The muon content of atmospheric air showers and the mass composition of cosmic rays

For the attainment of the academic degree of

Doctorate in Science
from the
Karlsruher Institut für Technologie (KIT)
and the
Universidad Nacional de San Martín (UNSAM)

accepted

Dissertation

of

Lic. Flavia Gesualdi
of Ciudad Autónoma de Buenos Aires

Day of the oral examination: December 15th, 2022

Referee: Prof. Dr. Alberto Daniel Supanitsky

Co-referee: Prof. Dr. Ralph Engel

Supervisor: Dr. Markus Roth

The muon content of atmospheric air showers and the mass composition of cosmic rays

Tesis aceptada 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 Institut für Technologie (KIT)

por

Lic. Flavia Gesualdi
de Ciudad Autónoma de Buenos Aires

Fecha de la defensa oral: 15 de Diciembre de 2022

Director: Prof. Dr. Alberto Daniel Supanitsky

Co-director: Prof. Dr. Ralph Engel

Colaborador: Dr. Markus Roth

The muon content of atmospheric air showers and the mass composition of cosmic rays

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

von der KIT-Fakultät für Physik des
Karlsruher Instituts für Technologie (KIT)

und der

Universidad Nacional de San Martín (UNSAM)

genehmigte

Dissertation

von

Lic. Flavia Gesualdi

aus Ciudad Autónoma de Buenos Aires

Tag der mündlichen Prüfung: 15. Dezember 2022

Referent: Prof. Dr. Alberto Daniel Supanitsky

Korreferent: Prof. Dr. Ralph Engel

Betreuer: Dr. Markus Roth

Abstract

Cosmic rays are messengers from outer space holding the answer to the Universe's deepest mysteries: What are they exactly? Where do they come from? How do they accelerate to such energies, and how are the laws governing their interactions? At the highest energies, these unubiquitous particles can only be detected through the showers of secondary particles that they generate in their interactions with the Earth's atmosphere. Large ground-based observatories, like the Pierre Auger Observatory, detect these extensive air showers and attempt to reconstruct as much information of the primary cosmic ray as possible. In particular, the number of secondary muons is a key observable because it is directly related to the atomic mass number of the primary that generated them. Understanding the mass composition as a function of the energy would shed light on various open questions strongly linked to the origin of cosmic rays. The Pierre Auger Observatory has dedicated scintillators buried underground to directly detect muons, the Underground Muon Detector (UMD).

This thesis is devoted to the accurate determination of the muon content of air showers and to the study of its composition implications. We analyze direct muon measurements of air showers with energies between $10^{17.22}$ eV and $10^{19.46}$ eV from two experiments. At lower energies we extensively analyze UMD data, which we complement at higher energies with measurements from the Akeno Giant Air Shower Array (AGASA). For the UMD data, we develop new methods that significantly improve the estimation of the muon number. These methods also allow for the reconstruction of the muon signal as a function of time with an unprecedented time resolution, opening the door to the reconstruction of new composition-sensitive observables. As a direct application, we study the measured lateral distribution of muons and the models that attempt to describe it.

Furthermore, we analyze the mass composition implications of the UMD and AGASA data. The composition interpretation of the data can only be inferred by a comparison against air-shower simulations. We therefore simulate single-proton, single-iron, and mixed composition scenarios based on the three newest-generation high-energy hadronic interaction models. To better compare the results against those of other experiments, we compute the so called z-values, a scale of the muon content in data relative to that of proton and iron simulations. The combined results offer a picture consistent with other experiments: the unexpectedly heavy composition constitutes evidence of a muon deficit in air-shower simulations that increases with the energy. These results can help to improve the high-energy hadronic interaction models, which in turn would improve the precision of the inferred mass composition.

Resumen

Los rayos cósmicos son mensajeros del espacio que contienen la respuesta a los misterios más profundos del Universo: ¿Qué son exactamente? ¿De dónde vienen? ¿Cómo se aceleran hasta tales energías, y cómo son las leyes que rigen sus interacciones? A las energías más altas, estas escasas partículas sólo pueden detectarse a través de las lluvias de partículas secundarias que son generadas en sus interacciones con la atmósfera terrestre. Grandes observatorios, como el Observatorio Pierre Auger, detectan estas lluvias atmosféricas e intentan reconstruir la mayor información posible del rayo cósmico primario. En particular, el número de muones secundarios es un observable clave porque está directamente relacionado con la composición química del primario que los generó. La comprensión de la composición química en función de la energía arrojaría luz sobre diversas preguntas no resueltas fuertemente vinculadas al origen de los rayos cósmicos. El Observatorio Pierre Auger dispone de detectores de centelleo bajo tierra para detectar los muones de forma directa, el "Underground Muon Detector" (UMD).

Esta tesis está dedicada a la determinación precisa del contenido de muones de las lluvias atmosféricas y al estudio de sus implicaciones sobre la composición química del primario. Analizamos mediciones directas de muones de lluvias atmosféricas con energías comprendidas entre $10^{17,22}$ eV y $10^{19,46}$ eV realizadas por dos experimentos. A las energías más bajas analizamos extensamente los datos del UMD, que complementamos a las energías más altas con mediciones de "Akeno Giant Air Shower Array" (AGASA). Para los datos del UMD, desarrollamos nuevos métodos que mejoran significativamente la estimación del número de muones. Estos métodos también permiten la reconstrucción de la señal de muones en función del tiempo con una resolución temporal sin precedentes, abriendo la puerta a la reconstrucción de nuevos observables sensibles a la composición química. Como aplicación, estudiamos la distribución lateral de muones medida con el UMD, así como los modelos que intentan describirla.

Además, analizamos las implicaciones de los datos del UMD y de AGASA con respecto a la composición química. La interpretación de la composición de los datos sólo puede inferirse mediante una comparación contra simulaciones de lluvias atmosféricas. Por lo tanto, simulamos escenarios de protón puro, hierro puro y de composición mixta, basados en los tres modelos de última generación de las interacciones hadrónicas de alta energía. Para comparar mejor los resultados con los de otros experimentos, calculamos los llamados valores z, una escala del contenido de muones en los datos respecto del de simulaciones de protón y hierro. Los resultados combinados ofrecen una imagen coherente con la de otros experimentos: la composición inesperadamente pesada constituye evidencia de un déficit de muones en las simulaciones de lluvias atmosféricas que aumenta con la energía. Estos resultados pueden ayudar a mejorar los modelos de interacciones hadrónicas de altas energías, lo que a su vez mejoraría la precisión de las inferencias de composición química.

Zusammenfassung

Die Teilchen der kosmischen Strahlung sind Boten aus dem Weltraum, die die Antworten auf die tiefsten Geheimnisse des Universums enthalten: Was sind sie? Woher kommen sie? Wie werden sie auf solche Energien beschleunigt, und was sind die Gesetze, die ihre Wechselwirkungen bestimmen? Bei den höchsten Energien können diese seltenen Teilchen nur durch die Schauer von Sekundärteilchen nachgewiesen werden, die sie in der Erdatmosphäre erzeugen. Große bodengestützte Observatorien wie das Pierre-Auger-Observatorium erfassen diese Luftschauer und versuchen, so viele Informationen über das Primärteilchen wie möglich zu rekonstruieren. Insbesondere die Anzahl der Myonen ist eine wichtige Größe, da sie direkt mit der Masse des Primärteilchens zusammenhängt. Ein Verständnis der Massenzusammensetzung in Abhängigkeit der Energie würde Aufschluss über offene Fragen geben, die eng mit dem Ursprung der kosmischen Strahlung zusammenhängen. Das Pierre-Auger-Observatorium verfügt zur direkten Messung der Myonen über vergrabene Szintillatoren, den "Underground Muon Detector" (UMD).

Diese Arbeit widmet sich der genauen Bestimmung des Myonengehalts von Luftschauern und den Implikationen desselben für die chemische Zusammensetzung der kosmischen Strahlung. Wir analysieren direkte Myonenmessungen von Luftschauern mit Energien zwischen $10^{17,22}$ eV und $10^{19,46}$ eV von zwei Observatorien. Bei niedrigeren Energien analysieren wir detailliert UMD-Daten, die wir bei höheren Energien durch Messungen des "Akeno Giant Air Shower Array" (AGASA) ergänzen. Für die UMD-Daten entwickeln wir neue Methoden, die die Präzision der gemessenen Myonenzahl deutlich verbessern. Diese ermöglichen auch die Rekonstruktion des Myonensignals als Funktion der Zeit mit einer noch nie dagewesenen Zeitauflösung, was die Rekonstruktion neuer sensitiveren Messgrößen ermöglicht. Als direkte Anwendung untersuchen wir die gemessene laterale Verteilung von Myonen und vergleichen sie mit Modellen.

Außerdem analysieren wir die Implikationen der UMD- und AGASA-Daten für die Massenzusammensetzung. Die Interpretation der Daten hinsichtlich der chemischen Zusammensetzung lässt sich nur durch einen Vergleich mit Simulationen ableiten. Wir simulieren daher Szenarien mit Protonen, Eisenkernen und einer gemischten Zusammensetzung auf der Grundlage der drei Modelle für hochenergetische hadronische Wechselwirkungen der neuesten Generation. Um die Ergebnisse besser mit denen anderer Experimente vergleichen zu können, berechnen wir die so genannten z-Werte, die eine Skala des Myonengehalts in den Daten im Verhältnis zu dem von Protonen- und Eisensimulationen bilden. Die kombinierten Ergebnisse ergeben stimmen mit denen anderer Experimente überein: Die unerwartet schwere chemische Zusammensetzung ist ein Beweis für ein Myonendefizit in Luftschauer-Simulationen, das mit der Energie zunimmt. Diese Ergebnisse können dazu beitragen, die Modelle der hadronischen Wechselwirkung zu verbessern, was wiederum die Genauigkeit der abgeleiteten Massenzusammensetzung erhöhen würde.

Contents

Chapter 1

Introduction

Cosmic rays are particles that arrive from outer space carrying incredibly high, macroscopic energies, orders of magnitude beyond what can be achieved in human-made accelerators. Despite decades of extensive research, the nature, sources, and acceleration mechanisms of cosmic rays are still not clear. To unravel these mysteries, cosmic rays are studied mainly through their mass, energy spectrum, and arrival directions. The spectrum drops steeply with energy, thus cosmic rays with energies above $\sim 10^{15}$ eV can only be detected indirectly. For this, large ground-based observatories study the cascades of secondary particles, known as extensive air showers, that the ultra-high energy cosmic rays produce in their interactions with the atmosphere. Among the various kinds of secondary particles, muons are of special relevance. Since they almost always reach ground before decaying, they are tracers of the hadronic interactions, which are the "backbone" of the air shower. The number of muons of an air shower is strongly related to the composition of the cosmic ray that originated it: lighter primaries produce less secondary muons than heavier primaries.

Knowing the mass composition as a function of the energy is essential to understanding key aspects of cosmic rays. Identifying their mass could help in determining their sources [1], in understanding the transition energy between galactic and extragalactic cosmic rays [2], or in ruling out models of the flux suppression at the highest energies [3] (more details in Chapter 2). Currently, composition identification is hampered by the limitations of the high-energy hadronic interaction models that have to be used to interpret the data. These models necessarily extrapolate lower-energy accelerator data to the ultra-high energies of cosmic rays. To test these models, the consistency of the composition implications of different observables is typically analyzed. There are two main composition-sensitive observables: the depth-of-the-shower maximum X_{max} and, as already mentioned, the number of muons N_{μ}. The former, X_{max}, is typically subject to smaller systematic uncertainties [4], and is thus taken as a reference in composition interpretations. If two observables offer inconsistent interpretations under a same model, it is understood that there is a problem in the model.

A muon deficit in air-shower simulations increasing with the energy ($\gtrsim 10^{16}$ eV) was reported by several experiments [5]. At ultra-high energies, the muon deficit can be considerably large. For example, the Pierre Auger Collaboration reports a deficit of 30 % to 80 % at 10^{19} eV [6]. It is still not known whether the muon deficit originates from new physics at the highest energies or by a partial mismodelling of hadronic collisions [7]. Characterizing the muon deficit would help to understand the aspects of hadronic interactions that can cause these deviations [5]. In this sense, composition studies can help to improve the high-energy hadronic interaction models, which in turn improve the precision of the inferred mass composition [8, 9].

Muons can be detected directly or indirectly. Indirect detection can be achieved by combining the responses of detectors with different relative sensitivity to the different air-shower components [10]. Another method of indirect detection relies on inclined events [8]. In contrast, direct detection can be achieved by using shielding material or by burying the detectors underground. For the latter, and for inclined events, most of the electromagnetic component of air showers is absorbed by the shielding material or by the atmosphere respectively, leaving essentially the muonic component to be detected. Among the different kinds of particle detectors, segmented scintillation detectors are the most widely used for muon detection. These kind of detectors can be designed to operate as calorimeters or as counters. For the former, a number of muons is estimated by transforming the integral of the output signal to a total deposited energy, which is then converted to a number of particles by knowing the energy deposited by a single particle. For the latter, the output signal of a segment is compared against a threshold, and the resulting binary signal is matched to a known muon signal. In this work, we use data from two kinds of segmented detectors that achieve direct muon detection: most prominently, from the Underground Muon Detector (UMD) of the Pierre Auger Observatory [1, 11] (described extensively in Chapter 3), but also from the muon detectors of the Akeno Giant Air Shower Array (AGASA) [12].

In this work, we study the muon content of air showers and its composition implications at energies between $10^{17.22}$ eV and $10^{19.46}$ eV using UMD and AGASA data. In order to have a high-quality UMD data set, we develop dedicated methods to accurately determine the number of muons. We show that the statistical problem of counting muons in a segmented detector is inherently difficult: If not considered in the design of the counting strategy, time-unresolved particles can cause severe under-counting biases [13, 14]. We explain how the problem of counting muons on a scintillation detector is equivalent to the problem of counting "balls in boxes". Applying this idea in a smart way, we develop a new method that exploits the full time resolution of the detector. With possible applications beyond muon counting, a great advantage of the presented strategy is that it provides estimates of the number of muons as a function of time as seen by the detector to a single time-bin resolution. In the future, this could be used to reconstruct another prominent composition-sensitive observable: the maximum of the muon production depth distribution $X^{\mu}_{\max}$. Moreover, to obtain an unbiased estimate of the number of muons, muons that transverse two neighbor scintillation bars have to be considered too. These tend to deposit enough signal in both scintillation bars so as to be counted twice, causing an over-counting bias. With detailed detector simulations, we parameterize the overall bias as a function of air-shower observables, including the reconstructed number of muons itself. In turn, we use the parameterization for correcting the reconstructed number of muons, thus obtaining reliable estimates accurate to $\pm 4\,\%$.

Having an unbiased estimate of the number of muons from UMD data, we continue with a study of the muon lateral distribution function (muon LDF) as measured by the UMD. The muon LDF is interesting in itself because it holds composition information both in its size and shape. We analyze the goodness-of-fit of different muon LDF models and provide parameterizations of the measured muon LDF. These can be useful for low-multiplicity event-wise fits, for building toy models, or as a reference for the design of future muon detectors. Furthermore, we analyze the composition interpretation from the muon content in UMD data by comparing it against single-proton and single-iron simulations, as well as against mixed composition scenarios derived from the fits to the $X_{\max}$ distributions. We study the evolution of the muon content as a function of the distance to the shower axis, of the zenith angle, and of the energy (between $10^{17.22}$ eV and $10^{18.42}$ eV), and we also compute the so-called z-values [5], a muon scale of data relative to simulations that enables comparisons between different experiments. To complement these results with measurements at higher energies (between $10^{18.83}$ eV and $10^{19.46}$ eV), we use the muon measurements performed by the AGASA collaboration. Finally, we put in perspective the results of this work by comparing them to further estimations from nine different analyses of six experiments [5, 15].

Chapter 2

Cosmic rays and extensive air showers

Cosmic rays were discovered in the early 20th century. Scientists observed that there was a source of ionizing radiation causing electroscopes to discharge. At that time, the dominant hypothesis was that they came from radioactive elements in the ground. Victor F. Hess tried to test this hypothesis by ascending in a hot air balloon to 5300 m with three electroscopes. To his surprise, the electroscopes did not discharge slower, but four times faster at that altitude. He also repeated the experiment during a solar eclipse re-obtaining the same result, ruling out the sun as the main source of this ionizing radiation [16]. He concluded that the ionizing radiation had to come "from above". In 1938, P. Auger, R. Maze, and T. Grivet-Meyer discovered that this ionizing radiation arrived as "air showers", as they observed coincidences between counters 5 m apart from each other [17, 18].

Years have passed and our knowledge of cosmic rays has evolved. In this chapter, we summarize the basic current information about cosmic rays and the extensive air showers that they produce. In Sec. 2.1 we give a brief introduction to what cosmic rays are. In Sec. 2.2 we describe air showers, its components, and the Heitler and Heitler-Matthews models of the evolution of a particle cascade. In Sec. 2.3 we describe different principles of detection of cosmic rays or air showers. In Sec. 2.4 we detail different possible sources, acceleration mechanisms, and the propagation of cosmic rays. In Sec. 2.5 we give an overview of the estimated mass composition of cosmic rays, and we discuss the existing problem of a muon deficit in air-shower simulations. Finally, in Sec. 2.6 we provide a brief summary.

2.1 Cosmic rays

Cosmic rays are known to be predominantly ionized nuclei ranging from hydrogen (light) to iron (heavy) [16]. There are also electrons, photons, and neutrinos, but not nearly as abundant as the ionized nuclei. Nuclei with an atomic mass number A larger than that of iron are less stable and thus, in their long journey to Earth, disintegrate to smaller nuclei.

The cosmic ray energies cover a very wide range of energies, from 10^9 eV to a few 10^{20} eV. Figure 2.1 shows the cosmic ray energy spectrum (multiplied by $E^{2.6}$ to enhance its features) as measured by different experiments. For reference, the equivalent laboratory energy of the Large Hadron Collider (LHC) is at about 10^{17} eV. Futhermore, above $\sim 10^{15}$ eV, the flux is so small that cosmic rays can only be detected indirectly through the extensive air showers they produce. Above 10^{13} eV, the cosmic ray spectrum can be very well modeled by a power-law $J \sim E^{-\alpha}$, where $\alpha \sim 3$ is the so-called spectral index. It is remarkable that α changes only very little in the eight decades of energy. Nevertheless, there are four important features at which the spectral index changes: the "knee", the "second knee", the "ankle", and the flux suppression. The knee is at $\sim 10^{15.6}$ eV, after which the flux steepens. The second knee is at $\sim 10^{17}$ eV, after which the flux steepens even more. The ankle is at $\sim 10^{18.7}$ eV, after which the flux becomes less steep. And finally the flux suppression occurs above $\sim 10^{19.7}$ eV, where the flux drops very steeply.

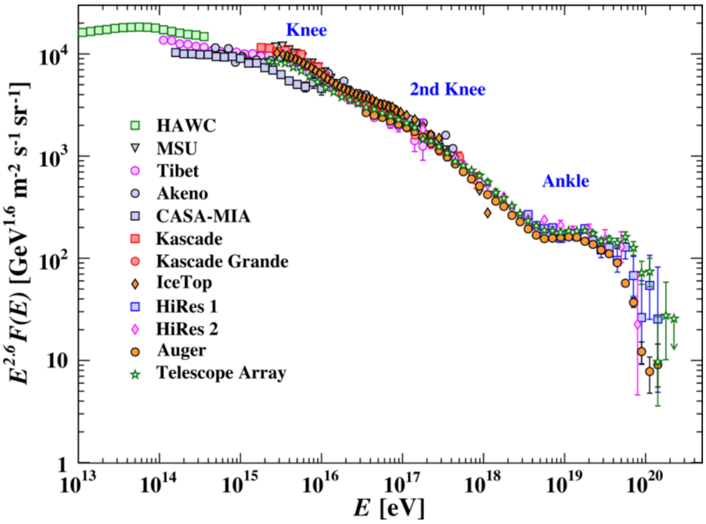


Figure 2.1: Differential energy spectrum of cosmic rays, multiplied by the energy to the power of 2.6, as a function of the energy (in logarithmic scale). The different experiments that contribute to the measurements are detailed in different markers. Figure extracted from Ref. [19].

It is also known that cosmic rays can be divided into a less energetic galactic component and a more energetic extra-galactic component [20]. However, the transition energy between these two is still not well determined [16].

2.2 Extensive air showers

When a primary cosmic ray interacts with the molecules of Earth's atmosphere, it produces a cascade of billions of secondary particles that conform an "extensive air shower" (EAS). These secondary particles can be divided into three components: the hadronic, the electromagnetic, and the muonic. The hadronic component is sometimes referred to as the backbone of the air shower. It is mainly comprised of charged pions, protons, antiprotons, and neutrons, which in their interaction and decays feed the muonic and electromagnetic components. The electromagnetic component is composed by electrons, positrons, and photons, and is the dominant component as it carries most of the energy of the shower. Finally, the muonic component, comprised of muons and antimuons, originates mainly from the decay of hadrons (only a very small fraction is produced from the electromagnetic component through photo-nuclear interactions). Since most of the muons reach ground before decaying, they operate as a tracer of the hadronic component.

In the electromagnetic component, the dominant interactions are pair creation ($\gamma \to e^+ + e^-$) and Bremsstrahlung ($e \to e + \gamma$), although photo-nuclear interactions are also present. Neglecting the latter, the Heitler [21] model provides a simple understanding of the electromagnetic component of a gamma-initiated air shower. The basic idea is illustrated in the left panel of Fig. 2.2. There we can see that a photon creates a e^+e^- pair, or an $e^\pm$ creates a photon, after one splitting length $l_{\mathrm{em}} = \ln 2\, X_0$, being $X_0 \sim 37\,\mathrm{g\,cm^{-2}}$ the radiation length in air. In one splitting length, the $e^\pm$ looses half of its energy, while the energy of the photon is assumed to be equally split between the e^+e^- pair. Furthermore, the number of particles duplicate after one splitting length. The multiplication process continues until the energy of the secondary particles reaches the critical energy $\xi_c^{\mathrm{em}} = 85\,\mathrm{MeV}$, where ionization and interactions with electrons of air molecules become dominant. We can estimate the maximum number of particles simply as $N_{\mathrm{max}} = E_0/\xi_c^{\mathrm{em}}$, being E_0 the energy of the primary. It also holds true that the maximum number of particles is $N_{\mathrm{max}} = 2^{n_c}$, where n_c is the number of splitting lengths that the cascade needs until reaching critical energy. It follows that $n_c = \ln(E_0/\xi_c^{\mathrm{em}})/\ln 2$. Finally, we can estimate the depth at which the number of particles is maximum as [16]

$$X_{\mathrm{max}}^{\mathrm{em}} = l_{\mathrm{em}} n_c = X_0 \ln\left(\frac{E_0}{\xi_c^{\mathrm{em}}}\right). \tag{2.1}$$

Although this simple model does not describe correctly the relative amounts of photons and electrons (there are much more photons than electrons in reality), the linearity between N_{max} and E_0 holds very well. Moreover, if we define the elongation rate $\Lambda = dX_{\mathrm{max}}/d\log_{10}(E_0)$, the rate of increase of X_{max} per decade of primary energy, we obtain from the Heitler model in air $\Lambda^{\mathrm{em}} \sim 85\mathrm{g\,cm}^{-2}$, which is also a realistic value [22].

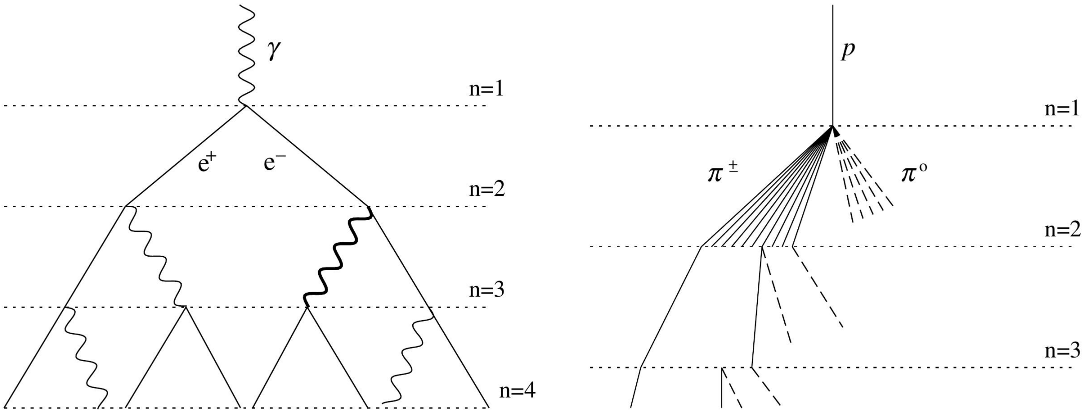


Figure 2.2: Schemes of an electromagnetic (left panel) and hadronic (right panel) air shower. In the electromagnetic air shower, photons produce pairs and electrons produce a photon through Bremsstrahlung after every splitting length. In the hadronic air shower, a proton produces several charged pions, which continue to interact in every splitting length, and several neutral pions that decay to photons (not depicted in the illustration). Figure adapted from Ref. [22].

The Heitler-Matthews model (see Ref. [22] and references therein) extends the Heitler model to hadron-initiated air showers. A sketch of this model can be seen on the right panel of Fig. 2.2. The basic idea is that particles interact in steps of one splitting length $l = \ln 2\,\lambda_\pi$, being $\lambda_\pi \sim 120\,\mathrm{g\,cm}^{-2}$ the interaction length of a pion in air. After each interaction, $N_{\mathrm{tot}} = 15$ particles are created, $N_{\mathrm{ch}} = 10$ charged pions, and 5 neutral pions. Each neutral pion is assumed to immediately decay into two photons, feeding the electromagnetic component. The charged pions propagate another splitting length and interact. The process continues until the pions reach their critical energy ξ_c^{π}, that occurs when the decay length to a muon is shorter than the splitting length. We can estimate the depth of the shower maximum, at which the number of photons and electrons are maximum, by focusing on the first generation of photons. For the purpose of the explanation, let us take the simplest primary: a proton. The first interaction occurs when the proton goes through $X_1 = \ln 2\lambda_{\mathrm{p}}$, being λ_{p} the interaction length of a proton. After the first interaction, $N_{\mathrm{tot}}/3$ neutral pions are generated, which decay to $2N_{\mathrm{tot}}/3$ photons. Each of these photons, with an energy of $E_0/2N_{\mathrm{tot}}$, generates an electromagnetic sub-shower. Then, using Eq. (2.1), the depth of the maximum for a proton shower is [16]

$$X^{\mathrm{p}}_{\mathrm{max}} = X_1 + X_0 \ln\left(\frac{E_0}{2N_{\mathrm{tot}}\xi^{\mathrm{em}}_c}\right). \tag{2.2}$$

Moreover, the number of muons of such an air shower is equal to the number of charged pions that reach critical energy $N_\mu = (N_{\mathrm{tot}}2/3)^{n_c}$. Since $n_c = \ln(E_0/\xi^{\mathrm{em}}_c)/\ln(N_{\mathrm{tot}})$, it follows that [16]

$$\begin{aligned}\ln N_\mu &= n_c \ln(N_{\mathrm{tot}}2/3),\\ &= \frac{\ln(E_0/\xi^{\mathrm{em}}_c)}{\ln(N_{\mathrm{tot}})}\ln(N_{\mathrm{tot}}2/3),\\ &= \beta \ln(E_0/\xi^{\mathrm{em}}_c),\end{aligned} \tag{2.3}$$

where in the last equation we grouped $\beta = \ln(N_{\mathrm{tot}}2/3)/\ln(N_{\mathrm{tot}})$. The Heitler-Matthews model yields $\beta = 0.85$, while air-shower simulations fit better a β between 0.85 to 0.92 [22].

To model heavier hadron-initiated air showers we can apply the superposition principle. The idea is that the air shower generated by a nucleus of atomic mass number A and energy E_0 is equivalent to the sum of A proton-initiated air showers of energy E_0/A. We can easily see how X_{max} is smaller (shallower) for a heavier primary by replacing E_0 with E_0/A in Eq. (2.2). The difference between the depth-of-the-shower maximum of a primary of mass number A and a proton primary of identical energy is [16]

$$X^{\mathrm{A}}_{\mathrm{max}} - X^{\mathrm{p}}_{\mathrm{max}} = -X_0 \ln A. \tag{2.4}$$

Regarding the number of muons, we have to consider that there are A times more pions at critical energy, but starting from an energy of E_0/A. This yields [16]

$$\ln N_\mu = (1-\beta)\ln A + \beta \ln(E_0/\xi^{\mathrm{em}}_c). \tag{2.5}$$

Apart from this simple, analytical models, the cosmic ray community profits from open software to simulate Monte Carlo air showers. The most prominent are CORSIKA [23], AIRES [24], and CONEX [25]. These implement the particle interactions following different hadronic interaction models. The most widely used high-energy hadronic interaction models are EPOS-LHC [26], QGSJetII-04 [27], and Sibyll2.3c [28]. These are referred to as post-LHC models due to their tuning to LHC data, which are extrapolated to higher energies. The most prominent low-energy hadronic interaction models are GEISHA [29], Fluka [30, 31], and UrQMD [32, 33].

2.3 Detection of cosmic rays

The way in which cosmic rays are detected depends on their energy. Up to some 10^{14} eV or 10^{15} eV, the cosmic ray flux is large enough to allow for the direct detection of cosmic rays. In this case, the typical detectors consist of spectrometers or calorimeters, which are put on balloons or satellites. Both can directly identify the primary, but spectrometers can measure energy much more accurately at the cost of measuring lower energy events (up to $\sim 10^{12}$ eV). This is because their maximum detectable energy is directly related to the maximum radius of curvature measurable by the tracker, as well as by the fiducial region of the magnetic field. The calorimeters (which do not put the particles through a magnetic field) measure energy with less precision because of shower-to-shower fluctuations in the development of the cascades in the calorimeter and due to the punch-through of particles at the bottom of the calorimeter. Some examples of such detectors are: the Alpha Magnetic Spectrometer-02

(AMS-02), which is inside the International Space Station [34]; CREAM [35], which has been carried on balloons; and DAMPE [36] and PAMELA [37], which are satellites.

Cosmic rays with energies above 10^{15} eV have such a low flux that they can only be detected through the EASs they produce. Large ground based observatories use the atmosphere as a calorimeter and detect either the secondary particles of the EAS, or the radiation produced by the interaction of the secondary particles with the atmosphere. Among the latter, detectors like the Pierre Auger Observatory [38] and Telescope Array [39] measure with dedicated telescopes the UV fluorescence light emitted by nitrogen molecules as they de-excite from the interactions with secondary particles. This technique is described in more depth in Sec. 3.2. Detectors like HESS [40] or TUNKA [41] use the Cherenkov light emitted by charged secondaries, although this technique is mostly used in gamma-ray astronomy. There are also radio detectors, a relatively new detection principle, which measure the radio pulses emitted by electrons as they deflect in the geo-magnetic field and due to the charge anisotropy in the development of the air-shower [42]. These are further described in Sec. 3.3.2. Examples of such detectors can be found in the Pierre Auger Observatory [38], Yakutsk array [43], TUNKA [41], and LOFAR [44]. For detecting particles at ground there are two kinds of detectors that are typically used: water Cherenkov detectors, and scintillators. The former measure the Cherenkov light of the charged secondaries as they interact with the water of the detector. The latter measure the scintillation light emitted by the secondary particles as they transverse the scintillation material (as minimum ionizing particles). These techniques are further described in Sec. 3.1 and Sec. 3.3.6 respectively. The Pierre Auger Observatory counts both with water Cherenkov detectors and scintillators [38]. Other experiments with scintillators are, for example, Telescope Array [39] and AGASA [45].

2.4 Acceleration mechanisms, propagation, and possible sources

The most supported mechanism through which cosmic rays reach non-thermal energies is the Fermi mechanism [46]. The basic idea is that a magnetized plasma would transfer macroscopic kinetic energy to charged particles in successive scattering encounters. In each encounter, the increase of energy ΔE is proportional to the energy E. Then, if the particle was injected with an energy E_0, after n encounters its energy is [16]

$$E_n = E_0 (1 + \xi)^n, \tag{2.6}$$

where ξ is the factor of proportionality. In the second-order Fermi mechanism, a magnetized plasma cloud moving at velocity v accelerates the charged particle as [16]

$$E_n = E_0 \left(1 + \frac{4}{3}\beta^2\right)^n, \tag{2.7}$$

where $\beta = v/c$. On the other hand, in the first-order Fermi mechanism, the plasma clouds are replaced by a shock front (like the ones produced by supernova explosions) with which the particle interacts multiple times, and here [16]

$$E_n = E_0 \left(1 + \frac{4}{3}\beta\right)^n. \tag{2.8}$$

The crucial difference between the two mechanisms is that in the interactions with the shock front, the particle always exits the shock in the upstream direction, hence there is always an energy gain. This makes the first-order Fermi mechanism more efficient.

It is interesting to analyze the spectrum implications of these mechanisms. From Eq. (2.6), the number of encounters needed to reach energy E_n is

$$n = \frac{\ln(E_n/E_0)}{\ln(1+\xi)}. \tag{2.9}$$

If the probability of escaping in one encounter is a constant $p_{\rm esc}$, the probability of remaining is $1 - p_{\rm esc}$, and the probability of remaining after m encounters is $(1 - p_{\rm esc})^m$. Then, the number of particles escaping with an energy $E \geq E_n$ is [16]

$$\begin{aligned} N(E \geq E_n) &\propto \sum_{m=n}^{\infty} (1 - p_{\rm esc})^m, \\ &= \frac{(1-p_{\rm esc})^n}{p_{\rm esc}}, \qquad (2.10) \\ &= \frac{1}{p_{\rm esc}} \left(\frac{E_n}{E_0}\right)^{-\gamma}, \qquad (2.11) \end{aligned}$$

where in Eq. (2.10) we replaced Eq. (2.9). In Eq. (2.11), $\gamma = -\ln(1-p_{\rm esc})/\ln(1+\xi)$. It is remarkable that the Fermi acceleration mechanism predicts a power-law spectrum of energies.

Moreover, on their journey from the source to the Earth, charged cosmic rays can be deflected by magnetic fields. Knowing that the Larmor radius ($r = p/(qB)$) is inversely proportional to the charge q, we can understand that lighter, less charged nuclei are less deflected than heavier nuclei, and that more energetic nuclei (with larger momentum p) also deflect less [47]. Furthermore, extra-galactic cosmic rays can interact, most notably, with photons of the radiation field present in the Universe (like the cosmic microwave background and the extragalactic background light). For a light cosmic ray, e.g. a proton, the dominant effect is photo-pion production ($p + \gamma \rightarrow p + \pi^0$). This process is endothermic and thus makes the proton loose energy. The effect is known as the Greisen, Zatsepin, Kuzmin (GZK) supression, since it limits the maximum energy that a cosmic ray can have. The steep fall of the spectrum after $\sim 5.5 \times 10^{19}$ eV could be explained by the GZK cutoff [48], and/or by a limit in the power of the sources. Moreover, electron positron pair production (known as Bethe-Heitler pairs) with photons of the radiation field is also possible and endothermic, but the loss of energy is considerably smaller and is thus not the dominant effect. For heavier nuclei, photo pion production is possible but the dominant effect is photo-disintegration [49].

Regarding cosmic ray sources, there are conditions on their size and magnetic field such that they can accelerate cosmic rays to ultra-high energies. The Hillas condition [50], derived from equating the Larmor radius of the accelerated particle to the size R of the acceleration region, can be used to estimate the maximum energy a candidate source can produce

$$E_{\rm max} \propto \beta Z e B R, \tag{2.12}$$

where Z is the charge number of the accelerated particle, e is the absolute value of the electron charge, and B is the magnetic field of the source. This is a necessary (but not sufficient) condition that a possible source has to fulfill. Figure 2.3 shows a so-called Hillas plot, a diagram of possible cosmic ray sources according to their magnetic field and characteristic distance. Any object above and to the right of the red dashed lines are able to confine a proton in the acceleration region up to the corresponding marked energy.

From Fig. 2.3 we can see that there is not one clear and unique candidate as a source of ultra-high energy cosmic rays (UHECRs, $E \geq 10^{18}$ eV). The prime candidates of galactic

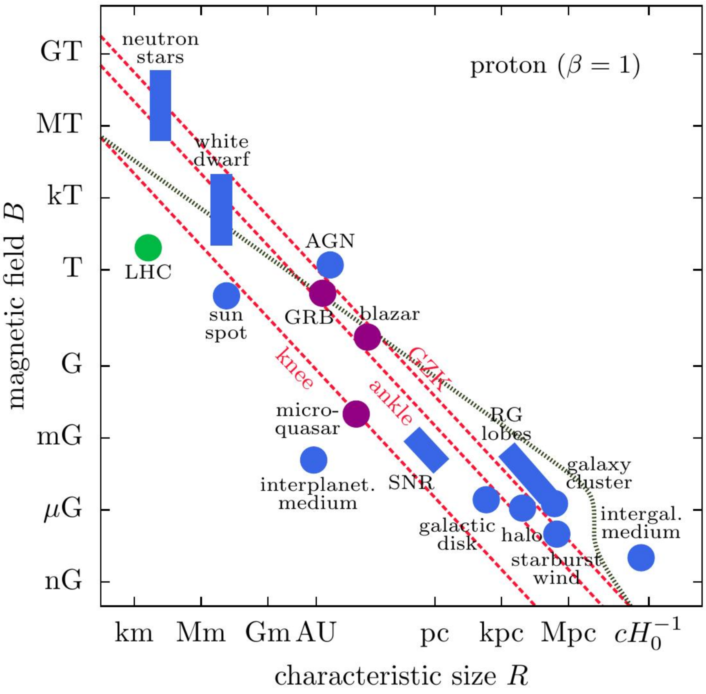


Figure 2.3: Hillas plot: Possible cosmic ray sources classified according to their size and magnetic field. The red dashed lines mark the regions above which a source would be able to accelerate a proton to the corresponding energy (knee, ankle, or GZK), assuming $\beta = 1$. The dotted gray line marks an upper limit from interactions with the microwave background and due to synchrotron losses in the source. Figure extracted from Ref. [51].

cosmic-rays are supernova remnants (SNRs), while the prime candidates for extra-galactic cosmic rays are gamma-ray bursts (GRBs), active galactic nuclei (AGNs), and starburst galaxies [20]. There is ongoing research dedicated to identifying possible concrete sources. On one hand, neutrinos, and also photons to some degree, would be great tracers of UHECR sources as they propagate unaffected by the intergalactic magnetic field in their journey to Earth. However, at ultra-high energies their fluxes are extremely low [52, 53]. Photons are also more likely to interact with the radiation field. An alternative would be to use the light component of UHECRs, which, as mentioned before, deflect less than heavier (and less energetic) cosmic rays [1]. However, the latter requires an event-by-event classification of the primary mass, which is still a challenge.

2.5 Mass composition and muon deficit in air-shower simulations

We understand for mass or chemical composition the identity of the primary cosmic ray. Understanding the mass composition of cosmic rays is one of the keys to solving the long standing mystery of their origin. On the one hand, as mentioned in Sec. 2.4, the light component of ultra-energetic cosmic rays could be used to trace back sources [1]. On the other hand, the transition between galactic and extragalactic sources is expected to be directly related to a change in composition as a function of the energy. This is because it is believed that the galactic sources reach lower acceleration limits than extragalactic sources. At the end of their capabilities, the galactic sources accelerate heavier, more charged nuclei to higher energies than lighter nuclei. Therefore, we expect to see a transition from the highest-energy galactic heavy component to the lowest-energy extragalactic light component [2]. In a similar way, knowing the mass composition of cosmic rays would help constrain models of the suppression, as it is expected that the highest-energy cosmic rays are mainly extragalactic intermediate mass or even heavy nuclei [3].

Furthermore, above the highest (laboratory) energy reachable by the Large Hadron Collider ($\sim 10^{17}$ eV), cosmic rays (or rather the EASs they produce) are the only way to test high-energy hadronic interaction models [54]. These models are typically tested by analyzing the consistency of the composition implications of different EAS observables [54]. Composition studies help to improve these models, which in turn improve the precision of the inferred mass composition [8, 9].

From Eqs. (2.4) and (2.5) we can see that the depth-of-the-shower maximum $X_{\max}$ and the muon content N_μ are parameters sensitive to the mass composition, as they depend on the mean logarithmic mass number $\ln A$. In fact, $X_{\max}$ and N_μ (or other measures of the muon content) are the most sensitive observables to the mass composition [55]. There are other composition-sensitive observables, like the slope of the lateral distribution function, the rise time of the signals at ground, and the maximum of the muon production depth distribution, but they do not significantly improve the discrimination power of the $X_{\max}$ and N_μ combination [56]. Moreover, not only the mean values of $X_{\max}$ and N_μ carry composition information, but also their fluctuations do. The first one, $X_{\max}$, is typically inferred from the longitudinal development of air showers measured with optical detectors, like fluorescence detectors or non-imaging Cherenkov detectors, but it can also be indirectly reconstructed using radio measurements [56]. The method for reconstructing $X_{\max}$ from fluorescence measurements is further explained in Sec. 3.2. The second one, the muon content, can be measured directly using shielded or buried detectors, or indirectly using inclined events [8]. In both cases, the much more abundant electromagnetic component has to be absorbed, either by the earth or shielding material, or by the atmosphere. After measuring samples of the muon lateral distribution function, the muon number or density at a specific distance to the shower axis can be estimated by evaluating a fitted function. This procedure is further described in Sec. 3.3.6.

The composition fractions f_A can be estimated from measurements of the $X_{\max}$ distributions by fitting it, at each energy, via a weighted sum of the single-primary distributions [57]

$$P(X_{\max}|E) = \sum_A f_A(E) P(X_{\max}|A, E), \tag{2.13}$$

where $\sum_A f_A(E) = 1$. Evidently, the inferred fractions depend on the assumed hadronic interaction model that provides the single-primary $X_{\max}$ distributions.

Figure 2.4 shows the comparison of measurements of $\langle X_{\max} \rangle$ (top panel) and $\sigma[X_{\max}]$ (bottom panel) from different experiments against proton and iron simulations of the three leading high-energy hadronic interaction models. There are evident differences between the predictions of different experiments (most notably at lower energies), implying that the estimations suffer from important systematic uncertainties. Nevertheless, the top panel of Fig. 2.4 suggests that there is a shift to a heavier composition that starts at $\sim 10^{16}$ eV, and another one at $\sim 10^{18.5}$ eV. The latter can also be observed in the bottom panel of Fig. 2.4. Between $\sim 10^{17.5}$ eV and $10^{18.5}$ eV the measurements are compatible with a light composition. Notice that a change in the slope of $\langle X_{\max} \rangle$ (a change in the elongation rate) is a clear evidence of a change in composition, as it can be derived from Eq. (2.4).

Moreover, the composition implications are typically expressed in terms of the mean-logarithmic mass $\langle \ln A \rangle$. By using Eq. (2.4) for a primary of mass number A and for iron ($A = 56$), and by taking the average, we can derive that [56]

$$\langle \ln A \rangle = \frac{\langle X_{\max} \rangle - \langle X^{\mathrm{p}}_{\max} \rangle}{\langle X^{\mathrm{Fe}}_{\max} \rangle - \langle X^{\mathrm{p}}_{\max} \rangle} \ln 56. \tag{2.14}$$

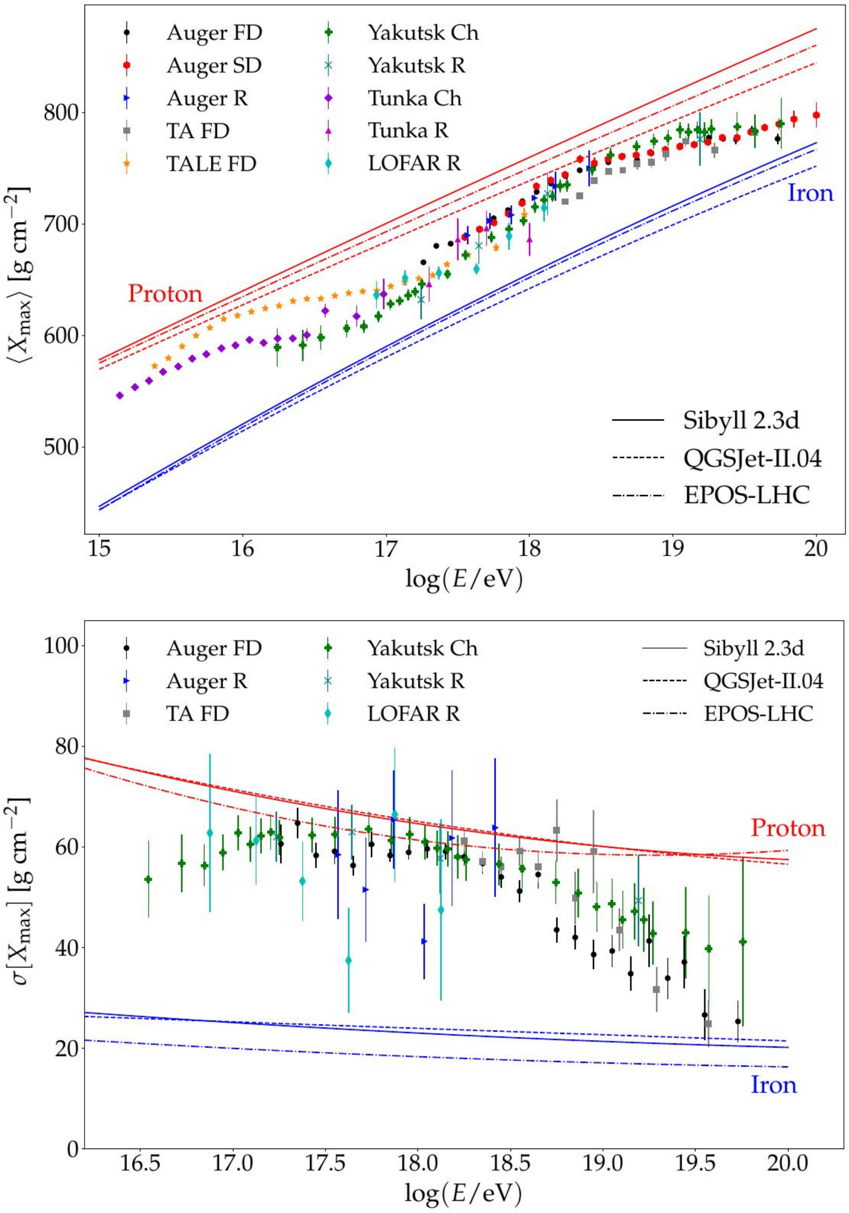


Figure 2.4: Average depth-of-the-shower maximum (top panel) and fluctuations in the depth-of-the-shower maximum (bottom panel) as a function of the logarithmic energy. Different detectors are identified with different markers (FD stands for fluorescence detectors, SD for surface detectors, R for radio detectors, and Ch for Cherenkov detectors). The predictions of EPOS-LHC, QGSJetII-04, and Sibyll2.3d for proton (red) and iron (blue) generated air showers are shown in dashed, dotted, and solid lines respectively. Figures extracted from Ref. [56].

Figure 2.5 shows $\langle \ln A \rangle$ (as inferred from $X_{\max}$ measurements) as a function of the logarithmic energy. The observations are equivalent as for the top panel of Fig. 2.4. It is relevant to add that the differences of $\langle \ln A \rangle$ between the different models are within $\sim \pm 0.8$, and fairly constant with energy.

As it can be seen in Figs. 2.4 and 2.5, air-shower simulations reproduce to a good extent the behavior of $X_{\max}$ (at least for post-LHC models), and the difference between models is relatively small. Moreover, the theoretical uncertainties of $X_{\max}$ are relatively small compared to those of N_μ (or ρ_μ) [4, 58]. Nevertheless, recent studies suggest that a shift in $X_{\max}$

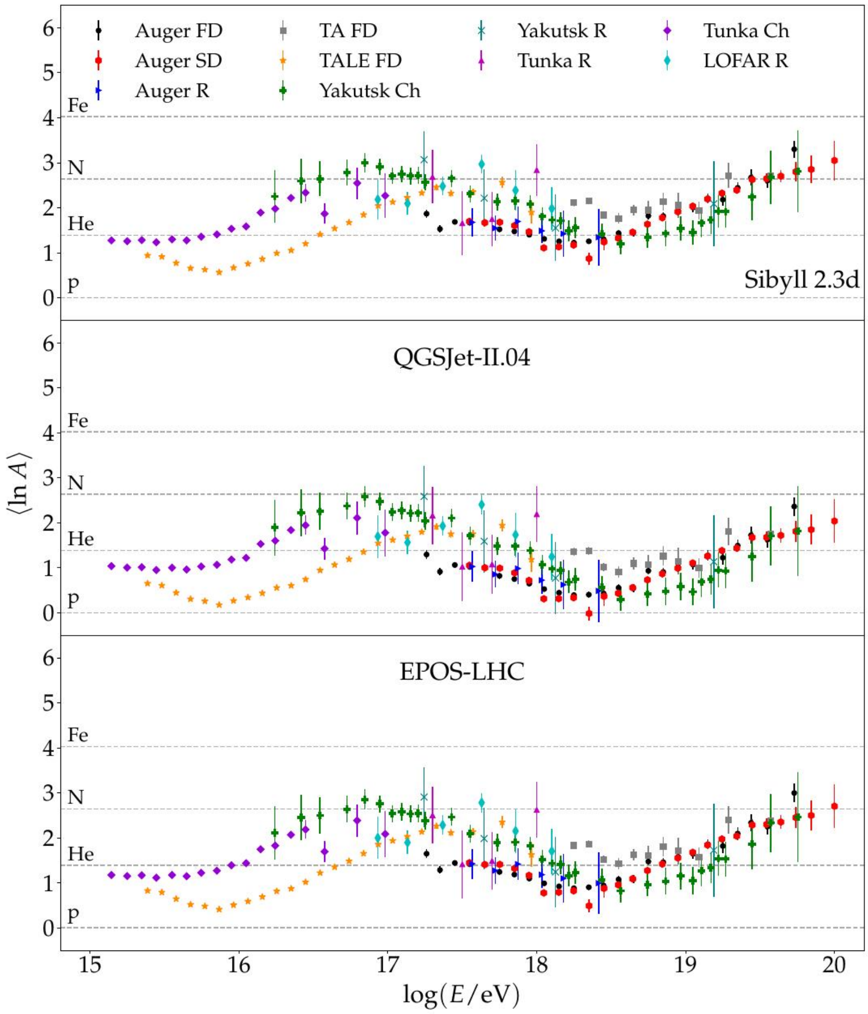


Figure 2.5: Mean-logarithmic mass inferred from X_{max} measurements (c.f. Eq. (2.14)) as a function of the logarithmic energy. The data are the same as of the top panel of Fig. 2.4. The top panel corresponds to the high-energy hadronic interaction model Sibyll2.3d, the middle panel to QGSJetII-04, and the bottom panel to EPOS-LHC. Figure extracted from Ref. [56].

in simulations may also be necessary to reproduce the observed behavior in data [56]. In spite of this, other EAS observables are typically tested by comparing their composition interpretation to the one obtained from X_{max}. An inconsistent interpretation would imply that the models do not correctly reproduce all EAS observables.

Muon density measurements of different experiments cannot be directly compared because they greatly depend on the energy, zenith angles, altitude of observation, distance to the shower axis, and muon energy threshold. One way to compare them is through the calculation of a so-called z-value, defined as [5]

$$z := \frac{\ln N_{\mu,\mathrm{data}}^{\mathrm{det}} - \ln N_{\mu,\mathrm{p}}^{\mathrm{det}}}{\ln N_{\mu,\mathrm{Fe}}^{\mathrm{det}} - \ln N_{\mu,\mathrm{p}}^{\mathrm{det}}}, \tag{2.15}$$

where the suffix “det” indicates that the proton and iron muon contents derive from full-detector simulations. At least to first order, detector effects and the effects of the specific conditions in which the measurements were taken should cancel out. The properties and uncertainties of this z-scale, as well as concrete formulas for its calculation in averaged data are extensively discussed in Secs. 7.1.4 and 8.2.4. Moreover, by replacing Eq. (2.5) into Eq. (2.15), we can see that the z-values should map to $\langle \ln A \rangle / \ln 56$. This enables the comparison of the composition implications of measurements of the muon content to those of measurements of $X_{\max}$.

Figure 2.6 shows the comparison of the z-values of eight experiments: EAS-MSU, IceCube Neutrino Observatory, KASCADE-Grande, NEVOD-DECOR, Pierre Auger Observatory, SUGAR, Telescope Array, and Yakutsk. We can see that data and simulations are consistent up to 10^{16} eV [5, 59]. Above this energy, there is a discrepancy between data and simulations, suggesting a deficit in simulations that increases with the energy. Additionally, the Pierre Auger Collaboration reported that the muon deficit is greater for larger values of the zenith angle [7], and Telescope Array observed a larger muon deficit at larger distances to the shower axis [60].

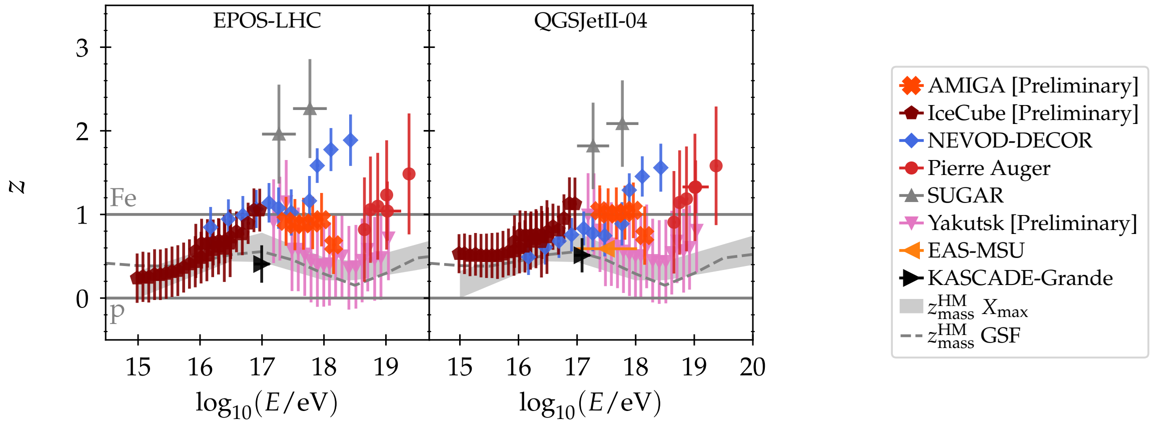


Figure 2.6: Values of z (c.f. Eq. (2.15)) as a function of the logarithmic energy, for EPOS-LHC (left panel) and QGSJetII-04 (right panel). Different markers denote different detectors. In gray dashed lines, the expected z-values from the Global Spline Fit composition model [61], and in a gray shaded area those expected from the Pierre Auger fits to $X_{\max}$ distributions [57]. Adapted from Refs. [5, 59].

It is still not understood whether the muon deficit is caused by a new phenomenon at high energies or by a partial mismodeling of hadronic interactions at high or low energies [7]. Understanding in which part of the phase-space the muon deficit is, can help model builders improve the high-energy hadronic interaction models until they reproduce the behavior of the muon content. If this was achieved, the muon content could be used to reduce the systematic uncertainties of mass composition analyses, opening the possibility of a better understanding of the origin of cosmic rays.

2.6 Summary

In this chapter we summarized the key aspects of cosmic ray and air-shower physics relevant to this work. We started with the description of what we know cosmic rays are: particles, mostly nuclei ranging from proton to iron, with non-thermal energies that come from outer space. We showed how the cosmic ray spectrum can be well described by a power-law with small changes in the spectral-index throughout almost 8 decades of energy. We explained how a cosmic ray generates an extensive air shower, which is a cascade of secondary particles, as it interacts with the Earth's atmosphere. We discussed the different components of an air shower (hadronic, electromagnetic, and muonic), and presented the Heitler and

Heitler-Matthews models of an electromagnetic and hadronic air shower (respectively). In particular, we analyzed the implications of the latter in the behavior of the depth-of-the-shower maximum $X_{\max}$ and N_{μ} with energy and primary mass.

Moreover, we presented the different cosmic ray and air-shower detection techniques. We then explained the Fermi acceleration mechanism and showed how the first-order Fermi mechanism predicts a power-law spectrum. As the cosmic rays propagate in the intergalactic medium, the interactions with the cosmic microwave background can be shown to predict a flux suppression at the highest energies. Moreover, we presented the Hillas criterion, a condition on the size and magnetic field that a possible source has to fulfill to be considered as a candidate source.

Finally, we discussed how knowing the mass composition of cosmic rays would help elucidate their origin, as well as improve hadronic interaction models. We explained that $X_{\max}$ and N_{μ} are the most sensitive composition observables, and show current results on their composition implications. In particular, we justified why inconsistent composition implications of N_{μ} measurements with respect to those of $X_{\max}$ are typically interpreted as a muon deficit in air-shower simulations. The muon deficit, starting at $\sim 10^{16}$ eV, is observed to increase with energy, zenith angle, and distance to the shower axis.

Chapter 3

The Pierre Auger Observatory

The Pierre Auger Observatory is the largest cosmic-ray observatory of the world. Situated in the southern hemisphere, in Malargüe, Argentina, it covers an area of $\sim$3000 km^2 [38]. Such large areas are necessary to measure with sufficient statistics the low flux of high-energy cosmic rays.

The observatory is designed as a hybrid detector: It combines fluorescence detectors (FD) with the surface detectors (SD). In Fig. 3.1 we show a map of the Observatory, where it can be noticed that the fluorescence telescopes, grouped in four sites, overlook the area covered by the SD. The fluorescence telescopes have a low duty-cycle ($\sim$ 15 % [38]), but set the energy scale of the Observatory by measuring calorimetrically the electromagnetic component of air showers. On the other hand, the surface array has a duty-cycle of almost 100 % [38], hence providing very large statistics, measuring the muonic and electromagnetic components of the air showers shortly after the shower maximum [62]. The energy of the SD is then cross-calibrated to the scale set by the FD [63]. In this way, the combined detectors provide measurements of extensive air showers (EASs) with low systematic uncertainties and high statistics.

In this chapter we explain the different detectors that compose the Pierre Auger Observatory. We describe the SD, the FD, as well as the detectors that integrate the upgrade of the Observatory (Auger Prime). These are the Surface Scintillator Detector (SSD), the Radio Detector (RD), and the Underground Muon Detector (UMD) (also referred to as Auger Muons and Infill for the Ground Array (AMIGA)). The latter is explained particularly in depth, as it is extensively used in this work.

3.1 The surface detector

The SD was designed to operate with a 100 % duty-cycle, to measure the time structure of the signals from EASs, to measure vertical as well as inclined events, and to have an aperture independent of the energy above $10^{18.5}$ eV [38]. These design goals have been met, with more than 1600 water Cherenkov detectors (WCDs) deployed which cover an area of $\sim$3000 km^2. The aperture achieved considering events with zenith angles below 60° is of 7350 km^{2}sr [65]. The stations have elevations ranging between 1300 m and 1600 m above sea level [66]. They are arranged in an isometric triangular grid of side 1500 m. A subset of the stations within the 1500 m array (usually referred to as SD-1500) are arranged more densely: one subset covers 28 km^2 with stations spaced by 750 m (SD-750), while the other one covers 1 km^2 with stations spaced by 433 m (SD-433). The objective of these denser arrays is to extend the sensitivity of the Observatory to lower energies [66].

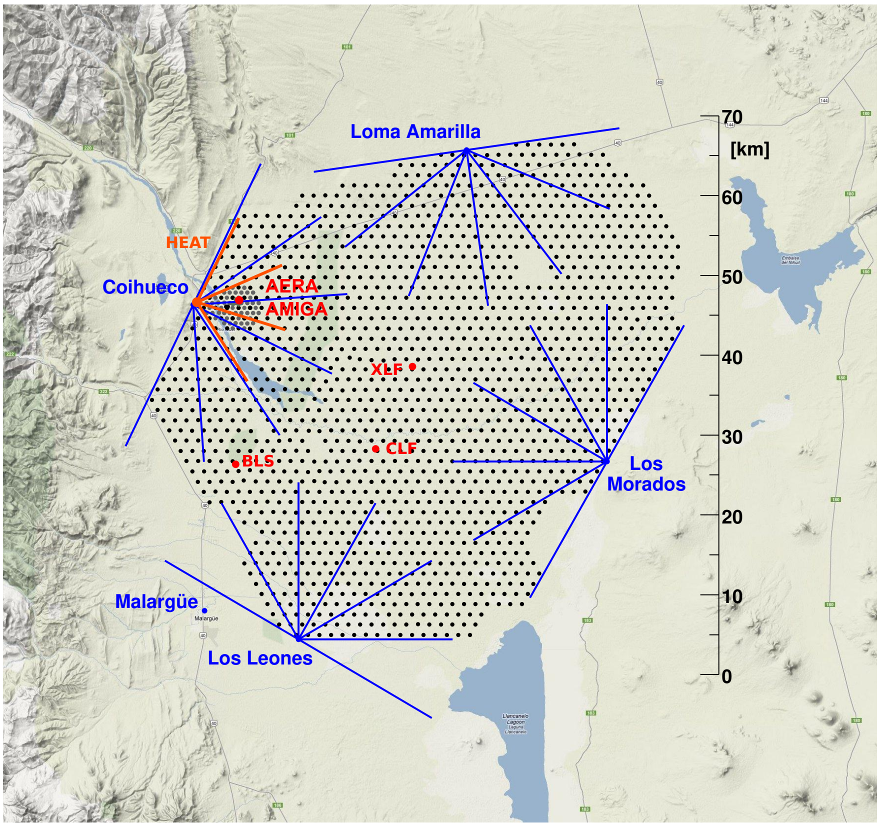


Figure 3.1: Layout of the Pierre Auger Observatory. On the right, a scale in kilometers can be seen. The surface detector stations (black dots) cover an area of $\sim 3000\,\mathrm{km}^2$. The fields of view of the fluorescence telescopes (blue lines) overlap with said area. The position of other detection systems, including the UMD (alias AMIGA), can also be seen (red dots and lines). Image taken from Ref. [64].

A picture of an SD-station can be appreciated in Fig. 3.2, showing its main components. The station consists of a cylindrical water tank of 3.6 m in diameter and 1.2 m in height. The tank is made of polyethylene, and contains a sealed multi-layered liner with a reflective inner surface. Inside the liner 12, 000 l of ultra-pure water are stored. The liner has also three windows of clear polyethylene at its top part, through which three photomultiplier tubes (PMTs) measure the Cherenkov photons produced by high-energy charged secondary particles as they pass through the water. The height of the WCD allows also to measure high-energy photons through the electron-positron pairs they produce inside the tank (or rather, by the Cherenkov light emitted by the latter). The PMTs (Photonis XP1805/D1) are 9 in. in diameter and are placed 1.2 m apart from the central axis of the tank.

Two signals are obtained from each PMT, which are tagged with a GPS time stamp thanks to the GPS antenna (see Fig. 3.2). The low-gain signal is taken from the anode of the PMT, while the high-gain signal is taken from the last dynode and amplified to be 32 times larger than the low-gain signal [66]. The signals are digitized using 10 bit 40 MHz semi-flash analog-to-digital converters (FADCs) [38]. These are then clocked into a programmable logic device (PLD) which implements the firmware that evaluates the trigger conditions on the FADC traces, stores the traces in a buffer memory, and communicates with the station microcontroller when there is a local or station trigger [38]. The station controller is an IBM PowerPC 403 GCX-80 MHz and lays together with the slow-control microprocessor in the unified board (UB)[1]. The station controller runs the data acquisition, transmitting through the communications antenna (visible in Fig. 3.2) the time stamps of the (level two) local triggers to the central data acquisition system (CDAS) and replying to the data requests from CDAS array-level triggers. On the other hand, the slow-control consists of DACs and ADCs that monitor voltages, currents, and temperatures that are used for calibration [38]. The electronics lie inside an enclosure which seals and protects them from dust, light, and water [67], and that can be seen in Fig. 3.2.

The data acquisition system is a hierarchical system consisting of local- and array-level triggers. The first two levels, T1 and T2, are local-level triggers formed at each station. The signals that pass the T2 trigger-level are sent to the CDAS. The CDAS searches for spatial-temporal coincidences among the T2 data in at least 3 neighbor stations. If the geometry and timing of the coincidences is compatible with a shower-front propagating at the speed of light, then a T3 array-level trigger is raised. When this happens, CDAS requests the signals, as well as calibration and monitoring data, of all stations lying within 6 crowns around the station with largest signal that had a T1 or T2 within 30 µs [65, 68].

The SD stations were purposefully designed to be self-powered. For this reason, they count with two solar panels as well as a battery box (see Fig. 3.2) that provide an average of 10 W, sufficient for powering all the components.

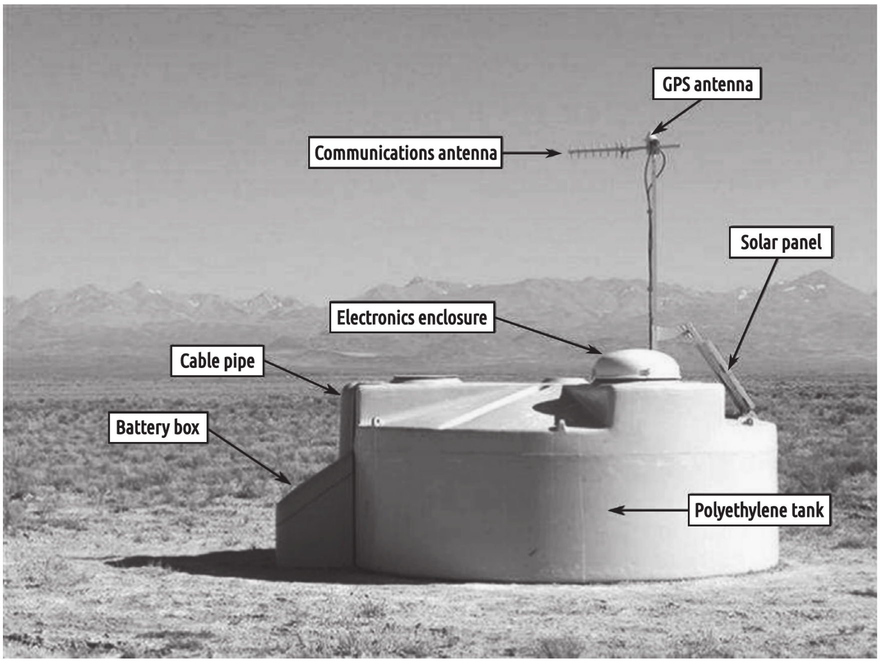


Figure 3.2: A picture of an SD-station showing its main components. Image taken from Ref. [67].

After understanding how the SD stations work, we summarize the main aspects of the SD event reconstruction, as it is used in the next chapters. A detailed explanation can be found in Ref. [66]. The SD events can be reconstructed to obtain mainly the shower geometry (arrival direction and impact position) and the shower size. The latter is then converted to an energy using a cross-calibration with FD measurements.

The first step is to obtain the signal at each station out of the FADC traces. The high-gain FADC traces are used if there is no saturation, and if not the low-gain FADC traces are used. The procedure consists of selecting the window with signal (the same for the three traces of

---

[1] At the moment of writing this thesis, the electronics of the Pierre Auger Observatory are also undergoing an upgrade. The unified boards (UBs) are being replaced by upgraded unified boards (UUBs). More information in Sec. 3.3.4.

an SD station), and integrating the ADC counts of the traces as a function of time to obtain the signal charge measured by each PMT. Then the average of the three charges is taken. The average charge is converted to units of Vertical Equivalent Muons (VEMs) by means of the detector calibration. The charge of one VEM is obtained from the fit to the charge histogram of background particles (which has a distinguishable muon contribution) [66].

The geometry is determined in an iterative procedure. A first estimate of the impact position is given by the signal barycenter, this is, the signal-weighted center-of-mass of the stations participating in the event. Then the first estimate of the direction of the shower axis (anchored at the barycenter) and barycenter time are obtained by assuming a plane shower-front propagating at the speed of light, solving for the three neighbor stations with highest average signals. With this initial values, a time model including curvature is fitted to data to determine the shower axis direction, the impact time of the shower core, and (if there are more than 5 stations) the radius of curvature. The fitting procedure consists of a chi-square minimization of predicted vs measured start-times, contemplating uncertainties in the measured start-times (typically dominated by sampling fluctuations) [66].

Knowing the shower geometry, the shower size is estimated from the fit to data of the lateral distribution function (LDF), i.e., the average signal as a function of the distance to the shower axis $S(r)$. The shower size is the expected average signal size at an optimal distance r_{opt}, defined as the distance that minimizes systematic uncertainties due to the unknown shape of the LDF. Its value depends almost exclusively on the spacing and shape of the array [69], and for the SD-1500 array of the Pierre Auger Observatory it is $\sim 1000\,\text{m}$, hence adopting $r_{\text{opt}} = 1000\,\text{m}$. The model of the signal as a function of the distance is

$$S(r) = S(r_{\text{opt}})\, f_{\text{LDF}}(r), \tag{3.1}$$

where $S(r_{\text{opt}})$ is the shower size estimator, and $f_{\text{LDF}}(r)$ is the LDF model, normalized such that $f_{\text{LDF}}(r_{\text{opt}}) = 1$. There is no analytical solution to hadronic-cascade equations, but there are different empirical LDF models. The one used in $\overline{\text{Off}}\underline{\text{line}}$ [70], the collaboration software for simulating the detector and reconstructing data and simulations, is a modified Nishimura-Kamata-Greisen (NKG) function (see Ref. [66] and references therein)

$$f_{\text{LDF}}(r) = \left(\frac{r}{r_{\text{opt}}}\right)^{\beta} \left(\frac{r + r_0}{r_{\text{opt}} + r_0}\right)^{\beta+\gamma}, \tag{3.2}$$

where $r_0 = 700\,\text{m}$. For events with "sufficient sampling" of the LDF around 1000 m, β is fitted. The criterium is that there have to be at least 2, 3, or more stations within a distance to the shower core on the shower plane of $400\,\text{m} < r < 1600\,\text{m}$, separated by at least 900 m, 800 m, or 700 m respectively. If the criterium for fitting β is not met, β is fixed to a value parameterized from events with good LDF sampling (high multiplicity and good spatial distribution). On the other hand, γ is very difficult to fit event-by-event because it is strongly correlated to β. For this reason, γ is always fixed to the value of a parameterization obtained from good LDF-quality events. The only difference is that this subset of good events are requested to sample well the tail, modifying the condition on the distances to the shower axis to be within $1000\,\text{m} < r < 2000\,\text{m}$. This is because γ modifies the standard NKG (a simple power-law) at large distances to the shower axis. Both parameterizations of β and γ are functions of the zenith angle and of the shower size [66].

With the signal model, the shower size $S(r_{\text{opt}})$ and position of impact $\vec{x}_c$ are obtained by minimizing a log-likelihood

$$\ln \mathcal{L} = \sum_i \ln P(S(r_{\text{opt}}), \vec{x}_c | S_i, \vec{x_i}), \tag{3.3}$$

where the sum runs over the stations that belong to the event, S_i is the signal of the i-th station, and $\vec{x_i}$ its position. In $\overline{\text{Off}}\underline{\text{line}}$, the model of P depends on the type of signal of the station [66]:

- For small signals of candidate stations ($S_i < 20\,\text{VEM}$), P is modelled as a Poissonian.
- For large signals of candidate stations or recovered saturated stations[2] ($S_i \geq 20\,\text{VEM}$), P is approximated as a Gaussian.
- For unrecoverable saturated stations, the signal S_i is taken as a lower limit, and thus P is the probability of detecting a signal larger than S_i, and approximated as the integral of a Gaussian from S_i to infinity.
- For non-triggered stations, P is the complement of the trigger probability.

Accidental stations, which trigger within the time window of an air-shower event but have no time and geometric correlation with the rest of the stations, are excluded from the likelihood reconstruction.

Figure 3.3 shows an example of an LDF fit performed using $\overline{\text{Off}}\underline{\text{line}}$.

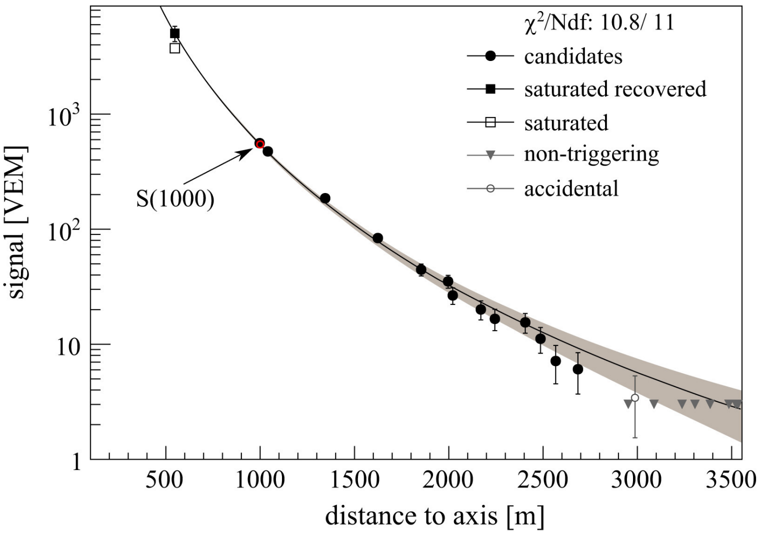


Figure 3.3: Measured signals (in VEM) as a function of the distance to the shower axis on the shower plane, for different types of stations (different markers), and the $\overline{\text{Off}}\underline{\text{line}}$ fit (solid line) that uses a modified-NKG model (see text for details). Taken from Ref. [38].

3.2 The fluorescence detector

The fluorescence detector consists of 24 air fluorescence telescopes that are grouped in 4 sites: Coihueco, Loma Amarilla, Los Morados, and Los Leones, as shown in Fig. 3.1. A picture of and FD building can be seen in the left panel of Fig. 3.4. The FD was designed to detect all showers with energies above 3×10^{18} eV that fall within the observatory during its data-taking time, i.e., clear moonless nights. Each telescope covers an elevation range from 1.5° to 30°, as well as 30° in azimuth. In combination, the 6 telescopes of each site cover 180° in azimuth facing towards the interior of the array. The fluorescence telescopes detect the fluorescence emission of the nitrogen in the atmosphere (in the ultra-violet (UV) range),

---

[2] Saturated signals are recovered by using the fact that the undershoot in the low-gain channel of the PMTs is proportional to the size of the signal itself.

which happens as a consequence of the excitation due to the interaction with the charged particles of air showers. The number of emitted photons is linearly related to the deposited energy of said charged particles. Overall, these electromagnetic energy losses amount to $\sim$ 90 % of the energy of the primary, making the fluorescence detector an ideal tool for estimating the energy of cosmic rays [71].

At one of the sites there are also three high-elevation auger telescopes (HEAT) that cover an elevation range from 30° to 58°. A picture of HEAT can be seen in the right panel of Fig. 3.4. The design is very similar to that of the standard FD telescopes, but lie in individual inclined housings. The site where it lies (Coihueco) is the closest to the SD-750 array, allowing the detection of hybrid, lower energy events. By detecting these air showers that develop higher in the atmosphere, HEAT extends the range of the FD to lower energy air showers down to $\geq 10^{17}$ eV [72].

Figure 3.4: Left: Picture of the FD building at Los Leones. Right: Picture of HEAT. Images retrieved from Ref. [38].

The different parts of a fluorescence telescope can be seen in Fig. 3.5. Fluorescence telescopes are formed by: a 13 m^2 segmented mirror, a PMT-camera at the focus surface with 440 pixels, a large UV-passing filter, and a Schmidt optics corrector ring. The mirror was designed to be segmented to reduce weight and costs. The PMT camera consists of a matrix of 440 PMTs of the model XP3062 and light collectors (22 rows by 20 columns). The camera provides a field of view of 30° in azimuth and 28.1° in elevation. The UV filter is made of a Schott MUG-6 glass, transmitting more than 80 % of the light of interest ($\sim$ 300 nm to 380 nm), and filtering out visible photons that would otherwise make impossible the measurement of the UV photons. The corrector ring serves to double the aperture of the FD, while keeping it light and minimizing costs, and it is the circumferential part of the corrector plate of a classical Schmidt camera. It can be added that each telescope counts with an automatic shutter system to protect it from harsh climate conditions, such as strong winds, as well as from too intense light sources that could saturate the electronics. The buildings where the telescopes stand are climate-controlled [71].

The FD electronics performs anti-alias filtering, and digitizes (every 100 ns) and stores the signals of the PMT cameras. Similar to the SD (see Sec. 3.1), the triggering system is also hierarchical. There are three levels of triggers at the detector. A level 1 trigger is a pixel threshold trigger implemented in firmware, a level 2 trigger is a track-shape over many pixels also implemented in firmware, and a level 3 trigger is at software level and it identifies non-lightning events. If an event passes the level 3 trigger, the event is merged by the PC of the telescope building. If the event passes high-quality cuts and is thus a candidate for an air shower, a hybrid T3 trigger is sent to CDAS, which requests the data at the SD stations that are expected to have signals from the same air shower [71].

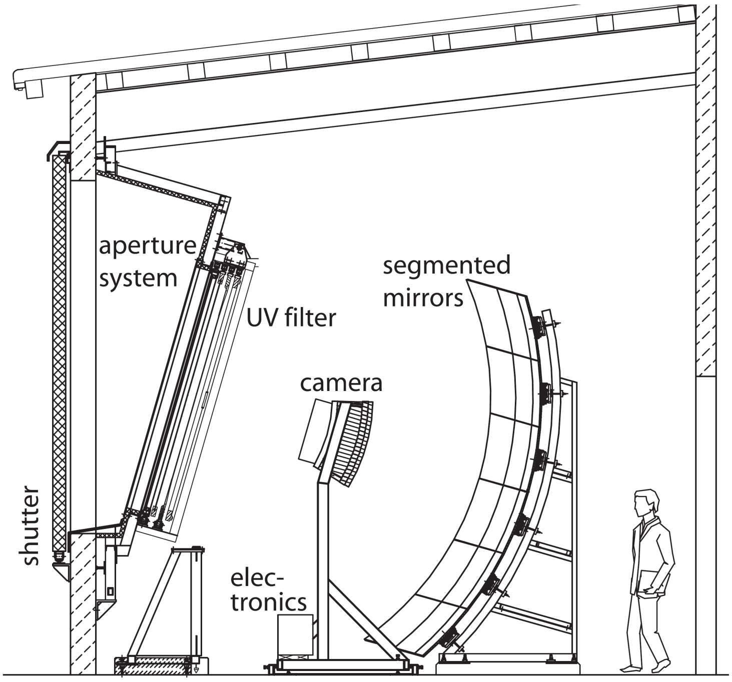


Figure 3.5: Scheme of a fluorescence telescope. The different parts are identified and a person is drawn for perspective. Image extracted from Ref. [38].

Knowing how the detector works, we now summarize the main aspects of the reconstruction of hybrid (SD-FD) events. For more details, see Ref. [71]. An air shower event appears on the PMT cameras as a line of activated pixels with a clear time sequence and in a track-like pattern. The geometry of the air shower can be determined by identifying the shower-detector plane, which is the plane that includes the eye and the line to the shower axis, and by using the timing information of the pixels. If there are SD stations participating in the event, which is mostly the case, the shower core is even more constrained, obtaining a better reconstructed geometry. The achieved resolution in hybrid reconstructions is of 50 m in the shower core and 0.6° in arrival direction. Furthermore, the depth of the shower maximum X_{max} can be also reconstructed using the FD. Knowing the geometry, the characteristics of the atmosphere, and the fluorescence yield (the number of photons emitted per unit of energy loss), the collected light as a function of time can be converted to energy deposit as a function of slant depth $dE/dX(X)$. Such a profile is fitted using a Gaiser-Hillas function [73, 74]

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

where X_0 and λ determine the shape of the function, and together with X_{max} are the parameters of the fit. As suggested by the notation, $(\frac{dE}{dX})_{\text{max}}$ is the energy deposit per slant depth at the maximum. An example of the fit to the shower profile can be seen in Fig. 3.6.

Finally, the total energy is estimated by integrating the profile, and by correcting by the so-called “invisible energy” carried by neutrinos and by high-energy muons. The energy estimated in this way has a resolution $\leq 10\,\%$ [71].

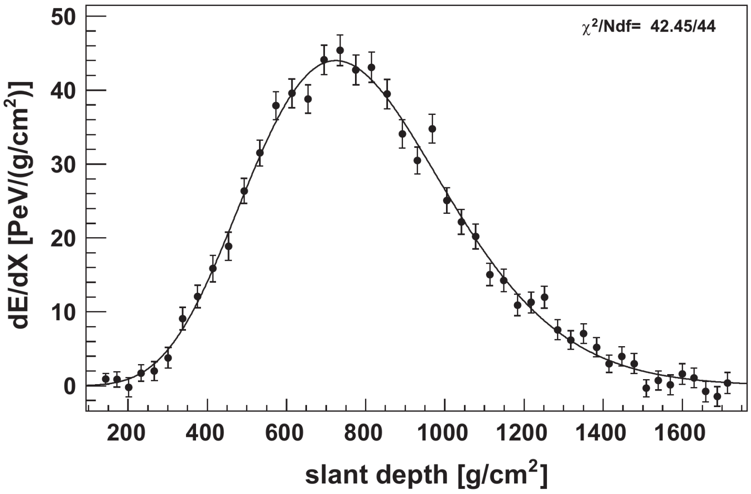


Figure 3.6: Energy deposit per slant depth as a function of the slant depth for an example event. The measured data points are in markers, and the fit of a Gaisser-Hillas function (see Eq. (3.4)) can be seen as a solid line. Figure extracted from Ref. [38].

3.3 Auger Prime

As stated in Chapter 2, there are several unresolved questions in the field of ultra-high energy cosmic rays. Their origin, nature, and the acceleration mechanisms that allow them to reach such high energies are still not completely understood. One of the key elements that would allow major steps forward is identifying event-wise the mass composition of the cosmic rays. To fulfill this purpose, the Pierre Auger Observatory is being upgraded by incorporating: the scintillator surface detector (SSD), the radio detector (RD), the extension of the range of the SD with a small PMT, the incorporation of new electronics or upgraded unified boards (UUBs), the extension of the up-time of the FD, and the underground muon detector (UMD). A scheme of an upgraded SD station can be seen in Fig. 3.7. The radio antenna and the SSD can be seen at the top of the tank.

Below, we summarize the main characteristics of the elements of the upgrade, and focus on the UMD as it is extensively used in this work.

3.3.1 Scintillator surface detector

The scintillator surface detectors (SSDs), which will lie on top of each SD station of the array, were designed to distinguish the different components of an air shower together with the SD. While the SSD is approximately equally sensitive to the electromagnetic and muonic components of air showers, the SD is more sensitive to the muonic component. The difference in sensitivity can be exploited to decompose the two measured signals, thus isolating the muonic part that is mass sensitive [75].

Each SSD module consists of 48 scintillator strips distributed in two equal panels, that add up to a total area of 3.8 m^2. Each strip is 1.6 m long, 5 cm wide, and 1 cm thick. Each strip has two straight extruded holes, inside which wavelength shifting (WLS) optical fibers lie. Each fiber threads two scintillator bars acquiring a "U" shape, and the ends of all fibers are bundled and collected in a cookie that connects them optically with a PMT. More information on the principle of detection of plastic scintillators and WLS fibers is given in Sec. 3.3.6. The data acquisition takes place simultaneously with the corresponding SD station, which triggers the SSD [76, 77].

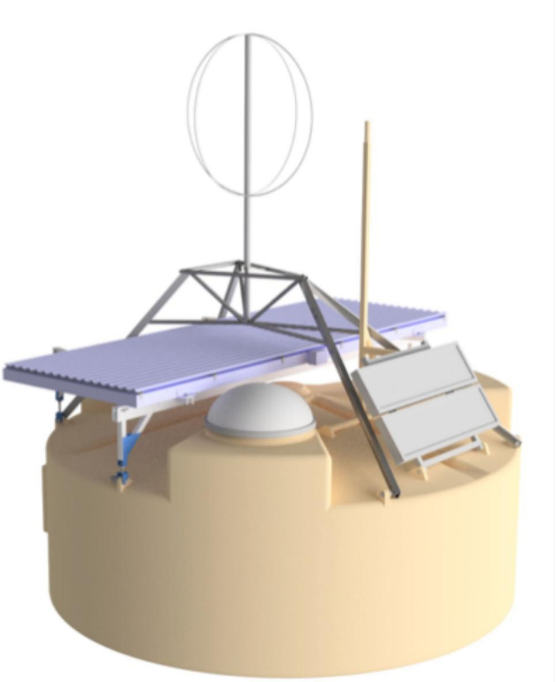


Figure 3.7: Scheme of an upgraded SD station. On top, the radio antenna and the SSD can be seen. Figure extracted from Ref. [10].

3.3.2 Radio detector

The Auger engineering radio array (AERA), a prototype radio array preceding the radio detector (RD), proved the capability of radio detection to separate the muonic and electromagnetic components of air showers, principally in very inclined air showers. The RD, with a $\sim 100\,\%$ duty cycle, will provide mass sensitivity in the zenith range of 65° to 85°, which can be useful for anisotropy studies. The expected capabilities of separating proton from iron air showers is a $\sim 90\,\%$ purity at an efficiency of $80\,\%$ [78].

The RD will consist of a dual-polarized Short Aperiodic Loaded Loop Antennas (SALLAs) installed on top of each SD station. Like the SSD, it will be triggered by its paired SD station [78].

3.3.3 Small photomultiplier tubes

Apart from the three PMTs of each SD station, a new small PMT (SPMT) will be installed in each SD station. The purpose is to extend the dynamic range of the SD stations, by being able to measure larger signals that saturate the standard PMTs, thus enabling measurements closer to the shower axis.

The SPMT are of the model Hamamatsu R8619-22, and are 30 mm in diameter, significantly small in comparison to the 9 in-diameter (~ 23 cm) standard PMTs [10].

3.3.4 Upgraded unified board

The incorporation of new detectors requires new electronics, because the unified board (UB) is only prepared to process 6 signals, which are the high- and low- gain outputs of the SD PMTs. The new electronics, dubbed upgraded unified board (UUB), allows to process the signals of all the detectors of the upgrade because it counts with 10 channels. It also increases the dynamic range and time resolution with respect to the previous electronics, by digitizing the signals at 120 MHz (instead of 40 MHz) with a 12-bit analog-to-digital converter (ADC) (instead of a 10-bit one) [79].

3.3.5 Extension of the fluorescence detector uptime

To acquire more data and increase statistics of hybrid events, the FD will have an additional new operation modality that allows to extend its duty cycle up to $\sim 29\,\%$. This will be achieved by reducing the gain of the PMT cameras of the FD telescopes, enabling them to operate under more intense light conditions, starting earlier and ending later each data-taking night [10].

3.3.6 The underground muon detector

The underground muon detector (UMD) consists of plastic scintillation detectors buried 2.25 m underground. Because the other components of the air showers are largely absorbed in the matter above the detectors (atmosphere and earth), the UMD provides a direct measurement of the muonic component of air showers with energies between $10^{16.5}$ eV and 10^{19} eV. The depth at which the detectors are and the density of the soil determine the mass overburden that the muons have to go through to reach the detector, of $540\,\mathrm{g/cm^2}$, implying a vertical muon energy threshold of ~ 1 GeV. Furthermore, with the developments of this work (see Chapter 4) the UMD can be used to measure the time structure of the muon signal, which allows to reconstruct mass-sensitive parameters.

The UMD will consist of 219 scintillation modules located at 73 positions: 61 locations in the SD-750 array, and 12 in the SD-433 array [80]. A scheme of an SD station together with its three UMD modules can be seen in Fig. 3.8. The set of three UMD modules that correspond to the same SD station are referred to as UMD counter. The modules will lie close to its paired SD station (within 20 m) to probe approximately the same shower density as the corresponding SD station, with its RD, and SSD. The multi-hybrid detection can be used to validate the estimation of the muon component from the rest of the detectors which always measure multiple shower components.

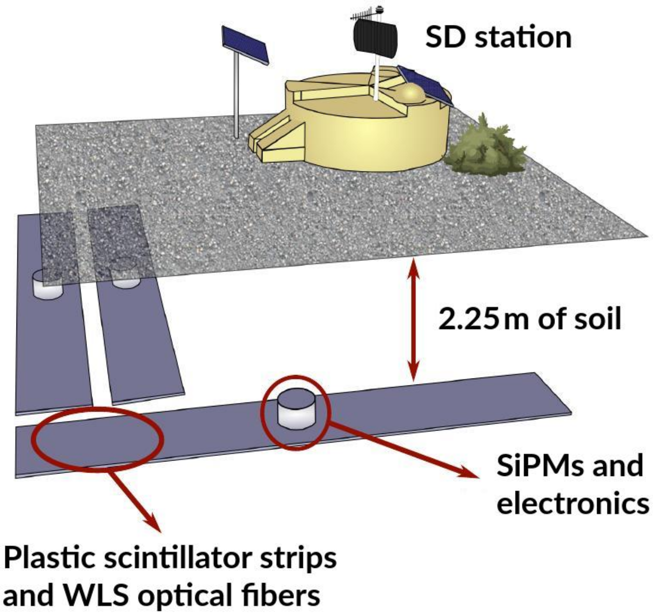


Figure 3.8: Scheme of an SD station and its three UMD modules buried underground. The different parts are identified. Image adapted from Ref. [81].

A UMD module consists of 64 strips of scintillation material (polystyrene doped with fluor) covering an area of 10 m^2. Each module is divided into two identical panels of 32 strips each. Every strip is 4 m long, 4.1 cm wide, and 1.1 cm thick, covered in the outside with a refractive titanium dioxide (TiO_2) layer. Inside each strip, there is a straight extruded hole with a WLS optical fiber. The optical fiber guides the photons towards a cookie, which makes the optical connection to a silicon photo-multiplier (SiPM). It can be added that a previous design of the detector used multipixel photo-multipliers instead of SiPMs [80, 82]. A picture of a UMD module during assembly can be seen in Fig. 3.9. There the scintillation bars (white on the outside) as well as the WLS optical fibers (shining in green) can be seen. The ends of the optical fibers are already placed in the cookie.

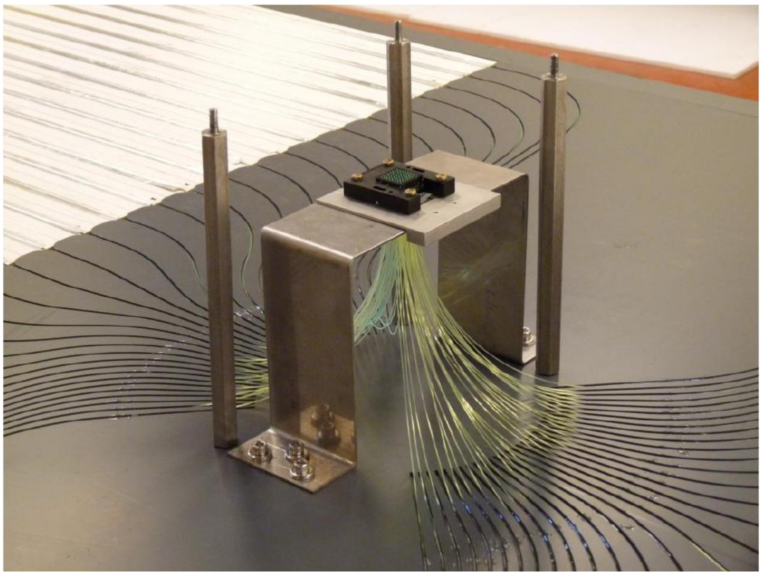


Figure 3.9: Picture of a UMD module under construction. Image extracted from Ref. [83].

The principle of detection of the scintillator detector with WLS fibers is trough fluorescence. The impinging muons excite the electrons of the scintillation material, which de-excite by emitting fluorescence photons. To be more precise, there is a chain of excitation and de-excitations in the scintillator: First, the polystyrene base (Dow Styron 663W) absorbs the incoming radiation and emits photons in the UV spectrum. These photons would attenuate very shortly if they were not absorbed by a first dopant (2,5-diphenyloxazole or PPO), which emits other UV photons with larger attenuation length. Finally the latter are absorbed by a second dopant (1,4-bis(5-phenyloxazol-2-yl) benzene or POPOP) that emits photons of blue color ($\sim$ 420 nm). The mean attenuation length of these fluorescence photons in the scintillator material is between 5 cm and 25 cm. Since the scintillator strips are 4 m long, it is necessary to introduce the WLS optical fiber to conduct the signal to the electronics of the detector. Trying to use a standard optical fiber would be futile, because no photons outside of it could be transported through the fiber. That is why the Saint-Gobain BCF-99-29AMC multi-clad WLS fiber has a scintillation core that absorbs the blue fluorescence photons of the scintillation material and emits green photons. Some of the latter ($\sim$ 10 %) remain inside of the fiber thanks to the two layers of transparent cladding material outside of it with lower refractive indexes than the core. These green photons propagate through the fiber core (by total internal reflection), and the ones that do not suffer attenuation reach the SiPM [81]. Fig. 3.10 shows a scheme of the scintillation strip with the WLS optical fiber (left panel), and a scheme of the fiber (right panel). In the latter, the refractive indexes of the core and claddings are indicated.

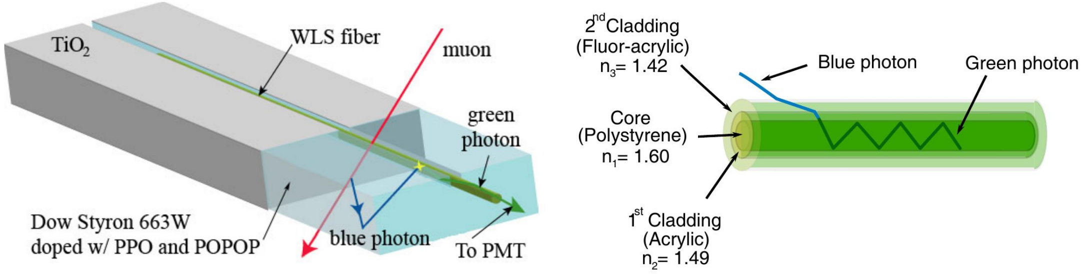


Figure 3.10: Left: Scheme of a scintillator strip. Right: Scheme of the WLS optical fiber. Images extracted from Ref. [83] and Ref. [81] respectively.

The SiPM is a solid state photodetector consisting of an array of avalanche photo-diodes implemented on common silicon substrate operated in Geiger mode. The model used in the UMD is HAMAMATSU S13361-2050NE-08. Each SiPM contains 1584 cells of photo-diodes to which a reverse voltage is applied. This voltage is larger than the breakdown voltage of the diodes to operate in Geiger mode. In this mode, the impinging photons can generate a multiplication of carriers with infinite gain (avalanche), constituting an electric current. The multiplication process eventually stops due to a resistor that quenches the otherwise self-sustaining avalanche process. It is relevant to add that we refer to the number of triggered cells in a SiPM as the number of photon-equivalents [81]. Fig. 3.11 shows a scheme and a picture of the cookie (left panel) and of the SiPM array (right panel).

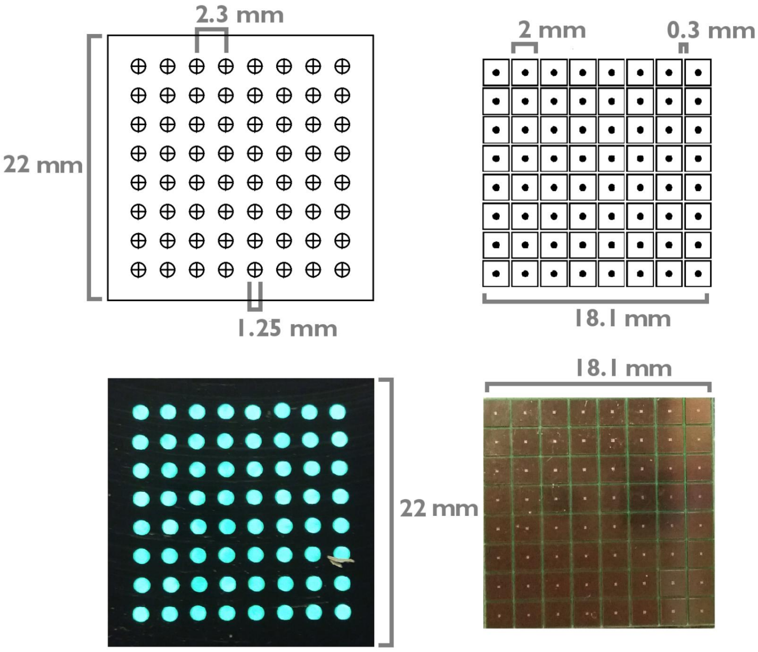


Figure 3.11: Left: Scheme and picture of the cookie or optical connector. Right: Scheme and picture of the SiPM array. Images extracted from Ref. [81].

The electric current produced by the SiPMs is processed with the readout electronics of the UMD. The readout electronics consists of two modes: the *counter* and the *integrator* modes. In combination, they achieve a broad dynamic range, being the counter mode appropriate for small signals (down to single muons), while the integrator mode can process large signals (hundreds of muons). Both modes work simultaneously measuring the very same muons.

In the counter mode, the signal of the 64 SiPMs are processed independently by a pre-amplifier, a fast-shaper, a discriminator, and a field programmable gate array (FPGA). The first three are built within each channel of two 32-channel application-specific integrated circuits (ASICs), model CITIROC from WEEROC. The latter, the FPGA, samples the 64 signals independently every 3.125 ns and converts them into 2048-bit long traces (i.e. 6.4 μs). Every bit of the trace can be a “1” or a “0” depending on whether the signal passes the 2.5 PE (77.5 mV) threshold. A muon then appears as a sequence of 1s in the trance of a channel. Fig. 3.12 shows the simulated response of the electronics to one impinging muon. The muon generates several photon-equivalents that appear as spikes in the top panel. The input current is amplified and inverted by the pre-amplifier (second to top panel). Then it is further amplified, and inverted by the fast shaper (second to bottom panel). Finally, it is digitized by the FPGA (bottom panel).

In the integrator mode, the 64 signals are analogically added and the sum is amplified with low- and high- gain amplifiers. The two resulting signals are sampled every 6.25 ns by two analog-to-digital converters (ADCs) that output two 1024-bit long traces (i.e. also 6.4 μs). In this mode, the signal charge (the sum of the ADC counts in a time window of the trace) is divided by the mean charge of a muon to convert it to a number of muons [81]. In this work, we focus on the counter mode of the UMD.

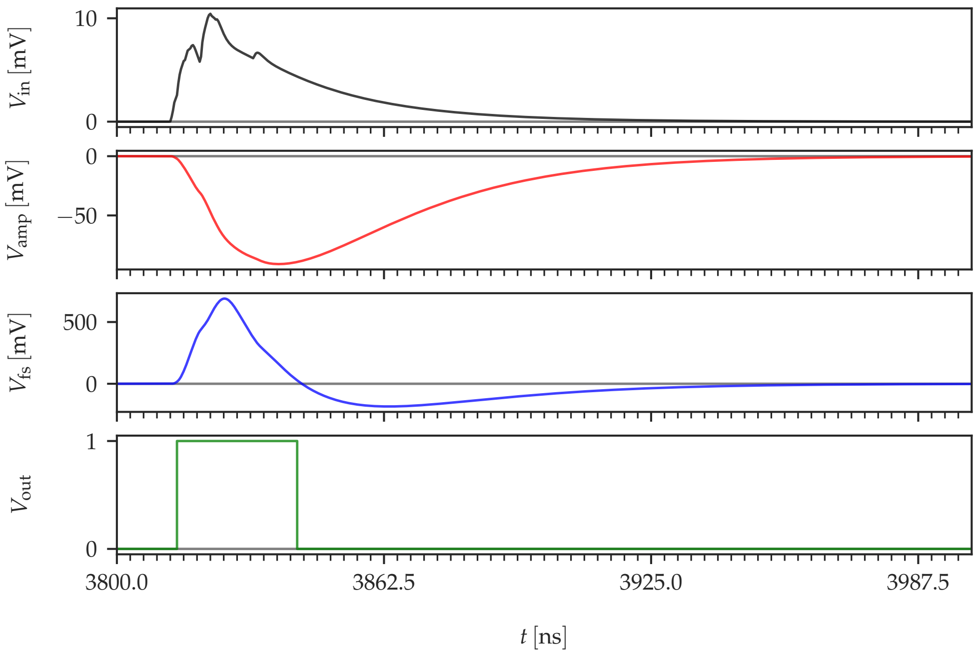


Figure 3.12: Simulation of the response of the electronics to one muon impinging on a scintillation bar. The muon generates photon-equivalents (top panel). The signal is then amplified and inverted by the pre-amplifier (second to top panel). Afterwards, the fast shaper further amplifies and inverts again the signal (second to bottom panel). Finally, the FPGA digitizes the signal (bottom panel).

Having understood how the UMD works, we discuss the UMD event reconstruction for the counter mode, to which a substantial part of this work is devoted. We can retrieve the binary traces of every active channel of the UMD modules that participate in an event. These binary signals are matched to a single-muon pattern of 0s and 1s of the form “1111xxxxxxxx”, where x can take the values 0 or 1 [1, 80]. This pattern maximizes the signal-to-noise ratio, and was the subject of the study of Ref. [81]. Fig. 3.13 presents a module-level simulated event, showing the impinging muons as a function of time, the binary traces, and the pattern matches for the channels with non-null signal.

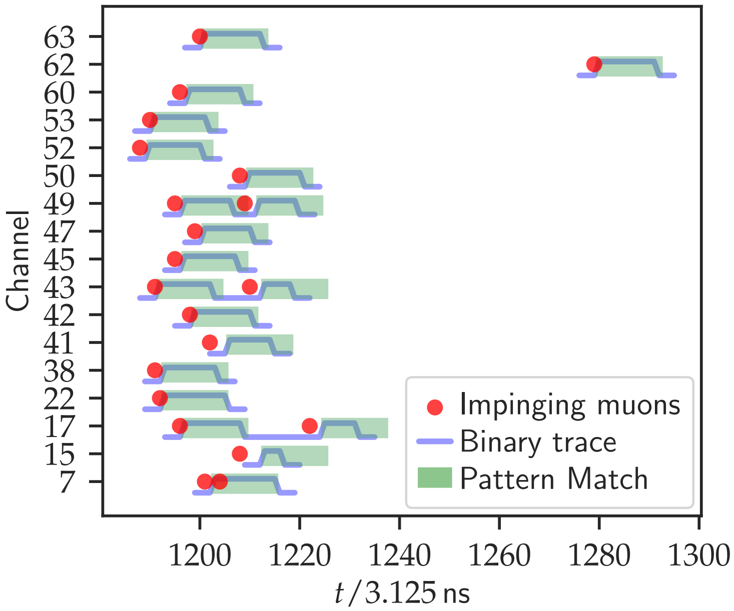


Figure 3.13: Simulation of an air shower event at module level. The muons (red circles) impinge on the different segments of the detector at a certain time (x-axis), which correspond to a given channel (y-axis). The muons generate binary signals on their corresponding channel (blue lines), which are matched to a pattern of the kind "1111xxxxxxxx" (green rectangles).

Obtaining an estimated number of muons from the number of pattern matches is done through a counting strategy. The number of pattern matches cannot be directly used as an estimate of the counts due to the pile-up effect. This happens when two or more particles hit one scintillator strip almost simultaneously, such that they are read out as only one particle. It constitutes an under-counting effect. In Fig.3.13 we can see an example of pile-up in channel 7, where the signal of two impinging muons is matched to only one pattern. The way to obtain a number of counts accounting for the effect of pile-up is treated in Chapter 4.

Even having a nearly pile-up-unbiased estimate of the number of muons impinging on a UMD module, the estimate can still suffer from biases due to the effect of corner-clipping muons. These are muons that transverse two neighboring scintillator strips and deposit enough signal in both strips such that one particle is read out as two, constituting an over-counting effect. The way to correct for this effect, as well as for other sources of bias, in treated in Chapter 5.

After having an unbiased estimate of the number of muons, an estimate of the muon density at each module is obtained simply by dividing the reconstructed counts by the effective area of the detector. Incorporating the geometry reconstructed using the SD, the distance to the shower axis on the shower plane can be known for each module. Each event provides a sample of the lateral distribution function (LDF) of muons, this is, the muon densities as a function of the distance to the shower axis. Typically the sampled distribution is fitted, and the fitted function is evaluated at a fixed distance to the shower axis (450 m for the UMDs in the SD-750 array), providing an event-wise estimate of the muonic component size which can be used for further analyses. A critical point to obtain reliable estimates of the muonic size is to use events that sample "well" the muon LDF around the fixed distance. To select such events in the $\overline{\text{Off}}\underline{\text{line}}$ framework, we developed a tool that is described in Appendix A. An example of a muon LDF fit to an event can be seen in the left panel of

Fig. 3.14. The event identification number is 180685784700, and it was detected on March 10th, 2018. For visualization, the average muon density of the three modules of a counter are displayed (instead of that of each module). The fitting function used here is a modified NKG (see Eq. 3.2) with fixed $\alpha = 1$, $\gamma = 1.85$, $r_s = 150\,\mathrm{m}$, and $r_{\mathrm{opt}} = 450\,\mathrm{m}$. There are different functions that can be used to describe the muon LDF, but we delay that discussion to Chapter 6.

As mentioned at the beginning of this section, the developments of this work allow to reconstruct the time structure of the muon signal (see Chapter 4). We show as an example in the right panel of Fig. 3.14 the reconstructed number of muons (as seen by the detector) as a function of time for one module of the counter with largest signal of the event shown in the left panel.

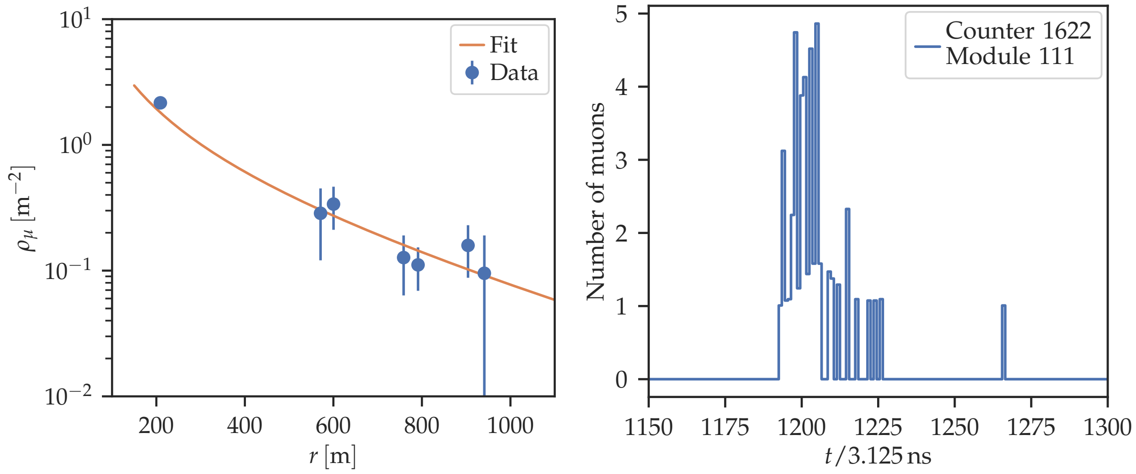


Figure 3.14: Left: Muon density as a function of the distance to the shower axis for an example event. On a solid line, the fit of a muon LDF. Right: Reconstructed muon density as a function of time for counter 1622 (the closest to the shower core), module 111 of the same event.

3.4 Summary

In this chapter we introduced the Pierre Auger Observatory and the detectors that conform it: The surface detector, the fluorescence detector, and within the upgrade of the observatory, known as Auger Prime, the scintillator surface detector, the radio detector, and the underground muon detector. We also described the upgraded electronics, the incorporated small photomultiplier tubes, and the extension of the fluorescence detector uptime, which are also part of the upgrade of the observatory.

We explained how the surface detector is an array of water Cherenkov detectors (the largest of its kind), and explained how the energy and geometry are reconstructed from the measured signals. With respect to the fluorescence detector, we explained that the telescopes measure the fluorescence light emitted by the de-excitation of the nitrogen in air, after the excitation due to interactions with the particles of air showers. We also described how the energy can be reconstructed from the measured light.

Moreover, we described with particular detail the underground muon detector. We discussed the principle of detection of scintillation detectors and detailed the parts of its modules and of its electronics. Finally, we explained the method used for the reconstruction of the muon content of measured air showers, which is directly relevant for Chapters 4 and 5.

Chapter 4

Muon counting strategy

As discussed in Sec. 3.3.6, when reconstructing the muon densities measured by the Underground Muon Detector (UMD) in the counter mode, the first step is to identify the single-muon pattern matches in each channel of each module that takes part of an event. Subsequently, the number of pattern matches are converted to an estimate of the number of muons via a counting strategy. The most trivial strategy would consist of taking the number of pattern matches as the estimate of the number of impinging muons. However, such an estimate would be biased by the effect of pile-up. This occurs when two or more muons transverse the same scintillator strip almost simultaneously, creating signals that cannot be resolved. Consequently, these muons are identified as only one pattern match, which can lead to under-counting. It follows that any reasonable counting strategy should consider the pile-up effect in its design, by understanding that one pattern match should account for one to infinite impinging muons.

In this chapter we show the development of a new counting strategy for the UMD. The motivation behind it is that the estimated number of muons in the $\overline{\text{Off}}\underline{\text{line}}$ UMD reconstruction was reported to have a significant bias which increased with the input number of muons [84]. Although a correction of this bias was proposed for the reconstruction of detector simulations, the same correction could not be applied to the reconstruction of data. This is because the correction was based on knowing the impinging time of the muons, information that is unknown for real data.

Figure 4.1 shows the relative bias of the reconstructed number of muons against the input (Monte-Carlo truth) number of muons, where the counting algorithm was the one used in $\overline{\text{Off}}\underline{\text{line}}$ before this work. The bias is computed by simulating impinging muons from the signals of proton EPOS-LHC [26] air shower simulations with $18.0 \leq \log_{10}(E/\text{eV}) \leq 18.2$ and $33^\circ \lesssim \theta \lesssim 39^\circ$ ($0.30 \leq \sin^2\theta \leq 0.40$). The details of how this algorithm works and how the simulations are performed are delayed to the following sections. At this point it is important to notice that the relative bias is significant, tending to more negative values for larger input number of muons. In this example we observe almost a $\sim -40\,\%$ bias for 200 input muons.

The developed strategy for counting muons in the UMD is actually an application of a solution to the more general problem of counting particles in segmented counters as a function of time. In other words, the new strategy could be used in other particle or astroparticle detectors. Furthermore, and most remarkably, it allows to reconstruct the muon signal as a function of time as seen by the detector to a single time-bin resolution. This can be used to reconstruct mass-sensitive observables from the timing information of the muon signal, opening the door to new analyses.

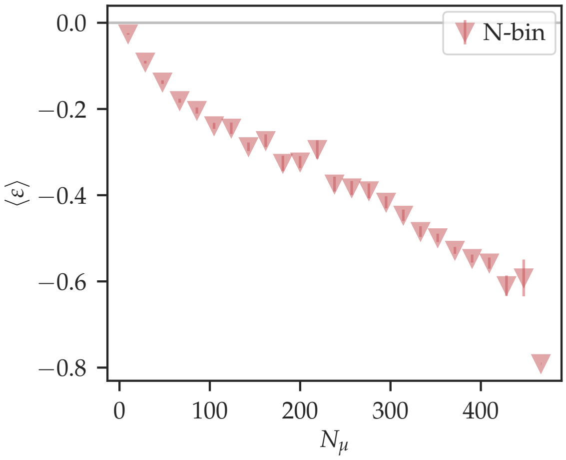


Figure 4.1: Relative bias in the reconstructed number of muons with respect to the input (Monte-Carlo truth) number of muons, as a function of the input number of muons, for the strategy that was implemented in $\overline{\text{Off}}\underline{\text{line}}$ before this work. The bias is calculated from different random samples generated from the average muon density as a function of time and distance to the shower axis of proton EPOS-LHC showers of $18.0 \leq \log_{10}(E/\text{eV}) \leq 18.2$ and $33^\circ \lesssim \theta \lesssim 39^\circ$ ($0.30 \leq \sin^2\theta \leq 0.40$). A gray horizontal line marks null bias.

In this chapter, we test and compare the performance of the new muon counting algorithm against three other methods, including the one previously used in $\overline{\text{Off}}\underline{\text{line}}$. The structure of the chapter is as follows. We explain the various statistical models or algorithms that provide estimates of the impinging number of muons in Sec. 4.1. In Sec. 4.2, we explain how we generate the air shower library and how we simulate the detector response. We compare the performance of the various methods in Sec. 4.3, where we also show an example of the reconstructed muon signal as a function of time. Finally, we summarize the results and provide an outlook in Sec. 4.4.

It is relevant to add that in the analysis of this chapter we do not consider corner-clipping muons as they are an independent source of bias, with an effect that can be (and is) corrected after the counting strategy is applied. This is the subject of study of Chapter 5.

The analysis shown in this chapter is largely based on the **publication**:

- *Estimation of the number of counts on a particle counter detector with full time resolution*, F. Gesualdi, A.D. Supanitsky, Eur. Phys. J. C 82, 925 (2022) (Ref. [85]),

and on the Pierre Auger **internal publication**:

- *A new pile-up correction strategy for the Underground Muon Detector*, F. Gesualdi, A.D. Supanitsky, GAP 2022 − 001 (Ref. [86]).

4.1 Counting strategies

On a first approximation, the problem of counting muons (or particles in general) on a segmented muon counter like the UMD (or in a segmented detector in general) is analogous to the statistical problem of counting balls in boxes. The terms "classical occupancy problem" and "classical shot problem" are also used to describe said statistical problem [87]. In this problem, a finite number of balls (the impinging particles) are allocated uniformly in a finite number of boxes (the segments of the detector). After one realization of the experiment, each box may have zero balls (no pattern match) or one or more balls (with pattern match or occupied). We define as occupancy the number of occupied boxes after one realization of the experiment.

Applied to the UMD counter, the balls would be the number of impinging muons N_μ. If we think of an air shower as a stochastic process, N_μ, a property of the event, would be the result of one realization of a Poisson distribution of mean μ, the mean number of particles expected at the UMD module for that air shower, which is a property of the air shower. The cosmic ray primary, energy, zenith angle, and the distance of the module to the shower axis on the shower plane determine μ. It can be expressed as $\mu = \rho_\mu A \cos\theta$, being ρ_μ the mean muon density of an air shower on the shower plane, A the active area of the UMD module, and θ the zenith angle of the air shower. Understanding the distinction between N_μ and μ is important: N_μ is the number of muons actually impinging the detector, while μ is the mean number. As discussed in Chapter 6, the muon lateral distribution function can be reconstructed using either one of them [13, 14, 55].

There are two questions that the counting strategies aim to answer: (1) What is the estimated number of impinging muons N_μ, given that we know the occupancy? (2) What is the estimated mean number of particles μ, given that N_μ is a realization of the Poisson distribution with mean μ? The counting strategies address these questions using the information of the pattern matches in an event trace, and as a result provide estimates of N_μ and μ.

In the following we describe the four counting different strategies considered in this analysis.

4.1.1 Infinite window strategy

This strategy was employed by the AGASA collaboration (see for example Ref. [12]). It is the simplest strategy because it does not use the timing information of the muon trace (hence its name). Instead, it only uses the occupancy k, which is computed as [13, 55, 88]

$$k = \sum_{i=1}^{n_s} \Theta(m_i), \tag{4.1}$$

where n_s is the number of active segments of the detector, Θ is the Heaviside step function, and m_i is the number of starting pattern matches of the i-th channel. If $m_i = 0$, $\Theta(m_i) = 0$, otherwise $\Theta(m_i) = 1$. For this strategy, k is equivalent to the number of channels with at least one pattern match.

The number of impinging muons on a scintillator strip follows a Poisson distribution of parameter μ/n_s. Therefore, the probability of a channel being empty is $q = \exp(-\mu/n_s)$, and the probability of a channel being occupied is $p = 1 - q$ [13, 14]. The probability of having k occupied channels (successes) out of n_s channels (trials) given μ (which determines the success probability) follows a binomial distribution [13, 14]

$$B(k|\,\mu) = \binom{n_s}{k} p^k q^{n_s-k} = \binom{n_s}{k} e^{-\mu} (e^{\mu/n_s} - 1)^k. \tag{4.2}$$

The expression for the likelihood of μ given k occupied channels is identical to Eq. (4.2). For $k < n_s$, the maximum likelihood estimator of μ is given by [13, 14]

$$\widehat{\mu} = -n_s \ln\left(1 - \frac{k}{n_s}\right). \tag{4.3}$$

Furthermore, the probability of k channels being occupied given that N_μ muons impinged the detector and that there are n_s scintillation bars follows the occupancy distribution [88]

$$Occ\,(k|\,N_\mu, n_s) = \binom{n_s}{k} \frac{S(N_\mu, k)}{n_s^{N_\mu}} \quad \forall\, k \in \mathbb{Z} \mid 1 \le k \le n_s,$$

where $S(N_\mu, k)$ are the Stirling numbers of the second kind (see also Ref. [87]). As shown in Refs. [55, 88], the maximum likelihood estimator of N_μ can be approximated as

$$\widehat{N}_\mu = \frac{\ln\left(1 - \dfrac{k}{n_s}\right)}{\ln\left(1 - \dfrac{1}{n_s}\right)}. \tag{4.4}$$

Eq.(4.3) and Eq.(4.4) are evidently very similar, and in the limit when n_s tends to infinity, $\widehat{N}_\mu$ tends to $\widehat{\mu}$.

It can be added that a module is said to be saturated when $k = n_s$. In such case, both $\widehat{\mu}$ and $\widehat{N}_\mu$ tend to infinity.

4.1.2 N-bin window strategy

This is the strategy that was implemented in the standard reconstructions in $\overline{\text{Off}}\underline{\text{line}}$ before this work [1], and it is the one used in Fig. 4.1. It was first described in Ref. [14] and has been used for analyses of UMD PMT data in Refs. [89, 90][2].

In this strategy, the event trace is partitioned into N-bin-long time windows, where N is equal to the number of bins in a single-muon pattern (12 time-bins of 3.125 ns for the UMD with SiPMs). The last window can actually be shorter if N is not an exact divider of the number of bins of the trace. This is the case for the UMD, where the trace is 2048-bins long. In such case, the last window is nonetheless taken into account in the calculation. We then compute for each j-th window the number of occupied channels, k_j, and use Eqs. (4.1), (4.3) and (4.4) to compute for each window the estimated number of muons, $\widehat{N}_{\mu,j}$, and the estimated mean muon number, $\widehat{\mu}_j$. The total $\widehat{N}_\mu$ and $\widehat{\mu}$ are calculated by adding the estimates of each window:

$$\widehat{N}_\mu = \sum_{j=1}^{n_w} \widehat{N}_{\mu,j}, \tag{4.5}$$

$$\widehat{\mu} = \sum_{j=1}^{n_w} \widehat{\mu}_j, \tag{4.6}$$

where j runs over the n_w number of windows of the trace, 171 in this case.

Saturation is flagged if at any window j the condition $k_j = n_s$ is fulfilled.

---

[1] In the $\overline{\text{Off}}\underline{\text{line}}$ framework, the standard application for reconstructing UMD data is `MdDataReconstruction`, and the standard applications for reconstructing detector simulations are `MdSdSSdReconstruction`, and `MdSdFdReconstruction`.

[2] As shown in the introduction to this chapter, or in Sec. 4.3, this strategy introduces significant biases. These are compensated in the analyses of Refs. [89, 90], at least to first order, by correcting against simulations, using a method very similar to the one presented in Chapter 5.

4.1.3 N-bin centered window strategy

This method is utilized in this study to understand the causes of the biases produced by the N-bin window strategy. It is also very similar to the original correction attempting to address those biases that was developed for the reconstruction of simulations [84].

It is very similar to the N-bin strategy, but it positions the windows such that one is centered at the peak of the signal. The idea is that there exists at least one 12 time-bin window j^* where the occupancy k_j is maximum. If there are multiple of them, we take the one that starts earliest. To locate it, we slide a 12 time-bin window over all the trace, computing k_j for every j-th possible window. The earliest window that fulfills $k_{j^*} = \max_j(k_j)$ is the "centered" window. The way to complete the partition of the trace into 12 time-bin windows is determined by this centered window. Since the event signal in the trace normally begins at around 3300 ns or 3800 ns [81], we ignore the first bins of the trace that do not complete a full 12-bin window, but keep the last bins in this condition. This approach is generalizable for different time window sizes, with the option to keep or not incomplete windows.

Having partitioned the trace into windows, we use Eqs. (4.5) and (4.6) to compute $\widehat{N}_\mu$ and $\widehat{\mu}$, respectively. Depending on the partition, the number of windows n_w can be 170 or 171.

Just like for the N-bin strategy, saturation is given if $k_j = n_s$ for any $1 \leq j \leq n_w$.

4.1.4 1-bin window strategy

This strategy is the one developed in this work. This approach exploits the entire time structure of the signal rather than in many-bin windows (like the N-bin or N-bin centered strategies) or as a whole (like the infinite window strategy). The idea is to determine for each j-th time-bin of the trace, not just the number of occupied channels k_j, but also the number of inhibited channels $n_{\mathrm{inhib},j}$. The latter represents the number of channels with a pattern match that began in an earlier time-bin. These are treated as inhibited channels since it would be impossible to measure or resolve a muon that fell within the inhibition window (i.e., the single-muon pattern match); in other words, inhibited channels are effectively dead channels. In the analogy with the balls in boxes problem, having channels that are inhibited is equivalent to having fewer boxes. The number of non-inhibited channels is equal to $n_s - n_{\mathrm{inhib},j}$.

Additionally, as the number of available (non-inhibited) channels decreases, so does the size of the detector, resulting in a fewer number of detectable muons. Evidently, the detection area changes from bin to bin. In order to determine the number of muons that would be observed for a constant detector area equal to the active area, we multiply by the number of active segments n_s and divide by the number of non-inhibited segments $n_s - n_{\mathrm{inhib},j}$. We then compute for each j-th bin of the trace

$$\begin{aligned}\widehat{\mu}_j &= -\left(n_s - n_{\mathrm{inhib},j}\right) \ln\left(1 - \frac{k_j}{n_s - n_{\mathrm{inhib},j}}\right) \times \frac{n_s}{n_s - n_{\mathrm{inhib},j}} \\ &= -\, n_s \ln\left(1 - \frac{k_j}{n_s - n_{\mathrm{inhib},j}}\right), \end{aligned} \tag{4.7}$$

and

$$\widehat{N}_{\mu,j} = \frac{\ln\left(1 - \dfrac{k_j}{n_s - n_{\mathrm{inhib},j}}\right)}{\ln\left(1 - \dfrac{1}{n_s - n_{\mathrm{inhib},j}}\right)} \times \frac{n_s}{n_s - n_{\mathrm{inhib},j}}. \tag{4.8}$$

The total $\widehat{N}_\mu$ and $\widehat{\mu}$ are computed as their sum across all the bins of the trace, i.e., as in Eqs. (4.5) and (4.6), taking the number of windows n_w equal to the trace length (2048).

In this strategy, the detector is said to be saturated if, at any time bin, the occupancy equals the number of available channels: $k_j = n_s - n_{\text{inhib},j}$ for any $1 \le j \le n_w$.

4.1.5 Summary of the counting strategies

The characteristics of the four strategies are summarized in Table 4.1. Although the number of windows n_w for each method are specific to SiPM UMD signals, the methods can be used to process the signals of other segmented counters mutatis mutandis.

Strategy	$\widehat{\mu}$	$\widehat{N}_\mu$	n_w	Centered
Infinite	$-n_s \ln\left(1 - \frac{k}{n_s}\right)$	$\dfrac{\ln\left(1 - \frac{k}{n_s}\right)}{\ln\left(1 - \frac{1}{n_s}\right)}$	1	-
N-bin	$-\sum_{j=1}^{n_w} n_s \ln\left(1 - \dfrac{k_j}{n_s}\right)$	$\sum_{j=1}^{n_w} \dfrac{\ln\left(1 - \frac{k_j}{n_s}\right)}{\ln\left(1 - \frac{1}{n_s}\right)}$	171	No
N-bin centered	$-\sum_{j=1}^{n_w} n_s \ln\left(1 - \dfrac{k_j}{n_s}\right)$	$\sum_{j=1}^{n_w} \dfrac{\ln\left(1 - \frac{k_j}{n_s}\right)}{\ln\left(1 - \frac{1}{n_s}\right)}$	170-171	Yes
1-bin	$-\sum_{j=1}^{n_w} n_s \ln\left(1 - \dfrac{k_j}{n_s - n_{\text{inhib},j}}\right)$	$\sum_{j=1}^{n_w} \dfrac{n_s}{n_s - n_{\text{inhib},j}} \dfrac{\ln\left(1 - \frac{k_j}{n_s - n_{\text{inhib},j}}\right)}{\ln\left(1 - \frac{1}{n_s - n_{\text{inhib},j}}\right)}$	2048	-

Table 4.1: List of the formulas and characteristics of the four counting strategies considered. The number of windows for each method n_w is calculated considering the length of the SiPM UMD trace and of a single-muon pattern. It is specified in the column "Centered" if the windows are positioned so that the center of one aligns with the signal peak.

It can be added that the performance of the different methods for the reconstruction of N_μ or μ in real data could possibly be optimized by selecting a window within the trace where the signal lies or is expected to lie. This could help improve the signal-to-noise ratio. However, defining the edges of such a window for all events is not trivial, and is a subject of future work. Moreover, if a variable window-size is used, background rates should be accounted for also with care. For the analysis of this chapter, where we do not simulate noise, this is unnecessary and we thus keep the whole trace. Furthermore, we also keep the full trace for the analysis of real data because the general bias correction described in Chapter 5 can easily account for the extra counts caused by background muons as long as the window is mostly of the same length for all traces (i.e. constant background rate). In the 1-bin strategy, the variation of the analyzed trace length are below 1 %, and thus we can well approximate the background rate as constant.

4.2 Simulation of the response of the underground muon detector to air showers

In order to compare the different counting strategies in a realistic scenario, we generate a library of Monte-Carlo air shower simulations, and we also simulate the response of the UMD to said air showers. In the following we first describe how the detector is modelled. We also dedicate a special subsection to showing the effects of the detector that influence the biases discussed in Sec. 4.3. Furthermore, we explain how the air shower library is generated, and how it is used in combination with the detector simulations. Finally, we show a validation of the simulation chain by studying the expected and observed saturation fraction.

4.2.1 Simulation of the detector response

In this subsection we describe the simulation of the counter mode of the UMD with SiPMs. The detector simulations take as input the number of impinging muons and their impinging times with respect to the beginning of the signal. With this information, the detector simulation outputs the event trace of the module, i.e., the digital signals of the 64 channels. The model of the detector follows Ref. [81, 91]. Assuming the given model, we derive an analytical solution for the detector response to one or more muons.

For deriving such analytical solution, we neglect noise, which is irrelevant for evaluating the effectiveness of the counting strategies. Taking into account all sources, background noise is estimated to produce $\sim 5.5\,\%$ of the counts in a module trace. The optimal single-muon pattern is affected by the background muon and noise count rates, which are inherent to the environment and the detector, and which is already optimized to such conditions. However, because the counting strategies use already-matched patterns as input, the background is irrelevant when evaluating them. Even if noise had any net effect in the estimation of the number of muons, it would be corrected in the general bias correction of Chapter 5. As a reminder, we reiterate that we also do not consider corner-clipping muons as it is a separate source of bias that can be also corrected for at the mentioned later step (c.f. Chapter 5).

We start with assigning the start-time of the UMD signal by sampling it from its distribution. The start-time of the UMD signal is not given at a fixed time-bin, but rather it is determined by the trigger time of the SD station. The latter is given by the timing of both the electromagnetic and muonic shower front, while the start-time of the UMD signal is only related to the muonic shower front, and additionally both are subject to Poission fluctuations. Furthermore, the effective delay between the SD trigger and the start-time of the SD signal is different for different types of triggers, because their sensitivity with respect to the different shower components is different. We sample the start-time of the UMD signal from a double Gaussian distribution $0.27 \times \mathcal{N}(3337\,\mathrm{ns}, 91\,\mathrm{ns}) + 0.73 \times \mathcal{N}(3757\,\mathrm{ns}, 70\,\mathrm{ns})$ [81]. The first Gaussian corresponds to the delay with respect to a T2 time-over-threshold trigger in the paired SD station, whereas the second Gaussian corresponds to that of a T2 threshold trigger.

Having defined the start-time of the signal, we introduce now the impinging muons of which we know the impinging time with respect to the start-time. Then, for each muon we assign a scintillator strip and the corresponding channel by sampling a discrete uniform distribution $\mathcal{U}\{1, 64\}$. We also assign an impinging position in the scintillator strip l by sampling a continuous uniform distribution $\mathcal{U}\{0\,\mathrm{m}, 4.01\,\mathrm{m}\}$. We compute the average number of photo-electrons $\langle N_{\mathrm{PE}} \rangle$ produced by said muon at position l that reach the SiPM as [81]

$$\langle N_{\mathrm{PE}} \rangle (l) = 17.4\, e^{-l/4.16\,\mathrm{m}} + (1.0 - 17.4)\, e^{-l/0.037\,\mathrm{m}}. \tag{4.9}$$

The actual number of photo-electrons reaching the SiPM produced by said muon N_{PE} is computed by sampling a Poisson distribution of mean $\langle N_{\mathrm{PE}} \rangle$.

To compute the time delay of the photo-electron with respect to the impinging time of the muon, we calculate the time it takes the photon to propagate from the impinging position in the optical fiber to the SiPM. For this we divide the distance l by the speed of light in the optical fiber ($0.60\,c$). The photo-electron has two other sources of time delay: the time delays due to the excitation and de-excitation of the scintillator and of the core of the optical fiber. These are sampled from two exponential distributions of parameters 3.7 ns and 3.5 ns respectively [81]. The resulting time of the photo-electron t_{PE} is the addition of the signal start time, the impinging time of the muon (measured with respect to the signal start time), the propagation time, and the scintillator and fiber delays.

Additionally, we analytically model the electronics response to each photo-electron. The electronics consist of: a pre-amplifier, a fast shaper, a discriminator, and a Field-Programmable Gate Array (FPGA). Whereas the mathematical models for the photo-electron pulse as well as for each component of the electronics are taken from Ref. [81], we provide the analytical solution of the electronics response to a photo-electron signal. For this we neglect baseline noise.

The input signal in the electronics as a function of time caused by a photo-electron, $V_{\mathrm{PE}}(t)$, is modelled as [91]

$$\begin{aligned} V_{\mathrm{PE}}(t) =& 0.29\,\mathrm{mV} \left(1 - e^{-\frac{t-t_{\mathrm{PE}}}{3.82\,\mathrm{ns}}}\right) \left[23.22\, e^{-\frac{t-t_{\mathrm{PE}}}{1.187\,\mathrm{ns}}} + 1.609\, e^{-\frac{t-t_{\mathrm{PE}}}{23.44\,\mathrm{ns}}} + e^{-\frac{t-t_{\mathrm{PE}}}{0.221\,\mathrm{ns}}}\right] \times \Theta(t - t_{\mathrm{PE}}), \\ =& \sum_{i=1}^{6} A_i\, e^{-\frac{t-t_{\mathrm{PE}}}{\tau_i}}\, \Theta(t - t_{\mathrm{PE}}). \end{aligned} \tag{4.10}$$

The next element is the pre-amplifier, which we model as a low-pass filter [81, 91]. The amplified signal $V_{\mathrm{amp}}(t)$ is obtained by first Fourier transforming $V_{\mathrm{PE}}(t)$, multiplying by the low-pass filter transfer function $H_I(\omega)$, and then inverse transforming. As a result

$$\begin{aligned} V_{\mathrm{amp}}(t) &= \sqrt{2\pi}\, \mathcal{F}^{-1} \left[\mathcal{F}[V_{\mathrm{PE}}(t)] \times H_I(\omega)\right], \\ &= \sqrt{2\pi} \sum_{i=1}^{6} \mathcal{F}^{-1} \left[\frac{A_i}{\sqrt{2\pi}} \frac{e^{-it_{\mathrm{PE}}\omega}}{(1/\tau_i + i\omega)} \frac{k_I}{\sqrt{2\pi}(1 + i\omega\tau_I)} \right], \\ &= \sum_{i=1}^{6} \frac{k_I A_i \tau_i}{\tau_i - \tau_I} \left(e^{-\frac{t-t_{\mathrm{PE}}}{\tau_i}} - e^{-\frac{t-t_{\mathrm{PE}}}{\tau_I}} \right) \Theta(t - t_{\mathrm{PE}}), \\ &= \sum_{i=1}^{6} \sum_{j=1}^{2} \tilde{A}_{i,j} e^{-\frac{t-t_{\mathrm{PE}}}{\tau_{i,j}}}\, \Theta(t - t_{\mathrm{PE}}), \end{aligned} \tag{4.11}$$

where $k_I = -17.5$ and $\tau_I = 17\,\mathrm{ns}$ [91]. For deriving Eq. (4.11) we used that the Fourier transform commutes with the sum. In the last equation we group the coefficients such that $\tilde{A}_{i,1} = k_I A_i \tau_i / (\tau_i - \tau_I)$, $\tilde{A}_{i,2} = -\tilde{A}_{i,1}$, $\tau_{i,1} = \tau_i$, and $\tau_{i,2} = \tau_I$. Notice that the functional form of $V_{\mathrm{amp}}(t)$ is identical to that of $V_{\mathrm{PE}}(t)$.

A fast-shaper processes the signal after the pre-amplifier. The former is modelled as a practical differentiator [81, 91]. To obtain the signal after the fast-shaper $V_{\mathrm{fs}}(t)$, we first transform Laplace $V_{\mathrm{amp}}(t)$, multiply by the transfer function of the practical differentiator $H_{II}(s)$, and inverse transform:

$$\begin{aligned} V_{\mathrm{fs}}(t) =& \mathcal{L}^{-1}\left\{\mathcal{L}[V_{\mathrm{amp}}(t)] \times H_{II}(s)\right\}, \\ =& \sum_{i=1}^{6}\sum_{j=1}^{2} \mathcal{L}^{-1}\left\{ \frac{\tilde{A}_{i,j} e^{\frac{t_{\mathrm{PE}}}{\tau_{i,j}}}}{s+1/\tau_{i,j}} \times \left[\Theta(t_{\mathrm{PE}})\, e^{-t_{\mathrm{PE}}(s+1/\tau_{i,j})} + \Theta(-t_{\mathrm{PE}})\right] \times \frac{(-s)k_{II}\tau_{II}}{(1+s\tau_{II})^2} \right\}, \\ =& \sum_{i=1}^{6}\sum_{j=1}^{2} \frac{-\tilde{A}_{i,j}k_{II}\tau_{i,j}}{\tau_{II}(\tau_{i,j}-\tau_{II})^2} \times \left\{ e^{-\frac{t-t_{\mathrm{PE}}}{\tau_{II}}} \left[(t-t_{\mathrm{PE}})(\tau_{i,j}-\tau_{II}) + \tau_{II}^2\right] - e^{-\frac{t-t_{\mathrm{PE}}}{\tau_{i,j}}} \tau_{II}^2 \right\} \Theta(t-t_{\mathrm{PE}}), \end{aligned} \tag{4.12}$$

where $k_{II} = 47.1$ and $\tau_{II} = 2.4\,\mathrm{ns}$.

By adding the contribution to $V_{\mathrm{fs}}(t)$ of all the photo-electron pulses caused by each muon impinging in the same scintillator strip, we may finally simulate the response of the discriminator and FPGA, yielding the output signal or binary trace $V_{\mathrm{out}}(t)$ for each channel. If $V_{\mathrm{fs}}(t)$ exceeds the threshold (77.5 mV) for more than 1.51 ns, the FPGA output is a 1 in the binary trace. Additionally, the FPGA samples the signal every 3.125 ns. By repeating the process for all the channels, we obtain the final event trace for a module.

We can see an example of the input, pre-amplified, after fast-shaper, and output signals as a function of time for one simulated muon in Fig. 4.2. The single photo-electron pulses can be recognized in the input signal.

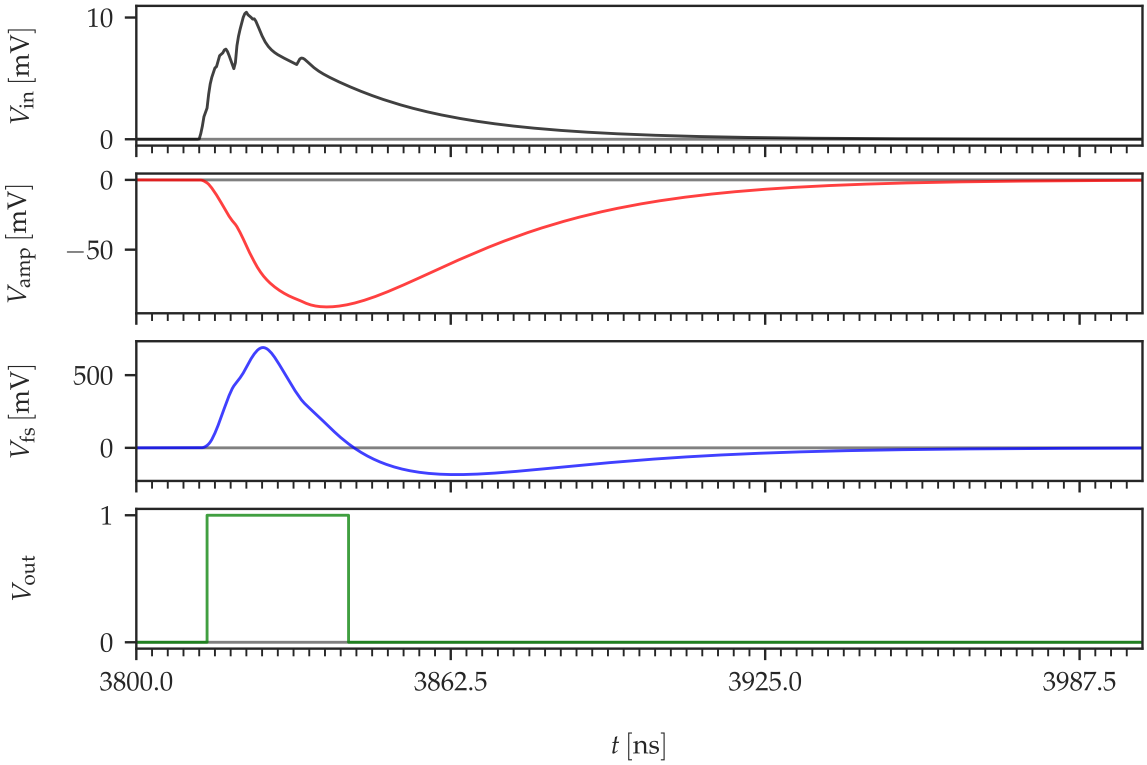


Figure 4.2: Model of the pulse generated by one muon impinging on a scintillator strip. The input signal (top) passes a pre-amplifier (second to top), a fast shaper (second to bottom), and a discriminator and FPGA (bottom). The FPGA digitizes the output signal in 3.125 ns-wide time bins.

After simulating the response of every channel of the UMD module to the given input muons, the binary signal of each channel is matched to the single-muon pattern (as mentioned before, "1111xxxxxxxx"). An example of a simulated module-level event is shown in Fig. 4.3. There we can see the impinging muons, the binary traces, and the matched patterns for all channels participating of the event. In this example, the effect of pile-up can be seen in channels 17 and 23, where the signal of two and three impinging muons respectively is matched only to one pattern.

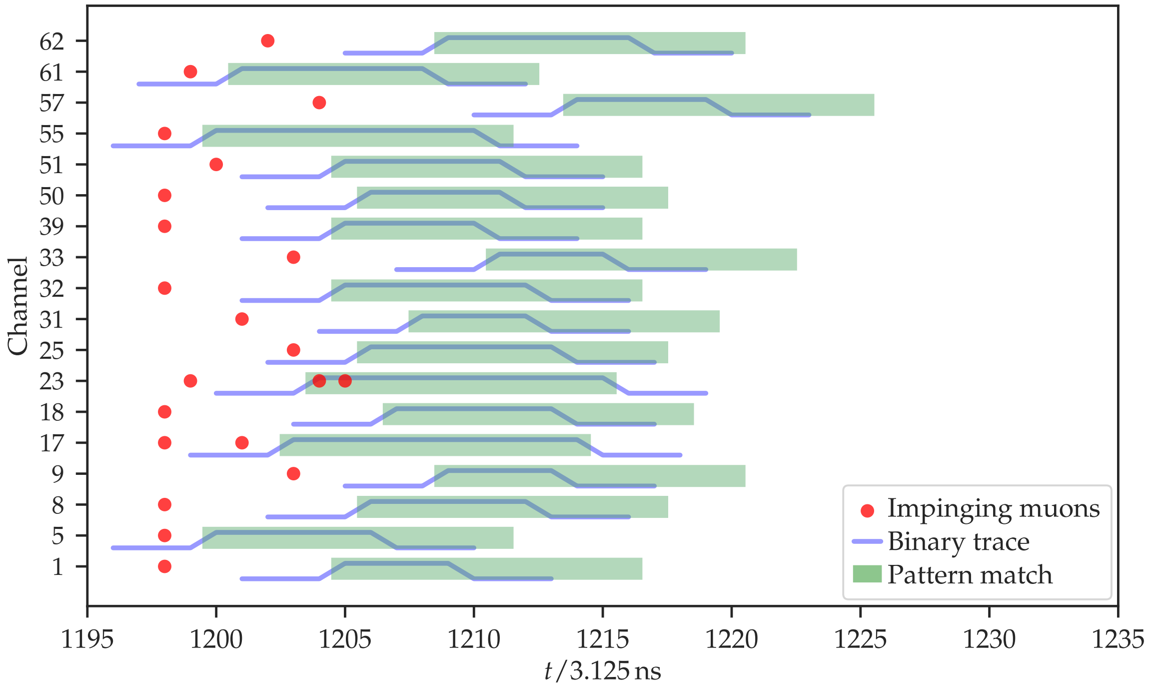


Figure 4.3: Simulated response of a module to the muons of an air shower simulation. Muons (circles) impinge at a given time (x-axis) on the different scintillator strips of the detector. They generate a binary signal (lines) in the corresponding channels (y-axis), which are matched to a single-muon pattern "1111xxxxxxxx" (rectangles).

4.2.2 Detector effects in the counting strategies

The $\widehat{\mu}$ and $\widehat{N}_\mu$ estimates, computed as discussed in Sec. 4.1 are prone to two sources of bias that are neither inherent to the counting strategies nor connected to the pile-up effect.

The first one is a detector and pattern-matching inefficiency that results in under-counting, and equally impacts all counting strategies. It occurs when one muon fails to generate an output signal that is strong or long enough to match the single-muon pattern.

The second one is caused by an undershoot of the signal after the fast-shaper, as visible in the third panel of Fig. 4.2. This produces an under-counting effect which is a result of the design of the electronics. It occurs when a later muon signal mounts on an undershoot caused by earlier muons in the same channel. The later muon may not be matched to the single-muon pattern if either the amplitude of its signal is too small or the absolute value of the amplitude of the undershoot is very large (the amplitude is very negative). The infinite window strategy is the only one that is insensitive to this effect because it suffices to match the early muons in the channel to tag it as occupied over the whole trace. All other strategies are equally affected by the under-counting due to undershoot.

We quantify the number of pattern matches lost as a result of the undershoot and of "small signals" to better understand their effects. In order to do this, we determine for each channel whether the muon signal would match a pattern if it were the only muon signal in it. If not, the muon is in principle lost due to its small signal. This would result in a loss of $\sim$ 3.6 % pattern matches, independent of the energy or zenith angle of the cosmic ray that caused it. However, we consider the muon recovered if the pattern match that the muon would have produced[3] overlaps with that of another muon of the channel. The latter occurs more frequently in more vertical and energetic air showers, which are associated with greater muon rates. The top left panel of Fig. 4.4 shows the net loss of pattern matches as a result of small signals. To obtain this, we employ proton initiated air showers with the typical signals present in an hexagonal array like the SD-750 (see Sec. 4.2.4). For higher-energy and more vertical air showers, the net loss is, as predicted, smaller. The net effect is less than 4 % lost pattern matches.

Moreover, we continue analyzing the signal of a single muon in a channel. If the muon would generate a pattern match were the only muon signal in the channel, we analyze whether that pattern match overlaps with the pattern matches of the total signal of the channel. If this is not the case, the muon is lost due to undershoot. The results are shown in the top right panel of Fig. 4.4. The muons lost because of the undershoot can amount from 0.1 % for low-energetic, inclined air showers, to 3.3 % for high-energetic, vertical air showers.

The lower panel of Fig. 4.4 displays the total detector effects, which are simply the sum of the two contributions (i.e., what is displayed in the top and middle panels of the figure). We can observe that the undershoot has a dominant effect, and more pattern matches are lost in high-energy, more vertical air showers. The total number of pattern matches lost as a result of detector effects range from 2.6 % to 5.0 %, reaching up to 5.3 % for iron-initiated air showers. We repeat that in the case of the infinite window strategy, the total detector effects are only the ones of the top panel, as undershoot does not affect this strategy.

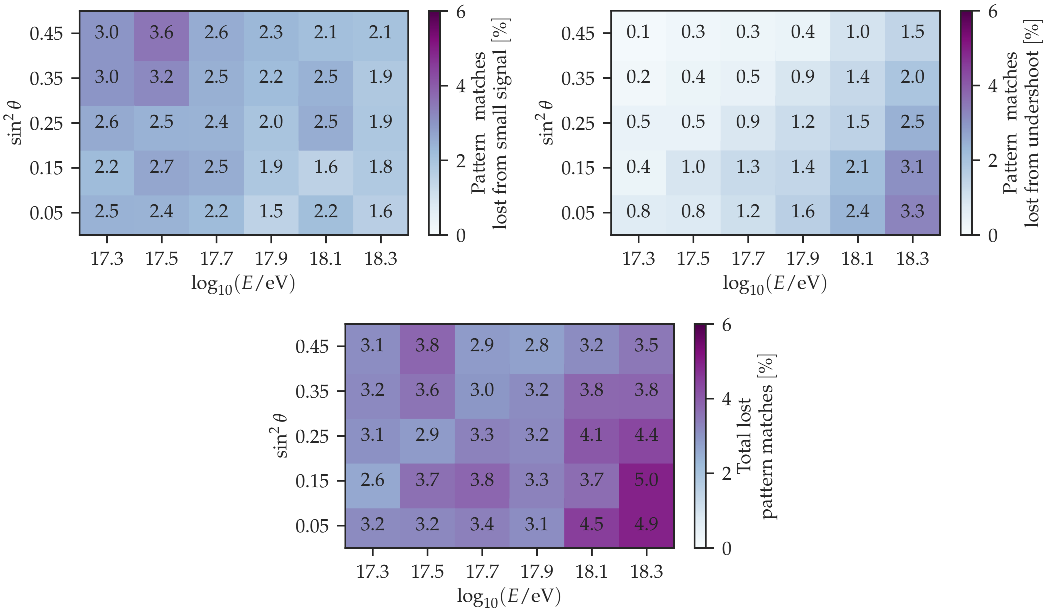


Figure 4.4: Mean percentage of lost pattern matches due to small signals (upper panel), to undershoot (middle panel), and the sum (bottom panel), as a function of logarithmic energy and the sine square of the zenith angle. We employ simulations of proton-initiated air showers using the distance distribution like the ones of the UMD array.

---

[3] We use the time of the maximum of the fast-shaper signal as the time of the beginning of the pattern that the muon would have produced.

4.2.3 Air shower simulations

In order to proceed with our goal of simulating a realistic scenario, we generated a library of ~ 7600 air showers of proton and iron primaries using the high- and low-energy hadronic interaction models EPOS-LHC [26] and UrQMD [32, 33], respectively. The air shower simulations were produced using CORSIKA v7.7402 [23]. Their arrival directions follow an isotropic distribution with zenith angles in $0° \leq \theta \leq 48°$, and they have a uniformly distributed logarithm of primary energy between $17.2 \leq \log_{10}(E/\mathrm{eV}) \leq 18.4$.

We separate the simulations by primary, into $\log_{10}(E/\mathrm{eV})$-bins with widths of 0.2 centered at 17.3, 17.5,..., 18.3, and in $\sin^2\theta$-bins of width 0.10 centered at 0.05, 0.15,..., 0.45. We subsequently compute an average profile of the number of muons as a function of the logarithmic distance to the shower plane and of the time $\mathrm{d}\mu/\mathrm{d}t \times \Delta t$ for each primary and $(\log_{10}(E/\mathrm{eV}), \sin^2\theta)$-bin. To do this, we first compute the profile of each air shower by extracting, for each muon that has sufficient energy to reach the ground, the distance to the shower axis on the shower plane and the time at which it arrives at the shower plane. Next, we calculate a weighted mean of the $\mathrm{d}\mu/\mathrm{d}t \times \Delta t$ profiles, for all showers that fall into a same $(\log_{10}(E/\mathrm{eV}), \sin^2\theta)$-bin, being the weight of the i-th shower its energy E_i times the cosmic ray flux evaluated at that energy $J(E_i)$. In this way we get the average profile that would be observed for air showers with energies following the distribution given by the cosmic ray flux. The model of the flux follows Ref. [92]. As an example, we show in Fig. 4.5 the mean profile $\mathrm{d}\mu/\mathrm{d}t \times \Delta t$ obtained from proton air showers with $17.8 \leq \log_{10}(E/\mathrm{eV}) \leq 18.0$ and $27° \lesssim \theta \lesssim 33°$ $(0.20 \leq \sin^2\theta \leq 0.30)$.

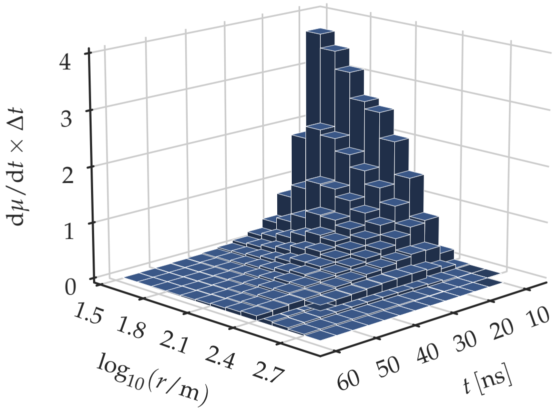


Figure 4.5: Mean number of muons per time-bin as a function of the shower plane arrival time and of the logarithmic distance to the shower axis on the shower plane. The average takes proton air showers with $17.8 \leq \log_{10}(E/\mathrm{eV}) \leq 18.0$ and $27° \lesssim \theta \lesssim 33°$ $(0.20 \leq \sin^2\theta \leq 0.30)$.

4.2.4 End-to-end simulation chain

Using the air shower library described in Sec. 4.2.3 as the input for the detector simulations described in Sec. 4.2.1, we generate two detector simulations sets.

The first detector-simulation set is obtained in the following way: For each primary and $(\log_{10}(E/\mathrm{eV}), \sin^2\theta)$-bin, the profile $\mathrm{d}\mu/\mathrm{d}t \times \Delta t$ is evaluated 10,000 times at a random logarithmic distance $\log_{10}(r^*/\mathrm{m})$ which we sample from a continuous uniform distribution $\mathcal{U}\{1.0, 3.5\}$. This determines the mean input number of muons as a function of time $\mathrm{d}\mu/\mathrm{d}t \times \Delta t\,|_{r=r^*}$ (which integrates to $\mu = \mu\,(\log_{10}(r/\mathrm{m}))\,|_{r=r^*}$). Next, we sample a Poisson distribution of mean μ to generate the number of impinging muons N_μ. Afterwards, the times of the N_μ muons are sampled from $\mathrm{d}\mu/\mathrm{d}t \times \Delta t\,|_{r=r^*}$. This information is the input to the detector simulation, which, as described in Sec. 4.2.1, allocates the muons randomly in the scintillator strips, simulates the electronics response, and outputs the 64 binary signals of the module channels. The binary signals are then matched to the pattern “1111xxxxxxxx”. With this, we estimate $\widehat{\mu}$ and $\widehat{N}_\mu$ for each of the four strategies described in Sec. 4.1 (see Table 4.1).

The second detector simulations set is identical to the first one, except for the distribution of the logarithmic distances $\log_{10}(r^*/\mathrm{m})$ to the shower axis. In this set we want this distribution to be realistic. For this purpose, we simulate an hexagonal array of 750 m spacing, like that of the SD-750 array where the UMD lies, and throw random event-core positions. From this we can calculate the distribution of distances to the shower axis. However, triggering has to be considered, because there would be increasingly more UMD modules at large distances otherwise. Using the previously parameterized muon profiles integrated in time for the different primaries, energies, and zenith angles, $\mu(\log_{10}(r/\mathrm{m}))$, we estimate the number of muons at the detector. By requiring the impinging number of muons to be ≥ 3 we simulate the triggering conditions. In this way we obtain the realistic distribution of the distances to the shower axis, as seen from the shower plane. We find that the distributions of the same energy at various zenith angles are indistinguishable from one another. However, there is a dependence on the energy and on the primary. This is because more energetic air showers produce more muons, and some of them still reach large distances from the shower axis with sufficient energy to trigger the detector. Additionally, stations that are farther from the shower axis can be still be triggered by air showers produced by heavier primaries because they produce more muons than lighter primaries of the same energy. In Fig. 4.6 we show the distribution of distances for the considered $\log_{10}(E/\mathrm{eV})$-bins, for a 50 % − 50 % mixture of proton and iron, for the mentioned hexagonal array. From Fig. 4.6 we can see that the distribution first rises linearly, peaks, and then decreases softly with distance, as expected.

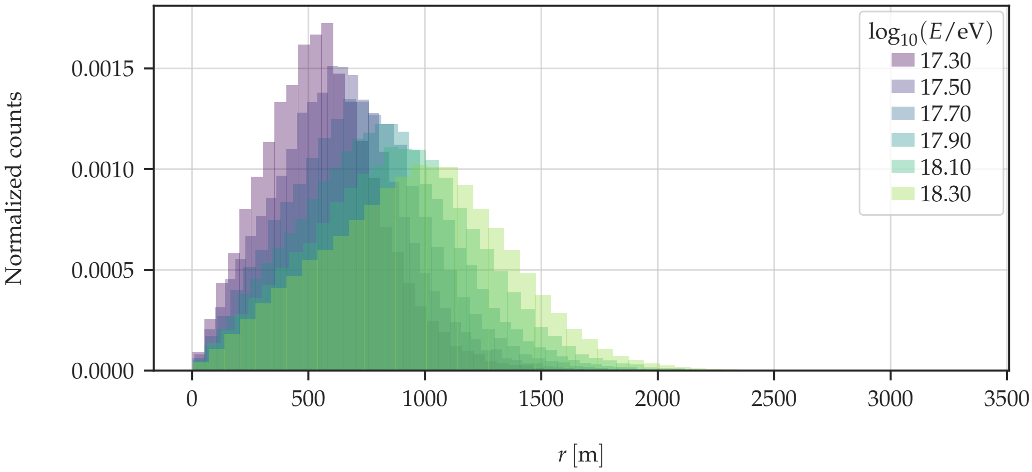


Figure 4.6: Normalized distributions of station distances to the shower axis in a simulated hexagonal array with 750 m spacing (like the one where the UMD lies), for the considered $\log_{10}(E/\mathrm{eV})$-bins, for a 50 % − 50 % mixture of proton and iron, and for zenith angles below 48°.

4.2.5 Validation of end-to-end simulations via the saturation fraction

For the purpose of validating our simulations, we estimate the expected fraction of saturated events and compare it to the fraction that we find in simulations.

Using our statistical model (c.f. Sec. 4.1), we are able to estimate the expected saturation fraction as a function of μ and N_μ for the infinite window method. For any other strategy, we would have to introduce a model or hypothesis on how μ evolves with time.

The expected saturation fraction as a function of μ follows a binomial distribution where the condition of saturation is met, this is, the number of successes (k) is equal to the number of trials (n_s)

$$B(n_s|\,\mu) = e^{-\mu}(e^{\mu/n_s} - 1)^{n_s}. \tag{4.13}$$

Moreover, the expected saturation fraction as a function of N_μ follows the occupancy distribution as described in Ref. [88] (see Eq. (4.4)), also evaluated at the condition of saturation ($k = n_s$)

$$Occ\,(n_s|\,N_\mu, n_s) = \frac{S(N_\mu, n_s)}{{n_s}^{N_\mu}}. \tag{4.14}$$

We repeat that $S(N_\mu, k)$ are the Stirling numbers of the second kind.

We reiterate that the UMD module is considered to be saturated in an event if, for any window j, all channels are occupied (infinite window, N-bin, and N-bin centered methods) or all channels are either occupied or inhibited (1-bin strategy). In this case there is no estimate of $\widehat{\mu}$ and $\widehat{N}_\mu$ as they tend to infinity. Notice that a module can be saturated for the infinite window strategy (if every channel has at least one pattern match), but it is not necessarily saturated for the other strategies (if the timing of the muons is sufficiently spread in time).

We employ the first set of simulations, which consists of detector simulations with a uniform distribution of the logarithmic distance from the shower axis (see Sec. 4.2.4). Then, we compute the fraction of saturated events as a function of μ and as a function of N_μ for all air showers of a same $(\log_{10}(E/\mathrm{eV}), \sin^2\theta)$-bin. Fig. 4.7 shows the predicted saturation fraction and the one found in simulations for each of the four counting strategies. Proton air showers with $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 18.2$ and $18^\circ \lesssim \theta \lesssim 27^\circ$ ($0.10 \leq \sin^2\theta \leq 0.20$) were used in this example. The general behavior is the same for other energies and zenith angles; the saturation model is exactly the same, but the range of values that μ or N_μ reach changes. We see that the saturation fraction of the infinite window strategy follows the models closely, while the 1-bin and N-bin centered methods follow them approximately, showing smaller saturation fractions than the infinite window strategy. It is expected that the N-bin centered strategy behaves in this way since a big part of the signal can be contained in a single window, making the strategy function almost as the infinite window for average air-shower signals. We can also see in Fig. 4.7 that the N-bin strategy reaches saturation at larger μ and N_μ values. This, however, is only an artifact caused by the strategy due to not taking inhibited channels into consideration.

4.3 Analysis of the performance of the counting strategies

In this section we present the evaluation of the performance of the presented counting strategies.

We present the results only for N_μ for simplicity, but the general trends are also valid for μ. The main difference is that the Poisson fluctuations in the number of injected muons affect the resolution in the estimation of μ. These play a dominant role when N_μ is much smaller than the number of segments n_s. In contrast, when N_μ is large, the detector segmentation predominantly determines the resolution in the estimators of N_μ and μ, which are thus similar [13].

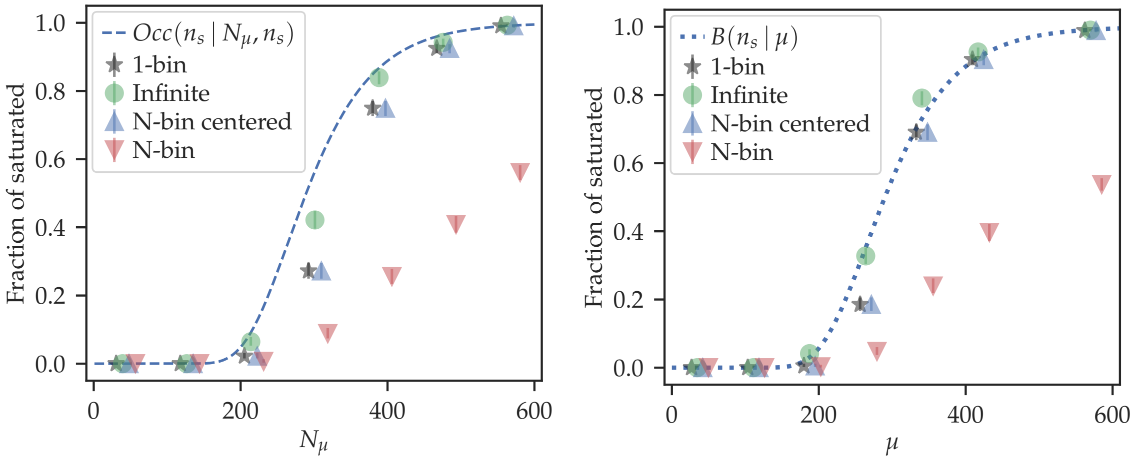


Figure 4.7: Saturation fraction for each of the four strategies under consideration, as a function of N_μ (left panel) and as a function of μ (right panel), as modeled (dotted line) and as calculated from simulations (markers). The detector simulations use proton air showers with a uniform distribution in the logarithmic distance to the shower axis, with $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 18.2$ and $18^\circ \lesssim \theta \lesssim 27^\circ$ ($0.10 \leq \sin^2\theta \leq 0.20$).

4.3.1 Bias as a function of the number of impinging muons

As a first step, we define the relative difference between the estimated and Monte-Carlo true number of muons as

$$\varepsilon = \frac{\widehat{N}_\mu - N_\mu}{N_\mu}. \tag{4.15}$$

In Fig. 4.8 we show ε as a function of the number of true impinging muons N_μ. This is an example computed from proton air showers with $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 18.2$ and $33^\circ \lesssim \theta \lesssim 39^\circ$ ($0.3 \leq \sin^2\theta \leq 0.4$). We signal with a small black line the average for each distribution. The saturated modules are excluded. When saturation is significant, at large N_μ values, the distributions appear as multi-modal. We observe a larger variance in the case of the N-bin strategy. This is because of the random placement of the signal in the partition of the trace: a set of otherwise identical binary signals with different start times will be reconstructed as different numbers of muons just because the signals are partitioned in different ways. In Fig. 4.9 we compare the relative bias $\langle\varepsilon\rangle$ (left panel) and the relative resolution $\sigma(\varepsilon)$ (right panel) of each strategy, which summarizes the information presented in Fig. 4.8. The error bars in $\langle\varepsilon\rangle$ are the standard deviation of the mean. The error bars in $\sigma(\varepsilon)$ were calculated using bootstrap. Compared to the other strategies, the N-bin strategy immediately stands out for having a substantially larger relative bias (to negative values) and larger standard deviation. Additionally, the N-bin centered, infinite window, and 1-bin strategies all perform roughly similarly in terms of mean relative bias: for an input of $N_\mu \lesssim 200$, they all have a small negative bias (contained within ±10 %) dominated by detector effects, while the bias increases above that as the detector begins to saturate. In addition, because of not considering inhibited channels, the mean relative bias of the N-bin centered strategy tends to negative values, similar to the N-bin strategy. The 1-bin strategy behaves like the infinite window strategy with a smaller detector, but suffering of slightly more negative biases due to the undershoot effects. The bias in the infinite window strategy is the least, yet it is still non-zero. We notice that the precision of the 1-bin strategy is similar to or greater than that of the other strategies when we compare the standard deviations.

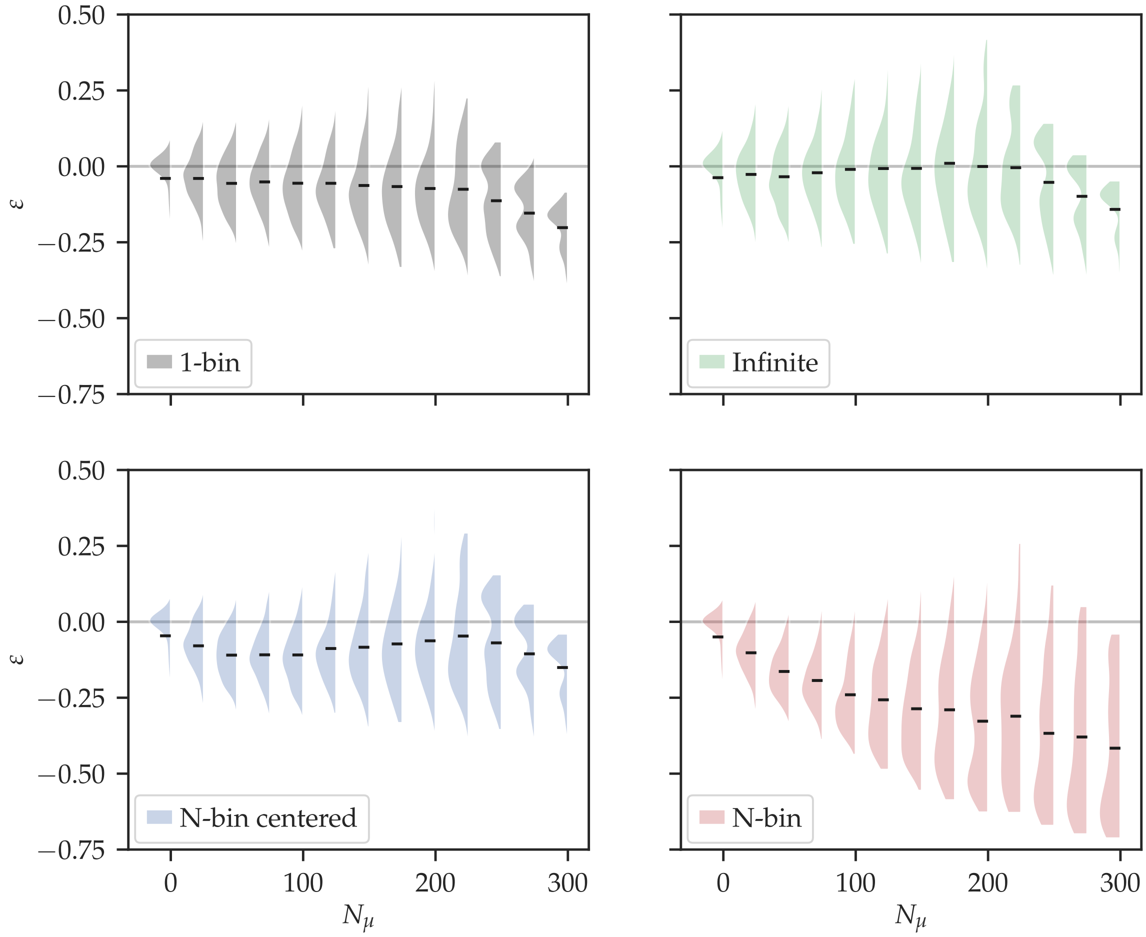


Figure 4.8: Relative difference of the estimated vs Monte-Carlo true number of muons (c.f. Eq. (4.15)) as a function of the Monte-Carlo true number of muons. The different panels show the different strategies: The 1-bin strategy is at the top left panel, the infinite window strategy is at the top right panel, the N-bin centered strategy is at the bottom left panel, and the N-bin strategy is at the bottom right panel. We used the first simulations set (uniform in $\log_{10}(r/\mathrm{m})$) taking proton air showers with $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 18.2$ and $33^\circ \lesssim \theta \lesssim 39^\circ$ ($0.30 \leq \sin^2\theta \leq 0.40$).

It is relevant to add that using iron simulations or other $(\log_{10}(E/\mathrm{eV}), \sin^2\theta)$-bins results in a comparable bias behavior as a function of the Monte-Carlo true number of muons for all strategies.

4.3.2 Average performance in an underground muon detector-like array

In this section we study how the average relative bias ϵ behaves in a UMD-like array, for the four considered strategies, and for different energies and zenith angles. This complements the analysis described in Sec. 4.3.1. The idea is that, in data, large values of impinging number of muons N_μ are not as frequent as low numbers, and they are only given at distances close to the core. For this analysis, we use the second detector-simulations set, that provide a realistic distribution of the logarithmic distances to the shower axis $\log_{10}(r^*/\mathrm{m})$, i.e., like that of the SD-750 array.

In Fig. 4.10 we show an example of the distribution of ε (c.f. Eq. (4.15)) computed from 1000 proton air showers with $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 18.2$ and $27^\circ \lesssim \theta \lesssim 33^\circ$ ($0.20 \leq \sin^2\theta \leq 0.30$), for the four strategies considered, for an array like the SD-750. For other energies and

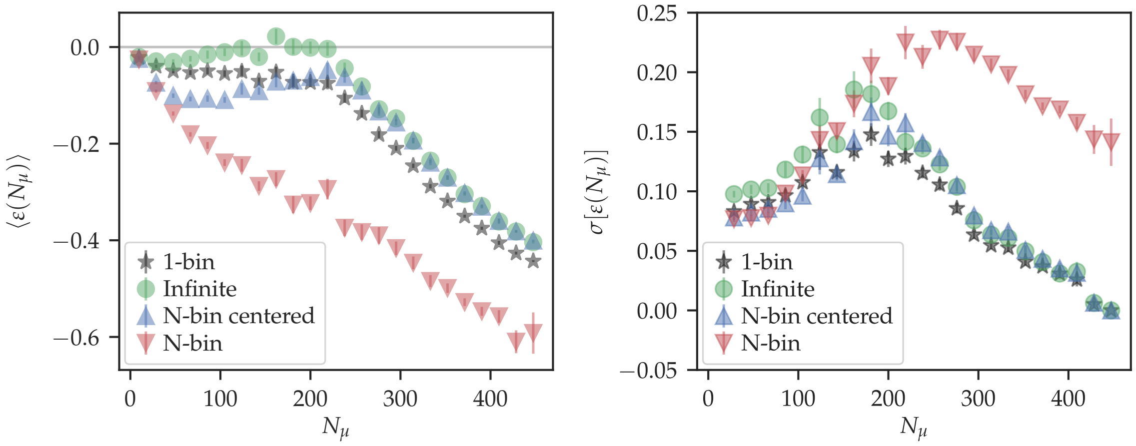


Figure 4.9: Relative bias (left) and relative standard deviation (right) in the reconstructed number of muons, as a function of the Monte-Carlo true number of muons, for the four strategies considered (in different markers). The detector simulations that we used are the same as those used for Fig. 4.8.

zenith angles the results are similar. The contributions to each bin of the histograms of the simulated module-level events that would not pass the distance cut $r > r_{\min}$, being $r_{\min}$ the farthest distance to the shower core at which there is saturation for the considered strategy, are differentiated using gray hatching patterns. In a light shaded area we mark the standard deviation and in a dark shaded area the standard deviation of the mean, both centered at the mean and calculated considering the distance cut. We can observe that, in this case, the N-bin strategy presents the largest biases. The mean ε is generally slightly more negative for all strategies when the distance cuts are ignored.

In Fig. 4.11 we summarize and compare the mean ε as a function of $\log_{10}(E/\mathrm{eV})$ and $\sin^2\theta$ for the four strategies, using the second detector-simulations set, with proton air showers and including the distance cut. In Fig. 4.12 we show the same but for iron-initiated air showers. We can observe that the bias is contained within ±3.9 % for the infinite window strategy, within ±4.2 % for the 1-bin strategy, within 4.9 % for the N-bin centered strategy, and within ±5.4 % for the N-bin strategy. Since most measurements take place at larger distances from the shower axis, as was previously indicated, there are often fewer input muons, which have small biases. As a result, the average biases are generally small. The bias is dominated by the detector effects mentioned in Sec. 4.2.2. We reiterate that the undershoot does not affect the infinite window strategy, which is only susceptible to the effects of the detector and pattern matching efficiency. On the contrary, the other three strategies are subject to both effects. From Fig. 4.12 we see that, for iron air showers, the biases are generally slightly larger than for proton air showers for all strategies. In both cases, the distance cut has only a very small effect: it can influence the mean bias in a $(\log_{10}(E/\mathrm{eV}), \sin^2\theta)$-bin at most by 0.2 %.

4.3.3 Performance under large or double-bump-like input signals

It was thought before that the underestimation of the number of muons using the N-bin strategy arose from the signal being “split” into two windows. The idea was that two parts of the signal were wrongly mathematically treated as independent, while they should be treated as simultaneous by containing (most of) the signal in one window. However, if this were the problem, the solution should be to center the window around the signal, as it is done in the N-bin centered strategy.

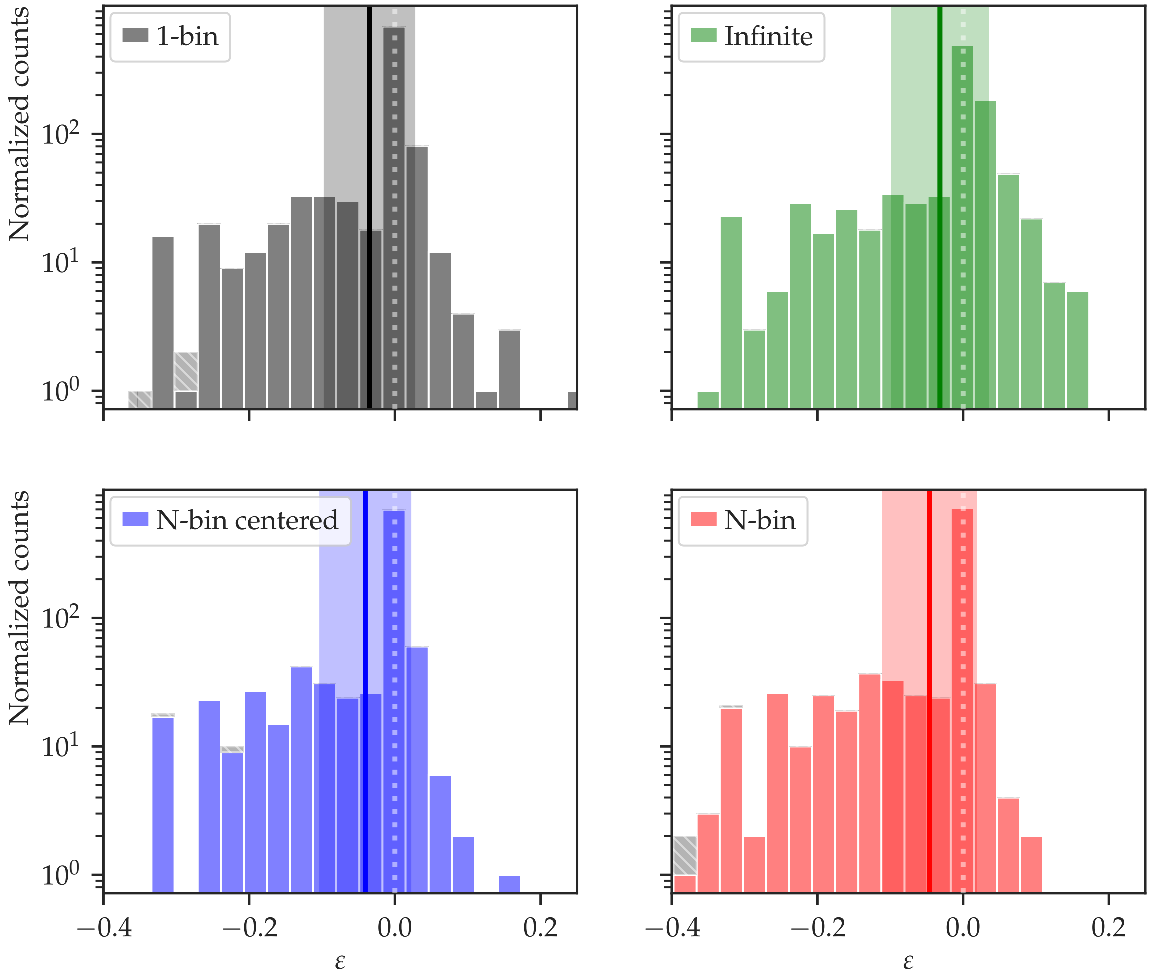


Figure 4.10: Histograms of ε (c.f. Eq. (4.15)) for the four strategies under consideration. The standard deviation of the mean is shown as a vertical dark shaded area centered at the mean, while the standard deviation is shown as a light shaded area. Null bias is indicated with a white dashed vertical line. The events that fail to pass the distance cut $r > r_{\min}$ are represented in hatched patterns (see text for details). We use the second detector-simulations set (i.e. distribution of the logarithmic distances to the shower axis as in the SD-750 array, c.f. Fig. 4.6), taking as input proton air showers with $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 18.2$ and $27^\circ \lesssim \theta \lesssim 33^\circ$ ($0.20 \leq \sin^2\theta \leq 0.30$).

In this section we test this hypothesis and throw light on the origin of the biases in the estimators of the N-bin strategy. For this purpose, we design two different input signals. These can be seen in Fig. 4.13. The first input signal (top panel) is a log-normal

$$\frac{\mathrm{d}\mu(t)}{\mathrm{d}t} = \Theta((t-t_0)/\mathrm{ns})\,\frac{A}{((t-t_0)/\mathrm{ns})\,\sigma\sqrt{2\pi}} \times \exp\left(-\frac{(\ln((t-t_0)/\mathrm{ns})-m)^2}{2\sigma^2}\right), \tag{4.16}$$

with the amplitude A equal to its integral $\mu \approx 337$, and being the scale, shape, and location parameters $m \approx 3$, $\sigma \approx 1$, and $t_0 \approx 0.6\,\mathrm{ns}$ respectively. To put it in perspective, the average proton air shower of $\log_{10}(E/\mathrm{eV}) = 18.1$ and $\theta \approx 30^\circ$ ($\sin^2\theta = 0.25$) has a scale parameter that is 4 times larger and an amplitude that is 25 times smaller than those of this input signal. Notice that it would be equivalent to consider an input signal of standard width while taking narrower windows (i.e. a shorter single-muon pattern). This implies that another detector

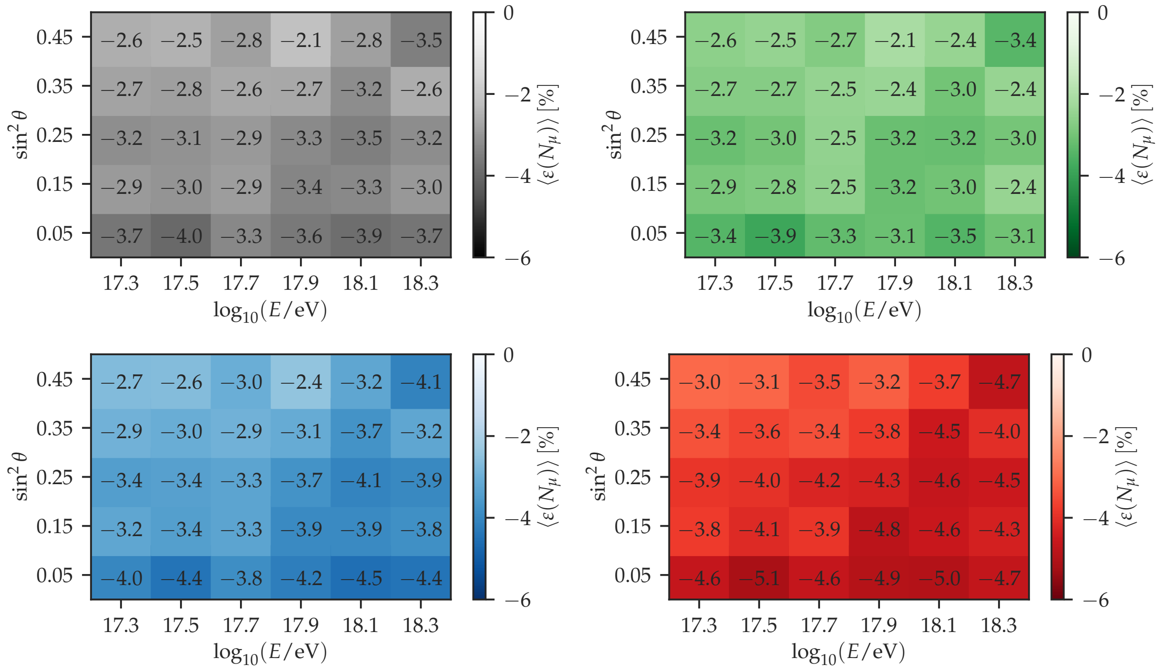


Figure 4.11: Mean ε as a function of the logarithmic energy (x-axis) and the sine square of the zenith angle (y-axis), for the 1-bin (top left), the infinite window (top right), the N-bin centered (bottom left), and the N-bin (bottom right) strategies. We use proton showers in an SD-750-like array, considering distance cuts (see details in text).

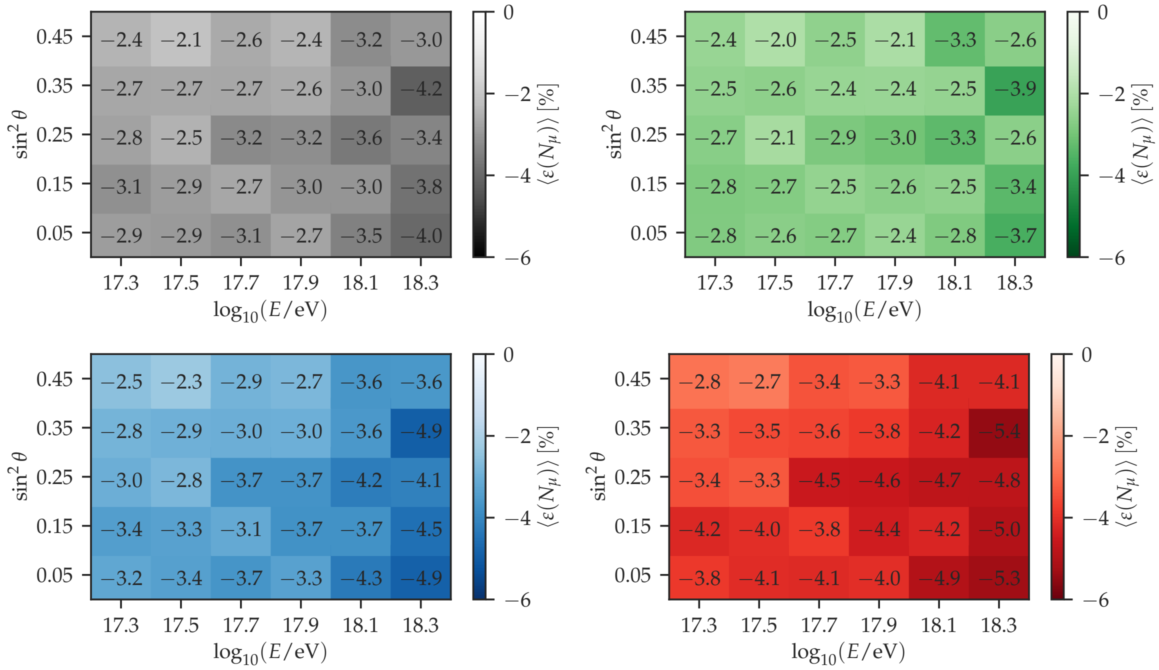


Figure 4.12: Same as Fig. 4.11, but using iron-initiated air showers.

with different electronics that has a shorter muon pattern would have a response to average air-shower signals similar to the response of the UMD to this wide signal.

The second input signal, shown in the bottom panel of Fig. 4.13, is a double-bump like signal. It is composed of two log-normals separated by a time interval of our choosing, Δt, which is equal to the difference between the two location parameters, $t_{0,2} - t_{0,1}$. Both log-normals have a proton air shower's shape and scale parameters of $\log_{10}(E/\mathrm{eV}) = 18.1$ and $\theta \approx 30°$ ($\sin^2\theta = 0.25$), but the second log-normal has a smaller amplitude than the first one by 25 %. It integrates in total to $\mu \approx 300$.

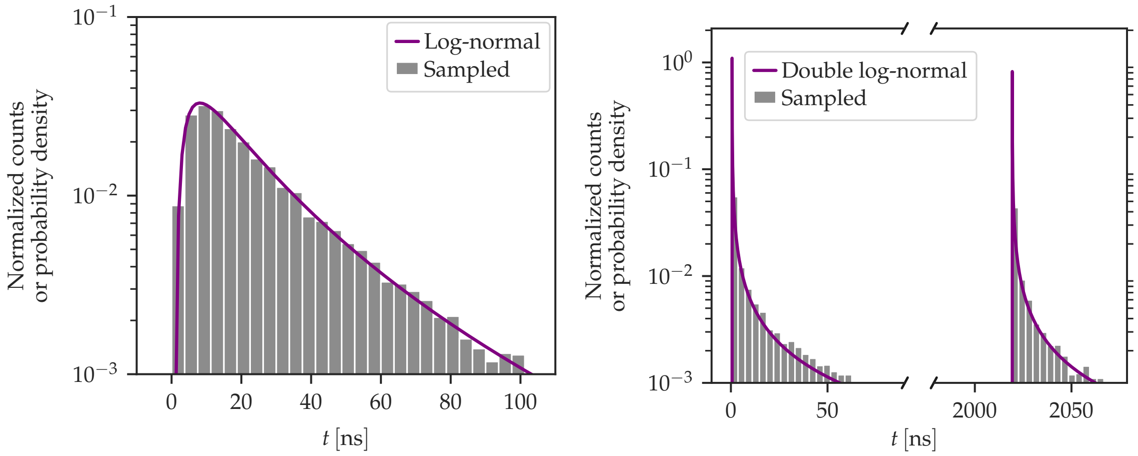


Figure 4.13: Signals used as input for testing the counting strategies. The left panel displays a wide log-normal signal with a scale, shape, and location parameters of $m \approx 3$, $\sigma \approx 1$, and $t_0 \approx 0.6$ ns respectively. The right panel displays two log-normals, one with location parameter $t_{0,1} \approx 0.6$ ns and the other with $t_{0,2} \approx 2019.35$ ns, with the second one having an amplitude 25 % smaller than the first one, and both with the shape and scale parameters of a proton air shower of $\log_{10}(E/\mathrm{eV}) = 18.1$ and $\theta \approx 30°$.

We proceed to generate 1048 events randomly sampling the wide signal, and 1300 events randomly sampling the double-bump signal. For the latter, we set Δt 100 times to each of the following values: 2000.000 ns, 2003.125 ns,..., 2037.500 ns. In greater detail, we set the number of impinging muons by sampling N_μ from a Poisson distribution of mean μ. Then we sample the timings of the N_μ muons from the discussed input signals $\mathrm{d}\mu(t)/\mathrm{d}t$.

Figure 4.14 shows the distribution of ε for each strategy, and for the two input signals. In the left panel (wide log-normal), we see that the infinite window strategy performs the best, in spite of suffering from significant saturation, and is followed closely by the 1-bin strategy. In the right panel (double log-normal), the 1-bin strategy is the best performing one. As mentioned before, the infinite window strategy is not affected by undershoot, only by the detector inefficiencies. It is also more prone to saturation under such kind of signals. These two effects explain its bias. Moreover, the 1-bin strategy is subject to detector inefficiencies and to the effect of undershoot. In both cases, the N-bin centered and the N-bin strategies have larger biases, as expected. This demonstrates that the N-bin centered strategy underperforms in the UMD when the signal is wide or when there are a substantial number of late particles. It also implies that the N-bin centered strategy would have large biases in average signals in other detectors where the muon pattern is shorter.

Since we observe that the biases of the N-bin strategy are not sufficiently solved by centering the window around the signal maximum, this analysis demonstrates that these biases cannot be explained from the signal being split in two windows. In Sec. 4.3 and Sec. 4.3.2 we observed that the bias of the N-bin centered strategy is relatively small for average signals. The reason is that a big percentage of the signal can be contained within one 12 time-bin window in many cases. This causes the N-bin centered strategy to effectively perform as the infinite window strategy. This strategy is subject to biases under wider signals or shorter single-muon patterns. This analysis proves that the actual reason of the biases in

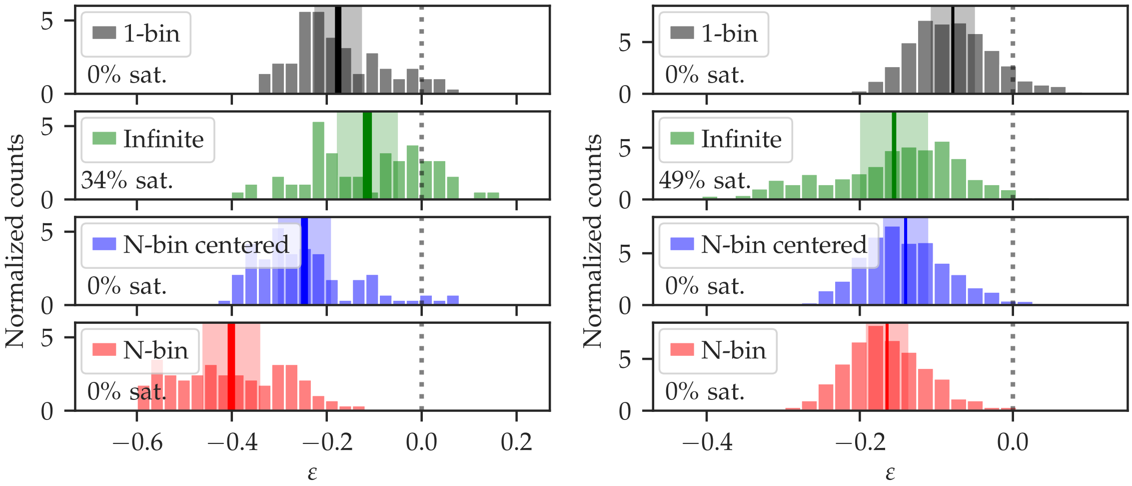


Figure 4.14: Distributions of ε (c.f. Eq. (4.15)) for each strategy, for a wide input signal (left) and a double-bump input signal (right). A vertical dark shaded area centered at the mean marks the standard deviation of the mean, while the standard deviation is included in the lighter shaded area. The percentage of saturated events for each strategy and input signal is indicated in each panel.

the N-bin and N-bin centered strategies is that they do not model the detector properly by not accounting for inhibited channels.

4.3.4 Reconstruction of the muon signal as a function of time

As mentioned before, a great advantage of the 1-bin strategy developed in this work is that it uses the complete timing information of the measured signals. This allows to reconstruct the number of muons as a function of time as seen by the detector to a single time-bin resolution (i.e. 3.125 ns).

Figure 4.15 shows an example of such reconstructed signal. There are three curves that represent the input muons as a function of time, the average input muons as a function of time as seen by the detector (i.e. adding the time delays of the detector to the arrival times of the muons), and the number of muons as a function of time estimated using the 1-bin strategy. In this example, the input number of muons as a function of time are taken from the average of proton air shower simulations with $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 18.2$ and $\theta \lesssim 18^\circ$ ($\sin^2\theta \leq 0.10$), at 450 m from the core. The timing of the reconstructed muon signal is consistent with that of the muon signal as seen by the detector. They are both affected by the time delays introduced by the detector, i.e. the propagation in the optical fibers and the de-excitation times of the scintillator and optical fiber core.

4.4 Summary and outlook

In this chapter we studied how to estimate the number of muons impinging on the UMD modules. The analyzed counting strategies directly generalize to the more general problem of estimating the number of impinging particles in a segmented detector. The main effect that has to be taken into account in any counting strategy is pile-up, which happens when two muons impinge a scintillator bar almost simultaneously such that they cannot be resolved in time.

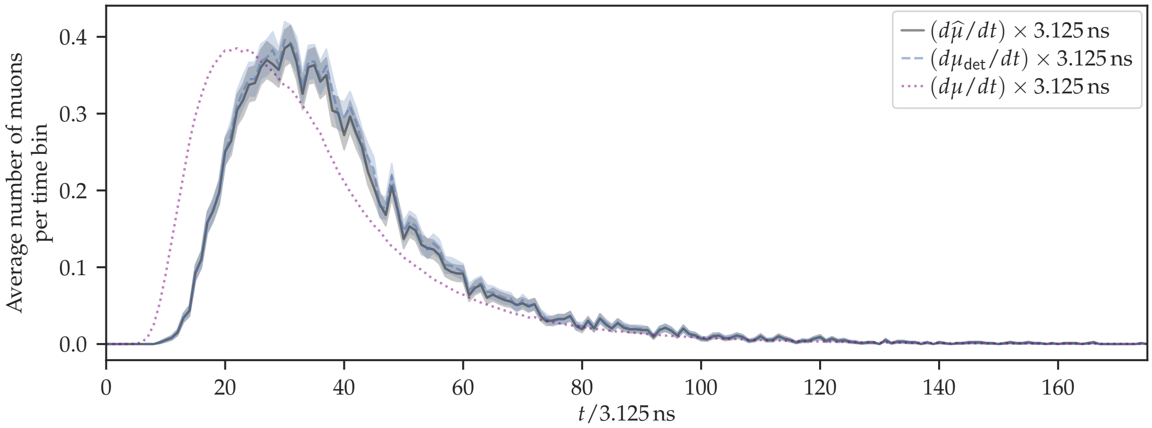


Figure 4.15: Average number of muons reconstructed using the 1-bin strategy (solid line), average input number of muons as detected (dashed line), and average input number of muons (dotted line), all as a function of time. The average number of muons as observed by the detector is calculated by accounting for the detector's time delays in the input muons. The average muon profile of proton air shower simulations with $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 18.2$ and $\theta \lesssim 18°$ ($\sin^2\theta \leq 0.10$) at 450 m from the core is used as the input number of muons as a function of time. Shaded areas represent $1\,\sigma$ uncertainties.

We described and compared four counting strategies: the infinite window, the N-bin, the N-bin centered, and the 1-bin strategies, the latter being the one developed in this work. We explained how they are all founded on the solution to the "classical occupancy problem", also referred to as the "balls in boxes problem". In our case, the balls represent the muons and the boxes represent the segments of the UMD module. The primary distinction between the strategies is the length of the window during which the single-muon pattern matches are regarded as simultaneous. For the infinite window strategy, this is the complete trace; for the N-bin and N-bin centered strategies, it is a window of the length of the single-muon pattern; and for the 1-bin strategy, it is a single time bin. The positioning of the windows is optimized for the N-bin centered strategy, in comparison to the N-bin strategy, making it so that the center of the output signal coincides with the center of a particular window. The 1-bin strategy's main distinction from the others is its consideration of inhibited segments.

We demonstrated that all strategies, aside from the N-bin one, perform well (within ±10 %) for typical air shower muon signals in a detector like the Auger UMD. The infinite window strategy is the one with the smallest bias, but the 1-bin strategy has an only slightly larger bias. The 1-bin strategy had the smallest standard deviation in the predicted number of muons, specially at high input muon levels, while the N-bin strategy had the largest. The bias of each strategy for each $(\log_{10}(E/\mathrm{eV}), \sin^2\theta)$-bin considering the typical distribution of input number of muons is observed to be contained within ±5.4 %, being the largest (more negative) for the N-bin strategy. Detector effects are the dominant cause of the bias: the signal undershoot causes up to a $-3.3\,\%$ bias, while the detector and pattern matching inefficiencies account for up to $-3.6\,\%$ of the bias.

To understand the origin of the biases in the N-bin strategy we analyzed the performance of the different counting strategies under a wide and a double-bump-like input signals. Under these conditions, we observed that the infinite window and 1-bin strategy were the best performing strategies. The results proved that the biases in the N-bin strategy, which extend to the N-bin centered strategy, are inherent to the methods, which do not treat inhibited channels as such.

We demonstrated that the 1-bin strategy may be used to reconstruct the time structure of the muon signal as observed by the detector down to a single time-bin resolution. No other strategy provides comparable opportunities. Studies on the temporal structure of the muon signal can now be performed thanks to the 1-bin counting strategy developed in this work. Future work will be devoted to deconvolving the detector effects in the reconstructed average muon signal to estimate the input muons as a function of time. This is a necessary step for one of the most prominent analysis that this strategy will allow: the reconstruction of the depth-of-the-shower maximum of muons, $X_{\mu\,\mathrm{max}}$ [93]. This parameter is highly sensitive to composition. In this way, this work constitutes a first big step towards new composition analyses using $X_{\mu\,\mathrm{max}}$.

Finally, we can generalize the conditions under which the counting strategy developed in this work is applicable: (1) There is a counter with sub-units, (2) in each sub-unit, the signal is processed with a discrimination threshold (counter-mode), and not by using the amplitude or deposited charge (integrator-mode), (3) a sub-set (larger than one) of counting sub-units expect the same particle rate, and (4) not all the sub-units have signal simultaneously (no saturation). One case in which this could possibly be applied is for counting photons produced in liquid scintillators with photo-multiplier tubes (PMTs), a typical design of neutrino studies. In this case, the energy of a neutrino event can be estimated by counting the produced scintillation photons [94]. More specifically, the counting strategy may be applied to each set of PMTs that are equidistant from the interaction vertex, which expect the same scintillation photon rate. In spherical detectors, like the Sudbury Neutrino Observatory + (SNO+) [95], and of the Jiangmen Underground Neutrino Observatory (JUNO) [96], the set of PMTs that expect the same rates lie in rings which represent the intersection of spheres centered in the vertex with the PMT spherical array. In cylindrical detectors, the set of PMTs that expect the same rates follow more complex curves given by the intersection of a sphere centered at the vertex of interaction with the PMT cylindrical array.

Chapter 5

Correction of the biases in the reconstructed number of muons

In Chapter 4 we showed a new counting strategy (the 1-bin counting strategy) that applied to the binary channel of the UMD provides useful estimates of the number or density of muons in air showers. In said analysis, we did not simulate all effects that affect the real reconstructed number of muons. It is of our interest to obtain an unbiased estimator of the reconstructed number of muons $N_{\mu\,\mathrm{rec}}$ that takes into account all effects, in order to use it for data reconstruction. In reality, the dominating bias in $N_{\mu\,\mathrm{rec}}$ is a double-counting effect caused by "corner-clipping muons", which are muons that transverse two neighbor scintillator strips. The double-counting effect is given when the muon, or its secondary electrons, produce enough signal in both scintillator strips such that the muon is counted in both strips[1]. Double counting is more likely when the impinging muon is more inclined (larger zenith angle θ_μ) and more perpendicular to the scintillator strips. The latter is given when the azimuth of the module, ϕ_{mod}, and the azimuth of the muon, ϕ_μ, have a difference $\Delta\phi_\mu = \phi_{\mathrm{mod}} - \phi_\mu$ close to 90° or 270°.

One method for correcting the biases is against detector simulations, comparing module-wise the number of reconstructed muons $N_{\mu\,\mathrm{rec}}$ to the number of injected muons N_{inj} (see for example Ref. [97]). In said approach, the bias $(N_{\mu\,\mathrm{rec}}/N_{\mu\,\mathrm{inj}} - 1)$ can be parameterized, and then the parameterization can be used to correct $N_{\mu\,\mathrm{rec}}$. Such method has the advantage that if $N_{\mu\,\mathrm{rec}}$ has biases from several sources, it corrects for them all, at least to the extent in which simulations reproduce the real behavior. This method fulfills the goal of having a final unbiased estimator.

In contrast, data-driven methods are limited by the impossibility of knowing the true number of injected muons N_{inj}. Trying to estimate the latter from the measured data to correct $N_{\mu\,\mathrm{rec}}$ is in essence equal to knowing the correction, which is evidently a problem.

In this chapter, we describe a new parameterization of the bias in the estimator of the number of reconstructed muons $N_{\mu\,\mathrm{rec}}$ derived from simulations. We propose that the bias can be parameterized as a function of the zenith angle of the shower θ and of the difference in azimuth between the module and the shower $\Delta\phi = \phi_{\mathrm{mod}} - \phi$ (similar to what was done for PMT UMD data in Refs. [90, 97]), and additionally on $N_{\mu\,\mathrm{rec}}$.

In Sec. 5.1 we outline the designed method, in Sec. 5.2 we show our estimates of the bias as well as its parameterization, and in Sec. 5.3 we show that the remaining dependencies in

---

[1] There is also the possibility that the muon or its secondary electrons produce enough signal in three or more neighbor scintillator strips, but this is so infrequent ($< 0.4\,\%$) that the effect is negligible.

the bias of the corrected number of muons are small. Finally, in Sec. 5.4 we summarize the conclusions.

5.1 Method

The objective of the method is to obtain a parameterization of the bias in the reconstructed number of muons $N_{\mu\,\mathrm{rec}}$. Ideally, we would make such parameterization a function of the impinging zenith angle of the muon θ_μ, of the difference between the module and the muon azimuth angles $\Delta\phi_\mu = \phi_{\mathrm{mod}} - \phi_\mu$, and of the number of injected muons $N_{\mu\,\mathrm{inj}}$. However, these quantities cannot be reconstructed from the detector information. We therefore approximate the angles of the muon to those of the shower: $\theta_\mu \approx \theta$ and $\phi_\mu \approx \phi$ (hence $\Delta\phi_\mu \approx \Delta\phi$). Additionally we parameterize in the reconstructed number of muons $N_{\mu\,\mathrm{rec}}$ instead of in the injected ones $N_{\mu\,\mathrm{inj}}$.

For each module, we estimate the relative difference between the reconstructed and injected number of muons as

$$\varepsilon(\theta, \Delta\phi, N_{\mu\,\mathrm{rec}}) = \frac{N_{\mu\,\mathrm{rec}}}{N_{\mu\,\mathrm{inj}}} - 1. \tag{5.1}$$

Notice that since only module-level events with $N_{\mu\,\mathrm{inj}} > 0$ are relevant, the fraction in Eq.(5.1) never takes infinite values. The bias is then the average ε over a certain bin over the parameters.

The proposed parameterization of the bias is the function

$$f_{\mathrm{bias}}(\theta, \Delta\phi, N_{\mu\,\mathrm{rec}}) = a(\theta) + b(\theta)\,|\sin(\Delta\phi)| + c(\theta)\,\log_{10} N_{\mu\,\mathrm{rec}}, \tag{5.2}$$

where $a(\theta)$, $b(\theta)$, and $c(\theta)$ are modeled phenomenologically as

$$a(\theta) = a_0 + a_1 \sin^2\theta, \tag{5.3}$$

$$b(\theta) = b_0 + b_1 \sec\theta, \tag{5.4}$$

$$c(\theta) = c_0 + c_1 \sin^2\theta, \tag{5.5}$$

being a_0, a_1, b_0, b_1, c_0, and c_1 parameters of the fit. The function of Eq. (5.2) reflects the nature of the phenomenon, which is mostly related to the geometrical projection of the impulse of the muon on the scintillator, and hence it is linear in $|\sin(\Delta\phi)|$, allowing the intercept $a(\theta)$ and slope $b(\theta)$ to vary with θ. The last term of Eq. (5.2), linear in $N_{\mu\,\mathrm{rec}}$, is introduced after noticing that for fixed θ and $\Delta\phi$, the bias depends on the number of injected or reconstructed muons (see Sec. 5.2). The need for this term can be understood from the fact that the probability of double-counting a corner-clipping muon depends on the number of available scintillator bars. If a corner-clipping muon transverses two scintillator bars, and either is already occupied by the pattern match of another muon that impinged it, the corner-clipping muon will not be counted twice. Furthermore, this last term can also account for the biases intrinsic to the estimation of $N_{\mu\,\mathrm{rec}}$.

A similar parameterization to Eq. (5.2) was used in Refs. [90, 97], without the correction term in $N_{\mu\,\mathrm{rec}}$, and using other models of $a(\theta)$ and $b(\theta)$. We keep the expressions of Eqs. (5.3-5.4) since they are comparably simpler and we find no significant improvement in the fit by using the expressions of Refs. [90, 97].

Finally, the corrected muon density $N_{\mu\,\mathrm{corr}}$ is computed as

$$N_{\mu\,\mathrm{corr}} = \frac{N_{\mu\,\mathrm{rec}}}{1 + f_{\mathrm{bias}}(\theta, \Delta\phi, N_{\mu\,\mathrm{rec}})}. \tag{5.6}$$

The reconstructed and injected number of muons per module are retrieved from detector simulations. To obtain these simulations, we employ a library of 3000 proton- and 3000 iron-initiated air-shower simulations, that use EPOS-LHC and UrQMD as the high- and low-energy hadronic interaction models respectively, ran using Corsika v7.7402 [23]. The logarithmic energy of the primary is uniformly distributed in $17.2 \lesssim \log_{10}(E/\mathrm{eV}) \lesssim 18.7$ and the arrival directions correspond to an isotropic distribution with zenith angles in $0° \lesssim \theta \lesssim 48°$. Each of these Monte Carlo simulations is thrown five times on the detector array, simulating and reconstructing the response of the UMD SiPM modules, using $\overline{\text{Off}}\underline{\text{line}}$ (git version eb5cd9ad[2]) [70]. The energy, core position, and geometry of the shower are reconstructed using the SD array. The module-wise estimates of the reconstructed number of muons $N_{\mu\,\mathrm{rec}}$ are obtained using the 1-bin strategy (see Chapter 4).

It is important to understand the difference between the detector simulations developed in Chapter 4 and those ran using $\overline{\text{Off}}\underline{\text{line}}$ for this chapter. The differences are that $\overline{\text{Off}}\underline{\text{line}}$ simulates the corner-clipping effect, background muons, and digital noise. Due to the latter, the solution to the response of the electronics to the signal caused by the photo-electrons and the noise is computed numerically.

The fit is performed as follows. We first select only modules with $1.01 < N_{\mu\,\mathrm{rec}} \leq 170$. The upper cut is set at the number of muons at which the expected saturation fraction of events is 1 % (in the data analysis of Chapter 6 we use an equivalent cut). The lower cut is explained from the formula for counting muons with the 1-bin counting strategy (c.f. Eq. (4.8)): we see that if there is only one pattern match ($k = 1$), the reconstructed number of muons is also 1. In such case there is no bias to correct. We set the cut at 1.01 to allow for numerical inaccuracies. We then proceed to group the module-wise estimates of ε (c.f. Eq. (5.1)) in $(\sin^2\theta, \Delta\phi, \log_{10} N_{\mu\,\mathrm{rec}})$-bins, and in each bin the mean and standard deviation of the mean of ε is computed. Then we build a global χ^2 function

$$\chi^2 = \sum_{i,j,k} \frac{\left(\langle\varepsilon\rangle\,(\theta_i, \Delta\phi_j, N_{\mu\,\mathrm{rec}\ k}) - f_{\mathrm{bias}}(\theta_i, \Delta\phi_j, N_{\mu\,\mathrm{rec}\ k})\right)^2}{\sigma\left[\langle\varepsilon\rangle\,(\theta_i, \Delta\phi_j, N_{\mu\,\mathrm{rec}\ k})\right]^2} \tag{5.7}$$

which we minimize using Minuit/Migrad in the software ROOT [98] to obtain the parameters of Eqs. (5.3-5.5).

It is relevant to add that the parameters in Eq. (5.2) will not necessarily result the same if other counting strategies, that provide different estimators of the number of muons, and therefore are subject to other intrinsic biases. We remind the reader that we use the 1-bin counting strategy (c.f. Chapter 4).

5.2 Parameterization of the bias

We show the results of the global fit in two-dimensional plots for an easier visualization. To achieve this, we plot the bias as a function of one variable, in bins of a second variable in different colors, and for a given bin of the third variable.

Figure 5.1 shows the dependence of the bias of Eq. (5.1) as a function of $|\sin\Delta\phi|$, for different $\sin^2\theta$ bins, at small $N_{\mu\,\mathrm{rec}}$ values on the left panel ($1.01 < N_{\mu\,\mathrm{rec}} \leq 10$), and at larger values on the right panel ($100 < N_{\mu\,\mathrm{rec}} \leq 125$). We also show the result of the global fit using the parameterization of Eqs. (5.2-5.5), which describes the observed biases to a good degree. As expected, the bias grows with $|\sin\Delta\phi|$, and grows more steeply the larger θ is. This is a purely geometrical consequence of clipping corner muons impinging on the modules.

---

[2] The used version incorporates separately the contents of commit bfecb9af, which fixes a bug in the reconstructed distances to the shower axis measured on the shower plane: before the fix, said distances were not properly projected onto the shower plane.

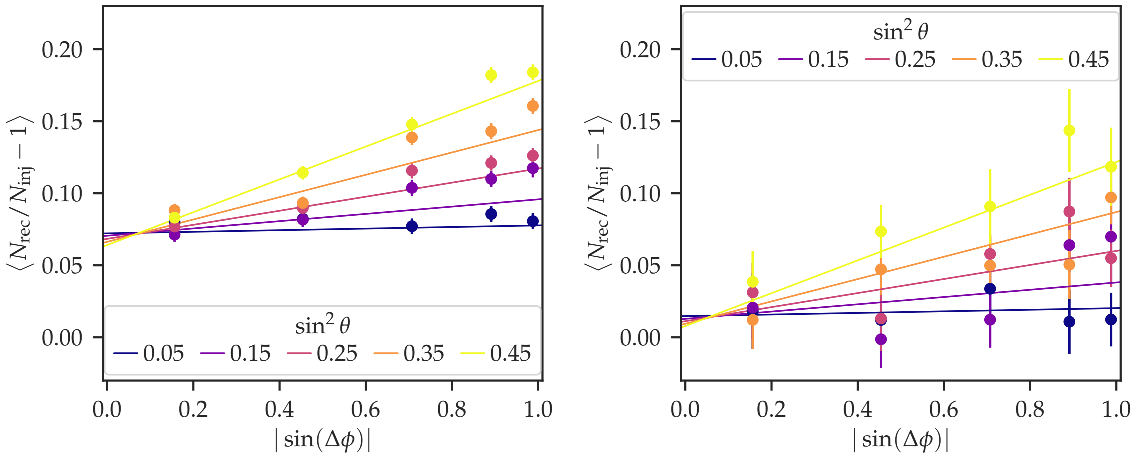


Figure 5.1: Bias in the reconstructed number of muons against the injected ones (see Eq. (5.1)) as a function of $|\sin\Delta\phi|$, for different $\sin^2\theta$ bins in different colors. In the left panel $1.01 < N_{\mu\,\rm rec} \leq 10$, while in the right panel $100 < N_{\mu\,\rm rec} \leq 125$. The global fit is superimposed.

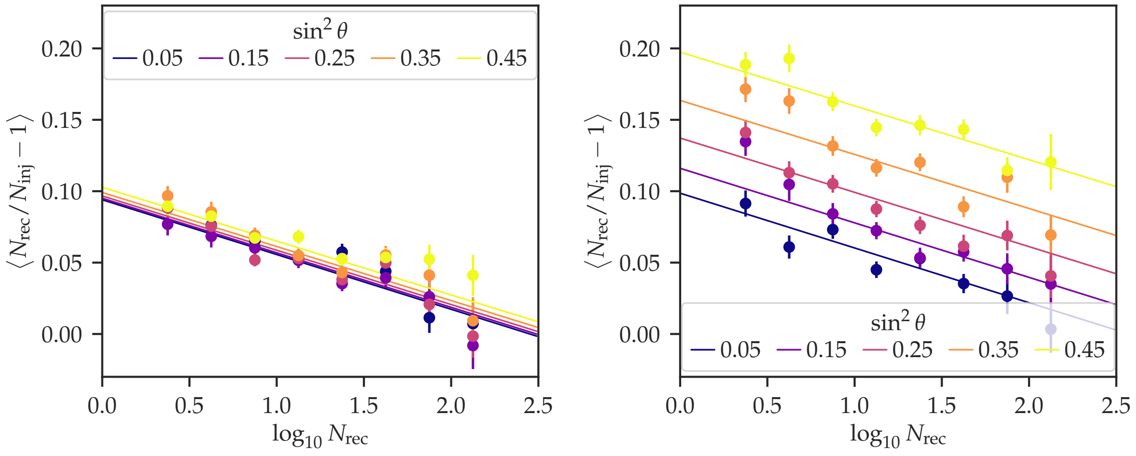


Figure 5.2: Bias in the reconstructed number of muons against the injected ones (see Eq. (5.1)) as a function of the logarithm of the reconstructed number of muons, for different $\sin^2\theta$ bins in different colors, and for two $\Delta\phi$ bins: $7^\circ \lesssim \Delta\phi \lesssim 14^\circ$ in the left panel, and $76^\circ \lesssim \Delta\phi \lesssim 83^\circ$ in the right panel. The global fit is superimposed.

The fact that the intercept is not zero can be attributed to the distribution of the zenith and azimuth angles of the muons around those of the shower. Therefore, even at $\theta = 0$ or $\Delta\phi = 0$, some muons can still transverse two bars.

Figure 5.2 shows the dependence of the bias (c.f. Eq. (5.1)) with $\log_{10} N_{\mu\,\rm rec}$ for different $\sin^2\theta$ bins (in different colors), and for $7^\circ \lesssim \Delta\phi \lesssim 14^\circ$ (left panel) and $76^\circ \lesssim \Delta\phi \lesssim 83^\circ$ (right panel). Superimposed, the result of the global fit using the parameterization, which, once again, we see that it describes the biases to a good degree. It can be seen that the bias is positive for small $N_{\mu\,\rm rec}$ values, dominated by the double counting of corner-clipping muons. The bias decreases with $N_{\mu\,\rm rec}$, as the probability of double-counting becomes smaller as more scintillator bars are occupied. As expected, the biases are larger for larger $\Delta\phi$ values. This figure evidences the need of including the linear term in $\log_{10} N_{\mu\,\rm rec}$ in the parameterization of the bias (c.f. Eq. (5.1)).

Finally, Fig. 5.3 shows the dependence of the bias (see Eq. (5.1)) with $\log_{10} N_{\mu\,\rm rec}$ for different $|\sin\Delta\phi|$ bins (in different colors), for $0.10 < \sin^2\theta \leq 0.20$ (left panel), and for

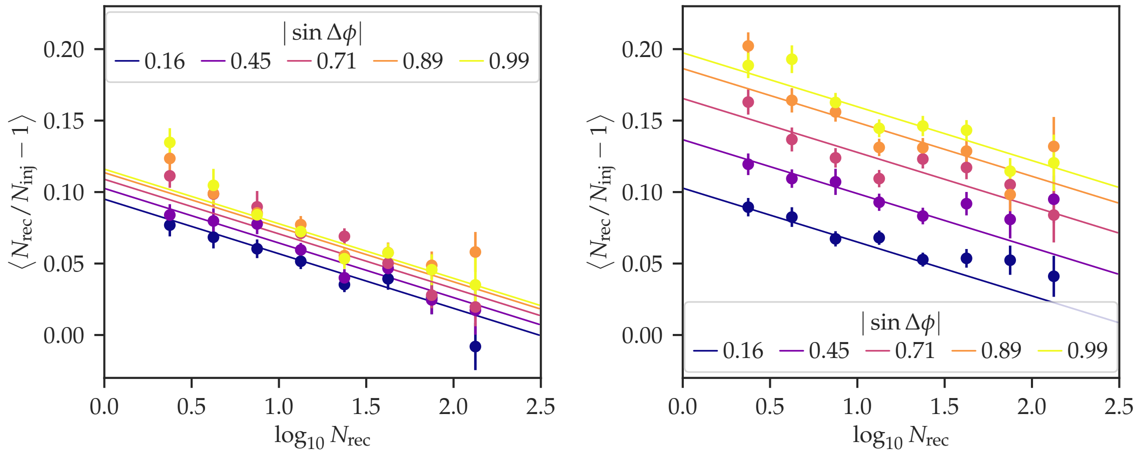


Figure 5.3: Bias in the reconstructed number of muons against the injected ones (c.f. Eq. (5.1)) as a function of the logarithm of the reconstructed number of muons, for different $|\sin\Delta\phi|$ bins in different colors, and for two $\sin^2\theta$ bins: $0.10 < \sin^2\theta \leq 0.20$ (left panel), and $0.40 < \sin^2\theta \leq 0.50$ (right panel). The global fit is superimposed.

$0.40 < \sin^2\theta \leq 0.50$ (right panel). Superimposed, the global fit. We observe the same as in Fig. 5.2 but highlighting a different dependency.

A list of the values of the fitted parameters can be found in Table 5.1.

a_0	a_1	b_0	b_1	c_0	c_1
0.094 ± 0.004	-0.021 ± 0.014	-0.34 ± 0.02	0.34 ± 0.02	-0.038 ± 0.002	0.002 ± 0.009

Table 5.1: Resulting parameters of the global fit to the bias in the reconstructed number of muons (see Eqs. (5.2-5.5)).

It is relevant to add that the results of Ref. [97] are not comparable to the ones of this work, not only for the extra correction term on $N_{\mu\,\mathrm{rec}}$ that we introduced, but because of several differences: (1) The analysis of Ref. [97] is of PMT simulations, while this analysis uses SiPM simulations. (2) The estimator of $N_{\mu\,\mathrm{rec}}$ is different to the one used in this work (for more information see Chapter 4), yielding different intrinsic biases. And (3) the module azimuths of said analysis were incorrect, artificially rising the estimated bias for azimuths close to 0° and 180° and lowering it for azimuths close to 90° and 270°.

5.3 Analysis of the remaining bias

Having the parameterization of the bias described in Sec. 5.2, we use it to correct the reconstructed number of muons via Eq. (5.6). In the following we analyze the dependencies of the remaining bias, i.e., the bias in the corrected number of muons, estimated from the mean of the relative difference between the corrected and injected number of muons:

$$\varepsilon^* = \frac{N_{\mu\,\mathrm{corr}}}{N_{\mu\,\mathrm{inj}}} - 1. \tag{5.8}$$

Figure 5.4 shows the remaining bias as a function of $\sin^2\theta$ (top left), of $|\sin\Delta\phi|$ (top right), of $\log_{10} N_{\mu\,\mathrm{rec}}$ (bottom). The points represent the bias (mean of ε^*) in the corrected number of muons for the bin in the x-axis variable, and the error bars represent the standard deviation

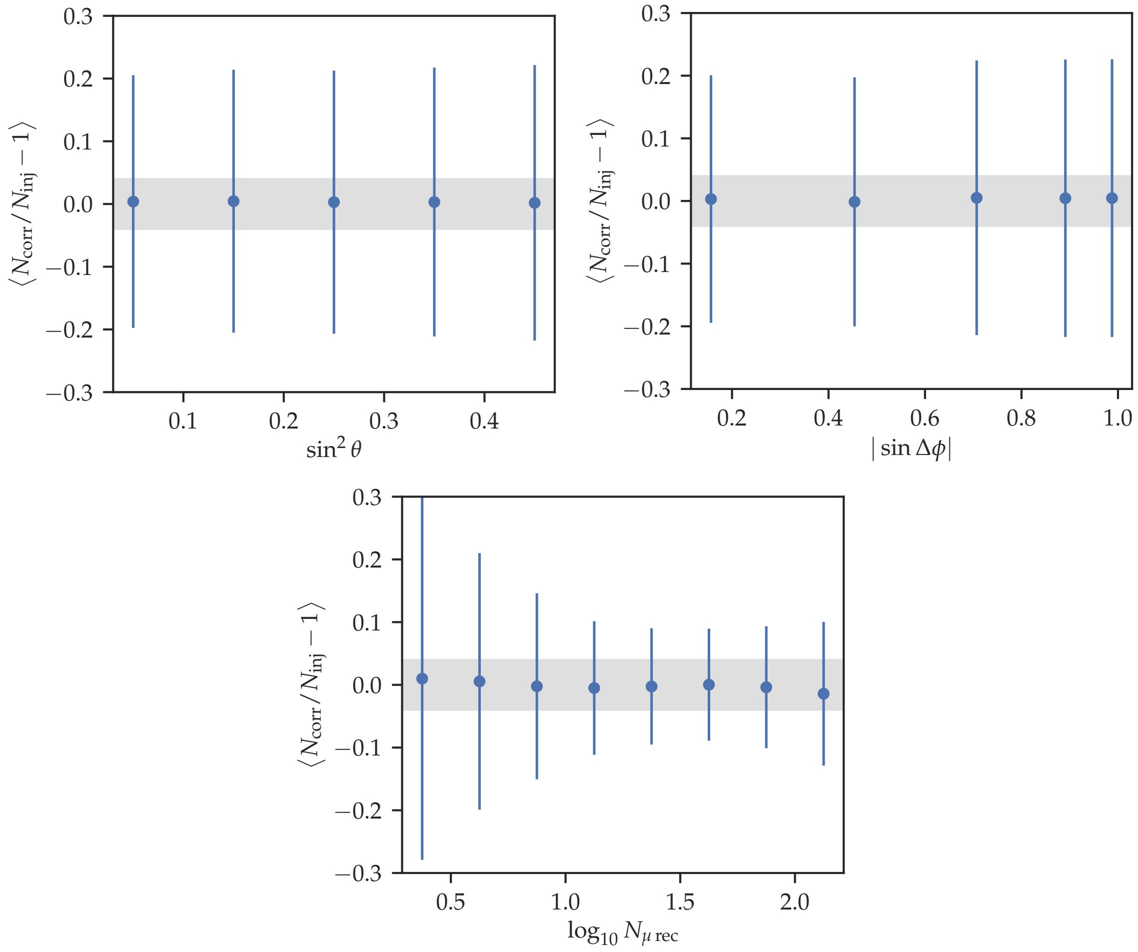


Figure 5.4: Bias in the corrected number of muons against the injected one (see Eq. (5.8)) as a function of $\sin^2\theta$ (top left), of $|\sin\Delta\phi|$ (top right), and of $\log_{10} N_{\mu\,\mathrm{rec}}$ (bottom). The points represent the mean of ε^* and the error bars the standard deviation of ε^*. The gray band marks ±4 %.

(not the standard deviation of the mean) of ε^*. In almost all cases, the standard deviation of the mean is contained in the marker. All figures have a gray band enclosing ±4 % bias. We observe that the remaining bias has negligible dependencies with the variables in which the bias is parameterized ($\sin^2\theta$, $|\sin\Delta\phi|$ and $\log_{10} N_{\mu\,\mathrm{rec}}$).

Furthermore, in Fig. 5.5 we analyze the dependence of the remaining bias as a function of the logarithmic reconstructed energy (top left), of the logarithmic distance to the shower core (top right), of the logarithmic number of injected muons $\log_{10} N_{\mu\,\mathrm{inj}}$ (bottom left), and as a function of the primary (bottom right). The remaining bias as a function of $\log_{10} N_{\mu\,\mathrm{inj}}$ is, as expected, not null, but it is well contained within ±4 %. The largest bias is given at short distances from the shower core, but it is also still contained within ±4 %. The remaining bias as a function of the primary is remarkably small. We therefore can expect that the parameterization of the bias is, to a good approximation, independent of the chosen high-energy hadronic interaction model. This is because the difference between proton and iron air-shower simulations of a same high-energy hadronic interaction model are much larger than the difference between air-shower simulations of a same primary using different high-energy hadronic interaction models.

It is expected that there are spurious dependencies of the corrected number of muons. This is because the chosen model (Eqs. (5.2-5.5)), linear in the parameters, is a good approx-

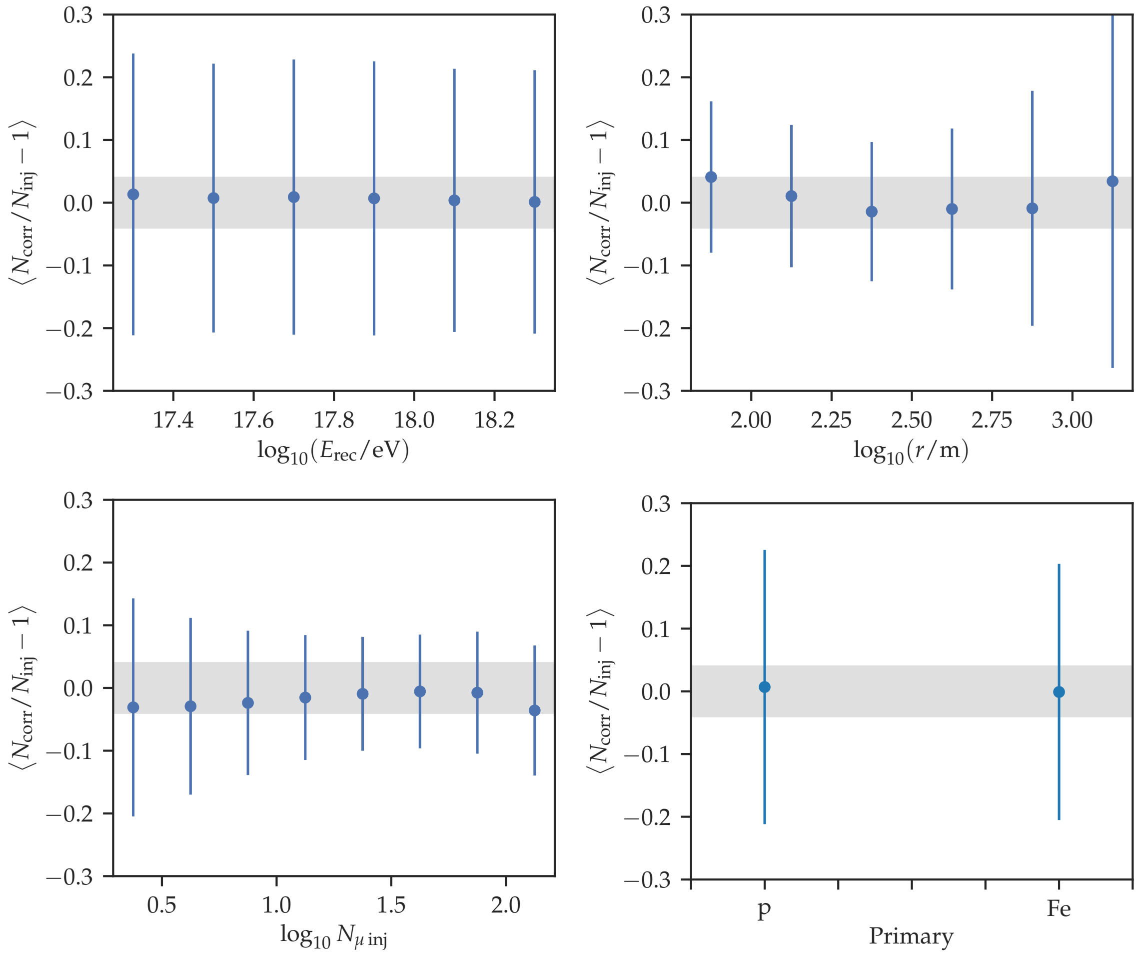


Figure 5.5: Bias in the corrected number of muons against the injected one (see Eq. (5.8)) as a function of $\log_{10}(E/\mathrm{eV})$ (top left), of $\log_{10}(r/\mathrm{m})$ (top right), of $\log_{10} N_{\mu\,\mathrm{inj}}$ (bottom left), and of the primary type (bottom right). Like for Fig. 5.4, the points represent the mean of ε^*, the error bars the standard deviation of ε^*, and the gray band encloses ±4 %.

imation but may not describe the full complexity of the dependencies of the bias. At this point we can add that we attempted to add a quadratic term in $\log_{10} N_{\mu\,\mathrm{rec}}$ which did not significantly improve the performance of the fit nor the remaining dependencies. The other reason to expect spurious dependencies is that the assumptions that the zenith and relative azimuth of the shower can be used in place of those of the muon are only approximately met. Nevertheless, it is remarkable that the remaining dependencies can be contained within only ±4 %. Hence, we find the performed parameterization sufficiently precise.

5.4 Summary

In this chapter we parameterized the bias in the reconstructed number of muons against the injected ones for the UMD (SiPM) using detector simulations ($\overline{\mathrm{Off}}\underline{\mathrm{line}}$) of Monte Carlo air showers. We saw that the bias in $N_{\mu\,\mathrm{rec}}$ relative to the injected number of muons $N_{\mu\,\mathrm{inj}}$ is clearly dominated by the double-counting of corner-clipping muons, i.e., muons that deposit enough signal in two neighbor scintillator bars such that they are counted twice. We proposed that the bias can be described as a function of the zenith angle of the shower θ, of the difference in azimuth between the module and the shower $\Delta\phi = \phi_\mathrm{mod} - \phi$, and also of the reconstructed number of muons $N_{\mu\,\mathrm{rec}}$. The first two variables, θ and $\Delta\phi$, are expected to be required, since the major source of bias in $N_{\mu\,\mathrm{rec}}$ is the (geometrically dependent) effect of corner-clipping muons. We showed that the last variable, $N_{\mu\,\mathrm{rec}}$, is also required, which is explained because the probability of double counting a corner-clipping muon depends on how many available scintillator bars there are. Furthermore, the last term of the parameterization can also account for intrinsic biases of $N_{\mu\,\mathrm{rec}}$.

The chosen parameterization proved to reproduce the dependencies of the bias to a good degree. Furthermore, we used the parameterization of the bias to correct the reconstructed number of muons, and we analyzed the remaining dependencies of the bias in the corrected number of muons. We found that the remaining dependencies are small. The largest one is with respect to the logarithmic distance to the shower core, specially very close to the core (below 100 m). On the contrary, we found that the dependence with the primary is remarkably small. All in all, the remaining dependencies are well contained within $\pm 4\,\%$.

Chapter 6

Characterization of the muon lateral distribution function

In Chapters 4 and 5 we developed a new counting strategy and corrected for the total bias, to obtain an unbiased estimate of the number of muons impinging on a module of the Underground Muon Detector (UMD) of the Pierre Auger Observatory.

In this chapter we use the direct measurements of the muons of air showers in order to characterize the muon lateral distribution function (muon LDF). The latter is a great tool for composition analyses, as both its size and slope carry information of the primary mass. We compute the average muon LDF from UMD data at different energies and zenith angles. We then study different muon LDF models by fitting them to the data, and analyzing the goodness-of-fit. We also provide parameterizations of the measured muon LDF that can be useful, for example, for building a toy model, for event-wise fits of the muon LDF, or as a reference for the design of future muon detectors.

The chapter is organized as follows: Sec. 6.1 discusses different muon LDF models found in literature, as well as previous parameterizations of UMD data. In Sec. 6.2 we analyze a particular muon LDF model. In Sec. 6.3 we show the method through which we select the data used in this work, which includes a selection of the energy and zenith angle of the events, as well as cuts in distance to the shower axis, and an outlier detection procedure. Furthermore, in Sec. 6.4 we describe the fitting procedure, the information criteria used to analyze the goodness-of-fit, and the test we use to analyze the uniformity of the residuals. Sec. 6.5 shows the results of the measured and fitted muon LDF, comparing different models. In Sec. 6.5.2 we compare the overall goodness-of-fit of the different models to find which muon LDF model best describes the data. In Sec. 6.6 we provide a parameterization of the muon LDF of data as a function of the energy and zenith angle, which can be useful for making realistic toy models. In Sec. 6.7 we use the aforementioned parameterization and make a toy model of the muon LDF to analyze the systematic uncertainties in the fitted muon LDFs introduced by the fitting procedure used in the work. Finally, Sec. 6.9 summarizes the key findings.

6.1 Previous studies and parameterizations of the lateral distribution of muons

There is no analytical derivation of the muon LDF from first principles. Instead, the most widely used model is the one derived by Greisen [99], adapting the analytical approximation of the electromagnetic LDF, a Nishimura-Kamata-Greisen (NKG) function, to the muon LDF. The NKG function can be expressed as

$$\rho^{\mathrm{NKG}}(r, s, N_e) = \frac{N_e}{r_M^2} \frac{\Gamma(4.5-s)}{2\pi\Gamma(s)\Gamma(4.5-2s)} \left(\frac{r}{r_M}\right)^{s-2} \left(1+\frac{r}{r_M}\right)^{s-4.5}, \tag{6.1}$$

where r_M is the Moliere radius (the radius of a cylinder containing on average 90 % of an the energy deposition of an electromagnetic shower), s is the shower age, N_e is the number of electrons, and Γ is the gamma function. Greisen's adaptation of the NKG function to the muon LDF is as follows [99]:

$$\rho_\mu^{\mathrm{NKG}}(r, E_{\mathrm{th}}) = \frac{14.4\,(r/\mathrm{m})^{-0.75}}{(1+r/320\,\mathrm{m})^{2.5}} \left(\frac{N_\mu}{10^6}\right)^{0.75} \frac{51\,\mathrm{GeV}}{E_{th}+50\,\mathrm{GeV}} \left(\frac{3\,\mathrm{GeV}}{E_{th}+2\,\mathrm{GeV}}\right)^{0.14\,(r/\mathrm{m})^{0.37}}, \tag{6.2}$$

where N_μ is the total number of muons, and E_{th} is the vertical energy threshold of the muons. For the UMD, we can replace E_{th} with 1 GeV. Then, reordering the terms conveniently:

$$\rho_\mu^{\mathrm{NKG}}(r) = 14.4 \left(\frac{N_\mu}{10^6}\right)^{0.75} (r/\mathrm{m})^{-0.75} \left(1+\frac{r}{320\,\mathrm{m}}\right)^{-2.5}$$

We can further introduce some additional terms such that the last two factors, which encapsulate the dependence with r, equal 1 when evaluated at $r = 450\,\mathrm{m}$:

$$\rho_\mu^{\mathrm{NKG}}(r) = 14.4\,(450)^{-0.75} \left(\frac{N_\mu}{10^6}\right)^{0.75} \left(\frac{r}{450\,\mathrm{m}}\right)^{-0.75} \left(\frac{1+r/320\,\mathrm{m}}{1+450\,\mathrm{m}/320\,\mathrm{m}}\right)^{-2.5}$$

Finally, we can rewrite the last expression in the more general form:

$$\rho_\mu^{\mathrm{NKG}}(r) = \rho_{r_{\mathrm{ref}}} \left(\frac{r}{r_{\mathrm{ref}}}\right)^{-\alpha} \left(\frac{1+r/r_0}{1+r_{\mathrm{ref}}/r_0}\right)^{-\beta}, \tag{6.3}$$

where we introduce the reference distance r_{ref}, the distance r_0, and the exponents α and β.

The KASCADE-Grande Collaboration, in their analyses of the muon LDF [100], added an additional term to the NKG to make the muon LDF smaller at larger distances. This function, originally proposed by Lagutin and Raikin for the lateral distribution of electrons [101, 102], is dubbed modified-NKG (mNKG):

$$\rho_\mu^{\mathrm{mNKG}}(r) = \rho_{r_{\mathrm{ref}}} \left(\frac{r}{r_{\mathrm{ref}}}\right)^{-\alpha} \left(\frac{1+r/r_0}{1+r_{\mathrm{ref}}/r_0}\right)^{-\beta} \left(\frac{1+(r/10r_0)^2}{1+(r_{\mathrm{ref}}/10r_0)^2}\right)^{-\gamma}. \tag{6.4}$$

KASCADE-Grande reported that the best-fit values of the parameterization to their measured data were $\alpha = 0.69$, $\beta = 2.39$, and $\gamma = 1.0$, with $r_0 = 320\,\mathrm{m}$, letting the size of the shower as a free parameter to fit all their measured showers. In said case the vertical muon energy threshold is of $E_{th} = 230\,\mathrm{MeV}$. They also reported that the function has problems in fitting data below 100 m [100].

Furthermore, several studies used Eq. (6.4) to fit event-wise the muon LDF computed from UMD Monte-Carlo simulations. In Refs. [13, 103], the exponent α was fixed to 0.75, using $r_0 = 320\,\mathrm{m}$, and ρ_{450}, β and γ were left as free parameters of the fit. In Ref. [14] the parameter γ is not let free but fixed to 2.95.

In another study, UMD PMT data were event-wise fitted using $r_0 = 280\,\mathrm{m}$, and fixing $\alpha = 0.3$ and $\gamma = 4.6$. The mentioned parameters were optimized from simulations. Here, β was parameterized from the average of data as $\beta = \beta(\theta) = 4.4 - 1.1\sec(\theta)$ [104]. Another study on UMD SiPM data used the same fixed parameters but obtained another parameterization of β, namely $\beta(\theta) = 7.65 - 6.62\sec(\theta) + 2.25\sec^2(\theta)$ [105]. However, in Sec. 6.5, we observe that the exponent β depends also on the energy.

It is important to mention at this point that the parameters of the modified-NKG are correlated, and in particular, the exponents are strongly correlated. For this reason, fixing the parameters to very different values does not necessarily imply that the functional shapes are very different.

For the UMD, the reference distance of 450 m was proved to be the optimal distance. This means that ρ_{r_ref} fluctuates the least at $r_\mathrm{ref} = 450\,\mathrm{m}$, compared to other reference distances, for different samples of the same simulated muon LDF [13, 14]. Nevertheless, fluctuations in the reconstruction of the core position were not considered in the derivation of such optimal distance.

Other experiments use a function with an exponential factor. This was initially proposed by Hillas [106] in a phenomenological attempt to parameterize the muon LDF:

$$\rho_\mu^{\mathrm{Hillas}}(r) = A\left(\frac{r}{r_0}\right)^{-\alpha}\exp\left(-\frac{r}{r_0}\right). \tag{6.5}$$

In his work he fixed $r_0 = 600\,\mathrm{m}$, and by fitting the function to the measured muon LDFs he found $\alpha \in [1.40, 1.54]$ and $A \in [0.00441, 31.2]$ for air showers with energies between 10^{15} eV and 10^{19} eV, with a vertical muon energy threshold of 1 GeV.

For example, Tien-Shan data was fitted using the Hillas function, obtaining $\rho(r) = 5.95 \times 10^{-4} r^{-0.7}\exp\left(-\frac{r}{80\,\mathrm{m}}\right)$ (see Ref. [107] and references therein).

We propose a change in Eq. (6.5) to normalize the dependence with r such that it is 1 at r_ref, to simplify the interpretation of the shower size:

$$\rho_\mu^{\mathrm{Hillas}}(r) = \rho_{r_\mathrm{ref}}\left(\frac{r}{r_\mathrm{ref}}\right)^{-\alpha}\exp\left(-\frac{r - r_\mathrm{ref}}{r_0}\right), \tag{6.6}$$

The parameter r_ref would be fixed for each experiment to minimize fluctuations in the muon density. The parameter ρ_{r_ref} of Eq. (6.6) can be expressed as a function of the parameters of Eq. (6.5) as $A(r_\mathrm{ref}/r_0)^{-\alpha}\exp(-r_\mathrm{ref}/r_0)$.

KASCADE also tried to fit muon density measurements with the semiempirical formula proposed by Linsley [108]:

$$\rho_\mu^{\mathrm{Linsley}}(r) = \frac{N}{r_0^2}\left(\frac{\eta - 2}{2\pi}\right)\left(\frac{r}{r_0}\right)^{-1}\left(1 + \frac{r}{r_0}\right)^{-(\eta-1)} \tag{6.7}$$

with $\eta = 3.3 + 1.3\cos\left(0.4\sec(\theta) + 0.9 - 0.4\log_{10}(N/10^8)\right)$. This can be matched to Eq. (6.3) with a specific parameterization of β and ρ_{r_ref} derived from that of η.

Finally, we introduce a function that was not found in literature. We believe it is worth attempting to fit it to UMD data due to its simplicity, as well as its application to electromagnetic LDFs. This is the log-log parabola, or more generally, a polynomial of degree D in $\ln(r)$:

$$\rho_\mu^{\text{log-log}} = \rho_{r_\mathrm{ref}}\exp\left(\sum_{i=1}^{\mathrm{D}} c_i \ln^i\left(\frac{r}{r_\mathrm{ref}}\right)\right), \tag{6.8}$$

where c_i is the coefficient preceding the logarithm to the i-th power.

In this analysis of the muon LDF from SiPM UMD data, we evaluate the functions of Eq. (6.3), Eq. (6.4), Eq. (6.6), and Eq. (6.8).

6.2 Considerations on the functional form of the modified NKG

In this section we analyze the functional form of the modified NKG or KASCADE-Grande function (c.f. Eq. (6.4)), since it is the most widely used model of the muon LDF. The purpose is to understand to which extent the parameters α, β, and γ influence the shape of the function.

We first start by grouping the factors of Eq. (6.4) according to their exponent:

$$\rho_\mu^{\mathrm{mNKG}}(r) = \rho_{r_{\mathrm{ref}}} \underbrace{\left(\frac{r}{r_{\mathrm{ref}}}\right)^{-\alpha}}_{f_\alpha(r)} \underbrace{\left(\frac{1+r/r_0}{1+r_{\mathrm{ref}}/r_0}\right)^{-\beta}}_{f_\beta(r)} \underbrace{\left(\frac{1+(r/10r_0)^2}{1+(r_{\mathrm{ref}}/10r_0)^2}\right)^{-\gamma}}_{f_\gamma(r)}. \tag{6.9}$$

Notice that the evaluation of any of the factors with an exponent equal to zero yields one ($f_x(r)|_{x=0} = 1$), which is equivalent to excluding the factor from the product. For the purpose of this simple analysis, we set $r_{\mathrm{ref}} = 450\,\mathrm{m}$ $r_0 = 320\,\mathrm{m}$, $\alpha = 0.75$, $\beta = 2.6$, and $\gamma = 3.0$. The effects that we describe then correspond to this specific set of values of the parameters, but since they are close to the fixed and fitted values in Sec. 6.5, they provide a useful picture of the behavior of the function.

The left panel of Fig. 6.1 shows a comparison between a modified NKG function with all its factors (black) and with the removal of the α (blue), β (orange) and γ (green) factors. We notice that removing $f_\alpha(r)$ has a very large impact at $r < r_{\mathrm{ref}}$, removing $f_\gamma(r)$ only impacts after $r > 1000\,\mathrm{m}$, and removing $f_\beta(r)$ has a large impact in the whole distance range. Of course, by the design of the modified NKG, the value of the function at $r = r_{\mathrm{ref}}$ (450 m in this case) is $\rho_{r_{\mathrm{ref}}}$ independently of the removal of any $f_x(r)$ factor. The right panel of Fig. 6.1 evidences the difference in the logarithm of the muon density between the partial and full muon LDFs. The deviations can be significant, larger than one order of magnitude, especially very close or very far from the core. In this plot it is easier to see that the $f_\gamma(r)$ factor only impacts at large distances from the shower core.

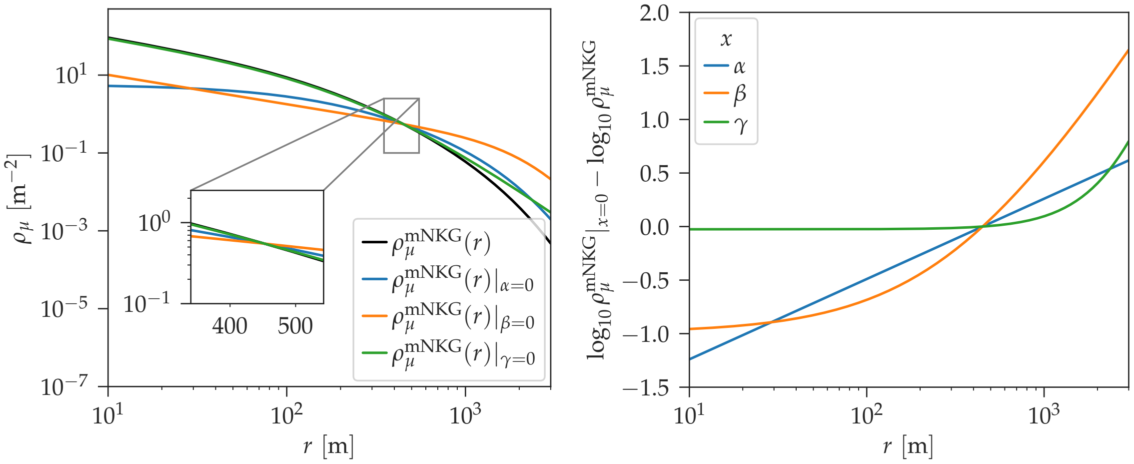


Figure 6.1: Left panel: Modified NKG muon LDF (Eq. (6.9)) with all its factors (black) compared to the same function removing the α (blue), β (orange), or γ (green) factor, in double logarithmic scale. The inset shows that all the plotted functions equal the same at $r = 450\,\mathrm{m}$, as by design the $\rho(r = 450\,\mathrm{m}) = \rho_{r_{\mathrm{ref}}}$. Right panel: Deviation in the logarithmic muon density with respect to the full modified NKG as a function of the distance (logarithmic scale), by removing the α (blue), β (orange), or γ (green) factor.

Furthermore, we want to know how much a change in one of the parameters (α, β, and γ) impacts the muon LDF. For this we plot in Fig. 6.2 the absolute value of the derivative of the muon density with respect to the parameter divided by the muon density as a function of the distance to the shower axis, evaluated at the aforementioned reference values. Notice that the derivatives are all positive-valued at $r < r_{\mathrm{ref}}$ and negative-valued at $r > r_{\mathrm{ref}}$; thus at $r = r_{\mathrm{ref}}$ the derivatives are zero which in log-scale drop to $-\infty$. This is because the parameterization at $r = r_{\mathrm{ref}}$ evaluates to $\rho_{r_{\mathrm{ref}}}$ irrespective of the values of α, β, and γ. We see that $|d\rho_\mu(r)/d\alpha|/\rho_\mu$ has the largest values, followed closely by $|d\rho_\mu(r)/d\beta|/\rho_\mu$, and lastly by $|d\rho_\mu(r)/d\gamma|/\rho_\mu$.

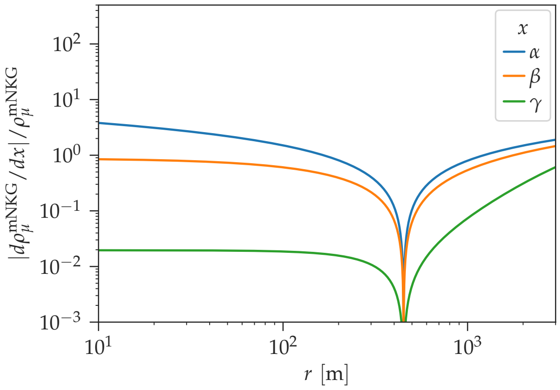


Figure 6.2: Absolute value of the derivative of the muon density with respect to α (blue), β (orange), and γ (green), divided by the muon density, as a function of the distance to the shower axis r, in double logarithmic scale.

We are now interested in knowing in which r range each factor is dominant for the slope of the muon LDF $d\rho^{\mathrm{mNKG}}(r)/dr$. The left panel of Fig. 6.3 shows a comparison of the (negative of the) slope of the modified NKG. The slope with all its factors is in black, the slope evaluated at $\alpha = 0$ is in blue, at $\beta = 0$ in orange, and at $\gamma = 0$ in green. The right panel of Fig. 6.3 evidences the deviation of the slope with one factor removed with respect to the slope with all the factors. For this we plot the absolute value of the difference between the slope without the different factors and the slope with all its factors. Close to the core all differences are negative, while far from the core the differences are positive. When the absolute value of the difference is equal to zero the log-scale shows a drop to $-\infty$. Similar to what we see in the left panel of Fig. 6.1 for the logarithmic muon density, the α and β factors have approximately the same impact, in this case on the slope, except at distances above 1000 m where the β factor dominates. In comparison, the γ factor has a very small impact.

Moreover, we analyze how much a change in one of the parameters (α, β, and γ) impacts the slope of the muon LDF. Figure 6.4 shows a comparison of $d^2\rho^{\mathrm{mNKG}}(r)/dr dx$ with $x = \{\alpha, \beta, \gamma\}$. We can see that $d^2\rho^{\mathrm{mNKG}}(r)/dr\, d\alpha$ dominates up to $r \sim 1000$ m, followed by the $d^2\rho^{\mathrm{mNKG}}(r)/dr\, d\beta$, and above $r \sim 1000$ m the latter is the dominant.

It is interesting to notice that in general $f_\alpha(r)$ is a significant factor. However α is in practice usually fixed to one value for all events in a wide range of energies and zenith angles. This is normally done because α and β are so strongly correlated that it would be very difficult to fit both. Nevertheless, the modified NKG function can still be well tuned at a fixed α by varying β. This point is illustrated in Fig. 6.5, where we plot a modified NKG with β taking the values 0, 1, 2, and 3.

Finally, we show how different values of r_0 affect the modified NKG in Fig. 6.6. Here we show the modified NKG with the reference parameters, and three values of r_0 found in different parameterizations from literature. From Fig. 6.6 we can see that the effect in the shape of the modified NKG is not large. With respect to fits to data, an analysis showed that a change in r_0 changes mostly the exact optimal values of the rest of the parameters as well as

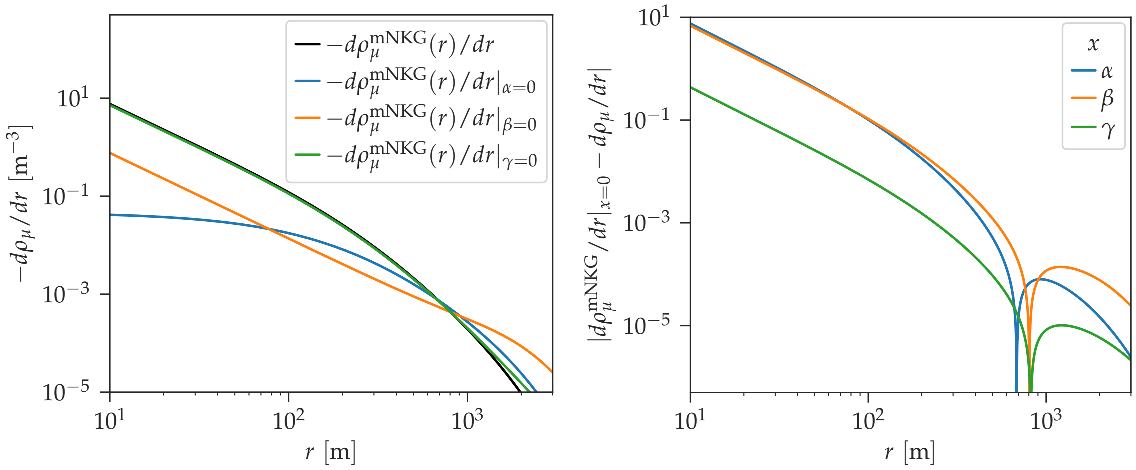


Figure 6.3: Left panel: Slope of the modified NKG muon LDF (Eq. (6.9)) with all its factors (black) compared to the same removing the α (blue), β (orange) or γ (green) factor, in double logarithmic scale. Right panel: Deviation in the slope of the muon density of the modified NKG muon LDF as a function of the distance (logarithmic scale), with respect to the slope of the slope including all factors, by removing the α (blue), β (orange) or γ (green) factor.

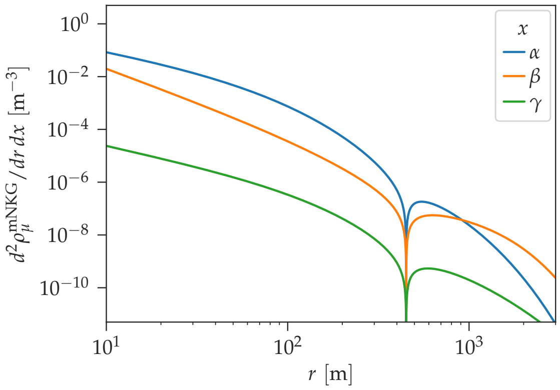


Figure 6.4: Second derivative of the muon density with respect to the distance to the shower axis r, and to α (blue), β (orange), or γ (green), as a function of r in double logarithmic scale.

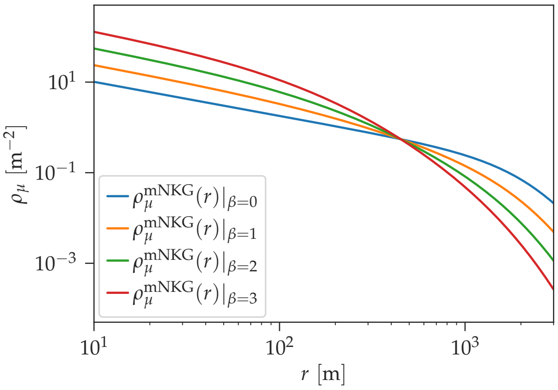


Figure 6.5: Muon density of a modified NKG muon LDF as a function of the distance to the shower axis. Different values of β are shown in different colors.

their correlation coefficient, but that it does not significantly change the quality of the fit (i.e.

the goodness-of-fit is very similar) [109]. This suggests that the rest of the parameters can "accomodate" to a specific value of r_0. Because of this, and because of the limited capability of fitting many parameters to data, we fix $r_0 = 320\,\text{m}$ in the rest of this work. This value was found to be optimal in fits to UMD simulations in Refs. [13, 103].

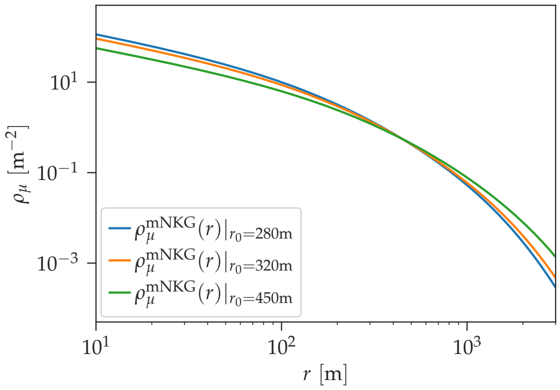


Figure 6.6: Muon density of a modified NKG muon LDF as a function of the distance to the shower axis. Different values of r_0 are shown in different colors.

6.3 Data selection

In this section we describe how the data is selected and prepared before the physics analyses are performed. The dataset consists of the reconstructed data acquired within the period from January 2018 to December 2021 by the UMD modules instrumented with SiPMs. For the reconstruction we use $\overline{\text{Off}}\underline{\text{line}}$ (git version eb5cd9ad, see Sec. 5.1). The muon counting strategy is the 1-bin strategy (see Chapter 4) used with a 12-bin (37.5 ns) single-muon pattern. We use the estimator of the mean number of muons μ (see Chapter 4). Additionally, the muon densities are also bias-corrected against simulations as described in Chapter 5.

Since the trigger of the UMD modules is subordinate to the trigger of its paired SD station, every event is a hybrid UMD-SD event. Nevertheless, the reconstruction of the UMD events could be performed independently of the SD event reconstruction. However, at the moment of writing this thesis, the geometry reconstruction is not implemented for SiPM data. Therefore, the distances from the UMD modules to the shower axis are inferred from the SD geometry reconstruction.

It is relevant to add that the reconstruction already excludes ill-behaving data ("bad periods") as described in Ref. [105]. Additionally, we exclude module 103 from counter 1764 in the period from 16.05.2020 to 01.06.2020 due to firmware problems [110]. Finally, we also exclude counter 1622 because, at the moment of writing this thesis, the azimuth angles of the modules have not yet been measured on the field, which are crucial for the bias correction (c.f. Chapter 5).

6.3.1 Minimum energy and zenith angle selection for full trigger efficiency

It is important to select events within the full trigger efficiency regime to ensure an unbiased muon density sample. If we selected events below full efficiency, the muon density distributions would be biased towards upward fluctuations.

Because the UMD is subject to the SD trigger, the trigger efficiency is determined by the latter. One of the most pessimistic estimates of the T4 efficiency of the 750 m array including TOTd/MoPS triggers (as this work does) is the parameterization provided in Ref. [111].

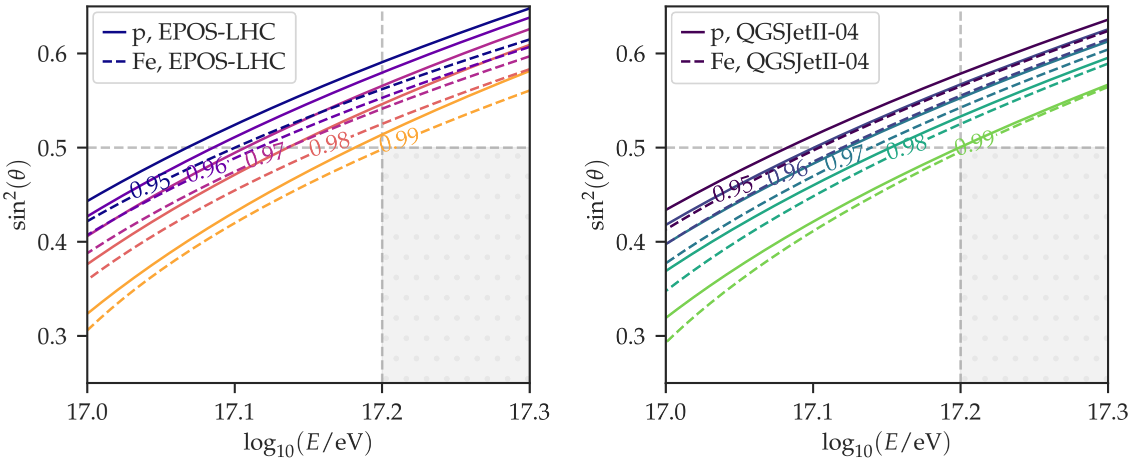


Figure 6.7: Constant differential T4 trigger efficiency (contours) as a function of $\log_{10}(E/\mathrm{eV})$ and $\sin^2\theta$, for simulations of proton (solid lines) and iron (dashed lines) primaries, performed with the high-energy hadronic interactions model EPOS-LHC (left panel) and QGSJetII-04 (right panel). These are parameterizations extracted from Ref. [111]. A label and a color identify each constant efficiency contour. Two gray dashed lines identify the minimum energy and maximum zenith angle of the selected data, which lie in the gray shaded area.

Figure 6.7 shows a contour plot of the efficiency as a function of $\log_{10}(E/\mathrm{eV})$ and $\sin^2\theta$, for proton and iron simulations, for the hadronic models EPOS-LHC and QGSJetII-04, following said parameterization.

Selecting an energy and zenith angle range within full trigger efficiency is rather arbitrary. However, it is convenient not to choose very high zenith angles θ. This is because the effective detection area is proportional to $\cos\theta$ ($A_{\mathrm{eff}} = A\cos\theta$), leading to larger statistical uncertainties for larger zenith angles. Therefore, we select events in the zenith angle range $0^\circ \leq \theta \leq 45^\circ$. With this zenith angle selection, we can see that requesting $\log_{10}(E/\mathrm{eV}) \geq 17.2$ ensures that the T4 efficiency is above 98 %. The selected data lie in the gray shaded area of Fig. 6.7.

It is relevant to add that we find artifacts in the SD geometry reconstruction for energies below the full T4-efficiency threshold. The artifact consists in less reconstructed cores on the imaginary lines that connect the counters. This constitutes another important reason to work only in the full efficiency regime. Since we do so, this artifact is not a problem in the selected data.

6.3.2 Outlier detection: Candidate counters with null and small signal

When analyzing the muon density as a function of the distance to the shower axis, outliers are clearly revealed. If not removed, they can bias the mean muon density, which can spoil the convergence of the fit.

Since the UMD trigger is subordinate to the SD trigger, it is possible that a candidate counter measures zero signal. In Sec. 6.3.3 we explain a very general outlier detection method, which is based on the logarithm of the measured muon density. However, because we cannot use the logarithm of a null value, we have to treat this case separately.

For the purpose of being conservative, we apply this outlier detection method (as well as the one in Sec. 6.3.3) on bias-uncorrected counter muon densities. Nevertheless, we afterwards repeat the procedures on bias-corrected module muon densities, which yield very few extra outliers.

A healthy counter with null signal should be paired to a small SD signal. In Fig. 6.8 we plot the muon density in non-null (blue) and null (orange) candidate counters as a function of the signal of the paired SD station. We find counters with null or small muon densities that have a paired SD counter with large signal. We exclude them with two conditions: the first one requires $\rho_\mu \leq 2.0\,\mathrm{m}^{-2}$ and $S_\mathrm{SD} \geq 1000\,\mathrm{VEM}$, and the second one requires $\rho_\mu \leq 0.25\,\mathrm{m}^{-2}$ and $S_\mathrm{SD} \geq 150\,\mathrm{VEM}$. These two conditions are shown as gray-shaded areas in Fig. 6.8.

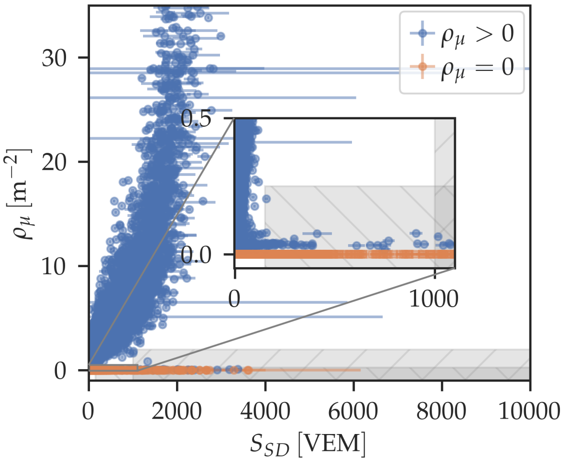


Figure 6.8: Muon density of non-null (blue) and null (orange) candidate counters as a function of the signal of the paired SD station. The inset highlights small muon densities and SD signals. The points that lie within the gray-shaded areas are tagged as outliers.

We detect a total of 34 outliers in uncorrected counter muon densities in this way, a small number compared to the total number of healthy counter muon densities, which is larger than 170, 000. Applying the method to bias-corrected module densities yields no extra outliers.

6.3.3 Outlier detection: Candidate counters with positive signal

To identify outliers in positive muon signals, several methods were attempted. The best results were obtained by using a modified z-score based on the logarithmic muon density. For every candidate counter with positive signal i, its z-score is computed as:

$$\text{z-score}_i = \frac{\log_{10}(\rho_{\mu,i}) - \text{median}[\log_{10}(\rho_\mu)]}{1.48\,\text{MAD}[\log_{10}(\rho_\mu)]}, \tag{6.10}$$

where the median and the median absolute deviation (MAD) are computed over the muon densities, for each candidate counter with positive signal, within a $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta,\ \log_{10}(r/\mathrm{m}))$-bin. The bin widths are $\Delta\log_{10}(E/\mathrm{eV}) = 0.2$, $\Delta\sin^2\theta = 0.1$, and $\Delta\log_{10}(r/\mathrm{m}) = 0.1$. The denominator, $1.48\,\text{MAD}[\log_{10}(\rho_\mu)]$, constitutes a robust estimate of the standard deviation.

The method for detecting outliers consists of three steps:

1. The median and the MAD are computed for each $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta,\ \log_{10}(r/\mathrm{m}))$ bin, using all data points that are not already tagged as outliers.

2. The z-score is computed for each data point that is not already an outlier.

3. Data points that have a z-score such that $|\text{z-score}_i| > 5$ are tagged as outliers.

For self-consistency, all steps are repeated until no outliers are detected in step 3. For our complete dataset we needed four iterations, although after the first one very few extra outliers are found (< 10).

Figure 6.9 shows an example of the outlier detection procedure in the first iteration, corresponding to events with energies in $17.8 \leq \log_{10}(E/\mathrm{eV}) \leq 18.0$ and zenith angles below 18° ($0.00 \leq \sin^2\theta \leq 0.10$). On the left panel we show the z-scores, and on the right panel we show the muon LDF. On the right panel, the profile corresponds to the median logarithmic muon density on a given logarithmic distance bin, and the error bars are its corresponding MAD. Note that these points map to (0 ± 1) in the z-score of the left panel. In both figures, the gray circles represent non-outliers. The green stars are the data points that fulfill the exclusion criterium and are therefore tagged as outliers. It is evident that the points constitute outliers and that the procedure correctly identifies them.

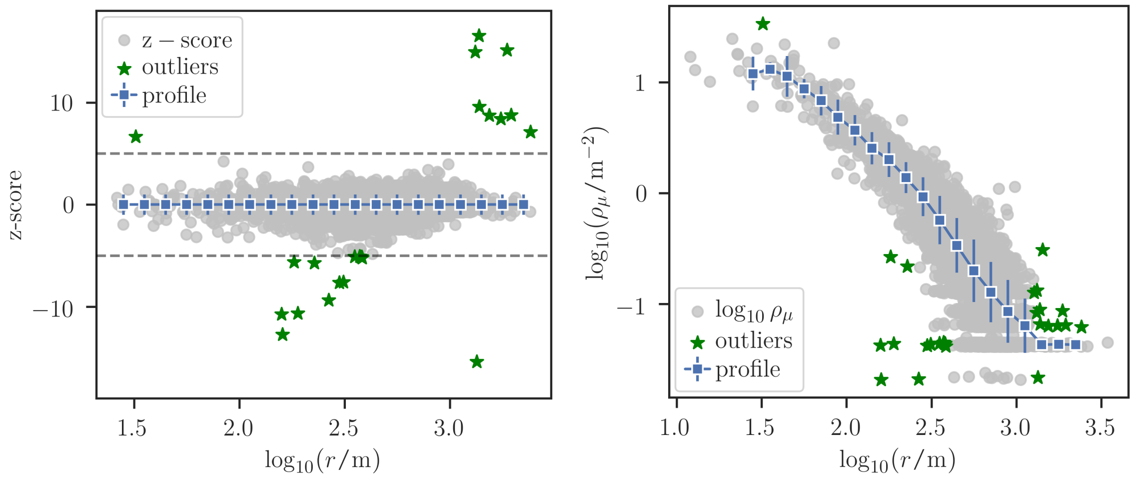


Figure 6.9: Z-score (c.f. Eq. (6.10)) (left panel) and logarithmic muon density (right panel) as a function of the distance to the shower axis, for all data points with $17.8 \leq \log_{10}(E/\mathrm{eV}) \leq 18.0$ and zenith angles below 18° ($0.00 \leq \sin^2\theta \leq 0.10$). Non-outliers are depicted with gray circles. Green stars represent points that fulfill $|\text{z-score}_i| > 5$, and are therefore tagged as outliers.

We tag a total of 1389 uncorrected counter muon densities as outliers with this procedure, which is still very small compared to the number of healthy counter measurements ($> 170{,}000$), constituting less than 1 % of the total counter measurements. Repeating the procedure on bias-corrected module densities yields 43 extra outliers, out of more than $500{,}000$ total module-level measurements.

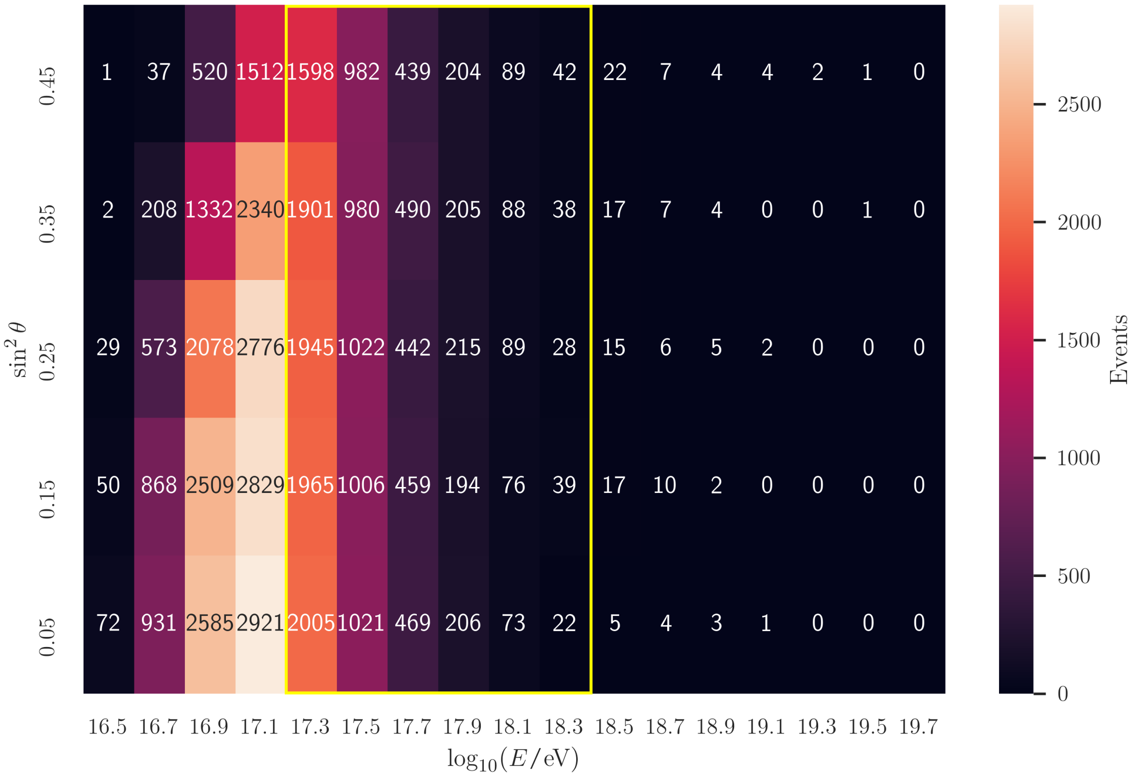
As a final remark regarding outliers, it would be desirable that faulty counters or modules were identified using monitoring data, instead of using the data itself. Nevertheless, finding these outliers can help identify problems in the functioning of the detectors that were not evident in monitoring data. The cause of these ill-behaved counters or modules is subject of further investigations in future work.

6.3.4 Upper energy cut and final event selection

Figure 6.10 shows the number of reconstructed events per bin of logarithmic reconstructed energy ($\log_{10}(E/\mathrm{eV})$) and sine square of the zenith angle ($\sin^2\theta$). The color scale represents the number of events, which is also written in each bin.

From Fig. 6.10 we can see that for $\log_{10}(E/\mathrm{eV}) \geq 18.4$ there is not enough statistics. We thus set this value as an upper energy cut. The selected events can be seen in Fig. 6.10

enclosed by a yellow rectangle. From a total of 42646 events, 18332 pass the selection criteria (43 %).



$\sin^2\theta$	16.5	16.7	16.9	17.1	17.3	17.5	17.7	17.9	18.1	18.3	18.5	18.7	18.9	19.1	19.3	19.5	19.7
0.45	1	37	520	1512	1598	982	439	204	89	42	22	7	4	4	2	1	0
0.35	2	208	1332	2340	1901	980	490	205	88	38	17	7	4	0	0	1	0
0.25	29	573	2078	2776	1945	1022	442	215	89	28	15	6	5	2	0	0	0
0.15	50	868	2509	2829	1965	1006	459	194	76	39	17	10	2	0	0	0	0
0.05	72	931	2585	2921	2005	1021	469	206	73	22	5	4	3	1	0	0	0

Figure 6.10: Number of events (color scale and labels) as a function of the logarithmic energy and sine-square zenith angle bin. A yellow rectangle encloses the selected events.

6.3.5 Distance cut: Saturation

At close distances from the shower axis, saturation becomes more likely. In this condition, the non-saturated muon densities are downward fluctuations of the true muon density distribution. In other words, when saturation is very likely, the muon density distribution is biased.

For this reason, we impose a minimum distance of the modules to the shower core, to take into account their measured muon densities. In Sec. 4.2.5 we observed that at 170 input muons the saturation probability is below 1 %. For a $10\,\mathrm{m}^2$ module, this corresponds to a muon density on the shower plane of $170/(10\,\mathrm{m}^2\cos\theta)$. If we imposed a cut on muon density, we would bias the muon density distributions. Instead, we impose a cut at the distance at which the muon density is expected to reach $170/(10\,\mathrm{m}^2\cos\theta)$. To compute the expected value, we divide the measured distances in $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bins, and we fit a simple power-law to the logarithmic mean muon densities as a function of the distance to the shower axis (measured on the shower plane)

$$\log_{10}(\rho_\mu/\mathrm{m}^{-2}) = \log_{10}\rho_{450} - \alpha\log_{10}(r/450\,\mathrm{m}), \tag{6.11}$$

where the intercept $\log_{10}\rho_{450}$ and slope α are the fitted parameters. We can solve Eq. (6.11) for r by setting $\rho_\mu = 170/(10\,\mathrm{m}^2\cos\theta)$, where $\cos\theta = \sqrt{1-\sin^2\theta}$ and $\sin^2\theta$ is the bin-center value. Because the power law function overestimates the expected muon density at short distances to the shower axis, it also overestimates the distance r at which the saturation probability is expected to be 1 %. In this way, we provide a conservative estimate of r.

Figure 6.11 shows an example muon LDF for events with reconstructed energy in $17.6 \leq \log_{10}(E/\mathrm{eV}) \leq 17.8$ and reconstructed zenith angles in $27^\circ \lesssim \theta \lesssim 33^\circ$ ($0.20 \leq \sin^2\theta \leq 0.30$). The orange squares represent the mean, and the standard deviation of the mean is contained within the marker. In gray circles we can see the candidate (bias-corrected) module densities. In this example the limit muon density on the shower plane is $\rho_\mu = 19.63\,\mathrm{m}^{-2}$ ($\log_{10}(\rho_\mu/\mathrm{m}^{-2}) = 1.29$), and yields a distance cut at $r = 95.3\,\mathrm{m}$ ($\log_{10}(r/\mathrm{m}) = 1.98$). This is illustrated in purple lines.

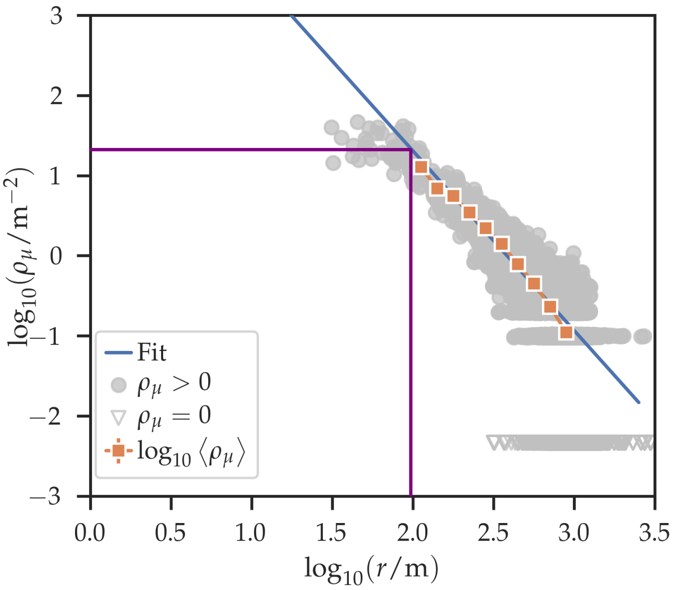


Figure 6.11: Logarithmic muon density as a function of the logarithmic distance to the shower axis. The data corresponds to reconstructed events with energies in $17.6 \leq \log_{10}(E/\mathrm{eV}) \leq 17.8$ and zenith angles in $27^\circ \lesssim \theta \lesssim 33^\circ$ ($0.20 \leq \sin^2\theta \leq 0.30$). In gray circles all candidate module densities can be seen, in orange squares their mean, which contains the standard deviation of the mean, and in a blue line the power-law fit. The purple lines illustrate how the distance cut is determined.

Figure 6.12 shows the distances that are used as lower cuts, where the saturation probability is equal or below 1 %.

6.3.6 Distance cut: Lateral trigger probability of the surface detector

The last cut we impose is a maximum distance cut, related to the fact that at large distances from the shower axis, the trigger probability decreases. Just as the trigger efficiency depends on the energy and zenith angle (c.f. Sec. 6.3.1), the trigger probability also decreases with the distance to the shower axis, where the secondary particles rate becomes smaller. The muon densities at large distances to the shower axis constitute upper fluctuations of the true muon density distribution, which cannot be unbiassedly sampled.

Following Ref. [112], we model the SD-array lateral trigger probability $\mathrm{LTP}(r)$ as

$$\mathrm{LTP}(r; E, \theta) = \begin{cases} \left(1 + \exp\left[-\frac{r - R_0(E,\theta)}{\Delta R(E,\theta)}\right]\right)^{-1}, & \text{if } r \leq R_0. \\ \left(1 + 2\exp\left[-\frac{r - R_0(E,\theta)}{2\Delta R(E,\theta)}\right]\right)^{-1}, & \text{if } r > R_0. \end{cases} \tag{6.12}$$

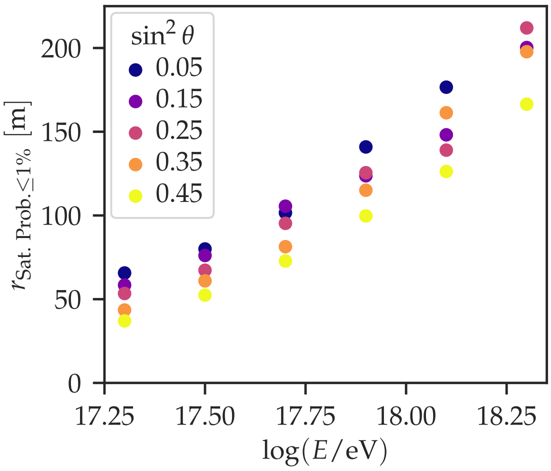


Figure 6.12: Lower cut distances at which the saturation probability is equal or below 1 %, as a function of the logarithmic energy, and for different zenith angles (different colors).

where R_0 and ΔR are parameterized as a function of the energy and zenith angle as in Ref. [112] (including all trigger types). This parameterization was done using hybrid SD-FD data. Fig. 6.13 shows the lateral trigger probability as a function of the distance, for different energies at $\sin^2\theta = 0.25$ on the left panel, and for different zenith angles at $\log_{10}(E/\text{eV}) = 17.9$ on the right panel. A gray dashed line marks 90 % lateral trigger probability.

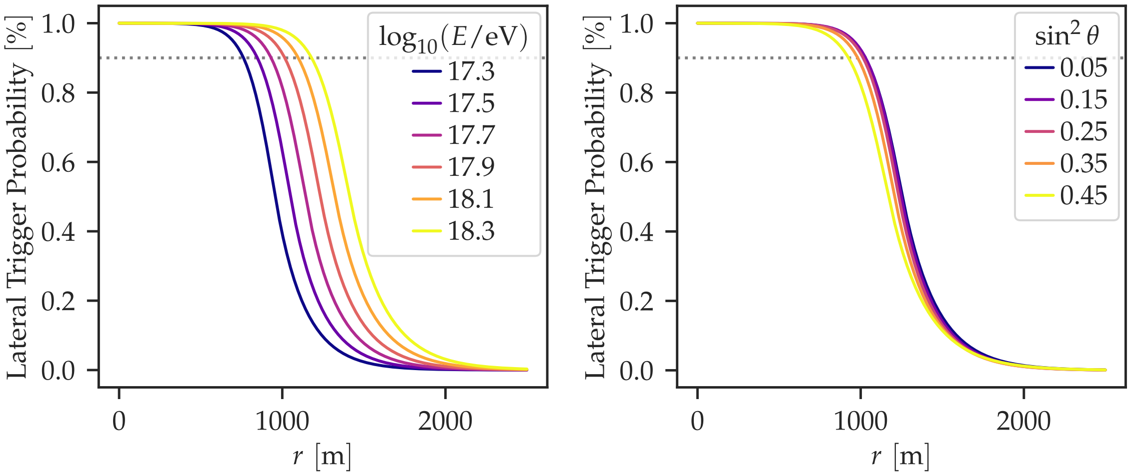


Figure 6.13: Lateral trigger probability (LTP) as a function of the distance to the shower axis, as parameterized in Ref. [112]. The left panels shows said curves for different energies at fixed $\sin^2\theta = 0.25$, while the right panel shows them for different zenith angles at $\log_{10}(E/\text{eV}) = 17.9$. Gray dashed lines mark 90 % lateral trigger probability.

Finally, Fig. 6.14 shows the distances that are used as upper cuts, where the lateral trigger probability equals 90 %.

6.4 Methods

In this section we describe the statistical method that we use for fitting the data, as well as for assessing the goodness-of-fit and the uniformity of the residuals.

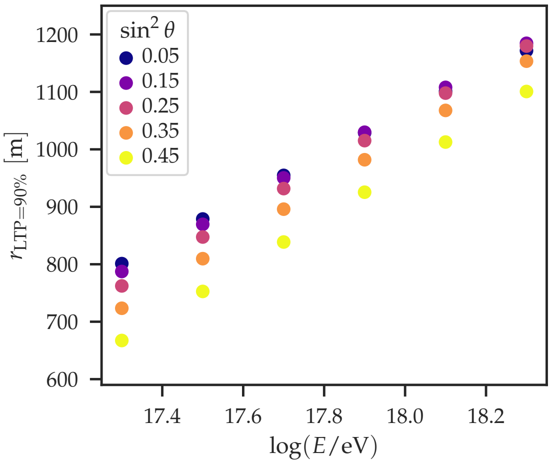


Figure 6.14: Upper cut distances at which the lateral trigger probability (LTP) equals 90 %, as a function of the logarithmic energy, and for different zenith angles (different colors).

6.4.1 Weighted least squares fit

We start by binning the data in logarithmic energy, in steps of $\Delta\log_{10}(E/\mathrm{eV}) = 0.2$. Since the muon density depends on the energy as $\sim E^{\beta}$, the muon densities in an energy bin have a contribution on the fluctuations due to the contribution of the different energies. Since there are more data at lower than at higher energies, this can also bias the mean muon densities if not accounted for. For this reason, we normalize the muon densities ρ_μ, with corresponding reconstructed energy E, to the center of the reconstructed energy bin E_{norm}, by computing

$$\rho_{\mu,\mathrm{norm}} = \rho_\mu \left(\frac{E_{\mathrm{norm}}}{E}\right)^{\beta}, \tag{6.13}$$

with $\beta = 0.9$ [5]. In any case, the normalization has a small effect on average. Although it can change individual densities in up to $\sim 20\,\%$, the average muon densities change in less than $2\,\%$.

Moreover, we bin the data also in the sine-square of the reconstructed zenith angle, and in logarithmic distance to the shower axis. Then, for every $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bin, we fit the logarithm of the mean normalized muon density as a function of the center of the logarithmic distance bin, via a weighted least-squares. The weights correspond to the standard deviation of the normalized mean. We use the different models discussed in Sec. 6.1, also varying which parameters are fixed or freed. For clarity, we provide the expression of the χ^2 [113]:

$$\chi^2 = \sum_{i=1}^{N} \frac{\left[\log_{10}\langle\rho_{\mu\,i}/\mathrm{m}^{-2}\rangle - \log_{10}\left(\rho_\mu^{\mathrm{model}}(r_i;\vec{p})/\mathrm{m}^{-2}\right)\right]^2}{\sigma^2\left[\log_{10}\langle\rho_{\mu\,i}/\mathrm{m}^{-2}\rangle\right]}, \tag{6.14}$$

where i runs over the different $\log_{10}(r/\mathrm{m})$ bins, and $\log_{10}\rho_\mu^{\mathrm{model}}(r_i;\vec{p})$ is the predicted logarithmic muon density by a given model, evaluated at the same distance bin, and with a parameter vector $\vec{p}$.

The weighted least-squares fits rely on Gaussian assumptions. In our case, most of the calculated points of the logarithm of the mean muon density are the result of an average over a large number of measurements. Thus, the central limit theorem allows us to approximate Gaussian errors in the mean muon density. Furthermore, we approximate that the errors in the logarithm of the mean muon densities are also Gaussian. Under these conditions, the χ^2 is a statistically accurate model [113].

It is relevant to add that, to achieve convergence consistently, we found necessary to follow four steps: (1) Use the data set with no outliers, (2) fit the mean of the muon densities, (3) express the models as linear in their parameters where possible, and (4) set accurate initial values. For the latter, we first fit a power-law function (c.f. Eq. (6.11)) to the data of each $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bin. The resulting values for the parameters $\log_{10}\rho_{450}$ and α were used to determine initial values of the parameters of the tested muon LDFs for each $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bin.

6.4.2 Goodness of fit: Penalized likelihood approach

One of the most customary ways to evaluate the goodness of fit of a least-squares fitted model is analyzing the χ^2 value (or its value divided by the number of degrees of freedom χ^2/N_{dof}). More specifically, the χ^2 can be tested against its expected distribution to obtain a p-value that can serve as a measure of the goodness of fit. Sometimes this step is skipped and χ^2/N_{dof} is directly used as a measure of goodness of fit, typically interpreting that the model over-fits if $\chi^2/N_{\mathrm{dof}} \ll 1$, under-fits if $\chi^2/N_{\mathrm{dof}} \gg 1$, and is appropriate if $\chi^2/N_{\mathrm{dof}} \sim 1$. However, the χ^2 does not provide a clear procedure for comparing the goodness of fit of different models. In particular, it is not clear how to use χ^2 to distinguish between the goodness-of-fit of a same model fitted with one or two free parameters.

In contrast, there are likelihood-based approaches that can be used for making such comparisons. At this point it is relevant to notice that the χ^2 (c.f. Eq. (6.14)) can be expressed as a Gaussian likelihood L [113] in the following way:

$$\chi^2 = -2\ln L + \sum_{i=1}^{N} \ln\left[\frac{1}{\sqrt{2\pi}\sigma\left[\log_{10}\langle\rho_{\mu\,i}/\mathrm{m}^{-2}\rangle\right]}\right], \tag{6.15}$$

where the sum of the last term is simply a constant value for each $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bin.

The problem we are interested in is testing how different functional forms fit to data, as well as comparing one same functional form fixing or freeing some of its parameters. The latter is an example of a nested model. The model M_1 is said to be nested in M_2, if some coordinates of the parameter vector $\vec{p}_1$ are fixed. Taking the Hillas function (c.f. Eq. (6.6)) as example, the parameter vector $\vec{p}_2$ is $(\log_{10}\rho_{450},\ \alpha,\ r_0^{-1})$, while $\vec{p}_1$ is $(\log_{10}\rho_{450},\ \alpha,\ r_{0\,\mathrm{fix}}^{-1})$, where $r_{0\,\mathrm{fix}}^{-1}$ is a fixed value. Then, the comparison of the models M_1 and M_2 can be viewed as a classical hypothesis testing problem of the null hypothesis H_0: $r_0^{-1} = r_{0\,\mathrm{fix}}^{-1}$. If M_1 is nested in M_2, then the largest likelihood achievable by M_2 will always be larger than that of M_1. Adding a penalty on models with more parameters would balance between over-fitting and under-fitting. This leads to the so called “penalized likelihood approach”.

To compare different models fitted using a penalized likelihood function, we use the Bayesian or Schwarz Information Criterion (BIC) [114] and the Akaike Information Criterion (AIC) [115]. The idea behind both AIC and BIC is that every model loses some information of the real process that it intends to represent. These information criteria quantify the relative amount of information lost by a model. The higher the quality of a model, the less information it looses, and hence the AIC and BIC values are smaller. With AIC or BIC, both model estimation and selection can be simultaneously accomplished.

The definition of the AIC is

$$\mathrm{AIC} = 2N_{\mathrm{par}} - 2\ln(\hat{L}), \tag{6.16}$$

and of the BIC is

$$\mathrm{BIC} = N_{\mathrm{par}}\ln(N_{\mathrm{obs}}) - 2\ln(\hat{L}), \tag{6.17}$$

where N_{obs} is the number of data points or observations in the data set, $\hat{L}$ is the maximized value of the likelihood function of the model M, and N_{par} is the number of parameters of the model.

Both serve as a goodness-of-fit measure as they are proportional to the log-likelihood. Since the likelihood can be increased by the addition of parameters, which might result in over-fitting, both criteria have a penalty term for the number of parameters. The main difference is that the penalty is larger in BIC than in AIC.

It is useful to notice that a constant factor C in the likelihood translates to a constant additive term $-2\ln(C)$ in the AIC/BIC; such a constant term is irrelevant in the comparison of two models.

In the case of a χ^2 fit, the AIC and BIC take the following form:

$$\text{AIC} = 2N_{\text{par}} + \hat{\chi}^2, \tag{6.18}$$

and of the BIC is

$$\text{BIC} = N_{\text{par}} \ln(N_{\text{obs}}) + \hat{\chi}^2, \tag{6.19}$$

where $\hat{\chi}^2$ is the minimized value of the χ^2, and where we neglect a constant additive term in both cases.

6.4.3 Uniformity of the residuals

Like the χ^2, the AIC and BIC are insensitive to the sign of the deviation of the data points with respect to the fitted function, as the three quantities are a function of the squared deviations. For this reason, it is necessary to test for the uniformity of the residuals separately. This is important to rule out models that consistently under-estimate the muon LDF at short distances from the core while they over-estimate it at large distances from the core, or that behave in the exact opposite way.

For this purpose we employ the Wald-Wolfowitz "runs" test [116]. This non-parametric test that can be applied to test the randomness of the signs of the residuals. Essentially, it tests the hypothesis that the signs of the residuals in a sequence are independent and identically distributed.

A "run" is a sequence of equal signs. The length of the run is the number of elements in it. For example, let us suppose that we analyze a fit where there are $N = 10$ distance bins where the residuals have the signs $+ + + - + + + - - -$. In this example there are $R = 4$ runs, with $n_+ = 6$ positive values and $n_- = 4$ negative values in the sequence. The Wald-Wolfowitz test statistic is then defined as

$$Z = \frac{R - \hat{R}}{\sigma_R}, \tag{6.20}$$

where R is the number of observed runs, $\hat{R}$ is the number of expected runs, and σ_R is the standard deviation of R. The two latter are computed as

$$\hat{R} = \frac{2n_+ n_-}{n_+ + n_-} + 1, \tag{6.21}$$

$$\sigma_R^2 = \frac{(\hat{R}-1)(\hat{R}-2)}{n_+ + n_- - 1}. \tag{6.22}$$

The distribution of Z is tabulated for small number of runs, while if $n_+ > 10$ and $n_- > 10$ it can be approximated to a standard normal distribution. In particular we compute the probability to observe $|Z| > |Z_{\text{obs}}|$. If the probability is close to zero, this supports rejecting the null-hypothesis that the signs of the residuals are a sample of a random sequence. In other words, if this probability is small, the model does not reproduce data well.

To combine the p-values of the Wald-Wolfowitz test of each $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bin, we use the Fisher combined probability test or Fisher method [117]. In this method, a χ^2 test statistic is built from the p-values p_i of each bin:

$$\chi^2 = -2\sum_{i=1}^{k} \ln(p_i), \tag{6.23}$$

where k is the number of p-values to combine (in this case it is the number of bins, 30). The resulting test statistic is distributed as a χ^2 with $2k$ degrees of freedom. Knowing this, we can compute a unique p-value from the probability to observe $|\chi^2| > |\chi^2_{\mathrm{obs}}|$.

6.5 Results

In this section we present the fits of the different models to data, and compare their goodness of fit.

6.5.1 Fits to data

In the following we show how the different models fit the events with $17.6 \leq \log_{10}(E/\mathrm{eV}) \leq 17.8$ and zenith angles in $27^\circ \lesssim \theta \lesssim 33^\circ$ ($0.20 \leq \sin^2\theta \leq 0.30$) as an example. In all applicable cases we set $r_{\mathrm{ref}} = 450\,\mathrm{m}$, as this is the distance at which fluctuations in the muon density are minimized for the SD-750 m array [14]. Figure 6.15 shows how the modified NKG or the NKG fit the aforementioned data. The positive muon densities are shown in gray circles, the null muon densities are shown in triangles at an arbitrary logarithmic density value, and the logarithm of the mean muon densities per $\log_{10}(r/\mathrm{m})$-bin are shown in orange squares. In all displayed fits we use $r_0 = 320\,\mathrm{m}$, as we do not see an improvement in the fit from fixing r_0 to other values. The top left panel is a modified NKG (c.f. Eq. (6.4)) with $\alpha = 0.75$ (as initially proposed by Greisen [99]), and $\gamma = 3.00$ (very similar to the parameters used in Ref. [14]). For comparison, the top right panel is an NKG with $\alpha = 0.75$ (equivalently, a modified NKG with the same α and $\gamma = 0$). The bottom left panel uses the values of the standard $\overline{\text{Off}}\underline{\text{line}}$ event-wise muon LDF reconstruction, $\alpha = 1.0$ and $\gamma = 1.85$. Once again, for comparison, the bottom right panel uses an NKG with $\alpha = 1.0$. A priori, all four fitted functions seem to model the muon LDF reasonably well. A detailed comparison of the goodness of fit is described in Sec. 6.5.2.

It can be added that we also attempted freeing different parameters of the modified NKG, or fixing them to other values, but we observed no evident improvement in the fits. In particular, when α and β are both let free, the absolute value of their correlation coefficient is above 0.96 in most fits, and the values to which they converge vary widely and depend on their assigned initial values. Furthermore, when let free, the fitted values of γ vary greatly too (from ~ 0 to ~ 7). This is because there is no (unbiased) statistics very close and very far from the shower core, which would offer a leverage for fitting α and γ respectively. For these reasons, it is not feasible to fit α and γ.

Figure 6.16 shows the same as Fig. 6.15, but fitting the data with log-log polynomials (c.f. Eq.(6.8)) of second (left panel) and third (right panel) degrees. They both exhibit an unphysical behavior: the second-order log-log polynomial predicts that the muon density decreases close to the core, while the third-order log-log polynomial presents an unjustifiable change in curvature. Four and five degree polynomials exhibit the same problems.

Finally, Fig. 6.17 displays the same data as Figs. 6.15 and 6.16, but fitted with a Hillas function (c.f. Eq. (6.6)). On the left panel we set all three parameters free ($\log_{10}\rho_{450}$, α, and

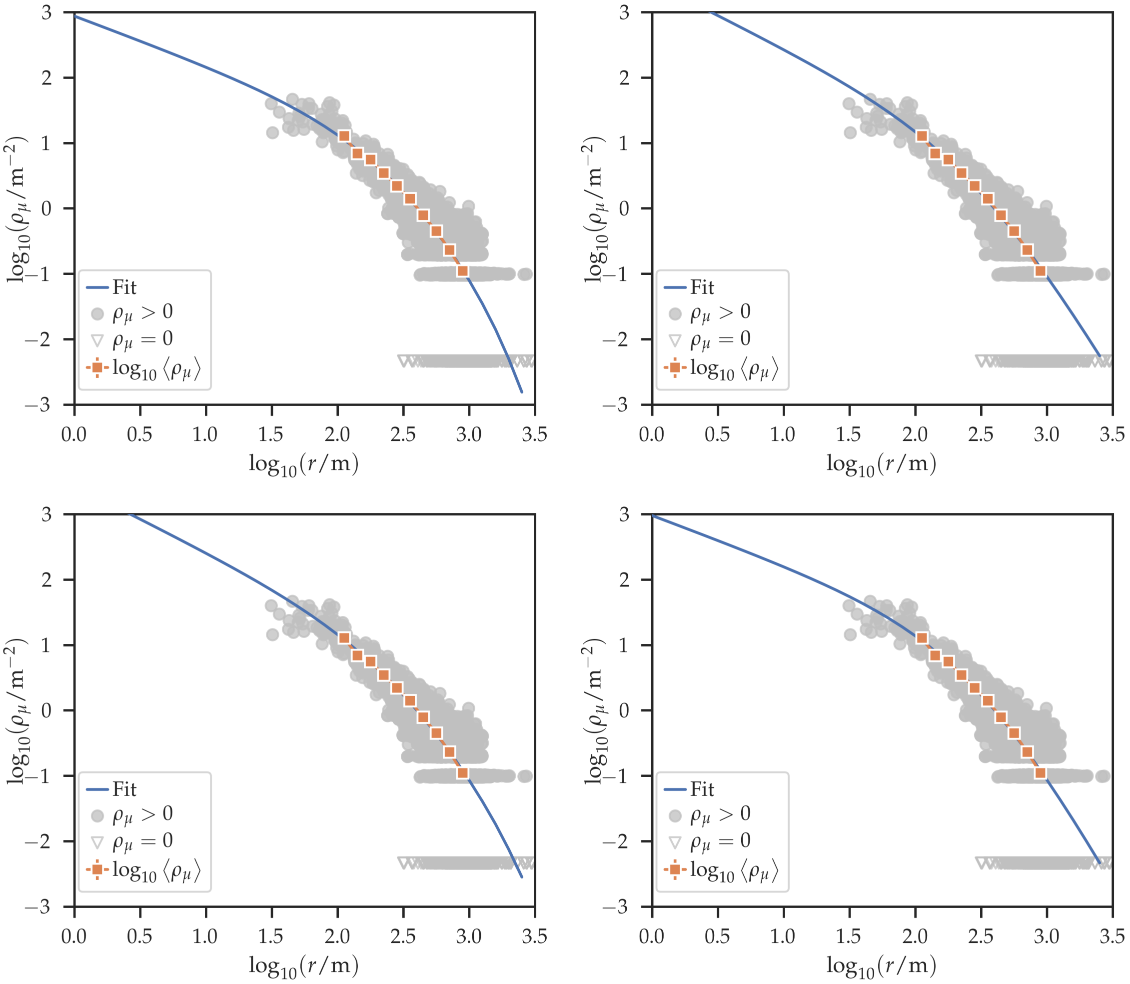


Figure 6.15: Logarithm of the muon density as a function of the logarithm of the distance to the shower axis on the shower plane. Candidate counters are represented with gray circles, saturated counters with gray upside triangles, null signal counters with gray downside triangles (at an arbitrary value of $\log_{10}(\rho_\mu/\mathrm{m}^{-2})$, the mean is represented with orange squares, and the fit with a blue line. The data points have energies in $17.6 \leq \log_{10}(E/\mathrm{eV}) \leq 17.8$, and zenith angles in $27° \lesssim \theta \lesssim 33°$ ($0.20 \leq \sin^2\theta \leq 0.30$). The fitted functions are modified NKGs (c.f. Eq. (6.4)) with $r_0 = 320\,\mathrm{m}$, and the following fixed parameters: (Top left) $\alpha = 0.75$ and $\gamma = 3.00$, (top right) $\alpha = 0.75$ and $\gamma = 0$, (bottom left) $\alpha = 1.0$ and $\gamma = 1.85$, and (bottom right) $\alpha = 1.0$ and $\gamma = 0$.

r_0^{-1}), while on the right panel we fix $r_0 = 750\,\mathrm{m}$. A priori, this exponential function fits reasonably well the data, both with r_0 fixed or free.

6.5.2 Comparison of the goodness-of-fit

In order to assess which is the best model of the muon LDF, for all the fits presented in Sec. 6.5.1 we compare the AIC, and BIC (see Sec. 6.4.2), and the p-value of the Wald-Wolfowitz runs test (c.f. Sec. 6.4.3). Since there are actually as many muon LDF fits as $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bins (30), we report on the sum of the individual values of AIC and BIC, and the p-value of the Fisher test that combines those of the individual Wald-Wolfowitz tests.

Figure 6.18 shows a comparison of the aforementioned quantities, for the models shown in Sec. 6.5.1. We include two additional models for comparison: the power law, and a modi-

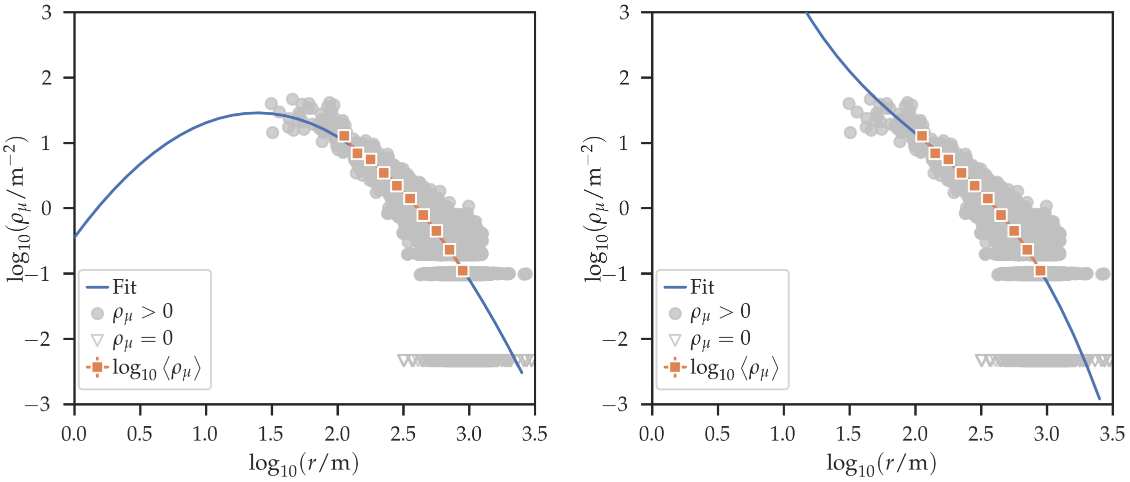


Figure 6.16: Same as Fig. 6.15, but the fitted functions are log-log polynomials (c.f. Eq.(6.8)) of second degree (left panel) or third degree (right panel).

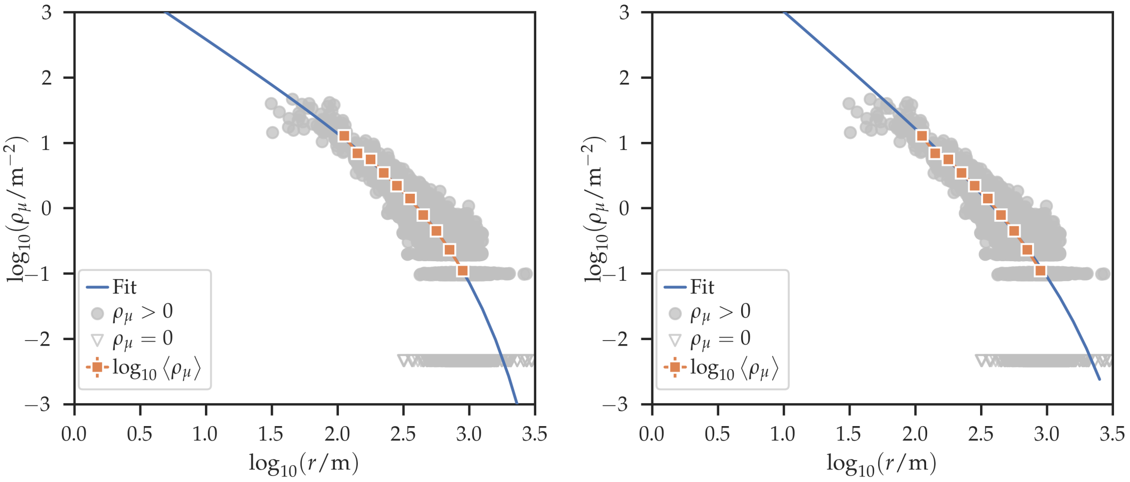


Figure 6.17: Same as Figs. 6.15 and 6.16, but the fitted functions are Hillas functions (c.f. Eq. (6.6)) with $\log_{10}\rho_{450}$, α, and r_0^{-1} as free parameters in the left panel, while in the right panel we fix $r_0 = 750\,\mathrm{m}$.

fied NKG with α, β, and γ as free parameters. It is easy to see that the power law does not model appropriately the muon LDF, as the signs of the residuals are highly not random (the p-value of the Wald-Wolfowitz test is close to 0), and also the AIC and BIC are comparably very high. Although the next two models, the log-log polynomials of degree 2 and 3, have good goodness-of-fit indicators, we know from Sec. 6.5.1 that they yield unphysical results. We then consider only the different variations of the Hillas and modified NKG (or NKG). Except for the Hillas function with fixed r_0, which has comparably bad goodness-of-fit indicators, the performance of these models is not very different from one another. The NKG, modified NKG, and Hillas (with free r_0) functions evidently constitute reasonable models of the muon LDF. It is interesting to notice that the modified NKG with α, β, and γ as free parameters does not have a better AIC and BIC than the modified NKG with fixed parameters $\alpha = 0.75$ and $\gamma = 3.0$. It has a p-value of the Fisher test of almost 1 (like the log-log polynomials), suggesting that it can be over-fitting data. In fact, the fitted modified NKGs with all its parameters free have a slope close to the core that changes significantly from bin to bin with no clear energy or zenith angle trend. Moreover, we observe that the p-value of the Fisher test is minimal for the NKG with $\alpha = 1$, and the AIC and BIC take larger values than for the other models and/or combinations of fixed parameters. Without considering the unphysical log-log polynomial models, the best AIC and BIC is achieved with a modified NKG of fixed parameters $\alpha = 0.75$ and $\gamma = 3.0$. In this case, the p-value of the Fisher test is reasonably good, not being too high or too low.

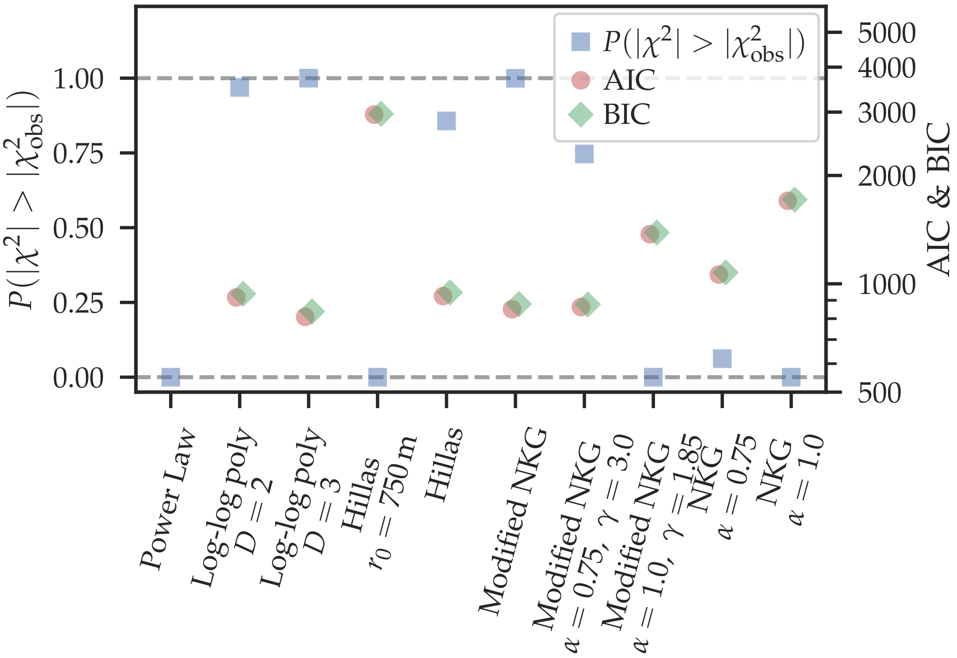


Figure 6.18: P-value of the Fisher test (squares, left scale), and AIC and BIC (circles and diamonds respectively, right scale), for the different tested models of the muon LDF (see text for details). The gray dashed lines mark p-values of the Fisher test of 0 and 1.

It is relevant to add that we scanned over the values to fix for α and γ. We allowed α to take the values $\{0.0, 0.5, 0.75, 1.0, 1.5\}$ while we allowed γ to take the values $\{0.00, 1.00, 1.85, 2.00, 3.00, 3.50\}$. We observed that, of all possible combinations, the one that minimized the AIC and BIC was $\alpha = 0.75$ and $\gamma = 3.0$.

We conclude that the Hillas, NKG, or modified NKG are all good models of the muon LDF. The model that best fits our data (in terms of AIC and BIC) is a modified NKG with fixed parameters $\alpha = 0.75$ and $\gamma = 3.0$. The modified NKG with all its parameters free has a similar AIC and BIC but tends to over-fit the data. Given this result, and the fact that the modified NKG is a physically motivated model, we adhere to this model for the rest of the analysis.

It is relevant to add that, in the future, the muon densities from the integrator mode of the UMD will complement the measurements of the counter mode reaching distances much closer to the core. With these muon densities, the evaluation of fitting or fixing the parameter α should be revisited.

6.6 Useful parameterizations of the muon LDF

In this section we provide two parameterizations of the muon LDF based on the fits that use the modified NKG with $\alpha = 0.75$ and $\gamma = 3.0$. Both of them provide a data-like muon density as a function of the distance to the shower axis, primary energy, and primary zenith angle. These parameterizations can be useful for making realistic toy models that make use of the muon LDF, for example, as used in Sec. 6.7.1.

Both parameterizations consist of expressing the two free parameters of the modified NKG, $\log_{10}\rho_{450}$ and β, as a biquadratic functions of two variables, which in the first case are $\log_{10}(E/\mathrm{eV})$ and $\sin^2\theta$:

$$\log_{10}\rho_{450} = a_\rho + b_\rho(\log_{10}(E/\mathrm{eV}) - 17.3) + c_\rho(\log_{10}(E/\mathrm{eV}) - 17.3)^2, \tag{6.24}$$

$$\beta = a_\beta + b_\beta(\log_{10}(E/\mathrm{eV}) - 17.3) + c_\beta(\log_{10}(E/\mathrm{eV}) - 17.3)^2, \tag{6.25}$$

where the parameters $a_\rho, b_\rho, c_\rho, a_\beta, b_\beta$, and c_β (generically x_ρ and x_β) are themselves quadratic functions of $\sin^2\theta$:

$$x_\rho = x_{\rho,0} + x_{\rho,1}(\sin^2\theta/0.5) + x_{\rho,2}(\sin^2\theta/0.5)^2, \tag{6.26}$$

$$x_\beta = x_{\beta,0} + x_{\beta,1}(\sin^2\theta/0.5) + x_{\beta,2}(\sin^2\theta/0.5)^2. \tag{6.27}$$

In fact, as we see later in this section, the quadratic term in $\log_{10}(E/\mathrm{eV})$ is not necessary for the case of $\log_{10}\rho_{450}$, in which case we fix $c_{\rho,i} = 0$ (for $i = \{0, 1, 2\}$).

The second parameterization is almost identical. The difference is that β is modelled as a biquadratic function of $\log_{10}\rho_{450}$ and $\sin^2\theta$:

$$\beta = \tilde{a}_\beta + \tilde{b}_\beta(\log_{10}\rho_{450}) + \tilde{c}_\beta(\log_{10}\rho_{450})^2, \tag{6.28}$$

where its parameters $\tilde{a}_\beta$, $\tilde{b}_\beta$, and $\tilde{c}_\beta$ (generically $\tilde{x}_\beta$) follow the function

$$\tilde{x}_\beta = \tilde{x}_{\beta,0} + \tilde{x}_{\beta,1}(\sin^2\theta/0.5) + \tilde{x}_{\beta,2}(\sin^2\theta/0.5)^2. \tag{6.29}$$

The parameterization of β can also be used for fixing its value in event-wise fits of the muon LDF where the sampling of distances to the shower core is not very good. Typically, in such cases only the shower size $\log_{10}\rho_{450}$ is fitted. In that scenario it is recommendable to use the second parameterization, i.e., $\beta(\log_{10}\rho_{450}, \sin^2\theta)$. This is because the size of the shower $\log_{10}\rho_{450}$ depends not only on the energy and zenith angle of the primary, but also on the primary mass. A set of showers of a same $\log_{10}\rho_{450}$ are expected to have more similar muon LDFs than a set of showers of a same energy, which can vastly differ due to shower-to-shower fluctuations.

Figure 6.19 shows the parameterization of $\log_{10}\rho_{450}$ as a biquadratic function of the logarithmic energy and the sine-square of the zenith angle. The points represent the values obtained from each fit of the muon LDF at each $(\log_{10}(E/\mathrm{eV}), \sin^2\theta)$-bin with the modified NKG function, while the surface and lines represent the fit of $\log_{10}\rho_{450}$ with the biquadratic function (c.f. Eqs. (6.24) and (6.26)). The upper left panel is a 3-dimensional plot, while the upper right panel is a 2-dimensional projection. The bottom left panel shows $\log_{10}\rho_{450}$ as a function of $\log_{10}(E/\mathrm{eV})$, with the different $\sin^2\theta$-bins in different colors, while the bottom right panel shows $\log_{10}\rho_{450}$ as a function of $\sin^2\theta$, with the different $\log_{10}(E/\mathrm{eV})$-bins in different colors. A linear function sufficiently describes the dependence of $\log_{10}\rho_{450}$ with $\log_{10}(E/\mathrm{eV})$. It is easy to see that $\log_{10}\rho_{450}$ monotonously increases with $\log_{10}(E/\mathrm{eV})$, and monotonously decreases with $\sin^2\theta$, as expected. The former we expect from the simple argument that a higher-energy air shower produces more particles, in particular muons. The latter we expect because of attenuation: although the atmosphere does not play a significant role in attenuating muons, the energy E_μ that the muons need to go through the soil on top of the UMD depends on the zenith angle of the shower ($E_\mu > 1\,\mathrm{GeV}/\cos\theta$) [97]. Furthermore, the biquadratic fit is clearly a good model.

Figure 6.20 shows the parameterization of β as a biquadratic function of $\log_{10}(E/\mathrm{eV})$ and $\sin^2\theta$ (first parameterization, c.f. Eqs. (6.25) and (6.27)). The figure is equivalent to Fig. 6.19

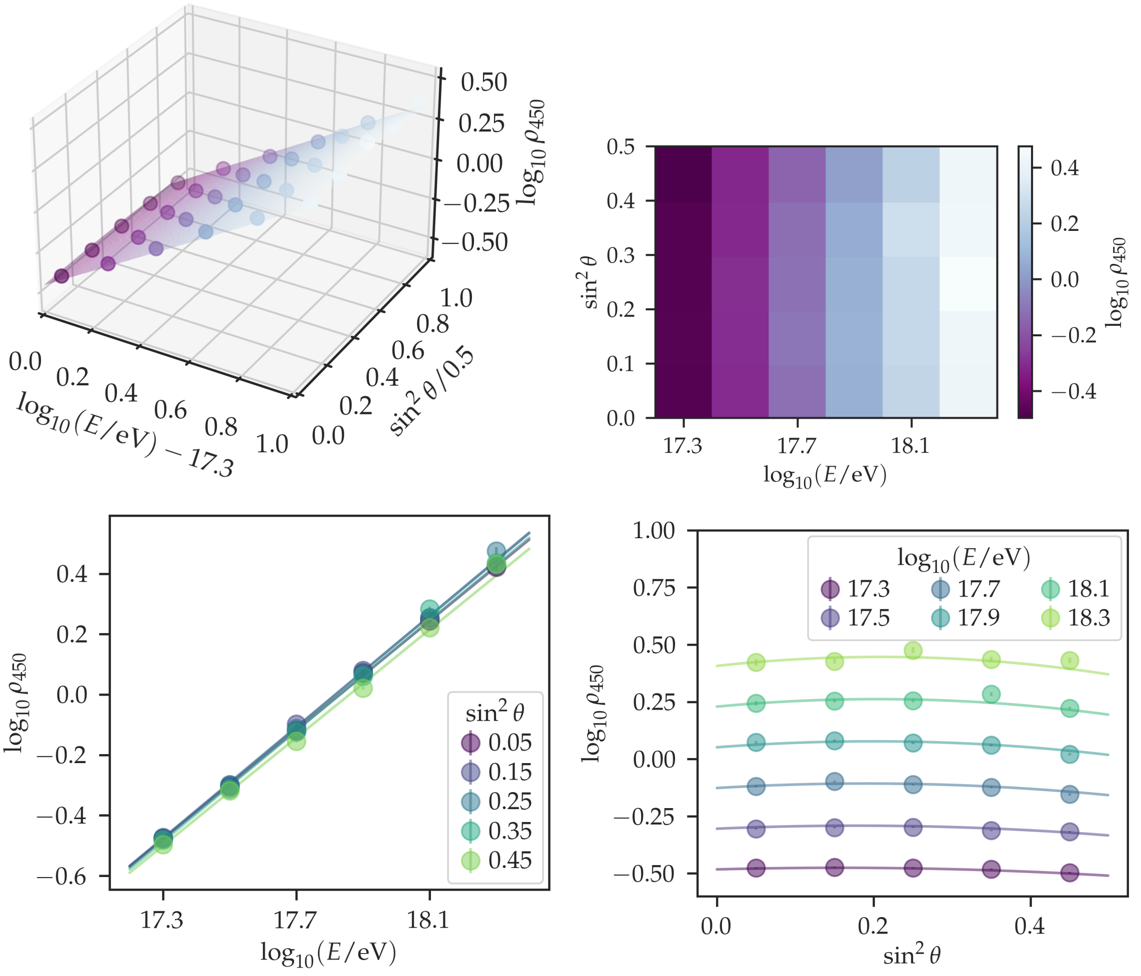


Figure 6.19: Parameter $\log_{10}\rho_{450}$ of the modified-NKG fits to data (see text for details), as a function of the $\log_{10}(E/\mathrm{eV})$ and $\sin^2\theta$ bins. The surface and lines are the result of the biquadratic fit (c.f. Eqs. (6.24) and (6.26)). Top left: 3-dimensional plot. Top right: 2-dimensional projection. Bottom left: $\log_{10}\rho_{450}$ as a function of $\log_{10}(E/\mathrm{eV})$, with the different $\sin^2\theta$-bins in different colors. Bottom right: $\log_{10}\rho_{450}$ as a function of $\sin^2\theta$, with the different $\log_{10}(E/\mathrm{eV})$-bins in different colors.

but plotting the parameter β. We can observe that β clearly decreases with the zenith angle. This means that the muon LDF is flatter the larger the zenith angle. This is because the maximum muon production depth ($X_{\mu\,\mathrm{max}}$) recedes for more inclined showers [118]. It is also clear that β increases with the energy, which can be understood from the deeper $X_{\mu\,\mathrm{max}}$.

Figure 6.21 shows the parameterization of β as a biquadratic function of the parameter $\log_{10}\rho_{450}$ and the $\sin^2\theta$-bin (c.f. Eqs. (6.28) and (6.29)). There is a tendency of β to increase with $\log_{10}\rho_{450}$, implying that air showers with larger size tend to have also steeper muon LDFs, as expected from the associated deeper $X_{\mu\,\mathrm{max}}$.

Finally, the values of the parameters of the biquadratic fits are presented in Table 6.1.

6.7 Analysis of the systematic uncertainties of the fits

A controlled experiment is necessary to quantify the systematic uncertainties that arise from the fitting procedure. We therefore develop a toy model to generate mock data from a user-

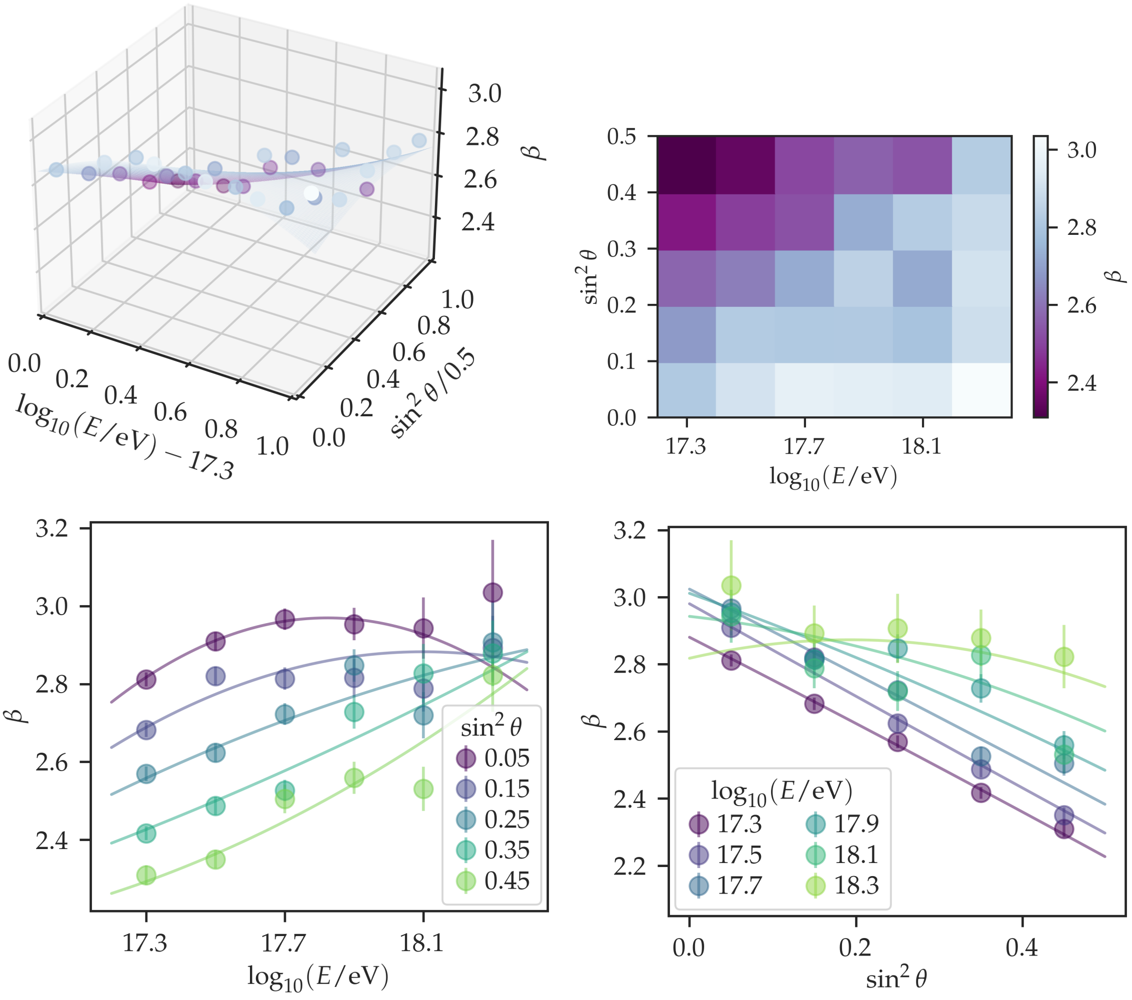


Figure 6.20: Same as Fig. 6.19 but for the parameter β. The parameterization is that of Eqs. (6.25) and (6.27).

defined muon LDF. These mock data is then fitted in the same way as real data. Comparing the generating and fitted muon LDFs we can understand the systematic uncertainties introduced by fixing parameters, while simultaneously accounting for all the characteristics of the fitting procedure that could also introduce systematic uncertainties (for instance, applying distance cuts, binning in $\log_{10}(r/\mathrm{m})$, and fitting the mean instead of all the data points).

6.7.1 Toy model

The toy model allows us to obtain a mock sampled muon LDF from a user-defined, “true” muon LDF. It consists, in first place, of an infill-like array of $10\,\mathrm{m}^2$ muon detector modules. A random position of the core is assigned, following a uniform distribution in the area between three neighboring detectors.

From the user-defined muon LDF and zenith angle, the true muon density is computed at every detector. Array trigger effects are not necessary to simulate since the whole analysis is performed above an energy that ensures full-efficiency. In contrast, the station trigger effect has to be simulated. Therefore, a detector is kept as part of the event randomly, following the lateral trigger probability described in Sec. 6.3.6.

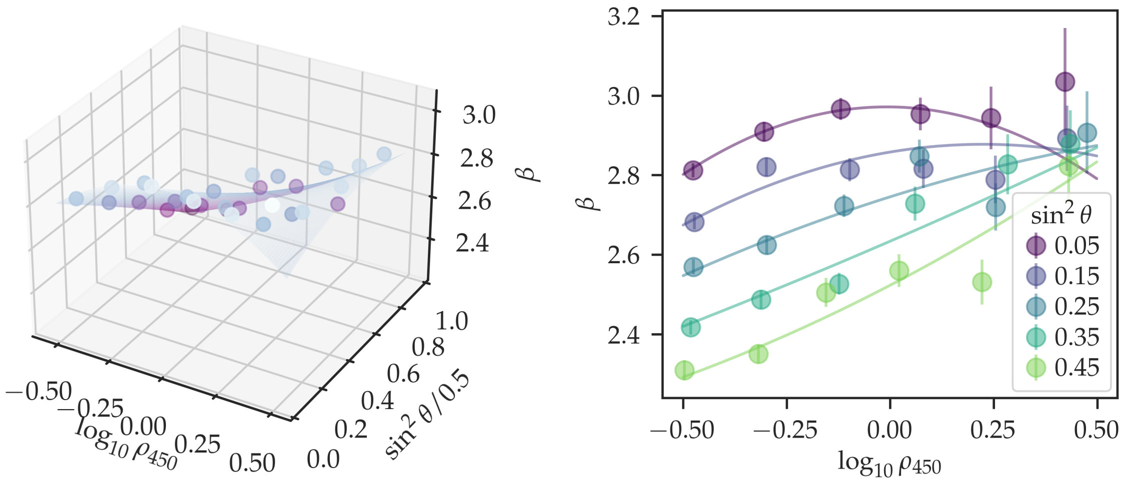


Figure 6.21: Parameter β of the modified-NKG fits to data (see text for details), as a function of the fitted parameter $\log_{10}\rho_{450}$ and the $\sin^2\theta$-bin. The biquadratic fit of β is that of Eqs. (6.28) and (6.29). The left panel is a 3-dimensional plot, while the right panel shows β as a function of $\log_{10}\rho_{450}$ for the different $\sin^2\theta$-bins in different colors.

The probability that a muon detector saturates is not simulated, but we apply the same minimum distance cuts as for data (c.f. Sec. 6.3.5) which ensure a probability of saturation below 1 %.

The true number of muons in each detector is sampled from a Poisson distribution with its parameter equal to the true number of muons, obtaining in this way the number of impinging muons. Then the mock muon density is computed by dividing the impinging number of muons by the effective detector area. To simulate a mis-reconstructed position of the core, the true position of the core is randomly fluctuated following an isotropic bivariate Gaussian distribution with a mean equal to the true position of the core and a width equal to the infill core resolution σ_{core}. The latter is parameterized as:

$$\sigma_{\text{core}}(E,\theta)/\text{m} = a(\theta) + b(\theta)(\log_{10}(E/\text{eV}) - 16.5) + c(\theta)(\log_{10}(E/\text{eV}) - 16.5)^2, \qquad (6.30)$$

where $a(\theta) = 93.53 + 94.32\sin^4\theta$, $b(\theta) = -92.52 - 13.87\sin^2\theta$, and $c(\theta) = 31.63 + 2.88\sin^2\theta$. This parameterization was obtained using QGSJetII-04 proton showers with energies in $16.5 \le \log_{10}(E/\text{eV}) \le 18.0$ and zenith angles $0° \le \theta \le 60°$ [119]. Figure 6.22 shows said parameterization of the core resolution as a function of the energy and for different zenith angles. Although we extrapolate the parameterization to the energies of this work, we do not expect that a small change in the parameterization of the dependence of the core resolution to greatly affect the toy model. What is important is to include this effect to replicate the dispersion of data, specially close to the core.

The left panel of Fig. 6.23 shows the muon density as a function of the distance to the shower axis on the shower plane for real data at energies in $17.6 \le \log_{10}(E/\text{eV}) \le 17.8$ and zenith angles in $0.2 \le \sin^2\theta \le 0.3$. The muon LDF fitted to the data in Fig. 6.23 is the one used to generate the mock data of the right panel of Fig. 6.23. This muon LDF is a modified NKG with fixed parameters $r_0 = 320\,\text{m}$, $\alpha = 0.75$, and $\gamma = 3.0$, (as mentioned in Sec. 6.5.2), and fitted parameters $\log_{10}\rho_{450} = -0.64$ and $\beta = 2.72$. We generate 442 mock events, which is the number of events in real data. The dashed vertical lines are the distance cuts that we set to data (see Secs. 6.3.5 and 6.3.6). It can be seen that the mock data reproduces very well most features of real data. At low muon densities there is a slight difference in the discrete values that the muon densities take for the mock data. In the latter case we assume a constant

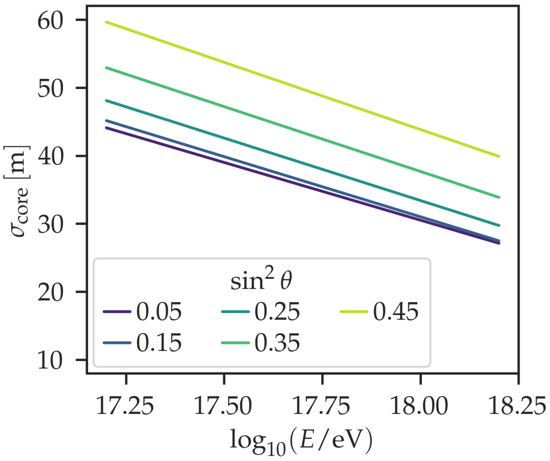


Figure 6.22: Resolution in the reconstructed position of the shower core as a function of the logarithmic energy, for different zenith angles in different colors. The parameterization, described in Ref. [119], uses QGSJetII-04 proton air-shower simulations with energies in $16.5 \leq \log_{10}(E/\text{eV}) \leq 18.0$ and zenith angles in $0° \leq \theta \leq 60°$.

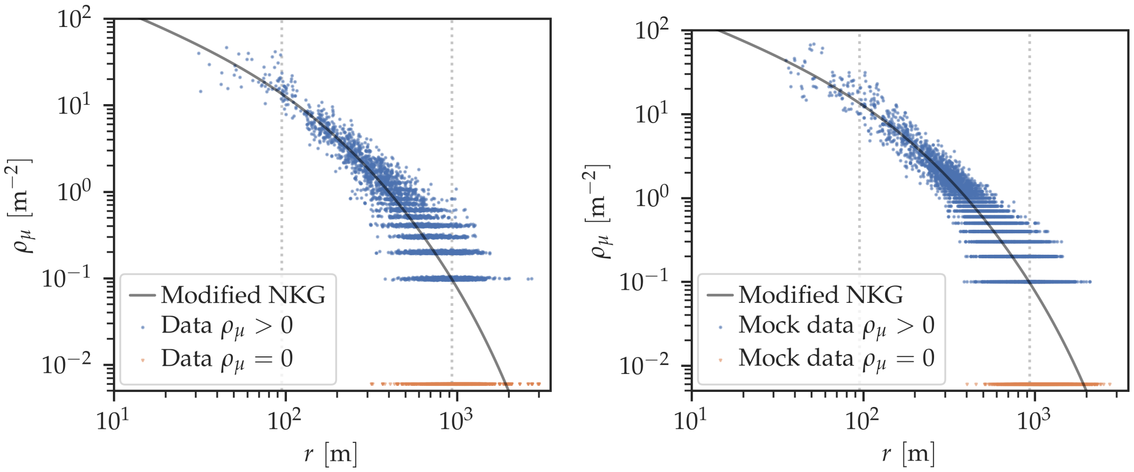


Figure 6.23: Muon density as a function of the distance to the shower axis for real (left panel) and mock (right panel) data. The muon LDF that is fitted to real data is the one used to generate the mock data, which is a modified NKG with parameters $\log_{10}\rho_{450} = -0.12$ and $\beta = 2.70$ (and the fixed parameters described in Sec. 6.5.2). The real data are events with reconstructed energies in $17.6 \leq \log_{10}(E/\text{eV}) \leq 17.8$ and zenith angles in $0.2 \leq \sin^2\theta \leq 0.3$. The dashed vertical lines are the minimum and maximum distance cuts that are used in data (see Secs. 6.3.5 and 6.3.6).

module area of $10\,\text{m}^2$; the muon density for real detectors is computed using the effective area, which can vary depending on the amount of working scintillation strips of each module. Furthermore, the bias correction makes the reconstructed muon densities of real data take continuous values. We already assessed the remaining biases in the reconstructed muons after the bias correction in Sec. 5.3. Hence, and for the purpose of this analysis, we deem our toy model sufficient.

6.7.2 Analysis of the bias from fixing the parameter α of the modified Nishimura-Kamata-Greisen function

As mentioned in Sec. 6.5.1, the parameters α and β from the modified NKG are largely correlated (with a correlation coefficient ≥ 0.96). Additionally, the saturation of the binary channel prevents us from having more information about the muon density close to the core which would constrain α better. Hence, if both α and β are let free, the values of the fitted parameters depend strongly on the initial values.We therefore decide to fix α in the reconstruction to a value of $\alpha_{\text{rec}} = 0.75$, as we show in Sec. 6.5.2 that it minimizes the overall AIC and BIC (in combination with fixing $\gamma = 3.0$). Nevertheless, fixing $\alpha_{\text{rec}} = 0.75$ does not restrict greatly the shape of the muon LDF if β is left as a free parameter of the fit, as shown in Sec. 6.2. In this section we study the bias that arises from fixing $\alpha_{\text{rec}} = 0.75$, given that the true α has a certain value α_{true}.

At the same time, we are studying the biases introduced from the chosen fitting method. In particular, we can estimate the bias that arises from fitting the logarithm of the mean muon density as a function of the logarithmic distance, with the chosen logarithmic distance binning ($\Delta \log_{10}(r/\text{m}) = 0.1$), and with the imposed distance cuts at short distances (to exclude bins where the probability of saturation is $> 1\,\%$) and at long distances (where the lateral trigger probability falls below $90\,\%$).

We generate mock data as follows: For each value of α_{true} in $\{0.6, 0.675, 0.75, 0.825, 0.9\}$, we generate $N_{\text{events}} = 442$ mock events from a modified NKG with parameters that correspond to realistic values and statistics for $17.6 \leq \log_{10}(E/\text{eV}) \leq 17.8$ and zenith angles in $0.2 \leq \sin^2\theta \leq 0.3$, of values $\log_{10}(\rho_{450}/\text{m}^2) = -0.12$, $\alpha = \alpha_{\text{true}}$, $\beta = 2.70$, $\gamma = 3.0$. These mock data are treated exactly as real data. The logarithm of the mean muon density as a function of the logarithmic distance is fitted letting $\log_{10}(\rho_{450}/\text{m}^2)$ and β free, and fixing $\alpha_{\text{rec}} = 0.75$ and $\gamma_{\text{rec}} = 3.0$. This constitutes the fit result of one experiment. We repeat 20 experiments for each value of α_{true}, and compute the mean fitted muon LDF and the standard deviation. In this way we can get an idea of the expected bias (and its variability) from fixing α in the fits to real data.

The left panel of Fig. 6.24 shows the relative error of the fitted muon LDF as a function of the logarithmic distance to the shower axis. Different colors represent muon LDFs generated from different values of α_{true}. The thick lines represent the average over the fitted muon LDFs of all experiments, and the shaded areas represent the standard deviation. The latter give an idea of how the bias could fluctuate in different realizations of the experiment. A gray dashed line locates $r = 450\,\text{m}$. A gray square encloses the inset that is plotted on the right panel. In the latter, we find the same variable plotted as a function of the distance to the shower axis.

From the left panel of Fig. 6.24 we can see that a bias of $+20\,\%$ ($-20\,\%$) in the fixed value of α translates into a bias in the muon LDF of at most $(-3.2 \pm 0.4)\,\%$ ($(+6.0 \pm 0.4)\,\%$) within the distance cuts. The uncertainties in the reported biases are computed from the plotted standard deviation, divided by the square root of the number of simulated experiments. The largest biases are attained at the smallest and largest distances from the shower axis. For the case $\alpha_{\text{true}} = \alpha_{\text{rec}} = 0.75$, the bias comes from the fitting method, and it is considerably small, of at most $(1.0 \pm 0.4)\,\%$, overlapping zero. From the right panel of Fig. 6.24 we can see that the bias in the muon LDF at $r = 450\,\text{m}$ is of at most $(+1.4 \pm 0.2)\,\%$ or (-0.9 ± 0.2), implying that $r = 450\,\text{m}$ is still an optimal distance considering systematics in the fit of the average muon LDF.

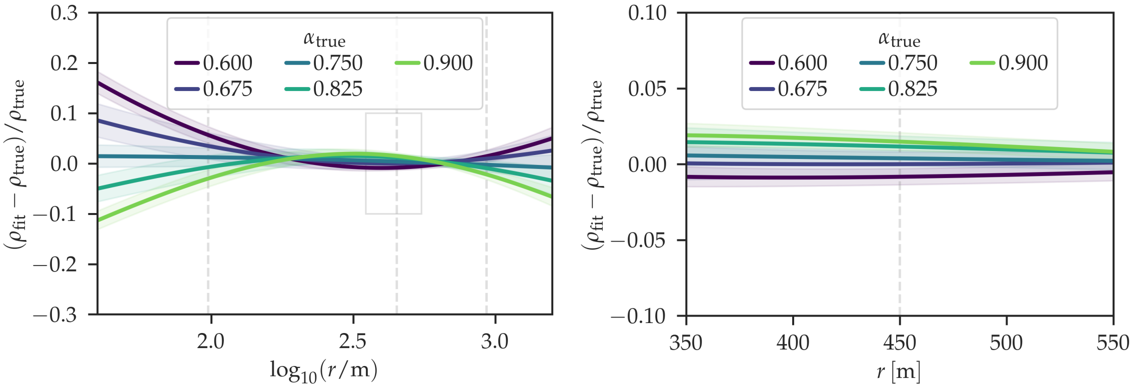


Figure 6.24: Left panel: Relative error of the fitted muon LDF as a function of the logarithmic distance to the shower axis. Right panel: An inset of the figure on the left panel around $r = 450\,\mathrm{m}$ in linear scale. Different colors represent different values of the parameter $\alpha = \alpha_{\mathrm{true}}$ of the generating muon LDF. In all cases, the muon LDF was fitted fixing $\alpha_{\mathrm{rec}} = 0.75$. The gray dashed lines locate $r = 450\,\mathrm{m}$, as well as the upper and lower distance cuts, and a gray rectangle on the left panel shows the inset that is plotted on the right panel.

6.7.3 Analysis of the bias from fixing the parameter γ of the modified Nishimura-Kamata-Greisen function

In Sec. 6.5.1 we observed that there is not enough sensitivity to fit γ, since the lateral trigger probability drops before or at the distances where this parameter becomes relevant (see Sec. 6.2). This prevents us from having (unbiased) statistics at large distances from the shower axis, making it unfeasible to fit γ. As a consequence, we have to fix γ in the reconstruction. We choose to fix it to $\gamma_{\mathrm{rec}} = 3.0$, which minimizes the overall AIC and BIC (in combination with fixing $\alpha = 0.75$), as shown in Sec. 6.5.2. As for α (c.f. Sec. 6.7.2), we study the bias that arises from fixing $\gamma_{\mathrm{rec}} = 3.0$, considering that the true γ has a value γ_{true}.

For this parameter, the distance at which the lateral trigger probability drops below 90 % ($d_{\mathrm{LTP}=90\,\%}$) plays an important role. Because this distance varies greatly within our dataset, we test how the bias in the muon LDF that arises from fixing γ depends on this distance.

In this case, we let γ_{true} take the values in $\{2.4, 2.7, 3.0, 3.3, 3.6\}$. For each γ_{true}, we test three realistic scenarios varying $d_{\mathrm{LTP}=90\,\%}$:

1. $d_{\mathrm{LTP}=90\,\%} = 667\,\mathrm{m}$, $N_{\mathrm{events}} = 1598$, $\log_{10}(\rho_{450}/\mathrm{m}^2) = -0.50$, $\beta = 2.34$ (this corresponds to realistic values and statistics for $17.2 \leq \log_{10}(E/\mathrm{eV}) \leq 17.4$ and zenith angles in $0.4 \leq \sin^2\theta \leq 0.5$),
2. $d_{\mathrm{LTP}=90\,\%} = 932\,\mathrm{m}$, $N_{\mathrm{events}} = 442$, $\log_{10}(\rho_{450}/\mathrm{m}^2) = -0.12$, $\beta = 2.70$ (this corresponds to realistic values and statistics for $17.6 \leq \log_{10}(E/\mathrm{eV}) \leq 17.8$ and zenith angles in $0.2 \leq \sin^2\theta \leq 0.3$),
3. $d_{\mathrm{LTP}=90\,\%} = 1172\,\mathrm{m}$, $N_{\mathrm{events}} = 22$, $\log_{10}(\rho_{450}/\mathrm{m}^2) = 0.40$, $\beta = 3.01$ (this corresponds to realistic values and statistics for $18.2 \leq \log_{10}(E/\mathrm{eV}) \leq 18.4$ and zenith angles in $0.0 \leq \sin^2\theta \leq 0.1$).

In every case, the generating muon LDF has parameters $\alpha = 0.75$ and $\gamma = \gamma_{\mathrm{true}}$. Once again, these mock data are treated exactly as real data. The average muon LDF is fitted letting $\log_{10}(\rho_{450}/\mathrm{m}^2)$ and β free, and fixing $\alpha_{\mathrm{rec}} = 0.75$ and $\gamma_{\mathrm{rec}} = 3.0$. This constitutes the fit result of one experiment. We repeat 20 experiments for each value of γ_{true} and each $d_{\mathrm{LTP}=90\,\%}$ scenario, and for each of these we compute the mean fitted muon LDF and the standard deviation.

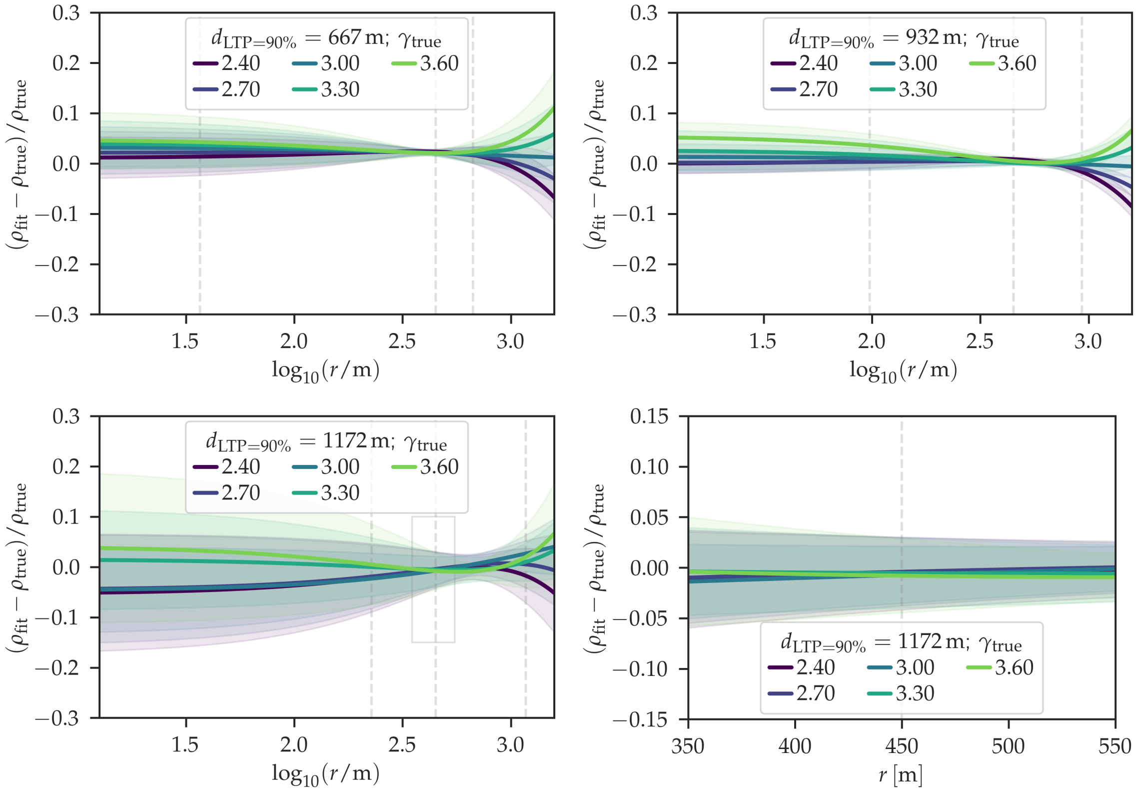


Figure 6.25: Relative error of the fitted muon LDF as a function of the (logarithmic) distance to the shower axis, for a distance where the lateral trigger probability falls below 90 % ($d_{\mathrm{LTP}=90\,\%}$) of 667 m (top left panel), 932 m (top right panel), and 1172 m (bottom left and right panels). Different colors represent different values of the parameter $\gamma = \gamma_{\mathrm{true}}$ of the generating muon LDF. In all cases, the muon LDF was fitted fixing $\gamma_{\mathrm{rec}} = 3.0$. The gray dashed lines locate $r = 450\,\mathrm{m}$, and a gray square on the bottom left panel shows the inset that is plotted on the bottom right panel.

Fig. 6.25 shows the relative error of the fitted muon LDF as a function of the (logarithmic) distance to the shower axis, for different values of $d_{\mathrm{LTP}=90\,\%}$: 667 m (top left panel), 932 m (top right panel), and 1172 m (bottom left and right panels). Different colors represent muon LDFs generated from different values of γ_{true}. The thick lines represent the average over the fitted muon LDFs of all experiments, and the shaded areas represent the standard deviation. A gray dashed line locates $r = 450\,\mathrm{m}$. A gray square encloses the inset that is plotted on the bottom right panel. In the latter, we find the same variable plotted as a function of the distance to the shower axis.

The top panels and left bottom panel of Fig. 6.25 show that a +20 % (−20 %) bias in the fixed value of γ translates into a bias in the muon LDF of at most $(+4.5 \pm 1.3)\,\%$ ($(-2.3 \pm 1.3)\,\%$) within the distance cuts. Once again, the largest biases are attained at the smallest and largest distances from the shower axis. The case where $\gamma_{\mathrm{true}} = \gamma_{\mathrm{rec}} = 3.0$ shows that the biases from the method itself are small (within $\leq 3\,\%$). In the bottom left panel of Fig. 6.25, we can see large standard deviations that result from the low number of events that appear in a realistic high-energy scenario (where $d_{\mathrm{LTP}} = 1172\,\mathrm{m}$), and remains practically unchanged when considering more experiment repetitions. Finally, the bottom right panel of Fig. 6.25 shows that the bias of the muon LDF at $r = 450\,\mathrm{m}$ is (in the worst case, $d_{\mathrm{LTP}} = 1172\,\mathrm{m}$) of at most $(-0.9 \pm 0.8)\,\%$. This implies that $r = 450\,\mathrm{m}$ is still an optimal distance considering systematics in the fit of the average muon LDF.

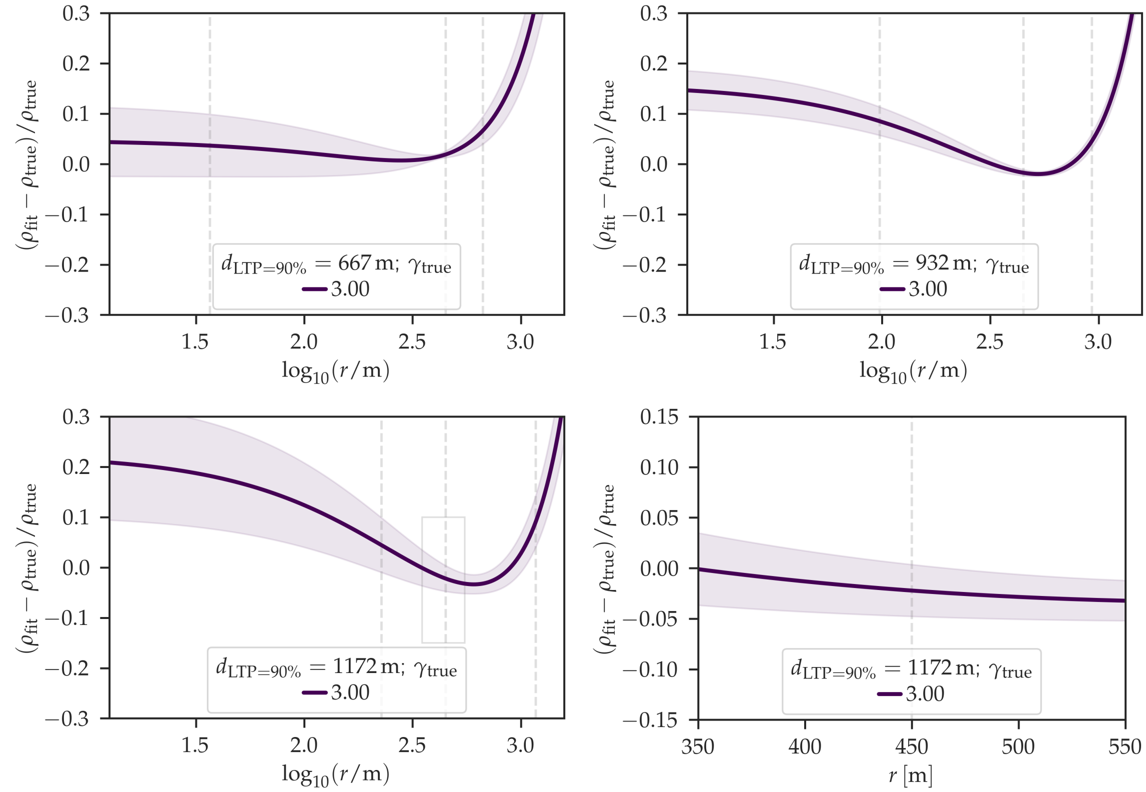


Figure 6.26: Relative error of the fitted muon LDF against the true one as a function of the (logarithmic) distance to the shower axis, for a distance where the lateral trigger probability falls below 90 % ($d_{\rm LTP=90\,\%}$) of 667 m (top left panel), 932 m (top right panel), and 1172 m (bottom left and right panels). In all cases, the true parameter of the muon LDF is $\gamma_{\rm true} = 3.0$ and it is fitted fixing $\gamma_{\rm rec} = 3.0$. The gray dashed lines locate $r = 450$ m, and a gray square on the bottom left panel shows the inset that is plotted on the bottom right panel.

In Sec. 6.2 we found that the term with γ does not have a large impact on the muon LDF in comparison to α and β. However, in Sec. 6.5.2 we showed that the modified NKG fits data significantly better than the NKG (or, equivalently, a modified NKG with $\gamma = 0$). This can be understood with a simple toy experiment: let us assume that the true underlying function is a modified NKG, generate mock data from it, and fit it with an NKG. We want to know what is the bias in the fitted muon density with respect to the true one, as a function of the distance to the shower axis. For this, we repeat the experiment of Fig. 6.25, for a $\gamma_{\rm true} = 3.0$ and $\gamma_{\rm rec} = 0$. The result is shown in Fig. 6.26. We can see that the mean bias can be as high as 11 %, and it can reach 15 % considering fluctuations (bottom left panel). In this case, the bias arises because the fitted parameters, $\log_{10}\rho_{450}$ and β, compensate for the absence of the γ term (or equivalently, for $\gamma = 0$). In conclusion, despite there is not enough sensibility to fit the value of γ, removing this term introduces a bias. Considering the latter, and the fact that we observed that a modified NKG with $\gamma = 3.0$ fits data better than with $\gamma = 0$ (c.f. Sec. 6.5.2), we find that it is better to keep this term of the modified NKG with a fixed value of $\gamma = 3.0$.

6.7.4 Analysis of the bias from cutting at different lateral trigger probabilities

As mentioned in Sec. 6.7.3, the cut at large distances from the shower axis due to the drop in the lateral trigger probability is necessary for keeping only unbiased samples of the muon density. In this section, we quantify and compare the biases introduced by applying this cut at different lateral trigger probabilities. The method is analogous to the one described in Sec. 6.7.3, only that both the γ_{true} and γ_{rec} are fixed to 3.0. In this case, for each of the three realistic scenarios, we generate a mock sample from a given modified NKG (with the same parameters as those described in Sec. 6.7.3). We then cut the mock samples at the distances at which the lateral trigger probability equals 90 %, 80 %, 50 %, 25 %, and 0 % (i.e., no cut). The samples with the distance cuts are fitted with a modified NKG letting $\log_{10} \rho_{450}$ and β free.

The corresponding distance cuts are:

1. $d_{\mathrm{LTP}=90\,\%} = 667\,\mathrm{m}$, $d_{\mathrm{LTP}=80\,\%} = 748\,\mathrm{m}$, $d_{\mathrm{LTP}=50\,\%} = 885\,\mathrm{m}$, and $d_{\mathrm{LTP}=25\,\%} = 1022\,\mathrm{m}$, for $17.2 \leq \log_{10}(E/\mathrm{eV}) \leq 17.4$ and $0.4 \leq \sin^2\theta \leq 0.5$,

2. $d_{\mathrm{LTP}=90\,\%} = 932\,\mathrm{m}$, $d_{\mathrm{LTP}=80\,\%} = 1009\,\mathrm{m}$, $d_{\mathrm{LTP}=50\,\%} = 1142\,\mathrm{m}$, and $d_{\mathrm{LTP}=25\,\%} = 1275\,\mathrm{m}$, for $17.6 \leq \log_{10}(E/\mathrm{eV}) \leq 17.8$ and $0.2 \leq \sin^2\theta \leq 0.3$,

3. $d_{\mathrm{LTP}=90\,\%} = 1172\,\mathrm{m}$, $d_{\mathrm{LTP}=80\,\%} = 1267\,\mathrm{m}$, $d_{\mathrm{LTP}=50\,\%} = 1429\,\mathrm{m}$, and $d_{\mathrm{LTP}=25\,\%} = 1592\,\mathrm{m}$, for $18.2 \leq \log_{10}(E/\mathrm{eV}) \leq 18.4$ and $0.0 \leq \sin^2\theta \leq 0.1$.

Fig. 6.27 shows the relative error of the fitted muon LDF as a function of the (logarithmic) distance to the shower axis, for different values of d_{LTP}, for different realistic scenarios: $17.2 \leq \log_{10}(E/\mathrm{eV}) \leq 17.4$ and $0.4 \leq \sin^2\theta \leq 0.5$ (top left panel), $17.6 \leq \log_{10}(E/\mathrm{eV}) \leq 17.8$ and $0.2 \leq \sin^2\theta \leq 0.3$ (top right panel), and $18.2 \leq \log_{10}(E/\mathrm{eV}) \leq 18.4$ and $0.0 \leq \sin^2\theta \leq 0.1$ (bottom left and right panels). Different colors represent muon LDFs fitted considering different distance cuts at different lateral trigger probabilities. The thick lines represent the average over the fitted muon LDFs of all experiments, and the shaded areas represent the standard deviation. Gray dashed lines mark the minimum distance cut, $r = 450\,\mathrm{m}$, and the different maximum distance cuts. On the bottom right panel, a gray square encloses the inset that is plotted on the bottom right panel. In the latter, we find the same variable plotted as a function of the distance to the shower axis.

In the left and right top panels of Fig. 6.27, we can see that the biases are below 5 %, and there is not a significant difference between the different considered cuts. However, in the bottom left panel, the biases are larger, and also the difference in the biases from cutting at different lateral trigger probabilities is larger. We thus focus on this case. We can see that at higher energies, where the distances at which the lateral trigger probability drops below the considered values are larger, a more conservative cut is beneficial. We find the behavior that we expected: if the data are cut at larger distances from the shower axis (or not cut at all), the densities are positively biased, and so is the fit. Because the model is not flexible enough, in order to fit the positively biased densities at large distances, it underestimates the muon densities at short distances. By choosing to cut at a 90 % lateral trigger probability, the bias is of at most $(+1 \pm 1)\,\%$ or $(-1 \pm 4)\,\%$. From the bottom right panel of Fig. 6.27, we can see that the more conservative cut at 90 % also helps reduce the bias at 450 m, which in this case is of only $(-0.4 \pm 1.1)\,\%$.

6.8 Comparison to previous studies

In this section we compare the muon LDF found in this work with that of a previous data analysis [97] of the prototype array of the UMD, where the modules were instrumented with PMTs instead of SIPMs. It is important to note that a direct comparison of the muon content or the muon LDF of different experiments is not possible. This is because the muon content greatly depends on: the muon vertical energy threshold, the atmospheric depth of the experiment, the primary energies and zenith angles, and the distances to the shower axis,

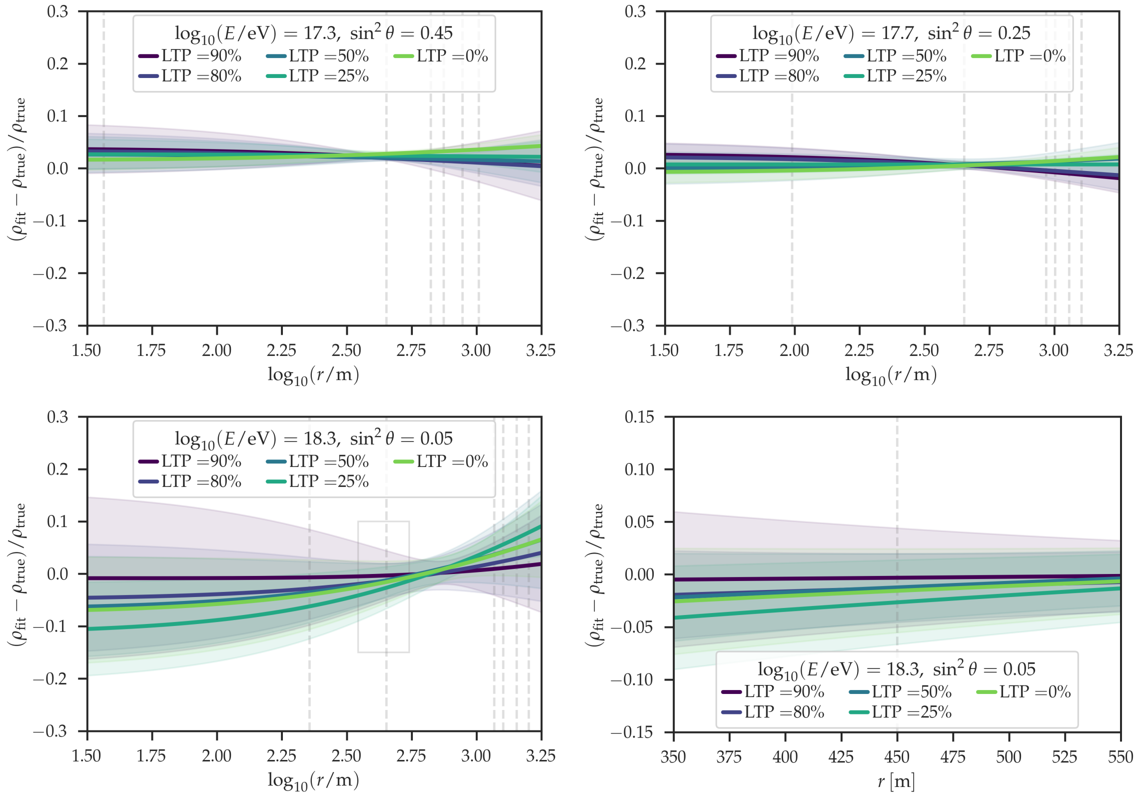


Figure 6.27: Relative error of the fitted muon LDF as a function of the (logarithmic) distance to the shower axis, applying a distance cut where the lateral trigger probability falls below 90 % (purple lines), 80 % (blue lines), and 50 % (green lines), for different realistic scenarios: $17.2 \leq \log_{10}(E/\mathrm{eV}) \leq 17.4$ and $0.4 \leq \sin^2\theta \leq 0.5$ (top left panel), $17.6 \leq \log_{10}(E/\mathrm{eV}) \leq 17.8$ and $0.2 \leq \sin^2\theta \leq 0.3$ (top right panel), and $18.2 \leq \log_{10}(E/\mathrm{eV}) \leq 18.4$ and $0.0 \leq \sin^2\theta \leq 0.1$ (bottom left and right panels). The gray dashed lines locate the minimum and maximum distance cuts, as well as $r = 450$ m. A gray square on the bottom left panel shows the inset that is plotted on the bottom right panel.

apart from being subject to different uncertainties from the different methods with which it is determined. Since the muon content depends almost linearly on the energy, systematic uncertainties in the energy can also greatly affect any comparison between experiments. Future studies could try to parameterize these dependencies using simulations, allowing for the comparison of the muon content of different experiments at least to the extent in which the simulations reproduce the actual behavior of air showers. Alternatively, one could define a scale of the muon content relative to simulations. This is in fact what is done in Chapters 7 and 8 to compare the muon content of different experiments. In this analysis, we limit the comparison to the previous instrumentation of the same experiment. This implies that the phase space for the UMD PMT and UMD SiPM data is the same, and the only difference arises from the response of the electronics of the detector, and from the method for the determination of the muon content.

The aforementioned study [97] determines the muon content using the N-bin strategy (see Chapter 4) and applies a correction of the overall biases based on detector simulations in a similar procedure as in Chapter 5. As mentioned there, there are several differences: at its moment, the dependence of the bias with the number of muons was not know, and so the parameterization of the bias of Ref. [97] does not include such a term. The analysis also suffered from errors in the assumed module azimuth angles. Furthermore, the response

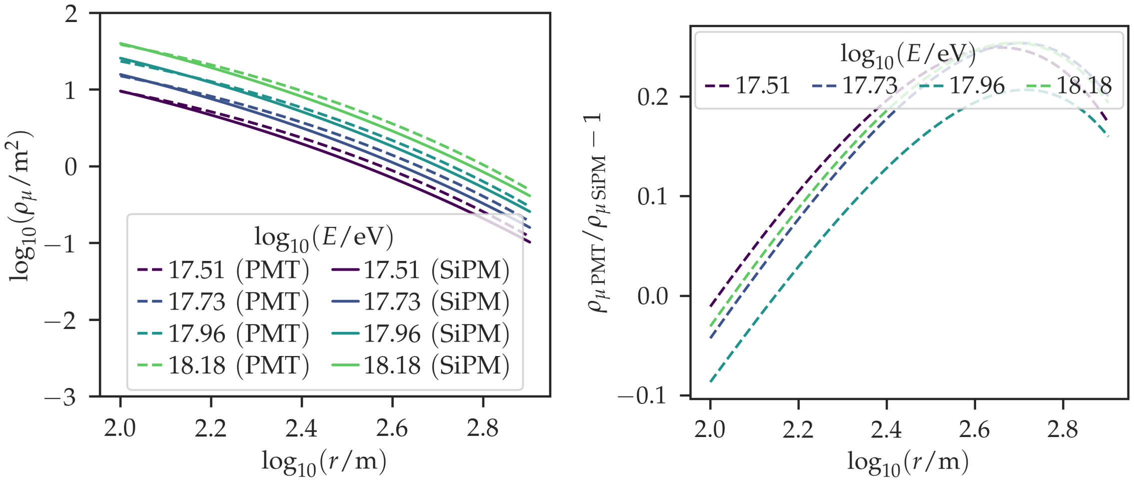


Figure 6.28: Comparison between the muon LDFs determined in this work (solid lines) and those from the analysis of UMD PMT data of Ref. [97] (dashed lines). Different colors identify different energies (see text for details).

and characteristics of the electronics of the SiPMs and PMTs are different. In particular, the efficiency of the modules instrumented with PMTs could possibly have been mismodelled, which could cause too large muon densities. Lastly, there are also concerns on whether the correction of the bias was applied with the correct sign. The bias correction should make the uncorrected muon densities smaller, but if wrongly applied, it would make them higher. The there calculated bias, and therefore the size of the correction, ranged between $\sim 0\,\%$ to $\sim 30\,\%$.

Figure 6.28 shows the comparison between the muon LDF as parameterized in this work against the one of Ref. [97]. The left panel shows the parameterized average muon LDFs from this work at different logarithmic energies, as well as the fitted muon LDFs to PMT data. Different logarithmic energies are shown in different colors. The PMT muon LDFs are fits to binned data, where the centers of the logarithmic energy bins are the ones indicated in the plot, with a bin width of $\Delta \log_{10}(E/\mathrm{eV}) = 0.11$, and with zenith angles in $0^\circ\theta \leq 24.62^\circ$ ($0 \leq \sin^2\theta \leq 0.17$). The UMD muon LDFs use the parameterization of the modified NKG of Eqs. (6.24) and (6.25) evaluated at the energies of the centers of the bins and at the mean $\sin^2\theta$ of the considered interval. The right panel shows the relative difference between the muon LDFs of PMT and those of UMD for the same cases. From Fig. 6.28 we can see that the muon content of both analysis is compatible at short distances from the shower axis (between $\sim$ 100 m to $\sim$ 160 m), but that at larger distances the PMT muon densities are larger than the ones found in this work by up to $\sim 25\,\%$. At 450 m, the difference is between $\sim 20\,\%$ to $\sim 25\,\%$.

Furthermore, in Ref. [97] a modified NKG is event-wise fitted to PMT data, obtaining estimates of ρ_{450} for each selected event. The shape of the fitted modified NKG is determined by its parameters, which are optimized for simulations: $r_0 = 280\,\mathrm{m}$, $\alpha = 0.3$, $\gamma = 4.6$ (like in this analysis, $r_\mathrm{ref} = 450\,\mathrm{m}$). For low-multiplicity events, the parameter β is fixed by the following parameterization, also optimized for simulations:

$$\beta = b_0 + b_1 \sec(\theta), \tag{6.31}$$

where $b_0 = 4.4$ and $b_1 = -1.1$. The obtained values of ρ_{450} at different zenith angles are then normalized to 35° via a constant intensity cut. This is a procedure that uses the fact that the expected rate of events is isotropic to estimate the zenith angle dependence of ρ_{450} (for more

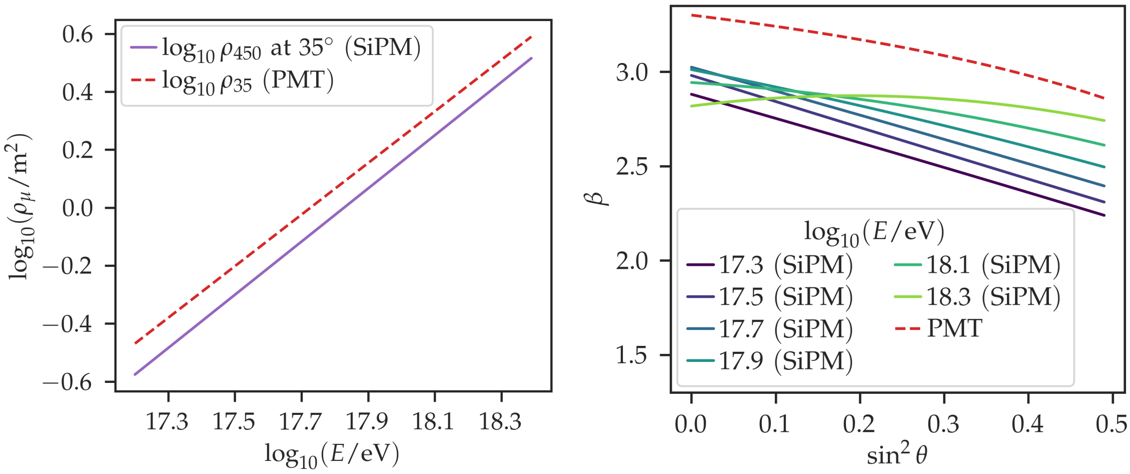


Figure 6.29: Left panel: Comparison between the parameterization of $\log_{10}\rho_{450}$ of this work (c.f. Eq. (6.24)) evaluated at $\theta = 35°$ as a function of the logarithmic energy (solid line), and the parameterization of $\log_{10}\rho_{35}$ from the UMD PMT data analysis (c.f. Eq. (6.32), dashed line). Right panel: Comparison between the parameterization of β from this work (c.f. Eq. (6.25)) evaluated at different energies (different colors, solid lines) as a function of the sine square of the zenith angle, and the one of the UMD PMT data analysis (c.f. Eq. (6.31), dashed lines).

information see for example Refs. [92, 120]). The values of ρ_{35} are fitted with a power-law in energy

$$\rho_{35} = A\left(\frac{E}{10^{18}}\right)^b, \tag{6.32}$$

and the reported values of the parameters are $A = (1.75 \pm 0.05(\text{stat.}) \pm 0.05(\text{syst.}))\,\mathrm{m}^{-2}$ and $B = 0.89 \pm 0.04$ (stat.) $\pm\, 0.04$ (syst.) [90].

The left panel of Fig. 6.29 shows the comparison of the parameterization of $\log_{10}\rho_{450}$ of this work (c.f. Eq. (6.24)) evaluated at $\theta = 35°$ as a function of the logarithmic energy, against the parameterization of $\log_{10}\rho_{35}$ reported in Ref. [97]. The right panel of Fig. 6.29 shows, for completeness, the comparison of the parameterization of β from this work (c.f. Eq. (6.25)) evaluated at different energies as a function of the sine square of the zenith angle, against the one of the UMD PMT data analysis (c.f. Eq. (6.31)). However, the difference in the rest of the parameters of the modified NKG between the two analyses (in particular the large difference in α) makes the comparison of β meaningless. Despite the muon LDFs from the PMT data analysis are less steep (see Fig. 6.28), the values of β are larger than those of this work, which are necessary to compensate the smaller value of α.

In the left panel of Fig. 6.29 we observe differences between the logarithmic muon densities ranging between 0.11 and 0.07. The latter translate to differences of 28 % to 17 % in muon density (respectively). A previous study of UMD SiPM data [121], reported that the UMD PMT muon densities were $\sim$ 18 % larger than those of the UMD SiPM data determined in that work. In other words, an independent analysis found differences compatible with the ones found in this work. Although we detailed the differences between the UMD PMT data analysis and the one of UMD SiPM data of this work, the exact cause of the difference is not yet known, and is subject of further investigations.

6.9 Summary

In this Chapter we analyzed the muon Lateral Distribution Function (muon LDF) estimated from the muon densities measured with the Underground Muon Detector of the Pierre Auger Observatory.

We started by introducing the different models of the muon LDF found in literature: the NKG, the modified NKG, and the Hillas functions; and we also introduced a simple log-log polynomial function. We particularly analyzed the shape of the modified NKG to understand how it depends on its parameters. We observed that the term with α is comparably important, but that the shape can still vary widely only by varying the β factor. We also observed that the term with γ affects the shape of the muon LDF only at large distances from the shower axis.

We then presented the way in which we select the data. On one hand, we showed that the array trigger efficiency imposes a minimum energy and maximum zenith angle of the events. On the other hand, we saw that the maximum energy that we can select is limited by statistics. With the selected ranges, we binned the data in logarithmic energy and sine-square of the zenith angle, and implemented an outlier detection procedure. For null muon densities we based the outlier detection method on the measured signal at the paired water-Cherenkov detector (WCD), flagging as outliers the points where the UMD muon density was null or very small, and the WCD signal was very large. For positive muon densities we used a robust iterative procedure based on the z-score computed from the median and median absolute deviation of the logarithmic muon density. Finally, it was also necessary to introduce cuts in the distance of the module to the shower axis to avoid sampling biases. We implemented a minimum distance cut based on the expected muon density, which ensures a probability of saturation below 1 %. At large distances, the distance cut is related to the decreasing lateral trigger probability, which, with the cut we introduce, lays above 90 %.

Furthermore, we described the fitting procedure, in which we group the data in bins of logarithmic energy, sine-square of the zenith angle, and also in logarithmic distance. We normalized the muon densities to the energy of the center of the bin to avoid spurious fluctuations or biases. We then computed in each bin the logarithm of the mean and standard deviation of the normalized muon densities. We inserted these in a χ^2 function which we minimize for each $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bin. Moreover, we introduced the information criteria used to analyze the goodness-of-fit, the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). We also showed how we can use the Wald-Wolfowitz test to analyze the randomness of the signs of the residuals of a fit, which constitutes a goodness-of-fit criterion that is complementary to the AIC and BIC. We combined the p-values of the Wald Wolfowitz test for each $(\log_{10}(E/\mathrm{eV}),\ \sin^2\theta)$-bin using the Fisher method.

We computed the muon LDF from data and compared how the different analyzed models fit it. We observed that fitting the parameters α and γ of the modified NKG is unfeasible due to the lack of unbiased statistics very close and very far from the shower core, respectively. We compared the AIC, BIC, and p-value of the Wald-Wolfowitz test for the different models and found that a modified NKG with fixed parameters $r_0 = 320\,\mathrm{m}$, $\alpha = 0.75$, and $\gamma = 3.0$ is the best overall fitting function. Moreover, based on the latter fits, we provided two parameterizations of the muon LDF of data as a function of the energy and zenith angle, or shower size and zenith angle, which can be useful for fitting event-wise measurements of the muon LDF, for making realistic toy models, or as a reference for the design of other muon detectors.

Moreover, we use the aforementioned toy model of the muon LDF to analyze the systematic uncertainties in the fitted muon LDFs introduced by fixing the parameters α and γ, and also by the overall characteristics of the fitting procedure used in the work. We observed that a +20 % (−20 %) bias in the fixed value of α translates into a bias in the muon LDF of at most (-3.2 ± 0.4) % ($(+6.0 \pm 0.4)$ %), and that a +20 % (−20 %) bias in the fixed value of γ translates into a bias in the muon LDF of at most $(+4.5 \pm 1.3)$ % ((-2.3 ± 1.3) %) considering distance cuts. We also showed that, despite not having enough statistics to fit γ, fixing it to the value $\gamma = 3.0$ introduces less biases than removing the factor from the modified NKG function. We proved that applying a maximum distance cut at a 90 % lateral trigger probability is conservative and helps keep biases small, specially at high energies. Since fixing these parameters and applying the aforementioned distance cut biases in less than ±2 % the muon density at 450 m, this distance proves to be the optimal one to minimize the uncertainties in the muon LDF, even when considering the systematics introduced by the fit of the average muon LDF.

Finally, we compared the muon LDF as determined in this work from UMD SiPM data to the one reported in Ref. [97] computed from UMD PMT data. We found that the muon content at 450 m from the shower axis and at a zenith angle of 35° of the PMT analysis is higher by 17 % to 28 % than the one determined in this work, in agreement with the results of an independent analysis of SiPM data [121]. Although the differences between the two analyses were explained, the exact cause of the discrepancy is not yet known and is subject of future investigations.

x	$a_{x,0}$	$a_{x,1}$	$a_{x,2}$	$b_{x,0}$	$b_{x,1}$	$b_{x,2}$	$c_{x,0}$	$c_{x,1}$	$c_{x,2}$
$\log_{10}\rho_{450}$	-0.482281	0.044838	-0.072323	0.889217	0.143859	-0.153548	0	0	0
β	2.880601	-0.634257	-0.020821	0.639228	-0.591242	0.264586	-0.701877	1.511456	-0.615173

x	$\tilde{a}_{x,0}$	$\tilde{a}_{x,1}$	$\tilde{a}_{x,2}$	$\tilde{b}_{x,0}$	$\tilde{b}_{x,1}$	$\tilde{b}_{x,2}$	$\tilde{c}_{x,0}$	$\tilde{c}_{x,1}$	$\tilde{c}_{x,2}$
β	3.028379	-0.572124	0.011611	-0.115562	1.080899	-0.389606	-0.883964	1.905430	-0.828611

Table 6.1: Values of the parameters of the biquadratic fits to $\log_{10}\rho_{450}$ as a function of $\log_{10}(E/\mathrm{eV})$ and $\sin^2\theta$ (first row of values), of the fit to β as a function of $\log_{10}(E/\mathrm{eV})$ and $\sin^2\theta$ (second row of values), and of the fit to β as a function of $\log_{10}\rho_{450}$ and $\sin^2\theta$ (third row of values).

Chapter 7

Comparison of the muon content of underground muon detector data and simulations

As discussed in Chapter 2, it is a known fact that current air-shower simulations fail to reproduce the muon content found in data. This is obtained from analyzing the consistency of the mass composition implications of X_{max} and of the number of muons N_μ (or equivalently the muon density ρ_μ). By understanding where the deficit is (i.e. in which part of the phase space), the hadronic interaction models can be improved, and conversely, the new hadronic interaction models can be used to provide better mass composition estimates.

In this chapter we compare the muon densities estimated from measurements of the Underground Muon Detector of the Pierre Auger Observatory (as described in Chapters 4 and 5) to muon densities obtained from simulations, both for single proton and single iron primaries, as well as for mixed compositions scenarios, for the high-energy hadronic interaction models EPOS-LHC, QGSJetII-04, and Sibyll2.3c. The aim is to understand whether the UMD data are consistent with a muon deficit in air shower simulations, and if so, to quantify it. In particular, we study how the difference between the muon content of data and simulations depends on the energy, zenith angle, and distance to the shower axis.

The structure of the chapter is as follows. In Sec. 7.1 we describe the analysis, including the analytical computation of energy resolution effects in the simulated muon densities, the transformation to the reference energy scale introduced by the *Spectrum Working Group* [122], the computation of the muon density (divided by the reconstructed energy) for pure proton, pure iron, and mixed composition scenarios, and the computation of a z-value. In Sec. 7.2 we show the comparisons on the muon denisty as a function of the distance, and as a function of the reconstructed energy, for different zenith angles. We also show the computed z-values and how they compare to the expected values predicted by the mixed composition models, as well as to those of other analyses of Pierre Auger data. Finally, in Sec. 7.3 we summarize the main conclusions.

It is important to mention that the methods used in this chapter are strongly based on the ones used in the following **publications**:

- *Muon deficit in air shower simulations estimated from AGASA muon measurements*, F. Gesualdi, A. D. Supanitsky, and A. Etchegoyen, Phys. Rev. D 101, 083025, 2020 [123],

- *Muon deficit in simulations of air showers inferred from AGASA data*, F. Gesualdi, A. D. Supanitsky, and A. Etchegoyen, Proc. 37th Int. Cosmic Ray Conf., 2021 [124],

- *On the muon scale of air showers and its application to the AGASA data*, F. Gesualdi, H. Dembinski, K. Shinozaki, A.D. Supanitsky, T. Pierog, L. Cazon, D. Soldin, and R. Conceição on behalf of the Working group on Hadronic Interactions and Shower Physics (WHISP), Proc. 37th Int. Cosmic Ray Conf., 2021 [125].

7.1 Analysis

In this section we describe how we analytically compute the detector effects in the simulated muon densities. We also introduce the reference energy scale. We show how we compute the muon density (divided by the reconstructed energy) for pure proton, pure iron, and mixed composition scenarios. Finally we describe the computation of a z-value.

7.1.1 Simulating detector effects in the muon density

In this analysis we do not use $\overline{\text{Off}}\underline{\text{line}}$ [70] detector simulations because, at the time of writing, they were not in a production-ready state. There are concerns regarding the normalization of the muon flux in the detector. More specifically, the number of muons injected in the detector seem to differ from the ones expected from the CORSIKA [23] simulations. Nevertheless, this does not affect (at least to a first order) detector-specific corrections like the one described in Chapter 5, where the reconstructed number of muons is corrected against the ones injected in the $\overline{\text{Off}}\underline{\text{line}}$ simulations, being the correct normalization of the simulations not crucial. This is not the case for this analysis, and for this reason we mimic the detector effects analytically.

When comparing the muon content of data against simulations, it is important to consider the effects of the energy reconstructions in simulations. This is because the number of muons is proportional to E^{β}, where $\beta \sim 0.9$. While the muon content of data is a function of the reconstructed energy E_R, that of simulations is a function of the Monte-Carlo or true energy E. These two cannot be directly compared, even if E_R estimates E with no bias [126].

To account for the energy reconstruction effects, we compute the average muon density at a distance r divided by the reconstructed energy to the power of α, in the i^{th} reconstructed energy bin, as [123]

$$\left\langle \frac{\rho_{\mu\, r}}{E_R^{\alpha}} \right\rangle (E_{Ri}) = \frac{\displaystyle\int_{E_{Ri}^-}^{E_{Ri}^+} \int_0^{\infty} \langle \widetilde{\rho}_{\mu\, r} \rangle (E)\, E_R^{-\alpha}\, J(E)\, G(E_R|E)\, \varepsilon(E) dE\, dE_R}{\displaystyle\int_{E_{Ri}^-}^{E_{Ri}^+} \int_0^{\infty} J(E)\, G(E_R|E)\, \varepsilon(E) dE\, dE_R}, \tag{7.1}$$

where E_{Ri}^- and E_{Ri}^+ are the lower and upper limits of the reconstructed energy bin of center E_{Ri}. Notice that if α is zero we obtain the mean value of $\rho_{\mu\, r}$. Furthermore,

- $\langle \widetilde{\rho}_{\mu\, r} \rangle (E)$ is the average muon density at distance r from the shower axis (measured on the shower plane) as a function of the Monte Carlo true energy of the simulation [1], which is obtained from power-law fits to air-shower simulations performed by using CORSIKA [23],

- $J(E)$ is the cosmic ray flux, which is obtained from the fit to the Pierre Auger cosmic ray flux presented in Ref. [92],

- $G(E_R|E)$ is the conditional probability distribution of E_R conditioned to E, which is modelled as a log-normal distribution with a standard deviation taken from Ref. [127],

---

[1] The tilde in $\langle \widetilde{\rho}_{\mu} \rangle (E)$ emphasizes that this quantity cannot be directly compared to the average muon density computed from data.

- $\varepsilon(E)$ is the trigger efficiency as a function of the Monte Carlo true energy, modelled as an error function [128].

We first explain the rationale behind Eq. (7.1) and afterwards the details of each of the functions involved. The idea is as follows [123]. The true energy of a real or simulated air shower, E, is reconstructed to a value E_R, which constitutes a fluctuation around E according to $G(E_R|E)$. Moreover, the cosmic ray flux $J(E)$ determines the natural distribution of the true energy E (normalized within a certain energy range). The product of the efficiency and the cosmic ray flux, $J(E)\varepsilon(E)$ gives the observed distribution of the energy E at the detector. Then, the product $G(E_R|E)J(E)\varepsilon(E)$ constitutes the joint probability distribution of E and E_R.

Let us first examine what happens when we calculate the mean muon density at the reconstructed energy E_R, $\langle\rho_{\mu\, r}\rangle(E_R)$, using Eq. (7.1) with $\alpha = 0$ and in the limit in which the reconstructed energy bin width tends to zero, at high and at low energies. In the former case, the efficiency is ~ 1. While $G(E_R|E)$ is roughly symmetric with respect to the energy, $J(E)$ is highly asymmetric as it drops steeply with energy. Then, the product $J(E)G(E_R|E)$ is also asymmetric, taking larger values for lower energies. This means that an event with reconstructed energy E_R comes, most likely, from an event with energy $E < E_R$. Then, $\langle\rho_{\mu\, r}\rangle(E_R)$ is calculated as the integration of the contributions of $\langle\widetilde{\rho}_{\mu\, r}\rangle(E)$ weighted by the product $J(E)G(E_R|E)$ (normalized within a certain energy range). We know that $\langle\widetilde{\rho}_{\mu\, r}\rangle(E)$ can be modelled a power law in energy, i.e. $\propto E^\beta$, with $\beta \sim 0.9$. This means that $\langle\widetilde{\rho}_{\mu\, r}\rangle(E)$ is smaller for lower energies, which weigh more in the integration. Therefore, evaluated at a specific numerical value E^*, $\langle\rho_{\mu\, r}\rangle(E_R = E^*) < \langle\widetilde{\rho}_{\mu\, r}\rangle(E = E^*)$. The difference increases where the flux $J(E)$ is steeper and where the energy resolution function $G(E_R|E)$ is broader. We now introduce the integration in a reconstructed energy bin, taking it into account in the normalization as well. This has the additional (smaller) effect of making the convolved muon density even smaller: at a bin centered at E_{Ri}, $\langle\rho_{\mu\, r}\rangle(E_{Ri} = E^*) < \langle\rho_{\mu\, r}\rangle(E_R = E^*)$. Furthermore, for lower energies, the efficiency $\varepsilon(E)$ can play a significant role. Going from higher to lower energies, the efficiency can drop more steeply than what the flux increases, making the product $G(E_R|E)J(E)\varepsilon(E)$ asymmetric in the opposite way as for higher energies, i.e., taking smaller values for lower energies. Following the same steps as before, mutatis mutandis, we can conclude that, evaluated at a specific numerical value E^*, $\langle\rho_{\mu\, r}\rangle(E_R = E^*) > \langle\widetilde{\rho}_{\mu\, r}\rangle(E = E^*)$, and also $\langle\rho_{\mu\, r}\rangle(E_{Ri} = E^*) > \langle\rho_{\mu\, r}\rangle(E_R = E^*)$ in a bin centered at E_{Ri}.

We now detail the functions used in Eq. (7.1). In first place, for every model, primary, and distance bin, we fit $\langle\widetilde{\rho}_{\mu\, r}\rangle(E)$ with the equation

$$\langle\widetilde{\rho}_{\mu\, r}\rangle(E) = \rho_{\mu\, r(19)} \left(\frac{E}{10^{19}\,\mathrm{eV}}\right)^{\beta}, \tag{7.2}$$

where $\rho_{\mu\, r(19)}$ is the muon density at distance r and at energy 10^{19} eV, and where β is the exponent. Both are left as free parameters of the fit.

The muon flux $J(E)$ follows the parameterization [92]

$$J(E) = J_0 \left(\frac{E}{10^{18.5}\,\mathrm{eV}}\right)^{-\gamma_1} \prod_{i=1}^{3} \left[1 + \left(\frac{E}{E_{ij}}\right)^{1/\omega_{ij}}\right]^{(\gamma_i-\gamma_j)\omega_{ij}}, \tag{7.3}$$

where $j = i+1$, and the parameters take the values $J_0 = 1.315 \times 10^{-18}\mathrm{km}^{-2}\mathrm{sr}^{-1}\mathrm{yr}^{-1}\mathrm{eV}^{-1}$, $E_{12} = 5.0 \times 10^{18}\mathrm{eV}$, $E_{23} = 13 \times 10^{18}\mathrm{eV}$, $E_{34} = 46 \times 10^{18}\mathrm{eV}$, $\gamma_1 = 3.29$, $\gamma_2 = 2.51$,$\gamma_3 = 3.05$, $\gamma_4 = 5.1$, and $\omega_{ij} = 0.05$. An explanation of the shape of the parameterization can be found in Ref. [92].

In Ref. [127], the energy resolution function $G(E_R|E)$ is modelled as a Gaussian distribution of mean $\mu(E) = E$ and standard deviation $\sigma(E)$

$$\sigma(E) = E\left[0.06 + 0.05\left(\frac{E}{10^{18}\,\mathrm{eV}}\right)^{-\frac{1}{2}}\right]. \tag{7.4}$$

However, we model the detector resolution as a log-normal distribution. On one hand, this is what is expected from first principles, as the signal arises from a multiplicative process. Hence, it is natural that the signal at 450 m from the shower axis (the energy estimator of the SD-750 array) and therefore the SD energy follows a log-normal distribution. For example, in Ref. [120] such model is used for the parameterization of the energy resolution for the SD-1500 array. On the other hand, the log-normal distribution is, in this case, more simple to use than the Gaussian, because it is analytically integrable even when it is divided by E_R. In any case, since the energy resolution is considerably good, the impact of the choice of the specific resolution function is negligible. The log-normal distribution is defined as

$$G(E_R|E) = \frac{1}{\sqrt{2\pi}\,\sigma_l(E)E_R}\,\exp\left[-\frac{1}{2}\left(\frac{\ln E_R - \mu_l(E)}{\sigma_l(E)}\right)^2\right], \tag{7.5}$$

where $\mu_l(E)$ and $\sigma_l(E)$ are the parameters of the distribution. These are determined by requiring that the mean of the log-normal distribution matches the mean $\mu(E) = E$ and variance $\sigma^2(E)$ of the Gaussian distribution:

$$\exp\left[\mu_l(E) + \frac{\sigma_l(E)^2}{2}\right] = \mu, \tag{7.6}$$

$$\left[\exp(\sigma_l(E)^2) - 1\right]\exp\left[2\mu_l(E) + \sigma_l(E)^2\right] = \sigma^2(E). \tag{7.7}$$

Solving the system of equations for $\mu_l(E)$ and $\sigma_l(E)$ we find

$$\mu_l(E) = \ln\mu(E) - \frac{1}{2}\ln\left[1 + \left(\frac{\sigma(E)}{\mu(E)}\right)^2\right], \tag{7.8}$$

$$\sigma_l(E) = \sqrt{\ln\left[1 + \left(\frac{\sigma(E)}{\mu(E)}\right)^2\right]}. \tag{7.9}$$

Finally, the 6T5 trigger efficiency is parameterized as [128]

$$\varepsilon(E) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left[3.71\log_{10}\left(\frac{E}{10^{16.93}\,\mathrm{eV}}\right)\left(1. + 0.187\log_{10}\left(\frac{E}{10^{16.93}\,\mathrm{eV}}\right)\right)\right], \tag{7.10}$$

which is very similar to the parameterization derived in Ref. [129] but for events with zenith angles $\theta \leq 47°$.

Figure 7.1 shows a comparison between $\langle\widetilde{\rho}_{\mu\,r}/E\rangle(E)$ and $\langle\rho_{\mu\,r}/E_R\rangle(E_{Ri})$ as a function of the corresponding energy for pure proton and pure iron EPOS-LHC simulations at 450 m from the shower axis. We can see that the difference between the curves is not large. $\langle\rho_{\mu\,r}/E_R\rangle(E_{Ri})$ can be up to 5 % smaller and up to 14 % larger (at the lowest energies) than $\langle\widetilde{\rho}_{\mu\,r}/E\rangle(E)$ in the analyzed energy range (from $10^{16.8}$ eV to $10^{18.6}$ eV). As explained before, at the lowest energies, the dominant effect is the steeply decreasing trigger efficiency, while at higher energies the resolution in the reconstructed energy dominates. Binning in reconstructed energy with a bin width of $\Delta\log_{10}(E_R/\mathrm{eV}) = 0.2$ has in comparison only a small effect.

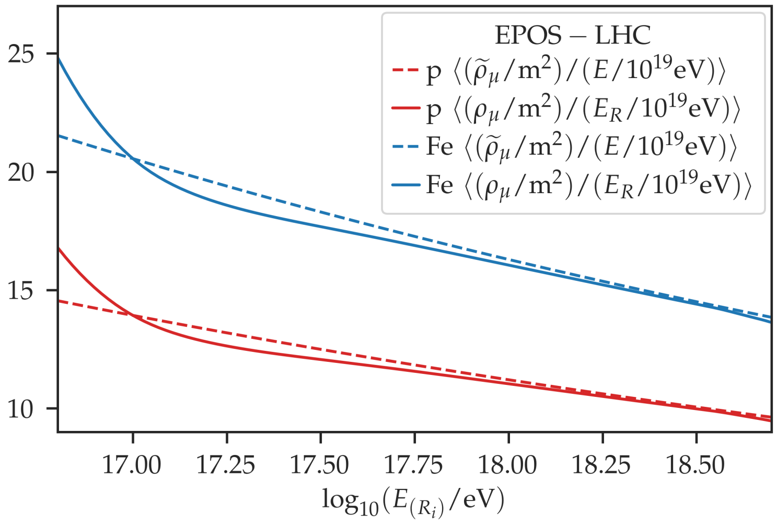


Figure 7.1: Average muon density at 450 m from the shower axis, divided by the energy, as a function of the logarithm of the Monte Carlo energy E (dashed lines) and average muon density at 450 m from the shower axis, divided by the reconstructed energy, as a function of the logarithm of the reconstructed energy bin E_{Ri} (solid lines). We used a bin width of $\Delta \log_{10}(E/\mathrm{eV}) = 0.2$. The high-energy hadronic interaction model used to generate the air showers is EPOS-LHC, and the primaries are proton (red) and iron (blue).

It is relevant to mention the fact that none of the functions in Eq. (7.1) are defined from 0 to ∞ in E. We restrict the integration range to $16.8 \leq \log_{10}(E/\mathrm{eV}) \leq 18.7$. With this energy interval we ensure to contain $\geq 97.5\,\%$ of the true energies E that are mapped to reconstructed energies E_R in the boundaries of the studied interval, $\log_{10}(E_R/\mathrm{eV}) = 17.22$ and $\log_{10}(E_R/\mathrm{eV}) = 18.42$. The impact of using a finite integration range as opposed to an infinite one is estimated to be $\sim 0.1\,\%$ in this energy range, which is negligible in comparison with other sources uncertainties. The integrals are computed numerically using ROOT [98].

7.1.2 Transformation to the reference energy scale

When comparing the analysis of different air-shower experiments, it is crucial to use a unified energy scale. On one hand, we want this analysis to be comparable to the analysis of AGASA data performed in Chapter 8. On the other hand, we use both Pierre Auger and Telescope Array mass composition fractions as a function of the energy, which is another good reason to work on a unified scale. Because of these reasons we work with the so-called reference energy scale introduced by the *Spectrum Working Group* [122].

Each experiment has its own energy scale, E^{data}, which has to be multiplied by a factor $f_{\mathrm{E}} = E^{\mathrm{ref}}/E^{\mathrm{data}}$ to transform the energies to the cross-calibrated energy scale E^{ref}. The factors f_{E} are estimated by matching the characteristic features of the different flux measurements, relying on the assumption that all experiments should observe the same cosmic ray flux because of its isotropic nature.

The correction factor for the Pierre Auger experiment is reported to be $f_{\mathrm{Auger}} = 1.052$ [122], while that of the Telescope Array experiment is reported to be $f_{\mathrm{TA}} = 0.948$ [122]. Furthermore, we take as systematic uncertainties of the Pierre Auger data in the reference energy scale, the ones that they report on their energy scale, i.e., $\pm 14\,\%$ [130].

It is relevant to add that the energy in Eq. (7.4) is in the Auger energy scale. Expressing the energy resolution as a function of the reference energy scale we obtain

$$\sigma(E_{\text{ref}}) = E_{\text{ref}}\, f_{\text{Auger}} \left[0.06 + 0.05 \left(\frac{E_{\text{ref}}\, f_{\text{Auger}}}{10^{18}\,\text{eV}} \right)^{-\frac{1}{2}} \right]. \tag{7.11}$$

7.1.3 Calculation of the muon density

Data

As mentioned in the introduction to this chapter, the muon densities are estimated from measurements with the Underground Muon Detector of the Pierre Auger Observatory equipped with SiPMs. The dataset is the same used in Chapter 6, i.e., they are measurements taken between January 2018 and December 2021. For the reconstruction we use $\overline{\text{Off}}\underline{\text{line}}$ (git version 5e0048b4) [70]. The energy and geometry is reconstructed from the SD information. The muon densities are reconstructed using the estimator of the mean number of muons μ of the 1-bin counting strategy described in Chapter 4 with a 12-bin (37.5 ns) single-muon pattern. The muon densities are also bias-corrected against simulations as described in Chapter 5. After the data selection described in Chapter 6, the muon densities are grouped in bins of logarithmic energy in the range $17.22 \leq \log_{10}(E_R/\text{eV}) \leq 18.42$ in bins of width 0.2, in zenith angles between $0^\circ \leq \theta \leq 45^\circ$ ($0 \leq \sin^2\theta \leq 0.5$) in two bins, being the bin edge 30° ($\sin^2\theta = 0.25$), and in logarithmic distance to the shower axis in the range $1.5 \leq \log_{10}(r/\text{m}) \leq 3.2$ in bins of width 0.1 (but not all distance bins are populated for every energy and zenith angle bin). We propagate the statistical uncertainty of the core, following the model described in Sec. 6.7.1, into the muon density; we also take into account a contribution from the distance bin width. Null measured muon densities are included in the analysis. With respect to what is shown in Chapter 6, there is a difference of 0.02 in the energy scale ($\log_{10} f_{\text{Auger}}$) because of using the reference energy scale in this case. Also, in contrast to Chapter 6, we do not normalize the measured muon densities to the center of the reconstructed energy bin, but rather take into account the effect of binning in reconstructed energy in the simulated muon densities (c.f. Eq. (7.1)).

Simulations

We generated a library of proton, helium, nitrogen, and iron air showers using the high-energy hadronic interaction models EPOS-LHC [26], QGSJetII-04 [27], and Sibyll2.3c [28], and the low-energy hadronic interaction model UrQMD [32, 33], within CORSIKA v7.7402 [23]. The generated arrival directions were isotropically distributed in the range $0^\circ \leq \theta \leq 45^\circ$ and with a uniform distribution in the logarithm of the (Monte Carlo) primary energy in $16.8 \leq \log_{10}(E/\text{eV}) \leq 18.7$. We simulated 3000 air showers per primary and per high-energy hadronic interaction model.

For each simulated air shower, the muon density at a given distance bin is calculated by counting the muons that fall within the ring on the shower plane that the distance-bin determines, which would reach ground with an energy $E_\mu \geq 1\text{GeV}/\cos(\theta_\mu)$, being θ_μ the zenith angle of the muon. Then the count is divided by the area of the shower-plane ring. Furthermore, for each energy, zenith angle, and distance bin, the muon densities of all showers that populate the bin are averaged. For each model, primary, zenith angle and distance bin, the average muon densities as a function of the Monte Carlo energies are fitted with the power law of Eq. (7.2). Finally, $\langle \rho_{\mu\,r}/E_R^\alpha \rangle (E_{Ri})$ is calculated via the numerical evaluation of Eq. (7.1).

Mixed composition scenarios

The muon densities for the mixed composition scenarios $\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle(E)$ are calculated as the linear combination of the single primary (proton, helium, nitrogen, and iron) muon densities, being the factors the mass fractions

$$\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle(E) = \sum_{A} f_A(E)\,\langle\widetilde{\rho}_{\mu r}^{\,A}\rangle(E), \tag{7.12}$$

where $A = \{\mathrm{p}, \mathrm{He}, \mathrm{N}, \mathrm{Fe}\}$ and $f_A(E)$ is the mass fraction as a function of primary energy. For energies above $10^{17.22}$ eV we use the mass fractions obtained from the fits to the X_{max} experimental distributions by the Pierre Auger Collaboration [57], while for lower energies we use the mass fractions estimated also from fits to the X_{max} distributions but by the Telescope Array Collaboration [131]. The mass fractions as a function of each experiment's energy are transformed to the reference energy scale, and are linearly interpolated between the discrete values in which they are given to obtain a continuous function. It is relevant to add that the Telescope Array mass fractions are estimated for EPOS-LHC, but we use them for QGSJetII-04 and Sibyll2.3c too considering an extra 10 % systematic uncertainties in the mass fractions. The mass fractions of Telescope Array only play a secondary role in this analysis, because they only partially influence the resulting convolved muon densities at the lower reconstructed energies.

The convolved $\langle\rho_{\mu r}^{\mathrm{mix}}/E_R^{\alpha}\rangle(E_{Ri})$ is calculated by introducing $\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle(E)$ (as defined in Eq. (7.12)) through Eq. (7.1). Notice that $\langle\rho_{\mu r}^{\mathrm{mix}}/E_R^{\alpha}\rangle(E_{Ri})$ is not equivalent to a linear combination of $\langle\rho_{\mu r}^{A}/E_R^{\alpha}\rangle(E_{Ri})$ because the mass fractions $f_A(E)$ depend on the energy E, that is an integration variable.

We treat the statistical and systematic uncertainties of $\langle\rho_{\mu r}^{\mathrm{mix}}/E_R^{\alpha}\rangle(E_{Ri})$ as follows: for each model, and for each discrete energy value, we select the combination of mass fractions within the boundaries of its uncertainties that maximize and minimize $\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle(E)$ (this over-estimates the uncertainties, but given that the covariance matrices of the mass fraction fits are not available this method is the best approach). In this way, we calculate $\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle(E) \pm \sigma[\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle](E)$ for each discrete energy value. Moreover, the values of $\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle(E) + \sigma[\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle](E)$ and $\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle(E) - \sigma[\langle\widetilde{\rho}_{\mu r}^{\,\mathrm{mix}}\rangle](E)$ are linearly interpolated in the energy range under consideration. Finally, the uncertainties on $\langle\rho_{\mu r}^{\mathrm{mix}}/E_R^{\alpha}\rangle(E_{Ri})$ are obtained by inserting each interpolated function in the convolution of Eq. (7.1).

7.1.4 Computation of a z-scale

According to the Heitler-Matthews model (see Sec. 2.2), the muon content N_μ of an air shower grows with the primary mass number A and with the primary energy E following

$$\ln N_\mu = (1-\beta)\ln A + \beta \ln(E/\xi_c), \tag{7.13}$$

being β the power-law index ($\beta \approx 0.9$), and ξ_c a critical energy constant [22].

As mentioned in Sec. 2.5, a muon scale to quantify the muon content measured by different experiments was defined in Ref. [5] as

$$z := \frac{\ln N_{\mu,\,\mathrm{data}}^{\mathrm{det}} - \ln N_{\mu,\,\mathrm{p}}^{\mathrm{det}}}{\ln N_{\mu,\,\mathrm{Fe}}^{\mathrm{det}} - \ln N_{\mu,\,\mathrm{p}}^{\mathrm{det}}}, \tag{7.14}$$

where the suffix "det" indicates that the proton and iron muon contents derive from full-detector simulations, which is useful to cancel first order detector effects. This muon scale has desirable properties: (1) The differences in the logarithms are, to first order, independent of the experimental conditions in which the measurements were taken (distance to the shower axis, zenith angle, etc...), at least to the extent to which the air shower simulations reproduce them; (2) It is independent of the energy to first order (the remaining energy dependence is only through the composition); (3) It should range between 0 and 1, being 0 for proton-like data and 1 for iron-like data, given that simulations correctly reproduce the muon content of real showers; (4) If the mass number A is known event-by-event, the Heitler-Matthews model predicts a value of z of $\ln A / \ln 56$. It is useful to add that z can be equivalently computed from the muon densities instead of muon numbers mutatis mutandis.

Furthermore, to evaluate the muon deficit, we need to consider a reference z-value

$$z_{\text{mass}} := \frac{\ln N^{\text{det}}_{\mu,\,\text{mass}} - \ln N^{\text{det}}_{\mu,\,\text{p}}}{\ln N^{\text{det}}_{\mu,\,\text{Fe}} - \ln N^{\text{det}}_{\mu,\,\text{p}}}, \tag{7.15}$$

where $N^{\text{det}}_{\mu,\,\text{mass}}$ represents the muon content from detector simulations, evaluated using the mass fractions of a given composition model.

To build an estimator of the muon scale z, it is necessary to introduce the mean within a given energy bin. We define an estimator of the muon scale by taking the logarithm of the mean of the muon number,

$$z_{\ln\langle\cdot\rangle} = \frac{\ln\langle N^{\text{det}}_{\mu,\,\text{data}}\rangle - \ln\langle N^{\text{det}}_{\mu,\,\text{p}}\rangle}{\ln\langle N^{\text{det}}_{\mu,\,\text{Fe}}\rangle - \ln\langle N^{\text{det}}_{\mu,\,\text{p}}\rangle}. \tag{7.16}$$

For this analysis we take the muon densities at $r \sim 450\,\text{m}$ from the shower axis, for events with zenith angles in $0^\circ \leq \theta < 45^\circ$ [2].

As a reference with which to compare the aforementioned z-values, we define the estimator of z_{mass} as

$$z_{\langle\ln\cdot\rangle\,\text{mass}} = \frac{\langle\ln N^{\text{det}}_{\mu,\,\text{mass}}\rangle - \langle\ln N^{\text{det}}_{\mu,\,\text{p}}\rangle}{\langle\ln N^{\text{det}}_{\mu,\,\text{Fe}}\rangle - \langle\ln N^{\text{det}}_{\mu,\,\text{p}}\rangle}. \tag{7.17}$$

Notice that in this case we use the mean of the logarithm instead of the logarithm of the mean. We delay the discussion of the effect of this difference to Sec. 8.2.4.

Moreover, an alternative expression to Eq. (7.17), which does not require to compute detector simulations, can be obtained using the Heitler-Matthews model. For this we take the average on both sides of Eq. (7.13), and evaluate it for proton, iron, and a mixed composition scenario. Replacing these expressions in $z_{\langle\ln\cdot\rangle\,\text{mass}}$ we obtain

$$z^{\text{HM}}_{\langle\ln\cdot\rangle\,\text{mass}} = \frac{\langle\ln A\rangle}{\ln 56}. \tag{7.18}$$

Finally, if the measured z-values follow z_{mass}, then the simulations have no muon deficit. Therefore, a scale of the muon deficit in simulations, Δz, is defined as

$$\Delta z := z - z_{\text{mass}}. \tag{7.19}$$

We delay a detailed discussion and analysis on alternative ways to compute z-values to Chapter 8. An alternative estimation of the z-values of the UMD SiPM data can be found in Appendix B.2.

---

[2] For this analysis we do not perform a constant intensity cut to obtain the equivalent muon densities at a given zenith angle, because $\rho_{\mu\,450}$ is almost independent of the zenith angle (see Sec. 6.6). Nevertheless, the results could be improved by including the aforementioned method [97].

7.2 Results

In this section we compare the muon content of air showers in data against proton, iron, and mixed composition scenarios of the three high-energy hadronic interaction models considered. We analyze the dependence with the distance to the shower axis and with the energy. Finally, we show the computed z-values.

7.2.1 Muon content as a function of the distance

The muon density covers several orders of magnitude as it decreases with the distance to the shower axis. For that reason, we take a reference muon density to facilitate the visualization of the comparison of the dependence with the distance to the shower axis. Figure 7.2 shows the average muon density divided by the average muon density of EPOS-LHC proton air showers as a function of the distance to the shower axis, for data, proton, iron, and mixed composition scenarios, at $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.52$. The results for other energies are included in Appendix B.1. The muon densities from simulations and the mixed composition scenarios are computed from Eq. (7.1) using $\alpha = 0$. The left panels show the results for more vertical zenith angles ($\theta < 30°$), while the right panels show the results for more inclined zenith angles ($30° \leq \theta < 45°$). The two upper panels show the comparison of data against EPOS-LHC muon densities, the central panels against QGSJetII-04, and the bottom panels against Sibyll2.3c. The error bars represent the statistical uncertainties, while the brackets represent the systematic uncertainties. In the case of single-primary muon densities the uncertainties are contained within the markers.

From the three right panels of Fig. 7.2 it is noticeable that the muon densities from data at large zenith angles and short distances to the shower axis fall below the proton predictions for the three considered models, although the data are mostly still consistent with simulations within uncertainties. At short distances from the shower axis, the largest source of statistical uncertainty is the propagated uncertainty in the position of the core. Additionally, there might actually be another source of uncertainty not taken into account, caused by the lack of correction of asymmetries in the data reconstruction. In inclined events, the shower core tends to be reconstructed systematically more upstream. This leads to biases in the reconstructed impinging point of the shower core. At shorter distances from the shower axis, the impact of this bias in the relative error of the reconstructed distances is larger. Estimating the introduced (systematic) uncertainty is far from trivial, but, as already mentioned, it is only relevant at short distances from the shower axis [132]. We observe this effect only at low energies, like in this case, where it is possible to measure muon densities close to the shower core without saturation. We restrict the rest of the analysis to distances above $\log_{10}(r/\mathrm{m}) > 2.40$, where, from the right panels of Fig. 7.2, we observe that this bias has no significant effect.

Furthermore, in Fig. 7.2, the slope of the muon LDF is not well reproduced at small zenith angles, but it is approximately reproduced at large zenith angles (for $\log_{10}(r/\mathrm{m}) > 1.75$) by EPOS-LHC and Sibyll2.3c. This is approximately the trend at all energies. At higher energies ($\log_{10}(E_R/\mathrm{eV}) \geq 18.12$) it is also noticeable that the muon densities of data are generally larger than those of the mixed composition scenarios for all models (see Appendix B.1).

From Fig. 7.2 we can conclude that the difference between simulations and data cannot be solved by correcting the normalization of simulations with a unique factor. There is an evident dependence with distance to the shower axis and with zenith angle. In the next section, we show that there is also a clear trend with energy.

7.2.2 Muon content as a function of the energy

Because the muon density grows almost linearly with energy, it increases considerably within the analyzed energy range ($10^{17.32}$ eV to $10^{18.32}$ eV). For that reason, we compare the average muon density divided by the reconstructed energy, which in the case of simulations and the mixed composition scenarios is computed from Eq. (7.1) using $\alpha = 1$. Figures 7.3, 7.4, and 7.5 show the average muon density divided by the reconstructed energy, as a function of the center of the reconstructed energy bin, for data, proton, iron, and mixed composition scenarios, at different distances from the shower axis: $r \sim 280\,\mathrm{m}$ in Fig. 7.3, $r \sim 450\,\mathrm{m}$ in Fig. 7.4, and $r \sim 710\,\mathrm{m}$ in Fig. 7.5.

The Figs. 7.3 to 7.5 show that the muon content of data is consistently above the predictions from the mixed composition scenarios, although mostly not significantly. In most cases the muon content of data is compatible with a heavier composition. This difference between data and the mixed composition scenarios is larger closer to the shower core, where most muons are produced by the high-energy hadronic interactions. The difference is slightly larger at smaller zenith angles, but the trend is not as clear as it is with the energy or distance to the shower axis. The difference is also smaller if the energy scale is increased considering its systematic uncertainties. It is also noticeable that, of the three analyzed models, QGSJetII-04 shows the largest discrepancies between data and the mixed composition predictions.

To further analyze the compatibility between data and simulations, we quantify the average muon deficit as a function of the reconstructed energy by calculating a correction factor F, defined as the ratio between the experimental average muon density divided by the reconstructed energy and the one obtained from simulations,

$$F = \frac{\langle \rho_{\mu\,r}^{\mathrm{data}} / E_R \rangle}{\langle \rho_{\mu\,r}^{\mathrm{S}} / E_R \rangle}, \tag{7.20}$$

where S denotes the scenario under analysis, i.e. S={mix, p, Fe}. We derive the uncertainties in F by propagating the uncertainties of $\langle \rho_{\mu\,r}^{\mathrm{data}} / E_R \rangle$ and $\langle \rho_{\mu\,r}^{\mathrm{S}} / E_R \rangle$. Figures 7.6, 7.7, and 7.8 show the values of the correction factor F, for the three considered models, and for the single nuclei and mixed composition scenarios, as a function of the logarithm of the reconstructed energy bin center. The statistical uncertainties of F are represented with error bars while its systematic uncertainties with brackets. The reconstructed energy bins in which the correction factor F is larger than 1 by more than 1σ (considering total uncertainties) are highlighted in gray and detail the significance.

The Figs. 7.6 to 7.8 show more clearly what was already seen in Figs. 7.3 to 7.5. The correction factors F are all above 1 for the mixed composition scenarios, although in most cases not significantly for EPOS-LHC and Sibyll2.3c. There is only in some cases a significant difference at higher energies for the aforementioned models. In contrast, QGSJetII-04 has in most cases correction factors $F > 1$ with a significance above 1σ. Although the values of F are larger for the larger zenith angles, there is not a significant difference. As stated in Ref. [123] (and references therein), it can be assumed that the composition model derived from X_{max} measurements are subject to smaller systematic uncertainties introduced by the hadronic interaction models. Hence, the observed incompatibilities (where found) can be interpreted as evidence of a muon deficit in air-shower simulations [3].

---

[3] Alternatively, if air-shower simulations had a deeper X_{max}, the implied composition would be heavier, and that would help alliviate the tension between data and simulations with respect to the number of muons [133].

7.2.3 Computed z-values

Figure 7.9 shows the z-values computed from the UMD SiPM data, as explained in Sec. 7.1.4, for the three considered models, as a function of the logarithm of the reconstructed energy bin center. The z-values are computed from the average muon densities at 450 m from the shower axis, measured on the shower plane, and for events with a zenith angle between $0^\circ \leq \theta < 45^\circ$. The gray shaded area corresponds to the predictions of the z-values using the mass fractions of the fits to the Pierre Auger X_{max} distributions [57], while the gray dashed lines use the mass fractions of the GSF composition model [61]. Gray horizontal lines mark $z = 0$ and $z = 1$, the expected values for a pure proton and pure iron composition respectively. The systematic uncertainties of the energy scale are not depicted here. We can see that the z-values of data are compatible within uncertainties with the predictions of both composition models at lower energies, but that there is a difference that grows with energy between the z-values from data and those expected from the mixed composition scenarios. This is consistent with what is shown in Fig. 7.7 for the mass composition model of the Pierre Auger Collaboration. Like in Fig. 7.7, the difference is the largest for QGSJetII-04, followed by Sibyll2.3c, and the smallest is for EPOS-LHC.

Considering total uncertainties, the z-values of the two highest energy bins (centered at $\log_{10}(E_{Ri}/\text{eV})$=18.12 and $\log_{10}(E_{Ri}/\text{eV}) = 18.32$) do not overlap the expected values from the GSF composition model, for the three high-energy hadronic interaction models. In contrast, when considering the composition model derived from the fits to the Pierre Auger X_{max} distributions, only the z-value at $\log_{10}(E_{Ri}/\text{eV}) = 18.32$ for QGSJetII-04 is found incompatible. In all other cases, the computed z-values are compatible within total uncertainties with both composition models.

As mentioned before, an alternative estimation of the z-values of the UMD SiPM data can be found in Appendix B.2.

7.2.4 Comparison to other Pierre Auger z-values

Figure 7.10 shows a comparison of the z-values of different analyses of Pierre Auger data. In this case, we only show the models EPOS-LHC and QGSJetII-04 since the other analyses do not consider Sibyll2.3c. The z-values calculated in this work are shown in full purple stars. Orange crosses show the values computed in an analysis of UMD PMT data (see Ref. [5] and references therein). We show in an olive octagon the value computed from hybrid measurements between the SD-750 array and the Heat-Coihueco fluorescence telescopes, extracted from Ref. [15] [4]. Although the latter has no computed error bars, it is a good reference for comparison as it is in the same energy range. Finally, the red circles show the Pierre Auger results at higher energies, from two different analyses: the first one is an analysis on hybrid data of the SD-1500 array and the fluorescence detectors, while the second one is an analysis of SD-1500 data at very inclined angles ($62^\circ \leq \theta \leq 80^\circ$), (see Ref. [5] and references therein).

We can see from Fig. 7.10 that the z-values computed in this work are in tension with those of the UMD PMT data (specially considering the common systematic uncertainties), but are compatible with those of hybrid infill data within uncertainties. The points computed from the UMD PMT data lie above those of this work. As shown in Sec. 6.8, the muon densities of the UMD PMT data are larger than the ones of this work by 17 % to 28 %. Such difference in the muon densities explains most of the difference between the two sets of z-values. As mentioned there, the cause of the differences between the estimated muon densities is not completely understood, and a deeper investigation should be undertaken.

---

[4] In Ref. [15], two types of z-values are computed considering different energy scale factors. We take for comparison the points compatible with the energy rescaling factor of this work, $f_{\text{Auger}} = 1.052$.

It is relevant to add that we delay to Chapter 8 the comparison of these z-values to the ones computed from AGASA data in this work, as well as to those of other experiments [5].

7.3 Summary and conclusions

The objective of this chapter was to analyze the compatibility between the muon content in UMD SiPM data and single proton, single iron, and mixed composition scenarios computed from three different high-energy hadronic interaction models: EPOS-LHC, QGSJetII-04, and Sibyll2.3c. We used two different composition models, one derived from the fits to the Telescope Array [131] and Pierre Auger [57] X_{max} distributions, and another one derived from the mass fractions of the Global Spline Fit composition model [61].

As first step, we showed a method to analytically compute the effects of the reconstruction in the average muon density (divided, or not, by the reconstructed energy), and by binning in reconstructed energy. The detector effects proved to be mostly small (between $-5\,\%$ and $+14\,\%$, the latter only at the lowest energies), due to the very good energy resolution of the Pierre Auger Observatory. Furthermore, we used the reference energy scale introduced by the *Spectrum Working Group* [122]. We also showed how to compute the so-called z-values from the UMD data.

We compared the average muon density as a function of the logarithmic distance to the shower axis between data and single proton, single iron, and mixed composition scenarios. We observed that the slope of the muon LDF is not well reproduced at small zenith angles, but that it is approximately reproduced by EPOS-LHC and Sibyll2.3c at large zenith angles (and for $\log_{10}(r/\mathrm{m}) > 1.75$). This showed that introducing a unique normalization correcting factor would not solve the difference between data and simulations.

Furthermore, we compared the average muon density divided by the reconstructed energy as a function of the logarithm of the reconstructed energy bin center between data and the single proton, single iron, and mixed composition scenarios. We performed the comparison for three distances to the shower axis, $r \sim 280\,\mathrm{m}$, $r \sim 450\,\mathrm{m}$, and $r \sim 710\,\mathrm{m}$, and for two zenith angle bins ($0^\circ \leq \theta < 30^\circ$ and $30^\circ \leq \theta < 45^\circ$). We found the UMD data in most cases compatible with a pure iron composition considering total uncertainties, and consistently above the predictions from the mixed composition scenarios, but in most cases compatible within uncertainties. This difference is larger if the energy scale is decreased, the larger the energy, the smaller the distance to the shower core, the smaller the zenith angles, and for the model QGSJetII-04.

Finally, we computed the z-values from the UMD data using the average muon densities at $r \sim 450\,\mathrm{m}$ and taking the whole zenith angle range ($0^\circ \leq \theta < 45^\circ$). Similarly, we observe that the difference between the z-values from data and those of the mixed composition scenarios increases with the reconstructed energy, but that in most cases the difference is not significant. We did find incompatibilities for all models with the GSF composition model in the two highest-energy bins ($\log_{10}(E_{Ri}/\mathrm{eV}) = 18.12$ and $\log_{10}(E_{Ri}/\mathrm{eV}) = 18.32$), while only the highest-energy z-value of QGSJetII-04 is incompatible with the composition model derived from the fits to the Pierre Auger X_{max} distributions. These incompatibilities between data and the mixed composition scenarios are interpreted as evidence of a muon deficit in air-shower simulations. Finally, we compared the z-values computed in this work with those of other analyses of Pierre Auger data. We found our z-values in tension with those of a UMD PMT analysis (see Ref. [5] and references therein), but compatible with those of a hybrid infill analysis [15]. In Chapter 8 we compare these z-values to those computed from AGASA data at higher energies, as well as to those of other experiments.

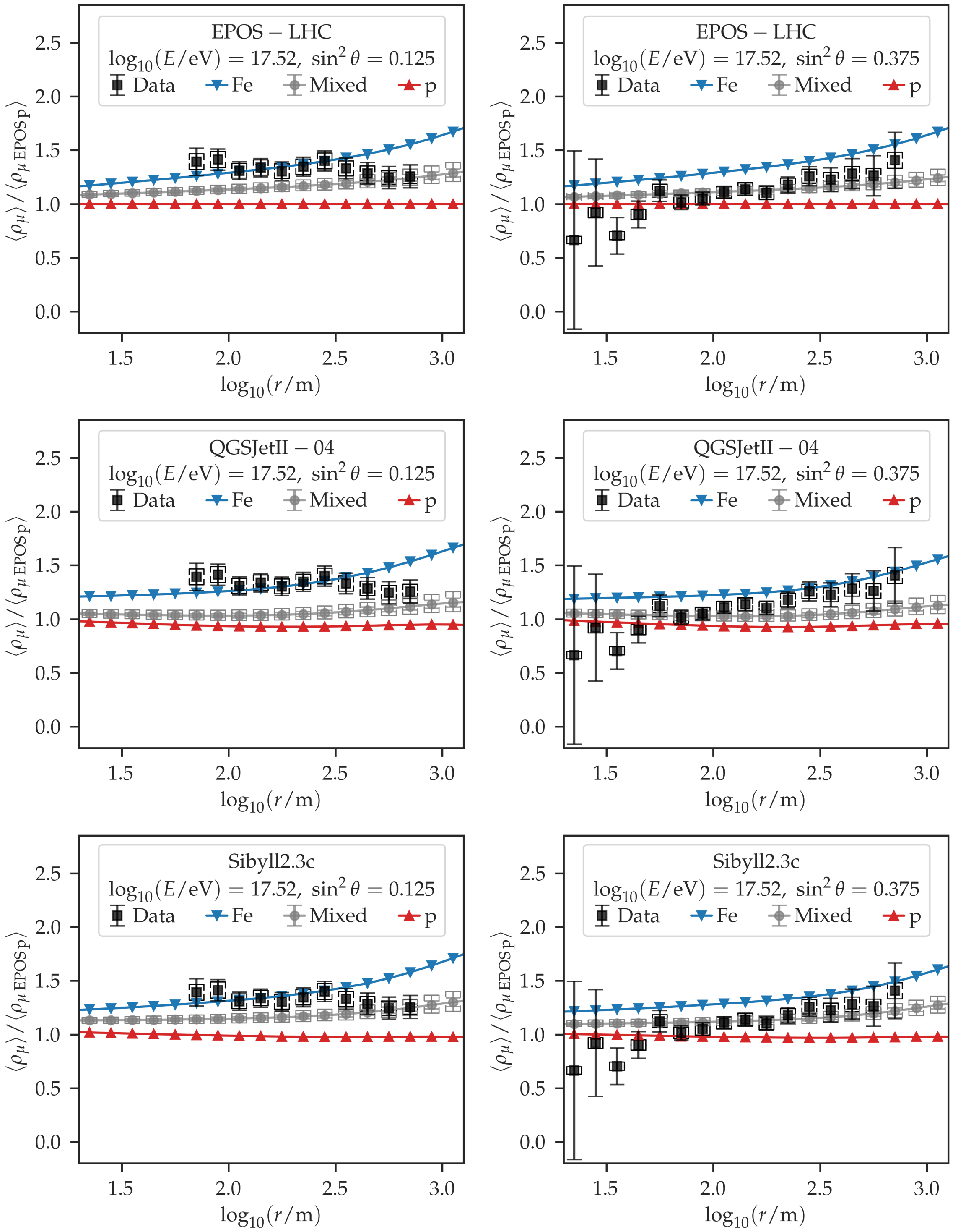


Figure 7.2: Average muon density divided by the average muon density of EPOS-LHC proton air showers as a function of the distance to the shower axis, for the reconstructed energy bin centered at $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.52$ with a bin width of $\Delta\log_{10}(E_R/\mathrm{eV}) = 0.2$. Square markers represent data, down triangles iron, circles represent the mixed composition scenarios, and up triangles proton. Error bars represent statistical uncertainties and brackets represent systematic uncertainties. These are contained within the markers in the case of proton and iron simulated densities. In the two upper panels the simulated muon densities and those of the mixed composition scenarios correspond to the model EPOS-LHC, in the central panels to QGSJetII-04, and in the bottom panels to Sibyll2.3c. The three left panels show the comparison for the smaller zenith angles ($\theta < 30°$), and the three right panels for the larger zenith angles ($30° \leq \theta < 45°$).

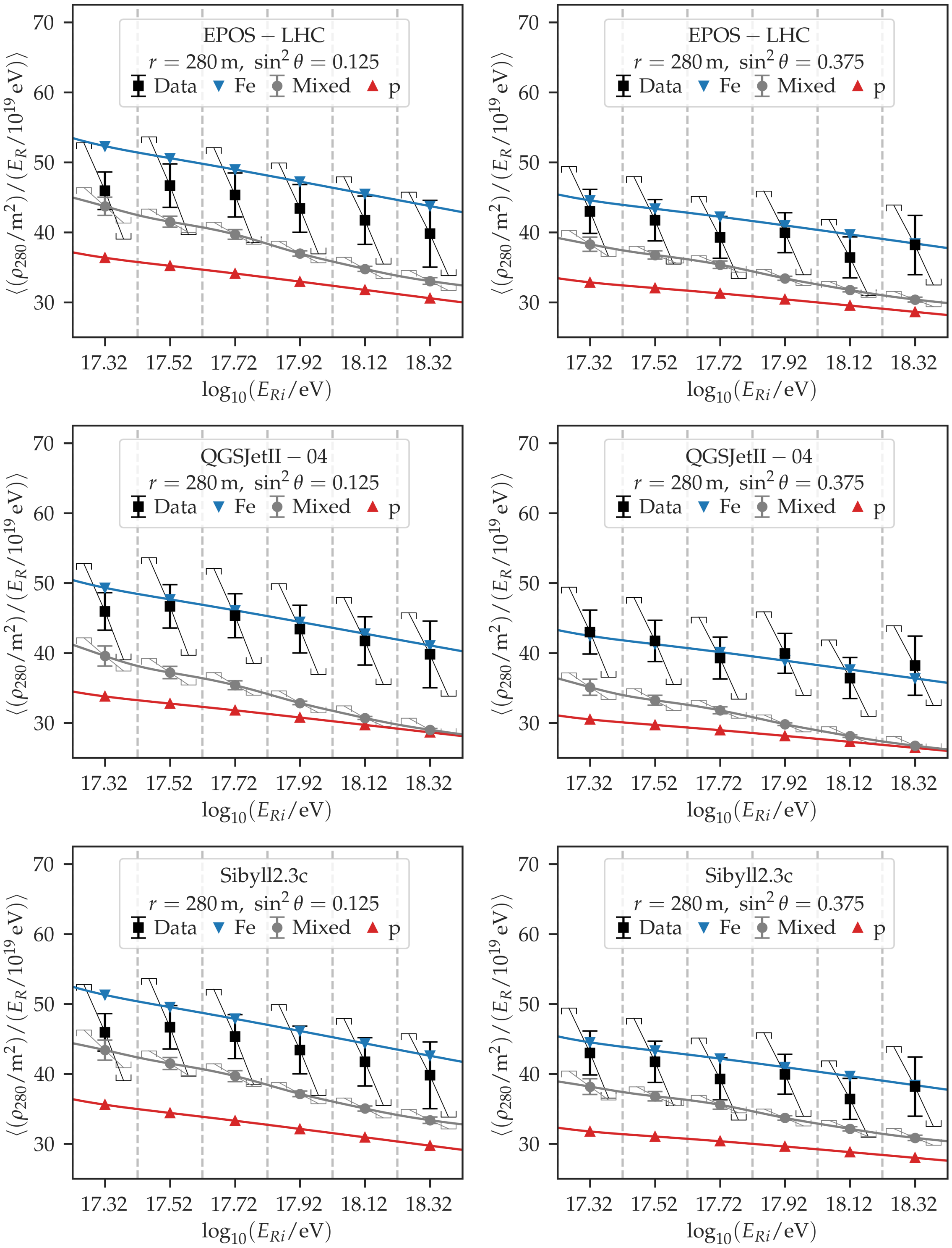


Figure 7.3: Average muon density divided by the reconstructed energy, as a function of the center of the reconstructed energy bin, for $2.40 \leq \log_{10}(r/\mathrm{m}) < 2.50$ ($r \sim 280\,\mathrm{m}$). The considered bin width is $\Delta \log_{10}(E_R/\mathrm{eV}) = 0.2$. Square markers represent data, down triangles iron, circles represent the mixed composition scenarios, and up triangles proton. Error bars represent statistical uncertainties and brackets represent systematic uncertainties. These are contained within the markers in the case of proton and iron simulated densities. In the two upper panels the simulated muon densities and those of the mixed composition scenarios correspond to the model EPOS-LHC, in the central panels to QGSJetII-04, and in the bottom panels to Sibyll2.3c. The three left panels show the comparison for the smaller zenith angles ($\theta < 30°$), and the three right panels for the larger zenith angles ($30° \leq \theta < 45°$).

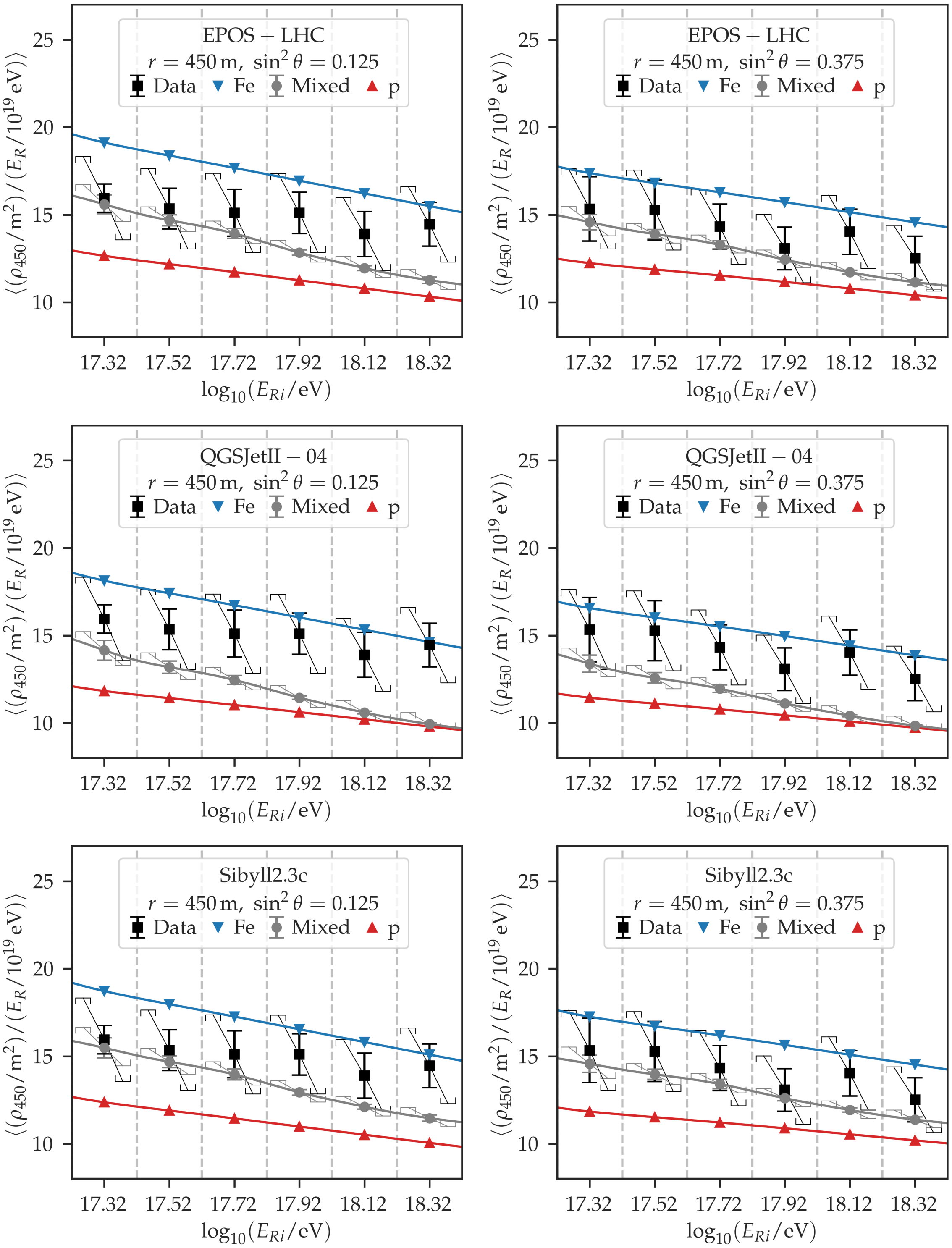


Figure 7.4: Same as Fig. 7.3, but for $2.60 \leq \log_{10}(r/\mathrm{m}) < 2.70$ ($r \sim 450\,\mathrm{m}$).

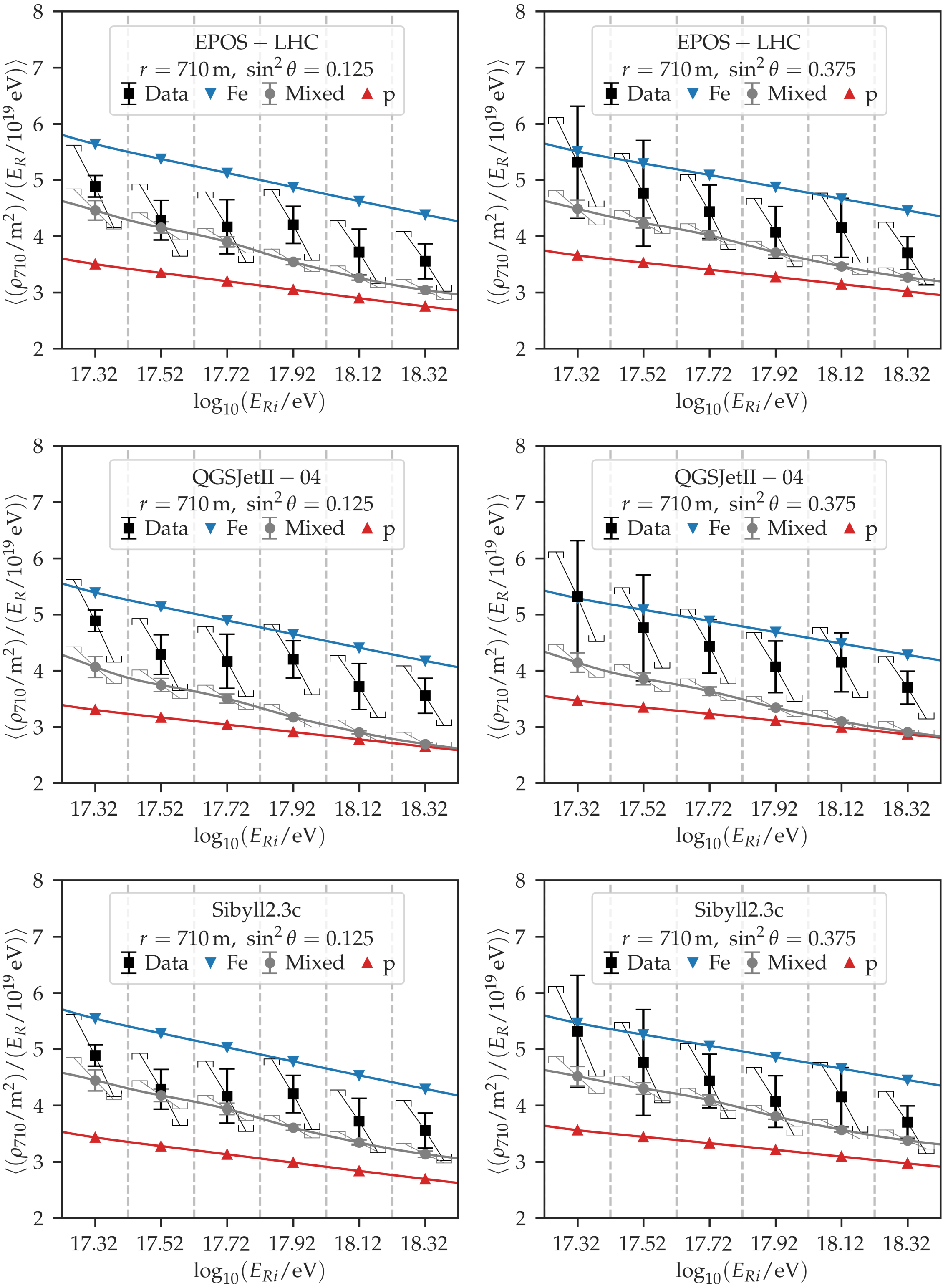


Figure 7.5: Same as Fig. 7.3, but for $2.80 \leq \log_{10}(r/\mathrm{m}) < 2.90$ ($r \sim 710\,\mathrm{m}$).

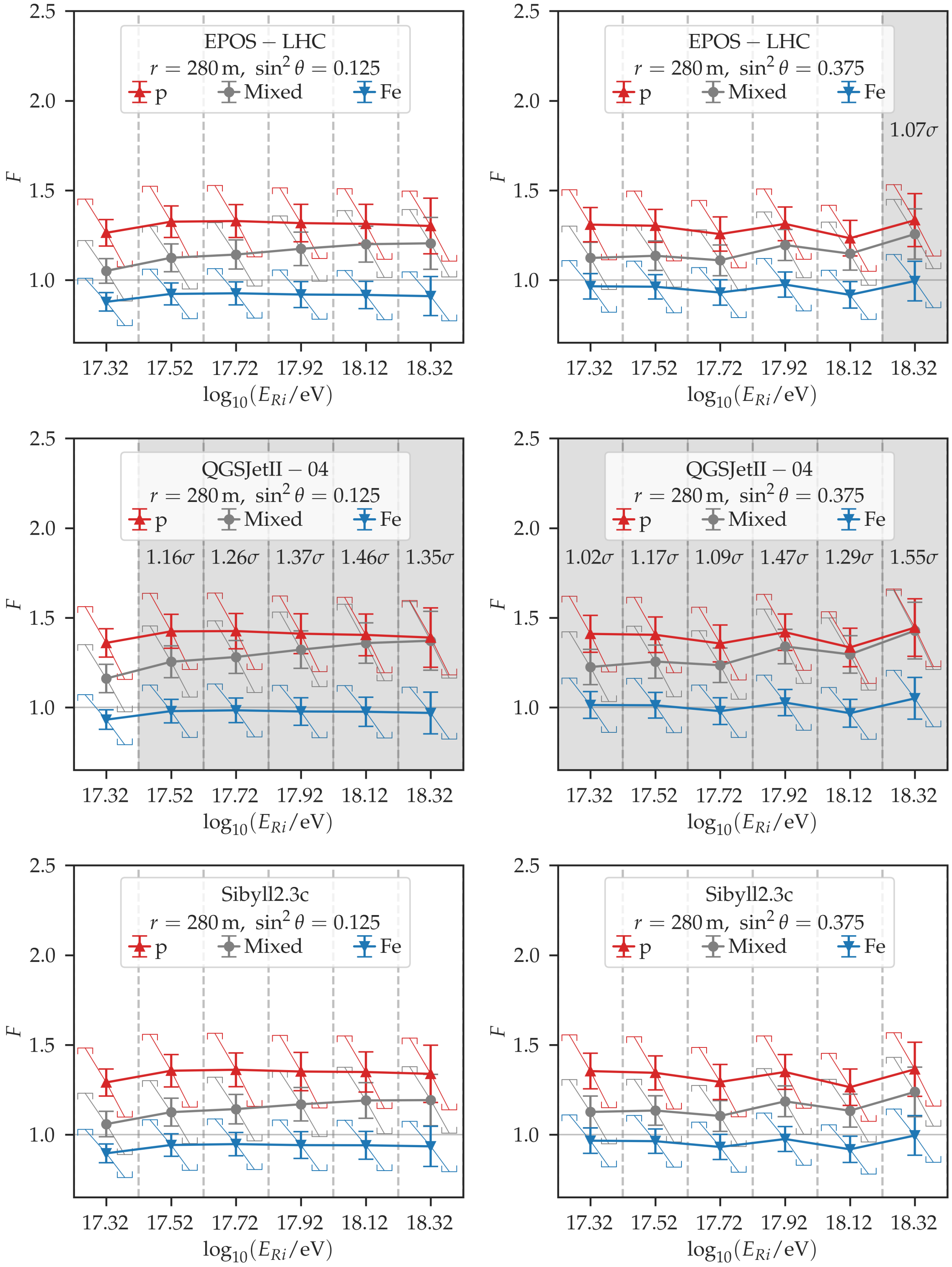


Figure 7.6: Correction factor F (see Eq. (7.20)), as a function of the center of the reconstructed energy bin, for $2.40 \leq \log_{10}(r/\mathrm{m}) < 2.50$ ($r \sim 280\,\mathrm{m}$), considering a bin width of $\Delta \log_{10}(E_R/\mathrm{eV}) = 0.2$. Down triangles represent iron, circles represent the mixed composition scenarios, and up triangles represent proton. Statistical uncertainties are shown with error bars and systematic uncertainties with brackets. In the two upper panels the model is EPOS-LHC, in the central panels it is QGSJetII-04, and in the bottom panels it is Sibyll2.3c. The three left panels show the comparison for the smaller zenith angles ($\theta < 30°$), and the three right panels for the larger zenith angles ($30° \leq \theta < 45°$). Highlighted in gray are the reconstructed energy bins where the correction factor F, that makes the mixed composition scenario match data, is larger than 1 with a significance of more than 1σ; in these cases the significance is detailed in the same bin. A gray horizontal line marks $F = 1$.

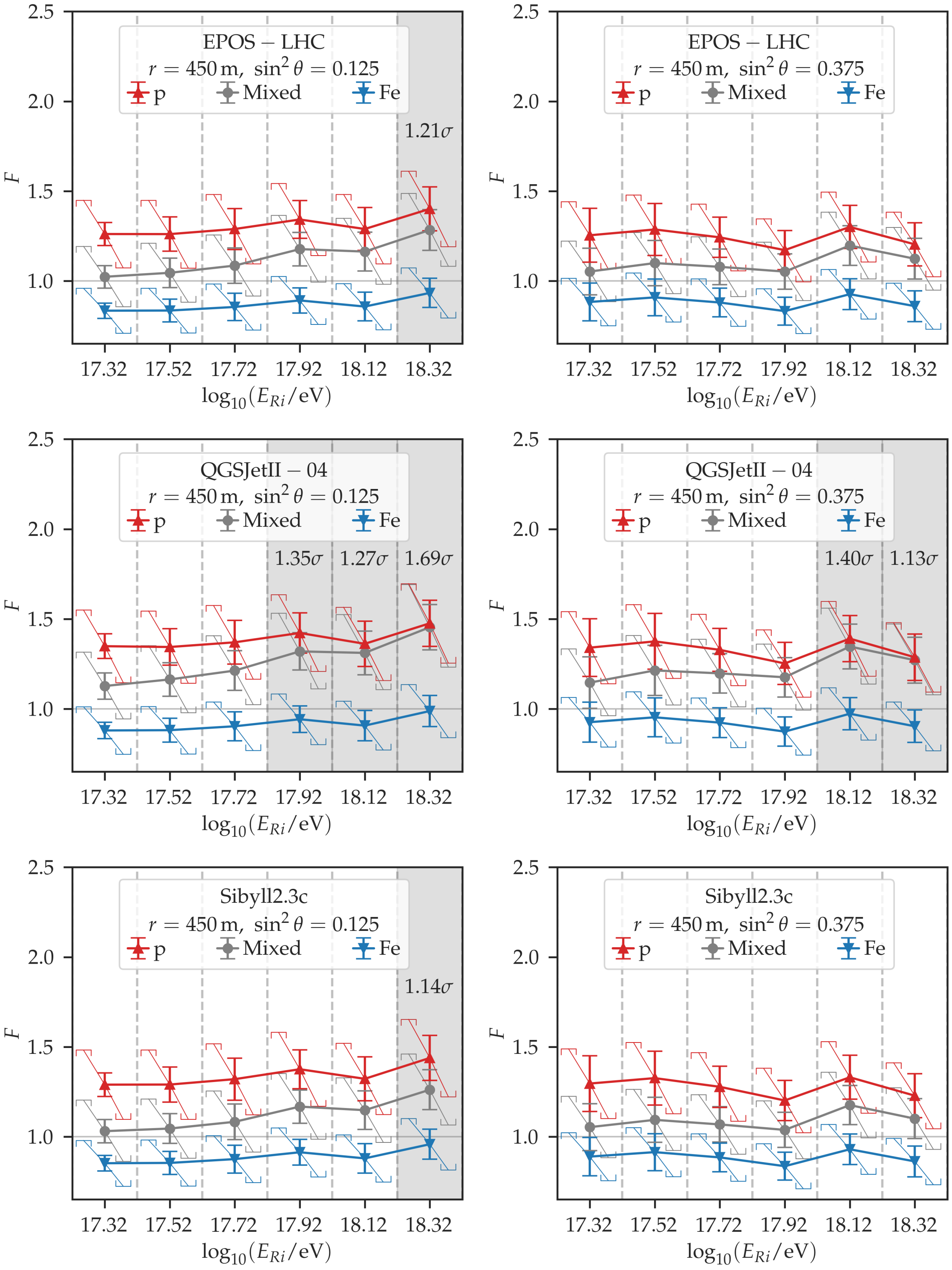


Figure 7.7: Same as Fig. 7.6, but for $2.60 \leq \log_{10}(r/\mathrm{m}) < 2.70$ ($r \sim 450\,\mathrm{m}$).

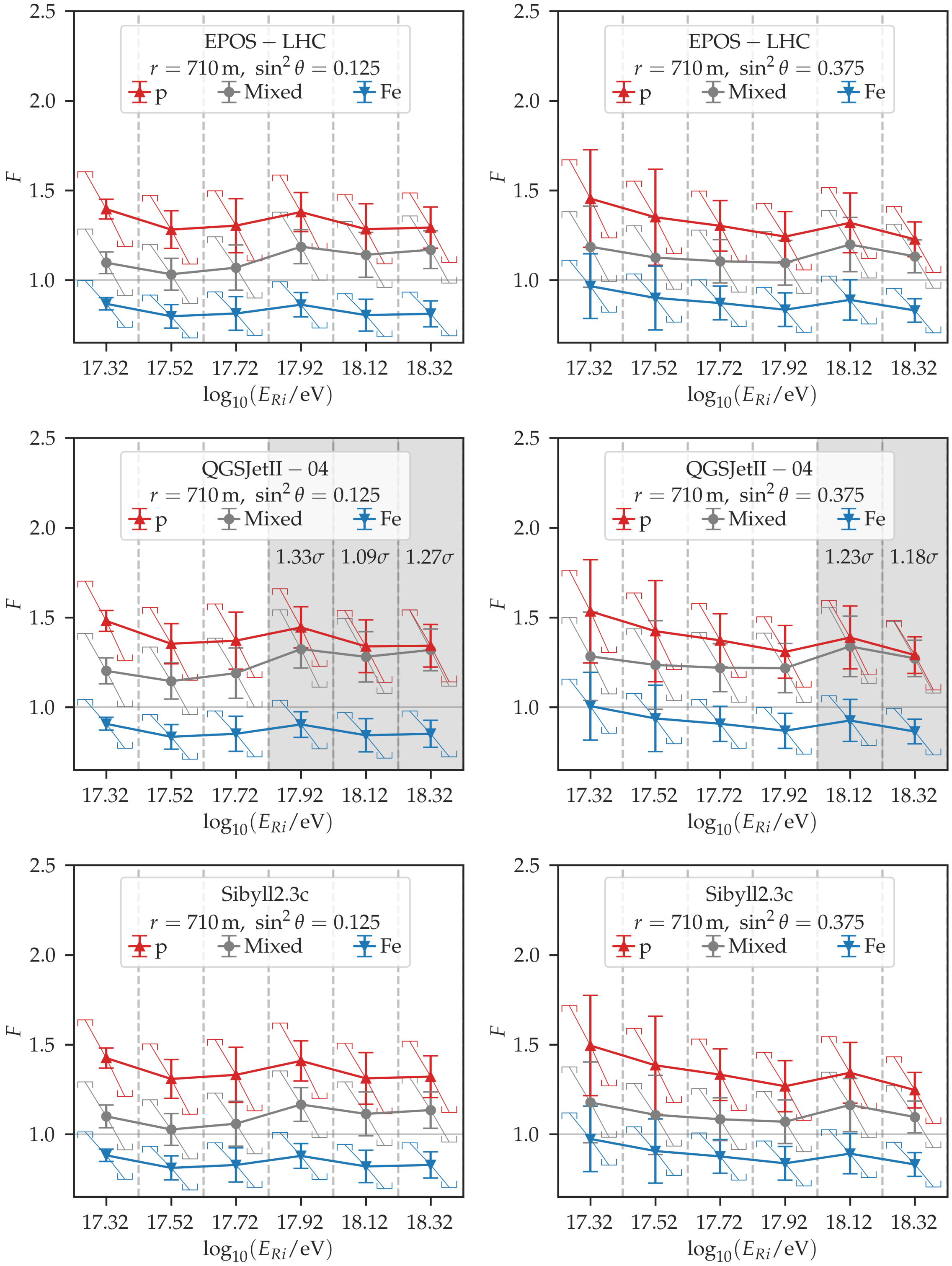


Figure 7.8: Same as Fig. 7.6, but for $2.80 \leq \log_{10}(r/\mathrm{m}) < 2.90$ ($r \sim 710\,\mathrm{m}$).

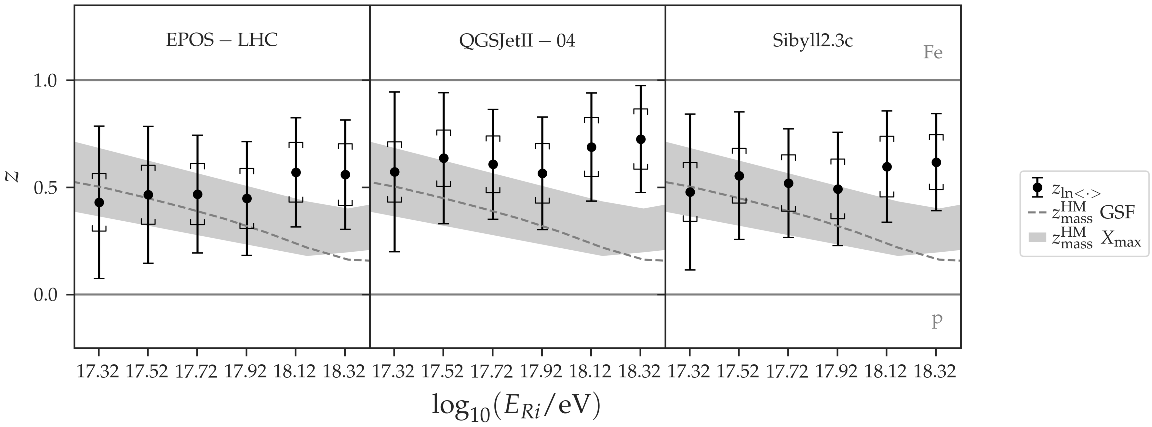


Figure 7.9: Values of z as a function of the logarithm of the center of the reconstructed energy bin for the three considered models. The left panel shows the values computed for EPOS-LHC, the middle panel for QGSJetII-04, and the right panel for Sibyll2.3c. Gray horizontal lines mark the expected values of z for pure proton and pure iron compositions (0 and 1 respectively). Dashed lines shows the z-value expected from mixed composition scenarios considering the mass fractions of the GSF model, while the gray shaded area represents the ones expected from the fits to the Pierre Auger $X_{\max}$ distributions. The computed z-values are shown in circles, with its statistical uncertainties represented with error bars and its systematic uncertainties with brackets.

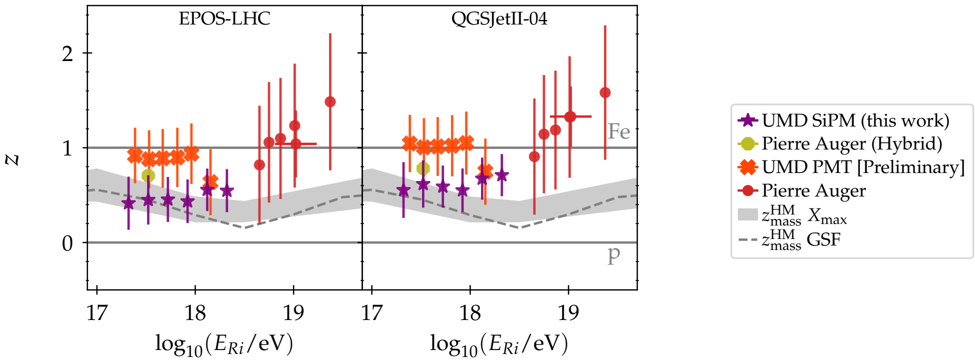


Figure 7.10: Same as Fig. 7.9, but comparing different analyses of Pierre Auger data. The models shown are EPOS-LHC (left panel) and QGSJetII-04 (right panel). The z-values computed in this work are shown in full purple stars, and the error bars represent total uncertainties. The orange crosses correspond to the results of UMD PMT data, while the red circles correspond to analyses of hybrid and inclined data (see Ref. [5] and references therein). The olive octagon corresponds to an analysis of hybrid data in the infill [15].

Chapter 8

Muon deficit in air-shower simulations estimated from AGASA data

In Chapter 7 we compared the muon content in UMD SiPM data of the Pierre Auger Observatory and in air-shower simulations at reconstructed energies in the range $17.22 \leq \log_{10}(E_R/\mathrm{eV}) \leq 18.42$. It is of our interest to compare the muon content between data and simulations also at higher energies. Since the UMD was not designed to measure higher energy events given its spacing and exposure, it does not provide enough statistics above the aforementioned energy.

For this reason, in this chapter we analyze direct measurements of the muon densities at 1000 m from the shower axis performed by the Akeno Giant Air Shower Array (AGASA). The AGASA data set is valuable and unique due to its ultra-high energies, only reached by a couple of other experiments (the Pierre Auger Observatory, Telescope Array, and Yakutsk Array). At the aforementioned energies, it is the only experiment that provides a direct detection of muons, while simultaneously measuring the primary energy. The aforementioned data set consists of events with zenith angles $\theta \leq 36^\circ$ and reconstructed energies in the range $18.83 \leq \log_{10}(E_R/\mathrm{eV}) \leq 19.46$. We compare the muon content of these data to that of air-shower simulations of proton, iron, and mixed composition scenarios, for three high-energy hadronic interaction models: EPOS-LHC, QGSJetII-04, and Sibyll2.3c. The methods used in this chapter are analogous to those of Chapter 7.

The structure of this chapter is as follows. We first start by describing AGASA in Sec. 8.1. In Sec. 8.2 we describe the analytical computation of detector effects in the simulated muon densities, the transformation of the AGASA energy scale to the reference energy scale, we describe the used data set, and describe the way to compute the z-values from AGASA data. In particular, we define two different estimators of the muon scale and muon deficit scale, analyze their biases, and provide expressions to compute them for the case of AGASA. In Sec. 8.3 we show the comparisons of the muon content as a function of the reconstructed energy of data to single-proton and single-iron simulations, as well as to mixed composition scenarios. Furthermore, we show the z-values computed from the two different estimators and how they compare to the expected values predicted by the mixed composition models. Moreover, in Sec. 8.4 we show how they compare to the z-values computed from UMD SiPM measurements in this work (c.f. Chapter 7), as well as to those of other experiments. Finally, in Sec. 8.5 we summarize the main conclusions.

This chapter is strongly based on the **publications**:

- *Muon deficit in air shower simulations estimated from AGASA muon measurements*, F. Gesualdi, A. D. Supanitsky, and A. Etchegoyen, Phys. Rev. D 101, 083025, 2020 [123],

- *Muon deficit in simulations of air showers inferred from AGASA data*, F. Gesualdi, A. D. Supanitsky, and A. Etchegoyen, Proc. 37th Int. Cosmic Ray Conf., 2021 [124],

- *On the muon scale of air showers and its application to the AGASA data*, F. Gesualdi, H. Dembinski, K. Shinozaki, A.D. Supanitsky, T. Pierog, L. Cazon, D. Soldin, and R. Conceição on behalf of the Working group on Hadronic Interactions and Shower Physics (WHISP), Proc. 37th Int. Cosmic Ray Conf., 2021 [125].

8.1 The Akeno Giant Air Shower Array

The Akeno Giant Air Shower Array (AGASA) experiment was located in Akeno (Hokuto since 2004), Yamanashi, Japan. It was conformed by an array of 111 scintillation detectors spread across $\sim 100\,\mathrm{km}^2$, as well as by 27 muon detectors. The experiment was able to measure events with energies above 3×10^{16} eV and with zenith angles $\theta \leq 45^\circ$ [45]. The average altitude of the detectors was 667 m above sea level. The detectors were decommissioned in 2004.

Similar to the Pierre Auger Observatory, AGASA counted with arrays of detectors separated by different distances, to cover a wider energy range. The data that we use correspond to the largest array, dubbed A100 as it covered $\sim 100\,\mathrm{km}^2$. Each position of the array was instrumented with a 5 cm-thick scintillation counter with an area of $2.2\,\mathrm{m}^2$. Similar to the surface scintillation detectors (SSDs) or the underground muon detectors (UMD) of the Pierre Auger Observatory, the scintillation detectors had optic fibers and a photomultiplier tube (PMT). The detectors had a broad dynamic range of 0.3 to 5×10^4 particles per detector. Data taking was triggered under the condition that 5 neighbor stations had a signal equal or larger than 0.3 particles within $2.5\,\mu$s [134].

The muon detectors were deployed in the southern part of the array. Each muon detector was formed by 50 proportional counters made of an iron pipe of dimensions $5\,\mathrm{m} \times 10\,\mathrm{cm} \times 10\,\mathrm{cm}$, coated inside with zinc. They were filled with a gas mixture of 90 % argon and 10 % methane at a pressure of 690 mmHg at 15 °C [135]. The muon detectors were shielded with 1 m of concrete or 30 cm of iron, implying a vertical muon energy threshold of 0.5 GeV [45].

Figure 8.1 shows a map of AGASA. The positions of the scintillation detectors can be seen in circles, and those that have a square on top were also instrumented with the muon detectors.

8.2 Analysis

In this section we describe the analytical computation of the detector effects in the simulated muon densities. We also discuss the transformation to the reference energy scale. We describe how we compute the muon density (divided by the reconstructed energy) for pure proton, pure iron, and mixed composition scenarios. Finally, we discuss on two different estimators of a z-value.

8.2.1 Simulating detector effects in the muon density

The way to mimic the detector effects is identical to what is described in Sec. 7.1.1. In particular, the way to obtain the average muon density at a distance r divided by the reconstructed

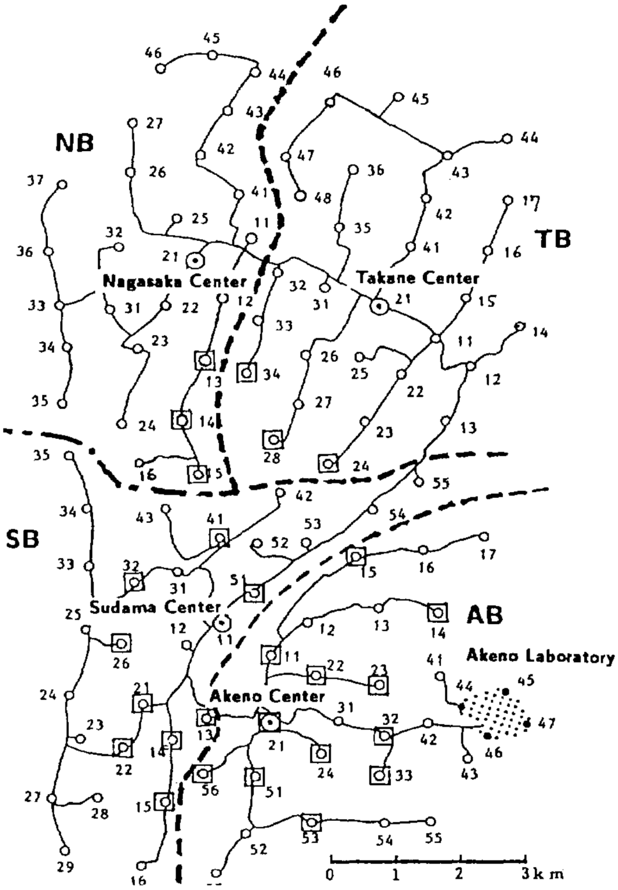


Figure 8.1: Map of the Akeno Giant Air Shower Array. The positions of the stations are marked in circles. Those that count with muon detectors have also a square. Image extracted from Ref. [136].

energy to the power of α, in the i^{th} reconstructed energy bin, $\left\langle \frac{\rho_{\mu\, r}}{E_R^{\alpha}} \right\rangle (E_{Ri})$, is as shown in Eq. (7.1), which we repeat here for clarity [123]:

$$\left\langle \frac{\rho_{\mu\, r}}{E_R^{\alpha}} \right\rangle (E_{Ri}) = \frac{\displaystyle\int_{E_{Ri}^-}^{E_{Ri}^+} \int_0^{\infty} \langle \widetilde{\rho}_{\mu\, r} \rangle (E)\, E_R^{-\alpha}\, J(E)\, G(E_R|E)\, \varepsilon(E) dE\, dE_R}{\displaystyle\int_{E_{Ri}^-}^{E_{Ri}^+} \int_0^{\infty} J(E)\, G(E_R|E)\, \varepsilon(E) dE\, dE_R}, \tag{8.1}$$

where E_{Ri}^- and E_{Ri}^+ are the lower and upper limits of the reconstructed energy bin of center E_{Ri}, $\langle \widetilde{\rho}_{\mu\, r} \rangle (E)$ is the average muon density at distance r from the shower axis (measured on the shower plane) as a function of the Monte Carlo true energy of the simulation, $J(E)$ is the cosmic ray flux, $G(E_R|E)$ is the conditional probability distribution of E_R conditioned to E, and $\varepsilon(E)$ is the trigger efficiency as a function of the Monte Carlo true energy. Naturally, since the AGASA muon detector is different from the UMD of the Pierre Auger Observatory, the difference between the calculations of Sec. 7.1.1 and this section is the way we model the aforementioned functions.

In first place, the simulated average muon densities at 1000 m from the shower axis, $\langle \widetilde{\rho}_{\mu\, 1000} \rangle (E)$, are fitted using a power law in energy (see Eq. (7.2)) for every primary and hadronic interaction model.

Futhermore, the cosmic ray flux (shifted to the reference energy scale as explained in Sec. 8.2.2) is obtained by fitting the Telescope Array measurements [137] with the following function [138]

$$J(E) = A \begin{cases} \left(\frac{E}{E_a}\right)^{-\gamma_1} & \log E \leq \log E_a \\ \left(\frac{E}{E_a}\right)^{-\gamma_2} \frac{1+(E_a/E_s)^{\delta\gamma}}{1+(E/E_s)^{\delta\gamma}} & \log E > \log E_a \end{cases}, \tag{8.2}$$

where A, E_a, E_s, γ_1, γ_2, and $\delta\gamma$ are free fit parameters. The resulting values of the parameters are given in Table 8.1 [123]. It is relevant to add that we also test the systematic uncertainties introduced by choosing a specific flux parameterization by using instead the fit to the flux measurements of the Pierre Auger Observatory as presented in Ref. [138].

Parameter	Fitted Value
A $[10^{-19}\text{eVkm}^2\text{ yr sr}]$	3.5 ± 0.5
$\log_{10}(E_a/\text{eV})$	18.71 ± 0.02
$\log_{10}(E_s/\text{eV})$	19.88 ± 0.09
γ_1	3.248 ± 0.012
γ_2	2.63 ± 0.06
$\delta\gamma$	2.4 ± 0.8

Table 8.1: Parameters of the fit to the UHECR flux measured by Telescope Array (see Eq. (8.2)).

The energy resolution function of AGASA, $G(E_R|E)$, is modelled as a log-normal distribution [139] (see Eq. (7.5)), where the parameter μ_l of the log-normal is $\ln E$, and the parameter σ_l is related to the standard deviation $\sigma(E)$ through,

$$\sigma_l(E) = \sqrt{\ln\left[\frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\,\frac{\sigma^2(E)}{E^2}}\right]}. \tag{8.3}$$

The way to derive Eq. (8.3) is equivalent to that of Eq. (7.9) in Sec. 7.1.1. $\sigma(E)$ is obtained from the signal resolution $\sigma[S_{600}]$ as a function of $\log_{10} S_{600}$ by using the S_{600} to energy conversion function, reported in Ref. [139], corrected to match the reference energy scale as explained in Section 8.2.2

$$E = f_{\text{AGASA}} \times 2.21 \times 10^{17}\, S_0(600)^{1.03}\,\text{eV}. \tag{8.4}$$

The S_{600} resolution as a function of $\log_{10}(S_{600})$, obtained from shower and detector simulations, for showers with zenith angles in $33^\circ \leq \theta \leq 44^\circ$, is taken from Ref. [136]. The resulting values for $\sigma(E)$ are then fitted using a second degree polynomial in $\log_{10}(E/\text{eV})$ given by

$$\sigma(E)/\text{eV} = (17 \pm 3) - (1.59 \pm 0.37)\log_{10}(E_R/\text{eV}) + (0.039 \pm 0.009)\log^2_{10}(E_R/\text{eV}). \tag{8.5}$$

Finally, we assume that the trigger efficiency is ~ 1 within the integration range ($18.0 < \log_{10}(E/\text{eV}) < 19.8$) since we did not find it reported in literature. Although this might not be the most accurate assumption, specially at the lowest end of the integration range, it most likely has a small effect in the analyzed reconstructed energy range ($18.83 \leq \log_{10}(E_R/\text{eV}) \leq 19.46$). It is relevant to add that the integration range is limited to the smallest definition range of all functions, which is that of the air-shower simulation library.

Figure 8.2 shows a comparison between $\langle \widetilde{\rho}_\mu / E \rangle (E)$ and $\langle \rho_\mu / E_R \rangle (E_{Ri})$. We observe that $\langle \rho_\mu / E_R \rangle (E_{Ri})$ can be 11 % to 22 % smaller than $\langle \widetilde{\rho}_\mu / E \rangle (E)$ in the analyzed energy range. At low energies, this difference is explained by the large uncertainty in the reconstructed energy ($\sim$28 % at $10^{18.83}$ eV), while at high energies the dominant effect is the flux suppression. The effect of binning in reconstructed energy with a bin width of $\Delta \log_{10}(E_R/\mathrm{eV}) = 0.2$ is in comparison small.

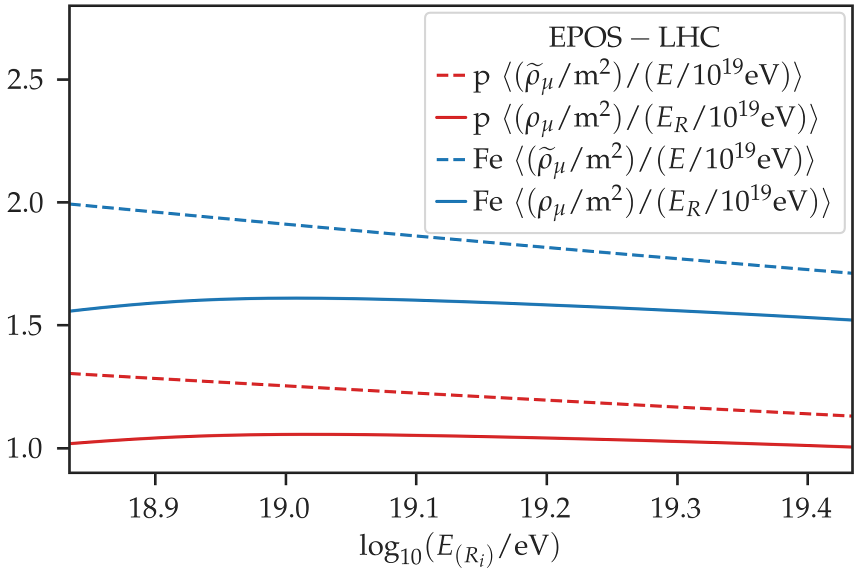


Figure 8.2: Average muon density at 1000 m from the shower axis, divided by the energy, as a function of the logarithm of the Monte Carlo energy E (dashed lines) and average muon density at 1000 m from the shower axis, divided by the reconstructed energy, as a function of the logarithm of the reconstructed energy bin E_{Ri} (solid lines). We used a bin width of $\Delta \log_{10}(E/\mathrm{eV}) = 0.2$. The air-shower simulations use the model EPOS-LHC and the primaries are proton (red) and iron (blue).

8.2.2 Transformation to the reference energy scale

As in Sec. 7.1.2, we work on the reference energy scale [122] to enable the comparison of the AGASA data to that of other experiments, including the UMD of the Pierre Auger Observatory. It is also fundamental since we use data from three different experiments in this analysis: the muon densities as a function of the energy from AGASA, the cosmic ray energy spectrum from Telescope Array, and the mass composition fractions as a function of the primary energy, obtained from the fits to the X_{max} experimental distributions, from the Pierre Auger Observatory.

In Sec. 7.1.2 we presented the energy correction factors f_{TA} of Telescope Array and f_{Auger} of Pierre Auger. In this section we compute the correction factor of AGASA. We remind the reader that the correction factors are found by matching flux measurements under the hypothesis that the cosmic ray flux is isotropic. The correction factor of AGASA is given by,

$$f_{\mathrm{AGASA}} = \frac{E^{\mathrm{ref}}}{E^{\mathrm{TA}}} \times \frac{E^{\mathrm{TA}}}{E^{\mathrm{AGASA}}} = 0.948 \times 0.72 = 0.68, \tag{8.6}$$

where $E^{\mathrm{TA}}/E^{\mathrm{AGASA}}$ is taken from Ref. [140]. The correction factor is considerably large, compared to that of Telescope Array or Pierre Auger. Most of the difference between the

AGASA and the reference energy scale arises from the muon deficit in the simulations used to calibrate the AGASA energy scale. Those simulations made use of older-generation high-energy hadronic interaction models.

The systematics in the reconstructed energy are computed as the sum in quadrature of three contributions. The first one arises from a possible bias in the lateral distribution of muons (from the used empirical function, the exclusion of zero-density data, and the absence of non-hit detectors), and amounts to a $\pm 7\,\%$ systematic in energy [141]. The second contribution arises from the constant intensity cut method in the zenith angle range of this data set ($0^\circ \leq \theta \leq 36^\circ$), and it is estimated to contribute between 17 % to 20 % to the systematic uncertainties in the energy (estimated from Fig. 17 of Ref. [141]). The third contribution comes from the exponent in the energy scale formula, and it contributes between 8 % and 10 %. This exponent is taken from simulations, and different high-energy hadronic interaction models predict different values of it [142]. Finally, there is also a $\sim 10\,\%$ uncertainty associated with the reference energy scale, which we treat separately as it is the same for any experiment in this scale. The energy systematics are added in quadrature with a factor $\beta = 0.9$ (see Eq. (7.2)) to the intrinsic systematics of the muon density, to obtain the total systematics of the muon density. The latter range between 18 % to 21 % without including the $\pm 10\,\%$ systematic uncertainty of the reference energy scale, or between $\pm 22\,\%$ to $\pm 26\,\%$ including the $\pm 10\,\%$ systematic uncertainty of the reference energy scale.

8.2.3 Calculation of the muon density

The muon densities in the AGASA data set are determined using the infinite window strategy (described in Sec. 4.1.1) [134][1]. They report that this is a good estimator if the showers are nearly-vertical (presumably to avoid double counting corner-clipping muons) and if the muon densities are $\lesssim 10\,\mathrm{m}^{-2}$ (such that the number of occupied segments is much smaller than the available ones) [45, 134]. Then, the muon density at 1000 m from the shower axis is determined from the fit of the measurements to a muon lateral distribution function [45]; its uncertainty is reported to be $\sim 40\,\%$ above 10^{19} eV (see Ref. [134] and references therein). The muon density values of the analyzed events are extracted from Fig. 7 of Ref. [134] [2] and are shown in Fig. 8.3. A table with the values of the energy and muon densities can also be found in Ref. [123]. The data set consists of events with zenith angles $\theta \leq 36^\circ$, and, as mentioned before, with a vertical muon energy threshold of 0.5 GeV [134]. The events with no muon detection, enclosed in a rectangle in Fig. 8.3, are included in the analysis. The energy cut at $\log_{10}(E_R/\mathrm{eV}) = 19.46$ is set due to the sharp drop in statistics beyond that energy.

Regarding simulations, we generated a library of proton, helium, nitrogen and iron initiated air showers, using the high-energy hadronic interaction models QGSJetII-04, EPOS-LHC and Sibyll2.3c, and the low-energy hadronic interaction model Fluka version 2011.2x [30, 31]. For each model and primary type, we simulated ~ 20 showers (~ 30 for proton primaries) per input energy, in the energy range $18.0 \leq \log_{10}(E/\mathrm{eV}) \leq 19.8$ and in steps of $\Delta \log_{10}(E/\mathrm{eV}) = 0.2$. A larger number of proton-initiated showers (with respect to iron-initiated showers) are simulated because shower-to-shower fluctuations are larger for lighter primaries. We also simulated additional showers for proton and iron primaries of the models QGSJetII-04 and EPOS-LHC extending the energy range to $19.8 \leq \log_{10}(E/\mathrm{eV}) \leq 20.8$ to validate the performance of the integral in Eq. (8.1) in a finite energy range. The altitude used for the simulations is the average altitude of the detectors, 667 m [139]. We set the x

---

[1] The AGASA collaboration refers to this way of computing the muon density as the "on-off density" [134].

[2] Previous versions of this data set can be found in Refs. [143, 144].

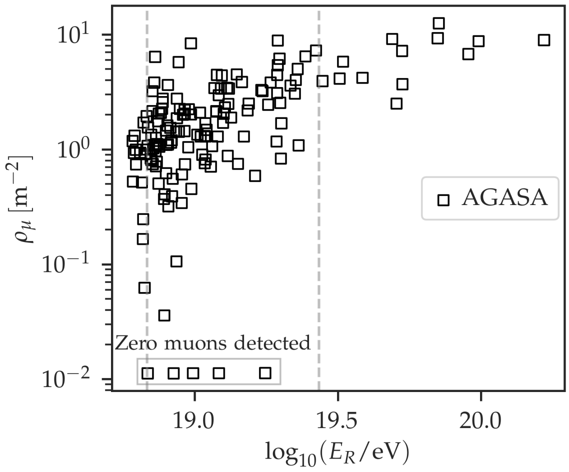


Figure 8.3: Muon density (in logarithmic scale) as a function of the logarithm of the reconstructed energy (in the reference scale). The gray dashed lines mark the analyzed interval of energies. For the events enclosed by the rectangle no muons were measured in any muon detector. The data points are extracted from Fig. 7 of Ref. [134].

and z components of the Earth's magnetic field at Akeno to $B_x = 30.13\,\mu\mathrm{T}$ and $B_z = 35.45\,\mu\mathrm{T}$ [145], where the coordinate system is that of CORSIKA [23].

For simulations and mixed composition scenarios, the way to compute the muon densities (divided by the reconstructed energy) at 1000 m from the shower axis is analogous to what is described in Sec. 7.1.3.

8.2.4 Computation of a z-scale

Definition of two muon scale estimators

In Sec. 7.1.4 we introduced the definition of the muon scale [5],

$$z := \frac{\ln N_{\mu,\mathrm{data}}^{\mathrm{det}} - \ln N_{\mu,\mathrm{p}}^{\mathrm{det}}}{\ln N_{\mu,\mathrm{Fe}}^{\mathrm{det}} - \ln N_{\mu,\mathrm{p}}^{\mathrm{det}}}. \tag{8.7}$$

To build an estimator of the muon scale z, it is necessary to introduce the mean within a given energy bin. We defined in Sec. 7.1.4 one estimator of z by taking the logarithm of the mean of the muon number,

$$z_{\ln\langle\cdot\rangle} = \frac{\ln\langle N_{\mu,\mathrm{data}}^{\mathrm{det}}\rangle - \ln\langle N_{\mu,\mathrm{p}}^{\mathrm{det}}\rangle}{\ln\langle N_{\mu,\mathrm{Fe}}^{\mathrm{det}}\rangle - \ln\langle N_{\mu,\mathrm{p}}^{\mathrm{det}}\rangle}. \tag{8.8}$$

In this section we introduce a second estimator of z, defined by taking the mean of the logarithm of the muon number,

$$z_{\langle\ln\cdot\rangle} = \frac{\langle\ln N_{\mu,\mathrm{data}}^{\mathrm{det}}\rangle - \langle\ln N_{\mu,\mathrm{p}}^{\mathrm{det}}\rangle}{\langle\ln N_{\mu,\mathrm{Fe}}^{\mathrm{det}}\rangle - \langle\ln N_{\mu,\mathrm{p}}^{\mathrm{det}}\rangle}. \tag{8.9}$$

At this point it is important to mention that any other estimator that “mixes” the logarithm of the mean with the mean of the logarithm would suffer from undesired biases.

These two estimators are in principle different, since the logarithm of the mean and the mean of the logarithm of the muon content are not the same. One can be expressed as a function of the other by using that $N_\mu = \langle N_\mu \rangle (1+\epsilon)$, with $\epsilon = (N_\mu - \langle N_\mu \rangle)/\langle N_\mu \rangle$ being the event-wise relative error or deviation. Then

$$\begin{aligned} \langle \ln N_\mu \rangle &= \ln\langle N_\mu \rangle + \langle \ln(1+\epsilon) \rangle = \ln\langle N_\mu \rangle + \left\langle \epsilon - \tfrac{1}{2}\epsilon^2 + \mathcal{O}(\epsilon^3) \right\rangle, & (8.10) \\ &\approx \ln\langle N_\mu \rangle - \tfrac{1}{2}\left(\mathrm{RSD_{tot}}[N_\mu]\right)^2, & (8.11) \end{aligned}$$

where $\langle \epsilon^2 \rangle = \left(\sigma_{\mathrm{tot}}(N_\mu)/\langle N_\mu \rangle\right)^2 = \left(\mathrm{RSD_{tot}}[N_\mu]\right)^2$ is the square of the relative standard deviation of N_μ [126]. The suffix “tot” marks that all sources of fluctuations have to be considered. In the approximation we drop the higher-order terms.

Biases of the muon scale estimators

The estimators of the muon scale, $z_{\ln\langle\cdot\rangle}$ and $z_{\langle\ln\cdot\rangle}$ (Eq. (8.8) and Eq. (8.9) respectively), are subject to several sources of bias. If the detector simulations mismodel the real detector effects in $\langle N^{\mathrm{det}}_{\mu,\{\mathrm{p,Fe}\}} \rangle$, or if there is a composition bias in $\langle N^{\mathrm{det}}_{\mu,\mathrm{data}} \rangle$, both estimators will be biased approximately to the same degree. These sources are already contemplated in the systematic uncertainties.

The advantage of $z_{\ln\langle\cdot\rangle}$ is that it does not explicitly depend on the detector resolution, which makes it better for comparisons among different experiments. In contrast, $z_{\langle\ln\cdot\rangle}$ suffers from systematics from a possibly mis-modeled (“mm”) detector resolution. To understand this effect, we assume that the mean values are not biased, $\left\langle N^{\mathrm{mm}}_{\mu,\mathrm{p}} \right\rangle \approx \langle N_{\mu,\mathrm{p}} \rangle$, since this source of bias is already accounted for in the systematic uncertainties. The total detector resolution depends on the muon number detector resolution $\mathrm{RSD_{det}}[N_\mu]$, and on the energy resolution $\mathrm{RSD_{det}}[E]$. Both affect the measured $N_\mu(E)$. We can use Eq. (7.13) to propagate the energy resolution and obtain an approximation of the total muon number resolution

$$\mathrm{RSD}_{\mathrm{det}\oplus}[N_\mu] \approx \sqrt{(\mathrm{RSD_{det}}[N_\mu])^2 + (\beta \cdot \mathrm{RSD_{det}}[E])^2}, \qquad (8.12)$$

where $\beta \approx 0.9$. If the detector simulations mismodel the total detector resolution, the estimator $z_{\langle\ln\cdot\rangle}$ is also mismodeled:

$$z^{\mathrm{mm}}_{\langle\ln\cdot\rangle} = \frac{\langle \ln N_{\mu,\mathrm{data}} \rangle - \left\langle \ln N^{\mathrm{mm}}_{\mu,\mathrm{p}} \right\rangle}{\left\langle \ln N^{\mathrm{mm}}_{\mu,\mathrm{Fe}} \right\rangle - \left\langle \ln N^{\mathrm{mm}}_{\mu,\mathrm{p}} \right\rangle}. \qquad (8.13)$$

Using Eq. (8.11) we can approximate $\left\langle \ln N^{\mathrm{mm}}_{\mu,\mathrm{p}} \right\rangle \approx \ln\left\langle N^{\mathrm{mm}}_{\mu,\mathrm{p}} \right\rangle - \frac{1}{2}\left(\mathrm{RSD^{mm}_{tot}}[N_\mu]\right)^2$ and $\langle \ln N_{\mu,\mathrm{p}} \rangle \approx \ln\langle N_{\mu,\mathrm{p}} \rangle - \frac{1}{2}\left(\mathrm{RSD_{tot}}[N_\mu]\right)^2$. Equivalent expressions can be obtained, mutatis mutandis, for iron-initiated showers. For this estimation, we assume that $\left\langle N^{\mathrm{mm}}_{\mu,\mathrm{p}} \right\rangle \approx \langle N_{\mu,\mathrm{p}} \rangle$, since this source of bias is already accounted for in the systematic uncertainties. Under this consideration we can approximate

$$\left\langle \ln N^{\mathrm{mm}}_{\mu,\mathrm{p}} \right\rangle \approx \langle \ln N_{\mu,\mathrm{p}} \rangle + \frac{1}{2}\left[\left(\mathrm{RSD_{tot}}[N_\mu]\right)^2 - \left(\mathrm{RSD^{mm}_{tot}}[N_\mu]\right)^2\right]. \qquad (8.14)$$

Replacing the last expression in Eq. (8.13) we obtain

$$z^{\mathrm{mm}}_{\langle\ln\cdot\rangle} \approx z_{\langle\ln\cdot\rangle} - \frac{\frac{1}{2}\left[\left(\frac{\sigma_{\mathrm{tot}}(N_\mu)}{\langle N_\mu \rangle}\right)^2 - \left(\frac{\sigma^{\mathrm{mm}}_{\mathrm{tot}}(N_\mu)}{\langle N_\mu \rangle}\right)^2\right]}{\langle \ln(N_{\mu,\mathrm{Fe}}) \rangle - \langle \ln(N_{\mu,\mathrm{p}}) \rangle}, \qquad (8.15)$$

where we grouped the terms with true number of muons in $z_{\langle \ln \cdot \rangle}$. Note that the bias in the denominator cancels out since it is a priori the same for proton and iron detector-simulated showers. Finally, we use that $\langle \ln(N_{\mu,\mathrm{Fe}}) \rangle - \langle \ln(N_{\mu,\mathrm{p}}) \rangle$ is expected to be $(1-\beta)\ln(56)$ by the Heitler-Matthews model (see Eq. (7.13)). For convenience, we can write the bias in $z_{\langle \ln \cdot \rangle}$ as a function of the difference between the mismodeled and true detector resolution:

$$z^{\mathrm{mm}}_{\langle \ln \cdot \rangle} - z_{\langle \ln \cdot \rangle} \approx \frac{1}{(1-\beta)\ln(56)} \Big[\mathrm{RSD}_{\mathrm{det}\oplus}[N_\mu] \left(\mathrm{RSD}^{\mathrm{mm}}_{\mathrm{det}\oplus}[N_\mu] - \mathrm{RSD}_{\mathrm{det}\oplus}[N_\mu] \right) + \tfrac{1}{2} \left(\mathrm{RSD}^{\mathrm{mm}}_{\mathrm{det}\oplus}[N_\mu] - \mathrm{RSD}_{\mathrm{det}\oplus}[N_\mu] \right)^2 \Big] . \tag{8.16}$$

Figure 8.4 shows $z^{\mathrm{mm}}_{\langle \ln \cdot \rangle} - z_{\langle \ln \cdot \rangle}$ (color scale and contours) as a function of the true total detector resolution $\mathrm{RSD}_{\mathrm{det}\oplus}[N_\mu]$ (x-axis) and as a function of the difference between the mismodeled and true total detector resolutions $\mathrm{RSD}^{\mathrm{mm}}_{\mathrm{det}\oplus}[N_\mu] - \mathrm{RSD}_{\mathrm{det}\oplus}[N_\mu]$ (y-axis). As expected, $z^{\mathrm{mm}}_{\langle \ln \cdot \rangle} - z_{\langle \ln \cdot \rangle} = 0$ when the true and modelled resolution are identical. Shaded in gray is the non-physical region where $\mathrm{RSD}^{\mathrm{mm}}_{\mathrm{det}\oplus}[N_\mu] \leq 0$. The thick contours correspond to a systematic uncertainty in $z_{\langle \ln \cdot \rangle}$ of ± 0.07. From the plot we can read, for example, that for a true detector resolution of 30 %, $z_{\langle \ln \cdot \rangle}$ suffers from a systematic of ± 0.07 when the modelled resolution is up to 12 % smaller or 8 % larger than the true one. It can be added that Ref. [9] is a recent example in which the mismodelling of detector fluctuations was investigated.

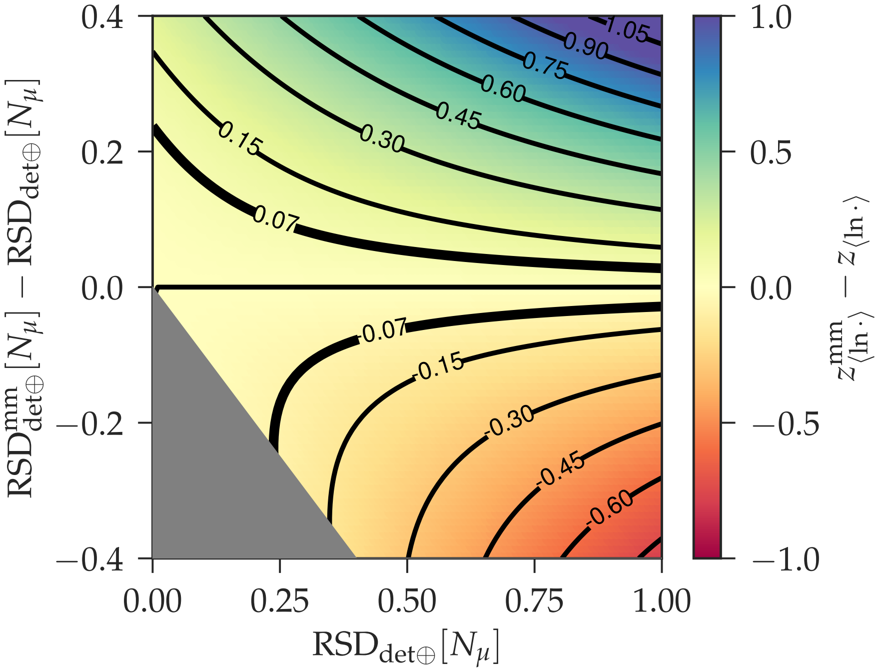


Figure 8.4: $z^{\mathrm{mm}}_{\langle \ln \cdot \rangle} - z_{\langle \ln \cdot \rangle}$ (contours and color scale) as a function of the true total detector resolution $\mathrm{RSD}_{\mathrm{tot}}[N_\mu]$ (x-axis), and of the difference between the mismodeled and true total detector resolutions $\left(\mathrm{RSD}^{\mathrm{mm}}_{\mathrm{tot}}[N_\mu] - \mathrm{RSD}_{\mathrm{tot}}[N_\mu]\right)$ (y-axis). The non-physical region where $\mathrm{RSD}^{\mathrm{mm}}_{\mathrm{det}\oplus}[N_\mu] \leq 0$ is shaded in gray. The thick countours mark $|z^{\mathrm{mm}}_{\langle \ln \cdot \rangle} - z_{\langle \ln \cdot \rangle}| = 0.07$.

Biases of the muon deficit estimators

As defined in Sec. 7.1.4, the muon deficit scale is

$$\Delta z := z - z_{\mathrm{mass}}. \tag{8.17}$$

In Sec. 7.1.4 we also showed that we can use the Heitler-Matthews model (c.f. Eq. (7.13)) to estimate z_{mass} as

$$z^{\text{HM}}_{\langle\ln\cdot\rangle\,\text{mass}} = \frac{\langle \ln A \rangle}{\ln 56}. \tag{8.18}$$

The advantage of using $z^{\text{HM}}_{\langle\ln\cdot\rangle\,\text{mass}}$ as reference is that it allows a direct comparison of muon measurements against the composition predictions ($\langle \ln A \rangle$) derived from other observables, most importantly, from fits to the X_{max} distributions.

We can define two estimators of the muon deficit scale Δz by subtracting $z^{\text{HM}}_{\langle\ln\cdot\rangle\,\text{mass}}$ from each of the two estimators of z, forming $\Delta z_{\ln\langle\cdot\rangle}$ and $\Delta z_{\langle\ln\cdot\rangle}$.

On one hand, the systematic uncertainties in $\Delta z_{\langle\ln\cdot\rangle}$ can be divided in two: the ones that propagate from $z_{\langle\ln\cdot\rangle}$, which are already understood (c.f. Sec. 8.2.4), and the ones that propagate from $z^{\text{HM}}_{\langle\ln\cdot\rangle\,\text{mass}}$. The latter suffers from a systematic uncertainty propagated from that of the composition (like any estimator of z_{mass}), and it has also an additional (smaller) uncertainty due to the Heitler-Matthews model not reproducing simulations perfectly.

On the other hand, $\Delta z_{\ln\langle\cdot\rangle}$ suffers from all of the systematic uncertainties described above. But in this case, $z^{\text{HM}}_{\langle\ln\cdot\rangle\,\text{mass}}$ introduces an additional systematic uncertainty, because it is the predicted value of $z_{\langle\ln\cdot\rangle\,\text{mass}} = (\langle \ln N^{\text{det}}_{\mu,\text{mass}}\rangle - \langle \ln N^{\text{det}}_{\mu,\text{p}}\rangle)/(\langle \ln N^{\text{det}}_{\mu,\text{Fe}}\rangle - \langle \ln N^{\text{det}}_{\mu,\text{p}}\rangle)$ instead of $z_{\ln\langle\cdot\rangle\,\text{mass}} = (\ln\langle N^{\text{det}}_{\mu,\text{mass}}\rangle - \ln\langle N^{\text{det}}_{\mu,\text{p}}\rangle)/(\ln\langle N^{\text{det}}_{\mu,\text{Fe}}\rangle - \ln\langle N^{\text{det}}_{\mu,\text{p}}\rangle)$. We can understand the difference between them by using Eq. (8.11) to approximate each term in $z_{\langle\ln\cdot\rangle\,\text{mass}}$ as

$$z_{\langle\ln\cdot\rangle\,\text{mass}} \approx \frac{\ln\langle N^{\text{det}}_{\mu,\text{mass}}\rangle - \ln\langle N^{\text{det}}_{\mu,\text{p}}\rangle - \frac{1}{2}\left[\left(\text{RSD}_{\text{sh-sh}}[N_{\mu,\text{mass}}]\right)^2 - \left(\text{RSD}_{\text{sh-sh}}[N_{\mu,\text{p}}]\right)^2\right]}{\ln\langle N^{\text{det}}_{\mu,\text{Fe}}\rangle - \ln\langle N^{\text{det}}_{\mu,\text{p}}\rangle - \frac{1}{2}\left[\left(\text{RSD}_{\text{sh-sh}}[N_{\mu,\text{Fe}}]\right)^2 - \left(\text{RSD}_{\text{sh-sh}}[N_{\mu,\text{p}}]\right)^2\right]}, \tag{8.19}$$

where we used that only shower-to-shower fluctuations depend on the primary, and all other sources of fluctuations cancel out. Note that the first two terms of the numerator divided by the first two terms of the denominator conform $z_{\ln\langle\cdot\rangle\,\text{mass}}$.

From Eq. (8.19) we see that $z_{\langle\ln\cdot\rangle\,\text{mass}}$ depends on the shower-to-shower fluctuations of the muon content, while $z_{\ln\langle\cdot\rangle\,\text{mass}}$ does not. These fluctuations affect $z^{\text{HM}}_{\langle\ln\cdot\rangle\,\text{mass}}$, which come in as a bias in $\Delta z_{\ln\langle\cdot\rangle}$. This bias in $\Delta z_{\ln\langle\cdot\rangle}$ is largest when $\left(\text{RSD}_{\text{sh-sh}}[N_{\mu,\text{mass}}]\right)^2$ is maximum, which is given approximately at a 50 % proton - 50 % iron mixture. It was reported to be of 0.07 in the worst case scenario [126]. This bias should be compared to the systematic uncertainties in $\Delta z_{\langle\ln\cdot\rangle}$ to understand which estimator provides the smaller systematic uncertainties for a given experiment. Essentially, if the experiment's resolution is very good and known in good detail, the estimator $\Delta z_{\langle\ln\cdot\rangle}$ has smaller biases; otherwise $\Delta z_{\ln\langle\cdot\rangle}$ has smaller biases. It is also relevant to add that this bias could be corrected using a parameterization of shower-to-shower fluctuations for single primaries (e.g. like the one found in Ref. [126]), as well as $\sigma(\ln A)$, typically predicted in composition models (e.g. by the Global Spline Fit [61] and by the Pierre Auger [57] composition models).

Computation of the two z-value estimators from AGASA data

With the muon density obtained from the simulations using Eq. (8.1) and the data, we can directly compute $z_{\ln\langle\cdot\rangle}$ as described in Sec. 8.2.4 (c.f. Eq. (8.8)). It should be noted that all equations of Sec. 8.2.4 are valid when N_μ is replaced with $\rho_{\mu\,r}$.

In contrast, like it was discussed in Sec. 7.1.4, it is not possible to directly compute $z_{\langle\ln\cdot\rangle}$ as defined in Sec. 8.2.4 (c.f. Eq. (8.9)). The reason is that the values of $\left\langle \ln[\rho_{\mu\,\{\text{p,Fe}\}}(E_{Ri})] \right\rangle$ cannot be analytically computed (like in Eq. (8.1)) without a very detailed knowledge of the detector resolution. Nevertheless, we can use Eq. (8.11) to give an estimate of $z_{\langle \ln \cdot \rangle}$. The derivation is analogous to that of Sec. 7.1.4. The approximation yields

$$z_{\langle \ln \cdot \rangle} \approx \frac{\left\langle \ln \rho_{\mu,\mathrm{data}}^{\mathrm{det}} \right\rangle - \ln \left\langle \rho_{\mu,\mathrm{p}} \right\rangle + \frac{1}{2}\left[\left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu,\mathrm{p}}] \right)^2 + \left(\mathrm{RSD}_{\text{not sh-sh}}[\rho_{\mu}] \right)^2 \right]}{\ln \left\langle \rho_{\mu,\mathrm{Fe}} \right\rangle - \ln \left\langle \rho_{\mu,\mathrm{p}} \right\rangle + \frac{1}{2}\left[\left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu,\mathrm{p}}] \right)^2 - \left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu,\mathrm{Fe}}] \right)^2 \right]}, \tag{8.20}$$

where we use in the denominator that all sources of fluctuations except shower-to-shower cancel out. In the numerator, we also split the total relative variance of proton air showers into shower-to-shower and not shower-to-shower fluctuations. The latter is computed as

$$\left(\mathrm{RSD}_{\text{not sh-sh}}[\rho_{\mu}] \right)^2 = \left(\mathrm{RSD}_{\mathrm{tot}}[\rho_{\mu,\mathrm{data}}] \right)^2 - \left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu,\mathrm{mass}}] \right)^2, \tag{8.21}$$

where $\mathrm{RSD}_{\mathrm{tot}}[\rho_{\mu,\mathrm{data}}]$ is computed from the data scatter within each reconstructed energy bin, while $\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu,\mathrm{mass}}]$ is computed from simulations assuming a mixed composition. As in Sec. 7.1.4, $\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu,\mathrm{mass}}]$ is obtained by fitting a normal distribution to the weighted sum of the single-primary distributions of the simulated muon densities, where the weights are the mass fractions given by the GSF model.

It is relevant to add that we propagate the uncertainties of the muon density into each estimate of z.

8.3 Results

In this section we compare the muon content as a function of the reconstructed energy of data to single-proton and single-iron simulations, as well as to mixed composition scenarios. We also show the z-values computed from the two different estimators and a comparison against the expected values predicted by the mixed composition models.

8.3.1 Muon content as a function of the energy

In Fig. 8.5 we plot the average muon density (at 1000 m from the shower axis) divided by the reconstructed energy, as a function of the logarithm of center of the reconstructed energy bin (in the reference scale), $\langle \rho_\mu / E_R \rangle (\log_{10}(E_{Ri}/\mathrm{eV}))$. We compare the values computed from AGASA data, from proton and iron simulations, and from the mixed composition scenarios. We group the data from the 120 events in three energy bins of a bin width of $\Delta \log_{10}(E_R/\mathrm{eV}) = 0.2$, which result in 67, 33, and 20 events from the lower to the higher energy bin. The systematic uncertainties of the AGASA data are represented in square brackets, and they correspond to the energy (hence they are diagonal) and muon density systematics. The square brackets corresponding to the mixed composition models account for the systematics in the energy, and also include the systematic uncertainties propagated from the mass fractions, which are smaller than the former.

From Fig. 8.5 we can see that the AGASA data are more compatible with a pure iron composition for all models, and that the data points lie above those of the mixed composition scenarios. In particular, the lowest energy data point is not compatible with the mixed composition scenarios for all three models. This also holds for the highest-energy data point in the case of QGSJetII-04. In contrast, the remaining data points are compatible with the mixed composition scenarios within total uncertainties.

Furthermore, we summarize the information in a unique value by computing $\langle \rho_\mu / E_R \rangle$ taking the average in the whole energy range ($18.83 \leq \log_{10}(E_R/\mathrm{eV}) \leq 19.46$). Notice that ρ_μ / E_R is nearly constant within the analyzed energy range, hence it is reasonable to compute

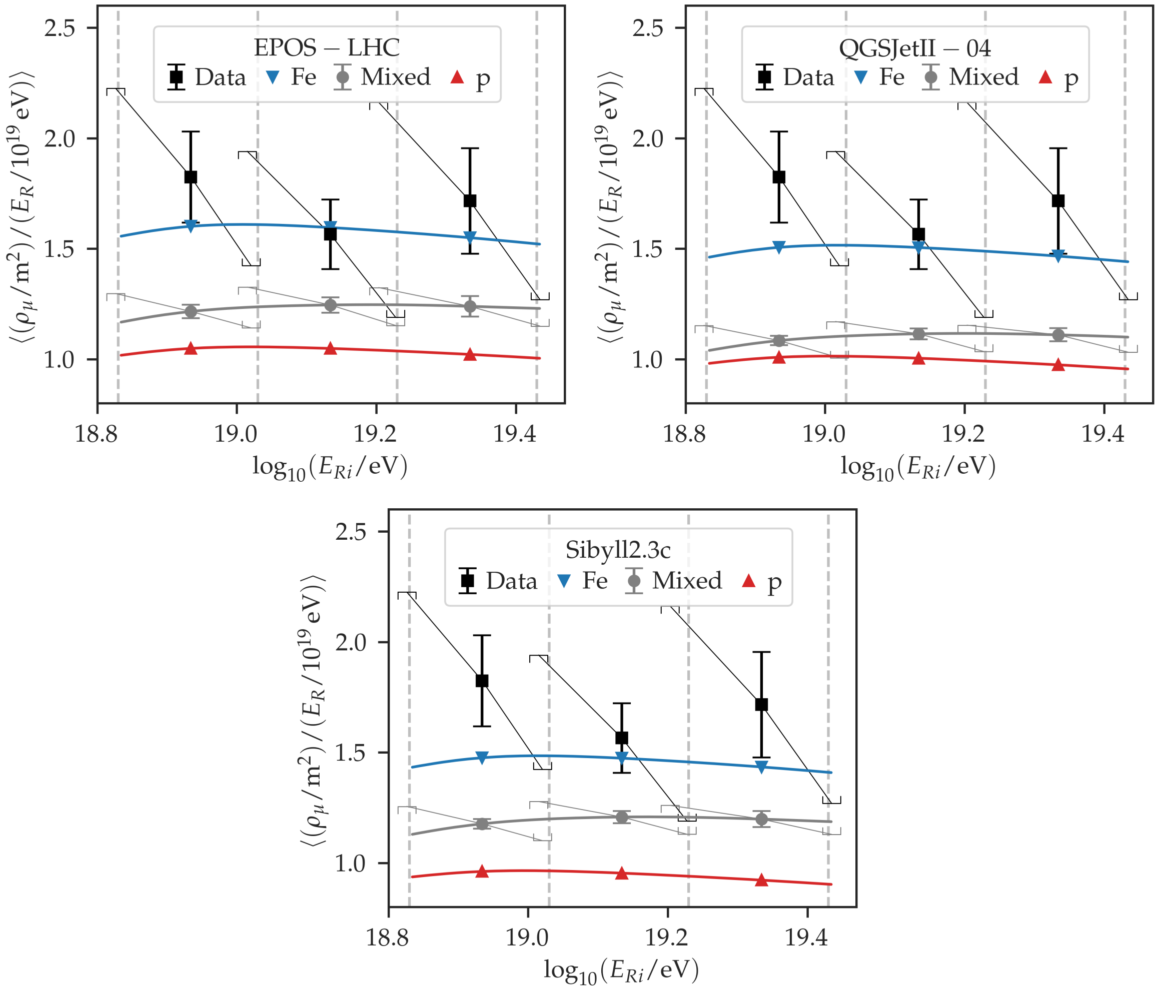


Figure 8.5: Average muon density (at 1000 m from the shower axis) divided by the reconstructed energy, as a function of the logarithm of the reconstructed energy in the center of the *i*-th bin (in the reference scale), $\langle\rho_\mu/E_R\rangle(\log_{10}(E_{Ri}/\mathrm{eV}))$. The AGASA data points [134] are shown in black squares, those of proton and iron simulations in upper red and lower blue triangles (respectively), and those of the mixed composition scenarios in gray circles. The top left panel corresponds to the model EPOS-LHC, the top right panel to QGSJetII-04 , and the bottom panel to Sibyll2.3c. In square brackets we depict the systematic uncertainties. The gray dashed lines mark the edges of the reconstructed energy bins.

such average. In Fig. 8.6 we show the values of the aforementioned average computed from AGASA measurements, and for proton, iron, and mixed composition scenarios for the three considered models.

From Fig. 8.6 we can see a similar behavior to what was seen in Fig. 8.5. Once again, the composition inferred from $\langle\rho_\mu/E_R\rangle$ from AGASA data is compatible with heavy primaries, for the three models considered. In this case, the values from AGASA data are larger than those of the mixed composition scenarios, and are not compatible within total uncertainties. Considering the upper uncertainties for the mixed composition models, and the lower uncertainties for the AGASA data point, the discrepancies are of 1.9σ for QGSJetII-04, 1.6σ for EPOS-LHC, and 1.7σ for Sibyll2.3c. Assuming that the mixed composition model is correct, this implies that the AGASA data constitute evidence of a muon deficit in simulations.

Moreover, as in Sec. 7.2.2 we quantify the average muon deficit in the reconstructed energy range $18.83 \leq \log_{10}(E_R/\mathrm{eV}) \leq 19.46$ with a correction factor $F = \langle\rho_\mu^{\mathrm{data}}/E_R\rangle/\langle\rho_\mu^{\mathrm{S}}/E_R\rangle$

		$\langle(\rho_\mu/\mathrm{m}^2)/(E_R/10^{19}\,\mathrm{eV})\rangle \pm (\mathrm{stat}) \pm (\mathrm{syst})$
AGASA	Data	$1.82 \pm 0.13 \pm^{0.58}_{0.35}$
	Fe	1.47
Sibyll2.3c	Mixed	$1.19 \pm 0.03 \pm^{0.07}_{0.08}$
	p	0.95
	Fe	1.59
EPOS - LHC	Mixed	$1.23 \pm 0.04 \pm^{0.08}_{0.08}$
	p	1.05
	Fe	1.50
QGSJetII - 04	Mixed	$1.10 \pm 0.03 \pm^{0.06}_{0.08}$
	p	1.00

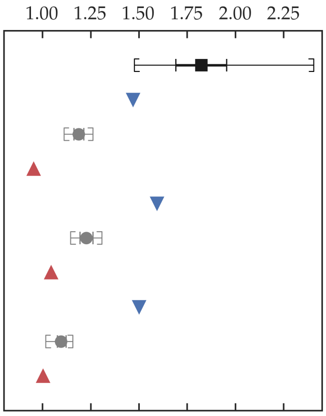


Figure 8.6: Average muon density (at 1000 m from the shower axis) divided by the reconstructed energy, $\langle \rho_\mu / E_R \rangle$, for AGASA data and for proton, iron, and mixed composition scenarios, for the three considered models. The average is taken in the energy range $18.83 \leq \log_{10}(E_R/\mathrm{eV}) \leq 19.46$. The values obtained are reported in the table (left) and are also plotted (right) on the same row. The square brackets show the systematic uncertainties.

(i.e. Eq. (7.20)), being S the scenario under consideration (S={mix, p, Fe}). We propagate the uncertainties of the numerator and denominator into F by standard uncertainty propagation. In Fig. 8.7 we can see the computed values of F, together with their statistic and systematic uncertainties, for the single nuclei and mixed composition scenarios, for the three considered models. Similar to what we observed in Fig. 8.6, we can see that the correction factors of the mixed composition scenarios do not overlap with 1 even considering total uncertainties.

		$F \pm (\mathrm{stat}) \pm (\mathrm{syst})$
	Fe	$1.24 \pm 0.09 \pm^{0.40}_{0.24}$
Sibyll2.3c	Mixed	$1.54 \pm 0.12 \pm^{0.50}_{0.31}$
	p	$1.91 \pm 0.14 \pm^{0.61}_{0.37}$
	Fe	$1.14 \pm 0.08 \pm^{0.37}_{0.22}$
EPOS - LHC	Mixed	$1.49 \pm 0.12 \pm^{0.49}_{0.30}$
	p	$1.74 \pm 0.13 \pm^{0.56}_{0.33}$
	Fe	$1.21 \pm 0.09 \pm^{0.39}_{0.23}$
QGSJetII - 04	Mixed	$1.66 \pm 0.13 \pm^{0.54}_{0.34}$
	p	$1.82 \pm 0.13 \pm^{0.58}_{0.35}$

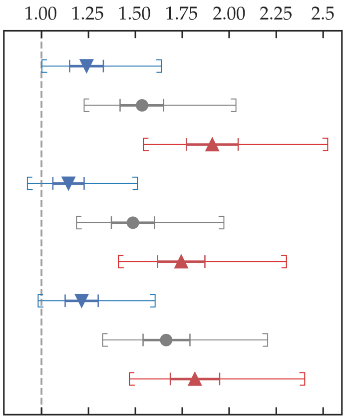


Figure 8.7: Correction factor F (see text for details) corresponding to the single nuclei and mixed composition scenarios, for the three considered models. The analyzed energy range is $18.83 \leq \log_{10}(E_R/\mathrm{eV}) \leq 19.46$. The obtained values are reported in the table (left) and are also plotted (right) on the same row. The systematic uncertainties are depicted with square brackets.

It is relevant to add that, when using the fit to the flux measurements of the Pierre Auger Observatory [138] instead of that of Telescope Array [123, 137], the values of F and $\langle \rho^S_\mu / E_R \rangle$ vary less than $\sim 1\,\%$. In other words, the results shown in Figs. 8.6 and 8.7 are basically independent of the chosen flux parameterization.

8.3.2 Computed z-values

Figure 8.8 shows the estimates of $z_{\ln\langle\cdot\rangle}$ (computed using Eq. (8.8)) and $z_{\langle\ln\cdot\rangle}$ (computed using Eq. (8.20)) as a function of the logarithmic reconstructed energy (in the reference scale). The relative difference of $z_{\langle\ln\cdot\rangle}$ to $z_{\ln\langle\cdot\rangle}$ ranges from $-9\,\%$ to $+10\,\%$, and is written next to the $z_{\langle\ln\cdot\rangle}$ values. We also show the values of $z^{\mathrm{HM}}_{\mathrm{mass}}$ from the GSF (gray dashed lines) and Pierre Auger $X_{\max}$ (gray shaded area) composition models. In this case, we do not include in the plotted systematic uncertainties those of the reference energy scale, as these values are compared (in Sec. 8.4) to those of other experiments that are subject to the same source of uncertainty.

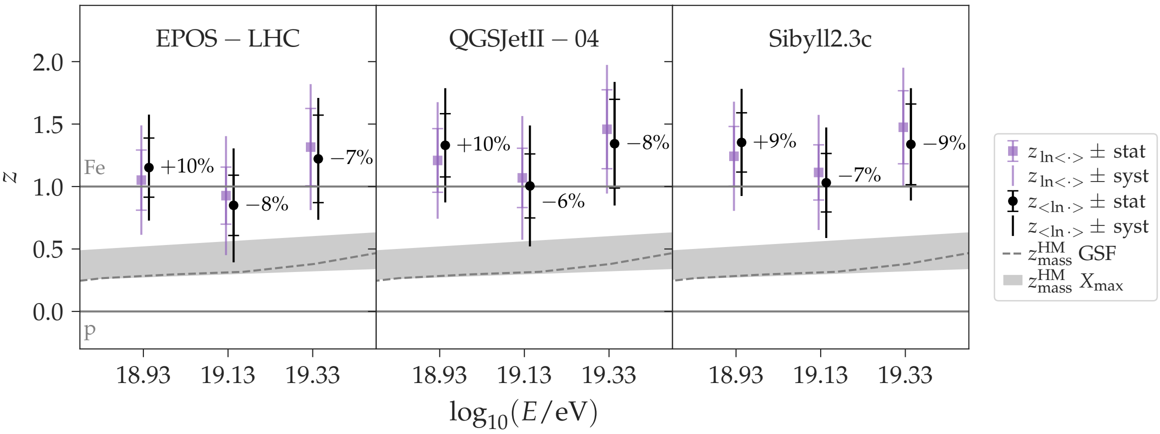


Figure 8.8: $z_{\ln\langle\cdot\rangle}$ (squares) and $z_{\langle\ln\cdot\rangle}$ (circles), for EPOS-LHC (left), QGSJetII-04 (center), and Sibyll2.3c (right) as a function of the logarithmic reconstructed energy (in the reference energy scale). For clarity, we introduce a horizontal displacement between the two sets of values. The relative difference of $z_{\langle\ln\cdot\rangle}$ to $z_{\ln\langle\cdot\rangle}$ is written next to each pair of points. We also show the values of $z^{\mathrm{HM}}_{\mathrm{mass}}$ from the GSF model (gray dashed lines) and from the Pierre Auger $X_{\max}$ (gray shaded area) composition models.

We can understand the difference between the $z_{\ln\langle\cdot\rangle}$ and $z_{\langle\ln\cdot\rangle}$ values from Sec. (8.2.4). As mentioned there, $z_{\langle\ln\cdot\rangle}$ is subject to systematic uncertainties due to mismodelling the detector resolution. We can see it directly in Eq. (8.20), where the estimated total detector resolution is part of $\left(\mathrm{RSD}_{\text{not sh-sh}}[\rho_\mu]\right)$; the other contributions to it are all primary-independent sources of fluctuations. We can obtain an estimate of the total detector resolution by normalizing the muon densities to the center of the reconstructed energy bin and subtracting $\left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu,\mathrm{mass}}]\right)^2$. With that, we can estimate the total detector resolution as $\mathrm{RSD}_{\mathrm{det}\oplus}[N_\mu] \approx \sqrt{(\mathrm{RSD}_{\mathrm{det}}[N_\mu])^2 + (\beta \cdot \mathrm{RSD}_{\mathrm{det}}[E])^2}$. The total detector resolution estimated in this way ranges between 50 % to 84 %, depending on the energy. We can read from Fig. 8.4 that a total detector resolution of 50 % mismodeled in as little as $\sim\pm 6\,\%$ translates into a systematic in $z_{\langle\ln\cdot\rangle}$ and $\Delta z_{\langle\ln\cdot\rangle}$ of already ± 0.07. The estimate of the the total detector resolution is not more precise than $\pm 6\,\%$. This implies that for the AGASA data analyzed in this work, $z_{\ln\langle\cdot\rangle}$ and $\Delta z_{\ln\langle\cdot\rangle}$ are better estimators of the muon scale and muon deficit, in comparison to $z_{\langle\ln\cdot\rangle}$ and $\Delta z_{\langle\ln\cdot\rangle}$.

At this point, it is relevant to add that this is also the case for the Pierre Auger data. We can estimate the total detector resolution in an analogous way as for AGASA, which for the Pierre Auger data ranges from 23 % to 54 %, depending on the energy. In this case we can read from Fig. 8.4 that a total detector resolution of 23 % (resp. 54 %) mismodeled in 10 % (resp. $\sim\pm 5\,\%$) translates into a systematic in $z_{\langle\ln\cdot\rangle}$ and $\Delta z_{\langle\ln\cdot\rangle}$ of already ± 0.07. The resolution in the distance to the shower core, that we propagate into the resolution in $\rho_{\mu\,450}$, is not known to such a good precision. Therefore, $z_{\ln\langle\cdot\rangle}$ and $\Delta z_{\ln\langle\cdot\rangle}$ are the best estimators in this case too. Nevertheless, we show the comparison of the two z-values in Appendix B.2. It is also worth mentioning that if the parameter of the muon LDF fit ρ_{450} was used instead of the average of muon densities around an interval at 450 m, the detector resolution might improve enough such that the estimators with least biases are then $z_{\langle \ln \cdot \rangle}$ and $\Delta z_{\langle \ln \cdot \rangle}$.

8.4 Comparison with the z-values from other experiments

Figure 8.9 shows the muon scale (computed from $z_{\ln \langle \cdot \rangle}$) and Fig. 8.10 shows the muon deficit scale (computed from $\Delta z_{\ln \langle \cdot \rangle}$) of several experiments: from AGASA data (computed in this work), from UMD SiPM data of the Pierre Auger Observatory (computed in this work in Chapter 7), from the SD-750 (infill) array and Heat-Coihueco hybrid analysis also from Pierre Auger Observatory data [15], from UMD PMT data also of the Pierre Auger Observatory [5, 59, 146], from the IceCube data [5, 59, 146], from NEVOD-DECOR data [5, 59, 146], from the SD-1500 array and FD hybrid measurements and from the SD-1500 inclined events at higher energies of the Pierre Auger data [5, 59, 146], from SUGAR data [5, 59, 146], from Yakutsk data [5, 59, 146], from the EAS-MSU data [5, 59, 146], and finally from the KASCADE-Grande data [5, 59, 146]. We make the comparison for two high-energy hadronic interaction models, EPOS-LHC and QGSJetII-04, as almost all experiments computed the z-values only for the latter.

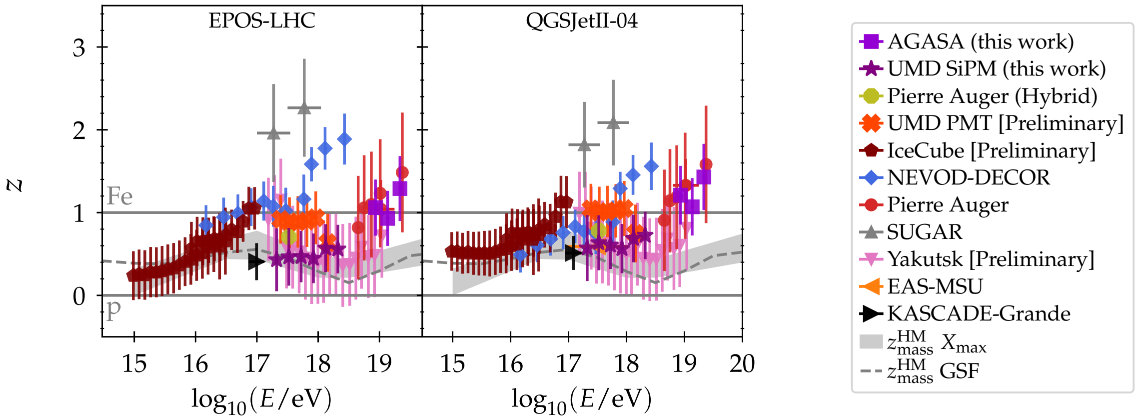


Figure 8.9: Values of z computed from the UMD data (c.f. Chapter 7), from AGASA data (this chapter), and from nine different data sets of seven experiments [5, 15, 59, 146], as a function of the logarithm of the center of the reconstructed energy bin. The left panel shows the values computed for EPOS-LHC, and the right panel for QGSJetII-04. Gray horizontal lines mark the expected values of z for pure proton and pure iron compositions (0 and 1 respectively). Dashed lines shows the z-value expected from mixed composition scenarios considering the mass fractions of the GSF model, while the gray shaded area represents the ones expected from the fits to the X_{max} distributions. The error bars represent total uncertainties (excluding the uncertainty of the reference energy scale).

Figures 8.9 and 8.10 show that the AGASA data presents a remarkable agreement with the Pierre Auger data. Furthermore, the AGASA data lie above the Yakutsk Array data but are compatible within total uncertainties. From the Fig. 8.10 we can also see that the $\Delta z_{\ln \langle \cdot \rangle}$ values from AGASA data support a muon deficit in simulations at the highest energies.

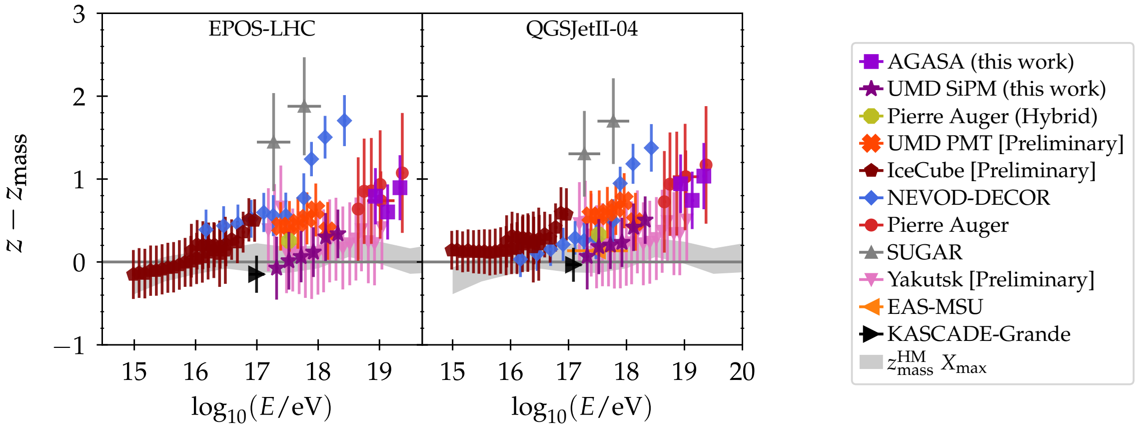


Figure 8.10: Values of Δz computed from the UMD data (c.f. Chapter 7), from AGASA data (this chapter), and from nine different data sets of seven experiments [5, 15, 59, 146], as a function of the logarithm of the center of the reconstructed energy bin. The left panel shows the values computed for EPOS-LHC, and the right panel for QGSJetII-04. A gray horizontal lines marks the expected value of Δz if there was a complete agreement between measurements and the expected values from the GSF model. The error bars represent total uncertainties (excluding the uncertainty of the reference energy scale).

8.5 Summary and conclusions

In this chapter we analyzed the measurements of the muon density at 1000 m from the shower axis reported by the AGASA experiment. We compared them to the muon content of air shower simulations generated with the high-energy hadronic interaction models QGSJetII-04, EPOS-LHC, and Sibyll2.3c. We simulated single proton and single iron-initiated air showers, as well as four-component mixed composition scenarios from two composition models: from the fits to the X_{max} distributions of the Pierre Auger Observatory, and from the Global Spline Fit model. Moreover, we adopted the reference energy scale defined by the *Spectrum Working Group* [122] to allow for the comparison with different experiments results. In order to compare the data to simulations, we analytically computed the effects introduced by the energy reconstruction and by binning in reconstructed energy.

We found that the AGASA measurements are compatible with a heavy composition for all the considered models, lying above the predictions of the mixed composition scenarios for the three considered interaction models. When analyzing the average throughout the complete energy range ($18.83 \leq \log_{10}(E_R/\text{eV}) \leq 19.46$), we found the estimate of $\langle \rho_\mu / E \rangle$ from AGASA data is significantly larger than those of the mixed composition scenarios. Furthermore, we computed a muon density correction factor F in the complete energy range for the three models considered. We found that, for the mixed composition scenarios to be in perfect agreement with AGASA measurements, the muon density should be incremented by a factor of $1.49 \pm 0.11\,(\text{stat}) \pm {}^{0.49}_{0.30}\,(\text{syst})$ for EPOS-LHC, $1.54 \pm 0.12\,(\text{stat}) \pm {}^{0.50}_{0.31}\,(\text{syst})$ for Sibyll2.3c, and $1.66 \pm 0.13\,(\text{stat}) \pm^{0.54}_{0.34}\,(\text{syst})$ for QGSJetII-04.

Moreover, to compare the AGASA composition or muon deficit implications to that of other experiments, we studied two estimators of the muon scale and computed their values from the AGASA data. The first estimator, $z_{\ln\langle\cdot\rangle}$, is computed from the logarithm of the mean of the muon number/density in data and simulations, while the second estimator, $z_{\langle\ln\cdot\rangle}$, is computed from the mean of the logarithm of the muon number/density. To estimate the muon deficit $\Delta z_{\ln\langle\cdot\rangle}$ or $\Delta z_{\langle\ln\cdot\rangle}$, we subtract $z^{\text{HM}}_{\text{mass}} = \langle \ln A \rangle / \ln 56$ from the corresponding estimator of z.

We showed that $z_{\ln\langle\cdot\rangle}$ and $z_{\langle\ln\cdot\rangle}$ are subject to different systematic uncertainties. On one hand, $z_{\langle\ln\cdot\rangle}$ (and hence $\Delta z_{\langle\ln\cdot\rangle}$) is subject to systematics from a mismodeled detector resolution. This systematic error depends on the total detector resolution and the degree to which it is mismodeled. On the other hand, $\Delta z_{\ln\langle\cdot\rangle}$ is biased from the shower-to-shower fluctuations in $z_{\mathrm{mass}}^{\mathrm{HM}}$ by at most ~ 0.07. For the case of AGASA, and in general when the detector resolution is not very good and/or known to a very good precision, $\Delta z_{\ln\langle\cdot\rangle}$ proves to be a better estimator of the muon deficit, with less systematic uncertainties.

Finally, we compared the values of $z_{\ln\langle\cdot\rangle}$ and $\Delta z_{\ln\langle\cdot\rangle}$ from the AGASA data to those of other experiments, including the values computed from the UMD SiPM data (c.f. Chapter 7). We observed that the AGASA data are in very good agreement with the Pierre Auger data, and are larger than the Yakutsk array values, but compatible within total uncertainties. The AGASA data support a muon deficit in simulations. These estimates add a valuable quantification of the muon scale, and constitute evidence of a muon deficit in simulations at the highest energies.

Chapter 9

Summary and outlook

This work is devoted to the precise determination of the muon content of air showers, and to the analysis of their mass composition or muon deficit implications. The muon content is a composition-sensitive observable par excellence, and knowing the mass composition as a function of the energy would shed light on a wide range of unsolved astrophysical problems, strongly linked to the still unknown origin of cosmic rays.

We first revisited the Heitler-Matthews model, a very simple yet powerful model of hadronic air showers. We showed how it predicts that the number of muons in an air shower is related to the atomic mass number of the cosmic ray that generated it. We also introduced the Pierre Auger Observatory, the largest cosmic-ray observatory in existence, and its multiple detectors. We meticulously described the Underground Muon Detector (UMD), the buried scintillation detectors of the Pierre Auger Observatory, the data of which we extensively analyzed.

In our aim to obtain an unbiased estimate of the number of muons for the UMD, we developed a novel muon counting strategy. It is based in the solution to the "balls in boxes problem", where the balls are the particles and the boxes are the scintillation bars. The strategy accounts by design for the effect of unresolved particles, i.e., the effect of two or more particles hitting the same scintillation bar almost simultaneously. We compared this strategy to three others from literature, including the one used in the UMD until this work. The key difference of the new strategy with respect to the rest is the consideration of inhibited scintillation bars, which allows for the exploitation of the full time resolution of the detector. Using realistic simulations of the detector response to air showers, we found that our strategy performs with a comparably small bias in the non-saturation regime, and significantly better than the previously used strategy. Moreover, it offers a unique possibility: to reconstruct the time structure of the muon signal as seen by the detector to a single time-bin resolution. Such a resolution is not achievable with any other known strategy.

Due to detector inefficiencies and an undershoot of the signal, we saw that the estimated number of muons is not exempt from bias. Additionally, muons traversing two neighboring scintillation bars are often a source of over-counting. Using full detector simulations, we developed a method to correct for this and remaining biases of the muon number estimator. We parameterized the bias as a function of the zenith angle of the shower, of the azimuth of the module with respect to the azimuth of the shower, and of the uncorrected estimator of the number of muons itself. The dependency on the number of muons was not reported before, but we showed that it is physically expectable. After using the parameterization to correct for the biases, the estimator of the number of muons was found to be accurate within $\pm 4\,\%$.

With reliable estimates of the number of muons, we went on to investigate the muon lateral distribution function (muon LDF) as measured with the UMD. The first step was to define the phase space within which the observed muon densities constitute unbiased samples. We applied distance cuts to avoid saturation of the muon detectors and to ensure a high station-level (lateral) trigger probability. The selected energies and zenith angles were determined by the array-level trigger probability and by the acquired statistics. A very robust outlier detection technique, based on a z-score on the logarithmic muon density, was implemented to exclude ill-behaving modules. With this, we binned the data in logarithmic energy, sine-square of the zenith angle, and logarithmic distance to the shower axis. We normalized the muon densities to the center of the corresponding reconstructed energy bins. We then performed a weighted least-squares fit on the logarithmic average of the normalized muon density, testing several models found in literature. To compare their goodness-of-fit we used the Akaike and Bayesian information criteria, and we also performed a Wald Wolfowitz test to check for the randomness of the signs of the residuals. We found that the best fitting function is the so-called modified Nishimura-Kamata-Greisen function with certain parameters fixed ($r_0 = 320\,\mathrm{m}$, $\alpha = 0.75$, $\gamma = 3.0$). Using this function, we provided parameterizations of the muon LDF as measured with the UMD. These could be used, for example, for fixing all parameters except the size of the muon LDF in low-multiplicity event-wise fits, or as a reference for the design of future muon detectors, or for building realistic toy models of the muon LDF, as we did. Using said toy model, we studied the systematic uncertainties introduced in the fitted muon LDFs by fixing the parameters α and γ, and by applying the distance cut at the chosen lateral trigger probability. A relevant conclusion of this analysis is that, for the 750 m surface detector array of the Pierre Auger Observatory, the distance that minimizes systematic uncertainties due to the lack of information of the muon LDF is 450 m. The systematic uncertainties introduced at this optimal distance by the fits of the muon LDF are within $\pm 2\,\%$.

In order to infer the composition interpretation of the UMD data, we compared its muon content to that of single-proton and single-iron simulations, as well as against mixed composition scenarios derived from the fits to the distributions of the Pierre Auger [57] and Telescope Array [131] depth-of-the-shower maximum. We used simulations of the three newest-generation high-energy hadronic interaction models: EPOS-LHC, QGSJetII-04, and Sibyll2.3c. We developed an analytical method to estimate the effects of energy reconstruction and energy binning (instead of normalizing the measured muon densities to the center of the reconstructed energy bin). A positive shift of 5.2 % in energy was considered in this analysis to work in the cross-calibrated energy scale described by the *Spectrum Working Group* [122]. This enabled the correct combination and comparison of results from different experiments. With this, we compared the muon content in data and simulations as a function of the distance to the shower axis, of the zenith angle, and of the energy (between $10^{17.22}$ eV and $10^{18.42}$ eV). We found that the slope of the muon LDF is not well reproduced by simulations for more vertical zenith angles ($\theta < 30^\circ$). This showed that introducing a unique normalization correcting factor would not suffice to solve the difference between data and simulations. We also observed that the muon deficit in simulations was larger at shorter distances to the shower core ($r \sim 280\,\mathrm{m}$), where most muons are produced by the high-energy interaction models. In agreement with previous Pierre Auger results [57], QGSJetII-04 presents significantly larger disagreements than the other models. Although in most cases the muon content of data was found compatible with the mixed composition scenarios within total uncertainties, the difference between them was clearly increasing with the energy. Furthermore, we computed the so-called z-values from these UMD data and simulations. We found our values compatible within total uncertainties with those of the

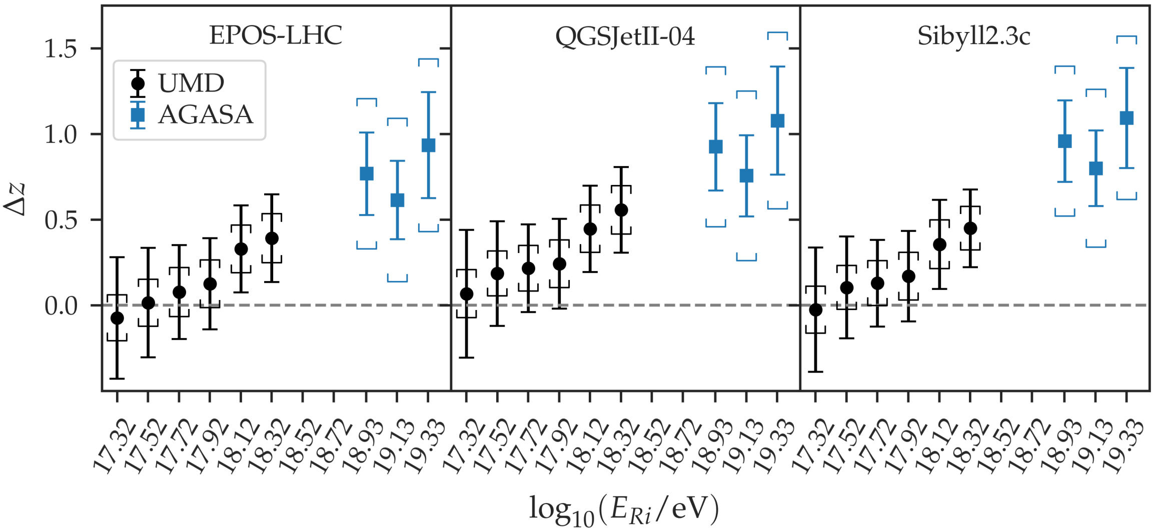


Figure 9.1: Difference between the computed z-values from UMD and AGASA data and the expected z-values from the Global Spline Fit composition model [61], as a function of the logarithmic reconstructed energy. The left panel corresponds to the high-energy hadronic interaction model EPOS-LHC, the middle panel corresponds to QGSJetII-04, and the right panel to Sibyll2.3c. A gray dashed line marks $\Delta z = 0$. Error bars represent statistical uncertainties while brackets represent systematic uncertainties.

UMD PMT data analysis [5] (although the latter are larger), as well as with those of a hybrid analysis of surface and fluorescence detector data of the 750 m array [15].

Finally, to complement the results of the UMD, we used the public dataset of direct muon measurements at 1000 m from the shower axis of the AGASA collaboration [134]. These hybrid data are complementary to the UMD data because they cover higher energies (between $10^{18.83}$ eV and $10^{19.46}$ eV). With analogue methods as the ones applied to the UMD data, we estimated the detector effects, and corrected the energy by a factor of 0.68 to work in the cross-calibrated energy scale. In this case, the energy correction factor was considerably larger mostly because the AGASA energy calibration, being simulation driven, relied in older hadronic interaction models which suffered from even greater muon deficits. We observed that the muon content of AGASA data was significantly larger than that of the mixed composition scenarios, being compatible with a heavier composition. The muon content in air-shower simulations should be increased by 54 % to 66 % (depending on the model) to match the muon content of data. This was interpreted as evidence of a muon deficit in air-shower simulations. Moreover, we extensively discussed two possible ways to compute the z-values and their uncertainties. We proved that, unless the detector resolution is very good and known to a very good detail, it is better to compute z from the logarithm of the mean of the muon densities (instead of from the mean of the logarithm of the muon densities).

Figure 9.1 shows the computed z-values from the direct muon measurements of the UMD and of AGASA, with the expected value from the Global Spline Fit composition model [61] subtracted. It can be seen that the Δz-values increasingly deviate from zero, which constitutes clear evidence of a muon deficit that increases with the energy. These results are largely consistent with other nine estimations from different experiments or analyses [5, 15].

The precise estimations of the muon content of UMD data developed in this work are now part of the standard reconstruction of UMD data and simulations. Future analyses of these data will profit from the techniques here presented. The developed counting strategy and the treatment of residual biases are general methods that can be applied to other counting detectors with counting sub-units, with possible applications, for instance, to neutrino experiments. This work opens the door to new research into the temporal structure of the muon component as a result of the here developed counting strategy. By deconvolving the detector effects in the muon temporal signal, the maximum of the muon production depth distribution $X^{\mu}_{\max}$ could be reconstructed [93]. This great composition-sensitive observable can be key for new mass composition analyses.

The analysis of the muon LDF of the UMD can be used as a point of comparison for other estimations. The provided parameterizations can be employed for fixing the shape of the muon LDF in low-multiplicity event-wise fits, or for building toy models as done in this work. The detailed comparisons of the muon content in UMD and AGASA data and simulations add valuable information of where the discrepancies appear, which can hint model builders towards the origin of the muon deficit in simulations. If the cause were found, the here computed z-values could be corrected by the difference in the muon content in the new simulations. In this way, the deficit-corrected z-values would constitute mass composition estimates that could reduce the overall uncertainties, ultimately contributing to a better understanding of the origin of cosmic rays.

Appendix A

Selecting Muon Detector data with the ADST event selection tool in Offline

The `selectADSTEvents` tool allows users to select reconstructed events from the Advanced Data Summary Trees (ADSTs) produced by $\overline{\text{Off}}\underline{\text{line}}$ [147]. It is designed to select events by using ASCII files, which the users can customize as needed. Selecting events is simple when using this tool, and so is comparing the criteria applied to different data sets. It also aids in the standardization of quality cuts. For instance, the use of `selectADSTEvents` is widespread for SD and FD, for which the ICRC ADST cut files are kept in record as the official cuts for physics analyses. Futhermore, the functionality of `selectADSTEvents` was recently extended for RD [148].

Since 29.06.2020[1], the functionality of the `selectADSTEvents` tool was also extended to MD. Applying this tool to standardize analyses is desirable, especially considering the many publications to come after years of stable data acquisition.

In Sec. A.1, we explain the available cuts for the MD. In Sec. A.2, we explain how to use them independently of and in combination with the SD cuts. In Sec. A.3, we discuss the application of cuts to obtain high-quality data sets for MD analyses. Finally, in Sec. A.4, we summarize the main conclusions.

This chapter is largely based on the Pierre Auger **internal publication**:

- *Selecting Muon Detector data with the ADST event selection tool in Offline*, F. Gesualdi, M. Roth, D. Schmidt, D. Veberič, GAP 2021 – 013 (Ref. [149]).

A.1 Description of the implemented Muon Detector cuts

The MD cuts were designed to complement the SD cuts. This is because both detectors are always used simulataneously as the MD trigger is subordinate to the SD trigger. The `allMD.cuts` file contains the description of all available MD cuts. This is the content of said file:

```
adst cuts version: 1.0
minRecLevel 1              # 0 = no Md event found, 1 = Md has triggered
                           # stations, 2 = has MLDF fit

minMdLDFStatus 1           # 0 = not reconstructed, 1 = reconstructed
```

---

[1] svn commit: r33637, revision: 3363.

```
fixBeta                      # (bool) select events with fix beta
                             # (cut or anti-cut)

minCandidateCounters 5       # minimum number of candidate counters
maxCandidateCounters 20      # maximum number of candidate counters
minCounters 5                # minimum number of total counters
maxCounters 20               # maximum number of total counters
hasCounter 1764              # select events with counterId

minZenithMD 0.               # minimum zenith angle [deg] from MD geometry
                             # reconstruction
                             # if there is no MD geometry reconstruction,
                             # the event is rejected
maxZenithMD 60.              # maximum zenith angle [deg] from MD geometry
                             # reconstruction
                             # if there is no MD geometry reconstruction,
                             # the event is rejected

hottestStationHasCounter     # (bool) select events where the hottest
                             # SD station is paired to a non-rejected
                             # MD counter (cut or anti-cut)

badPeriodsRejectionFromFile # reject bad periods defined in the file
                             # MDBadPeriodsFile in config
```

The functionality of the different cuts is the following:

- `minRecLevel`: Is a cut on the minimum reconstruction level of the MD event. Option 0 implies accepting every SD-MD event (even if the event contains no MD event), option 1 requires that at least one SD station that is paired to a non-rejected MD counter has triggered for the event, and option 2 requires that the event has a fitted muon LDF (see `MdRecLevel.h`).

- `minMdLDFStatus`: Is a cut on the minimum status of the muon LDF fit. Option 0 accepts all events regardless of their muon LDF status, and option 1 requires that the event has a fitted muon LDF. Notice that `minRecLevel` 2 and `minMdLDFStatus` 1 are equivalent.

- `fixBeta`: Is a boolean cut on the status of β, one of the parameters of the muon LDF fit. β can be fixed or free during the muon LDF fit depending on the number of counters that triggered within a distance around 450 m (see `MdLDFFinder.xml.in`). If the command `fixBeta` is excluded from the MD cuts file, all events are accepted regardless of whether β was fitted or taken fixed to the value from its parameterization. If `fixBeta` is included, only events with a fixed β are selected. If `!fixBeta` is included, only events with β having been a free parameter are selected.

- `minCandidateCounters`: Is the minimum number of candidate counters that an event needs to pass the selection.

- `maxCandidateCounters`: Is the maximum number of candidate counters that an event may have to pass the selection.

- `minCounters`: Is the minimum number of counters (saturated, candidate, silent[2], or rejected) that an event needs to pass the selection.

- `maxCounters`: Is the maximum number of counters (saturated, candidate, silent, or rejected) that an event needs to pass the selection.

- `hasCounter`: Selects events that contain the specified counter number.

- `minZenithMD`: Selects events that have a zenith angle reconstructed by the MD geometry reconstruction with a value above the specified minimum (stated in degrees). If no MD geometry reconstruction is available for the event, the event is rejected.

- `maxZenithMD`: Selects events that have a zenith angle reconstructed by the MD geometry reconstruction with a value below the specified maximum (stated in degrees). If no MD geometry reconstruction is available for the event, the event is rejected.

- `hottestStationHasCounter`: Is a boolean cut. The command `hottestStationHasCounter` selects events where the hottest SD station has a paired non-rejected MD counter. Conversely, the command `!hottestStationHasCounter` selects events where the hottest SD station is paired to a rejeced MD counter or not paired to any MD counter. If the command is excluded, all events are accepted regardless of whether the hottest SD station is paired or not to an MD counter.

- `badPeriodsRejectionFromFile`: Excludes all events from the bad periods that the user specifies in a file. Its format is identical to that of the SD bad periods files (see, for example, `ADST/Analysis/cuts/ICRC2019/BadPeriods_sd750.txt`). The MD bad periods file has to be specified as the `MDBadPeriodsFile` in the `.config` file.

It is relevant to add that the implementation of the cuts can be found in the Offline ADST trunk directory (`ADST/Analysis/src/`) in the files `MDSelection.h` and `MDSelection.cc`. The codes there implemented have to be kept in synchronization with the file `ADST/Analysis/cuts/allMD.cuts`. In other words, if a new cut is implemented, the latter three files should be modified accordingly. Finally, in order to support MD cuts, it was also necessary to adapt the following files: `ADST/Analysis/src/Analysis.cc`, `ADST/Analysis/src/Analysis.h`, `ADST/Analysis/CMakeLists.txt`, and `ADST/Analysis/examples/selectEvents.cc`.

A.2 Usage of the Muon Detector cuts

The usage of this code is analogous to that of the SD or FD cuts. A detailed description of how to use `selectADSTEvents` can be found in Refs. [150, 151]. Nevertheless, we explain the steps for applying the MD cuts, both independently of the SD cuts and in combination with them.

1. Start by having an ADST file of reconstructed SD-MD events, e.g., "`myADST.root`".

2. Create a file, e.g., "`mymdcuts.cuts`". This can be a customized version of the `allMD.cuts` file.

---

[2] We are aware that the term "silent station" in early days referred to something completely different. We still use the term since it has been widely used within the collaboration and, more importantly, within the Offline framework. We refer to silent MD counters as the counters that are paired to untriggered SD stations. Silent counters can hold zero or positive-valued signals, which cannot be accessed because their paired SD station did not trigger. Notice that MD counters that *measured* zero signal are among the candidate counters, for they are paired to SD stations that triggered.

3. (Optionally) create a file, e.g., "`mysdcuts.cuts`". This can be a customized version of the `allsdcuts.cuts` file. This step is not necessary, but would be the most common usage scenario.

4. (Optionally) create a file with the bad MD periods, e.g., "`mybadmdperiods.dat`".

5. Create a file, e.g., "`myanalysis.config`" with the line `MDCutFile mymdcuts.cuts`. In a simultaneous use of MD and SD cuts (step 3), add the line `SDCutFile mysdcuts.cuts`. If a file with bad MD periods was created (step 4), add also the line `MDBadPeriodsFile mybadmdperiods.dat`.

6. Open a terminal and run `selectADSTEvents -c myanalysis.config myADST.root`. The absolute number and percentages of the events that pass each selection cut are displayed in the output of the terminal. The output file "`selected.root`" contains the selected events.

It is important to notice that the periods contained in a bad MD periods file will exclude all SD-MD events belonging to that period. To exclude only a specific MD counter or a specific module of an MD counter during a certain period, the user should employ the `MdModuleRejector.xml` of the `MdDataReconstruction` application of $\overline{\text{Off}}\underline{\text{line}}$. Specifically, the user should add a line in that file specifying the counter id, module(s) id(s), reason of rejection, and period of rejection.

In normal usage, both SD and MD cuts have to be applied. For instance, the cuts in energy can only be set in the SD cuts file because the energy is reconstructed from SD data (see `minLgEnergySD` and `maxLgEnergySD` in the file `allSD.cuts`). Also, at the time of this GAP-note, the MD geometry reconstruction is functional for PMT data but not for SiPM data. (For this reason, in the `ModuleSequence.xml` of the `MdDataReconstruction` application, the `MdGeometryFitterAG` can be included for PMT data, but should not be included for SiPM data). Therefore, if the commands `minZenithMD` and `maxZenithMD` are included in the MD cuts file, and the cuts are applied to events with no MD geometry reconstruction, all those events will be rejected. To avoid this, the user simply has to exclude these commands from the MD cuts file. The data can be selected using the zenith angle reconstructed from SD, via the commands `minZenithSD` and `maxZenithSD` in the SD cuts file (see `allSD.cuts`).

A.3 Application to high-level analyses

As said in the previous section, the MD cuts should be used together with the standard SD cuts. The SD cuts used in the ICRC of 2015, 2017, and 2019, request events to have the highest-signal station surrounded by an hexagon of six active stations (i.e., a 6T5 trigger). This is achieved by setting `T5Trigger 2` in the SD cuts file. Said cut already ensures a good quality of the MD events if the MD modules behave normally. Unhealthy MD modules should be excluded from the analyses by having the `MdModuleRejector.xml` and the SD and MD bad period files up to date. Having said this, the most essential MD cut is `minRecLevel 1`, which ensures that there is at least one MD counter with information in the SD-MD event.

The latter strategy was used in the recently published analysis of the PMT data [90]. In that analysis, the 6T5 cut was combined with a cut on SD-reconstructed zenith angle ($\theta \leq 45°$), to minimize attenuation effects and to have a large effective detection area, and with a cut on the energy ($E \geq 2 \times 10^{17}$ eV), to ensure the full efficiency of the 750 m array [111].

An additional way of ensuring a high quality of the SD-MD events is by using the command `hottestStationHasCounter`. It is desirable that the core of an event lies within or in the direct vecinity of the region enclosed by the working detectors. Even if the SD event qualifies as 6T5, its core could fall outside the area enclosed by the MD counters, and still trigger some MD counters (more precisely, trigger their paired SD stations). Then, those counters would only sample the tail of the muon LDF of said event, and the estimated muon density at 450 m (ρ_{450}) would be obtained by extrapolation, which could lead to biases. A good way to avoid these cases is by requiring the hottest SD station to be paired to a non-rejected MD counter. This ensures that the core is within the region enclosed by the working MD counters, or in its direct vecinity, i.e., within an hexagonal cell which has in the center a working MD counter.

More restrictive conditions can be applied to ensure an even better sampling of the muon LDF. This could be relevant for studies of the muon LDF of individual events, or to have a reliable event-by-event estimate of ρ_{450}. One way to try to achieve this would be by using a cut on `minCandidateCounters`. However, a high number of minimum candidate counters can exclude entire periods of data acquisition. This can happen if that period is in coincidence with that of a rejected counter in the `MdModuleRejector.xml`, especially if the counter belongs to the inner unitary cell of the MD array. A more recommendable way to achieve this is to select events were β was set free during the muon LDF reconstruction, by setting `!fixBeta` in the MD cuts file. This implies a condition on the number of candidate counters depending on their distance to 450 m, the minimum being three candidate counters.

A.4 Conclusions and summary

The tool `selectADSTEvents` is now fully functional and ready to be applied for the selection of MD events. We provided a detailed description of the implemented cuts. We explained how to use them with or without SD cuts. In the most common usage conditions, the MD cuts are applied in combination with SD cuts, for example, to select on the SD-reconstructed energy and zenith angle. We also discussed on how to apply quality cuts for high-level physics analyses. For the latter, particularly relevant are the MD cuts: on MD event reconstruction level, on whether the hottest SD station is paired to a non-rejected MD counter (boolean cut), and on whether the slope of the muon LDF was free during the fit (boolean cut). We believe that the extended functionality of the `selectADSTEvents` tool helps users to apply cuts to MD data in a simpler way, and most importantly, facilitates the standardization of the MD data sets used for physics analyses.

Appendix B

Additional results of the muon deficit inferred from measurements of the Underground Muon Detector of the Pierre Auger Observatory

B.1 Muon density as a function of the distance to the shower axis

Figures B.1 to B.5 show the average muon density divided by the average muon density of EPOS-LHC proton air-showers as a function of the distance to the shower axis, for data, proton, iron, and mixed composition scenarios, for the reconstructed energy bins centered at $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.32$, $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.72$, $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.92$, $\log_{10}(E_{Ri}/\mathrm{eV}) = 18.12$ and $\log_{10}(E_{Ri}/\mathrm{eV}) = 18.32$. This is the same as what is shown in Fig. 7.2 but for the mentioned energies instead of $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.52$.

B.2 An alternative estimation of the z-values from the underground muon detector data

In Sec. 8.2.4 we defined an estimator of the muon scale by taking the mean of the logarithm of the muon number,

$$z_{\langle \ln \cdot \rangle} = \frac{\langle \ln N^{\mathrm{det}}_{\mu,\,\mathrm{data}} \rangle - \langle \ln N^{\mathrm{det}}_{\mu,\,\mathrm{p}} \rangle}{\langle \ln N^{\mathrm{det}}_{\mu,\,\mathrm{Fe}} \rangle - \langle \ln N^{\mathrm{det}}_{\mu,\,\mathrm{p}} \rangle}. \tag{B.1}$$

To be able to compute $z_{\langle \ln \cdot \rangle}$ for the UMD SiPM data, we need further approximations, because we cannot analytically compute $\langle \ln N^{\mathrm{det}}_{\mu,\,\{\mathrm{p,Fe}\}} \rangle$ using Eq. (7.1). We therefore approximate, in all terms

$$\langle \ln N_\mu \rangle \sim \ln \langle N_\mu \rangle - \tfrac{1}{2} \left(\mathrm{RSD}_{\mathrm{tot}}[N_\mu] \right)^2, \tag{B.2}$$

where $\mathrm{RSD}_{\mathrm{tot}}[N_\mu] = \sigma_{\mathrm{tot}}(N_\mu) / \langle N_\mu \rangle$, and where we drop $\mathcal{O}\left(\langle \mathrm{RSD}^3_{\mathrm{tot}}[N_\mu] \rangle \right)$ terms. Then, using the muon densities instead of the muon numbers, Eq. (B.1) can be approximated to

$$z_{\langle \ln \cdot \rangle} \approx \frac{\ln \left\langle \rho^{\mathrm{det}}_{\mu\,\mathrm{r},\,\mathrm{data}} \right\rangle - \ln \left\langle \rho_{\mu\,\mathrm{r},\,\mathrm{p}} \right\rangle + \frac{1}{2} \left[\left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu\,\mathrm{r},\,\mathrm{p}}] \right)^2 - \left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu\,\mathrm{r},\,\mathrm{mass}}] \right)^2 \right]}{\ln \left\langle \rho_{\mu\,\mathrm{r},\,\mathrm{Fe}} \right\rangle - \ln \left\langle \rho_{\mu\,\mathrm{r},\,\mathrm{p}} \right\rangle + \frac{1}{2} \left[\left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu\,\mathrm{r},\,\mathrm{p}}] \right)^2 - \left(\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu\,\mathrm{r},\,\mathrm{Fe}}] \right)^2 \right]}, \tag{B.3}$$

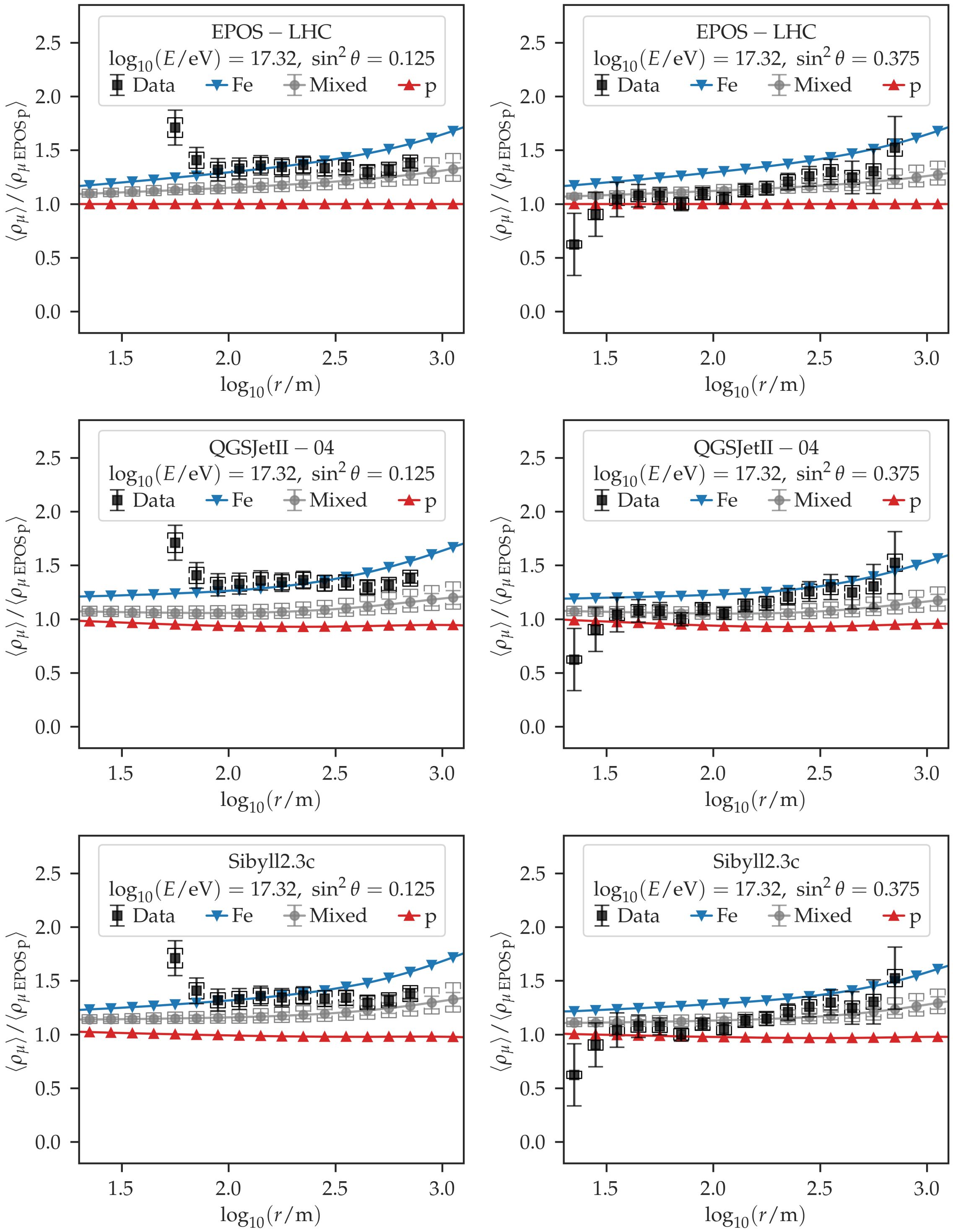


Figure B.1: Same as Fig. 7.2 but for the reconstructed energy bin centered at $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.32$.

where we use in the numerator and denominator that all sources of fluctuations except shower-to-shower cancel out. Although we could compute $\left\langle \ln \rho^{\mathrm{det}}_{\mu\,\mathrm{r,data}} \right\rangle$, this would require neglecting null muon density measurements. In this case, the effect of excluding these measurements introduces a considerable bias, and for this reason we compute $\ln \left\langle \rho^{\mathrm{det}}_{\mu\,\mathrm{r,data}} \right\rangle$

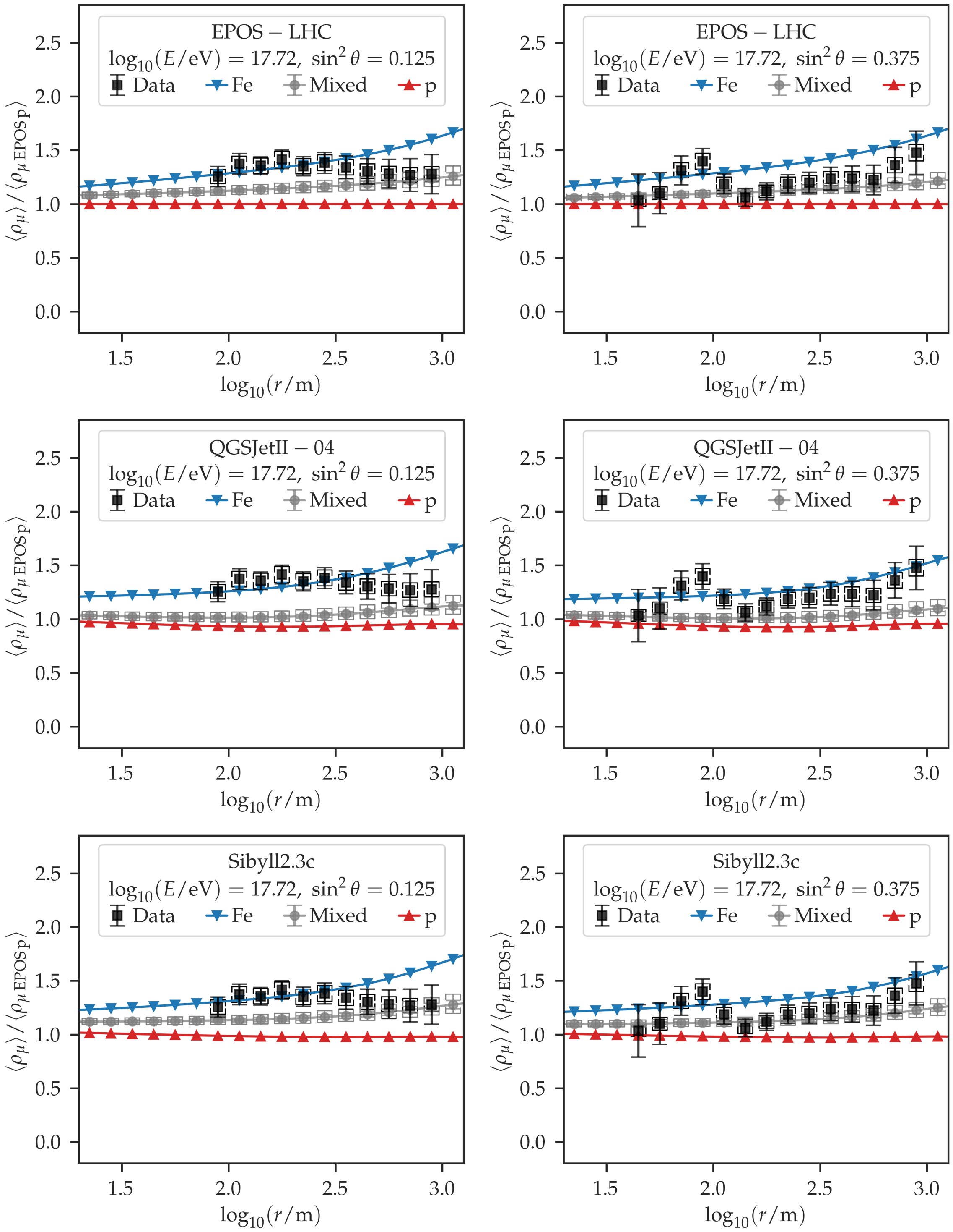


Figure B.2: Same as Fig. 7.2 but for the reconstructed energy bin centered at $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.72$.

instead, including the null density measurements. Like in Sec. 7.1.4, we take the muon densities at $r \sim 450\,\mathrm{m}$ from the shower axis, for events with zenith angles in $0^\circ \leq \theta < 45^\circ$. We compute $\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu\,\mathrm{r},\,\{\mathrm{p,Fe}\}}]$ from the proton and iron Monte Carlo simulations of each hadronic interaction model. Furthermore, we compute $\mathrm{RSD}_{\text{sh-sh}}[\rho_{\mu\,\mathrm{r},\,\mathrm{mass}}]$ from simulations assuming a mixed composition scenario, by fitting a normal distribution to the weighted

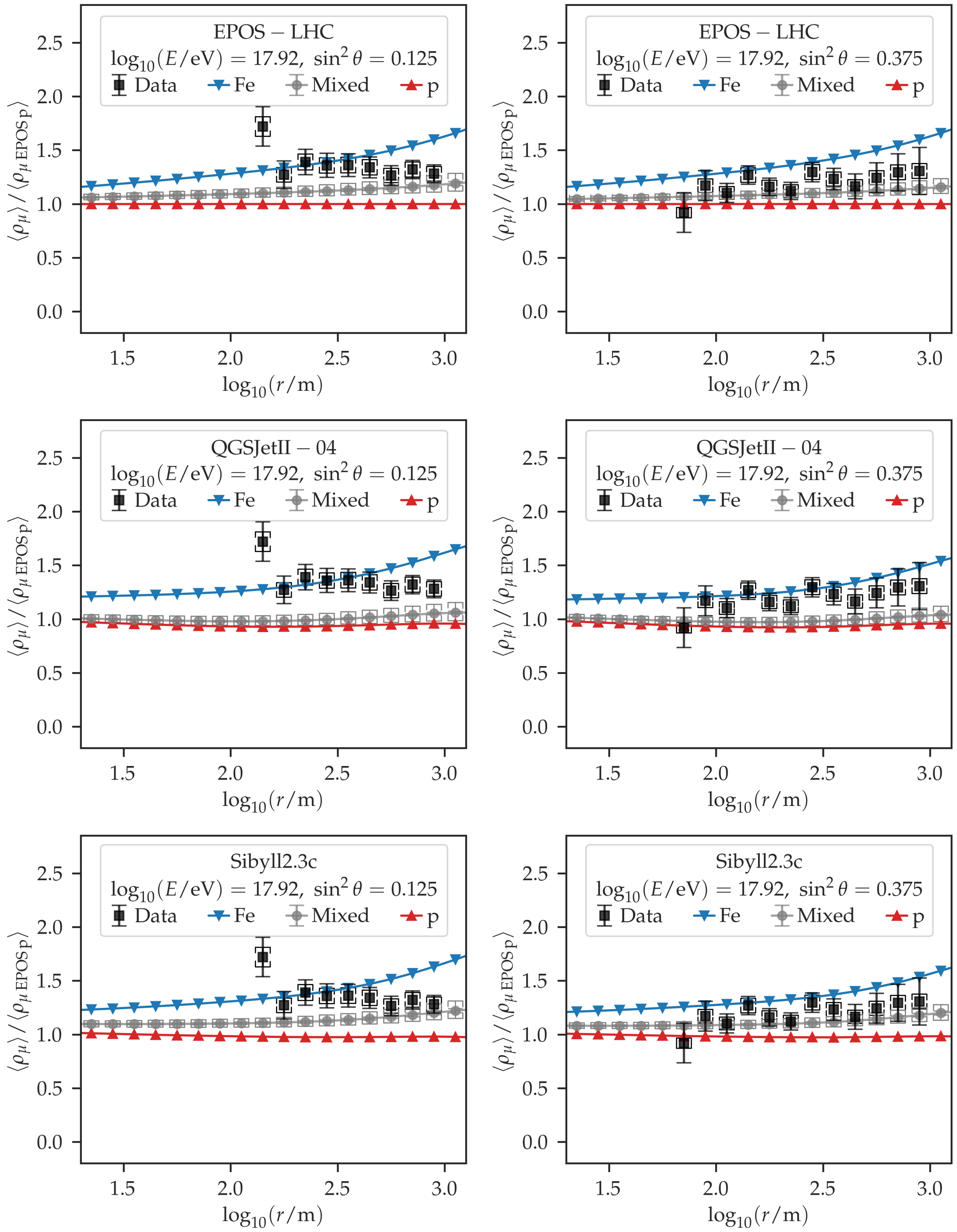


Figure B.3: Same as Fig. 7.2 but for the reconstructed energy bin centered at $\log_{10}(E_{Ri}/\mathrm{eV}) = 17.92$.

sum of the single-primary distributions of the simulated muon densities, where the weights are the mass fractions given by the Global Spline Fit (GSF) model [61]. We use the aforementioned composition model to allow a consistent comparison of the here calculated z-values with those of other experiments, as presented in Refs. [5, 59, 146].

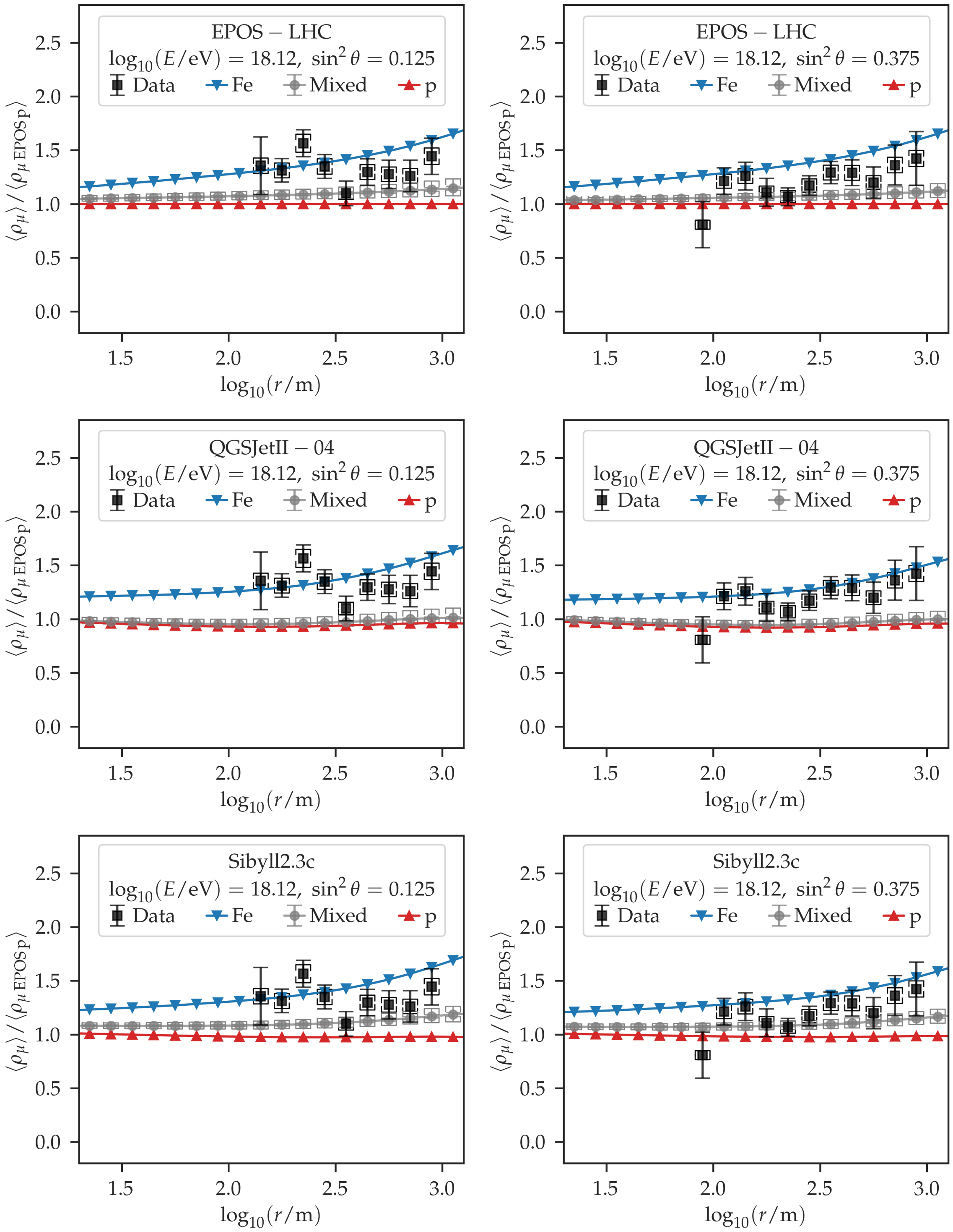


Figure B.4: Same as Fig. 7.2 but for the reconstructed energy bin centered at $\log_{10}(E_{Ri}/\mathrm{eV}) = 18.12$.

Figure B.6 shows the comparison between $z_{\langle \ln \cdot \rangle}$, as computed from Eq. B.3, and $z_{\ln \langle \cdot \rangle}$, as computed from Eq. 7.16. We can see that the difference between the two estimators is truly small compared to the uncertainties. In all cases the absolute difference is below ±0.04 in the z-scale, and in relative values, as it can be seen in the figure, below ±0.3 %.

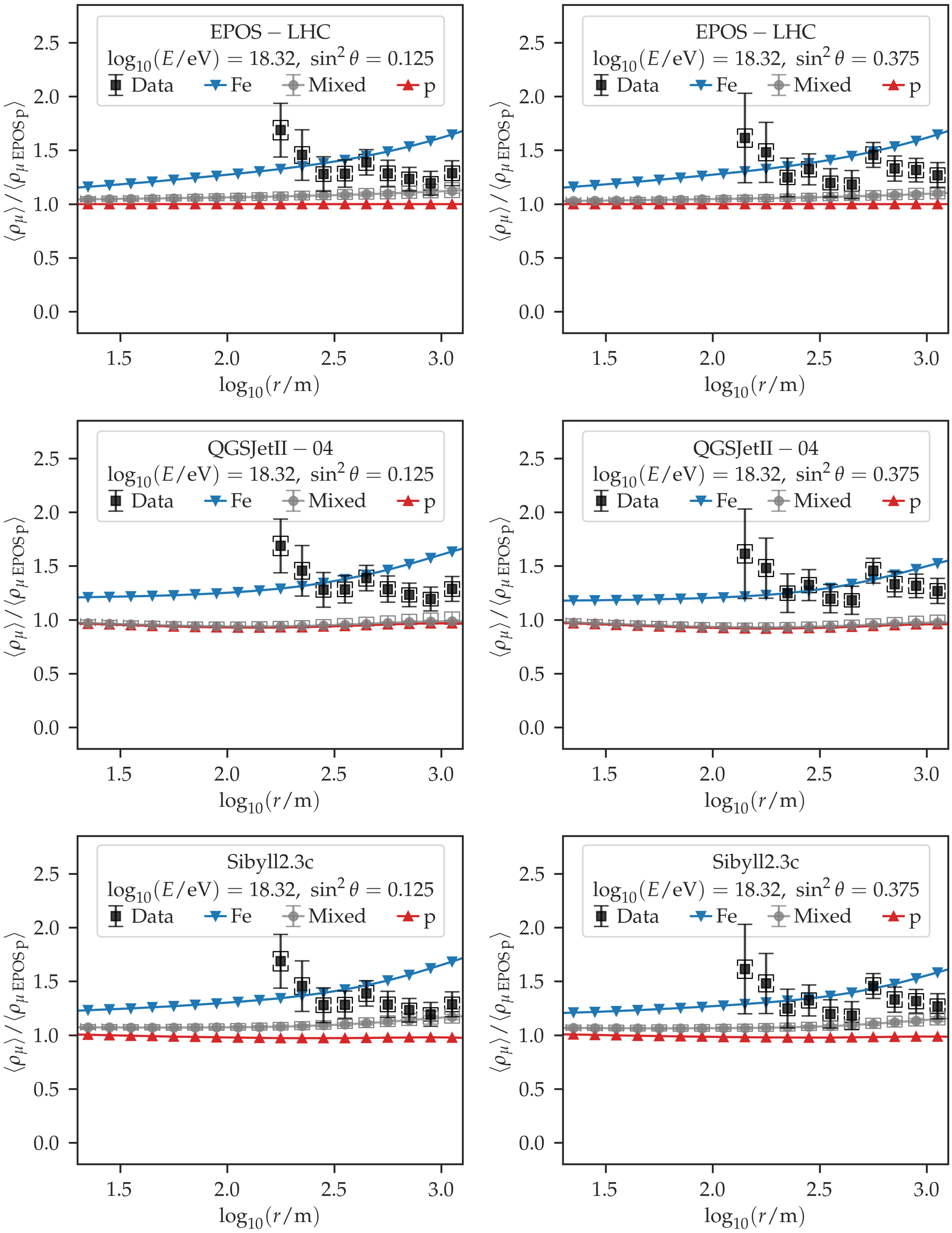


Figure B.5: Same as Fig. 7.2 but for the reconstructed energy bin centered at $\log_{10}(E_{Ri}/\mathrm{eV}) = 18.32$.

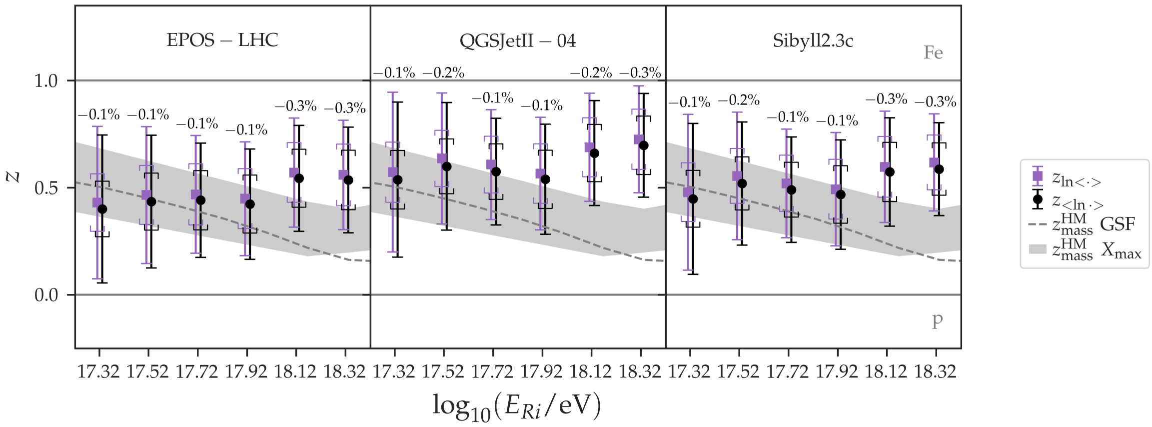


Figure B.6: $z_{\ln\langle\cdot\rangle}$ (squares) and $z_{\langle\ln\cdot\rangle}$ (circles), for EPOS-LHC (left), QGSJetII-04 (center), and Sibyll2.3c (right) as a function of the logarithmic reconstructed energy (in the reference scale). For clarity, we introduce a horizontal displacement between the two sets of values. The relative difference of $z_{\langle\ln\cdot\rangle}$ to $z_{\ln\langle\cdot\rangle}$ is written on top of each pair of points. We also show the values of $z_{\text{mass}}^{\text{HM}}$ from the GSF model (gray dashed lines) and from the Pierre Auger $X_{\max}$ (gray shaded area). Gray horizontal lines mark the expected values of z for pure proton and pure iron compositions (0 and 1 respectively). The statistic uncertainties are represented with error bars and the systematic uncertainties with brackets.

Acknowledgments

In first place, I would like to greatly thank my supervisor Alberto Daniel Supanitsky. I am grateful for his continuous guidance and support, for providing me the structure that allowed me to evolve into the scientist I wanted to become, for taking into account my points of view, and for giving me the freedom to develop my ideas while guiding me towards useful contributions. I feel truly grateful for having had the opportunity to work with him.

Furthermore, I would like to thank Ralph Engel for facilitating this work and for his guidance in relation to this thesis. I am also grateful to Alberto Etchegoyen and Johannes Blümer for having created this Double Doctoral degree in Astrophysics. I would also like to thank the members of the funding agents, UNSAM, DAAD, and HIRSAP, for acknowledging the relevance of this program and for supporting it. I am also thankful to the members of CNEA and Conicet for granting me with this PhD position and a workplace. Thanks to my advisor, Markus Roth, for his guidance and extensive support. Thanks to Sabine Bucher and Marie-Christine Mundt for being always there to help.

I am also grateful towards David Schmidt, for his support and help in understanding Offline. Thanks to Johannes Hulsman, Steffen Hahn, and Max Reininghaus, for introducing me to ROOT, Python, and Corsika respectively, and most importantly for all the good fun times. Thanks to Alexander Streich, Ana Laura Müller, Tobias Schultz, Kathrin Bismark, and Luca Deval for the nice moments shared. Thanks to Ana Martina Botti for passing on her knowledge to me, and to Joaquín De Jesús for useful discussions and for his openness and generosity. Thanks to Adrián Sedoski for being always available to help on any IT related matter.

I am grateful towards the people with whom I shared this journey, including the students, researchers, and staff of both ITeDA and KIT, the members of the Pierre Auger Collaboration, and the members of the WHISP group. In particular, I would like to acknowledge Kenji Shinozaki for generously sharing his knowledge of AGASA with me, as well as Diego Ravignani for useful discussions.

My sincere thanks go to Marcelo Solimano, for his amazing generosity of inviting me to his physics lectures while I was at high-school. His passion for physics left an unavoidable footprint in me that guided me towards pursuing this career.

Thanks to my friends from Instituto Balseiro, Marco Amelio, Agustín Dall'Alba, and Juan Manuel González, for being my friends. Thanks to my friend Agustín Somacal for always being there.

My deepest gratitude is to Martin Leo Schimassek, in first place, for the uncountable discussions about my work, listening to me and giving me valuable feedback, greatly helping me move forward faster and have more confidence in the design and interpretation of my analyses. But I want to thank him most importantly for supporting me so much every single day, for always helping me see the positive sides, for being so caring and making me feel at home, for making me a better person, and for the beautiful love we share. I love you! I am also very grateful to his siblings Rudolf Schimassek and Karin Schimassek, and his parents Irene Schimassek and Walter Schimassek, for being so immensely warm and generous to me. I want to thank my family, specially my aunt Liliana Teresa Ongini and my grandmother Ana María Robles, for always caring for me. Thanks to Pedro Gesualdi for his contagious joy and loyal company. I want to thank my brother Augusto Gesualdi for being an example for me and for guiding me towards broader horizons. Finally, I am immensely grateful towards my parents, Silvia Graciela Ongini and Ernesto Marcelo Gesualdi, for their unconditional love and day-by-day support, for always taking care of me and giving me everything I need and more, and for always encouraging me to follow what I love.

Thank you!

Bibliography

[1] A. Aab, et al., Pierre Auger, The Pierre Auger Observatory Upgrade - Preliminary Design Report (2016)`arXiv:1604.03637`.

[2] B. Peters, Primary cosmic radiation and extensive air showers, Il Nuovo Cimento 22 (4) (1961) 800–819. `doi:10.1007/BF02783106`.

[3] A. Aab, et al., Combined fit of spectrum and composition data as measured by the Pierre Auger Observatory, Journal of Cosmology and Astroparticle Physics 2017 (04) (2017) 038–038. `doi:10.1088/1475-7516/2017/04/038`.
URL `https://doi.org/10.1088/1475-7516/2017/04/038`

[4] R. R. Prado, Tests of hadronic interactions with measurements by pierre auger observatory, EPJ Web Conf. 208 (2019) 08003. `doi:10.1051/epjconf/201920808003`.
URL `https://doi.org/10.1051/epjconf/201920808003`

[5] H. P. Dembinski *et al.* for the EAS-MSU, IceCube, KASCADE-Grande, NEVOD-DECOR, PierreAuger, SUGAR, Telescope Array, Yakutsk EAS Array, Report on tests and measurements of hadronic interaction properties with air showers, EPJ Web Conf. 210 (2019) 02004. `doi:10.1051/epjconf/201921002004`.
URL `https://doi.org/10.1051/epjconf/201921002004`

[6] A. Aab, et al., Pierre Auger Collaboration, Muons in air showers at the pierre auger observatory: Mean number in highly inclined events, Phys. Rev. D 91 (2015) 032003. `doi:10.1103/PhysRevD.91.032003`.
URL `https://link.aps.org/doi/10.1103/PhysRevD.91.032003`

[7] A. Aab, et al., Pierre Auger Collaboration, Testing hadronic interactions at ultrahigh energies with air showers measured by the pierre auger observatory, Phys. Rev. Lett. 117 (2016) 192001. `doi:10.1103/PhysRevLett.117.192001`.
URL `https://link.aps.org/doi/10.1103/PhysRevLett.117.192001`

[8] A. Aab, et al., Pierre Auger Collaboration, Muons in air showers at the pierre auger observatory: Mean number in highly inclined events, Phys. Rev. D 91 (2015) 032003. `doi:10.1103/PhysRevD.91.032003`.
URL `https://link.aps.org/doi/10.1103/PhysRevD.91.032003`

[9] A. Aab, et al., Pierre Auger Collaboration, Measurement of the fluctuations in the number of muons in extensive air showers with the Pierre Auger Observatory, Phys. Rev. Lett. 126 (2021) 152002. `doi:10.1103/PhysRevLett.126.152002`.
URL `https://link.aps.org/doi/10.1103/PhysRevLett.126.152002`

[10] A. Castellina, Pierre Auger, AugerPrime: the Pierre Auger Observatory Upgrade, EPJ Web Conf. 210 (2019) 06002. `arXiv:1905.04472`, `doi:10.1051/epjconf/201921006002`.

[11] A. Aab, et al., Calibration of the underground muon detector of the pierre auger observatory, Journal of Instrumentation 16 (04) (2021) P04003. `doi:10.1088/1748-0221/16/04/p04003`.
URL `https://doi.org/10.1088/1748-0221/16/04/p04003`

[12] N. Hayashida, et al., Muons ($\geqslant$ 1 GeV) in large extensive air showers of energies between $10^{16.5}$ eV and $10^{19.5}$ eV observed at Akeno, J. Phys. G: Nucl. Part. Phys. 21 (1995) 1101–1119.

[13] D. Ravignani, and A.D. Supanitsky, A new method for reconstructing the muon lateral distribution with an array of segmented counters, Astroparticle Physics 65 (2014). `doi:10.1016/j.astropartphys.2014.11.007`.

[14] D. Ravignani, A.D. Supanitsky, and D. Melo, Reconstruction of air shower muon densities using segmented counters with time resolution, Astroparticle Physics 82 (2016) 108–116. `doi:https://doi.org/10.1016/j.astropartphys.2016.06.001`.
URL `https://www.sciencedirect.com/science/article/pii/S0927650516300809`

[15] B. Andrada and F. Sánchez, Top-Down analysis at $10^{17.5}$ eV, Auger internal note GAP-2021-023 (2021).

[16] T. K. Gaisser, R. Engel, E. Resconi, Cosmic Rays and Particle Physics, 2nd Edition, Cambridge University Press, 2016. `doi:10.1017/CBO9781139192194`.

[17] C. Grupen, Astroparticle Physics, Springer Nature, 2020.

[18] F. Gesualdi, Determination of the Energy Scale of the Pierre Auger Observatory with the Scintillator Surface Detector, Ph.D. thesis, Universidad Nacional de Cuyo, Instituto Balseiro (2017).
URL `http://ricabib.cab.cnea.gov.ar/676/1/Gesualdi.pdf`

[19] A. Kheirandish, Identifying galactic sources of high-energy neutrinos, Astrophysics and Space Science 365 (2020). `doi:10.1007/s10509-020-03816-3`.

[20] A. Aab, et al., Pierre Auger, Large-scale cosmic-ray anisotropies above 4 EeV measured by the Pierre Auger Observatory, Astrophys. J. 868 (1) (2018) 4. `arXiv:1808.03579`, `doi:10.3847/1538-4357/aae689`.

[21] W. Heitler, The quantum theory of radiation, Oxford University Press (1954).

[22] J. Matthews, A Heitler model of extensive air showers, Astroparticle Physics 22 (5) (2005) 387–397. `doi:https://doi.org/10.1016/j.astropartphys.2004.09.003`.
URL `https://www.sciencedirect.com/science/article/pii/S0927650504001598`

[23] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, T. Thouw, CORSIKA: A Monte Carlo code to simulate extensive air showers, Tech. Rep. FZKA-6019, Forschungszentrum Karlsruhe (1998). `doi:10.5445/IR/270043064`.

[24] S. J. Sciutto, AIRES: A system for air shower simulations (1999)`arXiv:astro-ph/9911331`, `doi:10.13140/RG.2.2.12566.40002`.

[25] T. Bergmann, R. Engel, D. Heck, N. N. Kalmykov, S. S. Ostapchenko, T. Pierog, T. Thouw, K. Werner, T. Bergmann, R. Engel, D. Heck, N. Kalmykov, S. Ostapchenko, T. Pierog, T. Thouw, K. Werner, One-dimensional Hybrid Approach to Extensive Air Shower Simulation, Astropart. Phys. 26 (6) (2007) 420 – 432, astro-ph/0606564. `doi:DOI:10.1016/j.astropartphys.2006.08.005`.

[26] T. Pierog, Iu. Karpenko, J. M. Katzy, E. Yatsenko, K. Werner, EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider, Phys. Rev. C92 (3) (2015) 034906. `arXiv:1306.0121`, `doi:10.1103/PhysRevC.92.034906`.

[27] S. Ostapchenko, Qgsjet-ii: physics, recent improvements, and results for air showers, EPJ Web of Conferences 52 (2013) 02001. `doi:10.1051/epjconf/20125202001`.
URL `https://doi.org/10.1051/epjconf/20125202001`

[28] F. Riehn, H. P. Dembinski, R. Engel, A. Fedynitch, T. K. Gaisser, T. Stanev, The hadronic interaction model Sibyll 2.3c and Feynman scaling, in: 35th International Cosmic Ray Conference (ICRC2017), Vol. 301 of International Cosmic Ray Conference, 2017, p. 301. `arXiv:1709.07227`.

[29] H. Fesefeldt, Report **PITHA-85/02**, RWTH Aachen (1985).

[30] T. T. Böhlen, F. Cerutti, M. P. W. Chin, A. Fassò, A. Ferrari, P. G. Ortega, A. Mairani, P. R. Sala, G. Smirnov, V. Vlachoudis, The FLUKA Code: Developments and Challenges for High Energy and Medical Applications, Nuclear Data Sheets 120 (2014) 211–214. `doi:10.1016/j.nds.2014.07.049`.

[31] A. Ferrari, P. R. Sala, A. Fassò, J. Ranft, FLUKA: A multi-particle transport code (program version 2005), CERN Yellow Reports: Monographs, CERN, Geneva, 2005. `doi:10.5170/CERN-2005-010`.
URL `https://cds.cern.ch/record/898301`

[32] S. Bass, M. Belkacem, M. Bleicher, M. Brandstetter, L. Bravina, C. Ernst, L. Gerland, M. Hofmann, S. Hofmann, J. Konopka, G. Mao, L. Neise, S. Soff, C. Spieles, H. Weber, L. Winckelmann, H. Stöcker, W. Greiner, C. Hartnack, J. Aichelin, N. Amelin, Microscopic models for ultrarelativistic heavy ion collisions, Progress in Particle and Nuclear Physics 41 (1998) 255–369. `doi:https://doi.org/10.1016/S0146-6410(98)00058-1`.
URL `https://www.sciencedirect.com/science/article/pii/S0146641098000581`

[33] M. Bleicher, E. Zabrodin, C. Spieles, S. A. Bass, C. Ernst, S. Soff, L. Bravina, M. Belkacem, H. Weber, H. Stöcker, W. Greiner, Relativistic hadron-hadron collisions in the ultra-relativistic quantum molecular dynamics model, Journal of Physics G: Nuclear and Particle Physics 25 (9) (1999) 1859–1896. `doi:10.1088/0954-3899/25/9/308`.
URL `https://doi.org/10.1088/0954-3899/25/9/308`

[34] R. Battiston, AMS 02, The antimatter spectrometer (AMS-02): A particle physics detector in space, Nucl. Instrum. Meth. A 588 (2008) 227–234. `doi:10.1016/j.nima.2008.01.044`.

[35] E. S. Seo, et al., CREAM: 70 days of flight from 2 launches in Antarctica, Adv. Space Res. 42 (10) (2008) 1656–1663. `doi:10.1016/j.asr.2007.03.056`.

[36] J. Chang, et al., DAMPE, The DArk Matter Particle Explorer mission, Astropart. Phys. 95 (2017) 6–24. `arXiv:1706.08453`, `doi:10.1016/j.astropartphys.2017.08.005`.

[37] O. Adriani, et al., The PAMELA experiment on satellite and its capability in cosmic rays measurements, Nucl. Instrum. Meth. A 478 (2002) 114–118. `doi:10.1016/S0168-9002(01)01726-0`.

[38] A. Aab, et al., Pierre Auger, The Pierre Auger Cosmic Ray Observatory, Nucl. Instrum. Meth. A 798 (2015) 172–213. `arXiv:1502.01323`, `doi:10.1016/j.nima.2015.06.058`.

[39] H. Kawai, et al., Telescope array experiment, Nuclear Physics B - Proceedings Supplements 175-176 (2008) 221–226, proceedings of the XIV International Symposium on Very High Energy Cosmic Ray Interactions. `doi:https://doi.org/10.1016/j.nuclphysbps.2007.11.002`.
URL `https://www.sciencedirect.com/science/article/pii/S0920563207007992`

[40] J. A. Hinton, H.E.S.S., The Status of the H.E.S.S. project, New Astron. Rev. 48 (2004) 331–337. `arXiv:astro-ph/0403052`, `doi:10.1016/j.newar.2003.12.004`.

[41] S. F. Berezhnev, et al., The Tunka-133 EAS Cherenkov light array: status of 2011, Nucl. Instrum. Meth. A 692 (2012) 98–105. `arXiv:1201.2122`, `doi:10.1016/j.nima.2011.12.091`.

[42] G. A. Askaryan, Excess negative charge of an electron-photon shower and its coherent radio emission, Soviet Physic JETP (1962).

[43] S. Knurenko, Z. Petrov, I. Petrov, Radio emission of air showers with extremely high energy measured by the yakutsk radio array, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 866 (2017). `doi:10.1016/j.nima.2017.04.033`.

[44] O. Scholten, et al., The initial stage of cloud lightning imaged in high-resolution, J. Geophys. Research Atm. 126 (4) (2021) e2020JD033126, e2020JD033126 2020JD033126. `doi:https://doi.org/10.1029/2020JD033126`.

[45] N. Hayashida, et al., Muons ($\geqslant$ 1 gev) in large extensive air showers of energies between $10^{16.5}$ ev and $10^{19.5}$ ev observed at akeno, J. Phys. G: Nucl. Part. Phys. 21 (1995) 1101–1119.

[46] E. Fermi, On the Origin of the Cosmic Radiation, Phys. Rev. 75 (1949) 1169–1174. `doi:10.1103/PhysRev.75.1169`.

[47] A. Arámburo-García, K. Bondarenko, A. Boyarsky, D. Nelson, A. Pillepich, A. Sokolenko, Ultrahigh energy cosmic ray deflection by the intergalactic magnetic field, Phys. Rev. D 104 (2021) 083017. `doi:10.1103/PhysRevD.104.083017`.
URL `https://link.aps.org/doi/10.1103/PhysRevD.104.083017`

[48] V. S. Berezinskii, S. I. Grigor'eva, A bump in the ultra-high energy cosmic ray spectrum, Astronomy and Astrophysics 199 (1-2) (1988) 1–12.

[49] G. Matthiae, The cosmic ray energy spectrum as measured using the pierre auger observatory, New Journal of Physics 12 (7) (2010) 075009. `doi:10.1088/1367-2630/12/7/075009`.
URL `https://dx.doi.org/10.1088/1367-2630/12/7/075009`

[50] A. M. Hillas, The origin of ultra-high-energy cosmic rays, Ann. Rev. Astron. Astrophys. 22 (1) (1984) 425–444.

[51] M. Aartsen, et al., Astrophysical neutrinos and cosmic rays observed by icecube, Advances in Space Research 62 (10) (2018) 2902–2930, origins of Cosmic Rays. `doi:https://doi.org/10.1016/j.asr.2017.05.030`.
URL `https://www.sciencedirect.com/science/article/pii/S0273117717303757`

[52] J. Abraham *et al.* (Pierre Auger Collab.), Upper limit on the cosmic-ray photon fraction at eev energies from the pierre auger observatory, Astroparticle Physics 31 (6) (2009) 399 – 406. `doi:DOI:10.1016/j.astropartphys.2009.04.003`.

[53] J. Abraham *et al.* (Pierre Auger Collab.), Upper limit on the diffuse flux of UHE tau neutrinos from the Pierre Auger Observatory, Phys. Rev. Lett. 100 (2008) 211101, arXiv astro-ph/0712.1909. `doi:10.1103/PhysRevLett.100.211101`.

[54] R. R. Prado, Tests of hadronic interactions with measurements by Pierre Auger Observatory, in: European Physical Journal Web of Conferences, Vol. 208 of European Physical Journal Web of Conferences, 2019, p. 08003. `arXiv:1810.00586`, `doi:10.1051/epjconf/201920808003`.

[55] A. Supanitsky, A. Etchegoyen, G. Medina-Tanco, I. Allekotte, M. Gómez Berisso, M. Medina, Underground muon counters as a tool for composition analyses, Astroparticle Physics 29 (6) (2008) 461–470. `doi:https://doi.org/10.1016/j.astropartphys.2008.05.003`.
URL `https://www.sciencedirect.com/science/article/pii/S0927650508000741`

[56] A. D. Supanitsky, Determination of the cosmic-ray chemical composition: Open issues and prospects, Galaxies 10 (3) (2022). `doi:10.3390/galaxies10030075`.
URL `https://www.mdpi.com/2075-4434/10/3/75`

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

[58] T. Pierog, Air Shower Simulation with a New Generation of post-LHC Hadronic Interaction Models in CORSIKA, in: 35th International Cosmic Ray Conference (ICRC2017), Vol. 301 of International Cosmic Ray Conference, 2017, p. 1100.

[59] L. Cazon, Working group report on the combined analysis of muon density measurements from eight air shower experiments, Vol. ICRC2019, 2019, p. 214. `doi:10.22323/1.358.0214`.

[60] R. U. Abbasi, et al., Telescope Array Collaboration, Study of muons from ultrahigh energy cosmic ray air showers measured with the telescope array experiment, Phys. Rev. D 98 (2018) 022002. `doi:10.1103/PhysRevD.98.022002`.
URL `https://link.aps.org/doi/10.1103/PhysRevD.98.022002`

[61] H. P. Dembinski, R. Engel, A. Fedynitch, T. Gaisser, F. Riehn, T. Stanev, Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV, PoS ICRC2017 (2018) 533. `arXiv:1711.11432`, `doi:10.22323/1.301.0533`.

[62] I. Allekotte *et al.* (Pierre Auger Collab.), The surface detector system of the pierre auger observatory, Nucl. Instrum. Meth. A586 (2007) 409–420, arXiv astro-ph/0712.2832. `doi:10.1016/j.nima.2007.12.016`.

[63] P. Sommers, First Estimate of the Primary Cosmic Ray Energy Spectrum above 3 EeV from the Pierre Auger Observatory, in: 29th International Cosmic Ray Conference (ICRC29), Volume 7, Vol. 7 of International Cosmic Ray Conference, 2005, p. 387.

[64] D. Veberič (2021). [link].
URL `https://web.ikp.kit.edu/darko/auger/auger-array/auger_array-pdf/auger_array-ad.pdf`

[65] J. Abraham *et al.* (Pierre Auger Collab.), Trigger and aperture of the surface detector array of the Pierre Auger Observatory, Nucl. Instrum. Meth. A613 (2010) 29–39.

[66] A. Aab, et al., Pierre Auger, Reconstruction of events recorded with the surface detector of the Pierre Auger Observatory, JINST 15 (10) (2020) P10021. `arXiv:2007.09035`, `doi:10.1088/1748-0221/15/10/P10021`.

[67] I. Allekotte, et al., Pierre Auger, The Surface Detector System of the Pierre Auger Observatory, Nucl. Instrum. Meth. A 586 (2008) 409–420. `arXiv:0712.2832`, `doi:10.1016/j.nima.2007.12.016`.

[68] Martin Leo Schimassek, Exploiting the trigger information of the pierre auger observatory, phdthesis (2022).

[69] D. Newton, J. Knapp, A. A. Watson, The Optimum Distance at which to Determine the Size of a Giant Air Shower, Astropart. Phys. 26 (2007) 414–419.

[70] S. Argiro, L. Nellen, T. Paul, T. Porter, L. Prado, The offline software framework of the pierre auger observatory, in: IEEE Symposium Conference Record Nuclear Science 2004., Vol. 3, 2004, pp. 1862–1866 Vol. 3. `doi:10.1109/NSSMIC.2004.1462607`.

[71] J. Abraham *et al.* (Pierre Auger Collab.), The fluorescence detector of the Pierre Auger Observatory, Nucl. Instrum. Meth. A620 (2010) 227 – 251.

[72] C. Meurer, N. f. t. P. A. C. Scharf, Heat - a low energy enhancement of the pierre auger observatory, Astrophysics and Space Sciences Transactions 7 (2011). `doi:10.5194/astra-7-183-2011`.

[73] T. K. Gaisser, A. M. Hillas, Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers, in: International Cosmic Ray Conference, Vol. 8 of International Cosmic Ray Conference, 1977, p. 353.

[74] A. Aab, et al., Pierre Auger, Depth of Maximum of Air-Shower Profiles at the Pierre Auger Observatory: Measurements at Energies above $10^{17.8}$ eV, Phys. Rev. D 90 (12) (2014) 122005. `arXiv:1409.4809`, `doi:10.1103/PhysRevD.90.122005`.

[75] A. Taboada, Pierre Auger, Analysis of Data from Surface Detector Stations of the AugerPrime Upgrade, PoS ICRC2019 (2020) 434. `doi:10.22323/1.358.0434`.

[76] R. Šmída, Pierre Auger, Scintillator detectors of AugerPrime Scintillator detectors of AugerPrime, PoS ICRC2017 (2018) 390. `doi:10.22323/1.301.0390`.

[77] G. Cataldi, et al., Pierre Auger, The upgrade of the Pierre Auger Observatory with the Scintillator Surface Detector, PoS ICRC2021 (2021) 251. `doi:10.22323/1.395.0251`.

[78] F. Schlüter, Pierre Auger, Expected performance of the AugerPrime Radio Detector, PoS ICRC2021 (2021) 262. `doi:10.22323/1.395.0262`.

[79] P. Abreu, et al., Pierre Auger, AugerPrime Upgraded Electronics, Vol. ICRC2021, 2021, p. 230. `doi:10.22323/1.395.0230`.

[80] A. Aab, et al., Calibration of the underground muon detector of the pierre auger observatory, Journal of Instrumentation 16 (04) (2021) P04003. `doi:10.1088/1748-0221/16/04/p04003`.
URL `https://doi.org/10.1088/1748-0221/16/04/p04003`

[81] 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. thesis, Karlsruher Institut für Technologie (KIT), 51.03.03; LK 01 (2019). `doi:10.5445/IR/1000100543`.

[82] A. Aab, et al., Pierre Auger, The Pierre Auger Observatory Upgrade - Preliminary Design Report (2016)`arXiv:1604.03637`.

[83] M. Platino, M. R. Hampel, A. Almela, A. Krieger, D. Gorbeña, A. Ferrero, G. D. L. Vega, A. Lucero, F. Suarez, M. Videla, O. Wainberg, A. Etchegoyen, AMIGA at the auger observatory: the scintillator module testing system, Journal of Instrumentation 6 (06) (2011) P06006–P06006. `doi:10.1088/1748-0221/6/06/p06006`.
URL `https://doi.org/10.1088/1748-0221/6/06/p06006`

[84] A. M. Botti, private communication.

[85] F. Gesualdi, A. D. Supanitsky, Estimation of the number of counts on a particle counter detector with full time resolution, Eur. Phys. J. C 82 (2022) 925. `doi:10.1140/epjc/s10052-022-10895-9`.

[86] F. Gesualdi, A. D. Supanitsky, A new pile-up correction strategy for the Underground Muon Detector, Auger internal note GAP-2022-001 (2022).

[87] Ben O'Neill, The classical occupancy distribution: Computation and approximation, The American Statistician 75 (4) (2020) 364–375. `arXiv:https://doi.org/10.1080/00031305.2019.1699445`, `doi:10.1080/00031305.2019.1699445`.
URL `https://doi.org/10.1080/00031305.2019.1699445`

[88] A. Supanitsky, Estimation of the number of muons with muon counters, Astroparticle Physics 127 (2021) 102535. `doi:https://doi.org/10.1016/j.astropartphys.2020.102535`.
URL `https://www.sciencedirect.com/science/article/pii/S0927650520301079`

[89] S. Müller for the Pierre Auger Collaboration, Direct Measurement of the Muon Density in Air Showers with the Pierre Auger Observatory, EPJ Web Conf. 210 (2019) 02013. `doi:10.1051/epjconf/201921002013`.
URL `https://doi.org/10.1051/epjconf/201921002013`

[90] A. Aab, et al., Pierre Auger, Direct measurement of the muonic content of extensive air showers between $\mathbf{2 \times 10^{17}}$ and $\mathbf{2 \times 10^{18}}$ eV at the Pierre Auger Observatory, Eur. Phys. J. C 80 (8) (2020) 751. `doi:10.1140/epjc/s10052-020-8055-y`.

[91] A. Botti, F. Sánchez, M. Roth, A. Etchegoyen, Development and validation of the signal simulation for the underground muon detector of the Pierre Auger Observatory, Journal of Instrumentation 16 (07) (2021) P07059. `doi:10.1088/1748-0221/16/07/p07059`.
URL `https://doi.org/10.1088/1748-0221/16/07/p07059`

[92] A. Aab, et al., The Pierre Auger Collaboration, Features of the energy spectrum of cosmic rays above 2.5×10^{18} eV using the pierre auger observatory, Phys. Rev. Lett. 125 (2020) 121106. `doi:10.1103/PhysRevLett.125.121106`.
URL `https://link.aps.org/doi/10.1103/PhysRevLett.125.121106`

[93] L. Cazón, R. A. Vázquez, E. Zas, Depth development of extensive air showers from muon time distributions, Astropart. Phys. 23 (4) (2005) 393–409.

[94] W. Wu, M. He, X. Zhou, H. Qiao, A new method of energy reconstruction for large spherical liquid scintillator detectors, Journal of Instrumentation 14 (03) (2019) P03009–P03009. `doi:10.1088/1748-0221/14/03/p03009`.
URL `https://doi.org/10.1088/1748-0221/14/03/p03009`

[95] S. Andringa, et al., Current Status and Future Prospects of the SNO+ Experiment, Advances in High Energy Physics 2016 (2016). `arXiv:1508.05759`, `doi:10.1155/2016/6194250`.

[96] A. Abusleme, et al., JUNO, JUNO physics and detector, Progress in Particle and Nuclear Physics 123 (2022) 103927. `doi:https://doi.org/10.1016/j.ppnp.2021.103927`.
URL `https://www.sciencedirect.com/science/article/pii/S0146641021000880`

[97] S. N. Müller, Measurement of the cosmic ray composition with air showers detected by the amiga extension at the pierre auger observatory, Ph.D. thesis, Karlsruher Institut für Technologie (KIT) (2019). `doi:10.5445/IR/1000089730`.

[98] R. Brun, F. Rademakers, ROOT: An object oriented data analysis framework, Nucl. Instrum. Meth. A389 (1997) 81–86. `doi:10.1016/S0168-9002(97)00048-X`.

[99] K. Greisen, Cosmic ray showers, Annual Review of Nuclear Science 10 (1) (1960) 63–108. `arXiv:https://doi.org/10.1146/annurev.ns.10.120160.000431`, `doi:10.1146/annurev.ns.10.120160.000431`.
URL `https://doi.org/10.1146/annurev.ns.10.120160.000431`

[100] J. van Buren, T. Antoni, W. Apel, A. Badea, K. Bekk, A. Bercuci, M. Bertaina, J. Blümer, H. Bozdog, I. Brancus, M. Brüggemann, P. Buchholz, A. Chiavassa, K. Daumiller, F. Di Pierro, P. Doll, R. Engel, J. Engler, F. Feßler, D. Zimmermann, Muon Size Spectrum measured by KASCADE-Grande 6 (2005) 301.

[101] A. Lagutin, R. Raikin, Lateral distribution of electrons in EAS at superhigh energies: predictions and experimental data, Nuclear Physics B - Proceedings Supplements 97 (1) (2001) 274–277. `doi:https://doi.org/10.1016/S0920-5632(01)01282-8`.
URL `https://www.sciencedirect.com/science/article/pii/S0920563201012828`

[102] W. A. et al, The KASCADE-Grande experiment, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 620 (2) (2010) 202–216. `doi:https://doi.org/10.1016/j.nima.2010.03.147`.
URL `https://www.sciencedirect.com/science/article/pii/S0168900210007734`

[103] A. Supanitsky, A. Etchegoyen, G. Medina-Tanco, I. Allekotte, M. Berisso, M. C. Medina, Underground Muon Counters as a Tool for Composition Analyses, Astropart. Phys. 29 (2008) 461–470. `arXiv:0804.1068`, `doi:10.1016/j.astropartphys.2008.05.003`.

[104] S. N. Müller, Measurement of the Cosmic Ray Composition with Air Showers Detected by the AMIGA Extension at the Pierre Auger Observatory, Ph.D. thesis, Karlsruher Institut für Technologie (KIT) (2019). `doi:10.5445/IR/1000089730`.

[105] N. González, F. Sánchez, J. Figueira, A. Almela, A. Botti, A. Etchegoyen, A. Fuster, M. Hampel, A. Lucero, M. Platino, D. Ravignani, B. Wundheiler, On the muon lateral distribution using Underground Muon Detector data (2020).

[106] A. M. Hillas, J. D. Hollows, H. W. Hunter, D. J. Marsden, Calculations on the particle and energy-loss densities in extensive air showers at large axial distances, in: International Cosmic Ray Conference, Vol. 29 of International Cosmic Ray Conference, 1970, p. 533.

[107] A. Haungs, H. Rebel, M. Roth, Energy spectrum and mass composition of high-energy cosmic rays, Reports on Progress in Physics 66 (7) (2003) 1145–1206. `doi:10.1088/0034-4885/66/7/202`.

[108] J. Linsley, Primary cosmic rays of energy 10^{17} to 10^{20} ev, the energy spectrum and arrival directions, in: International Cosmic Ray Conference, Vol. 4 of International Cosmic Ray Conference, 1963, p. 77.

[109] Q. Luce, Study and update of the shape of the nkg within offline framework, talk presented at Auger Collaboration Meeting, Malargue, Argentina (2019).

[110] J. D. Jesús, Private Communication (2020).

[111] A. Coleman, Measurement of the cosmic ray flux above 100 PeV at the Pierre Auger Observatory, Auger internal note GAP-2018-045 (2018).

[112] B. D. J.M. Albury, J.A. Bellido, Parameterisation of the "Lateral Trigger Probability" functions using hybrid data (2017).

[113] A. G. Frodesen, O. Skjeggestad, H. Tofte, Probability and Statistics in Particle Physics, Universitetsforlaget, Bergen, Norway, 1979.

[114] P. Stoica, Y. Selen, Model-order selection: a review of information criterion rules, IEEE Signal Processing Magazine 21 (4) (2004) 36–47. `doi:10.1109/MSP.2004.1311138`.

[115] H. Akaike, Information Theory and an Extension of the Maximum Likelihood Principle, Springer New York, New York, NY, 1998. `doi:10.1007/978-1-4612-1694-0_15`.
URL `https://doi.org/10.1007/978-1-4612-1694-0_15`

[116] J. V. Bradley, Distribution-Free Statistical Tests, Englewood Cliffs N.J: Prentice-Hall, 1968.

[117] F. Mosteller, R. A. Fisher, Questions and answers, The American Statistician 2 (5) (1948) 30–31.
URL `http://www.jstor.org/stable/2681650`

[118] T. Antoni, W. Apel, F. Badea, K. Bekk, K. Bernlöhr, H. Blümer, E. Bollmann, H. Bozdog, I. Brancus, A. Chilingarian, K. Daumiller, P. Doll, J. Engler, F. Feßler, H. Gils, R. Glasstetter, R. Haeusler, W. Hafemann, A. Haungs, D. Heck, T. Holst, J. Hörandel, K.-H. Kampert, J. Kempa, H. Klages, J. Knapp, D. Martello, H. Mathes, H. Mayer, J. Milke, D. Mühlenberg, J. Oehlschläger, M. Petcu, H. Rebel, M. Risse, M. Roth, G. Schatz, F. Schmidt, T. Thouw, H. Ulrich, A. Vardanyan, B. Vulpescu, J. Weber,

J. Wentz, T. Wiegert, J. Wochele, J. Zabierowski, Electron, muon, and hadron lateral distributions measured in air showers by the kascade experiment, Astroparticle Physics 14 (4) (2001) 245–260. `doi:https://doi.org/10.1016/S0927-6505(00)00125-0`.
URL `https://www.sciencedirect.com/science/article/pii/S0927650500001250`

[119] Q. Luce, Private Communication (2020).

[120] M. Schimassek, D. Veberič, R. Engel, Useful Parameterisations of SD Simulations, Auger internal note GAP-2019-075 (2019).

[121] M. Perlin, Core-corona hadronization model and its impact on muon content of extensive air showers, Ph.D. thesis, Karlsruher Institut für Technologie (KIT), 51.13.03; LK 01 (2022). `doi:10.5445/IR/1000143163`.

[122] D. Ivanov for the Pierre Auger Collaboration and the Telescope Array Collaboration, Report of the Telescope Array - Pierre Auger Observatory Working Group on Energy Spectrum, in: 35th International Cosmic Ray Conference (ICRC2017), Vol. 301 of International Cosmic Ray Conference, 2017, p. 498.

[123] F. Gesualdi, A. D. Supanitsky, A. Etchegoyen, Muon deficit in air shower simulations estimated from AGASA muon measurements, Phys. Rev. D 101 (2020) 083025. `doi:10.1103/PhysRevD.101.083025`.
URL `https://link.aps.org/doi/10.1103/PhysRevD.101.083025`

[124] F. Gesualdi, A.D. Supanitsky, and A. Etchegoyen, Muon deficit in simulations of air showers inferred from AGASA data, in: Proc. 37th Int. Cosmic Ray Conf., Berlin, Germany, 2021. `doi:https://doi.org/10.22323/1.395.0326`.

[125] F. Gesualdi, H. Dembinski, K. Shinozaki, D. Supanitsky, T. Pierog, L. Cazon, D.Soldin, and R. Conceição for the Working group on Hadronic Interactions and Shower Physics (WHISP), On the muon scale of air showers and its application to the AGASA data, in: Proc. 37th Int. Cosmic Ray Conf., Berlin, Germany, 2021. `doi:https://doi.org/10.22323/1.395.0473`.

[126] H. P. Dembinski, Computing mean logarithmic mass from muon counts in air shower experiments, Astroparticle Physics 102 (2018) 89–94. `arXiv:1711.05737`, `doi:10.1016/j.astropartphys.2018.05.008`.

[127] P. Abreu, et al., Pierre Auger, The energy spectrum of cosmic rays beyond the turndown around 10^{17} eV as measured with the surface detector of the Pierre Auger Observatory, Eur. Phys. J. C 81 (11) (2021) 966. `arXiv:2109.13400`, `doi:10.1140/epjc/s10052-021-09700-w`.

[128] R. Hofsaess, Private Communication (2021).

[129] R. Hofsaess, Event and Energy Reconstruction in the Transition Region from Galactic to Extra-Galactic Cosmic Rays with Data from the Pierre Auger Observatory, Master Thesis (2020).

[130] V. Verzi for the Pierre Auger Collaboration, The Energy Scale of the Pierre Auger Observatory, in: The Pierre Auger Observatory: Contributions to the 33rd International Cosmic Ray Conference (ICRC 2013), International Cosmic Ray Conference, 2013, pp. 7–10.

[131] R. U. Abbasi, et al., Telescope Array, The Cosmic-Ray Composition between 2 PeV and 2 EeV Observed with the TALE Detector in Monocular Mode, Astrophys. J. 909 (2) (2021) 178. `arXiv:2012.10372`, `doi:10.3847/1538-4357/abdd30`.

[132] Q. Luce, Private Communication (2022).

[133] J. Vicha, Pierre Auger, Adjustments to model predictions of depth of shower maximum and signals at ground level using hybrid events of the Pierre Auger Observatory, PoS ICRC2021 (2021) 395. `doi:10.22323/1.395.0310`.

[134] K. Shinozaki, M. Teshima, AGASA Results, Nuclear Physics B (Proc. Suppl.) 136 (2004) 18–27.

[135] N. Hayashida, T. Kifune, Proportional counter for air shower observation, Nucl. Instrum. Meth. 173 (1990) 431–437.

[136] S. Yoshida, et al., The cosmic ray energy spectrum above 3×10^{18} ev measured by the akeno giant air shower array, Astroparticle Physics 3 (1995) 105–123.

[137] C. Jui for the Telescope Array Collaboration, Summary of Results from the telescope Array Experiment, in: 34th International Cosmic Ray Conference (ICRC2015), Vol. 34 of International Cosmic Ray Conference, 2015, p. 35.

[138] F. Fenu for the Pierre Auger Collaboration, Pierre Auger, The cosmic ray energy spectrum measured using the Pierre Auger Observatory, in: The Pierre Auger Observatory: Contributions to the 35th International Cosmic Ray Conference (ICRC 2017), 2017, pp. 9–16, [PoSICRC2017, 486(2018)]. `doi:10.22323/1.301.0486`.

[139] M. Takeda, et al., Energy determination in the akeno giant air shower array experiment, Astroparticle Physics 19 (2003) 447–462.

[140] D. Ivanov, Energy spectrum measured by the telescope array surface detector, Ph.D. thesis, Rutgers U., Piscataway (2012).
URL `https://rucore.libraries.rutgers.edu/rutgers-lib/38839/`

[141] S. Yoshida, et al., Lateral distribution of charged particles in giant air showers above 1 EeV observed by AGASA, J. Phys. G: Nucl. Part. Phys. 20 (1994) 651.

[142] M. Takeda, et al., Energy determination in the Akeno Giant Air Shower Array Experiment, Astropart. Phys. 19 (2003) 446. `doi:10.1016/S0927-6505(02)00243-8`.

[143] K. Shinozaki, et al., Properties of EHE gamma-ray initiated air showers and their search by AGASA, Vol. ICRC2001, 2001, p. 346.

[144] K. Shinozaki, et al., Upper Limit on Gamma-Ray Flux above 10^{19} eV Estimated by the Akeno Giant Air Shower Array Experiment,, The Astrophysical Journal Letters 571 (2002) L117–L120. `doi:10.1086/341288`.

[145] National Centers for Environmental Information, National Oceanic and Atmospheric Administration, accessed September 2, 2019. [link].
URL `http://www.ngdc.noaa.gov/geomag/`

[146] D. Soldin, EAS-MSU, IceCube, KASCADE-Grande, NEVOD-DECOR, Pierre Auger, SUGAR, Telescope Array, Yakutsk EAS Array, Update on the Combined Analysis of Muon Measurements from Nine Air Shower Experiments, PoS ICRC2021 (2021) 349. `arXiv:2108.08341`, `doi:10.22323/1.395.0349`.

[147] S. Argiro, S. Barroso, J. Gonzalez, L. Nellen, T. Paul, T. Porter, L. Prado Jr., M. Roth, R. Ulrich, D. Veberič, The offline software framework of the Pierre Auger Observatory, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 580 (3) (2007) 1485–1496. `doi:https://doi.org/10.1016/j.nima.2007.07.010`.
URL `https://www.sciencedirect.com/science/article/pii/S0168900207014106`

[148] M. Gottowik, Extension of the ADST event selection tool for radio, Auger internal note GAP-2020-033 (2020).

[149] F. Gesualdi, M. Roth, D. Schmidt, and D. Veberič, Selecting Muon Detector data with the ADST event selection tool in Offline, Auger internal note GAP-2021-013 (2021).

[150] I. Mariş, S. Müller, F. Schüssler, R. Ulrich, M. Unger, ADST-Analysis Manual: Utilities for ADST-Analysis, Auger internal note GAP-2009-012 (2009).

[151] R. Ulrich, Basic ADST excercises, Offline Workshop, Karlsruhe Institute of Technology (2016).
URL `https://indico.nucleares.unam.mx/event/1126/session/4/material/0/0.pdf`