TÉCNICO LISBOA

Probing multiparticle production properties of the first ultra-high energy cosmic ray-Air interaction in Extensive Air Showers with low muon content

Miguel Alexandre Jesus da Silva Martins

Thesis to obtain the Master of Science Degree in

Engineering Physics

Supervisor: Prof. Lorenzo Cazon Boado
Prof. Ruben Maurício da Silva Conceição

Examination Committee

Chairperson:	Prof. Mário João Martins Pimenta
Supervisor:	Prof. Ruben Maurício da Silva Conceição
Member of the Committee:	Prof. Pedro Jorge Dos Santos de Assis

November 2021

Acknowledgments

First and foremost, I would like to thank Dr. Lorenzo Cazon and Dr. Ruben Conceição for their guidance and all the insightful discussions over the last 3 years. I cannot thank them enough for all the opportunities they created and their trust in my work. I must also thank them for the careful reading of this text and all their valuable suggestions.

I thank the Auger group of the Laboratório de Instrumentação e Física Experimental de Partículas for all the fruitful discussions and partial implementation of the Offline modules needed for this work. In particular, I would also like to thank Dr. Felix Riehn, Dr. Ruben Conceição and Dr. Lorenzo Cazon, whose work on probing the energy spectrum of hadrons at ultra-high energy through fluctuations of the muon content of EAS, set the foundations of the phenomenological studies presented in the first part of this thesis.

I thank the financial support by FCT under the project number CERN/FIS-PAR/0031/2019.

I will always be thankful to my colleagues, whose help, support, insight and friendship got me through many sleepless nights before important exams and demanding deadlines, throughout the 5 years of the Master's in engineering physics.

Lastly, on a personal level I am deeply thankful to my parents, brother, great grand-aunt, close friends and Diogo for their patience and encouragement, specially in stressful times.

Resumo

Os raios cósmicos de alta energia constituem uma oportunidade única para estudar interações hadrónicas a energias de centro de massa superiores a 100 TeV. Neste trabalho, mostra-se que o espectro de energia hadrónica da primeira interação p-Ar é restringido por medidas do declive da distribuição do número de muões à superfície da Terra, em cascatas atmosféricas com baixo conteúdo muónico. Prova-se ainda que este declive permite restringir a secção eficaz de produção de piões neutros em interaçãos p-Ar a altas energias, com uma dependência residual no modelo hadrónico usado. Adicionalmente, mostra-se que medidas precisas do declive da distribuição do número de muões para cenários de composição mista, dentro das incertezas experimentais atuais, estão apenas limitadas pelo número de eventos, desde que a referida distribuição seja corrigida para contabilizar a dependência do fluxo de raios cósmicos na energia do primário.

É ainda proposta a reconstrução da distribuição do número de muões a partir da sua correlação com o número esperado de segmentos ativos na área fiducial das RPCs de estações de MARTA a 450 m do eixo do chuveiro. Em condições ideais, mostrou-se que a reconstrução foi bem sucedida para primários a $10^{18.5}$ eV e ângulo zenital 40°, não obstante da contaminação electromagnética. No entanto, um efeito de limiar enviesa os estimadores do declive e dos dois primeiros momentos da ditribuição do número de muões. Os resultados conseguidos com RPCs em todo o array SD-750 do Observatório Pierre Auger são reprodutiveis com 7 estações de MARTA numa célula hexagonal unitária.

Palavras-chave: Raios cósmicos de alta energia, Cascatas amosféricas de partículas, Flutuações da componente muónica, Espectro de energia dos hadrões, Secção eficaz e produção de piões neutros, MARTA

Abstract

Ultra-high energy cosmic rays offer a unique opportunity to study hadronic interactions at center-of-mass energies above $100\,\mathrm{TeV}$. With this work, we show that the energy spectrum of hadronically interacting particles of the first p-Air interaction is constrained by measurements of the slope of the shower-to-shower distribution of the number of muons at the ground level in Extensive Air Showers with low muonic content. Similarly, we show that the mentioned slope can be used to constrain the production cross-section of neutral pions in ultra-high energy p-Air interactions, with a residual dependence on the hadronic interaction models. Furthermore, we prove that precise measurements of said slope in mixed composition scenarios are only limited by event statistics, within current experimental uncertainties and provided that the muon number distribution is corrected for the cosmic ray spectrum dependence on the primary energy.

We propose the reconstruction of the muon number distribution through its correlation with the expected number of active pads in the fiducial area of RPCs of MARTA stations at $450\,\mathrm{m}$ from the shower axis. Under ideal conditions, the reconstruction of that distribution is successful at $10^{18.5}\,\mathrm{eV}$ and zenith angle 40°, despite the electromagnetic contamination. However, a threshold effect prevented a completely unbiased measurement of the slope and of the first and second moments of the muon number distribution. Results achieved with RPCs over the SD-750 array of the Pierre Auger Observatory can be reproduced with a hexagonal unitary cell of 7 MARTA stations.

Keywords: Ultra-high energy cosmic rays, Extensive Air Showers, Muon Fluctuations, Energy Spectrum of hadrons, Production cross-section of neutral pions, MARTA

Contents

List of Tables

List of Figures

List of Abbreviations

AMIGA Auger Muons and Infill for the Ground Array.

CDAS Central Data Acquisition System.

EAS Extensive Air Shower.

FADC Flash Analog to Digital Converter.

FD Fluorescence Detector.

HIM Hadronic Interaction Model.

LDF Lateral Distribution Function.

LHC Large Hadron Collider.

LHCf Large Hadron Collider forward.

MARTA Muon Array with RPCs for Tagging Air Showers.

MLDF Muon Lateral Distribution Function.

MoPS Multiplicity of Positive Steps.

NKG Nishimura-Kamata-Greisen.

pdf Probability Density Function.

PMT Photomultiplier Tube.

QCD Quantum Chromodynamics.

RPC Resistive Plate Chamber.

SD Surface Detector.

SD-750 Infill array of the Pierre Auger Observatory.

SSD Scintillatior Surface Detector.

TH Trigger Threshold.

ToT Time over Threshold.

ToTs Time over Threshold deconvoluted.

UHECR Ultra-High Energy Cosmic Ray.

VEM Vertical Equivalent Muon.

WCD Water Čerenkov Detector.

Chapter 1

Introduction

1.1 Ultra-High Energy Cosmic Rays

Ultra-High Energy Cosmic Rays (UHECR) are relativistic charged particles, protons or light nuclei, accelerated by astrophysical sources at energies greater than 10^{18} eV. After being produced, cosmic rays propagate through the Universe, interacting with galactic and extragalactic magnetic fields which bend their trajectories according to their rigidity R (energy over charge). As such, charged UHECRs do not point directly to their sources.

As UHECRs reach the Earth's atmosphere, they interact with nuclei of Air molecules producing hundreds to thousands of particles, which further interact triggering macroscopic cascades of particles known as Extensive Air Showers (EAS). It is through the detection and reconstruction of EAS that the primary energy, E_0, mass composition and arrival direction of UHECRs, is determined. Once known, these quantities are used to constrain the possible astrophysical scenarios responsible for the production and acceleration of these high energy particles. For example, Rayleigh analysis in right-ascension and in the azimuthal angle of the arrival directions of UHECRs has shown that their distribution is anisotropic and compatible with a dipole located outside the galactic plane, which supports an extra-galactic origin for UHECRs at $E_0 > 8 \times 10^{18}$ eV [1].

The center-of-mass energies attained by UHECRs, $\sqrt{s} \sim 100$ TeV, surpass by an order of magnitude those reached by the largest human-made collider, the LHC. Thus, the interest in their study is, at least, two fold. On the one hand, these cosmic rays yield information about astrophysical sources in the nearby Universe, since, for example, their energy and composition greatly constrains the minimum size of the UHECR accelerator, via the Hillas condition: the minimum size of an accelerator is the Larmor radius of the charged cosmic ray in the presence of the accelerator's magnetic field [2]. On the other hand, UHECRs provide a unique opportunity to improve our description of the hadronic interactions at ultra-high energies.

However, there are a few caveats to the study of UHECRs. First, the flux of ultra-high energy events is extremely low, preventing their direct measurement with spacecrafts or balloons. Furthermore, even for large exposure times and ground experiments spanning over large areas, the collected data sets are

only now becoming statistically significant. Secondly, as a consequence of the first caveat, on which we elaborate on in the next section, UHECRs are measured through the reconstruction of Extensive Air Showers. Thus, the interpretation of their mass composition inherits the high systematic uncertainties stemming from the poor description of high energy interactions, provided by Hadronic Interaction Models (HIM).

1.1.1 Energy spectrum of Cosmic Rays

Let $\mathcal{N}$ be the number of cosmic ray events and E_0 be the primary energy. Then, the cosmic ray energy spectrum is given by a power law in energy

$$\Phi(E_0) = \frac{\mathrm{d}\mathcal{N}}{\mathrm{d}E_0} \propto E_0^{-\gamma}, \tag{1.1}$$

where γ is the spectral index, which, in fact, varies with energy [3]. Correspondingly, the event rate per unit area and solid angle, reads, for an energy bin $\mathrm{d}E_0$

$$J(E_0) = \frac{\mathrm{d}\Phi}{\mathrm{d}A\,\mathrm{d}t\,\mathrm{d}\Omega} = J(E_{\mathrm{ref}}) \left(\frac{E_0}{E_{\mathrm{ref}}}\right)^{-\gamma}, \tag{1.2}$$

where E_{ref} is some reference energy for which the flux is known.

Throughout the energy range of UHECRs, the spectral index γ varies from ~ 3.2, between the *knee* at $E_0 = 5 \times 10^{15}$ and the *ankle* at $E_0 = 5 \times 10^{18}$ eV, to ~ 2.5 before $E_0 = 1.3 \times 10^{19}$ eV. It then stays at ~ 3 until $E_0 = 5 \times 10^{19}$ eV, before increasing to ~ 5, highly suppressing the number of cosmic ray events [3]. This behaviour is shown in Figure 1.1, where the spectrum is weighted by E_0^3 and the data comes from the Pierre Auger Observatory.

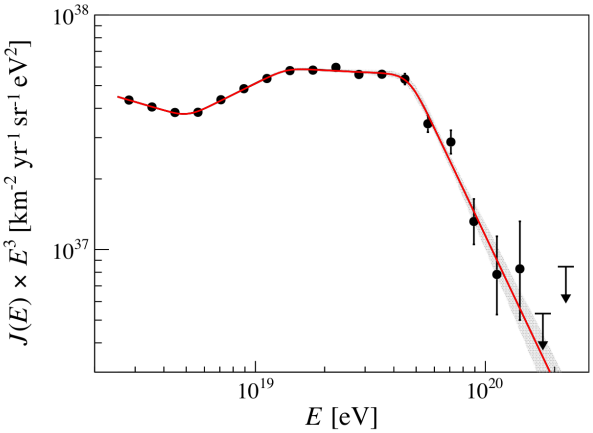

Figure 1.1: Fitted UHECR spectrum weigthed by E_0^3 above $E_0 = 10^{18}$ eV [3].

The large suppression of the cosmic ray energy spectrum above $E_0 = 5 \times 10^{19}$ eV can be due to source exhaustion or/and to propagation effects, such as the Greisen-Zatsepin-Kuzmin (GZK) cut-off and the giant dipole resonance [4]. The GZK effect was proposed independently by Greisen [5] and Zatsepin and Kuzmin [6] and corresponds to the attenuation of protonic cosmic rays with enough

energy to produce neutral pions via their interaction with the Cosmic Microwave Background (CMB): $p + \gamma_{CMB} \rightarrow p + \pi^0$. Thus, only cosmic rays with $E_0 = 5 \times 10^{19}$ eV closer than ~ 100 Mpc reach the Earth [2].

A rough estimate to bare in mind is that at $E = 5 \times 10^{19}$ eV the flux of UHECRs is about 1 km^{-2} century^{-1}.

1.1.2 Open questions

Currently, there are many open questions regarding ultra-high energy cosmic ray research, of which we highlight: the origin of the bulk of UHECRs, their mass composition, if the origin of the end of the cosmic-ray spectrum is a result of source exhaustion or propagation effects, the transition from galactic to extragalactic cosmic rays, the effect of galactic and extragalactic magnetic fields on the trajectories of charged UHECRs, anisotropy expectations for specific astrophysical scenarios and how hadronic interactions can be accurately described at such ultra-high energies [2].

1.2 Extensive Air Showers

Extensive Air Showers are cascades of interacting particles triggered by cosmic rays. In the next few sections, we briefly describe their physics, relevant observables, EAS simulation packages and corresponding hadronic interaction models, as well as some surprising recent results.

1.2.1 Air Shower Physics

After the first UHECR-Air interaction about 75% of the primary energy is carried by hadronically interacting particles, mainly charged pions, kaons and light baryons, while the remaining energy lies in the electromagnetic sector, as shown in the left panel of Figure 1.2. From this point on, the EAS behaves almost as two independent cascades: the electromagnetic cascade, and the hadronic cascade which eventually decays into muons. The right panel of Figure 1.2 displays a very simplified scheme of the first few particle generations [a] of an EAS with the shower components properly identified.

The allowed processes underwent by cascade particles are described by Quantum Chromodynamics (QCD) and by the Glashow-Weinberg-Salam theory of electroweak interactions. Whether a particle interacts with Air molecules or decays, is dependent on its energy: the energy below which a particle is more likely to decay than to interact is known as critical energy, ξ_c. Thus, since each interaction requires energy conservation and its multiplicity [b] is larger than the number of initial state particles, the average energy of shower particles of a given generation decreases with respect to the preceding generation and it becomes more likely that unstable particles decay rather than interact.

The electromagnetic component is a standard cascade of interacting electrons, positrons and photons, that reaches its maximum size at a given atmospheric slant depth, X_{max}, before degrading enough

[a] The first generation, $g = 1$, corresponds to particles emerging from the Cosmic-Ray-Air interaction; the second generation, $g = 2$ to particles emerging from the interaction of 1^{st} generation particles, and so on.

[b] Number of particles emerging from a given interaction or decay

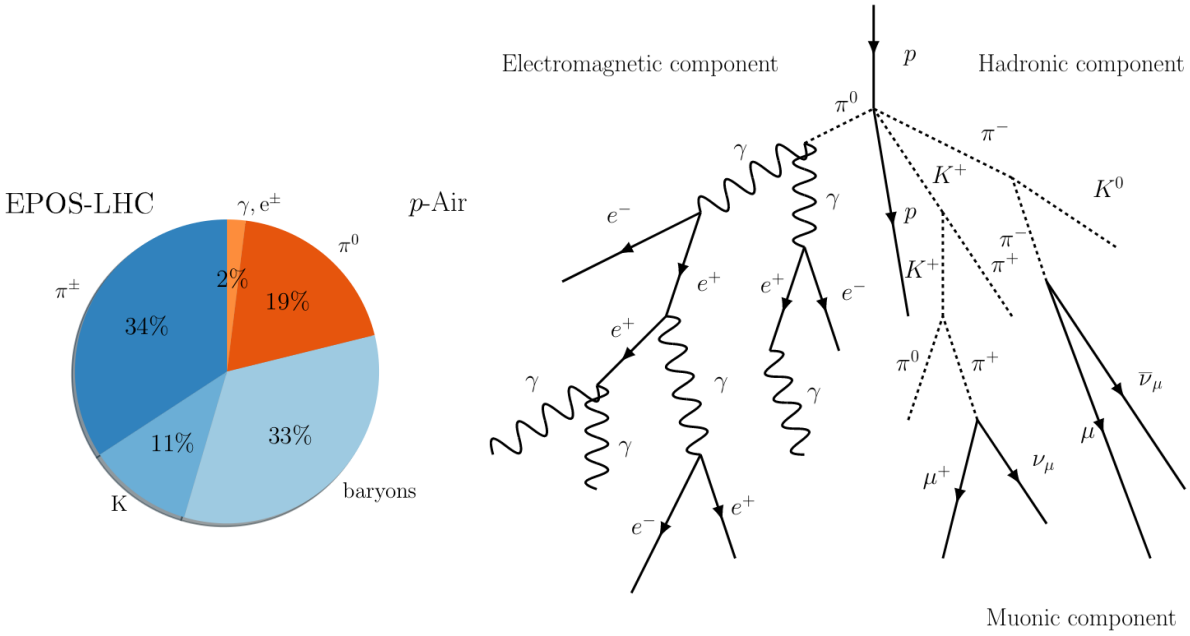

Figure 1.2: Left panel: Energy partition among particles emerging from a proton-Air interaction at $E_0 = 10^{19}$ eV as simulated by EPOS-LHC. Hadronically interacting particles are shown in shades of blue and those of the electromagnetic sector in shades of orange [7]. Right panel: Simplified scheme of an EAS displaying the hadronic, electromagnetic and muonic components. Note that some vertices refer to decays, while others refer to interactions with Air molecules. Furthermore, not all final state particles are represented in each interaction.

energy to interact with the electrons of Air molecules and energy losses through ionisation start to dominate. It accounts for 90% of the primary energy at the ground level and its evolution is rather well studied in the literature: its longitudinal profile is described by the Gaisser-Hillas function, and many of its features, such as the electron's energy spectra, lateral and angular distributions, can be computed analytically under some approximations [7]. The electromagnetic component accounts for about 25% of the energy released in each interaction and it is fed by the decay of neutral pions, π^0's. These mesons have a proper life time of $\tau = 8.43 \times 10^{-17}$ s [8], which even at such high energies prevents them from interacting hadronically. Instead, they decay into photons, $\pi^0 \to \gamma\gamma$, with a branching ratio of $\Gamma \sim 99\%$ [8]. These mesons are produced in every stage of the EAS, coupling the electromagnetic and hadronic cascades. However, the contribution of neutral pions emerging from the Cosmic Ray-Air interaction highly dominates that of the subsequent shower generations. Besides the decay of neutral pions, there are other subdominant processes that still couple the electromagnetic and hadronic/muonic showers, such as the muon decay $\mu^- \to e^- \overline{\nu}_e \nu_\mu$ and its conjugate, with a branching ratio of $\Gamma \simeq 100\%$ [8], and the electromagnetic component stemming from low energy hadrons. Another important contribution to muon production is photo-pion production, which is particularly relevant for showers with low muon content [9].

In turn, the hadronic cascade comprises the hadronically interacting particles and must be described by phenomenological Hadronic Interaction Models. The critical energies of the most common mesons in EAS are $\xi_c^{\pi^\pm} = \mathcal{O}(50\ \text{GeV})$, $\xi_c^{K^\pm} = \mathcal{O}(1\ \text{TeV})$, $\xi_c^{K_L^0} = \mathcal{O}(200\ \text{GeV})$ and $\xi_c^{K_S^0} = \mathcal{O}(30\ \text{TeV})$ [7], below which pions and kaons decay. The allowed decay channels with the largest branching ratios for these

mesons are: the decay $\pi^+ \to \mu^+\nu_\mu$ and its conjugate, have a branching ratio of $\Gamma \simeq 100\%$ [8]; the decay of $K^+ \to \mu^+\nu_\mu$ and its conjugate with $\Gamma \simeq 64\%$ [8] and $K^+ \to \pi^+\pi^0$ with $\Gamma \simeq 21\%$ [8]. Thus, the hadronic sector ultimately decays predominantly into muons.

Muons are charged leptons with $m_\mu = 105.7$ MeV$/c^2$ and a proper lifetime of $\tau = 2.20 \times 10^{-6}$ s [8]. Their relatively large mass means that the energy carried by shower muons is degraded mainly through ionisation, which combined with their appreciable lifetime scaled by a Lorentz factor $E_\mu/m_\mu = \gamma \gg 1$ at GeV energy scales, results in a high penetration length in the atmosphere. Hence, the muonic component, fed by hadronically interacting particles, can be detected at the ground level, where it accounts for $\sim 10\%$ of the primary energy. Thus, muons are messengers of the hadronic activity in all shower generations.

A simple, yet effective model relating the muon content of EAS, N_μ, with the primary energy and composition, is described below.

1.2.2 Heitler-Matthews model and muons

After Heitler had established a framework capable of describing the most prominent features of the electromagnetic cascade [10], Matthews developed a similar model for the hadronic cascade, where the only hadronically interacting particles are charged pions [11]. Within this framework, known as Heitler-Matthews model, the number of muons in Extensive Air Showers can be expressed as a power law function of the primary energy and mass composition. We follow closely the deductions of [11].

Consider a proton initiated EAS where only charged and neutral pions, $\pi^\pm$ and π^0 emerge from each interaction in proportion $2/3$ to $1/3$, respectively, due to isospin symmetry. While neutral pions immediately decay $\pi^0 \to \gamma\gamma$, feeding the electromagnetic cascade, charged pions further interact with Air molecules after travelling a fixed distance, $d = \lambda_{\text{Int}} \ln 2$, with λ_{Int} the strong interaction length. We assume that the same holds for all subsequent interactions, with fixed total and pionic multiplicities, m_{total} and m_{ch}, respectively, until charged pions reach a critical energy $\xi_c^{\pi^\pm}$. Afterwards, all charged pions decay into muons, a plausible assumption given the branching ratio presented earlier for this decay. Lastly, we start by assuming that all interactions are perfectly inelastic, before introducing a parameter κ to account for leading particle effects.

Let E_0 be the energy of the primary. After g shower generations the number of charged pions reads $N_{\text{ch}} = m_{\text{ch}}^g$, and the total multiplicity $N_{\text{total}} = m_{\text{total}}^g = \left(\frac{3}{2} m_{\text{ch}}\right)^g$. Assuming an equipartition of energy among emerging pions, each pion of generation g carries an energy

$$E_\pi = \frac{E_0}{N_{\text{total}}} = \frac{E_0}{\left(\frac{3}{2} m_{\text{ch}}\right)^g}. \tag{1.3}$$

The critical generation, g_c, is readily determined by making $E_\pi = \xi_c^{\pi^\pm}$ in the previous equality and solving with respect to g:

$$g_c = \frac{\ln\left(\frac{E_0}{\xi_c^{\pi^\pm}}\right)}{\ln\left(\frac{3}{2} m_{\text{ch}}\right)}. \tag{1.4}$$

Thus, the number of muons of the EAS can be cast as a power law in the primary energy by noting that $N_\mu = N_{\text{ch}}$ at the critical generation:

$$N_\mu = m_{\text{ch}}^{g_c} \iff \ln N_\mu = g_c \ln\left(m_{\text{ch}}\right) = \beta \ln\left(\frac{E_0}{\xi_c^{\pi^\pm}}\right), \tag{1.5}$$

where we defined

$$\beta \equiv \frac{\ln m_{\text{ch}}}{\ln\left(\frac{3}{2} m_{\text{ch}}\right)}, \tag{1.6}$$

to give

$$N_\mu = \left(\frac{E_0}{\xi_c^{\pi^\pm}}\right)^\beta. \tag{1.7}$$

Note that the above result also assumes that muons do not decay themselves. It is straightforward to generalise the above formula to a primary with mass A if we neglect the binding energy of the nucleons, and apply the superposition principle to conclude that a nucleus with mass number A behaves as A nucleons with energy E_0/A each. Therefore,

$$N_\mu = A\left(\frac{E_0}{A\xi_c^{\pi^\pm}}\right)^\beta = A^{1-\beta}\left(\frac{E_0}{\xi_c^{\pi^\pm}}\right)^\beta. \tag{1.8}$$

A more realistic approximation considers that most of the energy in a hadron-nucleus interaction is carried by a leading particle. To include this effect in the above model, we consider the inelastic coefficient κ, such that the leading particle takes a fraction $(1-\kappa)$ of the energy in each collision, while charged pions take $\frac{2}{3}\kappa$. As a result, we must alter Equation (1.6) by taking $\frac{3}{2} \to \left(1 - \frac{\kappa}{3}\right)^{-1}$ and $m_{\text{ch}} \to 1 + m_{\text{ch}}$.

Although fairly simple, the Heitler-Matthews model is able to reproduce the average number of muons in EAS simulations if we take $\beta \simeq 0.93$, even though the exact β value depends on the primary composition. Additionally, β increases if the proportion of hadronically interacting particles increases with respect to the total multiplicity. Ultimately the Heitler-Matthews model shows how the average value of N_μ varies with the primary energy and composition, albeit a full description of the hadronic interactions and shower development in the atmosphere requires the usage of Hadronic Interaction Models.

1.2.3 Air Shower Simulators and Hadronic Interaction Models

EAS simulators are used to perform simulations and to analyse experimental data, enabling shower reconstruction and the determination of the primary composition. These Monte Carlo simulators account for all the electromagnetic, strong and weak interactions either among shower particles or using Air atoms as a target. They also account for decays as well as for the propagation of secondary shower particles in the atmosphere, under the influence of the geomagnetic field, giving a complete picture of the shower evolution. Hence, they are not only able to yield average values of shower observables, but also their shower-to-shower fluctuations around the average value.

Air shower simulators are flexible, allowing for the selection of the primary (protons, photons or heavier nuclei until iron), its energy and arrival direction. They also allow the tuning of other shower parame-

ters and the inclusion of different models to simulate high and low energy hadronic interactions. Standard EAS simulator packages, used in this work, include CORSIKA (**Co**smic **R**ay **Si**mulations for **KA**SKADE) [12] and CONEX [13]. The later, combines Monte Carlo simulations with low energy cascade equations, to reduce the computation time and gather a large number of events, enhancing computational efficiency. In both air shower simulation packages, low energy hadronic interactions can be handled by FLUKA [14, 15] or by the microscopic UrQMD [16] routine, while electromagnetic interactions and particle transport can be described by the EGS4 package [17] or analytically by Nishimura-Kamata-Greisen (NKG) equations.

Currently, hadronic interactions are modelled by phenomenological models, which extrapolate from accelerator data and take into account different mechanisms and theoretical considerations in the description of QCD interactions. Examples of up-to-date hadronic interaction models are EPOS-LHC, QGSJET II-04 and SIBYLL 2.3d:

- **EPOS-LHC**: based on the Parton-Based Gribov Regge Theory [18], it is used in air shower simulations and heavy ion collisions. It allows for collective hadronization in $p-p$ scattering, and new versions include different parametrizations of flow in small volumes with high density of thermalized matter in $p-p$ collisions (core) when compared to a large volume produced in heavy ion collisions. It is able to reproduce LHC data in minimum bias hadronic interactions [19].

- **QGSJET II-04**: developed in the framework of Reggeon Field Theory (RFT) it treats non-linear interaction effects as Pomeron-Pomeron [c] interactions by re-summing enhanced RFT diagrams, as opposed to the energy-scale phenomenological parametrizations used by other HIMs [20].

- **SIBYLL 2.3d**: based on the Dual Parton Model (DPM) and the minijet model, it describes the interaction in terms of color fields and strings, using the Lund string fragmentation model to describe hadronization [21].

These phenomenological models are tuned with data from particle physics experiments, which cover a limited range in energy and kinematic phase space, rendering the models predictions increasingly uncertain at the ultra-high energies attained by cosmic rays. In fact, uncertainties in the determination of the moments of the distributions of X_{max} and muon related quantities are not dominated by detector uncertainties, but rather by the differences between HIMs [22]. Consequently, the inaccuracy of each model and inconsistencies between models hamper the interpretation of shower data, and are the source of recent unexpected results.

1.2.4 Simulation and data inconsistencies

It was shown in 2014 that the primary is heavier than proton at the highest energies through X_{max} measurements [23], which either points towards unconsidered astrophysical sources, acceleration mechanisms or physics in UHECR propagation, or, towards an incorrect modelling of the cascade physics. In

[c] Pomerons are phenomenological objects representing the elementary interaction when several parallel QCD interactions occur. They are also known as parton ladders [18].

fact, an accurate description of the highest energy hadronic interactions in EAS relies on the knowledge of the primary, which, in turn, is inferred from shower observables interpreted with inaccurate HIM. This creates a degeneracy that must be broken by considering new shower observables, independent of X_{max}. The fluctuations of the muon content in EAS constitute such an observable [24].

Another paradigmatic example of incorrect predictions of HIMs is the so-called muon puzzle. Measurements performed at the Pierre Auger Observatory show that shower simulations underestimate the muon content of EAS by $26 - 43$ % starting at primary energies of $E_0 \sim 10^{16}$ eV, despite correctly predicting its shower-to-shower fluctuations [25].

Now, for a fixed number of shower generations, highly constrained by X_{max} measurements, this means that there is more energy contained in the hadronic sector of EAS than what is predicted by simulations, resulting in a larger total multiplicity of muons. Many theoretical mechanisms have been proposed to explain the puzzle. For example, [26] proposes the formation of Quark-Gluon plasma (QGP) in p-Air ultra-high energy interactions, which would enhance the production of strange mesons and dump the production of neutral pions, increasing the energy contained in the hadronic sector of the cascade. Alternatively, Lorentz Invariance Violation (LIV) scenarios in the pionic and photonic sectors have been considered to justify a possibly larger lifetime of neutral pions [27], allowing them to interact hadronically, and ultimately contribute to the production of muons. Note, however, that the space of all possible theories is constrained by the requirement that the shower-to-shower fluctuations of the muon content still have to be compatible with the data, as they already are.

In any case, the deep connection between the hadronic activity in EAS, the muon content and its shower-to-shower fluctuations need to be further explored, both phenomenologically and experimentally, to improve our description of hadronic interactions at ultra-high energies.

1.3 The Pierre Auger Observatory

The Pierre Auger Observatory [28] is the world's largest UHECR observatory. Completed in 2008, it has been taking EAS data since 2004 with the goal of determining the energy spectrum, arrival direction and composition of cosmic rays attaining energies greater than $E_0 = 10^{17}$ eV. Run by the Pierre Auger Collaboration, the observatory is located near the city of Malargüe, in Mendoza, Argentina and comprises an hexagonal array of 1661 surface detector (SD) stations spaced by 1.5 km, which spread over an area of 3000 km^2, along with 27 fluorescence telescopes (FD) located in 4 stations surrounding the array. This hybrid detection technique allows for measurement redundancy and cross-calibrations, mitigating the systematic uncertainties of reconstructed shower observables and enabling the reconstruction of high quality events. The observatory sits at an average altitude of $1400\,\mathrm{m}$ above sea level, which corresponds to an atmospheric overburden of ~ 875 g cm^{-2}. Figure 1.3 shows a map of the Pierre Auger Observatory, along with pictures of fluorescence and surface detector stations.

Since its completion, the observatory underwent several upgrades, many of which with the goal of better disentangling the muonic and electromagnetic shower components. The AugerPrime upgrade includes the installation of Surface Scintillator Detectors (SSD) [29]; the addition of Radio detectors,

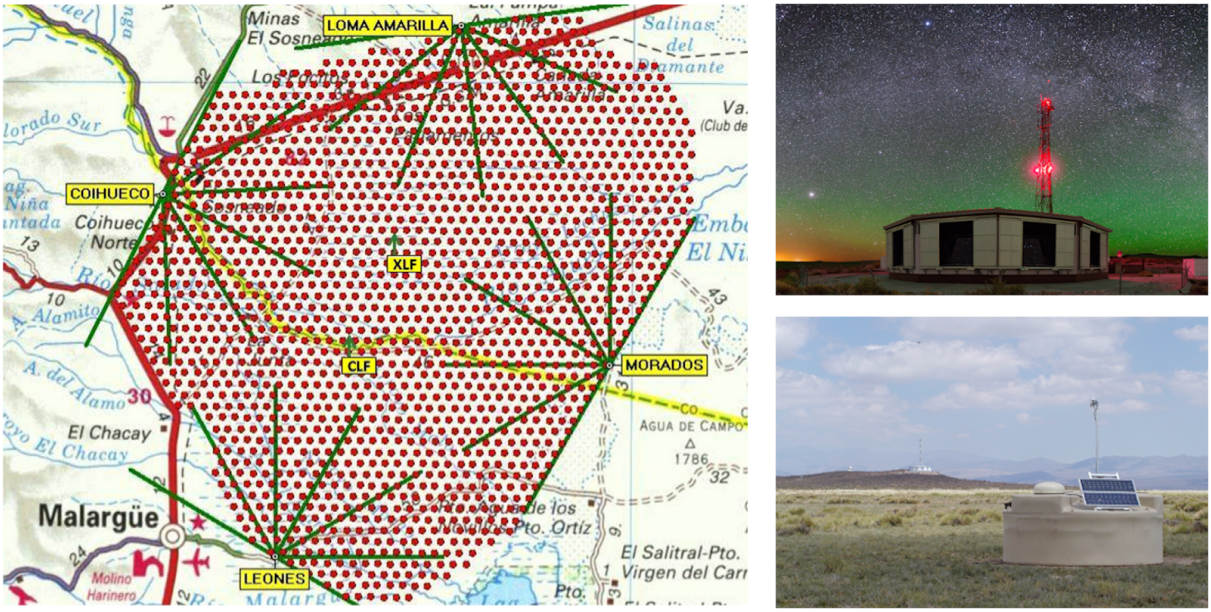

Figure 1.3: Left panel: Map of the Pierre Auger Observatory, with each red dot representing a surface detector station. The four fluorescence detector station sites are indicated by the names in the yellow boxes: Coihueco, Los Leones, Los Morados and Loma Amarilla [28]. Top right panel: a fluorescence detector station with 6 fluorescence telescopes. Bottom right panel: a surface detector station.

consisting of circular antennas placed on top of the water Čerenkov detectors (WCDs), to measure the radio emission of the electromagnetic component in inclined showers [30]; the AMIGA upgrade [31] and the HEAT installation [29]. Additionally, other proof-of-concept detector arrays, like the MARTA engineering array have been proposed and are now under installation.

1.3.1 Fluorescence detector

Each fluorescence detector station has 6 detector telescopes that measure the amount of fluorescence light emitted by excited nitrogen molecules at the passage of charged shower particles. The light is emitted isotropically and in the UV region of the spectrum, coming mostly from the electromagnetic component of the shower. For this reason, the FD allows the reconstruction of the longitudinal profile of the shower, from which X_{max} can be extracted. This is an important shower variable that correlates with the primary composition: its mean value is deeper and its distribution broader for lighter primaries. Furthermore, the intensity of the fluorescent emission is proportional to the energy dissipated by the shower, thus yielding a calorimetric measurement of the primary energy.

Due to the extreme sensitivity of the fluorescence telescopes, FDs can only operate in dark moonless nights with good weather conditions, resulting in a duty cycle of just $\sim 15\%$.

1.3.2 Surface detector

Each surface detector station consists of a cylindrical water tank with radius $r_{\mathrm{tank}} = 1.8$ m filled with 12 tons of ultra-pure water, the detection medium, up to a height $h_{\mathrm{water}} = 1.2$ m, and three 9 in. photo-

multiplier tubes (PMTs) viewing the water volume from above. The PMTs capture the Čerenkov light produced by relativistic charged particles travelling faster than light in the detection medium. Moreover, they measure the signal produced by high energy photons provided they convert to electron-positron pairs in the water volume. The reflexive inner surface of the tanks maximizes the capture of Čerenkov light. Each station is self-powered by solar panels, has a duty cycle of 100 % and communicates wirelessly with the central data acquisition system to register events. Additionally, the tank's cylindrical shape renders them sensitive to highly inclined showers, with zenith angles $\theta > 60^{\circ}$.

The water Čerenkov detectors (WCD) are calibrated *in situ* with atmospheric muons. They register the time structure of the charge deposited by shower particles, time trace, whose integral yields the measured signal. The total signal is the combination of the electromagnetic and muonic signals. Thus, the individual signals produced by each shower component cannot be easily disentangled from one another.

Due to its importance in describing hadronic interactions and determining the primary composition, accurate and precise measurements of the muon content of EAS are paramount. Thus, efforts have been made to isolate this component, either by using the existing WCDs and extracting the muon signal from the shape of its time trace using Recurrent Neural Networks station by station [32], or by enhancing the SD capabilities with other detectors such as AMIGA (Auger Muon and Infilled Ground Array) and MARTA (Muon Array with RPCs for Tagging Air Showers). The former relies on the placement of an underground scintillator with a mass overburden of $\sim 520\,\mathrm{g\,cm^{-2}}$, called the a muon detector (MD), to attenuate electromagnetic particles and better isolate the muon signal [28]. The latter, to be described shortly, consists of a rectangular lattice of sensitive elements of Resistive Plate Chambers (RPCs) placed underneath the WCDs, yielding a mass overburden of $\sim 170\,\mathrm{g\,cm^{-2}}$ [33].

Besides the regular array, where stations are spaced by $1.5\,\mathrm{km}$, there is a denser array area 6 km away from the fluorescence detector of Coihueco, where stations are spaced by $750\,\mathrm{m}$, known as the Infill array or SD-750. It covers an area of $23.5\,\mathrm{km^2}$ and its stations can measure showers down to $E_0 = 3 \times 10^{17}$ eV with 100 % efficiency, while those of the regular array only reach a 100% efficiency for showers with $E_0 > 10^{18.5}$ eV. A more detailed discussion of surface station and event triggers can be found in Chapter 3, for now it suffices to say that studying EAS at 10^{17} eV allows for cross-checks between shower and LHC data, while ensuring that the cosmic ray flux is larger, requiring less time to gather statistically significant data sets.

1.3.3 MARTA engineering array

MARTA stands for Muon Array with RPCs for Tagging Air showers, where RPCs, in turn, stand for Resistive Plate Chambers [33]. These detectors are low cost, have high spatial and time resolutions and allow for a direct measurement of the muon content of EAS.

Each RPC is composed of millimetre thick gaseous volumes embedded in a uniform electric field produced by high-voltage electrodes applied to two resistive parallel plates enclosing the gas chambers. As charged shower particles pass through the chambers, the gas is ionised creating a cascade of

electrons under the influence of the electric field. A signal is thus induced on the read out electrodes, and transported by coaxial cables to the Data Acquisition System (DAQ).

Structurally, RPC units are laminated into three $2\,\mathrm{mm}$ thick resistive plates, separated by $1\,\mathrm{mm}$ gaseous volumes, containing tetrafluorathane (R-134a). The high resistivity of the plates prevents electrical discharges. The readout plane is external and divided into 64 individual pick-up electrods or pads of $14 \times 18\,\mathrm{cm}^2$ cm separated by a $1\,\mathrm{cm}$ guard ring. Thus, each unit has an active area of $1.2 \times 1.5\,\mathrm{m}^2$. The segmentation level is adjusted to improve the detector's spatial resolution and the signal pick-up electrodes are physically separated from the gas chambers, resulting in high-voltage insulation and gas tightness. Additionally, the formation of multi-gap gaseous volumes ensures a faster detector response to charge avalanches, yielding a high time resolution of a few nanoseconds, while enhancing the detector's efficiency. The right panel of Figure 1.4 shows the read-out segmented plane of a RPC unit, with coaxial cables and auxiliary electronics.

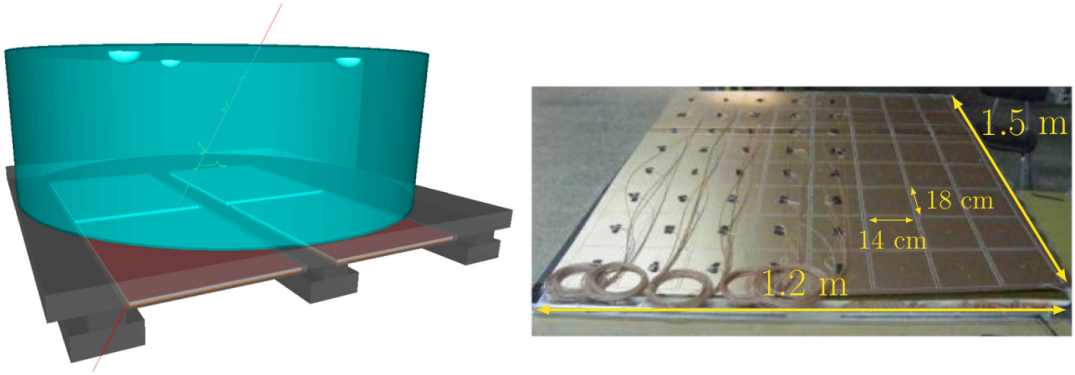

Figure 1.4: Left panel: Scheme of the implementation of MARTA stations with 4 RPCs (in red) hosted by a concrete structure (in black) below the WCD (in blue). Right panel: Photo of the RPC detector with the segmented read out plane, coaxial cables and auxiliary electronics. The picture shows the relevant dimensions. Both panels taken from [33].

Four RPC units, corresponding to an active area of $7\,\mathrm{m}^2$, are to be placed in a concrete structure underneath WCDs of the Infill array, at the Pierre Auger Observatory, forming hybrid surface stations where the water tank serves both an active calorimeter and trigger for the RPCs. In fact, the concrete thickness of $20\,\mathrm{cm}$ along with the water depth of $h_{\mathrm{water}} = 1.2\,\mathrm{m}$ correspond to a vertical mass overburden of $\sim 170\,\mathrm{g\,cm}^{-2}$, assuming a concrete density of $\rho_{\mathrm{concrete}} = 2.5\,\mathrm{g\,cm}^{-3}$. As such, most of the electromagnetic component is absorbed by the WCD, allowing a direct measurement of the muonic component. Note, however, that about 15% of the muons arriving at the WCDs completely degrade their energy within them, not being detected by the RPCs.

Due to the read-out plane segmentation we can define a fiducial region of the RPCs corresponding to the set of pads whose mass overburden, for a given shower inclination θ, exceeds $170\,\mathrm{g\,cm}^{-2}$. Within this region, the electromagnetic contamination is expected to be significantly attenuated, so we can further assume to first approximation that the number of muons is given by the number of active fiducial pads, in triggered stations. Another upshot of this segmentation is the capability of using the fiducial and non-fiducial areas of the RPCs as two separate detectors.

Finally, combining information from the WCDs and the RPCs allows for cross-calibrations and reduc-

tion of systematic uncertainties, making MARTA detectors serious candidates for measuring accurately the number of muons in EAS.

1.4 Objectives and Outline

With this work, we aim to:

- use the slope of the low tail of the distribution of the muon content of proton-initiated air showers over an ensemble of events to phenomenologically probe the first p-Air interaction;
- use the capabilities of the MARTA engineering array to reconstruct this slope and the first and second moments of the muon number distribution and determine the bias and resolution of these estimators.

This work is organised as follows. In Chapter 2 we motivate the connection between fluctuations of the muon content of EAS and those of a modified fraction of energy carried by hadronically interacting particles of the first generation of proton initiated showers, based on the Heitler-Mathews model and on the results of [9]. Moreover, we describe the re-sampling procedure used to constrain the forward production cross-section of neutral pions and the energy spectrum of hadrons using measurable features of the probability density function of the number of muons at the ground level in proton-initiated showers with low muonic content, irrespective of the used hadronic interaction model. Additionally, the feasibility of these measurements with enough precision to distinguish between hadronic interaction models, is discussed considering mixed composition scenarios, the cosmic ray flux dependence on the primary energy and current experimental uncertainties. The minimal number of events necessary to perform the described analysis is computed. In Chapter 3 we extend the results of the previous chapter to LHC equivalent center-of-mass energies, describe the simulation and reconstruction framework of Auger events, the surface detector trigger chain and explain in detail the reconstruction of the shower-to-shower distribution of the muon content of EAS from its correlation with the expected number of active fiducial pads of the resistive plate chambers of the MARTA engineering array at $450\,\mathrm{m}$ from the shower axis. Under ideal conditions, we show that the reconstruction method induces a large resolution at $E_0 = 10^{17.5}$ eV and $\theta = 30^\circ$, which prevents an unbiased measurement of the slope of the muon number distribution. The good quality and robustness of the reconstruction method are confirmed for a primary energy of $E_0 = 10^{18.5}$ eV and zentih angle $\theta = 40^\circ$, despite the residual electromagnetic contamination. The presence of a threshold effect which prevents a completely unbiased measurement of the features of the muon number distribution is addressed. Scenarios where the full Infill array is filled with RPCs or just an hexagonal conventional cell of the SD-750 array are studied and discussed separately. Finally, in Chapter 4 we draw conclusions and discuss future lines of work.

Chapter 2

Accessing the first interaction in proton induced EAS with low muonic content

As mentioned in Chapter 1, the first generations of extensive air showers induced by UHECRs offer a unique opportunity to study ultra-high energy hadronic interactions. However, these highly energetic events are extremely rare and occur at high altitudes, forbidding their direct access. Rather, the hadronic sector is probed through its decay into muons, which live long enough to be measured at the ground level. Since muons form at all stages of the showers development, the muonic component of EAS carries information about the hadronic activity in all shower generations, rendering the access to the first, most energetic, Cosmic-Ray-Air interaction very challenging. To circumvent this issue, [9] demonstrated that shower-to-shower fluctuations of the muon content of EAS are dominated by fluctuations of the fraction of energy contained in the hadronic sector of the first interaction in proton induced showers, α_1, while subsequent shower generations mostly contribute to the overall muon scale. Consequently, a way to probe the energy spectrum of hadrons at ultra-high energies was found.

In this chapter we elaborate on the correlation between the distributions of the muon content, N_μ, and α_1 over ensembles of proton induced EAS with low muonic content. After motivating the connection between the shower-to-shower fluctuations of these variables, we present the relevant features of their distributions and show how the slope of the low tail of the distribution of $\ln N_\mu$, denoted Λ_μ, can be converted into the corresponding slope of the α_1 distribution, Λ_α. Moreover, we show that the forward region of the energy spectrum of the leading neutral pion in p-Air ultra-high energy interactions can be equally described by an exponential slope, Λ_π, which we correlate with Λ_μ, effectively constraining the production cross section of π^0's at ultra-high energies through measurements of Λ_μ. The feasibility of the measurement of Λ_μ with enough accuracy to distinguish between hadronic interaction models is addressed in mixed primary composition scenarios, considering the power law dependence of the cosmic ray spectrum on the primary energy and experimental uncertainties.

Most of the findings presented in this Chapter were developed during the elaboration of the thesis and recently published in [34].

2.1 Shower-to-shower fluctuations of the muon content in proton induced EAS

The basic dynamics of the hadronic cascade can be described by the Heitler-Matthews model. Within this framework, the number of muons for a given primary energy, $N_\mu(E_0)$, obeys Equation (1.7), where the exponent, β, depends logarithmically on the total and charged pion multiplicity. This model assumes that in each interaction, the incoming energy is equally distributed among a fixed number of emerging pions. However, in reality, the multiplicity varies from interaction to interaction, throughout the cascade, and energy is shared unevenly among emerging particles, resulting in shower-to-shower fluctuations of the muon content of EAS. Thus, the simplistic Heitler-Matthews model must be modified to account for these effects

First, we establish that only the *average* number of muons, $\langle N_\mu(E_0)\rangle$, follows the power law expressed in Equation (1.7), where the exponent β should, in fact, read

$$\beta = \frac{\ln m}{\ln m_{\text{tot}}}, \tag{2.1}$$

where m_{tot} is the total multiplicity, instead of the total pion multiplicity, and m denotes the multiplicity of hadronically interacting particles, in place of the multiplicity of charged pions.

We now let m_g denote the average hadronic multiplicity of generation g, which can be written in terms of the number of hadronically interacting particles in generation g, M_g, and those of generation $g-1$, M_{g-1}, as

$$m_g = \frac{M_g}{M_{g-1}}. \tag{2.2}$$

Therefore, we can write the total number of muons as

$$N_\mu = M_{g_c} = \prod_{g=1}^{g_c} m_g, \tag{2.3}$$

where g_c is the average critical generation of an EAS. Note that in the previous equality we are assuming all hadronically interacting particles decay into muons, which, in turn, do not decay into other leptons. If we interpret the multiplicity as a random variable with probability density function $P(m)$, mean value m and dispersion $\sigma(m)$, and we assume that all realisations of m are independent, the dispersion of the average multiplicity of generation g is suppressed by

$$\sigma(m_g) = \frac{\sigma(m)}{\sqrt{M_{g-1}}}. \tag{2.4}$$

Thus, since M_{g-1} increases exponentially, and $\sigma(m)$ itself decreases with energy, the fluctuations of

the average multiplicity of deep shower generations are highly suppressed. Therefore, the fluctuations of the average multiplicity of the first interaction dominate the fluctuations of N_μ, via Equation (2.3).

Besides multiplicity fluctuations, the uneven and probabilistic way in which the energy is shared among emerging particles also induces fluctuations of the muon content of EAS. To see how they can be accounted for, we consider only fluctuations arising from the first interaction, $g = 1$, and assume that each of the m_1 resulting hadronic subshowers, carrying a fraction $x_i = \frac{E_i}{E_0}$ of the primary energy, decays, on average, into $\langle N_\mu(E_i)\rangle$ muons. Then, the muon content of a given EAS can be written as

$$N_{\mu,1}(E_0) = \sum_{i=1}^{m_1} \langle N_\mu(E_i)\rangle = \sum_{i=1}^{m_1} \mathcal{C} E_i^\beta = \sum_{i=1}^{m_1} \mathcal{C}(E_0 x_i)^\beta = \mathcal{C} E_0^\beta \sum_{i=1}^{m_1} x_i^\beta = \langle N_\mu(E_0)\rangle\, \alpha_1, \tag{2.5}$$

where we defined $\mathcal{C} = \xi_c^{-\beta}$ and

$$\alpha_1 \equiv \sum_{i=1}^{m_1} x_i^\beta. \tag{2.6}$$

This quantity can be more easily interpreted by taking the limits $\beta \to 0$ and $\beta \to 1$. In the former limit, $\alpha_1 = m_1 = 1$ is the hadronic multiplicity of the interaction [a], while in the later, $\alpha_1 = \sum_i x_i$ is the fraction of energy contained in the hadronic sector of the first generation. Thus, α_1 is a modified fraction of energy carried by hadronically interacting particles, that accounts for both multiplicity and energy fluctuations. The weight given to fluctuations of these quantities is determined by β.

Following the same reasoning, we can introduce fluctuations from subsequent shower generations, g, each with a given α_g value. Defining α_g recursively as

$$\alpha_g \equiv \frac{N_{\mu,g}}{N_{\mu,g-1}}, \tag{2.7}$$

for each particle generation, allows us to cast N_μ as

$$N_\mu(E_0) = \langle N_\mu(E_0)\rangle \prod_{g=1}^{g_c} \alpha_g = \alpha_1 \omega \quad \text{with} \quad \omega = \langle N_\mu(E_0)\rangle \prod_{g=2}^{g_c} \alpha_g. \tag{2.8}$$

Mind α_2 in the above equation. If the multiplicity of the first interaction is m_1, then there are m_1 subshowers contributing to α_2, so that its fluctuations with respect to α_1 are suppressed by a factor of $\sim 1/\sqrt{m_1}$, provided subshowers are independent from one another. Proceeding recursively, it is clear that fluctuations of α_g in deeper shower generations become smaller and smaller, and the distribution of α_g over an ensemble of showers narrower. Hence, we should expect that fluctuations of N_μ are mostly dominated by those of α_1, strongly correlating these quantities. On the other hand, the actual muon scale depends on all shower generations through Equation (2.8).

This reasoning was thoroughly tested and verified in proton induced EAS, simulated with CORSIKA and CONEX, at different primary energies and zenith angles, as well as for the hadronic interaction models EPOS-LHC, QGSJET II-04 and SIBYLL 2.3c, as discussed in [9]. There, it was shown that shower-to-shower fluctuations of the number of muons are dominated by fluctuations of α_1, proving

[a] Note that using Equation (2.1), $\beta = 0$ is equivalent to requiring $m_1 = 1$.

that the former can be used to access multiparticle production properties of the first cosmic-ray Air interaction.

2.1.1 A word on the EAS muon puzzle

The above analysis allows us to express the muon puzzle in a more quantitative manner. Writing the number of muons in terms of α_g, like in Equation (2.8), we can propose two scenarios to increase the number of muons in simulations: either some α_g is larger than simulated or all α_g are.

In the first scenario, the most likely α to be larger than predicted by simulations is that of the first generation, α_1, where the energy is highest and exotic phenomena are more likely to occur. Thus, the effect of $\alpha_1 \to \alpha_1 + \delta\alpha_1$ would affect N_μ linearly, while the effect on its fluctuations would depend on the considered hadronic interaction model. Note that perturbing just α_1 would imply a threshold for new phenomena.

In the second scenario, the total increase in N_μ would be $\propto \left(1 + \frac{\delta\alpha}{\alpha}\right)^g$, if all α up to generation g were increased. In this case, a small change in the energy imbalance towards the hadronic sector, propagates throughout the shower. Therefore, letting $\delta\alpha/\alpha \sim 5\%$ yields an overall $\sim 35\%$ increase in the muon content of EAS after six generations. The impact of this scenario in the fluctuations of N_μ would be small, since these are dominated by fluctuations of α_1 [9].

The fact that the measured fluctuations of the muon content of EAS are consistent with simulations, favours either the second scenario, where the change $\delta\alpha$ in each generation is small, or the first if $\delta\alpha_1$ is increased reasonably without changing its fluctuations. The latter case would require a very specific change in α_1, constraining the possible exotic mechanisms responsible for it.

2.2 Constraining the hadron energy spectrum of the first p-Air ultra high energy interaction

2.2.1 Correlation between α_1 and N_μ

As pointed out above, shower-to-shower fluctuations of N_μ correlate with those of a quantity of the first Cosmic-Ray-Air interaction, α_1, which is a modified fraction of the primary energy contained in the hadronic sector of the first shower generation. To deepen our study of these distributions and their connection, about 10^6 proton initiated showers with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67^\circ$ were simulated with CONEX, using three post-LHC hadronic interaction models: EPOS-LHC, QGSJET II-04 and SIBYLL 2.3d[b]. The number of muons reaching the ground, N_μ, is given at 1400 m above sea level, the average altitude of the Pierre Auger Observatory. We chose to use inclined showers at this stage since their electromagnetic component is almost completely absorbed by the larger slant depth of atmosphere, with respect to vertical showers, guaranteeing a purer measurement of the muonic component by the WCDs of the Pierre Auger Observatory.

[b]Note that SIBYLL 2.3d only became available after the submission of [34], where SIBYLL 2.3c was used.

For each event, the values of α_1 and N_μ were computed, and their distributions over the ensemble of simulated showers were built. Figure 2.1 shows the correlation between the distributions of α_1 and N_μ, over the simulated ensemble of showers, along with the correlation factor $\rho_{\mu,\alpha} = \mathrm{cov}(\alpha_1, N_\mu)/\sigma_{N_\mu}\sigma_{\alpha_1}$ for the EPOS-LHC model. The corresponding plots for QGSJET II-04 and SIBYLL 2.3d can be found in Appendix A.

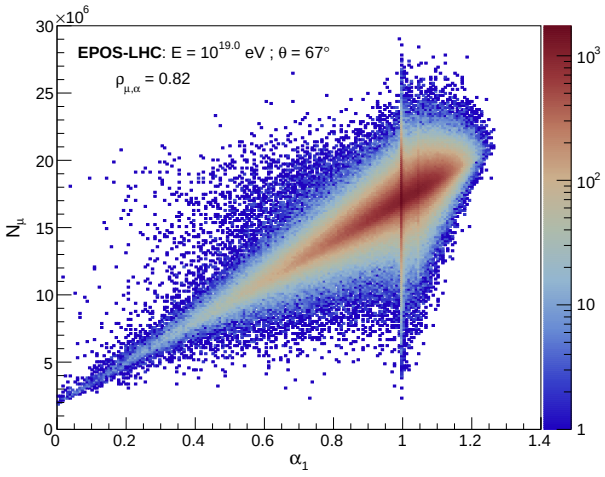

Figure 2.1: Correlation between the distributions of α_1 and N_μ over an ensemble of $\sim 10^6$ proton induced EAS, with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Simulations were performed with CONEX, using the hadronic interaction model EPOS-LHC. The correlation factor between the two quantities can be read in the upper left corner.

The positive correlation between the two variables displayed in Figure 2.1 is apparent from their high correlation factor $\rho_{\mu,\alpha} \simeq 0.82$. Similar values of $\rho_{\mu,\alpha} \simeq 0.80$ and $\simeq 0.77$ were found for QGSJET II-04 and SIBYLL 2.3d, respectively, in agreement with the results of [9]. Furthermore, $N_\mu > 0$ when $\alpha_1 = 0$, unlike what is expected from Eq.(2.8). This is due to photon-pion production, as discussed in [9]. The vertical structure along α_1, corresponds to diffractive events in which the incident proton scatters quasi-elastically of the air nucleus, only interacting with its color field. Given the low energy transfer, the scattered proton is still highly energetic and hence able to yield large multiplicities of hadronically interacting secondary particles in subsequent interactions. As a result, the number of muons reaching the ground level becomes more sensitive to deeper interactions, rather than to the first one. Thus, N_μ fluctuates considerably within a narrow band of α_1 values, locally breaking their correlation. A similar, but less pronounced, feature close to $\alpha_1 \simeq 1.05$ can be observed for all models. Weather this feature has physical meaning or whether it is just an artefact, is not known, and it appears to be irrelevant for our analysis.

2.2.2 Distributions of α_1 and $\ln N_\mu$

We now turn to the shape of the shower-to-shower distributions of α_1 and $\ln N_\mu$, i.e., their probability density functions if properly normalised. In fact, from Equation (2.8), we know these are related by

$$\ln N_\mu = \ln \alpha_1 + \ln \omega, \tag{2.9}$$

from where it follows that the probability density function of $\ln N_\mu$, $P(\ln N_\mu)$, is a convolution of the probability density function of $\ln \alpha_1$, $f(\ln \alpha_1)$, with that of $\ln \omega$, $g(\ln \omega)$, in the form

$$P(\ln N_\mu) = \int_{-\infty}^{\infty} f(\ln \alpha_1) g(\ln N_\mu - \ln \alpha_1)\, \mathrm{d} \ln \alpha_1 \,. \tag{2.10}$$

Figure 2.2 shows the mentioned distributions for the ensemble of $\sim 10^6$ proton initiated showers simulated with CONEX for the hadronic interaction models EPOS-LHC, QGSJET II-04 and SIBYLL 2.3d: the left panel corresponds to the probability density function of α_1 and the right panel corresponds to that of $\ln N_\mu$.

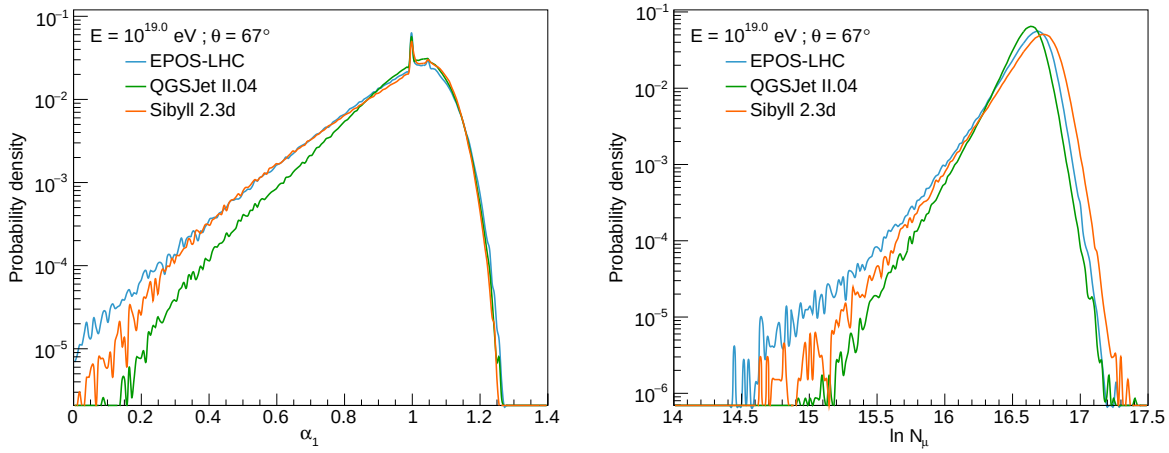

Figure 2.2: Distributions of α_1 (left panel) and $\ln N_\mu$ (right panel) over an ensemble of $\sim 10^6$ proton induced EAS, with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Simulations were performed with CONEX, using the hadronic interaction models EPOS-LHC (blue), QGSJET II-04 (green) and SIBYLL 2.3d (orange).

The α_1 distribution displays a left low tail corresponding to events where an appreciable fraction of the primary energy is carried by the electromagnetic sector of the first p-Air interaction. The sharp peak at $\alpha_1 \simeq 1$ corresponds to diffractive events, as mentioned earlier, where the incident proton scatters quasi-elastically of the Air nucleus. Similarly, the distribution of $\ln N_\mu$ displays low and high tails, corresponding to showers with low or high muon content. It is evident from Figure 2.2, that the steepness of the low tail of both distributions is model dependent, along with the average values and widths of the α_1 and $\ln N_\mu$ probability densities. Naturally, these differences stem from the different physical laws and assumptions employed by each model in simulating hadronic interactions at energies and in regions of the kinematic phase space beyond the reach of current human-man colliders. A thorough study on the

model dependence of the steepness of the low tail of these distributions must be done elsewhere.

The shape of the low tail of the α_1 and $\ln N_\mu$ distributions is approximately exponential. In order to characterise each tail in a quantitative manner, we fitted them to exponential functions of the form

$$\frac{d\mathcal{N}}{d\alpha_1} = \mathcal{C}_\alpha \exp\left(\frac{\alpha_1}{\Lambda_\alpha}\right), \tag{2.11}$$

and

$$\frac{d\mathcal{N}}{d\ln N_\mu} = \mathcal{C}_\mu \exp\left(\frac{\ln N_\mu}{\Lambda_\mu}\right), \tag{2.12}$$

where $\frac{d\mathcal{N}}{dx}$ is the number of events, $\mathcal{N}$, within a narrow bin of width dx and $\mathcal{C}_\alpha$, Λ_α, $\mathcal{C}_\mu$ and Λ_μ are free parameters. Henceforth, Λ_α and Λ_μ will be referred to as exponential slopes of the α_1 and $\ln N_\mu$ distributions, respectively.

Since the bulk and the end of the low tail of the α_1 and $\ln N_\mu$ distributions are not exponential, the fit window from which we extract the values of Λ_α and Λ_μ was chosen so that only 5% deviations from a pure exponential function were allowed. In practice, we choose a value for the upper limit of the fit a fixed distance below the maximum of either distribution, Δ_{max}, and with a fixed starting width that allows the convergence of the fit, and then decrease the lower bound of the fit window, bin by bin, until the vertical distance between the real distribution and the fit differs by more than 5% of the distribution's value at that point. This criterion has the advantage of being independent of the model. Additionally, we verified that small variations of the fit window, do not compromise the qualitative results obtained in this chapter.

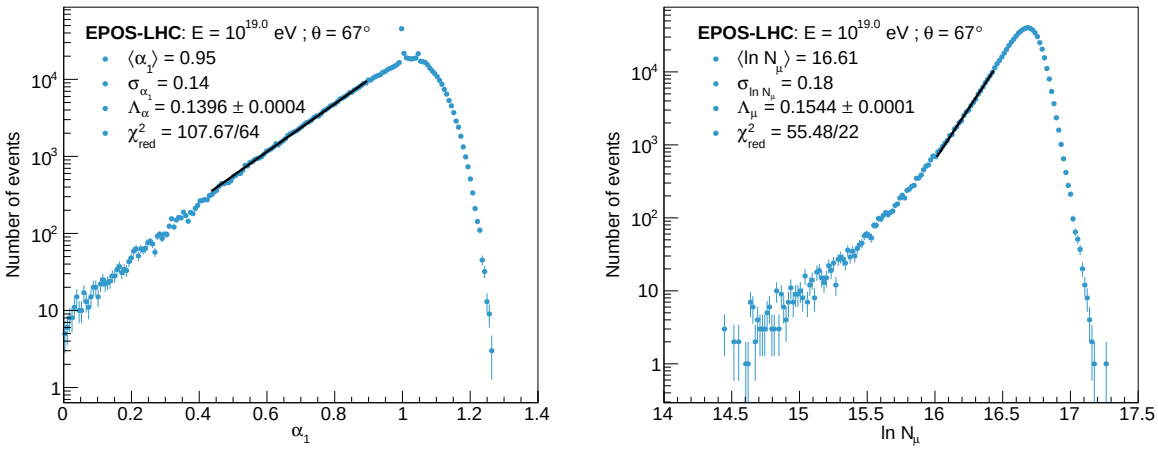

Figure 2.3: Fit to the tail of the α_1 (left panel) and $\ln N_\mu$ (right panel) distributions over an ensemble of $\sim 10^6$ proton induced EAS, with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Simulations were performed with CONEX, using EPOS-LHC (blue), and the fit curve is shown in black.

The fit was performed with the least squares method and implemented with the `MINUIT` minimisation package for all hadronic interaction models. Figure 2.3 shows fits to the low tails of the distributions of α_1 (left panel) and $\ln N_\mu$ (right panel) with the aforementioned criterion, for the EPOS-LHC model. The fit curve is shown in black and the distributions in blue. The corresponding fits to the tails of the distributions of α_1 and $\ln N_\mu$ for QGSJET II-04 and SIBYLL 2.3d can be found in Appendix A. In all cases,

the value of $\Delta^{\alpha}_{\mathrm{max}}$, the mean value of α_1, $\langle\alpha_1\rangle$, the root-mean square of the distribution of α_1, σ_{α_1}, its exponential slope, Λ_α, and the χ^2 value of the fit divided by the number of degrees of freedom, χ^2_{red}, are shown in Table 2.1, whereas the obtained values of $\Delta^{\mu}_{\mathrm{max}}$, $\langle \ln N_\mu\rangle$, $\sigma_{\ln N_\mu}$, Λ_μ and χ^2_{red} are displayed in Table 2.2.

Parameters	EPOS-LHC	QGSJET II-04	SIBYLL 2.3d
$\Delta^{\alpha}_{\mathrm{max}}$	0.10	0.10	0.10
$\langle\alpha_1\rangle$	0.95	0.97	0.96
σ_{α_1}	0.14	0.12	0.14
Λ_α	0.1396 ± 0.0004	0.1101 ± 0.0003	0.1497 ± 0.0006
χ^2_{red}	107.67/64	104.79/67	71.00/54

Table 2.1: Values of $\Delta^{\alpha}_{\mathrm{max}}$, $\langle\alpha_1\rangle$, σ_{α_1}, Λ_α and χ^2_{red} for the α_1 distribution in proton induced EAS with $E_0 = 10^{19}$ eV and $\theta = 67^\circ$ simulated with CONEX, for all hadronic interaction models.

Parameters	EPOS-LHC	QGSJET II-04	SIBYLL 2.3d
$\Delta^{\mu}_{\mathrm{max}}$	0.25	0.25	0.25
$\langle \ln N_\mu\rangle$	16.61	16.59	16.65
$\sigma_{\ln N_\mu}$	0.18	0.14	0.18
Λ_μ	0.1544 ± 0.0001	0.1277 ± 0.0001	0.1599 ± 0.0001
χ^2_{red}	55.48/22	112.30/24	21.13/27

Table 2.2: Values of $\Delta^{\mu}_{\mathrm{max}}$, $\langle \ln N_\mu\rangle$, $\sigma_{\ln N_\mu}$, Λ_μ and χ^2_{red} for the $\ln N_\mu$ distribution in proton induced EAS with $E_0 = 10^{19}$ eV and $\theta = 67^\circ$ simulated with CONEX, for all hadronic interaction models.

The average α_1 and $\ln N_\mu$ values are in agreement with the results obtained in [9] for the EPOS-LHC and QGSJET II-04 models. Note that we cannot compare the results obtained with SIBYLL 2.3d since at the time of the submission of [9] only SIBYLL 2.3c was available. The values of $\chi^2_{\mathrm{red}} \sim 1$ extracted from the fit to the tail of the α_1 distribution reflect its high quality, while showing that the criterion chosen to define the fit window is reasonable. Note that $\Delta^{\alpha}_{\mathrm{max}}$ was fixed to initiate the fit, in a model independent way. More importantly, the Λ_α values obtained for each model are incompatible with each other, characterising, in a model dependent way, the energy spectrum of hadrons of the first shower generation.

On the other hand, the higher χ^2_{red} values found when fitting the tail of the $\ln N_\mu$ distribution for the EPOS-LHC and QGSJET II-04 models reflect the fact that the tail is not perfectly exponential, but rather slightly concave closer to its low end. Additionally, the calculation of χ^2 assumes that the content of each bin fluctuates in a Gaussian way, and that bins fluctuate independently which is not the case in simulations, leading to an overestimation of the χ^2 value. In any case, the acceptable quality of the fit is clear, and the obtained values of Λ_μ for each model are incompatible with each other, re-reinforcing the fact that different hadronic models are instantiations of different physical laws.

2.2.3 Re-sampling the α_1 and $\ln N_\mu$ distributions

We now exploit the correlation between the distributions of α_1 and $\ln N_\mu$ to investigate the connection between Λ_μ and Λ_α, and determine if the conversion of the former to the latter is also model dependent. To do so, we perturb the tail of the α_1 distribution below $\alpha_1 = 1$ by re-sampling the ensemble of events and selecting pairs of $(\ln N_\mu, \alpha_1)$ with a probability density function $p(\alpha_1) \propto e^{(\alpha_1 - 1)/\delta\Lambda_\alpha}$, where $\delta\Lambda_\alpha$ parametrises the perturbation. Thus, the new α_1 distribution can be written in terms of the original one as

$$\left(\frac{d\mathcal{N}}{d\alpha_1}\right)_{\text{new}} = \begin{cases} \dfrac{d\mathcal{N}}{d\alpha_1} \exp\left(\dfrac{\alpha_1 - 1}{\delta\Lambda_\alpha}\right) & \text{for } \alpha_1 < 1 \\ \dfrac{d\mathcal{N}}{d\alpha_1} & \text{for } \alpha_1 \geq 1 \end{cases}, \tag{2.13}$$

Hence, the exponential slope of the α_1 distribution is effectively changed by $\Lambda_\alpha^{-1} \rightarrow \Lambda_\alpha^{-1} + \delta\Lambda_\alpha^{-1}$, while keeping its non-exponential features. To implement the re-sampling procedure, we partition the ensemble of events into subsets of $d\mathcal{N}/d\alpha_1$ elements where $\alpha_1^* \in [\alpha_1, \alpha_1 + d\alpha_1]$, and for each of these subsets we choose, at random and following a uniform probability density function, floor$\left[\left(d\mathcal{N}/d\alpha_1\right)_{\text{new}}\right]$ [c] pairs of $(\ln N_\mu, \alpha_1)$. Hence, through the correlation between the α_1 and $\ln N_\mu$ distributions, the slope of the latter is also changed $\Lambda_\mu^{-1} \rightarrow \Lambda_\mu^{-1} + (\delta\Lambda'_\mu)^{-1}$. After exhausting the ensemble of events, the tails of the new, re-sampled α_1 and $\ln N_\mu$ distributions are fitted with the criterion discussed previously, yielding pairs of slope values $(\Lambda_\mu, \Lambda_\alpha)$ for each perturbation $\delta\Lambda_\alpha$. It is important to note that these perturbations are small and assume that the hadronic properties of the interaction remain unchanged, covering a limited region around the nominal $(\Lambda_\mu, \Lambda_\alpha)$ values. Moreover, these perturbations do not try to emulate any physical process, which would ultimately change other properties of the multiparticle production in p-Air interactions. Figure 2.4 shows the nominal and re-sampled distributions of α_1 and $\ln N_\mu$, in grey and blue, respectively. The fit to the low tail region of both distributions is also shown, in black.

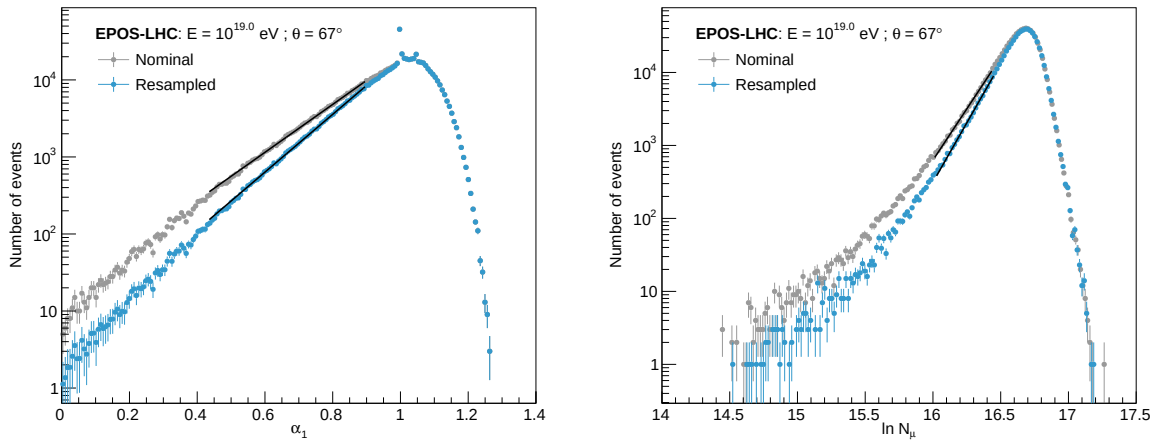

Figure 2.4: Left panel: Fitted nominal (grey) and re-sampled (blue) α_1 distribution. Right panel: fitted nominal (grey) and re-sampled (blue) $\ln N_\mu$ distribution. Both distributions are over an ensemble of 10^6 proton induced EAS, simulated with CONEX with the EPOS-LHC hadronic interaction model.

[c] The floor function is defined by floor: $\mathbb{R} \rightarrow \mathbb{N}$; floor$(x) = n$, if $n \leq x \leq n + 1$, for $n \in \mathbb{N}$

2.2.4 $\Lambda_\mu \leftrightarrow \Lambda_\alpha$ conversion

By varying $\delta\Lambda_\alpha$, we can parametrise curves in the $(\Lambda_\mu, \Lambda_\mu)$ space for each model, which allow the conversion of a measurement of Λ_μ into a value of Λ_α. These calibration curves are represented by the solid lines shown in Figure 2.5 for the interaction models EPOS-LHC (blue), QGSJET II-04 (green) and SIBYLL 2.3d (orange), where the filled circles indicate the nominal pair $(\Lambda_\mu, \Lambda_\alpha)$. Also displayed in Figure 2.5 are dashed lines corresponding to the same re-sampling prescription described in 2.2.3, but with perturbations $\delta\Lambda_\mu$ applied directly to the tail of the distribution of $\ln N_\mu$. Specifically, the distribution of $\ln N_\mu$, was modified below its maximum, $\ln N_\mu^*$, by selecting pairs $(\ln N_\mu, \alpha_1)$ for each event with a probability $p(\ln N_\mu) \propto e^{(\ln N_\mu - \ln N_\mu^*)/\delta\Lambda_\mu}$. The tail of the distribution of α_1 is thus affected by the re-sampling prescription via its correlation with N_μ. The circumferences represent the nominal $(\Lambda_\mu, \Lambda_\alpha)$ values. Note that in [34] SIBYLL 2.3c was used instead of SIBYLL 2.3d. Although the qualitative conclusions have not changed, we refer to Appendix B for a comparison between SIBYLL versions.

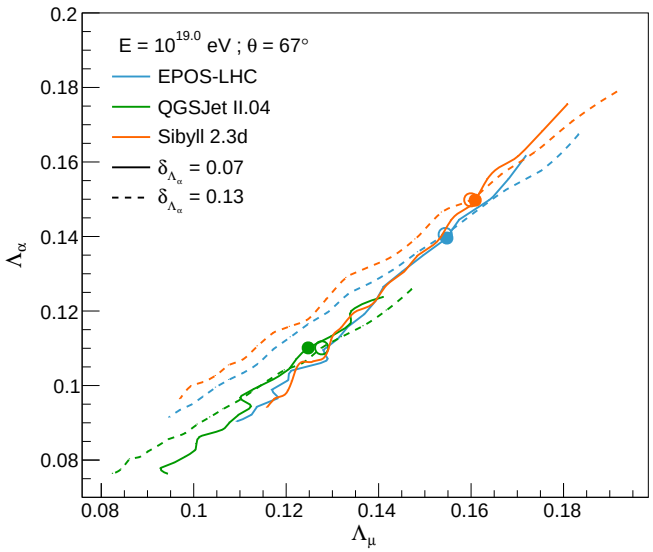

Figure 2.5: Conversion curves between Λ_μ and Λ_α, for the hadronic interaction models EPOS-LHC (blue), QGSJET II-04 (green) and SIBYLL 2.3d (orange). The solid lines show how Λ_μ changes if Λ_α is changed, while the dashed lines represent the converse situation. Solid dots and circumferences represent the nominal Λ_α and Λ_μ values.

From Figure 2.5 it is apparent that there is a monotonic relation between Λ_α and Λ_μ for all models, for both re-sampling procedures. The curves parametrised by $\delta\Lambda_\alpha$ are more inclined than those parametrised by $\delta\Lambda_\mu$. Furthermore, the maximum deviation in Λ_α across models is $\delta^\alpha_{\Lambda_\alpha} = 7\%$, when Λ_α is modified and Λ_μ is affected, and $\delta^\mu_{\Lambda_\alpha} = 13\%$, when Λ_μ is modified and Λ_α is affected. This difference is expected to be related with the imperfect correlation between the distributions of α_1 and $\ln N_\mu$, which in itself comes from the subdominant contribution of the deeper shower generations to the fluctuations of the muon content of EAS.

Mathematically, the mismatch between conversion curves parametrised by $\delta\Lambda_\alpha$ and $\delta\Lambda_\mu$ results from the interplay between at least two effects:

- For a fixed direct perturbation $\delta\Lambda$ to the low tails, the absolute variation of Λ_μ is larger than that of Λ_α, assuming a perfect correlation between the α_1 and $\ln N_\mu$ distributions.

  Let $\Delta\Lambda(\delta\Lambda)$ be the variation of Λ due to perturbation $\delta\Lambda$. Given the definition of $\delta\Lambda$ we have,

$$\Lambda^{-1} \to \tilde{\Lambda}^{-1} = \Lambda^{-1} + \delta\Lambda^{-1} \implies \Delta\Lambda(\delta\Lambda) = \tilde{\Lambda} - \Lambda = -\frac{\Lambda^2}{\Lambda + \delta\Lambda}. \tag{2.14}$$

  Thus, for fixed $\delta\Lambda > 0$, $\Delta\Lambda(\delta\Lambda)$ is a decreasing function of Λ, so that $|\Delta\Lambda(\delta\Lambda)|$ increases with Λ. Since $\Lambda_\mu > \Lambda_\alpha$, $|\Delta\Lambda_\mu(\delta\Lambda)| > |\Delta\Lambda_\alpha(\delta\Lambda)|$.

  Consequently, if the correlation between α_1 and N_μ was perfect, the slope, m, of the calibration curves for a given model and perturbation $\delta\Lambda$ would obey:

$$m = \frac{\Delta\Lambda_\alpha(\delta\Lambda)}{\Delta\Lambda_\mu(\delta\Lambda)} < 1, \tag{2.15}$$

  and would not depend on the perturbed distribution.

- If the tail of one distribution is perturbed by $\delta\Lambda$, the perturbation felt by the distribution affected by the re-sampling procedure is suppressed by the imperfect correlation between α_1 and $\ln N_\mu$.

  Selecting events after a direct perturbation to the tail of the α_1 ($\ln N_\mu$) distribution, might pick up some that do not contribute to the tail of the $\ln N_\mu$ (α_1) distribution, and leave behind events that do. Consequently, not only the tail of the distribution affected by the re-sampling procedure is changed, but also other parts of this distribution. As a result, there is an overall suppression of the effect of a given perturbation to the tail of the affected distribution, during the re-sampling procedure. In other words, for a fixed perturbation it is more efficient to directly perturb Λ_μ (Λ_α) than to perturb Λ_α (Λ_μ) and measure its effect on Λ_μ (Λ_α). Therefore, if $\delta\Lambda$ is a direct perturbation to the tail of the α_1 ($\ln N_\mu$) distribution, the perturbation induced on Λ_μ (Λ_α) is $\delta\Lambda'$ such that $|\Delta\Lambda_\mu(\delta\Lambda)| > |\Delta\Lambda_\mu(\delta\Lambda')|$ ($|\Delta\Lambda_\alpha(\delta\Lambda)| > |\Delta\Lambda_\alpha(\delta\Lambda')|$). If the correlation was perfect, both situations would be equally efficient.

  As a consequence, letting m be the slope of a hypothetical conversion curve between perfectly correlated distributions, $m_{\alpha\to\mu}$ be the slope of the calibration curves parametrised by $\delta\Lambda_\alpha$ and $m_{\mu\to\alpha}$ be the slope of the calibration curves parametrised by $\delta\Lambda_\mu$, we have

$$m_{\alpha\to\mu} = \frac{\Delta\Lambda_\alpha(\delta\Lambda)}{\Delta\Lambda_\mu(\delta\Lambda')} > m \quad ; \quad m_{\mu\to\alpha} = \frac{\Delta\Lambda_\alpha(\delta\Lambda')}{\Delta\Lambda_\mu(\delta\Lambda)} < m. \tag{2.16}$$

  In the limit where α_1 and N_μ are completely uncorrelated, $m_{\alpha\to\mu} \to \infty$ and $m_{\mu\to\alpha} \to 0$. As a result, the conversion curves parametrised by $\delta\Lambda_\alpha$ are steeper and correspond to an anti-clockwise rotation of the curves parametrised by $\delta\Lambda_\mu$, around the nominal values of $(\Lambda_\mu, \Lambda_\alpha)$. A sketch of this considerations is shown in Figure 2.6, whose legend contains the necessary information for its interpretation. Note that distributions affected by the re-sampling procedure are changed as a whole, even though the most prominent change is the steepness of their low tails, as a result of their good but imperfect correlation with distributions whose tail is directly perturbed.

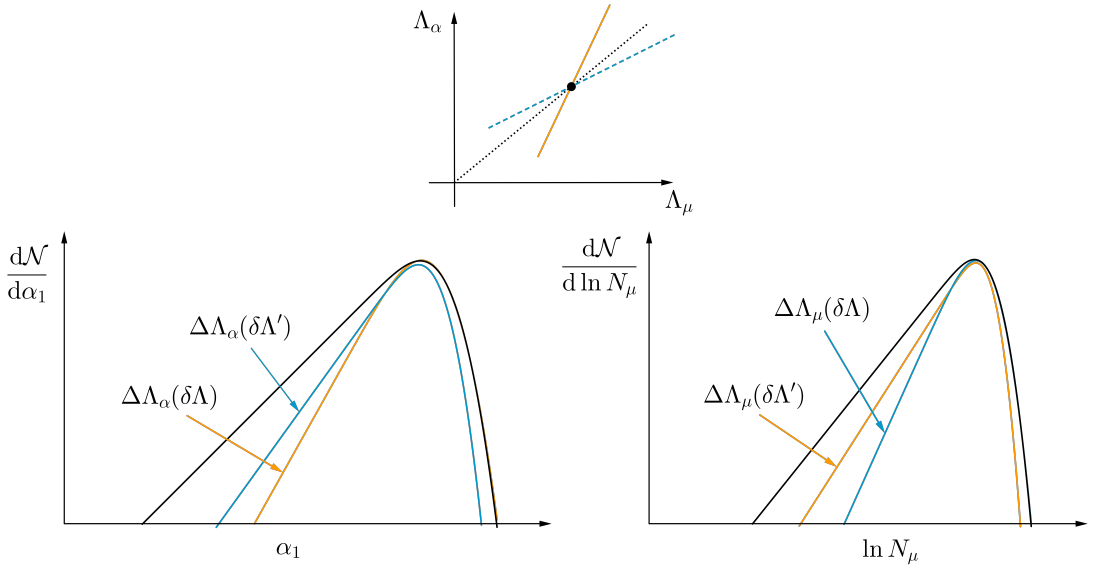

Figure 2.6: Sketch of the distributions resulting from the re-sampling procedure and respective calibration curves. Nominal distributions are shown in black. Orange and blue distributions of α_1 correspond to direct and induced perturbations, respectively, whereas blue and orange distributions of $\ln N_\mu$ correspond to the converse scenario. Direct perturbations are parametrised by $\delta\Lambda$, while perturbations due to the resampling procedure are represented by $\delta\Lambda'$. The orange solid conversion curve is parametrised by direct perturbations to the tail of the α_1 distribution, while the dashed blue calibration curve is parametrised by direct perturbations to the tail of the distribution of $\ln N_\mu$. The dotted black calibration curve corresponds to a hypothetical situation where the distributions of α_1 and $\ln N_\mu$ are perfectly correlated.

The result of the above effects is that the curves parametrised by $\delta\Lambda_\mu$ can be obtained from the curves parametrised by $\delta\Lambda_\alpha$ by a rotation followed by a stretch, since the total perturbation is finite. This clockwise rotation is centred at the nominal $(\Lambda_\mu, \Lambda_\alpha)$ values and drives the calibration curves apart, explaining why $\delta^{\mu}_{\Lambda_\alpha} > \delta^{\alpha}_{\Lambda_\alpha}$. Note that this *ad-hoc* considerations do not bring insight into the origin of the universality of the calibration curves across models. Naturally, this universality is dependent on the actual modification of the spectrum of hadrons of the first interaction. What we have established, was simply that Λ_μ is probe of Λ_α. In future works, the parameters of the simulations will be changed *a priori*, inducing a more physical change in Λ_α, and thus on Λ_μ, through what we have proved.

We also note that the solid circles and circumferences do not exactly coincide. This stems from the stochastic nature of the re-sampling process in which for each bin $\mathrm{d}\alpha_1$ ($\mathrm{d}\ln N_\mu$) we are selecting subsets of events with the same size as the original set, without caring for repetition. Thus, even if the distribution of α_1 is unperturbed, the re-sampling process might not pick up every event with the corresponding $\ln N_\mu$ value only once, but rather more than one time, leaving some events out. This reasoning holds if we instead consider direct perturbations to the tail of the distribution of $\ln N_\mu$.

At any rate, in showers with low muonic content, since α_1 refers to the first interaction and $\ln N_\mu$ to a quantity measured at the ground level, causality dictates that changes in the shower-to-shower distribution of α_1 affect the $\ln N_\mu$ distribution, and not the other way around. Hence the calibration curves parametrised by $\delta\Lambda_\alpha$ should be the ones used to convert a measurement of Λ_μ into a value of Λ_α, with a

maximum difference across models of $\sim 7\%$. Therefore, even though the different models are based on different physical laws, the conversion from Λ_μ to Λ_α is almost universal, i.e., model independent.

The above study has shown that the energy spectrum of hadronically interacting particles of the first p-Air interaction can be constrained by measurements of Λ_μ in showers with low muonic content, via its conversion into values of Λ_α. The latter quantity is more fundamental than a single or average value of α_1, since it is directly related to multiparticle production properties at the highest energies.

A few remarks are in need. By the superposition principle, for primaries with mass number A the distribution of $\ln N_\mu$ is a convolution of A independent $\ln N_\mu$ distributions for proton induced showers. Thus, the fluctuations of the muon content of EAS induced by primaries heavier than proton become more Gaussian and narrow by a factor of $\sim 1/\sqrt{A}$ as A increases. Consequently, the low tail of these distributions is non-exponential and hidden underneath the tail of the distribution of $\ln N_\mu$ in proton induced EAS. While this means that the $\Lambda_\mu \to \Lambda_\alpha$ calibration is only valid for proton primaries, it also means that Λ_μ could still be measured in a mixed composition scenario, provided it includes proton primaries and the fit range is judiciously chosen. Moreover, we have considered a fixed, well defined primary energy for the entire ensemble of events, while in reality we should consider an energy bin, over which N_μ varies with E_0^β and the cosmic ray spectrum varies in such a way that more energetic events are rarer. Finally, in real measurements one must consider the uncertainties in the reconstruction of the primary energy and muon content of EAS, yielding a smeared distribution of $\ln N_\mu$, due to the finite resolution of the detector array.

2.3 Constraining the energy spectrum of neutral pions in ultra-high energy p-Air interactions

We have already established the positive correlation between the modified fraction of energy contained in the hadronic sector of the first p-Air interaction, α_1, and the number of muons at the ground level N_μ in showers with low muonic content. To first approximation, the fraction of energy contained in the electromagnetic sector of the first interaction, f_{em} can be written $f_{\mathrm{em}} \simeq 1 - \alpha_1$, and about 90% of this fraction is carried by neutral pions, as represented in Figure 1.2.

By isospin symmetry, the number and energy carried by the bulk of neutral and charged pions is roughly the same, so that the energy imbalance toward π^0's in showers with low muon content is due to fast π^0's. These are connected with forward production and valence quarks, and by taking most of the primary energy, leave less energy in the hadronic shower to produce muons. Therefore, it is natural to study the connection between the low tail of the distribution of $\ln N_\mu$ and the forward region of the energy spectrum of neutral pions. In fact, we shall consider only the *leading* (most energetic) neutral pion and denote by x_{L} the fraction of the primary energy in the laboratory frame taken by these particles.

The choice of considering the energy spectrum of the leading neutral pion instead of the full spectrum is twofold. Firstly, in deep inelastic scattering it is common that the leading particle takes a significant portion of the available initial state energy, so that the features of the forward region of the energy

spectrum should be mostly accounted for by considering only the leading particle. Secondly, it would be impractical, if not impossible, to perform the *ad hoc* re-sampling method we described in Section 2.2.3 to further explore the constraining power of Λ_μ, since there would be an energy spectrum of neutral pions rather than single quantity per event, breaking the one-to-one relation between N_μ and the fraction of energy taken by a given neutral pion. A few caveats of this choice will be brought up throughout this section.

2.3.1 Correlation between x_L and N_μ

Figure 2.7 shows the correlation between x_L and N_μ, with the correlation factor $\rho_{\mu,\pi}$ for the EPOS-LHC model. The correlation between these two distributions for the remaining hadronic interaction models can be found in Appendix A.

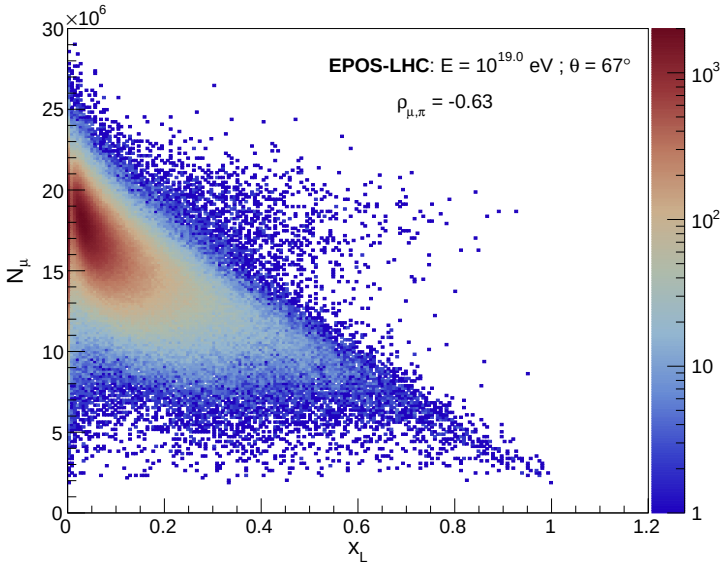

Figure 2.7: Correlation between the distributions of x_L and N_μ over an ensemble of $\sim 10^6$ proton induced EAS, with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Simulations were performed with CONEX, using the hadronic interaction model EPOS-LHC. The correlation factor between the two quantities can be read in the upper right corner.

The negative correlation between x_L and N_μ is apparent from the relatively high correlation factor of $\rho_{\mu,\pi} \simeq -0.66$. Similar values of $\rho_{\mu,\pi} \simeq -0.61$ and $\rho_{\mu,\pi} \simeq -0.60$ were found for QGSJET II-04 and SIBYLL 2.3d, respectively. This reasonable correlation was to be expected since x_L is almost complementary to α_1. However, $|\rho_{\mu,\pi}| < \rho_{\mu,\alpha}$ for all models. This might be due to the fact that we are only considering the leading neutral pion, rather than the full energy spectrum. Thus, in events where more than one neutral pion takes most of the energy, and hence with a relatively low x_L, the muon content of the shower can still be quite low due to a overall energy imbalance towards the electromagnetic sector. Once again, a vertical line at $x_L \simeq 0$ corresponding to diffractive events can be seen, worsening the correlation between this variable and N_μ. We must note, however, that we are only considering first generation neutral pions, which already shows its dominant role in muon production, with respect to deeper generations. A further

study of the actual impact of later shower generations was not conducted, and could be done elsewhere.

2.3.2 Distribution of x_L

The normalised energy spectrum, in the laboratory frame, of the leading neutral pion of the first p-Air interaction, $\frac{d\mathcal{N}}{dx_L}$, along with the same distribution weighted by x_L, $x_L\frac{d\mathcal{N}}{dx_L}$, are represented in Figure 2.8 by the solid and dashed blue lines, respectively, for the EPOS-LHC model.

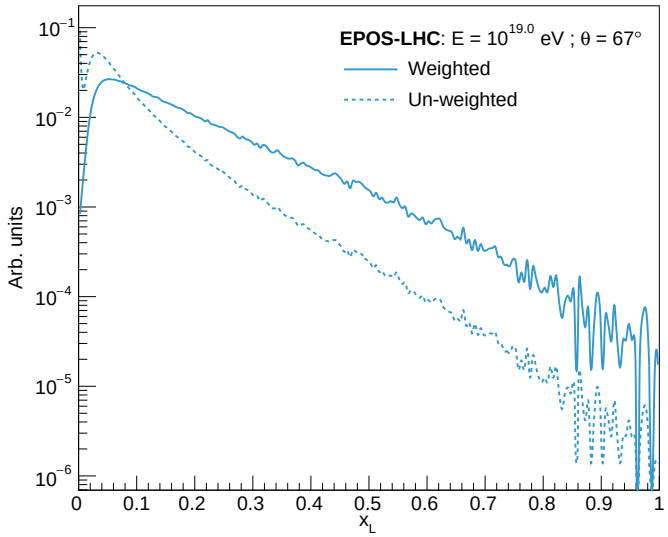

Figure 2.8: Probability density function of the fraction of energy, in the laboratory frame, carried by the leading neutral pion of the first p-Air interaction (solid blue line) along with same distribution weighted by x_L (dashed blue line). Proton induced showers were simulated with CONEX, for a primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$, using the hadronic interaction model EPOS-LHC.

The shape of the high tail of the energy spectrum of the leading neutral pion is not exponential all the way through as seen in Figure 2.8, so that fitting it to an exponential function would limit the fit range and compromise the quality of the fit. Therefore, we chose to fit the distribution of x_L weighted by x_L, $x_L\frac{d\mathcal{N}}{dx_L}$, since its high tail is visibly closer to a pure exponential function and equally correlated with the distribution of $\ln N_\mu$. This choice differs from the one taken in [34], although the qualitative results derived from weighted or un-weighted distributions of x_L are identical.

The high exponential tail of the x_L distribution weighted by x_L is visible for all hadronic interaction models, although its steepness and similarity with a perfect exponential function vary from model to model. Note that this feature is not surprising since the same is verified for the α_1 distribution, which is complementary, to first order, to the distribution of x_L. In Figure 2.9 the probability density function of the fraction of energy in the laboratory frame carried by the leading neutral pion of the first p-Air interaction weighted by that fraction is shown for all hadronic models.

The forward region of the distribution of x_L weighted by x_L was fitted to an exponential function of the form

$$x_L\frac{d\mathcal{N}}{dx_L} = \mathcal{C}_\pi \exp\left(\frac{x_L}{\Lambda_\pi}\right), \tag{2.17}$$

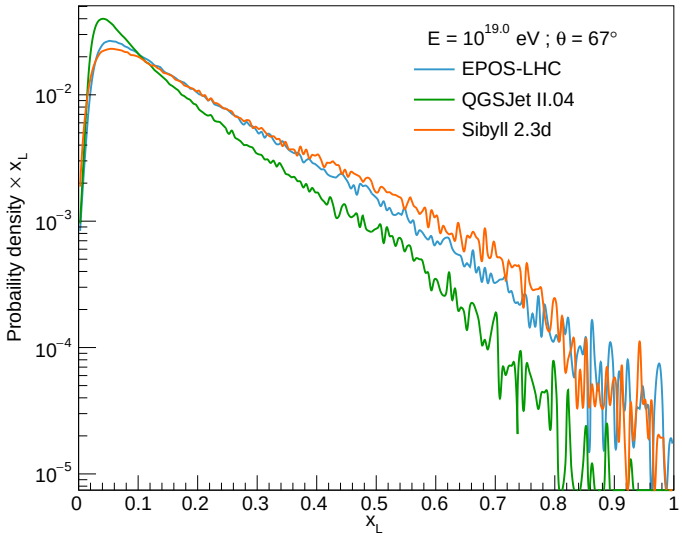

Figure 2.9: Probability density function of x_L weighted by x_L over an ensemble of $\sim 10^6$ of proton induced EAS with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$, simulated with CONEX, using the hadronic interaction models EPOS-LHC, QGSJET II-04 and SIBYLL 2.3d.

where $\mathcal{C}_\pi$ and Λ_π are free parameters. The latter is the exponential slope that characterises the forward region of the leading neutral pion's energy spectrum in the laboratory frame. We note once more that this Λ_π differs from the one used in [34], since there the used distribution of x_L is the pure energy spectrum, not the weighted one. We also recall that the criterion chosen to select the fitting range is the one described in Subsection 2.2.2: only 5% deviations from a pure exponential are allowed, starting a distance Δ^π_{max} above the maximum of the distribution. A fit to the forward region of the spectrum of the fraction of energy, in the laboratory frame, carried by the leading neutral pion weighted by this fraction of energy is shown in Figure 2.10 for the EPOS-LHC model. The plots made with the remaining hadronic models can be found in Appendix A.

The values of Δ^π_{max}, $\langle x_L \rangle$, σ_{x_L}, Λ_π and χ^2_{red} can be found in Table 2.3 for the considered hadronic interaction models. Note the the moments of x_L are taken with respect to the probability density function $x_L \frac{dN}{dx_L}$.

Parameters	EPOS-LHC	QGSJET II-04	SIBYLL 2.3d
Δ^π_{max}	0.15	0.15	0.15
$\langle x_L \rangle$	0.17	0.13	0.18
σ_{x_L}	0.14	0.12	0.15
Λ_π	-0.1524 ± 0.0015	-0.1250 ± 0.0012	-0.1714 ± 0.0018
χ^2_{red}	64.48/55	59.38/48	112.34/61

Table 2.3: Values of Δ^π_{max}, $\langle x_L \rangle$, σ_{x_L}, Λ_π and χ^2_{red} for the distribution of x_L weighted by x_L in proton induced EAS with $E_0 = 10^{19}$ eV and $\theta = 67°$ simulated with CONEX, for all hadronic interaction models.

It is apparent that the values of Λ_π for the different models are incompatible with each other. The determination of the physical mechanisms responsible for this difference must be studied elsewhere.

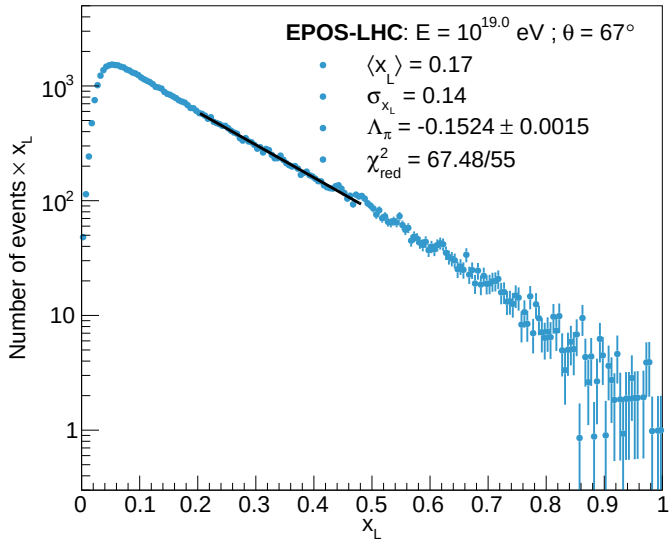

Figure 2.10: Fitted forward region of the distribution function of x_L weighted by x_L over an ensemble of $\sim 10^6$ of proton induced EAS with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$, simulated with CONEX, using EPOS-LHC.

Furthermore, the energy spectrum of the leading neutral pion in the laboratory frame is hardest for QGSJET II-04 and softest for the SIBYLL 2.3d model. We also note that the values of $\chi^2_{red} \sim 1$ confirm the quality of the fit from which Λ_π is extracted.

2.3.3 $\Lambda_\mu \leftrightarrow \Lambda_\pi$ conversion

Adapting the re-sampling procedure presented in Section 2.2.3, we perturbed the high tail of the distribution of x_L by selecting pairs of values $(x_L, \ln N_\mu)$ with a probability density function $p(x_L) \propto e^{(x_L^* - x_L)/\delta\Lambda_\pi}$, yielding a new distribution

$$\left(\frac{d\mathcal{N}}{dx_L}\right)_{new} = \begin{cases} \dfrac{d\mathcal{N}}{dx_L} & \text{for } x_L < x_L^* \\ \dfrac{d\mathcal{N}}{dx_L} \exp\left(\dfrac{x_L^* - x_L}{\delta\Lambda_\pi}\right) & \text{for } x_L > x_L^* \end{cases}, \tag{2.18}$$

where x_L^* is the critical point where the probability density function $x_L \frac{d\mathcal{N}}{dx_L}$ reaches its maximum. Thus, the slope of the high tail of the distribution of x_L is changed by $\Lambda_\pi^{-1} \to \Lambda_\pi^{-1} + \delta\Lambda_\pi^{-1}$. Via its correlation with x_L, the slope of the tail of the $\ln N_\mu$ distribution is also changed, with little effect on the rest of the distribution. As noted previously, the perturbations $\delta\Lambda_\pi$ are not an attempt to reproduce any specific physical process that could affect the forward region of the energy spectrum of the leading neutral pion, but rather a change in the frequency of events with high x_L, to study the way Λ_μ changes for each model.

The tails of the distributions arising from this re-sampling procedure were then fitted. An example is show in Figure 2.11 for a given perturbation $\delta\Lambda_\pi$ with the EPOS-LHC model.

As expected, the softer the energy spectrum of the leading neutral pion, the steeper the tail of the distribution of $\ln N_\mu$ as there is more energy available to produce muons, increasing the probability that

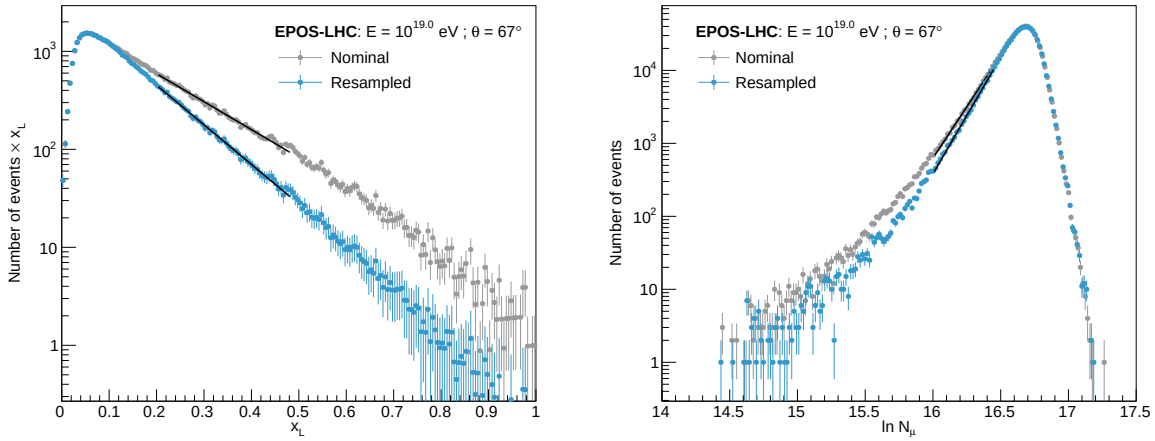

Figure 2.11: Left panel: Fitted nominal (grey) and re-sampled (blue) distributions of x_{L} weighted by x_{L}. Right panel: fitted nominal (grey) and re-sampled (blue) $\ln N_\mu$ distribution. Both distributions are over an ensemble of 10^6 proton induced EAS, simulated with CONEX with the EPOS-LHC hadronic interaction model.

a given shower has a larger muon content.

By varying $\delta\Lambda_\pi$, we obtain successive pairs of $(\Lambda_\mu, \Lambda_\pi)$ values, forming a calibration curve for each model. Once again, these curves allow the conversion of a measurement of Λ_μ into a value of Λ_π, and are represented by the solid lines shown in Figure 2.12 for the interaction models EPOS-LHC, QGSJET II-04 and SIBYLL 2.3d, where the filled circles indicate the nominal pair $(\Lambda_\mu, \Lambda_\pi)$. In the same figure, dashed lines are parametrised by direct perturbations to the low tail of $\ln N_\mu$ distribution, applied below its maximum value, $\ln N_\mu^*$. Once again, we re-sample the ensemble of events by selecting pairs $(x_{\mathrm{L}}, \ln N_\mu)$ with a probability density function $p(\ln N_\mu) \propto \mathrm{e}^{(\ln N_\mu - \ln N_\mu^*)/\delta\Lambda_\mu}$. The high tail of the weighted distribution of x_{L} is thus changed through its correlation with $\ln N_\mu$. The circumferences represent the nominal values of $(\Lambda_\mu, \Lambda_\alpha)$. As before, we refer to Appendix B for a comparison between SIBYLL versions.

From Figure 2.12 it is apparent that there is a monotonic relation between Λ_μ and Λ_π for all models. The curves parametrised by $\delta\Lambda_\pi$ have a remnant model dependence reaching $\delta_{\Lambda_\pi} \sim 28\%$. However, the curves parametrised by $\delta\Lambda_\mu$ have a much smaller model dependence, reaching a maximum deviation across models of $\delta_{\Lambda_\pi} \sim 10\%$. Mathematically, this result is explained by the combination of two effects: a direct perturbation $\delta\Lambda_\mu$ to the tail of the $\ln N_\mu$ distribution as more impact on the measured Λ_μ than applying the same perturbation to the weighted x_{L} distribution and measure its effect on the tail of the $\ln N_\mu$ distribution through the re-sampling procedure (and vice-versa); and directly perturbing Λ_π is less efficient in affecting Λ_μ than a direct perturbation to Λ_μ is in affecting Λ_π. As a result, the curves parametrised by $\delta\Lambda_\mu$ are obtained from the curves parametrised by $\delta\Lambda_\pi$ by a rotation about the nominal $(\Lambda_\mu, \Lambda_\pi)$ values followed by a stretch of the curves parametrised by $\delta\Lambda_\mu$, which brings the calibration curves for different models closer or drives them apart.

The above observations stem from the imperfect correlation between x_{L} and N_μ, meaning that values of x_{L} contributing to the forward tail of this distribution can correspond to values of N_μ outside the low tail region of this distribution, suppressing the effect of perturbations to the tail of a given distribution when

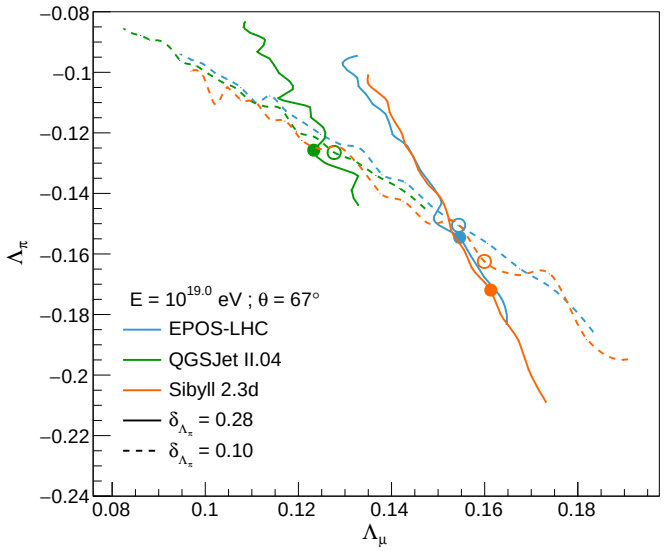

Figure 2.12: Conversion curves between Λ_μ and Λ_π, for the hadronic interaction models EPOS-LHC (blue), QGSJET II-04 (green) and SIBYLL 2.3d (orange). The solid lines show how Λ_μ changes if Λ_π is changed, while the dashed lines represent the converse situation. Solid dots and circumferences represent the nominal Λ_π and Λ_μ values.

the other is directly perturbed. Physically, this imperfect correlation has different origins, such as the subdominant but noticeable effect of deeper shower generations to the shower-to-shower fluctuations of N_μ, regardless of the x_L value of the first interaction. Moreover, the fact that we are only considering the leading neutral pion of the first interaction, rather than the full energy spectrum of these particles, means that the resulting muon content does not discriminate between an event where the leading pion takes most of the primary energy from an event where most of the primary energy is taken by more than one neutral pion with low x_L value. As mentioned before, diffractive events also play a role in breaking the correlation between the tail of N_μ and the forward region of the x_L distribution, since in these events $x_L \sim 0$ but $\ln N_\mu$ can still be small. Nevertheless, a thorougher study would have to be performed using the full energy spectrum of neutral pions. Such a study would likely imply running a full Monte Carlo simulation with a modified energy spectrum of neutral pions, rather than following an *ad hoc* re-sampling procedure. This will be addressed in a future work.

Causality dictates that the calibration curves used to convert a measurement of Λ_μ into a value of Λ_π are the ones parametrised by $\delta\Lambda_\pi$, i.e, the ones represented by the solid lines in Figure 2.12, for which the conversion is manifestly model dependent. However, one can still constrain the forward region of the energy spectrum of neutral pions, hence its production cross section, through measurements of Λ_μ, in ultra-high energy p-Air interactions. Thus, Λ_μ could be used to exclude the exotic physics phenomena affecting the production of neutral pions at the highest energies.

2.4 Measurement of Λ_μ

As previously stressed, X_{max} measurements are compatible with heavier primaries at the highest energies, for which the $\ln N_\mu$ distribution is an A-fold convolution of $\ln N_\mu$ distributions for proton. Moreover, the energy spectrum of cosmic rays is a power law in energy (see Equation 1.1) and the muon content of EAS is dependent on the primary energy 1.7. Thus, for an ensemble of events with primary energy within a small bin $\mathrm{d}E$, those with less energy are more abundant and have, on average, a lower muon content, yielding a distorted distribution of $\ln N_\mu$ when compared to the one obtained in the previous sections. Finally, the measured $\ln N_\mu$ distribution is convoluted with the finite resolution of the detectors and the uncertainty in the determination of the primary energy and the shower's muon content. As such, we are compelled to investigate under which conditions a measurement of Λ_μ can be performed with enough precision to distinguish between hadronic models.

2.4.1 Simple model of a mixed composition scenario

To illustrate the impact of mixing proton with a heavier primary on the distribution of the number of muons in EAS reaching the ground level, along with the effect of considering the cosmic ray flux dependence on the primary energy (see Equation (1.1)), we considered a 1:1 mixture of proton and Helium. About $\sim 700\,000$ proton initiated showers along with $\sim 100\,000$ helium initiated ones were simulated with CONEX using the post-LHC hadronic interaction model EPOS-LHC, for a primary energy of $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Once again, the number of muons is recovered at 1 400 m above sea level. Note that, at this stage, we are not considering experimental limitations nor more complex, extreme, mixed composition scenarios.

We emulate the mixed composition scenario by choosing a primary, either proton or Helium, with equal probability and a primary energy $E_0 \in [10^{18.9}, 10^{19.1}]$ eV following a probability density function $\propto E_0^{-2.7}$. We then select a given value of N_μ, at random and following the distribution of $\ln N_\mu$ corresponding to either proton or helium at $E_0 = 10^{19}$ eV. From Equation (1.7), the selected N_μ value is corrected for the new primary energy as

$$N_\mu(E_0) = N_\mu(10^{19}\ \text{eV}) \left(\frac{E_0}{10^{19}\ \text{eV}} \right)^{\beta}, \tag{2.19}$$

where E_0 is given in eV and $\beta = 0.93$. Note that this value of β is an average over the primaries from proton to iron, and it is approximately independent of the hadronic interaction model. Repeating the above procedure iteratively 10^6 times, yields an ensemble of EAS over which we build the distribution of $\ln N_\mu$. Figure 2.13 shows the distributions of $\ln N_\mu$ for pure proton and He with $E_0 = 10^{19}$ eV, and for a 1:1 mixture of proton and helium with primary energy $E_0 \in [10^{18.9}, 10^{19.1}]$ eV.

As mentioned earlier, the distribution of the number of muons at the ground level for helium initiated showers is narrower and closer to a Gaussian distribution. According to Equation (1.8), its average value is larger than that of the $\ln N_\mu$ for a proton initiated shower by

$$\langle N_\mu \rangle_{\text{He}} = 4^{1-\beta} \langle N_\mu \rangle_{\text{p}}, \tag{2.20}$$

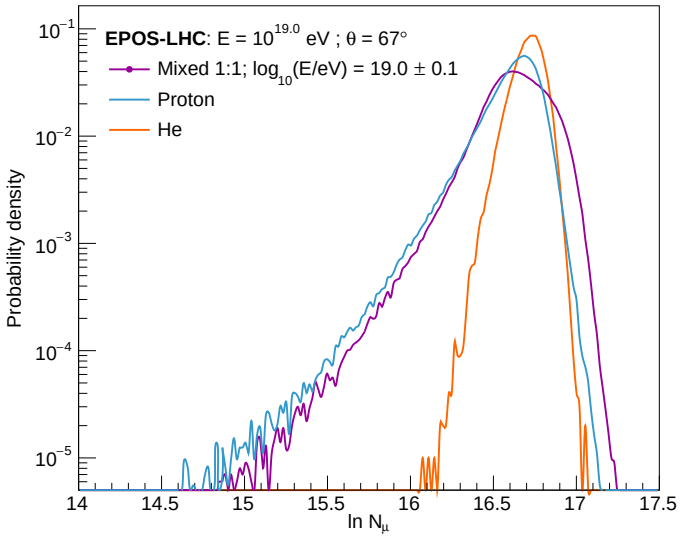

Figure 2.13: Probability density functions of $\ln N_\mu$ over ensembles of EAS with $\theta = 67°$ for pure proton (blue) and pure helium (orange) with $E_0 = 10^{19}$ eV, and for a mixed 1:1 composition of p:He (purple) with $\log_{10}(E_0/\text{eV}) = 19.0 \pm 0.1$. The showers were simulated with CONEX, for the post-LHC hadronic interaction model EPOS-LHC.

where 4 is the mass number of Helium, and the subscripts p and He refer to proton and helium induced EAS, respectively. The maximum of the distribution of $\ln N_\mu$ for helium is shifted to the right, with respect to the corresponding distribution for proton. The tail of the former distribution is much steeper and hidden underneath the tail of the $\ln N_\mu$ distribution for pure proton, mainly affecting the shape of the upper region of the low tail of the distribution for a mixed composition. Additionally, the latter distribution is wider than that of a proton induced shower, and its shape near the peak is distorted by the power law dependence of the primary energy spectrum. This distortion is a combination of two effects: on the one hand, the dependence of N_μ on the primary energy, through Equation (1.7), means that, when E_0 increases, the $\ln N_\mu$ distribution shifts to the right; while, on the other, the number of events with higher energy is suppressed by $E_0^{-\gamma}$, according to Equation 1.1. Therefore, we can envision the distribution of $\ln N_\mu$ for pure proton with a smearing in the primary energy as the superposition of increasingly suppressed $\ln N_\mu$ distributions ordered by each fixed energy. As a result, and adding the helium contamination, the shape of tail of the $\ln N_\mu$ distribution for the mixed composition scenario is visibly different from that of the corresponding distribution for proton.

To extract Λ_μ under these conditions, we must first account for the smearing in energy, by plotting the distribution of $\ln\left(N_\mu/E_0^\beta\right)$ instead of $\ln N_\mu$, since the former does not depend on the primary energy via Equation (1.7).

Figure 2.14 shows the distribution of $\ln\left(N_\mu/E_0^\beta\right)$ for proton and helium induced showers with $E_0 = 10^{19}$ eV and for a p:He 1:1 mixture with $E_0 \in [10^{18.9}, 10^{19.1}]$ eV, for the EPOS-LHC model. Additionally, a grey band showing the fitting range is displayed. The criterion to determine the fit range is the same as the one presented in Section 2.2.2. The upper limit of the fit region is at a distance $\Delta^\mu_{\text{max}} = 0.25$ from the

maximum of the muon distribution for the proton induced shower.

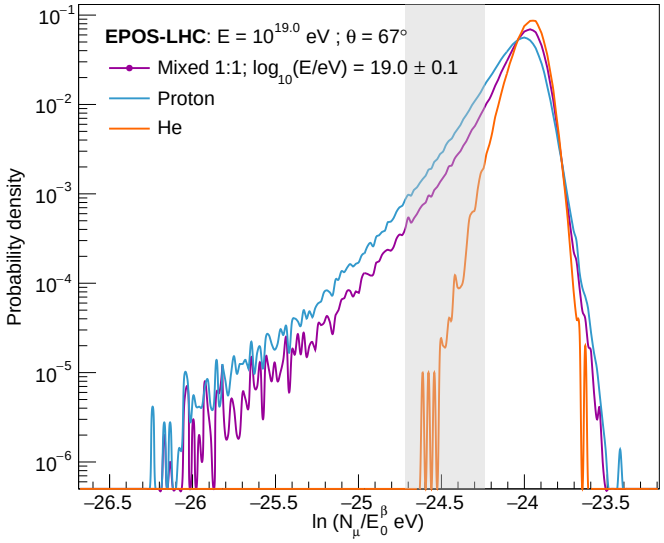

Figure 2.14: Probability density functions of $\ln\left(N_\mu/E_0^\beta\right)$ over ensembles of EAS with $\theta = 67^\circ$ for pure proton (blue) and helium (orange) with $E_0 = 10^{19}$ eV, and for a mixed 1:1 composition of p:He (purple) with $E_0 \in [10^{18.9}, 10^{19.1}]$ eV. The showers were simulated with CONEX, for the post-LHC hadronic interaction model EPOS-LHC.

Shifting each entry of the $\ln N_\mu$ distributions by $-\beta \ln E_0$ successfully mitigated the distortion induced by the power law dependence of N_μ and the comic ray energy spectrum on E_0. What is left, is a distribution with a tail similar to that of the proton distribution, whose peak is enhanced by the helium contribution. As expected, the rapid fall of the low tail of the $\ln N_\mu$ distribution for helium, preserves the features encountered in the corresponding distribution for proton showers, allowing the extraction of Λ_μ, with a given accuracy.

To quantify the relative accuracy of the measurement of Λ_μ for a 1:1 mixture of p:He, we define δ_μ:

$$\delta_\mu = 1 - \frac{\Lambda_\mu^{\text{mixed}}}{\Lambda_\mu^{\text{p}}}, \tag{2.21}$$

where Λ_μ^{p} is the slope of the tail of the $\ln N_\mu$ distribution for proton initiated EAS and $\Lambda_\mu^{\text{mixed}}$ is the slope of the tail of the corresponding distribution for the mixed composition scenario. For the chosen fit region we get $\delta_\mu \simeq 3\%$, which is enough to distinguish between hadronic interaction models [35]. Thus, conveniently choosing the fitting range and plotting $\ln\left(N_\mu/E_0^\beta\right)$ instead of $\ln N_\mu$ allows the extraction of Λ_μ with high accuracy, in the scenario considered in this section.

2.4.2 Realistic mixed composition scenarios and experimental uncertainty

We must now consider the effect of a Gaussian detector resolution in the determination of the muon content of EAS along with a more realistic mixed composition scenario. This was achieved by simulating an ensemble of EAS with primary energy $E_0 = 10^{18.7}$ eV and zenith angle $\theta = 67^\circ$, using different

primaries, with CONEX and the hadronic interaction model SIBYLL 2.3c. Note that we have not used SIBYLL 2.3d since this results are taken directly from [34]. To emulate a mixed primary composition, the different primaries, namely: proton, helium, nitrogen, iron and photons were combined in different proportions. To emulate the experimental resolution, a Gaussian smearing in N_μ of 20% (still larger than the current resolution attained by the Pierre Auger Observatory of 17% [36]) was applied. The resulting muon distribution for the composition p:He:N:Fe in proportion 2:1:1:0, henceforth designated by total distribution, along with a photon contamination of 0.5 % [37] is plotted in Figure 2.15. Also displayed in the same figure is a shaded area corresponding to the fit region.

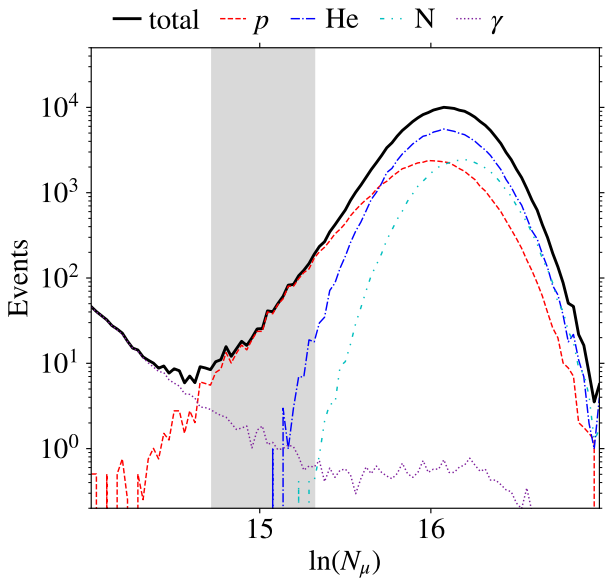

Figure 2.15: Distribution of $\ln N_\mu$ over an ensemble of EAS initiated by a mixture of p:He:N:Fe in proportion 2:1:1:0 with primary energy $E_0 = 10^{18.7}$ eV and zenith angle $\theta = 67°$, simulated with CONEX, using SIBYLL 2.3c. The distributions corresponding to proton, helium and nitrogen are shown in red, blue and cyan, respectively, along with a photon contamination shown in purple. The total distribution is shown in black. Additionally, a Gaussian smearing of 20% was applied to each distribution. Plot taken from [34].

From Figure 2.15 it is clear that the photon contamination affects only the low end of the tail of the $\ln N_\mu$ distribution, while the high tail region is mostly affected by the helium contribution, as the fluctuations of the muon content of heavier primary are suppressed by $1/\sqrt{A}$, where A is the mass number. Choosing the upper limit of the fit region below the helium contribution and the lower limit to avoid the photon contamination, allows for the extraction of Λ_μ, since the total and proton $\ln N_\mu$ coincide in that region. It is also worth noting that the increased width of all distributions stems from their convolution with a Gaussian resolution.

To quantify the bias and relative accuracy of the measurement of Λ_μ for the total distribution with respect to the proton distribution, the value of δ_μ, as defined in Equation (2.21), was computed for different mixed composition scenarios and different hadronic interaction models. The minimal number of events for each mass composition required to achieve a relative accuracy in the measurement of Λ_μ of 20% was determined. This threshold was chosen as it is the minimal δ_μ value for which the different

hadronic interaction models can be distinguished [38], provided the fit region coincides with the grey band in Figure 2.15. The results of this study are displayed in Table 2.4. Note that SIBYLL 2.3c was used since we have not repeated this study since the release of [34].

Scenario	Model	$\mathcal{N}_{\text{tail}}$	$\mathcal{N}_{\text{total}}$
1:1:1:1	QGSJET II-04	21	1 564
	EPOS-LHC	29	1 926
	SIBYLL 2.3c	30	1 667
1:2:1:0	QGSJET II-04	32	7 056
	EPOS-LHC	36	5 505
	SIBYLL 2.3c	33	4 411
1:6:2:0	QGSJET II-04	205	385 776
	EPOS-LHC	132	136 212
	SIBYLL 2.3c	123	78 482

Table 2.4: Number of events within the fit range, $\mathcal{N}_{\text{tail}}$, and in the total distribution of $\ln N_\mu$, $\mathcal{N}_{\text{total}}$, required to measure Λ_μ with a precision of $\delta_\mu = 0.2$, for distinct composition scenarios of the form p:He:N:Fe for the hadronic interaction models EPOS-LHC, QGSJET II-04 and SIBYLL 2.3c. Adapted from [34]

Regardless of the composition scenario, we verified that the bias in the measurement of Λ_μ was negligible [35], and even in extreme mass composition scenarios, it is still possible to measure Λ_μ with the required accuracy provided the number of events is large enough and that Λ_μ is extracted from the distribution of $\ln\left(N_\mu/E_0^\beta\right)$ rather than that of $\ln N_\mu$, as inferred in the previous section, with the assumption that the experimental resolution is Gaussian, not hiding completely the tail of the total muon distribution.

For the composition compatible with X_{max} measurements, the precise determination of Λ_μ requires about 5 000 events. For the energy range where the detectors of the regular array of the Pierre Auger Observatory have full efficiency, i.e., above $10^{18.5}$ eV, this is still a difficult number to reach, given the low UHECR flux and the exposure of surface array of the Observatory. Moreover, stations of the regular array measure the total time trace produced by the muonic and electromagnetic components, making it difficult to disentangle these components and achieve an accurate and precise measurement of N_μ, especially in vertical showers.

Besides the AugerPrime upgrades, focused on measuring the mentioned components individually, there are proof-of-concept detectors, such as MARTA which ought to do the same. Moreover, the latter engineering array is set to be placed underneath stations of the Infill array of the Pierre Auger Observatory, which are more closely spaced and equipped with a new set of triggers, allowing the measurement of showers at $E_0 = 3 \times 10^{17}$ eV, with 100% efficiency. In this region, the cosmic ray flux is about 10^3 larger than that at the threshold energy of the regular array. Therefore, a measurement of Λ_μ with enough precision to distinguish between hadronic interaction models, could, in principle, be performed at this lower energies. Furthermore, the forward region of the energy spectrum of neutral pions is already being measured in proton-lead collisions at the LHCf [39] at center-of-mass of mass energies equivalent to cosmic rays with primary energy of 10^{17} eV, allowing for important cross-checks.

From all of the above, the MARTA engineering array appears to be a suitable candidate to reconstruct

the muon number distribution in proton initiated Air showers and allow the extraction of Λ_μ. We explore this possibility in the next chapter.

Chapter 3

Reconstruction of the muon number distribution in proton induced EAS with the MARTA engineering array

In the previous chapter we have concluded that the slope of the muon number distribution can be used to constrain the production cross-section of neutral pions and the energy spectrum of hadronically interacting particles in ultra-high energy p-Air interactions, with a residual model dependence of $\delta_{\Lambda_\pi} \sim 28\%$ and $\delta_{\Lambda_\alpha} \sim 7\%$, respectively. Furthermore, we showed that this measurement could be achieved with enough precision to distinguish between hadronic interaction models in mixed composition scenarios and within current experimental uncertainties provided the number of events exceeds $\mathcal{N} \sim 5\,000$. In this chapter we use the information retrieved by MARTA stations, i.e., stations of the Infill array of the Pierre Auger Observatory with resistive plate chambers placed underneath the water Čerenkov detectors, to reconstruct the muon number distribution over an ensemble of proton induced vertical showers and determine to which precision the measurement of Λ_μ can be made. The bias in the reconstructed estimators of the average muon content and its fluctuations are also determined, as well as the experimental resolution of the measurement.

The structure of this chapter is as follows: in Section 3.1 we show that the conversions from Λ_μ to Λ_α and from Λ_μ to Λ_π can be performed at lower energies and zenith angles, namely $E_0 = 10^{17.5}$ eV and $\theta = 30^\circ$; in Section 3.2 we describe the $\overline{\text{Off}}\underline{\text{line}}$ software and discuss the station and event triggers; in Section 3.3 we describe the reconstruction of the number of muons in EAS from the normalisation of the lateral distribution function of the number of active fiducial pads of the RPCs, present the reconstructed distribution of the number of muons at the ground level, determine the reconstruction resolution and compute the bias in the reconstructed features of the distribution, namely Λ_μ^{rec}, and its first and second moments. Moreover, we repeat the reconstruction procedure at higher energies and zenith angles, namely $E_0 = 10^{18.5}$ eV and $\theta = 40^\circ$, separating the case where the MARTA engineering array covers the entire SD-750 array and just 7 stations in an hexagonal arrangement (1 in the centre + the first 6

neighbours).

3.1 Constraining the energy spectrum of neutral pions and hadrons in proton induced vertical showers at $10^{17.5}$ eV

In Chapter 2 we have argued that inclined showers were preferred to better measure their muon content, since the electromagnetic shower was almost completely absorbed in the atmosphere, yielding a total signal closer to the muonic signal measured by WCDs. However, to reconstruct the muon number distribution using the MARTA engineering array we have chosen to use showers with zenith angle $\theta = 30°$. The main reason for this choice is that RPCs of the engineering array are horizontal, so that their effective detection area, the area perpendicular to the shower axis, is suppressed by $\cos\theta$ for showers arriving with zenith angle θ, increasing particle losses with respect to vertical showers. Additionally, the reconstruction of inclined showers is more involved than that of vertical showers, $\theta < 60°$, in part because muons and anti-muons are deflected in opposite directions by the Earth's magnetic field, yielding a dipolar shower footprint, that breaks the rotational invariance and distorts the shape of the lateral distribution function, LDF, on which the reconstruction procedure is based. Considerations of this and other effects, would introduce unnecessary difficulties into our analysis. Lastly, the Infill array has just 61 stations spreading over an hexagonal area of $A = 23.5$ km^2, so that highly inclined showers could saturate the entire array, especially given the tight restrictions imposed by physics and quality triggers of the Observatory. Although this may not be a problem by lowering the shower energy, it reinforces the aforementioned arguments.

Moreover, we have decided to use showers initiated by protons with $E_0 = 10^{17.5}$ eV, rather than at $E_0 = 10^{19}$ eV. As mentioned in Chapter 2, this lower energy is comparable to the equivalent center-of-mass energy in $p-p$ collisions at the LHC. In fact, letting $m_p \simeq 938$ MeV$/c^2$ [8] be the proton mass, and E_0 be its energy in the laboratory frame, we have an equivalent center-of-mass energy, $\sqrt{s}$, of

$$\sqrt{s} = \sqrt{2m_p(m_p + E_0)} = \sqrt{2 \times 938 \times (938 + 3 \times 10^{11})} = 23.7 \times 10^6 \text{ MeV} = 23.7 \text{ TeV}. \tag{3.1}$$

This enables the comparison of measurements of Λ_μ with measurements of the forward energy spectrum of neutral pions performed at the LHCf, in $p-p$ and $p-$lead collisions [39].

Furthermore, MARTA stations are triggered by surface stations of the Infill array, which are fully efficient at primary energies of $E_0 = 3 \times 10^{17}$ eV, making the proposed measurement feasible in the desired energy range. Lastly, lowering the energy by a factor of $10^{1.5}$, results in a $\sim 10^4$-fold increase in the cosmic ray flux, according to Equation (1.1). Thus, the minimum number of events necessary to distinguish between hadronic models, $\mathcal{N} \sim 5\,000$, would be gathered faster.

In fact, we can estimate the number of events measured by a unitary hexagonal array of MARTA stations considering the energy bin $\log_{10}(E_0/\text{eV}) \in [17.4, 17.6]$ and $\theta \in [25°, 35°]$. The infinitesimal

exposure, $\mathrm{d}\Sigma$, of a horizontal unitary cell with area A_{cell} reads

$$\mathrm{d}\Sigma = \mathrm{d}A_{\perp}\,\mathrm{d}\Omega\,\mathrm{d}t = \cos\theta\,\mathrm{d}A\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\phi\,\mathrm{d}t\,, \tag{3.2}$$

where ϕ is the shower's azimuth, θ the zenith angle and t the observation time. Integrating over ϕ, a time interval Δt and from $\theta_1 = 25^\circ$ to $\theta_2 = 35^\circ$ gives

$$\Sigma = 2\pi A_{\text{cell}}\Delta t\int_{\theta_1}^{\theta_2}\cos\theta\sin\theta\,\mathrm{d}\theta = \frac{\pi}{2}A_{\text{cell}}\Delta t\left(\cos(2\theta_1) - \cos(2\theta_2)\right), \tag{3.3}$$

Given an hexagonal lattice with spacing d, the unitary cell has an area of

$$A_{\text{cell}} = \frac{\sqrt{3}}{2}d^2. \tag{3.4}$$

Thus, the number of events measured by the cell under the aforementioned conditions and considering Equation (1.2) is thus given by

$$\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}\Sigma\,\mathrm{d}E_0} = J(E_{\text{ref}})\left(\frac{E_0}{E_{\text{ref}}}\right)^{-\gamma} \implies \frac{\mathcal{N}}{\Delta t} = \frac{\sqrt{3}\pi d^2}{4}\left(\cos(2\theta_1) - \cos(2\theta_2)\right)\frac{J(E_{\text{ref}})E_{\text{ref}}^{\gamma}}{1-\gamma}\left(E_2^{1-\gamma} - E_1^{1-\gamma}\right), \tag{3.5}$$

which corresponds for the Infill array, $d = 0.75\,\text{km}$, to

$$\mathcal{N} \simeq 75 \text{ events per month}, \tag{3.6}$$

for $E_0 \in [10^{17.4}, 10^{17.6}]$ eV, spectral index $\gamma = 3.2$ and taking $J(E_{\text{ref}})$ from Figure 1.1. The expected number of events for the entire array scales with the number of unitary cells.

We must now verify if there is still a monotonic relation between Λ_μ and Λ_α, and between the former and Λ_π, for proton induced showers at $E_0 = 10^{17.5}$ eV and zenith angle $\theta = 30^\circ$, since in Chapter 2 we used inclined showers with $E_0 = 10^{19}$ eV and $\theta = 67^\circ$. To do so, about $\sim 100\,000$ proton initiated showers with energy $E_0 = 10^{17.5}$ eV and zenith angle $\theta = 30^\circ$ were simulated with CONEX, for the post-LHC hadronic interaction model EPOS-LHC. Once again, the number of muons, N_μ, is measured 1 400 m above sea level. For this ensemble of showers, the distributions of α_1, x_{L} weighted by x_{L} and $\ln N_\mu$ were built and re-sampled, according to the procedure described in Section 2.2.3 to construct conversion curves $\Lambda_\mu \to \Lambda_\alpha$ and $\Lambda_\mu \to \Lambda_\pi$.

The left panel of Figure 3.1 shows the calibration curves between Λ_μ and Λ_α, while the right panel shows curves for the conversion $\Lambda_\mu \to \Lambda_\pi$, for the two mentioned energies and zenith angles, for the EPOS-LHC model. The solid lines correspond to direct perturbations to the variable plotted in the y axis, while dashed lines correspond to direct perturbations to the variable plotted on the x axis. The solid dots and circumferences correspond to the nominal slope values.

Figure 3.1 shows that there is still a monotonic relation between measurements of Λ_μ and values $\Lambda_\alpha/\Lambda_\pi$, although the conversion as proven to be dependent on the primary energy and/or the zenith angle. In particular, the nominal value of Λ_α is slightly larger for showers with $E_0 = 10^{19}$ eV showing

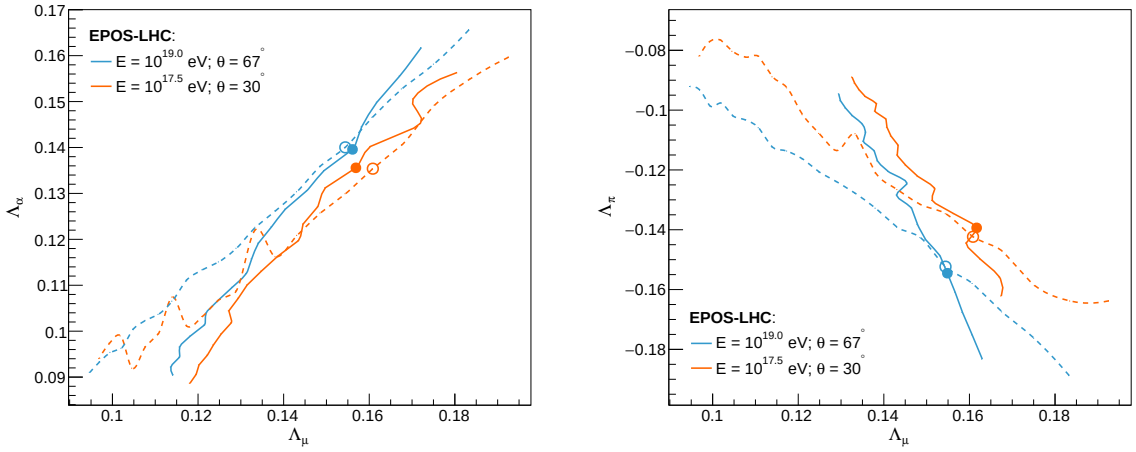

Figure 3.1: Calibration curves between Λ_μ and Λ_α (left panel) and between Λ_μ and Λ_π (right panel). Solid lines correspond to direct perturbations to the variable plotted on the y axis, while dashed lines correspond to direct perturbations to the variable plotted on the x axis. Curves for showers with $E_0 = 10^{19}$ eV and $\theta = 67°$ are shown in blue, while those for $E_0 = 10^{17.5}$ eV and $\theta = 30°$ are shown is orange. Both ensembles of showers were simulated with CONEX, for EPOS-LHC.

that the low tail of the energy spectrum of hadronically interacting particles is flatter at higher primary energies. In turn, the energy spectrum of neutral pions is harder for larger primary energies as indicated by the larger $|\Lambda_\pi|$ nominal value. Paradoxically, the value of Λ_μ is smaller for $E_0 = 10^{19}$ eV. This could mean that fluctuations of deeper shower generations could impact differently on the shower-to-shower fluctuations of the muon content, according to the primary energy and/or zenith angle. However, a detailed study on the dependence of the connection between Λ_μ and $\Lambda_\alpha/\Lambda_\pi$ on E_0 and/or θ, along with the mechanisms responsible for this dependence, is out of the scope of this work and must be done elsewhere.

In any case, we have proved that measurements of Λ_μ could still be used to probe the energy spectrum of hadronically interacting particles of the first p-Air interaction, along with the forward region of the production cross section of fast neutral pions arising from this interaction, at $E_0 = 10^{17.5}$ eV and $\theta = 30°$.

3.2 Auger events and SD trigger chain

To implement the reconstruction of the distribution of the number of muons based on the information retrieved by the RPCs, it is necessary to simulate the Auger event corresponding to each CORSIKA simulation, as well as the response of the RPCs to each event. That can be achieved by using the Offline framework which we describe below. Furthermore, since the RPCs are triggered by the surface detector, it is worth explaining in detail the trigger chain of the SD array of the Pierre Auger Observatory, which ultimately discerns background from high quality physics events.

3.2.1 The Offline framework

The Offline [40] is a general purpose framework developed by the Pierre Auger Collaboration which provides the infrastructure to support the computational tasks necessary to reconstruct and analyse shower events by emulating the response of the hybrid detection system of the Pierre Auger Observatory. The software is based on the premise that most tasks employed in the reconstruction of EAS can be factorised into sequences of self-contained processing steps, or modules, which can be implemented and sequenced by collaborators, to perform the desired data analysis. Hence, the framework must be robust, yet flexible, and extensible, being able to accommodate upgrades to the observatory instrumentation, without compromising its overall structure. These upgrades include AMIGA, Radio and MARTA. The Offline framework can handle various formats of input and output data, being able to read CORSIKA and CONEX files, and includes external packages like ROOT [41], for event serialisation, and GEANT4 [42], a toolkit designed to simulate the propagation of particles through matter. In fact, the latter package simulates the response of the water Čerenkov detectors and resistive plate chambers to the passage of secondary shower particles.

The Offline is organised into three main parts, which separate data from the algorithms used to treat them. They are:

- *processing modules:* factorised sequences of self-contained processing steps employed in simulation and reconstruction tasks. These algorithms are written in `C++`, taking advantage of its class oriented design, and registered with the Offline framework via a macro. In each module, the user can read information about the event and detector description, treat this information, and write the results back into the event or save them in a variety of formats. Instructions to assemble and sequence the modules are provided by `XML` files. The modular design of the framework allows collaborators to exchange code easily, improve upon existing analysis and easily pipeline simulation and reconstruction chains. Each module has a `Run()` method which is called once per event, along with `Init()` and `Finish()` methods, called only one time per job. The *Run Controller* invokes the algorithms written in each module, according to the instructions contained in the `XML` files. The *central configurator* points modules to the location of their configuration data files and creates parsers to assist in reading information at these locations, as specified by the *bootstrap* file.

- the *event data model*: the backbone of the framework, through which modules communicate. It contains the raw, calibrated, reconstructed and Monte Carlo information about the events. This information can be accessed by calling the required methods defined in suited classes of the code, which, in turn, contain other methods to verify the existence of the desired data. Classes are defined in different namespaces, according to the type of detector to which they correspond. Within the same namespace, classes follow the hierarchy normally associated with the observatory instruments.

- the *detector description*: a unified interface that contains slow changing information about the detector geometry, calibration constants and atmospheric conditions.

Furthermore, the Offline is built on a collection of utilities such as an `XML` parser, an error logger and various mathematical and physics services such as a geometry package capable of handling abstract manipulations of vectors. Further information about the Offline framework can be found in [28, 40].

3.2.2 Trigger of the surface detector of the Pierre Auger Observatory

As mentioned in Chapter 1, RPCs are triggered by water Čerenkov detectors. Thus, MARTA stations obey the strict hierarchical trigger chains built for the surface detector array. This chain is thoroughly described in [28, 43], while information about the new set of triggers can be found in [44].

The data acquisition, DAQ, trigger of the SD must fulfil physical and technical requirements. On the one hand, the rate of recordable events sustainable by the wireless communication between stations and the central campus is less than 1 event per hour, so that the system of triggers must reduce the counting rate of atmospheric muons of $\sim 3\,\mathrm{kHz}$. On the other hand, the trigger chain must maximise the storage of candidate physics events, while allowing data acquisition at the lowest possible energy. These requirements are met by introducing a hierarchical trigger chain such that the criteria to discriminate against background become increasingly stricter at each level. Furthermore, the ultimate discrimination between physics and chance events is performed off-line through a selection of participating detectors.

A few definitions are needed before describing the criteria applied at each trigger level. In each surface station, the signals captured by the PMT tubes are digitised by 40 MHz 10-bit Flash Analog to Digital Converters (FADCs) in bins of 25 ns, and sent to the central data acquisition system (CDAS) provided the event triggers the surface array. The calibration of these signals is performed *in situ* and automatically, using as a reference the average charge collected by a PMT from the Čerenkov light produced by a vertical and central through-going muon, denoted Q_{VEM} or VEM. In reality, atmospheric muons cross the WCD from all directions, producing a peak in the charge distribution, $Q^{\mathrm{peak}}_{\mathrm{VEM}}$, as well as a peak in the current distribution, $I^{\mathrm{peak}}_{\mathrm{VEM}}$, both proportional to Q_{VEM}. The signals recorded by the SDs are obtained by integrating the counts of the FADC bins of the time traces. Signals are then converted into units of Q_{VEM} using the calibration provided by atmospheric muons.

The trigger hierarchy of the SD array starts at station level, followed by array, physics and quality triggers, in this order. The first two levels of trigger, T1 and T2, are formed at each station. Four independent trigger modes are implemented as T1 to measure, in a complementary way, the electromagnetic and muonic components of the shower:

- *Trigger Threshold* (TH-T1): requires the coincidence of three PMTs each above $1.75\,I^{\mathrm{peak}}_{\mathrm{VEM}}$. This trigger selects large signals, regardless of their time signature, enabling the reduction of the rate due to atmospheric muons from 3 kHz to 100 Hz. It is particularly effective in detecting the muonic shower component;

- *Time-over-Threshold* (ToT): requires 13 bins (≥ 325 ns) in a sliding window of 120 FADC bins (3 µs) to be above $0.2\,I^{\mathrm{peak}}_{\mathrm{VEM}}$ in coincidence of 2 out of 3 PMTs. This trigger is effective in the detection of particles in vertical showers and away from the shower core [a], where their arrival is spread in time.

[a] Intersection point between the shower axis and the array plane

Furthermore, it discards efficiently the background signal produced by atmospheric muons, since their duration is only about 6 FADC bins, bringing the event rate down to 2 Hz;

- *Time-over-Threshold-deconvoluted* (ToTd): deconvolves the exponential tail of the reflected Čerenkov light inside the tank, with decay constant $\tau = 67$ ns, before applying the ToT condition. This trigger reduces the influence of muons, whose signals have a fast rise time and a typical decay constant of ~ 60 ns over a total duration of one or two FADC time bins. Starting with a trace $\{S_i\}$, where i indexes the FADC bin, the deconvolution algorithm reads

$$d_i = \frac{S_i - S_{i-1}e^{-\Delta t/\tau}}{1 - e^{-\Delta t/\tau}}, \tag{3.7}$$

  where $\{d_i\}$ is the deconvoluted trace and $\Delta t = 25$ ns is the FADC bin width.

- *Multiplicity-of-positive-steps* (MoPS): counts the number of instances in a sliding window of 120 FADC bins over which there is a monotonic increase of the signal amplitude. Each set of increasing signals is a positive step and for each one, the vertical increase, j, must be above the typical noise and bellow the characteristic amplitude of a vertical muon: $3 < j \leq 31$. If more than four positive steps fall within this range, the trigger condition is satisfied. This trigger is designed to select small, irregular signals, produced by many electromagnetic particles over a period of time longer than that of a typical muon pulse.

The last two trigger modes are specifically designed to be insensitive the atmospheric muons, improving the acceptance of small electromagnetic signals. Together, they reduce the trigger rate to about 0.3 Hz and accept signals as small as 1 VEM. Since their installation in 2013, they have been providing constant trigger rates, which extended the energy threshold of the SD-750 array down to $E_0 \sim 3 \times 10^{17}$ eV.

The T2 trigger is stricter than T1, and applied to the station controller to reduce to about 20 Hz the rate of events per detector. It also has 4 modes:

- *Threshold trigger* (TH-T2): it accepts TH-T1 triggers provided they exceed a higher signal threshold of $3.2\, I_{\text{VEM}}^{\text{peak}}$ in coincidence among three PMTs.

- ToT-T1, ToTd-T1 and MoPS-T1 are directly promoted to the T2 level.

Stations which fulfil the T2 trigger conditions are sent to CDAS, where the array trigger is formed. Figure 3.2 shows the logical scheme of the hierarchical trigger chain of the Auger surface detector.

The trigger of the array is known as T3. It is formed at the CDAS and combines the temporal and spatial information of T2 stations to look for correlations. Once it is formed, it passes stations which fulfil the T2 criteria to the CDAS along with T1 stations that are within 30 µs of the T3. The array trigger has two modes:

- *3-fold ToT2C$_1$&3C$_2$*: it requires the coincidence of at least three ToT, ToTd or MoPS detectors that meet a minimal compactness condition. Considering the hottest station[b] to be the center of

[b] Station with the largest signal

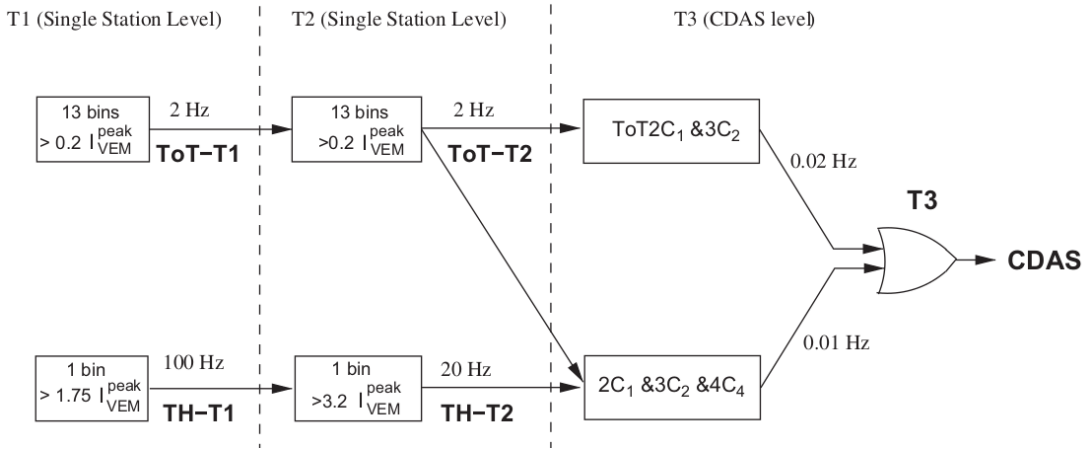

Figure 3.2: Logic scheme of the hierarchical trigger chain of the Auger surface detector, taken from [43].

successively larger concentric hexagons, C_n, the compactness condition is such that the other two stations lie in C_1, the closest ring of neighbouring stations, and in C_2, the second closest. Additionally, a temporal restriction is applied: each ToT must be within $(6 + 5C_n)$ µs of the first one. This trigger selects about 1 600 events per day, most of which are physical. A possible configuration is shown on the left panel of Figure 3.3.

- *4-fold $2C_1$&$3C_2$&$4C_4$*: it requires the coincidence of any four T2 stations, where one is the hottest one, the second is in C_1, the third in C_2 and the fourth one can be as far as in the C_4 ring, provided each station is within a $(6 + 5C_n)$ µs time window of the first one. This trigger is efficient in the detection of inclined showers. This trigger is more permissive, selecting about 1 200 events per day, of which 10% are real showers. A possible configuration is shown on the right panel of Figure 3.3.

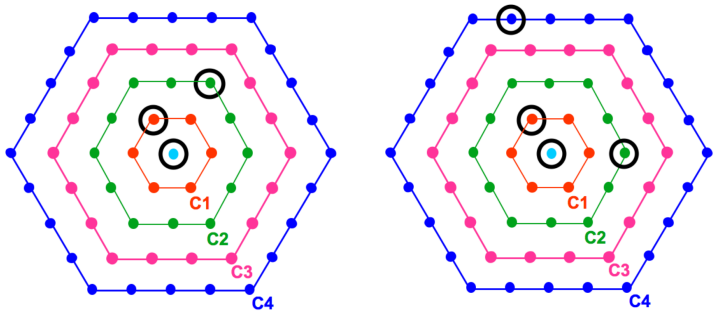

Figure 3.3: Left panel: Possible configuration of a T3-ToT$2C_1$&$3C_2$. Right panel: Possible configuration of a T3-$2C_1$&$3C_2$&$4C_4$. In both panels, four rings around the station with the highest signal are shown with different colours and identified by C_n, and the selected stations are within a $(6 + 5C_n)$ µs time window of the hottest station. Picture taken from [43].

The selection of physics showers and detectors belonging to each event is posterior to the data acquisition via the implementation of two successive trigger levels: the physics trigger, T4, which is based on the time and spatial configuration of the triggered detectors from a T3 event, and the fiducial/quality

trigger, T5, which ensures physics events are seen by functional stations surrounding the hottest one. The logic of this trigger chain is shown in Figure 3.4.

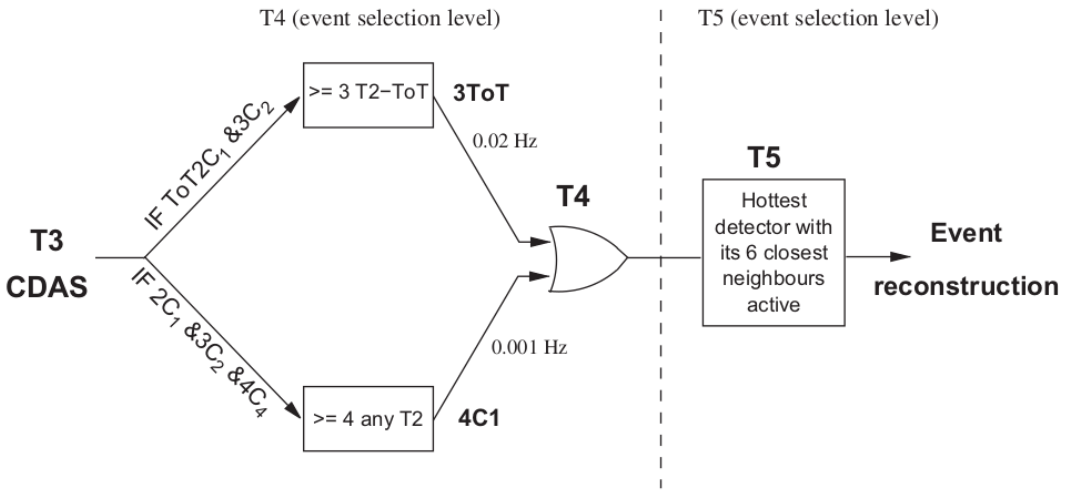

Figure 3.4: Logic sequence of the physics, T4, and fiducial/quality, T5, triggers of the Auger surface array taken from [43].

The physics trigger selects real showers from the T3 data. It operates in two levels:

- *3 ToT*: requires three nearby T2-ToT stations to form a triangular pattern with the constraint that their time signals fit to a plane shower front [c] moving at the speed of light. The selection of vertical showers with this trigger is up to 98%, reducing considerably the chance of coincidences passing the 3 ToT condition.

- *4C1*: requires four nearby stations regardless of their T2 trigger type with the additional constraint that their time signals fit to a plane shower front moving at the speed of light. This trigger brings to $\sim 100\%$ the efficiency for vertical showers.

Finally, the fiducial or quality trigger is applied. This is known as the 6T5 trigger and requires the station with the largest signal to be surrounded, at the time of triggering, by six fully functional stations, not all of them necessarily triggered. Not only does this trigger prevent the consideration of events too close to the perimeter of the array, but it also allows to discard non-operating stations, improving the overall quality of the reconstruction.

By the end of the trigger chain, the shower candidate can be reconstructed using the information provided by the selected stations. Henceforth, T2 stations belonging to a T4 event will be referred to as *candidate stations*, while nearby stations which are not T1 nor T2 will be referred to as *silent stations*. Note that the latter stations may have signals above the trigger thresholds, even though they were not flagged correctly due to the small trigger probability for low signals.

[c] Plane perpendicular to the shower axis

3.3 Reconstruction of muon number distribution in proton initiated showers using RPCs

To reconstruct the muon number distribution using the MARTA engineering array, about 6 000 full 3-D simulations of proton induced showers were run with CORSIKA v7.7410, for a primary energy of $E_0 = 10^{17.5}$ eV and $\theta = 30°$. These simulations were then run through the Offline software to emulate an Auger event. High energy hadronic interactions where simulated with EPOS-LHC. The analysis pipeline is configured in the file `ModuleSequence.xml.in`, which can be found in Appendix C. CORSIKA files are read with `EventFileReader.xml.in` and the event timestamp and shower core, chosen at random with a uniform p.d.f within a 1.5×1.5 km^2 square tile centred on station 4002 (well within the Infill array) are defined in `EventGenerator.xml.in`. A series of standard modules follow, which simulate the electronics of the water Čerenkov detector and partially that of the RPCs, as well as their response to shower particles, using the GEANT4 package. The selection of events and participating stations using the previously described trigger hierarchy is also performed, as well as the reconstruction of the shower parameters. Importantly, the modules developed in this work were `MARTALDFFinder.cc` and `MARTAAnalysis.cc`. The former executes the likelihood fit to the lateral distribution function, LDF, of the WCD signal, of the number of muons reaching the RPC pads and of the number of active fiducial pads of the RPCs, while the latter recovers and saves the relevant information in `.root` files for posterior analysis.

For the purposes of our study, we start by assuming that all stations of the Infill array are MARTA stations, before reducing the number of stations of the engineering array to just 7 in a minimal hexagonal configuration: 1 in the center and 6 first neighbours. Note that we saved all stations with SD and RPC simulation data, along with their trigger status. We have assumed the RPCs operate under ideal conditions. Moreover, we have not considered systematic uncertainties induced by the reconstruction of the shower geometry (core and shower axis) and primary energy. As such, the distance between each station and the shower axis, r, is the true distance. Furthermore, N_μ is the true number of muons measured at the ground level, i.e., 1400 m above sea level, with $E_\mu > 0.2$ GeV.

3.3.1 Impact of selecting a fiducial RPC area in electromagnetic contamination

The resistive plate chambers of the MARTA engineering array are encased in a concrete structure with a thickness of 20 cm, which is placed underneath the WCDs of the SD-750 array. This amounts to a vertical mass overburden of $\sim 170\,\mathrm{g\,cm^{-2}}$, which partially shields the RPCs from electromagnetic shower particles. Furthermore the RPCs are insensitive to photons, reducing drastically photon contamination. However, there are still sources of electromagnetic contamination. In particular, close to the shower axis, highly energetic electrons can completely traverse the water volume, preventing a pure measurement of the muonic component. Additionally, electrons arrive at surface stations preferentially in the direction of the shower axis, so that those which hit the lower portion of lateral surface of the WCD can have a track length inside the water volume short enough to still hit the RPCs. Finally, it is possible that muons

decay into electrons, after crossing the WCDs, hitting the RPCs even though their corresponding mass overburden could stop an electron from the pure electromagnetic component.

Now, RPCs count the number of charged particle hits, provided the spread in their arrival times is larger than the detector's dead-time, irrespective of the particle's type. Thus, in order to isolate the muonic hits, the electromagnetic contamination must be mitigated. This can be achieved by defining a fiducial region: a set of pads such that the water slant depth, in the direction of the shower axis, is larger than $120\,\mathrm{g\,cm^{-2}}$. This value was chosen since it corresponds to the vertical mass overburden of a volume of water with height $h_{\mathrm{water}} = 1.2\,\mathrm{m}$ and a density of $\rho = 1\,\mathrm{g\,cm^{-3}}$. Note that the definition of the fiducial area is purely geometrical, and that we are implicitly considering a planar shower front: all particles entering the station in the direction of the shower axis. This definition agrees with the used in [33] if we include the aforementioned concrete structure and extend the tolerance to $\geq 170\,\mathrm{g\,cm^{-2}}$.

To justify that choosing a fiducial area improves the purity of the measured muonic component, we recovered, for each event, the number of muonic, electromagnetic and total hits on the RPCs of candidate stations, as a function of the station's distance to the shower axis, r, for the entire and fiducial areas of the RPCs. Note that all hits were simulated, rather than reconstructed, meaning that all particles whose position lied inside some RPC pad were counted as hits, regardless of the detector's efficiency or dead time. Repeating this procedure for each candidate station of the ensemble of simulated events, and averaging the number of RPC hits produced by each component, yields the corresponding average lateral distribution function, LDF. The same procedure was implemented with the WCD signal, in VEM. The results are shown in Figure 3.5, where the actual average LDFs are plotted, and in Figure 3.6, where the percentage of the total signal or number of hits produced by each component, as a function of r, are plotted. Note that the plots have a grey band for $r < 100\,\mathrm{m}$. Although this value is somewhat arbitrary and was determined visually, it corresponds, approximately, to the distance before which the LDF of the number of muon hits stops being monotonic, either saturating or being suppressed. The physical meaning of this phenomenon close to the shower core is not clear, and it was excluded from the fits to the average LDFs, presented in the next section.

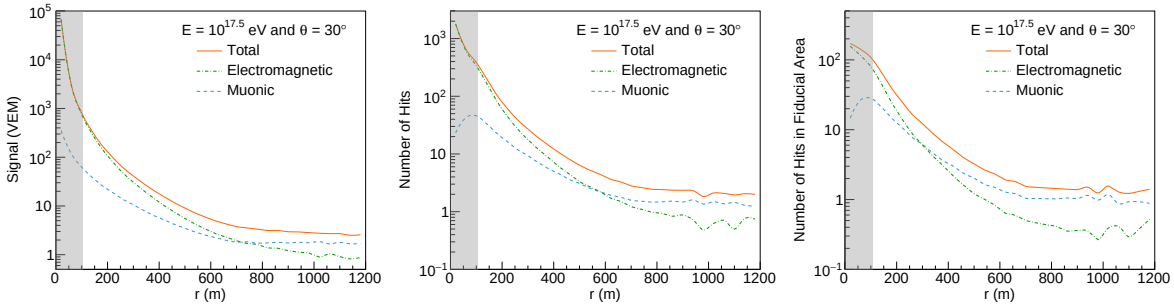

Figure 3.5: Left panel: average LDFs of the total (orange), electromagnetic (green) and muonic (blue) signals. Middle panel: average LDFs of the total, muonic and electromagnetic number of RPC hits, using the entire RPC area. Right panel: average LDFs of the total, muonic and electromagnetic number of RPC hits, restricted to the fiducial area. The ensemble of $\sim 6\,000$ proton initiated showers with primary energy $E_0 = 10^{17.5}$ eV and $\theta = 30°$, was simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

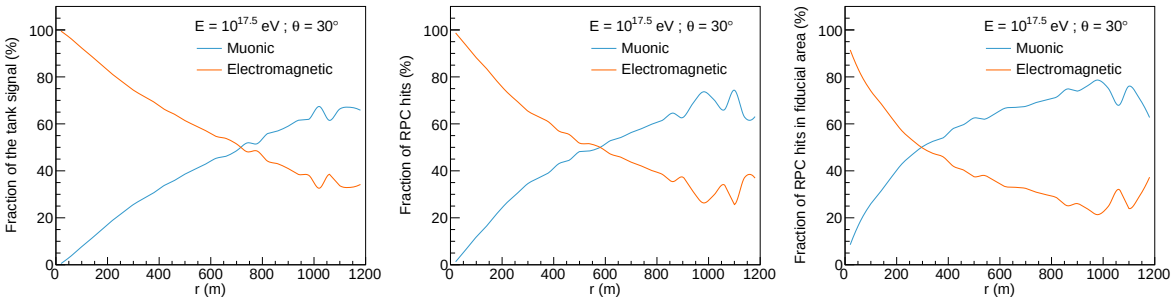

Figure 3.6: Left panel: fraction of the total WCD signal produced by the muonic (blue) and electromagnetic (orange) components. Middle panel: fraction of the total number of hits produced by muons and electrons. Right panel: fraction of the total number of hits produced by muons and electrons, in the fiducial area of the RPCs. The ensemble of $\sim 6\,000$ proton initiated showers with primary energy $E_0 = 10^{17.5}$ eV and $\theta = 30°$, was simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

Regarding the WCD signals, it is clear that the electromagnetic component dominates the measured signal for $r < 700\,\mathrm{m}$. This is expected since far from the shower axis, the depth of atmosphere traversed by shower particles is larger, absorbing partially or completely the energy carried by electromagnetic particles. On the contrary, muons are more penetrative particles, dominating the signal measured above $r > 700\,\mathrm{m}$. It is important to note that the signal produced by individual electromagnetic particles is, on average, much smaller than the signal produced by individual muons, so that the actual number of electromagnetic shower particles reaching the SD tanks is much larger than it appears from the LDFs. Note also that the total signal is not the reconstructed signal, which could be saturated closer to the shower core, but rather the sum of the simulated muonic and electromagnetic signals. Lastly, as r increases, the total signal tends to and stabilises at ~ 3 VEM, the threshold set by TH-T2, since we are only considering candidate stations, causing positive fluctuations far from the shower axis.

Regarding the number of RPC hits in all pads, it is clear that the muonic component dominates the total number of hits for $r \geq 550\,\mathrm{m}$, showing that the additional water and concrete overburden effectively mitigate the electromagnetic contamination, with respect to the WCD signals. However, the attenuation of electromagnetic particles is greatly improved by defining a fiducial area, within which muon hits already dominate the total number of hits above $r > 300\,\mathrm{m}$.

The results above show that at $r = 450\,\mathrm{m}$ about 60% of the simulated hits in the fiducial area of the RPCs are produced by muons. The importance of this distance will be made clear in the next sections. Nevertheless, we thus expect the number of hits in the fiducial area to be mostly sensitive to fluctuations of the muonic component, at the distances from the shower axis where the normalisation of the LDFs is typically extracted, for the SD-750 array. The remnant sensitivity to fluctuations of the electromagnetic component and its impact on the reconstruction of the muon distribution will be studied in the next sections.

The reconstructed detector variable closer to the actual number of simulated muon hits on the RPCs is the total number of reconstructed hits in the fiducial area corrected for pile-up effects. Pile-up occurs when the spread in arrival times of consecutive particles is smaller than the $\sim 50\,\mathrm{ns}$ dead-time of the

read-out of the RPCs [45] and although it can be accounted for using dedicated algorithms [46], they are not sufficiently studied. Thu, we have not implemented them in this study. Moreover, we have not attempted to further improve the purity of muonic signal by measuring their deposited charge on the RPCs, since a precise charge-number calibration is not yet developed and implemented in the Offline trunk [d] framework. Instead, we followed a different approach, using ideal and pessimistic case scenarios. For the pessimistic case, we approximate the number of muons hitting the RPCs by the number of active fiducial pads. This is the large dead-time limit, and very close to what could be immediately measured in an experiment. For the ideal case scenario, we use directly the simulated number of muons reaching the RPCs. Nevertheless, a more realistic analysis combining pile-up corrections to the number of measured hits with a calibration provided by the integrated charge collected by the RPCs, must be studied in a second iteration of this work.

The correlation between station-to-station fluctuations of the number of active fiducial pads and the number of muon hits in the fiducial area of the RPCs, for a distance to the shower axis in the range $100\,\mathrm{m} < r < 1400\,\mathrm{m}$, is shown in Figure 3.7 over the mentioned ensemble of 6 000 proton initiated showers. The lower limit on r was chosen to avoid phenomena close to the shower core, and the higher limit to avoid counting too many stations with positive fluctuations. In the same figure, the bin-by-bin average of the density plot is shown in black, along with the line $y = x$, corresponding to a $1:1$ ratio between the number of muon hits and active pads, in the fiducial area.

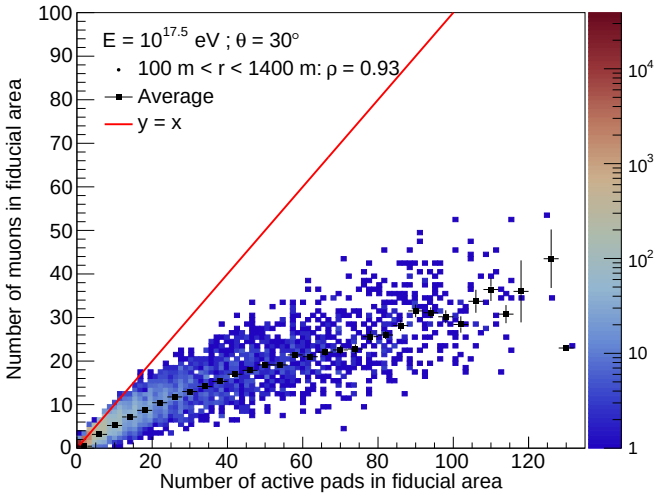

Figure 3.7: Correlation between the number of active fiducial pads and the number of muons hits in the fiducial area, for each station in the range $100\,\mathrm{m} < r < 1400\,\mathrm{m}$. The bin-by-bin average of the density plot is shown in black along with line $y = x$, corresponding to a $1:1$ ratio between the number of muon hits and active pads, in the fiducial area. The ensemble of $\sim 6\,000$ proton initiated showers with primary energy $E_0 = 10^{17.5}$ eV and $\theta = 30°$, was simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

From Figure 3.7, the very high correlation between the two quantities is apparent from their correlation factor, $\rho \sim 0.93$, indicating that, indeed, the number of active fiducial pads is a good proxy to

[d] The development area of Offline

the number of muon hits, station-wise. Additionally, we note that the number of active fiducial pads is usually larger than the number of muon hits, indicating the presence of a residual electromagnetic contamination. A fraction of this contamination comes from the pure electromagnetic shower component, while the remaining contamination is expected to come from the decay of $\mu^- \to e^- \overline{\nu}_e \nu_\mu$. The correlation between these two variables is further worsen by more than one muon hitting a given pad. This effect is predominant close to the shower core, where the muon density is larger the spatial resolution of the RPCs. Finally, the accumulation of points in $(0, 0)$ stems from the fact that we are considering stations with RPC and WCD simulation data, but not necessarily active fiducial pads.

Ultimately, the physical variable chosen to reconstruct the distribution of number of muons is the number of active fiducial pads of the RPCs.

3.3.2 Average Muon LDF

Shower-to-shower fluctuations of the true number of muons at the ground level correlate with fluctuations of the overall scale of the Muon Lateral Distribution Function, MLDF, that is, the number density of muons as a function of the distance to the shower axis r, $\rho_\mu(r)$:

$$\rho_\mu(r) = \frac{1}{2\pi r} \frac{\mathrm{d}N_\mu}{\mathrm{d}r}, \tag{3.8}$$

Note that in this formula, we assumed an axial symmetry around the shower axis. Over an ensemble of showers with the same primary energy and zenith angle, the shape of $\rho_\mu(r)$ is approximately the same, but its overall scale is sensitive to the muon content of the shower. So, the shape of the event-by-event MLDF is mostly fixed by the shape of the average MLDF, $\langle \rho_\mu(r) \rangle$. We can illustrate this correlation by taking the value of the true MLDF at $r = 450\,\mathrm{m}$ and the muon content of each event and produce a density plot, as shown in Figure 3.8, along with the value of the correlation factor, ρ.

Clearly, the correlation between N_μ and $\rho_\mu(450)$ is very strong, with $\rho \simeq 0.98$.

However, the true MLDF in real events is not known. Instead, it must be reconstructed from the data collected by surface detectors. As mentioned before, for the case of the RPCs this variable is the number of active fiducial pads in each station. Hence, we must compare the true average MLDF, $\langle \rho_\mu(r) \rangle$ with the average lateral distribution of the number density of active fiducial pads of the RPCs per station, $\langle M(r) \rangle$, before proceeding with an event-by-event analysis.

To plot $\rho_\mu(r)$, we recover the radial distance of each muon to the shower axis, and produce an histogram such that the content of each bin of width $\mathrm{d}r$ is the number of muons whose distance to the shower axis is between r and $r + \mathrm{d}r$, i.e., $\mathrm{d}N_\mu \,/\, \mathrm{d}r$, divided by $\pi \left((r + \mathrm{d}r)^2 - r^2 \right)$. In the limit where $\mathrm{d}r \to 0$ we recover the exact value of $\rho_\mu(r)$ at each point. Averaging, point-wise, over the ensemble of proton initiated events gives $\langle \rho_\mu(r) \rangle$.

To determine $\langle M(r) \rangle$, we count, event by event, the number of active fiducial pads in each RPC station, before dividing this number by the projection of the fiducial RPC area onto the shower plane. Indeed, the efective fiducial area, $A_{\mathrm{fid}}^{\mathrm{eff}}$, for a shower arriving with zenith angle θ, is related to the physical

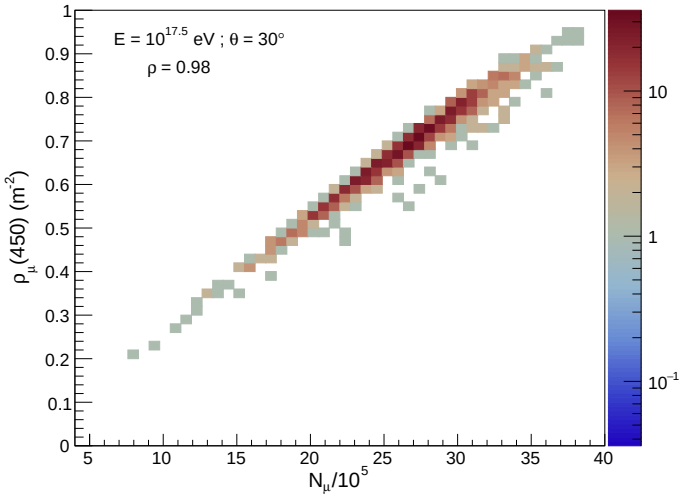

Figure 3.8: Correlation between $N_\mu/10^5$ and $\rho_\mu(450)$, over an ensemble of 1 000 CORSIKA simulations of proton initiated showers with $E_0 = 10^{17.5}$ eV and $\theta = 30°$, with EPOS-LHC. The corresponding Auger events were simulated and reconstructed using the Offline framework.

fiducial RPC area, A_{fid}, by

$$A_{\text{fid}}^{\text{eff}} = A_{\text{fid}} \cos\theta, \tag{3.9}$$

since the RPC plane is horizontal. Note that, as expected, for very inclined showers, $\theta \to \frac{\pi}{2}$, the effective fiducial area tends to zero. Additionally, we use two base cases to establish a comparison with the true MLDF: the number density of simulated muons whose trajectories cross WCDs, regardless of their trigger status, and the number density of muon RPC hits, which corresponds to the ideal situation where we could effectively measure this quantity. The effective detection area in the latter case is given by

$$A_{\text{RPC}}^{\text{eff}} = A_{\text{RPC}} \cos\theta, \tag{3.10}$$

On the other hand, the detection area of the WCDs projected onto the shower plane is given by

$$A_{\text{tank}}^{\text{eff}} = \pi r_{\text{tank}}^2 \cos\theta + 2 h_{\text{water}} r_{\text{tank}} \sin\theta, \tag{3.11}$$

given their non-zero height and circular horizontal cross section. For $\theta = 30°$ we have, approximately

$$\begin{aligned} A_{\text{tank}}^{\text{eff}} &\simeq 11.0\,\text{m}^2 \\ A_{\text{RPC}}^{\text{eff}} &\simeq 6.7\,\text{m}^2 \\ A_{\text{fid}}^{\text{eff}} &\simeq 5.0\,\text{m}^2, \end{aligned} \tag{3.12}$$

Additionally, we include the average LDF of the number of all active pads of the RPCs. The five aforementioned average LDFs for an ensemble of 1 000 events, are shown in the left panel of Figure 3.9. On the right panel, we plot the average LDFs of the number density of muons crossing the WCDs, the number density of muon RPC hits and the number density of active fiducial pads, divided, point-wise, by

the true MLDF.

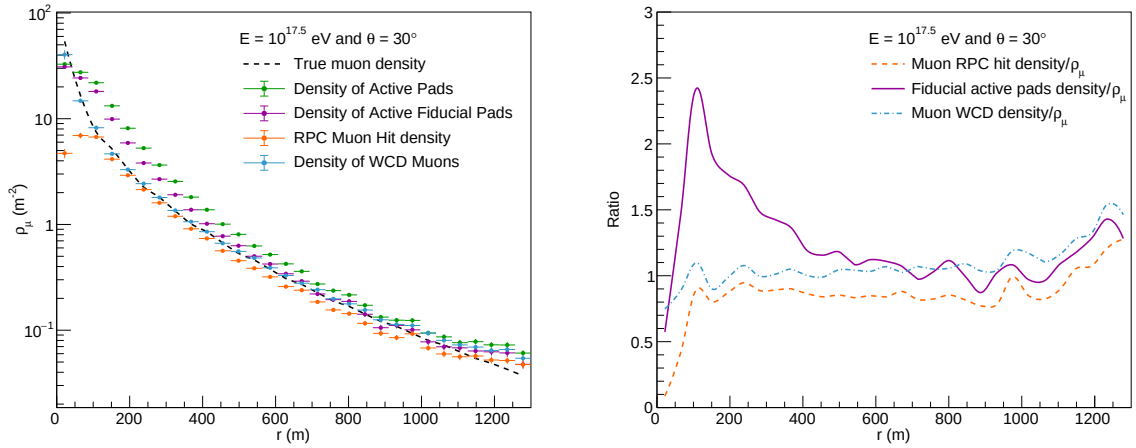

Figure 3.9: Left panel: average LDF of the true number density of muons at the ground level (black), of the number density of active RPC pads (green) and active fiducial RPC pads (purple), along with the number density of muon hits on the RPCs (orange) and of the muons crossing the WCD (blue). Right panel: point-by-point ratio between the average LDFs of the number density of muon hits on the RPCs (orange), of active fiducial pads (purple) and of muons crossing the WCD (blue), and the real MLDF. Both figures were produced using 1 000 CORSIKA-FLUKA simulations of proton initiated showers with $E_0 = 10^{17.5}$ eV and $\theta = 30°$, with EPOS-LHC. The corresponding Auger events were simulated and reconstructed using the Offline framework.

As expected, the average LDF of the number of simulated muons crossing the WCDs is very close to the true MLDF, except arbitrarily close to the shower core, where it underestimates the true number of muons, and above $r > 1000\,\mathrm{m}$, where positive fluctuations overestimate the true number density of muons. These fluctuations come from considering just stations with both SD and RPC simulation data, introducing a positive bias far from the shower core.

It is also clear that both the number density of active pads across the entire RPC and on the fiducial area overestimate, on average, the number of muons at each radial distance to the shower core, especially for $100 < r < 500$ m, due to electromagnetic contamination. Nevertheless, restricting the total active area to the fiducial region of the RPCs further mitigates this contamination. In particular, the ratio between the number of active fiducial pads and the muon LDF rises to 2.5 at $r = 150\,\mathrm{m}$, before steadily dropping to ~ 1 with r. At $r = 600\,\mathrm{m}$ the number of active fiducial pads is identical, on average, to the true number of muons at that distance. Note that for $r < 150\,\mathrm{m}$, where the density of shower particles is much larger than the spatial resolution of the RPCs, both LDFs saturate, since the number of available RPC pads is just 256. This explains the drop in the ratio between $\langle M(r)\rangle \,/\, \langle \rho_\mu(r)\rangle$ in this region. The existence of positive fluctuations is also evident for $r > 1000\,\mathrm{m}$.

Apart from the region close to the shower core, the LDF of the number of muon hits on the RPCs accounts, for each r value, for about $\sim 85\%$ of the real MLDF, indicating that about ~ 15 % of the muons completely degrade their energy inside the WCDs. We can present a rough estimation of this value considering the kinetic energy spectrum of the muons arriving at the WCDs and the muon stopping

power in pure water and concrete. The muon stopping power in water and concrete at minimum ionising energy is given by [8]

$$-\frac{1}{\rho}\left\langle\frac{\mathrm{d}E_\mu}{\mathrm{d}x}\right\rangle_{\mathrm{min}} \sim 2\,\mathrm{MeV\,cm^2\,g^{-1}}, \tag{3.13}$$

where ρ is the density of the medium, in this case either water or concrete, $\mathrm{d}x$ is an infinitesimal distance travelled by the muon over which its energy decreases by $\mathrm{d}E_\mu$. We now assume this value is constant throughout the kinetic energy spectrum of shower muons at the ground level, and that the angle subtended by the muon's arrival direction at the station and the vertical is the zenith angle of the shower axis $\theta = 30°$. Thus, considering a mass overburden of $170\,\mathrm{g\,cm^{-2}}$, the minimal energy necessary for a muon to completely traverse the water and concrete volumes is

$$E_{\mathrm{min}} \sim 170\,\mathrm{g\,cm^{-2}} \times \sec(30°) \times 2\,\mathrm{MeV\,cm^2\,g^{-1}} \simeq 390\,\mathrm{MeV}. \tag{3.14}$$

In Figure 3.10 we plotted the decimal logarithm of the kinetic energy spectrum of all muons arriving at the water Čerenkov detectors and that of muons with energy less than E_{min}, over an ensemble of $\sim 1\,000$ proton initiated showers with $E_0 = 10^{17.5}$ eV and $\theta = 30°$.

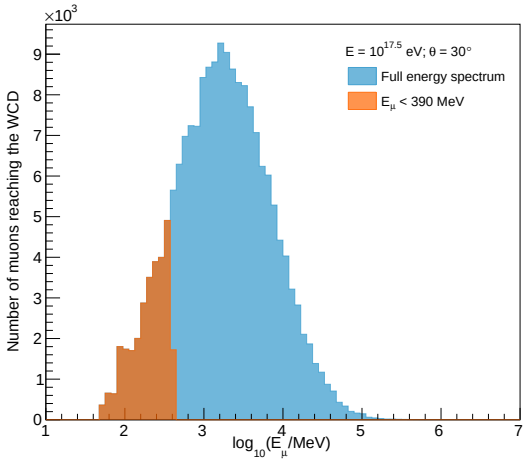

Figure 3.10: Kinetic energy spectrum of muons arriving the at the WCD. The full spectrum is in blue, while the spectrum for $E_\mu < 390\,\mathrm{MeV}$ is shown in orange.

Dividing the integrated kinetic energy spectrum of muons with $E < E_{\mathrm{min}}$, by the integral of the full kinetic energy spectrum yields, approximately, the percentage of muons stopped by the water and concrete volumes. As expected, this value is $\sim 16\%$, in good agreement with what we observed in the right panel of Figure 3.9.

We have established the similarity between the LDFs of the number of active fiducial pads and the true muon LDF above $r \sim 400\,\mathrm{m}$, that the overall scale of the LDF correlates with the muon content of the shower, and that there exists a high correlation between the number of muon RPC hits in the fiducial area and the number of active fiducial pads of the RPCs. Thus, we will take $M(450)$, the expected value of the LDF of the number of active fiducial pads, in one event, at $r = 450\,\mathrm{m}$, to probe the number of muons at the ground level, N_μ. The impact of the residual sensitivity of this variable to shower-to-shower

fluctuations of the electromagnetic component will be studied by comparing the obtained results with a base case: the number of expected muon hits at $r = 450\,\mathrm{m}$, $N_\mu^{\mathrm{Hits}}(450)$. Lastly, note that the total number of active fiducial pads in a given event would not be a good estimator of the muon content of the shower, since the position of the shower core is random.

3.3.3 Reconstruction of the event-by-event Muon LDF

To infer the number of muons at the ground level of a given event, it is necessary to fit the LDF of the number of active fiducial RPC pads, and determine its value at an optimal radial distance from the shower axis, $M(r_{\mathrm{opt}})$. As is the case with the true MLDF, it is expected that $M(r_{\mathrm{opt}})$ is sensitive to fluctuations of N_μ, in spite of the electromagnetic contamination. As a base case, and given its importance in the energy reconstruction of EAS, we shall also fit the LDF of the reconstructed signal measured by the SD, $S(r)$, in VEM, to determine the shower size $S(r_{\mathrm{opt}})$. The actual value of r_{opt} depends on the spacing of the array.

A modified Nishimura-Kamata-Greisen (NKG) type fit function was used to fit both lateral distributions [47]:

$$S(r) = S(r_{\mathrm{opt}}) \left(\frac{r}{r_{\mathrm{opt}}}\right)^{\beta} \left(\frac{r + r_{\mathrm{scale}}}{r_{\mathrm{scale}} + r_{\mathrm{opt}}}\right)^{\beta+\gamma} \tag{3.15a}$$

$$M(r) = M(r_{\mathrm{opt}}) \left(\frac{r}{r_{\mathrm{opt}}}\right)^{\beta_{\mathrm{M}}} \left(\frac{r + r_{\mathrm{scale}}}{r_{\mathrm{scale}} + r_{\mathrm{opt}}}\right)^{\beta_{\mathrm{M}}+\gamma_{\mathrm{M}}}, \tag{3.15b}$$

where $r_{\mathrm{scale}} = 700\,\mathrm{m}$, while β (β_{M}) and γ (γ_{M}) control the shape of the LDF. The β parameter controls the overall steepness of the LDF, while γ controls its shape far from the shower axis, where the muonic component dominates. Both shape parameters are a function of the zenith angle and shower energy. Furthermore, r_{opt} is the radial distance least dependent on the functional shape of the LDF, so that the evaluation of the LDF at this radius yields the most robust estimate of the shower size. While this value is usually determined, we fixed it to $450\,\mathrm{m}$ for the SD-750. In fact, the fit to $S(r)$ is usually done is several stages by the same module that fits the shower core and axis: `LDFFinder`. Here, since the latter quantities are fixed, we simplified this procedure, as we will describe below.

The shape parameters of the fit to the LDF of WCD signal, were parametrised by a polynomial in $\sec\theta$ and $\log_{10} S(450)$ [48]:

$$\beta = a_1 + b_1 \log_{10} S(450) + (c_1 + d_1 \log_{10} S(450)) \sec\theta + (e_1 + f_1 \log_{10} S(450)) \sec^2\theta. \tag{3.16a}$$

$$\gamma = a_2 + b_2 \log_{10} S(450) + (c_2 + d_2 \log_{10} S(450)) \sec\theta + (e_2 + f_2 \log_{10} S(450)) \sec^2\theta. \tag{3.16b}$$

The values of a_i, b_i, c_i, d_i, e_i and f_i can be found in Table 3.1.

Parameters	a_i	b_i	c_i	d_i	e_i	f_i
β	-2.379	-0.191	0.694	0.096	0.102	-0.182
γ	-1.047	0.983	-1.754	-0.786	0.715	0.621

Table 3.1: Values of the coefficients of the parametrisation of β and γ, according to [48].

Since we were only interested in finding the shower size $S(450)$, we have not considered the systematic uncertainties in β and γ.

The shape parameters for $M(r)$ were extracted from a χ^2-fit to the average LDF of the number of active fiducial pads, $\langle M(r)\rangle$, since all simulated showers have the same zenith angle and energy. While $r_{\text{opt}} = 450\,\text{m}$ and $r_{\text{scale}} = 700\,\text{m}$ were fixed, $\langle M(450)\rangle$, β_{M} and γ_{M} were left free. Note that we are assuming that the shape of the LDF is not strongly dependent on the muon content of the shower, only the event-by-event $M(450)$ is. Any residual dependence, could be accounted for by parametrising β_{M} and γ_{M} in terms of $\sec\theta$ and $\log_{10} M(450)$, or by freeing these parameters on a event-by-event basis, provided the number of candidate stations exceeds the number of free parameters, in this case, 3. However, we have not attempted to tune the event-by-event fit to the LDF beyond the necessary to reconstruct the muon number distribution. A study of the impact of further tuning the LDF fit could be done in a second iteration of this work.

Figure 3.11 shows the fit to the average LDF of the number of active fiducial pads of the RPCs over an ensemble of 1 000 proton induced showers. Additionally, the figure shows the station-to-station number of active fiducial pads of the RPCs as grey dots with an error bar of $\sqrt{M}$, where M is the number of active fiducial pads of a given station. Note that the fit window was limited to $r \in [100\,,\,1000]\,\text{m}$ to avoid the saturation near the shower core and positive fluctuations far from it. Again, these values are approximate and decided through a close inspection of the average LDF. Moreover, the fit includes all stations with RPC and WCD simulation data, not just candidate stations, to avoid any trigger effects and extract unbiased values of the shape parameters of the LDF.

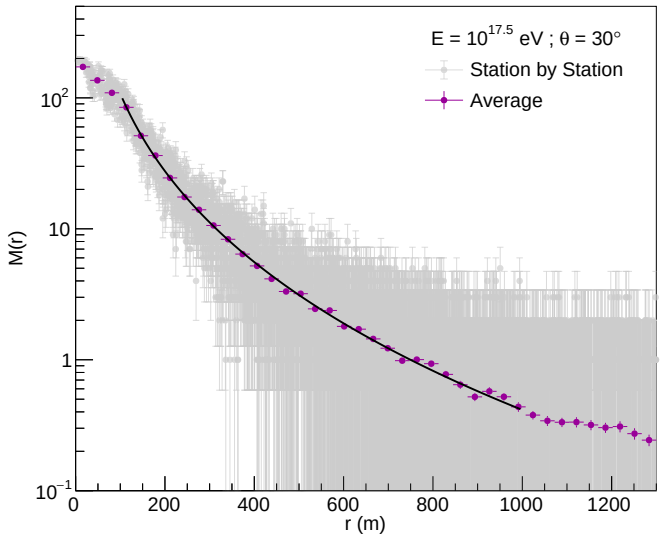

Figure 3.11: χ^2-fit (black) to $\langle M(r)\rangle$ (purple), along with the number of active fiducial pads of the RPCs per station (grey dots).

The values of the LDF parameters as presented in Equation (3.15b) are shown in Table 3.2 along with the value of χ^2_{red}.

For the determination of $S(450)$ and $M(450)$, the quantity sensitive to fluctuations of the muon content

Parameter	Value
r_{opt}	450 m
r_{scale}	700 m
$M(450)$	4.12 ± 0.47
β_M	-1.48 ± 0.53
γ_M	-1.4 ± 2.5
χ^2_{red}	$34.63/25$

Table 3.2: Results of the χ^2-fit performed to the average LDF of the number of active fiducial pads.

of EAS, for each event, a likelihood fit to the LDF of the WCD signal and number of active fiducial pads was performed. This fit method was chosen to take into account the trigger probability of SD stations, and include saturated and silent stations in the LDF. Otherwise, a regular χ^2 fit would assume a Gaussian likelihood, leading to biases in the values of the extracted parameters.

Suppose the station measures a signal S. We assume that S is a realisation of a random variable s with a piecewise continuous probability density function, according to the s value. In contrast, a simple χ^2-fit assumes that s follows a normal distribution. Instead, we determine, for each station i, the likelihood $\mathcal{L}_i$, which is proportional to the probability that a signal S is measured, given the set of parameters $S(r_{opt})$, β, γ, r_{opt} and r_{scale}. The most likely set of parameters is the one that maximises the total likelihood, $\mathcal{L}$, obtained by multiplying $\mathcal{L}_i$ over all the event stations. The same reasoning applies to the LDF fit to the number of active fiducial pads.

Before performing the likelihood fit to the reconstructed WCD signal, it is necessary to convert S, in VEM, into an effective number of particles, since, on average, the signal produced by single electrons and photons is much dimmer than the one produced by muons, and the considered p.d.fs apply to the number of particles. Following [49] the effective particle number, n_{eff}, is given by

$$n_{eff} = p \times \left(\frac{S}{\text{VEM}} \right), \tag{3.17}$$

where p is the Poisson factor, parameterised by

$$p = t + \frac{1-t}{\sqrt{f_S}}, \quad \text{with } t \equiv \frac{1}{1+e^z} \quad \text{and} \quad z \equiv 40\,(0.98 - f_S)\,, \tag{3.18}$$

where f_S is the signal uncertainty parametrised as

$$f_S \equiv \frac{\sigma(S)}{\sqrt{S}} = 0.865\,(1 + 0.593\,(\sec\theta - 1.22))\,. \tag{3.19}$$

The likelihood function for estimating the LDF of the WCD signal in VEM, is given by

$$\mathcal{L}_{SD} = \prod_{i=1}^{n_{cand}} P(n_i, \mu_i) \prod_{j=1}^{n_{sat}} P_{sat}(n_j, \mu_j) \prod_{k=1}^{n_{silent}} P_{zero}(n_k, \mu_k) \tag{3.20}$$

where n is the effective number of particles given in Equation (3.17) for a measured signal S, $\mu = \mu(S(450), \beta, \gamma)$ is the expected number of particles according to Equation (3.15a) with $S(450)$, β and

γ, i runs-over n_{cand} candidate unsaturated stations, j over the n_{sat} candidate saturated stations and k over the n_{silent} silent stations, stations that do not form a trigger. The probability density functions are presented below.

- For unsaturated signals measured by candidate stations, the p.d.f reads:

$$P(n_i, \mu_i) = \begin{cases} \dfrac{\mu_i^{n_i} \mathrm{e}^{-\mu_i}}{n_i!} & \text{if } \mu_i \leq 30 \\ \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(n_i-\mu_i)^2}{2\sigma^2}\right) & \text{if } \mu_i > 30 \end{cases} \tag{3.21}$$

  The lower branch takes into account the Poisson statistics followed by small numbers of counts, while the higher branch takes advantage of the Central Limit Theorem to approximate the p.d.f of large signals to a Gaussian.

- For saturated signals, we consider two cases: either the saturated signal can be recovered, in which case we adopt a normal distribution with a standard deviation that accounts for both the signal and recovery uncertainties

$$\sigma^2 = \sigma_{\text{S}}^2 + \sigma_{\text{recovery}}^2, \tag{3.22}$$

  or the signal cannot be recovered and the p.d.f is the probability of measuring an effective number of particles larger than n_i. That is

$$\frac{1}{\sqrt{2\pi}\sigma} \int_{n_i}^{\infty} \mathrm{d}n_i \exp\left(-\frac{(n_i-\mu_i)^2}{2\sigma^2}\right) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{n_i-\mu_i}{\sqrt{2}\sigma}\right)\right], \tag{3.23}$$

  where erf(x) denotes the error function. Hence,

$$P_{\text{sat}}(n_i, \mu_i) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(n_i-\mu_i)^2}{2\sigma^2}\right) & \text{if the recovery is successful} \\ \dfrac{1}{2}\left[1 - \text{erf}\left(\frac{n_i-\mu_i}{\sqrt{2}\sigma}\right)\right] & \text{otherwise} \end{cases} \tag{3.24}$$

  A detailed description of the recovery of the PMT signal can be found in [50].

- For silent stations, the probability density function must give the probability that the effective number of particles measured at the station is smaller than some threshold, taken to be $n_{\text{th}} = 3$ [48]. Thus, we must sum over the Poisson probabilities of having less than n_{th}:

$$P_{\text{zero}} = \sum_{i=1}^{n_{\text{th}}} \frac{\mu_i^{n_i} \mathrm{e}^{-\mu_i}}{n_i!}. \tag{3.25}$$

In turn, the measured number of fiducial active pads is already an integer close to the number of muon hits on the RPCs, so that the number of particles does not need to be parametrised in terms of the measured $M(r)$ value. Moreover, we have not considered the recovery of saturated RPC stations, as there is not an algorithm to perform this recovery in the simulation framework. Instead, we simply

discard events with saturated stations. A method to recover the number of saturated active fiducial pads of MARTA stations could be implemented in a second iteration of this work. Finally, the trigger of MARTA stations is provided by the SD, so that the likelihood function to estimate $M(r)$ must take into account stations that are flagged as silent if they don't belong to a T4. Hence, the likelihood function reads

$$\mathcal{L}_{\mathrm{M}} = \prod_{i=1}^{n_{\mathrm{cand}}} P(m_i, \mu_i) \prod_{k=1}^{n_{\mathrm{silent}}} P_{\mathrm{zero}}(n_k, \mu_k), \tag{3.26}$$

where m_i is the measured number of active fiducial pads of candidate station i, $\mu_i = \mu_i(M(450), \beta_{\mathrm{M}}, \gamma_{\mathrm{M}})$ is the expected number of active fiducial pads given Equation (3.15b). The probability $P_{\mathrm{zero}}(n_k, \mu_k)$ is exactly the same as given in Equation (3.25), while $P(m_i, \mu_i)$ is the one given in Equation (3.21), but with the number of active fiducial pads of the RPCs instead of the effective number of particles measured by the WCDs.

The actual fit is implemented in the module `MARTALDFFinder`, using the `MINUIT` minimisation package to minimise $-\ln \mathcal{L}$, by varying the shape parameters and shower sizes of the LDFs until the minimum -log-likelihood value is found. Note that, through the parameterisation of β and γ, and by taking β_{M} and γ_{M} from the fit to the average $M(r)$, the only free parameters, for each event, are $S(450)$ and $M(450)$. Additionally, the values of $\langle S(450) \rangle$ and $\langle M(450) \rangle$ are used as seeds to start the minimisation.

To emulate the ideal scenario in which we could have access to the number of muons reaching the RPC pads, N_μ^{Hits}, we have also performed a likelihood-fit to estimate $N_\mu^{\mathrm{Hits}}(r)$, using a NKG-type fit function, in order to extract the value of $N_\mu^{\mathrm{Hits}}(450)$, that is, the expected number of muon hits on the RPCs at $r = 450\,\mathrm{m}$ from the shower axis. Note that we could have used a different fit function, namely the MLDF parametrisation proposed by the KASKADE-Grande Collaboration: a modified Greisen formula [51]. However, the NKG function fits sufficiently well, and we did not attempt to tune the fit to the LDF beyond the necessary. The study of the impact of the functional form of the LDF on the reconstruction of the number of muons must be investigated in the future, although it is expected to be small since $\sim 450\,\mathrm{m}$ is very close to r_{opt}. To fit $N_\mu^{\mathrm{Hits}}(r)$, we have not considered trigger effects, so that the likelihood function to maximise reads

$$\mathcal{L}_\mu = \prod_{i=1}^{n_{\mathrm{total}}} P(n_i, \mu_i), \tag{3.27}$$

where i runs over all event stations, n_i is the number of simulated muon hits on the RPCs and μ_i is the expected number of muons hits for a given set of LDF shape parameters. As before, $P(n_i, \mu_i)$ is either a Gaussian or a Poisson distribution, depending on the value of n_i. Once again, β and γ were obtained from the χ^2-fit to the average LDF of the number of muon hits. Their values are displayed in Table 3.3, along with other LDF parameters and χ^2_{red}. The resulting plot is shown in Figure 3.12.

Note that in order to avoid the region too close to the shower core, and positive fluctuations far from it, the fit range was limited to $r \in [150\,,\,1000]\,\mathrm{m}$.

Proceeding in a event-by-event fashion, the values of $S(450)$, $M(450)$ and $N_\mu^{\mathrm{hits}}(450)$, along with their uncertainties were obtained from the likelihood fit. Figure 3.13 shows the fit to lateral distribution function of the WCD signal, of the number of active fiducial pads and of the number of muon hits, for one event.

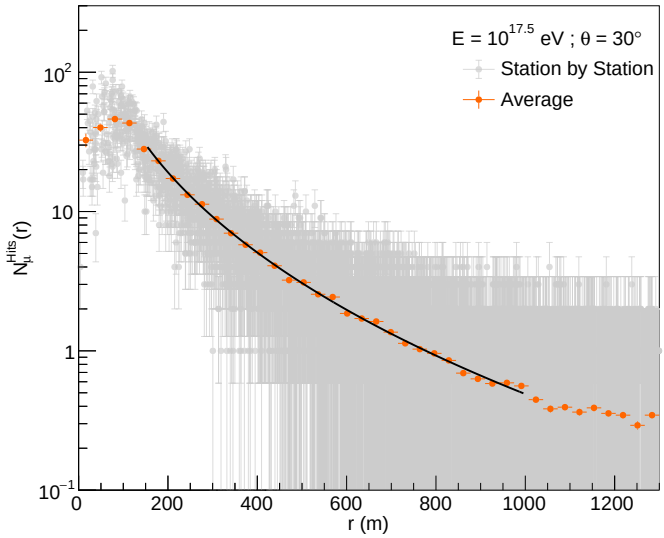

Figure 3.12: χ^2-fit (black) to $\left\langle N_\mu^{\text{Hits}}(r)\right\rangle$ (orange), along with the number of muon hits on the RPCs per station (grey dots).

Parameter	Value
r_{opt}	450 m
r_{scale}	700 m
$N_\mu^{\text{Hits}}(450)$	3.91 ± 0.56
β	-0.9 ± 1.1
γ	-2.5 ± 3.6
χ^2_{red}	$48.38/23$

Table 3.3: Results of the χ^2-fit performed to the average LDF of the number of muon hits on the RPCs.

In all cases, the quality of the fit is visibly good, although the number of candidate stations is small, in virtue of the relatively low primary energy and zenith angle.

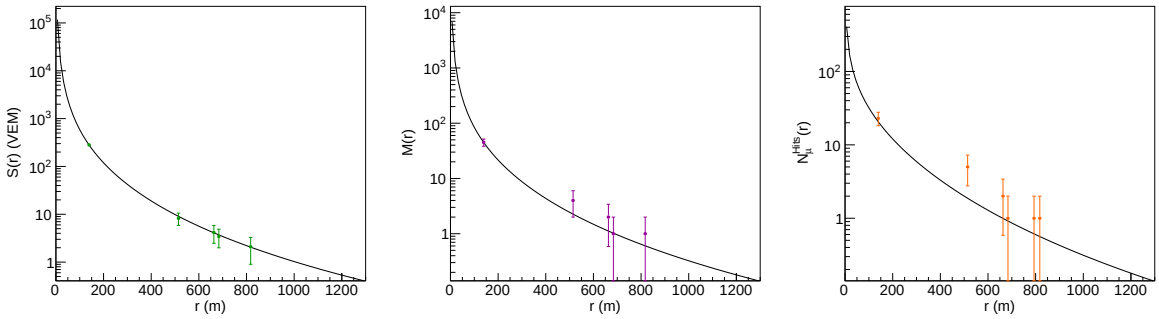

Figure 3.13: Left panel: LDF fit to the WCD signal, $S(r)$, in VEM. Middle panel: LDF fit to the number of active fiducial pads of the RPCs, $M(r)$. Right panel: LDF fit to the number of muon hits on the RPCs, $N_\mu^{\text{Hits}}(r)$. The shower with primary energy $E_0 = 10^{17.5}$ eV and $\theta = 30^\circ$ was simulated with CORSIKA using EPOS-LHC, and the corresponding Auger event simulated with Offline.

3.3.4 Reconstruction of the muon number distribution with the full SD-750

For each event, the values of $S(450)$, $M(450)$ and $N_\mu^{\text{Hits}}(450)$ were plotted against the true number of muons at the ground level, N_μ, scaled by 10^{-5}. Events with saturated stations were excluded, since the value of the LDF at $r = 450\,\text{m}$ would be underestimated. In the end, 5 105 of the original ensemble of 6 000 proton initiated events with $E_0 = 10^{17.5}$ eV and $\theta = 30°$, simulated with CORSIKA and EPOS-LHC were kept and used to produce the density plots shown in Figure 3.14. The correlation factor and a χ^2 linear fit with equation $y = m(N_\mu/10^5) + b$ are also displayed in each plot of the same figure. The factor of 10^{-5} can be absorbed by redefining the slope m: $\tilde{m} = m/10^5$. Note that the fit was performed to the scatter plot corresponding to the density plot.

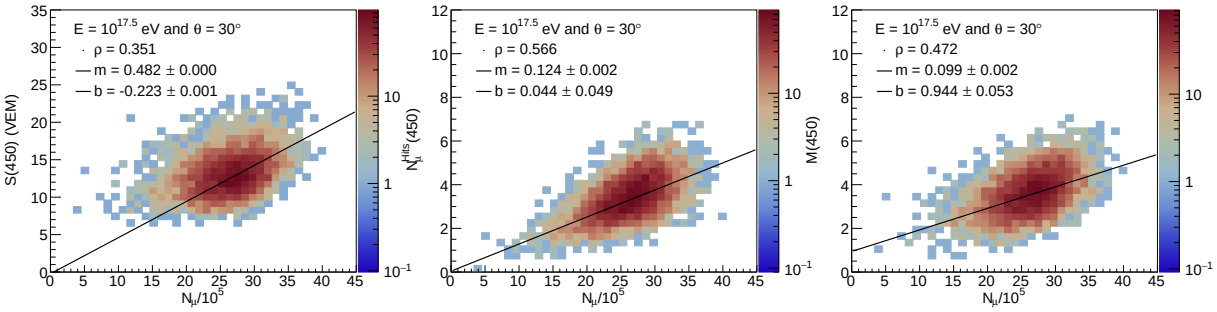

Figure 3.14: Left panel: Correlation between the distributions of $S(450)$ and N_μ. Middle panel: Correlation between the distributions of $N_\mu^{\text{Hits}}(450)$ and N_μ. Right panel: Correlation between the distributions of $M(450)$ and N_μ. Linear χ^2-fits of the form $y = m(N_\mu/10^5) + b$ are shown in black, and the fit parameters, along with the correlation factor, can be read off the top left of each panel. The ensemble of 5 105 showers with primary energy $E_0 = 10^{17.5}$ eV and $\theta = 30°$ was simulated with CORSIKA using EPOS-LHC, and the corresponding auger event simulated with Offline.

The values of $S(450)$ and N_μ are poorly correlated, with $\rho \simeq 0.35$, since the reconstructed signal measured by the WCD is also sensitive to shower-to-shower fluctuations of the pure electromagnetic component, which is significant at $\theta = 30°$. Furthermore, the flat shape of the density plot for low $S(450)$ values is indicative of a threshold effect, due to the low shower energy.

On the other hand, the correlation between $M(450)$ and N_μ is better, although worse than that of $N_\mu^{\text{Hits}}(450)$ and N_μ, probably due to the residual electromagnetic contamination. The shape of these density plots is also indicative a threshold effect, although not so significant. Moreover, the y-intercept of the fit curve $N_\mu^{\text{Hits}}(450) = \tilde{m}N_\mu + b$, is compatible with 0 at $1\,\sigma$, in contrast with the value of b for $M(450) = \tilde{m}N_\mu + b$. This might reflect the electromagnetic contamination. Nevertheless, it is apparent that even in the ideal case, where one could access directly to the number of muon hits, the array sensitivity to fluctuations of the muon content is low.

The reconstructed number of muons of each shower can be obtained from the value of $M(450)$ by inverting the linear relation

$$M(450) = \tilde{m}N_\mu^{\text{rec}} + b \iff \frac{M(450) - b}{\tilde{m}} = N_\mu^{\text{rec}}, \tag{3.28}$$

where $\tilde{m} \times 10^5$ and b are given in the right plot of Figure 3.14. The statistical uncertainty of the reconstructed number of muons is obtained by propagating the statistical errors of $\tilde{m}$, b and $M(450)$

$$\sigma^2_{N_\mu^{\text{rec}}} = \frac{1}{(\tilde{m})^2}\left(\sigma^2_{M(450)} + \sigma^2_b + \left(N_\mu^{\text{rec}}\right)^2 \sigma^2_{\tilde{m}}\right) \tag{3.29}$$

assuming $\tilde{m}$, b and $M(450)$ to be independent. Since the core, shower axis and primary energy are the true ones, we are not considering systematic uncertainties in the reconstruction of N_μ. By the same token, N_μ can also be obtained from $N_\mu^{\text{Hits}}(450)$, although this quantity is not physically accessible in real measurements. We have not used the shower-to-shower fluctuations of $S(450)$ to reconstruct the muon number distribution, since the correlation between these variables is poor and assuming a linear relation between the two would be erroneous.

To determine the resolution and bias induced by the reconstruction procedure in the measurement of N_μ, the residuals

$$r_\mu = \frac{N_\mu^{\text{rec}} - N_\mu}{N_\mu}, \tag{3.30}$$

were computed for each event, and their distribution over the ensemble of events was built. Note that our definition of r_μ absorbs the physical fluctuations of the muon content by dividing by N_μ. The resolution associated with the reconstruction procedure is the width of the distribution of residuals, $\sigma(r_\mu)$, while the bias is the mean value, $\langle r_\mu \rangle$. We extract these quantities by fitting the distribution of r_μ to a Gaussian function of the form

$$\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}r_\mu} = \frac{\mathcal{N}}{\sqrt{2\pi}\sigma_{\text{rec}}} \exp\left(-\frac{(r_\mu - \mu)^2}{2\sigma^2_{\text{rec}}}\right), \tag{3.31}$$

using the method of least squares. Hence, the resolution is given by the value of σ_{rec} and the bias by the value of μ. Moreover, to determine how the resolution and bias vary with the considered region of the true distribution of N_μ, the distribution of residuals was computed for the low tail region, $\ln N_\mu \in [0, 14.6]$, the bulk, $\ln N_\mu \in]14.6, 14.9]$ and the high tail region, $\ln N_\mu \in]14.9, 16.0]$.

Figure 3.15 shows the fitted distributions of residuals of the reconstructed number of muons at the ground level for the ensemble of 5 105 unsaturated events, considering the three aforementioned regions of the real distribution of $\ln N_\mu$, along with the distribution of the residuals over the entire ensemble. The left panel refers to the reconstruction based on $M(450)$ and the right panel to the reconstruction based on $N_\mu^{\text{hits}}(450)$. The values of the fit parameters can be found in Table 3.4 along with the value of χ^2_{red}.

The distributions of residuals are approximately Gaussian for all regions of the muon number distribution, as shown by the values of $\chi^2_{\text{red}} \sim 1$, although those referring to showers with low muonic content display a tail towards high r_μ values. Moreover, and regardless of the reconstruction variable, $M(450)$ or $N_\mu^{\text{Hits}}(450)$, N_μ^{rec} is a positively biased estimator. The reconstruction based of $M(450)$ induces a resolution of $\sigma_{\text{rec}} = 0.32$, which is larger than that based on $N_\mu^{\text{Hits}}(450)$, $\sigma_{\text{rec}} = 0.23$. Note that as $\ln N_\mu$ increases, the resolution improves, as expected. Thus, the reconstruction of the tail of $\ln N_\mu$ has the worst resolution, namely $\sigma_{\text{rec}} = 0.41$ for the physically accessible variable $M(450)$. The impact of this poor resolution will be evident when we discuss the reconstructed muon fluctuations.

The reconstructed distributions of $\ln N_\mu$, corresponding to $M(450)$ and $N_\mu^{\text{Hits}}(450)$, were built and the

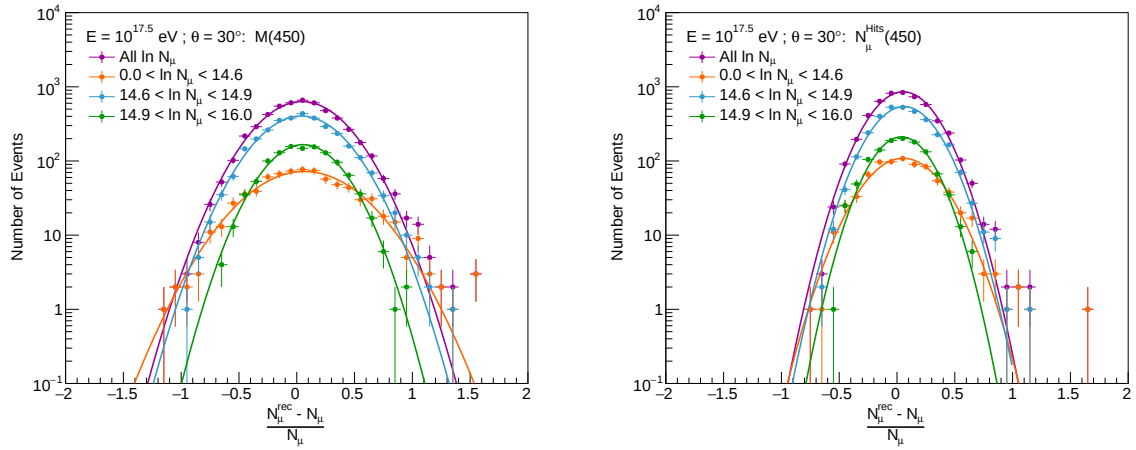

Figure 3.15: Gaussian fit to the distributions of the residuals of N_μ^{Rec} for different regions of the true distribution of $\ln N_\mu$: $\ln N_\mu < 14.6$ (orange), $14.6 < \ln N_\mu < 14.9$ (blue) and $14.9 < \ln N_\mu < 16.0$ (green), along with the residuals distribution over the entire ensemble of 5 105 unsaturated proton initiated events (purple) with $E_0 = 10^{17.5}$ eV and $\theta = 30°$. The left panel refers to the reconstruction based on $M(450)$ and the right panel on $N_\mu^{\text{Hits}}(450)$. Showers were simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

Variable	Region	μ	σ_{rec}	χ^2_{red}
$M(450)$	All $\ln N_\mu$	0.0458 ± 0.0046	0.3192 ± 0.0033	$37.59/25$
	$\ln N_\mu \in [0, 14.6]$	0.065 ± 0.016	0.405 ± 0.012	$20.61/24$
	$\ln N_\mu \in]14.6, 14.9]$	0.0388 ± 0.0057	0.3140 ± 0.0030	$26.63/20$
	$\ln N_\mu \in]14.9, 16.0]$	0.0554 ± 0.0082	0.2732 ± 0.0051	$9.32/14$
$N_\mu^{\text{Hits}}(450)$	All $\ln N_\mu$	0.0533 ± 0.0035	0.234 ± 0.023	$62.84/17$
	$\ln N_\mu \in [0, 14.6]$	0.053 ± 0.010	0.267 ± 0.069	$23.34/18$
	$\ln N_\mu \in [14.6, 14.9]$	0.0608 ± 0.0044	0.2330 ± 0.0030	$44.12/15$
	$\ln N_\mu \in]14.9, 16.0]$	0.0386 ± 0.0066	0.2113 ± 0.0041	$24.77/10$

Table 3.4: Values of the bias, μ, width, σ_{rec} and χ^2_{red} of the fit to the distributions of the residuals of N_μ^{rec} in different regions of the true distribution of $\ln N_\mu$, using $M(450)$ and $N_\mu^{\text{Hits}}(450)$ as reconstruction variables. The results refer to the ensemble of proton initiated events with primary energy $E_0 = 10^{17.5}$ eV and zenith angle $\theta = 30°$.

reconstructed slope of their low tails, Λ_μ^{rec}, was extracted by fitting the entire distribution to a convolution of an exponential with slope Λ_μ^{-1} and a Gaussian with mean μ and variance σ^2. A closed form of the fit function can be written as

$$\begin{aligned}\frac{\mathrm{d}\mathcal{N}}{\mathrm{d}\ln N_\mu} &\propto \int_{-\infty}^{\infty} \Theta(-x) \exp\left(\frac{x}{\Lambda_\mu}\right) \exp\left(-\frac{(\ln N_\mu - x - \mu)^2}{2\sigma^2}\right) \mathrm{d}x = \\ &= C_\mu \exp\left(\frac{\ln N_\mu - \mu}{\Lambda_\mu} + \frac{1}{2}\left(\frac{\sigma}{\Lambda_\mu}\right)^2\right) \operatorname{erfc}\left(\frac{\ln N_\mu - \mu}{\sqrt{2}\sigma} + \frac{\sigma}{\sqrt{2}\Lambda_\mu}\right),\end{aligned} \tag{3.32}$$

where $\Theta(x)$ is the Heaviside function, ensuring only the low tail of the distribution decays exponentially, $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ is the complementary error function and C_μ is an overall normalisation constant. The reason for preferring this parametrisation over the simpler exponential fit function used in Chapter 2 is two-fold. On the one hand, the smaller size of the ensemble of showers used in the reconstruction of

N_μ leads to larger fluctuations of the distribution of $\ln N_\mu$, preventing us from defining a good criterion to limit the range of the fit performed to the low tail of the mentioned distribution. On the other hand, the reconstructed $\ln N_\mu$ distributions are convolutions of the true $\ln N_\mu$ distribution with a Gaussian resolution induced by the reconstruction procedure, making it harder to isolate the tail. For coherence, the low tail of the true distribution of $\ln N_\mu$ was also fitted to Equation 3.32.

Figure 3.16 shows the true distribution of $\ln N_\mu$, obtained with EPOS-LHC, and the reconstructed distribution of $\ln N_\mu^{\text{rec}}$ using both $M(450)$ and $N_\mu^{\text{Hits}}(450)$, along with the values of Λ_μ and Λ_μ^{rec} for each case. The values of the relevant fit parameters, Λ_μ, μ, σ, along with the χ^2_{red} value, are presented in Table 3.5. The average and relative fluctuations of the true and reconstructed muon distributions, $\langle N_\mu \rangle$ and $\sigma(N_\mu)/\langle N_\mu \rangle$, are also shown in the same table.

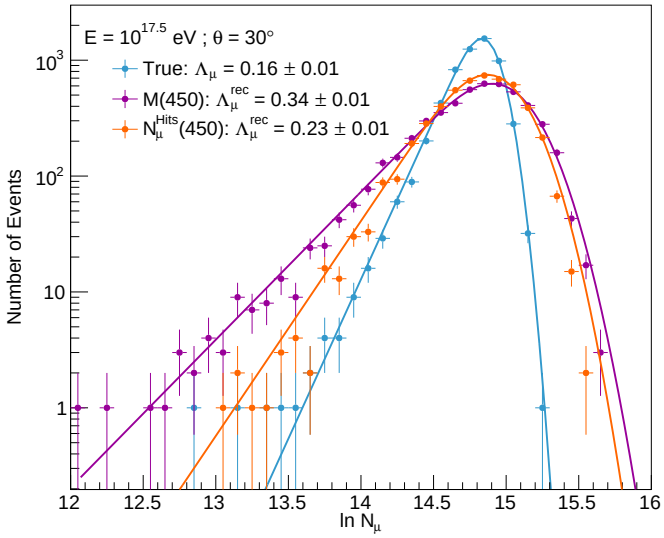

Figure 3.16: Fits to the true (blue) and reconstructed distributions of $\ln N_\mu$ from the linear conversion of $M(450)$ (purple) and $N_\mu^{\text{Hits}}(450)$ (orange) over an ensemble of 5 105 unsaturated proton induced events with primary energy $E_0 = 10^{17.5}$ eV and $\theta = 30°$. Showers were simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

Parameter	True N_μ	N_μ^{rec} $[M(450)]$	N_μ^{rec} $[N_\mu^{\text{Hits}}(450)]$
μ	14.9241 ± 0.0059	15.1008 ± 0.0062	15.0268 ± 0.0088
σ	0.1000 ± 0.0027	0.2188 ± 0.0056	0.2086 ± 0.0066
Λ_μ	0.1578 ± 0.0059	0.340 ± 0.012	0.235 ± 0.015
χ^2_{red}	$6258.28/18$	$3774.30/30$	$5318.47/22$
$\langle N_\mu \rangle$	2.632×10^6	$2.755 \pm 0.010 \times 10^6$	$2.7689 \pm 0.0080 \times 10^6$
$\sigma(N_\mu)/\langle N_\mu \rangle$	0.166	0.34	0.27

Table 3.5: Values of the relevant free parameters of the fit performed to the true and reconstructed muon distributions, along with the average N_μ and is relative fluctuations.

Evidently, the tail of the reconstructed $\ln N_\mu$ distribution using $M(450)$ is flatter than the tail of the true muon distribution, due to the poor resolution of the tail reconstruction. Consequently, the reconstructed slope $\Lambda_\mu^{\text{rec}} = 0.34 \pm 0.01$ is incompatible with the true slope at 18σ. Additionally, not only the average

value of N_μ^{rec} is about 5% larger than the true value, as a result of the positive reconstruction bias, but it is also incompatible with $\langle N_\mu \rangle$ at $\sim 12\sigma$. The relative fluctuations of the muon content of the showers are also overestimated, due to the large reconstruction resolution of $\sigma_{\text{rec}} = 0.34$. Importantly, even considering the ideal case where the number of muon hits on the RPCs could be measured, one obtains a reconstructed slope $\Lambda_\mu^{\text{rec}} = 0.23 \pm 0.01$, still incompatible with the true one at $\sim 5\sigma$, a positive bias of 5% on the average number of muons at the ground level, and an overestimation of the physical relative shower-to-shower fluctuations of the muon content, $0.27 > 0.17$. Thus, even in the absence of electromagnetic contamination, the reconstruction of the distribution of $\ln N_\mu$ for showers with $E_0 = 10^{17.5}$ eV and $\theta = 30^\circ$ using RPCs in the entire SD-750 array, does not allow for an unbiased measurement of Λ_μ. Moreover, the bias and resolution of the reconstruction prevent an accurate and precise determination of the first and second moments of the distribution of N_μ.

To improve the resolution of the reconstruction and verify the robustness of the reconstruction procedure, an ensemble of $\sim 1\,000$ proton initiated showers with $E_0 = 10^{18.5}$ eV and $\theta = 40^\circ$ was simulated with CORSIKA v7.7410, using the post-LHC hadronic interaction model EPOS-LHC. Each event was simulated 5 times, with different core positions, not only to avoid large statistical fluctuations, but also to maximise the probability of each shower having at least one corresponding unsaturated Auger event. The larger zenith angle was chosen to further mitigate the electromagnetic contamination, while keeping a reasonable effective RPC area. In turn, the higher primary energy was chosen to avoid the threshold effect noted in Figure 3.14 and improve the reconstruction resolution, while being low enough to prevent the saturation of the entire Infill array. Moreover, vertical events with higher energy and zenith angle trigger more stations, allowing for a better fit of the LDF. In fact, the average number of candidate stations increases from 4.4, for $E_0 = 10^{17.5}$ eV and $\theta = 30^\circ$, to 14.7 for a 10-fold increase in energy and $\theta = 40^\circ$.

The values of β_{M} and γ_{M}, as defined in Equation (3.15b), as well as the β and γ parameters of the LDF of the number of muon hits on the RPCs, were again extracted from the fits to their average LDFs, $\langle M(r) \rangle$ and $\langle N_\mu^{\text{Hits}}(r) \rangle$. The average LDFs can be found in Figure 3.17 and the values of the parameters can be read in Table 3.6.

Parameter	$\langle M(r) \rangle$	$\langle N_\mu^{\text{Hits}}(r) \rangle$
r_{opt}	450 m	450 m
r_{scale}	700 m	700 m
Value at 450 m	22.9 ± 2.1	27.5 ± 2.9
β	-0.45 ± 1.94	-0.42 ± 1.85
γ	-4.31 ± 6.09	-4.14 ± 3.82
χ^2_{red}	23.68/18	71.42/20

Table 3.6: Results of the χ^2-fit performed to the average LDF of the number of muon hits on the RPCs, $\langle N_\mu^{\text{Hits}}(r) \rangle$, and of the number of active fiducial pads of the RPCs, $\langle M(r) \rangle$.

Note that, indeed the values of the LDF parameters differ from those displayed in Table 3.2 and 3.3. Naturally, the values of $\langle M(450) \rangle$ and $\langle N_\mu^{\text{Hits}}(450) \rangle$ are larger than the ones found for showers with lower energies. At the same time, the average LDFs include many more stations and spread over a larger distance to the shower axis. Due to the increased number of shower particles, the ring of saturated stations around the shower axis is larger. Hence, the fit range of the LDF was changed to $r \in [300, 1\,600]$

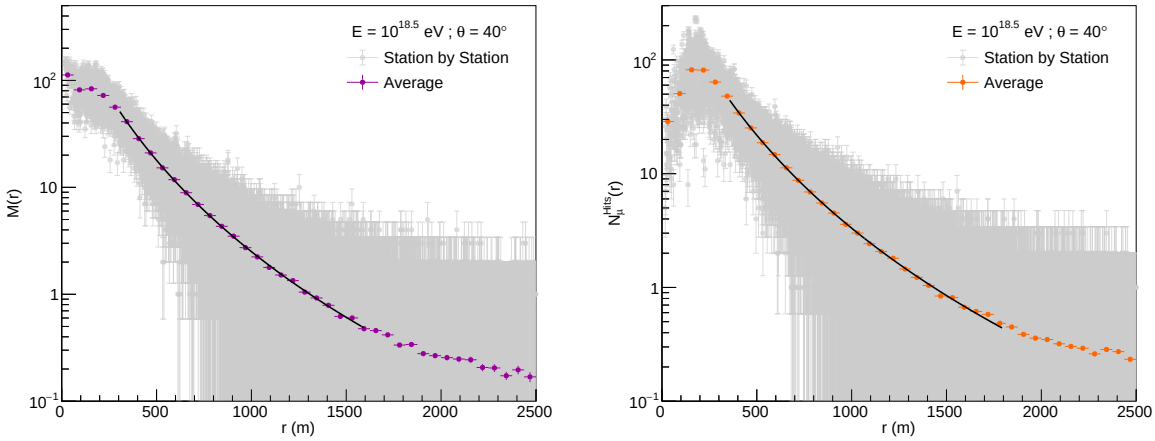

Figure 3.17: Left panel: χ^2-fit (black) to $\langle M(r)\rangle$ (purple), along with the number of active fiducial pads per station (grey dots). Right panel: χ^2-fit (black) to $\langle N_\mu^{\text{Hits}}(r)\rangle$ (orange), along with the number of muon hits on the RPCs per station (grey dots).

for $\langle M(r)\rangle$, and to $r \in [350, 1\,800]$ for $\langle N_\mu^{\text{Hits}}(r)\rangle$.

The values of $S(450)$, $M(450)$ and $N_\mu^{\text{Hits}}(450)$, along with their uncertainties were extracted from the event-by-event likelihood fit to the LDFs of the WCD signal, the number of active fiducial pads and number of muon hits on the RPCs, as described in Section 3.3.3, and plotted against the true values of N_μ. The obtained density plots were then fitted to a linear function of the form $y = m(N_\mu/10^5) + b$, using the least squares method. The results are shown in Figure 3.18. Events with saturated stations or abnormally large $\mathcal{L}_{\text{SD}}$ were discarded from the density plots, to ensure only quality events were considered in the reconstruction.

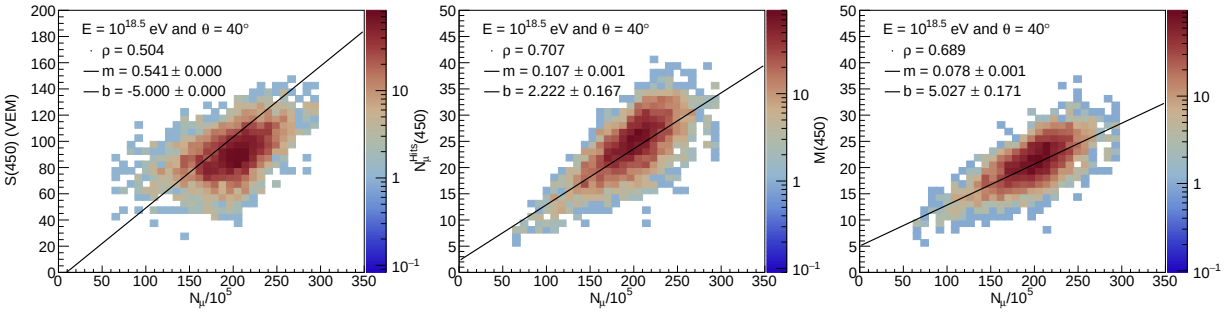

Figure 3.18: Left panel: Correlation between the distributions of $S(450)$ and N_μ. Middle panel: Correlation between the distributions of $N_\mu^{\text{Hits}}(450)$ and N_μ. Right panel: Correlation between the distributions of $M(450)$ and N_μ. Linear χ^2-fits of the form $y = m(N_\mu/10^5) + b$ are shown in red, and the fit parameters, along with the correlation factor, can be read off the top left of each panel. The ensemble of 5 105 showers with primary energy $E_0 = 10^{18.5}$ eV and $\theta = 40°$ was simulated with CORSIKA using EPOS-LHC, and the corresponding auger event simulated with Offline.

In all cases, the correlation between N_μ and the different reconstruction variables is stronger than the one obtained for the ensemble at lower primary energy and zenith angle. Once again, $S(450)$ is the least sensitive variable to fluctuations of N_μ, with $\rho \simeq 0.50$, since the WCD is the most exposed to the

electromagnetic component. The corresponding density plot as a more symmetrical shape around the fit curve, although the accumulation of events below this line still indicates that the fit to the LDF could be better tuned. Reinforcing this observation is the value of b for this plot. The correlation between $M(450)$ and N_μ is stronger with $\rho \simeq 0.69$, and the y-intercept of the linear regression is clearly incompatible with zero, being indicative of a residual electromagnetic contamination. Lastly, the correlation between $N_\mu^{\text{Hits}}(450)$ and the true N_μ distribution is the strongest, as expected, since this is the optimal case. However, the positive y-intercept may indicate some bias in the determination of $N_\mu^{\text{Hits}}(450)$.

Through Equation (3.28), the reconstructed number of muons, N_μ^{rec}, for each event was obtained from the values of $M(450)$ and $N_\mu^{\text{Hits}}(450)$. The residuals of N_μ^{rec} were computed using Equation (3.30), and the bias, μ, and resolution, σ_{rec}, of the reconstruction were extracted from a Gaussian fit to the distribution of r_μ over the ensemble of $\sim 3\,000$ events with unsaturated stations. The fit expression is given in Equation (3.31) and it was performed using the method of least squares. Once again, the bias and resolution where computed for different regions of the true $\ln N_\mu$ distribution: low tail $\ln N_\mu < 16.6$, bulk, $16.6 < \ln N_\mu < 16.9$ and high tail $16.9 < \ln N_\mu < 17.5$. The distributions of residuals are shown in Figure 3.19 and the fit parameters displayed in Table 3.7.

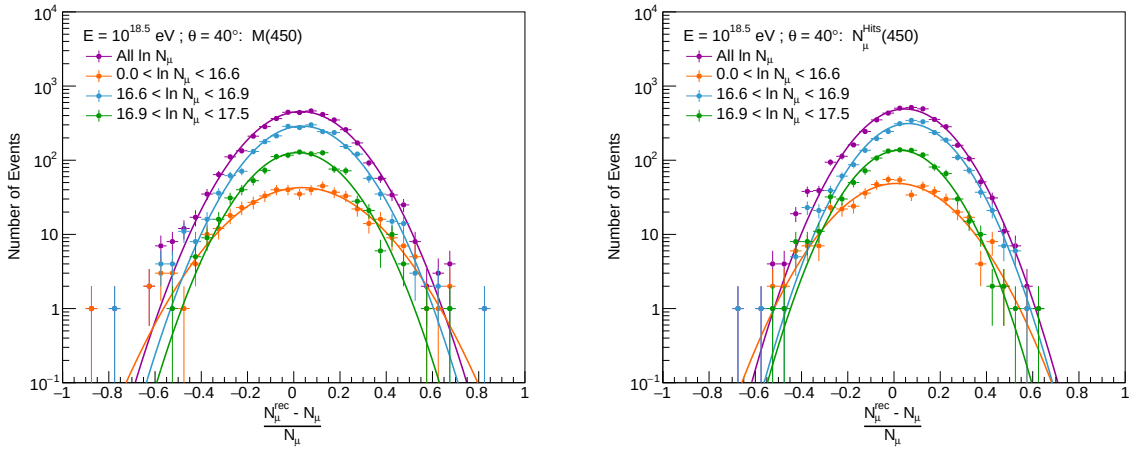

Figure 3.19: Gaussian fit to the distributions of the residuals of N_μ^{Rec} for different regions of the true distribution of $\ln N_\mu$: $\ln N_\mu < 16.6$ (orange), $16.6 < \ln N_\mu < 16.9$ (blue) and $16.9 < \ln N_\mu < 17.5$ (green), along with the residuals distribution over the entire ensemble of $\sim 3\,000$ unsaturated proton initiated events (purple) with $E_0 = 10^{18.5}$ eV and $\theta = 40°$. The left panel refers to the reconstruction based on $M(450)$ and the right panel on $N_\mu^{\text{Hits}}(450)$. Showers were simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

The obtained values of $\chi^2_{\text{red}} \sim 1$ show that the distributions of residuals are approximately Gaussian. However, the distribution of residuals for the tail region of the true distribution of the number of muons displays a tail towards low r_μ values, being indicative of a threshold effect. Notably, the bias and resolution of the reconstruction of N_μ decreased with respect to the values obtained for showers with $E_0 = 10^{17.5}$ eV and $\theta = 30°$. As before, the reconstruction of the tail of the distribution of N_μ has the worst resolution, $\sigma_{\text{rec}} = 0.21$ and $\sigma_{\text{rec}} = 0.17$, using $M(450)$ or $N_\mu^{\text{Hits}}(450)$, respectively. Since $M(450)$ is the experimentally accessible variable, we have an overall resolution induced by the reconstruction of

Variable	Region	μ	σ_{rec}	χ^2_{red}
$M(450)$	All $\ln N_\mu$	0.0315 ± 0.0028	0.1748 ± 0.0022	$41.66/27$
	$\ln N_\mu \in [0, 16.6]$	0.035 ± 0.010	0.2185 ± 0.0074	$16.775/25$
	$\ln N_\mu \in]16.6, 16.9]$	0.0363 ± 0.0035	0.1689 ± 0.0027	$37.45/24$
	$\ln N_\mu \in]16.9, 17.5]$	0.0187 ± 0.0052	0.1618 ± 0.0042	$27.01/20$
$N_\mu^{\text{Hits}}(450)$	All $\ln N_\mu$	0.0441 ± 0.0024	0.1472 ± 0.0020	$55.52/23$
	$\ln N_\mu \in [0, 16.6]$	0.0588 ± 0.0077	0.1650 ± 0.0057	$17.54/22$
	$\ln N_\mu \in [16.6, 16.9]$	0.0586 ± 0.0030	0.1404 ± 0.0024	$55.38/23$
	$\ln N_\mu \in]16.9, 17.5]$	0.0078 ± 0.0044	0.1372 ± 0.0035	$28.75/20$

Table 3.7: Values of the bias, μ, width, σ_{rec} and χ^2_{red} of the fit to the distributions of the residuals of N_μ^{rec} in different regions of the true distribution of N_μ, using $M(450)$ and $N_\mu^{\text{Hits}}(450)$ as reconstruction variables. The results refer to the ensemble of proton initiated events with primary energy $E_0 = 10^{18.5}$ eV and zenith angle $\theta = 40°$.

$\sigma_{\text{rec}} = 17\%$, while the bias is $\mu = 3.5\%$. Once again, N_μ is overestimated by the reconstruction method. In the best case scenario, where we could access $N_\mu^{\text{Hits}}(450)$, the resolution would be $\sigma_{\text{rec}} = 15\%$. Even so, the reconstruction of N_μ could be improved, to eliminate the positive bias observed. Namely, implementing algorithms to take into account any residual threshold effects.

The true and reconstructed distributions of $\ln N_\mu$ were built for the ensemble of $\sim 3\,000$ unsaturated events and fitted to Equation 3.32 to extract the value of Λ_μ. The values of the relevant fit parameters, Λ_μ, μ, σ, along with the χ^2_{red} value, are presented in Table 3.8. The average and relative fluctuations of the true and reconstructed muon distributions, $\langle N_\mu \rangle$ and $\sigma(N_\mu)/\langle N_\mu \rangle$, are also shown in the same table.

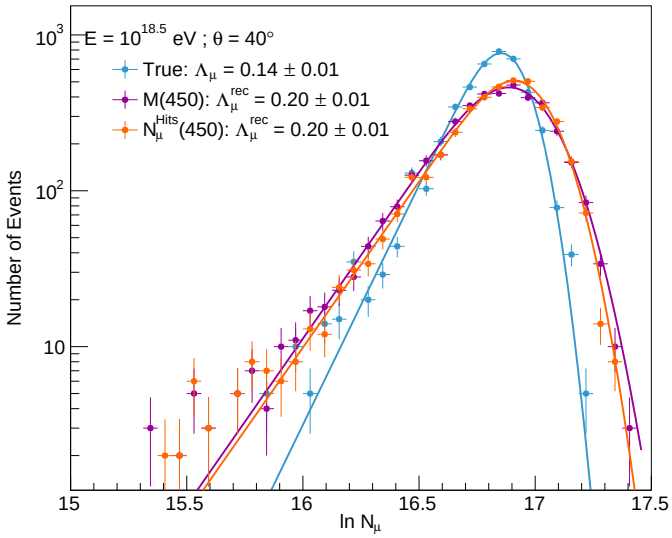

Figure 3.20: Fits to the true (blue) and reconstructed distributions of $\ln N_\mu$ from the linear conversion of $M(450)$ (purple) and $N_\mu^{\text{Hits}}(450)$ (orange) over an ensemble of $\sim 3\,000$ unsaturated proton induced events with primary energy $E_0 = 10^{18.5}$ eV and $\theta = 30°$. Showers were simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

Although the values of Λ_μ^{rec} for both reconstruction variables are closer to the true value of Λ_μ, they are still incompatible with it at 6σ. The fact that the tails of the reconstructed distribution of $\ln N_\mu^{\text{rec}}$ are flatter

Parameter	True N_μ	N_μ^{rec} $[M(450)]$	N_μ^{rec} $[N_\mu^{\text{Hits}}(450)]$
μ	16.9364 ± 0.0047	17.0134 ± 0.0066	17.0269 ± 0.0052
σ	0.0952 ± 0.0043	0.1552 ± 0.0053	0.1319 ± 0.0047
Λ_μ	0.1413 ± 0.0095	0.203 ± 0.012	0.204 ± 0.010
χ^2_{red}	$7462.10/19$	$4477.13/32$	$4437.89/29$
$\langle N_\mu \rangle$	1.992×10^7	$2.0480 \pm 0.0042 \times 10^7$	$2.0701 \pm 0.0032 \times 10^7$
$\sigma(N_\mu)/\langle N_\mu \rangle$	0.167	0.24	0.22

Table 3.8: Values of the relevant free parameters of the fit performed to the true and reconstructed muon distributions, along with the average N_μ and is relative fluctuations.

than that of the true distribution is the result of the Gaussian resolution induced by the reconstruction and its dependence on the value of N_μ. In particular, the resolution for the tail reconstruction is worse than that of the full distribution of $\ln N_\mu$. Furthermore, due to the reconstruction bias, the average value of N_μ obtained from $M(450)$ is overestimated by 3% and incompatible with the true $\langle N_\mu \rangle$ at 13σ. Even in the ideal case, the bias on the mean muon content is 4%, and the reconstructed $\langle N_\mu^{\text{rec}} \rangle$ is incompatible with the true one at 24σ. Although these discrepancies were improved by raising the primary energy and zenith angles, we conclude that the reconstruction of the distribution of N_μ using RPCs in the entire Infill array overestimates the mean muon number of EAS. The reconstructed relative shower-to-shower fluctuations of N_μ are about 24%, still larger than the physical shower-to-shower relative fluctuations of the muon content of EAS 17%. The improvement by using directly the muon hits on the RPCs, i.e., absence of electromagnetic contamination, is not significant.

In summation, we conclude that it is not possible to obtain an unbiased measurement of Λ_μ, using RPCs across the entire SD-750 array, even in the absence of electromagnetic contamination, and that there still exists a threshold effect hampering the reconstruction of the tail of the muon number distribution.

3.3.5 Reconstruction of the muon number distribution with an engineering array of 7 stations

So far, we have included all stations of the SD-750 array in the reconstruction of the muon number distribution. However, the MARTA engineering array is constituted by just 7 stations: a central one and the 6 stations of the closest ring, C_1. Thus, we must determine how the reconstructed muon distribution is changed when only the 7 stations of the engineering array are employed in the reconstruction. To do so, we choose the central station of the Infill array, with id 4001, and take it, along with the 6 first neighbours to form the MARTA engineering array. Then, using an ensemble of ~ 1000 CORSIKA showers with primary energy $E_0 = 10^{18.5}$ eV and $\theta = 40°$, for which the reconstruction had the best resolution and least bias, we considered two cases: for the first, the core was randomly simulated within a tile of $1.5 \times 1.5\,\text{km}^2$ around station 4001, while for the second case, we simulated the core within a tile of $3 \times 3\,\text{km}^2$ centred at the same station. The second case corresponds to a 4-fold increase of the exposure of the first one, and emulates a more realistic scenario where some events could have as little as 1 candidate MARTA station participating in the fit to the LDF. Both configurations, are shown in Figure 3.21, where MARTA

stations are shown in black and the remaining stations in grey. The position of each station is given with respect to station 4001.

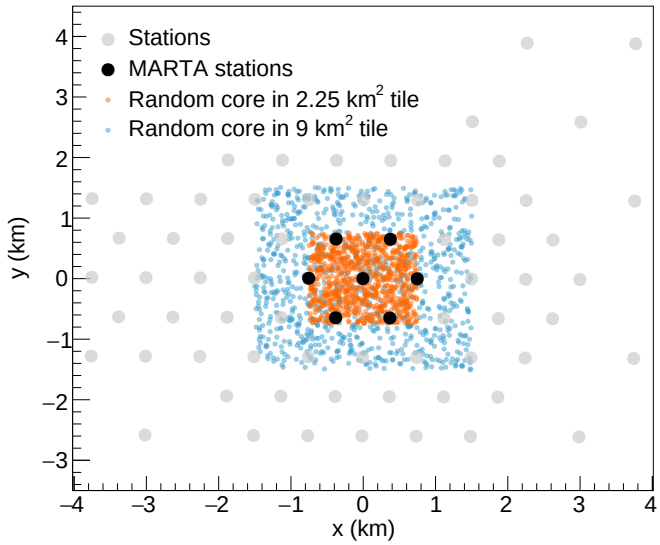

Figure 3.21: Stations of the MARTA engineering array, in black, and regular SD tanks, in grey. Blue dots represent the position of the core randomly simulated within a $9\,\mathrm{km}^2$ square, while orange dots represent the shower core within a square tile of $2.25\,\mathrm{km}^2$. The ensemble of $\sim 10^3$ showers with primary energy $E_0 = 10^{18.5}$ eV and $\theta = 40^\circ$ was simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

To extract the values of $M(450)$ and $N_\mu^{\mathrm{Hits}}(450)$ for each event, we performed a event-by-event fit to the LDF of the number of active fiducial pads and of the number of muon hits on the RPCs, as explained in Section 3.3.3. This time, we minimised the likelihoods of Equations (3.26) and (3.27), respectively, with the restriction that only stations of the MARTA engineering were considered. The shape parameters of the LDFs were taken from the χ^2-fits to the average LDFs, and their values can be found in Table 3.6.

The event-by-event values of $M(450)$ and $N_\mu^{\mathrm{Hits}}(450)$ were then linearly converted into N_μ^{rec} using Equation (3.28), with $\tilde{m}$ and b taken from the scatter plots between $M(450)/N_\mu^{\mathrm{Hits}}(450)$ and N_μ. Only events with unsaturated stations and with at least 1 station of the MARTA engineering array were considered. The distributions of $\ln N_\mu^{\mathrm{rec}}$ were then built over the aforementioned ensemble of CORSIKA showers. These distributions, along with the true distribution of $\ln N_\mu$ are shown in Figure 3.22, including the χ^2-fit to Equation (3.32). The fit parameters, including the slopes of the reconstructed tails, $\Lambda_\mu^{\mathrm{rec}}$, along with the average and relative fluctuations of N_μ are shown in Table 3.9, for both scenarios: the $2.25\,\mathrm{km}^2$ and $9\,\mathrm{km}^2$ tile of random core positions. The values of these parameters for the true muon distribution can be found in Table 3.8.

Additionally, the resolutions induced by the reconstruction using $M(450)$ and $N_\mu^{\mathrm{Hits}}(450)$ are $\sigma_{\mathrm{rec}} = 21\%$ and $\sigma_{\mathrm{rec}} = 18\%$, when the core is simulated inside the MARTA engineering array. This means that the resolution worsens by $\sim 25\%$ with respect to the one obtained when using the entire SD-750 array. However, the obtained values of $\Lambda_\mu^{\mathrm{rec}}$, $\langle N_\mu \rangle$ and $\sigma(N_\mu)/\langle N_\mu \rangle$ are very similar, whether one covers the entire Infill array with RPCs or just 7 stations in a hexagonal configuration, with one in the center,

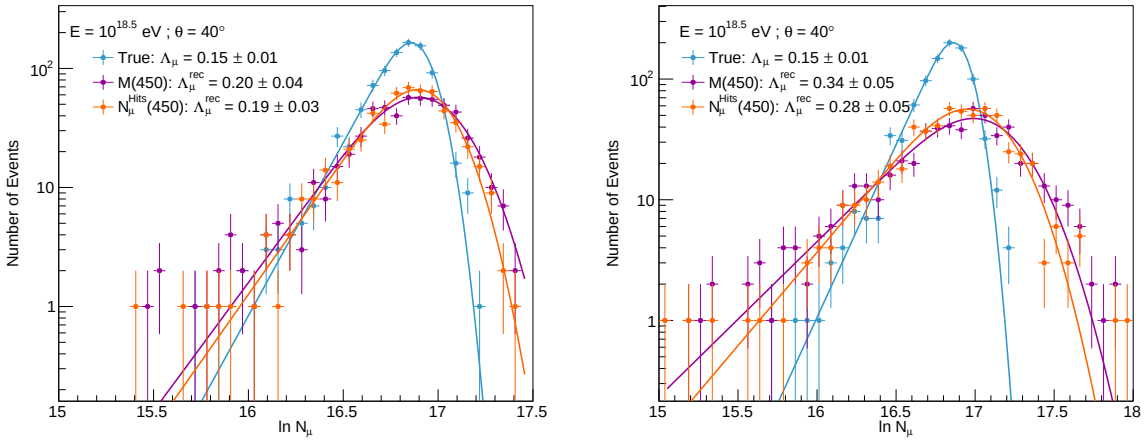

Figure 3.22: Fits to the true (blue) and reconstructed distributions of $\ln N_\mu$ from the linear conversion of $M(450)$ (purple) and $N_\mu^{\text{Hits}}(450)$ (orange), using only stations from the MARTA engineering array. Left panel: shower core within a tile of $2.25\,\text{km}^2$ in the center of the engineering array. Right panel: shower core within a tile of $9\,\text{km}^2$ in the center of the engineering array. The ensemble of $\sim 10^3$ unsaturated proton induced events with primary energy $E_0 = 10^{18.5}$ eV and $\theta = 30^\circ$ was simulated with CORSIKA, using EPOS-LHC, and the Auger event simulated with Offline.

Tile of core positions	Parameter	N_μ^{rec} $[M(450)]$	N_μ^{rec} $[N_\mu^{\text{Hits}}(450)]$
$2.25\,\text{km}^2$	μ	17.0035 ± 0.0022	17.014 ± 0.014
	σ	0.189 ± 0.016	0.152 ± 0.012
	Λ_μ	0.201 ± 0.037	0.192 ± 0.026
	χ^2_{red}	$332.46/26$	$354.81/24$
	$\langle N_\mu \rangle$	$2.089 \pm 0.013 \times 10^7$	$2.0824 \pm 0.0095 \times 10^7$
	$\sigma(N_\mu)/\langle N_\mu \rangle$	0.26	0.23
$9\,\text{km}^2$	μ	17.194 ± 0.022	17.128 ± 0.029
	σ	0.221 ± 0.022	0.210 ± 0.026
	Λ_μ	0.354 ± 0.047	0.291 ± 0.057
	χ^2_{red}	$324.05/31$	$553.97/28$
	$\langle N_\mu \rangle$	$2.275 \pm 0.035 \times 10^7$	$2.226 \pm 0.025 \times 10^7$
	$\sigma(N_\mu)/\langle N_\mu \rangle$	0.50	0.35

Table 3.9: Values of the relevant free parameters of the fit performed to the reconstructed muon distributions from the linear conversion of $M(450)$ and $N_\mu^{\text{Hits}}(450)$, along with the average N_μ and is relative fluctuations, using only stations of the MARTA engineering array.

provided that the core of the events lies inside the array. In this case, the shape of the LDF is mostly fixed by the stations of the engineering array, which are closer to the shower core. The remaining stations of the Infill play a sub-dominant role in fixing the shape of the LDF, mostly improving the overall resolution of the reconstruction and the bias in the moments of the distribution of N_μ. Once again, the tail reconstruction is not significantly improved by using directly the muon hits on the RPCs rather than the number of active fiducial pads, showing that the residual electromagnetic contamination is not the factor that dominates the quality of the reconstruction. Rather, the area of each of the RPCs should be increased, to boost its sensitivity to showers with low muonic content, within the Infill array.

On the other hand, when we consider events whose core can lie further way from the MARTA en-

gineering array, the quality of the reconstruction decreases considerably. In particular, the resolutions induced by the reconstruction are $\sigma_{\text{rec}} = 35\%$, and $\sigma_{\text{rec}} = 27\%$, from the linear conversion of $M(450)$ and $N_{\mu}^{\text{Hits}}(450)$, respectively. The value of $\Lambda_{\mu}^{\text{rec}}$ using $M(450)$ is 0.35 ± 0.05, which smaller and incompatible with the true tail slope at 4σ. Due to the very poor reconstruction resolution, the true relative fluctuations of N_{μ} are overestimated by 200%. The bias in the average value of N_{μ} is 15%, so $\left\langle N_{\mu}^{\text{rec}} \right\rangle$ is incompatible with the true value at 8σ. The results do not improve considerably when using the number of muon hits directly. These results are not surprising since this reconstruction includes events where very few (as little as 1) candidate stations of the MARTA engineering array contribute with low numbers of active fiducial pads, most, if not all of which, far from the shower core. Hence, the quality of the LDF fit worsens, and consequently the extracted value of $M(450)$, degrading its correlation with the true muon content of the shower.

We conclude that 7 stations of the MARTA engineering array are enough to reconstruct the distribution of the number of muons with a resolution similar to the one achieved by using the entire Infill array, provided we only consider events whose shower core lies inside the engineering array. At the same time, this equally means that we cannot get an unbiased estimation of the slope of the tail of the distribution of the number of muon, in either case. Note that the studied cases were ideal, in the sense that we have not considered systematic uncertainties from the reconstruction of the shower geometry and energy. Finally, we note that more sophisticated analysis may be implemented to take into account the verified threshold effects.

Chapter 4

Conclusions

In this work, we have shown that in proton initiated extensive air showers, the distribution of a modified fraction of the primary energy carried by hadronically interacting particles of the first generation, α_1, displays a low tail, which can be characterised by an exponential slope Λ_α. Similarly, the distribution of the number of muons in showers with low muon content can be characterised by a slope Λ_μ. We have demonstrated that Λ_μ is positively correlated with Λ_α, by re-sampling over an ensemble of proton initiated showers, and hence, that measurements of Λ_μ can be used to constrain the shape of the energy spectrum of hadrons in inclined showers with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67^\circ$. At a more fundamental level, we have established that the forward production cross-section of neutral pions arising from the first p-Air interaction can be characterised by a slope Λ_π, which can also be constrained by measurements of Λ_μ, in showers with low muon content. Importantly, we verified that these conclusions hold for vertical showers at equivalent center-of-mass energies of $p - p$ collisions at the LHC.

We showed that Λ_μ can still be accessed in mixed composition scenarios, since the dominant mass composition of showers with low muonic content is proton. Additionally, we concluded that, even in extreme composition scenarios, Λ_μ could be measured with enough precision, i.e, less than 20%, to distinguish between hadronic interaction models, within current experimental uncertainties, provided the number of events is large enough. In particular, for composition scenarios compatible with X_{max} measurements, we expect that about $5\,0000$ events guarantee a precise measurement of Λ_μ, if the distribution of the number of muons reaching the ground is corrected for its dependence on the primary energy.

Using the expected number of active fiducial pads of the MARTA engineering array at $r = 450\,\mathrm{m}$ from the shower core, we showed that the tail of the number of muons reaching the ground level in showers with $E_0 = 10^{17.5}$ eV and $\theta = 30^\circ$ can be reconstructed with a resolution of $\sigma_{\mathrm{rec}} = 41\%$, while the entire distribution is reconstructed with a resolution of $\sigma_{\mathrm{rec}} = 32\%$. As result, we obtain the value $\Lambda_\mu^{\mathrm{rec}} = 0.34 \pm 0.01$, which has a positive bias of 18σ with respect to the true $\Lambda_\mu = 0.16 \pm 0.01$ Moreover, $\langle N_\mu \rangle$ has a positive bias of 5% and the reconstructed muon fluctuations read 34%, while the physical true fluctuations read just 17%. Importantly, we showed that the resolution and bias induced by the

reconstruction of N_μ are not dominated by the residual electromagnetic contamination, but rather by a threshold effect.

By scaling the primary energy by a factor of 10 and setting $\theta = 40^\circ$, we demonstrated the robustness of the reconstruction procedure and improved the reconstruction resolution to $\sigma_{\text{rec}} = 17\%$. Once again, the resolution induced by the reconstruction of showers with low muonic content is larger, at $\sigma_{\text{rec}} = 22\%$, preventing an unbiased measurement of the slope of the tail of the distribution of the number of muons reaching the ground level, which reads $\Lambda_\mu^{\text{rec}} = 0.20 \pm 0.01$ while the true value is $\Lambda_\mu = 0.14 \pm 0.01$ The physical muon fluctuations are overestimated by 40% and the positive bias in the average value of N_μ is 13σ. Furthermore, we have showed that 7 stations of the MARTA engineering array with a minimal hexagonal configuration can reproduce the bias and resolutions achieved with the entire SD-750 array, provided the shower core falls within the engineering array.

In summation, the results show that, given the area of the RPCs, the resolution induced by the reconstruction procedure produces a bias in Λ_μ, even under ideal conditions, where the systematics of the geometry and energy reconstruction of the shower are not considered. Thus, it is necessary to introduce more sophisticated analysis which take into account the observed threshold effects and mitigate the bias in the estimator of the slope of the tail of the distribution of the number of muons. Finally, combining information of MARTA detectors with that of the corresponding WCDs could further improve our results and mitigate systematic uncertainties.

Bibliography

[1] A. Aab et al. “Observation of a large-scale anisotropy in the arrival directions of cosmic rays above 8×10^{18} eV”. In: *Science* 357.6357 (2017), pp. 1266–1270. eprint: https://science.sciencemag.org/content/357/6357/1266.full.pdf.

[2] R. Alves Batista et al. “Open Questions in Cosmic-Ray Research at Ultrahigh Energies”. In: *Frontiers in Astronomy and Space Sciences* 6 (2019), p. 23.

[3] A. Aab et al. “Measurement of the cosmic-ray energy spectrum above 2.5×10^{18} eV using the Pierre Auger Observatory”. In: *Phys. Rev. D* 102 (6 Sept. 2020), p. 062005.

[4] R. Aloisio. “Acceleration and propagation of ultra-high energy cosmic rays”. In: *Progress of Theoretical and Experimental Physics* 2017.12 (Nov. 2017). 12A102. eprint: https://academic.oup.com/ptep/article-pdf/2017/12/12A102/22075738/ptx115.pdf.

[5] K. Greisen. “End to the Cosmic-Ray Spectrum?” In: *Phys. Rev. Lett.* 16 (17 Apr. 1966), pp. 748–750.

[6] G. T. Zatsepin and V. A. Kuzmin. “Upper limit of the spectrum of cosmic rays”. In: *JETP Lett.* 4 (1966), pp. 78–80.

[7] L. Cazon. “Probing High-Energy Hadronic Interactions with Extensive Air Showers”. In: *PoS* ICRC2019 (2020), p. 005. arXiv: 1909.02962 [hep-ex].

[8] P. Zyla et al. “Review of Particle Physics”. In: *PTEP* 2020.8 (2020), p. 083C01.

[9] L. Cazon, R. Conceição, and F. Riehn. “Probing the energy spectrum of hadrons in proton air interactions at ultrahigh energies through the fluctuations of the muon content of extensive air showers”. In: *Physics Letters B* 784 (2018), pp. 68–76.

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

[11] J. Matthews. “A Heitler model of extensive air showers”. In: *Astroparticle Physics* 22.5 (2005), pp. 387–397.

[12] D. Heck, J. Knapp, J. N. Capdevielle, G. Schatz, and T. Thouw. “CORSIKA: A Monte Carlo code to simulate extensive air showers”. In: (Feb. 1998).

[13] M. Alekseeva et al. “Extensive Air Shower Simulation Program CONEX: Matching Monte Carlo and Numerical Methods”. In: *29th International Cosmic Ray Conference (ICRC29), Volume 7*. Vol. 7. International Cosmic Ray Conference. Jan. 2005, p. 139.

[14] A. Ferrari, P. R. Sala, A. Fasso, and J. Ranft. “FLUKA: A multi-particle transport code (Program version 2005)”. In: (Oct. 2005).

[15] T. Böhlen et al. “The FLUKA Code: Developments and Challenges for High Energy and Medical Applications”. In: *Nuclear Data Sheets* 120 (2014), pp. 211–214.

[16] S. A. Bass et al. “Microscopic models for ultrarelativistic heavy ion collisions”. In: *Prog. Part. Nucl. Phys.* 41 (1998), pp. 255–369. arXiv: nucl-th/9803035.

[17] I. Kawrakow et al. *EGSnrc toolkit for Monte Carlo simulation of ionizing radiation transport*. doi:10.4224/40001303 [release v2021]. 2000.

[18] H. Drescher, M. Hladik, S. Ostapchenko, T. Pierog, and K. Werner. “Parton-based Gribov–Regge theory”. In: *Physics Reports* 350.2 (2001), pp. 93–289.

[19] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko, and K. Werner. “EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider”. In: *Phys. Rev. C* 92 (3 Sept. 2015), p. 034906.

[20] S. Ostapchenko. “Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: QGSJET-II model”. In: *Phys. Rev. D* 83 (1 Jan. 2011), p. 014018.

[21] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser, and T. Stanev. “Hadronic interaction model Sibyll 2.3d and extensive air showers”. In: *Phys. Rev. D* 102.6 (2020), p. 063002. arXiv: 1912.03300 [hep-ph].

[22] A. Aab et al. “Direct measurement of the muonic content of extensive air showers between 2×10^{17} and 2×10^{18} eV at the Pierre Auger Observatory”. In: *The European Physical Journal C* 80.8 (2020), p. 751.

[23] A. Aab et al. “Depth of maximum of air-shower profiles at the Pierre Auger Observatory. II. Composition implications”. In: *Phys. Rev. D* 90 (12 Dec. 2014), p. 122006.

[24] K.-H. Kampert and M. Unger. “Measurements of the cosmic ray composition with air shower experiments”. In: *Astroparticle Physics* 35.10 (2012), pp. 660–678.

[25] A. Aab et al. “Measurement of the Fluctuations in the Number of Muons in Extensive Air Showers with the Pierre Auger Observatory”. In: *Phys. Rev. Lett.* 126 (15 Apr. 2021), p. 152002.

[26] L. A. Anchordoqui, H. Goldberg, and T. J. Weiler. “Strange fireball as an explanation of the muon excess in Auger data”. In: *Phys. Rev. D* 95.6 (2017), p. 063005. arXiv: 1612.07328 [hep-ph].

[27] S. Coleman and S. L. Glashow. “High-energy tests of Lorentz invariance”. In: *Phys. Rev. D* 59 (11 Apr. 1999), p. 116008.

[28] A. Aab et al. “The Pierre Auger Cosmic Ray Observatory”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 798 (2015), pp. 172–213.

[29] A. Castellina. “AugerPrime: the Pierre Auger Observatory Upgrade”. In: *EPJ Web Conf.* 210 (2019). Ed. by I. Lhenry-Yvon, J. Biteau, O. Biteau, and P. Ghia, p. 06002. arXiv: 1905.04472 [astro-ph.HE].

[30] Hörandel, Jörg R. and for the Collaboration. “Precision measurements of cosmic rays up to the highest energies with a large radio array at the Pierre Auger Observatory”. In: *EPJ Web Conf.* 210 (2019), p. 06005.

[31] A. Aab et al. “Design and implementation of the AMIGA embedded system for data acquisition”. In: *Journal of Instrumentation* 16.07 (July 2021), T07008.

[32] A. Aab et al. “Extraction of the muon signals recorded with the surface detector of the Pierre Auger Observatory using recurrent neural networks”. In: *Journal of Instrumentation* 16.07 (July 2021), P07016.

[33] P. Abreu et al. “MARTA: a high-energy cosmic-ray detector concept for high-accuracy muon measurement”. In: *The European Physical Journal C* 78.4 (2018), p. 333.

[34] L. Cazon, R. Concei ção, M. A. Martins, and F. Riehn. “Constraining the energy spectrum of neutral pions in ultra-high-energy proton-air interactions”. In: *Phys. Rev. D* 103 (2 Jan. 2021), p. 022001.

[35] L. Cazon, R. Conceição, and F. Riehn. *An opportunity to measure the high-x region of π^0 production in UHE proton-Air interactions with Auger*. Auger internal note GAP-2019-001. 2019.

[36] F. Riehn. “Measurement of the fluctuations in the number of muons in inclined air showers with the Pierre Auger Observatory”. In: *PoS* ICRC2019 (2019), p. 404.

[37] A. Aab et al. “Search for photons with energies above 1018eV using the hybrid detector of the Pierre Auger Observatory”. In: *Journal of Cosmology and Astroparticle Physics* 2017.04 (Apr. 2017), pp. 009–009.

[38] L. Cazon, R. Conceição, M. Martins, and F. Riehn. “Probing the high energy spectrum of neutral pions in ultra-high energy proton-Air interactions”. In: *PoS* ICRC2019 (2020), p. 226. arXiv: 1908.09668 [hep-ph].

[39] O. Adriani et al. “Measurements of longitudinal and transverse momentum distributions for neutral pions in the forward-rapidity region with the LHCf detector”. In: *Phys. Rev. D* 94 (3 Aug. 2016), p. 032007.

[40] S. Argiro et al. “The Offline Software Framework of the Pierre Auger Observatory”. In: *Nucl. Instrum. Meth. A* 580 (2007), pp. 1485–1496. arXiv: 0707.1652 [astro-ph].

[41] R. Brun and F. Rademakers. “ROOT — An object oriented data analysis framework”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 389.1 (1997). New Computing Techniques in Physics Research V, pp. 81–86.

[42] S. Agostinelli et al. “Geant4—a simulation toolkit”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 506.3 (2003), pp. 250–303.

[43] J. Abraham et al. “Trigger and Aperture of the Surface Detector Array of the Pierre Auger Observatory”. In: *Nucl. Instrum. Meth. A* 613 (2010), pp. 29–39. arXiv: 1111.6764 [astro-ph.IM].

[44] P. Abreu et al. “The energy spectrum of cosmic rays beyond the turn-down around 10^{17} eV as measured with the surface detector of the Pierre Auger Observatory”. In: (Sept. 2021). arXiv: 2109.13400 [astro-ph.HE].

[45] L. Cazon et al. *Expected performance of MARTA in the measurement of muons*. Auger internal note GAP-2013-018. 2013.

[46] A. Supanitsky et al. “Underground muon counters as a tool for composition analyses”. In: *Astroparticle Physics* 29.6 (2008), pp. 461–470.

[47] K. Kamata and J. Nishimura. “The Lateral and the Angular Structure Functions of Electron Showers”. In: *Progress of Theoretical Physics Supplement* 6 (Feb. 1958), pp. 93–155. eprint: https://academic.oup.com/ptps/article-pdf/doi/10.1143/PTPS.6.93/5270594/6-93.pdf.

[48] A. Schulz. *Measurement of the Energy Spectrum of Cosmic Rays between 0.1 EeV and 30 EeV with the Infill Extension of the Surface Detector of the Pierre Auger Observatory*. Auger internal note GAP-2012-136. 2012.

[49] D. Mockler. “Measurement of the Cosmic Ray Spectrum with the Pierre Auger Observatory”. PhD thesis. Karlsruher Institut für Technologie (KIT), 2019. 173 pp.

[50] M. Aglietta et al. *Recovery of Saturated Signals of the Surface Detector*. Auger internal note GAP-2008-030. 2008.

[51] J. v. Buren et al. “Muon Size Spectrum measured by KASCADE-Grande”. In: *Proc. of the 29th Int. Cosmic Ray Conf., Pune*. Vol. 6. 2005, pp. 301–304.

[52] M. Ave, M. Roth, and A. Schulz. “A generalized description of the time dependent signals in extensive air shower detectors and its applications”. In: *Astroparticle Physics* 88 (2017), pp. 46–59.

[53] D. Góra et al. “Muon number rescaling in simulations of air showers”. In: *PoS* ICRC2021 (2021), p. 207. arXiv: 2108.07527 [astro-ph.HE].

[54] H. Fesefeldt. “The Simulation of Hadronic Showers: Physics and Applications”. In: (Dec. 1985).

Appendix A

Extra plots for QGSJET II-04 and SIBYLL 2.3d

A.1 Correlation between α_1 and N_μ

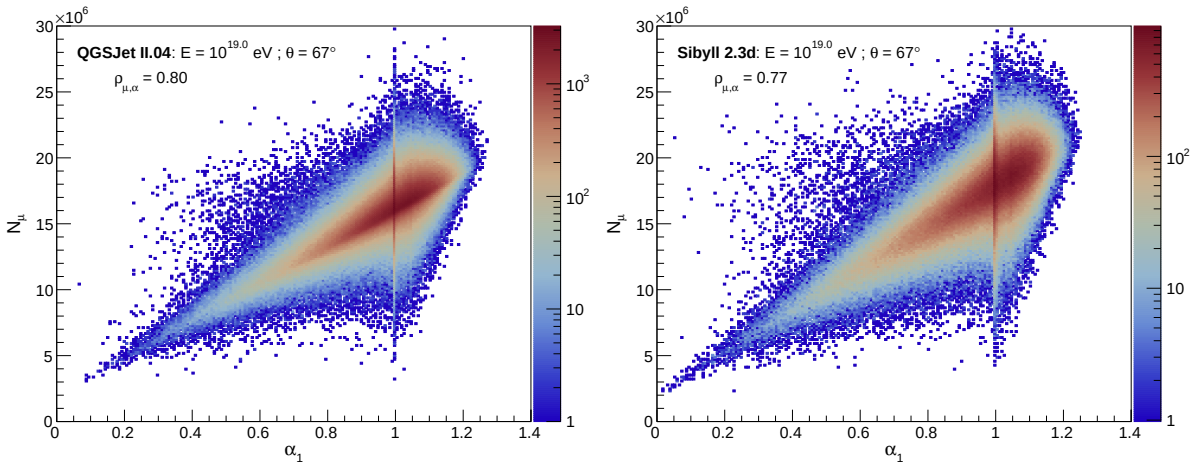

Figure A.1: Correlation between the distributions of α_1 and N_μ over an ensemble of $\sim 10^6$ proton induced EAS, with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Simulations were performed with CONEX, using the hadronic interaction model QGSJET II-04 (left panel) and SIBYLL 2.3d (right panel). The correlation factor between the two quantities can be read in the upper left corner.

A.2 Distributions of α_1 and $\ln N_\mu$

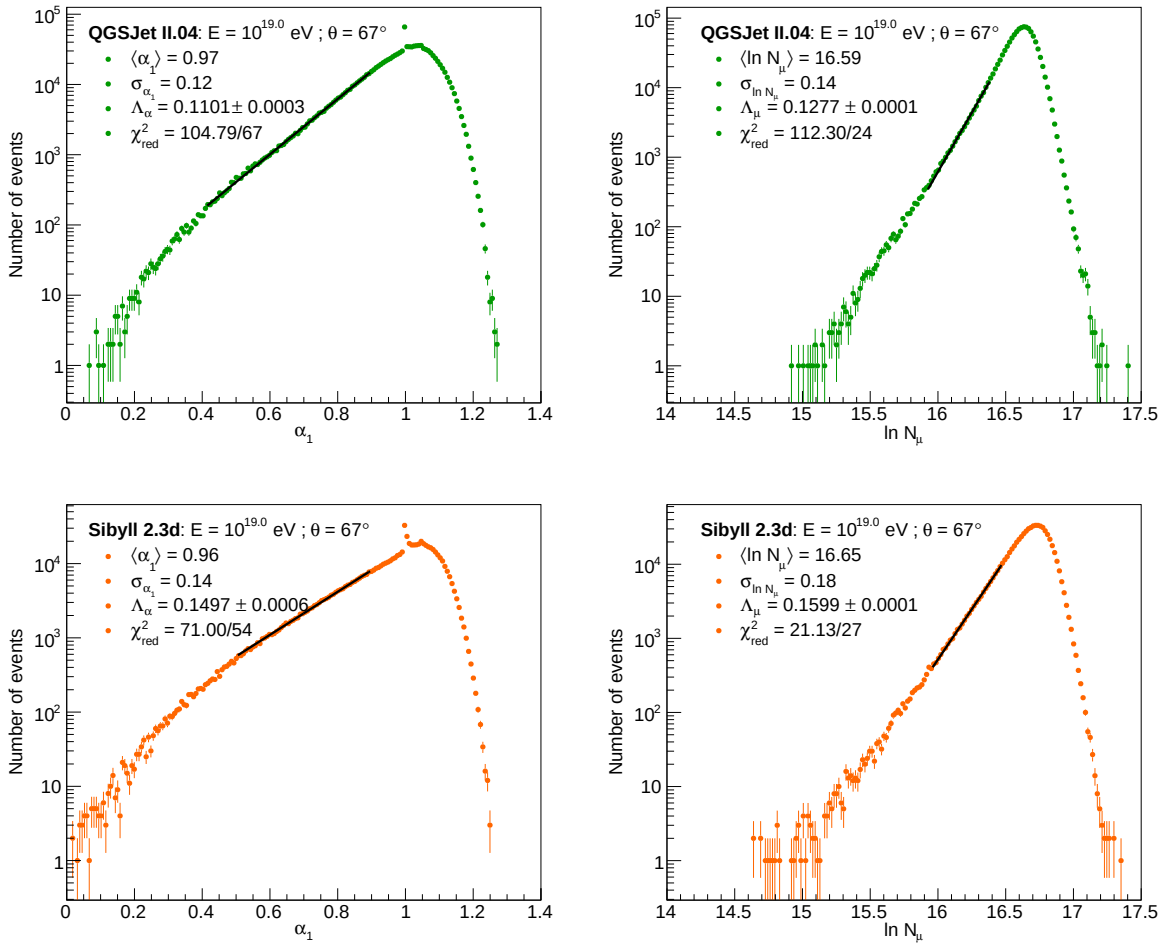

Figure A.2: Fit to the tail of the α_1 (left panels) and $\ln N_\mu$ (right panels) distributions over an ensemble of $\sim 10^6$ proton induced EAS, with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Simulations were performed with CONEX, using QGSJET II-04 (upper panels) and SIBYLL 2.3d (lower panels), and the fit curve is shown in black.

A.3 Correlation between x_L and N_μ

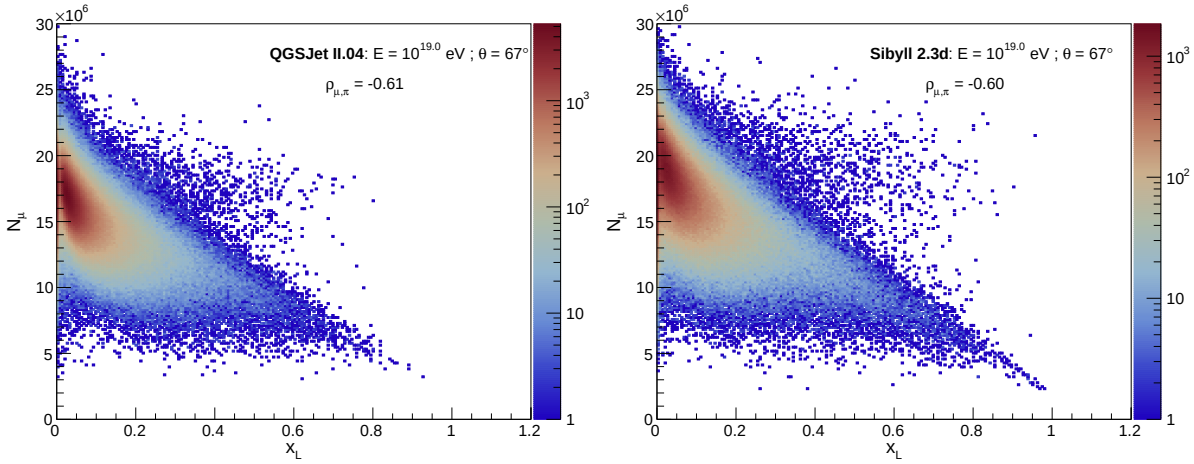

Figure A.3: Correlation between the distributions of x_L and N_μ over an ensemble of $\sim 10^6$ proton induced EAS, with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Simulations were performed with CONEX, using the hadronic interaction model QGSJET II-04 (left panel) and SIBYLL 2.3d (right panel). The correlation factor between the two quantities can be read in the upper left corner.

A.4 Distribution of x_L weighted by x_L

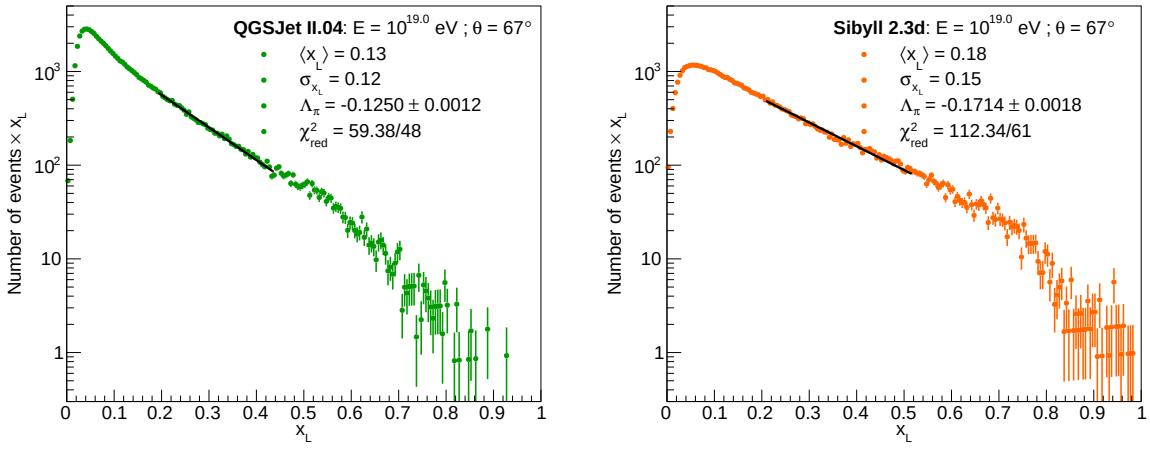

Figure A.4: Fit to the tail of the x_L weighed by x_L distributions over an ensemble of $\sim 10^6$ proton induced EAS, with primary energy $E_0 = 10^{19}$ eV and zenith angle $\theta = 67°$. Simulations were performed with CONEX, using QGSJET II-04 (left panel) and SIBYLL 2.3d (right panel), and the fit curve is shown in black.

Appendix B

Comparison between SIBYLL 2.3c and SIBYLL 2.3d

Since the submission of [34] the SIBYLL version was updated from 2.3c to 2.3d. As a result, we have chosen to use the latest version in this text. Although the qualitative results published in [34] still hold, a brief comparison between SIBYLL 2.3c and SIBYLL 2.3d is shown here for the sake of completeness.

B.1 Distributions

Figure B.1 shows the probability distributions of α_1, $\ln N_\mu$ and x_L weighted by x_L over an ensemble of $\sim 10^6$ showers for SIBYLL 2.3c and SIBYLL 2.3d(orange). Additionally, Table B.1 displays the values of $\langle \ln N_\mu \rangle$, $\sigma_{\ln N_\mu}$, Λ_μ, $\langle \alpha_1 \rangle$, σ_{α_1}, Λ_α, $\langle x_L \rangle$, σ_{x_L} and Λ_π for SIBYLL 2.3c and SIBYLL 2.3d. Note that the probability density function used to extract the moments of x_L is $x_\mathrm{L}\dfrac{\mathrm{d}\mathcal{N}}{\mathrm{d}x_\mathrm{L}}$.

Parameters	SIBYLL 2.3d	SIBYLL 2.3c
$\langle \alpha_1 \rangle$	0.96	0.97
σ_{α_1}	0.14	0.14
Λ_α	0.1497 ± 0.0006	0.1490 ± 0.0005
$\langle \ln N_\mu \rangle$	16.65	16.69
$\sigma_{\ln N_\mu}$	0.18	0.18
Λ_μ	0.1599 ± 0.0001	0.1623 ± 0.0001
$\langle x_\mathrm{L} \rangle$	0.18	0.19
σ_{x_L}	0.15	0.16
Λ_π	-0.1714 ± 0.0018	-0.1730 ± 0.0017

Table B.1: Values of $\langle \ln N_\mu \rangle$, $\sigma_{\ln N_\mu}$, Λ_μ, $\langle \alpha_1 \rangle$, σ_{α_1}, Λ_α, $\langle x_L \rangle$, σ_{x_L} and Λ_π for an ensemble of $\sim 10^6$ of proton induced EAS with $E_0 = 10^{19}$ eV and $\theta = 67^\circ$ simulated with CONEX, for SIBYLL 2.3c and SIBYLL 2.3d.

Qualitatively speaking, the plotted distributions with both models don't differ significantly from each other, although the energy spectrum of the leading neutral pions is harder for SIBYLL 2.3c, and the $\ln N_\mu$ is shifted toward lower values for SIBYLL 2.3d. The values of Λ_α and Λ_π of each model are compatible at 1σ, but the value of Λ_μ is not. The origin of this changes from SIBYLL version 2.3c to 2.3d is out of

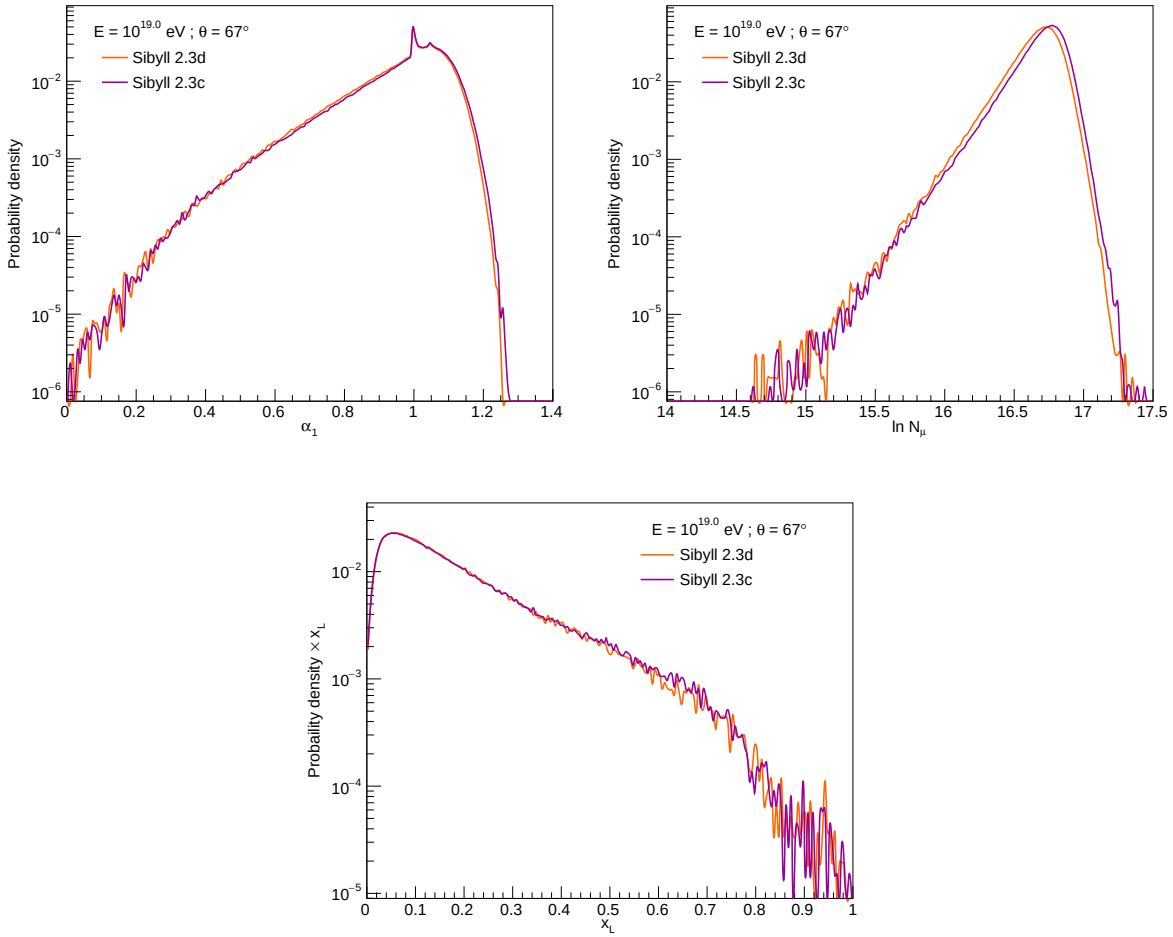

Figure B.1: Distributions of α_1 (top left panel), $\ln N_\mu$ (top right panel) and x_L weighted by x_L (bottom panel) over an ensemble of $\sim 10^6$ proton initiated showers with $E_0 = 10^{19}$ eV and $\theta = 67°$ simulated by CONEX with the hadronic interaction models SIBYLL 2.3c (purple) and SIBYLL 2.3d (orange).

the scope of this work. However, it can be due to some difference in the first interaction that does not affect Λ_α and Λ_π within the fitting range, but outside of it or/and that mainly affects other multiparticle production properties that change the slope of Λ_μ. It may also be due to cumulative differences in deeper shower generations, which do not affect much the first interaction, but affect the distribution of $\ln N_\mu$.

B.2 Conversion curves

Figure B.2 shows the conversion curves between Λ_μ and Λ_α (left panel) and between Λ_μ and Λ_π (right panel).

Regarding the $\Lambda_\mu \leftrightarrow \Lambda_\alpha$ conversion curves, its clear that the different versions yield very similar conversion curves, as manifested by the values of $\delta_{\Lambda_\alpha} \sim 0.04$. On the other hand, the calibration curves $\Lambda_\mu \leftrightarrow \Lambda_\pi$ parametrised by $\delta\Lambda_\mu$ differ more significantly. The origin of these difference is out of the scope of this work.

At any rate, these changes do not hamper the qualitative results obtained when choosing a SIBYLL

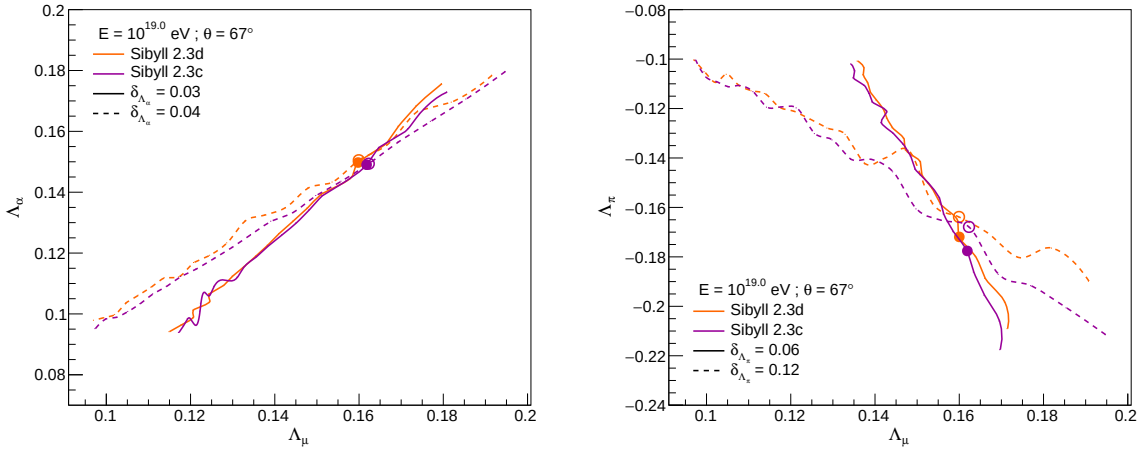

Figure B.2: Left panel: Conversion curves between Λ_μ and Λ_α. Right panel: Conversion curves between Λ_μ and Λ_π. Both sets of curves were obtained with CONEX simulations with the hadronic interaction models SIBYLL 2.3c(purple) and SIBYLL 2.3d(orange). The solid lines show how Λ plotted in the x axis is changed when Λ in the y axis changes, while the dashed lines represent the converse situation. Solid dots and circumferences represent the nominal values.

version over the other.

Appendix C

`ModuleSequence.xml.in` file with the reconstruction pipeline

```
<!DOCTYPE sequenceFile [
<!ENTITY % sd SYSTEM "@CONFIGDIR@/standardSdSequences.dtd">
%sd;
] >
<sequenceFile>
<enableTiming/>
<moduleControl>

<loop numTimes="unbounded" pushEventToStack="yes">
<module> EventFileReaderOG </module>
<loop numTimes="1" pushEventToStack="yes">

<module> EventGeneratorOG </module>
<module> SdAccidentalInjectorKG          </module>
<module> G4StationSimulatorOG            </module>

&SdSimulation;

<module> RPCSimulator                    </module>
<module> RPCElectronicsSimulator         </module>
<module> ClearParticleLists              </module>
<module> CentralTriggerSimulatorXb       </module>
<module> CentralTriggerEventBuilderOG    </module>
<module> RPCEventBuilder                 </module>
<module> EventBuilderOG                  </module>
<module> EventCheckerOG                  </module>
<module> SdCalibratorOG                  </module>
<module> EventCheckerOG </module>
<module> SdCalibratorOG </module>
<module> SdSignalRecoveryKLT </module>
<module> SdMonteCarloEventSelectorOG </module>
```

```
<module> SdEventSelectorOG </module>
<module> SdPlaneFitOG </module>
<module> LDFFinderKG </module>
<try>
<module> ScintillatorLDFFinderKG </module>
</try>
<module> Risetime1000LLL </module>
<module> SdEventPosteriorSelectorOG </module>

<module> MARTALDFFinder  </module>
<module> MARTAAnalysis  </module>

<module> EventFileExporterOG     </module>
<module> RecDataWriterNG                 </module>

</loop>
</loop>
</moduleControl>
</sequenceFile>
```