE-Grande data. The optimized methods are then applied to the KASCADE-Grande data, revealing evidence of a 3σ anisotropy at 33 PeV and suggesting Cygnus OB2 as a potential source region. Propagation studies from Cygnus OB2 to Earth likewise show consistency with the observed anisotropy at 33 PeV. In the end, this thesis suggests that pulsar wind nebulae (PWNe) associated with supernova remnants (SNRs) may act as natural accelerators of Galactic

cosmic rays in the PeV energy range, and thereby potentially provide an explanation for the origin of their observed anisotropies on Earth.

Zusammenfassung

Die Erforschung der Anisotropien hochenergetischer kosmischer Strahlung und ihres Ursprungs

In den letzten Jahrzehnten sind in der kosmischen Strahlungs-Physik mehrere offene Fragen deutlich geworden, insbesondere hinsichtlich des Ursprungs und der Beschleunigungsmechanismen hochenergetischer kosmischer Strahlung. Trotz verbesserter experimenteller Methoden bleiben die genauen Quellen von Teilchen mit Energien oberhalb des PeV-Bereichs unklar. Die Übergangsregion zwischen galaktischer und extragalaktischer kosmischer Strahlung – insbesondere um den Bereich des „Knies“ und des „Knöchels“ des Energiespektrums – wird weiterhin kontrovers diskutiert. Zudem ist der Einfluss galaktischer Magnetfelder auf die Ausbreitung und Anisotropien kosmischer Strahlung noch nicht vollständig verstanden. Die Lösung dieser Probleme ist entscheidend, um die grundlegenden Prozesse der Beschleunigung und Ausbreitung kosmischer Strahlung im Universum zu verstehen.

Kosmische-Strahlungs-Anisotropie ist ein weiterer wichtiger Aspekt zur Erklärung des Ursprungs kosmischer Strahlung und ihrer Bewegung im Raum. Anisotropien, die auf kleinen wie auch großen Skalen über verschiedene Energiebereiche hinweg beobachtet werden, liefern wertvolle Informationen über die Quellen kosmischer Strahlung und die Magnetfelder, die sie durchquert. Bei niedrigen Energien (unterhalb weniger TeV) werden großskalige Anisotropien wahrscheinlich durch die Verteilung naher Quellen und die Struktur des lokalen interstellaren Magnetfelds beeinflusst. Bei höheren Energien, insbesondere nahe dem „Knie“, werden die Anisotropien hingegen ausgeprägter und schwerer erklärbar, was auf komplexere Transportprozesse und mögliche nahe galaktische Quellen hinweist. Jenseits des „Knöchels“ nimmt die Anisotropie ab, was auf einen möglichen extragalaktischen Ursprung in diesem Bereich hindeutet, auch wenn dieser Übergang weiterhin untersucht wird. Zudem können kleinskalige Anisotropien, wie sie von Experimenten wie Tibet ASγ und IceCube beobachtet wurden, helfen, lokalisierte Quellen kosmischer Strahlung zu identifizieren und die Rolle der Turbulenz des Magnetfelds bei ihrer Ausbreitung zu beleuchten.

In dieser Dissertation wird zuerst die Empfindlichkeit des zukünftigen IceCube-Gen2-Oberflächenarrays für die Rekonstruktion der kosmischen Strahlungsanisotropie untersucht, und es wird gezeigt, dass IceCube-Gen2 wichtigen zusätzlichen Datenpunkten zum Gesamtbild beitragen wird. Anschließend optimiere ich analytische Methoden zur Rekonstruktion von Anisotropien der kosmischen Strahlung und zur Identifizierung signifikanter groß- und mittelskaliger Anisotropien oberhalb von 10 PeV in der Milchstraße. Die Empfindlichkeit der Daten für die Anisotropierekonstruktion wird zunächst mit traditionellen Methoden bewertet; anschließend werden mehrere optimierte Rekonstruktionsverfahren eingeführt, um statistische Einschränkungen zu überwinden und die Unsicherheiten in den rekonstruierten Dipolparametern zu verringern. Ein einheitliches Bild der kosmischen Strahlungsanisotropie wird durch die Kombination von Messungen großer Experimente gewonnen, die sich vom TeV-Bereich bis über EeV-Energien erstrecken. Dadurch ergibt sich eine kontinuierliche und globale Sicht auf die Entwicklung der Dipol-Anisotropien von galaktischen bis zu extragalaktischen Ursprüngen, wobei eine globale funktionale Anpassung der Dipolamplitude eine charakteristische „W-förmige“ Struktur zeigt. Aufgrund begrenzter

Statistik bleibt jedoch die Dipolphase zwischen 10^{15} eV und 10^{19} eV unsicher. Insbesondere im Bereich von 1 bis 100 PeV zeigen sich komplexe Phasenvariationen, insbesondere in den KASCADE-Grande-Daten. Die optimierten Methoden werden anschließend auf die KASCADE-Grande-Daten angewendet und zeigen Hinweise auf eine 3σ-Anisotropie bei 33 PeV, was auf Cygnus OB2 als mögliche Quellregion hindeutet. Ausbreitungsstudien von Cygnus OB2 zur Erde zeigen eine Übereinstimmung mit der beobachteten Anisotropie bei 33 PeV. Abschließend legt diese Dissertation nahe, dass Pulsarwindnebel (PWNe), die mit Supernovaüberresten (SNRs) assoziiert sind, als natürliche Beschleuniger galaktischer kosmischer Strahlen im PeV-Energiebereich wirken könnten und eine Erklärung für den Ursprung sowie die auf der Erde beobachteten Anisotropien liefern könnten.

Contents

Introduction

In recent decades, cosmic-ray physics has confronted several persistent challenges, particularly concerning the origin, acceleration mechanisms, and propagation of high-energy cosmic rays. Despite remarkable progress in experimental observations, the sources capable of accelerating particles beyond the PeV scale remain uncertain. The transition between Galactic and extragalactic components, especially near the "knee" and "ankle" features of the spectrum, remains a topic of debate, as does the role of Galactic magnetic fields in shaping cosmic-ray propagation and anisotropies. Resolving these questions is essential for understanding the fundamental processes that govern the production and transport of cosmic rays throughout the universe.

A central open problem in astroparticle physics is identifying the mechanisms and sources responsible for accelerating Galactic cosmic rays (GCRs) to PeV energies. The "knee" of the cosmic-ray spectrum, a steepening near 3 PeV, is generally interpreted as a limit to the acceleration capability of typical Galactic sources. Supernova remnants (SNRs) are widely regarded as the primary accelerators of GCRs via diffusive shock acceleration (DSA) [1]. However, whether SNRs alone can account for the highest-energy GCRs remains uncertain, as the efficiency of shock acceleration may decrease near PeV energies [2]. This has led to the hypothesis that additional accelerators, such as pulsar wind nebulae (PWNe) or other energetic astrophysical systems, might contribute to the Galactic cosmic-ray population [3, 4]. Sources capable of accelerating particles up to and beyond PeV energies are known as Pevatrons [5]. Although early gamma-ray observations had revealed promising PeVatron candidates, including the Galactic Centre region [6] and the Cygnus region [7], no definitive Galactic PeVatron had been confirmed at that time [4]. But the recent LHAASO results now provide strong evidence for super–PeV particle acceleration in several Galactic sources [8, 9]. Another major uncertainty concerns the energy at which the transition from Galactic to extragalactic cosmic rays occurs. While particles below the "knee" are generally attributed to Galactic sources, the origin of cosmic rays beyond this energy remains not fully resolved [3, 4, 10]. The "ankle" feature around 5×10^{18} eV, interpreted as the transition between Galactic and extragalactic cosmic rays, proven by Auger [11]. In many transition scenarios, an extragalactic component is already expected to contribute around the second knee at around $(4\text{–}8)\times10^{17}$ eV, where the Galactic iron component cuts off, and a lighter, more proton-dominated extragalactic flux begins to emerge [12, 13]. Above the ankle, at around $(3\text{–}5)\times10^{18}$ eV, the extragalactic cosmic rays likely dominate the flux, in the traditional ankle–transition picture, where a flatter extragalactic spectrum overtakes the steeper Galactic one [12, 13, 14]. Understanding this transition is key to determining whether PeV and EeV cosmic rays originate from Galactic PeVatrons or from extragalactic accelerators.

Cosmic-ray anisotropy provides an additional diagnostic of cosmic-ray origins and propagation. Large- and small-scale anisotropies observed over a broad energy range reveal the interplay between source distribution and magnetic field structure. At low energies (below a few TeV), the large-scale anisotropy likely reflects nearby sources and the configuration of the local interstellar magnetic field [15]. Moreover, small-scale sidereal anisotropies detected by the HAWC and IceCube experiments provide further insight into localised sources and the impact of magnetic-field turbulence on cosmic-ray propagation [16]

in this energy range. At higher energies, especially near the "knee" at 3 PeV, the anisotropy pattern becomes more complex, possibly indicating non-uniform diffusion or the influence of nearby Galactic sources [17]. Moreover, the observed large-scale anisotropy above the "ankle", at $(3\text{–}5)\times10^{18}$ eV, remains at the few-percent level, and its energy dependence and phase are agreed with a growing extragalactic contribution, although the exact origin and transition scenario remain model-dependent and not yet fully established [18].

The central objective of this thesis is to develop optimized analytical methods for reconstructing cosmic-ray anisotropies and to identify significant large- and medium-scale structures above PeV energies within the Milky Way. These efforts aim to trace the origin regions of Galactic cosmic rays, investigate their propagation to Earth, and propose a preliminary conceptual model for their acceleration to PeV energies. In Chapter 1, an introduction to the physics of cosmic rays will be given. Chapter 2 reviews existing methods for cosmic-ray anisotropy reconstruction. To address the lack of dipole anisotropy measurements between the PeV and EeV energy ranges, Chapter 3 investigates the sensitivity of observational data to anisotropy reconstruction, with particular attention to future observatories such as the IceCube-Gen2 surface array. Chapter 4 introduces optimized reconstruction methods for dipole anisotropy, designed to overcome statistical limitations and reduce uncertainties in the recovered parameters. In Chapter 5, results from major experiments, including IceCube and LHAASO, and so on, are combined to construct a unified, energy-dependent picture of dipole anisotropies up to 50 EeV. Chapter 6 applies the optimized approaches to KASCADE-Grande data to search for anisotropy signatures, investigate their possible origins, and study propagation features from a potential source region (Cygnus OB2) to Earth using the CRPropa software code. Finally, the chapter presents a potential proposal for interpreting the results, focusing on PeV particle acceleration in Galactic sources such as supernova remnants and their connection to the anisotropies observed by KASCADE-Grande.

Chapter 1

Cosmic Rays

Cosmic rays are high-energy charged particles and ionized nuclei, consisting of about 90% protons, 9% alpha particles, as well as the remaining heavier nuclei, and a minor component of electrons and positrons, with energies up to 10^{20} eV [19], spanning a broad energy spectrum. Because cosmic rays are charged, they are deflected by magnetic fields, primarily the Galactic magnetic field (GMF), and, therefore, at most energies, cannot be directly traced back to their sources. At low energies (GeV–TeV), the geomagnetic field and the solar wind (a magnetized plasma from the Sun) strongly affect what we observe near Earth. At higher energies (PeV), the deflections caused by the geomagnetic field or the solar wind per unit charge become smaller and can often be neglected, but the deflections due to the Galactic magnetic field remain significant.

Where do cosmic rays come from? Two related open questions remain: (i) what are the origins of cosmic rays across the full energy range, and (ii) what mechanisms are capable of accelerating them to very high energies (PeV) or even ultra-high energies (EeV) and above. The origin of cosmic rays is still not fully understood. Current candidates include the Sun, with its solar flares (a known source of low-energy cosmic rays); supernova remnants and pulsar-wind nebulae within our Galaxy (likely contributors up to at least the knee); and active galactic nuclei along with other extreme extragalactic environments at the highest energies [19]. The acceleration of cosmic rays occurs within astrophysical sources, while their subsequent propagation takes place in the surrounding interstellar medium (ISM) and, at higher energies, in the intergalactic medium (IGM).

The history of cosmic-ray physics is closely linked to the development of our understanding of the cosmos and the relevant high-energy astrophysics. One of the earliest recorded celestial events was the observation of the Crab Nebula in 1054 by the Chinese astronomer Weide Yang. This supernova explosion, documented in the section "Treatise on Astronomy" of the book "History of the Song Dynasty" nearly a thousand years ago, is now considered a possible source of cosmic rays. However, at that time, the tools for investigating such phenomena were limited to human sight, and the true nature of these powerful events remained a mystery for centuries, until the 20th century, with the development of modern physics, particularly breakthroughs in quantum mechanics, relativity, and particle physics. From then on, people began to uncover the mechanisms behind such high-energy processes. In 1912, Victor Hess discovered cosmic rays during a series of balloon flights, where he measured the ionization rate in the upper atmosphere [20]. Since then, it has been established that cosmic rays induce the atmosphere to produce cascades of secondary particles, known as air showers, which can be detected with modern ground-based instruments.

The detection and measurement of cosmic rays employ various techniques and instruments, including ground-based detectors, balloon-based experiments, and satellite-based observatories. Ground-based arrays such as the Pierre Auger Observatory [21], KASCADE (KArlsruhe Shower Core and Array DEtector) [22, 23, 24], LHAASO (Large High Altitude Air Shower Observatory) [25], the IceCube Neutrino Observatory [26], and its next-generation extension IceCube-Gen2 [27], detect extensive air showers produced when

high-energy cosmic rays interact with the Earth's atmosphere. Balloon-based experiments, such as CREAM (Cosmic Ray Energetics And Mass) [28, 29], have directly measured the elemental spectra of primary cosmic rays (from protons to iron) up to energies of about 10^{15} eV. Satellite-based experiments, such as the Alpha Magnetic Spectrometer (AMS) on the International Space Station [30] and DAMPE (DArk Matter Particle Explorer) [31], directly measure cosmic-ray properties in space, free from atmospheric interference, and provide unprecedented precision for protons, nuclei, electrons, positrons, and antiprotons up to the TeV range. Future experiments, including the LHAASO extensions [25], GRAND (Giant Radio Array for Neutrino Detection) [32], HERD (High Energy Cosmic-Radiation Detection) [33], GCOS (Global Cosmic-Ray Observatory) [34, 35], and SWGO (Southern Wide-field Gamma-ray Observatory) [36], will provide wider fields of view and extend the accessible energy range (LHAASO up to 10^{18} eV and HERD up to 10^{15} eV). These experiments are expected to improve cosmic-ray studies significantly.

Already now, many physical features of cosmic rays are being studied, including extensive air showers, the energy spectrum, mass composition, and anisotropies. These measurements are just the tip of the iceberg, providing deeper insight into the propagation of cosmic rays through the Galaxy and the mechanisms of their acceleration, which are closely related to the main topics explored in this thesis.

1.1 Discovery and History

At the beginning of the twentieth century, it was known that electroscopes discharged spontaneously in air, indicating the presence of an unknown penetrating radiation. The dominant view was that this radiation came from radioactive substances in the Earth's crust. Two scientists, Domenico Pacini and Victor Hess, developed independent experimental approaches that changed this view and led to the recognition of cosmic rays as an extraterrestrial phenomenon [37].

In 1911, Pacini performed a series of measurements of electroscope discharge rates over water and underwater (see the left panel of Figure 1.1). Using a sealed electroscope enclosed in a copper box, he compared measurements at the surface of the sea and at a depth of three meters in Livorno Bay [37]. Pacini observed that the ionization rate decreased when the device was immersed in water. He concluded that a significant portion of the radiation was independent of the direct action of radioactive substances in the soil, since water effectively shielded radiation coming from above. These measurements provided the first clear evidence for an origin of penetrating radiation from outside the Earth [37]. Almost at the same time, Hess performed a different experiment. In 1912, during a series of balloon flights to measure radiation levels at altitudes of up to about 5 km, he observed that the ionization rate increased with height (see the right panel of Figure 1.1), showing a strong signal. This proved that the radiation could not originate from the Earth's surface, but must come from space [20]. This marked the beginning of cosmic-ray research. For this discovery, Hess was awarded the 1936 Nobel Prize in Physics. In 1913, Kolhörster's high-altitude balloon measurements revealed an increase in ionizing radiation as the balloon climbed higher (altitude) into the atmosphere, thereby confirming the extraterrestrial origin of cosmic rays, and provided independent confirmation of Victor Franz Hess's 1912 discovery of cosmic rays [38, 39].

After these discoveries, many research groups tested how penetrating the new radiation was. In the 1920s, Millikan and his collaborators carried out a series of experiments

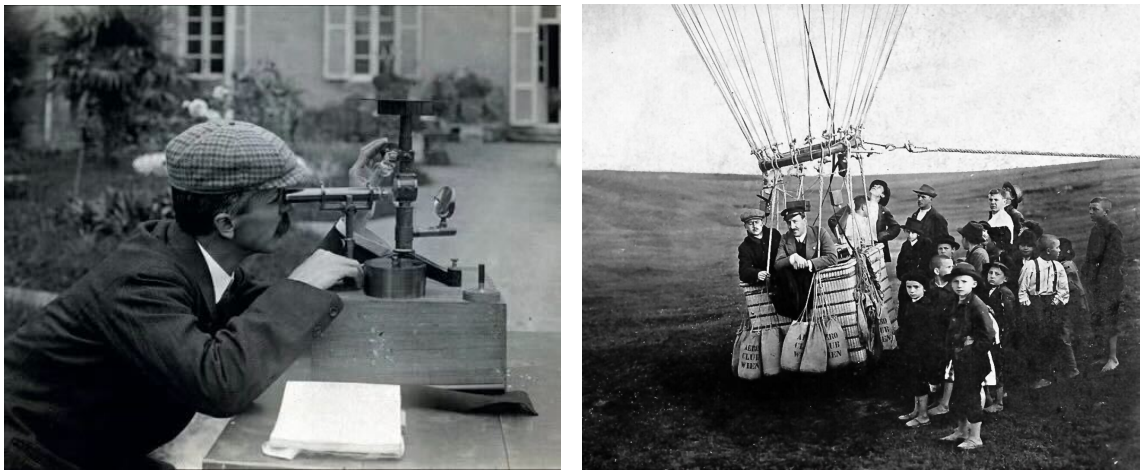


Figure 1.1: Left: Domenico Pacini performing a measurement in 1912 (image taken from the CERN webpage). Right: Victor Hess in a hydrogen balloon in 1912, during the experiments that led to the discovery of cosmic rays.

using sounding balloons and absorption measurements in air, lead, and water to study the penetrating power of what Millikan named cosmic rays [40, 41, 42, 43, 44]. Their lake absorption experiments showed that the radiation could even reach the bottom of deep water [41]. Besides, experiments in mines and tunnels showed that it could pass through thick layers of rock, much more than ordinary γ-rays [45]. From the depth–ionization curves, Millikan argued that the radiation had a wide spectrum of penetrating powers, which he interpreted as evidence for atom-building processes in space, possibly the formation of helium from hydrogen [40, 44]. They also strongly pointed out that cosmic rays do not simply come from stars or from the Earth, but from truly deep space regions with very low densities and high temperatures where atom-building may occur [44].

Further advancements came in the 1930s when people began to understand that cosmic rays were not merely a single type of radiation but a complex mix of high-energy particles from different directions. During this period, Bruno Rossi and Pierre Auger made significant contributions to the study of cosmic rays. Rossi developed the electronic coincidence method, which made it possible to register time-correlated particles across separated counters and opened the way to identifying extensive air showers (EAS) studies, and allowed for the detection of cosmic rays through multiple detectors and larger detection areas [46]. In the late 1930s, Pierre Auger used arrays of detectors spread over large areas and observed near-simultaneous signals across kilometre scales, which provided the first direct evidence of extensive air showers. Auger observed simultaneous signals that indicated the presence of large-scale particle cascades caused by high-energy primary cosmic rays interacting with the atmosphere. They concluded that a single high-energy primary could generate a vast atmospheric cascade of secondary particles, spreading over several square kilometres. Auger's pioneering experiments confirmed that extensive air showers were produced by primary cosmic rays with energies far beyond those achievable by any man-made accelerators of the time [47].

In the 1950s and 1960s, the development of new detector arrays allowed studies at higher energies with better statistics. At Volcano Ranch, John Linsley measured the cosmic-ray spectrum above 10^{17} eV and later reported events close to 10^{20} eV, opening the ultra-high-energy regime [48, 49]. Around the same time, studies of the knee in the

spectrum near 3 PeV, now known as the knee, were first observed with an air-shower array at Moscow State University [50]. Soon after, Peters suggested that the knee reflects a change of composition, with lighter nuclei cutting off first and heavier nuclei extending to higher energies according to their rigidity [51], now known as the Peters cycle. Other techniques were also developed, such as Cherenkov light measurements of air showers provided more information on the spectrum and mass composition [52]. During this time, many new particles were discovered using photographic emulsions and Wilson cloud chambers [53]. Among these were hyperons, pions, and kaons, forming what became known as the "particle zoo". These unexpected findings forced major changes in particle theory and accelerator-based particle physics. At the same time, studies of the primary cosmic radiation advanced, and it was shown that cosmic rays include nuclei much heavier than helium. With further improvements in experimental techniques, the relative abundances of different nuclei, and even some of their isotopes, were measured with good precision [53]. By the mid-1960s, theory predicted that the very highest-energy cosmic rays would be attenuated by interactions with the cosmic microwave background, a process now called the Greisen–Zatsepin–Kuzmin (GZK) cutoff [54, 55]. These results marked the move from early cosmic-ray studies to modern high-energy astroparticle physics.

1.2 Detection

In this section, we will review several key detection techniques, including satellite-based detectors, Cherenkov tanks, scintillator arrays, radio arrays, and digital optical modules (DOMs). Besides, we will provide an overview of the phenomena associated with extensive air showers (EAS), the cosmic-ray energy spectrum, the mass composition of cosmic rays, and the "muon puzzle" in cosmic-ray air shower detections.

1.2.1 Detectors and Experiments

The detection of cosmic rays relies mainly on two approaches: direct measurements of cosmic-ray particles using the compact multi-detector devices, with a spectrometer or a calorimeter above the atmosphere in space, and indirect measurements on the ground that capture the signatures of extensive air showers. Indirect detection, such as that performed by ground-based arrays, measures secondary air-shower particles (EAS) through Cherenkov radiation, scintillation, radio emission, or optical detection with photomultiplier tubes (PMTs), for example, in water tanks or in-ice detectors. Together, these methods cover the full cosmic-ray energy range, from a few MeV up to beyond 10^{20} eV. In the following, we discuss the different types of detectors commonly used in both direct and indirect cosmic-ray detection.

Direct measurements are usually carried out with detectors on balloons and satellites, which allow the study of primary particles before they interact in the atmosphere. Such instruments are composed of multiple subsystems working together. A magnetic spectrometer (as in AMS) measures the momentum or mass-to-charge ratio of charged particles by tracking their curvature in a magnetic field and using extra tracking layers to improve magnetic bending and track precision. It is considered the most precise technique for determining the momentum and the charge sign of cosmic rays under strong magnetic forces (see the left panel of Figure 1.2). The Alpha Magnetic Spectrometer (AMS-02) on the International Space Station in space: it uses high-resolution silicon tracking planes inside a strong magnetic field to measure the curvature of the particle track and thus its rigidity [56,

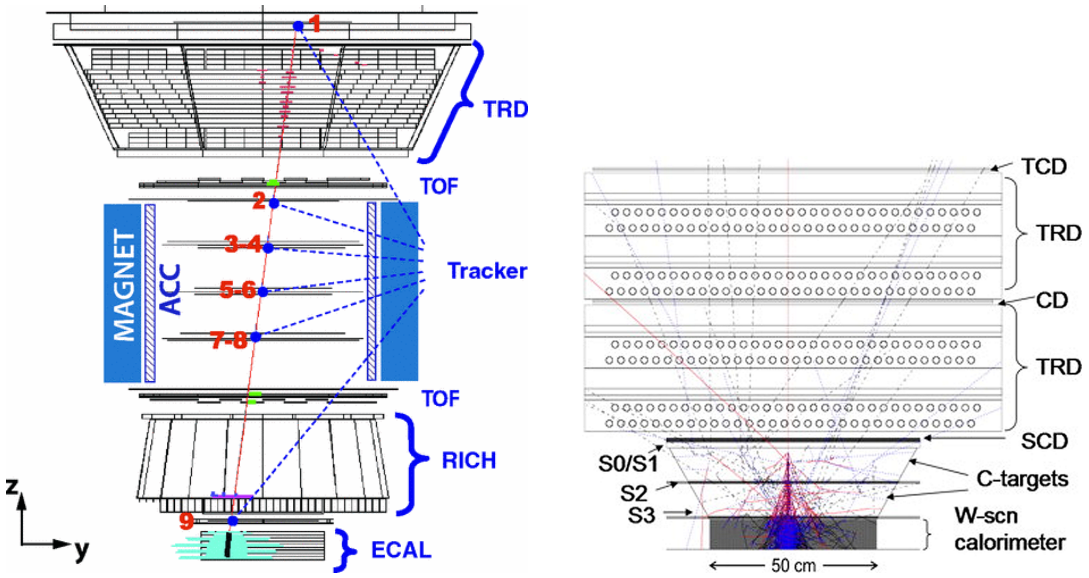


Figure 1.2: Left: The AMS-02 spectrometer detector installed on the International Space Station, illustrating the tracking of a TeV electron [56]. Right: A schematic cross-section of the CREAM instrument with a simulated proton-induced shower [58].

19]. Particle identification is achieved through several complementary subsystems. The transition radiation detector (TRD) discriminates between light leptons and hadrons, while the time-of-flight (TOF) counters measure velocity, direction, and charge via the ionization dependence on Z^2 [19]. An additional velocity and charge measurement is provided by the ring-imaging Cherenkov (RICH) detector. At the bottom of the instrument, the electromagnetic calorimeter (ECAL) records particle showers, confirming the identification of electrons and positrons and yielding an independent measurement of their energy [19]. The highest measurable energy in a spectrometer is limited by the accuracy of the track reconstruction, which sets the maximum detectable rigidity, and by the geometric acceptance defined by the size of the magnetic volume. With these capabilities, AMS-02 has extended direct measurements of protons and helium to above 1 TeV per nucleon, and of electrons and positrons to several hundred GeV. It measures the flux of cosmic rays from about 10^9 eV to 10^{12} eV [57] and also searches for antimatter and dark matter [56].

An alternative approach, which enables measurements at higher energies in direct cosmic-ray experiments, is to utilise a calorimetric detector without a magnetic spectrometer. A calorimeter measures the total energy of cosmic rays or their secondary particles by absorbing their energy, focusing on determining the particles' energy through energy deposition, as seen in DAMPE [31]. By removing the magnet, the instrument can achieve a much larger geometric acceptance, which is crucial for rare, high-energy events. In such detectors, the energy is determined from the development of the particle shower in the calorimeter. The energy resolution is less precise than in a spectrometer because of event-to-event fluctuations in shower development and because some particles escape through the bottom of the detector [19]. A well-known example is CREAM [28], which has flown several times on long-duration balloons over Antarctica (see the right panel of Figure 1.2). CREAM has measured the spectra of protons, helium, and heavier nuclei from above 1 TeV up to beyond 100 TeV per particle, extending the range of direct observations beyond that reached by AMS-02 [28, 29]. Besides, CALET on the International Space Station, as well,

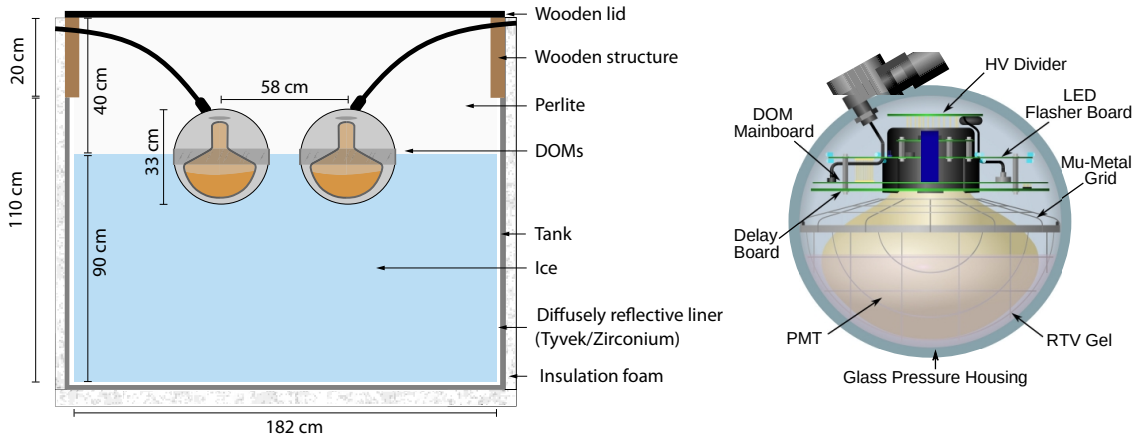


Figure 1.3: Left: Cross-sectional view showing the dimensions of a Water-Cherenkov tank in IceTop at IceCube. Right: Schematic illustration of a Digital Optical Module (DOM) used in IceCube. Figures adapted from [66].

reported the cosmic-ray spectrum from 10^{11} eV to 10^{14} eV [59]. Another example is the PAMELA satellite, which has contributed valuable data on cosmic rays up to the TeV range, in particular γ-rays [60]. Balloon experiments and satellite missions such as GREAM, DAMPE, AMS-02, CALET, and PAMELA have provided precise measurements of the fluxes of protons, helium, heavier nuclei, and electrons and positrons. Direct detection has therefore revealed the detailed energy spectra of individual elements and the ratios of secondary to primary nuclei (for example, boron-to-carbon). By this ratio, the lifetime of CR at low energies is Galactic as predicted by the leaky box model, which will be introduced in Section 1.4.1.

Indirect measurements are needed to study cosmic rays at the highest energies because the flux becomes very low [19]. The most practical approach is to build large detectors on the ground that can operate for many years and cover very large areas. These air-shower arrays extend over hundreds or even thousands of square kilometres [19]. They do not see the primary cosmic rays directly but instead record the showers of secondary particles produced when primaries hit the atmosphere, which provide indirect information about the incoming particle [19]. Surface detector arrays (air-shower arrays) use many particle detectors arranged in a grid, with spacings from about 13 m in KASCADE [22] and 15 m in Tibet AS-γ (Air-Shower Gamma Ray) [61] to more than 1 km in the Telescope Array [62] and Pierre Auger Observatory [21], depending on the energy range to be measured. Showers are identified by fitting the lateral distribution of particle densities in nearby detectors, and the arrival direction is reconstructed from the relative arrival times of the shower front [19]. The air-shower arrays have contributed to fundamental discoveries, such as the observation of the knee in 1958 [50] and the first event near 10^{20} eV at Volcano Ranch [49]. In the knee region, many arrays have provided data on the flux and composition of primary cosmic rays, including IceCube [63, 26], CASA-MIA [64], EAS-TOP [65], KASCADE [23], and GRAPES [19].

Air-shower arrays often employ more than one detector type. Water-Cherenkov tanks are surface detectors that use the Cherenkov effect. When charged shower particles travel faster than light in water (but not quicker than the speed of light in vacuum), they emit a cone of Cherenkov light [19]. Photomultiplier tubes (PMTs) inside the tank record this light

and convert it to electronic signals. Atmospheric Cherenkov techniques detect UV light (300–400 nm) from shower particles and help trace shower development [19]. At the Pierre Auger Observatory, about 1,600 water-Cherenkov tanks cover $\sim 3,000\,\mathrm{km}^2$ [67]. They are optimized for ultra-high-energy cosmic rays ($> 10^{18}$ eV), and provide precise energy and arrival-direction estimates with reliable detector responses. Auger combines surface data with fluorescence measurements to determine composition. Moreover, IceTop at IceCube Neutrino Observatory is an array of frozen water tanks, each equipped with two DOMs, placed above the strings of the in-ice array [68]. In addition to surface water-Cherenkov tanks, deep-ice optical modules (DOMs) can detect Cherenkov light from the high-energy (around 300 TeV) muon bundles produced by cosmic-ray air showers [69, 70]. IceCube contains 5,160 DOMs, each with a PMT, embedded in a cubic kilometre of Antarctic ice. When an air shower reaches the array, high-energy muons penetrate the ice and emit Cherenkov light, with the signal dominated at high energies by stochastic energy losses and energy deposits along the muon tracks [70]. On the surface, operating IceTop and IceCube together enables a coincident analysis, where the electromagnetic component of the shower is measured at the surface and the muon content deep in the ice. This combined method improves the energy reconstruction and provides composition-sensitive observables, providing new information on the primary cosmic rays [63, 26].

Fluorescence telescopes detect ultraviolet light from nitrogen molecules excited by charged particles in an air shower, an idea proposed in the 1960s by Greisen, Chudakov, Suga, and others [19]. The emission lies mainly between 300 and 400 nm, with about 4–5 photons per meter per charged particle at altitudes of 5–10 km [19]. This technique provides a nearly calorimetric energy measurement: about 90% of the shower energy is deposited in the atmosphere, while the remaining 10% (the missing energy) is carried by muons and neutrinos, with only a few percent model uncertainty [71, 19]. Because the light is emitted isotropically, showers can be observed from 5 to 35 km depending on geometry and energy [19]. Reconstruction requires shower-axis geometry, subtraction of the Cherenkov component, and atmospheric corrections. The method gives both the total energy and the depth of maximum development, $X_{\max}$, which is a key composition indicator [19]. Its main limitation is a low duty cycle of about 10-15% due to the need for dark, clear nights and careful monitoring of atmospheric conditions [19].

Scintillator arrays that measure charged particles in cosmic ray air showers. When a particle passes through a scintillator, it excites the material, producing a flash of light [19]. This light is then detected and converted into an electronic signal by photomultiplier tubes (or SiPM). Scintillators are particularly sensitive to muons, electrons, and gamma rays produced in air showers as secondary shower particles [22]. Experiments such as the KASCADE utilize scintillator arrays to study cosmic rays with energies ranging from 10^{14} eV to 10^{17} eV [22, 23]. Details on the scintillator array of KASCADE can be found at the website[1]. The current IceTop has the problem of snow accumulation on the tanks, which weakens and distorts the measured signals [72]. To fix this and improve composition studies, the planned upgrade will add an array of plastic scintillator panels above the tanks and thereby improve the detector response [72]. Other experiments also use scintillators. For example, the Telescope Array (TA) employs two-layer plastic scintillator stations, which provide accurate timing and signal density measurements [73]. In general, scintillators record fast timing and particle densities, which are then used to reconstruct the energy and arrival direction of the primary cosmic rays [22].

[1] https://kcdc.iap.kit.edu/information/aboutKASCADE/.

Radio arrays detect radio emissions from extensive air showers (EAS). When the charged particles in a shower move through the atmosphere, they produce coherent radio waves mainly through two processes: geomagnetic deflection of electrons and positrons in the Earth's magnetic field, and the Askaryan effect from the shower's negative charge excess [19, 74]. The emission is strongest below 100 MHz, where the signal scales nearly linearly with shower energy. The radio pulse is forward-beamed, with the field strength at about 100 m from the core giving the energy, and the lateral slope providing information on the depth of shower maximum, $X_{\max}$, which is linked to composition and air-shower geometry [19]. Radio pulses from air showers were first seen by Jelley in 1965 [75]. Early studies in the 1960s–70s confirmed the effect, but the method was not yet reliable [19]. Modern radio arrays have confirmed these effects and shown that radio detection can be a powerful tool, which was first shown by LOPES at KASCADE [19]. LOFAR (Low-Frequency Array) is especially effective in the range $10^{16.5}$–10^{18} eV [76], while the AERA (at Auger) covers 10^{17}–10^{18} eV [77]. Compared to optical methods, radio has the advantage of continuous operation in many weather conditions and can be deployed over large areas at relatively low cost. This makes radio detection good at measuring shower energy, arrival direction, and providing additional constraints on composition [74].

1.2.2 Future Experiments

What's next? Several upcoming projects and upgrades aim to extend the energy and particle coverage of cosmic-ray studies and introduce new detection techniques. These efforts will increase exposure, improve mass-composition sensitivity, and strengthen links to multi-messenger observations. In particular, some projects focus on improving the precision of measurements at lower energies (MeV), while others will open the PeV–EeV window with higher accuracy and larger statistics, providing new opportunities to uncover the origin of cosmic rays and the mechanisms that drive their acceleration.

Direct measurements. There is growing interest in planetary platforms. Recently, it was proposed to develop medium- and high-energy particle detectors tailored to the lunar surface radiation environment for Chang'e-7 [78]. The planned energy ranges are: protons 30 keV–300 MeV, electrons 30 keV–12 MeV, and heavy ions 8–300 MeV per nucleon [78]. In orbit, the HERD (High Energy Cosmic-Radiation Detection) experiment will be installed on the Chinese Space Station around 2027 [33]. HERD is designed to directly measure the spectrum and composition of protons, heavy nuclei, electrons, and gamma rays with high energy resolution and a large geometric acceptance [33, 79]. It will precisely measure primary cosmic rays from 30 GeV up to the PeV scale, probing the origin of the knee. Further details are available on the project webpages.[1][2]

Indirect measurements. Several new projects are planned. GRAND (Giant Radio Array for Neutrino Detection) will deploy large radio arrays totally about 200,000 antennas to detect air-shower signals from ultra-high-energy cosmic rays and extremely high-energy (EHE) neutrinos [32]. For cosmic rays, GRAND will measure inclined showers in the range 10^{17} to $10^{19.5}$ eV and aims to study the spectrum, large-scale anisotropies, and composition with unprecedented statistics [32]. In China, LHAASO is already operating, combining particle detectors, water-Cherenkov detectors, and Cherenkov telescopes. Planned extensions will expand sky coverage and dynamic range, improving composition and anisotropy studies in the TeV–PeV region and testing particle acceleration in Galactic sources up to the

[1] http://herd.ihep.ac.cn.
[2] https://herd.cloud.infn.it/en.

Figure 1.4: Left: Illustration of the HERD instrument, which will be installed on board the Chinese Space Station (CSS) around 2027, with a planned lifetime of 5–10 years (image taken from [86]). Right: Layout of the SWGO array with surface detectors arranged at a high-altitude site (4770 m) in Chile (credit: Richard White, MPIK).

"PeVatron" scale [25, 8]. In the US, the Telescope Array (TA) is expanding to four times its current size, providing much higher statistics above 10^{19} eV to confirm and study the northern-sky hotspot and to measure the highest-energy flux with reduced uncertainty [73]. GRAND, LHAASO, and TA, all located in the Northern Hemisphere, provide extensive data for cosmic-ray studies.

However, in the Southern Hemisphere, there is no LHAASO-like facility at present. The SWGO (Southern Wide-field Gamma-ray Observatory) is proposed for a high-altitude site in South America [80, 36]. SWGO will focus on gamma rays above a few hundred GeV and will also provide cosmic-ray measurements in the TeV–PeV range, including spectrum, anisotropy, and composition proxies over the southern sky, complementing northern arrays [80, 36]. In Argentina, AugerPrime upgrades Auger by adding scintillator detectors on each surface tank, underground muon detectors at selected sites, and radio antennas [81, 82]. Compared to Auger, which measured the spectrum and arrival directions with the largest exposure, but had limited composition sensitivity. These new detectors separate the electromagnetic and muonic components of each shower, enabling event-by-event composition tagging [81, 82]. AugerPrime aims to clarify the origin of the flux suppression above $10^{19.5}$ eV (maximum source energy and propagation effects) and to search for anisotropies with composition information. At the South Pole, IceCube-Gen2 will add a large surface array of scintillators and radio antennas working with the in-ice detector [83]. It will study showers in the PeV–EeV range with special sensitivity to muon content, and the surface array will also act as a veto for neutrino searches, while improving composition studies [83, 84, 27, 85]. Finally, GCOS (Global Cosmic-Ray Observatory) is a long-term proposal for a next-generation international facility designed to cover a very large area and collect large exposures above 10^{18} eV [34]. Its goal is to provide precise measurements of the spectrum, composition, and arrival directions of ultra-high-energy cosmic rays, ultimately aiming to identify their sources [34, 35].

To conclude, the future indirect experiments will mainly cover the PeV–EeV range, which is also the focus of this thesis and the key region for studying cosmic-ray anisotropies. At present, only KASCADE-Grande [87], IceCube [70] and Auger [88, 89] explore this range, and so far, no significant anisotropies have been observed.

1.2.3 Extensive Air Showers

The Earth is hit by cosmic-ray primaries every second, most often protons or nuclei accelerated in astrophysical sources such as the Sun, supernova remnants (SNR), and active galactic nuclei (AGN) [19], with energies ranging from below 1 GeV up to about 10^{20} eV [90]. When such a high-energy incoming particle enters the atmosphere, it interacts with a nucleus in the air, mainly nitrogen, oxygen, or argon nucleus at an altitude of about 15–35 km above the Earth's surface or inside, creating an initial collision at the primary interaction point [90]. This first interaction produces many secondary particles, which in turn interact and decay. The cascade multiplies and develops into what is called an extensive air shower (EAS). The shower evolution grows, with shower front expansion, reaching a maximum number of particles at some depth, and then gradually fades as the particles lose energy and are absorbed by the atmosphere, influenced by particle scattering effects, until reaching the final atmospheric depth [47, 19].

In an air shower, the main secondary pions ($\pi^{\pm}$ and π^0), kaons ($K^{\pm}$ and K^0), and other hadrons are produced. Neutral pions (π^0) decay almost instantly into two photons,

$$\pi^0 \to \gamma + \gamma, \tag{1.1}$$

which drives the electromagnetic cascade through pair production ($\gamma \to e^+e^-$) and $e^{\pm}$ bremsstrahlung [19, 90]. The γ-rays then propagate and contribute to the electromagnetic component of the air shower through pair production and further interactions. The charged pions (π^+ and π^-) are responsible for generating the muonic component of the air shower. Charged pions ($\pi^{\pm}$, with $c\tau \approx 7.8\,\mathrm{m}$) usually re-interact before they can decay when their energy is above about $E_\pi \gtrsim 30\,\mathrm{GeV}$ [90]. At lower energies, their decay produces muons and neutrinos, a channel that happens almost every time a charged pion decays,

$$\pi^+ \to \mu^+ + \nu_\mu\ , \quad \pi^- \to \mu^- + \bar{\nu}_\mu \quad (\sim 100\%), \tag{1.2}$$

where the muons, being relatively long-lived particles, can travel significant distances through the atmosphere and often reach the Earth's surface before decaying [19]. Besides, neutral kaons (K^0) are relatively rare in extensive air showers compared to charged pions and kaons. Charged kaons ($K^{\pm}$, with $c\tau \approx 3.7\,\mathrm{m}$) behave in a similar way, but due to their shorter lifetime, they tend to decay at somewhat higher energies [90], with decay channels that resemble those of muons and specifically follow:

$$K^+ \to \mu^+ + \nu_\mu\ , \quad K^- \to \mu^- + \bar{\nu}_\mu \quad (\sim 63.5\%), \tag{1.3}$$

where this channel occurs only about 63.5% of the time, while the remaining ($\sim 36.5\%$) goes into other decay channels. The left panel in Figure 1.5 illustrates the main decay channels of pions and kaons in extensive air showers. The decay of muons from the charged pions and kaons has a similar chain, namely

$$\mu^+ \to e^+ + \nu_e + \bar{\nu}_\mu\ , \quad \mu^- \to e^- + \bar{\nu}_e + \nu_\mu, \tag{1.4}$$

where e^+ and e^- are positrons and electrons, and ν_e and $\bar{\nu}_e$ are electron neutrinos and antineutrinos. When conditions are such that all particles decay, one therefore expects

$$(\nu_\mu + \overline{\nu}_\mu)/(\nu_e + \overline{\nu}_e) \sim 2\ , \quad \nu_e/\overline{\nu}_e \sim \mu^+/\mu^-. \tag{1.5}$$

Moreover, in pion–muon decay, the three neutrinos, each carrying a similar share of the pion's original energy [19].

A cosmic-ray–induced extensive air shower develops through three principal components: hadronic, electromagnetic, and muonic. The hadronic core consists primarily of nucleons,

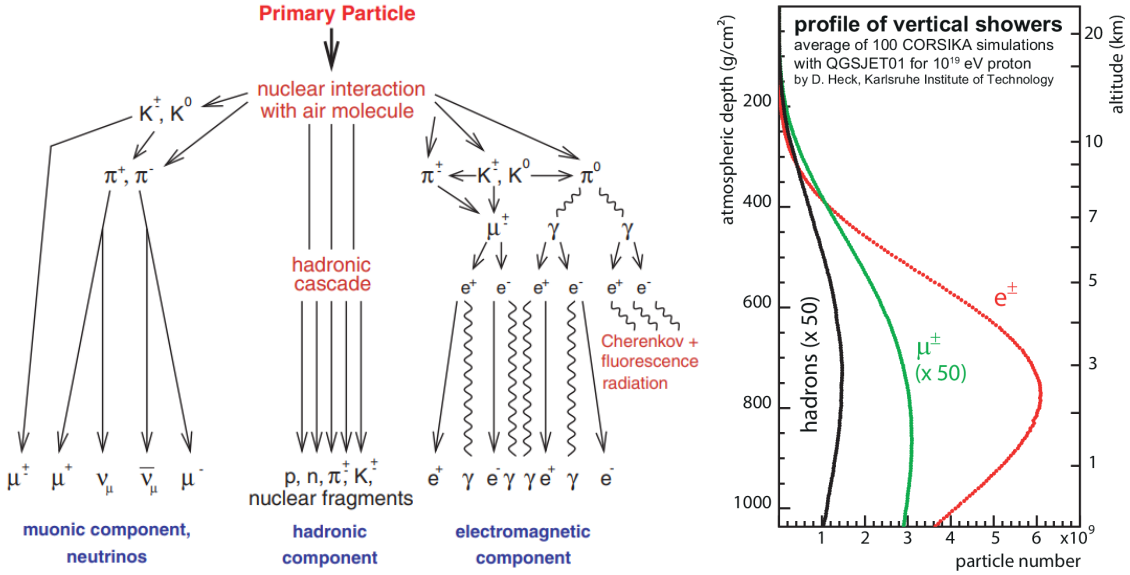


Figure 1.5: Left: The extensive air shower (EAS) profile with a schematic illustration of an air shower with various detection channels. Collisions in the atmosphere create many secondary hadrons, mainly pions and kaons. Charged ones decay into muons and neutrinos, forming the long-range muon component. Neutral pions decay into photons, which start the electromagnetic cascade and spread the primary energy into many particles. (taken from [91]). Right: Longitudinal EAS profile for different particle types, taken from [92].

charged pions, kaons, and other high-energy hadrons that undergo successive inelastic interactions, thereby sustaining the particle cascade. Neutral pions (π^0) decay almost instantaneously into two photons, which trigger electromagnetic subshowers through alternating processes of pair production and bremsstrahlung. The evolution of these electromagnetic cascades can be effectively approximated by the Heitler model [93, 19].

In an air-shower, the longitudinal profile (see the right panel in Figure 1.5) can be expressed by depth, which is not in meters but in column density, called slant depth. It is defined as $X(l) = \int_0^l \rho(h)\, dl$ (unit in $\mathrm{g\,cm^{-2}}$), where $\rho(h)$ is the atmospheric mass density at altitude h, and X meaning the mass of air contained in a column of $1\,\mathrm{cm}^2$ cross section along the trajectory, describing depth development. At all energies, electrons lose energy by ionization and atomic excitation, which gives an almost constant loss rate per depth, written as $-\alpha(E)$. At higher energies, bremsstrahlung becomes dominant, tthe radiated power increases with the square of the electron's energy, leading to a term $-E/X_0$, where X_0 is the radiation length of the medium. Since these two processes act independently, the total mean loss is simply the sum of the two contributions. The critical energy E_c is then defined at the point where ionization and bremsstrahlung losses are equal, $\alpha(E_c) = E_c/X_0$ [90]. For air, this condition gives $E_c \simeq 87$ MeV, meaning that electrons above this energy mainly lose energy by bremsstrahlung, while those below it lose energy mainly through ionization [90]. Thus, the total energy loss of electrons is obtained as $dE/dX = -\alpha(E) - E/X_0$. In air, the radiation length is $X_0 \simeq 37\,\mathrm{g\,cm^{-2}}$. For the case $E \gg E_c$, the radiative term dominates and energies decrease exponentially with depth. In a minimal branching picture (Heitler), interactions "split" after a depth λ_e, namely, one particle of energy E yields two particles of energy $E/2$ [94]. After depth $X = n\lambda_e$, the particle count and typical energy are $N(X) = 2^{X/\lambda_e}$ and $E(X) = E_0/2^{X/\lambda_e}$,

producing secondary particle cascade growth. Multiplication stops when $E(X) \approx E_c$, giving the scalings $N_{\max} = E_0/E_c$ and $X_{\max}^{\mathrm{EM}} \sim \lambda_e \ln(E_0/E_c)$ [90, 19]. The size of the cascade at maximum grows in direct proportion to the primary energy E_0, whereas the position of the maximum shifts only slowly, increasing like the logarithm of E_0, which describes the overall shower evolution. This leads to the well-known result for photon primaries,

$$\langle X_{\max}^{\mathrm{EM}} \rangle \simeq X_0 \ln\big(E_0/E_c\big) + \frac{1}{2}, \tag{1.6}$$

while the scaling of the shower size, $N_{\max} \sim E_0/E_c$, remains unchanged. The lateral development is governed by multiple Coulomb scattering, with a characteristic Molière radius $r_1 = (21\,\mathrm{MeV}/E_c)X_0 \simeq 9.3\,\mathrm{g\,cm^{-2}}$, which translates to about 80 m at sea level and increases with altitude [90].

A hadron entering the atmosphere collides after roughly one interaction length λ_{int} and produces many secondaries, predominantly pions. Neutral pions decay promptly to photons and seed EM sub-showers; charged pions mostly re-interact while energetic, and later decay to muons and neutrinos once their energy falls below a characteristic decay energy E_{dec}. The Heitler–Matthews model assumes each hadronic interaction yields n_{tot} pions, with n_{ch} charged and the rest neutral [93]. Because π^0 decay transfers energy to the EM channel each generation, after n generations, the energies are

$$E_{\mathrm{had}} = \left(\frac{2}{3}\right)^n E_0, \quad E_{\mathrm{EM}} = \left[1 - \left(\frac{2}{3}\right)^n\right] E_0, \tag{1.7}$$

so about 90% of the primary energy ultimately feeds the EM component after $n \sim 6$ generations [93, 90, 19]. Assuming that the first hadronic collision transfers its energy only into electromagnetic sub-showers through neutral pion decay, the position of the hadronic shower maximum can be estimated as

$$X_{\max}^{\mathrm{had}}(E_0) \;\approx\; \lambda_{\mathrm{int}} + X_{\max}^{\mathrm{EM}}\left(\frac{E_0}{2\,n_{\mathrm{tot}}}\right) \;\sim\; \lambda_{\mathrm{int}} + X_0 \ln\left(\frac{E_0}{2n_{\mathrm{tot}}E_c}\right), \tag{1.8}$$

where λ_{int} represents the length of hadronic interactions. The total number of muons produced in the air shower at ground level (on average) follows from the number of charged hadrons that reach E_{dec} before decaying:

$$N_\mu \;\simeq\; \left(\frac{E_0}{E_{\mathrm{dec}}}\right)^{\alpha}, \quad \alpha = \frac{\ln n_{\mathrm{ch}}}{\ln n_{\mathrm{tot}}} \approx 0.82\text{–}0.94, \tag{1.9}$$

where E_{dec} is the energy below which charged pions prefer to decay rather than re-interact. So that N_μ rises faster than linearly in E_0 on a log–log scale, with α reflecting hadronic multiplicities. Laterally, hadrons emerge with nearly energy-independent transverse momenta $\langle p_\perp \rangle \sim 0.35 - 0.4$ GeV, which makes the hadronic and muonic components broader on the ground than the EM one [90]. Most ground-level muons come from low-energy pion decays, making N_μ sensitive to low-energy interaction modeling.

The interactions and decays of charged pions and kaons feed the muonic component, which provides about 90% of the muons observed at ground [90]. Other photon-induced processes, such as photoproduction or muon-pair production, add only a small contribution [90, 19]. Muons lose little energy while traversing the atmosphere, so most survive to reach the ground; in very inclined showers ($\theta > 65°$) the electromagnetic part is largely absorbed, leaving mainly muons and their decay electrons [90]. Because they penetrate deeply, even into underground detectors, muons are a powerful probe of the primary cosmic ray. The total number of muons grows with energy and also depends on the primary mass. For air-shower development, a nucleus with mass A and energy E_0 can be approximated as

A independent nucleons of energy E_0/A [19]. This "superposition" approach yields simple scaling rules for inclusive observables:

$$\begin{aligned} N_{e,\max}^{(A)}(E_0) &\approx A\, N_{e,\max}^{(p)}(E_0/A) \approx N_{e,\max}^{(p)}(E_0), \\ \langle X_{\max}^{(A)}(E_0)\rangle &= \langle X_{\max}^{(p)}(E_0/A)\rangle, \\ N_{\mu}^{(A)}(E_0) &= A^{1-\alpha}\, N_{\mu}^{(p)}(E_0), \end{aligned} \tag{1.10}$$

where the superscripts (p) and (A) refer to showers initiated by a proton and by a nucleus with mass number A, respectively. For a fixed total energy, heavier primaries produce shallower showers and a larger number of muons, because the primary energy is distributed among many nucleons, resulting in a lower center-of-mass energy per nucleon and, consequently, an earlier shower development. Iron showers reach a maximum 80–100 $\mathrm{g\,cm^{-2}}$ higher in the atmosphere than proton showers of the same energy and contain $\sim 40\%$ more muons [19]. The superposition picture works well for mean quantities like $\langle X_{\max}\rangle$ and N_μ, but it does not capture correlation observables or higher moments that depend on nuclear breakup details [19]. In the superposition approximation, the number of particles at the shower maximum depends solely on the primary energy, not on the mass of the nucleus [95]. The position depth of the maximum, however, changes with mass and is expressed as a mass-dependent development with observable depth differences, i.e.,

$$X_{\max} \propto \lambda \ln\!\left(\frac{E_0}{A\,E_c}\right), \tag{1.11}$$

where E_c is the critical energy defined at the point where ionization and bremsstrahlung losses are equal (as mentioned previously). The expression (1.11) above is a form of the elongation-rate theorem [19]. As a result, showers from heavy nuclei develop faster and reach a maximum at a higher altitude in the atmosphere than proton showers of the same energy, although the dependence is only logarithmic. The first interaction of a heavy nucleus also happens very early, for iron, $\lambda_{\mathrm{int}} \sim 2.3\ \mathrm{g\,cm^{-2}}$ [19]. With increasing energy, the nucleon–air cross section rises, so the first interactions take place even higher in the atmosphere. The longitudinal and lateral development of these air showers is subsequently used to reconstruct the energy, mass, and arrival direction of the primary cosmic ray.

1.2.4 Muon Puzzle

For extensive air showers (EAS) produced by high-energy cosmic rays, the number of muons observed at the ground is larger than predicted by simulations. This discrepancy is known as the "muon puzzle" (see Figure 1.6). It has important consequences for hadronic interaction models and for systematics in reconstructing the EAS, since it points to missing physics at energies beyond the reach of accelerators [96, 97]. Underestimating the muon component leads to an incorrect energy reconstruction and mass composition, especially in inclined showers where muons dominate the shower structure. Inaccurate muon modeling may also bias the reconstructed shower core, primary energy, and arrival direction of cosmic rays, thereby affecting the accuracy of anisotropy studies and the determination of dipole amplitudes and phases.

Experiments measure N_μ with $\sim 10\%$ uncertainty, yet the N_μ-based composition band is 2.5–4 times wider, showing that hadronic model systematics dominate, while different air-shower codes agree at the $\sim 5\%$ level [98, 97]. Even LHC-tuned models predict fewer muons than observed [99]. The muon deficit was first identified by the Pierre Auger Observatory through a nearly model-independent hybrid analysis [100, 101], and

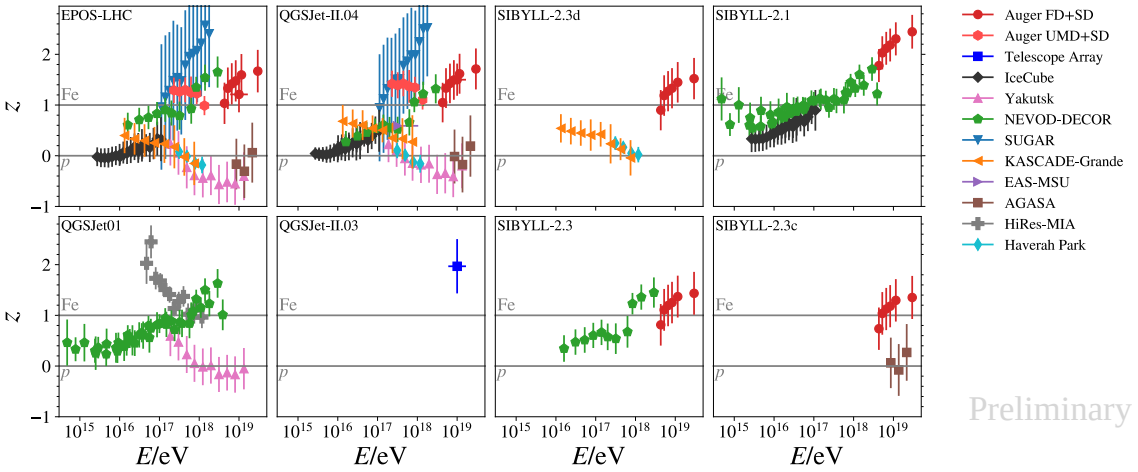


Figure 1.6: Comparison of muon density measurements from different experiments using the z-scale (see [96]), dependent on the hadronic interaction model. It shows data after energy-scale cross-calibration, except for KASCADE-Grande and EAS-MSU, which are included for reference. Also shown are z-values expected for a mixed composition based on optical measurements (band) and the GSF model (dashed line). Plot taken from [96]

the Telescope Array later reported consistent evidence using a more model-dependent surface-detector approach [102]. Auger has since extended its study to shower-to-shower fluctuations [103], while IceCube observed a similar trend at lower energies [104]. A detailed meta-analysis by the WHISP group further demonstrated a consistent muon excess across experiments, reaching 8σ significance above 10 PeV [96]. Comparisons between measured data and MC simulations are often performed using the z-scale [96, 105],

$$z = \frac{\ln\langle N_\mu^{\mathrm{det}}\rangle - \ln\langle N_{\mu,\mathrm{p}}^{\mathrm{det}}\rangle}{\ln\langle N_{\mu,\mathrm{Fe}}^{\mathrm{det}}\rangle - \ln\langle N_{\mu,\mathrm{p}}^{\mathrm{det}}\rangle}, \tag{1.12}$$

where N_μ^{det} is the muon density estimate observed in the detector, while $N_{\mu,\mathrm{p}}^{\mathrm{det}}$ and $N_{\mu,\mathrm{Fe}}^{\mathrm{det}}$ are the simulated muon density estimates for proton and iron showers, respectively, after full detector simulation, as shown in Figure 1.6. It reveals large discrepancies between experiments above 10 PeV and up to the highest energy.

Within the LHC energy range, the discrepancy grows smoothly with energy and appears already at $\sqrt{s} \sim 8$ TeV [97]. The most natural explanation is a change in secondary particle production that reduces the neutral pion fraction, increasing the hadronic over electromagnetic energy and thus raising the muon yield, while leaving X_{max} consistent [97]. At the LHC, an enhancement of strangeness production has been observed in high-multiplicity events [99], which could provide such a mechanism if it also occurs in the forward region that dominates air-shower development [106, 107]. Dedicated forward measurements and future p–O collisions at the LHC [108] are crucial to test this. If the puzzle is solved, the composition uncertainties (the N_μ bands) will be reduced by factors of 2.5–4 and improve predictions of atmospheric lepton fluxes, which form the main background for neutrino observatories, will improve [97].

The discrepancy indicates that present hadronic interaction models (EPOS-LHC, QGSJET, SIBYLL) may not correctly describe secondary production above accelerator energies [109, 97]. In particular, current models may channel too much energy into neutral

pions, while in reality more baryons, kaons, or strange hadrons could be produced [97]. This gap limits accurate modeling of particle production at high energies. These soft-QCD processes occur in the forward region, which is poorly constrained by existing data [97]. New collider measurements are therefore essential to refine the models. Although exotic scenarios beyond the Standard Model have been suggested, such as new particles decaying into muons, current evidence favours incomplete knowledge of soft hadronic interactions as the main cause [97]. Upcoming upgrades such as AugerPrime [21] and next-generation observatories like GRAND [32] will enable more accurate measurements of $X_{\max}$ via radio techniques for mass-composition studies, while LHC data, particularly from p–O collisions, will directly test the physics underlying the muon puzzle.

1.2.5 Composition

The composition of cosmic rays can be determined through direct and indirect detection methods. Direct measurements are performed by satellite or balloon-based detectors, such as PAMELA [60], AMS-02 [56], or CREAM [28], which can directly measure the charge (q) and mass of individual cosmic ray particles (measure the rigidity ($R = p/q$), where p is the particle momentum). For high-energy cosmic rays, where direct detection becomes challenging, ground-based air shower arrays, such as KASCADE-Grande [110], Pierre Auger Observatory [111, 21], and IceTop [70], infer the mass composition through measurements of extensive air showers (EAS).

For the direct measurement, typically in lower energies, cosmic-ray abundances and solar abundances, both show the odd–even pattern: even-Z nuclei are more common because they are more tightly bound. Two clear differences between them stand out. First, heavy nuclei ($Z > 1$) are far more abundant relative to protons in cosmic rays than in the solar system. This may reflect injection physics (hydrogen is harder to ionize and inject) or a true source-composition difference; the exact reason is not settled [19]. Second, several nuclei that are rare in stellar nucleosynthesis, Li, Be, B and the sub-Fe group, Sc, Ti, V, Cr, and Mn are strongly enhanced in cosmic rays [19]. These are secondaries made by spallation of heavier primaries (notably C, O, Fe) on interstellar gas. Using spallation cross sections, one infers the average grammage traversed by most cosmic rays to be $X \sim 5\ \mathrm{g\,cm^{-2}}$. For a disk density $\rho_N \sim 1\ \mathrm{cm^{-3}}$, this corresponds to a path length

$$l \sim X/(m_p \rho_N) \approx 3 \times 10^{24}\ \mathrm{cm} \simeq 10^3\ \mathrm{kpc}, \tag{1.13}$$

which is obtained based on propagation and extracted using B/C measurements of AMS-02. Since this greatly exceeds the disk half-thickness (~ 0.1 kpc), cosmic rays cannot move ballistically. Instead, their trajectories are strongly bent by the turbulent Galactic magnetic field and effectively follow a random walk. This indicates that cosmic-ray confinement occurs through diffusion, with particles scattering many times before leaving the Galaxy [19]. Low-energy cosmic rays are more likely to interact with matter because their small Larmor radius in the Galactic magnetic field leads to longer confinement times and higher column densities. In contrast, high-energy cosmic rays, with energies from the TeV up to the PeV range, are less confined and escape the Galaxy more quickly, reducing the effective column density, as expected from rigidity scaling.

Air-shower experiments cannot measure the mass of each primary directly, but they can use observables that depend on composition. The most important are the depth of the shower maximum, $X_{\max}$, and the number of muons at ground. In the superposition picture, a nucleus of mass A and energy E_0 behaves like A nucleons of energy E_0/A. As a result,

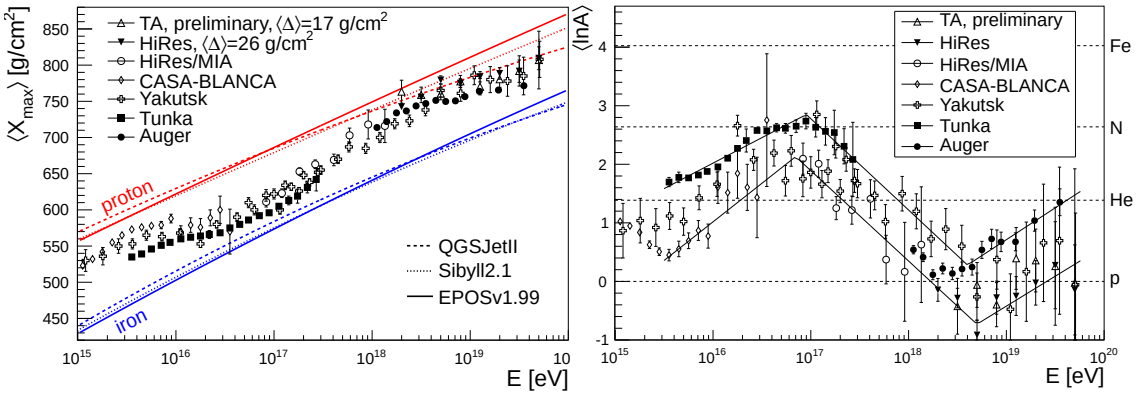


Figure 1.7: Left: measurements of $\langle X_{\max} \rangle$ with non-imaging Cherenkov detectors. The depth of shower maximum is shown as a function of primary energy, compared to the model simulation for protons and iron. Right: Mean logarithmic mass as a function of energy from $X_{\max}$ data. Both figures are adapted from [98], and the paper demonstrates that $\langle X_{\max} \rangle$ depends on the composition through the mean logarithmic mass, $\langle \ln A \rangle = \sum_i f_i \ln A_i$.

heavy primaries interact earlier and develop faster, so their showers reach the maximum in the higher atmosphere. Proton showers, by contrast, penetrate deeper and fluctuate more, while heavy nuclei show smaller fluctuations because they act like many nucleons together, and consequently produce narrower profiles [19]. These features lead to the scaling laws of Eq. (1.11), which predict smaller $X_{\max}$ and more muons for heavier primaries at the same energy as expected. The average depth $\langle X_{\max} \rangle$ is measured with fluorescence telescopes above EeV energies and with air-Cherenkov detectors at lower energies [19], see the left panel of Figure 1.7. To estimate the mass composition, data are compared with simulations with different hadronic interaction models for protons and iron, which provide reference curves. The measured $\langle X_{\max} \rangle$ is then interpolated between these two extremes to obtain the mean logarithmic mass, as follows the relation:

$$\langle \ln A \rangle = \frac{\langle X^{\mathrm{p}}_{\max} \rangle - \langle X^{\mathrm{data}}_{\max} \rangle}{\langle X^{\mathrm{p}}_{\max} \rangle - \langle X^{\mathrm{Fe}}_{\max} \rangle} \ln 56, \tag{1.14}$$

where $\langle X^{\mathrm{p}}_{\max} \rangle$ and $\langle X^{\mathrm{Fe}}_{\max} \rangle$ denote the model predictions for proton and iron primaries, respectively. This relation provides a model-dependent estimate of the mean logarithmic mass of primary cosmic rays [112, 98], as illustrated in the right panel of Figure 1.7. Measurements indicate that the average mass of CRs changes with energy. Between the knee at about 10^{15} eV and the second knee near 10^{17} eV, the composition becomes gradually heavier, consistent with lighter nuclei cutting off first in Galactic sources. From the second knee to the ankle at around 3×10^{18} eV, the trend reverses and the composition appears lighter, which may point to the growing contribution of an extragalactic component dominated by protons and helium. Above the ankle, some data suggest another increase of mass, but the uncertainties are still large, and the picture is not yet settled [19]. If confirmed, such a rise of $\langle \ln A \rangle$ reflects a Peters cycle of an extragalactic component exhibiting a rigidity cutoff near 3×10^{18} V [19].

The mass composition also provides clues about the sources of cosmic rays. For Galactic cosmic rays up to the knee, the presence of heavy primaries (e.g., Fe) is consistent with supernova remnants as sources, though injection biases and propagation also shape abundances [3]. The He/p ratio and element-dependent spectra constrain rigidity-dependent

acceleration and escape in shocks [3]. At higher energies, the evolution of composition helps distinguish the fading Galactic component from the emerging extragalactic one. At ultra-high energies, the observed mix is set by the sources' maximum rigidity and by propagation effects, namely, photo-pion production for protons and photo-disintegration of nuclei on the CMB/EBL [12, 113, 114].

1.2.6 Energy Spectrum

The cosmic-ray energy spectrum describes the flux of cosmic rays as a function of their energy. It spans a vast range of energies, from 10^9 eV to beyond (10^{20}) electron volts (eV). The spectrum is fundamental to cosmic-ray physics, as it provides critical insights into the origin, acceleration mechanisms, and propagation of cosmic rays through the galaxy and beyond [3]. The spectrum has been measured by various experiments across a wide range of energies, consistently demonstrating a steeply falling behavior above the GeV range [23, 115, 116, 117]. It can be well-approximated by a power-law function over many decades in energy, reflecting the underlying physical processes governing CR acceleration and propagation [90]. The differential CR energy spectrum can be obtained by:

$$J(E) = \frac{dN}{\mathrm{d}E\,\mathrm{d}A\,\mathrm{d}t\,\mathrm{d}\Omega} \propto E^{-\gamma}, \tag{1.15}$$

where E is the energy, N is the total number of cosmic-ray particles at energy E, A is the effective area of the observatory, Ω is the solid angle covered on the sky, and γ is the spectral index, typically around 2.7 for energies above a few GeV and up to about 4.2 at the highest energies. The power-law shape of the cosmic-ray spectrum is modified by several notable features: the "knee", the "ankle", and the ultra-high-energy cutoff, as Figure 1.15 shows. The combined fits of the scaled mass-dependent cosmic-ray energy spectrum from direct and indirect measurements are shown in Figure 1.8.

When Eq. (1.15) is plotted as $E^2J(E)$, a break in the power-law curve is seen at $E \simeq 3 \times 10^{15}$ eV, and this feature is called the "knee", which marks a steepening of the spectrum [118]. At this energy, the index changes from 2.7 to 3.1. The knee was first observed with an air-shower array by Kulikov and Khristiansen [50]. The KASCADE experiment provided the first high-precision measurements of the all-particle cosmic-ray spectrum in the knee region [23] and clarified the spectral behaviour. The results showed that the knee at $E \simeq 3 \times 10^{15}$ eV is caused primarily by a steepening of the light components (protons and helium), while heavier nuclei exhibit knee-like features at proportionally higher energies, consistent with a rigidity-dependent cutoff [22]. In addition, later measurements by other air-shower arrays are consistent with a rigidity-dependent behaviour of the knee [98].

Several air-shower experiments have measured the spectrum in the range above the knee with sufficient precision to reveal additional structure beyond a single power law. Results from KASCADE-Grande [119], Tunka-133 [120], and IceTop [63] show that the spectrum deviates from a simple index of $\gamma \simeq 3.0$, indicating further features (additional spectral breaks) in this transition region [19]. At $\sim 10^{17}$ eV, the steepening of the spectrum was found [121, 122]. KASCADE-Grande extended the energy range up to 10^{18} eV, revealing a "second knee" in the spectrum at $E \sim 10^{17}$ eV and providing evidence for the transition from Galactic to extragalactic cosmic rays [116, 119]. At 10^{18}-10^{19} eV, which is the start of ultra-high energy, the spectrum becomes flat, known as the "ankle", and above 7×10^{19} eV a suppression of the flux is observed [90]. HiRes Fly's Eye provided the first clear measurement of the ankle in the UHECR spectrum at $\sim 3 \times 10^{18}$ eV [123]. In 2020, Auger confirmed the flattening of the spectrum near 5×10^{18} eV at the "ankle" [117].

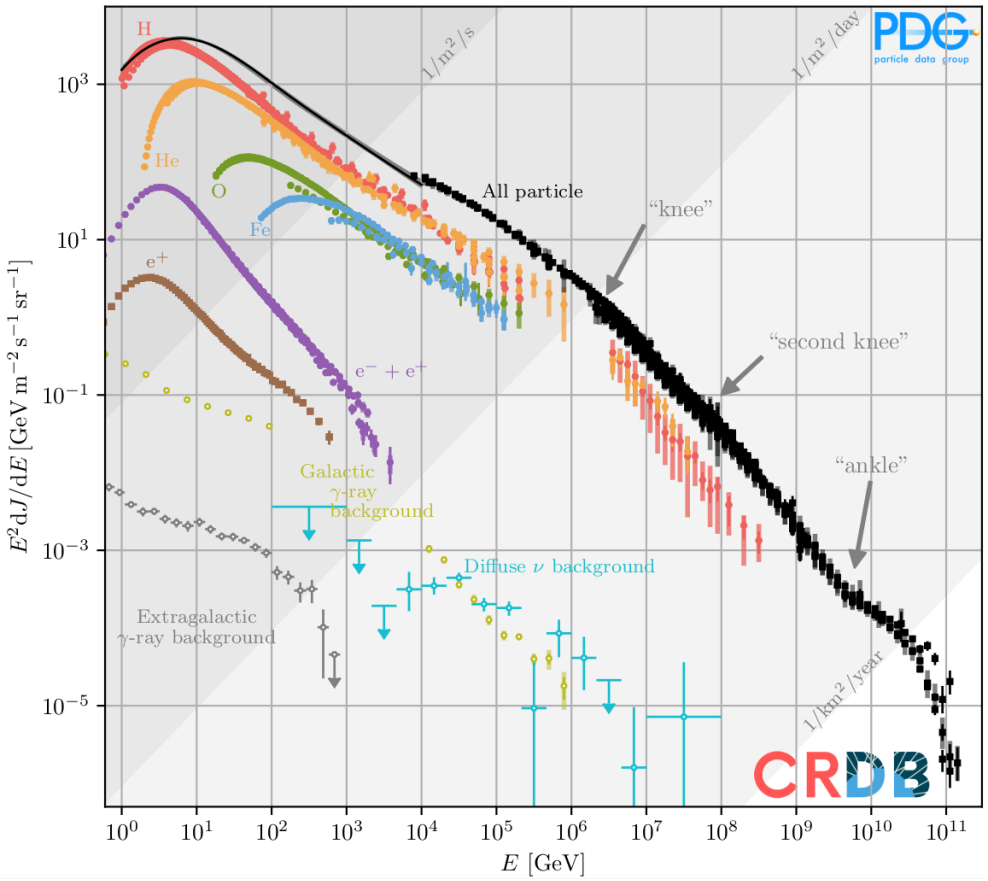


Figure 1.8: Energy-weighted cosmic ray spectrum above 10 TeV, illustrating the features: the knee, the second knee, and the ankle. The knee at 3 PeV marks a spectral steepening, dominated by Galactic cosmic rays. The second knee around EeV causes further steepening, while the ankle beyond EeV marks a spectral flattening. The diffuse γ-ray flux and the diffuse neutrino background are also included. Figure taken from [118]

.

Above $10^{19.5}$ eV, the spectrum exhibits a suppression, often associated with the GZK cut-off [54, 55]. In 1963, John Linsley at the Volcano Ranch array reported the detection of an air shower corresponding to a primary energy of about 10^{20} eV, which was the first experimental evidence of an ultra-high-energy cosmic ray (UHECR) at this energy [49]. HiRes was the first experiment to find a sharp suppression in the flux of ultrahigh-energy cosmic rays at about 6×10^{19} eV, consistent with the expected cutoff, while the ankle of the spectrum was also seen at about 4×10^{18} eV [115, 67], thereby confirming the expected high-energy features. Measurements by Auger in 2008 showed that the spectral index changes from $\gamma = 2.69$ in the range 4×10^{18}–4×10^{19} eV to $\gamma = 4.2$ above 4×10^{19} eV, also providing clear evidence of flux suppression consistent with the GZK effect [67]. More recently, in 2020, Auger confirmed the steepening of the spectrum at around 5×10^{19} eV finally [117], further supporting these observational trends. Moreover, in 1991, Fly's Eye (HiRes) recorded an extensive air shower whose primary energy was reconstructed at 3.2×10^{20} eV [124] (far exceeding the threshold expected for the GZK limit), so-called "Oh-My-God" event, which remains the highest-energy cosmic-ray particle ever reliably detected and provided the first direct evidence that nature can produce particles with energies far

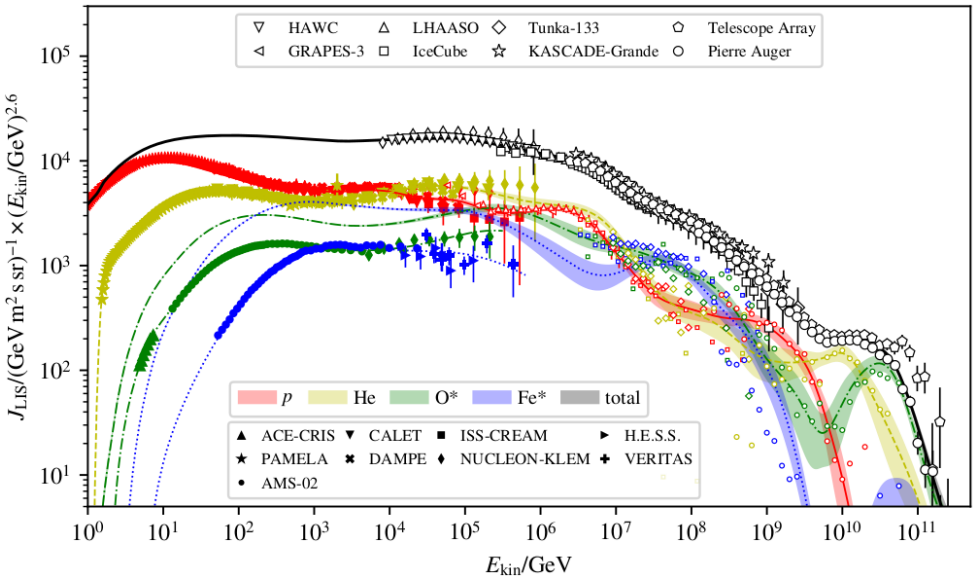


Figure 1.9: Combined fit of the scaled cosmic-ray energy spectrum from direct and indirect measurements. Colored data points and fit curves correspond to different mass groups. Filled markers show direct observations, while open markers indicate indirect air-shower observations. Figure adapted from [126].

beyond terrestrial accelerators, demonstrating the end of the cosmic-ray spectrum.

This power-law behavior arises from shock acceleration at the sources and from diffusive escape in turbulent Galactic magnetic fields, which steepens the spectrum during propagation [3, 19]. The knee is typically associated with a rigidity-dependent limit of Galactic accelerators (and possibly changes in propagation), where particles are no longer accelerated efficiently [98]. The ankle is often interpreted as the transition to an extragalactic population that overtakes the Galactic component [113, 12], as mentioned at the end of Section 1.2.5, further supporting this interpretation. At the highest energies (above 5×10^{19} eV), energy losses in background radiation shape the spectrum: protons lose energy by photo-pion production ($\sim 20\%$) and Bethe–Heitler pair production on the CMB/EBL, while nuclei are photo-disintegrated; a contribution from finite source rigidity limits is also possible [115, 125]. Detailed theoretical models of acceleration mechanisms and propagation effects can explain the observed features of the cosmic-ray spectrum, especially across these transitions, which will be discussed in the next sections.

1.3 Anisotropy

The arrival directions of cosmic rays on Earth are almost isotropic, but small anisotropies exist at the level of about 10^{-4}–10^{-2} from TeV energies up to above 50 EeV, increasing with energy. These appear as tiny modulations in the otherwise nearly uniform distribution of arrival directions. Early searches looked in the solar frame and found the Compton–Getting effect, a small diurnal variation caused by Earth's orbital motion. This indicated that cosmic rays are almost isotropic, with modulations at the $\sim 10^{-4}$ level [127, 128, 129]. The first hints of large-scale anisotropy were already reported in the 1930s, when early experiments found small deviations from an isotropic distribution of arrival directions [130].

However, the detectors then had limited stability, low statistics, and poor control of atmospheric and instrumental effects, so the measurements carried large uncertainties and did not allow firm conclusions about the presence or structure of a genuine cosmic-ray anisotropy.

With the development of larger ground arrays, systematic studies of small anisotropy signals became possible in the 1950s through data collected by large underground muon detectors and extended ground-based arrays [5]. From the 1970s to the 1990s, the focus shifted to the sidereal frame, since any steady anisotropy observed there should arise from Galactic transport and sources [131, 132]. Clear evidence of multi-TeV anisotropy was later reported by the Tibet ASγ experiment, which identified the "tail-in" and "loss-cone" features of TeV cosmic rays. The "tail-in" region corresponds to a broad excess, likely caused by cosmic rays streaming along the local interstellar magnetic field, while the "loss-cone" region represents a deficit resulting from magnetic-field scattering or reflection effects [15]. These results were confirmed by Milagro [133], Super-K [134] and ARGO-YBJ [135] in the northern hemisphere, revealing large- and medium-scale structures at the 10° level with amplitudes of 10^{-3}. HAWC [136, 16] and LHAASO [25, 137] further improved sensitivity at multi-TeV to PeV energies, showing the energy dependence of the anisotropy at the 10^{-4}–10^{-3} level. In the southern hemisphere, IceCube and IceTop cover the TeV–PeV range [138, 139, 140], and have observed anisotropy of about 10^{-4} to 10^{-3} below 1 PeV, with an approximate dipole pattern [141]. These structures evolve with energy, roughly consistent with anisotropic diffusion and local magnetic-field effects [5]. Figure 1.10 shows an example of full-sky anisotropy, which combines HAWC and IceCube data at a median energy of 10 TeV. In these energy ranges, different experiments have reported dipoles with large discrepancies in both sidereal amplitudes and phases below the PeV range, which may arise from differences in hadronic interaction models used for energy and direction reconstruction, as well as from differences in the energy-sample distributions of the median energies (excluding the effect of partial-sky coverage). Therefore, a global fit of these dipoles is presented in Chapter 5. Moreover, in the PeV range, KASCADE-Grande also reported dipole anisotropies [87], but with amplitudes and phases clearly shifted relative to those measured by IceTop in the same energy interval [142]. This mismatch indicates that the observed anisotropy at PeV energies is not yet consistently reproduced across experiments, and is therefore one of the motivations of Chapter 6. At even higher energies, above 1 EeV, Auger [88] and TA [102] have searched for dipole and large-scale patterns, and linked anisotropy to the transition from Galactic to extragalactic cosmic rays. Besides, Auger has reported a significant dipole above 8 EeV at the $\sim 6\%$ level, which is consistent with an extragalactic origin at the highest energies [88, 117, 89].

To explain the origin of cosmic-ray anisotropies, one has to connect them to the distribution of sources and their propagation through the Galaxy. At low energies ($\lesssim 10$ GeV), a kinematic dipole appears from the Compton–Getting effect [127], i.e., $\Delta I/\langle I\rangle = (\gamma+2)(v/c)\cos\theta$, with $\gamma \simeq 2.7$ the cosmic-ray spectrum index, θ is the angle between the CR arrival direction and the solar system motion. For the solar system velocity $v \simeq 220$ km/s, the expected amplitude is $\sim 0.35\%$, giving an excess near 290°–340° and a deficit near 110°–160° [143]. At sub-PeV energies, the anisotropy is likely dominated by local effects, such as the heliosphere and nearby supernova remnants (SNRs), which are thought to be primary accelerators of Galactic cosmic rays [3]. In this regime, cosmic-ray trajectories are strongly randomized by the Galactic magnetic field, resulting in only a small anisotropy. The dipole anisotropy is of particular interest in this energy range because it reflects large-scale directional patterns, often linked to nearby sources or magnetic structures [5].

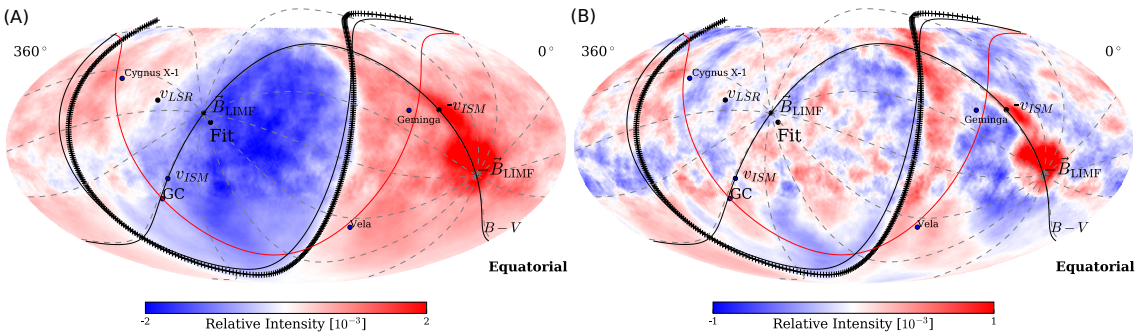


Figure 1.10: Combined data of IceCube and HAWC at a median energy of 10 TeV. (A) Relative intensity of cosmic rays at 10 TeV median energy and (B) corresponding small angular-scale anisotropy. Figures taken from [16]

In theory, if cosmic rays come from sources that are not uniformly distributed, there should be a gradient in their density, which, through diffusion, produces a net flux and a certain dipole amplitude [5]. This flux would appear as a dipole in the sky, pointing back toward the dominant sources [5]. Several studies have suggested that such a signal could reveal the location of nearby or major sources [15, 16]. However, observations so far do not show this clearly, and the dipole seen at TeV–PeV energies is much smaller than what simple diffusion models predict [5]. This discrepancy is known as the "cosmic-ray anisotropy problem" [144, 5]. However, at higher energies, particularly for ultra-high-energy cosmic rays (UHECRs), the deflection by magnetic fields becomes weaker, since their velocities approach the speed of light c, and cosmic rays tend to follow straighter paths from their sources, thereby improving directional sensitivity, making it easier to detect anisotropies [145]. The observations of the Pierre Auger Observatory suggest that the sources of UHECRs are extragalactic, likely associated with nearby large-scale structures such as AGNs and galaxy clusters within a few hundred megaparsecs [111].

1.4 Propagation

Cosmic-ray propagation involves complex processes that influence their energy, direction, and composition before they reach Earth. These processes are essential for interpreting cosmic-ray spectra and anisotropies observed on Earth, as they are shaped by the interactions cosmic rays undergo during their journey through the Galaxy [5, 3]. From the 1960s to the present day, people have developed many methods and models to explain the abundances, spectra, and anisotropies of cosmic rays on Earth.

1.4.1 Leaky-Box Model

Early work noticed that Li, Be, and B were far more abundant in cosmic rays than in the general elemental composition of the Universe. These light elements were understood as spallation products of heavier nuclei like C, N, and O during their passage through interstellar matter [146]. The first quantitative attempts to explain this were made by Appa Rao and Kaplon [147] and by Balasubrahmanyan et al. [148], who assumed that cosmic rays crossed a fixed thickness of matter, similar to a slab, with little spread in path length. Such models could reproduce either the (Li+Be+B)/(C+N+O) ratio or the CNO

spectra at low energy, but not both together, because ionization losses at low energies flattened the spectra too strongly [146].

To resolve this, the leaky-box model was proposed in the mid-1960s [149, 146]. Instead of assuming a fixed slab thickness, it introduced a distribution of path lengths (often exponential), allowing one to average over different traversals of matter and obtain realistic predictions. Two early tests illustrated the model: the effect of the newly discovered $\sim 3\,\mathrm{K}$ cosmic microwave background on the electron spectrum [149], and the Li, Be, B abundances relative to C, N and O [150]. In this framework, the Galaxy is treated as a uniform box where cosmic rays are injected into a uniform gas and radiation field, propagate freely, and then escape after a mean time $\tau_{\rm esc}(E)$. Diffusion and convection are not explicitly included, but are approximated by a single leakage term. The grammage $X(E)$, the column density traversed, naturally appears in this picture and decreases with rigidity, explaining secondary-to-primary ratios [150, 151]. Thus, particles inside the box have a constant probability per unit time to escape, $\tau_{\rm esc}^{-1} \ll c/h$, and follow

$$V \cdot \nabla N_i(E,\mathbf{x}) - \nabla \cdot \big[D(r)\,\nabla N_i(E,\mathbf{x})\big] \;\rightarrow\; N_i/\tau_{\rm esc}, \tag{1.16}$$

which is valid only when $c\,\tau_{\rm esc} \gg h$, that means in the *leaky box* model, cosmic-ray particles move at nearly the speed of light and have a small probability of escaping each time they reach and cross the boundary of the Galaxy disk [19]. The survival probability after a time t is, then $\exp(-t/\tau_{\rm esc})$. In this picture, $\tau_{\rm esc}$ is the mean residence time and $\lambda_{\rm esc} = \rho\,\beta c\,\tau_{\rm esc}$ is the average amount of matter traversed.

A more general treatment, the *weighted slab technique*, was later introduced to allow for a broader path-length distribution and to include possible re-acceleration during propagation [152]. In the equilibrium regime, meaning that the density $N(E,\mathbf{x})$ becomes time-independent and reaches a steady state, the leaky-box equation (1.16) reduces to

$$\frac{N_i(E)}{\tau_{\rm esc}(E)} = Q_i(E) - \left[\frac{\beta c\rho}{\lambda_i} + \frac{1}{\gamma\tau_i}\right] N_i(E) + \frac{\beta c\rho}{m_p}\sum_{k\geq i}\sigma_{i,k}\,N_k(E), \tag{1.17}$$

where m_p denotes the proton mass, and $\sigma_{i,k}$ is the spallation cross-section for the reaction $k \rightarrow i$. This form is used for describing primary and secondary nuclei, which keep the same energy per nucleon during fragmentation [19].

The boron-to-carbon (B/C) ratio can be estimated using Eq. (1.17), assuming that no boron from the sources, and the decay term is neglected since it is stable [19]. The source term is mainly from carbon and oxygen and $N_{\rm B} \simeq N_{\rm O}$, the B/C ratio becomes [19]:

$$\frac{N_{\rm B}}{N_{\rm C}} \;\simeq\; \frac{\lambda_{\rm esc}(E)}{1+\lambda_{\rm esc}(E)/\lambda_{\rm B}} \cdot \frac{\sigma_{\rm C\to B}+\sigma_{\rm O\to C}}{m_{\rm p}}, \tag{1.18}$$

where the escape length $\lambda_{\rm esc}(E) = \rho\,\beta c\,\tau_{\rm esc}(E)$. For boron production, typical values are $\sigma_{\rm C\to B} \approx 73$ mb, $\sigma_{\rm O\to B} \approx 30$ mb and $\lambda_{\rm B} \approx 7.1\ \mathrm{g\,cm^{-2}}$, but in precision fits one should use updated cross-section sets [153]. Using modern B/C data from AMS-02 [154], the inferred escape length decreases with rigidity (R) and is well described at high R by

$$\lambda_{\rm esc}(R) \;\approx\; 19\,\beta^3\left(\frac{R}{3\ \mathrm{GV}}\right)^{-\delta}, \tag{1.19}$$

with $\delta \simeq 0.40$ at high R ($\gtrsim$tens of GV) [154], $\lambda_{\rm esc} \propto R^{-\delta}$. At low R, with $R \lesssim 3$ GV, a common form is $\lambda_{\rm esc} = 19\,\beta^3\ \mathrm{g\,cm^{-2}}$. Solar modulation also shifts low-energy spectra between local interstellar space and near-Earth. For a primary nucleus P, the contributions

from fragmentation of heavier nuclei can be ignored, Eq. (1.17) reduces to

$$N_{\rm P}(E) = \frac{Q_{\rm P}(E)\,\tau_{\rm esc}(R)}{1+\lambda_{\rm esc}(R)/\lambda_{\rm P}}, \tag{1.20}$$

where ($\lambda_{\rm P} \gg \lambda_{\rm esc}$), which leads to $N_p(E) \simeq Q_p(E)\,\tau_{\rm esc}(R)$. Thus, if the observed spectrum follows $N(E) \propto E^{-(\gamma+1)}$, the source spectrum has to be steeper due to propagation,

$$Q(E) \propto E^{-\alpha}, \tag{1.21}$$

with $\alpha = \gamma + 1 - \delta \simeq 2.3$ when $\delta \simeq 0.40$ and $\gamma + 1 = 2.7$. For heavy nuclei like iron, where $\lambda_{\rm P} \simeq 2.3$ g cm^{-2}, the change from interaction-dominated to escape-dominated happens around 20 GeV/nucleon at low energies. The iron spectrum is observed to be flatter than protons and helium, in line with this expectation. The AMS-02 data [155, 154] extend B/C up to the TeV region and give $\delta \approx 0.4$ at high rigidity. This value lies between Kolmogorov (1/3) [156, 157] and Kraichnan (1/2) [158] turbulence predictions, with signs of a break in rigidity rather than a single slope over the full range.

The leaky-box framework quickly became the main and standard tool to interpret cosmic-ray propagation and secondary-to-primary ratios [159, 160, 152, 161, 162, 163, 164, 165]. Later, space-based missions such as PAMELA [60], AMS-02 [56], and DAMPE [31] provided high-precision measurements over a broad energy range. These data made clear that the simple leaky-box description was no longer sufficient, and motivating the use of modern transport and diffusion models [146].

1.4.2 Transport Model

The Ginzburg–Syrovatskii (G-S) equation, known as the "transport equation", was formulated by Ginzburg and Syrovatskii in the 1960s [166]. The transport equation describes the balance between sources, diffusion, convection, energy changes, and losses in the propagation of cosmic rays. However, most studies before the 1980s still employed the simpler leaky-box model. The reason was that the available data were limited, and simple relations in terms of escape time and grammage were enough to describe secondary-to-primary ratios [150, 151]. The leaky-box equations can be solved using basic algebra, whereas the full transport equation requires solving differential equations in space and momentum, which need powerful computing. An exception was for electrons, where energy losses and the discrete nature of sources made diffusion necessary [146], as studied by Shen and Mao [167], Cowsik and Lee [168].

The Galactic magnetic field is approximately 3 μG, roughly following the local spiral arm but with large fluctuations. Because the field is frozen into the ionized gas, motions of the plasma create turbulence and distort the field [19]. Energy input from stellar flares, pulsar winds, and supernova explosions maintains this turbulence active. As a result, charged particles scatter on these irregular fields and thus diffuse through the Galaxy [19]. The diffusion equation follows from particle conservation. The continuity equation is $\partial N/\partial t + \nabla \cdot \mathbf{J} = 0$, where $N(\mathbf{r}, t)$ is the particle density and $\mathbf{J}$ the flux. Assuming Fick's law (steady state), that particle flow follows the gradient of density, flux is obtained as $\mathbf{J} = -D(\mathbf{r})\nabla N(\mathbf{r}, t)$, where D is the diffusion coefficient. Substituting gives the diffusion equation in a static medium, $\partial N/\partial t = \nabla \cdot \big[D(\mathbf{r}), \nabla N\big]$. If the medium has a bulk velocity field $\mathbf{V}(\mathbf{r})$, the time derivative becomes a convective derivative, $\partial/\partial t \to \partial/\partial t + \mathbf{V} \cdot \nabla$, and the equation generalizes to a more complete form:

$$\frac{\partial N}{\partial t} + \mathbf{V} \cdot \nabla N - \nabla \cdot \big[D(\mathbf{r})\,\nabla N\big] = 0, \tag{1.22}$$

where $N_i(\mathbf{x}, E)$ is the density of type i particle at position $\mathbf{x}$ and energy E. The equation includes only diffusion and convection, without energy gains, losses, or explicit sources.

The cosmic-ray transport equation for a given particle species can be written in a general form. Let $\psi(\mathbf{r}, p, t)$ denote the cosmic-ray density of particles with momentum p at position $\mathbf{r}$ and time t. Its evolution is then described by [151]

$$\begin{aligned}\frac{\partial \psi(\mathbf{r},p,t)}{\partial t} &= Q(\mathbf{r},p,t) + \nabla\cdot\left(D_{xx}\nabla\psi - \mathbf{V}\psi\right) \\ &\quad + \frac{\partial}{\partial p}\left[p^2 D_{pp}\frac{\partial}{\partial p}\left(\frac{\psi}{p^2}\right)\right] - \frac{\partial}{\partial p}\left(\dot{p}\,\psi - \frac{p}{3}\,\psi\,\nabla\cdot\mathbf{V}\right) - \frac{\psi}{\tau_f} - \frac{\psi}{\tau_r},\end{aligned} \tag{1.23}$$

where $\psi(\mathbf{r}, p, t)$ related to the phase-space density by $\psi(p)\,\mathrm{d}p = 4\pi p^2 f(p)\,\mathrm{d}p$. The source term $Q(\mathbf{r}, p, t)$ accounts for primary injection by supernova remnants and other accelerators, as well as secondary contributions from spallation and decay. Spatial diffusion is described by $D_{xx}(\mathbf{r}, p)$, which reflects scattering on magnetic turbulence, while $\mathbf{V}$ represents a convection velocity that pushes particles away from the disk. Diffusive reacceleration appears as momentum-space diffusion with coefficient $D_{pp}(\mathbf{r}, p)$, and $\dot{p} \equiv dp/dt$ gives the continuous momentum gain or loss from processes such as ionisation and synchrotron cooling [169, 1, 151]. The last two terms τ_f and τ_r are the characteristic timescales for fragmentation and radioactive decay. In general, the source term Q can be viewed as a sum over many discrete injection events in space and time. Since the lifetime and size of individual accelerators are much shorter than the Galactic propagation scales, each source can be treated as instantaneous and point-like [167, 168]. This is written as

$$Q(t, E, \mathbf{r}) \;=\; \sum_j Q_j(E)\,\delta^3(\mathbf{r} - \mathbf{r}_s)\,\delta(t - t_s), \tag{1.24}$$

where $Q_j(E)$ is the injection spectrum of source j, located at position $\mathbf{r}_s$ and born at time t_s. The cosmic-ray flux observed at Earth then follows from a convolution of this source term with the Green's function G [167, 168]. For a set of burst-like sources, the flux can be expressed as an integral over the Green's function:

$$\phi(E, \mathbf{r}_\odot, t) \;=\; \frac{c}{4\pi}\sum_j \int dE_s\; G(t, E, \mathbf{r}_\odot \leftarrow t_s, E_s, \mathbf{r}_s)\; Q_j(E_s), \tag{1.25}$$

where G is the propagator, which gives the probability for a particle injected at $(\mathbf{r}_s, E_s, t_s)$ to be found at Earth $(\mathbf{r}_\odot, E, t)$. In steady state, the time derivative vanishes, and the equation is solved either analytically in simple geometries or numerically with codes such as GALPROP and DRAGON, with free-escape boundary conditions (e.g., $\psi(R_h, z, p) = \psi(R, \pm z_h, p) = 0$ for a halo radius R_h and half-height z_h). From the 1980s, diffusion models with a Galactic halo and reacceleration became standard for nuclei [170, 171]. In the late 1990s, numerical codes such as GALPROP [172, 173, 174] and later DRAGON [175, 176] made it possible to solve the full transport equation in 2D or 3D with realistic gas and radiation fields, reproduce secondary-to-primary ratios, and explain the observed spectrum below the ankle [173, 151]. With data from PAMELA [177], AMS-02 [57, 155, 154], and DAMPE [31], the transport equation is now the main framework, since it can describe nuclei, radioactive isotopes, electrons, antiprotons, and diffuse γ-rays together.

Diffusion models with spatially varying coefficients are more physically realistic than the leaky-box picture, although both can yield similar results. In a leaky-box model, the density is uniform inside the volume, while diffusion produces gradients and anisotropies. The diffusion coefficient D links the particle flux to the spatial gradient of density. By

including a source term, the particle conservation becomes $\dot{N} = -\nabla \cdot \mathbf{J} + Q(\mathbf{r}, t)$, i.e.,

$$\frac{\partial N}{\partial t} = \nabla \cdot (D \, \nabla N) + Q, \tag{1.26}$$

where only source and diffusion are included as in the standard transport equation Eq. (1.23). The solution of Eq. (1.26) takes the form of a Green's function,

$$G(\mathbf{r}, t) = \frac{1}{(4\pi D \, t)^{3/2}} \exp\left(-\frac{r^2}{4D \, t}\right), \tag{1.27}$$

which gives the probability to find a particle at $\mathbf{r}$ at time t if injected at the origin [19]. The mean 3D distance from the source is expressed as $\langle|\mathbf{r}|\rangle = \sqrt{6D \, t}$, where the diffusion coefficient D can be estimated from the mean displacement $\langle|\mathbf{r}|\rangle$ if the propagation distances at time t are known, for example, using the code CRPropa [178]. For a 1D diffusion (such as z, distance from the plane), one has $\langle|z|\rangle = \sqrt{4D \, t/\pi}$ [19].

A detailed description of cosmic-ray 1D diffusion is provided by [179], building on earlier work by Ginzburg and Syrovatskii [180]. A common picture is that cosmic rays move in a thin gaseous disk of height $h \sim 100$ pc, surrounded by a larger diffusive halo with scale H. Since the ionized ISM extends to about 1 kpc, it is reasonable to take H at least of this size. The time it takes a particle to diffuse out to a height H is $t_H \sim H^2/D$, where D is the diffusion coefficient [19]. At the halo edge ($z = H$), the particle density is set to zero, meaning free escape. One can then define an effective escape velocity, $v_D \sim H/t_H \sim D/H$. To connect with the leaky-box idea, one can average the disk gas density ρ_g over the whole halo [19]. This gives an effective mean density, $\rho_H = \rho_g h/H$ [19]. A cosmic ray moving at speed βc for a time t_H then crosses a grammage,

$$\lambda_{\rm esc} = \rho_g \, \beta \, c \, \frac{h \, H}{D} = 19\beta^3 \left(\frac{R}{3\,\mathrm{GV}}\right)^{-\delta} \mathrm{g/cm^2}. \tag{1.28}$$

where $\beta \equiv v/c$, R is the rigidity in GV. It is comparable to Eq. (1.19) from the leaky-box model. For the Galaxy, $h = 100\,\mathrm{pc}$, gas density in the disk is $n_H = 1\,\mathrm{cm^{-3}}$, one has

$$\frac{D}{H} = \frac{0.8 \times 10^6\,\mathrm{cm/s}}{\beta^2} \left(\frac{R}{3\,\mathrm{GV}}\right)^{-\delta}, \tag{1.29}$$

which gives an escape speed of order 10^6 cm/s for relativistic particles. At higher energies, the diffusion coefficient rises with rigidity as $D \propto E^\delta$, and at a fixed height H [19]. Note that in the 1D case, the Green's function must be written in its 1D form, while in 3D, the diffusion expressions have the same structure but allow escape both vertically and radially. The cosmic-ray diffusion explains why energetic charged particles are observed with nearly isotropic distributions and why they remain confined in the Galaxy [151]. The new AMS-02 fits prefer $\delta \approx 0.43$-0.53 and diffusion coefficients at $\geq 10\,\mathrm{GV}$ in the range $D_{xx} \sim (8\text{-}11) \times 10^{28}\ \mathrm{cm^2\,s^{-1}}$ [154].

In diffusion theory, cosmic rays are almost isotropic, but small gradients in their density lead to a dipole anisotropy. The anisotropy from discrete sources can be estimated as the ratio of the diffusive flux to the isotropic flux [181, 182]. Specifically, we start with the intensity in a direction $\hat{n}$, which can be written as

$$I(\hat{n}) = \frac{v}{4\pi} \, N(\mathbf{r}) \, \Phi(\hat{n}), \tag{1.30}$$

where $N(\mathbf{r})$ is the particle density and $v = \beta c$ is the velocity. To describe a nearly isotropic distribution, the angular dependence is expanded to first order as $\Phi(\hat{n}) \simeq 1 + \boldsymbol{a} \cdot \hat{n}$, where

$\boldsymbol{a}$ is the dipole vector. The magnitude of this vector defines the dipole amplitude,

$$A = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} = |\boldsymbol{a}|. \tag{1.31}$$

The flux of particles is obtained by taking the first angular moment of the intensity. In this approximation, it becomes $\mathbf{J} \simeq v\, N\, \boldsymbol{a}/3$, which directly relates the dipole to the ratio of flux and density, $A = 3|\mathbf{J}|/(vN)$. In the diffusion picture, the flux is driven by spatial gradients according to Fick's law, $\mathbf{J} = -D\,\nabla N$, so the anisotropy can be written as:

$$A = \frac{3D}{\beta c}\,|\nabla \ln N|, \tag{1.32}$$

where the anisotropy amplitude and direction depend on D and the local gradient [5]. In the anisotropy formula, the relevant scale is set by the gradient length of the cosmic-ray density, not by the full 3D radius. For the Milky Way, its halo radius is larger ($R_h \sim 15$–$20\,\mathrm{kpc}$) compared to its vertical half-height ($H \sim$ a few kpc), so vertical escape dominates, and the anisotropy is expressed in terms of H. Thus, for a one-dimensional disk-halo geometry, the density decreases over the scale of the halo height H. This gives a typical gradient of order $1/H$, and the anisotropy reduces to $A \simeq 3D/(\beta c\, H)$, where for high-energy cosmic rays, $\beta \equiv v/c \approx 1$. Finally, assuming the diffusion coefficient increases with rigidity (R), follows $D = D_0\,\beta\,(R/R_0)^{\delta}$. Then the anisotropy is expected as,

$$A(R) \simeq \frac{3D_0}{c\,H}\left(\frac{R}{R_0}\right)^{\delta}. \tag{1.33}$$

which decreases when the distance from the Galactic plane increases. The measured dipole anisotropy is much smaller than simple diffusion predicts: below (10^{-3}) at 10–20 TeV (Super-K [134], IceCube [140]) and only ($\sim 0.3\%$) at 2 PeV (IceTop [63]). This discrepancy has been mentioned in Section 1.3, i.e., the "cosmic-ray anisotropy problem" [144, 5]. One possible reason could be that the basic diffusion model is too simple: the sources are not uniform, and strong scattering in turbulent fields blurs the link to the dipole. Two solutions are often discussed. Reacceleration lowers the effective slope to ($\delta \simeq 0.3$), reducing anisotropy [183]. Alternatively, diffusion may be faster in source-rich spiral arms and slower locally, giving a smaller dipole at Earth [184]. Moreover, one can use the code CRPropa [178] to study the anisotropy at Earth by injecting cosmic rays from single or multiple sources, and then analyzing their propagation. We will apply this later in Chapter 6 to compare with the 33 PeV cosmic rays observed at KASCADE-Grande.

1.5 Sources

Cosmic-ray particles are accelerated by a range of mechanisms within diverse environments, both within our Galaxy and beyond. Cosmic rays can only be produced if charged particles are accelerated; hence, hadronic interactions at the sources are required. Their origins and the processes responsible for their acceleration have been subjects of extensive scientific investigation. While cosmic rays are detected from all directions, their sources vary and depend on factors such as energy levels, cosmic environments, and astrophysical processes. Within the Galaxy, sources such as supernova remnants, pulsars, and stellar clusters are thought to accelerate most cosmic rays of lower energies, while extragalactic objects like gamma-ray bursts (GRBs) and active galactic nuclei (AGN) are likely responsible for the highest-energy events [3]. Although the detailed acceleration processes of these sources remain uncertain, cosmic rays are generally modeled as being continuously injected over a

long period (millions of years) from all sources with an average spectrum and then diffusing through the Galactic magnetic field.

In 1934, Baade and Zwicky first proposed that supernovae (SNe) could power Galactic cosmic rays [185, 186]. They noted that, first, the flux of cosmic rays remains nearly constant over time, implying that they must be continuously produced throughout the Galaxy rather than originating from the Sun or ordinary stellar radiation [186]. Second, the lower intensity of cosmic rays near the equator can be explained if some of them are very energetic charged particles. Such particles must come from outside the Earth, since only then would their path be long enough for Earth's magnetic field to cause the observed dip at the equator [186].

The sources of Galactic cosmic rays are constrained by the energy required to maintain the observed cosmic-ray density in the Galaxy. The local cosmic-ray energy density has been measured by Voyager as $\rho_{\rm CR} \simeq 0.7$ eV/cm^3 [187]. If cosmic rays are confined for a time τ inside a volume V of the Galaxy and containing gas mass M, they cross a grammage $X = c\tau M/V$ [188]. The corresponding cosmic-ray luminosity is

$$L_{\rm CR} = \rho_{\rm CR} \frac{cM}{X} \simeq 5 \times 10^{40} \text{ erg/s}, \tag{1.34}$$

where $M = 5 \times 10^9 M_\odot$ and $X \simeq 10$ g/cm^2. A core-collapse SN ejects about $10\,M_\odot$ with typical velocities of $v \sim 5 \times 10^8$ cm/s. With a Galactic rate of about one per 30 years, the average kinetic energy injection is $L_{\rm SN,kin} \sim 3 \times 10^{42}$ erg/s. If only a few percent of this energy is converted into cosmic rays, they can explain the observed Galactic cosmic-ray luminosity in Eq.(1.34), at energy up to 10^{17} eV [144]. Magnetic-field amplification at supernova shocks allows particles to be accelerated up to the knee. Cosmic rays may take about 20–30% of the kinetic energy of the explosion [189, 190], thereby making supernovae a likely source of Galactic cosmic rays. A well-known example is the Crab Nebula supernova remnant, illustrated in the left panel of Figure 1.11. The distribution of core-collapse supernovae in the Milky Way is not uniform, but usually occurs in OB associations [188]. Multiple SNe within a few million years create super bubbles, with turbulence and multiple shocks that can accelerate CRs up to PeV energies [191]. This scenario also explains the efficient production of Li, Be, and B in the early Galaxy [192], the anomalous ^{22}Ne/^{20}Ne ratio [193], and the hard spectra of heavy nuclei [194, 195].

Similar to SN expanding, a nova explosion also produces expanding shells but with smaller energies (10^{46}-10^{47} erg) [188]. Their Galactic rate is about 100 per year, so the total energy budget is comparable to that of SNe. However, their maximum rigidity is limited to ~ 200 GV, so they may only contribute a steep, low-energy component [196]. Besides, explosions of massive stars into their winds can drive cosmic rays to energies above the knee, and possibly up to the ankle, if the explosion energy is above 10^{51} erg and the magnetic field is strong [197, 198]. Besides, winds of Wolf-Rayet stars may accelerate particles up to about 10^{18} eV, possibly marking the end of the Galactic CR spectrum [199, 200, 201]. The Galactic center (GC) has a supermassive black hole (Sgr A*), which may also accelerate CRs [188]. Past active episodes, possibly responsible for the Fermi bubbles, released on average $(1\text{–}7) \times 10^{42}$ erg/s [202]. Such outbursts can contribute significantly to CRs up to the knee if the diffusion halo is large and the diffusion coefficient slope is steep ($\delta \sim 0.5$) [188]. Because electrons and positrons lose energy quickly, the GC is expected to contribute mainly to hadrons [188]. All the sources mentioned above are from the Galactic region.

For extragalactic sources, a typical candidate is Gamma-ray bursts (GRBs), which

Figure 1.11: Left: Crab Nebula, an expanding SNR, recorded by a Chinese astronomer in 1054 CE (Credit: NASA). Right: Composite color image of Centaurus A (an AGN), showing the jets and lobes powered by the central black hole (Credit: ESO and NASA).

are among the brightest events in the universe, releasing enormous energy in only a few seconds, but are rare in the Galaxy ($\lesssim 10^{-4}$ per year). They are short-lived and thought to result from the collapse of massive stars or the merger of compact objects such as neutron stars, producing relativistic jets and strong magnetic fields that can accelerate particles to ultra-high energies [13]. GRBs have been proposed as sources of cosmic rays above the knee [203] and even ultra-high-energy cosmic rays (UHECRs) [204, 205, 206, 207]. However, subsequent studies indicate that their total energy output and acceleration efficiency may be insufficient to account for the observed UHECR flux [208]. High-luminosity GRBs roughly meet the energy budget for UHECRs, while low-luminosity GRBs (LLGRBs) occur more often and have weaker radiation fields, allowing heavy nuclei to survive [209, 210]. Studies of both types [211, 212, 213, 214, 215, 216, 217] show that in certain regimes, GRBs could reproduce the observed UHECR flux and composition. Ordinary supernovae are too weak to reach such energies, but engine-driven or trans-relativistic supernovae with mildly relativistic ejecta may do so [217, 218, 219, 220, 221]. Their predicted heavy composition agrees with Auger data. The 2017 detection of a neutron-star merger and its short GRB confirmed such events as real sources [222, 223]. They could explain cosmic rays below the ankle [224] or near the second knee [225].

In addition, active galactic nuclei (AGNs) are among the most promising candidates for accelerating UHECRs. A well-known example is Centaurus A (Cen A), the nearest radio galaxy to Earth and a long-standing UHECR candidate, illustrated in the right panel of Figure 1.11. AGNs are powered by supermassive black holes at the centers of galaxies, which accrete matter and eject powerful jets of particles at relativistic speeds. These jets, combined with the intense magnetic fields in the central regions of AGNs, are capable of accelerating particles to energies well beyond 10^{18} eV through processes like Fermi acceleration. Observational studies of UHECR arrival directions by the Pierre Auger Observatory have revealed a correlation between UHECRs and the distribution of matter in the large-scale structure of the universe [226], suggesting that these particles likely originate from nearby extragalactic structures, possibly including AGNs. AGNs with strong jets are leading candidates for the origin of ultra-high-energy cosmic rays (UHECRs) [13]. When

the jets point toward Earth, they appear as blazars, which could accelerate and beam UHECRs directly [227, 228, 229]. However, the number of nearby blazars is too small to explain the total flux. The radio galaxies, which are the parent population of blazars with jets not aligned with the line of sight, are also strong UHECR candidates [13]. The nearest example is Cen A, which is a long-standing candidate [230, 199, 231]. Recent models show that radio galaxies can produce the observed flux and heavy composition through shear or re-acceleration processes [232, 233]. In jetted AGN, a large fraction of the energy goes into radio lobes, which are extended regions with relatively weak magnetic fields ($B \sim 10^{-5}\,\mathrm{G}$), where energies up to $10^{20}\,\mathrm{eV}$ are possible [234, 235]. In addition, radio-quiet and low-luminosity AGN or quasar outflows could contribute collectively, as they are much more numerous [236, 237, 13]. Besides, large-scale structure (LSS) shocks formed during cluster and filament growth are natural sites for diffusive shock acceleration and may contribute to high-energy cosmic rays. However, limited magnetic fields ($\sim \mu$G) and energy losses restrict the maximum energy to about 10^{18}–$10^{19}\,\mathrm{eV}$, which makes their contribution to UHECR production uncertain compared with AGNs and GRBs [238].

1.6 Acceleration

The acceleration of cosmic rays should explain the broad energy range observed in their spectra, extending from a few MeV up to ultra-high energies exceeding $10^{18}\,\mathrm{eV}$ [19]. In 1949, Fermi proposed that charged particles can gain energy through random interactions with moving magnetic irregularities, a stochastic process now known as *second-order Fermi acceleration* [239]. This mechanism produces a power-law energy spectrum, $N(E) \propto E^{-\gamma}$, and is expected to operate in turbulent environments such as the downstream regions of supernova-remnant shocks, within active galactic nucleus (AGN) jets, and in intracluster turbulence. Subsequently, the theory of diffusive shock acceleration (DSA), known as the *first-order Fermi mechanism*, was developed as an efficient process that naturally generates power-law spectra [240, 241, 242, 2]. It operates at collisionless shocks, such as those in supernova remnants, in shocks within AGN jets, and in internal or external shocks of gamma-ray bursts. Although both mechanisms rely on repeated interactions with magnetic fields, they differ fundamentally in their energy gain per cycle.

1.6.1 Fermi Acceleration

Fermi's 1949 idea proposed that cosmic rays gain energy through repeated elastic scatterings with moving magnetic irregularities. This stochastic process increases their energy gradually and does not yet distinguish between different orders of acceleration [239]. In this context, the "order" refers to how the average fractional energy gain per interaction depends on the velocity of the scattering centers, usually expressed as $\beta = v/c$. Later developments clarified that such interactions fall into two regimes: second-order Fermi acceleration, where the energy gain is stochastic and scales as $\langle \Delta E/E \rangle \propto \beta^2$, and first-order Fermi acceleration, which occurs at shock fronts and gives a systematic gain scaling as $\langle \Delta E/E \rangle \propto \beta$.

Fermi proposed that cosmic rays could gradually gain energy through repeated elastic scatterings off moving magnetic irregularities, often referred to as "magnetic clouds" in the interstellar medium [239]. During each individual encounter, the particle's energy changes only by a small amount, but when one averages over many random head-on collisions and overtaking interactions with these magnetic irregularities, the statistics yield a consistently positive net energy gain for the particle. In the case where the scattering centers move

non-relativistically, such that their speed satisfies $\beta \equiv V/c \ll 1$, the average fractional energy gain per interaction is $\langle \Delta E/E \rangle \simeq 4\beta^2/3$, so the growth is second order in β and thus slow per interaction. Nevertheless, if one denotes the average fractional energy gain in a single encounter by $\xi \equiv \langle \Delta E/E \rangle$, then after n such scatterings the particle's energy follows a simple multiplicative growth law, given by

$$E_n \;=\; E_0(1+\xi)^n, \tag{1.35}$$

and with escape probability per encounter $P_{\rm esc}$ the survival probability after n encounters is $(1-P_{\rm esc})^n$. Eliminating $n = \ln(E/E_0)/\ln(1+\xi)$ gives a power-law integral spectrum:

$$N(>E) \propto \frac{(1-P_{\rm esc})^n}{P_{\rm esc}} = \frac{1}{P_{\rm esc}} \left(\frac{E}{E_0}\right)^{-\gamma}, \quad \gamma = \frac{\ln(1/P_{\rm esc})}{\ln(1+\xi)}, \tag{1.36}$$

where E_0 is the particle's initial energy, and the equation gives the proportion of particles reaching energies greater than E, which can be found by summing the probabilities of all possible encounters beyond n. This relation is the basic Fermi result (see also [243, 19]).

Consider a relativistic particle with lab-frame energy E_1 entering a scattering region, moving at velocity $V = \beta c$ at an angle θ_1 with respect to the flow direction. After several scatterings within the moving magnetic cloud, the particle's momentum distribution becomes isotropic in the cloud frame. In the cloud's rest frame, the particle's total energy, including both rest mass and kinetic contributions, is denoted by E_1'. If the scattering is elastic, the energy remains unchanged in the moving frame ($E_2' = E_1'$), and transforming back to the lab frame gives E_2. They are expressed as,

$$E_1' = \Gamma\, E_1(1-\beta\cos\theta_1), \quad E_2 = \Gamma\, E_2'(1+\beta\cos\theta_2'), \tag{1.37}$$

where $\Gamma = (1-\beta^2)^{-1/2}$ is the Lorentz factor, θ_2' is the exit angle of the particle relative to the direction of motion of the scattering region. Combining the two equations above, the single-encounter fractional energy change is obtained as

$$\frac{\Delta E}{E_1} = \frac{1-\beta\cos\theta_1+\beta\cos\theta_2'-\beta^2\cos\theta_1\cos\theta_2'}{1-\beta^2} - 1, \tag{1.38}$$

which is valid for $E \approx pc$ due to the approximate relativistic condition.

Second-order Fermi mechanism: scattering from plasma clouds. During a scattering event with such a cloud, the distribution of outgoing particle directions is isotropic in the cloud's rest frame, whereas the distribution of incoming directions is not uniform but instead weighted by the relative velocity between the particle and the moving cloud, which biases head-on encounters [19]. This isotropy of the outgoing directions implies that the differential distribution satisfies $\mathrm{d}n/(\mathrm{d}\cos\theta_2') = \text{const.}$, over the full range $-1 \le \cos\theta_2' \le 1$, and therefore the average value vanishes, giving $\langle \cos\theta_2' \rangle = 0$, as expected for a symmetric angular distribution. When one averages Eq. (1.38) over the relevant angular distributions in the non-relativistic limit $\beta \ll 1$, the incident-angle average becomes $\langle \cos\theta_1 \rangle = -V/(3c)$, and the mean fractional gain per encounter is [19]

$$\xi = \left\langle \frac{\Delta E}{E} \right\rangle \;\simeq\; \frac{4}{3}\,\beta^2. \tag{1.39}$$

where $\beta = V/c$ denotes the velocity of the plasma flow relative to the speed of light, rather than the velocity of the cosmic-ray particles themselves. Because the gain scales as β^2, many encounters are required for substantial acceleration [19].

First-order Fermi mechanism: encounters with a plane shock front. In the shock case, particles scatter on turbulence upstream and downstream and execute repeated

shock crossings [19]. For particles transmitted downstream in the shock frame, the angular distribution is no longer isotropic but weighted toward forward directions, with $\mathrm{d}n/\mathrm{d}\cos\theta_2' = 2\cos\theta_2'$ over the interval $0 \leq \cos\theta_2' \leq 1$, and this leads directly to an average value of $\langle\cos\theta_2'\rangle = 2/3$ [19]. To characterize the dynamics, we denote by u_1 and u_2 the plasma flow velocities upstream and downstream of the shock, respectively, in the shock-rest frame, with $u_1 > u_2$ as required for a compressive shock [19]. Similarly, averaging (1.38) over angles gives $\langle\cos\theta_1\rangle = -2/3$ and obtain the gain fraction [19]

$$\xi = \left\langle \frac{\Delta E}{E} \right\rangle \simeq \frac{4}{3}\frac{u_1 - u_2}{c} \approx \frac{4}{3}\frac{u_s}{c}, \tag{1.40}$$

where u_s is the shock speed. For non-relativistic, strong shocks ($u_s \simeq u_1$), we set $\beta \simeq \beta_s = u_s/c$. This linear scaling makes diffusive shock acceleration far more efficient than stochastic scattering, allowing particles to reach higher energies [19].

Energy spectrum. If the average fractional gain per *cycle* is $\xi \equiv \langle\Delta E/E\rangle$ and the escape probability per cycle is $P_{\rm esc}$, then after n cycles E_n follows Eq. (1.35). Hence, $\gamma \simeq P_{\rm esc}/\xi$, where $\xi, P_{\rm esc} \ll 1$, because Fermi acceleration assumes small fractional energy gains and small escape probabilities per cycle, which allows the continuous and power-law limit [19]. For a plane shock, projecting an isotropic flux onto the shock plane gives an encounter rate $c\rho_{\rm CR}/4$, while downstream advection removes particles at a rate $\rho_{\rm CR}u_2$, so the escape probability per cycle is $P_{\rm esc} = 4u_2/c$. Combining $P_{\rm esc}$ with Eq. (1.40), yields [19]

$$\gamma = \frac{P_{\rm esc}}{\xi} = \frac{3}{u_1/u_2 - 1}, \tag{1.41}$$

where $r \equiv u_1/u_2$ is the compression ratio. To get the differential spectrum, we first differentiate $N(>E)$ as in Eq. (1.35), one obtains

$$\frac{\mathrm{d}}{\mathrm{d}E}N(>E) = \frac{d}{\mathrm{d}E}\left[\int_E^\infty \frac{\mathrm{d}N}{\mathrm{d}E'}\mathrm{d}E'\right] = -\frac{\mathrm{d}N}{\mathrm{d}E}. \tag{1.42}$$

Therefore, the differential spectrum and the corresponding differential source index are

$$\frac{\mathrm{d}N}{\mathrm{d}E} \propto E^{-(\gamma+1)}, \quad s \equiv \gamma + 1 = \frac{r+2}{r-1}. \tag{1.43}$$

A shock forms when the upstream plasma velocity exceeds the local sound speed, $u_1 > c_1$, corresponding to a Mach number $M = u_1/c_1 > 1$. From the conservation of mass, momentum, and energy across the shock, the ratio of downstream to upstream densities (or equivalently the inverse velocity ratio) is [244, 19]

$$\frac{\rho_2}{\rho_1} = \frac{u_1}{u_2} = \frac{(\gamma_g+1)M^2}{(\gamma_g-1)M^2+2}, \tag{1.44}$$

where $\gamma_g = c_p/c_v$ is the adiabatic index of the gas. In the strong-shock limit ($M \gg 1$), this expression approaches $\rho_2/\rho_1 = (\gamma_g+1)/(\gamma_g-1)$. For a monoatomic ideal gas, where $\gamma_g = 5/3$, the compression ratio becomes $r = \rho_2/\rho_1 \approx 4$. This value is widely adopted for non-relativistic shocks, and through diffusive shock acceleration, to the canonical source (acceleration) spectrum $\mathrm{d}N/\mathrm{d}E \propto E^{-2}$.

Equations (1.39) and (1.40) show why shock acceleration is generally dominant, i.e., the gain is first order in β rather than second order in β^2. Hillas' criterion provides a general consideration for estimating the limited energy to which cosmic rays can be accelerated in sources such as SNRs and AGNs [245, 19]. The requirement is that a particle must remain magnetically confined within the acceleration region. The Larmor (gyration) radius of a

charged particle with energy E in a magnetic field B is

$$r_{\mathrm{L}} = \frac{p_{\perp} c}{ZeB} \simeq \frac{E}{ZeB}, \tag{1.45}$$

where Ze is the particle charge (Z is the atomic number) and $p_{\perp}$ is the component of its momentum perpendicular to the magnetic field. For effective acceleration by a shock, the condition $r_{\mathrm{L}} \lesssim R$ must be satisfied, where R is the size of the acceleration region in the source. This Hillas criterion gives an upper bound on the maximum energy:

$$E_{\max} \lesssim Ze\,\beta\,B\,R, \tag{1.46}$$

where $\beta = u_s/c$, the inclusion of β is necessary because the acceleration rate depends on the speed of the shock or plasma motion, which represents the dimensionless velocity of the scattering or acceleration region.

Galactic supernova remnants (SNRs) are widely considered the main accelerators of cosmic rays up to the knee energy, $E_{\mathrm{knee}} \sim 3\times10^{15}\,\mathrm{eV}$. Classical estimates of diffusive shock acceleration (DSA) without strong magnetic-field amplification limit the maximum proton energy in typical SNRs to $E_{\max} \lesssim 10^{14\text{–}15}\,\mathrm{eV}$, often referred to as the Lagage–Cesarsky limit [246, 247], also verified by Morlino [248]. Efficient magnetic-field amplification by streaming instabilities, such as the non-resonant hybrid (Bell) instability, can increase the field to $B \sim 100\,\mu\mathrm{G}$, allowing young SNRs with fast shocks ($u_s \sim 5\text{–}10\times10^3\,\mathrm{km\,s^{-1}}$) to reach proton energies of a few PeV for a short period, consistent with the Hillas scaling [190, 3, 232], as Eq. (1.46) shows. Observation aspect, the knee and its rigidity-dependent steepening are consistent with a Galactic proton cutoff near a few PeV and heavier nuclei extending up to Z times higher energies [23, 19, 245]. Recent γ-ray observations of the Galactic Centre region also indicate the presence of a long-lived Galactic PeVatron capable of accelerating protons to PeV energies [6]. In addition, the LHAASO observatory has detected a γ-ray source (a $< 6^\circ$ bubble) above PeV energies in the Cygnus OB2 region, providing further evidence of active Galactic PeVatrons [8].

Besides, the ultrarelativistic shocks in GRBs or powerful AGN shocks are under the Fermi first-order (diffusive shock) acceleration mechanism. When two collisionless plasma flows collide with a Lorentz factor $\Gamma \gg 1$, a shock forms and accelerates particles through repeated shock crossings. Under small-angle, quasi-isotropic scattering, this process produces a nearly universal power-law spectrum with index $s \simeq 2.2\text{–}2.3$ [249, 250, 251]. The maximum energy is bounded by confinement (Hillas criterion), available time, and losses. In favorable conditions, GRBs and powerful AGN shocks can reach EeV energies for protons [207, 206, 204]. So, the highest energies still stay as a mystery.

1.6.2 Other Acceleration Mechanisms

In cosmic-ray sources, besides diffusive shock acceleration, several other processes can work together with or replace Fermi acceleration, depending on the environment and evolutionary stage [3]. In complex environments such as SNRs, stellar clusters, and relativistic jets, shocks where flows collide, magnetic reconnection where fields reverse, shear at velocity boundaries, and turbulence throughout may all contribute to particle energization. Potential cosmic-ray accelerators can be classified by their characteristic size and magnetic-field strength [245], as shown in the Hillas diagram (Figure 1.12). The dominant mechanism depends on the plasma magnetization, geometry, available acceleration time, and the ability of particles to remain confined long enough to reach the observed energies [3, 204].

For example, *magnetic reconnection* provides an additional acceleration channel in

highly magnetized plasmas; oppositely directed magnetic fields can reconnect, releasing magnetic energy and generating strong electric fields that accelerate particles directly [252]. In relativistic jets and magnetized coronae, reconnection can operate alongside shock acceleration and may feed the particle population that shocks further energize. This process is relevant in pulsar wind nebulae (PWN), AGN jets and coronae, and accretion flows [252]. Giannios found that, in relativistic jets, magnetic reconnection can accelerate protons up to about $E \sim 10^{20}\,\mathrm{eV}$ in GRBs and powerful AGN jets, while iron nuclei can reach similar energies in moderately luminous AGN jets [253].

Besides, a rotating, magnetized compact object sets up a global electromotive force that extracts charges and drives currents along open magnetic field lines, which is the *Goldreich–Julian framework* [254]. This framework fixes the corotation charge density, opens field lines beyond the light cylinder, and provides a large potential drop across the open zone. In principle, this voltage can accelerate particles to very high energies. Meanwhile, pair creation and radiation losses reduce the energy. Within the Goldreich–Julian picture, the polar-cap model was proposed [255, 256], which is a specific local mechanism. Above the magnetic poles, charge-starved gaps can form where $E_{\parallel} \neq 0$, directly accelerating particles along open field lines [255, 256, 257]. The induced potential, roughly $\Phi \sim (\Omega B R^2)/c$, can in principle accelerate particles up to $E_{\max} \sim Ze\Phi$. For the typical properties of millisecond pulsars, protons could be accelerated to the energy $E_{\mathrm{p}} \simeq 1\,\mathrm{PeV}$, with luminosities of $L_{\mathrm{p}} \simeq 5 \times 10^{35}\,\mathrm{erg\,s^{-1}}$, using particle-in-cell simulations of the aligned pulsar magnetosphere [258]. For a typical pulsar, the energy corresponds to $\sim 10^{14\text{–}16}$ eV [255, 257], while young/fast SNRs (e.g., Crab Nebula) can reach $\sim 10^{15\text{–}16}$ eV [259].

Moreover, *unipolar inductors* provide an alternative way to accelerate particles to ultrahigh energies [179]. They arise in fast-rotating, magnetized systems such as neutron stars or black holes with accretion disks that lose rotational energy through jets. In neutron stars, rotation and strong magnetic fields induce electric fields that can accelerate charged particles along magnetic lines [204]. This process was first discussed for ordinary pulsars [261], but typical pulsars cannot reach energies above 10^{20} eV. Young magnetars with millisecond rotation periods and very strong surface magnetic fields can, however, reach such energies [262]. As the magnetar spins down, the energy of the accelerated particles decreases, producing a power-law energy spectrum [262, 263]. Ultra–high energy particles are thought to be produced during the early life of a magnetar, within a few days after its birth, making the emission short and impulsive [263]. This model was first proposed to explain the lack of a GZK cutoff, but the very hard injection spectrum it predicts does not match current data. A range of initial voltages among magnetars could produce a softer spectrum and may also generate detectable gravitational-wave signals [204].

Lastly, *wakefield acceleration* (WFA) forms when strong waves create charge separation in a plasma, which was found by Tajima in 1979 [264]. When intense electromagnetic or plasma waves propagate through a plasma, they push electrons and ions differently, creating strong charge separation and driving a wakefield behind the wave. The wakefield produces a ponderomotive force, a time-averaged, nonlinear force in an oscillating, inhomogeneous field, that can trap particles and accelerate them efficiently once the wave potential exceeds their initial momentum [264]. In this way, energy stored in collective plasma oscillations is converted into particle kinetic energy. In laboratory studies, plasma wakefields can be driven by high-energy beams, such as the 400 GeV proton beam used in CERN's AWAKE experiment [265]. Chen proposed that strong Alfvén shocks in relativistic astrophysical plasmas can drive plasma wakefields capable of accelerating particles to ultra-high energies [266]. Acceleration by the wake's longitudinal field can reduce some radiative losses

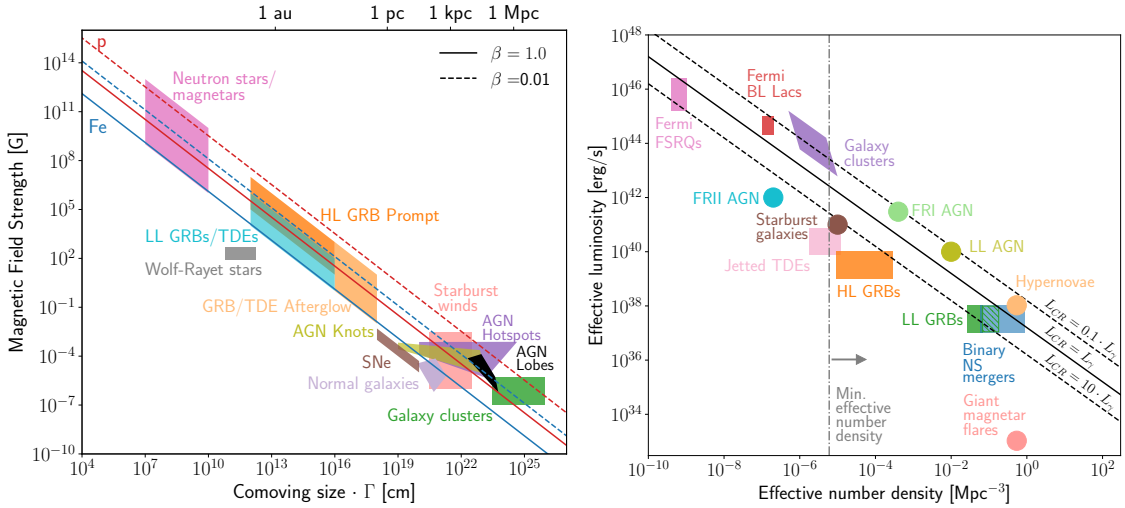


Figure 1.12: Left: Hillas diagram of the distribution of potential CR accelerators, where objects are positioned according to their characteristic size R and magnetic-field strength B [245, 13]. The shown B values correspond to the comoving frame of the outflow. The horizontal axis denotes R, interpreted as the distance from the central engine or, equivalently, the commoving source size scaled by the bulk Lorentz factor Γ. Solid and dashed curves represent the confinement limits for protons (red) and iron nuclei (blue) at energies of 10^{20} eV, assuming relativistic ($\beta_{\rm sh} = 1$) and sub-relativistic ($\beta_{\rm sh} = 0.01$) shock velocities, respectively. Sources located to the left of these boundaries cannot accelerate particles to 10^{20} eV. Right: Energy budgets of various candidate classes derived from infrared, radio, X-ray, and gamma-ray observations, compared with the ultra-high-energy cosmic-ray production rate estimated in [260]. Both panels are adapted from [13].

compared to shock cycles and can yield power-law spectra under suitable conditions [266]. Ebisuzaki and Tajima [267] suggested that intense Alfvén waves from magnetized accretion disks around supermassive black holes can drive ponderomotive and wakefield forces in relativistic jets, accelerating particles up to ZeV energies.

Chapter 2

Reconstruction of Anisotropies

Cosmic-ray arrival directions are not perfectly isotropic. Multiple experiments have now observed both large- and small-scale modulations across TeV–EeV energies [136, 268, 135, 269, 138, 139, 270, 137]. The typical amplitudes are at the level of 10^{-4}–10^{-3} in the TeV–PeV range, hence, careful control of exposure and systematics is essential [5].

These anisotropies appear on different angular scales. Large-scale features (dipole and low-ℓ multipoles) are consistent with a Galactic gradient shaped by diffusion and local magnetic geometry [16]. In contrast, small-scale structures suggest additional effects such as local magnetic turbulence and, possibly, nearby sources imprinting patterns that standard, smooth diffusion does not capture [5, 136]. Together, these observations indicate that cosmic-ray transport is more complex than simple homogeneous diffusion. At PeV energies and above, deflections in the Galactic magnetic field decrease, so arrival directions retain more information about their sources. This makes the anisotropy a potential way to probe the origin of the highest-energy Galactic cosmic rays, although limited statistics and non-uniform detector exposure still make reconstruction difficult [270]. Since the main goal of this thesis is to investigate detailed cosmic-ray anisotropies at high energies and to search for clues about Galactic cosmic-ray sources, this chapter focuses on introducing the existing methods used to reconstruct small deviations in cosmic-ray arrival directions from isotropy before presenting the main analyses in the following chapters. We will explain the coordinate transforms used to place events in the sidereal frame, describe reference-map building with time scrambling, and review both harmonic (Rayleigh/multipole) and likelihood approaches that fit sky intensity and detector acceptance [143, 5].

2.1 Reconstruction Procedures

Reconstruction of cosmic-ray anisotropies from isotropy involves two main steps for the data. First, each event is placed in the sidereal frame by converting event times to local sidereal time and mapping local directions to equatorial coordinates. Second, the detector exposure must be modelled carefully so that instrumental effects do not mimic real sky patterns. The observed event map is then compared with an exposure-based reference. To build this reference, time-scrambling or direct-integration methods are used to estimate the isotropic expectation under the detector's live time and acceptance. Weather, pressure, and live-time variations can produce artificial day-scale modulations, so solar and anti-sidereal sidebands are checked to monitor possible systematics [143, 271]. Analyses typically use sky-map projection, sidereal harmonics, or full-sky likelihood for large scales, apply East–West or maximum-likelihood fitting for reconstructions, and verify results in solar and anti-sidereal bands. Amplitudes, phases, and post-trial significances are then compared across different methods and energy bins [143, 5]. For the reconstruction of small-scale anisotropies, the spherical harmonic expansion, combined with the likelihood method, provides the most effective way to recover all possible structures [5, 16]. Based on this, an extended method is developed and applied to KASCADE-Grande in Chapter 6.

2.1.1 Coordinate Systems

Many cosmic-ray observatories record each event's arrival time in Coordinated Universal Time (UTC), providing a uniform global reference for timestamping large data sets. For example, the KASCADE-Grande experiment and the Pierre Auger Observatory supply data sets in which every event contains a UTC timestamp. For anisotropy analyses, however, the physical modulations of interest are expressed not in UTC but in sidereal, solar, and anti-sidereal time, so each recorded event must be transformed into these different time frames to separate signals from seasonal or instrumental effects. This conversion requires applying a well-defined time-transformation procedure based on the Greenwich Mean Sidereal Time (GMST), which relates the Earth's rotation to celestial coordinates.

To transform to equatorial coordinates, which are independent of the observer's position, we begin by computing the altitude h, which is related to the zenith angle, i.e., $h = 90° - \theta$. Using the observer's latitude ϕ and azimuth A, the declination δ, which is the angular distance of the object from the celestial equator [272, 273, 274], can be calculated as:

$$\sin\delta = \sin\phi\sin h + \cos\phi\cos h\cos A. \tag{2.1}$$

After determining the declination, the next step is to compute the hour angle H. The hour angle represents the time since the celestial object last crossed the observer's local meridian. It can be calculated through the following relations:

$$\cos H = \frac{\sin h - \sin\phi\sin\delta}{\cos\phi\cos\delta}, \quad \sin H = -\frac{\cos h\sin A}{\cos\delta}, \tag{2.2}$$

where the hour angle H can then be obtained by applying the arctan2 function, to combine both the sine and cosine components [273]:

$$H = \operatorname{arctan2}\left(-\cos h\sin A,\ \sin h - \sin\phi\sin\delta\right). \tag{2.3}$$

To find the right ascension α, the Local Sidereal Time (LST) must first be determined.

We adopt the azimuth convention A measured from North through East (0° =N, 90° =E). The GMST expression gives hours and is defined with respect to UT1, the formal timescale for sidereal time. In practice, UTC is used as an approximation (the UT1–UTC difference is typically less than one second). All angles and times are wrapped to their ranges. The LST depends on the observer's longitude λ and the Greenwich Sidereal Time (GST) [272, 275] through the equation:

$$\mathrm{LST} = \mathrm{GMST} + \lambda/15, \tag{2.4}$$

where the GMST is derived from the Julian Date and is calculated at 0h Universal Time (UT) using the following formula:

$$\mathrm{GMST\,[h]} = 18.697374558 + 24.06570982441908\,D, \quad D = \mathrm{JD} - 2451545.0. \tag{2.5}$$

where D represents the number of days since the reference epoch J2000, and $\mathrm{JD} = t_{\mathrm{UTC}}/86400 + 2440587.5$, where t_{utc} is simply the event timestamp expressed in Unix time, i.e., the number of seconds (some data may also include microseconds) recorded since the time 00 : 00 : 00 UTC in 1970-01-01. Finally, the right ascension α is calculated by subtracting the hour angle H from LST [272, 275]:

$$\alpha = \mathrm{LST} - H, \tag{2.6}$$

result wrapped to $[0, 24)$ h or degree. We must emphasize that "Sidereal time" usually refers to GST, which is a clock that tells us which part of the celestial sphere or right ascension

(RA) is crossing the meridian at Greenwich (0° longitude). However, what really matters for a given observatory (e.g., KASCADE-Grande) is the sidereal time at the observatory's longitude, which gives the current RA crossing its local meridian. If one uses GST instead, it would shift all phases of the detector longitude, and any true RA modulation will be washed out. Therefore, LST, as defined in Eq. (2.4), should be used for anisotropy studies.

2.1.2 Time-Scrambling of Data

In cosmic-ray anisotropy studies, the main goal is to detect deviations from an isotropic distribution of arrival directions. However, the observed anisotropies in real data may come from two different sources: instrumental exposure variations across the field of view, and true anisotropies produced by cosmic-ray sources and propagation. The exposure-related modulation is fixed in the detector's local coordinate frame, while an astrophysical signal is stable in the sidereal frame. The challenge is therefore to disentangle these two effects [139]. The time-scrambling algorithm addresses this by constructing an exposure-matched isotropic reference for comparison with the observed data [139]. It keeps each event's local coordinates fixed and resamples their arrival times in UTC (which closely approximates UT1, the timescale based on Earth's true rotation). Right ascension (R.A.) is then recomputed from the new LST. This procedure effectively randomizes celestial phases while preserving the detector's local response, including zenith and azimuth acceptance, event rate, and acquisition gaps, thereby removing any real astrophysical anisotropies but retaining detector-related exposure effects.

Specifically, the data from an experiment contains the measured directions in equatorial coordinates (α, δ) obtained from local coordinates (θ, ϕ) and event time t. The reference data is built by keeping each event's local coordinates fixed and assigning k new UTC times $t'_{i,j}$ drawn from the data's own time distribution within a window Δt centred on t_i (thus preserving live time and gaps) [139, 138, 88]. With (θ_i, ϕ_i) fixed, δ does not change, and the obtained new R.A. follows from the LST:

$$\alpha'_{i,j} = \mathrm{LST}(t'_{i,j}) - H(\theta_i, \phi_i, \phi_{\mathrm{site}}), \tag{2.7}$$

equivalently, we replace H, and obtain

$$\alpha'_{i,j} = \alpha_i + \left[\mathrm{LST}(t'_{i,j}) - \mathrm{LST}(t_i)\right] \pmod{360^\circ}, \tag{2.8}$$

where α_i is the initial R.A. All sidereal quantities are derived from the resampled UTC times (UT1 in principle). The scrambled data is then $(\alpha, \delta) \leftarrow (\alpha'_{i,j}, \delta_i)$, which preserves detector conditions (including dead time) by sampling from the data's own time distribution. In a simple picture, the live-time intervals define the frame within which the scrambling is performed; all resampled timestamps remain confined to this frame, ensuring that none fall outside the valid data-taking periods (no time gaps inside). For two-dimensional sky-map analyses, no additional exposure weighting is needed since the data and reference maps are constructed under identical detector conditions and can be compared pixel by pixel [139]. However, for the standard one-dimensional Rayleigh analysis, an exposure correction as a function of local sidereal time or R.A. must be applied to account for non-uniform coverage in R.A., which will be discussed in Chapter 6. The choice of window Δt sets the R.A. mixing scale. Scrambling over Δt mixes, R.A. by approximately

$$\Delta\alpha \simeq 15^\circ \times \frac{\Delta t}{1\,\mathrm{hour}}, \tag{2.9}$$

acting as a high-pass filter that suppresses structures broader than $\Delta\alpha$ and retains smaller-

scale features [88]. At the South Pole, stable conditions allow $\Delta t = 24\,\mathrm{h}$ (UTC), which robustly captures exposure over a full sidereal rotation [139]. This method is widely used by IceCube and Auger to identify large- and small-scale anisotropies [138, 139, 88].

2.1.3 Anisotropy and Significance

The anisotropies of cosmic-ray arrival directions are usually expressed in terms of the relative intensity, denoted by I or its differential part δI [5]. The relative intensity can be defined as the ratio of the observed flux in a given direction to the isotropic expectation. It is defined as, $I(\mathbf{n}) = \phi(\mathbf{n})/\phi_{\mathrm{iso}}$, where $\phi(\mathbf{n})$ is the directional flux and ϕ_{iso} is the isotropic mean [5], the unit vector $\mathbf{n}$ is defined in the equatorial coordinates,

$$\mathbf{n}(\alpha, \delta) = (\cos\alpha\cos\delta, \sin\alpha\cos\delta, \sin\delta). \tag{2.10}$$

By this definition, $I(\mathbf{n}) = 1$ corresponds to perfect isotropy. Often one writes: $I(\mathbf{n}) = 1 + \delta I(\mathbf{n})$, where $\delta I(\mathbf{n})$ is the anisotropy map (or residual map), satisfying $\langle \delta I(\mathbf{n}) \rangle = 0$. For N events with uniform exposure in α, the Rayleigh amplitude can be obtained, which gives a fast test for a first harmonic in sidereal time [131, 5].

From the observation aspect, based on the time-scrambling procedures as described in Section 2.1.2, the scrambled event distribution $N_{\mathrm{s}}(\alpha_i, \delta_i)$ at the i-th pixel on the reference map is calculated as $N_{\mathrm{exp}}(\alpha_i, \delta_i) = \sum_{j=1}^{N} \mathcal{E}(t'_{i.j})/\mathcal{E}(\alpha, \delta)$, where N is the total number of events in the scrambled dataset, $\mathcal{E}(\alpha_i, \delta_i)$ is the relative exposure in the direction (α_i, δ_i), $\mathcal{E}(t'_i)$ is the exposure at the scrambled time t'_{ij}. This approach effectively normalizes the scrambled events by exposure at their new, randomized arrival times, resulting in an isotropic reference map that accounts for detector effects [139]. Once the reference map is constructed, the anisotropy in cosmic-ray arrival directions is detected by comparing the data map $N_{\mathrm{obs}}(\alpha, \delta)$ (number of observed events in a given direction) with the expected reference map $N_{\mathrm{exp}}(\alpha, \delta)$. The relative intensity $I(\alpha, \delta)$, and the corresponding statistical significance of the anisotropy are computed as given by

$$\delta I(\alpha, \delta) = \frac{N_{\mathrm{obs}} - N_{\mathrm{exp}}}{N_{\mathrm{exp}}}, \quad S(\alpha, \delta) = \frac{\delta I(\alpha, \delta)}{\sigma_{\delta I}}, \tag{2.11}$$

where S is the simple z-score, $\sigma_{\delta I} \simeq 1/\sqrt{N_{\mathrm{exp}}}$ gives the uncertainty (standard deviation) of the relative-intensity map, which is obtained from the fluctuation (Poisson) of N_{obs}, i.e., $\sigma_{\delta I} \simeq \sqrt{\mathrm{Var}(N_{\mathrm{obs}})}/N_{\mathrm{exp}}$, and $\mathrm{Var}(N_{\mathrm{obs}}) \simeq \sqrt{N_{\mathrm{obs}}}$. The significance map $S(\alpha, \delta)$ is the simplest form, defined as the direct deviation of event distributions from the null-hypothesis (isotropy). This helps identify regions in the sky with excesses or deficits in the cosmic-ray flux compared to an isotropic background in a simple way.

For cases with limited statistics, the background in the signal (ON) region is not precisely known and must be estimated from data. It is standard to use the Li–Ma likelihood significance [276], which defines a separate background (OFF) region, chosen such that it contains only background events and no signal. The event counts in the ON and OFF regions are treated as independent Poisson variables that share the same background rate, scaled by an exposure factor α. This method gives properly calibrated p-values, unlike simple Gaussian estimates that assume a known background. The significance S is

$$S = \sqrt{2}\sqrt{N_{\mathrm{on}} \ln\left[\frac{1+\alpha}{\alpha}\frac{N_{\mathrm{on}}}{N_{\mathrm{on}} + N_{\mathrm{off}}}\right] + N_{\mathrm{off}} \ln\left[(1+\alpha)\frac{N_{\mathrm{off}}}{N_{\mathrm{on}} + N_{\mathrm{off}}}\right]}, \tag{2.12}$$

where N_{on} is the number of observed events in the signal region, N_{off} is the number

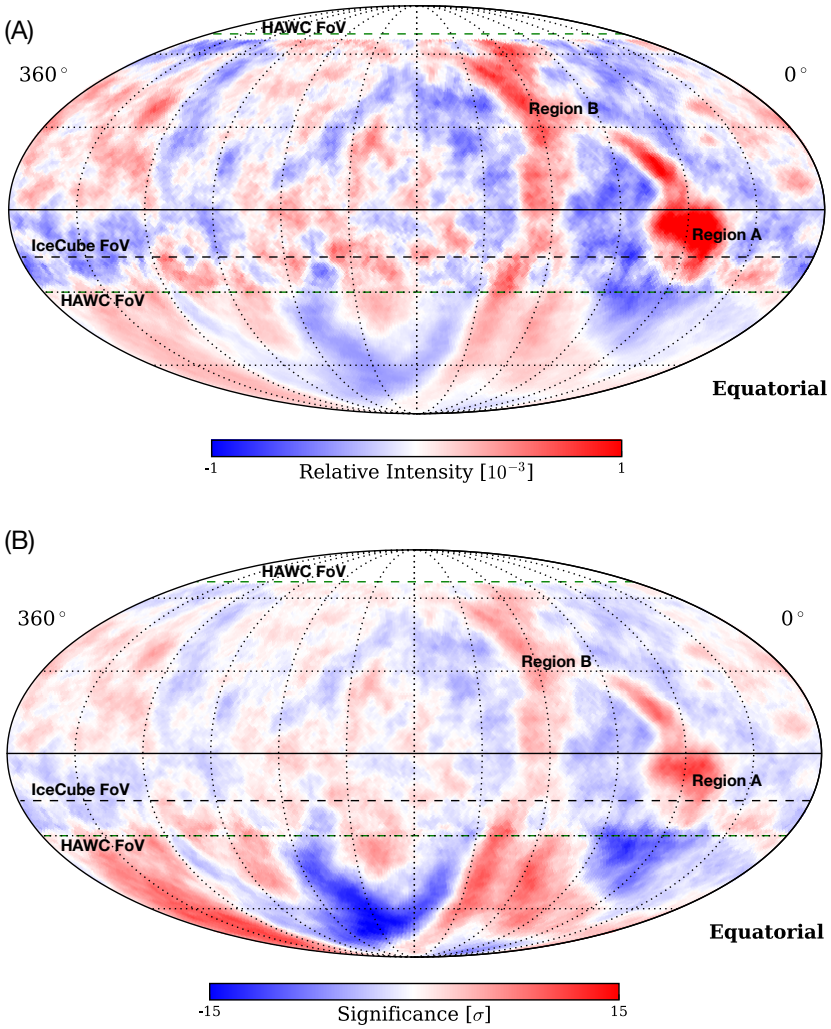


Figure 2.1: Example of combined IceCube and HAWC sky maps after subtracting the multipole fit from the large-scale map at a median energy of 10 TeV. Panel (A) shows the relative intensity $\delta I(\alpha, \delta)$, and panel (B) shows the corresponding signed statistical Li-Ma significance $S(\alpha, \delta)$. Figure adapted from [16].

of events in the background region, and α is the ratio of their exposures. The Li–Ma significance assumes Poisson statistics and quantifies the likelihood that the observed excess or deficit arises from random fluctuations rather than a true signal. These values represent local significance, meaning they do not account for the effect of scanning many sky positions. When multiple regions are tested, random fluctuations may appear significant by chance, a phenomenon known as the look-elsewhere effect. The global significance can be approximated as $p_{\text{global}} = 1 - (1 - p_{\text{local}})^{N_{\text{eff}}}$, where $N_{\text{eff}} \simeq \Delta\Omega_{\text{FoV}}/\Delta\Omega_{\text{bin}}$ is the number of trials, given by the ratio of the solid angle of the FoV to the top-hat smoothing scale.

This statistical method is widely used to assess the significance of excesses or deficits in anisotropies and source searches. It has been applied to γ-ray source identification by H.E.S.S. [277], and to cosmic-ray source searches by HAWC [136, 16] and the Pierre Auger Observatory [278, 279]. Figure 2.1 shows the combined IceCube and HAWC sky maps at a median energy of 10 TeV, which illustrates the residual small-scale anisotropy after

removing the large-scale components up to $\ell = 3$, revealing structures smaller than about 60°, displaying $\delta I(\alpha, \delta)$ and the corresponding significance $S(\alpha, \delta)$ [16]. A clear excess appears near $\alpha = 50°$, known as Region A, first reported by the Milagro Collaboration [280]. It extends roughly from $(54°, -16°)$ to $(78°, 18°)$. Region B marks the boundary between the excess and deficit areas in the northern sky [16].

2.1.4 Fit of Anisotropies

A simple way to obtain large-scale anisotropies from the distribution (a sky map using HEALPix [281]) of cosmic-ray arrival directions is to perform a projection on the sky map. Once we obtain a relative intensity map, $\delta I(\alpha, \delta)$, we can extract the dipole and multiple information by projecting the sky map onto the equatorial plane. By doing so, the two-dimensional map $\delta I(\alpha, \delta)$ becomes a one-dimensional map $\delta I(\alpha)$. Specifically, we bin α into n bins, then average the value of δI for all pixels within each α bin to obtain $\delta\overline{I}(\alpha)$. To quantify these deviations, the observed $\delta I(\alpha, \delta)$ distribution is often fitted with a combination of dipole and higher-order multipoles, representing different angular scales of anisotropy [138]. The fit of n different $\delta\overline{I}(\alpha)$ points can be written as a function of α,

$$I(\alpha) = 1 + \sum_{n=1}^{\infty} A_n \cos(\alpha - \phi_n), \tag{2.13}$$

where 1 represents the monopole ($n = 0$), $I = 1 + \delta\overline{I}$, A_n is the harmonic amplitude, and ϕ_n is the phase [5], assuming a full-sky coverage. The dipole fit ($n = 1$) gives a global measure of the anisotropy, while the multiple expansion reveals fine structures in the cosmic-ray distribution. The above expression can also be written in an equivalent sine–cosine (Fourier) form, known as the Rayleigh analysis, which will be introduced in a later section. The average uncertainty $\sigma_{\delta I}$ follows standard error propagation in the form $\sigma_{\delta I} = \sqrt{(1 + 1/k)N_{\rm exp}}$, where k is the number of scramblings used for the reference maps. For $k = 20$, the correction is only about +2.5%. This approach has been applied in experiments, such as IceCube and HAWC, to study both large- and small-scale anisotropies in cosmic-ray arrival directions [63, 282, 283, 88, 139, 136, 140].

Many observatories today can reconstruct the dipole anisotropy directly from two-dimensional sky maps of relative intensity $I(\alpha, \delta)$. However, some experiments still use the right ascension of the above projection method, where the partial-sky coverage will lead to a "fake" dipole amplitude. Now, we apply a new form for the projection as [5]

$$\tilde{A}_1 e^{i\tilde{\alpha}_1} = \frac{1}{\pi} \int_0^{2\pi} d\alpha\, e^{i\alpha} \left[\frac{1}{s_2 - s_1} \int_{s_1}^{s_2} ds\, I(\alpha, \delta) \right], \tag{2.14}$$

where $s = \sin\delta$ and the limits $s_{1,2} = \sin\delta_{1,2}$ define the declination range $[\delta_1, \delta_2]$ covered by the detector. If the relative intensity follows a pure dipole, this projection gives the correct dipole phase, $\tilde{\alpha}_1 = \alpha_1$, but the amplitude is reduced by a geometric factor,

$$\tilde{A}_1 \simeq \frac{\delta_1 - \delta_2 + c_1 s_1 - c_2 s_2}{2(s_1 - s_2)} A_1, \tag{2.15}$$

where A_1 and α_1 are the true dipole amplitude and phase, and $c_i = \cos\delta_i$. In most cases, the measured projection amplitude $\tilde{A}_1$ is reported without this correction, which generally leads to an underestimate of the true dipole strength by a factor of order one [5]. This partially explains the scatter in reported dipole amplitudes between different experiments. Moreover, this one-dimensional projection can also mix in contributions from higher multipoles ($\ell > 1$), adding further uncertainty to the reconstructed dipole amplitude [5].

2.1.5 Harmonic Analysis

The relative intensity of cosmic rays observed across the sky can be regarded as a scalar function defined on the celestial sphere and expressed in the usual spherical coordinate system. Any such angular function can be decomposed into spherical harmonic modes, i.e., spherical-harmonic components $Y_{\ell m}(\theta, \phi)$. In this standard representation, the angular position of each sky direction is specified by the polar angle θ and the azimuthal angle ϕ. These spherical coordinates are related to equatorial coordinates (δ, α) by $\theta = \pi/2 - \delta$ and $\phi = \alpha$. For convenience, we therefore introduce the notation,

$$Y_{\ell m}(\delta, \alpha) \equiv Y_{\ell m}(\theta = \pi/2 - \delta, \alpha), \tag{2.16}$$

which does not redefine spherical harmonics. The anisotropy (i.e., the relative intensity $I(\delta, \alpha)$) can then be expressed as the cumulative sum of contributions from all scales,

$$\begin{aligned} I(\alpha, \delta) =& \Gamma_{00} Y_{00}(\delta, \alpha) + \Gamma_{10} Y_{10}(\delta, \alpha) + \Gamma_{1,-1} Y_{1,-1}(\delta, \alpha) + \Gamma_{11} Y_{11}(\delta, \alpha) \\ &+ ... + \Gamma_{\ell,-\ell} Y_{\ell,-\ell}(\delta, \alpha) + \Gamma_{\ell\ell} Y_{\ell\ell}(\delta, \alpha), \end{aligned} \tag{2.17}$$

where the degree $\ell \to \infty$ depends on the anisotropy scale present in the observations or the underlying assumptions, $\Gamma_{\ell m}$ representing the coefficient corresponding to each moment of the degrees of scale structures. The functions $Y_{\ell m}(\delta, \alpha)$ are derived from the associated Legendre polynomials. Specifically,

$$Y_{\ell m}(\delta, \alpha) = \sqrt{\frac{(2\ell + 1)}{4\pi} \frac{(\ell - m)!}{(l + m)!}} P_\ell^m(\sin \delta)\, e^{im\alpha}. \tag{2.18}$$

For $\ell = 0$, we have $Y_{00} = 1/\sqrt{4\pi}$. For $\ell = 1$, the expressions are given as follows:

$$Y_{1,-1} = \sqrt{\frac{3}{8\pi}} \cos \delta e^{-i\alpha}, \quad Y_{10} = \sqrt{\frac{3}{4\pi}} \sin \delta, \quad Y_{1,1} = -\sqrt{\frac{3}{8\pi}} \cos \delta e^{i\alpha}. \tag{2.19}$$

For higher-degree harmonics, explicit expressions are listed in Appendix A. More generally, $I(\delta, \alpha)$ in Eq. (2.17) admits a complete spherical-harmonic expansion, starting with the monopole term ($\ell = 0$) and extending, in principle, to infinitely high multipole orders:

$$I(\alpha, \delta) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} a_{\ell m}\, Y_{\ell m}\,(\delta, \alpha)\,, \tag{2.20}$$

where $a_{\ell m}$ are the spherical harmonic coefficients, which quantify the strength of each multipole mode, i.e., for every degree ℓ and order m, they specify the amplitude and phase of the corresponding spherical-harmonic component $Y_{\ell m}(\delta, \alpha)$. Note that $\Gamma_{\ell m}$ in Eq. (2.17) is actually equivalent to $a_{\ell m}$. Together, these coefficients describe the full angular structure of the cosmic-ray anisotropy. Considering the well-defined relative intensity as shown in Eq. (2.20), the harmonic coefficient is expressed as

$$a_{\ell m} = \int_0^{2\pi} \int_{-\pi/2}^{\pi/2} \delta I(\alpha, \delta)\, Y_{\ell m}^* \cos \delta \, \mathrm{d}\delta \, \mathrm{d}\alpha, \tag{2.21}$$

where the complex conjugate follows the relation $Y_{\ell -m}(\delta, \alpha) = (-1)^m Y_{\ell m}^*(\delta, \alpha)$, ℓ' ranges from 0 to infinity when handling integrals of spherical harmonics over real domains, any imaginary components that arise from the complex exponential should either cancel out or be disregarded, as they do not contribute to the final real-valued outcome.

The coefficients $a_{\ell m}$ are determined by projecting the relative-intensity map $\delta\mathcal{I}(\delta, \alpha)$ onto the basis functions $Y_{\ell m}$. This projection is defined through the usual inner product on the

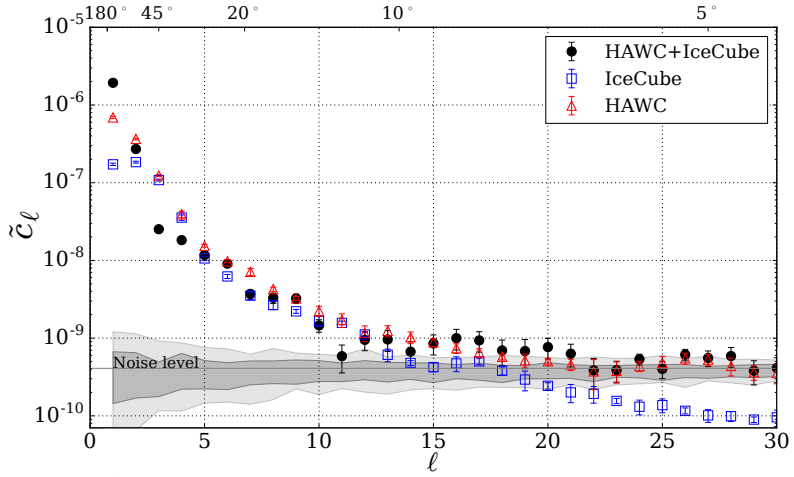


Figure 2.2: Combined (IceCube+HAWC) angular power spectrum of the cosmic ray anisotropy at 10 TeV. Figure taken from [16].

sphere, where the function is multiplied by the complex conjugate of the spherical harmonic and integrated over the full solid angle. Through this process, the projection extracts the "amount" of each spherical harmonic $Y_{\ell m}(\delta, \alpha)$ that is present within the function $I(\delta, \alpha)$. Each coefficient $a_{\ell m}$ carries both amplitude and phase information regarding both the amplitude and phase of the corresponding spherical harmonic component within $I(\delta, \alpha)$ and indicates the weight or significance of a particular spherical harmonic in the function's expansion. Larger values of $a_{\ell m}$ imply that the corresponding spherical harmonic $Y_{\ell m}(\delta, \alpha)$ contributes more significantly to the overall function. The angular power spectrum, in this context, represents the average power of all orders m for a specific degree (ℓ), i.e.,

$$\mathcal{C}_\ell = \frac{1}{2\ell + 1} \sum_{m=-\ell}^{\ell} |a_{\ell m}|^2, \tag{2.22}$$

where $2\ell+1$ is the number of independent modes for a given multipole ℓ. For a full-sky case, the variance of the spectrum $\mathcal{C}_\ell$ is $\mathrm{Var}(\mathcal{C}_\ell) = 2\mathcal{C}_\ell^2/(2\mathcal{C}+1)$, and the uncertainty (standard error) is obtained as $\sigma_{\mathcal{C}_\ell} = \sqrt{\mathrm{Var}(\mathcal{C}_\ell)} = \mathcal{C}_\ell\sqrt{2/(2\ell+1)}$.

An illustrative example is provided in Figure 2.2, which shows the combined IceCube–HAWC angular power spectrum of the cosmic-ray anisotropy at a median energy of 10 TeV. In the context of cosmic-ray relative-intensity studies, the monopole term simply corresponds to the overall isotropic mean flux. Since the relative intensity is defined regarding this mean, the monopole contribution vanishes, i.e., $I_\mathrm{m}(\alpha, \delta) = 0$, which directly implies that the monopole coefficient satisfies $a_{00} = 0$. Now, we assume that the relative intensity, denoted by $\mathcal{I}_\mathrm{d}(\delta, \alpha)$, is well-described as a function of dipole distribution only, specifically expressed as $\delta I_\mathrm{d} = a_{10}Y_{10} + a_{1,-1}Y_{1,-1} + a_{11}Y_{11}$. Then, $\mathcal{C}_1$ can be obtained. The dipole's amplitude is encapsulated within the components a_{10}, $a_{1,-1}$, and a_{11}. Now, let us consider our description of the dipole,

$$\delta I_\mathrm{d}(\alpha, \delta) = \mathcal{A} \sin\delta \sin\delta_\mathrm{d} + \mathcal{A} \cos\delta \cos\delta_\mathrm{d} \cos(\alpha - \alpha_\mathrm{d}). \tag{2.23}$$

where $(\alpha_\mathrm{d}, \delta_\mathrm{d})$ is the dipole's orientation. The amplitude $a_{\ell m}$ is then rewritten as

$$a_{10} = \sqrt{\frac{4\pi}{3}}\,\mathcal{A}\,\sin\delta_\mathrm{d}, \quad a_{11} = -\sqrt{\frac{2\pi}{3}}\,\mathcal{A}\,\cos\delta_\mathrm{d}\,e^{+i\alpha_\mathrm{d}}, \quad a_{1,-1} = \sqrt{\frac{2\pi}{3}}\,\mathcal{A}\,\cos\delta_\mathrm{d}\,e^{-i\alpha_\mathrm{d}}. \tag{2.24}$$

These coefficients describe the directional dependence of the dipole: a_{10} corresponds to the contribution along the declination axis, while $a_{1,\pm1}$ represent the dipole projection in right ascension, carrying the phase information through the complex exponentials $e^{\pm i\alpha_d}$. The amplitude of the multipole at degree ℓ is then obtained from the harmonic coefficients

$$\mathcal{A}_\ell = \sqrt{\sum_{m=-\ell}^{\ell} |a_{\ell m}|^2} = \sqrt{(2\ell+1)\,\mathcal{C}_\ell}, \tag{2.25}$$

$\mathcal{A}_1 = \sqrt{3\,\mathcal{C}_1}$ for dipole. For the case of a partial-sky-coverage sky map, one can directly apply Eq. (2.15) for a dipole. For the cases of $\ell > 1$, one can use the MASTER algorithm [284], which relates the observed pseudo-power spectrum (partial-sky) to the true sky spectrum through a coupling matrix.

To reveal smaller-scale features in the data, one can model the dominant large-scale pattern with a monopole, dipole, and quadrupole, then remove that model from the relative-intensity map and study the residual. A convenient approach, widely used in CMB analyses (see, e.g., [285]) and adopted in cosmic-ray studies, expresses the fractional intensity map $\delta I(\alpha, \delta)$ in a real spherical harmonic basis [286]:

$$\begin{aligned}\delta I(\alpha,\delta) \;&= m_0 + p_x\,\cos\delta\,\cos\alpha + p_y\,\cos\delta\,\sin\alpha + p_z\,\sin\delta \\ &\quad + Q_1\,(3\cos^2\delta - 1) + Q_2\,\sin(2\delta)\cos\alpha + Q_3\,\sin(2\delta)\sin\alpha \\ &\quad + Q_4\,\cos^2\delta\cos(2\alpha) + Q_5\,\cos^2\delta\sin(2\alpha),\end{aligned}$$

where m_0 is a constant offset (monopole), (p_x, p_y, p_z) are the dipole components, and $(Q_1, \ldots, Q_5)$ are the five independent quadrupole coefficients. The angular coordinates are right ascension α and declination δ. For fitting, one can select the analysis region (sky pixels within the experiment's field of view). Then, fit Eq. (2.26) to the relative-intensity map by least squares (or maximum likelihood with Poisson weights) [286]. Next, build the large-scale model from the best-fit parameters and subtract it from the data to obtain a residual map. Finally, evaluate goodness of fit (e.g., χ^2/ndf or likelihood ratio). If significant residual structure remains, higher multipoles ($\ell \geq 3$) are required. The dipole and quadrupole usually describe most of the small anisotropies at the 10^{-3} level, although the quadrupole can sometimes be as strong as or stronger than the dipole, depending on the detector's sky coverage and acceptance [286]. A poor fit probability after removing $\ell = 1, 2$ indicates unresolved structure and motivates extending the model (e.g., octuple or direct angular–power–spectrum methods) to describe the residual anisotropy [286]. For higher multipoles $\ell \geq 3$, the residual structures can provide additional insights into the influence of the local magnetic field and possible point-like sources, as shown in Figure 2.2 [16].

2.2 Other Reconstruction Methods

Cosmic-ray anisotropy reconstruction employs several techniques to identify and quantify directional variations in cosmic-ray flux. Each method provides unique insights into different scales of anisotropy, from large-scale structures spanning the entire sky to smaller, more localized features. The Rayleigh analysis is effective for identifying large-scale dipole anisotropies by examining variations in right ascension. Large-scale features are described with sidereal harmonics such as the Rayleigh first harmonic or low-ℓ multipoles, while full-sky likelihood methods recover two-dimensional structures and medium-scale features. Besides, the exposure can also be handled through the East–West method or

fitted together with the sky in a maximum-likelihood approach that simultaneously solves for relative intensity and detector acceptance [143, 5], which leverages the Earth's rotation to detect asymmetries in cosmic-ray flux between the eastern and western hemispheres. The generalized maximum-likelihood method allows for detailed analysis of both large- and medium-scale anisotropies, combining data from multiple observatories, calibrating the detector's exposure automatically. In this section, we will review these methods and their specific applications in reconstructing cosmic-ray anisotropies.

2.2.1 Rayleigh Analysis

The Rayleigh analysis is an effective method for reconstructing large-scale anisotropies in cosmic-ray arrival directions. This method is closely linked to discrete Fourier analysis, where the arrival direction of cosmic rays, given by the right ascension α, is interpreted as the polar angle in the xy-plane. The goal of the Rayleigh analysis is to extract the first harmonic modulation from the distribution of these arrival directions, which helps identify any dipole anisotropy in the cosmic-ray flux. The advantage of Rayleigh is that it reduces errors that come from small scales, thus it is good for large-scale anisotropies, such as a dipole.

Specifically, the distribution of cosmic-ray arrival directions in right ascension can be described by a periodic function $f(\alpha)$, which can be expanded into a Fourier series as

$$f(\alpha) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \left(a_k \cos(k\alpha) + b_k \sin(k\alpha)\right). \tag{2.26}$$

where a_k and b_k are the Fourier coefficients. For cosmic-ray anisotropy studies, we are primarily interested in the first harmonic term ($k = 1$), which represents the dipole component. Then, a_1 and b_1 are given by the integrals [1]:

$$a_1 = \frac{1}{\pi} \int_0^{2\pi} f(\alpha) \cos(\alpha)\, d\alpha, \quad b_1 = \frac{1}{\pi} \int_0^{2\pi} f(\alpha) \sin(\alpha)\, d\alpha. \tag{2.27}$$

In the simplest case, where we assume isotropy (i.e., no directional dependence in cosmic-ray arrival directions), the probability distribution $f(\alpha)$ is taken to be constant. Under this assumption, and in the context of real-world cosmic-ray data, where the number of events N is finite, these coefficients are expressed in discrete terms, and the sums over $\cos(\alpha_k)$ and $\sin(\alpha_k)$ in the Fourier series expansion simplify to [131]:

$$a_1 = \frac{2}{N} \sum_{k=1}^{N} \cos(\alpha_k), \quad b_1 = \frac{2}{N} \sum_{k=1}^{N} \sin(\alpha_k). \tag{2.28}$$

These discrete sums arise by replacing the continuous integrals with sums over individual events, where each right ascension value α_k corresponds to the observed arrival direction of the k-th cosmic-ray particle. Utilizing these values, the Rayleigh method identifies the amplitude A_{D} and phase α_{D} of the first harmonic as the key indicators of anisotropy, which indicate the strength and direction of any detected anisotropy. These are computed using the Fourier coefficients a_1 and b_1 as follows [131]:

$$A_{\mathrm{D}} = \sqrt{a_1^2 + b_1^2}, \quad \alpha_{\mathrm{D}} = \arctan2\left(\frac{\mathrm{b_1}}{\mathrm{a_1}}\right), \tag{2.29}$$

where the amplitude A_{D} represents the strength of the dipole anisotropy, while the phase α_{D} points to the right ascension of maximum cosmic-ray intensity. This direction provides insight into the source or modulation effect causing the anisotropy. Given that cosmic-ray

anisotropy studies rely on large datasets, one can estimate the uncertainties in both the amplitude and the phase. The uncertainties for the coefficients a_1 and b_1, and thus for $A_{\rm D}$ and $\alpha_{\rm D}$, depend on the total number N of observed events. The uncertainty in the Fourier coefficients is given by [131, 88, 89], $\sigma_{a_1} = \sigma_{b_1} \simeq \sqrt{2/N}$, which leads to uncertainties in the amplitude and phase, and are expressed as:

$$\sigma(A_{\rm D}) \simeq \sqrt{\frac{2}{N}}, \quad \sigma(\alpha_{\rm D}) \simeq \frac{1}{A_{\rm D}}\sqrt{\frac{2}{N}}, \tag{2.30}$$

and the Rayleigh p-value is $P(> A_{\rm D}) = \exp[-NA_{\rm D}^2/4]$. Thus, the larger the amplitude, the smaller the uncertainty in the phase measurement. Worth noting that the Rayleigh analysis is insensitive to structure orthogonal to the RA first harmonic. Meanwhile, small-scale features that modulate RA can leak into the dipole ($k = 1$) case.

2.2.2 East-West Method

The East–West (EW) method provides an exposure-independent way to measure large-scale anisotropies in right ascension by exploiting the approximate east–west symmetry of ground-based arrays [287]. Events are divided into two azimuthal sectors, i.e., the east and west. For each zenith range, the difference between the counting rates from both sectors removes common exposure effects such as atmospheric and detector variations, leaving a residual signal proportional to the derivative of the cosmic-ray intensity with respect to right ascension. If the acceptance is nearly symmetric and stable in time, this difference isolates the celestial modulation, where small asymmetries (e.g., from local obstacles or slow drifts) can be tested through stability checks.

Specifically, for each event with time t_i and azimuth ϕ_i, the right ascension of the local zenith is given by $\alpha_0(t_i)$. Events are recorded as coming from the east if $0 < \phi_i < \pi$ and from the west if $-\pi < \phi_i < 0$. The harmonic coefficients of the East–West modulation are then computed as the summations of all the events that fall in the corresponding sectors,

$$a_{\rm EW} = \frac{2}{N}\sum_{i=1}^{N}\cos\big[\alpha_0(t_i) - \xi_i\big], \qquad b_{\rm EW} = \frac{2}{N}\sum_{i=1}^{N}\sin\big[\alpha_0(t_i) - \xi_i\big], \tag{2.31}$$

where $\alpha_0(t_i)$ is the R.A. of the zenith at time t_i (the local sidereal time), and $\xi_i = 0$ for eastward events and π for westward events. The amplitude and phase of the right-ascension modulation are obtained from the EW coefficients as [287]

$$r_{\rm EW} = \sqrt{a_{\rm EW}^2 + b_{\rm EW}^2}, \quad \varphi_{\rm EW} = \arctan2\big(b_{\rm EW}, a_{\rm EW}\big). \tag{2.32}$$

They relate to the physical first-harmonic in R.A. as [287]

$$r = \frac{\pi}{2}\frac{\langle\cos\delta\rangle}{\langle\sin\theta\rangle}\, r_{\rm EW}, \quad \varphi = \varphi_{\rm EW} + \frac{\pi}{2}, \tag{2.33}$$

where $\langle\cos\delta\rangle$ and $\langle\sin\theta\rangle$ represent event-averaged factors that account for the detector's field of view. Measuring an EW amplitude equal to or greater than $r_{\rm EW}$ for an isotropic distribution is given by the so-called Rayleigh probability:

$$P(\geq r_{\rm EW}) = \exp\!\left(-\frac{N\, r_{\rm EW}^2}{4}\right). \tag{2.34}$$

This directly estimates the equatorial component of the dipole without requiring an explicit exposure model. It has been widely applied in large-scale anisotropy studies by experiments such as ARGO-YBJ, IceCube, and the Pierre Auger Observatory [287, 88]. The EW

method remains particularly valuable at low energies, where the trigger efficiency varies with time or azimuth, since such effects largely cancel in the east–west difference.

Bonino et al. [287] were the first to introduce this technique for analyzing the equatorial dipole component in a way that minimizes the dependence on detailed exposure modeling and therefore avoids many detector-related systematics. The Auger experiment applied the EW method to long-term data to control exposure systematics, finding a significant dipole above $\sim 8\,\mathrm{EeV}$ with an amplitude of $\sim$6% and a stable phase, while using EW below 1 EeV where trigger efficiency varies [88]. Overall, the method is suitable for large-scale anisotropy studies and complements full-sky maximum-likelihood or Rayleigh analyses [3].

2.2.3 East-West Derivative Method

The standard East-West (EW) count difference gives a direct variation of anisotropy (dipole only), but can be biased by slow variations in detector acceptance. The EW derivative method instead estimates $\mathrm{d}I/\mathrm{d}\alpha$ (the RA gradient), which is much less sensitive to acceptance drifts and therefore yields a cleaner first-harmonic (Rayleigh) reconstruction of the dipole in the equatorial plane [287, 288]. From the first harmonic of the EW signal, one obtains the dipole amplitude and phase in RA.

Using the collected cosmic-ray event data, the flux can be expressed in a general form,

$$\phi(\alpha, \delta) = \phi_0\, I(\alpha, \delta), \tag{2.35}$$

where ϕ_0 is the angular-averaged flux, which represents a monopole component in the flux (isotropic). The intensity, denoted by $\phi(\alpha, \delta)$, is a function of right ascension α and declination δ, while $\delta I = I - 1$ represents anisotropy. The intensity is defined as

$$I(t, \theta, \varphi) = 1 + \delta I(t, \theta, \varphi), \tag{2.36}$$

where δI is the relative intensity, which is always in the level $10^2 - 10^3$ for the ground-based observatories for cosmic-ray detections. Assuming the detector exposure $\mathcal{E}$.

At a given sidereal time t, under the local coordinate system, the arrival directions of cosmic rays can be naturally separated into two parts, covering the east ($0 < \varphi < \pi$) and west ($-\pi < \varphi < 0$) sectors. The number of events observed within a short sidereal time interval Δt in the east (+) and west (-) sectors can be expressed as

$$N_{\pm}(t) = \phi_0\, \Delta t\, E(t) \int_0^{\pi} \mathrm{d}\varphi \int_0^{\theta_{\max}} \mathrm{d}\theta\, \sin\theta\, \mathcal{A}(\pm\varphi, \theta)\, I(t, \pm\varphi, \theta), \tag{2.37}$$

where $\sin\theta$ is the geometric factor, and $\mathcal{E}$ is decomposed into the time-dependent exposure $E(t)$ and the acceptance $\mathcal{A}$. The detector acceptance $\mathcal{A}$ can be written as $\mathcal{A} = \mathcal{A}^0(1 + \delta\mathcal{J})$, where $\mathcal{A}^0$ is the average acceptance and $\delta\mathcal{J}$ is the deviation of the acceptance in the east or west sector from $\mathcal{A}^0$. Therefore, Eq. (2.37) can be rewritten as

$$\begin{aligned} N_{\pm}(t) &= \phi_0\, \Delta t\, E(t) \int_{\Omega_{\pm}} \mathcal{A}^0_{\pm}(1 + \delta\mathcal{J}(\pm\varphi, \theta))\,(1 + \delta I(t, \pm\varphi, \theta))\, \mathrm{d}\Omega, \\ &= \phi_0\, \Delta t\, E(t)\, \mathcal{A}^0_{\pm} \left(\int_{\Omega_{\pm}} \mathrm{d}\Omega + \int_{\Omega_{\pm}} \delta\mathcal{J}\, \mathrm{d}\Omega + \int_{\Omega_{\pm}} \delta I\, \mathrm{d}\Omega + \int_{\Omega_{\pm}} \mathcal{J}\, \delta I\, \mathrm{d}\Omega \right), \end{aligned} \tag{2.38}$$

where $\Omega_{\pm}$ denotes the solid angle of the east and west sectors, and $\mathcal{A}_{\pm}$ their corresponding acceptances. The east-west asymmetry at sidereal time t is then defined as

$$A_{\mathrm{EW}}(t) \equiv \frac{N_+(t) - N_-(t)}{N_+(t) + N_-(t)}. \tag{2.39}$$

Because the average acceptance $\mathcal{A}^0$ is even for the two sectors, i.e., $\mathcal{A}^0_+ = \mathcal{A}^0_-$, and the deviation $\delta\mathcal{J}$ is odd, i.e., $\delta\mathcal{J}_+ = -\delta\mathcal{J}_-$, inserting Eq. (2.38) into Eq. (2.39) yield

$$A_{\rm EW}(t) = \frac{\int_+ \delta\mathcal{J}\mathrm{d}\Omega - \int_- \delta\mathcal{J}\mathrm{d}\Omega + \int_+ \delta I\mathrm{d}\Omega + \int_- \delta I\mathrm{d}\Omega + \mathcal{O}_1(\delta\mathcal{J}\delta I)}{2\Omega + \mathcal{O}_2(\delta\mathcal{J}\delta I)}, \tag{2.40}$$

where $\Omega \equiv \Omega_+ = \Omega_-$. The deviation of acceptance follows $\delta\mathcal{J} \ll 1$ and $\delta I \ll 1$ for ground-based observatories, thus, $\mathcal{O}_1(\delta\mathcal{J}\delta I)$ and $\mathcal{O}_2(\delta\mathcal{J}\delta I)$ are nothing but the second-order "cross" terms, which approximately zero. Therefore, Eq. (2.40) is reduced to

$$A_{\rm EW}(t) = \langle\delta\mathcal{J}\rangle + \frac{1}{2}\left(\langle\delta I(t,\varphi,\theta)\rangle - \langle\delta I(t,-\varphi,\theta)\rangle\right). \tag{2.41}$$

Based on the relation between local sidereal time t and right ascension α, i.e., $\alpha = t - H$, as shown in Eq. (2.6), one can convert t to α in the data. Every particle from azimuth φ in the east sector is sampling the sky at R.A. $\alpha + \delta\alpha(\varphi,\theta)$, and the West sector at $\alpha - \delta\alpha(\varphi,\theta)$, where $\delta\alpha \ll 1$ is the small EW hour-angle shift is within a sidereal time bin. The first order Taylor expansion at each δI around the central R.A. α is $\delta I(\alpha \pm \delta\alpha) \approx \delta I(\alpha) \pm \delta\alpha\partial_\alpha\delta I(\alpha)$. When averaging $\delta I(\alpha \pm \delta\alpha)$ over the east and west sectors,

$$\langle\delta I(\alpha \pm \delta\alpha)\rangle \approx \langle\delta I(\alpha)\rangle \pm \langle\delta\alpha\partial_\alpha\delta I(\alpha,0)\rangle, \tag{2.42}$$

where $\delta I(\alpha,0)$ corresponds to the anisotropy evaluated at the central azimuth ($\varphi = 0$) where the shift is measured. We simply replace all those tiny, direction-dependent shifts $\delta\alpha(\theta,\varphi)$ by one "effective" shift, $\Delta\alpha$. Thus, one has

$$A_{\rm EW}(t) = \langle\delta\mathcal{J}\rangle + \Delta\alpha\,\partial_\alpha\delta I(\alpha,0), \tag{2.43}$$

where $\langle\Delta\mathcal{J}\rangle$ denotes the sector-averaged East–West acceptance asymmetry, which arises because no ground-based detector is ever perfectly symmetric with respect to azimuth, small odd-in-azimuth differences in effective area or trigger efficiency always remain and therefore do not cancel in the EW subtraction. The effective hour-angle offset $\Delta\alpha$ represents the acceptance-weighted mean eastward projection of all reconstructed arrival directions and quantifies how the map's local azimuthal asymmetries result in a shift in right ascension. With these definitions, the differential function of the relative intensity is then expressed as $\partial_\alpha\delta I(\alpha,0) = \left(A_{\rm EW}(t) - \langle\delta\mathcal{J}\rangle\right)/\Delta\alpha$. In local (east–north–up) coordinates, for a direction specified by zenith angle θ and azimuth φ (measured from north increasing toward east) has unit-vector components $v_x = \sin\theta\sin\varphi$ points toward the east (which we denote $v_{\rm E} \equiv v_x$), $v_y = \sin\theta\cos\varphi$ toward the north and $v_z = \cos\varphi$ points upward (toward the zenith). Computing the average of $v_{\rm E}$ over all events in the east sector and subtracting the corresponding average in the west sector $v_{\rm W}$, one finds

$$\Delta\alpha = \frac{1}{2}\left(\langle\sin\theta\sin\varphi\rangle_{\rm E} - \langle\sin\theta\sin\varphi\rangle_{\rm W}\right) = \langle\sin\theta\sin\varphi\rangle, \tag{2.44}$$

since the small acceptance deviation $\delta\mathcal{J}$ is antisymmetric (odd) in azimuth between the east and west sectors, as noted earlier, the same odd symmetry carries over to the geometric eastward projection. Therefore, when averaging the eastward component $\sin\theta\,\sin\varphi$ over the west sector, one obtains the opposite of the value in the east sector, namely $\langle\sin\theta\sin\varphi\rangle_{\rm W} = -\langle\sin\theta\sin\varphi\rangle_{\rm E}$, which leads to the simplified expression for $\Delta\alpha$.

From a data-analysis perspective, the measured cosmic-ray events are initially grouped into $N_{\rm sid}$ discrete sidereal-time bins τ, and within each time bin, the sky is further divided into $N_{\rm pix}$ celestial bins i. Using this two-dimensional binning in time and local direction, one can then construct an estimator for the East–West asymmetry by explicitly comparing

the event counts in the corresponding eastward and westward subsets of local-angle bins. Specifically, this estimator is defined as

$$\widehat{A}_{\mathrm{EW},\tau} = \left(\sum_{i\in\mathcal{D}_+} n_{\tau i} - \sum_{i\in\mathcal{D}_-} n_{\tau i} \right) \Bigg/ \sum_i n_{\tau i}, \tag{2.45}$$

where $\mathcal{D}_\pm$ are the sets of all bins in the east (+) or west (-). The deviation of the EW asymmetry $\delta\mathcal{J}$ as Eq. (2.40) and (2.41) shown can be estimated as

$$\widehat{\langle \delta\mathcal{J} \rangle} = \frac{1}{N_{\mathrm{sid}}} \sum_\tau \widehat{A}_{\mathrm{EW},\tau}, \tag{2.46}$$

where the summation extends over all events in both the east and west sectors for each sidereal-time bin. The effective right-ascension step $\Delta\alpha$ is then determined from the acceptance-weighted average of the eastward projection of the arrival directions, which reflects the detector's geometrical response to an infinitesimal shift in hour angle, given by

$$\widehat{\Delta\alpha} = \frac{1}{2N_{\mathrm{sid}}} \sum_\tau \left(\Delta\alpha_{+,\tau} + \Delta\alpha_{-,\tau} \right), \tag{2.47}$$

where $\Delta\alpha_{+,\tau} = \langle \sin\theta_i \sin\varphi_i \rangle_+$ and $\Delta\alpha_{-,\tau} = -\langle \sin\theta_i \sin\varphi_i \rangle_-$ based on the fact that $\Delta\alpha_{\pm,\tau}$ is always positive. Consequently, within a given sidereal-time bin τ, the effective right-ascension step is about averaging over all pixels within the range, namely

$$\Delta\alpha_{\pm,\tau} = \sum_{i\in\mathcal{D}_\pm} n_{\tau,i} \sin\theta_i |\sin\varphi_i| \Big/ \sum_{i\in\mathcal{D}_\pm} n_{\tau,i}. \tag{2.48}$$

The leading-order statistical uncertainty in each sidereal time bin τ is given by

$$\sigma(\Delta\alpha \partial_\alpha \delta I(\alpha_\tau)) \simeq \left(\sum_\tau n_{\tau,i} \right)^{1/2}, \tag{2.49}$$

where, for notational simplicity, we restrict our consideration to the equatorial plane (i.e., $\varphi = 0$) and introduce $\delta I(\alpha_\tau) \equiv \delta I(\alpha_\tau, \varphi = 0)$. Assuming that the underlying anisotropy is dominated by a dipolar modulation in right ascension, we parametrize the relative intensity at sidereal time τ as $\delta I(\alpha_\tau) = A_{\mathrm{sid}} \cos(\alpha_\tau - \alpha_{\mathrm{d}})$, where A_{sid} is the amplitude of the equatorial dipole and α_{d} is its phase. To fit the data to this pure dipole, we perform

$$\partial_\alpha \delta I(\alpha_\tau) = \frac{\widehat{A}_{\mathrm{EW}}(\tau) - \widehat{\langle \delta\mathcal{J} \rangle}}{\widehat{\Delta\alpha}}, \tag{2.50}$$

at each time bin τ, where $\partial_\alpha \delta I(\alpha_\tau) = -A_{\mathrm{sid}} \sin(\alpha_\tau - \alpha_{\mathrm{d}})$ from model, and we then perform the least-squares fit of this form as given in Eq. (2.50).

It is important to emphasize that this method is sensitive only to anisotropy in the equatorial plane, ignores any azimuthal structure in each sidereal-time bin τ, and cannot detect higher-order or latitudinal anisotropy components [288].

2.2.4 Maximum-Likelihood Method

The generalized maximum-likelihood method developed by Ahlers et al. [143] is a statistical technique for reconstructing cosmic-ray anisotropies using data from multiple observatories with overlapping fields of view. This method provides an optimal solution for reconstructing the anisotropies of cosmic rays, especially for large-scale and medium-scale anisotropies across the sky. This method is simplified and summarized in paper [143, 16].

Because cosmic-ray arrival times are widely spread due to large diffusion, the flux can be treated as effectively continuous over the full observation period of a ground-based detector. For a fixed energy range, the differential flux (in units of $\mathrm{cm}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$) is:

$$\phi(\alpha, \delta) = \phi_{\mathrm{iso}}\, I(\alpha, \delta) \tag{2.51}$$

Here, ϕ_{iso} denotes the average isotropic flux level, while $I(\alpha, \delta)$ represents the relative intensity at celestial coordinates (α, δ). The deviation from isotropy is then defined as $\delta I = I - 1$, which is expected to be small, i.e., $|\delta I| \ll 1$ [288]. In the local horizontal coordinate system of the observatory, each cosmic-ray direction is specified by the azimuth angle φ, zenith angle θ, and the local sidereal time t. A unit vector $\mathbf{n}'(\theta, \varphi)$ in this frame transforms to equatorial coordinates through $\mathbf{n} = \mathbf{R}(t)\,\mathbf{n}'$, where $\mathbf{R}$ is a rotation matrix dependent on sidereal time t and the geographic latitude Φ of the detector [143]. The detector's field of view is limited to zenith angles $\theta < \theta_{\max}$, resulting in a declination band over time $\delta_{\min} = \max(-\pi/2, \Phi - \theta_{\max})$ and $\delta_{\max} = \min(\pi/2, \Phi + \theta_{\max})$.

The method assumes that the detector's exposure $\mathcal{E}(\theta, \phi, t)$, per solid angle (θ, ϕ) and sidereal time t, accumulated over many sidereal days, can be factored into an angular-integrated exposure $E(t)$ and a relative acceptance $A(\theta, \phi)$, which is normalized as

$$\mathcal{E}(\theta, \phi, t) \simeq E(t) A(\theta, \phi), \tag{2.52}$$

where $A(\theta, \phi)$ is normalized such that $\int A(\theta, \phi) d\Omega = 1$. It is valid when the detector's directional response changes slowly with time compared to its diurnal modulation. This expression assumes that the relative acceptance of the detector does not vary significantly with sidereal time [143, 16]. For a given observatory, the number of cosmic rays expected from an angular element $\Delta\Omega_i$ of the local coordinate sphere corresponding to coordinates (θ_i, ϕ_i) during a sidereal time interval Δt_τ is given by:

$$\mu_{\tau i} \simeq I_{\tau i} \mathcal{N}_\tau \mathcal{A}_i, \tag{2.53}$$

where $\mathcal{N}_\tau \equiv \Delta t \phi^{\mathrm{iso}} E(t_\tau)$ represents the expected number of isotropic events in sidereal time bin τ, and $I_i(\tau)$ is the relative intensity observed in the local coordinates during the same time bin [143]. The relative acceptance, defined as $\mathcal{A}_i \equiv \Delta\Omega_i \mathcal{A}(\theta_i, \varphi_i)$, represents the binned detector acceptance for a given angular element. To relate the observed sky in local coordinates to the celestial sphere, the relative intensity $I_i(\tau)$ is defined by a coordinate transformation that maps the local coordinates (θ_i, ϕ_i) to equatorial coordinates (α, δ) through a time-dependent transformation $R(\tau)$, which relates the unit vector n in the local coordinate system to the corresponding vector in the celestial system. Thus, the intensity can be modelled as $I_i(\tau) \equiv I(R(n))\mathbf{n}'(\Omega_i)$. Each bin (τ, i) corresponds to a small time interval (typically a few minutes) and a small solid-angle element of the local sky, within which the detector records $n_{\tau i}$ events. Since each detected cosmic ray is an independent event, the number of observed events in this bin is random. The Poisson distribution is the natural model for this counting process and is given by:

$$P(n_{\tau i} \mid \mu_{\tau i}) = \frac{e^{-\mu_{\tau i}} (\mu_{\tau i})^{n_{\tau i}}}{n_{\tau i}!}, \tag{2.54}$$

which quantifies how (un)likely our observed data $n_{\tau i}$ is, given the model prediction $\mu_{\tau i}$. If $\mu_{\tau i}$ is large, and we observe many fewer events, then the probability is low, which means a deficit. Conversely, if we observe more than expected, that also reduces the likelihood, possibly indicating an excess or a feature in the anisotropy. The likelihood of observing n_i cosmic rays from a pixel i is described by the product of Poisson probabilities for all time

bins τ, and the likelihood for the entire data set is given by [143]:

$$\mathcal{L}(n|I,\mathcal{N},\mathcal{A}) = \prod_{\tau i} \frac{(\mu_{\tau i})^{n_{\tau i}} e^{-\mu_{\tau i}}}{n_{\tau i}!}. \tag{2.55}$$

The likelihood is maximized when the predicted values $\mu_{\tau i}$ best explain the observed data $n_{\tau i}$ across all times and angular bins. The anisotropic cosmic-ray intensity $I(\theta,\phi)$ can be reconstructed by mmaximizing this likelihood function. To determine the best-fit values of $I(\theta,\phi)$, the likelihood ratio λ is calculated as the ratio of the likelihood for the null hypothesis of no anisotropy, ($I^{(0)}$ [143], $\mathcal{N}^{(0)}$, $\mathcal{A}^{(0)}$),

$$\lambda = \frac{\mathcal{L}(n|I,\mathcal{N},\mathcal{A})}{\mathcal{L}(n|I^{(0)},\mathcal{N}^{(0)},\mathcal{A}^{(0)})}\,, \tag{2.56}$$

with $I^{(0)}_{\mathfrak{a}} = 1$. The estimators for the isotropic expectation $\mathcal{N}_\tau$ and relative acceptance $\mathcal{A}_i$ are determined iteratively by the following relations [143]:

$$\mathcal{N}^{(0)}_\tau = \sum_i n_{\tau i}, \quad \mathcal{A}^{(0)}_i = \sum_\tau n_{\tau i} \Big/ \sum_{\kappa j} n_{\kappa j}\,, \tag{2.57}$$

subject to the normalization condition $\sum_i \mathcal{A}_i = 1$.

In the combined analysis of data from HAWC and IceCube, the likelihood function (2.55) is extended by incorporating individual detector exposures for each observatory, while assuming a common relative intensity across both. The total exposure $\mathcal{E}$ is generalized by summing over disjoint sky sectors, which together form the entire field of view [16]. Each sector corresponds to the integrated field of view of a detector, and the exposure for each sector is modelled as a product of its angular-integrated exposure $E^s(t)$ and the relative acceptance $\mathcal{A}^s(\varphi,\theta)$, specifically [143],

$$\mathcal{E}(t,\varphi,\theta) \simeq \sum_{\text{sector } s} E^s(t)\mathcal{A}^s(\varphi,\theta)\,. \tag{2.58}$$

where the values of I, $\mathcal{N}$, and $\mathcal{A}$ of the maximum likelihood ratio ($I^\star$, $\mathcal{N}^\star$, $\mathcal{A}^\star$), as (2.56) shown, must satisfy the following implicit equations [143]:

$$I^\star_{\mathfrak{a}} = \frac{\sum_\tau n_{\tau\mathfrak{a}}}{\sum_{s\kappa} \mathcal{A}^{s\star}_{\kappa\mathfrak{a}}\mathcal{N}^{s\star}_\kappa}, \quad \mathcal{N}^{s\star}_\tau = \frac{\sum_i w^s_i n_{\tau i}}{\sum_j \mathcal{A}^{s\star}_j I^\star_{\tau j}}, \quad \mathcal{A}^{s\star}_i = \frac{\sum_\tau w^s_i n_{\tau i}}{\sum_\kappa \mathcal{N}^{s\star}_\kappa I^\star_{\kappa i}}, \tag{2.59}$$

where the window function w^s_i indicates whether a pixel i belongs to sector s (with a value of 1 if it is in the sector, and 0 otherwise). The binned quantity $\mathcal{A}^s_{\tau\mathfrak{a}}$ represents the relative acceptance for the sector s in equatorial coordinates at the pixel $\mathfrak{a}$ during time bin τ [16]. Since these are nonlinear equations, they cannot be solved explicitly but must be approached iteratively to find the best-fit solution.

Unlike traditional methods like the East-West derivative approach, the maximum-likelihood reconstruction does not need to assume azimuthal symmetry but corrects for detector systematics directly and reconstructs the full two-dimensional anisotropy pattern on the sky. It also makes it possible to combine data from different observatories and check how significant features are across different angular scales. This approach is particularly advantageous when combining data from multiple observatories, like HAWC [136] and IceCube [288], which together provide near-complete sky coverage, which would remove the biases caused by partial sky coverage.

Chapter 3

Sensitivity to Cosmic-Ray Anisotropy

Cosmic-ray anisotropy presents a key challenge in understanding cosmic-ray origin and propagation, as previously discussed. At lower energies (below a few TeV), anisotropies formed by nearby sources and the structure of the local interstellar magnetic [15, 136, 16]. As the energy increases from the TeV to the PeV range, both the amplitude and the phase of the anisotropy evolve: a phase shift appears around tens to hundreds of TeV, and then approaches the knee, the measured amplitudes are at or below a few 10^{-3}, with several results giving upper limits rather than a clear increase. These features point to changes in transport and a possible role of nearby Galactic sources, but the interpretation remains uncertain [17]. Composition changes around the knee, as well as energy-dependent diffusion, further complicate the connection between anisotropy and sources. Progress, therefore, requires energy-resolved, full-sky combinations across experiments with careful control of systematics and declination coverage.

However, the lack of full sky coverage can introduce biases in the reconstruction of anisotropies, such as dipoles or multipoles. Statistically significant measurements of the projected dipole amplitude are missing in the energy range between 2 PeV and 8 EeV (gap). The Pierre Auger Observatory has not yet observed a significant dipole below 8 EeV [270], and KASCADE-Grande did not find evidence for a dipole anisotropy in the data with median energy from 2.7 PeV to 33 PeV [268, 87]. Before proposing optimized reconstruction approaches (Chapter 4) for anisotropies and performing analysis with observational data, we will first investigate the sensitivity of an experiment (or multiple experiments) to anisotropies within this energy gap, using simulations with specific statistical and energy parameters. These sensitivities are both statistically and energy dependent. Following this, we will develop a theoretical framework to estimate the sensitivity of the surface array to cosmic-ray anisotropies using statistical methods and theoretical models. Finally, we will compare the sensitivity results from both simulations and theoretical calculations.

In Section 3.1, we begin by presenting an example using the planned IceCube-Gen2 surface array, followed by a detailed study of air-shower reconstruction efficiencies with this experiment. Then, we will evaluate its sensitivity to anisotropies in the cosmic-ray arrival directions at different primary energies and zenith angles of injected artificial dipoles, using dipole sky maps generated from Monte Carlo simulations. In Section 3.2, we aim to propose a theoretical framework for efficiently estimating the sensitivity of a surface observatory, such as IceCube-Gen2, to cosmic-ray large-scale anisotropies. By focusing on the significance function of the estimated one-dimensional projected dipole amplitude, we can develop equations to quantify uncertainties, allowing for an accurate assessment of dipole detection capabilities, namely, the sensitivity to dipole anisotropy. The consistency of the theoretical calculations with Monte Carlo simulation results reinforces the reliability of our approach. Additionally, the methodology for reconstructing the dipole phase and its significance was introduced. This work provides a basis for understanding the sensitivity of cosmic-ray observatories to anisotropic patterns in arrival directions, and can offer a quick estimate of their dipole sensitivity without the need for Monte Carlo simulations.

3.1 IceCube-Gen2 Surface Array

In this section, we present an analysis of the surface array's zenith-dependent and energy-dependent sensitivity to cosmic-ray anisotropies by examining the detector's response, specifically the cosmic-ray reconstruction efficiencies of the surface array, to primary particles such as protons and iron nuclei across a wide range of energies, and by reconstructing their arrival directions. The results of these simulations also suggest details of the array's design to improve its sensitivity to cosmic-ray anisotropies.

3.1.1 The Experiment

IceCube-Gen2 is a planned next-generation extension of the existing IceCube Neutrino Observatory [83], which is located at the geographic South Pole in Antarctica, one of the most stable and radio-quiet environments on Earth, making it an ideal site for high-energy particle astronomy. IceCube has already contributed to several significant discoveries in both astroparticle physics and particle physics (such as high-energy neutrinos, cosmic rays, and fundamental interactions). The current detector features an in-ice array of optical modules placed 1.5–2.5 kilometers deep in the ice, covering a total instrumented volume of about 1 km^3. Above this deep array lies the IceTop surface detector, an array of ice-Cherenkov tanks distributed across a $\sim$ 1 km^2 footprint, providing measurements of extensive air showers from cosmic rays. Both the in-ice optical array and the surface detector will be significantly expanded in IceCube-Gen2, improving the observatory's sensitivity, effective volume, and energy reach [83].

The high-energy optical array (Fig. 3.2) consists of 120 new strings (orange), spaced by 240 m, each instrumented with 80 advanced optical modules over a vertical extent of about 1.25 km. IceCube-Gen2 in the reference design comprises 361 stations, including both shallow and hybrid types of radio antennas. Adjacent to this, the optical high-energy array is depicted in Figure 3.2, consisting of 120 new strings (indicated by orange points) that are distributed 240 meters apart. Each string is equipped with 80 advanced optical modules spanning a vertical length of 1.25 kilometers. The optical detector's total instrumented

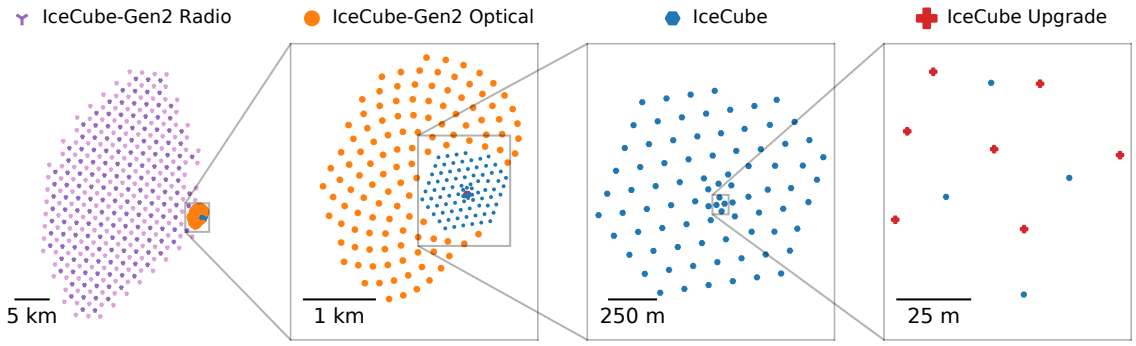


Figure 3.1: Top view of the proposed layout for the IceCube-Gen2 Neutrino Observatory at the South Pole station in Antarctica is presented. From left to right: The in-ice radio array in the reference design comprises 361 stations, an optical high-energy array consisting of 120 new strings, and the surface array. On the far right, the seven strings of the planned IceCube Upgrade are also shown, illustrating their position relative to the new arrays.

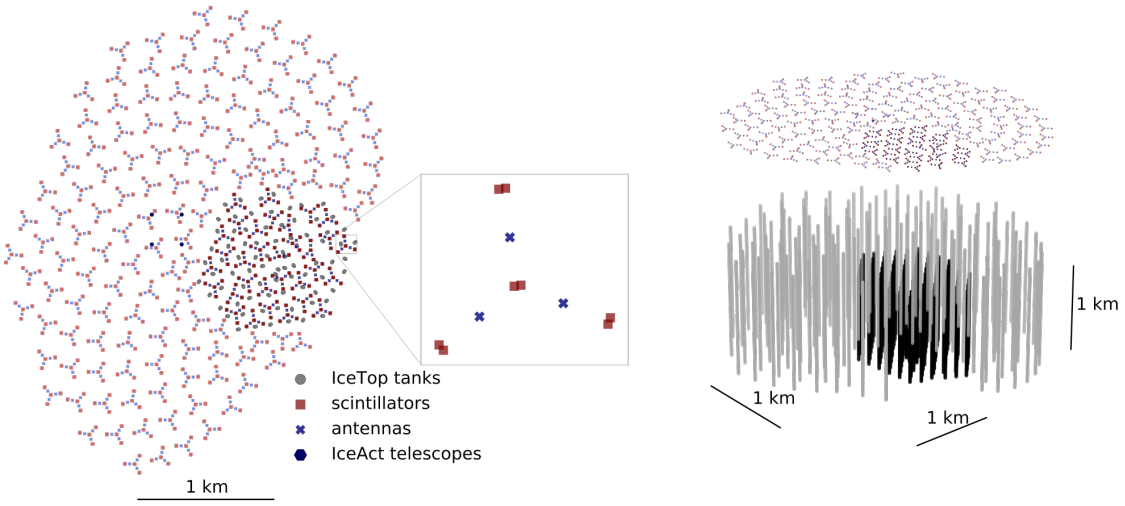


Figure 3.2: Illustration of IceCube-Gen2 surface array stations with elevated scintillator panels and radio antennas. The figure is taken from [289].

volume in this configuration is approximately 7.9 times greater than that of the existing IceCube detector array (illustrated by blue points). To the extreme right, the positioning of the seven IceCube Upgrade strings in relation to the existing IceCube strings is illustrated. The expansion of the observatory to include a surface array builds on its past successes and its crucial role in enhancing our understanding of cosmic rays as well as air-shower physics [85]. To detect extremely high-energy (EHE) and PeV-scale neutrino events, IceCube-Gen2 will include an in-ice radio array (left panel), which is optimized for $\gtrsim 10^{17}$ eV neutrino detection (not for CR detection). The design uniquely allows the detection of low-energy

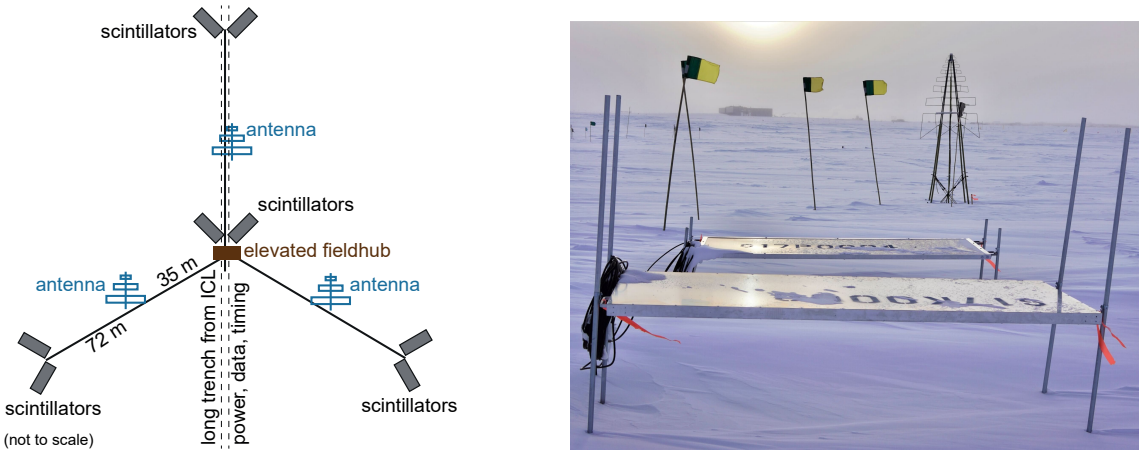


Figure 3.3: (Left) Schematic layout of a surface station developed for the IceTop enhancement, serving as the reference design for the IceCube-Gen2 surface array. Each station includes four pairs of scintillation detectors and three radio antennas linked to a central local data-acquisition unit. (Right) Photograph of the prototype station deployed at IceTop, where both scintillators and antennas are mounted on adjustable stands to mitigate snow accumulation. Plot and figure are taken from [27].

particles like GeV muons and electromagnetic particles at the surface, while muons above 300 GeV are detected in the ice. This makes the observatory a leader in studying particle physics through air showers [85].

The planned IceCube-Gen2 surface array will include stations equipped with elevated scintillator panels and radio antennas, as illustrated in Figure 3.1 and Figure 3.3. These stations follow the design proposed for the surface array extension, consisting of eight scintillator panels arranged in paired units that form a triangular geometric layout. Each of the three arms of this layout will have a radio antenna positioned approximately halfway along its length. These stations will be strategically placed above the in-ice optical strings, maintaining an average inter-station spacing of about 250 meters. Additionally, the center of the array will host four IceAct stations, each separated by about 200 meters and equipped with seven Cherenkov telescopes [27, 26]. For the study of anisotropies in the arrival directions of cosmic rays, the detections utilizing the scintillator panels are of particular interest. These scintillation detections will provide valuable data to help identify and analyze the subtle variations in the directional distribution of incoming particles, which have not been done by IceCube previously. The IceCube-Gen2 surface array will have a larger area with scintillator panels and will obtain more statistics than the current IceTop.

3.1.2 Simulation of Air Showers

To evaluate the full scientific potential of the planned IceCube-Gen2 surface array, an extensive and carefully constructed library of air-shower simulations was used for this study [289]. These existing simulations employed CORSIKA [290] to model the detailed distribution of secondary particles at the surface and CoREAS [291] to compute the associated radio emission produced by the electromagnetic component of the showers. For the detailed physics interactions, CORSIKA was compiled with FLUKA [292] to handle low-energy interactions and Sibyll 2.3d [293] for high-energy processes. Given the extensive computational demands of radio simulations and in order to reduce the need for thinning the simulations, two different approaches were implemented. For cosmic ray energies between 10^{13} and 10^{17} eV and zenith angles up to 51° (72° for energies up to $10^{16.5}$ eV in dedicated trigger-performance studies), showers were simulated without thinning using only the CORSIKA particle routines together with the modeled response of the scintillator array [289]. The resulting energy–zenith distributions for simulated proton and iron primaries are displayed in Figure 3.4, while the corresponding sky maps of their arrival-direction distributions are shown in Figure 3.6, illustrating the coverage expected for the surface array. For radio emission simulations, the thinning algorithm was applied. Proton and iron primaries were used to model extreme cases [289].

Air showers were injected multiple times within a 1.5 km radius from the array center, and the detector response was simulated. Secondary particles were injected into scintillator panels, with energy losses calculated using Geant4 [295], and signals expressed in MIPs [289]. The main observables were derived from signal amplitudes and timestamps, enabling precise shower geometry reconstruction using scintillator data. The study confirmed that for the larger IceCube-Gen2 surface footprint, the shower geometry could be accurately estimated using scintillator reconstruction. IceTop tanks, shielded by several meters of snow, functioned as muon-sensitive detectors, particularly effective at distances greater than 600 meters from the shower axis [296].

The planned surface area covered by the IceCube-Gen2 surface array is 6.6 km^2 (depending on the definition of the containment for reconstruction). The corresponding

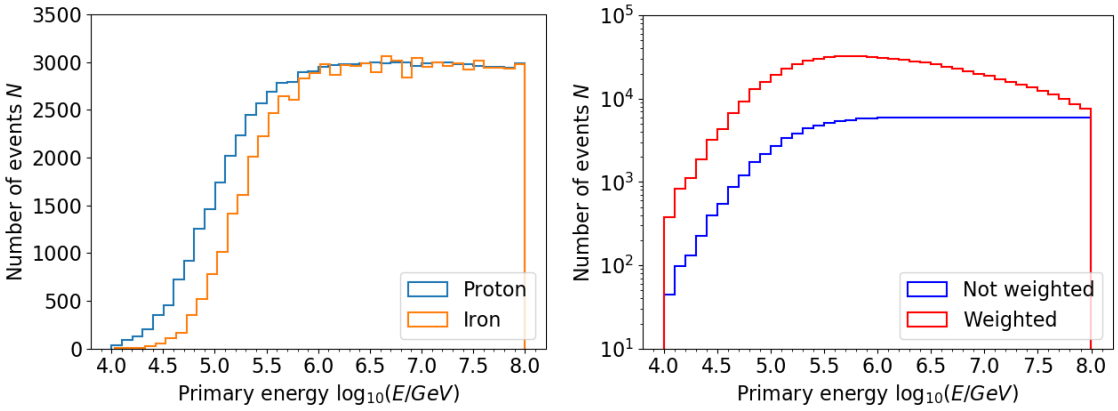


Figure 3.4: Histograms of simulated proton and iron energies. The left plot shows their energy distributions, while the right plot shows the combined distribution. The red line is weighted to the H4a flux model [294], and the blue line is unweighted.

CORSIKA simulations of the scintillator array response were performed for proton- and iron-induced air showers with $4 \leq \log_{10}(E/\mathrm{GeV}) \leq 8$ and zenith angles (θ) up to 51°, and $4 \leq \log_{10}(E/\mathrm{GeV}) \leq 7.5$ with θ up to 63° [289]. To simulate a large sample of contained events, the shower cores were randomly distributed within a 1.5km radius from the center of the surface array. The detector response was then simulated for the scintillators, where the secondary particles on the ground were injected into the scintillator panels.

To show the realistic capabilities of the IceCube-Gen2 surface array, we perform a selection process that involves choosing true air shower core locations (x, y) within a distance of 100 meters from the polygonal boundary of the array, ensuring that the full geometric footprint of the shower is properly accounted for. The scintillator array is fully efficient (100%) in triggering on air showers above 0.5 PeV for vertical events ($\theta = 0$). For a simple air shower trigger, at least three scintillator hits inside the array are required in order to achieve a minimal reconstruction of the shower direction and geometry. For a more realistic requirement, at least five scintillator stations should be triggered. Considering the scintillator-triggered multiplicities ≥ 5, we show the reconstruction efficiencies for proton and iron primaries and make the histograms for the two cosmic-ray primary particles. To account for the incomplete coverage in the region $0.6 \leq \sin^2(\theta) \leq 0.8$ and $7.5 \leq \log_{10}(E/\mathrm{GeV}) \leq 8$ we fill the missing bins by assuming a conservative 100% efficiency in that phase space (see Figure 3.5). Finally, we fit the unbroken histograms using a modified two-dimensional error function parameterized in terms of the primary energy E and the zenith angle $z = \sin^2\theta$, which is expressed as

$$\epsilon(E,z) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{E - (\mu_0 + \mu_1 z + \mu_2 z^2 + \mu_3 z^3)}{\sigma_0 + \sigma_1 z}\right)\right]. \tag{3.1}$$

where μ_0, μ_1, μ_2, μ_3, σ_0 and σ_1 are constant. These parameters for both proton and iron in Eq. 3.1 are listed in Table 3.1. We estimate the reconstruction efficiencies for helium, nitrogen, and aluminum using the natural logarithm of their mass numbers, denoted as $\ln A$. This estimation is based on the logarithmic mass dependence of the cosmic-ray primaries. The reconstruction efficiency for helium, nitrogen, and aluminum can be written as

$$\epsilon(A_i,E,z) = \frac{\ln A_i - \ln A_\mathrm{P}}{\ln A_\mathrm{Fe} - \ln A_\mathrm{P}}\left[\epsilon_\mathrm{Fe}(E,z) - \epsilon_\mathrm{P}(E,z)\right] + \epsilon_\mathrm{P}(E,z), \tag{3.2}$$

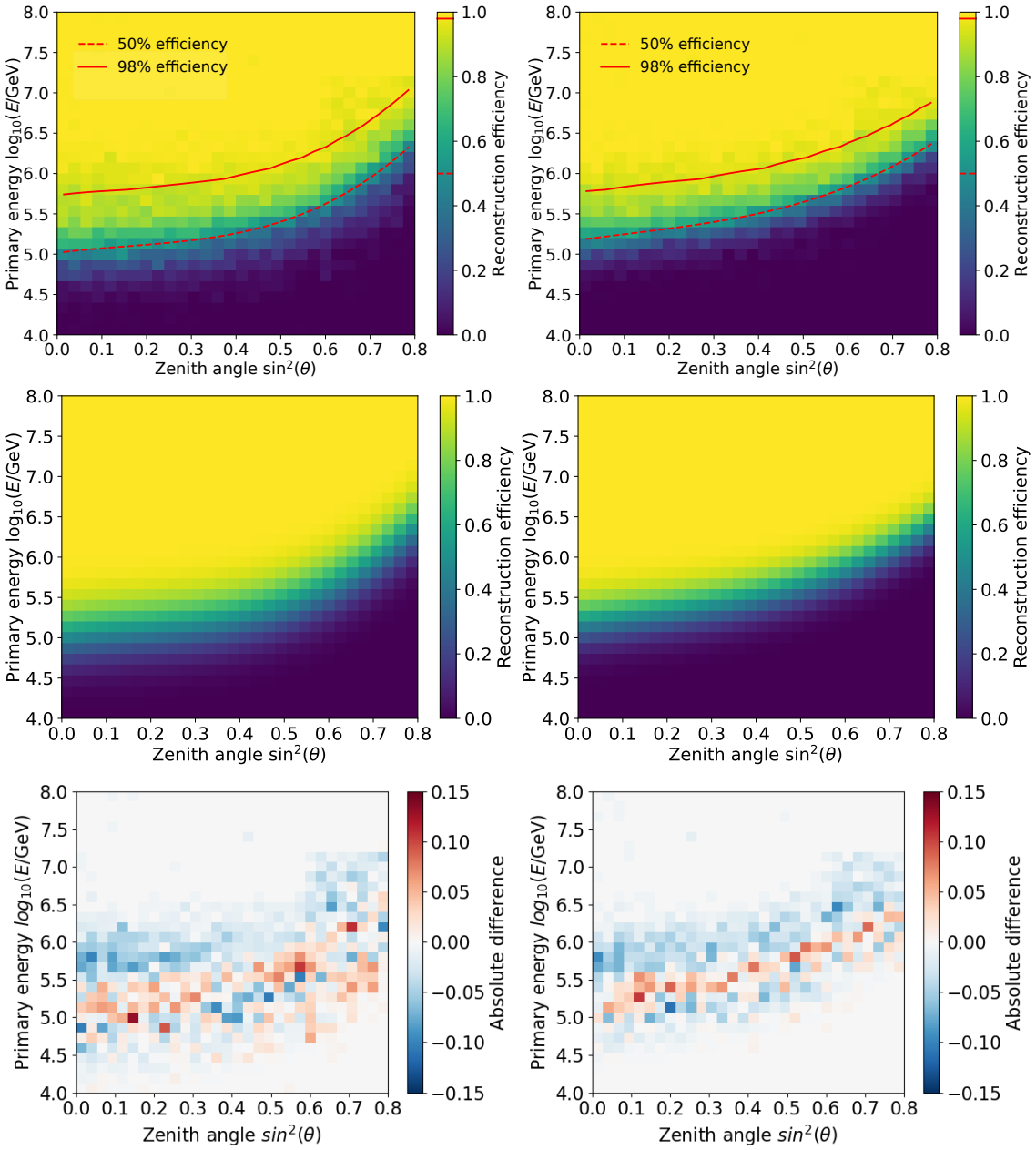


Figure 3.5: Two-dimensional histograms of the reconstruction efficiency for proton-induced (top left) and fitted (center left), and iron-induced (top right) and fitted (center right) air showers for scintillators in the IceCube-Gen2 surface array. The energy range is from 10^6 GeV to 10^8 GeV, with zenith angles up to 63° in $\sin^2\theta$ scale, and a scintillator multiplicity of ≥ 5 to trigger the event. The bottom two histograms show the absolute differences between the original and fitted histograms, respectively.

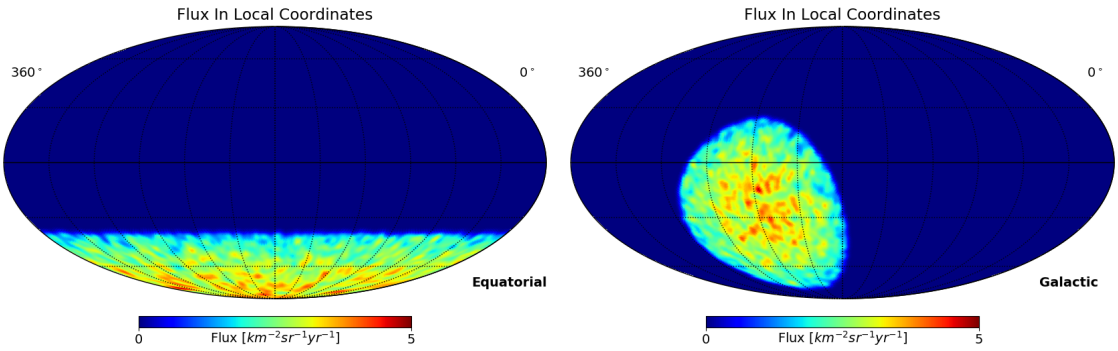


Figure 3.6: Sky maps showing the total simulated proton and iron flux for IceCube-Gen2, without weighting by the H4a flux model.

where $\ln A_{\mathrm{P}} = 0$ for protons. The index i ranges from 1 to 3, representing the estimated helium, nitrogen, and aluminium, respectively. Polynomials up to the third degree (z^3) can capture sufficient details for the efficiencies. We plot the contour lines at 50% and 98% reconstruction efficiency using Eq. (3.1), as shown in the left panel of Figure 3.7. The total efficiency of all particles utilizes Eq.(3.1), (3.2) and the H4a flux model [297].

The scintillators will be the most sensitive component of the surface hardware of IceCube-Gen2, and thereby determine the energy threshold necessary for cosmic-ray detection. In the performance studies of the reconstructed arrival direction presented in paper [289], it is shown that for air showers with threshold energies around 1 PeV, the incoming direction can already be reconstructed with an accuracy of a few degrees, even for events arriving at zenith angles as large as 45°. As the primary energy increases, the reconstruction improves, and above 10 PeV, the angular resolution reaches the sub-degree level. The accurate estimation of the shower geometry and the reliable determination of the arrival direction make it highly valuable for studying cosmic-ray anisotropies.

3.1.3 Simulation of Arrival Directions

The section aims to simulate the arrival directions of cosmic rays and generate synthetic sky maps to evaluate the sensitivity of the IceCube-Gen2 surface array to cosmic-ray anisotropy studies. We utilize the H4a flux model to estimate the number of cosmic-ray events detected in different energy bins. By integrating reconstruction efficiency and the number of cosmic rays over a decade of exposure, we can simulate the expected arrival directions. This involves complex calculations incorporating various factors such as detector efficiency, geographical location, and anisotropic distribution patterns.

By understanding how the array responds to different directional inputs, we can better predict its performance in identifying cosmic-ray dipoles across the sky. Therefore, we can test the IceCube-Gen2 array's ability to detect cosmic-ray anisotropies. The ability to

Table 3.1: Parameters in the fit function Eq.(3.1) of cosmic ray reconstruction efficiency of the IceCube-Gen2 surface array. Only proton and iron primaries are covered in this table.

Particle	μ_0	μ_1	μ_2	μ_3	σ_0	σ_1
Proton	5.017 ± 0.008	0.711 ± 0.083	-1.791 ± 0.240	3.824 ± 0.198	0.483 ± 0.008	0.001 ± 0.000
Iron	5.177 ± 0.006	0.753 ± 0.063	-0.665 ± 0.182	2.073 ± 0.149	0.404 ± 0.006	-0.067 ± 0.012

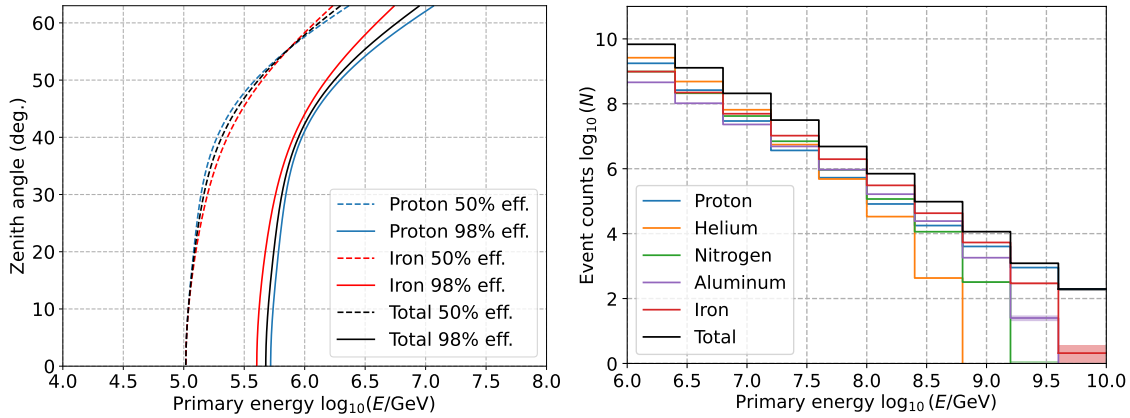


Figure 3.7: Left: contour lines with 50% and 98% reconstruction efficiency of the IceCube-Gen2 surface array are shown for proton and iron, and total efficiency of all particles utilizing the H4a flux model [274]. Right: histogram (with error bands) of the event counts of the IceCube-Gen2 surface array with a 10-year exposure, considering the H4a flux model and the reconstruction efficiency with 3 additional energy bins.

simulate different scenarios and predict outcomes enhances our capability to distinguish between random fluctuations and genuine anisotropic patterns. These simulations help us understand the directional dependencies and the potential sensitivity of the IceCube-Gen2 array to cosmic-ray anisotropies, highlighting the IceCube-Gen2 array's sensitivity to dipole anisotropies. The following parts will detail the methodologies used in the simulation, including the computational models and algorithms employed. We will also discuss the interpretation of the results, focusing on how well the IceCube-Gen2 array can detect and analyze cosmic-ray anisotropies.

Using the H4a flux model [297], the expected number of arrival directions in each energy bin can be obtained by integrating the two-dimensional function defined in Eq. (3.1) and (3.2), which describe the reconstruction efficiency and the number of cosmic rays with 10 years of exposure of the IceCube-Gen2 surface array. This calculation is a function of E and $z = \sin^2\theta$ for each primary mass group. Therefore, we have the event counts for each of the 5 primary components in the energy bins (see the right panel of Figure 3.7):

$$N_{\mathrm{eff},ij} = \int_{\theta_{\min}}^{\theta_{\max}} \int_{E_{j-1}}^{E_j} \epsilon_i\left(E,\theta\right) N_{\mathrm{H4a},i}\left(E\right) dE d\theta, \tag{3.3}$$

where i runs from 1 to 5, corresponding to the five different primary particles, while j runs from 1 to 8, denoting the upper and lower edges of the seven reconstructed energy bins used in the simulation. The quantity $N_{\mathrm{H4a},i}$ refers to the number of expected events for each primary component, as determined by the mass-dependent H4a flux model. Using this procedure, the full simulation produces a total of 8.355 billion simulated cosmic ray events distributed across all seven energy bins. Taking into account the ratio of different primaries weighted by the mass-dependent H4a flux expressions [297] and Eq. (3.1) and (3.2), the total efficiency averaged over all primaries is written as

$$\epsilon_{\mathrm{tot}}\left(E,\theta\right) = \frac{\sum_{i=1}^{5} \epsilon_i\left(E,\theta\right) \phi_{\mathrm{H4a},i}\left(E\right)}{\sum_{i=1}^{5} \phi_{\mathrm{H4a},i}\left(E\right)}, \tag{3.4}$$

where $\epsilon_i(E,\theta) \simeq 1$ for $\log_{10}(E/\mathrm{GeV}) > 7$, as shown in Figure 3.5.

Now, let us examine how one can simulate cosmic-ray arrival directions on the celestial sphere using equatorial coordinates at a given observatory. To obtain a realistic sky map, several key factors need to be taken into account in this model to ensure accurate simulation results. First, the total efficiency of the detector, as calculated in Eq. (3.4), is considered. This efficiency encompasses the detector's ability to capture and accurately record cosmic ray events from various directions. It also accounts for how features such as detector orientation, trigger conditions, and threshold effects influence the probability of detecting particles from different zenith and azimuth angles. Second, the relative exposure of the observatory needs to be taken into account, since the observatory's latitude, field of view, and duty cycle determine how long each sky region is observable and therefore contribute unevenly to the overall directional sensitivity. Lastly, the anisotropic distribution of the arrival directions is modelled and given by:

$$\omega_{\rm obs}\left(E,\theta,\delta,\theta_{\rm max},\mathcal{A},\delta_{\rm d},\delta_{\rm obs}\right)=\omega\left(\theta,\delta,\theta_{\rm max},\delta_{\rm obs}\right)\times\epsilon_{\rm tot}\left(E,\theta\right)\times\mathcal{D}\left(\mathcal{A},\delta,\delta_{\rm d},\delta_{\rm obs}\right), \quad (3.5)$$

where δ refers to the declination of the simulated cosmic-ray arrival directions, while $\delta_{\rm d}$ specifies the declination of the injected dipole. The quantity $\delta_{\rm obs}$ represents the declination of an observatory, and ω is the relative acceptance (exposure) of IceCube-Gen2 surface array at declination $\delta_{\rm obs}=-89.99^{\circ}$ without detector efficiencies.

The probability density function (PDF) describing this declination-dependent relative instantaneous acceptance, which characterizes the real-time directional sensitivity of the detector, is expressed by the following relation [274]:

$$\omega(\delta)=\frac{1}{\pi}\left(\cos\delta_{\rm obs}\cos\delta\sin\alpha_{\rm m}+\alpha_{\rm m}\sin\delta_{\rm obs}\sin\delta\right), \quad (3.6)$$

where π is used for normalizing the maximum possible value to 1, $\alpha_{\rm m}$ is given by a clip function which ensures the geometric factor (ξ) that links the observatory's latitude stays within the valid range for the arccos function, namely, $\alpha_{\rm m}=\arccos\left(\mathrm{clip}(\xi,-1,1)\right)$, which determines whether a celestial object at a given declination (δ) is within the observable zenith angle from a specific latitude (a_0). It is basically the furthest left or right the observatory can look to see a celestial object before it moves out of view. The quantity ξ is used to calculate the maximum observable azimuth angle at which an astronomical object can be observed from a specific location on Earth, and is a geometric relation between the observatory's position and the celestial object's declination. Taking into account the Earth's curvature and the observatory's field of view, ξ is expressed as:

$$\xi=\frac{\cos\theta_{\rm max}-\sin\delta_{\rm obs}\sin\delta}{\cos\delta_{\rm obs}\cos\delta}, \quad (3.7)$$

which relates how celestial coordinates (α,δ) are converted into local coordinates, such as altitude and azimuth, at a given time and location. For the IceCube-Gen2 surface array, the relative exposure is well approximated by $\omega(\delta)\simeq\cos(\delta-\delta_{\rm obs})$ due to the geometric symmetry. Besides, the function $\mathcal{D}\left(\mathcal{A},\delta,\delta_{\rm d},\delta_{\rm Gen2}\right)$ in Eq. (3.5) provides the dipole distribution, taking into account the relative orientation of the IceCube-Gen2 surface array to the dipole. The algorithm for configuring the dipole is as follows.

In this study, the primary focus is on dipole anisotropies, which describe how the observed cosmic-ray flux varies with direction across the sky. The resulting full distribution function, therefore, determines the probability of detecting cosmic rays from different directions. It is given by a comprehensive formula that incorporates detector efficiency, relative exposure, and anisotropy parameters. By integrating these key factors into the simulation, we can create a realistic model of how cosmic rays arrive at the observatory,

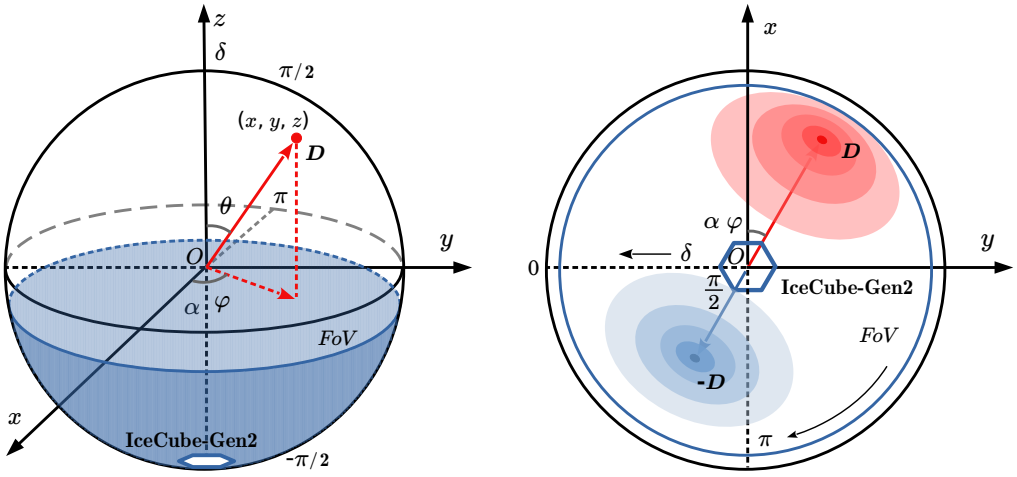


Figure 3.8: Combined 3D spherical coordinates and equatorial coordinates. The left plot shows the 3D FoV of IceCube-Gen2, where a point $\mathcal{D}$ represents the excess of a dipole, and the hexagon at the bottom represents the location of IceCube-Gen2. The right plot is the 2D projection of the celestial sphere on the xy plane, with IceCube-Gen2 at the centre.

which allows us to generate synthetic sky maps that accurately reflect the expected distribution of cosmic rays, thereby enabling detailed studies of cosmic-ray anisotropies and the effectiveness of the IceCube-Gen2 array in detecting these phenomena.

To simulate CR dipole anisotropy, we developed a mathematical model to simulate the behavior of dipole distributions on a spherical surface, which is particularly relevant for studying phenomena such as Cosmic Microwave Background (CMB) anisotropies. This model aims to create a realistic representation of how dipole patterns manifest on a celestial sphere due to varying intensities and directional influences. The model begins with the definition of a dipole vector in three-dimensional Cartesian coordinates (xyz) and is characterized by a direction vector, namely, $\overrightarrow{\mathcal{D}} = (x, y, z)$, which forms a full 3D unit vector of the arrival directions and defines the orientation of the dipole pattern on the sphere, as shown on the left plot of Figure 3.8. Next, the sphere is divided into discrete points or pixels, each representing a segment of the sphere's surface. The angle between the dipole direction vector $\overrightarrow{\mathcal{D}}_{\rm d}$ and the vector pointing to the center of each pixel, $\vec{v}_i$, is calculated, denoted by Θ_i, which determines the intensity of the dipole at each point, and the sum of all $\cos\Theta_i$ are calculated as follows: $\cos\Theta_i = \overrightarrow{\mathcal{D}} \cdot \vec{v}_i$ and $\cos\Theta_{\rm d} = \sum_{i=1}^{N_{\rm pix}} \overrightarrow{\mathcal{D}} \cdot \vec{v}_i$, where $N_{\rm pix}$ is the total number of pixels, $\cos\Theta_i$ yields a scalar field over the sphere ranging from -1 to 1, and the value of $\cos\Theta_{\rm d}$ represents the angular deviation relative to the dipole's direction. The coordinates of specific pixels can be obtained utilizing the transformation function in the HEALPix library [281], which transforms the HEALPix grid pixels to three-dimensional Cartesian coordinates, based on the principle of equal-area division on the sphere. This transformation begins by identifying each pixel with a unique identifier, which corresponds to a distinct area on the sphere. It then maps these identifiers to the geometric centre of their respective areas, calculating the Cartesian coordinates (x, y, z) that represent these centres. Theoretically, $\Theta_{\rm d}$ can be expressed as

$$\cos\Theta_{\rm d} = \cos\theta\cos\theta_{\rm d} + \sin\theta\sin\theta_{\rm d}\cos(\phi - \phi_{\rm d}), \tag{3.8}$$

where θ and ϕ are the spherical coordinates of the vectors on the celestial sphere, θ_d and

$\phi_{\rm d}$ are of the dipole. To transform this dipole distribution into a modulation pattern, it is scaled by an amplitude factor $\mathcal{A}$, which controls the strength of the dipole influence. This scaling is essential for adjusting the visibility and contrast of the dipole pattern across the sphere. The mathematical representation of this scaling is $\mathcal{D}_{\rm map} = 1 + \mathcal{A} \cdot \cos\Theta_{\rm d}$, which assigns to each pixel a value that reflects the dipole intensity at that point on the sphere. Finally, to ensure the values across the sphere sum to one, making it a valid probability distribution, the entire map is normalized by its total sum,

$$\mathcal{D}_{\rm pdf} = \mathcal{D}_{\rm map} \Big/ \sum\nolimits_i \mathcal{D}_{\rm map}, \tag{3.9}$$

which is critical as it converts the dipole map into a probability density function (PDF), and is used for statistical analysis and simulations.

To proceed with the simulation of dipole patterns on the celestial sphere, it is essential to translate the dipole orientation, which is originally specified in celestial coordinates such as right ascension (RA) and declination (Dec), into a corresponding set of three-dimensional Cartesian coordinates. This transformation relies on the standard definition of spherical coordinates (r, θ, ϕ), where the Cartesian components are expressed as $x = \sin\theta\cos\phi$, $y = \sin\theta\sin\phi$ and $z = \cos\phi$, where θ is colatitude, i.e., the angle measured southward from the North Pole, ranging from 0 to π, and ϕ is longitude, i.e., the angle measured eastward, ranging from 0 to 2π. RA (α) is equivalent to the longitude (ϕ) component in degrees, while Dec. (δ) ranges from -90 degrees (South Pole) to +90 degrees (North Pole). Colatitude, however, is measured from the North Pole downward. Therefore, the conversion from equatorial coordinates (right ascension and declination) to spherical coordinates (longitude and colatitude) is $\phi = \alpha$, $\theta = \pi/2 - \delta$, which can also be expressed in hours, if needed. By specifying the equatorial coordinates of a dipole, we can straightforwardly determine its Cartesian coordinates on the celestial sphere. This simplifies the input procedure of dipole injections in the simulation. Based on the setup of all the above and the combination of them as described in Eq. (3.5), the cosmic-ray arrival directions are simulated by randomly selecting pixels from these weighted distributions.

For each pixel of the sky map, the orientations of the individual direction vectors within that pixel are assigned randomly, with the total number of vectors corresponding to the expected number of cosmic ray events over the specified years of observatory exposure. This process is repeated across multiple maps to systematically compile and analyze the spatial distribution of cosmic rays detected by the observatory setup. Each step in the process, from generating weighted distributions to mapping and counting events, is strictly defined to ensure an accurate representation of the observatory's detection capabilities under the studied conditions. In this way, we obtain the effective probability for injecting a given dipole component into the simulated arrival-direction distribution as it would be observed by the IceCube-Gen2 surface array.

3.1.4 Dipole Injections

To evaluate the observable dipole anisotropy of IceCube-Gen2 and its sensitivity, we take into account the reconstruction efficiency discussed earlier, even though it approaches nearly 100% above 10^7 GeV. In the simulation framework, we inject cosmic-ray arrival directions into the geometry of the IceCube-Gen2 surface array, assuming event counts that are roughly consistent with what is expected from a ten-year observational exposure. The simulated dataset is divided into six logarithmic energy bins, spanning the range from 10^6 GeV to $10^{8.4}$ GeV, with a bin size of 0.4 in $\log_{10}(E/\text{GeV})$. Here, as an example, injected

dipoles have an amplitude of 10^{-3}, consistent with the approximate magnitudes reported for reconstructed dipoles in multiple experiments.

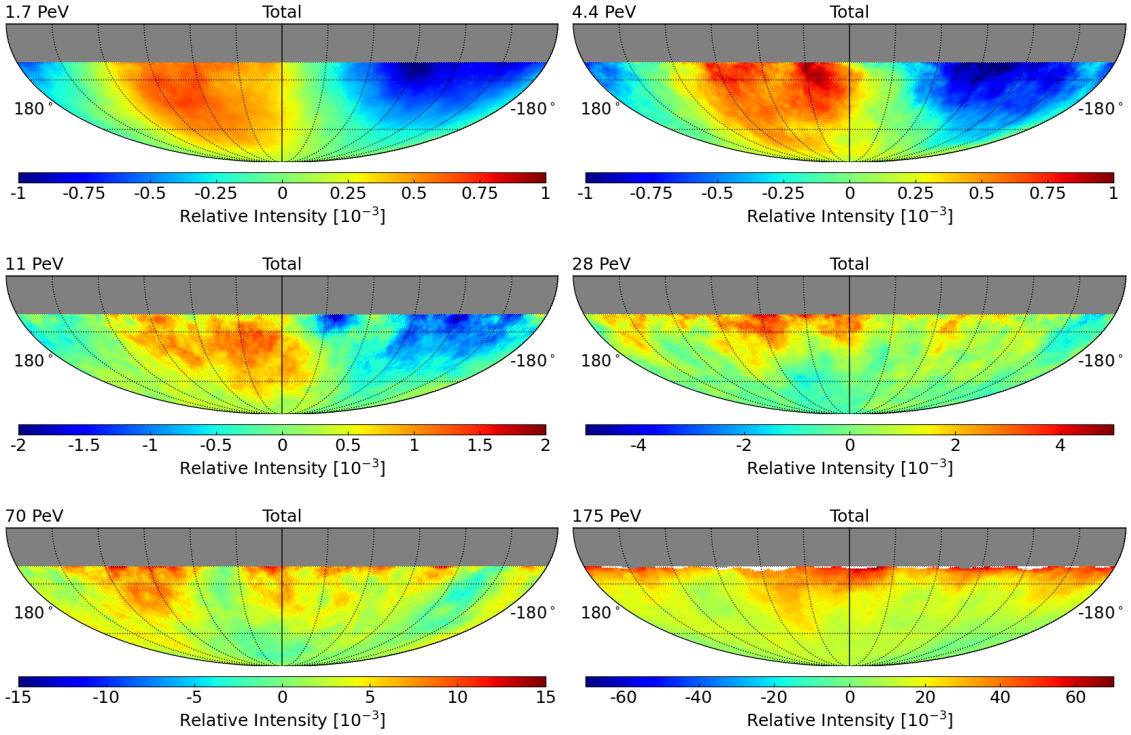


Figure 3.9: Sky maps of relative intensity illustrating a simulated pure dipole anisotropy injected at R.A. $\alpha = 60°$ and declination $\delta = -30°$ with an amplitude of 10^{-3}. The distributions are shown for all six representative energy bins of IceCube-Gen2, demonstrating the detector's sensitivity to large-scale anisotropy patterns across different energies. Each map has been smoothed using a 20° top-hat filter to emphasize large-scale structures.

The sky maps for a representative example of the simulated arrival directions are displayed in Figure 3.9. In these maps, the injected dipole patterns are visibly pronounced in the first three energy bins, while they disappear in the higher energy bins. This disappearance is attributed to the lack of statistical significance at higher energies. Figure 3.12 presents the corresponding angular power spectrum for the first energy bin, illustrating that the dipole term with $\ell = 1$ is clearly more prominent than the medium-scale anisotropies ($\ell = 2, 3$) and the small-scale anisotropies ($\ell \geq 4$). A more detailed analysis, including an initial test of how the reconstructed significance varies for different injected dipole declinations and amplitudes, is shown in Figure 3.10. The observed symmetry along the declination is a result of the partial sky coverage of the observations, as well as the inherent projection effects of the dipole. For these reasons, we intend to perform more precise analyses to determine the sensitivity of Gen2. This will include injecting additional sets of dipoles with larger amplitudes to investigate how the statistical significance propagates through the reconstruction. Through these tests, our goal is to derive a more accurate and detailed sensitivity curve that reflects the realistic performance expected for Gen2, particularly in the energy range where current experiments are still limited by statistics.

To get a precise and stable value of the reconstructed dipole amplitude, we choose dipole amplitudes depending on the declination, focusing on amplitudes that can be detected at

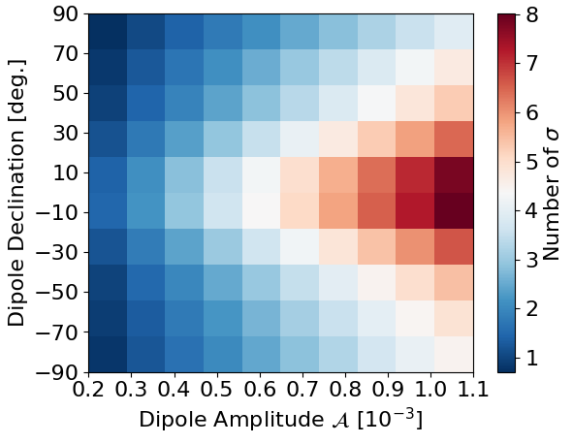


Figure 3.10: Two-dimensional histogram illustrating the reconstruction significance of simulated dipole anisotropies for various injected amplitudes and declinations. The simulation covers dipoles at declinations from $-80°$ to $80°$. The color scale indicates the statistical significance derived from the injected dipole maps shown.

$> 10\sigma$ within 10 years of operation. Otherwise, it would be required to simulate the same dipole repeatedly to capture the uncertainties adequately. We now apply the reconstruction efficiency to simulate the arrival directions for the IceCube-Gen2 surface array with a 10-year exposure. The simulation is divided into 7 energy bins ranging from 10^6 GeV to $10^{8.8}$ GeV with a bin size of 0.4 in $\log_{10}(E/\mathrm{GeV})$. In each energy bin, we inject 15 different dipoles at declinations ranging from $-80°$ to $80°$. In total, we have 1785 injected dipoles and their corresponding sky maps. Next, we randomly inject the CR arrival directions using the relative acceptance of the detector for all energy bins and all zenith angles below the threshold and scan over the dipole declination from $-80°$ to $80°$ with a bin size of $10°$. It is irrelevant whether the dipole orientation is within or outside the field of view (FoV) of Gen2. Furthermore, any observatory situated on Earth will invariably detect a partial sky coverage CR dipole, with the dipole orientation, or 'hotspot', potentially located either within or outside the FoV. In the simulation, we have a total of 1785 injected dipole maps, each with amplitudes detectable at levels exceeding 10σ. Our goal is to study how dipole significance propagates at IceCube-Gen2, so we consistently use larger amplitudes as inputs. Therefore, the amplitudes are chosen from 7×10^{-3} up to 9.56×10^{-1} for the injected dipole, with different energy bins covering different ranges of the amplitude.

3.1.5 Reconstruction of Dipole Anisotropy

To properly reconstruct the cosmic-ray dipole anisotropies at a partial-sky coverage observatory, it is essential to compare the actual sky map of injected CR arrival directions (data map with a simulated dipole in Eq. (3.5)) with a sky map that reflects the detector's response to an isotropic CR flux (reference map without dipole). Note that, in our case, the reference map is constructed to be purely isotropic once the detector exposure is applied. In particular, no dipole component is included, i.e., $\mathcal{D}\left(\mathcal{A}, \delta_{\mathrm{d}}\delta_{\mathrm{Gen2}}\right) = 1$. This comparison allows us to evaluate the deviation from isotropy and quantify the presence of the dipole anisotropy in the simulated data by Eq. (3.5). Therefore, the residual between the data maps and the reference maps, which are obtained by normalizing each declination band independently, is sensitive only to anisotropy in right ascension (α). In particular, the

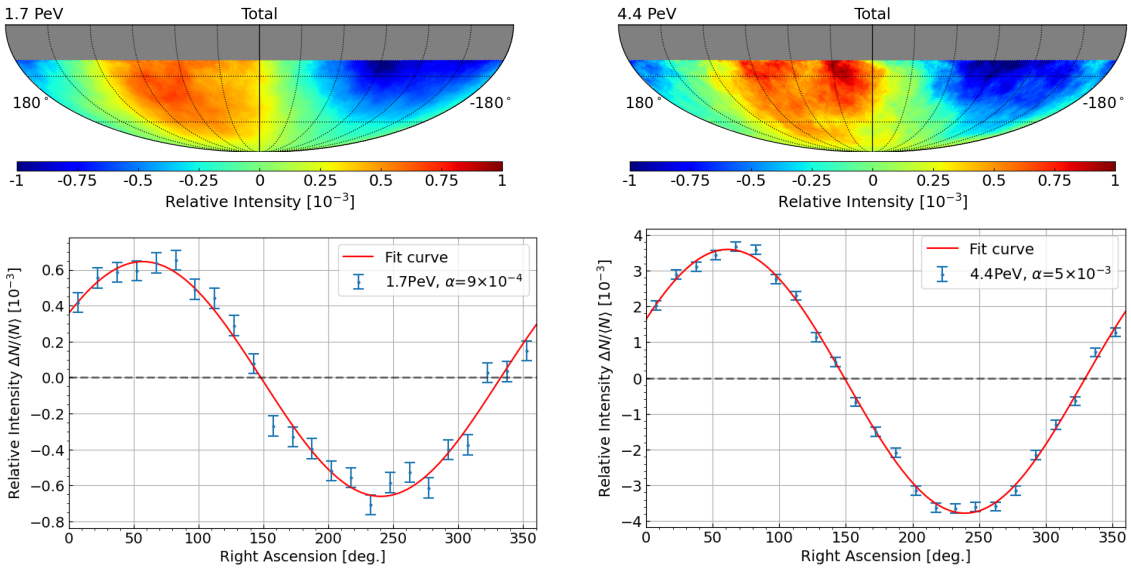


Figure 3.11: Upper panel: sky maps display the relative intensity. Lower panel: corresponding one-dimensional projections of these maps are shown with the first-harmonic fits, which extract the dipole parameters in each scenario.

fitting of the dipole's projection in the equatorial plane with the first harmonic allows for a direct measurement of a dipolar cosmic ray distribution. We ignore acceptance effects and assume an ideal homogeneous exposure in the injection procedure in Section 3.1.3, so that any remaining modulation in right ascension arises solely from large-scale anisotropies and statistical fluctuations. Based on the Monte-Carlo simulation as described, we generate sky maps of relative intensity defined by the equation

$$\delta I = \frac{N_{\mathrm{pix},i} - \langle N \rangle_{\mathrm{pix},i}}{\langle N \rangle_{\mathrm{pix},i}}, \tag{3.10}$$

where $N_{\mathrm{pix},i}$ represents the observed data map while $\langle N \rangle_{\mathrm{pix},i}$ represents the expected background map. This calculation allows us to visualize deviations from the expected isotropic background, highlighting regions of relative intensity differences.

Next, we divide the right ascension into n_α bins. We then perform a one-dimensional (1D) projection of the sky map and fit it with a first-harmonic function, given by

$$\delta I_{\mathrm{reco}} = \mathcal{A}_{\mathrm{reco}} \cos(\alpha - \varphi) + B, \tag{3.11}$$

where $\mathcal{A}_{\mathrm{reco}}$ is the amplitude of the reconstructed dipole, ϕ is the phase, and B is a constant. This fitting procedure helps us identify and quantify the dipole anisotropy. The reconstruction ratio, defined as the ratio between the reconstructed dipole amplitude and the true injected amplitude ($\mathcal{A}$reco/$\mathcal{A}$), typically remains below about 80%. This reduced ratio arises mainly from the limited field of view (FoV) of the surface array, which restricts the portion of the sky that is observed and therefore limits how accurately the injected anisotropy can be recovered. It is important to note that $\mathcal{A}$ represents the input dipole, which is a spherically distributed dipole, given that cosmic-ray arrival directions are inherently based on spherical coordinates, rather than the projected dipole we observe. Figure 3.11 shows the sky maps of relative intensity and their corresponding one-dimensional projection fits. The bottom panels display the 1D projections of the two maps, fitted with

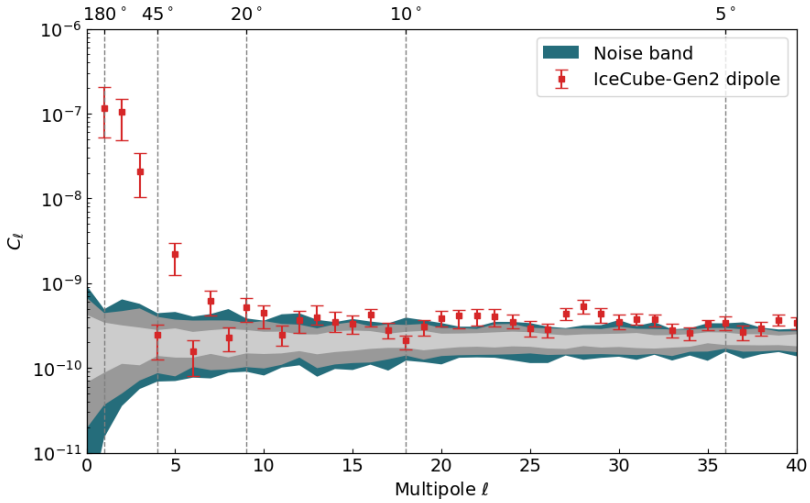


Figure 3.12: An example of the power spectrum of a dipole distribution. The wave-like points are generally from the limited field-of-view of IceCube-Gen2.

the first-harmonic function defined in Eq. (3.11). Figure 3.12 shows an example of the power spectrum of a dipole distribution, where the dipole ($\mathcal{C}_1$) point is significantly far away from the noise band. This is what we need, i.e., the dipoles with large significance.

3.1.6 Sensitivity to Dipole Anisotropy

Now we assess the sensitivity of a partial-sky coverage observatory to a dipole anisotropy. To assess the significance of a dipole deviation from isotropy, we consider a null hypothesis in the number of sigmas (i.e., a simple z-score)

$$n_\sigma = \frac{\mathcal{A}_{\rm reco} - \mathcal{A}_{\rm hypo}}{\sigma_{\mathcal{A}}}, \tag{3.12}$$

where $\mathcal{A}_{\rm hypo}$ is set to 0, corresponding to the value expected under the null hypothesis of no dipole signal, which differs from the idealized full-sky situation discussed in [274]). Under this hypothesis, the true amplitude equals the mean of the underlying Gaussian distribution, so its expected value remains zero. The variance, however, is determined entirely by the statistical uncertainty of the measurement and is given by $\sigma_{\alpha\rm bin} \propto 1/\sqrt{N_i}$, where N_i is the total number of arrival directions in each right ascension bin while $i = 1, 2, ...18$. The total number of arrival directions (N) in all energy bins. It is easy to get the approximate uncertainty of the map at each pixel, namely,

$$\sigma_{\rm map} \propto \sqrt{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}}, \tag{3.13}$$

where n is the total number of right-ascension bins used in the one-dimensional projected map, which in this analysis is fixed to 20 in this work. Consequently, we assume that the significance of the reconstructed dipole amplitude, as well as that of its one-dimensional harmonic extraction, scales proportionally to the inverse square root of the total number of events on the sky map, namely $\sigma_{\mathcal{A}} \propto 1/\sqrt{N}$, which is verified through Monte Carlo simulations of the first harmonic fit of the 1D projected map. In principle, the coefficient should be relevant to the observatory's orientation, particularly concerning declination,

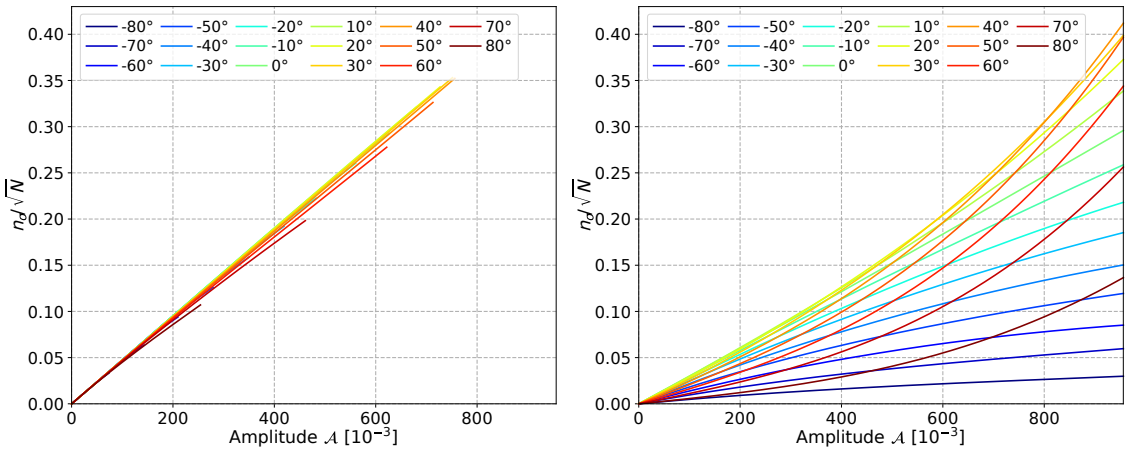


Figure 3.13: The fitting curves of significance σ and $\mathcal{A}$ of the reconstructed dipole (left) and the corresponding true dipole (right) are shown based on the sampled data points $(n_\sigma/\sqrt{N}, \mathcal{A})$ obtained from the simulation in Section 3.1.3. The colored lines are for the various declinations of the dipole.

which provides different FoVs and thereby affects the significance of dipole anisotropy. Furthermore, it also depends on the maximum zenith angle up to which cosmic-ray events are accepted, since this parameter controls the range of arrival directions contributing to the anisotropy reconstruction, because different dipole orientations project differently onto the observable sky. A larger exposure or wider zenith acceptance consequently improves the reconstruction significance by increasing the effective event statistics and sky coverage.

Therefore, based on Eq. (3.12), the sensitivity function (expressed in units of statistical significance, i.e., number of σ) can be written as follows:

$$n_\sigma = \mathcal{S}\left(\mathcal{A}, E, \theta_{\max}, \delta_{\rm d}, \delta_{\rm Gen2}\right) \mathcal{A}\sqrt{N}, \tag{3.14}$$

where $\mathcal{S}$ is defined as a sensitivity coefficient which is the inverse of the coefficient of $\sigma_{\mathcal{A}}$. The sensitivity to the reconstruction of the one-dimensional sidereal dipole, denoted as,

$$n_\sigma = \mathcal{S}_{\rm reco}\mathcal{A}_{\rm reco}\sqrt{N}. \tag{3.15}$$

which will be obtained from the observation by the IceCube-Gen2 surface array in the future. Similarly, the sensitivity of Gen2 against the true dipole can be expressed as,

$$n_\sigma = \mathcal{S}_{\rm true}\mathcal{A}_{\rm true}\sqrt{N}. \tag{3.16}$$

It is important to clarify that the sensitivity to the true dipole is not obtained directly from the reconstruction of the input dipole amplitude ($\mathcal{A}$). Instead, it represents the true dipole amplitude that would yield a reconstructed dipole detected at a given significance level, typically 3σ or 5σ, under the same observational conditions and detector response. To quantify this relationship, we scatter the data points $(n_\sigma/\sqrt{N}, \mathcal{A})$ obtained previously (see Section 3.1.4) for both the reconstructed and true dipole cases, and fit these points with a polynomial function of the form $\sum_{i=1}^{3} \lambda_i \mathcal{A}^i$, where λ_i are the polynomial coefficients (see Figure 3.13). The slope of these fitted curves then defines the sensitivity coefficients for both the reconstructed and true dipole scenarios, which can be expressed as

$$\mathcal{S}(\mathcal{A}, E) = \lambda_1(E) + 2\lambda_2(E)\mathcal{A} + 3\lambda_3(E)\mathcal{A}^2, \tag{3.17}$$

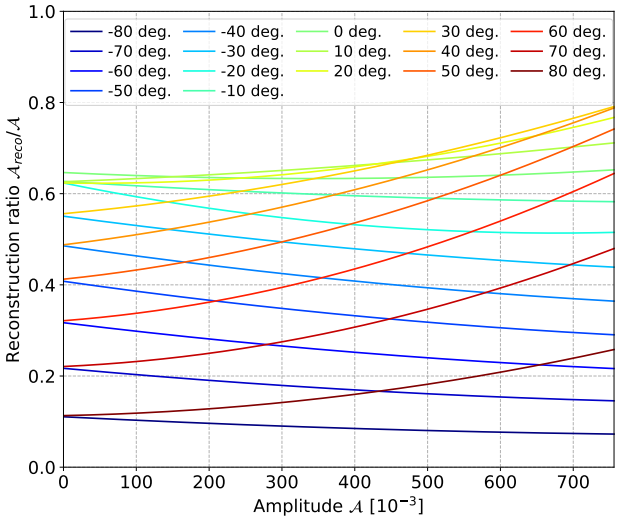


Figure 3.14: The ratio between the reconstructed dipole and the true dipole in the Monte-Carlo simulation with the same significance, while the amplitudes are chosen from 7×10^{-3} up to 9.56×10^{-1} for the true dipole. However, different energy bins cover different ranges of the true dipole amplitude. The curves show the propagation of the ratio with the declination of the true dipole from 80° to 80° with an angle bin size of 10°.

where λ_1, λ_2, and λ_3 are energy-dependent parameters that differ for proton and iron primaries. Figure 3.14 illustrates the ratio between the reconstructed dipole amplitude and the corresponding true input dipole. The sensitivity curves and bands for both 3σ and 5σ are shown in Figure 3.15, the corresponding points are in Table 3.2.

The sensitivity curves for the reconstructed dipole amplitudes, along with the corresponding bands that represent the true full-sky dipole amplitudes expected for the IceCube-Gen2 surface array, are displayed in Figure 3.15. These curves illustrate how significantly a measured dipole would deviate from the null hypothesis of perfect isotropy. Because the surface array observes only part of the sky, the reconstructed amplitudes $\mathcal{A}_{\text{reco}}$ are lower than the bands of true amplitude due to the field of view (FoV) of the IceCube-Gen2 surface array. When selecting the relevant energy interval for this study, we account for the region above 10^6 GeV and below $10^{8.8}$ GeV (overlap region with Auger), where upper limits exist, as well as the highest-energy points currently provided by IceCube. A dipole located close to either celestial pole becomes increasingly difficult to reconstruct with a one-dimensional projection method, since the projection distorts the amplitude and

Table 3.2: Shown are the points indicating the 3σ and 5σ sensitivity level bands for the reconstructed dipole for each energy bin E_i ($i = 1$ to 7), the upper and lower boundaries of the corresponding true dipole for both cases (sensitivity bands).

Median energy E_i	1.8 PeV	4.4 PeV	11 PeV	28 PeV	70 PeV	176 PeV	441 PeV
$\mathcal{A}_{\text{reco}}$ ($\times 10^{-3}$)	0.1224	0.2818	0.6981	1.8004	4.5926	12.0377	32.4111
	0.7067	1.6294	4.0435	10.4918	27.1849	74.3954	231.7128
$\mathcal{A}_{\text{reco}}$ ($\times 10^{-3}$)	0.2037	0.4694	1.1626	2.9980	7.6521	20.0448	53.8664
	1.1786	2.7160	6.7524	17.6033	46.0785	130.4091	483.0933
$\mathcal{A}_{\text{true}}$ ($\times 10^{-3}$)	0.7700	1.7720	4.3935	11.3220	28.9098	75.7195	204.2397
$\mathcal{A}_{\text{true}}$ ($\times 10^{-3}$)	1.2834	2.9524	7.3194	18.8748	48.1612	126.2120	342.2035

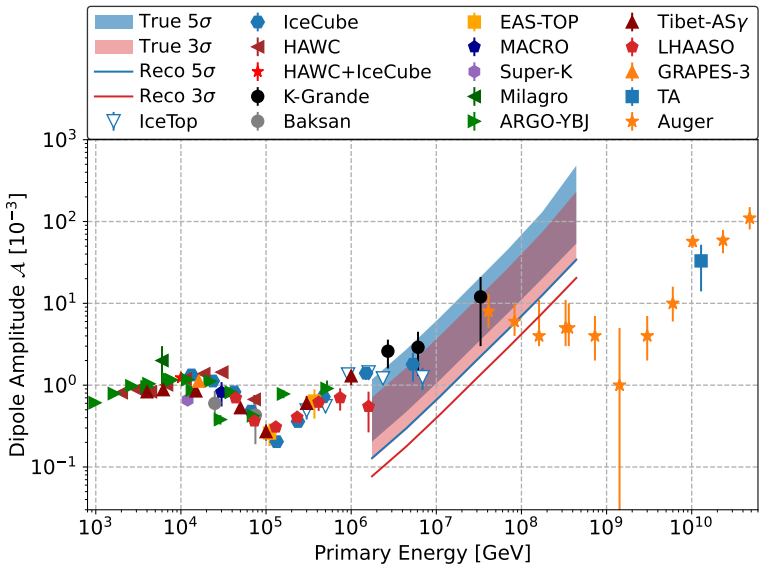


Figure 3.15: The curves show the 3σ and 5σ sensitivity of the IceCube-Gen2 surface array to the reconstructed dipole, while the bands represent the corresponding injected true dipole of 3σ and 5σ. The data points shown are reconstructed dipole amplitudes from various experiments [136, 268, 298, 299, 134, 133, 135, 269, 138, 140, 270, 300, 301]. The upper limits of KASCADE-Grande and Auger cover the range from 2.7 PeV to 8 EeV, where a significant energy gap remains.

phase. Consequently, a joint analysis combining IceCube-Gen2 with observatories in the Northern Hemisphere, such as LHAASO [137], will likely be necessary to achieve full-sky dipole sensitivity. Furthermore, very inclined dipole declinations must be explored, because present observations provide little information about the true dipole orientation and allow for a wide range of possibilities. Using Eq. (3.14) and Eq. (3.17), we also generate a sky map of the relative intensity at 176 PeV under the assumption of a reconstructed dipole detected at the 5σ level. In this case, the reconstructed amplitude is set to 9.029×10^{-3}, and with 20° top hat smoothing, as shown in Figure 3.16. The dipole orientation is set as $(270°, -10°)$ for the right ascension and declination. The corresponding amplitude of the injected true dipole ($\mathcal{A} = 1.437 \times 10^{-2}$) can be derived using the reconstruction ratio $\mathcal{A}_{\text{reco}}/\mathcal{A}$. We use HEALPix [281] (with $N_{\text{side}} = 64$) for the sky map; the size of each pixel tile in the sky map is approximately $(0.84°)^2$.

To conclude, we present the two-dimensional reconstruction-efficiency function for air showers induced by proton and iron primaries in the IceCube-Gen2 surface array, derived directly and consistently from the full set of existing CORSIKA simulations to capture the detector response. Building on this, we estimate the corresponding efficiencies for helium, nitrogen, and aluminium primaries by interpolating smoothly in the natural logarithm of their mass numbers, and then determine the total, composition-weighted efficiency using the H4a flux model to represent the expected cosmic-ray mixture. Next, we simulate cosmic-ray arrival directions using a dedicated toy Monte-Carlo framework, in which we inject a set of fifteen dipoles that span declinations from $-80°$ to $80°$ across seven energy bins that range from 10^6 GeV up to $10^{8.8}$ GeV in steps of 0.4 in $\log_{10}(E/\text{GeV})$. Altogether, this procedure produces 1,785 injected dipoles, with different anisotropy configurations. To study the anisotropy, we compare the actual sky map generated for each injected dipole

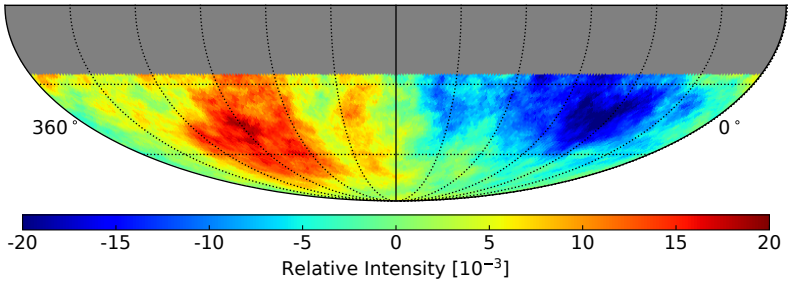


Figure 3.16: Simulated relative-intensity map as an example for the IceCube-Gen2 surface array, showing a 5σ reconstructed dipole with an amplitude of 9.029×10^{-3} at 176 PeV. The injected dipole orientation is set as $(270°, -10°)$ for the right ascension and declination. The corresponding amplitude of the true dipole is 1.437×10^{-2}.

with a corresponding reference map constructed without any dipole for all 1,785 cases, produce sky maps of the relative intensity together with their one-dimensional projections, and then fit these projections with first-harmonic functions to obtain the reconstructed amplitudes and phases. To assess the sensitivity of the IceCube-Gen2 surface array to a dipole anisotropy, we consider a null hypothesis and the propagation of the sigmas. Finally, we get the sensitivity function and show 3σ and 5σ sensitivity to the CR dipole anisotropy with curves and bands for the surface array. These results reveal the sensitivity of IceCube-Gen2 to cosmic-ray large-scale anisotropies, and are published at [84, 302].

In order to improve the accuracy of the sensitivity analysis for the IceCube-Gen2 surface array, we can extend the threshold of the zenith angle in the CORSIKA simulation up to 80° for scintillators, IceTop in IceCube-Gen2, in the future, which can provide a more accurate estimation of the reconstruction efficiency. With IceCube-Gen2, it will be possible to improve the sensitivity of the CR anisotropy in the energy range of the galactic-extragalactic transition of the CR origin. The sensitivity studies will be extended in the future by more realistic simulations.

3.2 Theoretical Considerations

In the previous section, we estimated the sensitivity of IceCube-Gen2 to high-energy cosmic-ray large-scale anisotropies. Here, we develop a general theoretical framework to assess an observatory's response to a dipolar anisotropy by evaluating the relative intensity of individual pixels on the sky map. This involves calculating the uncertainty associated with the projected dipole amplitude and deriving the corresponding sensitivity using statistical methods and theoretical models. Moreover, we aim to conduct a detailed and systematic analysis of the uncertainties involved in measuring the dipole amplitude, to understand how statistical noise and observational limitations propagate through the reconstruction. By these, one can derive the standard errors of the estimated parameters and quantify the sensitivity of both the dipole amplitude and phase with greater precision than before. By considering the observational parameters of the observatory, such as the declination and field of view of IceCube-Gen2, we can estimate its sensitivity to cosmic-ray anisotropies. We thus validate these theoretical results with MC simulations to ensure their consistency and accuracy. Additionally, we will compare the derived sensitivity function with the results obtained from simulations in the previous section to check for consistency.

3.2.1 Uncertainty of Dipole Amplitude

To approximately examine and quantify the uncertainty associated with the projected dipole amplitude, our analysis begins by evaluating the statistical fluctuations at each individual pixel on the sky map. The map is divided into $n_{\rm p}$ bands along the right ascension (RA) axis to ensure uniform coverage and statistical consistency. For each RA band, we compute the relative intensity for the i-th pixel within the k-th band, denoted by δI_{ki}, defined as the deviation of the observed event counts from the average counts at that pixel,

$$\delta I_{ki} = \left(N_{k,i} - \langle N_{k,i}\rangle\right) / \langle N_{k,i}\rangle \,, \tag{3.18}$$

where $N_{k,i}$ is the number of detected events, and $\langle N_{k,i}\rangle$ is the expected count for an isotropic flux. The corresponding uncertainty associated with each relative intensity measurement, which arises from both the observed and expected counts at the pixel,

$$\sigma_{ki} = \sqrt{\left(\frac{\partial \delta I_{ki}}{\partial N_i}\sigma_{N_i}\right)^2 + \kappa\left(\frac{\partial \delta I_{ki}}{\partial \langle N_i\rangle}\sigma_{\langle N_i\rangle}\right)^2}, \tag{3.19}$$

where $\sigma_{N_i} = \sqrt{N_i}$, $\sigma_{\langle N_i\rangle} = \sqrt{\langle N_i\rangle}$. The calibration factor κ is set as $1/n_s$, where n_s represents the number of samples in the reference map used to determine expected counts at each pixel. In this theoretical calculation, $n_s = 1$. However, in practical experimental analyses, a larger n_s leads to more accurate estimations of relative intensity. The uncertainty calculation can be simplified as follows:

$$\sigma_{ki} = \sqrt{\left(\frac{1}{\langle N_i\rangle}\sigma_{N_i}\right)^2 + \kappa\left(-\frac{N_i}{\langle N_i\rangle^2}\sigma_{\langle N_i\rangle}\right)^2} = \frac{N_i}{\langle N_i\rangle}\sqrt{\frac{1}{N_i} + \frac{\kappa}{\langle N_i\rangle}}. \tag{3.20}$$

In the 1D projection method, as discussed in Section 3.1.5, we average δI_{ki} over the entire k-th RA band to obtain $\delta\overline{I}_k$. To determine the corresponding overall uncertainty for each right ascension band, we calculate the average uncertainty in each right-ascension bin (α_k) as the square root of the average uncertainty with all pixels within this bin, namely,

$$\overline{\sigma}_k = \sqrt{\frac{1}{n^k_{\rm pix}}\sum_{i=1}^{n^k_{\rm pix}}\sigma^2_{ki}}, \tag{3.21}$$

which is due to the nonuniform arrangement of pixels on a sky map, a geometric factor (see the next section) should be considered in a continuous scenario for integration.

Sensitivity to ideal cosmic-ray anisotropy depends on the total statistics, the statistical methods used and the approach to fitting the anisotropy. For the method with 1D projection of δI and fitting $\delta\overline{I}_k$ to the first harmonic function (which yields the sidereal dipole amplitude, $\mathcal{A}_p$), the variance of a parameter, such as the projected dipole amplitude denoted as $\sigma_{\mathcal{A}}$, obtained from the fitting process provides a quantitative measure of the uncertainty in the estimation of $\mathcal{A}_p$. This variance is directly derived from the covariance matrix produced by the fitting process, specifically from the diagonal element corresponding to $\mathcal{A}$. To symbolically derive the full covariance matrix cov($\mathbf{X}$) for the parameters of the fit function as specified in Eq. (3.11), and corresponding to the combination of all the parameters, $\mathcal{A}$, φ, and $\mathcal{B}$, which are represented by the parameter vector, given by

$$\mathbf{X} = [\mathcal{A}, \varphi, \mathcal{B}], \tag{3.22}$$

whose best-fit values describe the reconstructed dipole from the 1D harmonic fit.

We begin the procedure by first evaluating all components of the Fisher information

matrix, denoted as $\boldsymbol{\mathcal{F}}$. This calculation is carried out by constructing the Jacobian matrix $\boldsymbol{\mathcal{J}}$, which contains the partial derivatives of the fit function with respect to all parameters of interest, and quantifies how sensitively the model responds to variations in each parameter. Once the Fisher matrix has been fully assembled from these gradients, the next step is to compute its inverse. Inverting $\boldsymbol{\mathcal{F}}$ transforms it directly into the covariance matrix, cov($\mathbf{X}$), which yields the estimated uncertainties of the fitted parameters [303, 304, 305]. The Jacobian matrix of the fit function $\delta\mathcal{I}(\alpha_i; \mathbf{X})$ with respect to the parameters $\mathbf{X}$ is a matrix of first-order partial derivatives. Each element of this matrix is defined as

$$\boldsymbol{\mathcal{J}}_i = \left[\frac{\partial\delta\mathcal{I}}{\partial\mathcal{A}}, \frac{\partial\delta\mathcal{I}}{\partial\varphi}, \frac{\partial\delta\mathcal{I}}{\partial\mathcal{B}}\right], \tag{3.23}$$

where each element corresponds to a derivative with respect to one of the parameters for each data point i involved in the fit. The Fisher Information Matrix for the least-squares fitting process is approximated by utilizing the Jacobian matrix, namely,

$$\boldsymbol{\mathcal{F}} = \boldsymbol{\mathcal{J}}^T \boldsymbol{\mathcal{W}} \boldsymbol{\mathcal{J}}, \tag{3.24}$$

where $\boldsymbol{\mathcal{W}}$ is the weight matrix, a diagonal matrix where each element $\boldsymbol{\mathcal{W}}_{kk} = 1/\sigma_k^2$ corresponds to the inverse of the squared standard deviation of the k-th data point. Here, $\boldsymbol{\mathcal{J}}^T$ is the transpose of the Jacobian matrix. Therefore, the Fisher Information Matrix is constructed by calculating the derivatives between any pair of parameters of the relative intensity, including the derivatives of the same parameter with itself, as follows:

$$\boldsymbol{\mathcal{F}}_{ij} = \sum_{k=1}^{n_{\rm p}} \frac{1}{\sigma_k^2} \frac{\partial\delta\mathcal{I}}{\partial\mathbf{X}_i} \frac{\partial\delta\mathcal{I}}{\partial\mathbf{X}_j}, \tag{3.25}$$

where n_p is the total number of data points in the fitting [303, 306]. The covariance matrix cov($\mathbf{X}$) of the parameter vector $\mathbf{X}$ is calculated as the inverse of the Fisher information matrix ($\boldsymbol{\mathcal{F}}_{ij}^{-1}$). This inversion process converts the Fisher Information Matrix into its corresponding covariance matrix, which quantifies the variance and covariance of the parameter estimates [307, 308]. Each element of this matrix reveals the uncertainty and the relationships between pairs of parameters within the model, as expressed by:

$$\mathrm{cov}(\mathbf{X}) = \begin{bmatrix} \boldsymbol{\mathcal{F}}_{\mathcal{A}\mathcal{A}} & \boldsymbol{\mathcal{F}}_{\mathcal{A}\varphi} & \boldsymbol{\mathcal{F}}_{\mathcal{A}\mathcal{B}} \\ \boldsymbol{\mathcal{F}}_{\varphi\mathcal{A}} & \boldsymbol{\mathcal{F}}_{\varphi\varphi} & \boldsymbol{\mathcal{F}}_{\varphi\mathcal{B}} \\ \boldsymbol{\mathcal{F}}_{\mathcal{B}\mathcal{A}} & \boldsymbol{\mathcal{F}}_{\mathcal{B}\varphi} & \boldsymbol{\mathcal{F}}_{\mathcal{B}\mathcal{B}} \end{bmatrix}^{-1}. \tag{3.26}$$

The standard error of estimation of parameters ($\mathbf{X}$) in a fitting process is calculated via the square root of the corresponding covariance matrix, namely,

$$\sigma_{\mathbf{X}} = \sqrt{\mathrm{cov}(\mathbf{X})}. \tag{3.27}$$

Therefore, the uncertainty of the estimation of the one-dimensional projected dipole amplitude is given by $(\boldsymbol{\mathcal{F}}_{\mathcal{A}\mathcal{A}})^{-1/2}$ on the whole equatorial plane with n_α RA band, namely,

$$\sigma_{\mathcal{A}} = \left(\sum_{k}^{n_\alpha} \frac{\cos^2(\alpha_k - \varphi)}{\overline{\sigma}_k^2}\right)^{-1/2}, \tag{3.28}$$

where $\overline{\sigma}_1^2 \simeq \overline{\sigma}_2^2 \simeq ... \simeq \overline{\sigma}_{n_p}^2$, since the difference of relative intensity and number in each RA band is tiny when the statistics are large, and the anisotropies are weak. Therefore, one can independently calculate the uncertainty $\overline{\sigma}_k^2$ from Eq. (3.28), which is useful for estimating the significance and the observatory's sensitivity to cosmic-ray anisotropies.

3.2.2 Sensitivity of Observations

To efficiently estimate the capabilities of a specific ground-based observatory with respect to cosmic-ray large-scale anisotropies, whether still in the planning stages or lacking sufficient operational time to obtain statistics, we are currently exploring a straightforward and calculable method. This method is designed to provide an approximate estimation of the anisotropy sensitivity of an observatory by considering the overall significance function of the estimated one-dimensional projected dipole amplitude, as shown in Eq. (3.28).

Now, continuing the discussion developed in the previous section, we shift our focus toward examining a sky map characterized by an anisotropic (dipole) distribution of cosmic-ray arrival directions. To quantify this distribution accurately, we begin by estimating the total number of detected events corresponding to a particular right ascension bin α_k on the sky map. Specifically, we consider the following expression:

$$N_k = \sum_{i=1}^{N_\delta} N_0 \epsilon(\delta_i) \cos\left(\delta_i - \delta_{\rm obs}\right), \tag{3.29}$$

where $\epsilon(\delta_i)$ denotes the air-shower reconstruction efficiency for cosmic rays arriving at the observatory from a specific declination interval δ_i, and the term $\cos(\delta_i - \delta_{\rm obs})$ represents the geometric acceptance of the observatory at that declination. The factor N_0 is the number of events in the pixel corresponding to the observatory's vertical maximum acceptance, where the zenith angle is 0. In this section, we assume 100% reconstruction efficiencies for IceCube-Gen2. Clearly, N_k, which quantifies the total number of events at a specific right ascension, depends critically on both the number of pixels and the event counts within each pixel that fall under the maximum acceptance of IceCube-Gen2 that corresponds to the declination. Specifically, this relationship can be expressed as follows:

$$N_k = \lambda n_\delta N_0, \tag{3.30}$$

where $n_\delta = n_{\rm pix}$, the parameter λ is interpreted as the weighting factor applied across all pixels at α_k, taking into account Eq. (3.29). This approach applies not only in scenarios where pixel size decreases but also in an equivalent continuous scenario where integration is used, and the pixel size is exceedingly small, namely,

$$\lambda = \frac{\sum_i^{n_\delta} \cos\left(\delta_i - \delta_{\rm obs}\right) \cos\delta_i}{\sum_i^{N_\delta} \cos\delta_i}, \tag{3.31}$$

where $\cos\delta$ serves as the calibration weighting for averaging over the celestial sphere, and $\delta_{\rm obs}$ represents the declination of the observatory. To transition this discrete setup into a continuous representation, the expression can be restructured into an integral form. By multiplying both the numerator and the denominator by the interval, denoted as $\Delta\delta$ and letting $\Delta\delta$ approach zero, we achieve the following integration:

$$\begin{aligned} \lambda &\simeq \frac{\int_{\delta_{\rm L}}^{\delta_{\rm U}} \cos\left(\delta - \delta_{\rm obs}\right) \cos\delta d\delta}{\int_{\delta_{\rm L}}^{\delta_{\rm U}} \cos\delta d\delta} \\ &= \frac{2(\delta_{\rm U} - \delta_{\rm L}) \cos\delta_{\rm obs} + \sin\left(2\delta_{\rm U} - \delta_{\rm obs}\right) - \sin\left(2\delta_{\rm L} - \delta_{\rm obs}\right)}{4(\sin\delta_{\rm U} - \sin\delta_{\rm L})}, \end{aligned} \tag{3.32}$$

where $\delta_{\rm L}$ and $\delta_{\rm U}$represent the lower and upper limits of the field of view (FoV) of a given ground-based detector, respectively. Referring to Eq. (3.20), when N_{ki} becomes significantly larger than the ratio between the event counts in any given pixel, and the average counts in that pixel approaches unity, $N_{ki}/\langle N_{ki}\rangle \simeq 1$. Furthermore, the expectation value, $\langle N_{ki}\rangle$

is a δ-dependent function, taking the form

$$\langle N_{ki} \rangle = N_0 \epsilon(\delta_{ki}) \cos(\delta_{ki} - \delta_{\rm obs}), \tag{3.33}$$

where we let $\epsilon(\delta_{ki}) = 1$ in this calculation to account for a cutoff at the zenith where efficiency is less than 1. Under these simplified yet representative conditions, and by incorporating the relationships defined in Eq. (3.29) and Eq. (3.30), Eq. (3.21) can be reformulated to more accurately describe the dependence on these parameters, as follows:

$$\overline{\sigma}_k \simeq \left(\frac{\lambda N_0}{N_k} \sum_{i=1}^{n_\delta} \frac{2}{\langle N_{ki} \rangle} \right)^{1/2}. \tag{3.34}$$

Now, assuming that the pixel size approaches zero, we can replace the discrete summation by a continuous integral, and take into account the spherical calibration by multiplying the factor $\cos\delta$, which leads to the approximation:

$$\overline{\sigma}_k^2 \simeq \frac{2\lambda}{N_k \mathcal{W}} \int_{\delta_{\rm L}}^{\delta_{\rm U}} \frac{\cos\delta}{\cos(\delta - \delta_{\rm obs})} d\delta = \frac{\gamma}{N_k}, \tag{3.35}$$

where γ represents the calibration coefficient, and $\mathcal{W} = \sin\delta_{\rm U} - \sin\delta_{\rm L}$ provides the normalization factor for the spherical calibration. This coefficient encapsulates the combined influence of various observational parameters, particularly the observatory's geographic location (declination) and its FoV, both of which directly affect the measurement uncertainty and directional sensitivity. It serves as a correction term that scales the reconstructed amplitude to account for geometric and exposure-related effects, and is given by

$$\gamma = \frac{2\lambda}{\mathcal{W}} \left[(\delta_{\rm U} - \delta_{\rm L}) \cos\delta_{\rm obs} + \ln\left(\frac{\cos\delta_{\mathcal{U}}}{\cos\delta_{\mathcal{L}}} \right) \sin\delta_{\rm obs} \right], \tag{3.36}$$

where we define $\delta_{\mathcal{L}} = \delta_{\rm L} - \delta_{\rm obs}$ and $\delta_{\mathcal{U}} = \delta_{\rm U} - \delta_{\rm obs}$ for simplicity. We obtain the numerical results $\lambda = 0.587$ and $\gamma = 2.487$ by considering $\delta_{\rm L} = -90°$, $\delta_{\rm U} = -10°$ and $\delta_{\rm obs} = -90°$ for the IceCube-Gen2 geometry in Eq. (3.32) and Eq. (3.36).

Furthermore, the uncertainty of the reconstructed dipole amplitude, as outlined in Eq. (3.28), can be reduced to a more straightforward form, providing an interpretation of the amplitude's variability. The simplified expression for this uncertainty is:

$$\sigma_{\mathcal{A}}^{-2} = \frac{N_k}{\gamma} \sum_{k}^{n_\alpha} \cos^2(\alpha_k - \varphi), \tag{3.37}$$

where n_α is the number of right ascension (RA) bands, the sum equivalent to an integration concerning α from 0 to 2π and equal to π. This integration is normalized by the sum of the derivative of α across the entire right ascension plane, specifically, $\Delta\alpha = 2\pi/n_\alpha$ for the discrete case and $\mathrm{d}\alpha = 2\pi/\mathrm{n}_\alpha$ for the integration. Consider the averaged uncertainty Eq. (3.34) in each right ascension bin, one has

$$\sigma_{\mathcal{A}} = \sqrt{\frac{2\gamma}{N_{\rm tot}}}, \tag{3.38}$$

with $N_{\rm tot} = n_\alpha N_k$ representing the total number of events. Finally, we examine the propagation of the significance of a projected dipole deviation from isotropy by considering a null hypothesis that varies solely with respect to the observed dipole amplitude and the statistics. This is described in Equation (3.12), which follows the form:

$$n_\sigma(\mathcal{A}_{\rm p}) = \frac{1}{\sqrt{2\gamma}} \mathcal{A}_{\rm p} \sqrt{N_{\rm tot}}. \tag{3.39}$$

In the case of IceCube-Gen2, where we have assumed a 100% reconstruction efficiency throughout this study, as discussed earlier, the resulting estimate yields $1/\sqrt{2\gamma} \simeq 0.448$. This value is in very good agreement with the expectations derived from our analytical framework as presented in Section 3.1.5, aligning well with the expectations, as shown in the left panel of Figure 3.13, which illustrates the close agreement.

Moreover, the orientation of the dipole is characterized by the right ascension (phase) and declination (Dec) of the dipole's location within the distribution of cosmic rays. While we do not have a method to accurately determine the dipole's Dec (which will be discussed in the next chapter), we can reconstruct the dipole phase. Similar to the estimation of amplitude, the uncertainty of the dipole phase is quantified as follows:

$$\sigma_{\varphi}^{-2} = \sum_{k}^{n_\alpha} \frac{\mathcal{A}_p^2 \sin(\alpha_k - \varphi)^2}{\overline{\sigma}_k^2}, \tag{3.40}$$

the sum equivalent to an integration concerning α from 0 to 2π. This integration is normalized by the sum of the derivative of α across the entire right ascension plane,

$$\sigma_{\varphi} = \frac{1}{\mathcal{A}_p}\sqrt{\frac{2\gamma}{N_{\rm tot}}}. \tag{3.41}$$

Since the dipole phase, φ is a variable measured in radians ranging from 0 to 2π, it differs fundamentally from $\mathcal{A}$, whose significance is derived from the deviation from the null hypothesis. For the significance of the dipole phase, we propose that it be defined as the deviation from half of the RA plane, specifically calculated as ($n_\sigma(\varphi) = \pi/\sigma_\varphi$),

$$n_\sigma(\varphi) = \frac{\pi}{\sqrt{2\gamma}}\mathcal{A}_{\rm p}\sqrt{N_{\rm tot}}, \tag{3.42}$$

which suggests a direct relationship to the significance of the amplitude, $n_\sigma(\varphi) = \pi\, n_\sigma(\mathcal{A}_{\rm p})$. The same idea for the significance of the reconstruction of angles can be applied to the dipole's declination in subsequent chapters, e.g., for the analysis presented in Chapter 6.

In summary, we have derived closed-form expressions for the uncertainties and significances of the one-dimensional, projected dipole amplitude and phase, Eqs. (3.38), (3.39), and (3.41), that depend solely on the total event statistics $N_{\rm tot}$ and a geometric–exposure calibration factor γ (itself determined by the observatory's declination coverage and FoV via λ, refer to Eqs. (3.32) and (3.36). These relations reproduce the Monte Carlo trends (Section 3.1.5) and therefore provide a reliable, fast estimator of dipole sensitivity and parameter precision without requiring new MC simulations. Practically, they enable rapid forecasting and optimization of observatory performance (e.g., trade-offs between exposure, FoV, and energy binning), quick consistency checks against data, and straightforward propagation of expected improvements as $N_{\rm tot}$ grows, making them especially convenient for observatory design studies and early-stage analyses.

Chapter 4

Optimized Anisotropy Reconstruction

The sensitivity of ground-based cosmic-ray observatories is hindered by the lack of statistics at high energies, particularly above the PeV range, as discussed in Section 3.1. Over the past few decades, experiments have provided long-term and significant observations of subtle sidereal anisotropies. However, these observations are often limited by the restricted coverage area of the observatories and the scarcity of high-energy data. As a result, there are substantial uncertainties in the reconstructed dipole amplitudes and corresponding data. Moreover, when the amplitude is minimal, extensive statistical data still lead to considerable uncertainties in both the reconstructed dipole amplitudes and phase. Therefore, it is necessary to develop improved methods for reconstructing the large-scale anisotropies against the lack of statistics and to reduce the instabilities of the reconstruction of anisotropies, such as the dipole.

4.1 Reconstructions with Partial Sky

The lack of full sky coverage can introduce biases in the reconstruction of anisotropic signals, such as dipoles or multipoles. To overcome these limitations, various techniques have been developed to project the observed sky map onto equatorial coordinates and properly account for missing regions. These reconstructions aim to minimize the impact of partial sky coverage on the analysis of cosmic-ray arrival directions. In the following sections, we will delve into the mathematical framework used to achieve these reconstructions, starting with an introduction to spherical angles and their relevance in understanding the distribution of cosmic rays in equatorial coordinates.

4.1.1 Spherical Angles

Cosmic-ray anisotropy analyses are usually based on 3-dimensional distributions with equatorial coordinates (α, δ), where α represents right ascension (RA) and δ denotes declination (Dec). To achieve precise theoretical analyses of anisotropies, we have to develop suitable distribution functions using equatorial coordinates. The function will enable us to accurately describe the relative intensity at any point within the three-dimensional distributions of cosmic-ray arrival directions, characterized by dipole or multipole orientations.

Now, let's consider a celestial sphere in spherical coordinates defined by xyz axes. Within the framework of spherical trigonometry, we explore the connections among three unit vectors: $\overrightarrow{OA}$, $\overrightarrow{OC}$ and $\overrightarrow{OD}$. These vectors originate from a common point and extend to the vertices of a spherical triangle on the celestial sphere. The orientation and Cartesian coordinates of these vectors are on the Northern Hemisphere and defined as follows: $\overrightarrow{OA}$ is aligned along the z-axis and sits on the North Pole, $\overrightarrow{OC}$ is a random vector on the Northern Hemisphere, and $\overrightarrow{OD}$ is positioned to an orientation with positive x coordinate in the xz plane for simplifying the analysis. The points A, C and, D form a spherical

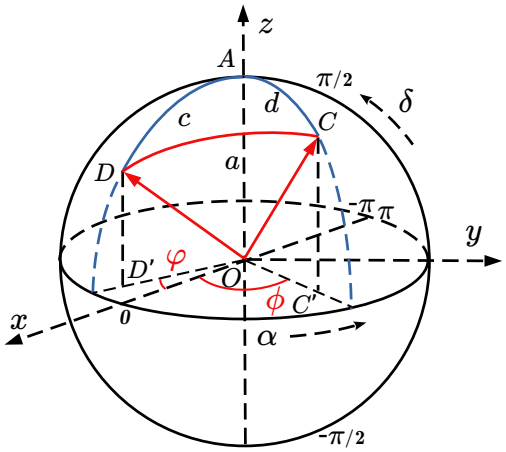


Figure 4.1: Spherical trigonometry, illustrating both the spherical and equatorial coordinate systems. The diagram shows the geometric relation between two points on the sphere: a random point C and the dipole maximum point D.

triangle $\triangle ACD$. The sides opposite these points are labelled a, c and d corresponding to the arcs $\widehat{CD}$, $\widehat{AD}$ and $\widehat{AC}$, respectively. These sides correspond to the two-dimensional plane angles $\angle COD$, $\angle AOD$ and $\angle AOC$. The spherical angles are represented by the points $\angle A$, $\angle C$ and $\angle D$ or simply as A, C and D for brevity. The vectors $\overrightarrow{OA}$, $\overrightarrow{OC}$ and $\overrightarrow{OD}$ can be expressed utilizing both the plane angles and the spherical angles within the three-dimensional Cartesian coordinates (xyz), as previously defined, specifically:

$$\begin{aligned} \overrightarrow{OA} &= (0, 0, 1), \\ \overrightarrow{OD} &= (\sin c \cos \varphi, \sin c \sin \varphi, \cos c), \\ \overrightarrow{OC} &= (\sin d \cos \phi, \sin d \sin \phi, \cos d), \end{aligned} \tag{4.1}$$

where the spherical angle A corresponds to the sum of ϕ and φ, can be obtained from the spherical law of sines, and is also equal to the projected angle on the xy plane, represented by $\angle C'OD'$. Furthermore, taking the scalar (dot) product of the two vectors $\overrightarrow{OD}$ and $\overrightarrow{OC}$ yields $\overrightarrow{OD} \cdot \overrightarrow{OC} = \cos a$, which directly leads to the expression for the angle $\angle COD$ as:

$$\cos a = \cos d \cos c + \sin d \sin c \cos A, \tag{4.2}$$

which also represents the spherical angle of the arcs $\widehat{CD}$, or in other words, the angular distance between points C and D. This expression is precisely what we expect for the distribution function of a dipole with relative intensity.

Now, let's consider the scenario of anisotropic distributions of cosmic rays in their arrival directions. We start with an original dipole located at the point D on the same celestial sphere, denoted as $(x_\mathrm{d}, y_\mathrm{d}, z_\mathrm{d})$, as we previously discussed. This point is characterized by spherical coordinates (x, y, z). Our goal is to describe the distribution of cosmic rays at the random point C on an equatorial spherical surface. This requires converting spherical coordinates (x, y, z) into equatorial coordinates (α, δ) with right ascension and declination. Now, we assign the point $(0, 1, 0)$ to represent $\alpha = 0$, $\delta = 0$. The point $(0, -1, 0)$ corresponding to $\alpha = -\pi$ or π, $\delta = 0$. The northern pole (A) is represented by $\alpha = 0$, $\delta = \pi/2$. The point C is denoted by (α, δ), and the dipole locals at D, is specified

as $(\alpha_{\rm d}, \delta_{\rm d})$. Eventually, the spherical angles of arcs c and d together with the angle A can be transformed into their equivalents in the equatorial coordinate system, namely,

$$d \to \frac{\pi}{2} - \delta, \quad c \to \frac{\pi}{2} - \delta_{\rm d}, \quad A \to \alpha - \alpha_{\rm d}, \tag{4.3}$$

which is expressed as a function of the relative intensity of the 3D spherical cosmic-ray anisotropic distribution in equatorial coordinates (α, δ) with the orientation of an original dipole $(\alpha_{\rm d}\delta_{\rm d})$. The spherical angle between the random point C and the dipole D corresponds to the angular distance between two points. Considering Eq.(4.2), we obtain the distribution function, which is characterized as

$$\mathcal{D}(\alpha, \delta) = \sin\delta \sin\delta_{\rm d} + \cos\delta \cos\delta_{\rm d} \cos\left(\alpha - \alpha_{\rm d}\right), \tag{4.4}$$

where $\mathcal{D}$ is defined as the distribution function of the original unit dipole. When $\alpha = \alpha_{\rm d}$ and $\delta = \delta_{\rm d}$, $\mathcal{D} = 1$, indicating the point of maximum relative intensity in the distribution of the unit dipole, known as the dipole hotspot, located at the point D, which consistent with the dipole's setup that we assumed. Observably, the first term in the expression represents the relative declination distance between the random point (α, δ) and the dipole $(\alpha_{\rm d}, \delta_{\rm d})$, while the second term indicates the relative RA distance. The relative intensity across the entire celestial sphere is therefore defined as $\delta\mathcal{I}(\alpha, \delta) = \mathcal{A}\mathcal{D}(\alpha, \delta)$, namely,

$$\delta\mathcal{I}(\alpha, \delta) = \mathcal{A}\sin\delta \sin\delta_{\rm d} + \mathcal{A}\cos\delta \cos\delta_{\rm d} \cos\left(\alpha - \alpha_{\rm d}\right), \tag{4.5}$$

where $\mathcal{A}$ denotes the amplitude of the original dipole. The first term concerns only declination and represents the vertical shift in the dipole's harmonic distribution of relative intensity, relative to the equatorial plane. Meanwhile, the second term describes the distribution along the equatorial plane, which relates to both right ascension and declination.

4.1.2 Dipole Projections

In cosmic-ray anisotropy studies, the relative intensity of cosmic-ray distributions on the celestial sphere is displayed on sky maps and usually plotted in equatorial coordinates with right ascension (α) and declination (δ), using the HEALPix mapping method [281], such that the areas of all pixels at a given resolution are identical. The one-dimensional projection of the sky map involves averaging the physical quantity, in this case, the relative intensity, of all pixels at certain right ascensions across all declinations within the Field of View (FoV). In this study, a proper theoretical method for averaging the dipole relative intensity, denoted as $\delta\mathcal{I}$, as specified by Eq. (4.5) over the range of declinations, is needed. To properly calculate the averaged dipole intensity over declination, we should consider the relative scale of the spherical surface area elements and average over the declination with appropriate weighting to obtain a function of right ascension alone.

Specifically, this geometric consideration requires that the averaging procedure include an additional weight factor to ensure that every region of the celestial sphere contributes in proportion to its physical area. This follows from the fact that the width of a band of constant declination decreases progressively as one moves away from the equator toward either pole, so each latitude strip covers a smaller surface area than a corresponding strip near the equator. To account for this shrinking area, the correct weight for an average over declination is the factor $\cos\delta$, since the surface-area element on the sphere is proportional to $\delta d \cos\delta$. Therefore, when projecting the two-dimensional sky map onto the equatorial plane in one dimension, the projection should be formed by integrating the relative intensity over declination with the appropriate $\cos\delta$ weight and then normalizing by the full range

of δ. Thus, the one-dimensional projection of the relative intensity sky map onto the equatorial plane can be expressed as follows:

$$\delta\hat{\mathcal{I}}(\alpha) = \frac{\int_{\delta_{\rm L}}^{\delta_{\rm U}} \delta\mathcal{I}(\alpha,\delta)\cos\delta d\delta}{\int_{\delta_{\rm L}}^{\delta_{\rm U}} \cos\delta d\delta}, \tag{4.6}$$

where $\delta\mathcal{I}(\alpha,\delta)$ represents the original dipole before projection, as defined by Eq. (4.5), $\delta\hat{\mathcal{I}}(\alpha)$ donates the sidereal projected dipole after the integration over δ, while $\delta_{\rm L}$ and $\delta_{\rm U}$ represent the lower and upper limits, respectively, of the certain FoV of a surface observatory on Earth. This projection method has been proven through MC simulations on a sky map with an injected dipole. By inserting Eq. (4.5) into Eq. (4.6), we derive the expression of the projected dipole, which is distributed along the equatorial plane only and includes a shift term along the amplitude axis, namely,

$$\begin{aligned}\hat{\mathcal{D}}(\alpha) =& \frac{1}{2}(\sin\delta_{\rm U} + \sin\delta_{\rm L})\sin\delta_{\rm d} \\ &+ \frac{\delta_{\rm U} - \delta_{\rm L} + \sin\delta_{\rm U}\cos\delta_{\rm U} - \sin\delta_{\rm L}\cos\delta_{\rm L}}{2(\sin\delta_{\rm U} - \sin\delta_{\rm L})}\cos\delta_{\rm d}\cos(\alpha - \alpha_{\rm d}),\end{aligned} \tag{4.7}$$

which is a function of right ascension only. This corresponds to the experimental data of anisotropies and represents the true modulated dipole projected with respect to right ascension, specifically, $\delta I(\alpha) = \mathcal{A}\hat{\mathcal{D}}(\alpha)$, where $\mathcal{A}$ is the original 3-dimensional dipole amplitude. The projection shifts along the declination with a coefficient,

$$c_{\rm s} = \frac{1}{2}(\sin\delta_{\rm U} + \sin\delta_{\rm L}), \tag{4.8}$$

which is a relative shift, and becomes the actual shift when it is multiplied by the vertical components of the dipole's self-effect, specifically, $\sin\delta_{\rm d}$. The projection term, moving along right ascension and similar to the shift term, involves coefficients that are determined by the observatory's FoV and depend on its upper and lower limits, namely,

$$c_{\rm p} = \frac{\delta_{\rm U} - \delta_{\rm L} + \sin\delta_{\rm U}\cos\delta_{\rm U} - \sin\delta_{\rm L}\cos\delta_{\rm L}}{2(\sin\delta_{\rm U} - \sin\delta_{\rm L})}, \tag{4.9}$$

which also involves the actual modulation after being multiplied by the parallel component, $\cos\delta_{\rm d}$. For a given dipole declination $\delta_{\rm d}$, this projection factor approaches unity in the limiting case where the field of view becomes extremely narrow and effectively collapses onto the equatorial plane, i.e., when $\delta_{\rm U} \to 0$ and $\delta_{\rm L} \to 0$, the projection coefficient approaches one. This term was also derived in the paper [309] using a different approach with the East-West method, but without the shift term, as shown in Eq. (4.8). Furthermore, when combined with the projection factor $c_{\rm p}$, the expression provides a straightforward way to reconstruct the dipole declination in a simplified one-dimensional framework. The equations above can be validated using an MC simulation that involves injecting a dipole with a specific amplitude into a sky map. This dipole is then projected onto the equatorial (right ascension) plane, covering a specific observatory's FoV, and fitted to a first harmonic function. The parameter functions c_s and c_p are confirmed by comparing the ratio of the projected amplitude to the injected amplitude.

4.1.3 Reconstruction of Dipole Declinations

In this section, we examine the theoretical framework for reconstructing the dipole declination and highlight the importance of including a shift term for a more accurate and

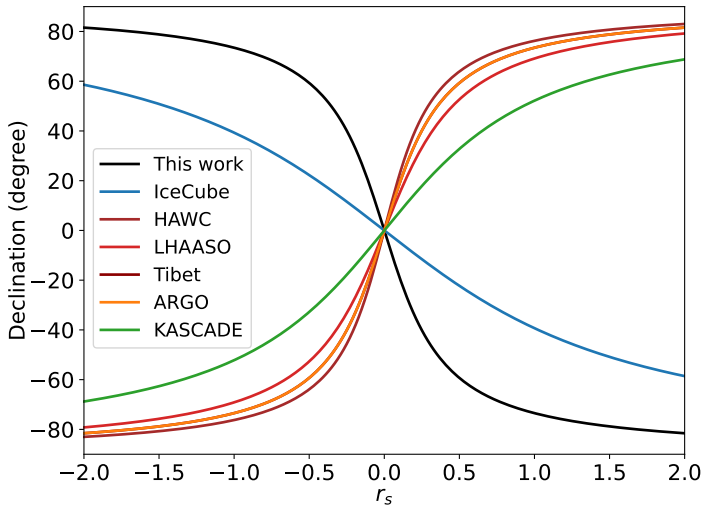


Figure 4.2: Expected reconstruction of the dipole declination ($\delta_{\rm d}$) from the ratio between the monopole shift and the projected dipole amplitude. Vertical shift terms at fixed energies, once measured, align along the corresponding theoretical curves.

stable result. The reconstruction is based on fitting a specific function to the observed data (arrival directions), in which both the projected dipole amplitude, $\mathcal{A}$p, and the monopole shift term, $\mathcal{B}$, are extracted through a first-harmonic analysis. These parameters are essential for correctly describing the dipole structure in the observed field of view. Once these parameters are determined, the dipole declination δd can be calculated by taking the arctangent of the ratio between the shift term $\mathcal{B}$ and the projected dipole amplitude $\mathcal{A}_{\rm p}$.

In principle, the most appropriate method for reconstructing a dipole anisotropy is to explicitly incorporate the shift term into the fit, since neglecting it leads to biased or unstable results, especially for dipoles located away from the equatorial plane. The corresponding fit function used in this reconstruction is

$$\delta I(\alpha) = \mathcal{A}_p \cos(\alpha - \alpha_d) + \mathcal{B}, \tag{4.10}$$

where the parameters $\mathcal{A}_p$ and $\mathcal{B}$ can be determined directly from the harmonic fitting procedure applied to the one-dimensional projection of the sky map. These two parameters specify a complete description of the projected dipole, as well as the true three-dimensional dipole structure within the limited field of view. By substituting the explicit expressions for the projection coefficients $c_{\rm s}$ and $c_{\rm p}$, and using the relation $\mathcal{B} = (c_{\rm s}/c_{\rm p})\mathcal{A}_p \tan\delta_d$, we can rewrite the problem in a way that allows us to recover the dipole declination. Thus, the reconstructed dipole declination is obtained as

$$\delta_d = \arctan\left(\frac{c_p \mathcal{B}}{c_s \mathcal{A}_p}\right), \tag{4.11}$$

as shown in Figure 4.2, the x-axis is defined by the experimental parameters c_p, c_s, $\mathcal{B}$ and $\mathcal{A}_p$. The uncertainty of δ_d depends on the uncertainties of $\mathcal{B}$ and $\mathcal{A}_p$, denoted as $\sigma_{\mathcal{B}}$ and $\sigma_{\mathcal{A}_p}$, respectively. The uncertainty in δ_d is calculated as follows:

$$\sigma^2_{\delta_d} = \left(\frac{\partial\delta_d}{\partial\mathcal{B}}\sigma_{\mathcal{B}}\right)^2 + \left(\frac{\partial\delta_d}{\partial\mathcal{A}_p}\sigma_{\mathcal{A}_p}\right)^2, \tag{4.12}$$

where the intermediate variable $u = c_{\rm p}\mathcal{B}/c_{\rm s}\mathcal{A}_{\rm p}$ is defined for convenience, and where the

derivative is written as $\partial\delta_d$ as $\partial \arctan(u)$. Carrying out the differentiation yields

$$\sigma^2_{\delta_d} = \left(\frac{u}{1+u^2}\right)^2 \left(\frac{\sigma^2_{\mathcal{A}_p}}{\mathcal{A}_p^2} + \frac{\sigma^2_{\mathcal{B}}}{\mathcal{B}^2}\right), \tag{4.13}$$

which can also be expressed as a function of the significance of $\mathcal{A}_\mathrm{p}$ and $\mathcal{B}$, denoted by $\mathcal{S}_\mathcal{A}$ and $\mathcal{S}_\mathcal{B}$, respectively. Additionally, this expression can also be rewritten in terms of the significances of $\mathcal{A}_\mathrm{p}$ and $\mathcal{B}$, denoted by $\mathcal{R} = \mathcal{B}/\mathcal{A}_\mathrm{p}$. It is specifically expressed as

$$\sigma^2_{\delta_d} = \left(\frac{c\mathcal{R}}{1+c^2\mathcal{R}^2}\right)^2 \left(\frac{1}{\mathcal{S}^2_\mathcal{A}} + \frac{1}{\mathcal{S}^2_\mathcal{B}}\right). \tag{4.14}$$

where the covariance term $\mathrm{Cov}(\mathcal{A}_\mathrm{p}, \mathcal{B})$ is small and approximates 0, and is ignored.

The large FoV tends to increase the sensitivity of the dipole's declination to $\mathcal{R}$, which consequently leads to greater uncertainty in the reconstruction process. Consequently, when an observatory has an exceptionally wide FoV, it can be advantageous to subdivide the observable sky into two distinct FoV segments, so that the dipole declination can be reconstructed independently in each region. By performing a separate analysis in both segments and subsequently averaging the two reconstructed values, one obtains a more stable and reliable estimate of the true dipole orientation. This approach provides a reliable estimate of the true orientation of the dipole when performing a one-dimensional projection of the sky map, especially under a broad FoV, and the projection effects become weak.

4.2 Optimized Estimation of Significance

Building on the estimation techniques discussed in Section 4.1, the next step focuses on optimizing the estimation of the significance of cosmic-ray anisotropy detection. In this section, we further refine this approach by introducing an optimized method for estimating significance. We incorporate the weights related to the number of events observed in each pixel, leading to a more accurate uncertainty estimation.

4.2.1 Weighted Uncertainties

By taking into account variations in the number of events across pixels, we ensure that the dipole detection is both robust and sensitive to actual anisotropic features in the cosmic-ray flux. Considering the weights in the averaged uncertainty to consider the number of events in each of the n_pix pixels, given by, $\overline{\sigma}_k = \sqrt{\sum_i \sigma^2_{ki} N_i / \sum_i N_i}$, namely,

$$\overline{\sigma}^2_k = \frac{1}{\sum_{i=1}^{N_\mathrm{pix}} N_i} \sum_{i=1}^{N_\mathrm{pix}} \frac{{N_i}^2}{\langle N_i \rangle^2} \left(\frac{1}{N_i} + \frac{1}{\langle N_i \rangle}\right) N_i. \tag{4.15}$$

Taking into account the definition of relative intensity, one has

$$\frac{{N_i}^2}{\langle N_i \rangle^2} \left(\frac{1}{N_i} + \frac{1}{\langle N_i \rangle}\right) N_i = \left(1 + \delta\mathcal{I}_{ki}\right)^2 \left(2 + \delta\mathcal{I}_{ki}\right), \tag{4.16}$$

where $\mathcal{I}_{ki}$ is the dipole distribution function as discussed in Chapter 4, namely,

$$\delta\mathcal{I}_{ki} = \mathcal{A} \sin\delta_i \sin\delta_\mathrm{d} + \mathcal{A} \cos\delta_i \cos\delta_\mathrm{d} \cos\left(\alpha_k - \alpha_\mathrm{d}\right). \tag{4.17}$$

In the next step, we incorporate Eq. (4.17) into Eq. (4.16), and we additionally introduce a weighting factor to account for the nonuniform distribution of events across declination

more accurately. Consequently, the uncertainty associated with a specific right-ascension band, denoted by α_k, can be reformulated into a more optimized and physically meaningful expression: $\hat{\sigma}k^2 = \int \delta\overline{\sigma}k^2 w, d\delta/\mathcal{W}$. This transformation from discrete to continuous is denoted as the transition from $\delta\mathcal{I}_{ki}(\alpha_k, \delta_i) \to \delta\mathcal{I}(\alpha, \delta)$, can give a smooth description of the weighted uncertainty across the sky, and $\hat{\sigma}_k^2$ is expressed as follows:

$$\hat{\sigma}_k^2 = \frac{1}{N_k \mathcal{W}} \int_{\delta_{\rm L}}^{\delta_{\rm U}} (1 + \delta\mathcal{I})^2 \, (2 + \delta\mathcal{I}) \, w(\delta) d\delta, \tag{4.18}$$

where N_k denotes the total number of events contributing to the right-ascension band α_k. To move from a discrete sum over declination bins to a continuous formulation, we replace δ_i with the continuous variable δ. The factor $w(\delta) = \cos\delta$ serves as the appropriate spherical-geometry weight, ensuring that each declination contributes in proportion to its actual surface area on the sky. The quantity $\mathcal{W}$ is the sum of the weight over the effective field of view of an observatory, namely $\mathcal{W} = \int_{\delta_{\rm L}}^{\delta_{\rm U}} \cos\delta d\delta$.

To obtain an optimized fit for the dipole parameters, it is necessary to incorporate the vertical shift of the fitting function, which arises from the combined effects of the observatory's restricted field of view and the orientation of the dipole in declination. This shift naturally appears in the dipole expression,

$$\delta I(\alpha) = \mathcal{A}_p \cos(\alpha - \alpha_d) + \frac{c_s}{c_p} \mathcal{A}_p \tan\delta_d, \tag{4.19}$$

which expresses the projected dipole as a combination of a cosine modulation and a constant offset determined by observational geometry, can also be written in the equivalent form

$$\delta I(\alpha) = \mathcal{A}\, c_{\rm p} \, \cos(\alpha - \alpha_d) + \mathcal{A}\, c_{\rm s}, \tag{4.20}$$

where the coefficients c_s and c_p depend directly on the field of view and latitude of the observatory, as derived in Chapter 5 and shown explicitly in Eq. (4.8) and (4.9). The uncertainty associated with the reconstructed amplitude $\mathcal{A}_p$ can then be determined using the covariance matrix derived in Chapter 3. By applying this formalism to Eq. (4.19), the inverse squared uncertainty takes the form

$$\sigma_{\mathcal{A}}^{-2} = \sum_{k}^{n_{\rm p}} \frac{(\cos(\alpha_k - \varphi) + \hat{c_s}/\hat{c_p})^2}{\hat{\sigma_k}^2}, \tag{4.21}$$

where $\hat{c_s} = c_s \sin\delta_d$, $\hat{c_p} = c_p \cos\delta_d$ and n_p is the number of right ascension bands, as well as the number of α points used for the harmonic fit of the sidereal dipole.

Now, let's look at details about N_k (in Eq. (4.18)) of each right ascension band on the projected map. Link to the expression of the projected dipole, Eq. (4.19), the number of events on each band follows the distribution of the projected dipole, namely,

$$\delta I_k = \frac{N_k - \langle N_k \rangle}{\langle N_k \rangle} = \mathcal{A}_p \cos(\alpha_k - \alpha_d) + \frac{\hat{c_s}}{\hat{c_p}} \mathcal{A}_p, \tag{4.22}$$

where $\langle N_k \rangle$ is the average number of events at every band (whole band as one "pixel"), δI_k is the relative intensity at the k-th band, meaning, the deviation of the counts at the k-th band from the average counts at the same band and therefore δI_k also equivalent to Eq. (4.19). The counts at the k-th band read,

$$N_k = \langle N_k \rangle \left(\mathcal{A}_p \cos(\alpha_k - \alpha_d) + \frac{\hat{c_s}}{\hat{c_p}} \mathcal{A}_p + 1 \right), \tag{4.23}$$

where $\langle N_k \rangle$ itself is approximately proportional to the total number of events on the sky,

weighted by the width of each right ascension interval. Specifically,

$$\langle N_k \rangle = \frac{N_{\rm tot}}{2\pi}\Delta\alpha_k, \quad \langle dN \rangle = \frac{N_{\rm tot}}{2\pi}d\alpha, \tag{4.24}$$

where the second expression provides the continuous differential version of the same relation and is useful for analytical treatments of the projected event distribution.

The uncertainty $\hat{\sigma}_k(\alpha_k) = \hat{\sigma}(\alpha)$ appearing in Eq. (4.21) is defined in the limit where the width of the right-ascension band becomes infinitesimally small, that is, when $\Delta\alpha_k = \Delta\alpha \to 0$ and $\Delta\alpha = d\alpha$. Under the same limiting treatment, Eq. (4.23) can be rewritten as

$$N_k = \langle N_k \rangle \left(\mathcal{A}_p \cos(\alpha_k - \alpha_d) + \frac{\hat{c_s}}{\hat{c_p}}\mathcal{A}_p + 1 \right). \tag{4.25}$$

Now, insert Eq. (4.17) into Eq. (4.18) and consider the mentioned approximation as well as for the declination bin δ_i. Now, we transition from discrete indices to continuous variables by replacing, namely, δ_i with δ, α_k with α, $\langle N_k \rangle$ with $\langle dN \rangle$. In this continuous formulation, the uncertainty associated with a right-ascension value α, corresponding to a band of infinitesimal width $\mathrm{d}\alpha$, is expressed as

$$\hat{\sigma}^2 = \frac{2\pi}{\mathcal{W}N_{tot}(\delta I + 1)d\alpha} \int_{\delta_{\rm L}}^{\delta_{\rm U}} (1 + \delta\mathcal{I})^2 \, (2 + \delta\mathcal{I}) \, w(\delta) d\delta. \tag{4.26}$$

The declination integral above, leading to the expression

$$\begin{aligned} f(\alpha) &= 2(\sin\delta_{\rm U} - \sin\delta_{\rm L}) + \frac{5}{2}\mathcal{A}\big[(\sin^2\delta_{\rm U} - \sin^2\delta_{\rm L})\sin\delta_d \\ &\quad + (\delta_{\rm U} - \delta_{\rm L} + \sin\delta_{\rm U}\cos\delta_{\rm U} - \sin\delta_{\rm L}\cos\delta_{\rm L})\cos\delta_d\cos(\alpha - \alpha_{\rm d})\big] + \mathcal{O}(\mathcal{A}^2) \\ &\simeq (\sin\delta_{\rm U} - \sin\delta_{\rm L})(2 + 5\delta I(\alpha)), \end{aligned} \tag{4.27}$$

where second- and third-order contributions in $\mathcal{A}$ have been omitted, since both $\mathcal{A}$ and $\mathcal{A}_p$ are of the magnitudes ranging from 10^{-3} to 10^{-1}. The second term in this expression is simply the projected dipole $\delta I(\alpha)$ multiplied by the constant factor $(\sin\delta_{\rm U} - \sin\delta_{\rm L})$, which is equal to $\mathcal{W}$. Using Eq. (4.19), the uncertainty associated with the estimation of $\mathcal{A}_p$, as introduced in Eq. (4.21), can finally be reformulated into the compact expression

$$\sigma_{\mathcal{A}}^{-2} = \frac{N_{tot}}{2\pi\mathcal{A}_p^2} \int_{-\pi}^{\pi} \frac{(\delta I(\alpha) + 1)\delta I^2(\alpha)}{2 + 5\delta I(\alpha)} d\alpha, \tag{4.28}$$

where one sees that the weight factor $\mathcal{W}$ naturally cancels out of the expression.

We define the integral term as $g(\alpha)$, which can be approximated by expanding it into a Taylor series around $\delta I = 0$, noting that δI varies only within a relatively small range, typically between 10^{-3} and 10^{-1}. In this expansion, we treat δI as the expansion variable

$$g(\alpha) = \frac{\delta I^2}{2} - \frac{3\delta I^3}{4} + \mathcal{O}(\delta I^4), \tag{4.29}$$

where we retain only the leading and next-to-leading terms of the series, as higher-order contributions are negligible for the magnitude of anisotropies considered. Making use of the functional form as δI described in Eq. (4.22), we obtain the result of the integration:

$$G(\mathcal{A}_p) = \frac{\pi}{4}\left[(2 + 4\hat{c}^2)\mathcal{A}_p^2 - (9\hat{c} + 6\hat{c}^3)\mathcal{A}_p^3\right], \tag{4.30}$$

where, for convenience, we have introduced the ratio $\hat{c}_{\rm s}/\hat{c}_{\rm p} = \hat{c}$. Next, by substituting the integrated result $G(\mathcal{A}_{\rm p})$ into the uncertainty expression given in Eq. (4.28), and evaluating the significance of a dipole's deviation from isotropy under a null hypothesis, we can

derive the final results concerning the observational significance of the sidereal dipole. This sensitivity, denoted as $n_\sigma(\mathcal{A}_\mathrm{p})$ and analogous to Eq. (3.39), can be written as

$$n_\sigma(\mathcal{A}_\mathrm{p}) = \sqrt{\frac{N_\mathrm{tot}}{8}}\left[(2+4\hat{c}^2)\mathcal{A}_\mathrm{p}^2 - (9\hat{c}+6\hat{c}^3)\mathcal{A}_\mathrm{p}^3\right]^{1/2}, \tag{4.31}$$

which is a function of $\mathcal{A}_\mathrm{p}$. This significance is described as a function of the one-dimensional projected dipole amplitude, $\mathcal{A}_\mathrm{p}$, which, as calculated in Chapter 3, is not a linear approximation but has been refined to be more precise and optimized.

4.2.2 Angular Resolution of Dipole

The orientation of the dipole is characterized by the right ascension and declination of the dipole's location in the cosmic-ray arrival-direction distribution, which together specify the phase and latitude of the dominant dipole component on the sky and thus determine its observable projection. Based on Eq. (3.40),

$$\sigma_\varphi^{-2} = \sum_k^{n_\alpha} \frac{\mathcal{A}_p^2 \sin(\alpha_k - \varphi)^2}{\hat{\sigma_k}^2}, \tag{4.32}$$

which quantifies the statistical uncertainty of the reconstructed dipole phase. We apply the same procedure used previously to convert discrete sums over right ascension into continuous integrals, thereby obtaining a compact expression that is directly comparable across different binning choices and observational layouts; in particular, following the transformation used for Eq. (4.28), namely,

$$\sigma_\varphi^{-2} = \frac{N_\mathrm{tot}\mathcal{A}_p^2}{2\pi}\int_{-\pi}^{\pi} \frac{(\delta I(\alpha)+1)\sin^2(\alpha-\varphi)}{2+5\delta I(\alpha)} d\alpha, \tag{4.33}$$

where we employ the same small-anisotropy expansion as before to retain analytic control. Integrating with the integral term $g(\alpha)$, we obtain the series

$$G(\mathcal{A}_p) = \pi\left(\frac{1}{2} - \frac{3}{4}\hat{c}\mathcal{A}_p + \frac{3}{8}\mathcal{A}_p^2 + \frac{15}{8}\hat{c}^2\mathcal{A}_p^2 + \mathcal{O}(\mathcal{A}_p^3)\right), \tag{4.34}$$

where we drop higher-order terms in $\mathcal{A}p$ since these contributions are subleading in the parameter range of interest for cosmic-ray dipole studies. This immediately yields,

$$\sigma_\varphi = \frac{2}{\mathcal{A}_p\sqrt{N_{tot}}}\left(1 - \frac{3}{2}\hat{c}\mathcal{A}_p + \frac{3}{4}\mathcal{A}_p^2 + \frac{15}{4}\hat{c}^2\mathcal{A}_p^2\right)^{-1/2}, \tag{4.35}$$

which represents the uncertainty of the reconstructed dipole phase after the full projection-and-fit procedure is applied. Since the angular variables (α and φ) vary on the compact interval from $-\pi$ to π, the most reliable way to evaluate phase sensitivity is therefore to consider its uncertainty σ_φ directly.

Now, let's deal with the declination of the reconstructed dipole in greater detail. By considering the Fisher information matrix within the least-squares fitting framework, along with the corresponding covariance matrix derived from the fit function as shown in Eq. (4.19), we can obtain the statistical uncertainty of the estimated dipole declination, δ_d, through the inverse-square-root of its diagonal Fisher element, $(\boldsymbol{\mathcal{F}}_{\delta\delta})^{-1/2}$, namely,

$$\sigma_\delta^{-2} = \sum_k^{n_\mathrm{p}} \frac{c_s^2}{c_p^2}\frac{\mathcal{A}_p^2}{\cos^4\delta_d}\frac{1}{\hat{\sigma_k}^2} = \frac{N_{tot}}{2\pi}\frac{c_s^2}{c_p^2}\frac{\mathcal{A}_p^2}{\cos^4\delta_d}\int_{-\pi}^{\pi}\frac{\delta I(\alpha)+1}{2+5\delta I(\alpha)}d\alpha, \tag{4.36}$$

where we have transformed the discrete sum into a continuous integral consistent with

earlier sections. Next, by expanding the integrand in a Taylor series and retaining only the first two nontrivial orders that contribute to the significance, we obtain an analytic approximation for the declination uncertainty:

$$\sigma_\delta = \frac{\sqrt{2}\cos\delta_d \sin\delta_d}{\hat{c}\mathcal{A}_p\sqrt{N_{\rm tot}}}\left(1 - \frac{4}{3}\hat{c}\mathcal{A}_p + \frac{15}{8}\mathcal{A}_p^2 + \frac{15}{4}\hat{c}^2\mathcal{A}_p^2 + \mathcal{O}(\mathcal{A}_p^3)\right)^{-1/2}, \tag{4.37}$$

where we retain only the dominant contributions controlling the scaling with respect to $\mathcal{A}p$ and the total number of detected events Ntot, while noting that more precise higher-order terms may be added later if necessary. If desired, one may also illustrate the behavior of this uncertainty by plotting the corresponding significance curves for different total event counts, such as 10^4, 10^5, 10^6, 10^7, 10^8, respectively.

4.2.3 Uncertainty with Ideal Exposure

Since the exposure of surface detectors depends on the zenith angle through a characteristic $\cos\theta$ modulation, the resulting average event rate recorded across the sky naturally acquires a dependence on declination. In the simplest and most idealized approximation, assuming uniform operational conditions and neglecting detector irregularities, this effect can be expressed as a proportionality to $\cos(\delta_j - \delta_{\rm obs})$, where $\delta_{\rm obs}$ denotes the geographical latitude of the observatory. Therefore, the distribution of the event rate can be taken into account when estimating the dipole uncertainty, as this directly affects the calculation of the statistical significance and the comparison with observed data.

We begin with the discrete form of the dipole distribution function, denoted as $\delta\mathcal{I}(\alpha_i, \delta_j)$, which corresponds to the relative intensity in the pixel located at the i-th right ascension band and the j-th declination bin. This function is given by,

$$\delta\mathcal{I}(\alpha_i, \delta_j) = \mathcal{A}\sin\delta_j \sin\delta_{\rm d} + \mathcal{A}\cos\delta_j\cos\delta_{\rm d}\cos(\alpha_i - \alpha_{\rm d}), \tag{4.38}$$

which represents the relative intensity (RI) in each pixel. It is fully equivalent to the continuous formulation $\delta\mathcal{I}(\alpha, \delta)$, which was introduced earlier in Section 4.1.2, but written here in discrete form to match the structure of sky maps produced by the HEALPix binning. Following the same procedure as in Eq. (3.33), we can obtain the discrete representation of the full event count distribution, i.e., $\delta\mathcal{I}$ on a sky map, by incorporating the dipole term, the exposure, and the efficiency of the observatory. The resulting expression for the expected number of cosmic-ray events in each pixel is

$$N(\alpha_i, \delta_j) = N_{\rm d}\,\epsilon(\theta)\Big[1 + \delta\mathcal{I}(\alpha_i, \delta_j)\Big]\cos(\delta_j - \delta_{\rm obs}), \tag{4.39}$$

where $N_{\rm d}$ is the number of events in the pixel with the dipole maximum (hotspot), $\epsilon(\theta)$ is the air-shower reconstruction efficiency at an observatory. The expression of $N_{\rm d}$ can be obtained by summing the both sides of Eq. (4.39) with respect to i and j, namely,

$$N_{\rm tot} = \sum_{i=1}^{n_\delta}\sum_{j=1}^{n_\alpha} N_{\rm d}\epsilon(\theta)\Big[1 + \delta\mathcal{I}(\alpha_i, \delta_j)\Big]\cos(\delta_j - \delta_{\rm obs}), \tag{4.40}$$

where $N_{\rm tot} = \sum_i\sum_j N(\alpha_i, \delta_j)$, n_α represents the total bins of right ascension (RA), while n_δ donates as the number of pixels in every RA bin. To get the integral form, Eq. (4.40) should be multiplied by $d\delta$ and divided by $d\delta$, as well as $d\alpha$, which are given by

$$\mathrm{d}\delta = \frac{1}{n_\delta}\int_{\delta_{\rm L}}^{\delta_{\rm U}} \mathrm{d}\delta\cos\delta = \frac{\mathcal{W}}{n_\delta}, \quad \mathrm{d}\alpha = \frac{1}{n_\alpha}\int_{-\pi}^{\pi}\mathrm{d}\alpha = \frac{2\pi}{n_\alpha}, \tag{4.41}$$

where $\mathcal{W} = \sin\delta_{\rm U} - \sin\delta_{\rm L}$ accounts for the spherical geometry weighting needed when averaging or integrating over declination. Therefore, by multiplying by $d\delta$ and dividing by $\mathcal{W}/n_\delta$, and multiplying by $\mathrm{d}\alpha$ and dividing by $2\pi/n_\alpha$, the sum of Eq. (4.39) becomes equivalent to the integration with respect to α and δ, specifically,

$$N_{\rm tot} = \frac{N_{\rm d} n_{\rm pix}}{2\pi\mathcal{W}} \int_{-\pi}^{\pi}\int_{\delta_{\rm L}}^{\delta_{\rm U}} \Big[1 + \delta\mathcal{I}(\alpha,\delta)\Big] \cos\left(\delta - \delta_{\rm obs}\right)\cos\delta\,\mathrm{d}\delta\,\mathrm{d}\alpha, \tag{4.42}$$

where the total number of pixels within the FoV is given by $n_{\rm pix} = n_\delta n_\alpha$ and the geometric calibration has already been included in λ and $\mathcal{W}$.

Taking into account the expression for $\delta\mathcal{I}(\alpha,\delta)$, as specified in Eq. (4.5), the total number of events can be obtained by utilizing the fraction of the count in the hotspot, namely, $N_{\rm tot} = N_{\rm d}\, n_{\rm pix}\, \lambda$, where λ is the result of the integration decided by $2\pi\mathcal{W}$, specifically:

$$\begin{aligned}\lambda = {} & \frac{1}{2\mathcal{W}}(\delta_{\rm U} - \delta_{\rm L})\cos\delta_{\rm obs} + \frac{1}{4\mathcal{W}}\Big[\sin\left(2\delta_{\rm U} - \delta_{\rm obs}\right) - \sin\left(2\delta_{\rm L} - \delta_{\rm obs}\right)\Big] \\ & + \frac{1}{4\mathcal{W}}\mathcal{A}\sin\delta_{\rm d}\Big[\cos\left(\delta_{\rm L} + \delta_{\rm obs}\right) - \cos\left(\delta_{\rm U} + \delta_{\rm obs}\right)\Big] \\ & + \frac{1}{12\mathcal{W}}\mathcal{A}\sin\delta_{\rm d}\Big[\cos\left(3\delta_{\rm L} - \delta_{\rm obs}\right) - \cos\left(3\delta_{\rm U} - \delta_{\rm obs}\right)\Big].\end{aligned} \tag{4.43}$$

Therefore, we can obtain the expression of $N_{\rm d}$ and replace it in Eq. (4.39), we can obtain:

$$N(\alpha,\delta) = \frac{N_{\rm tot}}{n_{\rm pix}\lambda}\epsilon(\theta)\Big[1 + \delta\mathcal{I}(\alpha,\delta)\Big]\cos\left(\delta - \delta_{\rm obs}\right), \tag{4.44}$$

where ϵ may be taken as unity for many experiments that operate above their full-efficiency threshold for reconstructed cosmic-ray arrival directions. The quantity $N_{\rm tot}$ can be obtained from the cosmic-ray flux within a known exposure, or the experiments' data. Equation (4.44) therefore provides a description of the expected statistical distribution of arrival directions for a dipole with a given orientation and amplitude on the sky. Meanwhile, the corresponding statistical uncertainty can be estimated from the counts in each pixel,

$$\sigma_{ij}^2 = \frac{N_{ij}^2}{\langle N_{ij}\rangle^2}\left(\frac{1}{N_{ij}} + \frac{1}{\langle N_{ij}\rangle}\right) = \frac{1}{N_{ij}}\left(1 + \delta\mathcal{I}_{ij}\right)^2\left(2 + \delta\mathcal{I}_{ij}\right), \tag{4.45}$$

where N_{ij} is given by $N(\alpha_i,\delta_j)$, as expressed in Eq. (4.44). Therefore, σ_{ij}^2 can be rewritten in terms of global observatory parameters and acceptance as:

$$\sigma_{ij}^2 = \frac{n_{\rm pix}\lambda}{N_{\rm tot}}\frac{\left(1 + \delta\mathcal{I}_{ij}\right)\left(2 + \delta\mathcal{I}_{ij}\right)}{\epsilon(\theta)\cos\left(\delta_j - \delta_{\rm obs}\right)}, \tag{4.46}$$

where λ is the declination-weighting factor defined previously. To obtain the uncertainty for a one-dimensional projection, the next step is to compute the averaged uncertainty $\hat{\sigma}_k^2$ for each right-ascension band. This requires summing (or integrating) over all declination pixels in that band, incorporating both the event-count distribution and the acceptance modulation. The transition from the discrete form $\delta\mathcal{I}_{ki}(\alpha_k,\delta_i)$ with $\delta\mathcal{I}(\alpha,\delta)$, allowing for smooth integration over the full declination range.

To convert the discrete summation into a continuous integral form, we begin by multiplying Eq. (4.46) by an infinitesimal declination interval $\mathrm{d}\delta$, and then divide by the corresponding normalization factor taken from the right-hand side of Eq. (4.41), where $\mathrm{d}\delta = \mathcal{W}/n_\delta$, following directly the same logic applied earlier in the conversion step of Eq. (3.35), which ensures a consistent transition from discrete binning to continuous angular integration. After performing this transformation, we replace the pixel-level uncertainties

with their averaged values across each full RA bin, adopting precisely the same procedure described in Eq. (3.21), with the substitution $n_\delta = n_{\rm pix}^k$ to denote the total number of pixels contained within the k-th RA strip. A crucial constraint in this construction is that the relation $n_\delta = n_{\rm pix}/n_\alpha$ must always remain satisfied, since, as emphasized previously, the geometric dependence of the FoV is already absorbed into the weighting factors λ and $\mathcal{W}$, and the pixel grids in RA and Dec are chosen to have equal angular spacing. With these, the average statistical uncertainty associated with all pixels contained within a given RA bin can ultimately be expressed in the following continuous integral form:

$$\hat{\sigma}_k^2 = \frac{\lambda\, n_\alpha}{N_{\rm tot}\mathcal{W}} \int_{\delta_{\rm L}}^{\delta_{\rm U}} \frac{(1+\delta\mathcal{I})\,(2+\delta\mathcal{I})}{\cos\left(\delta-\delta_{\rm obs}\right)} \cos\delta\, {\rm d}\delta, \tag{4.47}$$

where N_k denotes the total number of events recorded within the RA band centered on α_k, and the ϵ term has been neglected. Once a specific observatory latitude $\delta_{\rm obs}$ is chosen, one can directly determine the statistical uncertainties associated with both the dipole amplitude and the dipole phase by employing Eq. (3.28) together with Eq. (3.40). For convenience, we introduce the notation $F(\delta_{\rm L}, \delta_{\rm U}, \delta_{\rm obs}; \alpha)$. Next, based on Eq. (3.28), we convert it into an integral form by using ${\rm d}\alpha = 2\pi/n_\alpha$ from Eq. (4.41), i.e.,

$$\sigma_{\mathcal{A}}^{-2} = \frac{N_{\rm tot}\mathcal{W}}{2\pi\lambda} \int_0^{2\pi} \frac{\cos^2(\alpha-\varphi)}{F(\delta_{\rm L}, \delta_{\rm U}, \delta_{\rm obs}; \alpha)}\, {\rm d}\alpha, \tag{4.48}$$

where the factors of n_α cancel out, and where $\varphi = \alpha_{\rm d}$ denotes the input dipole phase. Since the phase merely shifts the integrand along the α-axis and because the integration extends over the entire interval $[0, 2\pi]$, the integral is invariant under such a shift. Hence, without loss of generality, we set $\varphi = 0$ in both numerator and denominator. With this simplification, the original sensitivity expression in Eq. (3.39) can be recast into

$$n_\sigma(\mathcal{A}_{\rm p}) = \sqrt{\frac{\mathcal{W}M}{2\pi\lambda}}\mathcal{A}_{\rm p}\sqrt{N_{\rm tot}}, \tag{4.49}$$

where M is a function that depends on the parameters $\delta_{\rm d}$, $\mathcal{A}_{\rm p}$, $\delta_{\rm L}$, $\delta_{\rm U}$ and $\delta_{\rm obs}$, represents the part that is integral over the entire α plane. Similarly, following the structure of Eq. (3.40), the uncertainty associated with the dipole phase can be estimated through

$$\sigma_\varphi^{-2} = \frac{N_{\rm tot}\mathcal{W}}{2\pi\lambda} \int_0^{2\pi} \frac{\mathcal{A}_{\rm p}^2 \sin^2(\alpha-\varphi)}{F(\delta_{\rm L}, \delta_{\rm U}, \delta_{\rm obs}; \alpha)}\, {\rm d}\alpha, \tag{4.50}$$

where the sin term arises naturally from the covariance elements of the Fisher information matrix associated with variations in the dipole phase. Since the phase significance is defined as $n_\sigma(\varphi) = \pi/\sigma_\varphi$, as introduced earlier in Section 3.2, we obtain an improved and more

Table 4.1: Shown are values of $\sqrt{\mathcal{W}M/2\pi\lambda}$. Top panel: dipole declination $\delta_{\rm d}$ from $-80°$ to $80°$ in $20°$ steps, for amplitudes $\mathcal{A}_{\rm p} = 0.01$ and $\mathcal{A}_{\rm p} = 0.1$. Bottom panel: for $\delta_{\rm d} = 0°, 30°$ and amplitudes $\mathcal{A}_{\rm p}$ from 0.01 to 0.10 in steps of 0.01.

$\delta_{\rm d}$ (deg)	$-80°$	$-60°$	$-40°$	$-20°$	$0°$	$20°$	$40°$	$60°$	$80°$
$\mathcal{A} = 0.01$	0.4448	0.4451	0.4458	0.4468	0.4478	0.4489	0.4498	0.4505	0.4509
$\mathcal{A} = 0.1$	0.4189	0.4226	0.4296	0.4391	0.4499	0.4606	0.4700	0.4769	0.4805

$\mathcal{A}$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
$\delta_{\rm d} = 0°$	0.4478	0.4479	0.4480	0.4482	0.4483	0.4486	0.4488	0.4491	0.4495	0.4499
$\delta_{\rm d} = 30°$	0.4494	0.4510	0.4527	0.4544	0.4561	0.4579	0.4598	0.4616	0.4636	0.4656

Table 4.2: Shown are values of $\sqrt{\pi \mathcal{W} V/2\lambda}$. Top panel: dipole declination $\delta_{\rm d}$ from -80° to 80° in 20° steps, for amplitudes $\mathcal{A}_{\rm p} = 0.01$ and $\mathcal{A}_{\rm p} = 0.1$. Bottom panel: values for $\delta_{\rm d} = 0^\circ, 30^\circ$ and amplitudes $\mathcal{A}_{\rm p}$ from 0.01 to 0.10 in steps of 0.01.

$\delta_{\rm d}$ (deg)	-80°	-60°	-40°	-20°	0°	20°	40°	60°	80°
$\mathcal{A} = 0.01$	1.3992	1.4003	1.4025	1.4054	1.4088	1.4121	1.4151	1.4173	1.4185
$\mathcal{A} = 0.1$	1.3176	1.3285	1.3492	1.3777	1.4109	1.4449	1.4757	1.4989	1.5114

$\mathcal{A}$	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	0.10
$\delta_{\rm d} = 0^\circ$	1.4088	1.4089	1.4090	1.4091	1.4093	1.4095	1.4098	1.4101	1.4105	1.4109
$\delta_{\rm d} = 30^\circ$	1.4137	1.4187	1.4237	1.4289	1.4341	1.4393	1.4446	1.4500	1.4555	1.4610

symmetric expression for the corresponding sensitivity. Specifically, this yields

$$n_\sigma(\varphi) = \sqrt{\frac{\pi \mathcal{W} V}{2\lambda}} \mathcal{A}_{\rm p} \sqrt{N_{\rm tot}}, \tag{4.51}$$

where V represents the part that is integrated over the entire α plane, analogous to W.

For the case of IceCube-Gen2, and assuming a 100% reconstruction efficiency, meanwhile, the FoV is limited by $\delta_{\rm L} = -90^\circ$ and $\delta_{\rm U} = -10^\circ$, as mentioned previously. Using Eq. (4.49), we perform a systematic scan of the dipole declination $\delta_{\rm d}$ from -80° to 80° in steps of 20° for two representative amplitudes, $\mathcal{A}_{\rm p} = 0.01$ and $\mathcal{A}_{\rm p} = 0.1$, with the results summarized in the top panel of Table 4.1. Furthermore, for fixed dipole declinations $\delta_{\rm d} = 0^\circ$ and 30°, we vary the amplitude from 0.01 to 0.10 in increments of 0.01, yielding the results shown in the bottom panel of Table 4.1. The corresponding coefficients $\sqrt{\pi \mathcal{W} V/(2\lambda)}$, which determine the phase sensitivity, are listed in Table 4.2. We find that at larger dipole amplitudes ($\mathcal{A}_{\rm p} \gtrsim 0.1$), the sensitivities of both the amplitude and phase vary more rapidly with the dipole declination. When the dipole is located on the equatorial plane, i.e., $\delta_{\rm d} = 0$, the sensitivities remain relatively stable as the amplitude changes. However, at non-zero declinations, such as $\delta_{\rm d} = 30^\circ$, the sensitivities increase noticeably for larger amplitudes. Compared to the rough estimate in Section 3.2, the framework developed here directly includes the declination-dependent acceptance $\cos(\delta - \delta_{\rm obs})$ and the dipole in the pixel and band weights, and gives a more accurate and reliable uncertainty evaluation.

4.3 Gradient Method

Inspired by the contour (isoheight) maps commonly used in geography and geophysics, we propose applying an analogous concept to the study of cosmic-ray large-scale anisotropies, focusing in particular on variations associated with dipole structures. Specifically, the gradient method we construct here is a novel approach for reconstructing the dipole anisotropy of cosmic rays, utilizing the gradient normal to isoheight lines on the three-dimensional celestial sphere to precisely measure and characterize the intensity of cosmic rays across different regions of the sky. This method leverages the analysis of isoheight lines, which are conceptualized as contours of constant cosmic-ray intensity, analogous to isobar lines in meteorology. These lines facilitate detailed mapping of anisotropic patterns on the celestial sphere. By examining the normals (i.e., local gradients) to these isoheight lines, one can obtain a clearer distribution of cosmic-ray arrival directions, particularly the amplitude and coordinates of the dipole's orientation. All averages below are exposure-weighted using a reference map that encodes detector acceptance.

4.3.1 Interpretation of the Gradient Method

To conduct a systematic analysis, the celestial sphere in equatorial coordinates is divided into arc bands, similar to slices of a spherical "toast." These bands are kept fixed in counts while the coordinate belts are rotated (rather than rotating the events themselves) and segmented along right ascension or declination. The exposure-weighted average of the relative intensity is then computed within each band, using inverse-variance (reference-count) weights.

The main idea of the "Gradient Method" is to search for the maximum and minimum average amplitudes within these spherical bands. We begin by selecting a central point characterized by its median declination and right ascension, around which a horizontal band is formed. This band is then rotated from horizontal to vertical, with rotation angles ranging from -90° to 90°, and each segment has a bin size of 9°. For each rotation, we calculate the (exposure–weighted) average relative intensity $\delta\overline{I}_i$ across the resulting ring belts, identifying the maximum and minimum values, denoted as $\delta\overline{I}_{\min}$ and $\delta\overline{I}_{\max}$ respectively, at certain rotation angles, specifically $\Delta\theta|_{\delta\overline{I}_{\min}}$ and $\Delta\theta|_{\delta\overline{I}_{\max}}$. Clearly, there exists a relationship between the rotation angles and the dipole's declination, which can be expressed as $\delta_{\rm d} = \pi/2 - \theta_{\rm d}$, where $\theta_{\rm d}$ represents the rotation angle corresponding to the absolute maximum (by magnitude) of the exposure–weighted average relative intensity. The rotation angle $\theta_{\rm d}$ corresponding to the dipole's maximum intensity can be specified as

$$\theta_{\rm d} = \theta\big(\text{argmax}\left\{\left|\delta\overline{I}_{\min}\right|, \left|\delta\overline{I}_{\max}\right|\right\}\big) \times \begin{cases} -1, & \text{if } \left|\delta\overline{I}_{\min}\right| > \left|\delta\overline{I}_{\max}\right|, \\ 1, & \text{if } \left|\delta\overline{I}_{\min}\right| \leq \left|\delta\overline{I}_{\max}\right|. \end{cases} \tag{4.52}$$

Following the selection of the belt that exhibits either the minimum or maximum average relative intensity, as indicated by the rotation angle $\theta_{\rm d}$, we proceed to rotate this chosen belt along the right ascension axis (α) over a range from 0° to 180°. This rotational procedure allows us to compute, for every position of the belt, the exposure–weighted average relative intensity $\delta\overline{I}(\alpha)$ for each segment within the belt. We then determine the right ascension of the dipole, $\alpha_{\rm d}$, by fitting $\delta\overline{I}(\alpha)$ with a first harmonic,

$$\delta\overline{I}(\alpha) = A_1 \cos(\alpha - \alpha_{\rm d}) + B_1 \tag{4.53}$$

which uniquely fixes $\alpha_{\rm d}$, up to the expected 180° dipole ambiguity.

Clearly, the sky is segmented into $n_{\rm bin}$ belts, with each belt being associated with a particular rotation angle that determines either a minimum or maximum value. The best-fitted estimate of the dipole amplitude, which is expressed as

$$\mathcal{A} = \frac{\delta\overline{I}_k}{F(\delta_k, \Delta\delta_k)\, \cos(\delta_k - \delta_{\rm d})}, \tag{4.54}$$

where k denotes the gradient-belt index, $\delta\overline{I}_k$ signifies the exposure–weighted average relative intensity of the belt, δ_k is its central declination, and $F(\delta_k, \Delta\delta_k)$ is a finite-belt geometric factor (accounting for the belt width and spherical weighting) that equals 1 in the infinitesimal-belt limit. Additionally, δ represents the declination of any point situated on the belts within the field of view. This setup is akin to the dipole function presented in Eq. (4.5). Therefore, the distribution of the best-fitted dipole on the celestial sphere can be described by Eq. (4.5), which captures the original cosmic-ray arrival directions with relative intensity across the entire celestial sphere, unlike the 1D projected dipoles typically reported by most experiments, which usually omit the dipole declinations. Now, let's explore the weighted relative intensity method.

Accordingly, the expression for the amplitude of a true dipole can be rewritten in a form that includes exposure weighting across the sky map, namely

$$\mathcal{A}_{\rm true} = \sum_{i=1}^{n_{{\rm pix},k}} \delta I_i \, \mathcal{W}_i \Bigg/ \left(\sum_{i=1}^{n_{{\rm pix},k}} \mathcal{W}_i \, \cos(\delta_k - \delta_{\rm d}) \right), \tag{4.55}$$

where $\delta I_i \equiv (N_i - N_{{\rm ref},i})/N_{{\rm ref},i}$ denotes the relative intensity in pixel i, and $\mathcal{W}_i = N_{{\rm ref},i}$ serves as the appropriate inverse-variance weight for Poisson-distributed counts. Here, N_i represents the observed number of events in a pixel i, while $N_{{\rm ref},i}$ gives the corresponding isotropic expectation (i.e., the reference or exposure map). The quantity δ_k denotes the central declination associated with the k-th belt. Furthermore, this approach compensates for uneven sky coverage, a common issue in observational cosmic-ray studies. In areas of the sky where the exposure is higher, corresponding to larger values of $N_{\rm ref}$, the method automatically accounts for the increased statistical precision by assigning proportionally larger weights. As a result, the weighted sum appearing in Eq. (4.55) adjusts itself in a manner that ensures all sky regions contribute in proportion to their statistical significance, thereby yielding a more stable and unbiased estimate of the dipole amplitude.

4.3.2 Optimized Gradient Method

The optimized gradient method further generalizes the approach described above so that it can be effectively applied to data sets in which the available statistics are limited. To minimize noise coupling, the event map is kept fixed, and only the coordinate belts are rotated during analysis, using small angular steps of 2°. This is coupled with the use of larger pixel scales for constructing three-dimensional sky maps, specifically choosing a small value of $n_{\rm side}$ (= 16 in this study), which defines the HEALPix grid resolution used to divide the sphere. These settings allow different belt orientations to sample comparable event populations within the same pixels across maps, thereby improving stability.

Focusing on a broad belt located in the median plane, we rotate only the belt coordinate frame (not the events). This process unfolds in several stages. First, belt orientations are adjusted by declination offsets δ_R in small δ bins, denoted as $\delta_{{\rm R},i}$. Each rotation yields a new coordinate distribution of the (fixed) arrival directions, visualized on 3D sky maps via

$$\delta_{{\rm CR},i} = \widetilde{\delta}_{{\rm CR},i} - \delta_{{\rm R},i}, \tag{4.56}$$

where $\widetilde{\delta}_{{\rm CR},i}$ is the declination of the event i in the rotated belt frame. Next, the maps are segmented into six vertical belts, facilitating a comparative analysis across different spatial sections of the sky. For each rendition of the map, the (exposure–weighted) average relative intensity $\delta\overline{I}_i$ is calculated across n belts, and the maximum and minimum averages are identified; the corresponding rotation angle defines $\theta_{\rm d}$, which correlates with the dipole declination via $\widetilde{\delta}_{\rm d} = \theta_{\rm d}$. We then refine the estimate by averaging the six values of $\widetilde{\delta}_{\rm d}$ obtained from all belts. Subsequently, the procedure is repeated with a coarser rotation step of 3° to check stability and consistency; following the same steps yields a second set of angles. The average of these angles, denoted by $\overline{\theta}_i$, determines the average dipole declination, $\overline{\delta}_{\rm d} = \overline{\theta}_i - \delta_{{\rm R},i}$. The corresponding algorithmic expression is

$$\delta\overline{\delta}_{\rm d} = \frac{1}{n_{\rm R} n_{\rm b}} \sum_{i=1}^{n_{\rm R}} \left[\sum_{j=1}^{n_{\rm b}} \theta_{ij} \left(\max\left(\left| \delta\overline{I}_{\rm min} \right|, \left| \delta\overline{I}_{\rm max} \right| \right) \right) - \delta_{{\rm R},i} \right] \mathbb{C}, \tag{4.57}$$

where $n_{\rm R}$ is the number of belt rotations, $n_{\rm b}$ is the number of belts, and $\mathbb{C}$ is a sign factor

fixing the north/south dipole hemisphere,

$$\mathbb{C} = \begin{cases} -1, & \text{if } \left|\delta\overline{I}_{\min}\right| > \left|\delta\overline{I}_{\max}\right|, \\ 1, & \text{if } \left|\delta\overline{I}_{\min}\right| \leq \left|\delta\overline{I}_{\max}\right|. \end{cases} \tag{4.58}$$

In the next step, we apply an analogous procedure to the right ascension coordinate. Keeping δ_{d} fixed, we rotate the belts along the R.A. axis and estimate the dipole phase from the locations of the extrema

$$\overline{\alpha}_{\mathrm{d}} = \frac{1}{n_{\mathrm{R}} n_{\mathrm{belt}}} \sum_{i=1}^{n_{\mathrm{R}}} \left[\sum_{j=1}^{n_{\mathrm{belt}}} \varphi_{ij} \left(\max\left(\left|\delta\overline{I}_{\min}\right|, \left|\delta\overline{I}_{\max}\right| \right) \right) - \alpha_{\mathrm{R},i} \right] \mathbb{D}, \tag{4.59}$$

with an east/west ambiguity captured by

$$\mathbb{D} = \begin{cases} 2, & \text{if } \left|\delta\overline{I}_{\min}\right| > \left|\delta\overline{I}_{\max}\right|, \\ 1, & \text{if } \left|\delta\overline{I}_{\min}\right| \leq \left|\delta\overline{I}_{\max}\right|. \end{cases} \tag{4.60}$$

Alternatively, the procedure can start without any initial global rotation of the 3D map. The celestial sphere is divided into six belts, each belt is rotated along R.A. in small angle bins, and for every rotation the (exposure–weighted) average relative intensity $\delta\overline{I}(\alpha)$ is computed. A first–harmonic fit, $\delta\overline{I}(\alpha) = A_1 \cos(\alpha - \alpha_{\mathrm{d}}) + B_1$, then yields the dipole phase α_{d}, and the dipole amplitude is obtained by fitting the weighted averages with Eq. (4.55). Partially overlapping slices matched to the pixel scale can be introduced to increase robustness, while keeping the event map fixed and applying only belt–frame rotations to avoid noise coupling.

This section presents the optimized gradient method as a proposed framework for reconstructing large-scale cosmic-ray anisotropies, particularly under conditions of limited statistics. By fixing the event map and rotating only the coordinate belts, this approach reduces statistical noise and improves the stability of dipole reconstruction across declinations. While the method provides a conceptually straightforward and computationally flexible way to extract dipole parameters from isoheight gradients, it remains primarily a proof of concept rather than a finalized analysis tool. In Chapter 6, we will develop and test a more precise analytical method, building on this idea to create a complete and quantitative framework for reconstructing anisotropy and evaluating sensitivity.

Chapter 5

Unified Cosmic-Ray Anisotropy

In many experiments, the reconstructed dipole amplitude $\mathcal{A}$ and phase $\alpha_{\rm d}$ are shifted because the field of view (FoV) for each individual experiment covers only a portion of the sky, which is different depending on the local experiment. In addition, the results are influenced by event samples (see Figure 5.2) based on other variables and by varying median energies for these samples. Therefore, the projected sidereal dipoles are not the original dipoles and lose some key information about the original distribution of cosmic-ray arrival directions, such as the declinations and amplitudes of the dipoles. The cosmic-ray anisotropies reported by experiments depend invariably on the observatory's location, its field of view, the efficiency, and the accuracy (both energy and direction) of reconstructing the primary cosmic-ray particles, as well as the corresponding event samples. This situation directly impedes the exploration of the cosmic-ray anisotropies and the studies investigating the direction of cosmic rays. That's the main reason discrepancies exist in the data on reconstructed dipoles from different experiments, even at the same or similar median energies. Therefore, we are motivated to extract the data reported by major experiments to reconstruct the "real" cosmic-ray dipoles as a function of energy. In this chapter, we will state that the unified real cosmic-ray dipole varies solely with energy, serving as a baseline in the interpretation of the CR anisotropies over a wide energy range.

We will present the amplitudes and phases of the projected dipoles, which are accurately reconstructed from the selected data after applying the partial-sky correction (discussed in Chapter 5) together with the calibrated fitting procedure. This includes detailed analyses of the event sample distributions, the overlap principle of cosmic-ray anisotropies, interpolation for cosmic-ray anisotropy distributions, the appropriate approach for one-dimensional projections of sky maps, and the algorithm to restore the reconstructions of the original cosmic-ray dipoles with their energies. In general, this work will not only recover the global evolution of cosmic-ray anisotropies but also provide some information on the origins and propagation of cosmic rays in our Galaxy.

5.1 Overlap of Anisotropies

We have developed a method to optimize and accurately reconstruct the large-scale anisotropies of cosmic rays, especially dipoles, at observatories with certain and different fields of view (FoV) and specific energy samples. This approach utilizes optimized algorithms to calibrate for the large discrepancies, which are from observations and partial sky coverages that change the original cosmic-ray dipole characteristics. By restoring the original amplitude and phase information, this method facilitates more precise comparisons between different experiments. It effectively gets rid of the impacts of geographical locations and observational constraints. Within this approach, we propose a new method for reconstructing the declination of dipoles by analyzing the characteristics of the dipole distribution and the corresponding fitting results. Furthermore, it could also allow for a refined analysis of the cosmic-ray spectrum, revealing subtle variations in dipole features

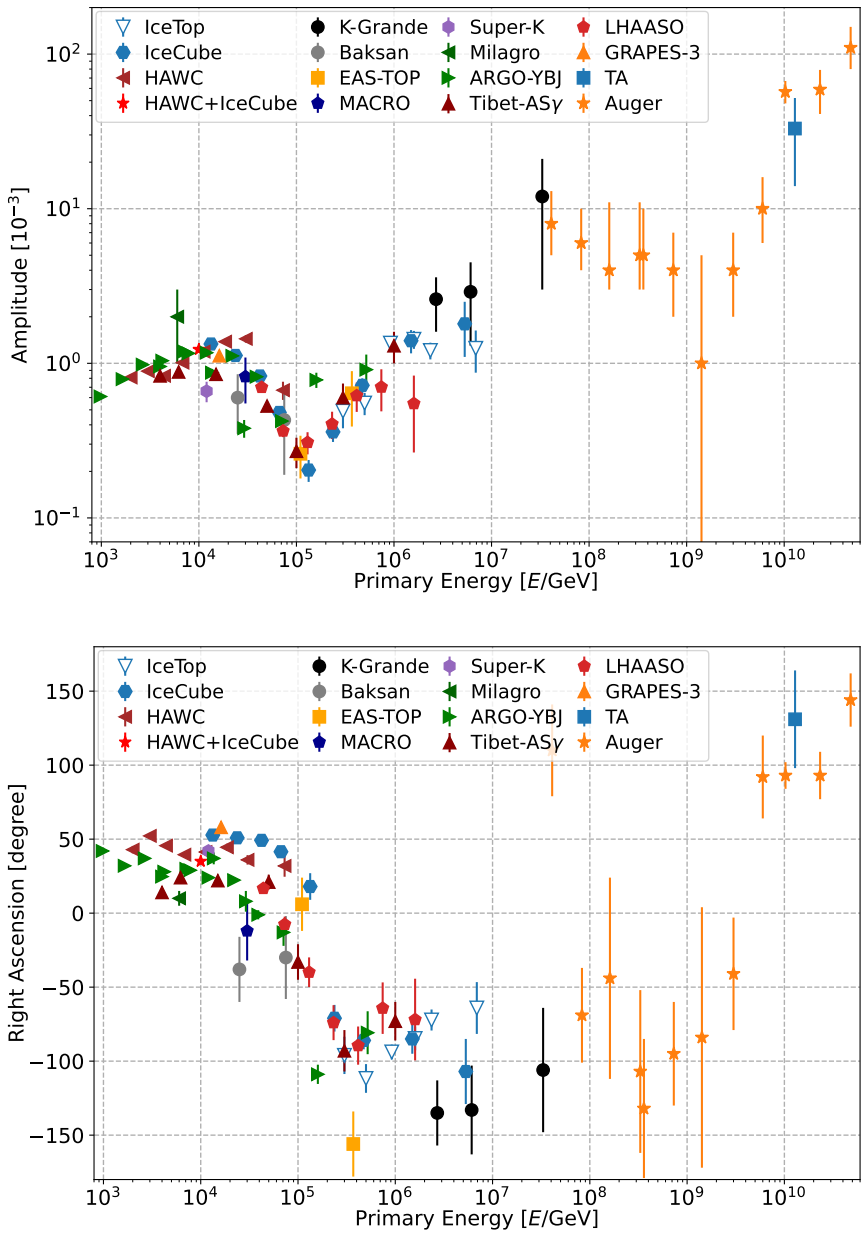


Figure 5.1: Reported dipole anisotropy. The data points shown are reconstructed sidereal dipole amplitudes (Top) and phases, i.e., right ascensions (Bottom), from various experiments [136, 268, 298, 299, 134, 133, 135, 269, 142, 140, 270, 300, 301]. Some of these results have been calibrated for partial-sky field-of-view (FoV) effects and thus report comparable sidereal dipole amplitudes (e.g., IceTop, IceCube, Auger, HAWC, and HAWC+IceCube), while the others have not and require further correction. The IceCube, IceTop, Auger, LHAASO, and GRAPES-3 data are updated dipole results as of 2025. Both plots reveal large discrepancies among different experiments in the energy-dependent reconstruction of dipoles, arising from various factors.

that were previously obscured.

The principle behind the observation of cosmic-ray anisotropies is based on accumulating large numbers of cosmic-ray events in their arrival-direction coordinates across different observation times, energy ranges, and sky regions. Consequently, these yield cosmic-ray anisotropies and spatial distributions on the celestial sphere. To restore the true dipole behind, let's first start from the relative intensity of the j-th pixel on this sky map, which is defined by $\delta I = (N_j - \langle N_j \rangle)/\langle N_j \rangle$ at a certain energy, where N_j represents the observed number of cosmic rays in the j-th pixel, and $\langle N_j \rangle$ denotes the average expected number of cosmic rays for that pixel. The quantity of δI therefore measures how much the observed intensity deviates from the expected average at a given energy. Concerning the fact that the dipole anisotropy of cosmic rays is known to vary with energy, and because published dipole measurements are frequently obtained from partially overlapping energy samples (as illustrated in Figure 5.2), we aim to understand the underlying principle governing the combination and interpretation of these continuously sampled dipole measurements.

Based on the assumption that a single, globally valid dipole exists as a function of energy, we partition the full sky map into n overlapping sub-maps, where each sub-map contains a subset of events whose distribution matches the event selection corresponding to a particular energy sample and its representative median energy. The total number of events across all n sub-maps is constructed to be equal to the total number of events in the full energy range, thereby guaranteeing complete statistical consistency. To clearly illustrate this concept, we first consider a simplified example in which the sky is divided into only two overlapping sub-maps. The shared region between these two maps, together with their respective relative intensities, denoted by δI_1 and δI_2 for each map, respectively, can then be expressed in the forms:

$$\delta I_1 = \frac{N_{1j} - \langle N_{1j} \rangle}{\langle N_{1j} \rangle}, \quad \delta I_2 = \frac{N_{2j} - \langle N_{2j} \rangle}{\langle N_{2j} \rangle}, \tag{5.1}$$

where the index j is used to specify the j-th pixel on the corresponding sky maps, thereby enabling a pixel-by-pixel comparison of the relative intensity fluctuations in the two data sets. The overlap of the two maps at pixel j, denoted by $\delta I_{2\mathrm{o}}$, can be expressed as follows:

$$\delta I_{2\mathrm{o}} = \frac{(N_{1j} + N_{2j}) - (\langle N_{1j} \rangle + \langle N_{2j} \rangle)}{\langle N_{1j} \rangle + \langle N_{2j} \rangle}, \tag{5.2}$$

which is obtained by summing the recorded counts $N_j = N_{1j} + N_{2j}$ for the combined observations and the corresponding expected average counts $\langle N_j \rangle = \langle N_{1j} \rangle + \langle N_{2j} \rangle$, representing an approximation to the mean number of events that would statistically fall into pixel j when the two sky maps are combined. Focusing on this expression, the mean number of events in the j-th pixel of each map can be estimated by summing the average counts of that pixel from both maps and multiplying by the ratio of the event counts in that pixel to the total counts of both maps, acting as a weighting or fractional factor, namely,

$$\begin{aligned} \langle N_{1j} \rangle &= \frac{N_{1j}}{N_{1j} + N_{2j}} \left(\langle N_{1j} \rangle + \langle N_{2j} \rangle \right), \\ \langle N_{2j} \rangle &= \frac{N_{2j}}{N_{1j} + N_{2j}} \left(\langle N_{1j} \rangle + \langle N_{2j} \rangle \right), \end{aligned} \tag{5.3}$$

where the expression $(\langle N_{1j} \rangle + \langle N_{2j} \rangle)$ can also be obtained for each individual sub-map. Utilizing Eq. 5.3, the relative intensity of the overlapped map given in Eq. (5.2) can be reformulated into a more explicit form, one that represents the weighted sum of the relative intensities associated with the j-th pixel in map 1 and map 2, where the weights

are determined by the respective event counts in that pixel. This leads to

$$\delta I_{2\mathrm{o}} = \frac{N_{1j}}{N_{1j}+N_{2j}}\delta I_1 + \frac{N_{2j}}{N_{1j}+N_{2j}}\delta I_2, \tag{5.4}$$

the appearance of δI_i results from substituting the term $(\langle N_{1j}\rangle + \langle N_{2j}\rangle)$ in Eq. (5.2). This indicates that the δI_{2o} is not merely the sum of the δI_j from the two sub-maps.

Similarly, let's explore a situation that is closer to a continuous distribution, moving beyond just two intervals to include extremely tiny energy intervals denoted by E_i (where i ranges from 1 to n) within the energy sample under analysis. Following the process of overlaying the j-th pixel from each of the n sub-maps to create composite event statistics, we can determine the relative intensity of this overlapped j-th pixel, as follows,

$$\delta I_{no} = \frac{(N_{1j} - \langle N_{1j}\rangle) + (N_{2j} - \langle N_{2j}\rangle) + ... + (N_{nj} - \langle N_{nj}\rangle)}{\langle N_{1j}\rangle + \langle N_{2j}\rangle + ... + \langle N_{nj}\rangle}, \tag{5.5}$$

where we apply the same assumption as discussed in Eq. (5.3), and considering n counts with their corresponding average counts across the n energy intervals, donates as:

$$\langle N_{ij}\rangle = \frac{N_{ij}\left(\langle N_{1j}\rangle + \langle N_{2j}\rangle + ... + \langle N_{nj}\rangle\right)}{N_{1j} + N_{2j} + ... + N_{nj}}. \tag{5.6}$$

Subsequently, we further rewrite Eq. (5.5) by replacing the sum of $\langle N_{ij}\rangle$ over i from 1 to n, denoted by $\sum_{i=1}^{n}\langle N_{ij}\rangle$ from Eq. (5.6), and resulting in:

$$\delta I_{no} = \frac{N_{1j}\left(N_{1j} - \langle N_{1j}\rangle\right)}{\langle N_{1j}\rangle \sum\limits_{i=1}^{n} N_{1j}} + \frac{N_{2j}\left(N_{2j} - \langle N_{2j}\rangle\right)}{\langle N_{2j}\rangle \sum\limits_{i=1}^{n} N_{2j}} + ... + \frac{N_{nj}\left(N_{nj} - \langle N_{nj}\rangle\right)}{\langle N_{nj}\rangle \sum\limits_{i=1}^{n} N_{nj}}, \tag{5.7}$$

where elements denoted by δI_{ij}, defined as $\delta I_{ij} = \left(N_{ij} - \langle N_{ij}\rangle\right) / \langle N_{ij}\rangle$ exist in the sum. Therefore, we derive the expression of relative intensity (δI_{no}) of the sky map, resulting from the overlapping of δI_{ij} across n sub-maps, specifically:

$$\delta I_{no} = \frac{\sum\limits_{i=1}^{n} N_{ij} - \sum\limits_{i=1}^{n}\langle N_{ij}\rangle}{\sum\limits_{i=1}^{n}\langle N_{ij}\rangle} = \frac{\sum\limits_{i=1}^{n} N_{ij}\delta I_{ij}}{\sum\limits_{i=1}^{n} N_{ij}}, \tag{5.8}$$

which clearly illustrates the impact of varying event counts on statistical observations. This demonstrates that the overlapped δI_j in a specific pixel is not merely the sum of the δI_{ij} values. Instead, the fractional differences in event counts across different segments of a distinct energy sample directly influence the observed patterns of cosmic-ray anisotropies.

For example, if two dipole maps overlap, with certain orientations $(\alpha_{\mathrm{d}}, 0)$ on the equatorial plane, amplitude ($\mathcal{A}$), and total number of events (N), and the dipole distribution can be expressed as $\delta I = \mathcal{A}\cos\left(\alpha - \alpha_{\mathrm{d}}\right)$, we can analyze their combined contribution. Specifically, assume that dipole 1 has orientation at $(0, 0)$, with $\mathcal{A}_1 = 10^{-2}$ and $N_1 = 10^8$, while dipole 2 has orientation at $(\pi/2, 0)$ with $\mathcal{A}_2 = 10^{-3}$ and $N_2 = 2\times10^8$; their overlap leads to $\delta I_{\mathrm{o}} = 0.0034\cos\left(\alpha - \pi/16\right)$. Now, in the continuous case, we represent a specific location within the sample by the energy (E), with an infinitesimally small interval dE, and the integral form of δI_{no} in Eq. (5.8) for a specific energy sample can be expressed as:

$$\delta I_{\mathrm{m}} = \frac{\int \mathcal{N}(E)\delta I(E)dE}{\int \mathcal{N}(E)dE}, \tag{5.9}$$

where δI_{m} represents the relative intensity of the entire energy sample, $\mathcal{N}(E)$ donates the probability density function (PDF) of the distribution of events in the energy sample, as a

function of energy, and, in principle, the integral (accumulation function) of $\mathcal{N}(E)$ equals 1 over the sample's entire energy range.

Up to now, we have understood the principle behind the formation of the observed dipole anisotropy, as shown in Eq. (5.9). This approach enables us to evaluate the variation in the relative intensity of cosmic-ray arrival directions by integrating data across different and continuous energies, and subsequently, to restore the energy-varying dipole anisotropy.

5.2 The Unified Anisotropy

The reported sidereal dipoles by the experiments are invariably multi-overlay results, which are overlapped concerning the dipoles with continuously varying amplitudes and orientations at continuously varying energies. The overlap of the dipole with different amplitudes at different energies and different orientations, with different event counts, will lead to varying observations, as we discussed before. Therefore, the reported data of these main experiments are usually not comparable. We now aim to reconstruct the full process of a global, continuous and energy-dependent cosmic-ray dipole as it would be detected and reconstructed through projection procedures by any observatory on Earth.

The reconstruction of dipole anisotropies involves projecting the 3D distributions of relative intensities, denoted by $\delta\hat{\mathcal{I}}$, which are, in principle, integrations over the experiments' field of view (FoV), specifically the declination range, and the specific energy range of the energy samples. Based on the discussions in Section 5.1, $\delta\hat{\mathcal{I}}$ is given by:

$$\delta\hat{\mathcal{I}}(E,\alpha) = \frac{1}{\mathcal{N}_0\mathcal{W}} \int_{E_\mathrm{L}}^{E_\mathrm{U}} \int_{\delta_\mathrm{L}}^{\delta_\mathrm{U}} \mathcal{N}(E)\mathcal{A}(E)\mathcal{D}(E,\alpha,\delta)w(\delta)\,\mathrm{dE}\,\mathrm{d}\delta, \tag{5.10}$$

where $\mathcal{N}_0 = \int_{E_\mathrm{L}}^{E_\mathrm{U}} \mathcal{N}(E)\,\mathrm{dE}$, and $\mathcal{N}_0$ is equal to 1 normally, meaning, the full accumulation of the probability density function of any energy sample which obtained by Monte-Carlo simulation is approximately normalized, both the dipole's right ascension and declination are functions of energy in the projected unit dipole $\mathcal{D}$, $w(\delta) = \cos\delta$ in this study, is the declination weight for the spherical surface area, $\mathcal{W}$ is the sum of the weight over the effective field of view. The integration over δ means the projection of $\mathcal{I}(E,\alpha,\delta)$ onto RA plane, becomes $\hat{\mathcal{I}}(E,\alpha)$. After integrating with respect to declination (δ), we can obtain

$$\delta\hat{\mathcal{I}}(\alpha) = \int_{E_\mathrm{L}}^{E_\mathrm{U}} \mathcal{N}(E)\mathcal{A}(E)\hat{\mathcal{D}}(E,\alpha)\,\mathrm{dE}, \tag{5.11}$$

where energy E changes to be median energy after the integration, denoted by E_m, $\hat{\mathcal{D}}(\alpha, E_\mathrm{m})$, which is the projected sidereal dipole. Since we don't have information about the δ_d in data, therefore we can only consider a joint expression, namely, $\mathcal{A}c_p\cos\delta_\mathrm{d} = \mathcal{A}_p$. $\mathcal{A}\cos\delta_\mathrm{d} = \mathcal{A}_\mathrm{p}$ represents the α-component of the original dipole amplitude $\mathcal{A}$. When the FoV is maximum, namely, $\delta_\mathrm{L} = -\pi$ and $\delta_\mathrm{U} = \pi$, c_p reaches the maximum value, denoted by $c_\mathrm{pm} = \pi/4$. Therefore, we can also plot the curve with the maximum projection of α-component of $\mathcal{A}$ by $A_\mathrm{p}c_\mathrm{pm}/c_\mathrm{p}$. The corresponding reconstructed Relative Intensity (RI) from the experiments can be expressed as follows:

$$\delta I(E_\mathrm{m},\alpha) = \mathcal{A}_p(E_\mathrm{m})\cos\big[\alpha - \alpha_\mathrm{d}(E_\mathrm{m})\big] + \mathcal{B}, \tag{5.12}$$

where δI represents the fitted relative intensity that has already been determined, while E_m, $\mathcal{A}_\mathrm{p}$, and αd are taken from the reported data of various experiments. The term $\mathcal{B}$ can be safely neglected in this study, since there are currently no available published data regarding the shift term in the harmonic fits, similar to the situation in Eq. (4.7).

5.2.1 Energy Distributions of Data

The data samples from various experiments consistently exhibit a log-normal distribution when the categorization of primary cosmic-ray particles is based on the number of hits recorded by detectors, designated as N_{channel}. This categorization can refer, for example, to specific channels like the muon channel in the case of the IceCube experiment [139] and LHAASO [137], or the aggregated detected particle density, $\Sigma\rho$, as observed in experiments like Tibet [310]. As contained in Eq. (5.10), the number density distributions of cosmic-ray events at each median energy are approximately expressed as

$$\mathcal{N}(E) = \frac{\mathcal{N}_0}{E\sigma\sqrt{2\pi}} e^{-\frac{(\ln E-\mu)^2}{2\sigma^2}}, \tag{5.13}$$

where $\mathcal{N}_0$ is a constant, σ is the standard deviation of the log-normal distribution, and μ is its mean, given by the natural logarithm of the median energy E_{m}, i.e., $\ln E_{\text{m}}$. For LHAASO [137], as shown in Figure 5.2, and Auger [311, 312], the energy resolutions are modelled as Gaussian distributions in logarithmic energy space, making them special cases of the log-normal distribution and thus approximately consistent with Eq. (5.13). The corresponding 68% containment interval of the event energies is bounded by $e^{\mu-\sigma}$ and $e^{\mu+\sigma}$, respectively. Therefore, the standard deviation σ can be obtained from the lower and upper bounds as $\overline{\sigma} = \ln(E_{\text{U}}/E_{\text{L}})/2$. The total number of events in an event sample is obtained by integrating the number density concerning primary energy within the whole energy range from E_{L} to E_{U}, namely, $\int_E \mathcal{N}(E)dE = \mathcal{N}_0$, and $\mathcal{N}_0$ represents the total event count for the given energy sample. By normalizing $\mathcal{N}_0$ to 1, Eq. (5.13) becomes a probability density function (PDF). This normalization allows dipole results from different energy samples and experiments to be directly compared in this study.

In the specific case of KASCADE-Grande, the reconstructed energy distribution associated with the median energy of 2.7 PeV follows a log-normal distribution within the range $(6.22, 6.60)$ in $\lg(E/\text{GeV})$. For higher energies, such as 6.1 PeV and 33 PeV, corresponding to the intervals $(6.60, 7.41)$ and $(7.41, 8.84)$ respectively in $\lg(E/\text{GeV})$, the distributions are well described by an exponential tail function, namely,

$$\mathcal{N}(E) = \mathcal{N}_0 e^{-k(x_0-x)}, \tag{5.14}$$

where x_0 denotes the starting point of the exponential tail, and $\mathcal{N}_0$ is a normalization constant that can be adjusted to yield a proper probability density function (PDF) within the specified energy range. Further details on the corresponding energy distributions, fitting parameters, and the overall fitting procedure are discussed in Section 6.1.

5.2.2 Analytical Reconstruction Method

Now, we construct the fitting functions for the α-component of the original amplitude, denoted by $\mathcal{A}^*$, and the right ascension of the original dipole. These functions, which depend solely on energy, are expressed as suitable combinations of polynomials and Gaussian functions to capture both smooth trends and localized variations, specifically,

$$\mathcal{A}^*(E) = \sum_{i=1}^{n} \lambda_i E^{\beta_i} e^{\gamma_i E}, \quad \alpha_{\text{d}}(E) = \sum_{j=1}^{n} \epsilon_j E^{\rho_j} e^{\eta_j E}. \tag{5.15}$$

According to the data shown in Fig. 5.1, the number of polynomial terms in all three functions is chosen as $n = 4$. The parameters are set as follows: $\beta_1 = \gamma_1 = \rho_1 = \eta_1 = 0$, $\beta_2 = \rho_2 = 1$, and $\gamma_2 = \eta_2 = 0$. In most experimental dipole reconstructions, the projected

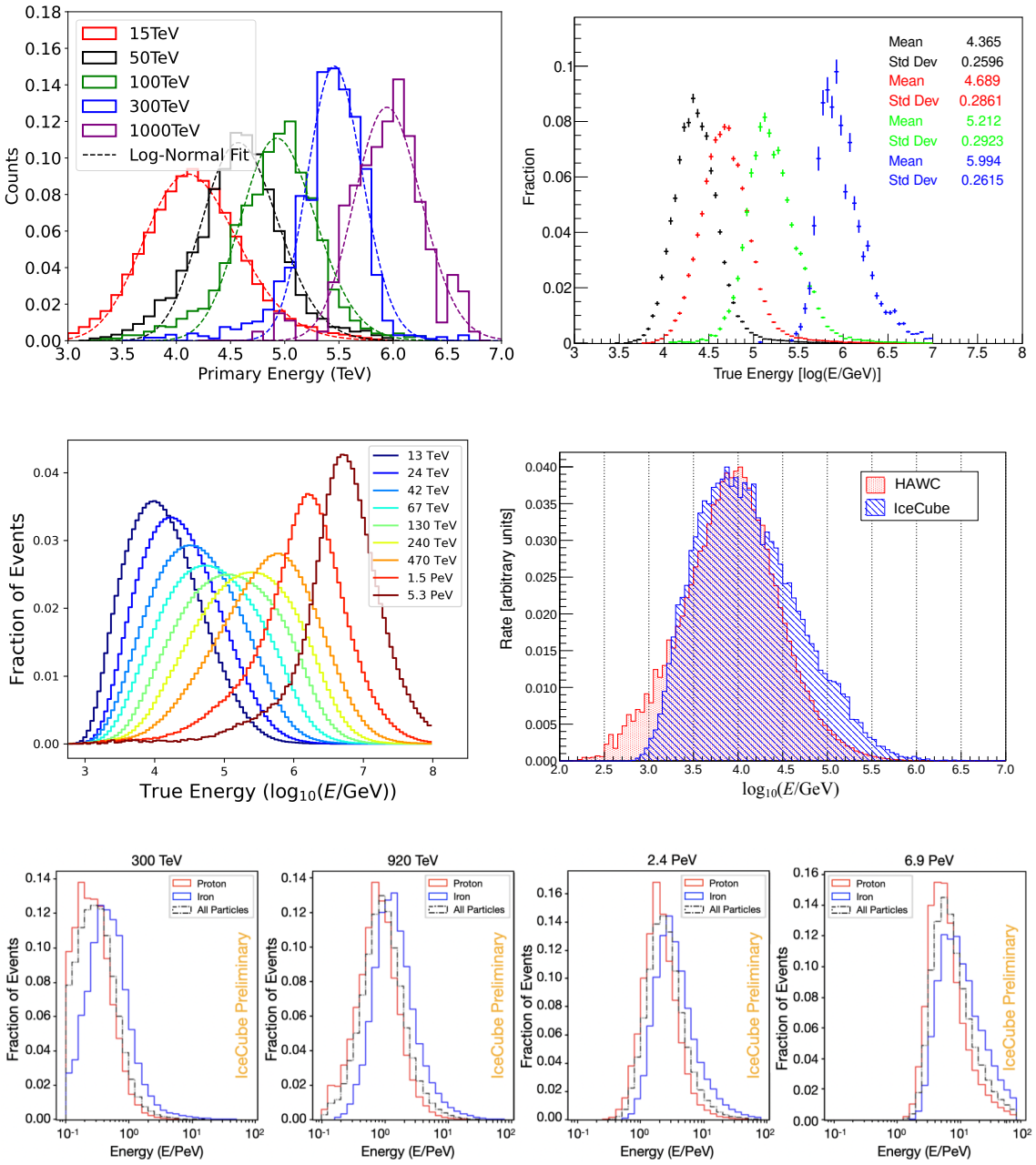


Figure 5.2: Illustration of simulated true energy distributions (i.e., median energy) for the assumed events. Shown are five examples of energy distributions for Tibet [310], LHAASO [300], IceCube [140], IceCube+HAWC [16], and IceTop [142] used in anisotropy studies. The IceTop simulations assume iron (blue), proton (red), and all-component (dashed) compositions. Other experiments report their energy distributions mainly in tables, without providing plotted representations. All energy distributions correspond to all-particle events. At a fixed energy, different experiments exhibit significant differences in their reconstructed energy distributions.

sidereal dipole includes only the right ascension term, and the fitting function is

$$\delta I(\alpha) = \mathcal{A}_{\rm p} \cos(\alpha - \alpha_{\rm d}) + \mathcal{B}, \tag{5.16}$$

where $\mathcal{A}_p$ denotes the projected dipole amplitude. Due to the dipole reconstructions performed by the experiments, which were based on harmonic fits that did not include the shift term in Eq. (5.15), we set $\mathcal{B} = 0$. The corresponding theoretical expression for the projected dipole, denoted by $\delta\hat{\mathcal{I}}(\alpha)$ for each energy sample, is detailed in Eq. (5.11). To assess the goodness of fit of the unified anisotropy model relative to a reference function with known parameters, we apply the least-squares method to minimize the difference between the observed and predicted values. Referring to Eq. (5.15), we fit the initial guesses of the fit functions in Eq. (5.12) to the standard first harmonic. This involves comparing the relative intensity from the data, given by Eq. (5.16), to our model, given by Eq. (5.11), across all data points from the selected experiments, expressed as

$$\delta I_i(\alpha) = \mathcal{A}_{{\rm p},i} \cos\left(\alpha - \varphi_i\right), \quad \delta\hat{\mathcal{I}}_i(\alpha) = \int_{E^i_{\rm L}}^{E^i_{\rm U}} \mathcal{N}(E)\mathcal{A}^*(E) \cos\left(\alpha - \alpha_{\rm d}(E)\right) dE, \tag{5.17}$$

at a fixed energy, where i is the index of the data points, $\mathcal{A}_{{\rm p},i}$ and φ_i represents the dipole amplitudes and phases from the data. Now, let $\mathcal{P}$ denote the vector in the least squares expression used in the minimisation process, namely,

$$\boldsymbol{\mathcal{P}} = \min\left(\sum_{i=1}^{N_{\rm D}} \frac{\left(\delta I_i - \delta\hat{\mathcal{I}}_i\right)^2}{\hat{\sigma}_i^2}\right), \tag{5.18}$$

where $\hat{\sigma}_i$ represents the universal uncertainty, which is given by integrating the error of δI_i, namely, $\Delta\delta I_i$, over the whole right ascension plane from $-\pi$ to π and weighted by 2π, taking into account both uncertainties of $\mathcal{A}_p$ and $\alpha_{\rm d}$ (represented by φ_i here) from the reported data. Consequently, the universal uncertainty can be written in the form:

$$\hat{\sigma}_i^2 = \int_{-\pi}^{\pi} \left(\sigma^2_{\mathcal{A},i} \cos^2(\alpha - \varphi_i) + \mathcal{A}^2_{{\rm p},i}\sigma^2_{\varphi,i} \sin^2(\alpha - \varphi_i)\right) d\alpha. \tag{5.19}$$

The difference between observation and the model is defined as $\Delta\delta I_i = \delta I_i - \delta\hat{\mathcal{I}}_i$, which should use the integral form since both of them are multi-curves (harmonics), not points, in principle. However, the positive and negative regions would be eliminated. Therefore, we seek the discrete calculation, which is expressed as:

$$\Delta\delta I_i(\alpha_k) = \delta I_i(\alpha_k) - \delta\hat{\mathcal{I}}_i(\alpha_k) \tag{5.20}$$

where the index $k = 1, 2, 3, \ldots, \mathfrak{n}$, and α_k spans the full interval from $-\pi$ to π. The modified least-squares minimization function is now written as:

$$\boldsymbol{\mathcal{P}} = \min\left(\sum_{i=1}^{N_{\rm D}} \sum_{k=1}^{\mathfrak{n}} \frac{\left(\delta I_i(\alpha_k) - \delta\hat{\mathcal{I}}_i(\alpha_k)\right)^2}{\hat{\sigma}_i^2(\alpha_k)}\right), \tag{5.21}$$

$\hat{\sigma}_i^2(\alpha_k)$ denotes the discrete value of the universal uncertainty $\hat{\sigma}_i$ at α_k. Using this data, one can determine the best-fitted parameters for the assumed functions of the dipole amplitudes and their corresponding orientations. Additionally, the parameters specified in Eq. (5.15) can be derived. Ultimately, we generate the evolution curves for the original dipole's amplitude, right ascension, and declination, which vary with energy. This represents the first method to reconstruct the dipole anisotropy of cosmic rays.

Now, let's consider the dipole declinations, δ_d, which in theory can be derived utilising

Equations (5.16) and (4.11). However, the data reported by experiments like HAWC and IceCube often omit details of the shift term in the first harmonic fit. Although these reports include 2D fits for the maximum sidereal dipoles, they lack information on declinations. This is because $\delta\mathcal{I} = \mathcal{A}^* \hat{c}_p \cos(\alpha - \varphi)$, where $\mathcal{A}^* = \mathcal{A}\cos\delta_d$ is considered the α-component of the original dipole $\mathcal{A}$. Additionally, while $\mathcal{A}_p = \mathcal{A}c_p \cos\delta_d$ and $\mathcal{A}_p = \mathcal{A}^* c_p$, are reported for 2D and 1D fits respectively, the actual declination δ_d remains unreported. The reconstructed Relative Intensity (RI) for a specific energy sample is then expressed as an integration over the energy range of that sample, specifically,

$$\delta\hat{\mathcal{I}}(\alpha, E_{\rm m}) = \int_{E_{\rm L}}^{E_{\rm U}} \mathcal{N}(E; E_{\rm m}, \sigma)\mathcal{A}_{\rm p}^*(E)\cos(\alpha - \alpha_{\rm d}(E))dE, \tag{5.22}$$

and $E_{\rm m}$ is the median energy of the energy sample.

Since the fit of dipole anisotropy is based on the existing data of sidereal dipole amplitudes $\mathcal{A}_{\rm p}$ and phases $\alpha_{\rm d}$, and is sensitive to both parameters, even small changes in either can introduce significant bias in the resulting fit, namely, the inferred dipole amplitudes and phases. A more reliable approach is to fit the $\mathcal{A}_{\rm p}$ and $\alpha_{\rm d}$ data to the most likely curves first, using a maximum-likelihood method, before extracting the dipole parameters with our technique. This curve fitting assumes that the dipole varies smoothly in both amplitude and orientation (in right ascension only, as declination is not included in this analysis) with changing energy. To determine smooth energy-dependent trends of the cosmic-ray dipole anisotropy, we apply a maximum-likelihood fitting procedure to the projected dipole parameters, the projected dipole amplitude $\mathcal{A}_{\rm p}$ and the phase $\alpha_{\rm d}$, which are reconstructed at different energy bins. These two parameters characterise the relative intensity modulation in right ascension α, given by the sidereal dipole form,

$$\delta I(\alpha) = \mathcal{A}_{\rm p}\cos(\alpha - \alpha_{\rm d}), \tag{5.23}$$

where α represents the right ascension plane, ranging from $-\pi$ to π. This expression for $\delta I(\alpha)$ has also been discussed earlier. The projected amplitude $\mathcal{A}_{\rm p}$ is the projection of a full dipole, with the entire celestial sphere as the FoV. The observed projected amplitude $\mathcal{A}_{\rm p}^\dagger = {\rm c_p}\mathcal{A}_{\rm p}$ involves coefficients that are determined by the observatory's FoV and depend on its upper and lower limits, as shown in Eq. (4.9). Thus, $\mathcal{A}_{\rm p}$ is calculated as

$$\mathcal{A}_{\rm p} = \frac{\mathcal{A}_{\rm p}^\dagger}{c_{\rm p}}, \quad c_p = \frac{2\delta_{\rm U} - 2\delta_{\rm L} + \sin 2\delta_{\rm U} - \sin 2\delta_{\rm L}}{4(\sin\delta_{\rm U} - \sin\delta_{\rm L})}, \tag{5.24}$$

where $c_{\rm p}$ is the calibration factor for the FoV. The amplitude $\mathcal{A}_{\rm p} = \mathcal{A}\cos\delta_{\rm d}$ represents the horizontal component of the dipole amplitude $\mathcal{A}$, projected by the factor $\cos\delta_{\rm d}$.

The measured values of $\mathcal{A}_p$ and α_d, together with their associated uncertainties $\sigma_{\mathcal{A}_p}$ and σ_{α_d}, are obtained in several energy bins. Because the measured amplitude and phase are derived from the same underlying function and are not statistically independent, we do not fit them separately. Instead, we construct a log-likelihood based on the full dipole modulation $\delta I(\alpha)$, sampled across a set of right ascension values α_k. Assuming that the dipole amplitude and phase evolve smoothly with energy, they can be modelled as polynomials times exponentials as functions of E, in logarithmic scale,

$$\mathcal{A}_{\rm p}^{\rm M}(E) = \sum_{i=1}^{n} a_i E^i \exp(\lambda_i E), \quad \alpha_{\rm d}^{\rm M}(E) = \sum_{i=1}^{n} b_i E^i \exp(\gamma_i E), \tag{5.25}$$

where $i = 0, 1, 2$, and the coefficients a_i, b_i, together with the exponential parameters λ_i and γ_i constitute the full set of free parameters to be determined through the global fit.

The resulting model prediction for the phase is subsequently wrapped into the interval $[-\pi, \pi]$ to get a valid phase angle. The observed dipole function for each energy bin E_i and right ascension sample α_k, and the corresponding model prediction, are given by

$$\delta I_{ik}^{\mathrm{D}} = \mathcal{A}_p^{(i)} \cos\left(\alpha_k - \alpha_d^{(i)}\right), \quad \delta I_{ik}^{\mathrm{M}} = \mathcal{A}_{\mathrm{p}}^{\mathrm{M}}(E_i) \cos\left(\alpha_k - \alpha_{\mathrm{d}}^{\mathrm{M}}(E_i)\right), \tag{5.26}$$

where α is discretized into a set of uniformly spaced bins. The uncertainty associated with each relative-intensity value, denoted by σ_{ik}, is obtained by propagating the measurement uncertainties of the dipole amplitude $\mathcal{A}_p$ and α_d according to

$$\sigma_{ik}^2 = \left[\cos\left(\alpha_k - \alpha_d^{(i)}\right) \cdot \sigma_{\mathcal{A}_p}^{(i)}\right]^2 + \left[\mathcal{A}_p^{(i)} \cdot \sin\left(\alpha_k - \alpha_d^{(i)}\right) \cdot \sigma_{\alpha_d}^{(i)}\right]^2. \tag{5.27}$$

We assume Gaussian uncertainties on $\mathcal{A}_p$ and α_d, the likelihood for a measurement is:

$$\mathcal{L}_{ik}\left(\delta I_{ik}^{\mathrm{D}} \,|\, \delta I_{ik}^{\mathrm{M}},\, \sigma_{ik}\right) = \frac{1}{\sqrt{2\pi}\,\sigma_{ik}} \exp\left(-\frac{(\delta I_{ik}^{\mathrm{D}} - \delta I_{ik}^{\mathrm{M}})^2}{2\sigma_{ik}^2}\right), \tag{5.28}$$

and the corresponding total log-likelihood is given by:

$$\ln \mathcal{L} = -\frac{1}{2} \sum_{i,k} \left[\ln(2\pi\,\sigma_{ik}^2) + \frac{\left(\delta I_{ik}^{\mathrm{D}} - \delta I_{ik}^{\mathrm{M}}\right)^2}{\sigma_{ik}^2}\right], \tag{5.29}$$

where only the second term (residual squared over variance) is relevant for optimisation, as the logarithmic prefactors are independent of the model parameters, thus, the log-likelihood maximisation reduces to minimising the weighted sum of squared residuals, which is a generalized χ^2 function.

However, after applying the necessary smoothing procedure to the measured dipole data, it becomes evident that estimating the best-fit parameters directly from Eq. (5.22) does not provide a sufficiently accurate or stable result. This limitation arises because the initial guesses for the model parameters $\mathcal{A}_{\mathrm{p}}^{\mathrm{M}}(E)$ and $\alpha_{\mathrm{d}}^{\mathrm{M}}(E)$ require fine-tuning and perfectly reconstructed dipoles, which do not reflect the reality of experimental results, where larger discrepancies appear at nearby energies. The sensitivity of the fitting procedure is further constrained by the relatively small number of available measurements, namely the sets of observed amplitudes $\mathcal{A}_{\mathrm{p}}^{\mathrm{D}}$ and $\alpha_{\mathrm{d}}^{\mathrm{D}}$. Consequently, the resulting estimates of $\mathcal{A}_{\mathrm{p}}^{\mathrm{M}}(E)$ and $\alpha_{\mathrm{d}}^{\mathrm{M}}(E)$ may deviate noticeably from the true, smoothly varying dipole behaviour that the model aims to describe. Therefore, a more reliable method is required.

5.2.3 Grid Iteration Method

To obtain a more accurate reconstruction of the dipole anisotropy and to extract the underlying dipole signal from the superposition of contributions at different energies within the effective energy range of a given energy bin, we perform an iterative procedure on each point, corresponding to a small energy interval within the data's effective energy range (energy samples, see Figure 5.2), where the measurements refer to the projected sidereal dipole amplitude A_{p} and phase Φ_{p}. Specifically, each energy point corresponds to a weighted average over dipole vectors, with weights determined by the discrete energy distribution of the sample. These dipoles overlap and collectively contribute to a mean dipole, which, in principle, corresponds to the measured dipole signal $\delta I = A_{\mathrm{p}} \cos(\alpha - \Phi_{\mathrm{p}})$ associated with the energy range centred at the median energy E_{m}. This forms the basic idea for extracting the energy-dependent dipole from data corresponding to a single energy sample. Meanwhile, this procedure can be applied to all $(A_{\mathrm{p}}, \Phi_{\mathrm{p}})$ data pairs that include a given energy point, ultimately yielding the best estimate of the dipole at that energy.

To represent the measured dipole anisotropy at an energy E (whatever median energy or energy), we express each dipole distribution (relative intensity) as a complex form,

$$\delta I_k = \frac{A_k}{2}\left(e^{i(\alpha-\phi_k)} + e^{-i(\alpha-\phi_k)}\right), \tag{5.30}$$

which representation is mathematically equivalent to the traditional cosine formulation, and captures only the real part of the complex exponential. Simply, we can express the dipole distribution as $\delta I = A_k e^{i(\alpha-\phi_k)}$ allowing us to manipulate the complete signal directly in the complex plane, while the measured quantity corresponds to its real projection, $\Re[\delta I] = A\cos(\alpha-\phi)$. The introduction of this complex representation naturally leads to the definition of a complex dipole vector $D_k = A_k e^{i\phi_k}$, which compactly includes both the amplitude A_k and phase ϕ_k as a linear system. The magnitude $|D_k|$ of this vector straightforwardly yields the dipole amplitude, whereas the complex argument $\arg(D_k)$ corresponds to the phase. This representation eliminates the need to split the expression into positive and negative frequency components, as required in the cosine form. The measured dipole vectors at a given energy sample, with median energy $E_{\rm m}$, are characterized by their amplitude A_i and phase Φ_i. We then express each in complex form, i.e.,

$$D_i(E_{\rm m}) = A_i e^{i\Phi_i}, \tag{5.31}$$

here, the index i (with i=1, 2,..., M) labels the dipole measurements from all experiments, each corresponding to a median energy $E_{\rm m}$ with an 68% energy range (energy resolution) bounded by $E_{\rm L}$ and $E_{\rm U}$, and energies lying within the overall range $[E_{\rm min}, E_{\rm max}]$. The range covered by the data with energy is approximated continuously. Rather than dividing the energy range of each sample individually, the global energy range is discretized into small intervals. using a grid of equally spaced bins E_j, and $j = 1, 2, ..., N_g$ referred to as grid points, at which the underlying dipole signal is estimated. In this analysis, we use $N_{\rm g} = 200$ grid points spanning from 2.9 to 10.8 in $\lg(E/{\rm GeV})$, corresponding to energies from 0.8 TeV to 60 EeV. Each grid point corresponds to a dipole characterized by the sidereal amplitude and phase, denoted by $\mathcal{D}_j = \mathcal{A}_j e^{i\varphi_j}$, at a specific energy $E = E_j$. When the grid spacing is sufficiently small, this provides a good approximation of the continuous energy dependence of the dipole.

We now propose a model by assuming that each measured dipole D_i can be described as a sum (or an integral, in the continuous case) of contributions from the dipoles $\mathcal{D}_j$ over the energy range corresponding to D_i. The sum of all dipoles ($\sum_j \omega_{ij}\mathcal{D}_j$) at all grid points within $[E_{\rm min}, E_{\rm max}]$ of the i-th energy sample yields a model estimated instantaneous pseudo-dipole $\mathbb{D}_i$, which corresponds to the measured dipole at the median energy $E^i_{\rm m}$. Specifically, the pseudo-dipole vector $\mathbb{D}_i$ is given by

$$\mathbb{D}_i(E_{\rm m}) = \sum_{j=1}^{N_g} \omega_{ij}\mathcal{A}_j e^{i\varphi_j} = \mathbb{A}_i e^{i\Psi_i}, \tag{5.32}$$

where the sum runs over all grid points with ω_{ij}, noting that $\omega_{ij} = 0$ if $E^j_{\rm grid} \notin \left[E^i_{\rm min}, E^i_{\rm max}\right]$. The value of $\mathbb{D}_i$ approximates the measured dipole D_i, but they are not necessarily equal. The weight ω_{ij} at the grid point j for the i-th measured dipole is defined by

$$\omega_{ij} = \frac{\mathcal{N}(E_j; E^i_{\rm L}, E^i_{\rm U}, E^i_{\rm m})\Delta E}{\sum_{k=1}^{N_g}\mathcal{N}(E_j; E^i_{\rm L}, E^i_{\rm U}, E^i_{\rm m})\Delta E}, \tag{5.33}$$

where ΔE_j represents the width of the energy grid point centred at E_j, $\mathcal{N}$ is the probability density function (PDF) describing the energy distribution within the 68% containment

range. It is modelled as either a log-normal or exponential-tail distribution, depending on the experiment, and is defined in terms of $E^i_{\rm L}$, $E^i_{\rm U}$, and $E^i_{\rm m}$, as given by by Eq. (5.13) and Eq.(5.14). For each dipole i, the weight ω_{ij} is normalized so that $\sum_j w_{ij} = 1$ since it's based on the discrete form of the PDF. To extract the dipoles, an inversion process, which is referred to as deconvolution, is performed, based on the exponential linear model given in Eq. (5.32), or, in an equivalent matrix form,

$$\mathbb{D} = \mathbf{W}\boldsymbol{\mathcal{D}}, \tag{5.34}$$

where $\mathbb{D}$ is a $M \times 1$ vector of the observed dipoles, where $\boldsymbol{\mathcal{D}}$ is a $N \times 1$ vector of unknown global dipoles at energy grid points E_j (with $j = 1, \ldots, N$) (which include both amplitude and phase, and can therefore be regarded as a complex-valued vector), and $\mathbf{W}$ is a $M \times N$ weight matrix that contains all rows of ω_{ij}. In the continuous form, the measured dipole is expressed as a weighted integral over energy, representing the contributions of the dipole across the energy range, $\mathbb{D}_i = \int_E \mathcal{N}\mathcal{D}\mathrm{d}E$, which makes the estimation of the dipole more challenging due to the continuous and energy-dependent fact of the integrated. The integration is performed over the full energy range $[E^i_{\rm min}, E^i_{\rm max}]$ corresponding to the i-th measurement. If multiple pseudo-dipole matrices $\mathbb{D}$ cover the same energy range, they independently constrain the value of the dipole $\mathcal{D}_j$ at the corresponding grid point. For example, we suppose the energy range of measurement 1 overlaps the grid points E_2, E_3, and E_4, while the range overlaps E_3, E_4, and E_5 of measurement 2. These are independent data points, but because their energy ranges overlap (not double-counting), both provide information about the dipoles $\mathcal{D}(E_3)$ and $\mathcal{D}(E_4)$ through their overlapping contributions. In the deconvolution process, the contributions to each energy grid point from all measurements are combined according to their respective weights. That is, the system of equations takes the form

$$\begin{aligned} \mathbb{D}_1 &= w_{1,2}\mathcal{D}(E_2) + w_{1,3}\mathcal{D}(E_3) + w_{1,4}\mathcal{D}(E_4) + \cdots, \\ \mathbb{D}_2 &= w_{2,3}\mathcal{D}(E_3) + w_{2,4}\mathcal{D}(E_4) + w_{2,5}\mathcal{D}(E_5) + \cdots, \end{aligned} \tag{5.35}$$

where we propose the ansatz $\mathbb{D}_i \approx D_i$, identifying the model prediction with the measured dipole. Both measurements are sensitive to the same underlying $\mathcal{D}_j$, but each with different nonzero weights and additional contributions from other energy grids E_j.

The deconvolution involves more than simply summing the overlapping values to obtain $\mathcal{D}_j$. Instead, it determines a set of values $\mathcal{D}_j$ (one for each energy grid point) that best satisfy all the independent constraints simultaneously through an iterative optimization. To reconstruct the energy-dependent dipole vector $\mathcal{D}_j$ from measured dipole data D_i, we assume that the observed dipole at each energy bin is a weighted average of the dipole signal over a finite energy range. With respect to the pseudo-dipole matrix in Eq. (5.34), the matrix formulation of the ansatz $\mathbb{D}_i \approx D_i$ can then be written as

$$\mathbf{D} \approx \mathbf{W}\,\boldsymbol{\mathcal{D}}, \tag{5.36}$$

where the vector $\mathbf{D}$ contains M complex-valued dipole components ($\mathbf{D} \in \mathbb{C}^M$), $\mathcal{D}$ contains the dipoles at N discrete energy grid points ($\mathcal{D} \in \mathbb{C}^N$), and W is a $M \times N$ weight matrix. Each element ω_{ij} describes the contribution of the dipole at the energy grid point E_j to the measurement i, based on the energy response of that measurement. The inversion (or deconvolution) procedure solves Eq. (5.36) to obtain the full vector of dipole values $\boldsymbol{\mathcal{D}}$. Specifically, it solves for a set of underlying dipoles such that the model prediction, $\mathbb{D}_i = \sum_j \omega_{ij}\,\mathcal{D}_j$, closely approximates the observed dipoles D_i for all i. At each iteration, the full set of $\mathcal{D}_j$ values is updated simultaneously to minimize the difference between

predicted and measured dipoles. This process continues until convergence, i.e., until $\mathbb{D}_i \approx D_i$ is satisfied across all measurements. To determine the matrix $\boldsymbol{\mathcal{D}}$, we define a cost function that combines a data misfit term:

$$\mathcal{C}(\mathcal{D}) = \sum_{i=1}^{M} \left| D_i - \sum_{j=1}^{N_g} \omega_{ij} \mathcal{D}_j \right|^2 + \lambda \mathcal{R}(\mathcal{D}). \tag{5.37}$$

The first term corresponds to a least-squares data misfit, $\|\mathbf{D} - \mathbf{W}\boldsymbol{\mathcal{D}}\|^2$, which is equivalent to a chi-squared statistic under the assumption of uniform uncertainties, where all σ_i are taken to be equal and absorbed into an overall normalization. The second term introduces regularization, where $\mathcal{R} = \sum_{j=1}^{N_g-1} |\mathcal{D}_{j+1} - \mathcal{D}_j|^2$ denotes the regularization term and λ controls its strength and is taken to be 0.8. Since multiple measurements contribute to a single energy grid point, the inversion procedure can cause misfits in the reconstructed dipoles. To address this, we solve for $\boldsymbol{\mathcal{D}}$ iteratively by minimizing the cost function $\mathcal{C}$ using a gradient descent algorithm [313], which also regularizes the solution. Let the vector of dipoles be expressed in the form:

$$\boldsymbol{\mathcal{D}} = (\boldsymbol{\mathcal{D}}(E_1), \boldsymbol{\mathcal{D}}(E_2), \ldots, \boldsymbol{\mathcal{D}}(E_N))^T. \tag{5.38}$$

At each iteration n, the estimate is updated for the full system [313], which is given by:

$$\boldsymbol{\mathcal{D}}^{(n+1)} = \boldsymbol{\mathcal{D}}^{(n)} + \eta \left[\mathbf{W}^T \left(\mathbf{D} - \mathbf{W}\boldsymbol{\mathcal{D}}^{(n)} \right) - \lambda \nabla \mathcal{R} \right], \tag{5.39}$$

where η is the learning rate and is set to 0.3, and the second term is the gradient of the data misfit. The smooth gradient term $\nabla\mathcal{R} = \mathcal{D}_1 - \mathcal{D}_2$ for $j = 1$, $\nabla\mathcal{R} = 2(2\boldsymbol{\mathcal{D}}_j^{(n)} - \boldsymbol{\mathcal{D}}_{j-1}^{(n)} - \boldsymbol{\mathcal{D}}_{j+1}^{(n)})$ for all others j. The extracted dipole curves (amplitude and phase) are stabilized by reducing fluctuations between neighboring energy bins. This is approximately equivalent to local averaging that pushes each point toward the mean of its neighbors.

The initialization of $\boldsymbol{\mathcal{D}}$ is performed by averaging all measured dipoles whose energy range $[E^i_{\rm min}, E^i_{\rm max}]$ includes the grid point E_j. Specifically, the initial value at each grid point is given by $\mathcal{D}_j^{(0)} \equiv \sum_i D_i / N_{\rm g}^{\rm R}$, where $N_{\rm g}^{\rm R} = \{i | E_j \in [E^i_{\rm min}, E^i_{\rm max}]\}$ is the set of all measurements whose energy range includes E_j, and $N_{\rm g}^{\rm R}$ is the number of such measurements. The iteration continues until convergence is achieved, which is defined as the condition that the norm of the update vector is smaller than a convergence criterion ϵ:

$$\left\| \boldsymbol{\mathcal{D}}^{(n+1)} - \boldsymbol{\mathcal{D}}^{(n)} \right\| < \epsilon. \tag{5.40}$$

here we set the convergence threshold to $\epsilon = 10^{-12}$. After convergence is reached, the dipole amplitude and phase at each energy grid point are extracted from the complex solution, i.e., the best-estimated vector $\boldsymbol{\mathcal{D}}$, as $\mathcal{A}_j = |\mathcal{D}_j|$ and $\varphi_j = \arg(\mathcal{D}_j)$, where $\mathcal{D}_j$ is the j-th component of $\boldsymbol{\mathcal{D}}$. Results are shown in Figure 5.3.

The uncertainties on the deconvolved dipole amplitude and phase are estimated using a resampling (Monte Carlo) method that propagates measurement uncertainties through the inversion process in a statistically consistent manner. The core idea is to generate many samples of the measured dipole data by perturbing each measurement according to its known uncertainties, applying the deconvolution to each sample, and quantifying the statistical spread of the reconstructed dipoles at each energy grid point. Each measurement i has associated uncertainties $\sigma_{A,i}$ and $\sigma_{\Phi,i}$. For each resampling iteration r (out of a total of B), the measured amplitude and phase are perturbed as:

$$A_i^{(r)} = A_i + \epsilon_{A,i}^{(r)}, \quad \Phi_i^{(r)} = \Phi_i + \epsilon_{\Phi,i}^{(r)}, \tag{5.41}$$

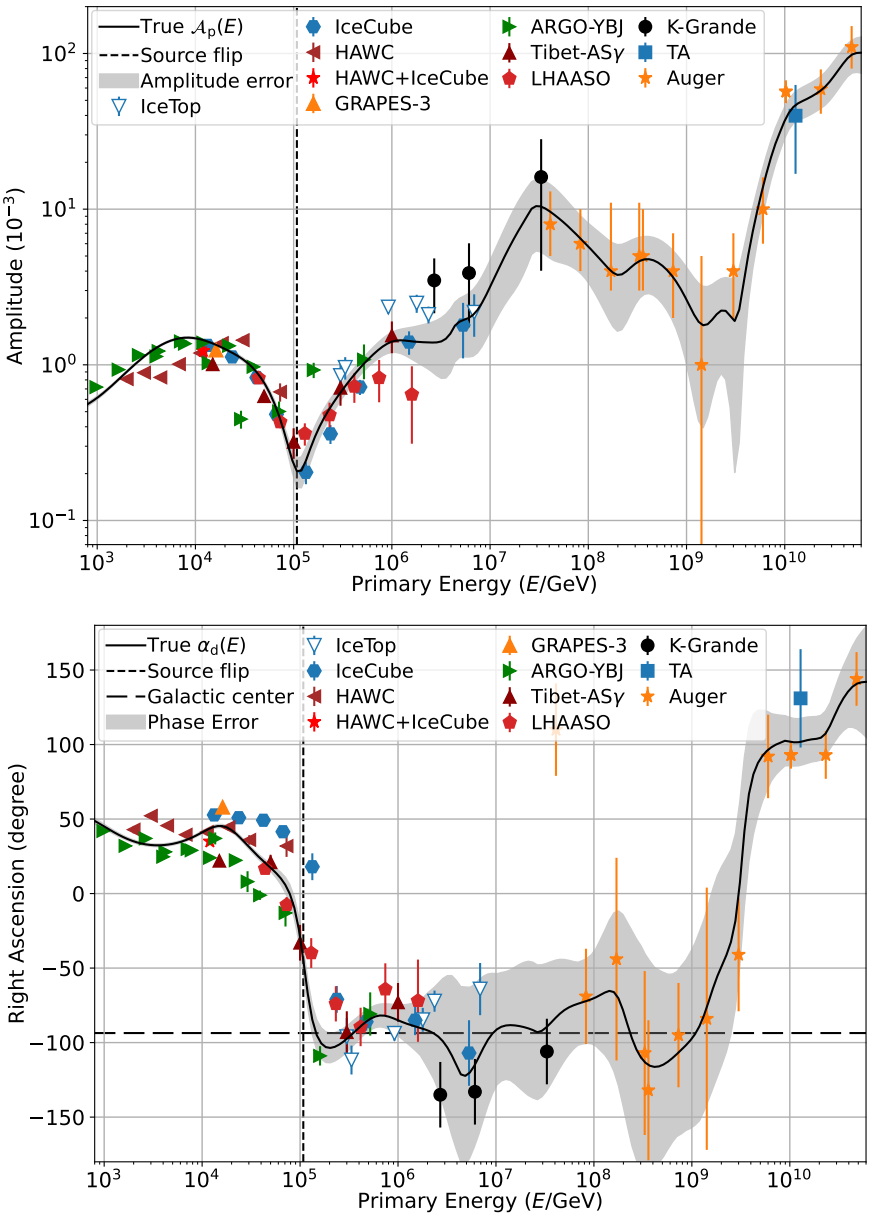


Figure 5.3: Unified anisotropy based on reported data from major experiments ([136, 87, 298, 299, 135, 269, 142, 140, 270, 300, 301]) spanning 1 TeV to 50 EeV, with all amplitudes calibrated by the FoV factor c_p as defined in Eq. (4.9) and (5.24). Note that the first Auger data point is excluded due to the large biased phase from the others. Top: extracted sidereal dipole amplitude, $\mathcal{A}_p(E)$. Bottom: the phase of the extracted dipole, $\alpha_d(E)$. Both fit curves are obtained as a function of energy. The error bands of $\mathcal{A}_p(E)$ and $\alpha_d(E)$ shown are by using data resampling. The short dashed line in both panels marks the flip energy at 107.9 TeV, with amplitude $\mathcal{A}_p = (2.1 \pm 0.40) \times 10^{-4}$ and phase $\alpha_d = 315.3° \pm 7.7°$, which may indicate a shift in the dominant cosmic-ray sources. The long dashed lines represent the galactic centre. The global fit curve of the amplitude exhibits a characteristic "W"–shaped structure, while the phase curve shows a twist in the middle.

where the perturbations $\epsilon_{A,i}^{(r)} \sim \mathcal{N}(0, \sigma_{A,i})$ and $\epsilon_{\Phi,i}^{(r)} \sim \mathcal{N}(0, \sigma_{\Phi,i})$ are independently drawn from Gaussian distributions centred at zero with standard deviations equal to the experimentally reported uncertainties. These perturbed values are converted into complex dipoles and used as inputs for the deconvolution. At each energy grid point E_j, only the measurements overlapping with that energy range, i.e., where $E_j \in [E_{\min}^i, E_{\max}^i]$, contribute to the estimation. This process quantifies the variability in the reconstructed dipoles due to measurement noise, with the spread across the B samples providing the statistical error bands on both amplitude and phase. Using the perturbed amplitude and phase, we construct the corresponding perturbed measured dipole

$$D_i^{(r)} = A_i^{(r)} e^{i\Phi_i^{(r)}}. \tag{5.42}$$

Using these perturbed dipole vectors as inputs, we repeat the full deconvolution procedure to obtain a resampled estimate of the underlying dipole vector, denoted by $\mathcal{D}^{(r)}$ for each iteration r. The deconvolution procedure is then repeated using the perturbed dataset, yielding a resampling estimate of the dipole vector at each energy grid point $\mathcal{D}_j^{(r)}$. Once all B resampling realizations have been processed, we compute the statistical uncertainty on the dipole amplitude and phase at each energy grid point j. The uncertainty on the amplitude is calculated as the standard deviation over the B resampling samples

$$\sigma_{\mathcal{A},j} = \sqrt{\frac{1}{B-1}\sum_{r=1}^{B}\left(\left|\mathcal{D}_j^{(r)}\right| - \bar{\mathcal{A}}_j\right)^2}, \quad \bar{\mathcal{A}}_j = \frac{1}{B}\sum_{r=1}^{B}\left|\mathcal{D}_j^{(r)}\right|. \tag{5.43}$$

where $\bar{\mathcal{A}}_j$ represents the mean reconstructed amplitude at the energy point. To determine the corresponding uncertainty in the reconstructed phase, we employ circular statistical methods, which are required because phases lie on a compact angular domain and cannot be treated with ordinary linear statistics. Specifically,

$$\sigma_{\varphi,j} = \sqrt{-2\ln(R_j)}, \quad R_j = \left|\frac{1}{B}\sum_{r=1}^{B} e^{i\varphi_j^{(r)}}\right|, \tag{5.44}$$

where R_j denotes the mean resultant length of the ensemble of phase vectors, giving a natural measure of phase around their mean direction. The error bands are obtained by repeatedly resampling the measured dipoles, performing the reconstruction 100 times, and averaging the results. To reduce statistical noise in the displayed uncertainty bands, the resulting standard deviations are additionally smoothed using a moving average in energy. Specifically, the standard deviation in the energy bin i is replaced by the average over the bin i and its four neighboring bins, which suppresses statistical fluctuations.

Based on this approach, we can extract the global and continuous dipole as a function of energy. We must emphasize that all reported anisotropy data from the main experiments have already incorporated this 3D geometric calibration, since they average data within specific right ascensions based on the HEALPix mapping method. This pixelation produces a subdivision of the spherical surface where each pixel covers the same surface area as every other pixel. Thus, direct averaging without additional declination-specific weighting maintains the integrity of the data by ensuring that each unit area of the celestial sphere is equally represented. However, we notice that the equal size and solid angle of each pixel can cause discrepancies in capturing large-scale anisotropic patterns, which vary from the equatorial plane to the poles. Near the poles, patterns are stretched and magnified, resulting in a disproportionately higher contribution to the average. Therefore, to achieve precise projections with minimal distortion, it's necessary to adjust the relative intensity,

δI, to decrease near the poles by a factor of $\cos\delta$. Alternatively, creating bands of equal width for the RA projection might be a more effective solution.

Based on the extracted dipole as a function of energy, a dipole flip is confirmed to occur at the energy of 107.9 TeV, where the dominant cosmic-ray sources appear to transition, corresponding to a sidereal amplitude of $\mathcal{A}_{\rm p} = (2.1\pm0.4)\times10^{-4}$ and phase $\alpha_{\rm d} = 315.3^\circ\pm7.7^\circ$. This flip may indicate a shift from nearby sources, such as the γ-ray source Crab Nebula (83.6°, 22°) and the Geminga supernova remnant [282, 135] (98.5°, 17.8°), to a nearby source region, where medium-scale anisotropies have been reported by ARGO-YBJ [282, 135]. The sidereal amplitude $\mathcal{A}_{\rm p}(E)$ of the dipole, as shown in Figure 5.3, exhibits a steeper energy dependence than that seen in the measured data in the range of 10^3 to 10^5GeV. This difference arises from the flattening effect in dipole reconstructions, where each measured dipole is a superposition of the dipole and contributions from neighbouring energy intervals. The resulting dipole phase tends to lie between the individual dipole phases, and the amplitude is correspondingly reduced. This averaging effect causes the measured dipole amplitude and phase to vary more slowly with energy compared to the dipole. Consequently, the observed flips in both amplitude and phase appear less sharp than those recovered through uncompress, i.e., the deconvolution.

Meanwhile, the global fit of the amplitude shows a characteristic "W-shaped" structure, as illustrated in Figure 5.3. The amplitude and phase shift around 3×10^{18} eV, corresponding to the energy of cosmic-ray ankle [123]. At this energy, the trend may indicate a growing contribution from an extragalactic component dominated by protons and helium. Overall, due to limited statistics, the fits of the dipole phase between (10^{15} eV and 10^{19} eV remain uncertain. In particular, between 1 PeV (Tibet-ASγ and IceCube) and 200 PeV (KASCADE-Grande), the dipole phases show complex variations and significant discrepancies, most notably in the KASCADE-Grande data. Therefore, a detailed new analysis of the KASCADE-Grande data is necessary to clarify the anisotropy in the PeV range, which is discussed in Chapter 6.

Chapter 6

Anisotropy Studies with KASCADE-Grande

In the reported anisotropy studies of KASCADE-Grande so far, no significant dipole anisotropy (more than 5σ) has been found [87, 288]. One reason for this is the lack of sufficient statistics, and in combination with methodical uncertainties, even though the dipole amplitudes in the respective energy range of the experiment are larger than 10^{-3}. Interestingly, recent work [8] from LHAASO found a gamma-ray bubble spanning at least 100 deg^2 in ultra-high energy (UHE), up to a few PeV, in the direction of the Cygnus X star-forming region. This bubble implies the presence of super PeVatrons capable of accelerating protons to at least 10 PeV. This energy range is also covered by KASCADE-Grande, which strongly motivates us to improve the significance of dipoles and medium-scale anisotropies using KASCADE-Grande data.

To reduce systematic effects and enhance the significance of the reconstructed dipole, we plan to apply the previously discussed angular-vector method to the measured sky maps, which should help minimize uncertainties in the reconstructed dipoles. Besides, to visualize the distributions of 2D equatorial dipoles, 3D celestial dipoles, and higher order multipole harmonics, we will employ an optimized maximum likelihood (ML) method based on that developed by Ahlers [143], and verify the large- and medium-scale features found in the 33 PeV bin by propagating CR particles from a potential source (Cygnus) using the CRPropa software [178] in this analysis.

6.1 The Experiment and Data Selection

The KASCADE-Grande experiment was an advanced air shower experiment dedicated to observing cosmic rays (CRs) by detecting charged particles generated in extensive air showers. The footprint of these showers and, in principle, the arrival times of particles on the ground enable us to reconstruct the arrival directions of the CRs. The number of charged particles (N_{ch}) in the shower serves as an indicator of the initial energy of the CRs. It was designed to study the composition and interactions of cosmic rays within the energy range of 10^{16} to 10^{18} eV, situated at the Karlsruhe Institute of Technology (KIT) in Karlsruhe, Germany (49.1°N, 8.4°E,110 m), KASCADE-Grande expanded upon the initial KASCADE experiment ($10^{14} - 10^{17}$ eV), which had been operational since 1996 [110]. This extension involved integrating 37 additional stations, each equipped with 10 square meters of plastic scintillation detectors, thereby enlarging the experimental area of KASCADE to approximately 0.5 square kilometres [110], as shown in Figure 6.1.

Operating in tandem with the existing KASCADE detectors until data collection ended in January 2013, KASCADE-Grande was arranged on an irregular triangular grid with an average station spacing of 137 meters. The stations were grouped into 16 hexagons, each consisting of seven modules that triggered upon full coincidence. This arrangement enabled precise measurements of cosmic ray shower parameters, including arrival direction, core location, and the total number of charged particles (N_{ch}) [110] as well as the muon number N_{μ} with the shielded detector of KASCADE. The main scientific objective of

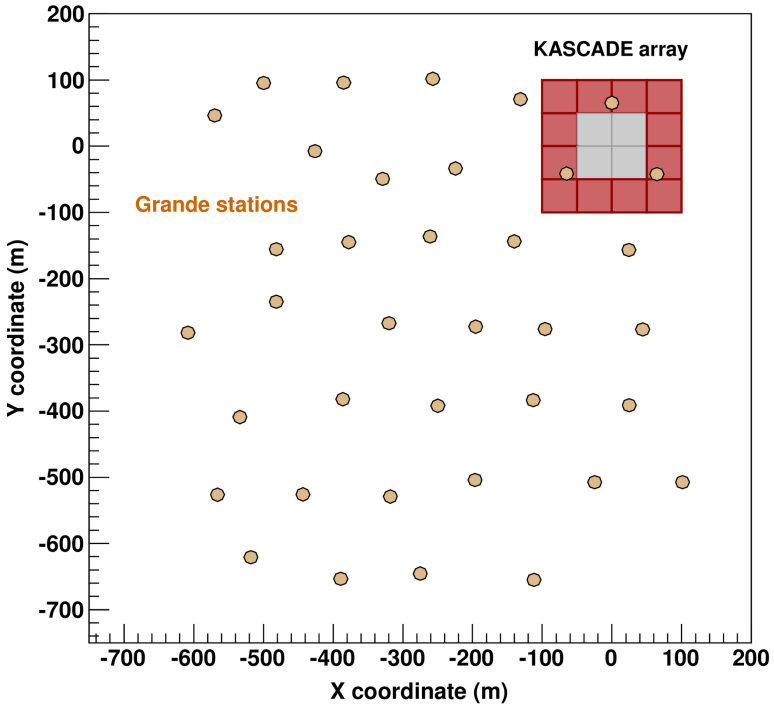


Figure 6.1: Layout of the KASCADE and the Grande arrays. The 37 Grande stations are marked by circles, and the 16 KASCADE detector clusters by squares. Each of the 12 outer KASCADE clusters includes 192 shielded plastic scintillator stations for measuring N_μ. This figure is taken from [314].

KASCADE-Grande was to investigate the transition from galactic to extragalactic cosmic rays, characterized by variations in the energy spectrum and mass composition. The experiment identified a knee-like structure in the energy spectrum of lighter elements, such as hydrogen and helium, around 10^{15} eV, and a similar feature for heavier elements, such as iron, near 80 PeV [23]. Additionally, an ankle-like structure was detected at around 100 PeV in the spectrum of light cosmic rays [23].

A notable discovery from KASCADE-Grande was that the muon attenuation length in the atmosphere exceeded predictions from existing hadronic interaction models [91]. This finding has significant implications for our understanding of cosmic ray air-shower development and interaction mechanisms. KASCADE-Grande's capability to measure muon numbers with high precision provided critical data for testing and refining these models [91]. The extensive dataset gathered over more than two decades has been made publicly accessible through the KASCADE Cosmic Ray Data Centre (KCDC[1]), in line with the Berlin Declaration on Open Data and Open Access. The ongoing analysis of this valuable dataset continues to improve our understanding of cosmic ray physics and contributes to the field of high-energy astroparticle physics [24].

For the current anisotropy study, we utilized data collected from March 2004 to October 2012, available through KCDC [24]. In the updated preselected datasets[2] from KCDC, we choose `ReducedData-GRANDE_runs_4775-7398_HDF5`, which limits event arrival directions

[1] See https://kcdc.iap.kit.edu.
[2] See https://kcdc.iap.kit.edu/datashop/fulldata/.

to zenith angles below 40°. To facilitate comparison with previous anisotropy studies by the KASCADE-Grande Collaboration [268, 87], we selected high-energy events with $N_{\rm ch} \geq 10^{5.2}$, where $N_{\rm ch}$ is the reconstructed total number of charged particles at the observation level of KG. The data is categorized into three $N_{\rm ch}$ bins corresponding to median energies of 2.7 PeV, 6.1 PeV, and 33 PeV (see Figure 6.2 and 6.3). These bins are defined by $\log_{10} N_{\rm ch}$ ranges of $[5.2, 5.6)$, $[5.6, 6.4)$, and ≥ 6.4, respectively, with median energies derived from Monte Carlo simulations [268]. All events are reconstructed with zenith angles $\theta \leq 40°$ and $N_{\rm ch} \geq 10^{5.2}$ are considered. The primary CR energy is estimated from $N_{\rm ch}(\theta)$, which is adjusted to a reference zenith angle ($\theta_{\rm ref} = 20°$) using the constant intensity cut technique [268]. This energy estimate is based on a calibration function for primary protons [315] using the QGSJetII-02 hadronic interaction model [316]. The total systematic uncertainty is about 20% at $E = 10^{16}$ eV and rises to $\sim 30\%$ at $E = 10^{18}$ eV, nearly independent of the hadronic interaction model [317]. These energy estimates are conservative, as the actual CR composition is likely heavier than pure protons. The dataset was also cleaned up by removing unnecessary array information for the data, meaning that only events with valid information are retained, enhancing data quality and relevance.

Since the primary energies in the anisotropy studies using KASCADE-Grande data are derived by converting the number of charged particles, $N_{\rm ch}$, into energy, we examine our data based on the $N_{\rm ch}$ distribution, as shown in Figure 6.2, and extract the corresponding energy distribution around each energy bin. The distributions of $N_{\rm ch}$ of the three energy bins can be fitted using two different functions:

$$\begin{aligned} N_1(x) &= \sum_k c_k x^k + \frac{A_{\ln}}{x\,\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \\ N_{2,3}(x) &= \lambda \exp[-k(x - x_0)], \end{aligned} \tag{6.1}$$

where N_1 corresponds to the fit for the first energy bin, while $N_{2,3}$ are used for the second and third energy bins. For N_1, the fit consists of a second-order polynomial combined with a log-normal distribution $F_{\ln}(x; \mu, \sigma)$. The coefficients are given by $c_1 = 1.212 \times 10^{10}, c_2 = -4.711 \times 10^9, c_3 = 4.498 \times 10^8$, with a log-normal amplitude $A_{\ln} = 2.646 \times 10^8$, mean $\mu = 1.663$, and standard deviation $\sigma = 0.075$. For the second energy bin, the distribution is fitted with an exponential tail of the form, where $\lambda_2 = 2.725 \times 10^7$, $k_2 = 4.294$, and $x_0^2 = 5.6$. For the third bin, the parameters are $\lambda_3 = 6.669 \times 10^5$, $k_3 = 4.575$, and $x_0^3 = 6.4$. Note that N_1 and $N_{2,3}$ are not normalized as probability density functions (PDFs), since the amplitudes are included to represent the full event distributions.

The primary energies are estimated from the correlation between the true primary energy and the reconstructed number of charged particles, which is carefully derived and experimentally validated through extensive Monte Carlo simulation studies [315]. This correlation considers both primary protons and iron nuclei, as well as different hadronic interaction models. Under the assumption of a linear relationship on a logarithmic scale, as obtained from fits to simulated data, the correlation is given by

$$\lg(E) = a_i + b_i \cdot \lg(N_{\rm ch}), \tag{6.2}$$

where E is in GeV, and the index i (ranging from 1 to 5) corresponds to proton, helium, carbon, silicon, and iron, respectively. For protons and iron nuclei, the fit parameters are $a_{\rm H} = 1.23$, $b_{\rm H} = 0.93$, and $a_{\rm Fe} = 1.75$, $b_{\rm Fe} = 0.90$, respectively, based on a specified primary composition [315]. The fit is performed within the range of full trigger and reconstruction efficiencies with zenith angle $\theta \leq 40°$. Since the energy calibration relies on simulations, it inherently depends on the choice of the hadronic interaction model. For the anisotropy

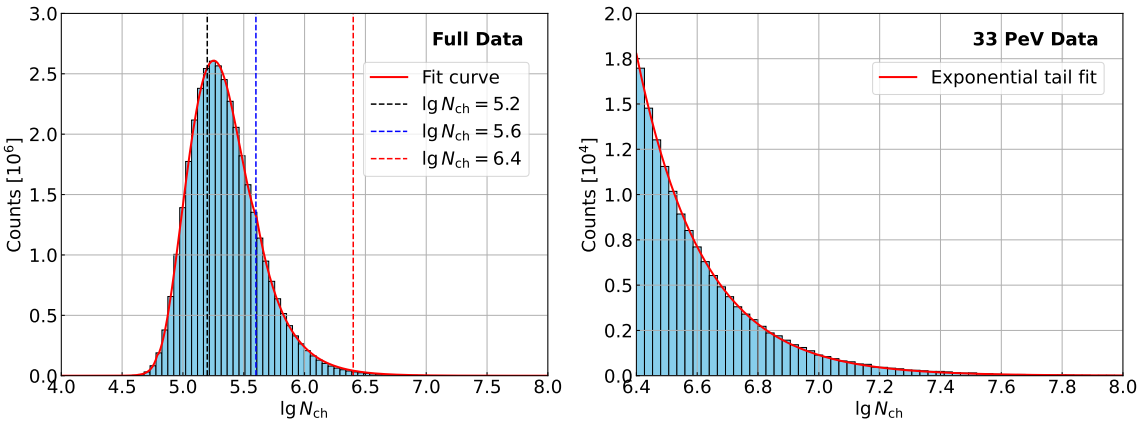


Figure 6.2: Distributions of KASCADE-Grande charged particles, shown for the full dataset (left) and zooming in on the high-energy (33 PeV) subset (right). The three lines indicate the boundaries of the three energy bins used in our analysis. The full distribution is fitted using a composite function: a log-normal distribution, a second-order polynomial in the range $5.2 \leq \lg(N_{\rm ch}) \leq 5.6$, and an exponential tail for $\lg(N_{\rm ch}) \geq 5.6$.

studies, the QGSJetII-02 model is used for energy estimation [315]. For other nuclei, we consider the interpolation between proton and iron by $\ln A$, where A is the nuclear mass number. Then we can get the correlation for each nucleus given by $\lg E_A = a_A + b_A \cdot \lg(N_{\rm ch})$, where $a_A = a_{\rm H} + (a_{\rm H} - a_{\rm Fe}) \ln A / \ln 56$, and $b_A = b_{\rm H} + (b_{\rm H} - b_{\rm Fe}) \ln A / \ln 56$. Finally, we obtain the estimated fit parameters for helium, carbon, and silicon as follows: $a_{\rm He} = 1.41$, $b_{\rm He} = 0.92$; $a_{\rm C} = 1.55$, $b_{\rm C} = 0.91$; and $a_{\rm Si} = 1.66$, $b_{\rm Si} = 0.91$, respectively.

Now we transform the measured all-particle charged particle distribution into an energy distribution. This can be achieved by applying an "average" or reference energy calibration. Alternatively, we can bin the data in energy by inverting the calibration relation. Using the charged particle distribution $N_{\rm ch}$ (Figure 6.2), and the fit function in Eq. (6.1), we convert the differential distribution $dN/\mathrm{d}\lg(N_{\rm ch})$ into the corresponding energy distribution $dN/\mathrm{d}\lg(E)$, and keep the binning in $\lg(E)$ space is properly handled. Now, by differentiating the calibration relation Eq. (6.2) with respect to $\lg(E)$ and $\lg(N_{\rm ch})$, we obtain the corresponding transformation of the differential distribution for a specific component i, which is in a Jacobian form:

$$\frac{\mathrm{d}N_i}{\mathrm{d}\lg(E)} = \frac{\mathrm{d}N_i}{\mathrm{d}\lg(N_{\rm ch})} \frac{\mathrm{d}\lg(N_{\rm ch})}{\mathrm{d}\lg(E)} = \frac{1}{b_i} \frac{dN_i}{\mathrm{d}\lg(N_{\rm ch})}. \tag{6.3}$$

Assume the original histogram is binned uniformly in $\lg(N_{\rm ch})$ with a bin width of $\Delta\lg(N_{\rm ch})$, (e.g., 0.1). To convert the binning into $\lg(E)$ space, the bin edges are mapped using the calibration relation. For a bin in $\lg(N_{\rm ch})$ with edges x_i and $x_{i+1} = x_i + \Delta x$, the corresponding edges in $\lg(E)$ space are $y_i = a_i + b_i\, x_i$ and $y_{i+1} = a_i + b_i\,(x_i + \Delta x)$. The bin width in $\lg(E)$ is then given by $\Delta\lg(E) = b_i\,\Delta\lg(N_{\rm ch})$. This bin mapping guarantees that the total number of events is conserved:

$$\sum_i \left(\frac{dN}{d\lg(N_{\rm ch})} \Delta\lg(N_{\rm ch}) \right) = \sum_j \left(\frac{dN}{d\lg(E)} \Delta\lg(E) \right). \tag{6.4}$$

In our data, the all-particle $\lg(N_{\rm ch})$ distribution originates from a mix of different components, each of which would be converted with its own slope b_i. When only the combined distribution is available, one can define an effective slope $b_{\rm eff}(E)$ as a weighted average of

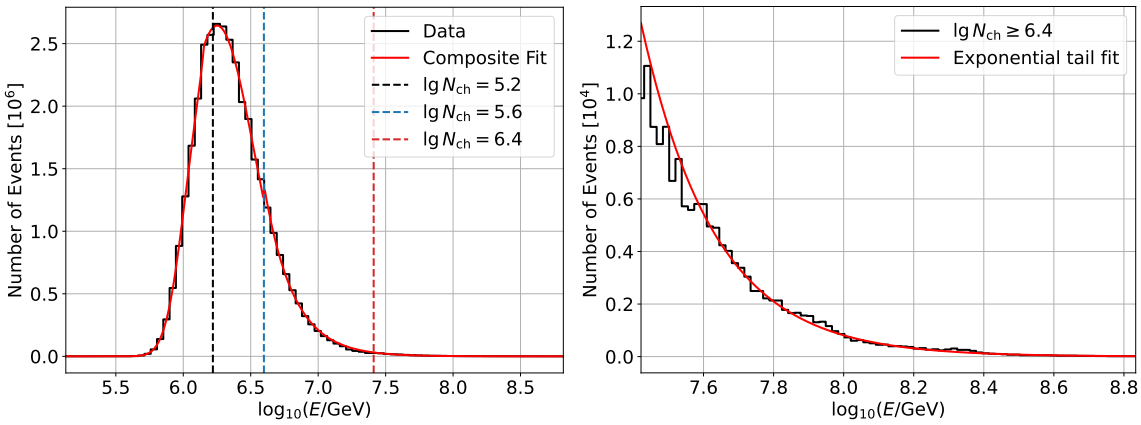


Figure 6.3: Distributions of the KASCADE–Grande energy samples extracted from the $\lg(N_{\mathrm{ch}})$ histograms shown in Figure 6.2, displayed for both the full dataset (left) and a zoomed-in view of the 33 PeV subset (right). The three vertical lines indicate the boundaries of the energy bins used in our analysis. The full distribution is fitted using a composite function consisting of a log-normal function in the range $5.2 \leq \lg(N_{\mathrm{ch}}) \leq 5.6$, while only exponential tail fits are applied to the $\lg(N_{\mathrm{ch}}) \geq 5.6$ distributions.

the slopes. Let $f_i(E)$ denote the fraction of events from the component i at the energy bin E (with $\sum_i f_i(E) = 1$). An effective calibration slope is defined as

$$b_{\mathrm{eff}}(E) = \sum_i f_i(E)\, b_i, \tag{6.5}$$

where $i = 1, 2..., 5$ and the functions $f_i(E)$ can be derived directly from the published KASCADE and KASCADE–Grande all-particle spectrum measurements [119] and available KCDC dataset[1]. The full set of resulting curves for $f_i(E)$, which quantify the effective spectral slopes for each of the defined energy intervals, is shown in Section 6.5 (see Figure 6.28). Using this effective slope, we can directly transform the all-particle distribution $\mathrm{d}N/\mathrm{d}\lg(N_{\mathrm{ch}})$ into the corresponding energy distribution consistently and continuously without needing first to separate the sample into individual mass components. As a result, the resulting energy distribution for the combined mass composition, valid over a given interval in $\lg(N_{\mathrm{ch}})$ range, is given by:

$$\frac{dN}{d\lg(E)} = \frac{1}{b_{\mathrm{eff}}(E)} \frac{dN}{d\lg(N_{\mathrm{ch}})}, \tag{6.6}$$

with the new bin width in energy space given by $\Delta\lg(E) = b_{\mathrm{eff}}(E)\,\Delta\lg(N_{\mathrm{ch}})$. After applying all the above steps, the total event count remains conserved as mentioned previously. The result is presented in the left panel of Figure 6.3.

Finally, the full reconstructed distribution is fitted using a composite functional form that combines three components: a log-normal distribution describing the lower part of the spectrum, a second-order polynomial valid in the intermediate range $5.2 \leq \lg(N_{\mathrm{ch}}) \leq 5.6$, and an exponential-tail function applied to the region $\lg(N_{\mathrm{ch}}) \geq 5.6$, which can capture the steeply falling behaviour at lower particle counts. From this fit, the lower and upper bounds of the central 68% containment of the distribution are 6.00 and 6.55 in $\lg(E/\mathrm{GeV})$, respectively. The resulting effective energy interval associated with the 2.7 PeV sample

[1] See https://kcdc.iap.kit.edu/spectra/.

is found to be $(6.22, 6.60)$ in $\lg(E/\mathrm{GeV})$. For the 6.1 PeV sample, the parameters of the exponential-tail probability density function, as defined in Eq. (6.1), are obtained from the fit to be $\lambda = 4.99$, $k = 4.66$, and $x_0 = 6.60$. Similarly, for the 33 PeV sample, the corresponding fit parameters are $\lambda = 1.61$, $k = 4.75$, and $x_0 = 7.41$. As part of these procedures, the median primary energy of each bin can also be estimated. For example, the median energy for the 33 PeV bin is determined using the condition $\lg N_{\mathrm{ch}} \geq 6.4$, with an upper limit of $(\lg N_{\mathrm{ch}})\mathrm{max} \sim 8$, which corresponds to the lower boundary of the true 33 PeV true energy bin, reported in [87]. Based on this, the energy range in $\lg E_A$ for each nuclear species is estimated to be: $\lg(E/\mathrm{GeV}) = (7.18, 8.67)$ for protons, $(7.30, 8.77)$ for helium, $(7.37, 8.83)$ for carbon, $(7.48, 8.94)$ for silicon, and $(7.51, 8.95)$ for iron.

6.2 East-West and Rayleigh Analysis

The KASCADE–Grande data set used in this analysis covers a total observation period of roughly eight years, as mentioned. During this period, the detectors experienced partial operations and shutdowns, along with environmental effects such as variations in temperature and atmospheric pressure. Consequently, these time-dependent effects can lead to non-negligible biases in the inferred anisotropy signals, particularly when studying small amplitudes at the 10^{-3} level or above. Therefore, it is necessary to examine the data in different time frames, such as solar time, anti-sidereal time and segmented local sidereal time. In parallel, we apply both the East–West method and the classical Rayleigh analysis to cross-check the stability of any detected modulation in the anisotropy studies.

6.2.1 Time Frames and Dipole Anisotropy

Sidereal time is a star-based astronomical timing system, in which the local sidereal time (LST) at an observatory corresponds to the right ascension of the celestial objects that are crossing the local meridian at that particular moment [272]. In cosmic-ray anisotropy studies, one challenge is variations in the data, which are caused by natural and human activities that directly affect the observed anisotropy patterns and introduce additional components. Fortunately, when the distribution of solar days in the data is sufficiently uniform, and the dominant systematic effects originate from natural (non-human) environmental variations, the East–West method can cancel the distorted components of the sidereal anisotropy [287, 288]. However, the large time gaps in data can still cause inconsistencies in the anisotropies. Therefore, we will examine the data in the solar, sidereal, and anti-sidereal time frames to better understand the sources of these inconsistencies.

Solar time is the standard time based on the Sun, consisting of 24 hours per day. The analysis of data with solar time can reveal systematic effects, including those caused by human activities and the Sun's regular influence on the environment, such as temperature and pressure variations. Additionally, many observatories, including KASCADE-Grande, have temperature-sensitive electronic components. If these electronics respond slightly in sync with the Sun, they may introduce artificial anisotropies due to changes in response, trigger, and threshold stability. If anisotropy is detected in solar time, it suggests local environmental effects rather than true cosmic-ray anisotropies.

A sidereal day is not exactly 24 hours; it's about 23 hours, 56 minutes, and 4 seconds long (or a sidereal day $D_{\mathrm{sid}} = 23.93447$ hours), roughly 4 minutes shorter than the 24-hour solar day. This small difference must be considered when searching for genuine cosmic-ray anisotropies. The relationship between solar time and sidereal time is $T_{\mathrm{sid}} = T_{\mathrm{sol}}(D_{\mathrm{sol}}/D_{\mathrm{sid}})$.

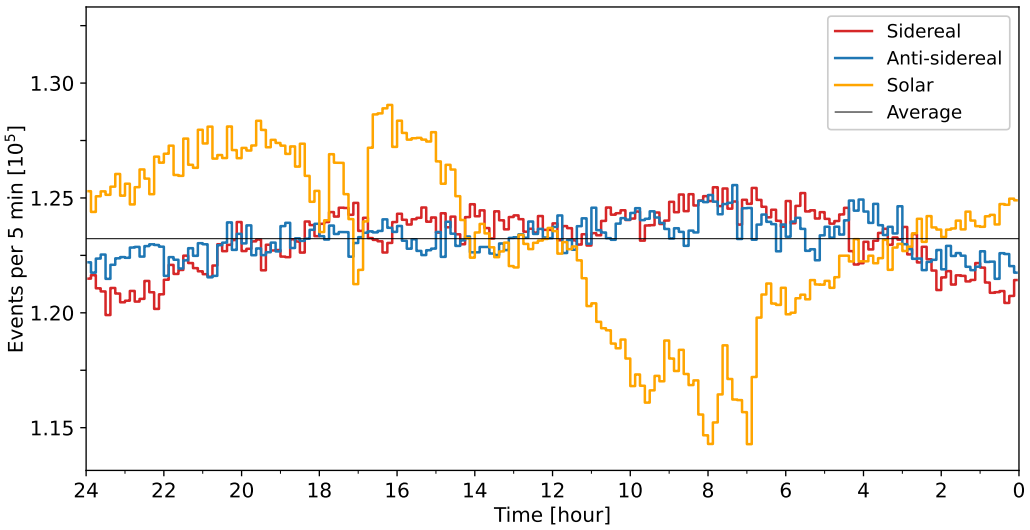


Figure 6.4: Event distributions (23.7 million events in total) from the selected KASCADE–Grande dataset (see Figure 6.3), each folded into 0–24 h cycles and displayed in the sidereal, anti-sidereal, and solar time reference frames.

Anti-sidereal time is a totally "fake" time system, which is the opposite way of the sidereal time and a bit longer than a solar day, since the anti-sidereal timescale is defined as $T_{\mathrm{anti}} = T_{\mathrm{sol}}(D_{\mathrm{sid}}/D_{\mathrm{sol}})$, which corresponds to 24 hours and 4 minutes per day (or 24.0657 hours). This corresponds to a bias in the reference frame, where a day is 4 minutes longer than a solar day. This defines the so-called anti-sidereal time [318, 319], which is a non-physical time frame. If an anisotropy appears in anti-sidereal time, it cannot originate from real astrophysical sources but may instead result from systematics related to non-regular environmental conditions, instrumental effects, or exposure variations. For example, detection efficiency may fluctuate throughout the day due to operational or environmental changes. Additionally, periods of detector operation and shutdowns can lead to irregular exposures. A strong signal in anti-sidereal time indicates potential issues with data-taking procedures or exposure modelling. The distribution of KASCADE-Grande data with $N_{\mathrm{ch}} \geq 10^{5.2}$ in sidereal, anti-sidereal, and solar time (day) is shown in Figure 6.4. A noticeable variation in solar time suggests a strong influence from the Sun on the environment or daily human activities, particularly those of researchers at the observatory doing maintenance of the detectors. Meanwhile, sidereal and anti-sidereal time exhibit similar variation amplitudes, but their periods and phases are difficult to determine. Although the solar-time variations are strong, they reflect instrumental and environmental effects rather than true signals. True cosmic-ray anisotropy is locked to the sidereal period,

Table 6.1: Amplitudes and phases of sidereal dipoles for the two subsets of each energy bin, shown for both the seasonally divided data and the time-ordered data splits.

Energy	$\mathcal{A}_{\mathrm{p}}^{\mathrm{sid}}(10^{-2})$	$\alpha_{\mathrm{d}}^{\mathrm{sid}}(^\circ)$	$\mathcal{A}_{\mathrm{p}}^{\mathrm{anti}}(10^{-2})$	$\alpha_{\mathrm{d}}^{\mathrm{anti}}(^\circ)$	$\mathcal{A}_{\mathrm{p}}^{\mathrm{sol}}(10^{-2})$	$\alpha_{\mathrm{d}}^{\mathrm{sol}}(^\circ)$
Total	0.57 ± 0.15	339.60 ± 15.28	2.55 ± 0.74	137.59 ± 16.53	1.90 ± 0.45	191.59 ± 13.46
2.7 PeV	0.48 ± 0.17	335.97 ± 20.03	2.51 ± 0.78	140.17 ± 17.72	1.79 ± 0.45	191.55 ± 14.43
6.1 PeV	0.86 ± 0.25	349.56 ± 16.43	2.70 ± 0.69	132.42 ± 14.50	2.20 ± 0.57	189.30 ± 14.80
33 PeV	2.37 ± 1.78	239.66 ± 42.99	3.57 ± 2.06	75.25 ± 33.26	4.49 ± 2.28	253.34 ± 29.28

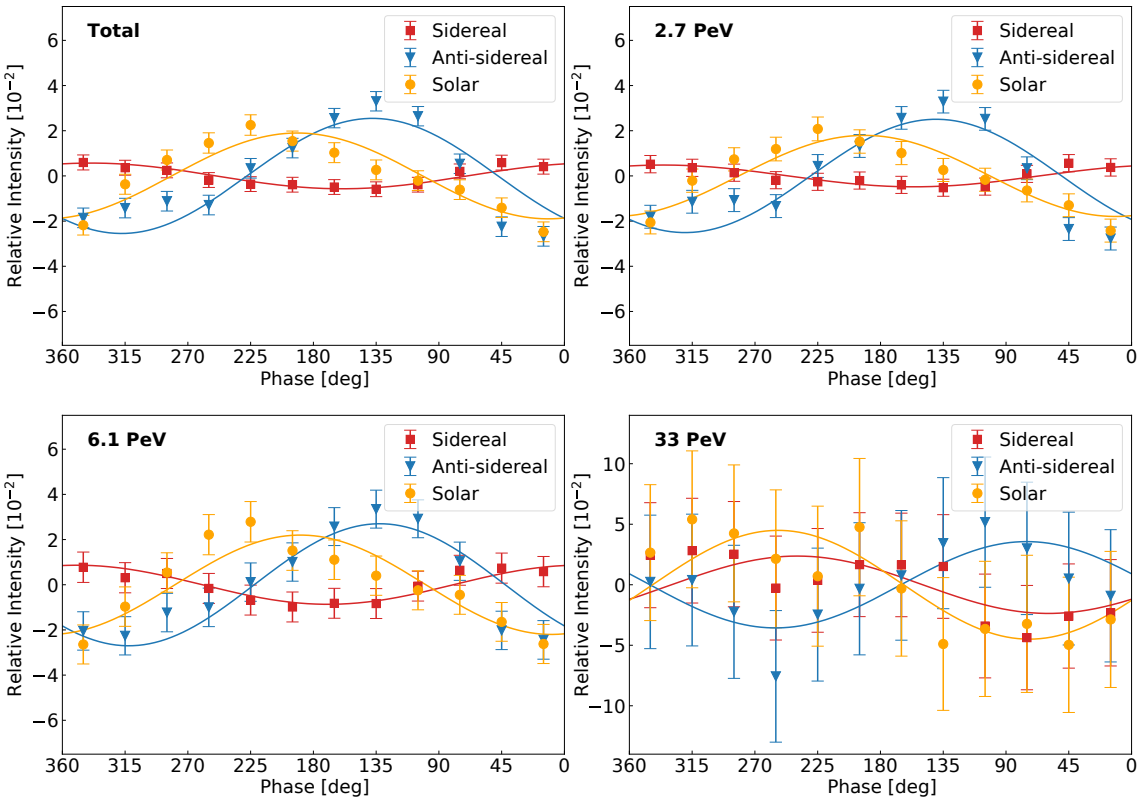


Figure 6.5: Relative intensity of cosmic-ray arrival directions in sidereal, anti-sidereal, and solar time frames using the East-West method, using a first-harmonic function. The dipole amplitude in the anti-sidereal time frame exceeds the statistical expectation from the sidereal frame, and its phase appears largely inverted, indicating residual systematics.

so solar-day fluctuations should weaken when the data are folded in sidereal or anti-sidereal time. However, the anti-sidereal distribution still exhibits large variations, comparable to those in sidereal time, revealing residual systematics.

To reduce acceptance-variation effects in the data, we apply the East–West derivative method [287, 288] to analyze anisotropies in each energy bin. To fit the data to a pure dipole, we perform the least-squares fit of Eq. (2.50), which models the differential relative intensity at each local sidereal time bin τ as

$$\partial_\alpha \delta I(\alpha_\tau) = -A_{\rm sid} \sin(\alpha_\tau - \alpha_{\rm d}), \tag{6.7}$$

where $A_{\rm sid}$ is the amplitude of the equatorial dipole and $\alpha_{\rm d}$ is its phase, and assuming the pure dipole $\delta I(\alpha_\tau) = A_{\rm sid} \cos(\alpha_\tau - \alpha_{\rm d})$. The observed values of $\partial_\alpha \delta I(\alpha_\tau)$ are those calculated from the data via the right-hand side of Eq. (2.50). As mentioned in Section 2.2.3, this method derives the relative intensity with respect to the binned and modelled right ascension (R.A.), automatically accounting for time-dependent variations in detector exposure by assuming equal acceptances from the eastern and western half-sectors of the sky. Therefore, in principle, it can smooth out long-term systematic biases and diurnal variations, especially in cases of slight directional or seasonal shifts, while remaining less sensitive to detector acceptance variations or atmospheric effects. Figure 6.5 presents the reconstructed dipole anisotropies of the arrival directions across the three time frames using the East-West derivative method, with detailed results provided in Table 6.1. However,

across the three time frames, the full dataset as well as the 2.7 PeV, 6.1 PeV and 33 PeV bins all show solar and anti-sidereal dipole amplitudes larger than the corresponding sidereal amplitudes, revealing significant time variations or systematic biases in the data that the East-West method cannot completely eliminate. For the three bins, each sidereal dipole always exhibits inverted phases compared to the corresponding anti-sidereal dipole, suggesting that systematic effects dominate the signal rather than a true cosmic anisotropy. Because true anisotropy only appears at the sidereal period, any signal that also shows up at the anti-sidereal period, especially with opposite phase, likely reflects solar-day systematics leaking into the observed dipole phases. Therefore, splitting the data in time is necessary to find, understand, and eventually mitigate these systematic effects.

6.2.2 Data Splitting Checks

To examine systematic effects in the data with details on sidereal anisotropy, we split the datasets for each energy bin into two independent time-based subsets. We then reconstructed the sidereal anisotropy separately for each subset to compare their consistency. One approach is to check for seasonal dependence, as done by IceCube, where the data were divided into austral summer and austral winter sets [319]. Another approach is to split the dataset into two time-ordered bins, as performed by the Pierre Auger Observatory [89]. These two ways would help identify potential systematic effects that could or noticeably influence the observed anisotropy.

We first split the total dataset and each energy bin (2.7 PeV, 6.1 PeV and 33 PeV) into winter (1 October–31 March) and summer (1 April–30 September) subsets to investigate possible seasonal influences and environmental dependencies in the measured anisotropy signals. For each of these subsets, we then repeat the full dipole reconstruction using the East–West method exactly as described previously, applying identical binning and fitting procedures to ensure direct comparability between the two subsets. The corresponding results are shown in the top four panels of Figure 6.6. From this comparison, we observe that, for the total dataset as well as for the 2.7 PeV and 33 PeV bins, energy bins, the dipole amplitudes obtained in winter and summer remain broadly consistent with one another, although the associated dipole phases show the largest degree of variation between the two seasons. In contrast, the 6.1 PeV bin shows a noticeable difference in amplitude between the two seasons: $\mathcal{A}_{\mathrm{p}}$ decreases from 0.0145 in winter to 0.0037 in summer, yielding $\Delta\mathcal{A}$ of 0.0108 (corresponding to roughly an 3σ effect), while its phase remains largely unchanged. These results indicate significant seasonal variations in dipole phase across all samples, which may still reflect residual atmospheric or detector-related systematics not fully corrected by the derivative East–West method.

To investigate possible long-term effects in the data, we next divided each energy sample into two chronologically ordered halves (2004–2008 and 2008–2012) to examine potential temporal drifts and stability in the reconstructed anisotropy parameters over the full observation period. We find that the dipole amplitudes remain consistent only for the 33 PeV, while all three bins exhibit phase shifts of $20^\circ - 40^\circ$ (exceeding their combined uncertainties), as shown in the bottom four panels of Figure 6.6. This indicates that the observed variations are not random but arise from underlying systematic effects, such as hardware upgrades or calibration changes, or evolving data-quality cuts that alter the detector's exposure profile over time and cannot be removed by simple seasonal splitting. Detailed dipole amplitudes and phases of the 3 bins are listed in Table 6.2. Therefore, a more detailed examination of the filtered data is necessary to identify the specific causes.

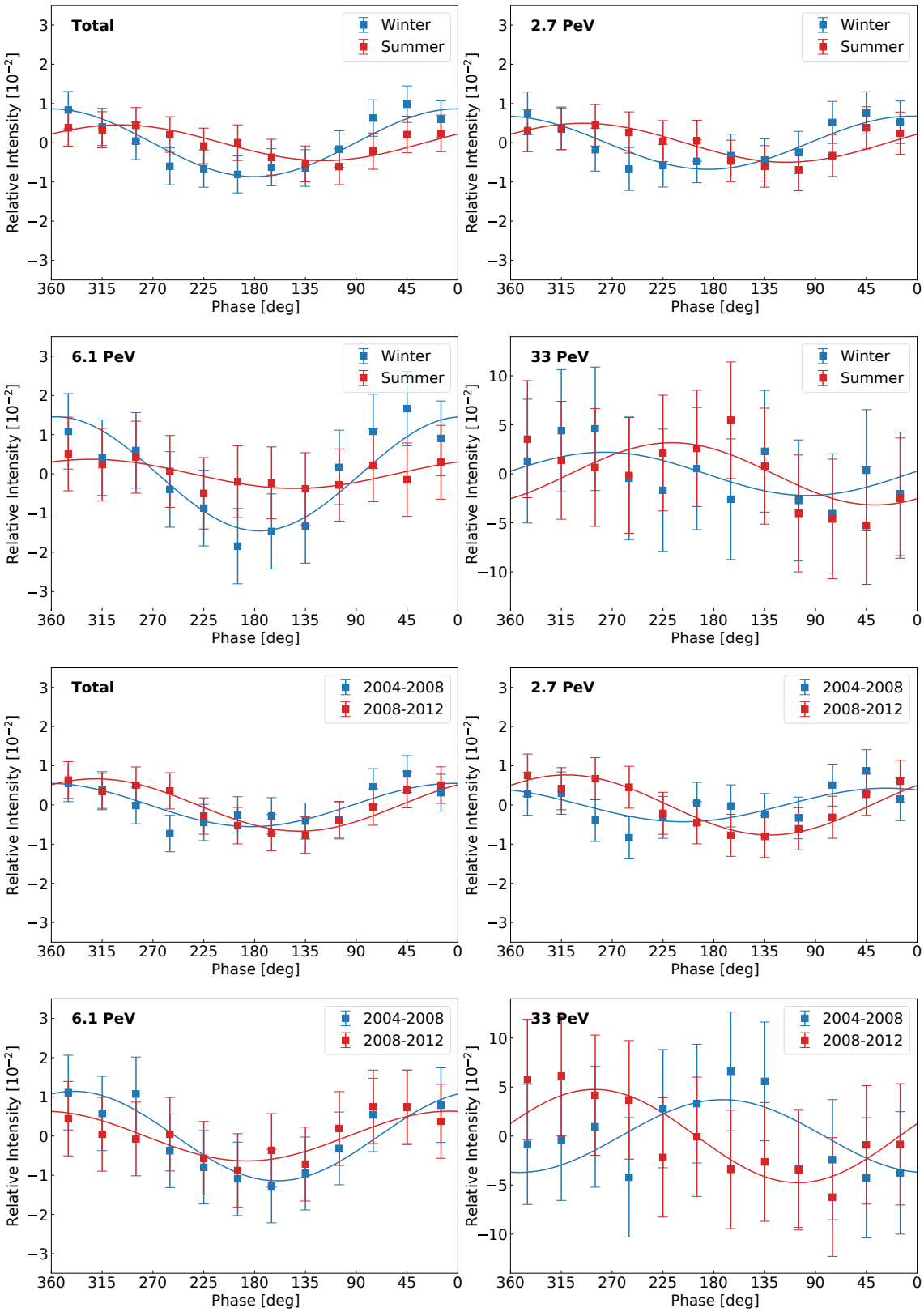


Figure 6.6: Comparison of the sidereal dipole for all three energy bins and the full dataset, using the derivative East-West method, under both seasonal and time-ordered data splits.

Table 6.2: Amplitudes and phases of sidereal dipoles for the two subsets of each energy bin, including both seasonal (winter and summer) and time-ordered data splits.

Median E	Season	$\mathcal{A}_{\rm p}(10^{-2})$	$\alpha_{\rm d}(^\circ)$	t-order	$\mathcal{A}_{\rm p}(10^{-2})$	$\alpha_{\rm d}(^\circ)$
Total	Winter	0.86 ± 0.21	0.40 ± 14.17	Bin 1	0.55 ± 0.29	3.05 ± 30.57
	Summer	0.45 ± 0.15	299.22 ± 18.33	Bin 2	0.67 ± 0.15	320.62 ± 13.31
2.7 PeV	Winter	0.68 ± 0.22	5.22 ± 18.40	Bin 1	0.43 ± 0.37	26.87 ± 49.08
	Summer	0.50 ± 0.20	295.98 ± 22.47	Bin 2	0.77 ± 0.17	310.60 ± 12.67
6.1 PeV	Winter	1.45 ± 0.45	355.88 ± 17.84	Bin 1	1.14 ± 0.36	340.07 ± 18.28
	Summer	0.37 ± 0.26	324.77 ± 39.53	Bin 2	0.63 ± 0.32	6.93 ± 28.67
33 PeV	Winter	2.21 ± 2.82	276.90 ± 73.02	Bin 1	3.71 ± 3.14	172.32 ± 48.54
	Summer	3.17 ± 2.38	217.08 ± 43.07	Bin 2	4.76 ± 2.44	285.51 ± 29.36

6.2.3 Data Quality Filtering

Since the above analyses reveal significant anisotropies in all time frames, one can examine the distribution of events across all sidereal days to better visualize temporal and directional variations in detector exposure. This can be done using a 2D histogram that maps the counts of arrival directions at each sidereal hour (or in half-hour intervals) throughout all recorded sidereal days and observation periods for the full data sample of a given energy range, and the details of the data collection can be directly displayed for further comparison between different energy bins and observation seasons.

First, we remove all sidereal days that are essentially empty, as such days do not contribute meaningful exposure and can distort the temporal structure of the dataset. Next, we clean the full dataset by excluding sidereal days that contain one or more substantial time gaps, thereby ensuring a more uniform exposure and a reliable temporal coverage across the entire observation period. Specifically, each 24-hour sidereal day is divided into 48 half-hour bins, and any day with a total gap (i.e., bins containing fewer than 100 events) exceeding 2 hours is removed, improving data stability. The data distributions before and after filtering are shown in Appendix C. Moreover, Figure 6.7 shows that, after filtering, sidereal and anti-sidereal variations are reduced while the solar variation remains large, reflecting normal daily fluctuations; overall variability is lower than before. The solar-time variations can be suppressed in the sidereal fold used for anisotropy studies. Still, any residual solar-day signal that isn't perfectly removed will leak into the sidereal analysis at a reduced amplitude (typically at the 10^{-4} level). We therefore fit the data to the dipole model to dig into these details.

Next, we split the filtered data and performed dipole reconstruction via the EW derivative method, utilizing Eq. (6.7) for fitting. After reconstruction, the R.A. phases for the total dataset and the 2.7 PeV and 6.1 PeV bins agree within 1σ uncertainties in the two-part comparison, whereas the 33 PeV bin remains a large shift that persists across multiple checks. We then compared these results with those from the unfiltered dataset (see Figure 6.8) to evaluate the impact of data filtering. The reconstructed dipole amplitudes are consistent between the two time-ordered subsets for all energy bins. The phase biases are reduced in the 2.7 PeV and 6.1 PeV bins but remain too large as before in the 33 PeV bin. This persistent disagreement most likely indicates additional time-dependent effects in the data, even though the EW method already corrects for angular acceptance and livetime variations [288]. Alternative anisotropy reconstruction methods are needed to fix the time gaps and further subtract the time-dependent variations from the data.

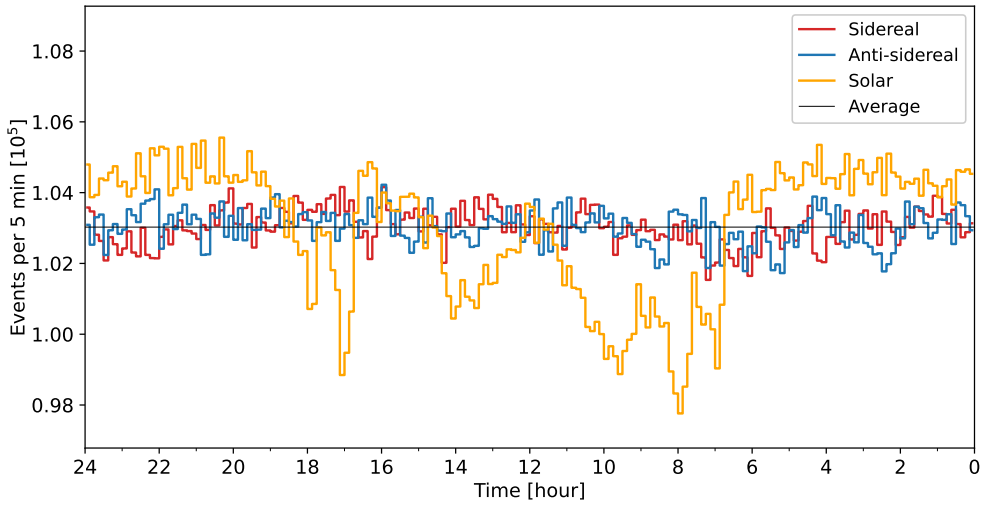


Figure 6.7: Event distributions (19.8 million events in total) of the filtered KASCADE-Grande data in sidereal, anti-sidereal, and solar time over a 24-hour period. The variations in all three time frames are reduced compared with the original data (see Figure 6.4), where the anti-sidereal frame shows a lower level of variation, while the solar-time frame still exhibits strong and twisted fluctuations.

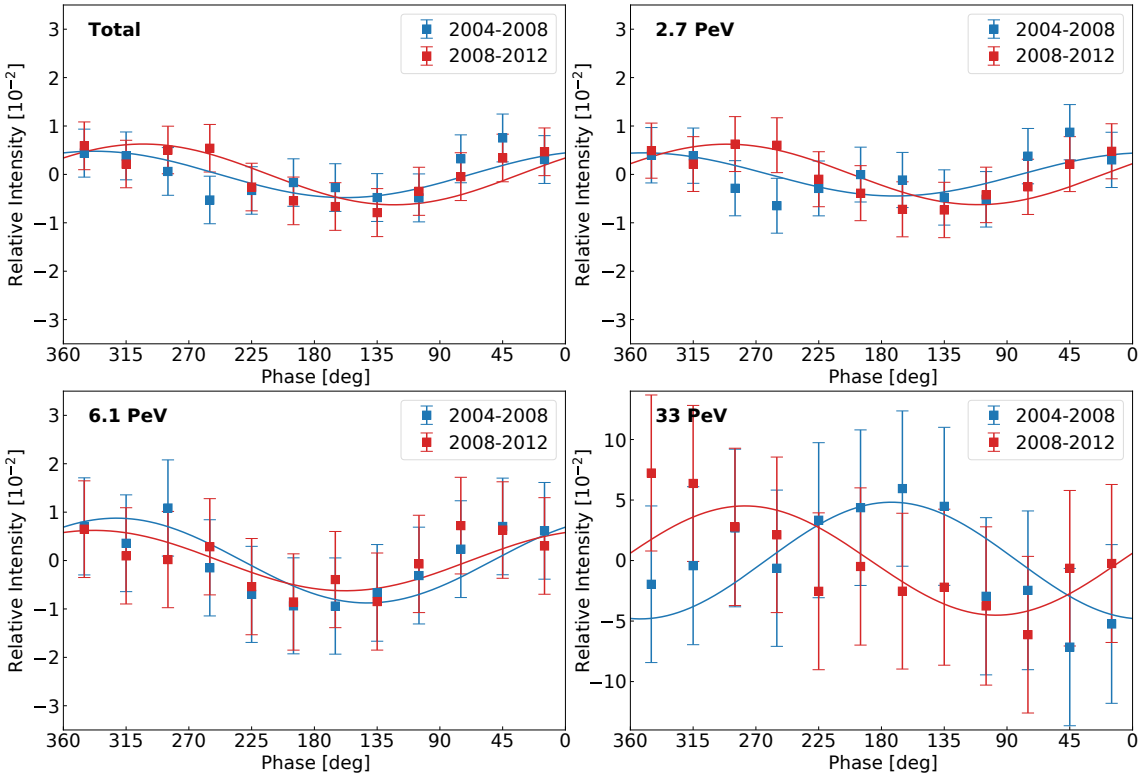


Figure 6.8: Comparison of the sidereal dipole across time-ordered data splits for all three energy bins and the full dataset, using filtered data analyzed with the derivative East–West method. The 2.7 PeV and 6.1 PeV bins now exhibit reduced biases between the two time-ordered subsets, while the 33 PeV bin still shows a large remaining bias.

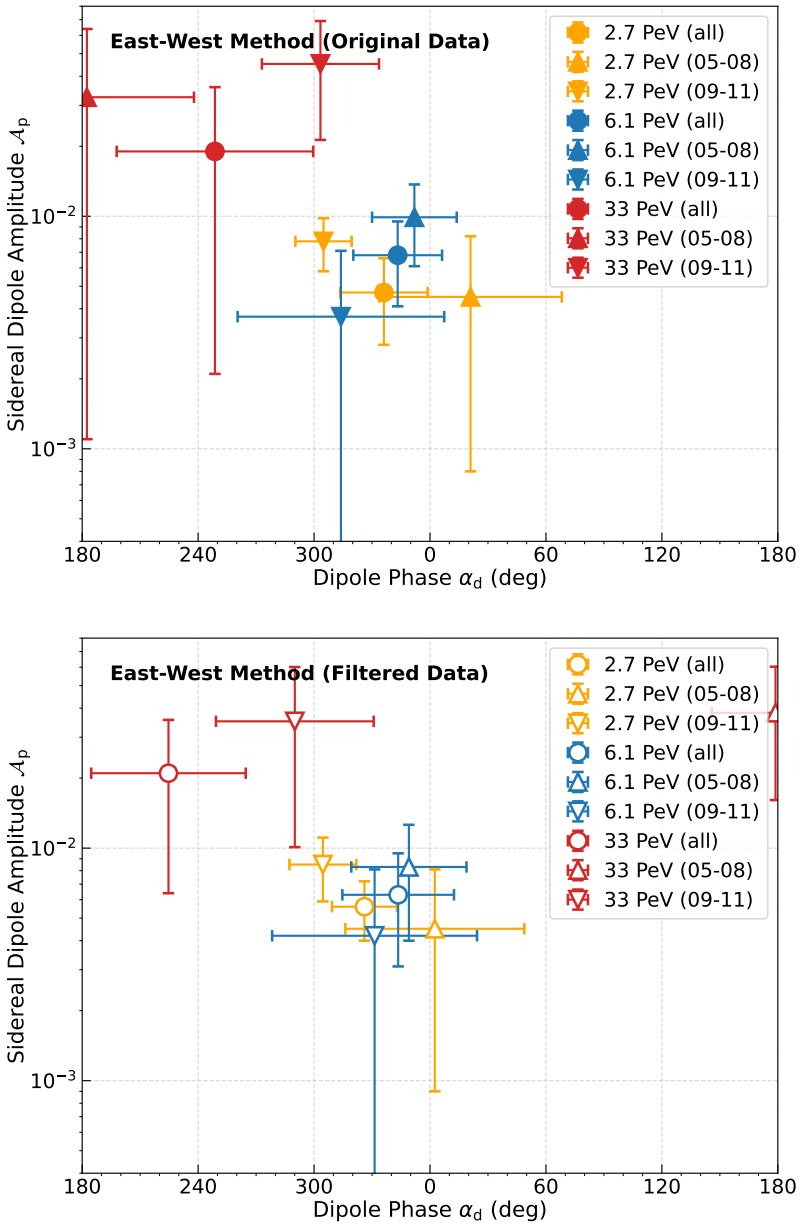


Figure 6.9: Comparison of the sidereal dipole in the phase–amplitude plane for time-ordered data splits (2005–2008, 2009–1011), together with the full data set (labeled as "all") in all three energy bins and in the full data set, using original unfiltered data (top) and filtered data (bottom), analyzed with the derivative East–West method. It appears that large biases are observed in both the dipole amplitudes and phases in all three energy bins. After data filtering, these biases are slightly reduced but remain substantial.

6.2.4 Rayleigh Analysis with Time-Scrambling

For reconstruction methods other than the East–West technique, it becomes essential to account explicitly for data calibration effects. Since our dataset contains sidereal gaps and many days have time gaps, where certain hours are underrepresented or even missing, the effective exposure is nonuniform over time, as seen in the 2D event distributions along sidereal days shown in Appendix C. Filtering the dataset improves the situation, but cannot eliminate these exposure variations, since gaps of different magnitudes persist across the entire observation period. To further reduce them, we apply a time-scrambling technique to the data. In this approach, each recorded event retains its local arrival direction (zenith and azimuth angles), but its R.A. is recalculated using event times randomly selected from a predefined time window ΔT around the original event time. In our analysis, we use a 24-hour window, following the method applied by IceCube [139], which maintains sensitivity to angular structures on all scales. Notably, scrambling event times while preserving the zenith angle in local detector coordinates is equivalent to randomly altering the R.A. within the same declination band, with the width determined by the chosen pixelization [139].

When performing the scrambling, live-time gaps are preserved in the resulting dataset, thus making the background data directly comparable. Each recorded event is characterized by event time t_i (seconds), original right ascension α_i, local arrival direction zenith θ_i and azimuth ϕ_i. To generate a scrambled dataset, we randomly assign new times t_i' to each event within a predefined time window

$$t_i' = t_i + \delta t, \quad \delta t \sim \mathcal{U}(-\Delta T/2, \Delta T/2), \tag{6.8}$$

where δt is uniformly drawn from the time interval $(-\Delta T/2, \Delta T/2)$. Following the method applied by IceCube [139], where $\Delta T = 24$ sidereal hours. The sidereal time at t, denoted by $\Theta_{\text{sid}}(t)$, increases linearly due to the Earth's rotation:

$$\Theta_{\text{sid}}(t) = \Theta_{\text{sid}}(t_0) + \omega_{\text{sid}}(t - t_0), \tag{6.9}$$

where $\Theta_{\text{sid}}(t_0)$ is the sidereal time at the reference epoch t_0, $\omega_{\text{sid}} \simeq 2\pi/23.93$ is the Earth's sidereal angular velocity, $(t - t_0)$ is the elapsed time in seconds. The reference epoch t_0 is typically chosen based on astronomical standards or the dataset's initial timestamp. Using the newly assigned event time t_i', the sidereal time is updated as:

$$\Theta_{\text{sid}}(t_i') = \Theta_{\text{sid}}(t_0) + \omega_{\text{sid}}(t_i' - t_0). \tag{6.10}$$

Given the hour angle H_i of the event, the new right ascension is

$$\tilde{\alpha}_i = \Theta_{\text{sid}}(t_i') - H_i \pmod{360^\circ}, \tag{6.11}$$

and local arrival direction (θ_i, ϕ_i) remains unchanged while only the RA is randomized. The scrambled data is then $(\alpha, \delta) \leftarrow (\tilde{\alpha}_i, \delta_i)$. It is worth noting that scrambling the local sidereal time of arrival events is equivalent to scrambling the RA coordinates within the same declination ring as the original arrival directions, making it a simpler approach compared to the time-scrambling, namely

$$\tilde{\alpha}_i = \alpha_i + \left| \frac{t_i + \delta t}{3600\,\text{s}} \right|_{24\text{h}} \times 15^\circ, \tag{6.12}$$

where the $|.|$ symbol denotes that the scrambled time is constrained within the range of 0 to 24 hours. As a result of this construction, the difference between the true distribution of arrival directions and the corresponding reference maps, where each declination band is independently normalized, remains sensitive to anisotropy in right ascension only.

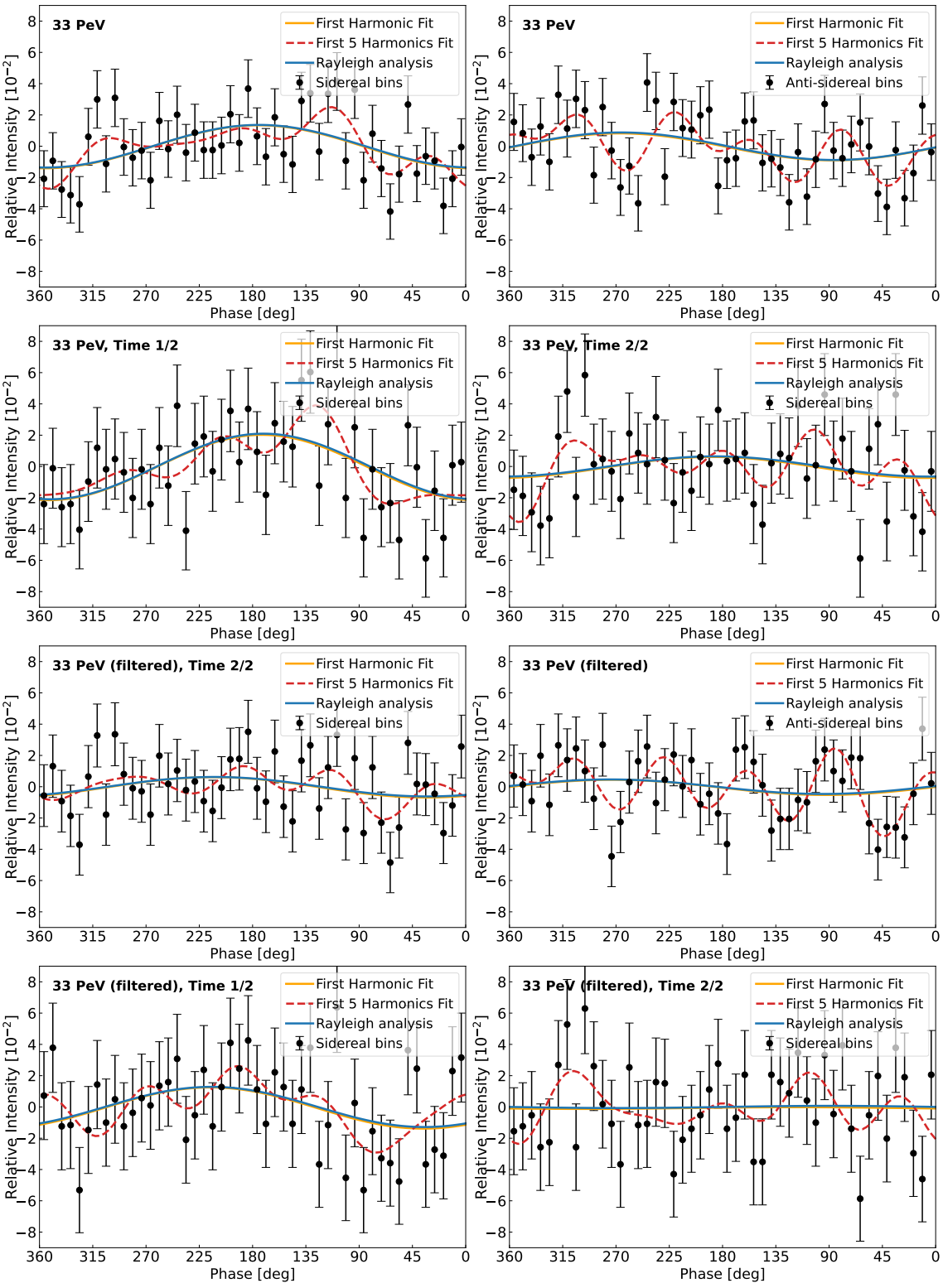


Figure 6.10: Comparison of the Rayleigh and Fourier analyses using right ascension bins. The top four plots show the original data without filtering, while the bottom four plots show the filtered results. Both without time-scrambled data as the background.

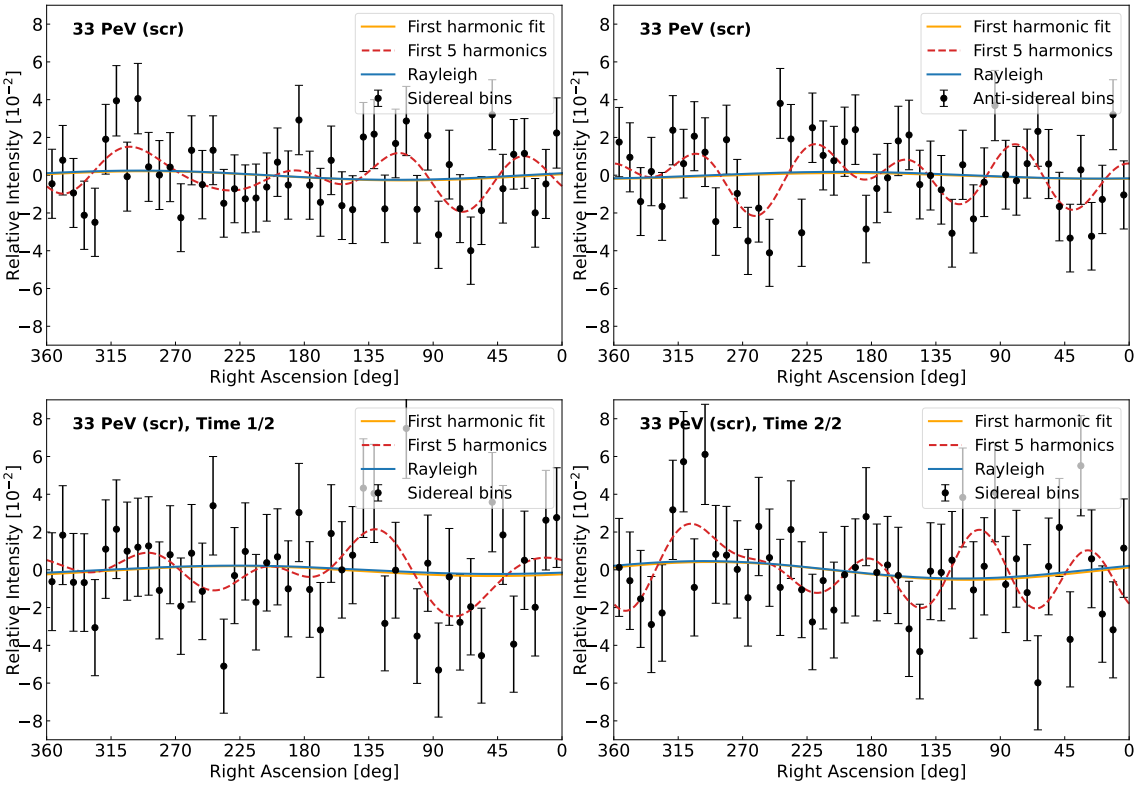


Figure 6.11: Comparison of Rayleigh and Fourier analyses using right ascension bins. The data are unfiltered, with time-scrambled data used as the background. The top right panel shows a smaller anti-sidereal dipole amplitude than the sidereal case. The bottom-left and bottom-right panels correspond to the first and second halves of the dataset, respectively, split in sidereal time, and reveal a large bias in dipole phases.

The standard Rayleigh analysis (see Section 2.2.1) for large-scale anisotropy studies does not account for acceptance or exposure variations, and therefore requires explicit corrections for these data fluctuations [278]. When the FoV approaches full-sky coverage, the obtained amplitude corresponds to the actual projected equatorial dipole $\mathcal{A}_{\rm p}^{\dagger} = \mathcal{A}\, p \cos\delta_{\rm d}$, where $p = \langle \cos\delta \rangle$; otherwise, it will be shifted and requires a 3D reconstruction model for precise dipole recovery. In our dataset, sidereal gaps and uneven event distributions over sidereal time introduce nonuniform exposure, as mentioned, as illustrated in Appendix C; therefore, we employ time-scrambling. After time scrambling, the dataset is used in the weighted Rayleigh analysis, where each scrambled event is assigned a background. Then, we subtract the background from the data, the effective Fourier coefficients $\tilde{a}_k$ and $\tilde{b}_k$ are computed as $\tilde{a}_k = a_k^{\rm d} - a_k^{\rm b}$ and $\tilde{b}_k = b_k^{\rm d} - b_k^{\rm b}$, namely,

$$\tilde{a}_k = \frac{2}{N}\sum_{i=1}^{N} \cos k\alpha_i - \frac{2}{nN}\sum_{j=1}^{nN} \cos k\alpha_j^{\rm scr}, \quad \tilde{b}_k = \frac{2}{N}\sum_{i=1}^{N} \sin k\alpha_i - \frac{2}{nN}\sum_{j=1}^{nN} \sin k\alpha_j^{\rm scr}, \quad (6.13)$$

where $a_k^{\rm d}$ and $b_k^{\rm d}$ denote the harmonic coefficients computed directly from the data using the original right-ascension values α_i, while $a_k^{\rm b}$ and $b_k^{\rm b}$ are obtained from the scrambled α_j, denoted by $\alpha_j^{\rm scr}$. Here, N is the total number of cosmic-ray arrival directions, and n represents the number of times the data is scrambled to obtain a background. For the 33 PeV energy sample, we $n = 200$ for scrambling to achieve a smooth and high-precision

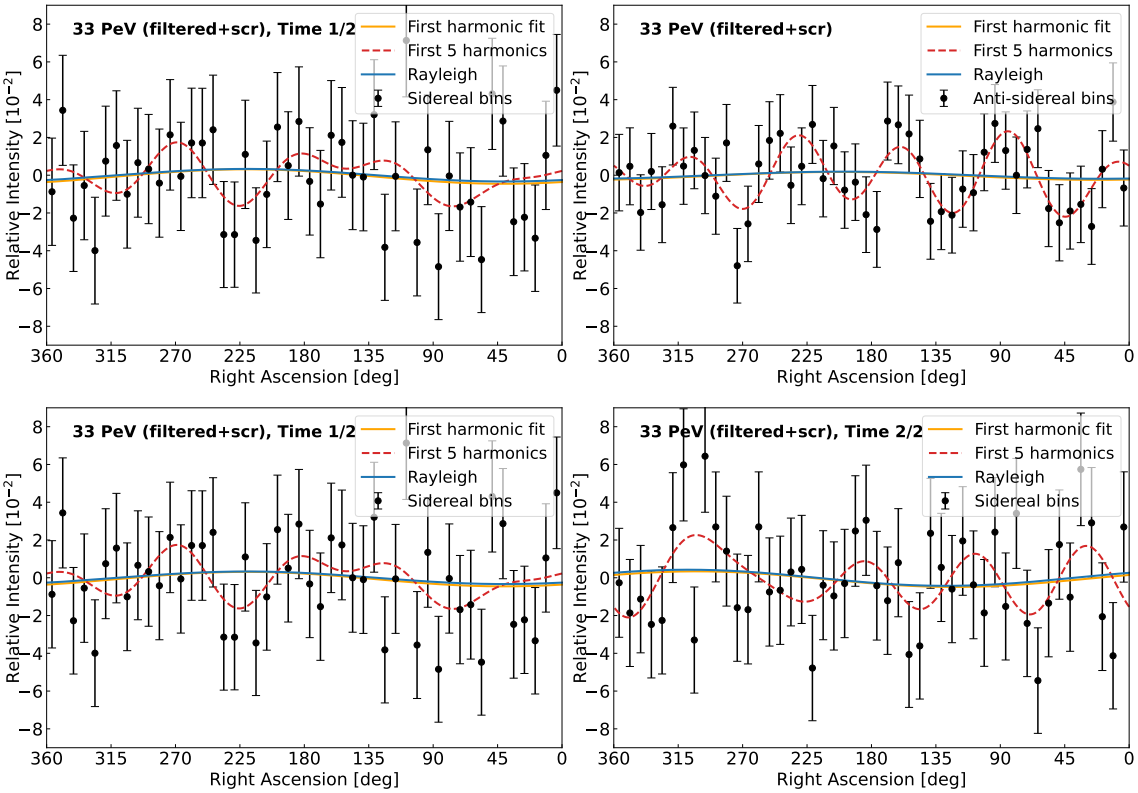


Figure 6.12: Comparison of Rayleigh and Fourier analyses using right ascension bins. The data is filtered, and time-scrambled data are used as the background. The top right panel shows a smaller anti-sidereal dipole amplitude than the sidereal case, while the bottom two plots of the split data reveal a large bias in dipole phases.

background. Once the difference between the data and background coefficients is obtained, the harmonic amplitude and corresponding phase are obtained as

$$r_k = \sqrt{\tilde{a}_k^2 + \tilde{b}_k^2}, \quad \varphi_k = \frac{1}{k} \arctan \frac{\tilde{b}_k}{\tilde{a}_k}, \tag{6.14}$$

with $k = 1$, it yields the dipolar component in the harmonics. Considering the standard propagation of errors, we obtain that the uncertainties of r_k and φ_k are

$$\sigma_r = \sqrt{\frac{2(1+1/n)}{N}}, \quad \sigma_\varphi = \frac{1}{k r_k} \sqrt{\frac{2(1+1/n)}{N}}, \tag{6.15}$$

which shows that the background subtraction increases the variance slightly by a factor of $(1 + 1/n)$, thereby accounting for the additional uncertainty introduced through the finite number of scrambling realizations used to construct the background. The corresponding p-value $P(\geq r_k)$ in this analysis is the probability of obtaining an amplitude equal to or larger than the observed sidereal dipole amplitude r under the null hypothesis of isotropy [131], which is the tail probability of the Rayleigh distribution. Assuming $r_x = a$ and $r_y = b$ are the independent Gaussian random variables with zero mean $\mu = 0$ (because the vectors are uniformly distributed) and variance σ^2, their joint probability density function is the product of two 1D Gaussian: $f(r_x, r_y) = (1/2\pi\sigma^2) \exp(-(r_x^2 + r_y^2)/2\sigma^2))$, which satisfy $\iint f(r_x, r_y)\, \mathrm{d}r_x\, \mathrm{d}r_y = 1$. Using the polar coordinates (r, θ) with $r_x = r \cos\theta$

and $R_y = r \sin\theta$, the area element becomes $\mathrm{d}r_x\,\mathrm{d}r_y = r\,\mathrm{d}r\,\mathrm{d}\theta$. Integrating out θ then gives the Rayleigh PDF for the magnitude r:

$$P(\geq r) = \int_r^\infty \int_0^{2\pi} \frac{1}{2\pi\sigma^2}\, \exp\!\Big(-\frac{r^2}{2\sigma^2}\Big)\, r\,\mathrm{d}\theta\,\mathrm{d}r = \exp\left(-\frac{r^2}{2\sigma^2}\right), \tag{6.16}$$

where the Gaussian's variance $\sigma^2 = 1/N$ for the standard Rayleigh analysis. In our case, where the background is subtracted, substituting σ_r from Eq. (6.15), one obtains:

$$P(\geq r_k) = \exp\left(-\frac{N r_k^2}{4(1+1/n)}\right), \tag{6.17}$$

where $1/n$ contributes a weight to each scrambled event. If n is large (say $n = 20$), then $(1 + 1/n) \approx 1.05$, which is only a small correction compared to the standard case $P = \exp(-Nr_k^2/4)$. A small value of P, therefore, indicates significant deviation from isotropy. The corresponding significance in standard deviations is $\sigma = \sqrt{2} \cdot \mathrm{erfc}^{-1}(p)$.

In addition to the Rayleigh analysis, we apply a Fourier analysis to the binned right ascension (RA) distribution to investigate the first harmonic and the first 5 components of harmonics. The Fourier analysis involves binning the cosmic-ray arrival directions in RA and computing the relative intensity within each bin. The number of events in each RA bin is determined from the observed dataset, while a background distribution is generated using time-scrambled data. By comparing the observed event counts to the expected background, we compute the relative intensity as $\delta I = \mathcal{N}_i^{\mathrm{d}}/\mathcal{N}_i^{\mathrm{scr}} - 1$ for each RA bin. The data is then

Table 6.3: Shown are all four cases: (1) original data without filtering or scrambling, (2) filtered data without scrambling, (3) original data without filtering but with scrambling, and (4) filtered data with scrambling. The table lists amplitudes, phases, and p-values from the Rayleigh and Fourier analyses using right ascension bins. Results are presented for sidereal and anti-sidereal anisotropies, as well as for the first and second halves of the dataset split in sidereal time. The last case provides a more stable reconstructed dipole amplitude; however, the phase still shows a large bias between the two split parts.

33 PeV	$r_1^{\mathrm{Ray}}(10^{-2})$	$\varphi_1^{\mathrm{Ray}}(^\circ)$	P-Ray	$r_1^{\mathrm{Fou}}(10^{-2})$	$\varphi_1^{\mathrm{Fou}}(^\circ)$	P-Fou
Sid (original)	1.4 ± 0.4	174.56 ± 15.53	0.0011	1.4 ± 0.4	175.28 ± 16.46	0.0010
Anti (original)	0.9 ± 0.4	265.61 ± 24.05	0.0586	0.9 ± 0.4	266.70 ± 28.03	0.0653
1/2 (original)	2.1 ± 0.5	170.88 ± 14.28	0.0003	2.1 ± 0.5	171.94 ± 14.76	0.0003
2/2 (original)	0.6 ± 0.5	186.57 ± 46.29	0.4649	0.7 ± 0.5	188.41 ± 45.15	0.4364
Sid (filter)	0.6 ± 0.4	212.31 ± 37.24	0.3061	0.6 ± 0.4	213.04 ± 36.56	0.2743
Anti (filter)	0.5 ± 0.4	275.07 ± 49.97	0.5183	0.5 ± 0.4	273.13 ± 50.18	0.4856
1/2 (filter)	1.3 ± 0.6	214.89 ± 25.51	0.0803	1.3 ± 0.5	215.64 ± 23.83	0.0707
2/2 (filter)	0.1 ± 0.6	88.55 ± 470.91	0.9926	0.1 ± 0.6	297.99 ± 586.74	0.9954
Sid (ori+scr)	0.2 ± 0.4	295.64 ± 87.07	0.8053	0.2 ± 0.4	290.28 ± 87.05	0.8053
Anti (ori+scr)	0.2 ± 0.4	201.69 ± 118.94	0.8905	0.2 ± 0.4	198.30 ± 136.58	0.9159
1/2 (ori+scr)	0.2 ± 0.5	227.11 ± 135.48	0.9145	0.3 ± 0.5	229.03 ± 114.75	0.8818
2/2 (ori+scr)	0.5 ± 0.5	297.15 ± 65.09	0.6788	0.5 ± 0.5	292.56 ± 63.52	0.6681
Sid (filter+scr)	0.4 ± 0.4	297.50 ± 63.93	0.6692	0.4 ± 0.4	291.91 ± 63.24	0.6650
Anti (filter+scr)	0.2 ± 0.4	199.59 ± 125.71	0.9013	0.2 ± 0.4	202.78 ± 113.02	0.8798
1/2 (filter+scr)	0.3 ± 0.6	219.74 ± 100.75	0.8507	0.4 ± 0.6	220.54 ± 87.64	0.8059
2/2 (filter+scr)	0.4 ± 0.6	308.15 ± 78.65	0.7669	0.4 ± 0.6	302.26 ± 81.09	0.7790

fitted to a harmonic expansion, $\delta I = \sum_k A_k \cos(k\alpha - \varphi_k)$, where k ranges from 1 to 5, as the monopole term in this analysis is zero.

The results are shown in Figure 6.11 for the 33 PeV energy bin only, to address the issue of large biased phases as illustrated in Figure 6.8, without filtering, and in Figure 6.12 with filtering applied (days with total gaps exceeding two hours, i.e., bins containing fewer than 100 events, are removed). The first-harmonic fits from the Fourier analysis closely overlap with the dipole curves obtained from the Rayleigh analysis in all cases. The anti-sidereal dipole amplitude falls within the 1σ uncertainty range of the sidereal dipole, but the phases remain inverted. Additionally, the first half of the dataset always exhibits a significant dipole phase bias compared to the second half. Moreover, medium-scale anisotropies from higher-order harmonic fit curves exhibit significant instability across the two time periods. Furthermore, as seen in the bottom-right plot of Figure 6.8, the split data analyzed with both Fourier and Rayleigh methods display smoother variations along right ascension, but show shifted dipole phases compared to results obtained with the derivative East-West method. Detailed values obtained from the Rayleigh analysis are provided in Table 6.3, corresponding to the plots shown in Figures 6.10, 6.11, and 6.12.

To conclude, neither the East–West derivative method nor the optimized Rayleigh analysis can fully resolve the time–variation issues observed in the reconstructed dipoles across the three considered energy bins. Firstly, this may be because the time gaps continue to introduce artificial effects and uncertainties in exposure. In particular, if missing or removed sidereal days, as well as the non-uniform detector acceptance, are distributed unevenly throughout the year, they can create a seasonal imbalance that also leaks into the sidereal and anti-sidereal frames, shifting the residual dipole. Besides, we suppose the data filtering does not perfectly remove solar diurnal or seasonal modulations. In that case, some of that power can also leak into the sidereal and anti-sidereal frames, introducing long-term systematics that were previously averaged out over time [320]. Therefore, rather than discarding entire days with long gaps, it is preferable to assign weights to each sidereal day based on the background, which has an effective time-dependent acceptance, and reduce these biases without losing valuable exposure. A more precise approach is therefore required, one that balances limited statistics with data stability.

6.3 Analysis with Angular-Vector Method

To achieve a higher level of precision in the reconstruction of potential anisotropies—and to further suppress the systematic effects that particularly affect dipole measurements, as discussed in Section 6.2, we perform an analysis using declination-dependent angular vectors and self-defined coordinate angular vectors within a full 3D dipole framework. This method reconstructs cosmic-ray dipole anisotropies, which is particularly effective in low-statistics scenarios. It analyzes the angular distribution of vectors and propagates uncertainties from vector components to the resultant dipole. The process begins by summing individual vectors on a celestial belt to obtain a resultant vector representing the pseudo-dipole moment. Two cases are considered: the first involves declination belts perpendicular to the z-axis, which are used to extract the dipole phase (right ascension) and the combined dipole amplitude of the vertical and horizontal components of the belts. The second case involves solid angle belts that are perpendicular to the y-axis, which incorporates both the vertical and horizontal components, but within a different geometric framework. These belts can be well modelled under the assumption of a pure dipole, which

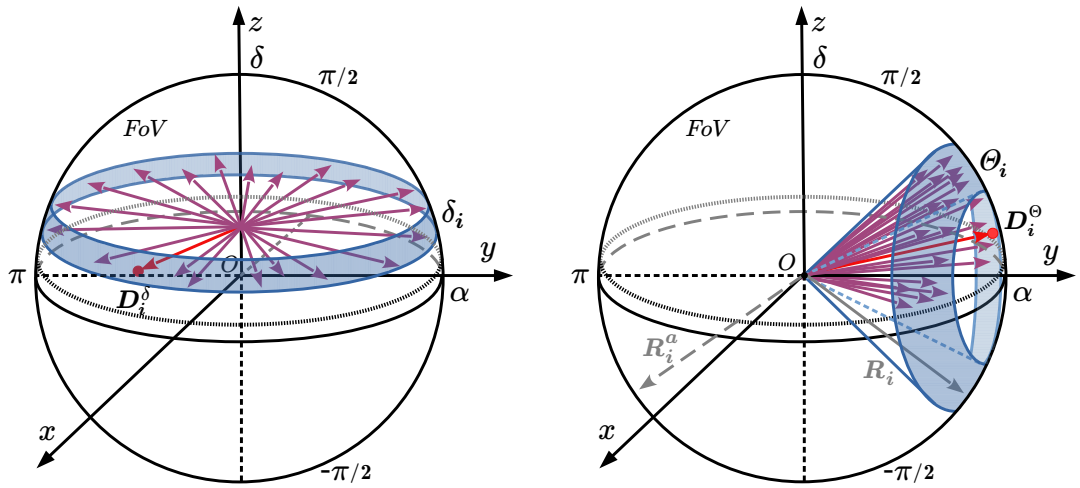


Figure 6.13: Declination bands in the angular-vector method: The left plot represents the δ-belt approach, while the right plot represents the Θ_Ω-belt approach. The purple vectors in both cases represent the data, while the red vectors, $\mathcal{D}_i^\delta$ and $\mathcal{D}_i^\alpha$, represent the resultant vectors obtained from the averaged data vectors.

yields an integrated dipole distribution to serve as the theoretical model and is directly comparable to the summed data vectors.

In both cases, we include the reference (background) from time-scrambled data and apply acceptance weights to both the data and reference vectors, strengthening the true signal while reducing natural and artificial noise. By comparing these two integrals, the method uncovers the relationship between the three-dimensional dipole amplitude and the declination, allowing for an accurate and self-consistent determination of both quantities. It is important to emphasize that the standard Rayleigh analysis [131, 321] corresponds to a special case of our method in which no weights are applied to the vectors and the reference (background) vectors are not subtracted, and it is therefore unsuitable for precise reconstructions for anisotropy studies.

For the first case, namely, the δ-belt shown in the left panel of Figure 6.13, we project the three-dimensional dipole distribution onto each declination belt and integrate it over the full right-ascension range, α from 0 to 2π. These δ-belts lie within the FoV of an air-shower observatory such as KASCADE–Grande, and, for simplicity, their acceptance is normalized through the application of the time-scrambling procedure. The method starts by analyzing 2D vectors within a declination ring, $\mathbf{u} = (\mathrm{u}_x, \mathrm{u}_y)$, where the dipole direction and amplitude are encoded in the average resultant vector. This approach considers two cases: one using only pure 2D vectors, and the other performs dipole reconstruction using those 2D vectors together with each arrival direction's individual z components, though the combination is not treated as a full 3D vector. The analysis focuses on unit vectors within δ-belts, where the dipole excess is extracted by summing and averaging vectors and performing self-cancellation across the full plane. When the entire FoV covers the full sky and is treated as a single belt, with a large event count and no weights or background subtraction, the result aligns with the standard Rayleigh analysis [131]. In both cases, the dipole excess is expressed in terms of vertical and parallel dipole components. The relationship between $\mathcal{A}$ and δ_d is then derived by equating the data average to the average of the integral dipole distribution function. In addition, the dipole phase α_d.

Since the true dipole orientation is not necessarily confined to the equatorial plane, estimates of the dipole phase and declination can become biased at high declinations due to the limited FoV and the predominance of equatorial dipole components in the measurement. Therefore, as an important cross-check, we perform a second reconstruction using a different class of belts, denoted by Θ_Ω-belts, defined in our custom coordinate system (see the right-hand plot in Figure 6.13). In this configuration, the Θ_Ω-belt examines dipole distributions within solid angle belts that are oriented perpendicular to the y-axis on the celestial sphere, utilizing spherical coordinates. Each such belt is constructed by selecting a specific angular interval Θ_Ω, around the y-axis, forming rings and belts on the celestial sphere with corresponding 2D vectors in the xz-plane. The FoV can be divided into two or more belts for analysis, depending on the available statistics and the existence of higher-order harmonics. As Θ_Ω varies, a narrow ring extends into a belt, enabling the analysis of vectors within specific angular regions. To reconstruct the dipole amplitude and orientation on a given Θ_Ω-belt, we evaluate the two-dimensional vectors in the xz-plane alongside the individual y-component of each event's arrival direction, which includes the three-dimensional dipole information within that band. The vector intensity on the belt is calculated as the deviation from a corresponding reference belt, derived from time-scrambled data and located symmetrically opposite the Θ_Ω-belt. The resultant vector on the belt is obtained by averaging the vectors and subtracting the summed reference vectors, incorporating weight as before, which accounts for acceptance variations. The magnitude of each vector is corrected for geometric projection effects at different Θ_Ω along a belt. The summation of vectors is mathematically equivalent to integrating a dipole, representing an integral component that forms an equation. The equations derived from different belts are used to determine the dipole amplitude and orientation, ultimately leading to the recovery of the global dipole. In the following, the method will be introduced in detail and then applied to the selected KASCADE-Grande dataset.

6.3.1 Vectors on Declination Belts

Now, we approach the problem from a theoretical standpoint to determine the components of the dipole amplitude for the averaged relative intensity on both the δ-ring and belt. The 2D vectors within a δ-ring, denoted by $\mathbf{u} = (\mathrm{u}_x, \mathrm{u}_y)$, follows a dipole distribution. Although the observatory's acceptance rate is not uniform across the δ range within the δ-ring, specifically within a small δ interval ring, this non-uniformity in acceptance primarily influences the uncertainty and precision associated with dipole estimation, rather than altering the fundamental characteristics or main features of the dipole itself. Each δ-ring yields the same dipole phase, and its amplitude is independent of the ring's declination coordinate, varying as $\sin\delta$. Moreover, the discussions regard the symmetry of the dipole's excess and deficit as equivalent to fitting only the first harmonic for the dipole, thereby allowing the small-scale anisotropies to be disregarded. When taking into account the spherical geometry associated with equatorial coordinates, the distribution of vectors across the entire celestial sphere follows an intensity function, which is represented as:

$$\mathcal{I}(\alpha, \delta) = 1 + \mathcal{A} \sin\delta_\mathrm{d} \sin\delta + \mathcal{A} \cos\delta_\mathrm{d} \cos\delta \cos(\alpha - \alpha_\mathrm{d}), \tag{6.18}$$

where δ_d and α_d represent the dipole's declination and right ascension, respectively, while $\mathcal{A}$ denotes the dipole amplitude. A detailed discussion on this is given in Chapter 5. The dipole amplitude of the α-component (which is projected onto the equatorial plane, specifically the xy-plane) with maximum field of view is denoted as $\mathcal{A}_\alpha = \mathcal{A} \cos\delta_\mathrm{d}$, whereas

the δ-component (which is projected onto the z-axis) is given by $\mathcal{A}_\delta = \mathcal{A}\sin\delta_{\rm d}$. In the scenario where the δ interval is finite, the δ-ring becomes a δ-belt.

We now begin the reconstruction using the δ-belts, where each event is represented by a vector corresponding to its cosmic-ray arrival direction. In the three-dimensional dipole reconstruction, the 2D vectors in the equatorial plane are combined with the z-component of each arrival direction, allowing the dipole amplitude and phase to be recovered using the information contained within a single belt. To carry out this procedure, we first divide the full FoV into a series of δ-belts, each containing 2D vectors ($\mathbf{u}$). Focusing on one single belt, if its δ interval approaches zero, it becomes a δ-ring, labelled as δ_i. The coordinates of a vector positioned on a given δ-ring are represented by right ascensions (α) as $\mathrm{u}_x = \cos\alpha$ and $\mathrm{u}_y = \sin\alpha$, forming a unit vector. The unit vectors ($\mathbf{u}_{ij}$) on a given δ-ring are anisotropically distributed in a two-dimensional plane and can be expressed in the $(\mathrm{u}_x, \mathrm{u}_y)$ coordinate system. These unit vectors ($\mathcal{N}_i$ in total) are equivalent to isotropically distributed non-unit vectors, denoted by $\mathbf{u}^*_{ij}$. The magnitude of each $\mathbf{u}^*_{ij}$ represents the event density at the point j. Worthing to note that $\mathcal{N}_i \neq \mathcal{N}^*_i$, because the sum of $\mathbf{u}_{ij}$ and $\mathbf{u}^*_{ij}$ are equal, but their counts are not.

Before introducing our main reconstruction approach, we first revisit the traditional dipole reconstruction method (see Chapter 3) as applied to belt-based vectors. To accurately represent relative intensity (RI), denoted as $\delta I_{ij} = I_{ij} - 1$ at any position on a δ-ring or belt. It is essential to account for sidereal variations in event counts in both the data and the reference (background), which may be influenced by various factors such as human activity and atmospheric effects, as previously mentioned. A reliable method for addressing this is to scramble the local sidereal time of arrival events to create a background for comparison with the data. As described in Eq. (6.12), scrambling the local sidereal time is equivalent to scrambling the right ascension (RA) coordinates within the same declination ring as the original arrival directions, denoted by $\tilde{\alpha}$, with each declination band independently normalized. Moreover, the horizontal RI (δI_{ij}) at the j-th point on the i-th δ-ring can be estimated by the deviation of the corresponding equivalent data vector at this point from the reference vectors $\mathbf{r}_{ij}$ at the same location on a δ-ring with $\mathcal{N}^{\rm r}_i$ equally spaced non-unit vectors $\mathbf{r}^*_{ij}$, where the reference δ-ring or belt indicates a null hypothesis, which corresponds to the magnitude of the average vector on the belt, assuming an isotropic distributed dipole. It is important to note that the total number of data vectors $\mathbf{u}^*_{ij}$ on the i-th ring or belt is equal to that of the corresponding reference ring or belt in the time-scrambling case, i.e., $\mathcal{N}^*_i = \mathcal{N}^{\rm r}_i$, but not in the isotropic background case, where $\mathcal{N}^*_i \neq \mathcal{N}^{\rm r}_i$ (as will be discussed in detail later). For consistency across different background constructions, we may choose to enforce $\mathcal{N}^*_i = \mathcal{N}^{\rm r}_i$ and calibrate the magnitude of each reference vector accordingly. Specifically, at the j-th point, the relative intensity is defined traditionally:

$$\delta I_{ij} = \frac{|\mathbf{u}^*_{ij}| - |\mathbf{r}^*_{ij}|}{|\mathbf{r}^*_{ij}|} \simeq |\mathbf{u}^*_{ij}| - |\mathbf{r}^*_{ij}|, \tag{6.19}$$

where $|\mathbf{r}^*_{ij}| = 1$ in the case of an isotropic background, but $|\mathbf{r}^*_{ij}| \neq 1$ when using time-scrambling. Eq. (6.19) is valid for both cases: reference belts with an isotropic distribution and those constructed using time-scrambling, since the denominator $|\mathbf{r}^*_{ij}|$ remains close to 1 in the time-scrambling case, with variations much smaller than those of δI_{ij}. The numerator, $(|\mathbf{u}^*_{ij}| - |\mathbf{r}^*_{ij}|)$, represents the relative difference between a data vector and its corresponding reference vector, typically on the order of $10^{-3} \sim 10^{-2}$ mostly, and reflects the specific anisotropy at the position of the data vector. The average of δI_{ij} over a δ-ring isolates only the vertical component of the dipole, as the equatorial component cancels out

under the assumption of a pure dipole. This is equivalent to the continuous average of the dipole's intensity function $\mathcal{I}(\alpha, \delta)$, as given in Eq. (6.18), over the equatorial plane within the full range $\alpha \in (0, 2\pi)$ and a specified declination interval $(\delta_{\rm L}, \delta_{\rm U})$, i.e., from the lower to upper bounds of the field of view for a given δ-belt. Consequently, this approach does not permit reconstruction of the local dipole on the ring through simple averaging, but instead requires fitting the first harmonic. However, from a data-driven perspective, this method becomes less reliable when statistical precision is limited.

Alternatively, and differing from the standard definition in Eq. (6.19), we can analyze the difference in magnitude between the average of the non-unit data vectors $\mathbf{u}^*_{ij}$ and the average of the corresponding reference vectors $\mathbf{r}^*_{ij}$, where j runs from 1 to $\mathcal{N}^*_i$, representing the residual vectors in the dipole excess region. Fortunately, each 2D data vector within a δ-ring or belt deviates (local imbalance) from its corresponding reference vector, which is not cancelled out but instead individually compared. The resultant vectors, denoted by $\boldsymbol{I}_i$, effectively counterbalance each other across the entire plane over a full 2π rotation. This yields an alternating pattern of excess and deficit along the ring, and after summation over the full 2π range, only the dipole excess remains, as the deficit components cancel out. Moreover, the resultant vector consistently points in the same averaged principal direction of the dipole. Equivalently, under the assumption of an isotropic distribution of non-unit vectors, one has $\mathbf{u}^*_{ij} = n^{\rm d}_{ij}\mathbf{u}^0_{ij}$ and $\mathbf{r}^*_{ij} = n^{\rm r}_{ij}\mathbf{u}^0_{ij}$, where $n^{\rm d}_{ij}$ represents the total number density of data vectors at j-th equally spaced position of the i-th ring or belt. The number density or magnitude of vectors follows the relations

$$n^{\rm d}_{ij} = n^{\rm sig}_{ij} + n^{\rm m}_{ij}, \quad n^{\rm r}_{ij} = n^0_{ij} + n^{\rm m}_{ij}, \tag{6.20}$$

where $n^0_{ij} = 1$ contributes the absolute monopole part in the reference vectors. This number density only weights the magnitudes of each vector. Therefore, one has

$$\mathbf{u}^{\rm sig}_{ij} = n^{\rm sig}_{ij}\,\mathbf{u}^0_{ij} = (n^{\rm d}_{ij} - n^{\rm r}_{ij} + n^0_{ij})\,\mathbf{u}^0_{ij} \tag{6.21}$$

where $\mathbf{u}^0_{ij}$ represents the equal spaced unit vectors. Clearly, Eq. (6.21) yields an intensity vector, which we denote by $\boldsymbol{I}_{ij}$, such that $\boldsymbol{I}_{ij} = \mathbf{u}^{\rm sig}_{ij}$, which can also be defined as

$$\boldsymbol{I}_{ij} = \mathbf{u}^*_{ij} - \mathbf{r}^*_{ij} + \mathbf{r}^0_{ij}, \tag{6.22}$$

where $\mathbf{u}^0_{ij} = \mathbf{r}^0_{ij}$. The reference vectors under the assumption of isotropic distribution of non-unit vectors can be expressed as $\mathbf{r}^*_{ij} = \mathbf{r}^0_{ij} + \mathbf{r}^{\rm m}_{ij}$, which is equivalent to the anisotropic distributed non-unit vectors $\mathbf{r}_{ij}$ from the scrambled data vectors; $\mathbf{r}^0_{ij}$ represents the pure isotropic distributed vectors, namely, monopole, on the reference ring, while $\mathbf{r}^{\rm m}_{ij}$ is the anisotropic part (multipole) in the reference vectors, namely, the residue multipoles without monopole and dipole. The data vectors $\mathbf{u}^*_{ij}$ share the same principle as $\mathbf{r}^*_{ij}$, which makes $\mathbf{u}_{ij}$ and $\mathbf{r}_{ij}$ are comparable individually at the micro level. The corresponding average of $\boldsymbol{I}_{ij}$, with its amplitude denoted by $\hat{I}_i$, is expressed as

$$\begin{aligned} \hat{I}_i &= \left| \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \mathbf{u}^{\rm sig}_{ij} \right| = \left| \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} n^{\rm d}_{ij}\,\mathbf{u}^0_{ij} - \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} n^{r}_{ij}\,\mathbf{u}^0_{ij} \right| \\ &= \left| \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \mathbf{u}_{ij} - \frac{1}{\mathcal{N}^{\rm r}_i} \sum_{k=1}^{\mathcal{N}^{\rm r}_i} \mathbf{r}_{ik} \right|, \end{aligned} \tag{6.23}$$

where the sum of $\mathbf{r}^0_{ij} = n^0_{ij}\mathbf{u}^0_{ij}$ cancelled out, due to its circular symmetry. The sum of $\mathbf{u}^*_{ij}$ equals the sum of $\mathbf{u}_{ij}$, as well as the sum of $\mathbf{r}^*_{ij}$ and $\mathbf{r}_{ij}$. When the number of time-

scrambling iterations $n_{\rm scr} > 1$, the total number of reference vectors becomes $\mathcal{N}_i^{\rm r} = n_{\rm scr}\mathcal{N}_i$, while the average of $\mathbf{r}_{ij}$ remains effective and yields improved statistical stability. To reduce fluctuations in the reference vectors, we choose $n_{\rm scr} = 20$ for the scrambling.

We perform exposure and acceptance calibration on the data before averaging the vectors and applying time-scrambling to generate reference vectors, unlike papers [278, 89]. Omitting this step can increase the contribution of higher-order multipole components ($\mathbf{r}_{ij}^{\rm m}$), primarily due to time gaps and variations in exposure. Equivalently, the local number density $n_{ij}^{\rm m}$ increases, which may lead to stronger signal cancellation and reduced precision in estimating the dipole amplitudes and orientations. Proper calibration, therefore, increases the effective contribution of signal-relevant vectors and reduces the uncertainty in $\hat{\mathcal{I}}$ and dipole reconstruction. The amplitude of the calibrated resultant vector is

$$\hat{I}_i = \left| \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \mathrm{w}_j \, \mathbf{u}_{ij} - \frac{1}{\mathcal{N}_i^{\rm r}} \sum_{k=1}^{\mathcal{N}_i^{\rm r}} \mathrm{w}_k \, \mathbf{r}_{ik} \right|, \tag{6.24}$$

where w_k denotes the weight w_j repeated across the $n_{\rm scr}$ time-scramblings of the data. The resulting weighted reference vectors are identical whether one applies the weights before time scrambling or applies time scrambling before weighting. The weight w_{ij} is defined as the inverse of the efficiencies in data, namely

$$\mathrm{w}_j = \left[\epsilon_{\rm T}(\tau_j) \, \cos\theta_j \, \eta(\tau_j)\right]^{-1}, \tag{6.25}$$

where $\epsilon_{\rm T}(\tau_j) = n_\ell(\tau)/\langle n_\ell(\tau)\rangle$ denotes the time efficiency, defined as the ratio of the effective number of sidereal bins ($n_{\rm T}$), accumulated over all sidereal days in the data, to the average number within the full 0–24 sidereal hour range. In this analysis, the 24 hours are divided into $\ell = 360$ time bins. The quantity $\eta(\tau_j) = N_\tau^{\rm scr}/\langle N_\tau^{\rm scr}\rangle$ represents the acceptance efficiency at each sidereal time bin, derived from the local sidereal time (LST) histogram obtained by applying $n_{\rm scr}$ time-scrambling to the data. Note that the effect of $\epsilon_{\rm T}$ is absorbed into η. Since the original time stamps are preserved during scrambling, all detector live-time gaps remain encoded in the data. Meanwhile, any true sky anisotropy is effectively smeared out due to the randomization of right ascension, making $N_\tau^{\rm scr}$ an unbiased estimator of the detector's relative live-time in each LST bin ℓ. The effective event rate becomes approximately uniform in LST, so that any first-harmonic signal extracted can be attributed solely to the sky.

For convenience in the subsequent derivations, we introduce the symbol $\mathfrak{U}_i \equiv \hat{I}_i$ for all subsequent calculations. The amplitude of the resultant vector, denoted by either $\hat{\boldsymbol{I}}$ or, equivalently, $\boldsymbol{\mathfrak{U}}$) is then expressed in the usual form $\mathfrak{U} = \sqrt{\mathfrak{U}_x^2 + \mathfrak{U}_y^2}$. Since the covariance between the two components is zero, $\mathrm{Cov}(\mathfrak{U}_x, \mathfrak{U}_y) = 0$, the variance of the amplitude $\mathfrak{U}$ is approximated by $\mathrm{Var}(\mathfrak{U}) \approx (\partial_x\mathfrak{U})^2\mathrm{Var}(\mathfrak{U}_x) + (\partial_y\mathfrak{U})^2\mathrm{Var}(\mathfrak{U}_y)$, with $\mathrm{Var}(\mathfrak{U}_x) = \mathrm{Var}(\mathfrak{U}_y)$ and $\mathrm{Var}(\mathfrak{U}_x) = (W/\mathcal{N}^2 + W^{\rm r}/\mathcal{N}^{\rm r,2})\mathrm{Var}(\cos\varphi)$, where $W = \sum_i \mathrm{w}_i^2$, $W^{\rm r} = \sum_k \mathrm{w}_k^2 = n_{\rm scr}W$, and $\mathrm{Var}(\cos\varphi) = \mathrm{Var}(\sin\varphi) = \langle\cos^2\varphi\rangle - \langle\cos\varphi\rangle^2 = 1/2$ for φ ranging from 0 to 2π. Based on Eq. (6.24) and the standard propagation of errors, the uncertainty of $\mathfrak{U}_i$ is then given by

$$\sigma_{\mathfrak{U},i} = \sqrt{\frac{W}{2\,\mathcal{N}_i^2}\left(1 + \frac{1}{n_{\rm scr}}\right)}. \tag{6.26}$$

As the scramble times $n_{\rm scr}$ increase, $\sigma_{\mathfrak{U}}$ decreases slightly. The above uncertainty expression is similar to Eq. (6.15) in the scrambled Rayleigh analysis, i.e., equivalent to a 3D version of the Rayleigh analysis, and differs only by a factor of $2\sqrt{W/\mathcal{N}_i}$. The tail probability of the Rayleigh distribution, i.e., the probability of obtaining an amplitude equal to or larger

than the observed amplitude $\mathfrak{U}_i$ under the null hypothesis of isotropy. Inserting Eq. (6.26) into Eq. (6.16), the tail probability is then obtained,

$$\mathcal{P}(\geq \mathfrak{U}_i) = \exp\left(-\frac{\mathcal{N}_i^2\,\mathfrak{U}_i^2}{W(1+1/n_{\rm scr})}\right), \tag{6.27}$$

which is similar to the form obtained in the Rayleigh analysis, as shown earlier in Eq. (6.17). Once two or more δ-belts are available, the individual phase estimates $\alpha_{\rm d}^i$ extracted from each belt can be combined and fitted to yield a single best estimate of the dipole phase, since under the assumption of an ideal, purely dipolar sky distribution, all belts should in principle return identical phase values. For the case involving higher harmonics beyond the dipole, one has $\mathrm{u}_x^n = \sin n\alpha$ and $\mathrm{u}_y^n = \cos n\alpha$. The resultant amplitude is given by $\mathfrak{U}_n = \sqrt{(\mathfrak{U}_x^n)^2 + (\mathfrak{U}_z^n)^2}$, and the corresponding phase by $\varphi_{\rm n} = \tan^{-1}(\mathfrak{U}_x^n/\mathfrak{U}_z^n)/n$, with $n = 1$ yields the dipolar component in the harmonics. The uncertainty of $\sigma_{\mathfrak{U}}$ retains the same form as in the dipole case, while the uncertainty in the phase is given by $\sigma_{\alpha_n} = \sigma_{\mathfrak{U}}/n\mathfrak{U}$.

Up to now, the relative intensity has been defined based on vectors in the declination plane (and belts) after subtracting the scrambled data as a reference. The next step, the reconstruction of the dipole within the declination belts, will be discussed.

6.3.2 Discrete Equivalence on Declination Belts

To achieve high precision when the total number of events in the FoV is small (e.g., $\mathcal{N} < 10^7$) and the relative intensities are at the $10^{-3} \sim 10^{-2}$ level, the average can be computed using a discrete series of points from the pure dipole excess. The $\mathcal{N}_\vartheta$ points are equally spaced within the interval $[-\pi/2, \pi/2]$ on the δ_i-ring, with angles denoted by ϑ_k. A more convenient form for ϑ_k (including both endpoints) is

$$\vartheta_k = -\frac{\pi}{2} + \frac{k\pi}{\mathcal{N}_\vartheta - 1}, \quad k = 0, 1, 2, ..., \mathcal{N}_\vartheta - 1. \tag{6.28}$$

At each ϑ_k, we evaluate the pure dipole excess scaled by the vertical component ($\cos\vartheta_k$) of the vectors, which contributes solely to the summation, while the horizontal components cancel out. The summation over the entire ϑ range is

$$\delta\hat{\mathcal{I}}_i(\delta) = \frac{1}{\mathcal{N}_\vartheta}\sum_{k=0}^{\mathcal{N}_\vartheta-1}\mathcal{A}_\delta\cos^2\vartheta_k = \frac{1}{\mathcal{N}_\vartheta}\sum_{k=0}^{\mathcal{N}_\vartheta-1}\mathcal{A}_\delta\left[\frac{1}{2} + \frac{1}{2}\cos\left(2\vartheta_k\right)\right], \tag{6.29}$$

where $\mathcal{A}_\delta = \mathcal{A}\cos\delta_{\rm d}\cos\delta$ is the local dipole amplitude on the δ-ring. When $k = 0$, or $k = \mathcal{N}_\vartheta - 1$, $\sin\vartheta = 0$, so the endpoints do not contribute to the overall dipole sum. Inserting Eq.(6.28) into Eq.(6.29), one has

$$\delta\hat{\mathcal{I}}_i^\dagger(\delta_i) = \frac{\mathcal{A}_\delta}{2} + \frac{\mathcal{A}_\delta}{2\mathcal{N}_\vartheta}\sum_{k=0}^{\mathcal{N}_\vartheta-1}\cos\left(-\pi + \frac{2k\pi}{\mathcal{N}_\vartheta - 1}\right). \tag{6.30}$$

Use the known identity of the finite sum of cosines in arithmetic progression, namely,

$$\sum_{k=0}^{n-1}\cos\left(\alpha + k\beta\right) = \sin\left(n\beta/2\right)\sin^{-1}\left(\beta/2\right)\cos\left(\alpha + (n-1)\beta/2\right), \tag{6.31}$$

in our case, $n = \mathcal{N}_\vartheta$, $\alpha = -\pi$, $\beta = 2\pi/(\mathcal{N}_\vartheta - 1)$. Thus, the sum of cosines in Eq. (6.30) is

$$\sum_{k=0}^{\mathcal{N}_\vartheta-1}\cos\left(-\pi + \frac{2k\pi}{\mathcal{N}_\vartheta - 1}\right) = \sin\left(\frac{\mathcal{N}_\vartheta\pi}{\mathcal{N}_\vartheta - 1}\right)\sin^{-1}\left(\frac{\pi}{\mathcal{N}_\vartheta - 1}\right) = -1, \tag{6.32}$$

for all integers $\mathcal{N}_\vartheta \geq 3$, and $\mathcal{N}_\vartheta\pi/(\mathcal{N}_\vartheta - 1)$ is actually $\pi + \pi/(\mathcal{N}_\vartheta - 1)$. Therefore, we get the precise expression of $\partial_\delta \delta\hat{\mathcal{I}}_i(\delta)$, an equatorial relative intensity at δ, follows the form

$$\delta\hat{\mathcal{I}}_i^\dagger(\delta) = \frac{(\mathcal{N}_\vartheta - 1)\mathcal{A}_\delta}{2\mathcal{N}_\vartheta}, \tag{6.33}$$

where $\mathcal{N}_\vartheta = \mathcal{N}_i/2$, as mentioned. Recalling the average of the dipole excess from data, the local pseudo-dipole amplitude on the ring is obtained as

$$\mathcal{A}_\delta^\dagger = \frac{2}{\mathcal{N}_i - 2}|\mathbf{u}_i^*|_\mathrm{r}, \quad \hat{\mathcal{A}}_\delta^\dagger = \frac{2}{\mathcal{N}_i - 2}|\mathbf{u}_i^*|, \tag{6.34}$$

where $\hat{\mathcal{A}}_\delta^\dagger$ denotes the averaged local dipole amplitude evaluated over the full δ-interval of the corresponding δ-belt, and $\mathcal{N}_i$ remains the unchanged form for simplicity. Because $\hat{\mathcal{A}}_\delta^\dagger$ varies only weakly from one belt to another, its belt-to-belt deviation can safely be neglected, especially given that the vertical dipole component contributes only at the level of $10^{-2} - 10^{-3}$. As $\mathcal{N}_i$ becomes very large, one has $\delta\hat{\mathcal{I}}_i(\delta) = \mathcal{A}_\delta^\dagger/2$ and $\mathcal{A}_\delta^\dagger \approx 2|\mathbf{u}_i^*|_\mathrm{r}/\mathcal{N}_i$, a result that can be explicitly verified by considering the continuous integral representation of the dipole excess over the angular range ϑ. Note that within any given declination belt, the operators satisfy $\hat{\mathcal{A}}_\delta^\dagger \neq \hat{\mathcal{A}}_\delta$due to the contribution of the vertical dipole component. However, when averaged over the full sky, $\hat{\mathcal{A}}_\delta^\dagger = \hat{\mathcal{A}}_\delta$, since the vertical term cancels out.

6.3.3 Reconstruction with Declination Belts

We are now focusing on the approximate continuous case. Since the difference between the average of the data vectors and the average of the reference vectors yields the mean of the $\cos\alpha$ and $\sin\alpha$ components of the intensity vector $\boldsymbol{I}$, the amplitude of the resultant vector, $\hat{I}$, serves as a pseudo-average of the intensity, as shown in Eq. (6.24). This resultant vector always points toward $\alpha = \alpha_\mathrm{d}$ within the accessible region. In the continuous case, assuming a pure dipolar distribution, the two-dimensional intensity vector of the data is denoted by $\boldsymbol{\mathcal{I}} = (\mathcal{I}_x, \mathcal{I}_y)$, which is equivalent to the intensity vector defined in Eq. (6.22). The x and y components are given by $\mathcal{I}_x = \mathcal{I}\cos\alpha$ and $\mathcal{I}_y = \mathcal{I}\sin\alpha$, where $\mathcal{I}(\alpha, \delta)$ denotes the data intensity distribution, which follows the dipole form defined in Eq. (6.18). In addition, the vertical component of the dipole is captured by $\mathcal{I}_z = \mathcal{I}\cos\delta$, which will be used in subsequent averaging procedures. Specifically, under partial-sky coverage with $\alpha \in (0, 2\pi)$ and $\delta \in (\delta_\mathrm{L}, \delta_\mathrm{U})$, the integrals of $\mathcal{I}_x$, $\mathcal{I}_y$, and $\mathcal{I}_z$ are expressed as

$$\begin{aligned}
\hat{\mathcal{I}}_x &= \int_{\delta_\mathrm{L}}^{\delta_\mathrm{U}} \int_0^{2\pi} \mathcal{I}(\alpha, \delta)\cos\alpha\cos\delta\,\mathrm{d}\alpha\,\mathrm{d}\delta = \pi\mathcal{A}_\mathrm{p} P\cos\alpha_\mathrm{d}, \\
\hat{\mathcal{I}}_y &= \int_{\delta_\mathrm{L}}^{\delta_\mathrm{U}} \int_0^{2\pi} \mathcal{I}(\alpha, \delta)\sin\alpha\cos\delta\,\mathrm{d}\alpha\,\mathrm{d}\delta = \pi\mathcal{A}_\mathrm{p} P\sin\alpha_\mathrm{d}, \\
\hat{\mathcal{I}}_z &= \int_{\delta_\mathrm{L}}^{\delta_\mathrm{U}} \int_0^{2\pi} \mathcal{I}(\alpha, \delta)\sin\delta\cos\delta\,\mathrm{d}\alpha\,\mathrm{d}\delta = 2\pi(S_1 + \mathcal{A}_\mathrm{s} S_2),
\end{aligned} \tag{6.35}$$

where the dipole shift term is defined as $\mathcal{A}_\mathrm{s} = \mathcal{A}\sin\delta_\mathrm{d}$, and the equatorial dipole component is $\mathcal{A}_\mathrm{p} = \mathcal{A}\cos\delta_\mathrm{d}$. The constants S_1, S_2, and P are given by:

$$S_1 = \int_{\delta_L}^{\delta_U} \sin\delta\cos\delta\,\mathrm{d}\delta, \quad S_2 = \int_{\delta_L}^{\delta_U} \sin^2\delta\cos\delta\,\mathrm{d}\delta, \quad P = \int_{\delta_L}^{\delta_U} \cos^2\delta\,d\delta. \tag{6.36}$$

The averages of $\hat{\mathcal{I}}_x$, $\hat{\mathcal{I}}_y$, and $\hat{\mathcal{I}}_z$ are obtained by normalizing each component by the integral of the intensity distribution $\mathcal{I}(\alpha, \delta)$ over the entire field of view. This integral yields a

monopole term, as it effectively averages out the dipole anisotropy. Namely,

$$\hat{\mathcal{I}}_0 = \int_{\delta_\mathrm{L}}^{\delta_\mathrm{U}} \int_0^{2\pi} \mathcal{I}(\delta,\alpha)\cos\delta\,\mathrm{d}\alpha\,\mathrm{d}\delta = 2\pi(C_0 + \mathcal{A}_\mathrm{s} S_1), \tag{6.37}$$

where $C_0 = \int_{\delta_L}^{\delta_U} \cos\delta\,\mathrm{d}\delta$. Analogous to the vector averaging procedure described in Eq. (6.24) for the data, the average values of $\mathcal{I}_x$, $\mathcal{I}_y$, and $\mathcal{I}_z$ within a given δ-belt are approximated by $\hat{\mathcal{I}}_x$, $\hat{\mathcal{I}}_y$, and $\hat{\mathcal{I}}_z$, respectively, each normalized by the integral of $\mathcal{I}(\alpha,\delta)$ over the field of view. This normalization yields the idealized vector average for a pure 3D dipole distribution. The resulting average components are given by

$$\frac{\hat{\mathcal{I}}_x}{\hat{\mathcal{I}}_0} = \frac{\mathcal{A}_\mathrm{p}\, p\cos\alpha_\mathrm{d}}{2(1+\mathcal{A}_\mathrm{s}\, s)}, \quad \frac{\hat{\mathcal{I}}_y}{\hat{\mathcal{I}}_0} = \frac{\mathcal{A}_\mathrm{p}\, p\sin\alpha_\mathrm{d}}{2(1+\mathcal{A}_\mathrm{s}\, s)}, \quad \frac{\hat{\mathcal{I}}_z}{\hat{\mathcal{I}}_0} = \frac{s+\mathcal{A}_\mathrm{s}\, t}{1+\mathcal{A}_\mathrm{s}\, s}, \tag{6.38}$$

where we define the parameters p, s and t as

$$p \equiv \frac{P}{C_0} = \langle\cos\delta\rangle, \quad s \equiv \frac{S_1}{C_0} = \langle\sin\delta\rangle, \quad t \equiv \frac{S_2}{C_0} = \langle\sin^2\delta\rangle, \tag{6.39}$$

where the quantities $\langle\sin\delta\rangle$, $\langle\cos\delta\rangle$ and $\langle\sin^2\delta\rangle$ represent the average over the entire FoV, weighted by the geometric factor $\cos\delta$, and are expressed as:

$$\begin{gathered}\langle\sin\delta\rangle \equiv \frac{\cos(2\delta_\mathrm{L}) - \cos(2\delta_\mathrm{U})}{4\,[\sin(\delta_\mathrm{U}) - \sin(\delta_\mathrm{L})]}, \quad \langle\cos\delta\rangle \equiv \frac{2(\delta_\mathrm{U}-\delta_\mathrm{L}) + \sin(2\delta_\mathrm{U}) - \sin(2\delta_\mathrm{L})}{4\,[\sin(\delta_\mathrm{U}) - \sin(\delta_\mathrm{L})]}, \\ \langle\sin^2\delta\rangle \equiv \frac{1}{3}\big(\sin^2\delta_\mathrm{U} + \sin^2\delta_\mathrm{L} + \sin\delta_\mathrm{U}\sin\delta_\mathrm{L}\big).\end{gathered} \tag{6.40}$$

When dealing with data within a limited FoV, it should be noted that the horizontal dipole component is always present and contributes to the observed modulation, whereas the vertical shift term becomes strongly suppressed because of the factor $\mathcal{A}_\mathrm{s}\, s \ll 1$ in Eq. (6.38). In the case where a single δ-belt spans the entire FoV of KASCADE-Grande, that is to say, $\delta_\mathrm{L} = 9.1°$ and $\delta_\mathrm{U} = 89.1°$, we can therefore obtain $\langle\sin\delta\rangle = 0.579$, $\langle\cos\delta\rangle = 0.746$ and $\langle\sin^2\delta\rangle = 0.394$.

To connect the theoretical model with the measured data, we define $\hat{\mathcal{I}}_i$ as the amplitude of the intensity vector $\boldsymbol{\mathcal{I}}$ averaged over the i-th δ-belt. Formally, it is given by,

$$\hat{\mathcal{I}}_i = \sqrt{\left(\frac{\hat{\mathcal{I}}_x}{\hat{\mathcal{I}}_0}\right)^2 + \left(\frac{\hat{\mathcal{I}}_y}{\hat{\mathcal{I}}_0}\right)^2} = \frac{\mathcal{A}_\mathrm{p}\, p}{2(1+\mathcal{A}_\mathrm{s}\, s)}, \tag{6.41}$$

which serves as the model. This serves as the model. It differs from the Rayleigh analysis by a factor of 1/2 because we do not employ the Fourier-expansion method that introduces a factor of 2 in the first-harmonic coefficients a and b. Those factors cancel out, and the two approaches are therefore equivalent. Therefore, Eq. (6.41) is comparable to analyses in the spherical polar coordinate system with zenith angle and azimuth, as mentioned in the papers [321, 287], where the dipole model is based on 3D direction vectors, whereas our formulation uses (α,δ) coordinates as shown in Eq. (6.18). Equivalently, we set the data amplitude $\hat{I}_i$, with $\hat{I}_i \equiv \mathfrak{U}_i = \hat{\mathcal{I}}_i$ as defined in Eq. (6.24), equal to the model amplitude $\hat{\mathcal{I}}_i$. Both the data and model amplitudes represent the sum of the dipole's horizontal component (the amplitude of the local dipole projected onto the right ascension plane) and the vertical shift term when considering partial-sky coverage of the entire FoV or a single δ-belt. The 3D dipole amplitude and phase at the i-th declination belt can be reconstructed. The dipole amplitude $\mathcal{A}_i$ can be obtained using Eq. (6.41), while the dipole phase $\alpha_{\mathrm{d},i}$ can be estimated from a δ-belt is given by the direction in which the resultant

vector $\mathfrak{U}_i$ points. The reconstructed amplitude $\mathcal{A}_i$ and phase $\alpha\mathrm{d}, i$ are obtained as follows:

$$\mathcal{A}_i = \frac{2\,\mathfrak{U}_i}{p_i \cos\delta_\delta - 2\,\mathfrak{U}_i\, s_i \sin\delta_\mathrm{d}}, \quad \alpha_{\mathrm{d},i} = \arctan\left(\frac{\mathfrak{U}_{y,i}}{\mathfrak{U}_{x,i}}\right), \tag{6.42}$$

where p_i and s_i can be obtained from Eq. (6.39). Using standard error propagation, the uncertainties on the reconstructed $\mathcal{A}_i$ and $\alpha_{\mathrm{d},i}$ are therefore given by

$$\sigma_{\mathcal{A},i} = \frac{2\,p_i \cos\delta_\mathrm{d}\,\sigma_{\mathfrak{U},i}}{(p_i \cos\delta_\mathrm{d} - 2\,\mathfrak{U}_i\, s_i \sin\delta_\mathrm{d})^2}, \quad \sigma_{\alpha_\mathrm{d},i} = \frac{\sigma_{\mathfrak{U},i}}{\mathfrak{U}_i}. \tag{6.43}$$

where $\sigma_{\mathfrak{U},i}$ is obtained from Eq. (6.26). The dipole phase and its error are comparable to those in Rayleigh analysis, as presented in Eq.(6.14).

If we have multiple δ-belts, we obtain a dipole phase $\alpha_{\mathrm{d},i}$ for each belt. These values of $\alpha_{\mathrm{d},i}$ can be combined using the weighted circular mean approach. Instead of fitting the angles directly, each angle is first converted into a 2D unit vector on the circle: $x_i = \cos\alpha^i_\mathrm{d}$ and $y_i = \sin\alpha^i_\mathrm{d}$, which correspond to the previously obtained $\mathfrak{U}_{x,i}$ and $\mathfrak{U}_{y,i}$, respectively. To compute the weighted average direction while accounting for uncertainties, we calculate the vector sum $\mathrm{U}_x = \sum \omega_i \mathfrak{U}_{x,i}$ and $\mathrm{U}_y = \sum \omega_i \mathfrak{U}_{y,i}$, where the weight is given by $\omega_i = 1/\sigma^i_{\alpha_\mathrm{d}}$, with $\sigma^i_{\alpha_\mathrm{d}}$ being the uncertainty of each dipole phase. The weighted circular mean of the dipole phase, along with the uncertainty on the mean (the circular standard error), is computed using the weighted resultant vector $\mathbf{U} = (\mathrm{U}_x, \mathrm{U}_y)$ as follows:

$$\alpha_\mathrm{d} = \mathrm{arctan2}\left(\frac{\mathrm{U}_y}{\mathrm{U}_x}\right), \quad \sigma_{\alpha_\mathrm{d}} = \sqrt{-\frac{2}{n_\mathrm{eff}} \ln\left(\frac{\mathrm{U}}{\sum \omega_i}\right)}, \tag{6.44}$$

where α_d is modulated within the 2π range, and $\mathrm{U} = \sqrt{\mathrm{U}_x^2 + \mathrm{U}_y^2}$ is the magnitude of the weighted resultant vector, $\sigma_{\alpha_\mathrm{d}}$is estimated as $\sigma_{\alpha_\mathrm{d}} \approx \sigma_\mathrm{cir}/\sqrt{n_\mathrm{eff}}$, where the common circular standard deviation is given by $\sigma_\mathrm{cir} = \sqrt{-2\ln(\mathrm{U}/\sum\omega_i)}$. The global tail probability ($\mathcal{P}$-value) of the vector distribution can be computed by combining each value from $\mathcal{P}$, i.e., (6.27), using Fisher's method. The combined test statistic is given by $X = -2\sum \ln \mathcal{P}_i$. which, under the null hypothesis, follows a chi-squared distribution with $2n_\mathrm{eff}$ degrees of freedom, where n_eff is the number of belts. The global $\mathcal{P}$-value from the angular-vector method is given by δ-belts is then $\mathcal{P}_\mathrm{global} = 1 - F_{\chi^2}(X; 2n_\mathrm{eff})$, where $F_{\chi^2}(X; 2n_\mathrm{eff})$ is the cumulative distribution function (CDF) of the chi-squared distribution.

For the reconstruction of $\mathcal{A}$ and δ_d, one may treat $\mathcal{A}_i$ and $\delta_{\mathrm{d},i}$ as common parameters and solve the full set of amplitude–declination equations, i.e., Eq. (6.42), simultaneously across all δ-belts. For the special case where the FoV is divided into just two δ-belts, the dipole declination and amplitude follow directly from the condition $\mathcal{A}_1 = \mathcal{A}_2$. One has

$$\delta_\mathrm{d} = \arctan\left(\frac{\mathfrak{U}_2\, p_1 - \mathfrak{U}_1\, p_2}{2\,\mathfrak{U}_1\mathfrak{U}_2(s_1 - s_2)}\right), \quad \mathcal{A} = \frac{2\,\mathfrak{U}_1}{p_1 \cos\delta_\mathrm{d} - 2\,\mathfrak{U}_1\, s_1 \sin\delta_\mathrm{d}}, \tag{6.45}$$

where the coefficients s_1, s_2, p_1 and p_2 depend on the specific FoVs (δ ranges) of the respective belts. Owing to the complexity of the calculations, the uncertainties in $\mathcal{A}$ and δ_d are evaluated numerically using computational methods.

Furthermore, as we obtain a set of Eq. (6.42) at different δ-belts, it becomes necessary to estimate the best-fit values of $\mathcal{A}$ and δ_d using a least-squares (LS) approach. Now, let us detail the process for estimating the parameters $\mathcal{A}$ and δ_d from the given dataset $(\mathfrak{U}_i, s_i, p_i)$, with associated uncertainties $\sigma_{\mathfrak{U}_i}$, using the LS fitting approach. The equation (6.41) is expressed with linearity in terms of $\mathfrak{U}_i$ and $\mathfrak{B}_i$ as follows:

$$\mathfrak{U}_i = \mathcal{A}\,\mathfrak{B}_\mathrm{i}, \quad \mathfrak{B}_i = p_i \cos\delta_\delta - \mathfrak{U}_i\, s_i \sin\delta_\mathrm{d}, \tag{6.46}$$

where $\mathfrak{B}_i$ is just the parts in the parenthesis of $\mathfrak{U}_i$. In matrix notation, this problem can be represented as $\mathbf{U} = \mathcal{A}\mathbf{B}$, where $\mathbf{U} = (\mathfrak{U}_1\, \mathfrak{U}_2\, \cdots \mathfrak{U}_n)^{\mathbf{T}}$ and $\mathbf{B} = (\mathfrak{B}_1\, \mathfrak{B}_2\, \cdots \mathfrak{B}_n)^{\mathbf{T}}$. To determine $\mathcal{A}$ and $\delta_{\rm d}$, we minimize the sum of squared residuals, which is given by:

$$\mathcal{S} = \sum_{i=1}^{n} \left(\frac{\mathfrak{U}_i - \mathcal{A}\mathfrak{B}_i}{\sigma_{\mathfrak{U}_i}} \right)^2, \tag{6.47}$$

which is the weighted sum of squared residuals, incorporating the inverse of the variance, $\sigma_{\mathfrak{U}_i}$, as given in Eq. (6.26). Since $\mathfrak{B}_i$ is a function of $\delta_{\rm d}$, the optimization problem becomes non-linear. The fitting process begins with an initial guess for $\mathcal{A}$ and $\delta_{\rm d}$, typically chosen based on simple approximations derived from sky maps of the datasets. Using a non-linear optimization algorithm, we fit the model to the data by minimizing the objective function, and the process adjusts $\mathcal{A}$ and $\delta_{\rm d}$ utilizing Eq. (6.47). By minimizing the objective function $\mathcal{S}$ and iteratively determining the best fit. Next, one can estimate their uncertainties using the covariance matrix derived from the fitting procedure. The covariance matrix, denoted by $\mathbf{C}$, contains information on the variances and covariances of the estimated parameters. Specifically, for the parameters $\mathcal{A}$ and $\delta_{\rm d}$, the covariance matrix is given by

$$\mathbf{C} = \begin{pmatrix} \sigma_A^2 & \sigma_{A\delta_{\rm d}} \\ \sigma_{A\delta_{\rm d}} & \sigma_{\delta_{\rm d}}^2 \end{pmatrix}, \tag{6.48}$$

where $\sigma_{\mathcal{A}}^2$ and $\sigma_{\delta_{\rm d}}^2$ represent the variances of $\mathcal{A}$ and $\delta_{\rm d}$, respectively, while $\sigma_{\mathcal{A}\delta_{\rm d}}$ denotes the covariance between these two parameters. The variances $\sigma_{\mathcal{A}}^2$ and $\sigma_{\delta_{\rm d}}^2$ are obtained from the diagonal elements of the covariance matrix, which provides a complete and quantitative description of the statistical uncertainties in the fit. The uncertainty in the dipole amplitude is therefore determined by $\sigma_{\mathcal{A}} = \sqrt{\mathbf{C}_{11}}$, and it directly reflects the precision. Similarly, the uncertainty in the parameter $\delta_{\rm d}$ is calculated as $\sigma_{\delta_{\rm d}} = \sqrt{\mathbf{C}_{22}}$, which describes the stability. The covariance $\sigma_{\mathcal{A}\delta_d}$, which quantifies the degree of correlation between $\mathcal{A}$ and $\delta_{\rm d}$, is directly taken from the off-diagonal element of the covariance matrix, specifically $\sigma_{\mathcal{A}\delta_{\rm d}} = \mathbf{C}_{12}$. By choosing 4 δ-belts segmentations of the FoV, the fitting results are shown.

Alternatively, an improved method is to compute each belt's dipole declination, $\delta_{{\rm d},i}$, by using not only the right ascension (RA) α_i, included in $\mathfrak{U}_i$ for each event, but also the declination δ_i, and then averaging all values of $\delta_{{\rm d},i}$. Next, substitute each $\delta_{{\rm d},i}$ into Eq. (6.54) to obtain the corresponding dipole amplitudes $\mathcal{A}_i$, and average these amplitudes to determine the final values of $\delta_{\rm d}$ and $\mathcal{A}$. Because the individual declinations δ_i are ignored in the previous RA-only analysis, this approach extracts additional anisotropy information from the full three-dimensional event distribution. Moreover, it permits an independent, simultaneous fit of both $\mathcal{A}$ and $\delta_{\rm d}$ within a single δ-belt, thereby improving accuracy and overall model consistency. Specifically, we define $\mathfrak{U}_{z,i}$ as the averaged z-component of the arrival directions of events, namely, $\sin\delta$, which is not included in the 2D data vectors $\mathbf{u}_{ij}$ or the corresponding reference vectors $\mathbf{r}_{ik}$. The relation between data and theory yields $\mathfrak{U}_{z,i} = \hat{\mathcal{I}}_z/\hat{\mathcal{I}}_0$. Similar to Eq. (6.24), the effective average of this $\sin\delta$ over the FoV is obtained by the weighted summation

$$\mathfrak{U}_{z,i} = \left| \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \mathrm{w}_j\, \sin\delta_{ij} - \frac{1}{\mathcal{N}_i^{\rm r}} \sum_{k=1}^{\mathcal{N}_i^{\rm r}} \mathrm{w}_k\, \sin\delta_{ik}^{\rm r} + \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \sin\delta_{ij}^0 \right|, \tag{6.49}$$

where $\delta_{ik}^{\rm r}$ denotes the declination of each event in the background sample (i.e., after time-scrambling. Let $\hat{r}_z^0$ be the third term of Eq. (6.49), i.e., the average of all z-components within the i-th belt, $\sin\delta_{ij}^0$ denote the isotropic background (as a monopole) for the z-

component of CR arrival directions. Equivalently, it can be written as the mean value of the z-components integrated over the δ and α ranges of the belt, given by

$$\hat{r}^0_{z,i} = \frac{1}{\Delta\Omega}\int_0^{2\pi}\int_{\delta_\mathrm{L}}^{\delta_\mathrm{U}} \sin\delta\cos\delta\,\mathrm{d}\delta\,\mathrm{d}\alpha = \langle\sin\delta_{ij}\rangle = s_i, \tag{6.50}$$

where s is defined in Eq. (6.39), and $\Delta\Omega = \int_{\delta_L}^{\delta_U}\cos\delta\,\mathrm{d}\delta$. The variance of $\mathfrak{U}_z$ is $\mathrm{Var}(\mathfrak{U}_z) = (W/\mathcal{N}^2 + W^\mathrm{r}/\mathcal{N}^{\mathrm{r},2})\mathrm{Var}(\sin\delta)$. The uncertainty of $\mathfrak{U}_i$ is given by

$$\sigma_{\mathfrak{U}_z,i} = \sqrt{\frac{W\sigma^2_{\sin\delta}}{\mathcal{N}_i^2}\left(1+\frac{1}{n_\mathrm{scr}}\right)}, \tag{6.51}$$

where the variance of $\sin\delta$, denoted by $\sigma^2_{\sin\delta}$, is calculated as

$$\sigma^2_{\sin\delta} = \frac{\sin\delta_\mathrm{U}+\sin\delta_\mathrm{L}}{2} - \frac{\sin^2\delta_\mathrm{U}+\sin\delta_\mathrm{U}\sin\delta_\mathrm{L}+\sin^2\delta_\mathrm{L}}{3} = t_i - s_i^2. \tag{6.52}$$

As the scramble times n_scr increase, $\sigma_{\mathfrak{U}_{z,i}}$ decreases slightly.

Based on Eq. (6.38) and (6.41), $\mathcal{A}_\mathrm{s}$ and $\mathcal{A}_\mathrm{p}$ can be expressed as functions of the resultant amplitude $\mathfrak{U}_i$ and the FoV parameters p, s and t, namely

$$\mathcal{A}_\mathrm{s} = \frac{\mathfrak{U}_{z,i}-s_i}{t_i - s_i\,\mathfrak{U}_{z,i}}, \quad \mathcal{A}_\mathrm{p} = \frac{2\,\mathfrak{U}_i(1+\mathcal{A}_\mathrm{s}s_i)}{p_i}. \tag{6.53}$$

Given the relation $\tan\delta_\mathrm{d} = \mathcal{A}_\mathrm{s}/\mathcal{A}_\mathrm{p}$, the dipole declination δ_d is then reconstructed as

$$\delta_{\mathrm{d},i} = \arctan\left(\frac{p\,(\mathfrak{U}_{z,i}-s_i)}{2\,\mathfrak{U}_i(t_i-s_i^2)}\right). \tag{6.54}$$

Then, $\mathcal{A}$ can be obtained by Eq. (6.42). If δ-belts suffer from large statistical fluctuations, we may limit the field of view to just two or even a single belt when estimating $\delta_{\mathrm{d},i}$. The statistical uncertainty on $\delta_{\mathrm{d},i}$ is obtained by standard error propagation, yielding

$$\sigma_{\delta_d,i} = \frac{1}{1+\mathcal{Q}^2}\sqrt{\left(\frac{\mathcal{Q}}{\mathfrak{U}_i}\right)^2\sigma^2_{\mathfrak{U},i} + \left(\frac{p}{2\,\mathfrak{U}_i(t_i-s_i^2)}\right)^2\sigma^2_{\mathfrak{U}_z,i}}, \tag{6.55}$$

where denote $\mathcal{Q}$ as the argument of the arctan in Eq. (6.54), for simplicity. Subsequently, using standard error propagation with respect to $\mathfrak{U}_i$ utilizing Eq. (6.41) and δ_d, the uncertainty of $\mathcal{A}_i$ in this method is obtained as

$$\sigma_{\mathcal{A},i} = \frac{\sqrt{\left(4\,p_i^2\cos\delta^2_{\mathrm{d},i}\,\sigma^2_{\mathfrak{U},i} + \left[2\,\mathfrak{U}_i\,(p_i\sin\delta_{\mathrm{d},i} + 2\,\mathfrak{U}_i\,s_i\cos\delta_{\mathrm{d},i})\right]^2\sigma^2_{\delta_\mathrm{d},i}}}{(p_i\cos\delta_{\mathrm{d},i} - 2\,\mathfrak{U}_i\,s_i\sin\delta_{\mathrm{d},i})^2}, \tag{6.56}$$

where the uncertainty in δ_d, i.e., $\sigma_{\delta_\mathrm{d},i}$, is obtained from Eq. (6.55).

However, this approach has the potential to result in significant biases in the estimation of both $\mathcal{A}$ and δ_d, particularly in situations where the observed cosmic-ray dipoles are positioned at relatively lower declinations and exhibit nonuniform exposure across the field of view. In such cases, the small gradients observed in the $\mathfrak{U}_i$ values along the declination can therefore lead to poor and uncertain reconstructions of both $\mathcal{A}$ and δ_d, thus affecting the overall reliability of the results. As a result, the precision and accuracy of these estimations are both significantly compromised under such observational limitations. For this reason, an independent analysis using an alternative approach is thus necessary to provide a cross-check and to complement the δ-belt method through joint analysis. This would hopefully lead to more accurate and less biased reconstructions of $\mathcal{A}$, α_d and δ_d using the same dataset. In the next step, we choose a different set of belts for the FoV.

6.3.4 Vectors on Solid-Angle Belts

Unlike the case of the δ-belts, we now introduce a family of solid-angle belts that are oriented perpendicularly to the y-axis, thereby providing an alternative geometric configuration that enables an independent cross-check of the dipole reconstruction. Each of these belts can still be expressed in terms of the usual equatorial coordinates, including both right ascension (α) and declination (δ), which is similar to the case of declination belts, but with a focus on their orientation relative to the xz plane instead of the yz plane. For a given solid angle Θ_Ω, which is oriented perpendicularly to the y-axis and extending over the interval from 0 to π along the y-axis, the boundary of this solid-angle region forms a circular ring on the celestial sphere. The angle Θ_Ω is specifically defined as the angle subtended by this ring, as depicted in Figure 6.14. A self-defined coordinate system (φ, Θ_Ω) is introduced, where φ (ranging from 0 to 2π) represents the plane angle on the ring, i.e., the angle between a vector and the y-axis, while Θ_Ω (ranging from 0 to π) represents the opening angle of the ring relative to the y-axis. The origin of this coordinate system corresponds to the centre of the celestial sphere.

The vectors on the Θ_Ω-ring and its associated belts remain two-dimensional, a choice that greatly simplifies the averaging procedure along the Θ_Ω direction on the celestial sphere, in close analogy to the approach previously applied to the δ-belts. More specifically, a two-dimensional unit vector $\mathbf{v} = (\mathrm{v}_x, \mathrm{v}_z)$ with $|\mathbf{v}| = 1$ on a Θ_Ω-ring is geometrically expressed in terms of the angles Θ_Ω and φ at the orientation by dividing the factor $\cos\theta$, namely in $\mathrm{v}_x = \cos\varphi$ and $\mathrm{v}_z = \sin\varphi$ to define its orientation more precisely. The relative acceptances at the positions of these vectors are not 100% due to zenith angles in the data, ranging from 0° to 40°. To represent the vectors on these solid angle belts using data, we employ both spherical (x, z) and equatorial coordinates (α, δ), as the arrival direction data are usually given in equatorial coordinates. Specifically, the relationships are given by

$$\begin{aligned} \mathrm{v}_x &= \frac{x}{\sqrt{x^2+z^2}} = \frac{\cos\delta\sin\alpha}{\sqrt{\sin^2\delta+\cos^2\delta\sin^2\alpha}}, \\ \mathrm{v}_z &= \frac{z}{\sqrt{x^2+z^2}} = \frac{\sin\delta}{\sqrt{\sin^2\delta+\cos^2\delta\sin^2\alpha}}, \end{aligned} \tag{6.57}$$

and then $\mathbf{v}$ on a Θ_Ω-ring can be well expressed. Most importantly, the vectors $\mathbf{v}_{ij}$ must be calibrated to account for the non-uniform relative acceptance $\epsilon = \cos\theta$ if the arrival directions in data, where θ denotes the zenith angle at the given position. This is because the acceptance is not uniform along the φ direction within the same Θ_Ω-ring or belt, making vectors at different locations not directly comparable. Since both coordinate systems can equivalently represent a given point on the celestial sphere, their corresponding x, y, and z coordinates must be equal. Thus, one obtains the relationships between angles, namely,

$$\sin\Theta_\Omega\cos\varphi = \cos\delta\sin\alpha, \quad \cos\Theta_\Omega = \cos\delta\cos\alpha, \quad \sin\Theta_\Omega\sin\varphi = \sin\delta. \tag{6.58}$$

All angular relationships in this construction can be derived directly from the equations presented above. When a Θ_Ω-ring continuously varies its opening angle Θ_Ω as it moves across the celestial sphere, the result is the creation of what we define as a Θ_Ω-belt. In this scenario, when the mean angle is equal to Θ_Ω, the corresponding ring becomes identical to a specific slice of the belt at that same opening angle, thereby preserving its geometric interpretation. Θ_Ω. This allows for a precise recovery of both the amplitude and orientation of the global real dipole later. Before performing the dipole reconstruction, we partition the celestial sphere into a sequence of Θ_Ω-belts based on a relationship triangle, with the

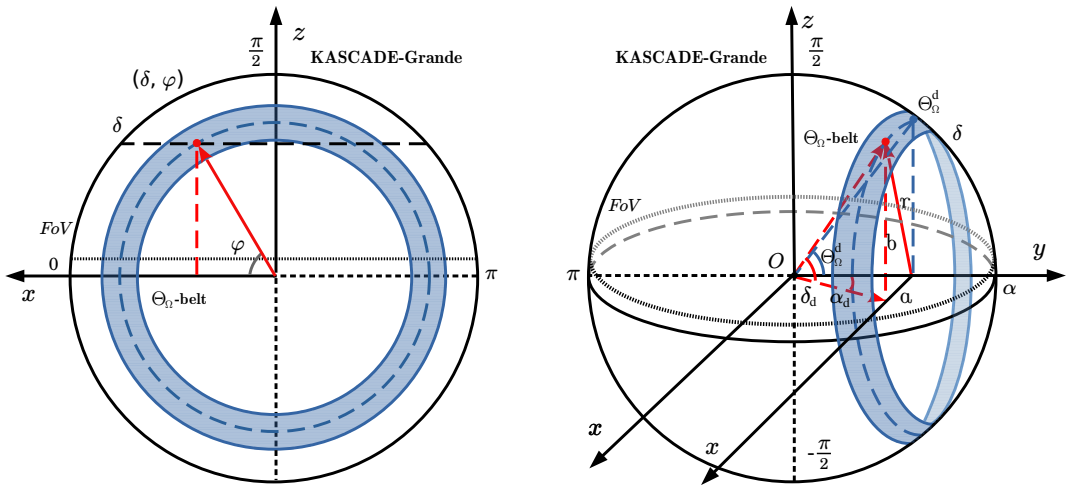


Figure 6.14: The relationship between different coordinate systems on a solid angle belt is illustrated. The right plot depicts a 3D belt perpendicular to the y-axis, while the left plot presents the 2D plane projection of the belt along with the spherical coordinate system. The red lines indicate a vector from the data on the belt.

solid angle Θ_Ω, right ascension α and declination Θ_Ω, as illustrated in the right plot of Figure 6.14. This division is performed specifically according to the following relation:

$$\sin^2\Theta_\Omega = \sin^2\delta + \cos^2\delta\sin^2\alpha, \tag{6.59}$$

where Θ_Ω, α, and δ span specific ranges, and the relationship described in Eq. (6.59) is based on a random point located on the Θ_Ω-ring, as depicted by the dashed circle. Due to the limited FoV of the observation, the rings, and belts are no longer complete. As a result, the local dipole maximum excess will not always lie on a broken Θ_Ω-ring or belt. To reconstruct the dipole anisotropies from data, we first configure the distribution of a dipole's relative intensity over the full celestial sphere using the defined (φ, Θ_Ω) coordinate system, similar to the representation in equatorial coordinates shown in Eq. (6.18). Based on spherical trigonometry, the dipole distribution function is expressed as

$$\mathcal{I}(\varphi, \Theta_\Omega) = 1 + \mathcal{A}\cos\Theta^{\rm d}_\Omega\cos\Theta_\Omega + \mathcal{A}\sin\Theta^{\rm d}_\Omega\sin\Theta_\Omega\cos(\varphi - \varphi_{\rm d}), \tag{6.60}$$

where the dipole amplitude in the φ-plane (which corresponds to the xz-plane) at declination δ, namely the parallel component, is denoted as $\mathcal{A}_\varphi = \mathcal{A}\sin\Theta^{\rm d}_\Omega\sin\Theta_\Omega$, analogous to the case of a one-dimensional projection. The term $\cos(\varphi - \varphi_{\rm d})$ accounts for the dipole variation in the φ-plane, which, in fact, corresponds to a δ-ring. Meanwhile, the shift term, denoted as $\mathcal{A}_\Theta = \mathcal{A}\cos\Theta^{\rm d}_\Omega\cos\Theta_\Omega$, is aligned with the y-axis in the Θ_Ω direction.

As in Section 6.3.3, we also review the traditional dipole reconstruction method using the vectors defined above. To accurately reconstruct the dipole amplitude and orientation on a Θ_Ω-belt, we start by considering a reference belt, which serves as the null hypothesis in our analysis. This reference belt is carefully selected so that it exhibits a rotational symmetry relationship with the Θ_Ω-belt (for the case of time-scrambled data as background, not symmetric). The relative intensity, denoted as $\delta I_{ij} = I_{ij} - 1$, where I_{ij} is intensity of vector $\mathbf{v}_{ij}$. It is associated with the arrival directions on the Θ_Ω-belt is then defined as the deviation in the distribution density of $\mathbf{v}_{ij}$, from what is expected under the null hypothesis, namely, $\delta\mathcal{I}(\varphi) = (\rho_{\rm D} - \rho_{\rm R})/\rho_{\rm R}$, where $\rho_{\rm D}(\varphi)$ represents the density of events

at a specific location on the Θ_Ω-belt, while $\rho_{\rm R}(\varphi)$ refers to the density at a corresponding location on the reference belt.

At a given position of a vector on the i-th belt, the reference vector $\mathbf{r}_{ij}$ is obtained from the averaged vector using the time-scrambling approach as we did for the δ-belts previously. The key point here is to translate the information of event density into vectors in the data. Note that the total number of data vectors $\mathbf{v}_{ij}$ on the i-th ring or belt is generally not equal to that of the corresponding reference ring or belt, i.e., $\mathcal{N}_i \neq \mathcal{N}_i^{\rm R}$ in both the time-scrambling case and isotropic background case. To ensure consistency in the number of vectors between the data belt and the reference belt, whether for the isotropic background or the time-scrambling case, we assume $\mathcal{N}_i = \mathcal{N}_i^{\rm R}$ by appropriately scaling the magnitude of each reference vector, as was done in the δ-belt case, shown in Eq. (6.19). This adjustment is not necessary for the data vectors themselves, as the averaging procedure, discussed later, automatically accounts for this. Similar to the δ-ring case discussed in Eq. (6.19), the unit vectors ($\mathbf{v}_{ij}$) on a given Θ_Ω-ring ring are distributed anisotropically in the xz-plane, expressed in $(\mathrm{v}_x, \mathrm{v}_z)$. There are $\mathcal{N}_i$ such unit vectors in total, which can be mapped to $\mathcal{N}_i^*$ non-unit vectors $\mathbf{v}_{ij}^*$ that are isotropically distributed, and note that $\mathcal{N}_i \neq \mathcal{N}_i^*$. Each $\mathbf{v}_{ij}^*$ has a magnitude proportional to the local event density at point j. The relative intensity at the j-th position on a Θ_Ω-ring is now defined as:

$$\delta\mathcal{I}_{ij} = \frac{|\mathbf{v}_{ij}^*| - |\mathbf{r}_{ij}^*|}{|\mathbf{r}_{ij}^*|} \simeq |\mathbf{v}_{ij}^*| - |\mathbf{r}_{ij}^*|, \tag{6.61}$$

where $|\mathbf{r}_{ij}^*| = 1$ in the case of an isotropic background with well-calibrated and uniform acceptance in both data and background, resulting in isotropically distributed non-unit vectors. Therefore, Eq. (6.61) is also valid in the time-scrambling case, since the denominator $|\mathbf{r}_{ij}^*|$ remains close to 1, with variations much smaller than those of $\delta\mathcal{I}_{ij}$. The numerator, $(|\mathbf{v}_{ij}^*| - |\mathbf{r}_{ij}^*|)$, quantifies the RI between a data vector and its corresponding reference vector (typically ranging from 10^{-3} to 10^{-2}) and captures the small-scale localized anisotropy at the position of the data vector with improved directional resolution accuracy. Eq. (6.61) is equivalent to the standard definition, $\delta\mathcal{I} = N_j/\langle N\rangle - 1$, when the pixel scale is small, as commonly used in most sky map analyses. However, this approach requires fitting the first harmonic and, from a data-driven perspective, becomes increasingly less reliable when statistical precision is limited, especially in regions of sparse coverage, as discussed for the δ-belts in Section 6.3.1.

Similar to the intensity vector defined in Eq. (6.22) for the δ-belt case, we define the intensity vector to imitate the average of vectors on the Θ_Ω-belts, and analyze the difference in magnitude between the average of $\mathbf{v}_{ij}$ and the average of $\mathbf{r}_{ij}$. Under the assumption of an isotropic distribution of non-unit vectors, we define $\mathbf{v}_{ij}^* = n_{ij}^{\rm d}\mathbf{v}_{ij}^0$ and $\mathbf{r}_{ij}^* = n_{ij}^{\rm r}\mathbf{v}_{ij}^0$, where $n_{ij}^{\rm d}$ and $n_{ij}^{\rm r}$ are the number densities of data and reference vectors, respectively, at the at j-th equally spaced position of the i-th ring or belt. Equivalently, these densities serve only to weight the magnitudes of each vector. Therefore, one has

$$\mathbf{v}_{ij}^{\rm sig} = n_{ij}^{\rm sig}\,\mathbf{v}_{ij}^0 = (n_{ij}^{\rm d} - n_{ij}^{\rm r} + n_{ij}^0)\,\mathbf{v}_{ij}^0, \tag{6.62}$$

where $\mathbf{v}_{ij}^0$ denotes the equally spaced unit vectors, while setting $n_{ij}^0 = 1$ accounts for the absolute monopole term in the reference vectors. The intensity vector $\boldsymbol{I}_{ij}$, which by definition equals $\boldsymbol{I}_{ij} = \mathbf{v}_{ij}^{\rm sig}$, can then be written as

$$\boldsymbol{I}_{ij} = \mathbf{v}_{ij}^* - \mathbf{r}_{ij}^* + \mathbf{r}_{ij}^0, \tag{6.63}$$

where $\mathbf{v}_{ij}^0 = \mathbf{r}_{ij}^0$. As in the δ-belt case, we model the reference vectors under an isotropic

distribution of non-unit vectors by $\mathbf{r}^*_{ij} = \mathbf{r}^0_{ij} + \mathbf{r}^{\rm m}_{ij}$, where $\mathbf{r}^0_{ij}$ the pure isotropic component (monopole) on the reference ring and $\mathbf{r}^{\rm m}_{ij}$ captures the anisotropic residuals (higher-order multipoles). The scrambled data vectors $\mathbf{r}_{ij}$ follow the same decomposition, ensuring that $\mathbf{u}^*_{ij}$ and $\mathbf{r}^*_{ij}$ share the same structure and can be compared directly at each position. Denoting the average of $\boldsymbol{I}_{ij}$ by $\hat{I}_i$. By applying Eq. (6.25) to weight both the data and reference vectors, similar to Eq. (6.24), the amplitude $\hat{I}_i$ is calculated as follows:

$$\hat{I}_i = \left| \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \mathbf{v}^{\rm sig}_{ij} \right| = \left| \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \mathrm{w}_j \, \mathbf{v}_{ij} - \frac{1}{\mathcal{N}^{\rm r}_i} \sum_{k=1}^{\mathcal{N}^{\rm r}_i} \mathrm{w}_j \, \mathbf{r}_{ik} + \frac{1}{\mathcal{N}_i} \sum_{j=1}^{\mathcal{N}_i} \mathbf{r}^0_{ij} \right|, \tag{6.64}$$

where the sum of $\mathbf{v}^0_{ij}$ does not cancel out. We define $\mathfrak{V}_i = \hat{I}_i$ for clarity, as before, the uncertainty of $\mathfrak{V}_i$ remains the same as Eq. (6.26). The sum of $\mathbf{u}^*_{ij}$ equals the sum of $\mathbf{u}_{ij}$, and likewise for $\mathbf{r}^*_{ij}$ and $\mathbf{r}_{ij}$. Moreover, due to the broken belt, the resultant vector consistently fails to point in the direction of the dipole.

To reconstruct the 3D dipoles with Θ_Ω-belts, we first consider the ideal case in which the reference belts are drawn from an isotropic background $\mathbf{r}^0_{ij}$. Let $\mathcal{N}$ denote the total number of vectors within the entire field of view, and let n_0 be the vector density on the reference celestial sphere. Although the reference δ-belt and Θ_Ω-belt differ only in their geometric shape, the number of vectors on a reference Θ_Ω-belt is then given by

$$\mathcal{N}^{\rm R}_i = n_0 \int_{\Theta^{\rm L}_\Omega}^{\Theta^{\rm U}_\Omega} \int_{\varphi_{\rm L}}^{\varphi_{\rm U}} \sin\Theta_\Omega \, {\rm d}\varphi {\rm d}\Theta_\Omega = n_0 \Delta\Omega, \tag{6.65}$$

where $\sin\Theta_\Omega$ represents the geometric calibration on the celestial sphere, $\Theta^{\rm L}_\Omega$ and $\Theta^{\rm U}_\Omega$ represent the lower and upper bounds of the Θ_Ω-belt, respectively. Based on Eq. (6.58), the range of φ is determined as follows:

$$\varphi_{\rm L} = \arcsin\left(\frac{\sin\delta_{\rm L}}{\sin\Theta_\Omega}\right), \quad \varphi_{\rm U} = \pi - \arcsin\left(\frac{\sin\delta_{\rm L}}{\sin\Theta_\Omega}\right), \tag{6.66}$$

where $\varphi_{\rm L}$ and $\varphi_{\rm U}$ are symmetric around the z-axis. Given that the angle of the ring varies along the belt, the average is taken over the entire angular range Θ_Ω of the belt. The solid angle $\Delta\Omega$ covered by the reference belt (equal to that of the data belt) is:

$$\Delta\Omega = \int_{\Theta^{\rm L}_\Omega}^{\Theta^{\rm U}_\Omega} \sin\Theta_\Omega \left(\pi - 2\arcsin\left(\frac{\sin\delta_{\rm L}}{\sin\Theta_\Omega}\right)\right) {\rm d}\Theta_\Omega, \tag{6.67}$$

which does not have a symbolic result, but only numerical results for certain given $\Theta^{\rm L}_\Omega$ and $\Theta^{\rm U}_\Omega$. Different from the δ-belt case, a dipole excess cannot be obtained by directly comparing with the reference vectors on a Θ_Ω reference belt. This is because vector cancellation between the data and the reference does not leave only the dipole excess on the broken Θ_Ω-belts. The vector $\hat{\mathbf{r}}^0_i$ represents the average of all vectors $\mathbf{r}^0_{ij}$ on the reference belt. Specifically, $\hat{r}^0_{ij}(x) = 0$ and $\hat{r}^0_{ij}(z) = \sin\varphi$, since the symmetry of $\mathbf{r}^0_{ij}$ on the reference belt causes all $\cos\varphi$ components to cancel out. Integration of the remaining z-components of the vectors over φ and Θ_Ω yields a resultant vector, which is given by

$$\hat{r}^0_i = \frac{1}{\Delta\Omega} \int_{\Theta^{\rm L}_\Omega}^{\Theta^{\rm U}_\Omega} \int_{\varphi_{\rm L}}^{\varphi_{\rm U}} \sin\varphi \sin\Theta_\Omega \, {\rm d}\varphi \, {\rm d}\Theta_\Omega = \frac{2}{\Delta\Omega} \int_{\Theta^{\rm L}_\Omega}^{\Theta^{\rm U}_\Omega} \sqrt{\sin^2\Theta_\Omega - \sin^2\delta_{\rm L}} \, {\rm d}\Theta_\Omega, \tag{6.68}$$

and $\hat{\mathbf{r}}^0_i = (0, 0, \hat{r}^0_i(z))$. To correctly perform the averaging of all vectors across an entire Θ_Ω-belt, a geometric calibration factor $\sin\Theta_\Omega$ must be included, analogous to the $\cos\delta$ factor used in equatorial coordinates. This calibration is essential because the Θ_Ω-belts are

broken due to the FoV and therefore are not full cycles. As a consequence, the natural self-cancellation that would occur for a full cycle is no longer valid, and the summed vector on the belt becomes biased by this geometric incompleteness. To compensate for this effect, we construct reference belts that have the same geometric shapes as the data belts, with identical ranges in both φ and Θ_Ω and which are generated under the null hypothesis of isotropy. The reference vectors, $\mathbf{r} = (r_x, r_z)$, are then defined to be equally spaced along this reference belt within the intervals $(\varphi_\mathrm{L}, \varphi_\mathrm{U})$ and $(\Theta_\Omega^\mathrm{L}, \Theta_\Omega^\mathrm{U})$. Even with this construction, the total summed vector obtained from a Θ_Ω-belt usually has a biased direction compared to the real dipole direction on the belt.

Similar to what we did in Section 6.3.4, we now introduce the symbol $\mathfrak{V}_i \equiv \hat{I}_i$ for all subsequent calculations. The amplitude of the resultant vector (i.e., $\hat{\boldsymbol{I}}$ or $\boldsymbol{\mathfrak{V}}$) is given by $\mathfrak{V} = \sqrt{\mathfrak{V}_x^2 + \mathfrak{V}_z^2}$. Since the covariance $\mathrm{Cov}(\mathfrak{V}_x, \mathfrak{V}_z) = 0$, the variance of $\mathfrak{V}$ is approximated by $\mathrm{Var}(\mathfrak{V}) \approx (\partial_x \mathfrak{V})^2 \mathrm{Var}(\mathfrak{V}_x) + (\partial_z \mathfrak{V})^2 \mathrm{Var}(\mathfrak{V}_z)$, with $\mathrm{Var}(\mathfrak{V}_x) \neq \mathrm{Var}(\mathfrak{V}_z)$ for φ ranging from φ_L to $\pi - \varphi_\mathrm{L}$ as Eq. (6.66) shows. The variance of $\mathfrak{V}_x$ and $\mathfrak{V}_z$ are $\mathrm{Var}(\mathfrak{V}_x) = (W/\mathcal{N}^2 + W^\mathrm{r}/\mathcal{N}^{\mathrm{r},2})\mathrm{Var}(\cos\varphi)$, $\mathrm{Var}(\mathfrak{V}_x) = (W/\mathcal{N}^2 + W^\mathrm{r}/\mathcal{N}^{\mathrm{r},2})\mathrm{Var}(\sin\varphi)$, where $W = \sum_i \mathrm{w}_i^2$, $W^\mathrm{r} = \sum_k \mathrm{w}_k^2 = n_\mathrm{scr} W$. Based on Eq. (6.24) and the propagation of errors, the error of $\mathfrak{V}_i$ is obtained by $\sigma_\mathfrak{V} = \sqrt{\mathfrak{V}_x^2\,\mathrm{Var}(\mathfrak{V}_\mathrm{x}) + \mathfrak{V}_z^2\,\mathrm{Var}(\mathfrak{V}_\mathrm{z})}\Big/\mathfrak{V}$, namely

$$\sigma_{\mathfrak{V},i} = \frac{1}{\mathfrak{V}_i}\sqrt{\frac{W}{\mathcal{N}_i^2}\left(1 + \frac{1}{n_\mathrm{scr}}\right)\left(\mathfrak{V}_{x,i}^2\,\sigma_{\cos\varphi}^2 + \mathfrak{V}_{z,i}^2\,\sigma_{\sin\varphi}^2\right)}, \tag{6.69}$$

where the variances of $\cos\varphi$ and $\sin\varphi$, denoted as $\sigma_{\cos\varphi}^2$ and $\sigma_{\sin\varphi}^2$ are given by

$$\begin{aligned}
\sigma_{\cos\varphi}^2 &= \langle\cos^2\varphi\rangle - \langle\cos\varphi\rangle^2 = \int_{\Theta_\Omega^\mathrm{L}}^{\Theta_\Omega^\mathrm{U}}\int_{\varphi_\mathrm{L}}^{\varphi_\mathrm{L}} (\cos^2\varphi - \cos\varphi)\,\sin\Theta_\Omega\,\mathrm{d}\varphi\,\mathrm{d}\Theta_\Omega, \\
\sigma_{\sin\varphi}^2 &= \langle\sin^2\varphi\rangle - \langle\sin\varphi\rangle^2 = \int_{\Theta_\Omega^\mathrm{L}}^{\Theta_\Omega^\mathrm{U}}\int_{\varphi_\mathrm{L}}^{\varphi_\mathrm{L}} (\sin^2\varphi - \sin\varphi)\,\sin\Theta_\Omega\,\mathrm{d}\varphi\,\mathrm{d}\Theta_\Omega.
\end{aligned} \tag{6.70}$$

As the number of scramblings n_scr increases, $\sigma_\mathfrak{V}$ decreases slightly, while the overall stability of the reference map improves. Inserting Eq. (6.69) into Eq. (6.16), the tail probability of the Rayleigh distribution is then obtained as

$$\mathcal{P}(\geq \mathfrak{V}_i) = \exp\left(-\frac{\mathcal{N}_i^2\,\mathfrak{V}_i^4}{2W(1 + 1/n_\mathrm{scr})(\mathfrak{V}_{x,i}^2\,\sigma_{\cos\varphi}^2 + \mathfrak{V}_{z,i}^2\,\sigma_{\sin\varphi}^2)}\right), \tag{6.71}$$

which differs from Eq. (6.27) for the δ-belt case, where the variance of cosines is not equal to the variance of sines. This arises because in the case of the broken Θ_Ω-belts, the value of $\mathcal{P}$ varies when the range of φ spanned by the belts changes, assuming $\mathcal{N}_i$ remains unchanged. Besides, when φ spans the full interval 0 to 2π, it reduces to Eq. (6.27). The overall $\mathcal{P}$ value for the entire FoV can be obtained by averaging the $\mathcal{P}$ values of all belts, appropriately weighted according to the number of vectors (i.e., event counts).

To conclude, the Θ_Ω-belt approach provides a reconstruction of dipole anisotropies that is fully comparable in performance and interpretation to the δ-belt that was discussed in the previous sections. They therefore form not a purely one-dimensional Rayleigh analysis, but rather a two-dimensional extension of it. However, it is important to maintain sufficiently high statistics in each belt and to ensure an accurately calibrated, declination-dependent exposure that properly reflects the zenith-angle acceptance of the detector when determining how many Θ_Ω-belts can be reliably defined within the available FoV. In the next step, the method of dipole reconstruction on the Θ_Ω is described in full detail and describes how the global dipole parameters are recovered from these quantities.

6.3.5 Reconstruction with Solid-Angle Belts

Next, we evaluate the average of the intensity vector $\boldsymbol{\mathcal{I}}_{ij}(\varphi, \Theta_\Omega)$, as defined in Eq. (6.60), over the entire Θ_Ω-belt. The resulting amplitude $\mathfrak{U}_i$ now represents the actual pseudo-average over the broken belt after acceptance calibration. However, its direction does not correspond to the true dipole direction, since the belt does not cover a full cycle. Since the magnitude of the difference between the averaged data vectors and reference vectors yields a residual vector containing the x- and z-components of the pseudo-average of intensity, this residual can then be fitted to the pure dipole distribution function by integrating and averaging over the entire Θ_Ω-belt. Based on the dipole distribution function, i.e., Eq. (6.60), the two components on a Θ_Ω-ring can then be expressed as

$$\mathcal{I}_x(\varphi, \Theta_\Omega) = \mathcal{I}(\varphi, \Theta_\Omega)\cos\varphi, \quad \mathcal{I}_z(\varphi, \Theta_\Omega) = \mathcal{I}(\varphi, \Theta_\Omega)\sin\varphi, \tag{6.72}$$

and after averaging $\mathcal{I}(\varphi, \Theta_\Omega)$ over a δ range, we obtain the pseudo-average of intensity on a Θ_Ω-belt. Based on the dipole distribution function Eq. (6.60), the x-, z- and y-components of $\hat{\mathcal{I}}$ as well as the monopole component $\hat{\mathcal{I}}_0$ are calculated as

$$\begin{aligned}
\hat{\mathcal{I}}_x &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} \int_{\varphi_{\mathrm{L}}}^{\varphi_{\mathrm{U}}} \mathcal{I}\cos\varphi\sin\Theta_\Omega \mathrm{d}\varphi\mathrm{d}\Theta_\Omega = P_1\mathcal{A}\sin\Theta_\Omega^{\mathrm{d}}\cos\varphi_{\mathrm{d}},\\
\hat{\mathcal{I}}_z &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} \int_{\varphi_{\mathrm{L}}}^{\varphi_{\mathrm{U}}} \mathcal{I}\sin\varphi\sin\Theta_\Omega \mathrm{d}\varphi\mathrm{d}\Theta_\Omega = S_1\mathcal{A}\cos\Theta_\Omega^{\mathrm{d}} + P_1\mathcal{A}\sin\Theta_\Omega^{\mathrm{d}}\sin\varphi_{\mathrm{d}},\\
\hat{\mathcal{I}}_y &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} \int_{\varphi_{\mathrm{L}}}^{\varphi_{\mathrm{U}}} \mathcal{I}\cos\Theta_\Omega\sin\Theta_\Omega \mathrm{d}\varphi\mathrm{d}\Theta_\Omega = S_2 + P_2\mathcal{A}\cos\Theta_\Omega^{\mathrm{d}} + 2P_3\mathcal{A}\sin\Theta_\Omega^{\mathrm{d}}\sin\varphi_{\mathrm{d}},\\
\hat{\mathcal{I}}_0 &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} \int_{\varphi_{\mathrm{L}}}^{\varphi_{\mathrm{U}}} \mathcal{I}\,\sin\Theta_\Omega\,\mathrm{d}\varphi\,\mathrm{d}\Theta_\Omega = \Delta\Omega + S_2\mathcal{A}\cos\Theta_\Omega^{\mathrm{d}} + 2\,P_0\mathcal{A}\,\sin\Theta_\Omega^{\mathrm{d}}\sin\varphi_{\mathrm{d}},
\end{aligned} \tag{6.73}$$

where, φ_{L} and φ_{U} are defined according to Eq. (6.66), the solid angle $\Delta\Omega$ is given by Eq. (6.67), the factor $\sin\Theta_\Omega$ provides the geometric calibration on the celestial sphere. The first and second integral terms in $\hat{\mathcal{I}}_x$ equal zero. The parameters S_1, S_2, $P1$, $P2$, P_3 and P_0 are obtained from the integrations over Θ_Ω, as follows:

$$\begin{aligned}
S_1 &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} 2\cos\Theta_\Omega\sin\Theta_\Omega\cos\varphi_{\mathrm{L}}\,\mathrm{d}\Theta_\Omega,\\
S_2 &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} (\pi - 2\varphi_{\mathrm{L}})\cos\Theta_\Omega\sin\Theta_\Omega\,\mathrm{d}\Theta_\Omega,\\
P_1 &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} \sin^2\Theta_\Omega\left(\frac{\pi}{2} - \varphi_{\mathrm{L}} - \frac{1}{2}\sin 2\varphi_{\mathrm{L}}\right)\mathrm{d}\Theta_\Omega,\\
P_2 &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} (\pi - 2\varphi_{\mathrm{L}})\cos^2\Theta_\Omega\sin\Theta_\Omega\,\mathrm{d}\Theta_\Omega,\\
P_3 &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} \sin^2\Theta_\Omega\cos\Theta_\Omega\sqrt{1 - \frac{\sin^2\delta_{\mathrm{d}}}{\sin^2\Theta_\Omega}}\,\mathrm{d}\Theta_\Omega,\\
P_0 &= \int_{\Theta_\Omega^{\mathrm{L}}}^{\Theta_\Omega^{\mathrm{U}}} \sin^2\Theta_\Omega\sqrt{1 - \frac{\sin^2\delta_{\mathrm{d}}}{\sin^2\Theta_\Omega}}\,\mathrm{d}\Theta_\Omega.
\end{aligned} \tag{6.74}$$

The angle Θ_Ω always satisfies the condition $\sin\Theta_\Omega > 0$, and the declination δ is restricted to the interval $\delta \in (0, \pi/2)$ due to the FoV of KASCADE-Grande. Within these constraints,

the relation $\sin\varphi_{\rm d} = \sin\delta_{\rm d}/\sin\Theta^{\rm d}_{\Omega}$ connects the dipole phase $\varphi_{\rm d}$ on a Θ_Ω to the true dipole declination δ_d and the corresponding belt opening angle $\Theta^{\rm d}_{\Omega}$. Using this geometric relation, the pseudo-average vector on the i-th Θ_Ω-belt can be expressed in terms of the contributions from its two fundamental components, namely

$$\mathfrak{V}_i = \hat{\mathcal{I}}_i = \sqrt{\left(\frac{\hat{\mathcal{I}}^i_x}{\hat{\mathcal{I}}^i_0}\right)^2 + \left(\frac{\hat{\mathcal{I}}^i_z}{\hat{\mathcal{I}}^i_0}\right)^2}, \tag{6.75}$$

where different belts yield different ranges of Θ_Ω for the parameters shown in Eq. (6.74).

Besides, we define $\mathfrak{V}_{y,i}$ as the averaged y-component of the arrival directions of events, namely, $\cos\Theta_\Omega$, which is not included in the 2D data vectors $\mathbf{v}_{ij}$ or the corresponding reference vectors $\mathbf{r}_{ik}$ but provides additional information for the dipole reconstructions. The relation between data and theory yields $\mathfrak{V}_{y,i} = \hat{\mathcal{I}}^i_y/\hat{\mathcal{I}}^i_0$. Similar to Eq. (6.49), the effective average of $\cos\Theta_\Omega$ over the FoV is then obtained by the weighted summation

$$\mathfrak{V}_{y,i} = \left| \frac{1}{\mathcal{N}_i}\sum_{j=1}^{\mathcal{N}_i} \mathrm{w}_j\, \cos\Theta_{ij} - \frac{1}{\mathcal{N}^{\rm r}_i}\sum_{k=1}^{\mathcal{N}^{\rm r}_i} \mathrm{w}_k\, \cos\Theta^{\rm r}_{ik} + \frac{1}{\mathcal{N}_i}\sum_{j=1}^{\mathcal{N}_i} \cos\Theta^0_{ij} \right|, \tag{6.76}$$

where $\Theta^{\rm r}_{ik}$ denotes the declination of each event in the background, and $\cos\Theta^0_{ij}$ represents the uniformly spaced unit vectors that serve as the background for each $\cos\Theta_{ij}$, which do not cancel out because each belt's Θ range is incomplete. The variance of $\mathfrak{V}_y$ is $\mathrm{Var}(\mathfrak{V}_y) = (W/\mathcal{N}^2 + W^{\rm r}/\mathcal{N}^{\rm r,2})\mathrm{Var}(\cos\Theta)$. The uncertainty of $\mathfrak{U}_i$ is then obtained as

$$\sigma_{\mathfrak{V}_y,i} = \sqrt{\frac{W\sigma^2_{\cos\Theta}}{\mathcal{N}^2_i}\left(1 + \frac{1}{n_{\rm scr}}\right)}, \tag{6.77}$$

where the variance of $\cos\Theta$, denoted as $\sigma^2_{\cos\Theta} = \langle\cos^2\Theta\rangle - \langle\cos\Theta\rangle^2$, is calculated as

$$\sigma^2_{\cos\Theta} = \frac{\cos^3\Theta_{\rm L} - \sin^3\Theta_{\rm U}}{3(\cos\Theta_{\rm L} - \cos\Theta_{\rm U})} - \left(\frac{\sin^2\Theta_{\rm U} - \sin^2\Theta_{\rm L}}{2(\cos\Theta_{\rm L} - \cos\Theta_{\rm U})}\right)^2. \tag{6.78}$$

As the scramble times $n_{\rm scr}$ increase, $\sigma_{\mathfrak{U}_{z,i}}$ decreases slightly.

Because the Θ_Ω-belts are broken, the pseudo-phase obtained at the i-th belt is equivalent to that of the integrated case and does not represent the actual phase of the local dipole on the belts. Based on Eq. (6.73) and Eq. (6.74), the theoretical expression of $\mathfrak{V}_{y,i}$ is thus obtained as:

$$\mathfrak{V}_{y,i} = \frac{\hat{\mathcal{I}}_{y,i}}{\hat{\mathcal{I}}_{0,i}} = \frac{S^i_2 + P^i_2\mathcal{A}\cos\Theta^{\rm d}_{\Omega} + 2P^i_3\mathcal{A}\sin\Theta^{\rm d}_{\Omega}\sin\varphi_{\rm d}}{\Delta\Omega^i + S^i_2\mathcal{A}\cos\Theta^{\rm d}_{\Omega} + 2\,P^i_0\mathcal{A}\,\sin\Theta^{\rm d}_{\Omega}\sin\varphi_{\rm d}}. \tag{6.79}$$

The dipole pseudo-phase $\varphi^*_{{\rm d},i}$ is estimated from a δ-belt as the direction of the resultant vector $\mathfrak{U}_i$ points, is given by $\tan(\varphi^*_{{\rm d},i}) = \mathfrak{V}_{z,i}/\mathfrak{V}_{x,i}$. Together with Eqs. (6.75) and (6.79), one can solve for $\mathcal{A}$, $\varphi_{\rm d}$ and $\Theta^{\rm d}_{\Omega}$. However, a more direct approach is to compare the average data components, i.e., $\mathfrak{V}_{x,i}$, $\mathfrak{V}_{z,i}$ from the 2D vectors and $\mathfrak{V}_{y,i}$ from Eq. (6.79), with those of the model. By considering the averages of $\hat{\mathcal{I}}_x$ and $\hat{\mathcal{I}}_z$ using Eq. (6.64) and Eq. (6.73), the two components $\mathfrak{V}_{x,i}$ and $\mathfrak{V}_{z,i}$ are expressed as

$$\begin{aligned}
\mathfrak{V}_{x,i} &= \frac{\hat{\mathcal{I}}_{x,i}}{\hat{\mathcal{I}}_{0,i}} = \frac{P^i_1\mathcal{A}\sin\Theta^{\rm d}_{\Omega}\cos\varphi_{\rm d}}{\Delta\Omega + S^i_2\mathcal{A}\cos\Theta^{\rm d}_{\Omega} + 2\,P^i_0\mathcal{A}\,\sin\Theta^{\rm d}_{\Omega}\sin\varphi_{\rm d}}, \\
\mathfrak{V}_{z,i} &= \frac{\hat{\mathcal{I}}_{z,i}}{\hat{\mathcal{I}}_{0,i}} = \frac{S^i_1\mathcal{A}\cos\Theta^{\rm d}_{\Omega} + P^i_1\mathcal{A}\sin\Theta^{\rm d}_{\Omega}\sin\varphi_{\rm d}}{\Delta\Omega + S^i_2\mathcal{A}\cos\Theta^{\rm d}_{\Omega} + 2\,P^i_0\mathcal{A}\,\sin\Theta^{\rm d}_{\Omega}\sin\varphi_{\rm d}}.
\end{aligned} \tag{6.80}$$

The uncertainties of $\mathfrak{V}_{x,i}$ and $\mathfrak{V}_{z,i}$, denoted by $\sigma_{\mathfrak{V}_x,i}$ and $\sigma_{\mathfrak{V}_z,i}$ correspond to the two terms in Eq. (6.69). At each belt, $\mathcal{A}$, φ_d and Θ^d_Ω can be solved using the three functions in Eq. (6.79) and Eq. (6.80). Furthermore, for simplicity, we can replace Θ^d_Ω and φ_d with δ_d and α_d using the angular relationships between the equatorial coordinates (δ, α) and our self-defined coordinates (Θ_Ω, φ), as shown in Eq. (6.58), namely

$$\sin\Theta^\mathrm{d}_\Omega \cos\varphi_\mathrm{d} = \cos\delta_\mathrm{d}\sin\alpha_\mathrm{d}, \quad \cos\Theta^\mathrm{d}_\Omega = \cos\delta_\mathrm{d}\cos\alpha_\mathrm{d}, \quad \sin\Theta^\mathrm{d}_\Omega \sin\varphi_\mathrm{d} = \sin\delta_\mathrm{d}, \tag{6.81}$$

which can be derived from solid geometry based on the equatorial coordinates and the self-defined coordinates, as shown in Figure 6.14. We note that the uncertainties of $\mathfrak{V}_{x,i}$, $\mathfrak{V}_{z,i}$ and $\mathfrak{V}_{y,i}$ are not equal to it in Eq. (6.26), because φ varies over an incomplete cycle rather than a full period, and similarly for $\mathfrak{V}_{x,i}$, since Θ_Ω does not span a full period either. The uncertainties of $\mathcal{A}$, α_d, and δ_d can also be determined in the fit through error propagation, taking into account the uncertainties from the data, i.e., $\sigma_{\mathfrak{V}_i}$ associated with $\mathfrak{V}_i$. Finally, the overall uncertainties of $\mathcal{A}$, α_d and δ_d for the entire FoV can then be obtained by taking a weighted average of the individual belt uncertainties.

The procedure described above can also be viewed as a Fourier-like "on-off" or signal–background comparison method, which is particularly advantageous in situations with limited statistics and intrinsically weak anisotropy signals. In the "on-off" method, one typically compares the number of counts (or events) in an "on" region (where both signal and background are present) with those in an "off" region (where only background is measured). The difference (after appropriate scaling) gives an estimate of the signal. By contrast, our angular-vector method is designed not merely to compare event numbers but to extract detailed directional information, specifically the amplitude and phase of an anisotropy, from the distribution of arrival directions, thereby achieving a much higher level of precision than the standard count-based approach. For instance, consider the case where ten events are recorded within a small directional patch, and the background model also contains ten events within the same patch. Using the traditional on–off method, one would conclude that the observed flux is perfectly isotropic, i.e., the anisotropy equals zero, since there is no difference in the event counts between observation and background. However, in our angular-vector analysis, the directional distributions of arrival directions in the observation and background are not identical. As a result, their respective resultant vectors differ, indicating the presence of anisotropies in the data.

6.3.6 Dipole Reconstructions

In this analysis, we exclude solar days from the early period when detector operation and data collection were unstable, and data coverage was largely non-continuous; see the distribution of solar days in Appendix C. Specifically, we include only the complete years from 2005 to 2011. The filtered datasets for all energy bins with sidereal days containing time gaps longer than two hours were removed. Furthermore, to improve overall consistency and suppress residual seasonal or instrumental effects, we first filter the data to retain only those sidereal years in which any day gaps are filled by the same sidereal days from other years. Figure 6.15 shows the event histograms for both the original and filtered data, where each 24-hour sidereal day is divided into 48 half-hour bins. Any day with a total gap (i.e., bins containing fewer than 100 events) exceeding two hours is removed to improve data stability. The anti-sidereal time histogram exhibits smaller variations than the sidereal one, indicating that systematic effects have been reduced.

To suppress statistical fluctuations in the construction of the reference (scrambled)

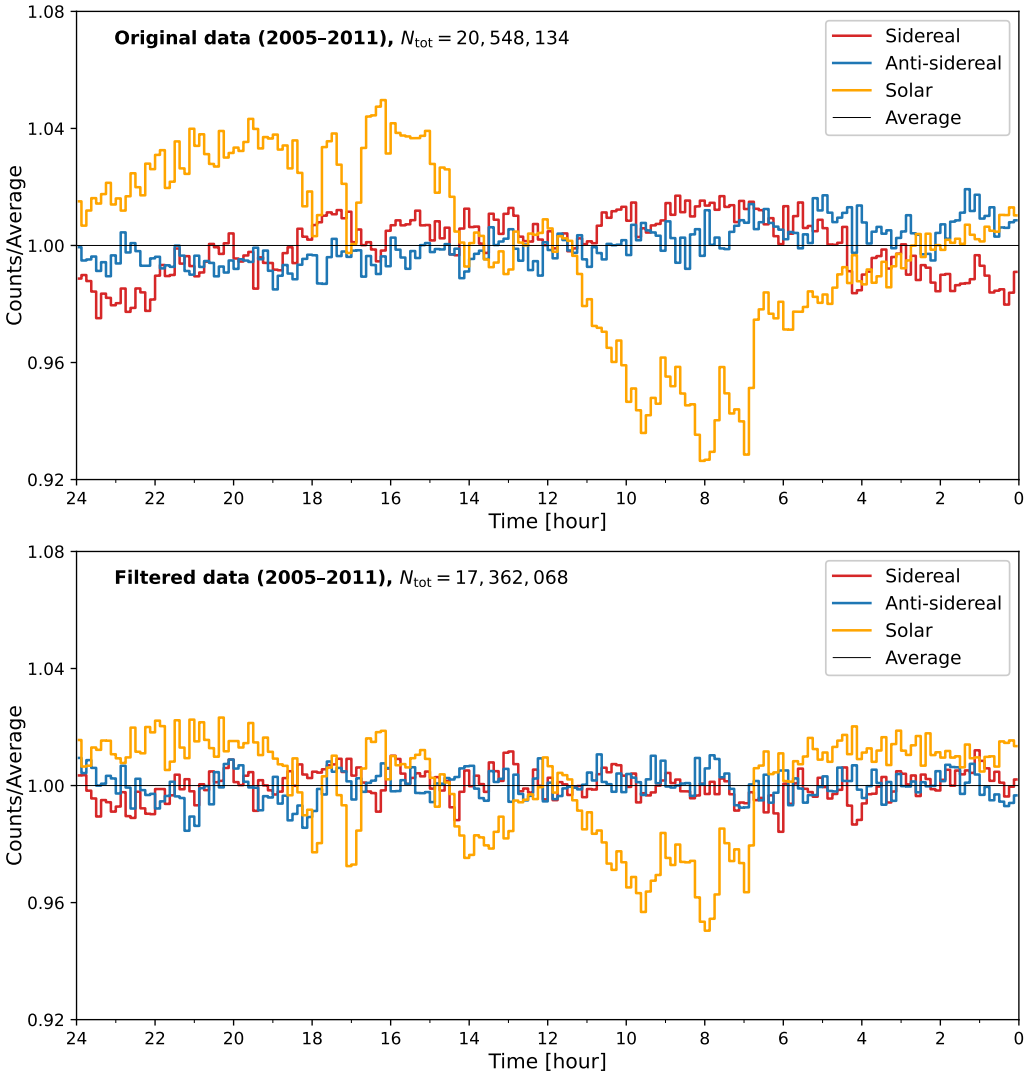


Figure 6.15: Event distributions of the selected KASCADE–Grande data from 2005 to 2011, folded into a 0–24 h cycle in sidereal, anti-sidereal, and solar time. The top panels show the original data, while the bottom panels display the filtered data after removing days with large time gaps, revealing reduced fluctuations across all three time frames.

vectors, we set the number of time scrambles according to the available statistics in each energy bin. Specifically, $n_{\mathrm{scr}} = 10$ for the high-statistics 2.7 PeV bin, $n_{\mathrm{scr}} = 20$ for the 6.1 PeV bin (limited by memory), and $n_{\mathrm{scr}} = 100$ for the low-statistics 33 PeV bin, before reconstruction. For the per-event time weight $w(\tau_j)$, we apply a two-layer UTC reweighting scheme. First, we build per-day histograms with $B = 48$ time bins per day (corresponding to ~ 30 min each) and define an effective day count

$$N_{\mathrm{eff,day}}(d) = N_{\mathrm{day}}(d)\,\frac{B}{n_{\mathrm{nonempty}}(d) + \epsilon}\,, \tag{6.82}$$

where ϵ is a small positive constant added to the denominator to avoid division by zero or unstable values when a day has very few or even no non-empty time bins. Then, the above expression 6.82 yields a day weight $w_{\mathrm{day}}(d) \propto \langle N_{\mathrm{eff,day}} \rangle / N_{\mathrm{eff,day}}(d)$, renormalized

Table 6.4: Angular-vector method with δ-belts and time-scrambling applied to the original data. Shown are the reconstructed sidereal dipole equatorial amplitudes, phases, and Rayleigh probabilities $P(\geq \mathfrak{U})$ obtained using a single δ-belt. The corresponding solar and anti-sidereal dipole amplitudes are also presented for all time-ordered subsets of each energy bin. The dipole phases derived from the time-ordered subsets are more stable across all bins compared with those in the previous section, as shown in Tables 6.2 and 6.3.

Energy	$\mathcal{N}$	$\mathcal{A}_{\rm p}(10^{-3})$	$\alpha_{\rm d}(^\circ)$	$P(\geq \mathfrak{U})$	$\mathcal{A}_{\rm p}^{\rm sol}(10^{-3})$	$\mathcal{A}_{\rm p}^{\rm anti}(10^{-3})$
Bin 1 (Total)	13,770,325	3.55 ± 0.57	207.1 ± 9.1	2.94×10^{-9}	0.12 ± 0.19	0.02 ± 0.19
Bin 1 (05–11)	13,117,461	2.40 ± 0.58	206.3 ± 13.8	1.88×10^{-4}	0.11 ± 0.20	0.02 ± 0.20
Bin 1 (05–08)	8,234,965	1.81 ± 0.73	209.3 ± 23.1	4.55×10^{-2}	0.13 ± 0.25	0.04 ± 0.25
Bin 1 (09–11)	4,882,496	4.84 ± 0.96	241.8 ± 11.3	2.87×10^{-6}	0.11 ± 0.32	0.01 ± 0.32
Bin 2 (Total)	4,737,418	5.44 ± 0.96	229.6 ± 10.2	1.23×10^{-7}	9.01 ± 0.33	4.86 ± 0.33
Bin 2 (05–11)	4,525,248	3.84 ± 0.99	239.5 ± 14.7	5.13×10^{-4}	8.49 ± 0.34	4.70 ± 0.34
Bin 2 (05–08)	2,782,856	3.88 ± 1.25	234.8 ± 18.5	8.13×10^{-3}	7.56 ± 0.43	5.41 ± 0.43
Bin 2 (09–11)	1,742,392	3.70 ± 1.61	225.8 ± 24.9	7.10×10^{-2}	18.9 ± 0.55	5.32 ± 0.55
Bin 3 (Total)	115,566	4.34 ± 5.76	293.4 ± 76.0	7.53×10^{-1}	2.41 ± 2.09	6.73 ± 2.09
Bin 3 (05–11)	110,303	5.09 ± 5.89	275.3 ± 66.3	6.89×10^{-1}	2.13 ± 2.14	6.03 ± 2.14
Bin 3 (05–08)	68,531	7.65 ± 7.46	260.3 ± 55.9	5.91×10^{-1}	3.52 ± 2.72	6.59 ± 2.72
Bin 3 (09–11)	41,772	2.90 ± 9.60	306.9 ± 189.6	9.55×10^{-1}	4.04 ± 3.49	5.08 ± 3.49

such that $\langle w_{\rm day} \rangle = 1$. Afterwards, combining all available days together (weighted by $w_{\rm day}$), we then form an effective occupancy per UTC time bin i and define a bin weight $w_i \propto \langle N_{\rm eff,bin} \rangle / N_{\rm eff,bin}(i)$, also renormalized to $\langle w_i \rangle = 1$. Each event occurring at UTC phase τ_j on day $d(j)$ is then assigned $w(\tau_j) = w_{i(\tau_j)}/w_{\rm day}(d(j))$, followed by a final normalization to $\langle w \rangle = 1$ (with small limits applied to remove extreme weight values). Finally, we can apply the same weights to both the data and the scrambled background when constructing the weighted angular vectors and dipoles.

Now, we proceed to the dipole reconstruction. For the δ-belt case in the angular-vector method, the time-scrambling procedure effectively accounts for daily variations while preserving sensitivity to angular structures at all scales. However, as discussed earlier, due to limited statistics, noisy data, and a partially covered sky, the dipole's vertical term $\mathcal{A}_{\rm s} = \mathcal{A} \sin\alpha$ becomes uncertain. Given that $\mathcal{A}_{\rm s} s \sim 10^{-2}$–$10^{-3}$ in this analysis, the denominator factor $(1 + \mathcal{A}_{\rm s} s) \approx 1$ of the denominator in Eq. (6.41), Eq. (6.41) is then reduced to $\mathcal{A}_{\rm p} = 2\,\mathfrak{U}/p$ with only the equatorial dipole component remaining. The corresponding error of $\mathcal{A}_{\rm p}$ is obtained as $\sigma_{\mathcal{A}_{\rm p}} = 2\sigma_{\mathfrak{U}}/p$. The reconstruction of dipole phase remains the same as Eq. (6.42) shown. By performing a single δ-belt reconstruction, the amplitude $\mathcal{A}_{\rm p}$ can be obtained with reduced fluctuation and improved statistical overall stability. Furthermore, the detailed sidereal amplitudes and phases for 2.7 PeV, 6.1 PeV, and 33 PeV are shown in Tables 6.4 and 6.5, corresponding to the original and filtered data, respectively. The phase uncertainties are reduced compared to the analysis in Section 6.2, while the equatorial amplitudes closely match those reported in the papers [87, 288]. Finally, the phase for the 6.1 PeV bin is smaller than reported in the literature, most likely due to remaining residual systematics. Similarly, in the scenario where a single Θ_Ω-belt spanning the full KASCADE–Grande FoV $\left(\Theta_\Omega^{\rm L} = \delta_{\rm L} = 9.1^\circ, \Theta_\Omega^{\rm U} = \pi - \delta_{\rm L}\right)$. However, this one-belt setup is largely insensitive to the dipole declination $\delta_{\rm d}$, while using multiple belts introduces additional noise due to the limited FoV of each belt. Therefore, this Θ_Ω-belt case is not used in the analysis presented in this thesis.

Table 6.5: Angular-vector method with δ-belts and time-scrambling applied to the filtered data. Shown are the reconstructed sidereal, solar, and anti-sidereal dipole equatorial amplitudes, phases, and Rayleigh probabilities $P(\geq \mathfrak{U})$ obtained using a single δ-belt. The dipole phases obtained from the time-ordered subsets show reduced stability in the second energy bin, while those in the first and third bins remain stable. The amplitudes, however, are more consistent across the last two bins, except for the first bin.

Energy	$\mathcal{N}$	$\mathcal{A}_{\rm p}(10^{-3})$	$\alpha_{\rm d}(^\circ)$	$P(\geq \mathfrak{U})$	$\mathcal{A}_{\rm p}^{\rm sol}(10^{-3})$	$\mathcal{A}_{\rm p}^{\rm anti}(10^{-3})$
Bin 3 (Total)	$11,924,773$	1.28 ± 0.61	174.5 ± 27.1	1.08×10^{-1}	0.07 ± 0.21	0.02 ± 0.21
Bin 3 (05–11)	$11,483,834$	0.96 ± 0.62	206.3 ± 37.1	3.04×10^{-1}	0.07 ± 0.21	0.02 ± 0.21
Bin 3 (05–08)	$7,034,107$	1.80 ± 0.79	106.8 ± 25.0	7.30×10^{-2}	0.09 ± 0.27	0.03 ± 0.27
Bin 3 (09–11)	$4,449,727$	3.14 ± 1.00	260.6 ± 18.2	7.10×10^{-3}	0.10 ± 0.33	0.03 ± 0.33
Bin 1 (Total)	$4,096,639$	2.11 ± 0.99	238.7 ± 26.9	1.03×10^{-1}	6.02 ± 0.35	1.99 ± 0.35
Bin 1 (05–11)	$3,954,264$	1.36 ± 1.01	244.6 ± 42.6	4.05×10^{-1}	5.72 ± 0.36	1.94 ± 0.36
Bin 1 (05–08)	$2,371,325$	1.52 ± 1.30	263.5 ± 48.7	5.01×10^{-1}	6.81 ± 0.46	2.85 ± 0.46
Bin 1 (09–11)	$1,582,939$	2.44 ± 1.60	248.8 ± 37.7	3.15×10^{-1}	6.30 ± 0.57	2.33 ± 0.57
Bin 3 (Total)	$115,567$	5.62 ± 6.21	281.6 ± 63.3	6.64×10^{-1}	0.56 ± 0.62	0.46 ± 0.62
Bin 3 (05–11)	$95,792$	5.50 ± 6.32	276.6 ± 65.9	6.85×10^{-1}	0.55 ± 0.63	0.40 ± 0.63
Bin 3 (05–08)	$58,024$	3.67 ± 8.11	247.0 ± 126.7	9.03×10^{-1}	0.37 ± 0.81	0.37 ± 0.81
Bin 3 (09–11)	$37,768$	3.92 ± 10.08	283.6 ± 147.4	9.27×10^{-1}	0.39 ± 1.01	0.39 ± 1.01

To assess the long-term stability of the reconstructed anisotropies, we divide the filtered event dataset into two time-ordered subsets, following the same procedure used earlier, and compare the corresponding results obtained from the East–West method and the optimized Rayleigh analysis, as shown in Figure 6.8 and Fig. 6.11 respectively. The reconstructions of sidereal 3D dipoles for the three energy bins, the results are presented in Tables 6.4 and 6.5. We observe that, for all three energy bins, the reconstructed dipole amplitudes and phases agree between the two time-ordered halves within their respective statistical uncertainties, although small residual biases remain. This indicates that the combination of time-dependent weights and the applied time-scrambling procedure does not fully account for variations in the effective number of active detector stations, nor does it completely correct for gradual long-term efficiency drifts or for seasonal fluctuations in event statistics. These incomplete corrections lead to residual modulations in the data, and such modulations leak into the sidereal frame, thereby contaminating the reconstructed anisotropies. Recall that experiments such as Auger [89] and HAWC [136] typically build reference maps (the isotropic expectation) by factoring the detector response into a local-angle acceptance and local angles. However, we do not have such detailed detector information. That is why we employ the time-scrambling approach, which works by randomly shifting events within a time window and averaging. This procedure smooths statistical noise and yields a stable background, but it cannot simultaneously smooth over all scales (from hours to months). Consequently, we obtained large biases in both dipole amplitude and phase for all three energy bins, as shown in Tables 6.4 and 6.5. Therefore, we must seek a more appropriate method to estimate the detector's solar (UTC) acceptance, rather than relying on time-scrambling.

Next, we smooth the UTC/Solar acceptance estimate by introducing a Dirichlet prior for the expected counts in each time bin. Specifically, we adopt the form $\alpha_t = \kappa\, \ell_t$, where ℓ_t denotes the relative detector exposure (livetime) in bin t, and κ is a global scale factor setting the effective sample size of the prior, which is the average over all UT bins across

Table 6.6: Angular-vector method with δ-belts and Dirichlet prior (see Eq. (6.83)) applied to the original data. Shown are the reconstructed sidereal dipole equatorial amplitudes, phases, and Rayleigh probabilities $P(\geq \mathfrak{U})$ obtained using a single δ-belt, together with the corresponding solar and anti-sidereal dipole amplitudes for all time-ordered subsets of each energy bin. The phases of the time-ordered subsets remain stable, whereas the amplitudes of the first two bins show noticeable bias, while the last bin remains consistent.

Energy	$\mathcal{N}$	$\mathcal{A}_{\rm p}(10^{-3})$	$\alpha_{\rm d}(^\circ)$	$P(\geq \mathfrak{U})$	$\mathcal{A}_{\rm p}^{\rm sol}(10^{-3})$	$\mathcal{A}_{\rm p}^{\rm anti}(10^{-3})$
Bin 1 (Total)	13,770,326	4.70 ± 0.53	311.2 ± 6.5	8.95×10^{-18}	0.12 ± 0.19	0.02 ± 0.19
Bin 1 (05–11)	13,117,462	3.52 ± 0.54	317.5 ± 8.8	7.68×10^{-10}	0.11 ± 0.20	0.02 ± 0.20
Bin 1 (05–08)	8,234,966	5.02 ± 0.69	321.6 ± 7.8	2.25×10^{-12}	0.13 ± 0.25	0.04 ± 0.25
Bin 1 (09–11)	4,882,496	1.64 ± 0.89	323.3 ± 31.2	1.85×10^{-1}	0.11 ± 0.32	0.01 ± 0.32
Bin 2 (Total)	4,737,419	4.67 ± 0.91	313.7 ± 11.1	1.81×10^{-6}	0.09 ± 0.33	0.01 ± 0.33
Bin 2 (05–11)	4,525,249	3.51 ± 0.93	320.6 ± 15.2	7.81×10^{-4}	0.09 ± 0.33	0.04 ± 0.33
Bin 2 (05–08)	2,782,857	5.81 ± 1.19	323.2 ± 11.7	5.90×10^{-6}	0.11 ± 0.42	0.02 ± 0.42
Bin 2 (09–11)	1,742,392	1.26 ± 1.50	351.9 ± 68.0	7.01×10^{-1}	0.07 ± 0.54	0.06 ± 0.54
Bin 3 (Total)	115,567	6.25 ± 5.83	313.4 ± 53.5	5.63×10^{-1}	0.37 ± 2.08	0.32 ± 2.08
Bin 3 (05–11)	110,304	5.94 ± 5.97	315.4 ± 57.6	6.10×10^{-1}	0.37 ± 2.13	0.37 ± 2.13
Bin 3 (05–08)	68,532	6.02 ± 7.58	301.1 ± 72.2	7.30×10^{-1}	0.54 ± 2.71	0.37 ± 2.71
Bin 3 (09–11)	41,772	6.52 ± 9.73	335.9 ± 85.5	7.99×10^{-1}	0.27 ± 3.48	0.17 ± 3.48

the whole timeline, i.e., $\kappa = N_{\rm tot}/\sum_j t_j$, where t_i represents the non-empty time bins. If bin t is a gap, $N_t^{\rm obs} = 0$, and $\ell_t = 0$ ensures that true gaps don't get artificial counts from the prior. Assuming uniform exposure at all time bins, $\ell_t = 1$. Follows from the Dirichlet–multinomial distribution [322, 323], and adding pseudo-counts proportional to the exposure, i.e. $N_t = N_t^{\rm obs} + \kappa\,\ell_t$. The posterior expectation is given by

$$\mathbb{E}[\,p_t \,|\, D\,] = \frac{N_t^{\rm obs} + \kappa\,\ell_t}{N_{\rm tot} + \kappa\,\Sigma_t \ell_t}, \tag{6.83}$$

where p_t denotes the probability that an event falls into a bin t, constrained by the normalization $\sum_t p_t = 1$, and where D represents the full dataset consisting of the observed counts $N_t^{\rm obs}$ in each bin. The above equation provides the smoothed probability fraction for bin t. This expression yields the smoothed fractional probability associated with bin t, and the corresponding smoothed expectation for the number of events in bin t is obtained from $E_t = N_{\rm tot}, \mathbb{E}[p_t]$. It is important to emphasize that E_t introduces a global suppression of the observed detector acceptance, regulated by the prior strength κ, which reduces large acceptance fluctuations. Unlike a local time-average procedure, the smoothing acts coherently across all bins and does not impose short-scale averaging. The efficiency (inverse of weight), as shown in Eq. (6.25), $\epsilon_{\rm T}(\tau_j)$ is now expressed as

$$\epsilon_{\rm T}(\tau_j) = \frac{N_t}{N_{\rm tot}\,\mathbb{E}(t)}, \tag{6.84}$$

which directly serves as the detector's relative exposure as a function of livetime. Based on the weights, Eq. (6.84) and the ratio of the effective number of sidereal bins $\eta(\tau_j)$ as described in Eq. (6.25), and by utilizing the previous angular-vector approach with a single belt, the results obtained from the time-split data and from the full data are presented in Tables 6.6 and 6.7. Importantly, the results obtained with the time-split data show improved consistency compared to the previous results derived using the direct

Table 6.7: Angular-vector method with δ-belts and Dirichlet prior (see Eq. (6.83)) applied to the filtered data. Shown are the sidereal, solar, and anti-sidereal dipole amplitudes and phases for all subsets. The amplitudes of the time-ordered subsets remain stable except for the second bin, while the phases of the first bin are consistent, whereas those of the last two bins still exhibit larger biases between the time-ordered subsets.

Energy	$\mathcal{N}$	$\mathcal{A}_{\rm p}(10^{-3})$	$\alpha_{\rm d}(^\circ)$	$P(\geq \mathfrak{U})$	$\mathcal{A}_{\rm p}^{\rm sol}(10^{-3})$	$\mathcal{A}_{\rm p}^{\rm anti}(10^{-3})$
Bin 1 (Total)	$11,924,774$	2.09 ± 0.57	304.2 ± 15.6	1.15×10^{-3}	0.04 ± 0.20	0.02 ± 0.20
Bin 1 (05–11)	$11,483,835$	1.71 ± 0.58	314.6 ± 19.5	1.32×10^{-2}	0.03 ± 0.21	0.03 ± 0.21
Bin 1 (05–08)	$7,034,108$	2.27 ± 0.74	325.5 ± 18.7	9.02×10^{-3}	0.06 ± 0.27	0.05 ± 0.27
Bin 1 (09–11)	$4,449,727$	1.11 ± 0.93	326.8 ± 47.9	4.90×10^{-1}	0.08 ± 0.34	0.04 ± 0.34
Bin 2 (Total)	$4,096,640$	2.13 ± 0.98	306.0 ± 26.2	9.21×10^{-2}	0.02 ± 0.35	0.04 ± 0.35
Bin 2 (05–11)	$3,954,265$	1.76 ± 0.99	314.0 ± 32.4	2.09×10^{-1}	0.01 ± 0.36	0.03 ± 0.36
Bin 2 (05–08)	$2,371,326$	3.11 ± 1.28	317.2 ± 23.7	5.31×10^{-2}	0.06 ± 0.46	0.02 ± 0.46
Bin 2 (09–11)	$1,582,939$	0.86 ± 1.57	9.2 ± 104.5	8.60×10^{-1}	0.06 ± 0.56	0.02 ± 0.56
Bin 3 (Total)	$99,291$	3.89 ± 6.29	308.7 ± 92.7	8.26×10^{-1}	0.24 ± 2.25	0.18 ± 2.25
Bin 3 (05–11)	$95,793$	4.24 ± 6.40	312.0 ± 86.5	8.03×10^{-1}	0.21 ± 2.29	0.23 ± 2.29
Bin 3 (05–08)	$58,025$	3.95 ± 8.23	286.8 ± 59.8	8.91×10^{-1}	0.36 ± 2.94	0.37 ± 2.94
Bin 3 (09–11)	$37,768$	4.74 ± 10.23	337.8 ± 61.8	8.98×10^{-1}	0.16 ± 3.66	0.31 ± 3.66

time-scrambling method, as shown in Tables 6.4 and 6.5.

We also evaluate the residual first-harmonic amplitudes in solar (SOL) and anti-sidereal (ASID) time using the same event weights that are applied for the sidereal analysis. After the UTC envelope correction, each event is assigned a combined weight w_i, and the residual amplitudes are obtained by projecting the weighted sample onto the corresponding temporal bases. As shown in Tables 6.6 and 6.7, the measured SOL and ASID dipole projected amplitudes $\mathcal{A}_{\rm p}^{\rm sol}$ and $\mathcal{A}_{\rm p}^{\rm anti}$ are much smaller than their statistical error, consistent with noise expectations. This confirms that the weighting procedure suppresses the daily or seasonal variation, as well as the artificial variations caused by time gaps. In principle, this approach is mathematically equivalent to smoothing and can be regarded as an optimised time-scrambling method when multi-scale time-scrambling is considered. At the same time, it overcomes the problem posed by the existence of numerous time gaps in the data, ranging from hours to months.

Besides, we also perform the log-likelihood iteration to achieve the best estimation of acceptance and anisotropies of the arrival directions, based on the weighted data using Eq. (6.84) and the ratio of the effective number of sidereal bins $\eta(\tau_j)$ as described in Eq. (6.25). We use the original data from 2005 to 2011, as shown in Figure 6.6, rather than the filtered dataset in Table 6.7, since filtering may increase instabilities by removing additional sidereal days. In fact, the Dirichlet prior in Eq. (6.84) already accounts for acceptance variations, making further filtering unnecessary. Firstly, we divide one sidereal day (23h 56m) into N_t equal phase bins $i = 1, \ldots, N_t$. We then fold all years of data into those N_t bins, so each bin i accumulates every event that occurred at that particular sidereal phase, regardless of which day it was. In the iteration, we fit one N_i per phase bin i. This captures the time-dependent (i.e., phase-dependent) exposure or overall event rate. We fit a single acceptance map A_j over the local-sky pixels j, each defined at a fixed zenith–azimuth (θ_j, φ_j). Here, A_j does not depend on i; it represents the average angular acceptance pattern of the detector. Considering the mean, i.e., $\mu_{ij} = N_i A_j I_j$ of the Poisson distribution in the iteration, as discussed in Section 2.2.4, where I_j is the

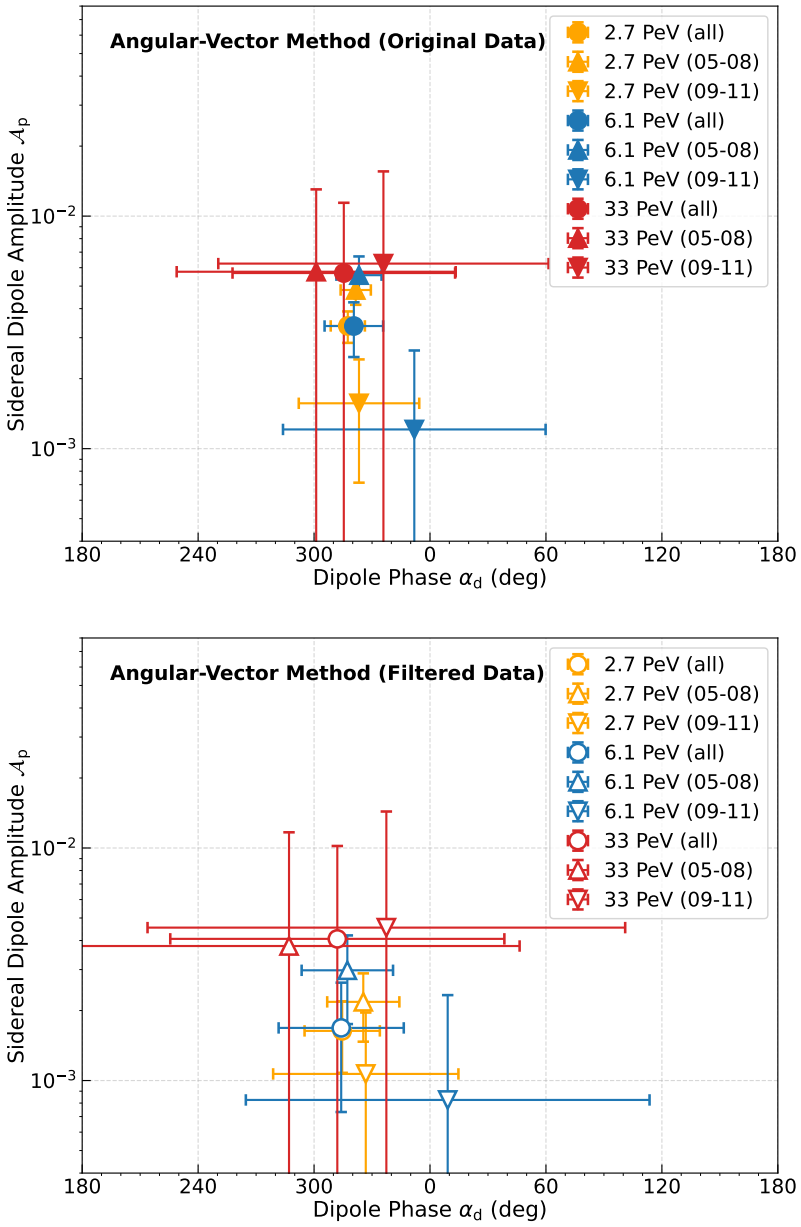


Figure 6.16: Comparison of the sidereal dipole in the phase–amplitude plane for time-ordered data splits (2005–2008, 2009–1011), together with the full data set (labeled as "all") in all three energy bins and in the full data set, using original unfiltered data (top) and filtered data (bottom), analyzed with the Angular-Vector method. Compared with Figure 6.9, which was obtained using the East–West method, the two plots above show reduced biases. It appears that the long-term variations (solar-time) are further disrupted after removing days, weakening the self-cancellation of the solar-time component. Consequently, the dipoles reconstructed from the original data (top) are more reliable.

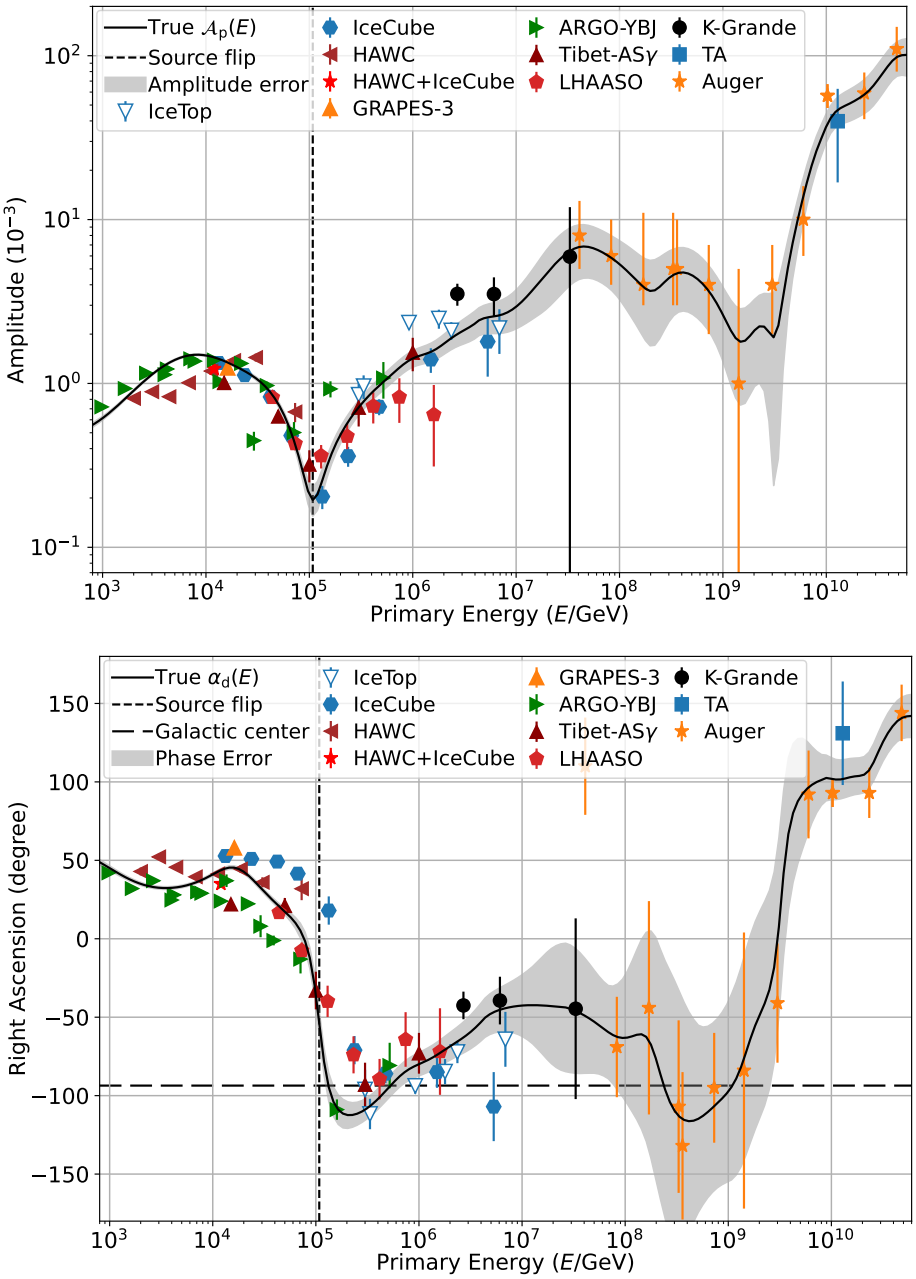


Figure 6.17: Updated unified anisotropy based on reported data from major experiments spanning 1 TeV to 50 EeV, with all amplitudes calibrated by the FoVs as discussed in Chapter 5. Note that the first Auger data point is excluded due to the large biased phase from the others. The KASCADE-Grande data used here come from the angular-vector method with a Dirichlet prior, whereas the dipoles in the older plots in Figure 5.3 were obtained using the East–West method as in the official paper [87]. Top: extracted sidereal dipole amplitude, $\mathcal{A}_p(E)$, considered for a visible global sky. Bottom: dipole phase, $\alpha_d(E)$. Both amplitude and phase curves are obtained as a function of true energy. The short dashed line in both panels marks the energy of 107.9 TeV, where the dominant CR sources may shift, with a corresponding amplitude $\mathcal{A}_p = (2.1 \pm 0.40) \times 10^{-4}$ and phase $\alpha_d = 315.3° \pm 7.7°$. The energy and amplitude of this flip remain unchanged from Figure 5.3. The long dashed lines mark the Galactic centre.

relative intensity, follows the 1D Fourier form of the data distribution (or vectors), namely, $I_j = 1 + a\cos\alpha_j + b\sin\alpha_j$. The results remain consistent with those obtained using the angular-vector approach, as shown in Tables 6.6 and 6.7, but they are not shown here.

Based on the dipole amplitudes and phases shown in Table 6.6, which are the final results with KASCADE-Grande data in this thesis, we then update the unified anisotropy through a joint analysis of different experiments, as studied in Chapter 5, and present the results in Figure 5.3. In particular, the updated results are shown in Figure 6.17. The extracted dipole variates exhibit a "W" shape in both amplitude and phase. Moreover, the fitted curves indicate that the dipole flip remains unchanged at 107.9 TeV, as confirmed in Chapter 5, suggesting a transition of Galactic cosmic-ray sources from local origins. As discussed earlier in Chapter 4, this flip may indicate a shift from nearby sources, such as the γ-ray source Crab Nebula (83.6°, 22°) and the Geminga supernova remnant [282, 135] (98.5°, 17.8°), not toward the Galactic Center (GC) region as previously proposed, but instead toward the Cygnus region. Furthermore, the fitted curve from around 1 PeV (Tibet-ASγ and IceCube) to around 100 PeV (KASCADE-Grande) shows a slight peak in phase at around 7 PeV, with dipole $\mathcal{A}_{\rm p} = 0.0025$ and phase $\alpha_{\rm d} \approx 315°$. The corresponding energy is close to the first cosmic-ray knee at 3 PeV. Consequently, this peak may be related to the deflection of the local magnetic field (including its turbulence), affecting cosmic rays propagating from the Cygnus region to Earth. Finally, between 6 PeV and 400 PeV, the next peak in amplitude at around 340 PeV, with amplitude $\mathcal{A}_{\rm p} \approx 0.006$, which may point to a transition of the cosmic-ray sources from the Cygnus region to extragalactic origins. This energy is also close to the second cosmic-ray heavy knee at 100 PeV.

It is important to emphasize that KASCADE–Grande is located at a latitude of 49.1°, and the data cover declinations from 9.1° to 89.1° due to a zenith threshold of 40°. The density of events decreases at lower declinations (closer to the equator) due to larger zenith angles, and this effect is further amplified by geometric stretching. Given that event statistics across all three energy bins are relatively low compared to sidereal dipole amplitudes, this may introduce large fluctuations in both event densities and anisotropies across the entire right ascension plane, especially at lower declinations where the field of view is narrower and exposure less uniform. Moreover, a detailed analysis of medium-scale anisotropy structures is necessary to determine whether the observed patterns indicate particular preferred orientations associated with potential cosmic-ray sources. This more refined investigation will be carried out in the following section.

6.4 Medium-Scale Anisotropy

Unlike large-scale anisotropies such as the dipole, which are typically associated with the overall cosmic-ray gradient in the Galaxy, the medium-scale structures are likely produced by scattering off local magnetic field irregularities or turbulence in the interstellar medium. These features may result from interference patterns as cosmic rays move through anisotropic magnetic turbulence. A nearby and recent source, such as a supernova remnant within a kiloparsec, could create an excess of cosmic rays in certain directions, but magnetic scattering distorts or shifts the excess, forming these medium-scale patterns. They don't necessarily point straight back to the source, but could still reflect its presence. Besides, these structures might also indicate a transition between different propagation regimes, i.e., different dominant sources located far from each other. In this section, we mainly focus on the KASCADE–Grande data to check the detailed anisotropies with sky maps.

6.4.1 Updated Maximum-Likelihood Method

Before the analysis, we validated the maximum-likelihood (ML) method by applying a time-scrambling procedure to the data for background estimation. The results are consistent with those obtained using the original method, in which the background in each pixel is iteratively updated based on an expected isotropic intensity map. The resulting maps will be shown later, and since they are identical, they are not repeated here. The underlying assumption of the maximum-likelihood method follows the same principles as direct integration or time-scrambling, as described in [143]. In principle, with a complete period (hours) of data, the analysis has equal sensitivity to anisotropic cosmic-ray arrival directions at all right ascensions. Therefore, as in Section 6.3, it is not necessary to exclude incomplete days with hour-scale gaps, as doing so would reduce the statistical power of the dataset. The only data excluded are those recorded before 27 November 2004, during which detector operation was unstable, and data coverage was largely non-continuous.

To investigate medium-scale anisotropies in the arrival directions of cosmic rays, we employ the maximum-likelihood reconstruction technique [143, 16], which was previously introduced in Section 2.2.4 Our goal is to recover the celestial relative intensity distribution $I(\alpha, \delta)$ on the celestial sphere while simultaneously accounting for the detector's time-dependent normalization $\mathcal{N}_\tau$ and its directional (local) acceptance $\mathcal{A}_i$, a probability density (or relative exposure) at the local direction i. The number of events observed in a time bin τ and local angular bin i, denoted by $n_{\tau i}$, is assumed to follow a Poisson distribution with mean $\mu_{\tau i} = \mathcal{N}_\tau \mathcal{A}_i I_{\tau i}$, where $I_{\tau i}$ represents the celestial intensity evaluated at the right ascension and declination associated with the local bin i at the sidereal time corresponding to τ. The overall likelihood, assuming independence among bins, is then expressed as

$$\mathcal{L}(I, \mathcal{A}, \mathcal{N}) = \prod_{\tau,i} \frac{(\mu_{\tau i})^{n_{\tau i}}\, e^{\mu_{\tau i}}}{n_{\tau i}!}. \tag{6.85}$$

Taking the logarithm simplifies this product into a sum:

$$\log \mathcal{L} = \sum_{\tau,i} \left[n_{\tau i} \log \left(\mathcal{N}_\tau\, \mathcal{A}_i\, I_{\tau i} \right) - \mathcal{N}_\tau\, \mathcal{A}_i\, I_{\tau i} - \log \left(n_{\tau i}! \right) \right]. \tag{6.86}$$

Since $\log(n_{\tau i}!)$ is independent of the parameters, we focus on maximising the effective log-likelihood of Eq. (6.86), namely

$$\log \mathcal{L}_{\text{eff}} = \sum_{\tau,i} \left[n_{\tau i} \log \left(\mathcal{N}_\tau\, \mathcal{A}_i\, I_{\tau i} \right) - \mathcal{N}_\tau\, \mathcal{A}_i\, I_{\tau i} \right], \tag{6.87}$$

where the parameters $I(\alpha, \delta)$, $\mathcal{A}_i$, and $\mathcal{N}_\tau$ are non-linearly coupled. The value of I_{τ_i} is obtained by transforming the celestial intensity map into local coordinates at sidereal time τ, and an analytical solution for the maximum-likelihood estimate is not feasible. Instead, we employ an iterative approach to obtain convergence. We begin by setting an initial guess based on the null hypothesis (an isotropic sky):

$$I^{(0)}(\alpha, \delta) = 1, \quad \mathcal{N}_\tau^{(0)} = \sum_i n_{\tau i}, \quad \mathcal{A}_i^{(0)} = \frac{\sum_\tau n_{\tau i}}{\sum_{\kappa j} n_{\kappa j}}, \tag{6.88}$$

where the condition $\sum_i \mathcal{A}_i^{(0)} = 1$, since it quantifies how likely (or how much relative exposure, assuming the total exposure is 1 of the entire FoV) the detector is to observe an event from direction i, assuming the sky is isotropic. This iterative procedure updates three sets of parameters, i.e., the intensity I, the time-dependent normalization factors $\mathcal{N}_\tau$ and the directional acceptance values $\mathcal{A}_i$. The updated value of the intensity in a given

celestial pixel a is obtained by gathering contributions from all local angular bins i, at all times τ a, that project onto the pixel at through the detector's coordinate transformation. The corresponding intensity update is therefore expressed as

$$I_{\mathfrak{a}}^{(n+1)} = \frac{\sum_\tau n_{\tau i}}{\sum_\tau \mathcal{N}_\tau^{(n)}\, \mathcal{A}_i^{(n)}}, \tag{6.89}$$

where the coordinates are transferred from (θ_i, ϕ_i) to the corresponding celestial coordinates (α_a, δ_a). Eq. (6.89) remains unchanged compared with the old expression, as Eq. (2.59) shows, since it is the ratio of observed to expected counts under the current model. We define $I(\alpha, \delta) = 1 + \delta I(\alpha, \delta)$ and also remove the $m = 0$ spherical harmonic components to remove any variations in declination, since the data are insensitive to them. Keeping I and $\mathcal{A}$ fixed, we first maximize $\log \mathcal{L}\text{eff}$ with respect to $\mathcal{N}_\tau$, then, holding I and $\mathcal{N}$ fixed, we take the derivative regarding $\mathcal{A}_i$ and obtain:

$$\frac{\partial \log \mathcal{L}_{\text{eff}}}{\partial \mathcal{N}_\tau} = \sum_i \left[\frac{n_{\tau i}}{\mathcal{N}_\tau} - \mathcal{A}_i I_{\tau i}\right], \quad \frac{\partial \log \mathcal{L}_{\text{eff}}}{\partial \mathcal{A}_i} = \sum_\tau \left[\frac{n_{\tau i}}{\mathcal{A}_i} - \mathcal{N}_\tau I_{\tau i}\right]. \tag{6.90}$$

Setting the derivatives to zero gives the update rules for the iterations, namely,

$$\mathcal{N}_\tau^{(n+1)} = \frac{\sum_i n_{\tau i}}{\sum_i \mathcal{A}_i^{(n)}\, I_{\tau i}^{(n+1)}}, \quad \mathcal{A}_i^{(n+1)} = \frac{\sum_\tau n_{\tau i}}{\sum_\tau \mathcal{N}_\tau^{(n+1)}\, I_{\tau i}^{(n+1)}}, \tag{6.91}$$

where the acceptance is renormalized as $\sum_i \mathcal{A}_i^{(n+1)} = 1$ after evaluating $\mathcal{A}_i^{(n+1)}$ for all i. We repeat the above steps until the solution of intensity at pixel $\mathfrak{a}$, $I_{\mathfrak{a}}$, converges, meaning that successive iterations yield only negligible changes in the reconstructed intensity.

To improve the numerical stability of the iterative maximum-likelihood procedure, particularly during early iterations or in the presence of statistical fluctuations, we introduce a learning rate parameter $\eta \in (0, 1]$ [324, 325, 326]. This learning rate controls the step size in updating each of the key parameters: the time-dependent normalization $\mathcal{N}_\tau$, the directional acceptance $\mathcal{A}_i$, and the cosmic-ray intensity $I_{\mathfrak{a}}$ in a given sky pixel $\mathfrak{a}$. Instead of applying the full maximum-likelihood update at each step, we use a convex combination of the current value and its update $\Psi^{(n+1)} = \Psi^{(n)} + \eta(\Psi_{\text{ML}}^{(n+1)} - \Psi^{(n)})$, namely

$$\Psi^{(n+1)} = (1 - \eta)\, \Psi^{(n)} + \eta\, \Psi_{\text{ML}}^{(n+1)}, \tag{6.92}$$

where $\Psi \in \{\mathcal{N}_\tau, \mathcal{A}_i, I_{\mathfrak{a}}\}$ denotes any of the iteratively updated quantities. Here, $\Psi_{\text{ML}}^{(n+1)}$ represents the value obtained from a full maximum-likelihood step, and $\Psi^{(n)}$ is the current value from iteration n. The parameter η controls how far the update proceeds toward the newly estimated solution. For $\eta = 1$, the procedure reduces to the original undamped iteration, as used in [143]. Choosing $\eta < 1$ helps suppress oscillations and avoid overshooting, which is particularly beneficial when working with low-statistics bins or complex detector-sky coupling, allowing for smoother convergence and improved stability, especially important when detector exposure is highly non-uniform or when Poisson fluctuations in sparse bins introduce noise. In our implementation, η is scheduled dynamically as a function of iteration number $\eta = 1/\sqrt{n}$ [325], allowing the method to take smaller steps at early stages and converge more confidently as the likelihood surface flattens, and improve the quality of the final anisotropy reconstruction.

Note that the intensity update function, i.e., Equation (6.89), is computed as the ratio of observed to expected counts across multiple bins. This multiplicative update is inherently de-normalising and, for this reason, is typically accompanied by an explicit projection step

that removes any $m = 0$ harmonics. As a result, the updates for $\mathcal{N}_\tau$ and $\mathcal{A}_i$ are expressed in the form of an additive convex combination, as shown in Equation (6.92). In principle, we do not need to include η in the intensity update, so it keeps the same as (6.89). For $\mathcal{N}_\tau$ and $\mathcal{A}_i$, their update with learning rate are

$$\begin{aligned} \mathcal{N}_\tau^{(n+1)} &= \left(1 - \eta^{(n)}\right) \mathcal{N}_\tau^{(n)} + \eta^{(n)} \frac{\sum_i n_{\tau i}}{\sum_i \mathcal{A}_i^{(n)}\, I_{\tau i}^{(n+1)}}, \\ \mathcal{A}_i^{(n+1)} &= \left(1 - \eta^{(n)}\right) \mathcal{A}_i^{(n)} + \eta^{(n)} \frac{\sum_\tau n_{\tau i}}{\sum_\tau \mathcal{N}_\tau^{(n+1)}\, I_{\tau i}^{(n+1)}}, \end{aligned} \tag{6.93}$$

followed by renormalization $\sum_i \mathcal{A}_i^{(n+1)} = 1$. The updated intensity is calculated as the ratio of observed to expected counts, based on the current estimates of $\mathcal{N}_\tau$ and $\mathcal{A}_i$. As a result, $I_{\mathfrak{a}}^{(n+1)}$ is already normalized within the iteration process.

6.4.2 Significance of Anisotropy

Medium-scale anisotropies were studied using the same likelihood-based method as in the large-scale analysis. However, instead of fitting a global dipole, the intensity I was reconstructed individually for each sky pixel within the three previously defined energy bins. Following Ahlers et al. [143, 16, 288], we applied a Gaussian smoothing (FWHM = 2°) during the iterations to stabilize the reconstruction and suppress numerical noise. To highlight structures on angular scales, top-hat smoothing with different radii is performed in the analysis for all energy bins in the analysis. For each celestial bin $\mathfrak{a}$, the total number of observed and expected events within this angular radius is computed based on the optimized iterations of $\mathcal{N}_\tau$ and $\mathcal{A}_i$ at each pixel, as shown in Eq. (6.93),

$$\tilde{n}_{\mathfrak{a}} = \sum_{\mathfrak{b}\in\mathcal{D}_{\mathfrak{a}}} \sum_\tau n_{\mathfrak{b}}^\tau, \quad \tilde{\mu}_{\mathfrak{a}} = \sum_{\mathfrak{b}\in\mathcal{D}_{\mathfrak{a}}} \sum_\tau \mathcal{A}_{\tau\mathfrak{b}}^\star\, \mathcal{N}_\tau^\star\, \mathcal{I}_{\mathfrak{b}}^\star, \quad \tilde{\mu}_{\mathfrak{a}}^{\mathrm{bg}} = \sum_{\mathfrak{b}\in\mathcal{D}_{\mathfrak{a}}} \sum_\tau \mathcal{A}_{\tau\mathfrak{b}}^\star\, \mathcal{N}_\tau^\star\, \mathcal{I}_{\mathfrak{b}}^{\mathrm{bg}}. \tag{6.94}$$

where $\mathcal{A}_{\mathfrak{b}}^\tau$ denotes the detector's relative acceptance in pixel $\mathfrak{b}$ and sidereal time bin τ, $\mathcal{N}^\tau$ represents the expected number of isotropic events in sidereal time bin τ, and $\mathcal{I}_{\mathfrak{b}}$ is the actual intensity at pixel $\mathfrak{b}$ [16, 288], the symbol $|^\star|$ denotes values after convergence. Assuming the background is approximately flat, $\mathcal{I}^{\mathrm{bg}} = 1$, the smoothed intensity becomes

$$\tilde{\delta I}_{\mathfrak{a}} = \tilde{\mu}_{\mathfrak{a}} / \tilde{\mu}_{\mathfrak{a}}^{\mathrm{bg}} - 1. \tag{6.95}$$

The top panels of Figures 6.19, 6.20, 6.21, 6.22, 6.23, and 6.24 show sky maps of intensity (anisotropy maps), illustrating the reconstructed anisotropies in three energy bins. To assess the statistical relevance of local excesses or deficits, a corresponding significance is constructed with Eq. (6.94), namely,

$$\tilde{S}_{\mathfrak{a}} = \sqrt{2\left(-\tilde{\mu}_{\mathfrak{a}} + \tilde{\mu}_{\mathfrak{a}}^{\mathrm{bg}} + \tilde{n}_{\mathfrak{a}} \log(1 + \tilde{\delta I}_{\mathfrak{a}})\right)}. \tag{6.96}$$

which yields a significance map. The value of $\tilde{S}_{\mathfrak{a}}$ is interpreted in units of Gaussian σ, provided that the smoothing radius is not too large, as also shown in the mentioned figures. All maps are constructed using top-hat smoothing with angles optimized for each energy bin, corresponding to the maximum pre-trial significance $\tilde{S}_{\mathfrak{a}}$ obtained from the smoothing-radius scan shown in Figure 6.18, for both the original and filtered data.

To effectively highlight the medium-scale anisotropy accesses on sky maps, and considering the limited statistics available from KASCADE-Grande, it is necessary to utilise both "pre-trial significance" and "post-trial significance". Pre-trial significance is the significance level calculated for a specific hypothesis or test without considering that multiple tests are

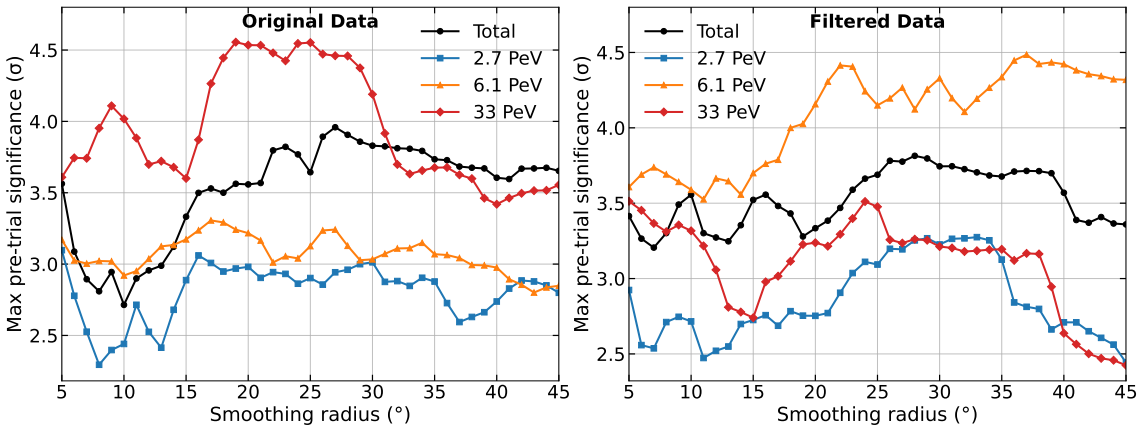


Figure 6.18: Top-hat smoothing-radius scan for the pre-trial significance sky maps using both the original and filtered data. The corresponding radius at which each energy bin reaches its maximum significance is taken as the smoothing radius.

being conducted, as Eq. (6.96) shows. It's the p-value obtained from a single hypothesis test [303, 327]. This is typically calculated using standard statistical tests (e.g., z-test, t-test, χ^2 test) and corresponds to the probability of observing the test statistic under the null hypothesis. Pre-trial significance is used when only one hypothesis is tested, reflecting the likelihood that the observed effect is due to chance. We employ the χ^2 test for the pre-trial significance in this study. The post-trial significance, also known as the adjusted significance, accounts for the fact that multiple hypotheses are being tested simultaneously [303, 327]. It provides a more accurate representation of the statistical significance by adjusting for the increased likelihood of Type I errors, thereby enhancing the reliability of the conclusions drawn from the data.

We follow the same procedure as in [288] to analyze the post-trial significances with the maximum-likelihood method [328]. We estimate the effective number of trials as $N_{\text{trial}} \simeq \Delta\Omega_{\text{FoV}}/\Delta\Omega_{\text{bin}}$, where $\Delta\Omega_{\text{FOV}}$ represents the size of the observatory's time-integrated field of view, and $\Delta\Omega_{\text{bin}}$ denotes the effective bin size corresponding to the top-hat smoothing scale. For the 33 PeV bin, for example, when using a smoothing radius of 19°, we obtain $N_{\text{trial}} \approx 15$. Various methods, such as Bonferroni correction, Holm's method, and FDR control [303, 327], are used to adjust the p-values to reflect the number of tests conducted. The post-trial p-value can then be approximated as follows:

$$p_{\text{post}} \simeq 1 - (1 - p)^{N_{\text{trial}}}, \tag{6.97}$$

where $p = \text{erfc}(\tilde{S}_{\mathfrak{a}}/\sqrt{2})/2$. The above function targets the multiple comparisons in statistical hypothesis testing. When conducting multiple independent hypothesis tests, the probability of making at least one Type I error (false positive) increases, necessitating an adjustment to the individual p-values to control the overall family-wise error rate (FWER). The derivation begins by recognizing that the probability of not observing a significant result in a single test is $1 - p_i$. When performing N_{trial} independent tests, the combined probability of not observing any significant results across all tests is $(1 - p_i)^{N_{\text{trial}}}$. The complement of this probability, $1 - (1 - p_i)^{N_{\text{trial}}}$, represents the probability of observing at least one significant result by chance, which serves as the adjusted p-value, and the FWER is properly controlled across multiple comparisons [303, 327]. Figures 6.19, 6.20, 6.21, 6.22, 6.23, and 6.24 also present the post-trial significance maps using Eq. 6.97, shown in the bottom panels for all

three energy bins and the time-ordered subsets of the KASCADE–Grande data, shown in equatorial coordinates.

The optimized method for iterative dipole reconstruction introduces a learning rate parameter η to the existing maximum-likelihood method by Ahlers [143, 288], which controls the update step size of $\delta_{0\mathrm{h}}$ and $\delta_{6\mathrm{h}}$, which correspond to $\mathcal{A}_\mathrm{s}$ and $\mathcal{A}_\mathrm{p}$ as discussed previously in our analysis. These components of the dipole are aligned along the declination and right ascension directions, respectively, and correspond to the strength of the dipole anisotropy in those directions. They are used to determine both the magnitude and phase of the 3D real dipole anisotropy. The dipole amplitude is given by $\mathcal{A} = \sqrt{\mathcal{A}_\mathrm{s}^2 + \mathcal{A}_\mathrm{p}^2}$ where $\mathcal{A}_\mathrm{s} = \mathcal{A}\sin\delta_\mathrm{d}$ is the vertical (declination) component, and $\mathcal{A}_\mathrm{p} = \mathcal{A}\cos\delta_\mathrm{d}$ is the equatorial projection. Note that $\mathcal{A}$ refers to the dipole amplitude, not to be confused with the relative acceptance $\mathcal{A}_i$. The dipole phase is given by $\alpha_\mathrm{d} = \arctan 2(\mathcal{A}_\mathrm{s}, \mathcal{A}_\mathrm{p})$. The highest pre-trial and post-trial significances reach approximately $3\,\sigma$ across all three energy bins. Specifically, for the original data, the post-trial significance reaches about $2.0\,\sigma$ at 2.7 PeV, $3.0\,\sigma$ at 6.1 PeV, and an excess of $4.0\,\sigma$ at 33 PeV, located near a Galactic longitude of about 80°. For the filtered data, the post-trial significances are $2.8\,\sigma$ at 2.7 PeV and $4.2\,\sigma$ at 6.1 PeV, and $2.8\,\sigma$ at 33 PeV.

In analyzing cosmic ray significance maps, it is crucial to consider the uncertainty in the position of detected hotspots, particularly when the data has been smoothed using a tophat function [329, 330]. The value at any given point on the map after smoothing represents the average significance over the surrounding region. While this process reduces noise, it also broadens features on the map, leading to an inherent uncertainty in pinpointing the exact position of a maximum (hotspot). To estimate the positional uncertainty due to smoothing, the tophat function can be approximated by a Gaussian function. Although the tophat is not exactly Gaussian, this approximation is useful for deriving an analytical estimate of the uncertainty. The Gaussian approximation is often used in various fields to simplify the analysis of smoothing functions and to relate the width of different types of kernels [329]. For a Gaussian distribution in two dimensions, the function is given by:

$$G(\theta) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{\theta^2}{2\sigma^2}\right), \tag{6.98}$$

where σ is the standard deviation, which determines the "width" of the distribution. The full width at half maximum (FWHM) of a Gaussian is related to σ by:

$$\mathrm{FWHM} = 2\sqrt{2\ln 2}\,\sigma \approx 2.355\,\sigma. \tag{6.99}$$

The smoothing process can be mathematically represented by a tophat function $W(\theta)$, defined as $W(\theta) = 1/\pi\theta_0^2$ for $\theta \le \theta_0$, and $W(\theta) = 0$ for $\theta > \theta_0$, where θ_0 is denotes the characteristic radius of the tophat and therefore determines the angular scale over which the smoothing operates. The diameter of the tophat function ($2\,\theta_0$) can be matched to the full width at half maximum (FWHM) of a Gaussian kernel to provide an equivalent effective "width" for comparing the angular resolution of the two smoothing schemes. By setting the FWHM of the Gaussian equal to the diameter of the tophat, namely $2\theta_0 = 2.355\,\sigma$, this leads to the relation $\sigma = 0.85\,\theta_0$. However, when considering the uncertainty in the position of the maximum introduced by the smoothing procedure, it is more appropriate to adopt the standard deviation that is directly associated with the radius of the tophat, rather than that derived from the FWHM–Gaussian equivalence. This approximation leverages the well-known correspondence between the smoothing behaviour of tophat and Gaussian filters for estimating angular-position uncertainties in cosmic-ray anisotropy

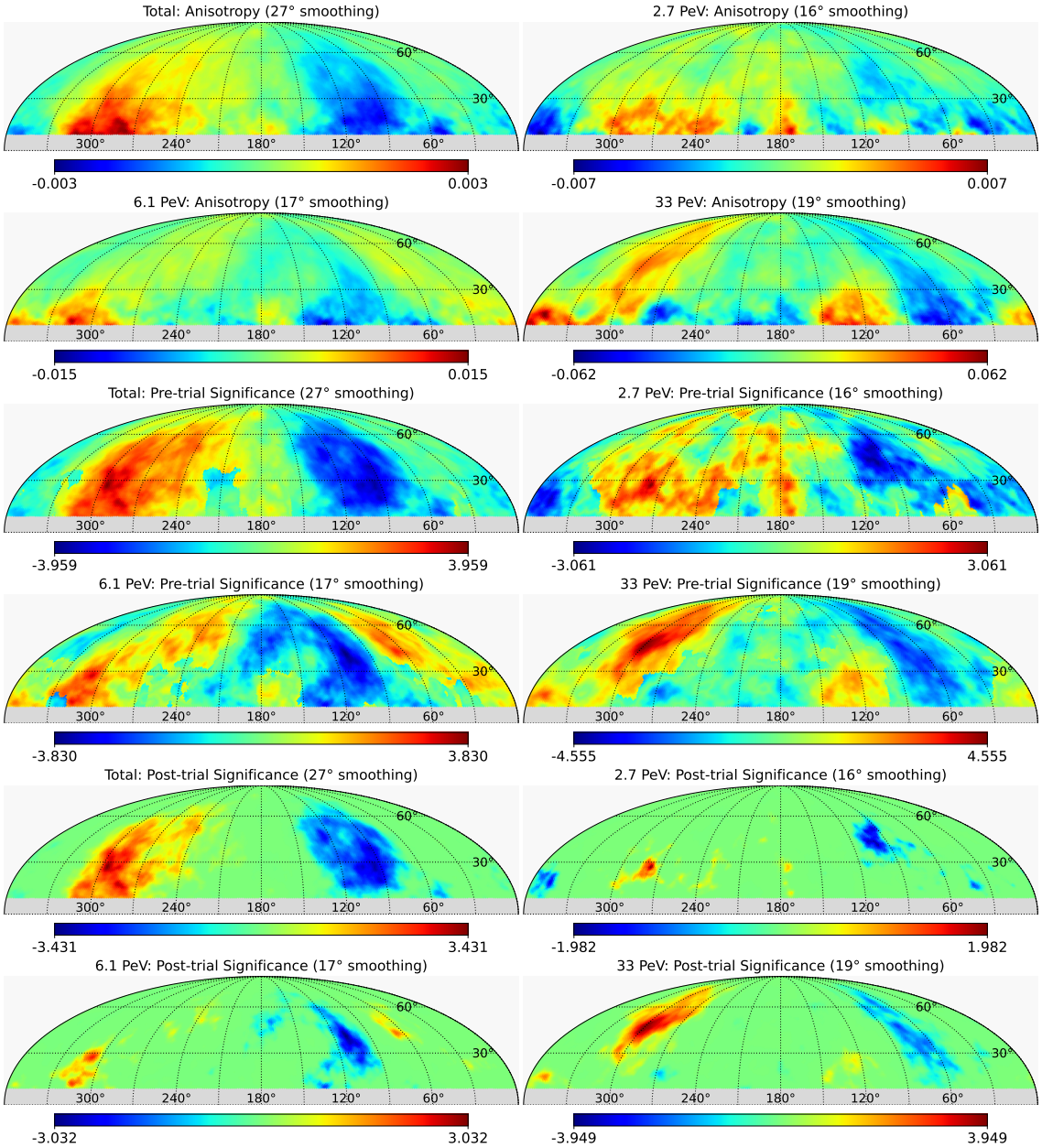


Figure 6.19: Sky maps of anisotropy, namely, relative intensity (top four panels), pre-trial significance (middle four panels), and post-trial significance (bottom four panels), obtained from the original KASCADE–Grande data. Each group of four maps corresponds to the total data, the 2.7 PeV bin, the 6.1 PeV bin, and the 33 PeV bin, respectively. Different energy bins use different top-hat smoothing radii, which are determined from the corresponding maximum pre-trial significance in the smoothing-radius scan. The detailed resulting values are accordingly shown in Figure 6.18.

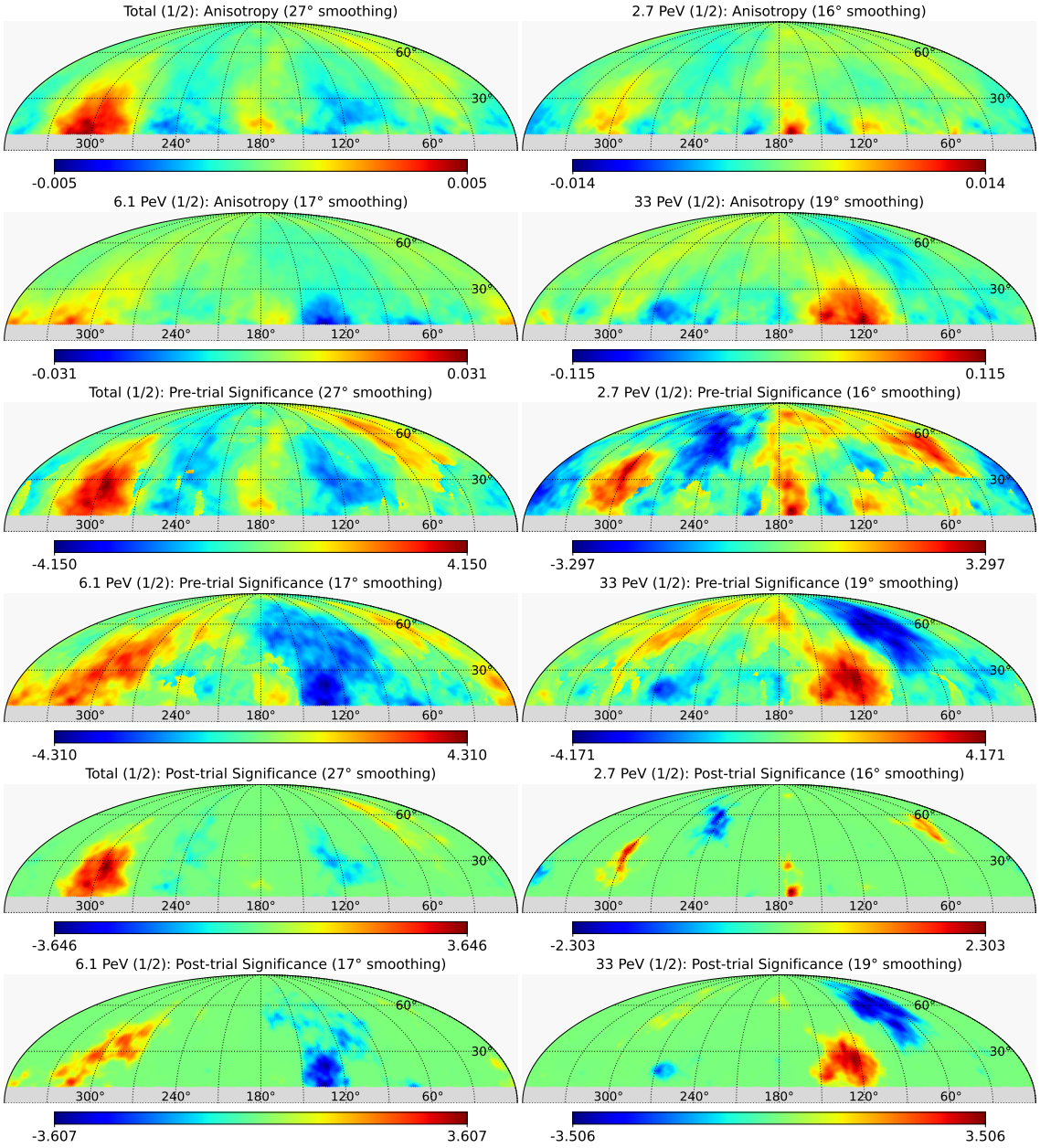


Figure 6.20: Shown are the original KASCADE–Grande data with time-ordered subsets, i.e., the 2005–2008 year range, used to check the consistency of anisotropies for each energy bin. Sky maps of anisotropy are presented, including relative intensity (top four panels), pre-trial significance (middle four panels), and post-trial significance (bottom four panels). Each group of four maps corresponds to the total data, the 2.7 PeV bin, the 6.1 PeV bin, and the 33 PeV bin, respectively. Different energy bins use different top-hat smoothing radii, which are determined from the corresponding maximum pre-trial significance in the smoothing-radius scan. The detailed resulting values are shown in Figure 6.18.

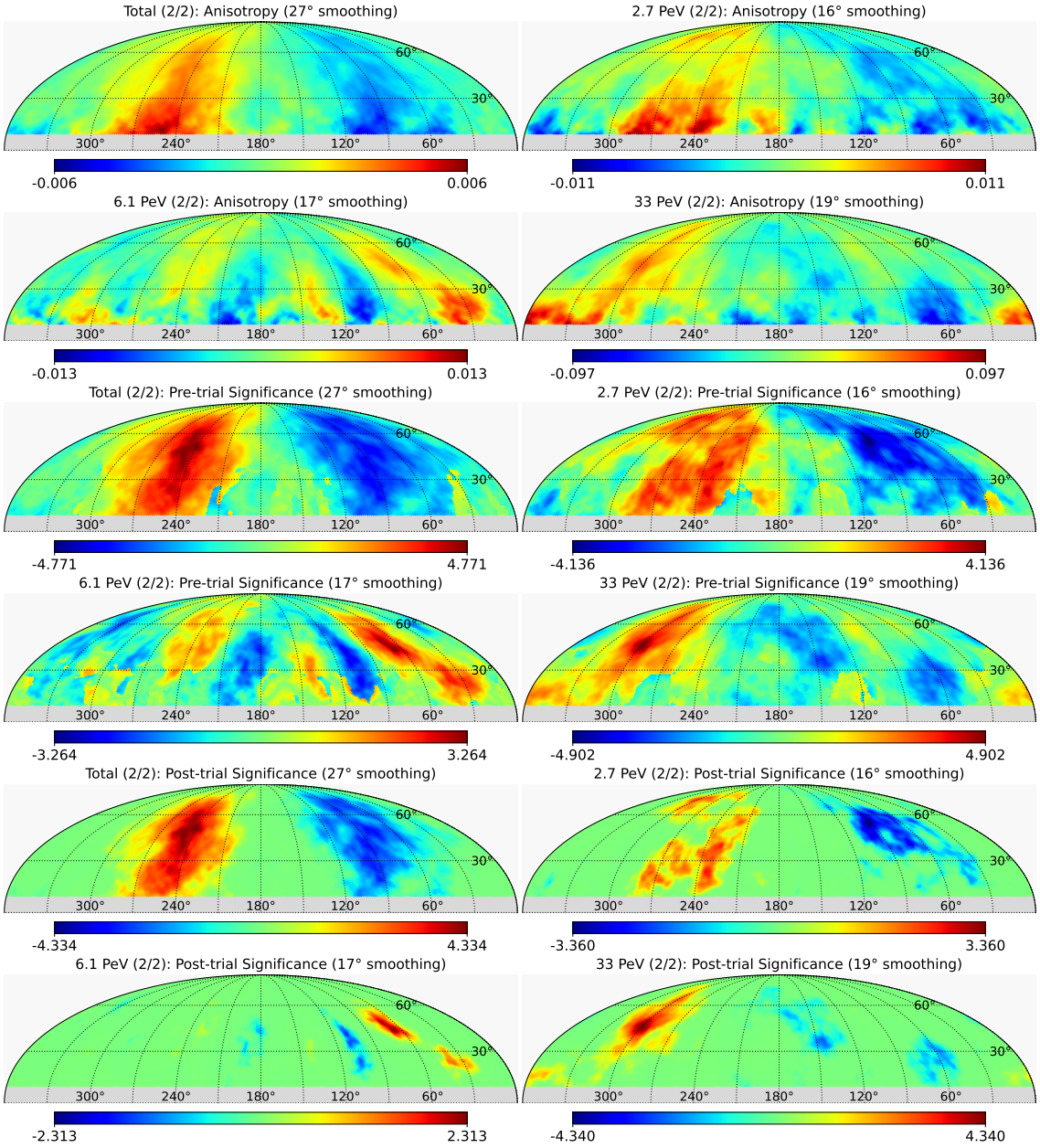


Figure 6.21: Shown are the original KASCADE–Grande data with time-ordered subsets, i.e., the 2009–2011 year range, used to check the consistency of anisotropies for each energy bin. Sky maps of anisotropy are presented, including relative intensity (top four panels), pre-trial significance (middle four panels), and post-trial significance (bottom four panels). Each group of four maps corresponds to the total data, the 2.7 PeV bin, the 6.1 PeV bin, and the 33 PeV bin, respectively. Different energy bins use different top-hat smoothing radii, which are determined from the corresponding maximum pre-trial significance in the smoothing-radius scan. The detailed resulting values are shown in Figure 6.18.

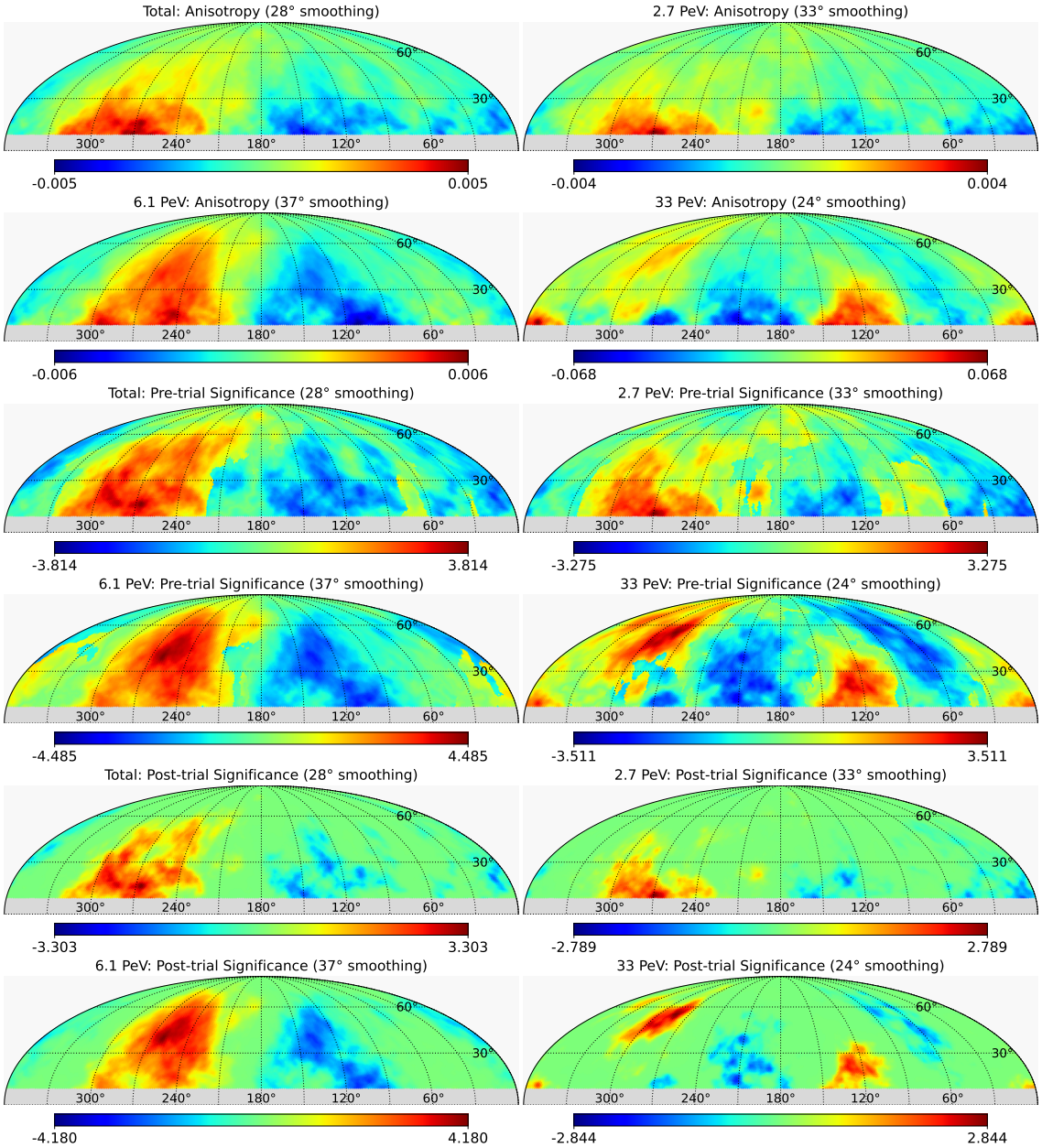


Figure 6.22: Sky maps of anisotropy, namely, relative intensity (top four panels), pre-trial significance (middle four panels), and post-trial significance (bottom four panels), obtained from the filtered KASCADE–Grande data. Each group of four maps corresponds to the total data, the 2.7 PeV bin, the 6.1 PeV bin, and the 33 PeV bin, respectively. Different energy bins use different top-hat smoothing radii, which are determined from the corresponding maximum pre-trial significance in the smoothing-radius scan. The detailed resulting values are accordingly shown in Figure 6.18.

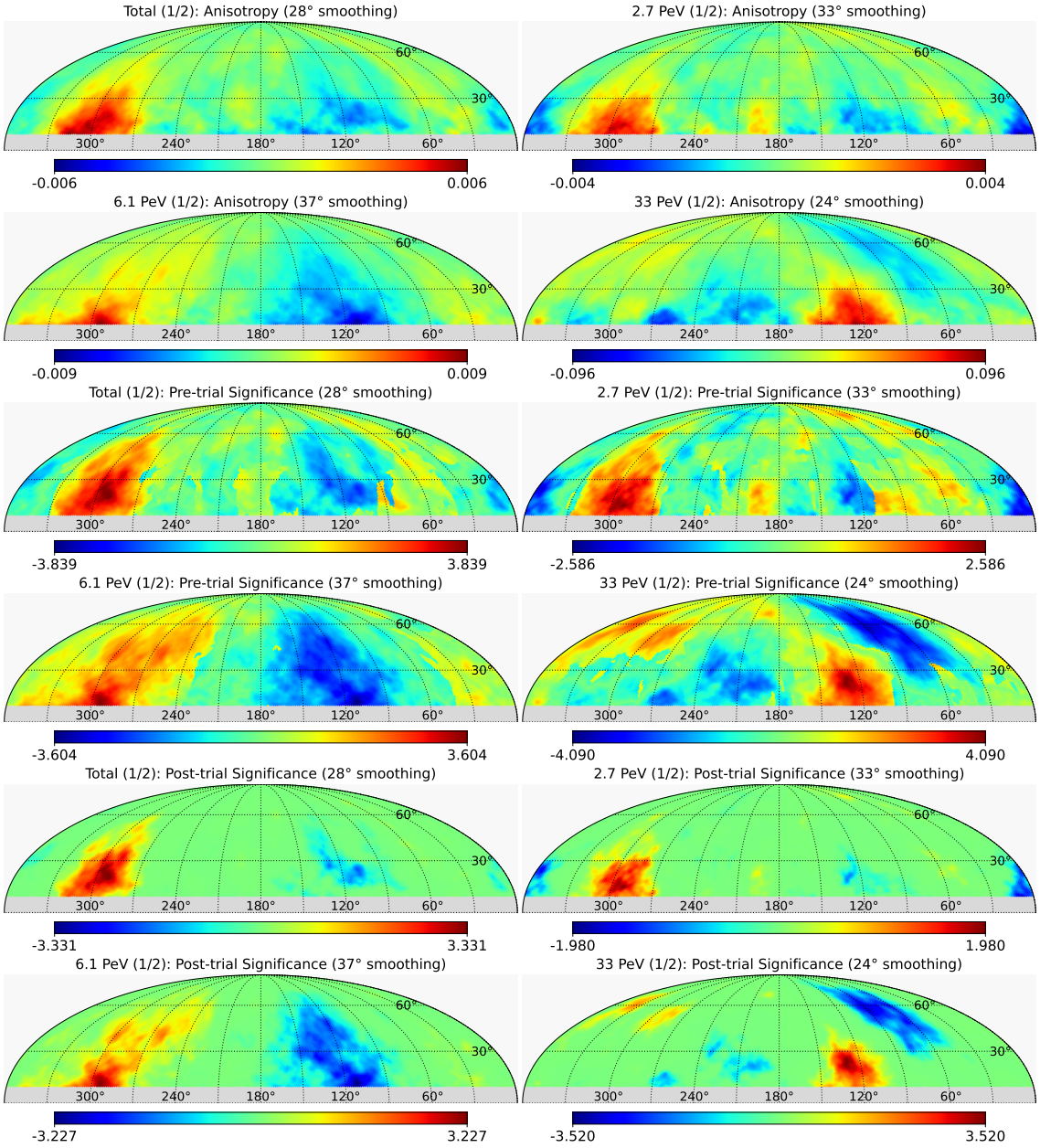


Figure 6.23: Shown are the filtered KASCADE–Grande data with time-ordered subsets, i.e., the 2005–2008 year range, used to check the consistency of anisotropies for each energy bin. Sky maps of anisotropy are presented, including relative intensity (top four panels), pre-trial significance (middle four panels), and post-trial significance (bottom four panels). Each group of four maps corresponds to the total data, the 2.7 PeV bin, the 6.1 PeV bin, and the 33 PeV bin, respectively. Different energy bins use different top-hat smoothing radii, which are determined from the corresponding maximum pre-trial significance in the smoothing-radius scan. The detailed resulting values are shown in Figure 6.18.

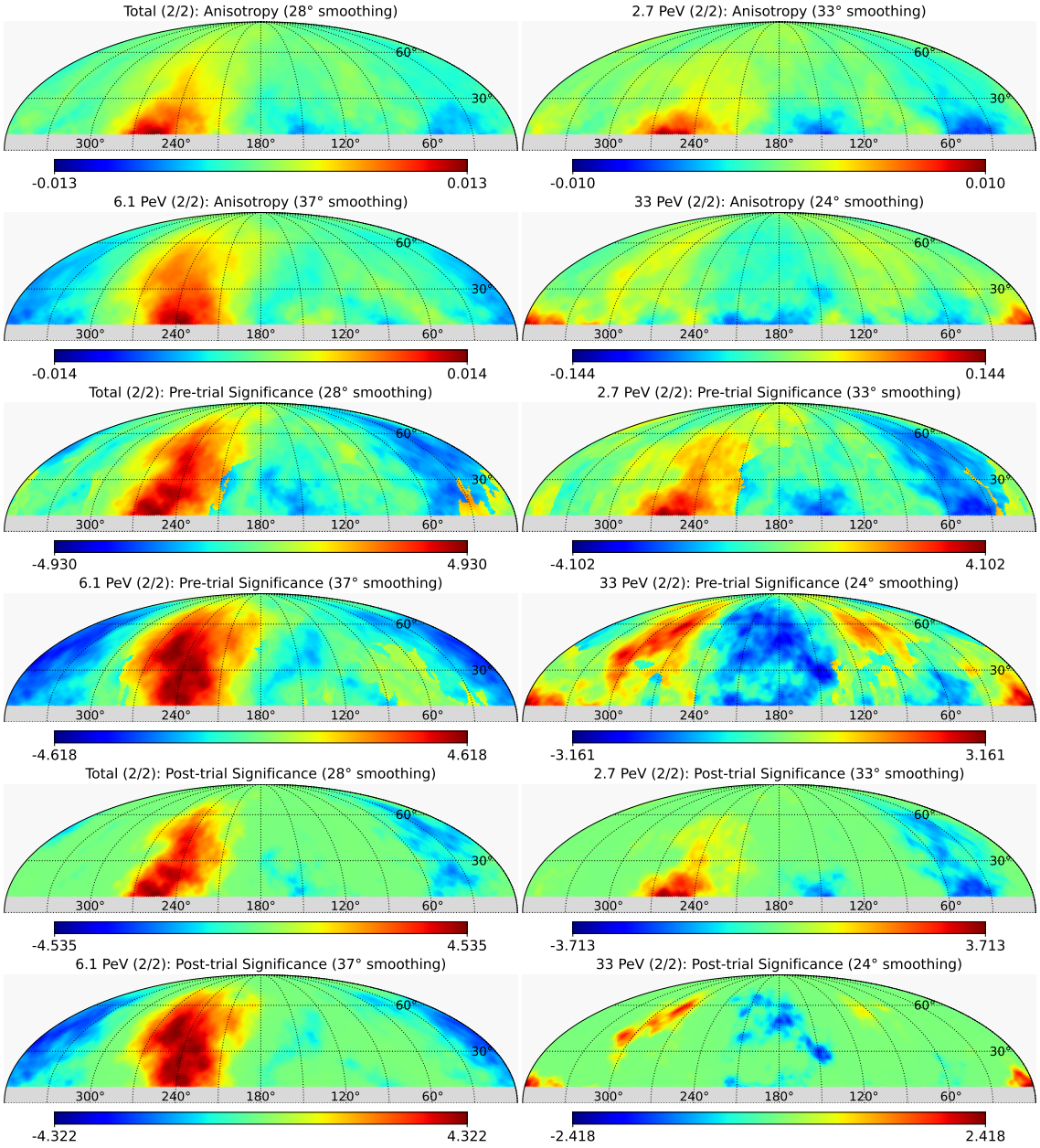


Figure 6.24: Shown are the filtered KASCADE–Grande data with time-ordered subsets, i.e., the 2009–2011 year range, used to check the consistency of anisotropies for each energy bin. Sky maps of anisotropy are presented, including relative intensity (top four panels), pre-trial significance (middle four panels), and post-trial significance (bottom four panels). Each group of four maps corresponds to the total data, the 2.7 PeV bin, the 6.1 PeV bin, and the 33 PeV bin, respectively. Different energy bins use different top-hat smoothing radii, which are determined from the corresponding maximum pre-trial significance in the smoothing-radius scan. The detailed resulting values are shown in Figure 6.18.

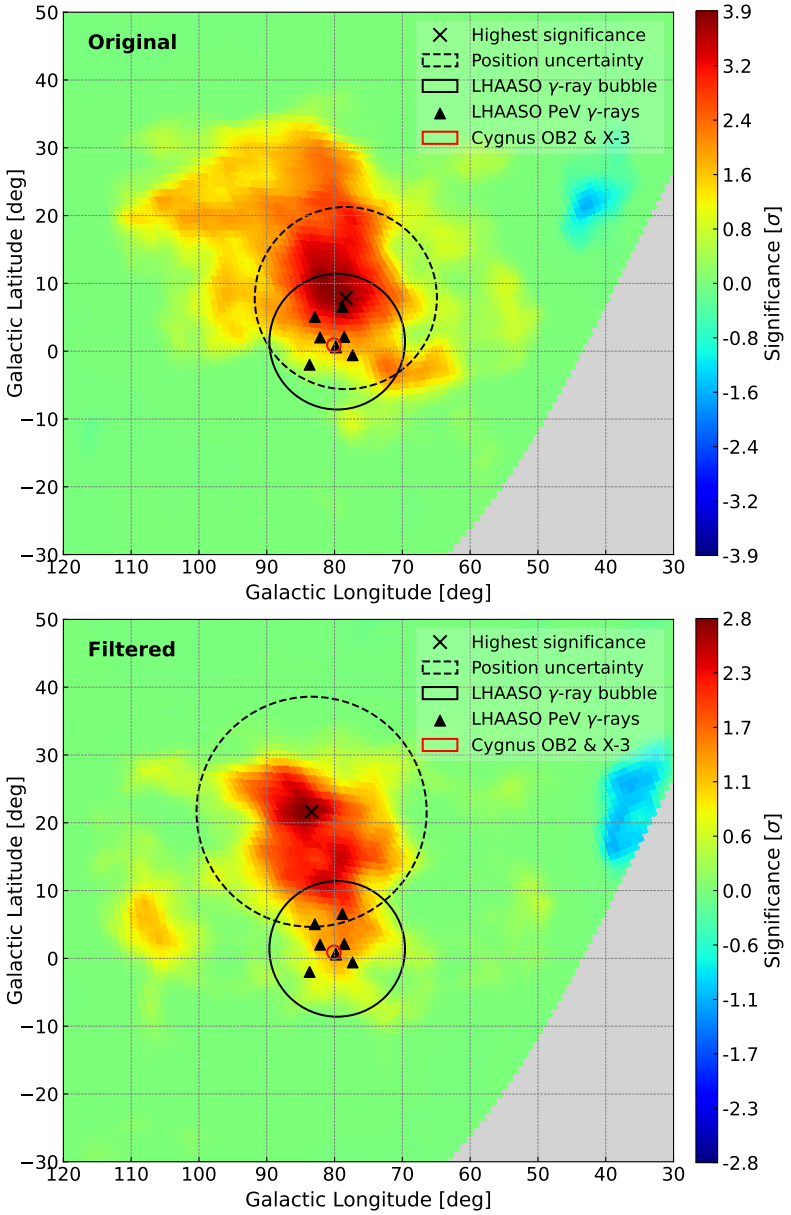


Figure 6.25: Shown are the post-trial significances obtained from the KASCADE–Grande data at the 33 PeV energy bin, together with eight γ-ray events above 1 PeV from the nearby source region Cygnus OB2, as observed by LHAASO. The dashed circles mark the orientation uncertainties, computed from Eq. 6.101 using the smoothing angles. For the top panel, the original data reaches a maximum pre-trial significance of $4.6\,\sigma$ with a smoothing angle of 19° and an uncertainty of $\sigma_s = 13.4°$; the corresponding post-trial significance is $3.9\,\sigma$. For the bottom panel, the filtered data reach $3.5\,\sigma$ with a smoothing angle of 24° and $\sigma_s = 17°$, with the corresponding post-trial significance being $2.8\,\sigma$. The hotspot (maximum) orientation for the original data (top) is located at Galactic coordinates $(l, b) = (78.3°, 7.8°)$, corresponding to equatorial coordinates RA = 298.8° and Dec = 44.2°. For the filtered data (bottom), the maximum is at Galactic coordinates $(l, b) = (83.4°, 21.6°)$, corresponding to RA = 282.2° and Dec = 54.3°.

maps, as discussed in papers [329, 330]. The RMS radius is a measure of the spread of a distribution around its mean, and it corresponds to the standard deviation for a uniformly distributed circular region, which directly corresponds to the standard deviation of the distribution after smoothing. The standard deviation σ_s represents the uncertainty in the position due to the smoothing effect of the tophat function. For a uniform distribution within a circular region of radius θ_0, the mean square radius $\langle\theta^2\rangle$ is calculated as

$$\langle\theta^2\rangle = \frac{1}{\pi\theta_0^2}\int_0^{\theta_0} 2\pi r \cdot r^2\, dr = \frac{\theta_0^4}{4\theta_0^2} = \frac{\theta_0^2}{2}. \tag{6.100}$$

The RMS radius is RMS $= \sqrt{\langle\theta^2\rangle}$, which also represents the effective positional uncertainty,

$$\sigma_s = \sqrt{\langle\theta^2\rangle} = \theta_0/\sqrt{2}. \tag{6.101}$$

The relationship between the RMS value and the standard deviation for uniformly distributed circular regions is well established in statistical mechanics and geospatial analysis [331]. Therefore, in our case, for example, when using the original data at 33 PeV, the maximum pre-trial significance of $4.56\,\sigma$ is reached at a smoothing angle of 19°, corresponding to an uncertainty of $\sigma_s \simeq 13.4°$.

In Galactic coordinates, we found that the hotspot region with maximum post-trial significance, shown in Figure 6.25, shifts by about 15° after the data filtering, i.e., from Galactic coordinates $(l, b) = (78.3°, 7.8°)$, corresponding to equatorial coordinates RA = 298.8° and Dec = 44.2°, to $(l, b) = (83.4°, 21.6°)$, corresponding to RA = 282.2° and Dec = 54.3°. This arises from time gaps in the data and the discontinuous sidereal days, which interrupt and distort the seasonal variations. The hotspot may reveal a correlation with the slight dip observed between the cosmic-ray knee and the second knee, which is strongly linked to the LHAASO γ-ray bubble region located near the Cygnus region [8], with a galactic longitude of approximately 80° and a latitude close to 0°. This could suggest that the origin of the observed excess is non-diffusive and may point to a local source of PeV γ-rays situated in the Cygnus region. These γ-rays could also be the result of high-energy cosmic ray interactions occurring in the vicinity of their sources. LHAASO reported 8 high-energy γ-ray photons with energy ≥ 1 PeV [8], including 2 located in the region of the massive star association Cyg OB2, roughly 0.5°. Specifically, the Cyg OB2 association has an approximate radius of about 2 degrees on the sky, the center is located at $(l, b) = (79.8°, 0.8°)$. This association is one of the largest and most massive known stellar associations, and it spans a significant area in the constellation of Cygnus. The core massive region of Cyg OB2, where the density of massive stars is highest, is often described with a smaller radius, around 0.5° [8]. This core region covers the known Cyg X-3 and a powerful pulsar PSR J2031+4127. As shown in Figure 6.25, the Cyg OB2 is within the KASCADE-Grande access region, which can be verified through the propagation of charged particles from the Cygnus region, such as Cyg X-3 or PSR J2031+4127, and reach on Earth and distribute the arrival directions of CRs on the celestial sphere.

6.4.3 Isotropy Test

The previous analytic method in Eq. (6.97) introduces an approximate trials factor by assuming N_{trial} independent tests within the significance map to compute p_{post}. This provides a quick estimate of how the global significance is reduced by multiple comparisons. However, since smoothing causes nearby pixels to become correlated and not fully independent, this method can slightly overestimate the real global significance and should

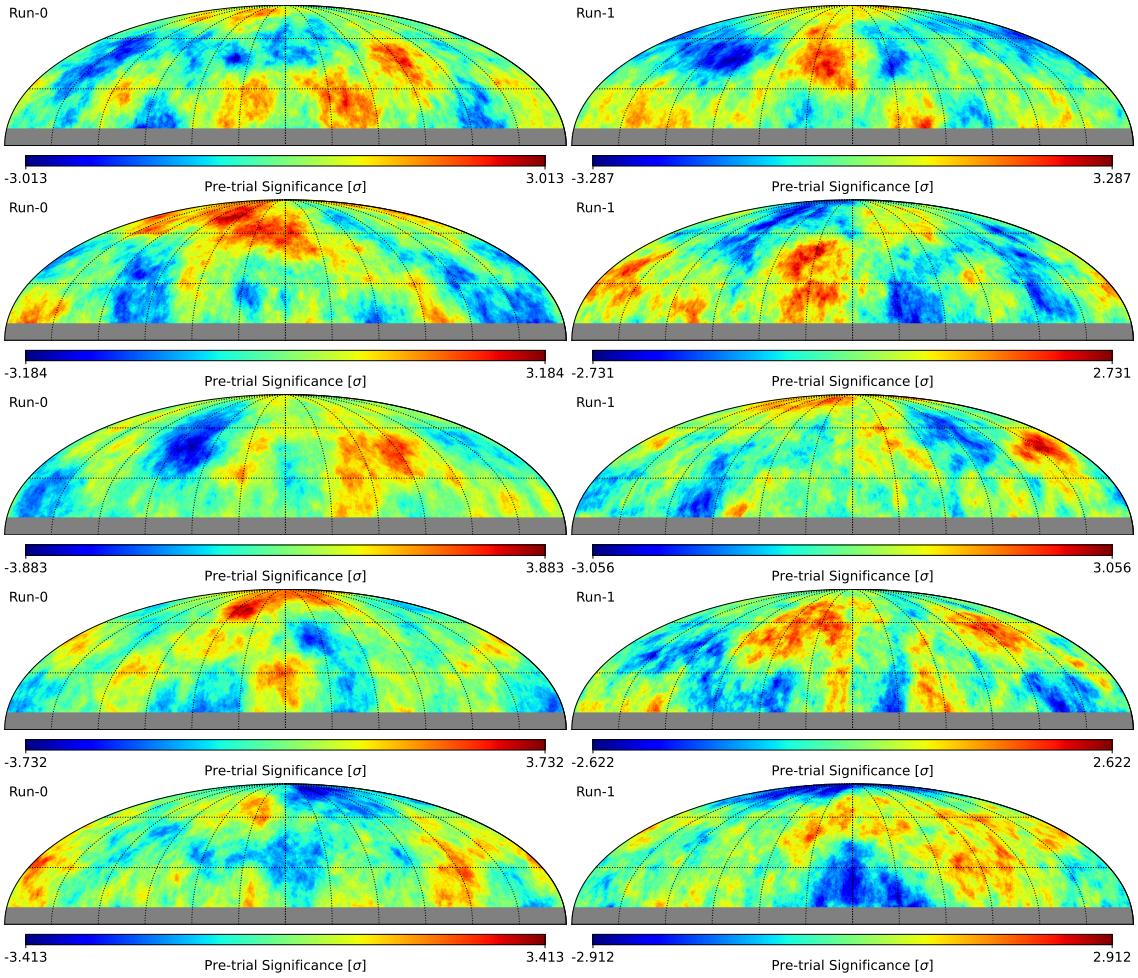


Figure 6.26: Sky maps of anisotropy (relative intensity), pre-trial significance and post-trial significance with KASCADE-Grande FoV, showing 10 examples out of 100,000 generated maps. All maps are smoothed using a 20° top-hat radius.

therefore only be regarded as an approximate and practical estimation.

A more accurate way to evaluate the global significance is to use Monte Carlo (MC) isotropy tests, following the approach used by the Auger Collaboration [279]. In this procedure, we generate a large ensemble of mock data sets with purely isotropic arrival directions and process each of them through the exactly same reconstruction and smoothing steps as the real data. For every simulated sky, we record the maximum significance found anywhere on the sphere after the full smoothing–radius scan, thereby effectively capturing the look–elsewhere effect (LE). Repeating this many times yields the sampling distribution of the maximum significance under isotropy, and we count how often an isotropic sky produces a maximum significance equal to or greater than that observed in the data. Concretely, let S_{data} denote the largest pre-trial significance found in the data after scanning over sky position and smoothing radius. From N isotropic realizations, let k of them produce a maximum significance $S \geq S_{\mathrm{data}}$. The global (post–trial) p–value is then $p_{\mathrm{post}} = k/N$, which we convert to a one-sided (excess only) Gaussian equivalent significance Z_{post} via $p_{\mathrm{post}} = 0.5\,\mathrm{erfc}(Z_{\mathrm{post}}/\sqrt{2})$. The binomial counting uncertainty on p_{post} is estimated as $\sigma_p \simeq \sqrt{p_{\mathrm{post}}(1-p_{\mathrm{post}})/N}$ and propagated to Z_{post} using $dp/dZ = -\exp(-Z^2/2)/\sqrt{2\pi}$.

In the 33 PeV bin, we identify hotspots with local pre-trial significance $S_{\rm local} = \tilde{S}_{\mathfrak{a}} > 4.6\sigma$ for the original data and $S_{\rm local} = \tilde{S}_{\mathfrak{a}} > 3.51\sigma$ for the filtered data, using the definition in Eq. (6.96). We then generate $N = 10^5$ isotropic maps with the same total event counts as the data sets ($N_{\rm evt} = 1.1 \times 10^5$ for the original and $N_{\rm evt} = 9.6 \times 10^4$ for the filtered sample). Each MC realization is processed identically to the data: same pixelization, same number of events (with acceptance ignored), and the same top–hat smoothing radius. Ten representative isotropic examples are shown in Figure 6.26. Among all maps, $k = 558$ and $k = 15,922$ of them produce a maximum significance $\geq S_{\rm data}$, yielding post-trial probabilities of $p_{\rm post} = 5.58 \times 10^{-3}$ and 0.159 for the original and filtered data, which correspond to $Z_{\rm post} \approx 2.8$ and 1.4. These translate into global significances of $2.8\,\sigma$ and $1.4\,\sigma$. This result indicates that the observed medium-scale excesses are compatible with rare statistical fluctuations in an isotropic sky, yet they provide a localized hint of a possible signal. After accounting for the look–elsewhere effect from scanning over both sky position and smoothing radius, the excess reaches global significances of $\sim 2.8\,\sigma$ (original data) and $\sim 1.4\,\sigma$ (filtered data). Further propagation simulations are necessary to evaluate the consistency of the results considering the biases present in the original and filtered data.

6.5 Simulation of Propagation

Although the exact flux fraction contributed by Cygnus OB2 relative to the diffuse Galactic background remains uncertain, the central question is whether the arrival directions expected from this region align with the KASCADE-Grande field of view (charged particles as CRs). Since any individual source can supply only a limited fraction of the total observed spectrum, it becomes essential to estimate the relative contribution that Cygnus OB2 could plausibly make. The previous anisotropy analysis demonstrated that the Cygnus OB2 region falls well within the location of the 3 PeV cosmic-ray excess observed by KASCADE-Grande, with the hotspot offset by only about 10° from Cygnus OB2. We therefore employ the CRPropa simulation framework[1] and its documentation[2], both of which are based on the methodology described in paper [178], to model the propagation of cosmic rays originating from the powerful pulsar PSR J2031+4127 in the Cygnus OB2 region, and compare the results with the 33 PeV excess (see Figure 6.25 as well as 6.19 and 6.22). Next, we generate sky maps of the simulated arrival directions exclusively for the 33 PeV bin only, since the 2.7 PeV and 6.1 PeV bins involve much slower cosmic-ray propagation, and compare these maps to the anisotropies observed in the same 33 PeV bin of the KASCADE-Grande data, focusing on the hotspot region as seen from Earth. Finally, we evaluate whether the distribution of cosmic-ray protons originating from the Cygnus OB2 region falls close to the field of view accessible to KASCADE–Grande to account for part of the observed excess.

As mentioned previously, Cygnus OB2 spans a radius of approximately 2° on the sky, centered at $(l, b) = (79.8°, 0.8°)$. It is one of the largest and most massive stellar associations in the Cygnus region. Its dense core is typically characterized by a smaller radius of about 0.5° [8], encompassing Cyg X-3 and the energetic pulsar PSR J2031+4127 in it. As shown in Figure 6.27, the central region of Cygnus OB2 is clearly visible within the field of view, together with diffuse hot gas and nearby structures, though not the full extent of the Cygnus OB2 association.

[1] See https://github.com/CRPropa.
[2] See https://crpropa.github.io/CRPropa3/pages/Installation.html.

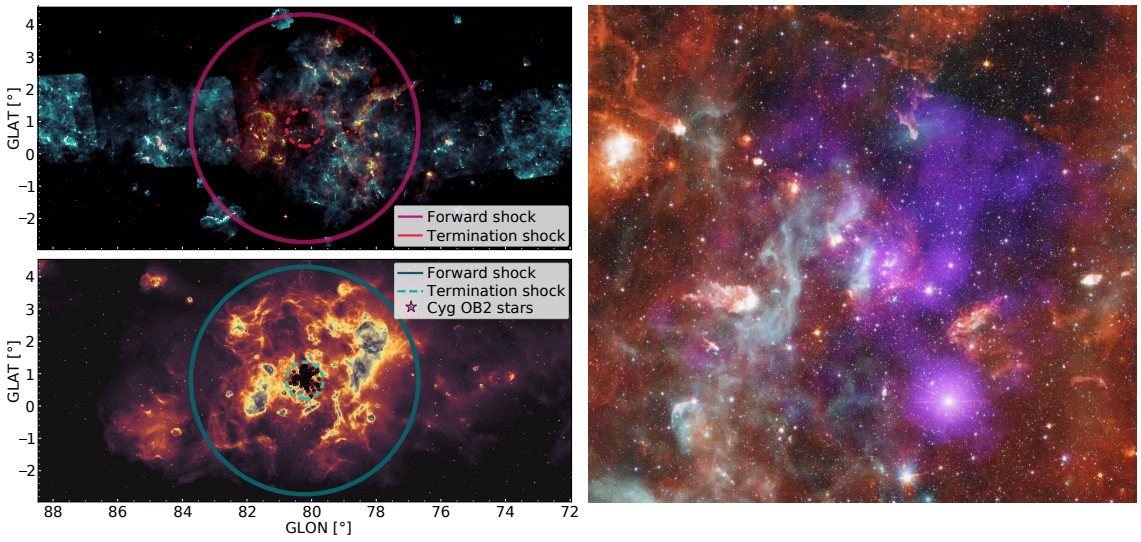


Figure 6.27: Left: Cygnus X and Cyg OB2. Top: 500 µm (blue) and 8 µm (orange) emission showing cold dust and photon-dominated regions, observed by the Herschel/SPIRE and MSX telescopes. Bottom: 1.42 GHz continuum map, mostly thermal free–free emission from gas likely ionized by Cyg OB2, from the Canadian Galactic Plane Survey. Both panels cover the same Galactic field. Image adapted from [332]. Right: Composite image from the Chandra X-ray Observatory centered on Cyg OB2, showing its dense stellar core (bright purple and blue-white region near the centre). Image credit: NASA.

6.5.1 Flux from Observation

Before estimating the relative fractions between the background and the total detected flux, we begin by examining in detail the nuclei fractions measured by KASCADE–Grande [119]. At around 33 PeV, the elemental abundances observed at Earth cannot be directly interpreted as the composition of particles arriving from any individual astrophysical region, including potential sources located in Cygnus OB2. This is because the measured flux at Earth is an integrated superposition of contributions from many sources spread across different distances and sky directions, each undergoing various levels of propagation effects during transport through the Galaxy, such as rigidity-dependent escape (Galactic leakage) and spallation interactions that break up heavier nuclei like silicon or iron. Consequently, the observed composition represents a complex, transport-weighted mixture of all contributing sources rather than the intrinsic composition of any single acceleration site.

The detailed elemental fractions are obtained from the values published in the KASCADE–Grande composition analysis [119], supplemented by the KCDC dataset[3] (see Section 6.1), and are shown in Fig 6.28. From these results, we observe that light nuclei dominate the spectrum at lower energies and continue to constitute a major fraction even above $E > 10^{7.2}$ GeV. Meanwhile, the proton component remains relatively small when compared to the contributions from heavier mass groups, whereas iron becomes increasingly prominent as energy increases. In addition, the inferred fractions depend on the hadronic–interaction model used in the air–shower reconstruction (e.g., QGSJet, EPOS), on the unfolding methodology used to recover the primary composition, and on uncertainties in the absolute energy calibration of the experiment. A rigidity-ordered

[3] See https://kcdc.iap.kit.edu/spectra/.

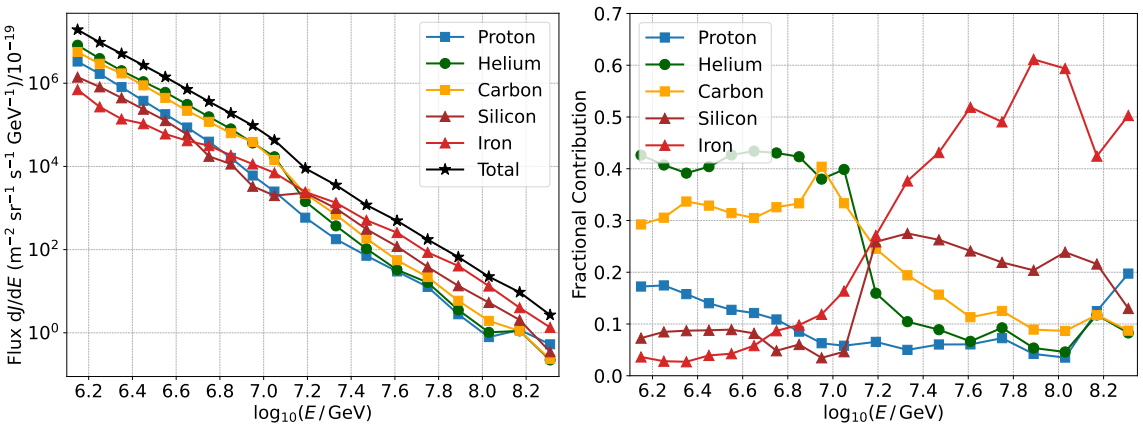


Figure 6.28: Left: cosmic-ray flux from the published KASCADE–Grande measurements [119], together with the data provided in the KCDC dataset. Right: the corresponding composition fractions derived from these datasets, illustrating the relative contributions of the various nuclear mass groups across the relevant energy range.

sequence of spectral knees, where the cutoff energies follow approximately $E_{\mathrm{knee}} \propto Z$, is consistent with the observed trends and further complicates a direct mapping from the Earth–observed mixture to any single source population.

Because the measured flux includes contributions from all directions in the sky, the specific part coming from the Cygnus OB2 region can only be studied using detailed modeling. In this approach, we simulate how particles are produced at the source (their spectra and cutoffs), how they travel through the Galaxy, taking into account diffusion and magnetic turbulence, but ignore their interaction with other nuclei on the way. These propagation simulations are then compared with the detector data to separate a possible Cygnus OB2 signal from the overall background of cosmic rays at $\mathcal{O}(1\text{–}100)$ PeV.

6.5.2 Propagation Settings

For the propagation simulations using CRPropa, to reduce CPU time in simulations of cosmic-ray propagation from Cygnus-OB2 to Earth, we adjusted the integration step sizes to 0.1 pc, which decreases the number of integration steps in smoothly varying regions. The integration step is set from 0.01 pc to 10 pc when propagating charged particles through the field. We set a spherical boundary centred at the Galactic center (the origin) with a radius of 20 kpc, removing any particles that propagate beyond this distance from the simulation. The rigidity of iron is very small compared to the lighter elements. It would experience large deflection angles continuously in the Galaxy, which are difficult to simulate, require a significant amount of CPU time and collect less from the Earth as an observer, being much more isotropic compared to the medium components (He+C+Si) group, whose flux is roughly the same as iron's. In contrast, the proton flux is 8 times less than that of iron and the light group. For heavier nuclei, we imposed effective propagation boundaries based on their Larmor radii to reduce CPU time.

In this analysis, we do not account for the Compton-Getting effect during cosmic-ray (CR) propagation. The cosmic rays reaching Earth from the Cygnus OB2 region can directly reveal the probability of CR distribution on the celestial sphere, influenced by Earth's location and the tilt of its rotational axis relative to the Galactic plane. Since the

Compton-Getting anisotropy is on the order of 10^{-3} (as discussed in Chapter 3), which is smaller than the 10^{-2} magnitude anisotropy observed at 33 PeV, its effect is negligible in our case. We injected 10^6 protons and helium and 10^5 carbon, silicon and iron nuclei from the source, assuming a 0.1 kpc boundary scale around Earth. For the propagation, we used CRPropa [178], incorporating both random striated and turbulent components of the Galactic magnetic field under the JF12 model [333] and the UF23 model [334]. In the UF23 case, we perform simulations using two different models: the "base" model and the "twistX" model, each of which adds additional grid-based turbulence components.

6.5.3 Turbulence in Magnetic Field

The Galactic magnetic field in our simulations is modelled as the superposition of a coherent, large-scale component and an isotropic, small-scale turbulent component [335, 333]. In the JF12 model [335], the turbulent contribution is implemented as a purely stochastic perturbation that does not possess an explicitly defined power spectrum or correlation length, meaning that it captures random magnetic fluctuations without following any physically motivated scale-dependent structure. The large-scale coherent, dust, and halo structures are specified, while the small-scale turbulence is added as a parametric random term. In CRPropa, if one needs explicit spectral control, one usually adds a synthetic turbulence module [333, 336]. It provides a statistical description of magnetic fluctuations but does not include resonant power at physical scales relevant for cosmic-ray scattering. In contrast, the UF23 model [334] provides updated coherent-field solutions constrained by recent data. In this work, we make use of both the UF23-*base* configuration, which provides the minimal coherent-field solution, and the UF23-*twistX* configuration

To describe the turbulent field more realistically, we use a hybrid approach that combines a Sun-centered grid turbulence for large eddies and a plane-wave superposition for small scales [178, 337] (also see the introduction webpages[1][2]). This setup captures field-line wandering on scales $\gtrsim$ 50 pc and pitch-angle scattering down to sub-pc scales, which are important for heavy nuclei at PeV–EeV energies [338, 339]. The turbulence grid is centered at the solar position $(x, y, z) = (-8.5,\ 0,\ 0)$ kpc. We use a cubic box of half-size 2 kpc (a 4 kpc cube) with 256^3 cells, giving a spatial resolution of $\Delta x \simeq 15.6$ pc. The grid turbulence follows a Kolmogorov spectrum, $P(k) \propto k^{-5/3}$, between $\ell_{\rm min} = 50$ pc and $\ell_{\rm max} = 300$ pc [339, 340], where ℓ denotes a spatial length scale (the inverse of the wavenumber $k = 2\pi/\ell$) in the turbulence spectrum. The minimum scale is chosen to be larger than 2–3 grid cells to avoid aliasing, while the maximum scale extends to large Galactic eddies. The RMS (root-mean-square) strength of the grid component is $B^{\rm grid}_{\rm rms}$, which represents the large-scale part of the turbulent field. Because the mean field $\langle \mathbf{B} \rangle$ in turbulence is zero, only the variance $\langle B^2 \rangle$ is physically relevant, giving $B_{\rm rms} = \sqrt{\langle B^2 \rangle}$. These keep the smallest turbulent structures well above the grid resolution limit and produce turbulence ~ 100 pc, which is relevant for large-scale trajectories near the Sun.

Small-scale scattering is added through a plane-wave superposition method in which N_m random modes are generated, each assigned an independently chosen phase, polarization state, propagation direction, and amplitude drawn from a common Kolmogorov spectrum extending over the interval $\ell^{\rm small}_{\rm min}$ to $\ell^{\rm small}_{\rm max}$. We set $\ell^{\rm small}_{\rm min} = 0.3$ pc (tightenable to 0.2 pc if needed) and $\ell^{\rm small}_{\rm max} = 50$ pc, so that the small-scale band overlaps the grid's lower end and extends well below the Larmor radii of heavy nuclei at tens of PeV. Unless noted otherwise,

[1] CRPropa MF: https://crpropa.github.io/CRPropa3/buildingblocks/MagneticFields.html.
[2] https://crpropa.github.io/CRPropa3/api/classcrpropa_1_1PlaneWaveTurbulence.html.

we use $N_m = 1024$, which provides an approximately isotropic realization and stable diffusion coefficients while keeping runtime reasonable. The plane-wave RMS strength is $B_{\rm rms}^{\rm pw}$. We adjust $(B_{\rm rms}^{\rm grid}, B_{\rm rms}^{\rm pw})$ so that the total turbulent RMS near the Sun satisfies,

$$B_{\rm rms}^{\rm tot} \approx \sqrt{\left(B_{\rm rms}^{\rm grid}\right)^2 + \left(B_{\rm rms}^{\rm pw}\right)^2} \simeq 3\,\mu{\rm G}, \tag{6.102}$$

which consistent with local ISM estimates [338, 339, 341]. We choose a balanced split, e.g., $B_{\rm rms}^{\rm grid} \approx B_{\rm rms}^{\rm pw} \approx 2.1\,\mu$G. This hybrid setup is motivated by considerations of rigidity. For a particle of charge Z and energy E, the Larmor radius in a μG field is

$$r_L \simeq 1.08\,{\rm pc}\,\frac{E/{\rm PeV}}{Z\,B_{\mu{\rm G}}}. \tag{6.103}$$

At $E = 33\,$PeV and $B \simeq 3\ \mu$G, we have $r_L \sim 12\,$pc for protons and $r_L \sim 0.5\,$pc for iron. The grid turbulence (with $\ell_{\rm min} = 50\,$pc) correctly captures large-scale meandering but contains little to no power at $\sim 12\,$pc and cannot represent sub-pc resonant modes. The plane-wave component fills this gap, supplying power from 0.3 pc to 50 pc and thereby enabling resonant pitch-angle scattering for both 33 PeV protons and iron [337, 339, 178]. Finally, one can compare different nuclei at fixed rigidity $R = E/Z$ [339, 342].

6.5.4 Injection Spectrum

Cygnus OB2 is a massive OB association located in the constellation Cygnus. It is one of the richest and most compact concentrations of young, massive stars in the Milky Way and is considered a potential source of high-energy cosmic rays. The source's characteristics—its location, age ($\sim 10^7$ years), stellar wind speed ($3000\,{\rm km\,s^{-1}}$), and collective mechanical power ($\sim 10^{39}\,{\rm erg\,s^{-1}}$), make it an ideal cosmic-ray accelerator. It can inject GeV–PeV protons into the surrounding medium, powering both the UHE γ-ray bubble and the 33 PeV cosmic-ray excess observed by KASCADE-Grande in this region [8]. In Galactic coordinates, Cygnus OB2 is located at approximately $l \simeq 80.22°$ and $b \sim 0.79°$ [332], while its equatorial coordinates are roughly R.A. $22^{\rm h}33^{\rm m}$ and Dec $41°28'$. It is positioned at $(-8.256, 1.427, 0.025)\,$kpc in the Galactocentric Cartesian coordinate system. A distance of 1.45 kpc from Earth is adopted for the propagation setup and subsequent analysis.

Before estimating the all-particle flux, we should start with the injection spectrum at the source. First, we take the γ-ray flux and the associated proton flux from the ultrahigh-energy (PeV) γ-ray bubble linked to Cygnus OB2, as reported by the LHAASO collaboration [8]. The model assumes a steady injection of energetic protons over the system age $t_{\rm age}$, originating from a point-like source at the center of the bubble [8]. If Cygnus OB2 is indeed the source of these energetic protons, the system age is estimated to be $t_{\rm age} = 2$–3 million years [8], through the γ-ray observation. The initial injection (or acceleration) spectrum of cosmic rays is modelled as

$$\dot{Q}_{\rm p} = Q_0 E_{\rm p}^{-s} \exp\left(-\frac{E_{\rm p}}{E_0}\right), \tag{6.104}$$

at the source, where $E_0 = 5\,$PeV is the spectral cutoff energy, s is the spectral index. This cutoff is treated as a fitting parameter and does not necessarily indicate the upper limit of the injection spectrum. The normalization factor Q_0 is constrained by the total proton injection luminosity, $L_{\rm p}$, which is estimated to be $L_{\rm p} \simeq 10^{37}\,{\rm erg\,s^{-1}}$, or equivalently $6.24 \times 10^{33}\,{\rm PeV\,s^{-1}}$ [8]. When the injected proton spectrum spans from GeV to multi-PeV energies, the normalization factor is approximately $Q_0 \approx 3.205 \times 10^{32}\,{\rm PeV^{-1}\,s^{-1}}$,

Table 6.8: Distribution of the case if we inject 10^7 protons (H) and 10^6 iron (Fe) per energy bin (other nuclei are not shown), including corresponding median energies.N denotes the number of injected primary particles. For protons, only the first five energy bins contribute effectively, whereas iron extends across all energy bins.

E Bin	1	2	3	4	5	6	7	8	9	10
$E_{\mathrm{m}}^{\mathrm{H}}$ [PeV]	17.20	25.30	35.40	50.10	70.00	98.00	138.00	197.00	280.00	400.00
$E_{\mathrm{min}}^{\mathrm{H}}$ [PeV]	15.20	21.41	30.17	42.53	59.89	84.36	119.00	167.40	235.90	334.80
$E_{\mathrm{max}}^{\mathrm{H}}$ [PeV]	21.41	30.17	42.53	59.89	84.36	119.00	167.40	235.90	334.80	468.00
N_{H}	8,500,000	1,370,000	130,000	5,750	400	50	0	0	0	0
$E_{\mathrm{m}}^{\mathrm{Fe}}$ [PeV]	17.68	24.89	35.02	49.26	69.24	97.25	136.42	191.08	267.10	372.46
$E_{\mathrm{min}}^{\mathrm{Fe}}$ [PeV]	15.20	21.41	30.17	42.50	59.87	84.34	118.82	167.39	235.81	332.20
$E_{\mathrm{max}}^{\mathrm{Fe}}$ [PeV]	21.41	30.17	42.50	59.87	84.34	118.82	167.39	235.81	332.20	468.00
N_{Fe}	411000	253000	152000	88770	49450	25840	12350	5210	1850	530

calculated using $Q_0 = L_{\mathrm{p}} / \int_{E_1}^{E_2} E\dot{Q}\mathrm{d}E$, with integration limits $E_1 = 1\,\mathrm{GeV}$ and $E_2 = 10\,\mathrm{PeV}$. Expressed entirely in GeV units, the normalization becomes $Q_0 \approx 1.602 \times 10^{39}\,\mathrm{GeV}^{-1}\,\mathrm{s}^{-1}$. The proton injection spectrum follows an exponential cutoff of E_0 around 5 PeV, $s = 2.25$, as LHAASO estimates. To extend this model to nuclei with charge Z_e, we assume that the acceleration mechanism is rigidity-dependent, such that the cutoff energy for a nucleus scales as $Z_e E_{\mathrm{cut,p}}$ [111], where $E_{\mathrm{cut,p}} = E_0$ is the cutoff energy of a proton. Under this assumption, the injection spectrum for a nucleus A can be written as

$$\dot{Q}_{\mathrm{A}} = \eta_A\, Q_0\, E^{-\alpha} \exp\left(-\frac{E}{Z_e E_0}\right), \tag{6.105}$$

where η_A is the abundance of nucleus A relative to protons, which can be influenced by R-dependent acceleration and escape, nuclear decay, and photo-disintegration processes. For instance, helium typically has an abundance in the range $\eta_{\mathrm{He}} \sim 0.1$–$0.2$, depending on the source environment. For simplicity in the propagation process, we set the number of total injected protons as 10^6, following the injection spectrum $\dot{Q}_{\mathrm{p}}$, with energy $\lg(E/\mathrm{GeV})$ within $(7.182, 8.670)$, i.e., from 15.2 PeV to 468 PeV to cover the median energy of 33 PeV. This range ensures that both the core and the tail of the expected source spectrum are sampled in the simulations, giving a direct comparison between the simulated and observed fluxes. The energy ranges of the medians for other nuclei are given at the end of Section 6.1.

6.5.5 Arrival Directions on Earth

After injecting and propagating protons from the Cygnus OB2 region, we collect all protons that reach the assumed detection sphere of radius (0.1 kpc centered on the solar position. We first examine these arrival directions before applying any anisotropy reconstruction to visualize the raw distributions. Figures 6.29 and 6.30 show these sky maps in Galactic and equatorial coordinates, using plain event counts in each pixel. The three panels in each figure compare the arrival distributions for the JF12 model and for the *base* and *twistX* configurations of the UF23 model. We observe clear differences between the arrival-direction distributions produced by JF12 and those obtained from the UF23 models. In the next section, we will analyze the detailed anisotropies of the simulated arrival directions using relative-intensity maps. Before doing so, we must estimate the signal fraction by studying the effective diffusion coefficient derived from the full set of simulated arrival directions.

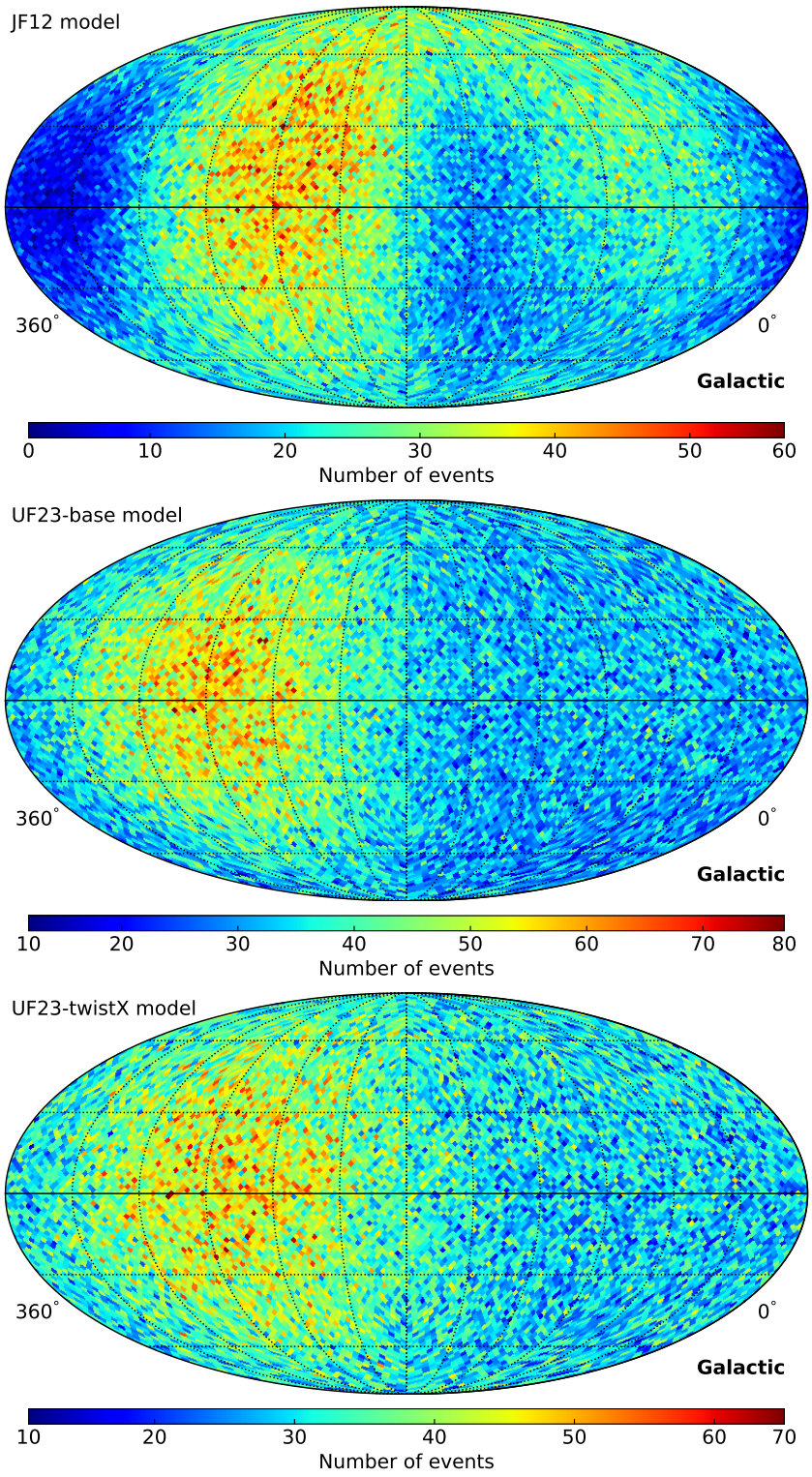


Figure 6.29: Collected protons on Earth, assuming an Earth-centered detection sphere with a radius of 0.1 kpc, obtained from protons injected and propagated from the Cygnus OB2 region. The sky maps are in galactic coordinates.

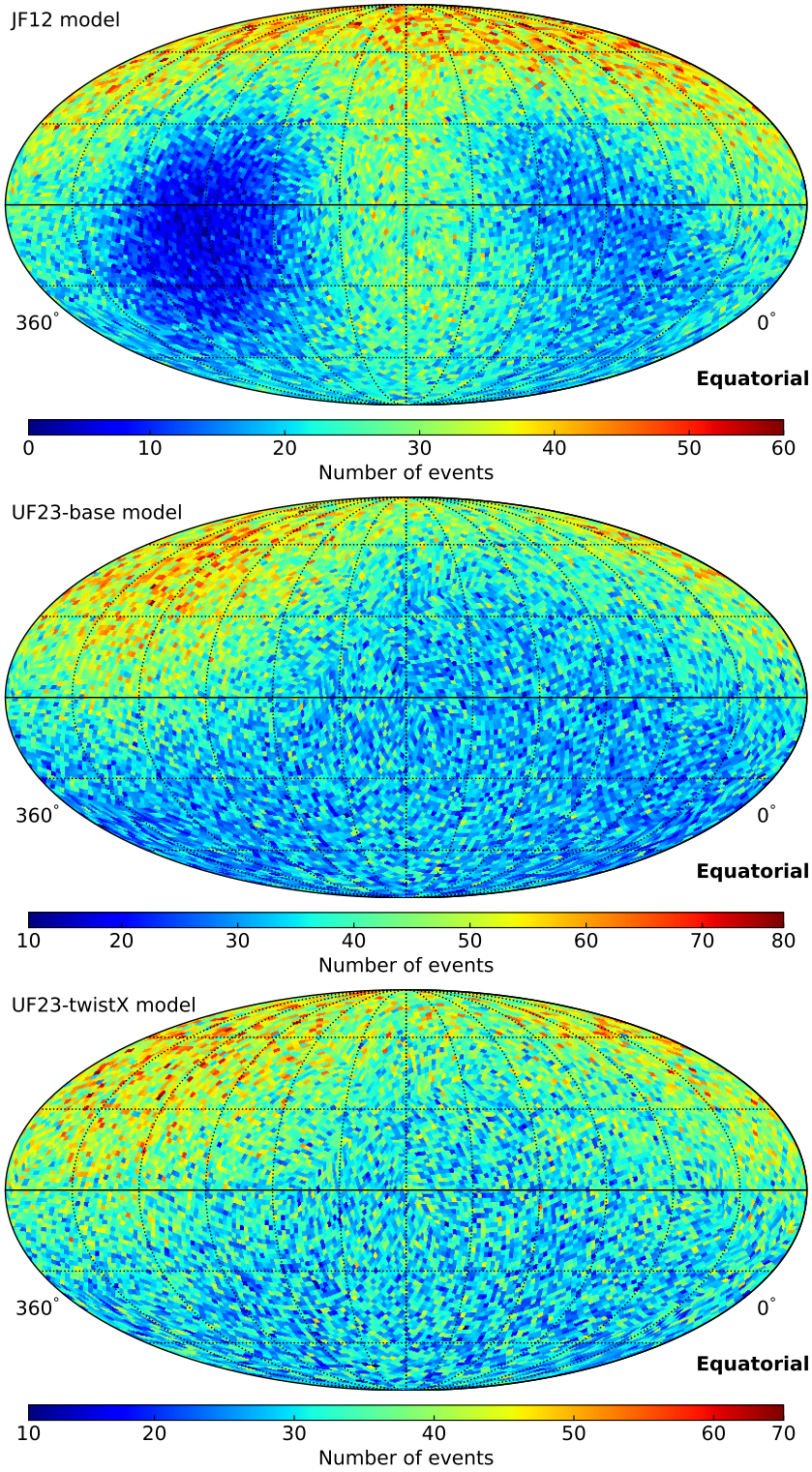


Figure 6.30: Collected protons on Earth, assuming an Earth-centered detection sphere with a radius of 0.1 kpc, obtained from protons injected and propagated from the Cygnus OB2 region. The sky maps are in equatorial coordinates.

6.5.6 Diffusion Coefficient

To estimate the flux from the source (Cygnus OB2 region) on Earth, we have to record the arrival time (i.e., the total propagation time starting from the source) of the particles. Because the numerical experiment models an impulsive, one-second release, the resulting distribution of arrival times corresponds directly to the Green function $G(E, t, \mathbf{r})$, i.e., the probability density that a single particle emitted at the source arrives on Earth after a delay t with energy E. In the case of isotropic diffusion in three spatial dimensions, the Green function represents the probability density that a particle released at the origin at time $t = 0$ is later found at the position $\mathbf{r}$ at time t [19, 343]. For an impulsive (instantaneous) point source, the Green function is given by

$$G(E, \mathbf{r}, t) = \frac{1}{(4\pi D(E)\, t)^{3/2}} \exp\left(-\frac{|\mathbf{r}|^2}{4D(E)\, t}\right), \tag{6.106}$$

where $\mathbf{r}$ is the position vector of the source, t is the time since injection, and D is the spatial diffusion coefficient. In realistic cosmic-ray scenarios, the diffusion coefficient is energy-dependent, typically modelled as

$$D(E) = D_0 \left(\frac{E}{E_0}\right)^{\delta}, \tag{6.107}$$

where D_0 and E_0 are normalization constants and δ is the spectral index, generally in the range 0.3–0.6 based on observational fits or turbulence models [19]. The typical parameters of the diffusion coefficient in Eq. (6.107) are $D_0 = 3 \times 10^{26}\,\mathrm{cm^2\,s^{-1}}$ and $\delta = 0.7$ in fits [8, 344]. For a source that follows an injection spectrum $\dot{Q}(E)$, the resulting cosmic-ray density at position $\mathbf{r}$ and time t is given by $n(E, \mathbf{r}, t) = \dot{Q}(E)G(\mathbf{r}, t, E)$.

In this analysis, cosmic-ray particles that propagate from a single source (SNRs) in the Cygnus OB2 region are assumed, and a total of 10^7 protons with energies ranging from 15.2 PeV to 468 PeV are injected for propagation, see Table 6.8. These arrival particles, shown in the histograms, are simulated assuming an Earth-centred bubble with a radius of 0.1 kpc. The energy-dependent diffusion coefficients can be fitted using the distribution of arriving cosmic-ray particles, as shown in Table 6.9. The parameter D_0 is obtained by fitting the arrival-time histogram (see Figure 6.31) to the Green's function using the first two energy bins and Eq. (6.107), resulting in $D_0 = (2.09 \pm 0.01) \times 10^{29}\,\mathrm{cm^2\,s^{-1}}$, with $E_0 = 1$ TeV and $\delta = 0.62 \pm 0.03$. This corresponds to a characteristic root-mean-square displacement in three-dimensional isotropic diffusion, $\langle|r|\rangle \equiv \sqrt{\langle r^2\rangle} = \sqrt{6D\,t}$, which is valid in the diffusion regime. For shorter times, the propagation follows a more straight-line trajectory and is bounded by $r \leq ct$. The expression of $\langle r^2\rangle$ is obtained by $\langle r^2\rangle = \langle x^2\rangle + \langle y^2\rangle + \langle z^2\rangle$, and $\langle x^2\rangle = \langle y^2\rangle = \langle z^2\rangle$ under isotropic diffusion, where $\langle x^2\rangle = \int x^2\, G(x, t)\, \mathrm{d}x$. If we assume the source is a plane source, the diffusion away from the plane source is given by the mean value of the distance from the plane $\langle|z|\rangle = 2\sqrt{D\,t/\pi}$ [19], using $\langle|z|\rangle = \int |z| G(z, t) \mathrm{d}z$.

To determine the energy dependence of the diffusion coefficient and recover the Green's function, we first fit the full expression in Eq. (6.106) using all arriving protons. From these fits, we extract the diffusion index δ. As shown in Table 6.8, the first four proton energy bins contribute most to the injected flux, and their corresponding Green's function fits are shown in Fig. 6.31. However, the derived δ values remain highly uncertain because the analytic Green's function assumes idealized conditions—an infinite, homogeneous medium with constant D, negligible energy losses, no boundaries, and fits to the full arrival-time or spatial distributions. A more reliable determination of δ would therefore require a much larger number of simulated particles per energy bin.

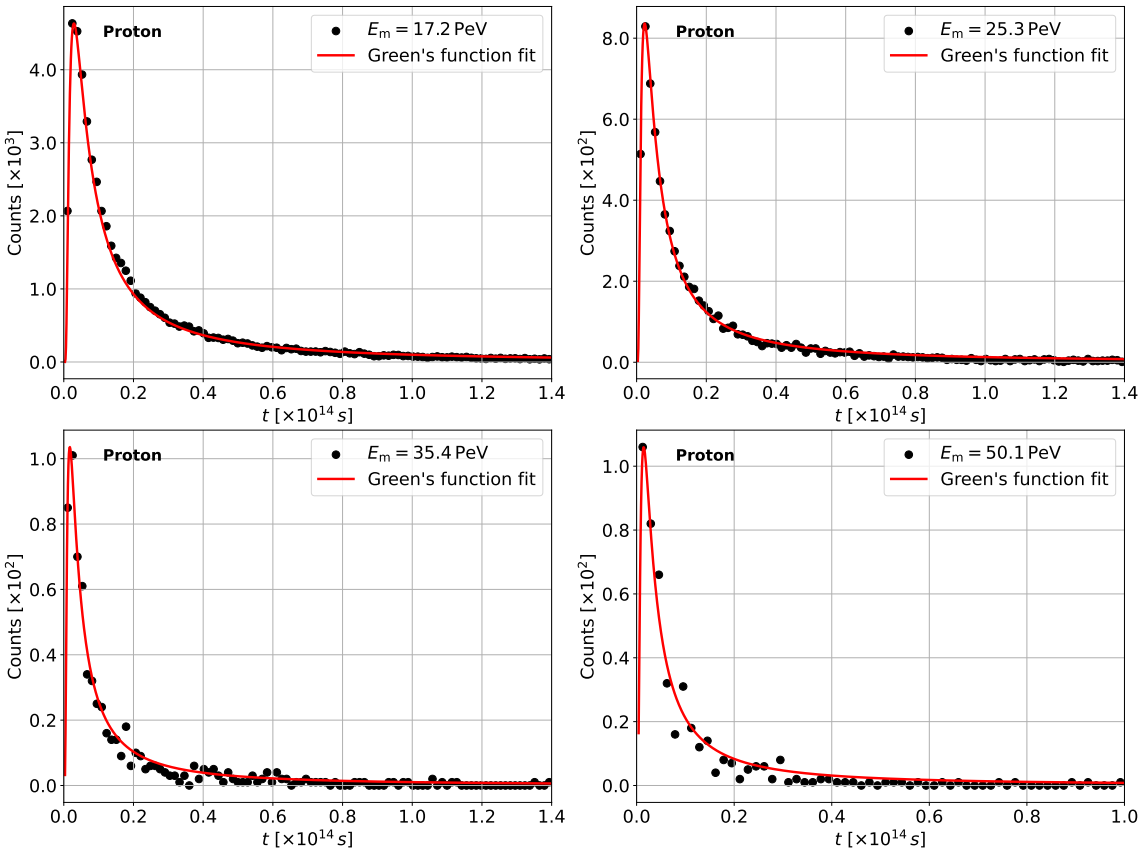


Figure 6.31: Fit of the arrival-time histogram of protons to the Green's function. The 50.1 PeV plot is obtained using 10^4 injected protons, applicable only to this energy bin.

Alternatively, the mean-squared displacement (MSD) approach is simpler and faster, which is the standard method for extracting diffusion coefficients from the propagated particles using CRPropa [345]. In the case of homogeneous and isotropic diffusion, the spatial dispersion of cosmic rays from a point source follows a simple relation between mean squared displacement and propagation time. In spatial d-dimensions, the mean MSD can be obtained via the Green's function, Eq. (6.106),

$$\langle r^2 \rangle = \int r^2 \, G(r,t) \, \mathrm{d}^3 r = 2\, d\, D\, t, \tag{6.108}$$

where D is the diffusion coefficient defined in Eq. (6.107) and t is the particle's arrival time [345]. If particles are emitted from a point source and are detected at t, each with a recorded trajectory r, then on average one has $D = \langle r^2 \rangle / 2\, d\, t$. In the case of three-dimensional space ($d = 3$), the diffusion coefficient simplifies to $D = \langle r^2 \rangle / 6\, t$. Then, we can directly estimate the diffusion coefficient. However, in our simulation, every proton is detected on the same spherical shell at a fixed radius of 0.1 kpc, but they arrive at different times t_i. Because each trajectory's endpoint displacement satisfies $r_i = r$ for all i. As a result, the mean-squared displacement becomes $\langle r^2 \rangle = r^2$, which contains no information about the diffusion process itself. The spread in arrival times, rather than the spread in spatial displacement, carries the relevant diffusion information in this case. Therefore, the standard MSD method cannot be used to estimate D, and a time-dispersion or flux evolution approach must be adopted instead for this fixed-boundary setup.

Table 6.9: Diffusion coefficient extracted from a Green's function fit to the full 33 PeV energy sample, covering 15.2–468 PeV and divided into 10 median energy bins. A total of $N_{\mathrm{KA}}^{\mathrm{tot}} = 146549$ events (unfiltered) are observed above 15.2 PeV, distributed across energy bins according to the fractions $f_{\mathrm{bin}}(E)$. In this analysis, both the Cygnus signal (N_{cyg}) and the KASCADE-Grande sample (N_{KA}) are assumed to consist entirely of protons. The total Cygnus contribution across the energy range is $N_{\mathrm{cyg}}^{\mathrm{tot}} = 2181$.

E Bin	1	2	3	4	5	6	7	8	9	10
$E_{\mathrm{m}}^{\mathrm{H}}$ [PeV]	17.2	25.3	35.4	50.1	70.0	98.0	138.00	197.0	280.0	400.0
$D(E)[10^{30}]$	1.22	1.55	2.07	2.46	×	×	×	×	×	×
$\sigma_D[10^{30}]$	0.0116	0.00754	0.0289	0.0711	×	×	×	×	×	×
$\Phi_{\mathrm{Cyg}}/\Phi_{\mathrm{KA}}$	2.52%	0.77%	0.13%	0.01%	×	×	×	×	×	×
$f_{\mathrm{bin}}(E)$	50.76%	25.06%	12.36%	6.07%	×	×	×	×	×	×
$N_{\mathrm{KA}}(E)$	74395	36729	18115	8902	×	×	×	×	×	×
$N_{\mathrm{Cyg}}(E)$	1876	281	23	1	×	×	×	×	×	×

Now, consider the problem from a different perspective. Each particle contributes exactly one measurement, i.e., its arrival time t_i, so our data set consists of the list t_i, rather than particle positions at a common time. In effect, we sample the diffusion process via the first arrival times at radius r (first passage to a sphere), which defines the appropriate arrival-time probability distribution derived from the diffusion Green's function in Eq. (6.106). The corresponding arrival-time distribution for isotropic diffusion in three dimensions is obtained from this Green's function and exhibits a long-time power-law tail. Similar to the MSD method, we can attempt to describe diffusion over time by using the spread of particle arrival times. However, the mean arrival time, $\langle t \rangle = \int_0^\infty t\, G\, \mathrm{d}t$ diverges at the long-time $t^{-3/2}$ tail of the arrival-time distribution. We therefore employ a likelihood-based approach to the problem. Now, if we detect N particles at times $t_1, t_2, \ldots, t_N$, the overall likelihood of measuring that particular set of times is given by the product of the individual probabilities $f(t_i \,|\, D, r)$. When integrating the Green's function $G(\mathbf{r}, t)$ over time up to infinity, one obtains $\int_0^\infty G(r,t)\mathrm{d}t = 1/(4\pi D r)$. Therefore, upon normalization, the arrival-time probability density can be written as

$$f(t \,|\, D, r) = \frac{r}{\sqrt{4\pi D(E)\, t^3}} \exp\left(-\frac{r^2}{4D(E)\, t}\right). \tag{6.109}$$

The likelihood for observing a particle with arrival times t_i is then given by

$$\mathcal{L}(D) = \prod_{i=1}^{N} f(t_i \,|\, D, r), \tag{6.110}$$

where $f(t \,|\, D, r)$ is the normalized arrival-time probability density derived from the Green's function Eq. (6.106). Taking the natural logarithm on Eq. (6.110) yields the log-likelihood:

$$\ln \mathcal{L}(D) = N \ln r - \frac{N}{2} \ln(4\pi D) - \frac{3}{2} \sum_{i=1}^{N} \ln t_i - \frac{r^2}{4D} \sum_{i=1}^{N} \frac{1}{t_i}. \tag{6.111}$$

To find the maximum-likelihood estimation for D, we differentiate $\ln \mathcal{L}(D)$ regarding D, dropping the terms independent of D, and set the derivative to zero:

$$\frac{\partial}{\partial D} \ln \mathcal{L}(D) = -\frac{N}{2D} + \frac{r^2}{4D^2} \sum_{i=1}^{N} \frac{1}{t_i} = 0. \tag{6.112}$$

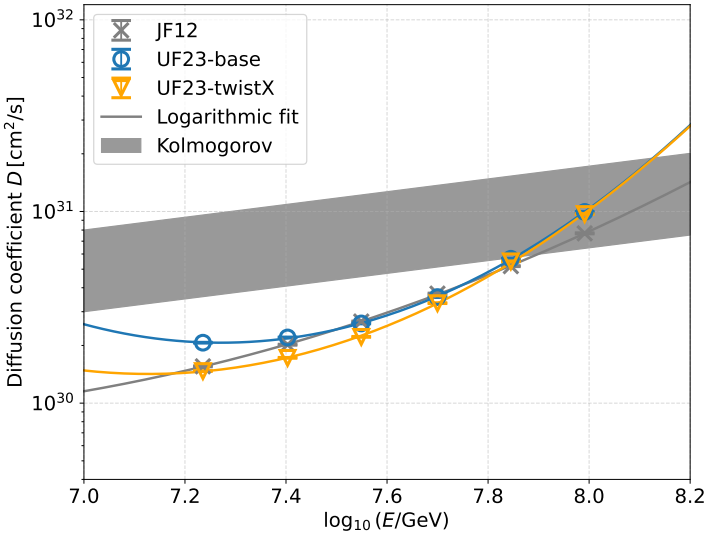


Figure 6.32: The energy-dependent (rigidity) diffusion coefficients using different field models are shown. The green band is the Kolmogorov extrapolation to high-energies based on B/C ratio, with $D_0 = (3-8) \times 20^{28}\,\mathrm{cm}^2\,\mathrm{s}^{-1}$, inferred at $E_0 = 10\,\mathrm{GeV}$ [151, 342].

By solving this equation, we obtain the maximum-likelihood estimator for D,

$$D = \frac{r^2}{2}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{t_i} = \frac{r^2}{2}\left\langle\frac{1}{t}\right\rangle, \tag{6.113}$$

where r is the fixed source-to-Earth distance used in the simulation. By introducing the harmonic mean of the arrival times, $\langle 1/t\rangle$, no divergence occurs. It is equivalent to fitting the full Green's-function shape by maximum-likelihood, but is much simpler to perform numerically and more precise, since it eliminates the need for fine-tuning the fit parameters, as shown in Figure 6.31. The resulting values of D and their corresponding uncertainties for the first four energy bins are shown in Table 6.9.

The ratio between the isotropic turbulence and the regular magnetic field strength is defined as $\eta = B_{\mathrm{rms}}/B_0$, where B_0 is the strength of the regular (uniform) magnetic field, $B_{\mathrm{rms}} = \sqrt{\langle B^2_{\mathrm{turb}}\rangle}$ is the root-mean-square of the turbulent (random) component of the magnetic field B_{turb}. The value of η gives how strong the turbulent field is relative to the regular field and determines the anisotropy level of cosmic-ray propagation through the interstellar medium. It has been found that, for isotropic turbulence in the presence of a uniform regular magnetic field, a turbulence level of $\eta = 0.5$ yields a parallel diffusion coefficient $D_{\parallel}$ that is consistent with values inferred from boron-to-carbon (B/C) ratio measurements [342, 151]. In contrast, higher turbulence levels lead to diffusion coefficients that are too low to match the observed data, strongly indicating inefficient large-scale transport and excessive scattering of charged particles [342, 151]. This suggests that cosmic-ray propagation must be anisotropic, requiring a sufficiently strong regular magnetic field component to reproduce observational constraints. However, the default Galactic magnetic field model in the CRPropa module (JF12Field) includes a turbulent component that is stronger than the regular field at both the Earth and the source location. Specifically, we find $\eta = 3.46$ at Earth and $\eta = 2.33$ at Cygnus OB2, implying that the local field environment is dominated by random fluctuations rather than by a coherent ordered field. Therefore, to reach the regime consistent with B/C-based diffusion, where $\eta = 0.5$, the

field configuration should be rescaled.

For this, we set $\eta \lesssim 0.5$ as a global parameter in the Galaxy. We consider a simulation box spanning $R \in [0.5, 15]$ kpc and $|z| \leq 5$ kpc, which is sufficiently large to capture the full propagation volume of PeV protons or iron nuclei. This choice is motivated by the smallness of the Larmor radius at these energies. For example, a 1 PeV proton in a magnetic field of 1 μG has a Larmor radius of approximately $r_{\rm L} \sim E/eB \sim 1$pc, while an iron nucleus at the same energy has $r_{\rm L}$ smaller by a factor $\sim Z$ (i.e., ~ 0.04, pc for $Z = 26$). Moreover, we chose vertical extent $|z| \leq 5$ kpc, which corresponds to the expected size of the Galactic cosmic-ray halo. Most importantly, the anisotropy systematics discussed in paper [346] are broadly consistent with an energy-dependent diffusion coefficient $D(E)$, as illustrated in Fig. 6.32. Through the relation in Eq. (1.33), an effective slope $\beta \lesssim 1/3$ reproduces the observed dipole-amplitude trend from 10 TeV up to nearly 40 PeV, as illustrated in Figure 6.17. It should be noted that δ in Eq. (1.33) is the transport index governing $D(E) \propto E^{\delta}$, whereas β is the observed dipole–amplitude index $A(E) \propto E^{\beta}$; typically $\beta \approx \delta$, unless energy-dependent source gradients or geometry break this equivalence.

6.5.7 Estimation of Source Flux

Formally, the cosmic-ray density at Earth can be expressed as a convolution of the source injection function $\dot{Q}(E)$ with the Green's function $G(E, \mathbf{r}, t)$ that describes propagation through the interstellar medium [343]. Specifically, it is given by:

$$n(E) = \int_{t_{\rm min}}^{T_{\rm age}} \dot{Q}(E)\, G(E, \mathbf{r}, t)\, {\rm d}t, \tag{6.114}$$

where $|\mathbf{r}| = 1.45$ kpc as mentioned previously. From the distribution of arrival time of protons in the simulation, the source age is adopted to be $T_{\rm age} = 3$–5 million years under the assumed injection spectrum of protons. The shortest possible propagation time for a PeV particle travelling in a straight line over a distance of approximately 1.5 kpc is about 1.54×10^{11} seconds, or 4.9×10^{3} yr. This time sets the lower bound, $t_{\rm min}$, for the Green's function in principle. The Green's function falls to zero extremely rapidly $t \to 0$ due to the exponential term $\exp(-r^2/4Dt)$. Physically, this reflects that cosmic rays cannot diffuse from the source to Earth in an infinitesimally short time. As a result, the earliest time bins, which are near $t \approx t_{\rm min}$, contain almost no events. The event counts then rise and reach a peak around $t_{\rm peak} = r^2/6D$, after which the behaviour is dominated by the power-law tail falling off as $t^{-3/2}$, as illustrated by the histograms and Green's function fit curves shown in Figure 6.31.

To compute the differential flux $\Phi(E)$ of cosmic rays incident on a surface, we consider the directional intensity over the upper hemisphere, i.e., over solid angles with $\theta \in [0, \pi/2]$. For an isotropic particle distribution with number density $n(E)$, $\Phi(E)$ is obtained as

$$\Phi(E) = \frac{{\rm d}N}{{\rm d}E\, {\rm d}A\, {\rm d}t\, {\rm d}\Omega} = \frac{c}{4\pi} n(E), \tag{6.115}$$

where c is the speed of light, and the number of particles is given by ${\rm d}N = n{\rm d}V{\rm d}E{\rm d}\Omega/4\pi$, since only the fraction $d\Omega/4\pi$ of particles from an isotropic distribution travel into the solid-angle element $d\Omega$. The differential volume element is ${\rm d}V = {\rm d}A\, c{\rm d}t \cos\theta$, where θ is the zenith angle. The factor $\cos\theta$ accounts for the projection of the particle's motion onto the surface of the Earth-centered bubble. The factor $1/4$ arises from the angular integration over the upper hemisphere. Based on Figure 6.32, we then calculate the arrival proton flux [343], which, following Eq. (6.115), represents the steady-state flux (without

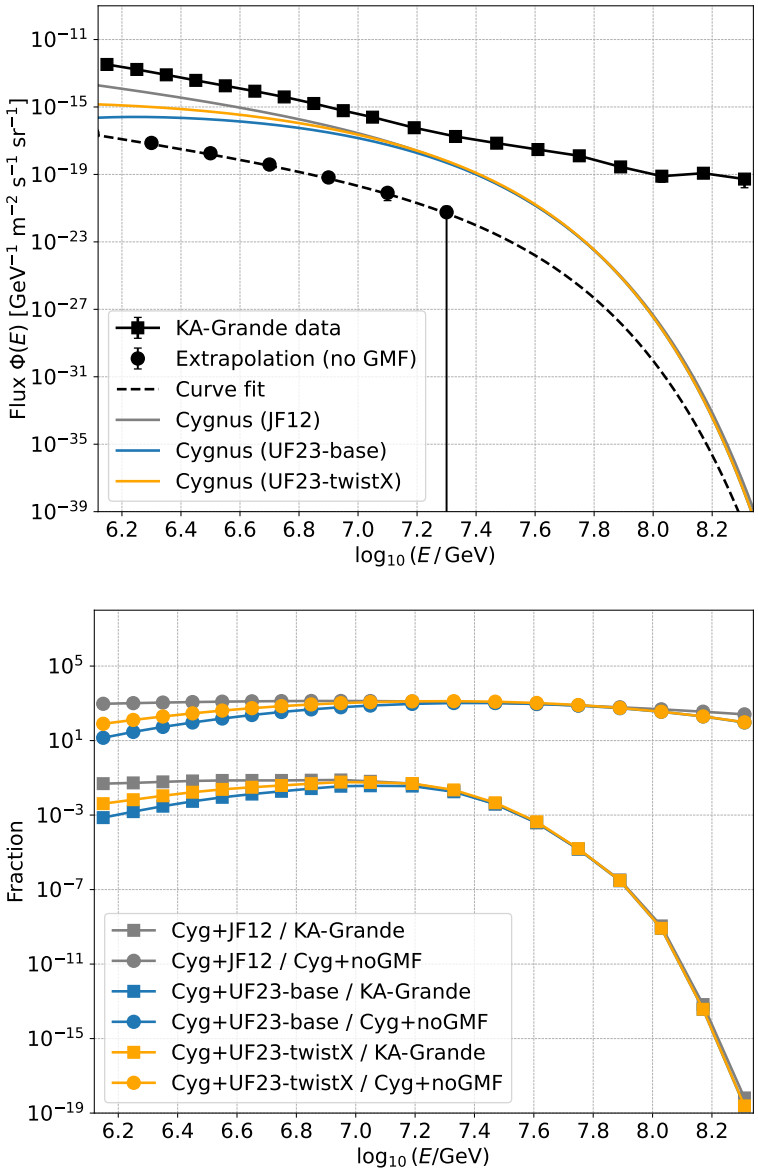


Figure 6.33: Top panel: differential fluxes at Earth for protons propagated from Cygnus (red solid curve), the reported KASCADE-Grande proton flux (blue square markers), and the Cygnus Bubble γ-ray flux within 6° (black solid curve), alongside an estimate of the Cygnus proton flux without Galactic magnetic-field and diffusion effects (red dashed curve). Bottom panel: fractions of propagated Cygnus protons relative to the KASCADE-Grande flux (blue) and to the undistorted proton flux (red). At true energies of 17.2, 25.3, 35.4, and 50.1 PeV (i.e., $\log_{10}(E/\mathrm{GeV}) = 7.24$ 7.40 7.55 and 7.70), relative to the KASCADE-Grande proton flux, the Cygnus contribution is 2.522%, 0.767%, 0.132% and 0.011%, respectively.

any solar or geomagnetic effects) at Earth. This flux can be predicted if the source has been emitting long enough that the local particle density no longer evolves with time:

$$\Phi(E) = \frac{c\,\dot{Q}(E)}{16\pi^2 |\mathbf{r}| D(E)} \left[\mathrm{erf}\left(\frac{|\mathbf{r}|}{2\sqrt{D(E) t_{\mathrm{min}}}} \right) - \mathrm{erf}\left(\frac{|\mathbf{r}|}{2\sqrt{D(E) T_{\mathrm{age}}}} \right) \right], \tag{6.116}$$

where $\dot{Q}(E)$ is obtained by Eq. (6.105), and $D(E)$ is obtained by Eq. (6.107). As $t_{\mathrm{min}} \to 0$, the argument of the error function diverges, and we find $\mathrm{erf}\left(r/2\sqrt{D\,t_{\mathrm{min}}}\right) \to 1$. Similarly, as $t_{\mathrm{age}} \to \infty$, the error function approaches $\mathrm{erf}\left(r/2\sqrt{D\,T_{\mathrm{age}}}\right) \to 0$, i.e., $\int_0^\infty G \mathrm{d}t \simeq 1/4\pi D|\mathbf{r}|$. However, the condition $T_{\mathrm{age}} \gg t_{\mathrm{min}}$ is not always satisfied, for example, in the case of protons with energies of several tens of PeV, the second error-function term may be exponentially small but still not entirely negligible.

The detailed distributions of the differential fluxes at Earth for protons propagated from Cygnus, along with the corresponding fractions relative to the KASCADE-Grande proton flux, are shown in Figure 6.33. In this analysis, both the Cygnus signal (N_{cyg}) and the KASCADE-Grande sample (N_{KA}) are assumed to consist entirely of protons. The number of expected events in the four effective energy bins is summarized in Table 6.9.

Since the diffusion coefficient D in Eq. (6.107) applies to protons only, and D is rigidity-dependent, we must apply the same methods to all other nuclei (helium, carbon, silicon, and iron) to obtain their signal ratios. To set their fractions relative to protons in the source spectrum, we need to estimate each element's abundance. It is important to emphasise that the abundance ratios of nuclei to protons from a single source differ from the Galaxy's overall composition and the observed composition on Earth due to effects during propagation. A reliable approach is to first estimate the composition of charged particles accelerated at the source (e.g., supernova remnants), and then account for propagation effects beyond diffusion, such as energy losses, decay, and so on.

6.5.8 Anisotropies on Earth

In statistical theory and probabilistic modelling, the Poisson distribution is a discrete probability law that describes the likelihood of observing a specific number of events within a fixed interval of time or space, assuming that individual events occur independently of one another and at a constant underlying average rate, without any dependence on the time elapsed since the previous event [347]. In this section, we adopt a pixel-based Poisson likelihood formulation to reconstruct the cosmic-ray relative-intensity map, under the assumption that the exposure is uniform across the entire sky

We begin by dividing the celestial sphere into N_{pix} equal-area HEALPix pixels. Denote by n_i the number of simulated arrival events that fall into pixel i, so that the total number of events within the entire FoV is given by $N_{\mathrm{tot}} = \sum_{i=1}^{N_{\mathrm{pix}}} n_i$. Under the assumption of uniform exposure, and each pixel shares the same average expected counts $\bar{\mu}$, we introduce a dimensionless relative-intensity map I_i to describe anisotropy deviates from the mean, which is defined as:

$$\mu_i = \bar{\mu}\, I_i, \quad \bar{\mu} = \frac{N_{\mathrm{tot}}}{N_{\mathrm{pix}}}, \quad \sum_{i=1}^{N_{\mathrm{pix}}} I_i = N_{\mathrm{pix}}. \tag{6.117}$$

If the expected number of events in a pixel i is $\hat{\mu}_i$, then the probability of observing exactly n_i events in that pixel is governed by the Poisson distribution, which takes the form:

$$P(n_i \,|\, \mu_i) = \frac{\mu_i^{n_i}\, e^{-\mu_i}}{n_i!}, \tag{6.118}$$

which represents the probability. Similar to the maximum-likelihood method we applied earlier to the observational data in Section 6.4.2, we now employ the same framework for the simulated propagation dataset, where the analysis is performed under the simplifying assumption of uniform exposure and partial-sky coverage, with no detector-related systematics included. Regarding the entire FoV, we assume statistical independence across all sky pixels, so that the joint likelihood of the model μ, given the observed counts n, is constructed as the product over all pixels of the corresponding Poisson probabilities for observing n_i vents when the expected number of events in that pixel is μ_i. Substituting Eq. (6.117) into the likelihood, one then obtains

$$\mathcal{L}(\hat{\mu}_i \,|\, n_i) = \prod_{i=1}^{N_{\rm pix}} \frac{\hat{\mu}_i^{n_i} e^{-\hat{\mu}_i}}{n_i!} = \prod_i \frac{\left(\bar{\mu}\, I_i\right)^{n_i} e^{-\bar{\mu}\, I_i}}{n_i!}\,. \tag{6.119}$$

The full log-likelihood (dropping the constant term $n_i!$) is then obtained as

$$\ln \mathcal{L} = \sum_i \left[n_i \ln(\bar{\mu}\, I_i) - \bar{\mu}\, I_i \right] + C, \tag{6.120}$$

where C is a constant. Maximizing $\ln \mathcal{L}$ with respect to I_i, namely, $\partial_{I_i} \ln \mathcal{L} = 0$, one finds the analytic maximum-likelihood estimator

$$I_i = n_i/\bar{\mu}, \quad \delta I_i = I_i - 1, \tag{6.121}$$

where $\bar{\mu} = \langle n_i \rangle$, δI_i denotes the relative deviation or fractional excess above isotropy, and measures the anisotropy contrast at each point. This is the special case of perfectly uniform exposure (no mask, no time or detector acceptance variations), the Poisson maximum-likelihood estimate for each pixel's intensity is simply its count divided by the all-sky mean. When the detector's exposure varies across the sky, due to limited field of view, zenith-angle cuts, or azimuthal asymmetries, the simple form Eq. (6.121) no longer holds. Instead, one must introduce a per-pixel acceptance A_i and, if the exposure also varies with time (e.g., sidereal rotation), a per-time-bin normalization N_t, as given in Section 6.3.3 and 2.2.4, and the theory is proposed in paper [143]. The sky maps of relative intensity, pre-trial significance and post-trial significance for collected protons (median energy 33 PeV) under the three different magnetic-field models are shown in Fig. 6.34.

To quantify the significance of any deviation from isotropy, we form the likelihood-ratio test statistic (TS) comparing the fitted anisotropy $\hat{I}_i$ to the null hypothesis $I_i^{\rm bg} \equiv 1$:

$$\mathrm{TS} = -2 \ln \frac{\mathcal{L}(I_i^{\rm bg})}{\mathcal{L}(\hat{I}_i)} = 2 \sum_{i=1}^{N_{\rm pix}} \left[n_i \ln \frac{\hat{\mu}_i}{\mu_i^{\rm bg}} + \mu_i^{\rm bg} - \hat{\mu}_i \right], \tag{6.122}$$

with $\mu_i^{\rm bg} = \bar{\mu}$ and $\hat{\mu}_i = \bar{\mu}\, \hat{I}_i$. Under Wilks' theorem [348], TS follows a χ^2 distribution with degrees of freedom equal to the number of free parameters in the anisotropy model (e.g., two for a dipole: amplitude and phase). The TS is the significance (in likelihood-ratio units) of that deviation, comparing the best-fit anisotropic model to the isotropic null-hypothesis. It reveals how unlikely the observed count in pixel i would be under the isotropic hypothesis. A per-pixel significance map is obtained as $\sigma_i = \sqrt{\mathrm{TS}_i}$, namely,

$$\sigma_i = \sqrt{2 \left(n_i \ln(\mu_i/\mu_i^{\rm bg}) + \mu_i^{\rm bg} - \mu_i \right)}, \tag{6.123}$$

which is equivalent to Eq. (6.96).

The post-trial significance map of KASCADE-Grande at 33 PeV highlights regions of high significance while suppressing low-significance regions due to the multi-trial test

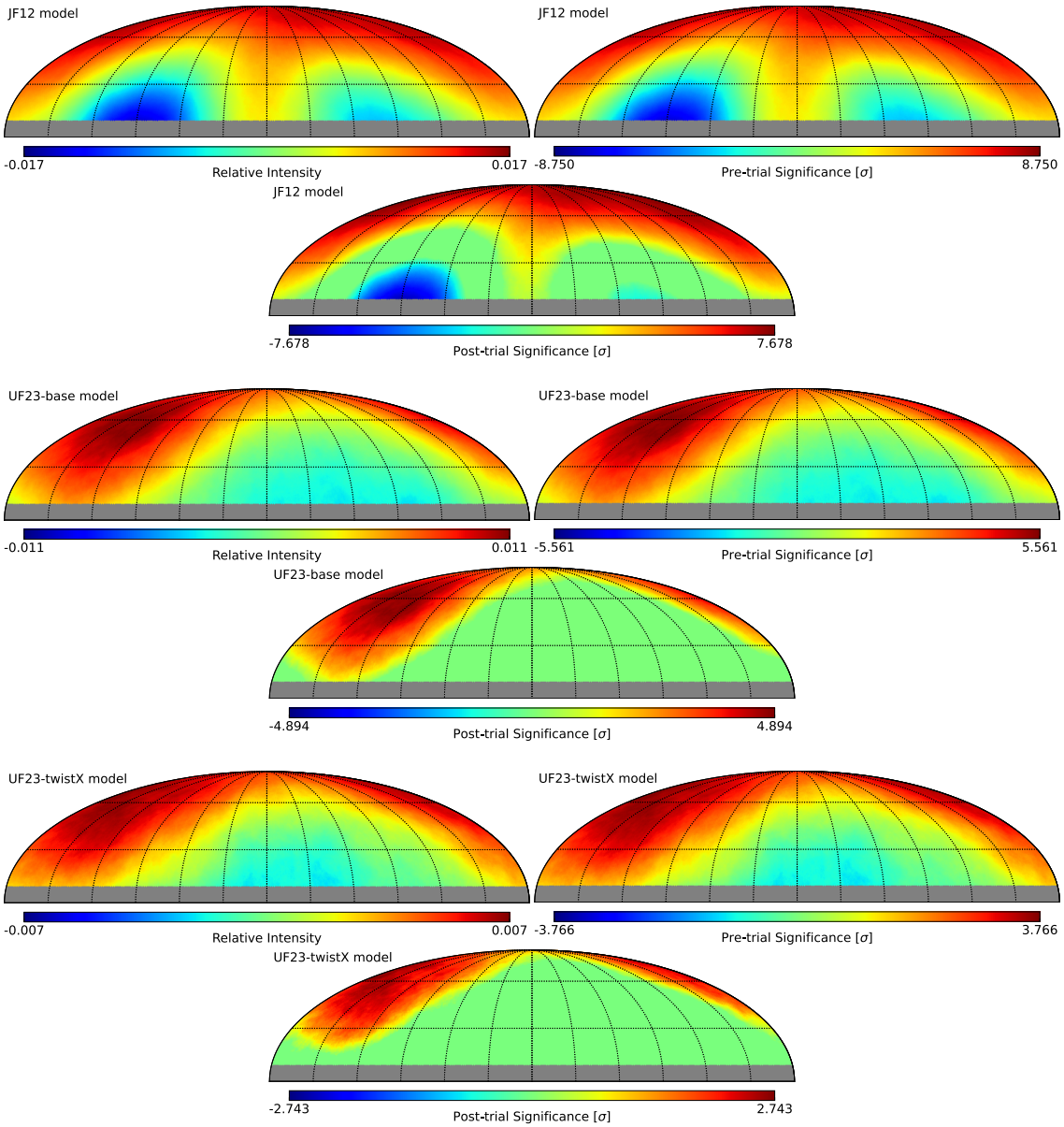


Figure 6.34: Sky maps of relative intensity, pre-trial significance, and post-trial significance in equatorial coordinates, obtained from propagated protons. From top to bottom are results using the JF12, UF23-base, and UF23-twistX Galactic magnetic-field models. The JF12 model exhibits large biased patterns, whereas the UF23-base and UF23-twistX models reproduce features more consistent with the anisotropies observed at 33 PeV by KASCADE–Grande (see Figures 6.19 and 6.22).

algorithm. This outcome differs from the cosmic-ray propagation results generated using CRPropa, which show the directly observed distribution from Earth. However, we can analyze the propagated cosmic rays and compare them with the anisotropy maps from KASCADE–Grande in both Galactic and equatorial coordinates, as shown in Figures 6.30 and 6.30. The differences in the excess regions obtained with different magnetic-field models, as illustrated in Figure 6.34, are likely due to the limited statistics of the KASCADE–Grande data at 33 PeV and a possible contribution from very high-energy γ-rays overlapping with the cosmic-ray excess in this region. Here, we conclude that the 33 PeV cosmic-ray excess observed by KASCADE-Grande is highly likely to originate from the Cygnus OB2 association. This may account for the significant discrepancy in dipole phases observed by IceCube and KASCADE-Grande at tens of PeV, as shown in Figure 5.1. This feature arises due to the deviation of Earth's rotational axis relative to the Galactic plane and the non-uniformity of the Galactic magnetic field (GMF), which behaves anisotropically for upstream and downstream cosmic rays originating from Cygnus OB2. In summary, sources within the Cygnus OB2 region, such as Cyg X-3, can plausibly account for the medium-scale anisotropies observed at 33 PeV by KASCADE–Grande (see Figure 6.25) from a propagation perspective, and cooperate with the very high-energy PeV γ-ray emission observed by LHAASO [8].

6.6 A Proposal for the Acceleration Mechanism

The analysis of the unified anisotropy with the mean experiments in Chapter 5, and the results of KASCADE-Grande in Chapter 6 show evidence of a cosmic-ray source at 33 PeV, which is close to the Cygnus region. This inspired us to investigate the mechanics of acceleration of high-energy charged particles, which can reach tens of PeV range in the Milky Way. However, while SNRs are capable of accelerating particles up to PeV energies, they are insufficient to explain the highest energy cosmic rays in our Galaxy, which necessitates considering other sources. In addition to supernova remnants, other astrophysical objects within the Milky Way, such as pulsar wind nebulae (PWNe) and massive star clusters, may also contribute to the Galactic cosmic-ray population. Pulsars, rapidly rotating neutron stars with intense magnetic fields, can accelerate particles in their magnetospheres and in the surrounding wind nebula. These objects are thought to be capable of contributing to the population of Galactic cosmic rays, especially at energies near the knee [3], at around 3 PeV.

The wakefield produces a ponderomotive force, a time-averaged, nonlinear force in an oscillating, inhomogeneous field, that can trap particles and accelerate them efficiently once the wave potential exceeds their initial momentum [264]. Charged particles can be accelerated by the wakefield in plasma via the Landau resonance, a process known as wakefield acceleration (WFA). This concept was first proposed in the study of laboratory plasmas [264]. In laboratory experiments, wakefield acceleration in plasma can be achieved using a laboratory-scale wakefield accelerator. This technique accelerates particles to high energies over short distances by utilizing a plasma wave or wakefield generated by either a particle beam or a laser pulse. In the first method, a high-energy electron beam (the "driver") generates a plasma wave in a gas, which then accelerates a secondary electron beam (the "witness"). In this way, energy stored in collective plasma oscillations is converted into particle kinetic energy. The wakefield in plasma can be driven by either high-energy particle beams or by intense laser pulses. In the beam-driven case, a high-energy "driver"

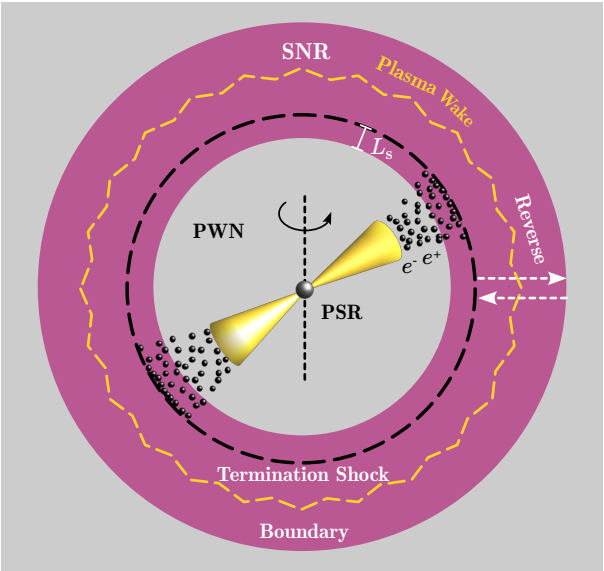


Figure 6.35: Schematic structure of a supernova remnant (SNR) containing a central pulsar (PSR) and its surrounding pulsar wind nebula (PWN). The yellow dashed circular wave illustrates the plasma wakefield that may arise from the interaction between the termination shocks of the pulsar wind and the reverse shocks propagating within the PWN.

beam propagates through plasma, generating a wake that accelerates a trailing "witness" beam. A notable example is CERN's AWAKE experiment [265], which uses a 400 GeV proton beam from the Super Proton Synchrotron (SPS) to drive a wakefield capable of accelerating electrons to GeV energies. The laser pulse must be coherent and have a duration which matches the plasma's duration of oscillation. This pulse drives a wake in the plasma, similar to a boat moving through the water, and the electric fields in the wake accelerate electrons to high energies.

In the context of cosmic-ray acceleration via the wakefield acceleration (WFA) mechanism, Chen [266] proposed that strong Alfvén shocks in relativistic astrophysical plasmas can drive a plasma wakefield capable of accelerating cosmic-ray particles to ultra-high energies. Acceleration by the wake's longitudinal field can reduce some radiative losses compared to shock cycles and can yield power-law spectra under suitable conditions [266]. Ebisuzaki and Tajima [267] further suggested that intense Alfvén waves generated in magnetized accretion disks around supermassive black holes can excite ponderomotive and wakefield forces within relativistic jets, enabling particle acceleration up to ZeV energies. However, such bursts from accretion disks are characteristic of extragalactic sources, such as active galactic nuclei (AGNs), and have no known analogue within the Milky Way. Other galactic processes are typically limited to energies up to the GeV range and therefore cannot account for the observed PeV cosmic rays. In recent years, the acceleration mechanisms of PeV galactic CRs have been focused on magnetospheres described in [258]. The work is relevant to proton acceleration in the pulsar magnetosphere, but only involves proton acceleration. It is worth noticing that some works have involved wakefield acceleration (WFA) of high-energy CRs, up to the PeV range, but the coherence and duration problems are not considered well.

For the first time, I propose that PeV cosmic rays in the Cygnus OB2 association may originate predominantly from pulsars. In contrast to previous interpretations, standard

Fermi acceleration in Galactic sources explains cosmic rays up to the knee ($\sim 3\,\mathrm{PeV}$ for protons, tens of PeV for Fe). Beyond this range, however, galactic contributions weaken, and cosmic rays with $> 10^{17}\,\mathrm{eV}$ are usually attributed to extragalactic accelerators. However, the KASCADE–Grande 33 PeV bin shows an excess near Cygnus OB2, which, as mentioned earlier, is supported by propagation simulations discussed earlier. High-energy cosmic-ray acceleration can involve wakefield acceleration, which, in principle, requires coherent emissions and short, matched durations relative to the plasma's frequency. Recent work primarily focuses on extragalactic sources such as active galactic nuclei (AGNs) [349, 266]. Similarly, I propose a cosmic-ray acceleration model through plasma wakefield acceleration, driven by high-energy coherent radiation from pulsars, as illustrated in Figure 6.35. In this framework, coherent radiation (ChR) generated by a relativistic electron beam or ChR pulse from polar-cap regions of neutron stars drives a wakefield in the surrounding PWN plasma. Pulses from pulsars, modulated during propagation, interact with the plasma in pulsar-wind nebulae (PWNe), generating wakefields capable of accelerating cosmic rays to PeV energies, which in turn can efficiently accelerate charged particles (such as protons) within the PWN wakefield. The detailed theory and mechanism will be presented in future work following this thesis.

Summary

This thesis focuses on high-energy cosmic-ray anisotropies and their origins, motivated by the existing observations from experiments such as IceCube, IceTop, LHAASO, Tibet, ARGO, HAWC, and Auger. In the PeV energy range, the observed dipole anisotropies reveal a significant signal across an energy gap that few observatories can currently access. After introducing the fundamentals of cosmic-ray physics and anisotropy reconstruction methods in Chapters 1 and 2, Chapter 3 presents my sensitivity studies for observatories (e.g., the IceCube-Gen2 surface array) in cosmic-ray anisotropy research. There, I developed both a Monte Carlo (MC) simulation framework and an equivalent theoretical model for evaluating sensitivity. Chapter 4 introduces optimized analytical methods for the reconstruction of cosmic-ray anisotropies. In Chapter 5, I combined and unified the dipole anisotropy measurements reported by major experiments to obtain a global fit of the dipole amplitude and phase from the TeV range up to beyond 100 EeV. The results reveal a distinct "W"-shaped dependence of the dipole amplitude on primary energy (and similarly for the phase), which may indicate a transition in the dominant cosmic-ray sources as well as rigidity-dependent propagation effects within our Galaxy. This feature provides important clues to the origins of cosmic rays in the Milky Way, particularly in the PeV energy range, above the so-called "knee" of the cosmic-ray spectrum. However, the dipoles reported in the official KASCADE-Grande analyses show a noticeable bias relative to those from IceCube and IceTop in overlapping energy ranges. This motivated a deeper investigation of dipole and medium-scale anisotropies using KASCADE-Grande data, which is described in Chapter 6. From this analysis, I found that the significance of the 33 PeV energy bin reaches 3σ near the Cygnus OB2 region, a result further supported by propagation studies using the CRPropa framework. Finally, I propose, for the first time, that ~ 33 PeV cosmic rays from within the Cygnus OB2 association may originate predominantly from pulsars and be possibly accelerated by wakefields formed within their pulsar wind nebulae (PWNe).

In the thesis, Chapter 3 quantifies the sensitivity of a partial-sky surface array using IceCube-Gen2 at the South Pole as the concrete example, to large-scale (dipole) anisotropy across 10^6–$10^{8.8}$ GeV using two approaches: (i) MC simulations of detector response and arrival directions and (ii) a theoretical frame for estimation based on the statistics of a first-harmonic fit in right ascension (RA). Both approaches agree on dominant scalings and limitations, yielding a coherent, practically useful picture of measurable signals and design trade-offs. On the simulation side, the detector's air-shower reconstruction efficiency is parametrized as a smooth function of energy and zenith, $\epsilon(E, z)$ with $z = \sin^2\theta$, folded over primaries with the H4a flux to obtain a total efficiency $\epsilon_{\mathrm{tot}}(E, \theta)$ (as an option). Arrival directions are drawn on HEALPix by combining ϵ_{tot}, the declination-dependent exposure, and an injected dipole modulation. Each map is converted to relative intensity, smoothed on large scales, projected into RA, and fit with the first harmonic $\delta I_{\mathrm{reco}}(\alpha) = \mathcal{A}_{\mathrm{reco}} \cos(\alpha - \varphi) + B$. Significance is summarized with the simple z-score $n_\sigma = (\mathcal{A}_{\mathrm{reco}} - \mathcal{A}_{\mathrm{hypo}})/\sigma_{\mathcal{A}}$ with $\mathcal{A}_{\mathrm{hypo}} = 0$. Partial-sky coverage compresses amplitudes, typically $\mathcal{A}_{\mathrm{reco}}/\mathcal{A} \lesssim 0.8$, with stronger reductions for dipoles near the poles or largely outside the field of view. By scanning amplitudes and declinations across energy bins and plotting $n_\sigma/\sqrt{N}$ versus $\mathcal{A}$, smooth fits provide effective slopes that convert a measured amplitude and total counts into an expected significance, enabling $3\sigma/5\sigma$ sensitivity curves for reconstructed

amplitudes and corresponding bands for the true (full-sky) amplitudes. The theoretical approach starts from the same RA fit but derives uncertainties and significances directly in terms of the total number of events and a single geometry–exposure factor γ that encodes latitude and declination coverage. In words, the projected amplitude uncertainty scales as $\sigma_{\mathcal{A}} \simeq \sqrt{2\gamma/N_{\rm tot}}$, and the amplitude significance scales as $n_\sigma(\mathcal{A}_p) = (1/\sqrt{2\gamma})\,\mathcal{A}_p\sqrt{N_{\rm tot}}$, as Eq. (3.39) shows. The phase precision improves in lockstep, $\sigma_\varphi = (1/\mathcal{A}_p)\sqrt{2\gamma/N_{\rm tot}}$, and a convenient phase "significance" proxy is $n_\sigma(\varphi) == \pi\, n_\sigma \mathcal{A}_p$, with equations (3.40) and (3.41). For the Gen2-like geometry adopted here, $\gamma \simeq 2.487$, so $1/\sqrt{2\gamma} \simeq 0.448$, in excellent agreement with the MC-measured slope $n_\sigma/\sqrt{N}$ versus $\mathcal{A}$. This allows all ground-based observatories to perform a fast and convenient estimation of their sensitivity to cosmic-ray dipole anisotropies in the future.

In Chapter 4, I present several optimized reconstruction methods developed to minimize biases in the recovered cosmic-ray dipoles and their uncertainties. These approaches account for partial-sky coverage and enhance the estimation of dipole significance. In addition, a "Gradient Method" is introduced to identify stable dipole directions under limited statistics, though it is not aimed at detailed reconstruction. The chapter establishes FoV-aware techniques for large-scale anisotropy reconstruction. The compact 3D dipole field on the sphere is expressed as $\delta I(\alpha,\delta) = \mathcal{A}\big(\sin\delta\,\sin\delta_{\rm d} + \cos\delta\,\cos\delta_{\rm d}\cos(\alpha-\alpha_{\rm d})\big)$. Projecting over declination yields the 1D modulation $\delta I(\alpha) = \mathcal{A}c_{\rm p}\cos(\alpha-\alpha_{\rm d}) + \mathcal{A}c_{\rm s}$, where $c_{\rm s}$ and $c_{\rm p}$ encode the observatory FoV. Using the fit model $\delta I(\alpha) = \mathcal{A}_{\rm p}\cos(\alpha-\alpha_{\rm p}) + \mathcal{B}$, the dipole latitude is recovered via $\delta_{\rm d} = \arctan\big(c_{\rm p}\mathcal{B}/(c_s\mathcal{A}_{\rm p})\big)$. For low-statistics data, the gradient method would increase the stability by scanning rotated spherical belts, applying inverse-variance weighting, and fitting iso-height contours to extract the dipole orientation $(\alpha_{\rm d},\delta_{\rm d})$ and amplitude $\mathcal{A}$.

In Chapter 5, I address why reported dipoles from different experiments disagree and how to recover the underlying, "real" global dipole as a function of energy. Partial-sky FoV, site geometry, sample definitions, and energy resolution bias the projected sidereal dipoles (amplitude and phase). I formalize how overlapping energy samples combine in relative intensity, model experiment-specific energy PDFs (log-normal or exponential tails), and calibrate FoV effects via Eq. (4.9). I then reconstruct the energy-dependent dipole using two approaches: (i) an analytical fit that projects a 3D dipole to right ascension and matches least-squares to reported $(\mathcal{A}_{\rm p},\alpha_{\rm d})$; and (ii) a deconvolution on an energy grid where each measured dipole is a weighted sum of complex dipole vectors $\mathcal{D}_j = \mathcal{A}_j e^{i\varphi_j}$, with weights from the sample's energy PDF. I solve Eq. (5.36) iteratively by gradient descent and propagate uncertainties by MC resampling. I also note practical projection details, such as applying $\cos\delta$ weighting or right-ascension banding for accurate large-scale patterns using HEALPix. The unified result yields continuous curves of amplitude and phase from 1 TeV to 50 EeV after FoV calibration, revealing a dipole flip at $E \simeq 107.9$ TeV with amplitude $\mathcal{A}_{\rm p} = (2.1\pm0.40)\times10^{-4}$ and phase $\alpha_{\rm d} = 315.3^\circ\pm7.7^\circ$. The amplitude shows a characteristic "W-shaped" structure, as shown in Figure 6.17, while the phase twists with energy. Around the ankle ($\sim 3\times10^{18}$ eV), the trends suggest an increasing contribution from extragalactic sources. Compared to raw measurements, the deconvolved dipole varies more sharply with energy because the superposition of wide energy bins in experiments flattens the true variations. These studies indicate that IceCube-Gen2 will provide important measurements for studying cosmic-ray anisotropies in the PeV–EeV range. However, some differences, especially in the 1-100 PeV range and in the KASCADE–Grande data, call for a more detailed analysis in the PeV range, which is presented in Chapter 6.

In Chapter 6, I analyze KASCADE–Grande data for dipole and medium-scale anisotropies

at median energies of 2.7, 6.1, and 33 PeV. Using KCDC public data, I select events with $\theta < 40^\circ$ and $N_{\rm ch} \geq 10^{5.2}$. I first survey systematics in solar, sidereal, and anti-sidereal time, then apply the East–West derivative method. A strong solar-time signal and anti-sidereal amplitudes comparable to the sidereal, with roughly inverted phases, which point to residual exposure and environmental effects. Seasonal and time-ordered splits show phase drifts of 20°–40°, with the 33 PeV bin particularly unstable, motivating stricter data-quality cuts. To stabilize exposure, I remove days with large gaps and run time-scrambling–based Rayleigh/Fourier analyses to subtract a high-statistics background. This filtering reduces variability and smooths the RA modulation, but anti-sidereal leakage and split-phase inconsistencies remain, especially at 33 PeV. In short, East–West and standard Rayleigh by themselves can't cleanly correct time-variation biases when live time is patchy; day-level acceptance weighting and FoV-aware 3D reconstructions are needed. I then reconstruct 3D dipoles by averaging exposure- and acceptance-weighted 2D vectors in belts and subtracting a time-scrambled reference to isolate the excess. I use declination belts to constrain the equatorial dipole via $(\mathfrak{U}_x, \mathfrak{U}_y)$ and the vertical component via the mean z term, and I also use solid-angle belts (Θ_Ω, φ) perpendicular to the y-axis as a geometric cross-check under partial FoV. This links the measured resultants to $\mathcal{A}$ and $(\alpha_{\rm d}, \delta_{\rm d})$ through FoV moments $p = \langle \cos\delta \rangle$, $s = \langle \sin\delta \rangle$, and $t = \langle \sin^2\delta \rangle$; uncertainties are propagated from the weights and the scramble count $n_{\rm scr}$. Phases from multiple belts are combined, and $\mathcal{A}$ and $\delta_{\rm d}$ are obtained per belt or via a global non-linear least-squares fit.

Moreover, I average the acceptance–calibrated intensity vector over broken Θ_Ω-belts to form pseudo–resultants $(\mathfrak{V}_{x,i}, \mathfrak{V}_{z,i})$ and a complementary y-component $\mathfrak{V}_{y,i}$, then fit these belt–averaged components to the integrated dipole model on (Θ_Ω, φ), noting that incomplete belts bias the resultant direction (a pseudo–phase). Using closed-form belt integrals (coefficients $S_1, S_2, P_0, P_1, P_2, P_3$) and the angle relations that map $(\Theta^{\rm d}_\Omega, \varphi_{\rm d})$ to $(\delta_{\rm d}, \alpha_{\rm d})$, I solve per belt for $(\mathcal{A}, \alpha_{\rm d}, \delta_{\rm d})$ with uncertainties propagated from $(\sigma_{\mathfrak{V}_x}, \sigma_{\mathfrak{V}_z}, \sigma_{\mathfrak{V}_y})$; in practice, a single Θ_Ω-belt spanning the KASCADE–Grande FoV is insensitive to $\delta_{\rm d}$ and multiple belts add noise, so I do not use this channel for final results, but used the declination δ-belt. For data quality, I remove unstable early solar days and large-gap sidereal days, apply $n_{\rm scr}$ time–scrambling to stabilize references, and introduce a Dirichlet prior on UTC occupancy to build a smoothed exposure weight that suppresses multi-scale time gaps; with these weights, residual solar/anti-sidereal amplitudes are consistent close to zero, validating the correction. The δ-belt angular–vector reconstructions (original and filtered, as well as time-split subsets) exhibit broadly stable phases and improved amplitude consistency, although limited statistics and partial sky coverage still induce biases, especially near low declinations, in the results. The updated FoV-calibrated anisotropy results indicate that the dipole amplitude changes with energy in a "W"-shaped pattern, as shown in Figure 6.17. A clear phase change appears around 107.9 TeV, and a small peak occurs near a few PeV (close to the knee). These features suggest that different sources dominate at different energies—possibly shifting toward the Cygnus region, and that magnetic fields affect the observed directions. Above a few PeV up to about 10^{17} eV, the results may indicate a gradual transition to extragalactic cosmic rays. More studies of medium-scale structures are needed to confirm the source directions.

Next, I searched for medium-scale structures in KASCADE–Grande using an improved maximum-likelihood method that fits the sky intensity, time normalization, and directional acceptance together, with Gaussian smoothing and a slower, more stable iteration. I map relative intensity and compute pre- and post-trial significances after top-hat smoothing optimized per energy bin, then validate global excesses with isotropy Monte Carlos to

capture the look–elsewhere effect. A hotspot at 33 PeV attains $\sim 4.6\,\sigma$ (pre-trial) and $\sim 3\,\sigma$ globally after trials, with location shifting by $\sim 15^\circ$ under data filtering, consistent with acceptance/coverage systematics. Its region (Galactic $l \sim 80^\circ$) overlaps with Cygnus, where PeV γ-ray activity has been observed, suggesting a possible local origin but not a confirmed discovery. The lower-energy bins show weaker post-trial features ($\lesssim 3\,\sigma$). Overall, the weighting, filtering, and ML stabilization help reduce solar and seasonal effects and random noise, but the limited FoV and statistics are still the main limitations. To study deeper the possible origins of the local enhancement, in Section 6.5, I simulate CR propagation from the Cyg OB2 region (PSR J2031+4127) with CRPropa and JF12/UF23 magnetic fields, using a hybrid turbulence model, to test whether source-born arrival directions populate the KASCADE–Grande FoV and reproduce the 33 PeV hotspot near $l \sim 80^\circ$. Protons (and nuclei) are injected with an LHAASO-motivated spectrum ($s \approx 2.25$, $E_0 \sim 5$ PeV, rigidity-scaled cutoffs); diffusion parameters are inferred from arrival-time fits to the Green function by using a maximum-likelihood estimator $D = r^2\langle 1/t\rangle/2$, as shown in Eq. (6.113), and giving $D(E) \propto E^\delta$ with $\delta \approx 0.6$ and $D_0 \sim 2 \times 10^{29}\,\mathrm{cm^2\,s^{-1}}$ at TeV reference. The predicted Cygnus proton flux contributes a few percent at $\sim$17–25 PeV, declining rapidly with energy, and sky maps under the model of the magnetic field named UF23 (base/twistX) align better with the observed 33 PeV excess than the model JF12. Altogether, the simulations support Cyg OB2 as a plausible contributor to the KASCADE–Grande medium-scale excess. The result supports a possible pulsar-driven wakefield acceleration scenario in pulsar-wind nebulae (PWNe), discussed in Section 6.6. The KASCADE–Grande measurements and our propagation simulations exhibit broadly consistent anisotropy morphologies. However, canonical Galactic accelerators struggle to exceed the PeV scale, leaving the energy range above 10 PeV under-explained. At the end of this thesis, I therefore propose wakefield acceleration in PWNe embedded in supernova remnants (SNRs): intense, magnetized, relativistic winds can drive longitudinal plasma waves whose ponderomotive fields trap and accelerate ions to multi-PeV rigidities. In this scenario, short-term particle injection, strong rigidity ordering, and magnetic beaming naturally produce the observed medium-scale excess near the Cygnus region while staying consistent with the large-scale dipole limits.

To give an outlook, future experiments and observations, such as IceCube-Gen2, LHAASO, CTA, AugerPrime/TA×4, and SWGO, as well as the Pierre Auger Observatory at the highest energies, will explore cosmic rays above PeV energies. They will also compare PeV γ-ray maps with medium-scale cosmic-ray excesses, search for possible neutrino signals from the Cygnus region, and allow three-dimensional anisotropy studies that take into account each experiment's field of view. Combined multi-messenger analyses and improved maps of the Galactic magnetic field will help reduce uncertainties in cosmic-ray propagation and reveal whether pulsar-wind nebulae are Galactic PeVatrons above 10 PeV.

Appendix

A Spherical Harmonics

The common spherical harmonics, namely, $Y_{\ell m}$ are based on spherical coordinates, zenith angle (θ) and azimuth angle (ϕ). The Spherical harmonics with equatorial coordinates, namely, right ascension (α), and (δ) are $Y_{\ell m}(\delta, \alpha)$. The azimuth angle ($\phi$) is equivalent to right ascension (α), while the transformation between the zenith angle (or polar angle) in spherical coordinates and declination in equatorial coordinates, namely $\theta = \pi/2 - \delta$. Their complex conjugates are $Y^*_{\ell m}(\delta, \alpha)$, which are follows a relation with $Y_{\ell m}(\delta, \alpha)$, namely:

$$Y_{\ell -m}(\delta, \alpha) = (-1)^m Y^*_{\ell m}(\delta, \alpha), \tag{A.1}$$

which follows from the general properties of the associated Legendre functions and ensures orthonormality on the sphere. In anisotropy analyses, the cosmic-ray relative intensity map $I(\alpha, \delta)$ is often expanded into these basis functions to isolate large- and small-scale angular structures. The coefficients $a_{\ell m}$ obtained from this expansion quantify the strength and phase of anisotropy modes of different angular scales, with the dipole ($\ell = 1$) and quadrupole ($\ell = 2$) terms being of large-scale anisotropy. The detailed $Y_{\ell -m}(\delta, \alpha)$ with degrees $\ell = 0, 1, 2, 3$ are given by the following:

For $\ell = 0$:

$$Y_{00}(\delta, \alpha) = \frac{1}{\sqrt{4\pi}}, \quad Y^*_{00}(\delta, \alpha) = \frac{1}{\sqrt{4\pi}}. \tag{A.2}$$

For $\ell = 1$:

$$\begin{aligned} Y_{1,-1}(\delta, \alpha) &= \sqrt{\frac{3}{8\pi}} \cos\delta e^{-i\alpha}, \quad Y^*_{1,-1}(\delta, \alpha) = \sqrt{\frac{3}{8\pi}} \cos\delta e^{i\alpha} \\ Y_{10}(\delta, \alpha) &= \sqrt{\frac{3}{4\pi}} \sin\delta, \quad Y^*_{10}(\delta, \alpha) = \sqrt{\frac{3}{4\pi}} \sin\delta \\ Y_{1,1}(\delta, \alpha) &= -\sqrt{\frac{3}{8\pi}} \cos\delta e^{i\alpha}, \quad Y^*_{1,1}(\delta, \alpha) = -\sqrt{\frac{3}{8\pi}} \cos\delta e^{-i\alpha}. \end{aligned} \tag{A.3}$$

For $\ell = 2$:

$$\begin{aligned} Y_{2,-2}(\delta, \alpha) &= \sqrt{\frac{15}{32\pi}} \cos^2\delta e^{-2i\alpha}, \\ Y_{2,-1}(\delta, \alpha) &= \sqrt{\frac{15}{8\pi}} \cos\delta \sin\delta e^{-i\alpha}, \\ Y_{20}(\delta, \alpha) &= \sqrt{\frac{5}{16\pi}} (3\sin^2\delta - 1), \\ Y_{2,1}(\delta, \alpha) &= -\sqrt{\frac{15}{8\pi}} \cos\delta \sin\delta e^{i\alpha}, \\ Y_{2,2}(\delta, \alpha) &= \sqrt{\frac{15}{32\pi}} \cos^2\delta e^{2i\alpha}. \end{aligned} \tag{A.4}$$

For $\ell = 3$:

$$
\begin{aligned}
Y_{3,-3} &= \sqrt{\frac{35}{64\pi}}\cos^3(\delta)e^{-3i\alpha}, \quad Y^*_{3,-3} = \sqrt{\frac{35}{64\pi}}\cos^3(\delta)e^{3i\alpha} \\
Y_{3,-2} &= \sqrt{\frac{105}{32\pi}}\cos^2(\delta)\sin(\delta)e^{-2i\alpha}, \\
Y_{3,-1} &= \sqrt{\frac{21}{64\pi}}\cos(\delta)(5\sin^2(\delta)-1)e^{-i\alpha}, \\
Y_{30} &= \sqrt{\frac{7}{16\pi}}(5\sin^3(\delta)-3\sin(\delta)), \\
Y_{3,1} &= -\sqrt{\frac{21}{64\pi}}\cos(\delta)(5\sin^2(\delta)-1)e^{i\alpha}, \quad Y^*_{3,1} = -\sqrt{\frac{21}{64\pi}}\cos(\delta)(5\sin^2(\delta)-1)e^{-i\alpha} \\
Y_{3,2} &= \sqrt{\frac{105}{32\pi}}\cos^2(\delta)\sin(\delta)e^{2i\alpha}, \\
Y_{3,3} &= -\sqrt{\frac{35}{64\pi}}\cos^3(\delta)e^{3i\alpha}, \quad Y^*_{3,3} = -\sqrt{\frac{35}{64\pi}}\cos^3(\delta)e^{-3i\alpha}.
\end{aligned}
\tag{A.5}
$$

B Coordinates

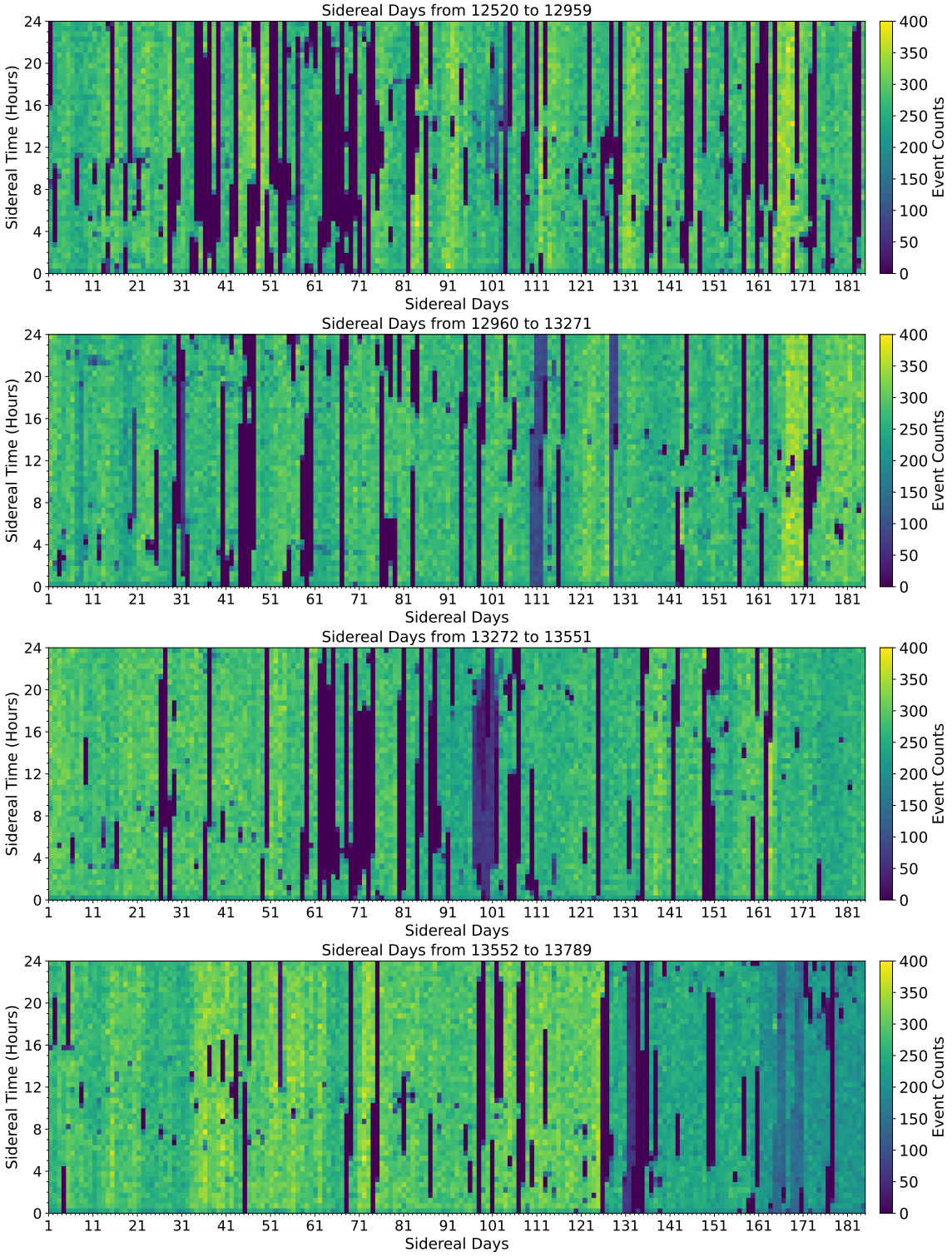
The function $\operatorname{arctan2}(y, x)$ can be formally defined as:

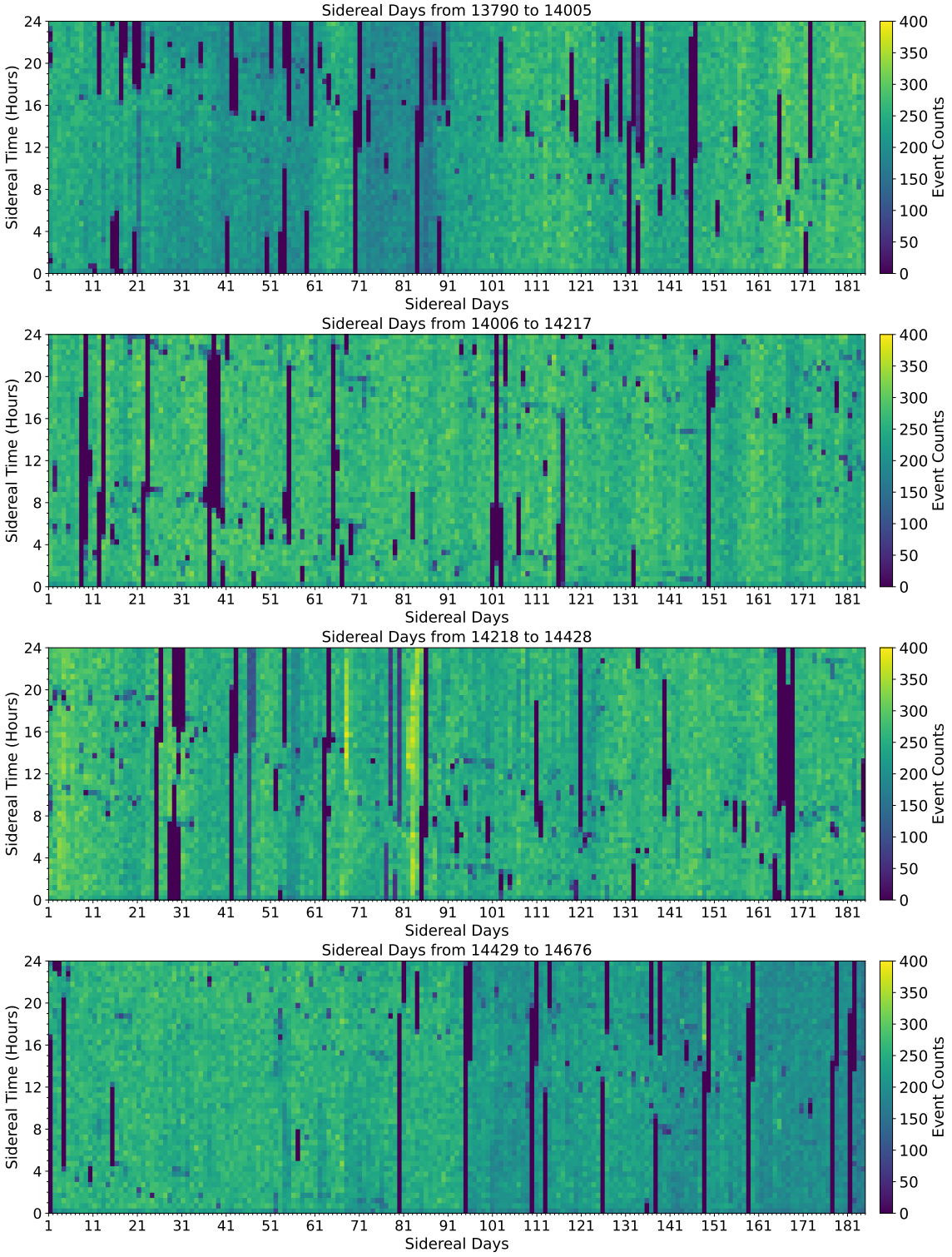
$$
\operatorname{arctan2}(y,x) = \begin{cases}
\arctan\left(\frac{y}{x}\right) & \text{if } x > 0, \\
\arctan\left(\frac{y}{x}\right) + \pi & \text{if } x < 0 \text{ and } y \geq 0, \\
\arctan\left(\frac{y}{x}\right) - \pi & \text{if } x < 0 \text{ and } y < 0, \\
\frac{\pi}{2} & \text{if } x = 0 \text{ and } y > 0, \\
-\frac{\pi}{2} & \text{if } x = 0 \text{ and } y < 0, \\
\text{undefined} & \text{if } x = 0 \text{ and } y = 0.
\end{cases}
\tag{B.1}
$$

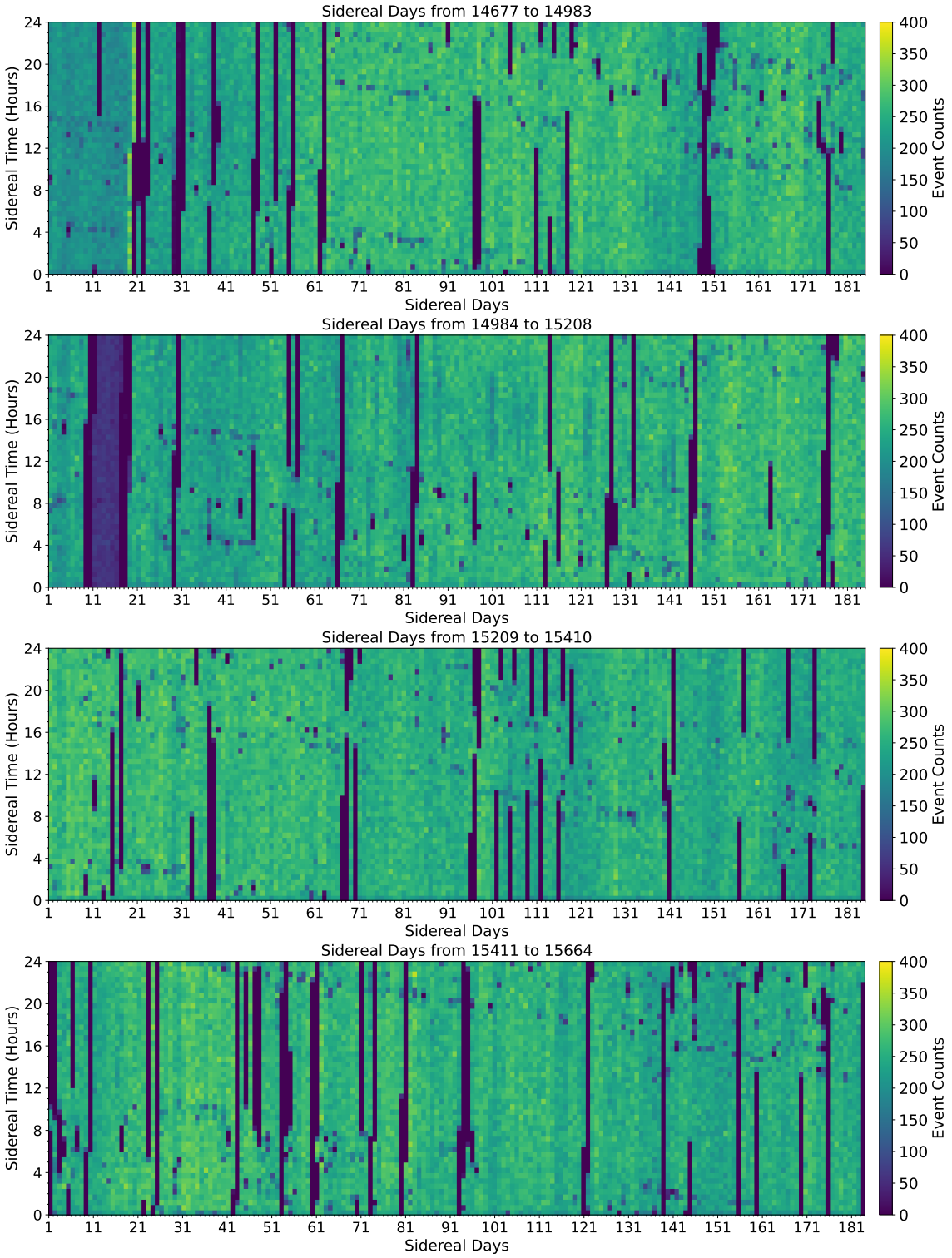
The traditional arctan function, $\alpha = \arctan(y/x)$, only returns values between $-\pi/2$ and $\pi/2$, which is insufficient to determine the correct angle for points in all four quadrants. By using arctan2, we ensure that the computed angle accurately reflects the position of the point (x, y) in the plane, taking into account the signs of both x and y.

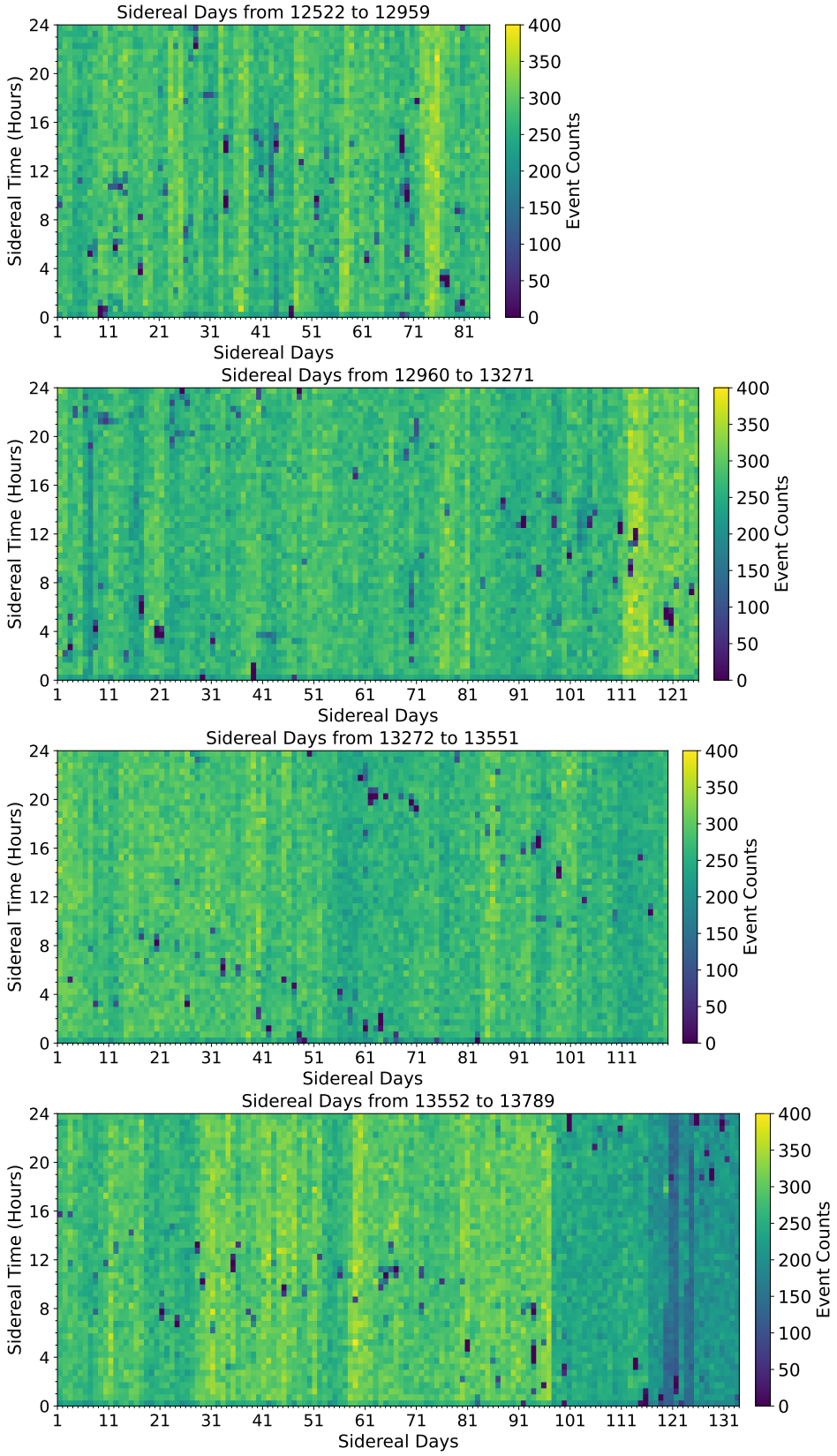
C KASCADE-Grande Data

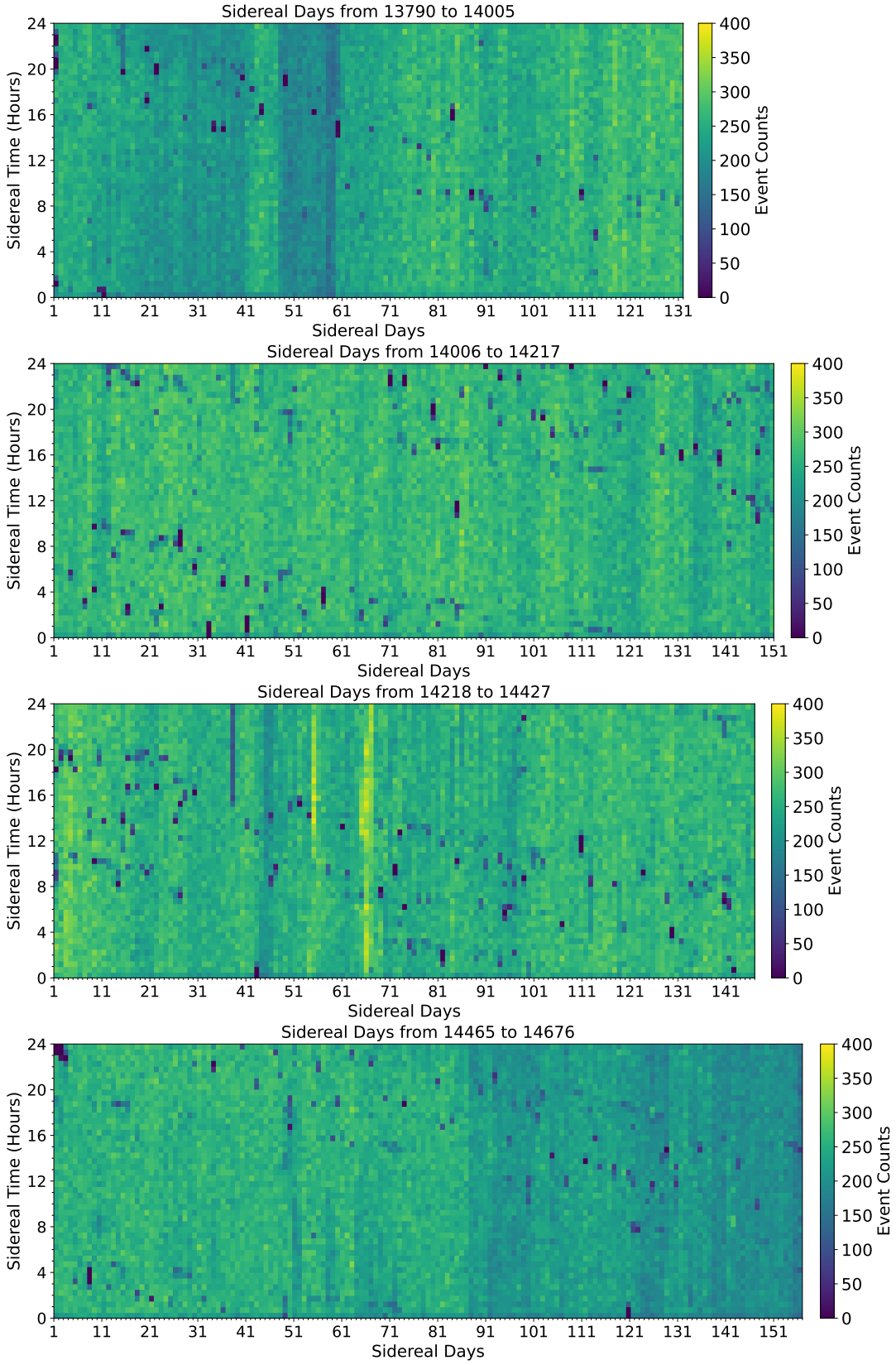
The distributions of the total data for KASCADE-Grande, collected from March 12, 2004, to October 9, 2012, are shown below, including only non-empty sidereal days. The raw data contain timestamp gaps due to detector downtime, maintenance periods, and acquisition interruptions, resulting in non-uniform time coverage. Non-valid sidereal days with a total gap of ≥ 4 hours and ≤ 100 events in any half-hour bin have been removed, and the cleaned data set is shown afterwards. The detailed time distributions in both Local Sidereal Time (LST) and Coordinated Universal Time (UTC) are shown below.

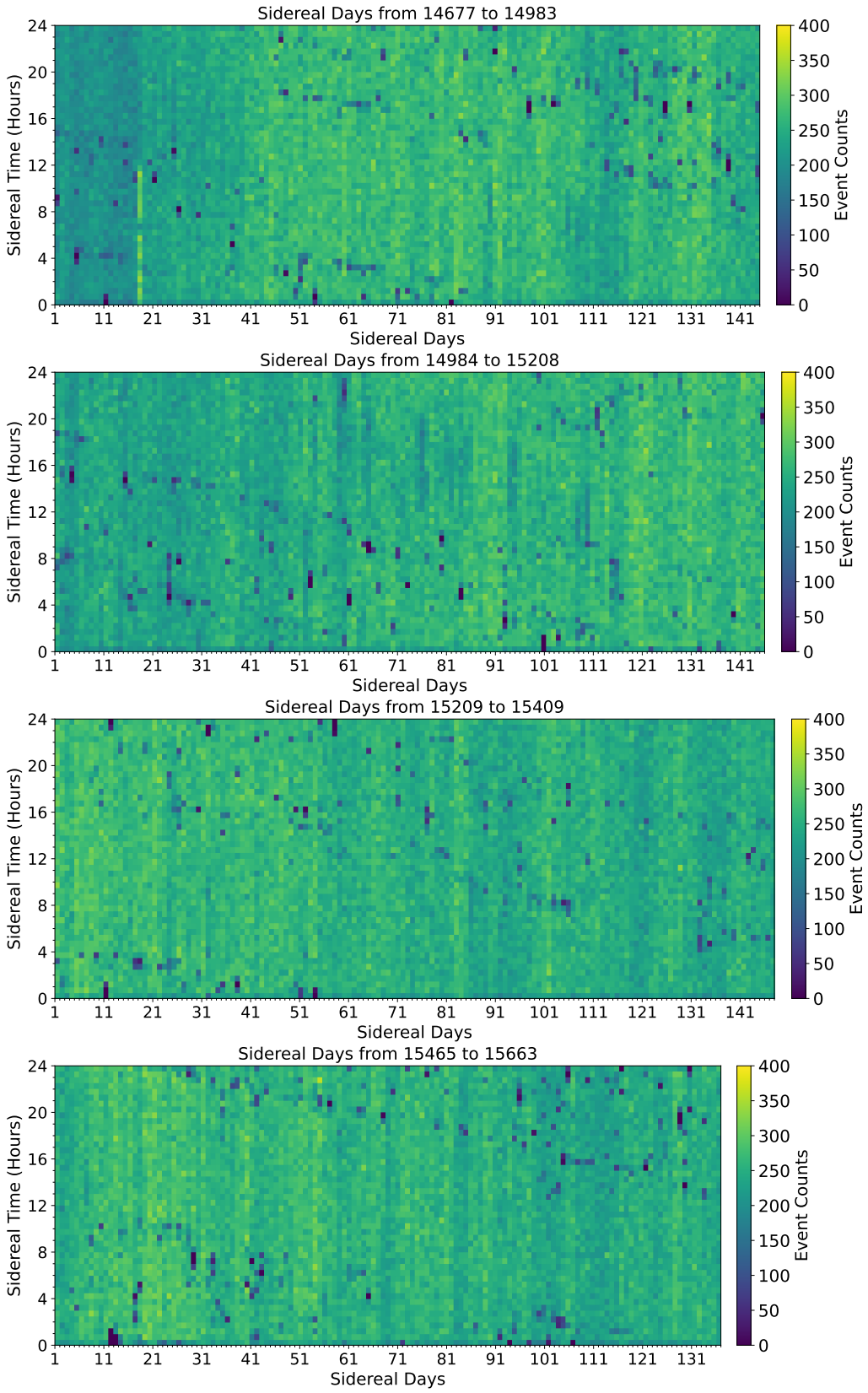

Sidereal Days from 12520 to 12959
Sidereal Days from 12960 to 13271
Sidereal Days from 13272 to 13551
Sidereal Days from 13552 to 13789
Sidereal Time (Hours)
Sidereal Days
Event Counts

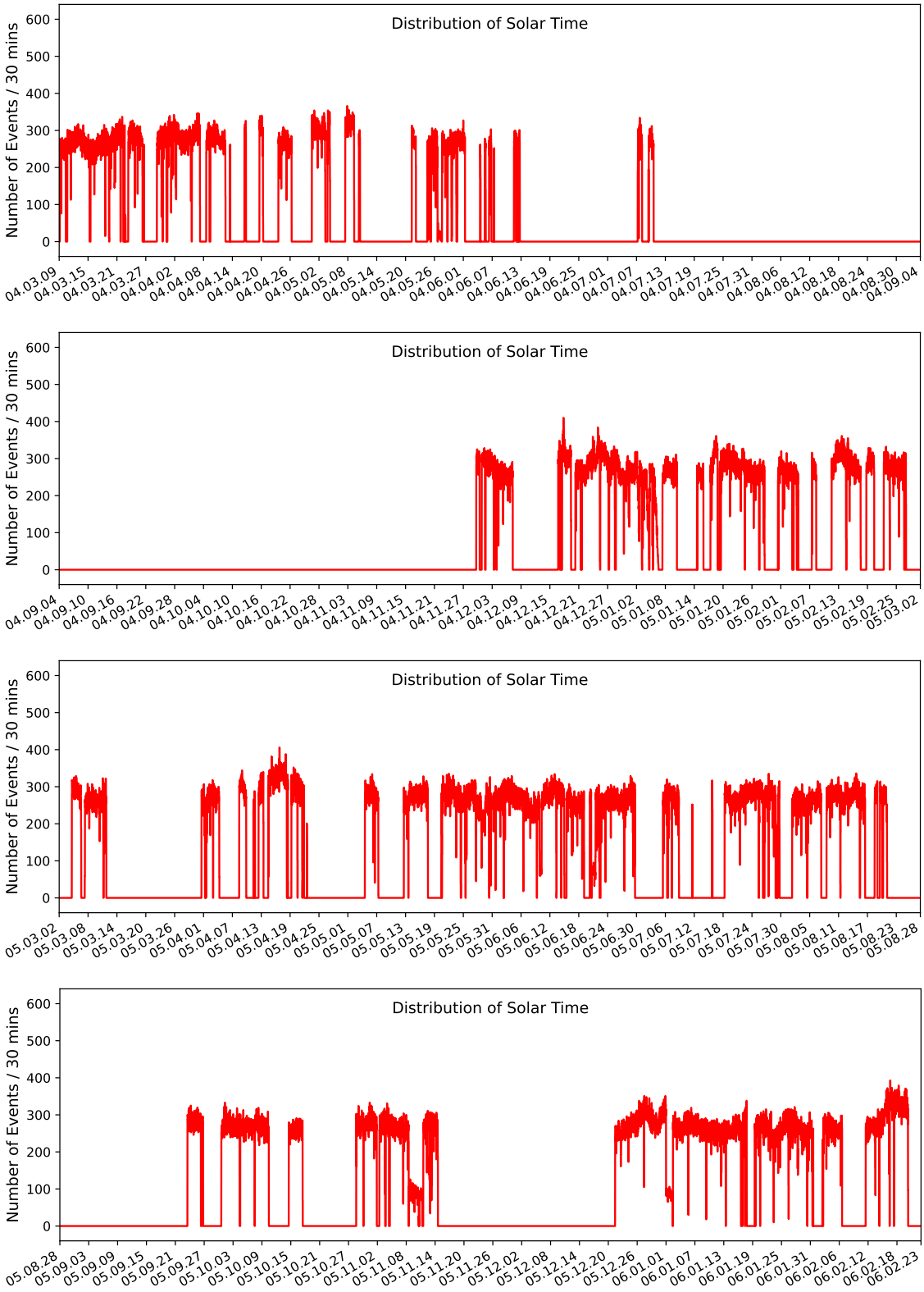

Sidereal Days from 13790 to 14005
Sidereal Days from 14006 to 14217
Sidereal Days from 14218 to 14428
Sidereal Days from 14429 to 14676
Sidereal Time (Hours)
Sidereal Days
Event Counts

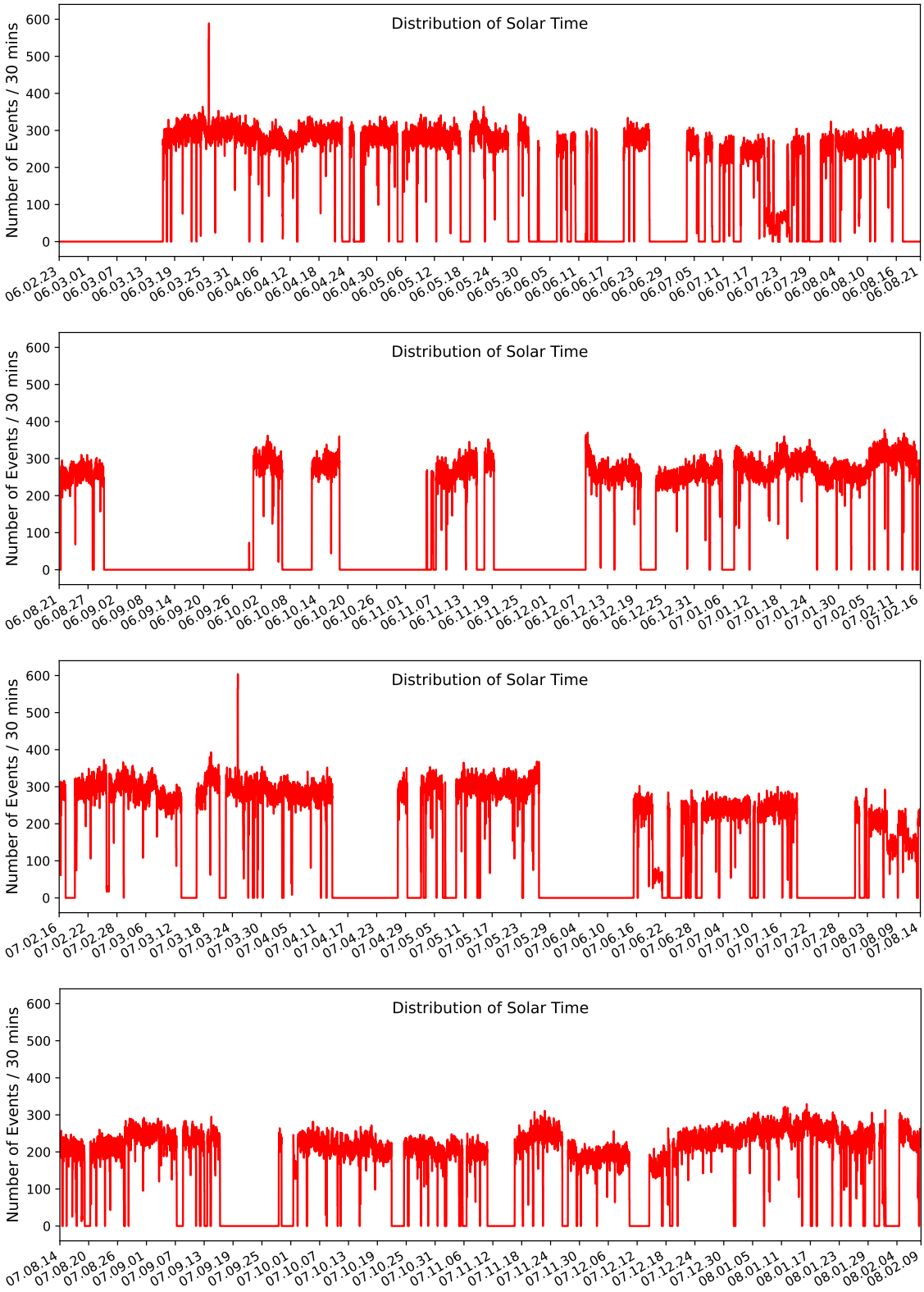

Sidereal Days from 14677 to 14983
Sidereal Days from 14984 to 15208
Sidereal Days from 15209 to 15410
Sidereal Days from 15411 to 15664
Sidereal Time (Hours)
Sidereal Days
Event Counts

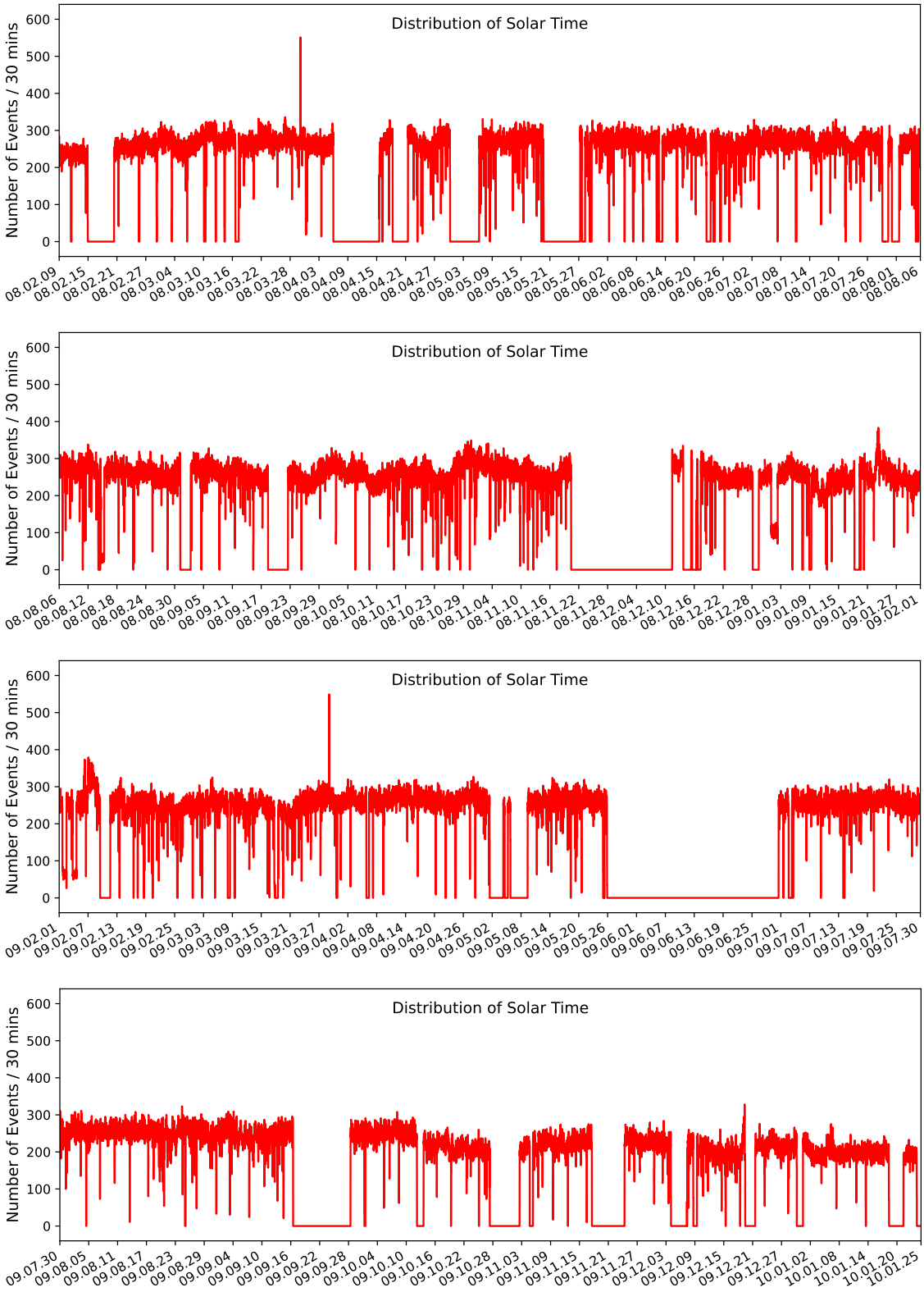

Sidereal Days from 12522 to 12959
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 12960 to 13271
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 13272 to 13551
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 13552 to 13789
Sidereal Time (Hours)
Sidereal Days
Event Counts

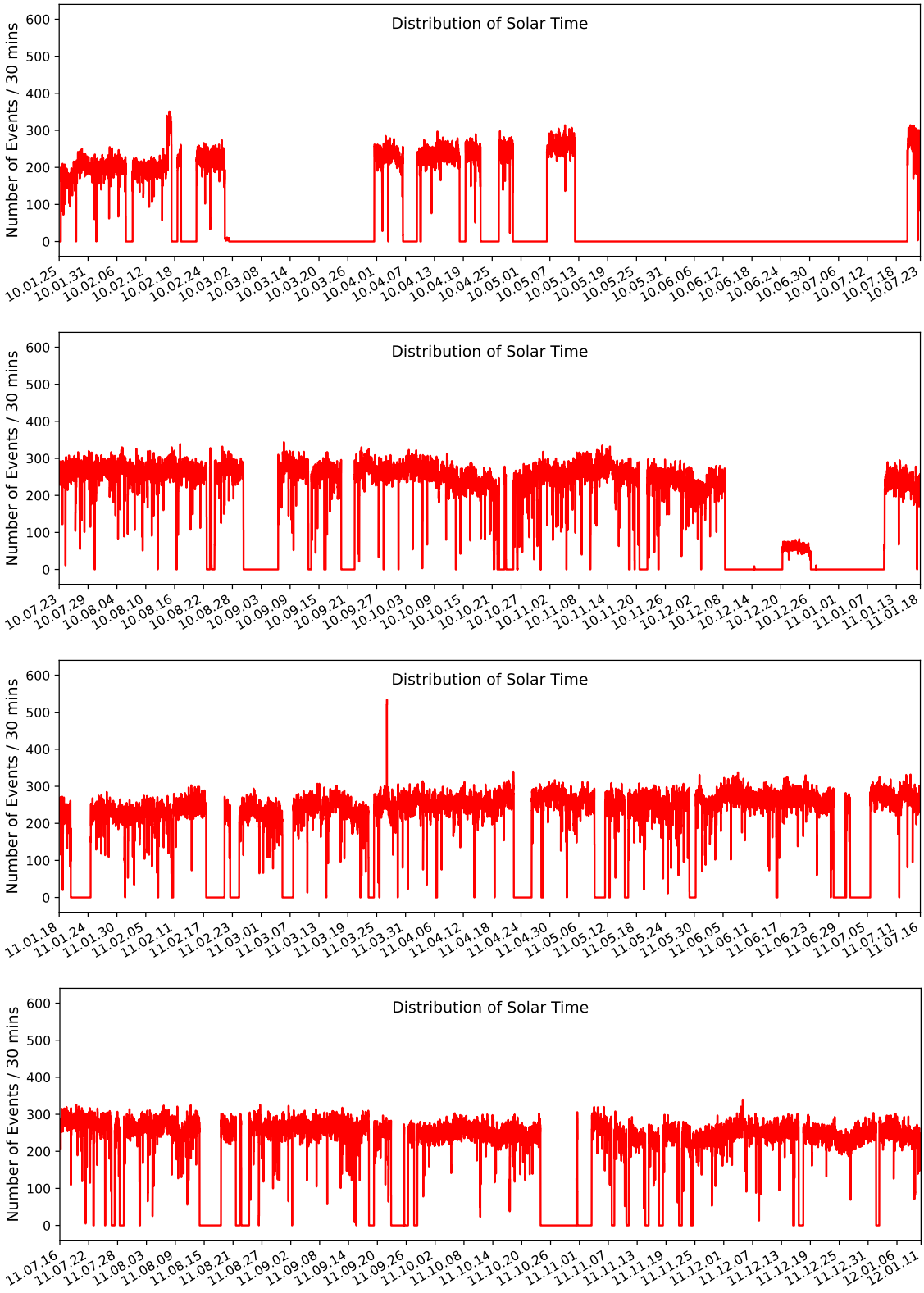

Sidereal Days from 13790 to 14005
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 14006 to 14217
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 14218 to 14427
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 14465 to 14676
Sidereal Time (Hours)
Sidereal Days
Event Counts

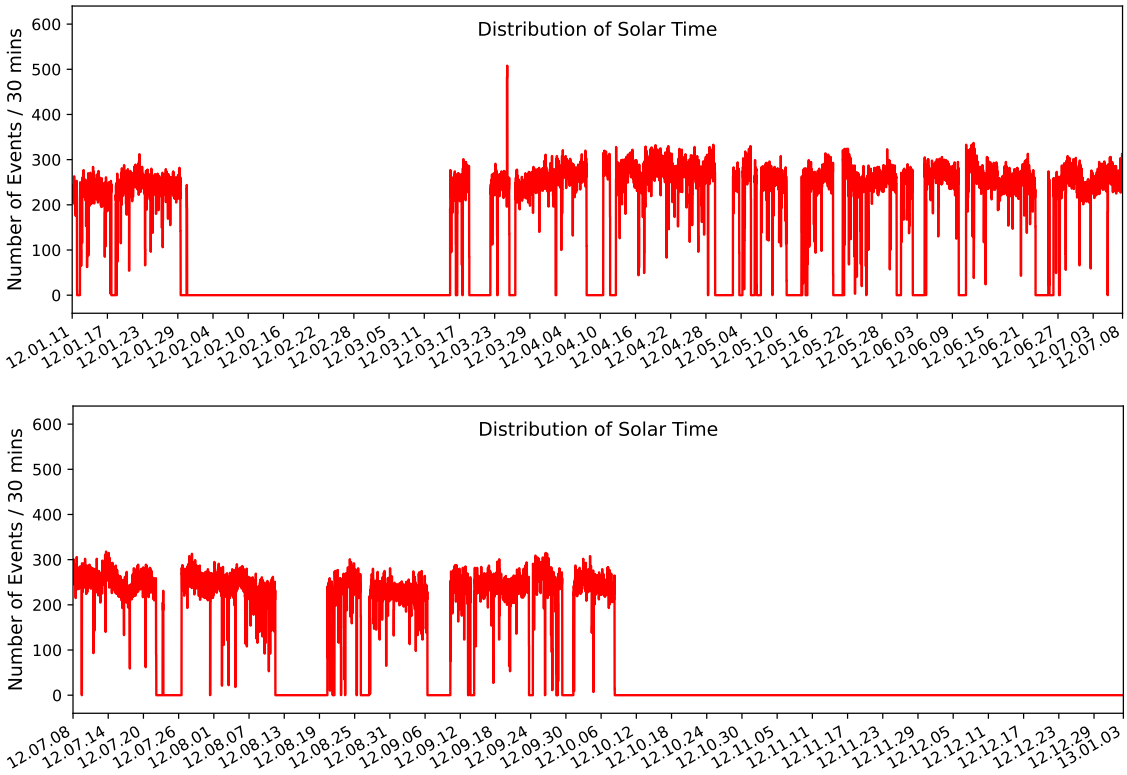

Sidereal Days from 14677 to 14983
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 14984 to 15208
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 15209 to 15409
Sidereal Time (Hours)
Sidereal Days
Event Counts
Sidereal Days from 15465 to 15663
Sidereal Time (Hours)
Sidereal Days
Event Counts


Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
04.03.09
04.03.15
04.03.21
04.03.27
04.04.02
04.04.08
04.04.14
04.04.20
04.04.26
04.05.02
04.05.08
04.05.14
04.05.20
04.05.26
04.06.01
04.06.07
04.06.13
04.06.19
04.06.25
04.07.01
04.07.07
04.07.13
04.07.19
04.07.25
04.07.31
04.08.06
04.08.12
04.08.18
04.08.24
04.08.30
04.09.04
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
04.09.04
04.09.10
04.09.16
04.09.22
04.09.28
04.10.04
04.10.10
04.10.16
04.10.22
04.10.28
04.11.03
04.11.09
04.11.15
04.11.21
04.11.27
04.12.03
04.12.09
04.12.15
04.12.21
04.12.27
05.01.02
05.01.08
05.01.14
05.01.20
05.01.26
05.02.01
05.02.07
05.02.13
05.02.19
05.02.25
05.03.02
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
05.03.02
05.03.08
05.03.14
05.03.20
05.03.26
05.04.01
05.04.07
05.04.13
05.04.19
05.04.25
05.05.01
05.05.07
05.05.13
05.05.19
05.05.25
05.05.31
05.06.06
05.06.12
05.06.18
05.06.24
05.06.30
05.07.06
05.07.12
05.07.18
05.07.24
05.07.30
05.08.05
05.08.11
05.08.17
05.08.23
05.08.28
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
05.08.28
05.09.03
05.09.09
05.09.15
05.09.21
05.09.27
05.10.03
05.10.09
05.10.15
05.10.21
05.10.27
05.11.02
05.11.08
05.11.14
05.11.20
05.11.26
05.12.02
05.12.08
05.12.14
05.12.20
05.12.26
06.01.01
06.01.07
06.01.13
06.01.19
06.01.25
06.01.31
06.02.06
06.02.12
06.02.18
06.02.23


Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
06.02.23
06.03.01
06.03.07
06.03.13
06.03.19
06.03.25
06.03.31
06.04.06
06.04.12
06.04.18
06.04.24
06.04.30
06.05.06
06.05.12
06.05.18
06.05.24
06.05.30
06.06.05
06.06.11
06.06.17
06.06.23
06.06.29
06.07.05
06.07.11
06.07.17
06.07.23
06.07.29
06.08.04
06.08.10
06.08.16
06.08.21
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
06.08.21
06.08.27
06.09.02
06.09.08
06.09.14
06.09.20
06.09.26
06.10.02
06.10.08
06.10.14
06.10.20
06.10.26
06.11.01
06.11.07
06.11.13
06.11.19
06.11.25
06.12.01
06.12.07
06.12.13
06.12.19
06.12.25
06.12.31
07.01.06
07.01.12
07.01.18
07.01.24
07.01.30
07.02.05
07.02.11
07.02.16
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
07.02.16
07.02.22
07.02.28
07.03.06
07.03.12
07.03.18
07.03.24
07.03.30
07.04.05
07.04.11
07.04.17
07.04.23
07.04.29
07.05.05
07.05.11
07.05.17
07.05.23
07.05.29
07.06.04
07.06.10
07.06.16
07.06.22
07.06.28
07.07.04
07.07.10
07.07.16
07.07.22
07.07.28
07.08.03
07.08.09
07.08.14
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
07.08.14
07.08.20
07.08.26
07.09.01
07.09.07
07.09.13
07.09.19
07.09.25
07.10.01
07.10.07
07.10.13
07.10.19
07.10.25
07.10.31
07.11.06
07.11.12
07.11.18
07.11.24
07.11.30
07.12.06
07.12.12
07.12.18
07.12.24
07.12.30
08.01.05
08.01.11
08.01.17
08.01.23
08.01.29
08.02.04
08.02.09


Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
08.02.09
08.02.15
08.02.21
08.02.27
08.03.04
08.03.10
08.03.16
08.03.22
08.03.28
08.04.03
08.04.09
08.04.15
08.04.21
08.04.27
08.05.03
08.05.09
08.05.15
08.05.21
08.05.27
08.06.02
08.06.08
08.06.14
08.06.20
08.06.26
08.07.02
08.07.08
08.07.14
08.07.20
08.07.26
08.08.01
08.08.06
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
08.08.06
08.08.12
08.08.18
08.08.24
08.08.30
08.09.05
08.09.11
08.09.17
08.09.23
08.09.29
08.10.05
08.10.11
08.10.17
08.10.23
08.10.29
08.11.04
08.11.10
08.11.16
08.11.22
08.11.28
08.12.04
08.12.10
08.12.16
08.12.22
08.12.28
09.01.03
09.01.09
09.01.15
09.01.21
09.01.27
09.02.01
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
09.02.01
09.02.07
09.02.13
09.02.19
09.02.25
09.03.03
09.03.09
09.03.15
09.03.21
09.03.27
09.04.02
09.04.08
09.04.14
09.04.20
09.04.26
09.05.02
09.05.08
09.05.14
09.05.20
09.05.26
09.06.01
09.06.07
09.06.13
09.06.19
09.06.25
09.07.01
09.07.07
09.07.13
09.07.19
09.07.25
09.07.30
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
09.07.30
09.08.05
09.08.11
09.08.17
09.08.23
09.08.29
09.09.04
09.09.10
09.09.16
09.09.22
09.09.28
09.10.04
09.10.10
09.10.16
09.10.22
09.10.28
09.11.03
09.11.09
09.11.15
09.11.21
09.11.27
09.12.03
09.12.09
09.12.15
09.12.21
09.12.27
10.01.02
10.01.08
10.01.14
10.01.20
10.01.25


Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
10.01.25
10.01.31
10.02.06
10.02.12
10.02.18
10.02.24
10.03.02
10.03.08
10.03.14
10.03.20
10.03.26
10.04.01
10.04.07
10.04.13
10.04.19
10.04.25
10.05.01
10.05.07
10.05.13
10.05.19
10.05.25
10.05.31
10.06.06
10.06.12
10.06.18
10.06.24
10.06.30
10.07.06
10.07.12
10.07.18
10.07.23
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
10.07.23
10.07.29
10.08.04
10.08.10
10.08.16
10.08.22
10.08.28
10.09.03
10.09.09
10.09.15
10.09.21
10.09.27
10.10.03
10.10.09
10.10.15
10.10.21
10.10.27
10.11.02
10.11.08
10.11.14
10.11.20
10.11.26
10.12.02
10.12.08
10.12.14
10.12.20
10.12.26
11.01.01
11.01.07
11.01.13
11.01.18
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
11.01.18
11.01.24
11.01.30
11.02.05
11.02.11
11.02.17
11.02.23
11.03.01
11.03.07
11.03.13
11.03.19
11.03.25
11.03.31
11.04.06
11.04.12
11.04.18
11.04.24
11.04.30
11.05.06
11.05.12
11.05.18
11.05.24
11.05.30
11.06.05
11.06.11
11.06.17
11.06.23
11.06.29
11.07.05
11.07.11
11.07.16
Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
11.07.16
11.07.22
11.07.28
11.08.03
11.08.09
11.08.15
11.08.21
11.08.27
11.09.02
11.09.08
11.09.14
11.09.20
11.09.26
11.10.02
11.10.08
11.10.14
11.10.20
11.10.26
11.11.01
11.11.07
11.11.13
11.11.19
11.11.25
11.12.01
11.12.07
11.12.13
11.12.19
11.12.25
11.12.31
12.01.06
12.01.11


Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
12.01.11
12.01.17
12.01.23
12.01.29
12.02.04
12.02.10
12.02.16
12.02.22
12.02.28
12.03.05
12.03.11
12.03.17
12.03.23
12.03.29
12.04.04
12.04.10
12.04.16
12.04.22
12.04.28
12.05.04
12.05.10
12.05.16
12.05.22
12.05.28
12.06.03
12.06.09
12.06.15
12.06.21
12.06.27
12.07.03
12.07.08

Distribution of Solar Time
Number of Events / 30 mins
0
100
200
300
400
500
600
12.07.08
12.07.14
12.07.20
12.07.26
12.08.01
12.08.07
12.08.13
12.08.19
12.08.25
12.08.31
12.09.06
12.09.12
12.09.18
12.09.24
12.09.30
12.10.06
12.10.12
12.10.18
12.10.24
12.10.30
12.11.05
12.11.11
12.11.17
12.11.23
12.11.29
12.12.05
12.12.11
12.12.17
12.12.23
12.12.29
13.01.03

Bibliography

[1] T. K. Gaisser, "Cosmic Rays and Particle Physics," Cambridge University Press, 1990.

[2] R. Blandford and D. Eichler, "Particle Acceleration at Astrophysical Shocks: A Theory of Cosmic Ray Origin," Phys. Rept. 154, 1–75 (1987).

[3] P. Blasi, "The Origin of Galactic Cosmic Rays," Astron. Astrophys. Rev. 21, 70 (2013), arXiv: 1311.7346.

[4] S. Gabici et al., "The origin of Galactic cosmic rays: challenges to the standard paradigm," Int. J. Mod. Phys. D 28, 1930022 (2019), arXiv: 1903.11584.

[5] M. Ahlers and P. Mertsch, "Origin of Small-Scale Anisotropies in Galactic Cosmic Rays," Prog. Part. Nucl. Phys. 94, 184–216 (2017), arXiv: 1612.01873.

[6] A. Abramowski et al., "Acceleration of petaelectronvolt protons in the Galactic Centre," Nature 531, 476–479 (2016), arXiv: 1603.07730.

[7] A. U. Abeysekara et al., "HAWC observations of the acceleration of very-high-energy cosmic rays in the Cygnus Cocoon," Nature Astron. 5, 465–471 (2021), arXiv: 2103.06820.

[8] Z. Cao et al., "An ultrahigh-energy γ-ray bubble powered by a super PeVatron," Sci. Bull. 69, 449–457 (2024), arXiv: 2310.10100.

[9] Z. Cao et al., "Ultrahigh-energy gamma-ray emission associated with black hole-jet systems," Nat. Sci. Rev., nwaf496 (2025), arXiv: 2410.08988.

[10] Z. Cao et al., "Precise measurements of the cosmic-ray proton energy spectrum in the 'knee' region," Sci. Bull., 2095–9273 (2025), arXiv: 2511.05013.

[11] A. Aab et al., "Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory," Phys. Rev. D 102 (2020), arXiv: 2008.06486.

[12] V. B. R. Aloisio and A. Gazizov, "Transition from galactic to extragalactic cosmic rays," Astropart. Phys. 39-40, 129–143 (2012), arXiv: 1211.0494.

[13] R. A. Batista et al., "Open Questions in Cosmic-Ray Research at Ultrahigh Energies," Front. Astron. Space Sci. 6, 23 (2019), arXiv: 1903.06714.

[14] R. Aloisio et al., "A dip in the UHECR spectrum and the transition from galactic to extragalactic cosmic rays," Astropart. Phys. 27, 76–91 (2007), arXiv: astro-ph/0608219.

[15] M. Amenomori et al., "Anisotropy and Corotation of Galactic Cosmic Rays," Science 314, 439–443 (2006), arXiv: astro-ph/0610671.

[16] A. U. Abeysekara et al., "All-Sky Measurement of the Anisotropy of Cosmic Rays at 10 TeV and Mapping of the Local Interstellar Magnetic Field," Astrophys. J. 871, 96 (2019), arXiv: 1812.05682.

[17] M. Ahlers, "Anomalous Anisotropies of Cosmic Rays from Turbulent Magnetic Fields," Phys. Rev. Lett. 117, 151103 (2016), arXiv: 1605.06446.

[18] A. Aab et al., "Large-scale cosmic-ray anisotropies above 4 EeV measured by the Pierre Auger Observatory," Astrophys. J. 868, 4 (2018), arXiv: 1808.03579.

[19] T. K. Gaisser, R. Engel, and E. Resconi, "Cosmic Rays and Particle Physics," Cambridge University Press, 2016.

[20] V. F. Hess, "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten," Phys. Z. 13, 1084–1091 (1912).

[21] A. Aab et al., "The Pierre Auger Cosmic Ray Observatory," Nucl. Instrum. Meth. A 798, 172–213 (2015), arXiv: 1502.01323.

[22] T. Antoni et al., "The cosmic-ray experiment KASCADE," Nucl. Instrum. Meth. A 513, 490–510 (2003).

[23] T. Antoni et al., "KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems," Astropart. Phys. 24, 1–25 (2005), arXiv: astro-ph/0505413.

[24] A. Haungs et al., "The KASCADE Cosmic-ray Data Centre KCDC: Granting Open Access to Astroparticle Physics Research Data," Eur. Phys. J. C 78, 741 (2018), arXiv: 1806.05493.

[25] A. Addazi et al., "The Large High Altitude Air Shower Observatory (LHAASO) Science Book (2021 Edition)," Chin. Phys. C 46, 035001–035007 (2022), arXiv: 1905.02773.

[26] M. G. Aartsen et al., "The IceCube Neutrino Observatory: Instrumentation and Online Systems," JINST 12, P03012 (2017), arXiv: 1612.05093.

[27] M. G. Aartsen et al., "IceCube-Gen2: the window to the extreme Universe," J.Phys.G 48, 060501 (2021), arXiv: 2008.04323.

[28] Y. S. Yoon et al., "Cosmic-Ray Proton and Helium Spectra from the First CREAM Flight," Astrophys. J. 728, 122 (2011), arXiv: 1102.2575.

[29] Y. S. Yoon et al., "Proton and Helium Spectra from the CREAM-III Flight," Astrophys. J. 839, 5 (2017), arXiv: 1704.02512.

[30] M. Aguilar et al., "Precision measurement of the proton flux in primary cosmic rays from rigidity 1 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station," Phys. Rev. Lett. 114, 171103 (2015).

[31] J. Chang et al., "The Dark Matter Particle Explorer mission," Astropart. Phys. 95, 6–24 (2017), arXiv: 1706.08453.

[32] J. Alvarez-Muniz et al., "The Giant Radio Array for Neutrino Detection (GRAND): Science and Design," Sci. China Phys. Mech. Astron. 63, 219501 (2020), arXiv: 1810.09994.

[33] S. N. Zhang et al., "The high energy cosmic-radiation detection (HERD) facility onboard China's Space Station," Proc. SPIE Int. Soc. Opt. Eng. 9144, 91440X (2014), arXiv: 1407.4866.

[34] A. Albert et al., "Science Case for a Wide Field-of-View Very-High-Energy Gamma-Ray Observatory in the Southern Hemisphere," (2019), arXiv: 1902.08429.

[35] M. Ahlers et al., "Ideas and Requirements for the Global Cosmic-Ray Observatory (GCOS)," (2025), arXiv: 2502.05657.

[36] P. Abreu et al., "Science Prospects for the Southern Wide-field Gamma-ray Observatory: SWGO," (2025), arXiv: 2506.01786.

[37] D. Pacini, "Penetrating Radiation at the Surface of and in Water," Nuovo Cim. 8, 93–100 (1912), arXiv: 1002.1810.

[38] W. Kolhörster, "Messungen der durchdringenden Strahlungen im Freiballon in größeren Höhen," Physik. Zschr. 14, 1153–1156 (1913).

[39] W. Kolhörster, "Messungen der durchdringenden Strahlungen bis in Höhen von 9,300 m," Verh. Dtsch. Phys. Ges. 16, 719–721 (1914).

[40] R. A. Millikan, "History of Research in Cosmic Rays," Nature 126, 14–16 (1930).

[41] R. A. Millikan, "High Frequency Rays of Cosmic Origin," Proc. Natl. Acad. Sci. USA 12, 48–55 (1926).

[42] R. A. Millikan and I. S. Bowen, "High Frequency Rays of Cosmic Origin I. Sounding Balloon Observations at Extreme Altitudes," Phys. Rev. 27, 353–361 (1926).

[43] R. A. Millikan and R. M. Otis, "High Frequency Rays of Cosmic Origin II. Mountain Peak and Airplane Observations," Phys. Rev. 27, 645–658 (1926).

[44] R. A. Millikan and G. H. Cameron, "The Origin of the Cosmic Rays," Phys. Rev. 32, 533–557 (1928).

[45] W. Bothe and W. Kolhörster, "The Nature of the Penetrating Radiation," Nature 123, 638 (1929).

[46] R. Bruno, "On the Magnetic Deflection of Cosmic Rays," Phys. Rev. 36, 606 (1930).

[47] P. Auger et al., "Extensive cosmic-ray showers," Rev. Mod. Phys. 11, 288–291 (1939).

[48] J. Linsley, L. Scarsi, and B. Rossi, "Extremely energetic cosmic-ray event," Phys. Rev. Lett. 6, 485–487 (1961).

[49] J. Linsley, "Evidence for a primary cosmic-ray particle with energy 10^{20} eV," Phys. Rev. Lett. 10, 146–148 (1963).

[50] G. V. Kulikov and G. B. Khristiansen, "On the size spectrum of extensive air showers," Sov. Phys. JETP 35, 441–444 (1958).

[51] B. Peters, "Primary cosmic radiation and extensive air showers," Nuovo Cim. 22, 800–819 (1961).

[52] A. E. Chudakov and N. M. Nesterova, "Cherenkov radiation of extensive air showers," Nuovo Cimento 8, Suppl. 2, 606–613 (1957).

[53] M. Friedlander, "Physics: A century of cosmic rays," Nature 483N7390, 400–401 (2012).

[54] K. Greisen, "End to the cosmic ray spectrum?," Phys. Rev. Lett. 16, 748–750 (1966).

[55] G. T. Zatsepin and V. A. Kuzmin, "Upper limit of the spectrum of cosmic rays," JETP Lett. 4, 78–80 (1966).

[56] M. Aguilar et al., "First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–350 GeV," Phys. Rev. Lett. 110, 141102 (2013).

[57] M. Aguilar et al., "Precision Measurement of the Proton Flux in Primary Cosmic Rays from Rigidity 1 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station," Phys. Rev. Lett. 114, 171103 (2015).

[58] E. S. Seo, "Direct measurements of cosmic rays using balloon borne experiments," Astropart. Phys. 39-40, 76–87 (2012).

[59] T. Shoji et al., "The CALorimetric Electron Telescope (CALET) on the International Space Station," Adv. Space Res. 64, 2531–2537 (2019).

[60] A. M. Galper et al., "The PAMELA experiment: a decade of Cosmic Ray Physics in space," J. Phys. Conf. Ser. 798, 012033 (2017).

[61] M. Amenomori et al., "Development and a Performance Test of a Prototype Air Shower Array for Search for gamma-ray Point Sources in the Very High-energy Region," Nucl. Instrum. Meth. A 288, 619 (1990).

[62] H. Kawai et al., "Telescope array experiment," Nucl. Phys. B Proc. Suppl. 175-176, 221–226 (2008).

[63] M. G. Aartsen et al., "Measurement of the cosmic ray energy spectrum with IceTop-73," Phys. Rev. D 88, 042004 (2013), arXiv: 1307.3795.

[64] T. Abu-Zayyad et al., "Evidence for Changing of Cosmic Ray Composition between 10^{17} eV and 10^{18} eV from Multicomponent Measurements," Phys. Rev. Lett. 84, 4276–4279 (2000), arXiv: astro-ph/9911144.

[65] M. Aglietta et al., "The cosmic ray primary composition in the 'knee' region through the EAS electromagnetic and muon measurements at EAS-TOP," Astropart. Phys. 21, 583–596 (2004).

[66] R. Abbasi et al., "IceTop: The surface component of IceCube," Nucl. Instrum. Meth. A 700, 188–220 (2013), arXiv: 1207.6326.

[67] J. Abraham et al., "Observation of the suppression of the flux of cosmic rays above 4×10^{19}eV," Phys. Rev. Lett. 101, 061101 (2008), arXiv: 0806.4302.

[68] R. Abbasi et al., "IceTop: The surface component of IceCube," Nucl. Instrum. Meth. A 700, 188–220 (2013), arXiv: 1207.6326.

[69] R. Abbasi et al., "Lateral Distribution of Muons in IceCube Cosmic Ray Events," Phys. Rev. D 87, 012005 (2013), arXiv: 1208.2979.

[70] M. G. Aartsen et al., "Cosmic ray spectrum and composition from PeV to EeV using 3 years of data from IceTop and IceCube," Phys. Rev. D 100, 082002 (2019), arXiv: 1906.04317.

[71] H. M. J. Barbosa et al., "Determination of the calorimetric energy in extensive air showers," Astropart. Phys. 22, 159–166 (2004), arXiv: astro-ph/0310234.

[72] A. Leszczyńska and M. Plum, "Simulation and Reconstruction Study of a Future Surface Scintillator Array at the IceCube Neutrino Observatory," PoS ICRC2019, 332 (2020), arXiv: 1909.02258.

[73] T. Abu-Zayyad et al., "The surface detector array of the Telescope Array experiment," Nucl. Instrum. Meth. A 689, 87–97 (2013), arXiv: 1201.4964.

[74] T. Huege, "Radio detection of cosmic ray air showers in the digital era," Phys. Rept. 620, 1–52 (2016), arXiv: 1601.07426.

[75] J. V. Jelley et al., "Radio Pulses from Extensive Cosmic-Ray Air Showers," Nature 205, 327–328 (1965).

[76] M. P. van Haarlem et al., "LOFAR: The LOw-Frequency ARray," Astron. Astrophys. 556, A2 (2013), arXiv: 1305.3550.

[77] T. Huege, "Radio detection of cosmic rays with the Auger Engineering Radio Array," EPJ Web Conf. 210, 05011 (2019), arXiv: 1905.04986.

[78] L. Wang et al., "Design and Development of Energy Particle Detector on China's Chang'e-7," Aerospace 11, 893 (2024).

[79] G. Silvestre, "The Silicon Charge Detector of the High Energy Cosmic Radiation Detection experiment," JINST 19, C03042 (2024).

[80] J. Hinton, "The Southern Wide-field Gamma-ray Observatory: Status and Prospects," PoS ICRC2021, 023 (2021), arXiv: 2111.13158.

[81] A. Aab et al., "The Pierre Auger Observatory Upgrade - Preliminary Design Report," (2016), arXiv: 1604.03637.

[82] J. Stasielak, "AugerPrime - The upgrade of the Pierre Auger Observatory," Int. J. Mod. Phys. A 37, 2240012 (2022), arXiv: 2110.09487.

[83] A. Ishihara et al., "The next generation neutrino telescope: IceCube-Gen2," PoS ICRC2023, 994 (2023), arXiv: 2308.09427.

[84] W. Hou et al., "Sensitivity of the IceCube-Gen2 Surface Array for Cosmic-Ray Anisotropy Studies," PoS ICRC2023, 354 (2023), arXiv: 2307.14655v1.

[85] A. Coleman et al., "The Surface Array of IceCube-Gen2," PoS ICRC2023, 205 (2023), arXiv: 2308.02990.

[86] D. Kyratzis et al., "The Plastic Scintillator Detector of the HERD space mission," PoS ICRC2023, 140 (2023).

[87] W. D. Apel et al., "Search for Large-scale Anisotropy in the Arrival Direction of Cosmic Rays with KASCADE-Grande," Astrophys. J. 870, 91 (2019).

[88] A. Aab et al., "Observation of a Large-Scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV," Science 357, 1266–1270 (2017), arXiv: 1709.07321.

[89] A. A. Halim et al., "Large-scale Cosmic-ray Anisotropies with 19 yr of Data from the Pierre Auger Observatory," Astrophys. J. 976, 48 (2024), arXiv: 2408.05292.

[90] R. Engel, D. Heck, and T. Pierog, "Extensive air showers and hadronic interactions at high energy," Annu. Rev. Nucl. Part. Sci. 61, 467–489 (2011).

[91] A. Haungs, H. Rebel, and M. Roth, "Energy Spectrum and Mass Composition of High-Energy Cosmic Rays," Rep. Prog. Phys. 66, 1145–1206 (2003).

[92] F. G. Schröder, "Radio Detection of Cosmic-Ray Air Showers and High-Energy Neutrinos," Prog. Part. Nucl. Phys. 93, 1–68 (2017), arXiv: 1607.08781.

[93] J. Matthews, "A Heitler model of extensive air showers," Astropart. Phys. 22, 387–397 (2005).

[94] W. Heitler, "The Quantum Theory of Radiation," Oxford University Press, 1944.

[95] J. Engel, T. K. Gaisser, T. Stanev, and P. Lipari, "Nucleus-nucleus collisions and interpretation of cosmic ray cascades," Phys. Rev. D 46, 5013–5025 (1992).

[96] H. P. Dembinski et al., "Report on Tests and Measurements of Hadronic Interaction Properties with Air Showers," EPJ Web Conf. 210, 02004 (2019), arXiv: 1902.08124.

[97] J. Albrecht et al., "The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider," Astrophys. Space Sci. 367, 27 (2022), arXiv: 2105.06148.

[98] K. Kampert and M. Unger, "Measurements of the Cosmic Ray Composition with Air Shower Experiments," Astropart. Phys. 35, 660–678 (2012), arXiv: 1201.0018.

[99] J. Adam et al., "Enhanced production of multi-strange hadrons in high-multiplicity proton-proton collisions," Nature Phys. 13, 535–539 (2017), arXiv: 1606.07424.

[100] A. Aab et al., "Muons in Air Showers at the Pierre Auger Observatory: Mean Number in Highly Inclined Events," Phys. Rev. D 91, 032003 (2015), arXiv: 1408.1421.

[101] A. Aab et al., "Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory," JCAP 04, 038 (2017), arXiv: 1612.07155.

[102] R. U. Abbasi et al., "Study of muons from ultrahigh energy cosmic ray air showers measured with the Telescope Array experiment," Phys. Rev. D 98, 022002 (2018), arXiv: 1804.03877.

[103] A. Aab et al., "Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory," Phys. Rev. Lett. 126, 152002 (2021), arXiv: 2102.07797.

[104] H. P. Dembinski et al., "Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV," PoS ICRC2017, 533 (2018), arXiv: 1711.11432.

[105] J. C. A. Velazquez, "A report by the WHISP working group on the combined analysis of muon data at cosmic-ray energies above 1 PeV," PoS ICRC2023, 466 (2023).

[106] S. Baur et al., "Core-corona effect in hadron collisions and muon production in air

showers," Phys. Rev. D 107, 094031 (2023), arXiv: 1902.09265.

[107] L. A. Anchordoqui et al., "Through the looking-glass with ALICE into the quark-gluon plasma: A new test for hadronic interaction models used in air shower simulations," Phys. Lett. B 810, 135837 (2020), arXiv: 1907.09816.

[108] Z. Citron et al., "Report from Working Group 5: Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams," CERN Yellow Rep. Monogr. 7, 1159–1410 (2019), arXiv: 1812.06772.

[109] T. Pierog et al., "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider," Phys. Rev. C 92, 034906 (2015), arXiv: 1306.0121.

[110] W. Apel et al., "The KASCADE-Grande experiment," Nucl. Instrum. Meth. A 620, 202–216 (2010).

[111] R. Aloisio, V. Berezinsky, and P. Blasi, "Ultra high energy cosmic rays: implications of Auger data for source spectra and chemical composition," JCAP 10, 020 (2014), arXiv: 1312.7459.

[112] L. Cazon and R. Ulrich, "The non-linearity between $\langle \ln A \rangle$ and $\langle X_{\max} \rangle$ induced by the acceptance of fluorescence telescopes," Astropart. Phys. 38, 41–45 (2012), arXiv: 1203.1781.

[113] A. Letessier-Selvon and T. Stanev, "Ultrahigh Energy Cosmic Rays," Rev. Mod. Phys. 83, 907–942 (2011), arXiv: 1103.0031.

[114] D. Biehl, D. Boncioli, A. Fedynitch, and W. Winter, "Cosmic-Ray and Neutrino Emission from Gamma-Ray Bursts with a Nuclear Cascade," Astron. Astrophys. 611, A101 (2018), arXiv: 1705.08909.

[115] R. U. Abbasi et al., "First observation of the Greisen-Zatsepin-Kuzmin suppression," Phys. Rev. Lett. 100, 101101 (2008), arXiv: astro-ph/0703099.

[116] W. D. Apel et al., "Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande," Phys. Rev. Lett. 107, 171104 (2011), arXiv: 1107.5885.

[117] A. Aab et al., "Measurement of the cosmic-ray energy spectrum above $2.5{\times}10^{18}$ eV using the Pierre Auger Observatory," Phys. Rev. D 102, 062005 (2020), arXiv: 2008.06486.

[118] S. Navas et al., "Review of Particle Physics," Phys. Rev. D 110, 030001 (2024).

[119] W. D. Apel et al., "KASCADE-Grande measurements of energy spectra for elemental groups of cosmic rays," Astropart. Phys. 47, 54–66 (2013), arXiv: 1306.6283.

[120] V. V. Prosin et al., "Tunka-133: Results of 3 year operation," Nucl. Instrum. Meth. A 756, 94–101 (2014).

[121] J. R. Horandel, "Cosmic rays from the knee to the second knee: 10^4 to 10^8 eV," Mod. Phys. Lett. A 22, 1533–1552 (2007), arXiv: astro-ph/0611387.

[122] R. D. Bergman and J. W. Belz, "Cosmic Rays: The Second Knee and Beyond," J. Phys. G 34, R359 (2007), arXiv: 0704.3721.

[123] D. J. Bird et al., "Evidence for correlated changes in the spectrum and composition of cosmic rays at extremely high-energies," Phys. Rev. Lett. 71, 3401–3404 (1993).

[124] D. J. Bird et al., "Detection of a cosmic ray with measured energy well beyond the expected spectral cutoff due to cosmic microwave radiation," Astrophys. J. 441, 144–150 (1995), arXiv: astro-ph/9410067.

[125] A. Esmaeili, A. Esmaili, and P. D. Serpico, "Pair production with capture by energetic cosmic ray nuclei in a photon background," Phys. Rev. D 111, 043004 (2025), arXiv: 2410.11958.

[126] H. Dembinski, R. Engel, A. Fedynitch, and K. Fujisue, "Global Spline Fit GSF-2025 - An update of the data-driven model of the cosmic-ray flux and its mass composition," PoS ICRC2025, 248 (2025).

[127] A. H. Compton and I. A. Getting, "An Apparent Effect of Galactic Rotation on the Intensity of Cosmic Rays," Phys. Rev. 47, 817 (1935).

[128] L. J. Gleeson and W. I. Axford, "The Compton-Getting Effect," Astrophysics and Space Science 2, 431–437 (1968).

[129] D. J. Cutler and D. E. Groom, "Observation of terrestrial orbital motion using the cosmic-ray Compton–Getting effect," Nature 322, 434–436 (1986).

[130] E. O. Wollan, "Present Status of Solar and Sidereal Time Variation of Cosmic Rays," Reviews of Modern Physics 11, 160–165 (1939).

[131] J. Linsley, "Fluctuation effects on directional data," Phys. Rev. Lett. 34, 1530–1533 (1975).

[132] K. Nagashima et al., "Galactic cosmic-ray anisotropy and its modulation in the heliomagnetosphere, inferred from air shower observation at Mt. Norikura," Il Nuovo Cimento C 12, 695–749 (1989).

[133] A. A. Abdo et al., "The Large Scale Cosmic-Ray Anisotropy as Observed with Milagro," Astrophys. J. 698, 2121–2130 (2009), arXiv: 0806.2293.

[134] G. Guillian et al., "Observation of the anisotropy of 10-TeV primary cosmic ray nuclei flux with the super-kamiokande-I detector," Phys.Rev.D 75, 062003 (2007), arXiv: astro-ph/0508468.

[135] B. Bartoli et al., "Argo-ybj Observation of the Large-scale Cosmic ray Anisotropy During the Solar Minimum Between Cycles 23 and 24," Astrophys. J. 809, 90 (2015), arXiv: 2112.14891.

[136] A. U. Abeysekara et al., "Observation of Anisotropy of TeV Cosmic Rays with Two Years of HAWC," Astrophys. J. 865, 57 (2018), arXiv: 1805.01847.

[137] W. Gao et al., "Observation of large-scale anisotropy in the arrival directions of cosmic rays with LHAASO," PoS ICRC2021, 351 (2021).

[138] M. G. Aartsen et al., "Observation of Cosmic Ray Anisotropy with the IceTop Air Shower Array," Astrophys. J. 765, 55 (2013), arXiv: 1210.5278.

[139] M. G. Aartsen et al., "Anisotropy in Cosmic-ray Arrival Directions in the Southern Hemisphere Based on six Years of Data From the Icecube Detector," Astrophys. J. 826, 220 (2016), arXiv: 1603.01227.

[140] R. Abbasi et al., "Observation of Cosmic-Ray Anisotropy in the Southern Hemisphere with 12 yr of Data Collected by the IceCube Neutrino Observatory," Astrophys. J. 981, 182 (2025), arXiv: 2412.05046.

[141] J. R. Hoerandel, "Models of the knee in the energy spectrum of cosmic rays," Astropart. Phys. 21, 241–265 (2004), arXiv: astro-ph/0402356.

[142] R. Abbasi et al., "Investigating Energy-Dependent Anisotropy in Cosmic Rays with IceTop Surface Array," PoS ICRC2025, 167 (2025), arXiv: 2507.23044.

[143] M. Ahlers et al., "A New Maximum-Likelihood Technique for Reconstructing Cosmic-Ray Anisotropy at All Angular Scales," Astrophys. J. 823, 10 (2016), arXiv: 1601.07877.

[144] A. M. Hillas, "Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays?," J. Phys. G 31, R95–R131 (2005).

[145] P. Abreu et al., "Update on the Correlation of the Highest Energy Cosmic Rays with Nearby Extragalactic matter," Astropart. Phys. 34, 314–326 (2010), arXiv:

1009.1855.

[146] C. Ramanath and H. Dawson, "The Correspondence between Leaky-box and Diffusion Models of Cosmic-Ray Propagation," Astrophys. J. 987, 64 (2025), arXiv: 2503.06281.

[147] M. V. K. A. Rao and M. F. Kaplon, "The influence of ionization loss on the shape of the differential primary cosmic ray energy-spectrum," Il Nuovo Cimento (1955-1965) 27, 700–717 (1963).

[148] V. K. Balasubrahmanyan, E. Boldt, and R. A. R. Palmeira, "Low-Energy Spectrum of Cosmic Rays as an Indicator of Primary Source Characteristics and Interstellar Propagation," Phys. Rev. 140, B1157–B1161 (1965).

[149] R. Cowsik, Y. Pal, S. N. Tandon, and R. P. Verma, "3° K Blackbody Radiation and Leakage Lifetime of Cosmic-Ray Electrons," Phys. Rev. Lett. 17, 1298–1300 (1966).

[150] R. Cowsik, Y. Pal, S. N. Tandon, and R. P. Verma, "Steady State of Cosmic-Ray Nuclei — Their Spectral Shape and Path Length at Low Energies," Phys. Rev. 158, 1238–1242 (1967).

[151] A. W. Strong et al., "Cosmic-ray propagation and interactions in the Galaxy," Ann. Rev. Nucl. Part. Sci. 57, 285–327 (2007), arXiv: astro-ph/0701517.

[152] S. P. Swordy et al., "Relative Abundances of Secondary and Primary Cosmic Rays at High Energies," Astrophys. J 349, 625 (1990).

[153] D. Maurin et al., "The importance of Fe fragmentation for LiBeB analyses: Is a Li primary source needed to explain AMS-02 data?," Astron. Astrophys. 668, A7 (2022), arXiv: 2203.00522.

[154] Y. Génolini et al., "Cosmic-ray transport from AMS-02 boron to carbon ratio data: Benchmark models and interpretation," Phys. Rev. D 99, 123028 (2019), arXiv: 1904.08917.

[155] A. Oliva, "Precision Measurement of Boron to Carbon flux ratio in Cosmic Rays from 2 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station." PoS ICRC2015, 265 (2016).

[156] A. Kolmogorov, "The Local Structure of Turbulence in Incompressible Viscous Fluid for Very Large Reynolds' Numbers," Akademiia Nauk SSSR Doklady 30, 301–305 (1941).

[157] A. Kolmogorov, "Dissipation of Energy in Locally Isotropic Turbulence," Akademiia Nauk SSSR Doklady 32, 16 (1941).

[158] R. H. Kraichnan, "Inertial-Range Spectrum of Hydromagnetic Turbulence," Physics of Fluids 8, 1385–1387 (1965).

[159] M. Meneguzzi, J. Audouze, and H. Reeves, "The Production of the Elements Li, Be, B by Galactic Cosmic Rays in Space and Its Relation with Stellar Observations," Astron. Astrophys. 15, 337 (1971).

[160] C. J. Cesarsky, "Galactic Cosmic Rays: Propagation and Origin," Springer Netherlands, 1987.

[161] R. A. Leske, "The Elemental and Isotopic Composition of Galactic Cosmic-Ray Nuclei from Scandium through Nickel," Astrophys. J 405, 567 (1993).

[162] N. E. Yanasak et al., "Measurement of the Secondary Radionuclides 10Be, 26Al, 36Cl, 54Mn, and 14C and Implications for the Galactic Cosmic-Ray Age," Astrophys. J 563, 768 (2001).

[163] M. Ave et al., "Propagation and Source Energy Spectra of Cosmic-Ray Nuclei at High Energies," Astrophys. J. 697, 106–114 (2009), arXiv: 0810.2972.

[164] N. E. Yanasak et al., "Elemental Composition and Energy Spectra of Galactic

Cosmic Rays During Solar Cycle 23," Astrophys. J. 698, 1666 (2009).

[165] S. Gabici, "Cosmic-Ray Physics (Direct, CRD): Rapporteur Talk," PoS ICRC2023, 030 (2024).

[166] V. L. Ginzburg and S. I. Syrovatskii, "The Origin of Cosmic Rays," Pergamon Press, 1964.

[167] C. S. Shen and C. Y. Mao, "Anisotropy of High Energy Cosmic-Ray Electrons in the Discrete Source Model," Astrophysical Letters 9, 169 (1971).

[168] R. Cowsik and M. A. Lee, "On the sources of cosmic ray electrons," Astrophysical Journal 228, 297–301 (1979).

[169] V. Ptuskin et al., "Diffusion and Drift of Galactic Cosmic Rays," Astron. Astrophys. 268, 726–734 (1993).

[170] E. S. Seo and V. S. Ptuskin, "Stochastic Reacceleration of Cosmic Rays in the Interstellar Medium," Astrophysical Journal 431, 705 (1994).

[171] D. Maurin, F. Donato, R. Taillet, and P. Salati, "Cosmic rays below Z=30 in a diffusion model: new constraints on propagation parameters," Astrophys. J. 555, 585–596 (2001), arXiv: astro-ph/0101231.

[172] I. V. Moskalenko and A. W. Strong, "Production and propagation of cosmic ray positrons and electrons," Astrophys. J. 493, 694–707 (1998), arXiv: astro-ph/9710124.

[173] A. W. Strong and I. V. Moskalenko, "Propagation of cosmic-ray nucleons in the galaxy," Astrophys. J. 509, 212–228 (1998), arXiv: astro-ph/9807150.

[174] T. A. Porter, G. Johannesson, and V. I. Moskalenko, "The GALPROP Cosmic-ray Propagation and Nonthermal Emissions Framework: Release v57," Astrophys. J. Supp. 262, 30 (2022), arXiv: 2112.12745.

[175] C. Evoli et al., "Cosmic-ray propagation with DRAGON2: I. numerical solver and astrophysical ingredients," JCAP 02, 015 (2017), arXiv: 1607.07886.

[176] C. Evoli et al., "Cosmic-ray propagation with DRAGON2: II. Nuclear interactions with the interstellar gas," JCAP 07, 006 (2018), arXiv: 1711.09616.

[177] O. Adriani et al., "PAMELA Measurements of Cosmic-ray Proton and Helium Spectra," Science 332, 69–72 (2011), arXiv: 1103.4055.

[178] B. Alves et al., "CRPropa 3.2 — an advanced framework for high-energy particle propagation in extragalactic and galactic spaces," JCAP 09, 035 (2022), arXiv: 2208.00107.

[179] V. S. Berezinskii, S. V. Bulanov, V. A. Dogiel, and V. S. Ptuskin, "Astrophysics of Cosmic Rays," North-Holland, 1990.

[180] V. L. Ginzburg, Y. M. Khazan, and V. S. Ptuskin, "Origin of cosmic rays: Galactic models with halo," Astrophys. Space Sci. 68, 295–314 (1980).

[181] V. S. Ptuskin, F. C. Jones, E. S. Seo, and R. Sina, "Effect of random nature of cosmic ray sources – Supernova remnants – on cosmic ray intensity fluctuations, anisotropy, and electron energy spectrum," Adv. Space Res. 37, 1909–1912 (2006).

[182] P. Blasi and E. Amato, "Diffusive propagation of cosmic rays from supernova remnants in the Galaxy. II: anisotropy," JCAP 01, 011 (2012), arXiv: 1105.4529.

[183] V. Ptuskin, "Propagation of galactic cosmic rays," Astropart. Phys. 39-40, 44–51 (2012).

[184] C. Evoli, D. Gaggero, D. Grasso, and L. Maccione, "A common solution to the cosmic ray anisotropy and gradient problems," Phys. Rev. Lett. 108, 211102 (2012), arXiv: 1203.0570.

[185] W. Baade and F. Zwicky, "Remarks on Super-Novae and Cosmic Rays," Phys. Rev. 46, 76–77 (1934).

[186] W. Baade and F. Zwicky, "Cosmic Rays from Super-Novae," Proc. Natl. Acad. Sci. U.S.A. 20, 259–263 (1934).

[187] L. Drury and A. W. Strong, "Power requirements for cosmic ray propagation models involving diffusive reacceleration; estimates and implications for the damping of interstellar turbulence," Astron. Astrophys. 597, A117 (2017), arXiv: 1608.04227.

[188] M. Kachelriess and D. V. Semikoz, "Cosmic Ray Models," Prog. Part. Nucl. Phys. 109, 103710 (2019), arXiv: 1904.08160.

[189] A. R. Bell and S. G. Lucek, "Cosmic ray acceleration to very high energy through the non-linear amplification by cosmic rays of the seed magnetic field," Mon. Not. Roy. Astron. Soc. 321, 433–438 (2001).

[190] A. R. Bell, "Turbulent amplification of magnetic field and diffusive shock acceleration of cosmic rays," Mon. Not. Roy. Astron. Soc. 353, 550–558 (2004).

[191] E. Parizot et al., "Superbubbles and energetic particles in the galaxy. 1. Collective effects of particle acceleration," Astron. Astrophys. 424, 747–760 (2004), arXiv: astro-ph/0405531.

[192] E. Parizot, "Superbubbles and the galactic evolution of li, be and b," Astron. Astrophys. 362, 786 (2000), arXiv: astro-ph/0006099.

[193] M. A. Duvernois, M. Garcia-Munoz, K. R. Pyle, J. A. Simpson, and M. R. Thayer, "The isotopic composition of Galactic cosmic-ray elements from carbon to silicon: The Combined Release and Radiation Effects Satellite investigation," Astrophys. J. 466, 457–472 (1996).

[194] Y. Ohira and K. Ioka, "Cosmic-ray helium hardening," Astrophys. J. Lett. 729, L13 (2011), arXiv: 1011.4405.

[195] Y. Ohira, N. Kawanaka, and K. Ioka, "Cosmic-ray hardenings in light of AMS-02 data," Phys. Rev. D 93, 083001 (2016), arXiv: 1506.01196.

[196] V. I. Zatsepin and N. V. Sokolskaya, "Three component model of cosmic ray spectra from 100 GeV up to 100 PeV," Astron. Astrophys. 458, 1–5 (2006), arXiv: astro-ph/0601475.

[197] P. L. Biermann et al., "Supernova explosions of massive stars and cosmic rays," Adv. Space Res. 62, 2773–2816 (2018), arXiv: 1803.10752.

[198] A. M. Soderberg et al., "A relativistic type Ibc supernova without a detected γ-ray burst," Nature 463, 513–515 (2010), arXiv: 0908.2817.

[199] P. L. Biermann and J. P. Cassinelli, "Cosmic rays. II. Evidence for a magnetic rotator Wolf Rayet star origin," Astron. Astrophys. 277, 691 (1993), arXiv: astro-ph/9305003.

[200] S. Thoudam et al., "Cosmic-ray energy spectrum and composition up to the ankle: the case for a second Galactic component," Astron. Astrophys. 595, A33 (2016), arXiv: 1605.03111.

[201] K. Murase and M. Fukugita, "Energetics of High-Energy Cosmic Radiations," Phys. Rev. D 99, 063012 (2019), arXiv: 1806.04194.

[202] E. Jaupart, E. Parizot, and D. Allard, "Contribution of the Galactic centre to the local cosmic-ray flux," Astron. Astrophys. 619, A12 (2018), arXiv: 1808.02322.

[203] A. Levinson and D. Eichler, "Baryon purity in cosmological gamma-ray bursts as a manifestation of event horizons," Astrophys. J. 418, 386–390 (1993).

[204] K. Kotera and A. V. Olinto, "The Astrophysics of Ultrahigh Energy Cosmic Rays,"

Ann. Rev. Astron. Astrophys. 49, 119–153 (2011), arXiv: 1101.4256.

[205] D. Eichler, N. Globus, R. Kumar, and E. Gavish, "Ultrahigh Energy Cosmic Rays: a Galactic Origin?," Astrophys. J. Lett. 821, L24 (2016), arXiv: 1604.05721.

[206] M. Vietri, "On the acceleration of ultrahigh-energy cosmic rays in gamma-ray bursts," Astrophys. J. 453, 883–889 (1995), arXiv: astro-ph/9506081.

[207] E. Waxman, "Cosmological gamma-ray bursts and the highest energy cosmic rays," Phys. Rev. Lett. 75, 386–389 (1995), arXiv: astro-ph/9505082.

[208] F. Samuelsson, D. Bégué, F. Ryde, and A. Pe'er, "The Limited Contribution of Low- and High-Luminosity Gamma-Ray Bursts to Ultra-High Energy Cosmic Rays," Astrophys. J. 876, 93 (2019), arXiv: 1810.06579.

[209] K. Murase, K. Ioka, S. Nagataki, and T. Nakamura, "High Energy Neutrinos and Cosmic-Rays from Low-Luminosity Gamma-Ray Bursts?," Astrophys. J. Lett. 651, L5–L8 (2006), arXiv: astro-ph/0607104.

[210] R. Y. Liu, X. Wang, and Z. G. Dai, "Nearby low-luminosity GRBs as the sources of ultra-high energy cosmic rays revisited," Mon. Not. Roy. Astron. Soc. 418, 1382 (2011), arXiv: 1108.1551.

[211] K. Murase, K. Ioka, S. Nagataki, and T. Nakamura, "High-energy cosmic-ray nuclei from high- and low-luminosity gamma-ray bursts and implications for multi-messenger astronomy," Phys. Rev. D 78, 023005 (2008), arXiv: 0801.2861.

[212] X. Wang, S. Razzaque, and P. Meszaros, "On the Origin and Survival of Ultra-High-Energy Cosmic-Ray Nuclei in GRBs and Hypernovae," Astrophys. J. 677, 432–440 (2008), arXiv: 0711.2065.

[213] S. Horiuchi, K. Murase, K. Ioka, and P. Mészáros, "The survival of nuclei in jets associated with core-collapse supernovae and gamma-ray bursts," Astrophys. J. 753, 69 (2012), arXiv: 1203.0296.

[214] P. Baerwald, M. Bustamante, and W. Winter, "Are gamma-ray bursts the sources of ultra-high energy cosmic rays?," Astropart. Phys. 62, 66–91 (2015), arXiv: 1401.1820.

[215] N. Globus, D. Allard, R. Mochkovitch, and E. Parizot, "UHECR acceleration at GRB internal shocks," Mon. Not. R. Astron. Soc. 451, 751–790 (2015), arXiv: 1409.1271.

[216] D. Boncioli, D. Biehl, and W. Winter, "On the common origin of cosmic rays across the ankle and diffuse neutrinos at the highest energies from low-luminosity gamma-ray bursts," Astrophys. J. 872, 110 (2019), arXiv: 1808.07481.

[217] B. T. Zhang, K. Murase, S. S. Kimua, S. Horiuchi, and P. Mészáros, "Low-luminosity gamma-ray bursts as the sources of ultrahigh-energy cosmic-ray nuclei," Phys. Rev. D 97, 083010 (2018), arXiv: 1712.09984.

[218] X. Y. Wang, S. Razzaque, P. Mészáros, and Z. G. Dai, "High-energy cosmic rays and neutrinos from semi-relativistic hypernovae," Phys. Rev. D 76, 083009 (2007), arXiv: 0705.0027.

[219] S. Chakraborty, A. Ray, A. M. Soderberg, A. Loeb, and P. Chandra, "Ultra-high-energy cosmic-ray acceleration in engine-driven relativistic supernovae," Nat. Commun. 2, 175 (2011), arXiv: 1012.0850.

[220] R. Y. Liu and X. Y. Wang, "Energy spectrum and chemical composition of ultrahigh-energy cosmic rays from semi-relativistic hypernovae," Astrophys. J. 746, 40 (2012), arXiv: 1111.6256.

[221] K. Fang, B. D. Metzger, K. Murase, I. Bartos, and K. Kotera, "Multimessenger implications of AT2018cow: High-energy cosmic-ray and neutrino emissions from magnetar-powered superluminous transients," Astrophys. J. 878, 34 (2019), arXiv: 1812.11673.

[222] B. P. Abbott et al., "Gravitational waves and gamma-rays from a binary neutron star merger: GW170817 and GRB 170817A," Astrophys. J. Lett. 848, L13 (2017), arXiv: 1710.05834.

[223] B. P. Abbott et al., "GW170817: Observation of gravitational waves from a binary neutron star inspiral," Phys. Rev. Lett. 119, 161101 (2017), arXiv: 1710.05832.

[224] X. Rodrigues, D. Biehl, D. Boncioli, and A. M. Taylor, "Binary neutron star merger remnants as sources of cosmic rays below the 'ankle'," Astropart. Phys. 106, 10–17 (2019), arXiv: 1806.01624.

[225] S. S. Kimura, K. Murase, and P. Mészáros, "Super-knee cosmic rays from Galactic neutron star merger remnants," Astrophys. J. 866, 51 (2018), arXiv: 1807.03290.

[226] J. Abraham et al., "Correlation of the highest-energy cosmic rays with the positions of nearby active galactic nuclei," Astropart. Phys. 29, 188–204 (2008), arXiv: 0712.2843.

[227] C. D. Dermer and S. Razzaque, "Acceleration of ultrahigh-energy cosmic rays in the colliding shells of blazars and gamma-ray bursts: Constraints from the Fermi Gamma-ray Space Telescope," Astrophys. J. 724, 1366–1372 (2010), arXiv: 1004.4249.

[228] E. Resconi, S. Coenders, P. Padovani, P. Giommi, and L. Caccianiga, "Connecting blazars with ultrahigh-energy cosmic rays and astrophysical neutrinos," Mon. Not. R. Astron. Soc. 468, 597–606 (2017), arXiv: 1611.06022.

[229] X. Rodrigues, A. Fedynitch, S. Gao, D. Boncioli, and W. Winter, "Neutrinos and ultrahigh-energy cosmic-ray nuclei from blazars," Astrophys. J. 854, 54 (2018), arXiv: 1711.02091.

[230] J. P. Rachen and P. L. Biermann, "Extragalactic ultrahigh-energy cosmic rays, I. Contribution from hot spots in FR-II radio galaxies," Astron. Astrophys. 272, 161–175 (1993), arXiv: astro-ph/9301010.

[231] G. E. Romero, J. A. Combi, L. A. Anchordoqui, and S. E. P. Bergliaffa, "Centaurus A as a source of extragalactic cosmic rays with arrival energies well beyond the GZK cutoff," Astropart. Phys. 5, 279–283 (1996), arXiv: gr-qc/9511031.

[232] D. Caprioli, ""Espresso" acceleration of ultrahigh-energy cosmic rays," Astrophys. J. Lett. 811, L38 (2015), arXiv: 1505.06739.

[233] S. S. Kimura, K. Murase, and B. T. Zhang, "Ultrahigh-energy cosmic-ray nuclei from black hole jets: Recycling Galactic cosmic rays through shear acceleration," Phys. Rev. D 97, 023026 (2018), arXiv: 1705.05027.

[234] F. Takahara, "On the origin of highest-energy cosmic rays," Prog. Theor. Phys. 83, 1071–1075 (1990).

[235] J. H. Matthews, A. R. Bell, K. M. Blundell, and A. T. Araudo, "Ultrahigh-energy cosmic rays from shocks in the lobes of powerful radio galaxies," Mon. Not. R. Astron. Soc. 482, 4303–4321 (2019), arXiv: 1810.12350.

[236] A. Pe'er, K. Murase, and P. Mészáros, "Radio-quiet AGNs as possible sources of ultrahigh-energy cosmic rays," Phys. Rev. D 80, 123018 (2009), arXiv: 0911.1776.

[237] X. Wang and A. Loeb, "Ultrahigh-energy cosmic rays from nonrelativistic quasar outflows," Phys. Rev. D 95, 063007 (2017), arXiv: 1611.07616.

[238] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva, "On astrophysical solution to ultrahigh-energy cosmic rays," Phys. Rev. D 74, 043005 (2006), arXiv: hep-ph/0204357.

[239] E. Fermi, "On the Origin of the Cosmic Radiation," Phys. Rev. 75, 1169–1174 (1949).

[240] W. I. Axford, E. Leer, and G. Skadron, "The Acceleration of Cosmic Rays by Shock

Waves," ICRC1977 11, 132 (1977).

[241] A. R. Bell, "The acceleration of cosmic rays in shock fronts – I." Mon. Not. R. Astron. Soc. 182, 147–156 (1978).

[242] L. O. Drury, "An introduction to the theory of diffusive shock acceleration of energetic particles in tenuous plasmas," Rep. Prog. Phys. 46, 973–1027 (1983).

[243] M. S. Longair, "High Energy Astrophysics," Cambridge University Press, 2011.

[244] L. D. Landau and E. M. Lifshitz, "Fluid Mechanics," Pergamon Press, 1982.

[245] A. M. Hillas, "The Origin of Ultrahigh-Energy Cosmic Rays," Ann. Rev. Astron. Astrophys. 22, 425–444 (1984).

[246] P. O. Lagage and C. J. Cesarsky, "The maximum energy of cosmic rays accelerated by supernova shocks," Astron. Astrophys. 125, 249–257 (1983).

[247] P. O. Lagage and C. J. Cesarsky, "Cosmic-ray shock acceleration in the presence of self-excited waves," Astron. Astrophys. 118, 223–228 (1983).

[248] G. Morlino and D. Caprioli, "Strong evidence of hadron acceleration in Tycho's supernova remnant," Astron. Astrophys. 538, A81 (2012), arXiv: 1105.6342.

[249] J. G. Kirk, A. W. Guthmann, Y. A. Gallant, and A. Achterberg, "Particle acceleration at ultrarelativistic shocks: An eigenfunction method," Astrophys. J. 542, 235–242 (2000), arXiv: astro-ph/0005222.

[250] A. Achterberg, Y. A. Gallant, J. G. Kirk, and A. W. Guthmann, "Particle acceleration by ultrarelativistic shocks: Theory and simulations," Mon. Not. R. Astron. Soc. 328, 393–408 (2001), arXiv: astro-ph/0107530.

[251] U. Keshet and E. Waxman, "The spectrum of particles accelerated in relativistic, collisionless shocks," Phys. Rev. Lett. 94, 111102 (2005), arXiv: astro-ph/0408489.

[252] S. Zenitani and M. Hoshino, "Particle Acceleration and Magnetic Dissipation in Relativistic Current Sheet of Pair Plasmas," Astrophys. J. 670, 702–726 (2007), arXiv: 0708.1000.

[253] D. Giannios, "UHECRs from magnetic reconnection in relativistic jets," Mon. Not. R. Astron. Soc. Lett. 408, L46–L50 (2010), arXiv: 1007.1522.

[254] P. Goldreich and W. H. Julian, "Pulsar electrodynamics," Astrophys. J. 157, 869 (1969).

[255] M. A. Ruderman and P. G. Sutherland, "Theory of pulsars: Polar caps, sparks, and coherent microwave radiation," Astrophys. J. 196, 51–72 (1975).

[256] J. Arons and E. T. Scharlemann, "Pair formation above pulsar polar caps: Structure of the low altitude acceleration zone," Astrophys. J. 231, 854–879 (1979).

[257] A. K. Harding and A. G. Muslimov, "Particle acceleration zones above pulsar polar caps: Electron and positron pair formation fronts," Astrophys. J. 508, 328–346 (1998), arXiv: astro-ph/9805132.

[258] C. Guépin, B. Cerutti, and K. Kotera, "Proton acceleration in pulsar magnetospheres," Astron. Astrophys. 635, A138 (2020), arXiv: 1910.11387.

[259] R. Buckley, "Acceleration of charged particles to extremely high energies in pulsar far fields," Nature 266, 37–38 (1977).

[260] A. Aab et al., "Testing Hadronic Interactions at Ultrahigh Energies with Air Showers Measured by the Pierre Auger Observatory," Phys. Rev. Lett. 117, 192001 (2016), arXiv: 1610.08509.

[261] S. L. Shapiro and S. A. Teukolsky, "Black Holes, White Dwarfs, and Neutron Stars: The Physics of Compact Objects," Wiley-Interscience, 1983.

[262] P. Blasi, R. I. Epstein, and A. V. Olinto, "Ultrahigh-energy cosmic rays from young neutron star winds," Astrophys. J. Lett. 533, L123–L126 (2000), arXiv: astro-ph/9912240.

[263] J. Arons, "Magnetars in the metagalaxy: An origin for ultrahigh-energy cosmic rays in the nearby universe," Astrophys. J. 589, 871–892 (2003), arXiv: astro-ph/0208444.

[264] T. Tajima and J. M. Dawson, "Laser electron accelerator," Phys. Rev. Lett. 43, 267–270 (1979).

[265] E. Gschwendtner et al., "AWAKE: The Advanced Proton Driven Plasma Wakefield Acceleration Experiment at CERN," Nucl. Instrum. Meth. A 829, 76–82 (2016), arXiv: 1512.05498.

[266] P. Chen, T. Tajima, and Y. Takahashi, "Plasma wakefield acceleration for ultrahigh-energy cosmic rays," Phys. Rev. Lett. 89, 161101 (2002), arXiv: astro-ph/0205287.

[267] T. Ebisuzaki and T. Tajima, "Astrophysical ZeV acceleration in the relativistic jet from an accreting supermassive black hole," Astropart. Phys. 56, 9–15 (2014), arXiv: 1306.0970.

[268] A. Chiavassa et al., "A study of the first harmonic of the large scale anisotropies with the KASCADE-Grande experiment," PoS ICRC2015, 281 (2015).

[269] M. Amenomori et al., "Large-scale sidereal anisotropy of Galactic cosmic-ray intensity observed by the Tibet air shower array," Astrophys. J. Lett. 83, 29–32 (2005), arXiv: astro-ph/0505114.

[270] A. Aab et al., "Cosmic-ray anisotropies in right ascension measured by the Pierre Auger Observatory," Astrophys. J. 891, 142 (2020), arXiv: 2002.06172.

[271] R. U. Abbasi et al., "Analysis of Large-Scale Anisotropy of Ultra-High Energy Cosmic Rays in HiRes Stereo Data," Astrophys. J. Lett. 713, L64–L68 (2010), arXiv: 1002.1444.

[272] J. Meeus, "Astronomical Algorithms," Willmann-Bell, 1998.

[273] P. Duffett-Smith and J. Zwart, "Practical Astronomy with your Calculator or Spreadsheet," Cambridge University Press, 2011.

[274] P. Sommers, "Cosmic ray anisotropy analysis with a full-sky observatory," Astropart. Phys. 14, 271–286 (2001), arXiv: astro-ph/0004016.

[275] S. E. Urban and P. K. Seidelmann, "Explanatory Supplement to the Astronomical Almanac," University Science Books, 2013.

[276] T. P. Li and Y. Q. Ma, "Analysis methods for results in gamma-ray astronomy," Astrophys. J. 272, 317–324 (1983).

[277] F. Aharonian et al., "Observations of the Crab Nebula with H.E.S.S," Astron. Astrophys. 457, 899–915 (2006), arXiv: astro-ph/0607333.

[278] A. Aab et al., "Large Scale Distribution of Ultra High Energy Cosmic Rays Detected at the Pierre Auger Observatory With Zenith Angles up to 80°," Astrophys. J. 802, 111 (2015), arXiv: 1411.6953.

[279] P. Abreu et al., "Arrival Directions of Cosmic Rays above 32 EeV from Phase One of the Pierre Auger Observatory," Astrophys. J. 935, 170 (2022), arXiv: 2206.13492.

[280] A. A. Abdo et al., "Discovery of Localized Regions of Excess 10 TeV Cosmic Rays," Phys. Rev. Lett. 101, 221101 (2008), arXiv: 0801.3827.

[281] K. M. Górski et al., "HEALPix - A Framework for high resolution discretization, and fast analysis of data distributed on the sphere," Astrophys. J. 622, 759–771 (2005), arXiv: astro-ph/0409513.

[282] B. Bartoli et al., "Medium scale anisotropy in the TeV cosmic ray flux observed by

ARGO-YBJ," Phys. Rev. D 88, 082001 (2013), arXiv: 1309.6182.

[283] P. Abreu et al., "Constraints on the Origin of Cosmic Rays above 10^{18} eV from Large Scale Anisotropy Searches in Data of the Pierre Auger Observatory," Astrophys. J. Lett. 762, L13 (2012), arXiv: 1212.3083.

[284] E. Hivon et al., "Master of the cosmic microwave background anisotropy power spectrum: a fast method for statistical analysis of large and complex cosmic microwave background data sets," Astrophys. J. 567, 2 (2002), arXiv: astro-ph/0105302.

[285] G. F. Smoot and P. M. Lubin, "Southern Hemisphere Measurements of the Anisotropy in the Cosmic Microwave Background Radiation," Astrophys. J. Lett. 234, L83 (1979).

[286] R. Abbasi et al., "Observation of Anisotropy in the Arrival Directions of Galactic Cosmic Rays at Multiple Angular Scales with IceCube," Astrophys. J. 740, 16 (2011), arXiv: 1105.2326.

[287] R. Bonino et al., "The East-West method: an exposure-independent method to search for large scale anisotropies of cosmic rays," Astrophys. J. 738, 67 (2011), arXiv: 1106.2651.

[288] M. Ahlers et al., "Large- and Medium-scale Anisotropies in the Arrival Directions of Cosmic Rays Observed with KASCADE-Grande," Astrophys. J. Lett. 886, L18 (2019), arXiv: 1909.09222.

[289] R. Abbasi et al., "Simulation study for the future IceCube-Gen2 surface array," PoS ICRC2021, 411 (2021), arXiv: 2108.04307.

[290] D. Heck et al., "CORSIKA: A Monte Carlo code to simulate extensive air showers," Report FZKA (1998).

[291] T. Huege, M. Ludwig, and C. W. James, "Simulating radio emission from air showers with CoREAS," AIP Conf. Proc. 1535, 128–132 (2013), arXiv: 1301.2132.

[292] T. T. Böhlen et al., "The FLUKA Code: Developments and Challenges for High Energy and Medical Applications," Nucl. Data Sheets 120, 211–214 (2014).

[293] F. Riehn et al., "Hadronic interaction model SIBYLL 2.3d and extensive air showers," Phys. Rev. D 102, 063002 (2020), arXiv: 1912.03300.

[294] T. K. Gaisser, T. Stanev, and S. Tilav, "Cosmic Ray Energy Spectrum from Measurements of Air Showers," Front. Phys. (Beijing) 8, 748–758 (2013), arXiv: 1303.3565.

[295] S. Agostinelli et al., "GEANT4: A Simulation toolkit," Nucl. Instrum. Meth. A506, 250–303 (2003).

[296] R. Abbasi et al., "Density of GeV Muons Measured with IceTop," PoS ICRC2021, 342 (2021), arXiv: 2107.09583.

[297] T. K. Gaisser, "Spectrum of cosmic-ray nucleons, kaon production, and the atmospheric muon charge ratio," Astropart. Phys. 35, 801–806 (2012), arXiv: 1111.6675.

[298] M. Aglietta et al., "Evolution of the cosmic ray anisotropy above 10^{14} eV," Astrophys. J. Lett. 692, 130–133 (2009), arXiv: 0901.2740.

[299] M. Ambrosio et al., "The Search for the sidereal and solar diurnal modulations in the total MACRO muon data set," Phys. Rev. D 67, 042002 (2003), arXiv: astro-ph/0211119.

[300] W. Zhang et al., "Evolution of cosmic ray anisotropy with energy up to PeV observed by LHAASO-KM2A," PoS ICRC2025, 264 (2025).

[301] B. Hariharan et al., "Investigating the influence of cosmic rays on atmospheric electric fields and tropical temperature anomalies using GRAPES-3," PoS ICRC2025, 180

(2025).

[302] W. Hou, "Sensitivity Study of IceCube-Gen2 Surface Array for Cosmic-Ray Anisotropy," PoS TAUP2023, 146 (2024).

[303] G. Casella and R. L. Berger, "Statistical Inference," Duxbury, 2002.

[304] T. Hastie, R. Tibshirani, and J. Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction," Springer, 2009.

[305] L. Wasserman, "All of Statistics: A Concise Course in Statistical Inference," Springer, 2004.

[306] J. A. Rice, "Mathematical Statistics and Data Analysis," Duxbury, 2006.

[307] E. L. Lehmann and G. Casella, "Theory of Point Estimation," Springer, 1998.

[308] B. R. Frieden, "Science from Fisher Information : A Unification," Cambridge University Press, 2004.

[309] M. Ahlers and P. Mertsch, "Origin of small-scale anisotropies in Galactic cosmic rays," Prog. Part. Nucl. Phys. 94, 184–216 (2017), arXiv: 1612.01873.

[310] M. Amenomori et al., "Northern Sky Galactic Cosmic Ray Anisotropy between 10 and 1000 TeV with the Tibet Air Shower Array," Astrophys. J. 836 (2017), arXiv: 1701.07144.

[311] P. Abreu et al., "The energy spectrum of cosmic rays beyond the turn-down around 10^{17} eV as measured with the surface detector of the Pierre Auger Observatory," Eur. Phys. J. C 81, 966 (2021), arXiv: 2109.13400.

[312] A. Aab et al., "Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory," Phys. Rev. D 102, 062005 (2020), arXiv: 2008.06486.

[313] B. Polyak, "Some methods of speeding up the convergence of iteration methods," Comput. Math. Math. Phys. 4, 1–17 (1964).

[314] W. D. Apel et al., "Probing the evolution of the EAS muon content in the atmosphere with KASCADE-Grande," Astropart. Phys. 95, 25–43 (2017), arXiv: 1801.05513.

[315] D. Kang et al., "All-particle energy spectrum of KASCADE-Grande based on shower size and different hadronic interaction models," J. Phys. Conf. Ser. 409, 012101 (2013), arXiv: 1309.4295.

[316] S. Ostapchenko, "QGSJET-II: Towards reliable description of very high energy hadronic interactions," Nucl. Phys. B Proc. Suppl. 151, 143–146 (2006), arXiv: hep-ph/0412332.

[317] W. D. Apel et al., "The KASCADE-Grande energy spectrum of cosmic rays and the role of hadronic interaction models," Adv. Space Res. 53, 1456–1469 (2014).

[318] F. J. M. Farley and J. R. Storey, "The Sidereal Correlation of Extensive Air Showers," Proc. Phys. Soc. 67, 996–1004 (1954).

[319] R. Abbasi et al., "Observation of an Anisotropy in the Galactic Cosmic Ray arrival direction at 400 TeV with IceCube," Astrophys. J. 746, 33 (2012), arXiv: 1109.1017.

[320] P. Abreu et al., "Search for First Harmonic Modulation in the Right Ascension Distribution of Cosmic Rays Detected at the Pierre Auger Observatory," Astropart. Phys. 34, 627–639 (2011), arXiv: 1103.2721.

[321] J. Aublin and E. Parizot, "Generalized 3D-reconstruction method of a dipole anisotropy in cosmic-ray distributions," Astron. Astrophys. 441, 407 (2005), arXiv: astro-ph/0504575.

[322] H. Steck and T. Jaakkola, "On the Dirichlet Prior and Bayesian Regularization," Adv. Neural Inf. Process. Syst., 713–720 (2002).

[323] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, "Bayesian Data Analysis," CRC Press, 2013.

[324] R. Herbert and M. Sutton, "A Stochastic Approximation Method," Ann. Math. Statist. 22, 400–407 (1951).

[325] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," CoRR (2014), arXiv: 1412.6980.

[326] S. Ruder, "An overview of gradient descent optimization algorithms," (2016), arXiv: 1609.04747.

[327] E. L. Lehmann and J. P. Romano, "Testing Statistical Hypotheses," Springer Cham, 2022.

[328] M. Ahlers, "Searching for All-Scale Anisotropies in the Arrival Directions of Cosmic Rays above the Ankle," Astrophys. J. 863, 146 (2018), arXiv: 1805.08220.

[329] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes: The Art of Scientific Computing," Cambridge University Press, 2007.

[330] I. J. Schoenberg, "Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions," Quart. Appl. Math. 4, 112–141 (1946).

[331] S. Chandrasekhar, "Stochastic Problems in Physics and Astronomy," Reviews of Modern Physics 15, 1–89 (1943).

[332] S. Menchiari et al., "Cygnus OB2 as a test case for particle acceleration in young massive star clusters," Astron. Astrophys. 686, A242 (2024), arXiv: 2402.07784.

[333] R. Jansson and G. R. Farrar, "A New Model of the Galactic Magnetic Field," Astrophys. J. 757, 14 (2012), arXiv: 1204.3662.

[334] M. Unger and G. R. Farrar, "The Coherent Magnetic Field of the Milky Way," Astrophys. J. 970, 95 (2024), arXiv: 2311.12120.

[335] R. Jansson and G. R. Farrar, "The Galactic Magnetic Field," Astrophys. J. Lett. 761, L11 (2012), arXiv: 1210.7820.

[336] G. R. Farrar, "The Galactic Magnetic Field and Ultrahigh-Energy Cosmic Ray Deflections," C. R. Phys. 15, 339–348 (2014), arXiv: 1405.3680.

[337] J. Giacalone and J. R. Jokipii, "The transport of cosmic rays across a turbulent magnetic field," Astrophys. J. 520, 204–214 (1999).

[338] M. Haverkorn, "Magnetic fields in the Milky Way," (2014), arXiv: 1406.0283.

[339] R. Beck, "Magnetic fields in spiral galaxies," Astron. Astrophys. Rev. 24, 4 (2015), arXiv: 1509.04522.

[340] A. N. Kolmogorov, "The local structure of turbulence in incompressible viscous fluid for very large Reynolds numbers," Proc. R. Soc. Lond. A 434, 9–13 (1991).

[341] M. Opher et al., "A strong, highly tilted interstellar magnetic field near the Solar System," Nature 462, 1036–1038 (2009).

[342] G. Giacinti et al., "Reconciling cosmic ray diffusion with Galactic magnetic field models," JCAP 7, 051 (2018), arXiv: 1710.08205.

[343] G. Bernard et al., "TeV cosmic-ray proton and helium spectra in the myriad model," Astron. Astrophys. 555, A48 (2013), arXiv: 1207.4670.

[344] F. Aharonian et al., "Massive Stars as Major Factories of Galactic Cosmic Rays," Nature Astron. 3, 561–567 (2019), arXiv: 1804.02331.

[345] A. P. Snodin et al., "Global diffusion of cosmic rays in random magnetic fields," Mon. Not. Roy. Astron. Soc. 457, 3975–3987 (2016), arXiv: 1509.03766.

[346] M. Ahlers, "Deciphering the Dipole Anisotropy of Galactic Cosmic Rays," Phys.

Rev. Lett. 117, 151103 (2016), arXiv: 1605.06446.

[347] F. A. Haight, "Handbook of the Poisson Distribution," J. Oper. Res. Soc. 18, 478–479 (1967).

[348] S. S. Wilks, "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," Annals Math. Statist. 9, 60–62 (1938).

[349] T. Ebisuzaki and T. Tajima, "Astrophysical Wake Acceleration Driven by Relativistic Alfvenic Pulse Emitted from Bursting Accretion Disk," Astropart. Phys. 128, 102567 (2021), arXiv: 1905.04506.

Acknowledgments

I would like to first thank my main supervisor, Dr. Andreas Haungs, and also my first supervisor, Prof. Ralph Engel, who accepted me as a PhD student at the Institute for Astroparticle Physics (IAP). It was a challenging time for me to transition from the Institute for Theoretical Physics (ITP) to IAP. Dr. Haungs has always been exceptionally kind and supportive, consistently providing detailed and constructive feedback on my work, even when I knew very little about the research topic at the beginning. His insights were always valuable and often aligned perfectly with my own research interests. Prof. Engel has also been very patient and thoughtful in offering his suggestions and guidance. From him, I learned a great deal—not only about scientific research but also about maintaining high standards of rigor and integrity in academic work.

My sincere appreciation goes to Dr. Donghwa Kang, who helped me greatly at the beginning of my work. Through our many discussions, I was able to make step-by-step progress in my research. Even though she was sometimes not very familiar with every detail of the project, she was always patient and willing to discuss them with me. She is also very kind and friendly; her care and guidance have felt like those of an older sister. I am also grateful to Dr. Frank Schröder for his support both before and after the start of my PhD. He guided me in understanding the fundamental aspects of the experiments and helped me integrate into the research group. I also want to thank my second supervisor, Prof. Thomas Schwetz-Mangold, for serving as my thesis reviewer. He is a very kind and approachable professor who is always smiling and offers valuable suggestions that have been very helpful to my work.

Many thanks go to Paras Koundal and Federico Bontempo, who helped me greatly with coding issues. I would also like to thank Mark Weyrauch and Julian Saffer for their assistance with code and for sharing valuable technical knowledge. My sincere thanks also go to Fahim Varsi, Shefali Shefali, Tista Mukherjee, and Megha Venugopal. I truly appreciate their patience whenever I asked for help, even when I disturbed them at inconvenient times. I am also grateful to Sabine Bucher for her kind help with coffee in the institute and for preparing the documents needed for conference participation. It was also my good fortune to share an office with Victoria Tokareva; her cheerfulness and positivity brought so much warmth and vitality to the workplace.

Most importantly, I would like to express my deepest gratitude to my family for their unwavering support and encouragement in allowing me to come to Germany to pursue my PhD. I would also like to thank Chao Zhang, who was always very friendly during his time at the institute. He shared many joyful moments with me, especially during our trips around Germany, and has continued to offer help and encouragement to this day. Many thanks also go to my roommate, Yunfei Miao, who was exceptionally kind and caring, especially when I was sick. We shared many moments of daily life in the same house during the pandemic, supporting each other through those difficult times. I am also deeply grateful to my landlords, Oma (Renata) and Opa (Otto), who were incredibly kind and always treated me like a member of their family, making me feel truly at home. Renata often cared for me during the pandemic, told me stories, and warmly invited me to dinners and celebrations on important holidays. Lastly, I would like to thank ChatGPT for assisting me in checking and improving the grammar and clarity of my thesis writing.

Declaration

Karlsruhe, 06.11.2025

Declaration of Independent Authorship of My Dissertation

I hereby declare that I have written the dissertation entitled

Research on the High-Energy Cosmic-Ray Anisotropies and Their Origins

independently and without any unauthorized external assistance. I have used only the sources and aids listed by me.

_____

Wenjie Hou

Photo credit: NASA

Far beyond cosmic rays, the universe is so deep and filled with countless mysteries. Humanity's desire and effort to uncover the truths of the cosmos are ongoing and unceasing.