Cross-Validation of Muon Content in Extensive Air Showers with Surface and Underground Detectors of AugerPrime

For the attainment of the academic degree of

Doctorate in Science

from the KIT-Fakultät für Physik of the
Karlsruhe Institute of Technology (KIT)
and the Instituto de Tecnología "Prof. Jorge A. Sábato" of the
Universidad Nacional de San Martín (UNSAM)

presented

Dissertation

of

Lic. Carmina Pérez Bertolli

from San Miguel de Tucumán

Day of the oral examination: 05. 12. 2025
Referee: Prof. Dr. Brian Wundheiler
Co-referee: Prof. Dr. Ralph Engel
Supervisors: Dr. Markus Roth, Dr. Nicolás González

Cross-Validation of Muon Content in Extensive Air Showers with Surface and Underground Detectors of AugerPrime

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 Instituto de Tecnología "Prof. Jorge A. Sábato" der
Universidad Nacional de San Martín (UNSAM)

angenommene

Dissertation

von

Lic. Carmina Pérez Bertolli

aus San Miguel de Tucumán

Tag der mündlichen Prüfung: 05. 12. 2025
Referent: Prof. Dr. Brian Wundheiler
Korreferent: Prof. Dr. Ralph Engel
Betreuer: Dr. Markus Roth, Dr. Nicolás González

Cross-Validation of Muon Content in Extensive Air Showers with Surface and Underground Detectors of AugerPrime

Tesis presentada para optar por el título de

Doctor en Astrofísica

de la KIT-Fakultät für Physik del
Karlsruher Institut für Technologie (KIT)
y del Instituto de Tecnología "Prof. Jorge A. Sábato" de la
Universidad Nacional de San Martín (UNSAM)

por

Lic. Carmina Pérez Bertolli

de San Miguel de Tucumán

Fecha de la defensa oral: 05. 12. 2025
Director: Prof. Dr. Brian Wundheiler
Co-director: Prof. Dr. Ralph Engel
Colaboradores: Dr. Markus Roth, Dr. Nicolás González

To my grandmother Clara, the kindest person I have ever known.

"Caminante, no hay camino, se hace camino al andar."[a]

[a] "Traveler, there is no path; the path is made by walking."
— Antonio Machado

Abstract

Ultra-high-energy cosmic rays (UHECRs), charged particles with energies exceeding $\sim 10^{18}$ eV, stem from the most energetic phenomena observed in nature. Despite more than half a century of study, their origin, acceleration mechanisms, and composition remain open questions in astroparticle physics. Understanding these aspects is crucial for unveiling the processes governing particle acceleration in the most extreme astrophysical environments and for testing hadronic interactions at energies far beyond those accessible in terrestrial accelerators.

Because of their extremely low flux at the highest energies, UHECRs cannot be detected directly. Instead, they are studied through extensive air showers (EAS), cascades of secondary particles generated when primary cosmic rays interact with the atmosphere of Earth. Among the various shower components, muons provide a particularly sensitive probe of the hadronic cascade and of the primary mass composition. Over the past decades, measurements have consistently revealed an excess of muons relative to predictions from contemporary hadronic interaction models. This long-standing discrepancy, known as the *Muon Puzzle*, points to an incomplete understanding of hadronic multiparticle production or possibly to new physics. Precise determination of the muon content in EAS as a function of primary energy and zenith angle is therefore essential to constrain interaction models and to improve the interpretation of the mass composition of cosmic rays.

The Pierre Auger Observatory in Argentina, the world's largest facility for UHECR research, combines a vast array of water-Cherenkov detectors (WCDs) in the Surface Detector (SD) with fluorescence telescopes to record the properties of EAS with unprecedented precision. Its ongoing upgrade, AugerPrime, enhances the Observatory's sensitivity to the primary composition by adding complementary detectors. Of particular importance is the Underground Muon Detector (UMD), designed to measure the muonic component of air showers by exploiting the natural shielding of the soil to suppress the electromagnetic component. The UMD thereby provides a direct measurement of high-energy muons that reach the ground, offering an independent handle on the muon content of air showers.

This work focuses on the estimation of the muon density at ground level using data from the UMD and on the development of calibration techniques that link underground and surface observables. The analysis combines detailed Monte Carlo simulations of extensive air showers generated with CORSIKA and processed within the official Auger simulation and reconstruction framework $\overline{\text{Off}}\underline{\text{line}}$, with real data collected by the AugerPrime detectors. The relationship between the muon density underground, ρ^{ug}, and the true simulated muon density at the surface, ρ^{og}, was characterized as a function of primary energy and zenith angle, resulting in a robust and unbiased parameterization valid over the full range of simulated conditions. Complementary studies of muon propagation through soil were carried out using both

the Continuous Slowing Down Approximation (CSDA) and full GEANT4 simulations to precisely quantify the energy threshold imposed by the UMD shielding.

Based on these results, calibration functions were developed to connect the estimated on-ground muon density to the simulated muonic signal in the WCDs. In addition, pre-trained neural network estimators developed within the Pierre Auger Collaboration were employed to infer the muonic component of the WCD signals, enabling application to real data where the pure muon signal cannot be directly extracted.

Finally, the established calibration was applied to real data collected by the Pierre Auger Observatory, providing for the first time a direct cross-comparison between muon observables derived from the SD and those measured by the UMD. The methods and results presented here demonstrate that the UMD can be effectively employed to estimate the muon density at ground level and to calibrate surface muon observables. The obtained results are consistent with the latest UMD-based measurements of mass composition and muon deficit, thus contributing to a unified understanding of the muon content in extensive air showers and to the ongoing efforts within AugerPrime to resolve the Muon Puzzle.

Zusammenfassung

Ultrahochenergetische kosmische Strahlen (UHECR), geladene Teilchen mit Energien über $\sim 10^{18}$ eV, stammen von den energiereichsten Phänomenen in der Natur. Trotz mehr als eines halben Jahrhunderts intensiver Forschung bleiben ihr Ursprung, ihre Beschleunigungsmechanismen, und ihre Zusammensetzung offene Fragen in der Astroteilchenphysik. Das Verständnis dieser Aspekte ist entscheidend, um die Prozesse zu entschlüsseln, die die Teilchenbeschleunigung in den extremsten astrophysikalischen Umgebungen bestimmen, und um hadronische Wechselwirkungen bei Energien zu testen, die weit über denjenigen terrestrischer Beschleuniger liegen.

Aufgrund des äußerst geringen Flusses bei den höchsten Energien können UHECR nicht direkt detektiert werden. Stattdessen werden sie über ausgedehnte Luftschauer (EAS) untersucht, Kaskaden sekundärer Teilchen, die entstehen, wenn primäre kosmische Strahlen mit der Atmosphäre wechselwirken. Unter den verschiedenen Komponenten dieser Schauer sind Myonen besonders entscheidend für Rückschlüsse auf die hadronische Entwicklung und die Primärzusammensetzung. Messungen der letzten Jahrzehnte haben wiederholt einen Überschuss an Myonen gegenüber den Vorhersagen aktueller Wechselwirkungsmodelle gezeigt. Diese anhaltende Diskrepanz, bekannt als das *Myon Puzzle*, weist auf ein unvollständiges Verständnis der hadronischen Mehrteilchenproduktion oder auf mögliche neue physikalische Prozesse hin. Eine präzise Bestimmung des Myonenanteils in Abhängigkeit von Primärenergie und Zenitwinkel eines EAS ist daher von zentraler Bedeutung, um Wechselwirkungsmodelle zu verifizieren und die Interpretation der Zusammensetzung kosmischer Strahlung zu verbessern.

Das Pierre-Auger-Observatorium in Argentinien, der weltweit größte Detektor zur Untersuchung von UHECR, kombiniert ein weitläufiges feld von Wasser-Cherenkov-Detektoren (WCD) als Oberflächendetektor (SD) mit Fluoreszenzteleskopen, um die Eigenschaften von EAS mit bisher unerreichter Genauigkeit zu rekonstruieren. Das laufende AugerPrime-Upgrade erhöht die Empfindlichkeit des Experiments gegenüber der Primärzusammensetzung durch zusätzliche Detektoren. Von besonderer Bedeutung ist der Untergrund-Myonendetektor (UMD), der entwickelt wurde, um die myonische Komponente der Luftschauer direkt zu messen, wobei die elektromagnetische Komponente durch die natürliche Abschirmung des Bodens unterdrückt wird. Dadurch ermöglicht der UMD eine direkte Messung hochenergetischer Myonen, die den Boden erreichen, und bietet eine unabhängige Möglichkeit, den Myonenanteil in Luftschauern zu untersuchen.

Diese Arbeit befasst sich mit der Bestimmung der Myonendichte am Boden mithilfe von UMD-Daten und der Entwicklung von Kalibrierungsverfahren, die unterirdisch und oberirdisch gemessene Observablen miteinander verknüpfen. Die Analyse kombiniert detaillierte Monte-Carlo-Simulationen ausgedehnter Luftschauer, die mit CORSIKA erzeugt und innerhalb der offiziellen Simulations- und Rekonstruktionssoftware der Pierre-Auger-Kollaboration, $\overline{\text{Off}}\underline{\text{line}}$, verarbeitet wurden, mit

realen Daten der AugerPrime-Detektoren. Die Beziehung zwischen der unterirdisch gemessenen Myonendichte, ρ^{ug}, und der simulierten Myonendichte am Boden, ρ^{og}, wurde als Funktion der Primärenergie und des Zenitwinkels charakterisiert, was zu einer robusten und unverzerrten Parametrisierung führt, die im gesamten Bereich der simulierten Schauerbedingungen gültig ist. Ergänzende Studien zum Myonentransport im Boden unter Verwendung sowohl der kontinuierlichen Energieverlustnäherung (CSDA) als auch vollständiger GEANT4-Simulationen ermöglichen eine präzise Bestimmung der durch die UMD-Abschirmung vorgegebenen Energieschwelle.

Aufbauend auf diesen Resultaten wurden Kalibrationen entwickelt, die die geschätzte Myonendichte am Boden mit dem simulierten myonischen Signal in den WCDs verbinden. Darüber hinaus wurden vortrainierte neuronale Netzwerke, die innerhalb der Pierre-Auger-Kollaboration entwickelt wurden, verwendet, um die myonische Komponente der WCD-Signale abzuleiten und auf reale Daten anzuwenden, bei denen das reine Myonensignal nicht direkt extrahiert werden kann.

Schließlich wurde die entwickelte Kalibrierung auf reale Daten des Pierre-Auger-Observatoriums angewandt und ermöglichte erstmals einen direkten Vergleich zwischen den aus dem SD abgeleiteten und den vom UMD gemessenen Myonensignalen. Die Methoden sowie Ergebnisse, die in dieser Arbeit vorgestellt werden, zeigen, dass der UMD effektiv zur Bestimmung der Myonendichte am Boden und zur Kalibrierung oberirdisch gemessener Myonenobservablen eingesetzt werden kann. Die erzielten Resultate stimmen mit den neuesten UMD-basierten Messungen zur Massenzusammensetzung und zum Myonendefizit überein und tragen so zu einem einheitlicheren Verständnis des Myoneninhalts in Luftschauern sowie zu den laufenden Bemühungen von AugerPrime bei, das Myonenrätsel zu lösen.

Resumen

Los rayos cósmicos de ultra alta energía (UHECR, por sus siglas en inglés), partículas cargadas con energías superiores a $\sim 10^{18}$ eV, se encuentran entre los fenómenos más energéticos observados en la naturaleza. A pesar de más de medio siglo de estudios, su origen, mecanismos de aceleración y composición química siguen siendo cuestiones abiertas en la física de astropartículas. Comprender estos aspectos es fundamental para desentrañar los procesos que gobiernan la aceleración de partículas en los entornos astrofísicos más extremos y para poner a prueba los modelos de interacciones hadrónicas a energías muy superiores a las alcanzables en los aceleradores terrestres.

Debido a su flujo extremadamente bajo a las energías más altas, los UHECR no pueden detectarse de forma directa. En su lugar, se estudian a través de las lluvias atmosféricas extendidas (EAS, por sus siglas en inglés), cascadas de partículas secundarias generadas cuando los rayos cósmicos primarios interactúan con la atmósfera. Entre las distintas componentes de la lluvia, el número de muones constituye un observable especialmente sensible a la composición en masa del primario. A lo largo de las últimas décadas, las mediciones han revelado de manera consistente un exceso de muones en comparación con las predicciones de los modelos contemporáneos de interacción hadrónica. Esta discrepancia persistente, conocida como el *Muon Puzzle*, señala una comprensión incompleta de la producción hadrónica múltiple o la posible existencia de nueva física. La determinación precisa del contenido de muones en función de la energía primaria y del ángulo cenital es, por lo tanto, esencial para restringir los modelos de interacción y mejorar la interpretación de la composición de los rayos cósmicos.

El Observatorio Pierre Auger, en Argentina, la instalación más grande del mundo dedicada al estudio de los UHECR, combina una extensa red de detectores Cherenkov con telescopios de fluorescencia para registrar las propiedades de las EAS con una precisión sin precedentes. Su actualización en curso, AugerPrime, mejora la sensibilidad del Observatorio a la composición primaria mediante la incorporación de detectores complementarios. De particular importancia es el Detector Subterráneo de Muones (UMD), diseñado para medir la componente muónica de las lluvias atmosféricas aprovechando el blindaje natural del suelo para suprimir el fondo electromagnético. El UMD proporciona así una medición directa de los muones de alta energía que alcanzan la superficie, ofreciendo una herramienta independiente para estudiar el contenido muónico de las lluvias.

Este trabajo se centra en la estimación de la densidad de muones a nivel del suelo utilizando datos del UMD y en el desarrollo de técnicas de calibración que vinculen los observables subterráneos y de superficie. El análisis combina simulaciones detalladas de lluvias atmosféricas generadas con CORSIKA y procesadas en el marco oficial de simulación y reconstrucción de Auger, Offline, junto con datos reales recopilados por los detectores de AugerPrime. La relación entre la densidad de muones

subterránea, ρ^{ug}, y la densidad real en superficie, ρ^{og}, se caracterizó como función de la energía primaria y del ángulo cenital, resultando en una parametrización robusta y libre de sesgos válida en todo el rango de condiciones simuladas. Estudios complementarios sobre la propagación de muones en el suelo se llevaron a cabo empleando tanto la Aproximación de Pérdida Continua de Energía (CSDA) como simulaciones completas con GEANT4, permitiendo cuantificar con precisión el umbral energético impuesto por el blindaje del UMD.

A partir de estos resultados, se desarrollaron funciones de calibración que conectan la densidad de muones estimada en superficie con la señal muónica simulada en los Detectores Cherenkov de Agua (WCD) del Detector de Superficie (SD). Además, se emplearon estimadores basados en redes neuronales previamente entrenadas y desarrolladas por la Colaboración Pierre Auger para inferir la componente muónica de las señales de los WCD, lo que permite su aplicación a datos reales en los que la señal pura de muones no puede extraerse directamente.

Finalmente, la calibración establecida se aplicó a datos reales recopilados por el Observatorio Pierre Auger, proporcionando por primera vez una comparación directa entre los observables muónicos derivados del SD y los medidos por el UMD. Los métodos y resultados presentados en esta tesis demuestran que el UMD puede utilizarse eficazmente para estimar la densidad de muones a nivel del suelo y calibrar los observables muónicos de superficie. Los resultados obtenidos son consistentes con las mediciones más recientes basadas en el UMD sobre composición y déficit de muones, contribuyendo así a una comprensión unificada del contenido muónico en las lluvias atmosféricas y a los esfuerzos en curso de AugerPrime por resolver el *Muon Puzzle*.

Contents

Chapter I

Introduction

Ultra-high-energy cosmic rays (UHECRs), charged particles with energies exceeding $\sim 10^{18}$ eV, are among the most energetic phenomena observed in nature. Despite more than half a century of study, their origin, acceleration mechanisms, and chemical composition remain open questions in astroparticle physics. Understanding these aspects is fundamental to unveiling the processes that govern particle acceleration in the most extreme astrophysical environments and to testing hadronic interactions at energies far beyond those accessible in terrestrial accelerators.

Because of their extremely low flux at the highest energies, direct detection of UHECRs is unfeasible. Instead, their study relies on observation of the cascades of secondary particles they produce when they interact with the atmosphere, known as *extensive air showers* (EAS). The energy, mass and arrival direction of the primary particle must be inferred indirectly from the measured distributions of particles on the ground and from the development of the shower in the atmosphere. Among the various shower components, muons play a central role: being less affected by multiple scattering, and because of their high mass ($\sim 200\, m_e$) which allows them to travel longer distances through the atmosphere, they carry direct information about the hadronic interactions in the shower core and the mass of the primary cosmic ray.

During the past decades, several experiments have revealed that the number of muons measured on the ground significantly exceeds the predictions from simulations based on current hadronic interaction models [1]. This persistent discrepancy, known as the *Muon Puzzle*, indicates either an incomplete understanding of hadronic multiparticle production or the presence of new physical processes. Precise measurements of the muon content as a function of primary energy and zenith angle are therefore crucial for constraining interaction models and for improving the interpretation of cosmic-ray composition. Understanding this discrepancy is critical not only for improving our knowledge of hadronic interactions but also for reliably inferring the mass composition of UHECRs. Accurate reconstruction of the muon content independently of the electromagnetic component is of great value to investigate hadronic physics and cosmic ray composition with unprecedented precision [2].

The Pierre Auger Observatory [3] in Argentina is currently the largest facility dedicated to the study of UHECRs. Its hybrid design, which combines a vast ar-

ray of water-Cherenkov surface detectors (SD) with fluorescence telescopes, allows a detailed reconstruction of EAS properties over a wide energy range. With the AugerPrime upgrade [2], the Observatory has been enhanced with additional detectors that increase its sensitivity to the mass composition at the highest energies. Of particular importance is the Underground Muon Detector (UMD) [4], designed to directly measure the muon content by exploiting the natural shielding of the soil, which suppresses the electromagnetic component of the shower and enables cleaner observation of high-energy muons.

This work focuses on the estimation of the muon density at ground level using data from the UMD and on the development of calibration techniques that relate the underground and surface observables. To achieve this, detailed Monte Carlo simulations of extensive air showers and detector responses were used, together with measurements from the AugerPrime detectors. The work aims to establish reliable estimators of the muon content, quantify their dependence on primary energy and zenith angle, and provide a basis for future physics analyses addressing the Muon Puzzle and composition studies.

Chapter 2 introduces the physics of cosmic rays, describing their energy spectrum, composition, and the main theoretical models proposed to explain their acceleration and propagation. It also discusses the key experimental challenges associated with their detection at ultra-high energies. Chapter 3 presents the Pierre Auger Observatory, its hybrid detection techniques, and the configuration of its main detector components. Special attention is given to the AugerPrime upgrade, particularly to the Underground Muon Detector (UMD), whose data form the basis of this thesis. Chapter 4 presents a complementary approach for estimating the on-ground muon density using muon stopping power values in soil, based on Continuous Slowing Down Approximation (CSDA) tables, and compares the results with full Geant4 simulations. Chapter 5 focuses on the estimation of the muon density at the surface from UMD measurements. It discusses the relation between the underground and on-ground muon densities and establishes a procedure to obtain the corresponding estimator. Chapter 6 details the development of the calibration functions that link the estimated on-ground muon density with the simulated muon density in the Water-Cherenkov Detectors (WCD) of the Surface Detector (SD). Subsequently, Chapter 7 presents an alternative calibration based on a neural network estimator of the muonic signal in the WCDs, a quantity that cannot be directly obtained from real data. Chapter 8 applies the developed estimators to real data collected by the Pierre Auger Observatory, providing for the first time a calibration between estimations of the muonic component derived from the SD and those obtained from the UMD measurements. Finally, Chapter 9 summarizes the main results and conclusions of this work, highlighting their relevance within the broader context of the AugerPrime program and outlining prospects for future analyses within the Collaboration.

Chapter II

Cosmic Rays

Ultra-high-energy cosmic rays (UHECRs), ionized nuclei with energies above $\sim 10^{18}$ eV, are the most energetic particles observed in nature, with a spectrum extending from below 1 GeV to about 10^{20} eV. Despite more than half a century of study, key questions regarding their origin, acceleration mechanisms, and elemental composition remain unresolved. The flux of UHECRs is extremely low: at 10^{17} eV the rate is only about 10 particles per km^2 per day and decreases rapidly with energy. Due to this rarity, UHECRs cannot be detected directly. Instead, they are studied through *extensive air showers* (EAS), cascades of billions of secondary particles generated when a cosmic ray interacts with the atmosphere. At the highest energies, these showers can spread over tens of km^2 and are therefore observed with large ground-based detector arrays. Interpreting such measurements requires comparison with air-shower simulations for different nuclear primaries. These simulations, in turn, rely on hadronic interaction models that extrapolate accelerator data far beyond the energies probed at the LHC [5], which introduces large systematic uncertainties.

Of particular interest are UHECRs with energies greater than $\sim 10^{18}$ eV, because they probe hadronic interactions at centers-of-mass energies vastly exceeding those accessible in human-made accelerators. Unlike photons or neutrinos, however, the trajectories of charged cosmic rays are strongly deflected by Galactic and extragalactic magnetic fields, preventing a straightforward association with astrophysical sources and making classical cosmic ray astronomy not yet feasible. While low-energy cosmic rays can be linked to solar and stellar phenomena such as coronal mass ejections [6], the sources of UHECRs remain uncertain and are the subject of intense research. A crucial step toward progress is the determination of the composition as a function of energy. Although protons dominate, heavier nuclei contribute significantly and their relative abundances vary with energy. Because the primary mass cannot be measured directly, it must be inferred from EAS observables. Two quantities are particularly sensitive to composition: the depth of maximum shower development, X_{max} and the number of muons in the cascade, N_μ. Both correlate with the primary mass but require detailed comparisons with simulations of different nuclei. Those simulations operate in an untested energy regime and are further complicated by intrinsic shower-to-shower fluctuations, which can obscure the relatively small differences between

neighboring mass groups. Consequently, precise composition measurements remain one of the central challenges in UHECR physics.

2.1. Energy Spectrum

The energy spectrum observed at Earth spans more than eleven decades, from $\sim 10^9$ eV to beyond 10^{20} eV, and, over most of this range, it is well described by broken power laws of the form $d\Phi/dE \propto E^{-\gamma}$ with spectral index $\gamma \approx 2.6$–3.3 depending on the interval. Formally, the differential flux

$$\frac{\mathrm{d}N}{\mathrm{d}E\,\mathrm{d}A\,\mathrm{d}t\,\mathrm{d}\Omega} \tag{2.1}$$

represents the number of particles detected per unit of energy E, area A, time t, and solid angle Ω. The all-particles spectrum compiled from several experiments is shown in Fig. 2.1.

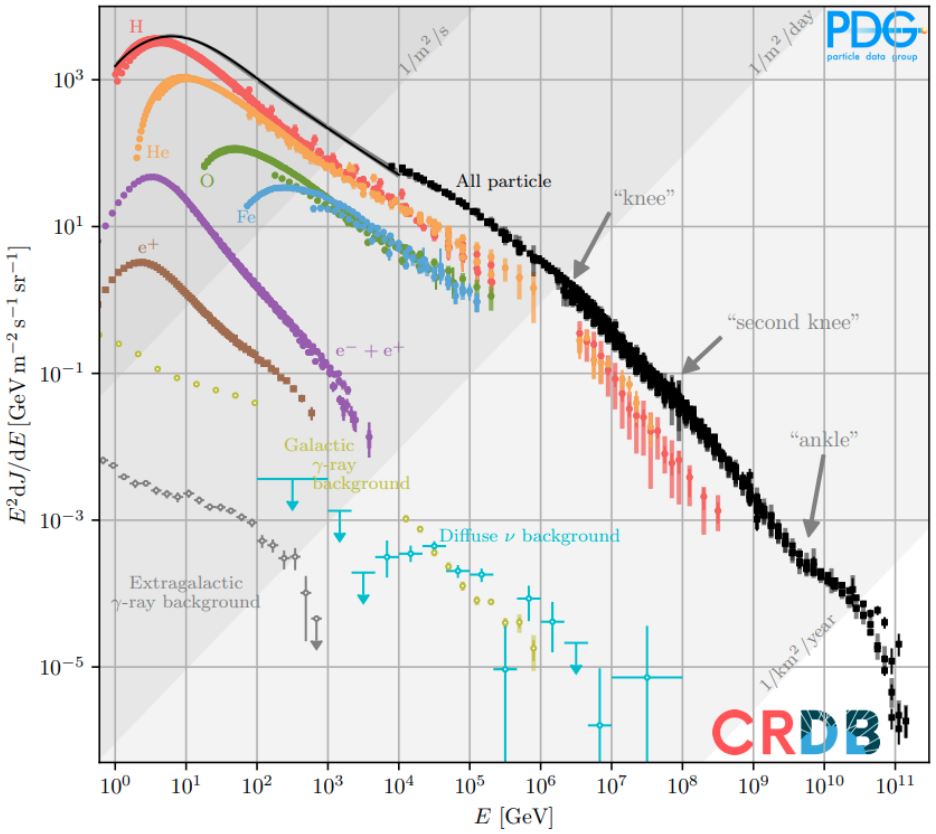


Figure 2.1: The spectrum of cosmic rays. Shown are measurements of the intensity of charged and neutral CRs, multiplied by kinetic energy squared. Figure extracted from [7].

Several distinct features are present in the steeply falling spectrum and are interpreted as transitions in the dominant sources or propagation effects. Around

$E \sim 10^{15}$ eV, the *knee* marks a steepening from $\gamma \sim 2.7$ to ~ 3.1. At $E \sim 10^{17}$ eV, a *second knee* appears and is often associated with the fading contribution of the heaviest Galactic nuclei. At $E \sim 5 \times 10^{18}$ eV, the *ankle* manifests itself as a spectral hardening and is commonly linked to the transition from Galactic to extragalactic dominance. Finally, above $E \gtrsim 5 \times 10^{19}$ eV, a pronounced *suppression* occurs. For protons, this suppression can be explained by the Greisen–Zatsepin–Kuzmin (GZK) effect—pion production in interactions with the cosmic microwave background (CMB) that limits the propagation distance to ~ 100 Mpc—whereas for nuclei ($A > 1$) photodisintegration through giant dipole resonance leads to emission of nucleons [8, 9]. Although substantial progress has been made, the detailed physical origin of these features remains under discussion. The prevailing picture is that cosmic rays up to the knee are predominantly Galactic—likely accelerated by first-order Fermi processes in supernova remnants—whereas above the ankle the flux is largely extragalactic. This view is supported by a large-scale dipolar anisotropy observed for $E > 8 \times 10^{18}$ eV, with the dipole direction pointing away from the Galactic plane [10]. The transition between the two components is expected between $\sim 10^{17}$ and $\sim 10^{18.5}$ eV, although its precise nature is still debated. At energies below $\sim 10^{14}$ eV the flux is high enough for direct detection by balloon- or satellite-borne instruments such as PAMELA [11], CREAM [12], and AMS [13]. Above this threshold, indirect techniques based on EAS become necessary.

2.2. Sources and Propagation

The spatial distribution and general characteristics of cosmic-ray sources can be constrained from the observed spectrum and from propagation physics. Nevertheless, the exact origins of individual ultra-high-energy cosmic rays (UHECRs) remain unknown, and no source has yet been conclusively identified. Since the seminal work of Fermi, theoretical models have established acceleration mechanisms that naturally yield power-law energy spectra. In the *second-order* process, charged particles gain energy through elastic scattering off moving magnetic irregularities within interstellar clouds, with the mean energy gain scaling with the square of the cloud velocity [14]. This process, however, is relatively inefficient. A more effective mechanism is *first-order* acceleration at shock fronts, known as diffusive shock acceleration (DSA), where the average fractional energy gain per cycle increases linearly with the shock velocity [15, 6].

In this framework, the energy after n shock crossings is given by

$$E_n = E_0 \left(1 + \frac{4}{3}\beta\right)^n , \tag{2.2}$$

where $\beta = v_{\text{shock}}/c$. Inverting this relation gives

$$n = \frac{\ln(E_n/E_0)}{\ln(1 + 4\beta/3)}. \tag{2.3}$$

If each cycle has an escape probability p_{esc}, the number of particles that remain

confined after n cycles is

$$N(E > E_n) = \sum_{m=n}^{\infty} (1 - p_{\text{esc}})^m = \frac{(1 - p_{\text{esc}})^n}{p_{\text{esc}}}. \tag{2.4}$$

Using Eq. 2.3, the resulting integral spectrum is

$$N(E > E_n) \propto \frac{1}{p_{\text{esc}}} \left(\frac{E}{E_0} \right)^{-\gamma}, \tag{2.5}$$

with spectral index $\gamma = -\ln(1-p_{\text{esc}})/\ln(1+4\beta/3)$. This mechanism naturally produces a power law consistent with the observed cosmic-ray spectrum.

Efficient acceleration requires that charged particles remain confined to the acceleration region by the magnetic field. This condition can be expressed by requiring that the Larmor radius R_L of a particle does not exceed the characteristic size of the acceleration region R_s. For a relativistic particle, the Larmor radius is given by

$$R_L = \frac{E}{ZeB}, \tag{2.6}$$

and the condition $R_L \lesssim R_s$ leads to

$$R_s \gtrsim \frac{E}{ZeB}. \tag{2.7}$$

Here, B denotes the magnetic field strength of the source, Z the particle charge number, e the elementary charge. This requirement, commonly referred to as the *Hillas criterion* [16], defines a necessary condition for acceleration to a given energy. However, satisfying Eq. 2.7 is not sufficient to ensure that such energies are reached, as additional constraints such as acceleration timescales, energy losses, and source energetics must also be fulfilled.

Figure 2.2 illustrates the Hillas diagram, which compares the sizes and magnetic field strengths of potential astrophysical accelerators. The lines represent the minimum product $R_s B$ required for the acceleration of particles to 10^{20} eV for various β values. From this diagram, it is evident that sources such as normal galaxies and supernova remnants do not meet the confinement condition of Eq. 2.7 and are therefore excluded as UHECR origins. Even for candidates that satisfy the Hillas criterion, sufficient energy budget and efficiency are required to reproduce the observed cosmic-ray flux.

During propagation, charged particles experience deflections by cosmic magnetic fields and suffer energy losses through interactions with background radiation fields. For protons at the highest energies, the dominant processes are *photo-pion production* and *Bethe–Heitler pair production*. In photo-pion production, a proton interacts with a photon of the cosmic microwave background (CMB), producing a Δ^+ resonance that decays as

$$p + \gamma_{\text{CMB}} \rightarrow \Delta^+ \rightarrow p + \pi^0, \tag{2.8}$$

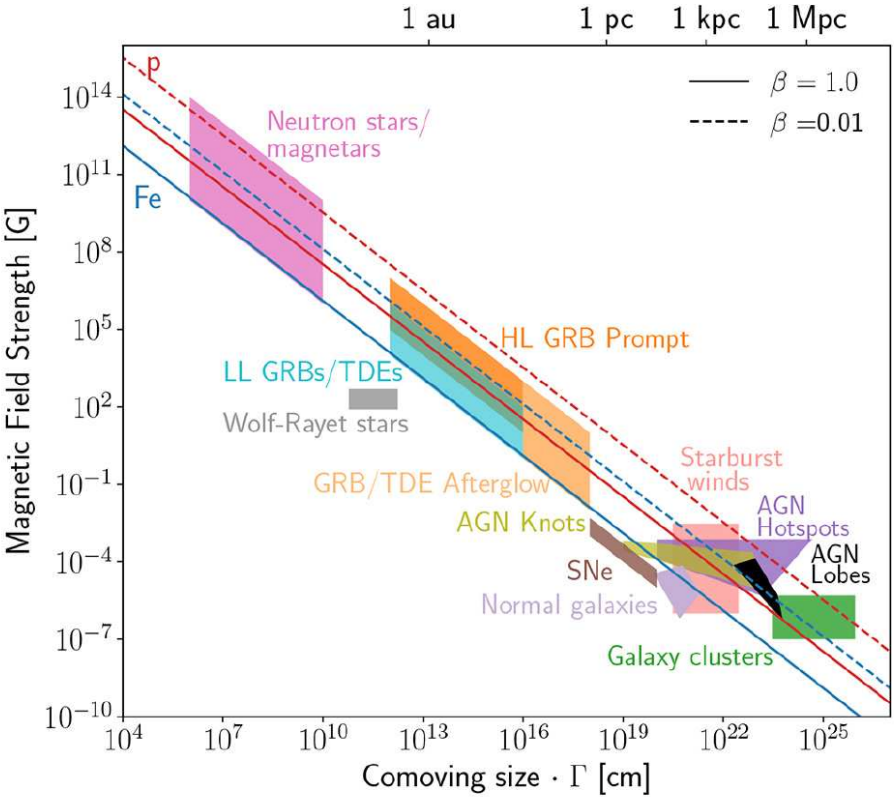


Figure 2.2: Sizes and magnetic field strengths of several candidate sources of cosmic rays. The solid and dashed lines represent the confinement limits derived from the Hillas criterion (Eq. 2.7), beyond which acceleration to 10^{20} eV is possible. The magnetic field strength is shown in the comoving frame of the source, and Γ denotes the Lorentz factor to account for relativistic motion. Figure reproduced from [17].

with the neutral pion carrying roughly 10–20% of the proton's initial energy and promptly decaying into two photons,

$$\pi^0 \rightarrow \gamma + \gamma. \tag{2.9}$$

If UHECRs were primarily protons, these interactions would generate a diffuse cosmogenic γ-ray flux around 10^{18} eV. Such a flux has not been observed, and only upper limits exist to date. Charged-pion production,

$$p + \gamma_{\mathrm{CMB}} \rightarrow \Delta^+ \rightarrow n + \pi^+, \tag{2.10}$$

leads to neutrino production through π^+ and subsequent μ^+ decays. Pair production,

$$p + \gamma_{\mathrm{CMB}} \rightarrow p + e^+ + e^-, \tag{2.11}$$

also contributes to continuous energy losses, though at lower rates.

For nuclei ($A > 1$), the dominant channel becomes *photodisintegration*:

$$X_A + \gamma_{\mathrm{CMB}} \rightarrow X_{A-k} + k, \tag{2.12}$$

triggered mainly by the Giant Dipole Resonance (8–30 MeV), the Quasi-Deuteron process (20–150 MeV), and the Baryonic Resonance ($\gtrsim$ 150 MeV) in the nuclear rest frame [18]. The cumulative attenuation from these processes defines the *GZK horizon*, the maximum distance from which UHECRs with $E \gtrsim E_{\text{th}}$ can reach Earth before losing most of their energy. Although originally formulated for protons, this concept extends naturally to heavier nuclei when photodisintegration is taken into account.

2.3. Extensive Air Showers

Because the flux above $\sim 10^{15}$ eV is too low for direct detection, cosmic rays at higher energies are studied via the extensive air showers they produce when interacting with the upper atmosphere. A high-energy primary initiates a chain of secondary interactions that multiply rapidly, generating a particle cascade whose footprint at ground level can span hundreds of meters to several kilometers, with the largest particle density near the shower axis. At the highest energies, the flux drops to about one particle per km^2 per century near the suppression region, making large-scale indirect detection indispensable. Ground-based observatories employ two complementary techniques: surface detectors that sample the secondary particles at ground and fluorescence telescopes that observe the faint ultraviolet light emitted by atmospheric nitrogen excited along the shower track. The shower development is commonly described as a function of atmospheric depth X, which accounts for the changing density along the path

$$X(h) = \frac{1}{\cos\theta} \int_h^\infty \rho(h')\, \mathrm{d}h', \tag{2.13}$$

where $\rho(h')$ is the atmospheric density at altitude h' and θ is the zenith angle, naturally incorporating the longer trajectories of inclined showers.

Figure 2.3 summarizes the main components: a hadronic core (pions, kaons, neutrons, and a small fraction of heavier mesons and baryons); an electromagnetic component (photons and $e^\pm$ largely from $\pi^0 \to \gamma\gamma$, pair production, and bremsstrahlung); a muonic component ($\mu^\pm$ from $\pi^\pm$ and $K^\pm$ decays, typically less than about 2% of the total particle number but crucial for composition); and an invisible component of neutrinos from meson and muon decays. Within this framework, two observables are directly sensitive to both energy and mass [20, 21]: the electromagnetic depth of maximum, X_{max}, and the muon content at ground, N_μ.

2.4. The Heitler Model

Although air-shower development can be modeled using cascade equations or Monte Carlo simulations, it is instructive to consider the Heitler model [22], which provides an intuitive analytic description of electromagnetic cascades. In this simplified picture, at each radiation length λ_{em} a particle splits into two particles of half the energy.

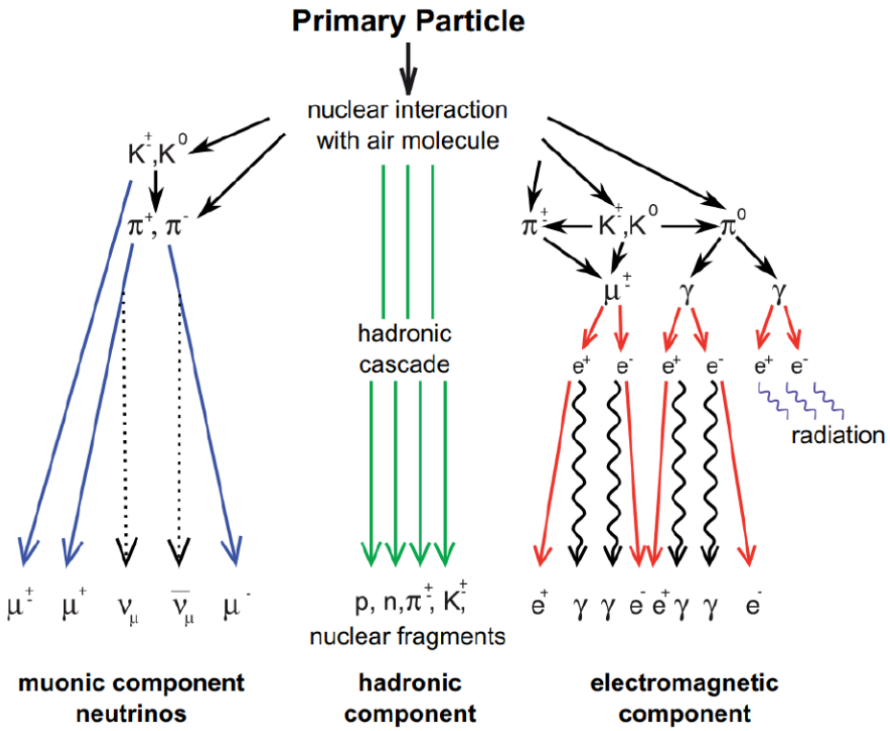


Figure 2.3: Schematic development of an extensive air shower based on the Heitler–Matthews model. Adapted from [19].

After n generations, the shower properties are given by

$$\begin{aligned} N_n &= 2^n, \\ E_n &= \frac{E_0}{2^n}, \\ X_n &= n\,\lambda_{\mathrm{em}}, \end{aligned} \tag{2.14}$$

where λ_{em} denotes the radiation length in air and E_0 is the primary energy. Particle multiplication ceases when the mean energy per particle reaches the critical energy ζ_{em} (for electrons in air, $\sim 84\,\mathrm{MeV}$), yielding

$$n_c = \frac{\ln(E_0/\zeta_{\mathrm{em}})}{\ln 2}. \tag{2.15}$$

At shower maximum,

$$\begin{aligned} N_{\max} &= \frac{E_0}{\zeta_{\mathrm{em}}}, \\ X_{\max} &= \lambda_{\mathrm{em}} \ln\!\left(\frac{E_0}{\zeta_{\mathrm{em}}}\right) \end{aligned} \tag{2.16}$$

Numerical factors of order unity, such as $\ln 2$, are absorbed into the effective radiation length in this simplified description. These relations encapsulate the essential behavior of electromagnetic cascades: exponential particle multiplication with depth,

followed by a well-defined maximum when energy degradation halts further multiplication.

2.5. The Heitler–Matthews Model

The Heitler model can be extended to hadron-induced air showers following the approach introduced by Matthews [20]. Consider a proton of energy E_0 entering the atmosphere. Hadronic interactions occur on a characteristic interaction length $\lambda_I \simeq 120\,\mathrm{g\,cm^{-2}}$. At each interaction, a total of N_{mult} pions are produced; a fraction α are charged and the remaining fraction $1-\alpha$ are neutral, such that $N_{\mathrm{ch}} = \alpha N_{\mathrm{mult}}$ and $N_0 = (1-\alpha)N_{\mathrm{mult}}$. Neutral pions decay promptly via $\pi^0 \to 2\gamma$, feeding the electromagnetic component of the shower, while charged pions re-interact after another interaction length λ_I, sustaining the hadronic cascade.

Assuming equal energy sharing among secondaries, the energy per pion after n generations is

$$E_n = \frac{E_0}{N_{\mathrm{mult}}^n},$$

while the number of charged pions evolves as

$$N_{\pi^\pm} = N_{\mathrm{ch}}^n.$$

When the pion energy drops below the critical value ζ, for which decay dominates over further interaction, the generation index

$$n_{\mathrm{dec}} = \frac{\ln(E_0/\zeta)}{\ln N_{\mathrm{mult}}} \tag{2.17}$$

marks the transition to muon production. Assuming that all charged pions decay, the total number of muons at ground is given by

$$N_\mu = \left(\frac{E_0}{\zeta}\right)^\beta, \tag{2.18}$$

where

$$\beta = \frac{\ln(\alpha N_{\mathrm{mult}})}{\ln N_{\mathrm{mult}}},$$

with simulations typically yielding $\beta \in [0.85,\, 0.92]$.

The partition of energy between the hadronic and electromagnetic components after n generations can be written as

$$E_{\mathrm{had}} = \left(\tfrac{2}{3}\right)^n E_0, \qquad E_{\mathrm{EM}} = \left[1 - \left(\tfrac{2}{3}\right)^n\right] E_0, \tag{2.19}$$

such that after approximately $n \simeq 6$ generations about 90% of the primary energy has been transferred to the electromagnetic component and dissipated through ionization losses in the atmosphere [6].

If only the electromagnetic subcascade initiated by photons from the first hadronic interaction is considered, the depth of shower maximum for proton-induced showers can be approximated by treating the electromagnetic cascade as initiated by a photon of energy $E_0/(2N_{\rm mult})$. Using the Heitler description of electromagnetic cascades, this leads to [6]

$$X^p_{\rm max} \simeq \lambda_p + \lambda_{\rm em} \ln\left(\frac{E_0}{2N_{\rm mult}\zeta_{\rm em}}\right), \tag{2.20}$$

where λ_p denotes the depth of the first interaction.

The logarithmic dependence of the depth of shower maximum on the primary energy defines the elongation rate,

$$D \equiv \frac{dX_{\rm max}}{d\ln E_0}. \tag{2.21}$$

In realistic hadronic showers, the elongation rate differs from the purely electromagnetic value $\lambda_{\rm em}$, reflecting the energy dependence of hadronic interaction lengths and particle multiplicities.

For a nucleus of mass number A and total energy E_0, the superposition model treats the shower as the incoherent sum of A nucleon-induced subshowers, each with energy E_0/A. The resulting muon number is

$$N^A_\mu = A\left(\frac{E_0/A}{\zeta}\right)^\beta = A^{1-\beta}\left(\frac{E_0}{\zeta}\right)^\beta, \tag{2.22}$$

while the depth of shower maximum becomes

$$X^A_{\rm max} = X^p_{\rm max} - D\ln A. \tag{2.23}$$

These relations lead to standard composition estimators:

$$\langle \ln A\rangle = \ln 56\,\frac{\langle \ln N^A_\mu\rangle - \langle \ln N^p_\mu\rangle}{\langle \ln N^{\rm Fe}_\mu\rangle - \langle \ln N^p_\mu\rangle}, \tag{2.24}$$

$$\langle \ln A\rangle = \ln 56\,\frac{\langle X^A_{\rm max}\rangle - \langle X^p_{\rm max}\rangle}{\langle X^{\rm Fe}_{\rm max}\rangle - \langle X^p_{\rm max}\rangle}. \tag{2.25}$$

These provide a practical bridge between EAS observables and the average mass composition.

2.6. The Muon Puzzle

Determining the mass composition above the ankle is pivotal because different astrophysical scenarios, such as maximum-rigidity acceleration at the source or photodisintegration during propagation, can reproduce the observed spectrum while implying different elemental mixtures. Converting air-shower observables into the mean logarithmic mass $\langle \ln A\rangle$ relies on detailed simulations (see Section 2.3). The two

most sensitive observables are $X_{\max}$ and N_μ. The muon content can be measured with high statistics using shielded detectors (e.g., the Underground Muon Detector at the Pierre Auger Observatory) or by analyzing very inclined showers with unshielded arrays where the electromagnetic component is largely absorbed [23, 24]. However, muon-based inferences carry larger systematic uncertainties, because muons originate from the hadronic sector of the cascade, which is more sensitive to model assumptions. Several experiments have reported that the composition inferred from muon measurements (via Eq. 2.24) is systematically heavier than that inferred from $X_{\max}$ (via Eq. 2.25), indicating an inconsistency between hadronic and electromagnetic modeling.

Key aspects of this discrepancy, often termed the *muon puzzle*, remain unsettled: the energy at which the deficit appears, its scaling with energy, and its magnitude are not yet firmly established; moreover, although many experiments see a deficit, not all do [25, 26]. Additional measurements over a broad energy range, preferably on a common energy scale, are needed to clarify the picture. The broader challenge is the persistent mismatch between simulations and data for the muon content of EAS at high and ultra-high energies [25]. Intriguingly, the effect seems to emerge already at TeV scales, suggesting that similar features should be visible at collider energies [27]. Addressing this requires both refinements to shower simulation frameworks and improved hadronic interaction models, especially in the forward phase space and in channels that affect baryon and strangeness production.

Uncertainties in present measurements arise from two sources: experimental errors and, more importantly, the model dependence in converting observables to mass-sensitive quantities such as $\langle \ln A \rangle$. The latter drives substantial spread among predictions, as illustrated in Figure 2.4.

The muon deficit has been reported across multiple experiments, e.g., The Pierre Auger Observatory, HiRes/MIA and NEVOD-DECOR [28], supporting the view that state-of-the-art simulations systematically underestimate muon production. Because the muon density at ground depends on primary energy, zenith angle, lateral distance, and detector threshold, direct inter-experiment comparisons are difficult. To mitigate this, measurements are often expressed relative to simulations using the z-scale [28],

$$z := \frac{\ln(N^{\text{data}}_{\mu,\text{det}}) - \ln(N^{p}_{\mu,\text{det}})}{\ln(N^{\text{Fe}}_{\mu,\text{det}}) - \ln(N^{p}_{\mu,\text{det}})}, \tag{2.26}$$

where $N^{\text{data}}_{\mu,\text{det}}$ is the measured muon density and $N^{p}_{\mu,\text{det}}$ and $N^{\text{Fe}}_{\mu,\text{det}}$ are the simulated densities for proton and iron primaries. Data–Monte Carlo comparisons also benefit from cross-calibrating energy scales across experiments by aligning all-particle fluxes [29, 30]. The resulting z-values compiled for different arrays are shown in Figure 2.5 for two hadronic models (EPOS-LHC and QGSJetII-04). Up to $\sim 10^{16}$ eV, measurements and simulations are broadly consistent; above this, the discrepancy grows, reinforcing the muon puzzle as a central challenge in cosmic-ray physics.

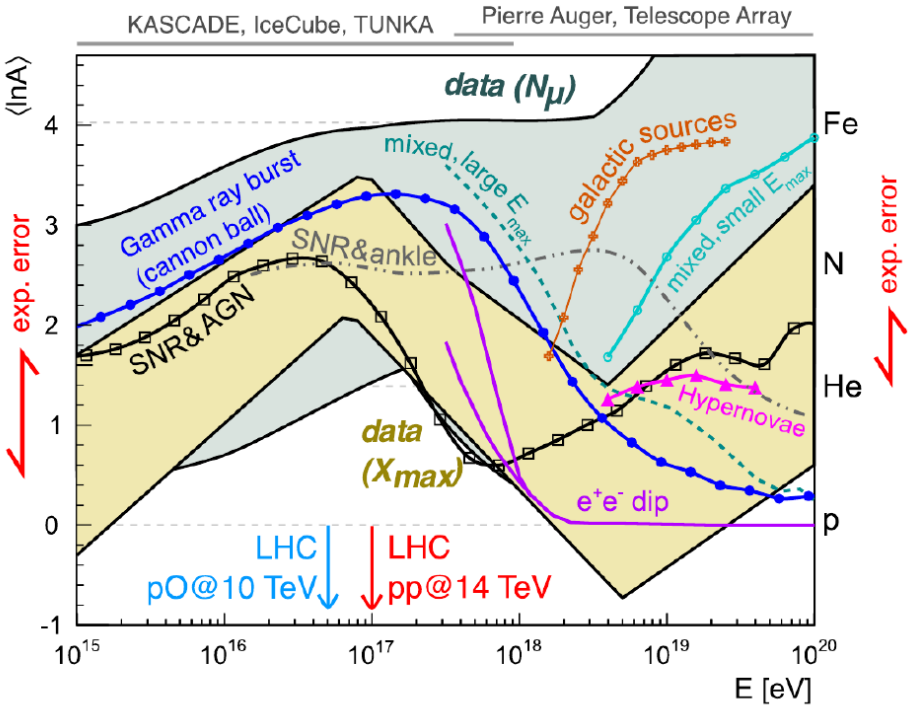


Figure 2.4: $\langle \ln A \rangle$ as a function of cosmic-ray energy E, used to infer the mass composition of cosmic rays. Vertical arrows at the plot edges indicate the instrumental uncertainties of the leading experiments at low and high energies. Figure adapted from [21, 28].

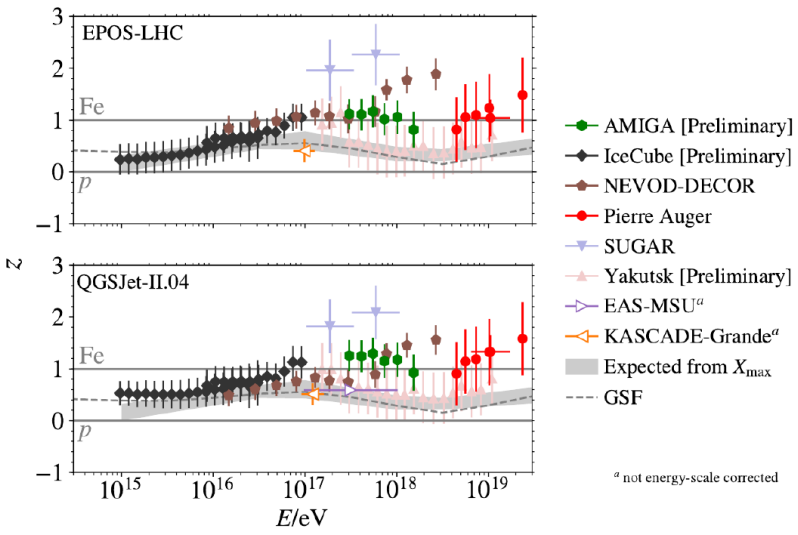


Figure 2.5: z-scale values for different experiments after applying energy-scale cross-calibration, shown for EPOS-LHC (top panel) and QGSJetII-04 (bottom panel). Figure extracted from [28].

Chapter III

The Pierre Auger Observatory

The Pierre Auger Observatory, located in the vicinity of Malargüe, Argentina, is the largest scientific facility ever built for the detection and study of ultra-high-energy cosmic rays (UHECRs). Covering an area of approximately 3,000 km^2, it was designed to collect sufficient statistics on the extremely low flux of the most energetic cosmic rays reaching Earth. Since the start of operations in 2004, the Observatory has continuously recorded data for more than two decades, providing an unprecedented dataset for the study of UHECRs and their associated air showers.

A key feature of the Pierre Auger Observatory is its *hybrid design*, which combines two complementary detection techniques: the Surface Detector (SD) and the Fluorescence Detector (FD). This concept allows simultaneous observation of the same extensive air showers (EAS) from the ground and from the atmosphere, greatly improving the accuracy of event reconstruction and the control of systematic uncertainties.

The SD consists of a vast array of 1,660 water-Cherenkov detectors (WCDs) arranged on a triangular grid with 1.5 km spacing, covering the entire 3,000 km^2 area. Each WCD is a cylindrical tank of 12 m^3 of purified water equipped with three photomultiplier tubes (PMTs) that record the Cherenkov light emitted by charged particles traversing the water volume. The SD operates nearly 100% of the time, continuously sampling the footprint of air showers at ground level and providing the exposure necessary to accumulate the large statistics required for studies at the highest energies.

The FD complements the SD by observing the longitudinal development of the same air showers through the detection of fluorescence light emitted by atmospheric nitrogen molecules excited by the passage of charged particles. It consists of 24+3 telescopes distributed across four sites located at the periphery of the SD array. Each telescope covers a field of view of 30° in azimuth and 28.6° in elevation, collectively providing almost complete coverage of the atmosphere above the array. The FD operates on clear, moonless nights, corresponding to a duty cycle of about 15%, but provides a near-calorimetric measurement of the shower energy. The hybrid reconstruction, which combines FD and SD information for the same events, allows for a precise calibration of the SD energy scale and improved determination of the

arrival direction and composition-sensitive observables.

A schematic layout of the Pierre Auger Observatory is presented in Figure 3.1. The black dots indicate the positions of the SD stations, while the blue lines mark the fields of view of the four fluorescence sites. The denser 750 m infill array, located near the Coihueco FD site, is also shown. This smaller array extends the measurement capabilities of the Observatory to lower energies, down to approximately 10^{17} eV, bridging the gap between direct cosmic-ray measurements and the main Auger energy range. At each station of this infill array, 30 m^2 underground muon detectors are being installed to measure the muonic component of air showers directly.

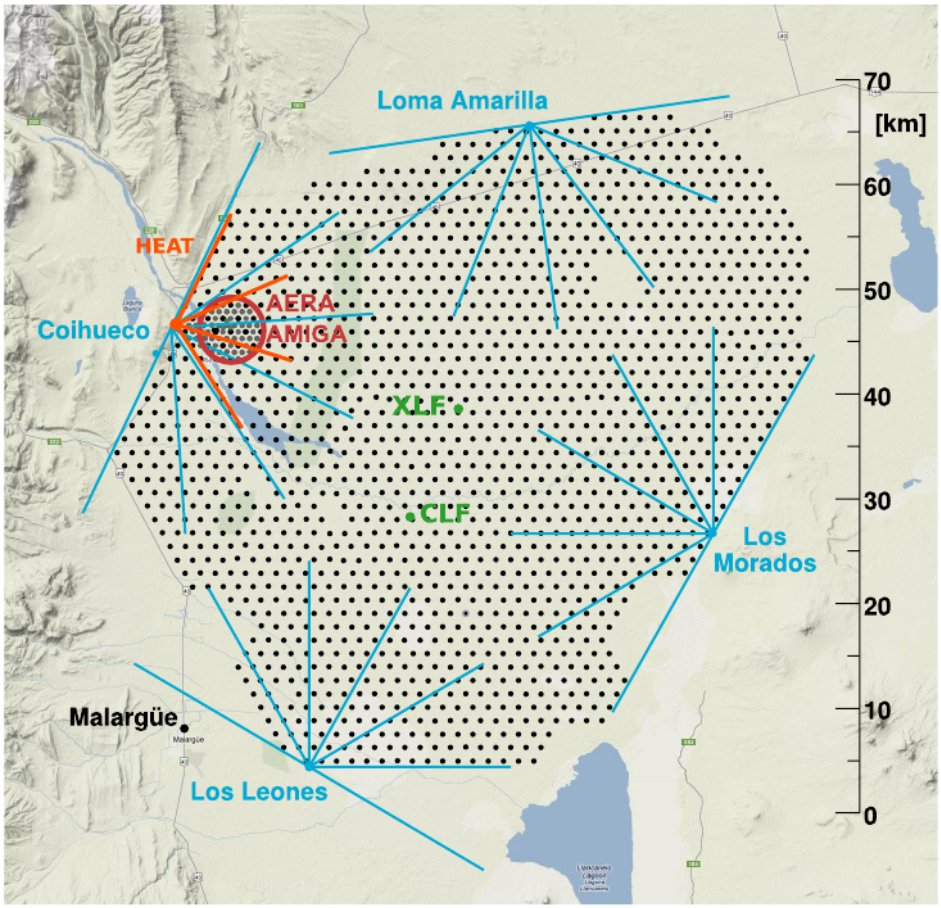


Figure 3.1: Schematic layout of the Pierre Auger Observatory, including a kilometric scale. Each black dot represents a Surface Detector (SD) station. The blue lines indicate the fields of view of the four Fluorescence Detector (FD) sites. The denser array with 750 m spacing between SD stations is also highlighted. At each of these stations, 30 m^2 Underground Muon Detectors (UMDs) are being installed to directly measure the muonic component of extensive air showers.

The Observatory has achieved major breakthroughs in the field of astroparticle physics, including the discovery of a large-scale dipole anisotropy in the arrival directions of UHECRs, high-precision measurements of the energy spectrum and mass composition, and stringent limits on photon and neutrino fluxes at the highest

energies. The success of these measurements stems largely from the hybrid approach and from the long-term stability and reliability of the detectors, which have operated under the harsh environmental conditions of the Pampa Amarilla with minimal downtime.

To further enhance its scientific capabilities, the Pierre Auger Observatory has recently been through a major upgrade known as *AugerPrime* [2]. The upgrade is designed to improve the determination of the mass composition of cosmic rays on an event-by-event basis, a key factor for understanding the origin and propagation of UHECRs. A central goal of AugerPrime is to enhance the discrimination between the electromagnetic and muonic components of air showers, thereby increasing sensitivity to the primary mass and improving the interpretation of observables such as the number of muons at ground and the depth of shower maximum.

The AugerPrime upgrade involves several major components. The most significant addition is the *Underground Muon Detector* (UMD), a network of buried scintillator modules designed to directly measure the muonic component of air showers. Each UMD station consists of an array of plastic scintillator strips covering an area of approximately $30\,\mathrm{m}^2$, installed at a depth of 2.3 m below the surface to shield it from the electromagnetic component of the shower. The UMD provides an unambiguous measurement of muon density at ground, which is crucial for disentangling the hadronic and electromagnetic components and thereby constraining the primary composition.

In addition to the UMD, AugerPrime includes several complementary upgrades. Each WCD station has been equipped with a thin scintillator detector on top (Surface Scintillator Detector, SSD), which enhances the separation between electromagnetic and muonic signals at the surface. New, faster electronics have been installed to improve time resolution and dynamic range, and an extended duty cycle for the fluorescence telescopes has been achieved through the implementation of the *FRAM* (Fluorescence Robotic Atmospheric Monitor) and the *FSA* (Fluorescence Subarray Auger) systems. Together, these improvements enable more precise hybrid reconstructions and extend the accessible energy range with better control of systematics.

The combination of WCD, SSD, FD, UMD enables the measurement of extensive air showers using multiple, complementary observables. This detector configuration provides sensitivity to both the electromagnetic and muonic components of the shower, allowing their partial separation and improving the reconstruction of primary mass composition over a broad energy range. Within the AugerPrime upgrade, these capabilities extend the potential of the Pierre Auger Observatory for detailed studies of ultra-high-energy cosmic rays, including investigations of their mass composition, energy spectrum, and the hadronic interactions governing shower development in the atmosphere.

3.1. The Surface Detector

The Surface Detector (SD) of the Pierre Auger Observatory consists of an array of 1660 water-Cherenkov detectors (WCDs) arranged in three nested triangular grids with spacings of 1500 m (SD-1500), 750 m (SD-750), and 433 m (SD-433). The SD plays

a central role in reconstructing extensive air showers (EAS) by recording the arrival times and lateral distribution of secondary particles at ground level. Among these particles, muons are particularly valuable: owing to their larger mass, they traverse the full water volume of the tank, producing Cherenkov light along their tracks. In contrast, electrons and other low-mass particles are more strongly attenuated and typically deposit their energy near the surface.

Each WCD is a light-tight cylindrical tank filled with $12\,\mathrm{m}^3$ of purified water. The inner surface is lined with reflective Tyvek material to ensure uniform light diffusion, while the outer black liner prevents contamination from ambient light. Three downward-facing photomultiplier tubes (PMTs) are mounted on the tank roof to collect Cherenkov photons, converting them into electrical signals that are digitized and analyzed to extract timing and amplitude information. A photograph and schematic representation of a WCD are shown in Figure 3.2.

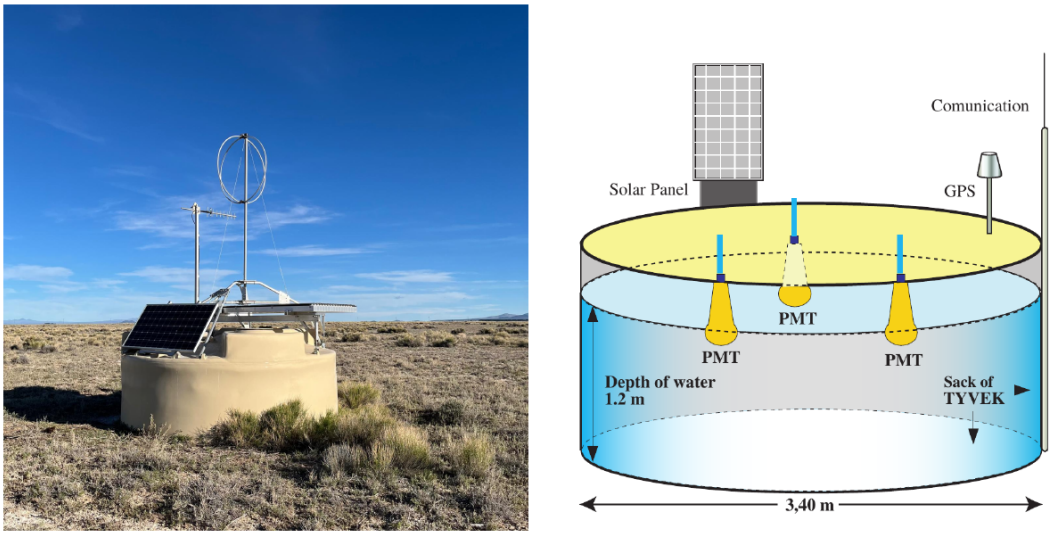


Figure 3.2: (Left) Photograph of a deployed Water-Cherenkov Detector (WCD), also showing the Scintillator Surface Detector (SSD) and Radio Detector (RD) mounted on top. (Right) Schematic cross-section of the WCD interior. Figure adapted from [31].

Each local station performs online calibration to adjust PMT gains to nominal levels and set trigger thresholds in physical units. These routines, implemented in the station's data acquisition (DAQ) software, maintain uniform PMT responses relative to a *Vertical Equivalent Muon* (VEM), defined as the most probable signal produced by a vertical muon crossing the central axis of the tank.

The station electronics, known as the Unified Board (UB), integrates the analog front-end, data acquisition, and trigger systems on a single board. It processes signals from the three large PMTs, which operate with a positive anode voltage due to their proximity to water. The high voltage is regulated by a module within the PMT base, and the signals are amplified and inverted, with the last dynode providing an output 32 times stronger than the anode. These signals are filtered and digitized by a 10-bit, 40 MHz semi-flash ADC.

The recorded traces are transmitted in raw ADC counts ("adc" units) to the

Central Data Acquisition System (CDAS) and require calibration before analysis [32]. After calibration, the traces are expressed in VEM units. The Offline framework allows access to simulated shower traces separated by particle component, such as the example muon trace shown in Figure 3.3 for a single PMT.

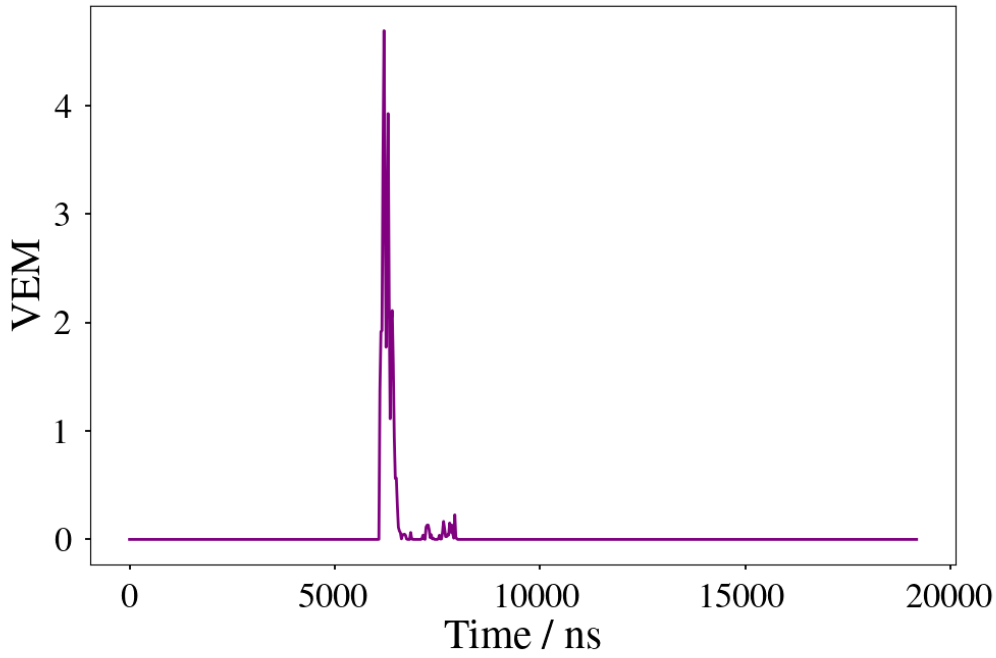


Figure 3.3: Example of a single muon trace in VEM units at 450 m from the shower core.

To obtain the total muon signal at each station, the trace from each PMT is integrated, multiplied by the reference peak value I_{VEM} (corresponding to vertical, through-going muons), and divided by Q_{VEM} (the charge of a Vertical Equivalent Muon). This yields the muonic signal in units of $\mathrm{VEM}_{\mathrm{charge}}$, commonly referred to simply as VEM.

For the Surface Detector, the lateral distribution of signals at ground level is described by a modified Nishimura–Kamata–Greisen (NKG) function:

$$S(r) = S(r_{\mathrm{opt}}) \left(\frac{r}{r_{\mathrm{opt}}} \right)^{\beta} \left(\frac{r + r_{\mathrm{s}}}{r_{\mathrm{opt}} + r_{\mathrm{s}}} \right)^{\beta+\gamma}, \tag{3.1}$$

where $r_{\mathrm{s}} = 700\,\mathrm{m}$ is fixed, $S(r_{\mathrm{opt}})$ is a free parameter, and the shape parameters β and γ are set to data-driven, parameterized values unless a sufficient number of triggered stations allow a free fit. The shower core position $\vec{x}_c$, initially estimated as the signal barycenter, is also fitted as a free parameter. The optimal distance r_{opt} minimizes the uncertainty arising from the unknown true LDF shape and depends primarily on detector spacing and array geometry [33].

The measured shower size $S(r_{\text{opt}})$ depends on the zenith angle of the incoming shower, since inclined events traverse a larger atmospheric depth and consequently produce smaller signals in the WCDs due to attenuation. To account for this effect, a zenith-dependent attenuation correction $f_{\text{att}}(\theta)$ is derived using the Constant Intensity Cut (CIC) method [34, 35]. This correction scales the measured signal to a reference zenith angle θ_{ref}, yielding

$$S_{\theta_{\text{ref}}} = \frac{S(r_{\text{opt}})}{f_{\text{att}}(\theta)}. \tag{3.2}$$

The reference angle θ_{ref} is defined as the median of the zenith distribution of recorded events: 38° for the SD-1500 array, which covers $0° \leq \theta \leq 60°$, and 35° for the SD-750 array, covering $0° \leq \theta \leq 55°$. The corresponding attenuation-corrected shower sizes are denoted S_{38} and S_{35}, respectively.

The attenuation-corrected shower size serves as an energy estimator for the primary cosmic ray. The absolute energy calibration is obtained by correlating SD and Fluorescence Detector (FD) measurements using hybrid events—air showers simultaneously reconstructed by both detectors. The FD provides a quasi-calorimetric measurement of the shower energy (see Section 3.2), allowing the SD signal to be calibrated against the FD energy. The resulting calibration relation is expressed as

$$E_{\text{FD}} = A \left(\frac{S_{38}}{\text{VEM}} \right)^{B}, \tag{3.3}$$

where A and B are calibration constants determined through a log-likelihood minimization procedure [36]. This relation enables the conversion of SD observables into an absolute energy scale anchored to the FD, ensuring consistency and reliability across the full energy range accessible to the Observatory.

3.2. The Fluorescence Detector

The Fluorescence Detector (FD) of the Pierre Auger Observatory consists of 24 telescopes distributed across four sites: Los Leones, Los Morados, Loma Amarilla, and Coihueco, overlooking the Surface Detector (SD) array (see Figure 3.1). The FD operates only during dark, moonless nights with favorable atmospheric conditions, resulting in a duty cycle of approximately 15%. Under these conditions, it measures the longitudinal development of extensive air showers (EAS) with energies above 3×10^{18} eV by detecting the faint fluorescence light emitted by atmospheric nitrogen molecules excited by the passage of charged shower particles.

Each FD telescope comprises three main components: a circular entrance aperture, a spherical mirror, and a camera located at the mirror's focal surface. The camera contains 440 photomultiplier tubes (PMTs), referred to as pixels, arranged in a matrix of 22 rows by 20 columns. An ultraviolet (UV) filter placed at the aperture suppresses background light and enhances the signal-to-noise ratio. A schematic view of an FD telescope, along with a photograph of its camera, is shown in Figure 3.4.

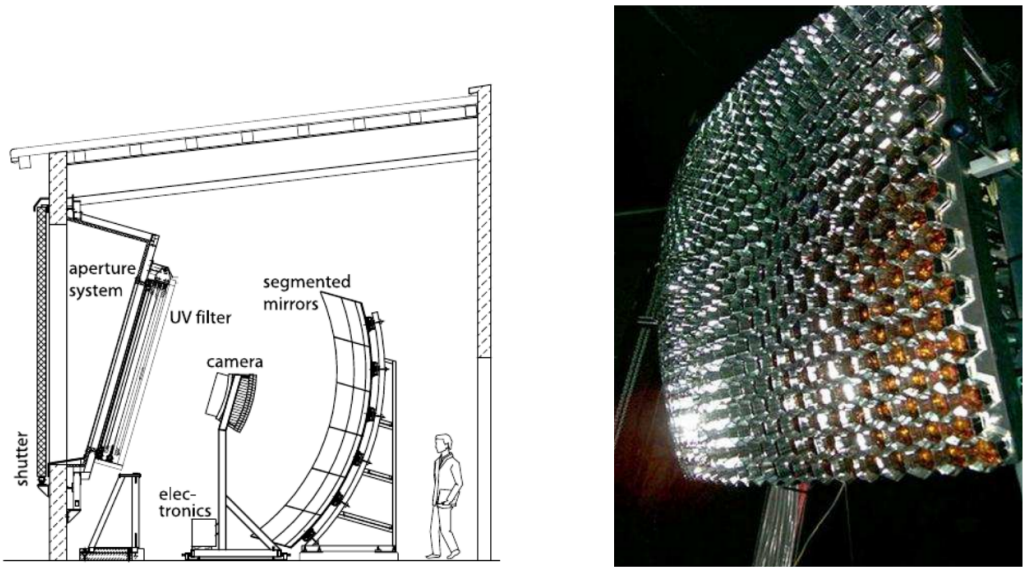


Figure 3.4: *Left*: Schematic view of an FD telescope and its main components. *Right*: Photograph of an FD camera showing its 440 PMTs. Figure adapted from [37].

As an air shower propagates through the atmosphere, its charged particles ionize the surrounding air molecules, losing energy in the process. When these ionized nitrogen molecules return to their ground state, a fraction of the released energy is emitted as fluorescence light, which is isotropic in nature. The fluorescence yield has been precisely measured under controlled laboratory conditions [38].

The fluorescence photons produced along the shower axis enter the telescope through the aperture, are reflected by the mirror, and focused onto the PMT camera. This produces a time-dependent signal in the activated pixels, as illustrated in Figure 3.5. From the timing information of these PMT signals, the geometry of the air shower can be reconstructed. When available, timing and spatial information from the SD stations are incorporated to further constrain and improve the reconstruction accuracy.

The measured pixel signals are projected onto the reconstructed shower axis, allowing the determination of the energy deposition profile as a function of slant depth, dE/dX. This longitudinal profile is fitted using the Gaisser–Hillas function:

$$f_{\mathrm{GH}}(X) = \left(\frac{dE}{dX}\right)_{\mathrm{max}} \left(\frac{X - X_0}{X_{\mathrm{max}} - X_0}\right)^{(X_{\mathrm{max}} - X_0)/\Lambda} e^{(X_{\mathrm{max}} - X)/\Lambda}, \tag{3.4}$$

where $\left(\frac{dE}{dX}\right)_{\mathrm{max}}$ denotes the maximum energy deposition, X_0 and λ are shape parameters, and X_{max} represents the depth at which the shower reaches its maximum development.

In addition, three high-elevation fluorescence telescopes, collectively referred to as the High-Elevation Auger Telescopes (HEAT), operate at the Coihueco site [39]. By extending the field of view to elevation angles between 30° and 58°, HEAT enables the

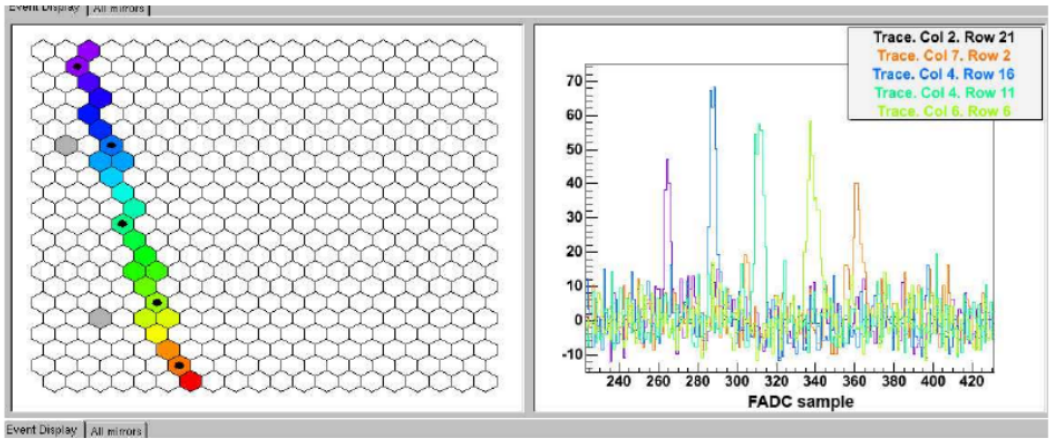


Figure 3.5: Example of an air-shower event as seen by the camera of an FD telescope [37]. *Left*: Pattern of triggered PMTs with color-coded timing. *Right*: Time-dependent signals recorded by the corresponding pixels.

observation of air showers that develop at higher altitudes and correspond to lower primary energies. In combination with the SD-750 array, this configuration extends the range of high-quality hybrid measurements down to energies of approximately $10^{17.2}$ eV.

3.3. The AugerPrime extension

As discussed in Chapter 2, achieving sensitivity to the primary mass on an event-by-event basis is essential for disentangling astrophysical scenarios of ultra-high-energy cosmic-ray sources and acceleration mechanisms. To enhance this sensitivity, the Pierre Auger Observatory has undergone a major upgrade known as *AugerPrime*. This upgrade encompasses several improvements: the installation of scintillator and radio detectors atop the existing WCDs, the replacement of the WCD electronics, the addition of a small photomultiplier tube (SPMT) to each WCD, and the deployment of the Underground Muon Detector (UMD).

The scientific goals of AugerPrime include a comprehensive investigation of the cosmic-ray mass composition and the hadronic interactions governing extensive air shower development, two aspects that are deeply interconnected. Central to these objectives is the precise estimation of the muonic component of air showers.

3.3.1. Scintillator Surface Detector

As part of the AugerPrime upgrade, each SD station was equipped with a Scintillator Surface Detector (SSD), mounted directly above the existing WCD. The SSD consists of an aluminum frame supporting a rectangular enclosure of 3.8 m × 1.3 m, containing two plastic scintillator panels with a combined active area of 3.8 m^2 (1.9 m^2 each). The SSD enhances the capability to discriminate between the electromagnetic and muonic components of air showers, thereby improving the determination of the

primary mass composition (see Figure 3.2).

3.3.2. Small Photomultiplier and Upgraded Electronics

The WCDs have been upgraded to extend their dynamic range and improve overall performance. A key improvement is the inclusion of a small photomultiplier tube (SPMT), the Hamamatsu R8619-22, with a diameter of less than 30 mm, significantly smaller than the ~23 cm of the standard PMTs. This SPMT enables the accurate measurement of large signals that would otherwise saturate the main PMTs, thus enhancing the detector's ability to record events near the shower core [2, 40].

The electronics of the WCDs have also been modernized. The original unified board (UB), featuring six channels, a 40 MHz sampling frequency, and a 10-bit ADC, has been replaced by the upgraded unified board (UUB). The new UUB provides 10 channels, a sampling rate of 120 MHz, and a 12-bit ADC, accommodating the additional inputs from the SPMT and SSD, and enabling communication with the Radio and Underground Muon Detectors. The operation of the Observatory with the original UB system is referred to as *Phase 1*, while the current operation with the UUB corresponds to *Phase 2*.

3.3.3. Radio Detector

Charged particles in extensive air showers emit coherent radio signals as they are accelerated in the geomagnetic field, primarily within the 30–80 MHz frequency range. A WCD equipped with both an SSD and a radio antenna is shown in Figure 3.2.

The radio signal complements the particle information obtained by the surface detectors. By combining the muon content measured with the WCD and the electromagnetic component reconstructed from the radio detector (RD), the energy dependence of the muon number in inclined showers can be studied up to the highest energies. This methodology extends the earlier WCD–FD hybrid approach, offering roughly an order of magnitude increase in event statistics.

While the WCD–SSD system provides an e/μ ratio measurement for vertical showers, the WCD–RD combination extends this sensitivity to more inclined events, effectively broadening the Observatory's aperture for mass-sensitive analyses.

3.3.4. Underground Muon Detector

The muonic component of EAS can be measured with higher precision using shielded detectors that suppress the electromagnetic background. For this purpose, the Underground Muon Detector (UMD) was implemented as part of the AugerPrime upgrade.

A dedicated subarray covering 27 km^2 was deployed to extend the detection level to lower energies. It comprises 49 additional SD stations arranged on a denser 750 m grid, forming the SD-750 array. This configuration achieves full detection efficiency for zenith angles below 55° and primary energies above 3×10^{17} eV.

Within a 23.5 km^2 region of the SD-750 array, underground scintillation detectors were installed to directly measure the muonic component of air showers. These

detectors, collectively referred to as the UMD, are present in 61 of the 71 SD-750 positions. Each UMD station comprises 30 m^2 of scintillator area buried to a depth of 2.3 m. The UMD operates in slave mode with its corresponding WCD, using the WCD trigger for data acquisition. The soil overburden acts as an electromagnetic shield, corresponding to a muon energy threshold of approximately 1 GeV for vertical incidence. Each station consists of three independent 10 m^2 modules, which together form the total active area of 30 m^2. A photograph of a station during installation is shown in Figure 3.6.

Figure 3.6: Installation of a UMD station with scintillation-based muon counters. Image from [41].

The final detector design, illustrated in Figure 3.7, places the three modules adjacent to the corresponding SD station and has been implemented in the SD-750 and SD-433 subarrays.

Each UMD module contains 64 plastic scintillator strips arranged in two panels of 32 strips. Wavelength-shifting (WLS) optical fibers run along the strips and are coupled to an array of 64 silicon photomultipliers (SiPMs). The SiPMs are solid-state photodetectors composed of avalanche photodiode microcells operating in Geiger mode on a common silicon substrate. During the engineering phase, multi-anode photomultiplier tubes (PMTs) were used instead of SiPMs.

The scintillator bars are manufactured at Fermilab [42]. Their operation relies on fluorescence: incident muons excite the scintillation material, whose electrons de-excite by emitting photons. Specifically, the polystyrene base (Dow Styron 663W) absorbs the incoming radiation and emits ultraviolet photons. These are absorbed by the primary dopant (2,5-diphenyloxazole, PPO), which re-emits photons at longer

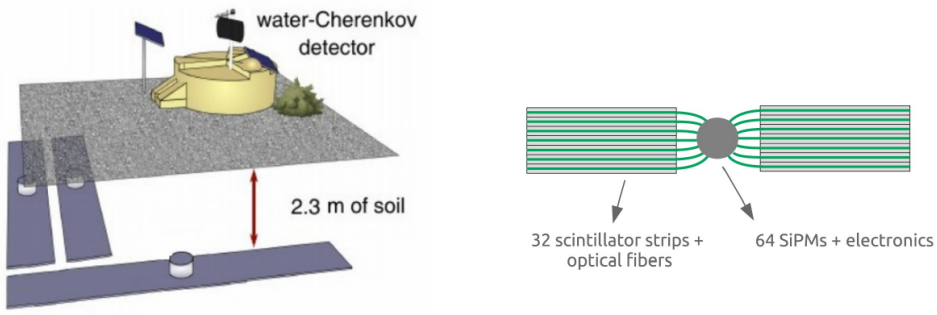


Figure 3.7: *Left*: Final design of a UMD station with three 10 m^2 modules. *Right*: Schematic of a single UMD module.

wavelengths, and subsequently by the secondary dopant (1,4-bis(5-phenyloxazol-2-yl)benzene, POPOP), producing blue photons with a wavelength of about 420 nm. The mean attenuation length of these photons in the scintillator ranges from 5 cm to 25 cm, much shorter than the 4 m length of the strips. Therefore, WLS fibers are used to efficiently collect and guide the light to the photodetectors.

The Saint-Gobain BCF-99-29AMC multi-clad WLS fiber features a scintillating core that absorbs the blue fluorescence photons and re-emits green photons, about 10% of which remain trapped by total internal reflection due to the two cladding layers with lower refractive indices. These photons propagate along the fiber and reach the SiPMs, where they are converted into electrical signals [43]. The basic scheme of a scintillator strip and its optical fiber is shown in Figure 3.8.

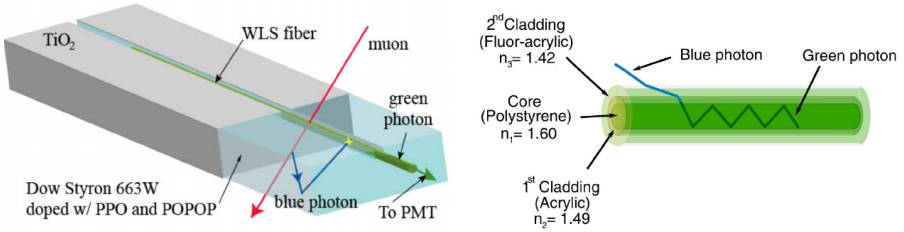


Figure 3.8: Left: Schematic of a scintillator strip. Right: Wavelength-shifting optical fiber. Figures adapted from [44, 43].

The UMD operates in two complementary readout modes: **binary** and **ADC**.

In *binary mode*, each of the 64 SiPM channels is processed independently through a preamplifier, fast shaper, and discriminator. The digital output is stored in a circular buffer of 2048 time bins, each 3.125 ns wide. Laboratory tests show that a single-muon pulse exceeding a 2.5-photoelectron threshold lasts about 12.5 ns. These studies [43] lead to the currently used counting strategies: muons are identified by searching for sequences of at least four consecutive "1" bits, distinguishing them from random noise. An example of a binary-mode trace is displayed in Figure 3.9.

In *ADC mode*, the total number of muons, N_μ^{ADC}, is estimated from the integrated charge as

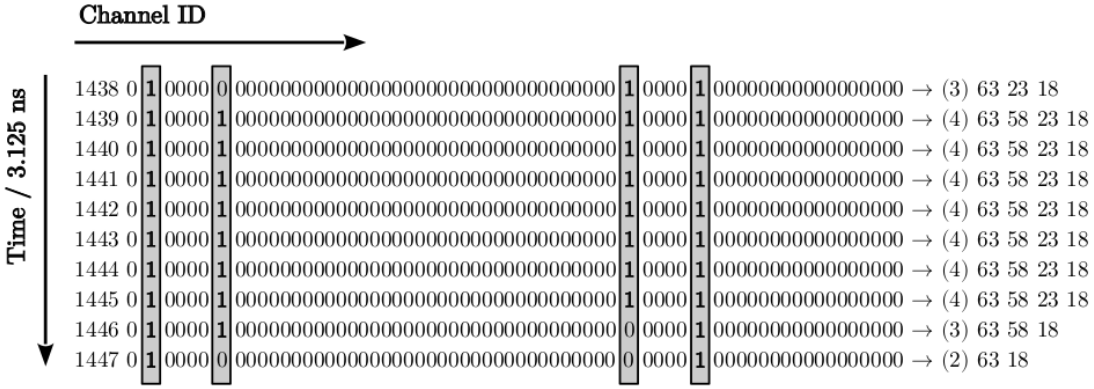


Figure 3.9: Example of a binary-mode output signal from a UMD station. Each column corresponds to one of the 64 channels, and each row represents a time bin. Figure adapted from [43].

$$N_{\mu}^{\mathrm{ADC}} = \frac{q_{\mathrm{meas}} \cos\theta}{\langle q_{1\mu}(\theta = 0)\rangle}, \tag{3.5}$$

where q_{meas} is the total measured charge per unit vertical path length, and $\langle q_{1\mu}(\theta = 0)\rangle$ is the mean charge deposited by a vertical muon.

The two readout modes provide complementary information on the muonic component of air showers and together enhance the mass-composition sensitivity of the upgraded array.

During the engineering phase, UMD modules employed multi-anode PMTs as photodetectors [4]. In the final design, these were replaced by SiPMs, as described above. Deployment of the UMD is proceeding progressively: the SD-433 subarray and roughly half of the SD-750 array are currently fully operational. The analyses presented in this thesis refer to the final SiPM-based configuration and are compared, in Section 8.7.3, with results from the engineering array.

3.4. $\overline{\text{Off}}\underline{\text{line}}$: The simulation and reconstruction environment of the Pierre Auger Collaboration

The $\overline{\text{Off}}\underline{\text{line}}$ framework is the official simulation and reconstruction environment of the Pierre Auger Collaboration. It provides a unified platform to simulate the detector response to EAS and to reconstruct real events recorded by the Observatory. The framework handles the complete analysis chain, from the particle cascades generated with external shower simulators such as CORSIKA to the final reconstruction of physical observables from experimental data.

$\overline{\text{Off}}\underline{\text{line}}$ consists of three main components: a modular processing pipeline, an event data model, and a detector description system [45, 46]. The modular pipeline allows users to configure and execute processing tasks through XML files, while the event data model accumulates all relevant information from both simulation

and reconstruction stages in a consistent structure. The detector description system provides access to the geometry and calibration of the detectors, as well as to environmental data such as atmospheric profiles, aerosol content, and long-term monitoring of detector conditions. The general structure of the framework is illustrated in Figure 3.10.

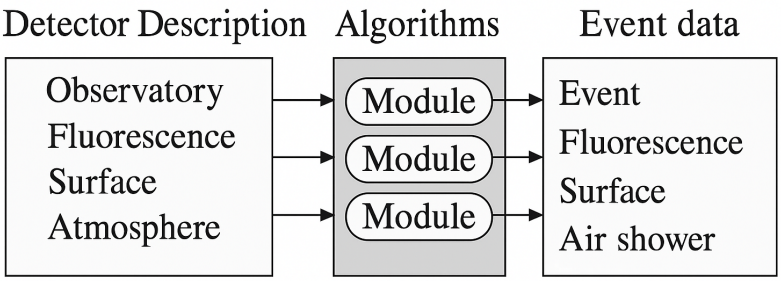


Figure 3.10: General structure of the $\overline{\text{Off}}\underline{\text{line}}$ framework. Simulation and reconstruction tasks are modularized. Figure extracted from [45].

This modular architecture ensures a clear separation between data and algorithms, allowing the same reconstruction procedures to be applied to simulated and experimental events. It also provides foundation classes for mathematical operations, error handling, and physics-related calculations, as well as a geometry package that manages coordinate transformations between different detector systems. By maintaining a consistent interface between simulated and measured data, $\overline{\text{Off}}\underline{\text{line}}$ enables direct comparison and validation of algorithms and detector configurations under identical conditions.

A key feature of $\overline{\text{Off}}\underline{\text{line}}$ is its integration with the Monte Carlo program CORSIKA (COsmic Ray SImulations for KAscade) [47, 48], which is used to simulate the development of EAS in the atmosphere. CORSIKA tracks the evolution of primary cosmic rays and the resulting particle cascades, describing electromagnetic and hadronic interactions across a wide range of energies and particle types. Within the Pierre Auger simulation chain, CORSIKA provides the list of secondary particles reaching the ground, which are then processed by $\overline{\text{Off}}\underline{\text{line}}$ to simulate their interactions with the detectors. In this way, the combination of both tools allows a realistic modeling of the full detection process, from the atmospheric development of the shower to the detector signals recorded at ground level.

In the case of the UMD, $\overline{\text{Off}}\underline{\text{line}}$ simulates the propagation of muons through the soil, the generation of scintillation light within the plastic scintillators, and the subsequent response of the silicon photomultipliers (SiPMs). Muons must have an

energy above approximately 1 GeV to reach the UMD depth, meaning that the low-energy portion of the muon spectrum is absorbed in the overburden. In contrast, the water-Cherenkov detectors of the Surface Detector are sensitive to muons with energies above ~240 MeV, which is the energy deposited by a Vertical Equivalent Muon (VEM), the signal unit for the WCD [32]. Accurately modeling the propagation and energy loss of muons through the soil is therefore essential to establish the connection between the on-ground and underground muon densities.

In this work, the injected muon densities in both the WCDs and the UMD were used to describe the muon content at the surface and at the UMD depth, respectively. The resulting simulations provide the basis for the calibration studies developed in the following chapters, linking the underground and surface muon observables within a consistent physical framework.

Chapter IV

Simulation of the Injected Muons in the UMD

In this chapter, a comparison is made between the Offline (see Section 3.4) implementation of muon propagation and injection and an independent method that estimates the density of muons reaching a depth of 2.3 m in soil. The alternative approach is based on the Continuous Slowing Down Approximation (CSDA), which describes the mean energy losses of charged particles as they traverse matter [49]. CSDA relies on stopping-power values that characterize energy loss per unit mass overburden.

Previous studies [50] indicate that the soil in the UMD region is predominantly composed of SiO_2. As an approximation, the stopping-power tables for this compound are therefore used to compute the energy losses of muons directly from simulated air showers.

4.1. Continuous Slowing Down Approximation

When a charged particle traverses matter, it loses energy primarily through ionization and excitation of the medium. The mean energy loss per unit path length, known as the *stopping power*, is defined as

$$S(E) \equiv -\frac{dE}{dx}, \tag{4.1}$$

where E is the particle energy and x the distance traveled. For high-energy muons and other heavy charged particles, the stopping power conveniently describes the gradual decrease of energy along the trajectory.

The average energy loss of muons can be expressed as

$$\left\langle \frac{dE}{dX} \right\rangle = a(E) + b(E)\,E, \tag{4.2}$$

where $a(E)$ represents the **ionization stopping power** due to ionization and atomic excitation, while $b(E)$ accounts for **radiative processes** such as bremsstrahlung, electron–positron pair production, and photonuclear interactions [51]:

$$b(E) = b_{\text{brems}} + b_{\text{pair}} + b_{\text{nucl.}}. \tag{4.3}$$

The functions $a(E)$ and $b(E)$ vary only slowly with energy in the regime where radiative effects become relevant. As illustrated in Figure 4.1, for energies below approximately 100 GeV the ionization term $a(E)$ dominates, while the radiative term $b(E)E$ contributes less than about 1% to the total stopping power. This corresponds to the **Bethe–Bloch region**, where the energy loss decreases with increasing energy until a minimum ionization point between 1-3 GeV. At even higher energies (hundreds of GeV to TeV), the **radiative regime** is reached, where bremsstrahlung, pair production, and photonuclear interactions become increasingly important and cause the total stopping power to increase approximately linearly with energy.

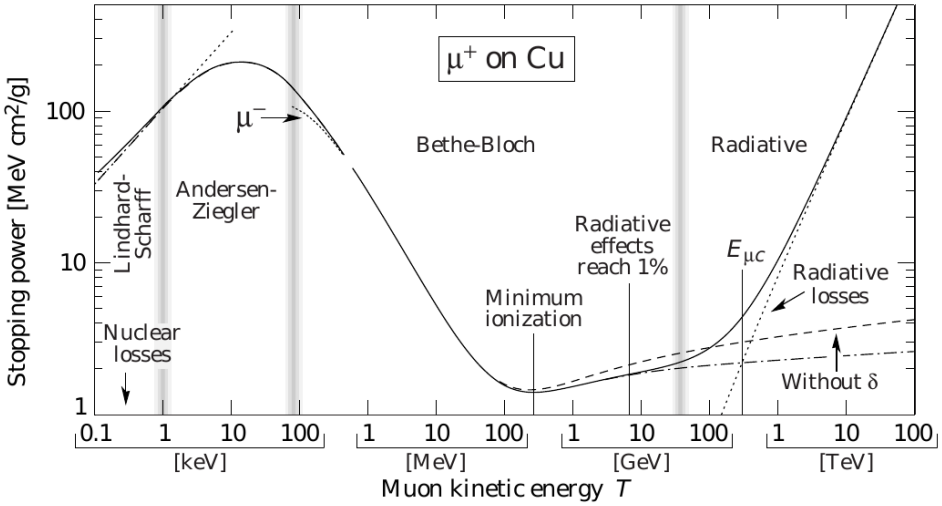


Figure 4.1: Mean stopping power, $\langle -dE/dX \rangle$, for positive muons in copper as a function of kinetic energy T. The solid curve indicates the total stopping power, with contributions from different processes labeled. The transition from the Bethe–Bloch region to the radiative regime occurs near $T \simeq 100$ GeV. Figure extracted from [49].

The **Continuous Slowing Down Approximation (CSDA) range** is defined as the total path length that a charged particle travels in a medium before stopping, assuming a continuous mean energy loss:

$$R(E) = \int_{E_0}^{E} \frac{dE'}{a(E') + b(E')\,E'}, \tag{4.4}$$

where E_0 is a sufficiently small cutoff energy such that the integral result is insensitive to its exact value.

At high energies, where both $a(E)$ and $b(E)$ can be treated as approximately constant, the integral simplifies to

$$R(E) \approx \frac{1}{b} \ln\!\left(1 + \frac{E}{E_{\mu c}}\right), \tag{4.5}$$

where $E_{\mu c} = a/b$ is the **muon critical energy**, defined as the energy at which electronic and radiative losses are equal.

Comprehensive tables of stopping powers and CSDA ranges for numerous materials, including elements, compounds, mixtures, and biological tissues, are provided in Ref. [49] for energies from 10 MeV up to 100 TeV. These tables also list the individual contributions to $b(E)$. It should be noted, however, that they do not include effects such as multiple scattering, energy-loss fluctuations, or range straggling in the electronic component, the latter referring to the statistical spread in particle penetration depth arising from stochastic variations in ionization energy loss.

4.2. Simulation of the UMD Shielding in Offline

High-precision detector studies rely on robust software frameworks for large-scale simulations. As discussed in Section 3.4, Offline uses the Geant4 toolkit to simulate particle interactions and the response of the detector. For the UMD, Geant4 models both the shielding provided by the surrounding soil and the interaction of muons with the scintillator bars.

Geant4 includes a wide range of electromagnetic, hadronic, and optical processes that span energies from a few hundred eV to several TeV [52]. Its versatility has made it a standard tool in particle and nuclear physics, accelerator design, space engineering, and medical applications.

During the initial deployment of the UMD, a dedicated Geant4 study [53] determined the minimum muon energy required to reach different burial depths. As shown in Figure 4.2, at a depth of 2.3 m the threshold is about 1 GeV.

Although this shielding is implemented in Offline via Geant4, an alternative method based on stopping-power tables was developed to estimate the muon flux reaching the UMD. The aim is not to replace Geant4 but to provide a transparent, easy-to-reproduce cross-check outside the full Offline environment. This approach retains the essential energy-loss physics while offering a clearer interpretability of muon propagation through the soil.

4.3. Simulations Set-Up

CORSIKA (v77420) was used to simulate air showers initiated by protons and iron nuclei at 10^{18} eV with $\theta = 0°$. EPOS-LHC and FLUKA served as hadronic interaction models: EPOS-LHC is based on the Gribov-Regge theory and is tuned to collider data from LHC and RHIC [54], while FLUKA provides detailed low-energy nuclear modeling [55]. A total of 240 independent showers were simulated for each primary[1].

The detector response was simulated with Offline Merge 2c590c2, assuming 218 SD stations, as this is the number of stations that compose the SD-750 array, each paired with a 30 m^2 UMD, arranged on a 750 m hexagonal grid. For each event, an inner ring of 12 SD+UMD stations was placed at 450 m from the shower axis, equally

[1] The plots presented here are available for other energy and zenith angle in Appendix A

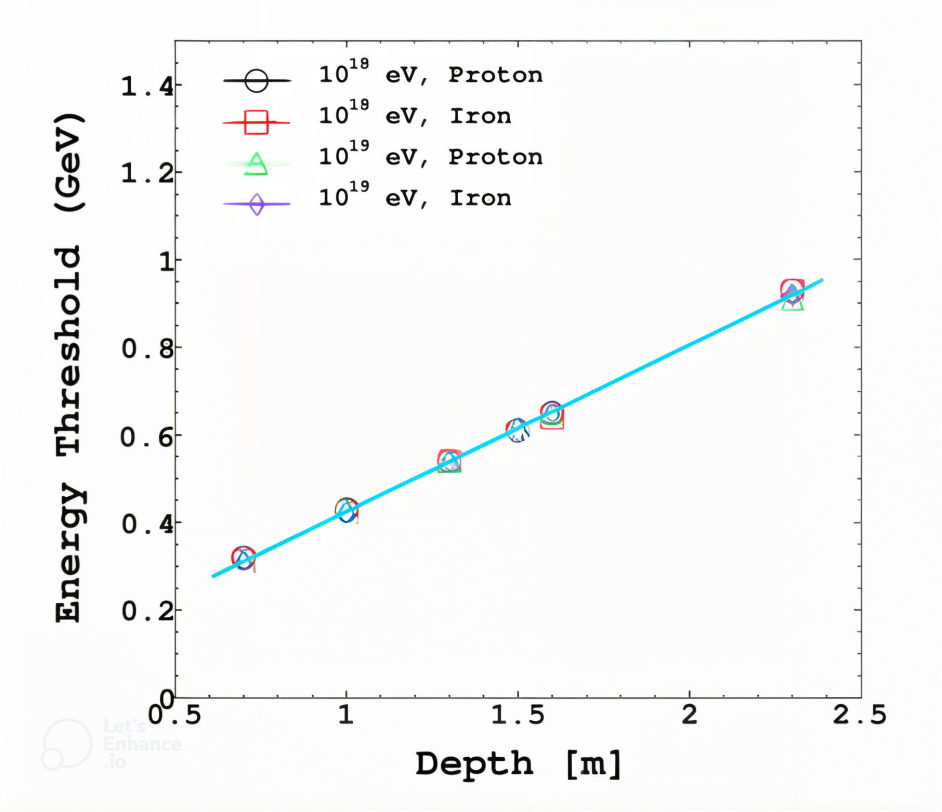


Figure 4.2: Muon energy threshold versus UMD burial depth [53].

spaced in azimuth. To improve statistics, each shower was resampled five times, yielding 1,200 events per primary.

4.4. On-Ground Muon Energy and Zenith Distributions

The energy distribution of muons at ground reflects both production and absorption during air-shower development. A key observable is the maximum muon production depth $X^{\mu}_{\max}$, the altitude at which the production rate peaks. Heavier nuclei interact earlier than lighter ones and thus have smaller (shallower) $X^{\mu}_{\max}$ values (Figure 4.3). Earlier development implies that a larger fraction of low-energy muons from iron showers are absorbed before reaching the ground, so the surviving iron muons are, on average, more energetic than those from proton showers.

Figure 4.4 shows the surface muon energy and zenith distributions for protons and iron at 10^{18} eV and $\theta = 0°$. Muons from Iron induced showers are on average about 20% more energetic: $\langle E^{\mathrm{p}}_{\mu} \rangle = (1.853 \pm 0.002)$ GeV and $\langle E^{\mathrm{Fe}}_{\mu} \rangle = (2.230 \pm 0.002)$ GeV.

In Figure 4.5 two examples of lower-energy and more inclined showers are shown. At lower primary energy ($\lg(E/\mathrm{eV}) = 17.5$), the mean-energy difference remains similar in relative terms ($\sim$ 19%) with $\langle E^{\mathrm{p}}_{\mu} \rangle = (1.942 \pm 0.002)$ GeV and $\langle E^{\mathrm{Fe}}_{\mu} \rangle = (2.394 \pm 0.002)$ GeV. At $\theta = 30°$ and 10^{18} eV the difference is $\sim$ 17% with $\langle E^{\mathrm{p}}_{\mu} \rangle = (2.366 \pm$

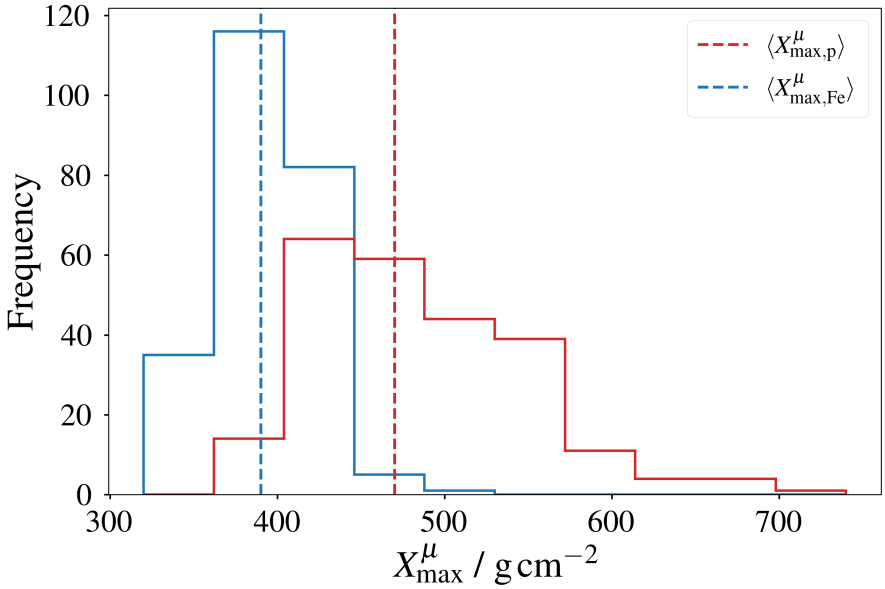


Figure 4.3: Maximum muon production depth $X^{\mu}_{\max}$ for different primaries.

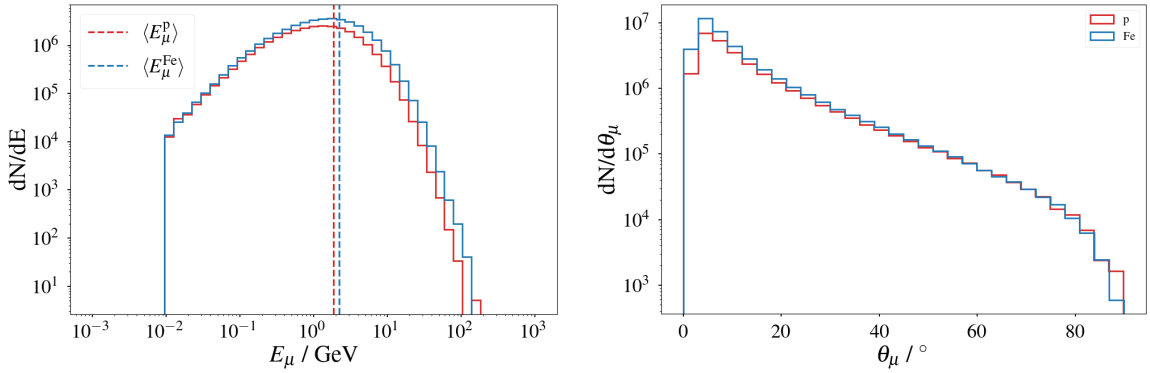


Figure 4.4: Surface muon energy (left) and zenith angle (right) distributions for proton and iron primaries at $\lg(E/\mathrm{eV}) = 18.0$, $\theta = 0°$ (left). It can be seen that the mean muon energy from Iron initiated showers is higher than the one fron proton initiated showers. It is also observed that muons tend to align with the shower axis, arriving with 0° in this case.

0.003) GeV and $\langle E^{\mathrm{Fe}}_{\mu} \rangle = (2.865 \pm 0.002)$ GeV. Thus, the *relative* offset between proton and iron primaries is nearly constant across energy and zenith [56], whereas the *absolute* mean muon energy increases for more inclined or lower-energy showers because only the most energetic muons survive propagation. This effect is further enhanced by the larger atmospheric slant depth traversed by inclined showers, which implies longer propagation paths through the dense lower atmosphere and leads to increased cumulative ionization energy losses, suppressing low-energy muons.

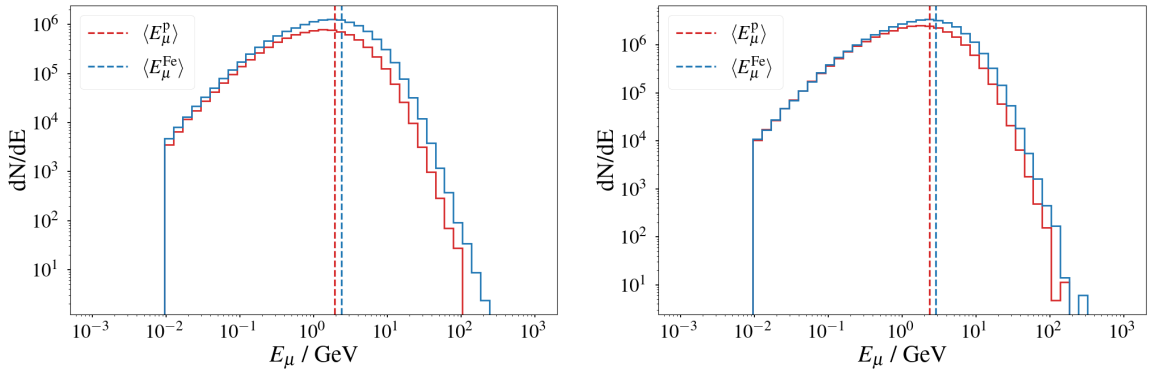


Figure 4.5: On-ground muon energies for proton and iron showers at $\lg(E/\text{eV}) = 17.5$, $\theta = 0°$ (left), and $\lg(E/\text{eV}) = 18.0$, $\theta = 30°$ (right).

4.5. Method to Estimate the Muon Density in the UMD Using Stopping Power Values

Once muons reach ground level, each surface muon with energy E_μ and zenith angle θ_μ is propagated through SiO_2 in steps of $\Delta x = 1\,\text{cm}$ along its track. At step i, the energy is updated as

$$E_{i+1} = E_i - \left(\frac{dE}{dX}\Big|_{E_i} \right) \rho_{\text{soil}}\, \Delta x, \tag{4.6}$$

where $\frac{dE}{dX}$ is the *mass stopping power* (e.g. MeV cm^2/g) interpolated from the table and $\rho_{\text{soil}} = 2.38\ \text{g/cm}^3$ [50]. The iteration continues until the muon stops or reaches a vertical depth of 2.3 m.

Figure 4.6 (left) illustrates two 1 GeV muons at zenith angles of 3° and 6°. The steeper trajectory produces a longer path length in soil, hence larger cumulative energy loss before reaching the detector. The right panel shows the stopping-power curve, where losses increase steeply below 1 GeV due to stronger ionization [49].

4.6. Muon Energy Loss and Underground Distributions

Figure 4.7 compares, for proton and iron primaries, the energy distribution of muons underground (colored) with their corresponding energies at the surface (gray). The gray histograms include only those muons that reach the UMD at a depth of 2.3 m, while lower-energy muons that stop in the soil before reaching this depth are excluded. The dashed line marks the approximate threshold of ~1 GeV required to traverse the overburden [53].

The zenith-angle distribution at UMD depth (Figure 4.8) most muons that arrive to the detector have a zenith angle $\theta \lesssim 60°$; at larger angles, the increased path length in soil and cumulative energy losses decreases the flux.

A small residual population is visible at very low underground energies, forming a tail in the spectra. These are muons that arrive with just enough energy to

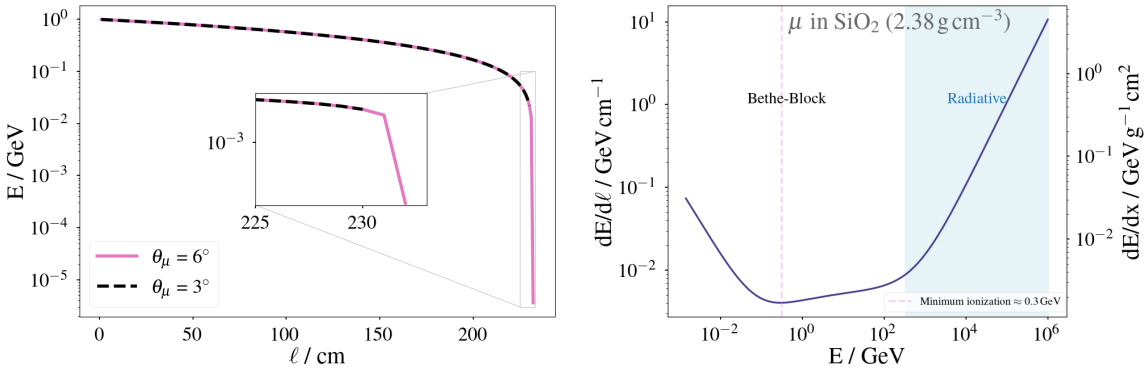


Figure 4.6: Left: energy loss for two 1 GeV muons at 3° and 6°; right: stopping-power curve with ionization and radiative regimes [49].

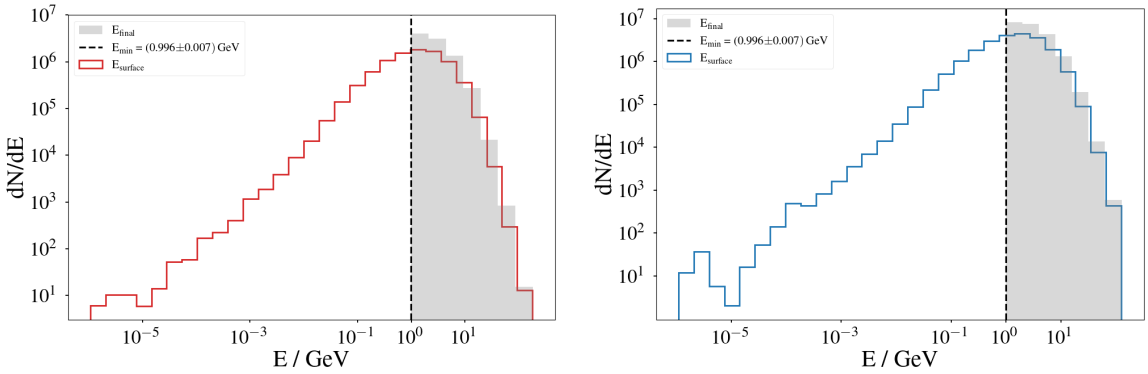


Figure 4.7: Underground muon energies (colored) versus surface energies (gray) for proton (left) and iron (right); dashed line marks ~1 GeV threshold.

reach the detector but effectively stop within or immediately before the scintillator, leaving negligible detectable signal. In particular, muons that reach the UMD depth with energies below ~10^{-5} GeV (~0.01 MeV) have essentially lost all kinetic energy, depositing at most 0.01 MeV in the scintillator. Such deposits would generate fewer than one photoelectron in the SiPM—far below the UMD trigger threshold (one-third of a MIP) [43]. Only muons depositing $\gtrsim$0.5 MeV—producing thousands of photons—reliably trigger the detector. This explains the apparent low-energy tail in the underground histograms without contributing to the detected muon count.

4.7. Comparison Between Stopping Power Method and Offline

Muon densities were compared at 450 m from the shower axis. For the stopping power method, the number of muons within (450±10) m was divided by the ring area projected on the shower plane to obtain the Stopping Power Density named ρ_{SP}. For Offline, the density in each 30 m^2 UMD station at the same distance was used

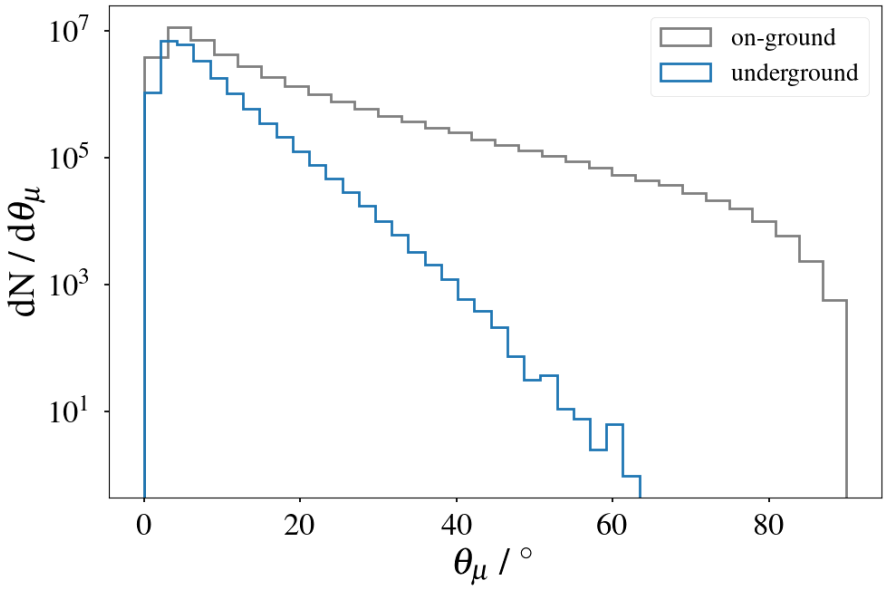


Figure 4.8: Zenith-angle distribution at UMD depth (blue) compared with surface muons (gray) for proton-initiated showers.

to obtain the Offline density named ρ_{off}. Figure 4.9 shows the relative differences between ρ_{SP} and ρ_{off}.

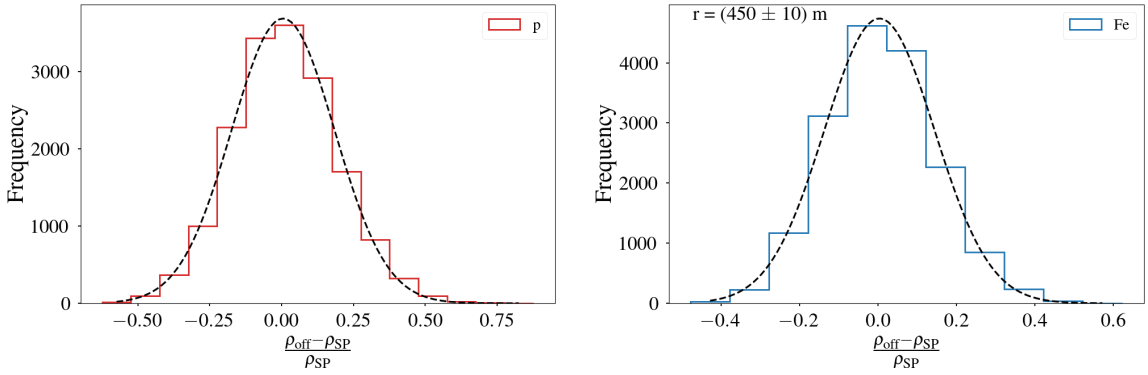


Figure 4.9: Relative difference between the muon density obtained using the Stopping Power Method (ρ_{SP}) and the one obtained using Offline (ρ_{off}) for proton (left) and iron (right) events.

The bias is (-0.004 ± 0.002) with $\sigma = (0.178 \pm 0.002)$ for protons and (0.005 ± 0.001) with $\sigma = (0.138 \pm 0.001)$ for iron. The dispersion arises mainly from Poisson fluctuations and from the simplified CSDA treatment, which neglects scattering and secondary production. Nevertheless, both methods agree very well, with biases below 0.5% and no dependence on primary mass.

4.8. Muon Lateral Distribution

To verify consistency over core distance, the muon lateral distribution function (LDF) was evaluated with both approaches using the same events. Figure 4.10 compares the muon density profiles, and Figure 4.11 shows the mean relative differences. Discrepancies remain below 10% up to about 1200 m, beyond which statistical fluctuations dominate.

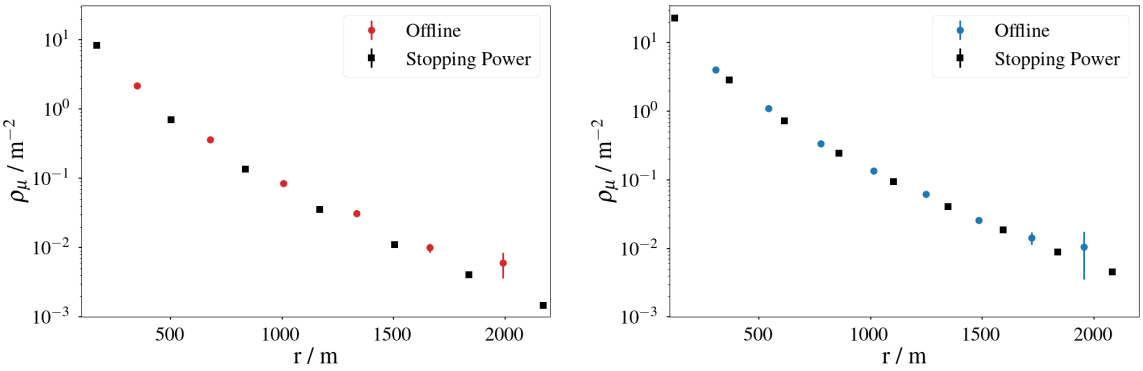


Figure 4.10: Muon LDFs from Offline and the CSDA method for proton (left) and iron (right) showers.

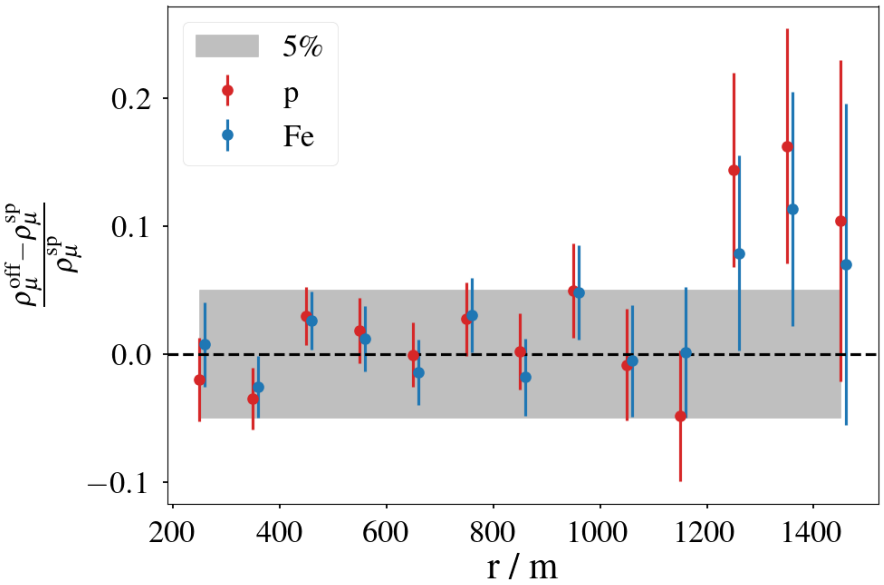


Figure 4.11: Mean relative difference between Offline and CSDA muon densities versus core distance.

4.9. Summary and Conclusions

An independent method based on the Continuous Slowing Down Approximation (CSDA) was developed to estimate the muon flux reaching the UMD depth. Using stopping-power data for SiO_2, each surface muon was propagated through the overburden to determine its residual energy and survival probability. This approach provides a detailed and physically transparent description of the energy losses experienced by muons as they traverse the soil, reproducing the expected 1 GeV transmission threshold, the angular dependence of muon attenuation, and the underground energy spectra obtained from detailed Geant4 simulations.

The comparison shows good consistency: the relative bias between the two approaches is below 0.5%, and the lateral distributions agree within 10% up to 1200 m. The remaining differences can be attributed to Poisson fluctuations and to the absence of multiple scattering and δ-ray production in the CSDA model. Despite these simplifications, the method captures the essential physics of muon energy loss and allows for an intuitive understanding of how the muon flux evolves with depth and angle.

In summary, the stopping-power-based approach offers a simple yet accurate framework to study muon attenuation in the UMD soil. It provides valuable physical insight into the mechanisms governing muon energy dissipation underground and establishes a consistent reference for the interpretation of simulated data used in the following chapters.

Chapter V

Estimation of Muons on Ground Using the UMD

The primary goal of this study is to develop a calibration that can be extrapolated to the main SD array, where no UMD is installed. A connection is needed between the estimated muon signal at the surface and the density of muons that actually traverse the WCD. As a secondary outcome, the calibration provides access to the characteristics of the muonic component in a range of muon energies that is usually inaccessible because detectors are often shielded, and it opens the door to investigate whether the muon deficit is uniform across the muon energy spectrum, an outlook that motivates future work.

The Pierre Auger Collaboration is actively working to obtain observables related to the muon content at the surface using the WCDs, now enhanced by the AugerPrime upgrade. However, a major challenge arises from the fact that WCDs do not discriminate against the electromagnetic component of the extensive air shower, making it difficult to isolate a purely muonic signal. In contrast, the UMD, shielded by soil, is specifically designed to detect only muons with energies above approximately 1 GeV, as discussed in the previous chapter.

Estimating the muon density at the surface is crucial because the UMD, by design, does not register low-energy muons that lose all their energy before reaching its depth. A reliable reconstruction of the on-ground muon density, ρ^{og}, including these low-energy particles, provides a more complete characterization of the shower development and enables additional physics analyses. In particular, it allows testing whether the long-standing *Muon Puzzle* [1], the observed excess of muons in data relative to simulations, is also related to a possibly different muon energy spectrum.

In this chapter, the relationship between the muon densities at the surface, ρ^{og}, and underground, ρ^{ug}, is parametrized and studied as a function of both the primary energy and the zenith angle of the incoming cosmic ray. The goal is to establish a robust empirical model linking the two observables, enabling the estimation of ρ^{og} from UMD measurements.

As the energy of the primary particle increases, the shower maximum develops deeper in the atmosphere [57]. Consequently, high-energy showers traverse less

atmospheric depth before reaching the ground, allowing a larger fraction of low-energy muons to survive. This leads to a decrease in the mean muon energy at ground level [58]. This behavior is illustrated in Figure 5.1, which shows the evolution of the shower maximum, $X_{\max}$, with energy (left) and the corresponding reduction of the mean muon energy near ground (right).

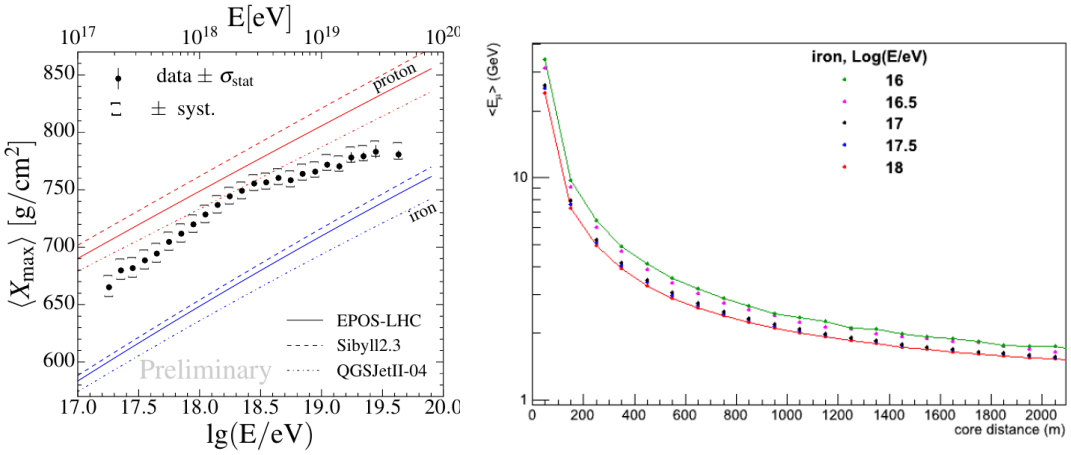


Figure 5.1: Left, depth of shower maximum $X_{\max}$ as a function of primary energy [57]. Right, mean muon energy at ground level as a function of distance to the shower axis [58].

A similar dependence is observed with the zenith angle. Muons originating from more inclined showers tend to be more energetic because they traverse a greater atmospheric depth, where low-energy muons are more likely to be absorbed [59]. This behavior is shown in Figure 5.2, where the mean muon energy increases with zenith angle, reflecting the progressive filtering of softer muons in more inclined showers.

For the simulations used in this work, the dependence of the relevant observables on energy and zenith angle is illustrated in Figure 5.3. The left panel shows the variation of ρ^{og} and ρ^{ug} with zenith angle for a fixed energy bin, and the right panel presents their dependence on energy for a fixed zenith interval.

Regarding the zenith dependence, ρ^{ug} remains nearly constant because it predominantly reflects the higher-energy muons capable of penetrating the soil. In contrast, ρ^{og} decreases at larger zenith angles, where the extended atmospheric path causes the absorption of lower-energy muons.

With respect to the energy dependence, both ρ^{og} and ρ^{ug} increase with the primary energy, following a power-law behavior consistent with previous studies [20]. The faster rise of ρ^{og} may be attributed to the trend observed in Figure 5.1, as the mean muon energy decreases with increasing primary energy, more muons reach the surface, although many lack sufficient energy to be detected by the UMD. This difference between the surface and underground muon densities provides direct sensitivity to the energy distribution of muons in air showers and serves as the foundation for the parameterization developed below.

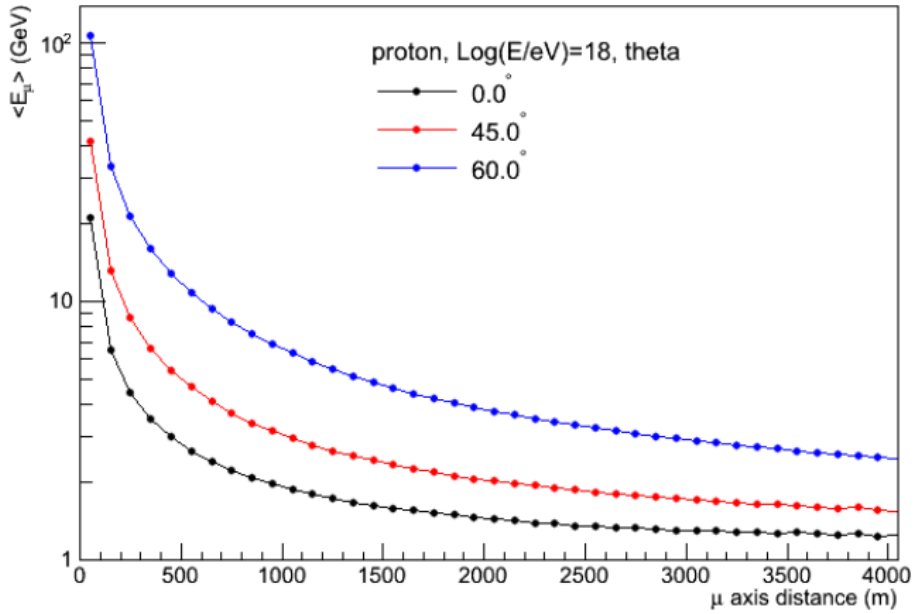


Figure 5.2: Mean muon energy as a function of the primary zenith angle [59].

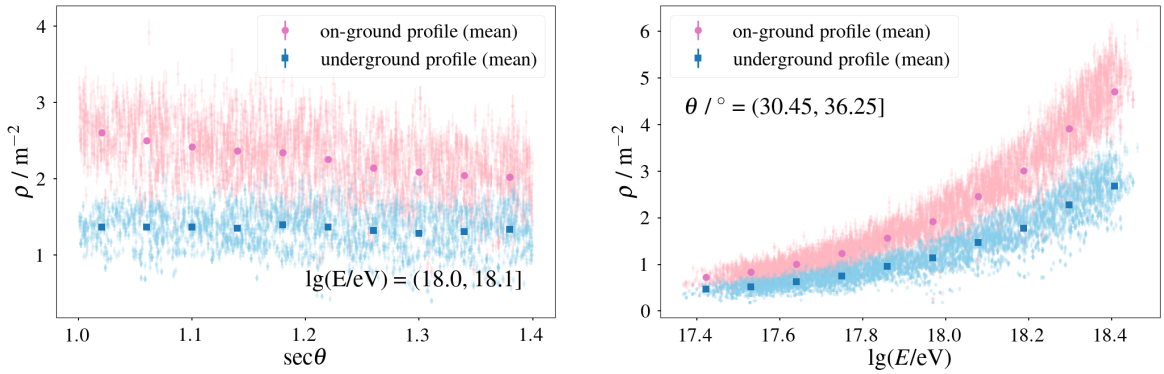


Figure 5.3: Left, dependence of ρ^{og} and ρ^{ug} on zenith angle for a fixed energy bin. Right, dependence on primary energy for a fixed zenith bin.

5.1. Simulation Library and Selection

In this chapter, a continuously distributed shower library in energy was employed, in contrast to the discrete-energy library used in the previous chapter.

The EAS simulations were generated with CORSIKA 7.6400, using the Fluka 2011.2x low-energy interaction model. These simulations were produced by the Monte Carlo task group of the Pierre Auger Collaboration on the local computing cluster of the Institute of Physics of the Czech Academy of Sciences (FZU) in Prague [60]. The simulated primary species included proton, helium, oxygen, and iron nuclei.

Each primary species and energy bin comprised 5000 simulated showers, gen-

erated following an energy distribution proportional to E^{-1} and a zenith-angle distribution proportional to $\sin\theta\cos\theta$, within the range $0° \leq \theta \leq 65°$. For a fixed hadronic model and energy bin, each primary type had 1250 showers simulated under four different atmospheric profiles, representing the seasonal conditions at the Pierre Auger Observatory site in Malargüe. These profiles correspond to the months of January (summer), March (autumn), August (winter), and September (spring).

For the present study, a subset of the full simulation library was selected with the following characteristics:

- **Primary species:** proton, helium, oxygen, and iron.
- **Energy range:** $\lg(E/\text{eV}) = [17.5, 18.0]$ and $[18.0, 18.5]$.
- **Zenith-angle range:** 0°–45°.
- **High-energy hadronic model:** EPOS-LHC.
- **Atmospheric profiles:** January and March.

The zenith-angle selection, limited to $\theta < 45°$, was chosen based on the characteristics of the UMD. Being a flat detector, the UMD experiences increasing reconstruction bias at larger zenith angles, where a significant fraction of muons fail to reach the scintillators. Moreover, the main systematic uncertainty in the reconstructed muon number, $N_{\mu,\text{rec}}$, originates from the so-called "corner-clipping muons", muons that traverse two neighboring scintillator strips [61]. This effect becomes more prominent for more inclined showers, where muons intersect the strips at nearly perpendicular angles. Although a correction algorithm for this bias is implemented in the Offline reconstruction chain, it remains insufficient for highly inclined events.

While the simulation library includes four pure primary compositions, subsequent analyses were performed using a mixed composition model. This mixture was constructed according to the composition fractions measured by the Pierre Auger Observatory from X_{max} observations with the Fluorescence Detector (FD) [57]. The adopted fractions of each primary type for the two energy bins are shown in Figure 5.4.

5.2. Detector Simulation and Muon Density Estimation

The response of the detectors was simulated using $\overline{\text{Off}}\underline{\text{line}}$, following the same procedure as in the previous chapter, each shower being re-sampled 10 times. Due to the 45° zenith cut, not all showers from the original library were included. Building the mass composition mix per energy bin further reduced statistics, leading to a total of ~28 000 events.

For each event, the muon density at 450 m from the shower core was determined using three different methods:

1. **Direct station injected muons (Monte Carlo values):** The density of muons was calculated using a dense ring at a fixed distance of 450 m of 12 WCD stations

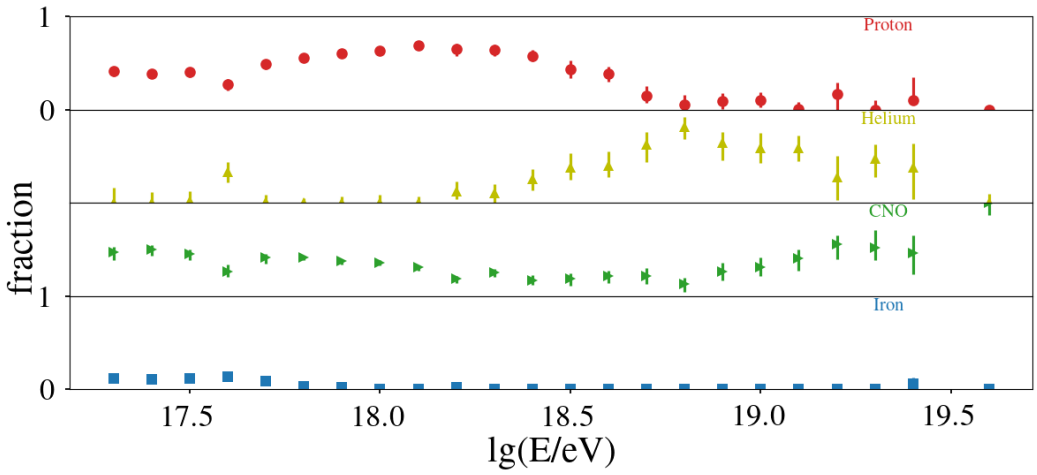


Figure 5.4: Fraction of events corresponding to each mass composition as a function of energy, based on Auger FD measurements [57] for the EPOS-LHC Hadronic Interaction Model.

with their UMD partner. The total number of injected muons in these 12 stations was divided by their total area.

2. **Muon density injected in all UMD stations of the dense ring (Monte Carlo values):** Number of muons that arrive to the detector after traversing the soil in Offline simulations. The total number of muons injected in all the UMD modules in each ring was divided by the sum of all the modules areas, obtaining a single value per shower as it was done in the WCD.

3. **Muon number reconstruction:** The number of muons reconstructed by the detector simulation in the dense ring and the parameter ρ_{450} obtained from the MLDF [62].

4. **Official MLDF fit:** The parameter ρ_{450} obtained from the MLDF [62] fit performed in Offline.

As mentioned before, for the parameterization from underground to on-ground, the Monte Carlo values were used, and then the systematics coming from using the MLDF fit and the station reconstructed density were evaluated.

Since the muon content is strongly correlated with the energy of the primary particle, both the ground-level and underground muon densities were normalized using a power-law dependence, as proposed in [20]. The resulting regressions, along with the corresponding power-law exponent β for each case, are shown in Figure 5.5. Here an Auger-Mix [57] composition for the model EPOS-LHC is used.

The purpose of this normalization is to remove the dominant energy dependence from both variables, thereby isolating the intrinsic correlation between them within each energy and zenith bin. For brevity, in the following sections ρ^{og} and ρ^{ug} denote the energy-normalized values.

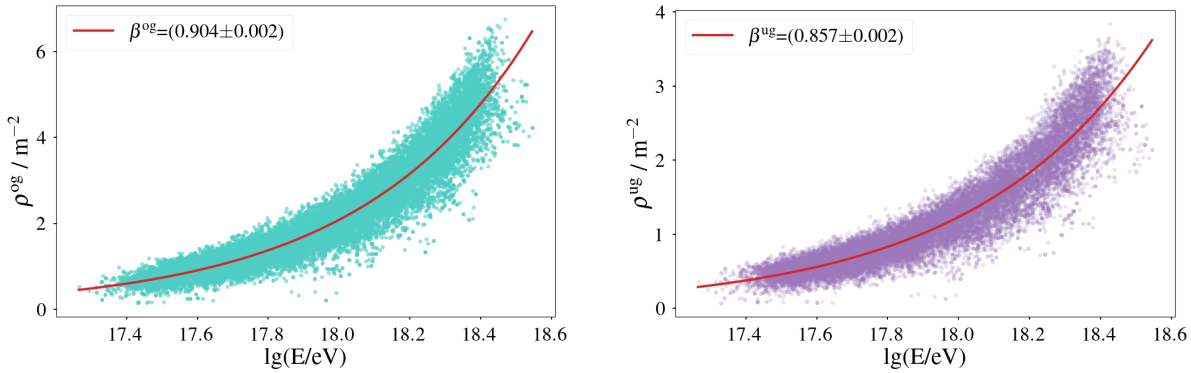


Figure 5.5: Power-law regressions of ρ^{og} (left) and ρ^{ug} (right) versus primary energy with Monte Carlo values.

5.3. Parameterization of Muon Density On-Ground

To investigate the relationship between the density of underground muons, ρ^{ug}, and the density of ground-level muons, ρ^{og}, ρ^{og} was plotted against ρ^{ug} to construct an empirical model. The analysis was carried out in discrete zenith and energy bins, using six uniform bins in $\sec(\theta)$ and energy intervals in $\lg(E/\text{eV}) = 0.1$.

The chosen functional form for the model is a combination of a power law and an exponential,

$$\rho^{og} = e^{\alpha} \, (\rho^{ug})^{\gamma}, \tag{5.1}$$

where γ and α are fit parameters. An example of the fit in a single bin is shown in Figure 5.6.

Fitting was performed independently in each energy–zenith bin across the full range, for an Auger-Mix-like composition at a fix distance of 450 m from the shower core, resulting in the values of γ and α shown in Figure 5.7.

Although empirical in nature, the model is motivated by the physical behavior of muon attenuation in soil. The observed non-linearity between ρ^{og} and ρ^{ug} reflects the fact that an increase in the density of muons on the surface does not translate linearly into the density underground. The parameter γ, which remains below 1 throughout the full range, captures this effect. This trend may be explained by the idea that higher surface muon densities are associated with broader energy spectra, which include a larger fraction of low-energy muons that are more likely to be absorbed before reaching the underground detector. As seen in Figure 5.7 (left), γ tends to larger values at higher zenith angles, consistent with improved underground penetration by more energetic muons.

The normalization factor α, shown in Figure 5.7 (right), also correlates with the energy and the zenith angle. It can be interpreted as an effective attenuation constant, representing the average survival probability of muons without explicitly modeling differences in their energy distributions between ground and underground levels.

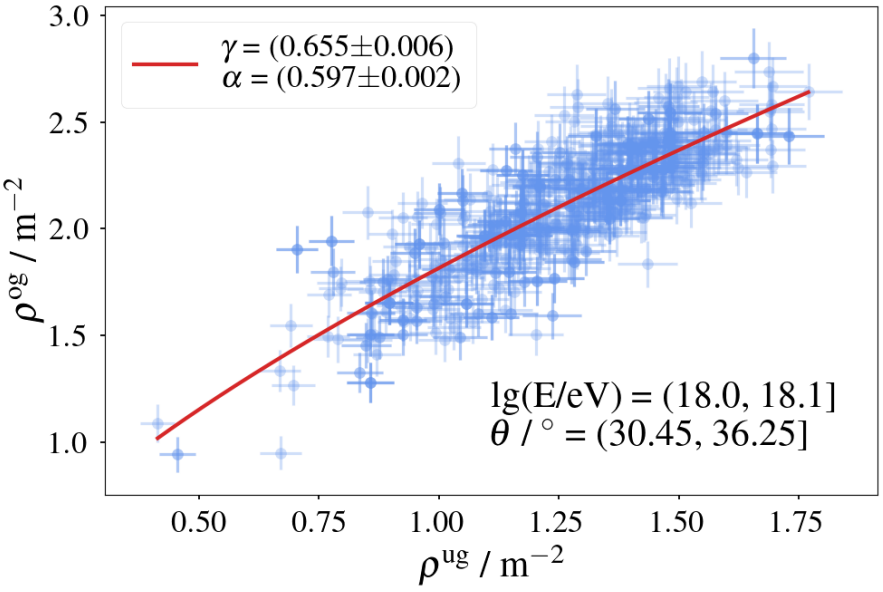


Figure 5.6: Scatter plot of ρ^{og} versus ρ^{ug} for a fixed energy and zenith bin, both quantities are energy-normalized as described in Section 5.1.

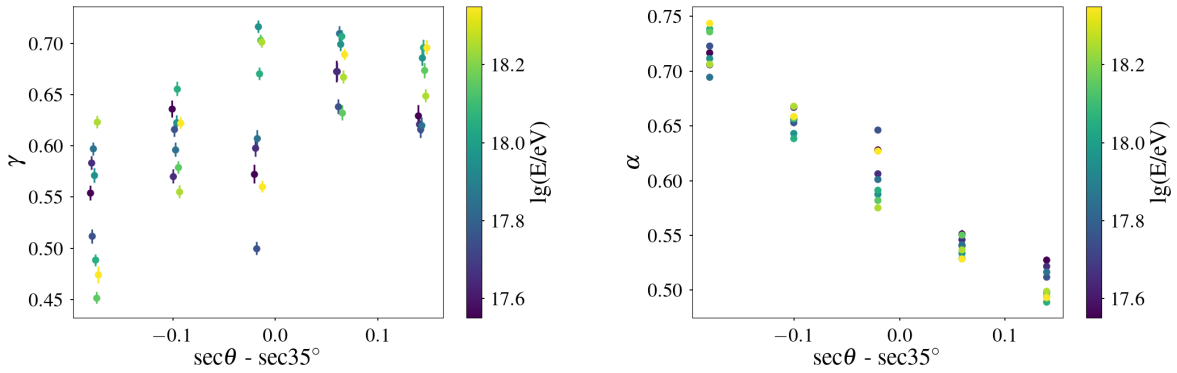


Figure 5.7: Fitted parameters γ (left) and α (right) as functions of zenith angle and energy.

For interpretability, the model can also be expressed in inverse form,

$$\rho^{\mathrm{ug}} = e^{-\alpha} \left(\rho^{\mathrm{og}}\right)^{1/\gamma}. \tag{5.2}$$

Note that the model does not accommodate cases where muons from the surface do not reach the underground detector. Incorporating such behavior would require a more complex formulation, possibly involving a probabilistic or offset-based approach. Given the energy range considered, the large detector area, and the fixed core distance of $r = 450$ m, such events are expected to be rare and to have negligible impact on the results.

5.4. Global Optimization of Parameters

Based on the preceding analyzes, the following global parameterization was optimized using a χ^2 minimization approach.

$$\ln(\rho^{\mathrm{og}}) = \alpha(e, s) + \gamma(e, s)\, \ln(\rho^{\mathrm{ug}}), \tag{5.3}$$

where $e = \lg(E/\mathrm{eV}) - 18.0$ and $s = \sec\theta - \sec 35^\circ$ represent the normalized energy and zenith dependencies, respectively. The functions $\alpha(e, s)$ and $\gamma(e, s)$ describe the energy- and angle-dependent transformation between the underground to on-ground muon densities. Both were initially modeled as second-order polynomials in e and s, including a linear correlation term.

$$\gamma(e, s) = \gamma_0 + \gamma_1\, e + \gamma_2\, s + \gamma_3\, e\, s + \gamma_4\, e^2 + \gamma_5\, s^2, \tag{5.4}$$

with a similar expression for $\alpha(e, s)$.

The complete data set was randomly divided into two subsets: 80% used for regression (training) and 20% reserved for independent testing. Optimal parameters were determined using **Ordinary Least Squares (OLS)** implemented in the *Statsmodels* library [63]. Terms exhibiting low statistical significance were removed iteratively without degrading the overall performance, yielding the reduced and more stable forms.

$$\gamma(e, s) = \gamma_0 + \gamma_1\, e\, s + \gamma_2\, s^2, \qquad \alpha(e, s) = \alpha_0 + \alpha_1\, s + \alpha_2\, e^2. \tag{5.5}$$

The final regression achieved a model p-value of approximately 0.5 and a determination coefficient $R^2 = 0.70$, indicating a satisfactory description of the data given the statistical fluctuations and limited sampling of extreme energies. The optimized coefficients and their statistical uncertainties are summarized in Table 5.1.

Table 5.1: Optimized coefficients of the $\alpha(e, s)$ and $\gamma(e, s)$ parameterizations obtained via χ^2 minimization. The uncertainties correspond to one standard deviation from the OLS fit.

Parameter	Value
γ_0	0.649 ± 0.003
γ_1	-0.601 ± 0.062
γ_2	-0.420 ± 0.173
α_0	0.578 ± 0.001
α_1	-0.613 ± 0.005
α_2	0.061 ± 0.007

The residual distributions for the training and testing subsets are shown in Figure 5.8. Both distributions are centered around zero with standard deviations of

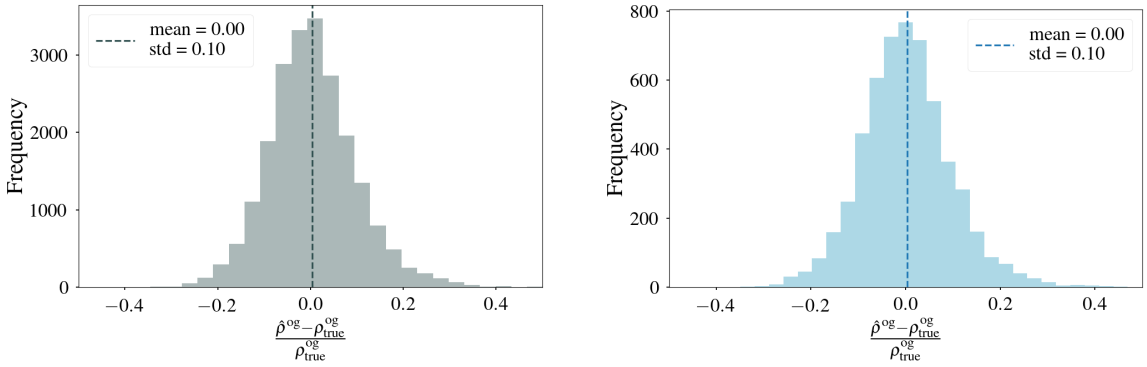


Figure 5.8: Residual distributions for the training (left) and testing (right) subsets. Both sets exhibit symmetric distributions centered at zero, with a standard deviation of ∼0.10, indicating a stable and unbiased fit.

approximately 0.10, confirming the absence of systematic bias and demonstrating consistent performance between the regression and validation data.

Finally, the regression was repeated using the complete data set to obtain the final set of coefficients listed in Table 5.1. The resulting global residual distribution and the correlation between the estimated and true on-ground densities are shown in Figure 5.9. The model reproduces the general trend with minimal bias and a correlation coefficient of $r \simeq 0.84$, confirming that the chosen functional form captures the main dependencies on the energy and the zenith angle.

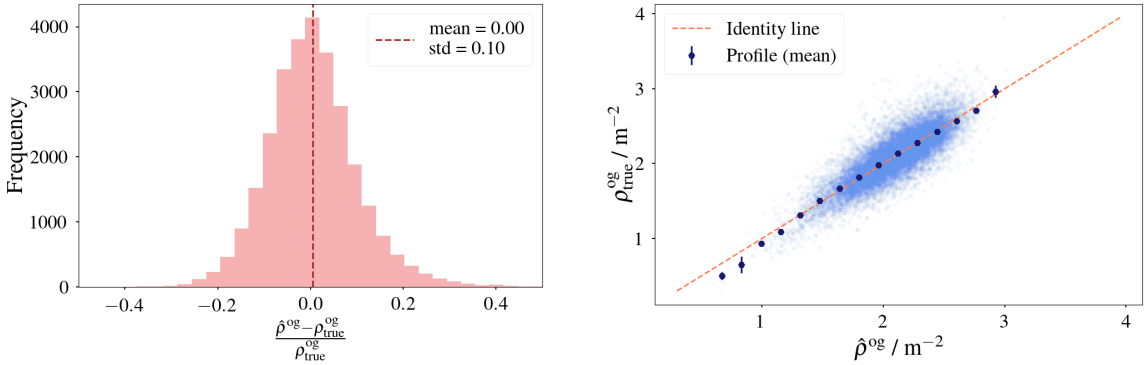


Figure 5.9: Left: Residual distribution for the entire dataset. The distribution remains symmetric and centered around zero, confirming the robustness of the global parameterization. Right: Correlation between the true ($\rho^{\mathrm{og}}_{\mathrm{true}}$) and estimated ($\rho^{\mathrm{og}}_{\mathrm{est}}$) on-ground muon densities. The solid line represents the identity relation, illustrating that the fitted model accurately reproduces the expected behavior over the full dynamic range.

5.5. Bias and Resolution of the Estimated Muon Density

The bias of the estimation was evaluated across the full energy and sec(θ) range using the test subset, as shown in Figure 5.10. The results indicate that the bias is largely independent of energy and zenith angle, remaining within 5%.

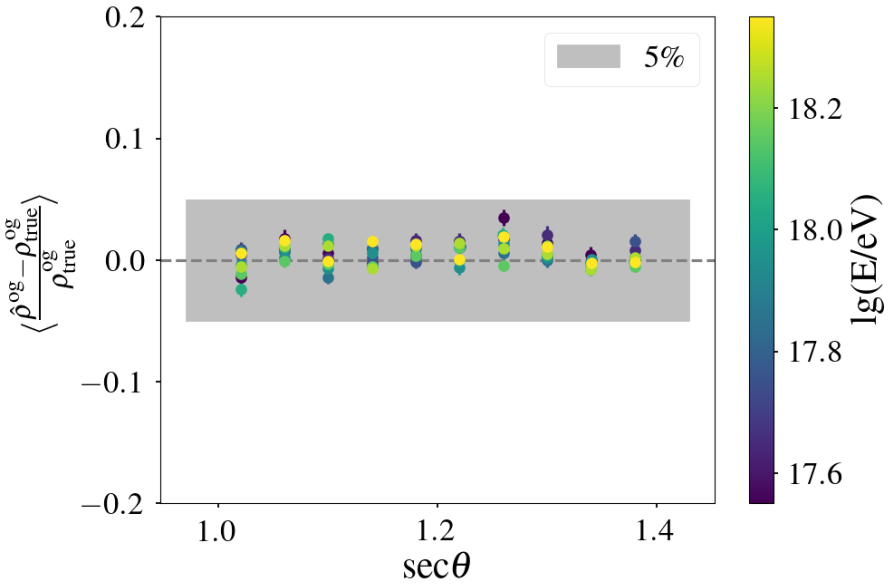


Figure 5.10: Bias across the full energy and sec(θ) range, computed as the mean residual in each bin.

To mitigate the influence of outliers, an Interquartile Range (IQR) filtering method was applied with a loosened multiplier of $k = 5$, whereby data points lying outside the interval $[Q_1 - k\,\mathrm{IQR},\ Q_3 + k\,\mathrm{IQR}]$ are excluded. This approach preserves a broader sample while minimizing the impact of extreme values. Figure 5.11 shows the fraction of data points removed per energy bin, which remains below 1% throughout the energy range.

The resolution of the muon density estimation, shown in Figure 5.12, remains better than 15% across the entire energy and zenith range, with values as low as 6% at the highest energies.

5.6. Systematic Uncertainties

In addition to the intrinsic resolution of the method, several sources of systematic uncertainty were evaluated: mass composition bias, hadronic model, energy resolution, soil density fluctuations, detector reconstruction, and the Muon Lateral Distribution Function.

As parameterization and optimization were performed using the Auger-Mix composition described in Section 5.1, potential biases arising from pure compositions

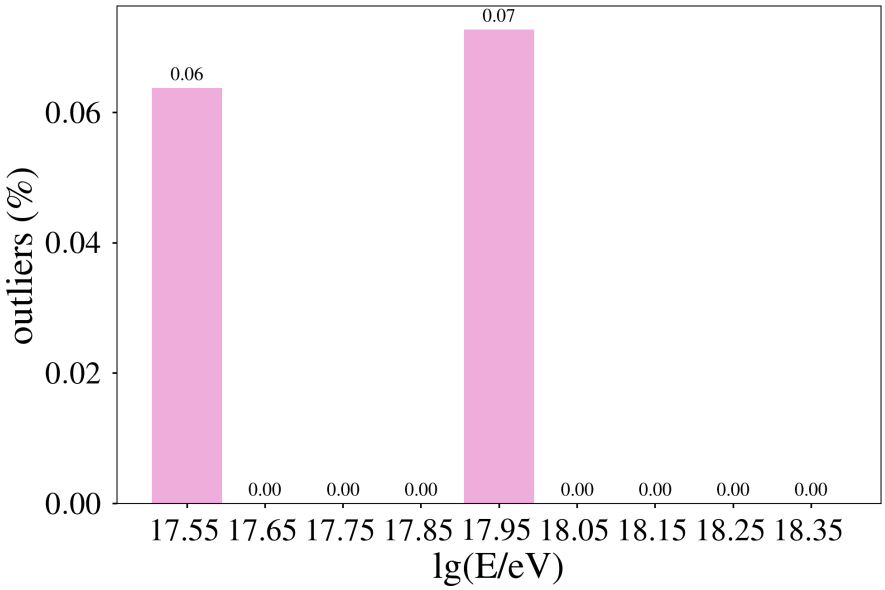


Figure 5.11: Percentage of outliers removed per energy bin.

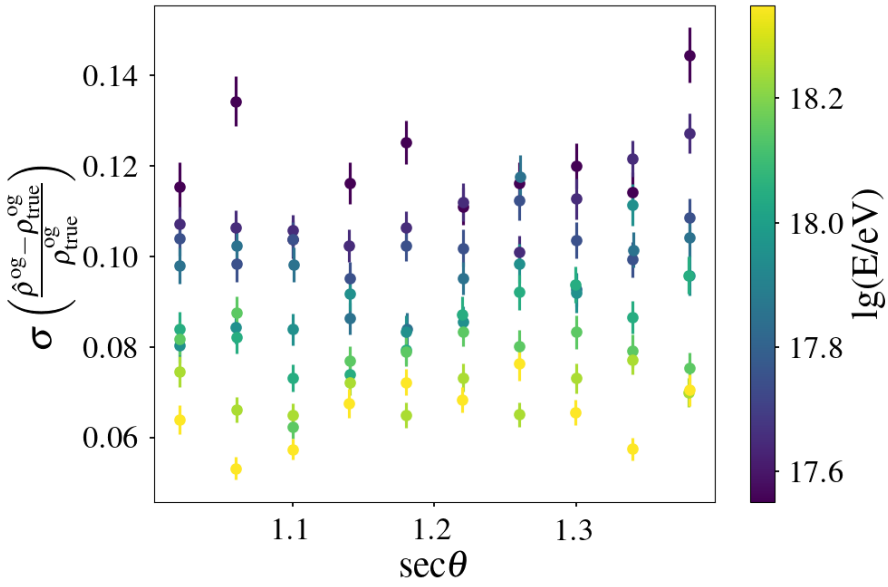


Figure 5.12: Resolution of the estimated muon density across the full energy and sec(θ) range.

were assessed. In Section 4, it was shown that protons and iron primaries exhibit different energy distributions. Consequently, the relationship between underground and on-ground muons may vary for these compositions due to the 1 GeV threshold of the UMD. In the left panel of Figure 5.13, the mass composition bias is shown for a specific energy bin. The right panel presents the mean bias as a function of energy for proton and iron. Throughout the energy range, the mean observed bias was +1% for proton and −1% for iron.

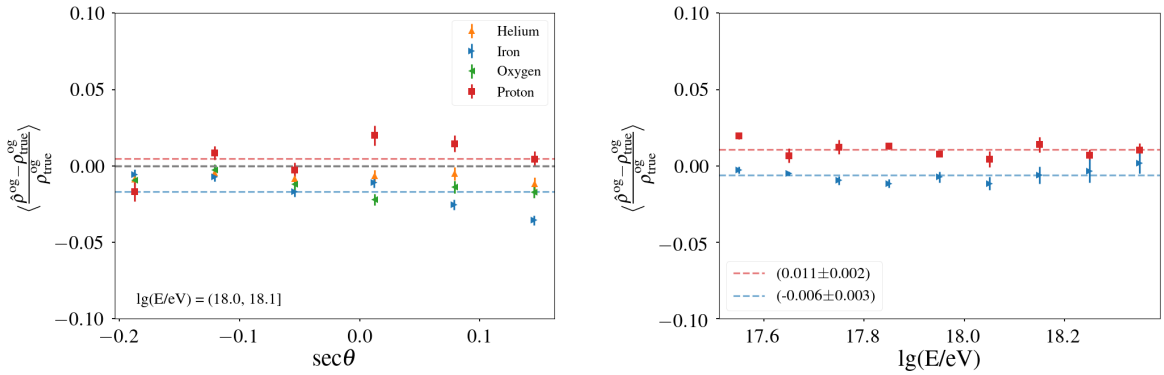


Figure 5.13: Systematic uncertainty due to mass composition.

As discussed in Section 2.6, there are long-standing discrepancies between the number of muons measured by cosmic-ray observatories and the predictions of current hadronic interaction models. To address this *Muon Puzzle*, several models are being refined. The energy distribution of muons varies among models, which directly influences the relationship between surface muons and those that penetrate to the UMD.

The parameterization of on-ground muon densities used here was derived with the EPOS-LHC model, and it was also tested against EPOS-LHC R [64] and Sibyll 2.3e [65]. Since Auger-Mix composition estimates are not yet available for these new models, EPOS-LHC R used the same Auger-Mix as the baseline, and Sibyll 2.3e used the Sibyll 2.3d composition fractions as an approximation. Figure 5.14 summarizes the hadronic-model bias. Across the entire energy range, the mean observed bias is approximately +4% for Sibyll 2.3e and −4% for EPOS-LHC R.

The energy scale of the Observatory is established through fluorescence observations, an inherently calorimetric method. A calibration between FD-measured energy and the surface size parameter S_{35} yields an energy systematic uncertainty of about 14% [66]. Since the muon estimation depends on the primary energy, this uncertainty was propagated, and the same shift was applied to the energy scaling of ρ^{ug} and ρ^{og}. Figure 5.15 shows the corresponding bias. The resulting systematic uncertainty from energy systematic is ±5%.

Variations in soil density across the 23.5 km^2 of the SD-750 array introduce further systematic uncertainties. The average measured soil density is $\delta_{\mathrm{soil}} = 2.380$ g/cm^3, with a standard deviation of $\sigma_\delta = 0.051$ g/cm^3 [50]. Averaged over all zenith angles, a systematic uncertainty of 2.8% was found [4]. Figure 5.16 shows the corresponding

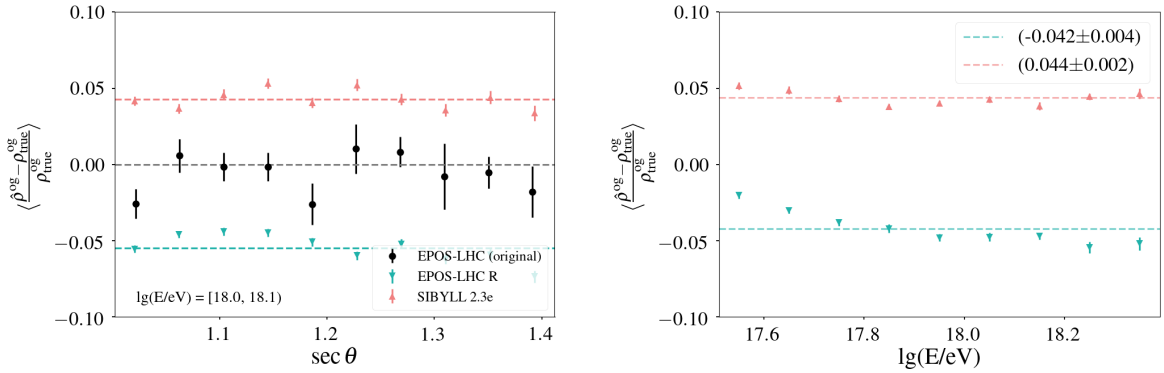


Figure 5.14: Systematic uncertainty associated with the choice of hadronic interaction model. Left, bias for a specific energy bin. Right, mean bias as a function of energy.

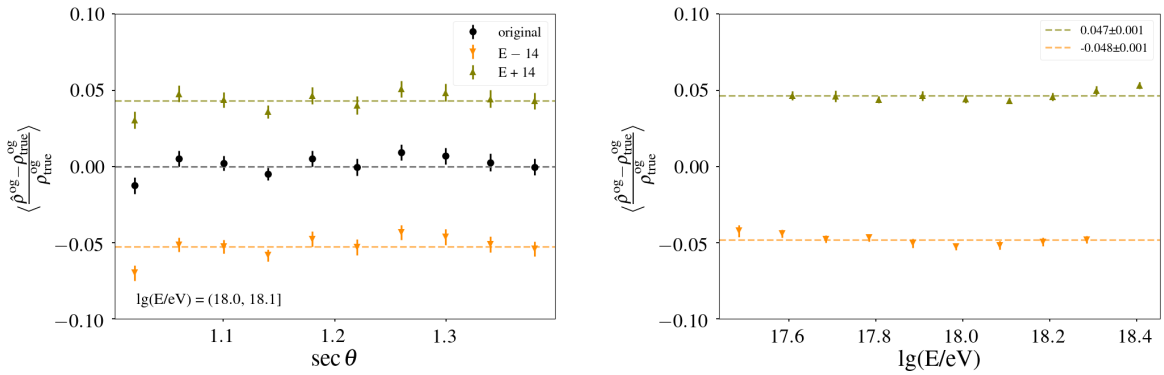


Figure 5.15: Systematic uncertainty due to energy systematic uncertainty.

bias. The resulting systematic ranges from −1% to +2%.

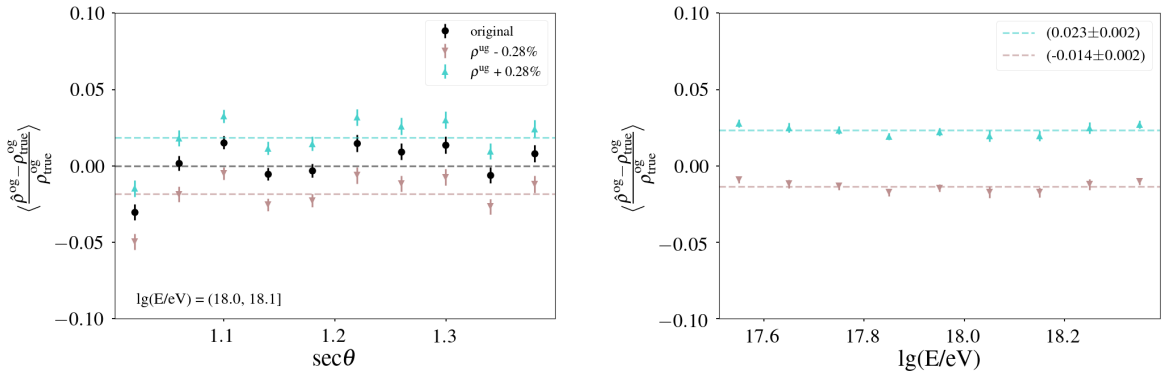


Figure 5.16: Systematic uncertainty due to soil density fluctuations.

As described above, the parameterization of ρ^{og} was based on ρ^{ug} measured using a dense ring of 12 stations at 450 m from the shower core, defined as the sum

Table 5.2: Summary of systematic uncertainties.

Source	Upper Bound	Lower Bound
Mass Composition	+1%	-1%
Hadronic Model	+4%	-4%
Energy Systematic Uncertainty	+5%	-5%
Soil Density Fluctuations	+2%	-1%
Detector Reconstruction	+1%	0%
MLDF fit	0%	-1%
Total	**+6.87%**	**-6.63%**

of injected muons in these stations divided by their total area.

Muon detection in the UMD is achieved by processing the output of each SiPM through a dedicated channel, yielding a binary trace of 2048 samples. A value of "1" or "0" is recorded depending on whether the processed signal exceeds a discriminator threshold. Muon signals are identified as a sequence of four or more consecutive "1"s, known as a muon pattern [67], resulting in a reconstructed number of muons, N_μ^{rec}. These reconstructed values are converted to densities by dividing by the effective area of each UMD module and subsequently fitted to a Muon Lateral Distribution Function (MLDF), as mentioned in Section 5.2.

We evaluate the potential bias from using reconstructed muon densities in the dense ring, $\rho_{\text{rec}}^{\text{ug}}$, instead of the injected Monte Carlo densities, ρ^{injected}, and the bias introduced by using ρ_{450} from the MLDF fit, which is relevant for applying this method to real data. Figure 5.17 shows these biases.

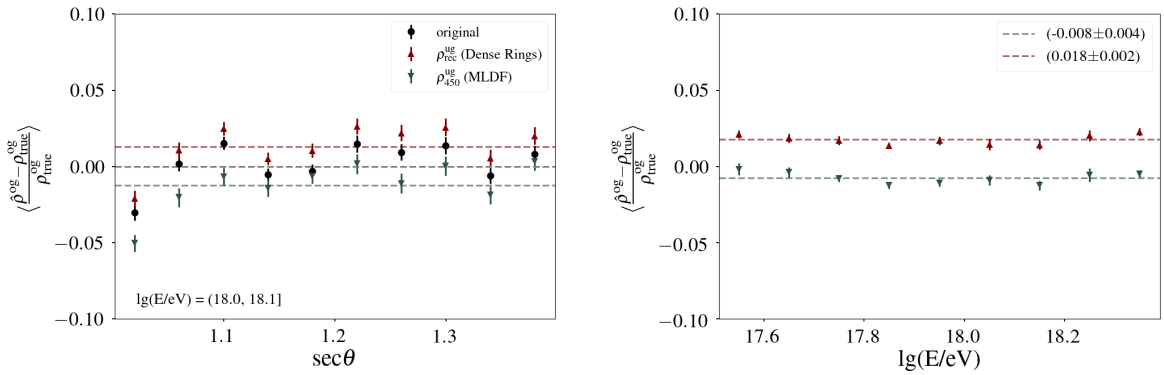


Figure 5.17: Systematic uncertainty due to detector reconstruction and MLDF fitting.

The uncertainty from using reconstructed rather than injected densities is approximately +1%, and the uncertainty from using the MLDF fit is −1%.

The combined systematic uncertainties are summarized in Table 5.2.

5.7. Summary and Conclusions

A method to estimate the muon density at the surface, ρ^{og}, was developed based on the injected muon density in the UMD, ρ^{ug}. Simulations of both the WCDs and the UMD were carried out using $\overline{\mathrm{Off}}\underline{\mathrm{line}}$. Previous composition estimates, based on the fraction of four pure primary masses, were used to construct an Auger-Mix-like composition.

ρ^{og} was defined as the total muon density crossing each WCD in a dense ring of stations located 450 m from the shower core. This quantity was fitted against ρ^{ug} for each simulated event. The variation of the obtained fit parameters with respect to the primary energy and zenith angle was studied, leading to a global parameterization of ρ^{og} as a function of ρ^{ug}, energy, and zenith angle.

The model parameters were optimized using least-squares minimization, resulting in relative residuals with a mean of zero and a standard deviation of 0.1. The regression yielded a p-value of 0.5 and a coefficient of determination $R^2 = 0.70$. The parameterization is robust, exhibiting a zero-centered bias that is independent of zenith angle and energy, with a resolution better than 15%. Five sources of systematic uncertainty were evaluated, leading to a total systematic uncertainty of +6.87% and −6.63%.

This study demonstrates the capability of the UMD to validate and calibrate surface muon content estimations through direct muon measurements, and it motivates the extrapolation of this calibration to the main array where underground detectors are not present.

Chapter VI

Validation and Proof of Concept in the 750 Infill

A central goal of the Pierre Auger Observatory is to achieve a reliable measurement of the muon content in ultra-high-energy cosmic-ray air showers, particularly at the highest energies. However, the 1500 m array of the SD lacks the UMD, which prevents a direct measurement of the muonic component. The method developed in this work provides a promising approach to estimate the number of muons at ground level, offering an observable that can be used to infer and cross-check the muonic signal in the WCDs. Nonetheless, this approach requires thorough validation to ensure its reliability and consistency before being applied to the full array.

In this chapter, a calibration test is presented using the pure muonic signal in the WCDs. This signal is first calibrated against the true density of muon injected into the ground, $\rho^{\mathrm{og}}_{\mathrm{true}}$, and subsequently against the estimated muon density, $\hat{\rho}^{\mathrm{og}}$, obtained using the method described in the previous chapter. Both calibrations are compared and their dependencies on the zenith angle (θ) and energy (E) are analysed.

6.1. Muon Density to Muon Signal

Muons traverse the water volume in the WCDs produce Cherenkov radiation with an intensity depending on E and θ. The Cherenkov radiation occurs when a charged particle moves through a medium at a speed greater than the speed of light in that medium [68]. Radiation is emitted in a coherent cone at an angle defined by the following:

$$\cos\theta_c = \frac{1}{n\beta}, \tag{6.1}$$

where n is the refractive index of the medium and β is the velocity of the particle expressed as a fraction of the speed of light in vacuum. The number of photons emitted per unit path length and per unit energy interval by a particle with charge $z_p e$ at the limiting angle is given by:

$$\frac{dN^2}{dEdx} \simeq \frac{\alpha z_p^2}{\hbar c} \sin^2\theta_c \simeq \left(\frac{370}{\text{eV cm}}\right) \sin^2\theta_c, \tag{6.2}$$

where α is the fine-structure constant. This implies that the number of photoelectrons detected is proportional to the track length of the muon through the detector medium (water).

The dependence of muon track length in WCDs on the zenith angle has been a subject of study since the early years of the Pierre Auger Observatory [69, 70, 71]. These studies have demonstrated a clear correlation between the track length and the angle at which muons enter the detector. Furthermore, as shown in Chapter 4, muons tend to arrive with a zenith angle similar to that of the primary cosmic-ray particle. Consequently, the muonic signal per unit muon density is expected to exhibit a dependence on the zenith angle.

Regarding primary energy, it was also discussed in Chapter 4 that higher-energy primaries tend to produce muons with lower average energies. This leads to a higher population of muons that would not trigger the WCD as the primary energy increases, which, as it will be shown, is translated to a negative offset in the signal to muon density curve.

In this chapter, the muonic signal, S_μ, is calibrated in simulations as it is considered the most accurate proxy of the muon content at the station level on the surface.

To establish the relationship between S_μ and the density of muons traversing the WCD, ρ^{og}, the mean value of S_μ in the 450 m dense ring per event is computed. Then, it is normalized by its dependence with the primary energy, which follows $S_\mu \sim E^{\beta_s}$, as was done before with the muon densities, as can be seen in Figure 6.1.

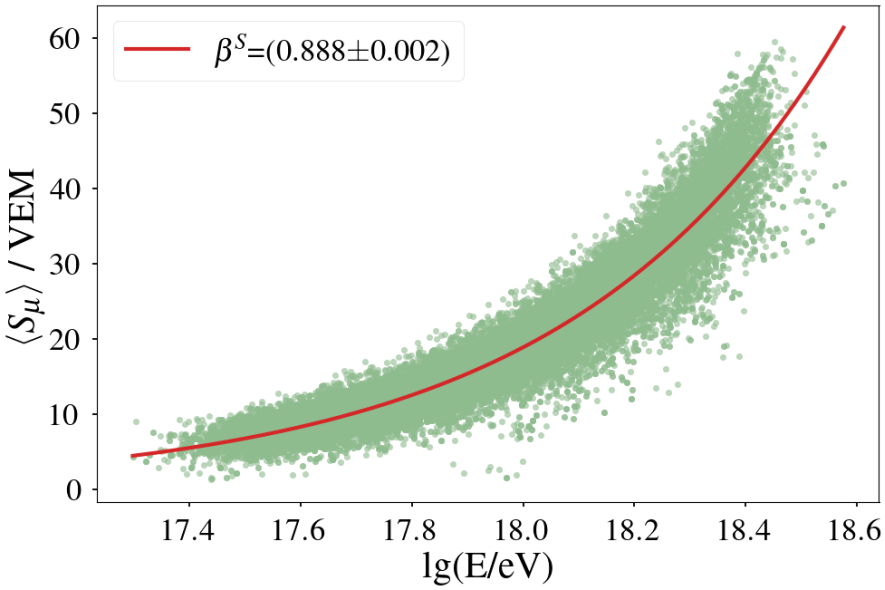


Figure 6.1: S_μ dependence on primary energy.

Furthermore, the total ρ^{og} that traverses all the stations in the ring is calculated as the sum of all injected muons in the ring divided by the total effective area of all the stations. The effective area of each WCD is given by:

$$A_{\text{eff}} = 10.18\,\text{m}^2\,\cos\theta \;+\; 4.32\,\text{m}^2\,\sin\theta \tag{6.3}$$

as defined in [72]. Here, θ is the zenith angle of the primary particle, and the values 10.18 m^2 and 4.32 m^2 correspond to the top area of the WCD and the longer side of the SSD, respectively (the latter serving as an approximation for the lateral area of the WCD).

Figure 6.2 shows the mean muon signal per ring plotted against the injected muon density.

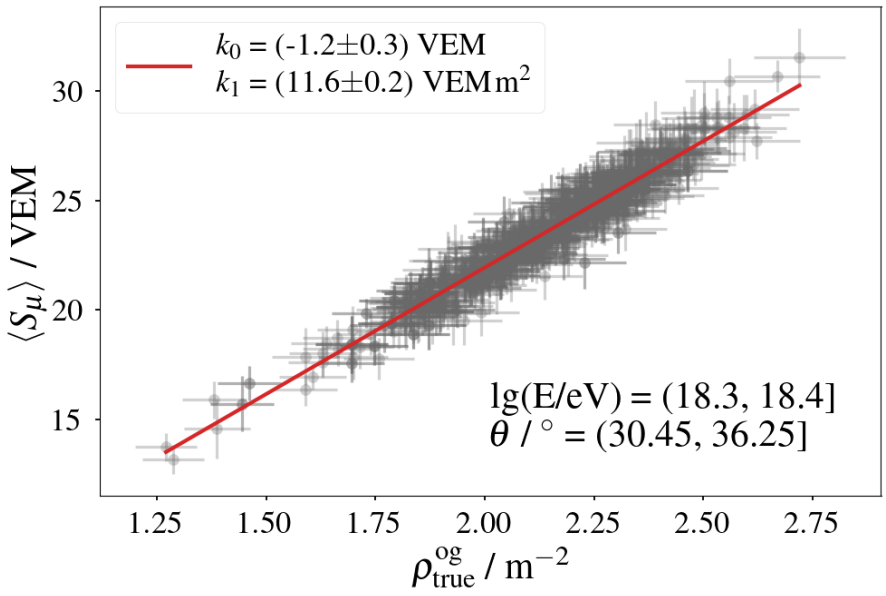


Figure 6.2: Mean muon signal per dense ring vs. injected muon density.

Figure 6.2 presents an example of the calibration using MC values of muon densities for a fixed energy/zenith bin. The linear dependence between both observables is evident, being:

$$\langle S_\mu \rangle = k_0 + k_1\,\rho^{\text{og}} \tag{6.4}$$

Linear parameters k_0 (offset) and k_1 (slope) were calculated for each energy / zenith bin, as depicted in Figure 6.3 with $s = \sec\theta - \sec 35°$ and $e = \lg\left(\frac{E}{10^{18}\,\text{eV}}\right)$.

- k_0**:** Displays random fluctuations but no significant dependence on E or θ. Its approximately constant and negative value can be interpreted as an offset accounting for muons that cross the WCD but fail to produce a full trigger signal, such as those entering at shallow angles or with energies just below the

detection threshold. This constant correction, typically of the order of one VEM (about 10% of the total signal), reflects an intrinsic detector inefficiency rather than a dependence on shower geometry or energy.

- k_1**:** The slope of the dependence, which quantifies the conversion between the traversing muon density and the measured signal. Its variation is primarily governed by the zenith angle, with a weaker dependence on the shower energy. At larger zenith angles, muons travel longer paths through both the atmosphere and the detector. The increased atmospheric path selectively attenuates low-energy muons, leading to a muon population with higher average energy at the ground. Consequently, each muon deposits more energy—and therefore produces a larger signal—when crossing the detector. This results in an overall increase of k_1 with zenith angle. The residual dependence on energy arises from the gradual change in the muon energy spectrum with the primary energy of the shower. However, this effect is smaller compared to the zenith dependence.

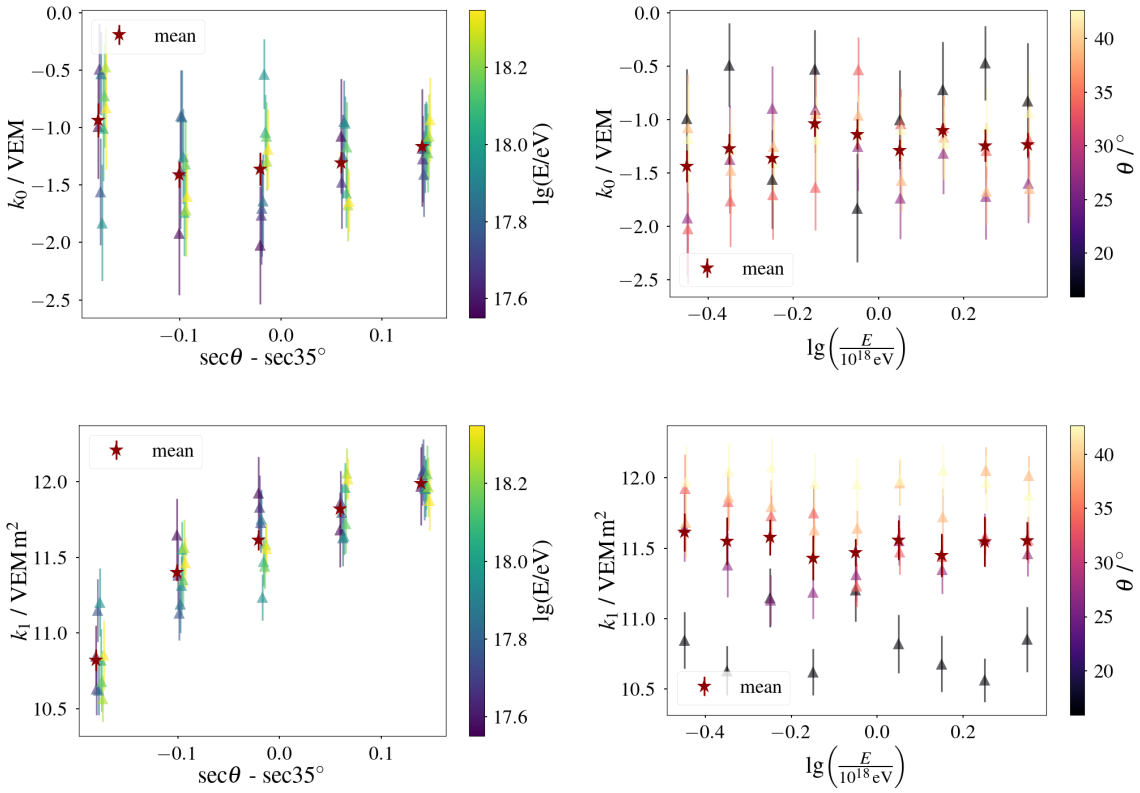


Figure 6.3: Parameters k_0 (offset) and k_1 (slope) as functions of s and e. Top: k_0 versus s (left) and e (right). Bottom: k_1 versus s (left) and e (right). Points are color–coded by energy and zenith angle, respectively.

6.2. Calibration using $\hat{\rho}^{og}$

After characterizing the relationship between the muon signal and the true traversing muon density, as well as the expected behaviour of the corresponding calibration parameters, the same procedure was repeated using the estimated density of muons on the ground $\hat{\rho}^{og}$. This estimator, obtained through the method described in the previous chapter, provides an indirect reconstruction of the muon density at the surface and, therefore, can serve as a practical input for the calibration of the mean muon signal $\langle S_\mu \rangle$. The fits were again performed in bins of energy and zenith angle, allowing for the study of possible dependencies on both variables. An example of the resulting linear calibration is shown in Figure 6.4.

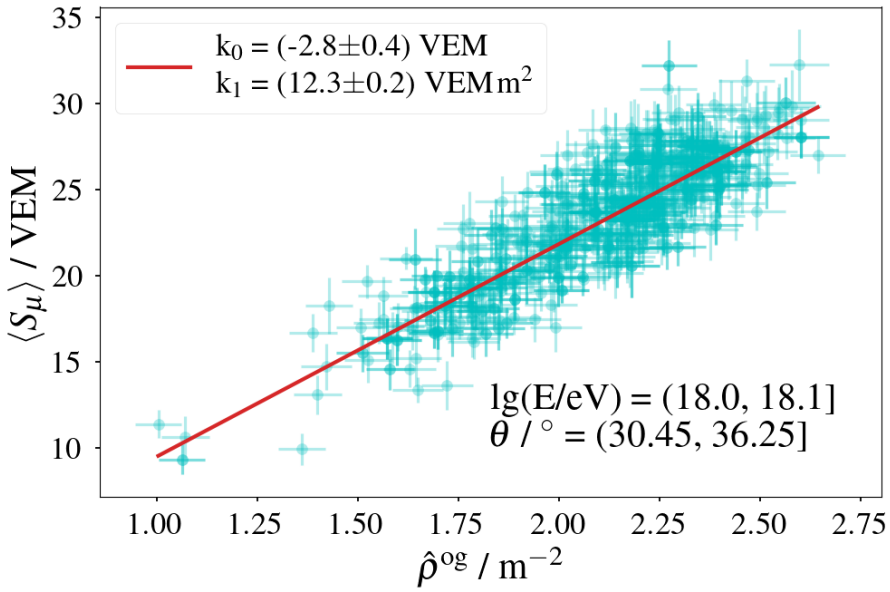


Figure 6.4: Example of the linear correlation between the mean muon signal per dense ring, $\langle S_\mu \rangle$, and the estimated on–ground muon density $\hat{\rho}^{og}$ for showers with $18.0 < \lg(E/\text{eV}) \leq 18.1$ and $30.4° < \theta \leq 36.3°$. Each point represents the average over all stations within each event, and the line indicates the best linear fit $\langle S_\mu \rangle = k_0 + k_1\, \hat{\rho}^{og}$.

For each energy–zenith bin, the calibration parameters k_0 (offset) and k_1 (slope) were extracted and then analysed as functions of the $s = \sec\theta - \sec(35°)$ $e = \lg\left(\frac{E}{10^{18}\,\text{eV}}\right)$

By construction, $\hat{\rho}^{og}$ is already normalized with respect to energy, while $\langle S_\mu \rangle$ was corrected for its intrinsic energy dependence using the exponential model introduced in Figure 6.1. The resulting dependencies of k_0 and k_1 on s and e are shown in Figure 6.5.

As illustrated in Figure 6.5, both calibration parameters display stronger scatter when using the estimated $\hat{\rho}^{og}$ instead of the true muon densities. This behaviour is expected since the estimator carries additional statistical and systematic uncertainties. However, the mean values of k_0 and k_1 in the zenith bins preserve the general

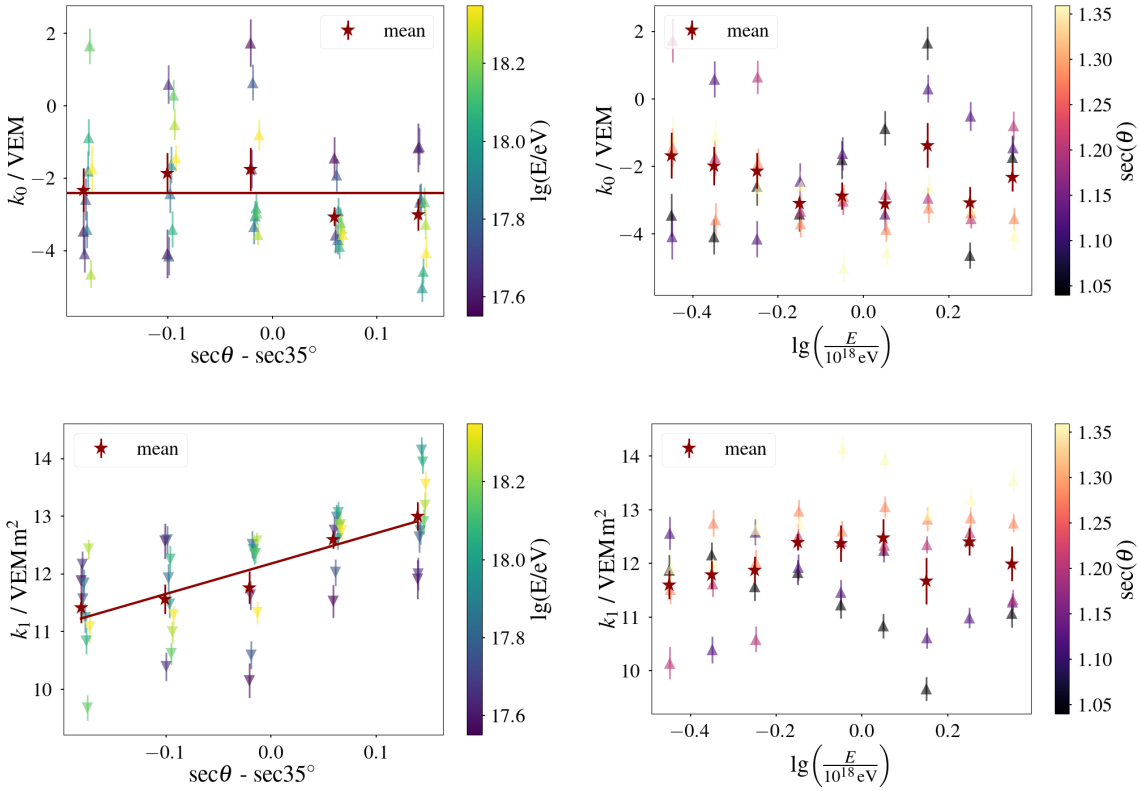


Figure 6.5: Calibration parameters k_0 (offset) and k_1 (slope) obtained from fits using the estimated on–ground muon density $\hat{\rho}^{\mathrm{og}}$. **Top:** k_0 as a function of the zenith variable s (left) and of the energy variable e (right). **Bottom:** k_1 as a function of s (left) and e (right). Each point corresponds to a single energy–zenith bin, and the color scale indicates the complementary variable (energy or zenith). The parameters show larger fluctuations than in the case using the true ρ^{og}, reflecting the propagation of uncertainties from the reconstruction process.

trends observed previously: k_0 remains approximately constant with s, while k_1 increases with increasing zenith angle, indicating a stronger muon contribution to S_μ for inclined showers.

To describe these dependencies quantitatively, a simple linear function was fitted to the behaviour of k_1 as a function of s:

$$k_1(s) = m_0 + m_1\, s\,, \tag{6.5}$$

where the mean value of k_0 was adopted as a fixed offset parameter. The resulting best–fit parameters are summarized in Table 6.1.

6.3. $\hat{\rho}^{\mathrm{og}}$ as a function of S_μ

After determining the calibration parameters, the final step in this proof of concept was to evaluate whether the density of muons on the ground, ρ^{og}, can be reliably

Table 6.1: Calibration parameters obtained from $\langle k_0 \rangle$ and the linear dependence of $k_1(s)$ in Eq. 6.5. Uncertainties correspond to the standard errors of the fits.

Parameter	Value
k_0	(-2.11 ± 0.18) VEM
m_0	(12.03 ± 0.09) VEM m^{-2}
m_1	(4.56 ± 0.74) VEM m^{-2}

estimated as a function of the muon signal S_μ. This derived quantity, denoted $\rho^{\mathrm{og}}(S_\mu)$, is validated against direct muon measurements from the UMD.

Based on the previously obtained calibration, $\hat{\rho}^{\mathrm{og}}(S_\mu)$ is estimated using:

$$\hat{\rho}^{\mathrm{og}}(S_\mu, s) = \frac{S_\mu - k_0}{k_1(s)} \tag{6.6}$$

where k_0 and $k_1(s)$ are the calibration parameters derived in the previous section using the estimated muon density $\hat{\rho}^{\mathrm{og}}$ obtained from the UMD simulations.

The bias and resolution of this parameterization were evaluated across the full energy and zenith range, as shown in Figure 6.6.

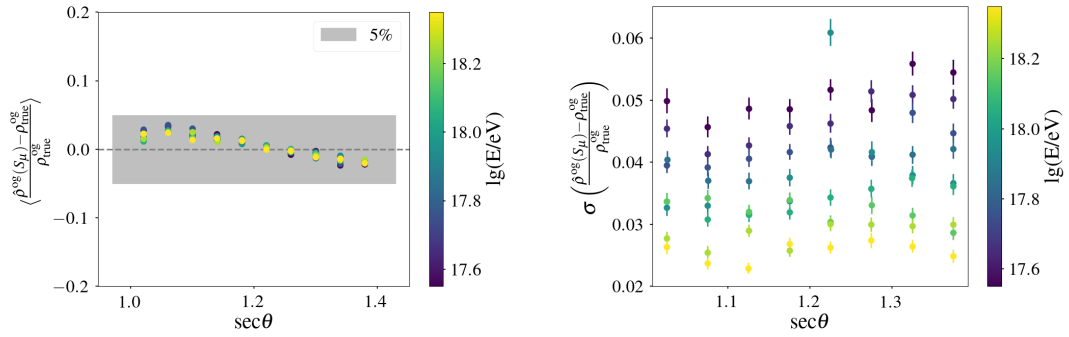


Figure 6.6: Bias (left) and resolution (right) of the estimated value $\hat{\rho}^{\mathrm{og}}$ using the k_0 and k_1 parameterization.

The results show that the bias is largely independent of energy with a slight zenith dependence, staying within ±5%. The resolution remains below 6%, reaching values as low as 3% at the highest energies.

6.4. Systematic Uncertainties of Calibration

In Chapter 5, the systematic uncertainties in the estimation of muons on the ground were analyzed and computed, yielding a total range of +6.87% to −6.63%. These uncertainties must be considered as a potential source of systematic error in the calibration itself.

Additionally, uncertainties related to the mass composition and energy scaling were revisited to assess their potential impact as a source of systematic bias in the calibration.

The value of $\hat{\rho}^{\text{og}}$ was shifted by its systematic uncertainty bounds (±) and the calibration procedure was repeated bin by bin using these new values. The resulting calibration parameters were compared with those obtained from the original data. The parameter k_1 exhibited a noticeable change under these shifts, while k_0 remained largely unaffected.

Figure 6.7 shows a representative example of the calibration curves for the nominal and systematically shifted values of $\hat{\rho}^{\text{og}}$. Figure 6.8 illustrates the variation of the k_0 and k_1 parameter across the full zenith range due to this source of uncertainty.

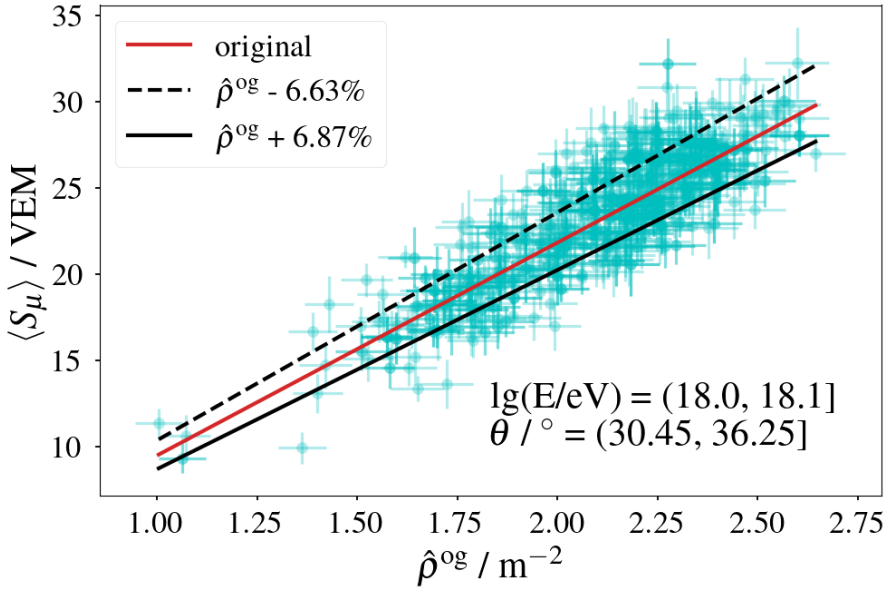


Figure 6.7: Example of the calibration in a single energy and zenith bin using the nominal $\hat{\rho}^{\text{og}}$ (red line) values shifted by their upper (dashed black line) and lower (full black line) systematic uncertainties. This demonstrates the impact of $\hat{\rho}^{\text{og}}$ uncertainty on the calibration curve.

The new parameterizations of k_1 were then used to re-estimate the value of $\hat{\rho}^{\text{og}}$ as a function of S_μ, allowing the quantification of the bias introduced by this source of uncertainty. The results are shown in Figure 6.9.

The systematic uncertainty propagated from the model of $\hat{\rho}^{\text{og}}$ was found to be in the range of +8% to −6%.

As in the previous chapter, an Auger-Mix composition model was used to perform the calibration in each bin. To check for potential mass composition systematics, the calibration was repeated assuming pure proton and iron compositions. The impact of this assumption is shown in Figure 6.10.

The uncertainty of the mass composition was found to exhibit a weak dependence on energy, with a mean bias of 0% for proton and approximately 1% for iron.

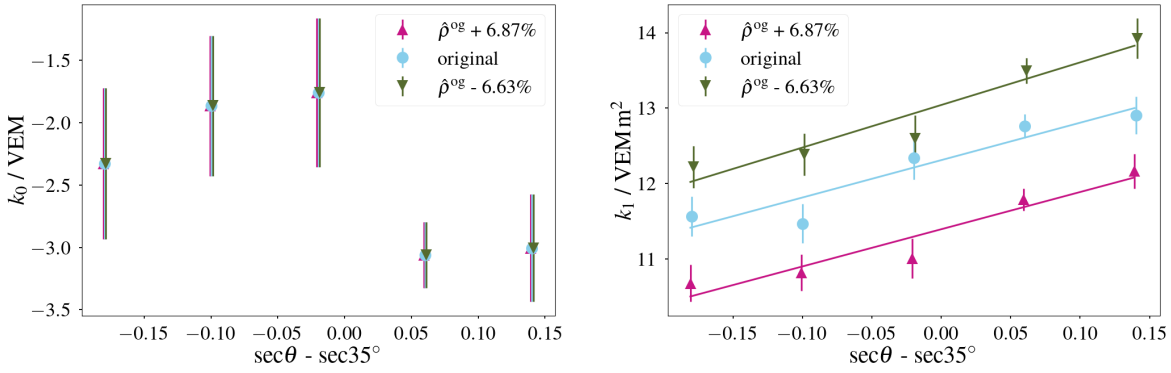


Figure 6.8: Left: Fitted parameter k_0 across zenith angle bins as a result of shifting $\hat{\rho}^{og}$: no sensitivity is observed. Right: Variation of the fitted parameter k_1 across zenith angle bins as a result of shifting $\hat{\rho}^{og}$, showing sensitivity of the slope to this source of uncertainty.

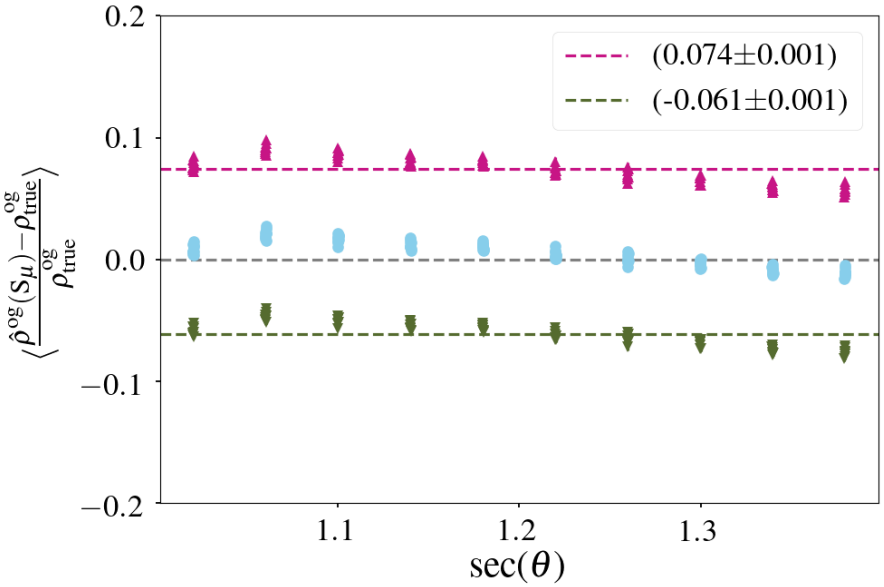


Figure 6.9: Bias in the estimated $\hat{\rho}^{og}$ as a function of S_μ due to systematic uncertainties in the $\hat{\rho}^{og}$ model. This quantifies how the model uncertainty propagates into the reconstructed observable. Color coded as Figure 6.8.

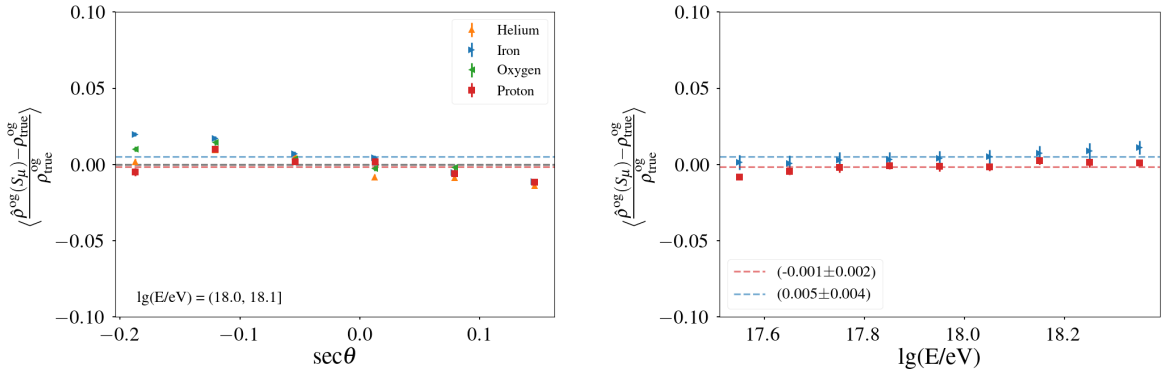


Figure 6.10: Left: Systematic bias in the $\hat{\rho}^{\mathrm{og}}$ vs. S_μ calibration for a representative energy bin, introduced by assuming different mass compositions. Right: Comparison of the calibration-derived bias for pure proton and pure iron assumptions, highlighting the effect of composition on the reconstructed observable.

The impact of the energy resolution was revisited following the procedure described in Chapter 5. Since both S_μ and $\hat{\rho}^{\mathrm{og}}$ are energy-scaled observables, the primary energy was systematically shifted by ±14% to evaluate the associated uncertainty.

Figure 6.11 presents an illustrative example of the calibration curves obtained in a single energy and zenith bin for the nominal energy and for the systematically shifted values. The variation of the calibration parameters k_0 and k_1 throughout the entire zenith range is summarized in Figure 6.12. Although k_0 remains largely unaffected by the energy shift, k_1 exhibits a clear sensitivity to this source of systematic uncertainty.

Modified parameterizations of k_1 were subsequently used to re–evaluate $\hat{\rho}^{\mathrm{og}}$ as a function of S_μ, allowing quantification of the induced bias. The resulting deviation is shown in Figure 6.13.

Overall, the propagated systematic uncertainty from the energy resolution was found to lie within the range of +7% to −3%.

The combined impact of the two sources of systematic uncertainties, as well as their sum in quadrature, are presented in Table 6.2.

Table 6.2: Summary of systematic uncertainties

Source	Upper Bound	Lower Bound
$\hat{\rho}^{\mathrm{og}}$ uncertainty	+8%	-6%
Mass composition	+1%	0%
Energy Resolution	+7%	-3%
Total	**+10.67%**	**-6.71%**

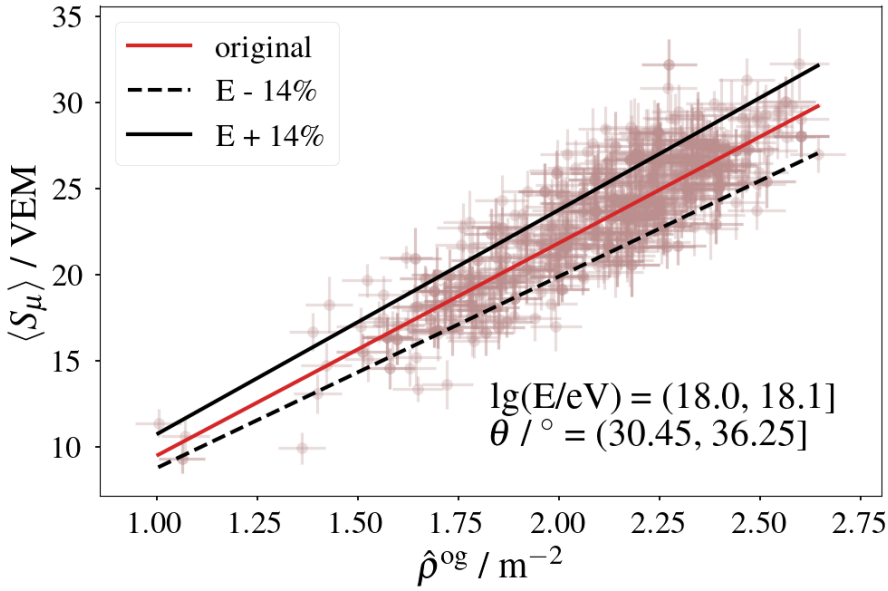


Figure 6.11: Example of the calibration in a single energy and zenith bin using the nominal energy (red line) and the values shifted by their systematic uncertainties: upper (solid black line) and lower (dashed black line). The plot highlights the effect of the $\hat{\rho}^{\mathrm{og}}$ energy scaling uncertainty on the calibration curve.

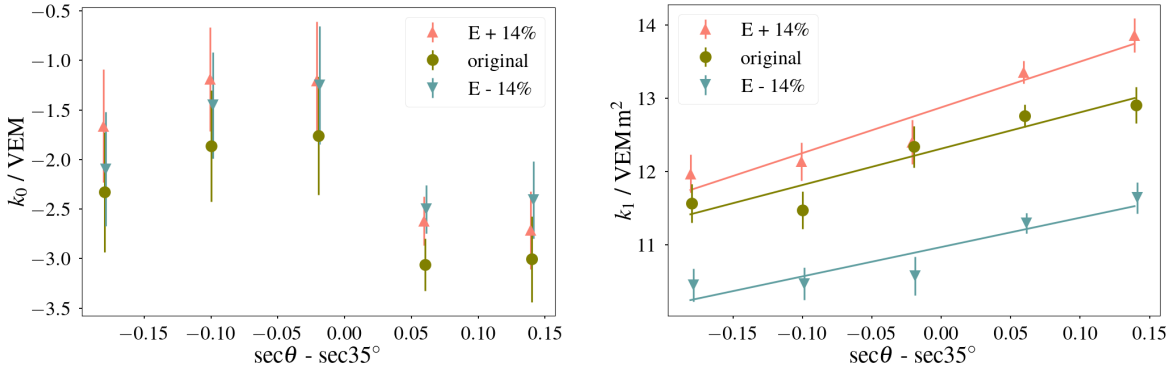


Figure 6.12: Left: Fitted parameter k_0 across zenith bins after shifting $\hat{\rho}^{\mathrm{og}}$; no significant dependence is observed. Right: Variation of the fitted parameter k_1 across zenith bins after applying the energy scaling shift, showing a clear sensitivity of the slope to this uncertainty.

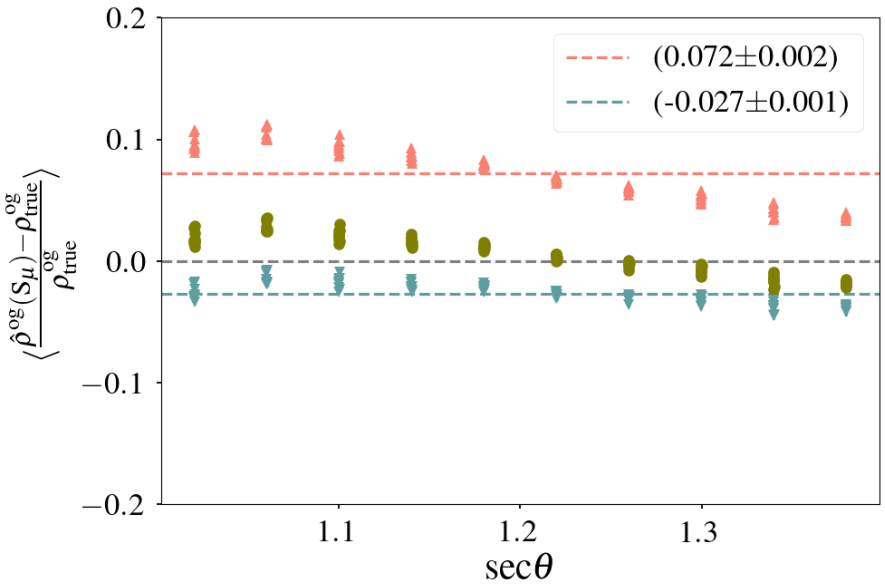


Figure 6.13: Bias in the reconstructed $\hat{\rho}^{\mathrm{og}}$ as a function of S_μ due to systematic uncertainties in the energy scaling color coded as Figure 6.12. This quantifies how the energy resolution propagates into the observable.

6.5. S_μ Artificial Fluctuations

As explained in the beginning of this chapter, the S_μ values used in this analysis were obtained by integrating the pure muon trace in the simulated WCD. However, in real measurements, this signal must be inferred using methods that separate the muonic component from the electromagnetic component of the total signal. One such method will be discussed in Chapter 7.

Since the goal is to eventually use the estimated ground-level muon content to calibrate surface-level muon signal estimations, it is important to quantify the resolution required from an estimator of S_μ for the calibration to remain reliable.

To investigate this, artificial Gaussian noise was added to the S_μ distributions within each energy and zenith bin. The noise had a standard deviation proportional to the true S_μ value:

$$S_\mu^{\mathrm{noisy}} = S_\mu + \mathcal{N}(0,\, (c\, S_\mu)^2) \tag{6.7}$$

where the dimensionless constant c defines *relative resolution* - that is, the standard deviation of the fractional error $(S_\mu^{\mathrm{noisy}} - S_\mu)/S_\mu$ is equal to c, this is explained in detail in Appendix B.

Figure 6.14 shows examples of the resulting S_μ^{noisy} distributions for various values of c.

The calibration procedure was repeated for each artificial resolution level, which yielded different values of the fit parameters k_0 and k_1, as shown in Figure 6.15. As

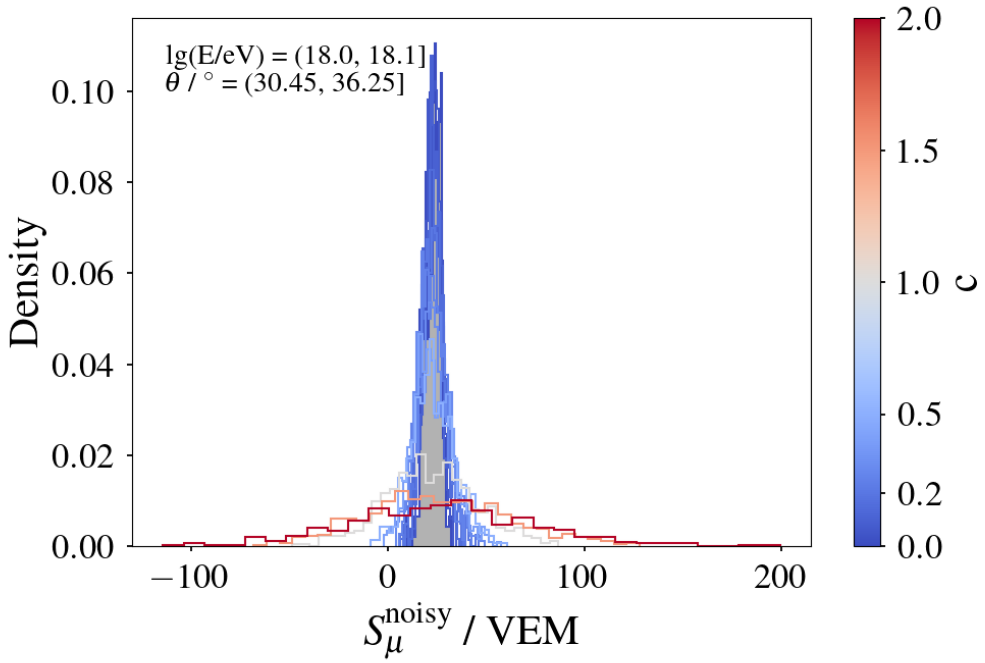


Figure 6.14: Distributions of S_μ for different resolution levels (i.e. different values of c).

expected, a poorer resolution (larger c) leads to increased scatter and bias in the fitted parameters.

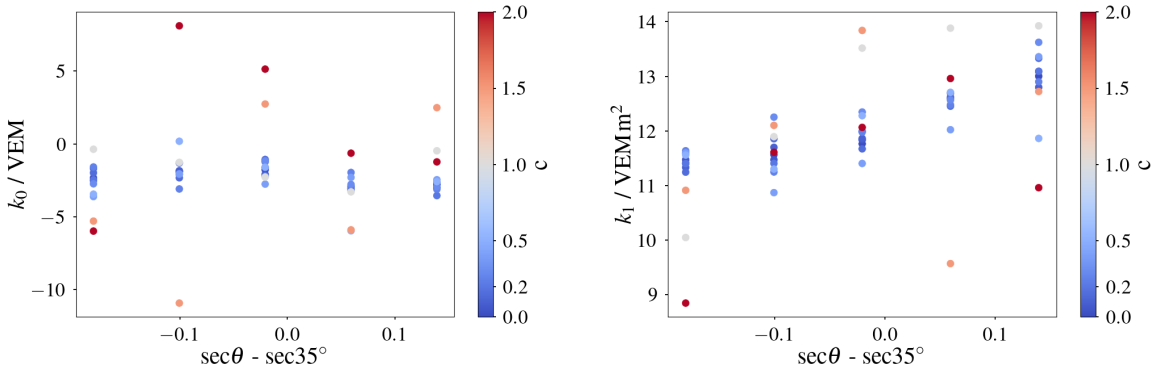


Figure 6.15: Values of k_0 and k_1 obtained for different S_μ resolution levels (i.e. different c).

Finally, the bias in the estimated $\hat{\rho}^{\text{og}}$ was calculated using the different (k_0, k_1) pairs. A maximum acceptable bias of 5% was chosen as a criterion to define the tolerance for S_μ resolution. As shown in Figure 6.16, this condition was satisfied up to a relative resolution of approximately $c \approx 0.3$, or 30%.

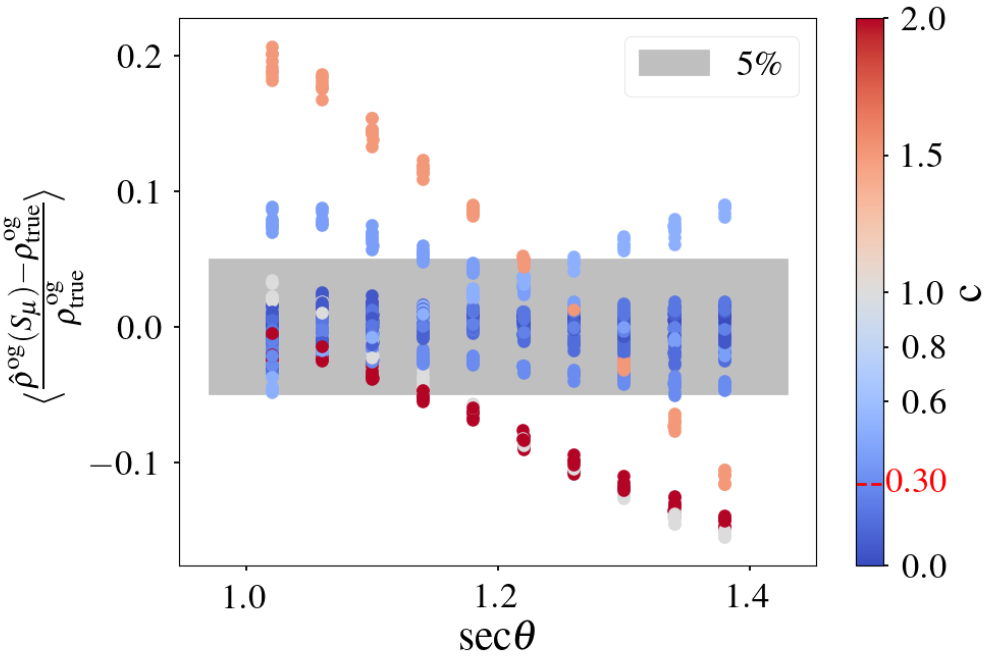


Figure 6.16: Bias in $\hat{\rho}^{\mathrm{og}}$ for different S_μ resolution levels. The relative resolution limit to obtain a bias below a 5% is shown as a red dashed line in the colorbar.

6.6. Summary and Conclusions

In this chapter, a calibration study was performed to relate the muonic signal in WCDs to the density of muons that arrive at ground level. This calibration is a crucial step toward validating the estimation of the muonic component in air showers, especially for the 1500 m array of the Pierre Auger Observatory, where the UMD is not available.

The calibration was first carried out using the true injected muon density, $\rho^{\mathrm{og}}_{\mathrm{true}}$, obtained directly from simulations. The procedure was then repeated using the estimated muon density, $\hat{\rho}^{\mathrm{og}}$, as described in the previous chapter. The goal was to determine whether the estimated observable could serve as a suitable proxy for calibration purposes.

The relationship between the muon signal, S_μ, and the muon density was found to be approximately linear, and was characterized by two parameters: the slope k_1 and the offset k_0. These parameters were extracted for each bin in zenith angle and primary energy, and their behaviors were studied as functions of the normalized zenith variable $s = \sec\theta - \sec 35^\circ$ and normalized energy $e = \lg\left(\frac{E}{10^{18}\,\mathrm{eV}}\right)$.

The slope k_1 showed a clear dependence on zenith angle, increasing for more inclined showers. This trend is consistent with the expected increase in muon track length through the WCDs at higher inclinations, which results in larger signals per muon. The energy dependence of k_1 was found to be less pronounced.

The offset k_0 exhibited more fluctuating behavior and no dependence on energy

and zenith.

Importantly, the calibration performed using the estimated muon density $\hat{\rho}^{\mathrm{og}}$ yielded results consistent with those obtained from the true muon density, supporting the validity of the estimation method presented earlier in this work. This consistency indicates that $\hat{\rho}^{\mathrm{og}}$ can reliably replace $\rho^{\mathrm{og}}_{\mathrm{true}}$ in calibration analyses, enabling the use of S_μ as a physics observable in data where the true muon density on the surface is inaccessible.

Overall, this chapter demonstrates that the reconstructed ground-level muon density can be successfully used to calibrate the muonic signal in WCDs, thereby supporting the broader objective of muon content characterization in ultra-high-energy cosmic ray showers at the Pierre Auger Observatory.

Chapter VII

Calibration Using Neural Network Outputs

Since the true muon signal, S_μ, cannot be directly measured from data, considerable effort has been devoted within the Pierre Auger Collaboration to develop reliable estimation methods [72, 73, 74, 75]. Among these, deep neural networks and other machine-learning techniques have proven particularly effective.

This chapter evaluates the performance of a neural-network-based estimator developed in [76, 77]. The estimator is treated as a proxy observable for S_μ and is validated and calibrated against the on-ground muon signal estimator introduced in Chapter 5. The network is trained using supervised learning, in which the algorithm learns a mapping between input features and target outputs (labels) using simulated data. Once trained, it can predict target values for previously unseen inputs.

In simulations, the total recorded signal in each WCD comprises four components:

- the muonic component,
- electromagnetic (EM) particles from muon decays and interactions,
- EM particles from neutral-pion decays, and
- EM particles produced in hadronic jets, i.e., high-transverse-momentum particles created during the final stages of the shower development.

The neural network described in [76, 77] was trained to isolate the muonic component within the first 200 FADC bins of the signal trace. This choice is motivated by the fact that, for showers with energies below 10^{19} eV, more than 90% of triggered stations contain the complete muon signal within these first 200 bins [75]. The contribution from later bins is therefore negligible for the purpose of muon-signal reconstruction.

The original network was trained using simulations from the 1500 m surface-detector array and achieved an S_μ estimation bias below 5%. For the present work, the model was retrained by the FZU group from the Pierre Auger Collaboration, using simulations corresponding to the denser 750 m array, employing the EPOS-LHC hadronic interaction model, an energy range of $\lg(E/\mathrm{eV}) \in [17.5, 18.5]$. As in the original setup, a dense ring of stations was simulated at a core distance of 450 m.

The output of the retrained network, hereafter denoted $\hat{S}_\mu$, was subsequently energy-normalised and calibrated following the procedure outlined in Chapter 6. Before applying this calibration, its resolution was evaluated across the relevant energy and zenith-angle ranges. Figure 7.1 shows the bias distribution and overall resolution of the estimator. The resolution of $\hat{S}_\mu$ remains below 14%, well within the 30% limit defined in Chapter 6 as the threshold required to keep the bias in the calibrated values $\hat{\rho}_\mu(S_\mu)$ below 5%. This confirms that the estimator is suitable for use in the calibration procedure.

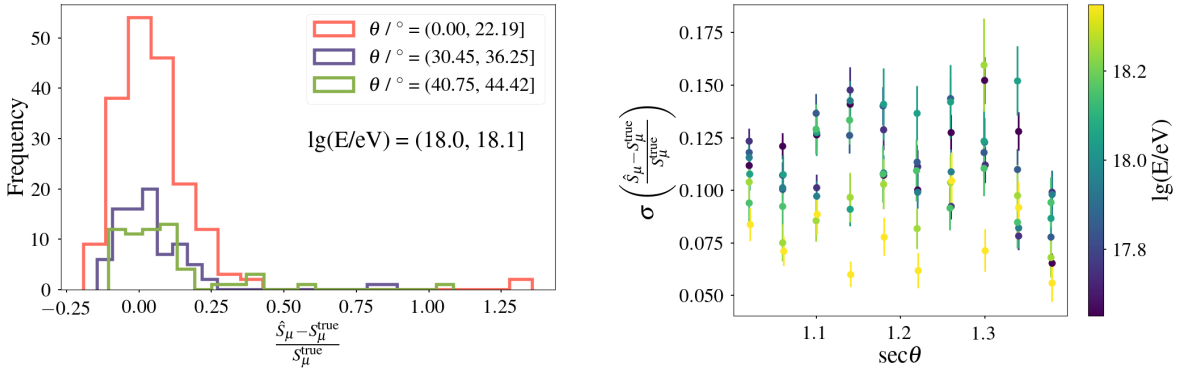


Figure 7.1: Bias (left) and resolution (right) of the $\hat{S}_\mu$ estimator obtained from the retrained model of [76].

The final calibration was performed using the two estimators $\hat{S}_\mu$ and $\hat{\rho}^{\text{og}}$. The motivation for this analysis lies in the fact that both quantities can be derived directly from the measurement capabilities of the detectors in the 750 m infill array of the Pierre Auger Observatory. Consequently, they can be obtained from real data without relying on additional simulation inputs, enabling a fully data-driven cross-check of the calibration. This approach ensures that the neural-network-based estimator $\hat{S}_\mu$ is physically anchored to observables accessible at the surface and can thus be meaningfully related to the on-ground muon density.

7.1. Injection Volume and Unthinning

In CORSIKA simulations, the computational time increases rapidly with primary energy, becoming prohibitively long for energies $E_0 > 10^{16}$ eV. To mitigate this, a technique known as "thinning" (or "importance sampling") is used. When thinning is enabled, all secondary particles with energies below a defined fraction of the primary energy, referred to as the thinning level $\varepsilon_{\text{th}} = E/E_0$, are subject to this algorithm. Only one particle from this group is selected and tracked, and its contribution is reweighted accordingly, while the rest are discarded [47].

For simulating the SD, the thinning procedure introduces artificial fluctuations in the particle densities at ground level. To smooth these effects, a *local sampling* technique is used. In this approach, all particles landing within a *sampling zone* (see

Figure 7.2) around a WCD are selected. Their weights are scaled by the ratio of the detector area to the sampling area and their arrival times are adjusted to preserve their time delay relative to the shower front. An unweighted set of particles is then randomly placed on the WCD surface for further simulation [78].

When the UMD is included in the simulations, a larger sampling zone encompassing both the WCD and the UMD must be used during the unthinning process.

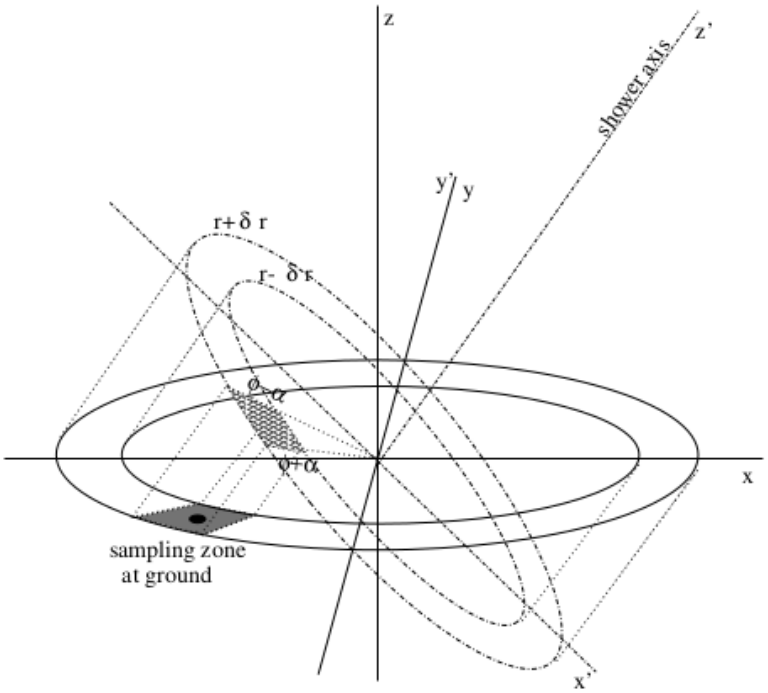


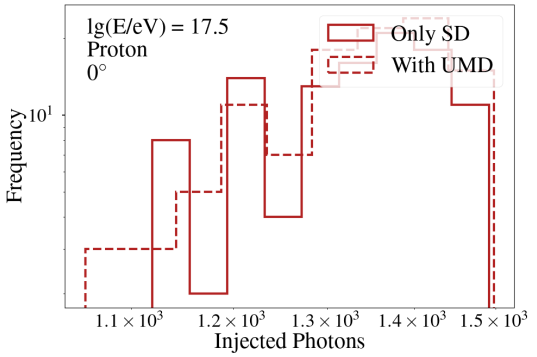
Figure 7.2: Example of a *sampling zone* associated with a WCD station. Figure adapted from [78].

Since neural networks (NN) were originally trained in simulations that included only WCD, a preliminary validation was performed to assess whether the enlarged sampling zone - required when simulating UMD - could influence the particle distributions used by the NNs. Figure 7.3 shows the distributions of injected photons, electrons and muons in WCDs, comparing simulations with and without the UMD.

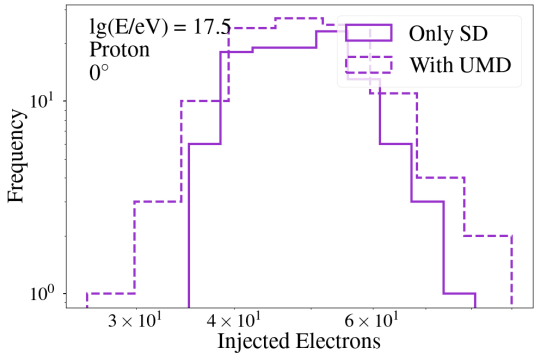
As shown in Figure 7.3, no significant qualitative differences are observed between the two simulation setups. Similarly, Figure 7.4 shows that the integrated signal per particle type in the WCDs remains consistent regardless of whether the UMD is included.

The same plots for inclined and Iron showers can be found in Appendix C.

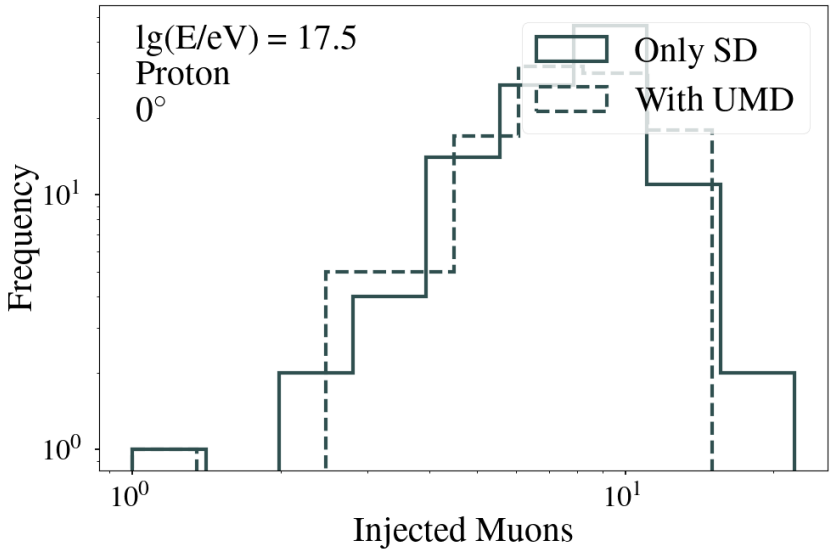
These findings are consistent with the earlier study presented in [79], reproduced in Figure 7.5, which demonstrated negligible bias in both particle counts and integrated signals across a range of sampling zone radii and distances from the shower core.



(a) Photons



(b) Electrons



(c) Muons

Figure 7.3: Distributions of injected particles in the WCDs for identical CORSIKA showers. Solid lines show simulations with only the WCD, dashed lines include the UMD.

7.2. $\hat{S}_\mu$ and Its Relation to On-Ground Muons

Since the neural networks (NNs) employed in this work are not yet optimized to process the WCD low-gain (LG) channel, only stations that did not saturate in the high-gain (HG) channel were used. This selection restricts the analysis to lower-energy events but ensures that the reconstructed observables are not affected by signal saturation.

Following the procedure described in Chapter 6, the relationship between $\hat{S}_\mu$ and the true muon density on the ground, ρ^{og}, was first examined. A clear positive correlation was observed, allowing for a linear calibration between both quantities,

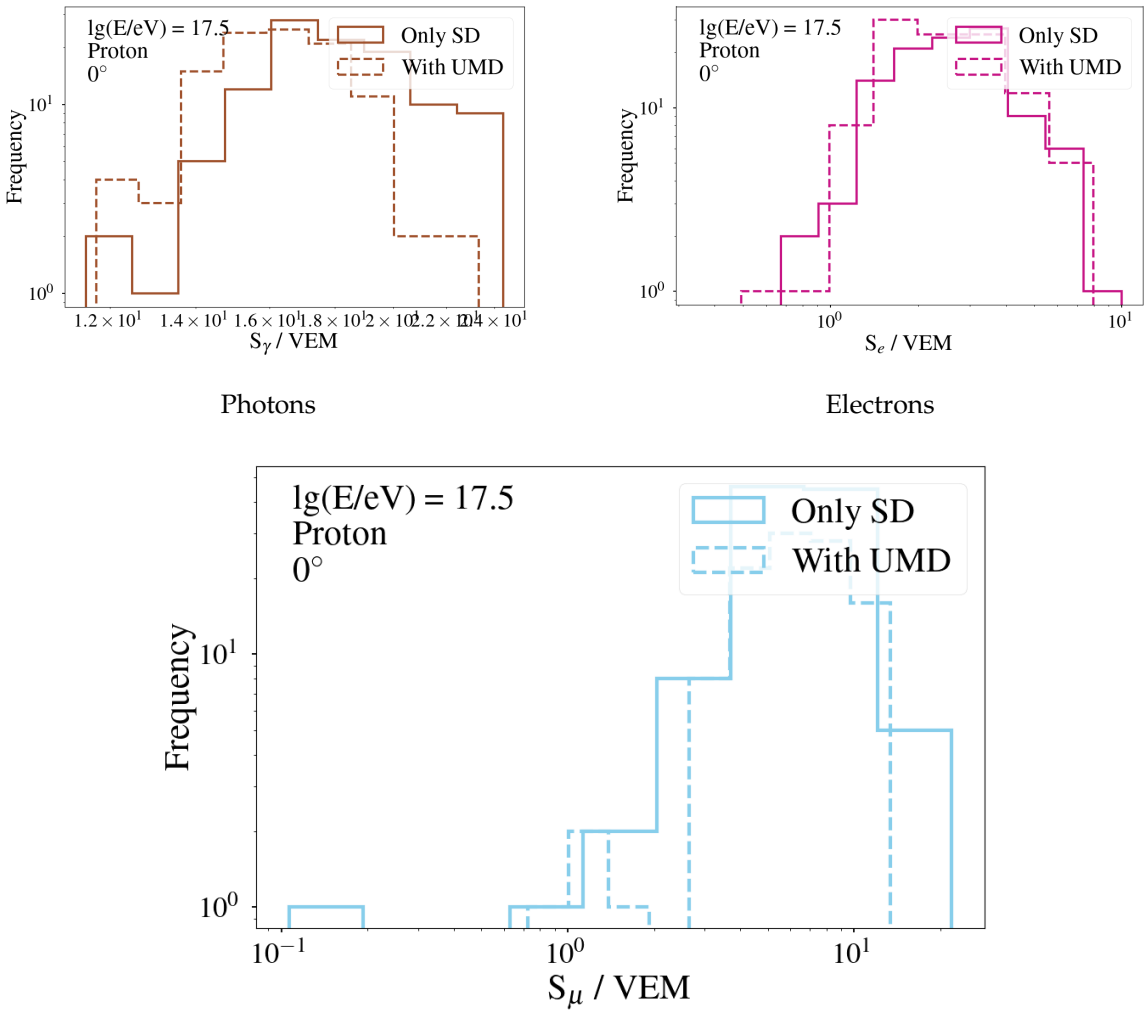


Figure 7.4: Integrated signals in the WCDs for photons, electrons, and muons. Solid lines show simulations using only the WCD, while dashed lines include both the WCD and the UMD.

as illustrated in Figure 7.6.

The fit parameters k_0 (offset) and k_1 (slope) were determined in each energy and zenith bin. Both exhibit a marked dependence on the zenith normalization s, as shown in Figure 7.7.

The main features of these parameters can be summarized as follows:

- k_0: Compared to the calibration using the true S_μ, k_0 shows a stronger dependence on the zenith angle and takes on positive values. This behavior is likely related to a residual composition bias in $\hat{S}_\mu$: within each bin, lighter primaries (e.g. protons) tend to yield lower ρ^{og} values, while heavier ones (e.g. iron) correspond to higher ρ^{og}. The resulting mixture slightly shifts the intercept of the

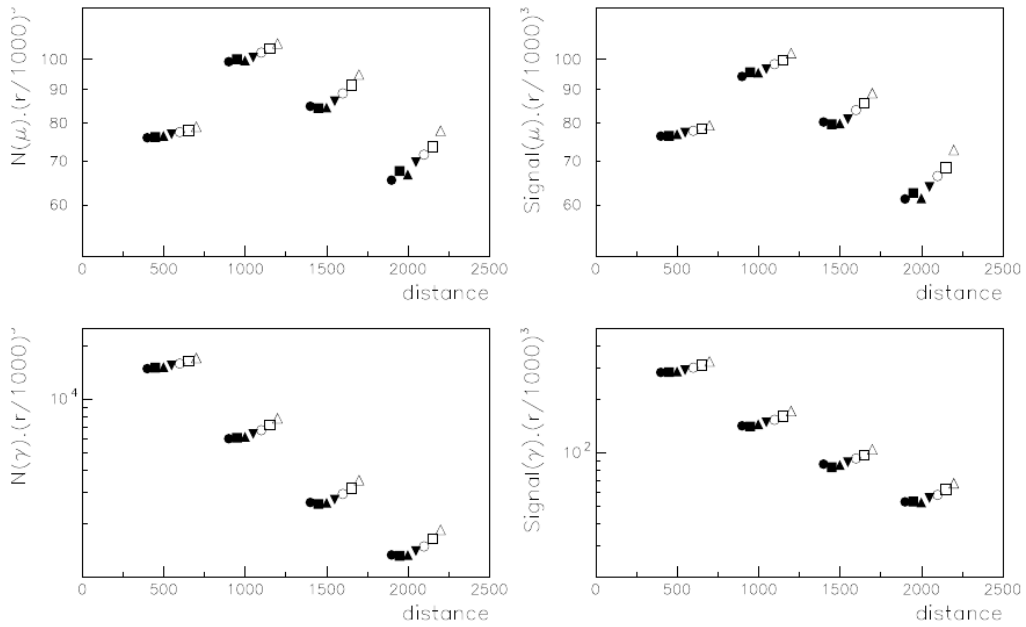


Figure 7.5: Bias in the number of muons and photons, as well as in the integrated signals, as a function of the sampling zone radius (δ = 0.025–0.3). Each column corresponds to a different distance from the shower core (500, 1000, 1500, and 2000 m). Figure adapted from [79].

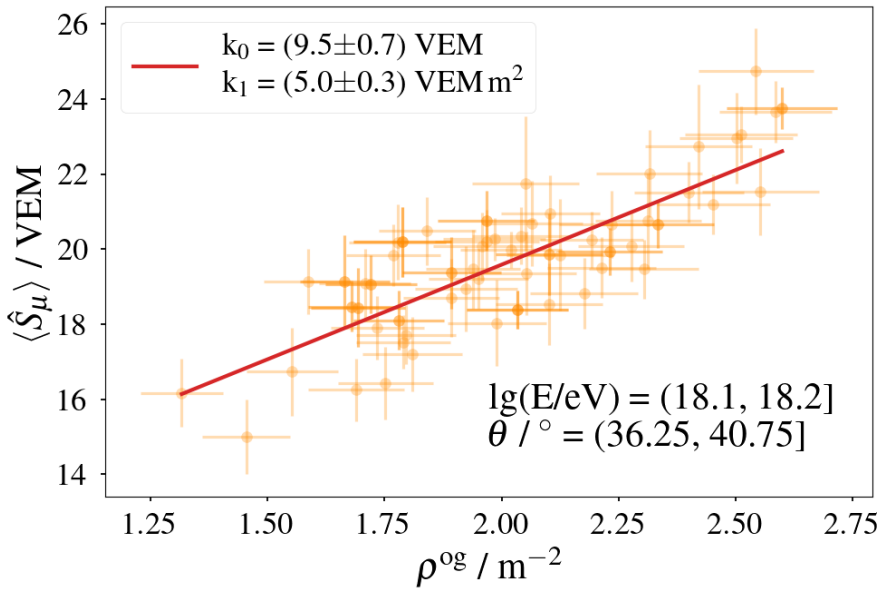


Figure 7.6: Example of the linear correlation between $\hat{S}_\mu$ estimated with NNs and the true Monte Carlo ρ^{og} for a selected energy and zenith bin in simulations.

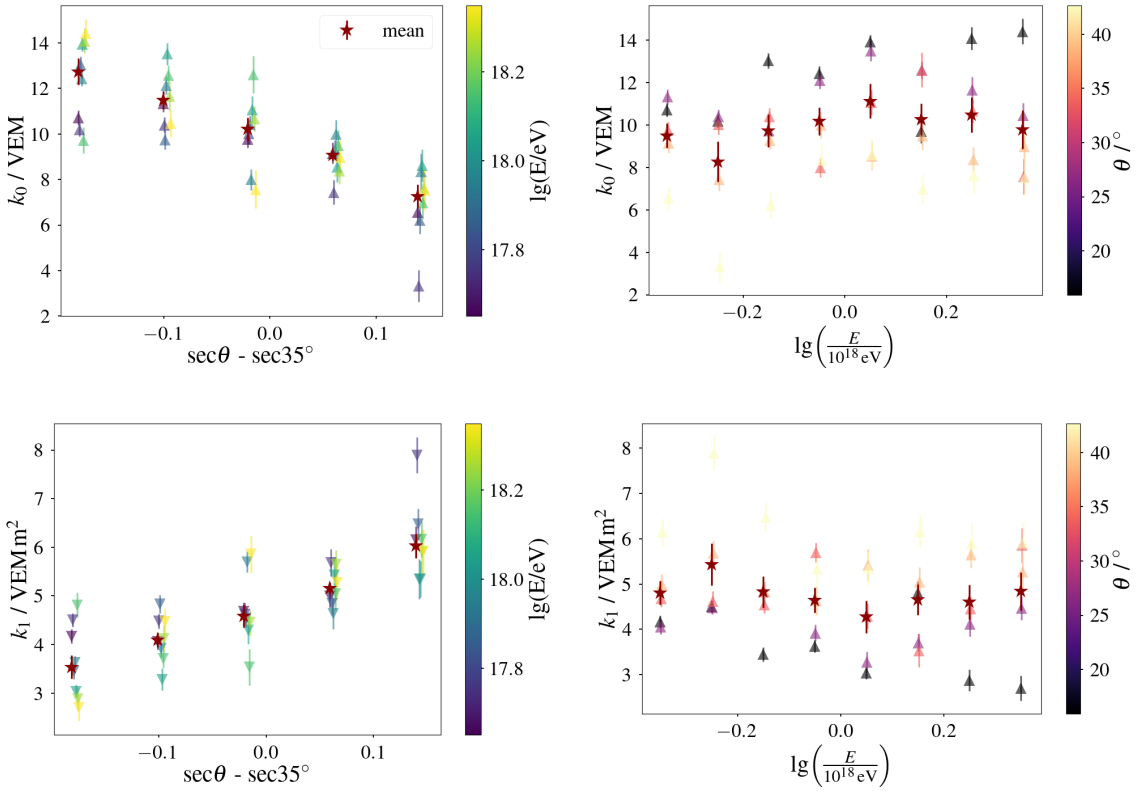


Figure 7.7: Parameters k_0 (offset) and k_1 (slope) as functions of s and e. Top: k_0 versus s (left) and e (right). Bottom: k_1 versus s (left) and e (right). Points are color–coded by energy and zenith angle, respectively.

fit.

- k_1: Although the absolute values of k_1 are smaller than those obtained with the true S_μ, its overall trend with zenith angle remains consistent, indicating that the estimator captures the correct physical dependence.

Despite the discrepancy reflected in the positive k_0 values, the estimator $\hat{S}_\mu$ preserves a strong and stable correlation with the true on–ground muon density. The resulting linear calibration remains reliable and physically meaningful, demonstrating that $\hat{S}_\mu$ provides a robust and effective proxy for ρ^{og} even under the current network limitations.

7.3. Calibration using $\hat{S}_\mu$

The final calibration was performed using the two estimators $\hat{S}_\mu$ and $\hat{\rho}^{\mathrm{og}}$. Figure 7.8 shows an example of the calibration using both estimators.

The correlation between the estimators is preserved when calibrating against $\hat{\rho}^{\mathrm{og}}$. Therefore, the linear parameters k_0 and k_1 were also extracted for each energy and

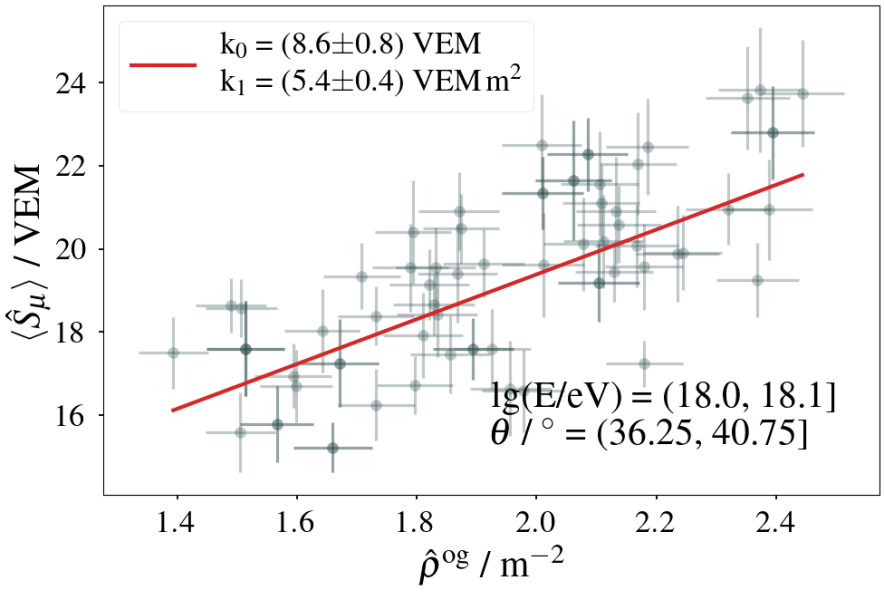


Figure 7.8: Example of the linear correlation between $\hat{S}_\mu$ from NNs and $\hat{\rho}^{\mathrm{og}}$ estimated with the method described in Chapter5 for a selected energy and zenith angle bin in simulations.

zenith bin. Their dependence on energy and zenith angle is shown in Figure 7.9.

Once the values of k_0 and k_1 were obtained across the full energy and zenith range, they were parameterized as linear functions s as before. The resulting expressions are:

$$k_0(s) = l_0 + l_1\, s \tag{7.1}$$

$$k_1(s) = m_0 + m_1\, s \tag{7.2}$$

The best-fit parameters are listed in Table 7.1.

Table 7.1: Calibration parameters given in Eqs. 7.1 and 7.2.

Parameter	Value
l_0	(9.1 ± 3.5) VEM
l_1	(-18.1 ± 3.5) VEM
m_0	(5.1 ± 0.2) VEM m^{-2}
m_1	(7.8 ± 1.5) VEM m^{-2}

These results yield a final calibration function to estimate ρ^{og} from $\hat{S}_\mu$:

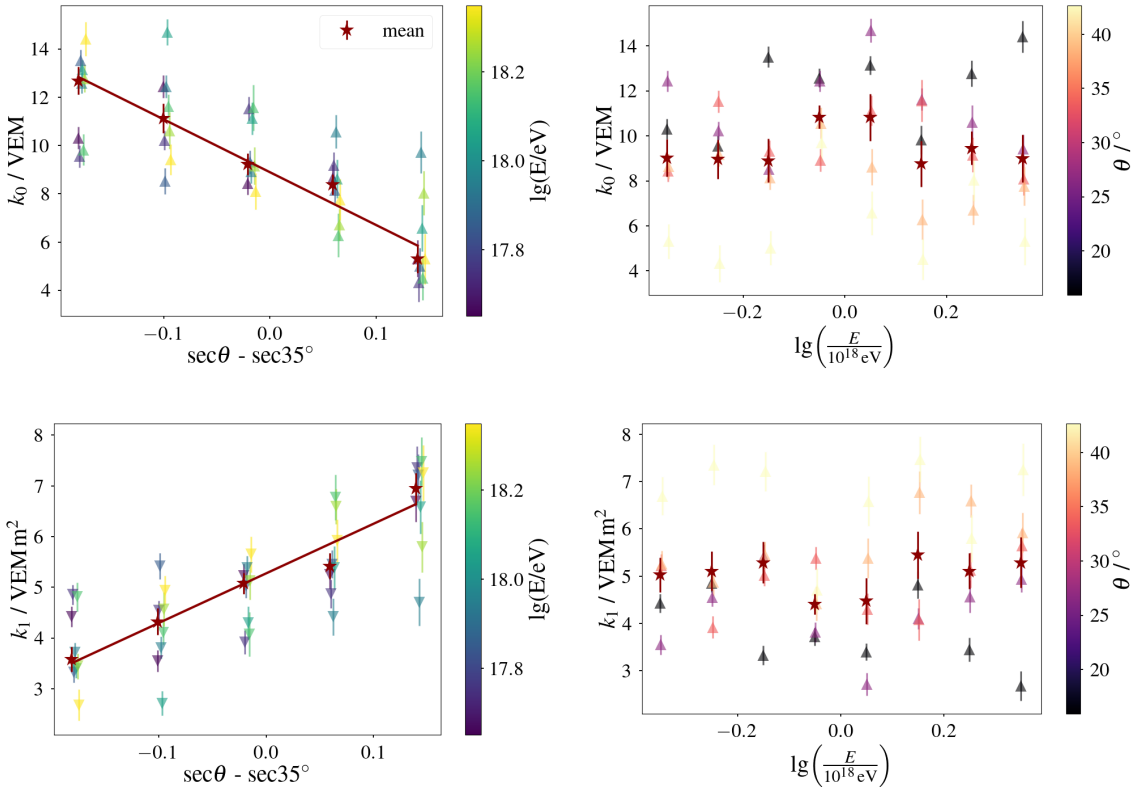


Figure 7.9: Parameters k_0 (offset) and k_1 (slope) as functions of s and e. Top: k_0 versus s (left) and e (right). Bottom: k_1 versus s (left) and e (right). Points are color–coded by energy and zenith angle, respectively.

$$\hat{\rho}^{\mathrm{og}}(\hat{S}_\mu, s) = \frac{\hat{S}_\mu - k_0(s)}{k_1(s)} \tag{7.3}$$

As a final step, the bias and resolution of this estimator were evaluated across the entire energy and zenith range, as shown in Figure 7.10.

As can be seen, the bias is largely independent of the zenith angle, with a mild energy dependence within a All data points lie within the ±5% band once their statistical uncertainties are considered, with the exception of a single outlier. the resolution remains below 20%, reaching values as low as 6% at the highest energies and zenith angles.

These results demonstrate that, given a muon signal estimator with sufficient resolution, it is possible to perform a reliable calibration using ground-level muon density obtained from direct UMD measurements, resulting in an unbiased estimate of surface muon density.

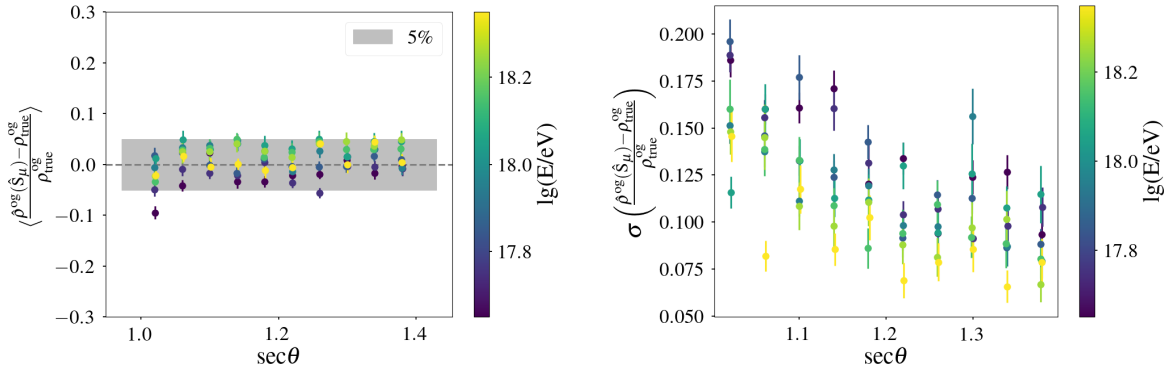


Figure 7.10: Bias (left) and resolution (right) of the estimator $\hat{\rho}^{\mathrm{og}}(\hat{S}_\mu, s)$ across the full energy and zenith range.

7.4. Summary and Conclusions

This chapter focused on establishing a reliable calibration method, using only Monte Carlo values, to convert the output of a neural network muon signal estimator, denoted as $\hat{S}_\mu$, into a physically meaningful estimate of the ground-level muon density, ρ^{og}. It was performed using the estimated muon density on-ground $\hat{\rho}^{\mathrm{og}}$ that relies on the direct measurements of muon density from the UMD

As done before with the true value of S_μ, the calibration approach involved fitting a linear model between $\hat{S}_\mu$ and $\hat{\rho}^{\mathrm{og}}$ within each bin of energy and zenith angle. The parameters of this model (the slope and offset) were found to vary systematically with energy and zenith. To enable a calibration across the full phase space, the slope and offset parameters were subsequently modeled as linear functions of the secant of the zenith angle. This provided a compact and flexible functional form that captures the main dependencies while avoiding overfitting.

After defining the global calibration function, its performance was assessed in terms of bias and resolution across the full energy and zenith range. The results indicate that the calibrated estimator is largely unbiased, with residuals close to zero, and zenith independent, with a soft energy dependence within a 5%. The resolution of the estimator, defined as the relative spread between the predicted and true values, remains below 20% throughout the parameter space and reaches values as low as 6% at higher energies.

These findings confirm that the neural network output, contains sufficient physical information to allow for a robust reconstruction of the surface muon density. The success of this calibration demonstrates the utility of combining machine learning techniques with physical detector constraints and motivates their use in future studies.

Overall, the calibrated estimator offers a practical and accurate tool for analyzing muon content in extensive air showers using the 750 m array of the Pierre Auger Observatory and for validating estimations of muon content by means of the SD.

Chapter VIII

Estimation of muon content on ground with Auger Data

In the previous chapter, parameterizations of the muon density at ground level derived from UMD measurements, $\hat{\rho}^{\mathrm{og}}$, and of the muonic signal in the WCDs, $\hat{S}_{\mu}$, were tested. In this final chapter, these parameterizations are applied to a selected subset of data acquired by the Pierre Auger Observatory. Both observables, $\hat{\rho}^{\mathrm{og}}$ and $\hat{S}_{\mu}$, are evaluated at a fixed core distance of 450 m, corresponding to the optimal reference distance for the 750 m SD array.

A central motivation for validating the SD-based estimators of the muon content through direct underground measurements is the long-standing discrepancy between simulations and observations in the muon component of extensive air showers. This comparison provides a direct test of whether the reconstructed muon densities at ground level are consistent with the independent information provided by the UMD.

As discussed in Chapters 4 and 5, the energy distribution of muons arriving at the surface depends strongly on the zenith angle, owing to the increasing atmospheric depth traversed by more inclined showers. By comparing the estimated $\hat{\rho}^{\mathrm{og}}$ in data with both simulations and underground measurements ρ^{ug} as a function of zenith angle, it becomes possible to investigate whether the observed discrepancies depend on the characteristic energy of the muons. A zenith-dependent deviation would indicate that the so-called Muon Puzzle does not uniformly affect the entire muon energy spectrum.

Finally, accounting for this possible dependence, the muon content measured in the data is compared with results from other experiments using the z-scale normalization, which provides a consistent framework for inter-experiment comparisons.

8.1. Data Quality

To ensure data quality, a selection of events was performed following the protocol established within the Pierre Auger Collaboration [80]. This procedure relies on

low-level detector observables to assess the proper functioning of the UMD in the field, ensuring that only high-quality events are included in the analysis.

The proper operation of each individual UMD module must be continuously verified. Dedicated tools and routines were therefore developed to monitor the performance of the modules on a regular basis. In this context, a new monitoring system was designed and implemented to provide a systematic module-wise evaluation of the UMD performance.

The Pierre Auger Observatory has established several procedures to monitor the performance of its various detector systems, which are crucial to maintaining optimal scientific operation. In the case of the UMD, specific monitoring tools were created to track the stability and efficiency of the individual detector modules. Two trace sensitive module-level variables, *trigger peak* and *intensity*, were introduced to allow long-term performance studies and identify possible changes in detector response over time.

- **Trigger peaks:** For each type of trigger of the WCD, as described in [81], this corresponds to the position of the time bin in the UMD signal buffer where the maximum signal is recorded.
- **Intensity:** Defined as the average length of positive traces (i.e., sequences of consecutive "1s") across all channels, see Figure 3.9.

Figure 8.1 shows a histogram of the number of 1s per time bin for a module operating in binary mode. From this histogram, the *trigger peaks* are identified as the maxima of each trace.

To model the signal trace shown in Figure 8.1, each peak was fitted using the following function:

$$f(x) = \frac{A_1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma_1^2}} + \frac{A_2}{\sigma_2\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma_2^2}} + \begin{cases} \dfrac{c}{\mu}, & \text{if } x < \mu, \\ \dfrac{d}{2048-\mu}, & \text{if } x > \mu, \end{cases} \tag{8.1}$$

where the first two terms correspond to Gaussian components that model the event region, and the piecewise terms describe the background before and after the peak.

The signal exhibits approximately Gaussian behaviour in the event region and a nearly uniform distribution in the background. A single Gaussian was insufficient to reproduce the transition between the event and the background; therefore, a second Gaussian with the same mean was introduced to better describe the base of the event peak.

The free parameters of the model are A_1, A_2, σ_1, σ_2, μ, c, and d. The parameters A_1 and A_2 represent the amplitudes of the two Gaussian components, while σ_1 and σ_2 are their respective standard deviations, defining the width of each contribution. The parameter μ denotes the position of the event peak, common to both Gaussian components. The coefficients c and d correspond to the background amplitudes before and after the event peak, respectively.

The background regions display a mostly uniform shape; however, in some modules the background amplitude slightly increases after the event peak (see Figure 8.2).

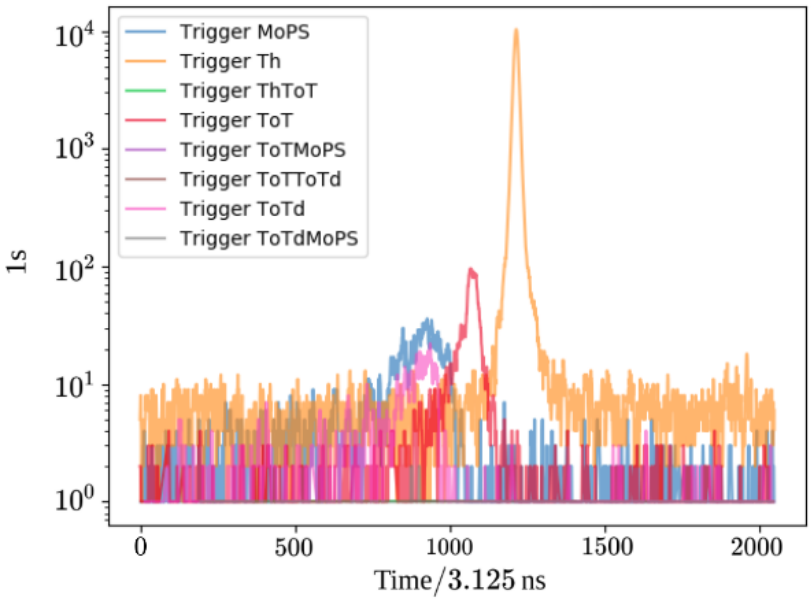


Figure 8.1: Histogram of 1s per time bin in the buffer for a module operating in binary mode. Each color corresponds to a different trigger type, following the trigger hierarchy. The regions near the peaks correspond to air-shower events, while the tails represent background signals. The differences in peak positions arise from the varying times each trigger takes to be generated in the WCD. Since UMD stations use circular buffers, slower triggers result in more accumulated signal and cause the corresponding peaks to appear sooner in the trace. Data correspond to the Kathy-Turner station, module 101, for March 2021.

To account for this, two independent background amplitudes were introduced. Each UMD module is fitted individually, as both the event peak and the background parameters are module-specific.

The position of each trigger peak remains approximately constant over time for a given module and trigger type. Among them, the Threshold trigger peak position is particularly relevant, as it reflects detector performance and is essential for calibration procedures [43]. Figure 8.3 shows the Threshold peak position for all modules across three months in 2021 (February, May, and December), demonstrating the parameter's stability. The peak position fluctuates within time-bin values of approximately 1170 to 1225.

The background signal is defined as the uniform portion of the histogram, corresponding to time-bin ranges 0–500 and 1500–2048, outside the peak regions. These peak regions are identified as $\mu \pm \sigma$ for each trigger. The background level is computed as the sum of the amplitudes of the two uniform components, normalized to the total signal. The *background percentage* is defined as the ratio of the uniform amplitudes to the total (Gaussian plus uniform) contribution. Figure 8.4 shows the background percentage for all modules across three months in 2021. Values typically range from 4% to 15%. No modules were found with background percentages outside this band.

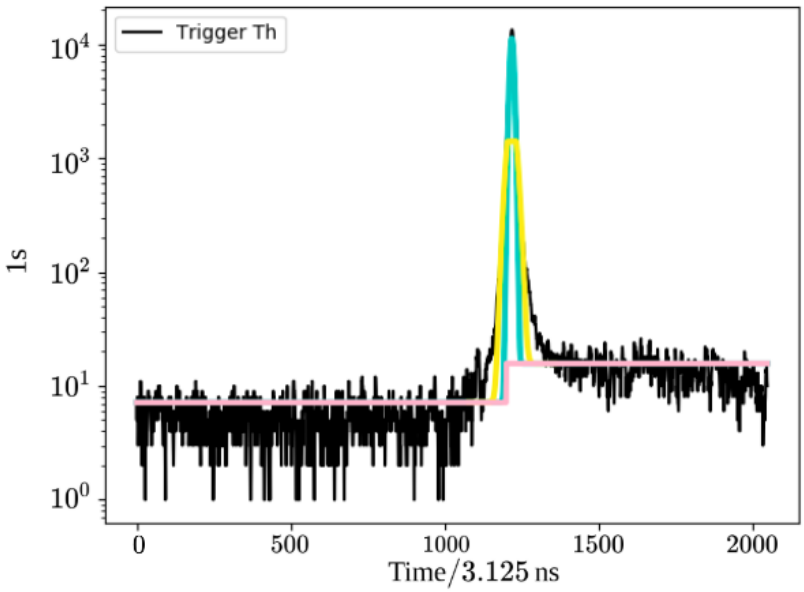


Figure 8.2: Black: events with Threshold trigger only. Colored lines: model fit as described in Eq. 8.1. Pink: background component; blue and yellow: the two Gaussian components with different amplitudes but the same mean. For this fit, the total background amplitude corresponds to approximately 7%, less than half of the 14.8% predicted from laboratory measurements with single muons [43].

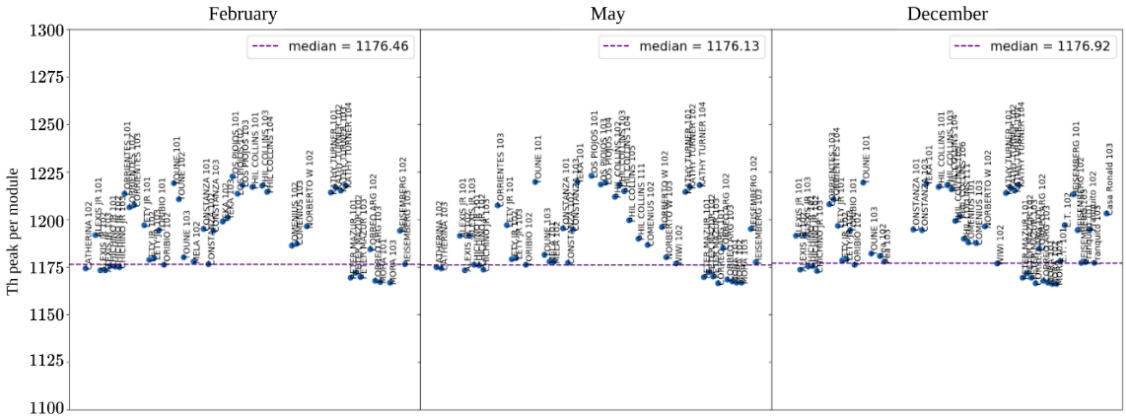


Figure 8.3: Threshold trigger peak position per module for three different months: February, May, and December 2021.

To detect anomalies in the background regions, an outlier detection method was implemented. If a signal exceeds the mean background level by more than 5σ, where σ is the standard deviation of the background, an alert is logged by the monitoring tool. These outliers may indicate excess noise outside the expected signal region and warrant further inspection. An example is shown in Figure 8.5.

Following the definition in [82], the intensity is defined as the mean number of consecutive 1s across all channels of each module. This variable is particularly

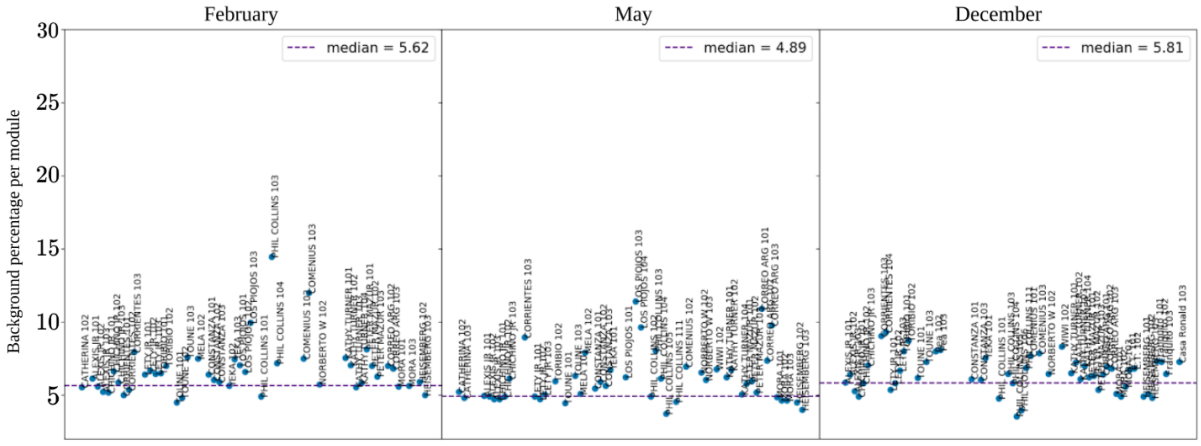


Figure 8.4: Background percentage per module for three different months: February, May, and December 2021.

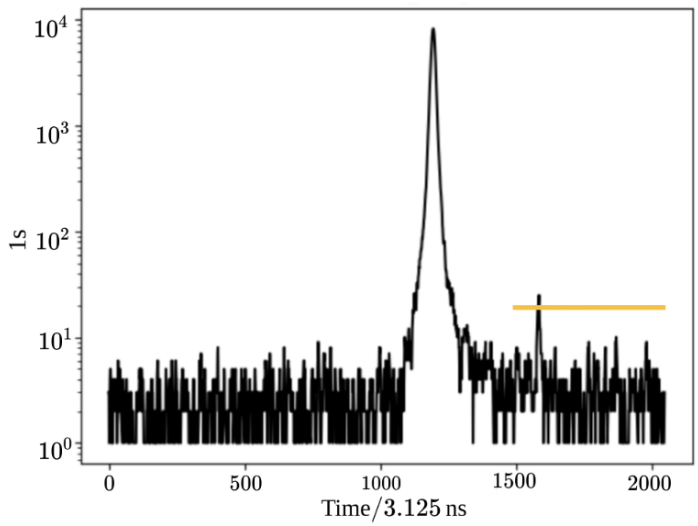


Figure 8.5: Black: events with Threshold trigger only. Any signal above the yellow line is flagged as an outlier.

relevant, as the muon-counting strategy directly relies on it. Laboratory studies indicate that the mean intensity for a single-muon trace is approximately 7.8, with a dispersion of 1.5 [43].

Three main observations can be made from Figure 8.6. First, the mean intensity aligns with the expected value from laboratory measurements of single muons. Second, the dispersion observed in measurements is roughly twice the laboratory value, where only single muons were considered. Third, an accumulation of short traces can be seen on the left side of the plot, corresponding to the background baseline.

The mean and standard deviation of the intensity per module are shown in Figures 8.7 and 8.8. Some particularly noisy modules are identified and accounted

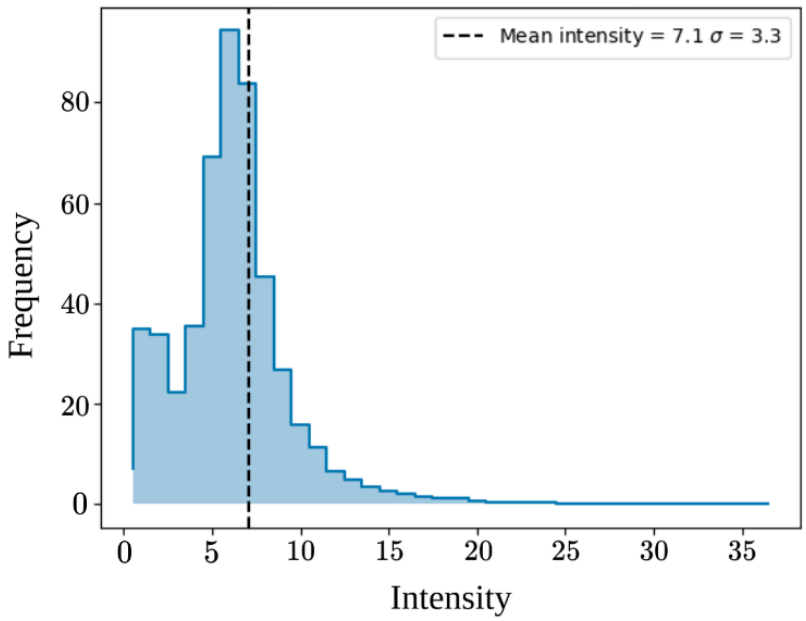


Figure 8.6: Histogram of mean intensity across all channels for one month of data from station Kathy-Turner m101.

for in the monitoring protocols.

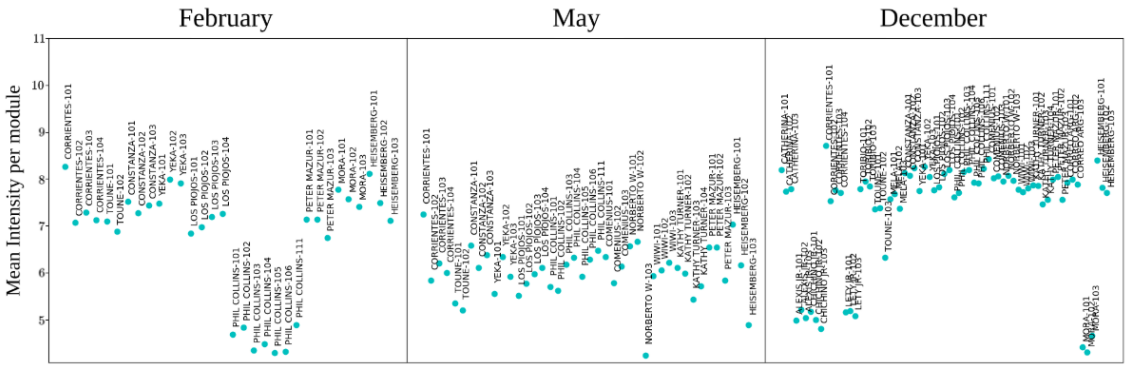


Figure 8.7: Mean intensity per module for three different months: February, May, and December 2021.

Monitoring the mean intensity serves as a complementary diagnostic to the monitoring of the Threshold (Th) peak position. In some cases, poor detector performance may not be revealed by the Th peak alone. For instance, Figure 8.9 shows the histograms of ones per bin and intensity for station Alexis Jr. m101.

As seen in Figure 8.3, a Threshold trigger peak value of around 1200 is expected. Based solely on this metric, the module in Figure 8.9 would appear to function normally. However, the low mean intensity observed indicates otherwise. This behavior suggests an increased presence of short traces, which can compromise the effectiveness of the muon-counting strategy.

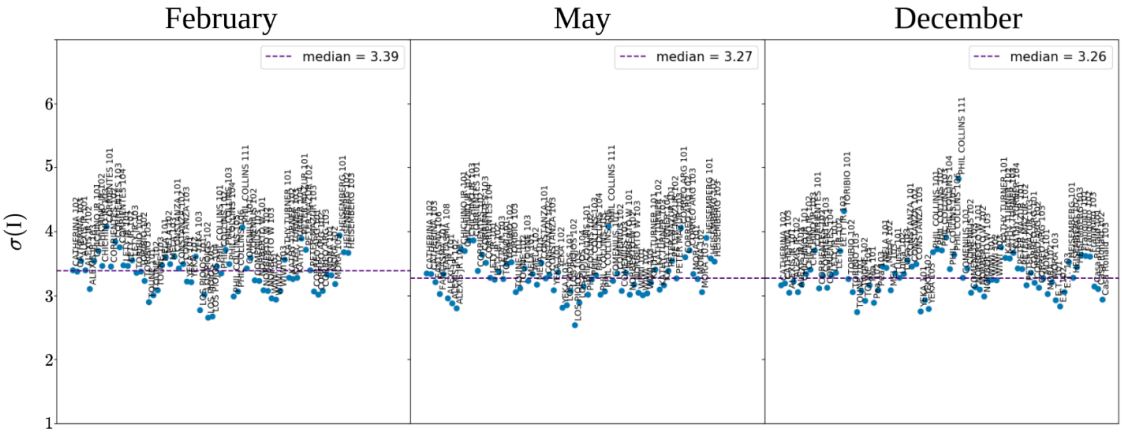


Figure 8.8: Standard deviation $\sigma(I)$ of intensity per module for three different months: February, May, and December 2021.

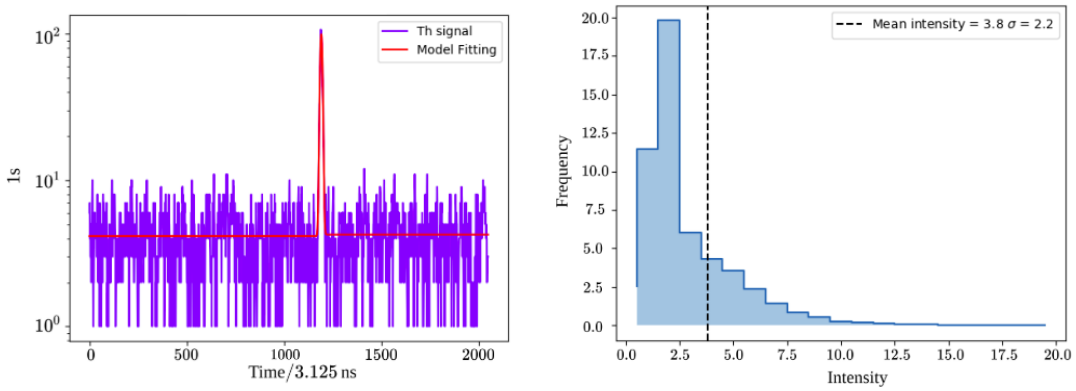


Figure 8.9: Left: histogram of ones per bin showing a Threshold trigger peak near the expected value of 1200. Right: histogram of intensities showing a lower-than-expected mean value of 3.8. Data correspond to December 2020.

To identify outliers, a station is considered anomalous if its mean intensity deviates by more than one standard deviation from the distribution of means across all stations, as illustrated in Figure 8.7.

8.2. Data Selection

This section outlines the selection and preparation of the dataset used for the physics analyses. The data corresponds to the *Phase I* of the observatory, meaning that data acquired by stations prior to the UUB electronics upgrade were considered. Event reconstruction was performed using the official Offline framework for both the WCD and UMD stations within the 750 m array.

Since the UMD trigger depends on the corresponding SD station trigger, all selected events are hybrid UMD-SD events.

It is important to note that the reconstruction process automatically excludes

problematic periods or malfunctioning stations ("bad periods"), as described in [83].

The analysis was restricted to events with reconstructed energies in the range $\lg(E/\mathrm{eV}) \in [17.5, 18.5]$ and zenith angles below 45°. Only showers satisfying the 6T5 trigger condition were selected, which requires that the station with the highest signal has six active neighbors. This condition ensures that the shower core is well contained within the array and not located near its boundaries or inactive regions [81]. After applying all quality criteria described below, a total of 8,041 events were retained for the analysis.

These criteria were chosen to match the phase space in which the muon-on-ground parameterization and the corresponding calibration from previous chapters were developed.

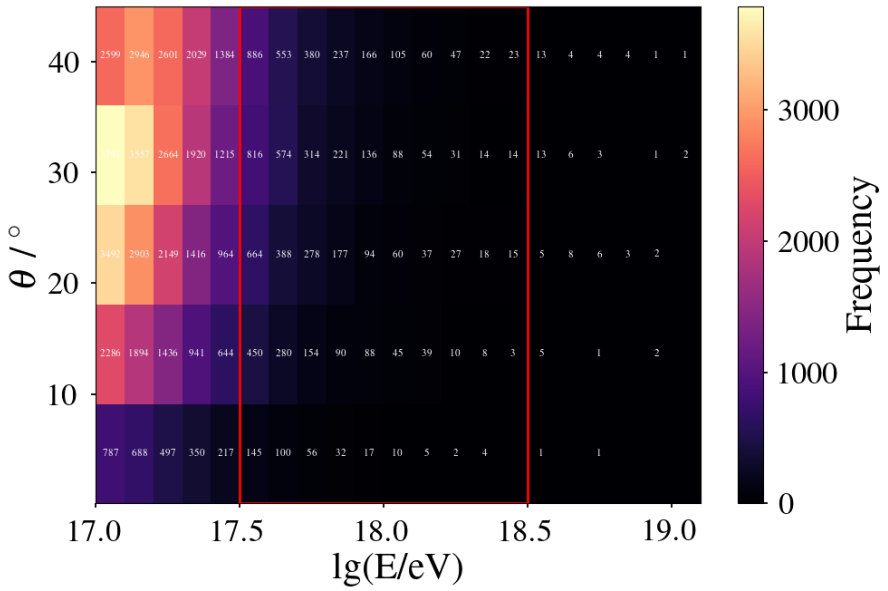


Figure 8.10: Number of reconstructed events as a function of logarithmic energy and zenith angle. The selected sample is highlighted in the red rectangle.

A second selection was then applied at the station level: Only stations with a reconstructed signal located between 400 and 500 m from the shower axis were used. This range was chosen to match the reference distance of 450 m used for parameterization and calibration in simulations. After this cut, a total of **2529 stations** were retained for the analysis.

8.3. Angular Dependence of the Muon Discrepancy

Since the on-ground muon density estimator presented in Chapter 5 was developed using simulations, it is necessary to verify whether the muon energy distribution is consistent between data and simulations before applying the method to the selected dataset. This verification was carried out using the underground muon density

reconstructed by the UMD as a proxy for the on-ground muon density, which is not directly measured by the WCD.

As discussed in Chapter 4, the mean muon energy at ground level strongly depends on the zenith angle of the primary particle. With increasing zenith angle, the shower front traverses a larger atmospheric depth, causing low-energy muons to lose all their energy before reaching the ground. Although a dependence on primary energy is also present, it is significantly weaker than the zenith dependence.

Section 2.6 summarized the long-standing discrepancy between the number of muons predicted by simulations and those measured in data. However, it remains unclear whether this discrepancy is uniform across the entire muon-energy spectrum, since muon detection techniques always involve an energy threshold imposed by the shielding used to filter out the electromagnetic component of the EAS.

As shown in Chapter 4, the muon detection threshold for the UMD approximately satisfies

$$E_\mu \simeq 1\ \mathrm{GeV}/\cos\theta_\mu, \tag{8.2}$$

where θ_μ is the incidence angle of the muon in the soil.

To quantify the correlation between zenith angle and muon energy, the energy distribution of muons was studied in proton-initiated showers with fixed primary energy of 10^{18} eV, selecting those arriving at the surface within 440–460 m from the shower axis. The left panel of Figure 8.11 shows the E_μ distributions for five fixed primary zenith angles, while the right panel displays the mean E_μ as a function of θ.

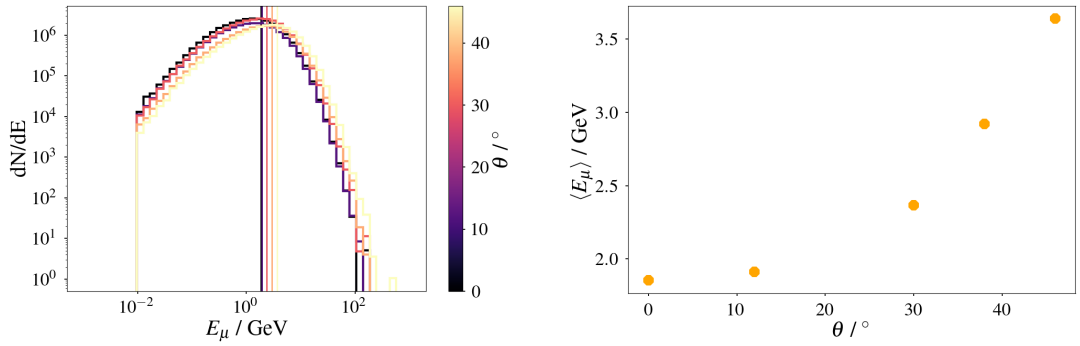


Figure 8.11: Left: E_μ distributions for different fixed primary zenith angles (proton showers), for muons arriving at ground within 440–460 m of the shower core. Right: mean E_μ as a function of θ.

From Figure 8.11 (right), an increase of approximately 50% in the mean muon energy is observed between 0° and 46°, matching the zenith range of the selected dataset.

Because the underground muon density depends not only on E_μ but also on the incidence angle through Eq. (8.2), the ratio between the number of muons above threshold and the total number of muons at the surface was also evaluated as a function of θ (Figure 8.12). This ratio increases by about 18% between 0° and 46°, emphasizing the role of the incidence angle.

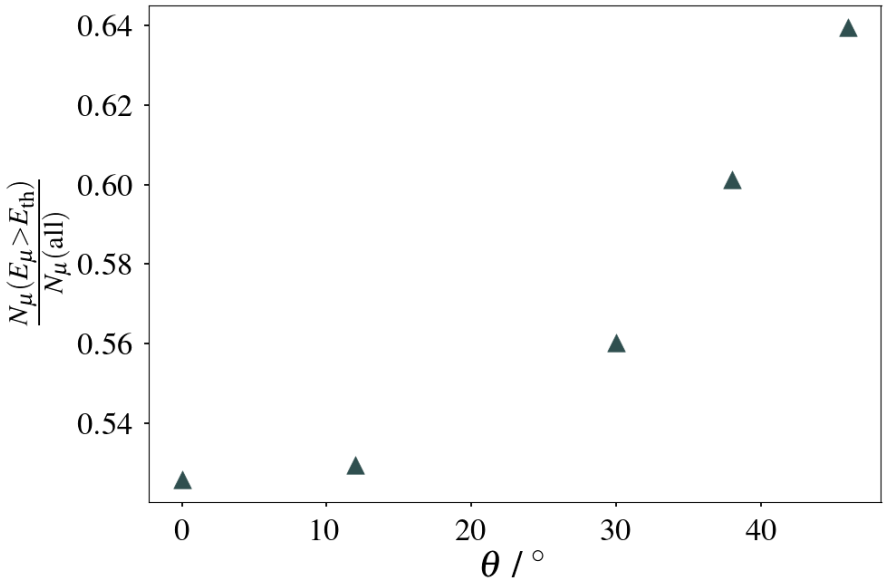


Figure 8.12: Ratio of muons reaching the UMD to the total number of muons on ground as a function of θ.

Given these trends, the energy distribution of muons above 1 GeV—which dominate the input to the on-ground muon density parameterization—was tested for consistency between data and simulations across the full zenith range. Specifically, the ratio

$$\frac{\langle \rho_{\rm ug}^{\rm data} \rangle}{\langle \rho_{\rm ug}^{\rm sim} \rangle} \tag{8.3}$$

was examined as a function of $\sec\theta$ for three hadronic interaction models: EPOS-LHC, EPOS-LHC R, and SIBYLL 2.3e. Weighted linear regressions of the ratio versus $\sec\theta$ were performed, and the slope significance was evaluated using Student's t-test with the corresponding p-value. The results are listed in Table 8.1, and the fitted curves are shown in Figure 8.13.

Table 8.1: Slope fits and corresponding statistical tests for different hadronic interaction models.

	EPOS-LHC	EPOS-LHC R	SIBYLL 2.3e
Slope (± stat.)	-0.13 ± 0.08	-0.07 ± 0.07	-0.10 ± 0.08
t statistic	-1.63	-0.99	-1.34
p-value	0.14	0.35	0.22

Using the conventional criterion ($p > 0.05$) for compatibility with a zero slope, the ratios for all three models are consistent with being constant across the full zenith

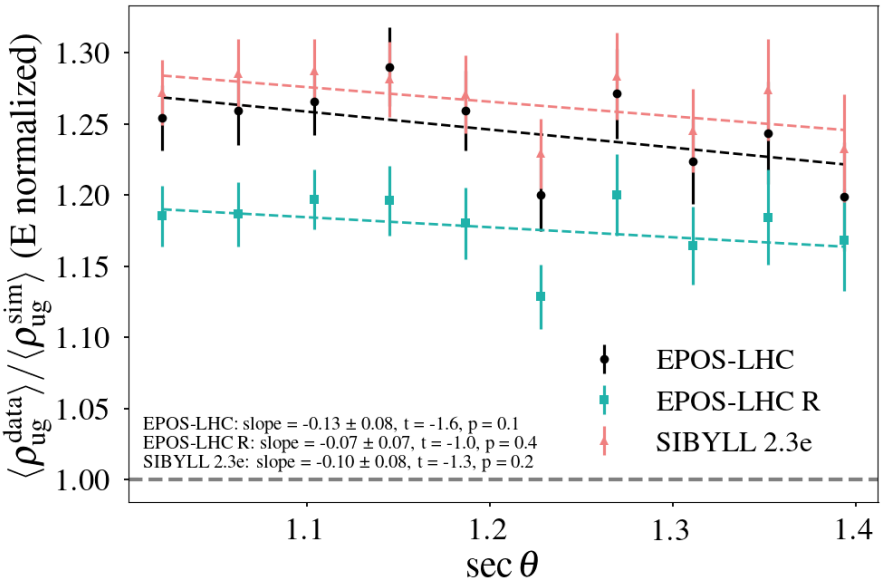


Figure 8.13: Ratio of underground muon densities in data to those in simulations (Auger-Mix) as a function of sec θ, with linear fits. The energy normalization of each muon density follows the same procedure described in Chapter 5.

range. This result indicates no significant data–simulation discrepancy in the energy distribution of muons above 1 GeV within the studied zenith interval. Among the models, EPOS-LHC R yields the smallest absolute slope and t value and the largest p-value, suggesting the closest agreement with the data.

Although all curves in Figure 8.13 exhibit an offset, indicating that more muons are measured in data than predicted by simulations, the constancy of the ratio with zenith angle implies that the muon deficit does not vary with muon energy above approximately 1 GeV. These findings validate the application of the on-ground muon parameterization to real data and support the robustness of the resulting on-ground muon density estimates.

8.4. Complete dataset correlation

Unlike simulations, real data do not provide truth-level (MC) quantities against which estimators can be directly validated. In addition, the persistent discrepancy between the muon content measured in data and that predicted by simulations limits the reliability of direct data–MC comparisons.

To assess the performance of the new estimators for the on-ground muon density and the muonic signal in the WCD, their behavior was evaluated in two complementary ways: (i) by studying their mutual correlation and (ii) by comparing their correlations with directly measured quantities. The direct observables considered are the underground muon density reconstructed by the UMD, ρ^{ug}, and the total

WCD signal, S_{tot}. The former provides a measurement of the muon density for muons with energies above approximately 1 GeV and serves as the physical input to the on-ground parameterization. The latter is the total signal recorded by the WCDs, which includes both electromagnetic and muonic components. The traces forming S_{tot} are the inputs to the neural networks (NNs) that estimate the muonic component $\hat{S}_{\mu}$. Since both ρ^{ug} and S_{tot} are directly measured and independent of simulation-based estimators, they offer a reliable reference for testing estimator performance.

As a first step, the correlation between the energy-normalized station-level values of S_{tot} and ρ^{ug} was calculated (Figure 8.14). For the complete dataset, the Pearson correlation coefficient is $r = 0.28$. Each point corresponds to a WCD–UMD pair located between 400 and 500 m from the shower core and is color-coded by the reconstructed zenith angle. A clear zenith dependence of S_{tot} is visible, which motivated a zenith-binned analysis.

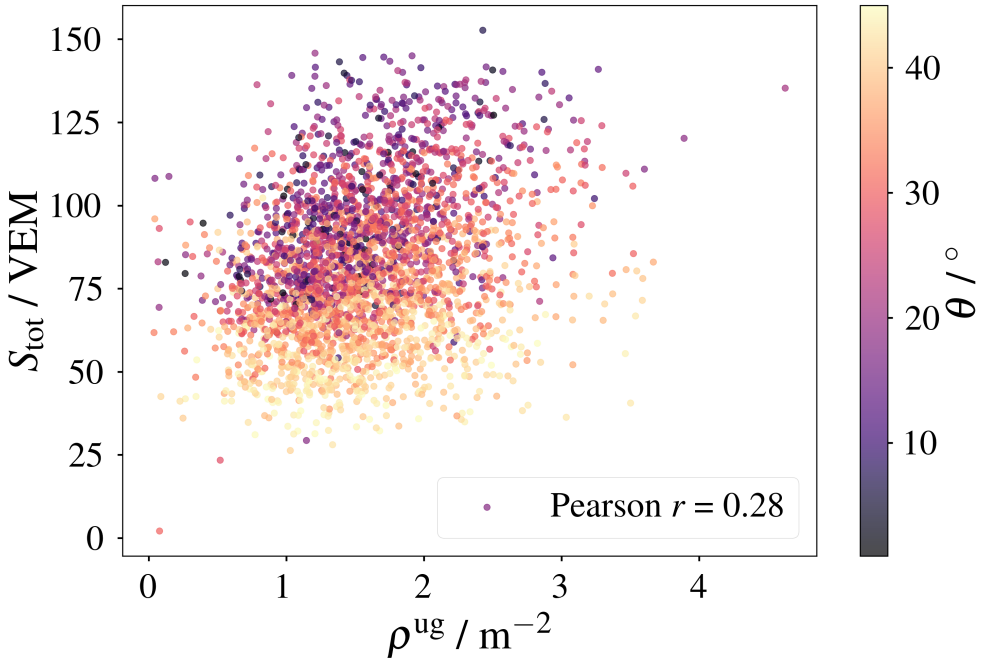


Figure 8.14: Correlation between S_{tot} and ρ^{ug} for the complete dataset, both energy-normalized as described in Chapter 5. Each point corresponds to a station located 400–500 m from the shower core and is color-coded by zenith angle. The Pearson coefficient is $r = 0.28$.

The neural networks were re-trained by the FZU group using EPOS-LHC and EPOS-LHC R simulations and subsequently applied to the data. For consistency, the on-ground muon parameterization was also derived using EPOS-LHC R. The corresponding coefficients and residual diagnostics are provided in Appendix D.

The correlations between the NN-based estimates of the muonic signal $\hat{S}_{\mu}$ and the reconstructed underground muon density ρ^{ug} are shown in Figure 8.15. The

resulting Pearson coefficients are $r = 0.30$ for EPOS-LHC and $r = 0.32$ for EPOS-LHC R. No significant zenith dependence is observed in either case. The equivalent study performed on simulated events is presented in Appendix E.

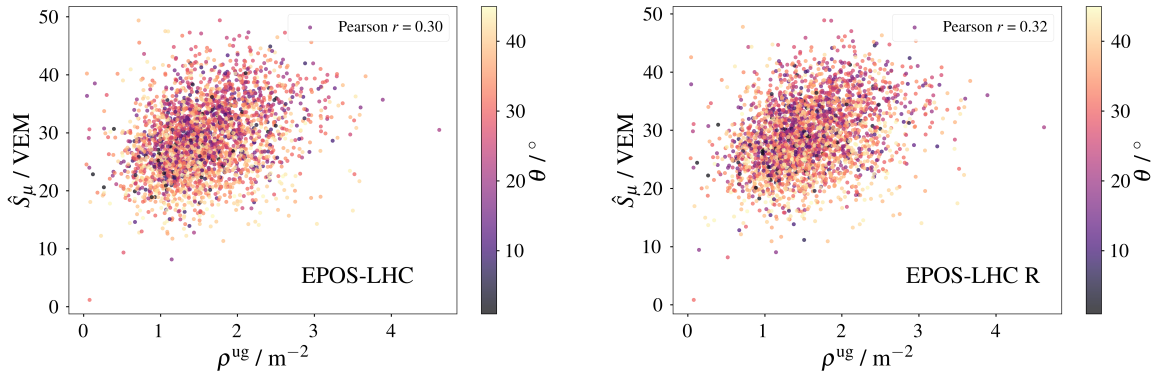


Figure 8.15: Correlation between the NN-based muonic signal $\hat{S}_\mu$ and the reconstructed underground muon density ρ^{ug} for the complete dataset. Pearson coefficients: $r = 0.30$ (EPOS-LHC, left) and $r = 0.32$ (EPOS-LHC R, right).

Because S_{tot} includes a significant electromagnetic component that decreases with atmospheric depth, correlations with ρ^{ug} were evaluated as a function of zenith angle (Figure 8.16). In these plots, data points represent station-level correlations obtained from real data, and dashed lines correspond to event-level correlations from simulations. This distinction is essential because the parameterization of $\hat{\rho}^{og}$ was developed at the event level, whereas only station-level quantities are available in the data.

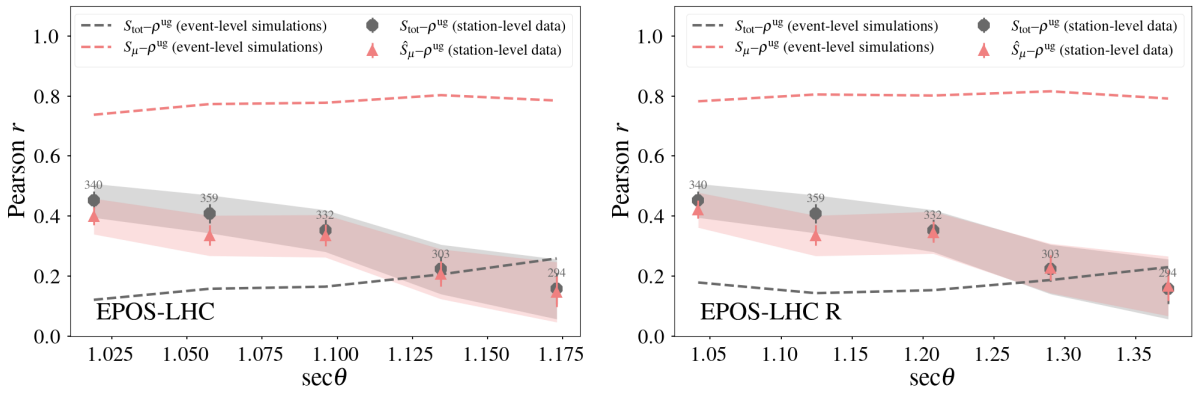


Figure 8.16: Pearson correlation versus zenith angle for S_{tot} (gray) and $\hat{S}_\mu$ (pink) with respect to ρ^{ug}. Data points correspond to station-level measurements, and dashed lines correspond to event-level correlations from simulations using EPOS-LHC (left) and EPOS-LHC R (right).

Figure 8.16 shows a clear difference in behavior between data and simulations. In simulations, the correlation with ρ^{ug} increases slightly with the zenith angle for both

S_{tot} and $\hat{S}_\mu$. In contrast, the data display a decrease in correlation from approximately $r = 0.4$ at small zenith angles to $r \approx 0.2$ at 45°. The similar evolution of S_{tot} and $\hat{S}_\mu$ is expected, since $\hat{S}_\mu$ is derived from the traces S_{tot}.

The difference between data and simulations arises mainly from the level at which the correlations are computed. In simulations, the analysis is performed at the event level, consistent with the development of the parameterization $\hat{\rho}^{\text{og}}$. In the data, only station-level quantities are available, as an event-level $\hat{S}_\mu$ reconstruction has not yet been developed. To confirm that energy normalization does not cause this difference, the same analysis was repeated using energy-unnormalized values of S_{tot} and ρ^{ug}. The results remained consistent, as discussed in the Appendix F.

Methodological effects also contribute to the observed divergence. At the station level, correlations are more sensitive to sampling fluctuations, the increasing average station radius with zenith angle. These effects reduce Pearson's r for inclined showers. Event-level observables, which are based on azimuthally averaged quantities, suppress such fluctuations and display a flatter zenith dependence.

The correlation between S_{tot} and the on-ground muon-density estimator $\hat{\rho}^{\text{og}}$ was then computed, yielding Pearson coefficients of $r = 0.43$ for EPOS-LHC and $r = 0.45$ for EPOS-LHC R (Figure 8.17).

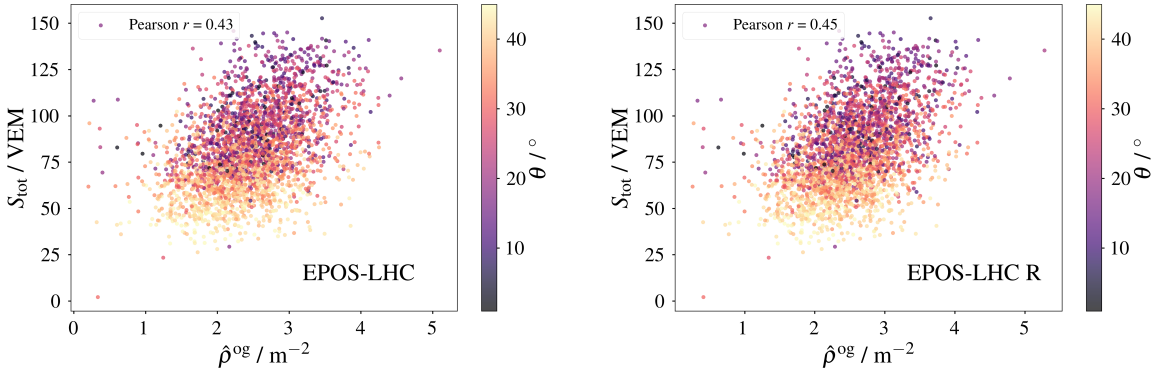


Figure 8.17: Correlation between S_{tot} and the on-ground muon-density estimator $\hat{\rho}^{\text{og}}$ for the complete dataset. Pearson coefficients: $r = 0.43$ (EPOS-LHC, left) and $r = 0.45$ (EPOS-LHC R, right).

The mutual correlation between $\hat{S}_\mu$ and $\hat{\rho}^{\text{og}}$ yields $r = 0.33$ for EPOS-LHC and $r = 0.36$ for EPOS-LHC R (Figure 8.18).

Finally, the correlations of S_{tot} and $\hat{S}_\mu$ with $\hat{\rho}^{\text{og}}$ were studied as a function of zenith angle and compared with event-level simulations (Figure 8.19). In these plots, data points correspond to station-level correlations, and dashed lines represent event-level results from simulations. This comparison allows identifying the zenith range in which both approaches are statistically compatible.

From Figure 8.19, the correlation between S_{tot} and $\hat{\rho}^{\text{og}}$ is consistent between data and simulations up to zenith angles of approximately 33° for both hadronic models. This indicates that the on-ground muon-density parameterization is reliable up to

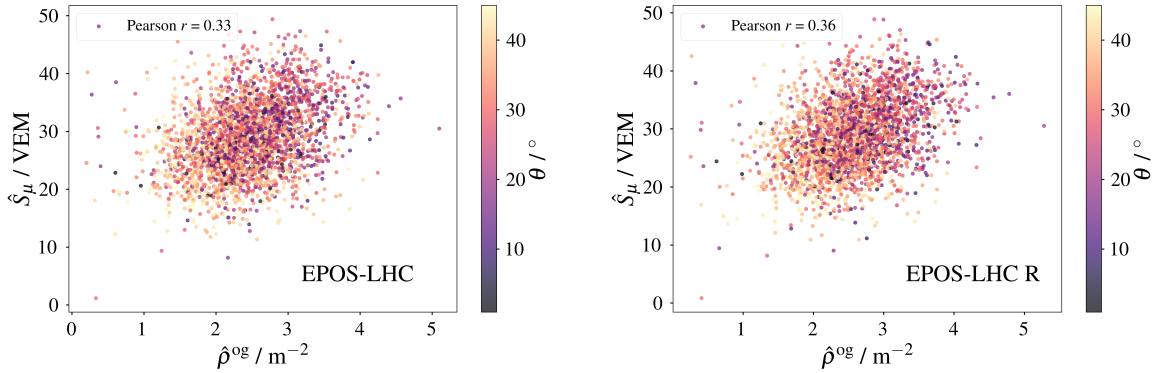


Figure 8.18: Correlation between $\hat{S}_\mu$ and $\hat{\rho}^{\mathrm{og}}$ for the complete dataset. Pearson coefficients: $r = 0.33$ (EPOS-LHC, left) and $r = 0.36$ (EPOS-LHC R, right).

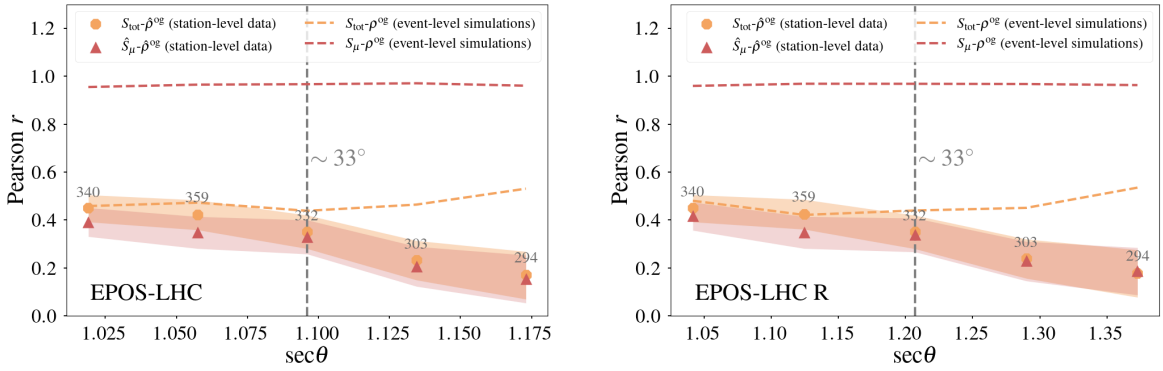


Figure 8.19: Pearson correlation versus zenith angle for S_{tot} and $\hat{S}_\mu$ with respect to $\hat{\rho}^{\mathrm{og}}$. Data points correspond to station-level measurements, and dashed lines correspond to event-level correlations from simulations using EPOS-LHC (left) and EPOS-LHC R (right).

that limit. For more inclined geometries, additional refinement is required to account for the increasing discrepancy between event-level and station-level behavior. In summary, the estimator was developed and validated at the event level using simulations, while the data analysis is necessarily performed at the station level. The comparison between both levels allows for the identification of the zenith range in which the parameterization remains applicable to real data. Since a fully developed method to estimate S_μ at the event level does not yet exist, which would, for example, require a detailed study of the S_μ LDF, this section presents a conservative method to define an angular region in which the parameterization developed to estimate ρ^{og} can be applied.

8.5. Binned calibration in data

Following the procedure described in Chapter 7, a calibration between estimators was also performed using data. Because the zenith cut at 33° reduced the available statistics to about 1400 events, and since the observables were already normalized by energy, only a binning in zenith was applied in order to retain sufficient statistics per bin. This differs from the case of simulations, where the binning was performed in both energy and zenith. For the calibration, the estimators obtained in simulations with the hadronic models EPOS-LHC and EPOS-LHC R were used as references.

Figure 8.20 shows an example of the calibration for a single zenith bin. The relation was obtained using the same linear regression function as employed in the simulations:

$$\hat{S}_\mu = k_1 \, \hat{\rho}^{\mathrm{og}} + k_0. \tag{8.4}$$

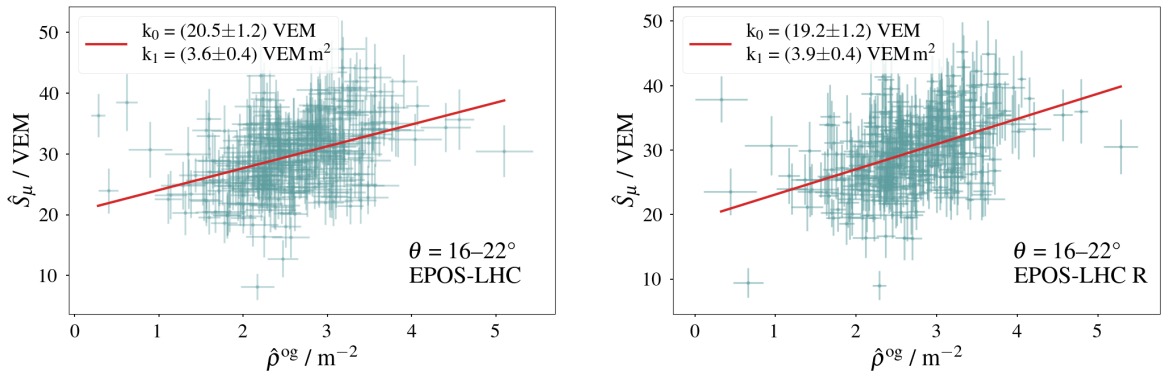


Figure 8.20: Linear calibration between $\hat{S}_\mu$ and $\hat{\rho}^{\mathrm{og}}$ for a single zenith bin, using EPOS-LHC (left) and EPOS-LHC R (right).

The parameters k_0 and k_1 were computed in each zenith bin for both hadronic models. The results are shown in Figure 8.21 for EPOS-LHC (top row) and EPOS-LHC R (bottom row).

From Figure 8.21, it can be observed that the zenith dependence of the calibration parameters, clearly visible in the simulations of Chapter 7, is not reproduced in the data. In simulations with EPOS-LHC, within the considered zenith range, the parameter k_0 remains around ~12 VEM, while in data it reaches mean values of $\langle k_0 \rangle = (20.1 \pm 0.5)$ VEM (EPOS-LHC) and $\langle k_0 \rangle = (19.8 \pm 0.6)$ VEM (EPOS-LHC R). In contrast, the slope k_1 shows good agreement: simulations yield ~4 VEM m^2, consistent with the values obtained in data, $\langle k_1 \rangle = (3.8 \pm 0.1)$ VEM m^2 (EPOS-LHC) and $\langle k_1 \rangle = (3.8 \pm 0.2)$ VEM m^2 (EPOS-LHC R). These results suggest that the relation between $\hat{S}_\mu$ and $\hat{\rho}^{\mathrm{og}}$ is consistent in both data and simulations, but that the baseline offset of the relation differs.

As a further test, the bias and resolution of the calibrated muon density were evaluated using

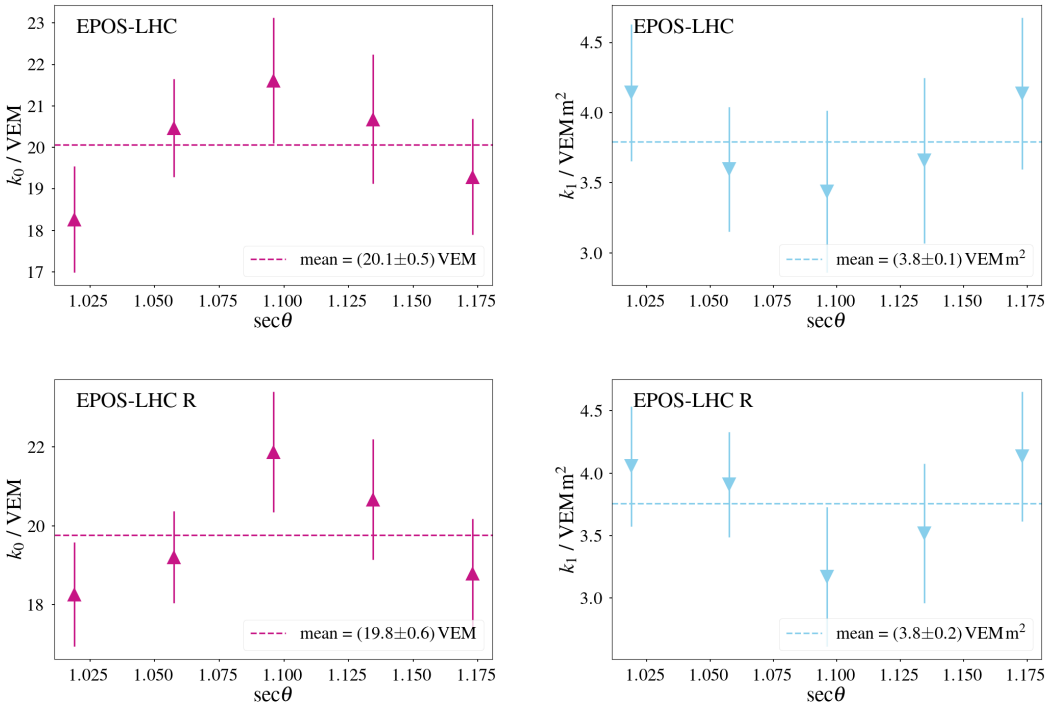


Figure 8.21: Calibration parameters obtained with EPOS-LHC (top) and EPOS-LHC R (bottom).

$$\hat{\rho}^{\mathrm{og}}(\hat{S}_\mu) = \frac{\hat{S}_\mu - k_0}{k_1}, \tag{8.5}$$

with the calibration constants taken as the mean values across the zenith range. The bias and resolution were computed as

$$\left\langle \frac{\hat{\rho}^{\mathrm{og}}(\rho^{\mathrm{ug}}, E, \theta) - \hat{\rho}(\hat{S}_\mu)}{\rho^{\mathrm{ug}}(E, \theta)} \right\rangle, \qquad \sigma \left(\frac{\hat{\rho}^{\mathrm{og}}(\rho^{\mathrm{ug}}, E, \theta) - \hat{\rho}(\hat{S}_\mu)}{\rho^{\mathrm{ug}}(E, \theta)} \right). \tag{8.6}$$

Figure 8.22 shows the resulting bias and resolution as a function of zenith angle (left) and primary energy (right), for both hadronic models.

Although the resolution remains limited ($> 50\%$), the bias is centered around zero and does not show significant dependence on the energy or the zenith. The bias also stays within 10% in most cases, except for the most vertical showers and the highest energies.

The main reason of the difference in resolution between data and simulations is due to the data treatment, from using dense rings at an event-level in simulations and single stations per events in data. This was proved in simulations using one station per event to compute the resolution as can be seen in the comparative plot of Figure 8.23.

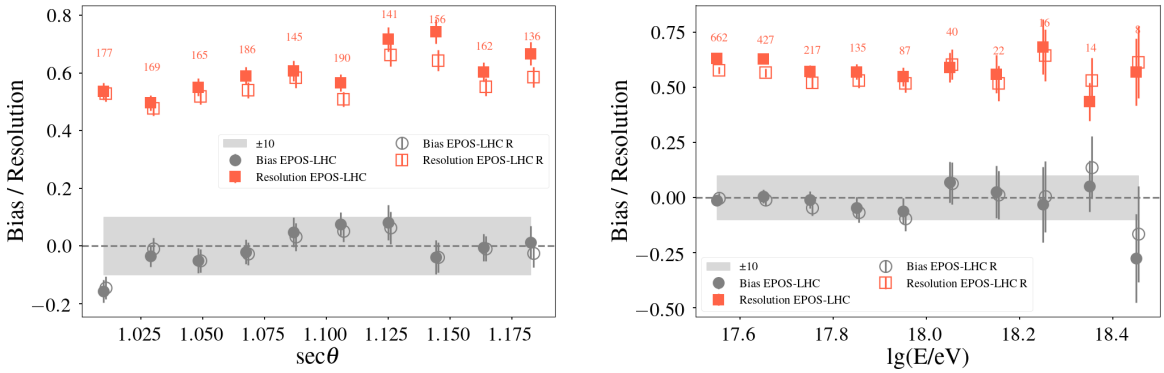


Figure 8.22: Bias and resolution of the calibrated muon density $\hat{\rho}(\hat{S}_\mu)$ compared to the parameterization of the UMD muon density.

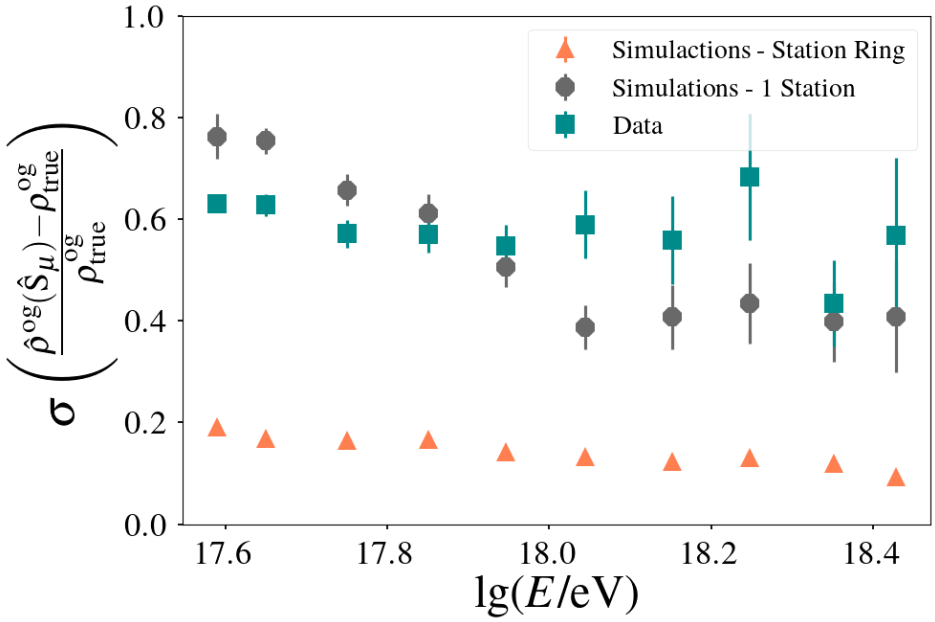


Figure 8.23: Resolution for different levels of data-taking: Station dense rings in simulations (pink triangles), simulations with one station per event (gray circles), data with ~1 station per event.

8.6. Zenith Correction

The goal of this analysis is to construct a zenith–independent estimator of the on–ground muon density by correcting the dependence of $\hat{\rho}^{\text{og}}$ on the arrival direction of the shower. Inclined showers traverse a larger atmospheric depth before reaching the ground. Consequently, fewer secondary particles survive to reach the detectors, producing a weaker footprint in the array compared to a more vertical shower initiated by the same primary with the same energy. This attenuation effect, commonly observed in surface observables, must be corrected to obtain a physically

meaningful, direction–independent muon estimator.

The study was restricted to showers with zenith angles between 0° and 33°, as defined in the previous section. Since the estimators were already corrected for their energy dependence, removing the residual zenith dependence is the final step toward obtaining a *pure* estimator. Such a corrected quantity enables direct comparisons between showers arriving at different angles and provides valuable input for composition analyses, since it reflects only intrinsic shower properties rather than geometric or atmospheric effects.

A reference angle of $\theta_{\text{ref}} = 23°$ was adopted, corresponding to the median of the event distribution within the selected range. The attenuation correction was defined as

$$\hat{\rho}_{23}^{\text{og}} \, f_{\text{att}}(\theta) = \hat{\rho}^{\text{og}}(\theta), \tag{8.7}$$

where $\hat{\rho}_{23}^{\text{og}}$ denotes the estimator normalized to the reference angle, obtained directly from the parameterization or after calibration.

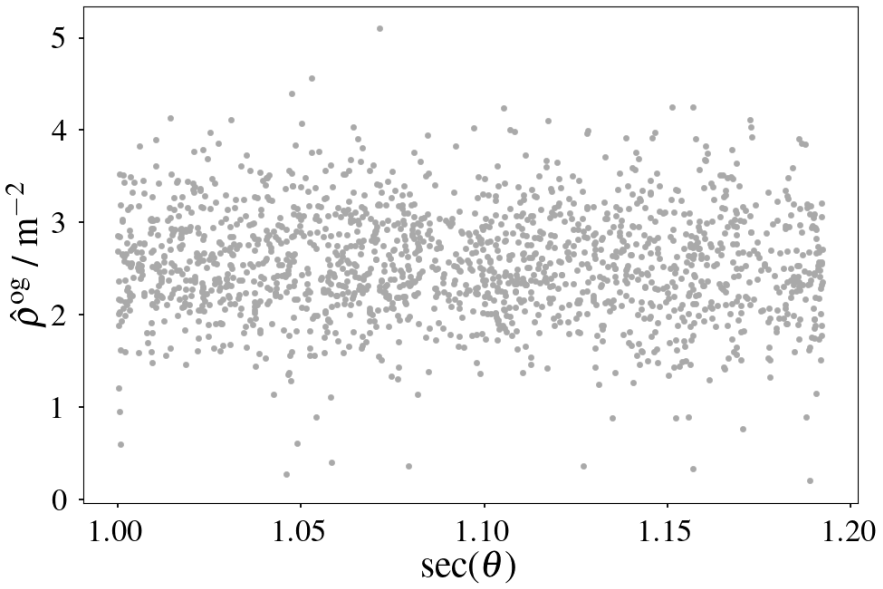


Figure 8.24: Dependence of the estimated on–ground muon density $\hat{\rho}^{\text{og}}$ on the zenith angle within the interval $0° < \theta < 33°$. Each point represents $\hat{\rho}^{\text{og}}$ per station. The mild negative slope indicates a gradual attenuation of the muon density with increasing atmospheric depth, consistent with the expected geometric and absorption effects for moderately inclined showers.

The functional form of the attenuation was determined empirically, motivated by the approximately linear decrease observed in Figure 8.24. It was modelled as

$$f_{\text{att}}(\theta) = \hat{\rho}_{23}^{\text{og}} + b \, (\sec\theta - \sec\theta_{\text{ref}}) \, , \tag{8.8}$$

where $\hat{\rho}_{23}^{\mathrm{og}}$ and b are free parameters obtained from a χ^2 minimization performed in each energy bin. The slope b quantifies the attenuation strength with the zenith angle, and its values as a function of energy are shown in Figure 8.25.

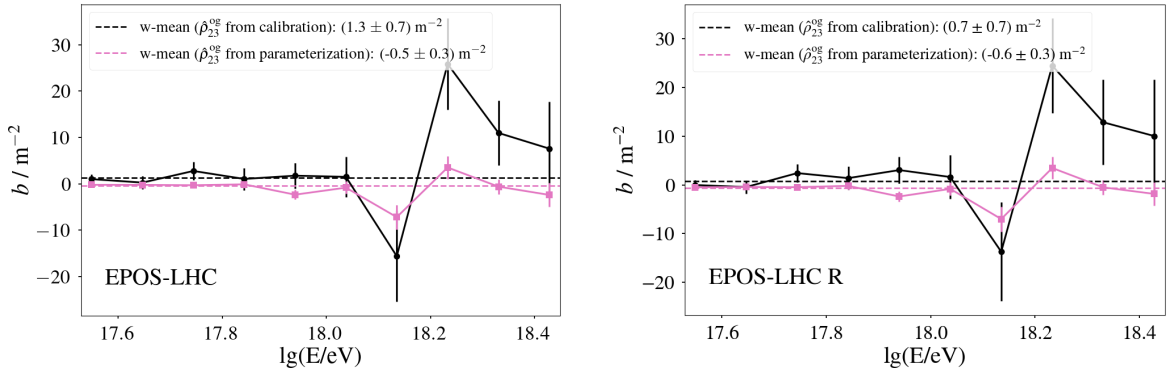


Figure 8.25: Slope b of the zenith dependence of the estimated on–ground muon density for the parametrized (pink) and calibrated (black) estimators, obtained in energy–binned fits for EPOS-LHC (left) and EPOS-LHC R (right). Negative values of b correspond to a decrease of $\hat{\rho}^{\mathrm{og}}$ with increasing zenith angle, consistent with physical attenuation. Positive slopes observed in the calibrated case at high energies suggest an unphysical trend, likely due to limited statistics or residual reconstruction biases.

For EPOS-LHC, the average slopes were found to be $\langle b \rangle = (-0.5 \pm 0.3)\,\mathrm{m}^{-2}$ for the parametrized estimator and $(1.3 \pm 0.7)\,\mathrm{m}^{-2}$ for the calibrated one. The corresponding values for EPOS-LHC R were $\langle b \rangle = (-0.6 \pm 0.3)\,\mathrm{m}^{-2}$ and $(0.7 \pm 0.7)\,\mathrm{m}^{-2}$. The positive slopes obtained in the calibrated cases are unexpected, indicating an unphysical increase of $\hat{\rho}^{\mathrm{og}}$ with the zenith angle, which may arise from limited event statistics or systematic effects in the calibration procedure. In the highest energy bins, the parameter b exhibits large fluctuations, consistent with the reduced number of events available.

Finally, a second fit iteration was performed with the slope b fixed to its corresponding mean value in each case. This procedure yielded consistent estimates of the muon density at the reference angle, producing direction–independent estimators. The resulting corrected quantities, denoted as $\hat{\rho}_{23}^{\mathrm{og}}$, provide the final zenith– and energy–independent muon density estimators used for subsequent composition analyses.

8.7. Mass composition and comparison with previous results

The connection between the muon content of extensive air showers and the primary mass is described by the Heitler–Matthews model (see Section 2.5). The total number of muons is expected to scale with mass A and primary energy E as

$$N_\mu = A \left(\frac{E}{AC}\right)^{\beta} = A^{1-\beta} \left(\frac{E}{C}\right)^{\beta}, \tag{8.9}$$

with power-law index $\beta \sim 0.9$ and energy constant C. Taking the logarithm and the average gives a linear relation,

$$\langle \ln N_\mu \rangle = (1-\beta)\langle \ln A \rangle + \beta \left\langle \ln\left(\frac{E}{C}\right)\right\rangle, \tag{8.10}$$

which directly links the mean logarithmic mass $\langle \ln A \rangle$ to the logarithm of the muon number.

8.7.1. The z-scale

To facilitate comparisons between different experiments and analyses, Dembinski et al. [28] introduced the *z-scale*, defined as

$$z := \frac{\ln(N^{\text{data}}_{\mu,\text{det}}) - \ln(N^{\text{p}}_{\mu,\text{det}})}{\ln(N^{\text{Fe}}_{\mu,\text{det}}) - \ln(N^{\text{p}}_{\mu,\text{det}})}, \tag{8.11}$$

where $N^{\text{p}}_{\mu,\text{det}}$ and $N^{\text{Fe}}_{\mu,\text{det}}$ are the proton and iron predictions from full-detector simulations. In this work, the values for $N^{\text{p,Fe}}_{\mu,\text{det}}$ are taken from Monte Carlo being the simulated muon density on ground.

Using the estimator $\hat{\rho}^{\text{og}}_{23}$, averaged in energy bins, the z-scale is computed as

$$z = \frac{\ln(\langle \hat{\rho}^{\text{og, data}}_{23} \rangle) - \ln(\langle \rho^{\text{og, p}}_{23} \rangle)}{\ln(\langle \rho^{\text{og, Fe}}_{23} \rangle) - \ln(\langle \rho^{\text{og, p}}_{23} \rangle)}. \tag{8.12}$$

Figure 8.26 shows the resulting z values for EPOS-LHC and EPOS-LHC R. The statistical uncertainties are represented by error bars, while systematic uncertainties are taken from Sections 5.6 and 6.4. At low energies both the parametrized, the density on ground that comes from parametrizing the UMD measurements, and calibrated, the ones that come from the calibration between the latter and $\hat{S}_\mu$, estimators give consistent results, while at higher energies the calibrated estimator fluctuates more strongly due to limited statistics. For EPOS-LHC, the results also agree with recent UMD measurements [84][1].

8.7.2. Translation to mass composition

Equation 8.10 and the z-definition (Eq. 8.11) can be combined to construct a mass-composition estimator,

$$z_{\text{mass}} = \frac{\langle \ln A \rangle}{\ln 56}. \tag{8.13}$$

Hence, the mean logarithmic mass is obtained as

[1] Thanks to M. Scornavacche for providing the data.

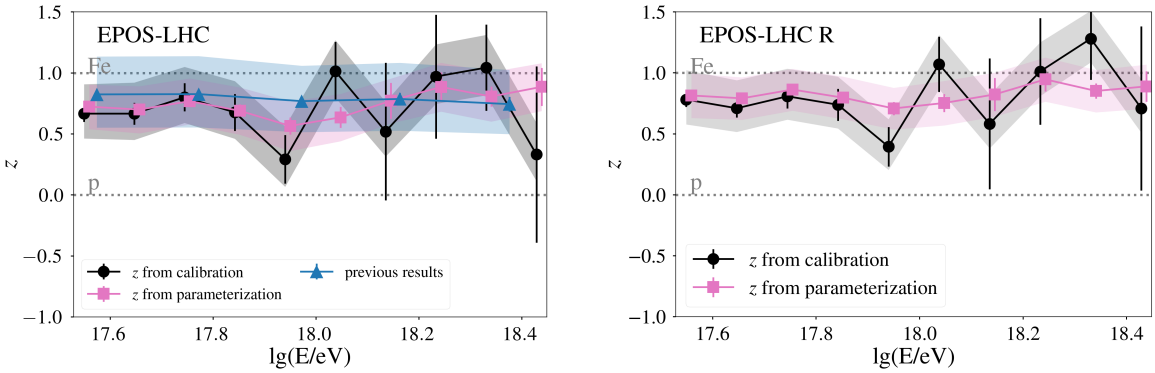


Figure 8.26: z-scale between data and simulations for EPOS-LHC (left) and EPOS-LHC R (right). In the case of EPOS-LHC the results are also compared with those obtained from ρ_{35}^{ug} in [84]. In both Figures, the **parametrization** (pink squares) corresponds to the values obtained with the method described in Chapter 5 from UMD measurements. The **calibration** (black circles) are the values obtained with $\hat{S}_\mu$ from NNs calibrated with the **parametrization** values as shown in Section 8.5.

$$\langle \ln A \rangle = \ln(56)\, \frac{\langle \ln(\hat{\rho}_{23}^{\mathrm{og,\,data}}) \rangle - \langle \ln(\rho_{23}^{\mathrm{og,\,p}}) \rangle}{\langle \ln(\rho_{23}^{\mathrm{og,\,Fe}}) \rangle - \langle \ln(\rho_{23}^{\mathrm{og,\,p}}) \rangle}. \tag{8.14}$$

The corresponding results are presented in Figure 8.27 for both interaction models. As in the z-scale, the parametrized and calibrated estimators are consistent at low energies, while at higher energies the calibrated estimator exhibits stronger fluctuations. The results remain compatible with earlier UMD measurements [84].

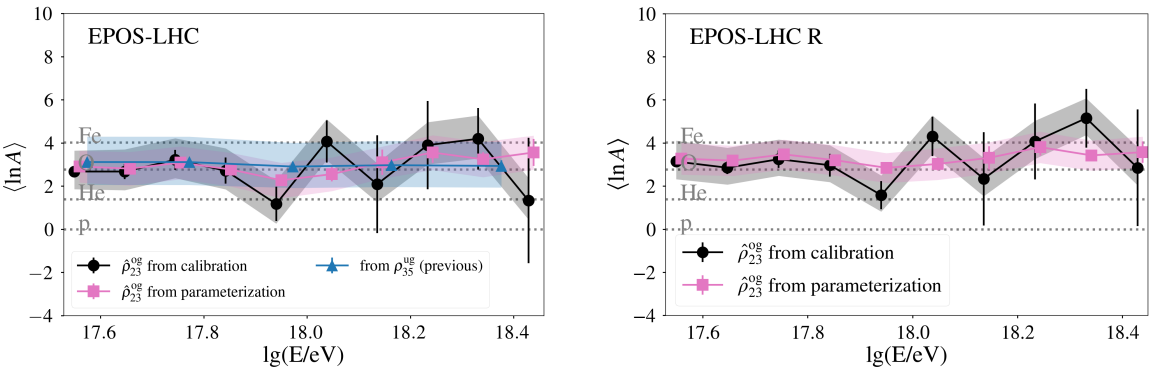


Figure 8.27: Mean logarithmic mass $\langle \ln A \rangle$ obtained from the z-scale using EPOS-LHC (left) and EPOS-LHC R (right). In the case of EPOS-LHC, results are compared to those from ρ_{35}^{ug} in [84].

8.7.3. Comparison with other experiments

Finally, Figure 8.28 compares the z-scale derived in this work with those obtained by other experiments. This step is essential since each experiment measures the muon content under different detector conditions, and the z-scale provides a common framework for comparison. Despite the shorter energy range considered here, the parametrized values follow a similar trend with energy as previous Auger UMD analyses [84, 62]. At higher energies, the calibrated values are limited by fluctuations, preventing a robust trend determination. As observed in earlier analyses, some tension remains with early UMD PMT results (purple symbols) and with Yakutsk (brown symbols).

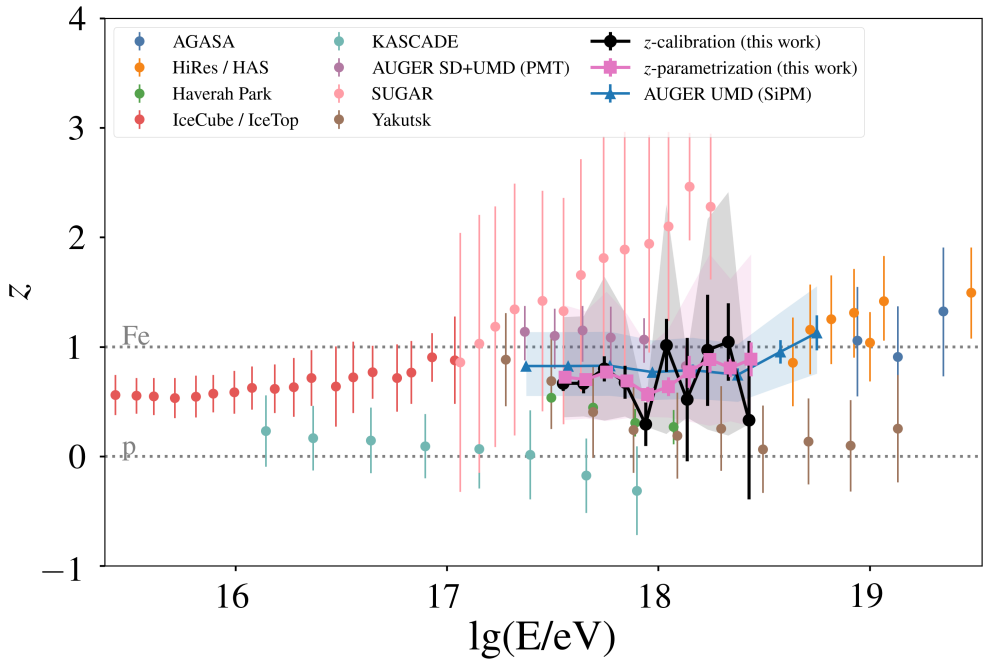


Figure 8.28: Comparison of the z-scale obtained in this work with other experiments, reproduced from [26].

8.8. Summary and conclusions

The estimator developed in chapter 5 was applied to real data from the Pierre Auger Observatory, and their performance with respect to the estimation of muon content on the ground was evaluated in this chapter, in a way that is comparable between experiments.

A dedicated UMD monitoring strategy was implemented, introducing trace–sensitive observables to track stability and flag anomalies. After quality selections, 2 529 stations within [400, 500] m of the core were retained, which corresponded to the phase

space in which the on–ground estimator and its calibration were constructed. Because the estimator $\hat{\rho}^{\mathrm{og}}$ was derived from simulations, it was necessary first to test whether the long–standing data–simulation tension in the muon content depends on the zenith angle, which would hint at a mismatch in the muon energy spectrum. Regressions of the data–to–simulation ratio of ρ^{ug} against $\sec\theta$ for EPOS-LHC, EPOS-LHC R and SIBYLL 2.3e yielded slopes compatible with zero at the level $p > 0.05$, supporting the applicability of the on-ground estimator in the data throughout the studied zenith range.

The correlations between the measured and estimated quantities were then characterized. The total WCD signal, S_{tot}, and the muon density of the UMD, ρ^{ug}, show a modest correlation ($r \simeq 0.28$), while $\hat{S}_\mu$ correlates with ρ^{ug} in $r = 0.30$ for EPOS-LHC and $r = 0.32$ for EPOS-LHC R. The on-ground estimator $\hat{\rho}^{\mathrm{og}}$ exhibits stronger correlations, with S_{tot} at $r = 0.43$ (EPOS-LHC) and $r = 0.45$ (EPOS-LHC R), and with $\hat{S}_\mu$ at $r = 0.33$ and $r = 0.36$, respectively, for the complete data set. However, it must be noted that the correlation with S_{tot} is affected by zenith dependencies, since this observable is particularly sensitive to the electromagnetic component of the shower. The apparent discrepancy between the zenith dependence of the correlations in data and in simulations can be attributed to the level of analysis: simulations were evaluated at the event level, whereas data correlations were computed at station level.

A calibration between $\hat{S}_\mu$ and $\hat{\rho}^{\mathrm{og}}$ was performed in zenith bins using a linear model, which yielded slopes in agreement with simulations ($\langle k_1 \rangle \simeq 3.8\,\mathrm{VEM\,m^2}$ for both models) but larger offsets in data ($\langle k_0 \rangle \simeq 20\,\mathrm{VEM}$ compared to $\sim$12 VEM in simulation), suggesting a baseline shift rather than a slope mismatch. The calibrated estimator $\hat{\rho}^{\mathrm{og}}(\hat{S}_\mu)$ showed biases consistent with zero (typically below 10%) and resolutions higher than 40%, without significant dependence on energy or zenith. To obtain a zenith–independent observable, the residual dependence was modelled as $\hat{\rho}^{\mathrm{og}}(\theta) \approx \hat{\rho}^{\mathrm{og}}_{23} + b(\sec\theta - \sec\theta_{\mathrm{ref}})$, with $\theta_{\mathrm{ref}} = 23^\circ$. The per-energy fits produced mean slopes close to zero for both EPOS-LHC and EPOS-LHC R, and fixing b to the weighted mean stabilized the intercepts, defining $\hat{\rho}^{\mathrm{og}}_{23}$ as the zenith–corrected estimator used for composition analyses.

Using $\hat{\rho}^{\mathrm{og}}_{23}$ together with Monte Carlo proton and iron anchors, the z scale was derived and subsequently translated into $\langle \ln A \rangle$. The parametrized and calibrated determinations were consistent at low energies, while the calibrated values fluctuated at the highest energies because of limited statistics. For EPOS-LHC, the results agree with recent UMD measurements within uncertainties. A comparison with other experiments in the common z–scale framework shows broadly consistent energy trends, while tensions remain with early Auger PMT results and with Yakutsk. The main limitations identified are the reduced statistics at the highest energies, the difference between the station-level treatment in the data and the event-level design of some tools, and a persistent offset in the $\hat{S}_\mu$–$\hat{\rho}^{\mathrm{og}}$ relation. These point to clear avenues for improvement, including implementing event level $\hat{S}_\mu$ in data, expanding statistics, refining muon-electromagnetic separation in neural network inputs, and reassessing low–energy hadronic and atmospheric modelling. In conclusion, within the selected phase space ($\lg(E/\mathrm{eV}) \in [17.5, 18.5]$, $\theta < 45^\circ$, with zenith–stable results

up to about 33°), the zenith–corrected on–ground estimator $\hat{\rho}_{23}^{\mathrm{og}}$ provides consistent z and $\langle \ln A \rangle$ determinations, and, together with the new monitoring and calibration, forms a coherent framework for stabilizing muon observables in data and allowing robust composition comparisons across experiments.

Chapter IX

Summary, Conclusions and Future Prospects

The study of the muon content in extensive air showers (EAS) remains one of the most sensitive probes for understanding hadronic interactions at ultra-high energies and for inferring the mass composition of cosmic rays. Persistent discrepancies between measurements and model predictions, commonly referred to as the *Muon Puzzle*, suggest that current hadronic interaction models underestimate muon production in the atmosphere. The AugerPrime upgrade of the Pierre Auger Observatory, with its enhanced detection capabilities, provides an unprecedented opportunity to investigate this issue by combining independent estimators of the muonic component derived from the Surface Detector (SD) and the Underground Muon Detector (UMD).

The work presented in this thesis aimed to develop and validate a consistent framework for estimating the muon density at the surface using the UMD, and to establish a calibration between underground and surface observables. This goal was addressed through a combination of detailed Monte Carlo simulations and analyses of real data collected by the AugerPrime detectors.

A first component of the study involved the use of simulation libraries of extensive air showers generated with CORSIKA and analyzed within the simulation and reconstruction framework of the Pierre Auger Collaboration, $\overline{\text{Off}}\underline{\text{line}}$. These simulations were employed to characterize the relation between the true on-ground muon density and the corresponding underground measurements, as well as to explore their dependencies on primary energy and zenith angle. Complementary simulations were performed to quantify muon energy losses in soil using both the Continuous Slowing Down Approximation (CSDA) and full Geant4 particle transport, enabling a precise description of the energy threshold imposed by the UMD shielding.

Using these tools, simulations of both the Water-Cherenkov Detectors (WCDs) and the UMD were carried out with the official $\overline{\text{Off}}\underline{\text{line}}$ software of the Pierre Auger Collaboration. Composition estimates based on the fractional contribution of four pure primary species were used to construct an Auger-Mix-like composition. The on-ground muon density, ρ^{og}, was defined as the total muon density crossing each WCD within a dense ring of stations located 450 m from the shower core. This quantity was fitted against ρ^{ug} for each simulated event. The variation of the obtained

fit parameters with respect to primary energy and zenith angle was then studied, leading to a global parameterization of ρ^{og} as a function of ρ^{ug}, energy, and zenith angle. The parameterization proved to be robust, exhibiting a zero-centered bias independent of both zenith angle and energy, with a resolution better than 15%. This analysis demonstrated the capability of the UMD to validate and calibrate surface muon content estimations through direct muon measurements.

In a second stage, calibration functions were established to connect the estimated on-ground muon density with the simulated muon density in the WCDs of the SD. These functions were derived for both direct simulation-based estimators and a neural-network model trained to reconstruct the muonic signal in the WCDs. The latter approach is particularly relevant for application to real data, since the true muonic component cannot be directly disentangled in experimental measurements. The neural-network estimator proved robust against fluctuations, offering improved generalization over the parameterized models.

The final part of the thesis applied these estimators to real data from the Pierre Auger Observatory. For the first time, a calibration was performed between independent estimations of the muonic component derived from the SD and the UMD measurements. This represents a major step toward an integrated interpretation of the surface and underground detectors, laying the foundation for future analyses aimed at determining the absolute muon content in air showers and testing the consistency of hadronic interaction models at the highest energies.

The methods and results presented here demonstrate that the UMD can be effectively used to infer the muon density at ground level. Its combination with SD observables has proved to be consistent with previous results on mass composition and with the observed muon deficit at the lowest primary energies studied in this work, although additional statistics are required to obtain conclusive results at the highest energies. The established calibration framework will enable future studies to quantify systematic differences between data and simulations, constrain model uncertainties, and investigate potential energy-dependent deviations in muon production.

Future work will focus on extending these analyses to the newest data set of the Pierre Auger Observatory, known as *Phase II*. Moreover, the research will continue with the extrapolation of the calibration to higher primary energies to make it compatible with the main SD array of the Pierre Auger Observatory, where the UMD is not deployed and no direct muonic component measurement is possible for vertical showers. Further refinement of the neural-network approach and the inclusion of additional observables, such as the SSD signal, will contribute to improving the precision and robustness of the muon estimators.

Appendix A

Stopping power for $10^{17.5}$ eV and 30° primaries

Distributions for $10^{17.5}$ eV

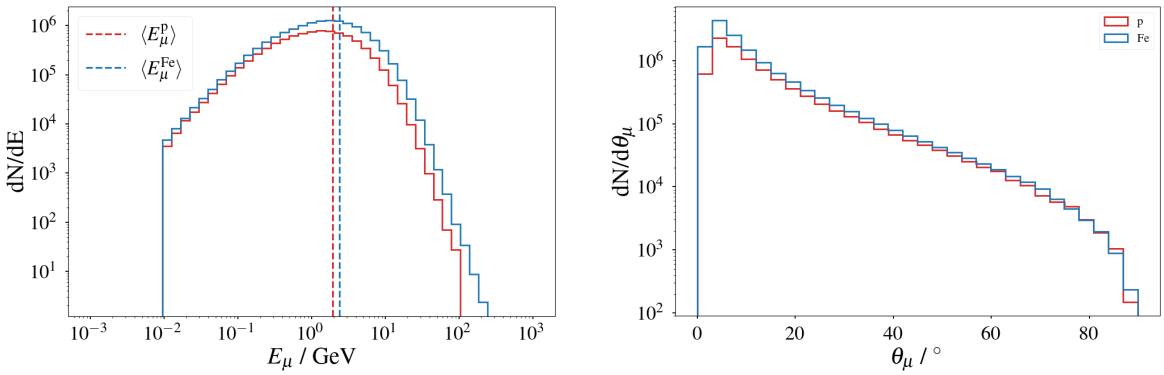


Figure A.1: Energy and zenith distributions for muons on-ground for vertical showers with an energy of $10^{17.5}$ eV

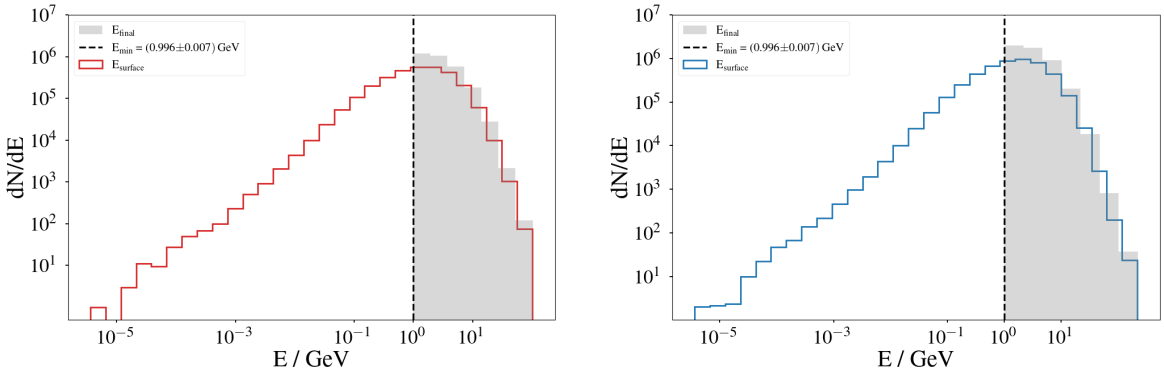


Figure A.2: Energy distributions of muons underground for vertical showers with an energy of $10^{17.5}$ eV compared to their original energy before trespassing the soil.

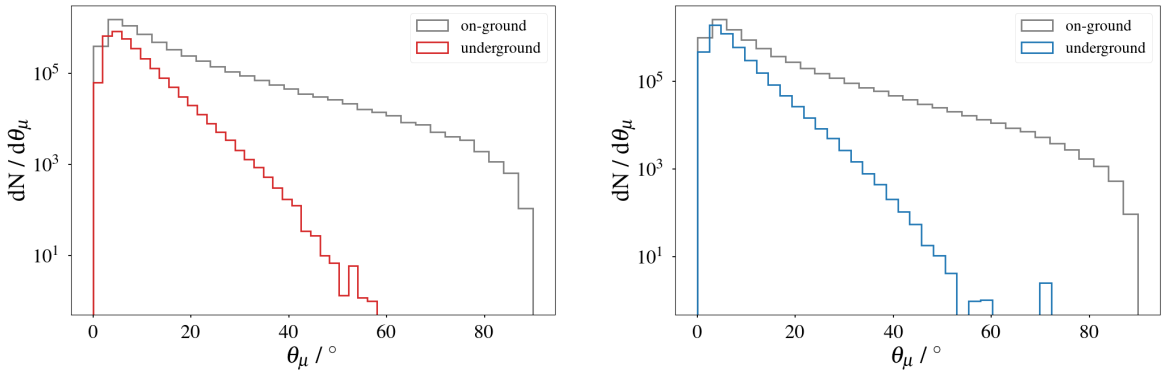


Figure A.3: Zenith distributions of muons underground for vertical showers with an energy of $10^{17.5}$ eV.

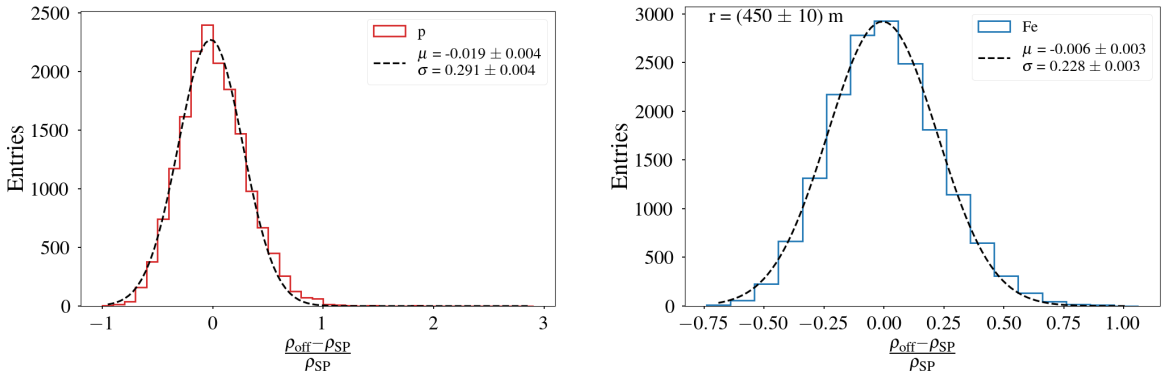


Figure A.4: Bias and resolution of the injected muons in the UMD Offline simulations for vertical showers with an energy of $10^{17.5}$ eV.

Distributions for 30°

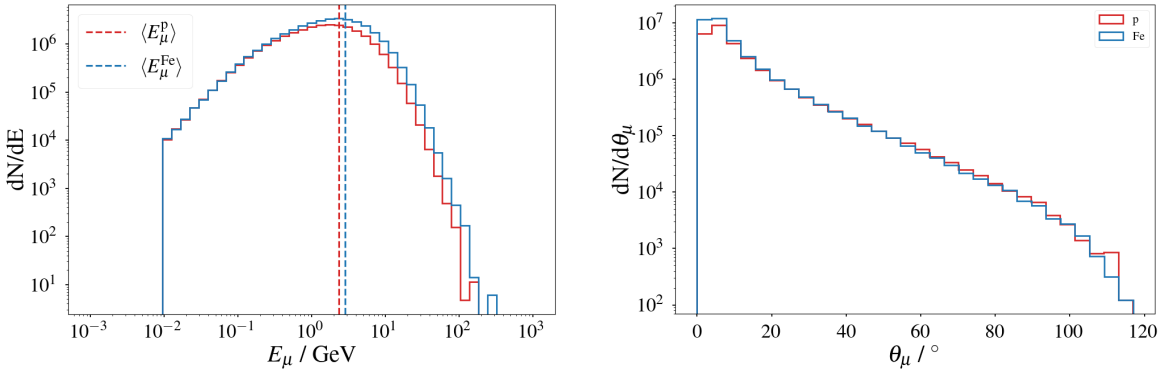


Figure A.5: Energy and zenith distributions for muons on-ground for 30 ° showers with an energy of 10^{18} eV

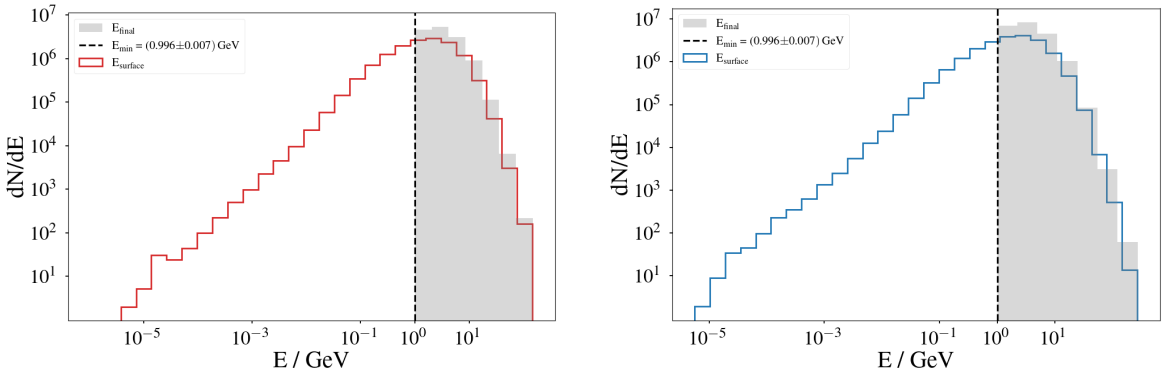


Figure A.6: Energy distributions of muons underground for 30 ° showers with an energy of 10^{18} eV compared to their original energy before trespassing the soil.

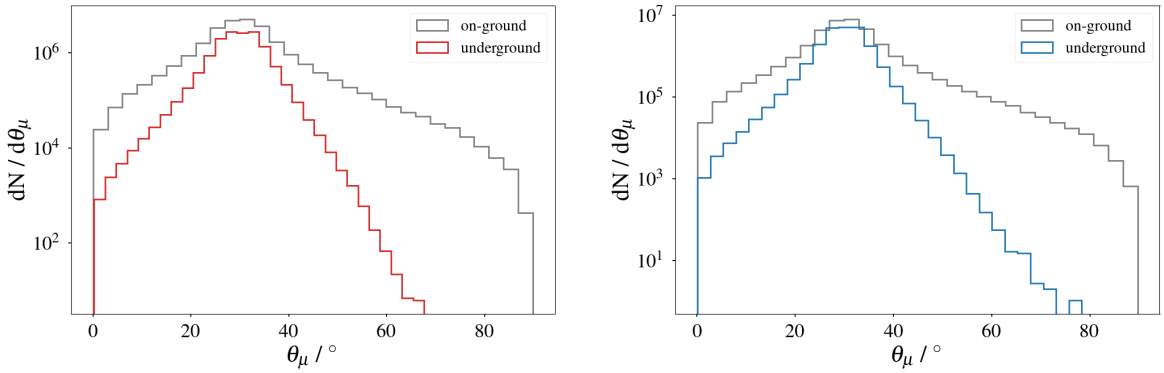


Figure A.7: Zenith distributions of muons underground for 30 ° showers with an energy of 10^{18} eV.

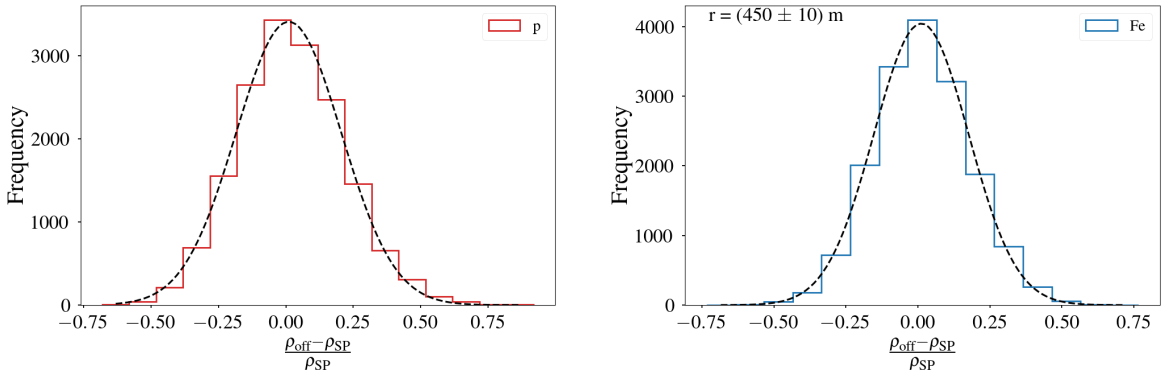


Figure A.8: Bias and resolution of the injected muons in the UMD Offline simulations for showers of 30 ° with an energy of $10^{18.0}$ eV.

Appendix B

Modeling S_μ Resolution with Artificial Noise

Artificial Noise Model

To study the effect of limited resolution on the calibration process, controlled artificial fluctuations were introduced into the S_μ values in simulations. These fluctuations mimic the statistical uncertainties one would expect from real-data estimators.

Let S_μ^{true} be the true muon signal in simulation. We define the noisy version as:

$$S_\mu^{\text{noisy}} = S_\mu^{\text{true}} + \mathcal{N}(0,\ (c\, S_\mu^{\text{true}})^2) \tag{B.1}$$

Here, $\mathcal{N}(0, \sigma^2)$ denotes a Gaussian distribution with mean 0 and standard deviation σ, and c is a dimensionless parameter that controls the relative resolution. In this formulation, the noise added to each event is drawn from a normal distribution whose standard deviation is proportional to the signal itself.

The relative (fractional) error is then defined as:

$$\delta = \frac{S_\mu^{\text{noisy}} - S_\mu^{\text{true}}}{S_\mu^{\text{true}}} \sim \mathcal{N}(0, c^2) \tag{B.2}$$

This follows from the fact that the added noise term, $\varepsilon = S_\mu^{\text{noisy}} - S_\mu^{\text{true}}$, is Gaussian-distributed:

$$\varepsilon \sim \mathcal{N}(0, (c\, S_\mu^{\text{true}})^2)$$

and dividing a normally distributed variable by a constant scales its standard deviation accordingly. Thus:

$$\delta = \frac{\varepsilon}{S_\mu^{\text{true}}} \sim \mathcal{N}\left(0,\ \frac{(c\, S_\mu^{\text{true}})^2}{(S_\mu^{\text{true}})^2}\right) = \mathcal{N}(0,\ c^2)$$

In other words, the distribution of relative errors is centered at 0 and has standard deviation c, making c the standard deviation of the relative error:

$$\text{std}\left(\frac{S_\mu^{\text{noisy}} - S_\mu^{\text{true}}}{S_\mu^{\text{true}}}\right) = c \tag{B.3}$$

This artificial noise model provides a simple and controlled way to study how estimator resolution impacts the calibration procedure.

Appendix C

Unthinning sanity check for Iron composition and inclined showers

Inclined Proton showers

For identical CORSIKA proton showers at $\sim 46°$, the injected-particle spectra recorded in the WCDs show the expected ordering of components. Photons dominate at low deposited energies, electrons populate an intermediate range, and muons form a harder tail. We compare simulations with only the WCD against configurations that also include the UMD in the station; the presence of the UMD changes only the shower geometry/material budget but leaves the injected-particle shapes consistent within expectations.

The corresponding integrated WCD signals reflect these component differences. We show photons and electrons on the top row and muons on the bottom panel; solid lines denote WCD-only simulations, dashed lines the WCD+UMD setup.

Vertical and inclined iron showers

For iron primaries, we repeat the checks at 0° and 46°. The injected-particle spectra (photons, electrons, muons) show the same qualitative behaviour as protons, with overall higher particle yields and harder muon components, as expected for heavier primaries. The comparison between WCD-only and WCD+UMD configurations again shows consistent shapes.

The integrated WCD signals for iron at 0° follow the same component ordering. Solid lines correspond to WCD-only simulations; dashed lines include the UMD.

At 46°, the injected-particle spectra preserve the same hierarchy, with the inclined geometry shifting the distributions as expected. The WCD-only vs. WCD+UMD comparison remains consistent.

Finally, the integrated WCD signals at 46° show the same trends and agreement between configurations.

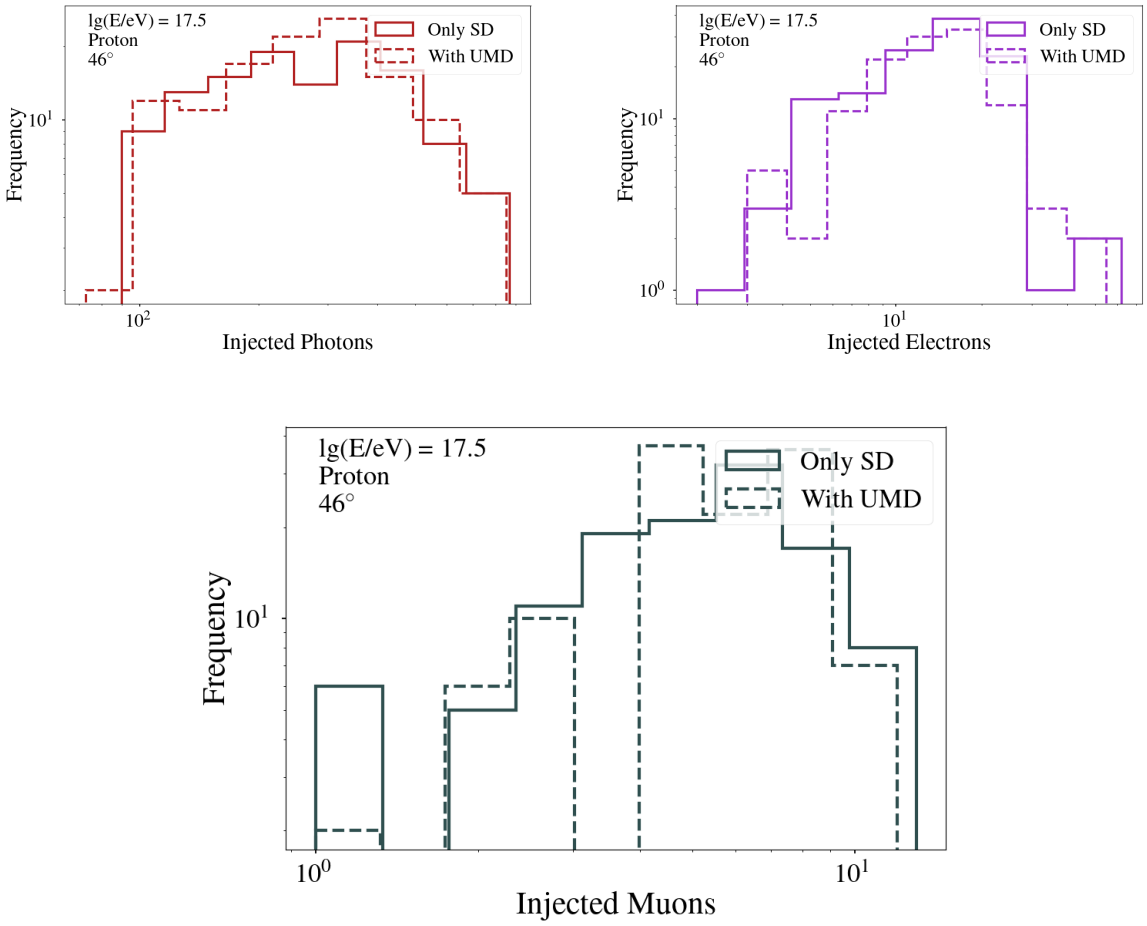


Figure C.1: Injected particle distributions (photons, electrons, muons).

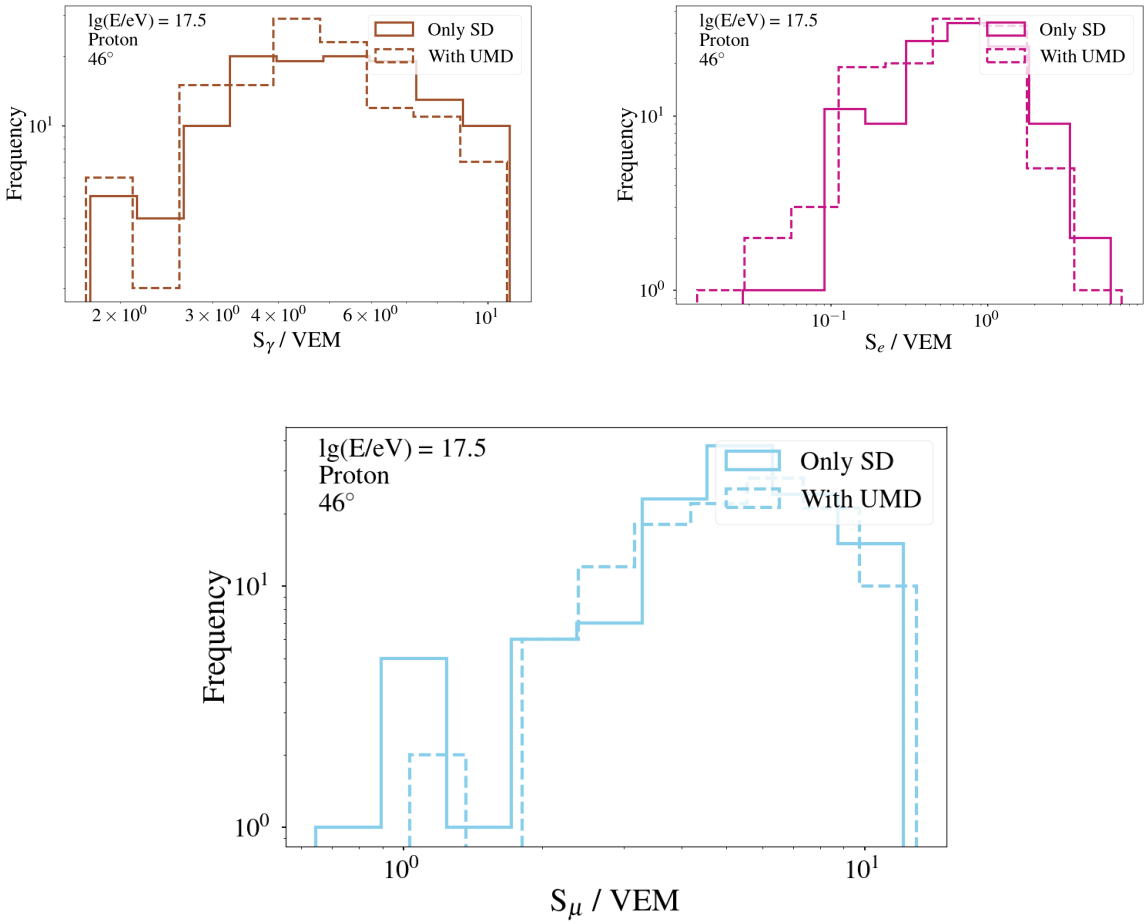


Figure C.2: Integrated WCD signals (photons, electrons, muons).

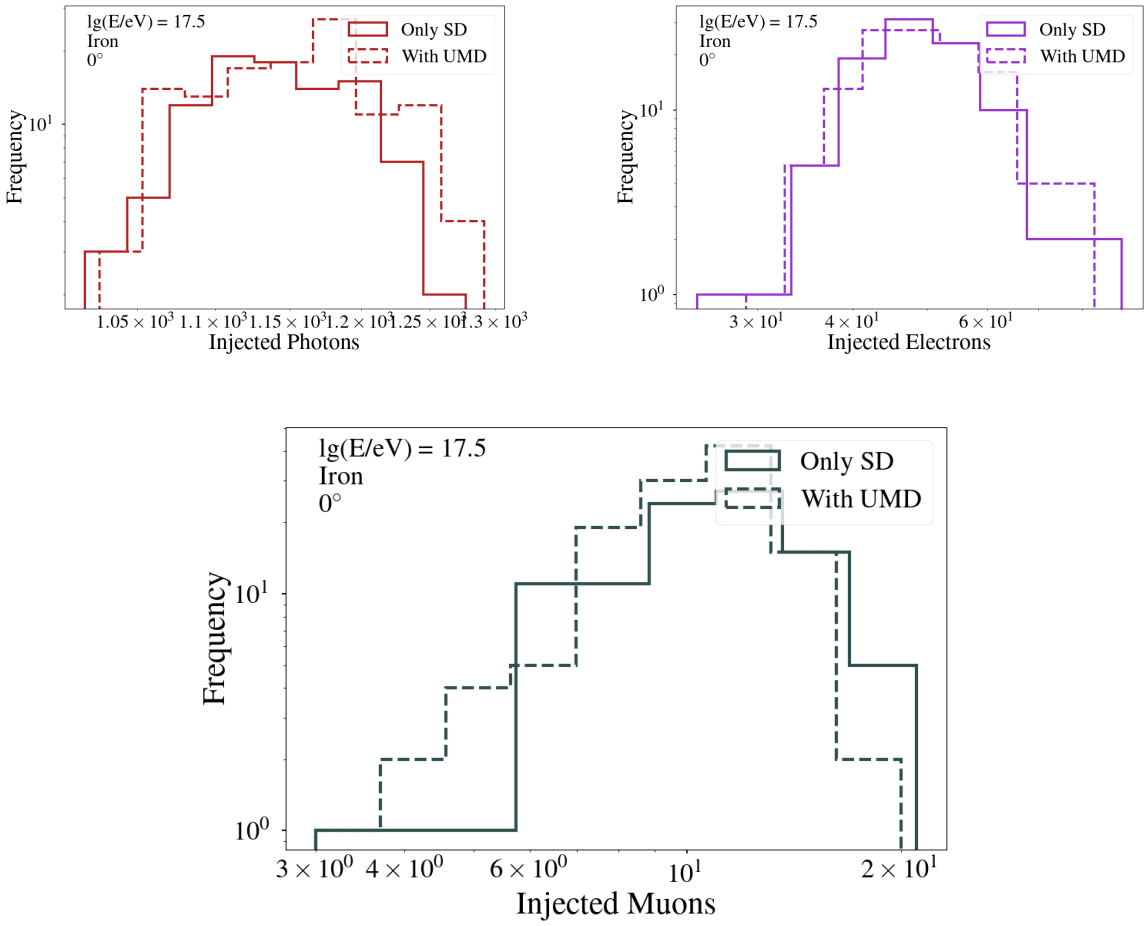


Figure C.3: Injected particle distributions at 0° (photons, electrons, muons).

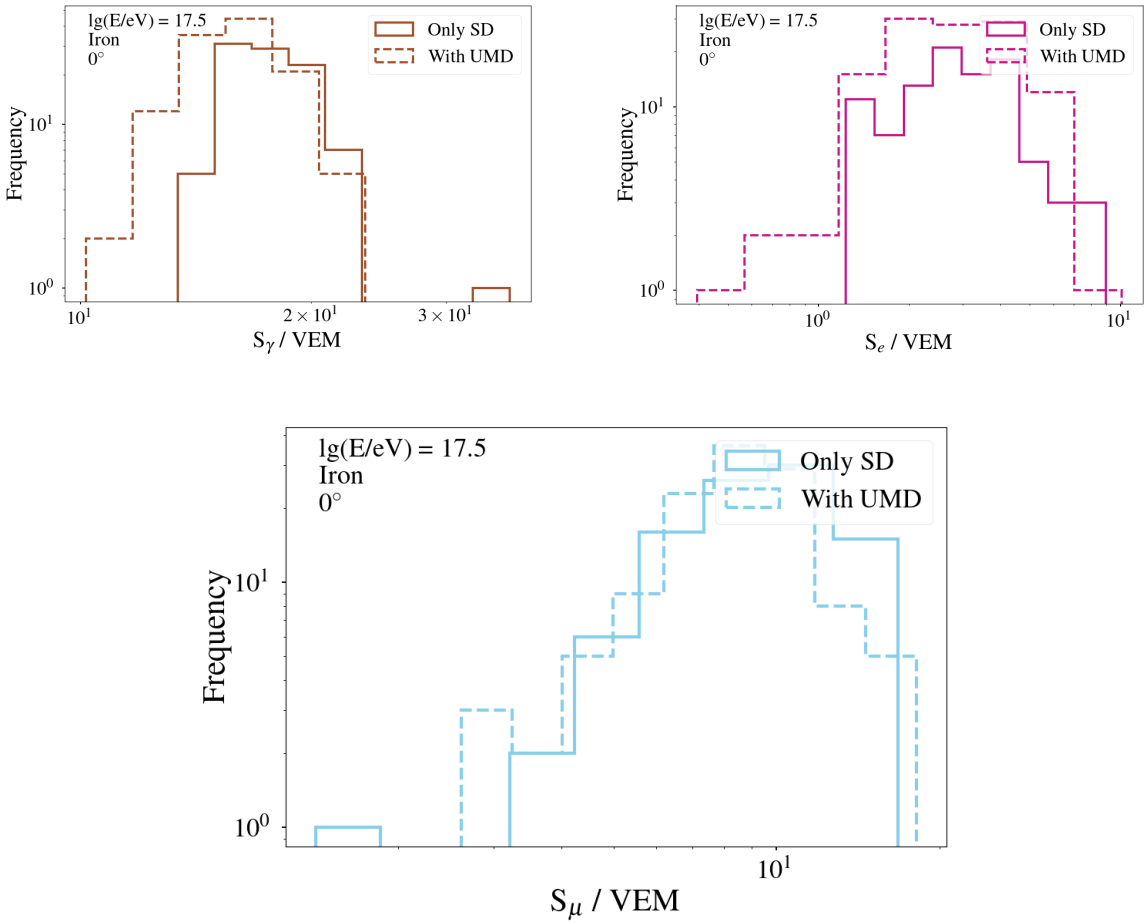


Figure C.4: Integrated WCD signals at 0° (photons, electrons, muons).

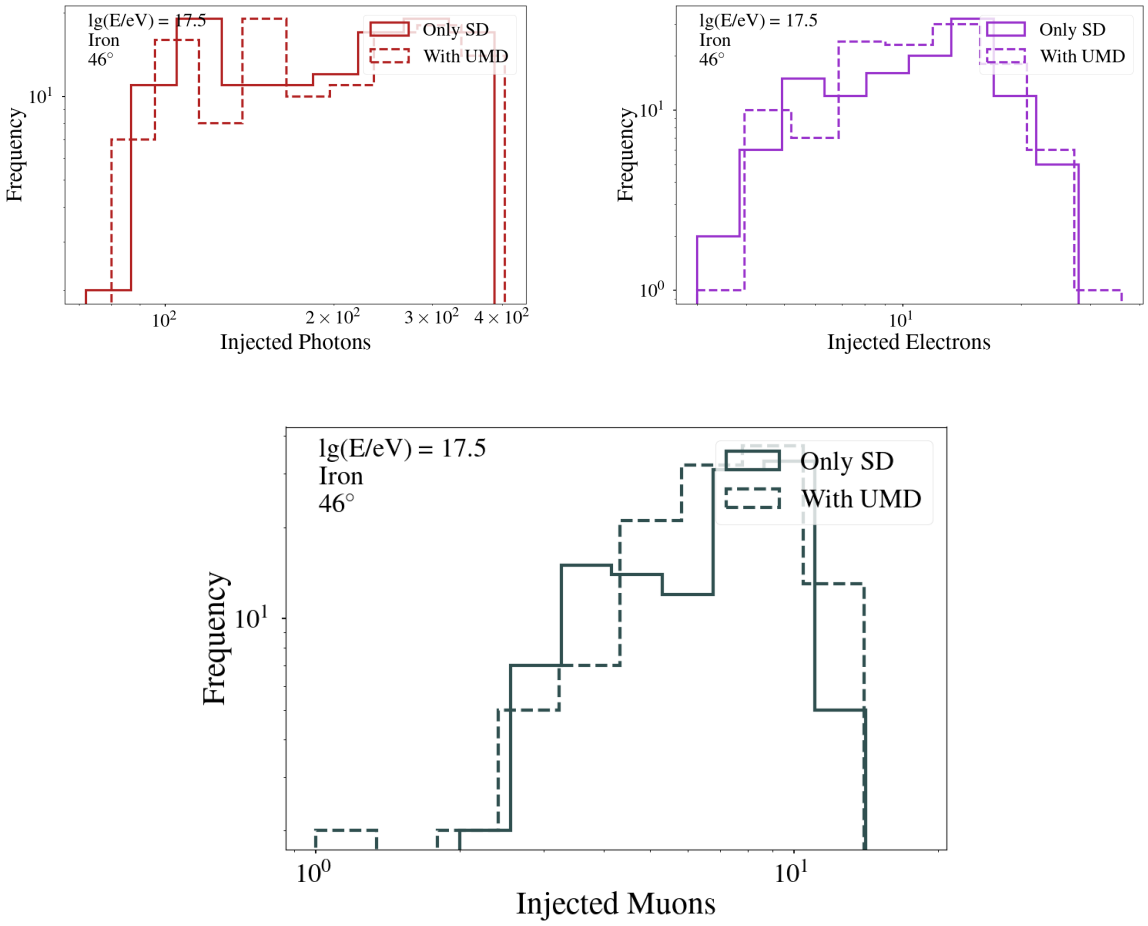


Figure C.5: Injected particle distributions at 46° (photons, electrons, muons).

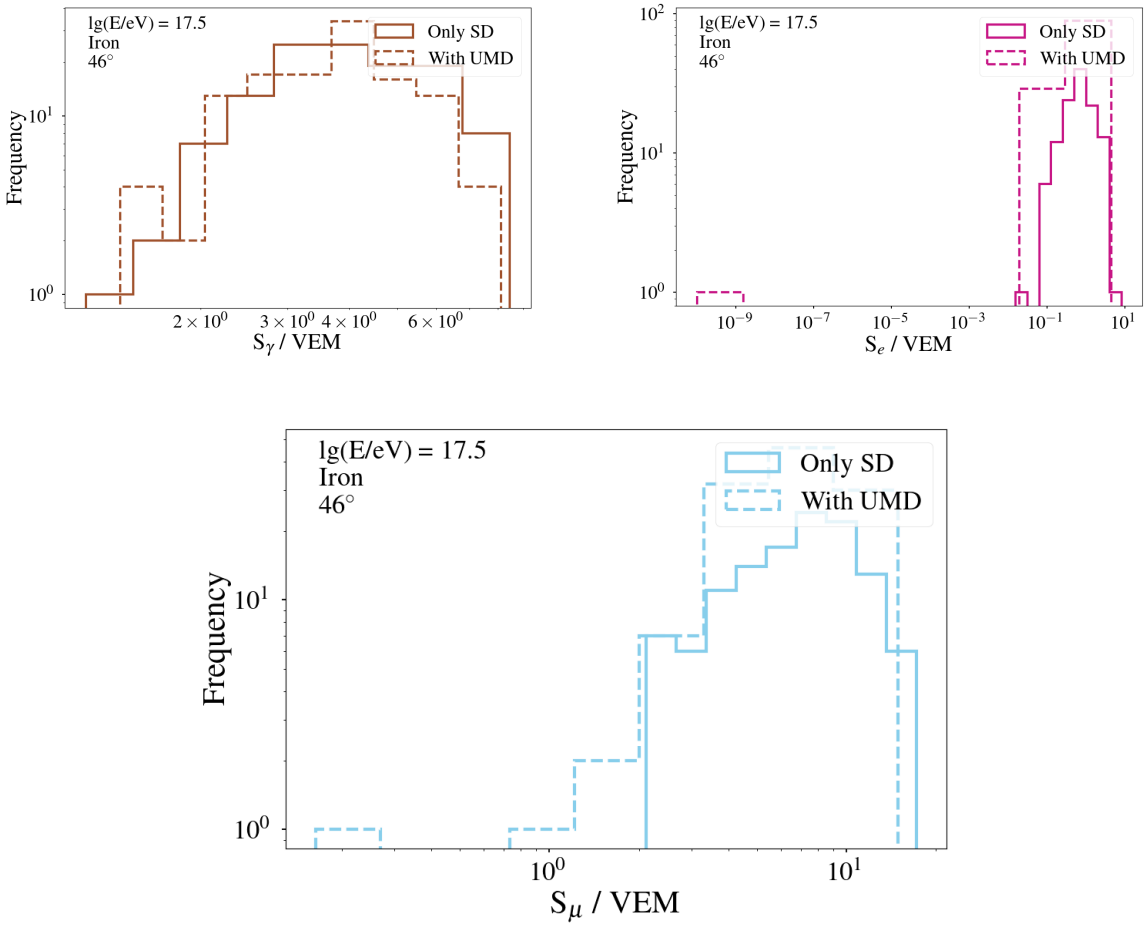


Figure C.6: Integrated WCD signals at 46° (photons, electrons, muons).

Appendix D

Parameterization of muons on-ground using EPOS-LHC R

For the model EPOS-LHC R, the same functional form presented in Chapter 5 was used to parameterize the muon density on-ground. The parameter values can be found in Table D.1.

Table D.1: Optimized parameters obtained via χ^2 minimization.

Parameter	Value
γ_0	0.621 ± 0.002
γ_1	0.041 ± 0.033
γ_2	-0.689 ± 0.088
α_0	0.646 ± 0.001
α_1	-0.574 ± 0.003
α_2	0.008 ± 0.005

The residuals were also calculated finding a mean=0 with a standard deviation of 0.9 as can be seen in Figure D.1 (left). The mean estimated and true values are compatible with the idensity as its shown in Figure D.1 (right).

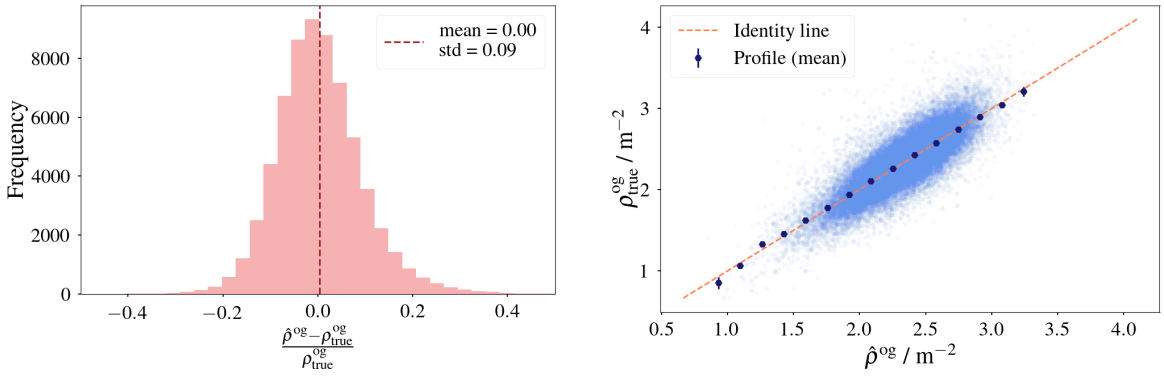


Figure D.1: Left: Residuals histogram for the parameterization of muon density on-ground using EPOS-LHC R. Right: True vs. estimated values.

Appendix E

Complete dataset correlation in simulations

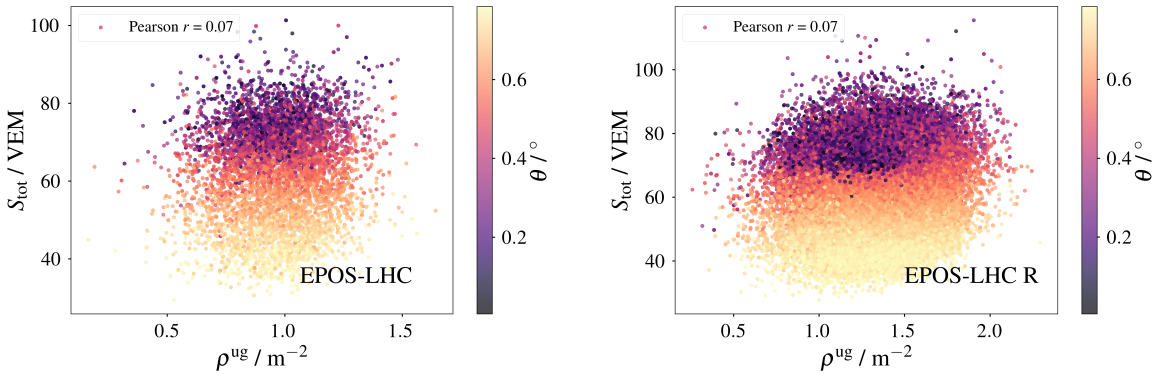


Figure E.1: Complete dataset correlation between S_{tot} and ρ^{ug} (MC). Pearson coefficients: $r = 0.05$ (EPOS-LHC, left) and $r = 0.08$ (EPOS-LHC R, right).

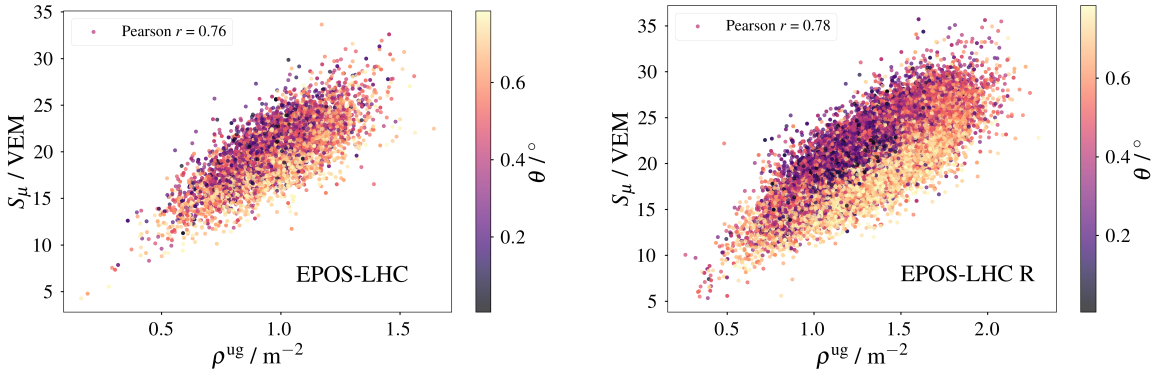


Figure E.2: Complete dataset correlation between S_μ (MC) and ρ^{ug} (MC). Pearson coefficients: $r = 0.76$ (EPOS-LHC, left) and $r = 0.79$ (EPOS-LHC R, right).

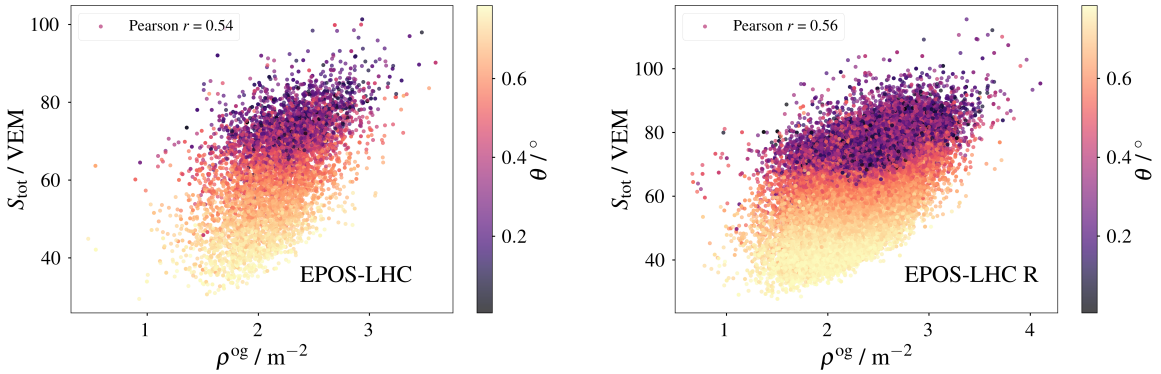


Figure E.3: Complete dataset correlation between S_{tot} and ρ^{og} (MC). Pearson coefficients: $r = 0.54$ (EPOS-LHC, left) and $r = 0.56$ (EPOS-LHC R, right).

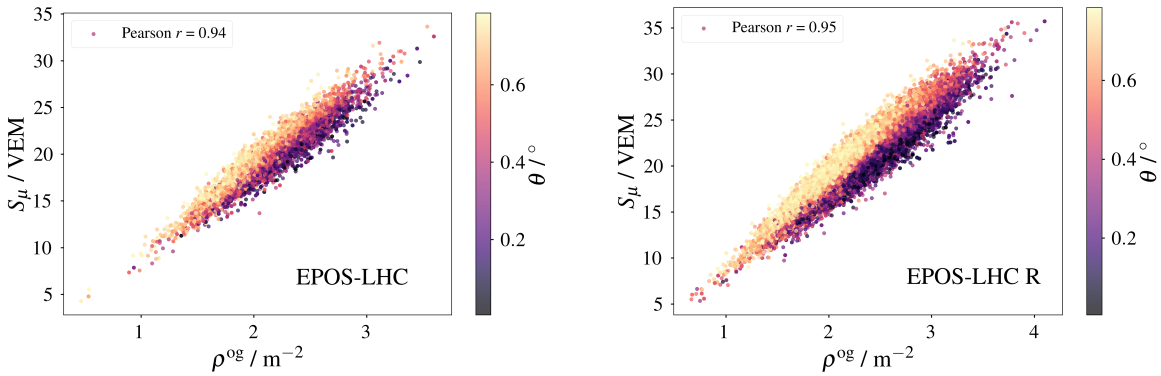


Figure E.4: Complete dataset correlation between S_μ (MC) and ρ^{og} (MC). Pearson coefficients: $r = 0.94$ (EPOS-LHC, left) and $r = 0.95$ (EPOS-LHC R, right).

Appendix F

Correlation trend with zenith angle of un-normalized variables

To evaluate how the correlation between the density of underground muons, $\rho^{\mu g}$, and the total signal in the water-Cherenkov detectors, S_{tot}, evolves with the zenith angle, the data set was divided into bins of reconstructed energy and zenith. For each energy bin, the Pearson's correlation coefficient r was calculated within narrower zenith intervals. The central value of each zenith bin was expressed in terms of $\sec\theta$, so that all energy bins could be compared on a common abscissa.

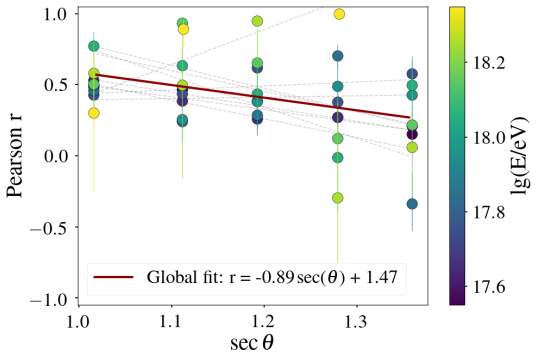
For every energy bin, the sequence of r values obtained as a function of $\sec\theta$ was fitted with a straight line,

$$r(\sec\theta) \simeq a + b\ \sec\theta,$$

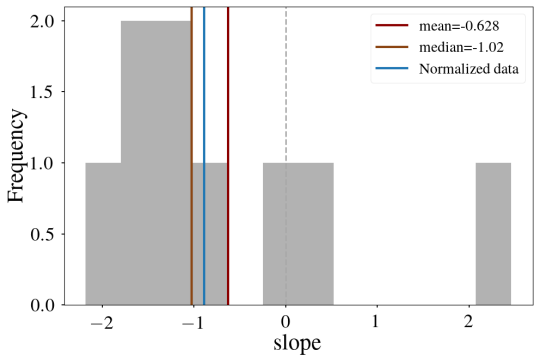
providing an estimate of the slope b that quantifies the trend of the correlation with increasing zenith. A negative slope indicates that the correlation weakens for more inclined showers. Repetition of this procedure across all energy bins yields a distribution of slopes $\{b\}$ that can be used to assess whether the downward trend is systematic.

The analysis yielded $N = 9$ independent energy bins with a mean slope of $\langle b\rangle = -0.63$ and a median slope of -1.02. In total, 66.7% of the slopes were negative, supporting the interpretation that the correlation generally decreases with zenith. As a complementary test, a pooled Spearman rank correlation was calculated between r and $\sec\theta$, giving $\rho = -0.41$ with $p = 6.3\times10^{-3}$. This confirms a significant monotonic decrease of the Pearson correlation with increasing zenith angle across the data set.

As a final sanity check, the ratio between S_{tot} and $\rho^{\mu g}$ was calculated for each energy and zenith bin finding that it decreases as expected given the stronger atmospheric attenuation of the electromagnetic component of the shower compared to the muonic component. This is shown in Figure F.2.



(a) Pearson r vs. $\sec\theta$ colored by energy, with per-energy (gray) and global (red) fits.



(b) Distribution of per-energy slopes b from $r(\sec\theta) \simeq a + b\ \sec\theta$.

Figure F.1: Correlation trend between underground muon density $\rho^{\mu g}$ and WCD total signal S_{tot} as a function of zenith. Left: Pearson correlation coefficient r versus $\sec\theta$ for each energy bin (points color-coded by bin center); thin gray lines show per-energy linear fits, and the crimson line the global fit. Right: histogram of per-energy slopes b. Over $N = 9$ energy bins we find a mean slope $\langle b \rangle = -0.63$ and a median slope of -1.02, with 66.7% of slopes negative, indicating that the correlation generally weakens with increasing zenith angle. A pooled Spearman rank test between r and $\sec\theta$ gives $\rho = -0.41$ with $p = 6.3 \times 10^{-3}$, confirming a significant monotonic decrease.

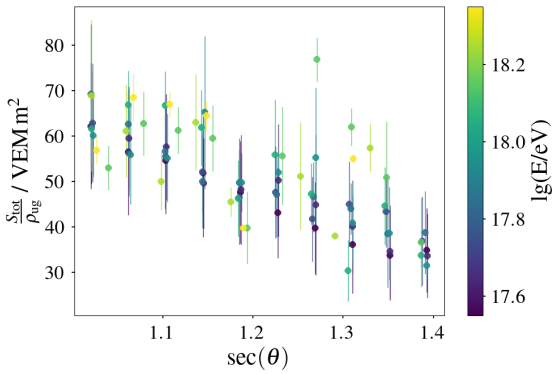


Figure F.2: Ratio between S_{tot} and $\rho^{\mu g}$ vs. zenith.

Acknowledgments

Amigo puede ser quien bien repara
En la musa o engendro que yo aporte
Amigo, sí, es también quien me soporte
Pero amigo mayor es quien me ampara[a]

[a] A friend can be one who carefully considers the muse or draft I bring; a friend, yes, is also one who puts up with me; but a greater friend is the one who shelters me. — from "Amigo mayor," Silvio Rodríguez.

I have been very fortunate to be surrounded by so many kind and generous people throughout this part of my life.

I express my sincere gratitude to my directors, Brian Wundheiler and Ralph Engel, for their guidance and support. I am equally grateful to my supervisors, Nicolás González and Markus Roth, for their dedication and invaluable advice.

I want to thank the members of my examination process committee: Juan Manuel Figueira, Andrés Arazi, Guido Drexlin and Torben Ferber, for their advice and their good treatment during such an important process.

I also thank Daniel Supanitsky, Diego Ravignani, David Schmidt, Darko Veberič, and Tanguy Pierog, whose insights and encouragement helped me at different stages of this work.

My heartfelt thanks go to María Cambon, Adrián Sedoski, Marie-Christine Mundt, Sabine Bucher, Anna Kendrick, Consuelo Vitón, Sabrina Coller and Tadeo Rodrguez for their constant kindness and willingness to help.

Vale Ferrari and Flor Thibaut, thank you for standing by me through difficult times.

I am deeply indebted to all the brilliant scientists and technicians of the Pierre Auger Collaboration for their inspiration and for pushing the boundaries of our understanding of cosmic rays.

Special thanks go to Margita Kubatová, Alexey Yushkov, and Steffen Hahn: the cross-calibration team! Working together with you made this thesis possible and reminded me that science is, above all, a collaborative endeavor.

To my lifelong friends, Anna and Lucio, thank you for always being there, near or far.

To my dear friends and companions on this journey: Marina, Ezequiel, Varada, Belén, Gabi, Joaquín, Fer, Rolando, José, Iván and Luca and to my friends on the German side: Sara, Emily, Fionna, Steffen, Karlita, Anton, Paul, Alan, Tobías, Olenna, Max, Jelena, Lukas, and Fabio, thank you for being like a family to me and for all the warmth, laughter, and support.

To my mother, Siomara, whose love and encouragement have always guided me; to my beloved brother, Mateo, who has stood by me in good and bad times; and to my sister, Catalina, who has given me strength when I could not find it myself. To my little ones (who are not so little anymore): Guada, Anto, and Vicen, thank you for your tenderness, joy, and love. To my aunt Simone, who is like a second mother to me; to my grandfather Dino, for his unwavering support; and to María, for caring for and loving my siblings and me.

And last but certainly not least, to my beautiful cat, Marie Curie, for being the best fluffy companion one could ask for.

Bibliography

[1] J. Albrecht, L. Cazon, H. Dembinski, A. Fedynitch, K.-H. Kampert, T. Pierog, W. Rhode, D. Soldin, B. Spaan, R. Ulrich, and M. Unger, "The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider," *Astrophysics and Space Science*, vol. 367, Mar. 2022.

[2] A. Castellina, "AugerPrime: the Pierre Auger Observatory upgrade," *EPJ Web of Conferences*, vol. 210, p. 06002, 2019.

[3] Pierre Auger Collaboration, "The Pierre Auger Observatory and its upgrade," *Science Reviews from the End of the World*, vol. 1, no. 4, pp. 1–26, 2020.

[4] Pierre Auger Collaboration, "Direct measurement of the muonic content of extensive air showers between 2×10^{17} and 2×10^{18} eV at the Pierre Auger Observatory," *Eur. Phys. J. C*, vol. 80, no. 8, p. 751, 2020.

[5] L. Evans, "The LHC machine," in *Proceedings of Science*, 2009.

[6] T. K. Gaisser, R. Engel, and E. Resconi, *Cosmic Rays and Particle Physics*. Cambridge University Press, 2016.

[7] J. Alvarez-Muñiz, Z. Cao, U. F. Katz, P. Mertsch, C. Spiering, and Particle Data Group, "Cosmic rays," *Phys. Rev. D*, vol. 110, p. 030001, 2024. Revised March 2024.

[8] F. W. Stecker and M. Salamon, "Photodisintegration of ultra-high-energy cosmic rays: A new determination," *The Astrophysical Journal*, 1999.

[9] L. N. Epele and E. Roulet, "Comment on the origin of the highest energy cosmic rays," *Phys. Rev. Lett.*, 1998.

[10] 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.

[11] O. Adriani *et al.*, "PAMELA measurements of cosmic-ray proton and helium spectra," *Science*, vol. 332, no. 6025, pp. 69–72, 2011.

[12] H. S. Ahn *et al.*, "Energy spectra of cosmic-ray nuclei at high energies," *The Astrophysical Journal*, vol. 707, p. 593, Nov. 2009.

[13] M. Aguilar *et al.*, "Precision measurement of the proton flux in primary cosmic rays from rigidity 1 GV to 1.8 TV with the Alpha Magnetic Spectrometer on the International Space Station," *Phys. Rev. Lett.*, vol. 114, p. 171103, Apr. 2015.

[14] R. J. Protheroe, "Acceleration and interaction of Ultra High Energy Cosmic Rays," in *Topics in Cosmic Ray Astrophysics* (M. A. DuVernois, ed.), New York: Nova Science Publishers, 1999. Also available as arXiv:astro-ph/9812055.

[15] E. Fermi, "Galactic magnetic fields and the origin of cosmic radiation," *The Astrophysical Journal*, 1954.

[16] A. M. Hillas, "The origin of ultra-high-energy cosmic rays," *Astronomy and Astrophysics*, 1984.

[17] 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 ultra-high energies," *Frontiers in Astronomy and Space Sciences*, vol. 6, p. 23, 2019.

[18] 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.

[19] A. Haungs *et al.*, "KCDC: the Kascade Cosmic Ray Data Centre," *Journal of Physics: Conference Series*, 2015.

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

[21] K.-H. Kampert and M. Unger, "Measurements of the cosmic ray composition with air shower experiments," *Astroparticle Physics*, vol. 35, pp. 660–678, May 2012.

[22] W. Heitler, *The Quantum Theory of Radiation*. Courier Corporation, 1984.

[23] Pierre Auger Collaboration, "Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events," *Phys. Rev. D*, vol. 91, no. 3, p. 032003, 2015.

[24] Pierre Auger Collaboration, "Measurement of the fluctuations in the number of muons in extensive air showers with the Pierre Auger Observatory," *Phys. Rev. Lett.*, vol. 126, no. 15, p. 152002, 2021.

[25] J. Albrecht, L. Cazon, H. Dembinski, A. Fedynitch, K.-H. Kampert, T. Pierog, W. Rhode, D. Soldin, B. Spaan, R. Ulrich, and M. Unger, "The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider," *Astrophysics and Space Science*, vol. 367, no. 3, 2022.

[26] J. Arteaga Velazquez, "A report by the WHISP working group on the combined analysis of muon data at cosmic-ray energies above 1 PeV," *PoS(ICRC2023)466*, 2023.

[27] Pierre Auger Collaboration, "Testing hadronic interactions at ultrahigh energies with air showers measured by the Pierre Auger Observatory," *Phys. Rev. Lett.*, vol. 117, p. 192001, Oct. 2016.

[28] H. Dembinski, J. Arteaga-Velázquez, L. Cazon, R. Conceição, J. Gonzalez, Y. Itow, D. Ivanov, N. Kalmykov, I. Karpikov, S. Müller, *et al.*, "Report on upgrades and measurements of hadronic interaction properties with air showers," *EPJ Web of Conferences*, 2019.

[29] J. R. Hörandel, "On the knee in the energy spectrum of cosmic rays," *Astroparticle Physics*, 2003.

[30] H. Dembinski, R. Engel, A. Fedynitch, T. Gaisser, F. Riehn, and T. Stanev, "Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV," 2017.

[31] B. G. Keilhauer, "Investigation of atmospheric effects on the development of extensive air showers and their detection with the Pierre Auger Observatory,". PhD thesis, Karlsruhe Institute of Technology, 2004.

[32] F. Convenga *et al.*, "Local-station calibration for dummies!." Auger internal note, not publicly available, 2023.

[33] D. Newton, J. Knapp, and A. A. Watson, "The optimum distance at which to determine the size of a giant air shower," *Astroparticle Physics*, vol. 26, no. 6, pp. 414–419, 2007.

[34] D. Veberič, B. Kegl, R. Engel, and M. Roth, "Constant Intensity Cut: Unbinned estimation of the signal attenuation function." Auger internal note, not publicly available, 2015.

[35] J. Hersil, I. Escobar, D. Scott, G. Clark, and S. Olbert, "Observations of extensive air showers near the maximum of their longitudinal development," *Phys. Rev. Lett.*, vol. 6, no. 1, p. 22, 1961.

[36] H. P. Dembinski, B. Kégl, I. C. Mariş, M. Roth, and D. Veberič, "A likelihood method to cross-calibrate air-shower detectors," *Astroparticle Physics*, vol. 73, pp. 44–51, 2016.

[37] J. Abraham *et al.*, "The Fluorescence detector of the Pierre Auger Observatory," *Nucl. Instrum. Meth. A*, vol. 620, no. 2, pp. 227–251, 2010.

[38] M. Ave *et al.*, "Precise measurement of the absolute fluorescence yield of the 337 nm band in atmospheric gases," *Astroparticle Physics*, vol. 42, pp. 90–102, 2013.

[39] C. Meurer and N. Scharf, "HEAT: A low-energy enhancement of the Pierre Auger Observatory," *arXiv preprint arXiv:1106.1329*, 2011.

[40] Pierre Auger Collaboration, "The dynamic range of the upgraded surface-detector stations of AugerPrime," Proceedings of Science, Sissa Medialab, Sept. 2024.

[41] Pierre Auger Collaboration, "Prototype muon detectors for the AMIGA component of the Pierre Auger Observatory," *Journal of Instrumentation*, vol. 11, no. 2, p. P02012, 2016.

[42] A. Pla-Dalmau, A. D. Bross, and V. V. Rykalin, "Extruding plastic scintillator at Fermilab," in *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No. 03CH37515)*, vol. 1, pp. 102–104, IEEE, 2003.

[43] A. M. Botti, "Determination of the chemical composition of cosmic rays in the energy region of 5 EeV with the AMIGA upgrade of the Pierre Auger Observatory,". PhD thesis, UNSAM, KIT, 2019.

[44] M. Platino, M. R. Hampel, A. Almela, A. Krieger, D. Gorbena, A. Ferrero, G. De La Vega, A. Lucero, F. Suarez, M. Videla, *et al.*, "AMIGA at the Auger Observatory: the scintillator module testing system," *Journal of Instrumentation*, vol. 6, no. 6, p. P06006, 2011.

[45] S. Argirò, S. L. C. Barroso, J. Gonzalez, L. Nellen, T. Paul, T. A. Porter, L. Prado Jr., M. Roth, R. Ulrich, and D. Veberič, "The $\overline{\text{Off}}\underline{\text{line}}$ software framework of the Pierre Auger Observatory," *Nucl. Instrum. Meth. A*, vol. 580, no. 3, pp. 1485–1496, 2007.

[46] E. Santos and for the Pierre Auger Collaboration, "Update on the $\overline{\text{Off}}\underline{\text{line}}$ framework for AugerPrime and production of reference simulation libraries using the VO Auger grid resources," in *Proceedings of the 38th International Cosmic Ray Conference (ICRC 2023)*, vol. PoS(ICRC2023)248, 2023.

[47] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw, *CORSIKA: A Monte Carlo code to simulate extensive air showers*. Forschungszentrum Karlsruhe GmbH, Karlsruhe, Germany, 1998. Report FZKA 6019.

[48] D. Heck and T. Pierog, *CORSIKA physics reference manual*. Forschungszentrum Karlsruhe GmbH, Karlsruhe, Germany, 1998. Report FZKA 6019.

[49] D. E. Groom, N. V. Mokhov, and S. I. Striganov, *Muon stopping-power and range tables 10 MeV–100 TeV*. 2001.

[50] B. Wundheiler, "Cosmic rays with energies between 10^{17} and 10^{19} eV,". PhD thesis, Universidad de Buenos Aires, 2014.

[51] P. H. Barrett, L. M. Bollinger, G. Cocconi, Y. Eisenberg, and K. Greisen, "Interpretation of cosmic-ray measurements far underground," *Rev. Mod. Phys.*, vol. 24, p. 133, 1952.

[52] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, *et al.*, "Geant4: A simulation toolkit," *Nuclear Instruments and Methods in Physics Research*, 2003.

[53] J. M. Figueira, "Simulations of AMIGA shielding, electromagnetic contamination, transmission and energy threshold." Presented at the internal meeting of the Pierre Auger Collaboration, 2016.

[54] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner, "EPOS LHC: Test of collective hadronization with LHC data," *Phys. Rev. C*, vol. 92, p. 034906, 2015.

[55] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft, "FLUKA: A multi-particle transport code," Tech. Rep. CERN-2005-010, SLAC-R-773, INFN-TC-05-11, CERN, SLAC, INFN, 2005.

[56] Pierre Auger Collaboration, "Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth," *Phys. Rev. D*, vol. 90, p. 012012, July 2014.

[57] J. Bellido and for the Pierre Auger Collaboration, "Depth of maximum of air-shower profiles at the Pierre Auger Observatory: Measurements above $10^{17.2}$ eV and composition implications," in *Proceedings of the 35th International Cosmic Ray Conference (ICRC 2017)*, vol. PoS(ICRC2017)506, 2017.

[58] M. Erfani, M. Risse, and M. Settimo, "Simulation study of muons in air showers (1)." Auger internal note, not publicly available, 2013.

[59] M. Erfani, M. Risse, M. Settimo, and A. Yushkov, "Simulation study of muons in air showers (2)." Auger internal note, not publicly available, 2014.

[60] E. dos Santos and A. Yushkov, "Extending the Naples CORSIKA shower library for auger studies at $16.5 \leq \log_{10}(e/\mathrm{eV}) \leq 18.0$." Auger internal note, not publicly available, 2018.

[61] J. de Jesús, J. M. Figueira, F. Sanchez, and D. Veberič, "Data-driven method to quantify and correct the corner-clipping effect in segmented muon counters," in *The 7th International Symposium on Ultra High Energy Cosmic Rays*, vol. PoS(UHECR2024)078, 2024.

[62] J. De Jesús, "Study of the mass composition of cosmic rays with the Underground Muon Detector of AMIGA,". PhD thesis, UNSAM, KIT, 2025.

[63] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with Python," in *Proc. of the 9th Python in Science Conf. (SciPy 2010)*, 2010.

[64] T. Pierog and K. Werner, "EPOS-LHC R: A global approach to solve the Muon Puzzle," 2025.

[65] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev, "Hadronic interaction model Sibyll 2.3d and extensive air showers," *Phys. Rev. D*, vol. 102, Sept. 2020.

[66] V. Verzi *et al.*, "The energy scale of the Pierre Auger Observatory," *Proceedings of the 33rd ICRC, Rio de Janeiro, Brazil*, 2013.

[67] F. Gesualdi and A. D. Supanitsky, "Estimation of the number of counts on a particle counter detector with full time resolution," *Eur. Phys. J. C*, vol. 82, p. 925, 2022.

[68] A. De Angelis and M. Pimenta, *Introduction to Particle and Astroparticle Physics: Multimessenger Astronomy and its Particle Physics Foundations*. Springer, 2 ed., 2018.

[69] A. Etchegoyen, "Track geometry and smearing of the bump calibration." Auger internal note, not publicly available, 2002.

[70] D. Supanitsky and X. Bertou, "Semi-analytical model of the three fold charge spectrum in a Water Cherenkov Detector." Auger internal note, not publicly available, 2003.

[71] B. Kégl and D. Veberič, "Single muon response: Tracklength." Auger internal note, not publicly available, 2009.

[72] T. Schulz, "Estimating mass composition with AugerPrime," Master's thesis, Karlsruher Institut für Technologie (KIT), 2020.

[73] A. Machado Payeras, "Potential of the AugerPrime surface detector for composition studies of high-energy cosmic rays,". PhD thesis, Universidade Estadual de Campinas, Instituto de Física Gleb Wataghin, 2023.

[74] S. Hahn, "Methods for estimating mass-sensitive observables of ultra-high energy cosmic rays using artificial neural networks,". PhD thesis, KIT and UNSAM, 2022.

[75] Pierre Auger Collaboration, "Extraction of the muon signals recorded with the surface detector of the Pierre Auger Observatory using recurrent neural networks," *Journal of Instrumentation*, vol. 16, p. P07016, July 2021.

[76] M. Kubatova, "Application of deep machine learning for studies of the muon air-shower component using the Pierre Auger surface detector data," Master's thesis, Czech Technical University in Prague, 2022.

[77] M. Kubatová, "Estimation of temporal muon signals in water-Cherenkov detectors of the surface detector of the Pierre Auger Observatory," 2025.

[78] P. Billoir, "Reconstruction of showers with the Ground Array: status of the "prototype" program." Auger internal note, not publicly available, 2000.

[79] P. Billoir, "Does the resampling procedure induce distortions in the FADC traces of the surface detector?." Auger internal note, not publicly available, 2005.

[80] C. Pérez Bertolli, F. Gollan, B. Wundheiler, and F. Sánchez, "Towards the development of a protocol for UMD shifts." Auger internal note, not publicly available, 2022.

[81] A. Coleman, "Measurement of the cosmic ray flux above 100 PeV at the Pierre Auger Observatory,". PhD thesis, The Pennsylvania State University, 2018.

[82] N. González, "Data quality observables for the Underground Muon Detector of AMIGA." Auger internal note, not publicly available, 2018.

[83] N. González, F. Sánchez, J. Figueira, A. Almela, A. Botti, A. Etchegoyen, A. Fuster, M. Hampel, A. Lucero, M. Platino, D. Ravignani, and B. Wundheiler, "On the muon lateral distribution using Underground Muon Detector data." Auger internal note, not publicly available, 2020.

[84] M. Scornavacche, "The muon content of extensive air showers measured by the Underground Muon Detector of the Pierre Auger Observatory,". PhD thesis, UNSAM, KIT, 2025.