MASS-DEPENDENT ANISOTROPY OF ULTRA-HIGH-ENERGY COSMIC RAYS

For the attainment of the academic degree of

Doctor of Science

from the
Department of Physics of the
Karlsruhe Institute of Technology (KIT)
and the
Institute of Technology "Prof. Jorge A. Sabato" of the
National University of General San Martin (UNSAM)

accepted

Dissertation

of

M. Sc. Edyvania Emily Pereira Martins

from Campina Grande

Day of the oral examination: 20. 12. 2024
Referee: Prof. Dr. Ralph Engel
Co-referee: Dr. Silvia Mollerach
Supervisors: Dr. Darko Veberič, Dr. Silvia Mollerach , Dr. Markus Roth

MASS-DEPENDENT ANISOTROPY OF ULTRA-HIGH-ENERGY COSMIC RAYS

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)
und des
Institut für Technologie ″Prof. Jorge A. Sábato″ der
Nationale Universität von San Martin (UNSAM)

genehmigte

Dissertation

von

M. SC. EDYVANIA EMILY PEREIRA MARTINS

aus Campina Grande

Tag der mündlichen Prüfung: 20. 12. 2024
Referent: Prof. Dr. Ralph Engel
Korreferentin: Dr. Silvia Mollerach
Betreuer: Dr. Darko Veberič, Dr. Silvia Mollerach , Dr. Markus Roth

MASS-DEPENDENT ANISOTROPY OF ULTRA-HIGH-ENERGY COSMIC RAYS

Tesis aceptada para optar por el título de

Doctora en Astrofísica

del Departamento de Fisica del
Instituto de Tecnología de Karlsruhe (KIT)
y del
Instituto de Tecnología ”Prof. Jorge A. Sábato” de la
Universidad Nacional de San Martín (UNSAM)

por

M. SC. EDYVANIA EMILY PEREIRA MARTINS
de Campina Grande

Fecha de la defensa oral: 20. 12. 2024
Director: Prof. Dr. Ralph Engel
Co-director: Dr. Silvia Mollerach
Colaboradores: Dr. Darko Veberič, Dr. Silvia Mollerach , Dr. Markus Roth

Abstract

The Pierre Auger Observatory has been collecting data for over 19 years, reaching more than 123 000 km^2 sr yr of accumulated exposure, with the surface detectors spread over 3000 km^2. The most significant discovery to this date is a dipole structure on the arrival directions of cosmic rays, with a total amplitude of approximately 7%. This results from the observed modulation in right ascension in the inclusive energy bin above 8 EeV, where the computed dipole equatorial component has a statistical significance of 6.8σ. A similar dipole pattern in the events with energies between 8 and 16 EeV has recently been observed with a statistical significance of over 5σ. The Pierre Auger Collaboration has also reported an increase in the dipole amplitude with energy. This anisotropy is understood to be of extra-galactic origin, as the maximum of the dipole points to a direction ~115° away from the Galactic center.

In the same energy range, the observed evolution of the depth of maximum shower development with energy indicates a progression towards heavier composition of cosmic rays with increasing energy.

In this thesis I present a novel approach to a search for composition-enhanced large-scale anisotropy. On the one hand, lighter events have higher rigidity than their heavier counterparts; therefore, their trajectories are less affected by the Galactic and extra-galactic magnetic fields. The expected effect is a higher anisotropy in the arrival direction of a subset of events with smaller mass and charge than the anisotropy in the overall flux of cosmic rays. On the other hand, the attenuation length is distinct for each mass group, leading to different horizons of cosmic ray sources for light and heavy particles at a given energy. Under a source-agnostic model, I investigate the dipole amplitude as a function of rigidity. Using a simulation library, I analyze the possibility of measuring a separation in total dipole amplitude between two A-distinct sub-populations of the Auger Phase I dataset.

After scrutinizing the proposed method, the Pierre Auger Collaboration has approved this analysis to *unblind* the mass information in a search for large-scale anisotropy. This novel approach indicates that a mass estimator can be used in searches for anisotropies already in Phase I of the Observatory data-taking. We have observed a separation in the dipole amplitudes of two A-distinct subsets of the data, which indicates a dependency of the measured dipole amplitude on the mass/charge. The most significant separation occurs in the 8 to 16 EeV energy range. There the computed chance probability is 2.9×10^{-5}, corresponding to a two-sided significance of 4.18σ. The additional information can help understand the potential sources of the ultra-high-energy cosmic rays, as a tighter constraint of the plausible source scenarios is needed to describe the measurements presented here.

This and future mass-informed anisotropy searches will be potentially improved in the context of AugerPrime, the upgraded stage of the Pierre Auger Observatory. In its second data-taking phase, additional detection techniques will improve the mass-sensitivity of the Observatory.

Zusammenfassung

Das Pierre Auger Observatory sammelt seit über 19 Jahren Daten und hat seitdem mehr als $123\,000\,\mathrm{km}^2$ sr yr *exposure* mit seinem Oberflächendetektoren, welcher $3000\,\mathrm{km}^2$ groß ist, erreicht. Die signifikanteste Entdeckung bis zum heutigen Tag ist eine Dipolstruktur in den Ankunftsrichtungen der kosmischen Strahlung, welche eine Gesamtamplitude von etwa 7% aufweist. Die Amplitude kommt von der beobachteten Modulation der Rektaszension in dem Energiebereich oberhalb von 8 EeV, ab welcher berechnete Dipol-Äquatorialkomponente eine statistische Signifikanz von 6.8σ aufweist. Ein ähnliches Dipolmuster in den Ereignissen mit Energien zwischen 8 und 16 EeV wurde kürzlich mit einer statistischen Signifikanz von über 5σ beobachtet. Die Pierre Auger Collaboration hat ebenfalls eine Zunahme der Dipolamplitude mit der Energie gefunden. Es wird davon ausgegangen, dass diese Anisotropie extra-galaktischen Ursprungs ist, da das Maximum des Dipols $\sim$115° vom galaktischen Zentrum zeigt.

Im gleichen Energiebereich deutet die beobachtete Entwicklung die Dicke der darueberliegenden Luftsehieh der maximalen Schauerentwicklung mit der Energie auf eine Zusammensetzung der kosmischen Strahlung mit zunehmender Energie.

In dieser Arbeit stelle ich einen neuen Ansatz für die Suche nach einer durch die Zusammensetzung *mass-enhanced* Anisotropie vor. Einerseits haben leichtere Kosmische Strahlen eine höhere magnetische Steifigkeit als schwere kosmische Strahlung; daher werden ihre Trajektorien weniger von den galaktischen und extra-galaktischen Magnetfeldern beeinflusst. Es wird erwartet, dass eine Population von leichteren kosmischen Strahlen eine höhere Anisotropie aufweist, als eine Population von schweren kosmischen Strahlen. Andererseits ist die Abschwächungslänge für jede Massengruppe unterschiedlich, was zu unterschiedlichen Maximaldistanzen zu Quellen der kosmischen Strahlung für leichte und schwere Teilchen bei einer bestimmten Energie führt. Im Rahmen eines quellenunabhängigen Modells untersuche ich die Dipolamplitude als eine Funktion der magnetischen Steifigkeit. Mit Hilfe einer Simulationsbibliothek analysiere ich die Möglichkeit, eine Trennung in der Gesamtdipolamplitude zwischen zwei A-unterschiedlichen Unterpopulationen des Auger Phase I Datensatzes zu messen.

Nach Prüfung der vorgeschlagenen Methode hat die Pierre Auger Collaboration diese Analyse genehmigt, um die Masseninformation bei der Suche nach großräumigen Anisotropien zu verdeutlichen. Dieser neuartige Ansatz zeigt, dass ein Massenschätzer bei der Suche nach Anisotropien bereits in Phase I der Datenerfassung des Observatoriums verwendet werden kann. Wir haben eine Trennung in den Dipolamplituden von zwei A-unterschiedlichen Teilmengen der Daten beobachtet, was auf eine Abhängigkeit der gemessenen Dipolamplitude von der Masse/Ladung hinweist. Die stärkste Trennung tritt im Energiebereich von 8 bis 16 EeV auf. Dort beträgt die berechnete Zufallswahrscheinlichkeit 2.9×10^{-5}, was einer zweiseitigen Signifikanz von 4.18σ entspricht. Die zusätzlichen Informationen können zum Verständnis der möglichen Quellen der ultrahochenergetischen kosmischen Strahlung

beitragen, da eine engere Eingrenzung der plausiblen Quellenszenarien erforderlich ist, um diese Messungen erklären zu können.

Diese und zukünftige masseninformierte Anisotropie-Suchen werden möglicherweise im Rahmen von AugerPrime, der die nächsten Generation der Datenaufnahme des Pierre Auger Observatory, verbessert werden. In der zweiten Phase der Datenaufnahme werden zusätzliche Nachweistechniken die Massensensitivität des Observatoriums verbessern.

Resumen

El Observatorio Pierre Auger lleva más de 19 años recogiendo datos, alcanzando más de 123 000 km^2 sr año de exposición acumulada, con los detectores de superficie distribuidos en 3 000 km^2. El descubrimiento más significativo hasta la fecha es la existencia de una estructura dipolar en las direcciones de llegada de los rayos cósmicos, con una amplitud total de aproximadamente el 7%. Esto resulta de la modulación observada en la ascensión recta en el intervalo de energía inclusiva superior a 8 EeV, donde la componente ecuatorial dipolar calculada tiene una significación estadística de 6.8σ. Recientemente se ha observado un patrón dipolar similar en los eventos con energías entre 8 y 16 EeV con una significación estadística de más de 5σ. La Colaboración Pierre Auger también ha reportado un aumento de la amplitud del dipolo con la energía. Se entiende que esta anisotropía es de origen extragaláctica, ya que el máximo del dipolo apunta a una dirección alejada $\sim 115^\circ$ del centro galáctico.

En el mismo rango de energía, la evolución observada de la profundidad del máximo desarrollo de la lluvia con la energía indica una progresión hacia una composición más pesada de los rayos cósmicos.

En esta tesis presento un método innovador para la búsqueda de anisotropía a gran escala basada en la composición. Por un lado, los eventos más livianos tienen mayor rigidez que sus contrapartes más pesadas; por lo tanto, sus trayectorias se ven menos afectadas por los campos magnéticos galácticos y extragalácticos. El efecto esperado es una mayor anisotropía en la dirección de llegada de un subconjunto de eventos con menor masa y carga que la anisotropía en el flujo total de rayos cósmicos. Por otra parte, la longitud de atenuación es distinta para cada grupo de masa, lo que conduce a diferentes horizontes de fuentes de rayos cósmicos para partículas livianas y pesadas a una energía determinada. Bajo un modelo agnóstico de las fuentes, investigo la amplitud del dipolo en función de la rigidez. Utilizando una biblioteca de simulaciones, analizo la posibilidad de medir una separación en la amplitud total del dipolo entre dos subpoblaciones con distinta massa A del conjunto de datos de Auger Phase I. Después de comprobar el método propuesto, la Colaboración Pierre Auger ha aprobado este análisis para utilizar la información de masa en una búsqueda de anisotropía a gran escala. Este nuevo método indica que se puede utilizar un estimador de masa en la búsqueda de anisotropías ya en la Phase I de toma de datos del Observatorio.

Observamos una separación en las amplitudes dipolares de dos subconjuntos con distinto A en los datos, lo que indica una dependencia de la amplitud del dipolo medido con la masa/carga. La separación más significativa ocurre en el rango de energía de 8 a 16 EeV. Allí, la probabilidad de que esta suceda por azar cuando ambas poblaciones tienen la misma distribución es de 2.9×10^{-5}, lo que corresponde a una significación bilateral de $4{,}18\sigma$. La información adicional puede ayudar a comprender las posibles fuentes de los rayos cósmicos de ultra-alta energía, ya que para describir con simulaciones las mediciones presentadas aquí se necesita una mayor restricción de las caracteristicas de las posibles fuentes. Esta y las futuras

búsquedas de anisotropía informadas por la masa probablemente se mejorarán en el contexto de AugerPrime, la fase ampliada del Observatorio Pierre Auger. En su segunda fase de toma de datos, las técnicas de detección adicionales mejorarán la sensibilidad del Observatorio a la masa de los rayos cosmicos.

CONTENTS

Chapter I

Introduction

Over a century ago, Victor Hess discovered that charged particles reaching Earth were the cause of the air ionization deep into our atmosphere [1]. This finding raised an essential question: where do these *cosmic rays* come from? Measurements on top of high mountains helped us understand how frequently these nuclei arrive and how their rate is related to their energy [2, 3]. The work developed by Heitler and later by Matthews [4] described the cascade of successive particle production due to a nucleus-atmosphere interaction. Those *extensive air showers* have spatial and temporal structures dependent on the parent nucleus energy and total mass. As a consequence, the measurement of some shower properties is crucial to recovering the initiating nucleus information. With this goal, satellite-born and ground-based observatories were built to learn more about cosmic rays and their origins.

The largest cosmic ray observatory is the Pierre Auger Observatory, covering an area larger than 3000 km^2 in Argentina. Due to its design, the 1660 surface detectors and 27 fluorescence telescopes can measure particles and radiation produced in extensive air showers initiated by cosmic rays with energies beyond those reachable in any human-made particle accelerator. More than that, these ultra-high-energy cosmic rays – UHECR – are not isotropically distributed in the sky. The high-quality data collected in Auger for almost two decades resulted in the discovery of a dipole structure in the arrival direction of UHECR. The interpretation is that the most energetic cosmic rays originate outside our galaxy.

Many are the unknowns surrounding the UHECRs, from the types of sources and the acceleration mechanisms pushing them to such high energies to the interactions with radiation and matter in the interstellar medium and the effects of magnetic fields in their trajectories. Nonetheless, from more universal physics knowledge, we expect that the nuclear mass and charge of these cosmic rays will relate to most of these open questions. For example, the effect of a magnetic field on a particle trajectory depends on its *rigidity*, the quotient between momentum and charge[1]. Therefore, it is clear that a fundamental piece of the "UHECR origin" puzzle is the mass composition of cosmic rays and how that changes with increasing energy.

---

[1] The rigidity of a particle with charge $Z\,e$ and momentum p is $R = pc/(Ze)$, with c the speed of light, and e the electric charge. In the relativistic regimen, $R \simeq E/(Ze)$, being E the total energy of the particle

In this thesis we investigate further the observed large-scale anisotropy . The observed dipole strength increases with energy in the UHECR regime while maintaining an approximately constant phase above 4 EeV that points away from our galaxy center. On the other hand, the Auger measurements indicate a progressively heavier mass composition of cosmic rays in the light of current hadronic interaction models. Therefore, we aim to combine these insights to probe a composition-dependent anisotropy in the arrival direction of UHECR.

For the first time, I used a mass estimator to subdivide the dataset into composition-distinct populations. I evaluated the differences and similarities in the reconstructed dipole in each subset. Due to the novelty of the analysis, we tested and scrutinized it using a simulation dataset that resembles the Auger measured data. With this, we could probe the quality of the analysis and, more importantly, evaluate its relevance under specific scenarios, even forecasting the reachable significance with the Phase I data of the Auger Observatory. With an exciting forecast, we proceeded with the *unblinding* of the data.

Following this introduction, the reader will be familiarized with fundamental concepts of cosmic rays in Chapter 2. A detailed description of the Pierre Auger Observatory is given in Chapter 3, describing the data used in this analysis. Then, Chapter 4 presents an overview of the anisotropy searches, focusing on the findings of the Pierre Auger Observatory. Before describing and probing the analysis proposal, the simulation dataset used to that end is described in Chapter 5, followed by the data-processing method adopted in Chapter 6. Then, with the data tools at hand, a description of the rigidity-dependent dipole model used to probe the relevance of the analysis is given and discussed in Chapter 7. In Chapter 8, the analysis parameters are fine-tuned using the simulation dataset and then applied to the rigidity-dependent model scenario. After ensuring the analysis is robust and reliable, the Pierre Auger collaboration agreed to apply it to the measured data. The data set with its important observed quantities are presented in Chapter 9 The results on the dipole amplitude of the defined populations are presented and discussed in Chapter 10. This thesis ends with Chapter 11, including a discussion summary and a preliminary interpretation of these results.

CHAPTER II

COSMIC RAYS

Since the discovery of cosmic rays the development of the astroparticle physics field has been promoted by the study of the possible sources, transport and acceleration mechanisms, and the interaction of such energetic particles with the interstellar medium, magnetic fields and Earth's atmosphere. The developments are not limited to astroparticle physics, as these studies have also allowed for the discovery of unstable particles, first reported by Lattes et al. [5]. In this chapter we will discuss the state of the art knowledge on the necessary conditions and possible environments that allow the nuclei to reach energies roughly one order of magnitude larger than human-made accelerators[1] [6]. We present an overview, but the in-depth discussion of these aspects is outside the scope of this thesis.

2.1. SOURCES AND ACCELERATION SITES

The processes causing the acceleration of the nuclei composing the flux of ultra-high-energy cosmic rays, UHECR, can be generalised in two categories: the direct, ″one-shot″ mechanisms, which depend on strong electric fields; and the stochastic acceleration in moving magnetized plasmas [6].

The first mechanisms relies on extreme scenarios, with high energy density and is thus associated with white dwarfs, pulsars or black holes. In the same environments, however, the energy losses processes may prevent the ultra-high-energies to be reached. See discussion in Ref. [6].

In the second class of processes, the energy increment occurs for each passage of the CR in the magnetised region, often called magnetic clouds. In the first description of such mechanism, Fermi [7] proposed the interaction of charged particles with wandering magnetic fields. The rate of energy gain is very slow, but this process gives rise to an inverse power law for the energy spectrum of the protons. Considering an isotropic cloud distribution and a relativistic cosmic ray, the average change in

---

[1] In the center-of-mass frame, the most energetic cosmic rays have an energy ~ 10 times larger than reachable by the LHC, the large hadron collider.

energy can be approximated to [8]:

$$\left\langle \frac{\Delta E}{E} \right\rangle \simeq \frac{8\, v_{\text{cloud}}^2}{3}, \tag{2.1}$$

where v_{cloud} is the speed of the cloud in units of the speed of light. Because the main dependence of the energy gain is on the cloud speed to the second power, this is known as the *second order Fermi acceleration mechanism*. As remarked by Fermi, this process is an incomplete explanation, since energy loss processes are competitive with the acceleration one, and only the particles above a certain energy threshold would be further accelerated. In addition, the time scale of such process in realistic astrophysical circumstances is of the order of$\sim 12 \times 10^6$ years [8].

A special scenario of this process considers 'head-on' collisions with the magnetized clouds. When the particles are traveling along the direction of the target's motion, the acceleration process becomes more efficient. An example of such scenario would be in the shock waves of matter ejection from a source, as in supernova remnants and active galactic nuclei [8]. The average energy gain is

$$\left\langle \frac{\Delta E}{E} \right\rangle \simeq \frac{4\, v_{\text{cloud}}}{3}, \tag{2.2}$$

thus being named the *first order Fermi acceleration mechanism*.

According to Ref. [8], a particle with initial energy of 1×10^9 eV would have to cross the magnetised region of a supernova remnant 100 times in order to reach an energy of $E \sim 1 \times 10^{15}$ eV. The process would take enough time that the majority of the sibling cosmic rays would have escaped the acceleration region. From that, one can expect that the supernova remnants can effectively accelerate cosmic rays up to energies of $E \sim 1$ PeV, several orders of magnitude smaller than the UHECR.

A common approach is to evaluate which acceleration sites would allow charged particles to achieve certain energy, by observing the "Hillas plot" . The minimal assumption is that the site should have a combination of dimension and magnetic field strength, sufficient to confine the particle until it reaches an energy E_{max}. This maximum energy achievable relates to a source with characteristic size, R, and magnetic field strength, B, via [8]

$$E_{\text{max}} =\sim Z\, B\, u\, L, \tag{2.3}$$

where Z is the charge of the particle/nucleus being accelerated, B is the strength of the magnetic field, u is the velocity of the shock, and L is the size of the magnetised region. In Fig. 2.1 we observe the relationship between the site size, here named R for the radius of such region, and the magnetic field intensity, B, for typical values of the source classes as annotated. In diagonal lines, the necessary condition for B and R in the case of a proton achieving a maximum energy of $E_{\text{max}} = 10^{21}$ eV (blue) and an iron with $E_{\text{max}} = 10^{20}$ eV (red).

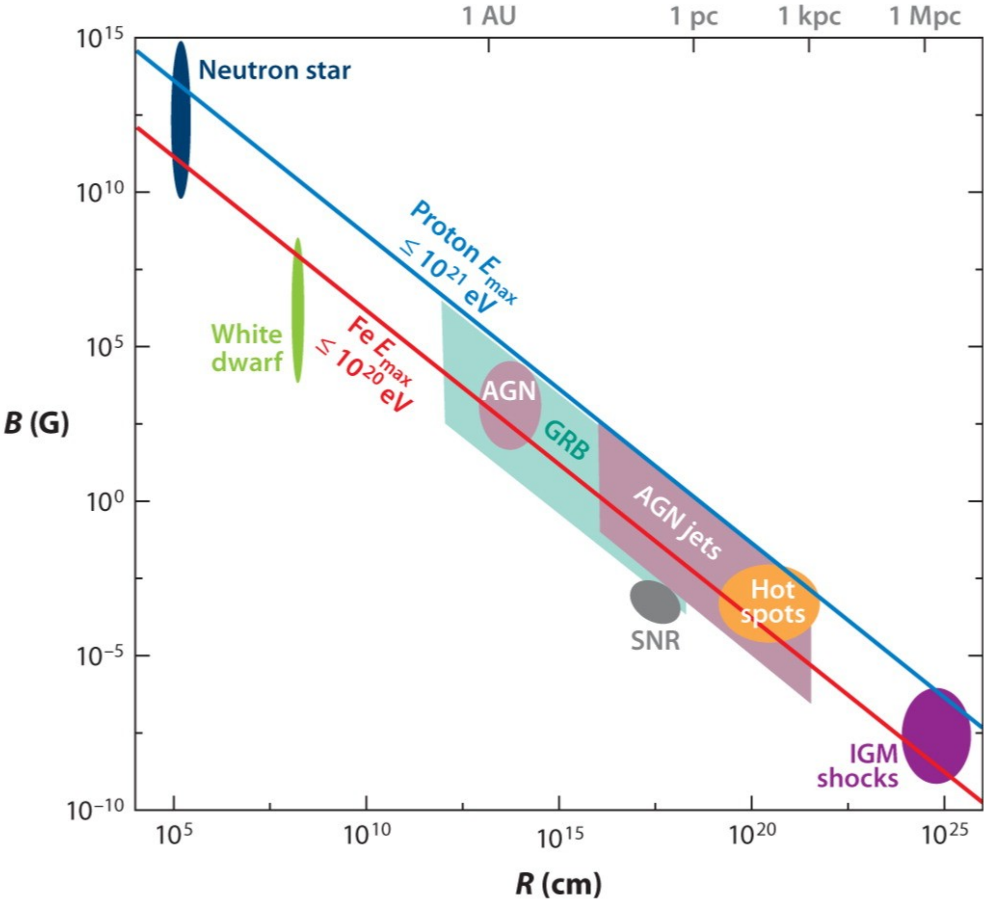


Figure 2.1: Typical values for the site size, R, and the magnetic field intensity, B, presented for distinct source classes, as annotated. In diagonal lines, the necessary condition for B and R in the case of a proton achieving a maximum energy of $E_{\text{max}} = 10^{21}$ eV (blue) and an iron with $E_{\text{max}} = 10^{20}$ eV (red). Abbreviations: AGN, active galactic nuclei; GRB, gamma-ray burst; IGM, intergalactic medium; SNR, supernova remnant. Reproduced from Ref. [9].

2.2. PROPAGATION

During their propagation in the interstellar medium, the UHECR will interact with the background radiation. The effect has first been pointed out by Greisen [10], and by Zatsepin and Kuzmin [11]; describing the photopion production in interactions of a highly energetic proton with the photons of the cosmic microwave background, CMB, and thus this is known as the GZK effect:

$$\begin{aligned} \gamma + p \longrightarrow \Delta^{+} \longrightarrow\ & p + \pi^{0} \\ & n + \pi^{+} \end{aligned} \tag{2.4}$$

Although the interaction with the CMB is dominant, photons with other wavelengths in the extragalactic background light, EBL, can also be part of such interactions. This is the main energy loss process for UHECR protons after they leave the acceleration region [12].

The interaction with the background radiation is not limited to photopion production, which dominates at energies $E > 10^{19}$ eV, but also via Bethe-Heitler [13] pair production at energies 10 times smaller. UHECR nuclei additionally undergo photo-disintegration, absorbing an impinging photon and subsequently fragmenting into an excited daughter nucleus and one or more nucleons. This is the dominating energy loss mechanism for UHECR nuclei [14].

In the Universe and specially in regions with ionized matter, magnetic fields in different configurations exist. Although the details of the magnetic fields are not known, measurements of the synchrotron radiation from the galactic plane region, for example, can help constrain the contributions from each component of the galactic magnetic field [15]. During the transport of charged CR, their trajectories are unavoidably affected by these fields. A coherent component of the magnetic field, that points to a single direction, will cause a deflection of the UHECRs that is inversely proportional to the nucleus rigidity [12]. This implies that particles with larger rigidity will "point" to regions closer to their sources than those with smaller rigidity. Other components to the magnetic field are the isotropic random, which points in all directions equally, and an anisotropic random component, that has a constant orientation but changes direction stochastically. The deflections due to these stochastic components of the magnetic field cause a scatter in the arrival directions of UHECRs [12].

The effects of the galactic magnetic fields on the UHECR trajectories have been largely studied, for example in Refs. [16, 17, 18]. In the quest to identify the possible sources and classes of sources of UHECR, all the propagation effects mentioned in this chapter need to be considered. Examples of such studies are the Refs. [19, 20, 21]. They often aim at reproducing with simulations the measurements by the Pierre Auger Collaboration while considering the propagation effects listed above. More about this approach will be discussed in Chapter 4.

2.3. A SUMMARY OF LATEST MEASUREMENTS

In Ref. [12] an extensive summary of the state of the art knowledge on UHECRs and related processes is provided. It provides also recommendations on strategies for upcoming observatories [22, 23] aiming to answer the open science questions. We reproduce here many of the results there reported.

The measurement of the all particle spectrum of CRs has paved the way of our knowledge. From the discovery of an increasing rate of ionizing particles with altitude by Hess [1], to the coinciding measurements of an array of detecting stations [2, 3] leading to the extensive air showers description [4]. The quantity is specially important as is related to the energy expenditure by the sources, but also carrying information on the acceleration mechanism itself and on the energy losses during propagation. The last is dependent on both the source distribution (characterising the distance traveled by the CRs) and the interstellar medium characteristics.

A compilation of the all-particle spectrum as measured by several observatories is presented in Fig. 2.2. One can see the intensity of cosmic rays multiplied by E^2 as a function of energy. This way the regions of change in spectral index, often called ″features″, are more pronounced. The steeply falling rate of events with increasing energy is a trend observed by measurements of all listed experiments. The gray areas seen indicate rate of 1 CR per squared meter per second, 1 per squared meter per year, and 1 per squared kilometer per year. The low rate of events per area and time at the highest energies is the reason why the direct detection of CR becomes unfeasible: if satellite-based detectors have an effective area of about 1 m^2, one event per second could be detected at energies around 10^{12} eV, while this rate would fall to one event per year for particles/nuclei with energies $\sim 10^{17}$ eV. Spectral features such as the knee ant the ankle are indicated in the figure, and will be further discussed in Section 5.2.

At the highest energies, the measurement of the mass or charge of the primary CR initiating an extensive air shower is not possible. Instead, an inference of the mean composition over an energy region can be obtained from different methods. This mean composition is commonly disentangled into contributions from four or five representative elements (for example, see Refs.[25, 26, 20, 21]. The method to obtain a mass estimator varies, however a key ingredient is the evaluation of the moments of the X_{max} distribution. X_{max} characterises the atmospheric depth where the maximal size of a shower is reached (see Chapter 3). Thus, it is important to evaluate this distribution as measured by the different observatories and techniques. In Fig. 2.3 we can observe the first (top) and second (bottom) moments, as reported in [27]. As will be discussed in the next chapter, the Pierre Auger Observatory provides measurements with different detector types: the fluorescence telescope, FD, and the surface detector, SD. Measurements by different observatories are shown. The general trend above $\sim 10^{17}$ eV is a dominance of light elements in the CR flux, while above~2 EeV the composition appears to become heavier for increasing energies. To interpret the trends, the continuous, dasehd and dotted lines indicate the expected dependence of X_{max} with energy for different hadronic interaction models. The expectations are shown for a proton (red) and an iron (blue) primary CR. More information can be found in Ref. [27].

As will be discussed in Chapter 5, there seems to be a steady progression towards heavier nuclei dominating the flux of the highest energies CRs. This can be a consequence of the ejection spectra at the sources or a hint of the propagation effects on heavier elements. Whether neither or both processes are promoting the composition transition, or other alternatives not yet explored, is as an open question deeply related to the sources of CR.

The next piece of information to uncover the possible sources of the CRs is the presence of anisotropies. These would indicate a dominant (class of) source over an isotropic background. As discussed in the previous sections, the propagation effects interfere with the measured energy and composition of CRs at Earth. Nonetheless, if one considers that the UHECR source distribution follows the large scale structure (LSS) of the universe, the expected outcome is an imprint of the LSS on the arrival direction of cosmic rays [28]. The dipole amplitude, for example, in a source scenario that follows the matter distribution in the universe is expected to be $\sim$ two times

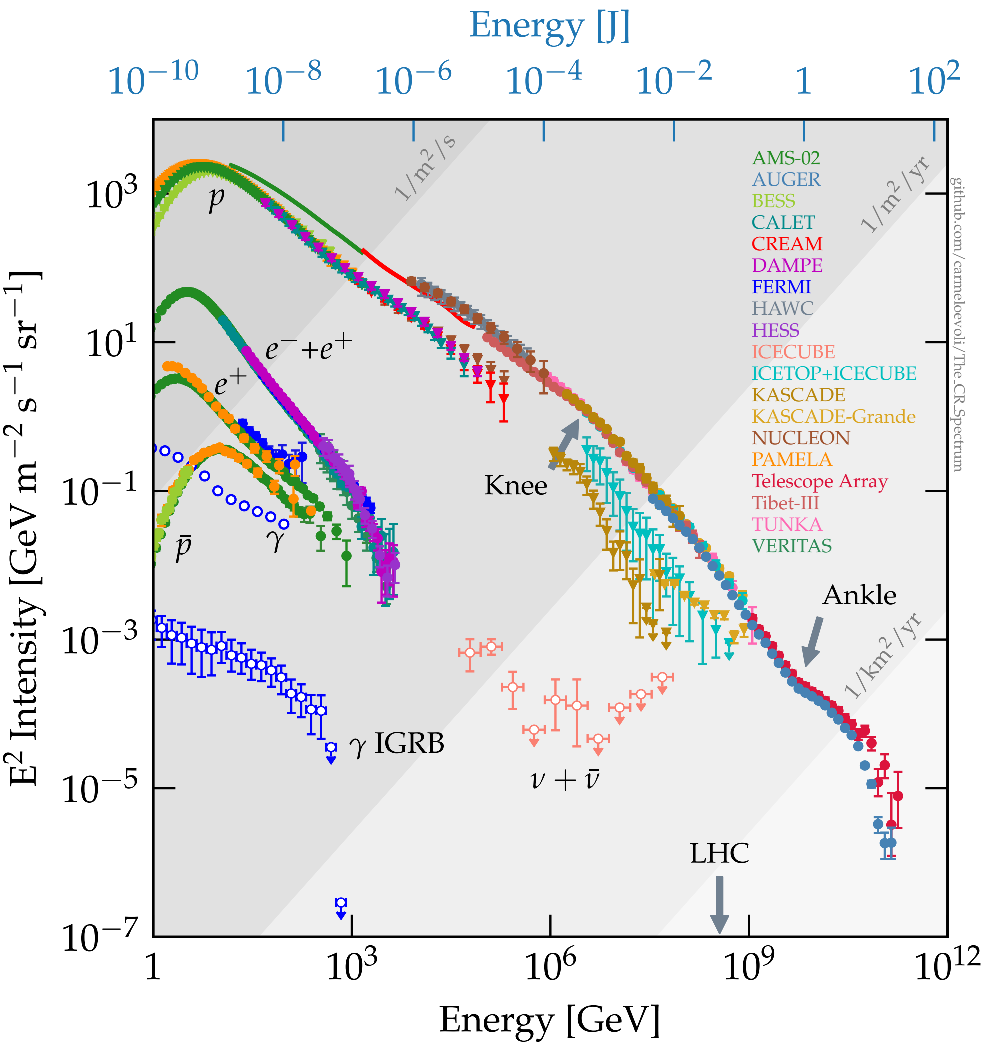


Figure 2.2: A compilation of the cosmic-ray energy spectrum measured by several experiments. Indicated with arrows, two well-established features can be seen: the *knee* and the *ankle*. At lower energies, the higher flux allows for direct detection, and when a particle identification is possible, the single-component spectrum is shown. At the highest energies, where an event-by-event identification of the CR is not yet possible, the all-particle spectrum is shown. Reproduced from Ref. [24].

larger than a scenario where sources are homogeneously distributed [29]. Whether one, a few or many sources dominate the flux of UHECR would have direct effects on the measured anisotropies at different scales. An illustration of the mass distribution in the local universe can be seen in Fig. 2.4, showing the matter density surveyed in the 2MASS survey [30].

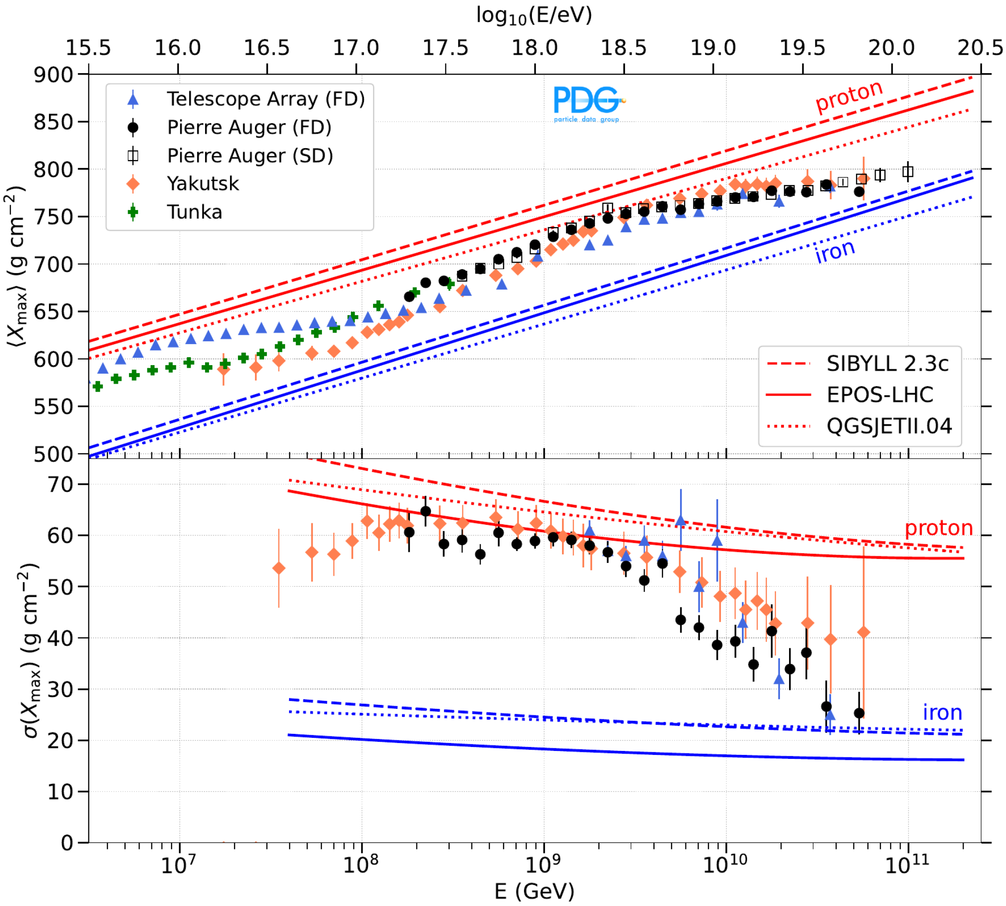


Figure 2.3: First (top) and second (bottom) moments of the $X_{\max}$ distribution, as measured by different observatories. The predictions for proton (red) and iron (blue) nuclei according to hadronic models are also shown. Different types of detectors can be used to measure $X_{\max}$, such as fluorescence telescope, FD, and the surface detector, SD. Reproduced from [27].

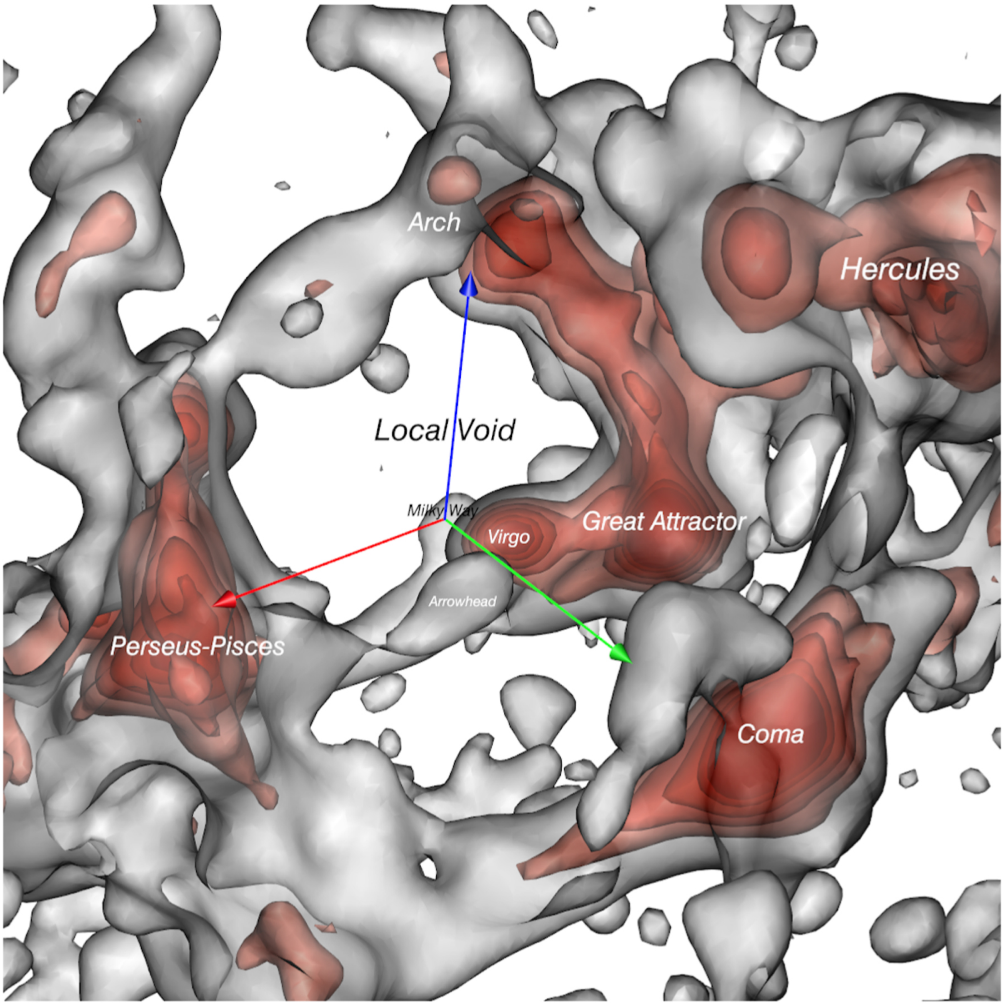


Figure 2.4: Matter distribution in close distance to our galaxy, according to measurements of the [30] survey and modeled in CosmicFlows [31]. Regions of same density present same color. Large nearby structures are also annotated. The Local Void fills the empty region above the Milky Way. Reproduced from Ref. [31].

CHAPTER III

THE PIERRE AUGER OBSERVATORY

3.1. THE SCIENTIFIC QUESTIONS

The flux of cosmic rays with energies above about 10^{14} eV is so low that it is unpractical to detect them directly either with balloon-borne or satellite-based experiments. Instead, measuring the secondary particles produced during the development of the extensive air showers through the atmosphere can aid in recovering some of the information of the *parent* nucleus. The interest in measuring the most energetic cosmic rays was motivated by the expectation that, at the highest energies, the charged nuclei would carry large rigidities, and thus the deflections by galactic or extra-galactic magnetic fields would be minimal. In other words, the hope was that by measuring the most energetic CRs those would point directly to their sources, allowing us to learn about the class of sources and acceleration mechanisms that lead to such highly-energetic nuclei. As reviewed by Alan Watson in Ref. [32], this expectation led to the construction of larger and larger experiments: Volcano Ranch, USA, covering 8 km^2; Haverah Park, UK, 12 km^2; SUGAR, Australia, 100 km^2; and Yakutsk, Siberia, 25 km^2.

In the 1970s, the observation of the atmospheric nitrogen excitation became a tool to measure the shower development. The electrons produced in the cascade of particles causes the excitation of N, and the posteriorly emitted fluorescence light can be observed at ~20 km distances on clear, dark nights. The process and observation technique of the atmospheric fluorescence will be discussed in a following section. For now suffices to say that new experiments emerged, such as Fly's Eye, and with that new technique not only the lateral spread of particles at a fixed height could be measured, but also the longitudinal development of the air showers. To this day, the most energetic cosmic ray ever measured was one named the 'Oh-my-God particle', measured in 1981 by Fly's Eye with an estimated energy of 3.2×10^{20} eV.

The Pierre Auger Observatory was designed to measure the extensive air showers of ultra-high-energy cosmic rays: those with energies above $\sim10^{18}$ eV=: 1 EeV [1]. Among the open questions motivating and guiding the design of the observatory were the ones listed by Sommers in Ref. [33]:

---

[1] EeV stands for Exa-electron-volt. Roughly equivalent to 0.1 Joules.

1. How are particles produced with such high energies?[2]
2. Why the measured UHECR do not clearly correlate to recognisable sources?
3. Why the flux of UHECR is not as suppressed as expected from the GZK effect?

The scrutiny of a pure proton composition assumption was one of the goals, together with the understanding of the origin of such energetic CRs. Because the flux of CRs decreases rapidly with energy, it was clear then that a large area should be covered by any observatory aiming at investigating the events with the highest energies. In 1995 a six-month design workshop was held to work on the proposal by Jim Cronin of a very large air shower array [36]. From this intensive work a Design Report was produced, including the scientific discussion, proposed design, and a cost estimate [37]. Following a research and development period with 32 prototype stations and two fluorescence telescopes in the Engineering Array, the validity of the design and the performance of the detectors was demonstrated. In the final form, a triangular grid of water-Cherenkov stations (WCD) regularly spaced 1500 m apart, was chosen as the best approach for the Pierre Auger Observatory. The deployment of the 1660 detecting stations in Malargüe, Argentina, was initiated in 2002 and finalised in 2008; with the data acquisition starting in 2004 with 154 active surface detector stations [36]. With the completed surface array, the surface detector spans over an area of $\sim$3000 km^2.

The surface array of the Observatory is overseen by four Fluorescence telescopes sites, each site originally containing 6 telescopes, and since 2010 three additional telescopes constitute an addition to the Coihueco site. The detection of events with both a telescope and a WCD constitues a *hybrid* event. In those cases, the angular resolution is $\sim 0.6°$, while it is 1.7° for events with energy between 3 and 10 EeV detected only by WCDs. The details of the detectors and detection modes will be discussed in the next section. The reader is invited to read Ref. [36], where a thorough description of detector components, the data acquisition system, the monitoring efforts and the data processing are detailed, among other important aspects of the sophisticated Auger Observatory.

Twenty years after the begin of data acquisition in the Auger Observatory, all the motivating questions remain open. We have learned that reality is far more complex than initially assumed, and unforeseen progress has been achieved. Remarkably, we have learned from Auger measurements that the composition of the most energetic particles is not proton-dominated over all energies, but rather progresses towards heavier elements. We have also learned that the spectrum of CRs has several features, including the suppression of the flux at energies beyond those stipulated by the GZK effect – which agrees with a heavier composition. And finally we have learned that it is quite difficult to pinpoint the sources of such CRs, but observing the distribution of the arrival directions of these nuclei, a large-scale dipole pattern was found with high significance.

---

[2] In the review by Sommers, making reference to the measurements by the Fly's Eye experiment [34] and by AGASA [35], with energies well beyond the suppression of extra-galactic protons due to interaction with the extra-galactic background radiation.

3.2. EXTENSIVE AIR SHOWERS

When a high-energy proton or heavier nucleus impinges Earth's atmosphere, a nuclear cascade will develop through the atmosphere [38], as a consequence of the collision with target nuclei[3]. The collection of particles and radiation produced in this cascading process is named *extensive air shower*, EAS. The collision processes therein are dominated by hadrons, and is common to call the collection of such byproducts the *hadronic component* of the air shower, which propagates longitudinally in the direction of the initial momentum vector of the initiating CR [40]. Because of the scattering processes and the perpendicular component of the initial momentum vector, the cascade also spreads laterally.

Another large part of the EAS is the *muon component*, which is mainly produced via the decay of charged pions and kaons. It is noteworthy to comment that the pion was one of the particles first discovered via the measurement of EAS at mountain altitudes [5]. The reader can find in Ref. [5] the iconic particle tracks in nuclear emulsions that led to the discovery of the charged pion.

The third "ingredient" of EAS is the *electromagnetic component*, composed of electrons, positrons and photons. This component dominates the EAS if it was initiated by photons and electrons, which often occurs at much lower energies than the UHECR regime. It is unclear if photons at the highest energies exist and could be differentiated from hadronic showers in the Pierre Auger Observatory [41], although machine learning techniques can aid in the search for photon-initiated showers. One of the "feeding" mechanisms for this component is the decay of a neutral pion, $\pi^0 \rightarrow 2\gamma$, which has an extremely short lifetime of 8.4×10^{-17} s [42], and thus the cascade initiated by this process will develop very high in the atmosphere.

While the charged pions are also mainly produced in the stratosphere, their decay products from $\pi^{+(-)} \rightarrow \mu^{+(-)} + \nu_\mu(\bar{\nu_\mu})$ [42], will transverse the atmosphere mostly without decaying or being brought to rest. While the muons, $\mu^{\pm}$, can be detected while passing through the surface stations, the neutrinos have a very small interaction cross-section and are unlikely to be recorded within the typical volume of 12 000 liters of water in these detectors [38].

As the majority of the particles produced in the showers are charged and propagate nearly at the speed of light, Cherenkov radiation, air fluorescence, and radio emission will also be byproducts of shower development through the atmosphere. The amount of each contribution, as well as the amount of particles, and the longitudinal and lateral developments of the EAS depend both on the primary mass and energy [40].

We present in Fig. 3.1 an illustration of an EAS with its components; and in Fig. 3.2 the contribution of each component to the total signal (depicted in black) in the station with highest signal in a simulated shower initiated by a proton with energy $E = 10^{19.4}$ eV and zenith angle of $\theta = 36°$. There we note contributions not only from the muonic (blue) and main electromagnetic[4] (red) components, but also the electro-

---

[3] The majority of our atmosphere is composed of the following gas molecules: nitrogen, N_2, oxygen, O_2, argon, Ar, and carbon dioxide, CO_2 [39].

[4] The main electromagnetic component is that produced in high-energy π^0 decays.

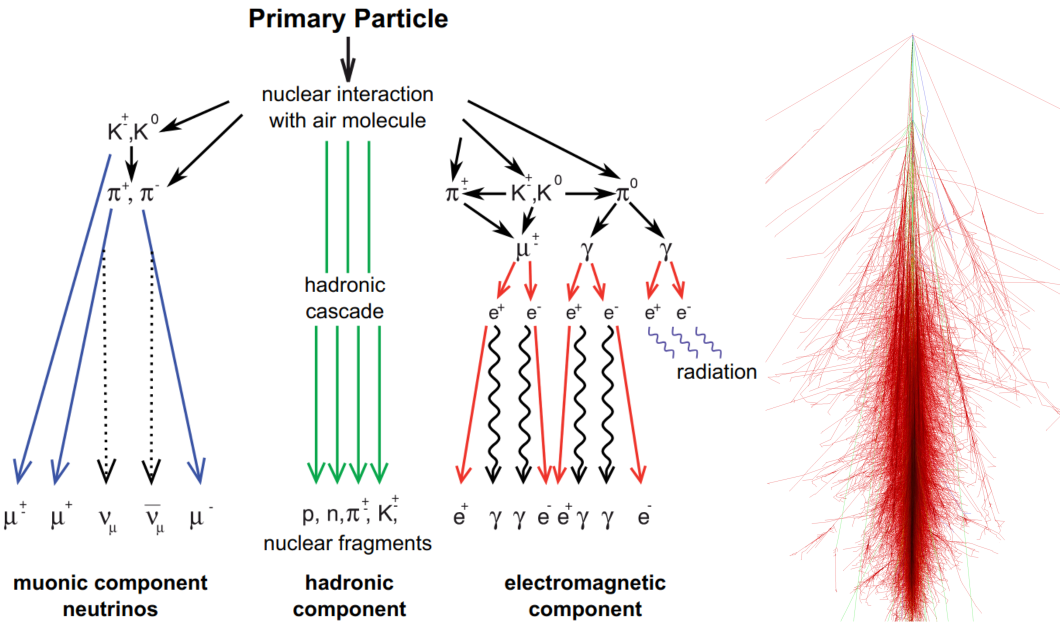


Figure 3.1: Illustration of the three components of an extensive air shower. The cascading process, initiated by the interaction of the primary particle with a constituent of the atmosphere, can be understood in terms of its muonic, hadronic, and electromagnetic components. On the left side, a schematic detail for the components and its constituents, reproduced from Ref. [43]. On the right side, a proton-initiated vertical shower with energy of 10^{12} eV as simulated with CORSIKA. The y axis in this image spans over 30 km of shower development, being the ground level at the bottom of the graph and the first interaction occurring at 30 km, on the top of the image. Same color coding for both images: blue represents the muonic, green the hadronic, and red the electromagnetic component. Reproduced from Ref. [44], where additional simulated showers are available.

magnetic particles from muon decay or interactions (green), and the electromagnetic particles from the decay of low-energy hadrons (yellow).

Although the EAS are highly complex processes better reproduced via thorough computational simulations, a generalised description can be achieved analytically. Adapting the electromagnetic cascade model by Heitler [46] to describe hadronic interactions, Matthews reported in Ref. [4] a set of cascade equations that provide predictions for measurable properties of extensive air showers.

Since the first interaction, the energy of the parent particle is split among the daughter particles. For the neutral pions, the particles decay almost immediately, giving rise to EM showers as discussed before. For the charged pions, the particles will travel through a column of matter before interacting and producing a new generation of pions. The process repeats until a critical energy is reached, where it becomes more likely that a daughter particle decays than interacts. At this point, all pions are assumed to produce muons, which will then reach the observation level.

The point in the shower development where the number of particles start to decrease define the shower maximum, and the atmospheric depth at which this

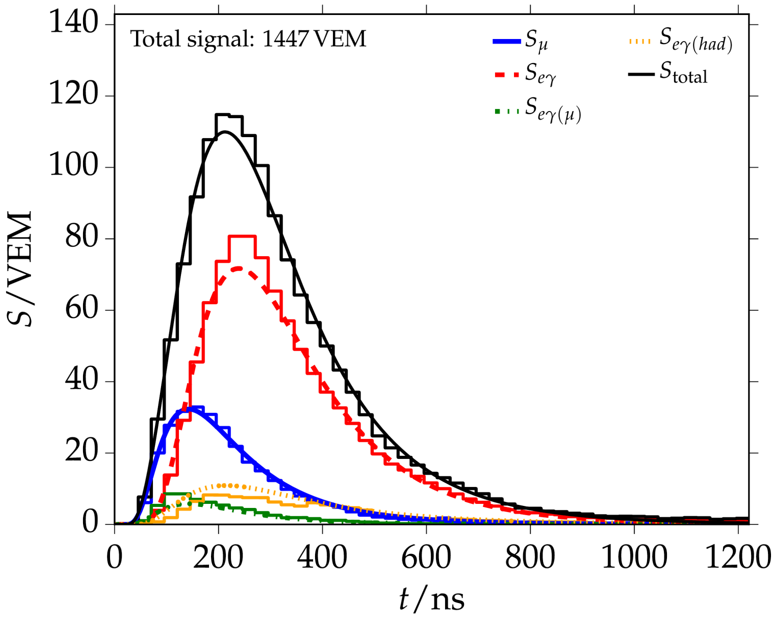


Figure 3.2: Contribution of each component to the total signal in the surface detector station with highest signal, from the fit of the total trace (black) according to the Universality model [45]. The simulated shower has as primary a proton with energy $E = 10^{19.4}$ eV and zenith angle of $\theta = 36°$. Model predictions are shown in smooth lines, while the components traces are shown as histograms. Reproduced from Ref. [45].

occurs, i.e. the integrated column of matter from the top of the atmosphere until that point, is the depth of maximum shower development, X_{max}, measured in $\text{g}\,\text{cm}^{-2}$.

For simplicity, we reproduce here the cascade equations that describe measurable quantities in nuclei-initiated shower, considering a simple superposition of A proton-initiated showers, as reported in Ref. [4]:

$$E_0 = 0.85\,\text{GeV}\,(N_\text{e} + 25N_\mu) \quad (3.1)$$

$$N_\mu^A = N_\mu^\text{p}\,A^{0.15} \quad (3.2)$$

$$X_{\text{max}}^\text{A} = X_{\text{max}}^\text{p} - \lambda_\text{r} \ln A \quad (3.3)$$

In Section 3.2, the energy E_0 of the CR is shown to be distributed over the electromagnetic and muonic components, proportionally to the number of particles in each: N_e and N_μ, respectively. From Section 3.2, the number of muons in a shower initiated by a nucleus increases with the atomic mass A. Thus, we can expect an iron-initiated shower to have ~1.8 times more muons than a proton-initiated cascade of same energy. Conversely, a shower initiated by a heavier primary will present an $X_{\text{max}} \approx 150\,\text{g cm}^{-2}$ shallower than the proton counterpart (this means higher in the atmosphere, farther from the observation level), according to Section 3.2. There, λ_r represents the radiation length in the atmosphere, which was reported in Ref. [4] to be $\lambda_\text{r} = 13.8\,\text{g}\,\text{cm}^{-2}$.

Not only the absolute content of particles is useful for retrieving the mass information of the primary, as described with the cascade equations, but also the temporal distributions of each shower component. The height where the bulk of particles in one component is produced is a reflection of the shower development stages. As a consequence, the particle arrival times at the detecting station level will carry this information, as can be seen in Fig. 3.2. A discussion of the *Universality* principle is carried out in Chapter 6.

3.3. DETECTORS AND DETECTION MODES

The Auger Observatory was designed to operate in hybrid mode [37], with the fluorescence detector (FD) providing a calorimetric measurement that sets the energy scale of the observatory [47]. While the fluorescence detector sites can detect events in clear, moonless nights and thus have a duty cycle of ~14%, the surface detector (SD) can detect at nearly 100% of the time [36]. In Fig. 3.3 we can see the map of the Observatory with the detecting stations and telescopes. The laser facilities XLF and CLF[5] are also shown at the center of the array; they are crucial for the identification of atmospheric parameters and for the calibration of the telescopes. The three bays of the high-elevation Auger telescopes, HEAT, are also shown. Additionally the Auger Engineering Radio Array, AERA, and the Underground Muon Detector, UMD, as a set of AMIGA[6] stations, are depicted. These are engineering arrays for testing the proposed detectors of the AugerPrime upgrade of the Observatory.

3.3.1. SURFACE DETECTOR

Each of the WCD stations is an autonomous unit, powered by solar panels and with batteries that allow for continued data-taking. When long periods of cloud coverage prevent the charging of the batteries and the charge reaches a critical value, a control board in the electronics has a function which allows the remote operator to set the station(s) into hibernation [36]. This is one possible cause for a station to become temporarily unavailable for data-taking, the status of which being recorded for proper accounting of the exposure of the Observatory.

The SD stations are cylindrical tanks with an inner liner made of flexible, reflective material, which diffusely reflects the Cherenkov light produced in the water volume when the energetic charged particles transverse it [36, 37]. The Cherenkov light can be detected by three photomultiplier tubes, which are connected to a unified board which hosts a station-level data acquisition system. Information recorded in each station is then shared with the central data acquisition system, CDAS, including calibration data for each transmitted event [36]. The details of the station-level conditions for the event recording do not pertain to the scope of this thesis. For further information, we direct the reader to Ref. [36].

The recorded events need to fulfill a set of requirements, both on station-level and on event-level. For example, one of the station-level trigger conditions evaluates

---

[5] CLF stands for *central laser facility*, while XLF indicates the *extreme laser facility*. More details can be found in Ref. [36].

[6] AMIGA stands for Auger Muon and Infilled Ground Array, although nowadays it is used to refer to one UMD station.

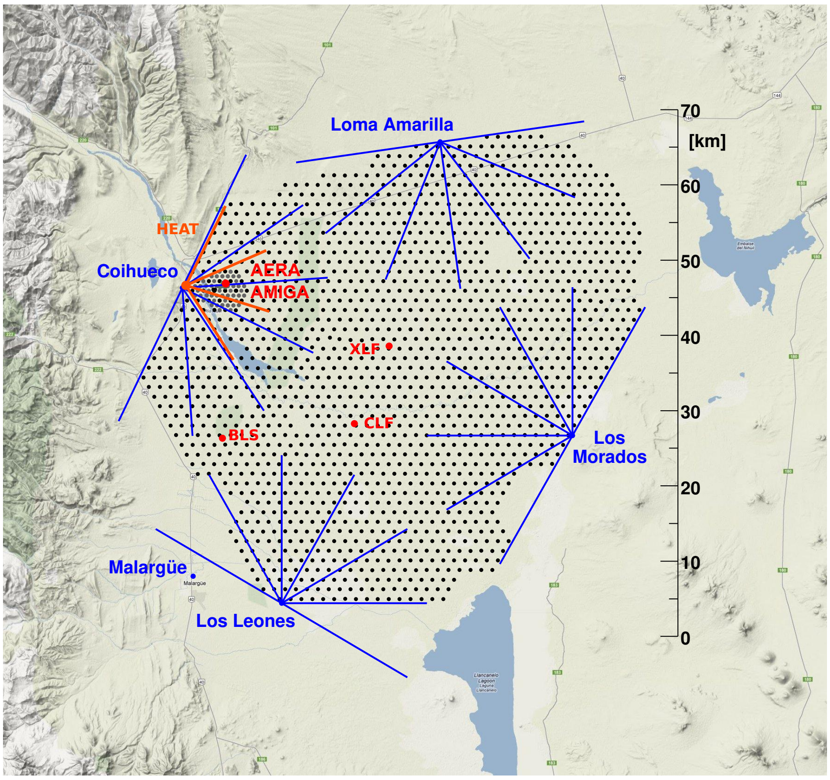


Figure 3.3: Schematic of the Pierre Auger Observatory, with the surface detector stations depicted in black dots, and each fluorescence telescope site marked with its corresponding name. The blue lines indicate the field of view of each telescope bay. Additional systems are annotated at the corresponding position in the Observatory. Centrally to the array, two laser facilities for atmospheric studies are shown. Reproduced from Ref. [48].

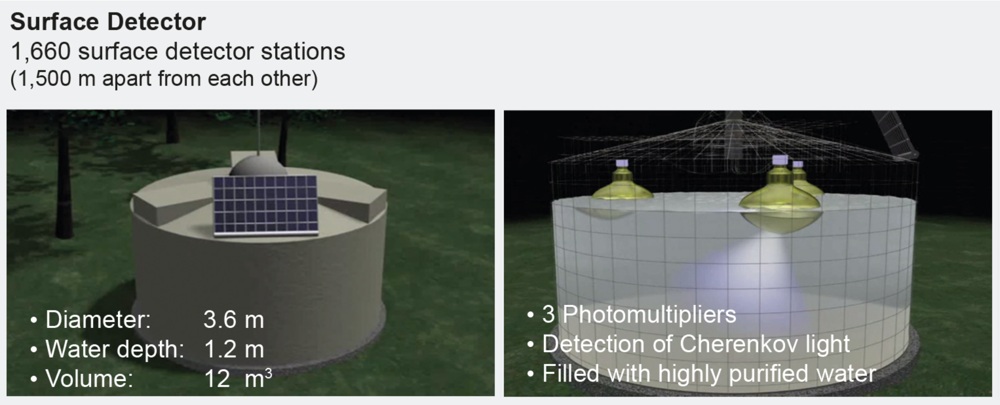


Figure 3.4: Visual representation of a Water-Cherenkov detector. On the right side, one can observe the display of the three Photomultiplier tubes in the detecting station. Reproduced from Ref. [49].

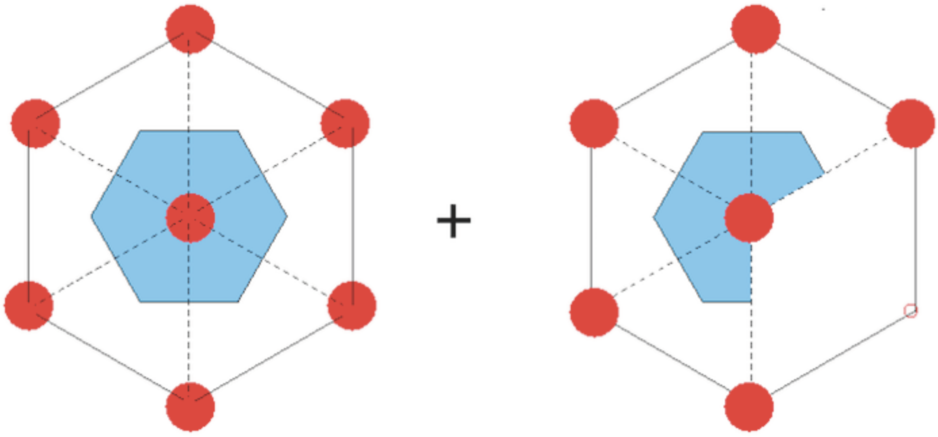


Figure 3.5: Schematic of the event-level T5 condition. On the left side, all 6 stations in the hexagon enclosing the station with highest signal are active (filled red circle). On the right side, 5 stations in the crown are available for data-taking. The blue area indicates the region where a shower core would be closer to the highest-signal, central station than to any other. Reproduced from Ref. [51].

if the recorded signal is above a certain threshold in a time window, while another aims at excluding muons from random background being recorded as an event (see Ref. [50] for more information). The first event-level trigger evaluates if the geometric location and timing of the triggered stations follow the expectations of a shower front moving at speed of light, i.e. if they are spatially and temporally correlated in the expected way. Additional event-level conditions might be required. For example, in the context of arrival directions analyses, the *crown* of stations surrounding that with highest signal must have 5 or 6 stations available for data-taking, fulfilling the *T5* condition. Additionally, in the case of 5 active stations, the shower core must fall within an isosceles triangle formed by active stations of the crown. An schematic of the event-level T5 condition can be seen in Fig. 3.5, where active stations are shown in filled red circles, and the blue shaded area indicates the region where a shower core would be closer to the highest-signal, central station than to any other. The yearly rate of events with energy above 3 EeV that fulfill the T5 condition is ~14 500 [36], which corresponds to roughly 40 events per day.

To probe the flux of CR at lower energies, the Pierre Auger Collaboration has built two denser surface arrays, nested in the standard one. The grids of stations spaced at 433 m (SD-433) and 750 m (SD-750) apart allows for a full detection efficiency to be reached at lower energies than in the standard array. In Ref. [52], the authors report for the first time the measurement of the *second knee* with the surface detector, observing the feature at ~230 PeV. Only the most dense array of surface stations, SD-433, reaches energies as low as $E = 63$ PeV, covering the region of this spectral feature with enough energy bins to characterize the transition region. The SD-433 is formed by 19 water-Cherenkov detectors identical to those deployed in the main array, and is nested within the SD-750 with which it shares seven detectors. The recording of events with the SD-433 began in 2013 [52]. An schematic view of the SD-433 and

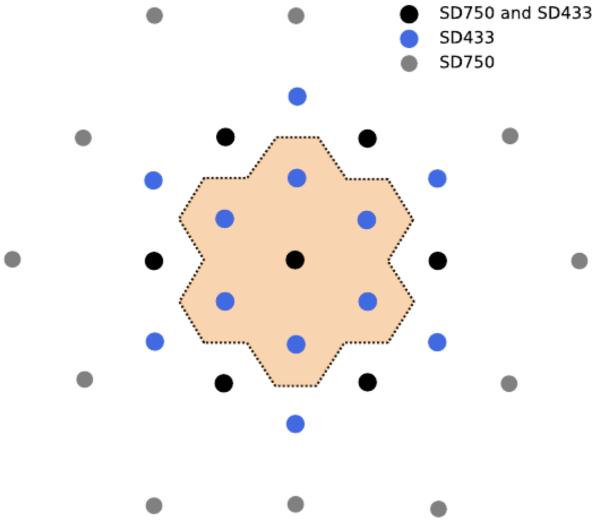


Figure 3.6: Schematics of the nested arrays SD-433 (blue) and SD-750 (gray), which have stations separated by 433 m and 750 m, respectively. Some stations are shared by both arrays (black). Reproduced from Ref. [52].

SD-750 nested arrays can be observed in Fig. 3.6.

Irrespective of the spacing of the grid of stations, the detection technique is the same. Once particles deposit Cherenkov radiation in the water, and a signal is recorded by the station, the information of all the triggered stations is used to determine the shower plane, and with that, the arrival direction of the cosmic ray. This is done by evaluating the time distances between the triggering of close-by stations.

Then, a fit of a lateral distribution function that describes the average rate of fall-off of the signals as a function of the distance from the shower core is performed. It takes into account the probabilities for the stations that did not trigger and the stations close to the shower axis with saturated signal traces [36]. From the fitted lateral distribution function, the signal reconstructed at 1000 m, $S(1000)$, is the estimator of the primary energy. A value of $S(1000) > 20\,\mathrm{VEM}$[7] corresponds to a primary CR energy $E > 3\,\mathrm{EeV}$, thus above the threshold of full-efficiency of the SD main array.

3.3.2. FLUORESCENCE DETECTOR

The Fluorescence Detector is formed by the four buildings that face towards the inner part of the array of WCD stations. With exception of HEAT, the FD sites hosts six bays, each containing an independent telescope, in a climate-controlled room. As each telescope has a field of view of 30° × 30° in azimuth and elevation, each site covers 180° in azimuth.

In clear moonless nights, the shutter and curtain that cover entrance of light to the bay are opened. The radiation reaching the bay first passes through a UV filter,

---

[7] VEM = vertical equivalent muon; is used as a unit of deposited signal in a WCD, and corresponds to the signal produced by one muon traversing the tank on a vertical trajectory [36].

effectively reducing the man-made background and effectively improving the signal-to-noise ratio of the air shower signal. The light is then reflected by a segmented spherical mirror and collimated in a set of photomultiplier tubes, constituting the pixels of the imaging camera.

The telescopes can collect radiation from the longitudinal development of showers, thus a calorimetric estimate of the primary CR is available by accounting for the deposited energy in the atmosphere, even at large distances. For a reliable and precise reconstruction of the air shower longitudinal profiles,the absolute calibration of the detector response is essential. This calibration determines the conversion of the analog counts to a light flux for each pixel that receives a portion of the signal from a shower [36]. Since 2013, only a relative calibration has been performed, which is only sensitive to changes of the telescope camera. Recently, a new end-to-end calibration tool was developed and tested: the XY-scanner fully described in Ref. [53]. Being portable and easy to mount on a telescope, this tool overcomes the largest issues of the previous calibration procedure: the high amount of work and person-power required. With the new calibration tool, a higher accuracy of the estimation of the cosmic-ray energies is reached.

The telescopes, in contrast to the surface detectors, are operated remotely by shift personnel. Located either at the central campus of the Observatory or in a remote shift room, the volunteering Pierre Auger collaborators are responsible for the preparation of the FD for a run, for making relative calibrations, starting and stopping runs, and online checking of the quality of measured data [36]. Additional monitoring equipment in each telescope site provide the shift personnel with atmospheric information such as temperature, rain, wind speed, and cloud coverage; all of which very important for the safety of the equipment.

As it is often possible to observe the majority or the entirety of the longitudinal development of EAS with the FD, the location of the shower development maximum is obtained at higher precision than that estimated from the measurements with the SD.

3.3.3. HYBRID DETECTION MODE

The hybrid events are those registered by the FD and having at least one triggered SD station. The event reconstruction is based on fluorescence detector data with additional timing information from the surface detector. This means that the arrival direction is more accurately measured using the SD timing information, while the energy and X_{max} estimators rely mostly on the FD counterpart. Those events detected in hybrid mode present an angular resolution on the arrival direction of ~0.6° and a resolution of the shower core position of 50 m; while the energy resolution, inherited from FD, is around 8%, and the X_{max} resolution is of $20\,\mathrm{g\,cm^{-2}}$.

While events recorded only by SD provide much larger statistics, the angular resolution in that case is at most 0.9° when the event contains more than 5 triggered stations, and 1.6° when only three stations contribute to the event. Similarly, the energy resolution of SD-only events is ~12% at the highest energies. For the in-depth discussion of the detection modes, we point to Ref. [36].

Aside from the increased resolution of hybrid events, another advantage of the

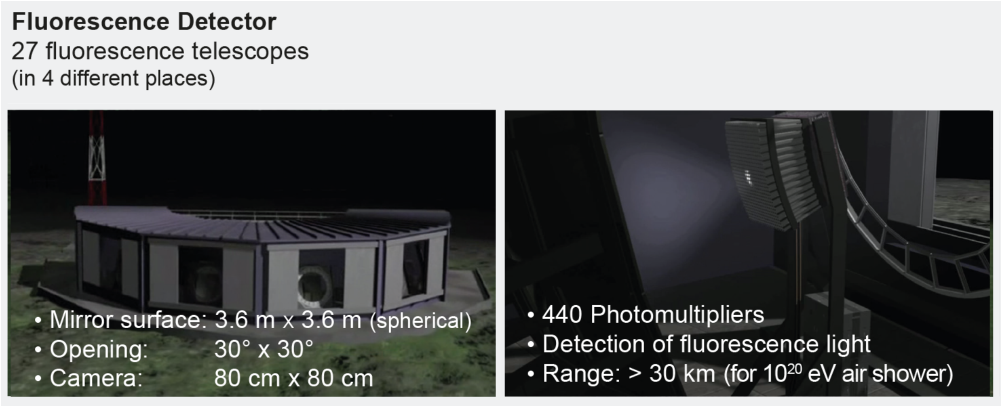


Figure 3.7: Visual representation of a Fluorescence Detector station. On the left side image, one can observe the arrangement of the six telescope bays, covering a total azimuth angle of 180°. On the right side image, the representation of the telescope mirror reflecting the fluorescence light on the 440 photomultipliers that constitute the telescope imaging pixels. Reproduced from Ref. [49]

hybrid detection mode is the possibility of calibrating the shower estimators in the two modes. For example, the SD energy scale is derived from the FD one, as the later has a more accurate energy estimator.

3.4. AugerPrime: the upgrade of the Pierre Auger Observatory

Following a decade of data-taking by the Pierre Auger Collaboration, the acquired knowledge of a flux suppression at energies above 5.5×10^{19} eV and the observation of a progressively heavier composition of UHECR above $\sim$2 EeV, the Pierre Auger Collaboration recognized that shower-by-shower inference of the mass of the primary cosmic ray is essential to bridge the gap in our knowledge on the origin of such energetic nuclei. A proposal to upgrade the Observatory was prepared in 2015 and made public afterwards. The reader can find the Preliminary Design Report in Ref. [54].

Other scientific goals of *AugerPrime*, the upgraded phase of data-taking of the Observatory, are the search for a (subdominant) proton flux at the highest energies, and the study of hadronic interactions in the shower development — as our current knowledge relies on modeled extrapolation of the measurements in particle accelerators.

Better estimation of the muon content of the air showers is the main goal of the upgrade of the observatory. With the improved information, we can rely on the surface detector to obtain the composition measurement up to the flux suppression region, aiding on the understanding of the origin of this spectral feature. With the muon-number measurement, one can also improve the discrimination power between hadron- and photon-initiated showers, increasing the current photon and neutrino sensitivity [54]. Most importantly for the topic of this thesis, the muon-content measurement will allow for a more accurate shower-to-shower identification

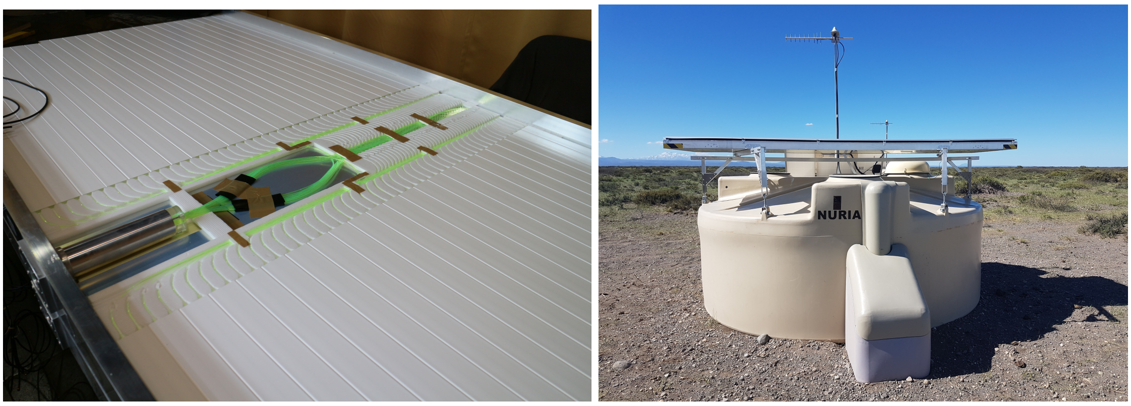


Figure 3.8: *Left*: Picture of a SSD detector with the green wavelenght-shifting fibers in the center that are guided to the photomultiplier tube. *Right*: SSD station mounted on top of a WCD station. The scintillation detector is housed in an aluminium box to protect its components from the wind and dust, and to facilitate the transportation during deployment [54]. Reproduced from Ref. [56].

of the CR initiating the shower, and with that further mass-informed anisotropy studies will be performed.

After a period of design, tests, and approval by Collaboration, the upgrade of the Observatory results in the addition of the scintillator surface stations (SSD), the radio detector (RD), and the underground muon detector (UMD). The upgrade also includes the exchange of the water-Cherenkov station electronics to a modern version thereof. The upgraded unified board, UUB, as the modern electronics for Auger-Prime is called, has more acquisition channels and therefore allows the processing of signals of all the detectors in a SD station, including those incorporated with the upgrade. The UUB also increases the time resolution of recorded signals [54].

The SSD stations are placed on top of the WCD stations. Each SSD consists of 48 scintillator strips, distributed over two panels as depicted in Fig. 3.8. Each strip is 1.6 m long, and hosts a wavelength shifting optical fiber. The fibers are placed in a “U” configuration that maximizes light yield and allows the use of a single photomultiplier [54]. Each SSD station covers an area of 3.8 m^2 and is more sensitive to the electromagnetic component of the shower. Because the WCD is more sensitive to the muonic component, the simultaneous measurement with SSD and WCD improves the sensitivity of the SD stations to the mass composition of UHECR [55].

The UMD consists of plastic scintillation detectors buried 2.25 m underground, detecting nearly exclusively muons from the showers as the other components will be largely absorbed in the soil. Distributed over 61 locations of the SD-750 array and 12 locations of the SD-433 array, a total of 219 UMD scintillation modules are deployed, and the set of three modules corresponding to an SD station is called a *UMD counter*. Each module consists of 64 scintillator strips, 32 on each side of the module. The strips, covered in reflective material [54], are 4 m long and host wavelength shifting optical fibers to guide the photons to the silicon photo-multipliers [57].

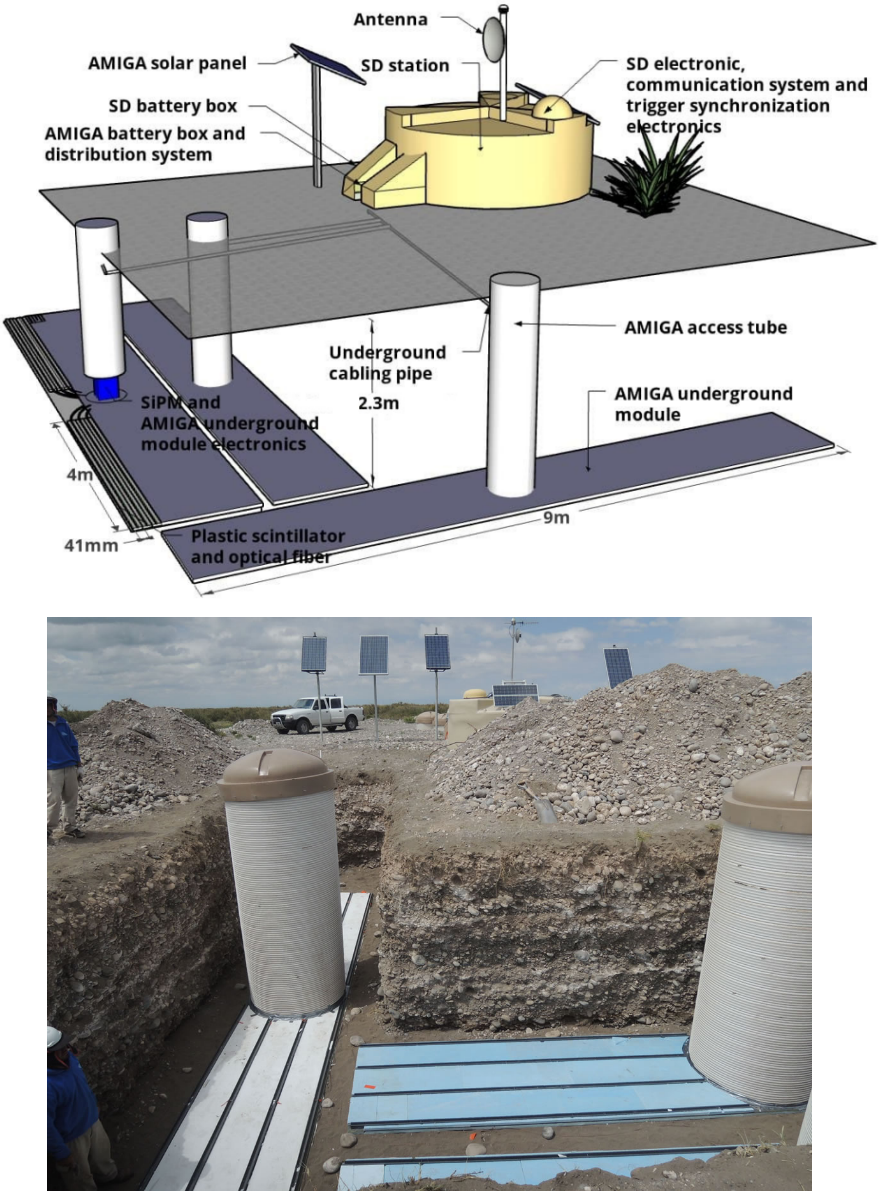


Figure 3.9: *Top*: schematic of a UMD counter with its three modules. Annotated, details of the scintillation strips and corresponding SD station. The counter is referred to as AMIGA (Auger Muon Infill for the Ground Array), the name adopted before the counters in the SD-433 were included. Reproduced from Ref. [58]. *Bottom*: Deployment of a UMD counter, with the exposed modules. The parts can be identified from the schematic version. Reproduced from Ref. [59].

Figure 3.10: Radio-detecting station in the Auger Engineering Radio Array, AERA. Reproduced from Ref. [56].

Since 2009, the Pierre Auger Collaboration examines the feasibility of radio observations of air showers with the Auger Engineering Radio Array (AERA). With the radio technique, the goals are: the calibration of the radio emission from the air showers; the demonstration of the radio technique for the measurement of extensive air showers; and the measurement of the cosmic-ray composition from 0.3 to 5 EeV, specially important in elucidating the transition from galactic to extragalactic cosmic rays [54]. Two components of the electric field measured at ground level have been identified [60]: the dominant one is perpendicular to Earth's magnetic field, while the second component is polarized radially with respect to the shower axis. In Ref. [60], the radio-energy estimator were shown to scale quadratically with the cosmic-ray energy by performing an absolute energy calibration against the surface-detector information in two years of data. More recently, the Collaboration has shown in Ref. [61] the agreements between FD and radio measurements of X_{max}, the shower maximum development depth, and for the first time the Collaboration reported a composition measurement with AERA [62], for cosmic rays with energies between $10^{17.5}$ and $10^{18.8}$ eV. The compatibility of the energy and X_{max} measurements using different techniques, together with the nearly 100% duty cycle of the radio detector with a X_{max} resolution better than $15\,\mathrm{g\,cm^{-2}}$ at the highest energies, makes for an exciting development of the scientific results by the Collaboration. The reader is invited to visit Refs. [60, 61, 62] to be informed about the radio technique in the context of the Auger Observatory. An example AERA station is depicted in Fig. 3.10.

The Pierre Auger Collaboration provides a plethora of publicly available material. From informative videos [63] and pictures [56] of both historical and scientific importance; to data structure information and insights on the main analysis performed by the Collaboration, together with 10% of its data [64].

CHAPTER IV

ANISOTROPIES ON THE ARRIVAL DIRECTION OF COSMIC RAYS

4.1. PREAMBLE

In 2017, a remarkable scientific breakthrough was the discovery of a large-scale structure in the arrival direction of ultra-high-energy cosmic rays by the Pierre Auger Collaboration [65]. The Collaboration first observed the dipole structure when considering all events with energies above 8 EeV. The amplitude of this modulation was sufficient to exclude a fluctuation of an isotropic distribution of the arrival directions as a source of the measured anisotropy. The isotropic scenario could originate such a structure with a chance probability of $P(\geq r) = 2.6 \times 10^{-8}$, being excluded with a corresponding two-sided significance of 5.2σ. Another remarkable feature of the measured dipole is its orientation: the excess of cosmic rays points 115° away from our galaxy center, indicating an extragalactic origin of this anisotropy.

With the measurements in the following years, a more extensive dataset allowed for the further evaluation of the large-scale anisotropy in the arrival direction of cosmic rays, now in intervals of the particles' energy. The Auger Collaboration reported a progression of the dipole amplitude with increasing energy and the excess of CRs consistently being away from the galactic center for energies above 4 EeV.

The searches for higher-order multipoles and correlations to specific, plausible sources have also been topics of high interest. The absence of a significant quadrupole amplitude is a unique feature in the data. The excess of cosmic rays around the Centaurus region for events with energy above 38 EeV is equally intriguing.

With the increasing quantity and quality of data, and the continuous development of more accurate techniques, the Pierre Auger Collaboration is actively engaged in its goal to identify the possible sources of the UHECRs. Considering galactic magnetic field models and distributions of plausible sources, researchers have probed which scenarios are most consistent with the measurements of the spectrum, arrival directions, and maximum shower development reported by the Auger Collaboration.

In the following sections, we will explore the methods and most recent results of the different approaches for searching for anisotropies in the arrival direction of CRs. The focus remains on the large-scale anisotropies, the approach adopted in this analysis. However, we will also discuss the results in synergy with the other scales of anisotropy.

4.2. LARGE-SCALE ANISOTROPY

Considering the rotation of the Earth and the translation around the Sun, an observatory at Earth collecting data over several years would have a uniform exposure as a function of right ascension, RA, for a given range of declination corresponding to its field of view. Once potential detector effects are excluded, a persisting modulation of the rate of events as a function of RA indicates a modulation of astrophysical origin.

The data acquisition at the Auger Observatory started before the full deployment of all 1600 detectors of the main array. Consequently, over a few years, the effective area — and therefore the Observatory's exposure— was not constant but increasing. Due to varying reasons, an SD detecting station can become temporarily unavailable for data-taking, reducing the Observatory's effective aperture. These modulations in aperture would cause a modulation in the measured equatorial dipole amplitude of $d_\perp < 4 \times 10^{-4}$ [66]. In addition, the array is tilted in the southeast direction. Together with the geomagnetic effects on the shower development, the tilt of the array can affect the d_z dipole component by +0.004 when considering events with zenith $\theta \leq 80°$ [66]. To account for the small nonuniformities in exposure in right ascension and for the tilt of the array, to each event, a weight w is assigned according to

$$w_i = [\Delta N_{\text{cell}}(\alpha_i^0)(1 + 0.003 \tan\theta_i \cos(\phi_i - \phi_0)]^{-1} . \tag{4.1}$$

In Eq. (4.1), $\Delta N_{\text{cell}}(\alpha_i^0)$ is the normalized rate of events as a function of the right ascension of the zenith of the Observatory α_i^0 at the time of detection of the i-th event. While ϕ_0 is the azimuth of the tilt of the array and θ_i and ϕ_i are the zenith and azimuth angles, respectively, of the CR initiating the measured air shower.

An angular distribution on a sphere is decomposable into spherical harmonics. Spherical harmonics are a complete set of orthogonal functions defined on the surface of a sphere; they are the analogous to the Fourier series of periodic functions, defined on a circle [67]. Therefore, the spherical harmonics can describe the CR arrival direction distribution in the celestial sky[1]. In the absence of full-sky coverage, a common practice is to combine the amplitudes of two Fourier series to obtain the dipole amplitude and direction. The weighted first harmonic, in terms of a measurable x, is described by the two amplitudes a_1^x and b_1^x as

$$a_1^x = \frac{2}{\mathcal{N}} \sum_{i=1}^{N} w_i \cos(x_i), \quad b_1^x = \frac{2}{\mathcal{N}} \sum_{i=1}^{N} w_i \sin(x_i) \tag{4.2}$$

where each event i is associated with a weight w_i, as defined in Equation 4.1, and $\mathcal{N}$ is the sum of weights over the N events. The amplitude in right ascension

---

[1] The celestial sky in this context refers to that observable at Earth. In other words, the abstract sphere concentric to Earth and with an arbitrary radius. This reference system is the equatorial coordinate system.

$r_1^\alpha = \sqrt{(a_1^\alpha)^2 + (b_1^\alpha)^2}$, while the phase[2] is obtained through $\varphi_1^\alpha = \arctan \frac{b_1^\alpha}{a_1^\alpha}$. The analogous is obtainable for a modulation in the events' azimuth angle, ϕ.

From the coefficients of the first harmonics in right ascension, α, and in azimuth, ϕ, the decomposition of a dipole amplitude into components that are perpendicular and parallel to the Earth rotation axis are given by

$$d_\perp \simeq \frac{r_\alpha}{\langle \cos\delta \rangle} \qquad d_z \simeq \frac{b_\phi}{\cos \ell_{\mathrm{obs}} \langle \sin\theta \rangle}, \tag{4.3}$$

respectively. In Eq. (4.3), $\langle \cos\delta \rangle$ is the mean cosine of the declination of the events, $\langle \sin\theta \rangle$ is the mean sine of the zenith angle of the events, and ℓ_{obs} corresponds to the average latitude of the Observatory. For the events measured at the Pierre Auger Observatory, $\ell_{\mathrm{obs}} \approx -35.2°$. The dipole direction in equatorial coordinates is defined via a phase α_d and a declination δ_d:

$$\alpha_d = \varphi_\alpha \qquad \tan\delta_d = \frac{d_z}{d_\perp}. \tag{4.4}$$

From the considerations of the uncertainties in the measurement of the dipole components, discussed in detail in Ref. [66], the statistical uncertainty dominates over the systematic effects. The spurious modulations in azimuthal angle, solar hours and anti-sidereal frequencies originating from the detector, atmospheric, and geomagnetic field effects will be discussed in Chapter 9.

The Auger Collaboration has used this robust harmonic analysis to report the measured large-scale anisotropies [69, 70]. The latest measurements are reproduced below from Ref. [70]. The detailed information for each relevant energy bin is reported in Table 4.1. The normalized rate as a function of right ascension and the related flux map for the events with energies above 8 EeV can be seen on the left and right sides of Fig. 4.1.

An illustration of some of the multipoles on the surface of a sphere can be seen in Fig. 4.2. The possible dipole patterns corresponds to $\ell = 1$, shown in the middle row. The quadrupole patters are seen in the bottom row, as $\ell = 2$, and present with five distinct modes.

---

[2] The *arctan2* function is used, where the resulting angle ϕ lies within $-\pi < \phi < \pi$. as the single argument arc-tangent function only returns angle measures in the interval $-\frac{1}{2}\pi < \phi < \frac{1}{2}\pi$. See Ref. [68].

Table 4.1: Latest results for the dipole analysis reported. For each energy bin denoted in the first column, the following information is presented: the number of events therein, N, the equatorial component of the dipole, $d_\perp$, the North-South component, d_z, the total amplitude d, the dipole direction via its right ascension, α_d, and declination δ_d, and the probability of obtaining, from fluctuations of an isotropic distribution in arrival directions of CR, a Rayleigh amplitude larger than the measured.

E/EeV	N	$d_\perp/\%$	$d_z/\%$	$d/\%$	$\alpha_d/^\circ$	$\delta_d/^\circ$	$P(\geq r_1^\alpha)$
$4-8$	118 722	$1.0^{+0.6}_{-0.4}$	-1.3 ± 0.8	$1.7^{+0.8}_{-0.5}$	92 ± 28	-52^{+21}_{-19}	0.14
≥ 8	49 678	$5.8^{+0.9}_{-0.8}$	-4.5 ± 1.2	$7.4^{+1.0}_{-0.8}$	97 ± 8	-38^{+9}_{-9}	8.7×10^{-12}
$8-16$	36 658	$5.7^{+1.0}_{-0.9}$	-3.1 ± 1.4	$6.5^{+1.2}_{-0.9}$	93 ± 9	-29^{+11}_{-12}	1.4×10^{-8}
$16-32$	10 282	$5.9^{+2.0}_{-1.8}$	-7.0 ± 3.0	$9.4^{+2.6}_{-1.9}$	93 ± 16	-51^{+13}_{-13}	4.3×10^{-3}
≥ 32	2 738	$11.0^{+4.0}_{-3.0}$	-13.0 ± 5.0	$17.0^{+5.0}_{-4.0}$	144 ± 18	-51^{+14}_{-14}	9.8×10^{-3}

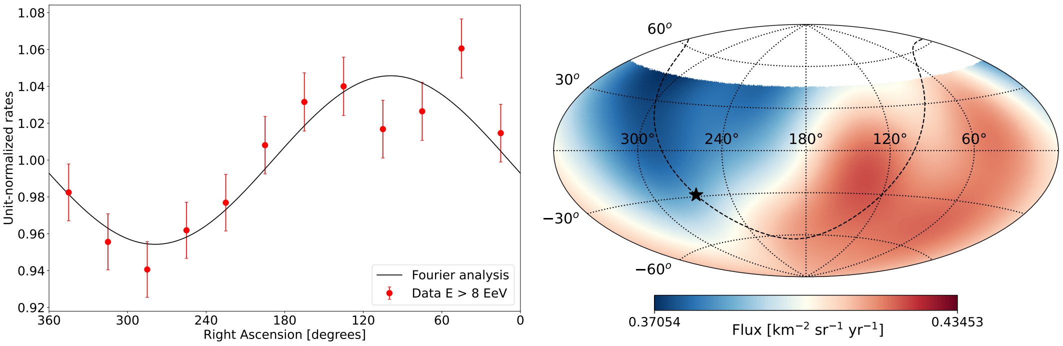


Figure 4.1: Normalized rate of events as a function of right ascension (*left*) and flux of cosmic rays in equatorial coordinates (*right*). A star locates the galactic center, while the dashed line represents the galactic plane. Both plots are representations of the dipole-like excess of UHECRs with energy above 8 EeV. Reproduced from Ref. [70].

Since its discovery, the measured dipole structure that points 115° away from our galaxy center supports the interpretation of an extragalactic origin of the UHECR. Remarkably, the excess of cosmic rays lies far from the galactic center, as seen in Fig. 4.1 and in the right plot of Fig. 4.3. Evaluating Fig. 4.3, the change in phase and the increase of the equatorial dipole amplitude for the energy bins above 2 EeV indicate that the transition between galactic and extragalactic origin of UHECR anisotropy occurs around this energy. The cause of such transition is not yet fully understood. The increase of the total dipole amplitude with energy is seen in the left plot of Fig. 4.4. Interestingly, the Auger Collaboration has reported a feature in the spectrum of cosmic rays at 5 EeV [52, 71, 72], which can also be associated with this transition. Notably, in a mixed composition scenario, the transition between the majority of UHECR having a galactic or extragalactic origin can occur at different energies for the nuclei, introducing substructures in the spectrum. The relationship between the

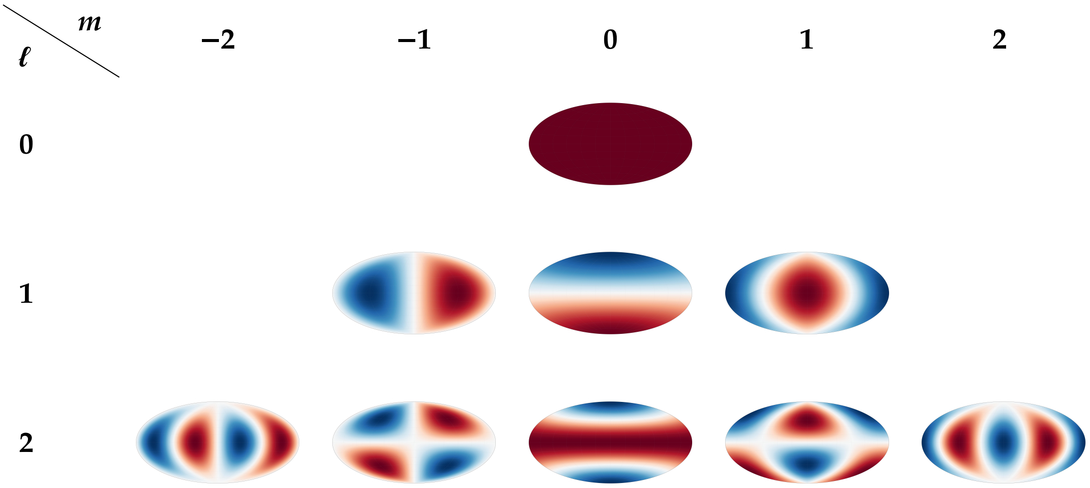


Figure 4.2: Illustration of multipoles on the surface of a sphere. On the first line, $\ell = m = 0$ corresponds to the monopole, where no localized excess is seen. This would correspond to an isotropic scenario in the context of the arrival direction of cosmic rays. The middle plots illustrate dipole patterns, as $\ell = 1$. The three possible modes, orthogonal to each other, are $m = \{-1, 0, 1\}$, from left to right. The last row of plots illustrates the case of $\ell = 2$, the quadrupole. The five possible modes, orthogonal to each other, are $m = \{-2, -1, 0, 1, 2\}$, from left to right.

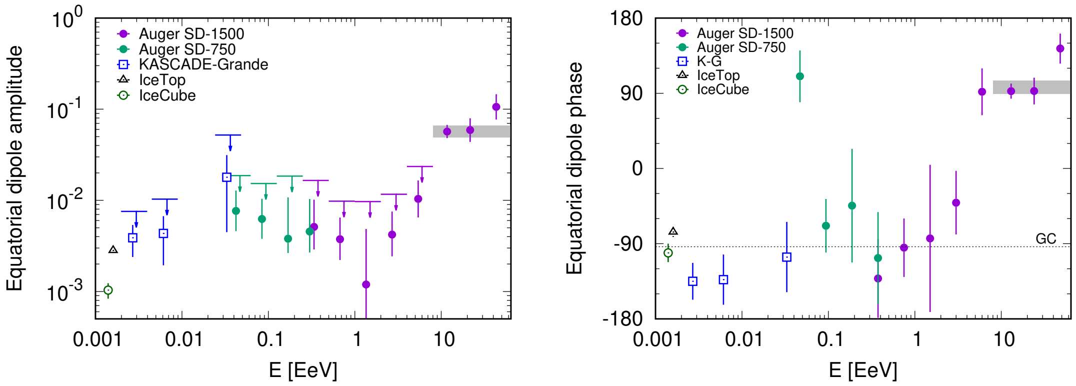


Figure 4.3: Equatorial dipole amplitude (*left*) and dipole phase (*right*) as a function of energy. In purple markers, the measurement with events detected with the SD-1500 array, and in green corresponding to measurements with the SD-750 array. Upper limits are shown with arrows corresponding to the amplitude where $P(> r_1^{\mathrm{UL}}) = 1\%$. Reproduced from Ref. [70].

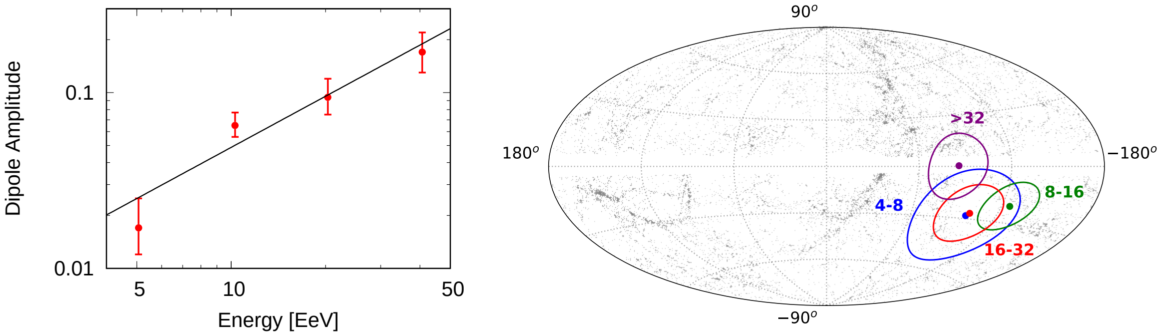


Figure 4.4: Total dipole amplitude, d, as a function of energy (*left*) and dipole direction in galactic coordinates (*right*) for each of the four energy bins: 4 to 8, 8 to 16, 16 to 32, and above 32 EeV. The dipole amplitude has been fitted as a function of energy in the form $d(E) = d_{10} \times \left(\frac{E}{10\,\mathrm{EeV}}\right)^{\beta}$, obtaining $d_{10} = 0.049 \pm 0.009$ and $\beta = 0.97 \pm 0.21$, in agreement with previous measurements. Reproduced from Ref. [70].

observed spectral features and the origin of CR, and even the characteristics of the UHECR sources and acceleration regions, is still to be understood.

Different scenarios have been probed in [73] to understand the source of the measured dipole amplitude. The authors showed, under the assumption of a pure proton composition, that even in the scenario of a homogeneous distribution of sources, an increasing-with-energy dipole amplitude is expected due to energy losses in the propagation of UHECR and the consequential dominance of closer sources to the flux of nuclei. It has also been shown that if the sources are not isotropically distributed, the expected anisotropy is larger than the homogeneous scenario. When considering that the source distribution is similar to that in the 2MASS catalog, an increase by a factor of 2 was found.

In Ref. [29], the authors extended the analysis to probe the expected dipole anisotropies in homogeneous and 2MASS-like scenarios in the case of a mixed composition of UHECRs. Again, the authors found a factor 2 enhancement of the dipole anisotropy for the 2MASS compared to the homogeneous scenario. The energy losses, nuclear fragmentation, and deflections in distinct extragalactic magnetic field strengths were considered as CR propagation effects. This showcased the transition between diffusive and quasi-rectilinear propagation of the UHECR. Evaluated under a composition scenario consistent with the spectrum and mass composition measured in Auger, the results for the 2MASS-like scenario indicate a trend of increasing dipole amplitude with energy compatible with the recent measurements. The analysis, however, did not include the propagation effects in the galactic magnetic field, which will dampen and shift the direction of the measured dipole.

In Ref. [70], a similar evaluation considers equal-luminosity sources distributed according to the 2MASS catalog within 120 Mpc of distance to Earth, while the contribution to the flux from farther away distances is taken as isotropic. Two source number density scenarios were considered, namely 10^{-5} Mpc^{-3} and 10^{-4} Mpc^{-3}, and the propagation of CRs from these source distributions were simulated with SimProp. In sequence, the dipole and quadrupole amplitudes were computed. This

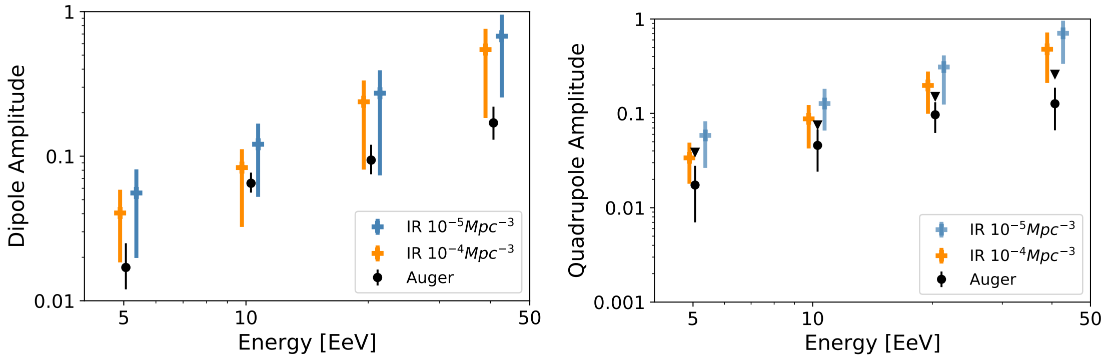


Figure 4.5: Total dipole (*left*) and quadrupole (*right*) amplitudes as a function of energy for the four energy bins: 4 to 8 EeV, 8 to 16 EeV, 16 to 32 EeV, and above 32 EeV. Measurements are shown in black closed markers, together with the 99% C.L. upper limit for the quadrupole amplitudes, in black triangles. Colored markers represent distinct source-density scenarios, where sources are distributed according to the 2MASS catalog, but with source densities of 10^{-5} Mpc^{-3} and 10^{-4} Mpc^{-3}, for blue and orange markers, respectively. Reproduced from Ref. [70].

way, the interactions of the charged nuclei with the extragalactic background radiation and with regular Galactic magnetic field could be taken into account. In Fig. 4.5, reproduced from Ref. [70], the dipole and quadrupole amplitudes are presented for each of the four energy bins in the two source-density scenarios. The non-homogeneous source distribution yields a compatible dipole amplitude, as the dipole expectations are consistent with the measurements within the uncertainties. Nonetheless, a tension to the measured quadrupole amplitude is seen, especially for the lower source density scenario, as can be observed in Fig. 4.5.

4.3. MULTIPOLAR ANALYSES AND INTERMEDIATE SCALE

In addition to the dipole and quadrupole moments, the power spectra for the different energy bins can also be evaluated in comparison to fluctuations of the isotropic scenario. When comparing to the 99% C.L. from Monte Carlo realizations of the isotropic scenario, the higher order multipoles powers that are larger than this upper-limit are C_{17} and C_8 in the 4 to 8 EeV and 16 to 32 EeV energy bins, respectively. The corresponding angular scales are $\sim 180°/\ell \approx 11°$ and $\approx 23°$, respectively. The power spectrum for these energy bins can be seen in Fig. 4.6. In agreement with the results from the Fourier analysis described in the previous section, the C_1 corresponding to the dipole pattern is well above the upper limit of isotropic fluctuation in the following energy bins: 8 to 16 EeV, 16 to 32 EeV and the above 8 EeV. However, after a statistical penalization for searches over different multipoles and energy bins, their significance is 3.3% and 26.5% respectively. All other multipoles in all energy bins are not above the upper limits obtained with the realizations of the isotropic scenario, described in Ref. [70]. In Fig. 4.7, the angular power spectrum measurements for the energy bins where a dipole structure is significant in the Rayleigh analysis.

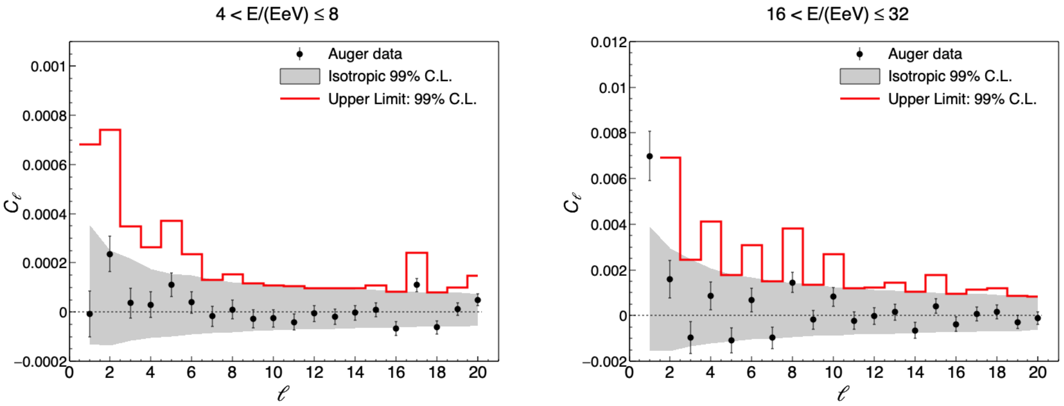


Figure 4.6: The angular power spectrum for the events in the energy bin 4 to 8 EeV (*left*) and 16 to 32 EeV (*right*). In these energy bins, the multipole coefficients C_{17} and C_8, respectively, are larger than the 99% C.L. obtained with simulations of isotropic background. The red line indicates the upper limit on the amplitudes with 99% C.L., taking into account exposure effects on the measurement of the amplitude C_ℓ^{rec} given a true amplitude C_ℓ. For more information and results on other energy bins, see Ref. [70].

Reproduced from Ref. [70].

In addition to evaluating higher-order multipoles, the Pierre Auger Collaboration also probes for localized excesses, either over the field of view of the Observatory or around regions of interest. The motivation is identifying possible sources, classes of sources, or acceleration regions.

In the first approach, the *overdensity* search is performed by computing the number of measured events in a region and comparing it to the expected number of events in the isotropic scenario. The loci of interest can be either known astrophysical objects or larger structures, like the galactic or supergalactic planes. Results reported in Ref. [74] are reproduced in Section 4.3.

As discussed in Section 4.2, the searches for large-scale anisotropies indicate that the origin of UHECR is most likely extragalactic. In agreement with that conclusion, the search for overdensities around the galactic center and galactic plane regions are not significantly distinct from isotropic expectations, as seen in Figs. 4.8a and 4.8b. The minimum p-value for the overdensity search around the supergalactic plane is found for a band of ±20° around the plane, and as with the galactic structures, does not present significant deviation from expectation in the isotropic scenario, as depicted in Fig. 4.8c.

The Centaurus region has long been of special interest [75]. In studies of correlation with catalogs of sources, described in the next session, there is an indication of an excess of events in this region, which hosts three galaxies at distances of ~ 4 Mpc: Centaurus A, NGC 4945, and M 83. The local p-value for an excess in this region can be seen in Fig. 4.8d, reproduced from Ref. [74].

Another approach, the *autocorrelation* search, is characterized by counting the number of event pairs separated by a small angular distance. Again, this quantity is compared to the expectation from an isotropic scenario to evaluate the significance of this excess, depicted in Fig. 4.8e. We refer the reader for more details on the searches

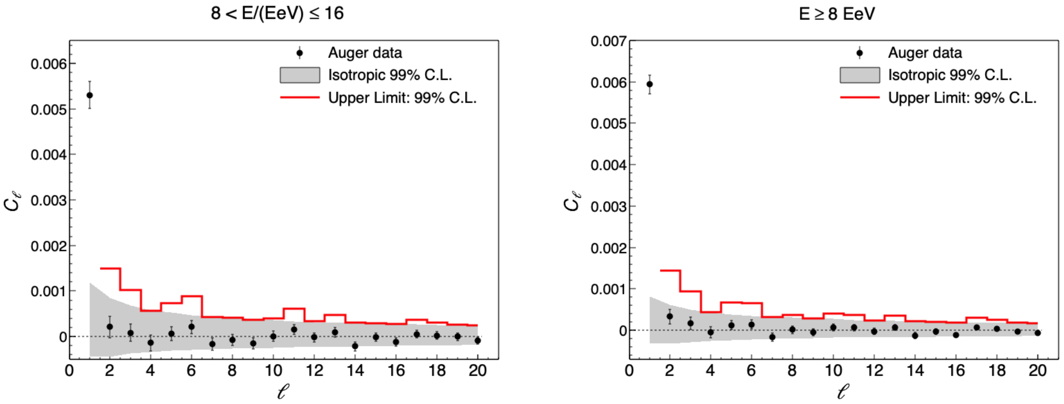
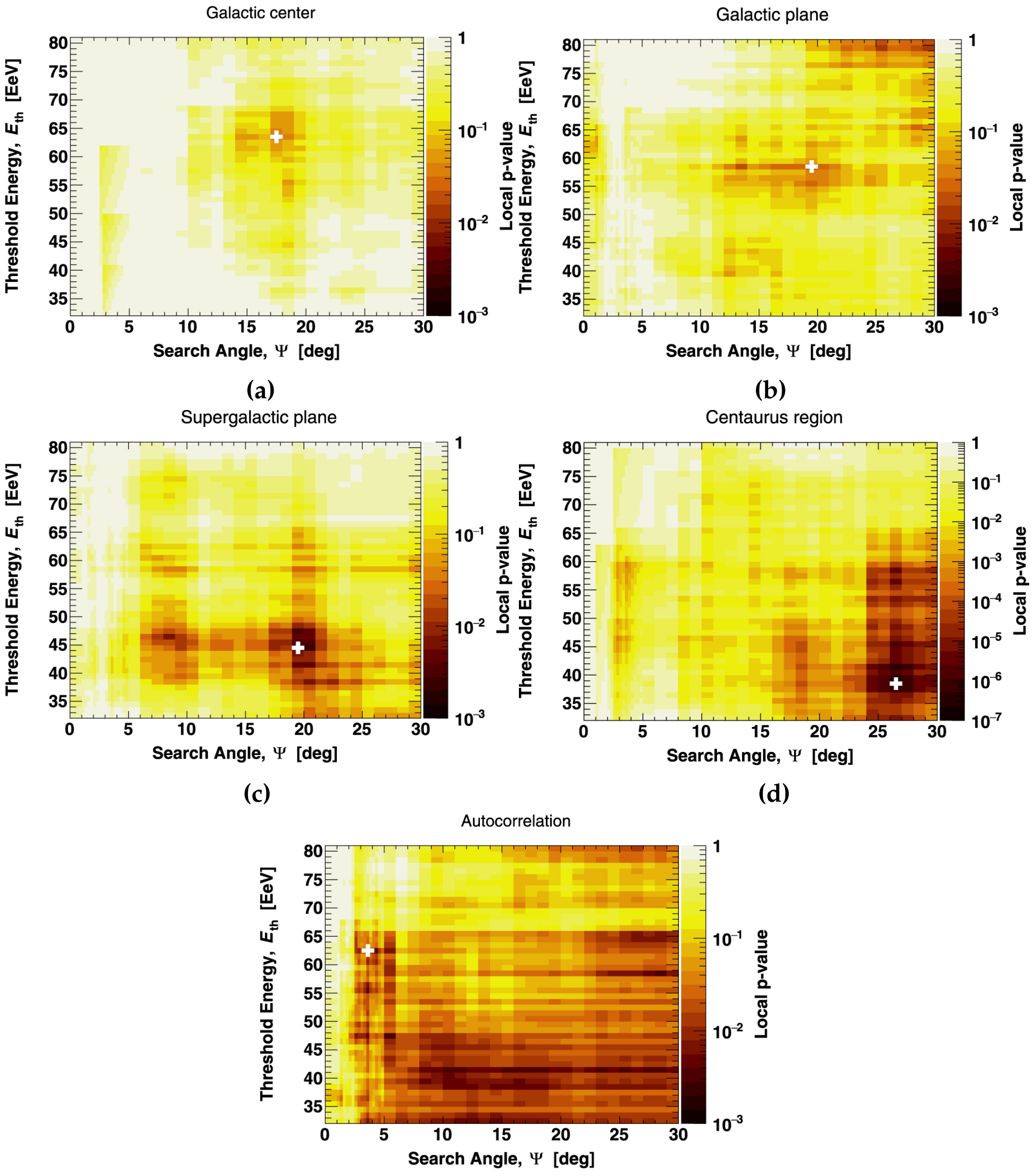


Figure 4.7: The angular power spectrum for the events in the energy bin 8 to 16 EeV (*left*) and above 8 EeV (*right*). In these energy bins, the results from the first harmonic analysis in right ascension allow for an exclusion of fluctuations of the isotropic scenario as source of the measured dipole structure with a significance of 5.7σ and 6.9σ, respectively. For more information about the Rayleigh analysis, see Table 4.1. The red line indicates the upper limit on the amplitudes with 99% C.L., taking into account exposure effects on the measurement of the amplitude C_ℓ^{rec} given a true amplitude C_ℓ. For more information and results on other energy bins, see Ref. [70].

for small and intermediate scales on [74], a very detailed report, which includes additional resources on the online version.



(a) (b) (c) (d) (e)

Figure 4.8: Local p-value as a function of the search angle, Ψ, and threshold energy,E_{th}.

4.4. CORRELATION TO CATALOGS

Through a likelihood-ratio test, the Pierre Auger Collaboration has tested different catalogs describing the set of possible sources of the observed UHECR anisotropies [74]. The known anisotropic large-scale distribution of matter, probed in the Two Micron All-Sky Survey (2MASS; [30]), was one of the probed catalogs. In this first scenario, the UHECR luminosity is assumed to be proportional to the stellar mass. A second scenario evaluated there consists of *starburst galaxies*, galaxies with a high star formation rate. There, the radio flux of the galaxies served as a proxy for the star formation rate, which is taken as proportional to the UHECR luminosity. The third catalog consisted of active galactic nuclei, AGNs, above a threshold flux in the hard X-ray emission, which served as a proxy for the UHECR flux. The last sample of possible sources, forming the fourth catalog, comprises radio galaxies and blazars from a Fermi-LAT catalog. The γ-ray flux is used there, assuming the UHECR luminosity is proportional to it. More detailed information, including the full list of sources in each catalog, can be found in [74].

In the likelihood-ratio analysis, the null hypothesis is the isotropic scenario. Each alternative scenario is evaluated via the flux expectation of each source catalog. This way, the test statistic, TS, is computed via

$$\mathrm{TS} = 2\ln(\mathcal{L}_1/\mathcal{L}_0) \;=\; 2\sum_i k_i \times \ln\frac{n^{\mathrm{H}_1}(\mathbf{u}_i)}{n^{\mathrm{H}_0}(\mathbf{u}_i)}; \tag{4.5}$$

with $\mathcal{L}_0$ and $\mathcal{L}_1$ the likelihood scores of the null and alternative hypothesis, respectively. In Eq. (4.5), n^{H_0} and n^{H_1} are the expected number of events of the models in the null- and alternative-hypothesis, respectively; while k_i is the count of measured events in the sky pixel i representing a direction $\mathbf{u}_i$.

The TS is maximized as a function of two free parameters, the search radius and the signal fraction, above successive energy thresholds. The maximum TS is obtained at roughly the same point of the parameter space for all four catalogs, with higher TS values for the starburst scenario. The best-fit parameters for the four scenarios are:

i. energy threshold $E_{\mathrm{th}} \sim 40\,\mathrm{EeV}$;

ii. Fisher search radius $\Theta = 14^\circ - 16^\circ$, which corresponds to a top-hat window of $\Psi = 23^\circ - 25^\circ$;

iii. signal fraction $\alpha \sim 10\%$.

It is remarkable to see that both the energy thresholds and search angles are consistent in the catalog search and in the overdensity searches around the supergalactic plane and the Centaurus region, depicted in Figs. 4.8c and 4.8d. In the overdensity search, the Centaurus A presents a post-trial p-value[3] of 3.0×10^{-5}, corresponding to a 4.0σ significance; while in the likelihood result for the starburst catalog, a post-trial p-value of 6.6×10^{-5} is found, corresponding to a 3.8σ significance.

A further step in the search for possible classes of sources of the most energetic cosmic rays is the combination of the arrival direction information with measurements of the energy spectrum and of the depth of maximum shower development, which is an indicator of the mass composition as discussed in Section 2.3. This approach, broadly referred to as *combined fit*, aims at constraining characteristics of the sources in the explored catalogs while regarding the effects of the propagation

---

[3] The p-value quantifies the probability of obtaining the observed results under the assumption of the null hypothesis being true. The post-trial p-value is computed as the fraction of isotropic simulations that have an equal or smaller probability under the same scan. [69].

of UHECR from source to Earth. These effects include, but are not limited to, interactions of the cosmic rays with the extragalactic background light and with the galactic and extragalactic magnetic fields. A common approach to simulating the propagation of the nuclei is to use Monte Carlo simulation codes, such as CRPropa and SimProp [76, 77].

In the combined-fit approach, source parameters such as the power-law spectral index γ, the rigidity cutoff R_{cut} defining the broken exponential tail of the spectrum, and the contribution of distinct nuclei in terms of their abundance in the overall flux, given as a fraction f_i^{nucleus}. A detailed description of the method can be found at [21]. The current state of the combined fit is an extension of previous analyses, which did not include the arrival direction information, as in [78, 20].

For each investigated scenario, the best-fit values for the parameters are those which minimize a deviance $D = -2\ln(L/L_{\text{sat}})$, with L and L_{sat} the likelihood of a model and of a saturated model that perfectly reproduces the data, respectively. The total deviance is a sum of two parts, one evaluating the agreement with the measured energy spectrum and the other with the maximum shower development.

It should be mentioned that this approach is mainly used to investigate intermediate-scale anisotropies. This implies that the dataset considered for the best-fit evaluation of the scenarios does not include the observed dipole structure[4]. Consequentially, the best-fit scenario found in the latest analysis is not expected to be extensible to lower energies while still reproducing the measurements reported by the Auger Collaboration.

The most recent findings reported in [21] indicate a preference for a scenario where the SBGs contribute to roughly 20% of the flux of CRs at 40 EeV, reaching a 1-sided significance of 4.5σ when considering the experimental systematic uncertainties. Additionally, the authors report that

> The total test statistic is dominated by a region in which two interesting source candidates reside, NGC 4945 and Centaurus A. [...] The contribution of Centaurus A to the total flux is estimated to be around 3% at 40 EeV [...] with a dominant nitrogen component reaching Earth above this energy. The signal fraction and blurring for the model with Centaurus A as a single source are in agreement with the ones obtained for NGC 4945 in the case of the SBG model.

Additionally, the authors report a dominance of the test statistic by the region around Centaurus A and NGC 4945, finding a source contribution of $\sim 3\%$ at 40 EeV to the overall flux in the single source and SBG scenarios, respectively.

4.5. FIRST APPROACH TO A MASS-SENSITIVE ANISOTROPY

Considering the extragalactic origin of UHECR, an effect of the galactic magnetic field on the trajectory of detected charged nuclei is unavoidable. The effect is larger for smaller-rigidity CR, provoking a deflection of its trajectory. The overall picture

---

[4] In [21], the authors take into account the arrival direction data above 16 EeV, and energy spectrum and shower depth distributions above 10 EeV.

gets more complex as thousands of CR arrive at Earth, following the steeply falling energy spectrum and with progressively heavier composition (which corresponds to higher atomic charges), possibly also from distinct sources at different distances to Earth.

As can be seen in Figs. 4.9 and 4.10 and discussed in Chapter 2, the galactic magnetic field affects the trajectories of charged particles up to very high rigidities. The lower the rigidity, the larger the shift in their trajectories.

In Fig. 4.9 cosmic rays with a rigidity of 20 EV had their trajectories simulated "backwards" from Earth to the Galactic/extra-galactic interface region to probe the effect of the galactic magnetic field on the observed arrival directions. Each filled circle corresponds to an arrival direction at Earth, while the open symbols indicate the location of the UHECR at this galactic-extra-galactic interface according to the magnetic field model, as annotated in the figure.

The impact of magnetic field deflections on a large-scale anisotropy is evaluated in Fig. 4.10 using the JF12 [79] galactic field model. In that figure, each open circles represents the direction of the maximum amplitude of an extragalactic dipole. The arrows originating from each location represents the location where the dipole pattern would point to, each segment of the arrow corresponding to the rigidity of the CR associated with that anisotropy: 32, 16, and 8 EV and the tip of the arrow for a rigidity of 4 EV. Interpreting this figure, one can argue that, under this field model a plethora of locations (outside of our galaxy) of the extragalactic dipole could give rise to the dipole direction measured at Earth, indicated with a '+' marker, within the 68% CL region shaded in gray. Additionally, the degree to which the anisotropy is "dissolved" due to the galactic magnetic field is also dependent on the extragalactic dipole location and on the rigidity of the CRs. The dipole amplitude measured at Earth, represented as a fraction of the original magnitude, is indicated by the color. Observing the tips of the arrows, for example, representative of CRs with rigidity of 4 EV, one can readily see that in some regions the color is lighter than in other regions of the galactic sky. A lighter color indicates that a smaller fraction of the original magnitude would be observed after the propagation of CR in galactic magnetic field.

A hypothesis that around the galactic plane an overall heavier composition of CR would be found when compared to far away from it has been evaluated in [80], and the most updated results of this analysis were reported in [81]. This is motivated by the expectation that, in a mixed composition scenario, the different mass components present at any given energy will be deflected in the galactic magnetic field to different degrees as they travel from their extragalactic sources to Earth [82]. Therefore the lighter particles at a given energy would be localized closer to their extragalactic sources, and therefore a region that does not include such sources would be dominated by the isotropized heavier component of the UHECR flux.

The current method for this analysis consists of the definition of *on-* and *off-plane* regions through the scanning of a sample of the data set, also defining the energy threshold to be considered for the analysis. The best parameters are those that maximize the difference in the composition in the two regions, evaluated via a two-sample homogeneity test. Then, the threshold in energy and the boundary of the on- and off-plane regions are applied to the full hybrid data set (see Section 3.3

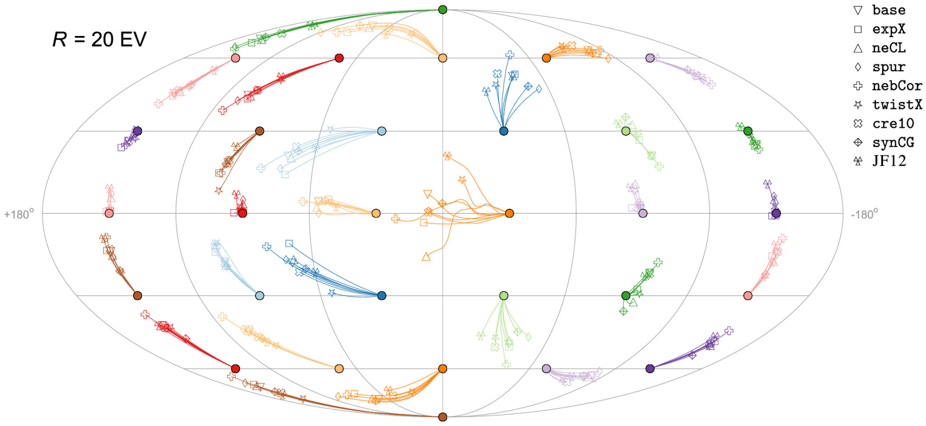


Figure 4.9: The deflection of UHECR at a fixed rigidity of 20 EV for eight model variations derived in Ref. [15] and for the JF12 model [79]. The CR were back-tracked from the filled circle positions, representing the arrival direction at Earth, to the edge of the galaxy where the open symbols represent the location of the UHECR at this galactic-to-extra-galactic interface.

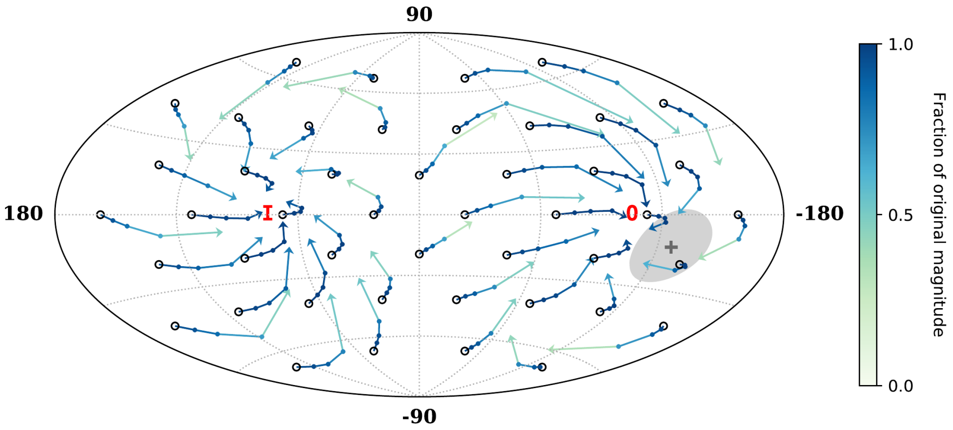


Figure 4.10: Illustration of the effect of the GMF on an extragalactic dipolar distribution of the UHECR flux and its rigidity dependence. Each black circle represents the direction of the extragalactic dipole, and the arrows originating from it indicate where the dipole component would point to when measured at Earth, when considering the JF12 regular field. The points in the arrow indicate distinct CR rigidity: 32, 16, and 8 EV, and the tip of the arrow, 4 EV. In addition to the shift in the direction of the extragalactic dipole, another effect is the dampening of its amplitude, indicated by the color of each segment of the arrow. The '+' marker indicates the direction of the measured dipole for energies above 8 EeV, while the gray area around it indicates the 68% CL region. The inner and outer spiral arms of our galaxies are indicated by the 'I' and 'O' labels, respectively.

for reference), and the composition is investigated in the regions. The difference between the two is evaluated through the first and second momenta of the depth of maximum shower development, $X_{\max}$.

Because of the high accuracy in the energy and $X_{\max}$ measurements in hybrid mode, it is possible to evaluate the energy dependence of the on-/off-plane difference in the proxy for mass composition. The latest results indicate that there is no longer strong support for a correlation between $X_{\max}$ and the galactic plane with the hybrid data alone, in contrast to previously reported [80, 81]. The reduction of the signal could be understood as a consequence of the changes in the reconstruction procedures, or in the atmospheric corrections. For further details, we recommend Ref. [81] and references therein.

Regardless of the differences between [81] and [80], the rate of signal growth indicates that a 5σ significance could not be reached with this analysis before 2035 when the decommissioning of the Pierre Auger Observatory is planned. The slow yearly increase of data in the hybrid mode is a decisive factor in that matter, thus is crucial to introduce an analysis with similar method but carried on SD-only dataset, far richer in statistics. This is under preparation within the Collaboration.

It is indubitable that adding the composition information to the anisotropy searches is of great interest. The community will benefit of such mass-dependent analysis, possibly helping understand further the processes occurring during CR transport, or even the types of sources contributing to the overall CR flux and to the measured anisotropies in different scales. Inspired by these remarkable findings, this thesis is a report on the first search of a composition signature on the observed dipole structures.

4.6. SUMMARY

In this chapter, I summarised the recent findings and developments in the search for anisotropies in UHECR's arrival directions. The ground method for the large-scale anisotropy search, in particular the dipole component of the overall flux of CR, was described in detail. The current knowledge is that fluctuations in an isotropic distribution of the arrival direction of UHECR is excluded as source of the measured anisotropy with large significances for events with energies above 8 EeV, and for events with energies between 8 and 16 EeV. An increase in the dipole amplitude with energy has been reported, although the steeply falling number of events at the highest energies prevent strong statements about the dipole component as in the energy ranges mentioned above.

For the most energetic events, however, smaller-scale structures become more significant, and the current status of over-density searches is not very far from the discovery level of 5σ statistical significance. Great effort has been made to understand the possible class of sources of the observed anisotropies, combining information from different measurements by the Auger Collaboration. Although a preference for the SBG class of sources can be indicated, the large number of unknowns regarding the propagation towards Earth, and even the lack of an accurate mass composition information of the detected CR, prevent a more elaborate conclusion. The Centaurus region stands as the most significant localized over-density of events, and is the scenario with larger likelihood in the combined fit approach, which includes information from the measured spectrum, composition and arrival direction distributions.

Although no strong statement can yet be drawn from the recent developments, it is clear that investigating the composition-dependence of the measured anisotropies in different scales will foster further development. Possible progress includes narrowing down the possible source scenarios and, hopefully, constraining the current models of the galactic magnetic field.

In the following chapter, I introduce the unprecedented search for a rigidity-dependent dipole signature in the arrival directions of CR.

CHAPTER V

THE AUGER-COMPATIBLE SIMULATION DATA SET

5.1. PREAMBLE

The main goal of my research project was to investigate signatures of a mass dependence of the observed large-scale anisotropy. We decided to do so by dividing the dataset into two composition-distinct subpopulations. Taking a conservative approach, we opt to define the subpopulations according to simulation expectations rather than a data-driven approach. In doing so, we can avoid an *a posteriori* penalization of the results due to excessive scanning of the parameters.

The composition of CRs and its evolution with energy are necessary information to reach the goal. A composition model is needed since no direct measurement of the CR composition is possible for the ultra-high range of energies. A continuous model is also favored when creating a library of simulated events compatible with the Auger Observatory measurements. To create such a library, a sampling from a pre-existing library will be performed; thus, the preference for continuous functions in defining the composition model.

Because the X_{max} is closely related to the atomic mass A of the cosmic ray initiating the extensive air shower, it has been used as a proxy for identifying the composition of the UHECRs. As reported in Refs. [83, 26], the fractional contribution of four representative nuclei can be estimated by matching the X_{max} expectations from hadronic interaction models for each assumed nuclei and the measurements.

In Section 5.3, we develop a four-component continuous, empirical model of the mass composition of CRs based on measurements by the Auger Collaboration but extended to higher energies. The *extended Auger mix*, the composition model adopted in this analysis, also incorporates a mass-ordering of the fractional contributions to the overall CR flux.

We construct a cumulative distribution function (CDF) from the energy spectrum described in Ref. [71] to reproduce the steeply falling energy spectrum. More details will be presented in Section 5.2.

5.2. SPECTRUM

The measurement of the spectrum of cosmic rays has allowed for the identification of two prominent features, namely the *ankle*, at $5{\times}10^{18}$ eV and the suppression of the CR flux at the highest energies, around $5{\times}10^{19}$ eV [84, 72, 71]. The suppression was first observed by the HiRes experiment [85], and confirmed in the same year by the Auger Collaboration, with twice as large exposure [86]. The features have also been later observed by the Telescope Array collaboration [87]. The latest measurements reported by the Auger Collaboration indicate additional features above the ankle region [72].

It is one of the current open questions whether changes in the overall composition of the CR flux can be related to the features observed in the spectrum [83, 88]. The relation could be a consequence of composition-dependent maximum energy of the acceleration mechanism or of the source, possibly in combination with transport effects [89].

As described in Section 4.4, the combined fit analyses takes into account the dependence of the measurements at Earth on the source information and the transport effects. When creating a simulation library, it is crucial to reproduce both the CR composition model and the observed features of the UHECR spectrum.

The spectral features described in Ref. [90] were reproduced in the creation of the simulation library. The results reported in Ref. [90] are in agreement with the previously published analysis for the highest energies [84, 71, 72], indicating the stability of the reported features. In Refs. [71, 84], the authors also reported the stability of the spectral features in different declination bands of equal exposure, discussing the compatibility with expectations from the observed dipole.

The function describing the spectrum of UHECRs is

$$J(E) = J_0 \left(\frac{E}{10^{18.5}\mathrm{eV}}\right)^{-\gamma_1} \prod_{i=1}^{3} \left[1 + \left(\frac{E}{E_{ij}}\right)^{1/\omega}\right]^{(\gamma_i-\gamma_j)\omega}, \tag{5.1}$$

where $j \coloneqq i+1$ and $\omega = 0.05$. The factor ω controls the width of the energy intervals over which the slope transitions occur. The three features reported in Refs. [71, 84] are introduced: the ankle is located at 5 EeV, the instep at 13 EeV, and the suppression at 46 EeV. Details of the spectral features are presented in Table 5.1.

A CDF of the spectrum is constructed for the energy range of interest in the dipole analysis (above 4 EeV) to create an Auger-compatible simulation library. The first step is to draw an energy value of the CR using the Monte Carlo method. Subsequently, another random number is drawn to match the CR to the specific nuclei according to the fractional contribution of each element in the given energy, as expected from the adopted composition model. In this way, events are selected from an existing library of simulated showers, creating a simulation library compatible with the measurements of the Auger Collaboration.

Table 5.1: Best-fit parameters for Eq. (5.1) describing the spectrum of CR, as reported in Refs. [71, 84]. The values, including statistical and systematic uncertainties, were obtained with a dataset of 215 030 events recorded between 1 January 2004 and 31 August 2018. The parameters E_{ij} indicate the energy at which a transition between regimes i and j is observed, indicated by a change of the spectral index γ_i to γ_j.

parameter	value $\pm\sigma_{\text{stat}} \pm \sigma_{\text{sys}}$
$J_0/10^{-18}\,\text{km}^{-2}\text{sr}^{-1}\text{yr}^{-1}\text{eV}^{-1}$	$1.315 \pm 0.004 \pm 0.400$
γ_1	$3.29 \pm 0.02 \pm 0.10$
γ_2	$2.51 \pm 0.03 \pm 0.05$
γ_3	$3.05 \pm 0.05 \pm 0.10$
γ_4	$5.1 \pm 0.3 \pm 0.1$
$E_{12}/10^{18}$ eV (ankle)	$5.0 \pm 0.1 \pm 0.8$
$E_{23}/10^{18}$ eV	$13 \pm 1 \pm 2$
$E_{34}/10^{18}$ eV (suppression)	$46 \pm 3 \pm 6$

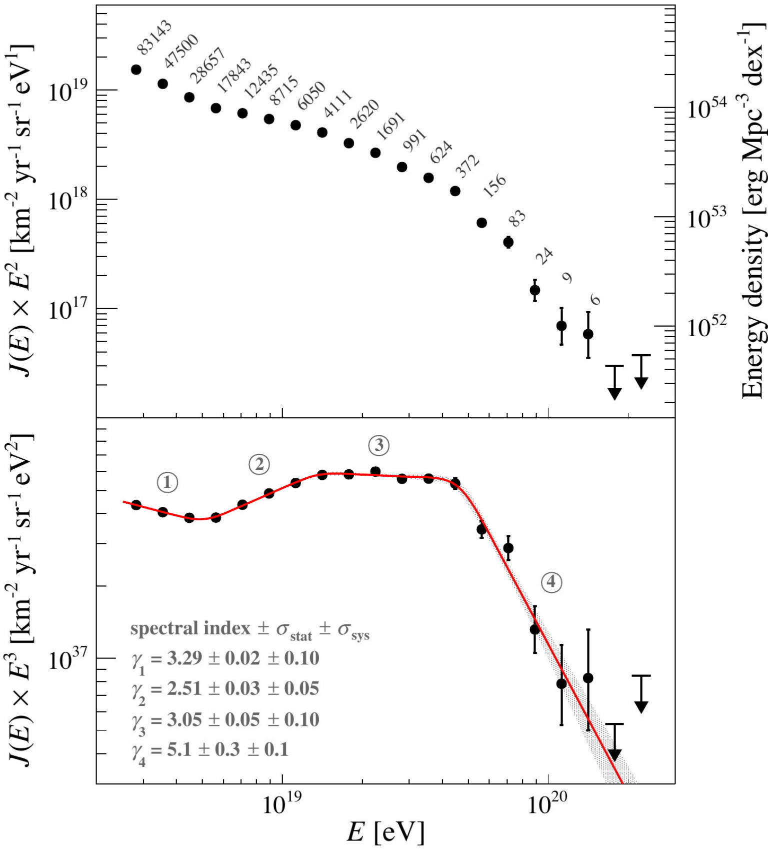


Figure 5.1: Energy spectrum and observed features. On the *top* panel, the flux multiplied by E^2 is shown, with annotated number of events in each energy interval of size 0.1 in $\lg E$, as detected between 1 January 2004 and 31 August 2018. On the *bottom* panel, the differential energy spectrum is multiplied by E^3, making the transition between the four regions annotated more visible. The spectral indexes in each region, reported in Table 5.1, are also annotated in the graph. Reproduced from Ref. [71].

5.3. FOUR-COMPONENT COMPOSITION MODEL

5.3.1. A FIT TO FD MEASUREMENTS

The cosmic-ray composition can be estimated by fitting the $X_{\max}$ distributions with the prediction from air shower simulations according to a hadronic interaction model [91, 26]. Using a binned likelihood fit, the authors evaluate the probability of obtaining a worse fit with the predicted $X_{\max}$ distributions than with the actual data. A novel approach was introduced in Ref. [26], using a Markov Chain Monte Carlo method, which avoids the global fit from being stuck in local minima.

In the composition fit of the $X_{\max}$ distribution with hybrid SD and FD measurements, it is a standard practice to perform it in terms of four nuclei components which are approximately equally spaced in $\ln A$, namely proton, indicated by p (or H), Helium (He), Nitrogen (N), and Iron (Fe). The latest results, reported in Refs. [26, 81], are reproduced in Fig. 5.2. Two hadronic interaction models were evaluated.

Due to the limited statistics of the FD, the last energy bin where an evaluation of composition is possible comprises all events with energies $\lg(E/\mathrm{eV}) > 19.6$. The analysis is thus not sensitive to the evaluation of the contribution of high energy components that lie between N and Fe, as is often discussed in the combined fit analysis described in Section 4.4.

The overall progression of the composition towards heavier nuclei with increasing energy is visible in Fig. 5.2 and agrees with Fig. 5.3. In Ref. [93], the authors investigated the evolution of CR composition according to a Peters cycle. In that regime, the maximum energy scales linearly with the charge Z of the CR (see Section 2.1). The authors compared this scenario with alternative astrophysically-motivated scenarios. The findings indicate that other scenarios better describe the spectral and composition features observed by the Auger Collaboration above the ankle of the spectrum.

The ordering in mass of the CR composition, according to the expectations from FD measurements of the depth of shower maximum $X_{\max}$, has been discussed within the Collaboration [94]. When compared to a charge ordering of the fractional contributions, corresponding to the mentioned Peters cycle, the evolution of CRs composition according to the atomic mass is in a better agreement with the measurements, also indicated by Ref. [93]. We opt to incorporate the apparent composition-ordering of the fractional contributions into the overall CR flux in the *extended Auger mix*, used as the composition model henceforth.

5.3.2. EXTENDED AUGER MIX

The proposed extended Auger mix empirically describes the contribution of four components in the total flux of CR observed in hybrid mode. Because the observed fractions present a rise followed by a fall, as seen in Fig. 5.2, we use a Gaussian function $\mathcal{N}(x;\mu,\sigma)$ as a simple descriptive model, where μ is the center and σ the width of the Gaussian. The choice to adopt a continuous function instead of the measured fractions favors and facilitates the adoption of the extended Auger mix when generating an Auger-compatible simulated dataset from available simulation

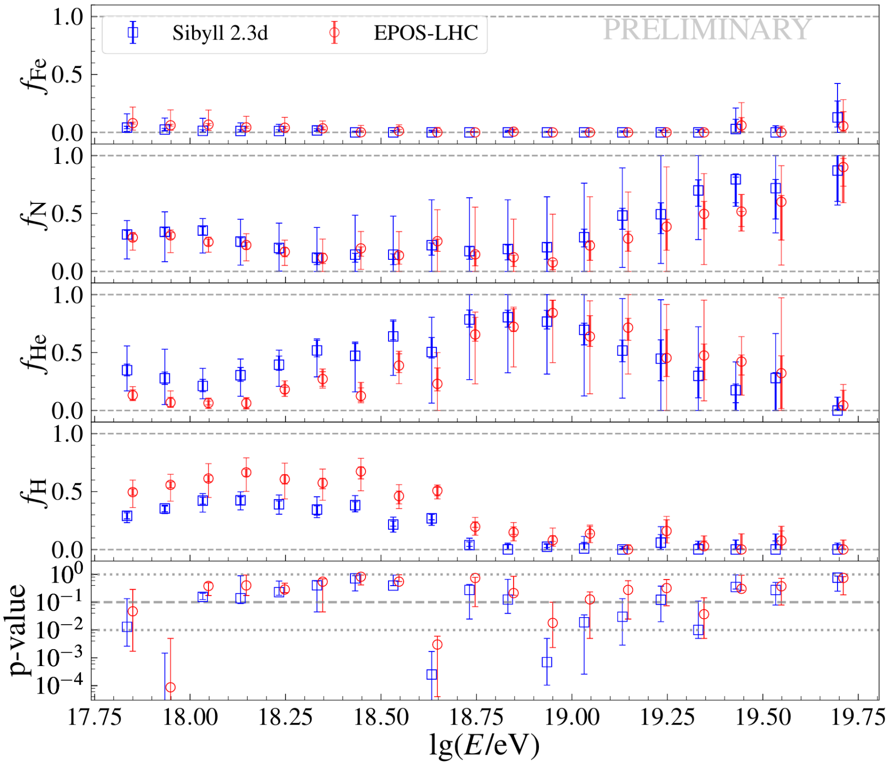


Figure 5.2: Composition fit for four elemental mass groups (top four panels). The fits presented in Ref. [26] reproduces the X_{max} distributions by adding the contribution of four representative nuclei: hidrogen (H), helium (He), nitrogen (N), and iron (Fe). The error bars denote statistical (inner cap) and total (outer cap) uncertainties. The bottom panel shows the p-values of the fit. Results are provided for two hadronic interaction models: Sibyll 2.3d in blue, and EPOS-LHC in red. Reproduced from Ref. [26].

libraries. More so, the extension to higher energies where the composition inference is not yet available can be of high interest for exploring the potentialities of Auger Phase I and of AugerPrime in searches for, for example, mass-enhanced anisotropies.

We adopt Gaussian functions to describe the component fractions as a function of energy E. The Gaussian functions are in a general form written as

$$a\,\mathcal{N}(\lg(E/\mathrm{eV}); m, w) + c. \tag{5.2}$$

Sometimes, the constant offset c and/or the center m are not fitted but are calculated or set to zero.

Because there are two peaks clearly distinguishable in the CNO component, we exceptionally allow for two such Gaussians. In addition, a constant factor is also allowed in CNO, as the published fraction fit seems to not be 0 anywhere in the energy range considered. Alternatively, a single Gaussian distribution is fitted to each p and He component. Finally, the Fe component is set as the remaining fraction, as the Auger mix is limited to four components only; therefore

$$f_{\mathrm{Fe}}(E) = 1 - \sum_{i\in\mathcal{P}} f_i(E), \tag{5.3}$$

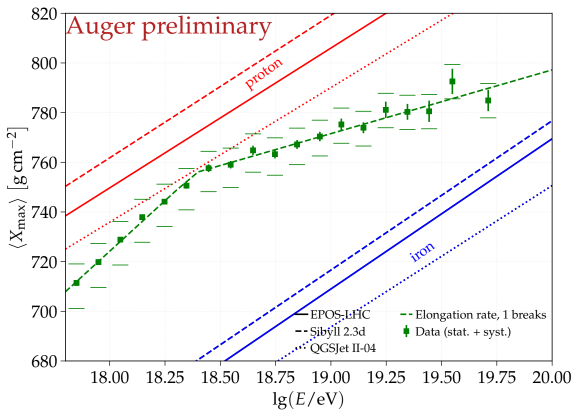


Figure 5.3: First (*left*) and second (*right*) moments of the $X_{\max}$ measurements with FD in hybrid events from years 2004 to 2021 [83]. Also shown are the predictions from different hadronic interaction models for primary proton (red) and iron (blue).

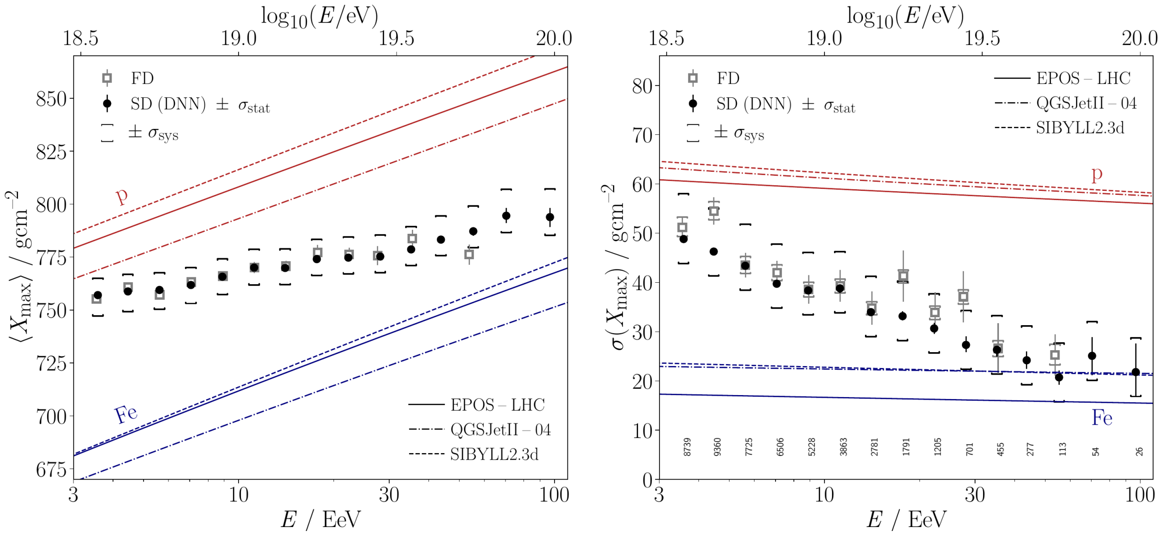


Figure 5.4: First (*left*) and second (*right*) moments of the $X_{\max}$ measurements. Square markers represent measurements with FD in hybrid events from 2004 to 2017 [92], while black markers indicate the DNN predictions from SD measurements recorded between 2004 and 2018 [88]. Also shown are the predictions from different hadronic interaction models for primary proton (red) and iron (blue).

where $\mathcal{P} = \{\mathrm{p}, \mathrm{He}, \mathrm{CNO}\}$.

To perform the fit, we define a loss function Q^2 to be minimized. In this loss function, we evaluate the deviation between the data points and the parameterized distribution for each component individually in terms of the asymmetric statistical uncertainty from the MCMC fitting procedure [26]. Additionally, there is a penalization term whenever the Fe component gets assigned either a negative or larger-than-one value to avoid such cases. The procedure is performed for the fractions obtained in Ref. [26] using EPOS-LHC and Sibyll 2.3d models. From the fitted

Gaussian parameters, detailed in Tables 5.2 and 5.3, we bring attention to the center of the proton component. Because we expect a mass (or charge) ordering, the center of the distributions of the other components are a simple function of the proton on.

We have explored two possibilities: (a) a mass or (b) a charge ordering of the maxima of fractional contributions of each component. In the mass-ordering case (a), we thus have

$$E^{\max}(A, Z) = A\, E^{\max}_{\text{proton}} \quad \Rightarrow \quad m(A, Z) = m^{A}_{\text{proton}} + \lg A, \tag{5.4}$$

while in the charge-ordering case (b), we have

$$E^{\max}(A, Z) = Z\, E^{\max}_{\text{proton}} \quad \Rightarrow \quad m(A, Z) = m^{Z}_{\text{proton}} + \lg Z. \tag{5.5}$$

With this simple description and the knowledge of a composition progression of component maxima, extending the Auger mix to the higher energies is straightforward, including the assumption of no contribution of elements heavier than Fe. We note that evaluating the iron contribution at higher energies depends on further measurements. We stand, however, on the assumption of a mass progression and interpret this increasing iron fraction at the highest energies as a possible contribution of elements heavier than the intermediate CNO group.

A comparison of the locations of maxima of each component in the two scenarios described by Eqs. (5.4) and (5.5) is depicted in Fig. 5.5. We bring attention to the small distance in the locations in the two scenarios due to the semi-universal relation between A and Z.

The results of the fit of the Gaussian function to the fractional contribution of each component for both hadronic models investigated are detailed in Table 5.2 and Fig. 5.6, for the A-ordering described in Eq. (5.4), and in Table 5.3 and Fig. 5.7, for the Z-ordering described in Eq. (5.5).

We find that a **mass** ordering of the four components is favored for both hadronic models considered. In this scenario, in both hadronic-model fractions, the Gaussian centers are very similar. However, in the Sibyll case the Helium component dominates over a large range of energies with a very broad distribution. In the EPOS-LHC case in the A-ordering scenario, the width of all components is very similar, indicating a more mixed composition than in the alternative hadronic-model case.

The best-fit results for each ordering scenario and hadronic model can be compared qualitatively. We see in Figs. 5.8 and 5.9 the alignment of the four components either in terms of the mass A or of the charge Z. It is notorious that the broadening of the He component in the charge-ordering scenario is in apparent disagreement with the fitted fractions depicted with the circle and square markers.

While we argue that mass-ordering seems to be favored, it is also noteworthy to mention that the limitation of four-component fit may have a large role in the evaluation of both scenarios.

In agreement with what has been discussed in Refs. [93, 94], we adopt the mass-ordering of the components as defined by Eq. (5.4). The parameters used to define the four-component continuous function, named *extended Auger mix*, are those in

Table 5.2, related to the EPOS-LHC hadronic model. The choice of the hadronic model is motivated purely by the availability of cascade simulations with that model in the energy range of interest and in all four components considered here.

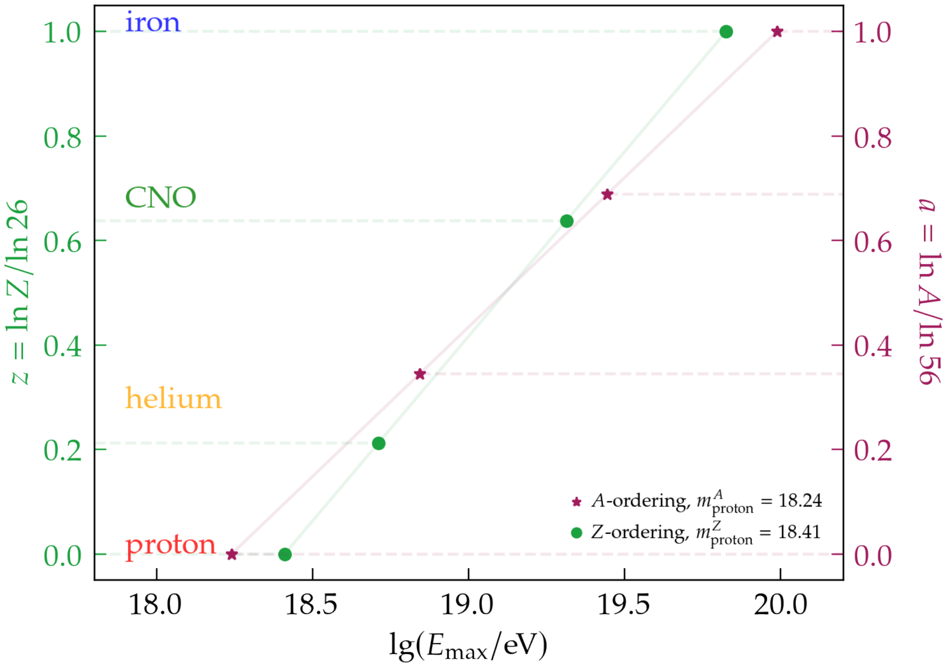


Figure 5.5: Illustration of A-ordering and Z-ordering, following the fits to the fraction of each of the four components contributing to the total UHECR flux. Results are based on EPOS-LHC model, with parameters detailed in Tables 5.2 and 5.3.

Table 5.2: Best-fit parameters for the four-component fractions based on two hadronic models, given an A-ordering of the composition fractions. Values in [ ] brackets are not fitted, but calculated from the proton-fitted value with $m(A, Z) = m^{A}_{\text{proton}} + \lg A$. A dash denotes that this parameter is fixed to zero. The CNO component has two Gaussian contributions and one constant.

	EPOS-LHC					Sibyll 2.3d				
primary	m	w	a	c	Q^2/n_{df}	m	w	a	c	Q^2/n_{df}
proton	18.24	0.45	0.74	—	0.16	18.19	0.38	0.39	—	0.06
helium	[18.85]	0.4	0.54	—	0.19	[18.79]	0.57	0.93	—	0.18
CNO	17.71	0.35	0.4	0.0	0.23	17.83	0.22	0.2	0.1	0.21
	[19.45]	0.34	0.61	—		[19.39]	0.34	0.37	—	
iron	—	—	—	—	0.09	—	—	—	—	0.15

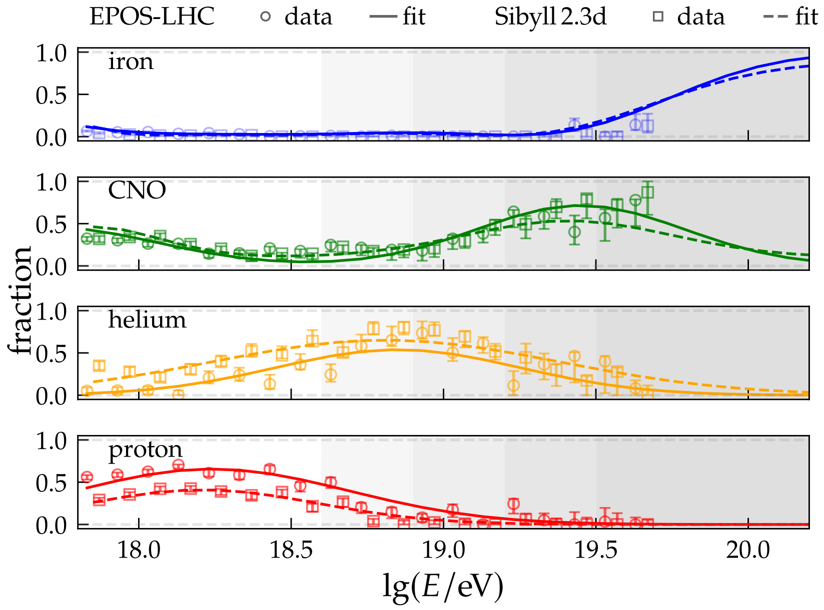


Figure 5.6: The composition-fraction fit using the EPOS-LHC (circles) and Sibyll 2.3d (squares) hadronic interaction models [26]. The continuous function for each component fitted to the EPOS-LHC fractions is shown with full lines. The continuous function for each component according to the Sibyll 2.3d fractions is shown with dashed lines. The maxima of distributions follow the mass ordering with respect to proton as $m(A, Z) = m^{A}_{\text{proton}} + \lg A$. The CNO component has an additional Gaussian at lower energies. The iron distribution is set as the remaining fraction of events. Parameters of the fitted distributions are displayed in Table 5.2.

Table 5.3: Best-fit parameters for the four-component fractions based on two hadronic models, given a Z-ordering of the composition fractions. Values in [ ] brackets are not fitted, but calculated from the proton fitted value as $m(A,Z) = m^{Z}_{\text{proton}} + \lg Z$. A dash denotes a parameter that is fixed to zero. The CNO component has two Gaussian contributions and one constant.

	EPOS-LHC					Sibyll 2.3d				
primary	m	w	a	c	Q^2/n_{df}	m	w	a	c	Q^2/n_{df}
proton	18.41	0.55	0.77	—	0.67	18.38	0.37	0.33	—	0.38
helium	[18.71]	0.8	0.68	—	0.6	[18.68]	0.83	1.15	—	0.36
CNO	17.78	0.25	0.2	0.08	0.19	17.83	0.26	0.27	0.03	0.32
	[19.31]	0.31	0.35	—		[19.28]	0.37	0.43	—	
iron	—	—	—	—	0.2	—	—	—	—	0.19

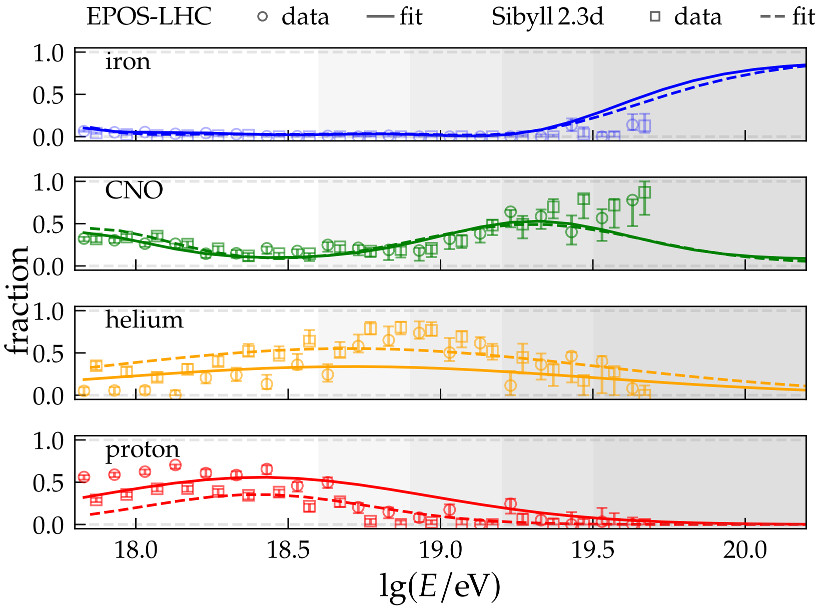


Figure 5.7: The composition-fraction fit using the EPOS-LHC (circles) and Sibyll 2.3d (squares) hadronic interaction models [26]. The continuous function for each component fitted to the EPOS-LHC fractions is shown with full lines. The continuous function for each component according to the Sibyll 2.3d fractions is shown with dashed lines. The maxima of distributions follow the charge ordering with respect to proton as $m(A,Z) = m^{Z}_{\text{proton}} + \lg Z$. The CNO component has an additional Gaussian at lower energies. The iron distribution is set as the remaining fraction of events. Parameters of the fitted distributions are displayed in Table 5.3.

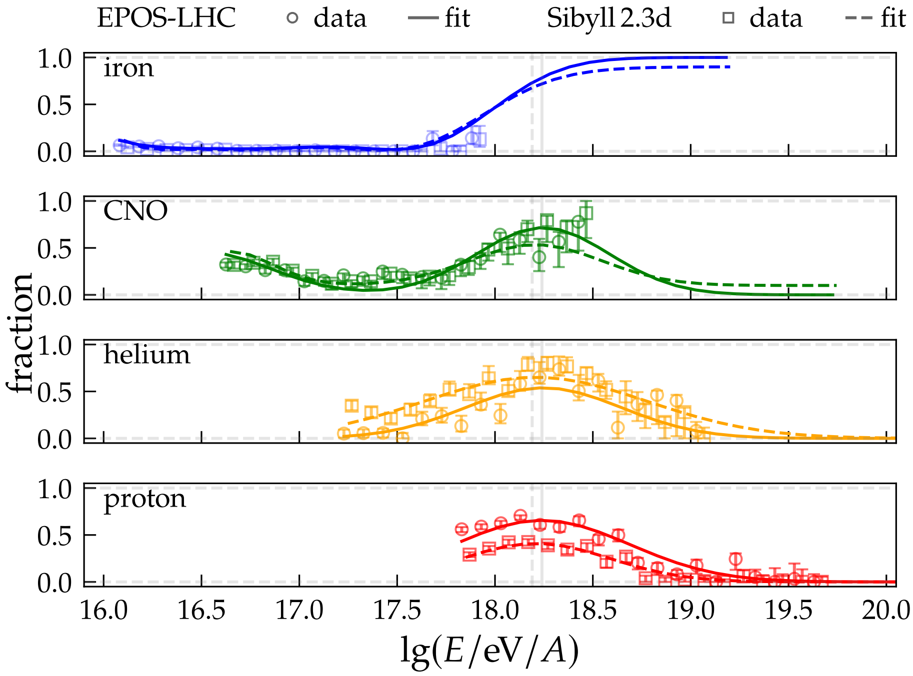


Figure 5.8: Illustration of A-ordering, following the fits to the fraction of each of the four components contributing to the overall UHECR flux. Results are based on EPOS-LHC and Sibyll 2.3d models, with parameters detailed in Table 5.2.

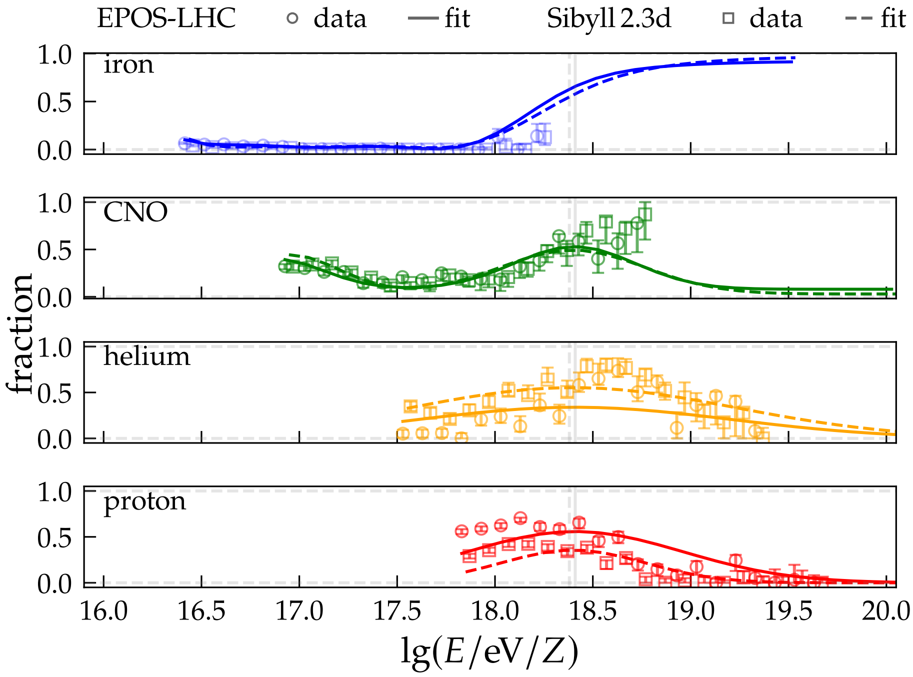


Figure 5.9: Illustration of Z-ordering, following the fits to the fraction of each of the four components contributing to the overall UHECR flux. Results are based on EPOS-LHC and Sibyll 2.3d models, with parameters detailed in Table 5.3.

5.4. COMPATIBILITY BETWEEN SIMULATION LIBRARY AND AUGER MEASUREMENTS

The simulation data set is created from a library of EPOS-LHC CORSIKA showers and reconstructed using the Universality module of $\overline{\text{Off}}\underline{\text{line}}$. The initial library follows an E^{-1} spectrum and has a nearly identical number of events for all components. Roughly 11 000 events are available for each component with energies between $10^{18.6}$ and $10^{20.3}$ eV, and zenith angles between 0° and 50°. The maximum zenith angle was smaller than the one adopted in the data analysis. This is a reflection of the dynamic evolution and improvement of the Universality module in parallel to the development and test of the proposed analysis. This point will be further discussed in the next chapter.

The simulation library was created following the unfolded spectrum described in Eq. (5.1). However, as discussed in Ref. [84], a correction factor c is applied to the raw spectrum to account for the detector effects. The correction factor is smaller for lower energies, indicating larger differences between raw and unfolded spectra. At energies larger than $E = 4{\times}10^{19}$ eV, the factor plateaus at $c \approx 0.97$.

Additionally, the unfolded spectrum is obtained considering only 6T5 events, while in the standard large-scale anisotropy search discussed in Section 4.2, both 6T5 and 5T5 events are used (see for reference Sections 3.3 and 4.2). The energy bias in 5T5 events has been evaluated, and dependencies with both zenith and azimuth angles were internally reported to the Collaboration [95]. The effect is larger at lower energies, of a couple of percent when considering events above 4 EeV, and below 1% for those above 8 EeV.

Considering a spectrum in the form $\mathrm{d}N/\mathrm{d}E \propto E^{-\gamma}$, we obtain that

$$\frac{\Delta N}{N} \propto (1-\gamma)\frac{\Delta E}{E}, \tag{5.6}$$

and can estimate that, for $\gamma = 3.29$ as reported in Table 5.1 below the ankle, the relative difference in the number of events above an energy E due to a systematic bias of $\Delta E/E \approx 2\%$ will be $\Delta N/N \approx -4.5\%$. To illustrate this effect, a band of ±5% is represented in green in the top panel of Fig. 5.10.

Comparisons between the Monte Carlo values of the Auger-compatible simulation library and the measurements with the FD are shown in Figs. 5.11 and 5.12. An overall agreement within the systematic uncertainties is seen, especially for lower energies.

The variance in the $\ln A$ distribution is an indicator of the composition mixing, with zero implying a pure composition beam. In Fig. 5.12, there is an indication that the used simulation library has a higher composition mix than the measurements according to the expectations from the EPOS-LHC hadronic model. We note, however, that the computation of $\langle \ln A \rangle$ and $\sigma(\ln A)$ in Fig. 5.3 takes into account shower-to-shower fluctuations and an energy-dependent parameter, while the one presented for the simulation library is computed in terms of the *true* A value, accounting for the contribution of each of the four components according to the composition model.

The compatibility of the moments in the X_{max} distribution between simulations and data is competitive to the ones obtained in the combined fits mentioned in

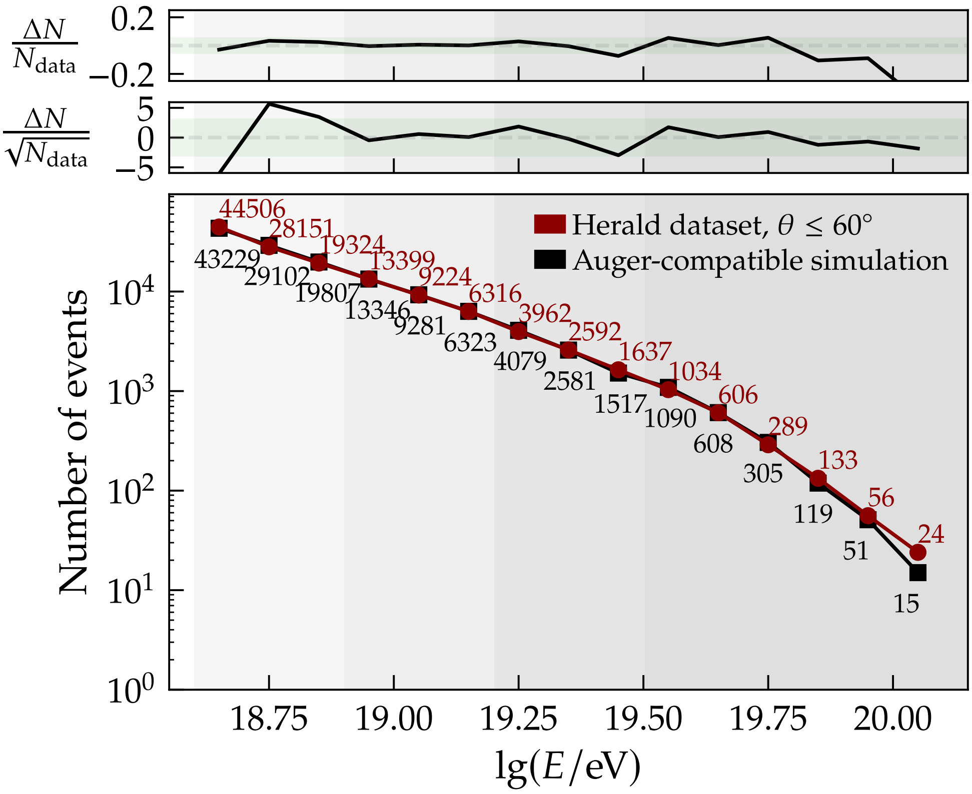


Figure 5.10: Number of events in Auger-compatible simulation library and in Herald dataset of events with zenith angle $\theta \leq 60°$. For the simulation library, E stands for the Monte Carlo energy; while for the Herald dataset, E represents the measured energy. Gray bands indicate the large energy bins used in the dipole analysis described in Section 4.2. Top panels indicate the difference in the number of events of the measured and simulated data sets relative to the number of measured events (*top*) or to the square root of that number (*mid*). The green bands indicate a 5% (*top*) or 3σ (*mid*) deviation.

Section 4.4 (see Refs. [21, 20]), even though the present approach is simpler both in terms of source assumptions and in terms of possible contributing components to the UHECR flux.

In Fig. 5.13, we present the decomposition of the fitted spectrum to the four contributing primaries, according to the extended Auger mix model. It is also possible to compare it with the results in the combined fit, where a qualitative agreement is noticeable. However, in the combined fit approach the contribution of distinct mass groups is evaluated, in contrast to the four components evaluated here.

The systematic effects on the measurement of the relevant quantities will be discussed in the next chapter.

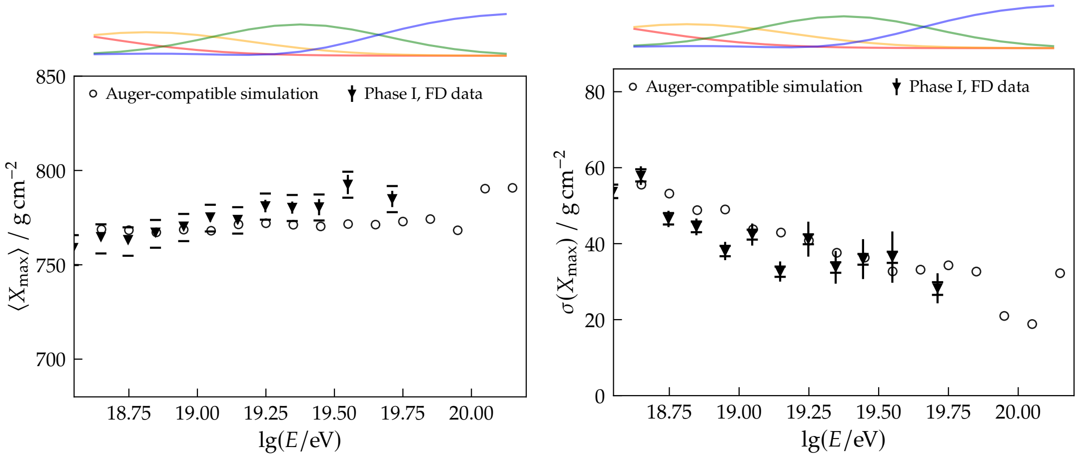


Figure 5.11: First (*left*) and second (*right*) moments of the MC X_{max} of the events in Auger-compatible simulation library and in the dataset of FD events reported in Ref. [83]. For the simulation library, E stands for the Monte Carlo energy; while for the FD dataset, E represents the measured energy. Statistical and systematic uncertainties are represented by the vertical line and the caps in each FD data point, respectively. The top panel schematically indicates the fractional contribution of the four components as a function of energy, as presented in Fig. 5.6 and Table 5.2 for the EPOS hadronic model.

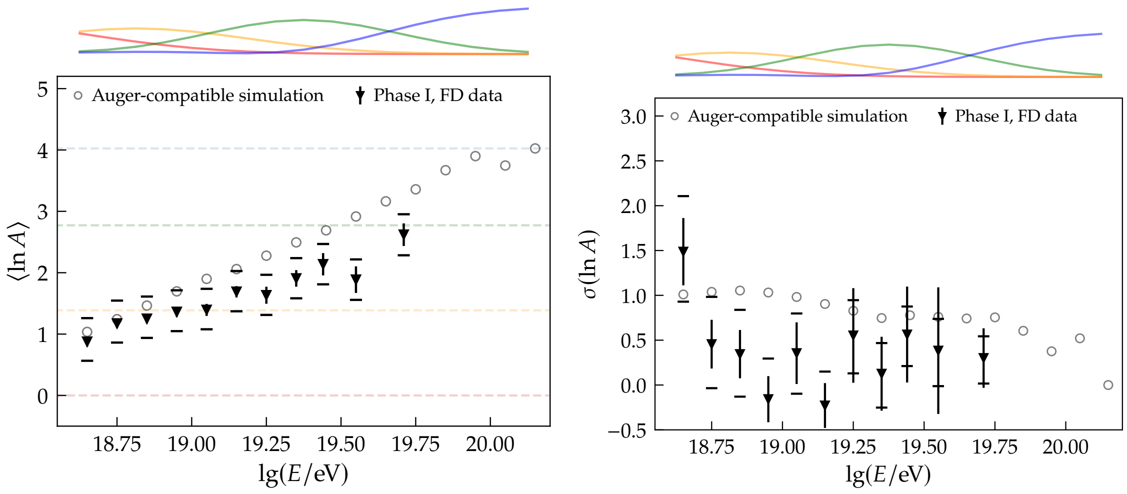


Figure 5.12: First (*left*) and second (*right*) moments of the MC $\ln A$ of the events in Auger-compatible simulation library and in the dataset of FD events reported in Ref. [83]. For the simulation library, E stands for the Monte Carlo energy; while for the FD dataset, E represents the measured energy. Statistical and systematic uncertainties are represented by the vertical line and the caps in each FD data point, respectively. The top panel schematically indicates the fractional contribution of the four components as a function of energy, as presented in Fig. 5.6 and Table 5.2 for the EPOS hadronic model.

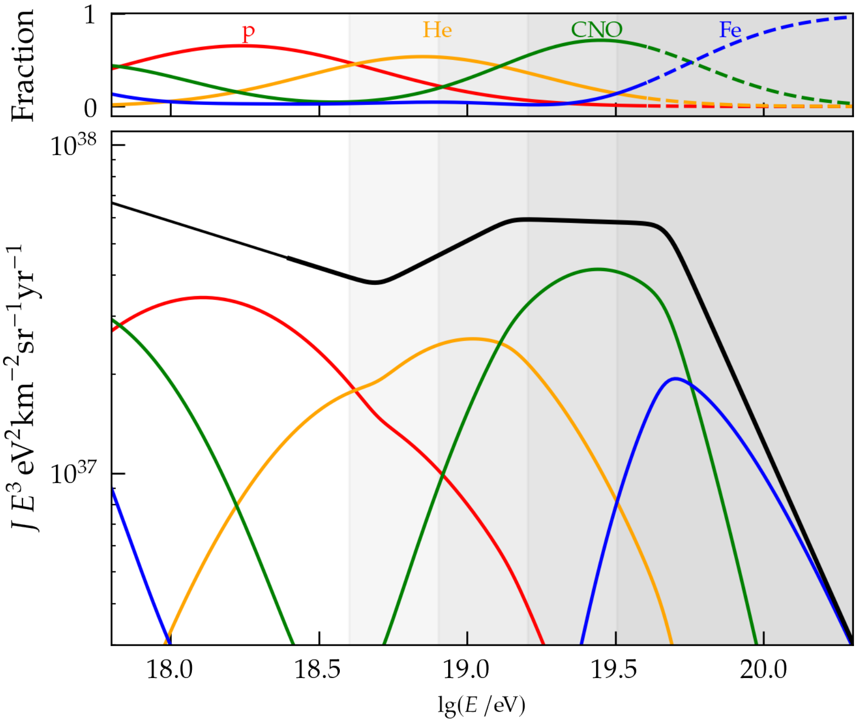


Figure 5.13: Spectrum and contributions from each of the four components considered in the extended Auger mix. In the *top* panel, the fractional contributions of four representative CR primaries as a function of energy, as parameterised in Table 5.2 for the EPOS-LHC model, are shown. The dashed lines indicate the region of extrapolation of the measurements; i.e. where the low statistic in the hybrid measurements prevents a composition estimate. In the *bottom* panel, the fitted spectrum and the contribution from each of the four components, color-coded as in the upper panel, are shown. The transition from the thinner to the thicker line indicates the transition between the data sets dominated by events detected with SD-750 and with SD-1500 arrays.

5.5. SUMMARY

The high number of events detected in the Pierre Auger Observatory during almost 20 years of Phase I allowed for the description of spectral features and for the indication of a progressively heavier mass composition.

Additionally, the observed large-scale anisotropy and the moderate significance of overdensities are indicative that Galactic and/or extragalactic magnetic fields significantly affect the distribution of arrival directions at Earth.

As discussed in the previous chapter, investigating the composition-dependence of the measured anisotropies in different scales will foster further development.

In order to probe the sensitivity of the current Auger measurements to a composition signature on the large-scale anisotropy, I have introduced a simulation library that is compatible with both the spectrum and the distribution of depths of shower maxima development X_{max}. The straight-forward Monte Carlo sampling of an existing simulation library allowed for the reproduction of the X_{max} and spectral features to a comparable level as done by more complex analyses described in Refs. [21, 20]. The advantage of the current approach is the minimization of the number of assumptions made, excluding the need for source models, extragalactic background light models, or magnetic-field models, for example. This will be further discussed in a following chapter.

In Fig. 5.14, I compile the information taken into account in this analysis. With that we can highlight how spectral features are potentially related to the changes in composition — indicated by the elongation rate breaks. I highlight also the main four large energy bins used in the dipole analysis. One can readily see that all but the 16 to 32 EeV energy bins include both a spectral feature and a possible composition change. Additionally, we currently interpret that a transition in the origin of the large-scale anisotropy occurs somewhere between 2 and 4 EeV, shaded in cyan in Fig. 5.14. The observation of an elongation rate break in this region brings attention to the possible and plausible relation between spectral and composition features.

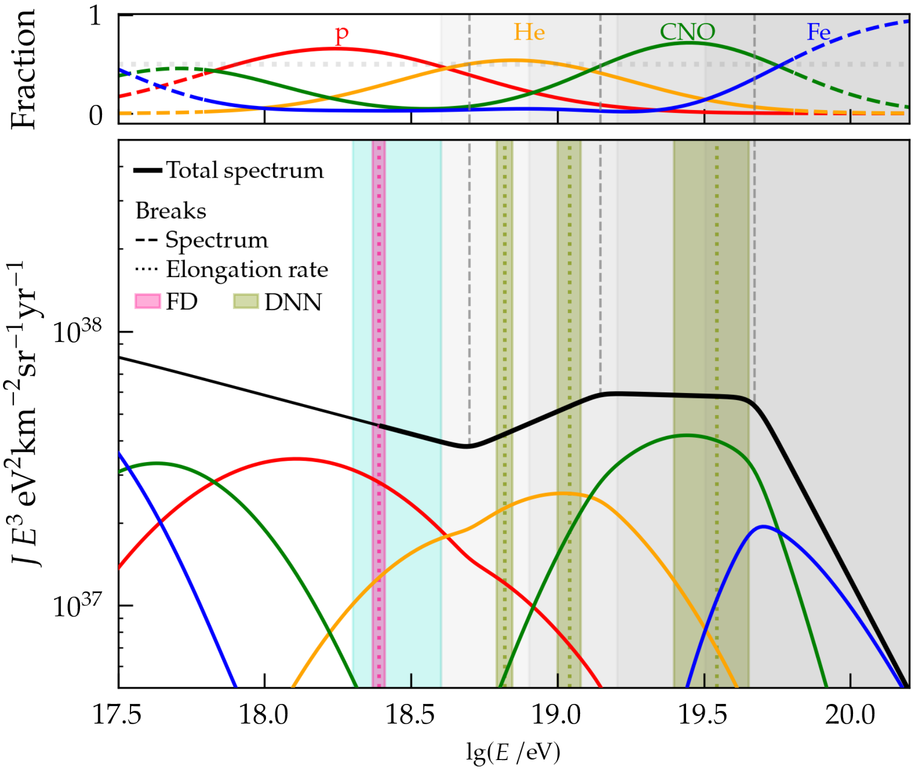


Figure 5.14: Same as in Fig. 5.13, additionally with the breaks in the energy spectrum and in the elongation rate. The later has been reported through measurements with the FD (in vertical dotted pink line) [83] and with SD data reconstructed with deep learning (in vertical dotted light green line) [88]. The bands around the vertical dotted lines represent the statistical uncertainty of the measurement. The spectral breaks are indicated with dashed gray lines. The cyan region between 2 and 4 EeV indicates the possible transition between galactic and extragalactic origin of the anisotropy as discussed in Section 4.2. Gray bands indicate each energy bin of the dipole analysis. The combination of features in the spectrum and in the elongation-rate evolution possibly provides interesting regions for future research.

CHAPTER VI

AIR SHOWER UNIVERSALITY AND A DERIVED MASS ESTIMATOR

The concept of air-shower universality has long been studied [96, 97, 98, 99]. It is based on the regularity of the energy spectra and angular distribution of most secondary particles existing at the point of maximum development of an extensive air shower, presenting the same shape and relative normalization irrespective of the initiating particle [96].

As the shower develops in the atmosphere, secondary particles will be produced, and those can be categorized into four groups [98]:

1. the muon component, μ
2. the electromagnetic component from high-energy neutral pion decay and the electromagnetic cascade thereof, $e\gamma$
3. the electromagnetic component from muon decay or interactions, $e\gamma(\mu)$
4. the electromagnetic component from the decay of low energy hadrons, $e\gamma(\text{had})$.

Exploring the spatial and temporal signatures of these components, shower-related quantities as the depth of maximum of shower development and the relative muon content of the shower can be recovered, from which the primary particle information can be inferred [98]. The *Universality* model then allows for the decomposition of a measured signal into the contributions from the subgroups of particles to uncover these quantities. In the context of the Pierre Auger Observatory, the universality principle has been deployed in the surface detector event reconstruction chain as an algorithm fine-tuned with extensive air shower and detector-response simulations [98, 97, 99]. The parametrization for both water-Cherenkov and scintillator detectors can be found in Ref. [100].

6.1. SHOWER-RELATED QUANTITIES

At the highest energies, an extensive air shower will be recorded by several detecting stations. For example, above 10 EeV the footprint of the air shower on the ground

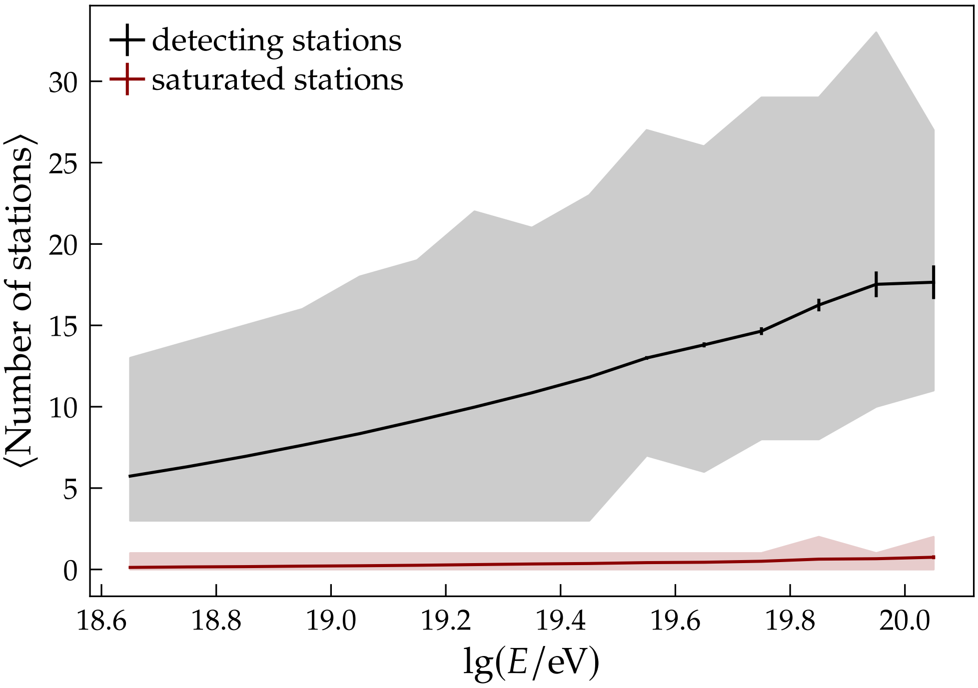


Figure 6.1: Average number of stations recording signal from air showers initiated by a UHECR with energy E, in black. In dark red, the average number of saturated stations as a function of the CR energy. The shaded regions delimit the minimum and maximum number of stations, following the same color code. Data depicted in this graph follow the quality criteria for arrival direction studies, which implies a minimum energy of 4 EeV and that at least 5 stations in the SD1500 array surrounding that with largest signal are available for data-taking.

extends over more than 25 km^2 [36], where then on average 7 or more stations will record a signal associated to the shower, as depicted in Fig. 6.1. As one of the trigger conditions, a minimum multiplicity of three stations is required.

The number of particles contributing to the signal in a station reflects the shower-to-shower fluctuations, but also the finite sampling area of the surface detector and the spacing between stations in the array. Therefore, the parametrization of the recorded signals is done relative to that at an optimal distance from the shower core. A general function that describes the lateral distribution of signals in the detecting stations at distance r from the shower core is a modification of the Nishimura-Kamata-Greisen function [36]:

$$S(r) = S(r_{\mathrm{opt}}) \left(\frac{r}{r_{\mathrm{opt}}}\right)^{\beta} \left(\frac{r + r_1}{r_{\mathrm{opt}} + r_1}\right)^{\beta+\gamma}, \qquad (6.1)$$

with $r_1 = 700\,\mathrm{m}$, $S(r_{\mathrm{opt}})$ the signal at an optimal distance to the shower core, and β a parameter that depends on the zenith angle and shower size. For the main array, with stations spaced 1500 m apart, the optimal distance is $r_{\mathrm{opt}} = 1000\,\mathrm{m}$, and the shower size is then $S(1000)$ [36].

The value of $S(1000)$ decreases with zenith angle due to the atmospheric attenuation of the shower and due to geometrical effects. The median zenith angle of the dataset is $\bar{\theta} = 38°$, and is used as a reference angle to convert $S(1000)$ in the CR energy estimator [36], as it correlates to the energy reconstructed with fluorescence telescope measurements in hybrid events, E_{FD}.

The longitudinal profile of a shower depicts the number of particles as a function of the atmospheric depth. As discussed in Chapter 3, the point where the maximum number of particles is reached, before it starts dropping, is denoted as X_{max}. It is useful to evaluate the shower development relative to this point, then discerning if the number of particles is increasing (before X_{max} is reached) or decreasing (after the maximum). The *age* parameter, s, has long been used for this purpose[96, 40], and is often described as

$$s \simeq \frac{3X}{X + 2X_{\mathrm{max}}} \,, \tag{6.2}$$

with X the atmospheric depth. At the depth of shower maximum, X_{max}, the shower age is $s = 1$, while it is smaller than one and greater than one before and after the maximum, respectively.

A parametrization of the contribution of each of the four secondary particle components can be done directly on the expected signal, without a need for modeling the detector response to each individual particle [100]. An example of the contributions of the four aforementioned shower components to the station signal is depicted in Fig. 3.2. In Ref. [100] a full description of the up-to-date version of Universality is discussed, including the parametrization for the scintillator detectors deployed for AugerPrime. The contribution from each component follows a modified Gaiser-Hillas function [99]

$$S(\Delta X) = S_{\mathrm{ref}}(r) \left(\frac{\Delta X - \Delta X_1}{\Delta X_{\mathrm{ref}} - \Delta X_1} \right)^{\frac{\Delta X_{\mathrm{max}} - \Delta X_1}{\lambda}} e^{\frac{\Delta X - \Delta X_{\mathrm{ref}}}{\lambda}} , \tag{6.3}$$

where ΔX is the atmospheric depth relative to X_{max} and ΔX_{ref} is a reference depth fixed at $200\,\mathrm{g\,cm^{-2}}$. Following the development in Ref. [99], ΔX_1 represents the depth at which the signal of a given component is expected to start for a hypothetical station close to the shower core. It is set to $-600\,\mathrm{g\,cm^{-2}}$ for the components 1 and 2, and to $-500\,\mathrm{g\,cm^{-2}}$ for the components 3 and 4 listed above.

For each radial distance to the shower axis r, a set of best fit values for $S_{\mathrm{ref}}(r)$, ΔX_{max}, and λ, is obtained by evaluating the signal response in a large simulation library.

Not only are the particles distributed spatially, but also temporally. The time of arrival of a particle at the detector is closely related to the point in the atmosphere where it was "created". Thus, the flux of particles measured at an instant t is related to the atmospheric depth of its production. The detailed development of the time dependence of different components of the showers is given in Ref. [99]. Typically, the particles arriving earliest are muons [98]. In Ref. [99], the author explores the containment of most significant information related to X_{max} in the 40% time percentile of the integrated signal, denoted t_{40}. It is a consequence of the longitudinal profile of the shower being contained in a similar proportion between the depth of first interaction, X_1, and that of maximum development, $X_{\max}$.

From the decomposition of the total signal into the contribution of the four listed components it is possible to estimate with Universality the muon content of the shower. The signal from the electromagnetic sub-shower (component 2) is weakly anti-correlated to the muon content, while the other components are strongly correlated to it [99]. Thus, due to the lack of a direct measurement of the muon number, the decomposition of the signal provides additional information of the relative muon content, R_μ. This quantity is often evaluated in comparison to the expectations of the number of muons in a proton-initiated shower as

$$R_\mu = \frac{N}{\langle N^\mathrm{p} \rangle} \, . \tag{6.4}$$

Following the cascade equations and the development in Ref. [99], a mass estimator is introduced in Eq. (6.5), which is analogous to the atomic mass number A,

$$\ln A = \frac{1}{\beta(\lambda + \phi_0)} \left[\phi_0 \left(\ln R_\mu - \ln R_\mu^\mathrm{p} \right) - \beta \left(X_{\max} - X_{\max}^\mathrm{p} \right) \right] \, . \tag{6.5}$$

While the λ parameter "weights" how much the $X_{\max}$ from a nucleus-initiated shower deviates from the proton expectations, the β parameter does so for the R_μ quantity [99],

$$-\lambda \ln A = X_{\max} - X_{\max}^\mathrm{p} \, , \qquad \text{and} \qquad \beta \ln A = \ln R_\mu - \ln R_\mu^\mathrm{p} \, . \tag{6.6}$$

From the evaluation of the simulation library described in Chapter 5, we found the following reference values and energy dependencies for the proton shower quantities.

$$X_{\max}^\mathrm{p} = 797.16 + 47.1 \times \lg(E/10^{19}\mathrm{eV}) \, , \tag{6.7}$$

$$\ln R_\mu^\mathrm{p} = 0.26 - 0.0176 \times \lg(E/10^{19}\mathrm{eV}) \, . \tag{6.8}$$

And for the relation of nuclei-initiated showers with the proton reference, the parameters λ and β could also be estimated, where we notice a small energy dependence, as also found in Ref. [99].

$$\lambda = 16.45 + 0.38 \times \lg(E/10^{19}\mathrm{eV}) \, , \tag{6.9}$$

$$\beta = 0.013 + 0.006 \times \lg(E/10^{19}\mathrm{eV}) \, . \tag{6.10}$$

In Eq. (6.5), the parameter ϕ_0 functions as a lever system between the contributions from the information on R_μ and $X_{\max}$ in order to determine $\ln A$. For example, if $\phi_0 = 0$, only the $X_{\max}$ information will be used to recover the mass estimator, while only R_μ information is used for estimating the atomic mass A in the case $\phi_0 \to \infty$. The optimal value for ϕ_0 is given by [99]

$$\phi_0 = \frac{\beta^2 \sigma^2_{X_{\max}}}{\lambda \sigma^2_{\ln R_\mu}} \, , \tag{6.11}$$

with $\sigma_{X_{\max}}$ and $\sigma_{\ln R_\mu}$ the standard deviation of the $X_{\max}$ and $\ln R_\mu$ distributions, respectively. The standard deviations can be approximated, under the assumption of a linear behavior [99],

$$\begin{aligned}\sigma_{X_{\max}}(X_{\max}) &= \sigma^{\mathrm{p}}_{X_{\max}} + (X_{\max} - X^{\mathrm{p}}_{\max})\,\frac{\Delta\sigma^{\mathrm{p,Fe}}_{X_{\max}}}{\lambda \ln 56}\,, & (6.12)\\ \sigma_{\ln R_\mu}(\ln R_\mu) &= \sigma^{\mathrm{p}}_{\ln R_\mu} + (\ln R_\mu - \ln R^{\mathrm{p}}_\mu)\,\frac{\Delta\sigma^{\mathrm{p,Fe}}_{\ln R_\mu}}{\beta \ln 56}\,. & (6.13)\end{aligned}$$

Again from the simulation library, we find the estimated behavior of Section 6.1 as

$$\sigma_{X_{\max}} = 68.2 + (X_{\max} - X^{\mathrm{p}}_{\max})\,\frac{23.1}{\lambda \ln 56}\,, \tag{6.14}$$

and of Section 6.1 as

$$\sigma_{\ln R_\mu}(\ln R_\mu) = 0.214 + (\ln R_\mu - \ln R^{\mathrm{p}}_\mu)\,\frac{-0.057}{\beta \ln 56}\,. \tag{6.15}$$

6.2. CORRECTIONS ON MEASURED DATA

The radial symmetry of the distribution of particles around the axis of an extensive air shower is one of the assumptions made for the interpretation and energy estimation with the signal recorded at the surface detector stations. However, the existence of a geomagnetic field affects the charged particle trajectories, and thus disturbs this symmetry. The effect of the Earth's magnetic field on the energy estimate with SD has been thoroughly investigated in Ref. [101].

The deflection of the relativistic charged particle due to the transverse component of the magnetic field will cause an azimuthal modulation of the particle density. As the particle density directly impacts the $S(1000)$ quantity, and therefore the energy estimate, the consequence is a modulation of the rate of events[1] as a function of the azimuthal angle, which left unaccounted for has an amplitude ~2%. Currently, the geomagnetic effects are directly accounted for in the event reconstruction [70].

As discussed in Chapter 4, the distribution of events in azimuth angle is necessary information for the computation of the dipole amplitude and direction. The modulation of the rate of events in azimuth angle impacts the dipole component parallel to the Earth's axis, d_z. Left uncorrected, the geomagnetic-field effect is a change in d_z by +0.011 [66]. Therefore, it is important to evaluate if a spurious modulation in the East-West direction exists, by observing the a_ϕ component of the Rayleigh analysis in azimuth angle as described in Section 4.2. The results are reported and discussed in Chapter 9.

As the atmospheric conditions affect the shower development, the evaluation of their effects on the energy estimate is discussed in detail in Ref. [102]. The air pressure and density, which varies throughout the day, are the quantities impacting

---

[1] Modulation on the rate of events *above* and/or *below* an energy threshold.

the most the development of extensive air showers. The pressure measured at ground level corresponds to the overburden matter traversed by the cascade of particles. Therefore, variations of that pressure implies changes in the stage of the longitudinal development of the shower when it reaches the ground, affecting the age parameter discussed previously.

The dominant effect is due to the change with the air density of the Molière radius[2] near the ground [102]. The lateral spread of particles is affected, therefore impacting the particle density and thus the measured signals, which become smaller as the density increases [102].

Additionally, the electromagnetic signal at a reference distance[3], to which the energy estimator is sensitive, corresponds mainly to the particles produced ~1 km above the ground. Although the daily air density variations are related to the ones measured at the ground, there is a delay in the response of the atmosphere to the temperature changes due to the heating or cooling of the ground. Therefore, the signal correction to account for atmospheric variations includes one factor dependent on the atmospheric pressure and two factors dependent on the air density: one the daily average of the density (at ground level), and another on the density that was determined two hours before. The two hours account for the time delay in response to the temperature changes. A consequence of these effects is a modulation of the rates of events at the solar hour frequency, presenting a typical amplitude of first harmonic of ~3.5% when not corrected for.

It is possible that the same mechanisms listed here could also affect other quantities related to the shower development, as X_{max} and R_{μ}, as the longitudinal and lateral distributions are deeply related to these quantities. Because mass-informed anisotropy searches have never before been performed, the evaluation of the atmospheric and geomagnetic effects on those quantities has also not been carried out. These effects have been studied by colleagues and reported within the Collaboration [103].

---

[2] By definition, the Molière radius is the radius of a cylinder containing ~90% of the electromagnetic energy deposition of a shower.

[3] The signal at an optimal distance to the shower core, as discussed previously.

CHAPTER VII

COMPOSITION DEPENDENCY OF A DIPOLE IN THE ARRIVAL DIRECTION OF UHECR

7.1. PREAMBLE

As discussed in Section 4.2, the most significant discovery to this date is a large-scale dipole structure with a total amplitude of approximately 7%. The Pierre Auger Collaboration has continuously observed the dipole amplitude as a function of energy, the significance surpasses 5σ only for energies above 8 EeV. Moreover, for energies above 4 EeV, the phase of the reconstructed dipole is pointing 115° away from the Galactic center, indicating that the sources of such anisotropy are extragalactic.

The following *energy* dependence of the observed dipole amplitude d was reported in Ref. [69],

$$d = d_{10} \left(\frac{E}{10\,\mathrm{EeV}} \right)^{\beta}, \tag{7.1}$$

with $d_{10} = 0.049 \pm 0.009$ and $\beta = 0.97 \pm 0.21$.

Considering that same-energy heavier elements will have lower rigidity than their lighter counterparts and that the magnetic-field deflection effects are rigidity-dependent, one can expect a more isotropic distribution of the arrival directions for the heavier CRs.

Together with the anisotropy signatures, the measurement of energy spectrum also provides insights on the phenomena taking place from the production and acceleration of UHECR to their transport and consequent measurement at Earth. The features seen in the spectrum and presented in Section 5.2 can be an indication of such effects. For example, the change of slope around 5 EeV, known as the *ankle* region, can be indicative of the dominance of the extragalactic component of the flux of UHECR. The suppression at higher energies, on the other hand, can be explained by the energy losses in interactions of the primary nuclei with the CMB, and/or by the efficiency loss on the acceleration of such energetic nuclei by the sources.

An example of such relation between spectrum and composition features can be discussed at lower energies. The measurements by Kascade-Grande [104, 105] indicate a spectral steepening of the heavy component at $\sim 80\,\mathrm{PeV}$, where the iron knee is expected to be in scenarios where the acceleration or confinement of CR in the galaxy are rigidity-dependent. The energy at which the heavier knee is expected, then, is related to the knee of the lighter counterpart, which can be expected around 3 to 5 PeV. The later has been supported by the latest measurements by LHAASO [106], which found a knee-like structure on the measured spectrum and a composition indicator compatible with a helium-dominated flux of CR around this energy.

Motivated by the knowledge of a changing composition of CRs with energy, together with the rigidity dependence of the curvature of such CR trajectories, along with potential composition-dependent effects of propagation and source ejection, we investigate a rigidity dependence of the dipole amplitude.

7.2. Astrophysical scenarios leading to a dipole structure

In an effort to explain the observed anisotropies at large scales, source scenarios combined with propagation effects have been discussed in Refs. [73, 29]. It is important to consider cases where either a few or many sources dominate the flux of nuclei at such energies and promote the overall anisotropy of the arrival direction of CRs. The scenarios were investigated in Refs. [73, 29] taking into account the possible spatial distribution of sources as well as the spatial density. It was reported that larger density of sources give rise to smaller dipole amplitudes. The authors found that, when the sources within 100 Mpc are distributed following the 2MRS catalog, the dipole amplitude increases by a factor 1.5 to 2 in comparison to the homogeneous scenario. More so, in that case the dipole structure would point towards the direction of the dipolar component of the matter distribution. A summary of the discussion and results of the Refs. [73, 29] was presented in Section 4.2.

We can gather that the spatial distribution of the extragalactic sources give rise to dipolar structures, and the galactic magnetic fields distort further the arrival direction of UHECR, both dampening the measured dipole amplitude and shifting the observed direction of such anisotropy.

As in Ref. [29], the attenuation length for each considered nuclei takes into account both photodisintegration, pair production, and adiabatic losses effects. The evaluation of Fig. 7.1 leads to the interpretation of the horizon of sources being distinct for each nuclei. For example, He nuclei with energies $\lg(E/\mathrm{eV}) > 19.6$ were most likely produced in sources within 20 Mpc; or are *daughter* nuclei of a heavier nuclei of farther origin.

In the presence of sizable turbulent magnetic fields, the propagation of cosmic rays is diffusive at low energies, for which the diffusion length ($l_D = 3D(E)/c$) is larger than distance traveled. For CRs with higher energy, the deflections are small within one coherence length l_c, and become sizable if the source distance is much larger than the coherence length. The rectilinear propagation of CR will then occur for the CRs with high energy over small distances.

With this considered, the dipole anisotropy Δ of UHECR in the diffusive regime can be estimated depending on the source(s) configuration[1]. For example, the dipole

---

[1] In this section, we use Δ to denote the dipole amplitude while in other parts of the thesis d is adopted. This

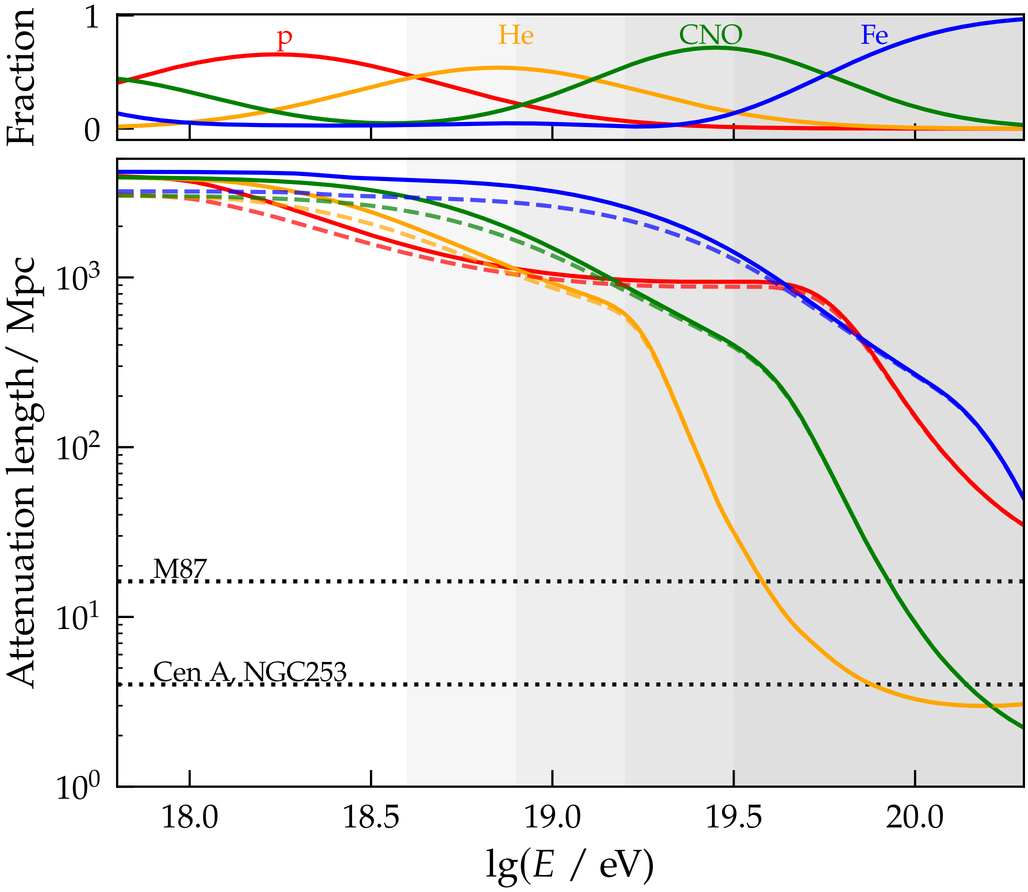


Figure 7.1: Attenuation length as a function of energy for the four representative nuclei. In the *top* panel, the fractional contributions of four representative CR primaries as a function of energy, as parameterised in Table 5.2 for the EPOS-LHC model, are shown. In the *bottom* panel, the dashed lines indicate the attenuation length for each nuclei for a redshift of $z = 0.5$, while the continuous line is that for $z = 0$. The horizontal lines indicate the distances of a few sources: Cen A and NGC 253 located ≃ 4 Mpc away from Earth, while M87 is located ≃ 17 Mpc away.

amplitude originating from a single source at a distance r_s is

$$\Delta = 3D(E)/c\, r_s, \tag{7.2}$$

being $D(E)$ the energy-dependent diffusion coefficient and c the speed of light in vacuum. $D(E)$ is, for $E > E_c$, $D(E) \propto (E/E_c)^2\, l_c$, with E_c being the critical energy where the radius of the curved trajectory of the charged particle in a turbulent magnetic field is equal to the coherence length of the field, l_c. In other words, the critical energy defines the transition between small and large deflections of a charged particle during propagation within the l_c distance in the presence of a turbulent magnetic field. For more information check Section II of Ref. [73], and Section III of

---

is to remind the reader that Δ is not equivalent to the dipole amplitude measured at Earth, d.

Ref. [29].

The dipole amplitude can be evaluated as $\Delta = 3 < \cos\theta >$, with θ being the angle between the final CR velocity and the position vector of the particle relative to the source. For a distribution of sources, the total dipole amplitude is the overlapped contribution from all sources. This implies a maximum dipole amplitude of $\Delta = 3$, corresponding to a rectilinear propagation from a single source.

Irrespective of a few sources or a distribution of many sources contributing to the UHECR flux, the dipole amplitude is expected to increase with energy. On the one hand, a particle/nuclei will present a higher rigidity with increasing energy, and as a consequence its trajectory will be less affected by magnetic fields, as the CR approaches the semi-ballistic regimen. This is also translated to the diffusion coefficient $D(E)$, being energy-dependent. On the other hand, the attenuation length decreases at higher energies as can be seen in Fig. 7.1, effectively reducing the horizon of sources or acceleration regions that can provide a nucleus at such energies. A higher contribution of nearby sources implies a higher anisotropy, and in the case these sources are distributed in a dipole pattern, the same patter should arise in the arrival direction of CRs. The dipole pattern observed at Earth can also be influenced by the configuration of the galactic magnetic field. The attenuation length as a function of the rigidity of the primary CR can be found in Appendix B., in Fig. 9.

In a consistent approach to the definition of the mass composition model, we aim at minimizing the number of assumptions while still reproducing well the measurements reported by the Auger collaboration. Instead of probing the sources distributions and considering the propagation of CR towards Earth, which requires the evaluation of the effects of magnetic fields and interactions with radiation, for example, we aim for a semiempirical approach.

7.3. A MODEL OF RIGIDITY-DEPENDENT DIPOLE AMPLITUDE

As a parameterization of dependence of the total dipole amplitude d on rigidity $R = E/Ze$, we introduce the following relation,

$$d(E, Z) = d_R \left(\frac{E/\mathrm{EeV}}{Ze}\right)^{\beta_R} \qquad \text{or} \qquad d(R) = d_R (R/\mathrm{EV})^{\beta_R}, \tag{7.3}$$

where Z is the nuclear charge ($Z = 1, 2, 8, 26$, for proton, helium, oxygen, and iron nuclei, respectively), and where d_R and β_R are the two model parameters. In this model, we additionally limit the dipole amplitude to a maximal upper-value $d_{\max}$, so that

$$d(R) = \begin{cases} d_R (R/\mathrm{EV})^{\beta_R} & ; \text{if } d(R) < d_{\max}, \\ d_{\max} & ; \text{otherwise.} \end{cases} \tag{7.4}$$

The introduction of the parameter $d_{\max}$ is related to the origin of the dipole structure, as discussed in the previous section. In the scenario where a single source is responsible for the anisotropy $d_{\max} = 3$, while in the case of a large number of sources contributing to the overall anisotropy $d_{\max} \simeq 1$ (see Ref. [29]).

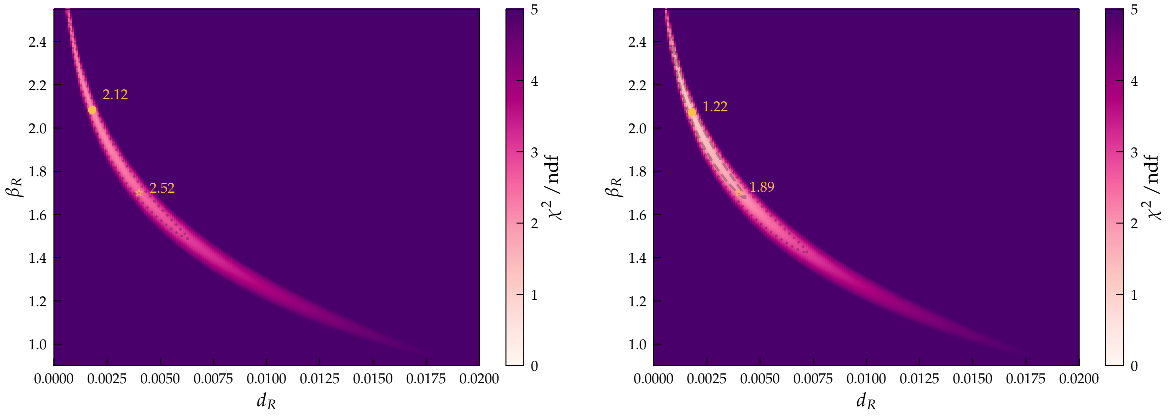


Figure 7.2: Maps of reduced χ^2 for the possible parameters in Eq. (7.3). The χ^2 is a goodness-of-fit evaluation of these parameters to describe the measured dipole amplitudes from Ref. [69], in each energy bin. The expected dipole amplitude in each energy bin is computed via Eq. (7.3), and the contribution from each component follows the composition fraction as described in the EPOS extended Auger mix, weighted by the spectrum cdf as described in Section 5.2. On the *left* the maximum dipole amplitude is $d_{\text{max}} = 1$, while on the *right* $d_{\text{max}} = 3$. Contours denote points with $\chi^2/\text{ndf} = 2$ (dashed line) and with $\chi^2/\text{ndf} = 3$ (dotted line). A full circle denotes the best parameters for the EPOS extended Auger mix composition, while a the empty star denotes an intermediate point, falling within the $\chi^2/\text{ndf} = 3$ contour for both composition models. The associated values are $d_R = 0.004$ and $\beta_R = 1.7$.

The parameters d_R and β_R can be estimated so that the model, under the described composition scenario, agrees with the measured total dipole amplitude. For the parameter estimation, we consider the three energy bins with $E > 8\,\text{EeV}$ since the dipole component in the energy bin between 4 and 8 EeV is not significant despite the large number of events therein. Additionally, the dipole amplitude in that energy range could be affected by the contribution from a low-energy source population needed to explain the flux below the ankle energy E_{ankle}, $\lg(E_{\text{ankle}}/\text{eV}) \approx 18.7$. We compute a *goodness-of-fit* to evaluate how well the parameterizations reproduced the three data points.

From the maps of reduced χ^2 in Fig. 7.2, we conclude that the parameters that allow for the best compatibility with measurements are independent of the choice of d_{max}. The values obtained by minimizing χ^2 for the EPOS Auger mix are $d_R = 0.0018$ and $\beta_R = 2.1$, adopted hereafter.

Fig. 7.3 shows the model of dipole amplitudes for each component, with colors representing the four nuclei of our UHECR composition. It portrays the equation Eq. (7.3) when using the parameters above. The horizontal lines depict the spectrum-weighted binned average of the expected total dipole amplitude. The gray boxes delimit the edges of the energy bins commonly used in the Rayleigh analysis: 4 to 8 EeV, 8 to 16 EeV, 16 to 32 EeV, and above 32 EeV.

Combining the EPOS extended Auger mix depicted in Fig. 5.5 with the rigidity-dependent dipole amplitude depicted in Fig. 7.3, we obtain the total dipole amplitude

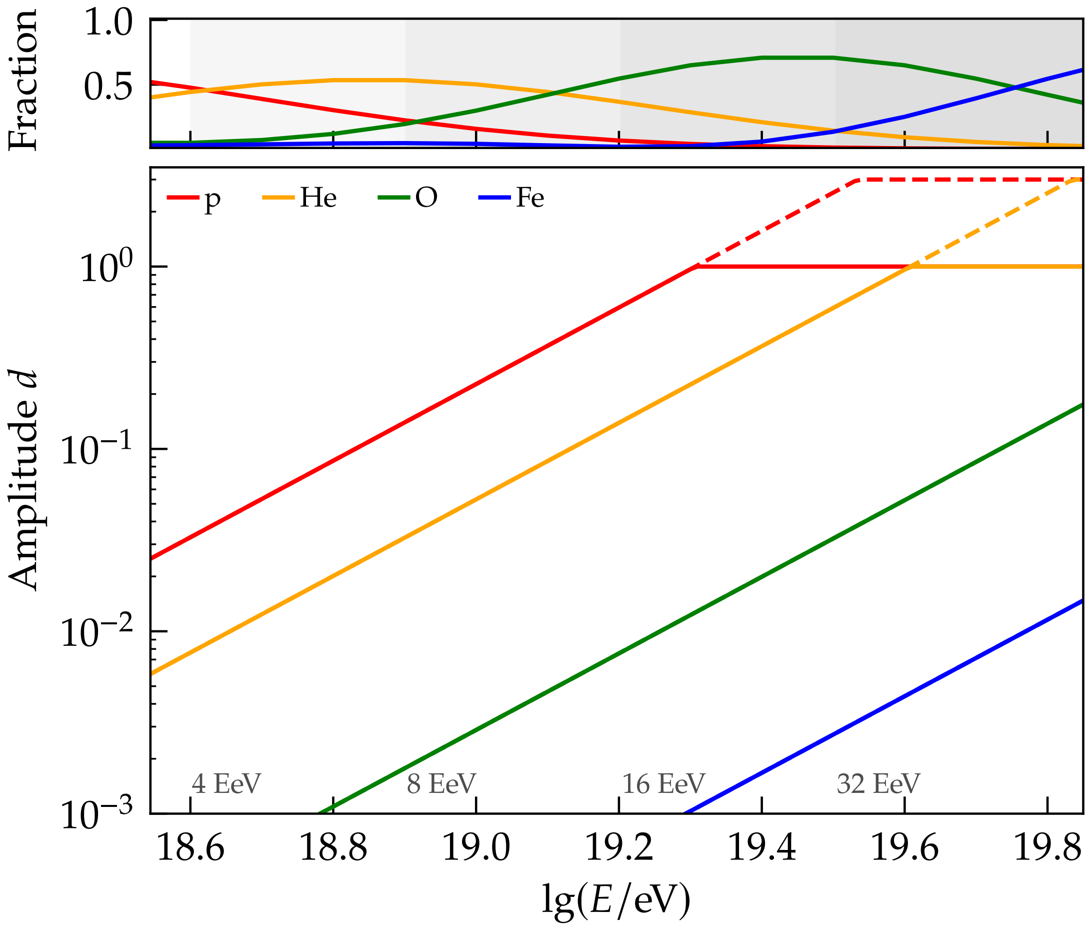


Figure 7.3: Model of the dipole amplitude d as a function of energy E and color coded for each primary particle/nucleus. The rigidity dependence of the dipole amplitude is assumed to be given by Eq. (7.3), with particular parameters indicated in Fig. 7.2. The two options for the maximum dipole amplitude, $d_{\max} = 1$ and at $d_{\max} = 3$, are shown with continuous and dashed lines, respectively. Gray boxes indicate the limits of the energy bins used for the large scale anisotrpy analysis. The top panel indicates the fractional contribution from each of the primary particle/nucleus to the overall CR flux, as described in the extended Auger mix composition model.

shown in Fig. 7.4. The colored lines are given by multiplying the component fraction for that energy by its corresponding dipole amplitude. Adding the contributions from each component results in the expected total dipole amplitude, shown in black.

In data, however, we obtain the dipole amplitude as a single measure in each energy bin. To translate the dipole amplitude model, a continuous function of energy, into the equivalent binned value, we have to consider the spectrum of the CR flux. We thus calculate the weighted average for the dipole amplitude in each energy bin, represented by the gray horizontal lines in Fig. 7.4. The weights are computed by the cdf of the spectrum function, defined in Ref. [84] and discussed in Section 5.2.

As described in Chapter 5, the Auger-compatible data set is constituted of EPOS-

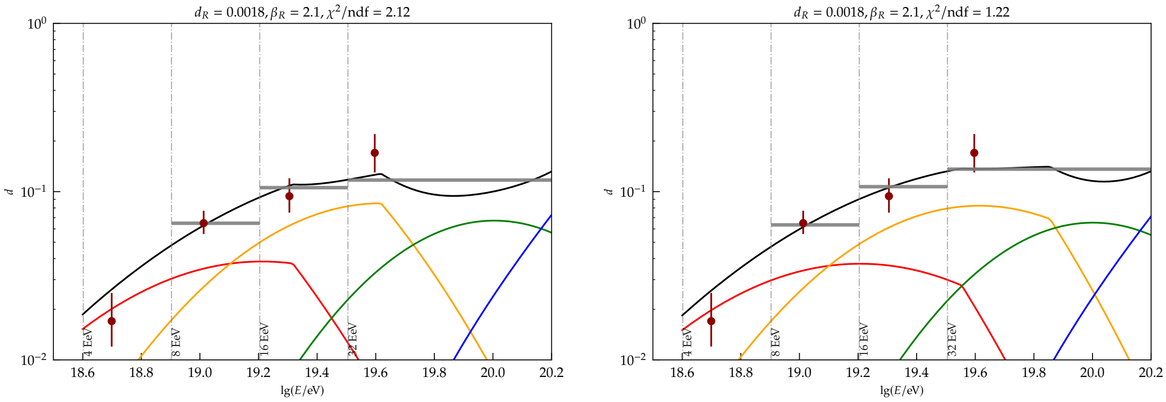


Figure 7.4: Contribution to the dipole amplitude by each component, for $d_{\max} = 1$ (left) and at $d_{\max} = 3$ (right), is given by lines in corresponding color. Summing all contributions, the total dipole amplitude is shown in black. Results are for the EPOS extended Auger mix model. Horizontal gray bars correspond to the spectrum-weighted total dipole amplitude in each large energy bin. Dark red markers represent the measured dipole amplitude as in Ref. [69]. Irrespective of the $d_{\max}$ value, the parameters that allow for a best description of the data, denoted with a circle in Fig. 7.2, are $d_R = 0.0018$ and $\beta_R = 2.1$.

LHC simulated events. Henceforth, we adopt the extended Auger mix and the d_R and β_R parameter values found in Figs. 7.2 and 7.4 to evaluate the separability of two populations in terms of their dipole amplitude.

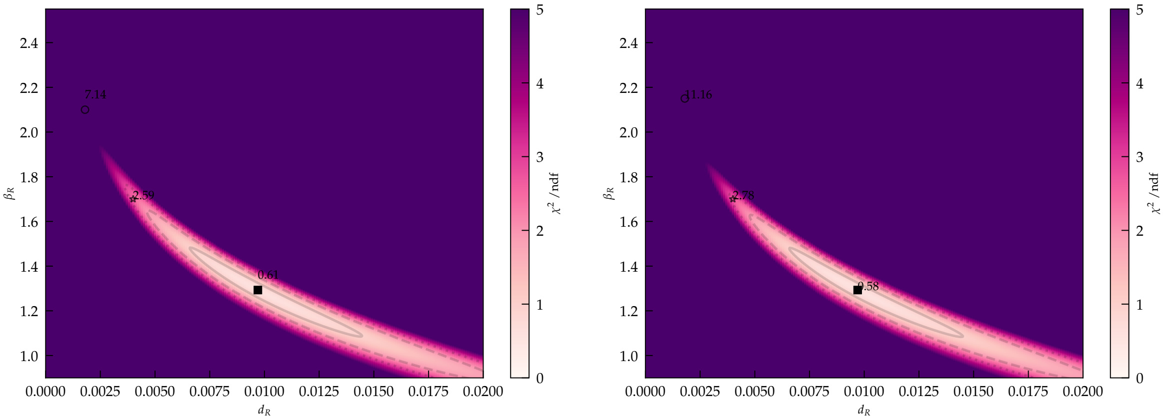


Figure 7.5: Same as in Fig. 7.2, but with Sibyll Auger mix. On the *left*, the maximum dipole amplitude is $d_{\text{max}} = 1$; while on the *right*, $d_{\text{max}} = 3$. Contours denote points with $\chi^2/\text{ndf} = 2$ (dashed line) and with $\chi^2/\text{ndf} = 3$ (dotted line). A full square denotes the best parameters for the Sibyll 2.3d extended Auger mix composition, while a circular marker represents the location of the best parameters found with the EPOS-based composition, annotated with the computed value of the reduced χ^2 when considering the Sibyll 2.3d extended Auger mix composition. The empty star denotes an intermediate point, falling within the $\chi^2/\text{ndf} = 3$ contour for both composition models, and the associated values are $d_R = 0.004$ and $\beta_R = 1.7$.

7.4. Scenario with alternative composition model

The same procedure is followed for the Sibyll 2.3d extendend Auger mix composition for completeness. First, the χ^2 scan in parameters d_R and β_R was also performed and is shown in Fig. 7.5. Following, one can observe in Fig. 7.7 the contribution from each component and the expected amplitude in each energy bin with the model parameters indicated in Fig. 7.5, which provide the best description of a rigidity dependence of the dipole amplitude under the Sibyll 2.3d Auger mix. We observe a dominance of the helium contribution to the overall dipole amplitude. This dominance is a consequence of the composition model itself, as the Sibyll 2.3d Auger mix is dominated by helium over the majority of the energy range observed, shown inFig. 5.6 with the dashed lines. In this composition scenario, the dipole amplitude should present a weaker dependence on rigidity in contrast with a more mixed composition model, such as the EPOS Auger mix.

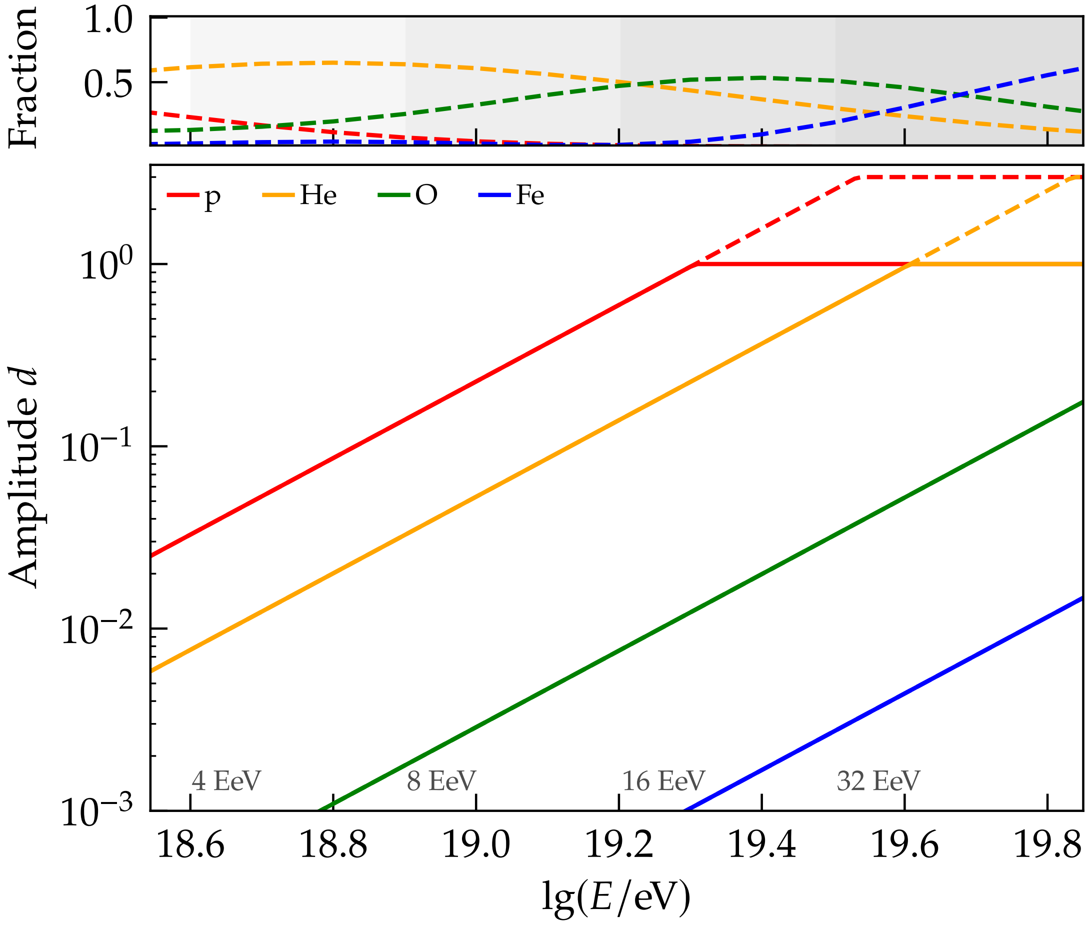


Figure 7.6: Model of the dipole amplitude d as a function of energy E and color coded for each primary particle/nucleus. The rigidity dependence of the dipole amplitude is assumed to be given by Eq. (7.3), with particular parameters indicated in Fig. 7.2. The two options for the maximum dipole amplitude, $d_{\text{max}} = 1$ and at $d_{\text{max}} = 3$, are shown with continuous and dashed lines, respectively. Gray boxes indicate the limits of the energy bins used for the large scale anisotrpy analysis. Dark red points indicate the reconstructed dipole amplitude as reported in Ref. [69]. Horizontal gray bars indicate the expectations for each energy bin, based on Eq. (7.3), the composition model, and weighted according to the spectrum measured by Auger. The top panel indicates the fractional contribution from each of the primary particle/nucleus to the overall CR flux, as described in the extended Auger mix composition model based on Sibyll expectations.

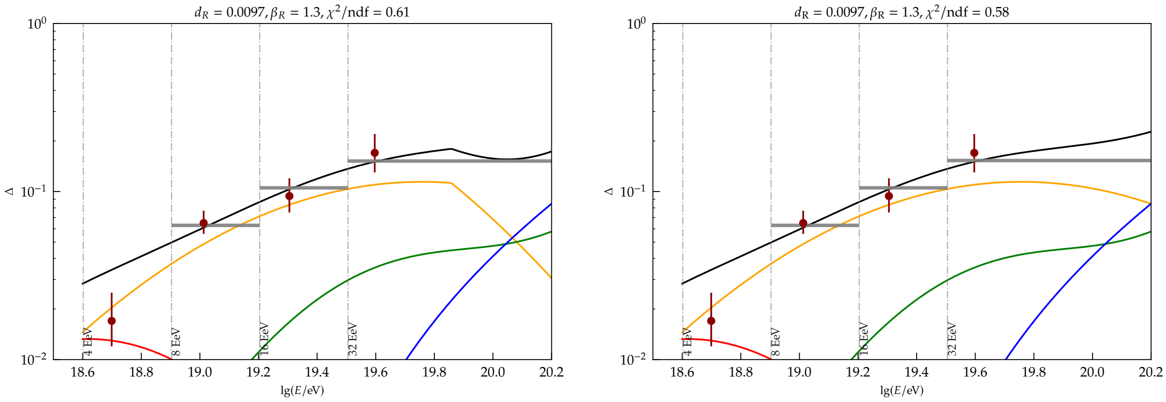


Figure 7.7: Same as in Fig. 7.4 but for the Sibyll 2.3d extended Auger mix. Irrespective of d_{max}, the parameters that allow for a best description of the data, marked with a square in Fig. 7.5, are $d_R = 0.0097$ and $\beta_R = 1.3$.

7.5. SUMMARY

The measurement of the dipole amplitude and phase are well established. Although there is a consensus of the extragalactic origin of such anisotropy, the source is not yet know. Several source models have been probed elsewhere (see Refs. [21, 20]) and I compiled in Section 7.2 the expectation for few source distribution scenarios under the mass composition adopted in this analysis.

Under a semi-empirical approach, I defined a phenomenological model capturing the qualitative expected growth of the dipolar anisotropies in the scenarios discussed above. The model does not make assumptions on source class or distribution, neither directly depends on effects of magnetic field or on propagation effects. Doing so, the definition of the parameters describing rigidity-dependent dipole amplitude model depends on the composition model adopted. It is a simple model, that could potentially be enhanced in the future by adding an A dependency, however as the adopted composition model itself takes the A progression information, we opted not to include such dependency.

With the model introduced in this chapter, that reproduces well the dipole amplitudes in the three highest energy bins, together with the Auger-compatible simulation library described in Chapter 5, we are equipped to probe the significance of the separation, in terms of the total dipole amplitude, of the two A-distinct populations.

The separation of the two sub-populations, light and heavy, will be defined in the next chapter through a scan in the Universality-reconstructed mass estimator $\ln A$. Then, I will compute for each energy bin the separation between those populations in terms of their dipole amplitude.

CHAPTER VIII

DISCOVERY POTENTIAL BASED ON SIMULATIONS

8.1. DEFINING LIGHT AND HEAVY POPULATIONS

Based on the model of rigidity dependence of the total dipole amplitude described in Chapter 7, together with the spectral features and composition evolution with energy, both described in Chapter 5, two subsets can be defined such that the difference in terms of the dipole amplitudes is maximized.

To probe how separate the two populations are, two approaches are introduced. In both approaches, all events with $\ln A$ below a threshold value $\ln A_{\text{light}}^{\text{thr}}$ constitute the *light population*, and all the events with $\ln A$ above a threshold value $\ln A_{\text{heavy}}^{\text{thr}}$ constitute the *heavy population*. No event can pertain to both populations, while some events do not pertain to either. In practice, this means $\ln A_{\text{heavy}}^{\text{thr}} > \ln A_{\text{light}}^{\text{thr}}$.

The first approach is to compute the difference between the cumulative distribution functions, CDFs, of the two populations. For each $\ln A$ threshold value, the events constituted each sub-population will be used to compute the expected dipole amplitude under this composition, according to the $d(R)$ model described previously. Then, taking the prediction for d_{light} and d_{heavy}, we compute the CDF for each assuming a Gaussian distribution centered at the computed amplitude and with width $\sigma = \sqrt{2/\mathcal{N}_{\text{pop.}}}$, with $\mathcal{N}_{\text{pop.}}$ the number of events in that population. The largest difference between the CDFs is then stored. The possible combinations of threshold values defining the light and heavy populations are evaluated and the pair that maximizes the largest difference of the CDFs is then the best threshold values for this approach.

In the second approach, the standardized mean difference, SMD, between the expected $d(R)$ for each population is evaluated. The SMD is computed with

$$SMD = \frac{|d_{\text{light}} - d_{\text{heavy}}|}{\sqrt{\sigma_{\text{light}}^2 + \sigma_{\text{heavy}}^2}}, \tag{8.1}$$

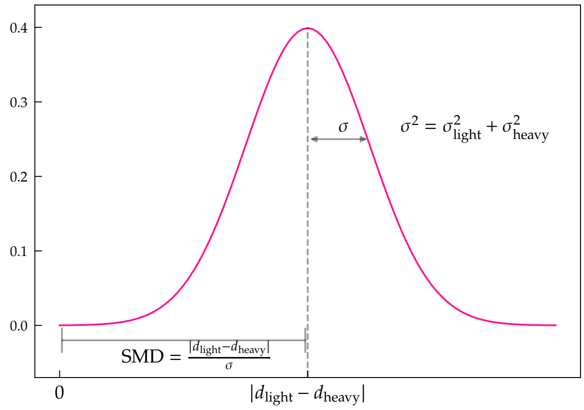


Figure 8.1: Illustration of the standardized mean difference, SMD.

with $\sigma_{\text{light/heavy}}$ being a representation of the uncertainty of such amplitude. Because in this model the total amplitude is not computed by the sum of the perpendicular and parallel amplitudes as described in Section 4.2, we use the statistical uncertainty as a proxy of the uncertainty in $d(R)$, to be computed as $\sigma = \sqrt{2/\mathcal{N}_{\text{pop.}}}$. Again, for each pair of possible threshold $\ln A$ values, the composition of the defined subpopulation is used to compute the corresponding expected dipole amplitude, and then the SMD is evaluated. The pair of values that maximizes this quantity is then the best set of threshold values for this approach in a given energy bin.

For both approaches, we simplistically assume a common direction of the light and heavy populations to compute the expected separation in total dipole amplitude. This is not be a valid assumption when the galactic (and extragalactic) magnetic field has sizable effects on the CRs trajectories. As discussed previously, the trajectories are affected according to the nuclei rigidity, and the resulting effect on the dipole amplitude depends on the length of the trajectories, being larger for further sources.

Observing Fig. 8.2, one can see that a wide variety of pair of values in that range result in a similar separation in the dipole amplitudes of each population. An overall agreement between the two approaches in observed. This can be understood when considering the Gaussian description of the distributions in d and that in both approaches the width of this distribution is taken into account. The noticeable trend of increase followed by decrease in SMD values for decreasing $\ln A^{\text{thr}}_{\text{light}}$ values, despite the increasing amplitude with smaller mass (thus higher rigidity), is due to the decreasing number of events remaining in the light supbopulation. With less events, the uncertainties in d become so large that the separation to another subpopulation is no longer significant.

Because the creation of the Auger-compatible simulation library involves a random processes in the selection of events such that the underlying distributions are followed, different realisations of this exercise will render slightly different X_{max} and thus $\ln A$ distributions thereof. Therefore, we expect a variation on the best $\ln A$ threshold values that define the light and heavy populations from the statistical fluctuations on distinct realizations of the simulation library. From Fig. 8.2 we can see

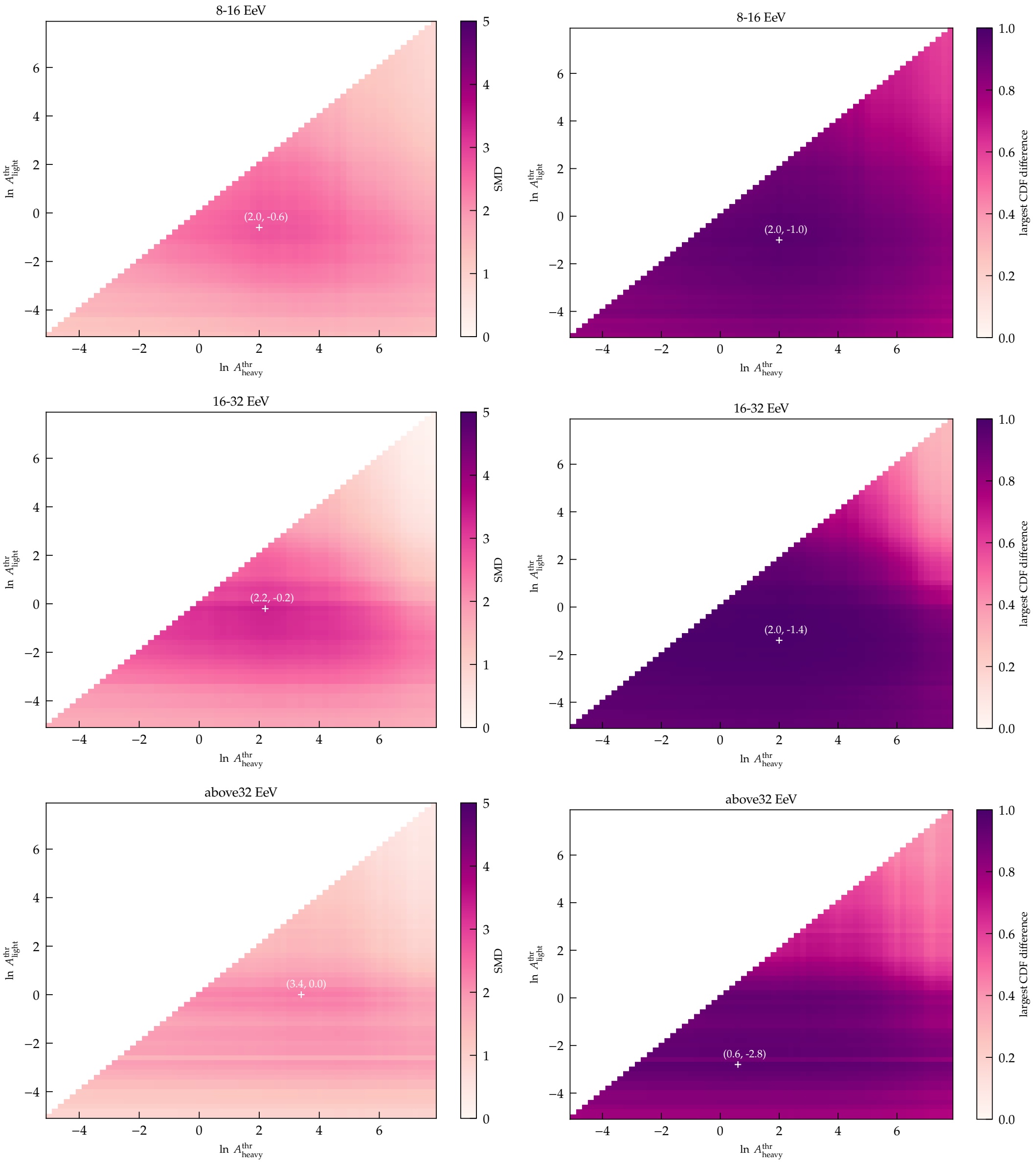


Figure 8.2: Definition of $\ln A$ threshold values on the Auger-compatible simulation library. White '+' indicates the pair of values that maximize the separation according to the approach: via the maximization of SMD (*left*) or of the largest CDF difference (*right*). In each graph, the x axis indicate the threshold values defining the heavy population, while the y axis indicate the threshold values defining the light population. Color coded is the separation power, in terms of the computed SMD on the left and largest CDF difference, on the right. Each row is representative of an energy range: 8 to 16 EeV (*top*), 16 to 32 EeV (*middle*), and above 32 EeV (*bottom*).

large regions of very similar SMD or CDF, which indicates that the effect of small variations in the threshold values on the separation power is not significant.

8.2. THE SEPARATION ON DIPOLE AMPLITUDE BETWEEN POPULATIONS

Having defined the parameters of the rigidity dependent dipole amplitude model that best describe our measurements under the assumption of a composition similar to the extended Auger mix, we set in the previous section pairs of $\ln A$ values, for each of the energy ranges of interest, that allow for the best possible separation of two mass distinct populations.

Once the *light* and *heavy* populations are defined for each energy, we can evaluate the composition therein. For example we show in Tables 8.1 and 8.2 the corresponding composition of each subset as defined with the threshold values in Fig. 8.2. Also show in the tables are the expected dipole amplitude for that population, given the composition therein and the rigidity-dependent model in Eq. (7.3). Then, we can compute how separable these two populations are in terms of the total dipole amplitude.

Summarising the findings, in Fig. 8.3 the expected dipole amplitude for the light and heavy populations. Also shown, are the measurements of the total dipole amplitude reported in Section 4.2. Additionally the horizontal lines indicate the expected dipole amplitude for a single-component CR beam, for proton (red) or iron (blue).

The very simple approach to describe the rigidity dependence of the dipole amplitude, under the assumption of the extended Auger mix composition model, allowed us to evaluate the discovery potential of a composition-informed anisotropy with Auger Phase I data. According to the prescription described in this thesis, we can expect a separation in terms of the dipole amplitude in all energy bins for two A-distinct populations. Since this approach does not rely on a scan of the parameters on data, no penalization is needed.

This proposed analysis went through the scrutiny of the Collaboration and together we discussed the possible outcome scenarios. Should we find no significant separation using the prescribed approach, it would not indicate a lack of rigidity dependence, but rather the need to move to another approach that does scan the mass-estimator parameter, in the fashion we scan the energy threshold and angular window in the analysis searching for overdensities in the sky. The reader should also consider that this model does not take any priors on the possible source distribution. It is clear from the current galactic magnetic field models that the location of the extragalactic dipole maximum, or in other words, the direction it ″points to″, affects also the dampening effect of the field (see Fig. 4.10).

Before the *unblinding* of the mass-information to the large scale searches using this analysis, several data quality and compatibility checks were performed. These included, but are not limited to, the discussion presented in the next chapter.

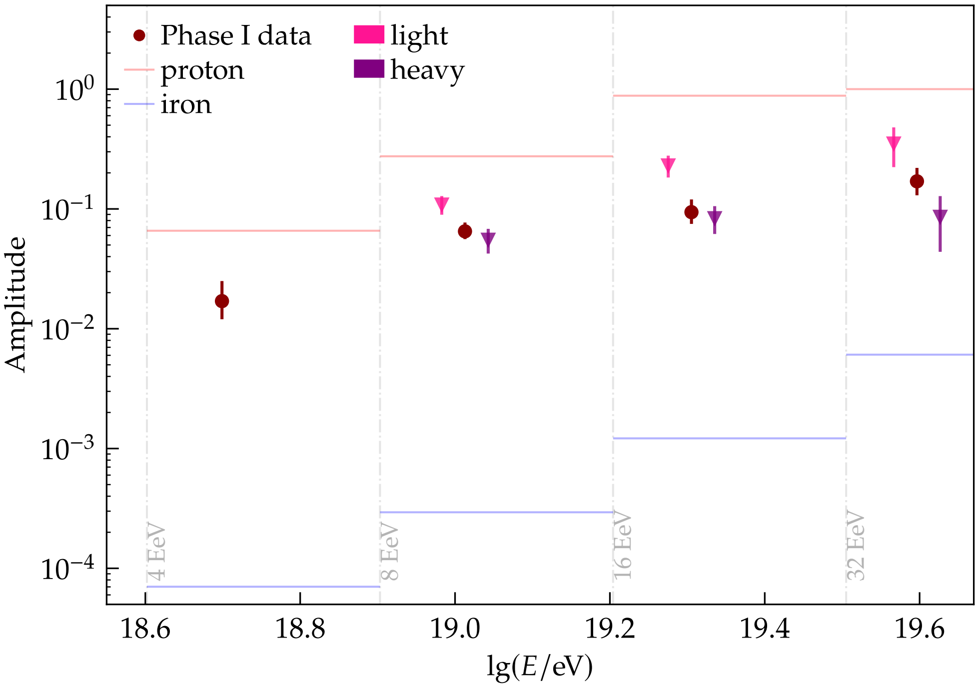


Figure 8.3: Expected dipole amplitude for two mass-distinct populations. The light (pink) and heavy (purple) populations were defined according to the threshold values that maximize the SM shown in Fig. 8.2. According to the composition in each sub-population, the expected dipole amplitude is then computed according to Eq. (7.3). The latest measurements of the dipole amplitudes are shown in dark red circles. For reference, the expected dipole amplitude for a single-component CR beam are shown for proton (red) and iron (blue).

Table 8.1: Information on the light and heavy populations, for each energy bin: 8 to 16 EeV (top), 16 to 32 EeV (middle), and above 32 EeV (bottom). Results are shown for the maximum dipole amplitude $d_{\max} = 1$ and the EPOS-LHC composition model. $N_{\text{light/heavy}}$ is the number of events , while $d_{\text{light/heavy}}$ is the total dipole amplitude in a given population. $f^{\mathcal{P}}_{\text{light/heavy}}$ is the fraction of a primary $\mathcal{P} = \{\text{p}, \text{He}, \text{CNO}, \text{Fe}\}$ in the light/heavy population.

Method	Value	$\ln A^{\text{thr}}_{\text{light}}$	N_{light}	d_{light}	$f^{\text{P}}_{\text{light}}(\%)$	$f^{\text{He}}_{\text{light}}(\%)$	$f^{\text{O}}_{\text{light}}(\%)$	$f^{\text{Fe}}_{\text{light}}(\%)$	$\ln A^{\text{thr}}_{\text{heavy}}$	N_{heavy}	d_{heavy}	$f^{\text{P}}_{\text{heavy}}(\%)$	$f^{\text{He}}_{\text{heavy}}(\%)$	$f^{\text{O}}_{\text{heavy}}(\%)$	$f^{\text{Fe}}_{\text{heavy}}(\%)$
SMD	2.74	−0.6	5571	0.1146	28.5	55.52	15.4	0.57	2.0	14316	0.0535	9.17	41.65	43.2	5.98
CDF	0.96	−1.0	4718	0.1182	30.08	54.45	14.92	0.55	2.0	14316	0.0535	9.17	41.65	43.2	5.98

Method	Value	$\ln A^{\text{thr}}_{\text{light}}$	N_{light}	d_{light}	$f^{\text{P}}_{\text{light}}(\%)$	$f^{\text{He}}_{\text{light}}(\%)$	$f^{\text{O}}_{\text{light}}(\%)$	$f^{\text{Fe}}_{\text{light}}(\%)$	$\ln A^{\text{thr}}_{\text{heavy}}$	N_{heavy}	d_{heavy}	$f^{\text{P}}_{\text{heavy}}(\%)$	$f^{\text{He}}_{\text{heavy}}(\%)$	$f^{\text{O}}_{\text{heavy}}(\%)$	$f^{\text{Fe}}_{\text{heavy}}(\%)$
SMD	3.39	−0.2	1078	0.2442	11.97	50.19	37.57	0.28	2.2	4572	0.0819	2.19	19.62	72.53	5.66
CDF	0.99	−1.4	512	0.2957	16.21	56.05	27.54	0.2	2.0	4846	0.0832	2.19	20.14	72.12	5.55

Method	Value	$\ln A^{\text{thr}}_{\text{light}}$	N_{light}	d_{light}	$f^{\text{P}}_{\text{light}}(\%)$	$f^{\text{He}}_{\text{light}}(\%)$	$f^{\text{O}}_{\text{light}}(\%)$	$f^{\text{Fe}}_{\text{light}}(\%)$	$\ln A^{\text{thr}}_{\text{heavy}}$	N_{heavy}	d_{heavy}	$f^{\text{P}}_{\text{heavy}}(\%)$	$f^{\text{He}}_{\text{heavy}}(\%)$	$f^{\text{O}}_{\text{heavy}}(\%)$	$f^{\text{Fe}}_{\text{heavy}}(\%)$
SMD	2.29	0.0	186	0.3467	1.08	33.87	58.6	6.45	3.4	1028	0.0885	0.29	5.45	50.0	44.26
CDF	0.95	−2.8	19	0.7381	0.0	84.21	10.53	5.26	0.6	1837	0.1072	0.33	7.02	57.38	35.27

Table 8.2: Information on the light and heavy populations, for each energy bin: 8 to 16 EeV (top), 16 to 32 EeV (middle), and above 32 EeV (bottom). Results are shown for the maximum dipole amplitude $d_{\max} = 3$ and the EPOS-LHC composition model. $N_{\text{light/heavy}}$ is the number of events , while $d_{\text{light/heavy}}$ is the total dipole amplitude in a given population. $f^{\mathcal{P}}_{\text{light/heavy}}$ is the fraction of a primary $\mathcal{P} = \{\text{p}, \text{He}, \text{CNO}, \text{Fe}\}$ in the light/heavy population.

Method	Value	$\ln A^{\text{thr}}_{\text{light}}$	N_{light}	d_{light}	$f^{\text{P}}_{\text{light}}(\%)$	$f^{\text{He}}_{\text{light}}(\%)$	$f^{\text{O}}_{\text{light}}(\%)$	$f^{\text{Fe}}_{\text{light}}(\%)$	$\ln A^{\text{thr}}_{\text{heavy}}$	N_{heavy}	d_{heavy}	$f^{\text{P}}_{\text{heavy}}(\%)$	$f^{\text{He}}_{\text{heavy}}(\%)$	$f^{\text{O}}_{\text{heavy}}(\%)$	$f^{\text{Fe}}_{\text{heavy}}(\%)$
SMD	2.74	−0.6	5571	0.1146	28.5	55.52	15.4	0.57	2.0	14316	0.0535	9.17	41.65	43.2	5.98
CDF	0.96	−1.0	4718	0.1182	30.08	54.45	14.92	0.55	2.0	14316	0.0535	9.17	41.65	43.2	5.98

Method	Value	$\ln A^{\text{thr}}_{\text{light}}$	N_{light}	d_{light}	$f^{\text{P}}_{\text{light}}(\%)$	$f^{\text{He}}_{\text{light}}(\%)$	$f^{\text{O}}_{\text{light}}(\%)$	$f^{\text{Fe}}_{\text{light}}(\%)$	$\ln A^{\text{thr}}_{\text{heavy}}$	N_{heavy}	d_{heavy}	$f^{\text{P}}_{\text{heavy}}(\%)$	$f^{\text{He}}_{\text{heavy}}(\%)$	$f^{\text{O}}_{\text{heavy}}(\%)$	$f^{\text{Fe}}_{\text{heavy}}(\%)$
SMD	3.91	−0.2	1078	0.2749	11.97	50.19	37.57	0.28	2.2	4572	0.0875	2.19	19.62	72.53	5.66
CDF	1.0	−1.4	512	0.3374	16.21	56.05	27.54	0.2	1.8	5084	0.0903	2.18	20.73	71.68	5.41

Method	Value	$\ln A^{\text{thr}}_{\text{light}}$	N_{light}	d_{light}	$f^{\text{P}}_{\text{light}}(\%)$	$f^{\text{He}}_{\text{light}}(\%)$	$f^{\text{O}}_{\text{light}}(\%)$	$f^{\text{Fe}}_{\text{light}}(\%)$	$\ln A^{\text{thr}}_{\text{heavy}}$	N_{heavy}	d_{heavy}	$f^{\text{P}}_{\text{heavy}}(\%)$	$f^{\text{He}}_{\text{heavy}}(\%)$	$f^{\text{O}}_{\text{heavy}}(\%)$	$f^{\text{Fe}}_{\text{heavy}}(\%)$
SMD	3.2	0.0	186	0.4713	1.08	33.87	58.6	6.45	3.6	990	0.1096	0.2	5.56	49.7	44.55
CDF	0.99	−1.8	46	0.7296	2.17	54.35	39.13	4.35	0.6	1837	0.1355	0.33	7.02	57.38	35.27

CHAPTER IX

AUGER PHASE I

9.1. THE DATA SET

The data set considered here is similar to the one reported in Ref. [70]. It consists of events measured with the main surface array SD-1500, where the water-Cerenkov stations are spaced 1500 m apart. The events were recorded between 01 January 2004 and 31 December 2022. In contrast to the dataset in Ref. [70], the events considered here have a maximum zenith angle of $\theta_{\text{max}} = 60^\circ$, allowing for a coverage of ~70% of the sky. As customary for the large-scale anisotropy analysis, the selected events must have at least five active stations surrounding the one with largest signal, and the reconstructed shower core must lie within an isosceles triangle of active stations. These conditions guarantee a proper reconstruction of the shower plane and thus the arrival direction of the primary CR, as well as reliable reconstruction of important information such as the primary CR energy. The total exposure for the events with $\theta \leq 60^\circ$ is $95\,118\,\text{km}^2\,\text{sr}\,\text{yr}$.

The energies of the events are corrected for atmospheric and geomagnetic effects to avoid spurious modulations in right ascension and azimuth [66, 107]. The systematic uncertainty of the absolute energy scale is ~14% [36] while the statistical uncertainty is ~7% for events with $E > 10\,\text{EeV}$. The angular resolution is better than 0.9° for events with $E > 10\,\text{EeV}$, and better than 1.6° for events with 3 stations, which is a more frequent scenario at lower energies. For additional information, see Ref. [36].

As the Auger Collaboration prepared for the second phase of data acquisition, AugerPrime, a more robust framework was introduced, making it possible to combine the different detection techniques for the surface array (see Section 3.3) and to evaluate the reconstruction on simulated events. For the first time, the new framework is being used for the large-scale anisotropy analysis. For purposes of consistency checks presented in this chapter, the previous framework will be named *Herald*, and the most recent framework, *Observer*. The checks evaluate long-term effects, as well as the agreement between important observed quantities such as energy, X_{max}, and the arrival directions. The consistency checks are performed for events that are successfully reconstructed in both frameworks, and the comparison is often performed on an event-by-event level, using for that a unique event identifier.

9.2. IMPORTANT OBSERVED QUANTITIES

As discussed in Section 5.2, the spectrum of cosmic rays has been continuously reported by the Auger Collaboration, and features have been identified. In the context of large-scale anisotropy analyses, the data is grouped in large energy bins, namely from 4 to 8, 8 to 16, 16 to 32 and above 32 EeV, often times also including a cumulative bin above 8 EeV. Because the energy ranges are so wide, smaller disagreements in the energy estimator of two frameworks are not expected to cause significant discrepancies. Nonetheless, migration of events to one or the other energy bin will occur whenever these estimators are not perfectly compatible.

Thus, the first compatibility check between the two frameworks is the number of events as a function of energy. In Fig. 9.1, a generally good agreement is observed. The differences are compatible with statistical fluctuations and lie well within 3σ. The differences are also within the known ~5% difference on the signal estimation between the two frameworks up to $E{\approx}10^{19.8}$eV (reported internally to the Collaboration in Ref. [108]).

Another important quantity for the purposes of this analysis is the depth of the shower maximum development, X_{max}, detailed in Chapter 6. The detection of events in hybrid mode, with information from both SD and FD as described in Chapter 3, allows for accurate measurement of X_{max}, with bias of few g/cm^2 and resolution that goes from ~22 to ~15 g/cm^2 with increasing energy when all the selection criteria described in Ref. [83] are met.

In Fig. 9.2, the measurements of the first two moments of the X_{max} distributions are presented. Measurements with the FD in the hybrid data set are shown with green square markers and the ones with the SD data set used in this analysis are shown with purple circles. Considering systematic uncertainties of the order of 30 g/cm^2 as evaluated in Ref. [99], an overall agreement between the two measurements is reached. The agreement between the X_{max} measurements with FD and SD detectors has also been discussed in Refs. [88, 109] using a deep-learning technique.

Additionally, quantities related to the arrival direction information and to the dipole amplitudes are of great importance and should be closely evaluated. As formulated in Eq. (4.3), the components of the dipole amplitude are related to the mean values of the sine of the zenith angle, and of the cosine of the equatorial declination of the events (see detailed explanation in Section 4.1 of Ref. [110]). Therefore, we present these distributions and average values for each energy bin in Fig. 9.3.

Evaluating Fig. 9.3, one can observe that the differences between the means of the distributions are small, below 1%. Consequentially, differences in the dipole amplitudes emerging from these discrepancies will be proportionally small. The values reported therein are in agreement with previous results (see Refs. [66, 107]); with the mean sine of the θ being affected by the maximum zenith angle adopted in this data set, $\theta \leq 60°$.

Spurious modulations in solar hour and annual frequencies can induce spurious modulations in the sidereal (RA) and antisidereal frequencies and as a consequence tamper the measured equatorial dipole amplitude (see Ref. [107]). Because no significant modulation of the CR flux should be present at these frequencies for an

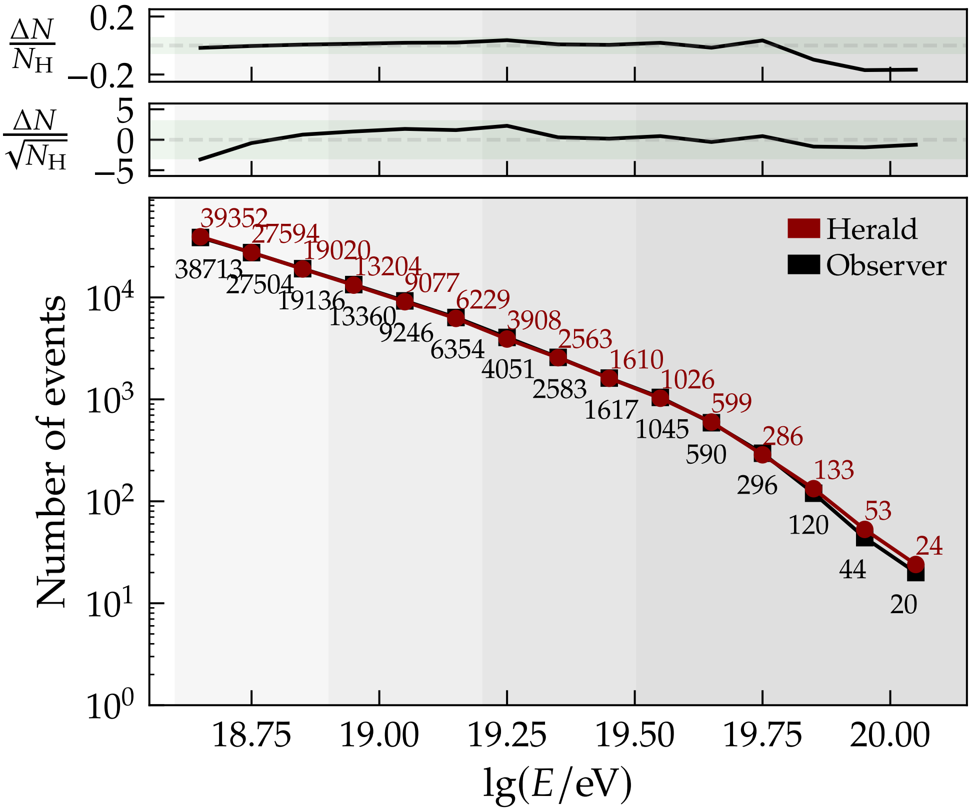


Figure 9.1: Number of events in Herald and in Observer frameworks. Comparison is performed for events that are successfully reconstructed in both frameworks, thus all events have a zenith angle $\theta \leq 60°$. As the total number of events is the same for both data sets, the differences in each energy interval is due to small differences in the reconstructed energy in each framework and consequential bin migration. Gray bands indicate the large energy bins used in the dipole analysis described in Section 4.2. Top panels indicate the difference in the number of events in each framework data set relative to the number of events in Herald, N_{H}, (*top*) or to the square root of that number (*mid*). The green bands indicate a 5% (*top*) or 3σ (*mid*) deviation.

anisotropy of astrophysical origin, the presence of a modulation in solar hour or anti-sidereal frequencies indicate that systematic effects were not properly accounted for. The description of the main sources of systematic effects, together with the impact on the dipole components can be found in Ref. [107], for example, and will be discussed in the following paragraphs.

One known effect is that of the geomagnetic field on the charged particles produced in the extensive air shower. The deflection of their trajectories will disturb the radial symmetry, biasing the energy estimator in an azimuth-dependent manner. The detailed description of this effect can be found in Ref. [101]. Left uncorrected, a modulation of the rate of events in azimuth angle is noticeable. To showcase that the related effect has been properly accounted for, we evaluate in Fig. 9.4 the normalized rate of events as a function of azimuth for each energy bin of interest. The normalized

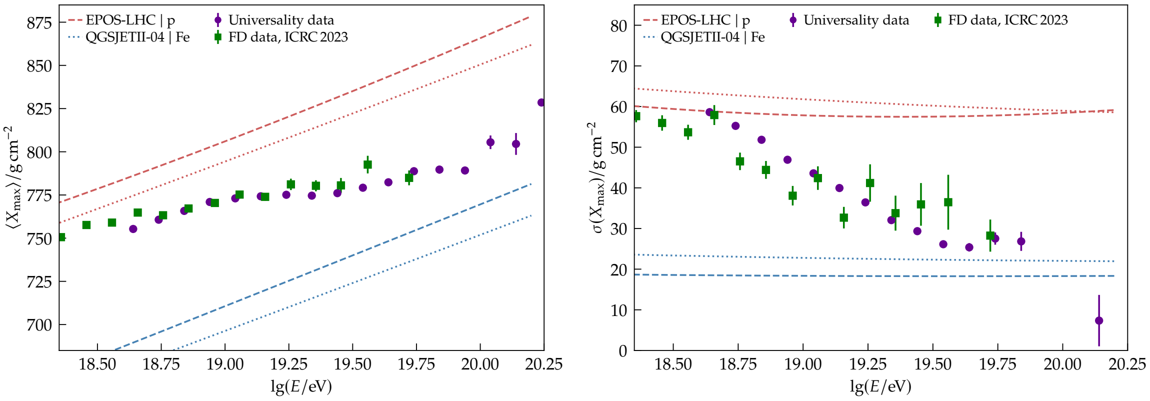


Figure 9.2: Mean (left) and standard deviation (right) of X_max distribution as a function of energy. Measurements by the fluorescence detector, FD, as reported in Ref. [83] are shown in green squares with the related statistical uncertainties. Measurements with the surface detector, SD, and reconstructed with the Universality approach as described in Chapter 6 are shown in purple circles with the related statistical uncertainties. As a reference, expectation for two primaries and two hadronic models are shown: red lines indicate proton expectations while blue, iron. Continuous lines indicate the expectations of EPOS-LHC model, and dashed line those of QGSJETII-04.

rate is computed as $N_i/\langle N\rangle$, where N_i is the number of events in a range of azimuth angle ϕ and $\langle N\rangle$ the mean number of events per angular interval in a energy bin. Annotated therein, the goodness of fit to a uniform distribution, evaluated via the reduced chi squared quantity.

To confirm that any remaining modulation in azimuth are not consequence of the effects of the geomagnetic field, we compute a Rayleigh analysis in ϕ as described in Section 4.2. In Fig. 9.5, we observe the negligible modulations caused by geomagnetic field effects on the energy by evaluating the a_ϕ amplitude. This parameter measures the East-West modulation and should be compatible with zero as flux modulations from the sky should be averaged out with the Earth's rotation. We can conclude that the spurious effect on the energy estimator has been well accounted for.

Another known effect is that of atmospheric conditions. The variations in air density and atmospheric pressure impact the shower development, and as a result a modulation of at most 2% is observed as a function of solar hours if this effect is not corrected for [102]. The energies are overestimated during afternoons when the temperature reaches a maximum and underestimated during the early mornings. This modulation, reported in Ref. [102], has a typical amplitude of ~3.5% in the event rate with energies above 2 EeV.

To evaluate if the atmospheric effects have been properly accounted for in the energy assignment, we evaluate the normalized rate of events as a function of solar hour for the energy bins of interest. In Fig. 9.6 we observe that the amplitudes are small and overall consistent with zero within the uncertainties for both frameworks. The largest differences between frameworks is observed in the 8 to 16 EeV energy

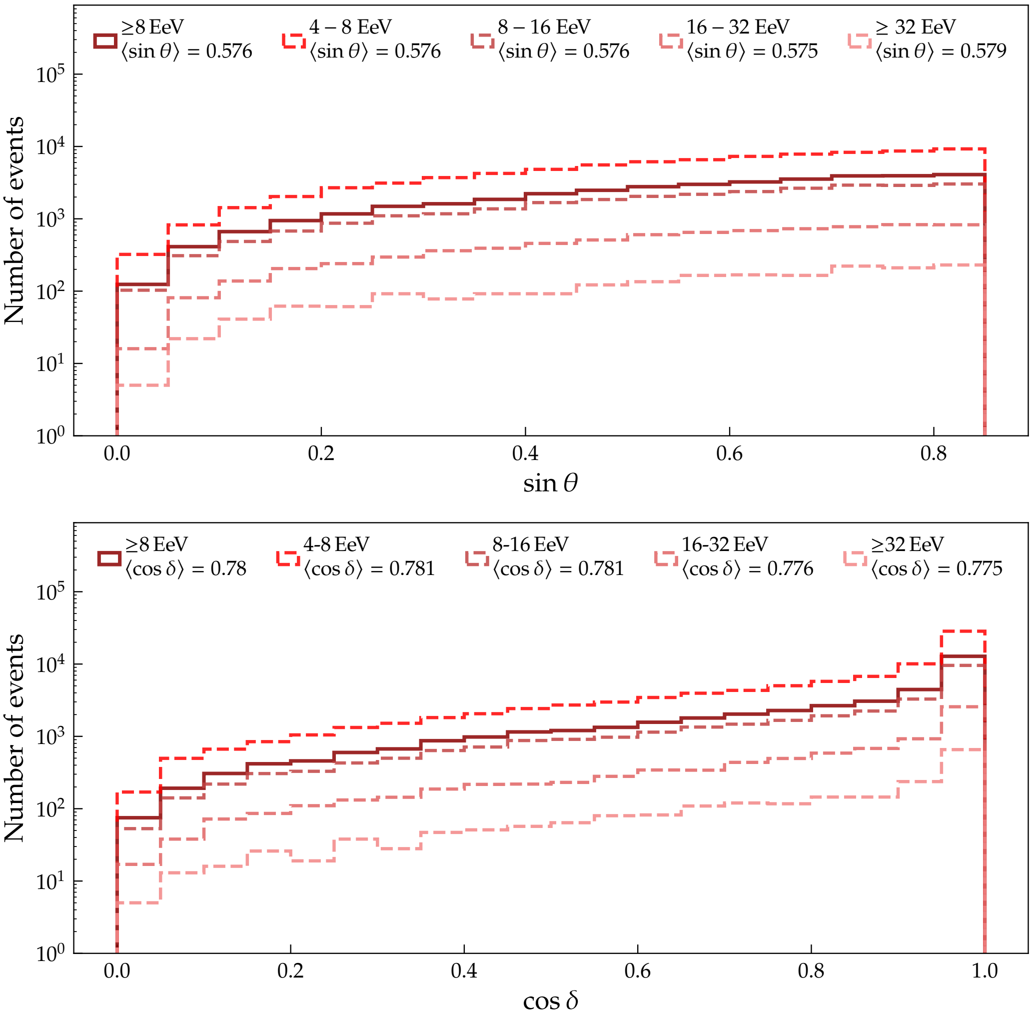


Figure 9.3: Number of events as a function of the sin of the zenith angle θ (top) and as a function of the cosine of the equatorial declination (bottom) of the events in each relevant energy bin. The darker, continuous line represents all events with energy above 8 EeV, while the dashed lines with lighter shades indicate the energy bins 4 to 8, 8 to 16, 16 to 32 and above 32 EeV. Annotated in each graph, the mean value of the distribution. This is an important quantity as is used for the computation of the dipole components.

range. There the amplitude is $r_1^{\text{solar}} = (1.6 \pm 0.9)\%$ when using the Observer-assigned energy, and $r_1^{\text{solar}} = (0.6 \pm 0.9)\%$ when using the Herald energy estimator. Nonetheless, the chance probability of obtaining such amplitude is of $P(r \geq r_1^{\text{solar}}) = 23\%$ in the Observer case.

In Fig. 9.7 we observe that the amplitudes are small and overall consistent with zero within the uncertainties for both frameworks. The largest differences between frameworks is observed in the 8 to 16 EeV energy range. There the amplitude is $r_1^{\text{antis}} = (1.0{\pm}0.9)\%$ when using the Observer-assigned energy, and $r_1^{\text{antis}} = (1.3{\pm}0.9)\%$

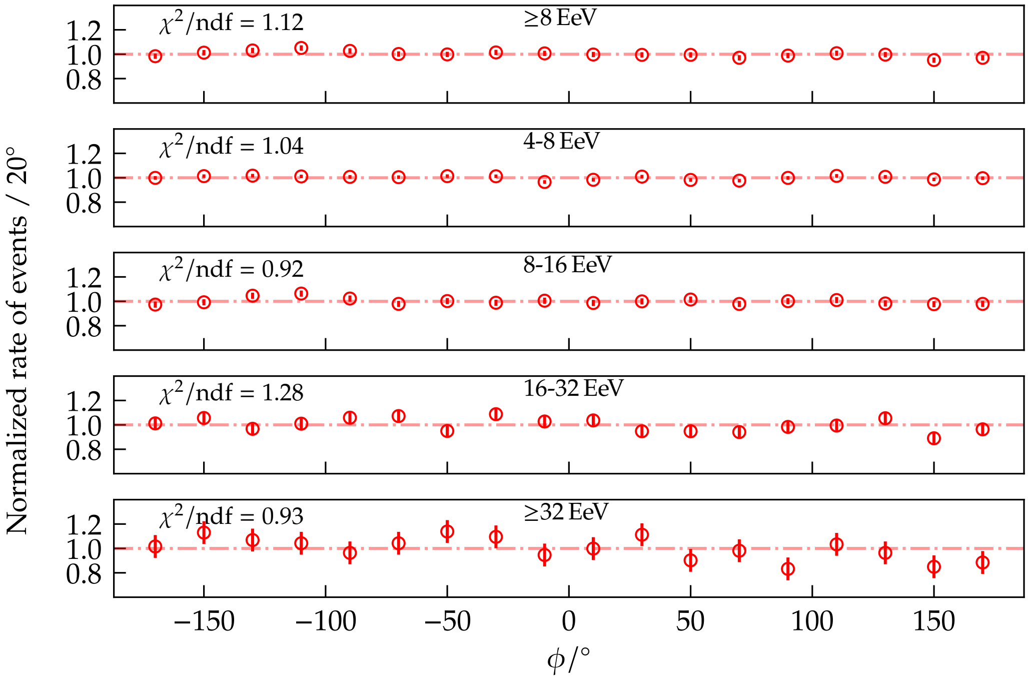


Figure 9.4: The normalized rate of events as a function of azimuth angle ϕ for each energy bin of interest. Annotated therein, the goodness of fit to a uniform distribution, evaluated via the reduced chi squared quantity.

when using the Herald energy estimator. Nonetheless, the chance probability of obtaining such amplitude is of $P(r \geq r_1^{\text{antis}}) = 38\%$ in the Herald case.

We conclude that the effects in the energy assignment that could cause a modulation in the event rate have been properly accounted for, and the remaining modulations are consistent with zero in both frameworks.

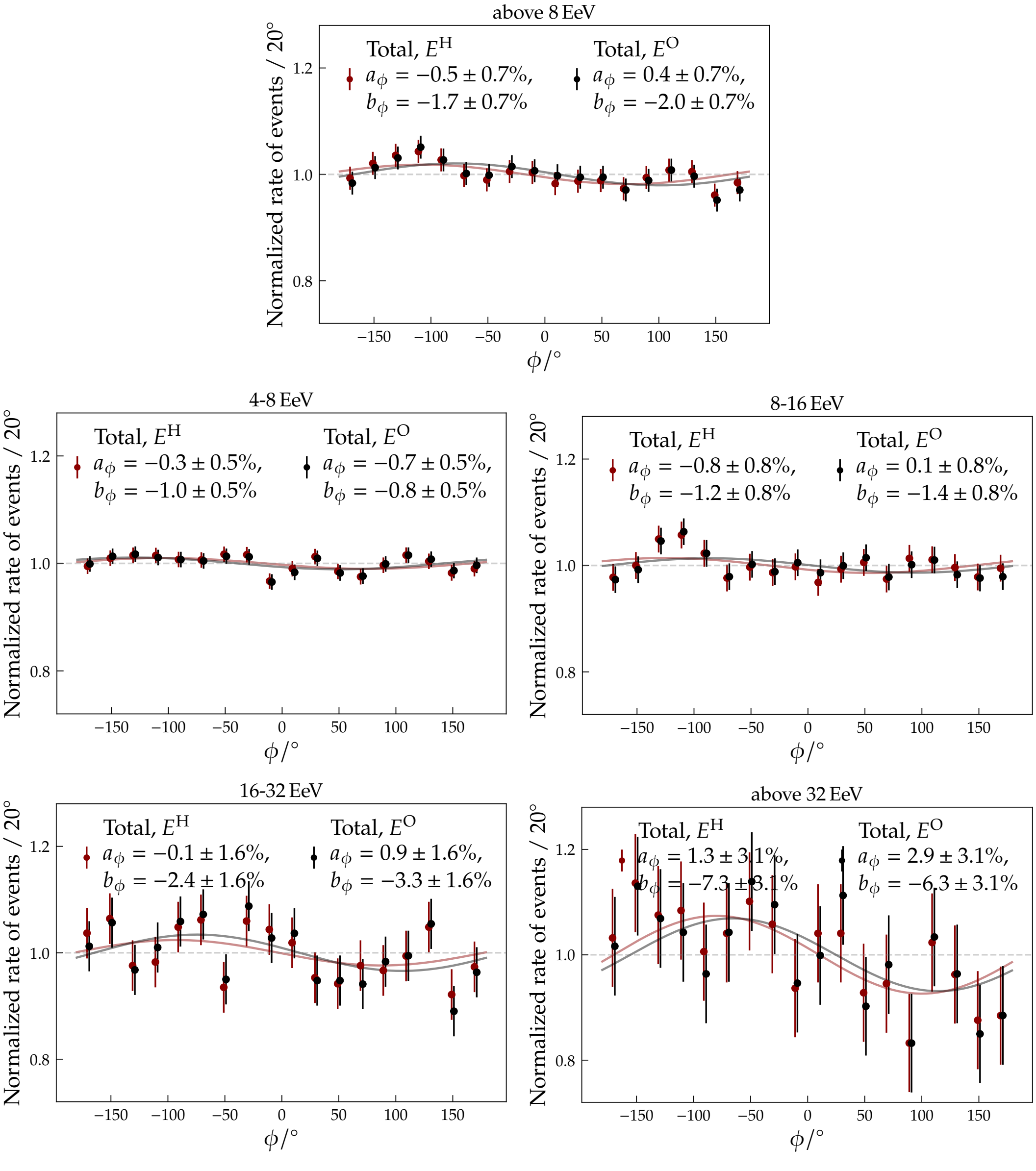


Figure 9.5: Rate of events as a function of the azimuth angle ϕ. Each graph corresponds to the modulation in an energy range of interest: all events above 8 EeV (top), 4 to 8 (middle left), 8 to 16 (middle right), 16 to 32 (bottom left) and above 32 EeV (bottom right). The first harmonic amplitudes of the event rate modulation in azimuth are annotated in each graph, together with the corresponding statistical uncertainty. Dark red markers indicate that the events were placed in each energy bin according to the Herald energy estimators,E^H, while in black using the Observer energy, E^O.

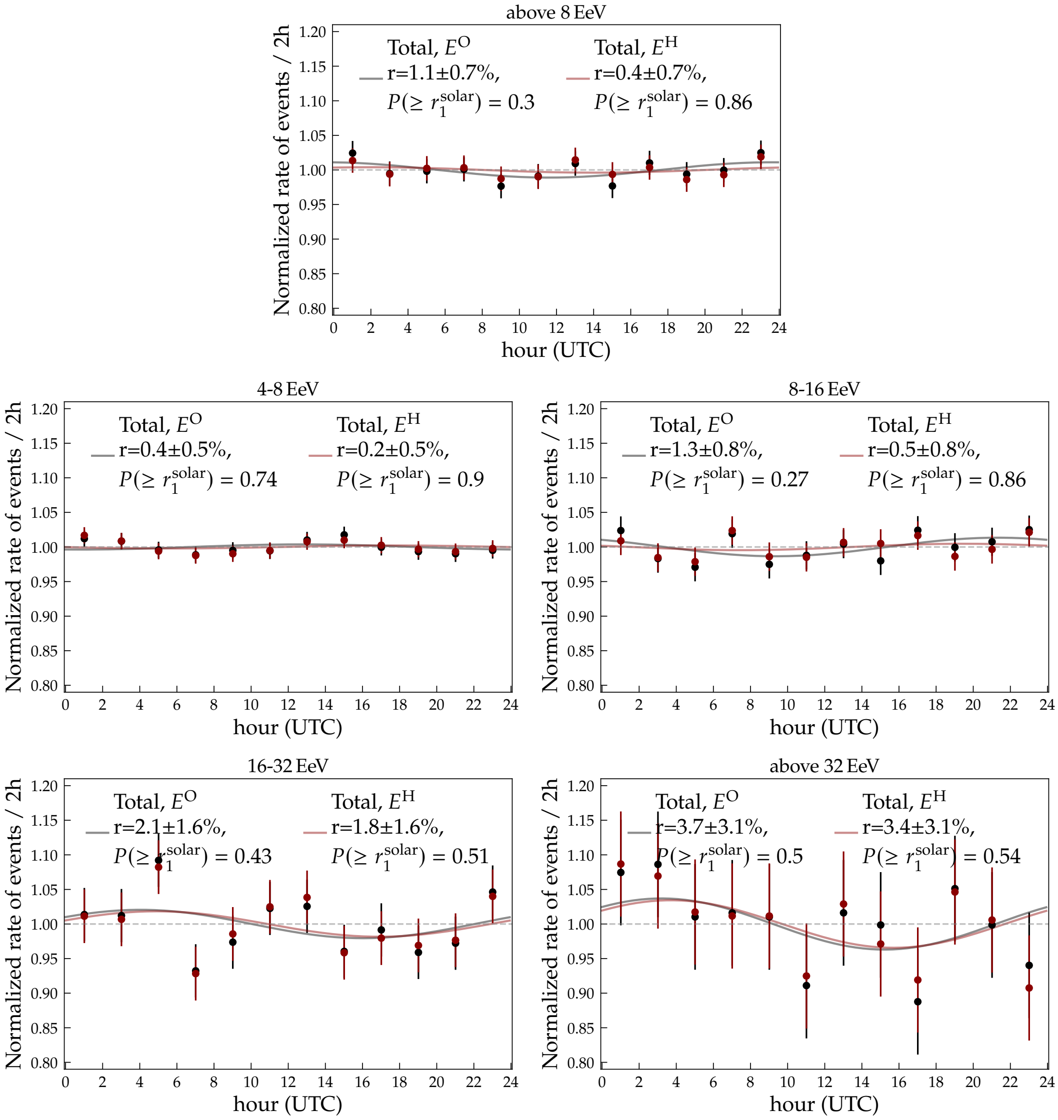


Figure 9.6: Rate of events as a function of the solar hour. Each graph corresponds to the modulation in an energy range of interest: all events above 8 EeV (top), 4 to 8 (middle left), 8 to 16 (middle right), 16 to 32 (bottom left) and above 32 EeV (bottom right). The first harmonic amplitudes of the event rate modulation in solar hour are annotated in each graph, together with the corresponding statistical uncertainty. Dark red markers indicate that the events were placed in each energy bin according to the Herald energy estimator, E^{H}, while in black using the Observer energy, E^{O}. The local time in Malargüe corresponds to UTC −3.

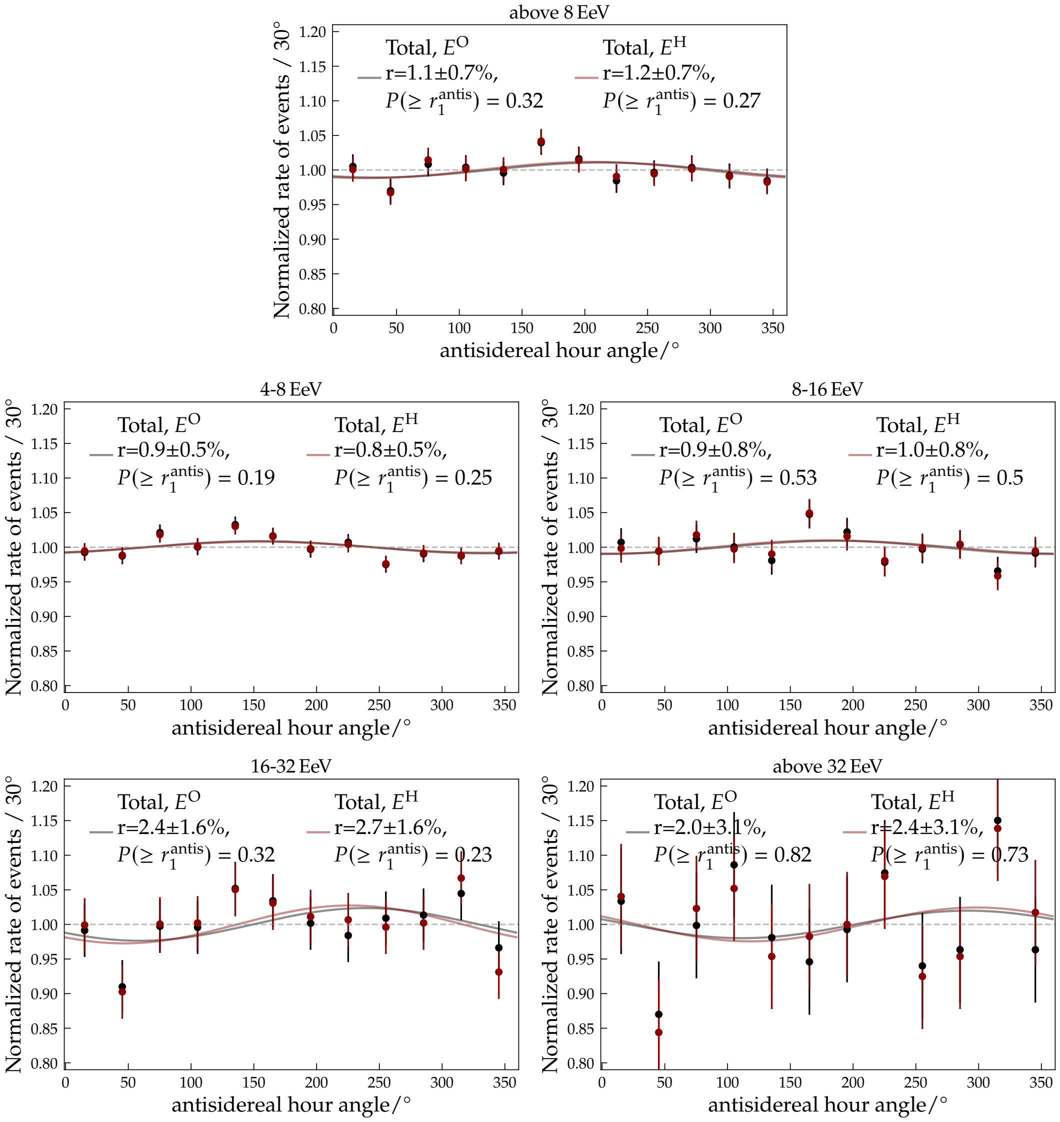


Figure 9.7: Rate of events as a function of the antisidereal hour angle. Each graph corresponds to the modulation in an energy range of interest: all events above 8 EeV (top), 4 to 8 (middle left), 8 to 16 (middle right), 16 to 32 (bottom left) and above 32 EeV (bottom right). The first harmonic amplitudes of the event rate modulation in antisidereal hour angle are annotated in each graph, together with the corresponding statistical uncertainty. Dark red markers indicate that the events were placed in each energy bin according to the Herald energy estimator, E^H, while in black using the Observer energy, E^O.

9.3. LIGHT AND HEAVY POPULATIONS

With the definition of *light* and *heavy* populations introduced in Section 8.1, we obtain two subsets of events in each energy bin. In this section we evaluate the relevant quantities in each energy range of interest for the large-scale analysis. A comparison to the expectations from the simulation library described in Chapter 5 allows us to evaluate the agreement before the *unblinding* of the mass-informed arrival directions and possible anisotropies.

In Fig. 9.8, we observe the distributions in X_{max} as a function of the corresponding $\ln A(X_{\mathrm{max}}, R_{\mu})$, for each energy bin. The events falling into the light population are represented in pink, while those in the heavy category are depicted in blue. The panels on the top and side of each graph represent the marginal distributions in $\ln A$ and X_{max}, respectively. On the top, the percentage of all events that fall into one or the other population is annotated with the corresponding color. We also observe the number of events in each subset of events, annotated in each central graph.

Upon closer inspection one can notice that some of the events in the heavy population present an X_{max} value compatible to that of the light population. This occurs in ~2.5% and ~1.6% of the events in the heavy subset in the energy ranges 4 to 8 and 8 to 16 EeV, respectively. The large $\ln A$ values for events with intermediate X_{max} values is potentially due to a poor R_{μ} reconstruction. Due to the small number of events in this cases, the systematic effect is expected to be much smaller than the statistical uncertainties in the dipole amplitudes.

Aside from these percent level disparity in the heavy population, no significant overlap in the X_{max} distributions is observed, further indicating that this is a relevant quantity for the mass discrimination of CR, as expected from the evaluation of cascade equations. The reader is reminded that the energy range of 4 to 8 EeV was not intended to be used in the mass-informed dipole amplitude analysis, and the inclusion of said energy bin in the tests performed in this Chapter are of learning purposes, should the Collaboration decide for future anisotropy analyses in this energy range.

In Table 9.1, we compare the fraction of events falling in each population for the measurements (depicted in Fig. 9.8) and for the Auger-compatible simulation library used to set the $\ln A$ threshold values defining the populations. The differences can be attributed to discrepancies between the mass composition model and the true, unknown mass composition, and to different distributions of the $\ln A$ when comparing simulations and measurements. The last effect could occur whenever there is a bias that has not been accounted for, shifting all $\ln A$ values in one direction. The width of the distribution also impacts the number of selected events above or below a threshold.

Comparing the simulations and measurements of the mass-estimator distribution in Fig. 9.9, it is noticeable that in the simulation library both first and second momenta present larger than the measured values, for most energies. This is an indication that most likely both effects have a role in the differences reported in Table 9.1. Together with the distribution of the simulations and measurements of $\ln A$, the empty circle markers indicate the true $\ln A$ distribution in the Auger-like simulation

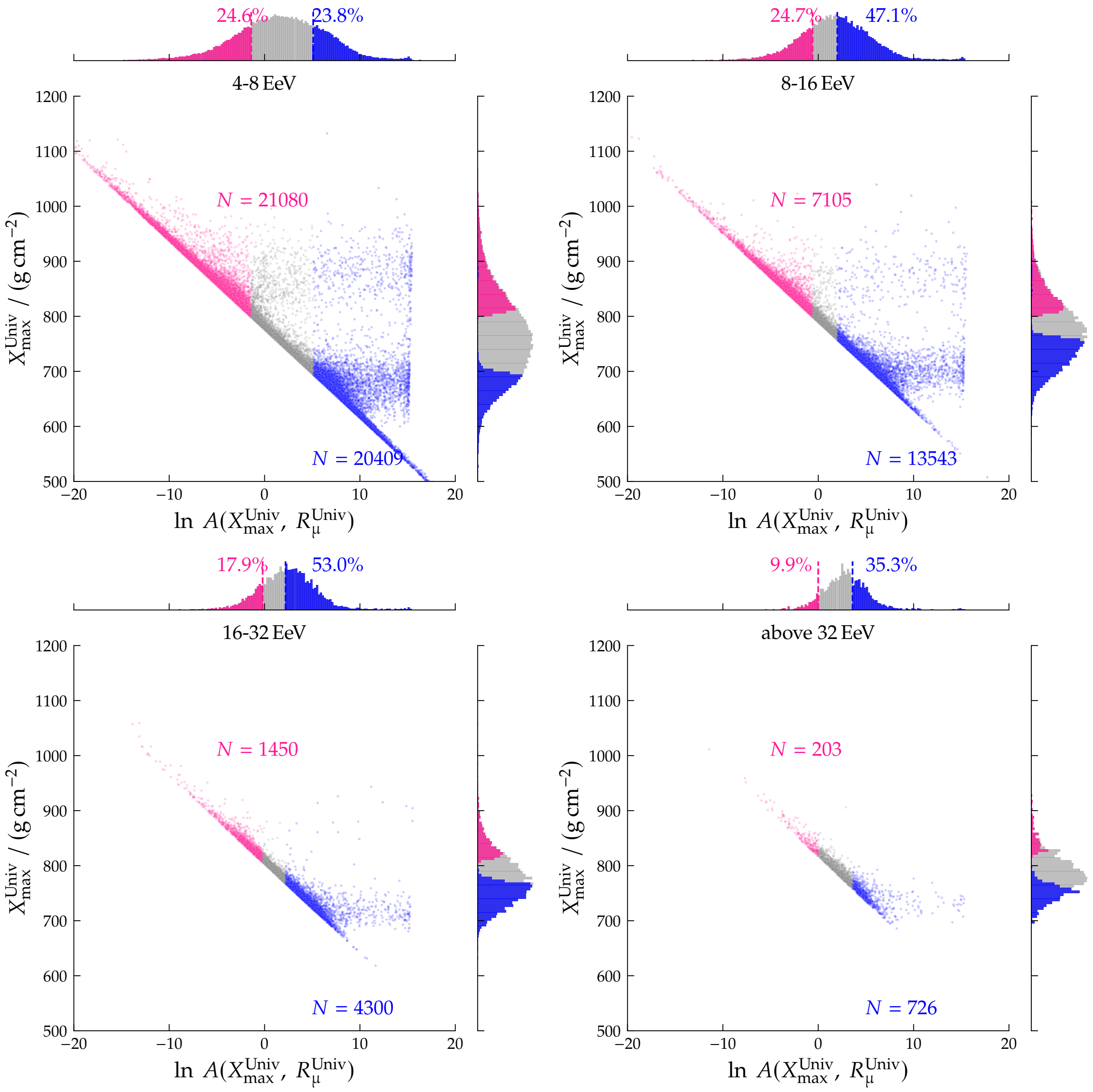


Figure 9.8: Distribution of the mass-related quantities: the mass estimator ln A on the x axis and the $X_{\max}$ on the y axis. The top and side panels on each graph indicate the marginal distributions of ln A and $X_{\max}$, respectively. While the pink color indicates the events and distributions are associated with the light population, the blue color is used to refer to the heavy population. Annotated in the same color scheme are the number of events in the population (in the central plots) and the percentage of the total set of event in said population (on the top panel). The distributions are shown for the following energy ranges: 4 to 8 EeV (top left), 8 to 16 EeV (top right), 16 to 32 EeV (bottom left), and above 32 EeV (bottom right).

Table 9.1: Events falling into the light and heavy populations of UHECR. In the first column, the energy range is indicated, followed by the description of the dataset. The third and fourth columns present the fraction of events in that range that are categorized as a *light* or *heavy* event. Lastly, the total number of events in that data set is given.

E/EeV	data set	$f_{\text{light}}(\%)$	$f_{\text{heavy}}(\%)$	N_{total}
8 to 16	simulations	19.5	50	28 612
	measurements	24.4	47.3	29 052
16 to 32	simulations	13.4	57	8 024
	measurements	17.4	54	8 191
above 32	simulations	8.7	46.3	2 136
	measurements	9.4	36.4	2 088

library, indicative of the composition mixing without any measurement effects. If the detection technique had perfect resolution and accuracy, the distributions of the true and reconstructed $\ln A$ values in the simulation library would be identical.

As in the previous section, we evaluate the distribution and mean values of $\sin\theta$ and $\cos\delta$ and these quantities are closely related to the components of the dipole. We repeat the exercise for each population, showing in Fig. 9.10 the distributions in the light population, and in Fig. 9.11 those related to the heavy population.

We can observe from Fig. 9.10 that the largest difference in $\langle\sin\theta\rangle$ occurs for the highest energy range, being ~8% larger than that of the lowest energy range. This can be understood as a consequence of the shower development and the relative position of the shower maximum development to the detection level (see the Appendix fr more information). The differences in $\langle\cos\delta\rangle$ are no larger than 1%, as found in Fig. 9.3.

Regarding the heavy population and its distributions in θ and in δ, the differences in $\langle\cos\delta\rangle$ are also of the order of 1%. On the other hand, the largest difference in $\langle\sin\theta\rangle$ is between the energy ranges of 4 to 8 EeV and 16 to 32 EeV, being ~5% larger at smaller energies. In this case, this could be an indication of the full efficiency of the Observatory being reached at higher energies for the heavier composition. This is further indicated by evaluating Fig. 9.12.

Next we evaluate the anisotropy-related checks for both populations in each energy bin. Namely, we evaluate the uniformity of the azimuth angle distribution and the modulation in the rate of events as a function of azimuth angle, solar hour, and anti-sidereal hour angle as in the previous section. For the sake of simplicity, in this section I refrain from the comparison of the two frameworks.

In Fig. 9.13 we observe a generally good agreement with a uniform distribution in azimuth angle for both populations in all energy bins. And in Fig. 9.14, we can evaluate if any remaining modulation in azimuth angle is not a consequence of the effects of the geomagnetic field. The a_ϕ parameter that measures the East-West

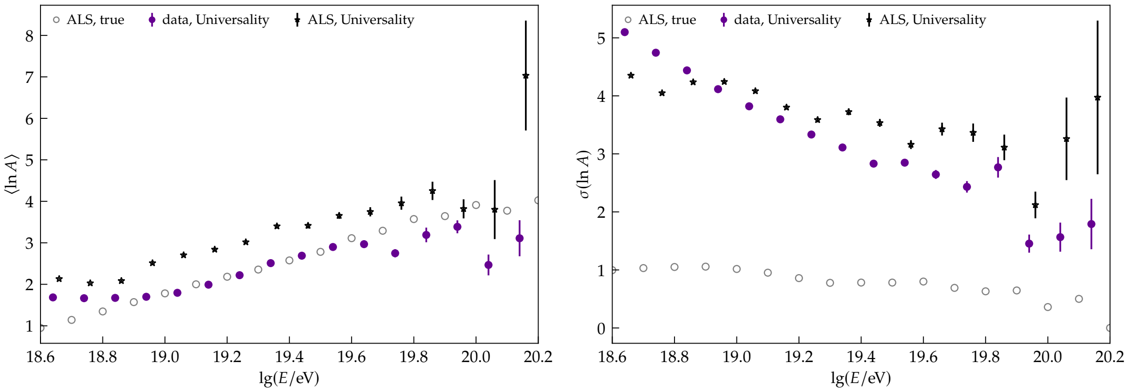


Figure 9.9: Mean (left) and standard deviation (right) of the mass estimator $\ln A$ as a function of energy. Shown are the results in the Auger-like simulation library (ALS) and the measurements in the Auger Phase I data. While the *true* value of $\ln A$ depends solely on the composition mixing, the reconstructed quantity with Universality has intrinsic dependency on both the shower-to-shower fluctuations and also on the reconstruction accuracy and resolution of the detection technique. Vertical bars indicate the statistical uncertainty of the Universality-reconstructed values. The true values are shown solely as reference and thus no uncertainty is attributed to it.

modulation is compatible with zero for both populations in all energy bins. We can conclude that the spurious effect on the mass-related quantities has been well accounted for.

To evaluate if the atmospheric effects have been properly accounted for in the mass-related quantities, we evaluate the normalized rate of events as a function of solar hour for the energy bins of interest. In Fig. 9.15 we observe that the amplitudes are small and overall consistent with zero within the uncertainties for both populations. The largest modulations are found for the light component, being greater at the lowest energy bin. There the amplitude is $r_1^{\text{solar}} = (2.5 \pm 1.1)\%$. Nonetheless, the chance probability of obtaining such amplitude is of $P(r \geq r_1^{\text{solar}}) = 6\%$.

Lastly, the rate of events in the antisidereal hour angle for both populations in each energy bin can be appreciated in Fig. 9.16. Again, all amplitudes are compatible with zero within the statistical uncertainties. Similar to Fig. 9.7, the largest modulation is found for the 8 to 16 EeV energy range in the heavy population, where the amplitude is $r_1^{\text{antis}} = (2.4 \pm 1.3)\%$. The chance probability of obtaining such amplitude is of $P(r \geq r_1^{\text{antis}}) = 18\%$.

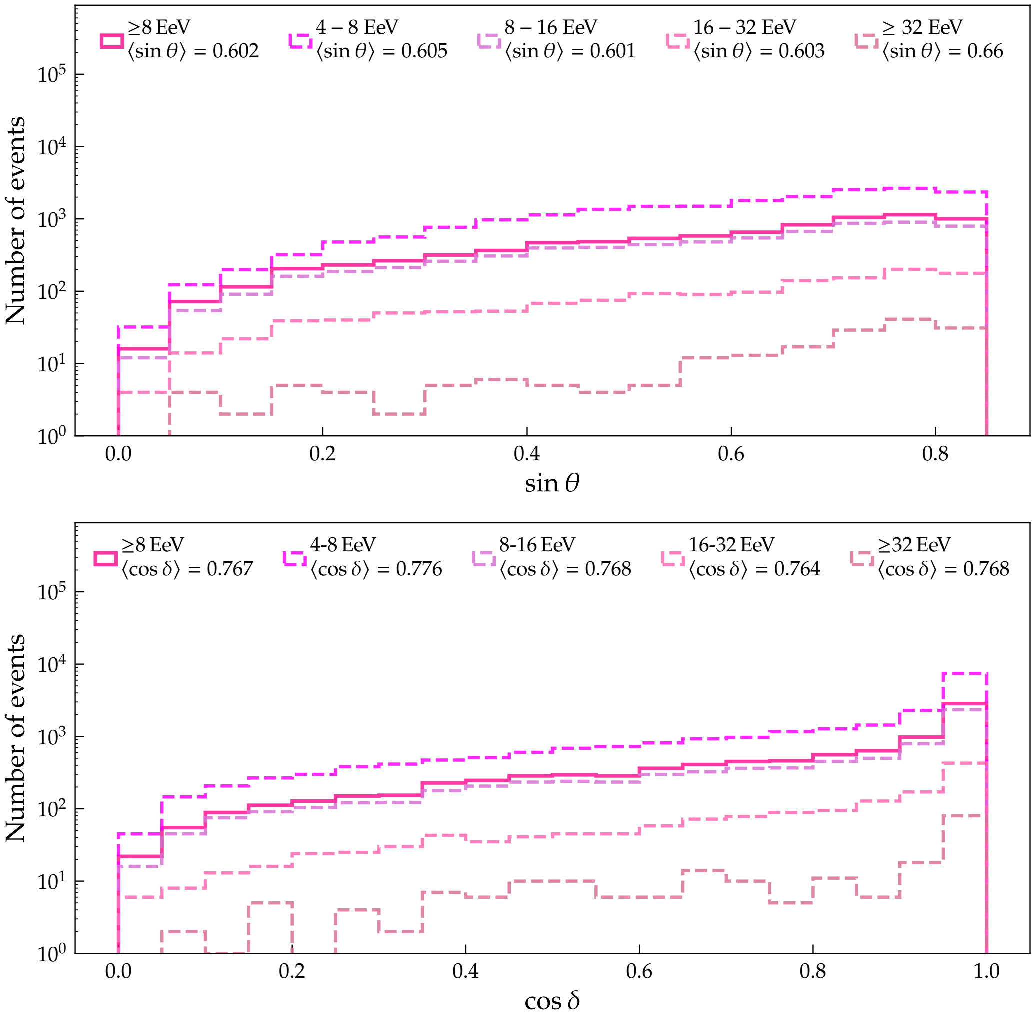


Figure 9.10: Number of events as a function of the sin of the zenith angle θ (top) and as a function of the cosine of the equatorial declination (bottom) of the events in the light population of each relevant energy bin. The darker, continuous line represents all events with energy above 8 EeV, while the dashed lines with lighter shades indicate the energy bins 4 to 8, 8 to 16, 16 to 32 and above 32 EeV. Annotated in each graph, the mean value of the distribution. This is an important quantity as is used for the computation of the dipole components.

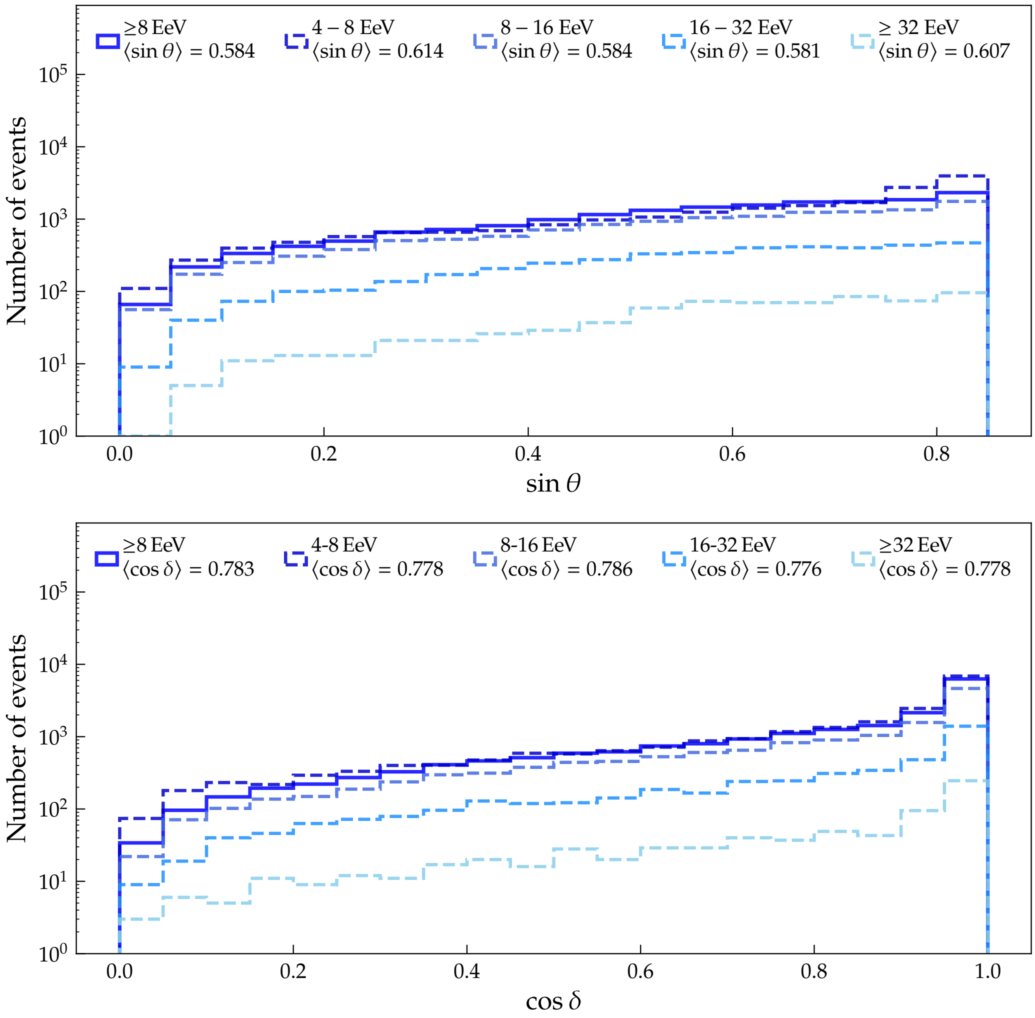


Figure 9.11: Number of events as a function of the sin of the zenith angle θ (top) and as a function of the cosine of the equatorial declination (bottom) of the events in the heavy population of each relevant energy bin. The darker, continuous line represents all events with energy above 8 EeV, while the dashed lines with lighter shades indicate the energy bins 4 to 8, 8 to 16, 16 to 32 and above 32 EeV. Annotated in each graph, the mean value of the distribution. This is an important quantity as is used for the computation of the dipole components.

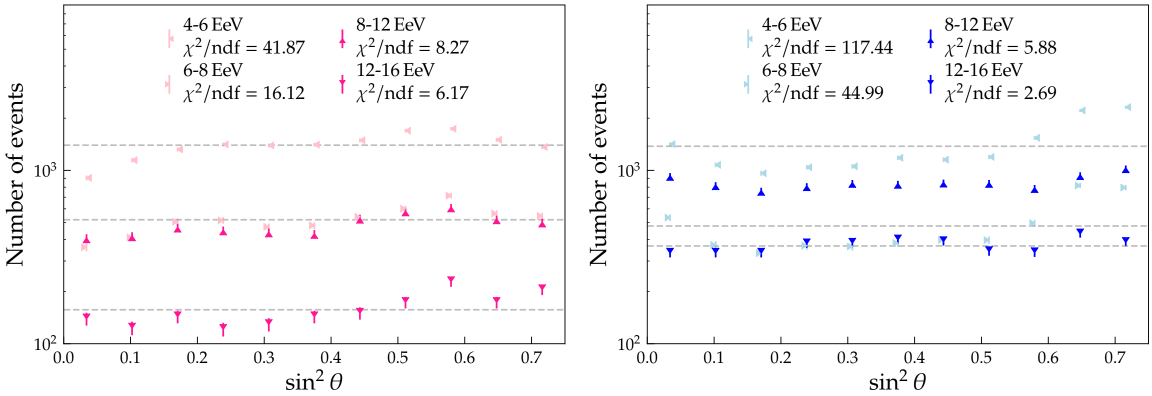


Figure 9.12: Number of events as a function of $\sin^2\theta$ for the light (left) and heavy (right) populations. The progressively smaller deviation from a uniform distribution with increasing energy, indicated by the reduced chi squared quantity, χ^2/ndf, can be an indicative of the composition effect on the energy estimate, which would causes migration of events into the other energy ranges also in mass-dependent manner. The effect is larger at energies closer to the full-efficiency energy threshold of the surface detector array, at 3 EeV. Small modulations in the $\sin^2\theta$ distribution are expected in the presence of a strong dipole.

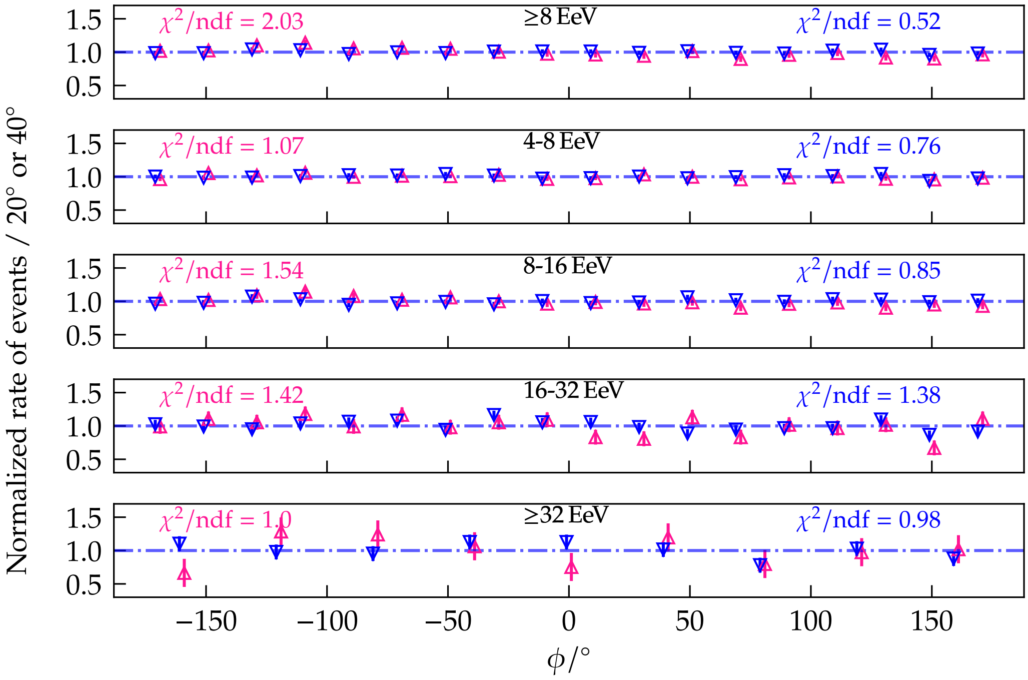


Figure 9.13: The normalized rate of events as a function of azimuth angle for each energy bin of interest. In pink, the events pertaining to the light population, while in blue those of the heavy population. Annotated therein with the same color scheme, the goodness of fit to a uniform distribution, evaluated via the reduced chi squared quantity, χ^2/ndf. Due to small statistic in the highest energy bin, this rate is evaluated in azimuth steps of 40° while this step size is 20° for all other energy intervals.

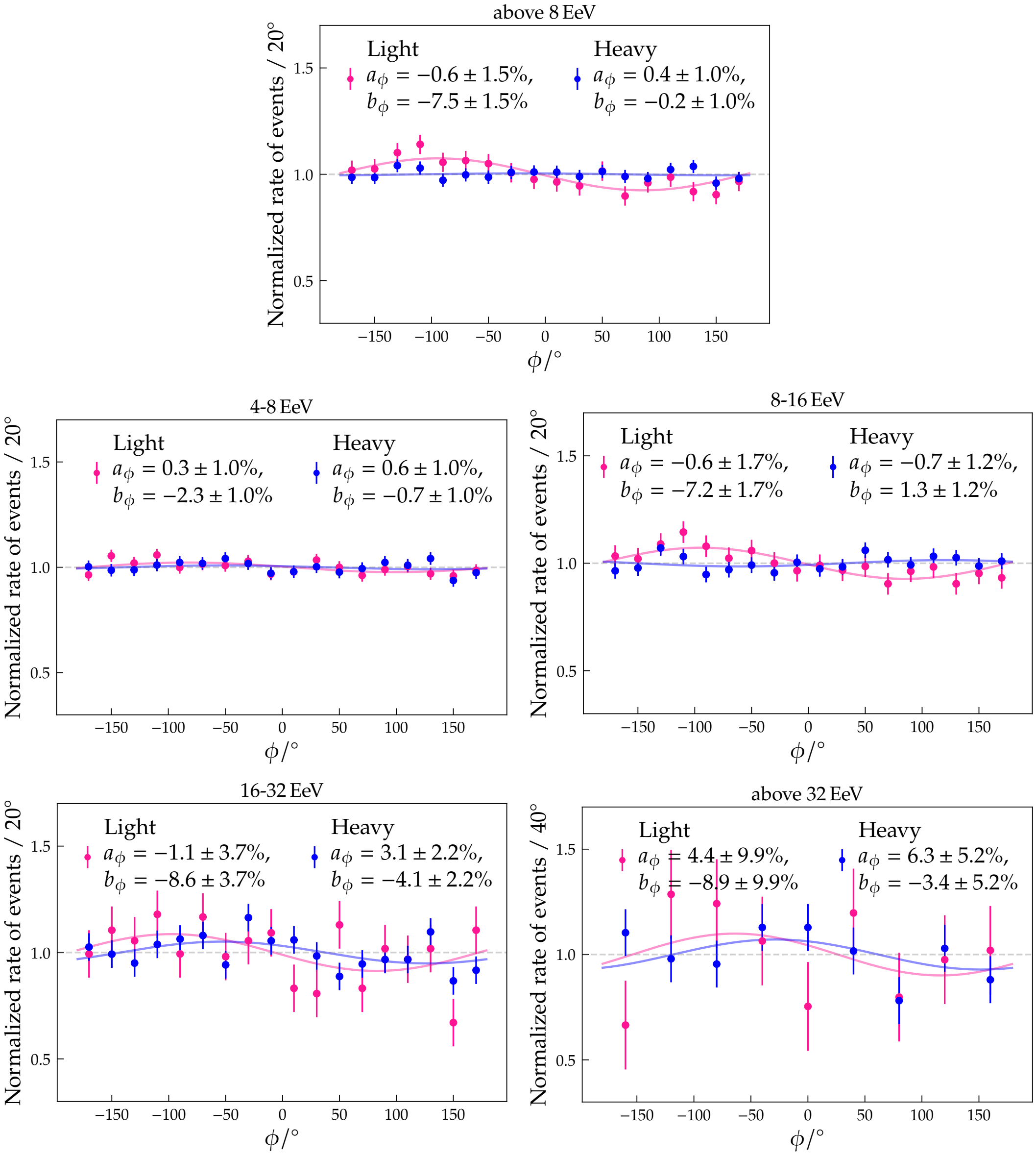


Figure 9.14: Rate of events as a function of the azimuth angle ϕ. Each graph corresponds to the modulation in an energy range of interest: all events above 8 EeV (top), 4 to 8 (middle left), 8 to 16 (middle right), 16 to 32 (bottom left) and above 32 EeV (bottom right). The first harmonic amplitudes of the event rate modulation in azimuth are annotated in each graph, together with the corresponding statistical uncertainty. The pink markers represent the events in the light population, while blue indicates those in the heavy population.

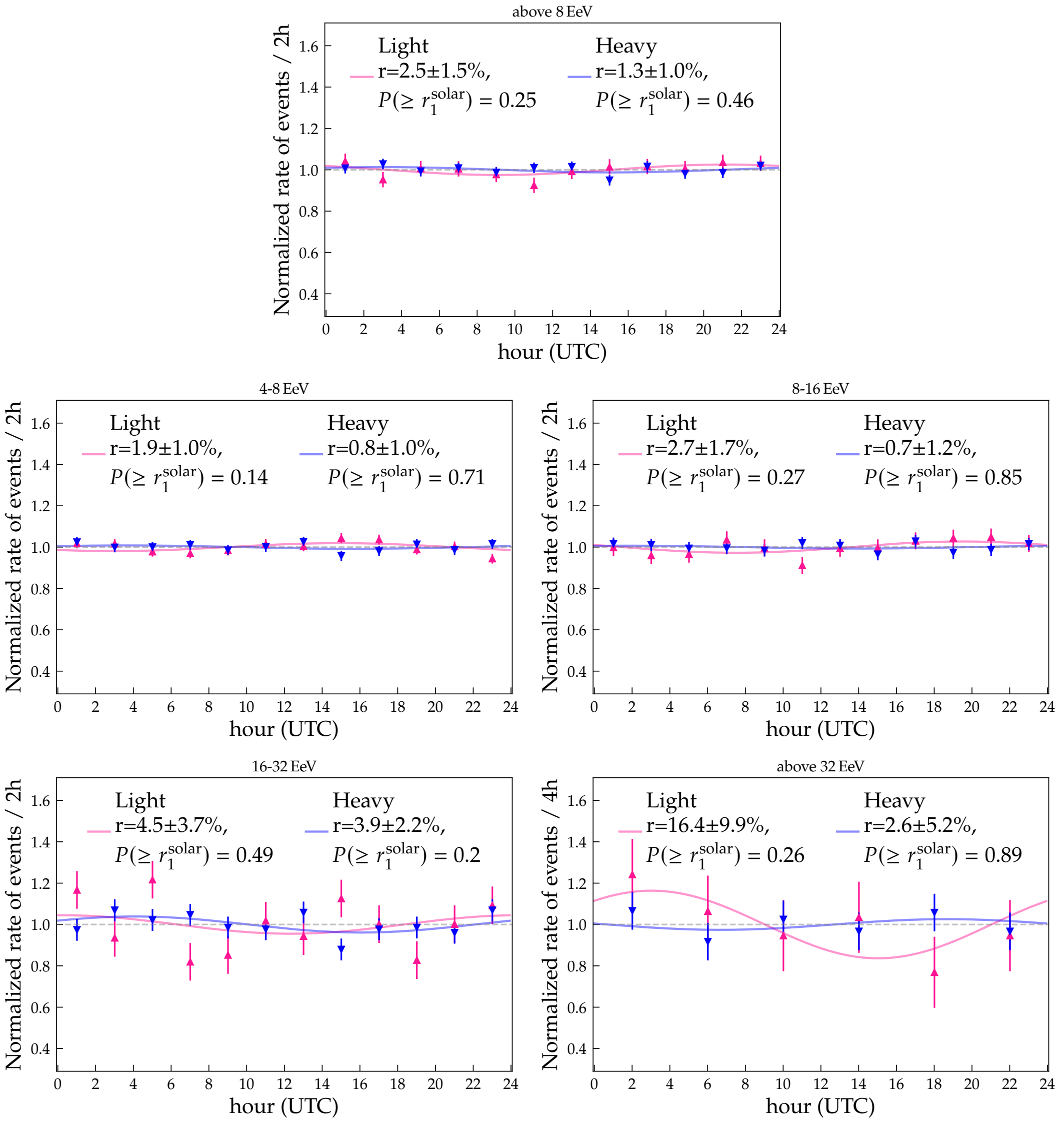


Figure 9.15: Rate of events as a function of the solar hour. Each graph corresponds to the modulation in an energy range of interest: all events above 8 EeV (top), 4 to 8 (middle left), 8 to 16 (middle right), 16 to 32 (bottom left) and above 32 EeV (bottom right). The first harmonic amplitudes of the event rate modulation in solar hour are annotated in each graph, together with the corresponding statistical uncertainty. The pink markers represent the events in the light population, while blue indicates those in the heavy population. The local time in Malargue corresponds to UTC −3.

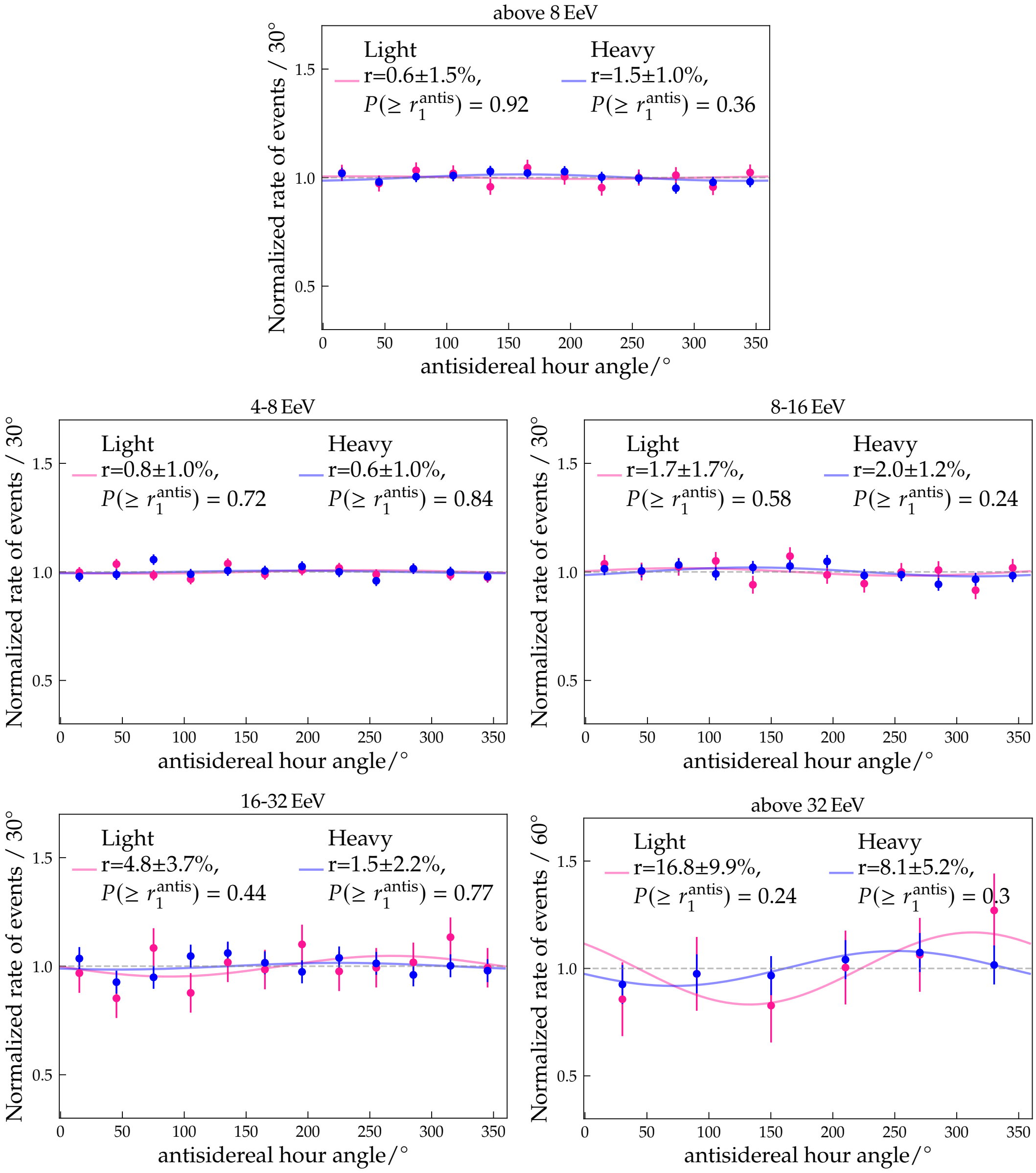


Figure 9.16: Rate of events as a function of the antisidereal hour angle. Each graph corresponds to the modulation in an energy range of interest: all events above 8 EeV (top), 4 to 8 (middle left), 8 to 16 (middle right), 16 to 32 (bottom left) and above 32 EeV (bottom right). The first harmonic amplitudes of the event rate modulation in antisidereal hour angle are annotated in each graph, together with the corresponding statistical uncertainty. The pink markers represent the events in the light population, while blue indicates those in the heavy population.

9.4. SUMMARY

The dipole amplitude of UHECR arrival directions provides the most significant measurement in the 20 years of data taking with the Pierre Auger Observatory. The modulation in the arrival direction, more pronounced in the distribution of the right ascension of the events, relies on the large period of data taking, covering a large portion of the sky. The known effects of the geomagnetic field and of the variations of the atmosphere in the air shower development and consequentially on the reconstruction quality of quantities such as the energy can cause a spurious modulation in the rate of detected events.

In this chapter I evaluated the existence of modulations in the rate of events in different quantities that impact the parallel and perpendicular components of the dipole. This is an important step in warranting the data quality and the interpretation of the measured dipole amplitude being of astrophysical origin.

For the first time in the context of large-scale anisotropy analysis, the Auger Collaboration is migrating from the previous to the most recent framework. Considering a known difference between the frameworks in the energy assignment, small differences could also arise in the full analysis. Therefore, the modulation of the rate of events is evaluated in the energy ranges of interest for the dipole analysis, which are defined according to the energy assigned in each framework. A good agreement has been observed, as expected, and no significant difference in the large-scale anisotropy results should arise thereof.

Applying the threshold values found in Chapter 8 to define the mass-distinct subsets, I also presented in this chapter the number of events in each population and the corresponding distribution in the mass-related quantities: $\ln A$ and X_{max}. For each population, in each energy range, I also evaluated the aforementioned modulations in the rate of events to verify if effects from the geomagnetic field or the atmosphere variability has been left unaccounted for. We can conclude that the modulations are negligible and should not have significant impact on the dipole amplitudes.

In the next chapter, I finally present the results of the mass-informed dipole amplitude. Additional data quality checks that could have a second order impact on the measured dipole amplitude will be reported in Chapter 11.

CHAPTER X

THE FIRST MEASUREMENT OF COMPOSITION-INFORMED DIPOLE AMPLITUDES

10.1. DIPOLE COMPONENTS IN TWO A-DISTINCT POPULATIONS

As described in Section 4.2, the dipole amplitude can be decomposed into two components that are parallel and perpendicular to the Earth's axis, d_z and $d_\perp$, respectively. Eq. (4.3) indicates that while d_z depends mostly on the distribution of events in azimuth angles, $d_\perp$ reflects the modulation in right ascension. We have shown in the previous chapter for each energy range of interest the b_ϕ, $\langle \sin\theta \rangle$, and $\langle \cos\delta \rangle$ quantities. In Fig. 10.1, we can see the normalized rate of events in right ascension for the full data set. The same exercise is repeated for the light and heavy populations.

In agreement with what has been reported in Ref. [69, 70], the modulation in right ascension (RA) with the largest significance occurs at the energy ranges of 8 to 16 EeV and above 8 EeV. This is true both for the full dataset presented in this analysis, with $\theta \leq 60°$, and for the subsets composing the light and heavy populations. While the two-sided significance reported in Ref. [70] has a value of 5.7σ for events with energies between 8 and 16 EeV, the value found here corresponds to 5.1σ. While the amplitudes of the modulation in RA are very similar, the condition of zenith angle $\theta \leq 60°$ limits the dataset to $\sim 78\%$ of the one adopted in Ref. [70], thus reducing its statistical power, reflected in the significance of the measurement.

When evaluating the significance of the amplitudes in the light and heavy populations in the 8 to 16 EeV range, we observe the two-sided significance of 4σ and 2.6σ, respectively. While the amplitude increases by a factor of ~ 1.6 for the light component in comparison to the total data set, again, the number of events plays its role: the light component comprises roughly 24% of the full dataset in that energy interval. We can estimate that the 5σ significance of the light component alone will be reached by 2030, assuming the obtained amplitude remains constant and the rate of events per year per population also remains constant.

A similar evaluation can be held for events with energy above 8 EeV. While the two-sided significance reported in [70] has a value of 6.8σ, the value found here corresponds to 6.3σ. In the light population, the two-sided significance is 4.6σ in

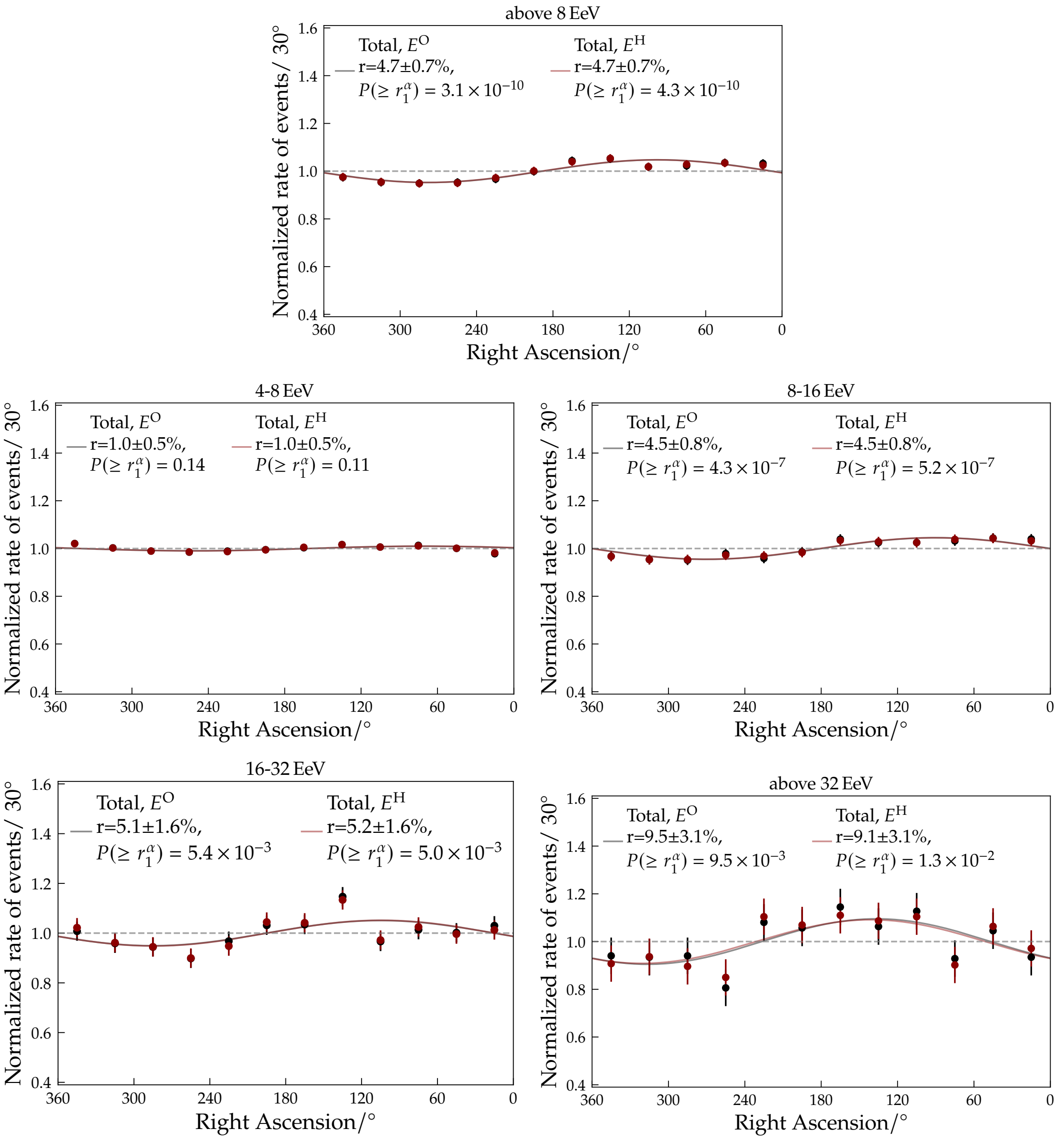


Figure 10.1: Rate of events as a function of the right ascension. Each graph corresponds to the modulation in an energy range of interest: all events above 8 EeV (top), 4 to 8 (middle left), 8 to 16 (middle right), 16 to 32 (bottom left) and above 32 EeV (bottom right). The first harmonic amplitudes of the event rate modulation in right ascension are annotated in each graph, together with the corresponding statistical uncertainty. The chance probability of obtaining an amplitude equal or larger than the one observed, $P(\geq r_1^{\alpha})$, is also annotated. Dark red markers indicate that the events were placed in each energy bin according to the Herald energy estimator, E^{H}, while in black using the Observer energy, E^{O}.

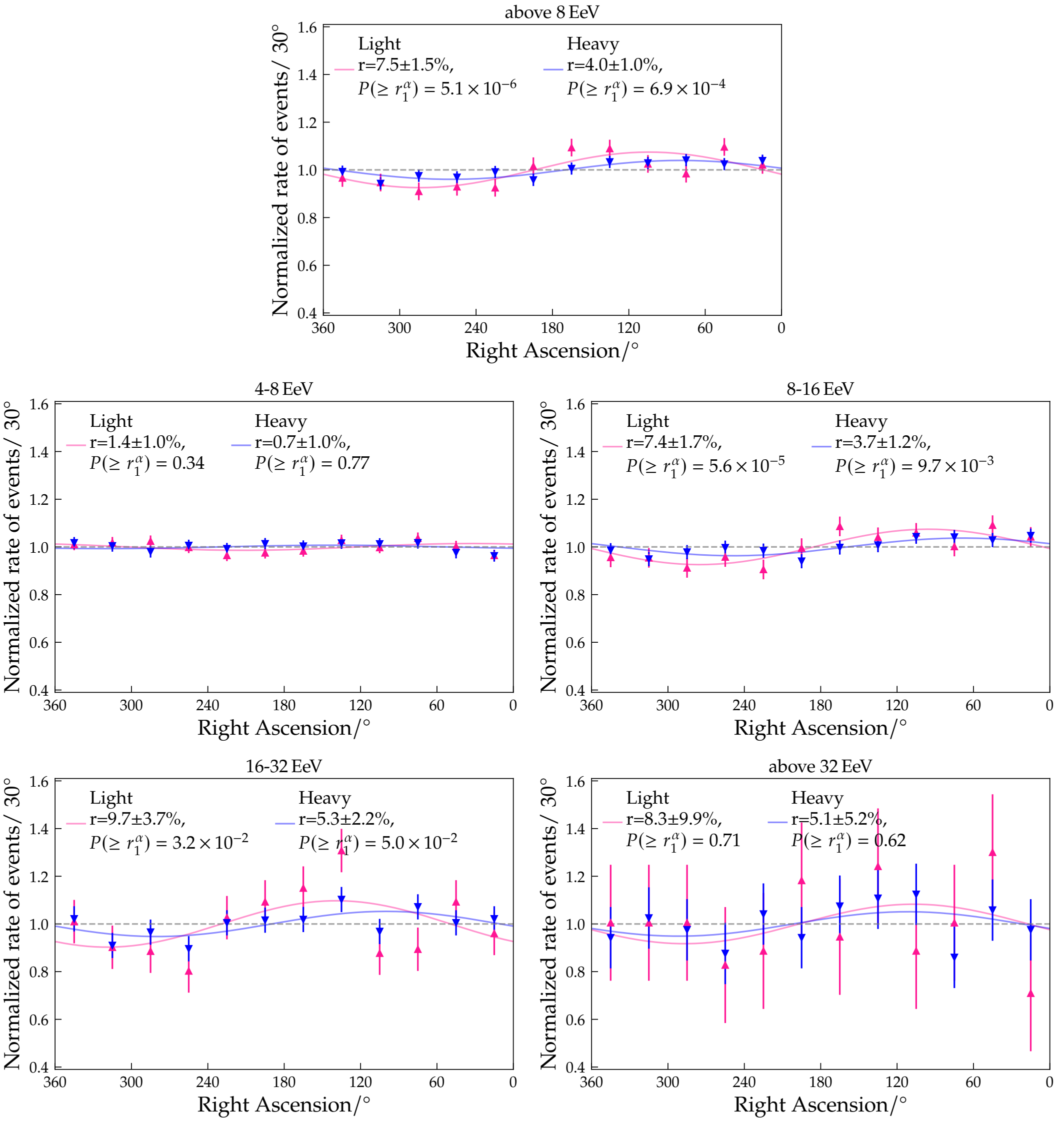


Figure 10.2: Rate of events as a function of the right ascension. Each graph corresponds to the modulation in an energy range of interest: all events above 8 EeV (top), 4 to 8 (middle left), 8 to 16 (middle right), 16 to 32 (bottom left) and above 32 EeV (bottom right). The first harmonic amplitudes of the event rate modulation in right ascension are annotated in each graph, together with the corresponding statistical uncertainty. The chance probability of obtaining an amplitude equal or larger than the one observed, $P(\geq r_1^\alpha)$, is also annotated. The pink markers represent the events in the light population, while blue indicates those in the heavy population.

this energy interval, although the amplitude is again 1.6 times larger than the total dataset presented here. Under the same considerations as before, we can estimate that the 5σ significance could be reached with three additional years of data taking (i.e. including data until December 2025).

However, the reader is reminded that the goal of this thesis is to evaluate the **separation** between the dipole modulations in two A-distinct populations, focusing on the significance of said separation rather than that of the populations individually.

Therefore, using Eq. (4.3), the components of the dipole are computed for each energy range and population. The left-side graph of Fig. 10.3 presents the equatorial dipole amplitude for the light (pink), heavy (blue), and total data (black) sets, computed in each energy range of interest as before: 4 to 8, 8 to 16, 16 to 32, and above 32 EeV. On the right-side graph of the same figure, the parallel component of the dipole, d_z, is shown following the same color scheme. For both graphs, the upper panel indicates the standardized mean difference between the two A-distinct populations as described by Eq. (8.1). We observe that the largest separation occurs in the 8 to 16 EeV range, in the d_z component, reaching a value of 4.08. If one adds in quadrature the SMD value of each energy bin, the combined values are 2.21 and 4.37 for the $d_\perp$ and d_z components, respectively. If only the three energy bins above 8 EeV are considered, these values are 2.15 and 4.22, respectively.

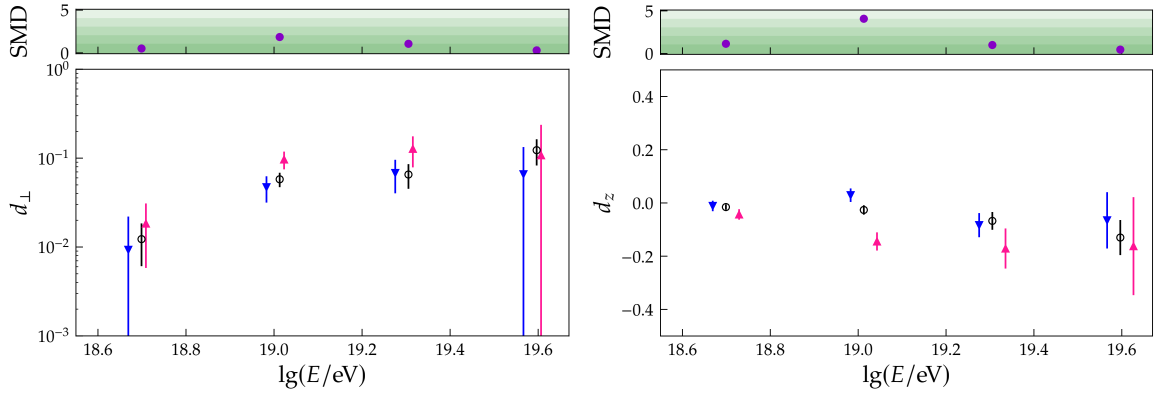


Figure 10.3: Components of the measured dipole amplitude as a function of energy; perpendicular component $d_\perp$ on the left side, and parallel component d_z on the right side. Shown are the amplitudes for the light (pink, up side triangle), heavy (blue, down side triangle), and total (black empty circle) populations. For both graphs, the upper panel indicate the standardized mean difference between the two A-distinct populations as described by Eq. (8.1).

We can notice that the light population presents a larger component $d_\perp$ than the total data set, and larger absolute d_z amplitudes, as well. This corroborates the initial hypothesis of a larger total dipole amplitude in a lighter subset of events in our full dataset. The results found in the novel mass-sensitive analysis indicate a larger separation between these populations in the d_z component, an unexpected feature given the small d_z amplitudes of the full data set. Under more careful consideration, this might indicate a combined effect of the source distribution and of the galactic magnetic field on the charged CR trajectories. A more detailed discussion of the possible interpretation will follow in the next section.

The observed dipole modulations are distinct for the two populations, not only in the amplitude of its components but also in the direction. Thus, an additional way of evaluating such separation is to compute the difference between the dipole vectors of the two subsets. First, each dipole vector is decomposed into three directional components according to

$$\begin{aligned} d_x &= d \times \cos(\delta_d) \times \cos(\alpha_d), \\ d_y &= d \times \cos(\delta_d) \times \sin(\alpha_d), \\ d_z &= d \times \sin(\delta_d), \end{aligned} \tag{10.1}$$

with α_d and δ_d the equatorial coordinates of the dipole direction, as defined in Eq. (4.4).

Then, the difference between the amplitudes in each component is used to compute the amplitude of the vector separation $\bar{r}$:

$$\bar{r} = \sqrt{(d_x^l - d_x^h)^2 + (d_y^l - d_y^h)^2 + (d_z^l - d_z^h)^2} \ , \tag{10.2}$$

where the quantities d_i^l and d_i^h correspond to the light and heavy populations, respectively, with $i = x, y, z$.

Lastly, the probability of obtaining such an amplitude of separation between the two populations in the case they come from a common parent distribution can be computed via Eq. (10.3), as presented and discussed in Ref. [111][1]. There, erfc stands for the complementary error function, and erfi represents the imaginary error function. σ and σ_z stand for the uncertainties in the perpendicular and z directions, respectively.

$$P(\geq \bar{r}) = \mathrm{erfc}\left(\frac{\bar{r}}{\sqrt{2}\sigma_z}\right) + \frac{\sigma}{\sqrt{\sigma_z^2 - \sigma^2}}\mathrm{erfi}\left(\frac{\bar{r}\sqrt{\sigma_z^2 - \sigma^2}}{\sqrt{2}\sigma\sigma_z}\right)\exp\left(-\frac{\bar{r}^2}{2\sigma^2}\right) \tag{10.3}$$

In Fig. 10.4, I present the total dipole amplitude of each subset of events, together with the two-sided significance of measuring such separation in the dipole vectors arising from the statistical fluctuation of two identical populations. The different sample sizes are taken into consideration in the uncertainties, as the uncertainty in the perpendicular direction is taken as $\sigma = \sqrt{(\sigma^l)^2 + (\sigma^h)^2}$, and the uncertainty in the parallel component is $\sigma_z = \sqrt{(\sigma_z^l)^2 + (\sigma_z^h)^2}$. The two-sided significance is computed via

$$\text{significance} = \sqrt{2} \times \mathrm{erfcinv}(P(\geq \bar{r})), \tag{10.4}$$

where erfcinv stands for the inverse of the complementary error function.

The most significant separation is observed in the energy range of 8 to 16 EeV. There, the computed chance probability is 2.9×10^{-5}, corresponding to a significance of 4.18σ. This result is slightly larger than the one obtained for the separation in

---

[1] The reader should notice that the Equation 24 presented in Ref. [111] presented a typo that has been corrected here.

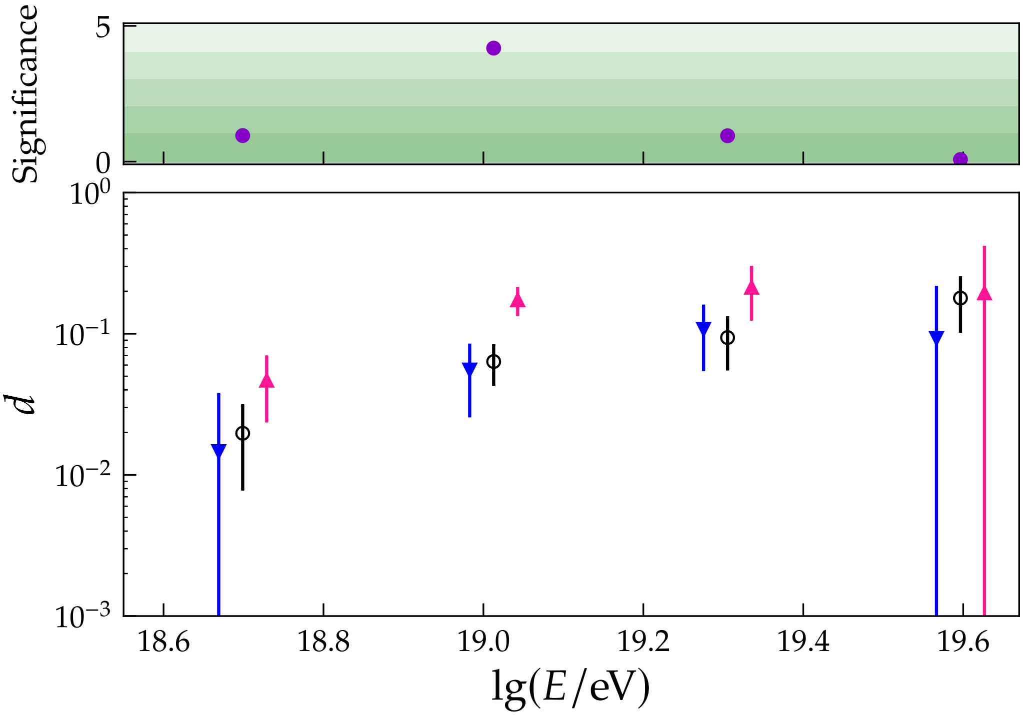


Figure 10.4: Total amplitude d of the dipole in the light (pink), heavy (blue), and total data set (black) as a function of energy. On the top panel, the two-sided significance of the separation between light and heavy populations occurring due to a statistical fluctuation of an isotropic distribution is evaluated according to the chance probability of measuring $\bar{r}$, Eq. (10.3).

the d_z component alone in the same energy bin, indicating the dominance of this component to the total separation presented in Fig. 10.4. If one adds in quadrature the significance of separation in all energy bins, a value of 4.40σ is found. When excluding the energy bin of 4 to 8 EeV, a combined significance of 4.29σ is found.

For completeness, we include in Fig. 10.5 the direction of the measured dipole in terms of its phase (left) and declination (right)in each energy range, and in Fig. 10.6 the direction of the dipole with the curve that corresponds to 68% confidence level in each direction. There, the color scheme is kept: pink, up-side triangle markers represent the light population; blue, down-side triangle markers indicate the heavy population; and the black circles represent the total data set in the corresponding energy range. Irrespective of which coordinate system we evaluate, equatorial in Fig. 10.5 or galactic as in Fig. 10.6, we note a smaller distance between the two populations with increasing energy. This can be understood, on the one hand, due to the rapidly decreasing number of events in each energy range, but also due to the progressively heavier and less mixed composition of UHECR, as indicated by Figs. 5.4, 5.6, 5.12, 9.2 and 9.8.

The relative flux of UHECRs for each population and energy range is depicted in Figs. 10.7 and 10.8, where the direction of the maximum of the measured dipole is also shown. The relative flux is computed as the ratio between the flux of events and the mean flux over the visible sky.

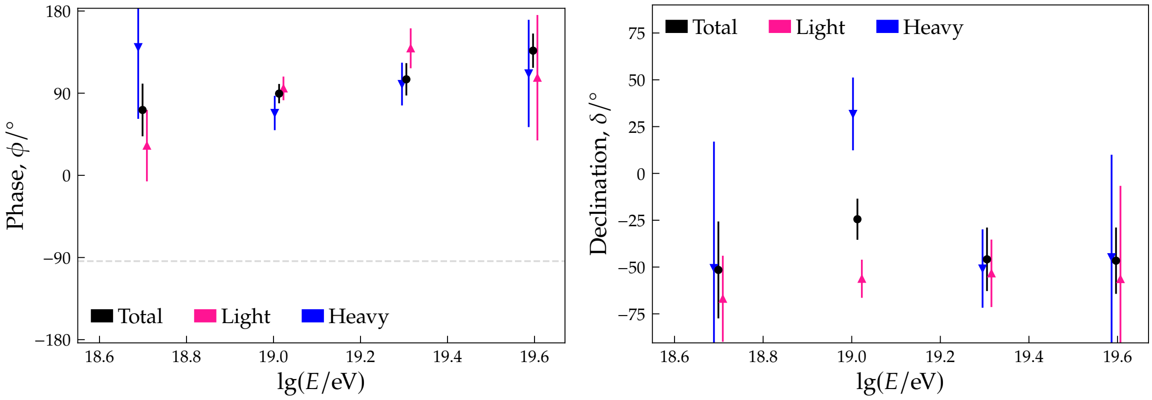


Figure 10.5: The direction of the dipole maxima for the light (pink, up-side triangle), heavy (blue, down-side triangle, and total (black circle) data sets. On the left side, the phase ϕ of the dipole is shown, with the dashed gray line indicating the direction of the galactic center. On the right side, the declination δ is presented.

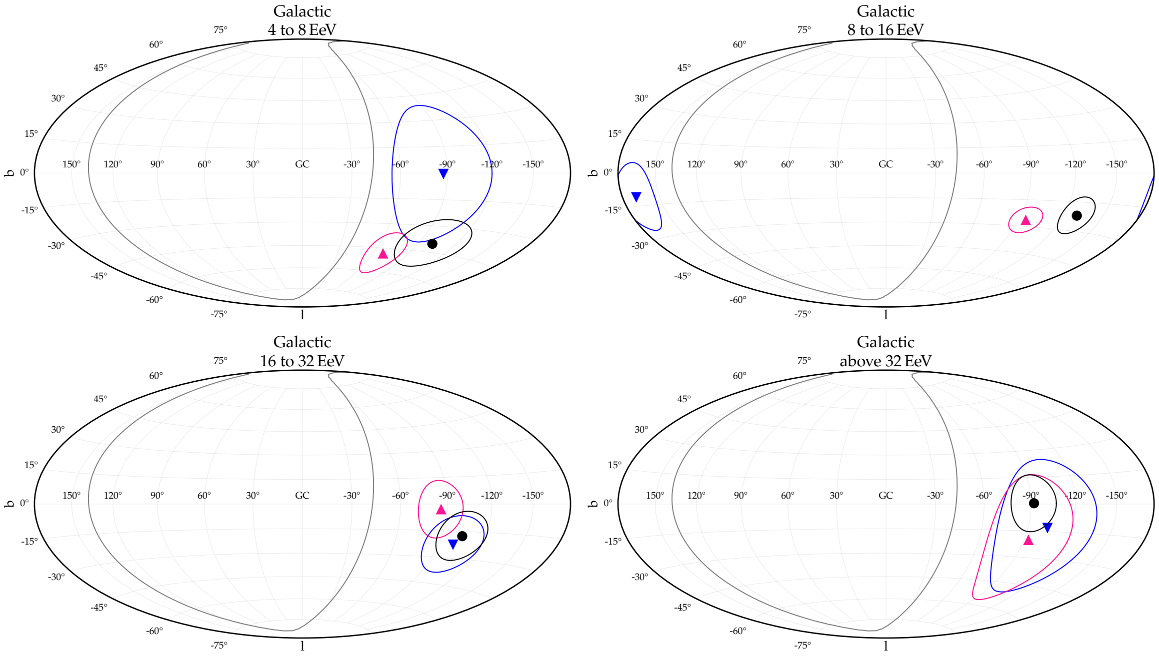


Figure 10.6: The direction of the maximum dipole amplitude in galactic coordinates. Different markers indicate the light (pink, up-side triangle), heavy (blue, down-side triangle, and total (black circle) data sets. The uncertainty curves corresponding to the 68% confidence level are indicated in the same color scheme. Shown are the results for each energy range: 4 to 8 EeV (top left), 8 to 16 EeV (top right), 16 to 32 EeV (bottom left), and above 32 EeV (bottom right).

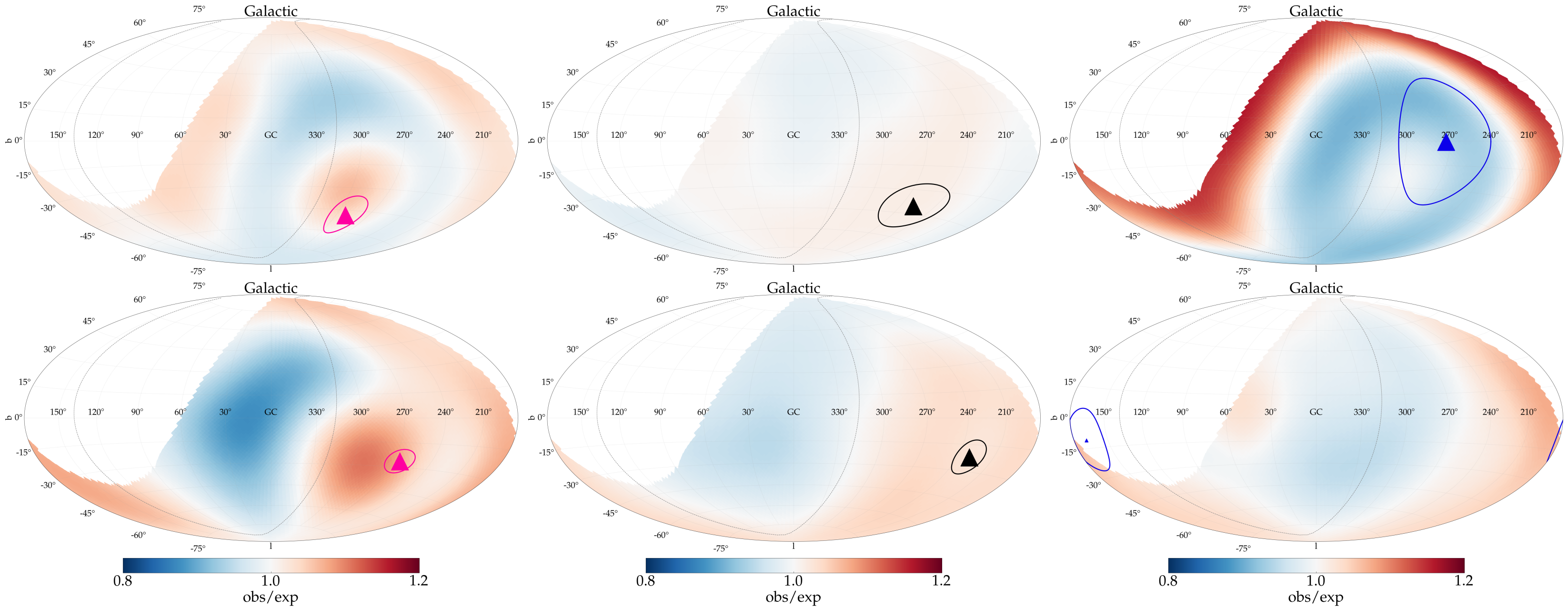


Figure 10.7: Ratio between the flux of events and the mean flux over the visible sky. Corresponds in first approximation to the ratio between the number of observed events and those expected from an isotropic distribution reported in Ref. [70]. In the first row, events with energy between 4 and 8 EeV; and in the second row, those with energy between 8 and 16 EeV. The graphs correspond to the light population (left), to the total dataset (center), and to the heavy population (right).

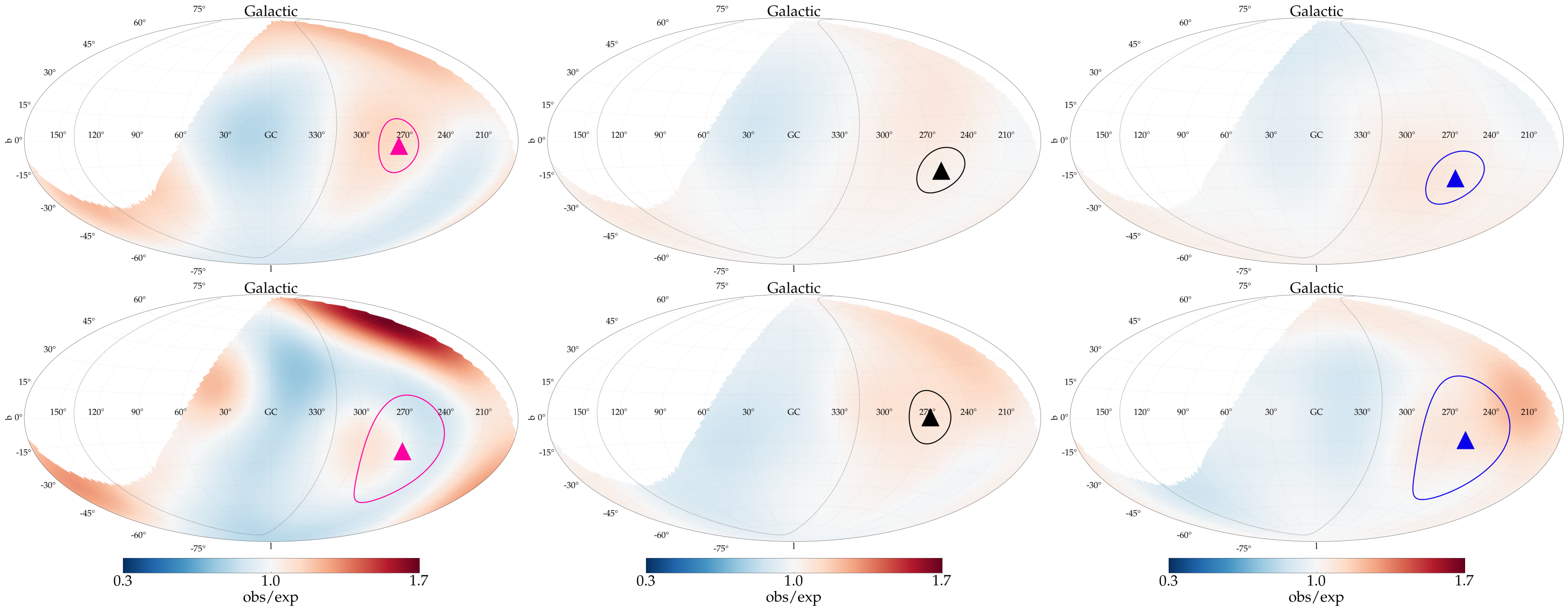


Figure 10.8: Ratio between the flux of events and the mean flux over the visible sky. Corresponds in first approximation to the ratio between the number of observed events and those expected from an isotropic distribution reported in Ref. [70]. In the first row, events with energy between 16 and 32 EeV; and in the second row, those with energy above 32 EeV. The graphs correspond to the light population (left), to the total dataset (center), and to the heavy population (right).

In Fig. 10.7, we observe on the top row a high concentration of flux of CRs for the light (left-side graph) and heavy (right-side) populations, specially closer to the edge of the field of view of the observatory. This effect is closely related to what was reported and discussed in Chapter 9, in special in Fig. 9.12. We observed a relative increase of 5% in the mean sine of the zenith angle of the events in the heavy population with energies between 4 and 8 EeV. A similar although smaller effect is noted in the same energy range for the light component. The excess of events at large zenith angles, illustrated in Fig. 9.12, is also seen in Fig. 10.7 as that corresponds to the border regions (and to the equatorial south pole). Nonetheless, we remind the reader that in the total data set, i.e. prior to the mass-informed selection, the dipole amplitude in that energy range presents a small statistical significance. Even with the systematic effect of excess of events at the large zenith angles, the observed dipole amplitude for the light and heavy populations are compatible to the one for the total data set in that energy range.

The relative flux of UHECRs for each population and energy range in equatorial coordinates can be found in Appendix B., in Figs. 10 and 11.

10.2. DISCUSSION AND INTERPRETATION

In this novel mass-enhanced search for dipole modulations in the arrival direction of cosmic rays, we have found indications that events with lower mass, constituting the *light* population, present with a larger anisotropy than the *heavy* counterpart. This is expected in the scenario of a dipole modulation of extra-galactic origin, where the charged cosmic rays have their trajectories modified by the galactic magnetic field, causing both a dampening of the dipole amplitude and a shift in its direction, as discussed in Section 4.2. However, to clarify the interplay of all factors causing the observed modulations, additional studies outside the scope of this thesis are needed. For example, the dipole declination difference between the two populations in the 8 to 16 EeV energy range could be explained by a scenario where individual sources produce distinct CR composition, by the rigidity-dependent effect of the galactic magnetic field on the CR trajectories, by a combination of these effects or other processes.

We have evaluated the main known sources of spurious modulations that could affect the measured dipole amplitudes. Since no significant modulation has been found in terms of the solar hour and anti-sidereal hour-angle, nor a relevant modulation in the East-West direction[2], we conclude that the measured dipole amplitudes are of astrophysical origin.

The efforts to understand the source of this interesting effect are underway in the Pierre Auger Collaboration. It is clear that the CR composition, along with the galactic magnetic field model and the assumed source distribution and characteristics, affect the measured anisotropies in different scales. The plausible source scenarios have long been investigated by the Collaboration, taking this information into account, for example, in Refs. [20, 21]. The results emerging from the analysis presented

---

[2] Evaluated through the amplitude a_ϕ of the Fourier analysis in azimuth, as shown in Fig. 9.14.

here will hopefully aid in the further development of such analyses and knowledge about plausible source scenarios and galactic magnetic field models. An example of such ongoing analysis is demonstrated in Fig. 10.10. The full description of it has been internally reported to the Collaboration [112]. It is largely based on [113], with the additional definitions of *light* and *heavy* populations. The simulated events are ordered according to the cosmic ray mass and selected such that the populations follow the same percentiles as found in Fig. 9.8. Different galactic magnetic field models can be used in the simulation process, and the effects on the observed dipole anisotropy can be appreciated in comparing left-side and right-side plot of that figure. For the Auger Collaboration members, we invite you to revisit the internal report available in [112].

In another ongoing analysis using the large-scale structure of mass distribution in the Universe, once again, it is possible to evaluate the dipole amplitudes. The interactions with both the background light and magnetic field are considered in the transport of UHECR. Some differences to the previously mentioned ongoing analysis should be commented. Firstly, the authors adopt the source distribution following the observed matter distribution in the universe for sources within 120 Mpc of distance to Earth, while beyond this distance the sources are considered to be uniformly distributed. Additionally, the CR composition and spectrum reported in Ref. [20], are applied to this analysis as benchmark. The simulated events are also ordered according to the cosmic ray mass and selected such that the populations follow the same percentiles as found in Fig. 9.8. The resulting dipole amplitude and directions can be seen in Fig. 10.9.

In a very preliminary discussion of the composition dependence of the measured dipole amplitudes, we remark that irrespective of the galactic magnetic field adopted, a separation in terms of amplitude is expected when selecting a parcel of the lightest events and another of the heaviest ones, in the four energy ranges of interest.

It is important to remark on the advantage of having used a simulation-based strategy to set the parameters defining the A-distinct populations since this allowed us to avoid penalization on the results due to extensive scanning of the parameters that would be otherwise needed. Thus, the significance of the results reported in the previous section is a strong indicator of this rigidity-/mass-dependence of the measured dipole amplitudes, while further data-taking is needed to reach a potential discovery.

As data taking in AugerPrime begins, the potentiality of mass-sensitive analyses increases. In AugerPrime, the cosmic ray mass estimators can be further informed by measurements of the SSD and the UMD, as described in Chapter 3, allowing for a more precise identification of the shower-initiating nucleus. This way, an analysis similar to the one proposed here could be designed for the AugerPrime data set, taking into consideration the improved mass estimator. Analyses in large- and small-scale anisotropies in that context can further aid in our understanding of source attributes and distribution and on the galactic magnetic field characteristics, for example. Full simulation chains, from the ejection of CR at sources with a given distribution, to the propagation in the extra-galactic background light and in magnetic fields, are currently being developed and are one of the approaches to

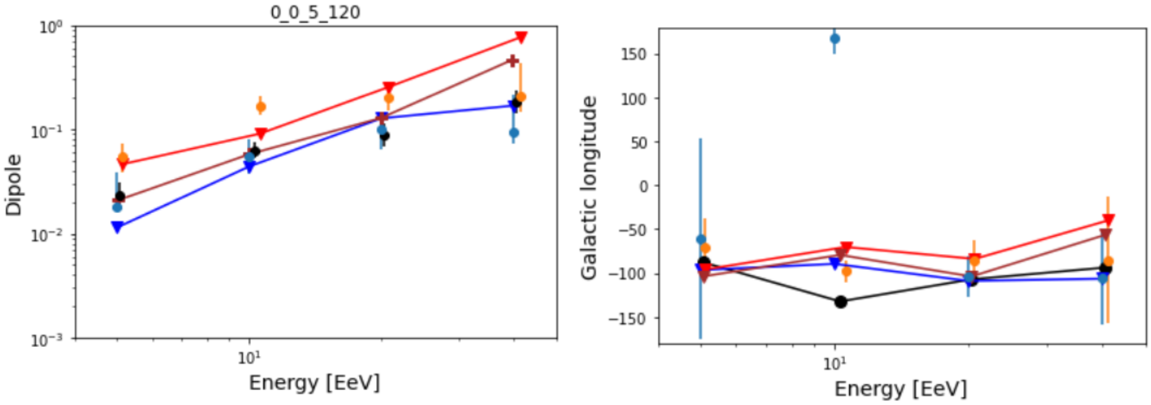


Figure 10.9: On the left side graphs, the measured dipole amplitudes are shown for a *light* (orange) and a *heavy* (light blue) population together with the expectations of dipole amplitude when the UHECR sources are distributed according to the large scale matter distribution in the Universe (based on the 2MASS measurements[30]). The expected amplitudes from the simulations are shown in red and dark blue, for the light and heavy populations respectively. The populations were set in the simulations by reproducing the percentage of measured events falling in each of these populations. More information on the text. On the right side plots, the galactic longitude of the maximum dipole amplitude are presented as a function of the energy. The composition of CR for this simulation study corresponds to that in Ref. [20], and the source distribution has a density of 10^{-5}Mpc^{-1}, selected from the 2MRS catalog up to 120 Mpc and uniformly distributed at larger distances and within the Galactic plane mask.

interpreting large- and intermediate-scale anisotropies results.

The analysis presented in this thesis indicates that already in Phase I of the Pierre Auger Observatory, one can add composition information to anisotropy searches and evaluate the effect of such on the measurements. The inclusion of CR mass information will open doors to a number of other studies, each potentially allowing us to further our understanding of the plausible sources of UHECR, the characteristics of the galactic magnetic field, and other yet unimaginable considerations. For example, the auto-correlation and catalogue-based searches described in Sections 4.3 and 4.4, respectively, can be once again evaluated but now using the light subset of events, which presents a larger mean rigidity than the total dataset. An intermediate-scale analysis using the composition information is currently being proposed and evaluated within the Collaboration. Other possibilities of mass-informed analyses are related to the transition of galactic to extra-galactic origin of the majority of UHECRs. For example, a search for the correlation of heavy events with the galactic plane at a few EeV could provide insightful information. There are many possibilities, and the near future of mass-informed anisotropy searches is exciting.

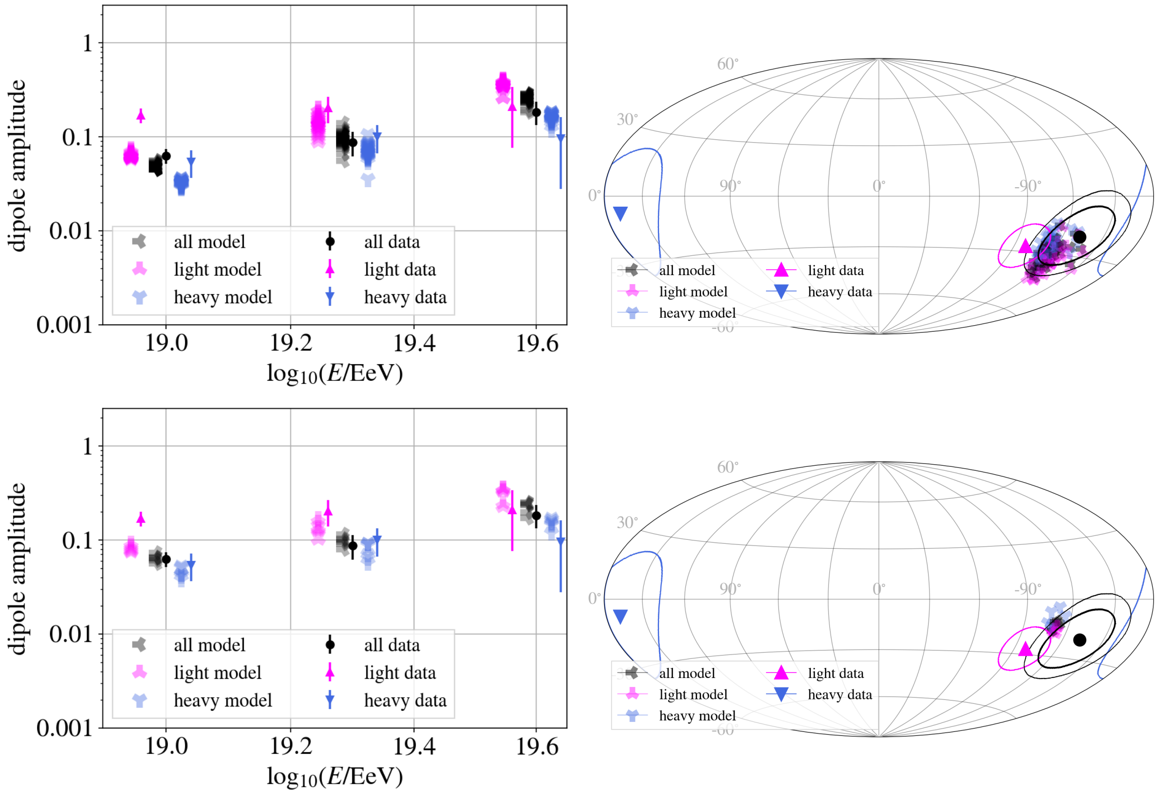


Figure 10.10: On the left side graphs, the dipole amplitudes are shown for a *light* (pink) and a *heavy* (blue) population when the UHECR sources are distributed according to the large scale matter distribution in the Universe (based on the 2MASS measurements[30]). The populations were set in the simulations by reproducing the percentage of measured events falling in each of these populations. More information on the text. The measured dipole amplitudes are shown in full triangle markers, while the best-fitting of 10 ,000 simulation realisations are shown with the translucent, y-shaped markers, following the same color scheme. On the right side plots, the direction of the maximum dipole amplitude in the 8 to 16 EeV energy range is shown for each data set in galactic coordinates, with the curves representing the 68% confidence level regions; the best-fitting simulations realisations are also shown, following the same color and markers scheme. The composition of CR for this simulation study was reported in Ref. [113], as well as all the parameters describing the sources of CRs. When separating the simulated events in two populations representing the same percentiles as in data, a separation in the dipole amplitudes is also observed. Different expectations could arise from a different composition and from different galactic magnetic field. On the first row, the results were obtained using the UF23 [15] base galactic magnetic field model, and on the second row, the JF12 [79] model. Presented internally to the Collaboration [112].

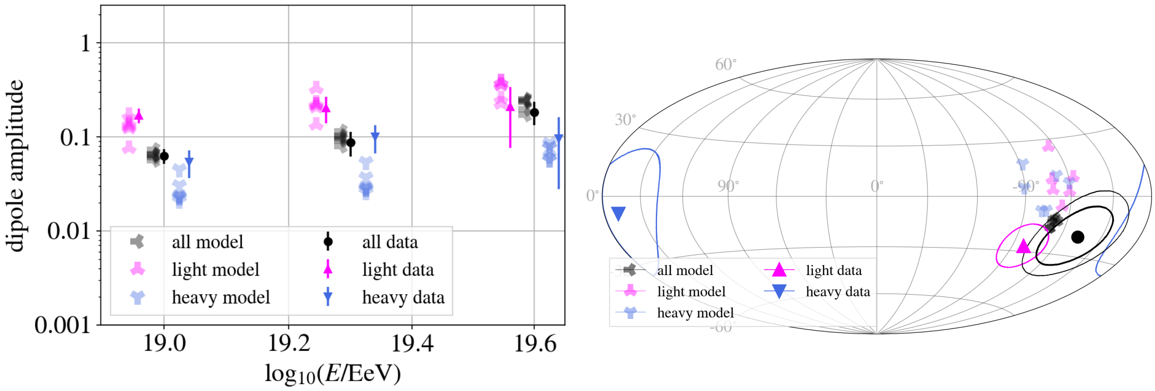


Figure 10.11: Same as in Fig. 10.10, however with a change in composition in the 8 to 16 EeV range. The composition found in Ref. [21] was modified to include a ∼ 10% proton contribution to the CR flux in that energy interval, in agreement to Ref. [26] and to the extended Auger mix. The consequence is an increase of the expected dipole amplitude of the light population, in a better agreement to the measurement. Presented internally to the Collaboration [112].

CHAPTER XI

SUMMARY AND CONCLUSIONS

In this work we have proposed, tested and applied a novel approach for the search of the composition signature on the dipole anisotropies measured in the Phase I of the Pierre Auger Observatory. From the measurements by the Auger Collaboration, we have learned that the composition of UHECR progresses towards heavier elements for energies above ~3 EeV, that the dipole amplitude increases with energy and the direction of this anisotropy points ~115° away from the galactic center for events with energy above 4 EeV. It has also been reported that the spectrum of cosmic rays presents features, indicating changes on the rate of events; out of which we highlight the *ankle* at ~ 5 EeV. A significant anisotropy in a dipole pattern is seen in the energy ranges of 8 to 16 EeV, a region of energy which encloses another feature in the spectrum, at 13 EeV, where the spectral index goes from 2.51 to 3.05, i.e. the rate of events per year, area, solid angle, and energy, is becoming smaller for $E > 13$ EeV. The intermediate-scale anisotropy searches at energies $E \geq 38$ EeV have probed the correlation to potential UHECR sources and regions of interest. And in this direction, from the simulation studies of the potential sources of such anisotropies we could evaluate the impact of the sources features and of the magnetic field on the observable anisotropy at Earth. The spectral features, the nuclei abundance in the UHECR flux and the anisotropy measurements seem to be correlated. Our current understanding of processes that impact CR trajectories, such as the interaction of CRs with the extra-galactic background light and with magnetic fields (galactic and extra-galactic), depends on both the energy and the mass/charge of the nuclei, and impacts the anisotropy measured at Earth. Thus, the inclusion of the composition information is the next necessary step for a better understanding of the interplay of effects that culminate in the observed anisotropies.

In order to probe the sensitivity of the current Auger measurements to a composition signature on the large-scale anisotropy, I have introduced an air shower simulation library that is compatible with both the spectrum and the distribution of depths of shower maxima development, $X_{\max}$, measured in the Pierre Auger Observatory. Next, I proposed a model to describe the rigidity dependence of the measured dipole amplitude. It is a simple, semi-empirical and source-independent model, that also does not directly depend on effects of magnetic field or any propagation effect, but is dependent on the composition model adopted. Then, I used the simulation library to find the criteria that define two mass-distinct subsets and allows for the largest possible separation in dipole amplitude, under the assumptions made. After the scrutiny of the Auger Collaboration, the analysis has been approved to be evaluated on the measured data set, which was followed by several tests, checking spurious modulations. Since no indication of a persisting spurious modulation that could fabricate a dipole pattern was observed, we have applied the harmonic analysis in right ascension and azimuth, and computed the dipole amplitude in the two composition-distinct subsets of the measured data.

In this work we have observed for the first time a strong indication of a composition dependence of the dipole amplitude measured in Phase I of the Pierre Auger Observatory. The significance of the separation in the dipole amplitude of two A-distinct subsets of the data reaches values of $\sim 4\sigma$ for the the events with energies between 8 and 16 EeV. These are exciting results that propel the Pierre Auger Collaboration to upcoming mass-informed analyses. The recent upgrade of the Observatory to the AugerPrime phase, with additional detection methods and potentially more accurate composition information, indicate that the next decade of data-taking with the Auger Observatory will render interesting results, indubitably fostering progress in our knowledge of the sources of ultra-high-energy cosmic rays.

Appendices

A. FURTHER COMPATIBILITY TESTS

It is important to evaluate the expected small differences between the frameworks within the large energy bins used in the large scale analysis. For that, it is then necessary to compare the events which not only are reconstructed by both frameworks, but with the additional condition that the events fall within the same energy bin. In Fig. 1, we evaluate the relative energy differences between the frameworks in each large energy bin, also highlighting the effect of saturated stations on the energy reconstruction. Events with one or more saturated stations are represented in black in Fig. 1, while the events without any saturated station are represented in dark red.

Distinct to the Herald framework, the Observer framework allows for the signal recovery in saturated stations, and thus allows for a better energy estimate in these cases. The number of events with one (or more) saturated station increases with increasing energy, and the aforementioned effect becomes more frequent, as can be seen in the right-hand panels of Fig. 1.

Nonetheless, the mean relative energy difference between frameworks is smaller than the statistical uncertainty and smaller than the absolute energy scale, for all energy bins. For the events with a saturated station in the energy ranges 16 to 32 EeV and above 32 EeV, there seems to exist a small dependency on the zenith angle of the event.

Another crucial quantity is the arrival direction of CRs. In Figs. 2 and 3, we evaluate the angular difference between the arrival directions of CRs between the two frameworks, in the relevant energy intervals. In order to do so, we now take the Observer energy estimator to separate the events into energy bins and exclusively compare the differences in arrival direction between frameworks.

Highlighting the events with (black) and without (dark red) a saturated station, we can interpret that there is no strong influence on the shower plane reconstruction and thus on the arrival direction when including events with a saturated station. The largest deviation is observed for the lowest energy bin of 4 to 8 EeV, where 2.5% of the events without a saturated station have an angular difference larger than 1°, and 8.2% of the events with a saturated station have an angular difference larger than 1°. Annotated in Fig. 2 are also the mean declination values in each energy range for those events with (black) and without a saturated station. A good agreement is observed, with differences of the order of the previously quoted angular resolution.

Analogously to Fig. 2, in Fig. 3 we evaluate the same angular difference but now

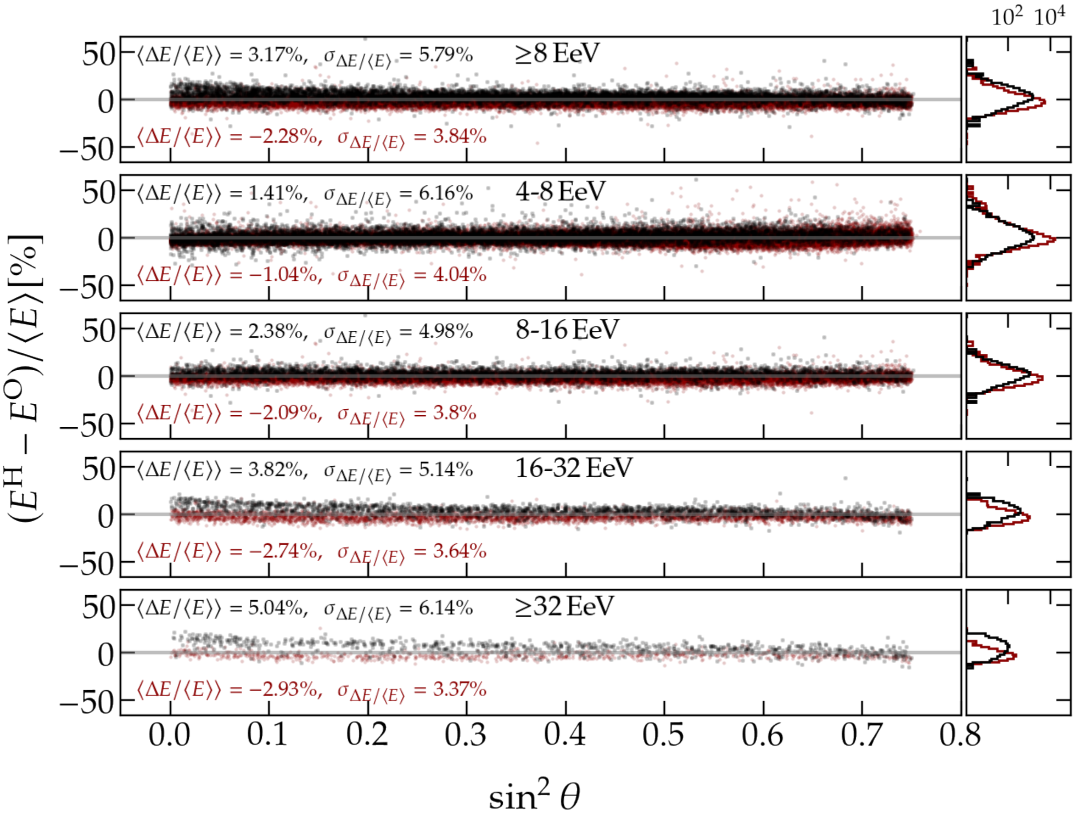


Figure 1: Difference of energy in Herald and Observer frameworks relative to the mean of both estimators. The relative difference is shown as a function of the squared of the sine of the zenith angle, in the left panels. On the right panels, the histogram of such relative differences. In black markers, the events with one or more saturated stations; in dark red markers, those without a saturated station. From top to bottom, each row indicate an energy range: above 8 EeV, 4 to 8, 8 to 16, 16 to 32 and above 32 EeV. Annotated and color coded, the mean and standard deviation of the relative energy difference between frameworks in the corresponding energy range with (black) or without (dark red) a saturated station.

in terms of the right ascension of the events. Annotated therein are the mean RA values in each energy range for those events with (black) and without a saturated station. A general agreement is observed, with differences increasing in the highest energy ranges, in a similar trend to Fig. 1.

It is interesting to notice that, although the rate of events with at least one saturated station increases at higher energies, the agreement between frameworks consistently improves with increasing energy. Thus, the small zenith dependency observed at the highest energies in Fig. 1 must not be correlated to the small angular distance between AD in the two frameworks in Fig. 2. Nonetheless, we can show that the small differences in terms of RA presented in Fig. 3 does not imply a fictitious modulation

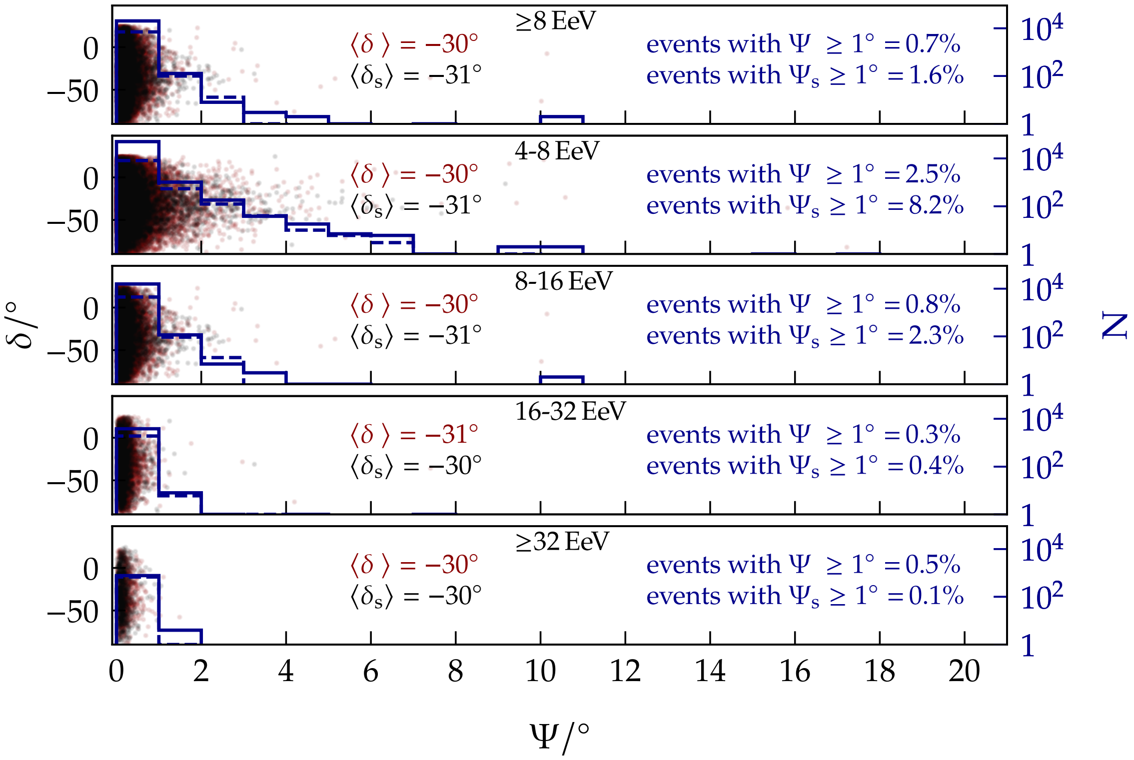


Figure 2: Angular difference of arrival direction in Herald and Observer frameworks as a function of the equatorial declination of the event (in the Observer framework). In black markers, the events with one or more saturated stations; in dark red markers, those without a saturated station. On the right-hand axis, the histogram of events as a function of the angular difference, with bin width of 1°. Continuous lines represent the events with one or more saturated stations; while dashed lies, those without a saturated station. From top to bottom, each row indicate an energy range: above 8 EeV, 4 to 8, 8 to 16, 16 to 32 and above 32 EeV. Annotated are the percentage of events in the corresponding energy range that have an angular difference larger than 1°. To differentiate the subset of events with and without a saturated station, we use Ψ_s and Ψ, respectively. Analogously, the mean right ascension is annotated on the left side with $\langle\delta_s\rangle$ and $\langle\delta\rangle$ the mean value for the events with and without a saturated station.

in RA by the introduction of events with a saturated station. To do so, we compare the equatorial dipole amplitude for the events with and without a saturated station in each energy bin, on the left plot of Fig. 4. The effects of a saurated station on the d_z component of the dipole is shown on the right plot of the same figure.

The stability of detection is also crucial, specially for large-scale anisotropy studies. In Fig. 5 we compare the rate of events per year observed with each framework, showing the full compatibility in this aspect. On the other hand, the time stability of a certain mass group is expected in non-transient scenarios. Therefore I present in Fig. 6 the number of events per year in the light and heavy subsets of the data, for each of the four energy bins. Apart from the initial growth in the number of

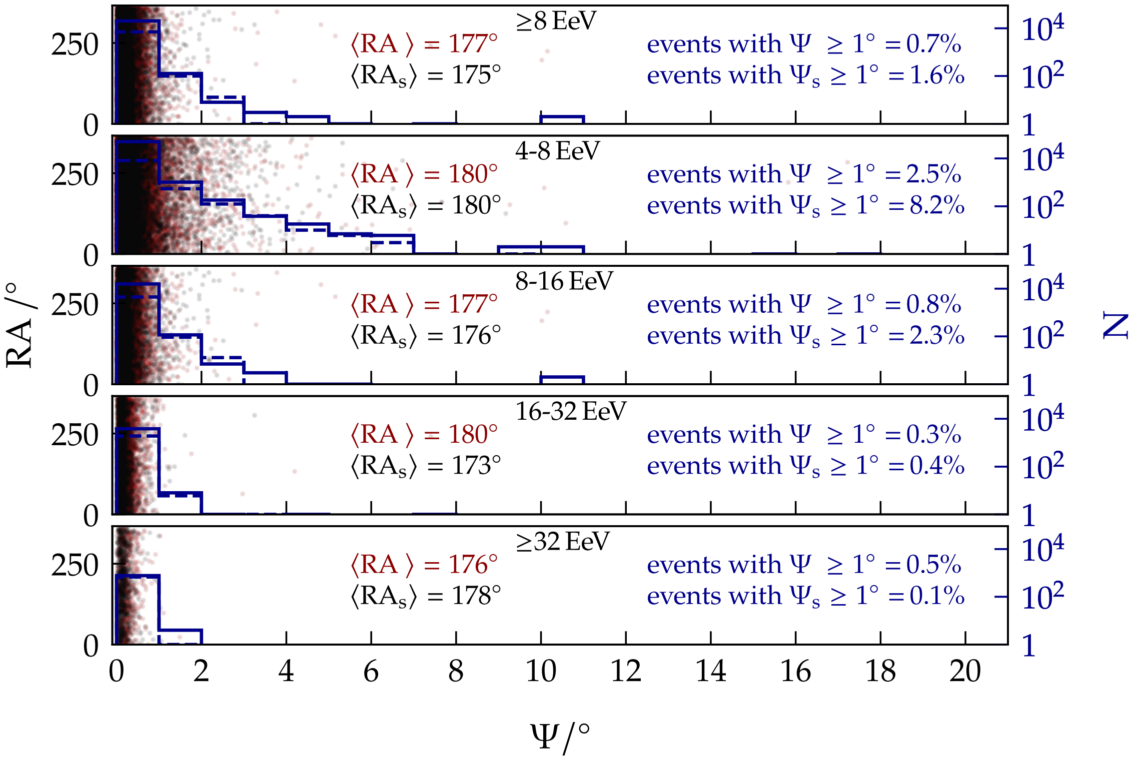


Figure 3: Angular difference of arrival direction in Herald and Observer frameworks as a function of the right ascension of the event (in the Observer framework). In black markers, the events with one or more saturated stations; in dark red markers, those without a saturated station. On the right-hand axis, the histogram of events as a function of the angular difference, with bin width of 1°. Continuous lines represent the events with one or more saturated stations; while dashed lies, those without a saturated station. From top to bottom, each row indicate an energy range: above 8 EeV, 4 to 8, 8 to 16, 16 to 32 and above 32 EeV. Annotated are the percentage of events in the corresponding energy range that have an angular difference larger than 1°. To differentiate the subset of events with and without a saturated station, we use Ψ_s and Ψ, respectively. Analogously, the mean right ascension is annotated on the left side with $\langle \mathrm{RA_s} \rangle$ and $\langle \mathrm{RA} \rangle$ the mean value for the events with and without a saturated station.

events, a consequence of the growth of the array, the rate of events appears to be fairly constant.

An important test to be performed is on the existence of spurious modulations of the mass estimator, as it is used to define the populations. Any modulation on the mass estimator itself would introduce a modulation in the arrival direction information and could potentially introduce a fictitious dipole pattern. We evaluate in Figs. 7 and 8 the distribution of the mass estimator in the two selected populations as a function of the azimuth angle and of the solar hour, as these are indicatives of unaccounted for corrections in the geomagnetic field and atmospheric effects. No

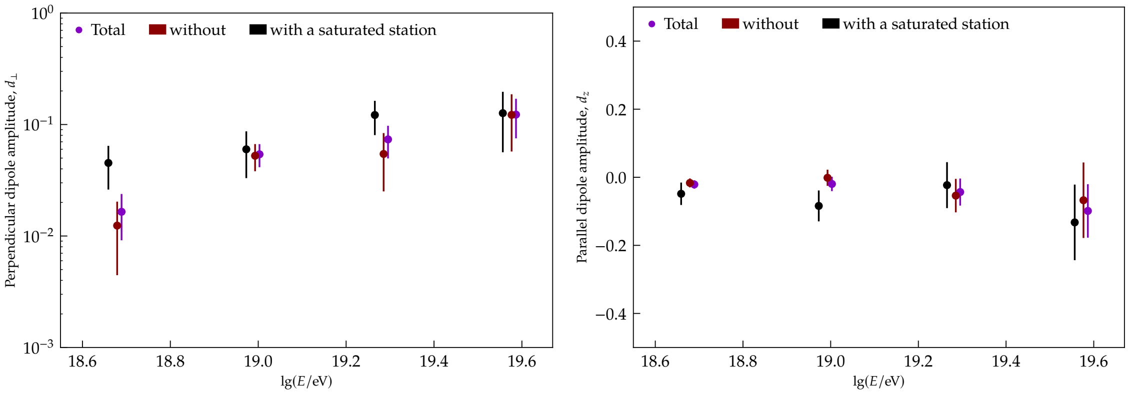


Figure 4: Dipole amplitude components in the perpendicular (left) and parallel (right) directions. The events without a saturated station, in dark red, are dominant and only a few percent of the selected events present a saturated station, shown in black markers. The rate of saturated events is seen in Fig. 6.1. Specially at lower energies, the events with a saturated station present a larger discrepancy in the dipole components with regards to the events without a saturated station.

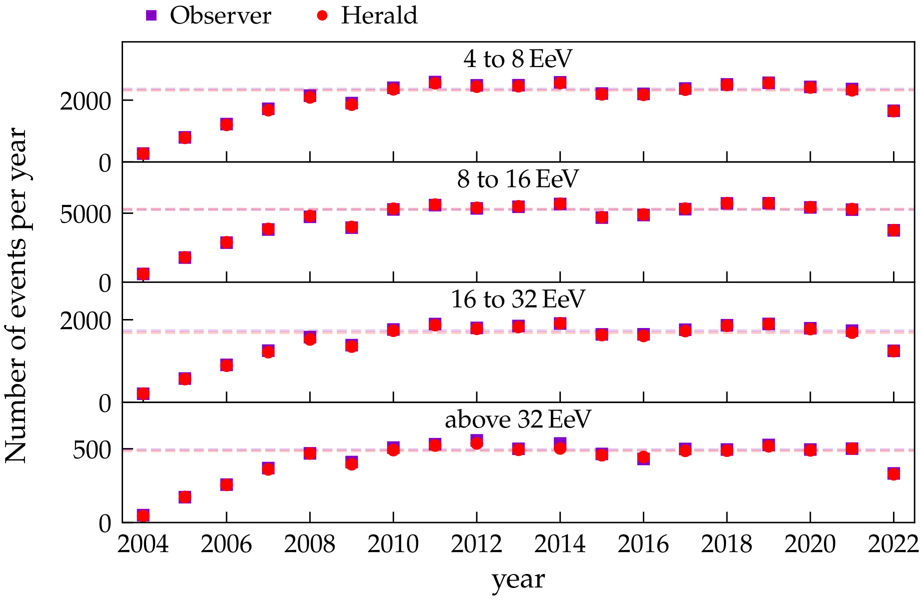


Figure 5: Number of events recorded per year. Each panel is representative of an energy range of interest to the large scale anisotropy. Observer (purple squares) and Herald (red circles) provide slightly different energy estimators. The rate of events per year is fairly compatible between both frameworks, and is remarkably constant between 2008 and 2021. In the early years, the detecting stations were being deployed, therefore the growth in teh rate of events. The deployment of stations for AugerPrime included the change of electronics. The events compatible with AugerPrime are not included in this dataset, and therefore we observe a decease in the rate of events in all energy ranges.

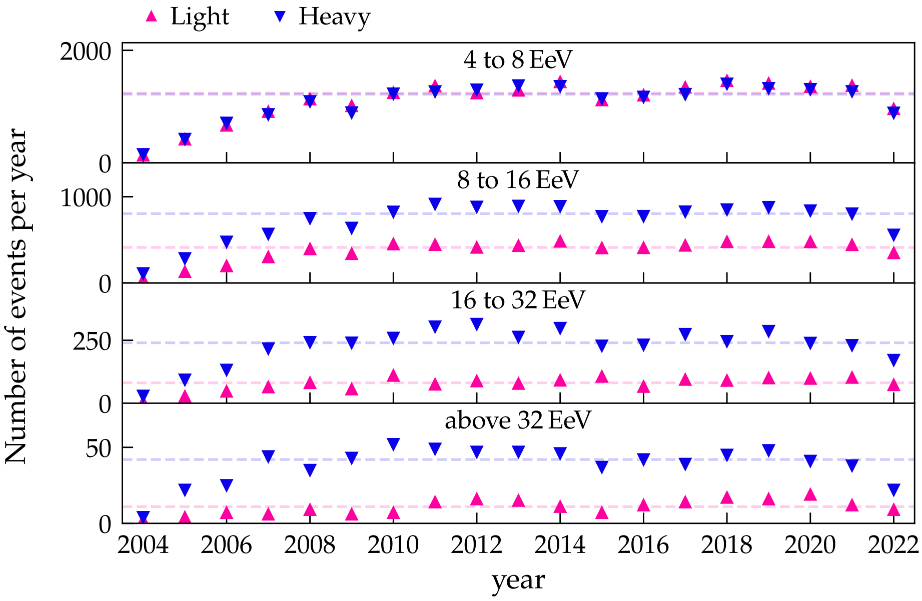


Figure 6: Number of events recorded per year in each mass-informed subset. Each panel is representative of an energy range of interest to the large scale anisotropy. Light (pink upwards triangles) and Heavy (blue downwards triangles) provide slightly different energy estimators. The rate of events per year is fairly compatible between both frameworks, and is remarkably constant between 2008 and 2021. In the early years, the detecting stations were being deployed, therefore the growth in the rate of events. The deployment of stations for AugerPrime included the change of electronics. The events compatible with AugerPrime are not included in this dataset, and therefore we observe a decease in the rate of events in all energy ranges.

strong modulation is observed. Further studies are needed to quantify the effect of this negligible modulation on the dipole components. Due to time constraints, this has not been included in this thesis, however will be included in the upcoming peer-reviewed publication of the mass-dependent dipole amplitude search developed in this thesis.

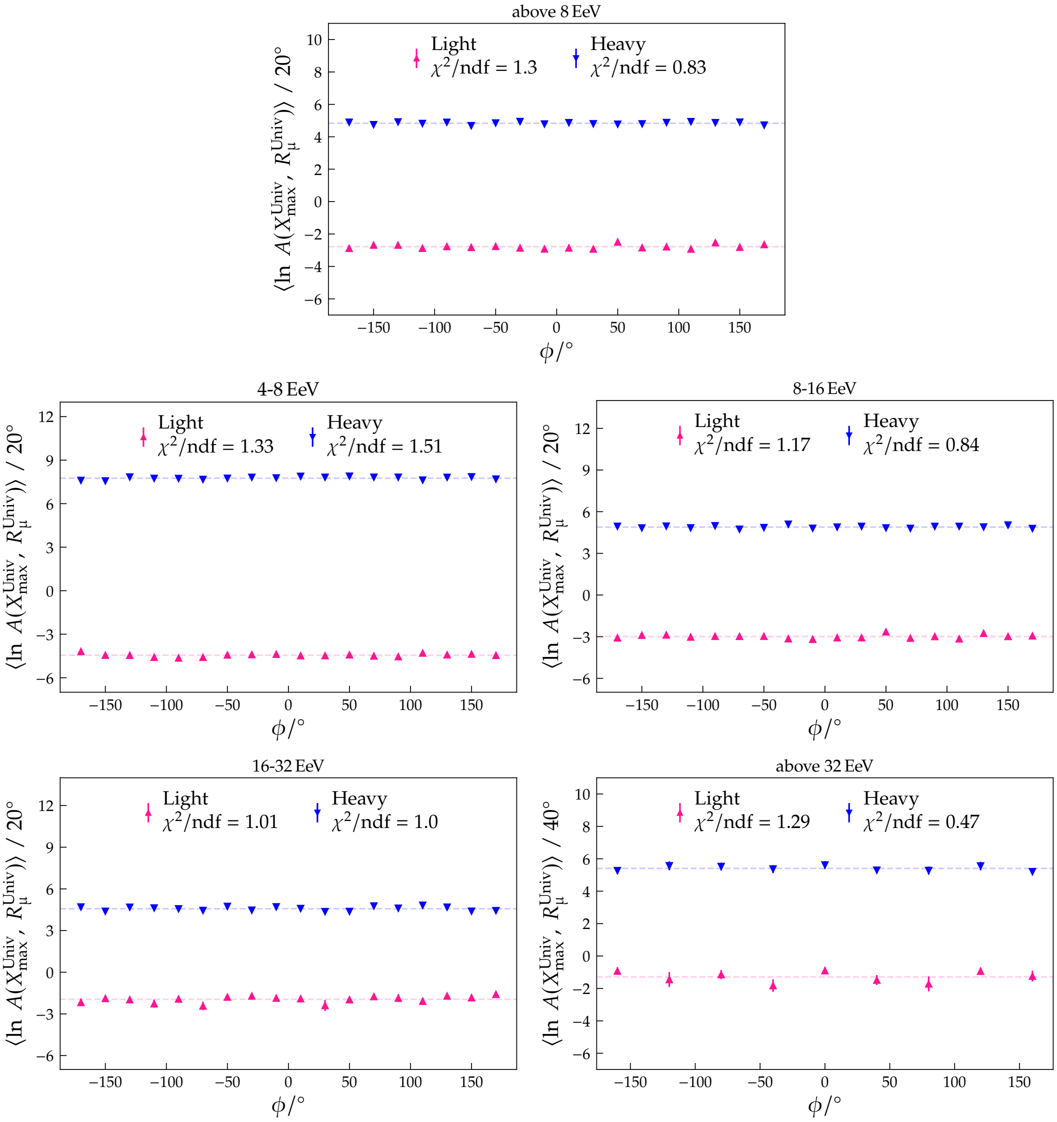


Figure 7: Behavior of the mass estimator $\ln A$ as a function of azimuth angle for the light (pink) and heavy (blue) populations, in each energy bin. Annotated are also the the reduced chi squared quantity of the agreement with a uniform distribution. This showcases that no strong spurious modulation of the mass estimator is observed in azimuth angle.

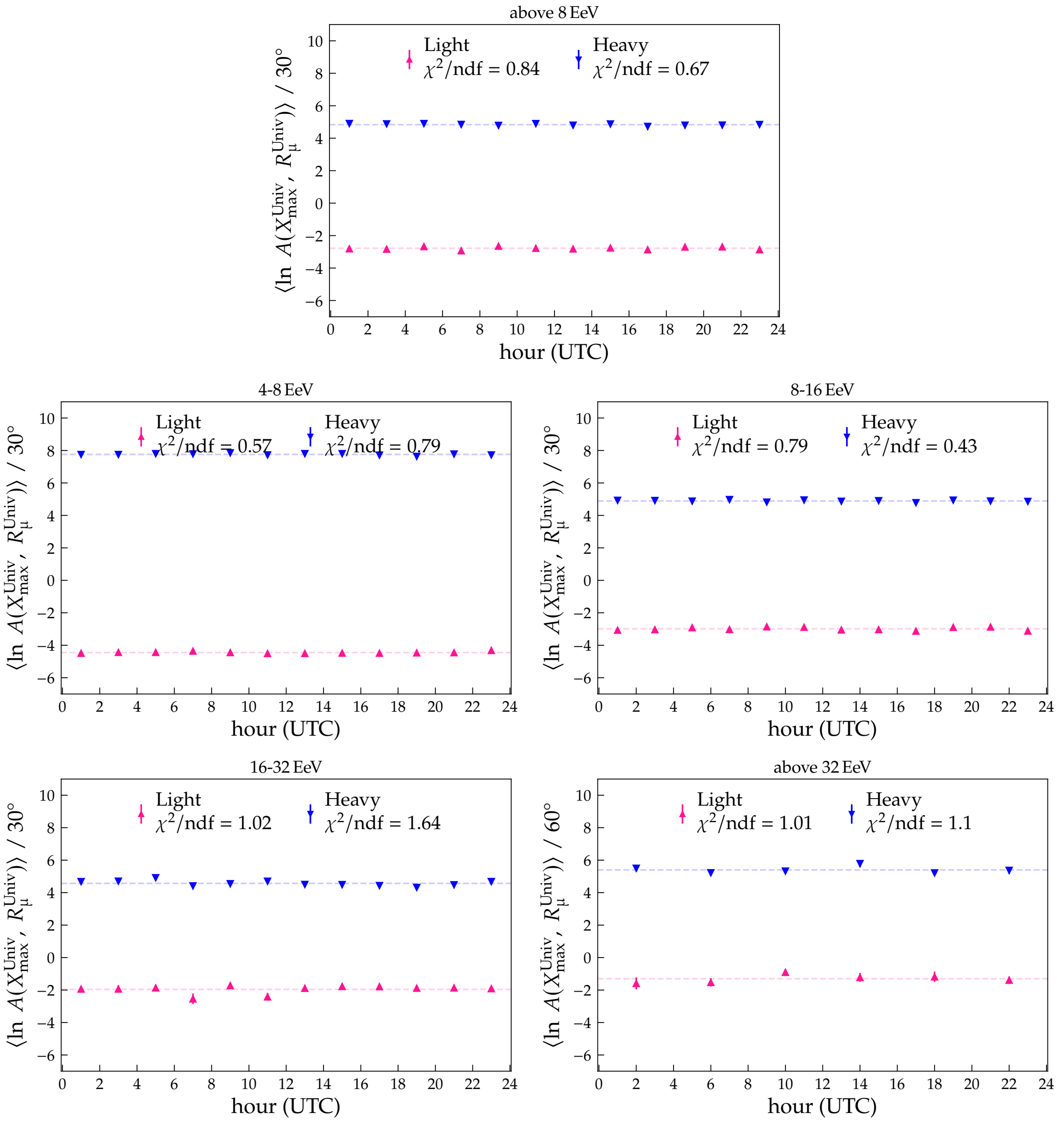


Figure 8: Behavior of the mass estimator $\ln A$ as a function of solar hour for the light (pink) and heavy (blue) populations, in each energy bin. Annotated are also the the reduced chi squared quantity of the agreement with a uniform distribution. This showcases that no strong spurious modulation of the mass estimator is observed in solar hour.

B. Complementary graphs

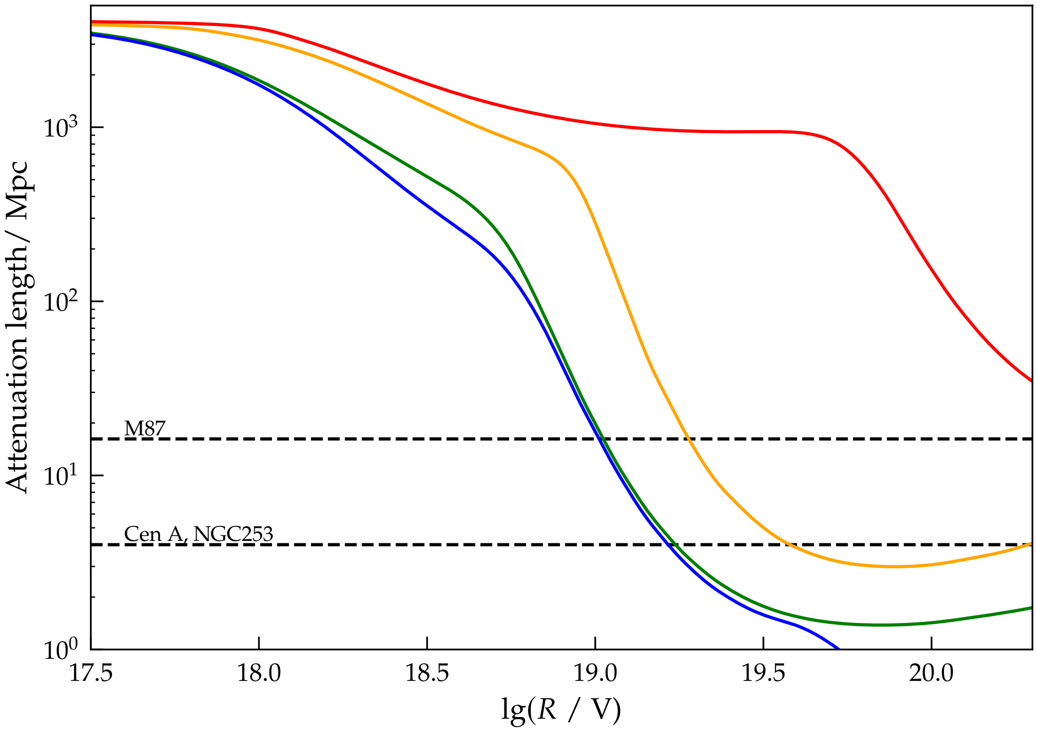


Figure 9: Attenuation length as a function of rigidity for the four representative primary cosmic rays: proton (red), helium (orange), nitrogen (green), and iron (blue). The dashed lines indicate the attenuation length for each nuclei for a redshift of $z = 0.5$, while the continuous line is that for $z = 0$. The horizontal lines indicate the distances of a few sources: Cen A and NGC 253 located ≃ 4 Mpc away from Earth, while M87 is located ≃ 17 Mpc away.

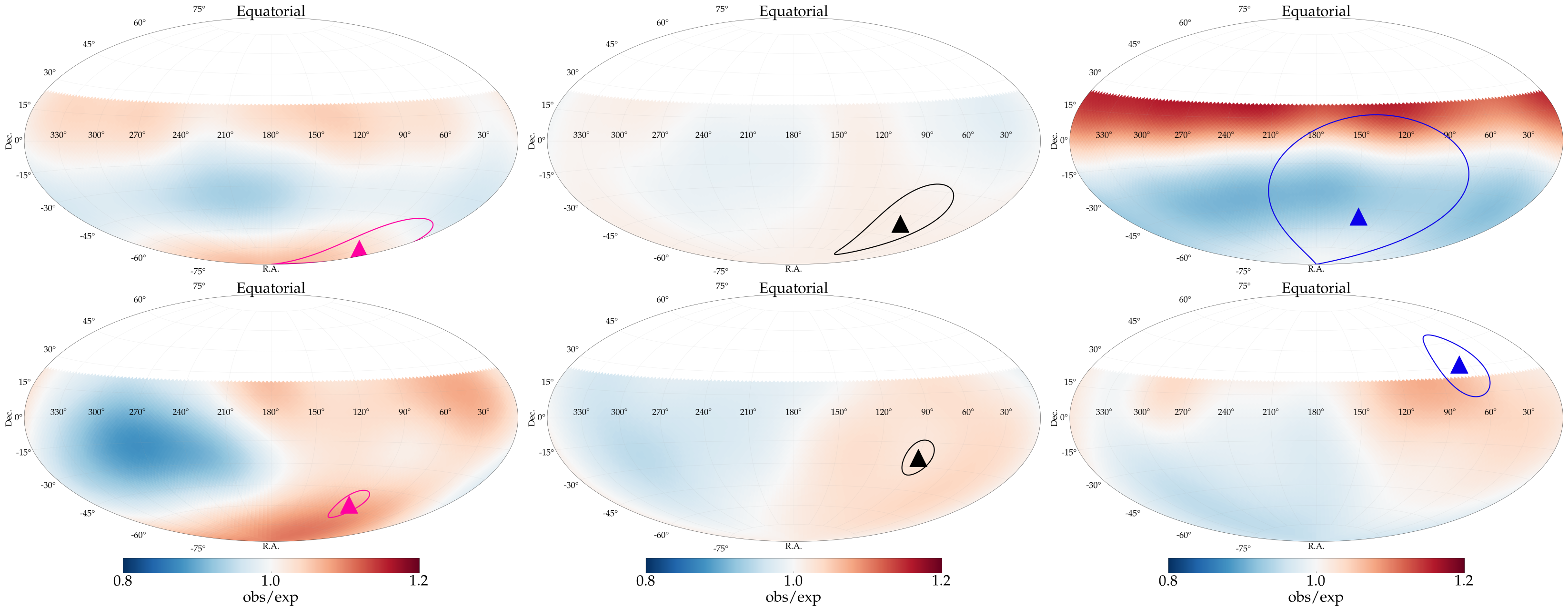


Figure 10: Ratio between the flux of events and the mean flux over the visible sky, in equatorial coordinates. Corresponds in first approximation to the ratio between the number of observed events and those expected from an isotropic distribution reported in Ref. [70]. In the first row, events with energy between 4 and 8 EeV; and in the second row, those with energy between 8 and 16 EeV. The graphs correspond to the light population (left), to the total dataset (center), and to the heavy population (right).

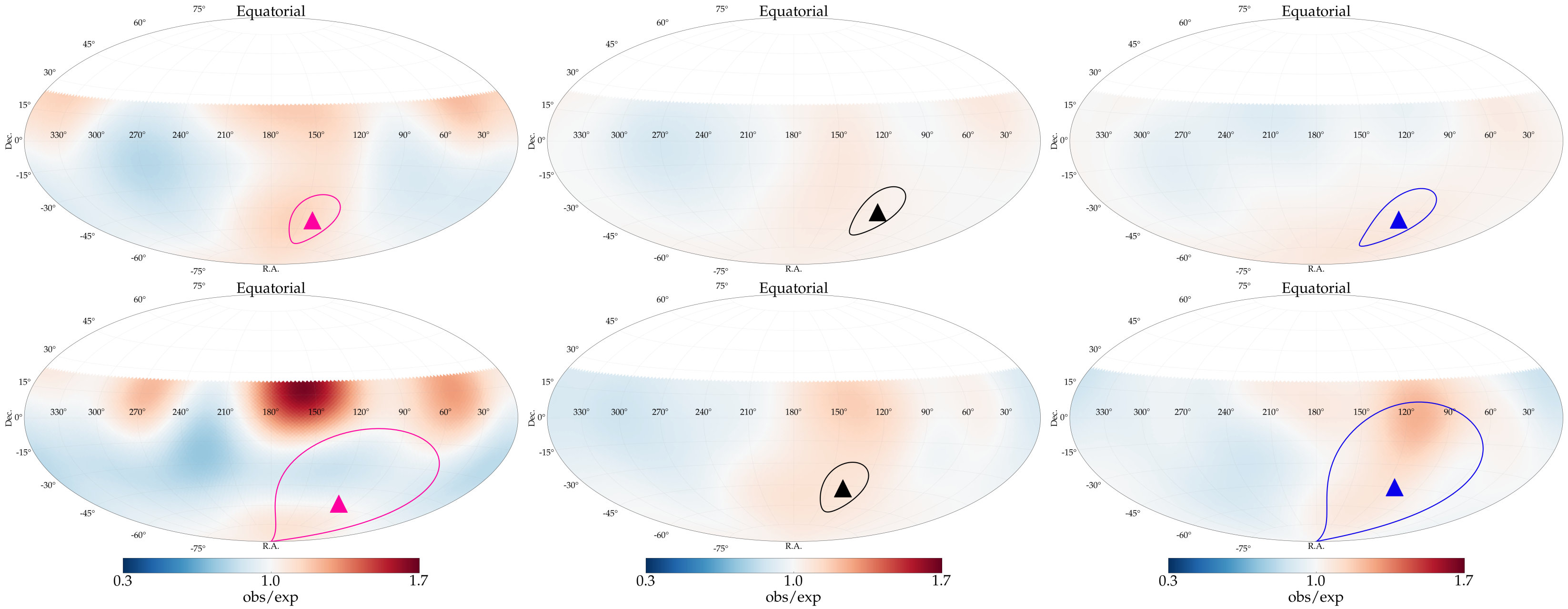


Figure 11: Ratio between the flux of events and the mean flux over the visible sky, in equatorial coordinates. Corresponds in first approximation to the ratio between the number of observed events and those expected from an isotropic distribution reported in Ref. [70]. In the first row, events with energy between 16 and 32 EeV; and in the second row, those with energy above 32 EeV. The graphs correspond to the light population (left), to the total dataset (center), and to the heavy population (right).

ACKNOWLEDGEMENTS

I am beyond grateful to my referees, Ralph Engel and Silvia Mollerach, for their support and incentive in the development of this work, especially for the guidance and welcome during my time in Bariloche. This thesis could not have been possible without Markus Roth welcoming me to the hardworking group at KIT and his broad view of the science being developed at the Pierre Auger Collaboration. I am also thankful for the numerous opportunities to join international conferences and schools, providing me with the knowledge and skills to develop this work. I also had intensive guidance from Darko Veberič, with whom many fruitful discussions were held – thank you for the opportunity to develop my writing and debate skills – and who always took care of the student's integration and development. To David Schmidt, I address my gratitude and appreciation, as his availability for discussion allowed me to explore the potential shortfalls of the proposed analysis and, with that, gain confidence in what is here presented. I cannot go further without thanking the gracious help over these four years: to Sabine Bucher, Katrin Link, Marie-Christine Mundt, and Anna Friedrich. Your work excellence allowed me to carry my work unworried, sure that whatever issue or concern on bureaucratic terms could be solved with your help. I am very grateful.

If life is a journey, the PhD process is definitely the part where one climbs up a steep mountain. I am fulfilled to have found my journey partner, my husband Damian Möhler, who has been by my side in all the ups and downs in these four years and for many more to come. I love you. You have also gifted me with a loving, welcoming, and adventurous family who helped me see Germany as a fun and enjoyable place to live. To Martin, Birgit, Paul, Kristin, and all the Möhler family: Vielen herzlichen Dank für eure Liebe und Unterstützung und dafür, dass ihr mich mit offenem Herzen aufgenommen habt.

I could not describe the journey without talking about my dear friends with whom I shared so many moments, meals, frustrations, and happiness. To you, my immense appreciation and the hope we share more of life: Allan Prayeras, Fabio Convenga, Steffen Hahn, Karla Ruiz Zavala, Sara Martinelli, and Olena Tkachenko. To Mariana Santana, you are my inspiration, and I can say you have trailed the path for me. I am so lucky to have your friendship and to be able to share all joy and sorrow; I admire you greatly and thank you for all the love. To Tobias Schulz: we did it!!! Thank you for so many nice moments and the special project of a dedicated relaxation room for

students; thank you for all the support especially in the last moments of the PhD.

To the colleagues who shared the struggle of starting a PhD during the coronavirus pandemic, I am thankful for the help handling all the initial bureaucracies: Nikos Karastathis, Max Büsken, Sara Martinelli, and Julian Saffer. To Max Stadelmaier, I thank you for the great help in the initial pandemics-filled months -from groceries to rides - and for the many fruitful discussions. To all the colleagues at KIT for the friendly environment and the shared fun outings: Daniel, Fiona, Quentin, Christoph, Philipp, Anton, Luca, Paul, Kathrin, and Berenika. To the colleagues in the DDAp: I admire every single one of you for giving the best of you to your thesis, dealing with all the stress and enjoying all of the fun of it. In special I send a hug to Eze, Gaby, Joaquin, Marina, and Varada. A special thank you goes to Carmina, with whom I shared the task of being students' representatives and through that got to esteem. To the colleagues in Bariloche, who made my stay there so much better and fun, a warm thank you, in special to Juan.

I would love to thank all the teachers and professors who supported and encouraged me. Pursuing a scientific career was not commonplace in my surroundings when I was younger, and your encouragement meant the world to me. Then, on the undergrad path, it is impossible not to thank Diego Cogollo and Douglas Vitoretti, who pushed me further and allowed me a beautiful working experience with welcoming people in Bucaramanga. I am thankful for my *alma mater*, Universidade Federal de Campina Grande, where I had the fortune to be instructed by professors who invited me to discuss various topics and from whom I had full support to pursue this path. And to the dear friends I made for life there, I hope to see you soon: Caio, Klecio and Tayse. For his continuous support and guidance, I thank Vitor de Souza for showing the role and duty of a good and caring advisor through his example. I am also happy for the friends I made there, Adriane Leal and Larissa Aquino, and for the colleagues who shared a lot of their workspace experience at the Instituto de Física de São Carlos; thank you for the good times.

Aos meus pais, Edmilson e Marivone Martins: vocês me deram a possibilidade de lutar por e alcançar os meus sonhos, e ensinaram através do exemplo que tudo que nos tem valor custa muito esforço e dedicação. Muito obrigada por todo o amor e oportunidades que me deram, desde as caronas para os inúmeros cursos que fiz, até as incriveis viagens que fizemos juntos. Que apesar da distância vocês saibam do meu amor e graditão. Essa conquista é nossa.

Às minhas irmãs, Emanuelly e Elayne Martins vocês são meus exemplos na vida, com toda força e ternura me ensinam a persistir e perserverar. Eu encontro minha força em vocês. À minha sobrinha Bianca, você é um suspiro de leveza e acalento para toda a família. Eu adoro ver você descobrindo novos mundos, e vou estar sempre aqui pra um abraço apertado e para as futuras aventuras. Ao cunhado Higor Vasconcelos, eu admiro suas mil fascetas e interesses, sou feliz de ter você em nossa famiília. Ao cunhado Pedro dos Anjos, *in memoriam*, sua alegria marcou a vida de todos que você cruzou, e me faz querer aproveitar a vida e os amores com a intensidade que você o fez.

Por fim, minha gratidão à minha terapeuta que através de seu ofício me ajudou a perseverar nessa escalada. Muito obrigada pela acolhida com carinho.

Mental health is a serious topic, and I hope whoever reads this has the help and support they need. Take some breaks and find ways to keep on going. You got this.

I spent some time selecting harmonic, distinguishable, and color-blind-friendly colors for the graphs I produced for this thesis in the hopes that they would be easily understandable for a broader audience. Instead of it becoming a dull commitment, I had fun thanks to the user-friendly online tools, all of which I recommend: coolors, paletton, and Coblis. Grammarly was a helpful tool for reviewing spelling, grammar, and punctuation. DeepL was used for the translation of the abstract, although the reviews from Steffen Hahn also need to be acknowledged. The invaluable computing tools used for this analysis were NumPy [114], SciPy [115], Uproot [116], Matplotlib [117], pandas [118], seaborn [119], and Astropy [120].

BIBLIOGRAPHY

[1] V. F. Hess, "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten," *Physikalische Zeitschrift*, vol. 13, pp. 1084–1091, 1912.

[2] B. Rossi, "Über die eigenschaften der durchdringenden korpuskularstrahlung im meeresniveau," *Zeitschrift für Physik*, vol. 82, no. 3, pp. 151–178, 1933.

[3] P. Auger, P. Ehrenfest, R. Maze, J. Daudin, and R. A. Fréon, "Extensive cosmic-ray showers," *Reviews of modern physics*, vol. 11, no. 3-4, p. 288, 1939.

[4] J. Matthews, "A Heitler model of extensive air showers," *Astroparticle Physics*, vol. 22, no. 5-6, pp. 387–397, 2005.

[5] C. M. G. Lattes, H. Muirhead, G. P. Occhialini, and C. F. Powell, "Processes involving charged mesons," *Nature*, vol. 159, no. 4047, pp. 694–697, 1947.

[6] L. A. Anchordoqui, "Ultra-high-energy cosmic rays," *Physics Reports*, vol. 801, pp. 1–93, 2019.

[7] E. Fermi, "On the origin of the cosmic radiation," *Phys. Rev.*, vol. 75, pp. 1169–1174, Apr 1949.

[8] D. Hooper, *Particle Cosmology and Astrophysics*. Princeton University Press, 2024.

[9] K. Kotera and A. V. Olinto, "The astrophysics of ultrahigh-energy cosmic rays," *Annual Review of Astronomy and Astrophysics*, vol. 49, no. 1, pp. 119–153, 2011.

[10] K. Greisen, "End to the Cosmic-Ray Spectrum?," *Phys. Rev. Lett.*, vol. 16, pp. 748–750, Apr 1966.

[11] G. T. Zatsepin and V. A. Kuzmin, "Upper limit of the spectrum of cosmic rays," *JETP Lett.*, vol. 4, pp. 78–80, 1966.

[12] A. Coleman, J. Eser, E. Mayotte, F. Sarazin, F. Schröder, D. Soldin, T. Venters, R. Aloisio, J. Alvarez-Muñiz, R. Alves Batista, D. Bergman, M. Bertaina, L. Caccianiga, O. Deligny, H. Dembinski, P. Denton, A. di Matteo, N. Globus, J. Glombitza, G. Golup, A. Haungs, J. Hörandel, T. Jaffe, J. Kelley, J. Krizmanic,

L. Lu, J. Matthews, I. Mariş, R. Mussa, F. Oikonomou, T. Pierog, E. Santos, P. Tinyakov, Y. Tsunesada, M. Unger, A. Yushkov, M. Albrow, L. Anchordoqui, K. Andeen, E. Arnone, D. Barghini, E. Bechtol, J. Bellido, M. Casolino, A. Castellina, L. Cazon, R. Conceição, R. Cremonini, H. Dujmovic, R. Engel, G. Farrar, F. Fenu, S. Ferrarese, T. Fujii, D. Gardiol, M. Gritsevich, P. Homola, T. Huege, K.-H. Kampert, D. Kang, E. Kido, P. Klimov, K. Kotera, B. Kozelov, A. Leszczyńska, J. Madsen, L. Marcelli, M. Marisaldi, O. Martineau-Huynh, S. Mayotte, K. Mulrey, K. Murase, M. Muzio, S. Ogio, A. Olinto, Y. Onel, T. Paul, L. Piotrowski, M. Plum, B. Pont, M. Reininghaus, B. Riedel, F. Riehn, M. Roth, T. Sako, F. Schlüter, D. Shoemaker, J. Sidhu, I. Sidelnik, C. Timmermans, O. Tkachenko, D. Veberic, S. Verpoest, V. Verzi, J. Vícha, D. Winn, E. Zas, and M. Zotov, "Ultra high energy cosmic rays the intersection of the cosmic and energy frontiers," *Astroparticle Physics*, vol. 149, p. 102819, July 2023.

[13] H. Bethe and W. Heitler, "On the stopping of fast particles and on the creation of positive electrons," *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, vol. 146, no. 856, pp. 83–112, 1934.

[14] D. Allard, "Extragalactic propagation of ultrahigh energy cosmic-rays," *Astroparticle Physics*, vol. 39–40, p. 33–43, Dec. 2012.

[15] M. Unger and G. R. Farrar, "The coherent magnetic field of the Milky Way," *The Astrophysical Journal*, vol. 970, no. 1, p. 95, 2024.

[16] The Pierre Auger Collaboration, "Search for magnetically-induced signatures in the arrival directions of ultra-high-energy cosmic rays measured at the Pierre Auger Observatory," *Journal of Cosmology and Astroparticle Physics*, vol. 2020, no. 06, p. 017, 2020.

[17] G. R. Farrar and M. S. Sutherland, "Deflections of uhecrs in the galactic magnetic field," *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 05, p. 004, 2019.

[18] G. Magkos and V. Pavlidou, "Deflections of ultra-high energy cosmic rays by the milky way magnetic field: how well can they be corrected?," *Journal of Cosmology and Astroparticle Physics*, vol. 2019, no. 02, p. 004, 2019.

[19] D. Walz, *Constraining models of the extragalactic cosmic ray origin with the Pierre Auger Observatory*. PhD thesis, Dissertation, RWTH Aachen University, 2016, 2016.

[20] The Pierre Auger Collaboration, "Constraining the sources of ultra-high-energy cosmic rays across and above the ankle with the spectrum and composition data measured at the Pierre Auger Observatory," *Journal of cosmology and astroparticle physics*, vol. 2023, no. 05, p. 024, 2023.

[21] The Pierre Auger Collaboration, "Constraining models for the origin of ultra-high-energy cosmic rays with a novel combined analysis of arrival directions,

spectrum, and composition data measured at the Pierre Auger Observatory," *Journal of cosmology and astroparticle physics*, vol. 2024, no. 01, p. 022, 2024.

[22] POEMMA collaboration, "The poemma (probe of extreme multi-messenger astrophysics) observatory," *Journal of Cosmology and Astroparticle Physics*, vol. 2021, p. 007, jun 2021.

[23] J. R. Hörandel, "GCOS - The Global Cosmic Ray Observatory," in *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*, vol. 395, p. 027, 2021.

[24] C. Evoli, "The Cosmic-Ray Energy Spectrum," Dec. 2020. `https://doi.org/10.5281/zenodo.4396125`.

[25] B. Bortolato, J. F. Kamenik, and M. Tammaro, "Learning the composition of ultrahigh energy cosmic rays," *Physical Review D*, vol. 108, no. 2, p. 022004, 2023.

[26] The Pierre Auger Collaboration, "Studies of the mass composition of cosmic rays and proton-proton interaction cross-sections at ultra-high energies with the Pierre Auger Observatory," in *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*, vol. 444, p. 438, 2023.

[27] S. Navas *et al.*, "Review of particle physics," *Phys. Rev. D*, vol. 110, no. 3, p. 030001, 2024.

[28] C. Ding, N. Globus, and G. R. Farrar, "The Imprint of Large-scale Structure on the Ultrahigh-energy Cosmic-Ray Sky," *The Astrophysical Journal Letters*, vol. 913, p. L13, May 2021.

[29] D. Harari, S. Mollerach, and E. Roulet, "Anisotropies of ultrahigh energy cosmic ray nuclei diffusing from extragalactic sources," *Physical Review D*, vol. 92, no. 6, p. 063014, 2015.

[30] J. P. Huchra, L. M. Macri, K. L. Masters, T. H. Jarrett, P. Berlind, M. Calkins, A. C. Crook, R. Cutri, P. Erdoğdu, E. Falco, *et al.*, "The 2MASS Redshift Survey—description and data release," *The Astrophysical Journal Supplement Series*, vol. 199, no. 2, p. 26, 2012.

[31] R. B. Tully, D. Pomarède, R. Graziani, H. M. Courtois, Y. Hoffman, and E. J. Shaya, "Cosmicflows-3: Cosmography of the local void," *The Astrophysical Journal*, vol. 880, p. 24, jul 2019.

[32] A. Watson, "The Pierre Auger Observatory – why and how," *2005 CERN-CLAF School of High-Energy Physics, CLASHEP 2005 - Proceedings*, 01 2006.

[33] P. Sommers, "Cosmic ray anisotropy analysis with a full-sky observatory," *Astroparticle Physics*, vol. 14, no. 4, pp. 271–286, 2001.

[34] D. J. Bird, S. C. Corbato, H. Y. Dai, J. W. Elbert, K. D. Green, M. A. Huang, D. B. Kieda, S. Ko, C. G. Larsen, E. C. Loh, M. Z. Luo, M. H. Salamon, J. D. Smith, P. Sokolsky, P. Sommers, J. K. K. Tang, and S. B. Thomas, "Detection of a Cosmic Ray with Measured Energy Well beyond the Expected Spectral Cutoff due to Cosmic Microwave Radiation," *The Astrophysical Journal*, vol. 441, p. 144, Mar. 1995.

[35] N. Hayashida, K. Honda, M. Honda, S. Imaizumi, N. Inoue, K. Kadota, F. Kakimoto, K. Kamata, S. Kawaguchi, N. Kawasumi, Y. Matsubara, K. Murakami, M. Nagano, H. Ohoka, M. Takeda, M. Teshima, I. Tsushima, S. Yoshida, and H. Yoshii, "Observation of a very energetic cosmic ray well beyond the predicted 2.7 k cutoff in the primary energy spectrum," *Phys. Rev. Lett.*, vol. 73, pp. 3491–3494, Dec 1994.

[36] The Pierre Auger Collaboration, "The Pierre Auger cosmic ray observatory," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 798, pp. 172–213, 2015.

[37] The Pierre Auger Collaboration, "Design report." `https://www.auger.org/index.php/admin-pages/design-report`. Accessed on 14 October 2024., 1996.

[38] D. H. Perkins, *Particle astrophysics*. OUP Oxford, 2009.

[39] National Oceanic and Atmospheric Administration, U.S. Department of Commerce, "The atmosphere," 2024. `https://www.noaa.gov/jetstream/atmosphere`. Accessed on 14 October 2024.

[40] P. K. F. Grieder, *Extensive air showers: High energy phenomena and astrophysical aspects. A tutorial, reference manual and data book. Vol. 1.* Heidelberg: Springer, 2010.

[41] The Pierre Auger Collaboration, "Search for photons above 10^{18} eV by simultaneously measuring the atmospheric depth and the muon content of air showers at the Pierre Auger Observatory," *Phys. Rev. D*, vol. 110, p. 062005, Sep 2024.

[42] S. Navas, C. Amsler, T. Gutsche, C. Hanhart, J. Hernández-Rey, C. Lourenço, A. Masoni, M. Mikhasenko, R. Mitchell, C. Patrignani, *et al.*, "Review of particle physics," *Physical Review D*, vol. 110, no. 3, p. 030001, 2024.

[43] A. Haungs *et al.*, "KCDC - The KASCADE Cosmic-ray Data Centre," *J. Phys. Conf. Ser.*, vol. 632, no. 1, p. 012011, 2015.

[44] F. Schmidt and J. Knapp, "CORSIKA Shower Images," 2005. `https://www-zeuthen.desy.de/~jknapp/fs/showerimages.html`. Accessed on 14 October 2024.

[45] A. Bridgeman, "Shower universality reconstruction of data from the Pierre Auger Observatory and validations with hadronic interaction models," in *Proceedings of 35th International Cosmic Ray Conference — PoS(ICRC2017)*, vol. 301, p. 323, 2017.

[46] W. Heitler, *The quantum theory of radiation*. Courier Corporation, 1984.

[47] B. Dawson, "The Energy Scale of the Pierre Auger Observatory," in *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*, vol. 358, p. 231, 2019.

[48] D. Veberič. `https://web.iap.kit.edu/darko/auger/auger-array/auger_array-pdf/auger_array-ad.pdf`. Accessed on 15 October 2024.

[49] The Pierre Auger Collaboration. `https://www.auger.org/index.php/observatory/auger-hybrid-detector`. Accessed on 15 October 2024.

[50] M. L. Schimassek, *Extending the Physics Reach of the Pierre Auger Observatory using Low-Level Trigger Data*. PhD thesis, Karlsruher Institut für Technologie (KIT), 2022. 51.13.03; LK 01.

[51] The Pierre Auger Collaboration. `https://www0.mi.infn.it/auger/hexagons.html`. Accessed on 15 October 2024.

[52] The Pierre Auger Collaboration, "The second knee in the cosmic ray spectrum observed with the surface detector of the Pierre Auger Observatory," in *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*, vol. 444, p. 398, 2023.

[53] C. M. Schäfer, *The XY-scanner: A Versatile Method for the Absolute End-to-End Calibration of Fluorescence Detectors*. PhD thesis, Dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), 2023.

[54] The Pierre Auger Collaboration, "The Pierre Auger Observatory Upgrade - Preliminary Design Report," *arXiv preprint arXiv:1604.03637*, 2016.

[55] A. Taboada, "Analysis of Data from Surface Detector Stations of the AugerPrime Upgrade," in *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*, vol. 358, p. 434, 2019.

[56] The Pierre Auger Collaboration. `https://www.flickr.com/photos/134252569@N07/albums/72157683547471100`. Accessed on 02 November 2024.

[57] F. Gesualdi, *The muon content of atmospheric air showers and the mass composition of cosmic rays*. PhD thesis, Dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), 2022.

[58] G. A. Anastasi, "AugerPrime: The Pierre Auger Observatory upgrade," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 1044, p. 167497, 2022.

[59] The Pierre Auger Collaboration. `https://visitantes.auger.org.ar/index.php/umd-underground-muon-detector/`. Accessed on 02 November 2024.

[60] The Pierre Auger Collaboration, "Energy estimation of cosmic rays with the engineering radio array of the pierre auger observatory," *Phys. Rev. D*, vol. 93, p. 122005, Jun 2016.

[61] The Pierre Auger Collaboration, "Demonstrating agreement between radio and fluorescence measurements of the depth of maximum of extensive air showers at the pierre auger observatory," *Phys. Rev. Lett.*, vol. 132, p. 021001, Jan 2024.

[62] The Pierre Auger Collaboration, "Radio measurements of the depth of air-shower maximum at the Pierre Auger Observatory," *Phys. Rev. D*, vol. 109, p. 022002, Jan 2024.

[63] The Pierre Auger Collaboration. `https://www.youtube.com/@pierreaugerobservatory9832`. Accessed on 02 November 2024.

[64] The Pierre Auger Collaboration. `https://opendata.auger.org`. Accessed on 02 November 2024.

[65] The Pierre Auger Collaboration, "Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV," *Science*, vol. 357, no. 6357, pp. 1266–1270, 2017.

[66] The Pierre Auger Collaboration, "Large-scale cosmic-ray anisotropies above 4 EeV measured by the Pierre Auger Observatory," *The Astrophysical Journal*, vol. 868, no. 1, 2018.

[67] Wikipedia contributors, "Spherical harmonics — Wikipedia, the free encyclopedia," 2024. `https://en.wikipedia.org/w/index.php?title=Spherical_harmonics&oldid=1241393683`. Accessed on 29 October 2024.

[68] Wikipedia contributors, "Atan2 — Wikipedia, the free encyclopedia," 2024. `https://en.wikipedia.org/w/index.php?title=Atan2&oldid=1252278261`. Accessed on 29 October 2024.

[69] The Pierre Auger Collaboration, "An update on the arrival direction studies made with data from the Pierre Auger Observatory," *PoS*, vol. ICRC2023, p. 252, 2023.

[70] The Pierre Auger Collaboration, "Large-scale Cosmic-ray Anisotropies with 19 yr of Data from the Pierre Auger Observatory," *The Astrophysical Journal*, vol. 976, p. 48, nov 2024.

[71] The Pierre Auger Collaboration, "Features of the Energy Spectrum of Cosmic Rays above 2.5×10^{18} eV Using the Pierre Auger Observatory," *Phys. Rev. Lett.*, vol. 125, p. 121106, Sep 2020.

[72] The Pierre Auger Collaboration, "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," *The European Physical Journal C*, vol. 81, pp. 1–25, 2021.

[73] D. Harari, S. Mollerach, and E. Roulet, "Anisotropies of ultrahigh energy cosmic rays diffusing from extragalactic sources," *Physical Review D*, vol. 89, no. 12, p. 123001, 2014.

[74] The Pierre Auger Collaboration, "Arrival directions of cosmic rays above 32 EeV from phase one of the Pierre Auger Observatory," *The Astrophysical Journal*, vol. 935, no. 2, p. 170, 2022.

[75] The Pierre Auger Collaboration, "Correlation of the highest-energy cosmic rays with nearby extragalactic objects," *Science*, vol. 318, no. 5852, pp. 938–943, 2007.

[76] R. Alves Batista, J. Becker Tjus, J. Dörner, A. Dundovic, B. Eichmann, A. Frie, C. Heiter, M. R. Hoerbe, K.-H. Kampert, L. Merten, G. Müller, P. Reichherzer, A. Saveliev, L. Schlegel, G. Sigl, A. van Vliet, and T. Winchen, "CRPropa 3.2 — an advanced framework for high-energy particle propagation in extragalactic and galactic spaces," *Journal of Cosmology and Astroparticle Physics*, vol. 2022, p. 035, Sept. 2022.

[77] R. Aloisio, D. Boncioli, A. Grillo, S. Petrera, and F. Salamida, "SimProp: a simulation code for ultra high energy cosmic ray propagation," *Journal of Cosmology and Astroparticle Physics*, vol. 2012, p. 007–007, Oct. 2012.

[78] The Pierre Auger Collaboration, "Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory," *Journal of Cosmology and Astroparticle Physics*, vol. 2017, no. 04, p. 038, 2017.

[79] R. Jansson and G. R. Farrar, "A new model of the galactic magnetic field," *The Astrophysical Journal*, vol. 757, no. 1, p. 14, 2012.

[80] The Pierre Auger Collaboration, "Indication of a mass-dependent anisotropy above $10^{18.7}$ eV in the hybrid data of the Pierre Auger Observatory," in *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*, vol. 395, p. 321, 2021.

[81] The Pierre Auger Collaboration, "Measurement of the mass composition of ultra-high-energy cosmic rays at the Pierre Auger Observatory," in *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*, vol. 444, p. 365, 2023.

[82] E. Mayotte and T. Fitoussi, "Update on the indication of a mass-dependent anisotropy above $10^{18.7}$ eV in the hybrid data of the Pierre Auger Observatory," *EPJ Web of Conferences*, vol. 283, p. 03003, 2023.

[83] The Pierre Auger Collaboration, "Depth of Maximum of Air-Shower Profiles above $10^{17.8}$ eV Measured with the Fluorescence Detector of the Pierre Auger Observatory and Mass Composition Implications," in *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*, vol. 444, p. 249, 2023.

[84] The Pierre Auger Collaboration, "Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory," *Physical Review D*, vol. 102, no. 6, p. 062005, 2020.

[85] R. Abbasi, T. Abu-Zayyad, M. Allen, J. Amman, G. Archbold, K. Belov, J. Belz, S. Ben Zvi, D. Bergman, S. Blake, *et al.*, "First observation of the Greisen-Zatsepin-Kuzmin suppression," *Physical review letters*, vol. 100, no. 10, p. 101101, 2008.

[86] The Pierre Auger Collaboration, "Observation of the suppression of the flux of cosmic rays above $4{\times}10^{19}$ eV," *Physical review letters*, vol. 101, no. 6, p. 061101, 2008.

[87] T. Abu-Zayyad, R. Aida, M. Allen, R. Anderson, R. Azuma, E. Barcikowski, J. W. Belz, D. R. Bergman, S. A. Blake, R. Cady, *et al.*, "The cosmic-ray energy spectrum observed with the surface detector of the telescope array experiment," *The Astrophysical Journal Letters*, vol. 768, no. 1, p. L1, 2013.

[88] The Pierre Auger Collaboration, "Inference of the mass composition of cosmic rays with energies from $10^{18.5}$ to 10^{20}eV using the Pierre Auger Observatory and Deep Learning," *arXiv preprint arXiv:2406.06315*, 2024.

[89] R. Alves Batista, J. Biteau, M. Bustamante, K. Dolag, R. Engel, K. Fang, K.-H. Kampert, D. Kostunin, M. Mostafa, K. Murase, *et al.*, "Open questions in cosmic-ray research at ultrahigh energies," *Frontiers in Astronomy and Space Sciences*, vol. 6, p. 23, 2019.

[90] The Pierre Auger Collaboration, "Energy spectrum of cosmic rays measured using the Pierre Auger Observatory," in *Proceedings of 37th International Cosmic Ray Conference — PoS(ICRC2021)*, vol. 395, p. 324, 2021.

[91] J. Bellido, "Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and Composition Implications," in *Proceedings of 35th International Cosmic Ray Conference — PoS(ICRC2017)*, vol. 301, p. 506, 2017.

[92] A. Yushkov, "Mass Composition of Cosmic Rays with Energies above $10^{17.2}$ eV from the Hybrid Data of the Pierre Auger Observatory," in *Proceedings of 36th International Cosmic Ray Conference — PoS(ICRC2019)*, vol. 358, p. 482, 2019.

[93] M. S. Muzio, L. A. Anchordoqui, and M. Unger, "Peters cycle at the end of the cosmic ray spectrum?," *Physical Review D*, vol. 109, no. 2, p. 023006, 2024.

[94] D. Boncioli, S. Petrera, F. Salamida, and C. Trimarelli, "Top-of-Atmosphere Combined Fit," *GAP-2022-007*, 2022.

[95] Mollerach, Silvia and Roulet, Esteban and Taborda, Oscar, "Using 5T5 events for Large Scale Anisotropy studies above 4 EeV," *GAP-2015-059*, 2015.

[96] P. Lipari, “Concepts of “age” and “universality” in cosmic ray showers,” *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, vol. 79, no. 6, p. 063001, 2009.

[97] A. Schulz, “Measurement of the energy spectrum and mass composition of ultra-high energy cosmic rays,” *Ph. D. Thesis*, 2016.

[98] A. Bridgeman, “Determining the mass composition of ultra-high energy cosmic rays using air shower universality,” *Ph. D. Thesis*, 2019.

[99] M. Stadelmaier, *On air-shower universality and the mass composition of ultra-high-energy cosmic rays*. PhD thesis, Karlsruhe Institute of Technology and Universidad de San Martin, 2022.

[100] M. Stadelmaier, R. Engel, M. Roth, D. Schmidt, and D. Veberič, “Model of the response of surface detectors to extensive air showers based on shower universality,” *Physical Review D*, vol. 110, July 2024.

[101] The Pierre Auger collaboration, “The effect of the geomagnetic field on cosmic ray energy estimates and large scale anisotropy searches on data from the Pierre Auger Observatory,” *Journal of Cosmology and Astroparticle Physics*, vol. 2011, p. 022, nov 2011.

[102] The Pierre Auger Collaboration, “Impact of atmospheric effects on the energy reconstruction of air showers observed by the surface detectors of the Pierre Auger Observatory,” *Journal of Instrumentation*, vol. 12, p. P02006, feb 2017.

[103] L. Apollonio, E. E. Martins, M. Roth, M. Stadelmaier, and D. Veberič, “Depth of the maxima of cosmic-ray showers above 4 EeV using air-shower universality techniques and the surface-detector data of the Pierre Auger Observatory,” *GAP-2024-057*, 2024.

[104] W. Apel, J. Arteaga-Velázquez, K. Bekk, M. Bertaina, J. Blümer, H. Bozdog, I. Brancus, P. Buchholz, E. Cantoni, A. Chiavassa, *et al.*, “The spectrum of high-energy cosmic rays measured with kascade-grande,” *Astroparticle Physics*, vol. 36, no. 1, pp. 183–194, 2012.

[105] W. Apel, J. Arteaga-Velázquez, K. Bekk, M. Bertaina, J. Blümer, H. Bozdog, I. Brancus, E. Cantoni, A. Chiavassa, F. Cossavella, *et al.*, “Kascade-grande measurements of energy spectra for elemental groups of cosmic rays,” *Astroparticle Physics*, vol. 47, pp. 54–66, 2013.

[106] Z. Cao, F. Aharonian, Axikegu, Y. Bai, Y. Bao, D. Bastieri, X. Bi, Y. Bi, W. Bian, A. Bukevich, *et al.*, “Measurements of All-Particle Energy Spectrum and Mean Logarithmic Mass of Cosmic Rays from 0.3 to 30 PeV with LHAASO-KM2A,” *Physical Review Letters*, vol. 132, no. 13, p. 131002, 2024.

[107] The Pierre Auger Collaboration, “Cosmic-ray anisotropies in right ascension measured by the Pierre Auger Observatory,” *The Astrophysical Journal*, vol. 891, no. 2, p. 142, 2020.

[108] O. Deligny and Q. Luce, "Spectral feature at 10 EeV and systematics uncertainties in $S(1000)$," *GAP-2019-021*, 2021.

[109] The Pierre Auger collaboration, "Deep-learning based reconstruction of the shower maximum Xmax using the water-Cherenkov detectors of the Pierre Auger Observatory," *Journal of Instrumentation*, vol. 16, p. P07019, jul 2021.

[110] The Pierre Auger Collaboration, "Large scale distribution of ultra high energy cosmic rays detected at the pierre auger observatory with zenith angles up to 80°," *The Astrophysical Journal*, vol. 802, no. 2, p. 111, 2015.

[111] The Pierre Auger Collaboration, "Large-scale distribution of arrival directions of cosmic rays detected above 10^{18} eV at the Pierre Auger observatory," *The Astrophysical Journal Supplement Series*, vol. 203, no. 2, p. 34, 2012.

[112] T. Bister, September 2024. The Pierre Auger Collaboration online meeting of the Arrival Directions task.

[113] T. Bister and G. R. Farrar, "Constraints on UHECR Sources and Extragalactic Magnetic Fields from Directional Anisotropies," *The Astrophysical Journal*, vol. 966, p. 71, Apr. 2024.

[114] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant, "Array programming with NumPy," *Nature*, vol. 585, pp. 357–362, Sept. 2020.

[115] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python," *Nature Methods*, vol. 17, pp. 261–272, 2020.

[116] J. Pivarski, H. Schreiner, A. Hollands, P. Das, K. Kothari, A. Roy, J. Ling, N. Smith, C. Burr, and G. Stark, "Uproot," *Zenodo*, 2017.

[117] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in Science & Engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[118] Wes McKinney, "Data Structures for Statistical Computing in Python," in *Proceedings of the 9th Python in Science Conference* (Stéfan van der Walt and Jarrod Millman, eds.), pp. 56 – 61, 2010.

[119] M. L. Waskom, "seaborn: statistical data visualization," *Journal of Open Source Software*, vol. 6, no. 60, p. 3021, 2021.

[120] Astropy Collaboration, A. M. Price-Whelan, P. L. Lim, N. Earl, N. Starkman, L. Bradley, D. L. Shupe, A. A. Patil, L. Corrales, C. E. Brasseur, M. Nöthe, A. Donath, E. Tollerud, B. M. Morris, A. Ginsburg, E. Vaher, B. A. Weaver, J. Tocknell, W. Jamieson, M. H. van Kerkwijk, T. P. Robitaille, B. Merry, M. Bachetti, H. M. Günther, T. L. Aldcroft, J. A. Alvarado-Montes, A. M. Archibald, A. Bódi, S. Bapat, G. Barentsen, J. Bazán, M. Biswas, M. Boquien, D. J. Burke, D. Cara, M. Cara, K. E. Conroy, S. Conseil, M. W. Craig, R. M. Cross, K. L. Cruz, F. D'Eugenio, N. Dencheva, H. A. R. Devillepoix, J. P. Dietrich, A. D. Eigenbrot, T. Erben, L. Ferreira, D. Foreman-Mackey, R. Fox, N. Freij, S. Garg, R. Geda, L. Glattly, Y. Gondhalekar, K. D. Gordon, D. Grant, P. Greenfield, A. M. Groener, S. Guest, S. Gurovich, R. Handberg, A. Hart, Z. Hatfield-Dodds, D. Homeier, G. Hosseinzadeh, T. Jenness, C. K. Jones, P. Joseph, J. B. Kalmbach, E. Karamehmetoglu, M. Kałuszyński, M. S. P. Kelley, N. Kern, W. E. Kerzendorf, E. W. Koch, S. Kulumani, A. Lee, C. Ly, Z. Ma, C. MacBride, J. M. Maljaars, D. Muna, N. A. Murphy, H. Norman, R. O'Steen, K. A. Oman, C. Pacifici, S. Pascual, J. Pascual-Granado, R. R. Patil, G. I. Perren, T. E. Pickering, T. Rastogi, B. R. Roulston, D. F. Ryan, E. S. Rykoff, J. Sabater, P. Sakurikar, J. Salgado, A. Sanghi, N. Saunders, V. Savchenko, L. Schwardt, M. Seifert-Eckert, A. Y. Shih, A. S. Jain, G. Shukla, J. Sick, C. Simpson, S. Singanamalla, L. P. Singer, J. Singhal, M. Sinha, B. M. Sipőcz, L. R. Spitler, D. Stansby, O. Streicher, J. Šumak, J. D. Swinbank, D. S. Taranu, N. Tewary, G. R. Tremblay, M. de Val-Borro, S. J. Van Kooten, Z. Vasović, S. Verma, J. V. de Miranda Cardoso, P. K. G. Williams, T. J. Wilson, B. Winkel, W. M. Wood-Vasey, R. Xue, P. Yoachim, C. Zhang, A. Zonca, and Astropy Project Contributors, "The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package," *The Astrophysical Journal*, vol. 935, p. 167, Aug. 2022.

ERKLÄRUNG

Karlsruhe, den 07.11.2024

Erklärung der selbständigen Anfertigung meiner Dissertationsschrift

Hiermit versichere ich, dass ich die Dissertationsschrift mit dem Titel

Mass-dependent anisotropy of ultra-high-energy cosmic rays

selbständig und ohne unerlaubte fremde Hilfe verfasst habe. Dabei habe ich keine anderen, als die von mir angegebenen Hilfsmittel benutzt.

Edyvania Emily Pereira Martins