

Search for PeV Photons with Machine Learning in the AugerPrime Era

For the attainment of the academic degree of

Doctorate in Science

from the

Karlsruher Instituts für Technologie (KIT)

and the

Universidad Nacional de San Martín (UNSAM)

approved

Dissertation

of

Lic. Ezequiel Enrique Rodriguez

from Ciudad Autónoma de Buenos Aires

Day of the oral examination: July 3rd, 2026

Referee: Prof. Dr. Brian Wundheiler

Co-referee: Prof. Dr. Ralph Engel

Supervisor: Dr. Nicolás González, Dr. Markus Roth



© 2026 Ezequiel Enrique Rodriguez

Except where otherwise noted, this work is licensed under the
Creative Commons Attribution 4.0 International License.

Search for PeV Photons with Machine Learning in the AugerPrime Era

Tesis aprobada para optar por el título de

Doctor en Astrofísica

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

y del

Karlsruher Instituts für Technologie (KIT)

por

Lic. Ezequiel Enrique Rodriguez

de Ciudad Autónoma de Buenos Aires

Fecha de la defensa oral: 3 de Julio de 2026

Director: Prof. Dr. Brian Wundheiler

Co-director: Prof. Dr. Ralph Engel

Colaborador: Dr. Nicolás González, Dr. Markus Roth

Search for PeV Photons with Machine Learning in the AugerPrime Era

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 der

Universidad Nacional de San Martín (UNSAM)

genehmigte

Dissertation

von

Lic. Ezequiel Enrique Rodriguez

aus Ciudad Autónoma de Buenos Aires

Tag der mündlichen Prüfung: der 3. Juli 2026

Referent: Prof. Dr. Brian Wundheiler

Korreferent: Prof. Dr. Ralph Engel

Betreuer: Dr. Nicolás González, Dr. Markus Roth

Abstract

Cosmic rays and ultra-high-energy photons provide complementary information on the origin, acceleration, and propagation of the most energetic particles in the Universe. In the energy range of a few tens of PeV, several models predict a diffuse photon flux hidden below the dominant background of hadron-induced extensive air showers. A measurement or improved constraint of this flux would provide insight into cosmic-ray interactions in the interstellar medium, the transition from Galactic to extragalactic cosmic rays, and possible exotic scenarios such as the decay of super-heavy dark matter. This thesis investigated the detector response, updated the relevant reconstruction sequences, and developed machine-learning methods required to improve PeV photon searches with the Pierre Auger Observatory in the AugerPrime era, exploiting the combined information of the Surface Detector (SD) and the Underground Muon Detector (UMD) in the 433 m configuration.

The first part of the work focused on hadron-induced showers, which constitute the dominant background for photon searches. The energy response of the SD-433 array was characterised in the region of the second knee, including its efficiency, energy bias, and energy resolution, providing the basis for quantitative comparisons between data and simulations for hadron-induced showers. Substantial effort was also devoted to the cleaning of the UMD data, including major improvements to the reconstruction sequence related to the rejection of faulty modules and the identification of spurious measurements.

The hadronic background was then characterised through a measurement of the muon content with the UMD-433. The mean shape of the muon lateral-distribution function (MLDF) was studied and found to be consistent with historical measurements of muons in the same energy range. An event-by-event muon observable was constructed from the event-level parametrisation of the MLDF and interpreted using detector-level proton and iron baselines. The resulting measurement around the second knee indicated a gradual evolution of the muon content from heavier to lighter primaries. The interpretation depends on the hadronic interaction model used to define the baselines, but the measured mean muon content is, on average, compatible with that expected for an oxygen-like composition over much of the studied range. The comparison with expectations based on shower-maximum measurements also suggests a residual muon deficit in simulations, although less severe than that inferred from some higher-energy observables. Comparisons with other experiments led to different levels of agreement, showing that compatibility depends strongly on the chosen composition references.

The second part of the thesis focused on photon-induced showers, corresponding to the signal component of the search. A dedicated reconstruction chain for photon-induced showers was developed, including a photon-specific lateral distribution function, geometry validation, and an energy calibration adapted to the predominantly electromagnetic nature of photon showers. In the reference phase space, the reconstructed photon energy was found

to be stable, with a mean bias consistent with zero and a relative resolution of about 14%. The energy response of the detector was therefore determined for both hadronic and photon showers, setting the baseline for the search analysis.

The central methodological contribution of this thesis was the development of a photon-hadron discriminator based on graph neural networks. Each event was represented as a detector graph constructed from the triggered surface stations and the available underground-muon information, allowing the classifier to exploit the spatial, temporal, and muonic structure of the shower footprint. A multi-view architecture, combining SD-only, UMD-only, and joint SD-UMD branches, was shown to provide strong discrimination power, with the joint branch yielding the best overall separation. Performance studies with proton and iron showers, together with tests under reduced detector information, indicated that the method was robust and largely interpretable. A comparison with a traditional observable employed by the Pierre Auger Collaboration for the same SD-433 and UMD-433 detector combination shows that, when an ideal UMD configuration is available, the gain in separation power is negligible. In that regime, adequate UMD area coverage already provides strong photon-hadron discrimination without the need for neural-network methods. The main advantage of the graph-based approach therefore lay in its robustness: it makes it possible to exploit less favourable SD-UMD configurations and thereby extend photon searches to a broader event selection.

Finally, the analysis chain was validated on a burn sample of real Phase I data as an intermediate step towards a photon search. The selected events populate the expected hadron-like region of the network-score distributions, with no unexpected instability observed as a function of energy, zenith angle, time, or score correlation. The graph-based selection doubled the number of usable burn-sample events relative to the previous SD-433+UMD selection, showing that the method can recover a substantially larger event sample while keeping the background contamination under control. Under a Phase I data unblinding scenario, the projected exposure achievable with this discriminator indicates that, in the absence of photon candidates, the existing photon limits in this energy range could be improved by approximately a factor of two.

Resumen

Los rayos cósmicos y los fotones de ultra alta energía proporcionan información complementaria sobre el origen, la aceleración y la propagación de las partículas más energéticas del Universo. En el rango de energía de unas pocas decenas de PeV, varios modelos predicen un flujo difuso de fotones oculto bajo el fondo dominante de cascadas atmosféricas extensas inducidas por hadrones. Una medición o una restricción más fuerte de este flujo permitiría obtener información sobre las interacciones de los rayos cósmicos en el medio interestelar, la transición entre rayos cósmicos galácticos y extragalácticos, y posibles escenarios exóticos, como el decaimiento de materia oscura superpesada. Esta tesis investigó la respuesta del detector, actualizó las secuencias de reconstrucción relevantes y desarrolló métodos de aprendizaje automático necesarios para mejorar las búsquedas de fotones de PeV con el Observatorio Pierre Auger en la era de AugerPrime, explotando la información combinada del Detector de Superficie (SD) y del Detector de Muones Subterráneo (UMD) en la configuración de 433 m.

La primera parte del trabajo se centró en las cascadas inducidas por hadrones, que constituyen el fondo dominante para las búsquedas de fotones. La respuesta en energía del arreglo SD-433 fue caracterizada en la región de la segunda rodilla, incluyendo su eficiencia, sesgo en energía y resolución energética, proporcionando la base para comparaciones cuantitativas entre datos y simulaciones de cascadas inducidas por hadrones. También se dedicó un esfuerzo considerable a la limpieza de los datos del UMD, incluyendo mejoras importantes en la secuencia de reconstrucción relacionadas con el rechazo de módulos defectuosos y la identificación de mediciones espurias.

El fondo hadrónico fue luego caracterizado mediante una medición del contenido muónico con el UMD-433. Se estudió la forma media de la función de distribución lateral de muones (MLDF), encontrándose que es compatible con mediciones históricas de muones en el mismo rango de energía. Se construyó un observable muónico evento por evento a partir de la parametrización de la MLDF a nivel de evento, y se interpretó utilizando referencias de protones y hierro a nivel de detector. La medición resultante alrededor de la segunda rodilla indicó una evolución gradual del contenido muónico desde primarios más pesados hacia primarios más livianos. La interpretación depende del modelo de interacción hadrónica utilizado para definir las referencias, pero el contenido muónico medio medido es, en promedio, compatible con el esperado para una composición similar al oxígeno en gran parte del rango estudiado. La comparación con expectativas basadas en mediciones del máximo de la cascada también sugiere un déficit residual de muones en las simulaciones, aunque menos severo que el inferido a partir de algunos observables muónicos a energías más altas. Las comparaciones con otros experimentos condujeron a distintos niveles de acuerdo, mostrando que la compatibilidad depende fuertemente de las referencias de composición elegidas.

La segunda parte de la tesis se centró en las cascadas inducidas por fotones, que corresponden a la componente de señal de la búsqueda. Se desarrolló una cadena de reconstrucción dedicada para cascadas inducidas por fotones, incluyendo una función de distribución lateral específica para fotones, una validación de la reconstrucción geométrica y una calibración de energía adaptada a la naturaleza predominantemente electromagnética de estas cascadas. En el espacio de fases de referencia, se encontró que la energía reconstruida de los fotones es estable, con un sesgo medio compatible con cero y una resolución relativa de aproximadamente 14%. Por lo tanto, la respuesta en energía del detector fue determinada tanto para cascadas hadrónicas como para cascadas fotónicas, estableciendo la base para el análisis de búsqueda.

La contribución metodológica central de esta tesis fue el desarrollo de un discriminador fotón–hadrón basado en redes neuronales de grafos. Cada evento fue representado como un grafo del detector construido a partir de las estaciones de superficie activadas y de la información disponible de muones subterráneos, permitiendo que el clasificador explote la estructura espacial, temporal y muónica de la huella de la cascada. Se mostró que una arquitectura de múltiples vistas, que combina ramas SD-only, UMD-only y conjunta SD–UMD, proporciona una fuerte capacidad de discriminación, siendo la rama conjunta la que produce la mejor separación global. Los estudios de rendimiento con cascadas de protones y hierro, junto con pruebas bajo condiciones de información reducida del detector, indicaron que el método es robusto y en gran medida interpretable. Una comparación con un observable tradicional empleado por la Colaboración Pierre Auger para la misma combinación de detectores SD-433 y UMD-433 muestra que, cuando se dispone de una configuración ideal del UMD, la ganancia en poder de separación es despreciable. En ese régimen, una cobertura adecuada de área del UMD ya proporciona una fuerte discriminación fotón–hadrón sin necesidad de métodos basados en redes neuronales. Por lo tanto, la principal ventaja del enfoque basado en grafos radica en su robustez: permite explotar configuraciones SD–UMD menos favorables y, de este modo, extender las búsquedas de fotones a una selección de eventos más amplia.

Finalmente, la cadena de análisis fue validada sobre una muestra de validación de datos reales de Phase I (*burnt sample*) como paso intermedio hacia una búsqueda de fotones. Los eventos seleccionados pueblan la región hadrónica esperada de las distribuciones de puntuación de la red, sin observarse inestabilidades inesperadas en función de la energía, el ángulo cenital, el tiempo o las correlaciones entre puntuaciones. La selección basada en grafos duplicó el número de eventos utilizables de la muestra de validación con respecto a la selección SD-433+UMD previa, mostrando que el método puede recuperar una muestra de eventos sustancialmente mayor mientras conserva un comportamiento controlado de la contaminación. Bajo un escenario de unblinding de los datos de Phase I, la exposición proyectada alcanzable con este discriminador indica que, en ausencia de candidatos a fotón, los límites existentes al flujo de fotones en este rango de energía podrían mejorarse aproximadamente en un factor dos.

Zusammenfassung

Kosmische Strahlung und ultrahochenergetische Photonen liefern komplementäre Informationen über den Ursprung, die Beschleunigung und die Propagation der energiereichsten Teilchen im Universum. Im Energiebereich von einigen zehn PeV sagen mehrere astrophysikalische Modelle einen diffusen Fluss von Photonen voraus, der unter dem überwiegenden Hintergrund aus hadronisch induzierten Luftschauern verborgen ist. Eine verbesserte Messsensibilität für diesen Fluss bietet Einblicke in die Wechselwirkungen kosmischer Strahlung im interstellaren Medium, in den Übergang von galaktischer zu extragalaktischer kosmischer Strahlung sowie in mögliche exotische Szenarien, wie den Zerfall superschwerer dunkler Materie. In dieser Arbeit wurden die Detektorantwort untersucht, die relevanten Rekonstruktionssequenzen aktualisiert und Methoden des maschinellen Lernens entwickelt, die zur Verbesserung von PeV-Photonensuchen am Pierre-Auger-Observatorium in der AugerPrime-Ära erforderlich sind. Dabei wurde die kombinierte Information des Oberflächendetektors (SD) und des unterirdischen Myonendetektors (UMD) in der 433 m Konfiguration genutzt.

Der erste Teil der Arbeit konzentrierte sich auf die Untersuchung hadronisch induzierter Luftschauer, die den dominierenden Hintergrund für Photonensuchen bilden. Die Energieantwort sowie Energieverzerrung des SD-433-Arrays wurde im Bereich des zweiten Knies charakterisiert, als auch die Messeffizienz bestimmt. Damit wurde die Grundlage für quantitative Vergleiche zwischen Daten und Simulationen hadroneninduzierter Schauer geschaffen. Ein erheblicher Aufwand wurde zudem der Bereinigung der UMD-Daten gewidmet, diese beinhaltet wesentliche Verbesserungen der Rekonstruktionssequenz im Zusammenhang mit der Zurückweisung fehlerhafter UMD-Module und der Identifikation fehlerhafter Messungen.

Der hadronische Hintergrund wurde anschliessend durch eine Messung des Myonengehalts mit dem UMD-433 charakterisiert. Die mittlere Form der lateralen Myon-Verteilungsfunktion (MLDF) wurde untersucht und als konsistent mit historischen Myonmessungen im gleichen Energiebereich befunden. Aus der Ereignisparametrisierung der MLDF wurde eine Myonobservablen konstruiert und mithilfe detektornaher Proton- und Eisenreferenzen interpretiert. Die resultierende Messung im Bereich des zweiten Knies zeigte eine graduelle Entwicklung des Myonengehalts von schwereren zu leichteren Primärteilchen. Die Interpretation dieses Zwischenergebnisses hängt vom hadronischen Wechselwirkungsmodell ab, das zur Definition der Referenzen verwendet wird; im Mittel ist der gemessene Myonengehalt jedoch über einen grossen Teil des untersuchten Bereichs mit dem vereinbar, der für eine Zusammensetzung von Sauerstoff oder vergleichbaren Elementen erwartbar ist. Der Vergleich mit Beobachtungen auf Basis von Messungen des Schauermaximums deutet zudem auf ein verbleibendes Myonendefizit von Luftschauer-Simulationen hin, wenn auch weniger ausgeprägt als dasjenige, das aus einigen Myonobservablen bei höheren Energien abgeleitet wird. Vergleiche mit anderen Experimenten führten zu unterschiedlichen Gra-

den der Übereinstimmung und zeigen, dass die Kompatibilität stark von den gewählten Zusammensetzungsreferenzen abhängt.

Der zweite Teil der Arbeit befasste sich mit photoneninduzierten Schauern, die der Signalkomponente der Suche entsprechen. Es wurde eine dedizierte Rekonstruktionskette für photoneninduzierte Schauer entwickelt, einschliesslich einer photonspezifischen lateralen Verteilungsfunktion, einer Validierung der geometrischen Rekonstruktion und einer Energiekalibration, die an die überwiegend elektromagnetische Natur von Photonenschauern angepasst ist. Im Referenzphasenraum erwies sich die rekonstruierte Photonenenergie nicht nur als stabil, sondern auch unverzerrt. Die relativen Auflösung der Kalibration beträgt etwa 14%. Damit wurde die Energieantwort des Detektors sowohl für hadronische als auch für photonische Schauer bestimmt und die Grundlage für die Suchanalyse geschaffen.

Der zentrale methodische Beitrag dieser Arbeit war die Entwicklung eines Photon-Hadron-Diskriminators auf Basis von Graph Neural Networks. Jedes Ereignis wurde als Detektorgraph dargestellt, der aus den beitragenden Sationen des SD und der verfügbaren Information über unterirdische Myonen konstruiert wurde. Dadurch konnte der Klassifikator die räumliche, zeitliche und myonische Struktur des Schauer-Fußabdrucks ausnutzen. Eine Multi-View-Architektur, die SD-, UMD- und den gemeinsamen SD-UMD-Zweig kombiniert, zeigte eine starke Diskriminationsleistung, wobei der gemeinsame Zweig die beste Gesamtrennung erreichte. Leistungsstudien mit Proton- und Eisenschauern sowie Tests unter reduzierter Detektorinformation zeigten, dass die Methode robust und weitgehend interpretierbar ist. Ein Vergleich mit einem traditionellen Observablen, das von der Pierre-Auger-Kollaboration für dieselbe SD-433- und UMD-433-Detektorkombination verwendet wird, zeigt, dass der Gewinn an Trennschärfe vernachlässigbar ist, wenn eine ideale UMD-Konfiguration verfügbar ist. In diesem Regime liefert eine ausreichende UMD-Flächenabdeckung bereits eine starke Photon-Hadron-Diskrimination, ohne dass neuronale Netze erforderlich sind. Der Hauptvorteil des graphenbasierten Ansatzes liegt daher in seiner Robustheit: Er ermöglicht es, unvollständige SD-UMD-Konfigurationen zu nutzen und dadurch Photonensuchen auf eine breitere Ereignisauswahl auszuweiten.

Abschliessend wurde die Analyseketten an einer Validierungsstichprobe realer Phase I-Daten (*burnt sample*) als Zwischenschritt hin zu einer Photonensuche überprüft. Die untersuchten Ereignisse bevölkern die erwarteten hadronähnlichen Bereiche der Netzwerk-Score-Verteilungen, ohne dass unerwartete Instabilitäten als Funktion der Energie, des Zenitwinkels, der Zeit oder der Score-Korrelationen beobachtet wurden. Die graphbasierte Auswahl verdoppelte die Anzahl nutzbarer Ereignisse in der Validierungsstichprobe im Vergleich zur vorherigen SD-433+UMD-Auswahl und zeigt damit, dass die präsentierte Methode eine deutlich grössere Ereignisstichprobe zurückgewinnen kann, während ein kontrolliertes Score-Verhalten erhalten bleibt. Unter einem Phase I-Entblindungsszenario deutet die mit diesem Diskriminator erreichbare projizierte Belichtung darauf hin, dass die bestehenden Photonenlimits in diesem Energiebereich, sofern keine Photonenkandidaten gefunden werden, um ungefähr einen Faktor zwei verbessert werden könnten.

To those who dare to pursue their curiosity

Contents

1	Introduction	1
2	Cosmic rays and gamma rays	3
2.1	Dawn of cosmic-ray research	3
2.2	Energy spectrum of cosmic rays	7
2.3	Ultra-High-Energy photons	12
3	Extensive air showers	21
3.1	Basic notions of cascade theory	21
3.2	Particle content and the Heitler-Matthews model	26
3.3	Simulation of extensive air showers	29
3.4	Features of hadronic- and gamma-induced air showers	32
4	The Pierre Auger Observatory	35
4.1	Auger early years	35
4.2	Surface Detector	37
4.3	SD Reconstruction	42
4.4	Fluorescence detector	45
4.5	AugerPrime	48
4.6	Underground Muon Detector	50
4.7	UMD Reconstruction	54
4.8	<u>Offline</u> Framework	60
5	SD-433 energy response	61
5.1	Efficiency of the array	63
5.2	Energy response	75
5.3	Summary	80
6	Muon content	83
6.1	Shape of the mean Muon Lateral Distribution Function	84
6.2	Muon Lateral Distribution Function for the UMD-433	90
6.3	The observable ρ_{21}	95
6.4	Compatibility with the UMD-750	99
6.5	Simulated baselines	101
6.6	Muon content around the second knee	104
6.7	Summary	107

7	Photon reconstruction	111
7.1	Efficiency	112
7.2	Lateral Distribution Function	113
7.3	Geometry reconstruction	119
7.4	Energy calibration	122
7.5	Summary	128
8	Photon-hadron discrimination	131
8.1	Formal statement of the classification problem	132
8.2	Event representation as graphs	133
8.3	Network architecture	136
8.4	Training stage	142
8.5	Discrimination performance	144
8.6	Comparison with M_1	149
8.7	Robustness	153
8.8	Explainability	154
8.9	Summary	160
9	Towards a photon search	163
9.1	Introduction	163
9.2	Validation on a burn sample	164
9.3	Search prospects	169
9.4	Systematic uncertainties of the scores	173
9.5	Summary	175
10	Summary and outlook	177
	Acronyms	193
	List of Tables	195
	List of Figures	197
A	Supplementary material for Chapter 5	203
A.1	Generalized Linear Model for trigger efficiency	203
A.2	Energy Scale Transformation	203
A.3	Fitted model for $\epsilon(E_{SD-433}, \theta)$	204
A.4	CIC consistency check	204
A.5	Energy resolution from joint SD-433 and SD-750 events	205
B	Supplementary material for Chapter 6	211
B.1	Data quality and updates of the reconstruction sequence	211
B.2	Technical details of pinocchio rejection	222
B.3	Shape of the mean MLDF validation	234
B.4	ρ_μ fluctuations	236
B.5	Alternative β parametrisation	237
B.6	Attenuation correction	239
B.7	Systematic uncertainties	241
B.8	Effect of the folding correction for $\ln\langle A \rangle$ and the z-scale	241
C	Statement on AI Use	245

Chapter 1

Introduction

- Where shall I begin, please your Majesty? – he asked.
- Begin at the beginning, – the King said gravely, – and go on till you come to the end: then stop.

Alice's Adventures in Wonderland, Lewis Carroll

Cosmic rays are highly energetic particles of extraterrestrial origin, composed primarily of protons and heavier nuclei, with a smaller contribution from electrons. Their energies span a vast range, from billions of electronvolts to beyond 10^{20} eV, making them a powerful probe of a wide variety of astrophysical phenomena. In the energy range from a few tens to a few hundred PeV, several models predict the existence of a diffuse photon flux concealed beneath the dominant cosmic-ray background. Constraining or measuring this diffuse flux can provide valuable information on the propagation of cosmic rays through the interstellar medium, their interactions with matter in the Galactic disc, and possible exotic scenarios such as the decay of super-heavy dark matter. A precise determination of this flux is therefore important both for understanding the interplay between cosmic rays and their astrophysical environments and for probing potential physics beyond the Standard Model.

The goal of this thesis is to improve photon searches at energies of a few tens of PeV at the Pierre Auger Observatory through the use of machine-learning techniques. This effort exploits the direct muon-detection capabilities of the Underground Muon Detector (UMD), one of the key components of AugerPrime, the upgrade of the Pierre Auger Observatory. The methodology developed in this work is intended to increase both the sensitivity and the robustness of gamma-ray detection in an energy range that is particularly relevant for understanding the transition between Galactic and extragalactic sources.

We begin by introducing the theoretical and experimental background needed to frame the objectives of this thesis. Chapter 2 presents the fundamentals of cosmic rays and gamma rays, including their physical properties, detection techniques, and the current state of the art in high-energy astrophysics. Chapter 3 reviews the physics of Extensive Air Shower (EAS), describing their development in the atmosphere, the production of secondary particles, and the simulation frameworks used to model these processes. Chapter 4 presents the Pierre Auger Observatory, its instrumentation, and the motivation for the AugerPrime upgrade, with particular emphasis on the UMD. It also introduces the Observatory's simulation and reconstruction framework, *Offline*, which plays a central role throughout this thesis.

We then turn to the main analyses and methodological developments of the thesis. Chapter 5 studies the energy response of the lowest-energy extension of the Surface Detector (SD), known as SD-433, in order to characterise its efficiency and sensitivity to cosmic rays. Chapter 6 investigates the mass composition of cosmic rays around the second *knee*,

providing important insight into the background against which photon searches must be performed. Chapter 7 examines the detector response to photon-induced showers, with particular attention to the array efficiency and the reconstruction sequence, addressing both the geometric reconstruction and the energy response. Chapter 8 introduces a photon–hadron discrimination method based on graph neural networks, combining the temporal signal traces recorded by the water-Cherenkov detectors with the muon information measured by the UMD. Finally, Chapter 9 validates the application of these methods on a burn sample of Phase I data and evaluates their implications for an eventual full-data unblinding. The results are discussed with particular emphasis on the exposure that could be recovered by using a broader class of events with incomplete UMD information.

Each chapter concludes with a brief summary that highlights the main results, and outlines possible directions for future work where necessary. The overall conclusions of the thesis are presented in Chapter 10, together with a discussion of their implications and prospects for future developments. In brief, the two main results of this thesis are:

- An evolution of the inferred mass composition from heavier to lighter primaries was identified around the second *knee*, in qualitative agreement with expectations for the Galactic-to-extragalactic transition; in addition, the mean measured muon content was found to be broadly compatible with oxygen-like primaries over much of the studied range, when averaged over the hadronic interaction models considered, although the exact interpretation remains model-dependent;
- A photon–hadron classifier based on graph neural networks was developed and shown to be sufficiently robust to extend photon searches with the combined SD–UMD detector configuration to a broader class of events than previously considered. Under a Phase I data unblinding scenario, the projected exposure achievable with this method indicates that, provided no photon candidates are found, the existing photon upper limits in this energy range could be strengthened by roughly a factor of two, corresponding to numerical limits about one half as large.

Cosmic rays and gamma rays

There is a theory which states that if ever anybody discovers exactly what the Universe is for and why it is here, it will instantly disappear and be replaced by something even more bizarre and inexplicable.

The Restaurant at the End of the Universe, Douglas Adams

2.1 Dawn of cosmic-ray research

The study of cosmic rays began at the turn of the 20th century, a period of intense scientific curiosity about the origins and nature of ionising radiation. This research laid the groundwork for understanding high-energy particles from outer space and led to significant developments in both astrophysics and particle physics. The discovery of radioactivity by Henri Becquerel in 1896, a landmark in the history of physics, can be regarded as the catalyst that ignited the dawn of cosmic-ray research.

Becquerel, while studying phosphorescent materials, observed that uranium salts emitted radiation capable of fogging photographic plates, even when shielded from light [1]. This unexpected result unveiled a previously unknown, penetrating form of radiation emanating from certain elements. Building on this work, Marie and Pierre Curie systematically studied radioactive substances and successfully isolated radium and polonium from pitchblende ore [2]. Their meticulous experiments not only quantified radiation intensity but also introduced the term *radioactivity* to describe this phenomenon. These investigations deepened the understanding of atomic structures and equipped scientists with tools to probe the mysterious radiation permeating Earth's atmosphere.

C.T.R. Wilson conducted groundbreaking experiments in the early 1900s to understand the ionisation of the atmosphere. He hypothesised that the observed ionisation might be caused by an external penetrating radiation and developed the cloud chamber to visualise ionisation tracks [3]. This invention not only provided direct evidence for ionisation but also became a critical tool in particle physics. Theodor Wulf refined this line of inquiry by constructing an electrometer sensitive enough to measure radiation at different altitudes. In 1910, Wulf conducted measurements at the base and summit of the Eiffel Tower. Contrary to expectations, he observed that the radiation intensity decreased less than predicted with altitude [4]. This finding suggested that a significant portion of the radiation might have an extraterrestrial origin, challenging the prevailing view that radiation was entirely terrestrial.



Figure 2.1: Victor Hess on a gondola in 1912, most likely on a test flight. The exact date and location are unknown. Taken from Ref. [7].

Victor Hess's balloon flights in 1911 and 1912 provided definitive evidence for the extraterrestrial origin of this radiation. Throughout 1911 and 1912, he completed a total of seven campaigns aboard helium balloons. Hess ascended to over 5000 metres equipped with highly sensitive detectors and discovered that radiation intensity increased significantly with altitude [5]. His experiments during a solar eclipse further excluded the Sun as the primary source [6]. Hess's findings earned him the Nobel Prize in Physics in 1936 and established the foundation of cosmic-ray research. Fig. 2.1 shows a photograph of a young Victor Hess in a balloon gondola in 1912.

In 1927, J. Clay's observations uncovered the latitude effect [8]. Measurements were conducted during voyages from Java, Indonesia, to the Netherlands, documenting variations in cosmic-ray intensity as a function of magnetic latitude. This study confirmed the charged nature of cosmic rays and their interaction with Earth's magnetic field. This foundational discovery highlighted how Earth's geomagnetic field modulates cosmic rays, deflecting lower-energy particles more effectively at equatorial regions.

Building on Hess's discoveries, E. Steinke conducted extensive absorption experiments in the early 1930s that correlated cosmic-ray intensity with atmospheric conditions. His findings demonstrated a strong anti-correlation between cosmic-ray intensity and atmospheric pressure, indicating that increased pressure reduced the penetration of cosmic rays through the atmosphere [9]. Additionally, Steinke observed an annual modulation in cosmic-ray intensity, which he attributed to variations in Earth's position relative to the Sun and



Figure 2.2: Regener (*right*) demonstrates his balloon electrometer to Hess (*left*) and Steinke (*centre*). Immenstaad/Lake Constance, August 1932. Taken from Ref. [11].

the heliospheric environment. These modulations provided early evidence of the Sun's influence on cosmic-ray propagation. Furthermore, his research identified a sidereal modulation, showing variations in intensity based on Earth's orientation relative to the fixed stars, suggesting potential anisotropy in cosmic-ray arrival directions. Steinke's work expanded the understanding of cosmic-ray interactions with Earth's atmosphere and highlighted the complex factors affecting their detection.

Around the same time, Erich Regener conducted experiments to study the absorption of cosmic rays using an ionisation chamber. His apparatus was designed to measure cosmic-ray intensity underwater and was deployed in Lake Constance [10]. Regener's experiments involved lowering the ionisation chamber into the lake to depths down to 250 metres, recording radiation intensity at various depths. His results provided one of the earliest absorption curves for cosmic rays, showing how their intensity diminished as a function of water depth, measured in metres of water equivalent (m.w.e). Regener interpreted his findings as strong evidence of the penetrating nature of cosmic rays and suggested that their absorption followed an exponential law. His meticulous measurements laid critical groundwork for understanding the interaction of cosmic rays with matter and their attenuation properties, influencing future cosmic-ray research. Figure 2.2 shows Regener interacting with Hess and Steinke during the measurements at Lake Constance.

The 1930s saw a leap in detection techniques with W. Bothe and W. Kolhörster's introduction of the coincidence method. This technique involved using two or more Geiger-Müller counters connected in such a way that they would only register an event if particles passed through all counters simultaneously. This approach significantly improved the accuracy of cosmic-ray measurements by filtering out background noise and enabling the identification of correlated particle interactions [12]. Bothe was later awarded the Nobel Prize in Physics in 1954 for his development of this method, which became a cornerstone in experimental

particle physics. High-altitude balloon flights led by Kolhörster extended measurements into uncharted territories, further advancing the understanding of cosmic rays [13].

Meanwhile, Bruno Rossi's studies of the East-West effect provided crucial insights into the nature of cosmic rays [14]. Rossi demonstrated that cosmic rays arriving from the west were more intense than those from the east, a phenomenon attributed to the deflection of positively charged particles by Earth's magnetic field. Notably, Rossi conducted these experiments in Asmara, Eritrea, taking advantage of its low geomagnetic latitude to maximise the observable effect [15]. This finding offered compelling evidence that cosmic rays are predominantly positively charged and highlighted the role of geomagnetic effects in their propagation.

The discovery of extensive air showers, large cascades of secondary particles produced when high-energy cosmic rays interact with the atmosphere, marked a significant milestone in cosmic-ray research. In 1937, Pierre Auger and his team studied the phenomenon of long-range coincidences involving muons and electrons, now recognised as air showers [16]. Their observations confirmed the presence of large-scale air showers and established that the primary cosmic rays responsible must possess energies exceeding 10^{15} eV. Additionally, they identified that the differential energy spectrum of cosmic rays follows a power-law behaviour, approximately described by $dN/dE \propto E^{-2}$. This research not only set the stage for contemporary detection methodologies but also highlighted the challenge of explaining the origins of such extreme energies, with Auger noting the difficulty of conceiving a single process capable of imparting such energy to a particle.

In 1947, Cecil Powell and collaborators uncovered the pion, a particle central to the theory of the strong nuclear force, using photographic emulsions exposed at high altitudes [17]. Subsequent studies in the 1950s identified additional mesons, including kaons and hyperons, through experiments relying on cosmic rays as natural particle accelerators. These discoveries laid the foundation for modern particle physics, enabling the exploration of particle interactions at energies far beyond what early accelerators could achieve. Cosmic-ray studies also enabled the exploration of high-energy phenomena, catalysing the discovery of the tau neutrino, charm quarks, and gluons in later decades with man-made accelerators and laboratories.

The first ground-based detector array specifically designed to study extensive air showers was developed in the 1950s under the leadership of Bruno Rossi [18]. The pioneering array in Harvard Agassiz Station consisted of 11 scintillators on a triangular grid covering 0.17 km^2 , and utilised synchronised detectors to measure the lateral spread of air showers, enabling the reconstruction of the energy and direction of primary cosmic rays. Rossi's work laid the foundation for modern observatories such as the Pierre Auger Observatory, significantly advancing the experimental capabilities in high-energy astrophysics. The work at Agassiz was the forerunner for the arrays of Chacaltaya Observatory and Volcano Ranch.

In the post-1950s era, cosmic-ray studies transitioned to the discovery and characterisation of cosmic-ray composition. John Simpson's development of space-based detectors in the late 1950s marked a pivotal shift in the precision measurement of cosmic-ray elemental and isotopic composition [19]. The Voyager missions in the 1970s and 1980s extended cosmic-ray research beyond the heliosphere, providing insights into the interstellar propagation of cosmic rays by measurement of composition abundances [20].

The early history of cosmic-ray research illustrates a remarkable interplay between observation and theory. From the revolutionary measurements of radiation to the discovery of fundamental particles, these efforts not only advanced astrophysics but also transformed our understanding of the building blocks of matter.

2.2 Energy spectrum of cosmic rays

The cosmic-ray spectrum observed at Earth spans a vast energy range, from approximately $\sim 10^9$ eV to beyond 10^{20} eV. Studying its features offers a natural entry point into the field of cosmic-ray research, as it encapsulates many of the central questions and challenges in the discipline. Explaining the energy distribution of these particles provides critical insights into the physical mechanisms that govern their origin, acceleration, and propagation through the interstellar medium.

The spectrum can be described by a power law, with the differential flux expressed as

$$\frac{dJ}{dE} = \frac{d^4N}{dt dA d\Omega dE} \propto E^{-\alpha}, \quad (2.1)$$

where α is the spectral index. From 10^9 eV to the suppression region above 10^{20} eV, α is roughly equal to 3. Due to the steep fall-off of the flux with increasing energy, the spectrum is often plotted multiplied by a power of E , which highlights deviations from pure power-law behaviour, often referred to as spectral features. This approach enhances the visibility of structures and transitions in the energy spectrum, as depicted in Fig. 2.3, taken from the cosmic-ray review of the Particle Data Group in 2024.

At the lower energy regime, such as $\sim 10^9$ eV, cosmic rays are abundant, with rates of thousands of particles per square metre per second. As the energy increases in the knee region around 10^{15} eV, the rate drops significantly to about one particle per square metre per year. In the ultra-high-energy (UHE) regime, above 10^{18} eV, particles become exceedingly rare, with fluxes corresponding to approximately one particle per square kilometre per year. At the highest energies, approaching the suppression region above 4×10^{19} eV, the expected rate decreases further to one particle per square kilometre per century. This dramatic decrease in flux underscores the challenge of detecting high-energy cosmic rays and necessitates the use of large-scale experiments and extensive observation times to study them.

The spectrum displays several distinct features across its energy range, including the *knee* at 10^{15} eV, the *second knee* around 10^{17} eV, the *ankle* above 10^{18} eV, and the suppression at 4×10^{19} eV. Intermediate structures, such as the *insep*, are also present and provide additional information about the energy transitions. The astrophysical interpretation of these features, including their implications for cosmic-ray origins and propagation, will be discussed in the following subsections.

2.2.1 Cosmic-ray propagation

Cosmic-ray propagation is crucial for understanding the origin, interactions, and arrival directions of cosmic rays at Earth. Propagation occurs through complex processes governed by magnetic fields, particle diffusion, and energy loss mechanisms, both within the Galaxy and in extragalactic space. We first introduce the intuitive leaky-box model and then turn to the transport equation, Galactic and extragalactic propagation mechanisms, and the role of magnetic fields.

2.2.1.1 Leaky-box model

The leaky-box model provides an intuitive framework for understanding how the energy spectrum of cosmic rays evolves as they propagate diffusively through the Galaxy. In this model, the Galaxy is treated as a box in which cosmic rays are confined, with a constant

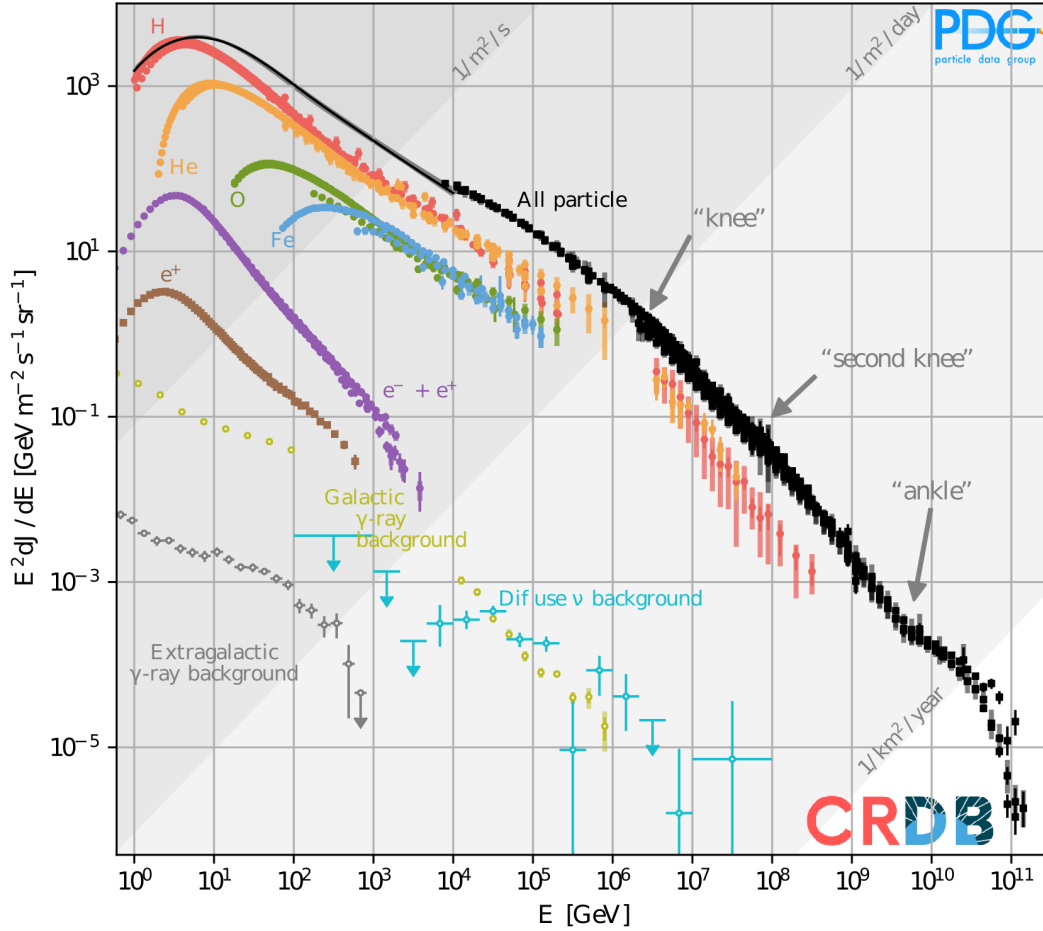


Figure 2.3: The spectrum of cosmic rays. Shown are measurements of the intensity of charged and neutral CRs, multiplied by kinetic energy squared. The data for the charged cosmic rays have been extracted from the Cosmic Ray Database [21]. Below 10^4 GeV, the all-particle spectrum is the sum of spline fits of the most important nuclear species. The figure contains diffuse gamma-ray fluxes and measurements of the diffuse neutrino background. Energy-integrated intensities are indicated by the various diagonal lines. Figure taken from Ref. [22].

probability of escape per unit time [23]. The stationary solution balances production and escape, leading to a steady-state density $n(E)$:

$$Q(E) = \frac{n(E)}{\tau_{\text{esc}}}, \quad (2.2)$$

where $Q(E)$ is the source term, τ_{esc} is the escape time, and $n(E)$ is the cosmic ray density. If the source spectrum follows a power law, $Q(E) \propto E^{-\alpha}$, and escape time τ_{esc} depends on energy as $\tau_{\text{esc}} \propto E^{-\delta}$, the resulting observed spectrum steepens to $n(E) \propto E^{-(\alpha+\delta)}$. For example, if the source spectrum has $\alpha = 2$ and diffusion dominates with $\delta = 0.7$, the observed spectrum at Earth becomes $n(E) \propto E^{-2.7}$. This steepening reflects the impact of diffusive propagation and escape from the Galaxy.

While the leaky-box model captures essential features of cosmic-ray propagation, it simplifies spatial and temporal variations. We then move to the transport equation, which accounts for diffusion, energy losses, and interactions in a unified framework.

2.2.1.2 Transport equation

The propagation of cosmic rays is quantitatively described by the transport equation, which accounts for various physical processes that affect their density and energy distribution [24, 25]. The equation is given as

$$\begin{aligned} \frac{\partial n(E, \vec{r}, t)}{\partial t} = & \nabla \cdot [D(E, \vec{r}) \nabla n(E, \vec{r}, t)] - \frac{\partial}{\partial E} [b(E) n(E, \vec{r}, t)] + Q(E, \vec{r}, t) \\ & - \frac{n(E, \vec{r}, t)}{\tau_{\text{decay}}} - \sigma_{\text{inel}}(E) n(E, \vec{r}, t) + \int_E^\infty \frac{n(E', \vec{r}, t)}{\tau_{\text{cascade}}(E', E)} dE'. \end{aligned} \quad (2.3)$$

Here, the left-hand side represents the time evolution of the cosmic ray density, while the right-hand side comprises several terms that shape the energy spectrum:

1. **Diffusion term:** $\nabla \cdot [D(E, \vec{r}) \nabla n(E, \vec{r}, t)]$ describes the spread of cosmic rays due to diffusion in a turbulent magnetic field. The diffusion coefficient $D(E, \vec{r})$ typically increases with energy as $D(E) \propto E^\delta$, where $\delta \sim 0.3\text{--}0.7$, depending on the turbulence regime. At higher energies, particles diffuse more rapidly, leading to a flatter spectrum in regions closer to the sources and a steeper spectrum at larger distances due to escape.
2. **Continuous energy loss term:** $-\frac{\partial}{\partial E} [b(E) n(E, \vec{r}, t)]$ accounts for continuous energy losses during propagation, such as ionisation, synchrotron radiation, and adiabatic expansion. The energy loss rate $b(E)$ depends on the interaction processes involved. Energy losses deplete high-energy particles, causing a softening of the spectrum over time and distance from the source.
3. **Source term:** $Q(E, \vec{r}, t)$ represents the injection of particles into the system by astrophysical sources like supernova remnants or active galactic nuclei. The source spectrum is typically modelled as a power law, $Q(E) \propto E^{-\alpha}$. The injection spectrum sets the initial shape of the energy distribution, which is later modified by propagation effects.
4. **Radiative decay term:** $-\frac{n(E, \vec{r}, t)}{\tau_{\text{decay}}}$ describes the loss of particles due to radiative decay with a characteristic decay time τ_{decay} . It is significant for unstable particles such as secondary cosmic rays.
5. **Inelastic interaction term:** $-\sigma_{\text{inel}}(E) n(E, \vec{r}, t)$ accounts for energy losses and particle depletion due to inelastic interactions with the interstellar medium. The cross-section $\sigma_{\text{inel}}(E)$ depends on the energy of the cosmic rays and the density of the medium.
6. **Cascade (or fed-down) term:** $\int_E^\infty dE' \frac{n(E', \vec{r}, t)}{\tau_{\text{cascade}}(E', E)}$ represents the production of particles at energy E by the decay or interaction of particles at higher energies E' . The timescale τ_{cascade} depends on the interaction or decay processes responsible for the cascading.

Additional terms, such as reacceleration due to magnetic turbulence, can also be included for a more comprehensive model. These effects may modify the energy spectrum and spatial distribution of cosmic rays, depending on the astrophysical environment.

2.2.1.3 Galactic and extragalactic propagation

In the Galaxy, cosmic rays propagate through turbulent magnetic fields with an average strength of $\sim 3 \mu\text{G}$ in the disc and comparable values in the halo. The diffusion coefficient follows:

$$D(E) \propto E^\delta, \quad (2.4)$$

where δ typically ranges from 0.3 to 0.7. Cosmic rays are confined within the Galactic halo, whose scale height ranges from 1 to 10 kpc. The residence time for GeV particles is approximately 2×10^7 years. During their confinement, cosmic rays interact with the interstellar medium (ISM), producing secondary particles, isotopes, and gamma rays, which provide indirect evidence of their propagation.

Cosmic rays escaping the Galaxy enter extragalactic space, where their propagation is governed by interactions with cosmic radiation fields, magnetic fields, and the expansion of the universe. These processes significantly alter their energy spectrum and observable properties.

Extragalactic magnetic fields are generally weaker than Galactic fields, with typical strengths of $B \sim 1$ nG in voids and up to $B \sim 1$ μ G in clusters and filaments. These fields cause cosmic rays to deviate from their original trajectories, with the deflection angle determined by their rigidity and the coherence length of the magnetic field. The Larmor radius, which quantifies the scale of deflection, is given by:

$$r_L = \frac{R}{B}, \text{ with } R = \frac{pc}{Ze}. \quad (2.5)$$

Here, R is the rigidity, B is the magnetic-field strength, p is the momentum, c is the speed of light, and Z is the atomic number. For high-rigidity particles, deflections are minimal, while low-rigidity particles experience significant isotropisation.

Cosmic rays also interact with the cosmic microwave background (CMB) and extragalactic background light (EBL), with the relevant propagation processes depending strongly on the mass and Lorentz factor of the primary particle. Protons lose energy through Bethe-Heitler pair production and, at the highest energies, through photopion production. The latter process gives rise to the conventional Greisen-Zatsepin-Kuzmin (GZK) effect [26, 27] and occurs predominantly through:

$$p + \gamma_{\text{CMB}} \rightarrow \Delta^+ \rightarrow \begin{cases} p + \pi^0 & (\text{neutral pion channel}), \\ n + \pi^+ & (\text{charged pion channel}). \end{cases} \quad (2.6)$$

Here, p is a proton, γ_{CMB} is a CMB photon, and Δ^+ is the intermediate resonant state. The effective threshold for photopion production on the CMB is approximately $E_{\text{GZK}} \sim 5 \times 10^{19}$ eV. Above this energy, the proton energy-loss length decreases to a few tens of Mpc, substantially limiting the distance from which protons can reach Earth.

For heavier nuclei, the corresponding attenuation is governed mainly by photodisintegration through interactions with the CMB and EBL [28, 29]. This process can be represented schematically as

$$\frac{A}{Z}X + \gamma_{\text{bg}} \rightarrow \frac{A'}{Z'}X + N_1 + \cdots + N_k, \quad A' < A, \quad (2.7)$$

where γ_{bg} represents a background photon and the N_i are emitted nucleons. At the giant-dipole resonance, one or a few nucleons are typically removed from the parent nucleus. The Lorentz factor is approximately conserved, such that the resulting fragments retain a similar energy per nucleon but have a lower total energy and mass. Repeated interactions can therefore generate a nuclear cascade, progressively transforming heavy primaries into lighter nuclei and secondary nucleons. At sufficiently high Lorentz factors, nuclei can also undergo photopion production through interactions involving their constituent nucleons.

These propagation effects produce a composition-dependent suppression of the UHE cosmic-ray flux. Strictly speaking, the GZK effect refers to the photopion attenuation of protons, while photodisintegration produces an analogous GZK-like attenuation for nuclei. The resulting cosmic-ray horizon depends on the primary species, energy, and radiation backgrounds, but can be reduced to distances of a few tens to approximately one hundred

Mpc at the highest energies. Early measurements by the Akeno Giant Air Shower Array (AGASA) [30] suggested the absence of a clear suppression, whereas more precise measurements by the Pierre Auger Observatory [31] and the Telescope Array (TA) [32] later established the presence of a suppression region.

The physical origin of the observed suppression cannot, however, be inferred from the energy spectrum alone. For a proton-dominated composition, it may be interpreted primarily as the conventional GZK effect, whereas for a mixed or heavier composition, photodisintegration becomes essential. In addition, a rigidity-dependent maximum acceleration energy at the sources can produce a similar spectral feature through a Peters-cycle scenario [33]. The observed suppression may therefore result from a combination of propagation effects and the maximum energies attainable at the sources. The combined effects of energy losses and magnetic deflections determine the observable propagation distance of cosmic rays. The resulting, composition-dependent horizon provides a window into the nearby universe and constrains both the distribution of the sources and the mass composition they inject.

2.2.2 Sources and acceleration mechanisms

The acceleration of cosmic rays is a fundamental process shaping their observed energy spectrum and providing insight into their astrophysical origins. Therefore, understanding which types of processes and environments could produce a power-law energy spectrum is imperative.

In systems where particles gain energy and have a probability of escape per interaction, a power-law spectrum naturally forms. Assume a particle with initial energy E_0 undergoes multiple cycles of acceleration, gaining energy by a factor $\kappa > 1$ per cycle, and escaping the system with probability P_{esc} . The particle energy after n cycles is:

$$E_n = E_0 \kappa^n. \quad (2.8)$$

The number of particles remaining after n cycles is:

$$N_n = N_0 (1 - P_{\text{esc}})^n. \quad (2.9)$$

where N_0 is the initial particle count. Eliminating n yields the energy distribution:

$$N(E) \propto E^{-\alpha}, \quad \text{where } \alpha = 1 + \frac{\ln(1/P_{\text{esc}})}{\ln(\kappa)}. \quad (2.10)$$

This derivation shows that the power-law index α depends on the energy gain κ and the escape probability P_{esc} , and it aligns with the observed cosmic-ray spectrum after accounting for propagation effects.

A good environment for this type of system is found in astrophysical shock-waves. In astrophysics, a shock-wave is a region where the flow of plasma undergoes sudden changes in density, velocity, and temperature. Such shocks are natural sites for particle acceleration. Two types of the so-called *Fermi acceleration* [34, 35] operate in these environments:

1. **First-order Fermi acceleration:** This mechanism occurs at strong shock fronts, where particles repeatedly cross the shock, gaining energy at each crossing due to the relative velocity difference between the upstream and downstream regions. The energy gain per cycle is proportional to $\Delta E/E \sim \Delta v/c$, where Δv is the velocity difference. The resulting spectrum follows a power law with an index $\alpha \sim 2$ for strong shocks.
2. **Second-order Fermi acceleration:** This mechanism operates in turbulent regions where particles scatter off moving magnetic clouds, gaining energy through random interactions. The energy gain per cycle is proportional to $(v/c)^2$, making it less efficient than first-order acceleration.

The high efficiency and simplicity of first-order Fermi acceleration make it a primary candidate for producing the observed power-law spectrum, which remains smooth and featureless, consistent with observations.

Supernova remnants (SNRs) are the most plausible candidates for Galactic cosmic-ray acceleration. Their shock waves provide the necessary energy and environment for first-order Fermi acceleration. The typical rate of supernova explosions is approximately 1–3 per century in the Milky Way, releasing $\sim 10^{51}$ erg of kinetic energy per event. Assuming 10% of this energy is transferred to cosmic rays, the resulting energy density aligns with the observed cosmic-ray energy density of ~ 1 eV/cm³.

For efficient particle acceleration, the following criteria must be met:

1. Particles must gain energy faster than they lose it to radiative or adiabatic processes.
2. Particles must remain confined within the acceleration region for sufficient time, which depends on the diffusion coefficient and the magnetic field structure.
3. The acceleration site must provide sufficient mechanical power to sustain the observed cosmic-ray flux.

The maximum energy $E_{\max}$ of accelerated particles is constrained by the size L , magnetic field strength B , and the velocity v of the shock. The Hillas criterion provides an estimate:

$$E_{\max} \sim ZeBL \frac{v}{c}. \quad (2.11)$$

For SNRs, this limit is $\sim 10^{15}$ eV, consistent with the knee of the cosmic-ray spectrum. In other words, SNRs are excellent candidates for Galactic cosmic-ray sources. Experimental evidence, such as gamma-ray observations from SNRs (e.g. SNR W44), supports the idea of these sources as cosmic-ray acceleration sites.

Beyond the Galaxy, more energetic processes are required to explain ultra-high-energy cosmic rays (UHECRs). Potential sources include:

1. **Active Galactic Nuclei (AGN):** Supermassive black holes with relativistic jets provide powerful environments for particle acceleration via shock waves and magnetic reconnection.
2. **Gamma-ray bursts (GRBs):** Explosive events releasing $\sim 10^{52}$ erg in seconds can accelerate particles to energies exceeding 10^{20} eV.
3. **Neutron stars and magnetars:** Compact objects with intense magnetic fields ($\sim 10^{13}$ G) and rapid rotation can accelerate particles through electromagnetic processes.

These sources are depicted on the Hillas diagram, which maps potential accelerators based on their size and magnetic field strength, providing a framework to evaluate their feasibility [36].

2.3 Ultra-High-Energy photons

Ultra-high-energy gamma rays are photons with energies above 100 TeV (0.1 PeV), corresponding to frequencies beyond 2.42×10^{28} Hz and wavelengths shorter than 1.24×10^{-20} m. Their study is important in cosmic-ray physics, as these neutral messengers carry information about the most energetic processes in the universe. Unlike charged cosmic rays, photons are not deflected by magnetic fields, and thus retain directional information about their sources. Similarly to neutrinos, they provide a unique opportunity for astronomy, allowing

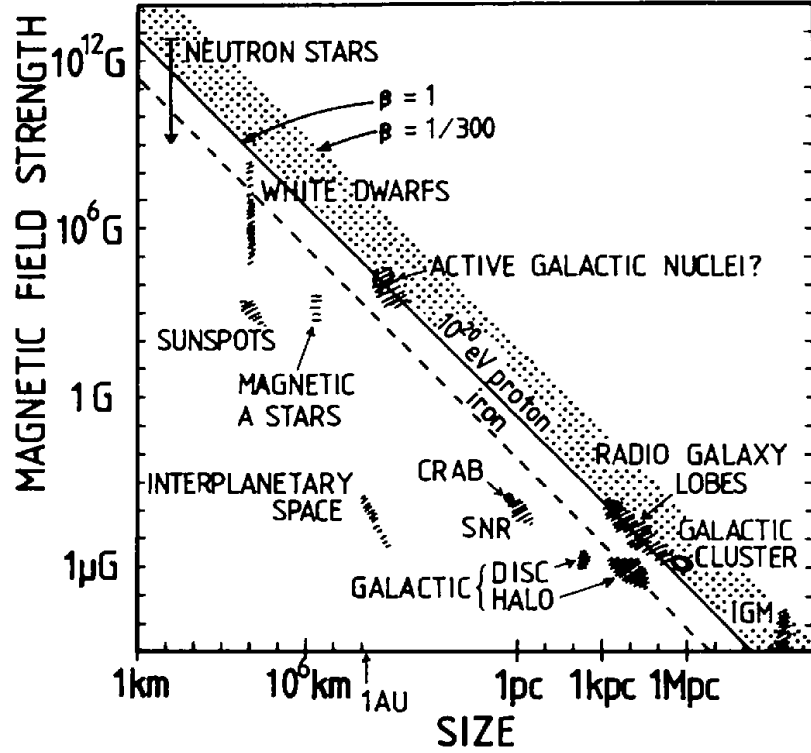


Figure 2.4: The Hillas plot of potential cosmic ray accelerators locates objects according to size and magnetic field. Objects to the left of the diagonals are not able to accelerate particles to 10^{20} eV total energy. Separate lines are displayed for protons (solid) and iron (broken), and the shaded region shows the effect of shock velocity. Figure taken from Ref. [36].

directional studies that may help uncover the origin of cosmic rays. This capability places UHE photons at the forefront of multi-messenger astrophysics.

The development of this field has been marked by a series of key observational breakthroughs. In 2019, the Tibet AS γ Collaboration reported the first statistically significant detection of gamma rays above 100 TeV from the Crab Nebula, confirming the existence of this high-energy regime [37]. Since then, the Large High Altitude Air Shower Observatory (LHAASO) has reported the observation of more than a dozen gamma rays with energies exceeding 1 PeV. Most recently, in 2024, LHAASO reported the detection of a 2.5 PeV photon originating from the Cygnus X region, setting a new record for the most energetic photon observed [38]. Sources capable of producing such energetic gamma-ray emission are usually referred to as *PeVatrons*.

To gain deeper insight into the origin of UHE photons, we first examine the broader mechanisms behind astrophysical gamma-ray production. A paradigmatic example is the Crab Nebula, a SNR and pulsar wind nebula (PWN) located in the constellation Taurus. Its broadband spectrum (Fig. 2.5) extends from radio waves to UHE gamma rays, offering a unique laboratory for studying the physical processes responsible for high-energy photon emission in astrophysical environments.

2.3.1 Physical processes behind photon production and sources

Photons are produced in astrophysical environments across a wide energy range. Broadly, two primary classes of mechanisms exist: leptonic and hadronic processes. These differ fundamentally in the particles involved and in the resulting photon energy spectra. Leptonic processes involve interactions of electrons or positrons with electromagnetic fields or low-

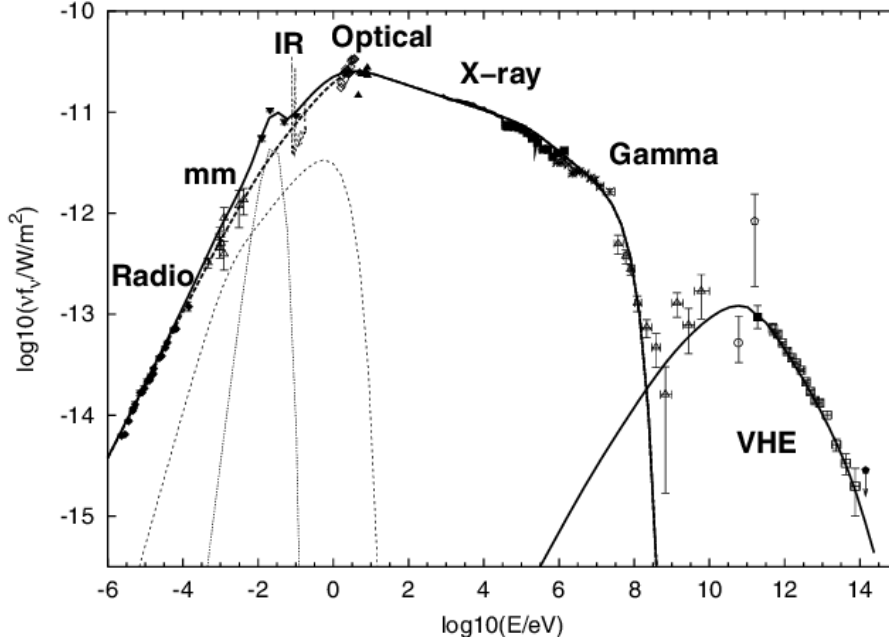


Figure 2.5: Spectral energy distribution of the Crab Nebula, showing emission across a broad range of wavelengths, from radio to very-high-energy (VHE) gamma rays. The spectrum exhibits two main components: a low-energy bump due to synchrotron radiation from relativistic electrons, and a high-energy bump attributed to inverse Compton scattering. Data points are collected from a wide range of observational instruments. Adapted from Ref. [39].

energy photons. Notable examples include synchrotron radiation and inverse Compton scattering. These processes often dominate in environments with strong magnetic fields or abundant low-energy photon fields.

Hadronic processes, in contrast, involve interactions of protons or heavier nuclei. Key examples include pion production and hadronic interactions with ambient matter or radiation fields. These processes are typically associated with environments rich in cosmic-ray activity, such as SNRs and AGNs. Pion decay is a key hadronic mechanism. Hadronic interactions generate neutral pions (π^0) and charged pions ($\pi^\pm$), which decay as follows:

$$\pi^0 \rightarrow \gamma + \gamma, \quad (2.12)$$

$$\pi^\pm \rightarrow \mu^\pm + \nu_\mu \rightarrow e^\pm + \nu_e + \nu_\mu + \bar{\nu}_\mu. \quad (2.13)$$

Gamma rays from π^0 decay manifest as a characteristic *pion bump* in spectra, typically peaking around 67.5 MeV.

The physical processes governing the Crab Nebula are, for the most part, interpreted within a *leptonic* framework. There is broad agreement that *synchrotron radiation* dominates the emission in the radio, optical, and soft X-ray bands. Similarly, *bremsstrahlung* is believed to contribute to the X-ray and lower gamma-ray regime. For gamma rays in the TeV range, *inverse Compton scattering* is considered the most likely mechanism. Fig. 2.6 presents a multi-wavelength view of the Nebula, in which different physical processes are responsible for the emission observed across various spectral regions. Such observations provide valuable insights into the physical conditions and particle populations within the Nebula.

Beyond nebulae, many other astrophysical environments support mechanisms for high-energy photon production. For instance, in SNRs, particles are accelerated via diffusive shock acceleration. Photons are subsequently generated through synchrotron radiation,

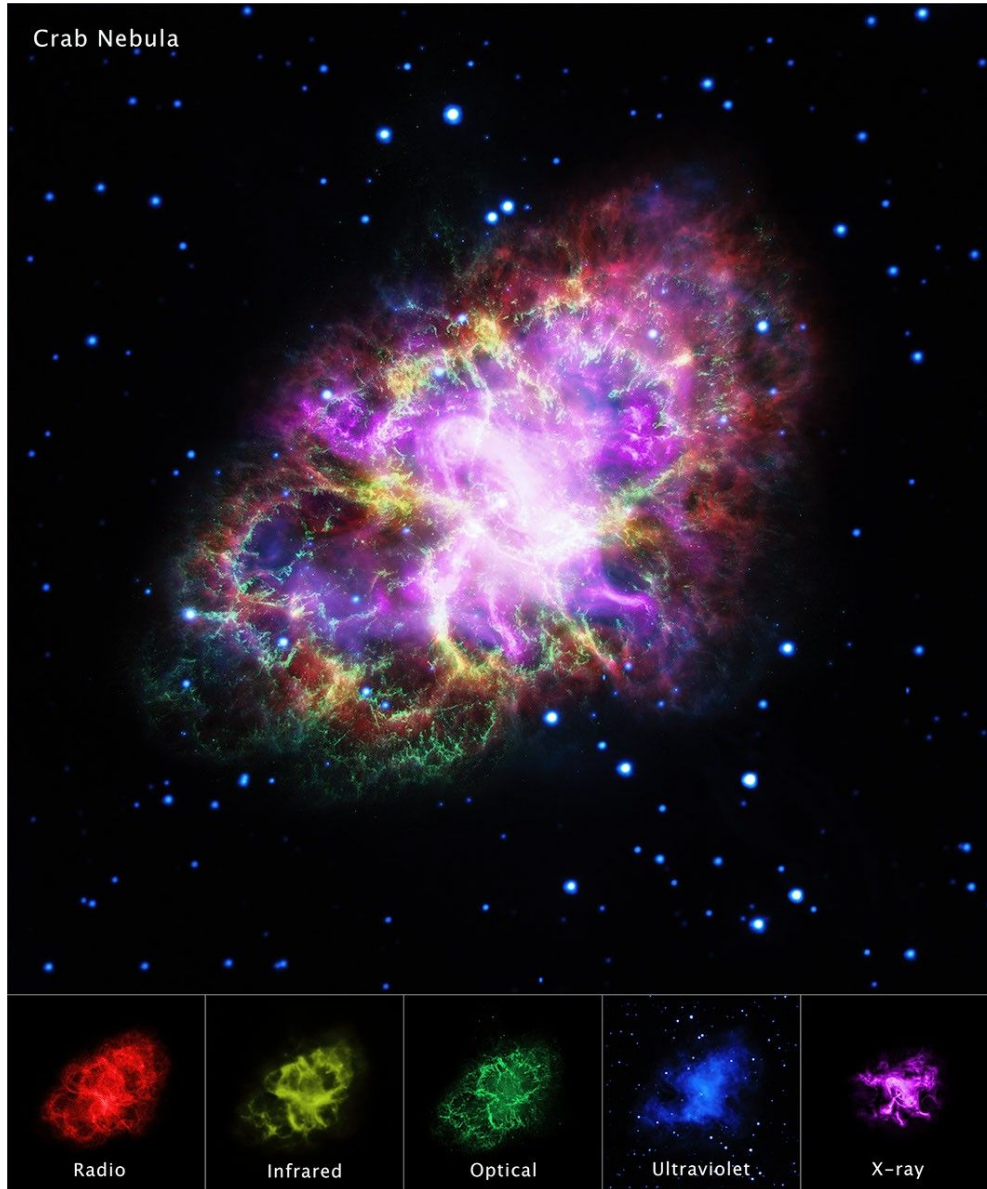


Figure 2.6: Multi-wavelength composite image of the Crab Nebula, combining data from five observatories spanning the electromagnetic spectrum. The image includes radio (red, VLA), infrared (yellow, Spitzer), optical (green, Hubble), ultraviolet (blue, XMM-Newton), and X-ray (purple, Chandra) observations. Visible structures include hot ionized filaments, thermal dust emission, and a central pulsar whose wind drives synchrotron emission from radio to X-rays. The Crab Nebula, located 6,500 light-years from Earth in the constellation Taurus, is the remnant of a supernova observed in 1054 CE. Image credit: NASA, ESA, G. Dubner (IAFE, CONICET-UBA) et al.; VLA/NRAO/AUI/NSF; Chandra/CXC; Spitzer/JPL-Caltech; XMM-Newton/ESA; and Hubble/STScI.

bremsstrahlung, and pion decay, producing a spectrum that extends from radio to gamma-ray energies.

Other notable astrophysical sources of photon emission include GRBs, AGNs, and pulsars [40]. GRBs are transient events of extreme luminosity. High-energy photons are produced in relativistic jets via synchrotron and inverse Compton processes. Their observed spectra typically follow a broken power law, shaped by the jet's Lorentz factor and the target photon population. AGN jets are powerful particle accelerators: synchrotron radiation

dominates at lower energies, while inverse Compton scattering and pion decay contribute at higher energies. The resulting spectrum depends on the jet's magnetic field strength and ambient photon density. Pulsars, too, accelerate particles in their magnetospheres, producing photons via curvature radiation and inverse Compton scattering. Their spectra often feature sharp peaks, tied to the geometry and energy scales of the magnetic field and accelerated particles.

A comprehensive discussion of photon production would be incomplete without addressing cosmic rays. Interactions of cosmic rays with the ISM generate diffuse gamma rays through bremsstrahlung and pion decay. The resulting spectrum is broad, reflecting both the energy distribution of the cosmic rays and the properties of the target medium. As discussed in the context of the GZK effect, interactions between ultra-high-energy cosmic rays and the CMB can also produce neutral pions, which subsequently decay into gamma rays (see Eqs. (2.6) and (2.13)). These scenarios, involving propagation rather than emission at the source, are typically referred to as *cosmogenic*.

2.3.2 State of the art

Detection techniques for photons across the electromagnetic spectrum have evolved to adapt to atmospheric transparency and physical constraints. As illustrated in Fig. 2.7, high-energy radiation such as X-rays or gamma rays is significantly absorbed by the Earth's atmosphere, requiring space-based telescopes for direct observation. For instance, satellites like Chandra or XMM-Newton operate in X-ray bands, while optical and infrared telescopes are ground-based, located at high altitudes or dry sites to minimise atmospheric absorption. At even longer wavelengths, radio telescopes can operate from the Earth's surface due to the high transparency of the atmosphere to radio waves.

In the case of UHE photons, the flux is extremely low, making direct detection unfeasible even from space. Instead, ground-based observatories infer their presence indirectly through the EAS they produce when interacting with the atmosphere. Two main strategies dominate this domain: the Imaging Atmospheric Cherenkov Technique (IACT) and large-scale surface detector arrays. These methods are designed to operate over wide areas and long periods, maximising exposure to compensate for the rarity of UHE photons.

The IACT technique detects Cherenkov light emitted by relativistic secondary particles in air showers. This brief flash of blue light is captured by fast optical systems during clear, moonless nights. Leading observatories employing this method include High Energy Stereoscopic System (H.E.S.S.), Major Atmospheric Gamma-ray Imaging Collaboration (MAGIC), and Very Energetic Radiation Imaging Telescope Array System (VERITAS), which have pushed the sensitivity frontier for photon energies in the TeV range and slightly beyond [42]. However, IACT is constrained by its limited duty cycle and relatively small field of view.

Surface detector arrays, on the other hand, operate continuously and can monitor extensive areas, making them well-suited for the search of rare UHE photons. These arrays consist of particle detectors spread over hundreds to thousands of square kilometres, enabling indirect detection of photons through the extensive air showers they induce. Notable examples include the High-Altitude Water Cherenkov (HAWC) Observatory in Mexico and LHAASO in China, both of which combine technologies like water-Cherenkov and scintillator detectors to probe photons in the TeV to PeV energy range [43]. Future projects such as *swgo* aim to provide similar coverage in the Southern Hemisphere.

The Pierre Auger Observatory is the largest facility dedicated to cosmic-ray and multi-messenger studies at ultra-high energies, employing a hybrid detection approach that combines surface water-Cherenkov detectors with fluorescence telescopes [44]. This thesis focuses on one of the denser infill arrays, called SD-433, which extends the sensitivity of the

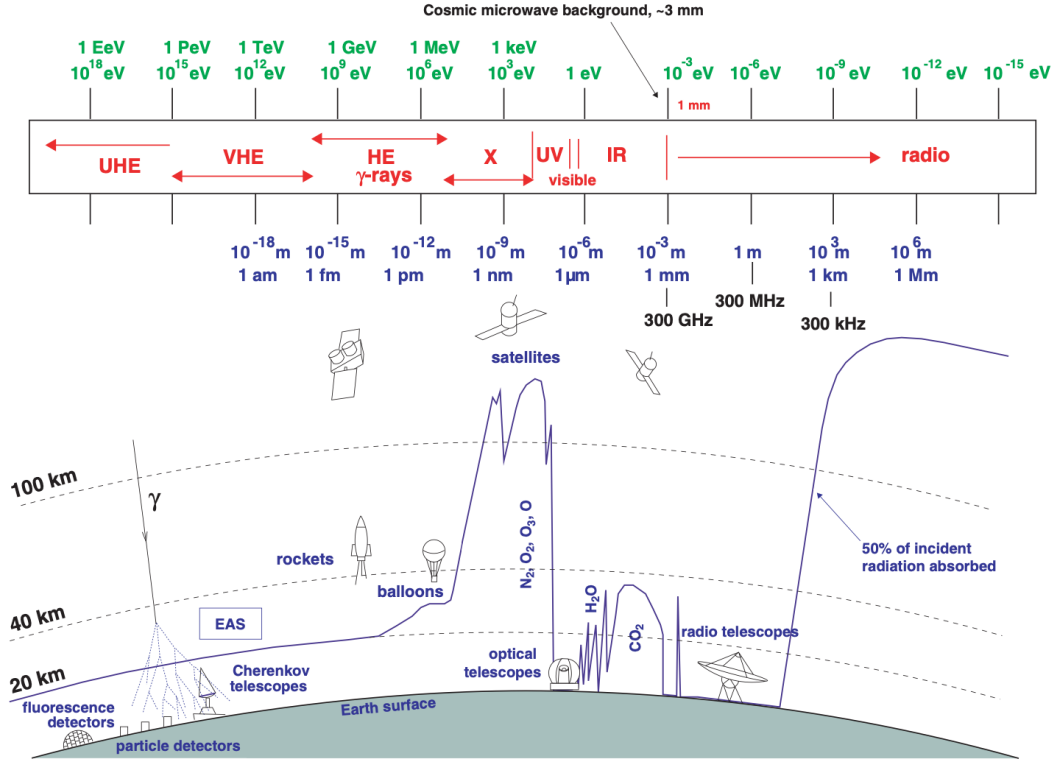


Figure 2.7: Atmospheric transparency across the electromagnetic spectrum. The blue curve indicates the altitude at which 50% of the incoming radiation is absorbed by the Earth’s atmosphere. Only optical and radio photons can reach ground-based detectors directly. Other wavelength bands, such as infrared, ultraviolet, X-rays, and gamma rays, are partially or entirely absorbed and require balloon-borne or satellite instruments. For VHE and UHE gamma rays, indirect detection via air-shower techniques is possible from the ground, including the IACT and particle detector arrays. Figure adapted from [41].

Observatory toward the tens of PeV energy range. Additionally, the information from the UMD is exploited to enhance photon–hadron discrimination.

Over the past years, the Pierre Auger Observatory has conducted several photon searches, including diffuse flux analyses, targeted transient events, and directional studies of candidate sources [45–52]. Transient searches focus on temporal coincidences with astrophysical events such as gamma-ray bursts or neutrino detections, while directional searches investigate excesses in specific regions of the sky associated with known or potential sources. Diffuse flux searches aim to constrain the integral photon flux from all directions and over extended periods, as shown in Fig. 2.8, where the latest upper limits are presented alongside various model predictions [53–57].

Predicted photon fluxes at ultra-high energies arise from several physical scenarios, each associated with different production mechanisms and energy spectra. *Cosmogenic photons* are produced in the interactions of ultra-high-energy cosmic rays with cosmic background radiation fields, such as the cosmic microwave background and the extragalactic background light [58, 59]. The resulting flux depends on the primary composition and cosmological source evolution, with separate models for proton-dominated or mixed-composition primaries (e.g., *GZK proton* and *GZK mixed* in Fig. 2.8). *Galactic contributions*, on the other hand, arise from interactions of cosmic rays with interstellar gas and radiation fields within the Milky Way [60], leading to a flux that falls steeply with energy [61]. At energies below 10^{17} eV,

this Galactic component may dominate, although it remains several orders of magnitude below the total cosmic-ray flux.

In addition to these standard scenarios, more speculative models predict enhanced photon fluxes at the highest energies. *Dark matter decay or annihilation*, particularly involving super-heavy dark matter (SHDM), can yield detectable gamma-ray fluxes due to the proximity of the Galactic centre, where attenuation is negligible. Various SHDM models differ in lifetime and decay channels, resulting in distinct spectral features labelled as SHDM Ia, Ib, and II in Fig. 2.8 [62, 63]. Observational constraints on these models therefore serve as powerful probes of both astrophysical and beyond-the-Standard-Model physics.

This thesis concentrates on improving diffuse photon searches in the 50–500 PeV range, although the machine learning techniques developed herein are generalisable to any of the aforementioned types of searches. The latest diffuse search at Auger using the UMD relied on a traditional analysis. In that work, an observable called M_1 , which quantifies the muon content relative to that expected from proton-initiated showers, was shown to be effective for photon–hadron discrimination, leading to the current upper limits shown as red markers in Fig. 2.8 [52]. In this thesis, we explore a complementary strategy based on graph neural networks, aiming to exploit the full spatial and temporal structure of the recorded data.

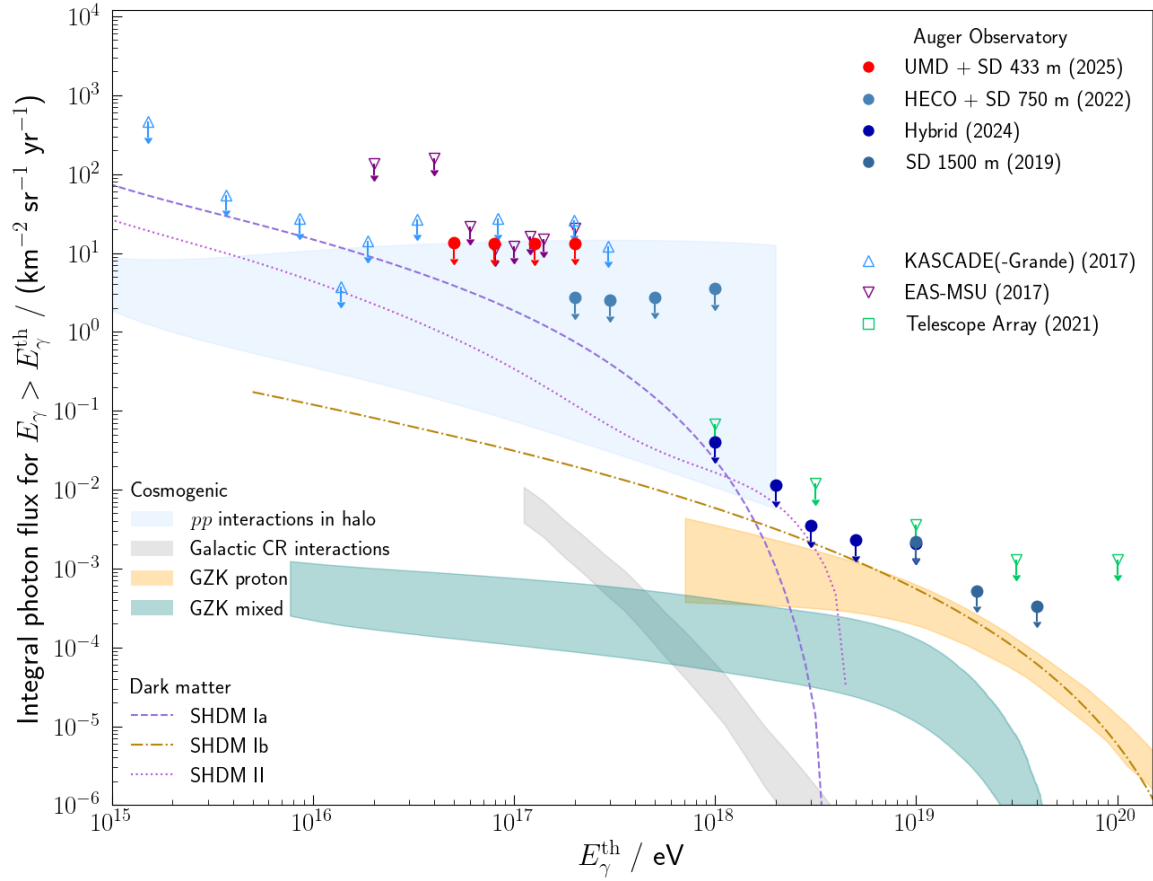


Figure 2.8: Integral upper limits on the diffuse photon flux for energies above threshold values $E_\gamma > E_\gamma^{\text{th}}$, as a function of E_γ^{th} . Results from several experiments are shown, including Auger (using different configurations), Karlsruhe Shower Core and Array DETector (KASCADE)(-Grande), Extensive Air Showers - Moscow State University (EAS-MSU), and TA. The current best limits are provided by the Pierre Auger Observatory using data from the SD arrays at 1500 m and 750 m spacing, hybrid observations, and more recently, the SD-433 array combined with the UMD. Coloured bands and lines represent predictions from various theoretical models of photon production.

Extensive air showers

Everything was intertwined, with the complexity of a three-dimensional puzzle, a puzzle in which truth was not necessarily fact and fact not necessarily truth.

The Wind-Up Bird Chronicle, Haruki Murakami

When a high-energy cosmic ray (CR) collides with an atomic nucleus in the Earth's atmosphere, it initiates a complex sequence of interactions that produces a large number of secondary particles. Observations of correlated particle detections over large ground areas led to the adoption of the term Extensive Air Shower (EAS) [16].

An EAS develops along the trajectory of the primary particle, known as the *shower axis*. The central region of the shower, or *core*, contains the most energetic hadrons, which sustain the cascade until it reaches the ground. As the shower propagates through the atmosphere, the primary energy is progressively dissipated, predominantly through ionisation losses and bremsstrahlung [64]. Only a small portion escapes in the form of weakly interacting particles, such as neutrinos, or in particles absorbed by the ground.

The shower's growth is governed by a balance between particle production and energy losses. This balance culminates at the *shower maximum*, the point at which the number of particles present simultaneously is largest. Beyond this stage, the shower attenuates as secondary particles are absorbed or decay.

In this chapter, we introduce the fundamental physics of extensive air showers, with emphasis on the main interaction processes and the characteristic stages of development. The initial structure follows closely the EAS introduction in Ref. [65]. The basic mathematical framework commonly employed to describe their evolution is presented in Section 3.1 and Section 3.2, following the notation and methodology of Refs. [64, 66], unless stated otherwise. A concise overview of current simulation techniques and the computational tools that form the basis for modern EAS research is provided in Section 3.3. Finally, the differences between hadronic- and photon-induced showers are discussed in Section 3.4.

3.1 Basic notions of cascade theory

When a high-energy particle enters the atmosphere, it initiates a cascade in which its energy is progressively transferred to secondary particles. In the case of EAS, the atmosphere acts as a calorimeter. Although the atmospheric density ρ is not homogeneous, the same

principles as in regular calorimeters apply: distances and distance-dependent quantities are expressed in terms of the *traversed matter* (or slant depth),

$$X(r) = \int_0^r \rho(\vec{r}(r')) dr', \quad (3.1)$$

which integrates the density along the (straight) path $\vec{r}$ of length r . The depth X , measured in g cm^{-2} , is defined from the entry point of the primary particle at the top of the atmosphere downwards along the shower axis.

The general cascade evolution can be described by

$$\frac{dN_i(E_i, X)}{dX} = -\frac{N_i(E_i, X)}{\lambda_i} - \frac{N_i(E_i, X)}{d_i} + \sum_{j=i}^J \int_{E_i}^{\infty} \frac{F_{ji}(E_i, E_j)}{E_i} \frac{N_j(E_j, X)}{\lambda_j} dE_j. \quad (3.2)$$

where N_i is the number of particles of type i at depth X and energy E_i , λ_i and d_i are the interaction and decay lengths, and F_{ji} is the yield of i from the interaction of j with an air nucleus. The first two terms describe losses of i due to interactions and decays, while the third term accounts for production of i from higher-energy parent particles.

For the electromagnetic component of an EAS, this system can be reduced to two coupled *cascade equations* for the photon spectrum $\gamma(W, X)$ and the electron/positron spectrum $\pi(E, X)$ [66], such that

$$\frac{d\gamma}{dX} = -\frac{\gamma(W, X)}{\lambda_{\text{pair}}} + \int_W^{\infty} \pi(E', X) \frac{d^2 n_{e \rightarrow \gamma}}{dW dX} dE', \quad (3.3)$$

$$\frac{d\pi}{dX} = -\frac{\pi(E, X)}{\lambda_{\text{brems}}} + \int_E^{\infty} \pi(E', X) \frac{d^2 n_{e \rightarrow e}}{dE dX} dE' + 2 \int_E^{\infty} \gamma(W', X) \frac{d^2 n_{\gamma \rightarrow e}}{dE dX} dW'. \quad (3.4)$$

In these equations, E denotes the kinetic energy of electrons or positrons, while W denotes the energy of photons. Both spectra are defined such that $\pi(E, X) dE$ (or $\gamma(W, X) dW$) gives the number of particles per unit energy interval at slant depth X , while λ_{pair} and λ_{brems} are the mean free paths for pair production and bremsstrahlung, respectively.

In Eq. (3.3), the first term represents photon loss due to pair production, and the second term is photon production via bremsstrahlung from electrons. In Eq. (3.4), the first term is electron loss via bremsstrahlung, the second term is electron production by electron scattering (with energy redistribution), and the third term is electron production from pair creation by photons.

This reduced system follows from the general cascade equation under the assumptions:

1. Only two processes are considered: bremsstrahlung ($e \rightarrow e + \gamma$) and pair production ($\gamma \rightarrow e^+ + e^-$).
2. Complete screening of atomic nuclei is assumed, so cross sections are energy-independent in the high-energy limit.
3. Energy loss of electrons by ionisation and collisions is neglected.
4. Electrons and positrons are treated identically and denoted collectively as e .

Under these conditions, each electron emits a photon after traversing one radiation length λ_{brems} , and each photon pair-produces after λ_{pair} , leading to a doubling of particle numbers per interaction length while energies remain above a critical threshold.

3.1.1 Longitudinal development for a single primary photon

Under direct substitution in Eqs. (3.3) and (3.4), it can be shown that elementary solutions exist with the factorised form of

$$\pi(E, X) = a E^{-(s+1)} f(X) \quad \text{and} \quad \gamma(W, X) = b W^{-(s+1)} f(X), \quad (3.5)$$

with s as a positive number, which at the same time implies the ratio $\lambda := \frac{1}{f} \frac{\partial f}{\partial X}$ is constant at a fixed energy and it follows that $f(X) \propto e^{\lambda X}$.

In this scenario, the number of particles changes exponentially with depth, while the energy spectra and relative abundance of photons and electrons remain unchanged. This does not happen in a realistic scenario. The energy spectra of particles are not constant and change with the shower development.

To move towards more realistic solutions, we consider the classic case of a single photon of energy W_0 injected at the top of the atmosphere. After a sequence of several approximations and mathematical tricks, one can solve the cascade equations. The corresponding initial condition is

$$\gamma(W, 0) = \delta(W - W_0)$$

and $\pi(E, t) dE$ will denote the number of electrons with energies in $[E, E + dE]$ at depth $t = X/X_0$. The quantity X_0 is the radiation length; i.e., the energy-loss length of charged particles, or the mean length after which only a fraction of $1/e$ of the initial energy remains.

The electron number for a cascade induced by a single photon can be written as

$$\pi(E, t) dE \approx \frac{dE}{E} \frac{1}{2\pi i} \frac{B(\bar{s})}{\sqrt{5\lambda_1(\bar{s})}} \int_{-i\infty+\bar{s}}^{i\infty+\bar{s}} ds \exp \left[\lambda_1(s) t + s y + \frac{1}{2} \ln s \right] \quad (3.6)$$

$$\approx \frac{dE}{E} \frac{B(\bar{s})}{\sqrt{2\pi} \left(\bar{s} \left(\frac{3}{2} \bar{s} t - \frac{1}{2\bar{s}^2} \right) \frac{1}{2} \lambda_1(\bar{s}) \right)} \exp \left[\lambda_1(\bar{s}) t + \bar{s} y + \frac{1}{2} \ln \bar{s} \right] \quad (3.7)$$

where $B(\bar{s}) = 2 \frac{\sigma_0}{\bar{s}+1}$, and σ_0 represents the dimensionless cross section for pair production in air $X_0/\lambda_{\text{pair}} \sim 7/9$. Additionally, λ_1 is the first root of a quadratic equation for λ that follows from the consideration of radiation probabilities of the terms in the coupled equations. Finally, the value y is defined by $y \equiv \ln \left(\frac{W_0}{E} \right)$. The expression is only valid *above a low-energy cutoff* of $E_{\text{cut}} = 100$ MeV. A more involved derivation, without falling into cumbersome mathematics, can be found in Ref. [65].

Two implications arise from this result. First, the existence of a (energy-dependent) shower maximum. For $E \ll W_0$, the electron spectrum $\pi(E, t)$ attains a maximum as a function of depth at

$$t_{\text{max}}(E) = y - 1 \approx y, \quad (3.8)$$

i.e. at the depth where the exponent in the solution is maximal. Defining the longitudinal profile

$$\Pi(t) = \int_{E_{\text{cut}}}^{W_0} \pi(E, t) dE, \quad (3.9)$$

the maximum of $\Pi(t)$ is the (global) *shower maximum*.

The second implication is that a parameter called the *shower age* s allows a universal description of the longitudinal development of the shower. Depth and energy dependencies can be recast through the *shower age* parameter

$$s = \frac{3t - 1}{t + 2y} \approx \frac{3t}{t + 2y}, \quad (3.10)$$

meaning that, to a good approximation, $\pi(E, t)$ may be expressed in terms of s rather than E and t separately. Equivalently,

$$s(t) = \lambda_1^{(-1)} \left(\frac{1}{\Pi} \frac{d\Pi}{dt} \right), \quad (3.11)$$

linking s to the fractional growth rate of the longitudinal profile. As s increases the spectrum steepens and high-energy particles become less abundant; in general $s > 0$ and $s = 1$ mark the shower maximum.

The cut-off of 100 MeV that ensures validity of Eq. (3.7) was introduced as a safeguard for the validity of the approximation where electron loss by ionisation and collisions is neglected. When Coulomb scattering and ionisation losses are included, the longitudinal development of electromagnetic cascades can be modelled more faithfully. In this regime, particle production effectively ceases once the bremsstrahlung and pair production probabilities vanish. This occurs when the particle energy falls below the *critical energy* for electrons in air,

$$\varepsilon_c^{e\gamma} \simeq 87 \text{ MeV},$$

where ionisation losses dominate over radiative losses. This description in cascade theory is particularly important for the late stages of shower development, when most particles have $E \lesssim \varepsilon_c^{e\gamma}$.

The *Greisen profile* [67, 68] provides a phenomenological description of the longitudinal cascade evolution, incorporating the critical energy cut-off. Defining

$$\beta_0 = \ln \left(\frac{E_0}{\varepsilon_c^{e\gamma}} \right), \quad y = \beta_0,$$

and using the shower age s from Eq. (3.10), the Greisen profile is given by

$$N(t) = 0.31 \beta_0^{-1/2} \exp \left[t \left(1 - \frac{3}{2} \ln s \right) \right]. \quad (3.12)$$

In this description of the cascade, the shower age s is mainly driven by the low-energy component of the particle distribution, and the depth of shower maximum depends on both the critical energy and the energy of the primary photon, rather than only on the primary energy.

The shower maximum occurs at $s = 1$ ($t_{\max} = y$), corresponding to a depth

$$X_{\max} = X_0 \ln \left(\frac{E_0}{\varepsilon_c^{e\gamma}} \right). \quad (3.13)$$

At $t_{\max}$, the number of particles is

$$N_{\max} = 0.31 \frac{E_0 / \varepsilon_c^{e\gamma}}{\sqrt{\ln(E_0 / \varepsilon_c^{e\gamma})}} \approx \frac{E_0}{\text{GeV}}, \quad (3.14)$$

valid to order-of-magnitude for a primary energy $E_0 \gtrsim 10^{16}$ eV. Several profiles at different energies are shown in Fig. 3.1.

The knowledge of the number of particles N with respect to the primary energy E_0 at a specific shower age s is particularly useful for the energy calibration of surface detector arrays. For example, in a detector array of slightly shielded scintillators (e.g. Volcano Ranch), electrons and positrons with $E \gtrsim 20$ MeV contribute to the measured signal, so that $N(E > 20 \text{ MeV})$ can be taken as the total number of electrons and the energy of the primary particle can be roughly estimated. This approach is the baseline for more involved energy calibration methods that will be discussed in Chapter 4 and Chapter 5.

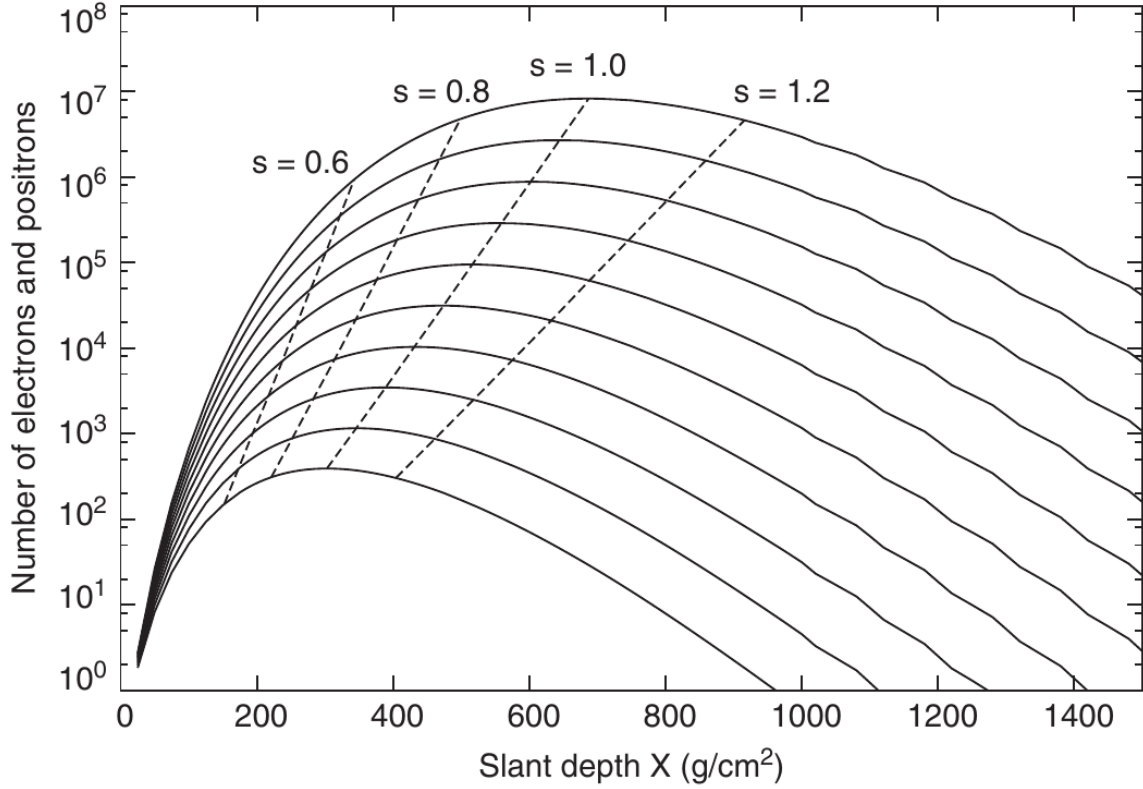


Figure 3.1: Longitudinal development of the number of electrons and positrons $N(X)$ in photon-initiated electromagnetic air showers, as a function of slant depth X in g/cm^2 . Solid curves correspond to half-decade intervals of the primary photon energy E_0 , ranging from 316 GeV (lowest curve) to 10^7 GeV (highest curve), calculated using the Greisen profile. The vertical axis shows $N(X)$ on a logarithmic scale, while the horizontal axis gives the slant depth. Dashed lines trace the locus of equal shower age s (Eq. (3.10)) across the range of primary energies. The line $s = 1.0$ marks the shower maximum X_{max} , where $N(X)$ peaks; $s < 1$ describes the growth phase dominated by particle multiplication, and $s > 1$ corresponds to the attenuation phase dominated by energy losses and particle absorption. This representation illustrates both the increase of X_{max} with primary energy and the universality of the longitudinal shape when expressed in terms of s . Figure taken from Ref. [64].

3.1.2 Lateral development

So far, we have focused mainly on the evolution of particles along the shower axis. In practice, particles are also produced at non-zero angles with respect to the axis and undergo deflections as they travel through the atmosphere. This angular divergence leads to a finite *lateral extension* of the shower front, which ultimately defines the footprint of the cascade at ground level. For high-energy photon- or hadron-induced showers, this footprint can range from a few tens of metres to several kilometres in diameter, depending on the primary energy, zenith angle, and altitude of observation.

While the production angles in pair creation and bremsstrahlung contribute to the spread, the dominant mechanism controlling the transverse size of the electromagnetic component is *multiple Coulomb scattering*. This process is quantified through the *Molière radius* r_M , which represents the characteristic scale containing the majority ($\sim 90\%$) of $e^\pm$ in the shower front. For electrons and positrons in air, the Molière radius is given by

$$r_M = \frac{E_s X_0}{\epsilon_c \rho}, \quad (3.15)$$

where X_0 is the radiation length, ρ the local density of air, ε_c the electron critical energy, and E_s is a constant related to the mean scattering loss ($\approx 21\text{--}27$ MeV). At the altitudes typical of ground arrays, r_M corresponds to $\mathcal{O}(100\text{ m})$ in physical distance. Particles with energies much larger than ε_c have proportionally smaller lateral spreads, scaling roughly as $R(E) \propto r_M \varepsilon_c / E$.

A full description of the lateral profile requires solving the three-dimensional cascade equations for the particle density $n(E, r, \alpha, t)$, as first carried out by Kamata and Nishimura [69–71]. Because their exact solutions are cumbersome for experimental use, Greisen [72] proposed a simplified parametrisation, now known as the Nishimura–Kamata–Greisen (NKG) function. In its standard form, the particle density of an electromagnetic (EM) shower at a distance r from the shower core is written as

$$\rho_{\text{NKG}}(r, t) = \frac{N_e(t)}{2\pi r_M^2} \frac{\Gamma(4.5 - s)}{\Gamma(s) \Gamma(4.5 - 2s)} \left(\frac{r}{r_M}\right)^{s-2} \left(1 + \frac{r}{r_M}\right)^{s-4.5}, \quad (3.16)$$

where $N_e(t)$ is the total number of electrons at depth t , and s is the *shower age*. The profile becomes narrower for $s < 1$ (younger showers) and broader for $s > 1$ (older showers).

Measurements with surface arrays of extensive air showers induced by hadronic nuclei have shown that the pure electromagnetic NKG form is too narrow to describe the observed particle distributions in hadronic showers. Moving beyond the simple case of a primary photon, other shower components appear and increase the complexity of the phenomenon. The presence of muons, which lose less energy to scattering and therefore spread over larger distances, requires more flexibility in the description of the lateral spread. To address this, modified empirical profiles have been proposed, for example:

$$\rho_{\text{mod}}(r) = \frac{N}{r_M^2} \left(\frac{r}{r_M}\right)^{-0.75} \left(1 + \frac{r}{r_M}\right)^{-3.25} \left(1 + \left(\frac{r}{a r_M}\right)^2\right)^{-1}, \quad (3.17)$$

where a is a scaling constant and N is the total particle count at ground level. Such functional forms, or variants thereof, are widely used in reconstructing air-shower events in modern observatories, especially in analyses that combine electromagnetic and muonic components. In this thesis, several functions describing the lateral spread of particles are used.

As an illustration, Fig. 3.2 shows the lateral and longitudinal development of a 10^{19} eV proton-induced air shower, with the contributions of different particle species providing a detailed view of the cascade structure. Accurate measurements of observables that characterise the evolution of EASs are essential for reliable energy calibration and for inferring the mass composition of primary cosmic rays.

Up to this point, we have mostly focused on electromagnetic cascades; i.e., showers composed exclusively of electrons, positrons, and photons. In contrast, air showers initiated by hadronic UHECRs exhibit a much richer particle content. The enormous centre-of-mass energy of the first interaction is sufficient to produce the full spectrum of Standard Model particles and to trigger a wide range of possible interactions. Under such conditions, the cascade equations can no longer be solved in closed form, making a detailed description of the particle content heavily reliant on numerical simulations.

3.2 Particle content and the Heitler-Matthews model

When a high-energy hadron enters the upper atmosphere, typically between 15 and 35 km above sea level, it collides with a nucleus of nitrogen, oxygen, or argon, initiating a cascade of secondary particles. Among these, charged and neutral pions are produced most frequently. Neutral pions, with an extremely short lifetime ($c\tau \approx 25\text{ nm}$), decay almost

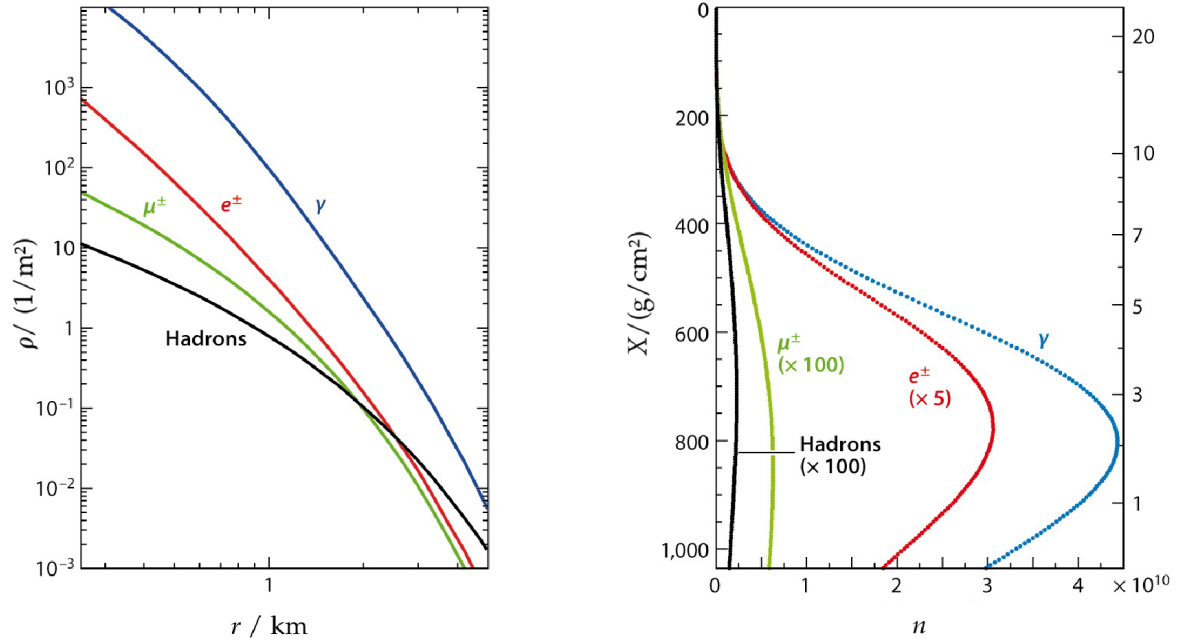


Figure 3.2: Average (*left*) lateral and (*right*) longitudinal profiles of vertical, proton-induced air showers with a primary energy of 10^{19} eV. The lateral distribution at ground level is evaluated at a depth of 870 g cm^{-2} , corresponding to the atmospheric depth at the Pierre Auger Observatory site. Simulation energy thresholds are 0.25 MeV for photons and electrons/positrons, and 0.1 GeV for muons and hadrons. Figure adapted from [73].

instantaneously into two photons, which seed the electromagnetic component of the shower. Charged pions ($c\tau \approx 7.8 \text{ m}$) and kaons ($c\tau \approx 3.7 \text{ m}$) can either interact again or decay into muons and neutrinos; for pion energies above $\sim 30 \text{ GeV}$, further interactions dominate over decay.

The particle content of an extensive air shower can be broadly classified into three components: electromagnetic (EM), muonic, and hadronic. The EM component, driven mainly by photons from π^0 decay, dominates in both particle number and energy [73]. It develops primarily through bremsstrahlung and pair production, producing photons and electrons with nearly identical spatial distributions. The hadronic core, consisting mainly of baryons, charged pions, and kaons, remains concentrated near the shower axis but can generate secondaries with large transverse momenta, thereby broadening the footprint of particles at ground. Muons, produced predominantly in the decay of charged pions and kaons, are comparatively long-lived, lose little energy in transit, and propagate essentially rectilinearly to the surface.

Numerical simulations reveal distinct differences between the lateral and longitudinal profiles of these components. The EM component is most concentrated near the shower core, the muonic profile is broader and more extended in depth, and the hadronic component remains localised but governs the early development of the cascade. Neutrinos, produced in pion and muon decays, interact so weakly that they do not affect the shower's visible evolution; instead, they carry away a fraction (~ 10 – 15% above 10^{19} eV) of the primary energy, which is invisible to calorimetric measurements [74, 75].

Although numerical simulations are needed for an in-depth description of hadronic showers, there is useful intuition to be gained from a simplified branching picture known as the Heitler–Matthews model. The classic Heitler scheme [76, 77] for EM cascades was extended by Matthews [78] to include hadronic production and decay, allowing a compact

account of how energy flows between the hadronic and EM components and how muons arise from the charged pion channel.

Consider a primary of energy E_0 that interacts after a mean inelastic depth λ_{int} , producing n_{tot} secondaries per generation, of which n_{ch} are charged ($\pi^\pm$) and the remainder are neutral (π^0). Neutral pions promptly feed the EM component through $\pi^0 \rightarrow 2\gamma$, while charged pions continue the hadronic chain until their energy falls below a characteristic decay scale E_{dec} , where $\pi^\pm \rightarrow \mu^\pm + \nu$ dominates. In this picture, a fraction $\simeq 1/3$ of the energy is transferred to EM at each generation, so after n generations the energy budgets read

$$E_{\text{had}}(n) = \left(\frac{2}{3}\right)^n E_0, \quad E_{\text{EM}}(n) = \left[1 - \left(\frac{2}{3}\right)^n\right] E_0, \quad (3.18)$$

capturing the gradual dominance of the EM channel as the cascade proceeds. This model is illustrated in the right panel of Fig. 3.3, while the left panel shows the already familiar treatment of pure EM showers.

The shower maximum of a hadron-induced cascade is controlled by the EM subshowers seeded by the early π^0 decays. Approximating with the first generation only, one finds

$$X_{\text{max}}^{(\text{had})}(E_0) \approx \lambda_{\text{ine}} + X_0 \ln\left(\frac{E_0}{2 n_{\text{tot}} E_c}\right), \quad (3.19)$$

with X_0 the EM radiation length, the mean interaction length λ_{int} , and E_c the EM critical energy.

Using the Matthews choice of explicitly tracking the charged multiplicity N_{ch} in the first interaction leads to the equivalent logarithmic dependence

$$X_{\text{max}} = X_1 + X_0 \ln\left(\frac{E_0}{3 N_{\text{ch}} \varepsilon_c^{e\gamma}}\right), \quad (3.20)$$

where X_1 is the first interaction depth and $\varepsilon_c^{e\gamma}$ the EM critical energy in air.

Muon production is tied to the charged hadrons that survive to the last generation. The stop condition is

$$\frac{E_0}{(n_{\text{tot}})^n} = E_{\text{dec}}, \quad N_\mu = (n_{\text{ch}})^n,$$

which eliminates n to give the familiar scaling

$$N_\mu = \left(\frac{E_0}{E_{\text{dec}}}\right)^\alpha, \quad \alpha = \frac{\ln n_{\text{ch}}}{\ln n_{\text{tot}}}. \quad (3.21)$$

Reasonable values of α lie in the interval $\alpha \simeq 0.82\text{--}0.94$, reflecting plausible ranges of charged/total multiplicities and the effective decay scale in air. Equivalently, in the Matthews parameterisation with charged multiplicity N_{ch} and the pion decay energy ε_c^π ,

$$N_\mu = \left(\frac{E_0}{\varepsilon_c^\pi}\right)^{1-\beta}, \quad 1-\beta = \frac{\ln N_{\text{ch}}}{\ln(\frac{3}{2} N_{\text{ch}})}, \quad (3.22)$$

making explicit that N_μ grows sub-linearly with E_0 because only a fraction of the energy remains in the charged branch at each step.

For primaries heavier than protons, the description can follow simply by treating a nucleus of mass number A as A independent nucleons, each initiating a proton-like shower with energy E_0/A (superposition model). This assumption is not far-fetched due to the fact that the binding energy of 5 MeV per nucleon is much smaller than the typical interaction energies. The superposition principle implies

$$X_{\text{max}}^{(A)}(E_0) \approx X_{\text{max}}^{(p)}(E_0/A), \quad (3.23)$$

$$N_\mu^{(A)}(E_0) = A^{1-\alpha} N_\mu^{(p)}(E_0), \quad (3.24)$$

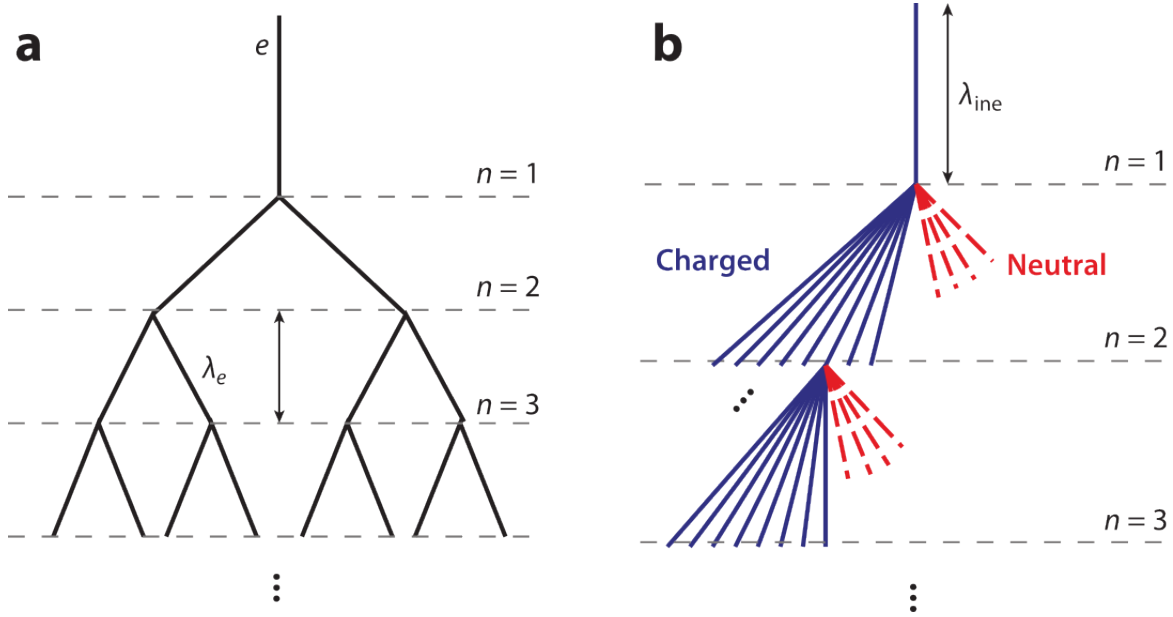


Figure 3.3: Illustration of the Heitler–Matthews branching model. *Left (a):* Purely electromagnetic cascade: each particle produces two secondaries after a mean free path λ_e , doubling the particle number at each generation n . *Right (b):* Hadronic cascade: after a hadronic interaction length λ_{ine} , the primary produces a total multiplicity of secondaries, split into *charged* pions (blue) that continue the hadronic chain, and *neutral* pions (red) that decay promptly into photons, seeding EM subshowers. Figure taken from [73].

and similarly $N_{\text{EM,max}}^{(A)}(E_0) \simeq N_{\text{EM,max}}^{(p)}(E_0)$ up to small shifts from cross-section effects. In this setting, heavier primaries generate more muons at the same E_0 , and reach shower maximum at shallower depths, consistent with detailed simulations. Iron showers possess roughly 40 % more muons than proton-initiated showers of the same energy, and their shower maximum occurs at an atmospheric depth about 80–100 g cm^{-2} smaller.

3.3 Simulation of extensive air showers

Extensive air showers are governed by highly stochastic processes in the creation and annihilation of secondary particles. Even if one were to solve a complete set of cascade equations that includes all Standard Model processes, the shower-to-shower fluctuations observed in real events would not be fully captured. For this reason, numerical simulations have become an indispensable tool in air-shower physics, complementing the analytical approaches introduced in Section 3.1 and Section 3.2.

The development of modern simulations was historically limited by computing power, but the increasing availability of resources has enabled detailed studies of both lateral and longitudinal shower profiles. Frameworks such as COsmic Ray SIMulations for KAScade (CORSIKA) [79] and CONEX [80] are widely used. The former performs a full Monte Carlo (MC) simulation of particle interactions and decays, accurately reproducing fluctuations at the cost of vast computing effort. The latter combines MC techniques with numerical solutions of cascade equations, which greatly reduces CPU demands but yields only expectation values for particle densities rather than explicit realisations of individual secondaries.

In complete MC simulations, the CPU time grows approximately linearly with the energy of the primary particle. To make simulations feasible at the highest energies, a statistical sampling method known as *thinning* is employed: only a representative fraction of particles

is tracked, each assigned a weight. A corresponding *de-thinning* procedure reconstructs physical particle distributions. Although thinning reduces computation time by several orders of magnitude, it can introduce artificial fluctuations if care is not taken [81].

A central ingredient of any air-shower simulation is the hadronic interaction model that governs multiparticle production. At energies beyond those accessible at the Large Hadron Collider (LHC), extrapolations are unavoidable, and the choice of model directly affects predictions for key observables such as the depth of the shower maximum (X_{max}), the number of muons, and the lateral distribution of particles at ground. Three models have been extensively used in cosmic-ray physics: QGSJET [82, 83], SIBYLL [84, 85], and EPOS [86, 87].

The *Quark–Gluon String Model with JETs* implements a microscopic treatment of multiple pomeron exchanges. QGSJET II includes saturation effects through a detailed theoretical model for pomeron–pomeron interactions, and still leaves only a small number of free parameters. It tends to predict conservative muon numbers and shower maxima at shallower depths compared to EPOS. Earlier versions (QGSJET 01) relied on outdated parton densities and neglected saturation, but QGSJET II and the newest QGSJET III remain a reference standard.

SIBYLL 2 represents a more economical approach, implementing only the essential features required for air-shower modelling. It describes hadron production through many concepts of the Dual Parton Model [88], but with fewer microscopic details than QGSJET or EPOS. SIBYLL generally predicts the lowest number of muons among the three models.

EPOS incorporates a more sophisticated treatment of parton dynamics, including pomeron–parton coupling and momentum sharing. It also includes the implementation of collective flow effects in interactions with high parton density. It systematically reproduces a wide range of accelerator measurements, but extrapolates more aggressively to ultra-high energies. EPOS tends to predict a higher muon content and a broader lateral distribution than either QGSJET or SIBYLL.

Although the electromagnetic component of showers is described similarly by all three models, differences become pronounced in the hadronic sector. These discrepancies stem not only from multiplicities but also from the fraction of baryons in secondary production: subshowers seeded by baryons generate $\sim 30\%$ more muons than those initiated by mesons. Consequently, model-dependent differences in baryon production propagate directly into uncertainties in mass composition analyses.

The longitudinal development is also model-dependent. The mean depth of shower maximum $\langle X_{\text{max}} \rangle$ varies significantly between models and even between different versions of the same model. Fluctuations in X_{max} , which are critical for composition studies, are consistently larger for proton primaries than for iron, but the absolute scale differs across models.

In practice, no single model can reliably describe hadronic interactions across the entire energy spectrum relevant for air-shower development. Therefore, high-energy interaction models are typically combined with specialised low-energy models that treat particle production close to threshold, up to laboratory energies of about 200 GeV. Among the most commonly employed are GHEISHA [89], FLUKA [90], and URQMD [91], each optimised for different aspects of low-energy hadron–nucleus interactions. By coupling these with high-energy generators such as QGSJET, SIBYLL, or EPOS, simulation frameworks can achieve a consistent description of extensive air showers over the full energy range of cosmic rays. Representative comparisons of such hybrid setups are discussed in the literature, highlighting the impact of the chosen low-energy treatment on observables like lateral distributions and muon yields [92–94].

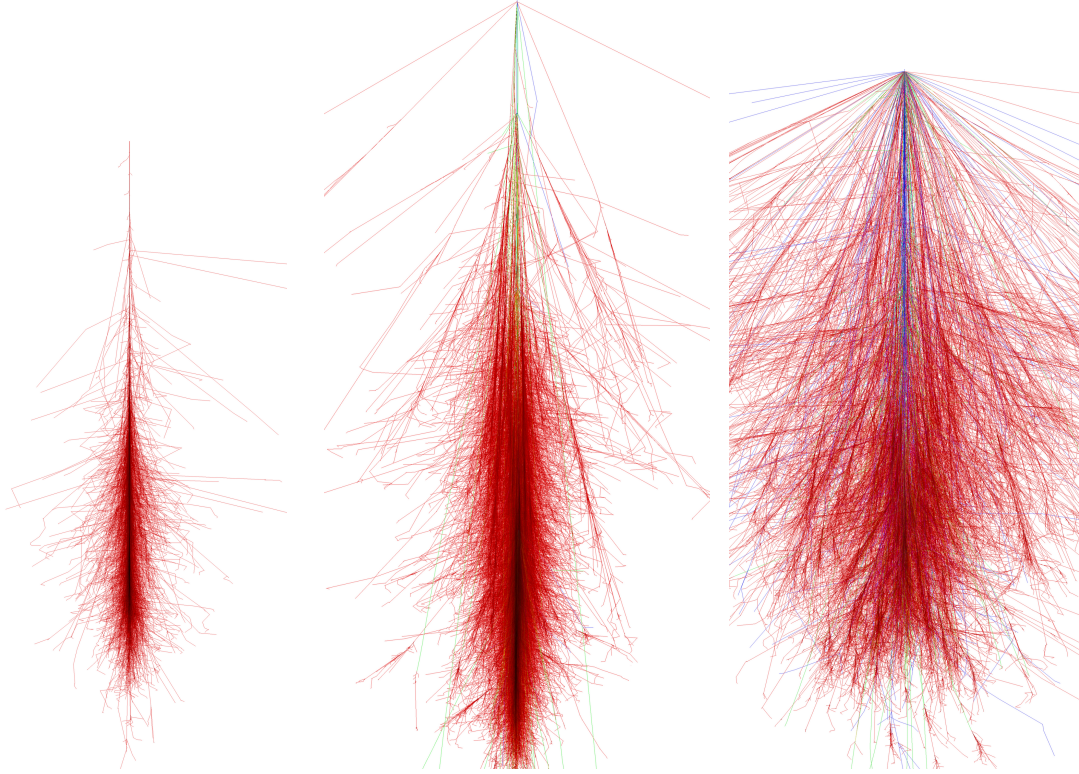


Figure 3.4: Visualisation of three simulated extensive air showers of 1 TeV primary energy, compiled by F. Schmidt and J. Knapp [95]. Shown from left to right are showers initiated by a photon, a proton, and an iron nucleus. Particle tracks are projected onto the simulation coordinate planes using the CORSIKA option `PLOTSH2` and `map2png`. Colours encode particle type and density on a logarithmic scale: red = electromagnetic component (electrons, positrons, gammas), green = muons, blue = hadrons. Colour mixing indicates overlapping particle types (e.g., red + green $\rightarrow$ yellow).

At the time of writing, the Pierre Auger Collaboration has recently adopted EPOS-LHC-R, SIBYLL 2.3D/E, and QGSJET-III.01 as the reference hadronic interaction models for air-shower simulations. These are all available in the official release of [CORSIKA 7.8010](#). Since the work presented in this thesis was initiated prior to this adoption, some of the earlier analyses rely on previous model versions. For clarity, each chapter or section explicitly specifies the versions of the hadronic models used in the corresponding simulations.

To illustrate the visual differences between primary particle types, Fig. 3.4 presents simulated air showers initiated by a photon, a proton, and an iron nucleus, each with a primary energy of 1 TeV. The images were generated with the CORSIKA option `PLOTSH2` and the `map2png` utility developed at the University of Leeds [95]. Particle tracks are projected onto the main coordinate planes of the simulation, with a logarithmic colour scale encoding particle densities: red corresponds to the electromagnetic component (electrons, positrons, and photons), green to muons, and blue to hadrons. Overlaps of colours indicate regions with multiple particle types, where, for example, red and green yield yellow. The figure highlights how the shower morphology depends strongly on the nature of the primary: photon-induced showers develop almost purely electromagnetically, proton showers contain a significant hadronic and muonic component, and iron showers exhibit an even more complex structure with higher track multiplicities and broader lateral spread.

3.4 Features of hadronic- and gamma-induced air showers

EAS initiated by photons differ markedly from those induced by hadronic primaries such as protons. These differences arise from the distinct underlying interactions that govern the cascade development. Proton-induced showers are characterised by an early onset of hadronic multiparticle production, leading to a large number of secondary mesons and, through their decay, an abundant muonic component. In contrast, photon-induced showers develop predominantly through purely electromagnetic processes, namely pair production and bremsstrahlung, which result in a much smaller number of muons and a more elongated longitudinal development [47].

One of the most relevant observables to distinguish the two cases is the depth of the shower maximum, X_{max} . Photon-induced showers penetrate more deeply into the atmosphere before reaching their maximum, as the first interaction occurs via electromagnetic processes with smaller cross sections compared to hadronic interactions. At ultra-high energies, this effect is further amplified by the suppression of cross sections due to the Landau–Pomeranchuk–Migdal (LPM) effect [96], which delays the development of the cascade. As a consequence, photon showers exhibit systematically larger X_{max} values and larger shower-to-shower fluctuations compared to proton showers of the same primary energy.

The lateral distributions of secondary particles also differ. Proton showers exhibit broader muon distributions at ground level due to the hadronic interactions that generate pions and kaons, whereas photon showers contain significantly fewer muons. This muon deficit is a robust signature of photon-induced showers and serves as a powerful discriminator in searches for ultra-high-energy cosmic photons. Furthermore, at energies above ~ 50 EeV, photon showers are subject to pre-showering in the geomagnetic field, where interactions with the Earth’s magnetic field induce early pair production, effectively modifying the longitudinal development before the atmosphere is reached [97].

The analyses in this thesis are restricted to primary energies of tens of PeV and above. In this regime, the influence of the LPM effect and of geomagnetic preshowering is significantly reduced compared to the extreme ultra-high-energy domain. Nevertheless, the general features that distinguish photon-induced from proton-induced showers remain valid: photon primaries still produce deeper shower maxima and exhibit a strong deficit of muons relative to hadronic showers. Figure 3.5 summarises qualitatively the typical differences between proton- and photon-induced showers. While the electromagnetic component dominates both cases, the hadronic nature of proton showers results in an earlier, more compact shower maximum and a significant muonic tail, in contrast to the deeper development of photon showers.

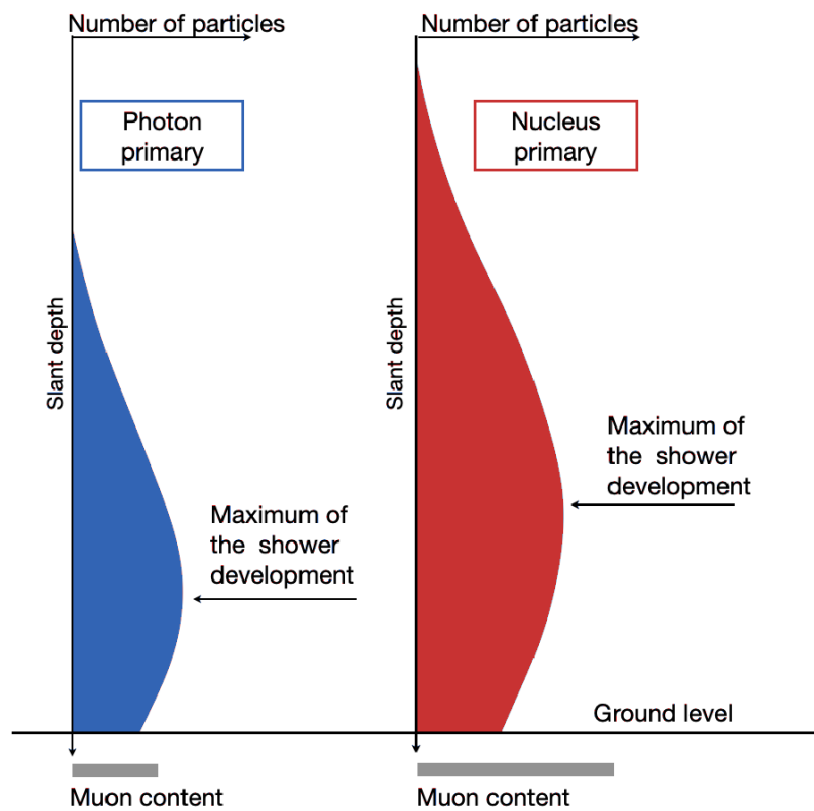


Figure 3.5: Longitudinal development of extensive air showers initiated by protons and photons of the same primary energy. The photon-induced shower develops deeper in the atmosphere, while the proton-induced shower reaches maximum earlier and produces a significant muonic component. Adapted from [47].

The Pierre Auger Observatory

All men have the stars but they are not the same things for different people. For some, who are travellers, the stars are guides. For others they are no more than little lights in the sky. For others, who are scholars, they are problems. For businessmen they were wealth. But all these stars are silent. You—you alone—will have the stars as no one else has them.

The Little Prince, Antoine de Saint-Exupéry

4.1 Auger early years

The idea of the Pierre Auger Observatory was conceived in 1991 with the aim of addressing the persistent mystery of UHE cosmic rays, particularly those exceeding 10^{19} eV. At such extreme energies, the cosmic-ray flux becomes exceedingly rare, requiring detectors of enormous size to collect statistically significant data. The scientific challenge was not only to detect these rare events, but also to unveil their origins, which remain an extensive topic of discussion by conventional astrophysical models as of today.

It was during the 1991 International Cosmic Ray Conference in Dublin that Nobel Laureate Jim Cronin and cosmic ray physicist Alan Watson jointly proposed a bold initiative: to construct a cosmic ray detector array spanning $5,000 \text{ km}^2$, an order of magnitude larger than existing experiments such as AGASA in Japan.

The first detailed planning efforts took place during Cronin's extended visit to Leeds in late 1991, where he collaborated closely with Watson. Drawing on their experience from the Haverah Park array, they explored the design of a large-scale detector, including possibilities for remote data acquisition and the use of GPS for precise timing. These technologies were just becoming viable thanks to pioneering work by Clem Pryke and others.

An international workshop held in Paris in April 1992 attracted over 100 scientists and became a turning point in the formation of the Collaboration. The workshop focused on detection strategies for cosmic rays above 10^{19} eV. Two primary techniques were discussed: ground-based surface detection, and fluorescence observation. However, a third, hybrid approach, originally proposed by Bruce Dawson and Alan Lee, was also presented. Although initially met with scepticism, this method, which combined the calorimetric energy capabilities of fluorescence telescopes with the high-statistics ground sampling of surface detectors, eventually gained traction and became central to the Observatory's design.

Subsequent design meetings, including one in Adelaide in January 1993 and another at Fermilab in 1995, solidified the technical blueprint: an array of 1.2-metre-deep water-



Figure 4.1: Jim Cronin, Alberto Etchegoyen, and Alan Watson at the groundbreaking ceremony of the Pierre Auger Observatory in Malargüe, Argentina, on March 17, 1999. The plaque commemorates the official inauguration of the site and the international collaboration that made it possible. Image taken from [98].

Cherenkov detectors placed 1.5 km apart, covering an area of 3000 km², complemented by fluorescence telescopes placed at strategic locations on the array’s perimeter.

Site selection was a critical milestone in the project. Several candidate locations across Australia, South Africa, and Argentina were considered. Ultimately, a remote site in Argentina’s Mendoza Province, Pampa Amarilla, near the town of Malargüe, was selected. The region offered excellent atmospheric transparency, accessibility for construction and operations, and a welcoming local community.

A key influence in the selection of Argentina was the serendipitous involvement of physicist Alberto Etchegoyen, who encountered the project unexpectedly at a conference in Córdoba¹. His later meeting with Jim Cronin in Bariloche in 1995 led to strong Argentine engagement. The Argentine government pledged substantial funding to support the observatory, an offer that played a decisive role in the final site selection vote held at UNESCO in 1995.

On clear moonless nights, the hybrid configuration allows for event-by-event cross-calibration. Fluorescence data improve the energy resolution of the surface detector events, while surface detectors provide geometric constraints for the fluorescence observations. Together, they enhance the precision of energy reconstruction, arrival direction estimation, and mass composition analyses. A high-quality subset of events measured by the hybrid approach is particularly valuable to the Collaboration, as they are valuable for the calibration of the complete set of surface detector events and define the energy scale of the Observatory.

¹Alberto is always happy to recount the story of “the empty chair”, as he affectionately calls it.

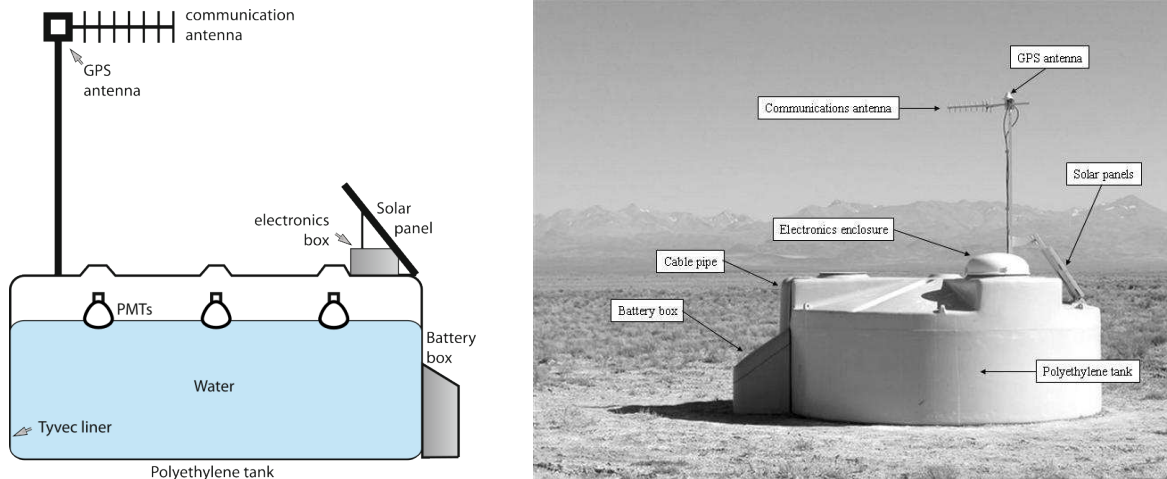


Figure 4.2: Left: Schematic cross-section of a surface detector tank, showing the three PMTs and other components. Adapted from Ref. [99]. Right: Photograph of an SD station installed in the field, taken from Ref. [100].

In the following subsections, we describe the Observatory’s main detection systems and their role in the analyses presented in this thesis.

4.2 Surface Detector

The SD array of the Pierre Auger Observatory consists of over 1,600 water-Cherenkov detectors (WCDs) deployed across a 3000 km² area in a triangular grid with 1.5 km spacing. These stations continuously sample the secondary particles of EAS at 1400 m above sea level with an almost 100% duty cycle, making them the core instrument for high-statistics cosmic-ray measurements at the Observatory.

The original array configuration is formally referred to as SD-1500, owing to its 1.5km spacing. It is commonly known within the Collaboration as the *main array*. Over time, two denser sub-arrays were added within its bounds: SD-750 and SD-433, with spacings of 750m and 433 m, respectively. These compact arrays include 81 and 19 stations, covering smaller areas but targeting higher cosmic-ray fluxes at lower energies. As a result, the SD-750 and SD-433 extend the Observatory’s sensitivity down to tens of PeV, complementing the main array’s coverage at ultra-high energies.

As the main component of each SD station, a cylindrical polyethylene tank (3.6 m in diameter and 1.2 m in height) with an internal Tyvek® reflective liner is filled with 12,000 litres of ultrapure water. The water is carefully purified to remove impurities, ions, and dissolved solids that could absorb or scatter photons, thereby ensuring maximum Cherenkov light yield. Three 9-inch photomultiplier tubes (PMTs), symmetrically mounted on the top surface and facing downward, detect the Cherenkov light produced by relativistic charged particles traversing the water volume and reflected inside the liner. A schematic cross-section of an SD station is displayed in the left panel of Fig. 4.2, while a photographic counterpart is available on the right panel.

Each station is autonomous, powered by solar panels and batteries, and includes a GPS timing module, a local CPU for signal processing, and a radio unit for data transmission to the Central Data Acquisition System (CDAS). To ensure uniform and accurate data, a robust automatic self-calibration procedure is required, since detectors are widely separated, operate in harsh environments, and must remain autonomous. The total bandwidth available

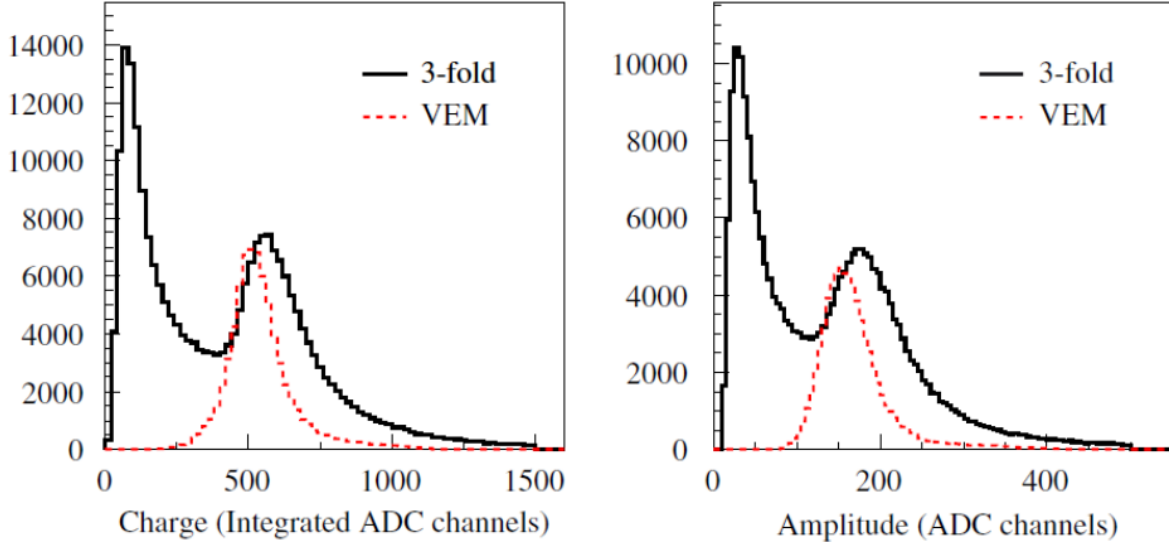


Figure 4.3: Charge (*left*) and pulse-height (*right*) histograms from an SD station, triggered by a 3-fold coincidence between all three PMTs at a trigger level of five channels above baseline, with the signals from all PMTs summed. The dashed histogram corresponds to an external muon telescope trigger selecting only VCT muons. The first peak in the black histogram arises from the convolution of the trigger on a steeply falling distribution of low-energy particles, while the second peak corresponds to vertical through-going atmospheric muons. Taken from Ref. [102]

from each station to the CDAS is 1200 bits per second [101], which implies calibration must be done by the local electronics.

All PMT signals are digitised at 40 MHz using 10-bit flash analog-to-digital converters (FADCs). In addition, two signals are taken from each PMT; one directly from the anode, and the other from the last dynode, amplified and inverted by the base electronics to a total of nominally 32 times the anode. The calibration unit is the Vertical Equivalent Muon (VEM) or Q_{VEM} , defined as the average signal produced by a vertical and central through-going (VCT) muon. The complete calibration procedure is complex and therefore its explanation is rather extensive. The details can be found in Ref. [102]; here, only a brief summary is provided.

During normal operation, the SD cannot select only vertical muons, but the continuous atmospheric muon flux (~ 2500 Hz) produces clear peaks in histograms that serve as calibration references. The *charge histogram* exhibits a peak at $Q_{\text{VEM}}^{\text{peak}}$, while the *pulse-height histogram* exhibits a peak at $I_{\text{VEM}}^{\text{peak}}$. These are directly related to Q_{VEM} and I_{VEM} , respectively, and allow calibration of signals and trigger thresholds. Using these muon-based references ensures uniform triggering conditions across the array. Figure 4.3 shows, in the left panel, the charge histogram and, in the right panel, the pulse-height histogram. The black contours represent the complete histograms, while the red ones correspond to the true VEM counts. The observed shift arises from the convolution of photoelectron statistics with an asymmetric peak in the track-length distribution, together with local light-collection effects.

The calibration consists of three steps. First, the *end-to-end gain setup* adjusts PMT high voltages so that each tube has $I_{\text{VEM}}^{\text{peak}} \approx 50$ channels, balancing the three PMTs within about 5%. Second, *continual online calibration* maintains stability through a convergence algorithm that iteratively adjusts the estimated value $I_{\text{VEM}}^{\text{est}}$ to match a singles rate of 70 Hz. This ensures uniform trigger conditions with $\sim 6\%$ accuracy and introduces only a minor atmospheric-pressure dependence ($\sim 0.1\%$ per g cm^{-2}). Third, *histogram analysis* provides high-statistics determinations of $Q_{\text{VEM}}^{\text{peak}}$ every minute ($\sim 1.5 \times 10^5$ entries), allowing Q_{VEM} to be obtained with $\sim 3\%$ precision.

Other parameters that are required for full VEM signal calibration are FADC baselines and the *dynode/anode gain ratio*. These are obtained by inspecting the shapes of the FADC traces and by using known relationships between the low-gain and high-gain responses of the PMT channels. Only then is a reliable conversion of signals across the full dynamic range achieved.

The SD employs several trigger modes designed to record signals from extensive air showers while also providing auxiliary information for calibration and space weather studies. The simplest of these is the scaler trigger, which records pulses with a very low threshold and is primarily used for auxiliary physics purposes such as monitoring space weather conditions. A complementary mode is the calibration trigger, which collects low-threshold pulses using 20 bins, one of them just above $0.1 I_{\text{VEM}}$. This mode provides high-rate cosmic ray data, with the three high-gain channels recorded from twenty samples before to twenty samples after the trigger time. The resulting data are used to build the calibration histograms for $Q_{\text{VEM}}^{\text{peak}}$ mentioned above, which are used to convert the six FADC traces into VEM units in offline analysis. During the deployment phase, the threshold was lowered to investigate signals from the tails of showers and to measure muon decay in the water volume of the SD stations [103].

The main shower trigger records 768 samples (corresponding to $19.2 \mu\text{s}$) of the six FADC traces and operates in two levels. The first level, denoted T1, has two independent modes. The first is a simple threshold (Th) trigger, which requires the coincidence of all three PMTs with signals above $1.75 I_{\text{VEM}}$. This condition selects large, compact signals, and is particularly effective for identifying very inclined showers, which after traversing large atmospheric depths are dominated by muons. The threshold has been tuned to reduce the rate of atmospheric muon triggers from approximately 3 kHz down to about 100 Hz. The second T1 mode is the time over threshold (ToT) trigger, which requires that at least thirteen bins within a $3 \mu\text{s}$ window (120 samples) exceed a threshold of $0.2 I_{\text{VEM}}$ in coincidence for at least two out of the three PMTs. This trigger efficiently selects signals spread over time, such as those produced by vertical events, stations near the core of low-energy showers, or stations far from the core of high-energy showers. The rate of ToT triggers depends on the muon pulse shape in the tank and averages about 1.2 Hz with a spread of roughly 1 Hz rms.

A second level, denoted T2, is applied to limit the global T1 rate to about 23 Hz. All T1-ToT triggers are promoted to T2-ToT, while T1-Th triggers are promoted only if they exceed a coincidence threshold of $3.2 I_{\text{VEM}}$ in the three PMTs, in which case they become T2-Th. All T2 triggers transmit their timestamp to the CDAS, where a global T3 trigger is formed. More details about the station local triggers are discussed in Ref. [104]. In Fig. 4.4, examples of waveforms that would pass these triggers are shown.

In June 2013, two additional T1 triggers were deployed across the array to improve the efficiency at lower energies and for photon- and neutrino-induced showers. These new triggers refine the ToT logic by applying more sophisticated analyses to the FADC traces. The Time-over-Threshold-deconvolved (ToTd) trigger first deconvolves the exponential tail of diffusely reflected Cherenkov light pulses before applying the ToT condition. This effectively compresses the fast muon signals, typically characterised by a rise time of a few nanoseconds and a decay constant of about 60 ns, into one or two time bins, thereby reducing their influence. The Multiplicity-of-Positive Steps (MoPS) trigger instead counts the number of positive-going signal steps in two of the three PMTs within a sliding $3 \mu\text{s}$ window. These steps are required to exceed a threshold of about five times the RMS noise while remaining below the level of approximately twelve vertical muon steps.

Both the ToTd and MoPS triggers require the integrated signal to be larger than 0.5 VEM . By minimising the contribution of single muons, these triggers reduce the effective energy threshold of the array and enhance the sensitivity to photon and neutrino showers, while

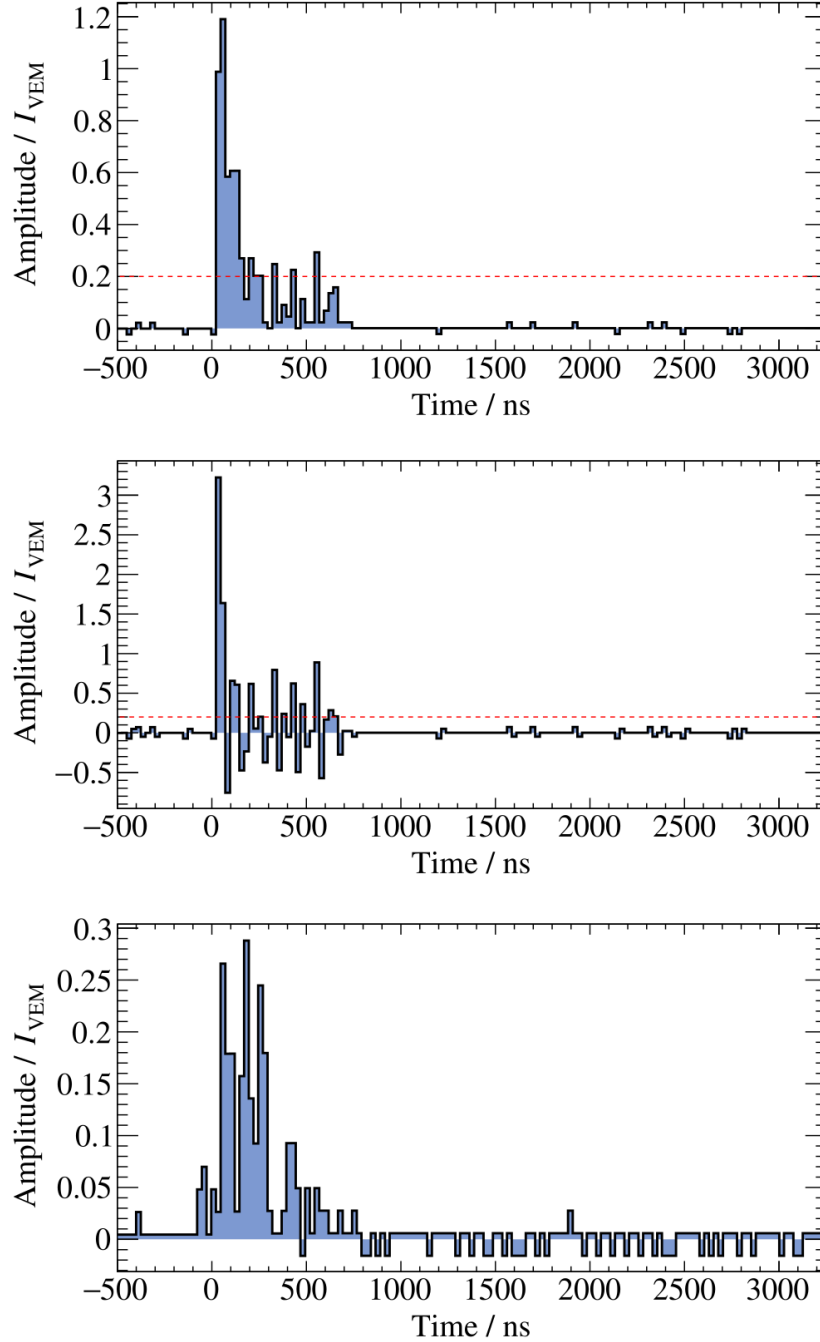


Figure 4.4: Illustration of the ToTd and MoPS trigger algorithms implemented in the SD stations. Originally published in Ref. [105]. *Top:* Example waveform passing the ToTd condition. *Middle:* Deconvolution of the same waveform, with the threshold required by the algorithm indicated by the dashed red line. *Bottom:* Example waveform fulfilling the MoPS condition.

maintaining an acceptable rate of random triggers. In the context of this thesis, where the targeted energy range is the lowest accessible at the Observatory and the focus is on photon signals, these triggers are especially important. Unless explicitly stated, all simulated and recorded events employed in the presented analyses include stations triggered by ToTd and MoPS, in addition to those selected by the original triggers.

Trigger data are continuously sent to CDAS, which looks for global patterns. The global or central trigger (T3) initiates the data acquisition of the array by combining the local

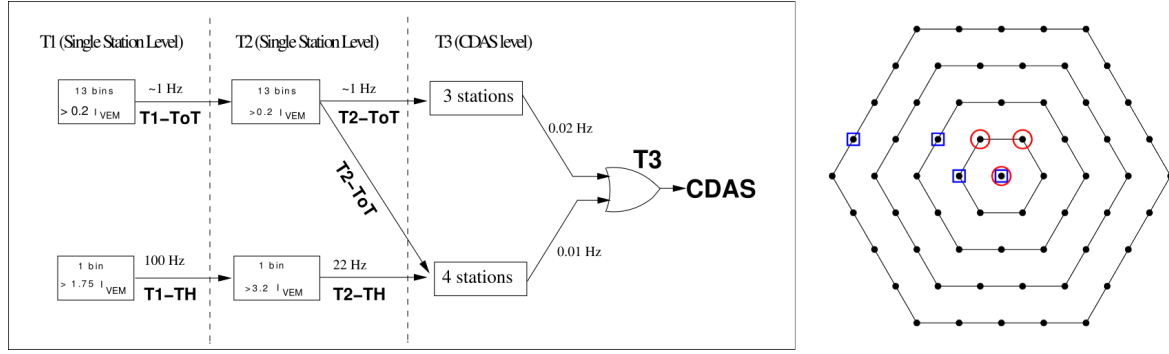


Figure 4.5: Hierarchy of the Surface Detector trigger system and geometric conditions for the global trigger definition. *Left:* Block diagram of the three trigger levels, from local T1 and T2 triggers to the global T3 trigger in CDAS, with the corresponding rates. *Right:* Illustration of the spatial compactness conditions: a 3-fold T2-ToT coincidence requires one neighbour in the first hexagon and one no further than the second, while a 4-fold coincidence of any T2 allows one station as far as the fourth hexagon. Examples of a 4-fold coincidence (blue squares) and a 3-fold coincidence (red circles) are highlighted. Figures taken from Ref. [100].

T2 triggers from individual stations. The main shower trigger condition is based on the identification of clusters of T2s that are close both in time and in space. To this end, the Central Trigger processor first searches for time clusters by centering a window of $\pm 25 \mu\text{s}$ on each T2. Clusters with a multiplicity of three or more are then examined for spatial coincidence. Two modes of coincidence are implemented. The first requires at least three T2-ToT triggers in compact configuration, namely one of the triggered stations must be within the first hexagon of neighbors of another, while a third must lie within the second hexagon.

The second mode requires a four-fold coincidence of any type of T2 (either ToT or Th) with slightly looser compactness: one of the four stations may be as far as the fourth hexagon, provided that another is within the first and one more within the second. Once the spatial condition is satisfied, a timing requirement is imposed: each T2 must lie within $(6 + 5n) \mu\text{s}$ of the central station, with n denoting the hexagon number. When these criteria are fulfilled, a T3 message is issued requesting all FADC traces recorded within $30 \mu\text{s}$ of the central T2 from every station located within six hexagons of those participating in the trigger. A schematic of the hierarchical trigger system and the corresponding geometric conditions for the T3 definition are shown in Figure 4.5.

The logic of the global T3 trigger in the SD-433 differs from the main array implementation. The adaptation is motivated by the higher expected rate of extensive air showers in the energy range targeted by the SD-433, together with the smaller physical extent of the array, which permits a simplified and faster handling of local triggers. In the current implementation, all 19 stations of the SD-433 participate in the T3 formation through the XbFast trigger algorithm. The algorithm is based on a straightforward triple-nested loop, which selects the first occurrence of any three T2s within a coincidence window of $10 \mu\text{s}$, regardless of their relative geometry. This coincidence is promoted to the only T3 of that GPS second.

This simplified scheme contrasts with the main array trigger, where spatial compactness requirements (hexagon-based topologies) and multiple T3s per second are possible. In the SD-433 case, two hard-coded constraints are present: (i) only the first 999 T2s per second are considered, and (ii) only one T3 per second can be issued. As a consequence, the first valid triple of T2s is always taken, without additional checks on geometry. While in normal conditions these restrictions are not critical, in situations where an individual station

develops abnormally high trigger rates, there is the potential that noise-induced coincidences dominate the first T3 of a second, thereby limiting the registration of genuine air shower events. Therefore, permanent monitoring and accurate tagging of bad acquisition periods are required for adequate operation of the array.

These limitations do not represent flaws, but rather reflect the pragmatic choice of an algorithm tailored to the unique operating conditions of the SD-433. The design prioritizes robustness and rate reduction in an array where higher shower frequencies are expected, while minimising computational overhead at the CDAS. To validate the algorithm, a dedicated check of the GPS nanosecond distribution of verified shower events reconstructed by the SD-433 was conducted. No accumulation of events was found at any particular GPS nanosecond, neither in the overall array distribution nor in the single-hexagon subsets. This confirmed that the trigger operates as expected and that the system does not reach data-transfer saturation.

The data recorded at each SD station include the GPS-synchronised start time of the signal, the full FADC traces from PMTs, as well as detailed amplitude and timing information for each PMT. In addition, local trigger flags and calibration metadata are stored to ensure the proper interpretation of the raw signals. These quantities constitute the fundamental inputs to the reconstruction chain, which transforms detector-level signals into estimates of extensive air shower observables, such as the shower core position, arrival direction, and shower size.

4.3 SD Reconstruction

This section summarizes the reconstruction of showers recorded with the SD, following the explanation in Ref. [106]. The paper describes EAS restricted to $\theta < 60^\circ$ for the SD-1500 and reflects the two validated frameworks, *Observer* and *Herald*, which implement functionally equivalent steps with small model differences that serve as cross-checks. Here, we describe only the *Observer* framework, as it is the current standard within the Collaboration. Moreover, the explanation is tailored to address the particular aspects of the SD-433 reconstruction, which is limited to vertical showers of $\theta < 45^\circ$.

4.3.1 Station-level reconstruction

We start at station level, where the digitised FADC traces are converted to the reference unit of a VEM. Then baselines are estimated independently for each PMT/channel along the trace to avoid signal bias. The algorithm searches for constant segments of at least 40 bins ($1 \mu\text{s}$) whose amplitudes lie within $\bar{b} \pm \sigma_b$ (with $\sigma_b = 2$ analog-to-digital converter (ADC) counts; increased if necessary for rare noisy cases). A global baseline is assembled by stitching these segments and linearly interpolating the gaps in *integrated-signal* space, yielding a baseline value for every FADC bin. A special treatment is applied to saturated PMT signals, where the undershoot of the signal is considered by different interpolation approaches.

After baseline subtraction, candidate signal *fragments* are consecutive bins at least 3 ADC counts above baseline. neighbouring fragments are merged if: (i) the start of fragment B occurs within $500 \text{ ns} + \text{duration}(A)$ after fragment A ends, and (ii) either $\int A \geq 0.3 \int B$ or the peak of B is below $5 I_{\text{VEM}}^{\text{peak}}$. The three PMTs are then averaged to a station trace; only the segment with largest signal is retained, enforcing a minimum window of 250 ns to contain any sub-threshold tails. The leading edge of this segment defines the station start time.

The station signal is the integral of the (high-gain by default; low-gain if saturated) averaged PMT trace within the selected window. If high-gain saturates ($\gtrsim 60$ VEM), low-gain

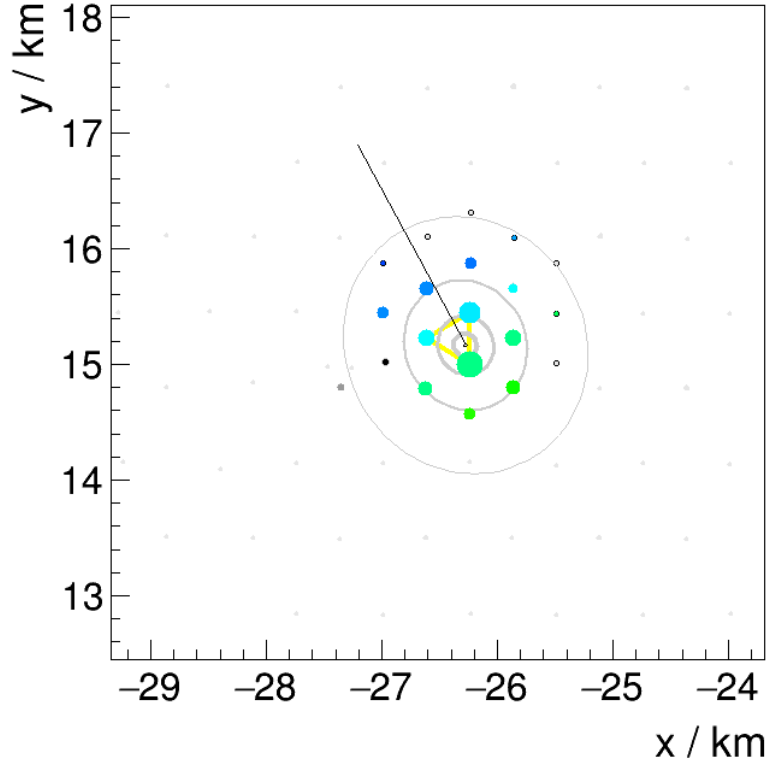


Figure 4.6: Footprint of the reconstructed SD event (auger ID 210807738800). Each marker corresponds to a triggered station; marker area scales with the measured signal (in VEM), and colour encodes the arrival time (blue = early, green = late). The yellow triangle is the seed triangle employed for the plane fit.

traces are used with the continuously monitored gain ratio and $Q_{\text{VEM}}^{\text{peak}}$. If low-gain also saturates ($\gtrsim 1000$ VEM), a recovery algorithm estimates the true signal from the undershoot amplitude; when PMT non-linearity curves are available, fits to non-saturated parts extend recovery up to $\sim 10^6$ VEM with $\sim 20\%$ resolution.

Twin-station data (~ 11 m separation) were employed to quantify time and signal fluctuations. The start-time variance adopts a general model

$$\sigma_{t_{\text{start}}}^2 = a^2 \left(\frac{2t_{50}}{n} \right)^2 \frac{n-1}{n+1} + b^2, \quad (4.1)$$

where t_{50} is the interval containing the first 50% of the integrated signal and $n = S/\bar{\ell}(\theta)$ is the effective particle count, with $\bar{\ell}(\theta)$ the mean track length in the cylindrical tank. Typical parameters are $a \simeq 0.7$ and $b \simeq 16$ ns. For non-saturated stations, the signal variance is nearly Poissonian:

$$\sigma_S^2 = f_S^2(\theta) S, \quad f_S(\theta) = 0.34 + 0.46 \sec \theta. \quad (4.2)$$

For recovered saturated signals, uncertainties grow from $\sim 5\%$ at ~ 1000 VEM up to $\gtrsim 50\%$ beyond 10^4 VEM due to PMT non-linearity and limited calibration. Once the signals and arrival times are reconstructed, the reconstruction sequence proceeds to event-level estimations and further refinements. Figure 4.6 shows the footprint on the ground of a reconstructed event.

4.3.2 Event-level reconstruction

Array triggering proceeds from local T1/T2 to central T3 coincidences. After a T3, stations upload traces and calibration/monitoring data. Offline, a physics trigger (T4) and a containment (T5) requirement are applied. For high-quality reconstruction, a strict 6T5 condition is always applied. All six neighbours of the hottest station should be active to ensure accurate core localisation. A *seed geometry* is built from the highest-signal, non-aligned seed triangle to fit a plane front and obtain initial estimates for the core position $\bar{x}_c$, and direction (u, v) of the unit vector $\hat{u} = (u, v, \sqrt{1 - u^2 - v^2})$; stations more than $-1 \mu\text{s}$ early or $+2 \mu\text{s}$ late are labelled accidentals and removed. Isolated triggers are also pruned; nearby non-triggered but active stations (up to $\sim 5 \text{ km}$) are attached for later likelihood constraints.

Then a new geometry fit minimises

$$\chi^2 = \sum_i \frac{[t_i - t_{\text{sh}}(\bar{x}_i)]^2}{\sigma_{t_i}^2}, \quad (4.3)$$

over the core time t_c , an enhanced estimation of the core position $\bar{x}_c$, direction (u, v) of the unit vector $\hat{u} = (u, v, \sqrt{1 - u^2 - v^2})$, and (when station multiplicity allows) a curvature parameter. The shower front spherical model is

$$c t_{\text{sh}}(\bar{x}) = c t_0 + |\bar{x} - \bar{x}_0|, \quad (4.4)$$

where t_0 is the start time at a virtual point $\bar{x}_0$. For multiplicity < 5 the curvature is fixed to a parameterisation from high-multiplicity events; otherwise it is fitted.

Signals are regressed versus perpendicular distance r using a fixed optimal distance $r_{\text{opt}} = 300 \text{ m}$, chosen to minimise shower-to-shower fluctuations for the 433 m triangular grid [107]. The size estimator is $S(300)$ with

$$S(r) = S(300) f_{\text{LDF}}(r), \quad f_{\text{LDF}}(300 \text{ m}) = 1, \quad (4.5)$$

where f_{LDF} is the lateral distribution function (LDF). The Observer uses a modified NKG [72] with a single β parameter:

$$f_{\text{LDF}}(r) = \left(\frac{r}{r_{\text{opt}}} \right)^\beta \left(\frac{r + r_s}{r_{\text{opt}} + r_s} \right)^\beta, \quad r_s = 700 \text{ m}. \quad (4.6)$$

Parameter β is determined from events with adequate lever arm (multiple stations between ~ 100 – 400 m , and separation $\gtrsim 175$ – 225 m) and parameterised versus θ and $S(300)$. No azimuthal asymmetry in the shower plane is considered at the time for the *Observer*.

The likelihood employed to fit the LDF combines information from all participating and nearby non-triggered stations given geometry:

$$\ln \mathcal{L} = \sum_i \ln P(S(300), \bar{x}_c | S_i, \bar{x}_i). \quad (4.7)$$

Observer factorises P into: (a) **small signals**, Poisson in an effective particle count matched to the variance model; (b) **large signals**, Gaussian $N(S_i; S(r_i), \sigma_i)$ with σ_i from the station-level variance; (c) **unrecoverable saturated**, one-sided integrals above the measured S_i ; (d) **non-triggered**, complementary trigger probability $1 - p_{\text{trig}}(S(r_i))$ derived from data. Figure 4.7 shows an example LDF fit with this likelihood, where markers denote stations, the solid line the best fit, and the shaded area the fit uncertainty.

When the lever-arm criterion is satisfied, the slope parameter β is left free; otherwise, fixed parameterisations are employed. The latter choice yields at most a $\lesssim 3$ – 4% bias in

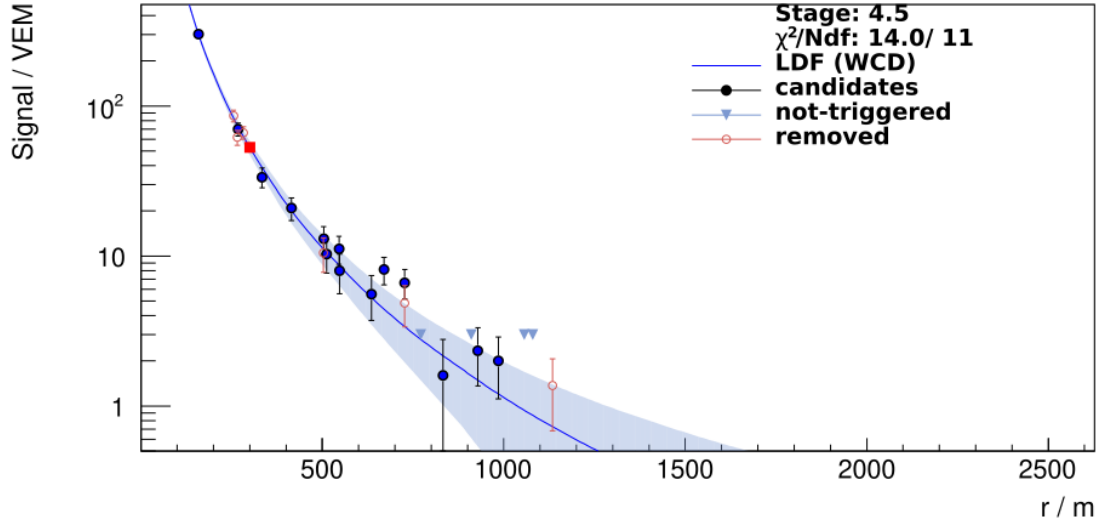


Figure 4.7: Lateral distribution function (LDF) fit for the SD-433 event (auger ID 210807738800). Markers denote stations: filled circles are stations used in the fit; open red circles are rejected stations because they belong to a different grid or have different electronics; and triangles are non-triggered SD-433 stations. The solid line shows the best-fit LDF, and the shaded band indicates the fit uncertainty.

$S(300)$ (slightly larger under saturation). Moreover, although all trigger types are used in the SD-433 reconstruction, dedicated likelihood terms for electromagnetic triggers are not yet included, from which additional, though still small, biases in $S(300)$ may be introduced. In practice, these effects remain limited and are largely absorbed by a subsequent energy calibration.

Details of the atmospheric attenuation correction applied to $S(300)$ and of the subsequent energy calibration of the corrected observable S_{30} , which defines the SD-433 energy estimator $E_{\text{SD-433}}$, are presented in Chapter 5, a chapter devoted to the detector’s energy response.

4.4 Fluorescence detector

The Fluorescence Detector (FD) of the Pierre Auger Observatory consists of twenty-seven wide-field Schmidt telescopes distributed across four sites surrounding the SD array: Los Leones, Los Morados, Loma Amarilla, and Coihueco (Fig. 4.8). Each site hosts six telescopes that overlook the array, except Coihueco, which includes the three additional High Elevation Auger Telescopes (HEAT) designed to observe showers at higher elevations and lower energies.

Each telescope covers a field of view of $30^\circ \times 30^\circ$ (azimuth $\times$ elevation). Fluorescence light in the $\sim 300\text{--}430\text{ nm}$ band enters through a large UV-pass window (Schott MUG-6) at the diaphragm, traverses an annular corrector ring that maintains small spot sizes across the focal plane, and is focused by a $\sim 10\text{ m}^2$ segmented spherical mirror onto a spherical focal surface instrumented with a camera of 440 hexagonal pixels. Fig. 4.9 shows the arrangement of the telescope components.

Pixels are 1.5° each and are implemented with Photonis XP3062 PMTs equipped with light collectors to minimise dead area. Signals are digitised every 100 ns and processed through a multi-level trigger chain that selects track-like patterns produced by extensive air showers. Absolute and relative calibrations, together with continuous atmospheric monitoring, enable calorimetric energy and X_{max} measurements [108].

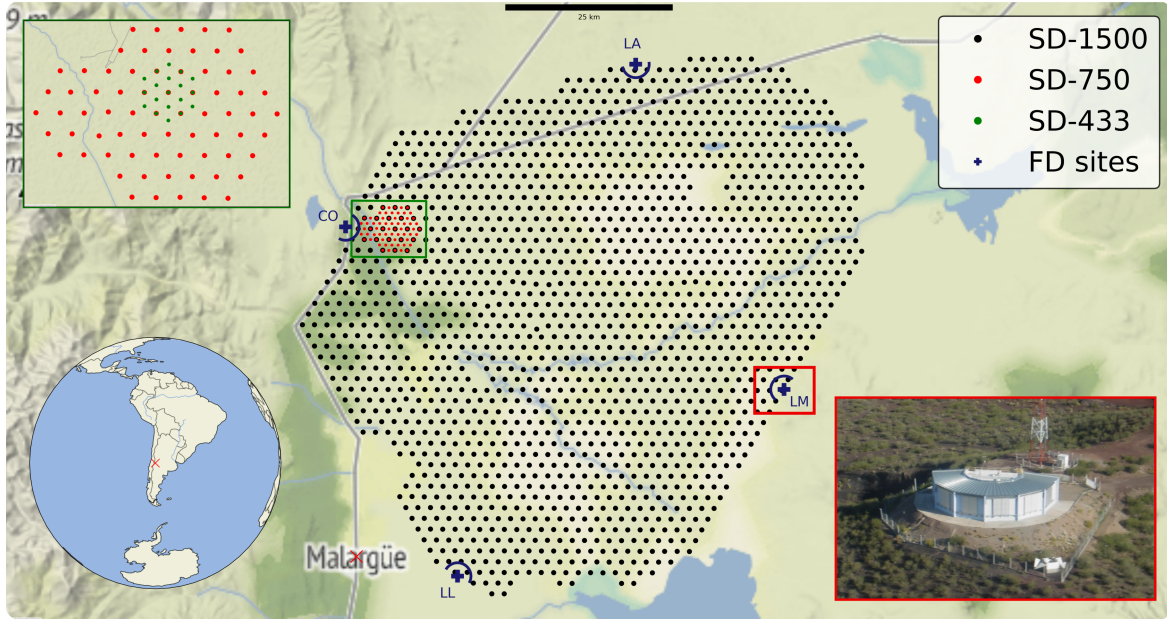


Figure 4.8: Layout of the Pierre Auger Observatory. The SD comprises ~ 1660 water-Cherenkov stations on a triangular grid with 1.5 km spacing (black dots), covering an area of about 3000 km^2 . The denser infill arrays, SD-750 (red) and SD-433 (green), are deployed near the Coihueco site to extend the energy range toward lower energies. The four FD sites, Los Leones (LL), Los Morados (LM), Loma Amarilla (LA), and Coihueco (CO), are shown as blue crosses on the periphery of the array. The inset on the top left highlights the nested infill configurations, while the lower-right inset shows one of the FD buildings.

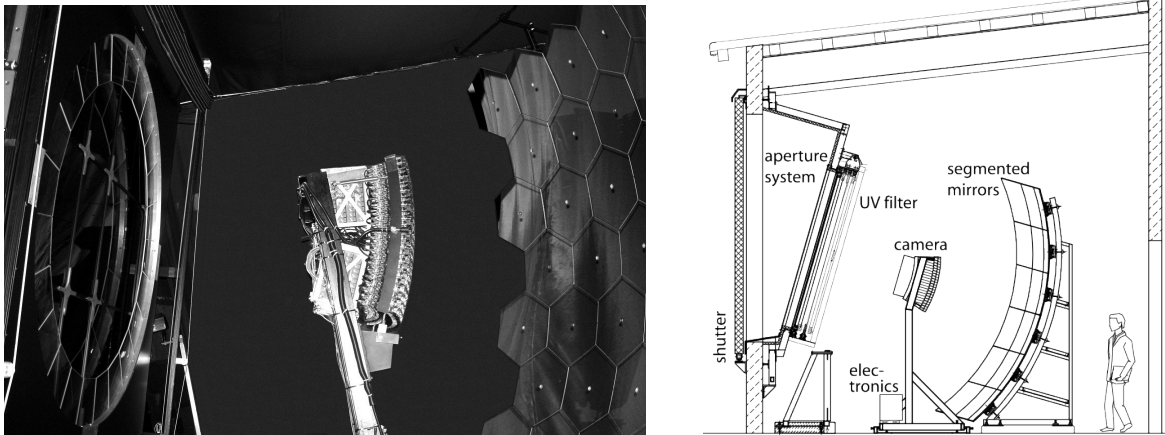


Figure 4.9: *Left:* Photo of a fluorescence telescope at Coihueco. *Right:* Schematic view of a fluorescence telescope with a description of its main components. Both figures are taken from Ref. [100].

The HEAT are three additional fluorescence telescopes located near Coihueco. They are housed in tiltable enclosures that raise the elevation field of view by roughly 30° (and up to $\sim 60^\circ$ elevation), enabling the observation of nearby, lower-energy showers and providing hybrid overlap with the infill arrays (i.e., SD-750 and SD-433). The HEAT telescopes share the core optical and camera design with the standard FD but is optimised for fast slewing and stable alignment at higher elevation. This extended coverage improves composition and spectrum measurements down to $\sim 10^{17}$ – 10^{18} eV and adds valuable cross-calibration with surface infill detectors.

Charged particles in an air shower excite atmospheric nitrogen, which subsequently emits fluorescence photons; the emitted light is approximately proportional to the local energy deposit. After correcting the recorded light for atmospheric transmission (molecular/aerosol scattering), subtracting the Cherenkov component, and applying the wavelength-dependent fluorescence yield, the *longitudinal energy-deposit profile* $dE/dX(X)$ is reconstructed. A Gaisser–Hillas function is typically fit to obtain the calorimetric energy and the depth of maximum, $X_{\max}$. Geometry is determined from the pointing directions and timing of triggered pixels: the shower-detector plane is defined by the camera track, and the axis within this plane is fixed by the pixel time sequence. Stereo observations or, more commonly, hybrid timing with one or more SD stations break degeneracies and yield sub-degree directions and metre-scale core resolutions at the highest energies [108]. A high-quality stereo event is shown in the left panel of Fig. 4.10.

The FD provides a nearly calorimetric energy scale that is used to *calibrate* the SD energy estimators. In hybrid events, the SD lateral signal at a reference distance (e.g., $S(1000)$ for the SD-1500) is corrected for zenith-angle attenuation to a reference zenith (e.g., S_{38} after correction for atmospheric attenuation and zenith-dependent effects). A power-law relation,

$$E_{\text{FD}} = A (S_{\text{ref}})^B, \quad (4.8)$$

is then fit to a high-quality hybrid sample to obtain (A, B) . The fit is applied to all SD events to assign energies on the FD scale while retaining the SD’s duty cycle. The scatter around the calibration captures shower-to-shower fluctuations and residual systematics; dominant systematics in the absolute scale arise from fluorescence yield, absolute calibration of the telescopes, and atmospheric transmission. The right panel of Fig. 4.10 illustrates typical SD observables regressed against E_{FD} . Observables S_{35} , S_{38} , and N_{19} correspond to SD-750 vertical events, SD-1500 vertical events, and SD-1500 inclined events, respectively [109]. The energy calibration and energy response of the SD-433 are discussed in Chapter 5.

Since the atmosphere acts as a vast and inhomogeneous calorimeter for extensive air showers, both the FD and SD measurements depend sensitively on the prevailing environmental conditions [111]. Although the impact is more critical for the FD, whose fluorescence light yield and transmission directly depend on atmospheric parameters, variations in temperature and air density also influence the SD. For instance, during cold days the denser atmosphere increases the slant depth above the array, so that showers reach a more advanced stage of development when they are detected at ground level. Consequently, continuous monitoring of the atmospheric state is an essential element of the Observatory’s operation.

Meteorological stations distributed across the site record temperature and pressure in real time at multiple locations. At the FD sites, infrared cloud cameras provide estimates of cloud coverage and transparency conditions for each telescope’s field of view. To characterise the vertical atmospheric profile, two laser facilities [112], the Central Laser Facility (CLF) and the eXtreme Laser Facility (XLF), emit calibrated ultraviolet laser pulses of known wavelength and intensity into the atmosphere. The light scattered from these shots is simultaneously detected by all FD eyes, providing an end-to-end calibration and a measure of the aerosol attenuation relevant for the energy reconstruction.

An independent cross-check of these measurements is provided by the Light Detection and Ranging (LiDaR) systems installed at each FD site [113]. Each lidar periodically scans the sky by sending short laser pulses and recording the backscattered signal, allowing the reconstruction of cloud coverage, cloud-base height, and the time evolution of aerosol profiles across the Observatory.

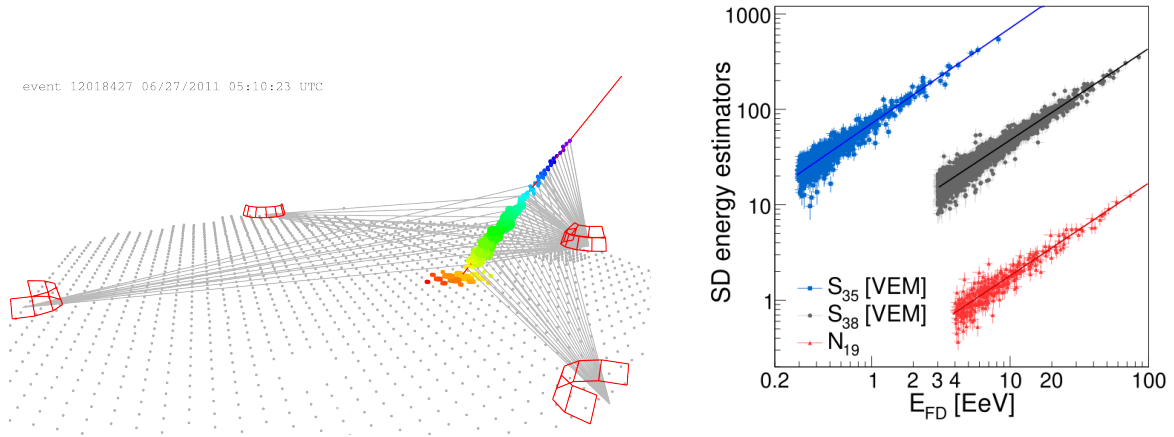


Figure 4.10: *Left:* Example of an UHE cosmic-ray event observed by the Pierre Auger Observatory with both the SD and the four FD sites, including the three HEAT telescopes. Colours indicate the signal times at the shower and at ground level, while the marker sizes are proportional to the magnitude of the measured signals on the ground and the reconstructed energy deposits along the shower track. The red squares outline the fields of view of the 27 fluorescence telescopes, with grey lines marking the viewing directions of the triggered pixels, and grey dots representing the SD stations. The combined reconstruction from the four fluorescence sites yields an average energy of $E = (4.7 \pm 0.1) \times 10^{19}$ eV and a depth of shower maximum of $X_{\max} = (730 \pm 3)$ g cm $^{-2}$. Extracted from Ref. [110]. *Right:* Energy calibration of the SD. The shower size measured for near-vertical events with the SD-1500 (S_{38}) and SD-750 (S_{35}) arrays, as well as for inclined showers (N_{19}), is shown as a function of the energy reconstructed by the fluorescence detector (E_{FD}). Each point represents a hybrid event observed simultaneously by the FD and SD, and the solid lines indicate the best-fit calibration relations used to transfer the nearly calorimetric FD energy scale to the surface arrays. Taken from Ref. [109].

4.5 AugerPrime

After more than two decades of successful operation, the Pierre Auger Observatory has reached a stage where many of the central questions in UHECR physics require new, more precise observables. The current surface and fluorescence detectors have provided decisive evidence for a suppression in the cosmic-ray flux above $10^{19.5}$ eV, yet the origin of this feature, whether it arises from the exhaustion of acceleration at astrophysical sources or from energy losses during propagation, remains uncertain. Moreover, discrepancies between data and simulations, particularly concerning the number of muons in EAS, highlight limitations in hadronic interaction models at energies beyond the reach of terrestrial accelerators [114].

To address these challenges, the Auger Collaboration initiated the AugerPrime upgrade programme. Its central goal is to improve mass composition sensitivity at the highest energies by distinguishing the electromagnetic and muonic components of air showers on an event-by-event basis. This enhanced capability is essential for identifying the nature of the particles responsible for the most energetic events observed on Earth, testing hadronic interaction models, and constraining the origin of the flux suppression [115]. Additionally, extending the lifetime of the Observatory well into the 2030s will triple the accumulated statistics and enable a broad range of multi-messenger studies.

AugerPrime introduces new detectors and upgraded electronics across the 3000 km 2 array while preserving continuous data taking. The main components of the upgrade are:

- a Scintillator Surface Detector (SSD), consisting of scintillation modules placed atop each existing WCD;

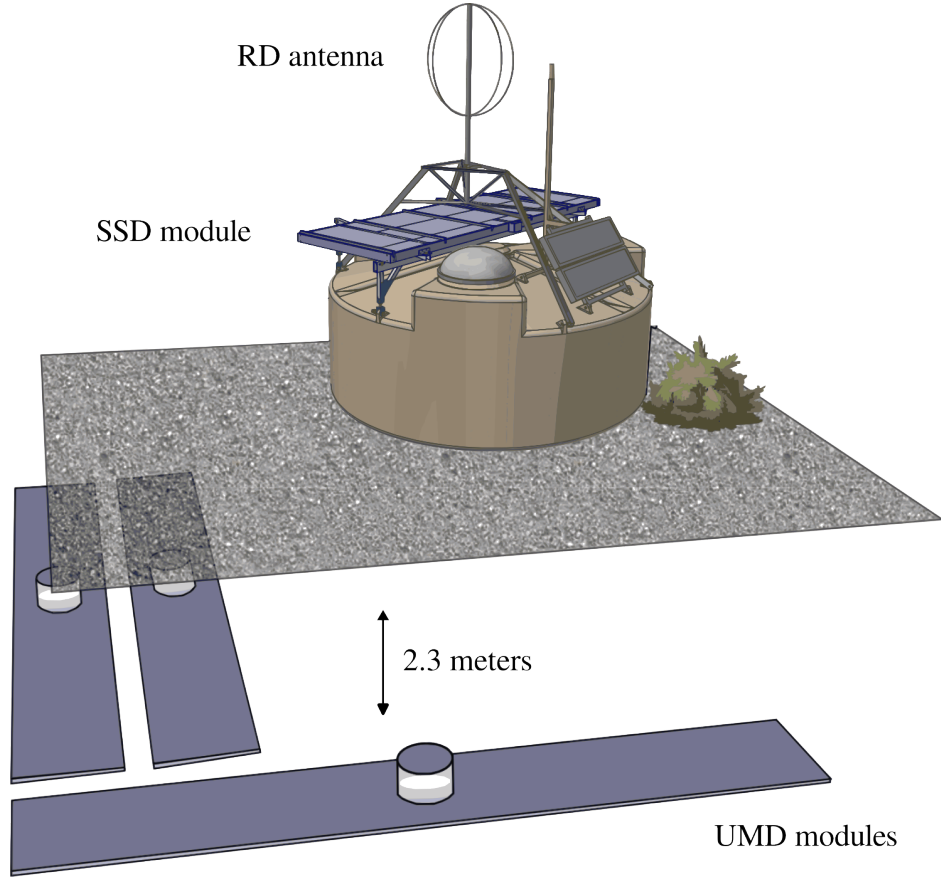


Figure 4.11: Schematic view of an AugerPrime station. Each surface detector WCD is complemented with a plastic SSD module placed on top and the RD antenna mounted on the supporting frame. Beneath the surface, three UMD modules are buried at a depth of approximately 2.3 m to shield them from the electromagnetic component of the shower. Together with the WCD, these detectors enable a multi-hybrid reconstruction of the air-shower components, combining electromagnetic, muonic, and radio observables in a single location.

- an Underground Muon Detector (UMD) for direct measurement of the muonic component;
- upgraded electronics and small PMTs (SPMTs) for increased dynamic range and advanced triggering;
- an extension of the FD duty cycle through modified PMT operation;
- and the addition of the Radio Detector (RD), which consists of antennae at each surface station for nearly 100% duty cycle composition measurements [116].

In the following subsections, we summarise these components, except for the UMD, which is treated separately because of its central role in this work. Figure 4.11 shows a schematic representation of an upgraded station.

Scintillator Surface Detector

The SSD consists of thin plastic scintillator planes ($4 \text{ m} \times 2 \text{ m}$) mounted on top of each WCD station. Each station unit consists of two scintillator modules with a combined active area of 3.8 m^2 , segmented into extruded strips with wavelength-shifting fibres guiding the

light to a single PMT [114]. This configuration provides a nearly uniform response to all charged particles and is therefore more sensitive to the electromagnetic component of the air shower, complementing the WCDs, which are more sensitive to muons.

By combining the SSD and WCDs signals, AugerPrime enables a decomposition of the shower footprint into its electromagnetic and muonic contributions. This synergy significantly enhances the accuracy of composition-sensitive observables and allows mass estimation on a shower-to-shower basis.

Radio Detector

The radio enhancement capitalizes on the success of the Auger Engineering Radio Array (AERA) and extends radio detection capabilities to the full array. Each SD-1500 station will host a radio antenna sensitive to the 30–80 MHz range. Radio detection complements the optical and particle measurements by providing access to the electromagnetic component with a duty cycle close to 100%. This capability is particularly valuable for inclined air showers, for which the scintillator upgrade is less effective. The radio data will improve the reconstruction of the shower direction, energy, and depth of maximum, and enhance the overall mass sensitivity of the Observatory [117].

Additional upgrades

Beyond new detector hardware, AugerPrime introduces several critical improvements to existing subsystems:

- **Electronics and Dynamic Range:** The original front-end electronics called Unified Board (UB) is replaced by the Upgraded Unified Board (UUB), tripling the sampling frequency (from 40 MHz to 120 MHz) and increasing the ADC resolution from 10 to 12 bits [118]. A SPMT is added to each WCD to extend the dynamic range up to signals corresponding to several thousand VEMs. Usually, changes in the electronics define what is referred to within the Collaboration as the data-taking periods known as Phase I and Phase II. This thesis focuses exclusively on Phase I data.
- **Uptime of the FD:** The telescope's PMTs will operate in an additional low-gain mode, allowing data acquisition during nights with higher background illumination (e.g., partial moonlight). This change increases the FD duty cycle from $\sim 19\%$ to about 29%, thereby improving hybrid statistics and mass discrimination at the highest energies [114].

Together, these upgrades constitute the most significant enhancement to the Pierre Auger Observatory since its construction. By combining multiple complementary detection techniques, AugerPrime will achieve unprecedented sensitivity to the mass composition and hadronic structure of ultra-high-energy cosmic rays, paving the way for the next decade of discovery in astroparticle physics.

4.6 Underground Muon Detector

The UMD provides a direct, background-suppressed measurement of the muonic component of EAS by means of segmented plastic-scintillator counters buried at a depth of approximately 2.3 m. The local soil density of $\sim 2.38 \text{ g cm}^{-3}$ [119] corresponds to a mass overburden of $\sim 540 \text{ g cm}^{-2}$ [120], which sets a vertical muon energy threshold of the order of 1 GeV. This effectively filters out the electromagnetic component of air showers, enabling a clean measurement of composition-sensitive observables derived from the muon content.

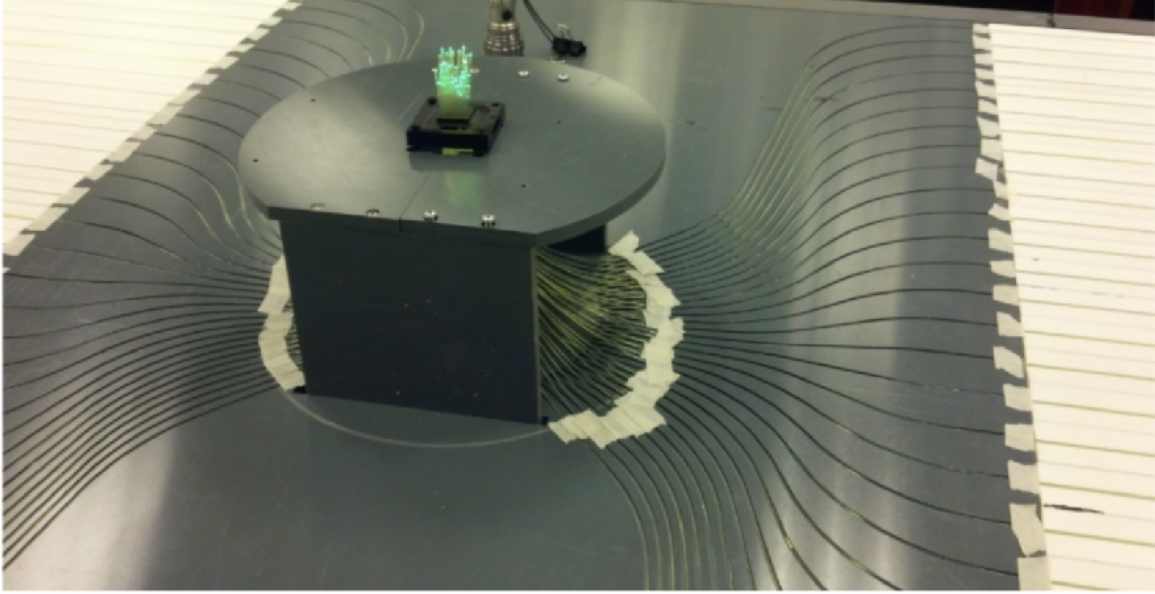


Figure 4.12: Assembly of a UMD module. The 64 WLS optical fibres from the scintillator strips converge toward the central optical coupling, or *cookie*, which will connect to the 8×8 SiPM array and front-end electronics. Each fibre collects the scintillation light from one strip and guides it to the photodetector. The photo shows a module during laboratory assembly before encapsulation in its protective casing. Taken from Ref. [121]

In this thesis, the UMD serves as the primary instrument for photon–hadron discrimination and for estimating mass-sensitive observables, with its hardware and acquisition specifics motivating the reconstruction approach detailed in Section 4.7.

Each UMD station comprises three identical modules with a total effective area of approximately 30 m^2 per station. Each module is a $\sim 10 \text{ m}^2$ plastic–scintillator counter segmented into 64 strips, arranged in two panels of 32 strips each. Every strip measures $\sim 4.0 \text{ m}$ in length, $\sim 4.1 \text{ cm}$ in width, and $\sim 1.1 \text{ cm}$ in thickness, and is externally coated with a 0.25 mm reflective layer of polystyrene loaded with 12% TiO_2 to minimise photon losses. A straight, extruded wavelength-shifting (WLS) fibre runs along each strip and guides light toward a central optical coupling, or *cookie*, which mates the fibres to an 8×8 silicon photomultipliers (SiPM) array. The base scintillator material is Dow Styron 663W polystyrene doped with 1,4-bis((2,5-diphenyloxazole (PPO) and 1,4-bis(5-phenyloxazole-2-yl)benzene (POPOP) in weight fractions of about 1% and 0.03%, respectively [121].

The scintillation process begins when a charged particle traverses the plastic, depositing energy in the polystyrene matrix and exciting its molecular states. The base polymer emits photons in the ultraviolet, but these would be strongly attenuated within a few millimetres if no additional dopants were present. The first dopant, PPO, acts as an intermediate wavelength shifter: it absorbs the ultraviolet photons from the base and re-emits them at slightly longer UV wavelengths with a much greater attenuation length. However, the plastic is still not transparent to this radiation. A second dopant, POPOP, therefore absorbs the PPO emission and re-emits photons in the blue–visible region, to which the subsequent WLS fibre is highly efficient. The Saint-Gobain BCF-99-29AMC fibre then performs the final wavelength shift from blue to green, transporting the light by total internal reflection toward the SiPMs, where it is converted into electrical pulses. This cascade of absorptions and re-emissions, illustrated schematically in Fig. 4.13, optimizes both the light yield and transmission efficiency of the UMD modules [119, 121].

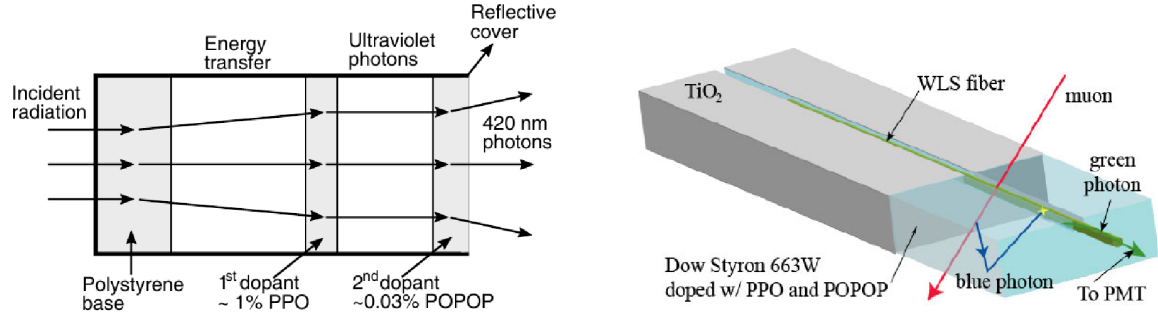


Figure 4.13: *Left:* Schematic representation of the scintillation and wavelength-shifting chain inside the plastic bars. The incident radiation excites the polystyrene base, which emits ultraviolet photons subsequently absorbed and re-emitted by the PPO and POPOP dopants at longer wavelengths (~ 420 nm). *Right:* Conceptual view of a scintillator strip coated with TiO₂ and traversed by a WLS optical fibre. Blue photons from the scintillator are absorbed by the fibre and re-emitted as green photons, which are transported by total internal reflection to the photosensor. Adapted from Refs. [119, 121]

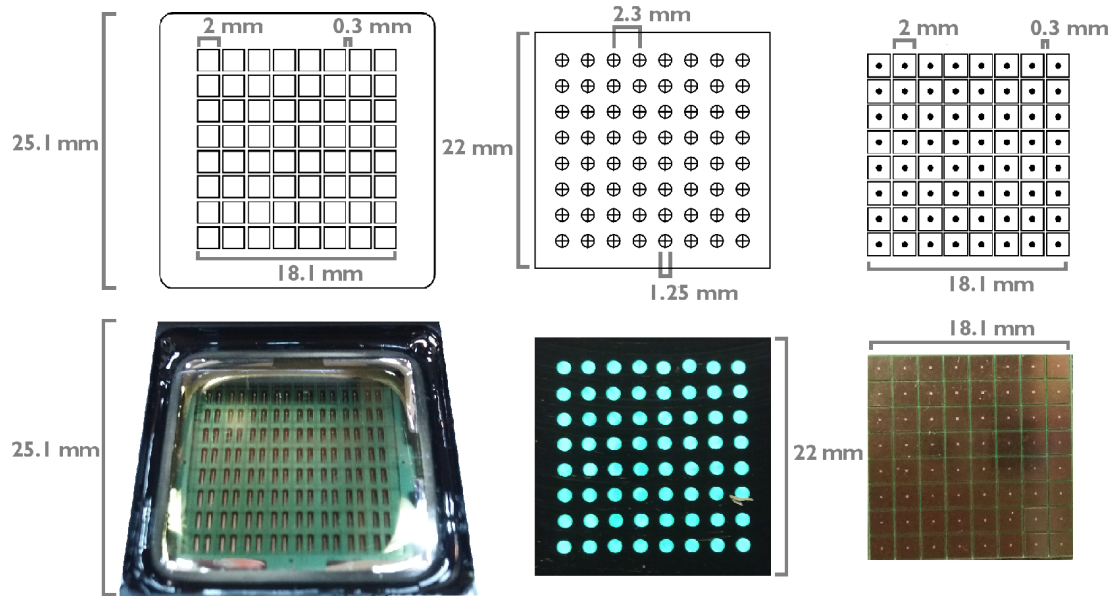


Figure 4.14: Comparison between the 8×8 multi-anode PMT (*left*), the optical connector (*middle*), and the SiPM array (*right*) used in the AMIGA/UMD modules. The schematics (*top*) show pixel layouts and external dimensions, while the photographs (*bottom*) display the corresponding physical devices.

During the engineering phase, the muon detector was called Auger Muons and Infill for the Ground Array (AMIGA) and operated with multi-anode PMT of type Hamamatsu H8804-200MOD, coupled to the optical connector through the same mechanical interface used today. More than one year of calibrated PMTs data was collected and analysed, leading to the first physics results from AMIGA [122]. After extensive evaluation, the PMTs were replaced by arrays of SiPMs, without requiring major modifications in the module structure [121]. Figure 4.14 compares the geometry and dimensions of the 8×8 multi-anode PMT, the optical connector, and the corresponding SiPM array, along with photographs of the devices.

The replacement was driven by several advantages of SiPMs over PMTs. In particular, SiPM arrays consume roughly ten times less power, cost up to four times less, and exhibit

higher photodetection efficiency at the emission peak of the WLS fibre (~ 492 nm): approximately 40% for the WLS versus $\sim 20\%$ quantum efficiency for the PMT. Furthermore, SiPMs eliminate inter-pixel crosstalk inherent to multi-anode PMTs. These benefits, combined with their compactness and robustness, made SiPMs an attractive choice for large-scale field deployment [123]. However, two limitations required dedicated solutions: the high thermal dark-count rate and the strong dependence of the SiPM gain on temperature. Both effects are mitigated by optimised discriminator thresholds, temperature monitoring, and an automatic bias-compensation circuit integrated in the front-end electronics [124].

Field tests were performed to validate the new design using prototype modules equipped with Hamamatsu S12572-100C SiPMs (3×3 mm² each) coupled to custom preamplifiers and shapers. Comparative studies of PMT and SiPM traces confirmed equivalent timing performance, with measured signal-width dispersions of $\sigma_{\text{PMT}} = (9.07 \pm 0.36)$ ns and $\sigma_{\text{SiPM}} = (9.37 \pm 0.32)$ ns [121].

In the production phase, UMD modules employ SiPM arrays operated in Geiger mode. A commonly used device is the Hamamatsu S13361-2050NE-08, featuring 1584 microcells with uniform breakdown characteristics. The bias voltage is actively adjusted to compensate for temperature variations, with an automatic gain correction of approximately 54 mV/°C around 25 °C. The dominant noise sources, namely dark counts and optical cross-talk, are largely suppressed by thresholding and by the binary pattern-recognition algorithm implemented in the digital front-end [124].

The UMD implements two complementary acquisition modes that operate simultaneously and record the same particles, as illustrated in Fig. 4.15.

- **Binary (counter) mode:** each of the 64 SiPM channels is processed by a CITIROC-based front-end consisting of a preamplifier, a fast shaper, and a discriminator. The discriminator output is sampled by an field-programmable gate array (FPGA) at 320 MHz (3.125 ns per sample), producing 6.4 μ s binary traces of 2048 bits per channel. A muon is identified as a short sequence of ones, and a pattern requirement of at least four consecutive ones (hereafter referred to as the *muon pattern*) provides robust separation from spurious hits due to dark noise or afterpulses. The nominal discriminator threshold is set to ~ 2.5 photon equivalents, efficiently suppressing dark-count-induced triggers.
- **ADC (integrator/calorimetric) mode:** the 64 analogue signals are summed and digitised by two ADCs operating at 160 MHz (6.25 ns sampling) into dual waveforms of 1024 samples each, corresponding to the same 6.4 μ s time window. The two amplification options, high gain (HG) and low gain (LG), extend the dynamic range, allowing accurate reconstruction of local muon densities over several orders of magnitude. The muon number is derived by normalising the integrated charge to the average single-muon response, as illustrated in Fig. 4.15.

Segmentation limits the instantaneous counting capacity of the binary mode, since coincident muons in the same strip may be recorded as one [120, 126]. The ADC mode therefore complements it by maintaining linearity in high-density regions [125]. Conversely, the binary mode provides superior statistical precision at low occupancies. In this work, we employ the *binary mode* as the primary acquisition channel for event-by-event muon counting and muon-LDF reconstruction, taking advantage of its per-strip segmentation and pattern-based muon identification. Given that photon-initiated showers exhibit low muon densities (Section 3.4) and that this thesis focuses on the lower-energy regime, the binary mode offers an optimal balance between simplicity and accuracy for the intended physical analyses.

The UMD is being deployed at 73 SD stations: 54 located within the SD-750 array, 12 within the SD-433 array, and 7 shared between both configurations (see inset in Fig. 4.8).

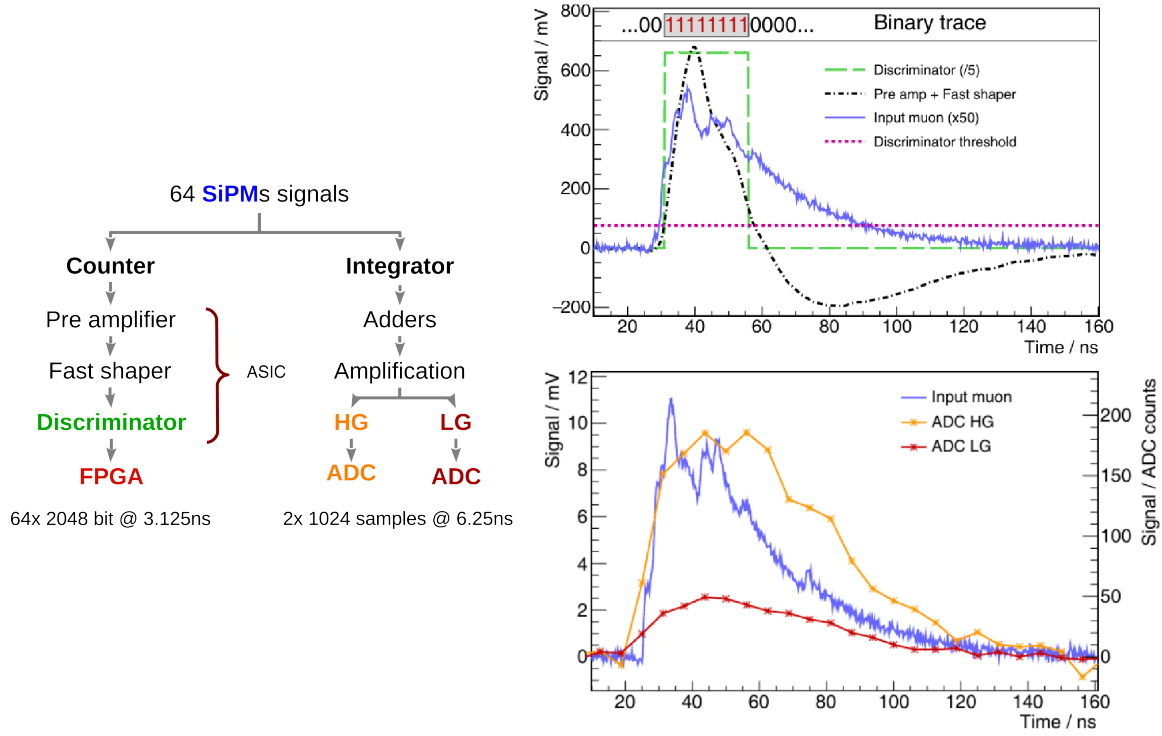


Figure 4.15: *Left:* Schematic diagram of the two UMD acquisition modes. Each of the 64 SiPM channels is exploited by the binary mode and an integrator/ADC mode. The counter mode processes signals through a preamplifier, fast shaper, and discriminator implemented in the CITIROC ASIC, with digital outputs sampled by an FPGA at 320 MHz. The integrator branch sums analogue signals and digitises them with high- and low-gain ADCs at 160 MHz. Taken from Ref. [121]. *Right:* Example waveforms for a simulated single muon pulse in both modes. The upper panel shows the fast-shaped signal and its corresponding binary trace; the lower panel shows the same muon pulse in high- and low-gain ADC channels. Taken from Ref. [125].

This corresponds to a total of 219 modules, with three buried scintillator modules per station. Each UMD station is installed within $\lesssim 20$ m of its paired WCD, enabling multi-hybrid operation with the SD, SSD, and RD at the same site, all synchronised by GPS timing. The UMD operates in slave mode, triggered by the local WCD trigger [127]. The close proximity ensures that all co-located detectors sample nearly identical lateral distances in each event, which is essential for composition-sensitive cross-checks and combined analyses.

At the time of writing, the deployment of the UMD is nearing completion. Figure 4.16 illustrates the ongoing installation process and the current deployment status across the array. The left panel shows modules being covered during the burial procedure at the Malargüe site, while the right panel displays the latest operational map: stations in green indicate modules currently taking data (*In ACQ*), and those in yellow correspond to stations already deployed but not yet in acquisition.

4.7 UMD Reconstruction

The UMD reconstruction proceeds in two main stages: a *module-level* stage, in which binary signals from individual scintillator modules are converted into local muon counts, and an *event-level* stage, in which these counts are combined to fit a global muon lateral distribution function (MLDF). In the following, we describe both stages in general terms to provide a complete overview of the reconstruction sequence. Further technical details,



Figure 4.16: *Left:* Installation of a UMD station during the burial phase at the Malargüe site. The three scintillator modules are positioned in a trench approximately 2.3 m deep, subsequently backfilled with local soil. The vertical conduit houses the optical and power connections to the surface electronics. *Right:* Deployment status of the UMD as of November 2025. Each circle represents one surface detector station: green markers denote stations in acquisition (*In ACQ*), yellow indicate stations deployed but not yet operational, and grey markers correspond to locations awaiting installation.

refinements, and the specific developments carried out as part of this thesis are presented in Chapter 6.

4.7.1 Module-level reconstruction

The purpose of the module-level reconstruction is to convert the binary traces of the scintillator strips into an estimate of the number of muons crossing each UMD module. Let $n_s = 64$ be the number of scintillator bars and k the number of bars that register at least one single-muon pattern within the full $6.4 \mu\text{s}$ acquisition window, sampled every 3.125 ns . Each muon traversing the detector can activate one or more bars, and the resulting pattern encodes the *occupancy* of the module.

Two related quantities are central to the reconstruction:

- μ : the *expected number of muons* impinging on the module, i.e. the mean of the underlying Poisson distribution that governs the muon arrivals;
- N_μ : the *actual number of muons* that crossed the module during the trace, a discrete realisation of that Poisson process, $N_\mu \sim \text{Poisson}(\mu)$.

Assuming uniformity in the local muon flux and statistical independence among bars, the probability of observing k fired bars for a given μ follows the *binomial occupancy likelihood* [128],

$$L(\mu) = P(k | \mu) = \binom{n_s}{k} e^{-\mu} (e^{\mu/n_s} - 1)^k. \quad (4.9)$$

Alternatively, when expressed in terms of the discrete number of muons N_μ , the likelihood takes the form [129]

$$L(N_\mu) = P(k | N_\mu) = \binom{n_s}{k} \frac{S(N_\mu, k)}{k!} n_s^{-N_\mu}, \quad (4.10)$$

where $S(N_\mu, k)$ are Stirling numbers of the second kind, representing the number of ways in which N_μ indistinguishable muons can occupy k distinct bars.

The corresponding maximum-likelihood estimators are

$$\hat{\mu} = -n_s \ln \left(1 - \frac{k}{n_s} \right), \quad (4.11)$$

$$\hat{N}_\mu \simeq \frac{\ln(1 - k/n_s)}{\ln(1 - 1/n_s)}. \quad (4.12)$$

These estimators inherently account for and correct the *pile-up* effect. This effect arises from the finite temporal resolution of the detector and occurs when two or more muons traverse the same bar within its ~ 37.5 ns inhibition time. In such cases, only one pattern is registered, leading to an undercount that increases with μ .

Since the early operation of the UMD, several time-windowing strategies have been explored to evaluate k from the binary traces [120]. They all rely on Eqs. (4.11)–(4.12), differing only in how the trace is divided into temporal windows:

- **Infinite-window:** the entire trace is treated as a single window. A strip contributes to k if it contains a single-muon pattern anywhere within the trace (four positive bins in a row). This approach provides a simple and direct estimate of the total number of muons impinging on the module and is the only strategy unaffected by undercounting due to undershooting in the fast-shaper signal. An example of a binary trace used in this procedure is shown in Fig. 4.17, where each strip contributes to the occupancy k if a single-muon pattern appears anywhere within the trace.
- **N-bin:** the trace is divided into windows equal in length to the single-muon pattern (12 bins). The occupancy is computed independently for each window and then summed,

$$\hat{\mu} = -\sum_{j=1}^{n_w} n_s \ln \left(1 - \frac{k_j}{n_s} \right). \quad (4.13)$$

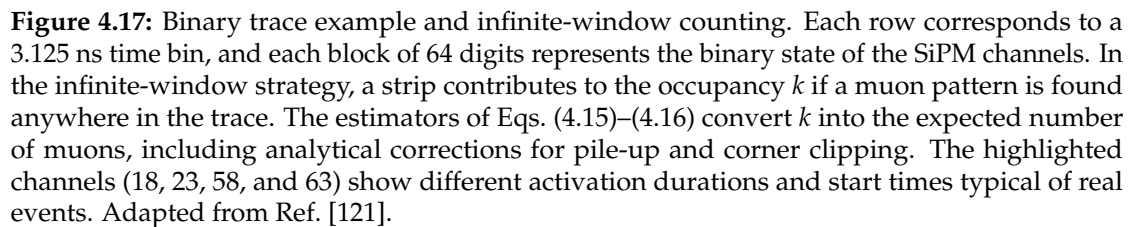
This method introduces correlations when muons traverse the same bar in consecutive windows.

- **N-bin centred:** a refinement of the previous approach, aligning one window with the signal maximum to reduce edge effects. It still suffers from artificial segmentation of the signal and limited applicability in low-statistics traces.
- **1-bin:** each time bin is treated as a separate window, explicitly correcting for inhibited (dead) bars due to the finite pattern duration. If $n_{\text{inh},j}$ bars are inactive in bin j , the effective segmentation becomes $n_s - n_{\text{inh},j}$, yielding

$$\hat{\mu}_j = -n_s \ln \left(1 - \frac{k_j}{n_s - n_{\text{inh},j}} \right), \quad \hat{\mu} = \sum_j \hat{\mu}_j. \quad (4.14)$$

This strategy allows time-resolved muon counting but remains sensitive to electronic undershoot and statistical fluctuations.

Comparative studies using both simulated and real traces demonstrated that, for realistic muon densities, all methods except the N-bin scheme perform within $\pm 10\%$ of the true value [120]. The **infinite-window** method was ultimately selected for the official reconstruction because it yields the smallest bias, is the least sensitive to electronic artefacts, and naturally aligns with the statistical likelihood used at the event level.


$$\hat{\mu} = -\frac{n_s}{1+p_{\text{cc}}} \ln\left(1 - \frac{k}{n_s}\right), \quad (4.15)$$

The presence of p_{cc} shifts $P(k)$ toward larger k values; applying this correction lowers the resulting $\hat{\mu}$ and $\hat{N}_{\mu}$ accordingly. A key advantage of this correction is that p_{cc} can be determined directly from data, making it less dependent on potential inaccuracies in simulations.

1. *Full detection efficiency*: every muon traversing a scintillator bar is assumed to produce the expected binary pattern. Effects such as light attenuation in the wavelength-shifting fibres or inefficiencies in the SiPM–CITIROC chain are not explicitly modelled at this stage.
2. *Negligible optical and electronic cross-talk*: light leakage between adjacent SiPM pixels and potential electrical interference between channels are considered sufficiently small to have no measurable impact on the number of bars flagged as active.

These approximations are justified for the current module-level reconstruction, where the leading sources of bias are already corrected analytically. The residual effects associated with optical attenuation and cross-talk are under investigation and are expected to introduce only second-order deviations from the reconstructed muon counts.

For $k = n_s$, the estimators of Eqs. (4.11) and (4.12) (and their p_{cc} -corrected forms) are not defined, as the likelihood increases monotonically with μ . To avoid divergence, a saturation threshold μ_{sat} is introduced, defined as the value of μ for which the probability of full saturation reaches 1 % [128]; i.e.

$$P(k = n_s \mid \mu_{\text{sat}}) = 0.01. \quad (4.17)$$

Solving for μ_{sat} using the occupancy model yields

$$\mu_{\text{sat}} = -n_s \ln\left(1 - 0.01^{\frac{1}{n_s}}\right), \quad (4.18)$$

following the criterion originally adopted in Ref. [128]. After incorporating the corner-clipping correction, the same condition is applied to the effective mean $\mu(1 + p_{cc})$, giving

$$\mu_{\text{sat}} = -\frac{n_s}{1 + p_{cc}} \ln\left(1 - 0.01^{\frac{1}{n_s}}\right). \quad (4.19)$$

This definition provides a consistent numerical threshold above which the probability of complete saturation exceeds 1 %. In practice, modules with $k = n_s$ are flagged as saturated and handled through dedicated likelihood terms at event level, where they constrain the fitted MLDF normalisation near the shower core.

In the event-level reconstruction, UMD modules are grouped into two classes: (i) *saturated* when all bars in the module are triggered ($k = n_s$); and (ii) *candidate* otherwise (modules that are not saturated and are not rejected by any quality criteria). These definitions are those used in the official reconstruction and determine how each module contributes to the likelihood at event level.

4.7.2 Event-level reconstruction

The goal at event level is to infer the MLDF, i.e., the expected muon density on the shower plane as a function of the shower-axis distance r . Beyond providing a physically interpretable description of the muon lateral spread, the MLDF enables a stable, event-wise summary variable at a fixed reference distance $\rho(r_{\text{ref}})$, which is widely used for composition-sensitive studies because it minimises model- and geometry-driven systematics when r_{ref} is appropriately chosen (typically around a few hundred metres for the UMD geometry). Parameterisations of the MLDF are commonly written in terms of $\rho(r_{\text{ref}})$ and a small set of shape parameters, so that $\rho(r_{\text{ref}})$ acts as the event-level normalisation to be fitted.

We first consider a flexible and frequently used family of functions derived from modified NKG forms and their re-parameterisations in terms of (α, β, γ) and a lateral scale r_0 ; see, for instance, the modified NKG variant:

$$\rho(r) = \rho(r_{\text{ref}}) \left(\frac{r}{r_{\text{ref}}}\right)^{-\alpha} \left(\frac{1 + r/r_0}{1 + r_{\text{ref}}/r_0}\right)^{-\beta} \left(\frac{1 + (r/10r_0)^2}{1 + (r_{\text{ref}}/10r_0)^2}\right)^{-\gamma}. \quad (4.20)$$

This form, employed also by the KASCADE-Grande Collaboration in the past, is widely accepted in the community as a suitable description of the MLDF [130–132].

Before fitting the MLDF, standard SD-level quality criteria are applied: the event must pass the physical T4 condition, have a successful LDF and energy reconstruction, and occur during stable acquisition periods. On the UMD side, at least three modules from distinct

stations with non-zero muon counts must be present. Additionally, the fiducial 6T5 condition, requiring the hottest SD station to be surrounded by six active neighbours, ensures a reliable core estimate, while analyses are limited to $\theta < 45^\circ$ to preserve sufficient projected area of the modules. In practice, the UMD fit treats $\rho(r_{\text{ref}})$ as a free parameter, fixes the slope β to an energy–zenith parametrisation derived from data, and fixes the core to the position reconstructed by the SD.

Several functional forms have been explored, motivated by NKG-like shapes and their modifications for muons. These can be cast into reference-distance parameterisations with fixed (α, γ) and a floating normalisation plus a limited number of shape terms. Such families have been tested using information criteria (AIC/BIC) and goodness-of-fit diagnostics, leading to practical recommendations on fixing a subset of shape parameters and choosing r_{ref} to minimise biases on $\rho(r_{\text{ref}})$ [120]. Unless explicitly indicated, the parameters (α, γ) are assumed to remain fixed at pre-validated values chosen to ensure the robustness of the MLDF fit.

The official UMD event-level reconstruction employs a binomial likelihood [126] for the MLDF fit. The key idea is to maximise the *joint* likelihood of the raw number of active bars k_i across all UMD modules, using the occupancy model that naturally incorporates segmentation and includes a correction for corner clipping via $p_{\text{cc},i}$. In this framework, *candidate* and *saturated* modules contribute with the same likelihood for k_i given the expected mean number of muons $\mu(r_i; \mathbf{p})$ at the module location r_i .

Writing n_s for the number of available strips per module, the event likelihood reads

$$\mathcal{L}_{\text{UMD}}(\mathbf{p}) = \prod_{i \in \text{cand} \cup \text{sat}} \binom{n_s}{k_i} e^{-\mu_i(1+p_{\text{cc},i})} \left(e^{\mu_i(1+p_{\text{cc},i})/n_s} - 1 \right)^{k_i}, \quad (4.21)$$

where $\mu_i = \mu(r_i; \mathbf{p})$ and $\mathbf{p}$ denotes the set of MLDF parameters (e.g., $\rho(r_{\text{ref}})$, (x, y) , and possibly β). This formulation operates directly on the binary observables k_i , avoiding an intermediate N_μ estimation for candidate and saturated modules. Saturated modules ($k_i = n_s$) behave as *lower-limit* constraints on the local expected muon count: their likelihood becomes step-like, effectively bounding μ_i from below and thus constraining the MLDF normalisation from the low side.

To mitigate biases arising from uncertainties in the shower-core position, the reconstruction can either (i) keep the UMD core fixed to the SD core, or (ii) allow the UMD core coordinates (x, y) to vary, constrained by a Gaussian penalty centred on the SD estimate and using the covariance matrix from the SD fit. The latter option improves the MLDF reconstruction when UMD stations lie close to the shower axis, as it naturally propagates the core uncertainty into $\rho(r_{\text{ref}})$. This is implemented by multiplying the likelihood by a penalty term of the form $\exp[-\chi^2(x, y)/2]$, leading to

$$\mathcal{L}_{\text{tot}}(\mathbf{p}, x, y) = \mathcal{L}_{\text{UMD}}(\mathbf{p}) \times \exp\left[-\frac{1}{2}\chi^2(x, y)\right], \quad (4.22)$$

where $\chi^2(x, y)$ is computed from the SD core covariance matrix [126]. This constrained-core option is particularly useful when a sufficient number of UMD modules are available, but the number of degrees of freedom remains limited by the total module count participating in the fit.

The slope parameter β can also be left free when the station sampling around the reference distance is dense enough and sufficiently well distributed in shower-axis distance. Otherwise, β is fixed to an empirical parameterisation that depends on both energy and zenith angle, obtained from the subset of well-reconstructed events where β could be reliably fitted. This adaptive approach ensures stable convergence for events with low multiplicity of modules, while retaining sensitivity in high-quality events.

Selecting a reference distance where the MLDF is least sensitive to the specific choice of functional form and to core-position uncertainties provides a robust event-level observable $\rho(r_{\text{ref}})$. Fixing a subset of shape parameters and using an optimised r_{ref} reduces systematic biases on $\rho(r_{\text{ref}})$ to the percent level, enabling precise physics analyses and consistent comparisons across data-taking periods and detector configurations. In practice, fixing β enhances stability and uniformity across events, while allowing it to vary in well-sampled cases provides valuable information on the lateral shape evolution of the muon component.

4.8 Offline Framework

The Offline Analysis Framework [133, 134] constitutes one of the cornerstones of this thesis, as it provides the most reliable and up-to-date simulation and reconstruction tools of the Pierre Auger Observatory. Developed in C++ within the Collaboration, it offers a unified environment for detector simulation, event reconstruction, and data analysis. Its design follows a modular philosophy: individual processing tasks are encapsulated in *modules* that can be flexibly combined through XML-based steering files, commonly referred to as *bootstraps*. This architecture allows collaborators to assemble complex applications, from raw data decoding to hybrid event reconstruction, by sequencing reusable and independently validated components.

The framework integrates several key systems: a centralised configuration mechanism that guarantees reproducibility through XML validation and logging; a comprehensive detector description that provides access to the evolving geometry, calibration, and atmospheric conditions; and an event data model that serves as a communication backbone among modules. All these are complemented by an extensive suite of utilities, including a geometry package supporting abstract coordinate transformations, and dedicated tools for serialisation, unit testing, and version control. Data persistence is based on the ROOT format, with the so-called Auger Data Summary Trees (ADSTs) providing an efficient structure for high-level physics analysis.

Over more than two decades of continuous operation, Offline has demonstrated exceptional flexibility and longevity. Its infrastructure has successfully evolved to accommodate major observatory upgrades, from the original SD and FD to the inclusion of AMIGA underground scintillators, the HEAT telescopes, and the AugerPrime extensions. This adaptability, combined with an emphasis on transparency and reproducibility, has made it a reference framework for cosmic-ray experiments beyond Auger.

In this thesis, we devoted a substantial portion of the work to the development, validation, and maintenance of Offline modules related to the SD-433 and UMD. A major effort was also dedicated to the standardisation of the corresponding reconstruction and simulation applications, ensuring their robustness and consistency across analyses. As a result, the *Standard Applications* developed as part of this work are now routinely employed in other studies within the Collaboration, well beyond the scope of the analyses presented in this thesis. The specific developments, extensions, and validation procedures are detailed in later chapters and complemented with dedicated appendices where appropriate.

SD-433 energy response

He who stands on tiptoe doesn't stand firm. He who rushes ahead doesn't go far.

Tao Te Ching, Lao Tzu

As the lower-energy extension of the Surface Detector, the SD-433 array plays a central role in characterising cosmic rays in the energy region of the second knee. It consists of 19 water-Cherenkov stations arranged in a hexagonal pattern nested within the SD-750 array, providing sensitivity to primary energies from a few tens to several hundreds of PeV. In this energy regime, the cosmic-ray flux exhibits a steeply falling power-law behaviour. Subtle features in the spectral index, such as the second knee, are believed to reflect the properties of astrophysical accelerators, the mass composition of cosmic rays at the sources, and the transition between galactic and extragalactic components.

To define a robust energy scale, it is necessary to fully characterise how the SD-433 array responds to air showers induced by different primaries, with diverse energies and zenith angles. Two detector-level ingredients are particularly important:

1. The detection efficiency, which quantifies the probability that a cosmic ray with energy E and zenith angle θ triggers and is successfully reconstructed by the array.
2. The energy response function, which describes how the true primary energy E is mapped into a reconstructed energy estimator E^{rec} . Following the formalism adopted for the SD-1500 array, the response function can be written as

$$r(E^{rec} | E) = \varepsilon(E) s(E^{rec} | E), \quad (5.1)$$

where $\varepsilon(E)$ denotes the trigger-and-reconstruction efficiency, and s is the spread function, which incorporates shower-to-shower fluctuations and detector effects. The spread is commonly described by a Gaussian distribution centred on the expected shower signal, optionally including a reconstruction bias.

Hadronic and photon-induced air showers produce markedly different particle content at ground level: while hadronic showers contain a significant muonic component contributing to the surface signal, photon showers are almost purely electromagnetic and lead to smaller signals for the same primary energy. As a consequence, a photon with true energy E_γ would be systematically underestimated if reconstructed using a calibration based on hadronic simulations alone. Accounting for these differences is crucial when deriving physics quantities such as integral photon-flux limits in terms of photon energy.

In addition, photon searches require accurate modelling of the background cosmic-ray flux, since rare fluctuations in hadronic showers can mimic photon-like signatures. Any mischaracterisation of the detector response to cosmic rays directly impacts the estimated rate of false positives, and thus the confidence in identifying a photon-candidate population. Therefore, the response-function studies presented in this chapter provide necessary input not only for a precise characterisation of cosmic rays near the second knee, but also for genuinely constraining photon detection prospects in the SD-433 energy range.

The energy estimation of the SD-433 array follows the same conceptual strategy adopted for the larger Auger surface arrays, while requiring specific adaptations due to its lower energy range and limited hybrid statistics. The reconstructed shower size is obtained by fitting the lateral distribution of the WCD signals using a modified NKG parameterisation, and evaluating the fitted function at a reference distance of 300 m from the shower axis, defining the observable $S(300)$.

As the surface signal attenuates with increasing zenith angle, a correction is required to define an energy estimator that is independent of the arrival direction. This correction is derived using the Constant Intensity Cut (CIC) method, which exploits the approximate isotropy of the cosmic-ray flux. Reconstructed events are grouped into bins of equal width in $\sin^2 \theta$ over the range $\theta \leq 45^\circ$. Within each zenith-angle bin, the events are ordered by decreasing $S(300)$, and several fixed integrated intensities are selected. Under the assumption of isotropy and full detection efficiency, events corresponding to the same intensity in different angular bins are associated with the same primary energy. The attenuation curve is then obtained by fitting the evolution of $S(300)$ with zenith angle at fixed intensity.

For the SD-433 array, the attenuation of the shower size is parametrised as a second-order polynomial in the variable $x = \sin^2 \theta - \sin^2 30^\circ$, defining a reference shower size S_{30} corresponding to the signal the shower would have produced at a zenith angle of 30° , which is the median angle of an isotropic distribution up to 45° . Small residual dependencies of the attenuation curve on the shower size are accounted for by allowing the linear coefficient of the polynomial to vary logarithmically with S_{30} . This procedure yields a zenith-angle-independent energy estimator suitable for calibration.

In contrast to the SD-1500 and SD-750 arrays, it is not possible to directly calibrate the SD-433 energy scale using the fluorescence detector. The lower primary energies accessed by the SD-433 array, together with the limited geometrical overlap with the fluorescence telescopes, result in a very small number of hybrid events with sufficient quality for a reliable calibration. As a consequence, we adopt an indirect calibration approach.

The SD-433 energy scale is instead derived by correlating S_{30} with the energy assigned to the same air showers by the SD-750 array, using events reconstructed simultaneously by both detectors. Since the SD-750 array is itself calibrated against the fluorescence detector using high-quality hybrid events, the SD-433 energy scale effectively inherits the calorimetric energy scale of the fluorescence detector through this transfer. Only events with reconstructed energies above the full-efficiency threshold of the SD-750 array are used in the calibration, ensuring that trigger and reconstruction biases are negligible.

The calibration relation is described by a power-law function of the form

$$E_{\text{SD-433}}^{\text{rec}} = A \left(\frac{S_{30}}{30 \text{ VEM}} \right)^B, \quad (5.2)$$

where the parameters A and B are determined by maximising a likelihood function that incorporates the energy resolutions of both surface arrays, the steeply falling cosmic-ray spectrum, the effect of the energy threshold, and the intrinsic correlation between S_{30} and the SD-750 energy due to the shared sampling of the same shower front. This approach ensures a statistically consistent transfer of the energy scale while minimising model-dependent

assumptions. The current energy calibration and attenuation-function parametrisation are documented in section A.2.

Given the calibration strategy outlined above, a detailed characterisation of the SD-433 detector response must rely predominantly on air-shower simulations. While the absolute energy scale of the array is inherited from the fluorescence detector through the SD-750 calibration transfer, several key ingredients entering the response function, such as the trigger or reconstruction efficiency, the energy resolution, and potential reconstruction biases, cannot be determined with sufficient precision from data alone in the SD-433 energy range. This limitation is a direct consequence of the reduced hybrid statistics at low energies and the steeply falling cosmic-ray flux.

In this chapter, we therefore characterise both the efficiency ε and the spread function s using a comprehensive set of Monte Carlo simulations, including pure proton, helium, oxygen, iron, and a mixed-composition scenario. The use of multiple primary masses allows us to assess the robustness of the response characterisation against uncertainties in the mass composition, which are particularly relevant near the second knee. Wherever possible, the simulation-based results, most notably those related to the detection efficiency (and indirectly for the energy resolution), are validated against measured data to ensure their reliability.

The response-function modelling presented in this chapter constitutes a necessary intermediate step between the experimental observables measured by the SD-433 array and the physics interpretations derived from them. In particular, it provides the quantitative framework required for the reconstruction of the cosmic-ray energy spectrum, and as a consequence, for the background estimation of the elusive photon flux. Conversely, the implications of this response modelling for photon energy reconstruction, and the choice of the most appropriate energy scale for photon searches, are addressed in detail in Chapter 7.

5.1 Efficiency of the array

5.1.1 Simulation production

The Napoli-Praha simulation library of proton, helium, oxygen, and iron primaries was utilised to simulate the SD-433 at the Lyon Computing Centre IN2P3 [135]. The simulations are based on CORSIKA 7.7 showers and cover an energy range from 10^{16} to $10^{17.5}$ eV, divided into logarithmic energy bins spanning half a decade each, following an E_{MC}^{-1} distribution (i.e. flat distribution in $\lg E_{MC}$). A total of 5000 showers were generated for each combination of energy bin and particle, with zenith angles uniformly distributed in $\sin^2 \theta_{MC}$ for $\theta_{MC} < 65^\circ$. These showers were simulated using the EPOS-LHC and FLUKA-INFN models for high- and low-energy hadronic interactions, respectively. The transition energy between these two regimes was set at 80 GeV, and a thinning level of 10^{-6} was applied.

Event simulation and reconstruction were performed by the Monte Carlo Task using the Offline tag v4r0p2-pre3. The events were simulated in a surface array where the detectors

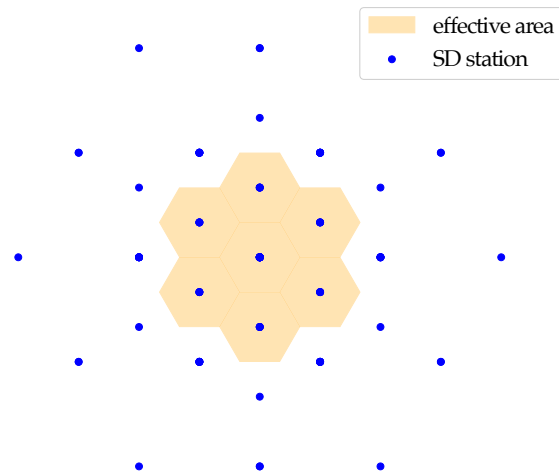


Figure 5.1: Geometric layout of the SD-433 array, displaying only the first crown of the SD-750. The effective area is indicated by the coloured shading.

were positioned at intervals of 433 m, and two outer crowns were formed by stations separated by 750 m, as shown in Fig. 5.1.¹ Each shower was thrown to the array ten times, and its core was randomly placed within the outer crown of the SD-433 array.² Stations that were triggered using the ToTd and MoPS algorithms were incorporated into the analysis. The simulation and reconstruction applications developed as part of this thesis have been incorporated into the *Offline* framework as Standard Applications, ensuring their long-term availability and usage within the Collaboration.

Only events satisfying the 6T5 condition in the SD-433 array were considered. This selection criterion is roughly equivalent to requesting reconstructed cores that fall within the effective area, which is shaded in Fig. 5.1. The possibility of events with Monte Carlo cores falling inside the effective area and being reconstructed outside it, and vice versa, was taken into account. For showers with zenith angles up to 45° , this migration is negligible.

5.1.2 Composition

The composition of cosmic rays plays a crucial role in efficiency calculations because the attenuation of air showers varies with the mass of the primary particle. The Fluorescence Detector (FD) directly measures $X_{\max}$, a proxy for mass composition, which can be integrated in SD efficiency studies to account for this systematic effect. For the main array, the efficiency can also be estimated using joint data with the SD-750. Neither of these approaches is feasible for the SD-433. For energies below 2×10^{17} eV, the low statistics prevent reliable fits of $X_{\max}$ templates, which could provide insights into the fractions of different particle masses.

In this scenario, we use the Global Spline Fit (GSF). The GSF is a data-driven model, based on measurements from many experiments, of the cosmic-ray flux and its composition from 10 to 10^{11} GeV [136]. The cosmic-ray flux is divided into four mass groups, which cover roughly equal ranges in logarithmic mass. The model relies on spline amplitudes and energy-scale offsets derived from a global least-squares fit. In this work, we adopt the GSF Mix as an assumption of mass composition for efficiency computations. The mass fractions of the GSF are shown in Fig. 5.2 in conjunction with the Auger mass fractions derived from template fitting of $X_{\max}$ data [109]. The energy regime targeted by this study is shaded in orange.

Four simulated datasets with proton, helium, oxygen, and iron primaries were produced. The GSF Mix was generated by sampling from these original four datasets in a manner that closely follows the fractions described by the composition model.

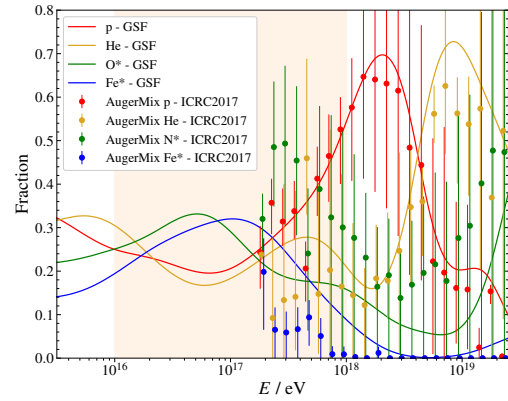


Figure 5.2: Mass fractions as a function of the primary energy according to the Global Spline Fit [136] (solid lines) and the Auger Mix [109] (markers). The Auger-Mix markers are centred at the mean normalised fractions coming from $X_{\max}$ template fits conducted with different hadronic models (EPOS-LHC, QGSJETII-04 and Sibyll 2.1). In the same fashion, the errors are set as the minimum and maximum fractions predicted by the different $X_{\max}$ template fits.

¹Because even small deviations from an ideal triangular layout can affect the detection efficiency, the simulated SD-433 array reproduces the actual station positions, including local elevation differences and minor geometric irregularities present in the real deployment.

²The complete production description is available at: [AugerWiki: NapoliLibrary](https://www.auger.wiki/NapoliLibrary)

5.1.3 Trigger efficiency model

The efficiency of the SD-433 array, ε , quantifies the probability p that a shower initiated by a primary of mass A and energy E , arriving with zenith angle θ , is reconstructed by the SD-433. It considers the LDF is fitted, and the primary energy can be estimated. Put differently, we assume that the efficiency depends on (a) the energy of the primary particle E , as the number of particles on the ground, their energy spectrum, and therefore, the signals produced in tanks are mainly driven by this quantity; (b) the zenith angle θ , because it defines the traversed path along the atmosphere, allowing for modelling the shower absorption and detector geometric effects to a great extent; and (c) the composition described by A , which is an important factor for the balance of the electromagnetic and muonic components of the showers.

The triggering of the array can be viewed as a random variable Y , with its outcome y representing the result of a Bernoulli trial: either the array is triggered ($y = 1$) or it remains inactive ($y = 0$). As a consequence of the definition for y , the efficiency can be expressed as:

$$\varepsilon = E[y|X], \quad (5.3)$$

where X represents the explanatory variables significantly influencing the expected value of y . Due to the binary nature of y , we model $y|X$ as a Bernoulli distribution with mean $E[y|X]$. While y can take values of 0 and 1, its expected value can lie in a continuum between these limits.

We model the efficiency ε using a Generalized Linear Model (GLM) [137] where the independent variables were $\lg E$, $\sin^2 \theta$, and $\ln A$, depending on the case. The dependent variable y was linked to the predictors with the probit link function Φ^{-1} . The model parameters were estimated by minimising the deviance, which is equivalent to maximising the likelihood function. More details of the GLM framework are available in section A.1.

5.1.4 Trigger efficiency with simulations

An initial intuition of the effect of the primary features on the array trigger efficiency is gained by looking at Fig. 5.3. The array efficiency increases with energy and decreases with zenith angle. In addition, a difference between proton- and iron-induced showers is observed. Past studies have found that lighter primaries up to 45° in zenith are more efficient than heavier ones in the SD-433 energy regime [138]. This was also the case for EPOS-LHC simulations used for the SD-750 array [139]. In this work, this behaviour was corroborated as shown in Fig. 5.3.

The definition of the array trigger includes the original station triggers, ToT and Th, as well as the newer ToTd and MoPS triggers. By construction, the latter two are more sensitive to the EM component of showers than to muons. As the relationship between the EM and muonic component varies with the primary mass, so will the trigger efficiency. Additionally, lighter primaries penetrate deeper into the atmosphere compared to primaries with heavier compositions. The combination of these effects causes the efficiency curves to change with composition. For vertical showers, proton showers reach full efficiency at lower energies than iron ones. For showers with increasing zenith angles, most of the EM component is absorbed by the atmosphere, and the muonic component tends to drive the efficiency.

The first and simplest model only explicitly considers the energy dependence of the array efficiency. Indirectly, the model also accounts for the zenith angle and the mass composition. The trigger probability can be thought of as the expected value of an efficiency

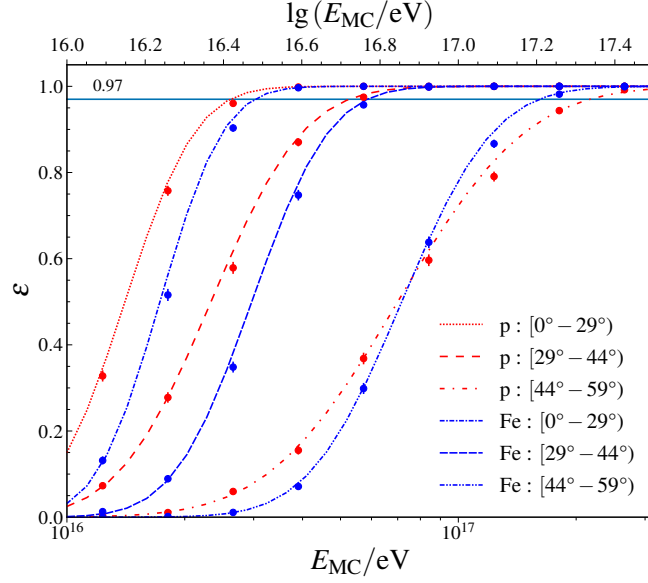


Figure 5.3: Binned estimates of ε for proton (red markers) and iron (blue markers) primaries. The lines with different styles correspond to unbinned fits, following the trigger model of Eq. (5.6).

function weighted by the zenith and mass composition populations. Mathematically, it can be expressed as:

$$\varepsilon(E) = \int_{\Delta \sin^2 \theta} \int_{\Delta \ln A} P(\sin^2 \theta, \ln A) \varepsilon(E, \sin^2 \theta, \ln A) d \sin^2 \theta d \ln A \quad (5.4)$$

where the efficiency function $\varepsilon(E, \sin^2 \theta, \ln A)$ is weighted by the joint probability distribution $P(\sin^2 \theta, \ln A)$, in the spirit of past unfolded estimations of efficiency [140]. The distribution can be factorised as

$$P(\sin^2 \theta, \ln A) = P(\sin^2 \theta)P(\ln A) = \sin \theta \cos \theta P(\ln A) \quad (5.5)$$

if the marginal distributions are considered independent and any joint effect between both variables is contained in ε . At least to first order, the assumption is not far-fetched, for an isotropic flux irrespective of the primary mass.

It is worth noting that analytical expressions for the joint and marginal distributions are not utilised, but these are enforced in the datasets by (a) the library itself, which follows a uniform zenith distribution,³ (b) in the case of a mixed flux, the dataset was created following the proportions of the GSF, where $P(\ln A)$ is implicitly a function of the energy.

We first choose a simple model with an intercept and a linear term for re-centred logarithmic energy as the linear part of ε . Therefore, we have

$$\Phi^{-1}(\varepsilon(E_{MC})) = a_0 + a_1 \lg \left(\frac{E_{MC}}{10^{17} \text{ eV}} \right), \quad (5.6)$$

where a_i are the parameters that are fitted by minimisation of the deviance computed from the model prediction $\hat{y}_i$ and labels y_i .

The results of the fit are shown in Section 5.1.4, and the parameters and associated errors are available in Table 5.1. To cross-check the unbinned model estimation with a simpler proxy for efficiency, the χ^2 between the fitted GSF model (solid black line in Section 5.1.4)

³In this study, efficiency calculations are limited to showers with a zenith angle up to 45° , as is customary for the SD-433. The quality of the standard geometry reconstruction degrades for higher zenith angles.

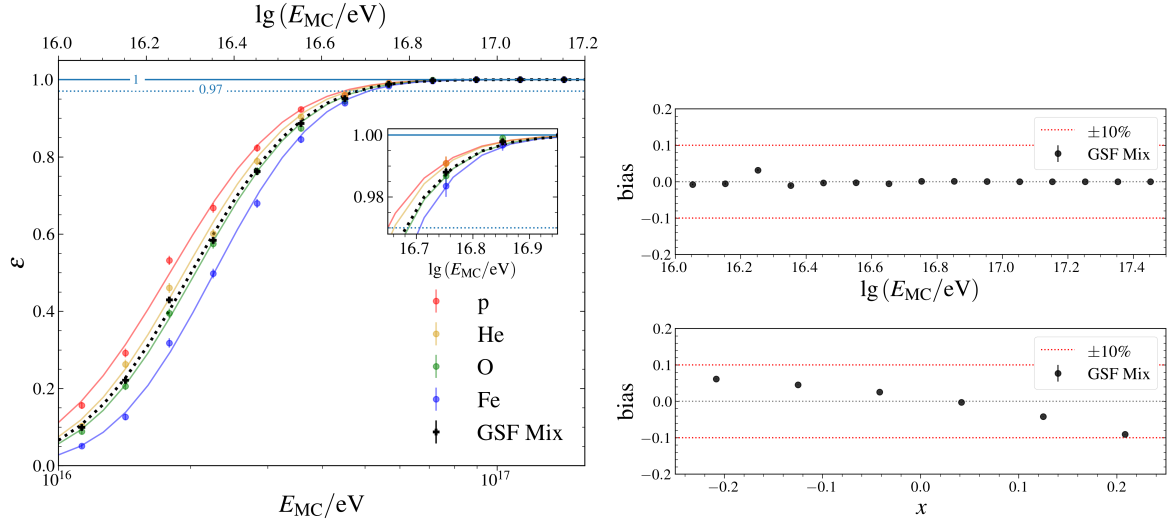


Figure 5.4: Left: $\epsilon(\lg(E_{MC}))$ for different composition scenarios. Markers correspond to binned estimates of ϵ from the fraction of triggered events, and the lines correspond to the unbinned fits from Eq. (5.6). The inset zooms into the high-efficiency region. The GSF mix is depicted in black. Right: Upper and lower panels present the bias for the GSF mix as a function of $\lg(E_{MC})$ and x , accompanied by the $\pm 10\%$ band in red dotted lines.

Table 5.1: Parameters for ϵ models following Eq. (5.6).

	a_0	a_1
p	3.564 ± 0.027	4.781 ± 0.038
He	3.587 ± 0.027	5.025 ± 0.038
O	3.483 ± 0.025	5.064 ± 0.037
Fe	3.469 ± 0.025	5.380 ± 0.038
GSF	3.479 ± 0.013	4.990 ± 0.018

and efficiency binned estimations (black markers in Section 5.1.4) was computed. With a value of ~ 0.003 and assuming 13 degrees of freedom (number of bins in the fitted energy range minus the number of parameters), the binned estimations are in good agreement with the fitted model. After this initial test, the biases from binned estimations are used as a goodness-of-fit measure.

The efficiency of lighter primaries exceeds that of heavier ones in the studied energy region. The efficiency of the GSF mix, as expected, falls below the pure proton efficiency curve. Near the full efficiency region, the curves flatten such that the transition between an efficiency of 0.97 and 1.0 extends up to three energy bins in the logarithmic scale⁴. The black markers in the upper panel of Section 5.1.4 represent the bias and show no apparent trend with the energy. This is not the case for the zenith angle, which exhibits a decreasing trend. Thus, the inclusion of the zenith angle as a predictor for the efficiency of the array is pursued.

⁴A typical uncertainty of 3–4% is assigned to the exposure of the array over a given data-taking period, which relaxes the strict requirement of reaching exactly full efficiency, since the dominant limitation arises from the exposure determination rather than from $\epsilon(E, \theta)$ itself.

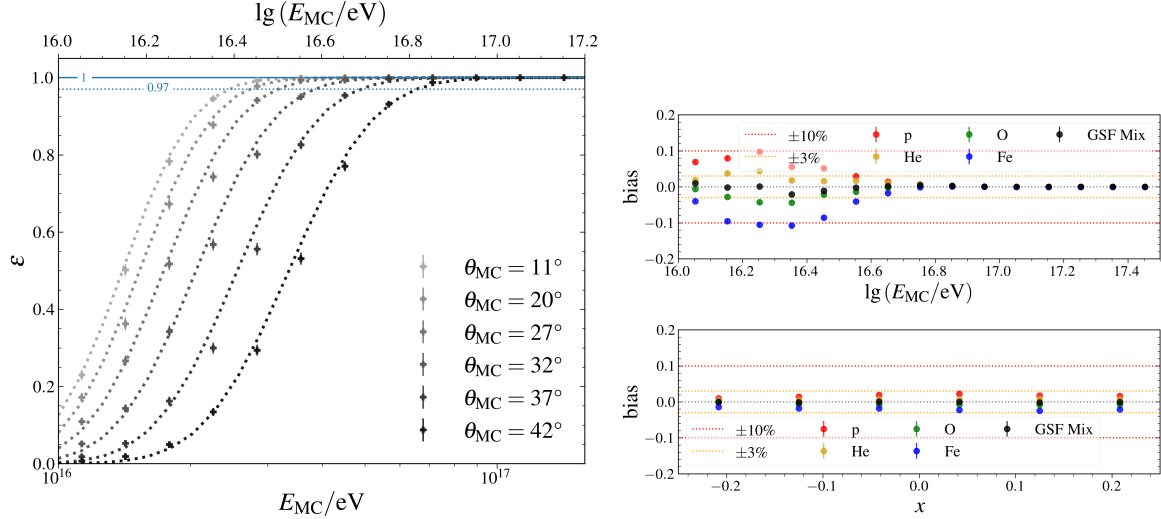


Figure 5.5: *Left:* $\varepsilon(\lg(E_{\text{MC}}), \theta)$ as a function of $\lg E_{\text{MC}}$ for different zenith bins. The labels represent the centre of the zenith-angle bins. Markers correspond to the binned estimation of ε from trigger to non-triggered event ratios for the GSF mix, and the lines correspond to the estimations from the unbinned fit. *Right:* Upper and lower panels present the bias for the GSF mix and pure fluxes as a function of $\lg E_{\text{MC}}$ and x , respectively. The $\pm 10\%$ band is represented by red dotted lines, while the $\pm 3\%$ band uses yellow dotted lines.

Table 5.2: Parameters for ε models following Eq. (5.8).

	a_0	a_1	a_2	a_3	a_4
p	5.284 ± 0.049	6.794 ± 0.063	-8.695 ± 0.307	-8.475 ± 0.532	-3.769 ± 0.406
He	5.488 ± 0.049	7.230 ± 0.064	-9.288 ± 0.321	-6.778 ± 0.554	-4.159 ± 0.431
O	5.387 ± 0.046	7.615 ± 0.064	-8.176 ± 0.304	-8.006 ± 0.571	-1.609 ± 0.436
Fe	5.383 ± 0.046	8.013 ± 0.066	-8.960 ± 0.308	-7.450 ± 0.610	-2.416 ± 0.455
GSF	5.193 ± 0.023	7.157 ± 0.031	-8.698 ± 0.148	-7.140 ± 0.274	-3.293 ± 0.205

By explicitly modelling the zenith dependence of the efficiency, we avoid averaging over the isotropic distribution. The mass-composition evolution with energy is again taken into account by the proportions of the GSF. We therefore propose a linear model that, in addition to the energy term, includes a term with $x = \sin^2 \theta - \sin^2 30^\circ$, a term with x^2 , and also an interaction term between E_{MC} and x that accounts for joint effects arising from these variables. Therefore, this time the efficiency curve represents:

$$\varepsilon(E, \sin^2 \theta) = \int_{\Delta \ln A} P(\ln A) \varepsilon(E, \sin^2 \theta, \ln A) d \ln A \quad (5.7)$$

and we choose as the linear part of the model

$$\mathbf{X}\beta = a_0 + a_1 \ell + a_2 x + a_3 x^2 + a_4 x\ell, \quad (5.8)$$

where $\ell = \lg(E_{\text{MC}}/10^{17} \text{ eV})$. Simpler models without the x^2 term, or the interaction term, proved to be insufficient to capture the variation of the efficiency. The final parametrisation was decided based on goodness-of-fit criteria, and the parameters are presented in Table 5.2.

The fit for the GSF mix, evaluated for different zenith bin centres, is shown in Section 5.1.4. The model captures the the zenith-angle dependence of the efficiency of the simulated SD-433 array. Inspecting the bias as a function of energy and zenith angle in Section 5.1.4, no

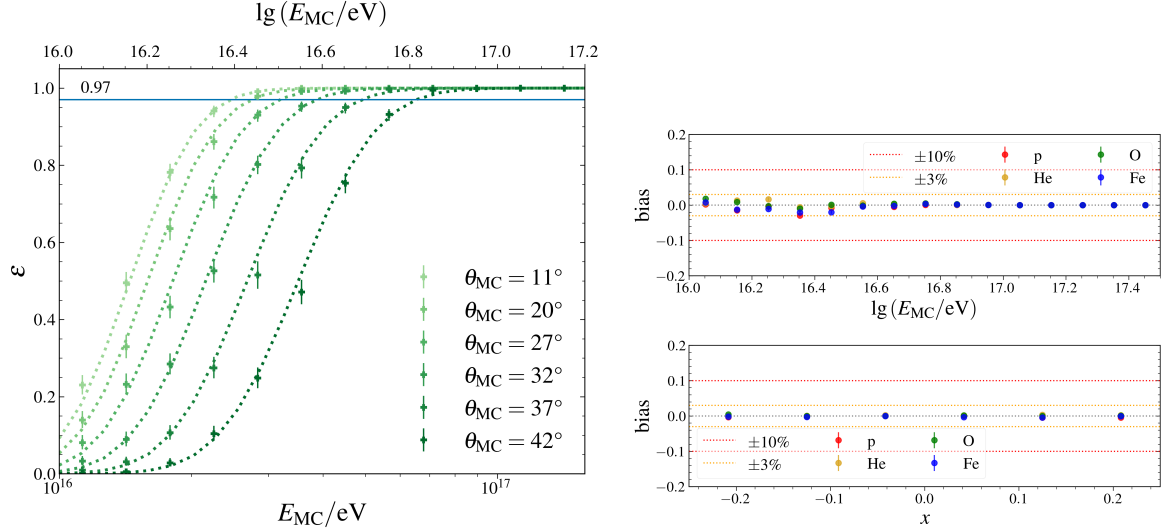


Figure 5.6: Left: $\varepsilon(\lg(E_{MC}), \theta, \ln(A) \approx 2.77)$ for an oxygen primary as a function of $\lg(E_{MC})$ for different zenith bins. The labels represent the centre of the zenith-angle bins. Markers correspond to binned estimations of ε from trigger to non-triggered event ratios, and the lines correspond to the estimations from the unbinned fit. Right: Upper and lower panels present the bias for the GSF mix and pure fluxes as a function of $\lg(E_{MC})$ and x , respectively. The $\pm 10\%$ band is represented by red dotted lines, while the $\pm 3\%$ band uses yellow dotted lines.

Table 5.3: Parameters for the ε model of Eq. (5.9).

a_0	a_1	a_2	a_3
5.250 ± 0.036	6.740 ± 0.049	-8.151 ± 0.169	7.567 ± 0.281
a_4	a_5	a_6	a_7
-3.023 ± 0.214	0.047 ± 0.013	0.319 ± 0.019	-0.304 ± 0.026

obvious trend is apparent for the mix. The mix bias is also well contained within the 3% band. However, the composition bias becomes apparent when the different primaries are considered separately. This finding leads the way to model an efficiency function depending on the energy, the zenith angle, and the composition.

At this stage, the focus is on a model such as $\varepsilon(E, \sin^2 \theta, \ln A)$. The intention is to estimate the probability parameter of a Bernoulli distribution for a particle with mass A , energy E_{MC} , which reaches the atmosphere with a zenith angle θ . This efficiency function does not take into account any weighting due to the underlying probability distributions.

The mass composition is modelled by a linear term with $\ln A$, and interaction terms between this variable and both the energy and zenith angle are also considered. The final parametrisation for the linear part of the model is

$$\mathbf{X}\beta = a_0 + a_1 \ell + a_2 x + a_3 x^2 + a_4 x \ell + a_5 \ln A + a_6 \ell \ln A + a_7 x \ln A, \quad (5.9)$$

and the fitted parameter values are provided in Table 5.3.

Because CNO is expected to be the most abundant mass group around the second knee, the results from the fit for oxygen are shown in Section 5.1.4 for different zenith-angle bins.

Although not shown, the other mass groups are also in good agreement with the model. This can be seen from Section 5.1.4, where the biases for different primaries are presented. For all primaries, the bias is contained within the 3% band. Only a subtle trend with the logarithm of energy can be glimpsed around $10^{16.3}$ eV, which is regarded as negligible. With this parametrisation, a closed expression for the SD-433 efficiency as a function of useful predictors is obtained.⁵

5.1.5 Efficiency in the SD-433 energy scale

Air-shower simulations provide events with a known primary energy E_{MC} , but they do not reproduce the measured detector response perfectly. In particular, the long-standing discrepancy in muon production between simulations and data implies that, at fixed energy, both the apparent shower size $S(300)$ and its attenuation with zenith angle can differ between Monte Carlo and observations [141]. Since the SD-433 energy estimator is constructed from $S(300)$ via a CIC correction and an energy calibration, these differences translate into a mismatch between the Monte Carlo energy scale and the energy scale used for data. Consequently, response quantities inferred from simulations (including the efficiency) must be expressed in the SD-433 energy scale before being used in data analyses.

We correct the mean data–MC discrepancies by reassigning each simulated event to an equivalent true energy $E_{\text{SD-433}}$. Conceptually, the reassignment follows the same chain that defines the SD-433 energy assignment in data,

$$E_{\text{MC}} \rightarrow S_{30}^{\text{MC}} \rightarrow S(300)^{\text{MC}} \rightarrow S_{30}^{\text{data}} \rightarrow E_{\text{SD-433}}, \quad (5.10)$$

and it depends on both zenith angle and energy through the attenuation corrections and calibration relations.

The attenuation-corrected shower size is defined as

$$S_{30} = \frac{S(300)}{f_{\text{att}}(x)}, \quad x = \sin^2 \theta - \sin^2 30^\circ, \quad (5.11)$$

where $f_{\text{att}}(x)$ is normalised at 30° . For the SD-433 array, f_{att} is modelled as a second-order polynomial,

$$f_{\text{att}}(x) = 1 + a x + b x^2. \quad (5.12)$$

In this work, f_{att} is determined both from data and from several simulation sets (pure primaries and the GSF mix). Figure 5.7 compares the corresponding CIC curves. Systematic differences are observed among primaries and with respect to data, increasing with zenith angle; these differences directly enter the energy reassignment.

The Monte Carlo calibration relating S_{30} to E_{MC} is parametrised as a power law,

$$E_{\text{MC}} = A_{\text{MC}} \left(\frac{S_{30}}{30 \text{ VEM}} \right)^{B_{\text{MC}}}. \quad (5.13)$$

Figure 5.8 shows the calibration obtained for the simulated primaries and for the GSF mix. The calibration of the mixed sample closely follows those of the pure primaries, indicating that realistic composition mixtures do not destabilise the calibration procedure.

⁵The expression for ϵ can be employed to assess scenarios that do not assume isotropic cosmic-ray arrival directions, or to explore the efficiency under different mass-composition models. To achieve this, the expression for ϵ should be integrated with appropriate weighting functions.

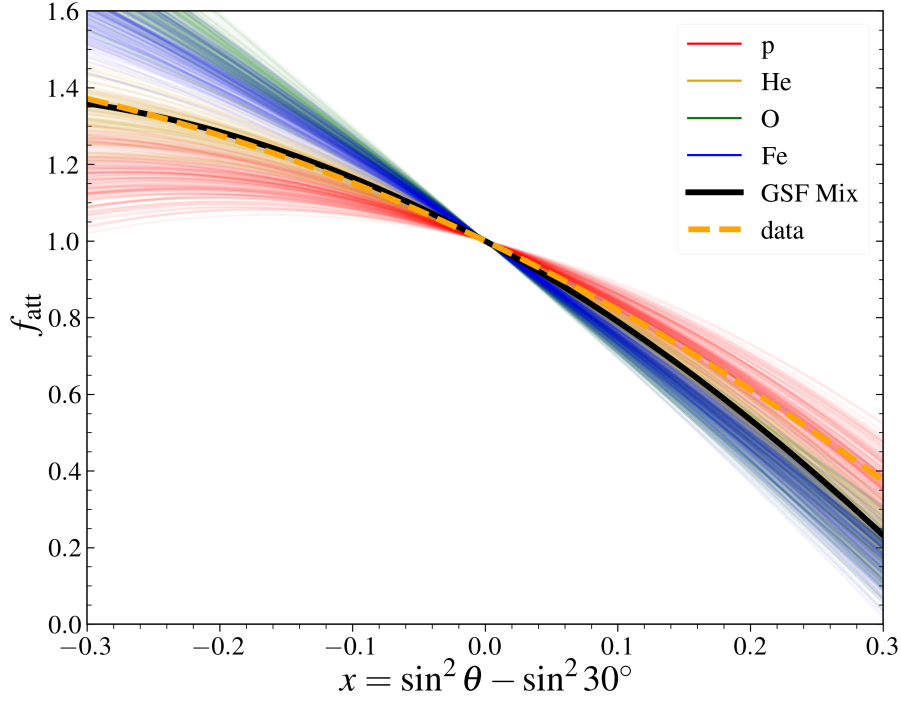


Figure 5.7: Monte Carlo CIC attenuation functions $f_{\text{att}}(x)$ for proton, helium, oxygen, iron, and the GSF mix, compared to the CIC obtained from data. The variable $x = \sin^2 \theta - \sin^2 30^\circ$ defines the reference angle of 30° . The different curves for each primary arise from bootstrapping and reflect the uncertainty in the attenuation function.

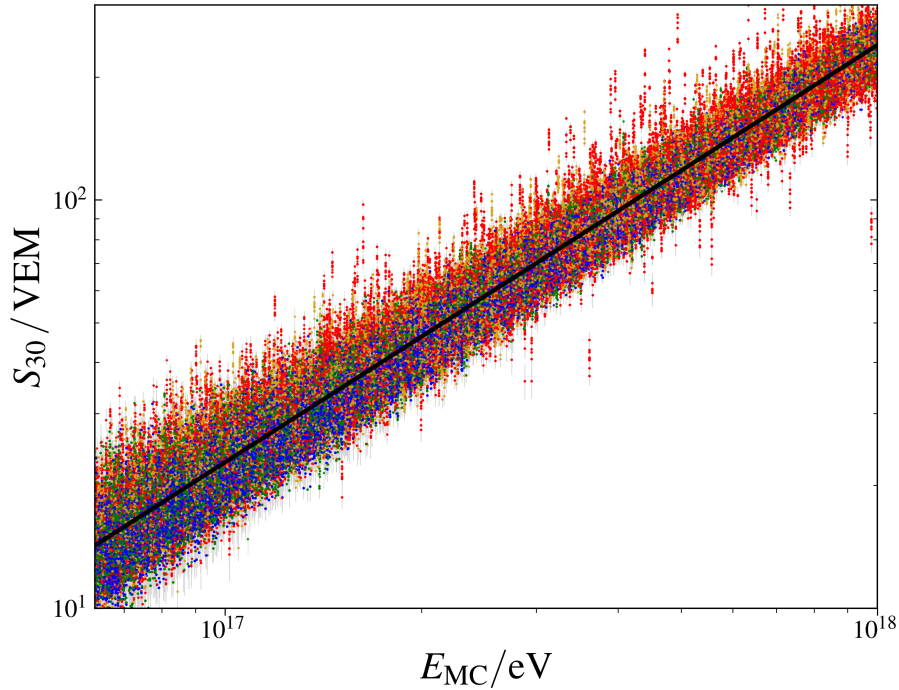


Figure 5.8: Energy calibration in simulations: correlation between S_{30} and E_{MC} , with the fitted calibration curve from Eq. (5.13). The colours of the markers indicate the mass composition. A slightly heavier composition is expected at lower energies, followed by a gradual transition towards lighter primaries (iron/blue to proton/red).

Combining the attenuation correction and calibration relations for simulations and for data yields an explicit reassignment $E_{\text{SD-433}}(E_{\text{MC}}, \theta)$. Following the derivation in [139], the mapping can be written as

$$E_{\text{SD-433}} = A \left[A_{\text{MC}} \left(\frac{E_{\text{MC}}}{100 \text{ PeV}} \right)^{B_{\text{MC}}} \left(\frac{f_{\text{att}}^{\text{MC}}(x)}{f_{\text{att}}(x)} \right) \right]^B, \quad (5.14)$$

where $f_{\text{att}}^{\text{MC}}$ and $(A_{\text{MC}}, B_{\text{MC}})$ are the CIC and calibration parameters obtained in simulations, while f_{att} and (A, B) correspond to the data-driven CIC and SD-433 calibration (see section A.2). On average, the correction is at the level of $\sim 12\%$ for $\theta \lesssim 30^\circ$, and increases for more inclined showers due to the growing impact of attenuation differences. A mild energy dependence is also expected due to differences in calibration slopes; both effects and the employed parametrisations are documented in section A.2.

Because the reassignment in Eq. (5.14) corrects the *mean* behaviour extracted from data, it does not explicitly include a dependence on $\ln A$. Therefore, it is most appropriate for a mixed-composition scenario such as the GSF flux. Nevertheless, corrected-energy results for pure primaries are still shown in the figures as reference bounds around the mixed case.

After applying Eq. (5.14) event-by-event, the efficiency is recomputed using $E_{\text{SD-433}}$ as the energy coordinate. In practice, this amounts to refitting the same functional form used for $\varepsilon(E_{\text{MC}})$ (cf. Eq. (5.6)), but now as a function of the corrected energy scale. The resulting energy-dependent efficiency for events with $\theta \leq 45^\circ$ is shown in Fig. 5.9. The fit yields $a_0 = 4.213 \pm 0.032$ and $a_1 = 5.496 \pm 0.044$ for the GSF mix, which are larger than the corresponding parameters in the Monte Carlo scale (Table 5.1). Since the probit model is monotone, this implies a faster rise of $\varepsilon(E)$ with energy in the SD-433 scale. The curve is also shifted towards lower energies, consistent with a scenario where the additional muon content observed in data effectively lowers the energy threshold relative to simulations.

The inset of Fig. 5.9 highlights that, even after the energy-scale correction, the transition from $\varepsilon \simeq 0.97$ to $\varepsilon \simeq 1$ extends over several energy bins. This behaviour is consistent with the shallow approach to full efficiency expected for sparse arrays near threshold.

Finally, while the efficiency is modelled as $\varepsilon(E, \theta)$ in the general framework, the present analysis adopts the energy-only dependence in the SD-433 scale as the standard input for studies within the Collaboration. A zenith-dependent parametrisation of $\varepsilon(E, \theta)$ is provided in section A.3 for completeness, but is not further discussed here.

5.1.6 Validation with data

We conducted additional checks with data from the ICRC23 SD-433 production [142]. To validate the simulation-based corrected efficiency, we perform a self-consistency check. Isotropy is employed as the full-efficiency criterion ($\varepsilon = 1$), probing the threshold via statistical tests with data. The CIC procedure and energy calibration have predetermined the attenuation and energy scale, so the tests only confirm the energy threshold where full efficiency is achieved.

Assuming isotropic event arrival, the zenith angle θ follows a uniform distribution in $\sin^2 \theta$, where $\theta \in (0^\circ, \theta_{\text{max}} = 45^\circ)$. The cumulative distribution of events forms a line from $(0, 0)$ to $(\sin^2 \theta_{\text{max}}, 1)$. To assess deviation, we compare this line with a step function derived from observed values, each step corresponding to an event with a height increase of 1 over the sample size. The area between the ideal and observed cumulative distributions, weighted by closeness to the tails, is quantified by the Anderson-Darling statistic (AD). We employ this discrepancy measure here for hypothesis testing, following its previous use in attenuation-function fits. For an illustrative computation example, refer to Ref. [143].

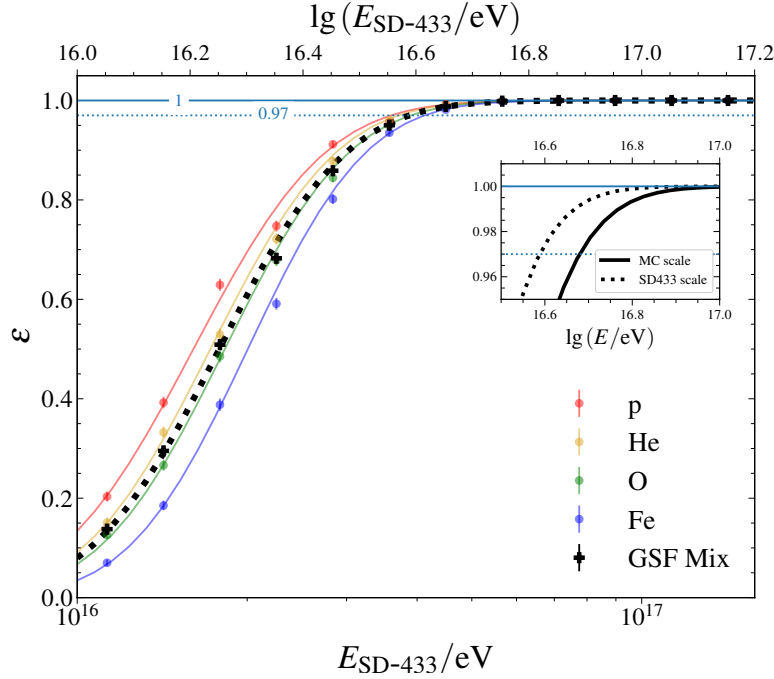


Figure 5.9: ε as a function of $\lg E_{\text{SD-433}}$. Markers correspond to binned estimates of $\varepsilon(E)$ from triggered to non-triggered event ratios, and the lines correspond to the unbinned fits. The inset zooms into the high-efficiency region, exhibiting the shift of the efficiency curves after the change in energy scales. The usual colour code adopted in the Collaboration for proton, helium, oxygen, and iron is employed. The GSF mix is depicted in black.

Using a uniform random number generator, we construct the null hypothesis distribution for the AD statistic, from instances drawn from a uniform distribution. In the limit of large samples, the H_0 distribution is independent of the sample size [144]. We employed samples of 4×10^5 instances (around the size of events above $10^{16.7}$ eV in data) from the random number generator and constructed the distribution by computing 10^5 statistics. Computing the statistic using $\sin^2 \theta_i$ from events above a determined energy threshold reveals proximity to isotropic arrival and, by extension, full efficiency. To probe for $\varepsilon = 1$, the chosen thresholds were $10^{16.7}$, $10^{16.8}$, and $10^{16.9}$ eV. Illustrated in Fig. 5.10, the AD statistic for each data sample is contrasted with the isotropic H_0 distribution.

p -values were computed as the ratio of the number of simulated AD values from the null distribution above the observed AD value to the total number of simulated samples. Requiring a p -value of 0.01 ($\text{AD} = 3.8$) to reject the H_0 , only the events above $10^{16.9}$ eV can be considered to arrive isotropically. However, requesting a p -value of 0.005 ($\text{AD} = 4.5$) already is compatible with isotropic arrival directions above $10^{16.8}$ eV. Therefore, it is safe to assume the transition to full efficiency is completed between $10^{16.8}$ and $10^{16.9}$ eV.

The motivation for this study is ultimately to both establish the full efficiency threshold for the SD-433 array and provide an estimate of the efficiency as a function of energy, considering events with θ up to 45° . These two ingredients aid in the computation of the cosmic-ray spectrum, as they are necessary inputs for the forward-folding procedure that corrects the observed flux by detector effects. As simplicity is sometimes the best approach, the last sanity check involves inspecting the cumulative number of events for independent zenith bands. The curves of $N_{\text{events}} E_{\text{SD-433}}^{\gamma-1}$, which are quantities proportional to the raw flux, are plotted for the measured indices before (γ_0) and after (γ_1) the second knee [142]. The zenith-angle limits are chosen to ensure a uniform distribution in $\sin^2 \theta$, and each curve is normalised to account for different exposures. The curves are shown in Section 5.1.6. The left

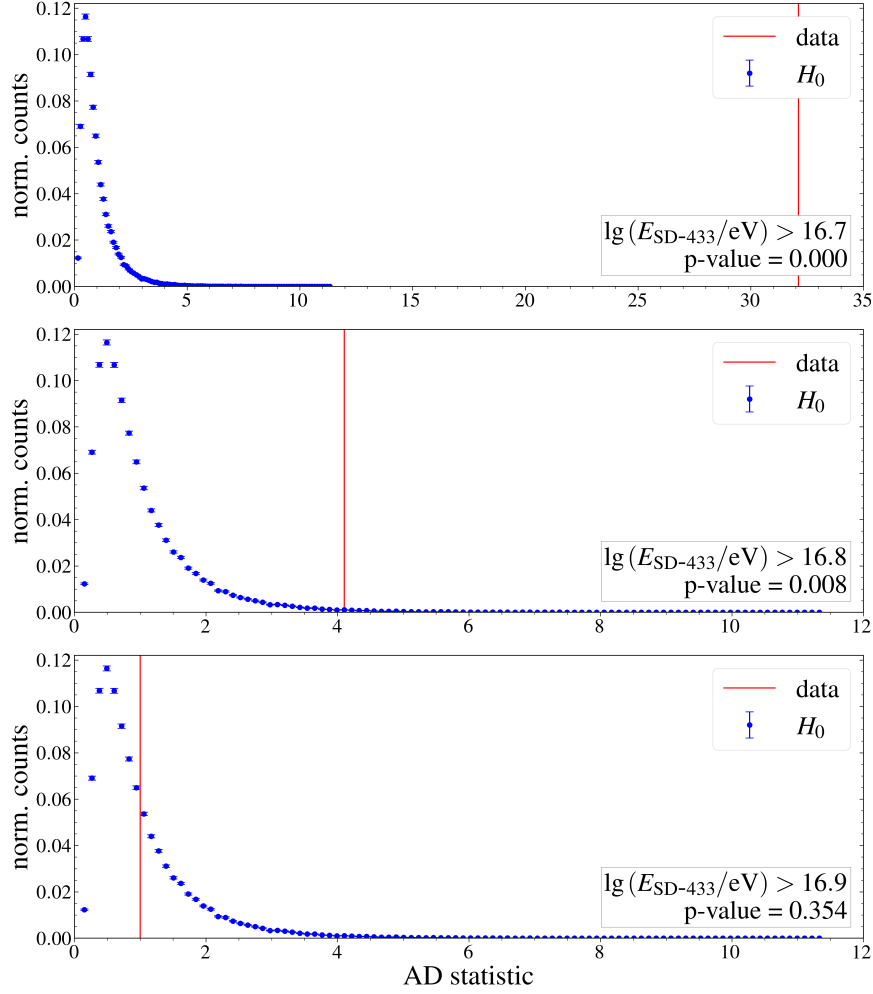


Figure 5.10: Hypothesis tests for isotropic arrival of events in the SD-433 array. From top to bottom, the tests are associated to thresholds at $\lg(E_{\text{SD-433}}/\text{eV})$ equal to 16.7, 16.8, and 16.9, respectively. The simulated null hypothesis is shown on each panel with blue markers and error bars, and the AD statistic found for the data sample is portrayed by a vertical red line. p -values are also shared in the lower right corner of each panel. Note that $E_{\text{SD-433}}$ refers to the energy estimator, as the true energy is unknown.

sides of both figures show inefficiency effects, as the curves corresponding to higher zenith angles slowly but systematically separate from the bulk of curves. On the right sides of both plots, fluctuations arising from low statistics are amplified by the factor of $E_{\text{SD-433}}^{\gamma-1}$. This is checked again in Section 5.1.6, where the histogram of cumulative events does not show strong fluctuations. In each of the figures, two thirds of the cumulative curves are mainly contained by the 1σ envelope with respect to the middle zenith-angle bin, without evident ordering or trend.

Overall, the curves suggest that setting the energy threshold to $\lg(E/\text{eV}) = 16.8$ largely avoids inefficiency effects. Extending the energy range of the array to lower energies, considering events with zenith angles up to 45° , reduces the flux due to the apparent loss of full efficiency. As a final check, the CIC correction is examined for different shower sizes in the section A.4, where it is shown that the attenuation correction is consistent with the suggested energy threshold.

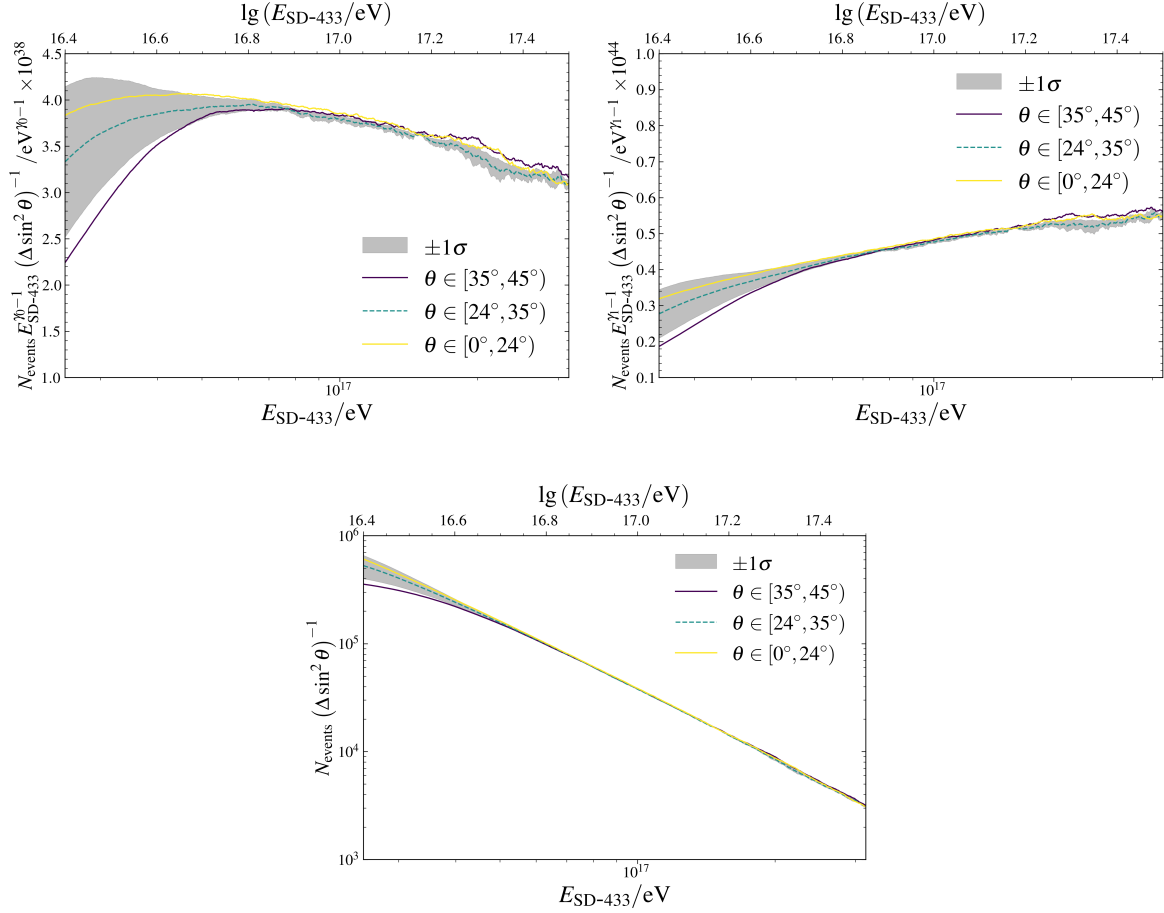


Figure 5.11: *Top left:* Cumulative number of events normalised by the corresponding width of the exclusive $\sin^2 \theta$ bin. A factor of $E_{\text{SD-433}}^{\gamma_0-1}$ makes the curves proportional to the flux before the second knee. *Top right:* Cumulative number of events normalised by the corresponding width of the exclusive $\sin^2 \theta$ bin. A factor of $E_{\text{SD-433}}^{\gamma_1-1}$ makes the curves proportional to the flux after the second knee. *Bottom:* Cumulative number of events normalised by the corresponding width of the exclusive $\sin^2 \theta$ bin. For all figures, the 1σ envelope shows the standard deviation from the middle zenith-angle bin (dashed line). Note that $E_{\text{SD-433}}$ refers to the energy estimator, as the true energy is unknown.

5.2 Energy response

We complete the characterisation of the SD-433 energy response in this section by studying the energy bias and energy resolution, which together define the spread function $s(E_{\text{SD-433}}^{\text{rec}} | E_{\text{SD-433}})$ introduced in Eq. (5.1). Because simulations and data differ in the mean shower-size response and in the zenith-angle attenuation, the true Monte Carlo energies of simulated events are mapped onto the SD-433 data energy scale.

After marginalising over zenith angle, the SD-433 energy-response spread function is modelled as a Gaussian in reconstructed energy,

$$s(E_{\text{SD-433}}^{\text{rec}} | E_{\text{SD-433}}) = \frac{1}{\sqrt{2\pi} \sigma_{\text{SD}}(E_{\text{SD-433}})} \exp \left[-\frac{(E_{\text{SD-433}}^{\text{rec}} - E_{\text{SD-433}} (1 + b_{\text{SD}}(E_{\text{SD-433}})))^2}{2 \sigma_{\text{SD}}^2(E_{\text{SD-433}})} \right], \quad (5.15)$$

where $b_{\text{SD}}(E_{\text{SD-433}})$ and $\sigma_{\text{SD}}(E_{\text{SD-433}})$ are the (energy-dependent) bias and resolution parameters of the zenith-marginalised response.

The marginalisation is performed assuming an isotropic flux observed by a horizontal detector, such that the zenith-angle distribution is weighted by the geometrical acceptance (projected area) up to $\theta_{\text{max}} = 45^\circ$. In this case,

$$s(E_{\text{SD-433}}^{\text{rec}} | E_{\text{SD-433}}) = \int_0^{\theta_{\text{max}}} s(E_{\text{SD-433}}^{\text{rec}} | E_{\text{SD-433}}; \theta) p(\theta) d\theta, \quad p(\theta) = \frac{2 \sin \theta \cos \theta}{1 - \cos^2 \theta_{\text{max}}} \text{ for } \theta \in [0, \theta_{\text{max}}], \quad (5.16)$$

which corresponds to a distribution uniform in solid angle, weighted by the projected area of the array.

5.2.1 Energy bias

The energy bias quantifies the mean systematic offset between reconstructed and true energies. In this chapter, it is defined as

$$b(E_{\text{SD-433}}) \equiv \mathbb{E} \left[\frac{E^{\text{rec}}}{E_{\text{MC}}} - 1 \mid E_{\text{SD-433}} \right], \quad (5.17)$$

where the expectation value is computed over events within a given true-energy interval expressed in the SD-433 data scale. We emphasise that the bias is evaluated using the reconstructed energy obtained from the simulation-based calibration, E^{rec} , and the true Monte Carlo energy, E_{MC} . In other words, the y-value is estimated in the MC scale, while only the x-scale is transformed to the data scale. This is necessary because the non-linear transformation of Eq. (5.14) does not conserve the relationship between reconstructed and true quantities.

The bias discussed here arises from detector and reconstruction effects and is most pronounced below the efficiency threshold, where the array is not yet fully efficient. In this regime, upward fluctuations are preferentially selected, leading to a positive bias. Once the full efficiency is reached, this selection effect becomes negligible, and the residual bias is expected to be consistent with zero for a properly calibrated estimator.

To describe the energy dependence of the bias in a compact form, we adopt the parametrisation

$$b(E_{\text{SD-433}}) = b_0 \exp \left[-b_1 \left(\lg(E_{\text{SD-433}}/\text{eV}) - 16 \right) \right] + b_2, \quad (5.18)$$

where b_0 , b_1 , and b_2 are fit parameters. The shift by 16 is introduced for numerical convenience, as it centres the argument in the energy range of interest.

We obtain the parameters from an unbinned maximum-likelihood fit assuming normally distributed residuals in

$$\delta \equiv \frac{E^{\text{rec}} - E_{\text{MC}}}{E_{\text{MC}}}, \quad (5.19)$$

as a function of $E_{\text{SD-433}}$. Using the functional form of Eq. (5.18), the best-fit values are

$$b_0 = 0.30 \pm 0.01, \quad b_1 = 10.0 \pm 0.3, \quad b_2 = 0.005 \pm 0.001, \quad (5.20)$$

where the quoted uncertainties correspond to the fit errors adopted for the likelihood evaluation.

Figure 5.12 presents the bias estimated from simulations after mapping to the SD-433 energy scale, together with the fit of Eq. (5.18). At energies relevant for most analyses, the bias remains small, while at lower energies it increases rapidly, as expected from the selection bias induced by the efficiency loss.

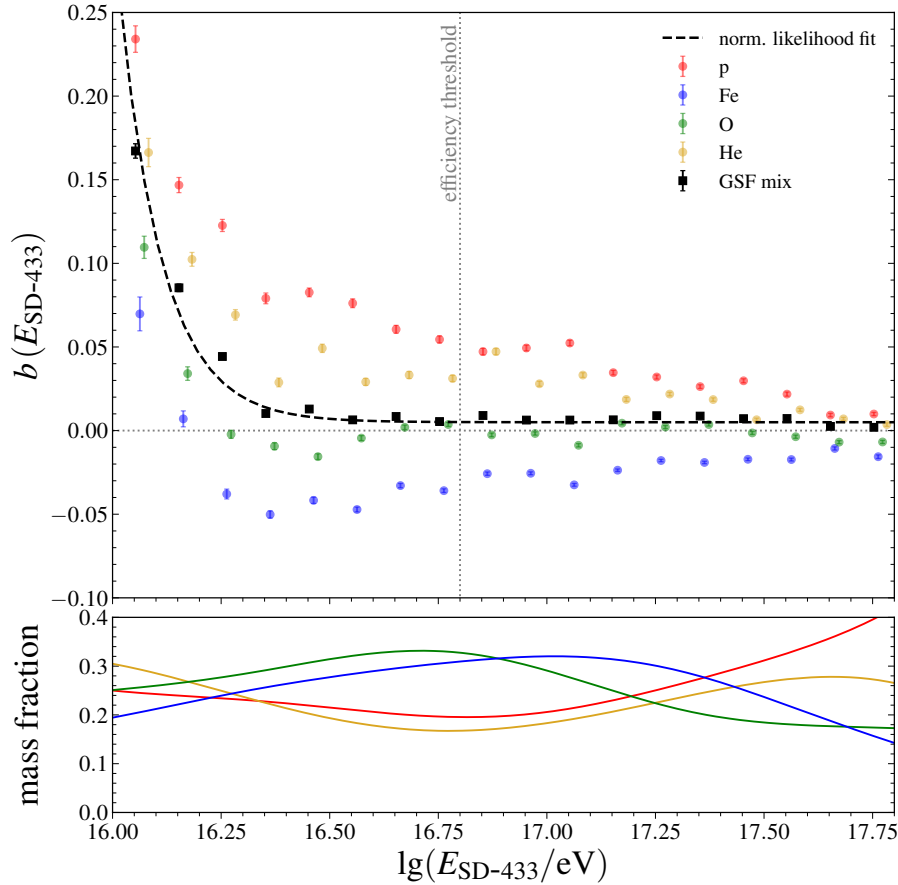


Figure 5.12: Energy bias as a function of reconstructed energy expressed in the SD-433 data scale, estimated from simulations. The dashed curve shows the best-fit parametrisation of Eq. (5.18). The vertical line marks the full-efficiency threshold. Coloured markers indicate the intrinsic composition dependence of the calibration: at fixed E_{MC} , different primaries (and their mixtures) exhibit systematic offsets in $E^{\text{rec}}/E_{\text{MC}} - 1$, reflecting the unavoidable mass dependence of the energy estimator. The lower panel shows the corresponding energy evolution of the GSF mass fractions used in the mixed-composition scenario.

The figure also highlights the intrinsic *composition bias* that arises when a single calibration is derived for a mixed set of primaries. At fixed true energy, showers initiated by lighter nuclei tend to produce a different ground signal (and thus a different reconstructed energy) than heavier primaries, so that the same estimator cannot be simultaneously unbiased for all masses. As a consequence, the reconstructed energy for light primaries can be systematically overestimated, while it can be underestimated for heavy primaries. This effect is unavoidable and is an expected feature of any calibration based on a composition mixture: by construction, the estimator is optimised to be (approximately) unbiased *on average* for the assumed mixture, rather than for each primary species individually. Above the full-efficiency threshold, the species-to-species separation visible in the bias curves provides a direct measure of this intrinsic spread in the energy assignment at fixed energy (or, equivalently, at fixed shower size).

A realistic evolution of the mass composition across the second knee does not introduce a significant additional bias in the energy assignment. This is illustrated in Fig. 5.12 by the bias estimated for the GSF mixture, which remains at the percent level across the energy range of interest. The corresponding mass-fraction evolution shown in the lower panel indicates that the calibration strategy can accommodate gradual changes in composition: despite

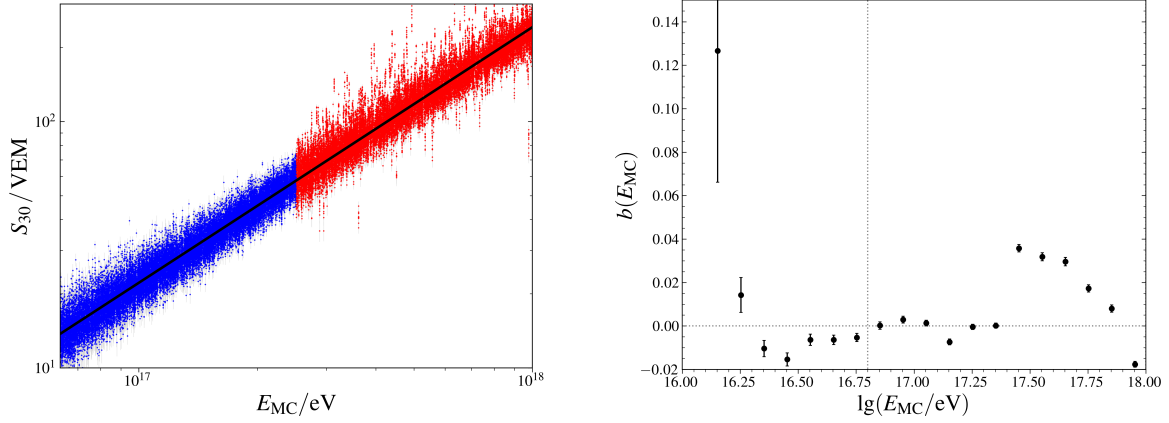


Figure 5.13: Stress test of composition-driven calibration artefacts using an extreme (unphysical) hard-break scenario. *Left:* Toy calibration sample in which the composition changes abruptly with energy: blue points correspond to iron primaries and red points to proton primaries, shown in the S_{30} – E_{MC} plane (black line: resulting calibration curve). *Right:* Induced energy bias $b(E_{MC})$ obtained when applying this calibration, showing a localised deviation around the break energy (vertical dotted line), while remaining close to zero away from the transition.

the continuous shift in the relative abundances of primaries with energy, the reconstructed energy scale remains stable, and the bias stays small once the array operates above full efficiency.

To further probe the robustness of the calibration against more extreme (and deliberately conservative) scenarios, we perform dedicated stress tests in which the composition is forced to change abruptly at a fixed energy. Only in the presence of such extremely hard (and therefore unphysical) breaks does a localised bias appear, reaching values of order $\sim 4\%$ around the break energy. This effect is shown in Fig. 5.13 and serves as a very conservative upper bound on potential composition-driven calibration artefacts.

5.2.2 Energy resolution

The energy resolution is quantified by the relative dispersion of reconstructed energies around the true energy. We define the SD-433 fractional resolution, reported as a function of true energy in the SD-433 scale, as

$$R(E_{SD-433}) \equiv \mathbb{E} \left[\frac{\sigma(E^{\text{rec}})}{E_{MC}} \mid E_{SD-433} \right], \quad (5.21)$$

where $\sigma(E^{\text{rec}})$ denotes the standard deviation of the reconstructed-energy distribution in a narrow true-energy interval, and the conditional expectation emphasises that the result is expressed as a function of the SD-433 reconstructed energy scale. In practice, the baseline model adopted in this chapter is obtained from simulations (after mapping to the SD-433 energy scale), and it is confronted with a data-driven estimate in the high-energy regime as a consistency check.

A direct estimate of R from data can be obtained using showers jointly reconstructed by the SD-433 and SD-750 arrays. This becomes feasible above the full-efficiency threshold, where the energy bias is negligible and the comparison is dominated by resolution effects. The method, summarised here and described in detail in Appendix A.5, uses the symmetric fractional difference

$$d \equiv \frac{2(E_{SD-433}^{\text{rec}} - E_{SD-750}^{\text{rec}})}{E_{SD-433}^{\text{rec}} + E_{SD-750}^{\text{rec}}},$$

in bins of the geometric-mean proxy $E_{\text{bin}} = \sqrt{E_{\text{SD-433}}^{\text{rec}} E_{\text{SD-750}}^{\text{rec}}}$. Under a Gaussian fractional-response model, the width of the d distribution constrains the combination $V_d = f_{\text{SD-433}}^2 + f_{\text{SD-750}}^2 - 2\rho f_{\text{SD-433}} f_{\text{SD-750}}$. Since $f_{\text{SD-750}}$ is taken from an external parametrisation, the SD-433 resolution $f_{\text{SD-433}}$ can be inferred by constructing a likelihood in each reconstructed-energy bin and profiling over nuisance parameters: the bin-wise mean of d (to absorb residual migration effects) and the unknown SD-433/SD-750 correlation ρ .

The resulting resolution estimates are shown in Fig. 5.14. The coloured markers display the simulation-based resolution for different primary masses after mapping to the SD-433 energy scale, while the black squares correspond to the GSF mixed-composition expectation. The dashed black line shows the fit of the parametrisation

$$R^2(E_{\text{SD-433}}) = R_{\text{sh}}^2 + \frac{R_0^2}{E_{\text{SD-433}}/10^{17} \text{ eV}}, \quad (5.22)$$

which separates an asymptotic high-energy term from a contribution that decreases with energy. In this form, R_{sh} represents the irreducible shower-to-shower fluctuation floor, while R_0 encodes the additional detector term that dominates towards lower energies. An unbinned likelihood fit to the simulation-based GSF mix yields

$$R_{\text{sh}} = 0.103 \pm 0.001, \quad R_0 = 0.096 \pm 0.005,$$

with both coefficients highly significant in the fit (p-values consistent with zero at numerical precision). The vertical dotted line in Fig. 5.14 indicates the full-efficiency threshold.

The light-blue band is the data-driven profile-likelihood envelope obtained from joint SD-433/SD-750 events (Appendix A.5). Although the envelope is broad, reflecting both limited statistics in the overlap sample and the strong degeneracy between $f_{\text{SD-433}}$ and ρ , its lower edge is compatible with the simulation-based model above full efficiency. This agreement supports the use of the simulation parametrisation as a baseline description, while the data-driven band provides a conservative cross-check on the attainable resolution from data.

The dashed green curve shows the SD-750 energy resolution reported in Ref. [139]. Over the energy range common to both arrays, we expect the SD-433 resolution to be comparable or better: at fixed primary energy, the denser grid provides a larger station multiplicity and therefore a more tightly constrained LDF fit. We thus adopt the SD-750 result as a conservative external reference, using it to further restrict the likelihood envelope and to tighten the range of plausible SD-433 energy resolutions.

It is worth noting that the simulation model assumes an idealised detector response in which all WCD stations are identical and calibrated uniformly. In data, additional station-to-station variability (e.g. residual calibration differences, ageing, and environmental effects) can introduce extra event-by-event dispersion that is difficult to capture perfectly in simulations. For this reason, a slightly worse resolution in data than in the nominal simulation model is plausible; the fact that the simulation curve lies near the lower boundary of the data-driven envelope is consistent with this expectation within the present uncertainties.

At high energies, R is dominated by shower-to-shower fluctuations and therefore depends on the intrinsic variability of shower development. Lighter primaries (e.g. protons) typically exhibit larger fluctuations, while heavier nuclei tend to have smaller relative fluctuations (superposition picture). In mixed-composition scenarios, the overall resolution is therefore influenced by the light component, which can dominate the variance even when it is not the dominant fraction by number. The lower panel of Fig. 5.14 shows the energy-dependent mass fractions assumed in the GSF mix used to construct the mixed-composition expectation.

For forward-folding applications, we construct the response function in Eq. (5.1) by combining: (i) the efficiency model $\varepsilon(E_{\text{SD-433}})$ from Section 5.1.5, (ii) the energy-bias model $b(E_{\text{SD-433}})$ from Eq. (5.18), and (iii) the resolution model $R(E_{\text{SD-433}})$ from Eq. (5.22). In practice, we model the spread function $s(E_{\text{SD-433}}^{\text{rec}} | E_{\text{SD-433}}; \theta)$ is modelled with a Gaussian distribution whose mean is shifted according to the bias and whose width is given by the resolution, with the understanding that above full efficiency the bias is small and the response is dominated by shower fluctuations.

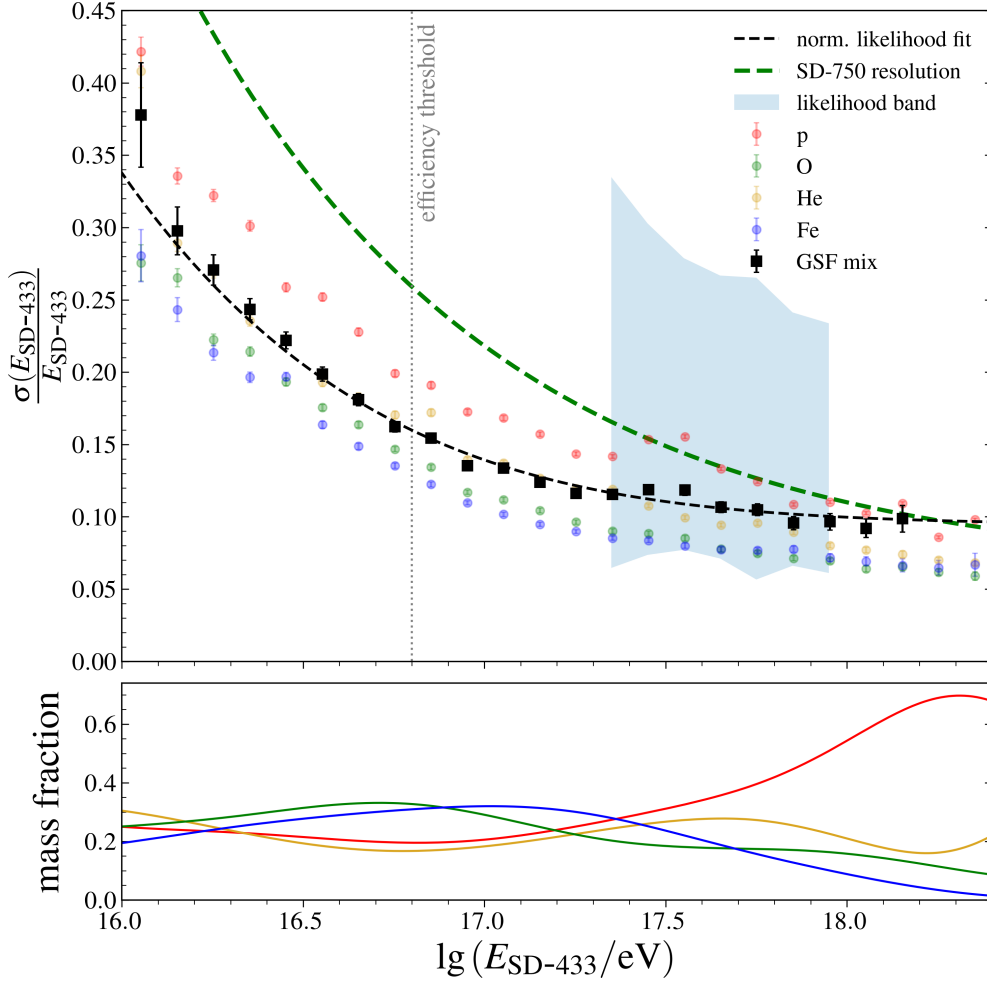


Figure 5.14: SD-433 fractional energy resolution as a function of $\lg(E_{\text{SD-433}}/\text{eV})$. Coloured markers show simulation-based estimates for different primaries after mapping to the SD-433 energy scale, while black squares correspond to the GSF mixed-composition expectation. The dashed black curve shows the fit of the parametrisation in Eq. (5.22). The dashed green curve shows the SD-750 energy resolution taken from Ref. [139]. The vertical dotted line indicates the full-efficiency threshold. The light-blue band is the data-driven profile-likelihood envelope obtained from joint SD-433/SD-750 events using the method described in Appendix A.5, which is applicable in the high-energy regime where the bias is negligible. The lower panel shows the energy-dependent mass fractions assumed in the GSF mix.

5.3 Summary

This chapter characterised the detector response of the SD-433 array in the energy region of the second knee, providing the key ingredients required to connect reconstructed

observables to the underlying primary-energy distribution. Because the SD-433 energy scale is obtained through an indirect transfer calibration, the response characterisation relies primarily on dedicated air-shower simulations, complemented by data-driven cross-checks in the high-efficiency regime.

The trigger-and-reconstruction efficiency was modelled with a probit GLM in terms of $\lg(E)$, $\sin^2 \theta$, and $\ln A$. After correcting simulated events to the SD-433 energy scale to account for mean data–MC discrepancies in shower-size response and attenuation, the resulting $\varepsilon(E_{\text{SD-433}})$ provides a practical, analysis-ready parametrisation for forward-folding applications. An isotropy-based validation with data indicates that the transition to full efficiency is completed between $10^{16.8}$ and $10^{16.9}$ eV for $\theta \leq 45^\circ$, motivating the standard threshold choice used in SD-433 analyses.

The energy response was then completed by quantifying the energy bias and resolution. The bias $b(E_{\text{SD-433}})$ is well described by an exponential approach to a small asymptotic offset, remaining at the percent level above full efficiency; the associated species-dependent separation illustrates the unavoidable composition dependence of a single mixed-composition calibration. The fractional energy resolution was parametrised as $R^2(E_{\text{SD-433}}) = R_{\text{sh}}^2 + R_0^2 / (E_{\text{SD-433}} / 10^{17} \text{ eV})$, yielding an asymptotic shower-to-shower floor and a decreasing low-energy term. A data-driven estimate from joint SD-433/SD-750 events provides a consistency check above full efficiency and is compatible with the simulation-based baseline within uncertainties.

Chapter 6

Muon content

It's the little things that count.

Popular saying

As discussed in Section 3.4, one of the key differences between hadron-induced extensive air showers and photon-induced ones is the muon content at ground. The goal of this chapter is therefore to estimate the muon content of the cosmic-ray background measured at tens of PeV and above. To do so, we exploit the capabilities of the UMD, which provides a direct handle on the muonic component of the shower. A reliable determination of the expected background is essential for assessing the discrimination power of any method aiming at identifying photon candidates and, ultimately, for the measurement of a photon flux or its upper limits.

The connection between muon content and mass composition follows from the hadronic-cascade picture summarised in Section 3.2. In the Heitler-Matthews framework, the total number of muons grows approximately as a power of the primary energy,

$$N_\mu = \left(\frac{E_0}{E_{\text{dec}}} \right)^\alpha, \quad \alpha = \frac{\ln n_{\text{ch}}}{\ln n_{\text{tot}}}, \quad (6.1)$$

where E_0 is the primary energy, E_{dec} is the characteristic energy at which charged mesons preferentially decay rather than re-interact, and n_{ch} and n_{tot} denote the charged and total pion multiplicities per hadronic interaction. For a nucleus of mass number A , the superposition model approximates the shower as the sum of A proton-like sub-showers of energy E_0/A , implying

$$N_\mu^{(A)}(E_0) = A^{1-\alpha} N_\mu^{(p)}(E_0). \quad (6.2)$$

This scaling makes explicit why muon measurements are intrinsically sensitive to composition: at fixed E_0 , heavier primaries yield a larger muon content. Taking logarithms highlights the relevant dependence,

$$\ln N_\mu = \alpha \ln E_0 + (1 - \alpha) \ln A - \alpha \ln E_{\text{dec}} + \text{const.} \quad (6.3)$$

so that, at fixed energy, composition information enters primarily through $(1 - \alpha) \ln A$.

In practice, experiments often express the measured muon content relative to simulated proton and iron baselines (including detector effects) through a dimensionless “muon scale”. In the notation used here, one may write

$$z = \frac{\ln \langle N_{\mu, \text{det}}^{\text{data}} \rangle - \ln \langle N_{\mu, \text{det}}^p \rangle}{\ln \langle N_{\mu, \text{det}}^{\text{Fe}} \rangle - \ln \langle N_{\mu, \text{det}}^p \rangle}, \quad (6.4)$$

for which $z = 0$ corresponds to proton and $z = 1$ to iron under the assumed interaction model and detector response [145]. Under the approximate linearity of $\ln N_\mu$ with $\ln A$ implied by Eq. (6.3), the mean logarithmic mass can then be written explicitly as

$$\langle \ln A \rangle = (\ln 56) \frac{\langle \ln N_{\mu,\text{det}}^{\text{data}} \rangle - \langle \ln N_{\mu,\text{det}}^p \rangle}{\langle \ln N_{\mu,\text{det}}^{\text{Fe}} \rangle - \langle \ln N_{\mu,\text{det}}^p \rangle} = (\ln 56) z, \quad (6.5)$$

which makes explicit how the measured muon content can be mapped into a physically interpretable composition proxy once the proton and iron baselines are defined consistently at detector level [146].

Since the beginning of the AMIGA project, substantial effort has been devoted to achieving a robust determination of the muon content with the UMD. This includes the development and validation of muon-counting strategies tailored to segmented scintillator detectors, as well as the implementation, maintenance, and continuous refinement of the reconstruction chain within the Offline framework. These long-term developments are essential because composition sensitivity and photon-hadron discrimination ultimately rely on controlling detector effects, reconstruction biases, and the stability of the muon measurement across time.

This chapter is organised as follows. Section 6.1 presents the validation of the shape of the mean MLDF, which serves as the first step in establishing the data quality of the UMD-433 and, at the same time, provides nontrivial physical insight. Section 6.2 then characterises the MLDF on an event-by-event basis within the energy range relevant to this work. The mass-sensitive estimator ρ_{21} is introduced in Section 6.3, and the simulated proton and iron baselines required for the physical interpretation of the measurements are presented in Section 6.5. The measured muon content and its implications for a photon search in the same energy range are then discussed in Section 6.6, followed by a brief summary in Section 6.7. Complementary material, including details of the data-cleaning campaign and the updates to the official reconstruction carried out in the context of this thesis, is collected in appendix B, together with additional technical aspects of the analysis that are omitted from the main text for the sake of brevity.

6.1 Shape of the mean Muon Lateral Distribution Function

Before drawing any conclusions about the mass composition of cosmic rays or about a possible excess of muons in data, it is first necessary to verify that the shape of the muon lateral distribution measured with the UMD is consistent with previous observations [147]. A first study in this direction was presented in the thesis of de Jesús [126], where the slope parameter β of the adopted UMD was corrected to account for differences in atmospheric depth between experiments. Here we revisit the same question from a complementary perspective, focusing instead on the characteristic muon-spread scale R_μ , following the discussion of Ref. [147]. The emphasis here is not on the event-by-event parametrisation itself, but on whether the *mean* muon lateral distribution is already consistent with the simple Greisen-based picture that underlies much of the historical discussion.

The present study was motivated by an exchange with Prof. Alan Watson, following a presentation at a Pierre Auger Collaboration meeting, where an event-by-event parametrisation of the UMD-433 muon lateral distribution based on a KASCADE-Grande-like function [148] was discussed:

$$\rho_\mu(r) = N_\mu \left(\frac{r}{R_\mu} \right)^{-\alpha} \left(1 + \frac{r}{R_\mu} \right)^{-\beta} \left(1 + \left(\frac{r}{10R_\mu} \right)^2 \right)^{-\gamma}. \quad (6.6)$$

In that presentation, the scale parameter was denoted by r_0 . Here, however, we adopt the notation R_μ , since r_0 is often used in other contexts for the Molière radius, whereas the quantity relevant for muon lateral distributions is the characteristic muon-spread scale. In this sense, R_μ defines the geometrical scale over which the muon density profile changes shape for a given atmospheric depth and zenith-angle interval.

Following from Ref. [147], the dependence of this scale on atmospheric depth can be estimated from the empirical relation inferred from AGASA data [149],

$$\lg(R_\mu/\text{m}) = 6.25 \times 10^{-4}(X - 920) + 2.39, \quad (6.7)$$

where X is the atmospheric depth in g cm^{-2} . In previous UMD analyses, the value $R_\mu = 320 \text{ m}$ was adopted, following the convention used by KASCADE-Grande. While this is a reasonable choice for the atmospheric depth relevant to that experiment, it is not expected to be optimal under Auger conditions.

The purpose of this section is therefore twofold. First, we test whether the average muon lateral distribution measured with the UMD is well described, in the radial range considered here, by the Greisen form, which also provides the baseline from which several historical muon-LDF parameterisations can be understood. Second, we compare the resulting mean shape with historical measurements from Volcano Ranch and AGASA after applying the corresponding correction in R_μ . These two comparisons are complementary: Volcano Ranch operated at an atmospheric depth similar to that of the Auger site, although with a lower muon-energy threshold of 200 MeV, whereas AGASA had a comparable threshold of 1 GeV but was located at a different depth. The shape test itself concerns the mean UMD distribution as a whole. In the main text, we use the UMD-433 as the reference case, since it provides the clearest and most economical illustration of the argument, while the corresponding validation for the UMD-750 is presented in a dedicated appendix to avoid unnecessary repetition.

6.1.1 Brief remarks on MLDF parameterisations

These remarks follow in part from [120], with additional context from [72, 147, 150]. The benchmark muon lateral distribution considered here originates in the early *Cornell* measurements of Bennett and Greisen. These data correspond to spectrograph measurements of muons, and the resulting parameterisation has remained a useful reference for comparisons below about 800 m. Their importance lies not only in their historical role, but also in the experimental technique itself.

In both the Cornell and Haverah Park spectrograph measurements [150, 151], the muon trajectories were reconstructed event by event, allowing unaccompanied muons to be identified with much greater confidence than in shielded-counter measurements. In the Haverah Park analysis, for example, candidate muons were required to satisfy strict selection criteria: they had to traverse the usable volume of the magnet, be free from confusion with neighbouring tracks, and be consistent with the incident direction of the recorded shower particles. The latter requirement was introduced specifically to reject chance unassociated muons, whose measured probability was reported to be below 0.2% per exposure [152]. For this reason, the spectrograph data provide particularly valuable benchmarks for the present discussion: they offer much better control of possible non-muonic contamination than other historical data sets, where such residual contributions are generally more difficult to assess.

The original Greisen parameterisation may be written as

$$\rho_\mu(N, r, \geq E_{\text{th}}) = \frac{14.4}{(1 + r/R_\mu)^{2.5}} \left(\frac{r}{\text{m}}\right)^{-0.75} \left(\frac{N}{10^6}\right)^{0.75} \frac{51 \text{ GeV}}{E_{\text{th}} + 50 \text{ GeV}} \left(\frac{3 \text{ GeV}}{E_{\text{th}} + 2 \text{ GeV}}\right)^{0.14 (r/\text{m})^{0.37}}, \quad (6.8)$$

where N is the shower size and E_{th} is the muon energy threshold. In this notation, R_μ denotes the *muon-spread factor*, that is, the characteristic geometrical scale governing the lateral spread of the muon distribution. In the original Cornell parameterisation this scale was fixed to 320 m.

A further point worth recalling is that the Cornell data were normalised to a reference shower size of 10^6 particles, reflecting the empirical observation that the muon content does not scale linearly with shower size, but approximately as $N^{0.75}$. This is the origin of the factor $(N/10^6)^{0.75}$ in Eq. (6.8). The remaining radial dependence should be understood primarily as an empirical representation of the observed shape. In this sense, the parameter $R_\mu = 320$ m is not merely a numerical constant, but a characteristic scale describing the lateral spread of the muon component. Indeed, as noted by Greisen in his original discussion, integration of the corresponding lateral distribution shows that half of the muons lie at distances greater than 320 m from the shower axis. This gives a direct physical interpretation to R_μ as a measure of the lateral dispersion of the muonic component.

For the UMD, one can take $E_{\text{th}} = 1$ GeV. In that case the threshold-dependent factors become unity,

$$\frac{51 \text{ GeV}}{E_{\text{th}} + 50 \text{ GeV}} = 1, \quad \left(\frac{3 \text{ GeV}}{E_{\text{th}} + 2 \text{ GeV}} \right)^{0.14 (r/\text{m})^{0.37}} = 1, \quad (6.9)$$

and the expression reduces to

$$\rho_\mu(r) = 14.4 \left(\frac{N}{10^6} \right)^{0.75} \left(\frac{r}{\text{m}} \right)^{-0.75} \left(1 + \frac{r}{R_\mu} \right)^{-2.5}. \quad (6.10)$$

This form already shows that, for the UMD threshold, the average radial behaviour is controlled by two simple factors: a leading power-law decrease with distance and a second term governed by the muon-spread factor R_μ . In other words, once the threshold-dependent terms are removed, the Greisen description reduces to a particularly transparent two-factor shape whose geometry is controlled explicitly by R_μ .

To make this structure more transparent, it is convenient to introduce a reference distance r_{ref} and factor the radial dependence relative to that point. Using

$$\left(\frac{r}{\text{m}} \right)^{-0.75} = \left(\frac{r_{\text{ref}}}{\text{m}} \right)^{-0.75} \left(\frac{r}{r_{\text{ref}}} \right)^{-0.75}, \quad (6.11)$$

and

$$\left(1 + \frac{r}{R_\mu} \right)^{-2.5} = \left(1 + \frac{r_{\text{ref}}}{R_\mu} \right)^{-2.5} \left(\frac{1 + r/R_\mu}{1 + r_{\text{ref}}/R_\mu} \right)^{-2.5}, \quad (6.12)$$

Eq. (6.10) can be rewritten as

$$\rho_\mu(r) = \rho(r_{\text{ref}}) \left(\frac{r}{r_{\text{ref}}} \right)^{-\alpha} \left(\frac{1 + r/R_\mu}{1 + r_{\text{ref}}/R_\mu} \right)^{-\beta}, \quad (6.13)$$

with

$$\alpha = 0.75, \quad \beta = 2.5, \quad (6.14)$$

and

$$\rho_\mu(r_{\text{ref}}) = 14.4 \left(\frac{N}{10^6} \right)^{0.75} \left(\frac{r_{\text{ref}}}{\text{m}} \right)^{-0.75} \left(1 + \frac{r_{\text{ref}}}{R_\mu} \right)^{-2.5}. \quad (6.15)$$

This representation is particularly convenient for the present discussion, since it preserves the reference-distance form while keeping the geometrical meaning of the muon-spread factor R_μ explicit. In this form, the normalisation is simply the density at the chosen reference distance,

whereas the remaining factors describe the radial shape relative to that point. This makes it especially useful when comparing measurements performed at different atmospheric depths: the overall scale and the shape can be discussed separately, while the role of R_μ remains physically transparent.

It is also worth noting that this two-factor form is already very close to the MLDF adopted by AGASA in the distance range $r \lesssim 800$ m [147, 149]. By contrast, including an additional term of the form $[1 + (r/10R_\mu)^2]^{-\gamma}$ leads to the shape used by KASCADE-Grande Eq. (6.6), where the extra factor was introduced to account for deviations of the MLDF at large distances from the shower axis [148]. In this sense, the Greisen form should not be viewed as one arbitrary parameterisation among many. Rather, it provides the common baseline from which several of the historical MLDF descriptions relevant to this analysis can be understood: AGASA is already close to the same two-factor structure over the radial interval studied here, while KASCADE-Grande mainly supplements that structure with an additional large-distance correction.

6.1.2 Mean-shape validation of the UMD: the UMD-433 reference case

The average UMD-433 MLDF is shown in Fig. 6.1¹. We built the profile from events above 100 PeV with zenith angles below 30° , using WCD-reconstructed cores and module distances between 100 and 800 m. Over this interval, the mean muon density is well described by the reduced Greisen form introduced in Eq. (6.10). The quality of the description is already evident from the figure: the fitted curve follows the average profile over the full radial range considered here.

An important feature of this parametrisation is that, once the muon-spread factor R_μ is fixed, the shape of the profile is essentially determined. The remaining free parameter then acts mainly as an overall normalisation, setting the scale of the muon density while leaving the radial dependence unchanged.

When the muon-spread factor derived from Eq. (6.7) is fixed to the value expected for Auger conditions, $R_\mu = 247$ m, the fitted normalisation is

$$N = (8.40 \pm 0.10) \times 10^7. \quad (6.16)$$

For comparison, imposing the historical KASCADE-Grande choice $R_\mu = 320$ m leads to a slightly poorer description in this restricted interval, with

$$N = (5.45 \pm 0.10) \times 10^7. \quad (6.17)$$

This illustrates a point already discussed: if the lateral scale is fixed to a value that is not optimal for the atmospheric depth of the data, part of the mismatch can be absorbed by the remaining fit parameters. In practice, the normalisation may shift so as to compensate, at least partially, for an inappropriate choice of R_μ . The same is true of additional shape parameters. If they are left free, they may also absorb part of the effect of a mis-specified R_μ . In particular, a biased choice of R_μ would most likely be accompanied by a compensating shift in β .

A more direct check is obtained by leaving R_μ free in the Greisen fit. In that case one finds

$$N = (8.35 \pm 0.60) \times 10^7, \quad R_\mu = (248 \pm 10) \text{ m}, \quad (6.18)$$

fully consistent with the fixed-value estimate derived from the atmospheric-depth scaling argument. This agreement is important. It shows that the mean UMD-433 distribution is

¹This is the reference case used to illustrate the mean-shape tests throughout the main body of this section. The corresponding analyses for the UMD-750 case are shown in section B.3.2.

not merely *compatible* with a Greisen-like curve, but that the preferred lateral scale extracted directly from the data coincides with the value expected for the Auger depth.

The same conclusion can be expressed in the reference-distance form of Eq. (6.13). Fixing $R_\mu = 247$ m and $\alpha = 0.75$, and fitting instead the density at $r_{\text{ref}} = 250$ m [153] together with the second exponent, yields

$$\rho(250) = (1.11 \pm 0.02) \text{ m}^{-2}, \quad \beta = 2.50 \pm 0.05. \quad (6.19)$$

This form makes the separation between normalisation and shape even more explicit: ρ_{250} fixes the overall scale of the profile, while β controls the residual curvature once R_μ is specified. The fitted slope is fully consistent with the canonical Greisen value $\beta = 2.5$, showing that, after adopting the physically motivated value of R_μ , the average UMD-433 shape does not require any appreciable modification of the standard two-factor form over the radial interval tested here. In particular, within the range effectively sampled in this study, there is no evidence that an additional large-distance correction term is needed at the level of the mean profile. This conclusion should nevertheless be understood as local to the present fit range. If the mean MLDF were to be followed significantly beyond 800 m, the simple two-factor form might no longer remain sufficient, and a third term analogous to that used in the KASCADE-Grande parametrisation could become necessary to describe the large-distance behaviour.

Taken together, these fits support a simple picture. The reduced Greisen description captures the mean UMD-433 data with a scale factor R_μ that agrees with the expectation for the Auger atmospheric depth, while the remaining fitted parameters behave essentially as normalisations of an already adequate shape. The event-by-event scatter around the mean profile is shown separately in section B.3.1. The analogous UMD-750 validation is presented in section B.3.2, where the same overall picture is recovered, and where the possible role of large-distance corrections becomes even more relevant because that configuration naturally probes higher energies and broader zenith angles.

6.1.3 Comparison with other experiments

Given the good agreement of the mean UMD-433 profile with the reduced Greisen form established in Section 6.1.2, the comparisons with earlier experiments naturally follow rather closely the Greisen-based comparisons already discussed in [147]. The fitted UMD-433 shape can be transported to the atmospheric depth relevant for a given historical data set by keeping the same functional form and replacing the muon-spread factor by the value implied by Eq. (6.7). This procedure isolates the geometrical effect associated with atmospheric depth and allows the resulting mean shape to be compared directly with previous measurements. The result is shown in Fig. 6.2 for Volcano Ranch and AGASA.

For Volcano Ranch, the UMD-433 curve shifted to $X = 834 \text{ g cm}^{-2}$ corresponds to $R_\mu \simeq 245$ m and is normalised around $r \approx 250$ m. Around this reference point the agreement is naturally good, and the overall decrease with distance is broadly reproduced. Away from that point, however, the two shapes are not identical: the Volcano Ranch measurements tend to lie somewhat above the shifted UMD curve at small distances and somewhat below it at larger distances, producing a visible trend in the ratio panel. This behaviour is not unexpected. The muon-energy threshold of Volcano Ranch was closer to ~ 200 MeV than to the 1 GeV relevant for the UMD, so some difference in shape should be anticipated even at similar atmospheric depth. In addition, the Volcano Ranch signal inside 100 m lies above the Greisen prediction, a behaviour that is at least qualitatively consistent with an additional near-axis contribution, potentially due to electromagnetic punch-through or, as examined in Ref. [125], knock-on electrons.

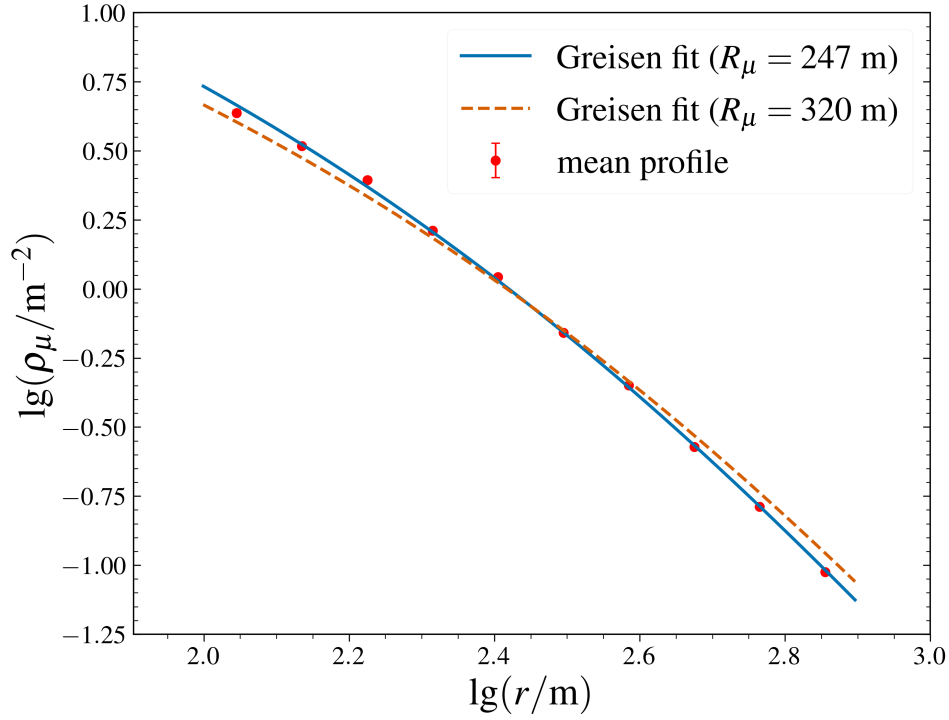


Figure 6.1: Mean UMD-433 muon density profile as a function of core distance together with two Greisen fits, one with fixed $R_\mu = 247$ m and one with fixed $R_\mu = 320$ m.

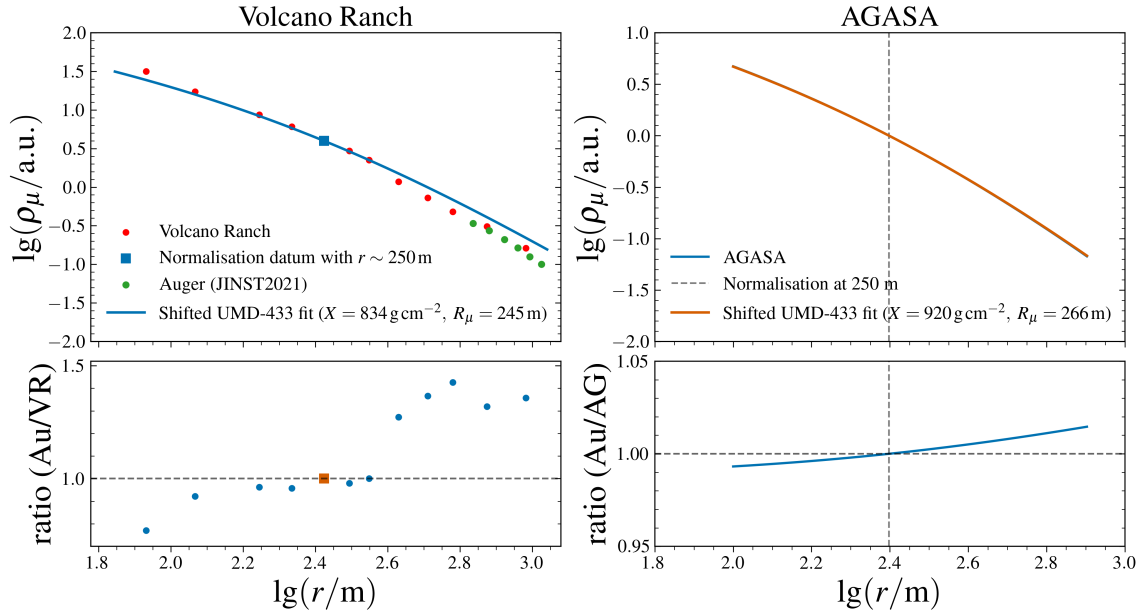


Figure 6.2: Comparison of the shifted UMD-433 Greisen fit with historical measurements from Volcano Ranch and AGASA. In the Volcano Ranch panel, green markers indicate the published Auger Water-Cherenkov neural-network reconstruction points shown for reference.

In this context, it is also useful to recall the complementary Auger neural-network reconstruction of the muon component from Water-Cherenkov Detector signals [154], discussed in Ref. [147]. There, the comparison with AGASA was found to worsen beyond about 1 km, whereas the agreement with Volcano Ranch was judged to be more satisfactory, despite the limited overlap in core distance. For illustration, the corresponding Auger points are also

shown as green markers in the Volcano Ranch panel of Fig. 6.2. A natural interpretation is that threshold effects play an important role: the neural-network analysis is sensitive to lower-energy muons than the 1 GeV threshold relevant for the UMD and AGASA, with the simulations used in that work understood to correspond to muons of order 300 MeV. In that sense, the Volcano Ranch data provide a useful intermediate benchmark for interpreting threshold-dependent differences in LDF shape.

The AGASA comparison is more stringent and, correspondingly, more reassuring. After shifting the UMD-433 fit to $X = 920 \text{ g cm}^{-2}$, which gives $R_\mu \simeq 266 \text{ m}$, the agreement with the AGASA parameterisation is remarkably close over the common radial range shown in Fig. 6.2. The ratio remains close to unity and shows only a mild trend with distance. This is exactly the behaviour anticipated in Section 6.1.1: below 800 m, the two-factor Greisen form and the AGASA parameterisation are expected to be nearly equivalent once the appropriate muon-spread factor is used.

6.2 Muon Lateral Distribution Function for the UMD-433

The extraction of a mass-sensitive muon observable with the UMD-433 requires a precise and stable reconstruction of the MLDF. In particular, before addressing the muon content of air showers in the SD-433 energy range, it is necessary to parametrize the *slope* of the MLDF, β , in a manner that remains robust under the limited station multiplicities characteristic of this array configuration. Since the mean MLDF was already examined in the previous section, we now focus on an event-by-event description suitable for the present analysis.

Guided by those results, we adopt Eq. (6.13) with $R_\mu = 247 \text{ m}$ and $\alpha = 0.75$. Accordingly, much of this section is devoted to developing a dedicated parametrisation of β for the UMD-433. The choice of the reference distance, r_{ref} , is also a crucial ingredient, since it should minimize the systematic uncertainty induced by the unknown true shape of the lateral distribution [155]. This distance is expected to depend mainly on the geometry of the array, in particular its spacing and layout, rather than on the primary energy or mass. In the case of the UMD-750, $r_{\text{ref}} = 450 \text{ m}$ was chosen as a compromise between Poissonian fluctuations and shower-to-shower fluctuations in the number of counted muons [128]. For the UMD-433, however, we adopt a data-driven determination of r_{ref} , presented in the next subsection, so as to reduce the dependence on hadronic-model assumptions before introducing the parametrisation of the MLDF slope β . Together, these elements provide the basis for defining a mass-sensitive observable and for propagating the systematic uncertainties associated with the MLDF shape.

Motivated by the smooth trends expected in the selected phase space [125, 126], we adopt a first-order expansion in $\lg(E)$ and $\sec \theta$,

$$\hat{\beta}(E, \theta) = \beta_0 + \beta_E \left[\lg\left(\frac{E}{\text{eV}}\right) - \lg\left(\frac{E_0}{\text{eV}}\right) \right] + \beta_\theta (\sec \theta - \sec \theta_0), \quad (6.20)$$

where (E_0, θ_0) are reference values chosen only to reduce correlations among fitted parameters. For the purpose of propagating MLDF-shape uncertainties, we encode the energy dependence of the resolution with the empirical form

$$\sigma_\beta \left(\lg\left(\frac{E}{\text{eV}}\right) \right) = a \exp \left[-b \left(\lg\left(\frac{E}{\text{eV}}\right) - \lg\left(\frac{E_0}{\text{eV}}\right) \right) \right] + c, \quad (6.21)$$

Finally, we introduce a few quantities used below. Assuming an uncertainty σ_β on the MLDF slope, we define $\beta_{\text{hi}} = \beta_{\text{nom}} + \sigma_\beta$ and $\beta_{\text{lo}} = \beta_{\text{nom}} - \sigma_\beta$, where β_{nom} is the nominal value given by the $\hat{\beta}$ parametrisation. Fitting the MLDF with each slope $\beta_i \in \{\beta_{\text{lo}}, \beta_{\text{nom}}, \beta_{\text{hi}}\}$

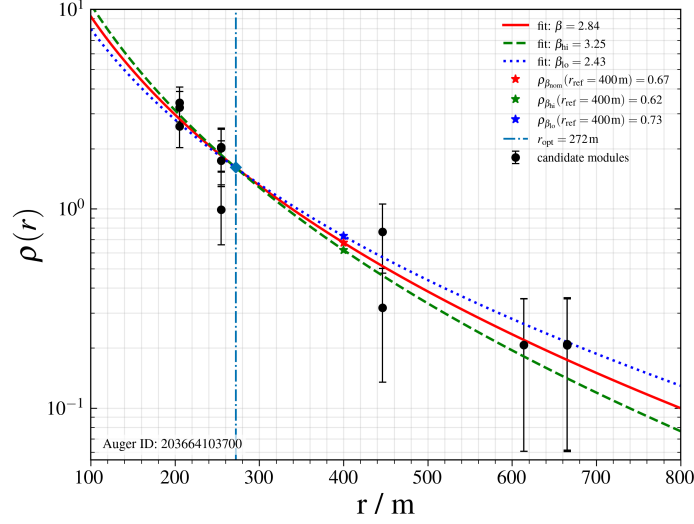


Figure 6.3: MLDF reconstruction for a representative UMD-433 event (Auger ID: 203664103700). Black markers indicate the muon-density estimates from the candidate UMD modules entering the likelihood fit. The curves show the best-fit NKG-based MLDF of Eq. (6.13) obtained when fixing the slope parameter to the nominal value β_{nom} (solid red), and to the perturbed values $\beta_{\text{hi}} = \beta_{\text{nom}} + \sigma_{\beta}$ (dashed green) and $\beta_{\text{lo}} = \beta_{\text{nom}} - \sigma_{\beta}$ (dotted blue). The stars mark the corresponding normalisations $\rho_i(r_{\text{ref}})$ evaluated at $r_{\text{ref}} = 400$ m, illustrating how the slope uncertainty propagates into the density estimator. The vertical dash-dotted line indicates the optimum distance $r_{\text{opt}} = 272$ m, defined as the distance where the spread among the three ρ_i estimates is minimal.

yields a corresponding density estimator ρ_i . The optimum distance r_{opt} is defined event by event (and is largely independent of r_{ref}); it is the shower-plane distance at which the spread among the three ρ_i estimates is minimal. Equivalently, r_{opt} is the distance where uncertainty in the MLDF slope has the smallest impact on the inferred density. An example with a real event and using $r_{\text{ref}} = 400$ m is presented in Fig. 6.3.

6.2.1 The reference distance

The goal of this subsection is to select the UMD-433 reference distance r_{ref} such that $\rho(r_{\text{ref}})$ is minimally affected by plausible variations of the MLDF shape. Operationally, we propagate the uncertainty of the slope parameter β into the normalisation $\rho(r_{\text{ref}})$. The starting point is a high-quality UMD-433 sample of 6T5 events where the hottest SD station is associated with at least one working UMD module. We further require at least five UMD stations, so enabling the event-by-event fit of the MLDF with a free β . Note this condition is less strict than the lever-arm criteria usually employed in SD studies [156], which will be applied in the last step of the slope parametrisation. Finally, to secure an accurate determination of the shower geometry, we require the closest counter to be located at least 50 m from the shower axis.

The resulting statistics of high-quality events as a function of energy and zenith angle are summarised in Fig. 6.4 (left). We also highlight two phase-space regions for the analysis: $\theta < 45^\circ$ for studies above $\lg(E/\text{eV}) \simeq 17.0$ and $\theta \lesssim 30^\circ$ to preserve sufficient station multiplicity at energies down to $\lg(E_{\text{SD-433}}/\text{eV}) \simeq 16.7$ without introducing strong near-threshold distortions. The high-energy threshold of $\lg(E_{\text{SD-433}}/\text{eV}) = 17.6$ is chosen to preserve sufficient event statistics for a reliable estimate of the spread in $\rho(r_{\text{ref}})$. The core positions of the selected events (restricted to the low-energy region and $\theta \lesssim 30^\circ$) are shown in Fig. 6.4 (right).

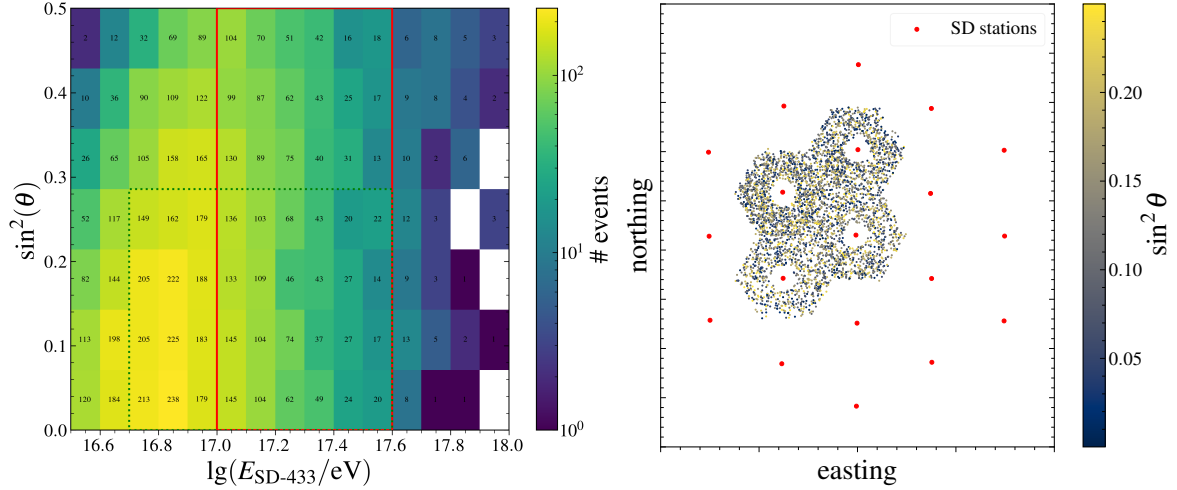


Figure 6.4: Overview of the UMD-433 event sample used to build the MLDF-shape study and the β parameterisation. *Left:* occupancy map in the $(\lg(E_{\text{SD-433}}/\text{eV}), \sin^2 \theta)$ plane after the baseline quality selections and requiring at least five candidate UMD stations. The solid red contour marks the broader comparison region ($\theta < 45^\circ$) used in a conservative energy range, while the dotted green contour highlights a reduced zenith range adopted to preserve UMD multiplicity and extend the analysis towards lower energies; the upper-energy bound is chosen to avoid the regime where the shower footprint increasingly extends beyond the instrumented area. *Right:* reconstructed core positions for events within the dotted green contour (see left panel), colour-coded by $\sin^2 \theta$; red markers indicate the SD-433 station locations. The concentration of cores within only a subset of four hexagons reflects the progressive deployment of UMD stations throughout Phase I.

We test a discrete set of potential reference distances,

$$r_{\text{ref}} \in \{200, 225, 250, 275, 300, 325, 350, 375, 400\} \text{ m}, \quad (6.22)$$

and, for each distance, we:

1. Construct an energy–zenith parametrisation $\hat{\beta}(E, \theta)$ using events reconstructed with free $(\rho(r_{\text{ref}}), \beta)$, with α fixed (Eq. (6.20)). The width of the relative residual $(\beta - \hat{\beta})/\beta$ provides an energy-dependent estimate of the β uncertainty we fit with Eq. (6.21).
2. Propagate the slope uncertainty into three representative MLDF shapes for each event: a nominal curve from $\hat{\beta}(E, \theta)$ and two variations obtained by shifting the slope to β_{lo} and β_{hi} . Keeping the shape fixed for each variant, we refit only the normalisation $\rho(r_{\text{ref}})$, yielding $(\rho_{\text{lo}}, \rho_{\text{nom}}, \rho_{\text{hi}})$ and, equivalently, three MLDF curves that bracket the expected shape uncertainty.
3. Estimate the optimum distance r_{opt} for each event by scanning $r \in [50, 800]$ m.

The resulting r_{opt} distributions are summarised, after binning in energy, by the mean profiles shown in Fig. 6.5 for both zenith selections ($\theta \leq 30^\circ$ and $\theta \leq 45^\circ$) and $r_{\text{ref}} = 325$ m. We checked that r_{opt} is largely independent of the particular normalisation distance r_{ref} used to define $\rho(r_{\text{ref}})^2$. Although $\langle r_{\text{opt}} \rangle$ shows a mild increase with energy, it remains confined to a narrow range, centred around ~ 250 m over the full phase spaces considered. Interestingly, an analysis of ρ_μ fluctuations around the mean, similar to Ref. [128], shows that the minimum (or initial plateau) extends to larger distances with increasing energy, consistent with the behaviour of r_{opt} (see section B.4).

²The spread of r_{opt} for the different values of r_{ref} considering all events is less than 1%.

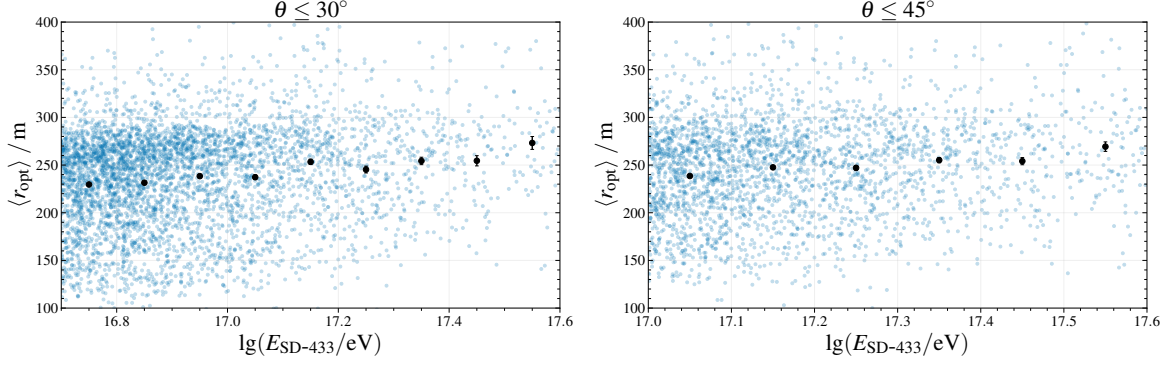


Figure 6.5: The optimum distance, r_{opt} , for the UMD-433 for $r_{\text{ref}} = 325$ m (see text for details). Blue points show the resulting r_{opt} values for individual events, while black markers indicate the mean in energy bins with statistical uncertainties, for the two quoted zenith angle ranges.

While $\langle r_{\text{opt}} \rangle$ provides the first cue for r_{ref} , we can further study the relative spread of the reconstructed muon-density estimator under MLDF-shape variations. For each tested r_{ref} we quantify the corresponding shape-induced spread of $(\rho_{\text{lo}}, \rho_{\text{nom}}, \rho_{\text{hi}})$ through the profile quantity

$$\delta\rho = \left\langle \frac{\rho_{\text{max}} - \rho_{\text{min}}}{\rho_{\text{nom}}} \right\rangle, \quad \rho_{\text{min}} / \rho_{\text{max}} = \min / \max(\rho_{\text{lo}}, \rho_{\text{nom}}, \rho_{\text{hi}}), \quad (6.23)$$

evaluated in bins of $\lg(E_{\text{SD-433}} / \text{eV})$. This is similar to the scan conducted to find r_{opt} , but here the spread of the muon estimator is computed only at r_{ref} . This observable directly measures the sensitivity of the reconstructed normalisation to plausible MLDF-shape variations driven by the uncertainty on β .

We summarise the propagation of the MLDF-slope uncertainty into the muon-density estimator in Fig. 6.6. At low energies the relative span of the reconstructed normalisation increases rapidly, with a slightly stronger effect for the broader zenith selection $\theta < 45^\circ$ at smaller reference distances. In both phase spaces considered, the reference distance that minimises the uncertainty on $\rho(r_{\text{ref}})$ is consistently found to be $r_{\text{ref}} \simeq 250$ m.

In previous UMD-750 studies [125, 126], *half* of this relative span was assigned as a systematic uncertainty on the muon-density estimator. Applying the same convention here, the lowest-energy bins would incur a systematic uncertainty that becomes comparable to that associated with the energy assignment (typically taken as $\sim 14\%$ in most analyses). Therefore, pushing the analysis to even lower energies would lead to a shape-driven systematic on the muon estimator that is no longer negligible within the overall uncertainty budget.

Given that the expected event statistics remain sufficient down to $\lg(E_{\text{SD-433}} / \text{eV}) \simeq 16.7$, even after the reduced aperture implied by the tighter zenith cut, we proceed with the reduced zenith selection. This choice keeps the propagation of β -uncertainties under control while still allowing the study to extend towards lower energies. Within this phase space, we adopt the UMD-433 reference distance $r_{\text{ref}} = 250$ m.

6.2.2 Slope parametrisation

In the previous subsection, we selected the UMD-433 reference distance $r_{\text{ref}} = 250$ m by minimising the shape-induced spread on the reconstructed muon-density estimator. We now set the *final* parameterisation of the MLDF slope $\hat{\beta}(E, \theta)$. This parameterisation allows us to (i) stabilise reconstructions in low-multiplicity regimes by fixing the MLDF shape, and (ii) propagate MLDF-shape systematics through controlled variations of β .

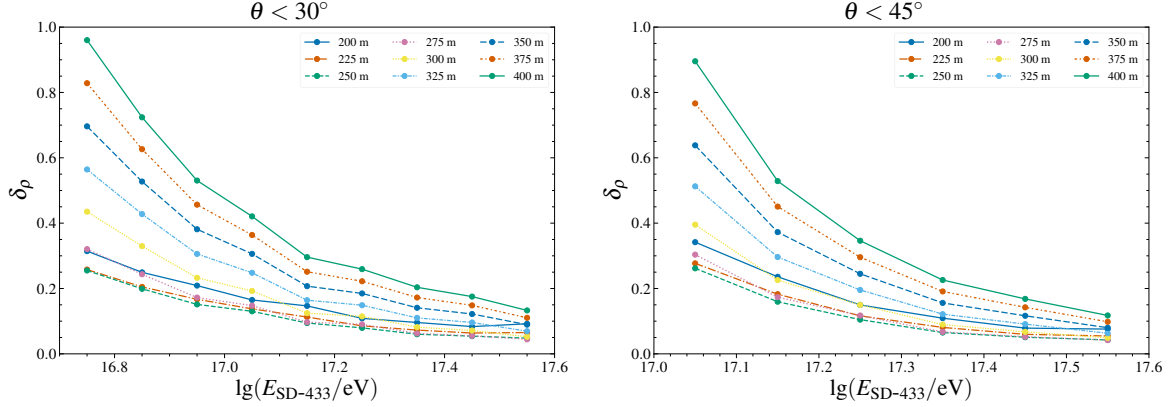


Figure 6.6: Mean shape-induced relative span of the reconstructed muon-density estimator, $\langle(\rho_{\max} - \rho_{\min})/\rho_{\text{nom}}\rangle$ (see Eq. (6.23)), as a function of $\lg(E_{\text{SD-433}}/\text{eV})$ for each tested normalisation distance r_{ref} . *Left:* reduced zenith selection ($\theta < 30^\circ$). *Right:* broader zenith selection ($\theta < 45^\circ$).

The slope parameter is finally parametrised from a set of high-quality events that satisfy a relaxed version of the lever-arm criteria³ employed for the SD-433 [156]. The criteria ensure that the lateral distribution is sufficiently sampled to constrain β even at the modest station multiplicities typical of UMD-433. We select candidate UMD stations within a radial window $r \in [50, 450]$ m and check if the maximum separation between stations is above a certain threshold Δr depending on the number of candidate UMD stations in the range, as specified in Table 6.1.

Table 6.1: Lever-arm selection used to decide whether the MLDF slope β is fitted freely. Distances are measured in the shower plane and the span is computed as $\Delta r = r_{\max}^{(\text{evt})} - r_{\min}^{(\text{evt})}$ using unique candidate stations within $[r_{\min}, r_{\max}]$.

Parameter	Value
Reference distance	$r_{\text{ref}} = 250$ m
Lever-arm radial window	$r \in [r_{\min}, r_{\max}] = [50, 450]$ m
Minimum span for $N_{\text{in range}} = 2$	$\Delta r \geq 150$ m
Minimum span for $N_{\text{in range}} = 3$	$\Delta r \geq 125$ m
Minimum span for $N_{\text{in range}} \geq 4$	$\Delta r \geq 100$ m

The event-by-event fitted β values and their trends with energy and zenith are shown in Fig. 6.7. The binned averages (black markers) indicate that the mean evolution is mild across the phase space considered, supporting the low-order form adopted in Eq. (6.20). The best-fit parameters are summarised in Table 6.2⁴. The reference values $\lg(E_0/\text{eV})$ and $\sec \theta_0$ are set to the medians of their respective distributions. In addition, an alternative parametrisation in terms of the shower size $S(300)$, rather than $\lg(E_{\text{SD-433}}/\text{eV})$, is provided in section B.5.

To quantify the quality of the parametrisation, we study the relative residual

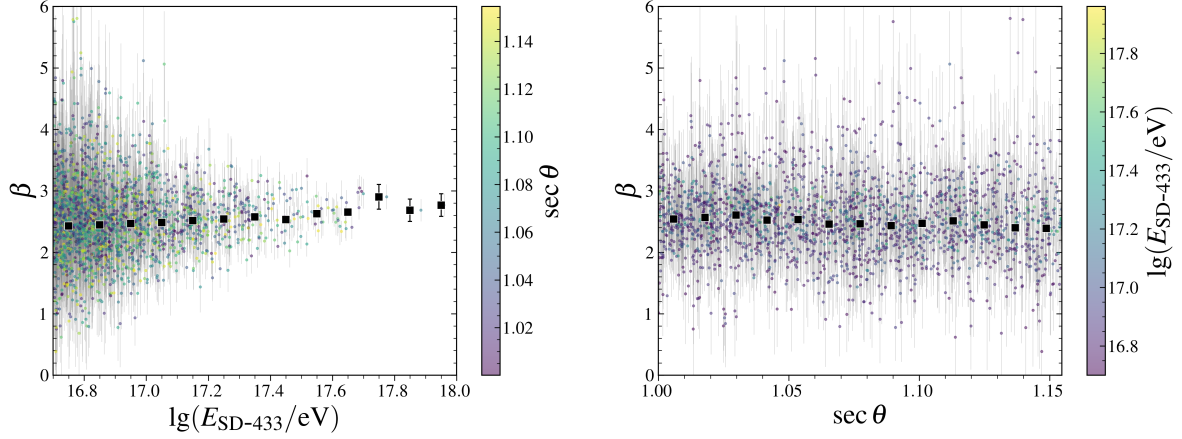
$$\delta_\beta \equiv \frac{\beta - \hat{\beta}}{\beta}, \quad (6.24)$$

³Because the muonic lateral distribution is intrinsically less steep than the LDF measured by the SD (where an electromagnetic component enhances the radial fall-off), and because the UMD-433 event sample is comparatively limited, adopting overly restrictive criteria would unduly reduce the statistics available for the shape study

⁴In practice, Eq. (6.20) is fitted with per-event weights set by their uncertainties σ_β returned by the MLDF fit.

Table 6.2: Parameters of the UMD-433 slope parameterisation Eq. (6.20) for $r_{\text{ref}} = 250$ m in the reduced-zenith phase space.

Quantity	Value
$\lg(E_0/\text{eV})$	16.9
$\sec \theta_0$	1.07
β_0	2.45 ± 0.06
β_E	-1.25 ± 0.10
β_θ	0.25 ± 0.08

**Figure 6.7:** Event-level MLDF slope β obtained from free-shape fits (with α and γ fixed) in the UMD-433 high-quality sample used for the parameterisation. *Left:* β versus $\lg(E_{\text{SD-433}}/\text{eV})$, coloured by $\sec \theta$. *Right:* β versus $\sec \theta$, coloured by $\lg(E_{\text{SD-433}}/\text{eV})$. Thin grey vertical bars indicate the fit uncertainty σ_β for each event. Black squares show binned, uncertainty-weighted averages, illustrating the mild mean dependence that is captured by Eq. (6.20).

and summarize its bias and resolution in energy bins. The corresponding results are shown in Fig. 6.8. Across the fitted range the bias is consistent with zero at the two percent level (orange band), while the resolution improves with energy as the average station multiplicity increases. The best-fit parameters for the resolution model of Eq. (6.21) are $a = 0.03 \pm 0.01$, $b = 2.6 \pm 0.3$, and $c = 0.03 \pm 0.02$. This prescription provides a compact, energy-dependent estimate of the MLDF-shape uncertainty induced by imperfect knowledge of β , which is propagated into the systematic uncertainty on $\rho(r_{\text{ref}})$.

As an independent validation of the reconstruction once β is fixed to Eq. (6.20), we study the residuals of the fitted MLDF with respect to the per-module muon-density estimates. Fig. 6.9 shows, in energy bins for the reduced-zenith phase space, the distribution of the relative residual

$$(\rho(r) - \text{MLDF}(r)) / \text{MLDF}(r),$$

where $\text{MLDF}(r)$ denotes the best-fit MLDF evaluated at the station distance r . In spite of the broad energy range considered, the description of the MLDF remains consistent with the density measurements around the reference distance.

6.3 The observable ρ_{21}

With the MLDF formalism in place, we can now pursue a measurement of the muon content using UMD-433 events, with particular emphasis on the energy region around the second knee. The results of this study can be reproduced using the standard Offline applica-

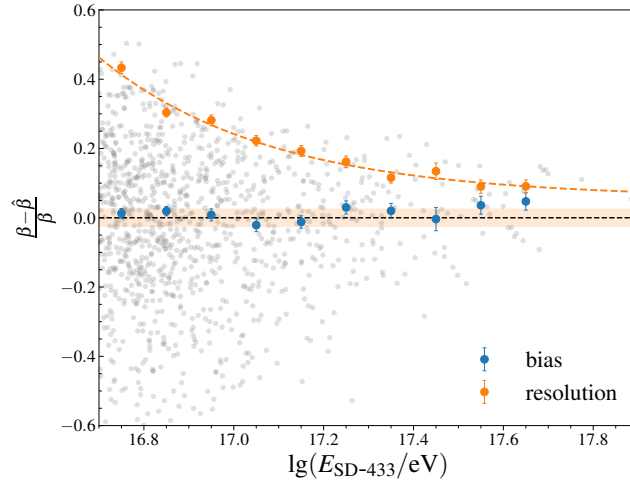


Figure 6.8: Relative residuals of the slope parameterisation, $\delta_\beta = (\beta - \hat{\beta})/\beta$ (grey points), as a function of $\lg(E_{SD-433}/\text{eV})$. Blue markers show the binned bias (mean of δ_β), while orange markers show the binned resolution (standard deviation of δ_β). The dashed curve is a fit of the resolution with the empirical model in Eq. (6.21).

tion MdSd433DataReconstruction with the corrected β parametrisation presented in [157]. In addition to the standard reconstruction settings, we adopt the following configurations and quality cuts:

1. Require 6T5 events.
2. Select events with $\theta \leq 30^\circ$ and $\lg(E_{SD-433}/\text{eV}) \geq 16.7$.
3. Restrict the data set to the period from 1 January 2018 to 21 December 2021 (Phase I).
4. Exclude SD periods flagged as bad.
5. Determine the core position in the MLDF fit.
6. Require at least three candidate UMD stations in the fit, including one associated with the hottest SD station and at least one belonging to an immediate neighbour.

These are not necessarily the same quality cuts as those employed for the UMD-750, but rather a relaxed selection intended to compensate for the limited statistics. Since only four hexagons were active during Phase I, stricter selection criteria can be applied in Phase II. The quality cuts employed here aim to balance statistics and reliability.

The UMD-433 reconstruction sequence provides, for each event, an estimator of the muon density at a fixed reference distance, $\rho(250)$, defined as the normalisation of the fitted lateral distribution at $r_{\text{ref}} = 250$ m. For composition studies it is desirable to remove the residual zenith-angle dependence induced by atmospheric (and soil) attenuation, so that the observable can be compared across arrival directions in a uniform way.

Figure 6.10 (left) shows $\rho(250)$ as a function of reconstructed SD-433 energy. The lower panel compares the average signal measured in a “vertical” ($\theta < \theta_{\text{ref}}$) to that of an “inclined” subsample ($\theta > \theta_{\text{ref}}$), where the reference angle is chosen as $\theta_{\text{ref}} = 21^\circ$ (the median zenith of the selected data set). A mild, but non-negligible attenuation trend is visible. To correct for it, we follow the Constant Intensity Cut (CIC) method of Ref. [143] and determine an attenuation function $f_{\text{att}}(\theta)$ such that, after applying an angle-dependent threshold, the

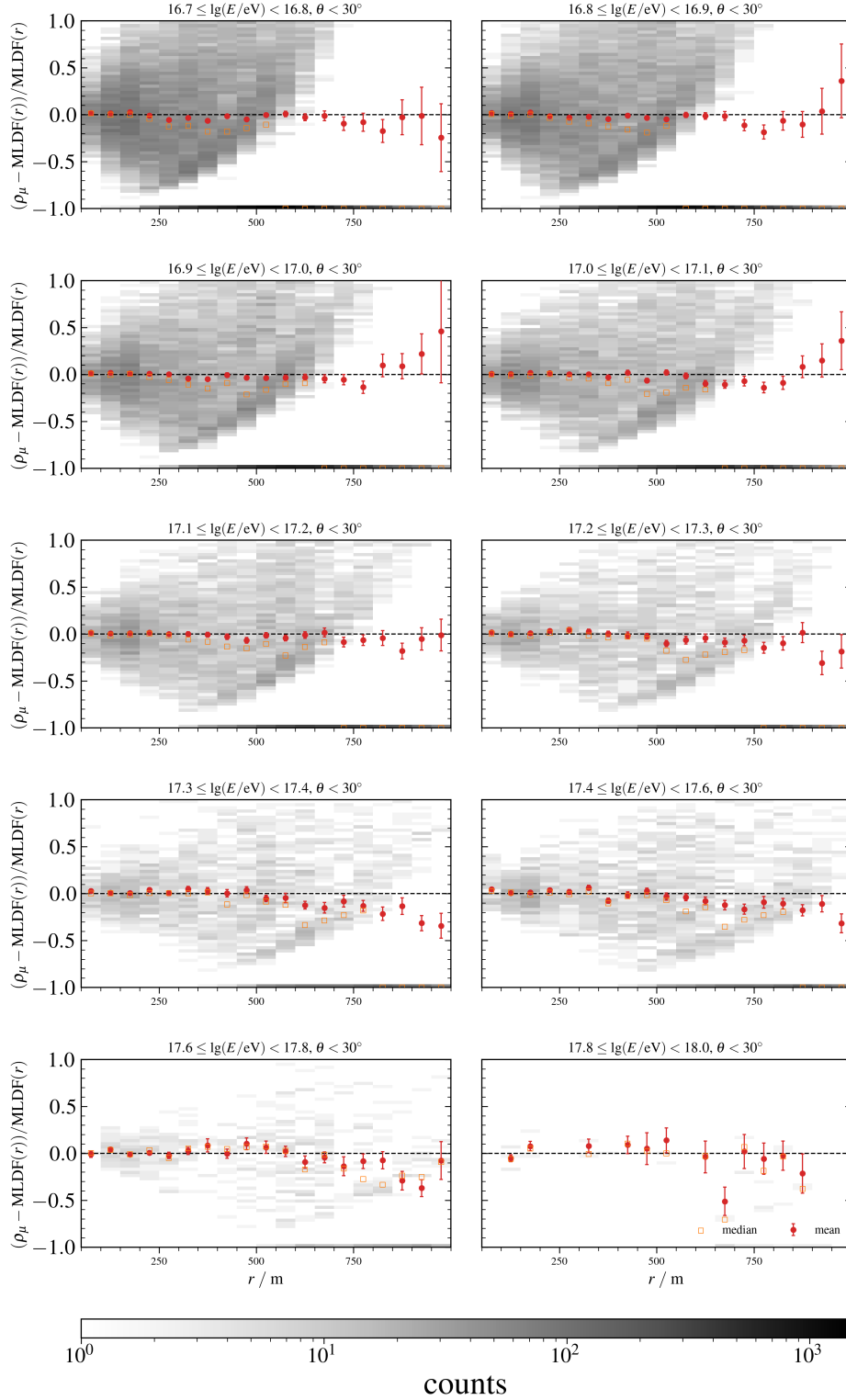


Figure 6.9: Residuals of the reconstructed MLDF as a function of shower-plane distance for the reduced-zenith phase space ($\theta < 30^\circ$), shown in bins of $\lg(E_{SD-433}/\text{eV})$. The grey density map represents the per-station residual distribution (colour scale in counts). Orange squares indicate the median residual, while red markers show the mean with its statistical uncertainty. The horizontal dashed line marks perfect agreement. The residuals are computed from the production in which the MLDF slope is fixed to $\hat{\beta}(E, \theta)$ from Eq. (6.20).

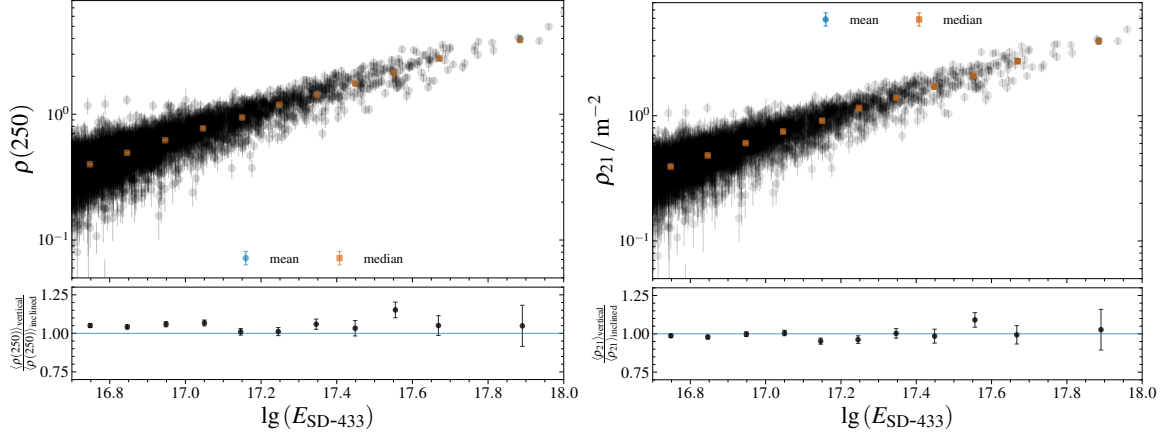


Figure 6.10: Attenuation correction for the UMD-433 muon-density estimator at $r_{\text{ref}} = 250$ m. *Left:* $\rho(250)$ as a function of reconstructed SD-433 energy. The cloud shows individual events, while blue circles (orange squares) indicate the binned mean (median). The lower panel shows the ratio of the average signal in a “vertical” subsample ($\theta < \theta_{\text{ref}}$) to that in an “inclined” subsample ($\theta > \theta_{\text{ref}}$), with $\theta_{\text{ref}} = 21^\circ$. *Right:* same representation after applying the CIC attenuation correction, i.e. using $\rho_{21} \equiv \rho(250)/f_{\text{att}}(\theta)$, for which the vertical/inclined ratio is consistent with unity within statistical fluctuations.

selected event distribution is compatible with isotropy in $\sin^2 \theta$. The implementation details and validation checks specific to this study are summarised in section B.6.

The energy and zenith dependence of the estimator is factorised as

$$\rho(250; E, \theta) = \rho_{21}(E) f_{\text{att}}(\theta), \quad (6.25)$$

and we parameterize the attenuation with the quadratic form

$$f_{\text{att}}(\theta) = 1 + a x + b x^2, \quad x \equiv \sin^2 \theta - \sin^2 \theta_{\text{ref}}, \quad (6.26)$$

with $\theta_{\text{ref}} = 21^\circ$ so that $f_{\text{att}}(\theta_{\text{ref}}) = 1$ by construction. The observable ρ_{21} corresponds to the muon density we would expect had the shower arrived at $\theta_{\text{ref}} = 21^\circ$. The chosen attenuation function is finally defined by

$$a = -0.43 \pm 0.05, \quad b = 0.86 \pm 0.08, \quad (6.27)$$

with $\text{corr}(a, b) = -0.35$.

The impact of the attenuation correction is illustrated in the right panel of Fig. 6.10. After mapping each event to the reference angle through Eq. (6.25), the energy dependence of the estimator is preserved, while the residual zenith-angle trend is removed: the ratio between the average signal in the “vertical” and “inclined” subsamples becomes compatible with unity across the full reconstructed-energy range, within statistical fluctuations. This confirms that the fitted $f_{\text{att}}(\theta)$ captures the relevant attenuation in the zenith phase space and provides a zenith-independent muon-density observable suitable for comparisons across arrival directions.

For visualisation purposes, we define a detrended version of the observable,

$$\frac{\rho_{21}}{(E/E_0)^{0.9}}, \quad (6.28)$$

where $E_0 = 10^{16.9}$ eV is a pivot energy (median of the energy distribution from selected events) and the exponent 0.9 is a convenient reference scaling motivated by the expectations

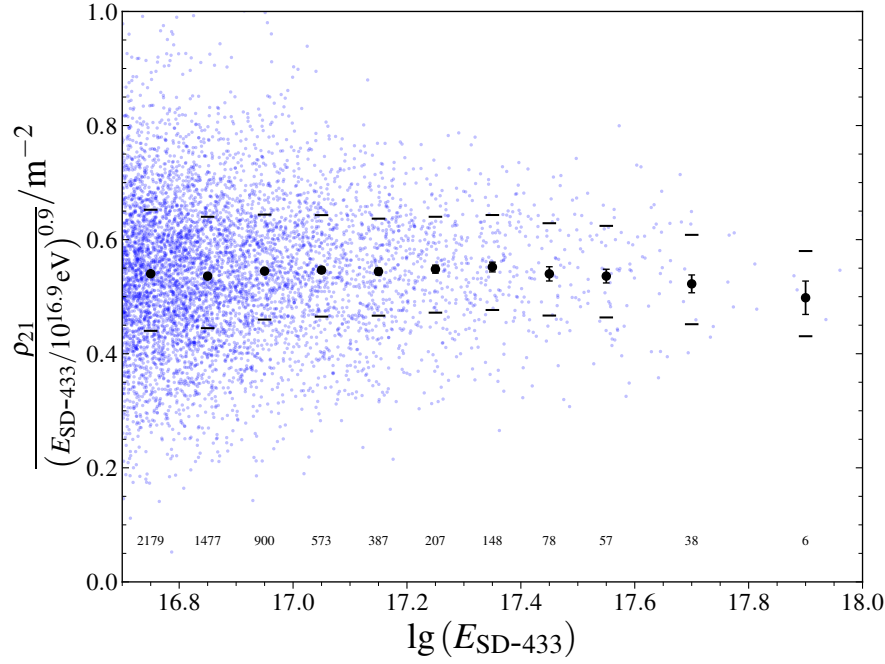


Figure 6.11: Detrended muon-density estimator $\rho_{21}/(E/10^{16.9} \text{ eV})^{0.9}$ as a function of reconstructed SD-433 energy. Event counts are reported per bin at the bottom. The last two bins are twice as wide to ensure sufficient statistics. The horizontal brackets represent the systematic uncertainty in each energy bin, as discussed in section B.7.

of the Heitler-Matthews model [78]. The corresponding representation is shown in Fig. 6.11. We stress that the physical information is carried by ρ_{21} ; the rescaling is introduced purely to tame the steep energy dependence of the muon density and thereby facilitate comparisons across energy bins.

Consistently with this intent, the binned points shown in Fig. 6.11 are constructed by first computing the mean muon density $\langle \rho_{21} \rangle$ in each energy bin and then applying the scaling factor evaluated at the bin centre energy E_c , i.e., $\langle \rho_{21} \rangle / (E_c/E_0)^{0.9}$. This procedure is, in general, not equivalent to taking the mean of the scaled quantity $\langle \rho_{21} / (E/E_0)^{0.9} \rangle$. The systematic uncertainties of ρ_{21} are introduced and briefly described in section B.7.

6.4 Compatibility with the UMD-750

Before turning to the physical interpretation of the UMD-433 measurement, it is worth checking its compatibility with the previous UMD-750 reconstruction in the energy range where both configurations overlap. Such a comparison is particularly valuable because it provides an empirical validation of the stability of the UMD response across two detector layouts, reconstruction chains, and reference observables.

Two complementary checks are shown in Fig. 6.12. The left panel compares the muon density at 450 m for events jointly seen by both arrays and reconstructed independently in each configuration. For the UMD-750, this observable is the standard reference estimator $\rho(450)$. For the UMD-433, we denote by $\tilde{\rho}(450)$ the muon density obtained by evaluating its reconstructed MLDF at the same distance. In this way, both reconstructions can be compared on a common footing at the UMD-750 reference distance. Event-by-event differences are naturally expected, since the two reconstructions are not based on exactly the same set of stations: some stations belong to both arrays, whereas others are excluded because they do not belong to the configuration under consideration. Even so, the average agreement is quite

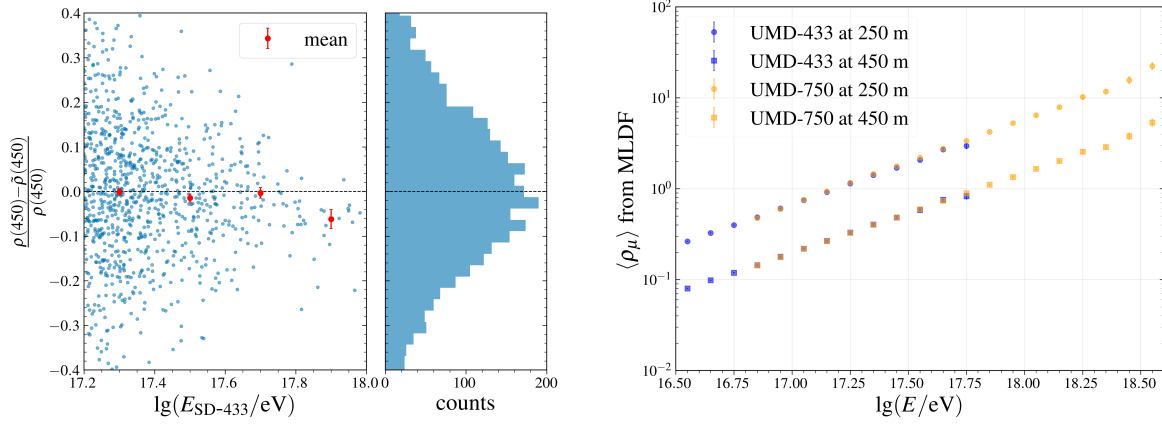


Figure 6.12: Compatibility checks between the UMD-433 and UMD-750 reconstructions for showers with $\theta < 30^\circ$ passing the quality cuts described in Section 6.3. *Left:* event-by-event comparison of the independently reconstructed muon density at 450 m for showers jointly seen by both arrays. The red markers show the mean difference in energy bins. The spread reflects the fact that the two reconstructions do not use exactly the same set of stations. *Right:* mean muon densities reconstructed by UMD-433 and UMD-750 at the reference distances 250 m and 450 m as a function of energy. The agreement is generally very good over a broad common range, with the best consistency observed at 450 m.

good up to $\lg(E/\text{eV}) \simeq 17.7$. Above this energy, the estimator obtained from the UMD-433 reconstruction begins to fall below the one from the UMD-750 reconstruction.

The right panel of Fig. 6.12 provides a broader comparison by showing the mean muon densities reconstructed by both arrays at two different reference distances, 250 m and 450 m. Once again, the agreement is generally good over a wide energy range. At 450 m, the relative difference is typically of order 2%. At 250 m, the discrepancy is somewhat larger, about 4% up to $\lg(E/\text{eV}) \simeq 17.7$, but still within statistical uncertainties. Since 250 m lies closer to the shower axis, where the reconstruction is more sensitive to systematic uncertainties in the core position, the comparison is expected to be less direct. At higher energies, the limited extent of the UMD-433, saturation effects, and the larger spacing of the UMD-750 also make the comparison less straightforward. Above $\lg(E/\text{eV}) \simeq 17.7$, the disagreement becomes too large to be attributed to a mere fluctuation, and the UMD-433 statistics also become very limited.

It is also remarkable that the UMD-750 reconstruction matches the UMD-433 result down to $\lg(E/\text{eV}) \sim 17.0$, even though the full-efficiency region of the UMD-750 is expected to lie at somewhat higher energies. In that sense, the UMD-433 measurement provides an important validation of the muon-content determination in this lower-energy regime. Since the muon content depends steeply on the primary energy, this agreement also supports a good relative alignment between the SD-433 and SD-750 energy scales in the overlapping range. We stress, however, that both checks are restricted to showers with $\theta < 30^\circ$ and passing the quality cuts introduced in Section 6.3.

A final cross-check is shown in Fig. 6.13, where we compare the observable ρ_{35} used in the UMD-750 analyses. The same quality cuts described above were applied. To build an estimator directly comparable to that of UMD-750, we use the UMD-750 attenuation correction to map the UMD-433 measurement into the corresponding reference observable. This test is deliberately demanding for the UMD-433: it requires extrapolating the measurement to zenith angles beyond those used to parametrize the MLDF, and it also probes the muon density at a reference distance that is not optimal for the UMD-433 configuration. It is therefore encouraging that the agreement remains good over a broad energy interval.

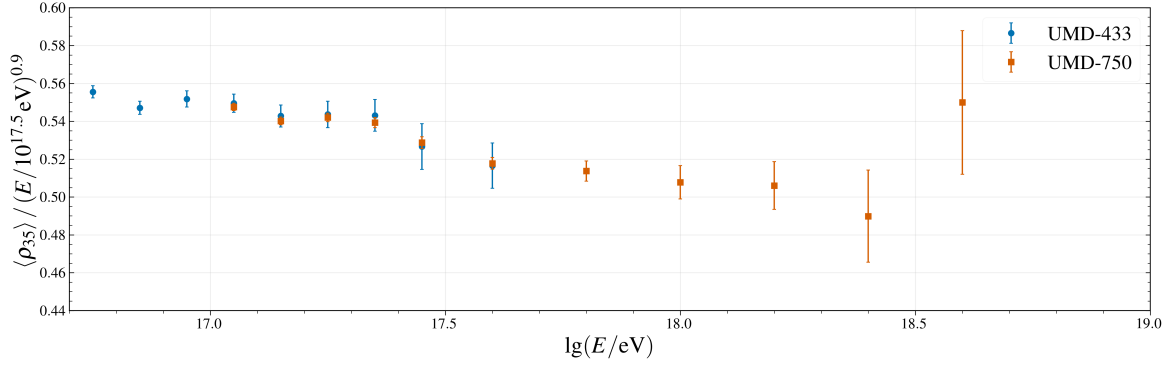


Figure 6.13: Comparison of the scaled mean observable $\langle \rho_{35} \rangle / (E / 10^{17.5} \text{ eV})^{0.9}$ for UMD-433 and UMD-750 as function of energy.

The main conclusion of these compatibility studies is that the UMD provides a stable and consistent measurement of the muon content of extensive air showers across the two detector configurations over a wide common energy range. The highest-energy region, above $\lg(E/\text{eV}) \simeq 17.7$, deserves additional scrutiny in future UMD-433 studies. Additional Phase II statistics may help reduce or clarify the observed difference.

6.5 Simulated baselines

In this section we use the Napoli-Praha library of simulated air showers stored on the Grid [135, 158]. In particular, we employ simulations produced with CORSIKA versions 7.7 and 7.8, thereby covering two sets of hadronic-interaction models. The former includes EPOS-LHC and QGSJET-II.04, which have been widely used in the literature and therefore provide a standard baseline for interpreting muon-number predictions across experiments. In addition, we use the more recent EPOS LHC-R and SIBYLL 2.3e samples from the latest production. The simulation set comprises proton and iron primaries in three energy bins spanning $\lg(E_{\text{MC}}/\text{eV}) = 16.5$ to 18.0.

The detector simulations and reconstructions are performed with the MdSd433SimRec application. The only difference with respect to the workflow described in Section 6.3 is that the corner-clipping correction is computed using the parametrisation derived from simulations [126]. This choice ensures an unbiased station-level estimate of the muon density. All subsequent reconstruction steps follow the same chain as in data, so that the simulated events are processed under conditions that closely emulate the data reconstruction. The CIC procedure described in section B.6 is repeated for all combinations of primary mass and hadronic-interaction model available in the library. For each (model, primary) pair we obtain an attenuation curve $f_{\text{att}}^{(\text{sim})}(\theta)$ and define the corresponding zenith-independent observable

$$\rho_{21}^{(\text{sim})} \equiv \rho^{(\text{sim})}(250) / f_{\text{att}}^{(\text{sim})}(\theta).$$

At this stage, it is useful to distinguish between two closely related summary quantities.

The mean density $\langle \rho_{21} \rangle$ and the mean logarithmic density $\langle \ln \rho_{21} \rangle$ are both proxies for the muon content. We use $\langle \rho_{21} \rangle$ mainly for qualitative visualisation, whereas the observables introduced in Eqs. (6.4) and (6.5) are constructed from $\langle \ln \rho_{21} \rangle$, following the recommendation of Ref. [146]. We parameterize these quantities as

$$\langle \rho_{21} \rangle(E_{\text{MC}}) = a \left(\frac{E_{\text{MC}}}{E_0} \right)^b, \quad (6.29)$$

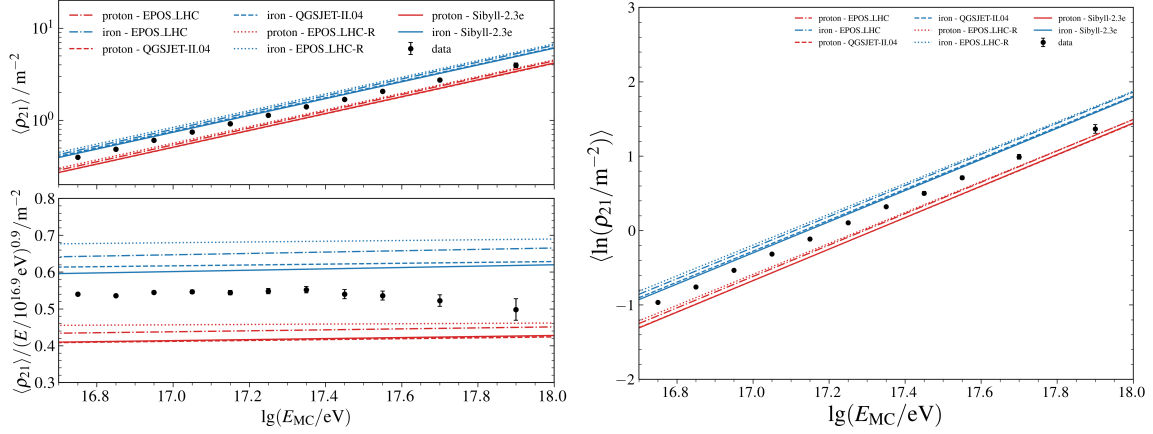


Figure 6.14: Fitted means of the simulated zenith-corrected muon density ρ_{21} for proton (red) and iron (blue) primaries, shown for each hadronic-interaction model (line styles as indicated in the legend). *Left:* mean density $\langle\rho_{21}\rangle$ as a function of $\lg(E_{\text{MC}}/\text{eV})$ (top, logarithmic y -axis), represented by the power-law parameterisation of Eq. (6.29). The lower panel shows the same information after detrending by $(E/E_0)^{0.9}$ with $E_0 = 10^{16.9}$ eV. *Right:* linear parameterisation of the mean logarithmic density $\langle\ln(\rho_{21}/\text{m}^{-2})\rangle$ given in Eq. (6.30).

and

$$\langle\ln\rho_{21}\rangle(E_{\text{MC}}) = m \left[\lg\left(\frac{E_{\text{MC}}}{\text{eV}}\right) - \lg\left(\frac{E_0}{\text{eV}}\right) \right] + c, \quad (6.30)$$

respectively, where $E_0 = 10^{16.9}$ eV is used as a pivot energy. The fitted mean values are shown in Fig. 6.14 for all hadronic-model and primary-mass combinations. The lower panel of the left-hand plot follows the detrending convention of Fig. 6.11: the means and fits are obtained for ρ_{21} , and the rescaling by $(E/E_0)^{0.9}$ is applied using the bin-center energy.

To compare the simulated baselines to data as a function of reconstructed energy, we account for energy-reconstruction effects and for the shape of the energy spectrum through a forward-folding procedure. Because the muon density rises rapidly with true energy, even an unbiased energy estimator induces a non-negligible distortion once convolved with the steeply falling cosmic-ray flux and the detector response [146].

Let $f(E_{\text{MC}})$ denote the simulated expectation for an observable as a function of the true Monte Carlo energy E_{MC} . The expected value of the simulated observable in a reconstructed-energy bin $E^- < E < E^+$ is obtained as

$$f(E_c) = \frac{\int_{E^-}^{E^+} \int_0^\infty f(E_{\text{MC}}) J(E_{\text{MC}}) G(E|E_{\text{MC}}) \varepsilon(E_{\text{MC}}) dE_{\text{MC}} dE}{\int_{E^-}^{E^+} \int_0^\infty J(E_{\text{MC}}) G(E|E_{\text{MC}}) \varepsilon(E_{\text{MC}}) dE_{\text{MC}} dE}, \quad (6.31)$$

where E_c is the bin centre [159]. The weight entering the integrand is often written as

$$w(E_{\text{MC}}) \equiv J(E_{\text{MC}}) G(E|E_{\text{MC}}) \varepsilon(E_{\text{MC}}). \quad (6.32)$$

The inputs to Eq. (6.31) are:

- $J(E_{\text{MC}})$: the cosmic-ray flux used to weight true energies. We use the spectrum reported in Ref. [142].
- $G(E|E_{\text{MC}})$: the SD-433 energy-resolution function, modelled as a Gaussian centred at the expected bias and with an energy-dependent width. The corresponding parametrisation is taken from Ref. [142].

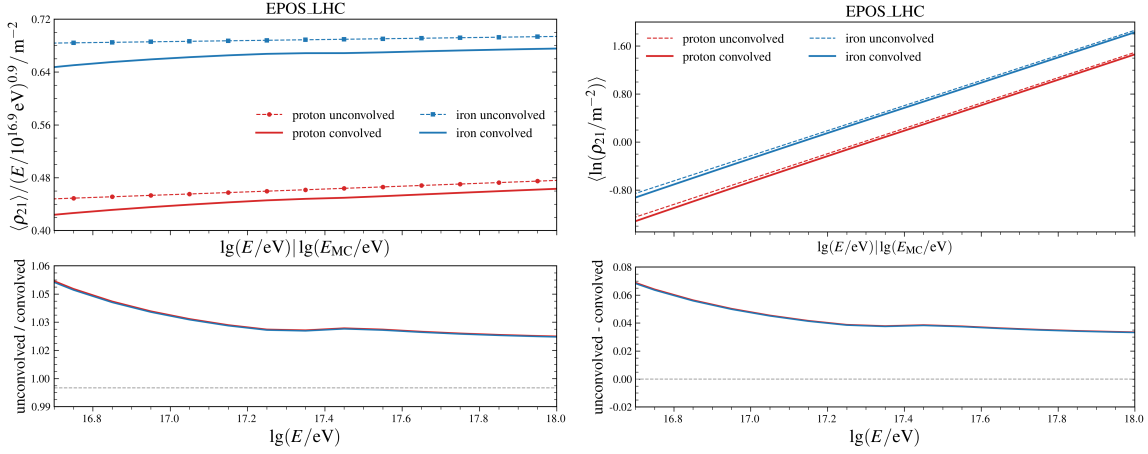


Figure 6.15: Forward-folding (convolution) correction applied to the simulated ρ_{21} expectations for EPOS-LHC. Dashed curves show the unconvolved predictions as a function of true energy E_{MC} , while solid curves show the forward-folded expectations in reconstructed-energy bins, obtained using Eq. (6.31) with the cosmic-ray spectrum, the SD-433 energy-response function, and the trigger efficiency. Red (blue) corresponds to proton (iron) primaries. *Left:* scaled mean density $\langle \rho_{21} \rangle / (E/10^{16.9} \text{ eV})^{0.9}$. *Right:* mean logarithmic density $\langle \ln(\rho_{21}/\text{m}^{-2}) \rangle$. The lower panels quantify the size of the correction: the ratio of the unconvolved to the convolved expectation (left) and the difference unconvolved – convolved (right), where the dashed horizontal lines indicate no correction (unity and zero, respectively).

- $\varepsilon(E_{MC})$: the effective trigger efficiency of the array. We compute it following the procedure described in Ref. [160].

A few remarks are in order. Restricting the analysis to nearly vertical showers ensures full efficiency over the relevant energy range, so the impact of the efficiency term on the folding is negligible. Although the response function $G(E|E_{MC})$ of Ref. [142] was derived up to $\theta = 45^\circ$, we tested the robustness of the folding by varying the resolution by $\pm 15\%$ and found consistent results. Only variations larger than $\pm 30\%$ in the resolution produced noticeable changes in the inferred $\langle \ln A \rangle$ and z-scale values, which we consider unlikely. We also assume the same energy resolution for proton and iron primaries, even though proton-induced showers are expected to have a poorer resolution. The net effect of this approximation should be a slightly lower proton baseline, but it is expected to remain small. The energy bias was also varied by $\pm 5\%$ with no significant impact, and such large shifts are unlikely. Overall, the dominant contribution to the forward-folding correction comes from the assumed spectrum, in particular from the low-energy spectral index, which controls the migration from higher to lower reconstructed-energy bins: the steeper the spectrum, the larger the expected correction in the lowest-energy bins.

Figure 6.15 illustrates the impact of the folding correction (cf. Eq. (6.31)) for the hadronic-interaction model EPOS-LHC. The dashed curves show the raw (unconvolved) Monte Carlo expectations as a function of true energy E_{MC} , whereas the solid curves show the corresponding forward-folded predictions in reconstructed-energy bins. We show the correction for both the mean muon density $\langle \rho_{21} \rangle$ (left) and its logarithmic counterpart $\langle \ln \rho_{21} \rangle$ (right). In both cases, the folding leads to systematically smaller expectations in reconstructed-energy space, as quantified by the ratio (left, bottom) and difference (right, bottom) panels. Since the corrections are very similar for the other hadronic models and primary-mass hypotheses considered, we show EPOS-LHC as a representative example.

The effect of the folding correction on the final composition estimates is shown explicitly in section B.8. There, the same comparisons are repeated using the direct simulation baselines,

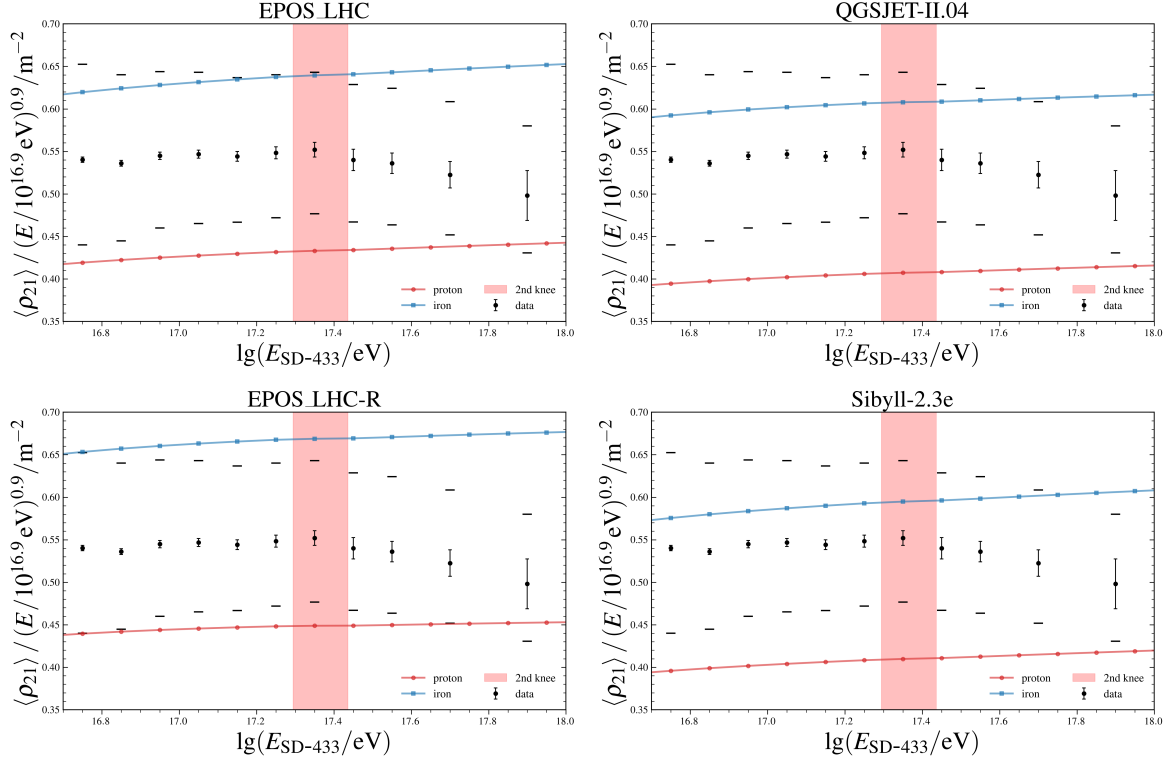


Figure 6.16: Scaled mean muon density $\langle \rho_{21} \rangle / (E/10^{16.9} \text{ eV})^{0.9}$ as a function of reconstructed SD-433 energy for different hadronic-interaction models. Black points show the UMD-433 measurement, while the solid red and blue curves show the forward-folded expectations for proton and iron primaries, respectively, after accounting for the cosmic-ray spectrum, the SD-433 energy response, and the trigger efficiency as described in Eq. (6.31). The vertical red band marks the energy range of the second knee. Short horizontal markers indicate the systematic uncertainty assigned to each datum.

without folding them with the cosmic-ray spectrum and detector response. The correction mostly affects the absolute normalisation of the inferred composition scale, with a slight effect on the slope of the first energy bins. By shifting the proton and iron muon-density baselines lower, it leads to higher nominal $\langle \ln A \rangle$ and z values for the measured UMD-433 densities.

6.6 Muon content around the second knee

Having defined the muon-content observable, demonstrated its compatibility in data with the twin UMD-750 reconstruction, and derived the corresponding simulated baselines, we can now turn to the interpretation of the UMD-433 measurement. Figure 6.16 shows the scaled quantity $\langle \rho_{21} \rangle / (E/10^{16.9} \text{ eV})^{0.9}$ together with the forward-folded proton and iron baselines for the different hadronic-interaction models. The figure is mainly useful as a qualitative guide. Even so, two robust features are apparent. First, over the full energy range considered, the measured muon content lies between the proton and iron expectations and, depending on the hadronic model, is either close to the midpoint or somewhat closer to the iron baseline. Second, the energy evolution shows a clear downward trend beginning immediately above the second-knee region indicated by the red band [142], pointing to a gradual transition towards lighter primaries beyond it. This trend is present independently of the absolute vertical scale assigned by the interaction model.

We quantify the same information in Fig. 6.17, where the detector-level baselines are converted into an estimate of the mean logarithmic mass $\langle \ln A \rangle$. For comparison, we also show the corresponding UMD-750 results and the published expectations inferred from $X_{\max}$ [161]. The decrease of $\langle \ln A \rangle$ observed with UMD-750 is also reproduced by the UMD-433 data, but the absolute scale is systematically lower in our analysis.

Since the previous section showed that the UMD-433 and UMD-750 measurements themselves are compatible, this offset most naturally points to differences in the simulated baselines. In this respect, it is worth noting that the correction introduced by the forward-folding procedure yields similar values to those reported in [126] for the UMD-750, suggesting that the main source of the discrepancy is unlikely to lie there. Instead, the remaining difference most likely originates in the shower library used to derive the baselines. Unfortunately, the library employed in the UMD-750 analysis is no longer available. A less likely, though still possible, explanation is that part of the discrepancy reflects the different radial scales at which the muon density is probed. A fully controlled comparison with UMD-750 would therefore require deriving both detector-level baselines from the same shower set and, ideally, from the same underlying simulated events. Accordingly, the baselines used here should not be interpreted as providing a uniquely correct mapping from muon content to mass; they are the ones obtained with the simulation library adopted in this work and the current UMD-433 reconstruction chain.

Within that caveat, a few conclusions are nevertheless robust. Except for EPOS-LHC-R, which indicates a lighter composition, the other hadronic models considered imply a muon content roughly at the level expected for oxygen up to $10^{17.4}$ eV, and then exhibit a lightening trend. At the same time, the inferred composition remains heavier than that implied by published $X_{\max}$ -based expectations in the same energy interval, with a stronger tension for QGSJET-II.04 than for EPOS-LHC. Although the difference is smaller than that inferred from the UMD-750 interpretation, this should not be read as evidence that the muon deficit disappears. Rather, if the baselines adopted here are approximately adequate, then the discrepancy with $X_{\max}$ in the UMD-433 energy range may be less dramatic than one might infer from some higher-energy muon-based observables alone.

A natural question is whether the observed lightening of $\langle \ln A \rangle$ already reflects the onset of the transition from a predominantly Galactic to a predominantly extragalactic cosmic-ray flux. Such an interpretation is qualitatively plausible. Galactic cosmic rays are generally expected to dominate the flux at low energies, whereas an extragalactic component should become increasingly important at the highest energies, since known Galactic accelerators are not thought to be capable of reaching the most extreme observed energies. In this context, results from the Pierre Auger Observatory have provided important guidance on the transition problem. In particular, the observation of a large-scale dipolar anisotropy above 8×10^{18} eV favours an extragalactic origin in that energy range [162], while searches for large-scale anisotropies above 10^{18} eV already constrain scenarios in which the flux is dominated by stationary Galactic sources densely distributed in the Galactic disc [163]. These results suggest that the transition from Galactic to extragalactic cosmic rays should occur somewhere between $\sim 10^{16}$ and $\sim 10^{18}$ eV, although its precise location remains an open question. In that broader picture, a gradual decrease of the mean mass above the second knee would not be unexpected, even if the interpretation of the absolute $\langle \ln A \rangle$ scale remains subject to the caveats discussed above.

To place the result in a broader context, Fig. 6.18 compares our $\langle \ln A \rangle$ values with a compilation of composition estimates obtained mostly from surface-detector observables. The comparison is adapted from Ref. [164], which compiles results based on Auger measurements of Δ_s , $X_{\max}^\mu$, R_μ , together with corresponding estimates from IceCube and AGASA. A special addition to our figure is the recent TALE result [165], which is based on hybrid

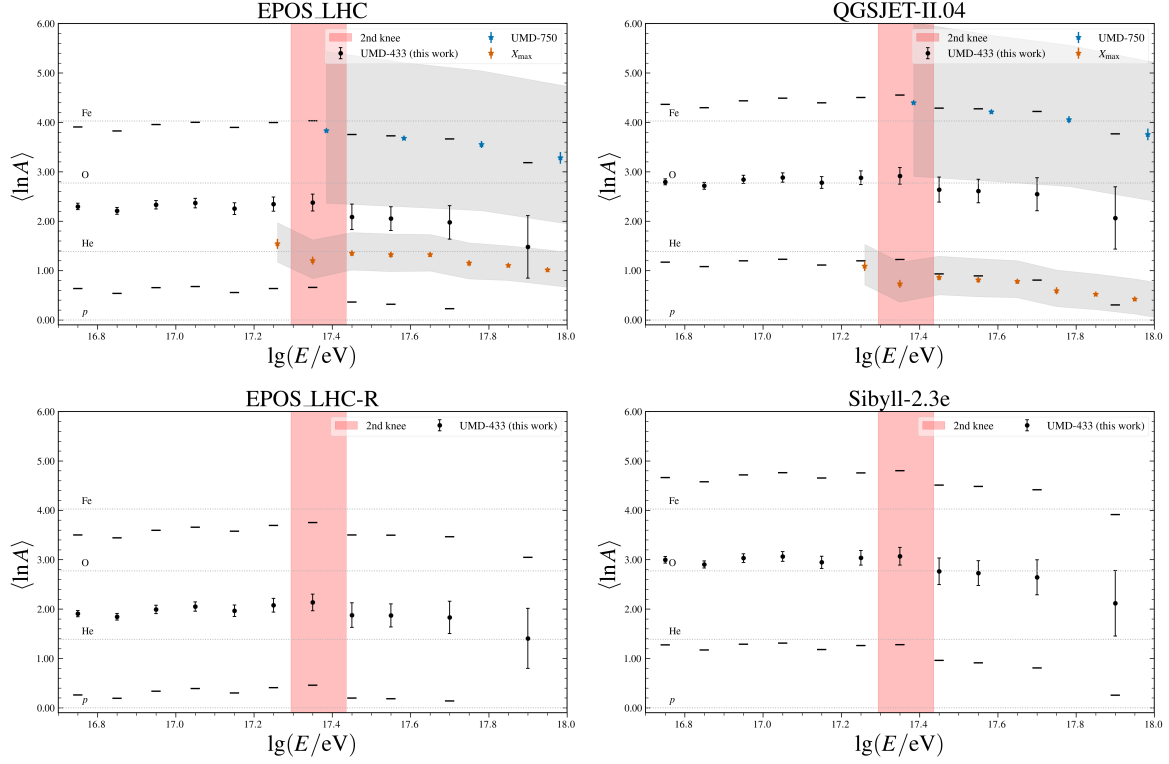


Figure 6.17: Mean logarithmic mass $\langle \ln A \rangle$ inferred from $\langle \ln(\rho_{21}) \rangle$ using the proton and iron detector-level baselines for each hadronic-interaction model. Black points show the UMD-433 result of this work, with horizontal bars indicating the systematic uncertainty. Blue markers show the corresponding UMD-750 result from Ref. [126], while orange markers show composition estimates inferred from published X_{max} measurements [161]. The shaded bands around the external points indicate their quoted systematic uncertainties. The vertical red band marks the energy range of the second knee, and the horizontal dashed reference levels labelled p, He, O, and Fe indicate the corresponding values of $\langle \ln A \rangle$.

events; we include it in the right-hand panel because it is directly relevant to the second-knee region. Seen from this broader perspective, our measurement helps bridge the gap between the IceCube/IceTop energy range and the higher-energy observables measured at Auger, although the agreement is far from perfect. The recent Telescope Array Lower energy Extension (TALE) hybrid result in the QGSJET-II.04 panel is in good agreement with our measurement, but reaches systematically lower $\langle \ln A \rangle$ above our estimate of the second-knee position. Nonetheless, it remains compatible within the quoted systematic uncertainties over the full energy range considered.

Finally, we express the result on the z -scale, shown in Fig. 6.19, to enable a direct comparison with previous muon measurements compiled in Ref. [141]. Since an up-to-date cross-experiment compilation is not yet available for the most recent hadronic models, we restrict this discussion again to EPOS-LHC and QGSJET-II.04. It should be kept in mind that the z -scale is particularly sensitive to the adopted reference energy scale. In our implementation, moving from the Auger reconstructed-energy scale to the reference scale used in the compilation corresponds to an energy rescaling by a factor of 1.054, which already lowers the derived z values by about 20% when expressed on that reference scale. Equivalently, a 1% change in the energy scale induces a change of about 4% in z . This strong sensitivity makes the z -scale somewhat less transparent than the direct $\langle \ln A \rangle$ comparison.

Even so, the overall evolution remains informative. The UMD-433 z values follow the same decreasing trend seen in other muon measurements, again indicating a transition from

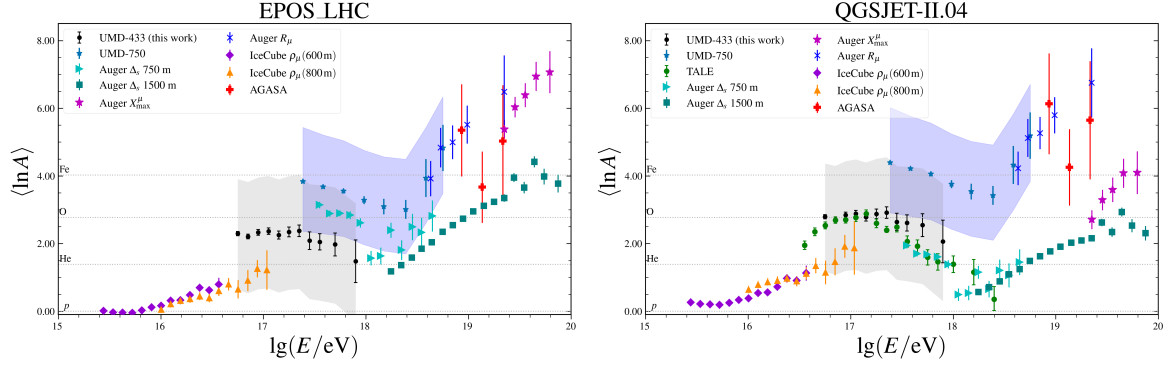


Figure 6.18: Broader comparison of the inferred mean logarithmic mass $\langle \ln A \rangle$ with previous composition estimates. The compilation is adapted from Ref. [164], which summarises results obtained mostly from surface or underground detector observables. In particular, it includes Auger estimates based on Δ_s , $X_{\max}^{\mu}$, R_{μ} , and the UMD, together with corresponding measurements from IceCube, AGASA, and the Telescope Array surface-detector analysis. The recent TALE hybrid result, shown only in the QGSJET-II.04 panel, is added from Ref. [165]. The UMD-433 points from this work are overlaid for comparison. Taken at face value, the figure suggests that the UMD-433 result follows the broad heavy-to-light evolution seen around the second knee and bridges the lower-energy measurements with the UMD-750 regime, albeit with a residual vertical offset.

heavier to lighter primaries above the second knee. In the EPOS-LHC interpretation, the result lies slightly below the GSF expectation [136], whereas for QGSJET-II.04 the offset is small but positive. Of course, the offset with respect to UMD-750 is still present. It is also noteworthy that, on the z -scale, the IceCube points lie above our measurement even though their corresponding $\langle \ln A \rangle$ estimates lie below it in the direct mass comparison. Inspection of the original IceCube publication [166] suggests that the $\langle \ln A \rangle$ values are broadly consistent with the general picture presented in Fig. 6.18. This suggests that the shift most likely originates in the energy correction applied in the z -scale comparison. In particular, the reported energy correction adopted for IceCube acts in the opposite direction to that used for Auger and amounts to nearly 20%. If the sensitivity is comparable to that observed for the UMD, such a shift could easily produce an increase of more than 60% in the reported z values. This substantially modifies the interpretation of the IceCube measurement. More generally, this apparent inconsistency once again illustrates the importance of using clear and transparent reference baselines, both within the collaboration and across experiments.

6.7 Summary

This chapter established the reconstruction and interpretation framework used to measure the muon content of air showers with the UMD-433 in the energy range of tens of PeV and above, with the specific goal of characterising the hadronic background relevant for photon searches. The chapter first motivated the analysis by recalling that the muon content at ground is intrinsically sensitive to the primary mass through the Heitler-Matthews picture, and that this information can be mapped into composition-sensitive observables such as $\langle \ln A \rangle$ and the z -scale once proton and iron detector-level baselines are defined consistently.

We then validated the shape of the mean MLDF measured with the UMD-433, since a reliable composition interpretation requires confidence that the detector reproduces the expected lateral behaviour of the muonic component. In the radial range effectively sampled by the array, the mean profile was shown to be well described by the reduced Greisen form. In particular, a free fit returned a muon-spread factor fully consistent with the value expected

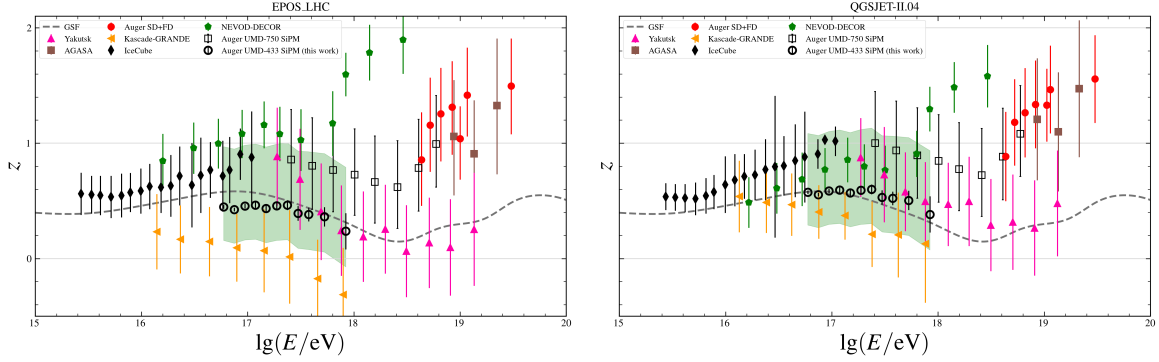


Figure 6.19: Muon content expressed on the z -scale as a function of energy for EPOS-LHC (left) and QGSJET-II.04 (right). The black open circles show the UMD-433 result from this work; the shaded green band indicates its systematic uncertainty. Other measurements are shown as in the compilation of Ref. [141]. The grey dashed curve denotes the Global Spline Fit (GSF) expectation [136].

for the atmospheric depth of the Auger site, $R_\mu \simeq 247$ m, while the fitted slope remained compatible with the canonical Greisen value. After applying the corresponding depth scaling, the comparison with historical measurements was found to be broadly consistent, with especially good agreement with AGASA, thereby reinforcing the interpretation that the UMD captures the expected mean muon-profile geometry under Auger conditions.

We then developed the event-by-event MLDF parametrisation required to extract a stable mass-sensitive observable in the low-multiplicity regime of the UMD-433. Adopting the Greisen-inspired reference-distance form with fixed $R_\mu = 247$ m and $\alpha = 0.75$, we performed a data-driven optimisation of the reference distance and found $r_{\text{ref}} = 250$ m to minimize the propagation of shape uncertainties into the reconstructed muon-density estimator. The residual dependence of the MLDF slope was then captured with a compact first-order parametrisation in energy and zenith angle, $\hat{\beta}(E, \theta)$, whose bias remains small over the fitted phase space and whose resolution improves with energy as the station multiplicity increases. Together, these results provide the practical reconstruction framework needed for the rest of the chapter: a fixed-shape MLDF that is physically motivated, empirically validated, and sufficiently stable to support the definition of $\rho(250)$, the attenuation-corrected estimator ρ_{21} , and the subsequent comparison with detector-level proton and iron baselines.

Using UMD-433 Phase I data, we measured the muon content of extensive air showers in the energy range around the second knee. The event-by-event muon lateral distribution was reconstructed with a fixed-shape MLDF and summarised by its normalisation at 250 m, $\rho(250)$. To enable uniform comparisons across arrival directions, we derived an attenuation correction with a Constant Intensity Cut procedure and defined the zenith-independent estimator $\rho_{21} \equiv \rho(250) / f_{\text{att}}(\theta)$ at the reference angle $\theta_{\text{ref}} = 21^\circ$. Simulated proton and iron baselines were built from the Napoli-Praha library (CORSIKA 7.7: EPOS-LHC, QGSJET-II.04; CORSIKA 7.8: EPOS-LHC-R and SIBYLL 2.3e) and were folded to reconstructed-energy space, accounting for the cosmic-ray spectrum, the SD-433 energy response, and the trigger efficiency.

The measured mean muon content was found to be consistent with that expected for oxygen primaries, or slightly lower, for all hadronic interaction models considered. The $\langle \ln A \rangle$ values derived from the baselines obtained in this work were compared both to the corresponding UMD-750 estimates and to composition expectations inferred from published X_{max} measurements. In the comparison with UMD-750, the UMD-433 values follow the same overall trend but are systematically lower. Since the two reconstructions were shown to

be compatible within their common phase space in data, the most likely origin of this discrepancy lies in the detector-level baselines used to interpret the measurements. Unfortunately, the shower library employed in the UMD-750 analysis is no longer available, so a direct one-to-one comparison is not possible. This emphasises the importance of deriving both baselines from the same shower library and, ideally, from exactly the same simulated events propagated through the detector and reconstruction chains. In contrast, the comparison with X_{max} -based estimates still points to a muon deficit, although in this energy range it appears less pronounced than that suggested by muon observables derived at higher energies.

For a broader comparison, the results were considered both in terms of $\langle \ln A \rangle$ and on the z -scale. In terms of $\langle \ln A \rangle$, the UMD-433 estimates appear to bridge the energy gap between the IceCube results at lower energies and the higher-energy measurements of our Observatory, although our results are slightly higher than would be required for a seamless match. In the QGSJET-II.04 interpretation, our values are systematically higher than the recent hybrid measurements reported by TALE, although still compatible within the quoted systematic uncertainties. However, the baseline issue becomes apparent again when the comparison is recast on the z -scale: in that representation, the offset with respect to IceCube changes sign. This once more highlights how sensitive muon-based interpretations are to the choice of reference baselines. The $\langle \ln A \rangle$ estimates from UMD-433 appear to follow the general composition picture more naturally, whereas on the z -scale the continuation from IceCube to UMD-750 is missed by a significant offset. In the latter representation, our estimates are in good agreement with the GSF model. More work is therefore needed to reconcile the interpretation of these measurements, both within the collaboration and across experiments. Nevertheless, the outlook is encouraging: a consistent measurement of the muon content with the UMD is now available over a wide energy range, even if a fully consistent interpretation with simulations still requires further work.

Taken together, these results provide the first determination of the muon content with the UMD-433 around the second knee using an updated reconstruction chain tailored to Phase I data. Beyond the composition result itself, the chapter delivers the background characterisation required for photon searches in the same energy range.

Photon reconstruction

I have been and still am a seeker, but I have ceased to question stars and books; I have begun to listen to the teaching my blood whispers to me.

Demian, Hermann Hesse

The physical differences between hadronic- and photon-induced showers have already been discussed in Section 3.4. In brief, the predominantly electromagnetic development of photon showers, together with their reduced muonic component, implies observables and reconstruction steps that cannot be transferred one-to-one from the standard hadronic reconstruction sequence. In particular, both the lateral distribution of the ground signal and the energy calibration must be adapted to photon-induced cascades. Moreover, the geometrical reconstruction, which enters all subsequent steps, must be explicitly validated to ensure that the photon reconstruction meets the minimum quality requirements used in later physics analyses.

A first prerequisite is to quantify the trigger efficiency of the SD-433 array for photon primaries. For that reason, we begin with an efficiency study in Section 7.1, closely mirroring the detector-characterisation approach previously performed for hadrons in Chapter 5. In the present case, since no photon event has yet been established experimentally in this energy range, the Monte Carlo simulations are taken as ground truth. This working hypothesis is well motivated: photon-induced showers are dominated by electromagnetic interactions, whose physics is understood more thoroughly than the modelling of the muonic component that drives a significant fraction of the systematic uncertainty in hadronic showers.

The simulation set used throughout this chapter is based on CORSIKA 7.7 with EPOS-LHC as the high-energy hadronic interaction model. While the choice of CORSIKA version is expected to have only a sub-dominant impact for photon-induced showers, whose development is largely electromagnetic and for which most generator updates concern hadronic interactions, we adopt EPOS-LHC consistently for all photon studies. This choice is further motivated by the results of Chapter 6, where the hadronic data are found to be most compatible with EPOS-LHC among the tested interaction models.

Once the efficiency is established, the next step is to define the photon LDF and to parametrise its slope, as presented in Section 7.2. A steeper lateral fall-off than for hadronic showers is expected: in the absence of a sizeable muonic component, the ground signal is largely electromagnetic and therefore decreases more rapidly with distance to the shower axis, whereas muons contribute a comparatively slower-varying component at large core distances. With a photon LDF in place, we then assess the geometrical performance of the reconstruction in Section 7.3. This includes dedicated checks of the angular and core

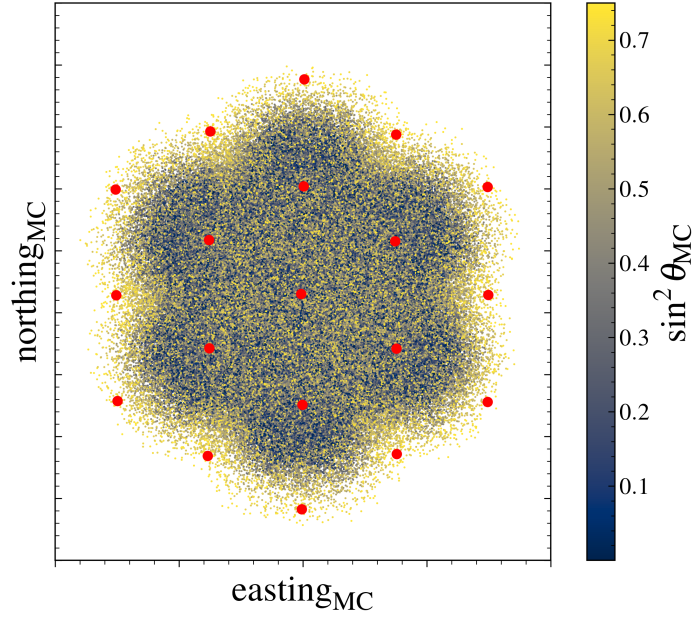


Figure 7.1: Example of the core-randomisation strategy used in the SD-433 photon simulation production. Each point corresponds to a simulated shower core position in the ground plane; colours indicate $\sin^2 \theta_{MC}$. Red markers indicate the station positions used to anchor the randomisation, ensuring that (geometrically) a 6T5 condition can be satisfied.

resolutions, as well as possible residual biases, to determine whether photon-induced events require additional quality criteria or special handling beyond those adopted for the hadronic SD-433 reconstruction.

Finally, in Section 7.4 we complete the photon reconstruction by defining the attenuation correction and the corresponding energy calibration. In particular, establishing a consistent photon energy scale is a prerequisite for any photon search with the array: the definition of E_γ conditions both the exposure calculation and the interpretation of photon-candidate observables, and must therefore be fixed before attempting to derive limits or claims on the photon flux. Most of the reconstruction aspects discussed here were originally developed and validated in the context of Ref. [52]. Since then, both the *Offline* framework and the associated detector simulations have evolved. This chapter therefore builds on that work and extends it by presenting an updated and internally consistent set of photon-reconstruction studies for Phase I with the SD-433.

7.1 Efficiency

The photon reconstruction efficiency is evaluated using dedicated SD-433 detector simulations based on the Napoli–Praha vertical library [135], with isotropic arrival directions and an E^{-1} spectrum. The simulated phase space covers $\theta \in (0^\circ, 60^\circ)$ and three energy intervals in $\lg(E/\text{eV})$ from 16.0 to 17.5. To increase statistics, multiple detector throws are generated for each shower and the core positions are randomised on the SD-433 layout. The randomisation strategy follows the one implemented for other arrays: cores are placed close to a randomly chosen active station such that, from purely geometrical considerations, a 6T5 configuration is possible. Overall, this production amounts to $\mathcal{O}(10^5)$ detector simulations.

Table 7.1: Parameters for the photon efficiency model in Eq. (7.2).

Term	Estimate	Std. error	z-value	95% CI
Intercept	+6.74	0.05	145.9	[6.65, 6.83]
x_E	+8.22	0.06	137.3	[8.11, 8.34]
x_θ	-10.6	0.2	-55.3	[-11.0, -10.3]
x_θ^2	-12.1	0.3	-40.6	[-12.7, -11.5]
$x_E x_\theta$	-6.7	0.2	-40.7	[-7.1, -6.4]

Initially, the reconstruction sequence follows the SD simulation and reconstruction sequences for hadronic showers. At this stage, we only require the reconstruction sequence to retain non-triggered events to allow an unbiased efficiency estimate¹. We define the trigger efficiency as the conditional probability for a photon shower with Monte Carlo truth $(E_{\text{MC}}, \theta_{\text{MC}})$ to produce an SD T3 trigger (see Section 4.2). Formally,

$$\varepsilon(E_{\text{MC}}, \theta_{\text{MC}}) \equiv \mathbb{E}[y \mid E_{\text{MC}}, \theta_{\text{MC}}], \quad y \in \{0, 1\}, \quad (7.1)$$

where $y = 1$ denotes an event that successfully triggered the array and $y = 0$ otherwise. In practice, ε is treated as a probability surface in the $(E_{\text{MC}}, \theta_{\text{MC}})$ plane.

We model the binary response y with a binomial likelihood in an unbinned fit analogous to Section 5.1. For a compact parametrisation of the efficiency surface, we adopt a GLM with a probit link,

$$\Phi^{-1}(\varepsilon) = \beta_0 + \beta_1 x_E + \beta_2 x_\theta + \beta_3 x_\theta^2 + \beta_4 x_E x_\theta, \quad (7.2)$$

where Φ^{-1} is the inverse cumulative distribution function of the standard normal distribution and

$$x_E \equiv \lg\left(\frac{E_{\text{MC}}}{\text{eV}}\right) - 17, \quad x_\theta \equiv \sin^2 \theta_{\text{MC}} - \sin^2(25^\circ). \quad (7.3)$$

The reference angle 25° is used to reduce parameter correlations by centring the zenith predictor around the bulk of the simulated distribution, assuming a photon spectrum proportional to E_γ^{-2} .

The best-fit coefficients for Eq. (7.2) are summarised in Table 7.1. The corresponding efficiency is shown in Fig. 7.2 as (i) ε versus $\lg(E_{\text{MC}}/\text{eV})$ in zenith slices and (ii) as a continuous 2D surface in the $(\lg(E_{\text{MC}}/\text{eV}), \sin^2 \theta_{\text{MC}})$ plane. In the context of a photon search, $\varepsilon(E_{\text{MC}}, \theta_{\text{MC}})$ is interpreted as the probability that a photon shower in this phase space triggers the SD-433 array. The efficiency falls rapidly with increasing zenith angle, as expected for predominantly electromagnetic air showers. As a validation of the parametric description, we inspect the residuals between the binary response and the fitted predictor. The marginalized expected residuals are found to lie within a $\pm 2\%$ band for both predictors, supporting the adequacy of the chosen GLM form in the simulated phase space.

7.2 Lateral Distribution Function

The reconstruction of photon-induced showers with the SD-433 requires a dedicated parametrisation of the LDF shape. Compared to hadronic showers, the photon ground signal is more strongly dominated by the electromagnetic component and therefore attenuates more rapidly with atmospheric depth. As a consequence, the slope parameter β exhibits a stronger coupled dependence on shower size and zenith angle than the analogous slope

¹Event building can be forced even in the absence of an SD/FD T3 by enabling the `ForceEventBuildWithoutT3` option in the `CentralTriggerEventBuilder` configuration.

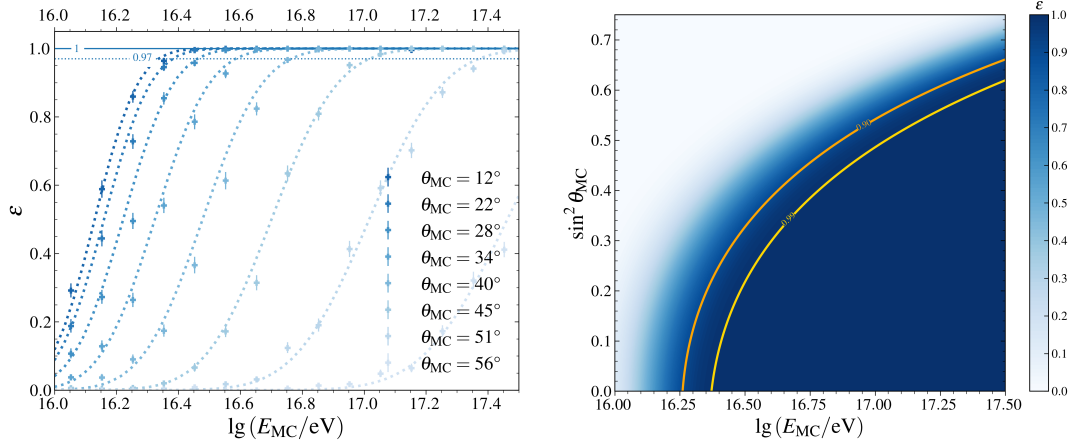


Figure 7.2: SD-433 photon trigger efficiency as a function of Monte Carlo truth. *Left:* binned efficiency estimates ε versus $\lg(E_{\text{MC}}/\text{eV})$ for several fixed zenith angles (markers), compared to the model prediction (dotted curves). *Right:* fitted efficiency surface in the $(\lg(E_{\text{MC}}/\text{eV}), \sin^2 \theta_{\text{MC}})$ plane; the overlaid contours indicate representative high-efficiency thresholds (e.g. 0.90 and 0.99).

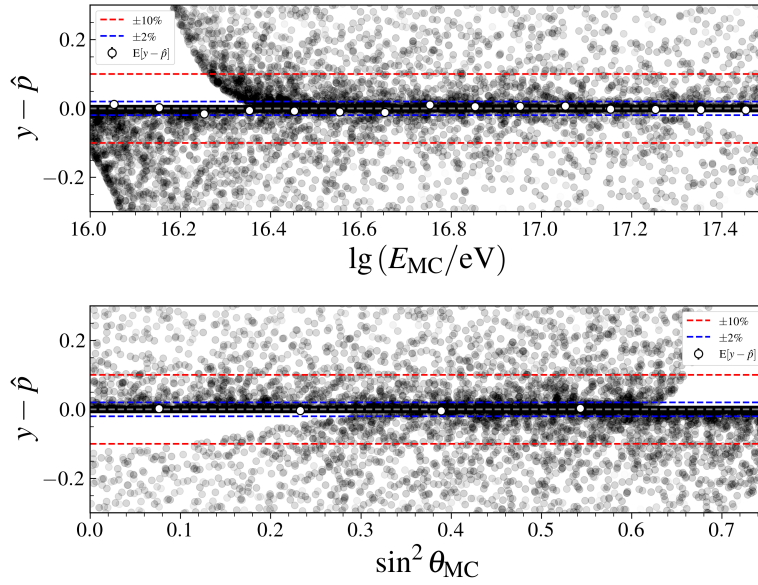


Figure 7.3: Validation of the photon efficiency parametrisation using the residual $y - \hat{p}$, where y is the binary trigger response and $\hat{p} \equiv \varepsilon(E_{\text{MC}}, \theta_{\text{MC}})$ is the fitted probability. Top: residuals versus $\lg(E_{\text{MC}}/\text{eV})$. Bottom: residuals versus $\sin^2 \theta_{\text{MC}}$. Grey points show event-by-event residuals, while white markers indicate the binned expectation $\mathbb{E}[y - \hat{p}]$. The dashed horizontal lines mark the $\pm 2\%$ (blue) and $\pm 10\%$ (red) bands used to assess the marginal agreement.

parametrisations used for hadronic observables (including the MLDF studies discussed in Chapter 6). In practice, this means that a direct fit $\beta = \beta(S(300), \theta)$ is unnecessarily correlated unless an initial attenuation-like correction is introduced first to absorb the dominant zenith trend of the size estimator.

For this reason, the strategy adopted here proceeds in three steps: (i) define and validate an event-by-event photon LDF shape and the reference size $S(300)$, (ii) derive a Monte Carlo-

Table 7.2: Lever-arm criteria adopted for the SD-433 sample used in the photon LDF slope (β) parametrisation. The criterion is considered fulfilled when at least $n_{\min}$ stations are found within $[r_{\min}, r_{\max}]$ and at least one pairwise distance is greater than $d_{\min}$.

r_{ref} [m]	$[r_{\min}, r_{\max}]$ [m]	$n_{\min}$	$d_{\min}$ [m]
300	[100, 400]	2	225
300	[100, 400]	3	200
300	[100, 400]	4	175

based attenuation-like correction for $S(300)$ (used only to stabilise the β parametrisation), and (iii) fit β as a function of the corrected size and zenith angle. This correction is an internal feature of the slope model and should not be confused with the final attenuation correction used later for the photon energy calibration in Section 7.4.

We adopt a modified NKG-like functional form for the photon LDF,

$$f_{\text{LDF}}(r) = S(r_{\text{ref}}) \left[\left(\frac{r}{r_{\text{ref}}} \right) \left(\frac{r + 700 \text{ m}}{r_{\text{ref}} + 700 \text{ m}} \right) \right]^{-\beta}, \quad (7.4)$$

where $S(r_{\text{ref}})$ is the size at the reference distance and β controls the slope. This variant has proven successful in modelling the LDFs of photon primaries [167].

As a working choice, we initially fix $r_{\text{ref}} = 300 \text{ m}$ and denote the corresponding size as $S(300)$. This choice is motivated primarily by reconstruction standardisation: the hadronic SD reconstruction is also formulated around a 300 m reference distance, so adopting the same value for photons simplifies the implementation and comparison of reconstruction applications. At this stage, however, $r_{\text{ref}} = 300 \text{ m}$ should be understood as a *candidate* reference distance that must be validated with the photon simulations, rather than as an a priori optimum.

Before discussing the slope parametrisation itself, it is useful to inspect the geometrical leverage of the sample used for that study. Figure 7.4 shows the reconstructed core positions for the subset of simulated photon events that pass the full selection used for the β parametrisation, including the lever-arm requirement. The detector simulations are generated with randomised core throws around SD stations, which provide broad geometrical coverage of the array neighbourhood. The structured annular patterns visible in the figure, however, are not a direct feature of the core randomisation itself; they arise from the selection of events that satisfy the lever-arm criteria (see Table 7.2). In practice, this selection preferentially retains events with station-to-core distance configurations that provide sufficient radial coverage for stable event-by-event LDF fits, while preserving the geometrical constraints used to emulate a possible 6T5 configuration.

A first validation of the form in Eq. (7.4), together with the provisional choice $r_{\text{ref}} = 300 \text{ m}$, is shown in Fig. 7.5. Using the lever-arm-selected subset (for which β can be fitted with sufficient radial coverage), we inspect the mean standardized residual $\langle (S - \text{LDF}(r)) / \delta S \rangle$ as a function of reconstructed distance to the shower axis. The residual profile remains close to zero over the core-distance range that dominates the fits, while a positive trend appears at large distances ($r \gtrsim 600 \text{ m}$), particularly at lower energies and larger zenith angles. The residuals around $r \simeq 300 \text{ m}$ remain small across the scanned energy and zenith intervals. This supports the use of $S(300)$ as a stable summary of the photon LDF normalisation and shows that adopting the hadronic reference distance is a reasonable choice for the photon reconstruction sequence.

To build a parametrisation for β , we first inspect the event-by-event slopes obtained from free- β LDF fits. The trends in Fig. 7.6 show that β depends on both energy and zenith angle, but the dominant effect is the zenith dependence. Moreover, the apparent zenith

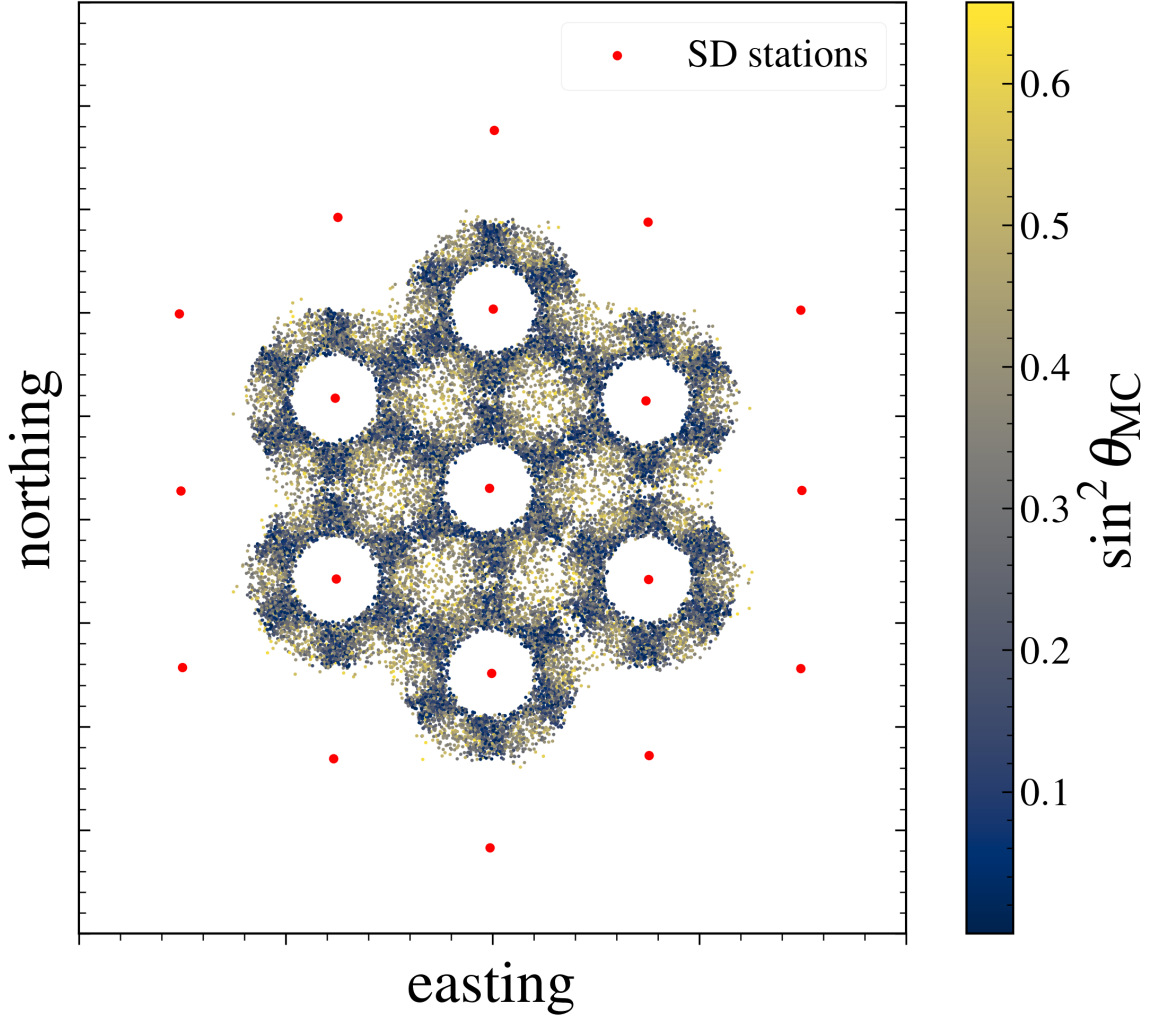


Figure 7.4: Ground distribution of simulated photon shower cores used for the LDF studies. Points show reconstructed core positions in the SD-433 plane and are coloured by $\sin^2 \theta_{MC}$. Red markers indicate SD station positions. The annular patterns around stations arise from the lever-arm selection applied to the randomised core throws used in the detector simulations.

trend becomes stronger at higher energies (bottom row of Fig. 7.6), which indicates that the size and zenith dependences do not factorise trivially in terms of the raw $S(300)$ variable. This feature motivates the intermediate attenuation-like correction introduced below, and the same coupling is also visible in the raw size–energy relation. As shown in Fig. 7.7, the mean $S(300)$ at fixed E_{MC} is strongly separated by zenith-angle bin. Therefore, we introduce a first-order attenuation-like factor $f(\theta)$ to define a corrected size proxy that removes the leading zenith dependence before fitting the slope model.

We model this preliminary attenuation trend using the size-per-energy proxy shown in the left panel of Fig. 7.8. Defining

$$x \equiv \sec \theta - \sec \theta_{\text{ref}}, \quad \theta_{\text{ref}} = 25^\circ, \quad (7.5)$$

the attenuation-like factor is parameterized as

$$f(\theta) = \frac{\tilde{f}(\theta)}{\tilde{f}(\theta_{\text{ref}})}, \quad \tilde{f}(\theta) = \frac{f_0}{1 + \exp\left(\frac{x-b}{c}\right)}. \quad (7.6)$$

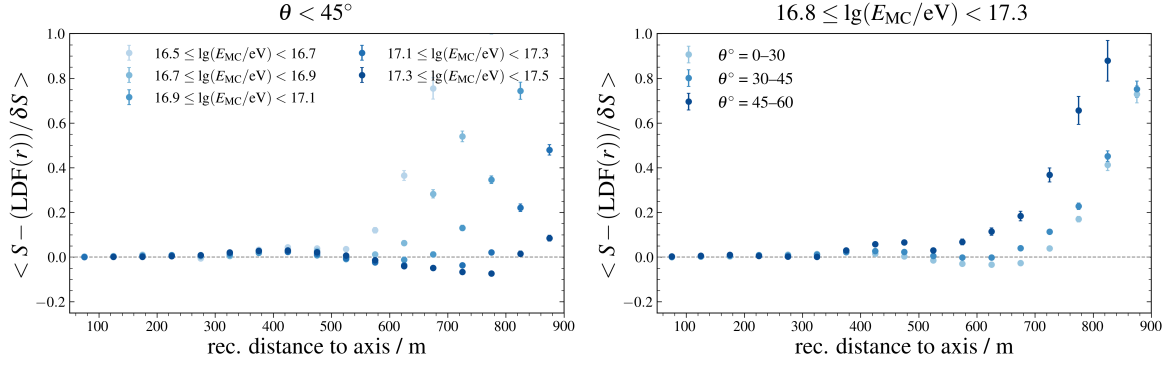


Figure 7.5: Validation of the modified photon LDF using the mean standardized residual $\langle [S - \text{LDF}(r)] / \delta S \rangle$ as a function of reconstructed core distance. *Left:* energy scan for events with $\theta < 45^\circ$. *Right:* zenith scan at fixed energy $16.8 \leq \lg(E_{\text{MC}}/\text{eV}) < 17.3$. The residuals remain close to zero over the central distance range, while a systematic positive trend appears in the far tail, motivating the use of $S(300)$ as the reference size.

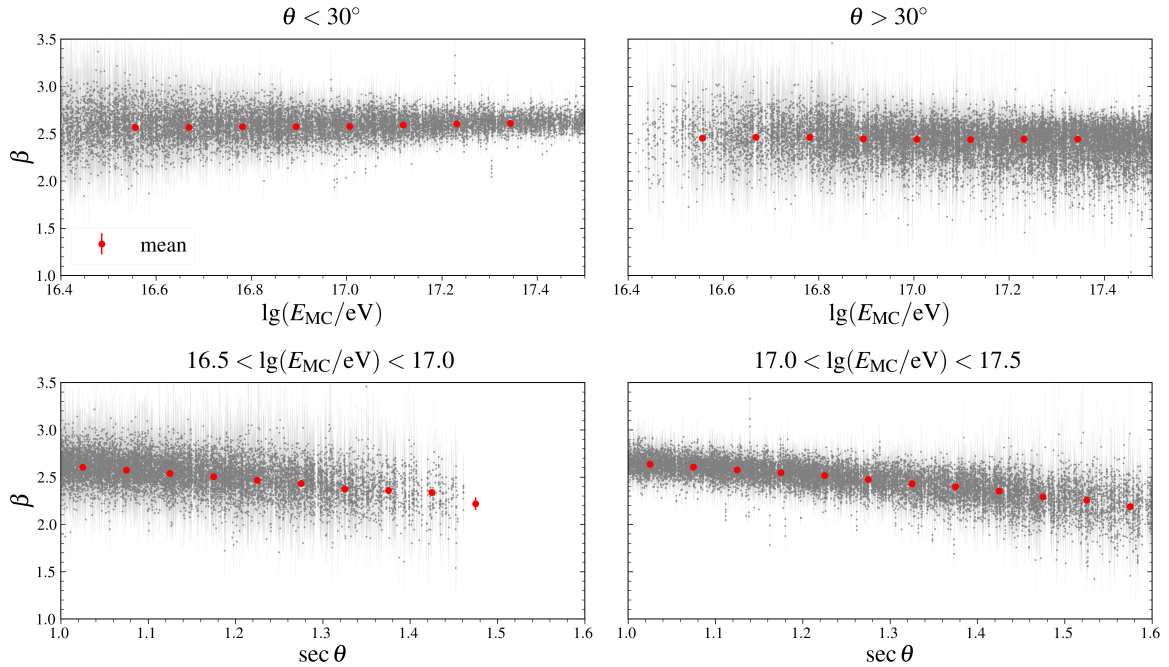


Figure 7.6: Trends of the free- β photon LDF slope obtained from event-by-event fits. Grey points show individual simulated events and red markers indicate weighted binned means. *Top row:* β as a function of $\lg(E_{\text{MC}}/\text{eV})$ for two broad zenith selections. *Bottom row:* β as a function of $\sec \theta$ for two energy ranges. The stronger zenith dependence at larger energies motivates an attenuation-like correction of $S(300)$ before parametrising β .

Here f_0 is the normalisation of the unnormalised trend $\tilde{f}$ (it cancels in the normalised correction f), while b and c control the position and width of the attenuation transition. The fitted parameters are listed in Table 7.3. Applying this correction defines the size proxy

$$S_{25} \equiv \frac{S(300)}{f(\theta)}, \quad (7.7)$$

whose energy dependence becomes approximately zenith-independent, as illustrated in the right panel of Fig. 7.8.

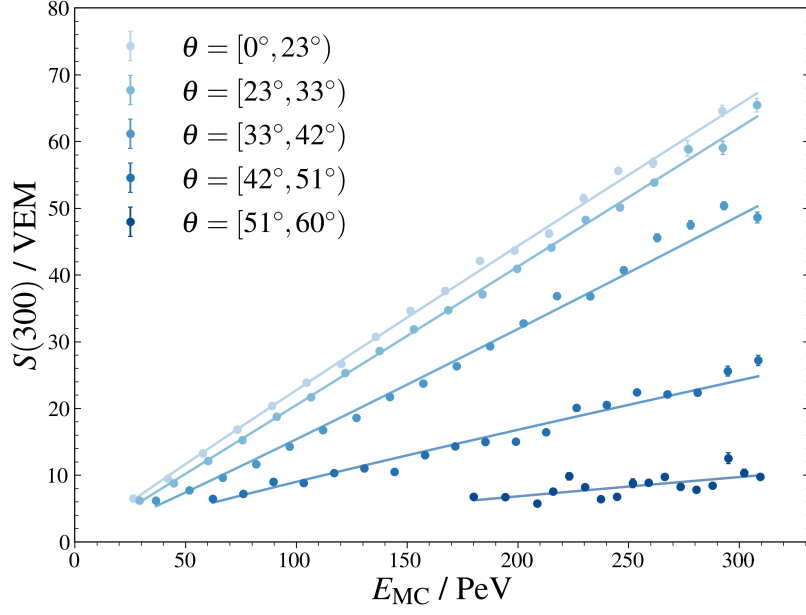


Figure 7.7: Mean photon shower size $S(300)$ as a function of E_{MC} in zenith-angle bins. Markers show binned means and the solid curves are simple empirical trends drawn separately for each zenith bin. The strong zenith separation at fixed energy motivates the introduction of an attenuation-like correction prior to the β parametrisation.

Table 7.3: Fit parameters of the preliminary attenuation-like model used to define S_{25} in Eq. (7.6). Uncertainties are statistical (fit covariance), quoted with one significant digit; parameter values are rounded accordingly.

Parameter	Estimate	Std. error
f_0	2.680	0.006
b	0.205	0.002
c	0.158	0.002

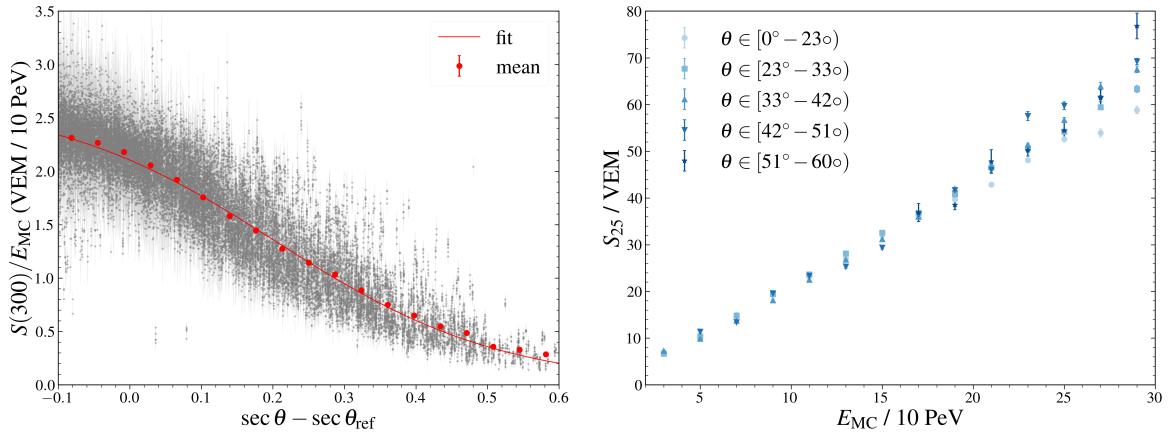


Figure 7.8: *Left:* mean trend of $S(300)/E_{MC}$ as a function of $x = \sec \theta - \sec \theta_{ref}$ (red markers), together with the sigmoid fit of Eq. (7.6) (red curve). Grey points show individual simulated events. *Right:* corrected size proxy $S_{25} = S(300)/f(\theta)$ as a function of E_{MC} in zenith bins. The strong zenith separation seen in Fig. 7.7 is largely removed, which stabilises the subsequent β parametrisation.

Table 7.4: Parameters of the photon slope model in Eq. (7.8) and Eq. (7.9).

Parameter	Estimate	Std. error
a_{00}	2.581	0.002
a_{01}	-0.61	0.01
a_{02}	-0.84	0.03
a_{10}	0.08	0.01
a_{11}	0.40	0.04
a_{12}	0.10	0.03

We then model the photon slope as

$$\beta(S(300), \theta) = \beta_0(\theta) + \beta_1(\theta) \lg\left(\frac{S(300)}{f(\theta) 30 \text{ VEM}}\right), \quad (7.8)$$

with $x \equiv \sec \theta - \sec \theta_{\text{ref}}$ and $\theta_{\text{ref}} = 25^\circ$. In the fitted range, the zenith dependences are well described by second-order polynomials in x ,

$$\beta_0(\theta) = a_{00} + a_{01}x + a_{02}x^2, \quad \beta_1(\theta) = a_{10} + a_{11}x + a_{12}x^2. \quad (7.9)$$

The fitted parameters are summarised in Table 7.4.

The performance of the parametrisation is shown in Fig. 7.9. The left panel indicates that the mean relative bias $\langle(\beta_{\text{pred}} - \beta)/\beta\rangle$ remains consistent with zero across the studied range of $\lg(S(300)/\text{VEM})$, with only small fluctuations at the lowest-statistics edges. The right panel shows the corresponding dispersion as a function of size for cumulative zenith selections. As expected, the relative spread decreases with increasing shower size, reflecting the improved radial leverage and signal statistics in larger events. For the full 0° – 60° sample, the envelope is well described by an exponential form

$$\sigma_\beta \equiv \sigma\left(\frac{\beta_{\text{pred}} - \beta}{\beta}\right) = A_\beta \exp(B_\beta s), \quad s \equiv \lg\left(\frac{S(300)}{\text{VEM}}\right), \quad (7.10)$$

with $A_\beta = 0.183 \pm 0.002$ and $B_\beta = -1.01 \pm 0.01$. These results support the use of Eq. (7.8) as the nominal photon-slope parametrisation for the subsequent reconstruction studies.

7.3 Geometry reconstruction

With the photon LDF and the fixed-slope parametrisation $\hat{\beta}(S(300), \theta)$ defined in Section 7.2, we now run a dedicated reconstruction production for the simulated photon sample in which the LDF slope is no longer fitted event-by-event, but fixed from the parametrisation. This configuration provides the appropriate setup to evaluate the geometrical performance of the photon reconstruction sequence under the same conditions that will be used in subsequent calibration and search studies.

In this section we compare reconstructed and Monte Carlo truth quantities to assess the performance of (i) the shower axis (arrival direction) and (ii) the shower core position. The former is particularly relevant for directional photon searches, because any excess or source association requires a reliable determination of the arrival direction. This is especially important in multimessenger studies, where the angular uncertainty directly enters the cross-correlation with external source catalogues or transient alerts.

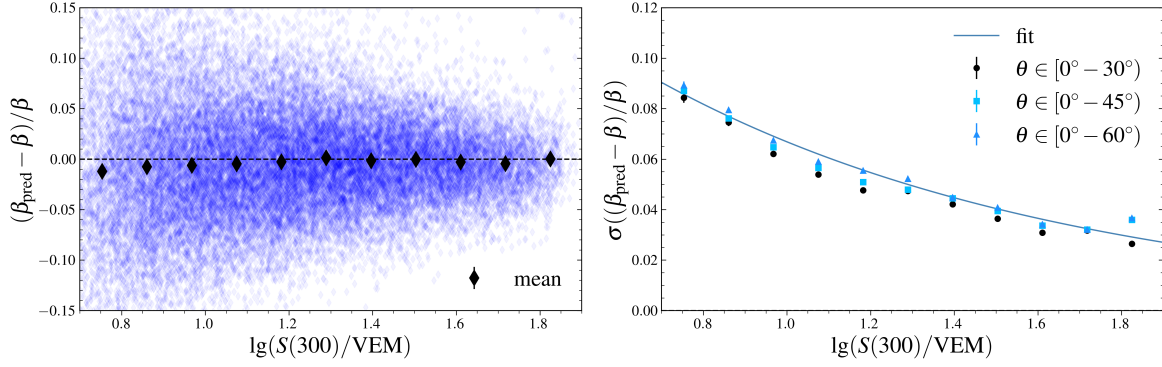


Figure 7.9: Validation of the photon β parametrisation. *Left:* event-by-event relative residual $(\beta_{\text{pred}} - \beta)/\beta$ (blue points) and binned means (black markers) as a function of $\lg(S(300)/\text{VEM})$. *Right:* dispersion $\sigma((\beta_{\text{pred}} - \beta)/\beta)$ versus $\lg(S(300)/\text{VEM})$ for cumulative zenith selections, together with the exponential fit used to summarise the size dependence of the relative slope resolution. The bias remains compatible with zero, while the resolution improves monotonically for larger shower sizes.

7.3.1 Axis reconstruction

As discussed in Section 4.3.2, the reconstruction of the shower axis is obtained from the station trigger times and the relative station positions. In this respect, the axis reconstruction for photon-induced showers does not differ from the standard hadronic SD reconstruction: the timing fit is the same, and no photon-specific LDF feature is required at this stage. Any difference in performance therefore arises from the different station multiplicities and signal patterns of photon showers, rather than from a different axis-fit formalism.

To quantify the angular performance, we define the angular resolution σ_α as the 68th percentile of the opening-angle distribution between the Monte Carlo and reconstructed axes,

$$\sigma_\alpha \equiv P_{68}(\alpha), \quad \alpha \equiv \arccos(\hat{u}_{\text{MC}} \cdot \hat{u}_{\text{rec}}), \quad (7.11)$$

where $\hat{u}_{\text{MC}}$ and $\hat{u}_{\text{rec}}$ are the unit vectors of the true and reconstructed shower directions, respectively [168].

The results are shown in Fig. 7.10. The angular resolution improves monotonically with increasing E_{MC} and with decreasing θ_{MC} , as expected from the larger number of triggered stations and improved timing leverage in better-sampled events. Over the phase space relevant for the present study, the typical angular resolution is approximately 0.7° – 1.5° , with the best performance reached at high energy and low zenith angle.

A similar trend is observed when the resolution is inspected as a function of the simulated distance between the shower core and the hottest station, r_{hottest} : events with more favourable geometries exhibit better directional reconstruction. Note that the shaded region corresponds to events in which the hottest station falls within the CORSIKA inner radius. In these cases, the particles reaching the WCDs are subject to more severe thinning/unthinning effects, and the corresponding results are therefore less reliable.

Overall, the axis reconstruction remains well behaved across the studied phase space, with the most stable performance found at low-to-intermediate zenith angles. This supports the use of the reconstructed arrival direction in photon-candidate analyses and source-oriented studies. In particular, the observed behaviour indicates that the directional reconstruction is not a limiting factor of the photon sequence in the angular range later retained for the energy calibration and flux studies.

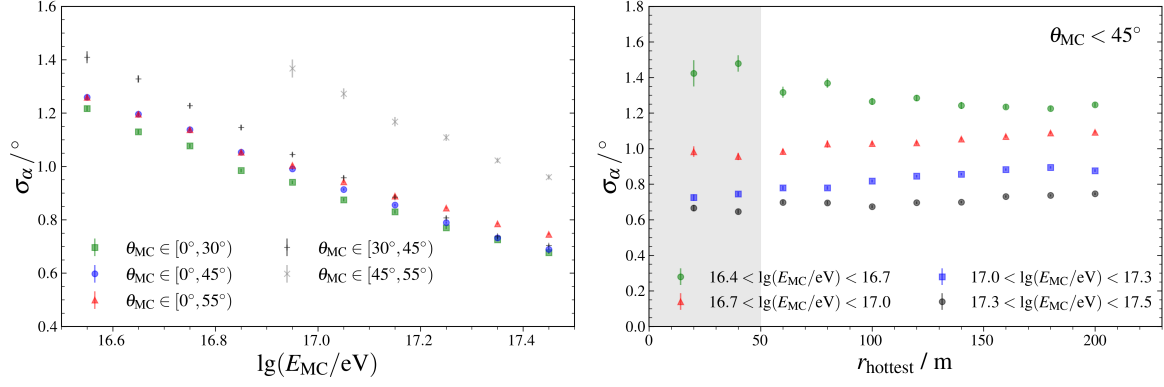


Figure 7.10: Angular resolution $\sigma_\alpha \equiv P_{68}(\alpha)$ of the photon reconstruction. *Left:* σ_α versus $\lg(E_{MC}/\text{eV})$ for different zenith-angle selections (including both disjoint and cumulative ranges). *Right:* σ_α versus the simulated core–hottest-station distance r_{hottest} for energy intervals, shown for the restricted sample $\theta_{MC} < 45^\circ$. The grey shaded region marks events with r_{hottest} inside the CORSIKA inner radius, where thinning/unthinning effects make the result less reliable.

7.3.2 Core reconstruction

Unlike the axis fit, the core reconstruction depends directly on the adopted photon LDF parametrisation. The quality of the reconstructed core position therefore provides an important validation of the fixed- β model introduced in Section 7.2. In particular, residual biases in the core can signal limitations of the assumed LDF shape at larger zenith angles, where shower-front and signal asymmetries become more relevant.

We define the core resolution from the distance between the Monte Carlo and reconstructed core positions in the shower plane [168],

$$\sigma_{\text{core}} \equiv P_{68}(r_{\text{core}}), \quad r_{\text{core}} \equiv |\vec{r}_{\text{MC}} - \vec{r}_{\text{rec}}|. \quad (7.12)$$

The resulting performance is shown in Fig. 7.11. As for the axis reconstruction, the core resolution improves with increasing E_{MC} and decreasing θ_{MC} . The overall event sample exhibits a characteristic resolution of order a few tens of meters; for the sample illustrated in Fig. 7.12, the global containment values are $r_{68} \simeq 18.4$ m and $r_{95} \simeq 42.9$ m.

A similar dependence is seen when the core resolution is plotted versus r_{hottest} (right panel of Fig. 7.11, shown for $\theta_{MC} < 45^\circ$): events with larger core–hottest-station separations outside the CORSIKA inner-radius region are reconstructed more accurately. As in the angular-resolution case, the shaded low- r_{hottest} region should be interpreted with caution because of enhanced thinning/unthinning artefacts.

A first inspection of the Cartesian residuals in detector coordinates may suggest an approximately symmetric behaviour, with no obvious global bias when projected on fixed ground axes (see Fig. 7.12). However, this representation is not optimal for diagnosing shower-related asymmetries, because it does not distinguish between the longitudinal (upstream/downstream) and transverse directions relative to the shower axis projected onto the ground.

The colour scale in Fig. 7.12 provides an additional clue: events with larger $\sin^2 \theta_{MC}$ (more inclined showers) populate a visibly broader halo in the residual plane, while more vertical events remain more concentrated around the centre. This makes the zenith-angle dependence of the core resolution apparent already in the 2D residual map, and is fully consistent with the trends quantified in Fig. 7.11.

To probe directional effects explicitly, we decompose the core residual into components parallel and perpendicular to the shower-axis projection on the ground. In the bias-component study (Fig. 7.13) we use the sign convention shown in the figure labels, i.e.

projections of $\vec{r}_{\text{rec}} - \vec{r}_{\text{MC}}$ onto the unit vectors $\hat{r}_{\parallel}$ and $\hat{r}_{\perp}$. The corresponding mean residuals reveal a clear non-zero bias along the longitudinal (upstream/downstream) direction, while the transverse component remains centred around zero within fluctuations.

More specifically, the perpendicular component $\langle (\vec{r}_{\text{rec}} - \vec{r}_{\text{MC}}) \cdot \hat{r}_{\perp} \rangle$ is consistent with zero across ϕ_{MC} , with fluctuations of order a few tenths of a metre and at most ~ 1 m depending on the binning and available statistics. By contrast, the parallel component $\langle (\vec{r}_{\text{rec}} - \vec{r}_{\text{MC}}) \cdot \hat{r}_{\parallel} \rangle$ shows a robust positive offset of several meters, with a much stronger dependence on zenith angle than on azimuth. In the two broad zenith-proxy intervals shown, the mean parallel bias is approximately at the level of ~ 5 m for $\sin^2 \theta_{\text{MC}} < 0.25$ and ~ 15 – 17 m for $0.25 < \sin^2 \theta_{\text{MC}} < 0.5$, with only modest modulation versus ϕ_{MC} .

The most plausible origin of this behaviour is a residual signal asymmetry already present at the LDF-fit stage, which is not fully captured by the current azimuthally symmetric parametrisation. A natural interpretation is an early–late (upstream/downstream) effect associated with the strong attenuation of the electromagnetic component along the shower-axis projection on the ground. In that case, stations at similar core distance but different azimuthal positions would sample different effective attenuation conditions and thus different signal amplitudes. Such asymmetries are expected to be more pronounced for photon-induced showers, whose ground signal is predominantly electromagnetic, than for hadronic showers with a larger muonic contribution. A more detailed future reconstruction should therefore consider an explicitly asymmetric (or direction-dependent) correction at the core/LDF stage.

From the perspective of defining a robust phase space for photon searches, these results identify $\theta < 45^\circ$ as the natural conservative reference range. While the axis reconstruction remains acceptable over a broad angular interval, the core performance, both in terms of resolution and longitudinal bias, degrades progressively toward larger zenith angles. The range $\theta < 45^\circ$ therefore provides the most stable and internally consistent baseline for the Phase I reconstruction studies, and this is also the region in which the later photon energy-calibration diagnostics are best behaved.

At the same time, the present geometry results do not suggest a catastrophic loss of performance immediately above 45° . If the priority is instead to maximise the available exposure, the reconstruction can plausibly be extended up to $\sim 52^\circ$, provided that any highly inclined photon candidate is subsequently subjected to a dedicated event-by-event review. Such an extension is particularly reasonable because, as will be discussed in Chapter 8, the photon-hadron discrimination performance is largely insensitive to the reconstructed quantities themselves.

7.4 Energy calibration

With the photon LDF and the corresponding geometrical reconstruction validated in Section 7.3, the final step of the photon reconstruction sequence is the definition of an energy estimator. As in the hadronic SD calibration (see the introduction for Chapter 5), the goal is to factorise the raw size observable into (i) an attenuation correction and (ii) a size–energy scaling law. For photon-induced showers, however, both ingredients must be modified with respect to the standard hadronic prescription. In particular, the attenuation trend is described with a different functional form, and the scaling exponent is not kept constant but allowed to vary with zenith angle [169]. The intuition behind this choice can already be anticipated from Section 3.1.1, where the single-photon cascade solution is discussed.

We model the photon energy estimator through the attenuation-corrected size

$$\frac{E_{\text{MC}}}{100 \text{ PeV}} = S_{25}^{\alpha(\theta)}, \quad S_{25} \equiv \frac{S(300)}{f_{\text{att}}(x)}, \quad x \equiv \sec \theta - \sec \theta_{\text{ref}}, \quad (7.13)$$

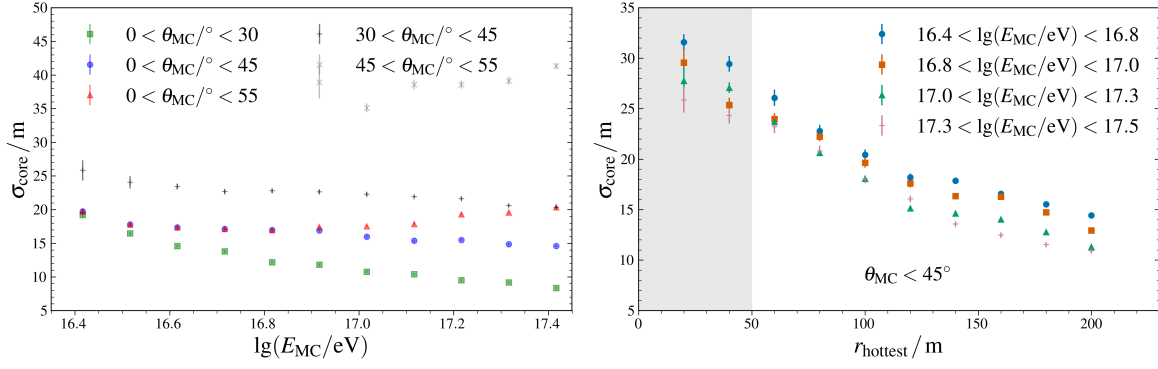


Figure 7.11: Core-position resolution for reconstructed photon showers, defined as $\sigma_{\text{core}} = P_{68}(r_{\text{core}})$ with $r_{\text{core}} = |\vec{r}_{\text{MC}} - \vec{r}_{\text{rec}}|$. *Left:* σ_{core} versus $\lg(E_{\text{MC}} / \text{eV})$ for several zenith-angle selections (disjoint and cumulative ranges). *Right:* σ_{core} versus the simulated core–hottest-station distance r_{hottest} for energy intervals, shown for $\theta_{\text{MC}} < 45^\circ$. The grey shaded region marks events with the hottest station inside the CORSIKA inner radius, where stronger thinning/unthinning effects make the reconstruction diagnostics less reliable.

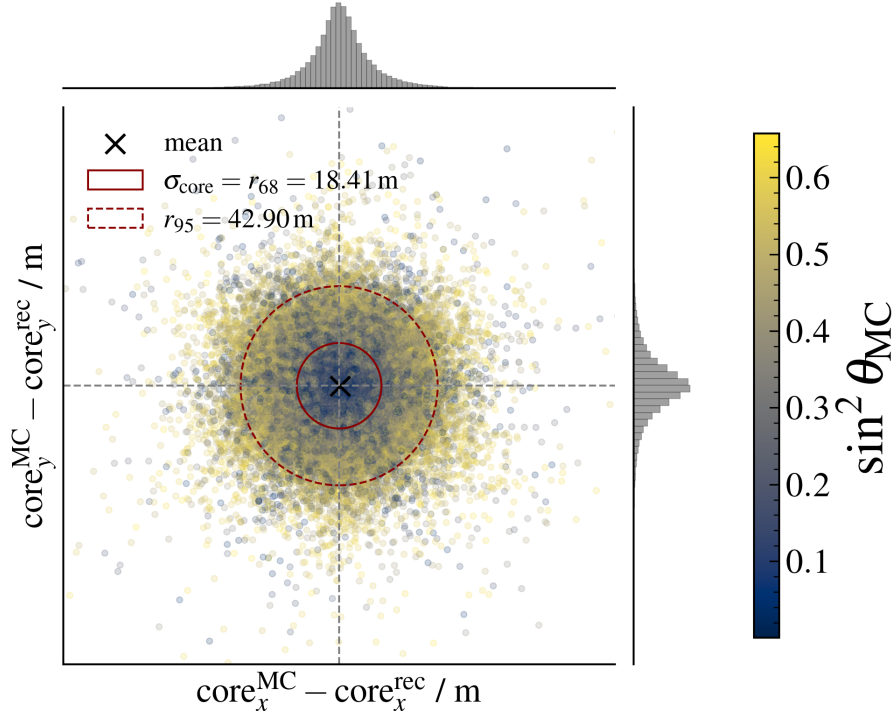


Figure 7.12: Distribution of core residuals in detector coordinates, shown in the ground plane as $(x_{\text{MC}} - x_{\text{rec}}, y_{\text{MC}} - y_{\text{rec}})$. The marker indicates the mean residual, and the circles show the radial containment scales r_{68} (solid) and r_{95} (dashed), with the values quoted in the legend. Marginal histograms along the top and right summarise the projected residual distributions. The colour coding of events follows $\sin^2 \theta_{\text{MC}}$.

with $\theta_{\text{ref}} = 25^\circ$. Unlike the standard hadronic calibration, we model the attenuation function $f_{\text{att}}(x)$ explicitly with a *Gaisser–Hillas* form (normalised at the reference zenith),

$$f_{\text{att}}(x) = \frac{\text{GH}(x)}{\text{GH}(0)}, \quad \text{GH}(x) = g_0 \left(1 + \frac{x - g_2}{g_1} \right)^{g_1/g_3} \exp\left(-\frac{x - g_2}{g_3} \right), \quad (7.14)$$

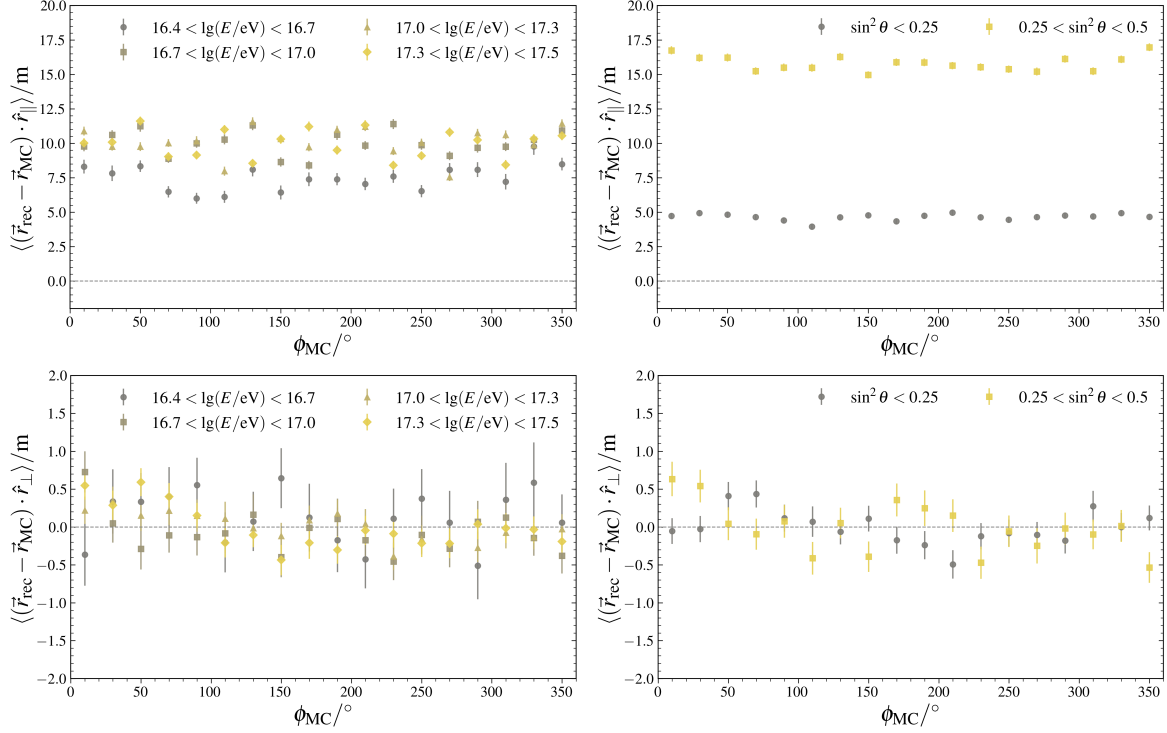


Figure 7.13: Mean projected core-residual components using the sign convention shown in the axis labels, i.e. projections of $\vec{r}_{\text{rec}} - \vec{r}_{\text{MC}}$ onto directions parallel ($\hat{r}_{\parallel}$) and perpendicular ($\hat{r}_{\perp}$) to the shower-axis projection on the ground. *Top row:* longitudinal/upstream-downstream component $\langle (\vec{r}_{\text{rec}} - \vec{r}_{\text{MC}}) \cdot \hat{r}_{\parallel} \rangle$. *Bottom row:* transverse component $\langle (\vec{r}_{\text{rec}} - \vec{r}_{\text{MC}}) \cdot \hat{r}_{\perp} \rangle$. In each pair, the left panel shows the dependence on ϕ_{MC} in energy bins, while the right panel shows the same quantity versus ϕ_{MC} for two broad zenith-proxy intervals in $\sin^2 \theta_{\text{MC}}$.

where the normalisation by $\text{GH}(0)$ guarantees $f_{\text{att}}(0) = 1$, i.e. $f_{\text{att}}(\theta_{\text{ref}}) = 1$.² This form provides a better description of the stronger zenith dependence induced by the predominantly electromagnetic signal of photon showers. In addition, the scaling index is explicitly zenith-dependent and is parametrised as

$$\alpha(\theta) = a_0 + a_1 \cos^2 \theta + a_2 (\cos^2 \theta)^2 + a_3 (\cos^2 \theta)^3, \quad (7.15)$$

thereby allowing the size-energy relation to deform smoothly across the angular range considered.

We organise the calibration procedure in two complementary stages: a sequence of *binned fits* used to construct stable intermediate estimates, followed by a *global unbinned fit* that provides the nominal parameter set. The role of the binned stage is not only diagnostic: it also establishes a clear correspondence between the observables and the subsets of parameters being constrained before the final joint optimisation.

To make this explicit, we first rewrite Eq. (7.13) in logarithmic form,

$$\lg S(300) = \alpha'(\theta) \lg \left(\frac{E_{\text{MC}}}{100 \text{ PeV}} \right) + \lg f_{\text{att}}(x), \quad \alpha'(\theta) \equiv \frac{1}{\alpha(\theta)}. \quad (7.16)$$

In the binned-fit strategy, an initial estimate of the slope term $\alpha'(\theta)$ is obtained from binned trends of $\lg S(300)$ versus $\lg(E_{\text{MC}}/100 \text{ PeV})$. Using this intermediate estimate, we then isolate the attenuation dependence and fit the Gaisser-Hillas attenuation function $f_{\text{att}}(x)$

²Equivalently, since $x = \sec \theta - \sec \theta_{\text{ref}}$, this corresponds to normalising the attenuation curve at $\theta = \theta_{\text{ref}}$.

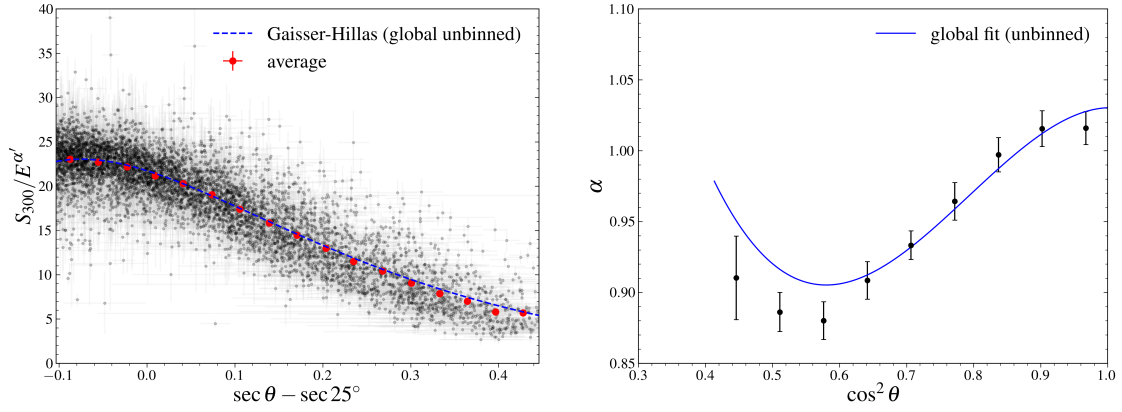


Figure 7.14: Compact summary of the photon energy calibration obtained with the staged fitting procedure. *Left:* attenuation diagnostic proxy $S(300)/E_{MC}^{\alpha'}$ as a function of $x = \sec \theta - \sec 25^\circ$, with event-by-event values shown in grey, bin-by-bin averages as red markers, and the Gaisser-Hillas attenuation model from the global unbinned fit as a blue dashed curve. *Right:* fitted values of the zenith-dependent exponent α as a function of $\cos^2 \theta$, compared with the cubic polynomial from the global unbinned fit.

from the proxy $S(300)/E_{MC}^{\alpha'}$, thereby obtaining the parameters $\{g_0, g_1, g_2, g_3\}$. With the attenuation correction fixed, we define $S_{25} = S(300)/f_{att}(x)$ and fit the corrected size-energy relation to determine the zenith-dependent exponent parameters $\{a_0, a_1, a_2, a_3\}$ in Eq. (7.15). From these binned-fit results we define the seeds for the nominal global unbinned fit, in which all calibration parameters are allowed to vary simultaneously. In this final step, the attenuation and exponent parameters are jointly refined, yielding the self-consistent calibration used in the reconstruction production. This staged procedure improves numerical stability and makes the origin of the final parameter values transparent.

The fitted functions are compared with the binned estimates in Fig. 7.14. The left panel shows the attenuation diagnostic proxy $S(300)/E_{MC}^{\alpha'}$ as a function of $x = \sec \theta - \sec \theta_{ref}$, including event-by-event values (grey), binned averages (red markers), and the global unbinned Gaisser-Hillas fit (blue dashed curve). The agreement confirms that the GH form captures the smooth attenuation trend used to define $f_{att}(x)$ over the relevant angular range. The right panel shows the binned estimates of the exponent α as a function of $\cos^2 \theta$, together with the cubic polynomial from the global unbinned fit. The smooth evolution of $\alpha(\theta)$, and the larger uncertainties in the most inclined bins, are consistent with the staged interpretation outlined above. Moreover, the unbinned fit departs slightly from the binned values in the most inclined bins. This difference is not treated as a limiting effect, since the energy bias shown below remains within acceptable limits.

The final fitted parameters of the global unbinned calibration are summarised in Table 7.5. For clarity, the attenuation and zenith-dependent scaling parameters are listed separately, but they are obtained from a single global fit of the full calibration model.

Before defining the final summary performance for the calibration, it is instructive to inspect the reconstructed-energy response in zenith bins (Fig. 7.15). In both panels, the intervals $\theta \in [0^\circ, 30^\circ)$, $\theta \in [30^\circ, 45^\circ)$, and $\theta \in [45^\circ, 55^\circ)$ are disjoint, while the grey series $\theta \in [0^\circ, 45^\circ)$ is cumulative and represents the conservative reference range used to define the baseline photon-reconstruction performance. The bias (left panel) remains close to zero for the two low-to-intermediate zenith bins and for the cumulative $[0^\circ, 45^\circ)$ sample, with fluctuations compatible with the bin uncertainties and no clear residual trend with

Table 7.5: Best-fit parameters of the photon energy calibration obtained from the global unbinned fit of Eqs. (7.13) to (7.15). The attenuation correction is modelled with a Gaisser–Hillas function in $x = \sec \theta - \sec \theta_{\text{ref}}$, while the scaling exponent $\alpha(\theta)$ is modelled as a cubic polynomial in $\cos^2 \theta$.

Parameter	Estimate	Std. error
GH(x)		
g_0	23.04	0.01
g_1	0.232	0.009
g_2	−0.075	0.001
g_3	0.171	0.003
$\alpha(\theta)$		
a_0	2.246	0.004
a_1	−5.72	0.01
a_2	7.78	0.01
a_3	−3.275	0.003

energy. By contrast, the inclined bin $\theta \in [45^\circ, 55^\circ)$ shows a systematically negative bias with substantially larger uncertainties.

A similar pattern is observed in the resolution (right panel of Fig. 7.15). For $\theta < 45^\circ$, the relative resolution stays in the range ~ 0.10 – 0.18 depending on zenith and energy, and the cumulative $[0^\circ, 45^\circ)$ sample remains comparatively stable across the calibration range. In contrast, the $\theta \in [45^\circ, 55^\circ)$ sample exhibits a marked degradation, with significantly larger dispersion and stronger bin-to-bin fluctuations. This behaviour is consistent with the geometrical diagnostics from Section 7.3, where the core reconstruction showed increasing longitudinal (upstream/downstream) bias towards larger zenith angles, plausibly linked to residual early–late asymmetries not captured by the current azimuthally symmetric photon LDF.

Taken together, the geometry and calibration diagnostics indicate that $\theta < 45^\circ$ is the most natural conservative reference range for the Phase I photon reconstruction. In this interval, both the geometrical reconstruction and the photon energy scale remain well controlled, and it therefore provides the cleanest phase space in which to define the baseline performance of the reconstruction chain.

However, the summary performance does not indicate a severe breakdown immediately above 45° . As shown by the extended bias–resolution comparison, enlarging the phase space up to $\sim 52^\circ$ produces only a moderate degradation: the reconstructed-energy bias remains small overall, and the worsening of the resolution is noticeable but not dramatic in the energy range where the array is most relevant for the search. At lower energies, moreover, highly inclined photon showers contribute little because the trigger efficiency falls rapidly with zenith angle. For this reason, if maximising the exposure is the primary objective, extending the search up to $\sim 52^\circ$ remains a plausible option, provided that any highly inclined photon candidate is subsequently reviewed in detail on an event-by-event basis.

The overall reconstructed-energy response is summarised in Fig. 7.16, where the conservative reference range $\theta < 45^\circ$ is compared directly with the extended selection $\theta < 52^\circ$. For the reference range, the mean bias remains consistent with zero, $\langle E_\gamma / E_{\text{MC}} - 1 \rangle = -0.004 \pm 0.002$, while the mean relative resolution is $\sigma(E_\gamma / E_{\text{MC}}) = 0.142 \pm 0.003$. The extension to $\theta < 52^\circ$ increases the dispersion, but the degradation remains moderate rather than pathological, and the bias stays comparatively well controlled. This reinforces the interpretation that $\theta < 45^\circ$ is the cleanest baseline choice, while an exposure-oriented extension to larger zenith angles remains defensible.

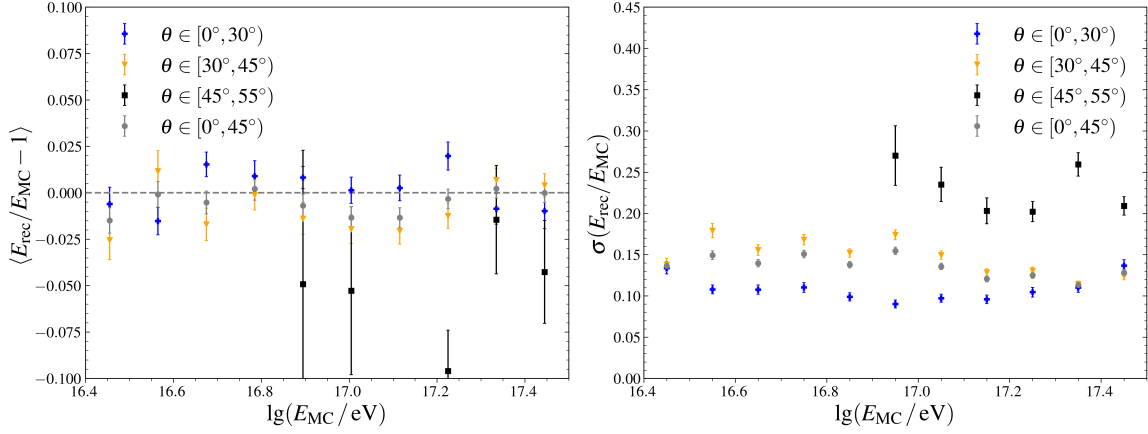


Figure 7.15: Reconstructed-energy response in zenith bins as a function of $\lg(E_{\text{MC}}/\text{eV})$. In both panels, the blue, orange, and black series correspond to the disjoint intervals $\theta \in [0^\circ, 30^\circ)$, $\theta \in [30^\circ, 45^\circ)$, and $\theta \in [45^\circ, 55^\circ)$, respectively, while the grey series is the cumulative sample $\theta \in [0^\circ, 45^\circ)$. *Left:* Mean relative bias, $\langle E_{\text{rec}}/E_{\text{MC}} - 1 \rangle$. *Right:* Relative resolution, $\sigma(E_{\text{rec}}/E_{\text{MC}})$.

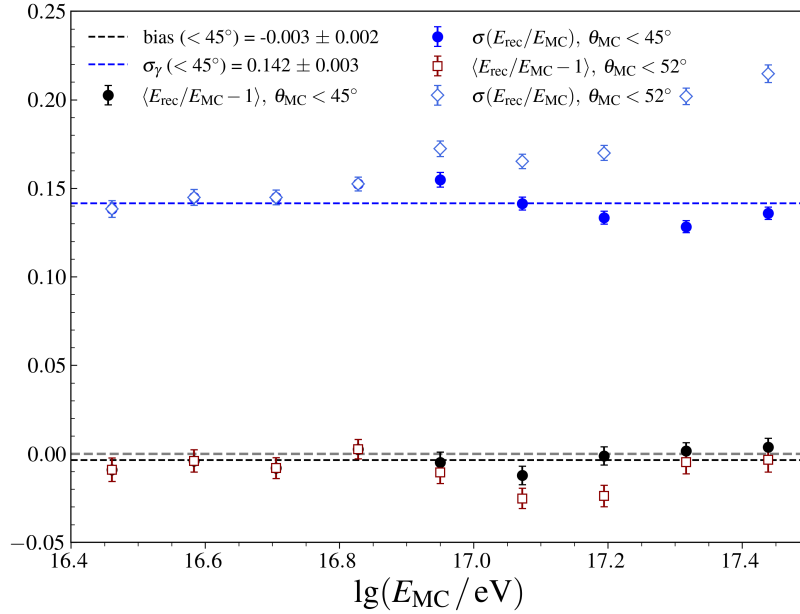


Figure 7.16: Compact summary of the photon energy response. Black markers show the bin-by-bin mean bias $\langle E_{\text{rec}}/E_{\text{MC}} - 1 \rangle$, and blue markers show the bin-by-bin relative resolution $\sigma(E_{\text{rec}}/E_{\text{MC}})$ as a function of $\lg(E_{\text{MC}}/\text{eV})$. The dashed horizontal lines indicate constant fits used to summarise the average performance over the fitted energy range, yielding a mean bias consistent with zero and a mean relative resolution of about 14%.

For completeness, Fig. 7.17 shows the response obtained when the photon energy calibration is applied to simulated hadronic primaries. The plotted quantity is $(E_{\text{MC}}/E_{\gamma}) - 1$, so negative values correspond to an overestimation of the true hadronic energy by the photon estimator. This behaviour is expected: the calibration was derived for photon-induced showers, whose ground signal is predominantly electromagnetic and whose attenuation and size-energy scaling differ from those of hadronic showers. The response for protons and iron is broadly compatible within the uncertainties, indicating that the dominant effect is the use of a photon-specific calibration rather than a strong composition-dependent shift. At low energies the photon estimator tends to assign substantially larger energies to hadronic events,

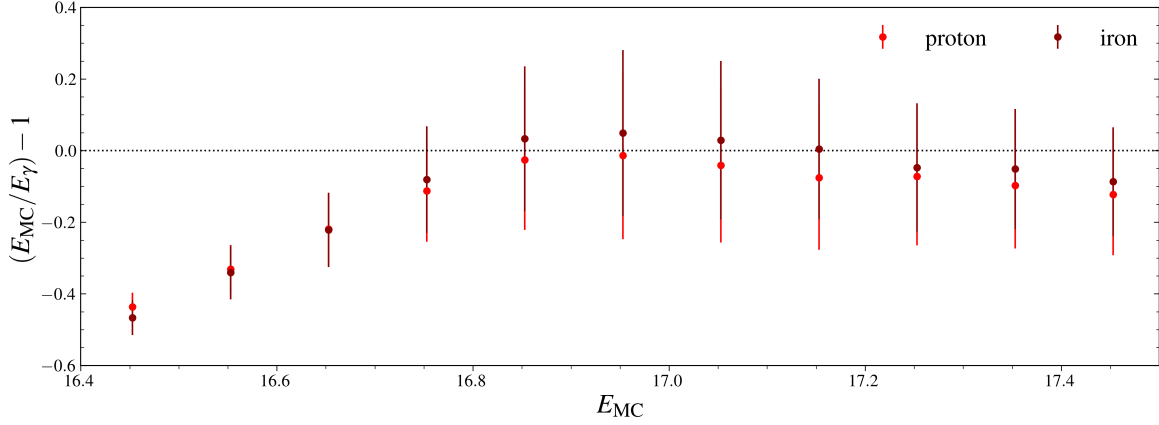


Figure 7.17: Energy response obtained when the photon energy calibration is applied to simulated hadronic primaries. The vertical axis shows $(E_{MC}/E_\gamma) - 1$, where E_γ is the energy reconstructed with the photon-specific calibration. Negative values therefore indicate that the photon estimator overestimates the true Monte Carlo energy of the hadronic shower. Red markers correspond to proton simulations and dark-red markers to iron simulations. The comparison illustrates the expected mismatch between the photon energy scale and hadron-induced showers, and is shown only as a diagnostic of the hadronic background under the photon hypothesis.

while around the central calibration range the bias is closer to zero and then becomes mildly negative again at higher energies. This mismatch is not expected to bias the photon–hadron discriminator, provided that the discriminator itself does not use the reconstructed energy estimator as an input. In that case, the classification is performed between photon events at a given reconstructed photon energy E_γ and hadronic background events which, under the same photon-energy assignment, generally correspond to lower true Monte Carlo energies.

7.5 Summary

In this chapter we established an updated and internally consistent reconstruction chain for photon-induced showers in Phase I with the SD-433. Starting from dedicated detector simulations, we first quantified the trigger/reconstruction efficiency in the (E_{MC}, θ_{MC}) plane and obtained a compact parametrisation suitable for exposure and sensitivity studies. This defines the simulated phase space in which the photon-search analysis can be formulated and provides the efficiency ingredient required for the later flux interpretation.

We then introduced a dedicated photon LDF and a fixed-slope parametrisation $\hat{\beta}(S(300), \theta)$, motivated by the stronger atmospheric attenuation of the electromagnetic component and the reduced muonic content of photon showers relative to hadronic events. The modified NKG-like form, together with the photon-specific slope model, provides a stable definition of the shower-size estimator $S(300)$ and enables a reconstruction configuration consistent with the geometry and calibration studies that follow.

Using this photon LDF, we validated the geometrical reconstruction by comparing reconstructed and Monte Carlo truth quantities for the shower axis and core position. While the angular reconstruction remains well behaved over a broad range, the core diagnostics reveal a zenith-dependent longitudinal (upstream/downstream) bias that increases towards larger inclinations, consistent with residual early–late asymmetries in the predominantly electromagnetic ground signal. These geometry checks already indicate that a conservative zenith restriction is needed to define a robust photon-reconstruction domain.

Finally, we derived the photon energy calibration by combining a photon-specific attenuation correction and a zenith-dependent size–energy exponent. Relative to the standard

hadronic calibration, the photon calibration differs from the standard hadronic one both in its attenuation model and in its zenith-dependent scaling exponent $\alpha(\theta)$. The calibration was constructed through a staged procedure: binned fits were used to obtain intermediate estimates, and these were used as seeds for the nominal global unbinned fit. The binned energy-response diagnostics in disjoint zenith intervals show that the reconstruction remains most stable below 45° , while the more inclined sample exhibits a moderate but visible degradation in bias and, more clearly, in resolution. Accordingly, $\theta < 45^\circ$ emerges as the conservative reference range for the Phase I photon reconstruction, i.e. the cleanest baseline in which to define the nominal reconstruction and calibration performance.

At the same time, the comparison with the cumulative $\theta < 52^\circ$ selection indicates that the degradation beyond 45° is not prohibitive. An extension to larger zenith angles therefore remains plausible if the gain in exposure is considered more important than maintaining the most conservative reconstruction domain, especially given that highly inclined low-energy photons contribute little to the effective acceptance and that any inclined candidate can be re-examined individually.

Taken together, the efficiency model, photon LDF, geometry validation, and photon energy calibration define most of the reconstruction ingredients required to perform a photon search with the SD-433. Within the conservative reference range ($\theta < 45^\circ$), this chapter establishes the baseline photon energy scale used to reconstruct candidate events, evaluate the exposure, and interpret the results in terms of photon flux constraints. For this reference sample, the reconstructed-energy response is summarised by a mean bias consistent with zero, $\langle E_\gamma/E_{\text{MC}} - 1 \rangle = -0.004 \pm 0.002$, and a mean relative resolution $\sigma(E_\gamma/E_{\text{MC}}) = 0.142 \pm 0.003$ over the fitted energy range. However, the broader comparison presented in this chapter also indicates that this conservative range need not be viewed as the only viable one: extending the search up to $\sim 52^\circ$ appears plausible when maximising exposure is the primary objective, since the additional degradation in reconstruction performance remains moderate and does not point to a breakdown of the method. The $\theta < 45^\circ$ range therefore serves best as the nominal reference configuration, while a more inclusive zenith interval can still be justified for the search itself.

Photon-hadron discrimination

The truth is rarely pure and never simple.

The Importance of Being Earnest, Oscar Wilde

In its most general form, *discrimination*, or pattern recognition, consists of assigning an observed object to one of a finite set of classes on the basis of a finite collection of measurements. In the probabilistic formulation adopted here, an air-shower event is represented by a set of detector observables, or more generally by a structured object built from the measured detector response, and the goal is to infer its unknown class from that information alone. For the purposes of this thesis, the relevant classes are *photon-induced* air showers and *hadron-induced* air showers. Since the energy range of interest starts at tens of PeV, the relevant detectors are the SD and the UMD in the 433 m configuration. The task of this chapter is therefore to construct and validate a rule that, given the measured detector response of a single event, quantifies how compatible that event is with either of these two classes.

The chapter combines two complementary parts. First, the classification problem is formulated, the event representation is defined, and the graph-neural-network architecture used in the analysis is described. Second, the resulting network scores are validated with independent simulations, compared with the M_1 muon-based observable [52], tested under degraded detector-information scenarios, and interpreted with event-level and dataset-level explainability tools. The search consequences of the method, including the validation on real burn-sample data and the corresponding exposure projection, are discussed separately in Chapter 9.

It is worth clarifying what will *not* be attempted in this chapter. A full introduction to deep learning is intentionally omitted. Instead, the objective is narrower and more directly relevant to the photon-search problem: to define the classification task precisely, to identify the physically meaningful inputs and labels, and to explain how a detector-level representation of an event can be mapped onto a photon-likeness score. Definitions and algorithmic details are introduced only when they are needed to justify the architecture, the training strategy, the validation procedure, or the interpretation of the final discriminator output.

8.1 Formal statement of the classification problem

Let X denote the information extracted from a reconstructed event, and let Y denote its (unknown) class label. In the binary setting relevant here,

$$Y \in \{0, 1\}, \quad (8.1)$$

where, by convention, $Y = 1$ corresponds to a photon-induced shower (signal) and $Y = 0$ to a hadron-induced shower (background). Following the probabilistic formulation of pattern recognition, the pair (X, Y) is regarded as a random object, and the goal is to infer Y from X alone.

A classifier is therefore a measurable decision function mapping the observation space into the label space. In practice, rather than considering all conceivable decision rules, one works with a restricted family of candidates,

$$\mathcal{H} = \{g_\alpha : \alpha \in \Lambda\}, \quad (8.2)$$

that is, a *hypothesis space* (or function class) indexed by parameters α . More generally, especially in modern machine-learning applications, one often constructs a real-valued score $s(X)$, from which a hard decision can be obtained by applying a threshold. In either case, the learning problem is to select, from $\mathcal{H}$, a rule that approximates the unknown class assignment as accurately as possible.

For a classifier $g \in \mathcal{H}$, the natural performance criterion under 0–1 loss is its *risk*, namely the probability of misclassification,

$$R(g) \equiv \mathbb{P}(g(X) \neq Y). \quad (8.3)$$

Equivalently, introducing the loss

$$\ell(g(X), Y) = \begin{cases} 0, & g(X) = Y, \\ 1, & g(X) \neq Y, \end{cases} \quad (8.4)$$

one may write

$$R(g) = \mathbb{E}[\ell(g(X), Y)]. \quad (8.5)$$

In this sense, the classification task is a particular instance of the general problem of risk minimisation: among all admissible hypotheses in $\mathcal{H}$, we seek one whose expected loss is as small as possible.

A central concept in this framework is the *Bayes* decision rule. If the full joint distribution of (X, Y) were known, then the optimal classifier under the misclassification criterion would assign to each observation the most probable class given the measured event. Equivalently, defining the posterior probability

$$\eta(x) \equiv \mathbb{P}(Y = 1 \mid X = x), \quad (8.6)$$

the ideal rule decides in favour of the photon class whenever $\eta(x) > 1/2$, and in favour of the hadron class otherwise. The corresponding minimum achievable error is the *Bayes risk*, denoted here by R^* . This optimal rule provides the natural theoretical benchmark for the problem. However, in realistic applications the joint distribution of (X, Y) is unknown, and neither the posterior probability nor the Bayes rule can be constructed directly. The practical objective is therefore not to attain the exact Bayes classifier, but to construct, from finite data, a decision function whose risk is as small as possible and ideally close to the best achievable within the chosen hypothesis space.

To do so, one introduces a finite labelled sample

$$D_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}, \quad (8.7)$$

from which a data-dependent classifier, here denoted by g_n , is constructed. At the sample level, the inaccessible true risk $R(g)$ is replaced by an empirical counterpart,

$$\hat{R}_n(g) = \frac{1}{n} \sum_{i=1}^n \ell(g(X_i), Y_i), \quad (8.8)$$

or, in practice, by a related optimisation objective. This makes explicit the central distinction emphasised in statistical learning theory: the classifier is built using finite training data, but its quality is determined by its performance on previously unseen events. Good behaviour on the training sample alone is not sufficient; the essential requirement is *generalisation*. In the present context, this means that the learned classifier should not merely reproduce features specific to the examples on which it was optimised, but should instead capture robust and transferable differences between photon- and hadron-induced showers that generalise to unseen events.

This statistical viewpoint is especially important for the present application because photon–hadron separation is not expected to be exact on an event-by-event basis. Shower development is intrinsically stochastic, detector measurements are noisy and incomplete, and hadronic fluctuations can partially mimic photon-like phenomenology [170]. The goal is therefore not to define an infallible tag, but to construct a discriminator that orders events according to their degree of photon-likeness while keeping the hadronic contamination under control. The quality of the classifier is determined by the trade-off between retaining genuine photon candidates and rejecting as much of the hadronic background as possible.

The previous formulation is deliberately abstract. In this chapter, the random object X is the detector-level representation of an extensive air shower as seen by the hybrid response of the SD and the UMD. It must encode, in a compact but sufficiently informative way, the observables that carry discrimination power between photon and hadron primaries. These include not only the WCD signals and their associated traces, but also their spatial organisation across the array and the complementary information on the muonic component provided by the UMD. The structure of X is therefore intrinsically relational, which motivates the graph-based representation introduced in the following section.

The *supervised* sample used to define D_n is obtained from Monte Carlo simulations. Measured data do not provide event-by-event class labels, and the diffuse photon flux, if present at all, is expected to be deeply subdominant with respect to the cosmic-ray background. Consequently, the classification problem must be formulated in terms of simulated signal (photon) and background (hadrons) classes, which serve as proxies for the unknown true populations. Measured data are then used not to provide labels, but as the final domain of application of the trained classifier and, in the case of a burn sample, as a benchmark against which possible mismodelling and systematic effects can be assessed.

8.2 Event representation as graphs

A common starting point in supervised classification is to represent each event as a fixed-length vector of handcrafted summary features, i.e. as *tabulated* data. This representation is well matched to many generic classifiers (including tree-based methods such as boosted decision trees), but it is not neutral: by construction it compresses an event into a list of global numbers and therefore risks discarding information carried by the *structure* of the detector response. For the SD–UMD hybrid reconstruction, the measured footprint is intrinsically

spatio-temporal: signals and times are recorded at discrete detector positions, and their correlations across the array encode physically meaningful patterns (front curvature, lateral fall-off, upstream/downstream asymmetries, muon-pooriness, etc.). This observation underlies the broader use of deep-learning methods at the Pierre Auger Observatory, motivated precisely by the fact that shower footprints contain complex spatio-temporal information that is challenging to exploit with traditional approaches [171–173].

Recently, several machine-learning analyses have incorporated detector-level structure explicitly. Neural networks have been trained to leverage the spatio-temporal information of SD shower footprints; they have been applied, among other tasks, to energy reconstruction using the time structure of water-Cherenkov detector (WCD) signals [174], to indirect estimation of $X_{\max}$ [175, 176], and to AugerPrime-era studies where transformers and convolutional networks improve the reconstruction of mass-sensitive observables such as $X_{\max}$ and the muon number [177]. These developments illustrate a recurring theme: when the relevant information is encoded in spatial and temporal correlations, architectures with built-in *spatiotemporal awareness* (e.g. convolutions on grids, attention over sequences, or message passing on graphs) can learn more informative representations than models that only see an unordered list of features.

For the specific problem of photon–hadron discrimination with SD-433 and UMD-433, a graph formulation is particularly natural. At detector level, an event is not a single fixed-length vector but a *set* of triggered stations embedded in the array plane, with node multiplicity varying strongly with energy, zenith angle, and trigger/quality conditions. This immediately creates three challenges for fixed-input models: (i) the input size is variable from event to event, (ii) the effective spatial sampling of the footprint could be irregular (because only a subset of stations triggers), and (iii) the ordering of stations is physically meaningless and should not affect the prediction (permutation invariance). Graph Neural Networks (GNNs) address these challenges by construction: they operate on variable-size graphs, aggregate information through permutation-invariant message-passing operators, and incorporate geometry through edges that encode spatial neighbourhoods in the detector grid. Moreover, the combined SD–UMD response is intrinsically *hybrid*: SD stations measure a mixture of electromagnetic and muonic components at ground, while the UMD provides a more direct handle on the muonic component and is therefore especially powerful for identifying muon-poor photon showers. A graph representation allows these complementary measurements to be fused locally and non-locally by propagating information across the footprint.

A second, practical motivation for using GNNs in this application is *incompleteness* and *irregularity*. Even though the underlying SD-433 layout is close to a regular triangular lattice, real and simulated events rarely populate that lattice as a complete image: stations may fail quality cuts, be temporarily inactive, or simply not trigger because the shower footprint is sparse, near the array boundary, or strongly attenuated. The incompleteness is even more pronounced for the UMD, since during Phase I it was deployed only in a fraction of the SD-433 array. A useful discriminator should therefore not rely on a rigid, fully observed grid; it must instead learn from partially observed detector configurations and degrade gracefully when information is missing.

In our implementation, this is achieved without imputing unmeasured quantities: a single graph encodes SD and UMD information, and binary validity masks are used to construct SD-only, UMD-only, and joint SD–UMD views on the fly during training and inference. This ensures that each information stream is exploited when present while remaining well defined when absent. The graph representation also makes it possible to include stations that do not lie on the nominal SD-433 grid. For instance, triggered stations from the first crown of SD-750 surrounding the SD-433 array are included as part of the event graph, since

exact alignment with the SD-433 lattice is not required for the definition of graph nodes and edges.

The supervised samples used throughout this chapter are based on a dedicated CORSIKA photon production released by the Auger MC Task. The air showers were generated with *corsika-7.7420*, with primary energies spanning the interval $\lg(E_{\text{MC}}/\text{eV}) \in [16.0, 18.0]$ in logarithmic bins of width 0.5, following an E^{-1} generation spectrum, and zenith angles up to 60° . The primary particles used here include photons, protons, and iron nuclei. In the supervised setting, protons are adopted as the reference (conservative) hadronic background, while iron is used for cross-checks that probe the behaviour of the discriminator in the presence of a substantially higher muon content.

The shower library was generated with EPOS LHC as the high-energy hadronic interaction model. The detector response of the SD-433 and the UMD-433 is simulated and reconstructed within the Auger *Offline* framework, using a modified version of the standard *MdSd433SimRec* application. In particular, the reconstruction sequence includes a dedicated photon energy calibration [178], implemented through custom modules, so that all events entering the machine-learning pipeline are reconstructed on a consistent photon energy scale.

Each reconstructed event is represented as a single undirected graph

$$G = (V, E), \quad (8.9)$$

where the node set V corresponds to triggered SD stations and the edge set E encodes neighbourhood relations in the SD-433 geometry. Each node carries local station-level information, while the edges specify which stations are allowed to exchange information in the subsequent graph-neural-network layers.

Let $N \equiv |V|$ denote the number of triggered stations in the event. When node features are present, the graph can be represented by an adjacency matrix

$$A \in \{0, 1\}^{N \times N}, \quad (8.10)$$

together with a node-feature matrix

$$X \in \mathbb{R}^{N \times F}, \quad (8.11)$$

where F is the number of features associated with each station. The i -th row of X , denoted x_i , contains the feature vector of station i . For the time-trace information, an additional matrix is used,

$$X_{\text{traces}} \in \mathbb{R}^{N \times T}, \quad (8.12)$$

where T is the number of time bins retained from the SD trace.

The adjacency matrix defines the neighbourhood of each station,

$$\mathcal{N}(i) \equiv \{j \in V : A_{ij} = 1\}. \quad (8.13)$$

In the implementation used here, two stations i and j are connected, $A_{ij} = 1$, if they are neighbours up to second order in the triangular SD-433 grid. This choice anchors the graph structure to the detector geometry: stations exchange information only with nearby stations, while the event as a whole remains a variable-size object whose number of nodes is determined by the triggered footprint. The precise neighbourhood definition is nevertheless a modelling choice. In particular, the graph provides the set of possible local interactions, whereas the attention mechanism, discussed in the next section, learns how strongly each neighbouring station contributes during message passing.

The node features are defined relative to the station with the largest measured SD signal in the event, which acts as a reference node. For each node, the feature vector includes the spatial and temporal offsets $(\Delta x, \Delta y, \Delta z, \Delta t)$ with respect to that reference, the number n_{PMT} of working PMTs in the WCD, the reconstructed SD signal, and, when available, UMD-related quantities such as the estimated muon density ρ_μ and the corresponding UMD effective area A_{eff} . This construction keeps the representation local at the station level while still encoding the relative geometry of the shower footprint.

The trace tensor X_{traces} contains the first $T = 60$ time bins, corresponding to 25 ns each, of the SD signal trace averaged over the available PMTs. Before being passed to the network, the trace amplitudes are logarithmically rescaled according to

$$\tilde{S}_i(t) = \frac{\lg[S_i(t)/\text{VEM} + 1]}{\lg(101)}, \quad (8.14)$$

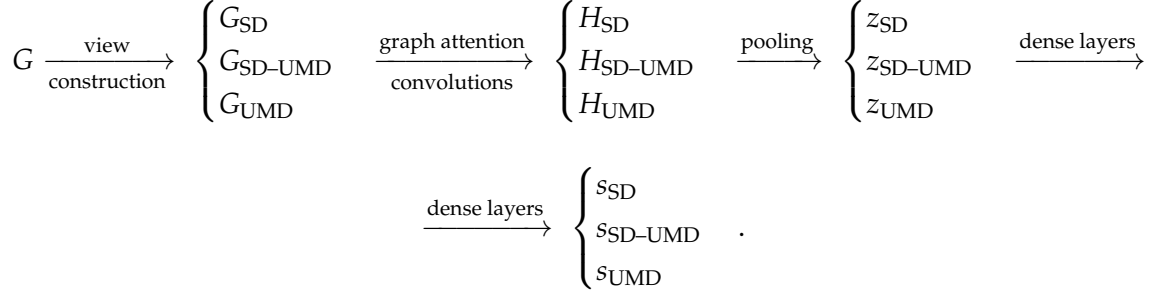
where $S_i(t)$ is the signal in time bin t of station i . The logarithmic scaling compresses the dynamic range of the traces while retaining the relative time structure of the signal. Together with the truncation to the first 60 bins, this preserves the most informative part of the waveform, particularly its early-time development [57]. Since photon-induced showers are expected to exhibit longer rise times in the VEM trace, retaining these initial bins provides a compact representation that still captures temporal information relevant for discrimination, while remaining computationally efficient for large-scale training. The traces are calibrated in VEM charge, which was considered to carry more direct physical meaning than the VEM-peak alternative. In addition, nominal traces are used for simulations, whereas data traces are deconvolved to reproduce the area-to-peak relationship observed in simulations. A more detailed explanation can be found in Ref. [179], where this technique is described and validated.

In both real data and realistically augmented simulations, some detector information can be absent. At station level, SD traces can be partially degraded, for example because one or more PMTs are missing, and UMD measurements may be unavailable because a station is not instrumented, not operational, or rejected by quality criteria. To handle this incompleteness without imputing unmeasured quantities, every node is associated with binary validity masks specifying which detector components are available. These masks indicate, for example, whether the SD information is present, whether a valid UMD measurement exists, and whether trace information can be used.

8.3 Network architecture

The neural network described below was implemented in PyTorch Geometric (PyG) [180], a widely used framework that extends the PyTorch interface with optimised graph data structures, message-passing layers, and batching utilities. Beyond its maturity and community support, PyG is particularly convenient in the present context because it allows fine-grained customisation at every stage of the training pipeline. This flexibility is especially valuable here because graph classification is still a relatively recent problem setting compared with more established machine-learning tasks, and its development continues to lie at the frontier of graph neural network research [181].

The full discriminator is a modular graph network that maps a reconstructed event graph G to *photon-likeness scores*. At a high level, the forward pass can be understood as the composition:



where G_i are the different views of the event graph, H_i are intermediate learned node-level representations [181], z_i are pooled graph-level embeddings [182], and each s_i is interpreted as a continuous photon-likeness score (not bounded between zero and one). The three scores correspond to three complementary information paths: an SD-only view, a UMD-only view, and a joint SD–UMD view. This design serves two purposes. First, it provides a natural mechanism to exploit whichever detector information is available on an event-by-event basis, without requiring imputation. Second, it makes the contribution of each sub-detector explicit, enabling controlled studies of the discrimination power of the SD and UMD separately, and of the gain achieved by their combination. In this sense, considerations of *interpretability* are built into the network architecture itself.

The graph branches do not operate directly on the raw node features at every stage. Instead, they construct progressively richer node-level representations. For a node i , we denote by

$$h_i^{(\ell)} \in \mathbb{R}^{d_\ell} \quad (8.15)$$

its learned embedding after layer ℓ , where d_ℓ is the embedding dimension at that depth. The initial representation $h_i^{(0)}$ is built from the input station features, including the trace embedding when applicable, and subsequent layers transform these embeddings into representations that are increasingly adapted to the photon–hadron discrimination task.

This representation-learning viewpoint is useful for interpreting the modular structure of the network. The quantities H_{SD} , H_{UMD} , and $H_{\text{SD-UMD}}$ in the schematic pipeline correspond to collections of such learned node embeddings for the three graph views. These node-level representations are later reduced to graph-level vectors and finally mapped to the scores s_{SD} , s_{UMD} , and $s_{\text{SD-UMD}}$.

A schematic summary of the full pipeline is shown in Fig. 8.1. The following subsections describe the individual components of the network in more detail.

8.3.1 Subgraph extractor

Although the SD-433 layout is a regular triangular lattice, a recorded event does not populate that lattice as a complete image. As explained above, to ensure that the network remains well defined under these realistic sources of incompleteness, each node carries binary validity flags indicating which detector components are available, e.g. whether the SD information is present, whether a UMD measurement exists, and whether trace information is usable.

A dedicated module, referred to as the *Subgraph Extractor*, uses these masks to construct three graph views from the same underlying event graph:

- G_{SD} : the SD-only view, containing nodes with valid SD information;

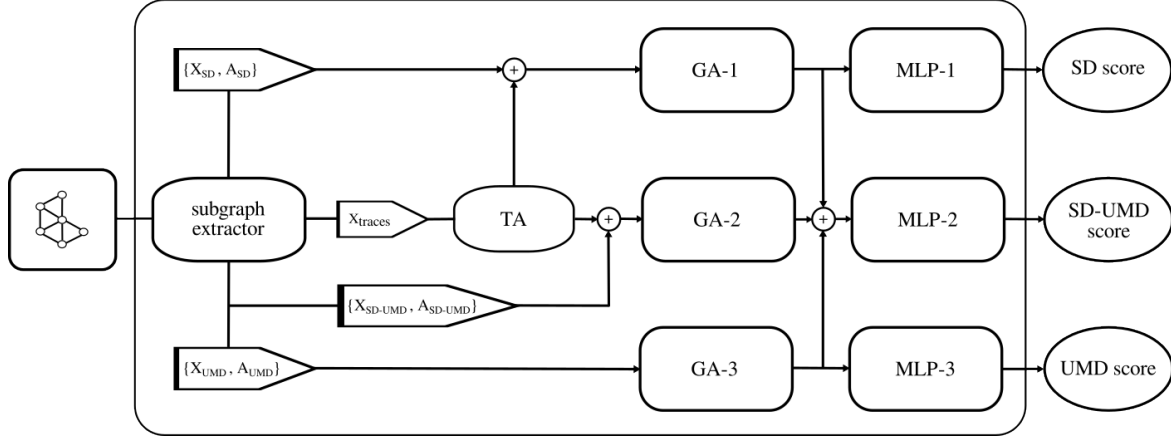


Figure 8.1: Schematic overview of the modular multi-view graph neural network used for photon–hadron discrimination. Starting from a reconstructed event graph, the *Subgraph Extractor* builds three views according to detector availability: an SD-only graph, a joint SD–UMD graph, and a UMD-only graph. For nodes with valid SD traces, the *Trace Analyzer (TA)* produces learned trace embeddings that are injected into the SD and joint branches. Each view is then processed by a dedicated *Graph Analyzer*, implemented with stacked graph-attention layers. The SD and UMD branches yield detector-specific graph-level representations and scores, while their pooled representations are also concatenated with that of the joint branch before the final multilayer perceptron head, producing the combined SD–UMD score. Here, X and A denote the node-feature matrix and adjacency structure of each view, respectively, and circled plus signs indicate feature-concatenation operations.

- G_{UMD} : the UMD-only view, containing nodes with valid UMD measurements;
- $G_{\text{SD-UMD}}$: the joint view, restricted to nodes for which both SD and UMD information are simultaneously available.

Each view inherits an edge structure consistent with the detector geometry, connecting neighbours up to second order in the SD-433 lattice, so that message passing remains meaningful in all branches. This design enforces, by construction, that UMD information is used only where it exists, while still allowing the network to extract discrimination power from the SD footprint when UMD coverage is incomplete.

This initial module is completely deterministic, with no learnable parameters involved. Its sole purpose is to organise the input information before further processing. Whether missing features result from data augmentation or from real data affected by non-functioning detectors, this part of the network separates the information accordingly. In principle, this operation could be implemented outside the network; however, training showed that keeping it within the compiled architecture is more efficient in terms of speed.

8.3.2 Trace Analyzer

The SD stations record not only an integrated signal but also a time trace whose shape carries information about the particle arrival-time distribution at ground. Rather than compressing this information into a small set of handcrafted summary variables, we feed a truncated and normalised trace representation into the network. Concretely, for each node with available trace information we use the first $T = 60$ time bins (25 ns each) of the trace, averaged over the available PMTs, thereby forming the per-node input

$$X_{\text{traces}} \in \mathbb{R}^{N \times T}.$$

A TA is then used to process these sequences with one-dimensional convolutional layers and to produce a learned trace representation for each node.

The TA is implemented as a one-dimensional convolutional network acting independently on each node trace. Each valid trace, originally represented by the first $T = 60$ time bins, is reshaped to a tensor of shape $(n_{\text{traces}}, 1, 60)$ and processed through three successive convolutional blocks. These blocks use, respectively, 64, 32, and 4 output channels, with kernel sizes 7, 6, and 5, and strides 3, 3, and 1. Each convolution is followed by a batch-normalisation layer [183] and a LeakyReLU activation. Since no padding or pooling is introduced, the temporal dimension is reduced solely by the convolutions themselves, following the sequence $60 \rightarrow 18 \rightarrow 5 \rightarrow 1$. The final output is therefore a four-dimensional embedding for each trace.

This architecture provides a compact summary of the waveform while preserving the temporal structures most relevant for discrimination. The first layer probes comparatively broad temporal windows and is therefore sensitive to coarse features such as the signal onset and rise-time development, whereas the deeper layers progressively compress this information into a low-dimensional representation. Because the final temporal dimension is reduced to unity, the output of the Trace Analyzer can be interpreted as an additional node-level descriptor and concatenated directly with the standard station features before graph processing.

The same Trace Analyzer is used in the two branches that contain SD information, namely the SD and SD-UMD views, while the UMD-only branch receives no trace input. This ensures that any differences between branches reflect the detector information available in each view, rather than differences in the trace-encoding stage itself.

8.3.3 Graph Analyzer

After node encoding, each graph view is processed by a dedicated Graph Analyzer (GA) branch based on stacked GATv2Conv layers [184, 185]. In the implementation adopted here, each branch consists of three successive graph-attention blocks, each formed by a GATv2Conv layer followed by batch normalisation. The role of the GA is to propagate information across the detector footprint. Most graph-neural-network layers can be understood within the message-passing framework.

At layer ℓ , each node updates its representation by combining its current embedding with information received from neighbouring nodes. Schematically,

$$m_{ij}^{(\ell)} = \text{MSG}^{(\ell)}(h_i^{(\ell)}, h_j^{(\ell)}), \quad (8.16)$$

$$a_i^{(\ell)} = \text{AGG}^{(\ell)}(\{m_{ij}^{(\ell)} : j \in \mathcal{N}(i)\}), \quad (8.17)$$

$$h_i^{(\ell+1)} = \text{UPD}^{(\ell)}(h_i^{(\ell)}, a_i^{(\ell)}). \quad (8.18)$$

Here, $m_{ij}^{(\ell)}$ denotes the message sent from node j to node i , $a_i^{(\ell)}$ is the aggregated information received by node i , and $h_i^{(\ell+1)}$ is the updated node embedding. Because the neighbours of a node form a set rather than an ordered sequence, the aggregation operation must be permutation invariant. This requirement is essential in the present application, since the prediction should not depend on the arbitrary ordering of stations in memory.

Stacking several graph layers increases the effective receptive field of each station. After one layer, a station representation can depend on its immediate neighbours. After two layers, it can also depend indirectly on neighbours of neighbours. With three graph-attention layers, as used here, the learned station embeddings can therefore encode information from a wider

portion of the shower footprint, while still respecting the local neighbourhood structure imposed by the SD-433 geometry.

A graph-attention layer extends this message-passing idea by learning how strongly neighbouring stations should contribute to each update. In a standard attention formulation, node embeddings are first linearly transformed,

$$z_i^{(\ell)} = W^{(\ell)} h_i^{(\ell)}, \quad (8.19)$$

and an unnormalised attention score is computed for each neighbouring pair,

$$e_{ij}^{(\ell)} = a^{(\ell)}(z_i^{(\ell)}, z_j^{(\ell)}). \quad (8.20)$$

The scores are then normalised over the neighbourhood of node i ,

$$\alpha_{ij}^{(\ell)} = \frac{\exp(e_{ij}^{(\ell)})}{\sum_{k \in \mathcal{N}(i) \cup \{i\}} \exp(e_{ik}^{(\ell)})}, \quad (8.21)$$

so that the updated node representation is obtained from an attention-weighted combination of neighbouring embeddings,

$$h_i^{(\ell+1)} = \sigma \left(\sum_{j \in \mathcal{N}(i) \cup \{i\}} \alpha_{ij}^{(\ell)} z_j^{(\ell)} \right). \quad (8.22)$$

The exact parametrisation used by GATv2Conv differs from the original graph-attention layer, but the same principle applies: neighbouring node embeddings are transformed, weighted through learnable attention coefficients, and aggregated to produce updated node representations.

This mechanism is well motivated for photon-hadron discrimination. In principle, the relevant information is neither purely local, *i.e.* contained in an isolated station, nor purely global, *i.e.* reducible to a single event-level summary. Rather, it lies in how signals, times, trace shapes, and muon content co-vary across neighbouring stations as a function of distance and shower geometry. The graph-attention mechanism therefore provides a learned, data-driven way of combining station information across the event footprint, while the fixed neighbourhood definition anchors that aggregation to the detector geometry.

In practice, graph-attention layers are implemented with several attention heads running in parallel. Each head computes its own set of attention coefficients and its own transformed aggregation. The resulting representations are then combined internally by the layer. The motivation is that different heads can respond to different aspects of the local structure, such as signal gradients, timing correlations, or muon-content patterns. Using several heads also stabilises the learned representation, since the update is not tied to a single attention pattern. The attention layers used here have four attention heads and include residual connections. This choice should be regarded as an architectural hyperparameter: increasing the number of heads did not improve the validation performance in the tests performed, while it increased the number of trainable parameters and the computational cost.

All GA branches share the same architectural design. That is, the SD, UMD, and SD-UMD paths use the same sequence of graph-attention and normalisation layers, as well as the same hyperparameter choices, although their trainable weights are learned independently. This makes the three paths directly comparable, since differences in performance can be attributed primarily to the information content of the corresponding graph views rather than to path-specific architectural variations. In addition, the SD and SD-UMD branches receive the learned trace embedding produced by the Trace Analyzer, whereas the UMD-only branch is processed without trace information, consistently with the detector content available in each case.

8.3.4 Dense layers and scores

Photon–hadron discrimination is ultimately an event-level task, so the node-wise representations must be reduced to fixed-size graph embeddings. This operation is usually referred to as a graph readout or global pooling. The readout must be permutation invariant, because the ordering of stations in the graph is arbitrary. For a set of node embeddings $\{h_i\}_{i \in V}$, typical pooling operations are

$$p_{\text{sum}} = \sum_{i \in V} h_i, \quad (8.23)$$

$$p_{\text{mean}} = \frac{1}{|V|} \sum_{i \in V} h_i, \quad (8.24)$$

$$p_{\text{max}} = \max_{i \in V} h_i, \quad (8.25)$$

where the maximum is taken component-wise.

These pooling functions emphasise complementary aspects of the graph. The sum retains information related to the accumulated activation over the footprint, the mean characterises the typical scale of the node embeddings, and the maximum highlights the strongest local response. For this reason, each branch uses the concatenation of global mean, max, and sum pooling rather than relying on a single readout operation.

To retain information from different depths of the message-passing stack, pooling is performed both after the first graph-attention layer and after the last one. The resulting pooled vectors are concatenated to form the graph-level representation of each branch. This gives the classifier access both to relatively shallow representations, which remain closer to the original node features, and to deeper representations, which encode more extensively propagated contextual information. Although this operation is not equivalent to a residual connection, it serves a related purpose: it preserves information from earlier stages of the network while still allowing the final score to exploit the more contextual features learned by deeper layers.

Once a graph-level embedding has been obtained, the final prediction can be produced by an ordinary multilayer perceptron. Schematically,

$$s = f_{\text{MLP}}(z), \quad (8.26)$$

where z denotes the pooled graph-level representation and s is the corresponding photon-likeness score. At this stage, the graph neural network has already converted the variable-size detector footprint into a fixed-size learned representation.

The pooled SD and UMD representations are each passed through a dedicated multilayer perceptron, yielding the detector-specific scores s_{SD} and s_{UMD} . For the joint branch, the pooled representation of the SD–UMD view is concatenated with the corresponding pooled representations of the SD and UMD branches, as indicated in Fig. 8.1, and this combined event-level representation is then mapped by a final multilayer perceptron to the score

$$s_{\text{SD-UMD}} = f_{\text{MLP}}(z_{\text{SD}} \oplus z_{\text{SD-UMD}} \oplus z_{\text{UMD}}), \quad (8.27)$$

where z_{SD} , $z_{\text{SD-UMD}}$, and z_{UMD} denote the pooled graph-level embeddings of the three branches, and $\oplus$ represents concatenation.

Keeping detector-specific branches while also constructing a joint score serves three purposes. First, it quantifies the standalone discrimination power of the SD and UMD information. Second, it isolates the gain obtained by combining both sub-detectors. Third, it makes the role of missing detector information explicit, thereby improving the interpretability and diagnosability of the architecture.

It is important to note that no final sigmoid activation is applied to any of the network outputs. The three scores are therefore treated as raw, unbounded logits rather than as quantities constrained to the interval $[0, 1]$. Although applying a sigmoid would make the outputs resemble class probabilities, such values should not in general be interpreted as calibrated probabilities without an additional calibration procedure [186]. Moreover, for a relatively well-separated classification task, a sigmoid activation can compress large regions of score space close to either 0 or 1, reducing the resolution available in the tails of the class distributions. Since the photon-search application is precisely sensitive to the extreme photon-like tail of the hadronic background, we keep the discriminator outputs in logit space.

8.4 Training stage

For the training stage, we adopt proton-induced showers as the reference hadronic background. This choice should not be interpreted as a statement that the real background in the relevant energy range is proton dominated. On the contrary, the composition study developed in Chapter 6 indicates that the background is expected to be heavier than a pure proton population, and therefore more compatible with an intermediate-to-heavy composition. Using protons as the background class is therefore a deliberately conservative choice. Physically, proton showers tend to be less muon-rich than showers induced by heavier nuclei and are consequently closer to photons in the observables most relevant for this analysis. As a result, training the classifier to separate photons from protons makes the discrimination problem more demanding than it would be for a heavier hadronic mixture, reducing the risk of overestimating the separation power.

The simulated sample is split into statistically independent training, validation, and test subsets. The class distributions are balanced, and care is taken to preserve the energy and zenith-angle distributions across all sets. The event counts are of order 2×10^5 for training and 6×10^4 for validation. A larger independent test sample is reserved for the final performance assessment. In addition to photons and protons, with approximately 7.5×10^4 events, this test sample also includes iron showers, with approximately 1.5×10^5 events, for studies beyond the binary training setup.

Earlier versions of the network were trained either by optimising the three heads jointly [187] or by following a staged procedure in which only one branch was updated at a time while the remaining branches were kept fixed [178]. In the final implementation adopted here, we use a global optimisation strategy. The three heads are trained simultaneously by minimising the sum of their binary-classification losses,

$$\mathcal{L} = \mathcal{L}_{\text{SD}} + \mathcal{L}_{\text{UMD}} + \mathcal{L}_{\text{SD-UMD}}. \quad (8.28)$$

This choice did not lead to a statistically significant degradation or improvement in the final separation power with respect to the staged procedure, but it reduced the computational cost and simplified the training workflow.

The network outputs are treated as raw logits, and each loss term is computed with the binary cross-entropy with logits. The trainable parameters are updated with the Adam optimiser using an initial learning rate of 10^{-4} [188]. A ReduceLROnPlateau scheduler monitors the validation loss and lowers the learning rate by a factor of 0.5 after five epochs without improvement. To stabilise the optimisation further, gradients are clipped to a maximum norm of unity, and automatic mixed-precision training is used whenever CUDA is available. The nominal training schedule corresponds to 100 epochs, although reasonable performance is already obtained after approximately 10 epochs. Model checkpoints are saved after every epoch, and the best model is selected according to the validation loss.

Besides the training and validation losses, we monitor the ROC-AUC throughout optimisation. In addition, for diagnostic purposes, we record a simple contamination proxy obtained by taking the median score of the simulated photon sample and counting the number of hadronic events above that threshold. This quantity is not used directly for optimisation, but it provides a convenient event-level summary of how the learned score separates signal-like and background-like events during training. Its interpretation is nevertheless limited by the finite statistics of the validation sample, especially in the most photon-like tail of the hadronic distribution.

A central objective of the training procedure is to obtain a discriminator that remains usable under incomplete detector configurations. The network is therefore not trained only under idealised detector conditions. Instead, the augmentation pipeline includes realistic degradations designed to mimic incomplete configurations encountered in operation. These include random masking of PMTs, masking of UMD modules, removal of complete stations, removal of UMD information at selected stations, and occasional restriction of the UMD footprint to the hottest counter. These operations are applied at batch-loading time through dedicated transformations passed to the PyG data loaders.

The PMT, module, and station masking probabilities used during training are intentionally aggressive compared with the typical conditions observed in the field. This choice forces the network to learn a discrimination rule that remains stable under partial detector incompleteness, rather than relying on a single idealised event representation. It also acts as a strong regularisation mechanism. Because the masking is randomised at station level, the network rarely sees exactly the same event twice during training: different PMT masks produce slightly different trace inputs, different UMD-module masks produce slightly different effective muon-density information, and station masking modifies the effective footprint seen by the graph branches.

All input features are standardised before being passed to the network. Initial checks with the burn sample indicated that the feature means and standard deviations obtained after augmentation are more compatible with those observed in data than the corresponding quantities obtained from the idealised simulated sample alone. This provides an additional motivation for including detector-level augmentation in the training pipeline, since it improves the compatibility between the simulated training domain and the real data domain to which the discriminator will eventually be applied.

The convergence history of the three output heads is shown in Fig. 8.2. This example considers a case where the augmentation is not especially aggressive towards the UMD, as other cases are discussed further in the chapter. The top panel displays the evolution of the binary cross-entropy loss for the training and validation samples, while the bottom panel shows the corresponding ROC-AUC values. The training and validation curves remain close to one another, indicating stable convergence and no evidence for strong overfitting. A clear performance hierarchy is already visible at this stage: the combined SD-UMD head achieves the best validation performance, followed by the UMD-only head, while the SD-only head remains less powerful. This ordering is physically sensible and anticipates the independent performance studies discussed below, where the direct muon information provided by the UMD is shown to play a central role and its combination with SD observables yields the strongest overall separation.

The preceding sections defined the event representation, the network architecture, and the training procedure. We now turn from construction to validation. The following sections summarise the discrimination power of the resulting scores, their comparison with a traditional muon-based observable, their robustness under degraded UMD information, and the physical interpretation obtained from explainability studies.

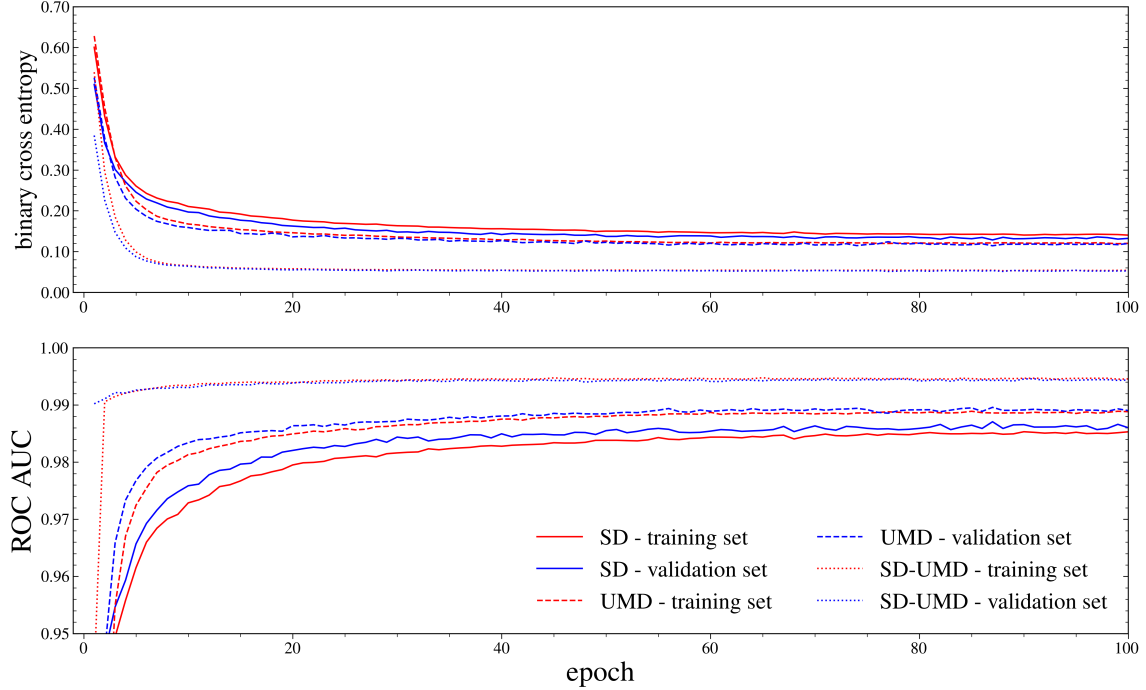


Figure 8.2: Example learning curves of the three discrimination branches: UMD-only, SD-only, and the combined SD-UMD path. The detector configuration considers the ideal case where all SD stations and UMD counters are available. The top panel shows the evolution of the binary cross-entropy loss, while the bottom panel shows the corresponding ROC-AUC, for both training and validation samples as a function of epoch.

8.5 Discrimination performance

Throughout this section and the following performance studies, all checks are performed on the photon reconstructed-energy scale, E_γ . Unless stated otherwise, the selected events satisfy $\theta < 52^\circ$ and have an estimated photon trigger efficiency larger than 0.9. This selection restricts the study to the region where the photon reconstruction is sufficiently efficient and where the comparison between signal efficiency and residual hadronic contamination is meaningful. The photon trigger-efficiency estimate and the energy reconstruction procedure, including the dedicated calibration used to assign E_γ , are described in Chapter 7.

We first evaluate the discrimination power of the three event-level scores defined by the architecture: the SD-only, UMD-only, and combined SD-UMD scores. Larger scores correspond to events that are more photon-like according to the network. The three branches are useful for different reasons. The SD-only score tests how much separation can be obtained from the spatial and temporal structure of the WCD footprint alone. The UMD-only score isolates the information carried by the direct muon measurements. The combined score tests the performance of the full event representation, in which both detector systems are allowed to contribute.

A first overview is provided by the score distributions in Fig. 8.3. In this figure, all available SD and UMD stations are used, corresponding to the most favourable detector-information scenario. In all three branches, photons and hadrons populate different regions of score space, with larger values corresponding to more photon-like events. The hierarchy of performance is already visible at this stage: the combined SD-UMD score provides the cleanest separation, the UMD-only score retains substantial discrimination power on its own, and the SD-only score shows a weaker but still non-negligible separation. This ordering is

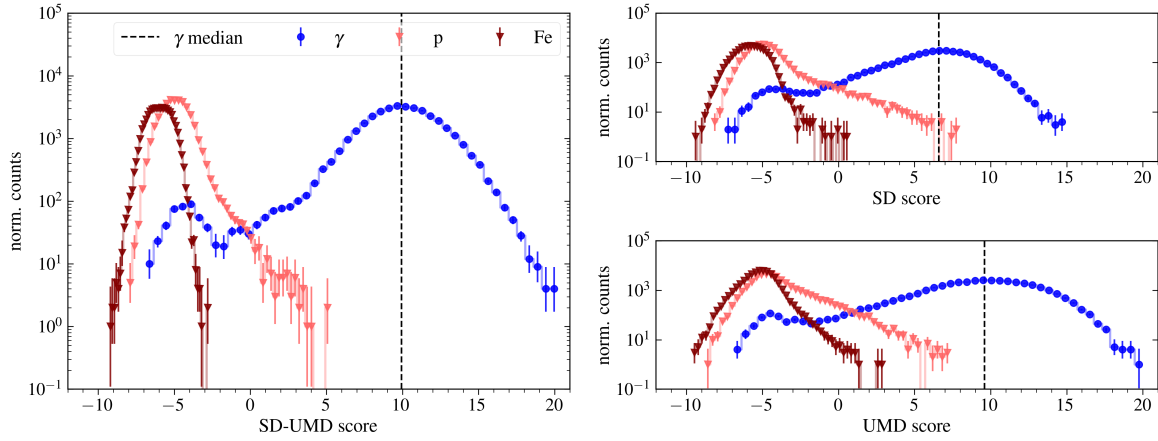


Figure 8.3: Distributions of the event-level discrimination scores for the three branches of the network on the independent test sample for events with $E_\gamma > 10^{16.5}$ eV. The large left panel shows the combined SD-UMD score, while the two panels on the right show the corresponding SD-only and UMD-only scores. Blue markers correspond to simulated photons, light red markers to simulated proton showers, and dark red markers to simulated iron showers. Larger scores indicate a more photon-like classification. The vertical dashed lines mark the median of the photon distribution in each case and are shown only as a visual reference.

physically sensible. The direct muon information measured by the UMD is expected to be one of the strongest handles for photon–hadron discrimination, while the SD contributes complementary information through the lateral, timing, and signal structure of the shower footprint.

The tails of these distributions are particularly important. A small fraction of photon showers can be assigned relatively hadron-like scores, while a small fraction of proton showers can be assigned relatively photon-like scores. These cases are not unexpected. Hadron-like photon showers are usually associated with photonuclear interactions in the first generations of the extensive air shower [189]. In this case, part of the primary photon energy is transferred to hadronic secondaries, producing charged mesons that can subsequently decay into muons. The resulting shower contains a larger muon component than a typical photon shower and can therefore appear more hadron-like to the network. Conversely, photon-like proton showers can arise from shower-to-shower fluctuations in the first hadronic interactions. In particular, a proton-induced shower has a non-negligible probability of producing a high-energy neutral pion early in the cascade. Since the π^0 decays promptly into photons, a large fraction of the shower energy can be channelled into the electromagnetic component, leading to a comparatively muon-poor shower footprint that resembles the photon hypothesis [170].

The inclusion of iron showers in the same comparison is also informative. Although the network was trained only against protons, the iron distributions are shifted even farther away from the photon population than the proton distributions. This behaviour is consistent with the expectation that protons constitute the most challenging hadronic background class. It also suggests that the network is learning physically meaningful correlations rather than features specific to the training label alone: showers that are more muon-rich and develop earlier in the atmosphere are assigned more hadron-like scores.

In the remainder of this chapter, we focus mainly on the combined SD-UMD score. This choice is motivated by the fact that it consistently provides the strongest photon–hadron separation among the three scores and is therefore the most relevant observable for the final search. The corresponding score distributions in reconstructed-energy intervals are shown

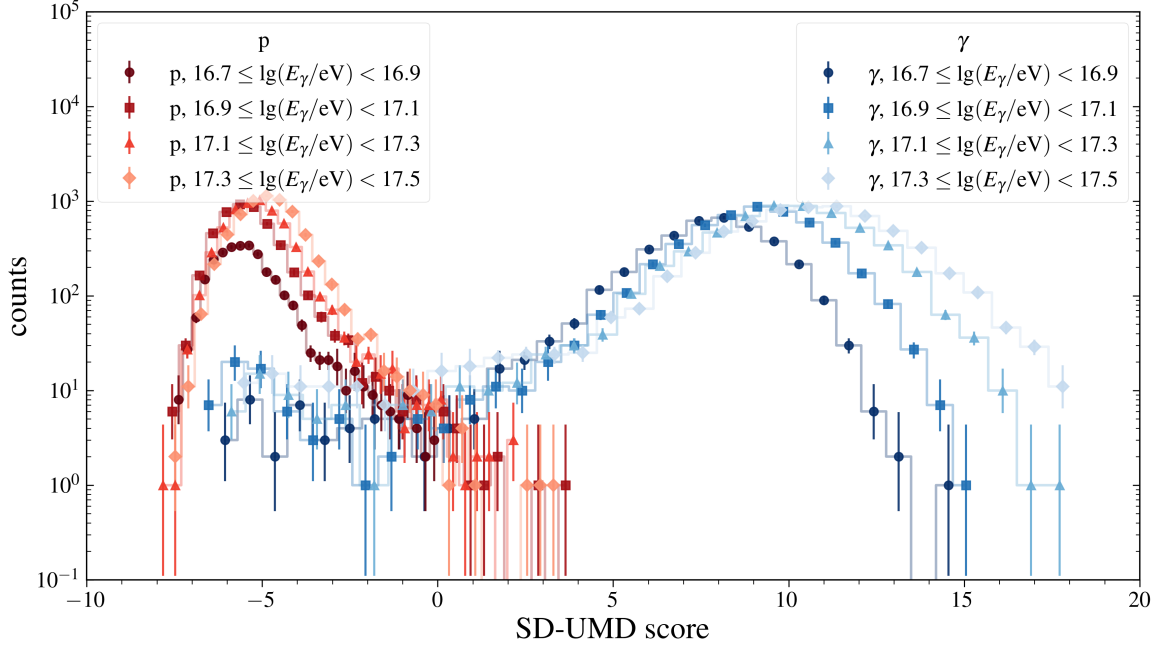


Figure 8.4: Energy-binned distributions of the combined SD-UMD discrimination score for simulated photons and proton showers. The events are grouped in reconstructed photon-energy intervals. Blue markers show photons and red markers show protons, with different marker styles denoting the different energy bins.

in Fig. 8.4. The same qualitative behaviour is observed in all bins: the proton distributions peak at lower, or even negative, scores, while the photon distributions are shifted towards larger, more photon-like values. The overlap between the two classes is mainly controlled by the extreme photon-like tail of the proton distribution and by the hadron-like tail of the photon distribution.

For both protons and photons, the score distributions shift slightly towards larger values with increasing energy, although the shift is somewhat more pronounced for photons. Part of the apparent energy dependence of the proton-score distributions may also be related to selection effects close to the trigger-efficiency threshold. In particular, at the lower edge of the analysed energy range, the selected proton sample need not be representative of all proton showers at the same primary energy. Showers that develop earlier in the atmosphere, corresponding to smaller $X_{\max}$, may leave a weaker electromagnetic footprint at ground and therefore be less likely to trigger or pass the event selection. If such events are correlated with more hadron-like network scores, the lower-score region of the selected proton distribution can be affected by this acceptance bias. This interpretation remains qualitative here, since a dedicated study of the correlation between the score, $X_{\max}$, and the trigger efficiency is beyond the scope of this chapter.

For a photon search, however, the most important performance quantity is not simply the visual separation between the bulk of the signal and background distributions. The relevant question is how many hadronic showers survive in the extreme photon-like region after imposing a threshold on the discriminator, while retaining a given fraction of true photons. This task is statistically demanding because the background leakage is controlled by the far tail of the hadronic score distribution, precisely where the number of available events is smallest.

This limitation from scarce statistics is especially important in machine-learning-based analyses, because proper performance evaluation must rely on simulated events that were

not used during either training or validation. To mitigate the limited tail statistics, the proton score distribution is modelled in reconstructed-energy bins of width $\Delta \lg(E_\gamma/\text{eV}) = 0.2$. In each energy bin, the proton tail in the vicinity of the photon distribution is described by an unbinned exponential fit to the most photon-like part of the proton distribution. The fitted conditional tail model is

$$f_{\text{tail}}(s \mid s \geq s_{\min}) = \lambda \exp[-\lambda(s - s_{\min})], \quad s \geq s_{\min}, \quad (8.29)$$

where s denotes the combined SD-UMD score, $s_{\min}$ is the lower bound of the fitted tail, and λ is the exponential slope. For a fixed value of $s_{\min}$, the unbinned maximum-likelihood estimator of the slope is

$$\hat{\lambda} = \frac{N_{\text{tail}}}{\sum_{i=1}^{N_{\text{tail}}} (s_i - s_{\min})}, \quad (8.30)$$

with the sum running over the proton events satisfying $s_i \geq s_{\min}$.

The lower bound $s_{\min}$ is determined from the proton scores themselves. Candidate values are obtained by scanning high-score quantiles of the proton distribution,

$$s_{\min}(q) = Q_q(\{s_p\}), \quad (8.31)$$

where Q_q is the empirical quantile of the proton scores in the corresponding reconstructed-energy bin. For each candidate q , the exponential slope is fitted with Eq. (8.30), and the agreement between the empirical tail distribution and the exponential model is evaluated with a Kolmogorov–Smirnov distance,

$$D_{\text{KS}}(q) = \sup_{s \geq s_{\min}(q)} |\hat{F}_{\text{tail}}(s) - [1 - \exp[-\hat{\lambda}(q)(s - s_{\min}(q))]]|. \quad (8.32)$$

The preferred tail start is then chosen as the candidate with the smallest D_{KS} , provided that the fitted region contains at least 50 tail events. Since the proton score distributions are highly concentrated at low scores, the relevant photon-like tails typically begin at quantiles above the 96th percentile in most reconstructed-energy bins.

The proton survival probability above a score threshold $s_{\text{cut}} \geq s_{\min}$ is then estimated as

$$P_p(s \geq s_{\text{cut}}) = \frac{N_{\text{tail}}}{N_p} \exp[-\lambda(s_{\text{cut}} - s_{\min})], \quad (8.33)$$

where N_{tail} is the number of proton events in the fitted tail and N_p is the total number of proton events in the corresponding energy bin.

An example of the tail fit is shown in the left panel of Fig. 8.5. The vertical dashed line marks the fitted value of $s_{\min}$. The fit describes only the proton tail to the right of this value; the lower-score part of the proton distribution is not used in the extrapolation. The uncertainty associated with the fit can appear relatively small when viewed directly on the score histogram. However, the relevant quantity for the search is the extrapolated survival probability, which depends exponentially on the fitted slope. Therefore, even a modest uncertainty in the tail description can propagate into a sizeable uncertainty in the estimated background contamination.

For the signal component, no tail model is used. The photon efficiency is computed directly from the empirical cumulative distribution of the simulated photon scores,

$$\varepsilon_\gamma(s_{\text{cut}}) = \frac{N_\gamma(s \geq s_{\text{cut}})}{N_\gamma}, \quad (8.34)$$

in the same reconstructed-energy bin. Combining Eq. (8.33) and Eq. (8.34) gives the relation between retained signal efficiency and residual proton contamination shown in the right

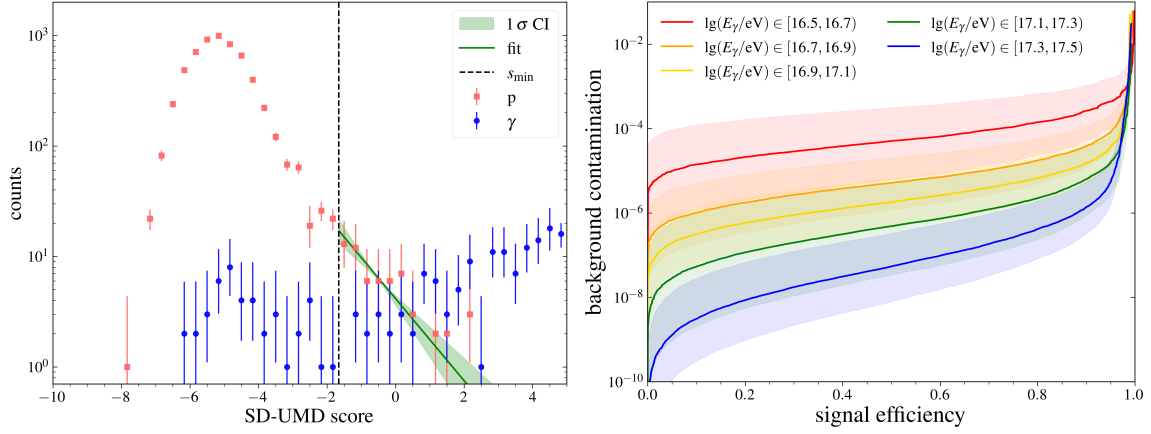


Figure 8.5: Left: example of the exponential fit to the photon-like tail of the proton SD-UMD score distribution for the reconstructed-energy interval $17.1 \leq \lg(E_\gamma/\text{eV}) < 17.3$. The dashed vertical line marks the minimum score $s_{\min}$ from which the tail model is applied. The red curve shows the fitted exponential model and the shaded region indicates the fit uncertainty. Right: estimated proton-background contamination as a function of photon signal efficiency for the combined SD-UMD discriminator, shown in reconstructed photon-energy intervals. The background survival probability is obtained from the exponential tail model, while the photon efficiency is computed directly from the empirical photon score distribution. The shaded bands show the propagated uncertainty from the tail-fit procedure.

panel of Fig. 8.5. The shaded bands correspond to the propagation of the tail-fit uncertainty through the contamination estimate. They show that, especially at low statistics, small differences in the fitted tail can translate into large changes in the inferred background rejection.

Several conclusions follow from these curves. First, the combined SD-UMD observable achieves a strong suppression of the proton background over the reconstructed-energy range considered here. Second, the rejection power generally improves with energy, consistently with the trend seen in the score distributions. This ordering is physically reasonable: higher-energy showers tend to have larger footprints and more information in both the SD and UMD components, while the relative difference between photon and hadron muon content also becomes easier to exploit.

At the same time, the uncertainty bands caution against overinterpreting small differences between neighbouring energy intervals. In the central curves, the higher-energy bins follow the expected ordering and usually show lower background contamination at fixed signal efficiency. However, the bands overlap substantially in parts of the high-efficiency region. This means that, with the present independent-test statistics, it is difficult to claim that the discriminator is strictly superior in one high-energy bin than in another based only on small separations between the central curves. The more robust conclusion is that the method provides strong background suppression and that the performance improves broadly with energy.

For reference, given the geometrical area spanned by the SD-433 array, the expected exposure over the lifetime of the experiment requires a discriminator with a residual background contamination of order 10^{-5} . The contamination curves show that, for the bins in the range $16.5 \leq \lg(E_\gamma/\text{eV}) < 17.5$, this level is reached around a photon efficiency of approximately 0.5, although the uncertainty of this estimate remains considerable. The lowest-energy bin also corresponds to less efficient showers and reduced statistics. For this reason, and to remain conservative, the rest of the analysis is restricted to energies above that bin.

8.6 Comparison with M_1

Having established the performance of the network under ideal detector-information conditions, the natural question is whether it improves over the existing M_1 observable used in the SD-433+UMD photon search [52]. This observable is based on the event-wise combination of the UMD muon-density measurements. In its general form,

$$M_b = \lg \left[\sum_i \frac{\rho_i}{\rho_{\text{pr}}(E, \theta)} \left(\frac{r_i}{r_{\text{pr}}} \right)^b \right], \quad (8.35)$$

where ρ_i is the muon density measured by the i -th UMD station at core distance r_i , ρ_{pr} is the expected muon density for proton-induced showers at the reference distance $r_{\text{pr}} = 200$ m, and b controls the radial weighting. The optimised choice $b = 1$ defines the M_1 observable. By construction, photon-like events tend to have lower M_1 values, because they contain fewer muons than the proton expectation, whereas hadronic events tend to have larger values.

A fully quantitative comparison is not straightforward. The published M_1 analysis used a specific simulation and reconstruction chain, including a different CORSIKA version and an Offline configuration tuned for that study. In addition, the published performance curves are presented in an equivalent energy scale, intermediate between the photon and hadron reconstructed-energy scales, while the present comparison is performed in the absolute photon reconstructed-energy scale, E_γ . This choice is deliberate. The reconstructed energy is not used as an input to the GNN, and therefore the discriminator itself is independent of the particular energy parametrisation used to bin events or define a final threshold. For the present photon-search application, it is therefore natural to express the comparison in terms of E_γ , consistently with the signal hypothesis.

For reference, we also tested a fast re-parametrised version of M_1 adapted to the present simulation library. This version is not intended to replace the published observable or to provide a full retuning of the M_1 analysis. Moreover, a redefinition for the Phase II has already been established and studied [190]. Here we only use it to study the qualitative behaviour of the observable under controlled modifications of the available UMD information. When using this recomputed M_1 observable on the present simulations, we restrict the comparison to events with $E_\gamma > 10^{17}$ eV and $\theta < 40^\circ$. This is the region in which we found that the recomputed observable approximately reproduces the behaviour reported in the published analysis. Outside this restricted range, differences between simulation libraries, detector simulations, reconstruction settings, and energy-scale conventions become more apparent.

Before comparing discrimination power, it is useful to inspect how M_1 correlates with the scores produced by one trained network. This comparison is shown in Fig. 8.6. The notation r_P denotes the Pearson linear correlation coefficient, while ρ_S denotes the Spearman rank correlation coefficient. The strongest correlation is observed for the UMD-only score, with $r_P = 0.917$ and $\rho_S = 0.853$. This is expected, because both M_1 and the UMD branch are driven directly by the measured muon content. The combined SD-UMD score is also strongly correlated with M_1 , although slightly less than the UMD-only score. This is consistent with the combined branch using the muon information together with complementary SD information.

The SD-only score also shows a sizeable correlation with M_1 , even though it does not directly use the UMD muon densities. This should be interpreted with some care. It does not imply that the SD branch measures the muon content directly. Rather, it indicates that the spatial, timing, signal, and trace information recorded by the SD carries shower-development information that is correlated with the muon content. This is physically plausible, since more hadron-like showers tend to be more muon-rich and to develop differently in the

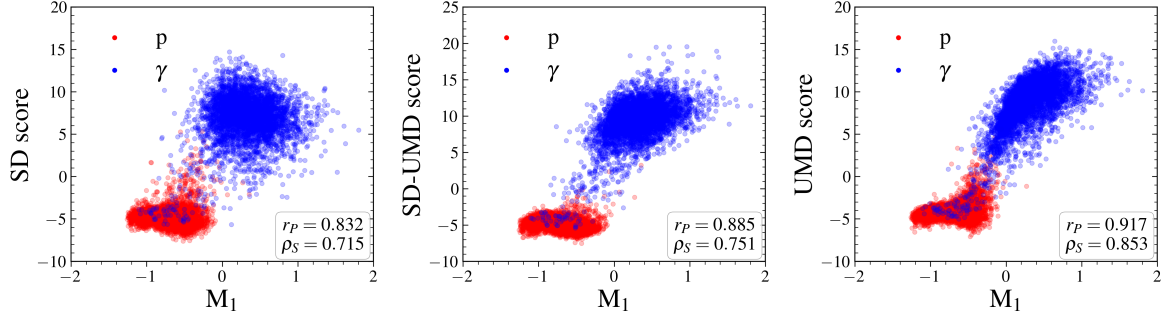


Figure 8.6: Correlation between the recomputed M_1 observable and the three scores produced by one trained network, evaluated on simulated photons and protons with $E_\gamma > 10^{17}$ eV and $\theta < 40^\circ$. The left panel shows the SD-only score, the centre panel shows the combined SD-UMD score, and the right panel shows the UMD-only score. Blue points correspond to photons and red points to protons. The quantities r_P and ρ_S denote the Pearson and Spearman correlation coefficients, respectively.

atmosphere, which can affect both the UMD densities and the SD footprint. The correlation therefore supports the interpretation that the network is learning physically meaningful shower features, but it should not be taken as evidence that the SD-only score is a substitute for a direct muon measurement.

To compare the combined SD-UMD score with M_1 , we use the difference between their estimated proton-background contamination curves, expressed in units of the uncertainty assigned to the SD-UMD contamination estimate. For each reconstructed-energy interval and photon signal efficiency ε_γ , we define

$$\Delta C(\varepsilon_\gamma) = \frac{C_{\text{SD-UMD}}(\varepsilon_\gamma) - C_{M_1}(\varepsilon_\gamma)}{\sigma_C(\varepsilon_\gamma)}, \quad (8.36)$$

where $C_{\text{SD-UMD}}$ is the proton contamination estimated from the combined network score, C_{M_1} is the corresponding contamination expected for the M_1 observable, and σ_C is the uncertainty assigned to the SD-UMD contamination curve from the tail-fit procedure. With this convention, $\Delta C < 0$ means that the network gives a lower estimated proton contamination than M_1 , while $\Delta C > 0$ means that M_1 gives the lower estimated contamination. The horizontal dashed lines at $\Delta C = \pm 1$ indicate the approximate one-standard-deviation range, while the dotted lines at $\Delta C = \pm 2$ are shown as an additional visual reference.

The resulting pull curves are shown in Fig. 8.7. The comparison should be interpreted cautiously, because the uncertainty entering the denominator is dominated by the limited statistics available in the photon-like tail of the independent proton test sample. Moreover, the pull is normalised only by the uncertainty assigned to the SD-UMD tail extrapolation. It should therefore be read as a diagnostic of the relative size of the two contamination estimates, rather than as a complete statistical test including all possible uncertainties of both observables.

The pull comparison shows that the relative behaviour of the two observables depends both on energy and on the photon efficiency at which the comparison is made. In the lowest reconstructed-energy bin, $16.7 \leq \lg(E_\gamma/\text{eV}) < 16.9$, the pull is positive over a broad low- and intermediate-efficiency range, reaching values above +1. This provides evidence that, in this part of the curve, the M_1 observable achieves a lower estimated proton contamination than the combined SD-UMD network score. However, the same bin also shows a clear change of behaviour at larger photon efficiencies: the pull crosses zero and becomes increasingly negative, eventually reaching values below -1 and -2 . In that regime, the network gives the lower contamination estimate.

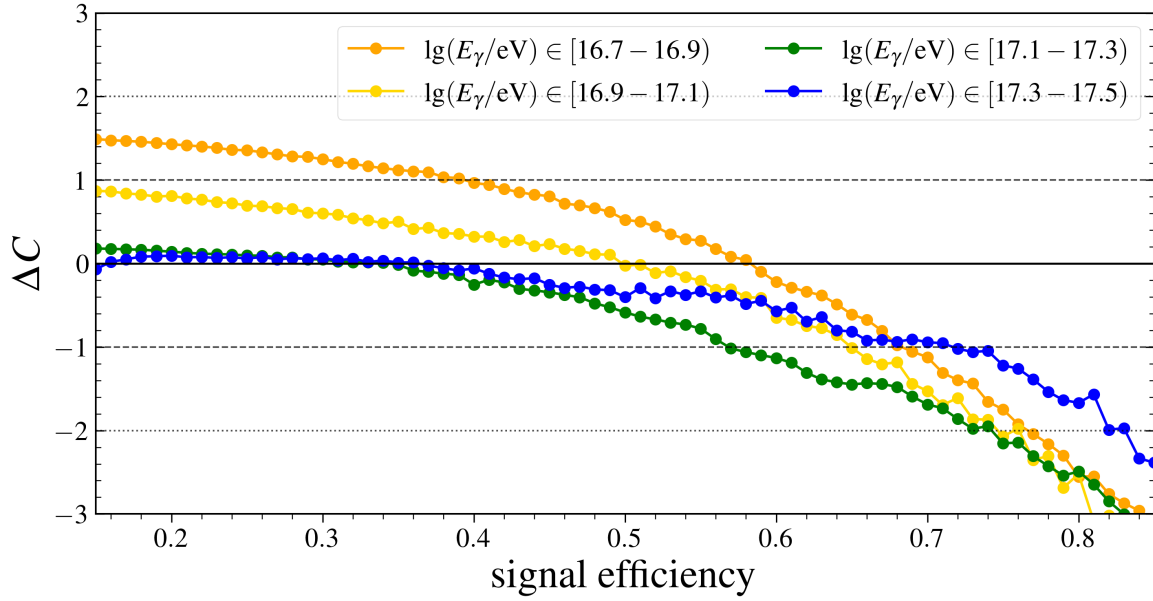


Figure 8.7: Pull comparison between the proton-background contamination obtained with the combined SD-UMD score and with the M_1 observable. The vertical axis shows $(C_{\text{SD-UMD}} - C_{M_1})/\sigma_C$, where $C_{\text{SD-UMD}}$ is the contamination estimated from the network score, C_{M_1} is the contamination estimated from the M_1 observable, and σ_C is the uncertainty of the network contamination curve from the tail-fit procedure. Positive values indicate that M_1 gives a lower estimated proton contamination than the network, while negative values indicate that the network gives a lower estimated contamination than M_1 . The dashed and dotted horizontal lines mark $\pm 1\sigma_C$ and $\pm 2\sigma_C$, respectively, as visual references.

The intermediate-energy bins show a less decisive pattern. For $16.9 \leq \lg(E_\gamma/\text{eV}) < 17.1$, the pull is positive at low efficiencies, remains close to zero around intermediate efficiencies, and becomes negative at higher efficiencies. For $17.1 \leq \lg(E_\gamma/\text{eV}) < 17.3$, the curve is close to zero at low efficiencies and then becomes increasingly negative as the photon efficiency increases. The highest-energy bin, $17.3 \leq \lg(E_\gamma/\text{eV}) < 17.5$, remains mostly within the $\pm 1\sigma_C$ band over a large part of the range, before becoming more clearly negative at the largest efficiencies. Therefore, except for the most pronounced regions of the curves, the uncertainty makes a definitive bin-by-bin claim unreliable.

The sign and evolution of the pull can also be interpreted in terms of the relative shapes of the proton tails for the two observables. Positive values of ΔC indicate that, at the threshold corresponding to a given photon efficiency, the proton tail selected by the network is larger than the one selected by M_1 . In those regions, M_1 provides the stronger background rejection. Negative values indicate the opposite: the network score assigns fewer proton events to the photon-like tail than M_1 does at the same photon efficiency. The observed sign changes therefore suggest that the two discriminators do not order the photon-like tail of the proton population in exactly the same way. In particular, M_1 appears more favourable in the stringent, low-efficiency part of the lowest-energy bin, while the network becomes more favourable once the selection is loosened towards higher photon efficiencies.

Under ideal detector-information conditions, the safest conclusion is not that one method is globally superior, but rather that the graph-network score and M_1 provide comparable discrimination power, with M_1 favoured in some low-efficiency regions and the network favoured at higher efficiencies, particularly where the pull becomes significantly negative. A stronger quantitative statement would require evaluating both methods with the same

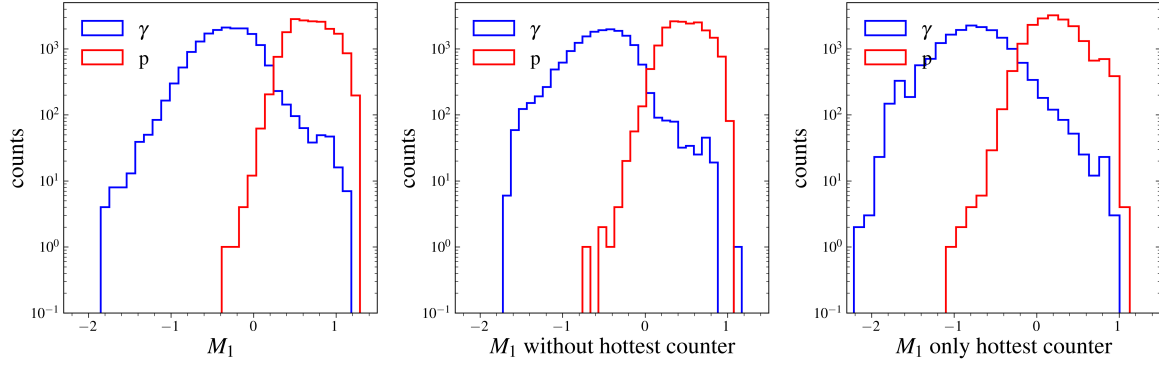


Figure 8.8: Qualitative behaviour of the re-parametrised M_1 observable under different UMD-information scenarios. Left: nominal M_1 using the available UMD counters. Centre: M_1 recomputed after removing the counter associated with the hottest SD station. Right: M_1 recomputed using only the hottest counter. Blue histograms correspond to simulated photons and red histograms to simulated protons. Photon-like events populate lower M_1 values, while proton-like events populate larger values. The increasing overlap in the reduced-information scenarios illustrates the dependence of M_1 on the available UMD sampling of the shower footprint.

shower library, detector simulation, reconstruction chain, event selection, and energy-scale convention, as well as a consistent treatment of the tail uncertainties for both observables.

If the nominal rejection power is comparable, the relevant question becomes different: why consider graph neural networks at all? The answer is that the network is not expected to improve the search primarily by producing a much stronger discriminator under ideal detector conditions. Its main potential advantage is operational. The M_1 observable relies directly on the measured UMD muon densities entering the event-wise sum in Eq. (8.35). It therefore depends on the available UMD sampling of the shower footprint and on the treatment of missing or incomplete counter information. This structure is physically well motivated, but it naturally ties the observable to detector configurations with sufficiently informative UMD coverage.

The network was designed to operate under more irregular detector patterns. It receives the event as a graph, uses validity masks to distinguish available from missing detector information, and combines direct muon information with SD signal and timing information. This design makes it possible, at least in principle, to retain useful discrimination power even when the UMD information is incomplete. The relevant comparison is therefore not only the best-case rejection power, but also the stability of the discriminator when the available detector information is degraded.

To illustrate this point, Fig. 8.8 shows three variants of the re-parametrised M_1 observable. The first uses the available UMD counters in the nominal configuration. The second removes the counter associated with the hottest SD station. The third keeps only the hottest counter. In the nominal case, the photon and proton distributions are clearly separated, with photons shifted towards lower M_1 values and protons towards larger ones. Removing the hottest counter changes the distributions but still leaves a visible separation. In the hottest-counter-only case, the separation is more strongly degraded: the distributions broaden and the overlap between photons and protons increases. This behaviour is expected, because M_1 was designed as a muon-content observable constructed from the sampled UMD footprint, not as a single-counter classifier.

The main motivation for the graph-based approach is therefore complementary. If the network remains reliable under less favourable UMD configurations, it may allow a larger fraction of the data set to be used in the photon search. In that case, the improvement would not necessarily come from a much lower background contamination at fixed exposure, but

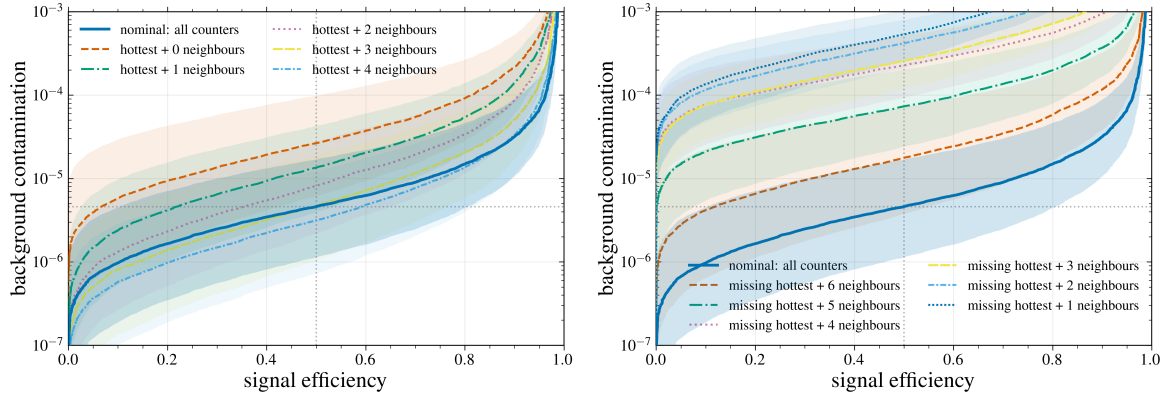


Figure 8.9: Proton-background contamination as a function of photon signal efficiency for the combined SD–UMD score under controlled modifications of the available UMD information. Left: configurations in which the UMD information is restricted to the hottest counter and progressively enlarged by adding neighbouring counters. Right: configurations in which the hottest counter is removed and the available UMD information is provided by neighbouring counters. The nominal case, using all available counters, is shown in both panels for reference. Shaded regions indicate the propagated uncertainty from the tail-fit procedure.

from a more flexible use of events with incomplete UMD information, potentially increasing the effective exposure while maintaining acceptable rejection power.

8.7 Robustness

In the present context, robustness refers to the operational stability of the discriminator. The score should preserve a meaningful ordering between photon- and hadron-induced showers across the relevant phase space and under plausible degradations of the detector information, rather than relying on a narrow region of phase space or on a single fragile feature. To test this behaviour, we repeat the contamination-curve procedure described in Section 8.5 after applying controlled modifications to the available UMD information. Two complementary extreme scenarios are considered. In the first one, shown in the left panel of Fig. 8.9, the UMD information is restricted to the counter associated with the hottest SD station and is then progressively enlarged by adding neighbouring counters. In the second one, shown in the right panel, the opposite situation is studied: the hottest counter is removed, and the discrimination is tested using configurations in which the available local UMD information comes from neighbouring counters instead.

The left panel shows that the degradation is surprisingly mild when the hottest counter is retained. As expected, the nominal configuration gives one of the strongest background rejection across the variants, and the performance improves as more neighbouring counters are added. However, even the most restrictive configurations retain substantial discrimination power. In particular, keeping only the hottest counter, or the hottest counter plus a small number of neighbours, still leads to background-contamination levels close to the range required for the search at moderate photon efficiencies. For reference, given the exposure expected for the combined SD-433 and UMD-433 array, a residual background contamination of order $2.5 \cdot 10^{-5}$ at an efficiency of 0.5 is a relevant target scale. The retained-hottest configurations remain close to this level, showing that the network does not require the full UMD footprint to preserve useful rejection power.

The right panel shows a much stronger degradation when the hottest counter is missing. In these configurations, the contamination curves move upwards over most of the

efficiency range, indicating that more proton events survive in the photon-like tail of the score distribution. Adding neighbouring counters partially recovers the performance, but the recovery is slower than in the retained-hottest case. This suggests that the information associated with the counter closest to the largest SD signal plays a special role in the learned discriminator. The effect of UMD coverage is still clearly visible, since the curves improve when more local muon information is available. Nevertheless, the comparison between the two panels indicates that the degradation is not controlled only by the number of counters. It also depends strongly on whether the local muon measurement associated with the hottest part of the SD footprint is present.

This behaviour is particularly interesting when compared with the M_1 observable discussed in Section 8.6. For M_1 , the hottest-counter-only configuration was strongly degraded, as expected for an observable designed as an event-wise sum over the sampled UMD footprint. The network shows the opposite qualitative trend: retaining the hottest counter while discarding much of the surrounding UMD information degrades the performance more slowly than removing the hottest counter and relying on neighbouring counters. This does not mean that the network is insensitive to UMD coverage. Rather, it suggests that the combined SD-UMD branch has learned a more local decision pattern, in which the coincidence between the largest SD signal and the corresponding muon information carries a large fraction of the separation power. The physical origin of this behaviour requires a more detailed inspection of what the network uses internally.

8.8 Explainability

Explainability refers to *post hoc* tools that help identify which stations, subdetectors, input features, or trace bins contribute most strongly to a given prediction, without implying that the full network can be reduced to a simple analytic rule. An informative discussion of explainability can be obtained by combining *event-level* explanation maps with *dataset-level* attribution summaries. These two levels answer different but complementary questions. At the event level, one asks why the network assigned a particular score to a specific reconstructed shower. At the dataset level, one asks whether stable patterns emerge when the same analysis is repeated over many showers. The first perspective provides concrete illustrative examples; the second shows whether those examples reflect a general behaviour of the model rather than an anecdotal one.

This distinction is especially important for the graph neural networks developed here, where the relevant information is distributed over stations, scalar node features, and SD time traces. A single-event visualization can reveal where the model is focusing, but its detailed structure may fluctuate from event to event. Dataset-level summaries are therefore needed to identify the patterns that persist statistically and can be interpreted as robust properties of the learned discriminator. For this reason, we proceed in two steps. We first construct explanation maps for representative individual events, and then average related quantities over the full test set in order to obtain a more stable picture of the network behaviour. A particular question is motivated by the robustness study in Section 8.7: why does the combined score degrade more slowly when the UMD counter associated with the hottest SD station is retained than when this counter is removed? The explainability analysis provides a way to address this question by checking whether the network assigns a special role to the hottest-station region, or whether the observed robustness is only an incidental consequence of the reduced-information tests.

Let $s^{(g)}$ denote the combined SD–UMD score assigned by the network to an event graph g . To estimate which inputs are most influential for that score, we use a local gradient-based attribution measure. For scalar node features we define

$$a_{if}^{(g)} \equiv \frac{\partial s^{(g)}}{\partial x_{if}^{(g)}} x_{if}^{(g)}, \quad (8.37)$$

where $x_{if}^{(g)}$ is the standardized value of feature f at node i . The score $s^{(g)}$ is first obtained from a standard forward pass of the network, and the derivatives $\partial s^{(g)} / \partial x_{if}^{(g)}$ are then computed by automatic differentiation in the corresponding backward pass. The resulting gradient $\times$ input quantity provides a computationally efficient local approximation to an integrated-gradient-like attribution: it measures how strongly the realised value of a given input component affects the score in the neighbourhood of the observed event.

Likewise, for the SD trace time bins we define

$$\tau_{it}^{(g)} \equiv \frac{\partial s^{(g)}}{\partial u_{it}^{(g)}} u_{it}^{(g)}, \quad (8.38)$$

where $u_{it}^{(g)}$ denotes the trace amplitude of node i in time bin t . The trace attribution therefore identifies which parts of the waveform most strongly influence the final score. At the event level we retain the sign of these quantities, because it is informative to know whether a given station, feature, or trace segment contributes in the photon-like or hadron-like direction. At the dataset level, by contrast, we work mostly with absolute values, $|a_{if}^{(g)}|$ and $|\tau_{it}^{(g)}|$, because the goal is to quantify how strongly the classifier tends to rely on a variable across many showers, rather than whether the contribution is positive or negative in one particular case.

A second ingredient of the event-level visualization is the node-importance map derived from graph-attention coefficients. For a non-expert reader, it is useful to think of these coefficients as describing how strongly the network weights the information exchanged between neighbouring stations when building its internal representation of the event. They do not by themselves explain the full final decision, but they do indicate where the model concentrates its message passing and which parts of the graph are most strongly involved in the learned representation. Here it is sufficient to note that graph-attention layers provide a natural way to identify the subset of stations on which the network is focusing most strongly.

If $\bar{a}_{ij}^{(\ell)}$ denotes the attention coefficient averaged over attention heads for edge (i, j) in layer ℓ , we define a node score

$$\omega_i^{(g)} = \frac{1}{L} \sum_{\ell=1}^L \frac{1}{2} \left[\sum_{j:(j,i) \in E_g} \bar{a}_{ji}^{(\ell)} + \sum_{j:(i,j) \in E_g} \bar{a}_{ij}^{(\ell)} \right], \quad (8.39)$$

which is then rescaled to its maximum value for visualization. The corresponding event-level attention-weighted attribution matrix is defined as

$$h_{if}^{(g)} \equiv \omega_i^{(g)} a_{if}^{(g)}. \quad (8.40)$$

The event-level scalar feature ranking is then obtained from

$$I_f^{(g)} = \frac{\sum_i |h_{if}^{(g)}|}{\sum_{i,f'} |h_{if'}^{(g)}|}. \quad (8.41)$$

This construction highlights the feature channels that matter *within the subset of nodes on which the network is actually concentrating*, while suppressing stations that play only a minor role in the decision.

We begin with a representative event taken from the photon-like tail of the proton distribution, shown in Fig. 8.10. This example is not intended to be statistically representative by itself. Rather, it is chosen to illustrate the kind of explanatory pattern that the network produces in individual showers and to motivate the dataset-level averages introduced afterwards. The explanation is strongly localized. The detector-plane panel shows the attention-derived node importance $\omega_i^{(g)}$, obtained from the averaged graph-attention coefficients, and indicates that the network does not distribute its focus uniformly over all triggered stations. Instead, it concentrates on a limited subset of nodes in the most informative region of the footprint. The hottest station is marked with a red circle, which helps relate the explanation to the reference frame used elsewhere in the analysis and to the reduced-UMD configurations studied in Section 8.7. The most important nodes are often located near the highest-WCD-signal region, indicating that the network tends to concentrate on the local neighbourhood around the dominant part of the shower footprint. At the same time, the remaining panels make clear that the network is not responding to signal size alone.

This is seen more explicitly in the feature-ranking and attribution panels. The bar chart shows the normalized event-level feature importance $I_f^{(g)}$, obtained from the attention-weighted attribution matrix $h_{if}^{(g)} = \omega_i^{(g)} a_{if}^{(g)}$. The heatmap displays the signed values of $h_{if}^{(g)}$ themselves, and therefore combines two ingredients: the local sensitivity of the score to each scalar node feature through $a_{if}^{(g)}$, and the node-level relevance through $\omega_i^{(g)}$. For this particular photon-like proton, the dominant scalar contribution is carried by the reconstructed muon density ρ_μ , while the remaining scalar variables are clearly subdominant.

In the trace panel, the leading station traces are shown in decreasing WCD signal. The grey curves correspond to the unnormalised traces, while the colour of each trace segment indicates the signed time-bin attribution $\tau_{it}^{(g)}$. The temporal attribution is concentrated mostly in the early part of the waveform and in a restricted subset of stations. For this event, part of the hadron-like contribution appears around relatively sharp structures in the traces, consistent with the expectation that muon-dominated signal components are more localized in time. The pattern should nevertheless be interpreted qualitatively, since the detailed trace-level attribution can vary from event to event.

This example supports an important qualitative conclusion. The combined SD-UMD branch is not behaving like a diffuse global average over the full graph. Rather, it behaves like a local-pattern recognizer that identifies a limited number of stations carrying the decisive joint SD-UMD information, and then combines scalar and temporal observables within those stations to form the final score. This provides a first qualitative explanation for the robustness behaviour observed previously: if much of the relevant information is concentrated around the hottest-station region, then retaining that local SD-UMD coincidence can preserve a large fraction of the discrimination power even when the surrounding UMD coverage is reduced.

Although such event-level views are highly informative, they remain anecdotal by construction. The next step is therefore to move from *how the network explains one event* to *how the network behaves on average*. To do so, we average unsigned attributions over the full test set in several complementary ways. These summaries allow us to answer progressively broader questions: which stations tend to dominate within a typical event, which scalar variables are most important overall, and which parts of the time traces carry the most useful discrimination information.

A first summary is obtained by ranking, within each event, the triggered stations according to decreasing WCD signal, so that rank $k = 1$ corresponds to the hottest station and

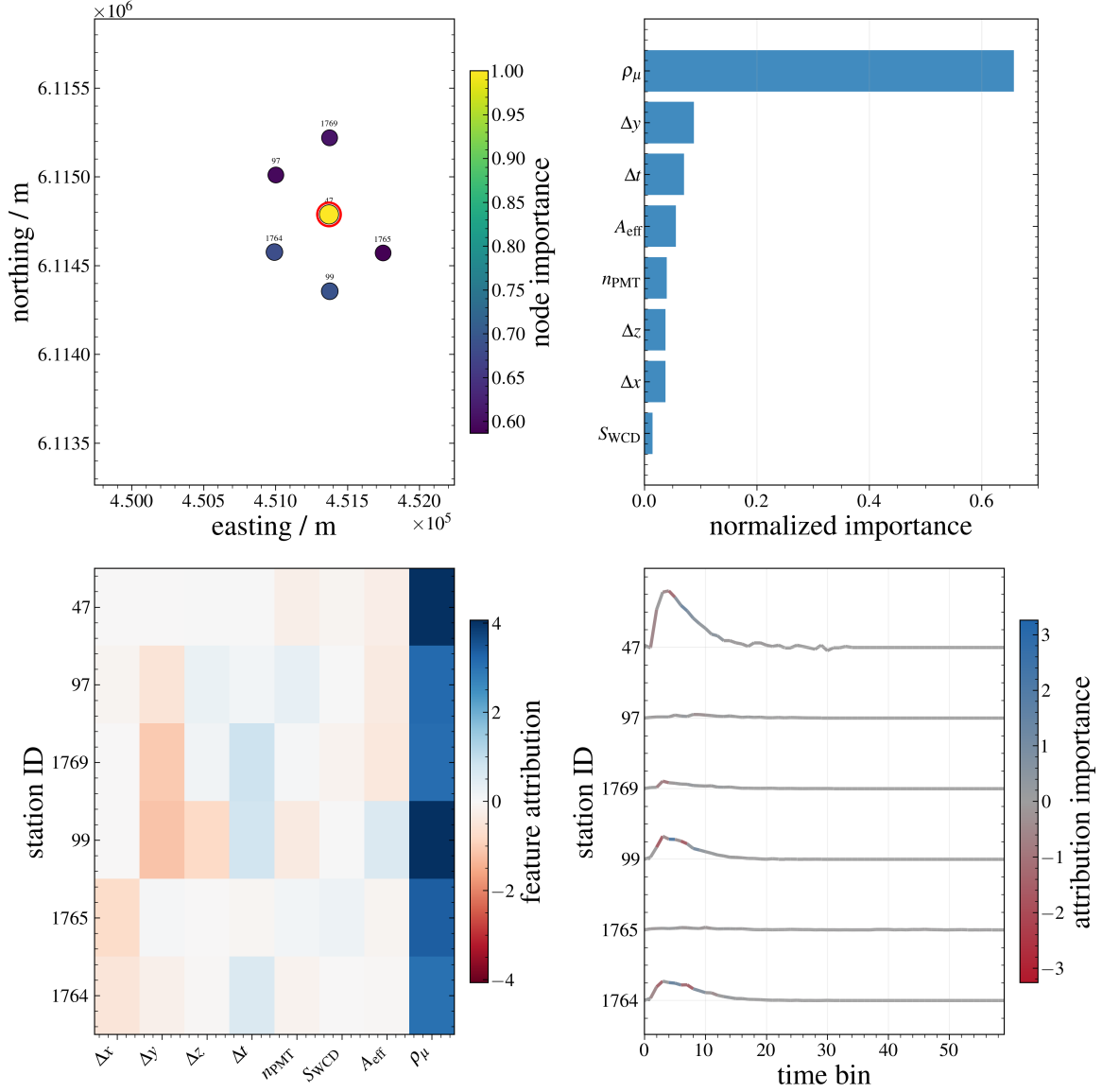


Figure 8.10: Event-level explanation for a representative simulated proton assigned a highly photon-like score by the combined SD-UMD branch. The event was measured by the north-east hexagon and belongs to the photon-like tail of the proton score distribution. Upper left: detector-plane projection coloured by the attention-derived node importance $\omega_i^{(g)}$; the hottest station is marked with a red circle. Upper right: normalized event-level scalar feature importance $I_f^{(g)}$. Lower left: signed attention-weighted node-feature attribution matrix $h_{if}^{(g)}$, with stations ordered by decreasing WCD signal. Lower right: unnormalised SD traces for the leading stations, also ordered by decreasing WCD signal; the colour of each trace segment indicates the signed time-bin attribution $\tau_{it}^{(g)}$.

larger k to progressively weaker stations. If $\pi_g(k)$ denotes the station of rank k in event g , the corresponding station-rank $\times$ feature attribution matrix is defined as

$$M_{kf} = \frac{1}{|\mathcal{D}_k|} \sum_{g \in \mathcal{D}_k} |a_{\pi_g(k)f}^{(g)}|, \quad (8.42)$$

where $\mathcal{D}_k$ is the subset of events with at least k valid stations. This construction does not follow a fixed detector identifier across events. Instead, it answers the more relevant question

of which variables dominate in the first, second, third, . . . most intense stations of a typical event. Because the ranking is based on the WCD signal, it also provides a natural bridge between the event-level visualization, where the hottest station was used as a reference, and the dataset-level picture.

A complementary summary is obtained by averaging the unsigned scalar attributions feature by feature over nodes and then over events. For each graph,

$$\bar{a}_f^{(g)} = \frac{1}{N_g} \sum_{i=1}^{N_g} |a_{if}^{(g)}|, \quad (8.43)$$

and the dataset-wide scalar feature attribution is

$$A_f = \frac{1}{|\mathcal{D}|} \sum_{g \in \mathcal{D}} \bar{a}_f^{(g)}. \quad (8.44)$$

The same construction is applied to the trace bins,

$$\bar{\tau}_t^{(g)} = \frac{1}{N_g} \sum_{i=1}^{N_g} |\tau_{it}^{(g)}|, \quad T_t = \frac{1}{|\mathcal{D}|} \sum_{g \in \mathcal{D}} \bar{\tau}_t^{(g)}. \quad (8.45)$$

These definitions give an event-weighted rather than node-weighted summary: each graph contributes once, regardless of how many stations it contains.

The station-rank heatmap in Fig. 8.11 shows that the explanatory weight is concentrated in the first few stations of each event, that is, in the subset of nodes with the largest WCD signal. This confirms at dataset level the localized behaviour already seen in the single-event map. The effect is particularly clear for the WCD signal itself: the attribution associated with S_{SD} is largest for the hottest station and decreases rapidly for lower-ranked stations. This indicates that, for the combined SD–UMD branch, the most informative signal-size information is mainly the signal measured by the hottest station, rather than a diffuse average over all stations.

The dataset-level matrix therefore suggests that the strongest separation is obtained from the combination of a large hottest-station signal and the corresponding local muon measurement. This interpretation is consistent with the robustness studies in Section 8.7, where the combined branch degraded more slowly when the hottest counter was retained than when it was removed. The explainability result therefore gives a natural explanation for the apparent reversal with respect to the M_1 observable: while M_1 benefits from summing over a broader sampled muon footprint, the network can exploit a localized SD–UMD pattern near the dominant part of the shower footprint. In that sense, the explainability analysis supports the same physical picture: a substantial part of the discrimination power appears to be carried by the local coincidence between the dominant SD signal and the muon information measured near the shower axis.

The timing offset Δt provides an additional secondary contribution, while the remaining geometry variables, PMT multiplicity, and UMD effective area are less important on average. This is an important result. It shows that the network is not merely selecting the hottest stations and then using all available inputs indiscriminately. Rather, it focuses on a restricted subset of physically informative stations and, within them, relies primarily on observables directly linked to muon content, the dominant WCD signal, and shower-front timing.

The dataset-wide scalar and trace summaries in Fig. 8.12 provide a complementary view. The scalar attribution ranking again identifies ρ_μ as the dominant input variable. The next most relevant quantities are Δt and S_{SD} , followed by weaker contributions from the relative station coordinates, A_{eff} , and n_{PMT} .

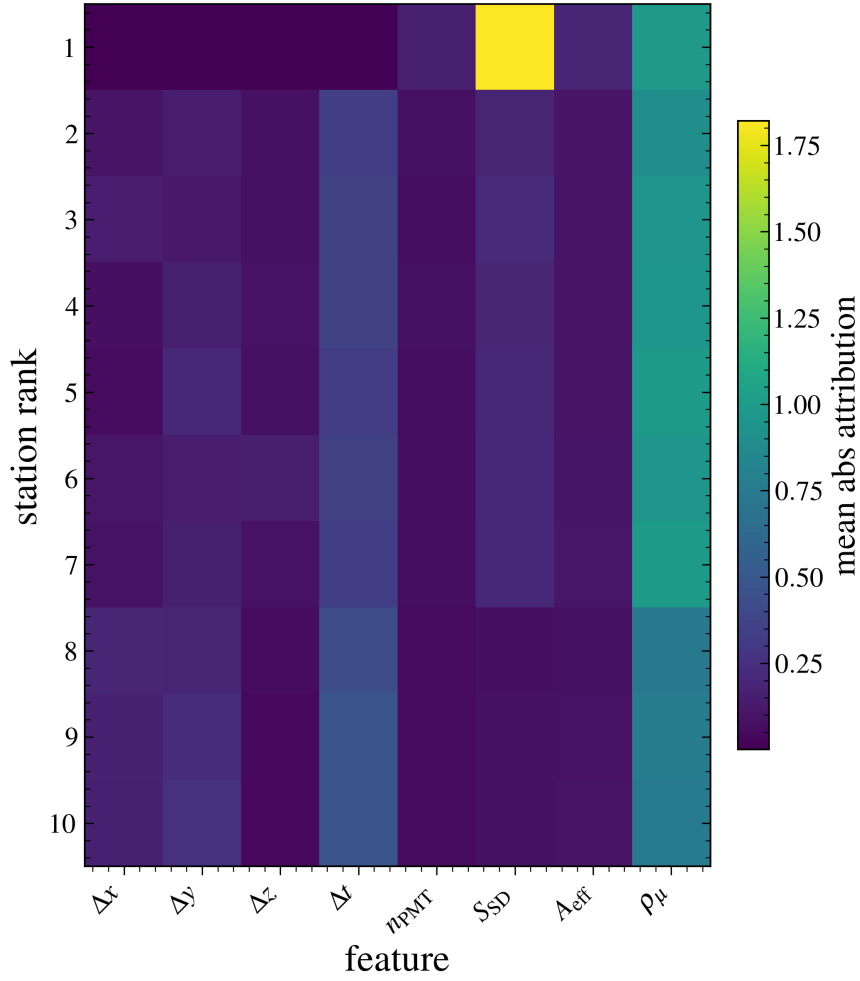


Figure 8.11: Dataset-wide station-rank scalar attribution matrix M_{kf} for the combined SD-UMD branch. For each event, stations are ranked by decreasing WCD signal, so that $k = 1$ corresponds to the hottest station. The colour scale shows the mean absolute gradient $\times$ input attribution of feature f at station rank k , averaged over events with at least k valid stations.

The trace-importance curve shows that the temporal attribution is not distributed smoothly over the full waveform. Instead, it is concentrated in a relatively narrow and structured region close to the beginning of the trace, with several pronounced peaks in the first time bins and a rapid decrease afterwards. This suggests that the network is particularly sensitive to localized early-time structures in the SD waveform. Such a pattern is consistent with the event-level maps discussed above, where some of the signed trace attribution was associated with sharp, localized components of the traces. These structures are likely related to muon signal contributions, which are expected to be more concentrated in time than the electromagnetic component.

At the same time, the fact that the attribution is concentrated near the trace onset remains closely connected to the usual risetime interpretation used in SD photon analyses. Photon-induced showers are expected to have broader time structures and a larger spread in the arrival times of the secondary particles than hadronic showers. The corresponding risetime observable increases when the electromagnetic component is larger, because electrons and secondary photons undergo scattering and attenuation in the atmosphere, unlike muons, which arrive more coherently near the shower front. It also increases when the shower develops deeper in the atmosphere, so that the difference in depth between $X_{\max}$ and the

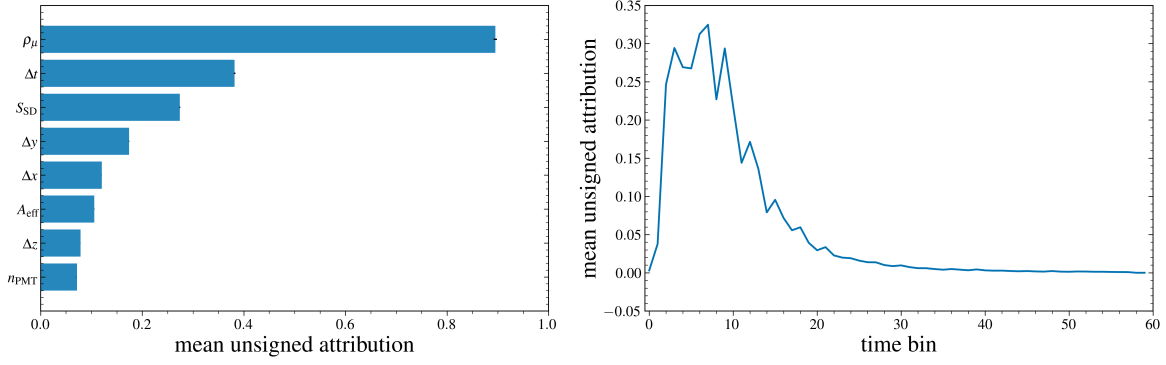


Figure 8.12: Dataset-wide unsigned attributions for the combined SD-UMD branch. Left: mean absolute scalar attribution A_f , averaged over stations and then over events. Right: mean absolute trace attribution T_t as a function of SD trace time bin, also averaged first over stations and then over events. The scalar ranking is dominated by the reconstructed muon density ρ_μ , while the trace attribution is concentrated in the early part of the waveform.

observation level becomes smaller. The attribution pattern found here therefore suggests that the network may be sensitive not only to broad trace-shape information, such as the effective width or rise of the signal, but also to localized early-time structures in the waveform. These structures are compatible with differences in the relative muonic and electromagnetic contributions to the SD signal, although a direct physical decomposition cannot be inferred from the attribution map alone.

8.9 Summary

This chapter introduced the graph-neural-network discriminator used for photon-hadron separation with the SD-433 and UMD detector information. The classification problem was formulated in probabilistic terms, with photon showers treated as signal and proton showers used as a conservative hadronic background during training. Each event was represented as a graph whose nodes correspond to triggered SD stations and whose features encode the local geometry, timing, WCD signal, UMD muon density, effective UMD area, and early-time trace information. Validity masks and subgraph extraction allow the same event representation to support SD-only, UMD-only, and combined SD-UMD views, which is essential for handling incomplete detector configurations.

The independent simulation tests show that the combined SD-UMD score gives the strongest photon-hadron separation, while the UMD-only score retains substantial discrimination power and the SD-only score provides additional complementary information. The background leakage relevant for a photon search was estimated by modelling the photon-like proton tail with an exponential fit in reconstructed-energy bins. Within the statistical limitations of the far tail, the graph-network score is comparable to the traditional M_1 observable under ideal detector-information conditions, with neither method being globally superior over the full efficiency range.

The main advantage of the graph-based approach appears in reduced-information configurations. The combined score degrades gradually when the UMD information is restricted to the counter associated with the hottest SD station and its neighbours, whereas the degradation is stronger when the hottest counter is removed. Even in adverse scenarios where only the UMD counter of the hottest SD station is active, the background contamination remains at 2.5×10^{-5} at a signal efficiency of 0.5. This is still at the baseline level required for photon searches with the SD and UMD in the 433-metres array. The explainability study gives a

consistent interpretation of this behaviour: the network concentrates much of its decision on the local region around the largest SD signal and assigns substantial importance to the corresponding muon-density information, while also using early-time SD trace structure.

Towards a photon search

There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after.

The Hobbit, J. R. R. Tolkien

9.1 Introduction

Diffuse photon searches at tens of PeV are limited not only by photon–hadron discrimination, but also by the exposure that remains after imposing the quality and detector-availability requirements needed for a reliable selection. Photon-induced showers are expected to be muon-poor and to develop differently from hadron-induced showers, so the combination of the SD and the UMD in the 433-metre configuration provides a powerful handle for this search. At the same time, the strict requirement of sufficiently complete hybrid SD–UMD information reduces the usable Phase I exposure, especially during the UMD deployment period.

The graph neural networks developed in Chapter 8 were designed to address this tension. They encode each event as a detector graph and combine SD footprint, timing, and trace information with direct UMD muon measurements, while explicitly handling irregular or incomplete detector configurations. Having established their ideal discrimination performance, their comparison with the M_1 observable, and their robustness under reduced UMD information, the focus now shifts to the consequences for a photon search.

For reference, in the absence of photon candidates, the standard cut-based approach employed within the Collaboration allows the integral upper limit to the diffuse photon flux to be written as

$$\Phi_\gamma < \frac{3.095}{(1 - f_{\text{burn}}) f_{\text{cut},\gamma} \mathcal{E}_\gamma}, \quad (9.1)$$

where 3.095 is the Feldman–Cousins upper bound for zero observed candidates at 95% confidence level, $1 - f_{\text{burn}}$ is the fraction of the data sample effectively available to the search, $f_{\text{cut},\gamma}$ is the average signal efficiency of the photon-selection cut, and $\mathcal{E}_\gamma$ is the accumulated exposure to photon primaries. This expression makes explicit that, once the background contamination is sufficiently suppressed, the sensitivity of the search is driven largely by exposure. A discriminator that remains reliable under incomplete UMD information or broader quality selections can therefore improve the search not only by rejecting background, but also by increasing the effective phase space over which upper

limits can be set. Alternative likelihood-template approaches, which use the full signal and background score distributions rather than a single hard threshold, may further improve the sensitivity in strongly background-dominated regimes [191]. Here, however, we restrict the discussion to the standard cut-based interpretation.

The chapter is organised as follows. In Section 9.2, we validate the score behaviour on a burn sample of Phase I data. In Section 9.3, we estimate the exposure associated with the selected SD-433 hexagons and discuss the impact of a graph-network selection on photon-search prospects. A short discussion of the main score-related systematic uncertainties is provided in Section 9.4.

9.2 Validation on a burn sample

As a final validation step, we apply the trained discriminators to a burn sample of real Phase I events. The purpose of this test is to verify that the score distributions obtained from real data behave as expected for a hadron-dominated sample and do not reveal unexpected systematic effects. To match the burn-sample definition used in the published SD-433+UMD photon search [52], we select events whose GPS second ends in 3. The selected events are required to satisfy the standard 6T5 quality condition, to have reconstructed photon energy $E_\gamma > 10^{16.7}$ eV, and to have zenith angle $\theta < 52^\circ$. Only events recorded between 1 January 2018 and 21 December 2021 are used, corresponding to the Phase I data-taking period considered here.

In contrast with the original search, we do not impose the UMD-coverage requirements used for the final photon selection. This choice is deliberate. The goal of the present test is to verify whether the graph neural networks remain applicable to real events with incomplete UMD information, and Phase I provides an ideal case study because the UMD array was still being deployed during this period. We do, however, require that the counter associated with the hottest SD station is a valid candidate. This requirement is consistent with the robustness studies reported in Chapter 8, where the local UMD information around the hottest station was found to play a central role in preserving the discrimination power. After these selections, the burn sample contains 3211 events, almost twice the 1644 events considered in the burn sample of the original search.

We first define an “ideal” subset of the burn sample by requiring at least three working UMD counters in the neighbourhood of the hottest counter. This criterion is motivated by the robustness study reported in Chapter 8, where configurations retaining the hottest counter together with at least three neighbouring counters still provided a separation close to the nominal case. The score distributions for this subset are compared in Fig. 9.1 with photon, proton, and iron simulations evaluated under ideal detector-information conditions. The comparison is shown separately for the combined SD–UMD score, the SD-only score, and the UMD-only score.

In all three branches, the burn-sample events populate the same broad region as the simulated hadronic background. They are concentrated far from the photon distributions and remain well below the photon-median reference thresholds shown by the dashed vertical lines. This behaviour is the expected one for a sample dominated by hadronic cosmic rays. It also provides a useful data-driven check that the network scores do not assign real Phase I events anomalously photon-like values when the local UMD information is sufficiently complete. Since this ideal subset already shows good separation and stable overall behaviour, no additional validation plots are shown for this configuration. Instead, the remaining checks focus on the reduced-information case, which provides a more stringent test of the method.

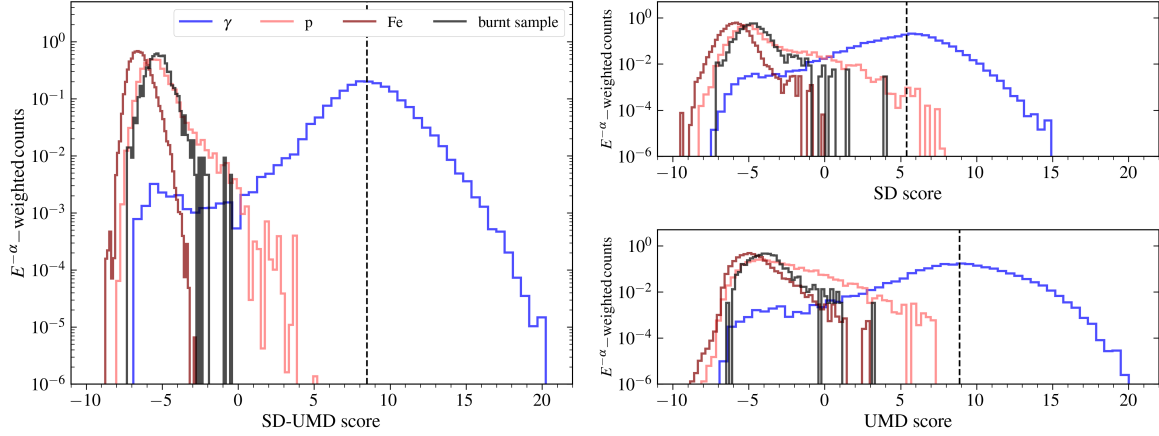


Figure 9.1: Distribution of the three network scores for the ideal subset of the burn sample, compared with simulated photons, protons, and iron showers evaluated under ideal detector-information conditions. The left panel shows the combined SD-UMD score, while the two panels on the right show the SD-only and UMD-only scores. Blue histograms correspond to simulated photons, light red histograms to simulated protons, dark red histograms to simulated iron showers, and black histograms to the burn-sample events. The simulated histograms are weighted according to the assumed energy spectra, using E^{-2} weights for photons and E^{-3} weights for hadrons. The vertical dashed lines indicate the median of the simulated photon distribution for each score.

We then repeat the same comparison for the broader burn-sample selection, in which the local UMD information is less complete. Inspection of the selected data shows that a large fraction of Phase I events contain a valid hottest counter but fewer than three working neighbouring UMD counters. To provide a simulation reference closer to this situation, the photon, proton, and iron samples are also evaluated under a reduced-information scenario in which the UMD information is strongly degraded: only the hottest counter is randomly retained in half of the events, and in other configurations many surrounding counters are also removed. This is intentionally far from the ideal detector configuration and is therefore a stringent test of the network scores.

The resulting score distributions are shown in Fig. 9.2. The combined SD-UMD score remains stable: the burn-sample events continue to populate the hadron-like region and stay well separated from the bulk of the photon distribution. The SD-only score also behaves consistently, as expected, since it does not directly depend on UMD counter availability. The UMD-only branch is the most affected by this reduced-information configuration. This is visible in the larger number of burn-sample events that extend towards, and in some cases beyond, the photon-median reference line. This behaviour is expected because the UMD-only score has access only to the degraded muon information, without the complementary SD signal and timing structure available to the combined branch. The comparison therefore reinforces the main conclusion of the robustness study: incomplete UMD coverage affects the discrimination, but the combined graph representation is substantially more stable than a purely UMD-based score.

A subtle feature of the comparison is that the mean of the burn-sample score distributions resembles the proton simulations more closely than the iron simulations, even though the composition study of Chapter 6 points to an intermediate composition in this energy range. This should not be overinterpreted, since the bulk position of the score distribution can be affected by residual simulation-data differences and by the absence of a composition-mixed reference sample. For the photon search, the critical region is instead the photon-like tail. In

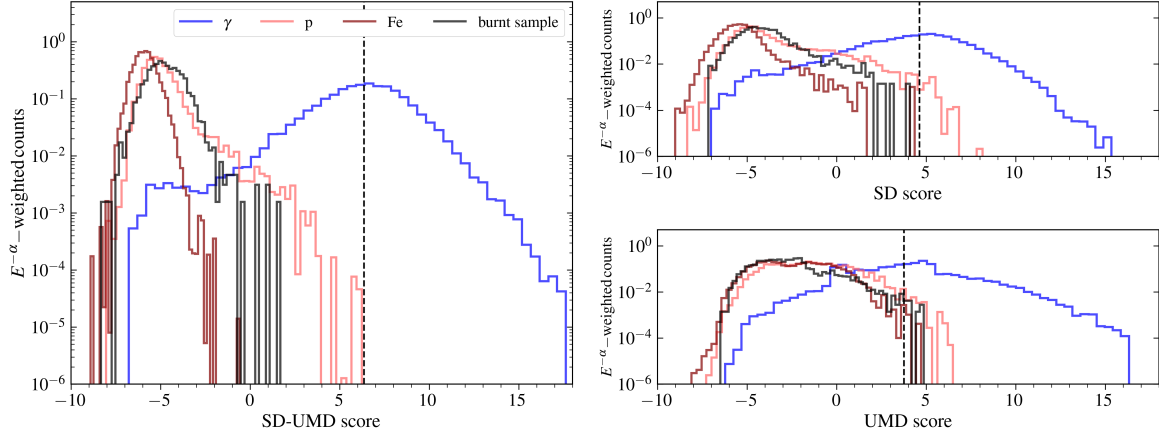


Figure 9.2: Distribution of the three network scores for a broader burn-sample selection with reduced UMD information, compared with simulated photons, protons, and iron showers evaluated under a matching degraded detector-information scenario. The left panel shows the combined SD-UMD score, while the two panels on the right show the SD-only and UMD-only scores. Blue histograms correspond to simulated photons, light red histograms to simulated protons, dark red histograms to simulated iron showers, and black histograms to the burn-sample events. The simulated histograms are weighted according to the assumed energy spectra, using E^{-2} weights for photons and E^{-3} weights for hadrons. The vertical dashed lines indicate the median of the simulated photon distribution.

that region, the burn sample remains compatible with an intermediate hadronic population and does not show an excess of events approaching the photon-median reference.

The energy dependence of the scores provides an additional validation of the method. It is shown in Fig. 9.3, where the burn-sample events are compared with simulated photons, protons, and iron showers as a function of reconstructed photon energy. For the simulated photons, the median score is computed in energy bins and fitted with cubic splines. This provides a smooth reference for the expected photon-like region of score space.

The three panels show that the burn-sample events follow the hadronic populations over the full energy range considered. If the usual photon-median criterion were used as a reference threshold, the burn-sample events would lie far from the candidate region. This is particularly clear for the combined SD-UMD score, where the simulated photon median remains well above the real-data distribution. The SD-only and UMD-only scores show the same qualitative behaviour, although with the expected weaker separation in the single-branch views.

The zenith-angle dependence of the scores is shown in Fig. 9.4. Here the scores are plotted as a function of $\sin^2 \theta$, which provides a convenient variable for inspecting trends across the zenith-angle range. The burn-sample events remain concentrated in the hadron-like region over the full range considered, following the simulated proton and iron populations rather than the photon population. No systematic drift towards photon-like scores is observed at large zenith angles. This is an important cross-check, because the effective area of the UMD modules depends on the shower geometry. Within the selected range, however, the score behaviour remains stable, supporting the use of the discriminator up to $\theta < 52^\circ$.

We also check the stability of the scores as a function of time. The result is shown in Fig. 9.5. The black markers correspond to individual burn-sample events, while the red curves show a 25-day running mean of the scores. The gaps in the distributions are expected and are associated with periods removed by the standard SD data-quality selection. Within the remaining data-taking periods, the score distributions do not show a sudden drift towards the photon-like region. Some changes in the running mean are visible, especially at

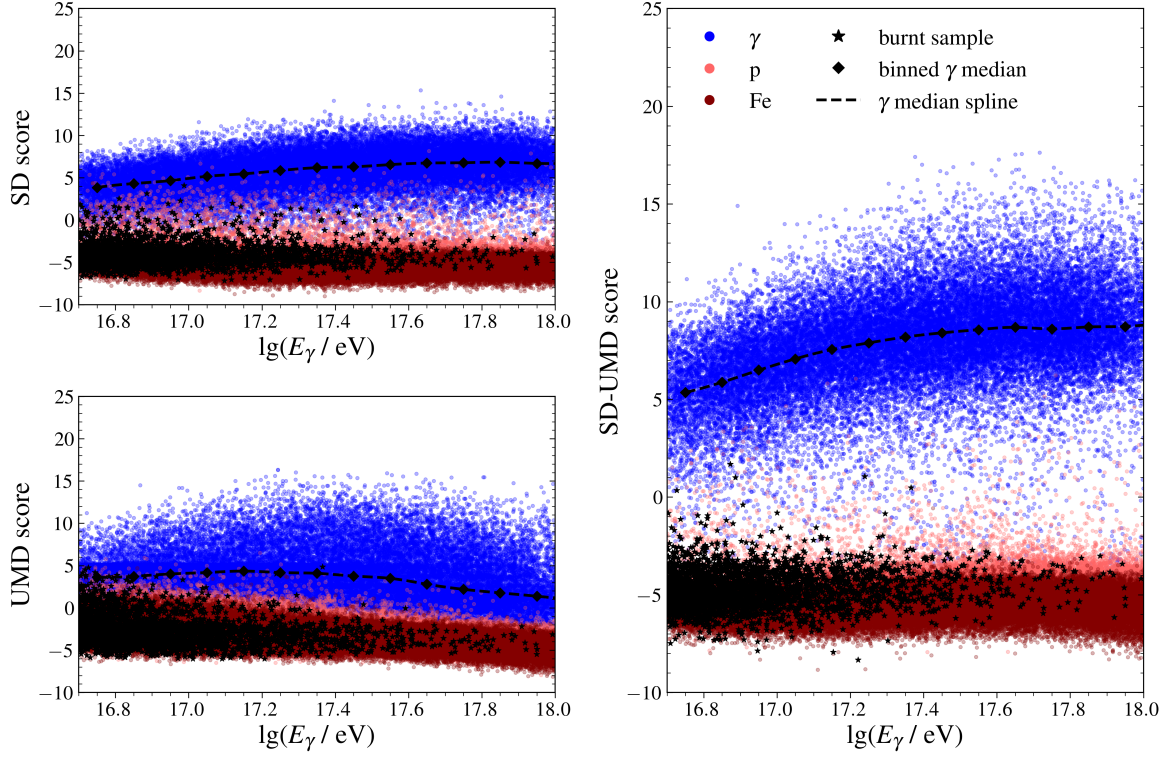


Figure 9.3: Network scores as a function of reconstructed photon energy for simulated photons, protons, iron showers, and burn-sample data. The left panels show the SD-only and UMD-only scores, while the right panel shows the combined SD–UMD score. Blue points correspond to simulated photons, light red points to simulated protons, dark red points to simulated iron showers, and black stars to the burn-sample events. Black diamonds indicate the binned median of the simulated photon distribution, and the dashed black line shows a cubic-spline fit to those medians.

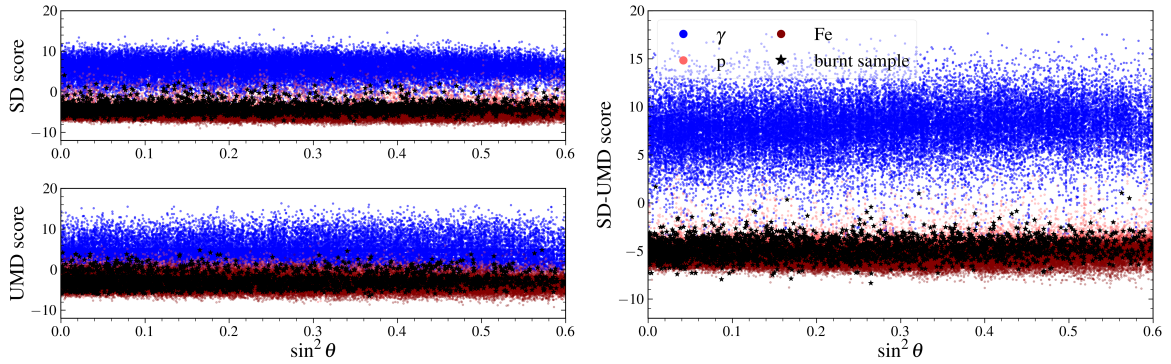


Figure 9.4: Network scores as a function of $\sin^2 \theta$ for simulated photons, protons, iron showers, and burn-sample data. The left panels show the SD-only and UMD-only scores, while the right panel shows the combined SD–UMD score. Blue points correspond to simulated photons, light red points to simulated protons, dark red points to simulated iron showers, and black stars to the burn-sample events.

the end of the Phase I period, but they remain within the hadron-like score region and are consistent with additional detector availability rather than with an unexpected failure of the discriminator. This provides an important check that the network response is not dominated by an unaccounted time-dependent artefact.

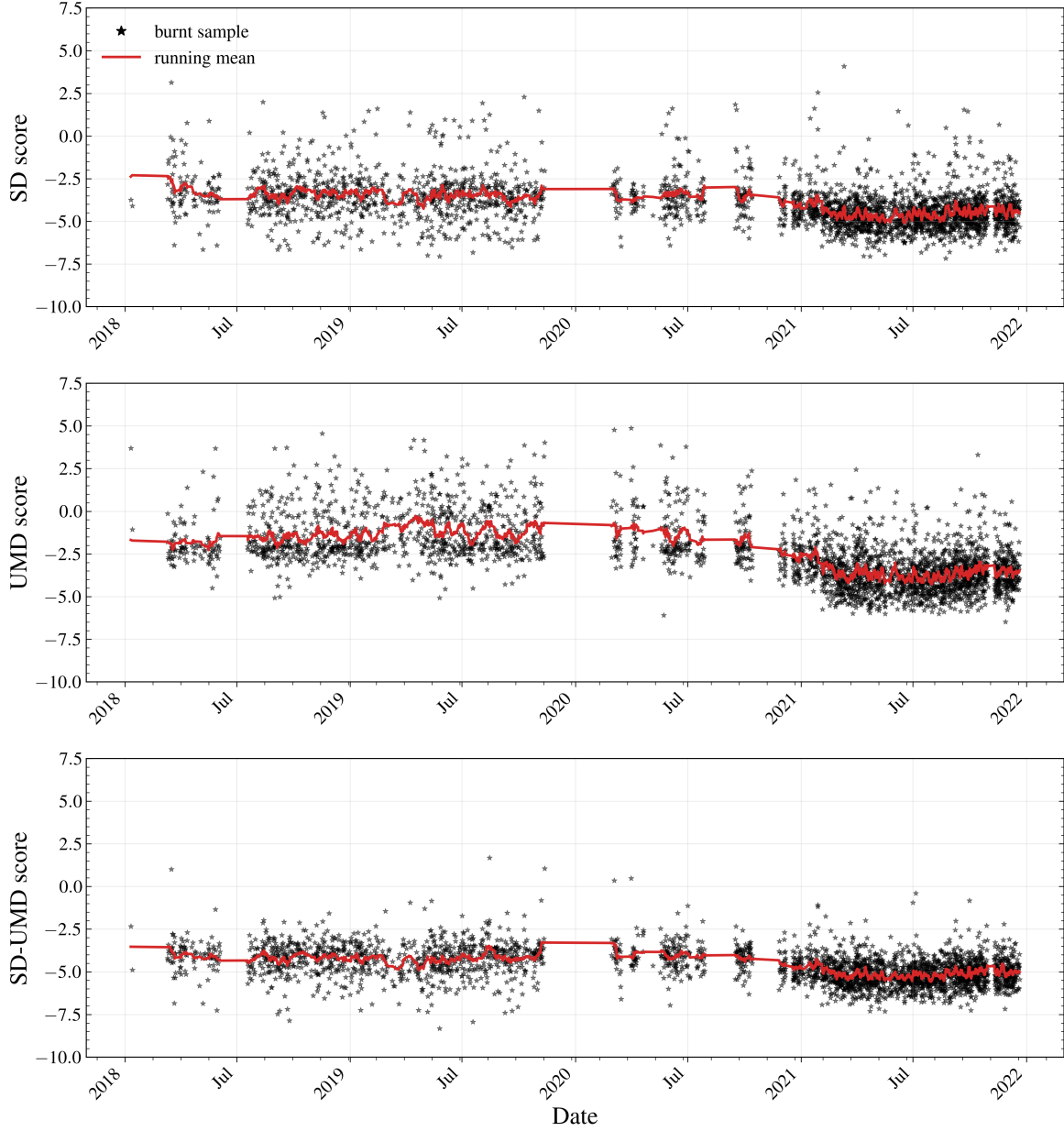


Figure 9.5: Time evolution of the three network scores for the burn-sample events. From top to bottom, the panels show the SD-only, UMD-only, and combined SD-UMD scores. Black stars correspond to individual burn-sample events, and the red curves show a running mean.

Finally, we verify that the correlations between the different network scores are consistent between simulations and real data. The result is shown in Fig. 9.6. The burn-sample events lie on the same hadron-like manifolds populated by the simulated proton and iron showers. In particular, the combined SD-UMD score remains correlated with both the SD-only and UMD-only scores, indicating that the joint branch behaves consistently when applied to real Phase I data. The data do not form a separate population in score space, nor do they show an anomalous displacement towards the photon region. This is an important cross-check because the combined branch uses both SD and UMD information: a significant simulation-data mismatch in either component could have produced a visible distortion in these two-dimensional score correlations.

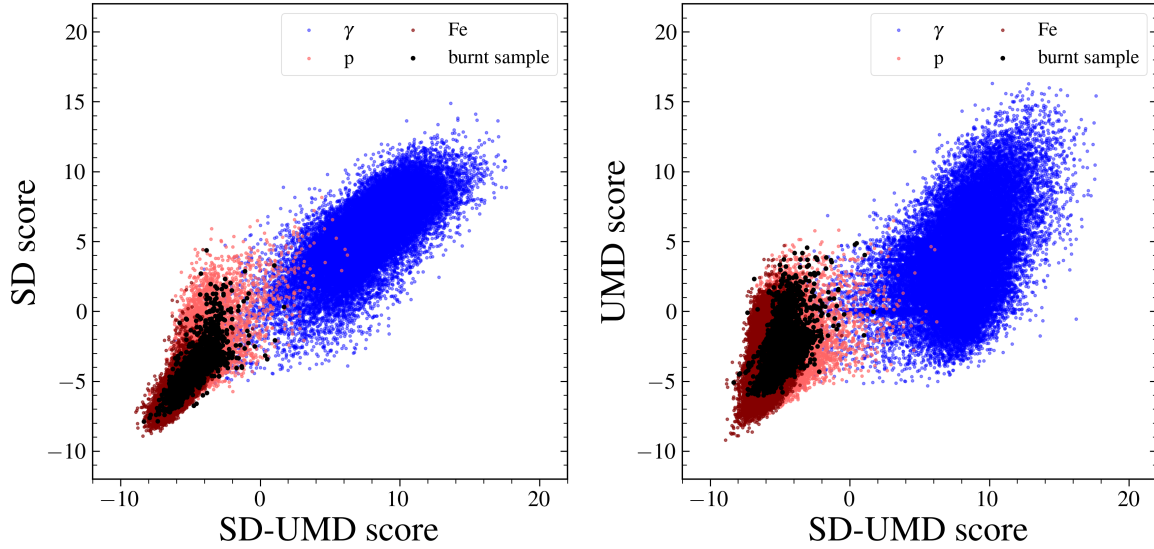


Figure 9.6: Correlation between the combined SD–UMD score and the single-branch scores for simulated photons, protons, iron showers, and burn-sample data. Left: SD-only score as a function of the combined SD–UMD score. Right: UMD-only score as a function of the combined SD–UMD score. Blue points correspond to simulated photons, light red points to simulated protons, dark red points to simulated iron showers, and black points to the burn-sample events. The burn-sample events follow the hadron-like simulation populations and do not form a separate or anomalously photon-like population in score space.

Taken together, these checks provide a data-driven validation of the graph-network discriminator on Phase I events. The burnt-sample distributions are compatible with a hadron-dominated population, remain far from the photon-median reference, and show no unexpected instability with energy, time, or score correlation. Most importantly, the method remains applicable to a substantially larger set of Phase I events than those satisfying the stricter UMD-coverage requirements of the previous search. This supports the main motivation for the graph-based approach: the gain is not necessarily a much stronger discriminator under ideal detector conditions, but a more flexible use of events with incomplete UMD information while maintaining controlled and physically interpretable score behaviour.

9.3 Search prospects

Having established the behaviour of the graph-network scores on the burn sample, we can now discuss the potential impact of the approach on a full unblinding of Phase I data. In the burn-sample study of Section 9.2, the number of usable events increased from 1644 events in the original SD-433+UMD selection to 3211 events in the graph-network selection. If the photon efficiency and the background rejection can be kept under control, this larger event sample directly increases the probability of finding a photon candidate. In the absence of candidates, the same increase in usable exposure would instead translate into stronger upper limits on the diffuse photon flux.

To quantify this effect more directly, we estimated the photon exposure for the four SD-433 hexagons whose centres were equipped at some point during Phase I with UMD counters. The selected hexagons are shown schematically in the left panel of Fig. 9.7. The calculation follows the approach of Ref. [192], in which the T2Life information is used to determine the time intervals during which a given hexagon is active. A hexagon is considered active when the central station and the six surrounding stations are alive.

The starting point is the time-dependent uptime of each selected hexagon. We restricted the calculation to the period from 1 January 2018 to 21 December 2021, and considered events with zenith angle $\theta \leq 52^\circ$. The SD uptime was corrected by subtracting the SD-433 bad periods identified from the time distribution of 6T5 events [192]. In addition, the first two weeks of January 2018 were excluded as an array-level bad period. We also applied station-specific rejection periods relevant for the selected hexagons. For station 97, these are 3–4 November 2020, 14 November–6 December 2020, and 15–17 December 2020. The long period from 14 November to 6 December 2020 is an additional *a posteriori* correction associated with unstable behaviour of Lety Jr. For station 30, the rejected periods are 26–27 May 2021 and 24–26 June 2021.

We then computed two exposure estimates. The first one, referred to as the pure SD exposure, only requires the SD hexagon to be active after the corrections described above. The second one, referred to as the hybrid exposure, imposes an additional UMD availability condition. For this hybrid exposure, the centre station of the active SD hexagon must be associated with at least one working UMD module, in line with the detector configurations considered in the previous sections. The relevant UMD counters are therefore paired with station 1764 for the central hexagon, station 30 for the south-west hexagon, station 12 for the north-west hexagon, and station 97 for the north hexagon. The beginning of the hybrid exposure for each centre was determined from the first available timestamp of a candidate UMD counter in the dataset. These start times are 1 January 2018 18:44:42 UTC for station 1764, 26 November 2020 00:33:06 UTC for station 12, 3 February 2021 07:42:54 UTC for station 97, and 10 June 2021 21:00:50 UTC for station 30. After these start times, UMD module-quality masks were applied. A centre counter was considered usable if at least one of its modules was not rejected by the UMD module-level bad periods [193].

For each threshold energy, the photon exposure was obtained by folding the corrected live time with the photon trigger efficiency from Section 7.1. The exposure above a photon threshold energy E_γ^{th} can be written as

$$\mathcal{E}_\gamma(E_\gamma^{\text{th}}) = \sum_h \frac{\int_{E_\gamma^{\text{th}}}^\infty E_\gamma^{-2} \left[\int_\Omega \int_t A_{\text{cell}} \cos \theta u_h(t) \tau(E_\gamma, \theta) dt d\Omega \right] dE_\gamma}{\int_{E_\gamma^{\text{th}}}^\infty E_\gamma^{-2} dE_\gamma}, \quad (9.2)$$

where the sum runs over the selected hexagons, $\tau(E_\gamma, \theta)$ is the photon trigger efficiency, and $u_h(t)$ is the corrected uptime of hexagon h . The angular integration is performed over the zenith-angle range $\theta \leq 52^\circ$. The factor E_γ^{-2} is used to average the trigger efficiency above each threshold, following the spectral assumption used in the photon-search exposure calculation. The area factor is

$$A_{\text{cell}} = \frac{\sqrt{3}}{2} (0.433 \text{ km})^2. \quad (9.3)$$

For the pure SD exposure, $u_h(t)$ is the SD uptime after subtracting the SD-433 bad periods and the relevant station-rejection periods. For the hybrid exposure, $u_h(t)$ is replaced by the product of the corrected SD uptime and the requirement that the centre counter of the corresponding hexagon has at least one available UMD module.

The resulting integral exposures, separated by hexagon centre and summed over the four selected hexagons, are given in Table 9.1. The quoted uncertainties follow the prescription of Ref. [192], using the relative spread of the event-rate distribution, 4%, as the exposure uncertainty. The pure SD exposure reaches 2.94–3.13 km² yr sr, while the corresponding hybrid exposure reaches 1.27–1.35 km² yr sr. The weak increase with threshold is expected: after averaging the trigger efficiency above E_γ^{th} , higher thresholds probe a region where the photon trigger is closer to full efficiency. Since the same photon-efficiency correction is

Table 9.1: Integral photon exposure above threshold for the four selected SD-433 hexagons centred on stations 1764, 30, 12, and 97. The SD exposure requires only the corrected SD hexagon uptime, while the hybrid exposure additionally requires at least one available UMD module at the centre counter. The exposures are averaged above each threshold assuming an E^{-2} photon spectrum and are integrated up to $\theta_{\max} = 52^\circ$.

Threshold $\lg(E_\gamma^{\text{th}}/\text{eV})$	Centre	SD exposure [km ² yr sr]	Hybrid exposure [km ² yr sr]	Hybrid / SD
16.7	1764	0.794 ± 0.032	0.685 ± 0.028	0.862
	30	0.666 ± 0.027	0.137 ± 0.006	0.205
	12	0.750 ± 0.031	0.248 ± 0.010	0.331
	97	0.732 ± 0.030	0.201 ± 0.008	0.274
	Total	2.942 ± 0.120	1.271 ± 0.052	0.432
16.9	1764	0.827 ± 0.034	0.713 ± 0.029	0.862
	30	0.694 ± 0.028	0.142 ± 0.006	0.205
	12	0.781 ± 0.032	0.259 ± 0.011	0.331
	97	0.762 ± 0.031	0.209 ± 0.009	0.274
	Total	3.064 ± 0.125	1.323 ± 0.054	0.432
17.1	1764	0.841 ± 0.034	0.725 ± 0.030	0.862
	30	0.706 ± 0.029	0.145 ± 0.006	0.205
	12	0.794 ± 0.032	0.263 ± 0.011	0.331
	97	0.775 ± 0.032	0.213 ± 0.009	0.274
	Total	3.116 ± 0.127	1.346 ± 0.055	0.432
17.3	1764	0.844 ± 0.034	0.728 ± 0.030	0.862
	30	0.709 ± 0.029	0.145 ± 0.006	0.205
	12	0.797 ± 0.032	0.264 ± 0.011	0.331
	97	0.778 ± 0.032	0.213 ± 0.009	0.274
	Total	3.129 ± 0.127	1.351 ± 0.055	0.432

applied to the pure SD and hybrid estimates, the ratio between them is controlled almost entirely by detector availability. The hybrid exposure is approximately 43% of the pure SD exposure at all thresholds, or equivalently the pure SD exposure is larger by a factor of about 2.3.

The size of this reduction is expected. The pure SD exposure only requires the corrected SD hexagon uptime, whereas the hybrid exposure additionally requires that the centre of the active hexagon has at least one working UMD module. The delayed deployment or availability of the UMD counters associated with stations 12, 97, and 30 therefore removes a substantial fraction of the Phase I live time from the hybrid calculation. This behaviour is visible in the cumulative-exposure curves shown in Fig. 9.7: station 1764 contributes to the hybrid exposure from the beginning of the period, while the other three centres begin to contribute only after their corresponding UMD start times. The ideal case of station 1764 suggests that the hybrid exposure could reach at least 86% of the pure SD exposure when relying only on the central UMD counter.

The projected limits are shown in Fig. 9.8. The black circles correspond to the published SD-433+UMD limits. The red stars show the approximate projection obtained by assuming no photon candidates during unblinding and employing the increased exposure, while

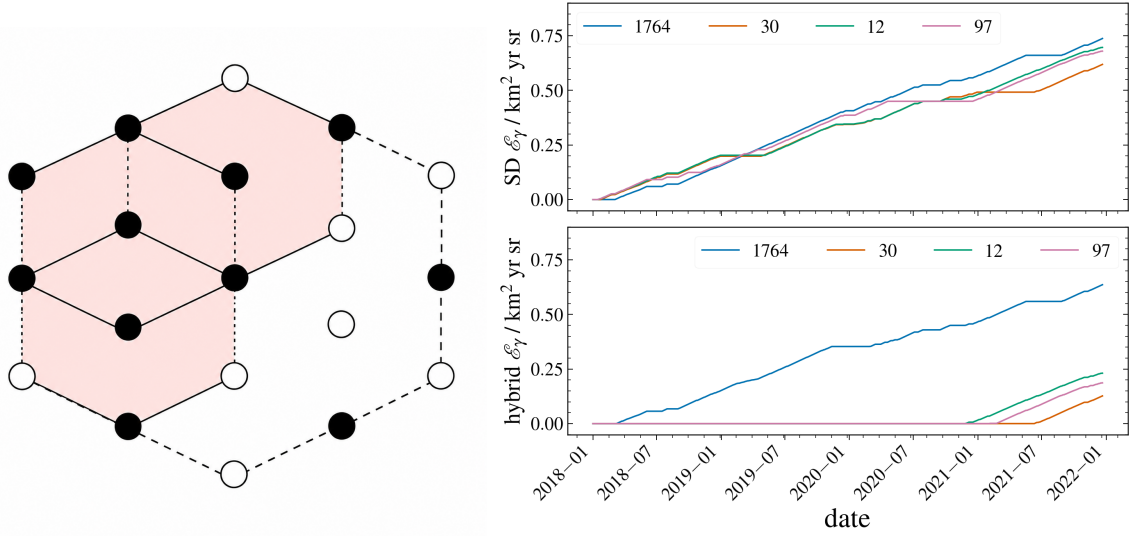


Figure 9.7: Left: schematic representation of the selected SD-433 hexagons used in the exposure estimate. The shaded region indicates the region covered by the four hexagons centred on stations 1764, 30, 12, and 97, for which the availability of the corresponding UMD centre counters was evaluated. Right: cumulative photon exposure for the same four hexagons. The upper panel shows the pure SD exposure, obtained from the corrected SD uptime after subtracting the SD-433 bad periods, the first two weeks of January 2018, and the relevant station-specific rejection periods. The lower panel shows the corresponding hybrid exposure, where the centre counter is additionally required to have at least one available UMD module. Each curve corresponds to one hexagon centre.

keeping the same signal efficiency. Under these assumptions, the integral limits improve by approximately the same factor as the exposure. In the lowest-threshold bin, this projected improvement brings the limit close to the first super-heavy-dark-matter prediction. The blue squares show, for reference, a projection based on the broader pure SD exposure for the chosen four hexagons. This comparison should not be interpreted as a complete SD-only photon-search result. Rather, it illustrates the scale of the limits achievable when the selection is not restricted to the subset of events with full hybrid SD–UMD information.

The projections highlight the main outcome of this analysis. Network-based discriminators do not need to outperform traditional observables dramatically under ideal detector conditions in order to be useful. Their main value lies in the possibility of extending the usable event sample by operating under incomplete detector configurations, while maintaining a photon–hadron separation compatible with the requirements of the search. The robustness tests show that the combined SD–UMD score degrades gradually when the UMD information is reduced, especially when the counter associated with the hottest SD station is retained. The explainability study gives a natural interpretation of this behaviour: a substantial part of the discrimination power appears to be carried by the local coincidence between the dominant SD signal and the nearby muon measurement. Finally, the burn-sample validation shows that the method behaves as expected on real Phase I data, without producing an unexpected population of photon-like events. Together, these results suggest that graph-based discriminators can provide a practical way to increase the effective exposure of photon searches starting at tens of PeV.

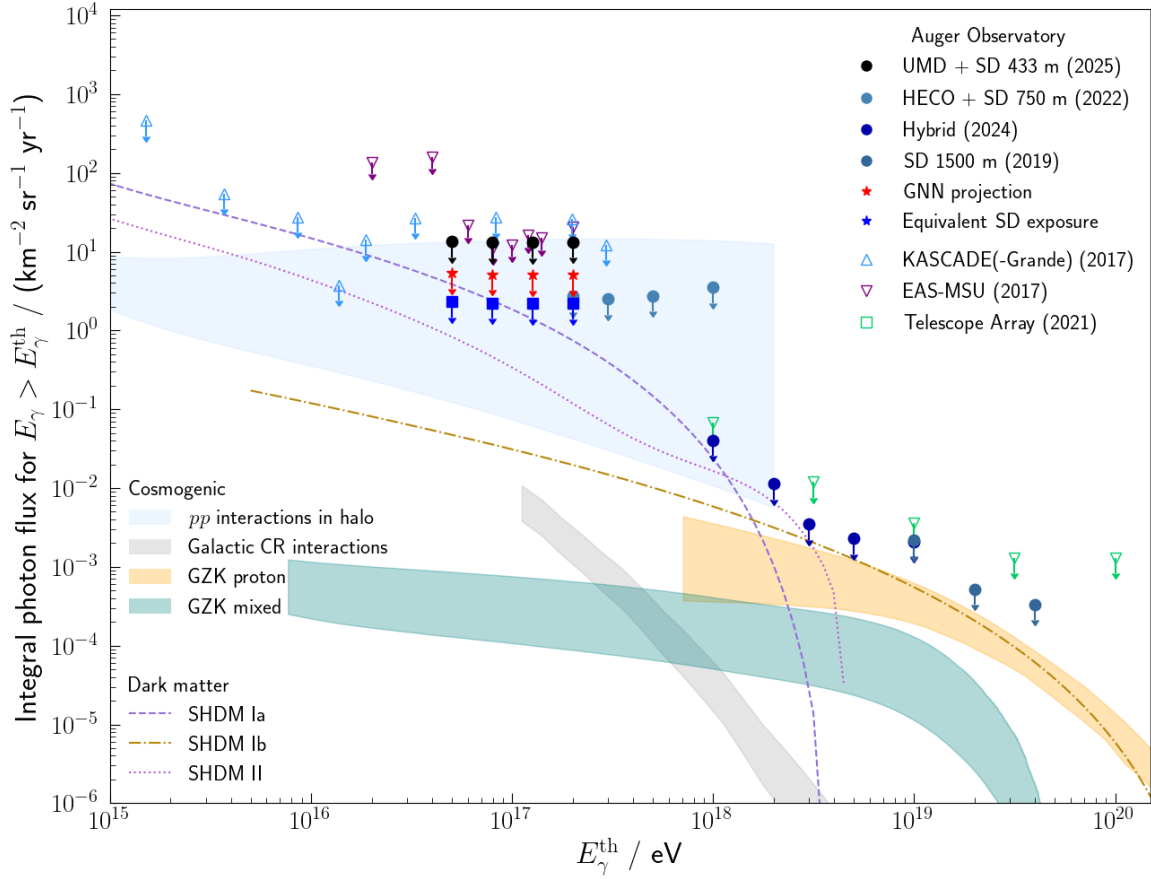


Figure 9.8: Integral photon-flux limits and projections above a threshold energy E_γ^{th} . The published SD-433+UMD limits from the Pierre Auger Observatory are shown as black circles [52]. Red stars indicate the approximate graph-network projection obtained by assuming an almost factor-of-two increase in effective exposure with no photon candidates and otherwise unchanged selection efficiency. Blue squares show, for reference, a projection based on the pure SD exposure for the same selected hexagons. Previous Auger limits at higher energies are shown together with limits from KASCADE-Grande, EAS-MSU, and Telescope Array [45, 53, 55–57, 194, 195]. The shaded regions show representative predictions for cosmogenic photon fluxes from interactions of ultra-high-energy cosmic rays with interstellar Galactic matter, background radiation fields, and hot gas in the Galactic halo [58–61]. Dashed and dotted curves indicate representative super-heavy-dark-matter scenarios [196, 197].

9.4 Systematic uncertainties of the scores

No discrimination observable intended for a photon search can be considered strictly free from systematic effects, and the network-based discriminator developed in this work is no exception. However, one of the guiding principles from the beginning of its development was to reduce, as much as possible, the dependence on reconstructed quantities that are known *a priori* to be especially vulnerable to model- or reconstruction-induced biases. The network is not explicitly provided with the reconstructed shower core, the reconstructed energy, or the zenith angle. Instead, the event is encoded in a local detector frame centred on the hottest SD station, and the geometric and timing variables are expressed relative to that reference. This is an important design choice, because it makes the score depend primarily on the internal structure of the measured detector footprint rather than on global

shower quantities whose systematic uncertainties would otherwise be inherited directly by the classifier.

A caveat is nevertheless required. The zenith angle is not completely absent from the event representation, because the reconstructed muon density ρ_μ is obtained from the measured counts through the corresponding effective area A_{eff} , which depends on the shower axis. This dependence is, however, qualitatively different from using the reconstructed zenith angle or reconstructed core as explicit inputs to the decision function. The shower axis enters only through an area-projection correction, and its determination relies primarily on the timing geometry of the SD stations, namely on the GPS-synchronised start times of the traces, which constrain the shower direction accurately. The behaviour of the three network scores as a function of energy and zenith angle is shown in Figs. 9.3 and 9.4. Within the available statistics, these distributions do not show evidence for a systematic trend that would spoil the discrimination.

A second major source of systematic uncertainty is the hadronic composition of the background itself. Throughout this thesis chapter, proton showers were adopted as the reference hadronic background during training and for the benchmark contamination estimates. This is a conservative choice. As shown in Chapter 8 and in Figs. 9.1 to 9.4 and 9.6, iron showers, which were not used during training, are shifted even farther away from the photon population than proton showers. The network therefore generalises in the physically expected direction: more muon-rich and earlier-developing hadronic cascades are assigned more hadron-like scores. Since the composition results discussed in Chapter 6 indicate that the mean mass in the relevant energy range is around oxygen or heavier, the real hadronic background is expected to be, if anything, less photon-like than the pure-proton benchmark adopted here. The validation with real events shows score-distribution tails compatible with an intermediate composition. From this standpoint, the discrimination performance estimated from proton-distribution tails should be interpreted as conservative with respect to composition.

The same logic extends, although less directly, to the choice of hadronic interaction model. No architecture choice can remove this irreducible source of uncertainty, and a dedicated comparison at the level of the final score distributions remains necessary. However, the effect should not be overstated. QGSJet-II.04 and Sibyll-2.3e predict slightly smaller muon contents than EPOS-LHC, but the corresponding shift is weaker than the separation induced by primary mass, as shown in Chapter 6. In other words, the model-to-model variation is expected to be subdominant with respect to the difference between protons and heavier primaries already explored in this study.

At detector level, additional mismatches between simulations and data may arise from incomplete UMD availability, masked channels, imperfect module-status information, or local response fluctuations. Here again, the discriminator was intentionally designed with such conditions in mind. Missing detector information is handled explicitly through validity masks and subgraph extraction, and the training stage includes random masking of PMTs, UMD modules, and even full stations. For this reason, the robustness tests conducted are directly relevant to the systematic-uncertainty discussion. The UMD counter coverage is important: reduced sampling of the muon footprint weakens the separation. However, the degradation is gradual when neighbouring counters are restored, and the discriminator retains useful separation even under strongly reduced UMD-information scenarios. This indicates that the score is not entirely dependent on an idealised detector configuration.

In summary, the discrimination observable developed here is not free from systematic effects, but several of the most immediate and avoidable ones were mitigated already at the level of the event representation and training strategy. The remaining relevant uncertainties are largely associated with the physics of hadronic air showers, the choice of hadronic

interaction model, and the detailed detector response in real data, rather than with an explicit use of reconstructed observables. Taken together, the proton-only training choice, the behaviour of the independent iron sample, the mild score trends across phase space, and the gradual degradation under reduced UMD information all suggest that the reported performance is likely to be conservative with respect to the dominant hadronic-background assumptions. The final validation with the burn sample did not reveal any unexpected systematic behaviour in the score distributions, providing an additional data-driven check of this interpretation.

9.5 Summary

This chapter used the graph-network discriminator developed in Chapter 8 to assess the practical search prospects for diffuse photons at tens of PeV. The burn-sample validation showed that real Phase I events populate the expected hadron-like region of the score distributions, both in an ideal local-UMD subset and in a broader reduced-information selection. The enlarged selection contains 3211 events, compared with 1644 events in the original SD-433+UMD burn-sample selection, showing that the graph-based treatment can recover a substantially larger usable event sample while retaining stable score behaviour.

The exposure estimate for the four selected SD-433 hexagons shows that, during the Phase I period considered here, the strict hybrid requirement reduces the usable photon exposure to about 43% of the corresponding pure-SD exposure. Under the assumption of unchanged photon efficiency and no photon candidates, the projected limits would be lowered by about 50%, corresponding to an almost factor-of-two improvement compared to the published limits. This projection shows the limits would lie closer to the most optimistic super-heavy-dark-matter scenarios in the lowest energy threshold bin.

The main conclusion is therefore operational. Networks are useful not only because they provide a powerful photon-hadron discriminator, but because they can exploit heterogeneous detector information without requiring a fully ideal hybrid configuration. More generally, these findings are relevant for AugerPrime independently of whether searches rely on deep-learning methods. The upgraded Observatory will provide a richer multi-hybrid event description, combining the SD information with additional mass-sensitive observables from the SSD, UMD, and RD. The results suggest that photon searches may benefit from strategies that retain the broad SD exposure while enhancing the photon-hadron discrimination with complementary detector information whenever it is locally available. In particular, observables or classifiers that exploit the detector information closest to the region of largest SD signal could recover much of the available exposure without requiring a fully hybrid event configuration. Such approaches would make it possible to extend the usable event sample for photon searches, increasing the chance of finding a candidate or, in the absence of candidates, improving the corresponding photon-flux limits.

Chapter 10

Summary and outlook

Tener una mentalidad científica es seguir planteándose preguntas sin esperar otra respuesta que una nueva pregunta.

Un inmenso azul: El mar, el abismo y la curiosidad humana, Patrik Svensson

This thesis was motivated by the problem of searching for diffuse photons starting at energies of a few tens of PeV, an energy domain in which the expected flux is extremely small and the dominant background is formed by hadron-induced extensive air showers. In this regime, any meaningful improvement in sensitivity requires not only larger exposure, but also a robust use of the available detectors. The central objective of this work was therefore to improve photon searches at the Pierre Auger Observatory by exploiting the joint capabilities of the SD-433 array and the Underground Muon Detector within the AugerPrime era, with particular emphasis on machine-learning methods able to use the full detector-level structure of the events.

The first contribution of the thesis was to establish the detector-response ingredients needed to formulate such a search on solid grounds. On the hadronic side, the response of the SD-433 array was characterised in the energy region of the second knee, including its efficiency, energy bias, and energy resolution. This provides the mapping between true and reconstructed energy that is required whenever simulations must be compared with data in a quantitative way. On the photon side, a dedicated reconstruction chain was developed for photon-induced showers, including a photon-specific lateral distribution function, geometry validation, and an energy calibration adapted to the predominantly electromagnetic nature of the signal. The reconstructed photon energy was shown to remain stable, with a mean bias consistent with zero and a relative resolution of about 14%. Taken together, these developments define the fundamental detector understanding needed to interpret both the signal expectation and the residual background in a future photon search.

A second major result of this thesis concerns the characterisation of the hadronic background itself through the muon content measured with the UMD-433. A substantial part of the work was devoted to the reconstruction problem: improving data-quality handling, clarifying the rejection logic at module, counter, and event level, and mitigating pathological cases capable of biasing the fitted muon lateral distribution function. On this basis, a stable event-by-event muon observable was defined and interpreted through detector-level proton and iron baselines obtained from simulations. The resulting measurement around the second knee indicates a gradual evolution from heavier to lighter primaries above that region, while remaining compatible with a muon content characteristic of oxygen-like or heavier compositions over much of the covered range. At the same time, the comparison

with $X_{\max}$ -based expectations still points to a residual muon deficit in simulations, although in this energy range the discrepancy appears less severe than what is sometimes suggested by higher-energy muon observables. This result is important not only as an independent composition measurement, but also because it clarifies the physical nature of the background against which photon candidates must be identified.

The core methodological contribution of the thesis is the development of a photon–hadron discriminator based on graph neural networks. Rather than reducing each event to a small set of handcrafted global variables, the analysis represents the multi-detector signals directly as a graph built from the triggered SD stations and the available UMD information. In this way, the classifier can exploit the spatial, temporal, and muonic structure of the shower footprint at detector level. The resulting multi-view architecture, combining SD-only, UMD-only, and joint SD–UMD information, was shown to provide useful discrimination power while remaining compatible with incomplete detector configurations. Equally important, the performance studies indicate that the method behaves conservatively: the use of protons as background, the observed response to iron showers, the mild trends across phase space, and the robustness under reduced UMD information all suggest that the quoted separation power should not be interpreted as artificially optimistic. In that sense, the graph-based approach developed here demonstrates that machine learning can be incorporated into a photon search in a way that is not only powerful, but also methodologically controlled.

The final chapter moved from detector studies and simulated performance to the question of practical applicability in data. The graph-network discriminator was first validated on a burn sample of real Phase I events. The score distributions were found to populate the expected hadron-like regions, with no unexpected instability as a function of energy, zenith angle, time, or score correlation. Most importantly, the graph-based selection increased the number of usable burn-sample events from 1644 to 3211 relative to the previous SD-433+UMD selection. This shows that the main practical advantage of the method is not necessarily a dramatic improvement in ideal detector-level separation, but the possibility of extending the usable event sample while retaining controlled photon–hadron discrimination.

The exposure study carried out for the selected SD-433 hexagons quantified this effect more directly. During the Phase I period considered here, the strict hybrid requirement reduces the usable photon exposure to about 43% of the corresponding pure-SD exposure. This loss is mainly driven by the delayed deployment or availability of the UMD counters associated with several of the selected hexagon centres. The graph-network approach provides a way to recover part of this lost exposure by using events with incomplete UMD information, especially when the counter associated with the hottest SD station is available. Under a Phase I data unblinding scenario, and assuming no photon candidates and unchanged signal efficiency, the projected exposure achievable with this discriminator indicates that the existing photon limits in this energy range could be improved by approximately a factor of two. In the lowest-threshold bin, this would move the projected limit closer to the most optimistic super-heavy-dark-matter scenarios considered.

In addition, the transition to Phase II opens the possibility of applying the same ideas to a larger and more mature multi-hybrid data set, provided that the response of the classifier is verified under the new acquisition conditions. More generally, the results suggest that future photon searches may benefit from strategies that retain the broad SD exposure while using complementary mass-sensitive information from the SSD, UMD, and RD whenever it is locally available. Beyond the immediate search itself, several further developments are natural: refining the treatment of systematic uncertainties in the tail of the discriminator distribution, comparing more extensively the response under different hadronic-interaction models and composition assumptions, and exploring whether the graph-based framework can be generalised to other classification or reconstruction tasks within AugerPrime.

Agradecimientos

Me resulta más natural dar las gracias en mi lengua materna. Por alguna razón, el español se siente más cálido que el inglés. A lo largo de estos años leí muchas tesis en las que las formas de agradecer son muy diversas. Algunos agradecimientos son cortos; otros, largos; algunos, prolijos; otros, más desprolijos, pero también más reales. Yo elijo dar las gracias como creo que deben darse, sin ahorrar ninguna palabra. Así que estás avisado: esta sección de agradecimientos es un poco más larga de lo común.

En primer lugar, quiero agradecer a mi familia. A mi papá, Daniel; a mi mamá, Andrea; y a mi hermana, Carolina, que me soportan desde que nací. Tengo un carácter difícil y soy muy exigente conmigo mismo y con ellos. Gracias por ayudarme, cada uno a su manera, incluso cuando yo no me dejo ayudar. Perdón por no permitírmelo más seguido. También le quiero agradecer a mi abuela Kela, que no pudo acompañarme en este camino, pero cuya memoria siempre me motiva a seguir adelante. A mi abuelo Quique, que, más allá de la distancia, siempre me hizo sentir que estaba orgulloso de mí. Perdón por no haber estado para despedirte. A mi familia de Buenos Aires y a mi familia de Córdoba, porque, aunque no compartamos mucho tiempo juntos, los tengo siempre presentes. Perdón por no visitarlos más seguido. Tengo que aprender a priorizar lo importante por sobre lo urgente.

A mis amigos de ITeDA. Gracias, Marina, por tu entendimiento. No voy a decir que somos parecidos, porque me vas a decir que me estoy copiando de algo que ya pensaste. Gracias, Gaby, por tu complicidad, porque fuimos codo a codo desde el principio hasta el final. Gracias, Carmi, por tu afecto, porque, en un ambiente donde la cercanía no siempre es habitual, las pequeñas muestras de cariño son una alegría. Gracias Varada y Joaquín por la serenidad que aportan, porque, cuando las emociones desbordan, la calma que transmiten con tanta naturalidad resulta reconfortante. Gracias a Fernando, porque las conversaciones sobre ejercicios, proteínas e inversiones, hacen que el ambiente laboral se sienta más relajado y humano. Gracias a Belén por las recomendaciones de cocina y lectura. Somos más personas que científicos, y ella refleja como nadie una curiosidad que va mucho más allá de la ciencia. A Rolando, José, Iván, y Luca, por todas las charlas y risas. Gracias a todos ustedes por hacer del día a día una aventura.

Existen amistades circunstanciales y amistades trascendentales. Estas últimas trascienden las circunstancias y perduran durante muchos años. Gracias a Lucas y a Nico por seguir en mi vida desde la secundaria, en las buenas y en las malas, en la cercanía y en la distancia. Gracias a Raúl, Aye, Mari y Lu, porque compartimos mucho más que entrenamientos y siempre me hacen sentir bienvenido cuando los visito en Bahía Blanca. Gracias también a mis amigos de los tiempos universitarios, Bambo, Facu, Marcos, Enzo, Pedrito, Mati, Sergio, Ro y Joaco, porque, por más ocupada que sea la vida adulta y por más difícil que sea reunirnos, valoro que se hagan el tiempo para verme cuando coincidimos en tiempo y espacio. Los mensajes que intercambiamos de vez en cuando para ponernos al día son un combustible

necesario para mantener el motor en marcha. Gracias a Flor, por aparecer en mi vida cuando menos lo esperaba y más lo necesitaba. Gracias por alegrar mis días, acompañarme y ser el refugio que no sabía que me hacía falta.

A mis directores del lado argentino del programa doctoral, Brian y Nico, porque no tardaron en entender que trabajo mejor con espacio y libertad. Eso no es algo trivial, porque las personas somos complejas. Terminar esta tesis no habría sido posible sin su ayuda y su comprensión. A mi director del lado alemán, Ralph, que con breves comentarios ocasionales es capaz de aportar más que en muchas charlas. Gracias también a todos los investigadores de ITeDA. Mis agradecimientos particulares a Diego, Daniel, Juan Manuel y Federico, porque escucharon mis ideas y las discutieron con seriedad.

A los investigadores de IAP, que me dieron la bienvenida y me hicieron sentir a gusto en un país cuya lengua y cultura me eran ajenas. Gracias, Markus, por valorar mi trabajo desde el primer día. Gracias, Darko, por las largas discusiones y enseñanzas en programación y estadística. Gracias, David, por darme oportunidades y por notar el trabajo que no siempre es visible para el ojo distraído. A mis colegas de Karlsruhe. Gracias, Steffen, Fiona, Berenika, Tobias, Paul y Emily por el tiempo compartido en KIT. Porque los intereses en común unen a las personas más allá de las fronteras geográficas y culturales. Agradezco en particular a Sara, que se preocupó por mí como nadie durante mis estadias en Alemania. Muchas gracias por hacerme sentir en familia cuando no era tu responsabilidad.

Tampoco pueden faltar agradecimientos más generales para todas las personas cuyo trabajo permite que los científicos no se ahoguen en un vaso de agua. Gracias a Tadeo, Consuelo, Sabrina y María, de ITeDA. Gracias a Marie-Christine, Sabine y Anna, de IAP. La burocracia es una jungla, y nunca habría sido capaz de atravesarla sin su ayuda y su constante buena predisposición. Gracias también a los miembros de la Colaboración Pierre Auger, porque las discusiones permanentes en llamadas, reuniones y conferencias enriquecieron profundamente mi paso por este excelente experimento. Hasta el día de hoy sigo maravillado por lo que puede lograr un grupo numeroso de personas semi-organizadas (somos físicos, no voy a fingir que sabemos organizarnos).

Quiero agradecerme también a mí mismo. Por haberme dado la oportunidad de perseguir mi curiosidad, más allá de todas las circunstancias. Por haber seguido adelante en momentos que nunca pensé que tendría que atravesar. Por haberme permitido conocer personas maravillosas y recorrer lugares que nunca imaginé pisar en la vida. Probablemente me esté olvidando de algún otro agradecimiento, así que también agradezco a todas las personas que se cruzaron en mi camino y me ayudaron, cada una a su manera. Finalmente te agradezco a vos por llegar hasta acá. Si dedicaste aunque sea unos minutos de tu vida a leer un fragmento de esta tesis, te lo agradezco de corazón.

Bibliography

- [1] H. Becquerel, Sur les radiations émises par phosphorescence, *Comptes Rendus de l'Académie des Sciences*, vol. 122, pp. 420–421, 1896.
- [2] P. Curie and M. Curie, Sur une nouvelle substance fortement radio-active, contenue dans la pechblende, *Comptes Rendus de l'Académie des Sciences*, vol. 127, pp. 175–178, 1898.
- [3] C. Wilson, On the ionisation of atmospheric air, *Proceedings of the Royal Society of London*, vol. 68, pp. 151–161, 1901.
- [4] T. Wulf, Beobachtungen über die Strahlung hoher Durchdringungsfähigkeit auf dem Eiffelturm, *Physikalische Zeitschrift*, vol. 11, pp. 811–813, 1910.
- [5] V. F. Hess, über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten, *Physikalische Zeitschrift*, vol. 13, pp. 1084–1091, 1912.
- [6] V. F. Hess, Untersuchungen über die Strahlung während einer Sonnenfinsternis, *Physikalische Zeitschrift*, vol. 13, pp. 1104–1110, 1912.
- [7] Y. Sekido and H. Elliot, Eds., *Early History of Cosmic Ray Studies: Personal Reminiscences with Old Photographs* (Astrophysics and Space Science Library). Dordrecht: D. Reidel Publishing Company, 1985, vol. 118, ISBN: 9789027720832. DOI: [10.1007/978-94-009-5434-2](https://doi.org/10.1007/978-94-009-5434-2).
- [8] J. Clay, Penetrating Radiation II. Latitude Effect, *Proceedings of the Royal Academy of Amsterdam*, vol. 30, pp. 1115–1127, 1927.
- [9] E. Steinke, über Schwankungen und Barometereffekt der kosmischen Ultrastrahlung im Meeresniveau, *Zeitschrift für Physik*, vol. 64, pp. 48–63, 1930. DOI: [10.1007/BF01397427](https://doi.org/10.1007/BF01397427).
- [10] E. Regener, Die Absorptionskurve der Ultrastrahlung und ihre Deutung, *Physikalische Zeitschrift*, vol. 34, pp. 306–323, 1933.
- [11] J. R. Hörandel, Early Cosmic-Ray Work Published in German, *arXiv preprint arXiv:1212.0706*, 2012.
- [12] W. Bothe and W. Kolhörster, Das Wesen der Höhenstrahlung, *Zeitschrift für Physik*, vol. 56, pp. 751–777, 1929. DOI: [10.1007/BF01340137](https://doi.org/10.1007/BF01340137).
- [13] W. Kolhörster, I. Matthes, and E. Weber, Gekoppelte Höhenstrahlen, *Naturwissenschaften*, vol. 26, p. 576, 1938. DOI: [10.1007/BF01773491](https://doi.org/10.1007/BF01773491).
- [14] B. Rossi, On the Magnetic Deflection of Cosmic Rays, *Physical Review*, vol. 36, pp. 606–607, 1930. DOI: [10.1103/PhysRev.36.606](https://doi.org/10.1103/PhysRev.36.606).

- [15] B. Rossi, Directional Measurements on the Cosmic Rays Near the Geomagnetic Equator, *Physical Review*, vol. 45, pp. 223–225, 1934. DOI: [10.1103/PhysRev.45.223](https://doi.org/10.1103/PhysRev.45.223).
- [16] P. Auger *et al.*, Extensive Cosmic-Ray Showers, *Reviews of Modern Physics*, vol. 11, pp. 288–291, 1939. DOI: [10.1103/RevModPhys.11.288](https://doi.org/10.1103/RevModPhys.11.288).
- [17] C. M. G. Lattes *et al.*, Processes Involving Charged Mesons, *Nature*, vol. 159, pp. 694–697, 1947. DOI: [10.1038/159694a0](https://doi.org/10.1038/159694a0).
- [18] A. A. Watson, A Brief History of the Study of High Energy Cosmic Rays using Arrays of Surface Detectors, in *EPJ Web of Conferences*, vol. 283, 2023, p. 01 002. DOI: [10.1051/epjconf/202328301002](https://doi.org/10.1051/epjconf/202328301002).
- [19] C. Y. Fan, P. Meyer, and J. A. Simpson, Preliminary Results from the Space Probe Pioneer V, *Journal of Geophysical Research*, vol. 65, no. 6, pp. 1862–1863, 1960. DOI: [10.1029/JZ065i006p01862](https://doi.org/10.1029/JZ065i006p01862).
- [20] W. R. Webber *et al.*, Voyager 1 Measurements Beyond the Heliopause of Galactic Cosmic Ray Helium, Boron, Carbon, Oxygen, Magnesium, Silicon and Iron Nuclei with Energies 0.5 to >1.5 GeV/nuc, *arXiv preprint arXiv:1712.02818*, 2017.
- [21] D. Maurin, A. Cheminet, and I. V. Moskalenko, Cosmic-ray databases: Overview, current status, and future prospects, *arXiv preprint*, 2023. arXiv: [2306.08901](https://arxiv.org/abs/2306.08901) [[astro-ph.HE](#)].
- [22] S. Navas *et al.*, Review of particle physics, *Physical Review D*, vol. 110, no. 3, 2024.
- [23] V. Ptuskin, Propagation of galactic cosmic rays, *Astroparticle Physics*, vol. 39, pp. 44–51, 2012.
- [24] A. W. Strong, I. V. Moskalenko, and V. S. Ptuskin, Cosmic-ray propagation and interactions in the Galaxy, *Annu. Rev. Nucl. Part. Sci.*, vol. 57, no. 1, pp. 285–327, 2007.
- [25] M. Kachelriess, Lecture notes on high energy cosmic rays, *arXiv preprint arXiv:0801.4376*, 2008.
- [26] K. Greisen, End to the Cosmic-Ray Spectrum?, *Physical Review Letters*, vol. 16, no. 17, pp. 748–750, 1966. DOI: [10.1103/PhysRevLett.16.748](https://doi.org/10.1103/PhysRevLett.16.748).
- [27] G. T. Zatsepin and V. A. Kuzmin, Upper Limit of the Spectrum of Cosmic Rays, *Soviet Journal of Experimental and Theoretical Physics Letters*, vol. 4, pp. 78–80, 1966.
- [28] J. Puget, F. Stecker, and J. Bredekamp, Photonuclear interactions of ultrahigh energy cosmic rays and their astrophysical consequences, *Astrophysical Journal*, vol. 205, Apr. 15, 1976, pt. 1, p. 638–654., vol. 205, pp. 638–654, 1976.
- [29] D. Allard, Extragalactic propagation of ultrahigh energy cosmic-rays, *Astroparticle Physics*, vol. 39, pp. 33–43, 2012.
- [30] N. Hayashida *et al.*, Updated AGASA event list above 4×10^{19} eV, *arXiv preprint astro-ph/0008102*, 2000.
- [31] Pierre Auger Collaboration, Measurement of the energy spectrum of cosmic rays above 10^{18} eV using the Pierre Auger Observatory, *Physics Letters B*, vol. 685, no. 4, pp. 239–246, 2010, ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2010.02.013>.
- [32] T. Abu-Zayyad *et al.*, The cosmic-ray energy spectrum observed with the surface detector of the telescope array experiment, *The Astrophysical Journal Letters*, vol. 768, no. 1, p. L1, 2013.
- [33] M. S. Muzio, L. A. Anchordoqui, and M. Unger, Peters cycle at the end of the cosmic ray spectrum?, *Physical Review D*, vol. 109, no. 2, p. 023 006, 2024.

- [34] E. Fermi, On the origin of the cosmic radiation, *Physical review*, vol. 75, no. 8, p. 1169, 1949.
- [35] L. Davis Jr, Modified Fermi mechanism for the acceleration of cosmic rays, *Physical Review*, vol. 101, no. 1, p. 351, 1956.
- [36] A. M. Hillas, The origin of ultra-high-energy cosmic rays, IN: *Annual review of astronomy and astrophysics. Volume 22. Palo Alto, CA, Annual Reviews, Inc., 1984, p. 425-444.*, vol. 22, pp. 425–444, 1984.
- [37] M. Amenomori *et al.*, First detection of photons with energy beyond 100 TeV from an astrophysical source, *Physical review letters*, vol. 123, no. 5, p. 051 101, 2019.
- [38] LHAASO Collaboration, An ultrahigh-energy γ -ray bubble powered by a super PeVatron, *Science Bulletin*, vol. 69, no. 4, pp. 449–457, 2024.
- [39] F. Aharonian, Very high energy gamma rays and origin of galactic cosmic rays, *eConf*, vol. 41213, p. 0043, 2004.
- [40] Z. Cao *et al.*, Ultra-high-energy gamma-ray astronomy, *Annual Review of Nuclear and Particle Science*, vol. 73, no. 1, pp. 341–363, 2023.
- [41] M. S. Longair, *High energy astrophysics*. Cambridge university press, 2011.
- [42] A. Hillas, Evolution of ground-based gamma-ray astronomy from the early days to the Cherenkov Telescope Arrays, *Astroparticle Physics*, vol. 43, pp. 19–43, 2013.
- [43] G. Di Sciascio, Ground-based gamma-ray astronomy: An introduction, in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1263, 2019, p. 012 003.
- [44] K.-H. Kampert *et al.*, Multi-messenger physics with the Pierre Auger Observatory, *Frontiers in astronomy and space sciences*, vol. 6, p. 24, 2019.
- [45] Pierre Auger Collaboration, Search for photons above 10^{18} eV by simultaneously measuring the atmospheric depth and the muon content of air showers at the Pierre Auger Observatory, *Physical Review D*, vol. 110, no. 6, p. 062 005, 2024.
- [46] Pierre Auger Collaboration, Search for ultra-high-energy photons from gravitational wave sources with the Pierre Auger Observatory, *The Astrophysical Journal*, vol. 952, no. 1, p. 91, 2023.
- [47] Pierre Auger Collaboration, Search for photons above 1019 eV with the surface detector of the Pierre Auger Observatory, *Journal of cosmology and astroparticle physics*, vol. 2023, no. 05, p. 021, 2023.
- [48] P. A. Collaboration, Searches for ultra-high-energy photons at the Pierre Auger Observatory, *Universe*, vol. 8, no. 11, p. 579, 2022.
- [49] Pierre Auger Collaboration, A search for photons with energies above 2×10^{17} eV using hybrid data from the low-energy extensions of the Pierre Auger Observatory, *The Astrophysical Journal*, vol. 933, no. 2, p. 125, 2022.
- [50] Pierre Auger Collaboration, Search for photons with energies above 10^{18} eV using the hybrid detector of the Pierre Auger Observatory, *Journal of Cosmology and Astroparticle Physics*, vol. 2017, no. 04, p. 009, 2017.
- [51] Pierre Auger Collaboration, A search for point sources of EeV photons, *The Astrophysical Journal*, vol. 789, no. 2, p. 160, 2014.
- [52] Pierre Auger Collaboration, Search for a diffuse flux of photons with energies above tens of PeV at the Pierre Auger Observatory, *arXiv preprint arXiv:2502.02381*, 2025.
- [53] W. D. Apel *et al.*, KASCADE-grande limits on the isotropic diffuse gamma-ray flux between 100 TeV and 1 EeV, *The Astrophysical Journal*, vol. 848, no. 1, p. 1, 2017.

- [54] Y. A. Fomin *et al.*, Constraints on the flux of (10 16- 1 0 17.5) eV cosmic photons from the EAS-MSU muon data, *Physical Review D*, vol. 95, no. 12, p. 123 011, 2017.
- [55] R. U. Abbasi *et al.*, Constraints on the diffuse photon flux with energies above 10^{18} eV using the surface detector of the Telescope Array experiment, *Astroparticle physics*, vol. 110, pp. 8–14, 2019.
- [56] M. Kuznetsov, O. Kalashev, and G. Rubtsov, Telescope Array search for EeV photons, *PoS ICRC2019*, vol. 326, 2020.
- [57] Pierre Auger Collaboration, Latest results from the searches for ultra-high-energy photons and neutrinos at the Pierre Auger Observatory, in *38th International Cosmic Ray Conference (ICRC2023)-Multi Messenger and Gravitational Wave (MM&GW)*, Sissa Medialab, 2024.
- [58] A. Bobrikova *et al.*, Predicting the UHE photon flux from GZK-interactions of hadronic cosmic rays using CRPropa 3, in *Proceedings of the 37th International Cosmic Ray Conference PoS (ICRC2021), Berlin, Germany*, 2021, pp. 12–23.
- [59] B. Sarkar, K.-H. Kampert, and J. Kulbartz, Ultra-High Energy Photon and Neutrino Fluxes in Realistic Astrophysical Scenarios, in *32nd International Cosmic Ray Conference*, vol. 2, 2011, p. 198. DOI: [10.7529/ICRC2011/V02/1087](https://doi.org/10.7529/ICRC2011/V02/1087).
- [60] C. Bérat *et al.*, Diffuse flux of ultra-high-energy photons from cosmic-ray interactions in the disk of the galaxy and implications for the search for decaying super-heavy dark matter, *The Astrophysical Journal*, vol. 929, no. 1, p. 55, 2022.
- [61] O. E. Kalashev and S. V. Troitsky, IceCube astrophysical neutrinos without a spectral cutoff and $10^{15} - 10^{17}$ eV cosmic gamma radiation, *Pisma Zh. Eksp. Teor. Fiz.*, vol. 100, no. 12, pp. 865–869, 2014. DOI: [10.1134/S0021364014240072](https://doi.org/10.1134/S0021364014240072). arXiv: [1410.2600](https://arxiv.org/abs/1410.2600) [astro-ph.HE].
- [62] M. Kachelriess, O. E. Kalashev, and M. Y. Kuznetsov, Heavy decaying dark matter and IceCube high energy neutrinos, *Phys. Rev. D*, vol. 98, no. 8, p. 083 016, 2018. DOI: [10.1103/PhysRevD.98.083016](https://doi.org/10.1103/PhysRevD.98.083016). arXiv: [1805.04500](https://arxiv.org/abs/1805.04500) [astro-ph.HE].
- [63] L. A. Anchordoqui *et al.*, Hunting super-heavy dark matter with ultra-high energy photons, *Astropart. Phys.*, vol. 132, p. 102 614, 2021. DOI: [10.1016/j.astropartphys.2021.102614](https://doi.org/10.1016/j.astropartphys.2021.102614). arXiv: [2105.12895](https://arxiv.org/abs/2105.12895) [hep-ph].
- [64] T. K. Gaisser, R. Engel, and E. Resconi, *Cosmic rays and particle physics*. Cambridge University Press, 2016.
- [65] M. Stadelmaier, On air-shower universality and the mass composition of ultra-high-energy cosmic rays, Ph.D. dissertation, Dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), 2022, 2022.
- [66] B. Rossi and K. Greisen, Cosmic-ray theory, *Reviews of Modern Physics*, vol. 13, no. 4, p. 240, 1941.
- [67] K. Greisen, The extensive air showers, *Progress in cosmic ray physics*, vol. 3, no. 1, 1956.
- [68] P. Lipari, Concepts of “age” and “universality” in cosmic ray showers, *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, vol. 79, no. 6, p. 063 001, 2009.
- [69] J. Nishimura and K. Kamata, The lateral and angular distribution of cascade showers, *Progress of Theoretical Physics*, vol. 5, no. 5, pp. 899–901, 1950.
- [70] J. Nishimura and K. Kamata, On the theory of cascade showers, i, *Progress of Theoretical Physics*, vol. 7, no. 2, pp. 185–192, 1952.

- [71] K. Kamata and J. Nishimura, The lateral and the angular structure functions of electron showers, *Progress of Theoretical Physics Supplement*, vol. 6, pp. 93–155, 1958.
- [72] K. Greisen, Cosmic Ray Showers, *Annual Review of Nuclear Science*, vol. 10, pp. 63–108, 1960. DOI: [10.1146/annurev.ns.10.120160.000431](https://doi.org/10.1146/annurev.ns.10.120160.000431).
- [73] R. Engel, D. Heck, and T. Pierog, Extensive air showers and hadronic interactions at high energy, *Annual review of nuclear and particle science*, vol. 61, no. 1, pp. 467–489, 2011.
- [74] Pierre Auger Collaboration, Data-driven estimation of the invisible energy of cosmic ray showers with the Pierre Auger Observatory, *Physical Review D*, vol. 100, no. 8, p. 082003, 2019.
- [75] A. G. Mariazzi, Determination of the invisible energy of extensive air showers from the data collected at Pierre Auger Observatory, in *EPJ Web of Conferences*, EDP Sciences, vol. 210, 2019, p. 02010.
- [76] W. Heitler, *The Quantum Theory of Radiation*: Oxford Univ, 1944.
- [77] J. Carlson and J. Oppenheimer, On multiplicative showers, *Physical Review*, vol. 51, no. 4, p. 220, 1937.
- [78] J. Matthews, A heitler model of extensive air showers, *Astroparticle Physics*, vol. 22, no. 5-6, pp. 387–397, 2005.
- [79] D. Heck *et al.*, CORSIKA: A Monte Carlo code to simulate extensive air showers, *Report fzka*, vol. 6019, no. 11, 1998.
- [80] T. Bergmann *et al.*, One-dimensional hybrid approach to extensive air shower simulation, *Astroparticle Physics*, vol. 26, no. 6, pp. 420–432, 2007.
- [81] P. Billoir, A sampling procedure to regenerate particles in a ground detector from a “thinned” air shower simulation output, *Astroparticle Physics*, vol. 30, no. 5, pp. 270–285, 2008.
- [82] S. Ostapchenko, Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model, *Physical Review D—Particles, Fields, Gravitation, and Cosmology*, vol. 83, no. 1, p. 014018, 2011.
- [83] S. Ostapchenko, QGSJET-III model of high energy hadronic interactions: The formalism, *Physical Review D*, vol. 109, no. 3, p. 034002, 2024.
- [84] F. Riehn *et al.*, A new version of the event generator Sibyll, *arXiv preprint arXiv:1510.00568*, 2015.
- [85] F. Riehn *et al.*, The hadronic interaction model SIBYLL 2.3 c and Feynman scaling, *arXiv preprint arXiv:1709.07227*, 2017.
- [86] T. Pierog *et al.*, EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider, *Physical Review C*, vol. 92, no. 3, p. 034906, 2015.
- [87] T. Pierog and K. Werner, EPOS. LHC-R: a global approach to solve the muon puzzle, *arXiv preprint arXiv:2508.07105*, 2025.
- [88] A. Capella *et al.*, Dual parton model, *Physics Reports*, vol. 236, no. 4-5, pp. 225–329, 1994.
- [89] H. Fesefeldt, Report PITHA-85/02, RWTH Aachen, vol. 647, p. 27, 1985.
- [90] A. Ferrari *et al.*, *FLUKA: A multi-particle transport code (Program version 2005)*. Cern, 2005.

- [91] M. Bleicher *et al.*, Relativistic hadron-hadron collisions in the ultra-relativistic quantum molecular dynamics model, *Journal of Physics G: Nuclear and Particle Physics*, vol. 25, no. 9, p. 1859, 1999.
- [92] C. Meurer *et al.*, Muon production in extensive air showers and its relation to hadronic interactions, *Czechoslovak Journal of Physics*, vol. 56, no. Suppl 1, A211–A219, 2006.
- [93] H.-J. Drescher and G. R. Farrar, Dominant contributions to lateral distribution functions in ultra-high energy cosmic ray air showers, *Astroparticle Physics*, vol. 19, no. 2, pp. 235–244, 2003.
- [94] H.-J. Drescher *et al.*, Model dependence of lateral distribution functions of high energy cosmic ray air showers, *Astroparticle Physics*, vol. 21, no. 1, pp. 87–94, 2004.
- [95] F. Schmidt and J. Knapp, CORSIKA Shower Images, <https://www-zeuthen.desy.de/~jknapp/fs/showerimages.html>, Accessed: 2025-08-26, 2005.
- [96] L. Landau and I. Pomeranchuk, *Limits of applicability of the theory of bremsstrahlung electrons and pair production for high energies*. United States Atomic Energy Commission, Technical Information Service, 1953, vol. 136.
- [97] B. McBreen and C. Lambert, Interactions of high-energy ($E > 5 \times 10^{19}$ eV) photons in the earth's magnetic field, *Physical Review D*, vol. 24, no. 9, p. 2536, 1981.
- [98] A. Watson, Development of the Pierre Auger Observatory, https://auger.org/images/PDFs/how-it-all-began_watson-2006.pdf, Based on a talk delivered in 2006 in honor of Jim Cronin's 75th birthday, 2006.
- [99] K.-H. Kampert and A. A. Watson, Extensive air showers and ultra high-energy cosmic rays: A historical review, *The European Physical Journal H*, vol. 37, no. 3, pp. 359–412, 2012.
- [100] Pierre Auger Collaboration and others, The pierre auger cosmic ray observatory, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 798, pp. 172–213, 2015.
- [101] P. D. J. Clark and D. Nitz for the Pierre Auger Collaboration, Communications in the Auger Observatory, *PoS ICRC2001*, vol. 2, 2001.
- [102] X. Bertou *et al.*, Calibration of the surface array of the Pierre Auger Observatory, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 568, no. 2, pp. 839–846, 2006.
- [103] P. Allison *et al.*, Observing muon decays in water Cherenkov detectors at the Pierre Auger Observatory, *arXiv preprint astro-ph/0509238*, 2005.
- [104] Pierre Auger Collaboration, Trigger and aperture of the surface detector array of the Pierre Auger Observatory, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 613, no. 1, pp. 29–39, 2010.
- [105] Pierre Auger Collaboration, The energy spectrum of cosmic rays beyond the turn-down around 10¹⁷ eV as measured with the surface detector of the Pierre Auger Observatory, *The European Physical Journal C*, vol. 81, no. 11, p. 966, 2021.
- [106] Pierre Auger Collaboration, Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory, *Journal of Instrumentation*, vol. 15, no. 10, P10021, 2020.
- [107] P. A. collaboration *et al.*, Performance of the 433 m surface array of the Pierre Auger Observatory, *Proceedings of Science*, vol. 395, p. 224, 2022.

- [108] Pierre Auger Collaboration, The fluorescence detector of the Pierre Auger Observatory, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 620, no. 2-3, pp. 227–251, 2010.
- [109] M. Unger for the Pierre Auger Collaboration, Highlights from the Pierre Auger Observatory, *arXiv preprint arXiv:1710.09478*, 2017.
- [110] M. Unger for the Pierre Auger Collaboration, Mass sensitive observables of the Pierre Auger Observatory, in *EPJ Web of Conferences*, EDP Sciences, vol. 53, 2013, p. 04 009.
- [111] Pierre Auger Collaboration, Description of atmospheric conditions at the Pierre Auger Observatory using the global data assimilation system (GDAS), *Astroparticle Physics*, vol. 35, no. 9, pp. 591–607, 2012.
- [112] Pierre Auger Collaboration, Techniques for measuring aerosol attenuation using the Central Laser Facility at the Pierre Auger Observatory, *Journal of Instrumentation*, vol. 8, no. 04, P04009, 2013.
- [113] Pierre Auger Collaboration, The lidar system of the Pierre Auger Observatory, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 574, no. 1, pp. 171–184, 2007.
- [114] Pierre Auger Collaboration, The pierre auger observatory upgrade-preliminary design report, *arXiv preprint arXiv:1604.03637*, 2016.
- [115] E. Waxman, Cosmological origin for cosmic rays above 10^{19} ev, *The Astrophysical Journal*, vol. 452, no. 1, p. L1, 1995.
- [116] A. M. Van den Berg for the Pierre Auger Collaboration, Radio detection of high-energy cosmic rays at the Pierre Auger Observatory, *arXiv preprint arXiv:0708.1709*, 2007.
- [117] T. Huege, The Radio Detector of the Pierre Auger Observatory—status and expected performance, in *EPJ Web of Conferences*, EDP Sciences, vol. 283, 2023, p. 06 002.
- [118] SD Electronics Hardware Specifications, https://atrium.in2p3.fr/nuxeo/nxdoc/default/90879bec-97c3-4dbc-a200-bfa7ec7bb231/view_documents, Accessed: 2025-01-27, Pierre Auger Collaboration.
- [119] B. Wundheiler, Rayos cósmicos con energías entre $1E17$ y $1E19$ eV, Ph.D. dissertation, Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales, 2013.
- [120] F. Gesualdi, The muon content of atmospheric air showers and the mass composition of cosmic rays, Ph.D. dissertation, Dissertation, Karlsruhe, Karlsruher Institut für Technologie (KIT), 2022, 2023.
- [121] 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, Ph.D. dissertation, 2019.
- [122] 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, *The European Physical Journal C*, vol. 80, no. 8, p. 751, 2020.
- [123] M. R. Hampel, Optoelectrónica con Fotomultiplicadores de Silicio para Detectores de Rayos Cósmicos, Ph.D. dissertation, Universidad Tecnológica Nacional, 2018.
- [124] A. Fuster, Diseño y calibración de un front-end analógico con fotomultiplicadores de silicio para detectores de rayos cósmicos, Ph.D. dissertation, Universidad Tecnológica Nacional, 2021.

- [125] M. A. Scornavacche, The muon content of extensive air showers measured by the Underground Muon Detector of the Pierre Auger Observatory, Ph.D. dissertation, 2025.
- [126] J. d. Jesús, Study of the mass composition of cosmic rays with the Underground Muon Detector of AMIGA, Ph.D. dissertation, 2025.
- [127] Pierre Auger Collaboration, Design and implementation of the AMIGA embedded system for data acquisition, *Journal of instrumentation*, vol. 16, no. 07, T07008, 2021.
- [128] D. Ravignani and A. D. Supanitsky, A new method for reconstructing the muon lateral distribution with an array of segmented counters, *Astroparticle Physics*, vol. 65, pp. 1–10, 2015.
- [129] A. Supanitsky, Estimation of the number of muons with muon counters, *Astroparticle Physics*, vol. 127, p. 102 535, 2021.
- [130] J. Van Buren, Muon size spectrum measured by KASCADE-Grande, 2005.
- [131] A. Lagutin and R. Raikin, Lateral distribution of electrons in EAS at superhigh energies: predictions and experimental data, *Nuclear Physics B-Proceedings Supplements*, vol. 97, no. 1-3, pp. 274–277, 2001.
- [132] e. W. Apel *et al.*, The KASCADE-grande experiment, *Nuclear Instruments and Methods in Physics Research Section A: accelerators, spectrometers, detectors and associated equipment*, vol. 620, no. 2-3, pp. 202–216, 2010.
- [133] S. Argiro *et al.*, The offline software framework of the pierre auger observatory, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 580, no. 3, pp. 1485–1496, 2007.
- [134] J. G. Gonzalez, The Offline Software of the Pierre Auger Observatory: Lessons Learned, in *2012 IEEE 10th International Symposium on Parallel and Distributed Processing with Applications*, IEEE, 2012, pp. 557–564.
- [135] V. Novotny *et al.*, Napoli + Praha library. Comparison between the Napoli and Praha CORSIKA productions, GAP-2019-068.
- [136] H. Dembinski *et al.*, Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV, *PoS*, vol. ICRC2017, DOI: [10.22323/1.301.0533](https://doi.org/10.22323/1.301.0533).
- [137] J. A. Nelder and R. W. Wedderburn, Generalized linear models, *Journal of the Royal Statistical Society Series A: Statistics in Society*, vol. 135, no. 3, pp. 370–384, 1972.
- [138] G. Silli *et al.*, Study of the efficiency for the SD-433 array, GAP-2021-001.
- [139] A. Coleman, Measurement of the Cosmic Ray Flux Above 100 PeV at the Pierre Auger Observatory, GAP-2018-045.
- [140] O. Deligny, Response function of the SD1500 array, GAP-2020-012.
- [141] J.C. Arteaga-Velázquez, A report by the WHISP working group on the combined analysis of muon data at cosmic-ray energies above 1 PeV, *PoS*, vol. ICRC2023, DOI: [10.22323/1.444.0466](https://doi.org/10.22323/1.444.0466).
- [142] G. Brichetto Orquera for the Pierre Auger Collaboration, The second knee in the cosmic ray spectrum observed with the surface detector of the Pierre Auger Observatory, *PoS*, vol. ICRC2023, DOI: [10.22323/1.444.0398](https://doi.org/10.22323/1.444.0398).
- [143] D. Veberič *et al.*, Constant Intensity Cut: Unbinned Estimation of the Signal Attenuation Function, GAP-2015-065.
- [144] M. Rahman, L. M. Pearson, and H. C. Heien, A modified Anderson-Darling test for uniformity, *Bulletin of the Malaysian Mathematical Sciences Society*, vol. 29, no. 1, 2006.

- [145] L. Cazon, Working group report on the combined analysis of muon density measurements from eight air shower experiments, *PoS*, vol. ICRC2019,
- [146] H. P. Dembinski, Computing mean logarithmic mass from muon counts in air shower experiments, *Astroparticle Physics*, vol. 102, pp. 89–94, 2018.
- [147] A. A. Watson, Lateral Distribution Measurements of the Muon and Electromagnetic Components of Showers and a Comment on some Measurements of the Muon Excess, GAP-2023-045, 2023.
- [148] W. D. Apel *et al.*, Probing the Evolution of the EAS Muon Content in the Atmosphere with KASCADE-Grande, *Astroparticle Physics*, vol. 95, pp. 25–43, 2017. DOI: [10.1016/j.astropartphys.2017.07.001](https://doi.org/10.1016/j.astropartphys.2017.07.001).
- [149] N. Hayashida *et al.*, Muons (≥ 1 GeV) in large extensive air showers of energies between 1016.5 eV and 1019.5 eV observed at Akeno, *Journal of Physics G: Nuclear and Particle Physics*, vol. 21, no. 8, pp. 1101–1119, 1995.
- [150] S. Bennett and K. Greisen, Energy Spectrum and Positive Excess of Muons in Cosmic-Ray Air Showers, *Physical Review*, vol. 124, no. 6, pp. 1982–1987, 1961. DOI: [10.1103/PhysRev.124.1982](https://doi.org/10.1103/PhysRev.124.1982).
- [151] H. Dixon *et al.*, Recent measurements of the muon component of large cosmic ray air showers, *Journal of Physics A: Mathematical, Nuclear and General*, vol. 7, no. 8, pp. 1010–1016, 1974.
- [152] J. Earnshaw *et al.*, The momenta of muons in extensive air showers, *Proceedings of the Physical Society*, vol. 90, no. 1, pp. 91–108, 1967.
- [153] E. Rodriguez, N. González, and B. Wundheiler, The Muon Lateral Distribution Function of the UMD-433, GAP-2026-011, 2026.
- [154] 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, no. 07, P07016, 2021.
- [155] 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. DOI: [10.1016/j.astropartphys.2006.08.003](https://doi.org/10.1016/j.astropartphys.2006.08.003).
- [156] P. Meder *et al.*, New Lever-Arm Criteria to Fit the LDF Slope Parameters for the SD-750 Array, GAP-2024-033, May 2024.
- [157] E. Rodriguez *et al.*, On the shape of the mean Muon Lateral Distribution Function of the UMD, GAP-2026-018.
- [158] E. dos Santos, Grid computing for VO Auger members. GAP-2020-011.
- [159] F. Gesualdi, A. D. Supanitsky, and A. Etchegoyen, Muon deficit in air shower simulations estimated from AGASA muon measurements, *Phys. Rev. D*, vol. 101, p. 083 025, 2020.
- [160] E. Rodriguez *et al.*, Revisiting the SD-433 efficiency, GAP-2024-051, Sep. 2024.
- [161] A. Yushkov for the Pierre Auger Collaboration, Mass composition of cosmic rays with energies above $10^{17.2}$ eV from the hybrid data of the pierre auger observatory, vol. ICRC2019.
- [162] 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. 6537, pp. 1266–1270, 2017. DOI: [10.1126/science.aan4338](https://doi.org/10.1126/science.aan4338).

- [163] Pierre Auger Collaboration, Large-Scale Distribution of Arrival Directions of Cosmic Rays Detected above 10^{18} eV at the Pierre Auger Observatory, *The Astrophysical Journal Supplement Series*, vol. 203, no. 2, p. 34, 2012. DOI: [10.1088/0067-0049/203/2/34](https://doi.org/10.1088/0067-0049/203/2/34).
- [164] A. D. Supanitsky, Determination of the cosmic-ray chemical composition: Open issues and prospects, *Galaxies*, vol. 10, no. 3, p. 75, 2022.
- [165] R. U. Abbasi, Telescope Array Collaboration, *et al.*, Cosmic ray mass composition measurement in the energy range from $10^{16.5}$ eV to $10^{18.5}$ eV observed with the TALE hybrid detector, 2026. arXiv: [2603.14804](https://arxiv.org/abs/2603.14804) [[astro-ph.HE](#)].
- [166] R. Abbasi *et al.*, Density of GeV muons in air showers measured with IceTop, *Physical Review D*, vol. 106, no. 3, p. 032 010, 2022.
- [167] G. Nicolás M. and M. Ioana. C., Parametrization of the lateral distribution function for photon primaries with the SD-433, GAP-2021-041, Oct. 2021.
- [168] G. N. *et al.*, Accuracy of the event geometry reconstruction with the SD-433, GAP-2020-038, Jul. 2020.
- [169] G. Nicolás M. and M. Ioana. C., Reconstruction of the primary energy for photon initiated air-showers simulated for the sd-433 configuration, GAP-2021-056, Aug. 2021.
- [170] J. Pawlowsky, K.-H. Kampert, and J. Rautenberg, Photon-likeness of hadron showers and impact of Lorentz boosting, *The Astrophysical Journal*, vol. 983, no. 1, p. 12, 2025.
- [171] E. Rodriguez, Overview of Machine Learning Applications at the Pierre Auger Observatory, in *EPJ Web of Conferences*, EDP Sciences, vol. 319, 2025, p. 13 006.
- [172] S. Hahn, P. A. Collaboration, *et al.*, Machine learning-based analyses using surface detector data of the Pierre Auger Observatory, in *Proceedings of 7th International Symposium on Ultra High Energy Cosmic Rays (UHECR2024)*. Ed.: I. Allekotte, 2025, p. 091.
- [173] M. Kubátová and P. A. Collaboration, Machine Learning Applications at the Pierre Auger Observatory, in *Journal of Physics: Conference Series*, IOP Publishing, vol. 3053, 2025, p. 012 011.
- [174] F. Ellwanger for the Pierre Auger Collaboration, Investigations of a novel energy estimator using deep learning for the surface detector of the Pierre Auger Observatory, in *38th International Cosmic Ray Conference (ICRC2023)-Cosmic-Ray Physics (Indirect, CRI)*, Sissa Medialab, 2024, p. 275.
- [175] Pierre Auger Collaboration, Deep-learning based reconstruction of the shower maximum $X_{\max}$ using the water-Cherenkov detectors of the Pierre Auger Observatory, *Journal of instrumentation*, vol. 16, no. 07, P07019, 2021.
- [176] Pierre Auger Collaboration, Inference of the mass composition of cosmic rays with energies from $10^{18.5}$ to 10^{20} eV using the Pierre Auger Observatory and deep learning, *Physical review letters*, vol. 134, no. 2, p. 021 001, 2025.
- [177] S. Hahn for the Pierre Auger Collaboration, Reconstruction of muon number of air showers with the surface detector of the Pierre Auger Observatory using neural networks, in *38th International Cosmic Ray Conference (ICRC2023)-Cosmic-Ray Physics (Indirect, CRI)*, Sissa Medialab, 2024, p. 318.
- [178] E. Rodriguez, Graph Neural Networks for Photon Searches with the Underground Muon Detector of the Pierre Auger Observatory, *arXiv preprint arXiv:2507.09287*, 2025.
- [179] F. Ellwanger *et al.*, Partial De/Convolution of SD Traces for Reversal of Ageing Effects, GAP-2022-039, 2022.

- [180] M. Fey and J. E. Lenssen, Fast graph representation learning with PyTorch Geometric, *arXiv preprint arXiv:1903.02428*, 2019.
- [181] L. Wu *et al.*, Graph neural networks: Foundation, frontiers and applications, in *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, 2022, pp. 4840–4841.
- [182] C. Liu *et al.*, Graph pooling for graph neural networks: Progress, challenges, and opportunities, *arXiv preprint arXiv:2204.07321*, 2022.
- [183] S. Ioffe and C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in *International conference on machine learning*, pmlr, 2015, pp. 448–456.
- [184] P. Veličković *et al.*, Graph attention networks, *arXiv preprint arXiv:1710.10903*, 2017.
- [185] S. Brody, U. Alon, and E. Yahav, How attentive are graph attention networks?, *arXiv preprint arXiv:2105.14491*, 2021.
- [186] C. F. Dormann, Calibration of probability predictions from machine-learning and statistical models, *Global ecology and biogeography*, vol. 29, no. 4, pp. 760–765, 2020.
- [187] E. Rodriguez, P. A. Collaboration, *et al.*, Neural Networks for Photon Searches with AugerPrime, in *Proceedings of 7th International Symposium on Ultra High Energy Cosmic Rays (UHECR 2024)*, 2025, p. 111.
- [188] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980*, 2014.
- [189] D. Heck, Production of Vector Mesons by Photonuclear Interactions in the Program CORSIKA, Tech. Rep. Forschungszentrum Karlsruhe, Tech. Rep., 2012.
- [190] N. González, A. Castellina, and M. Bertaina, Photon-hadron discrimination at tens of PeV: Revisiting the definition of Mb in Phase II, GAP-2025-051.
- [191] P. Savina, Enhancing our sensitivity on photon analyses, GAP2025-026.
- [192] N. González, C. Bonifazi, and I. Maris, The exposure of the 433 m array, GAP-2022-041.
- [193] N. M. González, I. C. Mariş, and F. A. Sánchez, A method to determine good data-taking periods for the underground muon detectors, GAP-2023-039.
- [194] Pierre Auger Collaboration, Search for photons above 10^{19} eV with the surface detector of the Pierre Auger Observatory, *JCAP*, vol. 05, p. 021, 2023. arXiv: [2209.05926 \[astro-ph.HE\]](#).
- [195] Y. A. Fomin *et al.*, Constraints on the flux of (10 16- 1 0 17.5) eV cosmic photons from the EAS-MSU muon data, *Phys. Rev. D*, vol. 95, p. 123 011, 2017.
- [196] R. Aloisio, V. Berezhinsky, and M. Kachelriess, Status of superheavy dark matter, *Phys. Rev. D*, vol. 74, p. 023 516, 2006.
- [197] O. E. Kalashev and M. Y. Kuznetsov, Constraining heavy decaying dark matter with the high energy gamma-ray limits, *Phys. Rev. D*, vol. 94, p. 063 535, 2016. arXiv: [1606.07354 \[astro-ph.HE\]](#).
- [198] M. Schimassek, D. Veberič, and R. Engel, Useful Parameterisations of SD Simulations, GAP-2019-075.
- [199] G. Brichetto Orquera, Diego Ravignani, Measurement of the SD1500 efficiency with the SD750, GAP-2021-066.

- [200] C. Pérez Bertolli *et al.*, Towards the development of a Protocol for UMD Shifts, GAP-2022-043, 2022.
- [201] M. Scornavacche *et al.*, Monitoring of the ADCT1 files in the Underground Muon Detector, Internal note, Pierre Auger Collaboration, 2024.
- [202] B. Andrada, F. Gollan, and F. A. Sánchez, UMD shift: behind the scenes, GAP-2024-038, Jul. 2024.
- [203] P. K. Grieder, *Cosmic rays at earth*. Elsevier Science BV, 2001.
- [204] A. V. Glushkov, K. G. Lebedev, and A. V. Sabourov, Content of Muons in Extensive Air Showers Measured by Yakutsk Array, Proceedings of the 37th ICRC; accepted for publication in Bulletin of the Russian Academy of Sciences: Physics, 2023. arXiv: [2301.12268 \[astro-ph.HE\]](#).
- [205] C. Pérez Bertolli, Cross-Validation of Muon Content in Extensive Air Showers with Surface and Underground Detectors of AugerPrime, Ph.D. dissertation, 2025.
- [206] European Commission, Directorate-General for Research and Innovation, Living guidelines on the responsible use of generative ai in research, ERA Forum Stakeholders' document, third version, May 2026, 2026.
- [207] OpenAI, Chatgpt, Accessed multiple times during thesis preparation. Statement of use in Appendix C.
- [208] Anthropic, Claude, Accessed multiple times during thesis preparation. Statement of use in Appendix C.

Acronyms

ADC analog-to-digital converter. 42, 50, 53, 54

ADST Auger Data Summary Tree. 60

AERA Auger Engineering Radio Array. 50

AGASA Akeno Giant Air Shower Array. 11, 85, 87, 89, 90, 105, 108, 236

AGN Active Galactic Nuclei. 12, 14, 15

AMIGA Auger Muons and Infill for the Ground Array. 52, 60, 84

CDAS Central Data Acquisition System. 37–39, 42, 211, 216, 217

CIC Constant Intensity Cut. 62, 70, 101

CLF Central Laser Facility. 47

CMB cosmic microwave background. 10, 16

CORSIKA COsmic Ray Simulations for KAscade. 29, 31, 111, 120, 121, 123, 135, 213

CR cosmic ray. 21

EAS Extensive Air Shower. 1, 16, 21, 26, 32, 37, 42, 48, 50

EAS-MSU Extensive Air Showers - Moscow State University. 19

EBL extragalactic background light. 10

EM electromagnetic. 26–28, 65

FADC flash analog-to-digital converter. 38, 39, 41, 42

FD Fluorescence Detector. 45–50, 60

FPGA field-programmable gate array. 53, 54

GA Graph Analyzer. 139, 140

GLM Generalized Linear Model. 65, 81, 113

GNN Graph Neural Network. 134, 149, 200, 201

- GRB** Gamma-ray burst. 12, 15
- GSF** Global Spline Fit. 64, 66–73, 77, 107, 109
- GZK** Greisen–Zatsepin–Kuzmin. 10, 11, 16, 17
- H.E.S.S.** High Energy Stereoscopic System. 16
- HAWC** High-Altitude Water Cherenkov. 16
- HEAT** High Elevation Auger Telescopes. 45, 46, 48, 60
- HG** high gain. 53
- IACT** Imaging Atmospheric Cherenkov Technique. 16, 17
- ISM** interstellar medium. 10, 16
- KASCADE** KArllsruhe Shower Core and Array DETector. 19, 58, 84, 85, 87, 88, 236
- LDF** lateral distribution function. 44, 45, 53, 58, 65, 79, 85, 90, 111, 113–117, 119–122, 126, 128, 129, 200
- LG** low gain. 53
- LHAASO** Large High Altitude Air Shower Observatory. 13, 16
- LHC** Large Hadron Collider. 30
- LiDaR** Light Detection and Ranging. 47
- LPM** Landau–Pomeranchuk–Migdal. 32
- MAGIC** Major Atmospheric Gamma-ray Imaging Collaboration. 16
- MC** Monte Carlo. 29, 135, 213
- MLDF** muon lateral distribution function. 54, 58–60, 84, 87, 88, 90–97, 99, 100, 107, 108, 114, 200, 211, 212, 216–221, 228, 236
- MoPS** Multiplicity-of-Positive Steps. 39, 40, 65
- NKG** Nishimura–Kamata–Greisen. 26, 44, 58, 59, 62, 218
- PMT** photomultiplier tube. 37–39, 42, 43, 45, 49, 50, 52, 53, 136, 212
- POPOP** 1,4-bis(5-phenyloxazole-2-yl)benzene. 51
- PPO** 1,4-bis((2,5-diphenyloxazole. 51
- PWN** pulsar wind nebula. 13
- PyG** PyTorch Geometric. 136
- RD** Radio Detector. 49, 54, 175, 178, 211

- SD** Surface Detector. 1, 2, 16, 19, 37–42, 45–48, 50, 53, 54, 58–66, 68, 70, 72–81, 90–92, 94, 96, 98–100, 102–104, 108, 111–113, 115, 120, 122, 128, 129, 131, 133–141, 143–160, 163–175, 177, 178, 197, 199–201, 211–213, 216, 217, 219
- SHDM** super-heavy dark matter. 18
- SiPM** silicon photomultipliers. 51–54, 212, 216, 217
- SNR** supernova remnant. 12–14
- SPMT** small PMT. 49, 50
- SSD** Scintillator Surface Detector. 48–50, 54, 175, 178
- TA** Trace Analyzer. 138, 139
- TA** Telescope Array. 11, 19
- TALE** Telescope Array Lower energy Extension. 106, 109
- Th** simple threshold. 39, 41, 65
- ToT** time over threshold. 39, 41, 65
- ToTd** Time-over-Threshold-deconvolved. 39, 40, 65
- UB** Unified Board. 50
- UHE** ultra-high-energy. 7, 10, 13, 16, 17, 35, 48
- UHECR** ultra-high-energy cosmic ray. 12, 26, 48
- UMD** Underground Muon Detector. 1, 2, 17–19, 49–56, 58–60, 83–96, 98–101, 104–109, 131, 133–141, 143–160, 163–175, 177, 178, 197, 200, 201, 211–213, 215–223, 234–236, 241, 243
- UUB** Upgraded Unified Board. 50
- VCT** vertical and central through-going. 38
- VEM** Vertical Equivalent Muon. 38, 39, 42, 50
- VERITAS** Very Energetic Radiation Imaging Telescope Array System. 16
- VHE** very-high-energy. 14, 17
- WCD** water-Cherenkov detector. 37, 48–50, 54, 62, 79, 87, 120, 133, 134, 136, 144, 156–160, 218, 222
- WLS** wavelength-shifting. 51–53
- XLF** eXtreme Laser Facility. 47

List of Tables

5.1	Parameters for ε models following Eq. (5.6)	67
5.2	Parameters for ε models following Eq. (5.8).	68
5.3	Parameters for the ε model of Eq. (5.9).	69
6.1	Lever-arm selection used for free- β fits	94
6.2	Best fit parameters of the UMD-433 β parameterisation	95
7.1	Parameters for the photon efficiency model	113
7.2	Lever-arm criteria used for the photon β parametrisation in SD-433	115
7.3	Parameters of the preliminary photon size attenuation model	118
7.4	Parameters of the photon slope parametrisation	119
7.5	Parameters of the photon energy calibration	126
9.1	Integral pure SD and hybrid photon exposures	171
A.1	Parameters of $\varepsilon(E_{\text{SD-433}}, \theta)$ for the GSF.	204
A.2	Profile-likelihood results per $\lg(E_{\text{bin}}/\text{eV})$ bin	209
B.1	Rejection status bit mask: enum value, integer encoding, and string label.	214
B.2	Statistics of rejected UMD modules in Phase I (SD-433 and SD-750 combined)	215
B.3	Enum encoding of MdLDFRecStage.	229
B.4	MdLDF reconstruction stage statistics in the SD-433 test production	230
B.5	MdLDF reconstruction stage statistics in the SD-750 test production	230
B.6	Alternative β parameters	239

List of Figures

2.1	Victor Hess on a gondola	4
2.2	Regener, Hess, and Steinke at Lake Constance	5
2.3	Cosmic-ray spectrum from the Particle Data Group	8
2.4	Hillas plot of cosmic-ray accelerators	13
2.5	Spectral energy distribution of the Crab Nebula	14
2.6	Multi-wavelength composite image of the Crab Nebula	15
2.7	Atmospheric transparency across the electromagnetic spectrum	17
2.8	Integral photon flux limits and model predictions at ultra-high energies	19
3.1	Greisen profiles	25
3.2	Average particle content in the lateral and longitudinal shower development	27
3.3	Heitler–Matthews schematic for EM and hadronic cascades	29
3.4	Simulated photon, proton, and iron showers	31
3.5	Proton vs. photon showers	33
4.1	Cronin, Etchegoyen, and Watson at the inauguration of the PAO	36
4.2	Surface Detector station	37
4.3	Charge and pulse-height histograms from a SD station	38
4.4	Illustration of ToTd and MoPS triggers	40
4.5	T3 hierarchy and spatial constraints	41
4.6	SD event footprint (auger ID 210807738800)	43
4.7	LDF auger id: 210807738800	45
4.8	Layout of the Pierre Auger Observatory	46
4.9	FD optics and camera	46
4.10	FD stereo and SD–FD calibration	48
4.11	AugerPrime station layout.	49
4.12	Assembly of a UMD module	51
4.13	Scintillation in the UMD modules	52
4.14	UMD opto-electronics	52
4.15	UMD acquisition modes	54
4.16	UMD deployment status	55
4.17	Binary trace example and infinite-window counting	57
5.1	Geometric layout of the SD-433 array	63
5.2	Mass fractions as a function of the primary energy from the Global Spline Fit	64
5.3	Binned estimates of ϵ for proton and iron	66
5.4	$\epsilon(\lg(E_{MC}))$ for different primaries and the GSF mix bias	67

5.5	$\varepsilon(\lg(E_{\text{MC}}), \theta)$ as a function of $\lg E_{\text{MC}}$ for different zenith bins and its bias . . .	68
5.6	$\varepsilon(\lg(E_{\text{MC}}), \theta, \ln(A) \approx 2.77)$ as a function of $\lg(E_{\text{MC}})$ and composition bias . .	69
5.7	MC attenuation functions $f_{\text{att}}(x)$ for different primaries and simulated GSF mix	71
5.8	Energy calibration for simulated GSF mix	71
5.9	ε as a function of $\lg E_{\text{SD-433}}$	73
5.10	Hypothesis tests for isotropic arrival of events in the SD-433 array	74
5.11	Normalised cumulative number of events	75
5.12	$b(E_{\text{SD-433}})$ for the GSF mix and composition bias	77
5.13	Stress test with hard break in composition	78
5.14	SD-433 fractional energy resolution as a function of $\lg(E_{\text{SD-433}}/\text{eV})$	80
6.1	Mean UMD-433 muon density profile with a Greisen fit	89
6.2	Mean UMD-433 profile vs Volcano Ranch and AGASA	89
6.3	Example MLDF fits	91
6.4	UMD-433 MLDF sample overview	92
6.5	Event-level optimum distance in the UMD-433	93
6.6	Shape-induced relative span on the muon-density estimator	94
6.7	Event-level UMD-433 MLDF slope values	95
6.8	Bias and resolution of the UMD-433 β parameterisation	96
6.9	MLDF residuals in the fixed- β production	97
6.10	Attenuation correction for the UMD-433 muon-density estimator	98
6.11	Detrended ρ_{21}	99
6.12	Compatibility checks between the UMD-433 and UMD-750 reconstructions .	100
6.13	Comparison of the ρ_{35} observable between UMD-433 and UMD-750	101
6.14	Simulation means of ρ_{21} and $\ln \rho_{21}$	102
6.15	Forward-folding correction for simulated ρ_{21} observables	103
6.16	Measured $\langle \rho_{21} \rangle$ compared to simulated baselines	104
6.17	Inferred $\langle \ln A \rangle$ compared to UMD-750 and X_{max}	106
6.18	Broader comparison of $\langle \ln A \rangle$ with previous composition estimates	107
6.19	z-scale comparison with other experiments	108
7.1	Core randomisation for SD-433 photon simulations	112
7.2	SD-433 photon trigger efficiency	114
7.3	Validation of the photon efficiency parametrisation	114
7.4	Photon simulation core throws on the SD-433 layout	116
7.5	Photon LDF residual profiles versus core distance	117
7.6	Event-by-event photon LDF slope trends	117
7.7	Raw photon size-energy relation in zenith bins	118
7.8	Attenuation-like correction used for the photon slope parametrisation	118
7.9	Validation of the photon slope parametrisation	120
7.10	Angular resolution σ_{α} of the photon reconstruction	121
7.11	Core resolution of the photon reconstruction	123
7.12	Core residuals in detector coordinates	123
7.13	Directional decomposition of the photon core bias	124
7.14	Photon energy calibration	125
7.15	Photon energy response in zenith bins	127
7.16	Photon energy-response	127
7.17	Hadronic response under the photon energy calibration	128
8.1	Overview of the multi-view GNN architecture	138
8.2	Learning curves of the three discrimination branches.	144

8.3	Distribution of discrimination scores from the GNN	145
8.4	Energy-binned distributions of the combined SD-UMD score	146
8.5	Exponential tail fit and signal-efficiency versus bkg.-contamination curves . .	148
8.6	Correlation between M_1 and the network scores	150
8.7	Pull comparison between the combined SD-UMD score and M_1	151
8.8	Effect of reduced UMD information on M_1	152
8.9	Robustness of the network score under reduced UMD information	153
8.10	Event-level explanation for a photon-like proton	157
8.11	Dataset-wide station-rank scalar attribution matrix	159
8.12	Dataset-wide scalar and trace attributions	160
9.1	Score distributions for the ideal burn-sample subset	165
9.2	Score distributions for the reduced-UMD burn-sample configuration	166
9.3	Energy dependence of the network scores in the burn sample	167
9.4	Zenith-angle dependence of the network scores in the burn sample	167
9.5	Time stability of the network scores in the burn sample	168
9.6	Correlation between network scores in simulations and burn-sample data . .	169
9.7	Selected SD-433 hexagons and cumulative photon exposure	172
9.8	Projected photon limits using the graph-network selection	173
A.1	Energy correction between E_{MC} and E_{SD-433}	204
A.2	ε as a function of E_{SD-433} and $\sin^2 \theta$	205
A.3	CIC angular dependence correction	205
A.4	Profile-likelihood estimate of the SD-433 fractional energy resolution	210
B.1	Muon overcounting example from a UMD module	217
B.2	Buried-pinocchio signature in the ρ_μ - S plane	218
B.3	Distance threshold for the pinocchio preselection	220
B.4	Impact of pinocchio rejection on $\rho(r_{ref})$	221
B.5	Weekly occurrence of the most evident buried pinocchios	222
B.6	Case study of a buried pinocchio in Auger ID 212851156000	223
B.7	Distributions of the module-level likelihood ratio $\lg(\Lambda)$	231
B.8	Clean inferred muon density ρ_μ versus SD signal S	232
B.9	Impact of rejecting incompatible module in event 213312731800	233
B.10	Extreme case of legacy fir in event 210711556400	233
B.11	UMD-433 additional shape validation	235
B.12	UMD-433 mean profile with Greisen fit	237
B.13	UMD-750 additional shape validation	237
B.14	Relative fluctuations of ρ_μ in the SD-433 phase space	238
B.15	Event-level SD-433 MLDF slope values	239
B.16	Bias and resolution of the SD-433 β parameterisation	240
B.17	CIC scan stability and attenuation-curve uncertainty	240
B.18	CIC validation plots	241
B.19	Systematic uncertainties on ρ_{21}	242
B.20	Analogous to Fig. 6.17 w. and w.o. folding correction	242
B.21	Analogous to Fig. 6.18 w. and w.o. folding correction	243
B.22	Analogous to Fig. 6.19 w. and w.o. folding correction	243

Supplementary material for Chapter 5

A.1 Generalized Linear Model for trigger efficiency

In this study, the independent variables of the set of predictors X are $\lg E$, $\sin^2 \theta$, and $\ln A$, or recentered variations of them. The efficiency, denoted by ε , is modelled by a Generalized Linear Model (GLM) describing the relationship between ε and a set of independent variables encoded in the design matrix $\mathbf{X}$. For n observations and p predictors, the design matrix $\mathbf{X}$ is defined as

$$\mathbf{X} = \begin{pmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}. \quad (\text{A.1})$$

The mean response $E[y|\mathbf{X}]$ is linked to a linear predictor $\mathbf{X}\beta$ through a link function. This means that the model fits a regression on a nonlinear transformation of the original dependent variable. It is worth mentioning the term *linear* in a *linear model* refers to the linearity in the model parameters, not in the variables. In this note, a probit link function Φ^{-1} was chosen to link ε with the design matrix $\mathbf{X}$ containing the predictors $\lg E$, $\sin^2 \theta$, and $\ln A$. In other words,

$$\Phi^{-1}(\varepsilon) = \mathbf{X}\beta. \quad (\text{A.2})$$

In this equation, β represents the vector of model parameters, $\mathbf{X}$ is the standard design matrix for a linear model, and Φ is the cumulative density function of a standardised normal distribution. This method is commonly referred to as probit regression. The parameters are estimated with an unbinned likelihood fit, with the deviance as the minimisation objective.

Other studies have employed an error function (erf) to model trigger efficiencies. This is essentially equivalent to using a probit function as a link function connecting the mean response of y to a function (not necessarily linear) of the predictors [198, 199]. This is the case due to

$$\text{probit}(p) = \Phi^{-1}(p) = \sqrt{2} \operatorname{erf}^{-1}(2p - 1). \quad (\text{A.3})$$

A.2 Energy Scale Transformation

To conduct the transformation between energy scales, the expression

$$E = A \left(\frac{S_{30}}{30 \text{ VEM}} \right)^B \quad (\text{A.4})$$

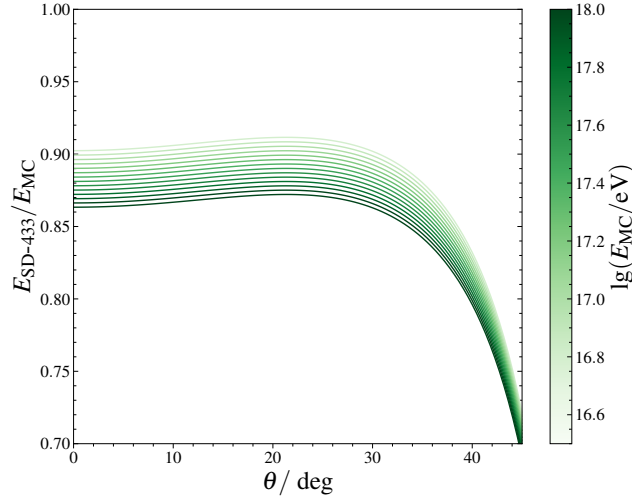


Figure A.1: Ratio between the SD-433 energy scale $E_{\text{SD-433}}$ and the Monte Carlo energy E_{MC} as a function of the zenith angle θ , for different E_{MC} values.

was used for the energy calibration. Meanwhile, the attenuation function f_{att} is described by

$$f_{\text{att}}(x) = 1 + [a_0 + a_1 x \lg(S_{30}/30 \text{ VEM})] + b x^2, \quad (\text{A.5})$$

where $x = \sin^2 \theta - \sin^2 30^\circ$. The coefficients employed are

	A	B	a_0	a_1	b
Data	117.0 ± 0.4	0.963 ± 0.003	-1.66 ± 0.06	0.19 ± 0.09	-1.41 ± 0.08
Monte Carlo	131.3 ± 0.4	0.978 ± 0.002	-1.87 ± 0.01	0.37 ± 0.01	-2.22 ± 0.03

The data parameters come from a new iteration of the values computed for Ref. [142], with small changes in the energy calibration. Monte-Carlo parameters were computed in the same fashion for the simulations described in this note.

A.3 Fitted model for $\varepsilon(E_{\text{SD-433}}, \theta)$

Following Eq. (5.8), the model parameters in the SD-433 energy scale are presented in Table A.1, and the function is displayed in Fig. A.2.

Table A.1: Parameters of $\varepsilon(E_{\text{SD-433}}, \theta)$ for the GSF.

a_0	a_1	a_2	a_3	a_4
5.440 ± 0.023	7.393 ± 0.032	-7.793 ± 0.160	-2.247 ± 0.259	-3.294 ± 0.215

A.4 CIC consistency check

In order to assess the consistency of the CIC procedure with the chosen energy threshold, the apparent shower size $S(300)$ and attenuation-corrected shower size S_{30} histograms are inspected. For energies above the suggested energy threshold of $10^{16.88}$ eV, the angular dependence seen for $S(300)$ is removed by the CIC procedure, as seen in the S_{30} histogram.

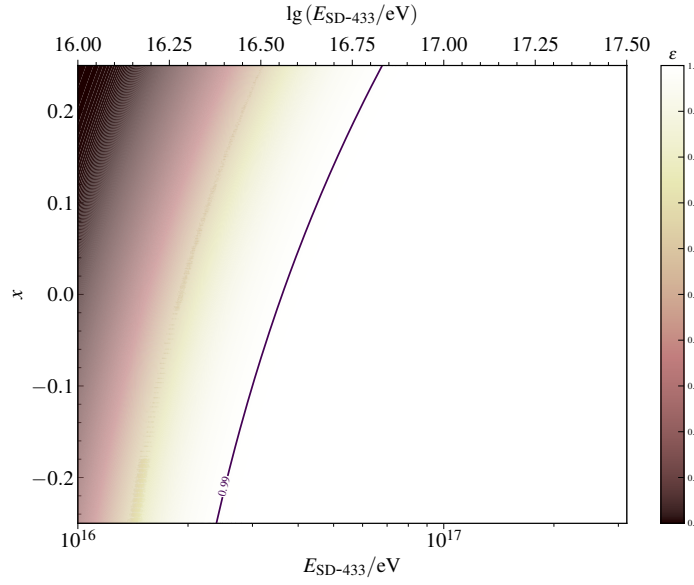


Figure A.2: ε as a function of $E_{\text{SD-433}}$ and $\sin^2 \theta$. The 0.99 boundary is depicted by a black solid line.

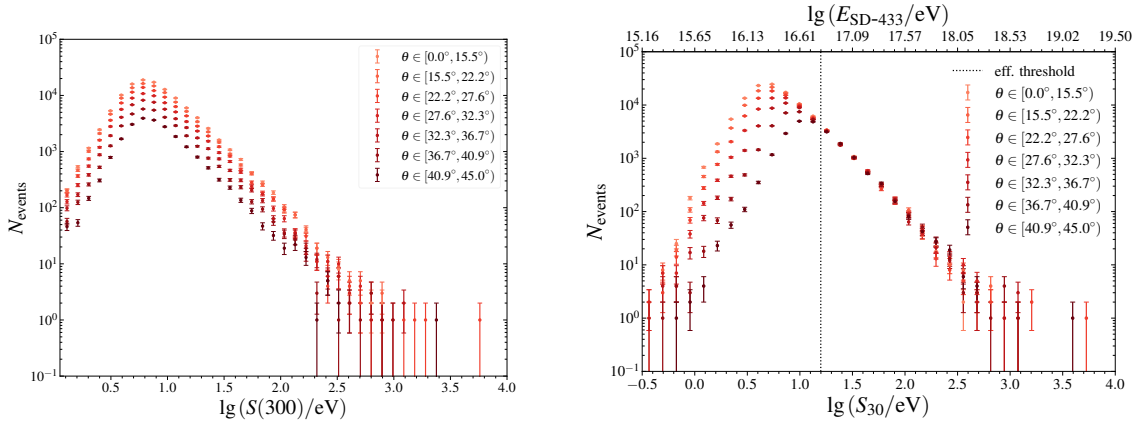


Figure A.3: *Left:* Apparent shower size histogram for independent zenith angle bins equally spaced in $\sin^2 \theta$. The darkness of the colour scale increases with the zenith angle. *Right:* Attenuation-corrected shower size histogram for independent zenith angle bins equally spaced in $\sin^2 \theta$. The darkness of the colour scale increases with the zenith angle.

A.5 Energy resolution from joint SD-433 and SD-750 events

This appendix describes a data-driven method to infer the SD-433 *fractional* energy resolution,

$$f_{\text{SD-433}}(E_{\text{true}}) \equiv \frac{\sigma(E_{\text{SD-433}} | E_{\text{true}})}{E_{\text{true}}}, \quad (\text{A.6})$$

from an event-by-event comparison between the SD-433 and SD-750 reconstructed energies in a common (*joint*) event sample.

For each selected joint event, two reconstructed energy estimators are available:

- $E_{\text{SD-433}}$: energy reconstructed with the SD-433 array;

- $E_{\text{SD-750}}$: energy reconstructed with the SD-750 array.

Throughout this appendix, $E_{\text{SD-433}}$ and $E_{\text{SD-750}}$ always denote *reconstructed* energies. This differs from parts of the main text where $E_{\text{SD-433}}$ may be used as a proxy for the true energy. The present notation keeps both estimators on equal footing.

Since E_{true} is unknown in data, the comparison is performed in bins of a reconstructed-energy proxy. Events are binned in the geometric mean

$$E_{\text{bin}} \equiv \sqrt{E_{\text{SD-433}} E_{\text{SD-750}}}. \quad (\text{A.7})$$

In the implementation, bins are defined in $\lg(E_{\text{bin}}/\text{eV})$ with fixed width $\Delta \lg E$.

The geometric mean is preferred over binning in either $E_{\text{SD-433}}$ or $E_{\text{SD-750}}$ alone because it is symmetric and reduces binning biases driven by fluctuations of a single estimator. At fixed E_{true} , both reconstructed energies fluctuate: if one bins using (say) $E_{\text{SD-433}}$, upward fluctuations of that estimator are more likely to migrate events into a higher reconstructed bin, while downward fluctuations migrate them out, thereby correlating the bin assignment with the very fluctuation one aims to measure. Using the geometric mean mitigates this effect because, to first order in fractional errors,

$$E_{\text{SD-433}} = E_{\text{true}}(1 + \varepsilon_{433}), \quad E_{\text{SD-750}} = E_{\text{true}}(1 + \varepsilon_{750}) \Rightarrow E_{\text{bin}} \simeq E_{\text{true}} \left(1 + \frac{\varepsilon_{433} + \varepsilon_{750}}{2} \right),$$

so the binning variable depends on the average fluctuation of the two reconstructions rather than on either fluctuation alone. Since the comparison observable introduced below depends primarily on the *difference* of fluctuations, binning in E_{bin} reduces the artificial coupling between bin membership and the difference mode that controls the inferred width.

To compare the two estimators without introducing an explicit dependence on the unknown true energy, we use the symmetric fractional difference

$$d \equiv \frac{2(E_{\text{SD-433}} - E_{\text{SD-750}})}{E_{\text{SD-433}} + E_{\text{SD-750}}}. \quad (\text{A.8})$$

This quantity is dimensionless, approximately equals the relative difference when $E_{\text{SD-433}} \simeq E_{\text{SD-750}}$, and is insensitive to an overall common rescaling of both energies.

To interpret the distribution of d within a narrow reconstructed-energy bin, both reconstructions are modelled as unbiased estimators with Gaussian *fractional* fluctuations at fixed truth,

$$E_i = E_{\text{true}}(1 + \varepsilon_i), \quad i \in \{\text{SD-433}, \text{SD-750}\}, \quad (\text{A.9})$$

where the fractional deviations are assumed bivariate normal,

$$\begin{pmatrix} \varepsilon_{\text{SD-433}} \\ \varepsilon_{\text{SD-750}} \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} f_{\text{SD-433}}^2 & \rho f_{\text{SD-433}} f_{\text{SD-750}} \\ \rho f_{\text{SD-433}} f_{\text{SD-750}} & f_{\text{SD-750}}^2 \end{pmatrix} \right), \quad (\text{A.10})$$

with RMS fractional resolutions $f_{\text{SD-433}}$ and $f_{\text{SD-750}}$ (at fixed truth) and correlation coefficient ρ . A parametrisation for $f_{\text{SD-750}}$ taken from Ref. [139] is employed here; thus, only $f_{\text{SD-433}}$ and ρ are unknown (but correlated).

Inserting Eq. (A.9) into Eq. (A.8) and expanding to first order in the fractional errors yields the core approximation

$$d \simeq \varepsilon_{\text{SD-433}} - \varepsilon_{\text{SD-750}}. \quad (\text{A.11})$$

Therefore, at fixed E_{true} the variable d is approximately Gaussian with variance

$$V_d(f_{\text{SD-433}}, f_{\text{SD-750}}, \rho) = f_{\text{SD-433}}^2 + f_{\text{SD-750}}^2 - 2\rho f_{\text{SD-433}} f_{\text{SD-750}}. \quad (\text{A.12})$$

If one could bin in E_{true} , the inference would reduce to measuring the Gaussian width of d in each truth bin and solving Eq. (A.12) for f_{433} . In data, bins are defined in E_{bin} rather than E_{true} , so it is necessary to construct a likelihood in each E_{bin} bin to extract $f_{\text{SD-433}}$, while treating the mean of d and the correlation ρ as nuisance parameters.

A forward-folding expression for the expected density of the selected joint sample in terms of the *reconstructed* energies is [140]

$$\frac{d^2 N}{dE_{\text{SD-433}} dE_{\text{SD-750}}} = \int dE_{\text{true}} \int d\Omega J(E_{\text{true}}) \mathcal{E}_{\text{joint}}(E_{\text{true}}, \theta) k(E_{\text{SD-433}}, E_{\text{SD-750}} | E_{\text{true}}, \theta), \quad (\text{A.13})$$

where $J(E_{\text{true}})$ is the differential cosmic-ray flux, $\mathcal{E}_{\text{joint}}(E_{\text{true}}, \theta)$ is the effective exposure of the joint sample (including acceptance, live time, and the joint-event requirement and analysis cuts), and $k(E_{\text{SD-433}}, E_{\text{SD-750}} | E_{\text{true}}, \theta)$ is the joint detector response kernel implied by the bivariate fractional-error model in Eq. (A.10).

A key practical consequence is that conditioning on a *reconstructed-energy* bin does not select a delta-function in E_{true} . Instead, each bin receives contributions from a range of true energies weighted by the effective prior $J(E_{\text{true}}) \mathcal{E}_{\text{joint}}(E_{\text{true}}, \theta)$. This leads to bin-migration effects whose detailed impact depends on the spectrum, exposure, and selection of the joint sample.

As a result, the comparison observable d can exhibit a non-zero conditional mean in a bin of E_{bin} even if both estimators are unbiased at fixed truth, i.e. $\mathbb{E}[\varepsilon_i | E_{\text{true}}] = 0$. To first order, $d \simeq \varepsilon_{\text{SD-433}} - \varepsilon_{\text{SD-750}}$, while the binning variable depends on the *sum* mode through $E_{\text{bin}} = \sqrt{E_{\text{SD-433}} E_{\text{SD-750}}} \simeq E_{\text{true}} (1 + \frac{\varepsilon_{\text{SD-433}} + \varepsilon_{\text{SD-750}}}{2})$. When $f_{\text{SD-433}} \neq f_{\text{SD-750}}$, the sum and difference modes are correlated under the bivariate Gaussian model, and selecting events by reconstructed energy can therefore bias the mean of d . For this reason, the present method does not impose any predicted mean shift. Instead, the mean of d is treated as a nuisance parameter and profiled analytically in each reconstructed-energy bin, so that the inference is driven primarily by the (core) width of the d distribution when constraining $f_{\text{SD-433}}$.

Consider one bin in $\lg(E_{\text{bin}}/\text{eV})$. After applying the event selection, the bin contains a sample $\{d_k\}_{k=1}^N$. The d distribution is modelled as Gaussian with variance $V_d(f_{\text{SD-433}}, f_{\text{SD-750}}, \rho)$ from Eq. (A.12) and an *unknown* mean μ_d ,

$$p(d | f_{\text{SD-433}}, \rho, \mu_d) = \mathcal{N}(d; \mu_d, V_d(f_{\text{SD-433}}, f_{\text{SD-750}}, \rho)). \quad (\text{A.14})$$

The corresponding negative log-likelihood (NLL) for the bin sample is

$$\text{NLL}(f_{\text{SD-433}}, \rho, \mu_d) = \frac{N}{2} \ln(2\pi V_d) + \frac{1}{2V_d} \sum_{k=1}^N (d_k - \mu_d)^2, \quad (\text{A.15})$$

where $V_d \equiv V_d(f_{\text{SD-433}}, f_{\text{SD-750}}, \rho)$.

For fixed $(f_{\text{SD-433}}, \rho)$, the maximum-likelihood estimator of the mean is the sample mean,

$$\hat{\mu}_d = \frac{1}{N} \sum_{k=1}^N d_k, \quad (\text{A.16})$$

so profiling over μ_d yields

$$\text{NLL}(f_{\text{SD-433}}, \rho) \equiv \min_{\mu_d} \text{NLL}(f_{\text{SD-433}}, \rho, \mu_d) = \frac{N}{2} \ln(2\pi V_d) + \frac{1}{2V_d} \sum_{k=1}^N (d_k - \hat{\mu}_d)^2. \quad (\text{A.17})$$

After profiling, the fit is driven by the centered second moment of the observed d distribution and by the model variance V_d .

The correlation coefficient ρ enters through Eq. (A.12) but is not the quantity of interest. It is therefore treated as a nuisance parameter and profiled in each energy bin. Let $\hat{\sigma}_d$ be the empirical standard deviation of the d sample in the bin. For a fixed input $f_{\text{SD-750}}$, the minimum variance compatible with a given ρ is

$$V_{d,\min}(\rho) = (1 - \rho^2) f_{\text{SD-750}}^2, \quad (\text{A.18})$$

which implies the feasibility condition

$$\hat{\sigma}_d^2 \geq (1 - \rho^2) f_{\text{SD-750}}^2 \implies \rho \geq \rho_{\min} \equiv \sqrt{1 - \left(\frac{\hat{\sigma}_d}{f_{\text{SD-750}}} \right)^2}. \quad (\text{A.19})$$

The argument of the square root is clipped to $[0, 1]$. In practice, ρ is scanned on a grid in $[\max(\rho_{\min}, \rho_{\text{floor}}), \rho_{\max}]$, with $\rho_{\max} < 1$ for numerical stability.

Equation (A.12) shows that, in a given energy bin, the Gaussian model constrains the data primarily through the variance of d . Since V_d depends on the two unknowns $(f_{\text{SD-433}}, \rho)$, these parameters are strongly correlated: for a fixed observed width $\hat{\sigma}_d$, many combinations of $(f_{\text{SD-433}}, \rho)$ provide the same V_d . Solving Eq. (A.12) for $f_{\text{SD-433}}$ illustrates this degeneracy,

$$f_{\text{SD-433}} = \rho f_{\text{SD-750}} \pm \sqrt{\hat{\sigma}_d^2 - (1 - \rho^2) f_{\text{SD-750}}^2},$$

which also motivates the feasibility bound in Eq. (A.19). For this reason, ρ is treated as a nuisance parameter and profiled in each bin.

For each trial value of $f_{\text{SD-433}}$, the profiled likelihood is obtained by minimising Eq. (A.17) over ρ ,

$$\text{NLL}_{\text{prof}}(f_{\text{SD-433}}) \equiv \min_{\rho \in [\rho_{\min}, \rho_{\max}]} \text{NLL}(f_{\text{SD-433}}, \rho). \quad (\text{A.20})$$

The best-fit estimate is

$$\hat{f}_{\text{SD-433}}^{\text{prof}} = \arg \min_{f_{\text{SD-433}}} \text{NLL}_{\text{prof}}(f_{\text{SD-433}}), \quad (\text{A.21})$$

and a confidence envelope in $f_{\text{SD-433}}$ is defined through a likelihood-ratio criterion,

$$\text{NLL}_{\text{prof}}(f_{\text{SD-433}}) - \text{NLL}_{\text{prof}}(\hat{f}_{\text{SD-433}}^{\text{prof}}) \leq \Delta \text{NLL}_{\text{thr}}. \quad (\text{A.22})$$

When Wilks' theorem applies for one parameter, $2\Delta \text{NLL}$ is approximately distributed as χ_1^2 , implying $\Delta \text{NLL}_{\text{thr}} = 0.5, 2.0, 4.5$ for approximate $1\sigma, 2\sigma, 3\sigma$ intervals.

A.5.1 Results

Table A.2 summarises the profile-likelihood results obtained in each $\lg(E_{\text{bin}}/\text{eV})$ bin, where $E_{\text{bin}} = \sqrt{E_{\text{SD-433}} E_{\text{SD-750}}}$. The fitted mean $\hat{\mu}_d$ is small in all bins (typically $|\hat{\mu}_d| \lesssim 0.02$), indicating that the two reconstructions are, on average, consistent within each reconstructed-energy interval.

The inferred SD-433 fractional resolution in the considered range $17.35 \leq \lg(E_{\text{bin}}/\text{eV}) \leq 17.95$ is of order 10–25%, with a mild, non-monotonic energy dependence. The broad shaded envelope in Fig. A.4 reflects the fact that the width of the comparison observable d constrains mainly the combination $V_d = f_{\text{SD-433}}^2 + f_{\text{SD-750}}^2 - 2\rho f_{\text{SD-433}} f_{\text{SD-750}}$: there is a strong degeneracy between $f_{\text{SD-433}}$ and the unknown correlation ρ , so different pairs $(f_{\text{SD-433}}, \rho)$ can reproduce a similar observed $\hat{\sigma}_d$. This is also visible in the large best-fit correlations, $\hat{\rho} \simeq 0.78\text{--}0.97$, and in the fact that the likelihood minimum often lies close to the feasibility bound $\rho_{\min}$ (Eq. A.19). In particular, in all bins $\hat{\sigma}_d < f_{\text{SD-750}}$, so a substantial positive correlation is required for consistency.

Table A.2: Profile-likelihood results per $\lg(E_{\text{bin}}/\text{eV})$ bin. N is the number of events in the bin, $f_{\text{SD-750}}$ is the external SD-750 fractional-resolution input, and $\hat{\mu}_d$ and $\hat{\sigma}_d$ are the empirical mean and standard deviation of d in the selected sample. The SD-433 resolution is reported as the profiled best-fit $\hat{f}_{\text{SD-433}}^{\text{prof}}$ with its profile-likelihood band $[f_{\text{lo}}, f_{\text{hi}}]$ obtained using the threshold of Eq. A.22 ($\Delta\text{NLL}_{\text{thr}} = 4.5$, corresponding approximately to a 3σ interval for one parameter). $\hat{\rho}$ is the correlation that minimises the NLL at fixed $f_{\text{SD-433}}$, and ρ_{min} is the feasibility bound from Eq. A.19.

$\lg(E_{\text{bin}}/\text{eV})$	N	$f_{\text{SD-750}}$	$\hat{\mu}_d$	$\hat{\sigma}_d$	$\hat{f}_{\text{SD-433}}^{\text{prof}}$	f_{lo}	f_{hi}	$\hat{\rho}$	ρ_{min}
17.35	802	0.202	-0.0045	0.128	0.140	0.065	0.335	0.777	0.772
17.45	460	0.190	-0.0006	0.107	0.117	0.074	0.303	0.864	0.827
17.55	306	0.180	0.0031	0.092	0.152	0.077	0.279	0.860	0.859
17.65	178	0.171	0.0111	0.086	0.240	0.071	0.267	0.969	0.865
17.75	94	0.163	0.0160	0.085	0.204	0.057	0.265	0.918	0.853
17.85	47	0.155	0.0112	0.065	0.099	0.066	0.241	0.969	0.909
17.95	27	0.149	-0.0112	0.057	0.150	0.061	0.234	0.931	0.925

At the highest energies the number of joint events decreases rapidly, and the profile bands remain wide; therefore, the apparent bin-to-bin fluctuations of $\hat{f}_{\text{SD-433}}^{\text{prof}}$ at the upper end of the range should not be over-interpreted as a physical feature, but rather as the combined effect of limited statistics and the intrinsic $(f_{\text{SD-433}}, \rho)$ degeneracy.

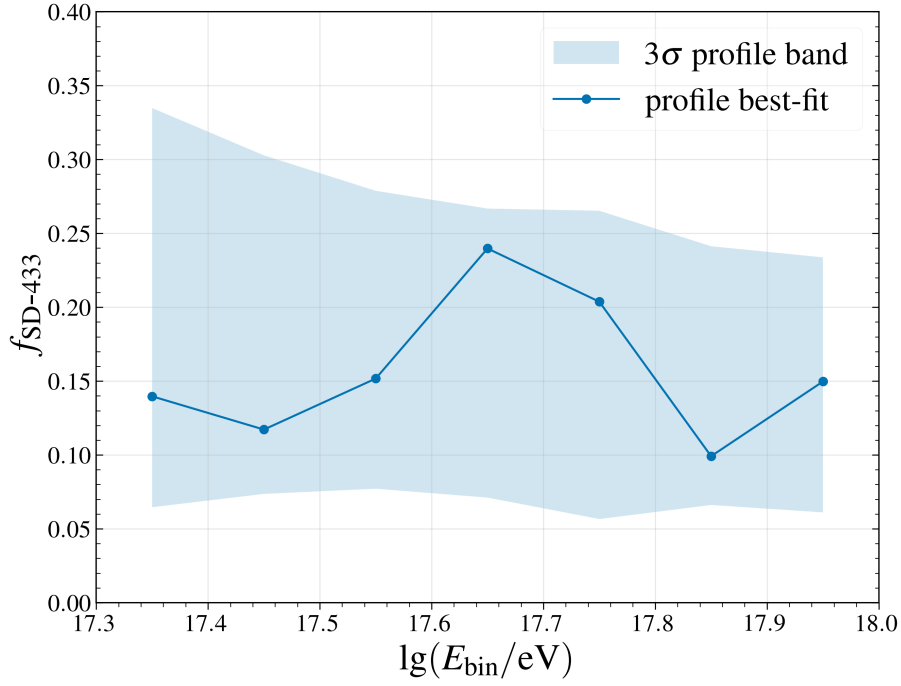


Figure A.4: Profile-likelihood estimate of the SD-433 fractional energy resolution as a function of the reconstructed-energy proxy $E_{\text{bin}} = \sqrt{E_{\text{SD-433}} E_{\text{SD-750}}}$. In each $\lg(E_{\text{bin}}/\text{eV})$ bin, the comparison observable $d = 2(E_{\text{SD-433}} - E_{\text{SD-750}})/(E_{\text{SD-433}} + E_{\text{SD-750}})$ is modelled by a Gaussian core with variance $V_d = f_{\text{SD-433}}^2 + f_{\text{SD-750}}^2 - 2\rho f_{\text{SD-433}} f_{\text{SD-750}}$, using an external parametrisation for $f_{\text{SD-750}}$ and profiling over the nuisance parameters μ_d (analytically) and ρ (numerically). Markers show the profiled best-fit $\hat{f}_{\text{SD-433}}^{\text{prof}}$, while the shaded band indicates the profile-likelihood envelope $[f_{\text{lo}}, f_{\text{hi}}]$ obtained from the criterion $\Delta\text{NLL} \leq \Delta\text{NLL}_{\text{thr}}$ with $\Delta\text{NLL}_{\text{thr}} = 4.5$ (approximately a 3σ interval for one parameter).

Supplementary material for Chapter 6

B.1 Data quality and updates of the reconstruction sequence

The mass-sensitive muon observable introduced in the chapter relies on a stable and well-understood response of the UMD over long time periods. In practice, this requires controlling both (i) *deterministic* sources of invalid measurements, such as known hardware or acquisition problems affecting specific modules or counters, and (ii) more subtle *reconstruction-level* pathologies that can bias the fitted MLDF even when standard quality flags are satisfied. This section summarises the data-quality strategy and the corresponding updates to the UMD reconstruction sequence implemented in the Offline framework in the context of this thesis, largely following the internal documentation developed in parallel within the Collaboration.

B.1.1 Module Rejection

The UMD reconstruction sequence interleaves dedicated quality-control steps with the actual reconstruction of the muon content at module, counter, and event level. In this context, “module rejection” denotes any decision that prevents a module (or its parent counter) from contributing to subsequent reconstruction steps, in particular to the MLDF fit. The logic is distributed over three places in the standard sequence: the CDAS-to-Offline conversion, the `MdModuleRejector` module, and the counter/event-level cuts implemented in `MdEventSelector`. In the context of this thesis, the `MdModuleRejector` module was rewritten from scratch, and the `MdEventSelector` module was updated significantly.

Operationally, each module carries a reconstruction status flag that distinguishes candidates from rejected modules, together with a more detailed rejection status that encodes the specific reason(s) for exclusion. This follows the methodology already adopted for the SD and RD reconstructions, adapted to the particularities of the UMD. Rejected modules are ignored by all downstream modules (`MdPatternFinder`, `MdPileUpCorrector`, `MdCornerClippingCorrector`, and `MdLDFFinder`), so that spurious or poorly defined muon-count estimates do not bias the reconstruction of the reference muon density.

B.1.1.1 Early module initialisation and hard rejections

The first set of module-level decisions is taken during the event conversion step, when CDAS events are translated into the internal Offline data model. At this stage, the converter iterates over all counters associated to triggered SD stations, creates the corresponding UMD modules, and assigns them an initial reconstruction status. Modules can be rejected already

here if severe on-ground or underground errors are detected in the trigger and data-request chain, or if inconsistencies between SD and UMD information indicate that a reliable module trace is not available.

In parallel, the front-end configuration of each module is recovered and used to determine the set of active channels via a bit mask. This information is employed by a channel-masking logic, and forms the basis for subsequent decisions on whether a module can deliver a usable muon measurement in a given event.

B.1.1.2 MdModuleRejector: static and time-dependent masks

The first explicit quality module in the UMD reconstruction sequence is MdModuleRejector. Its task is to apply all static and time-dependent module-level masks before any muon counting is attempted. In the current implementation, the module was rewritten from scratch to allow for a more transparent and flexible configuration.

The rejection logic is organised along three main axes:

- **Photodetector-based rejection.** Individual modules can be masked according to the status of their photodetectors (PMTs or SiPMs). This covers, for example, modules operated with prototype front-end boards, modules under test, or modules whose readout has been declared unreliable by hardware experts.
- **Geometrical and configuration-based rejection.** The module area and segmentation enter directly in the muon-density estimation and in the effective dynamic range of the binary counters. Modules that do not conform to the nominal 10 m² design (e.g. 5 m² modules) can therefore be rejected at this stage, unless a dedicated treatment is explicitly enabled in the configuration.
- **Time-dependent bad-period masks.** A central feature of the new implementation is the use of pre-computed lists of unstable acquisition periods for each module [193]. The resulting time intervals are stored in external configuration files and read by MdModuleRejector, which rejects any module whose event time falls into one of these flagged periods.

In practice, the configuration distinguishes between (i) custom rejections that are independent of time (e.g. prototypes, modules used only in dedicated studies), (ii) custom time-period masks defined by manual inspection, and (iii) automated unstable-period masks generated by the method described in Ref. [193]. This separation allows the lists to be maintained by different subgroups without requiring changes to the reconstruction code.

B.1.1.3 MdEventSelector: counter- and event-level quality cuts

After the module-level reconstruction steps and the corner-clipping correction, MdEventSelector applies the final quality cuts needed before the MLDF fit. This module operates both at counter and at event level and is highly configurable via XML, so that different analyses can enable or disable specific cuts without recompiling the reconstruction application.

At event level, the module can reject:

- saturated events, where at least one counter is saturated,
- events above a configurable maximum zenith angle, and
- events with an insufficient number of usable counters after all previous rejections.

At counter level, additional quality cuts decide whether the muon-density estimate of a given counter is allowed to enter the MLDF fit. The most relevant options are:

- **Time-quality cuts**, rejecting counters whose timing is inconsistent with the shower front.
- **Grid-type selections**, restricting the reconstruction to the desired sub-array (e.g. UMD-433 or UMD-750) in mixed configurations.
- **SD-linked rejection**, where counters can be discarded depending on the reconstruction status of their partner SD station. Since SD and UMD statuses are largely independent, only a subset of SD rejection reasons, those associated with erroneous, unphysical or clearly dubious triggers (e.g. lightning, random triggers, lonely stations, missing T3 triggers, out-of-time stations), are used to veto UMD counters.
- **Muon-content selections**, e.g. rejecting counters with extremely low reconstructed muon signal that are incompatible with the expectations from the shower geometry and lateral distance.

In addition, `MdEventSelector` includes specialised cuts that are only relevant for simulations. For instance, modules located within the CORSIKA MC radius, where only one out of every thousand particles is sampled to control file size, can be rejected to avoid inconsistencies in the effective muon density. Likewise, counters belonging to dense-array configurations that are not part of the nominal setup can be removed. These options ensure that simulated datasets are processed with the same reconstruction applications as real data, while avoiding configuration-dependent biases.

It is worth noting that several of the quality-based rejections implemented in the reconstruction sequence, in particular the rejection of modules during unstable acquisition periods, would be more appropriately applied at the stage where individual module data files from *md files* are merged into event-level files to produce the final *xad files*. Performing these checks earlier would reduce processing overhead. However, due to the time constraints associated with the preparation for ICRC-2025, the most expedient solution was to incorporate these rejections directly into the reconstruction chain. A future update should transfer the corresponding logic to the merging code, while preserving the detailed rejection bookkeeping already implemented here.

B.1.1.4 Technical details about module rejection

Module rejection is implemented as a bookkeeping layer that propagates along the full reconstruction chain: each module carries (i) a coarse reconstruction status flag, and (ii) a more detailed rejection status encoding the reason(s) for exclusion. This information is produced by the three layers already discussed (`CDAS-to-Offline` conversion, `MdModuleRejector`, and `MdEventSelector`) and ensures that rejected modules are consistently ignored by all downstream modules, so that pathological measurements cannot bias the Muon LDF fit. The `ReconstructionStatus` enum provides a minimal state machine: `eCandidate` modules are eligible for muon reconstruction, and `eRejected` modules are excluded from all subsequent reconstruction steps. This flag is intentionally simple and is used for fast branching in the reconstruction modules. The `RejectionStatus` enum is a *bit mask*: each rejection reason corresponds to one bit (`ePMT = (1<0)`, `eSiPM = (1<1)`, ...). This design allows a single module to accumulate multiple independent veto reasons without losing information. Operationally, modules can be rejected for a combination of causes spanning hardware status, geometry/-configuration, SD-MD inconsistencies, simulation-only selections, or time-dependent bad periods.

When multiple reasons apply, the stored mask is the bitwise OR of the corresponding flags, e.g.

$$R = \text{eT1NotFound} \mid \text{eSmallModule}.$$

This is particularly useful for post-mortem studies: even if a module is vetoed early, the bookkeeping can still retain additional “context” flags when these are known at the same stage.

For human-readable diagnostics and configuration I/O, the bit mask is complemented by a bidirectional mapping between `RejectionStatus` and stable string labels ("UnstablePeriod", "PrototypeSiPM", ...). The mapping is centralised in `RejectionStatusMapper`: it exposes a `status→string` map and its inverse `string→status`. Unknown labels are treated as configuration errors and are logged explicitly, preventing silent mis-configurations. Because the set of rejection reasons is part of the public reconstruction bookkeeping, whenever a new `RejectionStatus` entry is added, *both* the enum and the mapping must be updated (and, if the label is used in XML configuration, the corresponding `.xsd` must be extended as well). This keeps ADST-level outputs, XML configuration, and monitoring plots mutually consistent across releases. All reason codes and the statistics of module rejection from Phase I data are presented in the tables below.

Table B.1 lists the rejection bits used by the current reconstruction. This dictionary must remain synchronised with (a) the internal enumeration in the code, (b) the label-mapper utility, and (c) the XML schema definitions, to avoid silent mis-configuration.

Table B.1: Rejection status bit mask: enum value, integer encoding, and string label.

Bit	Enum	Value	Label
–	<code>eNoRejection</code>	0	NoRejection
0	<code>ePMT</code>	2^0	PMT
1	<code>eSiPM</code>	2^1	SiPM
2	<code>eT1LTSPairMissing</code>	2^2	MissingT1-LST-pair
3	<code>eT1NotFound</code>	2^3	MissingT1
4	<code>eSdMdInconsistency</code>	2^4	SD-MDInconsistency
5	<code>eMCNonTriggeredWCD</code>	2^5	Sim-NonTriggeredWCD
6	<code>eBadData</code>	2^6	BadData
7	<code>eBlackUMD</code>	2^7	BlackUMD
8	<code>eAMIGAGrande</code>	2^8	AMIGAGrande
9	<code>eShiftedLatchBin</code>	2^9	ShiftedLatchBin
10	<code>eFaulty</code>	2^{10}	Faulty
11	<code>eSmallModule</code>	2^{11}	SmallModule
12	<code>ePrototypeSiPM</code>	2^{12}	PrototypeSiPM
13	<code>eHighSdLowMd</code>	2^{13}	HighSdLowMd
14	<code>eOffGrid</code>	2^{14}	OffGrid
15	<code>eFarSilent</code>	2^{15}	FarSilent
16	<code>eInsideMCMinRadius</code>	2^{16}	InsideMCMinRadius
17	<code>eDenseArray</code>	2^{17}	Dense
18	<code>eTimeOutlier</code>	2^{18}	TimeOutlier
19	<code>eBinaryAndADCDiscrepancy</code>	2^{19}	Binary-ADC-Discrepancy
20	<code>eSDRejectionStatus</code>	2^{20}	SDRejectionStatus
21	<code>eUnstablePeriod</code>	2^{21}	UnstablePeriod
22	<code>eDuplicatedId</code>	2^{22}	DuplicatedId
23	<code>eIncompatibleZeroCount</code>	2^{23}	IncompatibleZeroCount

Table B.2: Statistics of rejected UMD modules in Phase I (SD-433 and SD-750 combined). Repeated modules are counted only once. Total $N = 1\,889\,981$.

#	Rejection reason(s)	N	(%)
1	SmallModule	1 077 110	56.991
2	PMT SmallModule	240 239	12.711
3	OffGrid	172 374	9.120
4	PMT	127 666	6.755
5	AMIGAGrande	74 085	3.920
6	SDRejectionStatus	38 129	2.018
7	PMT AMIGAGrande	33 084	1.750
8	PMT MissingT1	30 660	1.623
9	MissingT1 UnstablePeriod	24 361	1.289
10	PMT MissingT1 SmallModule	15 997	0.846
11	SmallModule OffGrid	13 059	0.691
12	IncompatibleZeroCount	8 797	0.465
13	UnstablePeriod	7 174	0.380
14	SmallModule UnstablePeriod	6 709	0.355
15	SmallModule SDRejectionStatus	6 635	0.351
16	DuplicatedId	5 707	0.302
17	PMT SmallModule SDRejectionStatus	1 834	0.097
18	PMT SDRejectionStatus	1 405	0.074
19	PMT MissingT1 SDRejectionStatus	925	0.049
20	MissingT1	913	0.048
21	AMIGAGrande SDRejectionStatus	891	0.047
22	SDRejectionStatus DuplicatedId	579	0.031
23	BadData	376	0.020
24	PMT AMIGAGrande SDRejectionStatus	236	0.012
25	MissingT1 SmallModule	218	0.012
26	OffGrid UnstablePeriod	199	0.011
27	MissingT1 SDRejectionStatus UnstablePeriod	159	0.008
28	SDRejectionStatus UnstablePeriod	135	0.007
29	PMT MissingT1 SmallModule SDRejectionStatus	108	0.006
30	SmallModule SDRejectionStatus UnstablePeriod	71	0.004
31	MissingT1 OffGrid	50	0.003
32	BadData UnstablePeriod	37	0.002
33	MissingT1 AMIGAGrande	29	0.002
34	PMT MissingT1 AMIGAGrande	15	<0.001
35	MissingT1 SDRejectionStatus	8	<0.001
36	MissingT1 SmallModule UnstablePeriod	2	<0.001
37	BadData SDRejectionStatus	2	<0.001
38	MissingT1 SmallModule SDRejectionStatus	2	<0.001
39	MissingT1 SmallModule OffGrid	1	<0.001

B.1.2 Data cleaning

A standardised set of UMD data-quality observables began to be used operationally with the start of official UMD shifts in 2024 [200–202]. To achieve a homogeneous treatment of the full data set, these shift observables were reproduced for Phase I data. The motivation is twofold. First, this reproduction provides a direct cross-check that the reconstruction inputs are consistent with the detector behaviour expected from operations and monitoring.

Second, it enables the identification of additional periods of degraded data taking that may not be captured by purely rate-based selections. In this workflow, the unstable periods described in Ref. [193] are treated as an *a priori* exclusion and are already subtracted from the analysis inputs; the monitoring exercise therefore focuses on residual instabilities beyond that baseline cleaning.

The monitoring reproduction was performed at module and counter level and targeted generic signatures of abnormal behaviour, such as changes in acquisition stability, atypical distributions of reconstructed signals, and inconsistencies in the expected time evolution of detector-level observables. Periods flagged by these criteria were handled conservatively: when a clear instrumental origin could be established, the corresponding intervals were incorporated into time-dependent masks (applied through the same machinery as in section B.1.1.2); otherwise, the behaviour was tracked but not used to define additional deterministic cuts. A detailed account of this monitoring exercise is beyond the scope of this thesis. Nevertheless, the problematic measurements found that are relevant for the muon-content reconstruction can be broadly grouped into two categories, and the main strategies adopted to mitigate them are described below.

The first category is *overcounting*, where a module reports an abnormally large number of fired segments; its treatment is summarised briefly in section B.1.2.1. The second category is *undercounting*, where module traces yield near-zero muon counts even in regimes where a non-zero muon density is expected. The method developed to detect and remove these cases, together with its integration into the reconstruction via iterative MLDF fits, is described in section B.1.2.2. For both classes of anomalies, the data-quality measures available at the time were not sufficient to suppress their impact on the reconstruction, motivating the dedicated treatments presented below.

B.1.2.1 Overcounting

The first class of pathological measurements for the UMD reconstruction is *overcounting*, in which the counting strategy yields an unphysically large number of muons for a given module/counter. Several mechanisms can lead to overcounting, ranging from instrumental instabilities to event-contamination effects. In the Phase I studies relevant for this thesis, most overcounting cases at the *module* level were ultimately traced back to instabilities in the measuring hardware, while a smaller (but still important) fraction of outliers with respect to the MLDF were caused by timing mismatches.

The UMD provides, in addition to the event data used for reconstruction, an alternative data stream sent to CDAS containing monitoring observables from the front-end electronics. These quantities are essential to diagnose instrumental overcounting. A representative failure mode is a malfunction of the SiPM temperature monitoring: because the bias voltage is regulated to compensate temperature variations, an incorrect temperature readout can trigger an immediate and artificial change in the applied voltage. Since the discriminator threshold is defined in voltage units, this modifies the effective threshold seen by the binary acquisition and can directly alter the fraction of “1”s in the trace, biasing the reconstructed muon count upward (see Fig. B.1). In practice, such instabilities are treated through the data-quality machinery described earlier: once identified, the affected modules and/or time intervals are incorporated into time-dependent masks.

A distinct source of overcounting arises at the *counter* level as an outlier w.r.t. the MLDF, even when the UMD electronics behave nominally. In these cases, the underlying mechanism is typically a timing mismatch: an SD station that is out of time with respect to the fitted shower front can be associated to the event due to a low-energy shower elsewhere in the array (outside the energy range of interest and effectively acting as noise), and the corresponding UMD counter may record muons that sporadically contaminate the reconstructed event.

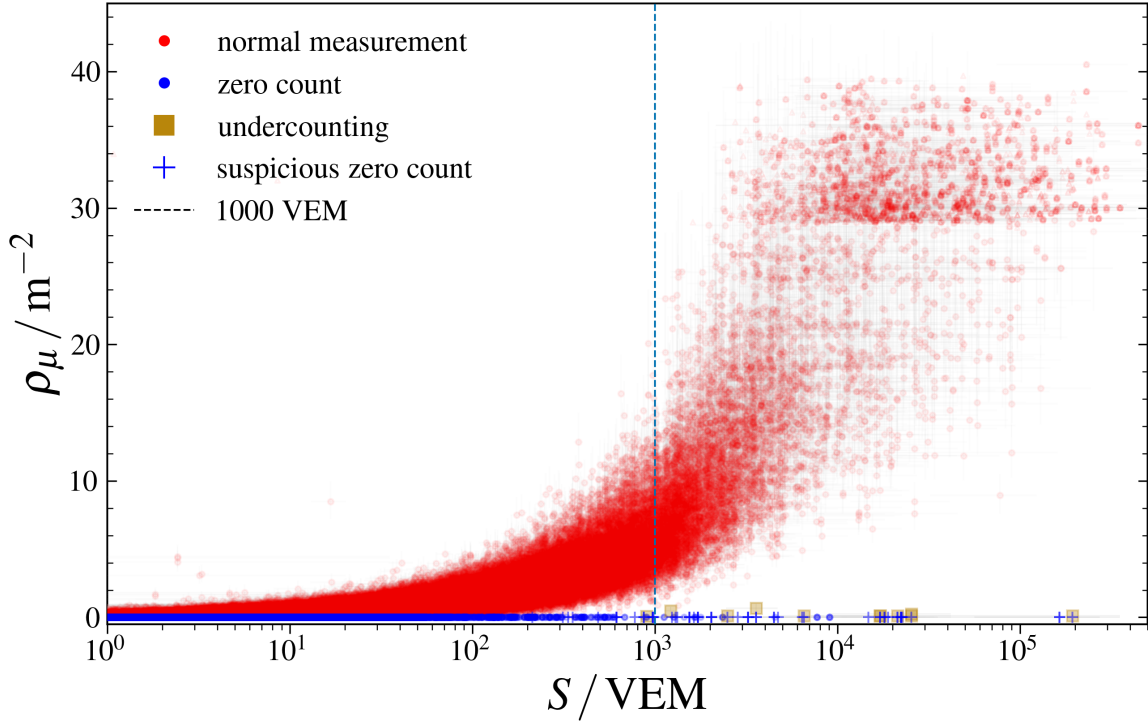


Figure B.2: Correlation between reconstructed UMD muon density and WCD signal for modules from UMD-750 and UMD-433 events. The dashed blue line marks the reference value of $S = 1000$ VEM, close to the onset of WCD saturation. All UMD zero-muon measurements are highlighted in blue. Among these, the most suspicious cases, occurring at core distances $\lesssim 100$ m where $\rho_\mu \approx 0$ is unphysical, are shown as blue crosses. Modules reconstructed with one, two, or three muons at similar core distances are displayed as yellow squares. Only measurements reaching the MLDF fit stage are shown here.

composition-sensitive observables based on the reconstructed muon density at a reference distance, $\rho(r_{\text{ref}})$.

A first indication of the problem is provided by the correlation between the reconstructed UMD muon density and the WCD signal. Figure B.2 shows that, besides the physically expected population in which ρ_μ increases smoothly with S , there exists a sparse set of anomalous measurements characterised by large S combined with $\rho_\mu \simeq 0$. The most unambiguous cases occur at small core distances ($r \lesssim 100$ m), where a null muon density is unphysical at the energies of interest. Dedicated case studies and monitoring checks show that these anomalies are typically coherent at the station level and intermittent in time, supporting an origin in trace-matching or timing-synchronisation issues rather than in a persistent, module-local hardware malfunction.

Away from the shower axis, genuine small densities become common and single-event diagnostics are intrinsically limited. The approach adopted is therefore statistical and proceeds in two steps: (i) an *inclusive* geometric preselection of potentially problematic modules, and (ii) an MLDF-based compatibility test that decides whether the corresponding counter(s) are consistent with the shower expectation¹.

¹In case of the UMD-433, the MLDF was not formally parametrised yet. Therefore, a NKG with the slope from the MLDF of the UMD-750, and a reference distance at 300 meters, was employed as a proxy.

We first characterise the distance dependence of the zero-muon fraction through the binned observable

$$\hat{R}(r, E) \equiv \frac{N_{\text{zero}}(r, E)}{N_{\text{tot}}(r, E)}, \quad (\text{B.1})$$

where $N_{\text{zero}}(r, E)$ and $N_{\text{tot}}(r, E)$ are, respectively, the number of zero-muon instances and the total number of module measurements in a given (r, E) bin. At small core distances, the observed zero-muon fraction does not vanish (presence of “buried pinocchios”), yielding a residual baseline probability p_0 for these modules.

The baseline component p_0 is estimated in a control region defined by $r < 100$ m and for events above the array’s full efficiency regime², where zero-muon instances can be unequivocally attributed to buried pinocchios. We compute

$$\hat{p}_0 \equiv \left. \frac{N_{\text{zero}}^{\text{ctrl}}}{N_{\text{tot}}^{\text{ctrl}}} \right|_{r < 100\text{m}} \quad (\text{B.2})$$

which yields

$$p_0 = \hat{p}_0 = (1.42 \pm 0.09) \times 10^{-2}, \quad (\text{B.3})$$

and keep p_0 fixed in the subsequent steps³. Beyond the baseline, $\hat{R}(r, E)$ increases with distance, as expected from physical intuition. In Fig. B.3 (left) we show examples of $\hat{R}(r, E)$ together with the reference level p_0 . To define a distance threshold for the preselection of candidate pinocchio modules, we proceed as follows.

For each (r, E) bin we also compute the statistical uncertainty $\sigma_{\hat{R}}(r, E)$ and define the per-bin significance relative to the baseline,

$$z(r, E) \equiv \frac{\hat{R}(r, E) - p_0}{\sigma_{\hat{R}}(r, E)}. \quad (\text{B.4})$$

Let $\{r_i\}_{i=1}^n$ denote the centres of the distance bins in increasing order and define $z_i(E) \equiv z(r_i, E)$. For a fixed threshold $z_{\text{cut}} = 3$, we define the transition radius as the smallest distance bin for which the excess is significant and remains significant in the immediately following bin,

$$r_{\text{th}}(E) \equiv \min \{r_i : z_i(E) > z_{\text{cut}} \wedge z_{i+1}(E) > z_{\text{cut}}\}. \quad (\text{B.5})$$

Requiring two consecutive bins above z_{cut} suppresses isolated statistical fluctuations. This transition radius marks the onset of a statistically significant excess above the buried-pinocchio baseline, and is used to preselect potentially problematic modules for the event-by-event compatibility test described in section B.2.2. In practice, we tag a module as suspicious if it reconstructs a number of muons compatible with zero within uncertainties and satisfies $r < r_{\text{th}}(E) + \epsilon$, where ϵ is a safety margin. This heuristic is deliberately permissive, since the final decision to reject the module is based on a stricter criterion.

The second step is an event-by-event hypothesis test that quantifies whether the strip-multiplicity data of a flagged UMD station are compatible with the MLDF prediction. The test is formulated at the level of the number of active strips K in the module trace window, and then propagated at the counter-level by aggregating the information of all involved modules. For a correctly synchronised module (“shower signal”), the expected number of shower muons is obtained from the MLDF prediction at the module position, including the corner-clipping correction, and the corresponding strip-activation likelihood $P(K = k \mid \mu, p_{\text{cc}}, \text{good})$

² $\lg(E/\text{eV}) > 16.8$ for the SD-433 and $\lg(E/\text{eV}) > 17.3$ for the SD-750 up to 45° in zenith.

³Note that this quantity is computed at the module level. At the counter level, the corresponding value is $(1.38 \pm 0.1) \times 10^{-2}$, which we consider equivalent. Since most counters comprise three modules, the same increase is expected in both the numerator and the denominator.

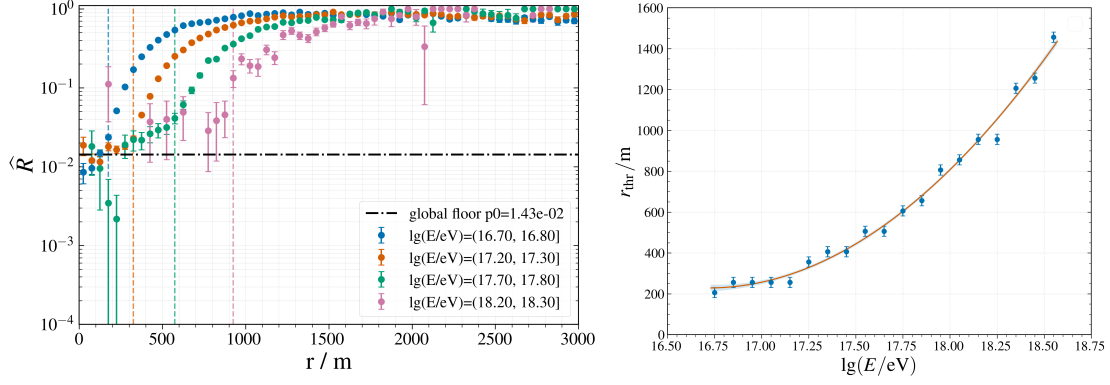


Figure B.3: *Left:* Fraction of modules with zero reconstructed muons, $\hat{R}(r, E)$, as a function of lateral distance for several energy bins. The baseline probability, p_0 , is shown as a horizontal dashed line. The thresholds $r_{\text{th}}(E)$ for different energy bins are denoted by a vertical dashed lines. *Right:* the transition radius, $r_{\text{th}}(E)$, parametrised with a polynomial fit (red line); the thin shaded band represents the fit uncertainty.

is modelled as a binomial process over the detector segmentation. For a trace-mismatched (“background”) module, the recorded window is assumed to contain no shower muons and the strip activations follow a noise-only likelihood $P(K = k \mid \text{bad})$ driven by atmospheric muons and instrumental noise.

The relative support for the two hypotheses is quantified by the likelihood ratio

$$\Lambda(k; \hat{\mu}) = \frac{P(K = k \mid \text{bkg})}{P(K = k \mid \mu, p_{\text{cc}}, \text{shower})}. \quad (\text{B.6})$$

with the prior trace-mismatch probability set to $\pi \simeq p_0$. In practice, the expected muon content used in the denominator is obtained from a *diagnostic* MLDF fit (performed after temporarily excluding the flagged counter under test, so that the test compares the counter to an MLDF prediction not constrained by its own data. The per-module likelihood ratios are then aggregated into a counter-level statistic (a product, or equivalently a sum in $\ln \Lambda$) and the counter is rejected only if the aggregated evidence exceeds a configurable threshold and at least a minimum number of modules are individually incompatible (for counters with three modules available, this requirement protects against isolated downward fluctuations). If a counter is rejected, the MLDF is refit on the cleaned measurement set and the result is stored; otherwise, the original fit parameters are retained. Each event is also assigned to an applicability class that records whether the procedure was enabled, whether suspicious counters existed, and whether a diagnostic fit could be performed.

The procedure was validated on dedicated UMD-433 and UMD-750 Phase I test productions. Among events that reach the MLDF stage, the reconstruction is dominated by cases with no suspicious counters, or with flagged counters found compatible with the MLDF prediction; therefore the final result remains the original fit solution. Only a small fraction of events requires the cleaned refit, corresponding to cases where at least one tested counter is classified as incompatible by the counter-level likelihood-ratio decision. This fraction is 0.88% for UMD-433 and 0.89% for UMD-750 after all filters. For UMD-433, an additional sizeable subset contains suspects but does not allow a diagnostic fit (typically because removing the flagged counter(s) leaves too few usable counters), so that no counter-level decision can be taken and the original fit result is kept. In the undiagnosed events, a sub-percent presence of pinocchios is expected.

The global impact of the cleaning on the reconstructed muon content is summarised in Fig. B.4. We plot the relative change at the corresponding reference distance, $\Delta_\rho \equiv$

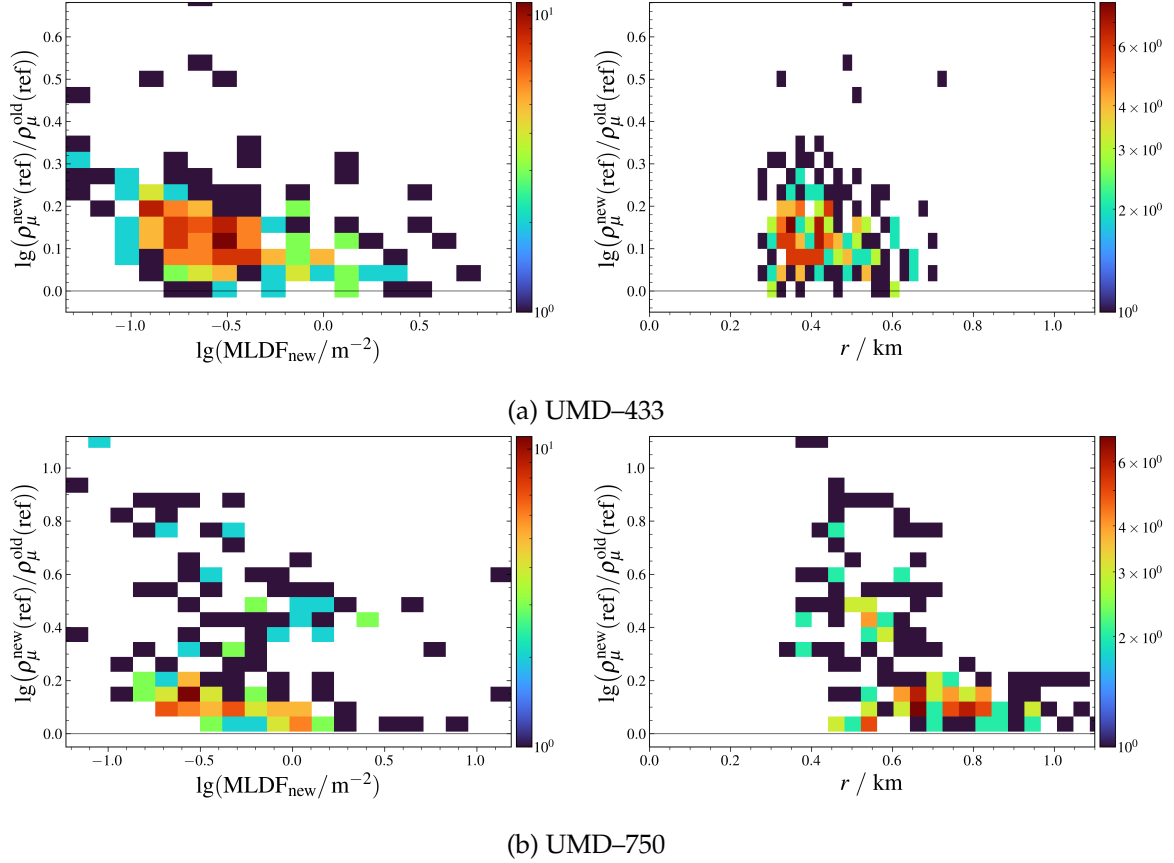


Figure B.4: Relative change in reconstructed muon density at the reference distance after cleaning, $\Delta\rho \equiv \rho_{\mu,\text{new}}(r_{\text{ref}})/\rho_{\mu,\text{old}}(r_{\text{ref}})$, shown as a function of the reconstructed MLDF and lateral distance. The reference distances differ by array: $r_{\text{ref}} = 300$ m for UMD-433 and $r_{\text{ref}} = 450$ m for UMD-750, so the two panels correspond to different characteristic distance scales. The distributions are built only from events that successfully reconstruct both in the legacy sequence (without rejection) and in the new sequence (with pinocchio exclusion); therefore, the most pathological cases that would fail standard ADST quality cuts after module rejection are not represented and the observed shifts should be regarded as conservative.

$\rho_{\mu}^{\text{new}}(r_{\text{ref}})/\rho_{\mu}^{\text{old}}(r_{\text{ref}})$, on a logarithmic scale. For UMD-433 we use $r_{\text{ref}} = 300$ m, while for UMD-750 we use $r_{\text{ref}} = 450$ m; therefore, the two panels probe different characteristic distances for the respective arrays. In both configurations the distribution is clearly shifted towards $\Delta\rho > 1$, indicating that the legacy reconstruction (*without* the rejection method) tends to underestimate the MLDF normalisation when undercounting anomalies are present. Despite the limited statistics, the observed cases show that malfunctioning counters can easily bias the reference muon density by $\gtrsim 30\%$ or more. The dependence of $\Delta\rho$ on $\lg(\rho_{\text{ref}}^{\text{new}})$ and on the lateral distance r further indicates that the effect is strongest for pinoccios located closer to the shower axis and for events with relatively small underlying muon densities.

The reconstructed events used for physics analyses in this thesis correspond to the updated sequence in which this undercounting mitigation is enabled by default. Additional material from, including detailed case studies, the full set of reconstruction-stage codes and statistics, the derivation of the counter-level likelihood product, and the default configuration block, is deferred to section B.2.

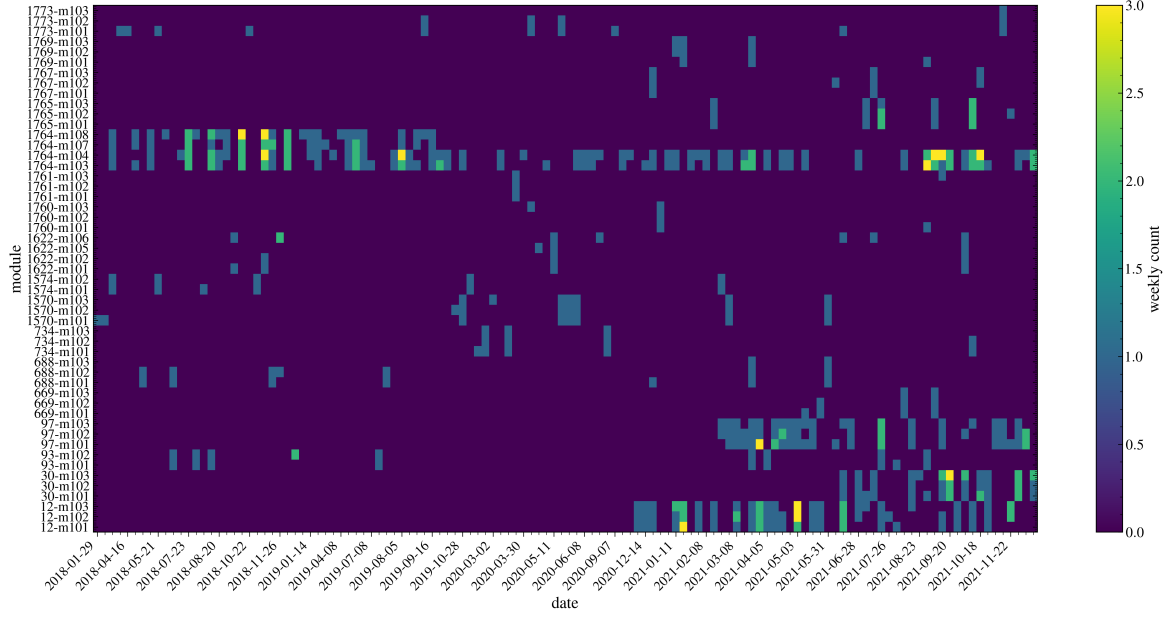


Figure B.5: Weekly occurrence of the most evident buried pinocchios for Phase I data, shown as a 2D histogram of module identifier (station_id–module_id) versus time. Only modules with reconstructed muon count $N_{\mu}^{\text{rec}} \leq 3.2$ and core distance $r \leq 100$ m are included. Five-square-meter modules are excluded, as they are not used in the current MLDF reconstruction, and modules known to be operating in confirmed bad periods are not shown.

B.2 Technical details of pinocchio rejection

This appendix collects the technical details, configuration blocks, reconstruction-stage bookkeeping, and the additional validation plots for the method to reject "pinocchio counters" described in (section B.1.2.2).

B.2.1 Time structure and an illustrative case study

To better understand the nature of these anomalies, we selected a sample of the most pronounced cases for investigation: events with core distances $\lesssim 100$ m, where vanishing muon densities are unphysical. For these events, we examined both high-level UMD reconstruction outputs and low-level monitoring observables. No clear signs of malfunction or systematic instability were found for the majority of the involved modules. In particular, the affected modules typically behave normally outside the event in question.

Figure B.5 shows a time histogram that reveals an additional pattern: the rate of the most evident buried pinocchios rises in late 2020 and early 2021. When the effect appears, it is typically coherent at the station level, affecting all UMD modules of a given station at the same time. This coherence produces clear vertical bands in the histogram and points to an origin in station-level operations, most plausibly in the exchange of timing and triggering information with the WCD.⁴ In such cases, recovering the correct signal becomes challenging because the full UMD readout of the station is compromised. Taken together, these observations could imply timing-synchronisation or trace-matching issues between the WCD and UMD data streams, rather than a persistent hardware malfunction confined to individual modules.

⁴A specific case is visible for station 1764: after the last pinocchio episodes seen in 2019, some of its UMD modules were reassigned to a neighbouring doublet station, resulting in a reduced vertical signature thereafter.

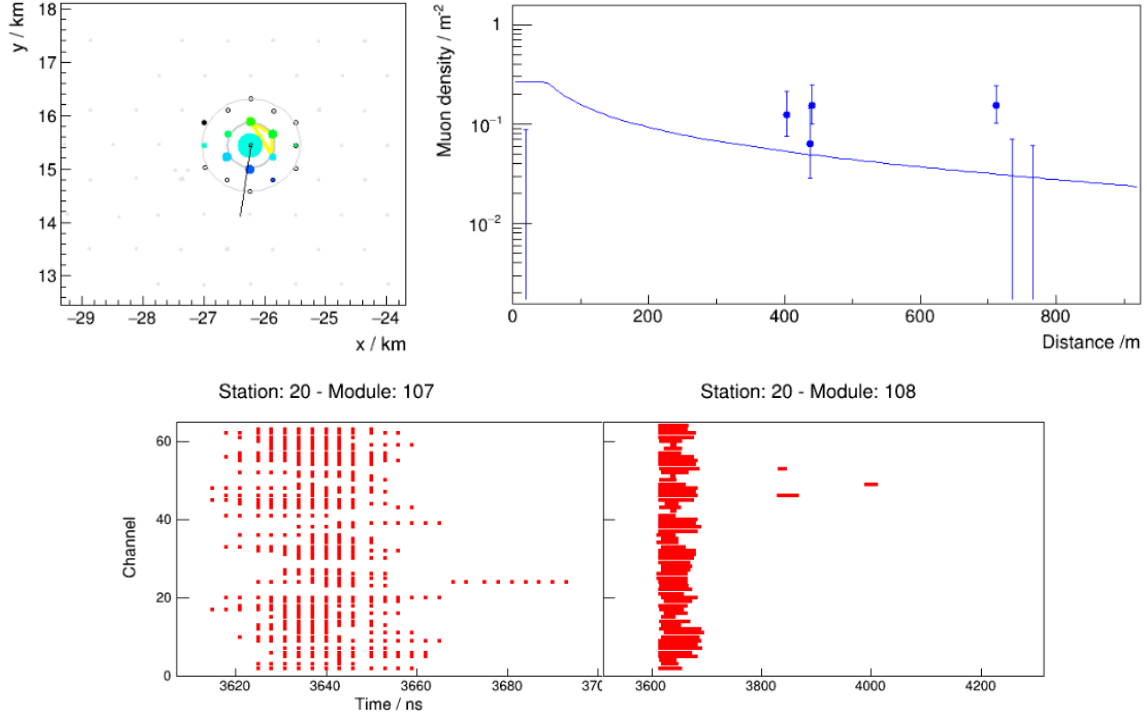


Figure B.6: Case study of a buried pinocchio in Auger ID 212851156000. *Upper panel:* Event footprint on the SD array (left) and the reconstructed MLDF (right). The closest UMD station to the shower core is station 1764, which recorded zero muons. *Lower panel:* The two non-rejected UMD modules of the nearby doublet station 20 show several muon hits in their traces, according to what is expected of a module close to the shower axis.

A representative example is shown in Fig. B.6, focusing on station 1764 and one of its neighbouring doublets. In this event the shower core lands almost directly above station 1764, yet the modules associated with its underground counterpart record zero muons (top right panel). In contrast, the doublet station 20, coupled to other UMD modules that are physically located beneath station 1764, measures the expected large number of muons (bottom right panel). This behaviour strongly supports the hypothesis that the traces in the modules from station 1764 assigned to the air-shower event do not correspond to the correct trigger window, and therefore do not encode a shower signal.

B.2.2 Likelihood model for “good” and “bad” modules

This section summarises the strip-multiplicity likelihoods used in the method. For a module that is correctly synchronised with the SD trigger (hereafter a module that encodes “shower” signals), the expected number of muons crossing the active area is

$$\hat{\mu} = \mu (1 + p_{cc}), \quad (\text{B.7})$$

where μ is the mean number of muons predicted by the MLDF at the module position and p_{cc} accounts for muons grazing the edges of the scintillator strips and potentially additional overcounting effects. Assuming that muons are independent and that each strip records at most a binary on/off response in a given time window, the probability that a single strip is activated is

$$p_{\text{seg}} = 1 - \exp\left(-\frac{\hat{\mu}}{n_s}\right), \quad (\text{B.8})$$

where n_s represents the detector segmentation. The number of active strips K follows a binomial distribution,

$$P(K = k \mid \hat{\mu}, \text{shower}) = \binom{n_s}{k} p_{\text{seg}}^k (1 - p_{\text{seg}})^{n_s - k}. \quad (\text{B.9})$$

For a pinocchio (hereafter a module encoding “background”), we assume that no shower muons are present in the recorded time window. However, several non-shower processes can still produce isolated strip activations: atmospheric muons crossing the scintillator, the dark rate of the SiPMs (including optical cross-talk), and spontaneous light emission in the optical fibres. Let R_{atm} denote the rate of atmospheric muons per strip, and R_{noise} the effective sum of instrumental noise contributions per strip above the binary threshold. The total non-shower rate per strip is then

$$R_b = R_{\text{atm}} + R_{\text{noise}}. \quad (\text{B.10})$$

For the UMD electronics configuration considered here, the trace window consists of 2048 time bins of width 3.125 ns, yielding

$$T = 2048 \times 3.125 \text{ ns} \approx 6.4 \mu\text{s}. \quad (\text{B.11})$$

The mean number of non-shower counts in the trace of a single strip is

$$\lambda_{\text{strip}} = R_b T, \quad (\text{B.12})$$

so that, under the assumption of a Poisson process per strip, the probability that a strip is activated at least once is

$$p_b = 1 - \exp(-\lambda_{\text{strip}}). \quad (\text{B.13})$$

Typical values inferred from the noise studies are $R_{\text{atm}} \approx 27 \text{ s}^{-1}$ [203] and $R_{\text{noise}} \sim 100\text{--}300 \text{ s}^{-1}$ at the binary threshold [121], leading to $\lambda_{\text{strip}} \sim 10^{-3}$ and $p_b \simeq \lambda_{\text{strip}}$. A module encoding background with n_s strips therefore has, under the noise-only hypothesis, a binomial distribution

$$P(K = k \mid \text{bkg}) = \binom{n_s}{k} p_b^k (1 - p_b)^{n_s - k}. \quad (\text{B.14})$$

For representative values $R_b \approx 200\text{--}250 \text{ s}^{-1}$ and $n_s = 64$, the expected number of spurious hits per module is $\mu_b = n_s p_b \simeq 0.1$ and

$$\begin{aligned} P(K = 0 \mid \text{bkg}) &\approx e^{-\mu_b} \simeq 0.91, \\ P(K = 1 \mid \text{bkg}) &\simeq 0.085, \\ P(K = 2 \mid \text{bkg}) &\simeq 4 \times 10^{-3}, \\ P(K = 3 \mid \text{bkg}) &\simeq 10^{-4}, \end{aligned} \quad (\text{B.15})$$

while the probability to observe $k \geq 4$ activations from non-shower processes alone remains very small ($\sim 10^{-6}$). The likelihood in Eq. (B.14) therefore provides a model for the strip multiplicities expected from trace mismatches, including the possibility of modules with one, two, or three active strips due to the combined effect of atmospheric muons and detector noise.

B.2.3 Mixture model and likelihood-ratio test

We now combine the two hypotheses into a simple mixture model. Using the baseline zero-muon fraction p_0 extracted in section B.1.2.2, we assign a prior probability

$$\pi \equiv P(\text{pinocchio}) \simeq p_0 \approx 1.4 \times 10^{-2}, \quad P(\text{not a pinocchio}) = 1 - \pi, \quad (\text{B.16})$$

for a given module to be affected by a trace mismatch. For an observed number of active strips $K = k$, the marginal probability under the mixture model is

$$P(K = k | \hat{\mu}) = (1 - \pi) P(K = k | \hat{\mu}, \text{shower}) + \pi P(K = k | \text{bkg}). \quad (\text{B.17})$$

Several key points should be mentioned at this stage. (i) We neglect the background contribution in the shower term: atmospheric contamination is assumed to be negligible compared to the shower signal, and part of the residual noise effects may also be effectively absorbed into the p_{cc} parameter. (ii) The background term relies on the hypothesis that pinocchios arise from a trace mismatch. (iii) Under the “shower” hypothesis, the expected number of muons is obtained from a diagnostic MLDF fit in which the modules of the counter under test are excluded. The relative support for the shower and background hypotheses is quantified by the likelihood ratio

$$\Lambda(k; \hat{\mu}) = \frac{P(K = k | \text{bkg})}{P(K = k | \hat{\mu}, \text{shower})}. \quad (\text{B.18})$$

Using Bayes’ theorem, the posterior probability that a given module is a pinocchio can be written in terms of Λ and the prior odds:

$$P(\text{pinocchio} | K = k, \hat{\mu}) = \frac{\pi \Lambda(k; \hat{\mu})}{(1 - \pi) + \pi \Lambda(k; \hat{\mu})}. \quad (\text{B.19})$$

A rejection criterion at module-level is then formulated as a likelihood-ratio test between the two hypotheses:

$$\Lambda(k; \hat{\mu}) > \Lambda_{\text{th}} \implies \text{module rejected as a buried pinocchio}, \quad (\text{B.20})$$

where Λ_{th} is a configurable threshold encoding the acceptable trade-off between false positives (rejecting genuinely shower-encoding modules) and false negatives (accepting background-encoding modules). For equal costs of misclassifying good and bad modules ($P(\text{pinocchio} | K) = 0.5$) the natural choice is

$$\Lambda_{\text{th}} \approx \frac{1 - \pi}{\pi} \sim 70\text{--}100, \quad (\text{B.21})$$

corresponding to rejecting a module whenever the noise-only hypothesis is tens of times more likely than the shower hypothesis. More conservative choices (e.g. targeting a posterior $P(\text{pinocchio} | K) \gtrsim 0.9$) translate into proportionally larger values of Λ_{th} .

Early versions applied the test strictly at the module level. In test productions, this occasionally removed isolated and valid downward fluctuations of a module in otherwise well-behaved counters. Since trace mismatches are expected to manifest at the counter level, we promote the final decision to the counter level by aggregating the module-level evidence. Consider a counter with M available modules. For each module $i = 1, \dots, M$ we compute $\Lambda_i \equiv \Lambda(k_i; \hat{\mu}_i)$ and define the counter-level likelihood ratio as

$$\Lambda_{\text{c}} \equiv \prod_{i=1}^M \Lambda_i \quad \left(\text{equivalently } \ln \Lambda_{\text{c}} = \sum_{i=1}^M \ln \Lambda_i \right). \quad (\text{B.22})$$

A counter is rejected if

$$\Lambda_c > \Lambda_{\text{th}} \quad \wedge \quad N_{\text{rej}} \geq N_{\text{min}}(M), \quad (\text{B.23})$$

where $N_{\text{rej}} \equiv \sum_{i=1}^M \mathbb{I}(\Lambda_i > \Lambda_{\text{th}})$ is the number of individually rejected modules and

$$N_{\text{min}}(M) \equiv \begin{cases} 1, & M \leq 2, \\ 2, & M \geq 3. \end{cases} \quad (\text{B.24})$$

The justification for the counter-level approach is described in section B.2.5⁵. In words: we require (i) sufficiently strong aggregated evidence against the shower hypothesis and (ii) for counters with three or more available modules, at least two modules to be individually incompatible with the shower expectation.

In practice, the likelihood ratio test is applied only to counters for which at least one module has been tagged as suspicious by the method of section B.1.2.2. For such a counter, *all* its available modules ($i = 1, \dots, M$) enter the likelihood-ratio computation in Eq. (B.22)–Eq. (B.23). Suspicious counters are evaluated in order of increasing core distance, since the inner-most stations dominate the MLDF fit. Given the expected rarity of this malfunctioning, once a counter is rejected in an event, no additional counters are tested; this prevents over-cleaning and limits the impact of the procedure to the measurements most relevant for the MLDF reconstruction.

The aggregation enhances robustness in two complementary ways. First, a *single* genuine local downward fluctuation (one module with unusually small k) is typically accompanied by other modules whose multiplicities remain compatible with the MLDF, yielding $\Lambda_i \lesssim \mathcal{O}(1)$ for those modules; these contributions counteract the outlier in the product and reduce the chance that a valid counter is removed. Second, when a trace mismatch affects the station coherently, multiple modules tend to exhibit $k = 0$ or very small k ; even if an individual module provides only moderate evidence ($k > 0$), the product Λ_c rapidly accumulates into strong counter-level evidence against the shower hypothesis. The additional requirement $N_{\text{rej}} \geq N_{\text{min}}(M)$ further suppresses accidental removals by demanding that the incompatibility is supported by *more than one* module for counters with $M \geq 3$.

B.2.4 Physical interpretation and trade-offs

The likelihood–ratio criterion admits a transparent physical interpretation in terms of the expected muon content of the shower and the residual noise in mismatched traces. For a given module, the shower–module likelihood in Eq. (B.9) increases when the observed strip multiplicity k is compatible with the MLDF prediction $\hat{\mu}$, whereas the background–module likelihood in Eq. (B.14) depends only on the combined R_b , the trace length, and the segmentation. The likelihood ratio $\Lambda(k; \hat{\mu})$ therefore quantifies, for each module, how much more likely the measurement is to arise from a mismatched trace than from a correctly matched shower signal.

For zero–strip modules ($k = 0$) the expressions simplify considerably. Using Eqs. (B.9) and (B.14), one finds

$$P(K = 0 \mid \text{shower}) = e^{-\hat{\mu}}, \quad P(K = 0 \mid \text{bkg}) \simeq (1 - p_b)^{n_s} \approx \exp(-\mu_b), \quad (\text{B.25})$$

where $\mu_b = n_s p_b$ is the mean number of non–shower counts per bad module. For representative values $R_b \approx 200\text{--}250 \text{ s}^{-1}$, one obtains $\mu_b \simeq 0.09$ and $P(K = 0 \mid \text{bkg}) \simeq 0.91$, so that

$$\Lambda(0; \hat{\mu}) \simeq P(K = 0 \mid \text{bkg}) e^{\hat{\mu}}. \quad (\text{B.26})$$

⁵Since the Λ_{th} threshold is driven primarily by the prior p_0 , which is roughly the same counter-level, the original value remains valid.

The “indistinguishability point” between good and bad zeros is given by $\Lambda(0; \hat{\mu}) = 1$, which using Eq. (B.7) yields

$$\mu_{\text{eq}} \simeq \frac{-\ln[P(K=0 \mid \text{bkg})]}{1 + p_{\text{cc}}} \approx 8.6 \times 10^{-2}. \quad (\text{B.27})$$

For an effective module area of $A_{\text{eff}} \simeq 10.5 \text{ m}^2$, mean p_{cc} values, and nearly vertical showers ($\cos \theta \simeq 1$), this corresponds to a muon density $\rho_{\mu} \simeq \mu_{\text{eq}} / A_{\text{eff}} \sim 8 \times 10^{-3} \text{ m}^{-2}$. Below this density scale, zero-strip modules are essentially indistinguishable from affected modules: the shower-module and background-module likelihoods are almost identical, and the posterior probability of being a pinocchio remains close to the prior π . In other words, in the far tail of the MLDF where the expected number of muons per module is much smaller than unity, zero-muon measurements carry little information about trace mismatches and should be retained.

As the expected muon content increases, the two hypotheses separate. For the likelihood-ratio threshold $\Lambda_{\text{th}} \approx (1 - \pi) / \pi \sim 70$ corresponding to a posterior $P(\text{pinocchio} \mid K) \simeq 0.5$, the condition $\Lambda(0; \hat{\mu}) = \Lambda_{\text{th}}$ gives

$$\mu_{\text{crit}}^{(0)} \simeq \frac{\ln[\Lambda_{\text{th}} / P(K=0 \mid \text{bkg})]}{1 + p_{\text{cc}}} \approx 4.0 \quad (\text{B.28})$$

or $\rho_{\mu} \approx 0.38 \text{ m}^{-2}$ for $A_{\text{eff}} \simeq 10.5 \text{ m}^2$. At such densities, the probability for a shower-encoding module to record $k=0$ becomes of order $e^{-4} \sim 2 \times 10^{-2}$ (and smaller as μ increases), while the background-module probability remains at $\simeq 0.91$. The resulting likelihood ratio implies that a zero-strip module at this density is far more likely to be a buried pinocchio than a genuine downward fluctuation. For larger μ the evidence for the background hypothesis grows rapidly, and zero-strip modules are almost certainly inconsistent with a correctly matched trace.

Low but non-zero multiplicities ($k=1, 2, 3$) exhibit a similar interplay. In a background-module, the expected number of spurious hits is $\langle k \rangle_{\text{bad}} \simeq \mu_{\text{b}} \sim 0.1$, so that one or two active strips occur at the percent level even in the absence of shower muons. In the far MLDF tail where $\mu \ll 1$, both models predict $k=0$ or $k=1$ with comparable probabilities and the likelihood ratio remains close to unity. In contrast, closer to the shower core where μ is of order a few or larger, the shower-module likelihood for $k=1$ or $k=2$ is strongly enhanced with respect to the broadened background distribution, and such modules are typically retained. The classification of $k=1, 2, 3$ modules therefore depends sensitively on the local MLDF prediction: they look noise-like in the extreme tail, signal-like in the intermediate regime where $\mu \sim 0.1$ – 1 , and become again increasingly incompatible with the shower hypothesis when μ is very large and the expectation moves towards many active strips.

It is important to stress that the background-module likelihood in Eq. (B.14) lumps the known non-shower contributions into an effective rate R_{b} , and may therefore miss additional sources of spurious strip activations. Any such contributions would broaden the distribution of K under the noise-only hypothesis, most noticeably for $k > 0$. A dedicated calibration of the noise rate at the level of individual strips, including its dependence on operating conditions, would refine the bad-module model and thus improve the quantitative accuracy and stability of the rejection criterion.

B.2.5 Counter-level aggregation derivation

Consider one counter with M available modules. For each module $i = 1, \dots, M$, let $K_i \in \{0, \dots, n_s\}$ be the observed number of active strips in the acquisition window, and

denote by $\mathbf{K} \equiv (K_1, \dots, K_M)$ the vector of active strip multiplicities for that counter. For each module i , the MLDF fit provides a prediction $\hat{\mu}_i$ for the expected number of shower muons at the module position.

We define two counter-level hypotheses:

$$H_S : \text{the counter is correctly synchronised,} \quad (\text{B.29})$$

$$H_B : \text{the counter is affected by a trace mismatch.} \quad (\text{B.30})$$

Under these hypotheses, the per-module likelihoods were already introduced in section B.1.2 where the “background” likelihood does not depend on $\hat{\mu}$.

The key point is that the MLDF parameters are obtained *excluding the full counter under test* from the MLDF fit. Consequently, the predictions $\hat{\mu}_i$ do not depend on $\mathbf{K}$ and can be regarded as fixed for the purpose of testing this counter. In other words, we use the orthogonality condition

$$\hat{\mu}_i \perp\!\!\!\perp \mathbf{K} \quad (\text{because the counter is excluded from the MLDF fit}). \quad (\text{B.31})$$

Assuming further that, conditional on the hypothesis $H \in \{H_S, H_B\}$ and on the fixed predictions $\hat{\mu}_i$, the M module measurements are independent, the joint likelihood factorises:

$$P(\mathbf{K} = \mathbf{k} \mid \hat{\mu}_i, H) = \prod_{i=1}^M P(K_i = k_i \mid \hat{\mu}_i, H). \quad (\text{B.32})$$

This conditional-independence assumption is justified as follows. For fixed shower geometry and a fixed “true” MLDF, the expected muon content at each module is fully specified by $\hat{\mu}_i$, and the remaining randomness entering K_i is dominated by (i) the local muon counting fluctuations at that module and (ii) the module’s instrumental noise process. Both sources are, to a good approximation, local to each module once $\hat{\mu}_i$ is described by the shower expectation. Any residual correlations between modules are precisely what is removed when we treat the MLDF prediction as fixed and do not re-fit it using the tested counter.

The counter-level likelihood ratio for the *vector* observation $\mathbf{K} = \mathbf{k}$ is defined as

$$\Lambda_c(\mathbf{k}; \hat{\mu}_i) \equiv \frac{P(\mathbf{K} = \mathbf{k} \mid H_B)}{P(\mathbf{K} = \mathbf{k} \mid \hat{\mu}_i, H_S)}. \quad (\text{B.33})$$

Using the factorisation in Eq. (B.32) under each hypothesis, we obtain

$$\Lambda_c(\mathbf{k}; \hat{\mu}_i) = \frac{\prod_{i=1}^M P(K_i = k_i \mid H_B)}{\prod_{i=1}^M P(K_i = k_i \mid \hat{\mu}_i, H_S)} = \prod_{i=1}^M \frac{P(K_i = k_i \mid H_B)}{P(K_i = k_i \mid \hat{\mu}_i, H_S)}. \quad (\text{B.34})$$

Using the per-module likelihood ratio we recover the product form used in the section B.1.2:

$$\Lambda_c = \prod_{i=1}^M \Lambda_i, \quad \ln \Lambda_c = \sum_{i=1}^M \ln \Lambda_i. \quad (\text{B.35})$$

If the tested counter were included in the MLDF fit, then $\hat{\mu}_i$ would become a function of the same data $\mathbf{K}$ being tested, i.e. $\hat{\mu}_i = \hat{\mu}_i(\mathbf{K}, \dots)$. In that case, the “shower” likelihood in the denominator of Eq. (B.33) would implicitly depend on $\mathbf{K}$, and the factorisation in Eq. (B.32) would no longer correspond to a proper likelihood for a fixed model. Formally, one would need to treat fitted parameters as additional nuisance parameters ϕ and compute

$$P(\mathbf{K} \mid H_S) = \int P(\mathbf{K} \mid \phi, H_S) p(\phi) d\phi, \quad (\text{B.36})$$

which is easier said than done.

B.2.6 Implementation and results on test production

The implementation of the method is designed to be fully backward compatible with the previous UMD reconstruction sequence. The feature can be enabled or disabled entirely via a configuration flag in the .xml file of the *MdLDFFinderAG* implemented starting from commit [5a0d9ff1](#). The default configuration is displayed in section B.2.7. When disabled, the reconstruction proceeds exactly as in the standard approach, and only the standard MLDF fit (*Fit A*) is performed and stored. The only difference is that the fit also optimises the shower core position [126], which is expected to become the default strategy for the MLDF fit in the near future⁶.

The algorithm is implemented as a three-stage pipeline:

1. The first stage performs the standard MLDF reconstruction (*Fit A*) using the usual candidate selection. If *Fit A* fails, the rejection logic is skipped and no MLDF result is stored. If *Fit A* succeeds and the method is enabled, the algorithm computes $r_{\text{th}}(\lg E)$ and scans candidate modules. A module is tagged as suspicious if $\hat{n}_\mu$ is below the configured threshold, and if it also lies within the region $r < r_{\text{th}} + \epsilon$. A counter is *flagged* if at least one of its modules is suspicious. Only the flagged counters closest to the shower axis are considered, up to a configurable maximum.
2. For each flagged counter, the counter is temporarily excluded *as a whole* (i.e. all its available modules are masked out) and a second reconstruction (*Fit B*) is performed. *Fit B* is *diagnostic only*: its purpose is to enforce the rejection criteria explained in section B.2.3.
3. If one counter is rejected and event selection remains viable, the reconstruction is repeated a third time (*Fit C*) on the cleaned candidate list (with the rejected counter excluded). *Fit C* is then stored as the final MLDF result. If no counter is rejected, the final MLDF result remains *Fit A*.

To track the reconstruction path and outcome downstream, the module stores an integer stage code, *MdLDFRecStage*, that encodes the status of *Fit A*, the rejection logic, the outcome of *Fit B*, the outcome of *Fit C*, and which fit (*A* or *C*) was selected as final. In hexadecimal form the code reads:

0x [Final] [C] [B] [ZR] [A],

where each field occupies 4 bits. The definitions of the enums employed for the encoding are presented in Table B.3.

Table B.3: Enum encoding of *MdLDFRecStage*.

Field	Values
Fit A status (A)	0 NotRun, 1 Ok, 2 InvalidMin, 3 InvalidCov
ZR status (ZR)	0 Off, 1 OnNoSuspects, 2 SuspectsButFitBNotViable, 3 SuspectsFitBAttempted, 4 OnButFitAFailed
Fit B result (B)	0 NA, 1 SuccessTestedCompat, 2 SuccessTestedIncompat, 3 SuccessButFitCNotViable, 4 Failed
Fit C result (C)	0 NA, 1 Failed, 2 Success
Final source (Final)	0 None, 1 A, 2 C

In the test productions, the reconstruction stage distribution is dominated by four codes for each array, summarised in Tables B.4 and B.5. Most events either contain no flagged counters, or contain flagged counters that are ultimately found compatible with the MLDF

⁶The MLDF of the SD-433 is not thoroughly parametrised yet. Therefore, the slope parametrisation (with $R_\mu = 320$) for the SD-750 array is used as a proxy and a fixed reference distance at 300 m compatible with the surface detector is enforced.

Table B.4: MdLDF reconstruction stage statistics in the SD-433 test production

Stage (dec)	Stage (hex)	Count	%	Decoded meaning
65553	0x10011	23906	53.14	FitA Ok; ZR enabled but no suspects; Final from Fit A
65569	0x10021	11630	25.85	FitA Ok; suspects present but Fit B not viable; Final from Fit A
65841	0x10131	9061	20.14	FitA Ok; suspects present; Fit B attempted; tested counters compatible; Final from Fit A
139825	0x22231	394	0.88	FitA Ok; suspects present; Fit B attempted; tested counter(s) incompatible; Fit C success; Final from Fit C

Table B.5: MdLDF reconstruction stage statistics in the SD-750 test production

Stage (dec)	Stage (hex)	Count	%	Decoded meaning
65553	0x10011	26094	74.68	FitA Ok; ZR enabled but no suspects; Final from Fit A
65841	0x10131	4839	13.85	FitA Ok; suspects present; Fit B attempted; tested counters compatible; Final from Fit A
65569	0x10021	3697	10.58	FitA Ok; suspects present but Fit B not viable; Final from Fit A
139825	0x22231	312	0.89	FitA Ok; suspects present; Fit B attempted; tested counter(s) incompatible; Fit C success; Final from Fit C

prediction, so the final result remains Fit A. Only a small fraction of events require the cleaned refit (Fit C), corresponding to cases where a flagged counter is classified as incompatible by the counter-level likelihood-ratio decision.

A notable difference arises for the SD-433 array, where a sizeable fraction of events falls into the “suspects but Fit B not viable” category (stage 65569). These are events in which a suspicious counter is present, but excluding it leaves too few remaining counters with usable signal to perform the diagnostic Fit B and, consequently, to take a counter-level decision. This limitation occurs predominantly at lower energies and is therefore more relevant for SD-433 than for SD-750 (see the stage fractions in Tables B.4 and B.5).

After applying the standard quality selections (full-efficiency regime, $\theta < 45^\circ$, and requiring hottest-station/counter association), the sample of fully diagnosed events (stages 65553, 65841, and 139825) amounts to $\mathcal{O}(3 \times 10^4)$ events for each array. For analyses requiring the cleanest possible dataset, the most conservative choice is to restrict to this diagnosed subset. However, using the full Phase I dataset is also acceptable in practice: for analyses that can tolerate a small residual contamination, the fraction of undiagnosed buried-pinocchio cases is expected to remain sub-percent. This residual level can be estimated by contrasting the expected pinocchio rate, 1.42%, with the fraction of counters rejected by the method in the test productions, 0.89%. The comparison implies that roughly 60% of the affected counters are successfully removed, while the remainder should predominantly belong to the undiagnosed category. In this regime, physics results that are robust against a low level of pinocchio contamination should not be measurably influenced.

The likelihood-ratio test yields the $\ln(\Lambda)$ distributions shown in Fig. B.7. The dotted histograms correspond to the subset of modules pre-selected as *suspicious* by the heuristic tagging described above, while the solid histograms include all reconstructed modules,

irrespective of whether their counters were flagged and tested. The vertical black line indicates the likelihood-ratio threshold Λ_{th} adopted in the counter-level decision.

Only suspicious modules with $\ln \Lambda$ to the right of this line provide evidence in favour of the buried-pinocchio (background-only) hypothesis when the per-module likelihood ratios are aggregated at the counter level; suspicious candidates to the left contribute instead to the shower-compatible hypothesis. This two-step structure is deliberate: the heuristic is kept intentionally inclusive to avoid missing pathological cases, and the likelihood-ratio test supplies the discriminating power needed to control false rejections. Despite small differences between the SD-433 and SD-750 productions, the preselection and the likelihood-ratio ranking remain broadly consistent, indicating that the geometric-like tagging captures most of the population that later accumulates at large Λ .

Figure B.8 shows the correlation between the reconstructed WCD signal and the UMD muon density *after* applying the rejection procedure. Only modules from successfully diagnosed events are shown in this plot. The cleaned sample is dominated by the physically expected population, in which ρ_μ increases smoothly with S over several decades, while the anomalous “pinocchio” pattern (large S combined with $\rho_\mu \simeq 0$) is strongly suppressed. The horizontal strip of blue markers indicates modules with a reconstructed zero measurement ($\rho_\mu \simeq 0$), which naturally extends to low signal values. In particular, below $S \simeq 45$ VEM it becomes increasingly difficult to diagnose whether a zero measurement corresponds to a genuine undercounting anomaly, because the expected MLDF prediction in this region approaches values that are compatible with null measurements within fluctuations.

The two gold square markers highlight measurements classified as *undercounting*. In these examples, a single suspicious module survives the procedure and a small number of residual zero counts remain after the method is applied. All such cases were re-examined and found

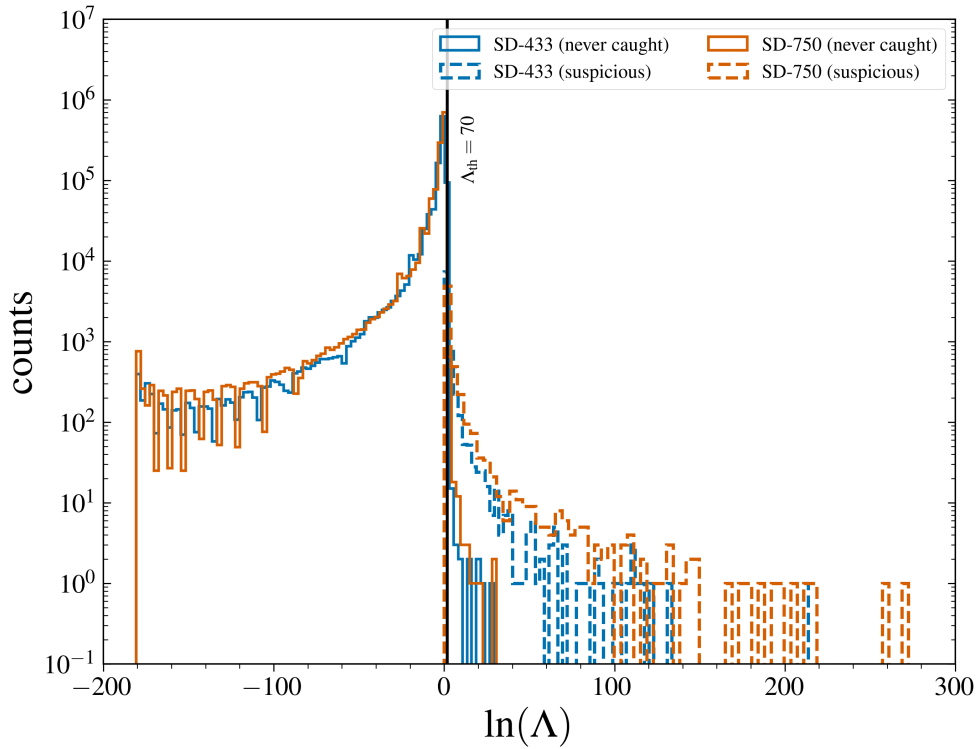


Figure B.7: Distributions of the module-level likelihood ratio $\lg(\Lambda)$ for events from UMD-433 and UMD-750. Solid histograms show all reconstructed modules, while dotted histograms show the subset preselected as suspicious by the geometric-like heuristic.

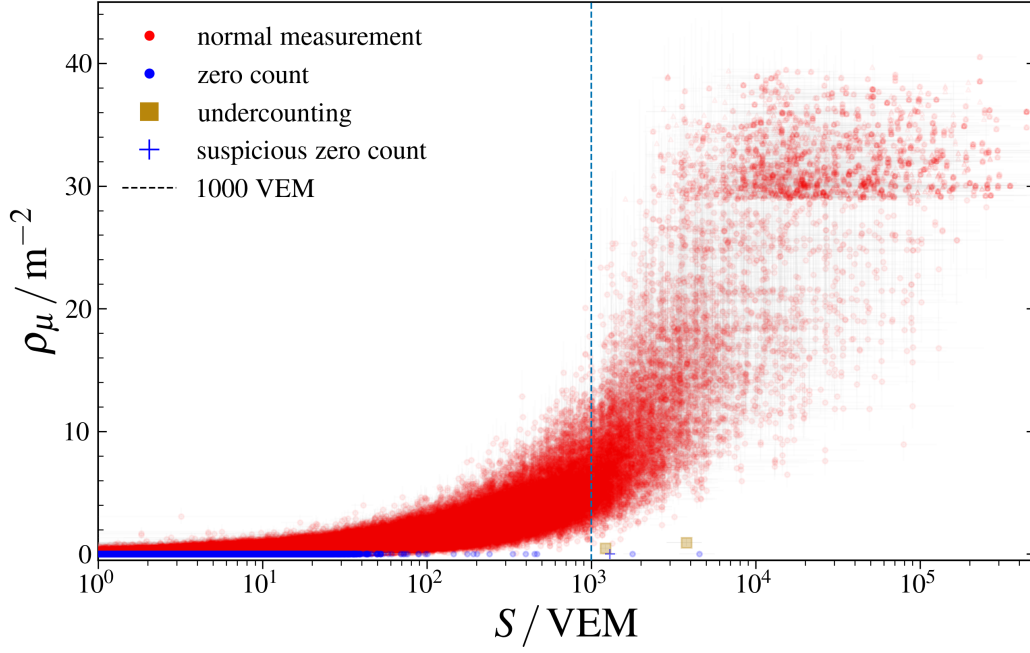


Figure B.8: Scatter of inferred muon density ρ_μ versus SD signal S after applying the rejection procedure. The cleaned sample is dominated by the physically expected population in which ρ_μ increases smoothly with S , while the anomalous “pinocchio” pattern (large S combined with $\rho_\mu \simeq 0$) is strongly suppressed. The horizontal band of near-zero measurements naturally extends to low S , where null muon densities become increasingly compatible with the MLDF expectation within fluctuations; only a small number of undercounting cases survive the criteria. Only measurements reaching the MLDF fit stage are shown here.

to correspond to isolated, single-module failures within otherwise well-behaved counters. The counter is therefore not rejected, because the remaining modules yield strip multiplicities consistent with the MLDF expectation. This behaviour is desirable: it shows that the counter-level decision avoids removing counters due to isolated downward fluctuations, which may still be physical.

In the inspected events, the malfunctioning modules were added to a dedicated list of bad acquisition periods so that they are rejected deterministically during reconstruction. The likelihood-ratio method described in this note is not intended to target these isolated module failures; rather, it is designed to identify the coherent counter-level undercounting pattern characteristic of buried pinocchios.

Figures B.9 illustrate a typical case in which a small number of suspicious candidates at relatively small lateral distances pull the Fit A solution towards a lower MLDF estimate. After applying the likelihood-ratio test, the incompatible counter is removed and the event is refit (Fit C), yielding an MLDF that is more consistent with the non-zero measurements. The inferred muon density at the reference distance increases from $\rho_\mu(r_{\text{ref}}=300 \text{ m}) = 5.8 \pm 0.55$ to 7.4 ± 0.66 , i.e. a relative change of about +28%. This emphasises that even when the visual change in the MLDF curve appears modest, the impact on the key physics proxy $\rho_\mu(r_{\text{ref}})$ can be substantial.

Figures B.10 show a more extreme failure mode, where the Fit A solution becomes strongly biased by an incompatible counter, producing an unphysical behaviour (here $\beta \simeq 0$). After rejection and the Fit C refit, both the slope and the reference density shift markedly ($\beta : 0.00 \rightarrow 4.39$ and $\rho_\mu(r_{\text{ref}}=450 \text{ m}) : 0.24 \pm 0.037 \rightarrow 1.0 \pm 0.27$). In this particular case, the rejected candidates include modules associated with the hottest station; in a standard

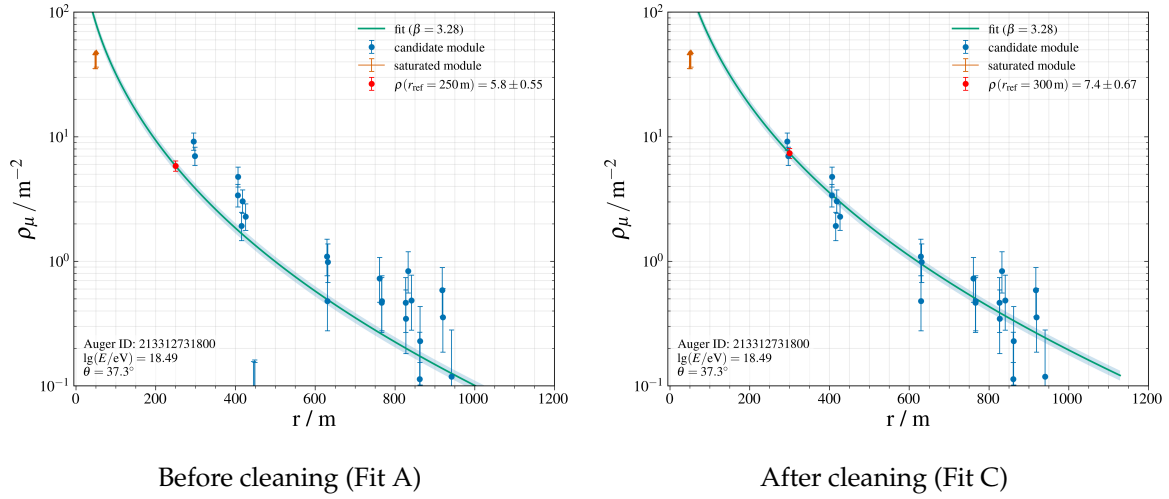


Figure B.9: Event 213312731800 illustrating the impact of rejecting incompatible module candidates. *Left:* Before cleaning (Fit A), suspicious zero-count modules at relatively small lateral distances bias the MLDF estimate downward. *Right:* After applying the likelihood-ratio test and refitting (Fit C), the MLDF becomes more consistent with the non-zero measurements and the inferred reference density increases from $\rho_\mu(r_{\text{ref}} = 300 \text{ m}) = 5.8 \pm 0.55$ to 7.4 ± 0.66 .

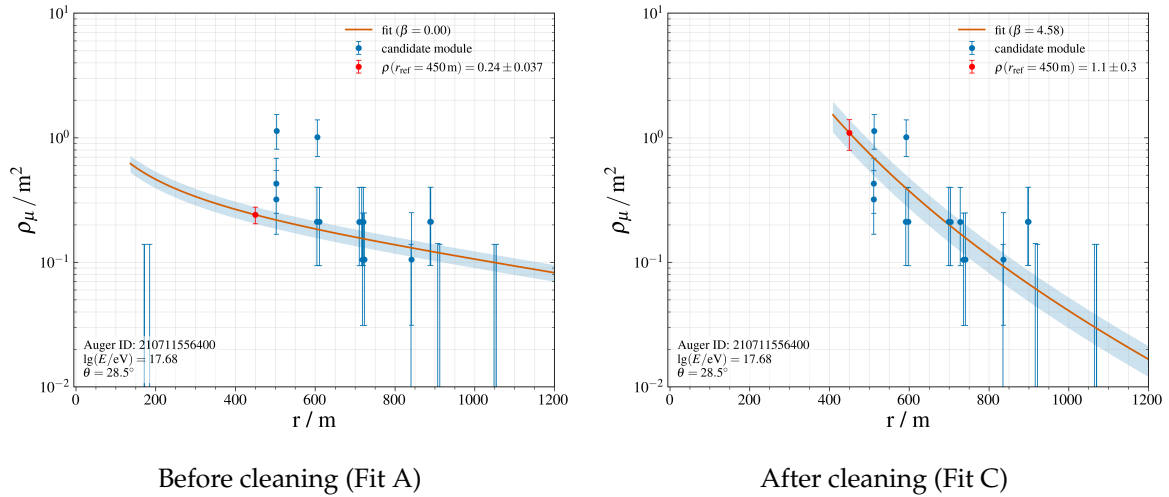


Figure B.10: Example event 210711556400 illustrating an extreme case of the legacy fit. *Left:* Before cleaning (Fit A), incompatible modules can strongly bias the MLDF, leading to unphysical behaviour (here $\beta \simeq 0$). *Right:* After rejection and the Fit C refit, both slope and normalisation shift markedly ($\beta : 0.00 \rightarrow 4.39$ and $\rho_\mu(r_{\text{ref}} = 450 \text{ m}) : 0.24 \pm 0.037 \rightarrow 1.0 \pm 0.27$). In this particular case, the rejected candidates include modules associated with the hottest station; in a standard analysis chain the event would ultimately be removed by ADST quality cuts

analysis chain, a common ADST quality cut (discarding events where the hottest station lacks an associated UMD module) would ultimately remove the event from high-quality datasets intended for physics measurements.

B.2.7 Default configuration block (MdlDFFinder.xml)

The default configuration excerpt for the MdlDFFinderAG module is reproduced below.

```
<zeroRejection>
  <enabled>1</enabled>
```

```

<!-- r_cut [m] = sum_k coeffs[k] * (lgE - polyLogEShift)^k -->
<polyLogEShift> 17.3 </polyLogEShift>
<coeffs>345.44272468 364.71226553 289.20024427 105.27267561</coeffs>
<!-- Tag a module as 'zero-like' if n_mu_hat <= threshold -->
<moduleZeroLikeThreshold>3.2</moduleZeroLikeThreshold>
<!-- Maximum number of suspicious counters to inspect -->
<inspectMaxNCounters>3</inspectMaxNCounters>
<!-- Prior for a module having a mismatched trace -->
<mismatchProbability>0.0138</mismatchProbability>
<!-- Poisson parameter for signal expected from atmospheric muons and noise -->
<poissonNoiseParameter>0.0015</poissonNoiseParameter>
<!-- Likelihood ratio threshold -->
<lrThreshold>70</lrThreshold>
<!-- Soft band around the cut (meters) to accept many modules.
      The restrictive cut is the p-value -->
<epsilonNearCut unit="m">100</epsilonNearCut>
</zeroRejection>

```

B.3 Shape of the mean MLDF validation

B.3.1 Complementary views of the UMD-433 sample

To complement the mean-profile fit discussed in Section 6.1.2, Fig. B.11 shows two event-level views of the UMD-433 sample from which the mean profile was constructed. In both panels, the red markers indicate the binned mean profile, while the black curve shows the Greisen fit with fixed $R_\mu = 247$ m. These plots rely on independent fits, and work as a visual guide to the population of measurements underlying the average shape discussed in the main text.

The left panel displays the energy-rescaled muon density,

$$\tilde{\rho}_\mu \equiv \frac{\rho_\mu}{(E_{\text{rec}}/10^{17.5} \text{ eV})^{0.9}}, \quad (\text{B.37})$$

using a two-dimensional histogram in the $(\lg r, \lg \tilde{\rho}_\mu)$ plane. This rescaling is introduced only as a visualisation aid: it reduces the broadening induced by the strong energy dependence of the muon density and makes the underlying mean trend easier to identify by eye. In this representation, events with zero reconstructed muons are displayed at the fixed value $\lg(\tilde{\rho}_\mu) = -2.0$. This convention does not assign a physical density to those events; it is used only to visualise compactly the zero-muon population, which becomes increasingly relevant at large core distances.

The right panel shows the corresponding event-by-event distribution for the raw muon density ρ_μ , with the points colour-coded by $\sin^2 \theta_{\text{rec}}$. This makes the zenith dependence easier to inspect in the same variable employed elsewhere in the note. Here, zero-muon measurements are displayed at the fixed value $\lg(\rho_\mu) = -1.5$, again purely for visualisation purposes. This panel also makes the origin of the horizontal banding at low densities more transparent. The effect is an expected consequence of the discrete nature of the reconstructed muon counts: when only a few muons contribute to the signal, changing the count by one produces a relatively large step in the inferred density. The spread in $\sin^2 \theta_{\text{rec}}$ further smooths this pattern through the projection factor entering the effective detector area, but does not remove it. At larger densities, where the underlying muon counts are higher, the

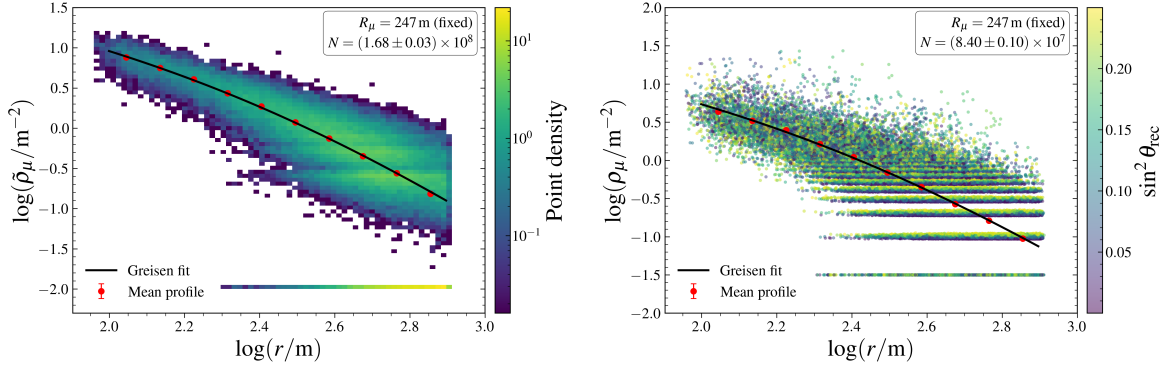


Figure B.11: Two complementary views of the UMD-433 sample used for the mean-shape validation. **Left:** two-dimensional histogram of the energy-rescaled muon density $\tilde{\rho}_\mu = \rho_\mu / (E_{\text{rec}}/10^{17.5} \text{ eV})^{0.9}$ as a function of shower-plane distance, with the mean profile and the Greisen fit for fixed $R_\mu = 247 \text{ m}$ superimposed. Zero-muon measurements are displayed at $\lg(\tilde{\rho}_\mu) = -2.0$. **Right:** event-by-event scatter plot of the raw muon density ρ_μ , colour-coded by $\sin^2 \theta_{\text{rec}}$, again with the mean profile and the same Greisen fit overlaid. Zero-muon measurements are displayed at $\lg(\rho_\mu) = -1.5$.

corresponding fractional steps become smaller and the point cloud gradually appears more continuous.

B.3.2 Shape validation for the UMD-750

This appendix shows that the mean-shape picture established in the main text is not peculiar to the UMD-433. The same exercise was repeated for the UMD-750, using showers above $10^{17.5} \text{ eV}$ and extending to $\theta < 45^\circ$. The resulting mean profile is shown in Fig. B.12. Overall, the agreement with the reduced Greisen form is slightly less close than for the UMD-433, but it remains very good over the fitted radial range and supports the same general interpretation.

Because the zenith interval is broader in this case, the corresponding atmospheric depth is larger, and therefore the value of the muon-spread factor inferred from Eq. (6.7) is also larger. Using that expression, we obtain $R_\mu = 278 \text{ m}$, to be compared with the historical value $R_\mu = 320 \text{ m}$. Since these two scales are closer than in the UMD-433 case, the difference between the corresponding fits is naturally smaller. Fixing $R_\mu = 278 \text{ m}$ yields

$$N = (2.47 \pm 0.05) \times 10^8, \quad (\text{B.38})$$

while fixing instead $R_\mu = 320 \text{ m}$ gives

$$N = (1.96 \pm 0.04) \times 10^8. \quad (\text{B.39})$$

A particularly encouraging result is obtained when R_μ is left free in the Greisen fit. In that case one finds

$$N = (2.5 \pm 0.2) \times 10^8, \quad R_\mu = (277 \pm 10) \text{ m}, \quad (\text{B.40})$$

which is essentially identical to the value inferred from Eq. (6.7). Thus, although the 320 m choice is less disfavoured here than for the UMD-433, the profile still prefers the value expected from the atmospheric-depth scaling. In other words, the geometrical scale extracted directly from the UMD-750 data agrees with the simple depth-based expectation at the level of the fit uncertainty.

The same conclusion is reached in the reference-distance form of Eq. (6.13). Fixing $R_\mu = 278$ m and $\alpha = 0.75$, and fitting the density at $r_{\text{ref}} = 450$ m together with the second exponent, yields

$$\rho(450) = 0.822 \pm 0.015, \quad \beta = 2.53 \pm 0.05. \quad (\text{B.41})$$

The fitted slope is therefore compatible, within the fit uncertainty, with the nominal Greisen value $\beta = 2.5$. As in the UMD-433 case, this indicates that once the physically motivated value of R_μ is adopted, the mean UMD-750 profile does not require any appreciable deformation of the standard two-factor shape **over the radial interval examined here**.

The corresponding scatter plots are shown in Fig. B.13. The same qualitative behaviour seen for the UMD-433 is present here as well. In the left panel, the energy-rescaled density

$$\tilde{\rho}_\mu \equiv \frac{\rho_\mu}{(E_{\text{rec}}/10^{17.5} \text{ eV})^{0.9}} \quad (\text{B.42})$$

again collapses the main trend into a compact ridge around the fitted curve. In the right panel, the raw muon density, colour-coded by $\sin^2 \theta_{\text{rec}}$, shows the familiar horizontal banding at low densities associated with the discrete character of the reconstructed muon counts.

The comparison with historical measurements leads to the same overall conclusion as in the main text. Since the mean UMD-750 profile is again very close to the Greisen shape, the shifted comparisons with earlier experiments follow the Greisen-based expectations rather closely as well. For that reason, and to avoid repeating the same line of argument, the corresponding comparison plots are not reproduced here. The main point is unchanged: once the correct geometrical value of R_μ is used, the mean UMD-750 shape also appears consistent with the historical benchmark.

A final caveat concerns the radial range explored here. In the present UMD-750 check, densities beyond 800 m have not been examined explicitly. This is potentially more relevant for the UMD-750 than for the UMD-433, since the higher energies probed by the former naturally extend the MLDF into a regime of larger core distances. In that region, the simple two-factor Greisen form may no longer provide a fully satisfactory description of the mean shape, and an additional large-distance term, analogous to the third factor used in the KASCADE-Grande parametrisation, may become necessary. This was the case for previous UMD studies [120, 125, 126]. More generally, the Greisen shape itself may cease to be optimal once the distribution is followed well beyond 800 m, as already suggested by the known discrepancy with AGASA in that regime [147]. In that case, and following the recommendation made in [147], a reasonable compromise would be to compare future UMD-750 measurements at large distances with an average of the AGASA and Yakutsk MLDFs [149, 204], rather than relying on the simple Greisen form alone.

Taken together, these checks show that the picture established in the main text is not peculiar to the UMD-433. Even in the broader zenith range and at higher energies relevant for the UMD-750, the mean profile remains well described by the reduced Greisen form, the preferred value of R_μ agrees with the atmospheric-depth expectation, and the fitted β remains compatible with the nominal one. This provides an additional and independent validation of the usefulness of the Greisen parametrisation as a simple and physically transparent reference description of the mean muon lateral distribution.

B.4 ρ_μ fluctuations

In the SD-433 energy range the expected muon densities are low, and the spread of ρ_μ is therefore dominated by Poissonian counting fluctuations (see Fig. B.14); in contrast, at higher muon statistics (closer to the shower axis) intrinsic shower-to-shower fluctuations become

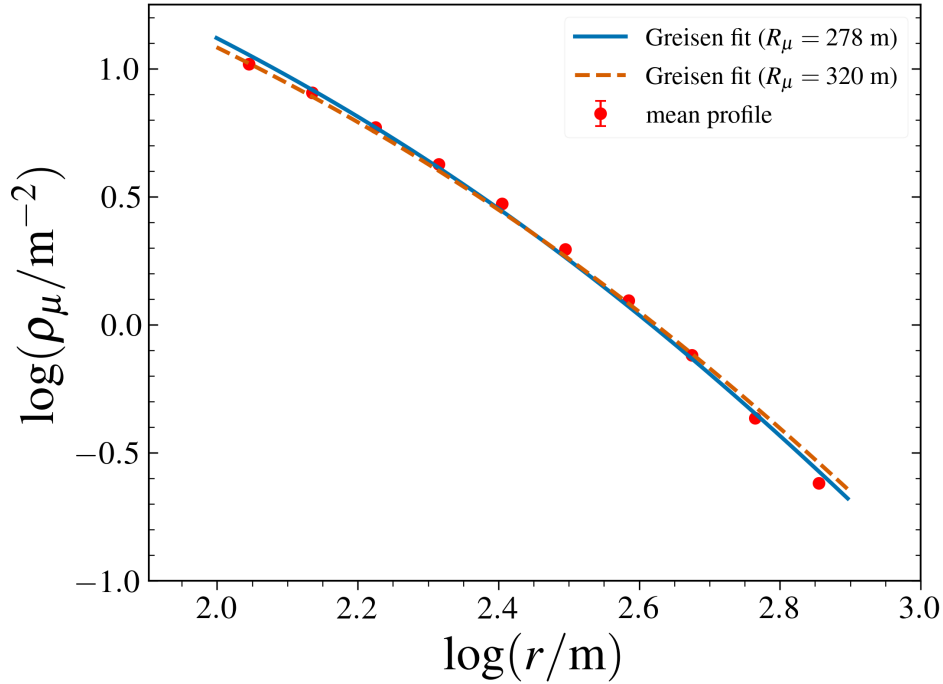


Figure B.12: Mean UMD-750 muon density profile as a function of core distance together with two Greisen fits, one with fixed $R_\mu = 278$ m and one with fixed $R_\mu = 320$ m.

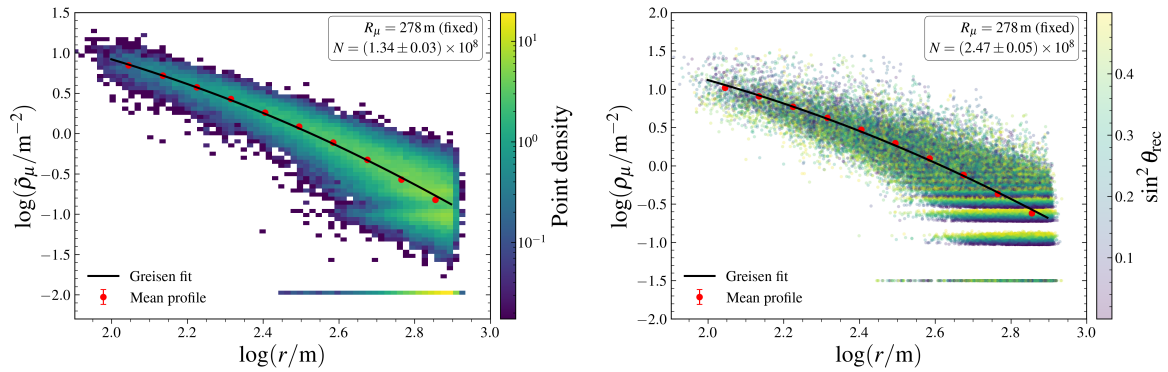


Figure B.13: Two complementary views of the UMD-750 sample used for the mean-shape validation. *Left:* two-dimensional histogram of the energy-rescaled muon density $\tilde{\rho}_\mu = \rho_\mu / (E_{\text{rec}}/10^{17.5} \text{ eV})^{0.9}$ as a function of shower-plane distance, with the mean profile and the Greisen fit for fixed $R_\mu = 278$ m superimposed. *Right:* event-by-event scatter plot of the raw muon density ρ_μ , colour-coded by $\sin^2 \theta_{\text{rec}}$, again with the mean profile and the same Greisen fit overlaid.

comparatively more important. As a consequence, the minima of $\sigma_{\rho_\mu} / \langle \rho_\mu \rangle$ are concentrated at small r , close to the shower axis, and shift to larger distances with increasing energy, following a trend similar to that observed for r_{opt} .

B.5 Alternative β parametrisation

The inferred muon content is sensitive to the adopted energy scale. Over the past few years, the energy calibration of the different arrays has been updated repeatedly; in that situation, the β parametrisation must in principle be refitted after each change. A formulation

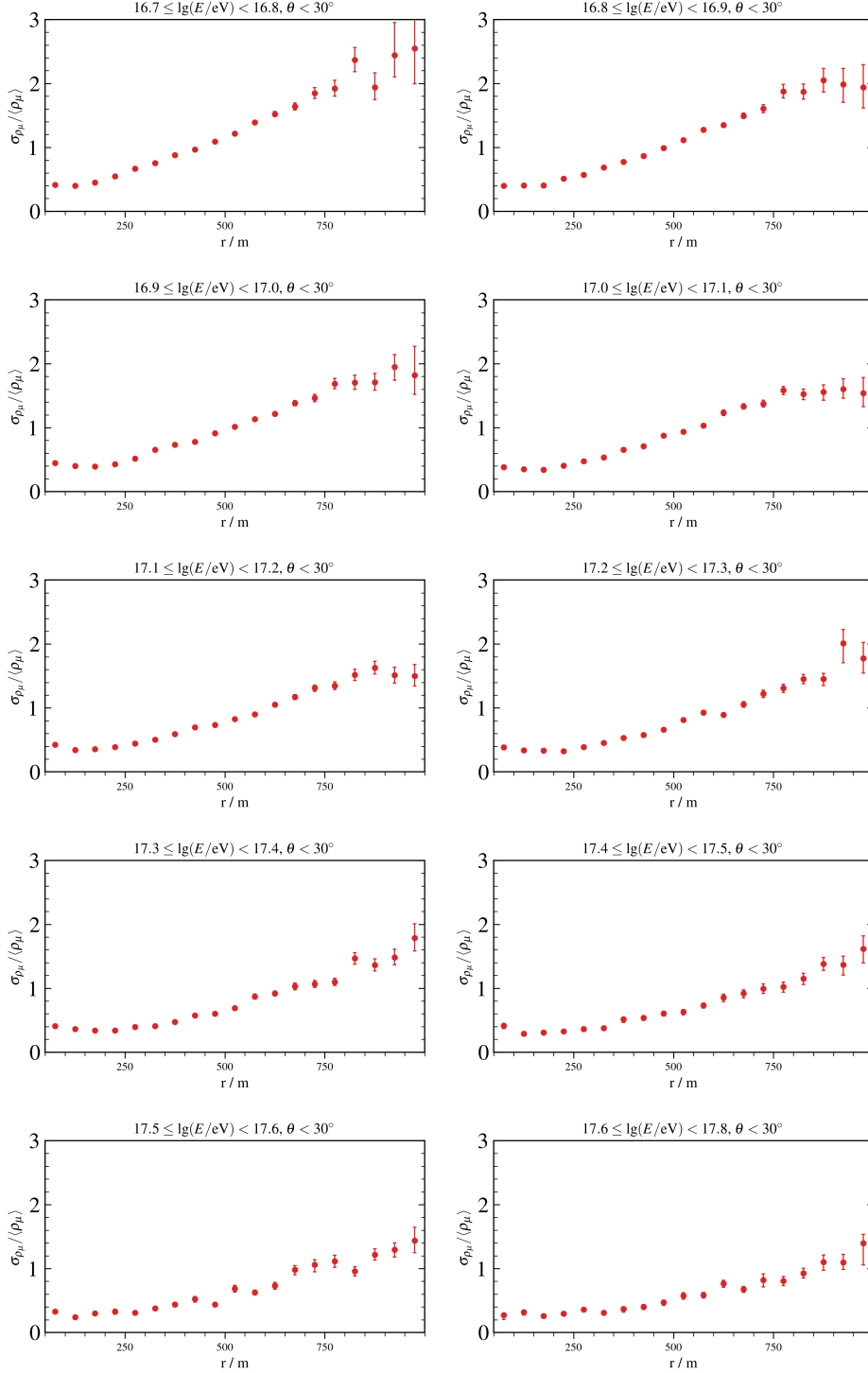


Figure B.14: Relative fluctuations of the reconstructed muon-density estimator, $\sigma_{\rho_\mu} / \langle \rho_\mu \rangle$, as a function of shower-plane distance r for the reduced-zenith phase space ($\theta < 30^\circ$), shown in bins of $\lg(E/\text{eV})$. Red markers indicate the mean relative fluctuation, with error bars reflecting the statistical uncertainty from bootstrapping.

based on $S(300)$ is therefore more stable over time⁷. This appendix summarises the shower-size-dependent slope parametrisation

$$\hat{\beta}(S(300), \theta) = \beta_0 + \beta_{SS} \left[\lg \left(\frac{S(300)}{\text{VEM}} \right) - \lg \left(\frac{S_{\text{ref}}}{\text{VEM}} \right) \right] + \beta_\theta (\sec \theta - \sec \theta_0). \quad (\text{B.43})$$

⁷We have implemented the option to use this parametrisation in Offline.

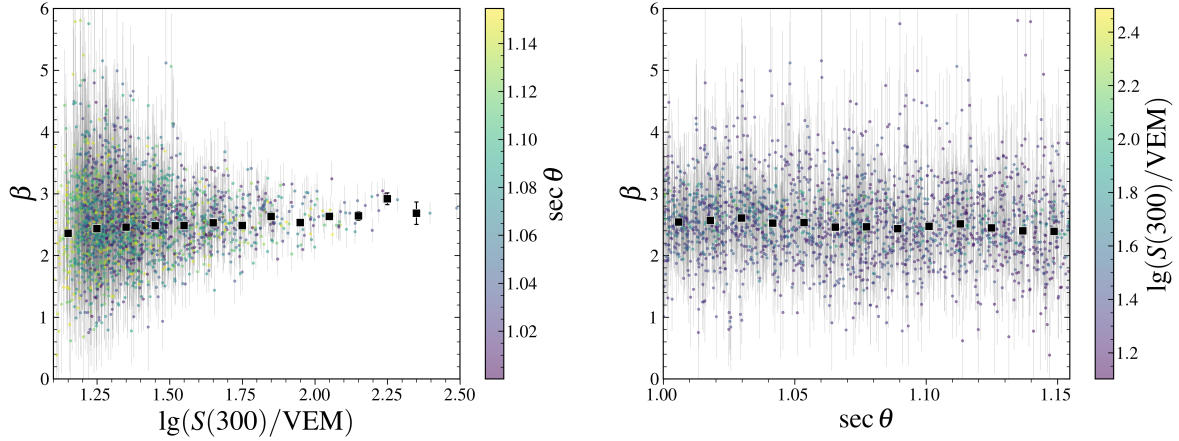


Figure B.15: Event-level MLDF slope β obtained from free-shape fits (with α and γ fixed) in the UMD-433 high-quality sample used for the parameterisation. *Left:* β versus $\lg\left(\frac{S(300)}{\text{VEM}}\right)$, coloured by $\sec\theta$. *Right:* β versus $\sec\theta$, coloured by $\lg\left(\frac{S(300)}{\text{VEM}}\right)$. Thin grey vertical bars indicate the fit uncertainty σ_β for each event. Black squares show binned, uncertainty-weighted averages, illustrating the mild mean dependence that is captured by Eq. (B.43).

Table B.6 presents the fitted parameters, while Figure B.15 shows the trends. Finally, Figure B.16 shows the expected bias and resolution of the slope estimator.

Table B.6: Parameters of the alternative UMD-433 slope parameterisation Eq. (6.20) for $r_{\text{ref}} = 250$ m in the reduced-zenith phase space.

Quantity	Value
$\lg(S_{\text{ref}}/\text{VEM})$	1.37
$\sec\theta_0$	1.07
β_0	2.45 ± 0.07
β_{SS}	-1.28 ± 0.10
β_θ	0.22 ± 0.10

B.6 Attenuation correction

We adopt the factorisation and quadratic attenuation model of Eqs. (6.25) and (6.26). In practice, the CIC procedure from Ref. [143] is repeated for a range of reference intensities, equivalently for a range of reference cuts $\rho_{21\text{-cut}}(\theta_{\text{ref}})$ defined in a narrow band around $\theta_{\text{ref}} = 21^\circ$. For each fixed cut we fit (a, b) by minimising an Anderson–Darling (AD) uniformity statistic computed on the CIC-selected sample in $x \equiv \sin^2\theta / \sin^2\theta_{\text{max}}$ (uniform for an isotropic flux). The scan is used both as a robustness check and as an empirical way to estimate the sensitivity of the fit to the chosen intensity. Figure B.17 summarises the results from this exercise.

While a is stable across the scan, b is weakly constrained and can exhibit large excursions and inflated uncertainties outside the stable region. This behaviour reflects the limited leverage of the quadratic term in the restricted zenith range ($\theta \leq 30^\circ$, for which $|x| \lesssim 0.13$ and $x^2 \lesssim 1.5 \times 10^{-2}$), so that the $b \cdot x^2$ contribution is much smaller than the linear $a \cdot x$ term.

The selected solution corresponds to $\rho_{21\text{-cut}}(\theta_{\text{ref}}) \sim 0.485$, yielding the parameter values reported in Eq. (6.27). Because b is weakly constrained by the AD criterion and partially

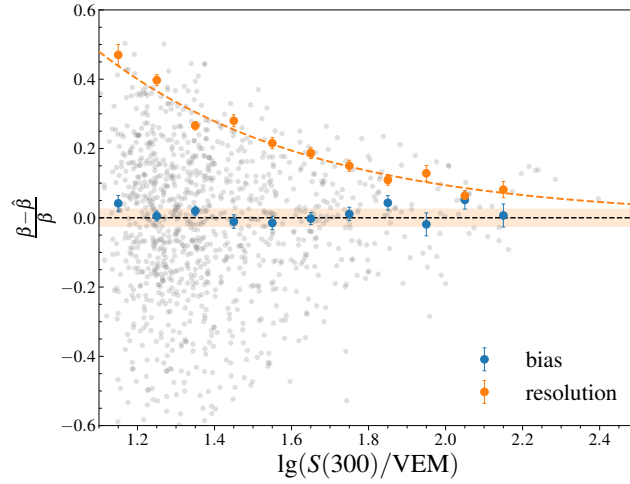


Figure B.16: Relative residuals of the slope parameterisation, $\delta_\beta = (\beta - \hat{\beta})/\beta$ (grey points), as a function of $\lg\left(\frac{S(300)}{\text{VEM}}\right)$. Blue markers show the binned bias (mean of δ_β), while orange markers show the binned resolution (standard deviation of δ_β). The dashed curve is a fit of the resolution with the empirical model analogous to Eq. (6.21).

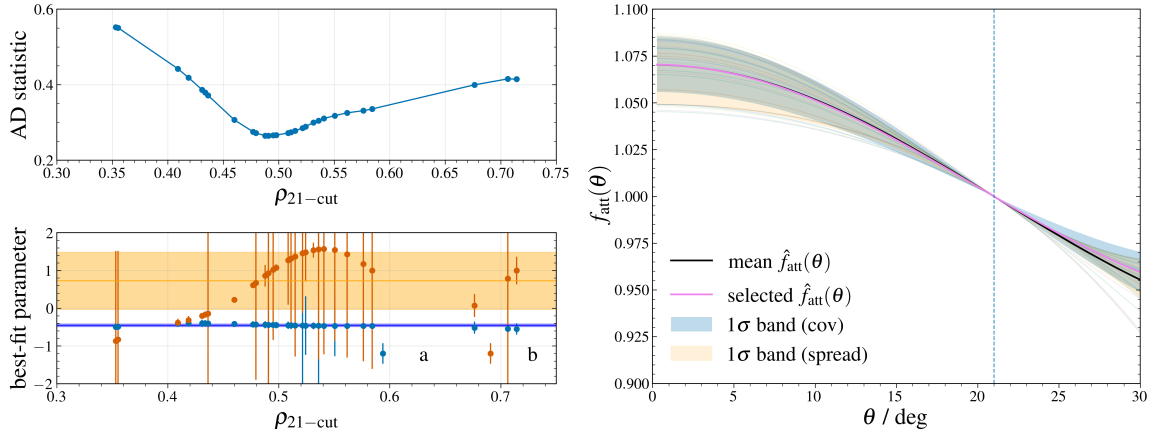


Figure B.17: CIC scan and uncertainty budget for the attenuation function. *Left:* results of fitting the attenuation model at fixed values of the reference cut $\rho_{21\text{-cut}}(\theta_{\text{ref}})$. The top panel shows the minimised Anderson–Darling (AD) statistic as a function of the cut, illustrating a broad “stable” region where the AD statistic remains approximately flat. The bottom panel shows the corresponding best-fit parameters a (blue) and b (orange), with their statistical uncertainties. The color bands reflect the uncertainties of associated parameters in the studied range. *Right:* ensemble of attenuation curves obtained from the scan. The black curve is the mean attenuation curve, and the magenta curve is the selected solution adopted in the analysis. The blue band shows the 1σ uncertainty propagated from the covariance of the selected (a, b) , while the orange band represents the spread across the scan. Over the full angular range considered, the mean and selected curves differ by less than 0.6%.

degenerate with the overall selection threshold, its impact on f_{att} is nevertheless limited compared to the a term in the SD–433 zenith range.

Finally, Fig. B.18 shows two complementary validation plots for the selected attenuation curve. The left panel illustrates how the fitted θ -dependent CIC threshold overlays the $(\sin^2 \theta, \lg \rho(250))$ event distribution, and the right panel shows that the empirical CDF of

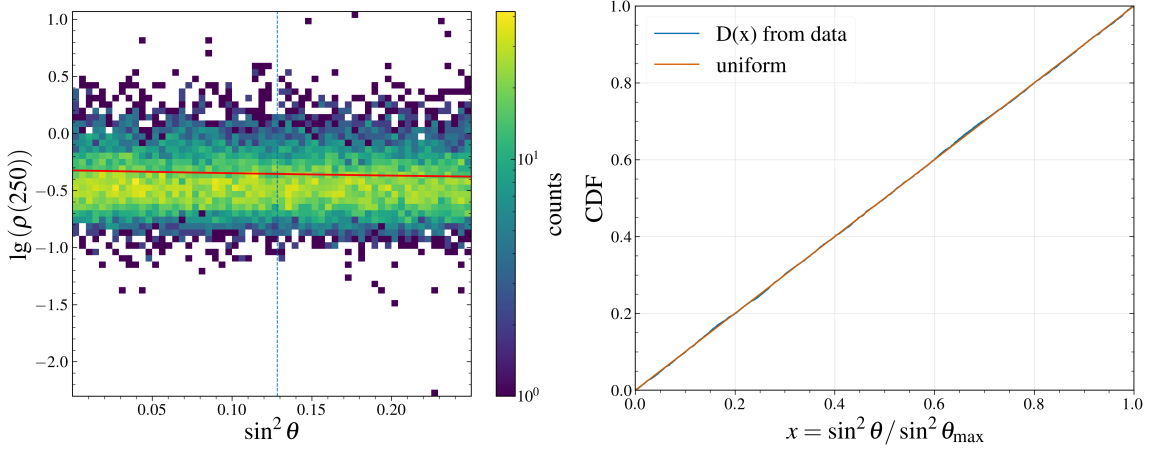


Figure B.18: Validation of the CIC attenuation correction for $\rho(250)$. *Left:* two-dimensional distribution of events in $(\sin^2 \theta, \lg \rho(250))$ (color scale in counts). The vertical dashed line marks $\sin^2 \theta_{\text{ref}}$ with $\theta_{\text{ref}} = 21^\circ$. The red curve shows the θ -dependent CIC threshold corresponding to the selected solution, i.e. the cut $\rho(250) = \rho_{21\text{-cut}}(\theta_{\text{ref}}) f_{\text{att}}(\theta)$. *Right:* empirical cumulative distribution function $D(x)$ of $x = \sin^2 \theta / \sin^2 \theta_{\text{max}}$ for the CIC-selected sample (blue), compared to the uniform expectation (orange). The close agreement indicates that the post-selection zenith distribution is compatible with isotropy within the analysed range.

$x = \sin^2 \theta / \sin^2 \theta_{\text{max}}$ in the selected sample closely matches the uniform expectation, as required by isotropy.

B.7 Systematic uncertainties

Several works have quantified the systematic uncertainties affecting the muon density measured with the UMD [120, 125, 126, 205]. In this work we retain the dominant sources and report their net contributions to ρ_{21} . Each contribution is evaluated through a dedicated one-at-a-time variation of the analysis, yielding an alternative estimate ρ'_{21} . We then compute the relative shift with respect to the baseline value, $(\rho'_{21} - \rho_{21}) / \rho_{21}$. Because a given variation can increase or decrease the reconstructed density, the resulting systematic uncertainty is, in general, asymmetric; we therefore quote separate upward and downward components.

The systematic uncertainties on ρ_{21} are summarised in Fig. B.19. The dominant contributions arise from the uncertainty in the energy scale and from the uncertainty in the MLDF shape (i.e., the slope choice). The energy-scale contribution is obtained by shifting the primary-energy estimate by $\pm 14\%$, while the MLDF-shape uncertainty follows the procedure described in Ref. [153]. The long-term contribution quantifies residual time-dependent effects (e.g., detector ageing), and the muon-pattern contribution is obtained by varying the muon pattern used in the infinite-window strategy; both are discussed in Ref. [126]. The resulting uncertainty is comparable to that obtained for the observable ρ_{35} in the UMD-750 analyses.

B.8 Effect of the folding correction for $\ln\langle A \rangle$ and the z-scale

The composition estimates presented in the main text use proton and iron reference baselines folded with the cosmic-ray spectrum, trigger efficiency, and detector response. This appendix shows the corresponding results obtained when this folding correction is omitted.

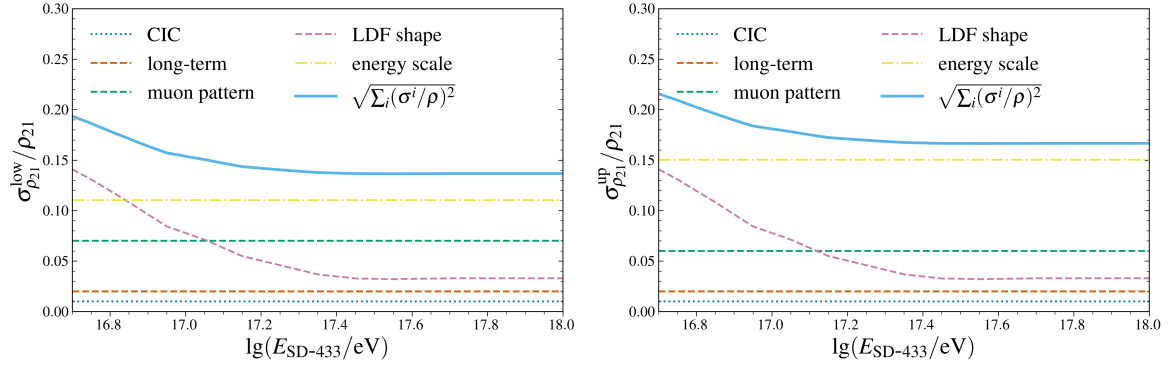


Figure B.19: Relative systematic uncertainties on ρ_{21} as a function of $\lg(E_{\text{SD-433}}/\text{eV})$. The left (right) panel shows the downward $\sigma_{\rho_{21}}^{\text{low}}/\rho_{21}$ (upward $\sigma_{\rho_{21}}^{\text{up}}/\rho_{21}$) contributions from one-at-a-time variations of the analysis: the attenuation correction (CIC), long-term stability (long-term), the muon-pattern choice in the infinite-window strategy (muon pattern), the MLDF-shape uncertainty (LDF shape), and the energy-scale uncertainty (energy scale). The solid line shows the total uncertainty computed as the quadratic sum of the individual contributions, $\sqrt{\sum_i (\sigma^i/\rho_{21})^2}$, evaluated separately for the upward and downward sides.

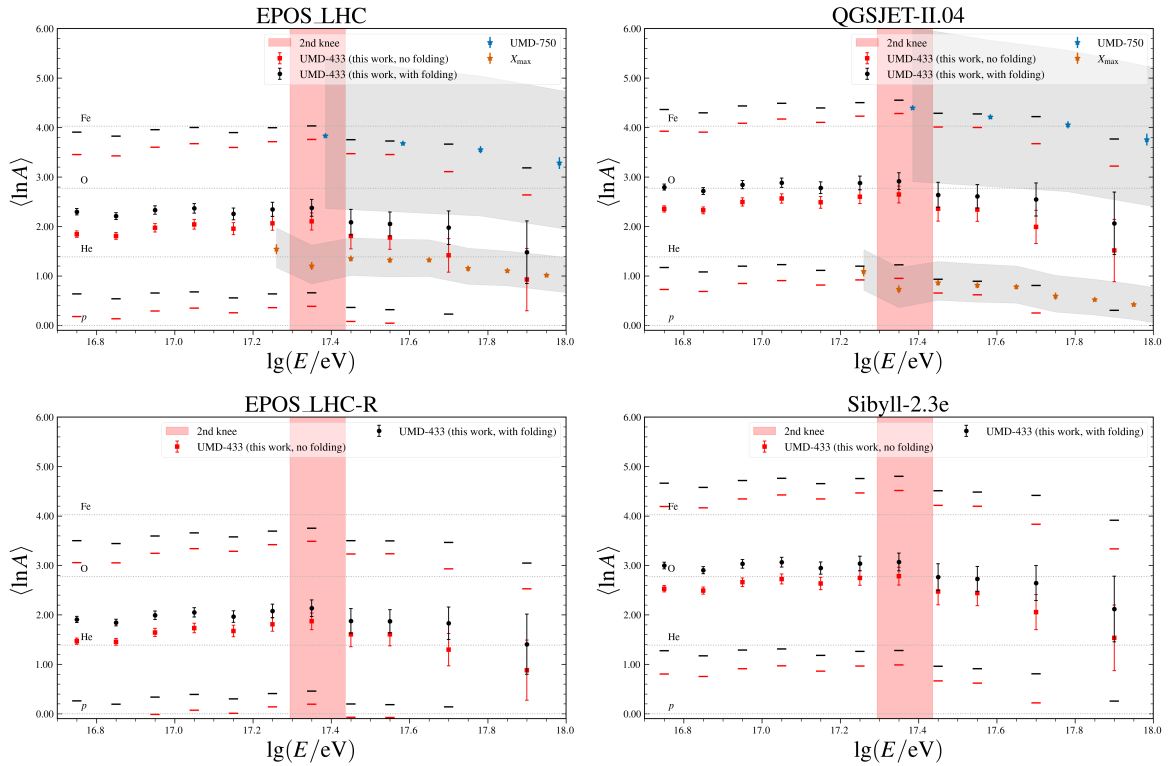


Figure B.20: Analogous to Fig. 6.17, with the addition of the UMD-433 results obtained without the folding correction.

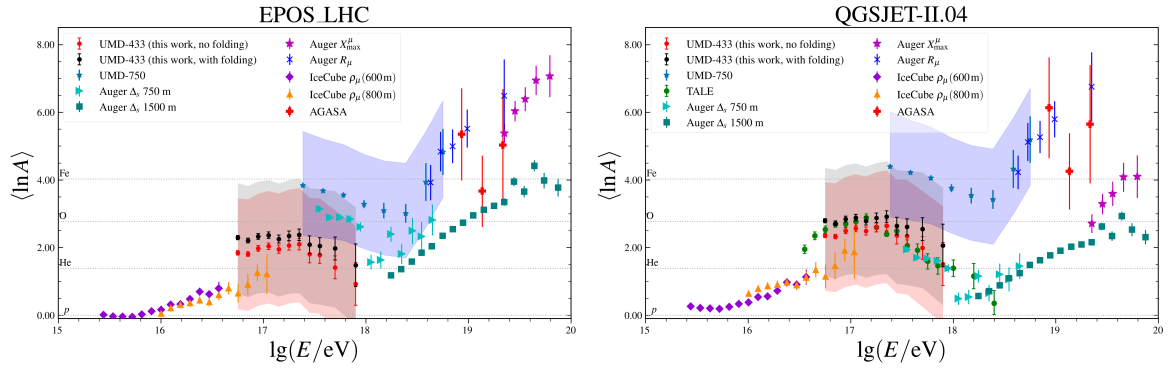


Figure B.21: Analogous to Fig. 6.18, with the addition of the UMD-433 results obtained without the folding correction.

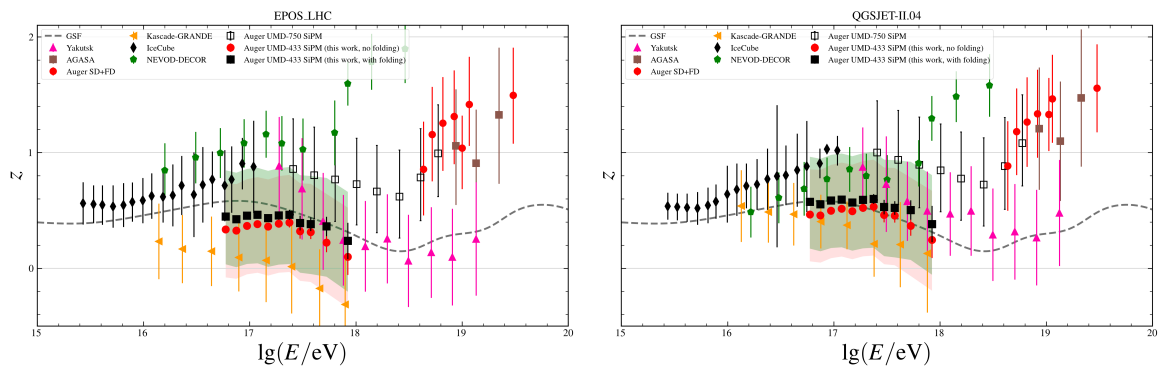


Figure B.22: Analogous to Fig. 6.19, with the addition of the UMD-433 results obtained without the folding correction.

Statement on AI Use

I declare that the scientific content of this thesis is my own work. The research questions, methodology, data analysis, figures, interpretation of results, conclusions, and selection and verification of references were carried out by me.

This declaration was prepared taking into account the European Commission's *Living guidelines on the responsible use of generative AI in research*, in particular their emphasis on transparency, human responsibility, and accountability when using generative AI tools in research [206].

During the preparation of this thesis, I used generative artificial-intelligence tools, in particular ChatGPT by OpenAI and Claude by Anthropic [207, 208], as editorial support tools. Their use was limited to linguistic and presentational assistance, including grammar and spelling correction, improvement of text flow and readability, suggestions for alternative wording, translation or polishing of self-written passages, and identification of typographical or stylistic inconsistencies.

These tools were not used to generate original scientific results, to perform or replace the data analysis, to draw conclusions, to fabricate or modify data, or to provide unverified references. All AI-assisted text was reviewed, edited, and approved by me. I remain fully responsible for the accuracy, integrity, and proper citation of the content of this thesis.