INTERNATIONAL DOCTORAL SCHOOL OF THE USC

Miguel Alexandre
Jesus da Silva Martins

PhD Thesis

Inferring hadronic interaction properties from the 2D distribution of muon content and depth of maximum of extensive air showers

Santiago de Compostela, 2025

Doctoral Programme in Nuclear and Particle Physics

ESCOLA DE DOUTORAMENTO
INTERNACIONAL DA USC

DOCTORAL THESIS

INFERRING HADRONIC INTERACTION PROPERTIES FROM THE 2D DISTRIBUTION OF MUON CONTENT AND DEPTH OF MAXIMUM OF EXTENSIVE AIR SHOWERS

Author

Miguel Alexandre Jesus da Silva Martins

Supervisors: Dr. Lorenzo Cazon Boado and Prof. Ruben Maurício da Silva Conceição

Tutor: Dr. Lorenzo Cazon Boado

PHD PROGRAMME IN NUCLEAR AND PARTICLE PHYSICS

SANTIAGO DE COMPOSTELA

"Compreendi, então, que um homem que houvesse vivido um único dia poderia sem custo passar cem anos numa prisão", Albert Camus em "O Estrangeiro"

Miguel Alexandre Jesus da Silva Martins acknowledges that the project that gave rise to these results received the support of a fellowship from “la Caixa” Foundation (ID 100010434). The fellowship code is LCF/BQ/DI21/11860033. The work was supported by the María de Maeztu grant CEX2023-001318-M funded by MICIU/AEI /10.13039/501100011033

Summary

As ultra-high-energy cosmic rays collide with nuclei in the Earth's atmosphere—at center-of-mass energies exceeding by more than an order of magnitude those achievable at current human-made colliders—they initiate vast cascades of secondary particles known as Extensive Air Showers.

These showers evolve through the re-interaction and propagation of secondary particles originating from hadronic interactions that cannot be calculated from first principles of Quantum Chromodynamics. Instead, these interactions are described by phenomenological hadronic interaction models that are poorly constrained by accelerator data. The uncertainty in the extrapolations of these models hampers the interpretation of the mass composition of the cosmic-ray flux, thereby hindering efforts to identify the sources of ultra-high-energy cosmic rays.

This thesis explores the link between the primary cosmic-ray–air interaction and the joint distribution of two key shower observables: the depth of the shower maximum, X_{max}, and the number of muons at ground level, N_μ. It culminates in a preliminary estimation of hadronic production characteristics in the primary interaction, derived from published data by the Pierre Auger Observatory.

We begin by decoupling the influence of hadron production stochasticity on the joint distribution of N_μ and X_{max} from other factors that shape it, including muon attenuation effects dependent on X_{max} and shower geometry, the composition and energy of the primary cosmic ray, and the energy threshold for muon detection in typical air-shower experiments. We demonstrate that the shape of the muon number distribution in muon-poor showers evolves as a function of X_{max}, and that this evolution is highly sensitive to the details of the primary interaction. Furthermore, we show that this feature is dominated by proton-induced showers and is measurable within the experimental precision of current air-shower observatories, despite the presence of heavier primaries in the cosmic-ray flux.

A new set of multiparticle production variables characterising the primary interaction is introduced by extending existing semi-analytic models. The variability and correlations among these variables are shown to drive fluctuations in both N_μ and X_{max}, enabling a physical interpretation of different regions of the N_μ–X_{max} distribution in terms of the properties of the primary interaction.

The derived set of production variables is then used to define a probabilistic mapping between the properties of the cosmic-ray-air interaction and the joint distribution of X_{max} and N_μ. The developed framework, formulated for proton-induced showers, allows any physically consistent prior on production variables to be propagated into predictions of the N_μ–X_{max} distribution, with systematic biases significantly smaller than the spread among hadronic interaction models and the typical experimental uncertainties. Such high-precision mapping can then be used to constrain the energy spectrum of hadrons produced in proton–air interactions in regions of kinematic phase space far beyond the reach of current accelerator experiments.

Finally, mean values of the proposed hadronic production variables were inferred using data published by the Pierre Auger Observatory, representing a promising first application of the framework developed in this thesis.

Published content in peer-reviewed journals

[1] Proton-air interactions at ultra-high energies in muon-depleted air showers with different depths

Lorenzo Cazon[1], Ruben Conceição[2,3], **Miguel Alexandre Martins**[1], Felix Riehn[1]

[1] *Instituto Galego de Física de Altas Enerxías (IGFAE), University of Santiago de Compostela, Rúa de Xoaquín Díaz de Rábago, Santiago de Compostela, Spain*
[2] *Laboratório de Instrumentação e Física Experimental de Partículas (LIP), Lisbon, Portugal*
[3] *Instituto Superior Técnico (IST), Universidade de Lisboa, Lisbon, Portugal*

Published in: *Phys.Lett.B* **859** (2024) 139115
ISSN: 03702693 (online)
DOI: https://doi.org/10.1016/j.physletb.2024.139115
Impact factor: 4.5 (2024)
SJR ranking: 1.383 (Q1)
Published by: Elsevier B.V.

Chapter: Certain contents of this publication are reproduced in Chapter 3.
Contribution of Ph.D. candidate: processed the Monte Carlo simulations used to extract the primary interaction quantities and shower observables needed for the study. Implemented suggestions and ideas of the co-authors and proactively conducted a significant part of the analyses. Produced all the figures.

[2] Probabilistic mapping between multiparticle production variables and the depth of maximum in proton-induced extensive air showers

Lorenzo Cazon[1], Ruben Conceição[2,3], **Miguel Alexandre Martins**[1], Felix Riehn[4]

[1] *Instituto Galego de Física de Altas Enerxías (IGFAE), University of Santiago de Compostela, Rúa de Xoaquín Díaz de Rábago, Santiago de Compostela, Spain*
[2] *Instituto Superior Técnico (IST), Universidade de Lisboa, Lisbon, Portugal*
[3] *Laboratório de Instrumentação e Física Experimental de Partículas (LIP), Lisbon, Portugal*
[4] *Technische Universität Dortmund, August-Schmidt-Straße 4, 44221 Dortmund, Germany*

Published in: *Phys. Rev. D* - 14 August, 2025
ISSN: 2470-0029 (online)
DOI: https://doi.org/10.1103/1664-pht6
Impact factor: 5.3 (2025)
SJR ranking: 1.458 (Q1)
Published by: American Physical Society.
Copyright permissions: Attached in Appendix E
Chapter: Certain contents of this publication are reproduced in Chapter 4.
Contribution of Ph.D. candidate: idealised and performed the semi-analytic calculations employed in the derivation of the central model of the publication. Processed the Monte Carlo simulations used to extract the primary interaction quantities and shower observables needed for the study. Produced all the figures.

Contents

1

INTRODUCTION

1.1 Ultra-high-energy cosmic rays

Ultra-high-energy cosmic rays (UHECRs) are protons or charged nuclei accelerated to energies, $E_0 > 1 \times 10^{18}\,\text{eV}$. The sources of these extremely energetic particles remain unknown [1], although large-scale anisotropies observed in their flux above $8 \times 10^{18}\,\text{eV}$ at Earth favour an extragalactic origin [2]. After being accelerated, these cosmic rays diffuse through the intergalactic medium, where they can interact with the Extragalactic Background Light (EBL) and the Cosmic Microwave Background (CMB) and their trajectories are deflected by the Galactic and Extragalactic magnetic fields. These deflections in both coherent and random magnetic fields isotropize the trajectories of these cosmic rays, leading to a relatively uniform distribution of their arrival directions at Earth.

As UHECRs interact with the nuclei of the Earth's atmosphere, at nucleon-nucleon center of mass energies of $\sqrt{s_{\text{NN}}} \sim 100\,\text{TeV}$, they trigger macroscopic cascades of particles known as Extensive Air Showers (EAS). The measurement of these showers allows for the reconstruction of key properties of the incident cosmic ray, or *primary particle*, such as its energy, E_0 and arrival direction, defined by the local zenith, θ and azimuthal, ϕ angles. It also enables the interpretation of their mass composition, which is highly dependent on the assumptions employed in the modelling of the succession of hadronic interactions driving the cascade. The low momentum transfer in hadronic-ion interactions with nitrogen and oxygen targets precludes their full treatment from the first principles of perturbative Quantum Chromodynamics (QCD). As a result, despite the improvement in experimental systematic uncertainties over the past decades [3], the large hadronic-interaction-model systematic uncertainty in the mass composition of UHECRs hampers the inference of their sources, in at least two ways. First, if the mass number A and charge Ze (with Z the atomic number and e the charge of the proton) of the incident nucleus are not precisely known, its rigidity $R = E_0/Ze$, is poorly constrained. Since, for a fixed magnetic field configuration, higher rigidities lead to smaller deflections, the imprecise determination of the rigidity results in large uncertainty regions for the backtracked direction of the UHECR. Second, the availability of nuclear species in the vicinity of the acceleration sites and the proposed acceleration mechanisms yield different expectations for the injected cosmic ray flux [4]. Hence, even if the change in composition during the diffusion of the cosmic ray flux is well accounted for, their composition at Earth is not accurately known, preventing a straightforward discrimination between different source classes. Additionally,

the modelling of the flux of astrophysical photons and neutrinos, produced from the interaction of the cosmic ray with particles in the vicinity of the source, the so-called *source environment*, also depends on the hadronic interaction model [5, 6], hindering efforts to constrain UHECR sources using a multi-messenger approach.

On the other hand, the unknown primary mass prevents the tuning of hadronic interaction models from measured shower observables. Therefore, there is a degeneracy between the knowledge of the primary mass and an accurate description of the interactions, which needs to be broken by combining several cascade observables. This challenge is one of the main motivators of this work.

1.1.1 Brief historical context

Cosmic rays were discovered by Victor Hess in 1912 [7], when he observed an increase in the air's ionisation rate with altitude during a balloon flight. In the early 1930s, this indirect measurement of cosmic radiation was supplanted by the direct measurement of high-energy cosmic particles in cloud chambers. This has led to the discovery of the positron by Anderson [8].

Using Geiger counts for triggering, the detection efficiency of cloud chambers could be increased, leading to the discovery of muons [9]. Later, charged pions [10] and kaons [11] were discovered with the advent of photographic emulsions. This tight connection between particle discovery and cosmic rays was not long-lasting, as in 1948 the first particle accelerator was built. Since then, accelerating and colliding particles to higher and higher energies has boosted the development of the Standard Model of Particle Physics and has been the main way of testing its predictions.

Already in the late 1930s, the coincident triggering of Geiger counters placed at distant locations, first by Rossi [12] and then by Pierre Auger [13], led to the discovery of Extensive Air Showers. In fact, the measurements of the experiment devised by the latter physicist led to the discovery of the first PeV cosmic-rays [14]. As a result, the focus of cosmic ray research shifted to understanding the origin of these particles. In decades that followed, larger and larger and more sophisticated cosmic ray experiments (see [15] for a review) enabled high-precision measurements of cosmic ray energy spectrum and flux.

Curiously, these high-precision measurements revealed tensions between the primary mass inferred from different air shower observables [16], motivating the study of UHECRs to probe high-energy hadronic interactions and test the limits of their description.

1.1.2 The energy spectrum of cosmic rays

The cosmic-ray flux, that is, the number of cosmic rays, $\mathcal{N}$, per unit area A, solid angle, Ω and time, t, with energy above E_0

$$\Phi(E_0) = \frac{\mathrm{d}\mathcal{N}}{\mathrm{d}\Omega\,\mathrm{d}A\,\mathrm{d}t}, \tag{1.1}$$

is related to the differential cosmic ray spectrum, $J(E_0)$, as

$$J(E_0) = \frac{\mathrm{d}\Phi}{\mathrm{d}E_0}. \tag{1.2}$$

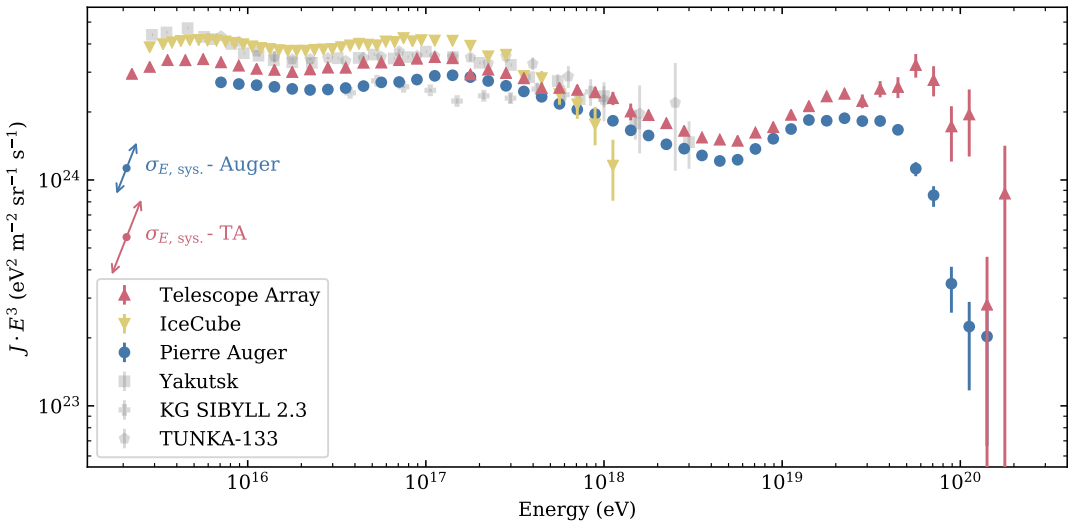

Figure 1.1: Cosmic ray energy spectrum above 1 PeV multiplied by E_0^3, where E_0 is the reconstructed primary energy, measured by several cosmic ray experiments. References to the data are in [3]. The figure was taken from [3], and is available on the repository `arxiv` under the Creative Commons license CC BY-SA 4.0 which can be accessed at `https://creativecommons.org/licenses/by-sa/4.0/`.

This differential spectrum is usually cast as an energy-dependent power law with spectral index $\gamma(E_0) > 0$:

$$J(E_0) \propto E_0^{-\gamma(E_0)}, \tag{1.3}$$

and is shown in Figure 1.1 multiplied by E_0^3 to emphasise its main features over several orders of magnitude in primary energy. The experimental points were obtained from data collected by multiple cosmic ray experiments, as described in [3].

The cosmic ray spectrum is steeply falling over several orders of magnitude in energy. The scarcity of cosmic rays with energies above 100 TeV implies that they cannot be directly measured using satellite or balloon experiments. Instead, they are indirectly measured through the extensive air showers they produce in the atmosphere in observatories extending over several square kilometres and operating over long periods of time. This results in exposures of $\sim 100\,000\,\mathrm{km}^2\,\mathrm{sr\,year}$, allowing for the precise measurement of the cosmic-ray flux at ultra-high energies, at which 1 cosmic-ray per square kilometre per century is expected.

The main features of the cosmic ray spectrum correspond to abrupt changes in the spectral index γ. This quantity is extracted by fitting the cosmic-ray spectrum to a collection of broken power-laws with smooth transitions at their juxtaposition points [17]. The key features of the cosmic ray spectrum are:

- *Knee*: corresponds to an increase of the spectral index from $\gamma = 2.7413 \pm 0.0004 \pm 0.0050$ to $\gamma = 3.128 \pm 0.005 \pm 0.027$, at $E_0 \simeq 3 \times 10^{15}$ eV, as measured by LHAASO [18]. The same observatory reported a hardening of the proton spectrum just below the knee, followed by an abrupt softening, possibly due to proton escape from the Galaxy [19]. This link between the all-particle and proton knee challenges the standard interpretation of the feature as the acceleration limit of heavy Galactic cosmic rays in supernova remnants (SNRs) [20], with a maximum energy $E_{\mathrm{max}} \propto Z$. For an

iron-dominated composition at the knee, the so-called Peters cycle [21] predicts a proton cutoff near 0.1 PeV, at odds with LHAASO results. A lighter composition at the knee is also consistent with the observed PeV gamma rays [22], likely produced by proton–photon interactions in the source environment.

- *Second knee*: at around $E_0 \simeq 1 \times 10^{17}$ eV, corresponds to the increase in spectral index from $\gamma = 2.85 \pm 0.01 \pm 0.05$ to $\gamma = 3.283 \pm 0.002 \pm 0.100$ [23]. The composition in this energy range is compatible with the mass number of the Carbon-Oxygen-Nitrogen (CNO) group [24], indicating that the second knee could be the high-rigidity counterpart of the knee. Alternatively, this feature can also be due to the transition between the Galactic and Extragalactic components of cosmic rays [25].

- *Ankle*: at $E_0 = (5.0 \pm 0.1) \times 10^{18}$ eV, it corresponds to a hardening of the spectrum marked by the change in spectral index from $\gamma = 3.29 \pm 0.02$ to $\gamma = 2.51 \pm 0.03$ [26]. The origin of this feature is still under debate, with suggestions of being connected with the transition between Galactic and Extragalactic cosmic rays. This hypothesis is coherent with the observation of the dipolar anisotropy in the arrival direction of cosmic rays above 8 EeV.

- *Instep*: at $E_0 = (1.3 \pm 0.1) \times 10^{19}$ eV, it corresponds to a flattening of the spectrum weighted by E_0^3. The spectral index after this transition energy reads $\gamma = 3.05 \pm 0.05$ [26], and it could be reproduced by assuming an energy-dependent mass composition of cosmic rays [27]. However, due to the recent discovery of this feature, further data is needed to understand its origin.

- *Suppression*: Corresponds to a sharp softening of the spectrum with $\gamma = 5.1 \pm 0.3$ above $E_0 = (46 \pm 3) \times 10^{18}$ eV [26]. Whether this suppression is due to the lack of sources capable of accelerating cosmic rays at much higher energies, to propagation effects, or both, is unclear and tied to multimessenger and mass-composition measurements at high energies. Multiple processes affecting cosmic-ray propagation at these energies have been proposed. Paradigmatic examples include the GZK cut-off [28, 29] and the Giant Dipole Resonance (GDR) [30]: a photo-excitation and dissociation of heavy nuclei. The GZK cut-off results from a shrinking of the cosmic ray horizon due to the absorption of protons in interaction with CMB photons. These processes include $p+\gamma_{\rm CMB} \to (\Delta^+)^* \to p+\pi^0$ and $p+\gamma_{\rm CMB} \to (\Delta^+)^* \to n+\pi^+$, and for a CMB temperature of $T = 2.7$ K, the minimum energy of these processes is indeed $\sim 5 \times 10^{19}$ eV. Taking into account the density of CMB photons and the cross-section for these processes allows the calculation of the typical attenuation length of the order of 100 Mpc, therefore limiting the number of sources contributing to the proton flux at Earth at ultra-high energies. Moreover, protons resulting from the decay of the Δ^+ resonance have lower energies than the initial ones, softening their injection energy spectrum. Data from the Pierre Auger Observatory suggests that the composition around the suppression is heavier than proton [31], the GZK effect is disfavoured relative to nuclear photo-dissociation processes.

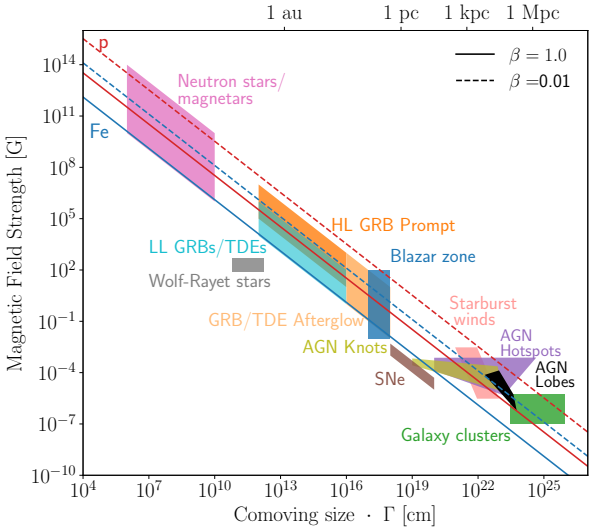

Figure 1.2: Hillas plot with candidate sources of ultra-high-energy cosmic rays. The proton and iron lines for primary energies of 10^{20} eV are indicated by the red and blue lines, respectively. The dashed and solid lines correspond to characteristic velocities of the bulk of plasma particles, relative to the speed of light, $\beta = 0.01$ and $\beta = 1$, respectively. The figure was taken from [3], and is available on the repository `arxiv` under the Creative Commons license CC BY-SA 4.0 which can be accessed at `https://creativecommons.org/licenses/by-sa/4.0/`.

1.1.3 Candidate sources for ultra-high-energy cosmic rays

Despite the precise measurement of the ultra-high-energy cosmic ray spectrum, the sources of these particles are not yet known. However, there are strong constraints on the possible sites of acceleration of EeV particles. One general constraint is the so-called *Hillas criterion* [32], that is, the condition that the gyro-radius of the cosmic ray in a magnetized region is smaller than the comoving size of the accelerator, L. In magnetized plasmas, with typical motions with velocities β and relativistic Lorentz factors Γ, the generated electric fields capable of gradually accelerating cosmic rays with charge Ze, yield a maximum energy of

$$E_{\max} = \eta^{-1} \beta Z e B L \Gamma, \tag{1.4}$$

where η is the efficiency of the accelerator and B is the intensity of the magnetic field. For a fixed cosmic ray composition and acceleration efficiency, Equation (1.4) linearly relates the intensity of the magnetic field B and the acceleration zone size $L\Gamma$, needed to reach each $E_{\max}$, as shown in Figure 1.2. The figure shows a collection of sources, capable of accelerating iron or proton nuclei up to $E_0 = 10^{20}$ eV, for different velocities of the bulk of the plasma.

Further constraints on these sources can be provided by the knowledge of the mass composition of the cosmic rays, as their mass composition at injection will discern between astrophysical objects. For example, the presence of heavy nuclei at the highest energies disfavours transient events such as the tidal disruption of stars or white dwarfs, as these objects mainly consist of light elements [4]. On the other hand, it would favour source classes such as neutron-star mergers [33], magnetars [34], long gamma-ray bursts [35]. Additionally, the reduced spread in the possible rigidities of UHECRs at Earth [36] favours classes of accelerators with little variability or even single sources like Centaurus A [37,

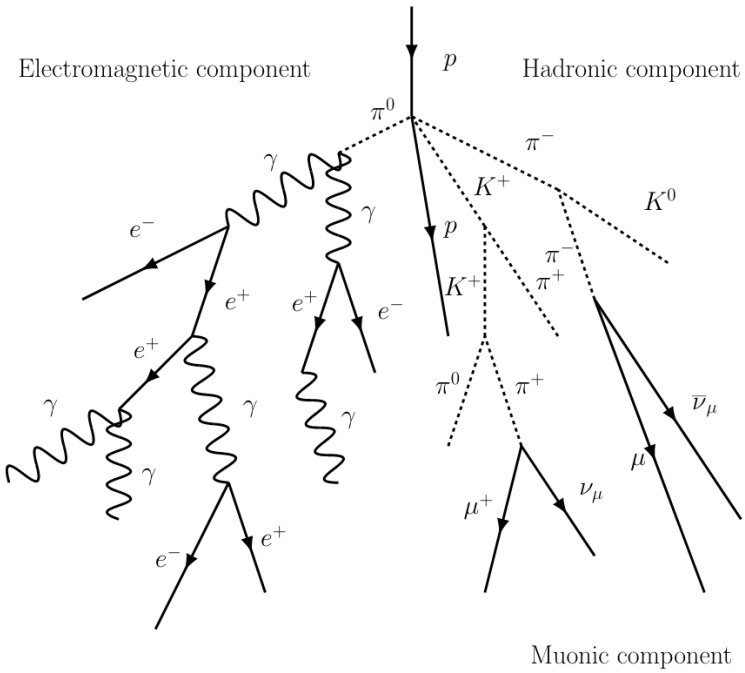

Figure 1.3: Simplified scheme of hadron production, propagation and decay in the first couple of generations of proton-induced extensive air showers. Common mesons, light baryons and leptons are represented. The electromagnetic, hadronic and muonic components are properly identified next to the group of particles constituting these components.

38]. The distinction between these two scenarios benefits from high-statistic measurement of anisotropies of the cosmic-ray flux.

Additionally, since the highest rigidity cosmic-ray events are less deflected by magnetic fields during propagation, the search for small and intermediate scale anisotropies in the cosmic-ray flux in correlation with source catalogues can favour certain catalogues. As an example, in [39] it is shown that the arrival directions of cosmic rays measured by the Pierre Auger Observatory above 32EeV are more correlated with Starburst Galaxies than with Active Galactic Nuclei.

Therefore, the accurate knowledge of the primary mass composition, via the interpretation of extensive air shower observables is essential to discern between the candidate sources of ultra-high energy cosmic rays.

1.2 Extensive Air Showers

In Section 1.1.2, we established that ultra-high-energy cosmic rays are measured via the cascades of particles they trigger when they interact with nuclei of the Earth's atmosphere. In this section, we provide a macroscopic view of these extensive air showers focused on the characterisation of their main components, depicted in Figure 1.3.

1.2.1 Electromagnetic and Hadronic Components

After the primary interaction, the primary energy, i.e., the energy of the incident cosmic ray, is shared among the final state secondary particles. These tend to either further interact or decay according to their proper lifetime, τ, and their boost factor, γ. Short-lived hadrons that decay into photons feed the so-called *electromagnetic sector* or *component* of the shower. That is, a collection of electromagnetic sub-cascades, triggered by each photon, which develop mostly via $e^{\pm}$-pair production and *bremsstrahlung*. The main hadrons feeding the electromagnetic cascade are:

- neutral pions, π^0, with a proper lifetime of solely $\tau_{\pi^0} \simeq 8.43 \times 10^{-17}\,\mathrm{s}$ [40] and a *branching ratio* of 99% for the decay

$$\pi^0 \to \gamma\gamma. \tag{1.5}$$

- η mesons, with a proper lifetime of $\tau_\eta \simeq 5 \times 10^{-19}\,\mathrm{s}$ [40], via the decays

$$\eta \to \gamma\gamma \quad ; \quad \eta \to 3\pi^0 \to 6\gamma, \tag{1.6}$$

  with branching ratios of 39.36% and 32.56% [40].

Longer-lived hadrons, such as light baryons and mesons, tend to further interact, according to their energy, and will be referred to as *hadronically interacting particles* throughout this work. These constitute the *hadronic sector* which is mostly composed of protons, anti-protons, neutrons ($\tau_n = 878.4\,\mathrm{s}$ [40]), charged pions $\pi^{\pm}$ ($\tau_{\pi^{\pm}} = 2.8 \times 10^{-8}\,\mathrm{s}$ [40]) and charged kaons $K^{\pm}$ ($\tau_{K^{\pm}} = 1.23 \times 10^{-8}\,\mathrm{s}$ [40]). On average, these hadrons carry $\approx 80\%$ of the projectile's energy, with the remaining $\approx 20\%$ mainly carried by neutral pions [41]. An interesting consequence of this fact is the power-law increase of the total amount of energy in the electromagnetic component of the shower. Let g denote the so-called shower generation, such that particles produced by the primary interaction belong to generation $g = 1$, those resulting from their further interaction belong to generation $g = 2$, and so on and so forward. If α_{had} denotes the fraction of the projectile's energy carried by the hadronically interacting secondaries of a given interaction, then the fraction of the primary energy in the hadronic sector at generation g, $\alpha_{\mathrm{had},g}$ reads

$$\alpha_{\mathrm{had},g} = \alpha_{\mathrm{had}}^{g} \implies \alpha_{\mathrm{EM},g} = 1 - \alpha_{\mathrm{had}}^{g}, \tag{1.7}$$

where, by conservation of energy, $\alpha_{\mathrm{EM},g} = 1 - \alpha_{\mathrm{had},g}$.

Throughout the cascade, the succession of hadronic interactions degrades the primary energy over an exponentially growing number of particles, causing the energy per particle to decrease, thereby increasing the probability for their decay. The energy at which the decay probability surpasses that of further interaction is called *the critical energy*. Therefore, after an initial increase over the first few interactions, the particle content of the cascade starts decreasing, as the energy of the interacting hadrons falls below their critical energy.

Critical energy

For a given hadron species k, its average interaction length, $d_{\text{int},k}$, at height h in the atmosphere is related to the hadron-air interaction cross section, $\sigma_{k-\text{air}}$, as

$$d_{\text{int},k} = \frac{\langle M_{\text{air}} \rangle}{\rho(h)\sigma_{k-\text{air}}} \implies \lambda_{\text{int},k} = \frac{\langle M_{\text{air}} \rangle}{\sigma_{k-\text{air}}}, \tag{1.8}$$

where $\langle M_{\text{air}} \rangle$ is the average mass of an air nucleus (an average of the masses of the individual elements of the atmosphere weighted by their abundances), $\rho(h)$ is the atmospheric density at height h and $\lambda_{\text{int},k} \equiv \rho(h) d_{\text{int},k}$ is the interaction length expressed as a grammage. When there is no ambiguity, all interaction lengths will be expressed in $\mathrm{g\,cm^{-2}}$, since these are independent of the local target density, or, in this case, the local density of the atmosphere. Due to the rise of the interaction cross-section with the projectile's energy, E_k, the interaction length is a slowly decreasing function of E_k: $\lambda_{\text{int},k}(E)$ [42].

On the other hand, the decay length of an unstable hadron k with mass M_k and proper lifetime τ_k depends on the density of the atmosphere as

$$\lambda_{\text{dec},k}(E, \rho) = \rho(h) c \gamma \tau_k, \quad \text{with} \quad \gamma = \frac{E_k}{M_k}, \tag{1.9}$$

with $c = 299\,792\,458\,\mathrm{m\,s^{-1}}$ [40] the speed of light in vacuum. Thus, the critical energy for species k, ξ_c^k, obeys the equation

$$\lambda_{\text{int},k}(\xi_c^k) = \lambda_{\text{dec},k}(\xi_c^k) \iff \lambda_{\text{int},k}(\xi_c^k) = \rho(h) c \left(\frac{\xi_c^k}{M_k} \right) \tau_k. \tag{1.10}$$

More intuition about the critical energy can be gained by considering a simple isothermal model of the atmosphere. With this assumption, the air density decreases exponentially with height h

$$\rho(h) = \rho_0 \mathrm{e}^{-h/h_0}. \tag{1.11}$$

As a result, the vertical depth or grammage down to height h, $X_{\text{vert}}(h)$ reads

$$X_{\text{vert}}(h) = \int_h^\infty \rho(h')\,\mathrm{d}h' = \rho_0 h_0 \mathrm{e}^{-h/h_0} = h_0 \rho(h). \tag{1.12}$$

Let θ denote the incidence angle of the primary cosmic-ray at the top of the atmosphere. In each interaction, the typical transverse momentum of the produced hadrons is of the order of ~ 250 MeV [20], so that, to a good approximation, we can assume that all hadrons retain the direction of the incident cosmic-ray θ. Moreover, for a flat atmosphere, this angle is constant at every height h and coincides with the zenith angle of the resulting shower. Therefore, the *slant depth* corresponding to $X_{\text{vert}}(h)$, is simply

$$X = X_{\text{vert}}(h) \sec\theta = h_0 \rho(h) \sec\theta, \tag{1.13}$$

allowing us to express Equation (1.9) as

$$\lambda_{\text{dec},k}(E) = \frac{c\tau_k}{h_0 m} E X \cos\theta = \frac{E}{\varepsilon_k} X \cos\theta \quad \text{with} \quad \varepsilon_k = \frac{h_0 M_k}{c\tau_k}. \tag{1.14}$$

Note that ε_k has the interpretation of an *attenuation energy*. Therefore, the critical energy defined in Equation (1.10) can be approximated by

$$\lambda_{\mathrm{int},k}(\xi_c^k) = \frac{\xi_c^k}{\varepsilon_k} X \cos\theta \iff \xi_c^k = \frac{\lambda_{\mathrm{int},k}(\xi_c^k)\varepsilon_k}{X} \sec\theta \tag{1.15}$$

Neglecting the slow variation of the interaction length with the hadron's energy, Equation (1.15) shows that critical energy tends to decrease with the slant depth X, since due to the increase in air density, hadrons tend to interact down to lower energies before decaying. Moreover, the critical energy increases with the zenith angle and with the attenuation constant ε_k. This constant provides the order of magnitude of the typical critical energy of each hadron species. For charged pions, we have $\varepsilon_{\pi^\pm} \simeq 100\,\mathrm{GeV}$ and for charged kaons $\varepsilon_{K^\pm} \simeq 850\,\mathrm{GeV}$. These energies also provide the typical energy for the decay products of these hadrons, which are mainly muons, as detailed in the next section.

1.2.2 Muon Component

The majority of the particles of the hadronic component, or hadronic core, that drive the cascade are charged pions, $\pi^\pm$, and kaons. The main decay channel of charged pions, with a branching ratio of 99.99% [40], is

$$\pi^\pm \to \mu^\pm + \nu_\mu/\overline{\nu}_\mu, \tag{1.16}$$

while the main decay channels of charged kaons are

$$K^\pm \to \mu^\pm + \nu_\mu/\overline{\nu}_\mu \quad \text{and} \quad K^\pm \to \pi^+ + \pi^0 \to \mu^\pm + \nu_\mu/\overline{\nu}_\mu + 2\gamma, \tag{1.17}$$

with branching ratios of 63.56% and 20.67%, respectively [40]. Therefore, the decay of the hadronic core dominantly feeds the muonic component of the shower. Henceforth, if there is no ambiguity, we will refer to muons and anti-muons as simply muons. These muons, together with prompt muons and photo-produced muons, constitute the *muon component*.

The mentioned leptons have a proper lifetime of $2.2 \times 10^{-6}\,\mathrm{s}$ [40] and a mass of $M_\mu = 105.65\,\mathrm{MeV}$ [40]. Therefore, for the typical critical energies of charged pions and kaons, the Lorentz boost factors of muons range from $\gamma \simeq 10^3$ to 10^4. Consequently, in the lab frame, their typical decay distances, $c\gamma\tau_\mu$, are of the order of tens of kilometres. Moreover, muons lose energy in the atmosphere essentially via ionisation, over a large energy range. In particular, the kinetic energy loss of muons per depth $\mathrm{d}X$ in the atmosphere at minimum ionising energy is given by [43]

$$\frac{\mathrm{d}E_\mu}{\mathrm{d}X} \simeq -2\,\mathrm{MeV\,g^{-1}\,cm^2}. \tag{1.18}$$

At sea level, the vertical atmospheric depth is about $1000\,\mathrm{g\,cm^{-2}}$, corresponding to energy losses of roughly 2 GeV. Consequently, most muons neither decay nor lose significant energy while traversing the atmosphere. They also experience negligible deflection from multiple scattering, and, in vertical showers, little deflection by the geomagnetic field, largely preserving the transverse momentum of their parent pion or kaon. This

momentum is approximately exponentially distributed around 250 MeV, so muons tend to follow the shower axis in nearly straight lines. As a result, muons generated over multiple shower generations can reach the ground level, retaining information on the hadron production history throughout the shower.

1.2.3 Cascade Equations

The interaction, decay and transport of shower particles in the atmosphere can be described by a set of coupled differential and integral equations known as Cascade Equations [20]. Let $\Phi_k(E, X)$ denote the flux of particles of type k with energy in the interval E to $E + \mathrm{d}E$ at slant depth X. The decrease of this flux between depths X and $X + \mathrm{d}X$ is due to: the interaction of particles of type k, their decay (in case these are unstable particles) and their attenuation as they propagate in the atmosphere. On the other hand, the interaction of other particles and their decay into particles of type k leads to an increase of $\Phi_k(E, X)$ over $\mathrm{d}X$. By accounting for each of these contributions, we can write

$$\begin{aligned}\frac{\mathrm{d}\Phi_k}{\mathrm{d}X} = &-\left[\frac{1}{\lambda_{\mathrm{int},k}(E)} + \frac{1}{\lambda_{\mathrm{dec},k}(E,X)}\right]\Phi_k(E,X) - \\ &- \frac{\mathrm{d}}{\mathrm{d}E}\left[\mu_k(E)\Phi_k(E,X)\right] + \\ &+ \sum_i \int_E^\infty \frac{1}{\sigma_{i-\mathrm{air}}(E_i)}\frac{\mathrm{d}\sigma_{i-\mathrm{air}\to k}}{\mathrm{d}E}\frac{\Phi_i(E_i,X)}{\lambda_{\mathrm{int},i}(E_i)}\,\mathrm{d}E_i + \\ &+ \sum_i \int_E^\infty \Gamma_{i\to k}(E_i,E)\frac{\Phi_i(E_i,X)}{\lambda_{\mathrm{dec},i}(E_i)}\,\mathrm{d}E_i\,.\end{aligned} \tag{1.19}$$

The first three terms of Equation (1.19) describe losses, with $\mu_k(E)$ encompassing all the energy loss processes undergone by particle k during its propagation in the atmosphere, including ionisation losses. The integral terms are summed over all particles i with energies greater than E, which contribute to the production of particles k. These are the so-called *source terms* and are responsible for coupling the individual cascade equations for each particle species. The production of particles of type k via interaction must take into account the differential yield of particles of type k in $i -$ air interactions. This yield is described in terms of the interaction cross-section $\sigma_{i-\mathrm{air}}$ and of the differential *inclusive cross-section* for the production of particles of type k in $i -$ air interaction, combined as $\frac{1}{\sigma_{i-\mathrm{air}}(E_i)}\,\mathrm{d}\sigma_{i-\mathrm{air}\to k}/\mathrm{d}E$. Likewise, the yield of particles of type k via decay of other particles, $\Gamma_{i\to k}(E_i, E)$, sums over the branching ratios of all decay channels of hadron i that lead to particles of type k.

As is, Equation (1.19) only permits computing the longitudinal development of the shower, that is, along the slant depth X. However, the formalism can be extended to compute the lateral distribution function of each particle, that is, the particle density as a function of the radial distance to the shower core, r, assuming an axially symmetric production and propagation of particles around the shower axis. For particular initial conditions, that is, an initial cosmic-ray flux at $X = 0$, these cascade equations can either be analytically solved (as in the case of electromagnetic showers under certain approximations [44]) or numerically solved by packages like CONEX [45, 46] and MCEQ [47].

Electromagnetic cascades

A paradigmatic example of solutions to Cascade Equations is that of the longitudinal development of photon-induced electromagnetic (EM) cascades [48]. The main processes driving these cascades are pair-production and *bremsstrahlung*. The typical interaction length of these processes is the so-called *radiation length*, $\lambda_r \simeq 37\,\mathrm{g\,cm^{-2}}$ in air [20]. Electromagnetic cascades grow until collisional losses of electrons in the atmosphere surpass radiative processes. This occurs at a critical energy $\xi_c^e \simeq 87\,\mathrm{MeV}$ [20]. In fact, it was found by Greisen [49], that the total number of electrons at the dimensionless depth $t = X/\lambda_r$, $N_e(t)$, of a cascade started by a photon with energy E_0 is well described by

$$N_e(t) = \frac{0.31}{\sqrt{\beta_0}} \exp\left\{ t \left(1 - \frac{3}{2} \ln s \right) \right\} \quad \text{with} \quad \beta_0 = \ln\left(\frac{E_0}{\xi_c^e} \right) \text{ and } s = \frac{3t}{t + 2\beta_0}. \tag{1.20}$$

The parameter s denotes the *shower age*, while β_0 denotes the dimensionless depth at which the number of electrons is maximal. In fact, we can identify this variable with the depth of the shower maximum $X_{\max}$

$$X_{\max} = \lambda_r \beta_0 = \lambda_r \ln\left(\frac{E_0}{\xi_c^e} \right). \tag{1.21}$$

Notably, this primary energy dependence of the depth of the shower maximum can also be derived in the framework of the Heitler model [50], which we derive, in detail, in Section 1.2.4.

Hadronic Cascades

The left panel of Figure 1.4 shows the longitudinal development of the photonic, electronic/positronic, hadronic and muonic components, obtained by solving cascade equations for a proton-induced shower with primary energy $E_0 = 10^{19}$ eV. For visualisation purposes, some longitudinal profiles were multiplied by a factor. The right panel of the figure shows the energy deposit of the shower per unit depth, $\mathrm{d}E/\mathrm{d}X$.

All profiles show the same general trend: an increase in the particle content up to a maximum, followed by a decrease, when either decay or collisional losses start dominating over further particle interactions. The maxima of each component are reached at different depths: the hadron profile tends to peak higher in the atmosphere due to the rapid dissipation of the primary energy and meson-decay occurring at $\sim 100\,\mathrm{GeV}$, while the electromagnetic profile tends to peak closer to the ground. After the maximum, the particle content of the electron and photon components decay exponentially. The muon profile, due to the relatively large lifetime of these leptons and their small energy losses in the atmosphere, peaks deep in the atmosphere and then declines more slowly after reaching its maximum depth. Moreover, the profile of charged particles is dominated by electrons and positrons.

Since most of the primary energy is carried by electromagnetic particles $(\gamma, e^{\pm})$, the depth of maximum of the energy deposit coincides with that of the electromagnetic profile.

Despite the many applications of cascade equations to model the development of hadronic and electromagnetic showers, as well as to compute neutrino and photon fluxes produced by the interaction of cosmic rays in the environment of their sources, they do not

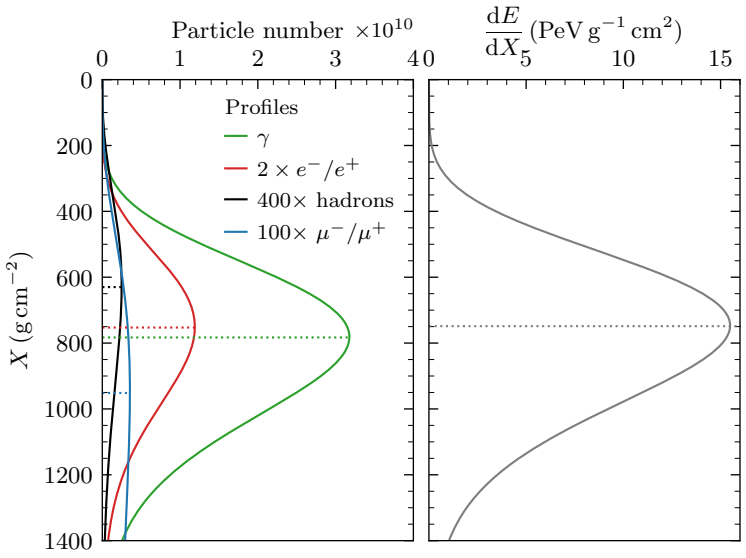

Figure 1.4: Left panel: Longitudinal profiles of the hadronic, electronic, photonic and muonic components for a proton-induced EAS. For visualisation purposes, some profiles were multiplied by an appropriate factor. The dotted colored lines indicate the depth of maxima of each profile. Right panel: energy deposit of the shower per unit depth, $\mathrm{d}E/\mathrm{d}X$. All profiles are obtained by solving cascade equations for proton-induced extensive air showers for $E_0 = 10^{19}$ eV.

account for shower-to-shower fluctuations. These fluctuations arise from the stochasticity of particle production, decay and transport in the atmosphere. Furthermore, the mentioned deterministic framework does not allow for the deduction of scaling laws for certain observables of EAS, for example, the depth of shower maximum $X_{\max}$ and the number of muons and electrons, as a function of basic shower parameters such as the primary energy E_0 and the primary mass number A (the number of nucleons of primary nuclei).

1.2.4 Cascades as splitting processes

Heitler model for electromagnetic cascades

The Heitler model [50] is a simplified description of photon-induced electromagnetic (EM) cascades as a branching process driven by pair production and *bremsstrahlung*. The model assumes that after a photon interacts, a pair of e^+/e^- is produced. In turn, these undergo bremsstrahlung, producing exactly one photon, continuing the cascading process. Either of these processes occurs after a depth $d = \lambda_r \ln 2$, where $\lambda_r \simeq 37\,\mathrm{g\,cm^{-2}}$ is the radiation length [20]. A scheme of such a cascade is shown in the left panel of Figure 1.5.

Since two particles are produced in each interaction, after g interactions or shower generations, the number of particles in the cascade is simply $N_g = 2^g$. Further assuming that, in each interaction, the energy of the incident particle is equally shared by both secondaries, implies that the energy per particle after g interactions is simply $E_g = E_0/2^g$, where E_0 denotes the energy of the primary photon. The multiplicative radiative processes driving the shower are assumed to be surpassed by collisional losses in the atmosphere at a critical energy $\xi_c^e \simeq 87\,\mathrm{MeV}$, corresponding to the critical generation or critical number

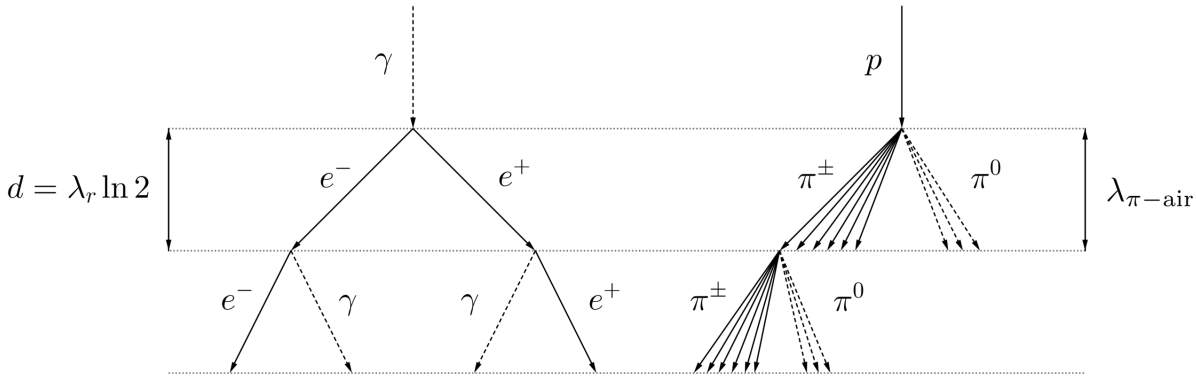

Figure 1.5: Left: Schematic representation of a photon-induced electromagnetic cascade according to the Heitler model. Right: Schematic representation of the proton-induced pionic cascade according to the Heitler-Matthews model.

of interaction steps

$$\xi_c^e = \frac{E_0}{2^{g_c}} \implies g_c = \frac{1}{\ln 2} \ln\left(\frac{E_0}{\xi_c^e}\right). \tag{1.22}$$

At the critical generation, the particle content of the electromagnetic shower reaches its maximum, given by

$$N_{\max} = 2^{g_c} = \frac{E_0}{\xi_c^e}. \tag{1.23}$$

Furthermore, the depth at which this maximum is reached, $X^\gamma_{\max}$, is given by

$$X^\gamma_{\max} = g_c \times d = \lambda_r \ln\left(\frac{E_0}{\xi_c^e}\right). \tag{1.24}$$

Remarkably, the logarithmic increase of $X^\gamma_{\max}$ with the primary energy of the incident photon reproduces fairly well the results of dedicated simulations of EAS [51]. In particular, for photon primaries, the so-called *elongation rate*, i.e., $\mathrm{d}X^\gamma_{\max}/\mathrm{d}\ln E_0$, is given by

$$\frac{\mathrm{d}X^\gamma_{\max}}{\mathrm{d}\ln E_0} = \lambda_r. \tag{1.25}$$

Heitler-Matthews model for hadronic cascades

The dynamics of proton-induced hadronic cascades can be understood by extending the Heitler model to pionic showers. This framework is known as the Heitler-Matthews model [51]. Its core assumptions are:

1. In each pion-air interaction, a fixed multiplicity of secondary pions arises, m_{tot}, independently of the projectile's energy;

2. One third of secondaries are neutral pions, $m_{\mathrm{EM}} = m_{\mathrm{tot}}/3$, which immediately decay into two photons, starting an electromagnetic cascade. The remaining two-thirds of secondaries are charged pions, $m_{\mathrm{had}} = 2m_{\mathrm{tot}}/3$, which further interact until reaching their critical energy ξ_c^π at which they instantly decay into muons;

3. In each pion-air interaction, the projectile's energy is equipartitioned among all secondary particles;

4. Charged pions interact after an energy independent interaction depth $\lambda_{\pi-\text{air}}$.

This model is schematically represented in the right panel of Figure 1.5. After g steps, or generations, the total number of charged pions, M_g, and the energy per charged pion, E_g, read

$$M_g = m_{\text{had}}^g \quad \text{and} \quad E_g = \frac{E_0}{m_{\text{tot}}^g}. \tag{1.26}$$

The number of muons, N_μ, is simply given by the number of charged pions at the critical generation, g_c. This generation is such that the energy per charged pion is the critical energy ξ_c^π, and is thus given by

$$\xi_c^\pi = \frac{E_0}{m_{\text{tot}}^{g_c}} \iff g_c = \frac{1}{\ln m_{\text{tot}}} \ln\left(\frac{E_0}{\xi_c^\pi}\right). \tag{1.27}$$

Hence, the number of muons reads

$$\ln N_\mu = g_c \ln m_{\text{had}} = \frac{\ln m_{\text{had}}}{\ln m_{\text{tot}}} \ln\left(\frac{E_0}{\xi_c^\pi}\right) \implies N_\mu = \left(\frac{E_0}{\xi_c^\pi}\right)^\beta, \text{ with } \beta = \frac{\ln m_{\text{had}}}{\ln m_{\text{tot}}}. \tag{1.28}$$

Detailed simulations of proton-induced EAS validated the power-law increase in the number of muons in proton-induced showers with $\beta \simeq 0.93$ [51].

In this framework, and letting the depth of the primary interaction point be X_1, then the depth of the maximum of the pionic cascade, X_{max}^π, amounts to

$$X_{\text{max}}^\pi = X_1 + g_c \times \lambda_{\pi-\text{air}} = X_1 + \frac{\lambda_{\pi-\text{air}}}{\ln m_{\text{tot}}} \ln\left(\frac{E_0}{\xi_c^\pi}\right) \tag{1.29}$$

In turn, the maximum of the electromagnetic cascades can be computed using the Heitler model, under the assumption that only neutral pions from the primary interaction contribute significantly to X_{max}. There are m_{EM} such neutral pions, resulting in $2m_{\text{EM}}$ photon-induced showers, which are assumed to peak all at the same depth, X_{max}. Moreover, taking the energy of each photon to be half of the parent neutral pion and using Equation (1.24), leads to

$$X_{\text{max}} = X_1 + \lambda_r \ln\left(\frac{E_0}{2m_{\text{tot}}\xi_c^e}\right). \tag{1.30}$$

The total energy carried by the hadronic component at the critical generation, E_{had} evaluates to

$$E_{\text{had}} = N_\mu \xi_c^\pi = \left(\frac{E_0}{\xi_c^\pi}\right)^{\beta-1} E_0, \tag{1.31}$$

while the energy carried by the electromagnetic component must be, by energy conservation,

$$E_{\text{EM}} = E_0 - E_{\text{had}} = E_0 \left[1 - \left(\frac{E_0}{\xi_c^\pi}\right)^{\beta-1}\right]. \tag{1.32}$$

The total number of produced electromagnetic particles is thus given by

$$N_{\mathrm{EM}} = \frac{E_0}{\xi_c^e}\left[1 - \left(\frac{E_0}{\xi_c^\pi}\right)^{\beta-1}\right]. \tag{1.33}$$

Given $\beta \simeq 1$, we can estimate the order of magnitude of N_{EM}/N_μ as

$$\frac{N_{\mathrm{EM}}}{N_\mu} \approx \frac{\xi_c^\pi}{\xi_c^e} \simeq 1000, \tag{1.34}$$

where we used $\xi_c^\pi \approx 100$ GeV and $\xi_c^e \approx 100$ MeV. Therefore, we indeed expect that the charged-particle profile of a proton-induced shower is dominated by the electromagnetic profile, reaching its maximum at a depth similar to that of the electromagnetic shower.

The Heitler-Matthews model also predicts the elongation rate of X_{max}, D_p. Since the models only consider mean values for each shower observable, we take $X_1 = \lambda_{p-\mathrm{air}}$, to obtain

$$D_p \equiv \frac{\mathrm{d}X_{\mathrm{max}}}{\mathrm{d}\ln E_0} = \lambda_r\left[1 + \frac{\lambda_{p-\mathrm{air}}}{\lambda_r}\frac{\mathrm{d}\ln\lambda_{p-\mathrm{air}}}{\mathrm{d}\ln E_0} - \frac{\mathrm{d}\ln m_{\mathrm{tot}}}{\mathrm{d}\ln E_0}\right] = \lambda_r\left(1 - D_\lambda - D_m\right), \tag{1.35}$$

where we defined

$$D_\lambda \equiv -\frac{\lambda_{p-\mathrm{air}}}{\lambda_r}\frac{\mathrm{d}\ln\lambda_{p-\mathrm{air}}}{\mathrm{d}\ln E_0} \quad \text{and} \quad D_m \equiv \frac{\mathrm{d}\ln m_{\mathrm{tot}}}{\mathrm{d}\ln E_0}. \tag{1.36}$$

Since the proton-air inelastic cross-section tends to increase with the primary energy, just like the multiplicity of secondary particles [42], we have $D_\lambda > 0$ and $D_m > 0$, so that $D_p < \lambda_r$. This important bound is a version of Linsley's elongation theorem [52].

The Heitler-Matthews model can be adapted to account for the so-called *leading particle effect* [20]. The leading particle is the secondary carrying the largest fraction of the incident particle's energy. In the lab. frame, this fraction is the elasticity κ_{el} of the interaction. In this extension, non-leading particles are assumed to still carry an equal fraction of the primary energy. Interestingly, adding this effect still yields a power increase of the muon content of the shower with the primary energy, albeit with an exponent

$$\beta = \frac{\ln m_{\mathrm{had}}}{\ln\left[3m_{\mathrm{had}}/(2+\kappa_{\mathrm{el}})\right]}. \tag{1.37}$$

As for X_{max}, considering that each neutral pion carries a fraction of the primary energy $(1-\kappa_{\mathrm{el}})E_0/(m_{\mathrm{tot}}-1)$, results in

$$X_{\mathrm{max}} = X_1 + \lambda_r \ln\left(\frac{(1-\kappa_{\mathrm{el}})E_0}{2(m_{\mathrm{tot}}-1)\xi_c^e}\right). \tag{1.38}$$

Despite the success of the Heitler-Matthews model in predicting the energy evolution of the mean values of N_μ and X_{max} for proton-induced showers, its main virtue is the ability to predict the primary mass dependence of these observables. This is achieved by assuming the *superposition principle*.

Superposition principle

Let us consider a nucleus with A nucleons. The typical binding energy of these nucleons is in the MeV scale, being therefore negligible compared to the cosmic-ray energy. As such, the *superposition* principle states that the interaction of a nucleus with energy E_0 and mass number A is equivalent to the interaction of A independent nucleons with energy E_0/A. Under this assumption, Equation (1.28) can be written for a primary with mass number A as

$$N_\mu(E_0, A) = A\left(\frac{E_0}{A\xi_c^\pi}\right)^\beta = A^{1-\beta}N_\mu^p, \tag{1.39}$$

where $N_\mu^p \equiv N_\mu(E_0, 1)$ is the Heitler-Matthews prediction for proton-induced showers. Since $\beta \simeq 0.93 < 1$, it follows that showers started by nuclear primaries tend to produce more muons than proton primaries of the same energy. For example, for an iron nucleus with $A = 56$, we expect $\approx 30\%$ more muons than for a proton-induced shower.

Applying the same principle, the depth of the shower maximum for such a nuclear primary reads

$$X_{\max}(E_0, A) = X_1 + \lambda_r \ln\left(\frac{E_0}{2m_{\text{tot}}A\xi_c^e}\right) = X_{\max}^p - \lambda_r \ln A, \tag{1.40}$$

where $X_{\max}^p \equiv X_{\max}(E_0, 1)$. Therefore, heavier primaries tend to trigger shallower showers. This result is intuitive: since the energy per nucleon is smaller, the primary energy degrades faster, so that the critical energy is reached higher in the atmosphere.

Allowing for a primary energy evolution of the primary composition, the elongation rate obtained from Equation 1.40 reads

$$\frac{\mathrm{d}X_{\max}(E_0, A)}{\mathrm{d}\ln E_0} = \frac{\mathrm{d}X_{\max}^p}{\mathrm{d}\ln E} - \lambda_r\frac{\mathrm{d}\ln A}{\mathrm{d}\ln E_0} = D_p - \lambda_r\frac{\mathrm{d}\ln A}{\mathrm{d}\ln E_0}. \tag{1.41}$$

If the primary mass composition does not evolve with energy, then the elongation rate for heavier primaries is the same as the elongation rate for protons, considering the simplified assumptions of the Heitler-Matthews model. Changes in composition with the primary energy are therefore observable via the change in the elongation rate. Moreover, given Linsley's elongation rate theorem, elongation rates larger than λ_r should always be connected with the lighting of the primary mass composition.

Despite the derived mass-trends of N_μ and $X_{\max}$, their interpretation in terms of the mass composition of ultra-high-energy cosmic rays relies on detailed Monte-Carlo simulations of extensive air showers, addressed in Section 1.3.

1.2.5 Shower-to-shower fluctuations

The predictions of the Heitler-Matthews model account for the energy and primary-mass dependence of the expected values of $X_{\max}$ and N_μ. However, particle production and propagation are stochastic processes, so that, even if the primary energy and mass are fixed, the shower observables follow distributions around their mean values. The shower-to-shower fluctuations of mass-sensitive observables prevent the shower-to-shower estimation of the primary mass, but contain important information about the hadronic interac-

tions driving the shower. Interestingly, this information can be qualitatively interpreted in light of Heitler-Matthews-like models.

A notable example was explored in [53], where the distribution of the muon content of extensive air showers was connected to fluctuations of the energy sharing and multiplicity of secondaries of hadronic interactions. Consider that the energy carried by each hadronically interacting particle produced in the primary interaction is allowed to vary. For hadron i, with energy $x_i E_0$ in the lab. frame, its corresponding muon yield is given in Equation (1.28). Therefore, the combined muon yield of all secondary hadrons of the primary interaction reads

$$N_\mu = \sum_{i=1}^{m_{\rm had}} N_\mu(x_i E_0) = \left(\frac{E_0}{\xi_c^\pi}\right)^\beta \sum_{i=1}^{m_{\rm had}} x_i^\beta = \left(\frac{E_0}{\xi_c^\pi}\right)^\beta \alpha_1 \quad \text{with} \quad \alpha_1 = \sum_{i=1}^{m_{\rm had}} x_i^\beta, \tag{1.42}$$

and varies from shower-to-shower according to the stochastic partition of the primary energy upon interaction. In fact, allowing for fluctuations in all deeper shower generations, for which we define a given α_g, we arrive at

$$N_\mu \propto \prod_{g=1}^{g_c} \alpha_g \implies \left(\frac{\sigma(N_\mu)}{N_\mu}\right)^2 = \sum_{g=1}^{g_c} \left(\frac{\sigma(\alpha_g)}{\alpha_g}\right)^2, \tag{1.43}$$

where $\sigma(\cdot)$ denotes the standard deviation of a random variable and we assumed independence of all α_g. Now, α_2 is computed for $m_{\rm had}$ sub-showers. Thus, assuming the independence of these sub-showers, the relative fluctuations of α_2 are suppressed by $\sqrt{m_{\rm had}}$ relative to those of α_1. This suppression is even stronger in deeper shower generations. Therefore, we expected that the main contribution for $\sigma(N_\mu)/N_\mu$ comes from the fluctuations of energy partition in the primary interaction. This is a general property of branching processes [54].

Importantly, in [53] it was shown that $\sigma(\alpha_1)$ accounts for $\sim 70\%$ of the fluctuations of the relative number content of EAS, linking the distribution of this shower observable to the energy spectrum of hadronically interacting particles of the primary interaction. As will be noted in Section 1.3, this spectrum cannot be computed from first principles, and is highly dependent on the assumptions employed in the modelling of hadronic interactions. Therefore, measuring $\sigma(N_\mu)$ provides constraints on particle spectra.

Note that the scale of the number of muons arises from the energy transfer into the hadronic channel in each stage of the shower development, which can be rewritten more explicitly as

$$N_\mu = \frac{E_0}{\xi_c^\pi} \prod_{g=1}^{g_c} \alpha_{{\rm had},g}. \tag{1.44}$$

Another notable example of the connection between particle physics and the fluctuations of shower observables is the connection between the proton-air cross-section and the distribution of $X_{\rm max}$. As noted in Equation (1.30), the depth of the primary-interaction point directly influences the value of $X_{\rm max}$, linking their distributions. In particular, X_1 is distributed according to

$$f(X_1) = \frac{1}{\lambda_{p-{\rm air}}(E_0)} \exp\left\{-\frac{X_1}{\lambda_{p-{\rm air}}(E_0)}\right\} \quad \text{with} \quad \lambda_{p-{\rm air}}(E_0) = \frac{\langle M_{\rm air}\rangle}{\sigma_{p-{\rm air}}(E_0)}. \tag{1.45}$$

This exponential distribution has a long tail towards large values of X_1 with slope $\lambda_{p-\mathrm{air}}$, leading to an exponentially falling tail of the distribution of X_{max} towards deep showers. The slope of this latter tail can then be fitted to constrain $\sigma_{p-\mathrm{air}}$, as demonstrated in [55, 56].

1.3 Fundamentals of the phenomenology of hadronic interactions

The interactions between the constituents of hadrons, that is, quarks and gluons, collectively known as partons, are described using Quantum Chromodynamics (QCD). However, hadron-hadron, $h - h$, interactions cannot be fully described from the first principles of this theory, in virtue of two of its key features [57]:

- *Color confinement*: color-charged particles, i.e., quarks and gluons, cannot exist freely, in virtue of the approximately linear increase of the strong force with distance. Therefore, these particles are ultimately bound in color-neutral objects: hadrons. The process by which partons evolve into hadrons is called *hadronization*;
- *Asymptotic freedom*: at low energy scales, such as those at which hadronization takes place, the strong coupling constant α_s becomes of the order of unity, preventing the calculation of the probability of a given final state of a hadronic interaction using perturbation theory.

Consequently, interaction cross-sections for $h - h$ processes, σ_{hh}, as well as the inclusive cross-sections for the production of a given particle c, $\sigma_{hh\to c}$, cannot be calculated explicitly using perturbation theory. Elastic cross-sections, $\sigma_{hh}^{\mathrm{el}}$, in particular, can not be calculated due to the inherent low-momentum transfer of these processes. The same limitation follows for the *inelastic* cross section:

$$\sigma_{hh}^{\mathrm{inel}} = \sigma_{hh} - \sigma_{hh}^{\mathrm{el}}. \tag{1.46}$$

Instead, these cross-sections must be described using phenomenological or effective theories. This is particularly relevant for Extensive Air Showers, which are driven by the final state hadrons of such interactions. Therefore, fundamental quantities like the differential expected particle yield of species c per unit volume of the phase-space d^3p_c

$$\frac{\mathrm{d}^3 m_c}{\mathrm{d}p_c^3} = \frac{1}{\sigma_{hh}} \frac{\mathrm{d}^3 \sigma_{hh\to c}}{\mathrm{d}p_c^3}, \tag{1.47}$$

or their energy spectrum

$$\frac{\mathrm{d}m_c}{\mathrm{d}E_c} = \frac{1}{\sigma_{hh}} \frac{\mathrm{d}\sigma_{hh\to c}}{\mathrm{d}E_c}, \tag{1.48}$$

where E_c is the energy of the produced hadron in the lab. frame (the relevant frame for air shower development) relies on particular stochastic hadronisation schemes and assumptions employed by different phenomenological Hadronic Interaction Models (HIMs). As it is clear from Equation 1.19, differences in inelastic cross-sections and particle yields will result in different showers, on average.

These hadronic interaction models, implemented as Monte Carlo simulators, handle hadronisation using probabilistic frameworks such as the Lund string fragmentation model [58]. In this framework, the color-field between quark-anti-quark pairs, $q\overline{q}$, is represented by a color string, forming a color-neutral object. While the energy density stored in the string is large, quantum fluctuations make the string break, forming new pairs of bound quarks. This process keeps repeating until the energy stored in the strings is too small to produce another hadron. The resulting color-neutral objects are hadrons. Beyond this hadronization scheme, certain hadronic interaction models include additional phenomena, such as collective hadronization, that modify the final state of the interaction.

These phenomenological models need to be tuned to accelerator data, to yield the correct cross-sections for particular hadronic processes. This data is obtained in restricted regions of the kinematic phase-space of hadron-hadron interactions, that is, at particular center of mass energies $\sqrt{s}$, and for limited emission angles of the final state hadrons. In fixed target experiments at low energies, it is possible to have measurements of particle yields in the so-called forward region. That is, for hadrons produced at high pseudorapidities η:

$$\eta = \frac{1}{2} \ln \left(\frac{|\mathbf{p}| + p_{\parallel}}{|\mathbf{p}| - p_{\parallel}} \right), \tag{1.49}$$

where $|\mathbf{p}|$ is the modulus of the momentum of the hadrons and $p_{\parallel}$ is the component of the momentum along the beam direction. On the other hand, in accelerator experiments, reaching centre of mass energies of $\sqrt{s} = 13$ TeV, such as the LHC, the particle detectors typically measure secondary charged particles with large transverse momentum $p_{\perp}$, and thus low η. Therefore, the description of ultra-high-energy hadron-air interactions lacks data in the high-energy and forward region of the kinematic phase space, where most of the hadrons further interacting are produced. This is clear from Figure 1.6 where the pseudorapidity distribution is weighted by the energy of each secondary particle. This is the so-called differential energy flow $\mathrm{d}E/\mathrm{d}\eta$ and, if the secondary particle hadronically interacts, then it is a proxy to the muon yield per pseudorapidity according to Equation (1.28). If the particle is a neutral pion, then it is a proxy to the energy of the photons produced by its decay, and therefore of the most energetic contributions to X_{max}. The figure also depicts shaded areas indicating the pseudorapidity range coverage of LHC experiments: CMS [59], LHCb [60] and LHCf [61]. Note that LHCf cannot measure charged particles as they are deflected by the strong magnetic fields of the hadron collider, and the LHCf detector is away from the beam line, in the forward direction. Hence, its data cannot directly constrain the production of the mesons and baryons that produce muons.

As a consequence, hadronic interaction models rely on different frameworks and assumptions to extrapolate from accelerator measurements into the relevant kinematic phase space for hadronic interactions in extensive air showers. Such a framework, common to all post-LHC hadronic interaction models, such as EPOS LHC-R [62], QGSJET-III.01 [63] and SIBYLL2.3e [64], is the Gribov-Regge field theory (GRFT). This theory treats the color-exchange between partons using a color-neutral phenomenological object called Pomeron. The amplitude for multiple Pomeron exchanges uniquely links the inelastic cross-section to particle production.

The mentioned models also employ other mechanisms that lead to differences in their predictions:

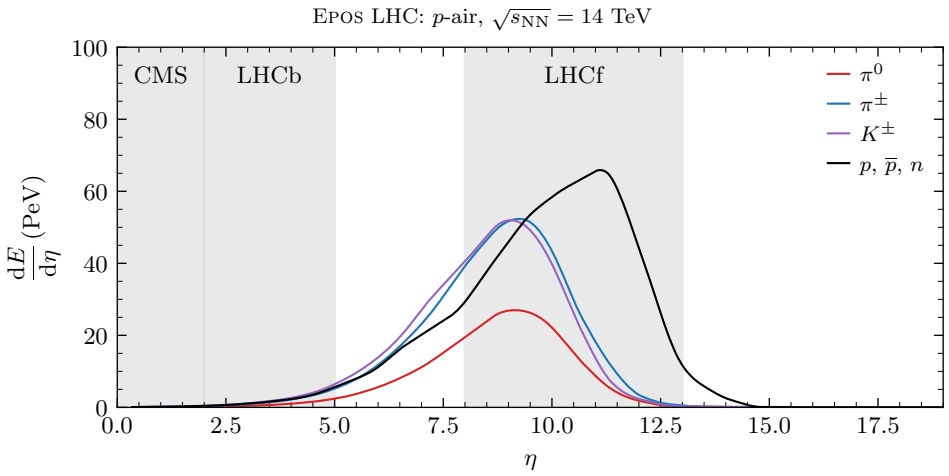

Figure 1.6: Differential energy flow for different secondary hadrons of proton-air interactions at $\sqrt{s_{NN}} = 14$ TeV, listed in the figure's legend. The pseudo-rapidity ranges covered by the CMS, LHCb and LHCf detectors are represented by the shaded grey bands. Note that only neutral particles are measured at the forward detector: LHCf.

- EPOS LHC-R: solves one of the main issues of energy conservation in GRFT by ensuring a consistent energy sharing in intermediate steps of hadronisation. It also accounts for collective effects in hadronisation via the core-corona approach [65, 66]. In this scheme, a core of Quark Gluon Plasma (QGP) formed in the high-density regions of the interaction is thermally hadronised accounting for correlations between particles. The corona is the low-density region where strings are fragmented in e^+e^- collisions. The core-corona approach is fit to pp data, and it naturally produces more baryon-anti-baryon pairs and strange particles such as Kaons. Moreover, and contrary to its older version. EPOS LHC [67], EPOS LHC-R includes hadronic re-scattering, that is, additional scattering of hadrons and resonances after parton hadronisation, influencing the final state particles that re-interact in a shower. Even though this effect does not influence the forward particle production, it is necessary to fit data at mid-rapidity, changing the connection between different regions of the kinematic phase-space

- QGSJET-III.01: It does not address the caveat of energy conservation in GRFT. Rather than parametrising the energy evolution of the parton density, it uses higher-order Pomeron graphs. These are resumed, ensuring conservation of unitarity.

- SIBYLL2.3e: similarly to QGSJET-III.01 it does not solve the energy conservation caveat of GRFT. It is solely designed for extensive air showers, and employes the mini-jet model [68] in treating high-momentum interactions, together with the Dual Parton Model [69].

Naturally, the mentioned hadronic interaction models must also support hadron-air and nucleus-air interactions. EPOS LHC-R and QGSJET-III.01 achieve this by extending the GRFT formalism, while SIBYLL2.3e uses the semi-superposition model. The latter model treats the interaction of a nucleus with A nucleons as A independent p-air interactions. The p-air interaction itself is treated by the classical Glauber [70] approach. Other

processes relevant for hadron-nucleus interactions, such as nuclear fragmentation, are also implemented in the generators (see [16] for a brief summary).

Additional important phenomena arise in p-air interactions, such as the diffractive dissociation of the air target. Such interactions are sometimes referred to as *diffractive* interactions. They represent up to $\approx 20\%$ of the value of the inelastic cross-section [16], and are essential to properly account for the elasticity κ_{el} of proton-air interactions. These interactions do not involve the exchange of quantum numbers between the beam and the target, only momentum transfer. In low-mass diffraction, the momentum transfer is of the order of a few GeV, while in high-mass diffraction, the momentum transfer can excite a Pomeron and produce a few secondary hadrons.

Finally, the different assumptions used by the models and the different datasets used for the tuning of their parameters lead to important differences in their predictions for the output of hadronic interactions in kinematic phase-space regions lacking accelerator constraints. As a result, even for a fixed primary mass composition, they lead to different predictions for air shower observables. Additionally, despite offering a consistent treatment of hadronisation, these models are phenomenological, so there is no guarantee that they can consistently describe extensive air shower data in terms of the primary characteristics.

1.4 Measuring extensive air showers: the Pierre Auger Observatory

Currently, the largest observatory of ultra-high-energy cosmic rays is the Pierre Auger Observatory [71]. Measuring extensive air showers since 2004, it follows other historically significant observatories such as Haverah Park [72] and AGASA (Akeno Giant Air Shower Array) [73], and coexists with the second largest cosmic ray observatory: the Telescope Array [74, 75].

The Pierre Auger Observatory is located in the Pampa Amarilla, near the city of Malargüe, Mendonza, Argentina, between latitudes 35.0° and 35.3° S and longitudes 69.0° and 69.4° W, at an average altitude of $\simeq 1\,400$ m above sea level. It consists of a triangular grid of 1 660 water-Cherenkov detectors (WCDs) covering an area of $3000\,\mathrm{km}^2$ overlooked by 27 fluorescence telescopes at four different sites. The set of WCDs is known as the Surface Detector array (SD) and the set of fluorescence telescopes is the Fluorescence Detector (FD). The Observatory is a hybrid instrument that combines information from both detection techniques to determine the energy and arrival direction of UHECRs through EAS measurements. The reconstructed shower observables can then be interpreted using hadronic interaction models to infer the mass composition of the cosmic-ray flux. The layout of the observatory, together with a schematic measurement of a real event, is shown in Figure 1.7.

1.4.1 The Fluorescence Detector

The Fluorescence Detector is composed of 24 fluorescence telescopes [77], six at each FD site, and three high-elevation telescopes known as HEAT (High Elevation Auger Telescopes) [71]. These telescopes are equipped with cameras with 440 photomultiplier tubes

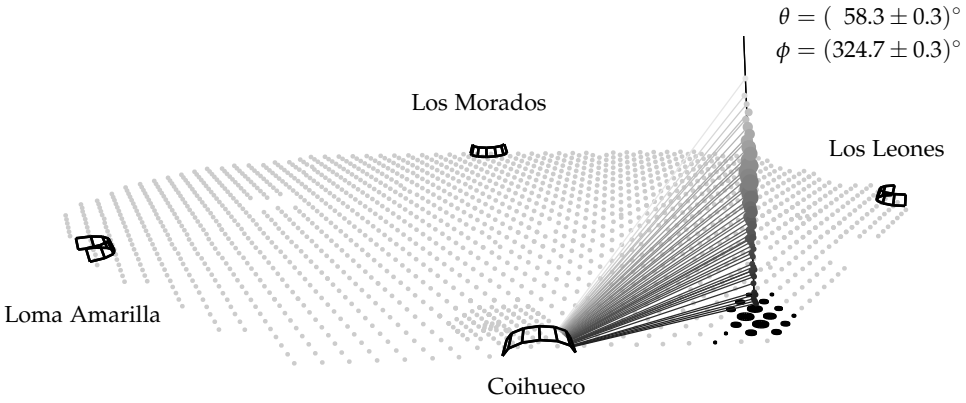

Figure 1.7: Schematic view of a measured event at the Pierre Auger Observatory, with event ID 15346477. The Surface Detector array is shown as a triangular grid of small grey dots. The four sites equipped with Fluorescence Detectors—Loma Amarilla, Coihueco, Los Leones, and Los Morados—are labeled accordingly. Water-Cherenkov Detectors triggered by the event are highlighted in black, with marker sizes proportional to the recorded signal. Slanted lines connect points along the reconstructed shower axis to the corresponding triggered pixels in the Fluorescence Detector cameras. Fluorescence emissions occurring earlier in the shower development are shown in lighter shades of grey, while later emissions, closer to the ground level, appear in darker tones. The reconstructed geometry of the shower is indicated by its zenith and azimuth angles, θ and ϕ. The figure was taken from [76], and is available on the repository `arxiv` under the Creative Commons license CC BY-SA 4.0 which can be accessed at `https://creativecommons.org/licenses/by-sa/4.0/`.

or pixels, covering a field of view of about 30° in elevation and 30° in the azimuthal direction. The telescopes measure the isotropic ultraviolet emission of nitrogen molecules excited by charged shower particles. Therefore, they can only operate in clear moonless nights with good atmospheric conditions, having a duty cycle of about 15%. The time evolution of the emission yield can be converted to the calorimetric energy deposited by the shower at each depth, $\mathrm{d}E_{\mathrm{cal}}/\mathrm{d}X$ [76], enabling the direct imaging of the longitudinal development of the shower in the atmosphere. An example of such a longitudinal profile of the calorimetric energy is shown in Figure 1.8.

The longitudinal profile is then fitted to a re-parametrised Gaisser-Hillas function [79], known as the Universal Shower Profile (USP) fit function [80]:

$$\frac{\mathrm{d}E_{\mathrm{cal}}}{\mathrm{d}X} = \left.\frac{\mathrm{d}E_{\mathrm{cal}}}{\mathrm{d}X}\right|_{\mathrm{max}} \left(1 + \frac{R}{L}\left(X - X_{\mathrm{max}}\right)\right)^{1/R^2} \exp\left\{\frac{X_{\mathrm{max}} - X}{RL}\right\}, \tag{1.50}$$

where R, L, $\left.\mathrm{d}E_{\mathrm{cal}}/\mathrm{d}X\right|_{\mathrm{max}}$ and X_{max} are free parameters. The total calorimetric energy E_{cal} is obtained via the integration of Equation (1.50). The reconstructed primary energy E_{FD} is then estimated as $E_{\mathrm{FD}} = E_{\mathrm{cal}} + E_{\mathrm{inv}}$, where E_{inv} is the *invisible energy*, i.e, the energy carried by neutrinos and muons that is not deposited in the atmosphere. This data-driven correction to the calorimetric energy is based on the Heitler-Matthews model and accounts for about $10 - 15$ % of the primary energy [81].

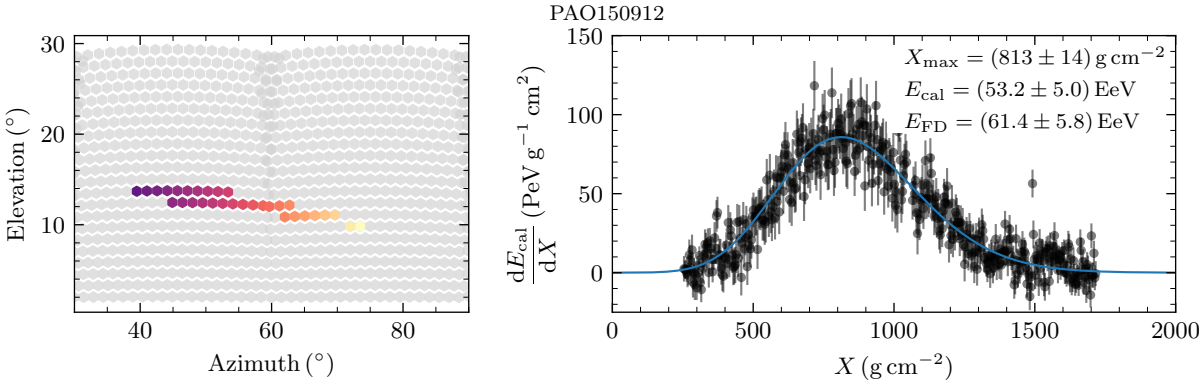

Figure 1.8: Left: Time sequence of triggered pixels of 2 FD cameras measuring the fluorescence emission of Auger event with ID PAO150912. The color of the triggered pictures ranges from dark purple to light yellow, indicating early and late trigger times, respectively. Background pixels are shown as a hexagonal grey grid. Right: fitted calorimetric energy deposited by the shower as a function of slant depth X in the atmosphere. Black markers represent the measured deposited energy in the FD pixels, while the blue curve is a fit to Equation (1.50). The extracted values of X_{max}, calorimetric energy, E_{cal}, and primary energy, E_{FD} are indicated in the legend. The data used to produce this figure is available at `https://opendata.auger.org/catalog/index.php?evid=150912` and presented in [78].

1.4.2 The Surface Detector array

The surface detector array [82] samples the particle content of the shower at the ground by recording the time structure of the Cherenkov emission of superluminal shower particles in the water volume of WCDs. The Cherenkov signal is converted to an electric signal using three 9-inch Photomultiplier Tubes (PMTs). The electric signal is then discretised in bins of 25 ns and calibrated in units of Vertical Equivalent Muons (VEM). That is, the average charge induced by a single vertical muon traversing the center of the WCD. This calibration is performed *in situ* using atmospheric muons.

Each WCD possesses a trigger system capable of discerning typical signals produced by shower particles from other signal sources. If a certain number of WCDs is triggered in coincidence and in particular geometric configurations, an SD event is registered. This event must then pass a hierarchical chain of sophisticated triggers to be identified as an extensive air shower and not the result of random coincidences. This chain terminates in a set of quality triggers that ensure the high quality of the reconstructed events used in various analyses. An example of such an event, taken from the Auger Public Data Catalogue [83] is shown in Figure 1.9.

The entire area of the SD array is covered by WCDs spaced by 1500 m, making up the SD-1500 array. Denser array areas, designed to detect lower energy showers, are composed of stations spaced by 750 m and 433 m, constituting the SD-750 and SD-433 arrays. The different arrays have different trigger efficiencies for showers of a given energy. Lower energy showers produce smaller footprints, which can only be measured if stations are sufficiently close to be sensitive to the low density of shower particles at the ground. The energy threshold for full detection efficiency is $\sim 10^{16.5}$ eV for SD-433, $\sim 10^{17.5}$ eV for SD-750 and $\sim 10^{18.4}$ eV for SD-1500.

The shower geometry defined by the local zenith and azimuth angles, θ and ϕ, re-

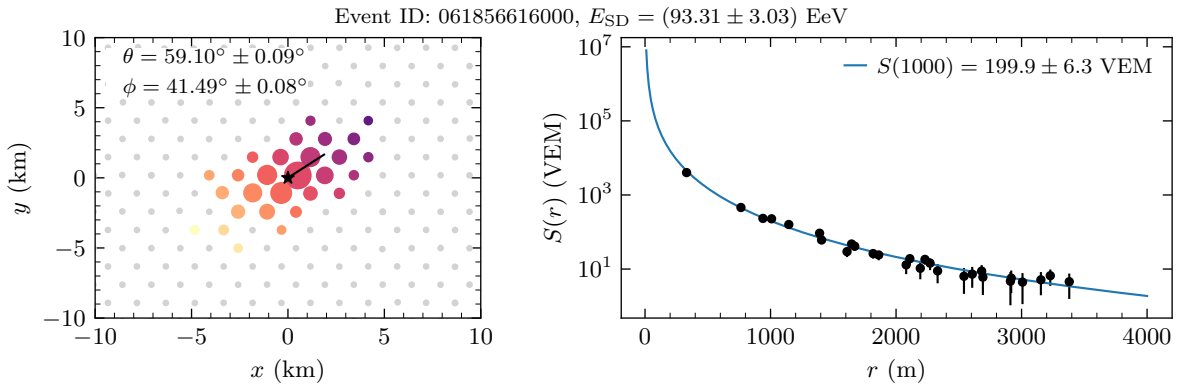

Figure 1.9: Left: Time sequence of WCD stations triggered by the Auger event with ID 61856616000. Marker color ranges from dark purple (early) to light yellow (late). Marker size scales with the logarithm of the measured signal. The star indicates the shower core, and the black line shows the ground-projected shower axis. Reconstructed zenith and azimuth angles are denoted by θ and ϕ. Non-triggered WCDs are shown in grey. Right: Lateral distribution of measured signals (black markers) and best-fit function (blue) from Equation (1.51). The variable r denotes the distance to the shower core in the plane perpendicular to the shower axis. The fitted $S(1000)$ is shown in the legend. The estimated energy E_{SD} appears in the title. Data from https://opendata.auger.org/data.php, presented in [83].

spectively, is reconstructed by fitting the starting times of the WCD signals to a curved shower front travelling at the speed of light. This fit also determines the impact point of the shower: the intersection of the shower axis with the ground, the so-called shower core. The time-integrated signals measured by triggered WCDs at positions $\mathbf{r}$ from the shower core are then used to build the lateral distribution function (LDF) of the WCD signal $S(\mathbf{r})$. In *vertical showers*, that is, in showers with $\theta < 60°$, this signal is dominated by electromagnetic particles and muons, while in *horizontal showers*, $60° < \theta < 80°$, the signal is mostly induced by muons. This is due to the attenuation of the electromagnetic component over the large atmospheric slant depths traversed by inclined showers. Moreover, in vertical showers, the particle density at ground is approximately axially-symmetric, so that the signal measured by the WCD is mostly a function of $r \equiv |\mathbf{r}|$. This lateral distribution is then fitted to the NKG function [84]:

$$S(r) = S(r_0) \left(\frac{r}{r_0}\right)^{\beta} \left(\frac{r + r_1}{r_0 + r_1}\right)^{\beta+\gamma} \quad \text{with} \quad r_1 = 700\,\text{m}, \tag{1.51}$$

by maximising a likelihood that combines the information from saturated stations near the shower core, triggered unsaturated stations and non-triggered stations. From this fit, the shower size, $S(r_0)$, can be extracted, together with the slope parameters β and γ. For the SD-1500 array, this shower size is denoted by $S(1000)$ (see the right panel of Figure 1.9). After the so-called constant intensity cut (CIC) [85], which accounts for the data-driven zenith-dependent attenuation of $S(1000)$, permits the computation of the event-by-event S_{38}: the value of shower size $S(1000)$ if the shower had $\theta = 38°$. This value is sensitive to the normalisation of the LDF, which in turn is highly sensitive to the primary energy E_0. Therefore, S_{38} is calibrated to the FD energy scale using a collection of high-quality events measured by both detectors: the golden-hybrid showers. Therefore,

all SD events are assigned an energy. Since the SD has a duty cycle of $\sim 100\%$, this allows the estimation of the cosmic ray spectrum with high statistical precision.

Furthermore, using Deep Neural Networks and frameworks like air shower Universality [86, 87], SD data can be used to estimate $X_{\max}$ [88] and the muon content of EAS [89, 90].

1.4.3 The detectors of AugerPrime

Currently, the Pierre Auger Observatory is concluding an upgrade known as *AugerPrime* [91]. The goals of this upgrade include: improving the precision of primary mass discrimination, enhancing the sensitivity to very-high-energy photon and neutrino searches, and improving the study of hadronic interactions. This is achieved through a new multi-hybrid design, which augments the available information per cosmic-ray event.

As a part of the upgrade, scintillator panels were placed on top of each WCD, constituting the Scintilator Surface Detector (SSD). The combination of the responses of the WCDs and SSDs to the electromagnetic and muonic components of the shower enables the inference of the shower's muon content, which is a powerful mass discriminator. New and faster electronics were installed in WCDs, permitting the sampling of the shower front with greater time resolution, possibly improving the training of Neural Networks using information from SD traces to estimate $X_{\max}$ and the contribution of the muons to these traces. Moreover, radio antennas were installed in each surface detector station, as part of the *Radio* upgrade. The energy fluence of the radio emission of a shower in the atmosphere can be measured by these antennas and used to estimate the primary energy. These radio antennas are mostly sensitive to inclined showers and have a duty cycle of 100%, enabling the estimation of primary energy for all inclined SD events. As such, the large collection of events with energy and muon content reconstructed independently will allow the high-precision measurement of several moments of the muon number distribution, providing powerful handles on hadronic interactions.

Moreover, as a part of the AMIGA (Auger Muons and Infill for the Ground Array) extension, Underground Muon Detectors (UMD) are being placed below all WCDs of the SD-750 array. These $20\,\mathrm{m}^2$ scintillators are buried under 2.3 m of soil, corresponding to a mass overburden of $\sim 540\,\mathrm{g\,cm}^{-2}$. This effectively shields the scintillators from most of the electromagnetic component of the shower and from muons below ~ 1 GeV, enabling the measurement of the muon lateral distribution function: $\rho_\mu(r)$. From this measurement, the muon scale can be calculated with high precision and then used in analyses that discriminate photon from hadron-induced showers or to compare the mass composition derived from muon measurements with that inferred from other shower observables, to probe inconsistencies in the description of hadronic interactions.

1.4.4 Inconsistent mass interpretation from different shower observables

Direct measurements of $X_{\max}$ using FD data [31] and Deep Neural Networks analysis using SD data [88, 92] show that for primary energies $E_0 > 3 \times 10^{19}\,\mathrm{eV}$ the interpreted composition of the cosmic ray flux becomes heavier and purer.

However, both direct muon measurements using UMDs [93] and indirect estimations

of the muon scale using inclined showers [94, 95], show that the number of muons at ground is between 30% and 45% larger than the predictions of the hadronic interaction models EPOS LHC [67], QGSJET-II.04 [96] and SIBYLL2.3d [64], if the composition derived from $X_{\rm max}$ measurements is assumed. Measurements of the muon scale using air shower universality show a similar trend [97]. To compare these results with each other and with those obtained by other experiments, which use different observables to estimate the muon content of EAS, we can define the quantity

$$z = \frac{\ln \langle x \rangle - \ln \langle x \rangle_p}{\ln \langle x \rangle_{\rm Fe} - \ln \langle x \rangle_p}, \tag{1.52}$$

where x is the value of the muon proxy, $\langle x \rangle$ is the data value, $\langle x \rangle_p$ is the expected value for a pure proton composition and $\langle x \rangle_{\rm Fe}$ is the expected value for a pure iron composition. Further defining $z_{\rm mass}$ as the z-scale value assuming the composition derived from $X_{\rm max}$, it follows that values of $\Delta z = z - z_{\rm mass}$ should be compatible with zero, for a consistent description of mass-sensitive air shower observables. However, it is clear from the left panel of Figure 1.10 that Δz increases with the primary energy, becoming incompatible with $X_{\rm max}$ expectations at the highest energies, for most of the measured data. This deficit in simulations is the so-called *Muon Puzzle* [16].

A clue about the origin of the muon puzzle was provided by measurements of the fluctuations of the muon content of inclined showers detected by Auger [95]. These are compatible with hadronic-interaction-model predictions assuming the mass composition derived from $X_{\rm max}$ (see the right panel of Figure 1.10). As shown in [53], the fluctuations of the muon content are mostly sensitive to the shape of the energy spectrum of hadrons produced in the primary interaction. Therefore, this result suggests that the Muon Puzzle arises from a cumulative mismatch in the description of hadronic interactions throughout the cascade, rather than from the appearance of exotic physics at the highest energies. This is somewhat supported by the increase in the muon deficit in simulations with the primary energy.

Several mechanisms can increase the muon content of EAS, such as an increase in baryon production, enhanced production of the vector meson ρ^0 [64] and strangeness enhancement [99]. However, this latter mechanism seems to be disfavoured by the latest results of FASER [100]. The challenge, however, is to consistently describe all shower observables. As an example, EPOS LHC predicts a larger muon scale than QGSJET-II.04, being in less tension with Auger data. However, Auger measurements suggest that EPOS LHC overestimates the mass-sensitive depth of maximum of the *apparent* muon production depth (MPD) distributions: $X^{\mu}_{\rm max}$ [101]. Note that the term *apparent* refers to the fact that the production depth distribution is derived from the muons arriving at the ground, rather than for all the muons produced in the shower. The tension between the mass composition derived from $X_{\rm max}$ and $X^{\mu}_{\rm max}$ is represented in the left panel of Figure 1.11.

Recently, the joint distributions of $X_{\rm max}$ and $S(1000)$ measured by Auger were fitted to Monte Carlo templates that allowed freedom in $\langle X_{\rm max} \rangle$ and in the normalisation of the muon signal $R_{\rm had}$, simultaneously for different zenith angles [102]. The analysis was performed for primary energies $E_0 \in [10^{18.5}, 10^{19.0}]$ eV. As shown in the right panel of Figure 1.11, the best fit to Auger data is achieved by deepening $\langle X_{\rm max} \rangle$ by $\approx 22{-}47\,{\rm g\,cm^{-2}}$,

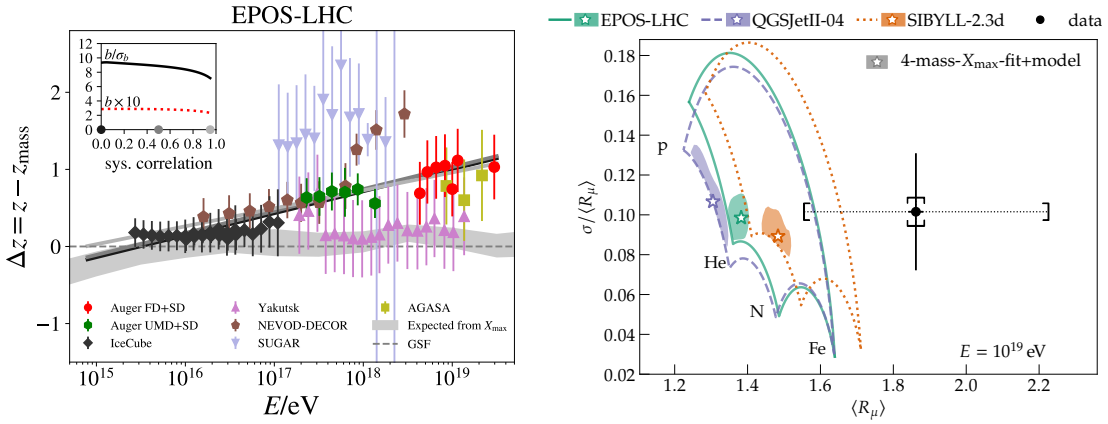

Figure 1.10: Left panel: Linear fits to the $\Delta z = z - z_{\rm mass}$ distributions. The z values are computed for different experiments based on their muon-proxy measurements. All references to the data can be found in [16]. The measurements are compared to the mass-expectation from $X_{\rm max}$ and to the predictions of the Global Spline Fit model [98]. Shown in the inset are the slope, b, and its deviation from zero in standard deviations for an assumed correlation of the uncertainties within each experiment. Figure was taken from [16], an *open access* article under the Creative Commons license CC BY-SA 4.0 which can be accessed at https://creativecommons.org/licenses/by-sa/4.0/. Right panel: Data (black points with error bars) and model predictions for the average and fluctuations of the number of muons in air showers with a primary energy of $\simeq 10^{19}$ eV. Statistical uncertainties are indicated by vertical error bars, while the total systematic uncertainty is represented by square brackets. The colored contours depict the range of predictions from hadronic interaction models for any mixture of four primary mass components: p, He, N, and Fe. Star symbols denote the predictions from the composition mixtures inferred from $X_{\rm max}$ measurements. The shaded regions represent the combined statistical and systematic uncertainties associated with the $X_{\rm max}$-based composition derived with EPOS LHC. The figure was taken from [95], an *open access* article under the Creative Commons license CC BY-SA 4.0 which can be accessed at https://creativecommons.org/licenses/by-sa/4.0/.

depending on the considered hadronic interaction model, and applying a rescaling $R_{\rm had}$ between 15% and 24%, depending on the hadronic interaction model and also on the zenith angle of the shower. In particular, the hypothesis of no shift in and no re-scaling of the muon component is excluded at more than 5σ. These results suggest a reinterpretation of the mass composition of the cosmic-ray flux: the primary mass must be heavier than what is currently inferred with air-shower simulations, somewhat alleviating the muon deficit.

Crucially, the latter result was derived by combining measurements of different shower observables, imposing tighter constraints on hadronic interactions at the highest energies.

1.5 Constraints on hadronic interactions from extensive air showers

The multiple inconsistencies in the description of the extensive air showers motivate the study of hadronic-interaction-model independent connections between shower observables

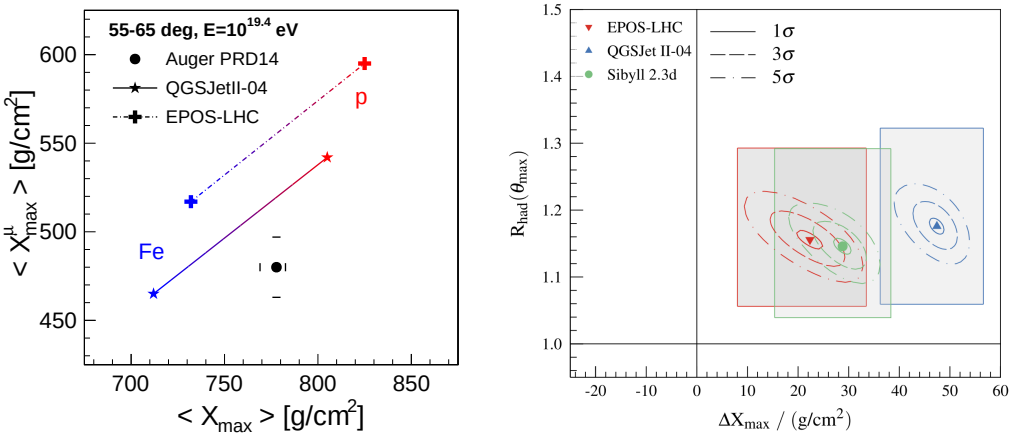

Figure 1.11: Left: measurement of the mean of the depth of maximum of the *apparent* MPD distribution, taken from [101], against the mean value of $X_{\max}$, taken from [76], for primary energy $E_0 = 10^{19.4}$ eV and $\theta \in [55°, 65°]$. The measurement is compared to the predictions of two hadronic interaction models for several mixtures of primaries, ranging from pure protons (red) to pure iron nuclei (blue). The figure was taken from [41], and is available on the repository `arxiv` under the Creative Commons license CC BY-SA 4.0 which can be accessed at `https://creativecommons.org/licenses/by-sa/4.0/`. Right: Optimal shift in $\langle X_{\max} \rangle$ and re-scaling of the muon signal, $R_{\rm had}(\theta_{\max} \approx 55°)$, derived from the fit of the joint distribution of $X_{\max}$ and $S(1000)$ obtained in [102]. The optimal values for the different hadronic interaction models are indicated by the symbols. The contours refer to the statistical uncertainty of the optimal values derived from the fitting algorithm, while the grey shaded regions represent the systematic uncertainties. The figure was taken from [102], an *open access* article under the Creative Commons license CC BY-SA 4.0 which can be accessed at `https://creativecommons.org/licenses/by-sa/4.0/`.

and relevant quantities describing hadron-air interactions. These quantities are generically called *multiparticle production variables*, and include the already defined multiplicity, $m_{\rm tot}$, elasticity, $\kappa_{\rm el}$, fraction of hadronic energy, $\alpha_{\rm had}$, and the primary interaction shower-to-shower estimator of the relative muon content: α_1. Another important parameter is the so-called charge ratio: the fraction of neutral pions relative to the total multiplicity of secondary particles.

The impact of changing the interaction parameters $m_{\rm tot}$, $\kappa_{\rm el}$, π^0-fraction and the proton-air cross-section, $\sigma_{p-\rm air}$ on the first and second moments of the distributions of N_μ and $X_{\max}$, was studied extensively in [55]. These parameters were changed by a multiplicative factor $f(E, f_{19})$ given by

$$f(E, f_{19}) = 1 + (f_{19} - 1) \times \begin{cases} 0 & \text{if} \quad E < E_{\rm th} \\ \dfrac{\log_{10}(E/E_{\rm th})}{19 - \log_{10}(E_{\rm th}/\text{eV})} & \text{if} \quad E \geq E_{\rm th} \end{cases} . \tag{1.53}$$

In practice, the $f(E, f_{19})$ scaling is implemented by resampling the particle arrays produced in SIBYLL2.1 [103], all the way down to a certain energy threshold, where hadronic interaction models are assumed to converge. This *ad hoc* resampling conserves kinematic quantities and approximately preserves some quantum numbers. The results of this procedure are depicted in Figure 1.12.

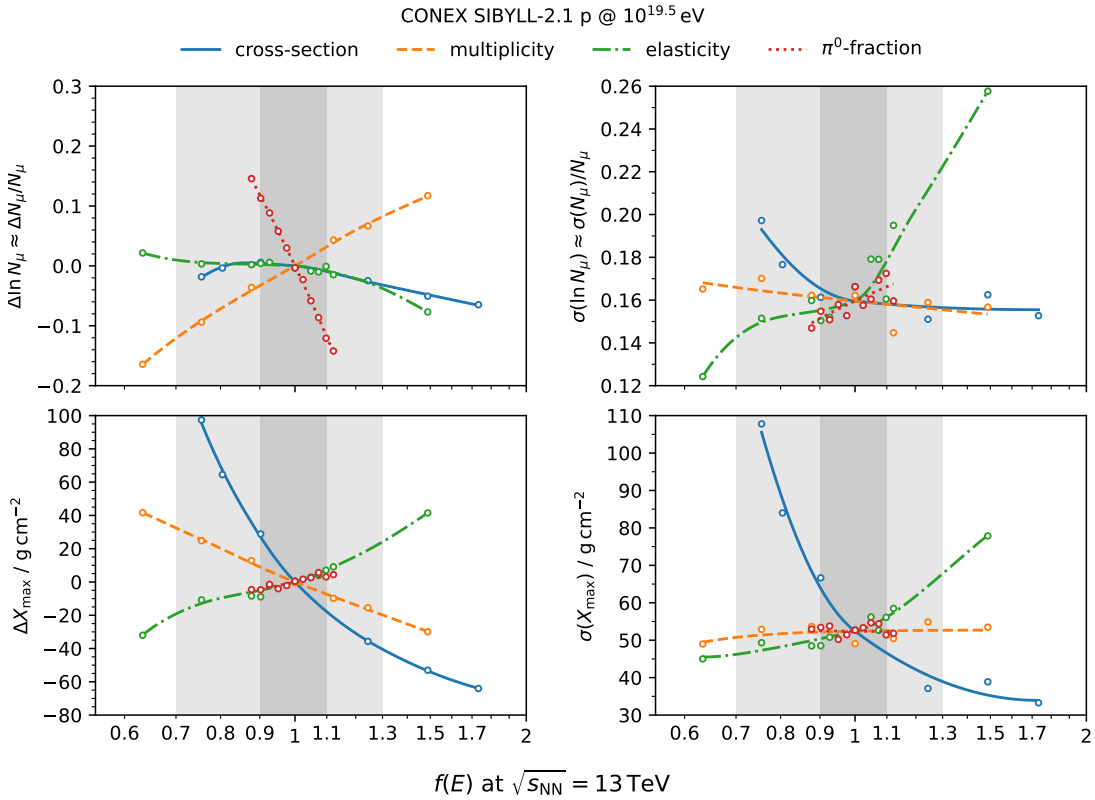

Figure 1.12: Left panels: absolute and relative difference in the expected values of $X_{\max}$ (bottom left) and $\ln N_\mu$ (top left), respectively, as a function of the scaling factor defined in Equation (1.53) multiplying different production variables. Right panels: evolution of the standard deviations of $\ln N_\mu$ (top right) and $X_{\max}$ (bottom right) with the modification factor defined in Equation (1.53). The modified production variables are indicated in the figure's legend and defined in the main text. The modifications were implemented in Sibyll2.1 for proton-induced showers at $E_0 = 10^{19.5}$ eV. The figure was taken from [16], an *open access* article under the Creative Commons license CC BY-SA 4.0 which can be accessed at https://creativecommons.org/licenses/by-sa/4.0/. The figure is based on the work presented in [55],

Decreasing multiplicity and increasing elasticity are very effective ways of deepening $\langle X_{\max} \rangle$, as the energy of the primary degrades more slowly in the atmosphere. Decreasing the proton-air cross-section deepens $\langle X_{\max} \rangle$ while greatly increasing its fluctuations, mainly due to the increased fluctuations of X_1, whose standard deviation, $\sigma(X_1) = \lambda_{p-\text{air}}$ (this follows directly from Equation (1.45)) changes as $\lambda_{p-\text{air}} \to f(E, f_{19})^{-1} \lambda_{p-\text{air}}$. Within the considered modifications, the most efficient way of increasing the muon scale is to decrease the π^0-fraction.

Since the resampling applied in this procedure is rather artificial, there is no guarantee that every combination of modifications leads to a physically consistent model of hadronic interactions. Furthermore, the main moments of the distributions of N_μ and $X_{\max}$ are affected by the mixture of primary masses. Thus, the application of this procedure to data would couple modifications of the physical parameters with the mass interpretation of the cosmic ray flux. Moreover, there is no proof that the modified parameters are the most efficient and consistent way of affecting the N_μ and $X_{\max}$ distributions. Lastly, this

procedure is rather computationally costly.

Despite these caveats, the method allowed for the measurement of proton-air cross-section using data collected by the Pierre Auger Observatory [56]. This result follows the pioneer measurements of the Fly's Eye observatory [104], and was succeeded by Telescope Array measurements [105]. All the results agree with hadronic interaction model extrapolations in the range $\sqrt{s} \sim 50\,\mathrm{TeV}$.

Finer modifications were explored in [106], where the distribution of the number of muons for muon-depleted showers was modified via an *ad hoc* resampling of the neutral pion spectrum of the primary interaction. This exercise is shown in the left panel of Figure 1.13. The softening of the spectrum decreases the likelihood of producing showers with low muon content. Therefore, measurements of the tail of the distribution of the number of muons towards low values constrain the differential inclusive cross-section for the production of neutral pions in p-air interactions at the highest energies. Importantly, this measurement has little sensitivity to the presence of heavier primaries in the cosmic-ray flux.

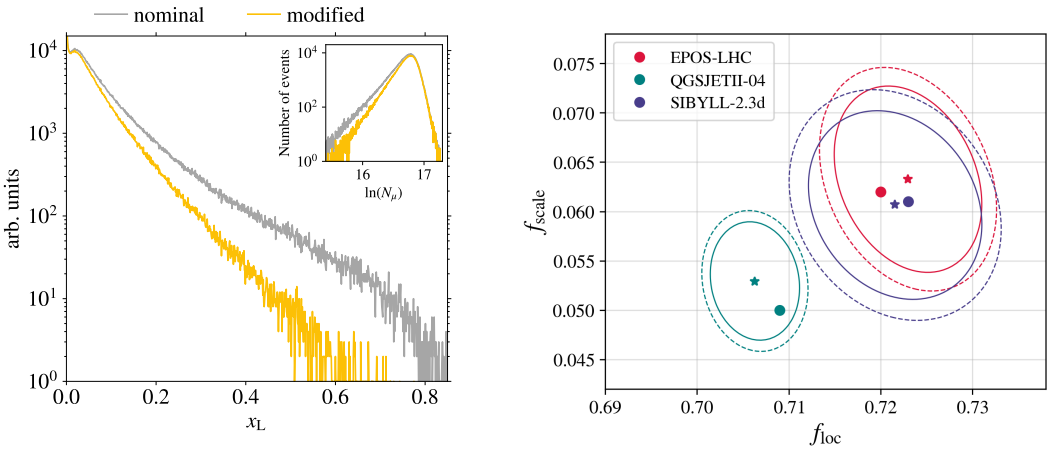

Figure 1.13: Left panel: Lab. frame energy spectrum of neutral pions produced in the primary proton-air interaction at $E_0 = 10^{19}$ eV before (grey) and after (yellow) being artificially softened by an *ad hoc* resampling procedure. The inset shows the corresponding modification of the shape of the muon number distribution. The employed hadronic interaction model was Sibyll2.3e and shower simulations were produced with Conex. The figure was taken from [106]. Right panel: nominal (full circles) and estimated (starred markers) predictions of the mean value, f_{loc}, and standard deviation, f_{scale}, of the energy-weighted average of the fraction of hadronic energy. The contours indicate the 1σ and 2σ uncertainties of the fit, and the estimated values are determined by a neural network applied to each hadronic interaction model. The figure was taken from [107], and is available on the repository `arxiv` under the Creative Commons license CC BY-SA 4.0 which can be accessed at `https://creativecommons.org/licenses/by-sa/4.0/`.

In [107], fully connected Neural Networks were used to convert combinations of multiparticle production variables into distributions of N_μ and X_{max}. The algorithm allows for a quick conversion of modified distributions of production variables into modified distributions of the mentioned shower observations, therefore enabling fitting the optimal moments of production parameters for a given measured distribution of N_μ and X_{max}. The validation of this fit is shown in the right panel of Figure 1.13, where the mean, f_{loc}, and the standard deviation, f_{scale}, of the fraction of hadronic energy in the first few shower

generations are fitted to the distributions of the number of muons and $X_{\max}$.

Although this approach seemed promising, the uniqueness of the found solutions and their dependence on the hadronic interaction model were not fully explored. Additionally, only the first and second moments of the distributions of N_μ and $X_{\max}$ were again considered, not taking advantage of all available information in their distributions, in particular, the correlations between these observables. Finally, the choices of the variables used in the fit, combining primary interaction information with that of deeper interactions, were not entirely physically motivated, and do not factorise the importance of the primary and deeper interactions in describing different features of the distributions of N_μ and $X_{\max}$.

A more global, physically motivated approach that takes into account all the information in the primary interaction in a consistent fashion and all the information in the full distribution of N_μ and $X_{\max}$ is lacking. This would allow for strong constraints on hadronic interactions in kinematic regimes not probed by human-made accelerators, using EAS data. This approach could then alleviate the tensions in the description of air shower observables and lead to the discovery of important physics mechanisms appearing at the highest energies.

1.6 Objectives

The main goal of this thesis is to infer properties of proton–air interactions at ultra-high energies using the information contained in the full two-dimensional distribution of two key shower observables: the depth of the shower maximum, $X_{\max}$, and the number of muons at ground level, N_μ. It is of fundamental importance to establish a framework that enables the probing of these interactions using air-shower data, rather than relying solely on the assumptions of the different hadronic interaction models. Such a framework can help alleviate the tensions between data and model predictions and provide better constraints on the sources of ultra-high-energy cosmic rays.

The main goal of the thesis is achieved through the use of Monte Carlo simulations of EAS and the development of semi-analytical models, by fulfilling the following objectives, each generally corresponding to a given chapter of the thesis:

1. Understand the effects that shape the two-dimensional distribution of N_μ and $X_{\max}$, and decouple those stemming from the stochasticity of hadron production from those arising from known effects of the cascade dynamics. To this end, the aforementioned two-dimensional distribution is studied as a function of the energy, direction of arrival, and mass composition of the cosmic ray triggering the cascade;

2. Study the $X_{\max}$-dependence of the shape of the muon number distribution for muon-depleted showers, to derive constraints on the inclusive cross section for the production of neutral pions in proton–air interactions;

3. Develop a probabilistic framework that maps the shape of the energy spectrum of secondary hadrons from the primary interaction onto the shape of the joint distribution of N_μ and $X_{\max}$, with minimal reliance on hadronic interaction models. This involves the derivation of multiparticle production variables, the modelling of the probabilistic response function of the shower to the primary interaction and a

thorough assessment of the precision and hadronic-interaction-model dependence of this response;

4. Provide the first measurements of the moments of the energy spectrum of secondary hadrons from the primary interaction using the recent results derived from Auger data and published in [102].

2

CHARACTERIZATION OF THE JOINT DISTRIBUTION OF N_μ AND $X_{\max}$

In this chapter, we describe the methodology used in this work and introduce the joint probability density function of the depth of shower maximum, $X_{\max}$, and the number of muons at ground level, N_μ: $f(N_\mu, X_{\max})$. We first describe the main features of the marginal distributions of these shower observables and motivate the study of their correlation. We then discuss key physical effects that shape the joint distribution $f(N_\mu, X_{\max})$ as a function of primary energy, zenith angle, and primary mass composition. These include the attenuation of the muon flux, modelled using a simplified approach, and the fluctuations induced by the truncation of the muon profile by the ground plane. We show that a simple superposition model does not fully capture the dependence of $f(N_\mu, X_{\max})$ on the primary mass number A, in particular due to the role of nuclear fragmentation. The connection between $f(N_\mu, X_{\max})$ and particle production in the primary interaction is deferred to later chapters. Finally, we demonstrate that, for state-of-the-art cosmic ray experiments, the energy threshold for muon detection has a minor impact on the form of $f(N_\mu, X_{\max})$.

2.1 Methodology

The shower observables analysed in Chapters 2 through 6, along with the multiparticle production variables introduced in later chapters, are obtained from several libraries of extensive air shower simulations. These libraries were generated using the simulation code CONEX versions 7.50 and 7.80 [45, 46]. Version 7.80 is used to simulate showers with up-to-date high-energy hadronic interaction models: EPOS LHC-R [62], QGSJET-III.01 [63], and SIBYLL2.3e [64]. Version 7.50 is employed to simulate showers with earlier versions of these models, namely EPOS LHC [67], QGSJET-II.04 [96], and SIBYLL2.3d [64].

Each simulation library is produced for fixed values of the primary energy E_0, zenith angle θ, and primary mass number. Particles and their interactions are tracked individually down to a transition energy defined as $E_{\text{trans}} = 0.005 \times E_0$. Below this threshold, the longitudinal shower development is calculated by numerically solving one-dimensional cascade equations.

The ground level is set to 1400 m above sea level, corresponding to the average altitude of the Pierre Auger Observatory [71], and to an average vertical atmospheric depth of $X_{\text{gr}} = 880\,\text{g}\,\text{cm}^{-2}$. The longitudinal profiles include all particles with energies above

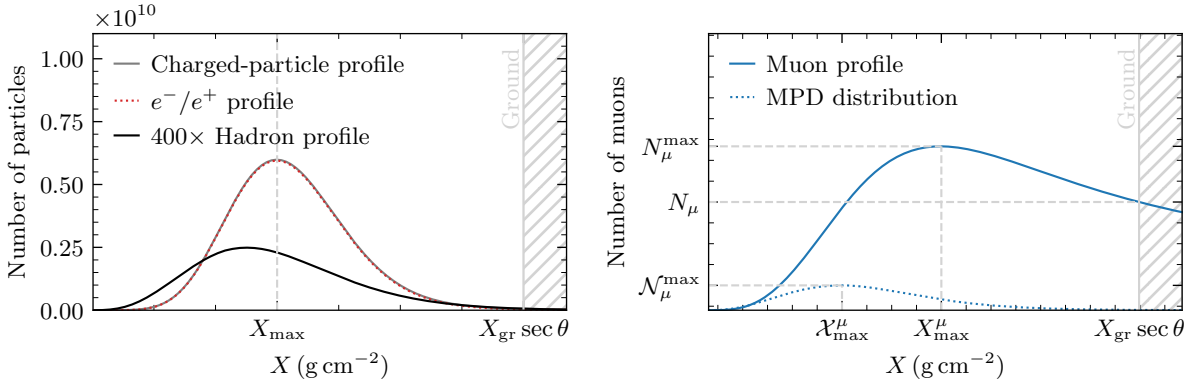

Figure 2.1: Left panel: Longitudinal profiles of the total number of charged particles (grey), e^-/e^+ (dotted red) and charged hadrons (black). The position of $X_{\max}$ is indicated by the vertical grey dashed line. Right panel: Longitudinal profile of the total number of muons (solid blue), and the Muon Production Depth (MPD) distribution (dotted blue). The main shower observables extracted from these longitudinal profiles are also depicted. Both panels show the ground level, at position $X_{\rm gr}\sec\theta$.

specific thresholds: 1 GeV for hadrons and muons, and 1 MeV for electrons, positrons, and photons.

Below, we introduce the main shower observables used in this analysis and describe how they are computed from the raw output of CONEX:

- $X_{\max}$: the atmospheric depth at which the longitudinal profile of all charged particles reaches its maximum. This profile is predominantly shaped by electrons and positrons, as illustrated in the left panel of Figure 2.1. The shower-by-shower value of $X_{\max}$ is extracted from a χ^2 fit to the number of charged particles as a function of atmospheric depth, $N(X)$, using the Gaisser–Hillas parametrization [79]

$$N(X) = N_{\max}\left(\frac{X - X_0}{X_{\max} - X_0}\right)^{\frac{X_{\max}-X_0}{\lambda}} \exp\left\{\frac{X_{\max} - X}{\lambda}\right\}, \tag{2.1}$$

  where X denotes the slant atmospheric depth, and the fit parameters are the maximum particle number $N_{\max}$, the free parameter X_0 (which is not the depth of the first interaction point), and the interaction length λ[1].

- $\mathcal{N}_\mu^{\max}$ and $\mathcal{X}_{\max}^\mu$: these denote, respectively, the number of muons produced at the maximum of the muon production depth (MPD) distribution, and the atmospheric depth at which this maximum occurs (see the right panel of Figure 2.1). These quantities are extracted by fitting a quadratic function to the number of produced muons, $N_\mu^{\rm prod}$, as a function of the traversed atmospheric slant depth X, only in the vicinity of the MPD peak at $(\mathcal{X}_{\max}^\mu, \mathcal{N}_\mu^{\max})$. Initial estimates for the fit are derived from the binned MPD histogram provided by the CONEX output: $\mathcal{N}_\mu^{\max}$ is initialized to the maximum bin content, and $\mathcal{X}_{\max}^\mu$ to the center of the corresponding bin. The fit is performed over a restricted range of $\pm 100\,\mathrm{g\,cm^{-2}}$ around the initial $\mathcal{X}_{\max}^\mu$ value. This simplified fit function was chosen over more complex parametrizations proposed

[1]Note that in CONEX λ is modelled as a quadratic function of depth: $\lambda = p_1 + p_2 X + p_3 X^2$.

in Refs. [108, 109] due to its faster convergence and the reduced sensitivity to the choice of initial parameter estimates. Moreover, a detailed modelling of the full MPD shape is not required for the purposes of this analysis.

- $N_\mu^{\max}$ and $X_\mu^{\max}$: analogous to ${\mathcal{N}_\mu}^{\max}$ and $\mathcal{X}_\mu^{\max}$, but extracted from the longitudinal profile of the number of muons: the total number of muons at each traversed depth, $N_\mu(X)$. This profile is directly read from the CONEX output (see the right panel of Figure 2.1) and can be related to the MPD distribution through the cascade equation (see Equation (1.19))

$$\begin{aligned}\frac{\mathrm{d}N_\mu}{\mathrm{d}X} = \frac{\mathrm{d}N_\mu^{\mathrm{prod}}}{\mathrm{d}X} &- \int_{E_\mu^{\mathrm{th}}}^{\infty} \left(\frac{1}{\lambda_{\mathrm{int}}(E)} + \frac{1}{\lambda_{\mathrm{dec}}(E,X)}\right) \frac{\mathrm{d}N_\mu(X,E)}{\mathrm{d}E}\,\mathrm{d}E \\ &- \int_{E_\mu^{\mathrm{th}}}^{\infty} \frac{\partial}{\partial E}\left(a(E)\frac{\mathrm{d}N_\mu(X,E)}{\mathrm{d}E}\right)\mathrm{d}E\,,\end{aligned} \tag{2.2}$$

  where $E_\mu^{\mathrm{th}} = 1$ GeV is the muon energy threshold for simulation, λ_{int} and λ_{dec} denote the interaction and decay lengths of muons in the atmosphere, respectively, and $a(E)$ describes energy loss processes, such as ionisation, as muons propagation in the atmosphere.

- N_μ: denotes the number of muons at the ground level. It is directly read from the CONEX output as the value of the muon longitudinal profile at a depth $X = X_{\mathrm{gr}} \sec\theta$. This is made explicit in the right panel of Figure 2.1.

- N_μ^{prod}: denotes the total number of muons produced down to the ground level:

$$N_\mu^{\mathrm{prod}} = \int_0^{X_{\mathrm{gr}}\sec\theta} \frac{\mathrm{d}N_\mu^{\mathrm{prod}}}{\mathrm{d}X}\,\mathrm{d}X\,. \tag{2.3}$$

Additionally, when more detailed information about the simulated cascades is needed, for example, access to the energy spectrum of muons at the ground level, we produced libraries of 1 000 proton-induced CORSIKA [110] simulations, using version 7.7402. Low-energy hadronic interactions were simulated with FLUKA [111, 112]. High-energy hadronic interactions were simulated with EPOS LHC, QGSJET-II.04 and SIBYLL2.3d. The transition between high- and low-energy hadronic interaction models was set to 80 GeV. The fraction of the primary energy below which the thinning algorithm becomes active was set to 10^{-6}, to mitigate the additional fluctuations of shower observables due to the thinning procedure. The kinetic energy thresholds for different shower particles were set to 50 MeV for muons and hadrons and to 250 keV for electrons/positrons and photons. The number of muons, N_μ, was obtained by integrating the muon flux on the ground level at a height of 1400 m a.s.l, above a certain energy threshold.

The data resulting from EAS simulations were analysed using standard tools such as `ROOT 6.32.06` [113] and several `Python 3.10.12` libraries, including: `Pandas 2.0.3` [114, 115] and `Scipy 1.11.1` [116]. All plots were produced using `Matplotlib 3.7.2` [117]. The particular statistical methods and algorithms used in different parts of the analyses are described in detail as deemed necessary and always preceding the results obtained in each chapter.

Analytical and semi-analytical derivations needed to explain particular features of the results, and to establish the connection between primary interaction properties and shower observables, are also carefully presented in the required chapters.

2.2 The joint distribution of N_μ and $X_{\max}$

As mentioned in Section 1.2.5, the shower-by-shower distribution of the number of muons, $f(N_\mu)$ is mostly shaped by the distribution of the primary interaction variable α_1 [53]. Note that, unless there is ambiguity, the probability density function of a random variable x, $f_X(x)$, will always be written as $f(x) \equiv f_X(x)$.

In turn, the distribution of $X_{\max}$, $f(X_{\max})$, is shaped by the fluctuations of the primary interaction point, X_1, and by the stochasticity of hadron production and propagation in the early stages of the shower development. These last two contributions to the fluctuations of $X_{\max}$ are captured by the variable

$$\Delta X_{\max} \equiv X_{\max} - X_1. \tag{2.4}$$

Moreover, the fluctuations of X_1 are independent of those of $\Delta X_{\max}$: $f(\Delta X_{\max} \,|\, X_1) \simeq f(\Delta X_{\max})$. The extraction of information about multiparticle production in the primary interaction from $f(X_{\max})$, via the probability density function of $\Delta X_{\max}$, $f(\Delta X_{\max})$, is the subject of Chapter 4.

Importantly, the fluctuations of N_μ and $X_{\max}$ are not independent, especially in proton-induced extensive air showers [107, 118]. Instead, the joint distribution of these shower observables, $f(N_\mu, X_{\max})$, which is built over an ensemble of proton-induced showers, has a rich structure, as shown in Figure 2.2. This figure was produced using a library of 10^5 CONEX simulations of proton-induced simulations as described in Section 2.1, with primary energy $E_0 = 10^{19}\,\text{eV}$ and $\theta = 60°$, using the high-energy hadronic interaction model QGSJET-III.01. The marginal probability density functions of N_μ, $f(N_\mu)$, and $X_{\max}$, $f(X_{\max})$, are shown in the same figure.

The marginal distribution $f(N_\mu)$ has a power-law tail towards low values composed of muon-depleted showers. As discussed in [106] and Section 1.5 this tail arises from primary interactions where neutral pions carry a large fraction of the primary energy, leaving less energy in the hadronic core to produce muons. The right tail of $f(N_\mu)$ is due to muon attenuation, as discussed later.

The distribution of $X_{\max}$ has a pronounced exponential tail towards deep showers. This tail is mostly due to the exponentially distributed depth of the primary interaction point, X_1 [56, 104]. That is, X_1 has a probability density function given by Equation (1.45). The other features of the distribution $f(X_{\max})$ stem for the shape of $f(\Delta X_{\max})$ which will be described in light of the stochasticty of particle production in the primary interaction in Chapter 4.

From the shape of the joint distribution $f(N_\mu, X_{\max})$ it is clear that N_μ and $X_{\max}$ are mildly correlated, as indicated by the Pearson correlation coefficient $\rho(N_\mu, X_{\max}) = \text{cov}\,[N_\mu, X_{\max}]\,/\sigma(N_\mu)\sigma(X_{\max}) = 0.11$, where $\text{cov}\,[N_\mu, X_{\max}] = \langle N_\mu X_{\max}\rangle - \langle N_\mu\rangle\,\langle X_{\max}\rangle$ denotes the covariance and $\sigma(\cdot)$ the standard deviation.

Furthermore, there is an exponential smear along the positive diagonal of N_μ-$X_{\max}$ plane and a hard boundary defined by the minimum $X_{\max}$ for each value of N_μ. These fea-

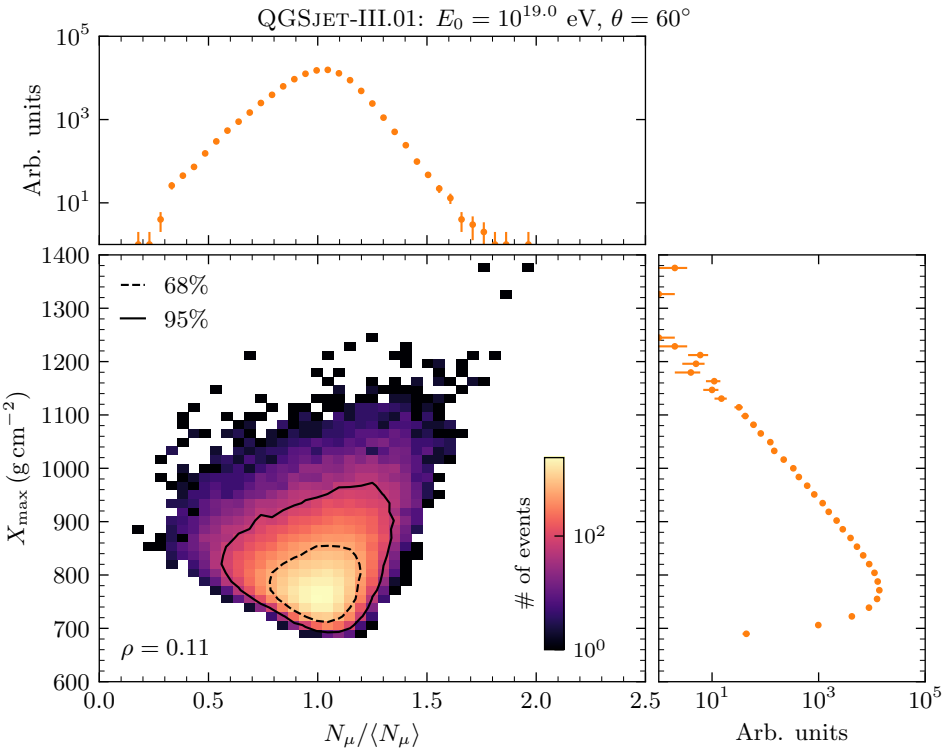

Figure 2.2: Central panel: Joint distribution of $N_\mu/\langle N_\mu \rangle$ and $X_{\max}$. The dashed and solid curves are the contours enclosing 68% and 95% of the events, respectively. Top panel: Marginal distribution of $N_\mu/\langle N_\mu \rangle$. Right panel: Marginal distribution of $X_{\max}$. This figure was produced with a library of 10^5 CONEX simulations of proton-induced showers with $E_0 = 10^{19.0}$ eV and $\theta = 60°$ using the high-energy hadronic interaction model QGSJET-III.01.

tures yield the triangular contours of $f(N_\mu, X_{\max})$. The values of $\log_{10}\langle N_\mu \rangle$, $\sigma(N_\mu)/\langle N_\mu \rangle$, $\langle X_{\max} \rangle$, $\sigma(X_{\max})$ and $\rho(N_\mu, X_{\max})$ can be read of Tables A.1, A.3 and A.5 for the high energy hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e, respectively.

The intricate physics of hadron production that correlates N_μ and $X_{\max}$ will be one of the main subjects of this work. As we will describe thoroughly in Sections 2.2.1 and 2.2.2, other effects shape $f(N_\mu, X_{\max})$ such as the fluctuations of X_1, the $X_{\max}$-dependent attenuation of muons before reaching the ground and the fluctuations induced by truncation of the muon profile by the ground. In shallower showers (those with lower values of $X_{\max}$), muons are produced closer to the top of the atmosphere. Therefore, assuming that the shape of the energy spectrum of muons at production varies little within the range of possible $X_{\max}$ values [109], a larger fraction of muons are attenuated in shallow showers relative to deeper ones. This $X_{\max}$-dependent attenuation will be analytically estimated in Section 2.2.1 and induces a positive correlation between N_μ and $X_{\max}$. Additionally, for zenith angles allowing full development of the muon component before reaching the ground, the additional fluctuations of N_μ due to the truncation of the muon profile by the ground correspond to the muon attenuation effect. However, in more vertical showers these additional fluctuations can further correlate or anti-correlate N_μ and $X_{\max}$, depending on whether the peak of the muon profile tends to be above or below the ground level. These effects affect the natural correlation between N_μ and $X_{\max}$ that arises from the

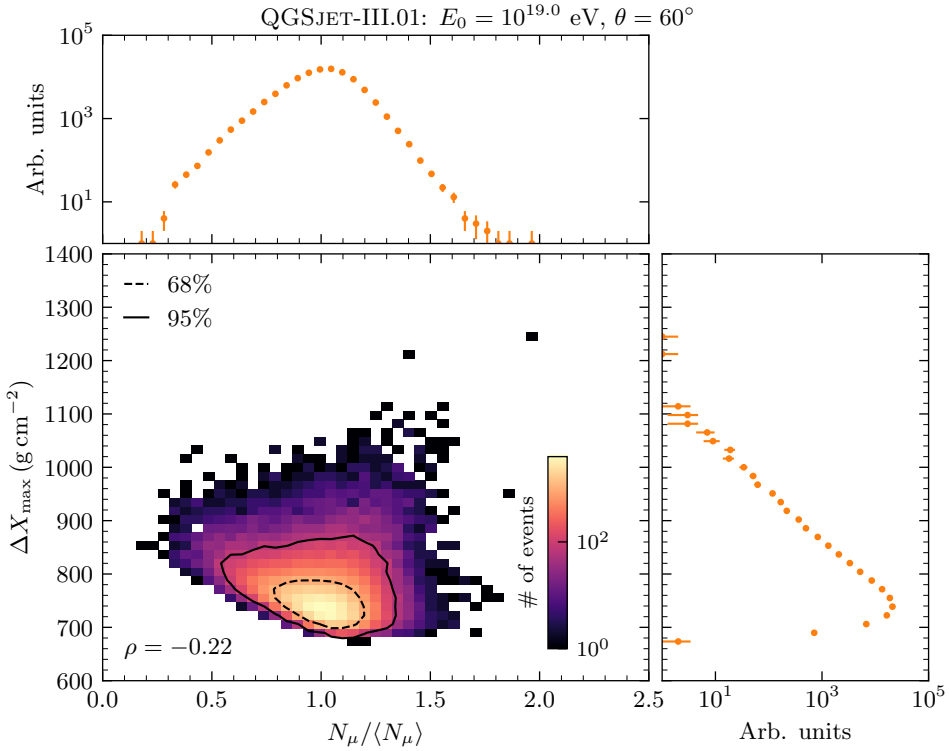

Figure 2.3: Central panel: Joint distribution of $N_\mu/\langle N_\mu \rangle$ and $\Delta X_{\max}$. The dashed and solid curves are the contours enclosing 68% and 95% of the events, respectively. Top panel: Marginal distribution of $N_\mu/\langle N_\mu \rangle$. Right panel: Marginal distribution of $\Delta X_{\max}$. This figure was produced with a library of 10^5 CONEX simulations of proton-induced showers with $E_0 = 10^{19.0}$ eV and $\theta = 60°$ using the high-energy hadronic interaction model QGSJET-III.01.

stochasticity of hadron production throughout the cascade, and are highly dependent on the shower geometry, as discussed in Section 2.2.2.

The fluctuations of X_1 can be removed by considering the distribution $f(N_\mu, \Delta X_{\max})$, as shown in Figure 2.3. The shape of this joint distribution is fully determined by the primary interaction and fluctuations in the production and propagation of hadrons in later shower generations. The observables N_μ and $\Delta X_{\max}$ are mildly anti-correlated, as indicated by their negative Pearson correlation coefficient $\rho(N_\mu, \Delta X_{\max}) = -0.22$. The fluctuations of X_1 thus introduce an additional smear along the positive diagonal of the N_μ-$X_{\max}$ plane, as they are not independent of the number of muons on the ground. This dependence is mostly driven by the truncation of the muon profile. Note that the distribution of $\Delta X_{\max}$, $f(\Delta X_{\max})$, still has an exponential tail towards deep showers. In Section 4.2.2, we will show that this tail mostly arises from diffractive primary interactions.

The effects of muon attenuation and truncation of the muon profile on $f(N_\mu, \Delta X_{\max})$ can be removed by considering the total number of produced muons, N_μ^{prod}. The joint distribution $f(N_\mu^{\text{prod}}, \Delta X_{\max})$ is shown in Figure 2.4, together with the marginal distributions $f(N_\mu^{\text{prod}})$ and $f(\Delta X_{\max})$.

The variables N_μ^{prod} and $\Delta X_{\max}$ are the most anti-correlated ($\rho = -0.39$). Moreover, the right portion of the marginal distribution $f(N_\mu^{\text{prod}})$ falls much more steeply than that of $f(N_\mu)$ while retaining the long left tail populated by muon-depleted showers. The values of $\log_{10} \langle N_\mu^{\text{prod}} \rangle$, $\sigma(N_\mu^{\text{prod}})/\langle N_\mu^{\text{prod}} \rangle$, $\langle \Delta X_{\max} \rangle$, $\sigma(\Delta X_{\max})$ and $\rho(N_\mu^{\text{prod}}, \Delta X_{\max})$ can

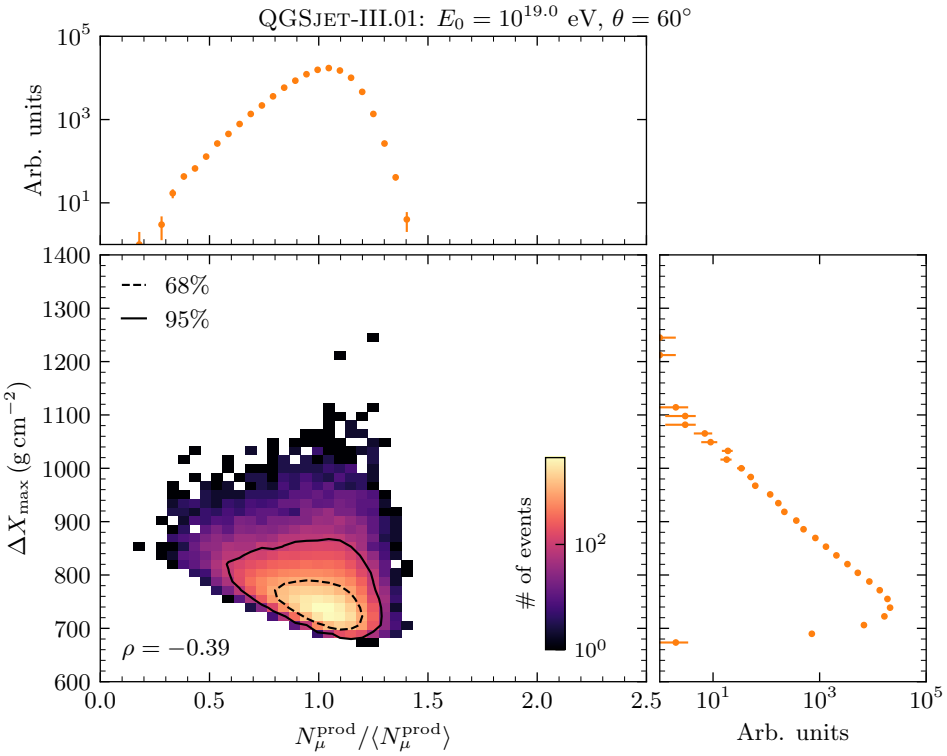

Figure 2.4: Central panel: Joint distribution of $N_\mu^{\text{prod}}/\langle N_\mu^{\text{prod}}\rangle$ and $\Delta X_{\max}$. The dashed and solid curves are the contours enclosing 68% and 95% of the events, respectively. Top panel: Marginal distribution of $N_\mu^{\text{prod}}/\langle N_\mu^{\text{prod}}\rangle$. Right panel: Marginal distribution of $\Delta X_{\max}$. This figure was produced with a library of 10^5 CONEX simulations of proton-induced showers with $E_0 = 10^{19.0}$ eV and $\theta = 60°$ using the high-energy hadronic interaction model QGSJET-III.01.

be read from Tables A.2, A.4 and A.6 for the high-energy hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e, respectively.

The shape of $f(N_\mu^{\text{prod}}, \Delta X_{\max})$ is mostly determined by the physics of hadron production in the early stages of the cascading process. Hence, the physics of hadron production in hadron-air interactions tends to anti-correlate N_μ and $X_{\max}$. The explanation of this anti-correlation is given in Chapter 5.

We now explore how the shape of $f(N_\mu, X_{\max})$ changes with the primary energy (Section 2.2.1), the zenith angle of the shower (Section 2.2.2), the primary mass composition (Section 2.2.3) and the energy threshold for muon detection (Section 2.2.4).

2.2.1 Dependence on the primary energy

The Heitler–Matthews model, introduced in Section 1.2.4, predicts that the mean number of muons in proton-induced showers, $\langle N_\mu \rangle$, scales with the primary energy E_0 as a power law with exponent β (see Equation (1.28)). It also predicts a linear increase of the average depth of shower maximum, $\langle X_{\max} \rangle$, with the logarithm of E_0, characterized by the elongation rate, D_p.

These relationships are illustrated in Figure 2.5 for three high-energy hadronic interaction models. The figure is produced using sets of 10^4 CONEX simulations of proton-induced simulations, generated using the setup described in Section 2.1, for primary energies ran-

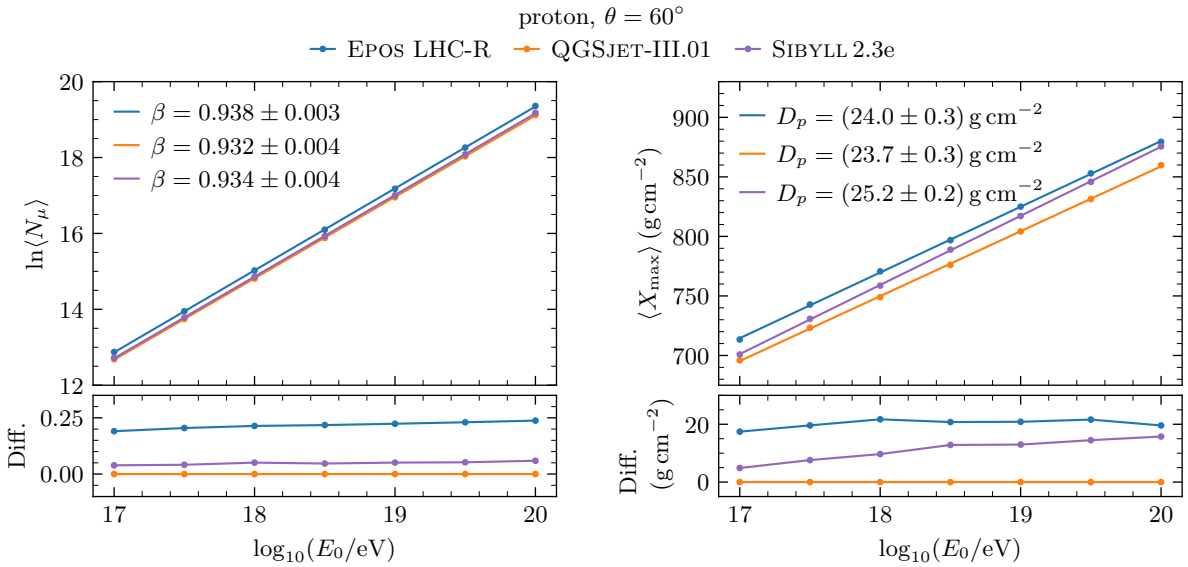

Figure 2.5: Upper left: Logarithm of the mean number of muons, $\ln\langle N_\mu \rangle$, as a function of the primary energy. The solid lines represent linear regression fits, with the corresponding Heitler-Matthews exponents β indicated in the legend. Lower left: Difference between $\ln\langle N_\mu \rangle$ for each model and the reference model QGSJET-III.01. Upper right: Mean depth of shower maximum, $\langle X_{\max} \rangle$, as a function of the primary energy. The solid lines show linear fits, with the corresponding elongation rates D_p given in the legend. Lower right: Difference between $\langle X_{\max} \rangle$ for each model and the reference model QGSJET-III.01. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60°$.

ging from $E_0 = 10^{17.0}$ to $10^{20.0}$ eV, in steps of $\Delta \log_{10}(E_0/\mathrm{eV}) = 0.5$. The values of the power-law index β are obtained from a χ^2 fit to the relation $\ln\langle N_\mu \rangle = \beta \ln E_0 + \ln \mathcal{C}$, while the elongation rate D_p is extracted from a χ^2 fit to $\langle X_{\max} \rangle = D_p \ln E_0 + b$.

The value of $\beta \simeq 0.93$ varies by $\approx 1\%$ on the high-energy hadronic interaction model, while $D_p \simeq 24.3\,\mathrm{g\,cm^{-2}}$ has a systematic model dependence of $\approx 1.5\,\mathrm{g\,cm^{-2}}$, in agreement with [119]. Hence, these parameters show little sensitivity to the different modelling approaches employed in the hadronic interaction models. By contrast, the absolute scale of the number of muons predicted by EPOS LHC-R is $\sim 25\%$ higher than the predictions of QGSJET-III.01, independently of the shower energy, considering $\theta = 60°$. Note that this discrepancy between the models has a zenith angle dependence due to the θ-dependence of the energy spectrum of muons. Similarly, the $X_{\max}$ scale predicted by EPOS LHC-R is $\sim 20\,\mathrm{g\,cm^{-2}}$ deeper than the scale derived with QGSJET-III.01 simulations, with no apparent energy dependence. The difference in $\langle X_{\max} \rangle$ between SIBYLL2.3e and QGSJET-III.01 has a strong energy dependence, starting at just $\approx 5\,\mathrm{g\,cm^{-2}}$ at $E_0 = 10^{17}$ eV and reaching $\approx 15\,\mathrm{g\,cm^{-2}}$ at $E_0 = 10^{20}$ eV.

Figure 2.6 shows $\langle N_\mu^{\mathrm{prod}} \rangle$ and $\langle \Delta X_{\max} \rangle$ as a function of the primary energy for three hadronic interaction models, together with their difference relative to the reference model QGSJET-III.01. The energy dependences of $\ln\langle N_\mu^{\mathrm{prod}} \rangle$ and $\langle \Delta X_{\max} \rangle$ are fitted to linear functions of the logarithm of E_0 to extract the slopes $\widetilde{\beta}$ and $\widetilde{D}_p$, respectively.

The evolution of $\ln\langle N_\mu^{\mathrm{prod}} \rangle$ mirrors that of $\ln\langle N_\mu \rangle$, yielding $\widetilde{\beta} \sim 0.93$ independently of the hadronic interaction model. The scale of N_μ^{prod} predicted by different models still differs by approximately 20%, considering the most extreme cases. This discrepancy

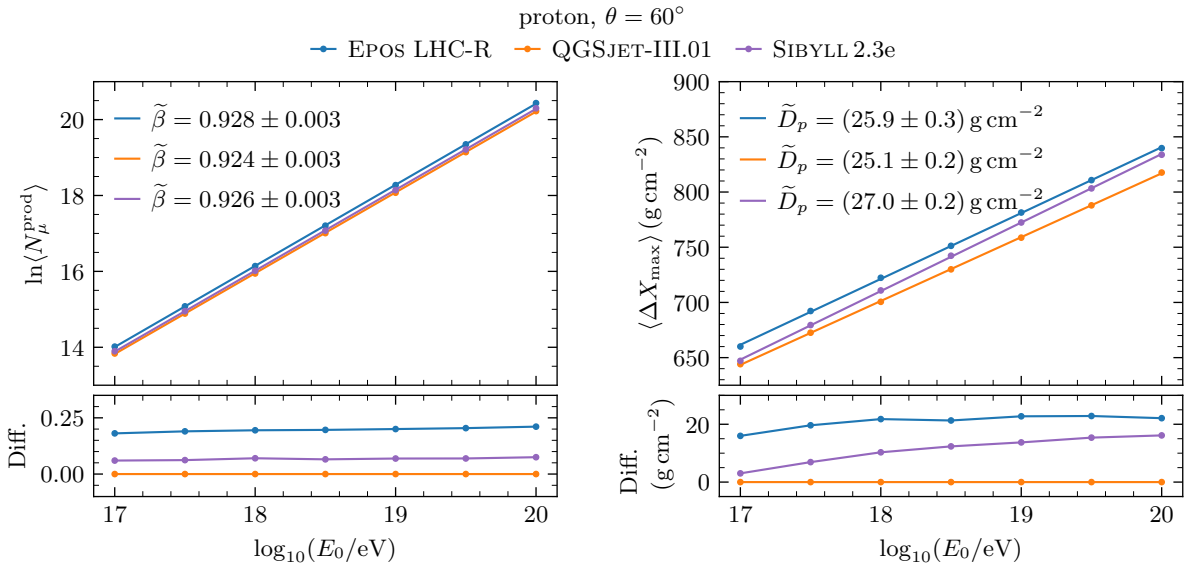

Figure 2.6: Upper left: $\ln\langle N_\mu^{\text{prod}}\rangle$ as a function of the primary energy. The solid lines represent linear regression fits, with the corresponding Heitler-Matthews exponents $\widetilde{\beta}$ indicated in the legend. Lower left: Difference between $\ln\langle N_\mu^{\text{prod}}\rangle$ for each model and the reference model QGSJET-III.01. Upper right: $\langle \Delta X_{\max}\rangle$ as a function of the primary energy. The solid lines show linear fits, with the corresponding elongation rates $\widetilde{D}_p$ given in the legend. Lower right: Difference between $\langle \Delta X_{\max}\rangle$ for each model and the reference model QGSJET-III.01. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60°$.

increases mildly with the primary energy. Similarly, the difference in $\langle \Delta X_{\max}\rangle$ between QGSJET-III.01 and SIBYLL2.3e increases with energy, ranging from about $3\,\text{g cm}^{-2}$ to $16\,\text{g cm}^{-2}$. This indicates that the model dependence of $\langle X_{\max}\rangle$ is primarily driven by particle production in the early stages of the shower, rather than by a strongly model-dependent energy evolution of the proton-air cross-section. Indeed, over 80% of the model dependence in $\langle X_{\max}\rangle$ originates from differences in $\langle \Delta X_{\max}\rangle$, highlighting the importance of describing the $X_{\max}$ distribution in terms of hadronic physics at the highest energies in the cascade. In particular, most of the model dependence in the scale of $X_{\max}$ arises from the primary interaction itself, rather than from highly distinct proton-air cross-sections predicted by different hadronic interaction models [120].

The standard deviations of the relative muon scale, $\sigma(N_\mu)/\langle N_\mu\rangle$, and of $X_{\max}$, $\sigma(X_{\max})$ are shown in Figure 2.7, for three high-energy hadronic interaction models.

Irrespective of the high-energy hadronic interaction model, both $\sigma(N_\mu)/\langle N_\mu\rangle$ and $\sigma(X_{\max})$ decrease with shower energy. The reduction in $\sigma(X_{\max})$ is predominantly driven by the increase of the proton–air cross section with energy, which reduces the interaction length $\lambda_{p\text{-air}}(E_0)$ and thus the variance $\sigma^2(X_1) = \lambda^2_{p\text{-air}}(E_0)$ (this result follows from Equation (1.45)). Since $\sigma^2(X_{\max}) = \sigma^2(X_1) + \sigma^2(\Delta X_{\max}) = \lambda^2_{p\text{-air}}(E_0) + \sigma^2(\Delta X_{\max})$, this leads to a decrease in $\sigma(X_{\max})$ as E_0 increases. Furthermore, as we verify in Section 4.2.3, $\sigma^2(\Delta X_{\max})$ itself also decreases with primary energy.

To understand the decrease in the relative fluctuations of N_μ with the primary energy,

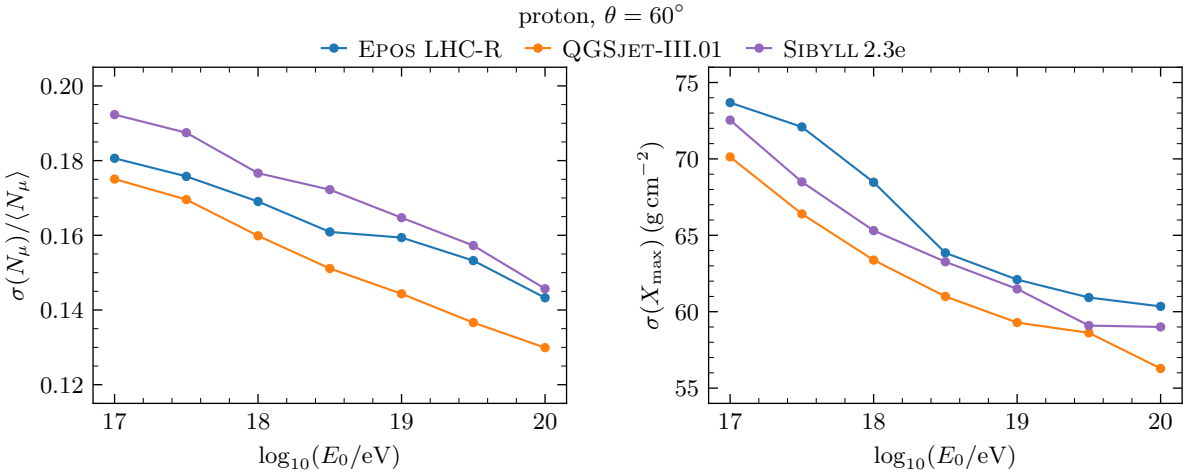

Figure 2.7: Left panel: Relative fluctuations of the muon content at the ground, $\sigma(N_\mu)/\langle N_\mu \rangle$, as a function of the primary energy. Right panel: Fluctuations of $X_{\max}$, $\sigma(X_{\max})$ as a function of the primary energy. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60°$.

we note that, under a first-order approximation and using Equation (1.42), we have

$$\frac{\sigma^2(N_\mu)}{N_\mu^2} = \beta^2 \left[\frac{\sigma(\xi_c^\pi)}{\xi_c^\pi}\right]^2 + \sigma^2(\alpha_1). \tag{2.5}$$

This expression follows the explorations presented in [53, 121, 122]. The variance of α_1 is approximately energy-independent, despite the small violation of Feynman scaling in the energy spectra of secondary particles of the primary interaction as the primary energy increases.

The energy dependence of the variance of the pion critical energy, ξ_c^π, can be derived by assuming that all muons reaching the ground originate from charged-pion decays at a fixed depth $\mathcal{X}_{\max}^\mu$ and that the energy dependence of the pion interaction length can be neglected. Inserting these assumptions into Equation (1.15) yields

$$\xi_c^\pi = \frac{\lambda_\pi\, \varepsilon_\pi\, \sec\theta}{\mathcal{X}_{\max}^\mu} \quad \Longrightarrow \quad \left[\frac{\sigma(\xi_c^\pi)}{\xi_c^\pi}\right]^2 = \left[\frac{\sigma(\mathcal{X}_{\max}^\mu)}{\mathcal{X}_{\max}^\mu}\right]^2. \tag{2.6}$$

Hence,

$$\frac{\sigma^2(N_\mu)}{N_\mu^2} = \beta^2 \left[\frac{\sigma(\mathcal{X}_{\max}^\mu)}{\mathcal{X}_{\max}^\mu}\right]^2 + \sigma^2(\alpha_1)\,. \tag{2.7}$$

The fluctuations in $\mathcal{X}_{\max}^\mu$ arise from two main sources: the variability of the depth of the first interaction and the stochasticity of the interaction lengths of the charged pions in subsequent interactions, encoded in $\sigma^2(\mathcal{X}_{\max}^\mu - X_1)$. The first contribution decreases with primary energy, as argued before. Regarding the second, as the primary energy increases, so does the number of secondary hadrons, which expands the space of all possible paths travelled by the charged pions, after the primary interaction. This larger ensemble of paths leads to an averaging effect that reduces the fluctuations in the total depth accumulated across these paths. Consequently, the relative fluctuations of $\mathcal{X}_{\max}^\mu - X_1$ also decrease

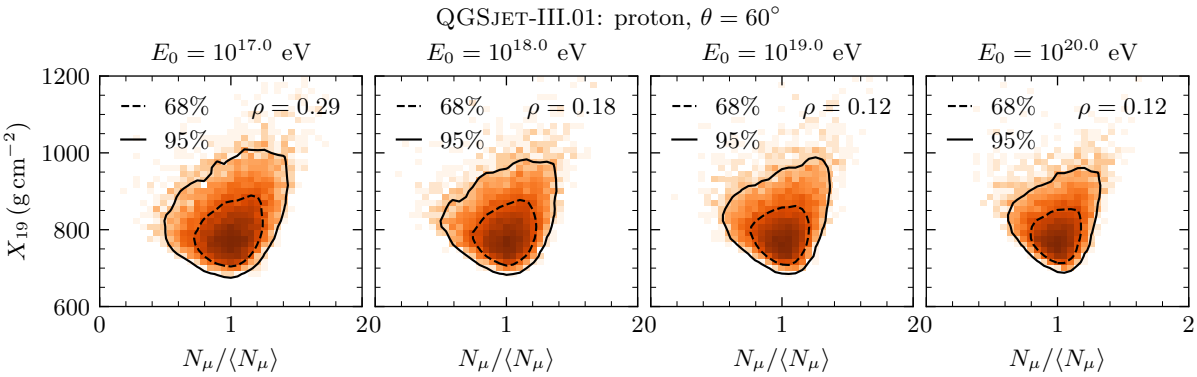

Figure 2.8: Joint distributions $f(N_\mu, X_{19} \,|\, E_0)$ as a function of the primary energy (increasing from left to right in steps of $\Delta \log_{10}(E_0/\text{eV}) = 1$). Each panel contains the Pearson correlation coefficient. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60°$, using QGSJET-III.01.

with energy, resulting in an overall decrease in the relative fluctuations of N_μ, as shown in the left panel of Figure 2.7. Using CONEX simulations, we find a Pearson correlation coefficient of approximately 0.8 between N_μ and $(\mathcal{X}^\mu_{\max})^\beta \alpha$, larger than the correlation between N_μ and α alone. This supports the simple model outlined above for describing N_μ fluctuations. Moreover, Equation (2.7) reproduces the energy dependence of $\sigma(N_\mu)/\langle N_\mu \rangle$ fairly well, up to a nearly constant offset of about 0.015. Finally, despite the crudeness of the calculation, our simple model predicts the ordering of $\sigma(N_\mu)/\langle N_\mu \rangle$ among different hadronic interaction models across all primary energies considered.

To characterize the energy dependence of the joint distribution of N_μ and $X_{\max}$, $f(N_\mu, X_{\max} \,|\, E_0)$, we define

$$X_{19} = X_{\max}(E_0) + D_p \ln 10 \times [19 - \log_{10}(E_0/1\,\text{eV})]\,, \tag{2.8}$$

since its average value is independent of the shower energy. This choice ensures that $\rho(N_\mu, X_{\max}) = \rho(N_\mu, X_{19}) = \rho(N_\mu/\langle N_\mu \rangle, X_{19})$ and facilitates the visual comparison of $f(N_\mu, X_{\max} \,|\, E_0)$ across the range of primary energies.

From Figure 2.8 it is apparent that $N_\mu/\langle N_\mu \rangle$ and X_{19} become less positively correlated with increasing primary energy. The same trend is observed when considering the relative number of produced muons, $N^{\text{prod}}_\mu / \langle N^{\text{prod}}_\mu \rangle$ and $\Delta X_{\max}$, as depicted in Figure 2.9.

As a measure of the changing shape of $f(N_\mu, X_{\max} \,|\, E_0)$ with E_0, we consider the energy dependence of $\rho(N_\mu, X_{\max})$. This dependence can be fully explained in light of the energy dependence of $\rho(N^{\text{prod}}_\mu, X_{\max})$, using the simple muon attenuation model developed in Section 2.2.1. In turn, the energy dependence of $\rho(N^{\text{prod}}_\mu, X_{\max})$ can be understood in terms of the energy dependence of $\rho(N^{\text{prod}}_\mu, \Delta X_{\max})$ and $\rho(N^{\text{prod}}_\mu, X_1)$. In fact, considering that

$$\sigma^2(X_{\max}) = \sigma^2(\Delta X_{\max}) + \sigma^2(X_1) + 2\text{Cov}[\Delta X_{\max}, X_1] \simeq \sigma^2(\Delta X_{\max}) + \sigma^2(X_1), \tag{2.9}$$

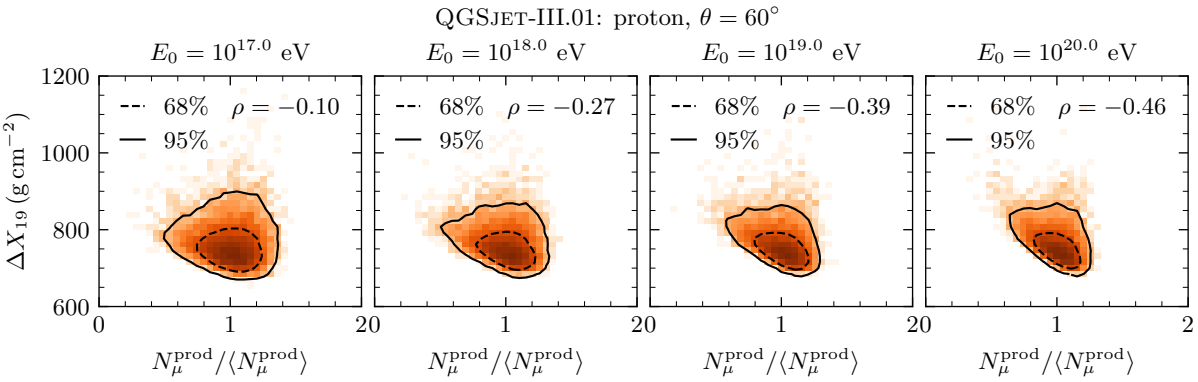

Figure 2.9: Joint distributions $f(N_\mu^{\rm prod}, \Delta X_{19} \,|\, E_0)$ as a function of the primary energy (increasing from left to right in steps of $\Delta \log_{10}(E_0/\text{eV}) = 1$). Each panel contains the Pearson correlation coefficient. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60°$, using QGSJET-III.01.

where we neglected the covariance term[2], the following relationship holds [3]

$$\rho(N_\mu^{\rm prod}, X_{\max}) = \left(1 + r_\sigma^2\right)^{-\frac{1}{2}} \left[\rho(N_\mu^{\rm prod}, \Delta X_{\max}) + \frac{\rho(N_\mu^{\rm prod}, X_1)}{r_\sigma}\right], \tag{2.10}$$

with $r_\sigma \equiv \sigma(X_1)/\sigma(\Delta X_{\max}) = \lambda_{p-\text{air}}(E_0)/\sigma(\Delta X_{\max})$.

The left panel of Figure 2.10 shows $\rho(N_\mu^{\rm prod}, X_1)$ as a function of the primary energy E_0, while the right panel shows $\rho(N_\mu^{\rm prod}, \Delta X_{\max})$, $\rho(N_\mu^{\rm prod}, X_{\max})$ and the predicted $\rho(N_\mu^{\rm prod}, X_{\max})$ using Equation (2.10).

There is a very mild positive correlation between the number of produced muons and X_1. This correlation is not due to the truncation of the MPD distribution by the ground, which truncates the integral that defines $N_\mu^{\rm prod}$ (see Equation (2.3)). Deeper showers would be truncated sooner, leading to a smaller number of produced muons down to the ground level and to an anti-correlation between X_1 and $N_\mu^{\rm prod}$. Despite the origin of this positive correlation being out of the scope of this work, we speculate that since cascades with deeper X_1 are prompt to develop in denser atmospheric layers, secondaries tend to further interact over more shower generations. Therefore, the critical energy for muon production decreases, leading to an enhanced muon content.

Nevertheless, the positive values of $\rho(N_\mu^{\rm prod}, X_1)$, which have a mild dependence on the hadronic interaction model, justify the more positive values of $\rho(N_\mu^{\rm prod}, X_{\max})$ relative to $\rho(N_\mu^{\rm prod}, \Delta X_{\max})$. In particular, Equation (2.10) predicts reasonably well the energy dependence of $\rho(N_\mu^{\rm prod}, X_{\max})$ in terms of the evolution of $\rho(N_\mu^{\rm prod}, \Delta X_{\max})$. The latter correlation coefficient describes the joint distribution $f(N_\mu^{\rm prod}, \Delta X_{\max})$, which is the most connected to the physics of hadron production driving the shower.

[2]Since $|\text{Cov}[\Delta X_{\max}, X_1]| = |\rho(\Delta X_{\max}, X_1)\sigma(X_1)\sigma(\Delta X_{\max})| \leq 0.05\sigma(X_1)\sigma(\Delta X_{\max})$, for all hadronic interaction models and primary energies in the range $E_0 \in [10^{17.0}, 10^{19.0}]$ eV, and therefore negligible compared to the variances of X_1 and $\Delta X_{\max}$.

[3]This is easy to see if we consider the linearity of the covariance operator: $\text{Cov}[N_\mu^{\rm prod}, X_{\max}] = \text{Cov}[N_\mu^{\rm prod}, \Delta X_{\max}] + \text{Cov}[N_\mu^{\rm prod}, X_1]$ and assume $\sigma(X_{\max}) = \sigma(\Delta X_{\max})\sqrt{1 + \sigma^2(X_1)/\sigma^2(\Delta X_{\max})} = \sigma(\Delta X_{\max})\sqrt{1 + r_\sigma^2}$, together with the definition of the Pearson correlation coefficient.

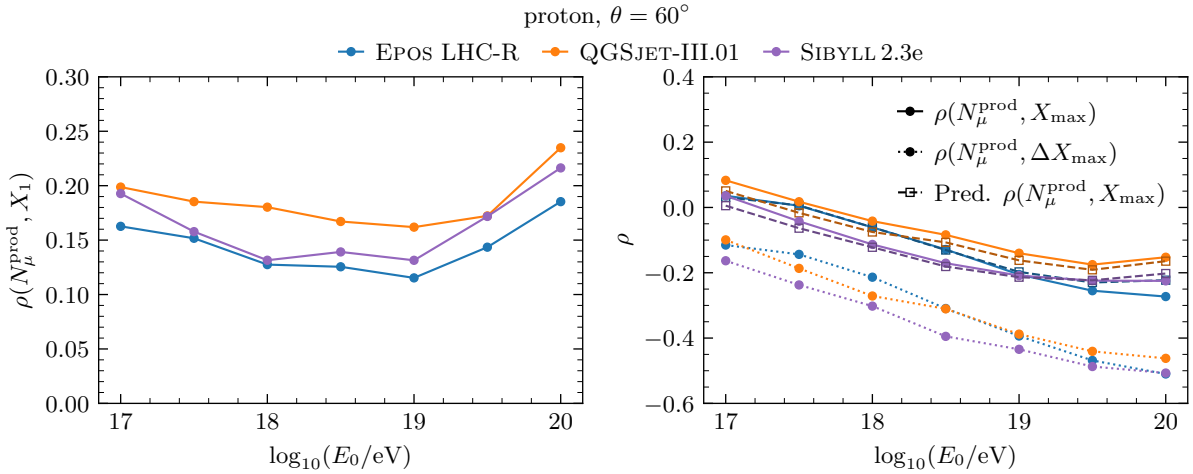

Figure 2.10: Primary-energy dependence of the Pearson correlation coefficients $\rho(N_\mu^{\rm prod}, X_1)$ (left) and $\rho(N_\mu^{\rm prod}, X_{\max})$ (right - solid light lines), $\rho(N_\mu^{\rm prod}, \Delta X_{\max})$ (right - dotted light lines) and predicted $\rho(N_\mu^{\rm prod}, X_{\max})$ from Equation (2.10) (right - dashed dark lines). This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60°$, for three hadronic interaction models.

We now describe the energy evolution of $\rho(N_\mu, X_{\max})$ in terms of the energy evolution of $\rho(N_\mu^{\rm prod}, X_{\max})$, by developing a simple attenuation model.

Simple muon attenuation model

We start by considering a single muon, i, produced with energy E_i and at depth X_i. In general, muons propagating in the atmosphere lose energy through various processes [123]. Here, we only consider ionisation energy losses at minimum ionising energy. That is, we assume that after travelling a depth $\mathrm{d}X$, the energy of the muon, E_μ, decreases as

$$\frac{\mathrm{d}E_\mu}{\mathrm{d}X} = -a \implies E_\mu(X) = -a(X - X_i) + E_i, \tag{2.11}$$

with $a = 2\,\mathrm{MeV\,g^{-1}\,cm^2}$ [43].

Furthermore, we neglect the muon decay probability, which depends exponentially on its kinetic energy, and consider that a muon does not reach depth $X > X_i$ if and only if it stops propagating, i.e., if its kinetic energy vanishes. This corresponds to the condition $E_\mu(X) = m_\mu$, where m_μ is the mass of the muon. This is equivalent to considering a Heaviside decay probability $p(E) = \Theta(E_\mu - m_\mu)$. Lastly, we suppose that $N_\mu^{\rm prod}$ muons are produced at the same depth X_i, with energy E_i following the power-law

$$\frac{\mathrm{d}N_\mu^{\rm prod}}{\mathrm{d}E_i} = CE_i^{-\gamma}\Theta(E_i - m_\mu), \quad \text{with } C = (\gamma - 1)\frac{N_\mu^{\rm prod}}{m_\mu^{1-\gamma}}, \tag{2.12}$$

where $\gamma > 1$ denotes the spectral index. The condition $\gamma > 1$ guarantees the integrability of $\mathrm{d}N_\mu^{\rm prod}/\mathrm{d}E_i$ over the domain $[m_\mu, \infty[$. Moreover, $\gamma = \gamma(E_0, \theta)$ [109, 123].

Now, the number of muons at a depth X, $N_\mu(X)$ is determined from

$$N_\mu(X) = \int_0^\infty \frac{\mathrm{d}N_\mu^{\rm prod}}{\mathrm{d}E_\mu}\,\mathrm{d}E_\mu\,. \tag{2.13}$$

Noting that the energy spectrum of muons at a depth X reads,

$$\frac{\mathrm{d}N_\mu^{\rm prod}}{\mathrm{d}E_\mu} = \frac{\mathrm{d}N_\mu^{\rm prod}}{\mathrm{d}E_i}\frac{\mathrm{d}E_i}{\mathrm{d}E_\mu} = \left.\frac{\mathrm{d}N_\mu^{\rm prod}}{\mathrm{d}E_i}\right|_{E_\mu} = C\,[E_\mu + a(X - X_i)]^{-\gamma}\,\Theta(E_\mu - m_\mu), \tag{2.14}$$

allows us to write

$$N_\mu(X) = C\int_{m_\mu}^\infty [E_\mu + a(X - X_i)]^{-\gamma}\,\mathrm{d}E_\mu\,, \tag{2.15}$$

which evaluates to

$$N_\mu(X) = N_\mu^{\rm prod}\left[1 + \frac{a\,(X - X_i)}{m_\mu}\right]^{1-\gamma}, \tag{2.16}$$

after using the definition of the constant C. Thus, the fraction of surviving muons at depth X, ε_μ, is given by

$$\varepsilon_\mu(X) \equiv \frac{N_\mu(X)}{N_\mu^{\rm prod}} = \left[1 + \frac{(X - X_i)}{\lambda_\mu}\right]^{1-\gamma}, \tag{2.17}$$

where we defined the characteristic attenuation length $\lambda_\mu = m_\mu/a \simeq 50\,\mathrm{g\,cm^{-2}}$: the depth undergone by a muon corresponding to a kinetic energy loss equal to the muon's mass.

We complete our approximations by considering that all muons are produced at $X_i = \mathcal{X}_{\rm max}^\mu$. Therefore, for the slant depth corresponding to the ground level $X = X_{\rm gr}\sec\theta$ we have

$$\varepsilon_\mu \equiv \varepsilon_\mu(X_{\rm gr}\sec\theta) = \left[1 + \frac{(X_{\rm gr}\sec\theta - \mathcal{X}_{\rm max}^\mu)}{\lambda_\mu}\right]^{1-\gamma}. \tag{2.18}$$

Note that $(X_{\rm gr}\sec\theta - \mathcal{X}_{\rm max}^\mu)$ is simply the slant depth between the depth of the maximum of the MPD distribution and the ground.

The value of the spectral index γ was obtained as follows. First, the shower-to-shower fraction of surviving muons ε_μ and the corresponding value of $1 + (X_{\rm gr} - \mathcal{X}_\mu^{\rm max})/\lambda_\mu$ was computed using the usual library of CONEX simulations. Then, γ was extracted from a linear regression to the pairs of points $(\ln\left[1 + (X_{\rm gr} - \mathcal{X}_\mu^{\rm max})/\lambda_\mu\right], \ln\left[N_\mu/N_\mu^{\rm prod}\right])$. The joint distribution of these pairs of points and the corresponding regression curves are depicted in Figure 2.11 for different primary energies, using the hadronic interaction model QGSJET-III.01. The values of γ for all up-to-date hadronic interaction models are shown in Table 2.1.

Despite the rudimentary assumptions taken in the derivation of the simple muon attenuation model, its high precision is evidenced by the low dispersion of the points about the regression line, and the high correlation between $\ln\varepsilon_\mu$ and $\ln\left[1 + (X_{\rm gr} - \mathcal{X}_\mu^{\rm max})/\lambda_\mu\right]$. The absolute value of the Pearson correlation coefficient between these quantities increases mildly with the primary energy from $\rho = -0.95$ to -0.97. Moreover, the precision of the muon attenuation model is independent of the high-energy hadronic interaction model. The value of γ is mildly dependent on the hadronic interaction model, evidencing the

$\log_{10}(E_0/\text{eV})$	EPOS LHC-R	QGSJET-III.01	SIBYLL2.3e
17.0	1.711 ± 0.002	1.627 ± 0.002	1.629 ± 0.002
17.5	1.734 ± 0.002	1.661 ± 0.002	1.658 ± 0.002
18.0	1.747 ± 0.002	1.681 ± 0.002	1.682 ± 0.002
18.5	1.752 ± 0.002	1.701 ± 0.002	1.704 ± 0.002
19.0	1.740 ± 0.001	1.710 ± 0.001	1.717 ± 0.001
19.5	1.734 ± 0.002	1.715 ± 0.002	1.733 ± 0.002
20.0	1.735 ± 0.002	1.735 ± 0.002	1.751 ± 0.002

Table 2.1: Spectral index γ extracted from a linear regression to the pairs of points $(\ln\left[1 + (X_{\text{gr}} \sec\theta - \mathcal{X}_\mu^{\max})/\lambda_\mu\right], \ln\left[N_\mu/N_\mu^{\text{prod}}\right])$ using Equation (2.18), as a function of the primary energy, E_0, for different hadronic interaction models. The values are obtained with libraries of 10^4 CONEX simulations of proton-induced showers with $\theta = 60°$.

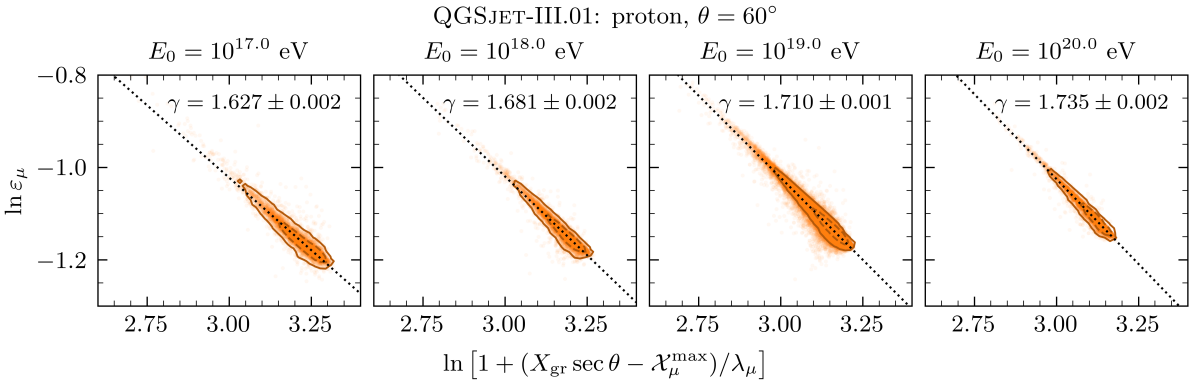

Figure 2.11: Calibration of the spectral index, γ, of the energy spectrum of muons at production given by Equation (2.18) as a function of the primary energy (increasing from left to right in steps of $\Delta \log_{10}(E_0/\text{eV}) = 1$). Each panel contains the shower-to-shower values of the true survival factor, $\ln \varepsilon_\mu$, and attenuation depth $\ln\left[1 + (X_{\text{gr}} \sec\theta - \mathcal{X}_\mu^{\max})/\lambda_\mu\right]$, and the regression curves, depicted as dotted black lines. The contours containing 68% and 95% of the events are shown in dashed and solid dark orange lines, respectively. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60°$, using QGSJET-III.01.

hadronic-model dependence of the high-energy portion of the energy spectrum of muons at production [109]. Furthermore, the increase in γ with the primary energy agrees with the hardening of the energy spectrum of muons at production. This is expected since lower energy showers develop faster, and therefore higher in the atmosphere, where the air density is smaller. Consequently, the critical energy for charged pions and kaons is higher, producing higher-energy muons upon decay.

We can now compute the predicted muon survival factor using Equation (2.18) and the optimal spectral index for each primary energy and each hadronic interaction model (see Table 2.1). Knowing N_μ^{prod} and $\mathcal{X}_{\max}^{\mu}$ for each shower, the corresponding number of muons at the ground can be estimated as $\varepsilon_\mu N_\mu^{\text{prod}}$, along with the Pearson correlation coefficient $\rho(\varepsilon_\mu N_\mu^{\text{prod}}, X_{\max})$. This is shown in Figure 2.12, as a function of the primary energy for different hadronic interaction models. The figure also shows the Pearson correlation coefficient between $(N_\mu, X_{\max})$ and $(N_\mu^{\text{prod}}, \Delta X_{\max})$.

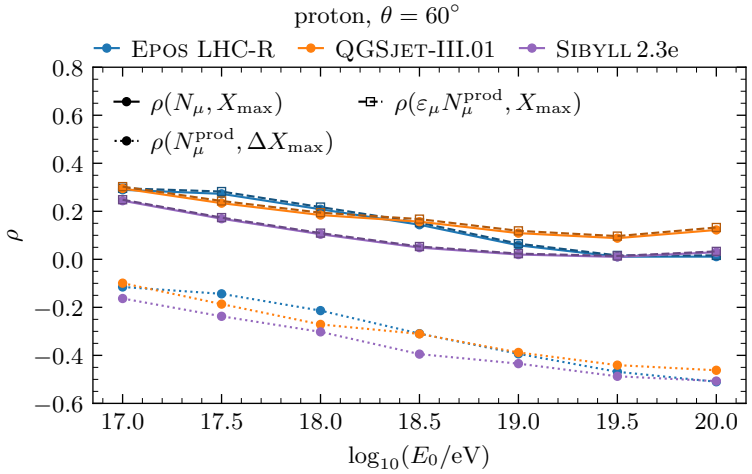

Figure 2.12: Primary-energy dependence of the Pearson correlation coefficients $\rho(N_\mu, X_{\max})$ (solid light lines), $\rho(N_\mu^{\text{prod}}, \Delta X_{\max})$ (dotted light lines) and $\rho(\varepsilon_\mu N_\mu^{\text{prod}}, X_{\max})$ (dashed dark lines). The shower-by-shower values of ε_μ are given by Equation (2.18), using the optimal values of γ for each hadronic interaction model. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60°$, for three different hadronic interaction models.

The presented muon attenuation model correctly predicts the correlation between N_μ and $X_{\max}$, from the correlation between N_μ^{prod} and $X_{\max}$. This is still the case if γ is averaged over the hadronic interaction models for each energy bin. Therefore, only the joint distribution of N_μ^{prod} and $X_{\max}$ needs to be described in terms of the primary proton-air interaction and X_1, since other features of the joint distribution of N_μ and $X_{\max}$ are accounted for by the attenuation of muons, for a fixed shower geometry.

We have also provided a possible tool to constrain the hardness of the energy spectrum of muons at production from the shape of the joint distribution of $f(N_\mu, X_{\max})$. The sensitivity of $f(N_\mu, X_{\max})$ to small variations in γ would need to be studied. This is out of the scope of this work. Furthermore, the model presented in Equation (2.18) shows that $f(N_\mu, X_{\max})$ is sensitive to $\mathcal{X}_{\max}^\mu$. This opens the possibility of constraining this important shower-observable [108] from a measurement of $f(N_\mu, X_{\max})$.

2.2.2 Dependence on the shower zenith angle

The shape of the joint distribution of N_μ and $X_{\max}$ is also highly dependent on the shower's zenith angle. This is clear from Figure 2.13, where $f(N_\mu, X_{\max} \,|\, \theta)$ was obtained from CONEX simulations of proton-induced simulations with fixed primary energy $E_0 = 10^{19}$ eV and $\theta \in \{15°, 30°, 45°, 60°\}$.

In particular, the Pearson correlation coefficient between N_μ and $X_{\max}$ becomes significantly more positive with increasing zenith angle. As a result, these shower observables are highly anti-correlated for nearly vertical showers.

The quantities $\mathcal{N}_\mu^{\max}$ (or N_μ^{prod}) and $\Delta X_{\max}$ are the most connected with hadron production in the early stages of the shower, which is largely independent of the zenith angle. Hence, it is not surprising that $f(\mathcal{N}_\mu^{\max}, \Delta X_{\max})$ has no evident zenith dependence, as shown in Figure 2.14. This zenith angle independence is verified for all hadronic

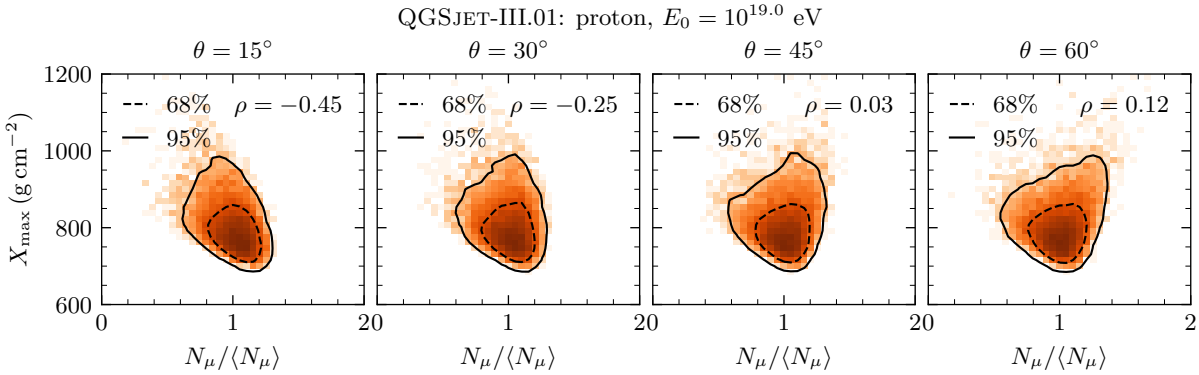

Figure 2.13: Joint distributions $f(N_\mu, X_{\max} \,|\, \theta)$ as a function of the shower's zenith angle, θ (increasing from left to right in steps of $\Delta\theta = 15°$). Each panel contains the Pearson correlation coefficient. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and different values of θ, using QGSJET-III.01.

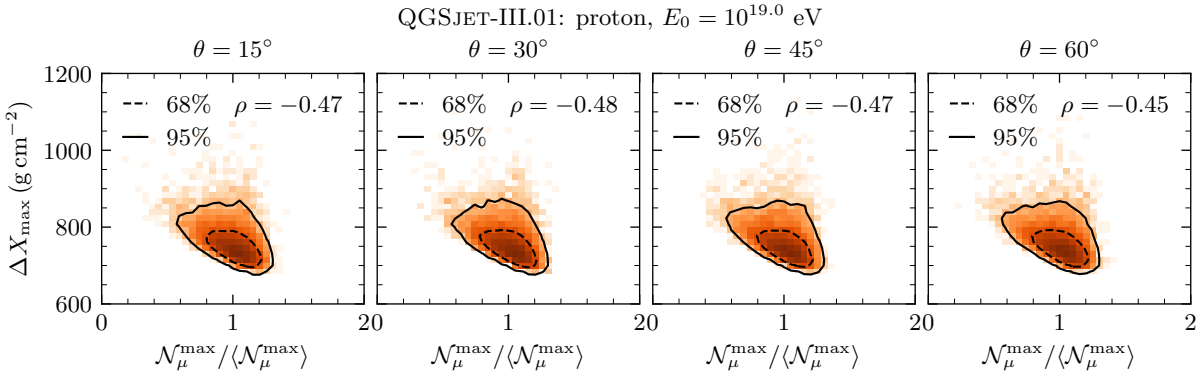

Figure 2.14: Joint distributions $f(\mathcal{N}_\mu^{\max}, \Delta X_{\max} \,|\, \theta)$ as a function of the shower's zenith angle, θ (increasing from left to right in steps of $\Delta\theta = 15°$). Each panel contains the Pearson correlation coefficient. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and different values of θ, using QGSJET-III.01.

interaction models. Note that we worked with $\mathcal{N}_\mu^{\max}$ instead of N_μ^{prod} because the ground truncates the deep tail of the MPD distribution in nearly vertical showers [109]. This would produce a zenith-dependent bias in the value of N_μ^{prod}, since more vertical showers are truncated closer to the peak of the MPD distribution.

The evolution of $f(N_\mu, X_{\max} \,|\, \theta)$ with zenith angle θ can be attributed to several factors, including fluctuations caused by the truncation of the muon profile at ground level, the θ-dependence of muon attenuation, and/or the increase in the critical energy of pions and kaons with θ. The latter effect is intuitive—more inclined showers develop in a thinner atmosphere, where the probability of meson decay exceeds that of interaction at higher energies compared to vertical showers, which develop in denser atmospheric layers. Therefore, fewer muons are produced in more inclined showers. If the variation of the critical energy with θ significantly influences the shape of $f(N_\mu, X_{\max} \,|\, \theta)$, a similar

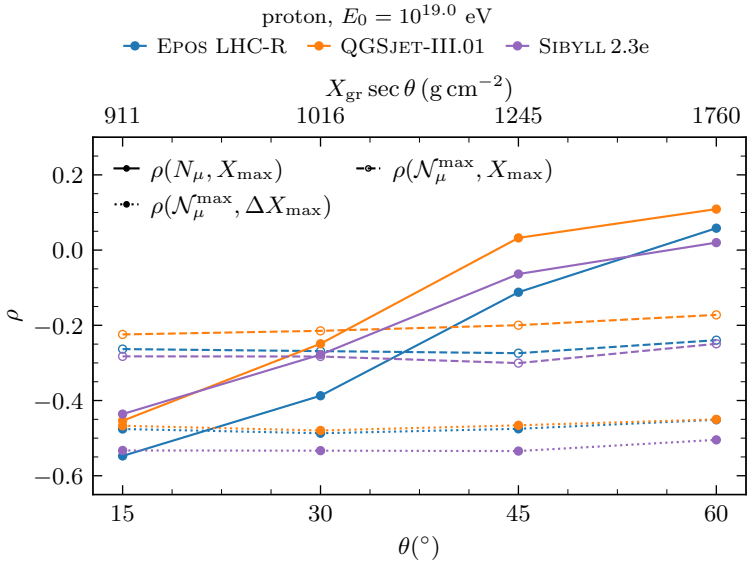

Figure 2.15: Zenith angle dependence of the Pearson correlation coefficients $\rho(N_\mu, X_{\max})$ (solid lines), $\rho(\mathcal{N}_\mu^{\max}, \Delta X_{\max})$ (dotted lines) and $\rho(\mathcal{N}_\mu^{\max}, X_{\max})$ (dashed lines). The top axis shows the slant depth of the ground corresponding to each zenith angle θ. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and different values of θ.

effect should be observed in $f(\mathcal{N}_\mu^{\max}, X_{\max}\,|\,\theta)$, since fluctuations in X_1 would also induce θ-dependent fluctuations of the critical energy, thereby affecting the muon production dynamics.

Figure 2.15 shows the Pearson correlation coefficients $\rho(N_\mu, X_{\max})$, $\rho(\mathcal{N}_\mu^{\max}, X_{\max})$ and $\rho(\mathcal{N}_\mu^{\max}, \Delta X_{\max})$ as a function of θ for libraries of 10^4 CONEX simulations of proton-induced simulations with $E_0 = 10^{19}$ eV produced with three different hadronic interaction models.

The value of $\rho(\mathcal{N}_\mu^{\max}, X_{\max})$ is independent of θ, showing that the effect of the change of critical energy with θ is negligible in shaping the joint distribution. Similar to what is done in Section 2.2.1, the constant decrease in $\rho(\mathcal{N}_\mu^{\max}, X_{\max})$ with respect to $\rho(\mathcal{N}_\mu^{\max}, \Delta X_{\max})$, which is independent of θ, is easily explained from the following relation

$$\rho(\mathcal{N}_\mu^{\max}, X_{\max}) = \left(1 + r_\sigma^2\right)^{-\frac{1}{2}} \left[\rho(\mathcal{N}_\mu^{\max}, \Delta X_{\max}) + \frac{\rho(\mathcal{N}_\mu^{\max}, X_1)}{r_\sigma}\right], \tag{2.19}$$

with $r_\sigma = \sigma(X_1)/\sigma(\Delta X_{\max}) = \lambda_{p-\mathrm{air}}(E_0)/\sigma(\Delta X_{\max})$. Once again, we assume that $\rho(X_1, \Delta X_{\max}) \simeq 0$, independently of θ. Moreover, r_σ is independent of θ. The correlation between $\mathcal{N}_\mu^{\max}$ and X_1 is mildly positive, $\rho(\mathcal{N}_\mu^{\max}, X_1) \simeq 0.13$, and weakly dependent on θ and the high-energy hadronic interaction model. This is in line with the results obtained in Section 2.2.1. Since both terms in the right-hand-side of Equation (2.19) are independent of θ, so must be the left-hand-side $\rho(\mathcal{N}_\mu^{\max}, X_{\max})$. Therefore, the θ dependence of $f(N_\mu, X_{\max}\,|\,\theta)$ arises in the passage from $\mathcal{N}_\mu^{\max}$ to N_μ.

Now, in proton-induced showers with $E_0 = 10^{19}$ eV, the median depth of the peak of the muon profile, $X_{\max}^\mu$, varies between $960\,\mathrm{g\,cm^{-2}}$ and $1070\,\mathrm{g\,cm^{-2}}$ according to the hadronic interaction model. These slant depths correspond to zenith angles $\theta \simeq 25^\circ$ and

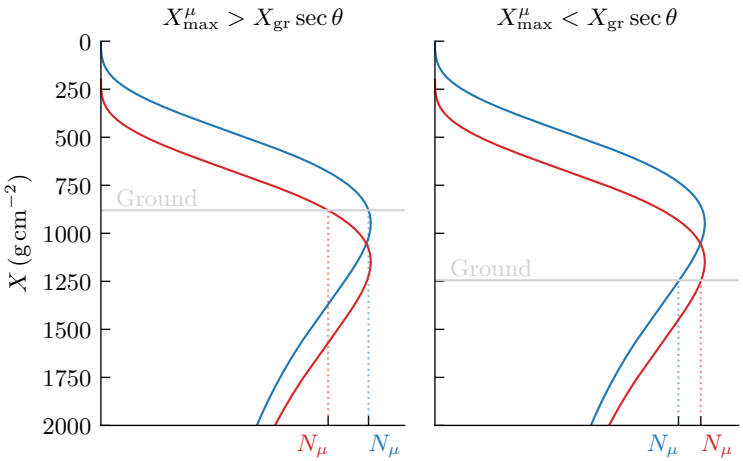

Figure 2.16: Illustration of the truncation of two muon profiles, peaking at different depths, for more vertical (left panel) and more inclined (right panel) showers. The deeper shower is depicted in red and the shallower in blue. The number of muons at the maximum of each profile, $N_\mu^{\max}$, is the same for both showers. The horizontal grey line denotes the ground level, which increases with $\sec\theta$.

$\theta = 35°$. Therefore, in showers with these zenith angles, it is equally likely that the peak of the muon profile lies below or above the ground. For more vertical showers, it becomes more likely that the peak of the muon profile lies below the ground. The opposite becomes true for more horizontal showers.

Therefore, consider an ensemble of showers for which $X^\mu_{\max}$ is above the ground: $X^\mu_{\max} < X_{\rm gr} \sec\theta$. Since the number of muons decreases monotonically between $X^\mu_{\max}$ and the ground level, for this set of showers, the deeper the shower, the larger the number of muons, for fixed $N_\mu^{\max}$. Hence, the truncation of the muon profile by the ground induces fluctuations that tend to positively correlate N_μ and $X_{\max}$. Now consider the opposite scenario: an ensemble of showers with $X^\mu_{\max} > X_{\rm gr} \sec\theta$. For this subset of showers, since the number of muons increases between $X_{\rm gr} \sec\theta$ and $X^\mu_{\max}$, deeper showers yield smaller values of N_μ, as the muon profile is cut earlier. Hence, for this population of showers, the truncation of the muon profile by the ground level negatively correlates N_μ and $X_{\max}$. These two scenarios are illustrated in Figure 2.16.

This effect is also clear if we divide the library of CONEX simulations into showers with $X^\mu_{\max} < X_{\rm gr} \sec\theta$ and those with $X^\mu_{\max} > X_{\rm gr} \sec\theta$, as shown in Figure 2.17.

All showers with $\theta = 15°$ have $X^\mu_{\max}$ located below the ground level, which enhances the anti-correlation between N_μ and $X_{\max}$. At $\theta = 30°$, the sample includes showers with $X^\mu_{\max}$ both below and above the ground. Consequently, for showers with $\theta \leq 30°$, the muon cascade is not yet fully developed before reaching the ground level. Thus, the muon attenuation model defined in Equation 2.2.1 cannot be applied to such showers.

In contrast, for zenith angles between $\theta = 45°$ and $\theta = 60°$, all showers satisfy $X^\mu_{\max} < X_{\rm gr} \sec\theta$. In this regime, the simplified muon attenuation model described by Equation (2.18) can be reliably applied to infer the transformation of $\rho(N_\mu, X_{\max})$ between $\theta = 45°$ and $\theta = 60°$, i.e., $\rho(N_\mu, X_{\max}\,|\,45°) \to \rho(N_\mu, X_{\max}\,|\,60°)$, using the corresponding value of $\rho(N_\mu^{\rm prod}, X_{\max})$. For showers at $\theta = 45°$, we perform the fitting procedure described in Section 2.2.1 to extract the spectral index γ. Using Equation (2.18), we

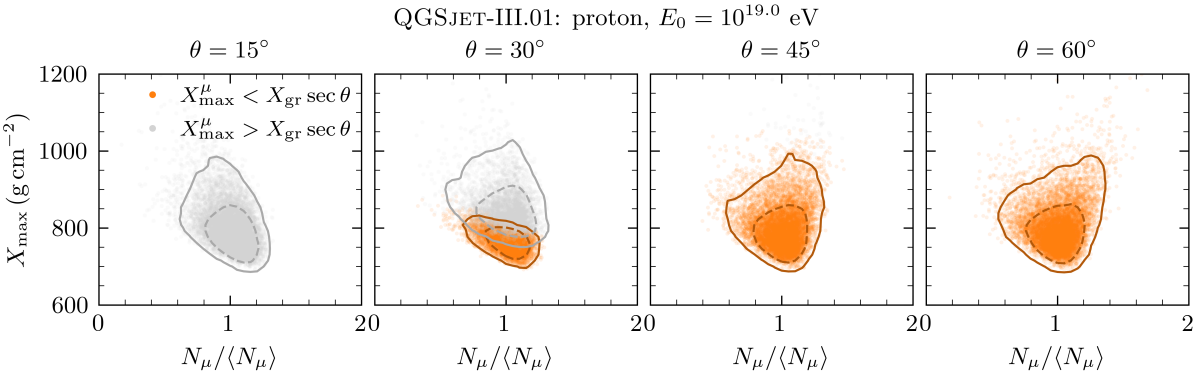

Figure 2.17: Joint distributions $f(N_\mu, X_{\max}\,|\,\theta)$ as a function of the shower's zenith angle, θ (increasing from left to right in steps of $\Delta\theta = 15°$) for two types of showers: $X^\mu_{\max} > X_{\rm gr}\sec\theta$ (peak of muon profile below ground — grey markers) and $X^\mu_{\max} < X_{\rm gr}\sec\theta$ (peak of muon profile above ground — orange markers. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and different values of θ, using QGSJET-III.01.

compute the expected number of surviving muons $\varepsilon_\mu N_\mu^{\rm prod}$ and $\rho(\varepsilon_\mu N_\mu^{\rm prod}, X_{\max})$. We obtain the values $\rho(\varepsilon_\mu N_\mu^{\rm prod}, X_{\max}\,|45°) = \{-0.12, 0.03, -0.06\}$, for EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e, respectively. These values differ by, at most, 0.01 from the values of $\rho(N_\mu, X_{\max}\,|45°)$, showing that the decreased muon attenuation is indeed responsible for the more negative value of $\rho(N_\mu, X_{\max}\,|\,45°)$ relative to $\rho(N_\mu, X_{\max}\,|\,60°)$. Lastly, we note that the evolution of $f(N_\mu, X_{\max}\,|\theta)$ is highly dependent on the hadronic interaction model.

Therefore, we have established that the change in the shape of the joint distribution of N_μ and $X_{\max}$ with θ is a geometrical effect driven by the additional fluctuations induced by the truncation of the muon profile by the ground. The fluctuations come both from muon attenuation and from the alternation of the depth of the peak of the muon profile below and above the ground. To avoid these effects, the main analyses of this work will be conducted with more inclined showers, having $\theta = 60°$, to provide a clearer picture of its connection with the primary interaction. In a realistic measurement, the mixing of zenith angles would greatly affect the shape of $f(N_\mu, X_{\max})$.

2.2.3 Dependence on the primary mass composition

The composition of the cosmic-ray flux $E_0 > 1$ EeV is expected to be mixed and heavier than proton [31, 124]. Therefore, we briefly address the joint distribution $f(N_\mu, X_{\max})$ as a function of the primary mass number: $f(N_\mu, X_{\max}\,|\,A)$.

According to the superposition principle introduced in Section 1.2.4, the average number of muons in an EAS induced by a nucleus with mass number A, $\langle N_\mu(E_0, A)\rangle$ relates to the number of muons in a proton-induced EAS, $\langle N_\mu(E_0, 1)\rangle \equiv \langle N_\mu^p(E_0)\rangle$ as

$$\langle N_\mu(E_0, A)\rangle = A^{1-\beta} N_\mu^p(E_0), \tag{2.20}$$

where $\beta \sim 0.93$ as found in Section 2.2.1. Analogously, the expected value of $X_{\max}$

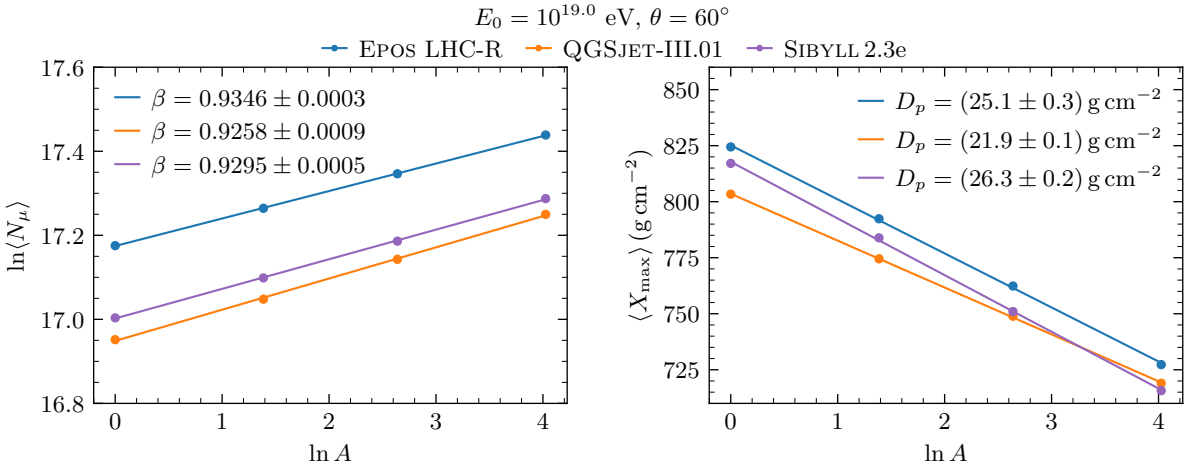

Figure 2.18: $\ln\langle N_\mu \rangle$ (left panel) and $\langle X_{\max} \rangle$ (right panel) as functions of the logarithm of the primary mass number A. The legends show the Heitler-Matthews exponent β (left) and the elongation rate D_p (right) extracted from the linear regression curves depicted as solid lines. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$.

for an shower induced by a nucleus with A nucleons, $\langle X_{\max}(E_0, A) \rangle$, relates to that of proton-induced EAS, $\langle X^p_{\max}(E_0) \rangle$ as

$$\langle X_{\max}(E_0, A) \rangle = \langle X^p_{\max}(E_0) \rangle - D_p \ln A, \quad \text{with} \quad D_p = \frac{\mathrm{d}\,\langle X^p_{\max} \rangle}{\mathrm{d} \ln E_0}, \tag{2.21}$$

where $D_p \simeq 24.3\,\mathrm{g\,cm^{-2}}$ denotes the elongation rate, as determined in Section 2.2.1.

Using libraries of CONEX simulations with 10^4 events, induced by the primaries proton ($A = 1$), helium ($A = 4$), nitrogen ($A = 14$) and iron ($A = 56$), with $E_0 = 10^{19}$ eV and $\theta = 60°$, we test the validity of Equations (2.20) and (2.21). Figure 2.18 shows $\ln\langle N_\mu \rangle$ (left panel) and $\langle X_{\max} \rangle$ (right panel) as a function of the logarithm of the primary mass number $\ln A$ for the three hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e. The panels also show linear fits to the points according to Equations (2.20) and (2.21), from which we extract the values of β and D_p.

Irrespective of the hadronic interaction model, the extracted values of β agree with those obtained in Section 2.2.1. However, the decrease of $\langle X_{\max} \rangle$ with $\ln A$ is slower than predicted by Equation (2.21) for QGSJET-III.01, despite being compatible with the result of Section 2.2.1 for the other hadronic interaction models. In particular, taking the extreme models, the value of $\langle X_{\max} \rangle$ for a proton-induced shower simulated with QGSJET-III.01 is $\sim 20\,\mathrm{g\,cm^{-2}}$ shallower than the EPOS LHC-R prediction, while being only $\sim 8\,\mathrm{g\,cm^{-2}}$ shallower for iron-induced showers. The determination of the reason for this discrepancy is out of the scope of this work.

In reality, the superposition principle, which assumes the independence of the nucleons of the projectile nucleus, is violated, even ignoring collective effects arising in heavy ion collisions. In particular, all nucleons are bounded before interacting at a depth $X_1(A)$, with $\langle X_1(A) \rangle = \lambda_{A-\mathrm{air}} < \lambda_{p-\mathrm{air}}$, making them maximally correlated before interaction. Moreover, the average number of nucleons participating in the primary interaction, also denominated *wounded nucleons*, $\langle \nu_A \rangle$, is usually smaller than A. In fact, the Glauber

theory predicts $\langle \nu_A \rangle = A\lambda_{A-\text{air}}/\lambda_{p-\text{air}}$ [70]. The incident nucleus is thus usually fragmented upon interaction, and the non-interacting nucleons, denominated *spectator* nucleons, keep propagating before interacting again. Within the Glauber framework, the depth of interaction of each nucleon of the projectile nucleus averages it out to $\lambda_{p-\text{air}}$, allowing the superposition principle to predict the correct A-evolution of $\langle X_{\max} \rangle$ [125].

Despite the aforementioned limitations, we investigate the extent to which the superposition principle remains valid in relating the fluctuations of $N_\mu(E_0, A)$ and $X_{\max}(E_0, A)$ to those of $N_\mu^p(E_0)$ and $X_{\max}^p(E_0)$, respectively. Under the superposition assumption, a shower initiated by a nucleus of mass number A and total energy E_0 is modelled as a superposition of A independent sub-showers, each induced by a nucleon with energy E_0/A. We allow both the muon content and the depth of shower maximum of each sub-shower to fluctuate according to the distributions obtained for proton-induced showers. Each sub-shower, indexed by $i \in \{1, \ldots, A\}$, is assumed to develop independently, producing $N_{\mu,i}^p(E_0/A)$ muons and reaching its maximum development at a depth $X_{\max,i}^p(E_0/A)$. The number of muons in the overall shower reads

$$N_\mu(E_0, A) = \sum_{i=1}^{A} N_{\mu,i}^p(E_0/A) = A^{-\beta} \sum_{i=1}^{A} N_{\mu,i}^p(E_0). \tag{2.22}$$

To estimate the overall value of $X_{\max}(E_0, A)$, we consider that each sub-shower has a maximum charged particle content of $N_{\max,i}$ and perform a weighted average of their values of $X_{\max,i}^p(E_0/A)$ using the value of $N_{\max,i}$ as weights. As a first approximation, we neglect the fluctuations of $N_{\max,i}$ and take $N_{\max,i} \propto E_0/A$, to get

$$X_{\max}(E_0, A) = \frac{\sum_{i=1}^{A} \frac{E_0}{A} X_{\max,i}^p(E_0/A)}{\sum_{i=1}^{A} \frac{E_0}{A}} = A^{-1} \left(\sum_{i=i}^{A} X_{\max,i}^p(E_0) \right) - D_p \ln A. \tag{2.23}$$

This shower-to-shower formulation reduces to Equations (2.20) and (2.21) when average values are applied to Equations (2.22) and (2.23), respectively. Since each value of $N_{\mu,i}^p(E_0)$ is sampled from a common distribution with standard deviation $\sigma\left[N_\mu^p(E_0)\right]$, the standard deviation of $N_\mu(E_0, A)$ is given by

$$\sigma[N_\mu(E_0, A)] = A^{\frac{1}{2}-\beta} \sigma\left[N_\mu^p(E_0)\right] \implies \frac{\sigma[N_\mu(E_0, A)]}{\langle N_\mu(E_0, A) \rangle} = \frac{1}{\sqrt{A}} \frac{\sigma\left[N_\mu^p(E_0)\right]}{\langle N_\mu^p(E_0) \rangle}. \tag{2.24}$$

Hence, the relative fluctuations of N_μ in showers induced by heavier primary are suppressed by $\sqrt{A}$. In the same fashion, it is trivial to show that

$$\sigma\left[X_{\max}(E_0, A)\right] = \frac{\sigma\left[X_{\max}^p(E_0)\right]}{\sqrt{A}}. \tag{2.25}$$

We test the superposition model by generating distributions of N_μ and $X_{\max}$ for nuclei with A nucleons from CONEX simulations of proton-induced simulations. This was achieved by sampling A values of $N_\mu^p(E_0)$ and $X_{\max}^p(E_0)$ from proton-induced showers, and computing $N_\mu(E_0, A)$ and $X_{\max}(E_0, A)$ using Equations (2.22) and (2.23). Note that during the resampling, the same events were used to compute simultaneously both shower observables, to keep their correlation.

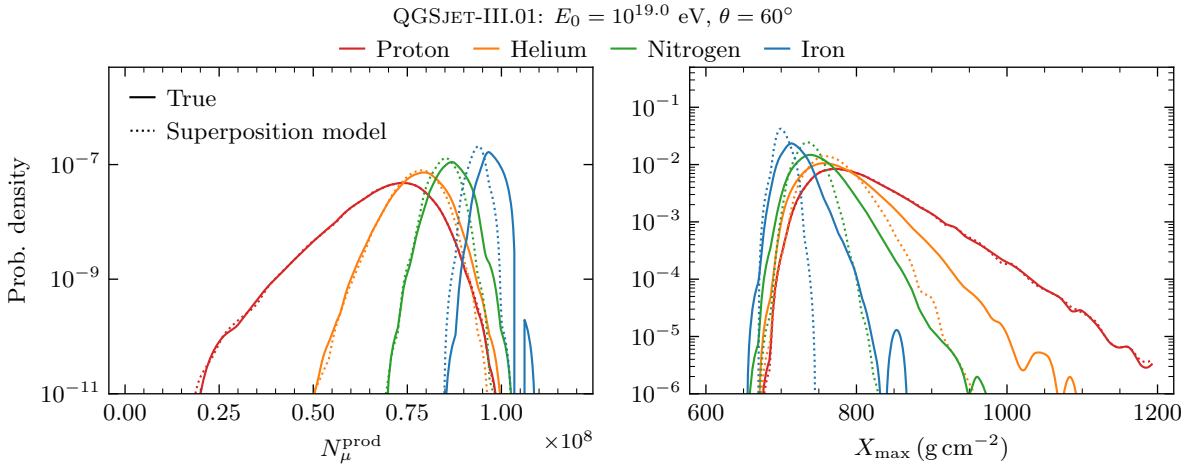

Figure 2.19: Marginal distributions $f(N_\mu^{\text{prod}} \,|\, A)$ (left panel) and $f(X_{\max} \,|\, A)$ (right panel) for different primary mass numbers $A \in \{1, 4, 14, 56\}$. Each panel contains the true distribution (solid curves) and the estimated distribution by the naive application of the superposition principle (dotted curves). This figure was produced with libraries of 10^4 CONEX showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$, using QGSJET-III.01.

The estimated distributions of N_μ^{prod}, $f(N_\mu^{\text{prod}} \,|\, A)$, and $X_{\max}$, $f(X_{\max} \,|\, A)$, obtained with this procedure, are shown in Figure 2.19, together with the true distributions of these quantities. Note that N_μ^{prod} was preferred over N_μ since the distribution of latter includes fluctuations due to muon attenuation and muon profile truncation. These effects depend on $\mathcal{X}_{\max}^{\mu}$ and the energy spectrum of muons at production, which depend on the primary mass [109].

The distribution of N_μ^{prod} for heavier primaries is well described by the superposition principle, as anticipated from the proof provided in [125]. Furthermore, we verified that the principle accurately describes the distribution of N_μ. On the other hand, the distribution of $X_{\max}$ is missing additional fluctuations. In particular, the exponential right tail. These additional fluctuations come from the nuclear fragmentation of the incident primary and the fluctuations of the number of interacting nucleons, as discussed in [119, 125]. A probabilistic treatment of such processes could improve our model and allow for a fast generalisation of distributions of shower observables induced by protons to nuclear primaries. Such a study is beyond the scope of this work.

The joint distributions $f(N_\mu, X_{\max} \,|\, A)$ for all primaries are shown in Figure 2.20, together with a dotted line parametrized by the A dependences of Equations (2.20) and (2.21). The mean values of N_μ and $X_{\max}$ for each primary are represented by the starred markers.

As previously reported in the literature [31, 97], an anti-correlation between N_μ and $X_{\max}$ emerges in the case of a mixed composition of primary particles. This behaviour is naturally explained by the superposition principle, which predicts that heavier nuclei produce showers that develop at shallower depths and contain more muons compared to those initiated by protons. In addition, the reduced shower-to-shower fluctuations in both N_μ and $X_{\max}$ for heavier primaries permit the separation between different mass groups. Specifically, the dependence of the correlation coefficient $\rho(N_\mu, X_{\max})$ on the relative abundances of different mass groups was used in [31] to infer that, within the

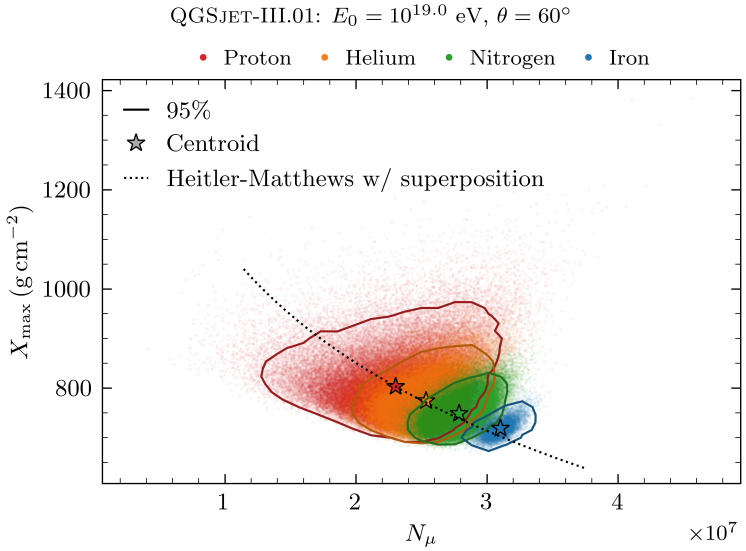

Figure 2.20: Joint distributions $f(N_\mu, X_{\max} \,|\, A)$ for proton ($A = 1$), helium ($A = 4$), nitrogen ($A = 14$) and iron ($A = 56$) primaries. The solid contour encloses 95% of the events. This figure was produced with libraries of 10^5 CONEX showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$, using QGSJET-III.01.

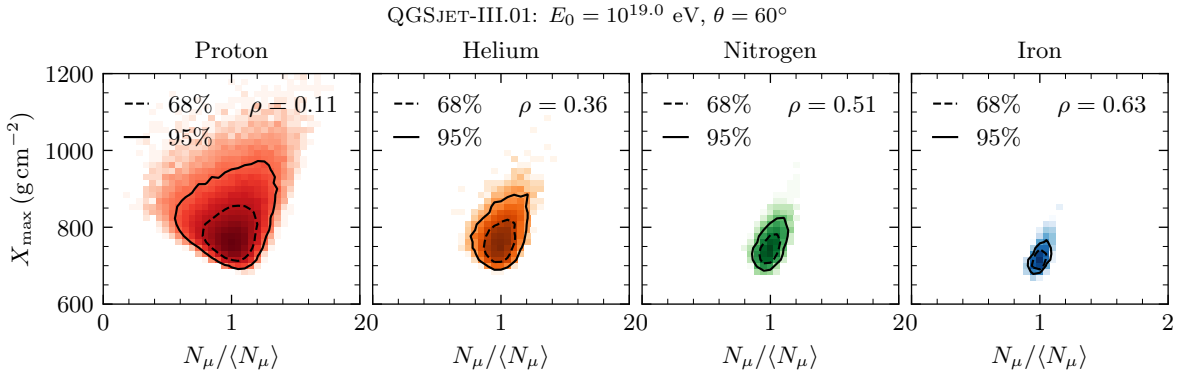

Figure 2.21: Joint distributions $f(N_\mu, X_{\max} \,|\, A)$ as a function of the primary mass number A (increasing from left to right with $A \in \{1, 4, 14, 56\}$). Each panel contains the Pearson correlation coefficient. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. This figure was produced with libraries of 10^5 CONEX showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$, using QGSJET-III.01.

energy range $E_0 = [10^{18.5}, 10^{19.0}]$ eV, there must be primaries with mass number $A > 4$.

The individual joint distributions $f(N_\mu, X_{\max} \,|\, A)$, shown in Figure 2.21, illustrate that the correlation between N_μ and $X_{\max}$ becomes increasingly positive with primary mass A for showers at $\theta = 60°$ and $E_0 = 10^{19}$ eV. This trend is already evident in the corresponding production-level distributions $f(N_\mu^{\text{prod}}, X_{\max} \,|\, A)$, displayed in Figure 2.22. It is worth noting that the variable $\Delta X_{\max}$ is not used in this analysis, since $X_{\max} - X_1$ neglects the presence of nuclear fragmentation processes. Hence, $\Delta X_{\max}$ lacks a useful physical interpretation for heavier primary compositions.

For each primary species, the observation that $\rho(N_\mu, X_{\max})$ is more positive than $\rho(N_\mu^{\text{prod}}, X_{\max})$ can be fully accounted for by the simplified muon attenuation model in-

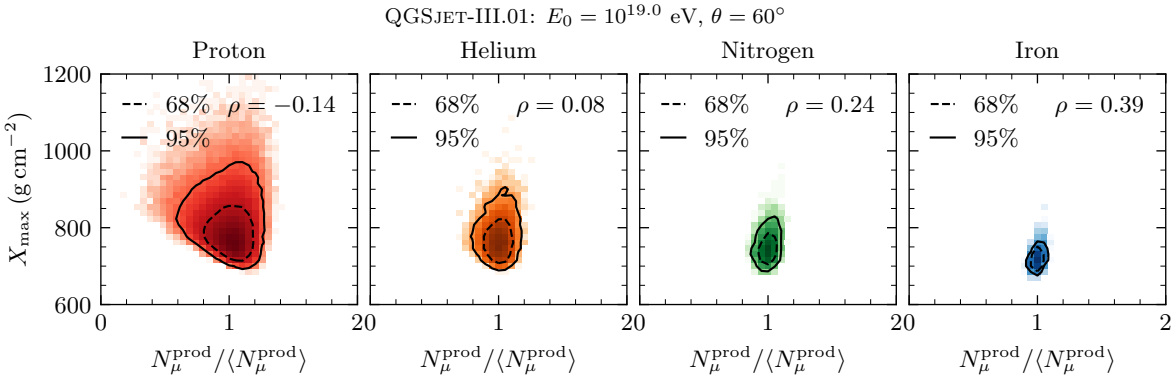

Figure 2.22: Joint distributions $f(N_\mu^{\text{prod}}, X_{\max} \,|\, A)$ as a function of the primary mass number A (increasing from left to right with $A \in \{1, 4, 14, 56\}$). Each panel contains the Pearson correlation coefficient. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. This figure was produced with libraries of 10^5 CONEX showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$, using QGSJET-III.01.

troduced in Equation (2.18). Specifically, by computing the muon survival factor ε_μ using the muon production depth $\mathcal{X}^\mu_{\max}$ for each primary and the proton spectral index γ evaluated at E_0/A, we obtain an estimate of the number of muons reaching the ground: $\varepsilon_\mu N_\mu^{\text{prod}}$. From this, the correlation $\rho(\varepsilon_\mu N_\mu^{\text{prod}}, X_{\max})$ can be computed, as outlined in Section 2.2.1, and directly compared to $\rho(N_\mu, X_{\max})$. Across all tested primary compositions and hadronic interaction models, the difference between $\rho(\varepsilon_\mu N_\mu^{\text{prod}}, X_{\max})$ and $\rho(N_\mu, X_{\max})$ remains below 0.01, confirming the adequacy of the attenuation model in reproducing the observed correlations.

The superposition principle cannot fully account for the increase in $\rho(N_\mu^{\text{prod}}, X_{\max})$ with growing primary mass. Letting $\rho(N_\mu^{\text{prod}}, X_{\max} \,|\, E_0, A)$ denote the Pearson correlation coefficient between N_μ^{prod} and $X_{\max}$ for a primary nucleus with mass number A and total energy E_0, we demonstrate in Appendix B.1 that applying the superposition principle yields

$$\rho(N_\mu^{\text{prod}}, X_{\max} \,|\, E_0, A) = \rho(N_\mu^{\text{prod}}, X_{\max} \,|\, E_0/A, 1). \tag{2.26}$$

Equation (2.26) shows that $\rho(N_\mu^{\text{prod}}, X_{\max})$ should solely depend on the energy per nucleon and thus remain unchanged when moving from proton to heavier nuclei, provided the scaling is applied.

In Section 2.2.1, we showed that, for proton-induced showers, the correlation between N_μ^{prod} and $X_{\max}$ becomes more positive as the primary energy decreases. Consequently, Equation (2.26) predicts a slight increase in the correlation $\rho(N_\mu^{\text{prod}}, X_{\max})$ with increasing mass number A, as the energy per nucleon scales as E_0/A. However, as illustrated in Figure 2.23, this effect is not the dominant contributor to the observed rise of $\rho(N_\mu^{\text{prod}}, X_{\max})$ with A. The values of $\rho(N_\mu^{\text{prod}}, X_{\max})$ used in the comparison were obtained by applying a cubic Akima interpolation [126] to the data shown in Figure 2.12, evaluated at E_0/A.

Importantly, the rate $\mathrm{d}\rho(N_\mu, X_{\max})/\mathrm{d}\ln A$ depends on the considered hadronic interaction model, highlighting differences in their treatment of nuclear interactions. This suggests that nuclear fragmentation and fluctuations in the number of participating nucleons play a leading role in shaping the joint distribution $f(N_\mu, X_{\max} \,|\, A)$ for nuclei with $A > 1$.

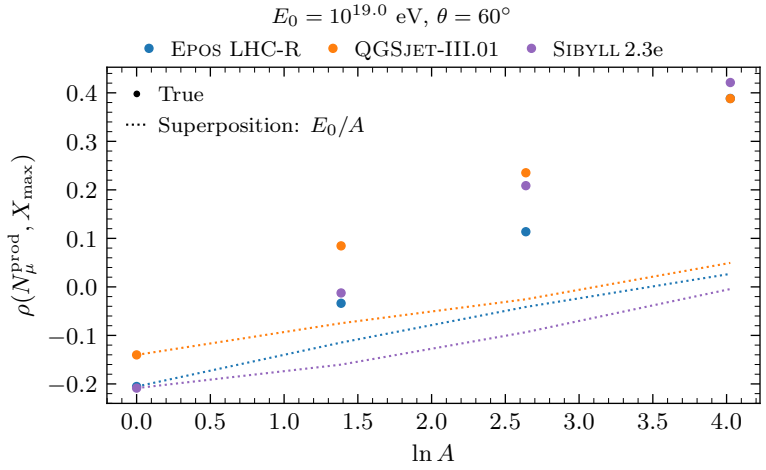

Figure 2.23: True (circular markers) and estimated (dotted lines) Pearson correlation coefficients $\rho(N_\mu^{\text{prod}}, X_{\max})$. The estimated values are obtained from the superposition principle using an interpolation of the values of $\rho(N_\mu^{\text{prod}}, X_{\max})$ shown in Figure 2.12. This figure was produced with libraries of 10^4 CONEX showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$.

As an illustrative example, consider an extreme case of nuclear fragmentation in which A_1 nucleons interact initially, while the remaining $A - A_1$ nucleons form a spectator nucleus that interacts subsequently. This scenario is expected to induce a positive correlation between N_μ and $X_{\max}$. The underlying mechanism can be motivated within the framework of the superposition principle: the muon yield from the resulting sub-cascades, initiated by A_1 and $A - A_1$ nucleons respectively, is given by

$$N_\mu(E_0, A_1) + N_\mu(E_0, A - A_1) = \left[A_1^{1-\beta} + (A - A_1)^{1-\beta}\right] N_\mu^p(E_0) > A^{1-\beta} N_\mu^p(E_0). \quad (2.27)$$

This inequality shows that partial fragmentation enhances the total muon content compared to a fully coherent nucleus of mass A, which is assumed in the classical superposition framework.

In addition, the two separate cascades are each likely to develop deeper in the atmosphere than in the case of a single, intact primary, from which all the nucleons interact. This results in an increased overall $X_{\max}$. Hence, such fragmentation leads to a population of showers with both deeper maxima and enhanced muon content, thereby contributing to a more positive correlation between N_μ and $X_{\max}$. A more realistic treatment of nuclear fragmentation would need to include its stochastic nature. Furthermore, the impact of other phenomena, like the formation of QGP in heavy-ion collisions, would need to be understood. Lastly, these processes are implemented differently in distinct hadronic interaction models.

In summary, the connection between the joint distribution of N_μ and $X_{\max}$ for proton-induced showers and that for heavier primaries is non-trivial. Accurately capturing this relationship necessitates, at least, a detailed modelling of nuclear fragmentation and its fluctuations. While this analysis lies beyond the scope of the present work, it represents a natural extension of the main results of this work, presented in later chapters, and derived for proton-induced showers.

2.2.4 Dependence on the muon energy threshold

Lastly, we investigate the sensitivity of the shape of $f(N_\mu, X_{\max})$ to the energy threshold for the simulation of muons, E_μ^{th}. The motivation for this study is mostly experimental since different cosmic-ray detectors have different thresholds for muon detection. For example, the UMD (see Section 1.4) has an energy threshold for muon detection of $\approx 1\,\text{GeV}$ [93], while a WCD (see Section 1.4) is sensitive to muons such that their velocity in water is larger than the speed of light. Taking the refractive index of pure water to be $n = 1.33$, then all muons having kinetic energy

$$E_\mu^{\text{th}} > \left[\frac{n}{\sqrt{(n^2-1)}} - 1\right] m_\mu \simeq 54.6\,\text{MeV}, \tag{2.28}$$

produce Cherenkov radiation in water, and are therefore detectable.

Furthermore, $X_{\max}$ is sensitive to the early stages of the shower development, while low-energy muons can be produced much later in the shower by processes such as photon $\mu^+\mu^-$-pair production ($\gamma + \text{air} \to \mu^+ + \mu^-$) and photo-pion production ($\gamma + p \to \pi^+ + n \to \mu^+ + \nu_\mu + n$ and $\gamma + n \to \pi^- + p \to \mu + \overline{\nu}_\mu + p$). Thus, fluctuations in the production of lower-energy muons affect the fluctuations of N_μ, for fixed values of $\sigma(X_{\max})$. It is reasonable to assume this could affect the shape of $f(N_\mu, X_{\max})$.

Lastly, it is known that the shape of the lateral distribution of muons at the ground is highly sensitive to E_μ^{th} [123] and to $X_{\max}$ [97]. In particular, the correlation between $X_{\max}$ and the muon density at a fixed distance from the shower core r_0, $\rho_\mu(r_0)$, is highly dependent on E_μ^{th} [127]. Hence, establishing whether this is also the case for $N_\mu = 2\pi \int \rho_\mu(r) r \, \mathrm{d}r$, is worthwhile.

Since the minimum threshold energy for muons in CONEX is $E_\mu^{\text{th}} = 1\,\text{GeV}$, we produced libraries of 1 000 proton-induced CORSIKA [110] simulations, using version 7.7402. Low-energy hadronic interactions were simulated with FLUKA [111, 112]. High-energy hadronic interactions were simulated with EPOS LHC, QGSJET-II.04 and SIBYLL2.3d. The transition between high- and low-energy hadronic interaction models was set to 80 GeV. The fraction of the primary energy below which the thinning algorithm becomes active was set to 10^{-6}, to mitigate the additional fluctuations of shower observables due to the thinning procedure. The kinetic energy thresholds for different shower particles were set to 50 MeV for muons and hadrons and to 250 keV for electrons/positrons and photons. The number of muons, N_μ, was obtained by integrating the muon flux on the ground level at a height of 1400 m a.s.l, above a certain energy threshold. The value of $X_{\max}$ is obtained as described in Section 2.1.

Figure 2.24 shows the joint distribution of N_μ and $X_{\max}$ for four values of E_μ^{th}, $f(N_\mu, X_{\max} \,|\, E_\mu^{\text{th}})$, along with the Pearson correlation coefficient $\rho(N_\mu, X_{\max})$, for proton-induced simulations with $E_0 = 10^{19}$ eV and $\theta = 67°$, using the high-energy hadronic interaction model EPOS LHC.

For $E_\mu^{\text{th}} < 1\,\text{GeV}$ the shape of $f(N_\mu, X_{\max} \,|\, E_\mu^{\text{th}})$ depends little on the energy threshold for muon detection. Thus, the contribution of fluctuations in the production of lower energy muons does not affect the shape of $f(N_\mu, X_{\max} \,|\, E_\mu^{\text{th}})$ in the range of the threshold energies covered by typical buried muon detectors and WCDs. On the contrary, for $E_\mu^{\text{th}} > 1\,\text{GeV}$, N_μ and $X_{\max}$ become increasingly anti-correlated with E_μ^{th}. This behaviour is common to all hadronic interaction models, as shown in Figure 2.25, where the connection

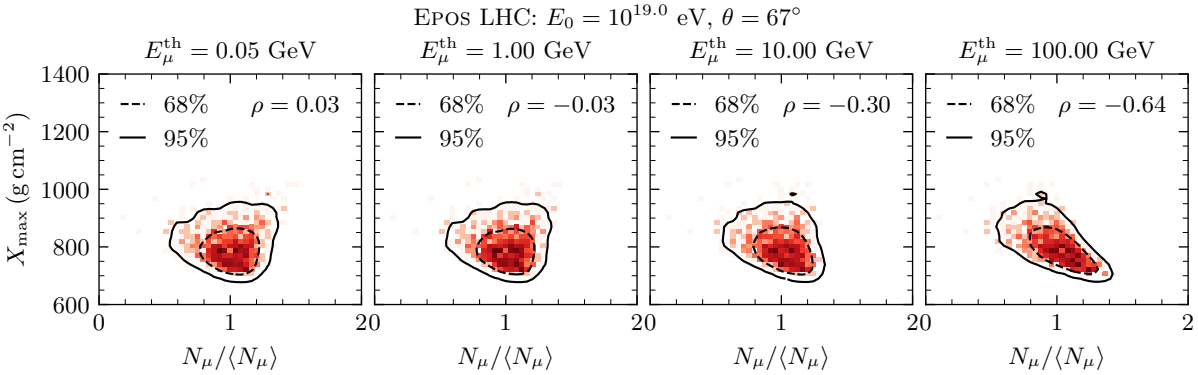

Figure 2.24: Joint distributions $f(N_\mu, X_{\max} \,|\, E_\mu^{\rm th})$ as a function of the muon energy threshold $E_\mu^{\rm th}$ (increasing from left to right with $E_\mu^{\rm th} \in \{0.05, 1, 10, 100\}$ GeV). Each panel contains the Pearson correlation coefficient. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. This figure was produced with libraries of 10^3 CORSIKA showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 67°$, using EPOS LHC.

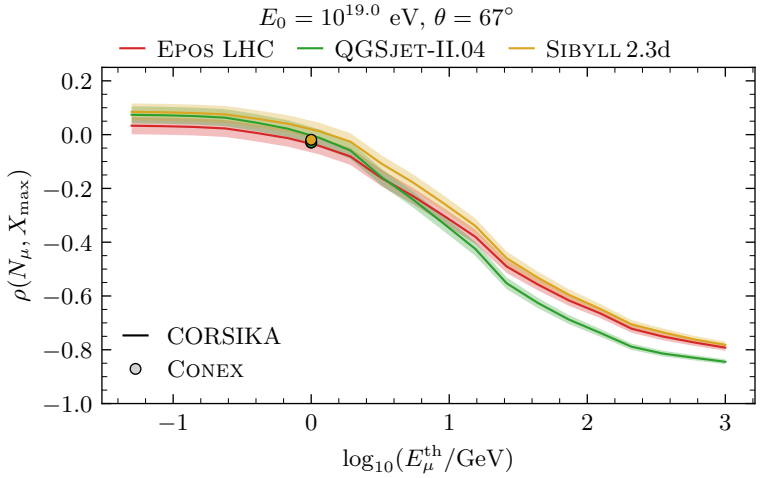

Figure 2.25: Dependence of the Pearson correlation coefficient $\rho(N_\mu, X_{\max})$ on the muon energy threshold, for three hadronic interaction models. This figure was produced with libraries of 10^3 proton-induced CORSIKA showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 67°$.

between N_μ and $X_{\max}$ is assessed through the $E_\mu^{\rm th}$-dependence of their Pearson correlation coefficient.

Figure 2.25 clearly shows that the correlation between N_μ and $X_{\max}$ obtained with CONEX and CORSIKA is consistent across different hadronic interaction models. This agreement is expected, given the relatively low transition energy between Monte Carlo simulations and cascade equations in CONEX, as well as the natural suppression of the impact of fluctuations from individual interactions occurring in the later stages of shower development.

Increasing the muon energy threshold $E_\mu^{\rm th}$ into the dozens and hundreds of GeV corresponds to selecting muons produced during the earlier stages of the shower, either prompt muons or those originating from the decay of higher-energy mesons. These muons are therefore sensitive to the mechanisms of particle production in the highest energy hadronic interactions, where the partitioning of energy between the hadronic and electromag-

netic components tends to induce an anti-correlation between N_μ and $X_{\max}$. Moreover, at these energies, particle production occurs far from the region of the kinematic phase space covered in accelerator experiments. This anti-correlation naturally emerges from the model presented in Chapter 5.

Moreover, the evolution of $\rho(N_\mu, X_{\max})$ with the muon energy threshold above 1 GeV depends on the hadronic interaction model. This opens the possibility of exploring the connection between $X_{\max}$ and N_μ using cosmic-ray experiments capable of separating muons by energy and simultaneously measuring $X_{\max}$. Such experiments could probe muon production at different stages of shower development, providing valuable constraints on hadronic interactions.

One such experiment is IceCube-Gen2 [128], which will be equipped with radio antennas for shower-by-shower estimation of $X_{\max}$; an extended surface array to estimate the muon density down to energies of $\sim 1\,\mathrm{GeV}$; and an enlarged in-ice detector to measure the number of high-energy muons. The feasibility of such analyses must be assessed in the context of a mixed cosmic-ray composition in the energy range $\sim [1, 100]$ PeV covered by the experiment, and the resolution achievable in the reconstruction of the relevant shower observables. Notably, measurements of GeV [129] and TeV [130] muons at IceCube [131] have already revealed inconsistent mass interpretations of the cosmic-ray flux [132].

2.3 Summary and Discussion

In this chapter, we have introduced and characterised the joint distribution $f(N_\mu, X_{\max})$, as well as the marginal distributions $f(N_\mu)$ and $f(X_{\max})$. We identified that the physics of hadron production, primarily connecting the number of produced muons N_μ^{prod} and $\Delta X_{\max}$, tends to anti-correlate these two quantities, and therefore also N_μ and $X_{\max}$. Furthermore, we have analysed the shape of $f(N_\mu, X_{\max})$ as a function of primary energy, zenith angle and mass composition, to characterise the effects of cascade dynamics that shape this distribution, without delving into the physics of hadron production. Such physics is explored in Chapter 5. Understanding these effects is crucial for applying the several results of this work to experimental data.

In particular, the energy evolution of $\langle N_\mu \rangle$ and $\langle X_{\max} \rangle$ revealed significant differences between hadronic interaction models, at the level of approximately 20% and $20\,\mathrm{g\,cm^{-2}}$, respectively. The analysis of the energy dependence of the relative fluctuations of N_μ showed that fluctuations in the charged pion critical energy contribute substantially to the variability of N_μ, alongside the fluctuations of the primary interaction variable α_1. Moreover, we showed that the transformation from $f(N_\mu^{\mathrm{prod}}, X_{\max})$ to $f(N_\mu, X_{\max})$ can be accurately reproduced with a simple muon attenuation model that accounts for the hardness of the muon spectrum at the depth of the muon production profile maximum, γ, and the depth at which this maximum occurs, $\mathcal{X}^\mu_{\max}$. Crucially, the observed increase in the anti-correlation between N_μ and $X_{\max}$ with primary energy is driven solely by the physics of hadron production.

We also inferred that the zenith angle dependence of $f(N_\mu, X_{\max})$ is due to a purely geometric effect. Besides muon attenuation, fluctuations arising from the truncation of the muon profile by the ground around the maximum of this profile contribute to a large anti-correlation between N_μ and $X_{\max}$.

Regarding the primary mass dependence of $f(N_\mu, X_{\max})$, we demonstrated that the anti-correlation between N_μ and $X_{\max}$ for a mixed primary beam can be understood within the Heitler-Matthews model and the superposition principle. Using this latter principle, we showed that $f(N_\mu)$ for a nucleus of mass number A can be approximated reasonably well by an A-fold convolution of the N_μ distribution for a primary energy scaled by $1/A$. However, nuclear fragmentation prevents an accurate description of $f(X_{\max})$ and, more importantly, of the full joint distribution $f(N_\mu, X_{\max})$ for heavier primaries and under this superposition hypothesis. This limitation will in turn restrict any straightforward generalisation of the conclusions derived for proton showers to showers initiated by heavier primaries.

Finally, we verified that the shape of $f(N_\mu, X_{\max})$ depends only weakly on the muon detection energy threshold for thresholds below 1 GeV, although a clear evolution is seen across several orders of magnitude in threshold. Moreover, this evolution depends on the hadronic interaction model, which provides an interesting opportunity for experiments that can separate GeV and TeV muons, such as IceCube-Gen2, to probe this behaviour and thus constrain hadronic interactions at PeV energies.

3

DISTRIBUTION OF N_μ IN MUON-DEPLETED SHOWERS AS A FUNCTION OF $X_{\max}$

In this chapter, we investigate the evolution of the distribution of the number of muons, N_μ, as a function of the depth of the shower maximum, $X_{\max}$. In particular, we focus on characterising the $X_{\max}$-dependence of the slope of the N_μ distribution in muon-depleted showers, denoted Λ_μ, and demonstrate that this observable is highly sensitive to the underlying hadronic interaction model. Specifically, while Λ_μ exhibits a universal value across models for shallow showers, it becomes increasingly model-dependent for deep showers.

We further show that $X_{\max}$ serves as a proxy for the *hadronic activity* of the primary proton–air interaction, providing access to the energy spectrum of neutral pions across distinct regions of the interaction's kinematic phase space via Λ_μ. Importantly, we demonstrate that an unbiased measurement of Λ_μ is achievable even under realistic, mixed mass-composition scenarios. Moreover, the statistical precision required to discriminate between hadronic interaction models can be attained at existing extensive air-shower observatories.

The results presented in this chapter for the hadronic interaction models EPOS LHC, QGSJET-II.04, and SIBYLL2.3d were originally published in [133]. Details of the article can be found in Chapter Published content in peer-reviewed journals. The article is published in an *open access* journal, which can be validated here https://www.sciencedirect.com/science/article/pii/S0370269324006737?via%3Dihub, and under the Creative Commons licence that can be consulted here: https://creativecommons.org/licenses/by/4.0/.

We also provide an updated analysis incorporating the newer models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e, which have been tuned to recent accelerator data. Notably, this includes new constraints from LHCf [134], whose forward neutral pion measurements are particularly relevant for refining predictions of Λ_μ.

3.1 Evolution of Λ_μ with $X_{\max}$

We begin by examining the connection between the shape of the distribution $f(N_\mu, X_{\max})$ and the physics governing primary proton-air interactions. Specifically, we analyse the distribution of $\ln N_\mu$ in bins of $X_{\max}$, henceforth denoted by $f(\ln N_\mu \mid X_{\max})$. As established in [106], the shape of the distribution $f(\ln N_\mu)$ for muon-depleted, proton-induced

showers is largely driven by the hardness of the neutral pion energy spectrum in the primary interaction. A harder spectrum increases the likelihood of producing highly energetic neutral pions, which channel more energy into the electromagnetic cascade. This, in turn, reduces the energy available for the hadronic cascade, leading to fewer muons from hadron decay. Consequently, the probability of muon-depleted showers increases, and the tail of $f(\ln N_\mu)$ becomes more prominent toward lower values of N_μ.

Since the decay of neutral pions initiates electromagnetic cascades, which rapidly decouple from the hadronic one, a larger energy transfer into neutral pions results in deeper electromagnetic sub-showers, as they are fed by more energetic photons [50]. This suggests a correlation between $f(\ln N_\mu)$ for muon-depleted showers and the depth of the shower maximum, $X_{\max}$.

This connection is illustrated in Figure 3.1, which shows the $f(\ln N_\mu, X_{\max})$ (lower panel) and the conditional distributions $f(\ln N_\mu \,|\, X_{\max})$ (upper panel). The figure was produced using a library of 10^5 CONEX simulations of proton-induced air showers, using the set-up described in Section 2.1, with $E_0 = 10^{19}$ eV and $\theta = 60°$. The colour of each distribution $f(\ln N_\mu \,|\, X_{\max})$ matches that of the marker indicating the centre of the corresponding $X_{\max}$-bin in the central panel. The chosen bin edges in $X_{\max}$ are {650, 750, 800, 850, 900, 1300} $\mathrm{g\,cm^{-2}}$. While this binning was not optimised, it was selected to ensure comparable statistics across all bins. Near the peak of the $X_{\max}$ distribution, bins are wider than the typical experimental resolution in $X_{\max}$, approximately $20\,\mathrm{g\,cm^{-2}}$ [76]. Toward the tails of the distribution, wider bins are required to obtain a statistically reliable shape of $f(\ln N_\mu \mid X_{\max})$, particularly given the limited number of events observed in cosmic-ray experiments. Additionally, the binning is chosen to be independent of the hadronic interaction model and the primary composition, and hence of the detailed shape of the $X_{\max}$ distribution. This allows for a robust characterisation of the features in $f(\ln N_\mu)$ and the application of the developed analysis to the measured cosmic-ray data.

The conditional distributions $f(\ln N_\mu \,|\, X_{\max})$ exhibit tails toward low values of N_μ, which correspond to muon-depleted showers. We formally define these tails as the subset of showers with a muon content below the 10% quantile of $f(\ln N_\mu \,|\, X_{\max})$, for each bin in $X_{\max}$. To quantify the flatness of these tails, we perform an unbinned likelihood fit to the functional form $y = B \exp\{\ln N_\mu / \Lambda_\mu(X_{\max})\}$, from which we extract the slope parameter $\Lambda_\mu(X_{\max})$. Steeper tails correspond to smaller values of Λ_μ. We have verified that the extracted value of Λ_μ is little dependent on the precise quantile used to define the tails, provided it remains below the bulk of the distribution.

The resulting fits are shown in the upper panel of Figure 3.1, illustrating that the tail of the muon number distribution becomes more prominent with the depth of the shower maximum. The dependence of Λ_μ on $X_{\max}$ is presented in Figure 3.2, for the earlier hadronic interaction models EPOS LHC, QGSJET-II.04, and SIBYLL2.3d (left panel), as well as their latest versions EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e (right panel). The insets display the distributions of $\ln N_\mu$ for the shallowest and deepest $X_{\max}$ bins.

Figure 3.2 clearly shows that Λ_μ increases monotonically with $X_{\max}$, regardless of the high-energy hadronic interaction model employed. Focusing on the models analysed in [133] (see the left panel of the same figure), we observe that the predicted values of Λ_μ are largely model-independent for shallower showers, but highly dependent on the hadronic interaction model in deeper showers. In particular, for the deepest $X_{\max}$ bin, the spread in Λ_μ relative to QGSJET-II.04 approaches $\sim 50\%$. For context, when Λ_μ is

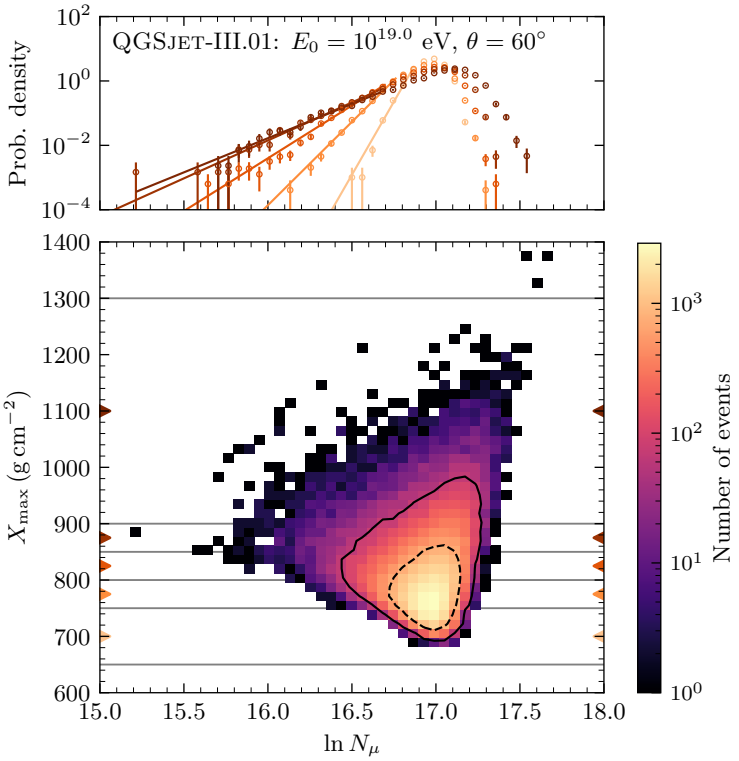

Figure 3.1: Lower panel: joint distribution of $\ln N_\mu$ and $X_{\max}$: $f(\ln N_\mu, X_{\max})$. The black contours contain 68% (dashed) and 95% (solid) of the shower simulations. The horizontal lines represent bins in $X_{\max}$ with edges $\{650, 750, 800, 850, 900, 1300\}\,\mathrm{g\,cm^{-2}}$ and centers indicated by the triangular markers. Higher values of $X_{\max}$ are represented by darker shades of orange and lower ones by lighter shades. Upper panel: distribution of $\ln N_\mu$ for each bin in $X_{\max}$, $f(\ln N_\mu \mid X_{\max})$, matching the colour of the corresponding $X_{\max}$-bin. This figure was produced with a library of 10^5 CONEX simulations of proton-induced showers using QGSJET-III.01.

extracted by fitting the distribution of N_μ integrated over all $X_{\max}$, the model dependence is approximately 20%. Most of the model-dependence of Λ_μ for deep showers is driven by EPOS LHC; however, its updated version, EPOS LHC-R, predicts a lower Λ_μ value in the deepest $X_{\max}$ bin, thereby reducing the overall model uncertainty to roughly 30% when using QGSJET-III.01 as a reference. Given the established connection between Λ_μ and the hardness of the neutral pion energy spectrum in the primary interaction, this reduction in model uncertainty is consistent with measurements of the forward neutral pion spectrum by LHCf [134], which condition the tuning of newer hadronic interaction models, particularly EPOS LHC-R. This sensitivity of Λ_μ to model tuning highlights its potential as a probe of hadron production mechanisms at the highest energies. Finally, we verified that both the monotonic increase of Λ_μ with $X_{\max}$ and the increasing divergence among model predictions remain qualitatively unchanged when varying the widths of the $X_{\max}$ bins, provided the same region of the $X_{\max}$ distribution, $f(X_{\max})$, is sampled.

To ensure that the observed evolution of Λ_μ with $X_{\max}$, as well as the growing model dependence, are not artifacts of the $X_{\max}$-dependent atmospheric attenuation of the muon flux (which influences $f(N_\mu, X_{\max})$, as discussed in Section 2.2.1), we repeated the analysis using the distribution of the number of produced muons, N_μ^{prod}, within each $X_{\max}$ bin. For clarity, we define $\Lambda_\mu^{\mathrm{prod}}$ as the slope characterising the tail of the distribution $f(\ln N_\mu^{\mathrm{prod}} \mid$

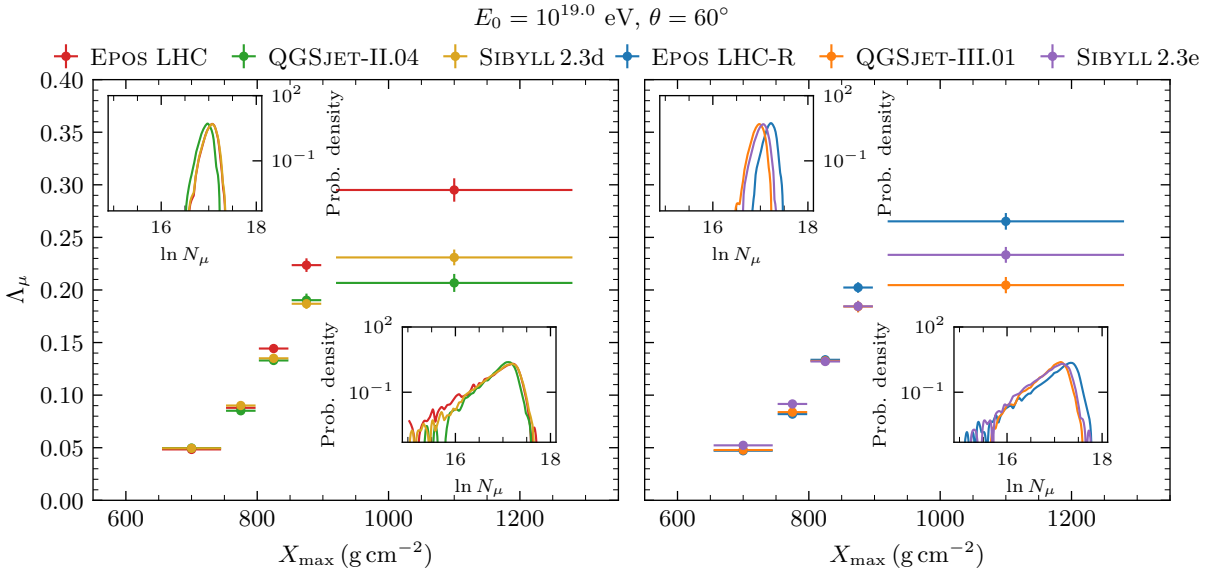

Figure 3.2: Λ_μ as a function of $X_{\max}$ for the hadronic interaction models EPOS LHC, QGSJET-II.04 and SIBYLL2.3d (left panel) and for EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e (right panel), for proton-induced EAS with $E_0 = 10^{19}$ eV and $\theta = 60°$. The distributions of $\ln N_\mu$ for the shallowest and deepest bins in $X_{\max}$ are shown in the left and right insets, respectively, of each panel.

$X_{\max}$) for muon-depleted showers. Figure 3.3 displays the evolution of both Λ_μ and $\Lambda_\mu^{\text{prod}}$ as functions of $X_{\max}$ for various hadronic interaction models. The shaded bands represent the uncertainty in Λ_μ. The lower panels of the figure show the spread in both Λ_μ and $\Lambda_\mu^{\text{prod}}$ across different hadronic interaction models, as a function of $X_{\max}$. Specifically, for a fixed $X_{\max}$-bin, let $\Lambda_\mu^{\max}$ and $\Lambda_\mu^{\min}$ denote, respectively, the maximum and minimum values of Λ_μ obtained among the various models. The spread in Λ_μ is then quantified by $\Delta\Lambda_\mu = \Lambda_\mu^{\max} - \Lambda_\mu^{\min}$.

Irrespective of the considered $X_{\max}$-bin and the hadronic interaction model, the value of $\Lambda_\mu^{\text{prod}}$ is lower than Λ_μ, indicating that the tail of $f(\ln N_\mu^{\text{prod}} \,|\, X_{\max})$ tends to be steeper than that of $f(\ln N_\mu \,|\, X_{\max})$. Otherwise, the trend of $\Lambda_\mu^{\text{prod}}$ with $X_{\max}$ is qualitatively similar to that of Λ_μ, with a similarly large model dependence for deep showers. Therefore, these behaviours are not due to the dependence of muon attenuation on $X_{\max}$.

3.1.1 Dependence on the threshold for muon-detection

The evolution of Λ_μ with $X_{\max}$ was further validated using the same library of CORSIKA simulations employed in Section 2.2.4. This air-shower generator provides a more comprehensive stochastic treatment of the cascade development and includes the simulation of particles down to lower kinetic energies. In particular, it allows for the investigation of the sensitivity of the $X_{\max}$-evolution of Λ_μ to fluctuations in the production and propagation of low-energy muons. These muons can be simulated down to kinetic energies of 50 MeV, and their shower-to-shower fluctuations are largely uncorrelated with those of $X_{\max}$, which are predominantly driven by high-energy hadronic interactions occurring in the early stages of the shower.

Figure 3.4 shows the evolution of Λ_μ with $X_{\max}$ for CORSIKA simulations with muon

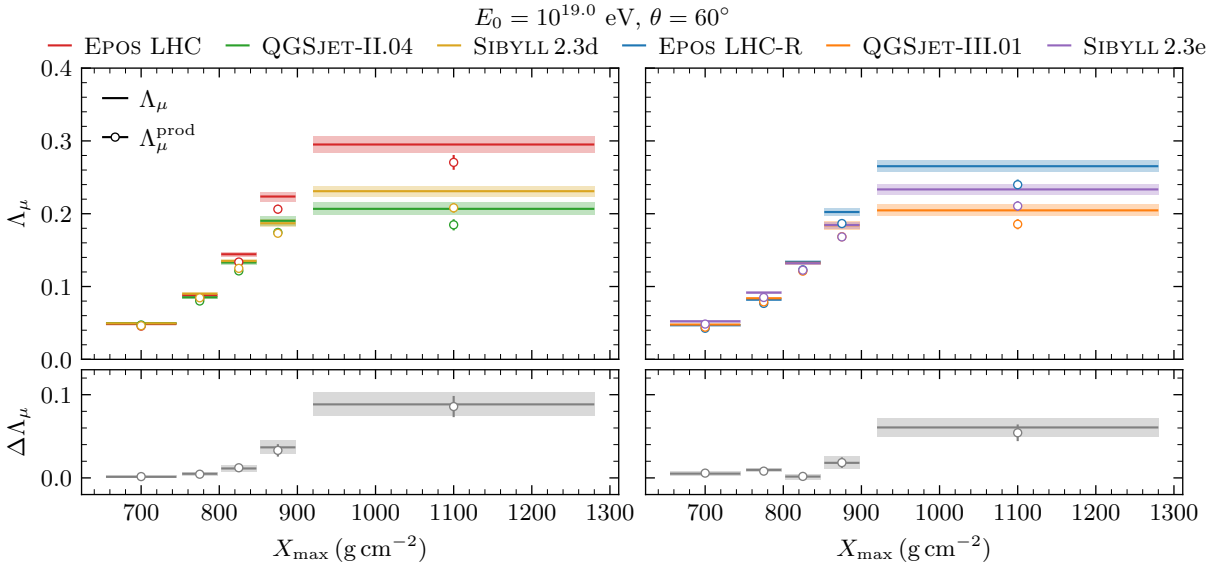

Figure 3.3: Upper panels: Λ_μ (steps with shaded uncertainties) and $\Lambda_\mu^{\text{prod}}$ (circular empty markers with error bars) as a function of $X_{\max}$ for the hadronic interaction models EPOS LHC, QGSJET-II.04 and SIBYLL2.3d (left panel) and for EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e (right panel), for proton-induced EAS with $E_0 = 10^{19}$ eV and $\theta = 67°$. Lower panels: spread in the hadronic model predictions of Λ_μ or $\Lambda_\mu^{\text{prod}}$.

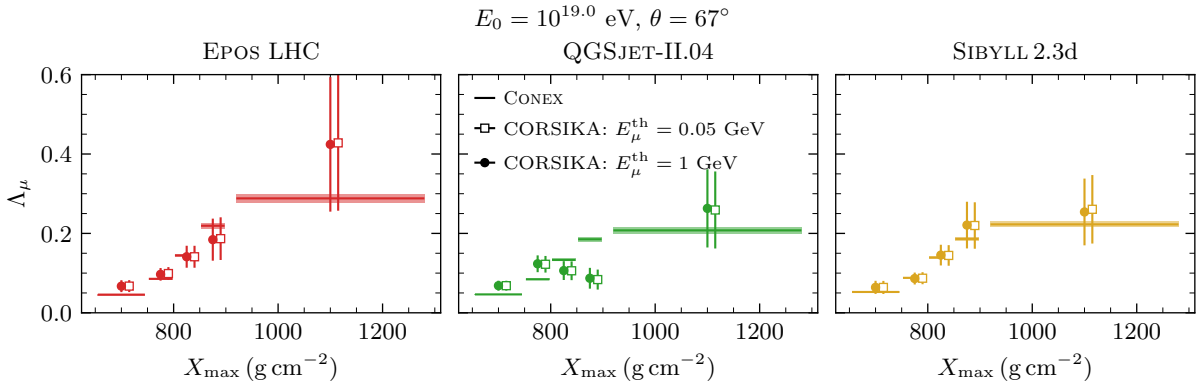

Figure 3.4: Λ_μ as a function of $X_{\max}$ for CONEX proton-induced showers (steps with shaded uncertainties) for CORSIKA simulations with kinetic energy thresholds for muon simulation: $E_\mu^{\text{th}} = 1\,\text{GeV}$ (solid circular markers) and $E_\mu^{\text{th}} = 0.05\,\text{GeV}$ (empty square markers, horizontally displaced to aid the visual separation from points obtained at higher energy threshold). This figure was produced using proton-induced EAS with $E_0 = 10^{19}$ eV and $\theta = 67°$, employing the high-energy hadronic interaction models EPOS LHC (left panel), QGSJET-II.04 (middle panel) and SIBYLL2.3d (right panel).

energy thresholds $E_\mu^{\text{th}} = 0.05\,\text{GeV}$ and $E_\mu^{\text{th}} = 1\,\text{GeV}$, as well as for CONEX-induced showers. Except for a few bins for the hadronic interaction model QGSJET-II.04, the values of Λ_μ obtained with both simulation codes are compatible across $X_{\max}$ bins. Moreover, the results for both energy thresholds are mutually compatible, indicating that the inclusion of additional low-energy muons has a negligible impact on the shape of the $\ln N_\mu$ distribution in the region dominated by muon-depleted showers. These findings are consistent with the reduced sensitivity of $f(N_\mu, X_{\max})$ to E_μ^{th} demonstrated in Section 2.2.4.

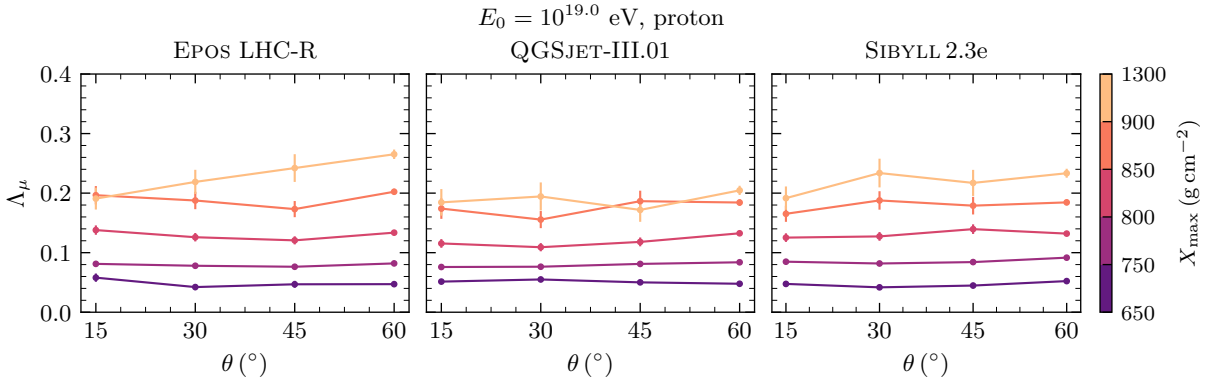

Figure 3.5: Λ_μ as a function of θ for the different bins in $X_{\max}$, indicated by the color scale. From left to right, each panel corresponds to the hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e, respectively. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and different values of θ.

3.1.2 Dependence on zenith angle

As discussed in Section 2.2.2, $f(N_\mu, X_{\max} \,|\, \theta)$, is highly dependent on the shower's zenith angle θ. On the contrary, $f(X_{\max})$ is independent of θ due to the fast decoupling of the electromagnetic cascade from the hadronic core of the shower. Thus, the relation

$$f(\ln N_\mu \,|\, X_{\max}) = \frac{f(\ln N_\mu, X_{\max})}{f(X_{\max})} = \frac{1}{N_\mu} \frac{f(N_\mu, X_{\max})}{f(X_{\max})}, \tag{3.1}$$

suggests that $f(\ln N_\mu \,|\, X_{\max})$ depends on θ, justifying the study of $\Lambda_\mu(X_{\max})$ as a function of θ. Figure 3.5 shows Λ_μ as a function of θ for each bin in $X_{\max}$, as indicated by the colour scale. Each panel corresponds to a different hadronic interaction model.

Across the hadronic interaction models, the variation of Λ_μ with zenith angle θ ranges from approximately 7% to 30%, depending on the specific $X_{\max}$ bin. In particular, the θ-dependence is most pronounced for the shallowest and deepest $X_{\max}$ bins. When restricting the analysis to showers with $\theta \geq 45°$, which ensures that the muonic component is fully developed before reaching the ground level, the variation in Λ_μ due to changes in θ is reduced to around 15%. Additionally, there is no clear monotonic trend of Λ_μ with θ, except in the case of the deepest $X_{\max}$ bin for EPOS LHC-R, where a systematic dependence is observed.

As shown in Figure 3.6, both the monotonic increase of Λ_μ with $X_{\max}$ and its growing dependence on the hadronic interaction model are consistently observed across all zenith angles. Considering the most up-to-date model, the differences between models are reduced in all bins in $X_{\max}$. In this figure, shaded bands represent the spread in Λ_μ values obtained by varying θ for each $X_{\max}$ bin. The coloured markers indicate the fitted Λ_μ for an ensemble of showers with zenith angles $\theta \in \{15°, 30°, 45°, 60°\}$ sampled according to the distribution $\cos\theta \sin\theta$, which approximates the expected angular distribution of extensive air showers from an isotropic cosmic-ray flux observed by a fully efficient horizontal detector array.

Therefore, despite the dependence of $f(\ln N_\mu \,|X_{\max})$ on θ, the measured values of Λ_μ and their evolution with $X_{\max}$, for a sample of cosmic ray events with different zenith

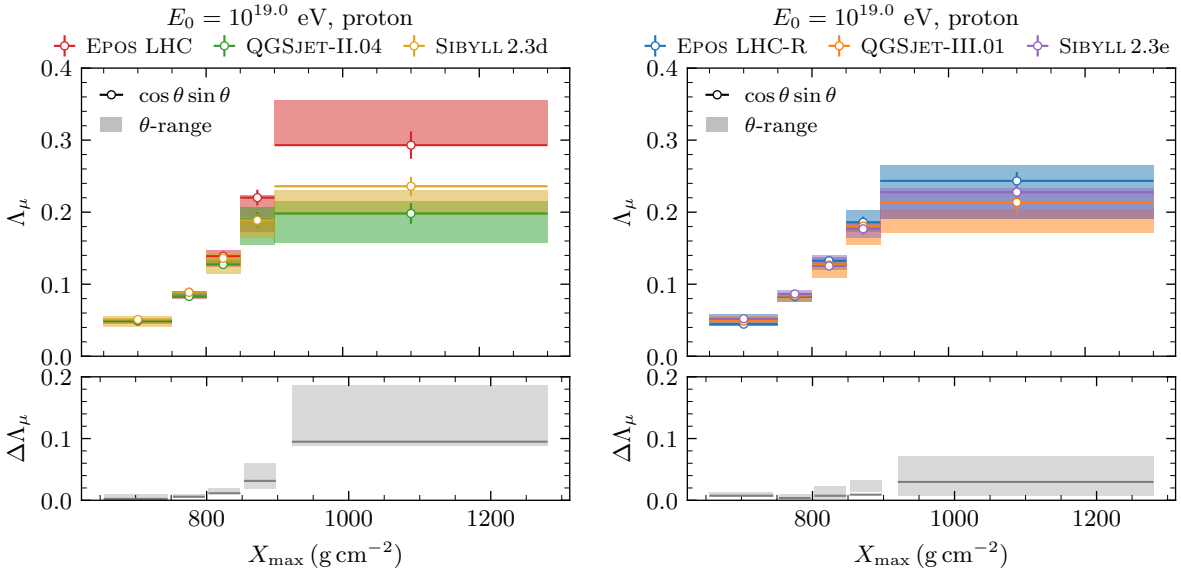

Figure 3.6: Upper panels: Λ_μ as a function of $X_{\max}$ for a sample of showers with zenith angles sampled from $\cos\theta \sin\theta$ using with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta \in \{15°, 30°, 45°, 60°\}$. The shaded bands indicate the spread in Λ_μ over the range of θ angles, for each $X_{\max}$-bin. Lower panels: spread in Λ_μ over the different hadronic interaction models, for each bin in $X_{\max}$. The shaded band is the spread in the model dependence of Λ_μ over the range of θ angles.

angles, are identical to the values obtained at fixed θ. This reinforces the message that the evolution of Λ_μ with $X_{\max}$ could be measured in cosmic ray observatories over a wide range of zenith angles, providing a good discrimination between hadronic interaction models.

3.1.3 Dependence on primary energy

As discussed in Section 2.2.1, the shape of $f(N_\mu, X_{\max} \,|\, E_0)$ is strongly dependent on the primary energy, a dependence that might be reflected in the evolution of Λ_μ with $X_{\max}$. Because $X_{\max}$ increases logarithmically with E_0, we analyse the evolution of $\Lambda_\mu(X_{19})$, as defined in Equation (2.8). Note that, since this definition involves only a constant shift, we have $f(\ln N_\mu, X_{\max}) = f(\ln N_\mu, X_{19})$. The binning in X_{19} for all primary energies is that of $X_{\max}$ defined at $E_0 = 10^{19}$ eV.

Figure 3.7 shows Λ_μ as a function of the primary energy for different bins in X_{19}, with the binning indicated by a corresponding colour scale. Each panel in the figure corresponds to one of the up-to-date hadronic interaction models. Moreover, we performed a χ^2 fit of Λ_μ as a function of E_0 to a linear model; the best-fit curves are superimposed as dashed lines.

For each bin in $X_{\max}$, except for the deepest one, Λ_μ decreases with increasing primary energy. This trend indicates that the distribution $f(\ln N_\mu \,|\, X_{\max})$ becomes broader at lower energies and is consistent with the observations discussed in Section 2.2.1. The slopes obtained from the linear fits to $\Lambda_\mu(E_0)$ are negative and significantly different from zero, with deviations exceeding 5σ for the first three bins in X_{19}. For the deeper bins, however, a constant value of Λ_μ as a function of E_0 cannot be excluded within statistical

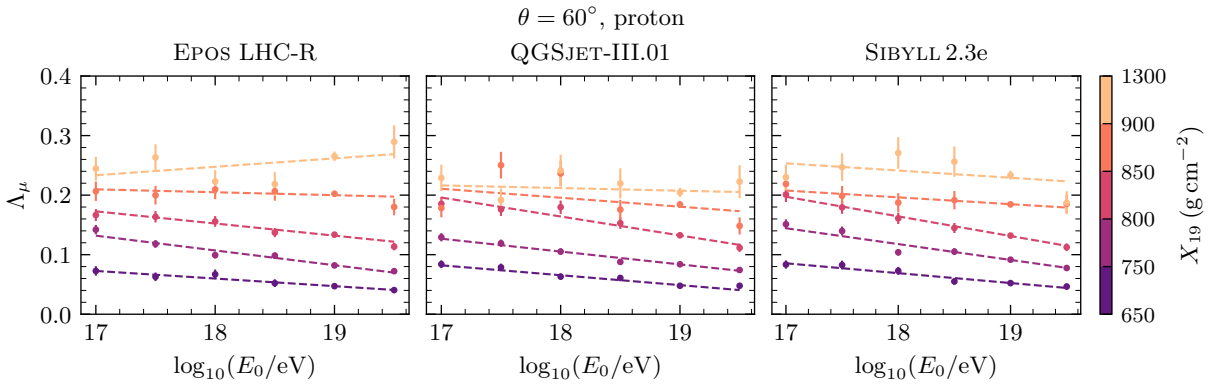

Figure 3.7: Λ_μ as a function of E_0 for the different bins in $X_{\max}$, indicated by the color scale. From left to right, each panel corresponds to the hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e. Dashed lines are the result of a χ^2 fit of $\Lambda_\mu(E_0)$ to a linear function, for each bin in $X_{\max}$. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with zenith angle $\theta = 60°$ eV.

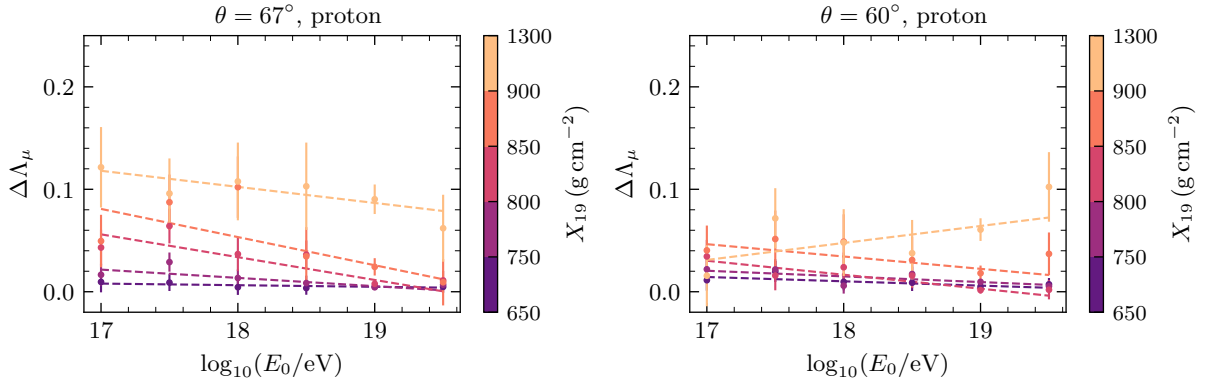

Figure 3.8: Model-spread in Λ_μ, $\Delta\Lambda_\mu$, as a function of $\log_{10}(E_0/\mathrm{eV})$ for the different bins in X_{19}, indicated by the color scale. The left panel referred to the high-energy hadronic interaction models EPOS LHC, QGSJET-II.04 and SIBYLL2.3d, and the right panel to their most up-to-date versions EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e. Dashed lines are the result of a χ^2 fit of $\Delta\Lambda_\mu(E_0)$ to a linear function, for each bin in $X_{\max}$. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with zenith angle $\theta = 60°$ eV.

uncertainties. Nonetheless, regardless of the primary energy, Λ_μ consistently increases with $X_{\max}$, reaffirming the trend already established for $E_0 = 10^{19}$ eV.

Figure 3.8 shows, for each bin in $X_{\max}$, the spread in Λ_μ, $\Delta\Lambda_\mu$, across hadronic interaction models, as a function of the primary energy. The left panel pertains to the hadronic interaction models used in [133] and the right panel to their up-to-date versions. For each $X_{\max}$ bin, the value of $\Delta\Lambda_\mu$ as a function of energy was fitted to a linear function of the form $y = mx + b$.

Regarding the older models, that is, EPOS LHC, QGSJET-II.04, and SIBYLL2.3d, the model dependence of Λ_μ increases with $X_{\max}$, irrespective of the primary energy. Moreover, for each fixed $X_{\max}$-bin, $\Delta\Lambda_\mu$ tends to decrease with increasing primary energy, although the slope of this decrease is compatible with zero given the large fit uncertainties. In contrast, the values of Λ_μ predicted by the most recent versions of the models exhibit a smaller spread, independently of the primary energy. Furthermore, no

statistically significant trend in $\Delta\Lambda_\mu$ with energy is observed for these newer models.

3.2 Connecting the evolution of Λ_μ with $X_{\max}$ with the primary interaction

The connection between Λ_μ and the hardness of the spectrum of neutral pions produced in the primary p-air interaction was established in [106], and summarised in Section 1.5.

Moreover, the connection between the moments of the distributions of $X_{\max}$ and $\ln N_\mu$ and those of the distributions of macroscopic interaction variables, such as the multiplicity, m_{tot} and elasticity, κ_{el} are studied in [55, 107], and summarised in Section 1.5

Here, we explore the connection between $X_{\max}$ and the *hadronic activity* of the primary interaction. High hadronic activity is characteristic of highly inelastic primary interactions, where a large number of secondary particles are produced and evenly disperse the primary energy. Furthermore, we connect the evolution of Λ_μ with $X_{\max}$ with the evolution of the energy spectrum of neutral pions with the degree of hadronic activity of the primary interaction.

3.2.1 Probing the hadronic activity of the primary interaction with $X_{\max}$

To study the evolution of Λ_μ as a function of $X_{\max}$, and given that Λ_μ is particularly sensitive to the characteristics of the primary interaction, we analyse the distributions of multiplicity, elasticity, and the fraction of hadronic energy, α_{had}, from the primary interaction, binned according to $X_{\max}$ as described in Section 3.1. These distributions are shown in Figure 3.9 for the hadronic interaction model QGSJET-III.01. The centres of the $X_{\max}$ bins are indicated by the colour scale, with deeper showers corresponding to darker shades of orange.

Furthermore, since fluctuations in X_1 are uncorrelated with particle production in the primary interaction, we also examine these primary interaction variables using an alternative binning in $\Delta X_{\max}$, shifted by $-25\,\text{g}\,\text{cm}^{-2}$ relative to the $X_{\max}$ binning. The corresponding distributions are shown in Figure 3.10.

Deeper showers are more likely to result from more elastic primary interactions with lower multiplicities, in which the primary energy is distributed more asymmetrically among the secondary particles. Such interactions are characterized by reduced hadronic activity. Conversely, shallower showers tend to arise from more deeply inelastic primary interactions with higher multiplicities, leading to increased hadronic activity. Intuitively, cascades initiated by primary interactions that produce a greater number of secondaries degrade the primary energy more rapidly, resulting in showers that reach their maximum development higher in the atmosphere [55]. Similarly, an increase in the elasticity of the most energetic proton-air interactions leads to slower shower development, with the shower maximum occurring deeper in the atmosphere. In particular, highly elastic primary interactions cause the leading proton to initiate a sub-shower, on average, an additional interaction length beyond X_1.

Also noteworthy is the steepening of the left tail of the α_{had} distribution with decreasing $X_{\max}$, accompanied by a rising frequency of interactions with $\alpha_{\text{had}} \simeq 1$. The former

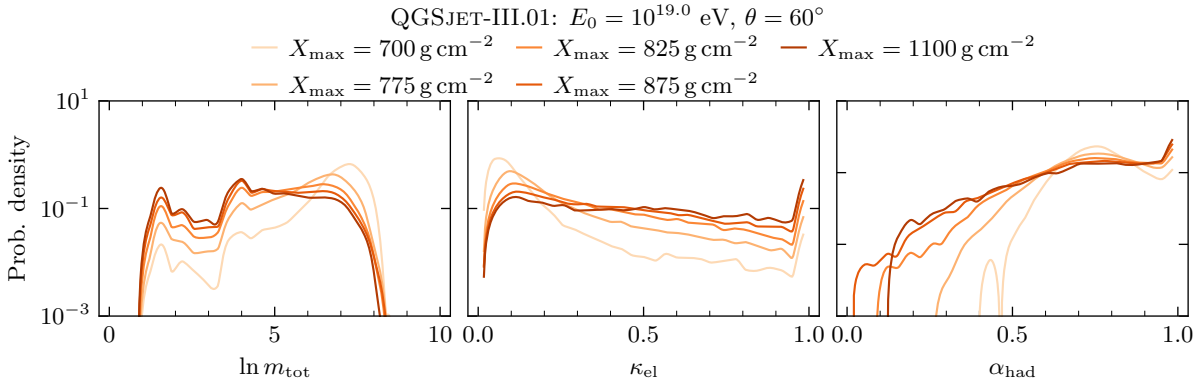

Figure 3.9: Distributions of the total multiplicity, $\ln m_{\mathrm{tot}}$ (left panel), elasticity, κ_{el} (middle panel) and fraction of hadronic energy of the primary p-air interaction (right panel), per bin in $X_{\max}$. This figure was produced with libraries of 10^5 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$.

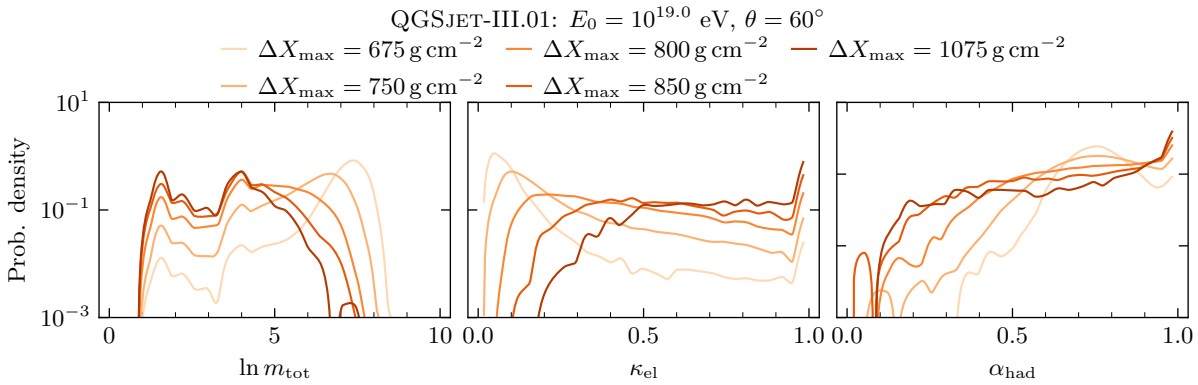

Figure 3.10: Distributions of the total multiplicity, $\ln m_{\mathrm{tot}}$ (left panel), elasticity, κ_{el} (middle panel) and fraction of hadronic energy of the primary p-air interaction (right panel), per bin in $\Delta X_{\max}$. This figure was produced with libraries of 10^5 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$.

corresponds to primary interactions in which a larger fraction of energy is transferred to neutral pions, while the latter reflects highly elastic interactions where the leading proton retains nearly the entire primary energy.

These trends are qualitatively independent of the hadronic interaction model, as illustrated in Figures 3.11 and 3.12, where $\langle \ln m_{\mathrm{tot}} \rangle$ and $\langle \kappa_{\mathrm{el}} \rangle$ are shown for each bin in $X_{\max}$ and $\Delta X_{\max}$, respectively. Furthermore, the separation between primary interactions with high and low hadronic activity becomes more pronounced when binning in $\Delta X_{\max}$ rather than in $X_{\max}$. This is due to fluctuations in X_1, which can result in primary interactions with appreciably different levels of hadronic activity falling within the same $X_{\max}$ bin. Nevertheless, such fluctuations do not preclude probing the hadronic activity of the primary interaction through measurements of $X_{\max}$.

The degree of hadronic activity in primary interactions can also be characterised by identifying the regions of kinematic phase space where most of the secondary particles responsible for driving the cascade are produced. These secondary hadrons typically have small transverse momentum, $p_\perp$, relative to their total momentum, $|\mathbf{p}|$. To determine the regions of phase space most relevant to shower development in each bin of $\Delta X_{\max}$ and

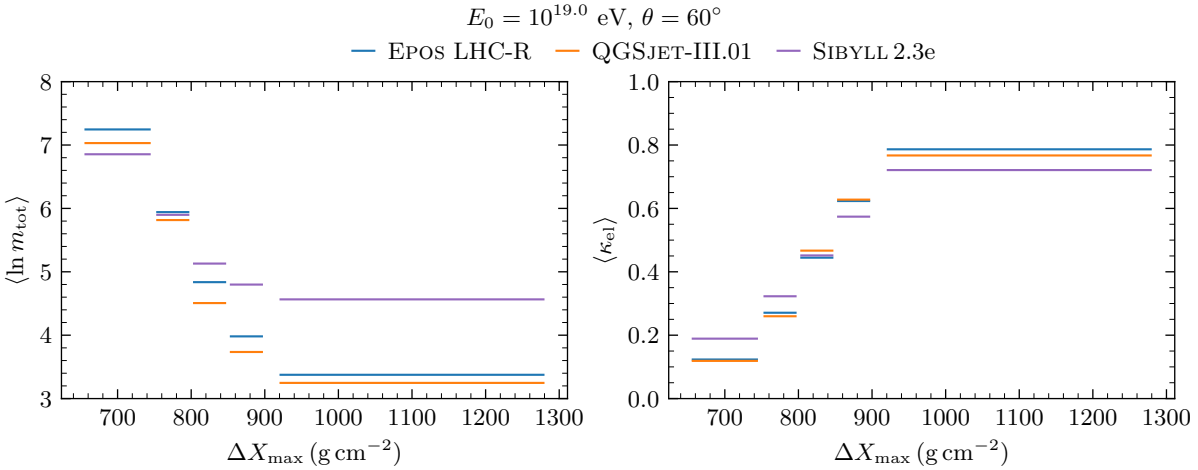

Figure 3.11: $\langle \ln m_{\rm tot} \rangle$ (left panel) and $\langle \kappa_{\rm el} \rangle$ (right panel) of the primary p-air interaction per bin in $\Delta X_{\max}$, for the hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e. This figure was produced with libraries of 10^5 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$.

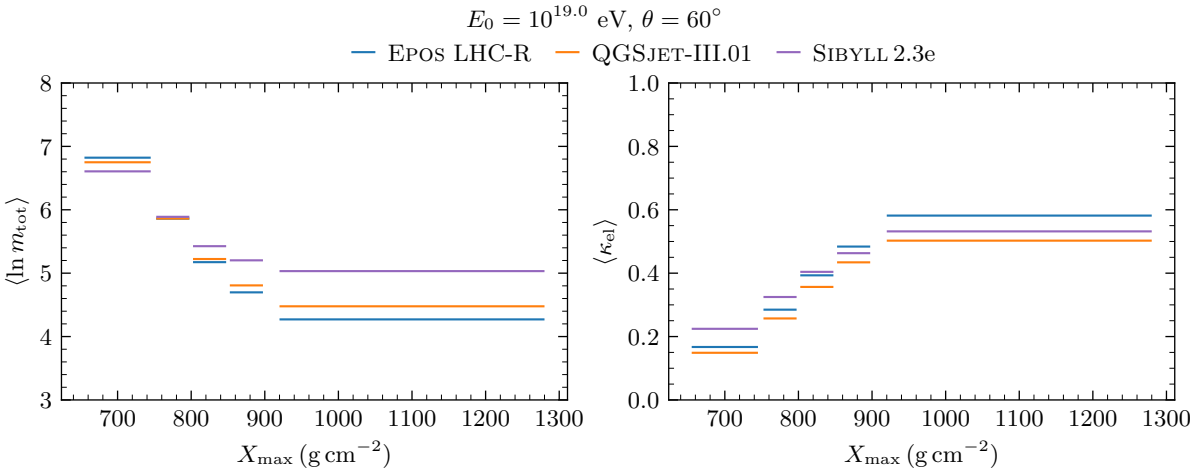

Figure 3.12: $\langle \ln m_{\rm tot} \rangle$ (left panel) and $\langle \kappa_{\rm el} \rangle$ (right panel) of the primary p-air interaction per bin in $X_{\max}$, for the hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e. This figure was produced with libraries of 10^5 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$.

$X_{\max}$, we compute the differential energy flow of the primary interaction, $\mathrm{d}E/\mathrm{d}\eta$, where $\eta = \frac{1}{2} \ln \left(\frac{|\mathbf{p}|+p_\parallel}{|\mathbf{p}|-p_\parallel} \right)$ is the pseudorapidity and $p_\parallel$ denotes the longitudinal momentum. In practice, the energy flow is obtained by weighting the pseudorapidity distribution of each primary interaction by the fraction of energy carried by each secondary hadron.

The differential energy flow for the hadronic and electromagnetic components, binned in $\Delta X_{\max}$ and $X_{\max}$, is shown in Figures 3.13 and 3.14, respectively.

In shallow showers, most of the primary energy is carried by secondary particles produced with lower pseudorapidities in the primary interaction. As progressively deeper $X_{\max}$ bins are considered, the bulk of the energy flow shifts toward higher values of η, indicating that the secondary particles are increasingly produced in the far-forward region.

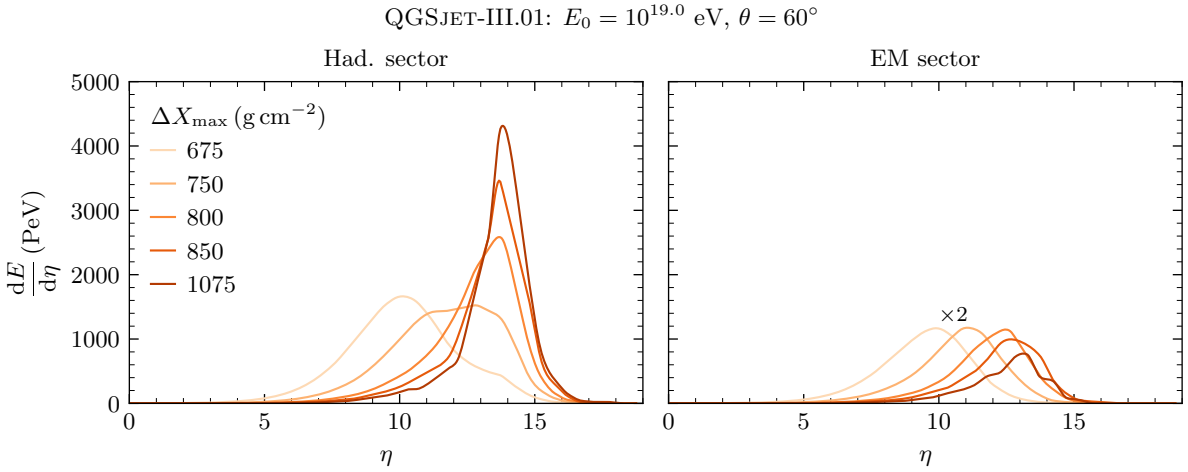

Figure 3.13: Differential energy flows into the hadronic (left panel) and electromagnetic (right panel) sectors per bin in $\Delta X_{\max}$. The energy flow into the electromagnetic sector is multiplied by 2 for visualisation purposes. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$, using QGSJET-III.01.

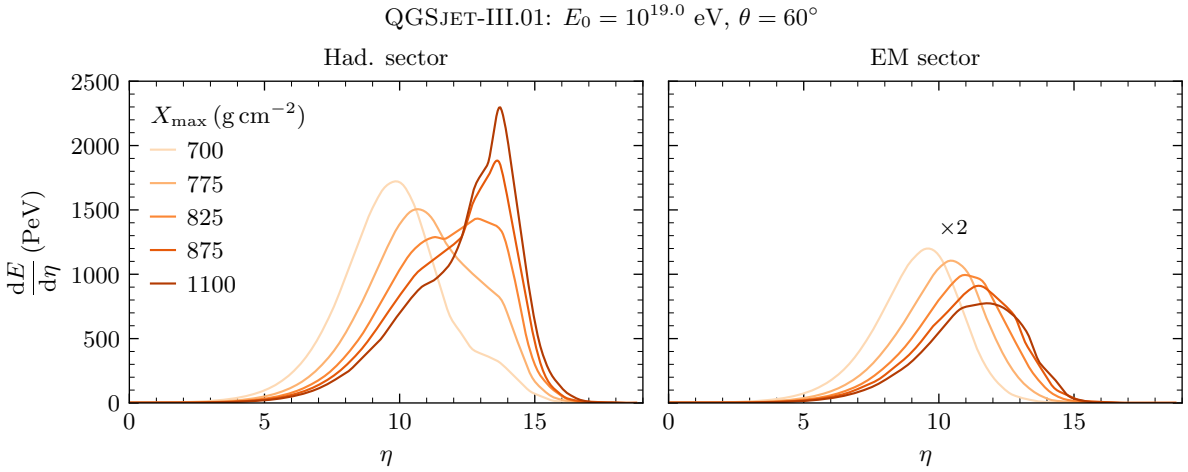

Figure 3.14: Differential energy flows into the hadronic (left panel) and electromagnetic (right panel) sectors per bin in $X_{\max}$. The energy flow into the electromagnetic sector is multiplied by 2 for visualisation purposes. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$, using QGSJET-III.01.

This behaviour is consistent with the macroscopic picture in which deeper showers are associated with more elastic primary interactions. Furthermore, the dispersion of the energy flow tends to decrease with increasing $\Delta X_{\max}$, likely due to the lower multiplicity of the corresponding primary interactions. Once again, binning in $\Delta X_{\max}$ provides a clearer separation between different primary interaction regimes.

These trends are qualitatively independent of the hadronic interaction model, as demonstrated by the evolution of the median of the differential energy flow, η_{m}, as a function of $X_{\max}$ and $\Delta X_{\max}$, shown in Figure 3.15. The median was chosen over the peak or the mean of the energy flow due to the irregular shape of the distributions, particularly in the deeper $X_{\max}$ bins, where the number of produced secondaries is typically smaller.

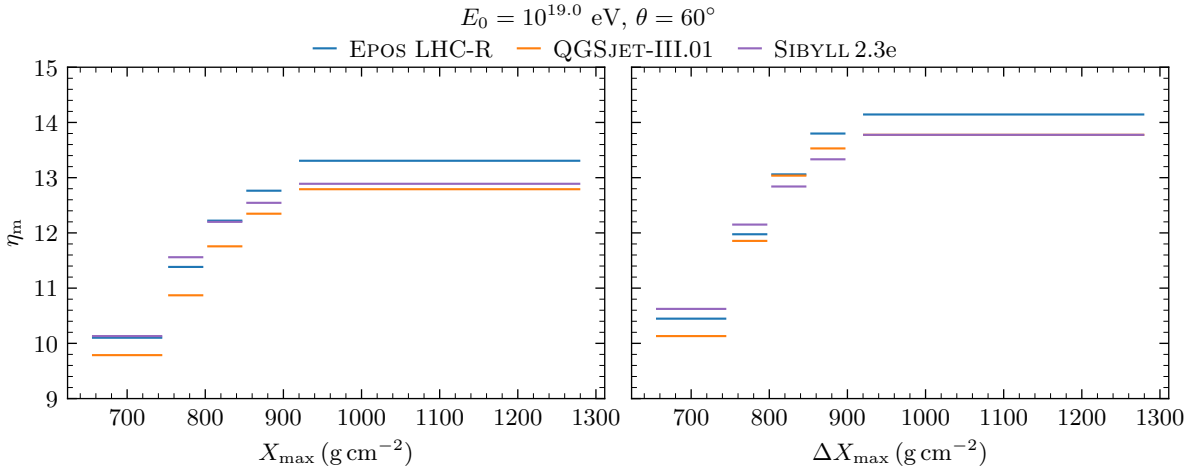

Figure 3.15: Evolution of the median of the energy flow η_m with $X_{\max}$ (left panel) and $\Delta X_{\max}$ (right panel), for EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$.

Therefore, selecting showers according to $X_{\max}$ allows for a continuous probing of the hadronic activity in the primary p–air interaction. More specifically, scanning in $X_{\max}$ corresponds to probing different regions of the kinematic phase space of the proton-air collision. This connection highlights that the evolution of Λ_μ with $X_{\max}$ is driven by a progressive reduction in the hadronic activity of the primary interaction.

3.2.2 The spectrum of neutral pions as a function of $X_{\max}$

As discussed in [106, 133], Λ_μ is sensitive to the hardness of the energy spectrum of neutral pions, denoted by $f(x_{\pi^0}) \propto \mathrm{d}m_{\pi^0}/\mathrm{d}x_{\pi^0}$, where m_{π^0} is the number of neutral pions produced in the primary interaction and x_{π^0} is the fraction of the primary energy they carry in the laboratory frame. Given the observed evolution of Λ_μ with $X_{\max}$, we characterise the shape of the neutral pion energy spectrum as a function of this shower observable, $f(x_{\pi^0} \,|\, X_{\max})$.

The weighted distribution $x_{\pi^0} f(x_{\pi^0} \,|\, X_{\max})$, which emphasises the tails of the spectrum, is shown in the left panel of Figure 3.16. The figure also displays χ^2 fits to the distributions using the functional form

$$f(x_{\pi^0} \,|\, X_{\max}) \propto (1 - x_{\pi^0})^{b(X_{\max})} x_{\pi^0}^{-1-a(X_{\max})}, \tag{3.2}$$

where $a(X_{\max})$ and $b(X_{\max})$ are positive, $X_{\max}$-dependent free parameters. Spectra with smaller values of b correspond to harder distributions.

The right panel of Figure 3.16 shows the analogous distribution, $x_{\pi^0} f(x_{\pi^0} \,|\, \Delta X_{\max})$, for the hadronic interaction model QGSJET-III.01.

Deep showers are associated with a harder energy spectrum of neutral pions produced in the primary interaction. These more energetic neutral pions promptly decay into high-energy photons, initiating electromagnetic sub-showers that tend to develop deeper in the atmosphere. The transition from a softer to a harder neutral pion spectrum becomes even more evident when binning showers in $\Delta X_{\max}$ rather than $X_{\max}$.

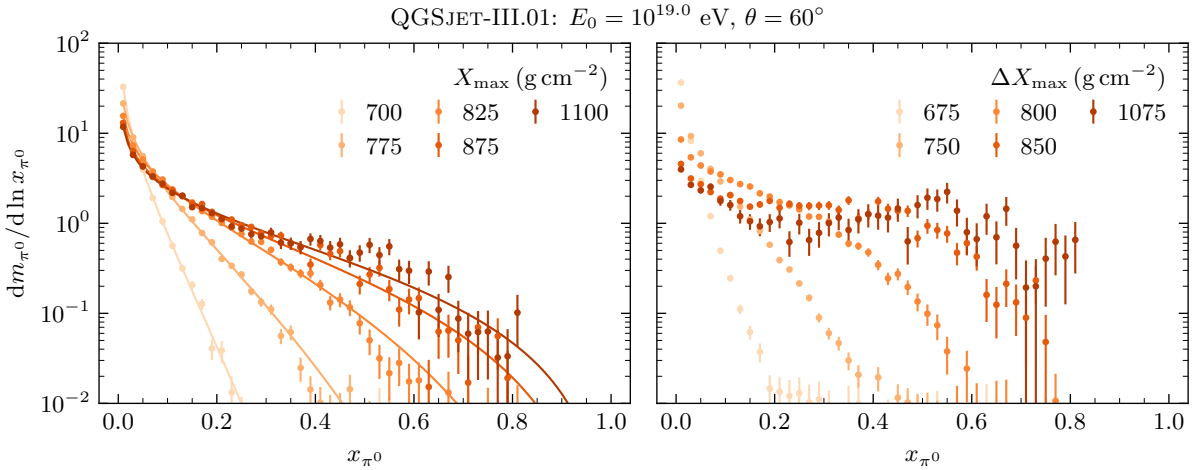

Figure 3.16: Energy weighted energy spectra of neutral pions of the primary p-air interaction per bin in $X_{\max}$ (left), $x_{\pi^0} f(x_{\pi^0} \mid X_{\max})$, and $\Delta X_{\max}$ (right), $x_{\pi^0} f(x_{\pi^0} \mid \Delta X_{\max})$.

This spectral hardening is directly linked to the hadronic activity of the primary interaction. Deeply inelastic interactions produce a large number of secondaries, reducing the available phase space per particle and leading to softer energy spectra. In contrast, harder energy spectra of neutral pions observed in deep showers result from lower multiplicity interactions, and contribute to a flattening of the tail in the $\ln N_\mu$ distribution.

To quantify the evolution of the hardness of the energy spectrum of neutral pions, we extract the parameter b from Equation (3.2) as a function of $X_{\max}$. This is shown in Figure 3.17, for the hadronic interaction models EPOS LHC, QGSJET-II.04, and SIBYLL2.3d (left panel), and their most-up-to-date versions EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e (right panel). The lower panels display the relative model spread in the fitted value of b, defined as $b_{\max}/b_{\min} - 1$, where $b_{\max}$ and $b_{\min}$ denote, respectively, the maximum and minimum values of b across the different hadronic interaction models.

The hardening of the neutral pion energy spectrum with increasing $X_{\max}$ is observed for all hadronic interaction models. Regardless of the specific version considered, the model dependence of the fit parameter b increases with $X_{\max}$, mirroring the behaviour observed for Λ_μ. Nevertheless, the spread in the predictions of b is reduced in the most recent versions of the models compared to earlier ones. This improvement can be attributed, at least in part, to the tuning of model parameters using forward particle production data from LHCf [134]. It is also likely that other mechanisms introduced or modified in the updated versions of the hadronic interaction models contribute to changes in the energy spectrum of neutral pions. A detailed discussion of these mechanisms is beyond the scope of the present work.

Importantly, we infer that the production spectrum of neutral pions can be constrained as a function of the hadronic activity in the primary p-air interaction, by measuring the evolution of Λ_μ with $X_{\max}$.

Finally, we suggest an explanation for the increase in the model dependence of Λ_μ with $X_{\max}$. As discussed, shallower showers are typically initiated by highly inelastic primary interactions, where secondary particles are produced with relatively lower rapidities and energies. This region of the kinematic phase-space is better constrained by accelerator data, leading to smaller differences in the predictions of Λ_μ across different models. In

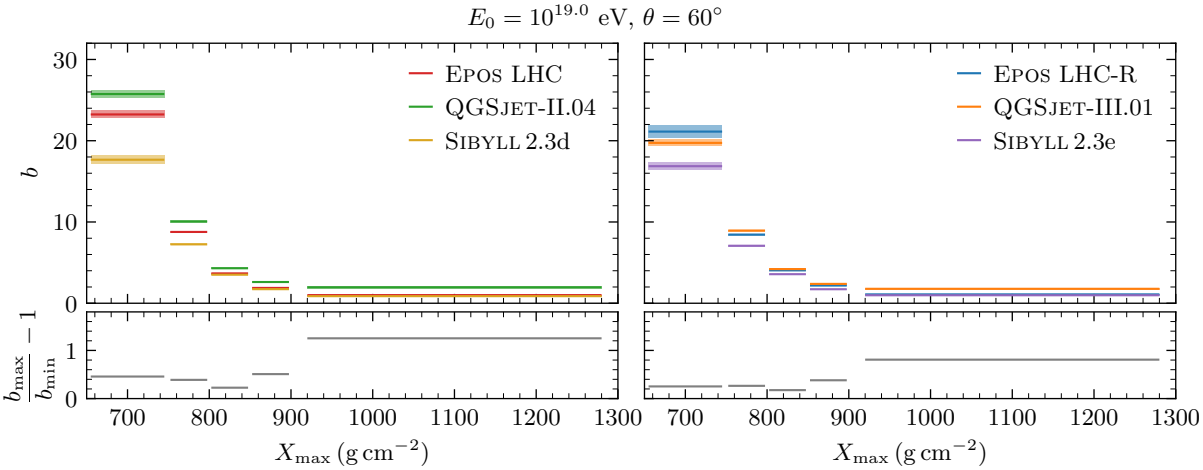

Figure 3.17: Upper panels: Hardness parameter of the energy spectrum of neutral pions of the primary interaction, b, as a function of $X_{\max}$. Lower panels: relative spread of the model predictions in b. The left panel refers to the hadronic interaction models EPOS LHC, QGSJET-II.04 and SIBYLL2.3d and the right panel to their most up-to-date versions EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e (right). This figure was produced with libraries of 10^5 CONEX simulations of proton-induced showers with primary energy $E_0 = 10^{19}$ eV and $\theta = 60°$.

contrast, deeper showers arise from more elastic interactions with reduced hadronic activity: fewer secondaries are produced, and the primary energy is more unevenly distributed among them. The leading particles tend to populate the far-forward kinematic region, which is poorly constrained by accelerator experiments. As a result, model predictions become increasingly dependent on their internal assumptions, leading to a growing model spread in Λ_μ. This interpretation is supported by the reduced model dependence observed in the newer versions of the hadronic interaction models, which have been tuned to data from the forward region of phase-space, such as those collected by LHCf [134, 135].

3.3 Experimental feasibility of the measurement of Λ_μ as a function of $X_{\max}$

The ability to discern between the values of Λ_μ predicted by the different hadronic interaction models, and therefore constrain the energy spectrum of neutral pions of the primary interaction as a function of its hadronic activity, depends on at least three aspects:

1. the composition of the flux of cosmic rays (and its evolution with the primary energy) since proton-induced showers yield the muon-depleted showers that populate the exponential tail of the distribution of $\ln N_\mu$;

2. the resolution in the reconstruction of the shower observables: $\ln N_\mu$, E_0 and $X_{\max}$. In particular, if $X_{\max}$ and N_μ are reconstructed separately, the resolutions on their reconstructions partially de-correlate these observables, widening the distribution of $\ln N_\mu$ and causing migrations between the bins in $X_{\max}$;

3. the number of events in the full ensemble of measured showers considered in the

analysis. In particular, fewer events lead to larger fit uncertainties, hampering the constraining power of a measurement of Λ_μ.

In this chapter, we do not consider the systematic uncertainties in the determination of $X_{\max}$ and N_μ (or their proxies), nor those affecting other shower observables that influence the estimation of $X_{\max}$ and N_μ, such as E_0. This is done to keep the discussion general, without tying it to specific reconstruction algorithms or air-shower detectors.

In this work, the feasibility of the measurement of Λ_μ in each bin of $X_{\max}$ was quantified for two different composition scenarios. For primary energies in $[10^{18}, 10^{18.5}]$ eV, these scenarios correspond to the ones with the lowest and highest proton fractions given by the 4-mass fraction fit to Auger data presented in [124]. Each scenario is mimicked by sampling from 10^5 showers induced by proton, helium, nitrogen and iron primaries in proportions p : He : N : Fe $= 7 : 1 : 2 : 0$ and $1 : 3 : 1 : 0$. Furthermore, we considered Gaussian resolutions in the reconstructions of $\ln N_\mu$ and $X_{\max}$, with $\sigma_{\rm rec}(N_\mu)/N_\mu = 20\%$ [94] and $\sigma_{\rm rec}(X_{\max}) = 20\,{\rm g\,cm^{-2}}$ [76], respectively. That is, for each event, the reconstructed number of muons, $N_\mu^{\rm rec}$, is sampled from a normal distribution of the form

$$G(N_\mu^{\rm rec}) = \frac{1}{\sqrt{2\pi}\sigma_{\rm rec}} \exp\left\{-\frac{(N_\mu^{\rm rec} - N_\mu)^2}{2\sigma_{\rm rec}^2}\right\}, \quad \text{with} \quad \sigma_{\rm rec} = \sigma_{\rm rec}(N_\mu)/N_\mu. \tag{3.3}$$

The values of $X_{\max}^{\rm rec}$ for each event are obtained from a similar procedure using $\sigma_{\rm rec} = \sigma_{\rm rec}(X_{\max})$.

In reality, the interpreted composition of the cosmic ray flux depends on the hadronic interaction model and evolves with the primary energy [124]. Moreover, the falling cosmic ray spectrum [17, 26] and the energy dependence of its composition could mean that the best energy bin to perform the analysis results from a compromise between maximising the fraction of proton-induced showers and the number of events. For the sake of simplicity, we study extreme composition cases in a small bin of primary energies. Furthermore, we perform our study with showers having $E_0 = 10^{19}$ and $\theta = 60°$, even though the mimicked composition scenarios are more realistic at lower primary energies.

The left panel of Figure 3.18 displays the distribution of $\ln N_\mu$ in each bin in $X_{\max}$ using the composition scenario $1 : 3 : 1 : 0$ and the true values of N_μ and $X_{\max}$. The right panel of the same figure shows the same joint distribution but using the reconstructed values of N_μ and $X_{\max}$ using the aforementioned resolutions. The tails of each distribution $f(\ln N_\mu \,|\, X_{\max})$ were fitted as described in Section 3.1. Despite the smearing introduced by the resolution in $f(\ln N_\mu \,|\, X_{\max})$, the tails due to proton primaries remain visible and evolve with $X_{\max}$ in agreement with the results presented in Section 3.1.

We now quantify the bias in Λ_μ introduced by the presence of heavier primaries in the shower samples and by the addition of detector resolutions in the reconstruction of the shower observables. We treat these two effects separately. That is, for the composition scenario $1 : 3 : 1 : 0$, we fit the tail of $f(\ln N_\mu \,|\, X_{\max})$ to extract the reconstructed value of Λ_μ, $\Lambda_\mu^{\rm rec}$, for two cases: using the Monte Carlo values of N_μ and $X_{\max}$, and using the reconstructed values with $\sigma_{\rm rec}(N_\mu)/N_\mu = 20\%$ and $\sigma_{\rm rec}(X_{\max}) = 20\,{\rm g\,cm^{-2}}$. The distributions $f(\ln N_\mu \,|\, X_{\max})$ obtained for the two cases are shown in Figure 3.19, together with the same distribution constructed from the proton-induced showers in the sample of showers with a mixed primary composition. Each panel is identified by the centre of the corresponding bin in $X_{\max}$ and contains the fraction of proton-induced showers, f_p. Note

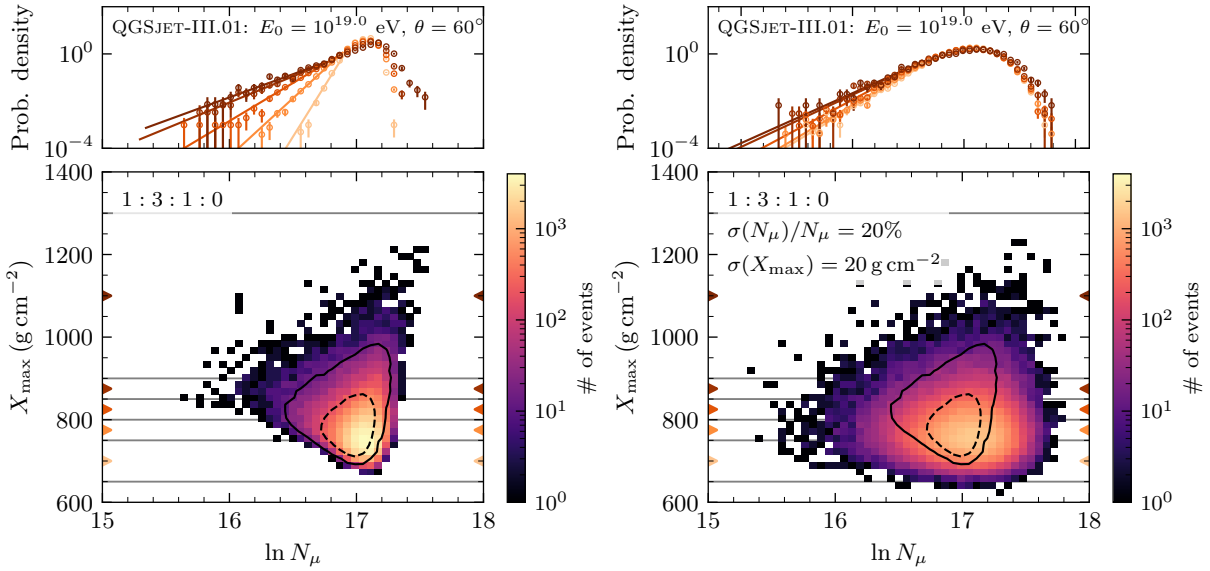

Figure 3.18: Lower panels: Joint distributions of $\ln N_\mu$ and $X_{\max}$: $f(\ln N_\mu, X_{\max})$ for the composition p : He : N : Fe = 1 : 3 : 1 : 0, obtained either directly from simulations (left) or reconstructed (right) by mimicking a Gaussian resolutions $\sigma_{\mathrm{rec}}(N_\mu)/N_\mu = 20\%$ and $\sigma_{\mathrm{rec}}(X_{\max}) = 20\,\mathrm{g\,cm^{-2}}$ (right). The black contours pertain to the true $f(\ln N_\mu, X_{\max})$ for proton-induced showers and contain 68% (dashed) and 95% (solid) of the simulations. Upper panels: distribution of $\ln N_\mu$ for each bin in $X_{\max}$, $f(\ln N_\mu \mid X_{\max})$, matching the colour of the corresponding $X_{\max}$-bin. This figure was produced by sampling from libraries of 10^5 CONEX-showers induced by primaries in different proportions, using the high-energy hadronic interaction model QGSJET-III.01.

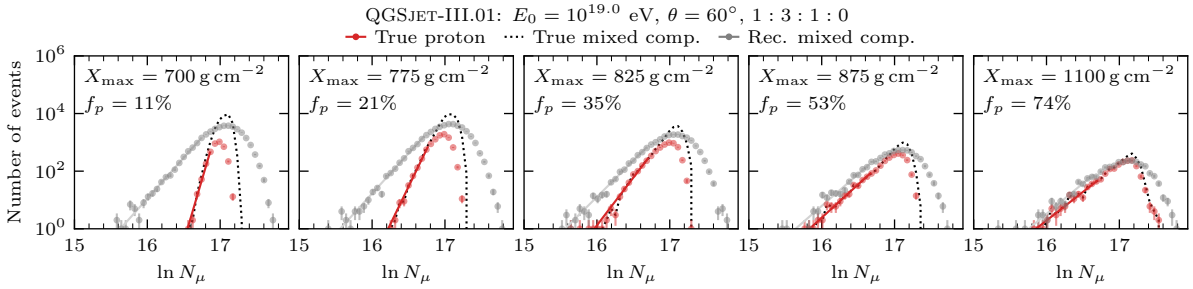

Figure 3.19: The Monte-Carlo (dotted grey line) and reconstructed (grey markers) distributions of $\ln N_\mu$ for a mixed composition of primaries with relative abundances p : He : N : Fe = 1 : 3 : 1 : 0. The reconstructed distribution is obtained by smearing the Monte Carlo values of N_μ and $X_{\max}$ with Gaussian resolutions of $\sigma_{\mathrm{rec}}(N_\mu)/N_\mu = 20\%$ and $\sigma_{\mathrm{rec}}(X_{\max}) = 20\,\mathrm{g\,cm^{-2}}$. The red markers depict the Monte-Carlo distribution of $\ln N_\mu$ proton-induced showers within the mixed sample. Each panel corresponds to a different bin in $X_{\max}$, increasing from left to right. The centre of each $X_{\max}$ bin and the corresponding fraction of proton-induced events, f_p, are indicated in the panels. The figure was produced using libraries of 10^5 CONEX showers induced by different primaries, employing the high-energy hadronic interaction model QGSJET-III.01.

that for this composition scenario, the fraction of proton showers should be close to 20%.

When using the Monte-Carlo values of $X_{\max}$ and N_μ, the left tail of the distribution $f(\ln N_\mu \,|\, X_{\max})$ obtained for a mixed primary composition visually coincides with that obtained from proton primaries alone. Moreover, the resolution in the reconstruction of

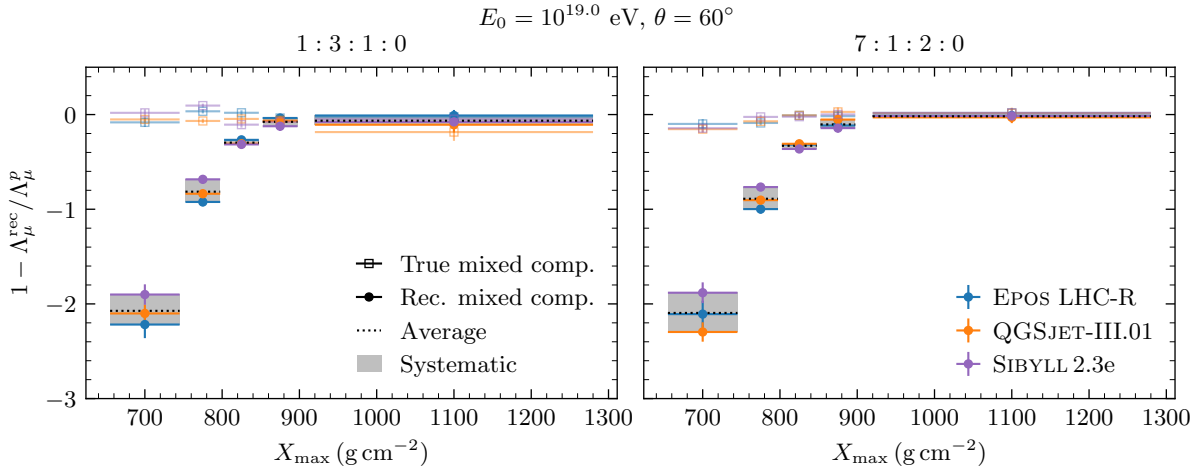

Figure 3.20: Bias in the reconstructed value of Λ_μ relative to the value obtained for proton-induced showers, $1 - \Lambda_\mu^{\rm rec}/\Lambda_\mu^p$, using the true (square empty markers) and reconstructed (solid circles) values of N_μ and $X_{\max}$. The average bias in the reconstruction is depicted by the dotted black line, while the bias spread between hadronic interaction models is depicted by the shaded grey band. The left panel refers to the primary mixing $1:3:1:0$ and the right panel to $7:1:2:0$. The figure was produced using libraries of 10^5 CONEX showers induced by different primaries, employing the high-energy hadronic interaction model QGSJET-III.01.

N_μ and $X_{\max}$ widens and flattens the left tail of $f(\ln N_\mu \,|\, X_{\max})$ especially for shallow showers, for which the true distribution is narrow compared with the mimicked detector Gaussian response. Additionally, and as expected, the fraction of proton showers increases with the depth of the considered bin in $X_{\max}$, as deep showers are dominantly induced by these primaries.

For clarity, we define $\Lambda_\mu^{\rm rec}$ as the value of Λ_μ extracted from the PDF $f(\ln N_\mu \,|\, X_{\max})$ pertaining to a mixed primary composition. We also define Λ_μ^p as the corresponding value obtained for proton-induced showers using the Monte Carlo values of the shower observables. The values of Λ_μ^p are the ones shown in Figure 3.2. Using the introduced notation, we determine, for each bin in $X_{\max}$, the relative bias, $1 - \Lambda_\mu^{\rm rec}/\Lambda_\mu^p$ obtained either directly using the Monte Carlo values of N_μ and $X_{\max}$ or the reconstructed ones using the mentioned resolutions. Both cases are shown in Figure 3.20 for the compositions $1:3:1:0$ (left) and $7:1:2:0$ (right), for the hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e. The figure also displays the model-averaged bias and the spread in bias across the considered hadronic interaction models. This spread is taken as a systematic uncertainty.

The bias induced by a mixture of primary masses on Λ_μ is less than 10%, irrespective of the considered bin in $X_{\max}$. This small bias can be explained by the superposition principle: a primary with a mass number A interacts as A independent nucleons with energy E_0/A, so that the resulting distribution of N_μ is approximately the A-fold self-convolution of the one present for proton showers. As verified in Section 2.2.3, multiple convolutions make $f(\ln N_\mu \,|\, X_{\max})$ increasingly Gaussian, reducing the low-N_μ tail associated with proton-induced muon-depleted showers. On the other hand, the convolution of the detector resolution with the distribution of $\ln N_\mu$ greatly biases Λ_μ in shallower showers, as noted in the analysis of Figure 3.19. Furthermore, for shallow showers, the

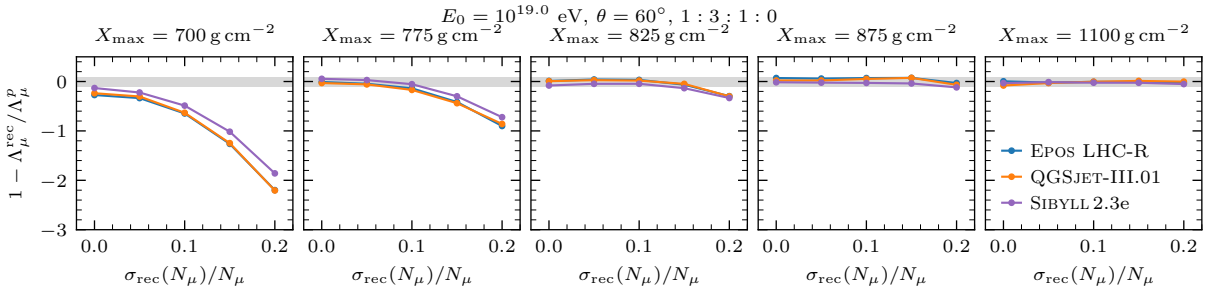

Figure 3.21: Bias in $\Lambda_\mu^{\mathrm{rec}}$ relative to Λ_μ^p as a function of the resolution in the reconstruction of N_μ, $\sigma_{\mathrm{rec}}(N_\mu)/N_\mu$, using $\sigma_{\mathrm{rec}}(X_{\max}) = 20\,\mathrm{g\,cm^{-2}}$. The shaded grey band corresponds to $\pm 10\%$ relative bias. The figure was produced using libraries of 10^5 Conex showers induced by different primaries, in proportions 1 : 3 : 1 : 0 employing the high-energy hadronic interaction models Epos LHC-R, QGSJet-III.01 and Sibyll2.3d.

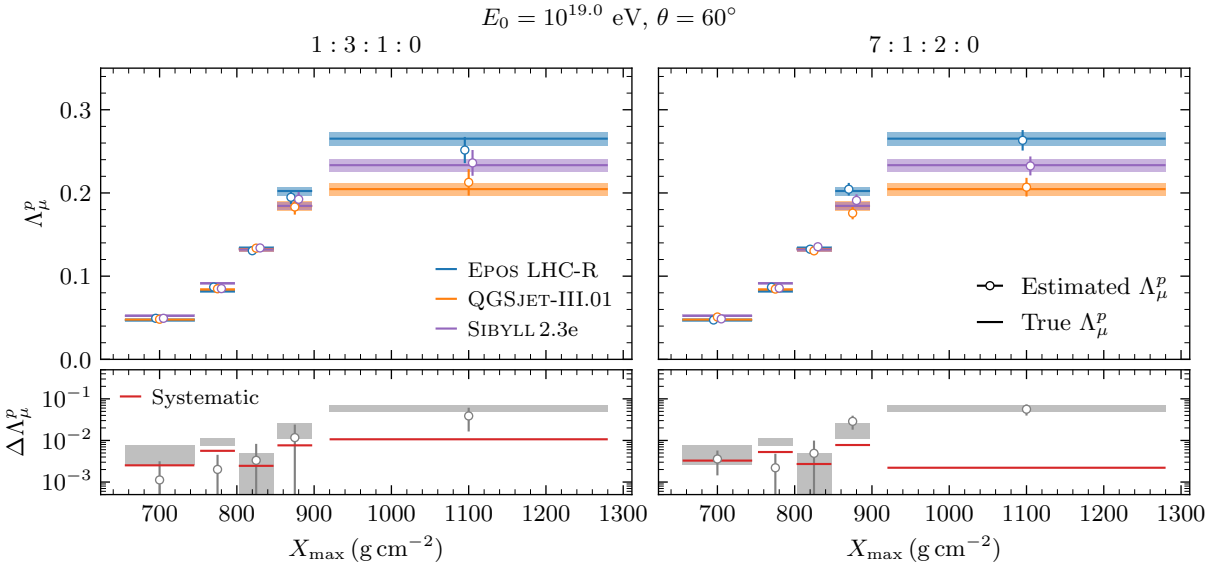

Figure 3.22: Upper panels: True (step function) and estimated (circular markers) values of Λ_μ^p, as a function of $X_{\max}$, obtained by applying an average bias correction to $\Lambda_\mu^{\mathrm{rec}}$. The latter parameter is extracted from an ensemble of showers induced by p, He, N and Fe in proportions 1 : 3 : 1 : 0 (left) and 7 : 1 : 1 : 0 (right) with N_μ and $X_{\max}$ reconstructed with resolutions $\sigma_{\mathrm{rec}}(N_\mu)/N_\mu = 20\%$ and $\sigma_{\mathrm{rec}}(X_{\max}) = 20\,\mathrm{g\,cm^{-2}}$, respectively. Lower panels: spread in true and estimated the values of Λ_μ^p, $\Delta\Lambda_\mu^p$, across hadronic interaction models. The step function depicted in red represents the systematic uncertainty of the average bias correction. The figure was produced using libraries of 10^5 Conex showers induced by different primaries, employing the high-energy hadronic interaction models Epos LHC-R, QGSJet-III.01 and Sibyll2.3e.

bias in Λ_μ is model-dependent, while for deeper ones, it becomes model-independent. In deep showers, the bias in Λ_μ is approximately 10%, consistent with the mixed-composition result, without introducing a reconstruction resolution. Note that these observations are qualitatively independent of the considered mixed composition scenario.

Figure 3.21 shows the increase of relative bias in Λ_μ with worsening resolution in the reconstruction of N_μ. The figure indicates that only the shallowest $X_{\max}$-bins are affected and that the hadronic-model dependence of the bias in Λ_μ increases with the resolution.

Detailed strategies to deconvolve the detector resolution from the physical fluctuations

of N_μ could mitigate the bias in Λ_μ for shallower showers. However, since such strategies are highly dependent on the specific response function of the detector, we apply, for the sake of simplicity, an average bias correction to Λ_μ^{rec} to recover an estimated value of Λ_μ^p. Furthermore, we take the model dependence of the relative bias in Λ_μ^p as a systematic uncertainty. This uncertainty in Λ_μ^p is greater than the difference between hadronic interaction models for the two shallowest bins in $X_{\max}$ that we considered. That is, for showers with $X_{\max} < 800\,\text{g}\,\text{cm}^{-2}$, thus preventing the discrimination between hadronic interaction model predictions. Nevertheless, a measurement of Λ_μ for shallow showers could still allow the simultaneous exclusion of all models.

Figure 3.22 depicts the true and estimated values of Λ_μ^p, with the latter determined from the average bias correction depicted as the black dotted steps in Figure 3.20. This average correction is applied to Λ_μ^{rec} for the three prior distributions $f(\ln N_\mu \,|\, X_{\max})$ provided by the different hadronic interaction models for the mixed composition scenarios $1:3:1:0$ (left panel) and $7:1:2:0$ (right panel). In general, the estimated values of Λ_μ^p are compatible with the true ones, although the model spread tends to be smaller in particular bins in $X_{\max}$.

Since the capacity to distinguish between model predictions of $\Lambda_\mu^p(X_{\max})$, and therefore add meaningful constraints on the neutral pion spectrum of the primary interaction, depends on the number of proton events in the sample, we determined the minimum number of events, $n_{\min}$, necessary to distinguish between the estimated values of Λ_μ^p obtained from different hadronic interaction models. The distinction threshold was set to 1σ and $3\,\sigma$, corresponding to $n_{\min}^{1\sigma}$ and $n_{\min}^{3\sigma}$ events, respectively. The error in Λ_μ^p was computed via a bootstrapping method to ensure its accurate estimation for low statistics. The upper panels of Figure 3.23 show the estimated values of Λ_μ^p as a function of the number of events in the full ensemble of EAS, N, for each bin of $X_{\max}$, using an injected primary composition with fractions $1:3:1:0$ and the reconstructed values of $X_{\max}$ and N_μ using the resolutions indicated in the figure. The lower panels show the relative error in the determination of Λ_μ^p, $\sigma(\Lambda_\mu^p)/\Lambda_\mu^p$, and a horizontal line indicating the threshold for model distinction. That is, for a spread between models $\Delta\Lambda_\mu^p$, we define the relative spread

$$\delta_\mu = \frac{\Delta\Lambda_\mu^p}{\Lambda_\mu^{p,\min}}, \tag{3.4}$$

where the minimum is taken over the hadronic interaction models. Visually, the minimum number of events for a 1σ distinction corresponds to the value of N where the intersection between $\sigma(\Lambda_\mu^p)/\Lambda_\mu^p$ and δ_μ occurs. The values of $n_{\min}$ for the composition scenarios $1:3:1:0$ and $7:1:2:0$ can be found in Table 3.1.

Since the models are close to being indistinguishable for showers in the bulk of the $X_{\max}$ distribution, the number of events needed for a 3σ model separation is greater than the number of events in the samples. However, for the deepest bin in $X_{\max}$, between 2 000 and 7 000 are sufficient to distinguish between the values of Λ_μ predicted by high-energy hadronic interaction models with a 1σ separation, depending on the considered composition and collection of hadronic interaction models. A 9-fold increase in statistics allows for a 3σ distinction in both cases. Note that these values are intended to represent orders of magnitude, more than an exact estimation, since the acceptance and quality cuts applied to data measured from different experiments might require raw samples with more events than those estimated here. Nevertheless, these numbers of events fall within

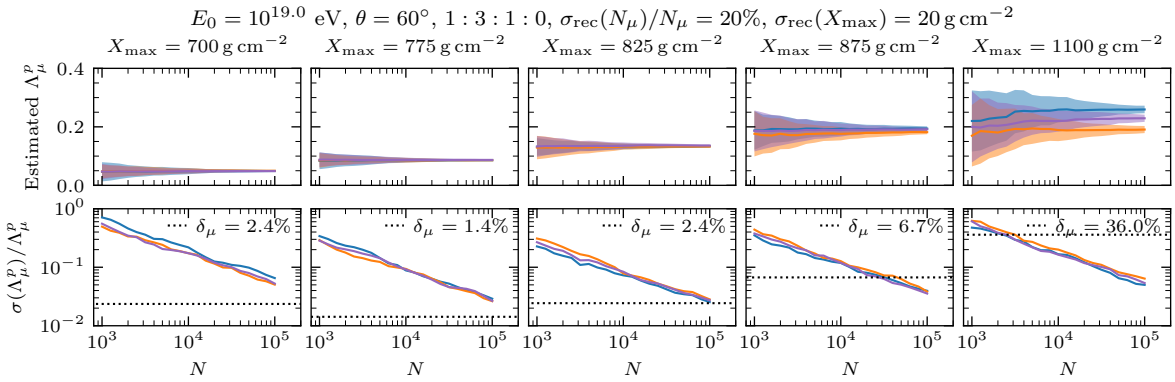

Figure 3.23: Upper panels: Estimated value of Λ_μ^p per bin in $X_{\rm max}$, as a function of the number of EAS simulations, N. The values of $X_{\rm max}$ increase from left to right, and the bin centres are indicated in each panel. The estimation is the result of applying an average bias correction to $\Lambda_\mu^{\rm rec}$, which is extracted from an ensemble of showers induced by p, He, N and Fe in proportions 1 : 3 : 1 : 0 with N_μ and $X_{\rm max}$ smeared with resolutions $\sigma_{\rm rec}(N_\mu)/N_\mu = 20\%$ and $\sigma_{\rm rec} = 20\,{\rm g\,cm^{-2}}$, respectively. Lower panels: relative uncertainty in Λ_μ^p, $\sigma(\Lambda_\mu^p)/\Lambda_\mu^p$, as a function of N, together with the threshold set by the relative difference between hadronic interaction models δ_μ.

	Most recent HIMs				Recently outdated HIMs			
	1 : 3 : 1 : 0		7 : 1 : 2 : 0		1 : 3 : 1 : 0		7 : 1 : 2 : 0	
$X_{\rm max}$ [g cm^{-2}]	$n_{\rm min}^{1\sigma}$	$n_{\rm min}^{3\sigma}$	$n_{\rm min}^{1\sigma}$	$n_{\rm min}^{3\sigma}$	$n_{\rm min}^{1\sigma}$	$n_{\rm min}^{3\sigma}$	$n_{\rm min}^{1\sigma}$	$n_{\rm min}^{3\sigma}$
700	—	—	—	—	—	—	—	—
775	—	—	—	—	—	—	—	—
825	100 000	100 000	100 000	100 000	13 030	100 000	18 478	100 000
875	45 231	100 000	14 229	100 000	5 080	54 393	3 519	29 587
1100	3 171	24 325	6 977	55 377	3 113	25 898	1 877	18 805

Table 3.1: Minimum number of events needed to distinguish between hadronic interaction models for different bins in $X_{\rm max}$ and two realistic mixed composition scenarios. The tag "Most recent HIMs" refers to EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e, and "Recently outdated HIMs" to EPOS LHC, QGSJET-II.04 and SIBYLL2.3d. The bins in $X_{\rm max}$ for which models are indistinguishable due to the model dependence of the bias in Λ_μ are indicated by —.

the current dataset of air showers recorded by the Pierre Auger Observatory [17, 26]. For this experiment, the shower-to-shower estimations of N_μ and $X_{\rm max}$ could be provided by neural networks [88, 89, 92], shower universality [97] and/or the improved detectors of AugerPrime [91].

3.4 Summary and Discussion

In this chapter, we demonstrated that the steepness of the distribution of the number of muons in muon-depleted showers, quantified by the parameter Λ_μ, evolves with the depth of the shower maximum, $X_{\rm max}$. This evolution is not a consequence of propagation

effects, but rather arises from the correlation between N_μ and $X_{\max}$ induced by hadron production in the early stages of shower development.

Furthermore, the dependence of Λ_μ on $X_{\max}$ reflects the hardness of the neutral pion energy spectrum resulting from different regions of kinematic phase space in the primary p-air interaction. These phase-space regions are indicative of primary interactions with varying degrees of hadronic activity and are effectively probed by $X_{\max}$. Notably, the value of Λ_μ for shallow showers is largely independent of the high-energy hadronic interaction model. In this regime, the primary interactions are deeply inelastic, producing a high multiplicity of secondaries that evenly distribute the primary energy. In contrast, for deep showers, predictions of Λ_μ exhibit significant model dependence, highlighting the influence of specific physical mechanisms implemented in each model. These effects are especially relevant in more diffractive or elastic interactions, probed in deeper showers.

Finally, we have shown that an unbiased measurement of Λ_μ is attainable with current extensive air shower detectors, given their resolution, exposure, and the interpretation of the mass composition of the cosmic-ray flux. The sensitivity of Λ_μ to the hadronic activity in ultra-high-energy cosmic-ray interactions, as probed through $X_{\max}$, establishes its potential as a tool to characterise hadronic physics at energy scales beyond the reach of current human-made accelerators.

4

PROBABILISTIC MAPPING BETWEEN MULTIPARTICLE PRODUCTION VARIABLES AND X_{max} IN PROTON-INDUCED EAS

In this chapter, we introduce a set of new multiparticle production variables—α_{had}, ζ_{had}, and ζ_{EM}—constructed from the energy spectra of secondary particles produced in the primary interaction of proton-induced extensive air showers. By extending the Heitler-Matthews model, we demonstrate that a specific linear combination of these variables, which we denote by ξ, accounts for more than 50% of the fluctuations observed in ΔX_{max}. We also quantify the amount of information about ΔX_{max} that originates from the primary interaction and is effectively captured by ξ. The causal link between ξ and ΔX_{max} is further examined as a function of the primary energy.

Crucially, we construct a probabilistic mapping based on this causal connection, which enables model-independent predictions of X_{max} moments with biases below $3\,\mathrm{g\,cm^{-2}}$, using the most up-to-date hadronic interaction models. This result demonstrates that the measured distribution of X_{max} can serve as a data-driven probe of the secondary hadron spectra from cosmic–ray–air interactions in proton-induced showers.

We also investigate the distributions of the newly introduced production variables and their correlations, interpreting them in terms of the kinematic properties of the primary interaction. In particular, the event-by-event values of ζ_{had} and ζ_{EM} are shown to reflect the degree of *hadronic activity* (as defined in Chapter 3) in the primary interaction, while their statistical moments relate to the shape of the secondary energy spectra.

Finally, we demonstrate that the distributions of the new production variables can be constrained using existing data from accelerator experiments in the far-forward rapidity region. This establishes a bridge between accelerator measurements and air shower observations, and illustrates the potential of the latter to probe hadronic interactions at energies beyond the reach of human-made colliders, through the distribution of X_{max} and the probabilistic mapping developed in this work.

The results presented in this chapter have been published [136]. Further details are provided in Chapter Published content in peer-reviewed journals. Reprinted excerpts and figures with permission from Lorenzo Cazon, Ruben Conceição, Miguel A. Martins, and Felix Riehn, *Phys. Rev. D - 14 of August, 2025*. Copyright 2025 by the American Physical Society. The license for the published article is in Appendix E.

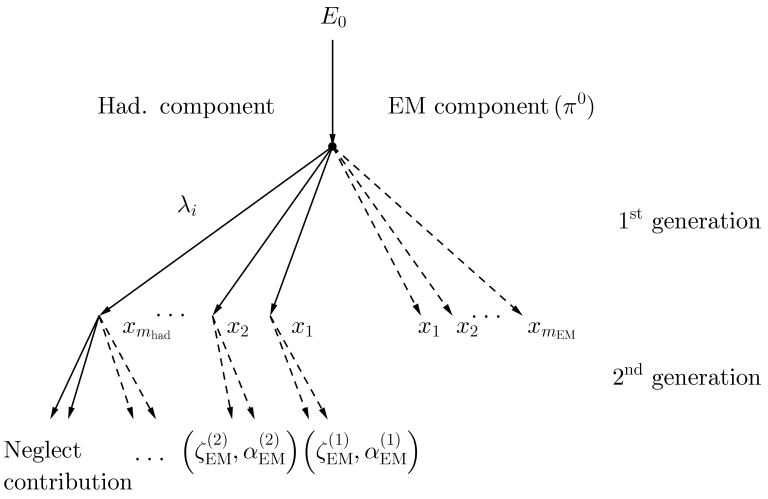

Figure 4.1: Scheme of the primary interaction, and first and second shower generations. The variable x designates the fraction of the primary energy carried by each secondary, in the laboratory frame. The number of hadronically interacting particles is denoted by $m_{\rm had}$, while the number of neutral pions is represented by $m_{\rm EM}$. Each hadronically interacting particle further interacts, after a depth λ_i, producing subsequent EM cascades characterised by the multiparticle production variables $\alpha_{\rm EM}$ and $\zeta_{\rm EM}$, defined in the main text.

4.1 Connecting fluctuations in the primary interaction to fluctuations of $X_{\max}$

In this section, we derive a multiparticle production variable of the primary interaction that causally connects the stochastic nature of hadron production to the shower-to-shower fluctuations of $\Delta X_{\max}$. This is achieved by applying the Heitler–Matthews formalism to each secondary particle produced in the primary proton–air interaction and estimating its contribution to the overall depth of shower maximum. A similar approach was used to study fluctuations in the number of muons, N_μ, in [53, 122], where the shape of the N_μ distribution was connected to the distribution of the variable α_1, as summarised in Section 1.2.5.

4.1.1 Derivation of a new multiparticle production variable of the primary interaction

We begin the derivation by separately considering the electromagnetic (EM) cascades initiated by neutral pions produced directly in the primary interaction (first-generation π^0), and the EM cascades triggered by neutral pions resulting from subsequent hadronic interactions of secondary particles of the first shower generation (second-generation π^0). This schematic distinction is illustrated in Figure 4.1.

We start by considering the EM cascade induced directly by secondary neutral pions of the primary p-air interaction, i.e., first-generation neutral pions. We only take into account the contribution of these light mesons, as they carry approximately 90% of the energy transferred to the EM sector [41] in the primary interaction. Each neutral pion, indexed by j, carries an energy $E_j = E_0 x_j$, in the laboratory frame and immediately

decays into two photons. We assume each photon carries half of the energy of the parent π^0, triggering an EM cascade with a depth of shower maximum given by Equation (1.24). This assumption neglects the stochastic fluctuations of the fraction of energy carried by each photon produced by the decay of the parent neutral pion. Moreover, it neglects the fluctuations in the development of the individual photon cascades. The caveats of these approximations are explored in Appendix C.1.

Under these assumptions, the depth of the maximum of the cascade triggered by a single neutral pion is

$$\lambda_r^{-1} X^{\pi^0}_{\max,j} \equiv t^{\pi^0}_{\max,j} = \ln\left(\frac{E_0}{2\xi_c^e}\right) + \ln x_j, \tag{4.1}$$

where we defined the dimensionless depth $t = \lambda_r^{-1} X$, following the convention in [48]. The first term in Equation (4.1) corresponds to the mean depth of the maximum of an EM cascade initiated by a photon with energy $E_0/2$. The second term is always negative and accounts for stochastic fluctuations in the fraction of the primary energy carried by neutral pions. For neutral pions with energy fractions $x_j < 2\xi_c^e/E_0$, the quantity $t^{\pi^0}_{\max,j}$ becomes negative, reaching a minimum value of $\lambda_r \ln(M_{\pi^0}/2\xi_c^e) \sim -10\,\mathrm{g\,cm^{-2}}$, where M_{π^0} is the neutral pion's mass. While these cases are unphysical, Equation (4.1) applies only when the decaying pion has enough energy to trigger an EM cascade. Moreover, the impact of such unphysical terms is negligible when considering contributions from other secondary particles of the primary interaction. This will become clear in the following paragraphs.

The stochastic fluctuations in x_j lead to significant variability in the values of $t^{\pi^0}_{\max,j}$. Each electromagnetic cascade develops independently, and their superposition gives rise to the longitudinal profile of the overall EM component initiated by the first-generation neutral pions. Rather than modelling each individual cascade with a Greisen function [137] and numerically determining the position of the maximum of their summed profile, we adopt a more useful and transparent approach. We estimate the depth of the maximum of the combined EM cascade, $t^{\mathrm{EM}}_{\max}$, as a weighted average of the individual maxima $t^{\pi^0}_{\max,j}$, using the Heitler-predicted normalisation of each profile, $N_{\max,j} = E_0 x_j / 2\xi_c^e$, as weights. This approach has been used in [122, 138]. An example of this procedure is illustrated in Figure 4.2 for a representative proton-induced shower.

In fact, using CONEX simulations of proton-induced showers, as described in Section 2.1, we verified that the weighted average provides an unbiased estimate of the position of the maximum of summed Greisen profiles. When the true values of $N_{\max,j}$, which fluctuate for a fixed photon energy, are used as weights, the estimation of $t^{\mathrm{EM}}_{\max}$ has an uncertainty of $5\,\mathrm{g\,cm^{-2}}$. If, instead, the approximation $N_{\max,j} \propto x_j$ is used, the uncertainty increases to $24\,\mathrm{g\,cm^{-2}}$. Letting m_{EM} denote the number of neutral pions produced in the primary proton-air interaction, we obtain

$$t^{\mathrm{EM}}_{\max} = \frac{\sum_{j=1}^{m_{\mathrm{EM}}} N_{\max,j} t^{\pi^0}_{\max,j}}{\sum_{j=1}^{m_{\mathrm{EM}}} N_{\max,j}} = \ln\left(\frac{E_0}{2\xi_c^e}\right) + \frac{\sum_{j=1}^{m_{\mathrm{EM}}} x_j \ln x_j}{\sum_{j=1}^{m_{\mathrm{EM}}} x_j} = \ln\left(\frac{E_0}{2\xi_c^e}\right) - \frac{\zeta_{\mathrm{EM}}}{\alpha_{\mathrm{EM}}}, \tag{4.2}$$

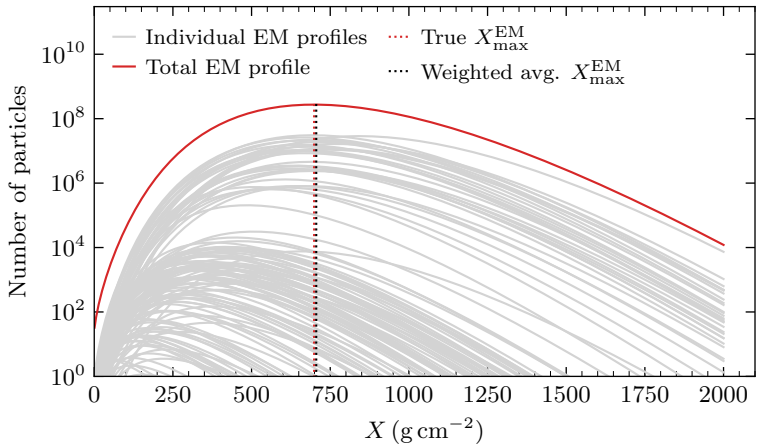

Figure 4.2: Example showing multiple longitudinal profiles of electromagnetic cascades initiated by the decay of neutral pions (in grey), sampled according to their energy spectrum in proton-air interactions. The individual cascades are modelled using Greisen functions, and their combined profile is shown in red. The vertical dotted lines indicate the true and estimated values of the depth of shower maximum, $X_{\max}^{\mathrm{EM}}$, where the estimated value is obtained from the energy-weighted average of the depths of the individual electromagnetic cascades.

where we identified the fraction of energy carried by all neutral pions of the first interaction $\alpha_{\mathrm{EM}} \equiv \sum_{j=1}^{m_{\mathrm{EM}}} x_j$ and defined the multiparticle production variable $\zeta_{\mathrm{EM}} \equiv -\sum_{j=1}^{m_{\mathrm{EM}}} x_j \ln x_j$, with $\zeta_{\mathrm{EM}} > 0$. We discuss the interpretation of ζ_{EM} in terms of its properties in Section 4.4.1. However, we already note that $\lim_{x\to 0} x \ln x = 0$. Hence, the contribution of low-energy neutral pions to $t_{\max}^{\mathrm{EM}}$ is negligible.

We now consider the electromagnetic showers initiated by the decay of neutral pions produced in the hadronic interactions of secondary particles from the primary interaction. These second-generation neutral pions are schematically represented in Figure 4.1. Suppose that each hadronically interacting secondary carries a fraction x_i of the primary energy (in the laboratory frame), such that the total energy retained in the hadronic sector is $\alpha_{\mathrm{had}} = \sum_i x_i$. Let m_{had} denote the total number of such hadrons. As a first approximation, we assume that the contribution of hadron i to $\Delta X_{\max}$ arises solely from the neutral pions produced in its first interaction with an air nucleus. We thus neglect neutral pions produced in subsequent generations, since the resulting photons are typically much less energetic than those from earlier stages of the shower. Let λ_i be the interaction length of hadron i, expressed in units of the radiation length λ_r. Applying Equation (4.2) to estimate the depth of the shower maximum of the electromagnetic cascade initiated by hadron i then yields

$$t_{\max,i}^{\mathrm{had}} = \ln\left(\frac{E_0}{2\xi_c^e}\right) + \lambda_i + \ln x_i - \frac{\zeta_{\mathrm{EM}}^{(i)}}{\alpha_{\mathrm{EM}}^{(i)}}. \tag{4.3}$$

Once again, the first term is the average value of $X_{\max}$ for a photon with energy $E_0/2$, and the others are stochastic corrections to account for the fraction of the energy of the secondary hadron inducing the EM cascade and the actual depth it traverses before interacting.

The combined $X_{\max}$ of the electromagnetic cascades initiated by each hadron is estimated by averaging the individual values $t_{\max,i}^{\mathrm{had}}$, weighted by the number of particles at each

shower maximum, $N_{\max,i}$. Using a set of proton-induced CONEX simulations (described in Section 2.1) for primary energies $E_0 \in [10^{12}, 10^{19}]$ eV, we verified that the energy dependence of $N_{\max}$ is well described by the fit $y = p_1 (E_0/\xi_c^e)^\gamma + p_2$, with $\gamma = 1.004 \pm 0.001$, in agreement with the findings of [51, 119]. This justifies the approximation $N_{\max,i} \approx E_0 x_i$, so that

$$t_{\max}^{\text{had}} = \ln\left(\frac{E_0}{2\xi_c^e}\right) + \frac{1}{\alpha_{\text{had}}}\left[\sum_{i=1}^{m_{\text{had}}} x_i \left(\lambda_i - \frac{\zeta_{\text{EM}}^{(i)}}{\alpha_{\text{EM}}^{(i)}}\right) - \zeta_{\text{had}}\right], \tag{4.4}$$

where we defined $\zeta_{\text{had}} \equiv -\sum_{i=1}^{m_{\text{had}}} x_i \ln x_i$, with $\zeta_{\text{had}} > 0$ in analogy to ζ_{EM}.

Finally, we compute the weighted average of the depths of the shower maxima estimated for the electromagnetic and hadronic sectors of the primary interaction, as given in Equations (4.2) and (4.4), respectively. The contribution of each sector to $\Delta X_{\max}$ is weighted by the corresponding fraction of the primary energy. This yields,

$$\begin{aligned}\lambda_r^{-1}\xi \equiv \alpha_{\text{had}} t_{\max}^{\text{had}} + (1-\alpha_{\text{had}}) t_{\max}^{\text{EM}} = \\ = \ln\left(\frac{E_0}{2\xi_c^e}\right) + \left[\sum_{i=1}^{m_{\text{had}}} x_i \left(\lambda_i - \frac{\zeta_{\text{EM}}^{(i)}}{\alpha_{\text{EM}}^{(i)}}\right)\right] - \zeta_{\text{had}} - \zeta_{\text{EM}},\end{aligned} \tag{4.5}$$

where ξ is the variable of the primary interaction that predicts the $\Delta X_{\max}$ of a given cascade, in a shower-by-shower fashion. A first shower-by-shower estimator of $X_{\max}$ is obtained by adding the independent fluctuations of the depth of the first interaction

$$X_{\max} = X_1 + \xi. \tag{4.6}$$

As defined in Equation (4.5), the variable ξ still encompasses shower-to-shower fluctuations arising both from the interaction depths of hadronically interacting secondaries of the primary interaction and from variations in their subsequent particle production. To isolate only the fluctuations originating from the stochastic energy partition among the secondaries in the primary interaction, we introduce the following additional approximations:

1. Parametrise the average value of λ_i as a function of incident's hadron energy $E_i^{\text{had}} = E_0 x_i$ as [42]:

$$\lambda_i(E_i^{\text{had}}) = \lambda_0 - \delta \ln\left(\frac{E_0}{E_{\text{ref}}}\right) - \delta \ln x_i, \tag{4.7}$$

   with $\delta > 0$, and where E_{ref} denotes a reference energy such that $\lambda_i(E_{\text{ref}}) = \lambda_0$. All quantities are expressed in units of the radiation length λ_r, making both λ_0 and δ dimensionless. The interaction length of a hadron depends on its type as well as on the hadronic interaction model. However, to simplify the expression for ξ, we neglect this dependence on the hadron species. The validity of Equation (4.7) is assessed in Section 4.2.1.

2. Parametrise the average ratio $\zeta_{\text{EM},i}/\alpha_{\text{EM},i}$ as a function of the energy of the parent hadron, E_i^{had}, in the form

$$\frac{\zeta_{\text{EM},i}}{\alpha_{\text{EM},i}} \simeq \left\langle \frac{\zeta_{\text{EM}}}{\alpha_{\text{EM}}} \right\rangle = a_0 + a_1 \ln\left(\frac{E_0 x_i}{E_{\text{ref}}}\right), \tag{4.8}$$

with $a_1 > 0$. This functional form is motivated by the fact that for equipartition of energy $\zeta_{\rm EM}/\alpha_{\rm EM} = \ln m_{\rm tot}$ and the average $m_{\rm tot}$ increases approximately as a power law in energy [42]. As before, the validity of Equation (4.8) is assessed in Section 4.2.1.

Inserting the parametrisations defined in Equations (4.7) and (4.8) into Equation (4.5), and defining the parameters $\omega = 1 - \delta - a_1$ and $\mathcal{C}_0 = \lambda_0 - a_0$, leads to Equation (4.9):

$$\lambda_r^{-1}\xi = \ln\left(\frac{E_0}{2\xi_c^e}\right) + \left[\mathcal{C}_0 + (\omega - 1)\ln\left(\frac{E_0}{E_{\rm ref}}\right)\right]\alpha_{\rm had} - \omega\zeta_{\rm had} - \zeta_{\rm EM}. \tag{4.9}$$

Therefore, ξ is a variable built from the energy spectra of secondaries of the first p-air interaction alone and that provides an event-by-event estimation of $\Delta X_{\max}$. Furthermore, the quantities $\alpha_{\rm had}$, $\zeta_{\rm had}$ and $\zeta_{\rm EM}$, whose linear combination yields ξ, are production variables built directly from the energy spectra of secondaries, therefore establishing a causal connection between these spectra and $\Delta X_{\max}$. These variables will be studied in depth in Section 4.4.1 .

Note that in the case where we take only Equation (4.2), and assume equipartition of energy among secondaries, the model reduces to the standard Heitler-Matthews prediction for $\Delta X_{\max}$ given in Equation (1.30). Below, we present other interesting limits derived from Equation (4.9).

Elastic proton-air interactions

For perfectly elastic proton-air interactions, all energy is retained in the hadronic sector and carried mostly by the scattered proton. Therefore, $\zeta_{\rm EM} = 0$ and $\alpha_{\rm had} = 1$. Furthermore, since there is only one outgoing hadron having a fraction of energy $x = 1$, $\zeta_{\rm had} = -1 \times \ln 1 = 0$. Hence,

$$\lambda_r^{-1}\xi = \ln\left(\frac{E_0}{2\xi_c^e}\right) + \mathcal{C}_0 + (\omega - 1)\ln\left(\frac{E_0}{E_{\rm ref}}\right). \tag{4.10}$$

Highly energetic neutral pions

In the case of primary interactions where most of the primary energy is taken by a single neutral pion, we trivially have $\alpha_{\rm had} = \zeta_{\rm had} = \zeta_{\rm EM} = 0$, reducing ξ to

$$\xi = \lambda_r \ln\left(\frac{E_0}{2\xi_c^e}\right). \tag{4.11}$$

Naturally, this limit, which is the average value of $X_{\max}$ given by the Heitler model for a photon primary with energy $E_0/2$, is never achieved due to the conservation of other quantum numbers of the proton-air interaction. However, we will see that this extreme case provides an upper limit on the value of ξ.

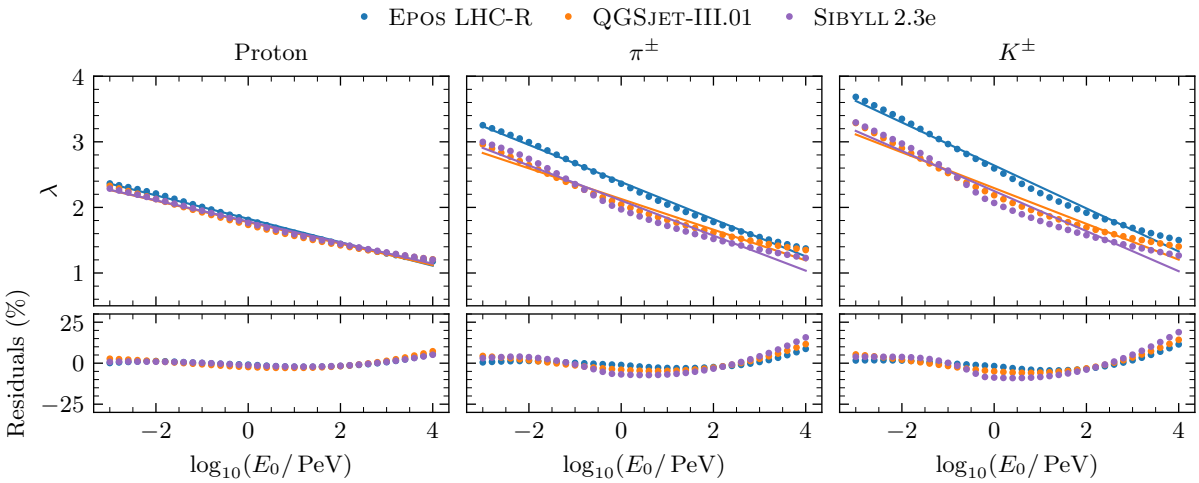

Figure 4.3: Energy dependence of the interaction length, λ, of protons (left), charged pions (center) and charged kaons (right) in air. The interaction length is relative to the radiation length $\lambda_r = 37\,\mathrm{g\,cm^{-2}}$. Circular markers refer to the predictions of the different hadronic interaction models and the full lines to linear fits performed with Equation (4.7).

Model	Proton-Air		$\pi^{\pm}$-Air		$K^{\pm}$-Air	
	δ	λ_0	δ	λ_0	δ	λ_0
EPOS LHC-R	0.08 ± 0.01	1.83 ± 0.01	0.12 ± 0.01	2.39 ± 0.01	0.14 ± 0.01	2.65 ± 0.01
QGSJET-III.01	0.07 ± 0.01	1.78 ± 0.01	0.10 ± 0.01	2.13 ± 0.01	0.12 ± 0.01	2.29 ± 0.02
SIBYLL2.3e	0.07 ± 0.01	1.79 ± 0.01	0.12 ± 0.01	2.11 ± 0.02	0.13 ± 0.01	2.25 ± 0.02

Table 4.1: Optimal parameters, δ and λ_0, extracted from fitting the energy-dependence of the interaction length of different hadrons in air to Equation (4.7).

4.2 Testing the causal connection between ξ and $\Delta X_{\max}$

The causal connection between ξ, introduced in Equation (4.9), and the shower observable $\Delta X_{\max}$ is quantified using libraries of 10^5 proton-induced CONEX showers as described in Section 2.1. To test this connection, we must first determine the values of the free parameters: $\mathcal{C}_0$ and ω, which were introduced in parametrisations expressed in Equations (4.7) and (4.8). These parameters do not vary from shower to shower, and their values change the relative contribution of the variables ζ_{had}, ζ_{EM} and α_{had} to ξ.

4.2.1 Optimization of the free parameters of ξ

Initial estimation of parameters

Using a set of CONEX simulations of proton-induced showers, with primary energies ranging from 10^{12} to 10^{19} eV, we obtain energy dependence of the interaction length of protons, charged pions and charged kaons in air, as shown in Figure 4.3. The energy-dependence of the interaction length of the different species is fit to Equation (4.7), to extract the values of λ_0 and δ for each projectile type and high-energy hadronic interaction model. The values of the optimal parameters are listed in Table 4.1 for $E_{\mathrm{ref}} = 1\,\mathrm{PeV}$.

Despite the non-linear evolution of the interaction length with the logarithm of the in-

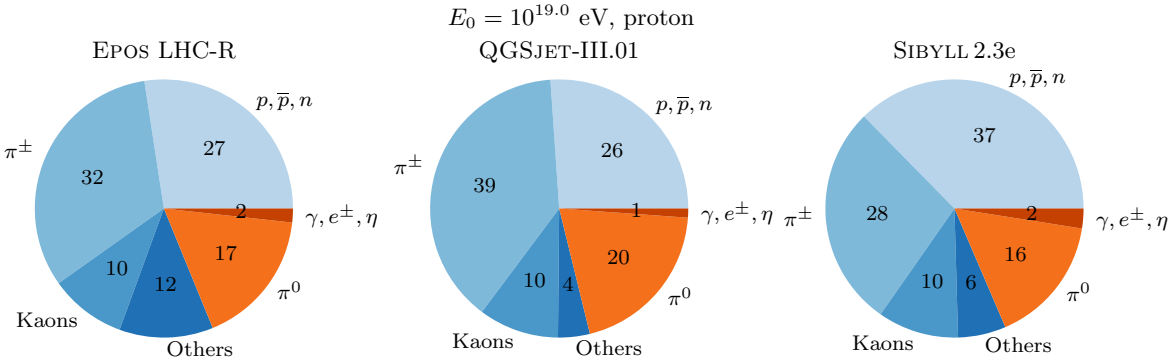

Figure 4.4: Average percentage of the primary energy carried by each particle sector in proton-air interactions. Blue shades refer to hadronically interacting particles, while orange shades refer to particles that feed the electromagnetic component of the shower.

Model	δ	λ_0
EPOS LHC-R	0.108 ± 0.001	2.200 ± 0.004
QGSJET-III.01	0.093 ± 0.002	2.027 ± 0.008
SIBYLL2.3e	0.095 ± 0.002	1.966 ± 0.008

Table 4.2: Initial estimations of the parameters λ_0 and δ, used to define $\mathcal{C}_0$ and ω (see Equation (4.9)).

cident energy, the relative deviation remains between 10% and 25% for proton and charged kaon projectiles, respectively. The discrepancy reaches its maximum at the highest energies. Nevertheless, in Section 4.2.1, we discuss the estimation and tuning of the parameters δ and λ_0, and validate the parameterisation of Equation (4.7) in an *a posteriori* fashion: by quantifying how well ξ reproduces the distribution of $\Delta X_{\max}$ after the tuning.

It is worth noting that the interaction length of protons in air exhibits only a weak dependence on the hadronic interaction model. This is largely due to the extensive constraints on the proton-proton cross-section over a wide energy range, provided by both accelerator data and air-shower measurements [56]. In contrast, the lack of data on the forward production of pions and kaons at energies relevant for ultra-high-energy cosmic rays results in a significantly larger model dependence for their predicted interaction cross-sections.

We compute the final values of the parameters δ and λ_0 by averaging their values for each particle species, weighting by the average fraction of the primary energy carried by each particle type in a proton-air interaction. The fraction of energy carried by the different particle sectors is shown as pie charts in Figure 4.4, for the different high-energy hadronic interaction models. The initial estimations of λ_0 and δ are displayed in Table 4.2.

In turn, the free parameters a_0 and a_1 are estimated by fitting the energy evolution of $\langle \zeta_{\rm EM}/\alpha_{\rm EM} \rangle$ to Equation (4.8), for three up-to-date hadronic interaction models. This is shown in Figure 4.5. The lower panel of the same figure shows the spread between the predictions of the different hadronic interaction models relative to the one predicting the lowest $\langle \zeta_{\rm EM}/\alpha_{\rm EM} \rangle$ in each energy bin. The model spread is defined as the difference between the maximum and minimum values of $\langle \zeta_{\rm EM}/\alpha_{\rm EM} \rangle$ across high-energy hadronic interaction models.

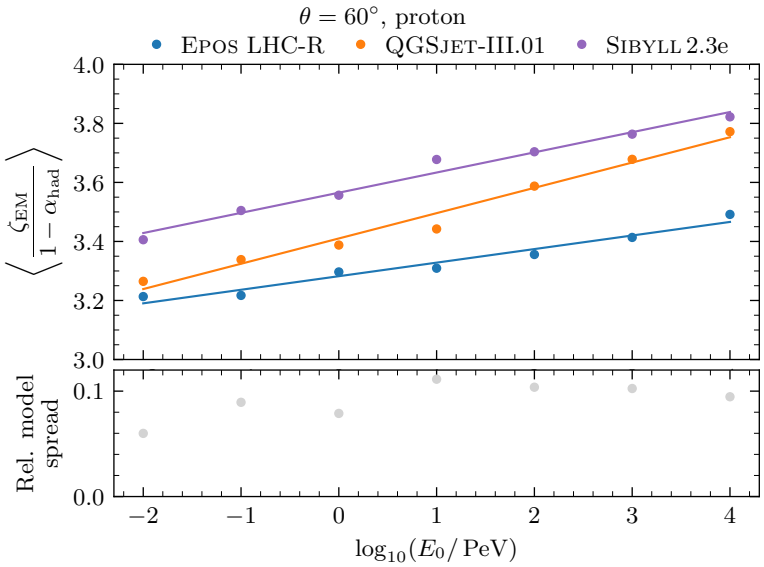

Figure 4.5: Top panel: Energy dependence of the ratio $\langle \zeta_{\rm EM}/1-\alpha_{\rm had}\rangle$ in proton-air interactions. Circular markers refer to the predictions of the different hadronic interaction models and the full lines to linear fits performed with Equation (4.8). Bottom panel: Relative model spread, computed as explained in the main text, between the hadronic model predictions for $\langle \zeta_{\rm EM}/1-\alpha_{\rm had}\rangle$, as a function of the primary energy.

Model	$\mathcal{C}_0$	ω
EPOS LHC-R	-1.08 ± 0.01	0.872 ± 0.002
QGSJET-III.01	-1.38 ± 0.02	0.870 ± 0.003
SIBYLL2.3e	-1.60 ± 0.01	0.875 ± 0.003

Table 4.3: Initial estimations of the free parameters in the definition of ξ, as given by Equation (4.9).

The evolution of $\langle \zeta_{\rm EM}/\alpha_{\rm EM}\rangle$ with the primary energy exhibits a strong dependence on the choice of high-energy hadronic interaction model. As expected from their tuning to lower-energy accelerator data, the model predictions begin to converge at lower energies. However, the absence of full convergence suggests that forward particle production remains insufficiently constrained. While the energy dependence of $\langle \zeta_{\rm EM}/\alpha_{\rm EM}\rangle$ could also have been examined for other relevant long-lived hadrons, such as charged pions and kaons, this was not pursued in the present study. Ultimately, the adequacy of the parametrisation in Equation (4.8) will be judged by the ability of the final variable, ξ, to accurately predict the value of $\Delta X_{\max}$ in the corresponding air shower cascade.

The initial estimations of $\mathcal{C}_0 = \lambda_0 - a_0$ and $\omega = 1 - \delta - a_1$, are shown in Table 4.3.

Optimization procedure

The optimisation of the free parameters $\mathcal{C}_0$ and ω must satisfy two key conditions:

- Minimise the contribution of the rest of the shower to $\Delta X_{\max}$, thereby ensuring a strong causal connection between this observable and ξ;

- Be independent of the hadronic interaction model, allowing the ξ-$\Delta X_{\max}$ relationship to be employed in a data-driven analysis.

In this work, these goals are achieved through a two-step procedure:

1. For each hadronic interaction model, we minimise a χ^2 function defined as

$$\chi^2 \equiv \sum_{k=1}^{N} \frac{(\Delta X_{\max,k} - \xi_k)^2}{\Delta X_{\max,k}}, \tag{4.12}$$

   where the index k runs over the N simulated showers. The resulting values of $\mathcal{C}_0$ are then averaged across the models, under the assumption that they converge at a reference energy E_{ref}.

2. With $\mathcal{C}_0$ fixed to its model-averaged value[1], $\mathcal{C}_0 = 1.25$, we repeat the χ^2 minimisation using the parametrisation

$$\omega = p_1 \left\langle \frac{\zeta_{\text{EM}}}{1-\alpha_{\text{had}}} \right\rangle + p_0, \tag{4.13}$$

   where the parameters p_0 and p_1 are optimised simultaneously across the three hadronic interaction models. In this way, model dependence is encoded solely through $\langle \zeta_{\text{EM}}/(1-\alpha_{\text{had}}) \rangle$, which is a function of the energy spectra of secondaries in the primary proton-air interaction. The effectiveness of this calibration is illustrated in Figure 4.6, where only the most recent hadronic interaction models, denoted by starred markers, are fitted with a linear function. The older versions of the models are used for validation purposes. Among them, the EPOS LHC model lies farthest from the calibration curve. The specific mechanisms within the model responsible for this discrepancy must be addressed elsewhere.

The optimal values of p_0 and p_1 for $E_0 = 10^{19}$ eV, listed in Table 4.6, allow Equation (4.9) to be expressed as

$$\xi \simeq 917 + \left[73 - 48 \left\langle \frac{\zeta_{\text{EM}}}{1-\alpha_{\text{had}}} \right\rangle\right] \alpha_{\text{had}} + \left[5 \left\langle \frac{\zeta_{\text{EM}}}{1-\alpha_{\text{had}}} \right\rangle - 40\right] \zeta_{\text{had}} - 37\zeta_{\text{EM}} \quad [\text{g cm}^{-2}]. \tag{4.14}$$

Note that, for the typical values of $\langle \zeta_{\text{EM}}/(1-\alpha_{\text{had}}) \rangle$ predicted by high-energy hadronic interaction models, the coefficients of the multiparticle production variables are negative. Since α_{had} is strongly correlated with N_μ, this provides a first indication of the origin of the observed anti-correlation between $\Delta X_{\max}$ and N_μ.

4.2.2 Evaluating the causal connection between ξ and $\Delta X_{\max}$

The strong connection between the shower-to-shower values of ξ and $\Delta X_{\max}$ is clearly illustrated in Figure 4.7, which shows their joint distribution for three high-energy hadronic interaction models, along with the corresponding Pearson correlation coefficient

[1] Only the most up-to-date hadronic interaction models, i.e., EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e, were included in the averaging and fitting procedures. Other models will later be used to validate the connection between ξ and $\Delta X_{\max}$.

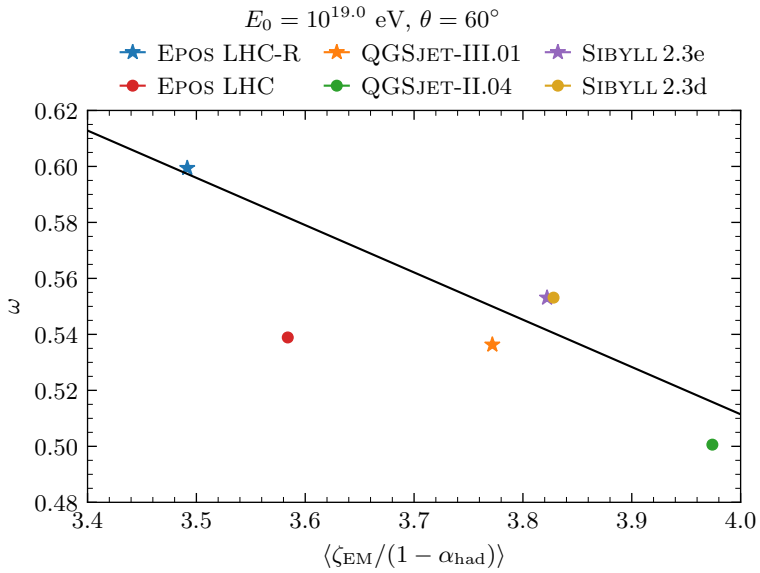

Figure 4.6: Calibration between the optimal value of the free parameter ω (see Equation (4.9)) and the value of $\langle \zeta_{\rm EM}/(1-\alpha_{\rm had}) \rangle$ for a collection of high-energy hadronic interaction models. The black line represents the regression curve obtained by fitting the model predictions to Equation (4.13). The figure was produced using libraries of 10^5 CONEX proton-induced showers, with $E_0 = 10^{19}$ eV and $\theta = 60°$.

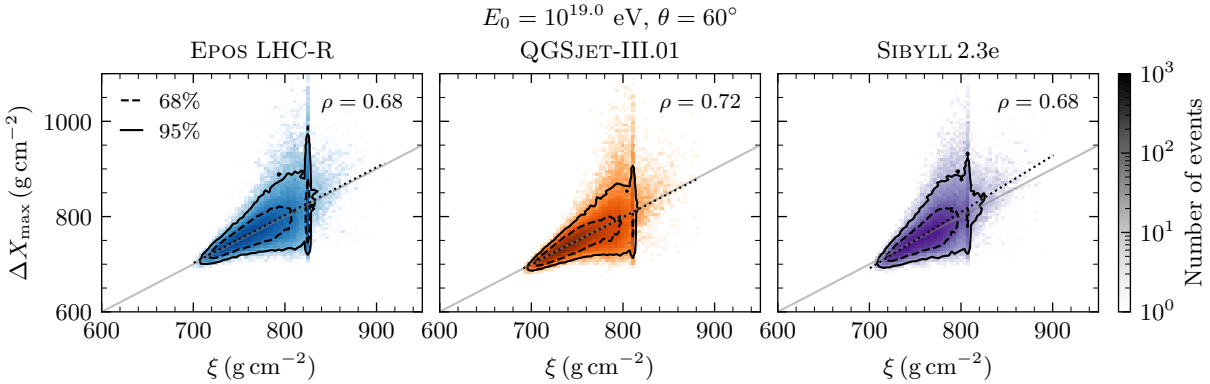

Figure 4.7: Correlation between the predictor of $\Delta X_{\max}$ from the primary p-air interaction, ξ, and the event-by-event values of $\Delta X_{\max}$. The contours containing 68 % and 95 % of the events are represented by the dashed and solid black lines, respectively. The 1:1 line is represented in solid grey and the linear regression curve as a dotted black line. This figure was produced using a library of 10^5 CONEX simulations of proton-induced showers, using up-to-date hadronic interaction models.

$\rho(\xi, \Delta X_{\max})$. The dashed and solid black contours enclose 68% and 95% of the events, respectively. For reference, the figure also includes the identity line (solid grey) and a fitted regression curve (dotted black), obtained from a linear fit to the pairs $(\xi, \Delta X_{\max})$.

The stochastic fluctuations of the primary interaction variable ξ are strongly correlated with those of the shower observable $\Delta X_{\max}$, as evidenced by their Pearson correlation coefficient, which ranges from $\rho = 0.68$ to 0.72 depending on the high-energy hadronic interaction model. For comparison, using more conventional production variables, such as the total multiplicity $m_{\rm tot}$ and elasticity $\kappa_{\rm el}$, combined as $\Delta X_{\max} \simeq \ln(E_0/2\xi_c^e) +$

Parameter	EPOS LHC-R	QGSJET-III.01	SIBYLL2.3e
$\sigma(\Delta X_{\max})$	45.1	40.1	43.0
$\sigma(\xi)$	30.3	29.3	26.6
ρ	0.68	0.72	0.68
$\langle \Delta X_{\max} - \xi \rangle$	2.8	-0.7	3.7
$\sigma_{\pm}(\Delta X_{\max} - \xi)$	(22.9, 25.2)	(18.9, 19.0)	(21.1, 28.4)

Table 4.4: Standard deviation of $\Delta X_{\max}$ and ξ, Pearson correlation coefficient between $\Delta X_{\max}$ and ξ, bias ($\langle \Delta X_{\max} - \xi \rangle$) and left and right resolutions $\sigma_{\pm}(\Delta X_{\max} - \xi)$, for three hadronic interaction models.

$\lambda_r \ln(\kappa_{\rm el}/m_{\rm tot})$ [119], results in significantly lower correlations, with $\rho \in [0.55, 0.63]$. This improvement highlights the benefit of a more detailed treatment of the information encoded in the energy spectra of secondary hadrons from the primary interaction, ultimately retained in $f(\Delta X_{\max})$. Additionally, fluctuations in ξ account for up to 56% of the variance in $\Delta X_{\max}$.

Two further features are evident in Figure 4.7. The first is a vertical band around $815\,\mathrm{g\,cm^{-2}}$, particularly visible in the EPOS LHC-R model, corresponding to highly diffractive events. The exact value of this vertical band is given by Equation (4.10). The second is the increasing spread in $\Delta X_{\max}$ with growing values of ξ, which is associated with the rising elasticity of the primary interaction, as discussed in Section 4.2.2.

Finally, we verified that the Pearson correlation coefficient between ξ and $\Delta X_{\max}$ is independent of the shower's zenith angle, varying by at most 0.02. For samples of 10^4 showers, and considering the uncertainty in ρ, given by $\sqrt{(1-\rho^2)/(10^4-2)} \simeq 0.007$, this variation is not statistically significant. This is expected due to the rapid decoupling of the electromagnetic component from the hadronic development in extensive air showers. The energy dependence of the strength of the causal link between ξ and $\Delta X_{\max}$ is thoroughly examined in Section 4.2.3.

The residuals $R_X \equiv \Delta X_{\max} - \xi$ capture both the intrinsic fluctuations in the shower development beyond the primary interaction and the approximations made in deriving the functional form of ξ. The marginal distribution of the residuals, $f(R_X) = \int f(R_X, \xi)\,\mathrm{d}\xi$, is shown in the upper panel of Figure 4.8 for the high-energy hadronic interaction models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e. Furthermore, the distributions of $\Delta X_{\max}$ and ξ, denoted by $f(\Delta X_{\max})$ and $f(\xi)$ respectively, are shown in the lower panel of the same figure.

The standard deviations of $\Delta X_{\max}$ and ξ, along with their Pearson correlation coefficient, are presented in Table 4.4 for the three high-energy hadronic interaction models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e. In addition, the bias and the asymmetric resolution in the prediction of $\Delta X_{\max}$ from ξ, denoted by $\langle \Delta X_{\max} - \xi \rangle$ and $\sigma_{\pm}(\Delta X_{\max} - \xi)$, respectively, are also reported in the table. Here, $\sigma_{-}(\Delta X_{\max} - \xi)$ represents the distance between the 0.16 quantile and the median of the residual distribution R_X, while $\sigma_{+}(\Delta X_{\max} - \xi)$ corresponds to the distance between the median and the 0.84 quantile. These quantiles correspond to the cumulative density function of a standardised normal distribution evaluated at -1 and $+1$, respectively.

The average value of ξ exhibits a bias with respect to the true mean of $\Delta X_{\max}$, ranging from approximately $-2\,\mathrm{g\,cm^{-2}}$ to $5\,\mathrm{g\,cm^{-2}}$, depending on the hadronic interaction model.

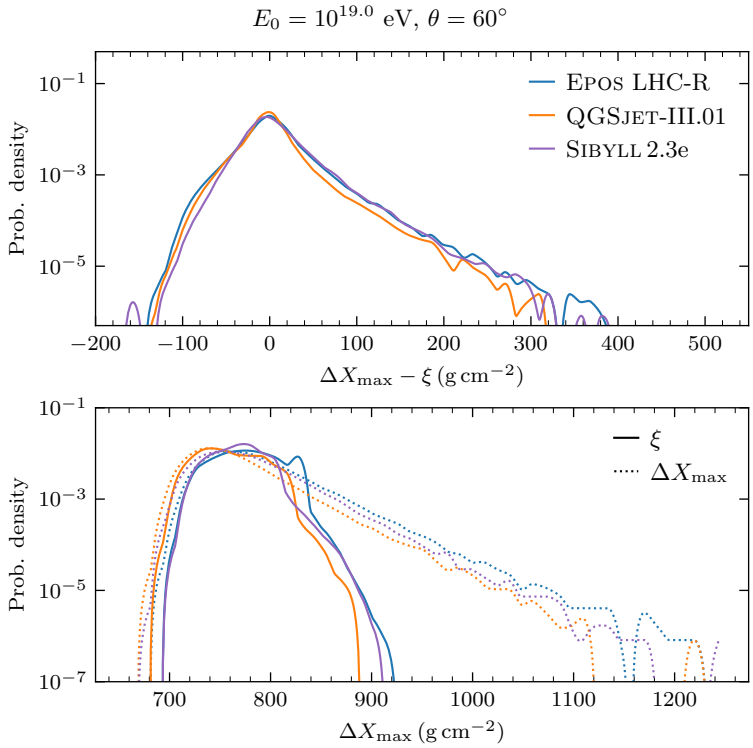

Figure 4.8: Upper panel: distribution of the residuals $\Delta X_{\max} - \xi$. Lower panel: distributions of $\Delta X_{\max}$ (dotted lines) and its predictor from the first interaction ξ (solid lines). The hadronic interaction models Epos LHC-R, QGSJet-III.01 and Sibyll2.3e are represented in blue, orange and purple, respectively. This figure was produced with the library of proton-induced Conex described in Section 2.1.

The resolution in the determination of $\Delta X_{\max}$ from ξ, varies between $\sigma_- \simeq 19\,\mathrm{g\,cm^{-2}}$ and $\sigma_+ \simeq 29\,\mathrm{g\,cm^{-2}}$, and it is mildly model-dependent. Notably, the shape of the distribution of residuals $f(R_X)$ is nearly identical across models and features an exponential tail toward positive values, which is responsible for the larger values of the right resolution with respect to the left resolution. These events correspond to the deep tail of the $\Delta X_{\max}$ distribution, shown in the lower panel of Figure 4.8, and are typically associated with highly elastic primary interactions. In such cases, the depth of the first interaction of the leading particle fluctuates significantly, strongly influencing the value of $\Delta X_{\max}$. In Section 4.2.2, we introduce an additional term to ξ to account for these fluctuations. Nevertheless, as will be shown in Section 4.2.2, the weak model dependence of $f(R_X)$ is sufficient to use the correspondence between ξ and $\Delta X_{\max}$, even without the inclusion of this term. Lastly, the agreement between the distributions of $\Delta X_{\max}$ and ξ for the bulk of proton-induced events is remarkably good.

Quantifying the useful information about primary interaction retained in ξ

As mentioned in the previous section, the fluctuations in particle production and propagation after the primary interaction and the approximations used in the derivation of ξ contribute to the variance of the residuals R_X. To isolate these contributions, we analyse a dedicated set of 10^4 Conex simulations of proton-induced showers, in which the transition

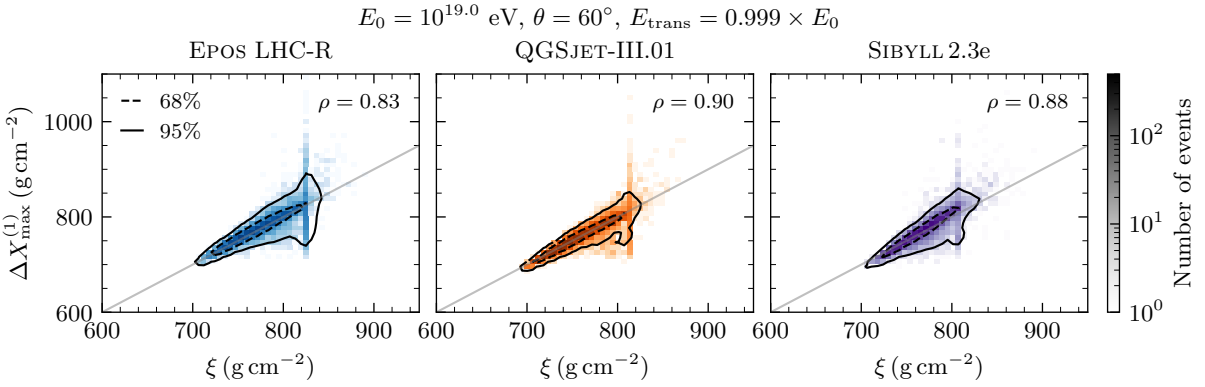

Figure 4.9: Correlation between ξ and $\Delta X^{(1)}_{\max}$, for stochastic primary proton-air interactions and a deterministic prediction of the rest of the shower. The distribution is obtained with CONEX proton-induced simulations as described in Section 2.1, but with the energy threshold between Monte-Carlo and Cascade Equations set to $E_{\text{th}} = 0.999 \times E_0$, using different high-energy hadronic interaction models.

threshold from Monte Carlo to cascade equations is set to $E_{\text{th}} = 0.999 \times E_0$. Under this configuration, stochastic fluctuations are restricted to the primary interaction, while the remainder of the shower evolves deterministically. An exception arises when the elasticity of the primary interaction exceeds 0.999, in which case the propagation and subsequent interaction of the leading proton are still handled with Monte Carlo. The resulting value of $\Delta X_{\max}$ for this constrained setup, denoted by $\Delta X^{(1)}_{\max}$, is shown as a function of ξ in Figure 4.9.

The Pearson correlation coefficient between ξ and $\Delta X^{(1)}_{\max}$, which ranges from 0.83 to 0.90 depending on the hadronic interaction model, demonstrates that ξ captures the majority of the variability in $\Delta X_{\max}$ originating exclusively from the stochasticity of the primary interaction. The lower correlation observed for EPOS LHC-R is attributed to its higher frequency of high-elasticity events, in particular with $\kappa_{\text{el}} > 0.999$, which degrade the strength of the causal link between ξ and $\Delta X_{\max}$.

The distributions of the residuals $\Delta X^{(1)}_{\max} - \xi$ and $\Delta X_{\max} - \xi$ are presented in the upper panel of Figure 4.10. The lower panel of the same figure shows the marginal distributions $f(\xi)$, $f(\Delta X_{\max})$, and $f(\Delta X^{(1)}_{\max})$, exemplified for the QGSJET-III.01 model. The residuals $\Delta X^{(1)}_{\max} - \xi$ are mainly due to the assumptions taken in the derivation of ξ, and their distribution is significantly narrower than $f(R_X)$, where the stochasticity of the rest of the shower is included.

The variance of $\Delta X_{\max}$ for a fully stochastic air shower can be expressed as the sum of the variance arising from the primary interaction alone, denoted $\sigma^2(\Delta X^{(1)}_{\max})$, and the variance associated with fluctuations in all subsequent generations of the shower, $\sigma^2(R)$:

$$\sigma^2(\Delta X_{\max}) = \sigma^2(\Delta X^{(1)}_{\max}) + \sigma^2(R), \tag{4.15}$$

where $R \equiv \Delta X_{\max} - \Delta X^{(1)}_{\max}$. Furthermore, the information loss incurred by the approximations made in the derivation of ξ, i.e., the portion of variability in $\Delta X^{(1)}_{\max}$ not captured by ξ, is quantified by the variance $\sigma^2(\Delta X^{(1)}_{\max} - \xi)$. This allows the total variance of $\Delta X_{\max}$

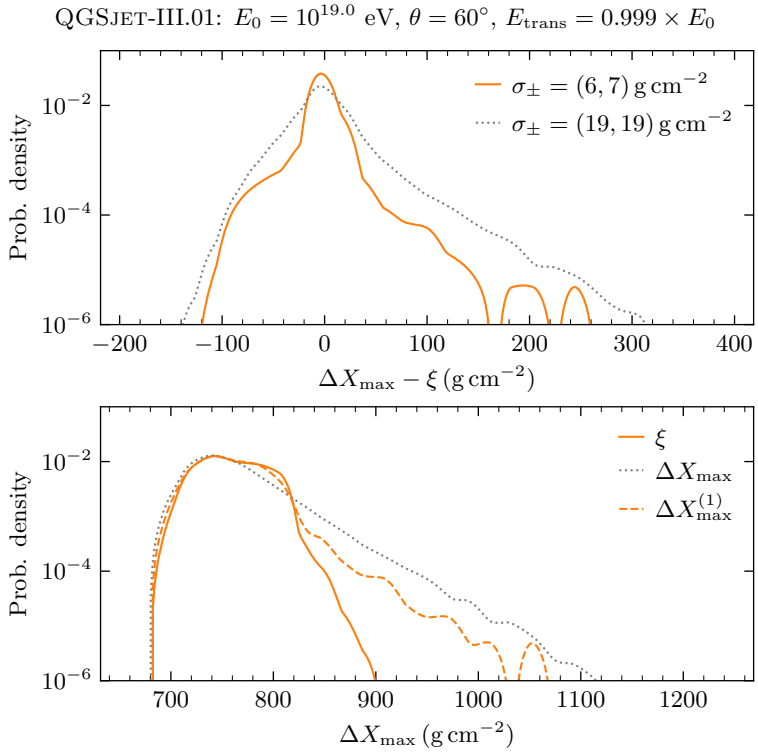

Figure 4.10: Upper panel: distribution of residuals $\Delta X_{\max} - \xi$ (dotted grey curve) and $\Delta X^{(1)}_{\max} - \xi$ (solid orange curve). The respective asymmetric resolutions are shown in the panel legend. Lower panel: distributions of ξ (solid orange), $\Delta X^{(1)}_{\max}$ (dashed orange) and $\Delta X_{\max}$ (dotted grey). The values of $\Delta X^{(1)}_{\max}$ are obtained from a library of CONEX simulations with stochastic primary interactions and a deterministic prediction of the rest of the shower, thanks to the high transition energy $E_{\text{trans}} = 0.999 \times E_0$, produced with the high-energy hadronic interaction model QGSJET-III.01.

to be approximately decomposed as

$$\sigma^2(\Delta X_{\max}) \simeq \sigma^2(\xi) + \sigma^2(\Delta X^{(1)}_{\max} - \xi) + \sigma^2_R, \tag{4.16}$$

where each term represents a distinct contribution to the fluctuations in $\Delta X_{\max}$: the fluctuations in ξ, those due to the model simplifications, and the cumulative variance from later shower generations, respectively.

The relative contributions of each term to $\sigma^2(\Delta X_{\max})$ are summarised in Table 4.5 for the considered high-energy hadronic interaction models.

In general, the stochasticity of the primary interaction alone accounts for approximately 65% of the total variance in $\Delta X_{\max}$. Of this, about 75% is captured by the variance of ξ. Thus, for a fully stochastic cascade, the fluctuations in ξ explain a significant fraction of the total variability in $\Delta X_{\max}$, reaching up to 56% in the case of QGSJET-III.01. Notably, with the exception of SIBYLL2.3e, the contribution of ξ to the total variance exceeds that from the remainder of the shower development beyond the first interaction.

These results indicate that the reduction in the precision of ξ in predicting $\Delta X_{\max}$ is primarily driven by fluctuations arising in the later stages of the shower development, rather than by limitations in the assumptions taken in the derivation of their connection.

Model	$\frac{\sigma^2(\Delta X^{(1)}_{\max})}{\sigma^2(\Delta X_{\max})}$	$\frac{\sigma^2(\xi)}{\sigma^2(\Delta X_{\max})}$	$\frac{\sigma^2(\Delta X^{(1)}_{\max}-\xi)}{\sigma^2(\Delta X_{\max})}$	$\frac{\sigma^2_R}{\sigma^2(\Delta X_{\max})}$
EPOS LHC-R	65%	47%	18%	35%
QGSJET-III.01	65%	56%	9%	35%
SIBYLL2.3e	58%	34%	24%	42%

Table 4.5: Relative contributions of the variances in ξ, $\sigma^2(\xi)$, the approximations taken in the derivation of $\xi, \sigma(\Delta X^{(1)}_{\max}-\xi)$, and the remaining interactions in the shower, σ^2_R, to the variance in $\Delta X_{\max}$, $\sigma^2(\Delta X_{\max})$, for the different hadronic interaction models.

One notable exception arises when an extremely energetic neutral pion is produced in the initial p-air interaction. In this scenario, described by Equation (4.11), ξ takes on a nearly fixed value, while the two dominant electromagnetic sub-showers can exhibit significant fluctuations in their maxima, thereby weakening the correlation between ξ and $\Delta X_{\max}$.

Impact of diffractive primary interactions in the tail of $\Delta X_{\max}$

As observed previously, the distribution of residuals, $f(R_X)$, exhibits a long exponential tail extending towards deep showers. This feature can be attributed to highly elastic primary interactions. In such cases, the primary proton-air interaction carries little information to explain fluctuations in shower observables, which become more sensitive to fluctuations in later interactions. In particular, their fluctuations will reflect, at least, the stochasticity of the propagation of the leading secondary of the cosmic ray-air interaction and hadron production upon its interaction. Furthermore, due to the low multiplicity of secondaries produced in these interactions, the fluctuations in the interaction depths of hadronically interacting particles are less suppressed [133]. Therefore, approximating these depths to their energy-dependent average, as done in the derivation of ξ, becomes less reliable. As a result, the variability in the residuals $R_X = \Delta X_{\max} - \xi$ increases with the elasticity of the first interaction, $\kappa_{\rm el}$, as shown in the left panel of Figure 4.11. There, the joint distribution $f(\kappa_{\rm el}, \Delta X_{\max} - \xi)$ is weighted by the logarithmic ratio between the true interaction depth of the leading particle in the primary interaction, λ_ℓ, and $\lambda(\kappa_{\rm el})$ (the predicted depth given by Equation (4.7) evaluated at $\kappa_{\rm el} E_0$).

The right panel of Figure 4.11 shows the evolution of the asymmetric resolution, $\sigma_\pm(R_X)$, relative to the median, as a function of $\lambda_\ell - \lambda(\kappa_{\rm el})$, for two populations of primary interactions: those with elasticity below the median, $\overline{\kappa_{\rm el}}$ (i.e., more inelastic interactions), and those above it.

Extreme values of the residuals $\Delta X_{\max} - \xi$ occur when the true interaction depth of the leading particle deviates significantly from the predicted value based on its energy $\kappa_{\rm el} E_0$. In particular, $\Delta X_{\max}$ tends to be smaller than its prediction ξ when the leading particle is very energetic and undergoes a shallow interaction depth. Conversely, $\Delta X_{\max}$ tends to be larger than ξ when the leading particle propagates over several interaction lengths before interacting again. Since the true value of λ_ℓ is exponentially distributed with a mean of $\lambda(\kappa_{\rm el} E_0)$, these events naturally produce the exponential tail towards positive values observed in $f(R_X)$.

The right panel of Figure 4.11 confirms that, for highly inelastic primary interactions, even significant misestimations of λ_ℓ have little impact on $\Delta X_{\max}$, resulting in precise

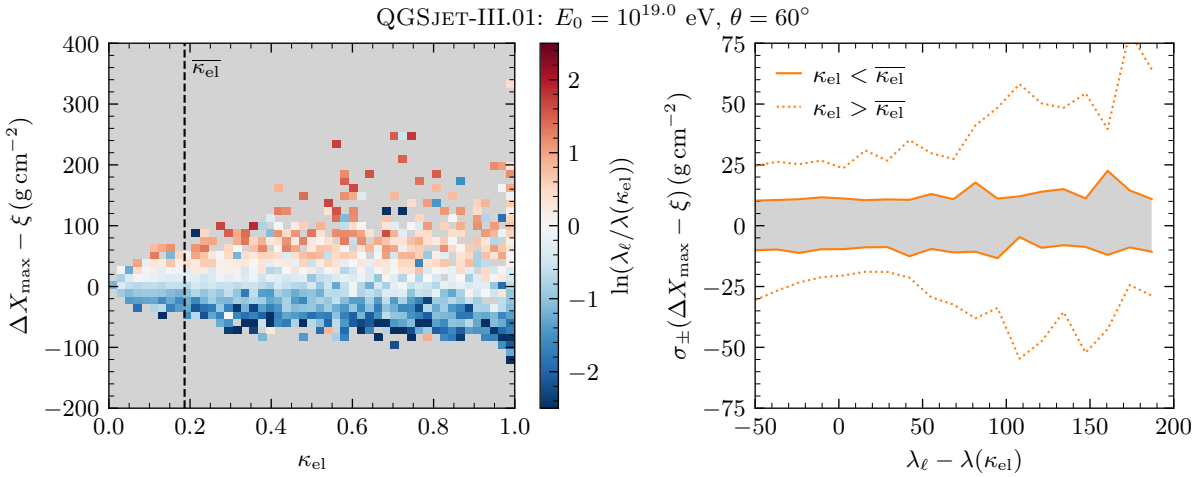

Figure 4.11: Left panel: Distribution of the residuals $\Delta X_{\max} - \xi$ as a function of the elasticity of the primary p-air interaction weighted by $\ln \lambda_\ell/\lambda(\kappa_{\rm el})$. The median value of $\kappa_{\rm el}$, $\overline{\kappa_{\rm el}}$, is represented by the vertical dashed line. Right panel: Asymmetric standard deviation of the residuals R_X, relative to its median, as a function of $\lambda_\ell - \lambda(\kappa_{\rm el})$, for primary interactions with: $\kappa_{\rm el} < \overline{\kappa_{\rm el}}$ (shaded grey area) and $\kappa_{\rm el} > \overline{\kappa_{\rm el}}$ (dotted red lines). This figure was produced with the library of CONEX simulations of proton-induced showers, as described in Section 2.1, using the hadronic interaction model QGSJET-III.01.

predictions from ξ. The opposite holds for highly elastic primary interactions: deviations of the actual interaction depth from the expected value lead to large fluctuations in $\Delta X_{\max}$ at fixed ξ. This is due to the large fraction of energy carried by the leading particle, which makes it dominant in determining the characteristics of the resulting shower.

This insight motivates the inclusion of a correction term in the definition of ξ (see Equation (4.9)) that accounts for the exponential fluctuations of λ_ℓ. The extended variable of the primary interaction, $\widetilde{\xi}$, becomes:

$$\widetilde{\xi} = \xi + \kappa_{\rm el}\,(\lambda_\ell - \lambda(\kappa_{\rm el}))\,, \tag{4.17}$$

where $\lambda(\kappa_{\rm el}) = \lambda_r\,[\lambda_0 - \delta \ln(\kappa_{\rm el} E_0/E_{\rm ref})]$.

For a library of 10^4 CONEX simulations in which the particle production of the leading particle in the primary interaction was recorded, we compute the joint distributions of ξ and $\Delta X_{\max}$, shown in Figure 4.12 for three hadronic interaction models. The parameters λ_0 and δ were optimised by minimising a χ^2 defined in Equation (4.12) with $\widetilde{\xi}$, while all other parameters used in the definition of ξ were held fixed to their optimal values.

The correlation between $\widetilde{\xi}$ and $\Delta X_{\max}$ is significantly improved relative to that between ξ and $\Delta X_{\max}$, across all hadronic interaction models considered. Naturally, the diffractive line observed in Figure 4.7 is no longer present in $f(\widetilde{\xi}, \Delta X_{\max})$.

The distributions of residuals $\Delta X_{\max} - \xi$ and $\Delta X_{\max} - \widetilde{\xi}$ are shown in the top panel of Figure 4.13, while the bottom panel displays the distributions $f(\xi)$, $f(\widetilde{\xi})$, and $f(\Delta X_{\max})$ for the hadronic interaction model QGSJET-III.01.

The distribution of $\widetilde{\xi}$ exhibits a more pronounced exponential tail at large depths, better matching the full distribution of $\Delta X_{\max}$ compared to ξ. Moreover, the diffractive peak present in the distribution of ξ is largely absent in $f(\widetilde{\xi})$. Additionally, the distribution

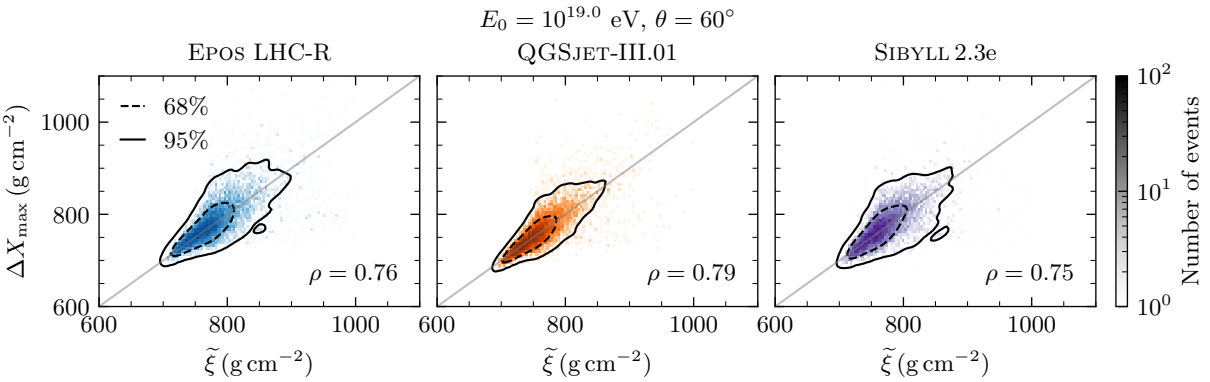

Figure 4.12: Correlation between $\widetilde{\xi}$ and $\Delta X_{\max}$. The contours containing 68% and 95% of the events are represented by the dashed and solid black lines, respectively. The 1:1 line is shown in solid grey. This figure was produced using a library of 10^4 CONEX simulations of proton-induced showers, using up-to-date hadronic interaction models.

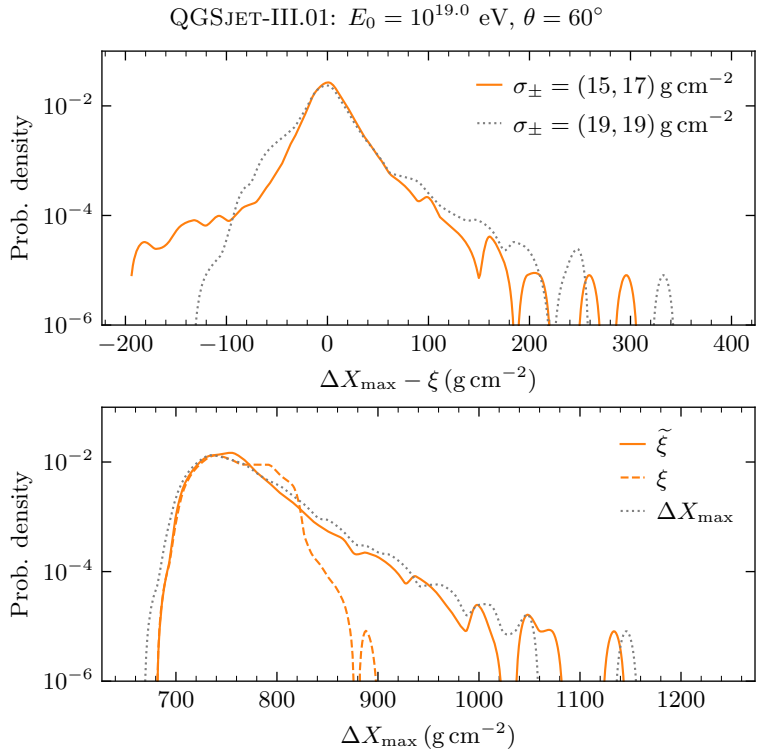

Figure 4.13: Upper panel: Distribution of the residuals $\Delta X_{\max} - \xi$ and $\Delta X_{\max} - \widetilde{\xi}$. Lower panel: Distributions of ξ, $\widetilde{\xi}$ and $\Delta X_{\max}$. This figure was produced with the library of CONEX simulations of proton-induced showers, as described in Section 2.1, using the hadronic interaction model QGSJET-III.01.

of the residuals for $\widetilde{\xi}$ is narrower, more symmetric, and exhibits a less prominent deep tail, leading to improved resolutions of $\sigma_\pm = (15, 17)\,\mathrm{g\,cm^{-2}}$.

In conclusion, the exponential tail of $\Delta X_{\max}$ primarily originates from the exponential fluctuations in the interaction depth of the leading particle in the primary interaction—a contribution that grows in importance with increasing elasticity. Therefore, fluctuations in particle production in later shower generations are not the main source of reduced

$\log_{10}(E_0/\mathrm{eV})$	p_1	p_0
17.0	-0.324 ± 0.003	1.57 ± 0.01
17.5	-0.316 ± 0.004	1.64 ± 0.01
18.0	-0.222 ± 0.003	1.34 ± 0.01
18.5	-0.153 ± 0.003	1.12 ± 0.01
19.0	-0.140 ± 0.004	1.09 ± 0.01
19.5	-0.122 ± 0.004	1.04 ± 0.01
20.0	-0.090 ± 0.004	0.93 ± 0.01

Table 4.6: Primary-energy dependence of the optimal parameters used in Equation (4.18).

precision in the prediction of $\Delta X_{\max}$ from ξ.

4.2.3 Energy evolution of the causal link between ξ and $\Delta X_{\max}$

We further validate the connection between ξ and $\Delta X_{\max}$ by computing the primary energy dependence of the mean of ξ in the energy range $\log_{10}(E_0/\mathrm{eV}) \in [17, 19]$ in steps of $\Delta \log_{10}(E_0/\mathrm{eV}) = 0.5$. For the considered primary energies, the parameter $\mathcal{C}_0$ in Equation (4.9) is fixed to its value obtained at $E_0 = 10^{19}\,\mathrm{eV}$, while ω is generalized as

$$\omega(E_0) = p_1(E_0)\,\langle \zeta_{\mathrm{EM}}/(1-\alpha_{\mathrm{had}})\rangle + p_2(E_0), \tag{4.18}$$

where $p_1(E_0)$ and $p_2(E_0)$ vary with E_0 as specified in Table 4.6. As before, the dependence on the hadronic interaction model is replaced by the dependence on $\langle \zeta_{\mathrm{EM}}/(1-\alpha_{\mathrm{had}})\rangle$, which is solely determined by the energy spectra of the secondary particles.

The energy evolution of $\langle \xi \rangle$ and $\langle \Delta X_{\max} \rangle$ is shown in Figure 4.14 for the high-energy hadronic interaction models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e. The same figure also includes fits to a linear function of the form $y = \widetilde{D}_p \ln E_0 + b$, where $\widetilde{D}_p$ represents the elongation rate.

The evolution of $\langle \xi \rangle$ is linear with $\ln E_0$, accurately reproducing the expected energy dependence of $\langle \Delta X_{\max} \rangle$ with biases smaller than $\pm 5\,\mathrm{g\,cm^{-2}}$ across two decades in primary energy. This bias amounts to approximately 20% of the spread among hadronic interaction model predictions for $\langle \Delta X_{\max} \rangle$. These results validate both the parameterisations used in the definition of ξ and the approximations adopted in its derivation. Furthermore, the elongation rate predicted from ξ differs by at most $2\,\mathrm{g\,cm^{-2}}$ from that of $\Delta X_{\max}$, depending on the hadronic interaction model. This value is comparable to the inter-model differences[2] in elongation rate.

The energy dependence of the standard deviations of $\Delta X_{\max}$ and ξ is shown in the left panel of Figure 4.15. The right panel displays the Pearson correlation coefficient between $\Delta X_{\max}$ and ξ, as well as between $\Delta X^{(1)}_{\max}$ and ξ, as a function of the primary energy.

As the primary energy decreases, so does the multiplicity of secondaries produced in the primary interaction. Consequently, fluctuations in later shower generations are less suppressed, leading to an overall increase in the variability of $\Delta X_{\max}$. As a result, $\sigma(\Delta X_{\max})$ decreases with increasing primary energy. In contrast, and as will be shown in Section 4.4.2, the distributions of the multiparticle production variables ζ_{EM} and ζ_{had}

[2] That is, the spread between the predictions of the different hadronic interaction models

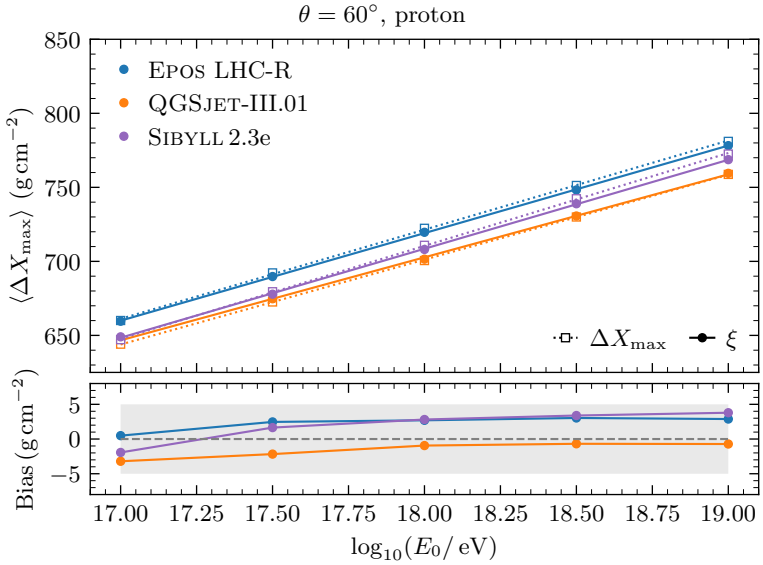

Figure 4.14: Top panel: Average values of ξ and $\Delta X_{\max}$ as a function of the primary energy. The linear fits to these dependencies are depicted by the dotted and solid lines. Bottom panel: Bias $\langle \Delta X_{\max} - \xi \rangle$ as a function of the primary energy, for the hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e. The grey band represents the $\pm 5\,\mathrm{g\,cm^{-2}}$ region, while the dashed grey line corresponds to a null bias.

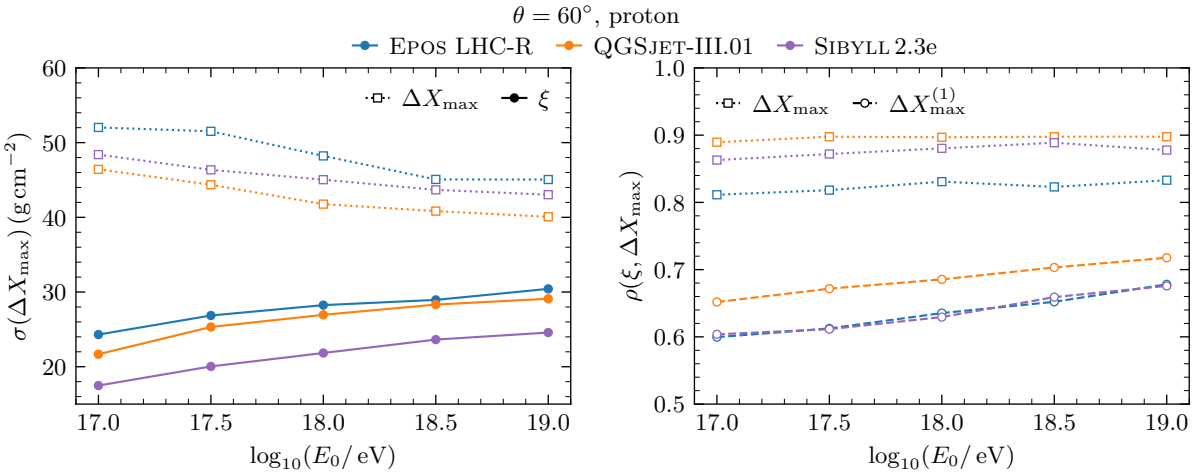

Figure 4.15: Left panel: Standard deviation of ξ and $\Delta X_{\max}$ as a function of the primary energy. Right panel: Primary-energy evolution of the Pearson correlation coefficient between ξ and $X_{\max}$, and between ξ and $X^{(1)}_{\max}$ (the values of $X_{\max}$ for a deterministic shower after a stochastic primary interaction). This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers using several primary energies and the hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e.

from the primary interaction become narrower at lower energies, resulting in a more peaked distribution of ξ. This effect is also evident in Figure 4.16, which shows the joint distribution $f(\xi, \Delta X_{\max} \,|\, E_0)$. The energy dependence of the values of $\Delta X_{\max}$ and ξ was removed by defining the quantities ΔX_{19} and ξ_{19} as in Equation 2.8, using the respective elongation rates. This facilitates the visual comparison of $f(\xi_{19}, \Delta X_{19})$ across primary

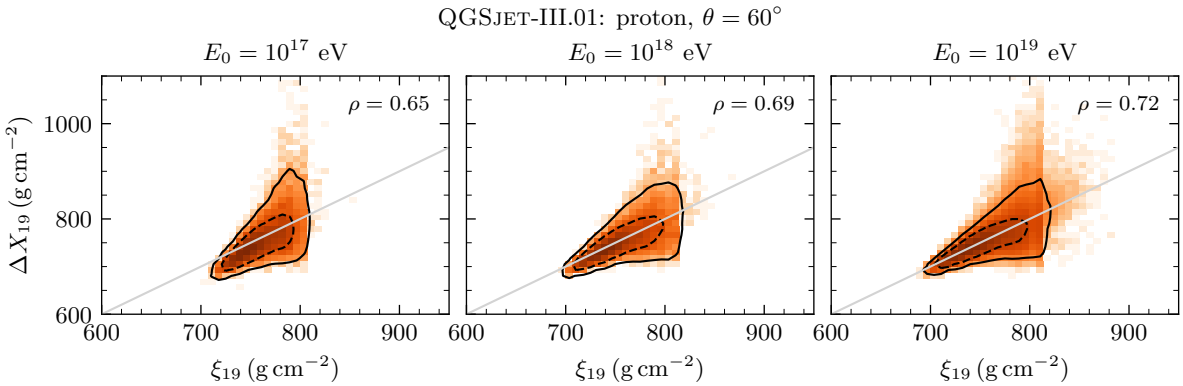

Figure 4.16: Joint distributions $f(\xi_{19}, \Delta X_{19} \,|\, E_0)$ as a function of the primary energy (increasing from left to right in steps of $\Delta \log_{10}(E_0/\text{eV}) = 1$). Each panel contains the Pearson correlation coefficient. The contours containing 68% and 95% of the events are shown in dashed and solid black lines, respectively. The 1-1 line is represented in light grey. This figure was produced with libraries of 10^4 CONEX simulations of proton-induced showers for different primary energies and $\theta = 60^\circ$, using QGSJET-III.01.

energies.

Finally, the decreasing importance of the primary interaction in determining the shower-to-shower value of $\Delta X_{\max}$ is reflected in the monotonic decrease of the Pearson correlation coefficient $\rho(\xi, \Delta X_{\max})$ with decreasing primary energy. However, the nearly constant correlation between ξ and $\Delta X^{(1)}_{\max}$ (see the right panel of Figure 4.15) confirms that the approximations made in the derivation of ξ remain valid even at lower primary energies.

4.3 A complete and universal model of the distribution of $X_{\max}$

4.3.1 Universality of the shower response to ξ

According to the results of Section 4.2, the fluctuations of ξ account for approximately 75% of the total information from the primary interaction that determines the shower's $\Delta X_{\max}$. Furthermore, we have established that the majority of the variability in $\Delta X_{\max}$ originates from the stochastic nature of ξ. We must now model the effects of particle production and propagation fluctuations in subsequent generations of the shower, as well as the information lost during the derivation of ξ due to model assumptions. These effects are encoded in the response of the shower to a given value of ξ from the primary proton-air interaction, described by the conditional probability density function $f(\Delta X_{\max} \,|\, \xi)$. This PDF can be re-centered and expressed in terms of the residuals $R_X = \Delta X_{\max} - \xi$, yielding the form $f(R_X \,|\, \xi)$, which is commonly referred to as the kernel of the integral transformation $f(\xi) \to f(\Delta X_{\max})$. This transformation is explicitly given by:

$$f(\Delta X_{\max}) = \int f(\xi) f(R_X \,|\, \xi, M) \, \mathrm{d}\xi \,, \tag{4.19}$$

where we have made explicit the kernel's dependence on the hadronic interaction model M.

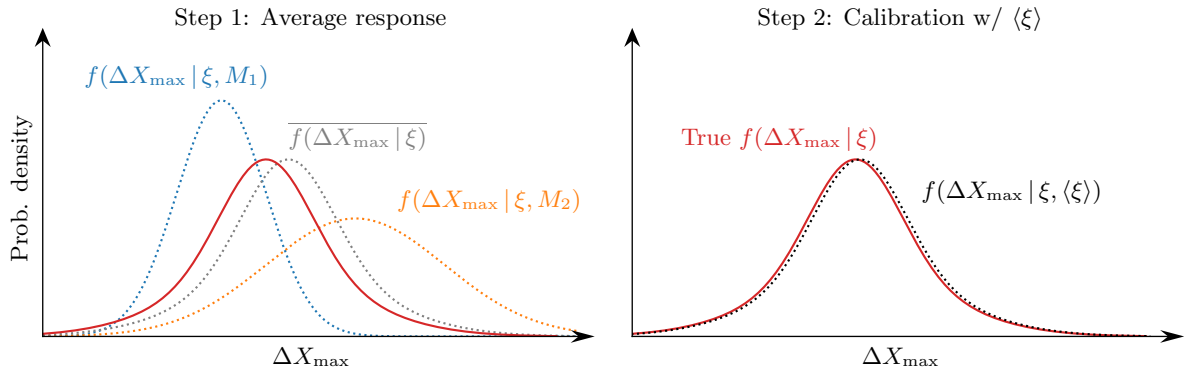

Figure 4.17: Deduction scheme for a universal shower response in two steps: modelling the response function $\overline{f(\Delta X_{\max}\,|\,\xi)}$ by averaging responses from two hadronic interaction models M_1 and M_2 (left panel); introducing a dependence on $\langle\xi\rangle$ to account for differences in shower modelling beyond the primary interaction (right panel).

While the role of $f(R_X\,|\,\xi, M)$ in shaping $f(\Delta X_{\max})$ is subdominant compared to that of $f(\xi)$, the ability to reconstruct the distribution of ξ from a measured distribution of $X_{\max}$ also depends on how much the response of the shower depends on the hadronic interaction model M. This motivates the study of the universality of the kernel $f(\Delta X_{\max}\,|\,\xi, M)$.

Let us adopt a shorthand notation where the PDF $f_X(x)$ of a random variable is denoted by the variable itself: $f_X(x) \equiv x$. Using this convention, Equation (4.19) can be rewritten as:

$$\Delta X_{\max} = \xi \otimes (R_X\,|\,\xi, M), \tag{4.20}$$

where $\otimes(R_X\,|\,\xi, M)$ denotes the convolution with the kernel $f(\Delta X_{\max} - \xi\,|\,\xi, M)$.

The model dependence of $f(\Delta X_{\max} - \xi\,|\,\xi, M)$ is addressed through a two-step procedure, using the CONEX simulation libraries described in Section 2.1. This process is illustrated in Figure 4.17 and consists of the following steps:

1. *Compute a model-averaged shower response to* ξ: For each hadronic interaction model $M \in \{$EPOS LHC-R, QGSJET-III.01, SIBYLL2.3e$\}$ and each fine bin in $f(\xi)$, we select the corresponding subset of $\Delta X_{\max}$ values. For each subset, the conditional PDF $f(\Delta X_{\max}\,|\,\xi, M)$ is estimated using a Gaussian Kernel Density Estimation (KDE). The bandwidth of the kernel was first estimated using the Silverman rule [139] and then optimised as will be described in Chapter 5, as the optimisation already involves the full distribution $f(N_\mu, X_{\max})$. To average over models, we first recentre each response at the common mean $\overline{\Delta X_{\max}} = \frac{1}{3}\sum_{M=1}^{3} \langle \Delta X_{\max} \rangle_M$ within each ξ bin, then compute the average PDF:

$$\overline{f(\Delta X_{\max}\,|\,\xi)} = \frac{1}{3}\sum_{i=1}^{3} f(2\Delta X_{\max} - \overline{\Delta X_{\max}}\,|\,\xi, M_i), \tag{4.21}$$

where $\overline{f(\Delta X_{\max}\,|\,\xi)}$ denotes the model-averaged response. Note that this binned approach is more easily generalisable to properly model $f(N_\mu, X_{\max})$ in Chapter 5. A caveat of this approach is that it implicitly assumes the shape of $f(\xi)$ can vary

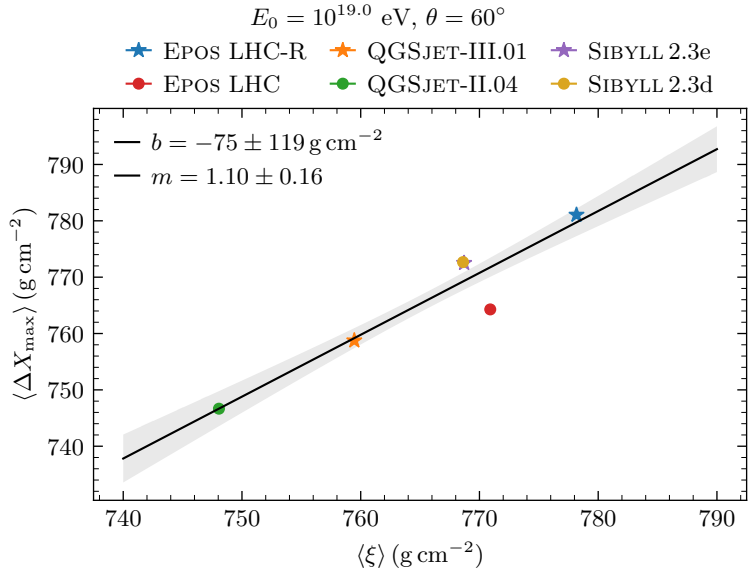

Figure 4.18: Scatter plot of $\langle \xi \rangle$ vs. $\langle \Delta X_{\max} \rangle$ for different hadronic interaction models. Newer models use starred markers; earlier versions use circles. The regression line and its uncertainty band are shown in black and grey, respectively. This figure was produced with the library of CONEX simulations of proton-induced showers, as described in Section 2.1, using $E_0 = 10^{19}$ eV and $\theta = 60°$.

independently of the subsequent particle production processes. In practice, this simplification would imply abrupt changes in secondary particle production as a function of the primary energy, which are difficult to justify without invoking exotic physics scenarios. In particular, without resorting to such scenarios, it was shown that the later stages of shower development also influence the $X_{\max}$ scale [120]. Consequently, using an average response must introduce biases in the moments of the distribution of $\Delta X_{\max}$, which we must correct for in the second step of the method, without relying on the hadronic interaction model dependence of the kernel.

2. *Calibrate the shower response using* $\langle \xi \rangle$: Assuming that changes in the primary interaction energy spectra propagate similarly through all hadronic interaction models, we hypothesise that $\langle \xi \rangle$ captures the model dependence of the full shower development. The use of this measure of the first moment of the secondary spectra is motivated by the fact that subsequent generations involve many interactions occurring in parallel, leading to a suppression of fluctuations in later shower generations. To test this hypothesis, we perform a linear calibration of $\langle \Delta X_{\max} \rangle$ as a function of $\langle \xi \rangle$ using several state-of-the-art hadronic interaction models. Figure 4.18 displays the resulting calibration, resulting from a linear regression of the form $\langle \Delta X_{\max} \rangle = m \langle \xi \rangle + b$.

Since the simulation libraries contain 10^5 events, the statistical uncertainty on $\langle \Delta X_{\max} \rangle$ is negligible. Therefore, to estimate realistic uncertainties in the regression parameters m and b, we perform a two-step χ^2 fit: first, using unitary uncertainties for each model point; then, estimating the standard deviation of the model residuals relative to the regression curve and using this value as the new error for each point. The final errors on m and b account for their covariance and are propagated to the

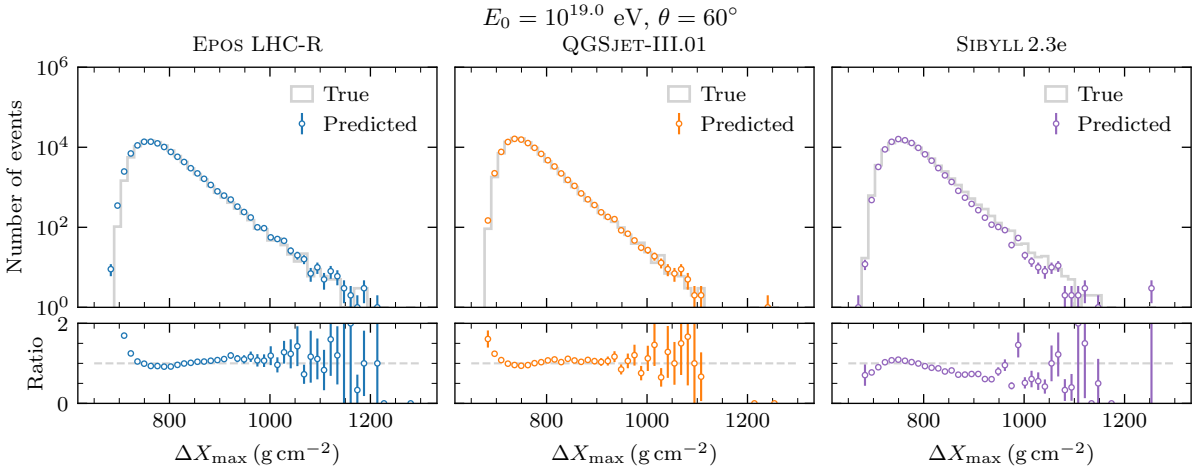

Figure 4.19: Upper panels: Distributions of $\Delta X_{\max}$ predicted by applying the universal shower response model ($R_X \,|\, \xi, \langle\xi\rangle$) to prior distributions of ξ produced with the hadronic interaction models EPOS LHC-R (left), QGSJET-III.01 (middle) and SIBYLL2.3e (right). The underlying grey histograms are true distributions of $\Delta X_{\max}$. Lower panels: ratio of the predicted to the true distributions $f(\Delta X_{\max})$. This figure was produced with the library of 10^5 CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$.

calibration curve as represented by the shaded area in Figure 4.18.

The linear relation between $\langle\xi\rangle$ and $\langle\Delta X_{\max}\rangle$ shows a dispersion of 8 g cm^{-2}, which is comparable with the systematic uncertainties in $X_{\max}$ measurements from Fluorescence Detectors [76], and represents about 20% of the total inter-model spread in $\Delta X_{\max}$. Notably, this relation holds even for models not used in the calibration of the free parameters in the definition of ξ, supporting the universality of the correspondence between $\langle\xi\rangle$ and $\langle\Delta X_{\max}\rangle$.

Using the calibration $\langle\Delta X_{\max}\rangle = m\,\langle\xi\rangle + b$, we redefine the kernel to explicitly depend on $\langle\xi\rangle$ rather than on M. In practice, we shift samples of events from the average kernel $\overline{f(\Delta X_{\max} \,|\, \xi)}$ so that their mean $\Delta X_{\max}$ matches the calibrated $m\,\langle\xi\rangle + b$. This final kernel is denoted by $f(R_X \,|\, \xi, \langle\xi\rangle)$ and allows us to express the predicted $f(\Delta X_{\max})$ as:

$$\Delta X_{\max} = \xi \otimes (R_X \,|\, \xi, \langle\xi\rangle). \tag{4.22}$$

We validate this model by generating prior distributions $f(\xi)$ from EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e, and predicting the corresponding $\Delta X_{\max}$ distributions using Equation (4.22). Results are shown in Figure 4.19, where the predicted distributions are compared to the true ones, with ratio plots in the lower panels.

The model successfully reproduces the true $\Delta X_{\max}$ distributions, including the tail towards deep showers, indicating that the differences across models are well captured by the priors $f(\xi)$. Thus, the $\Delta X_{\max}$ distribution effectively encodes the energy spectra of secondaries in the primary interaction. Note that the sharply falling tail of the distribution of $\Delta X_{\max}$ towards shallow showers cannot be well reproduced by the Gaussian KDE of the shower response. The impact of this mismatch on the main moments of the distributions is negligible, as shown in the next paragraph.

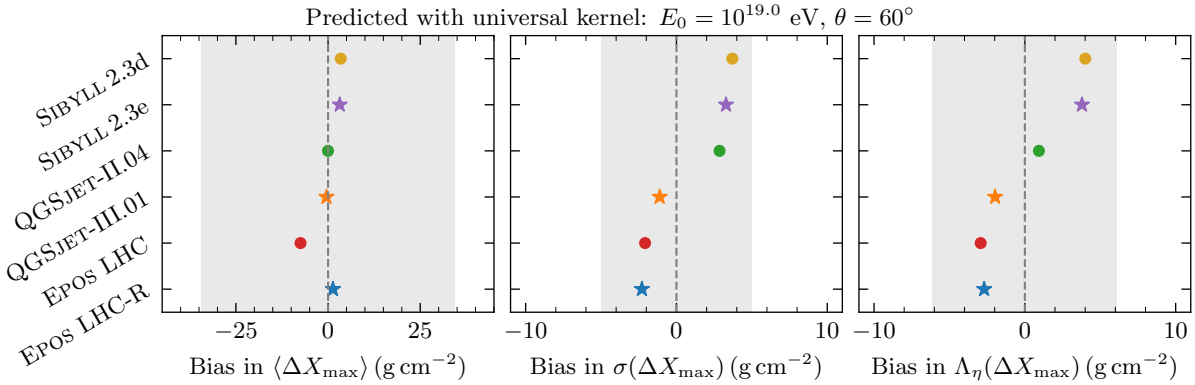

Figure 4.20: Biases in $\langle \Delta X_{\max} \rangle$ (left), $\sigma(\Delta X_{\max})$ (middle), and Λ_η (right) due to use of the universal shower response model ($R_X \,|\, \xi, \langle \xi \rangle$). Light grey bands depict the spread in the hadronic-interaction-model predictions of each moment. This figure was produced with the library of 10^5 CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$.

We also quantify biases in the mean and variance of $\Delta X_{\max}$, and in the exponential slope parameter Λ_η extracted from a likelihood fit to the 20% deepest showers [56]. These biases are shown in Figure 4.20, which also depicts, in grey bands, the spread between the predictions of the hadronic interaction models for the moments of the distribution of $\Delta X_{\max}$.

The spread of these biases across interaction models defines the systematic uncertainty in predicting moments of $f(\Delta X_{\max})$. For the mean, the uncertainty is below $8\,\mathrm{g\,cm^{-2}}$, or about 20% of the inter-model spread. When using only the most up-to-date models, the uncertainty decreases to $4\,\mathrm{g\,cm^{-2}}$. Therefore, applying the universal shower response model to a physically consistent prior for the primary proton-air interaction is expected to reproduce the resulting $\langle \Delta X_{\max} \rangle$ with sufficient precision to discriminate between the assumptions underlying different hadronic interaction models. The uncertainties in $\sigma(\Delta X_{\max})$ and Λ_η are $\sim 4\,\mathrm{g\,cm^{-2}}$, which are comparable to inter-model differences. Despite not allowing for model discrimination, the determination of higher-order moments of the distribution of $\Delta X_{\max}$ with reduced biases shows that its PDF is being accurately described.

4.3.2 Prediction of $X_{\max}$ from ξ

Thanks to several measurements of the proton–proton and proton–air cross-sections [56, 105] at center-of-mass energies relevant for proton–air interactions at the highest energies, the average depth of the first interaction point, X_1, and the corresponding interaction length, $\lambda_{p-\mathrm{air}}$, at $E_0 = 10^{19}$ eV, exhibit a model spread of only $3\,\mathrm{g\,cm^{-2}}$. As a result, differences in the shape of the distribution of $X_{\max}$ between hadronic interaction models can be investigated in terms of $f(\xi)$ and a universal shower response.

Repeating the two-step approach described in Section 4.3.1, we estimate the distribution of $X_{\max}$ as

$$X_{\max} = \xi \otimes (R_X \,|\, \xi, \langle \xi \rangle) \otimes X_1, \tag{4.23}$$

where the calibration of $\langle X_{\max} \rangle$ from $\langle \xi \rangle$ is derived from the linear regression shown in Figure 4.21.

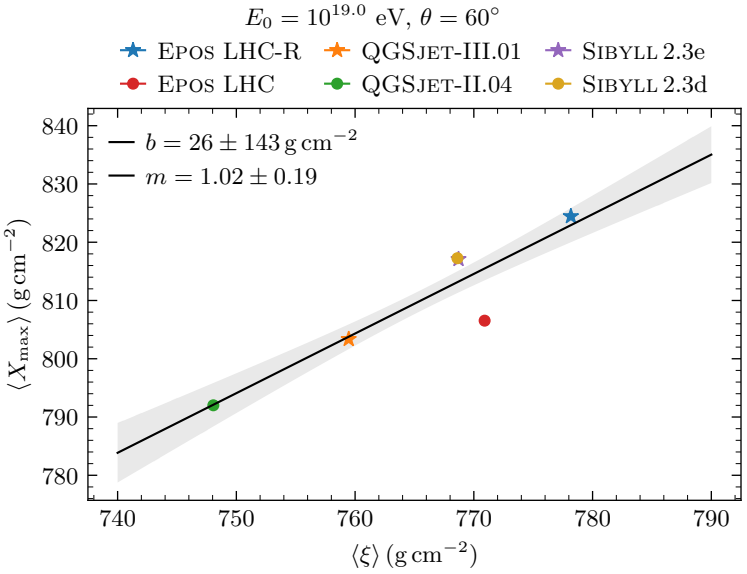

Figure 4.21: Values of $\langle \xi \rangle$ against $\langle X_{\max} \rangle$ for the hadronic interaction models EPOS LHC-R, QGSJET-III.01 and SIBYLL2.3e, EPOS LHC, QGSJET-II.04 and SIBYLL2.3d. The most up-to-date hadronic interaction models are depicted with stared markers, while their previous versions are represented by the circular markers. The black straight line results from a linear regression to the pairs of points $(\xi, X_{\max})$, while its uncertainty is represented by the shaded grey area. This figure was produced with the library of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$.

It is worth noting that Equation 4.23 inherently averages over the differences in the proton–air cross-section across models. This contributes to a slight increase in the dispersion of the model points about the calibration curve in the $\langle \xi \rangle$–$\langle X_{\max} \rangle$ plane.

The true and predicted distributions of $X_{\max}$, as obtained from Equation (4.23), are shown in Figure 4.22 for the hadronic interaction models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e.

The overall agreement between the true and predicted $X_{\max}$ distributions is remarkable, confirming the universality of the correspondence between ξ and $X_{\max}$. The biases introduced in the main moments of the $X_{\max}$ distribution by the use of the kernel $f(X_{\max} \,|\, \xi, \langle \xi \rangle)$ are presented in Figure 4.23.

The systematic uncertainty in estimating $\langle X_{\max} \rangle$ amounts to $10\,\mathrm{g\,cm^{-2}}$ when considering all hadronic interaction models, and is reduced to just $4\,\mathrm{g\,cm^{-2}}$ when restricting to the most recent model versions. For these up-to-date models, this systematic represents approximately 20% of the total spread between models, enabling discrimination based solely on differences in the energy spectrum of secondaries in the primary interaction. The corresponding biases in estimating $\sigma(X_{\max})$ and Λ_η are, at most, $3\,\mathrm{g\,cm^{-2}}$ and $4\,\mathrm{g\,cm^{-2}}$, respectively.

The strong causal relationship between ξ and $X_{\max}$, combined with the universal nature of the shower response to $f(\xi)$, implies that a measured $X_{\max}$ distribution can provide meaningful constraints on the energy spectra of secondary hadrons in kinematic regimes beyond the reach of man-made accelerators. Nevertheless, applying this framework to data is non-trivial: the mass composition of ultra-high-energy cosmic rays, as inferred through hadronic interaction models, is compatible with a mixed flux of light and heavy

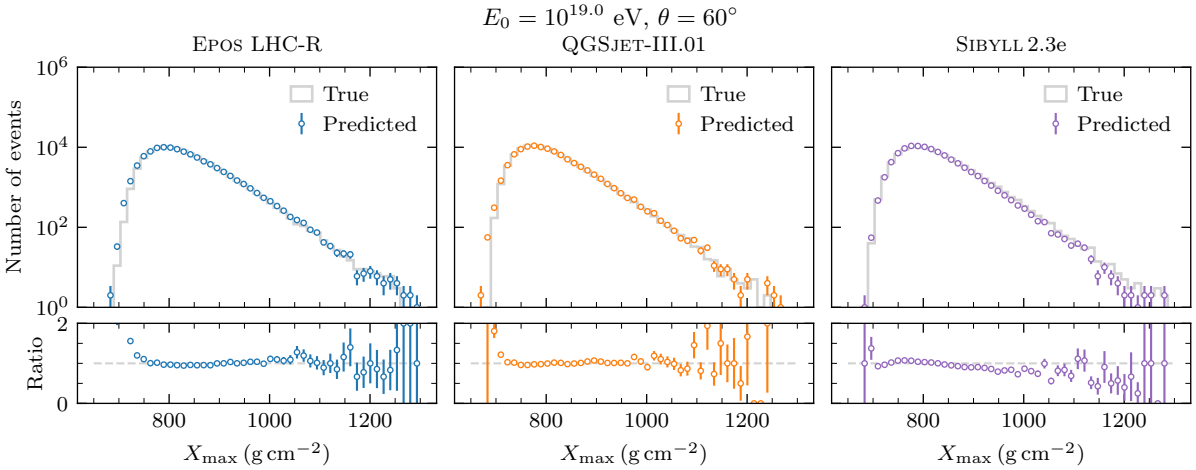

Figure 4.22: Upper panels: Distributions of $X_{\max}$ predicted by applying the universal shower response model$(R_X \,|\, \xi, \langle\xi\rangle) \otimes X_1$ to prior distributions of ξ produced with the hadronic interaction models EPOS LHC-R (left), QGSJET-III.01 (middle) and SIBYLL2.3e (right). The underlying grey histograms are true distributions of $X_{\max}$. Lower panels: ratio of the predicted to the true distributions $f(X_{\max})$. This figure was produced with the library of 10^5 CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$.

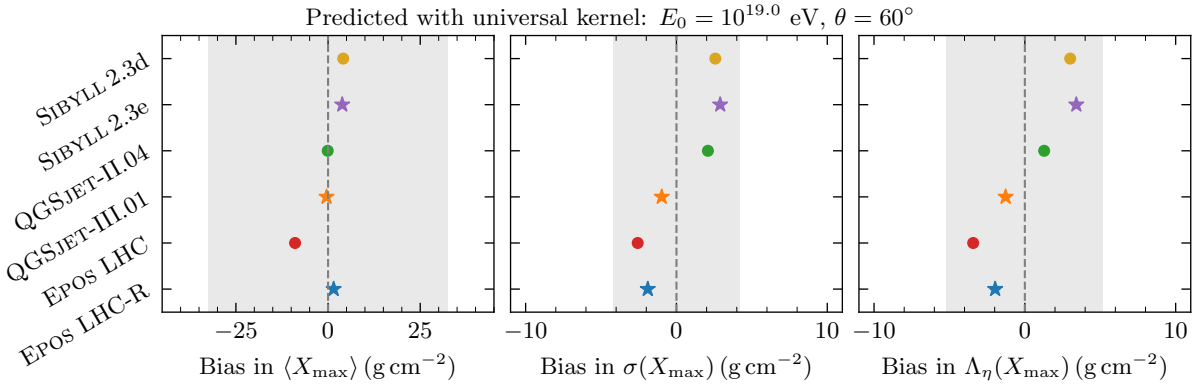

Figure 4.23: Biases in $\langle X_{\max}\rangle$ (left), $\sigma(X_{\max})$ (middle), and Λ_η (right) due to use of the universal shower response model $(R_X \,|\, \xi, \langle\xi\rangle) \otimes X_1$. Light grey bands depict the spread in the hadronic-interaction-model predictions of each moment. This figure was produced with the library of 10^5 CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$.

nuclei [140]. As shown in Section 2.2.3, the shape of the $X_{\max}$ distribution is significantly influenced by heavier primaries.

One potential strategy to address this challenge is to exploit correlations between shower observables, such as $X_{\max}$ and the number of muons N_μ, and constrain the analysis to regions of their joint distribution that are dominated by proton primaries. This subset of events, typically characterised by large $X_{\max}$ values and low muon content, allows the framework to be applied without modification. In Chapter 5 we take the first steps towards this outlined strategy.

Alternatively, the ξ–$X_{\max}$ correspondence could be generalised to heavier nuclei. This would require a careful probabilistic treatment of nuclear fragmentation and a dedicated assessment of the sensitivity of the $X_{\max}$ distribution to collective hadronisation effects

in the initial stages of the interaction. Nonetheless, this work lays the foundation for a phenomenological approach to probing the properties of the primary interaction via the full distributions of EAS observables.

4.4 New multiparticle production variables in extensive air showers

4.4.1 Interpretation of the new multiparticle production variables

The derivation of the functional form of ξ led to the introduction of new multiparticle production variables, namely ζ_{had} and ζ_{EM}, defined as:

$$\zeta_{\text{had}} \equiv -\sum_{i=1}^{m_{\text{had}}} x_i \ln x_i, \quad \zeta_{\text{EM}} \equiv -\sum_{j=1}^{m_{\text{EM}}} x_j \ln x_j. \tag{4.24}$$

These variables quantify the contribution of each secondary particle from the primary interaction to the overall electromagnetic longitudinal profile of the resulting cascade. Insight into their connection with the energy spectra of secondaries can be gained by examining their kinematic bounds. Consider ζ_{had}, a m_{had}-dimensional function of the energy fractions x_i, in the laboratory frame, of hadronically interacting particles: $\zeta_{\text{had}} = \zeta(x_1, x_2, \ldots, x_{m_{\text{had}}})$. For a fixed value of α_{had}, this function can be maximised under the energy conservation constraint $\alpha_{\text{had}} = \sum_i x_i$. As this constraint defines a m_{had}-dimensional manifold, the maximisation can be performed using Lagrange multipliers, yielding the upper bound:

$$\zeta_{\text{had}} \leq \alpha_{\text{had}} \ln\left(\frac{m_{\text{had}}}{\alpha_{\text{had}}}\right), \tag{4.25}$$

attained for $x_i = m_{\text{had}}^{-1}$, that is, for energy equipartition. The case $m_{\text{had}} = 1$ and $\alpha_{\text{had}} \neq 1$ provides a lower bound for ζ_{had}. Applying the same reasoning to ζ_{EM} leads to the following bounds:

$$\begin{aligned} -\alpha_{\text{had}} \ln \alpha_{\text{had}} \leq \zeta_{\text{had}} \leq \alpha_{\text{had}} \ln\left(\frac{m_{\text{had}}}{\alpha_{\text{had}}}\right), \\ -(1-\alpha_{\text{had}}) \ln(1-\alpha_{\text{had}}) \leq \zeta_{\text{EM}} \leq (1-\alpha_{\text{had}}) \ln\left(\frac{m_{\text{EM}}}{1-\alpha_{\text{had}}}\right). \end{aligned} \tag{4.26}$$

The lower bounds correspond to $m_{\text{had}} = 1$ and $m_{\text{EM}} = 1$ in each respective sector. If energy is instead equally partitioned among all secondaries regardless of sector, then:

$$0 \leq \zeta_{\text{EM}} + \zeta_{\text{had}} \leq \ln m_{\text{tot}}. \tag{4.27}$$

Since ζ_{had} and ζ_{EM} appear with negative coefficients in the definition of ξ, energy equipartition, for fixed α_{had}, minimises ξ and, through its causal connection with $\Delta X_{\max}$, also tends to minimise the latter. This is physically intuitive: with energy equipartition among secondaries, the primary energy is rapidly degraded, producing shallower showers.

Within these bounds, ζ_{had} and ζ_{EM} are sensitive to both the multiplicity and asymmetry in the energy distribution among secondary particles. This is illustrated in Figure 4.24, which shows the energy spectra of hadrons (left) and EM sector particles (right)

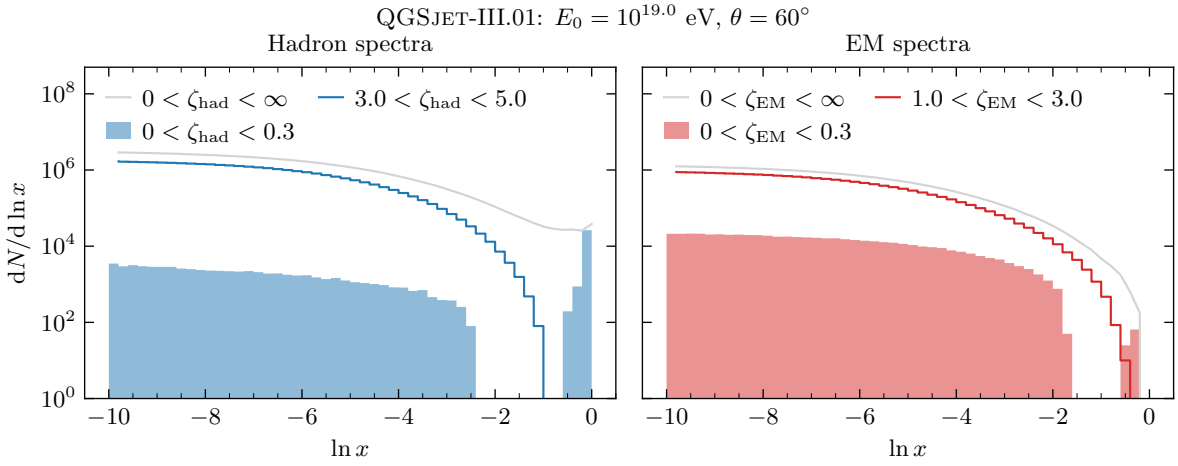

Figure 4.24: Left: energy spectra of hadrons from proton-air interactions at $E_0 = 10^{19}$ eV in the lab frame, for different ζ_{had} ranges. Right: the same for EM particles for ranges of ζ_{EM} values.

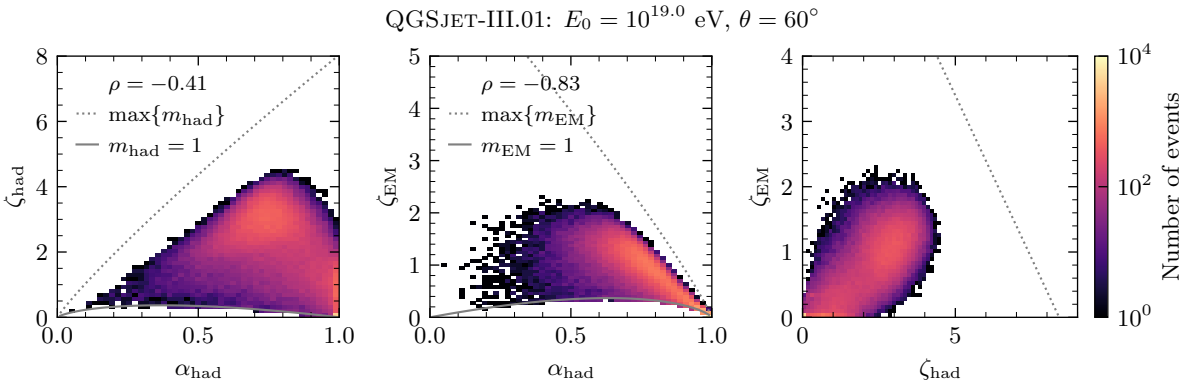

Figure 4.25: Joint distributions of ζ_{had} and α_{had} (left), ζ_{EM} and α_{had} (middle), and ζ_{had} and ζ_{EM} (right), including kinematic bounds.

for different ranges of ζ_{had} and ζ_{EM}, respectively.

A more asymmetric energy distribution among hadrons, as in diffractive events, leads to lower ζ_{had} values. In such cases, the leading particle carries a significant energy fraction, and, by energy conservation, the remaining secondaries carry little, resulting in a visible gap in the left panel of Figure 4.24. In contrast, highly inelastic interactions generate a larger number of secondaries with smaller energy fractions, leading to softer spectra and larger ζ_{had}. Hence, ζ_{had} is closely related to the degree of hadronic activity of a particular interaction, as defined in Chapter 3. A similar interpretation applies to ζ_{EM} and the spectrum of neutral pions.

The joint distributions $f(\zeta_{had}, \alpha_{had})$, $f(\zeta_{EM}, \alpha_{had})$, and $f(\zeta_{had}, \zeta_{EM})$ are shown in Figure 4.25 for the QGSJET-III.01 model. Kinematic bounds from Equation (4.26) are indicated with grey lines.

All multiparticle production variables are, by construction, different moments of the energy spectrum of secondaries, and are therefore mildly correlated. In fact, their joint distribution exhibits a complex structure, shaped by the interplay between elasticity and multiplicity, as well as by the kinematic bounds discussed above, which themselves depend

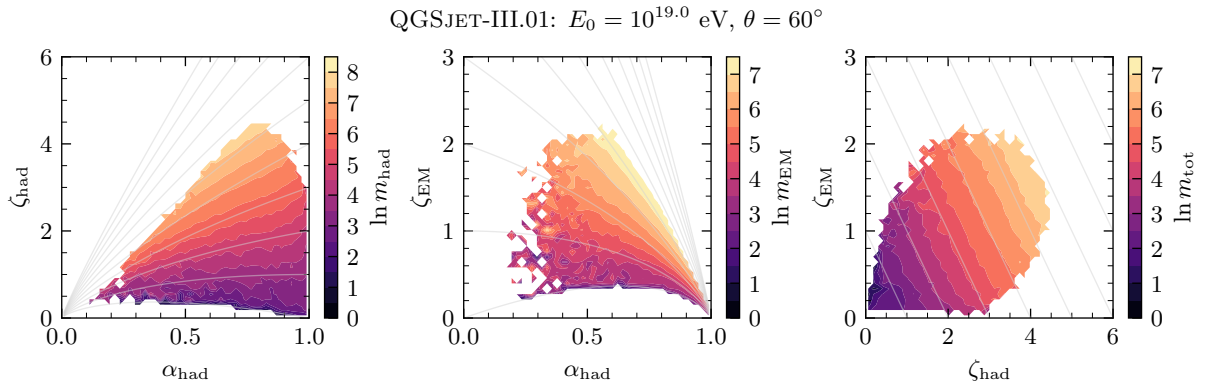

Figure 4.26: Same as Figure 4.25, weighted by $\ln m_{\rm had}$ (left), $\ln m_{\rm EM}$ (middle), and $\ln m_{\rm tot}$ (right). See the main text for further details.

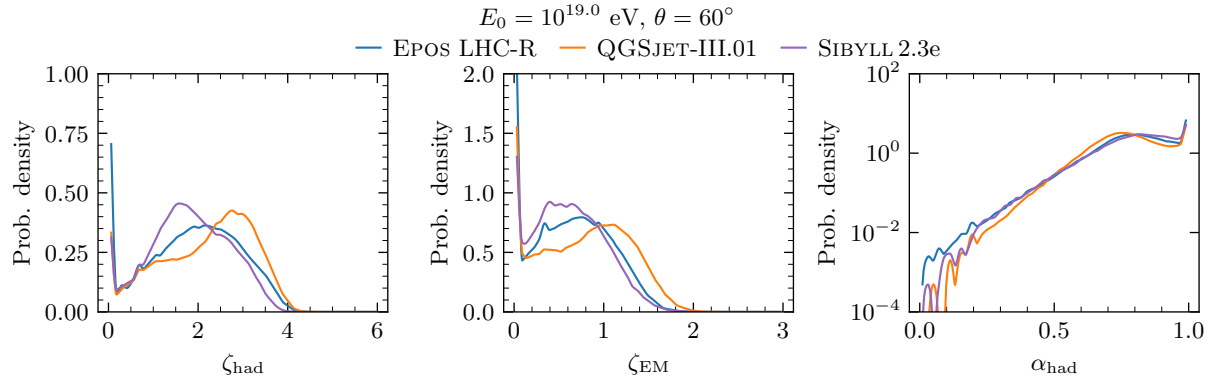

Figure 4.27: Distributions of $\zeta_{\rm had}$ (left), $\zeta_{\rm EM}$ (middle), and $\alpha_{\rm had}$ (right) for EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e models.

on these quantities. Weighting the joint distributions by the multiplicity of secondary particles accentuates regions associated with highly elastic or highly inelastic interactions, as illustrated in Figure 4.26. The overlaid grids correspond to the kinematic limits given by Equations (4.26) and (4.27), evaluated for different values of multiplicity.

As shown, the lowest and highest values of $\zeta_{\rm had}$ and $\zeta_{\rm EM}$ for each fixed $\alpha_{\rm had}$ correspond to the lowest and highest multiplicities, respectively.

The marginal distributions of $\zeta_{\rm had}$, $\zeta_{\rm EM}$, and $\alpha_{\rm had}$ from CONEX proton showers are displayed in Figure 4.27, for the EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e models.

The distribution of $\zeta_{\rm had}$ corresponds to the $m_{\rm had}$-fold weighted convolution of the energy spectra of hadronically interacting secondaries. As previously noted, higher $\zeta_{\rm had}$ values correspond to interactions with greater hadronic activity and higher multiplicity. Therefore, in the high multiplicity limit, the distribution of $\zeta_{\rm had}$ approaches a normal distribution due to the central limit theorem. The same reasoning applies to $\zeta_{\rm EM}$. Conversely, lower values of $\zeta_{\rm had}$ and $\zeta_{\rm EM}$ are associated with primary interactions with higher elasticity. The peaks of $f(\zeta_{\rm had})$ and $f(\zeta_{\rm EM})$ at zero are due to quasi-elastic proton-air scattering, and are the most prominent for the hadronic interaction model EPOS LHC-R.

Importantly, the shapes of the $\zeta_{\rm had}$ and $\zeta_{\rm EM}$ distributions are strongly dependent on the hadronic interaction model, reflecting differences in the treatment and parameterisation of secondary particle production. Constraining the shapes of these distributions thus provides a powerful tool for hadronic-interaction-model discrimination. The precise rela-

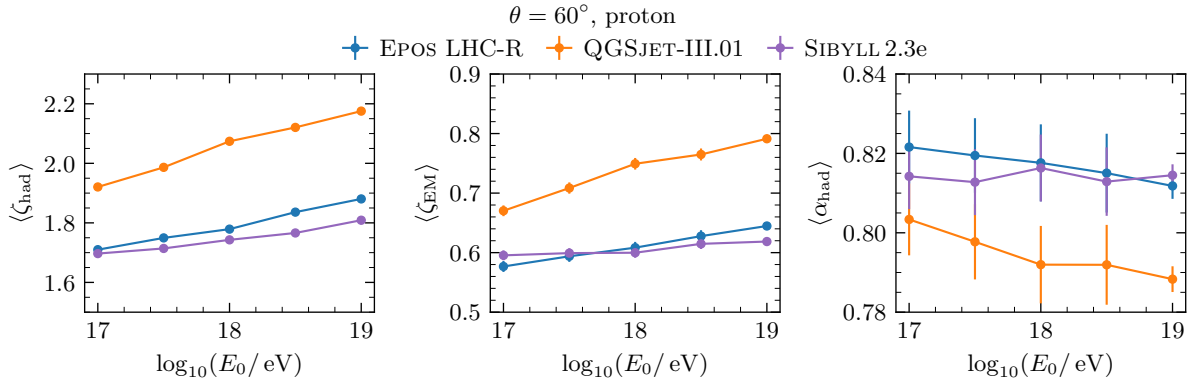

Figure 4.28: Energy dependence of the mean values of ζ_{had} (left), ζ_{EM} (middle), and α_{had} (right), based on libraries of CONEX simulations of proton-induced showers at various primary energies and a zenith angle $\theta = 60°$, using the high-energy hadronic interaction models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e.

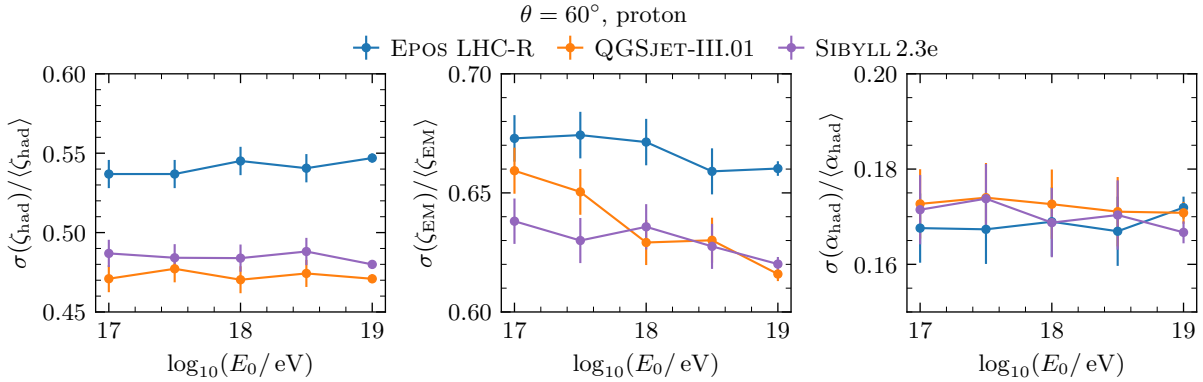

Figure 4.29: Energy dependence of the relative fluctuations of ζ_{had} (left), ζ_{EM} (middle), and α_{had} (right), for proton-induced showers simulated with CONEX at $\theta = 60°$, using the hadronic interaction models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e.

tion between the distribution shapes and the energy spectra of secondaries is currently under investigation. Finally, the properties of α_{had} are extensively discussed in [53, 106, 133].

As a final remark, we note that the production variables described in this section are strongly correlated with α_1, and consequently with fluctuations in N_μ. Their use in characterising regions of the joint distribution $f(N_\mu, X_{\mathrm{max}})$ is the subject of Section 5.1.

4.4.2 Energy dependence of new multiparticle production variables

The energy dependence of the mean values and relative fluctuations of the production variables ζ_{had}, ζ_{EM}, and α_{had} for proton-air interactions is shown in Figures 4.28 and 4.29, respectively, for different hadronic interaction models.

Regardless of the hadronic interaction model, the mean values of ζ_{had} and ζ_{EM} increase approximately linearly with the primary energy. This behaviour can be understood by recalling that the upper bounds of these variables are given by $\alpha_{\mathrm{had}} \ln(m_{\mathrm{had}}/\alpha_{\mathrm{had}})$ and $\alpha_{\mathrm{EM}} \ln(m_{\mathrm{EM}}/\alpha_{\mathrm{EM}})$, respectively, and that multiplicity increases approximately as a power

law with the projectile energy [42]. Moreover, both the absolute normalisations of $\langle \zeta_{\rm had} \rangle$ and $\langle \zeta_{\rm EM} \rangle$, as well as the slopes of their energy evolution, exhibit a strong dependence on the hadronic interaction model. This indicates that these observables are highly sensitive to the energy spectra of secondaries produced in hadronic interactions, and thus encode valuable information about the physics of hadron production.

In contrast, the evolution of $\langle \alpha_{\rm had} \rangle$ with energy varies between models. Both EPOS LHC-R and QGSJET-III.01 predict a decreasing trend in the mean fraction of energy transferred to hadronically interacting particles as the energy increases, whereas SIBYLL2.3e yields a nearly constant value.

The relative fluctuations of $\zeta_{\rm had}$ remain approximately constant across the entire energy range, though they are highly model-dependent. In contrast, the relative fluctuations of $\zeta_{\rm EM}$ tend to decrease with increasing primary energy, with the rate of decrease also depending on the hadronic interaction model. Interestingly, the relative fluctuations of $\alpha_{\rm had}$ show no significant energy dependence in any of the models considered. Understanding these trends from a microscopic point of view is a matter for future work.

Overall, the energy evolution of the main moments of $\zeta_{\rm had}$, $\zeta_{\rm EM}$, and $\alpha_{\rm had}$ exhibits a strong dependence on the chosen hadronic interaction model, making apparent their sensitivity to the specific mechanisms and assumptions implemented in each model. As such, these variables can provide valuable constraints on hadronic production across a broad range of primary energies.

4.4.3 Interpretation of new production variables in terms of kinematic variables

We can also express the variables $\zeta_{\rm EM}$ and $\zeta_{\rm had}$ in terms of the pseudorapidity η_i of each secondary particle i. If $\mathbf{p}_i$ denotes the momentum of particle i and $p_{\parallel,i}$ its longitudinal component, then the pseudorapidity is given by

$$\eta_i = \frac{1}{2} \ln \left(\frac{|\mathbf{p}_i| + p_{\parallel,i}}{|\mathbf{p}_i| - p_{\parallel,i}} \right). \tag{4.28}$$

Since we are interested in the interactions that drive the development of air showers, we restrict our derivation to forward-going particles, i.e., those with $p_\parallel \gg 0$ in the laboratory frame. If $p_{\perp,i}$ denotes the transverse momentum of particle i, then under the ultra-relativistic and small-angle approximations $E_i \simeq |\mathbf{p}_i| \simeq p_{\parallel,i}$, one finds

$$\eta_i \simeq \ln \left(\frac{2E_0}{p_{\perp,i}} \right) + \ln x_i, \tag{4.29}$$

from which it follows that

$$\begin{aligned} \zeta_{\rm had} &\simeq - \sum_{i=1}^{m_{\rm had}} x_i \eta_i + \sum_{i=1}^{m_{\rm had}} x_i \ln \left(\frac{2E_0}{p_{\perp,i}} \right), \\ \zeta_{\rm EM} &\simeq - \sum_{j=1}^{m_{\rm EM}} x_j \eta_j + \sum_{j=1}^{m_{\rm EM}} x_j \ln \left(\frac{2E_0}{p_{\perp,j}} \right). \end{aligned} \tag{4.30}$$

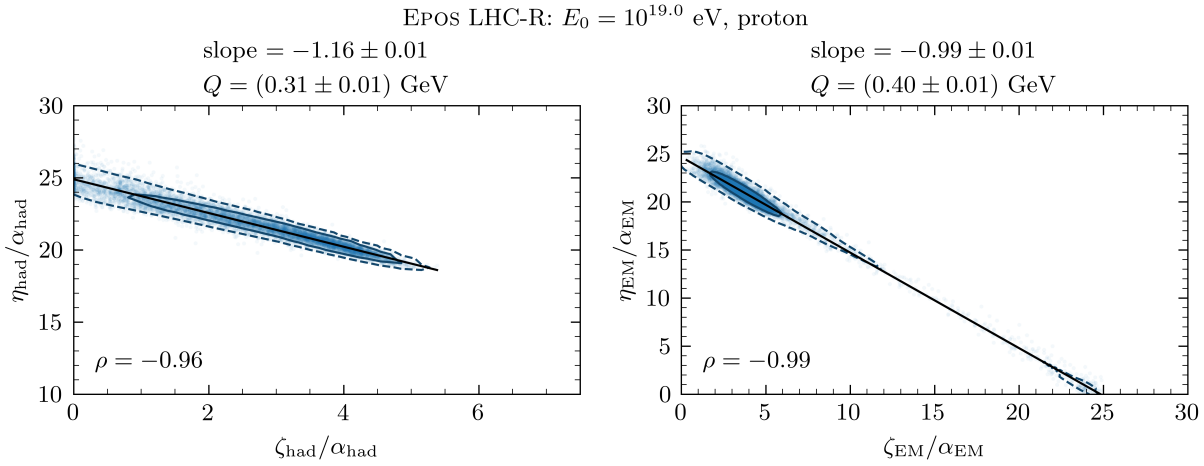

Figure 4.30: Left: joint distribution of $\eta_{\rm had}/\alpha_{\rm had}$ and $\zeta_{\rm had}/\alpha_{\rm had}$. Right: joint distribution of $\eta_{\rm EM}/\alpha_{\rm EM}$ and $\zeta_{\rm EM}/\alpha_{\rm EM}$. Both panels show contours containing 68% and 95% of the events (solid and dashed dark blue lines, respectively), along with a linear regression fit (solid black line). The slope m and transverse momentum scale Q are extracted from the fit.

Neglecting fluctuations in $p_{\perp,i}$ relative to those in the energy fractions x_i, we set $p_{\perp,i} \sim Q$, where Q denotes the characteristic transverse momentum scale of the interaction. This yields the approximations

$$\begin{aligned} \zeta_{\rm had} &\simeq -\sum_{i=1}^{m_{\rm had}} x_i \eta_i + \alpha_{\rm had} \eta_{\rm beam}, \\ \zeta_{\rm EM} &\simeq -\sum_{j=1}^{m_{\rm EM}} x_j \eta_j + (1-\alpha_{\rm had}) \eta_{\rm beam}, \end{aligned} \tag{4.31}$$

where $\eta_{\rm beam} = \ln(2E_0/Q)$ denotes the pseudorapidity characteristic of the incident proton. Defining $\eta_{\rm had} = -\sum_i x_i \eta_i$ and similarly for the electromagnetic sector, we obtain

$$\begin{aligned} \zeta_{\rm had} &\simeq -\eta_{\rm had} + \alpha_{\rm had} \eta_{\rm beam}, \\ \zeta_{\rm EM} &\simeq -\eta_{\rm EM} + (1-\alpha_{\rm had}) \eta_{\rm beam}. \end{aligned} \tag{4.32}$$

The validity of these approximations is confirmed by the distributions shown in Figure 4.30, from which the appropriate transverse momentum, Q, was extracted.

We have verified that the Pearson correlation coefficients between $\zeta_{\rm had}/\alpha_{\rm had}$ and $\eta_{\rm had}/\alpha_{\rm had}$ satisfy $\rho < -0.95$, independently of the hadronic interaction model. An analogous result holds for the electromagnetic component. Moreover, the transverse-momentum scales extracted from the regression, $Q \sim 150$–$250\,\text{MeV}$, are consistent with typical laboratory values of $p_\perp$ of secondaries of such proton-air interactions [109, 141]. The observed deviation of the slope from unity likely reflects the presence of a mixture of hadronically interacting particles contributing to $\zeta_{\rm had}$ and $\eta_{\rm had}$.

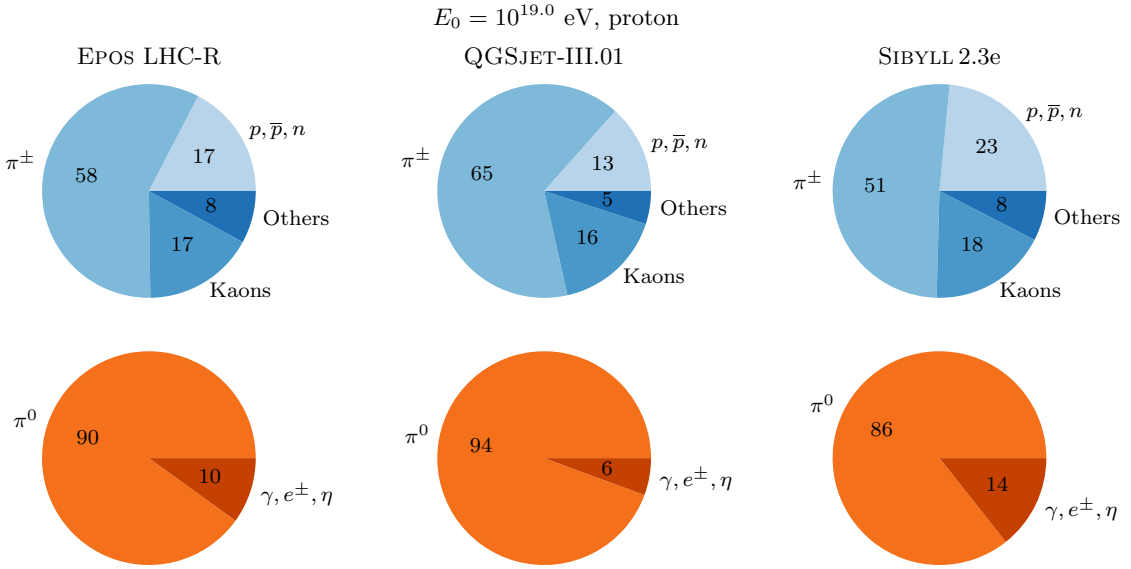

Figure 4.31: Average contribution of different secondary particles to ζ_{had} (upper panels) and ζ_{EM} (lower panels) for proton–air interactions at $E_0 = 10^{19}$ eV, simulated using the hadronic interaction models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e. Secondary particles are grouped into sectors: $p, \overline{p}, n$ (protons, anti-protons, and neutrons), kaons ($K^{\pm}$, K^0_L, K^0_S), $\pi^{\pm}$, and other baryons — which compose the hadronic sector — as well as π^0, and the electromagnetic group consisting of γ, $e^{\pm}$, and η.

4.4.4 Contribution of different particle sectors to ζ_{had} and ζ_{EM}

It is also instructive to examine the average contribution of different particle sectors to the production variables ζ_{had} and ζ_{EM}. For ζ_{had}, we consider three hadronic groups: the proton group (protons, anti-protons, and neutrons), the kaon group ($K^{\pm}$, K^0_L, and K^0_S), and the charged pion group. For ζ_{EM}, we include contributions from neutral pions and from the group consisting of photons, electrons, positrons, and η mesons.

The fractional contributions of each particle group to ζ_{had} and ζ_{EM} are shown in the upper and lower panels, respectively, of Figure 4.31, for three different hadronic interaction models.

Regardless of the hadronic interaction model, the dominant contribution to ζ_{had} comes from charged pions. This is likely due to their high multiplicity (recall that ζ_{had} is bounded above by a logarithmic function of the multiplicity) and the fact that these particles typically carry similar fractions of the primary energy. By contrast, despite the significant fraction of the primary energy carried by protons and neutrons in *quasi*-elastic interactions, the contribution of such collisions to ζ_{had} is negligible. The frequency of such *quasi*-elastic events is around 5%, effectively suppressing the average contribution of nucleons to ζ_{had}. As expected, neutral pions dominate the contribution to ζ_{EM}, highlighting the importance of accurately constraining their energy spectra to describe the distribution of ζ_{EM}.

Notably, the relative contributions of the different particle groups vary significantly across hadronic interaction models, reflecting the specific modelling of hadron production processes in each case.

Beyond mean values, one can also decompose the interaction-by-interaction distri-

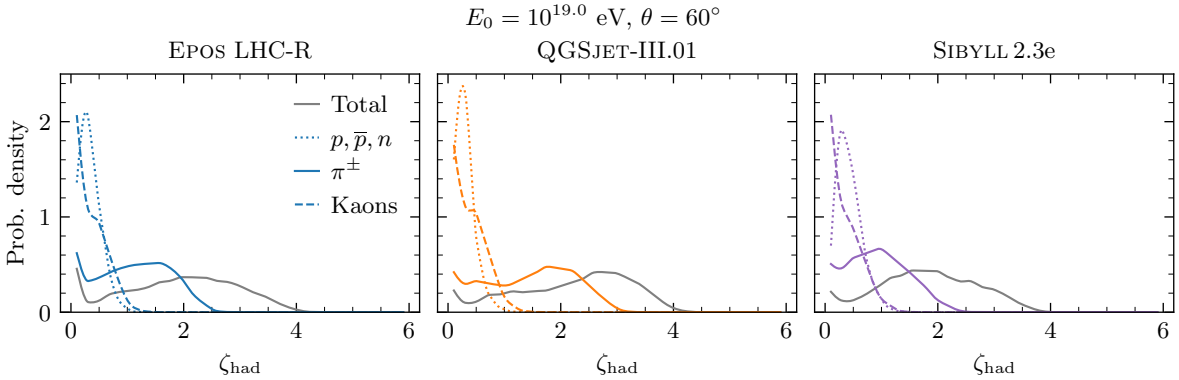

Figure 4.32: Distributions of ζ for different particle sectors, as defined in the opening paragraph of Section 4.4.4. The total distribution $f(\zeta_{\mathrm{had}})$ is labelled as 'Total'. Each panel corresponds to CONEX simulations of proton-induced showers using a different hadronic interaction model.

bution $f(\zeta_{\mathrm{had}})$ into contributions from the various particle sectors. This decomposition is valuable because the distributions $f(\zeta_k)$ for each group k are determined by the corresponding energy spectra, which may be measurable in accelerator experiments within specific regions of phase space. The decomposition follows directly from the additive structure of ζ_{had}:

$$\zeta_{\mathrm{had}} = -\sum_{i=1}^{m_{\mathrm{had}}} x_i \ln x_i = -\sum_{k=1}^{S}\sum_{i=1}^{m_k} x_i^{(k)} \ln x_i^{(k)} = \sum_{k=1}^{S} \zeta_k \quad \text{with} \quad \zeta_k \equiv -\sum_{i=1}^{m_k} x_i^{(k)} \ln x_i^{(k)}, \tag{4.33}$$

where k indexes the particle sector, m_k denotes the number of hadrons in sector k produced in a given interaction, S is the total number of sectors. If we neglect correlations between the ζ_k values across different sectors, the overall distribution $f(\zeta_{\mathrm{had}})$ can be expressed as an S-fold convolution:

$$f(\zeta_{\mathrm{had}}) = \bigotimes_{k=1}^{S} f(\zeta_k). \tag{4.34}$$

The distributions of ζ for the proton, kaon, and charged pion sectors are shown in Figure 4.32, together with the total distribution $f(\zeta_{\mathrm{had}})$, for three hadronic interaction models.

The overall shape of $f(\zeta_{\mathrm{had}})$ is largely determined by the energy spectrum and multiplicity of charged pions, encoded in the distribution $f(\zeta_{\pi^{\pm}})$. Consequently, a precise measurement of the charged pion energy spectrum is essential for constraining the shape of $f(\zeta_{\mathrm{had}})$ and, through the probabilistic mapping $\xi \to X_{\mathrm{max}}$, also the shape of $f(X_{\mathrm{max}})$. Section 4.5 discusses accelerator experiments capable of measuring hadron spectra in relevant regions of phase space for the development of extensive air showers.

4.5 New multiparticle production variables in accelerator experiments

The multiparticle production variables ζ_{had}, ζ_{EM} and α_{had} are constructed directly from the energy spectra of secondary particles of hadronic interactions. So, their values per

interaction can be readily computed in accelerator experiments, in the kinematic phase space covered by a particular detector. In particular, the proton-oxygen runs in Run 3 of the LHC [142] could provide important information to constrain these variables in proton-air interactions in the atmosphere.

We begin by identifying the regions of the kinematic phase space that contribute most significantly to each of the production variables. The variable $\alpha_{\rm had}$ is defined as the fraction of energy carried by hadronically interacting particles, excluding photons, electrons, positrons, π^0, and η mesons. Consequently, the region of maximal contribution to $\alpha_{\rm had}$ coincides with the peak of the differential energy flow, $\mathrm{d}E/\mathrm{d}\eta$. In practice, this energy flow is constructed by weighting the pseudorapidity distribution of hadrons by their fractional energy, $w_i = x_i$, in the lab. frame.

Analogously, the contribution of each hadronically interacting particle to $\zeta_{\rm had}$ is measured by the weight $w_i = -x_i \ln x_i$. To determine the region of phase space where the particles contributing most to $\zeta_{\rm had}$ are produced, we introduce the concept of the differential ζ-flow, defined as

$$\text{differential } \zeta\text{-flow} \equiv -\frac{\mathrm{d}}{\mathrm{d}\eta}\left(\sum_i x_i \ln x_i\right). \tag{4.35}$$

The same definitions apply to the electromagnetic component, specifically to the variables $1-\alpha_{\rm had}$ and $\zeta_{\rm EM}$.

The left panel of Figure 4.33 shows the energy and ζ flows into particles of the hadronic sector, while the right panel shows the flows into the EM sector, for proton-air interactions at $E_0 = 10^{17}\,\mathrm{eV}$, in the lab. frame. In the limit where the incident proton is ultra-relativistic, this energy corresponds to the nucleon-nucleon center-of-mass energy

$$s_{\rm NN} \simeq 2M_N E_0\left(1+\frac{M_p^2}{2E_0M_N}+\frac{M_N}{2E_0}\right) \implies \sqrt{s_{\rm NN}} \simeq \sqrt{2E_0M_N} \simeq 14\,\mathrm{TeV}, \tag{4.36}$$

where $M_N \simeq 938.92\,\mathrm{MeV}/c^2$ denotes the average nucleon mass and $M_p = 938.27\,\mathrm{MeV}/c^2$ denotes the mass of the proton [40]. Moreover, Figure 4.33 shows grey shaded bands indicating the pseudorapidity range covered by the particle detectors CMS, $\eta < 2$ [59], and LHCb, $2 < \eta < 5$ [60]. The panel pertaining to the EM sector displays an additional band showing the pseudo-rapidities covered by the LHCf detector [61], able to measure the spectra of neutral pions and neutrons in the pseudorapidity region $8.3 < \eta < 13$ [134, 135]. The total contribution of hadrons to $\alpha_{\rm had}$ and $\zeta_{\rm had}$ measurable by each detector is obtained by integrating the energy and ζ-flow, respective, over the covered pseudorapidity range. The normalized contributions are also shown in the same figure in roman and bold fonts, respectively.

The forward region of the kinematic phase-space contributes the most for the values of $\zeta_{\rm had}$ and $\alpha_{\rm had}$. The energy flow to the hadronic sector peaks at higher rapidities than the ζ flow. This is expected since the extreme pseudorapidity regions are populated by particles produced in *quasi*-elastic and diffractive interactions, where $\zeta_{\rm had} \to 0$ and $\alpha_{\rm had} \to 1$. Both peaks lie outside the region covered by current detectors. However, they lie inside the pseudorapidity region covered by the detectors of Forward Physics Facility (FPF) [143, 144]: $\eta > 7.2$. The FASER and FLARE experiments of this facility

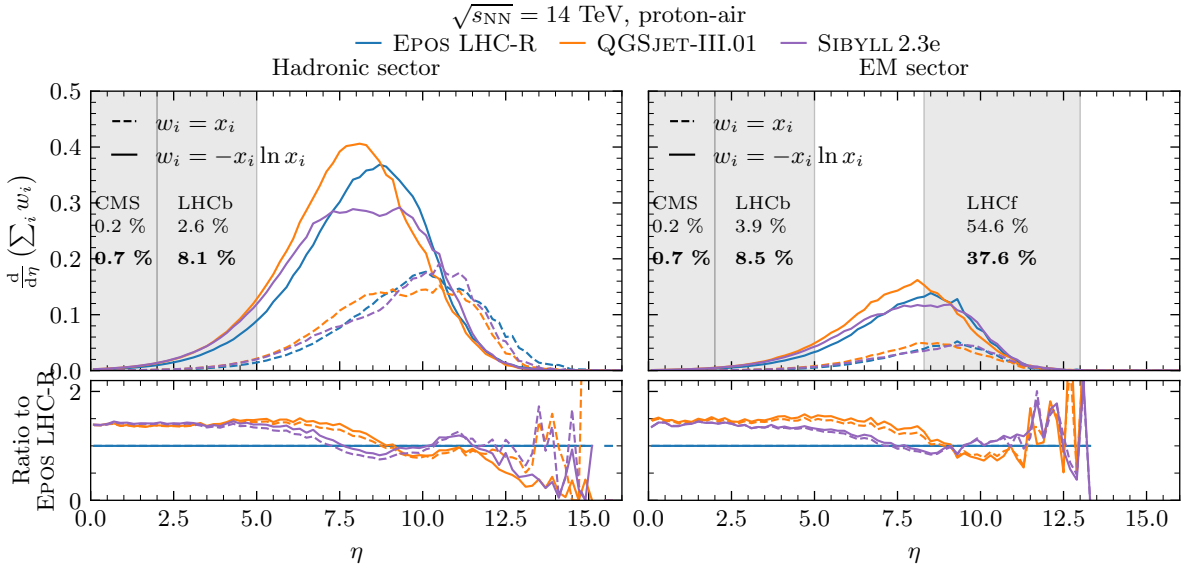

Figure 4.33: Upper panels: Contribution of each hadronically interacting particle to the values of α_{had} and ζ_{had} (left), and of each particle feeding the electromagnetic sector to α_{EM} and ζ_{EM} (right), as a function of pseudorapidity, for different hadronic interaction models. Lower panels: Ratios of the energy and ζ-flows predicted by each model to those predicted by EPOS LHC-R. Proton–air interactions were simulated with $E_0 = 10^{17}$ eV, corresponding to $\sqrt{s_{\text{NN}}} = 14$ TeV. The shaded grey bands indicate the pseudorapidity coverage of the CMS, LHCb, and LHCf detectors. The percentages in roman and bold font denote the fractions of energy and ζ-flows, respectively, covered by each detector.

are able to probe the energy spectrum of charged pions and kaons via their decays into neutrinos. Since these charged particles contribute, on average, to $\sim 60\%$ and $\sim 20\%$ of the values of ζ_{had}, measuring their spectra would constrain ζ_{had} in the kinematic region where particles most contribute to ζ_{had}. Above $\eta = 5$, the shapes of the energy and ζ-flows are highly dependent on the hadronic interaction model, especially at the highest rapidities, where accelerator data do not constrain the models. In particular, recent results from FASER [100] have already showed that hadronic interaction models tend to underestimate the rate of neutrino interactions from charged pion decays and overestimate that of neutrinos from kaon decays, in the pseudorapidity range $\eta > 8.5$.

Most features of the ζ-flow for the hadronic component are also present in the ζ-flow for the EM component. About 40% of the integrated ζ-flow can be measured at the LHCf, which additionally covers 55% of the primary energy flowing into forward neutral pions. In this far-forward pseudorapidity range, the ζ-flow is model-dependent in the lower pseudorapidity range covered by LHCf. However, better discrimination would be achieved for higher centre-of-mass energies in human-made accelerators, as shown in Figure 4.34. In this figure, the incident's proton energy is increased to $E_0 = 10^{18.7}$ eV to reach $\sqrt{s_{\text{NN}}} = 97$ TeV, the approximate center-of-mass energy proposed for the Future Circular Collider (FCC) [145]. Alternatively, such discrimination is already possible using ultra-high-energy proton-air interactions produced by cosmic rays. Therefore, both accelerator and cosmic-ray experiments offer complementary relevant information.

The distributions of ζ_{had} and ζ_{EM} can be directly constructed from the energy spectra of secondary particles measured within individual pseudorapidity intervals. Figure 4.35

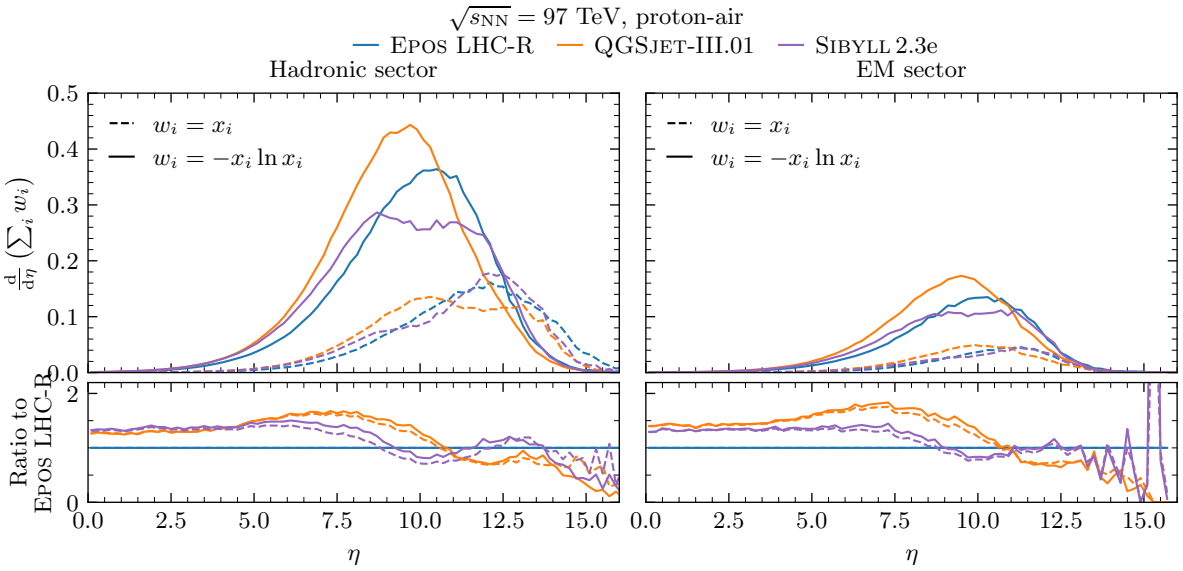

Figure 4.34: Upper panels: Contribution of each hadronically interacting particle to the values of $\alpha_{\rm had}$ and $\zeta_{\rm had}$ (left), and of each particle feeding the electromagnetic sector to $\alpha_{\rm EM}$ and $\zeta_{\rm EM}$ (right), as a function of pseudorapidity, for different hadronic interaction models. Lower panels: Ratios of the energy and ζ-flows predicted by each model to those predicted by EPOS LHC-R. Proton–air interactions were simulated with $E_0 = 10^{18.7}$ eV, corresponding to $\sqrt{s_{\rm NN}} = 97$ TeV.

presents these distributions for two nucleon–nucleon center-of-mass energies, $\sqrt{s} = 14$ TeV and $\sqrt{s} = 97$ TeV, across four pseudorapidity regions: $0 < \eta < 5$ (roughly corresponding to the combined acceptance of CMS and LHCb), $5 < \eta < 8$ (currently not covered by existing detectors), $\eta > 8$ (covered by LHCf and potentially by the Forward Physics Facility), and the full pseudorapidity range probed in proton–air interactions in extensive air showers (EAS).

As pseudorapidity and interaction energy increase, the differences between hadronic interaction models become more pronounced. This emphasises the potential of future high-energy facilities, such as the FCC, to probe particle energy spectra in the kinematic regions most relevant to EAS development. Current measurements by LHCf of the neutral pion energy spectrum in the forward region already provide valuable constraints on the shape of the $\zeta_{\rm EM}$ distribution at $\sqrt{s} = 13$ TeV. However, model discrepancies remain significant at the highest energies, where no direct experimental data are available. In the far-forward region ($\eta > 8$) at lower energies, QGSJET-III.01 and SIBYLL2.3e yield similar predictions for the $\zeta_{\rm had}$ distribution, while EPOS LHC-R differs substantially. This deviation suggests that hadron re-scattering — which modifies the correlation between mid- and forward-rapidity particle production — has a notable effect on the shape of $\zeta_{\rm had}$. Most importantly, the largest discrepancies between models appear in the pseudorapidity range accessible only through observations of extensive air showers, highlighting the critical role of cosmic-ray experiments in constraining hadronic interaction models.

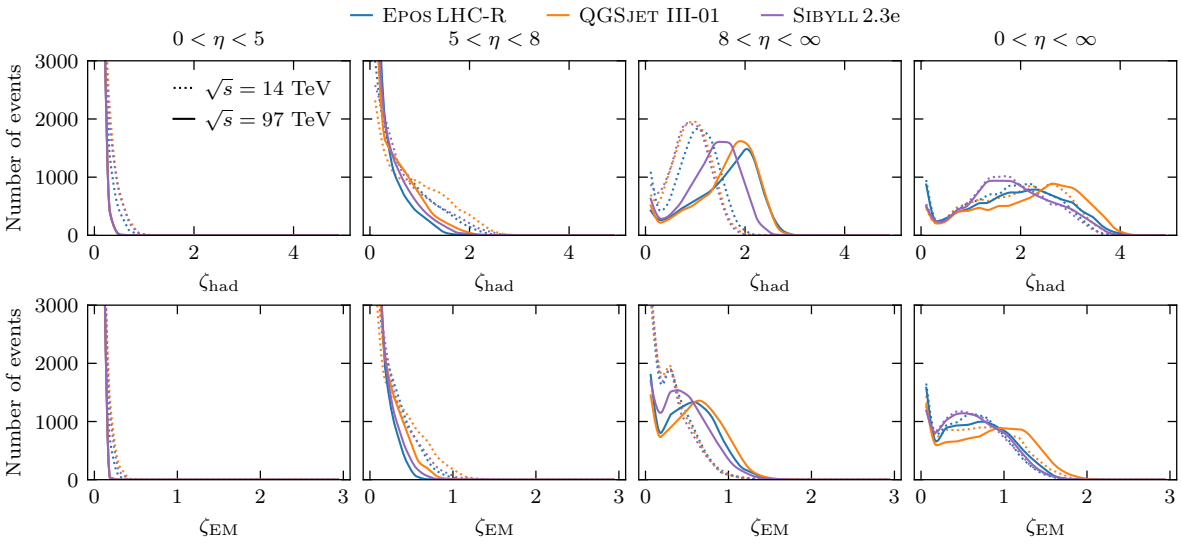

Figure 4.35: Distributions of ζ_{had} (upper panels) and ζ_{EM} (lower panels) for different pseudorapidity intervals. Dotted lines correspond to results at $\sqrt{s} = 14\,\mathrm{TeV}$ ($E_0 = 10^{17}\,\mathrm{eV}$), while solid lines correspond to $\sqrt{s} = 97\,\mathrm{TeV}$ ($E_0 = 10^{18.7}\,\mathrm{eV}$). The rightmost panels represent the pseudorapidity range covered in proton–air interactions within extensive air showers.

4.6 Summary and Discussion

In this chapter, we introduced a variable characterising the primary proton-air interaction, ξ, which governs the majority of fluctuations in ΔX_{max} of the resulting extensive air shower. This variable is defined as a linear combination of more fundamental quantities—ζ_{had}, ζ_{EM}, and α_{EM}—directly derived from the energy spectra of secondary particles produced in the first proton-air interaction.

We demonstrated that the event-by-event value of ξ provides a robust estimator of X_{max}, exhibiting minimal dependence on the hadronic interaction model. Notably, fluctuations in ξ account for approximately 50% of the total ΔX_{max} variance—remarkably close to the upper bound of 65% that would be achieved if complete information from the first interaction were retained. This confirms that fluctuations arising from subsequent shower development play only a secondary role in shaping the ΔX_{max} distribution. This behaviour was consistently observed across a wide range of primary energies.

Furthermore, we showed that extending the ξ variable to include fluctuations in the depth of the most energetic secondary interaction is essential for accurately reproducing the deep tail of the ΔX_{max} distribution, which is shaped by highly elastic cosmic-ray interactions. This opens the possibility of simultaneously constraining both the elasticity and the proton-air cross section from the tail of the X_{max} distribution.

Crucially, we developed a probabilistic framework that connects prior distributions of ξ to the observable distribution of X_{max}, using a universal model for the response function—or kernel—of the shower. In particular, the consistent variation of this kernel with changes in the secondary spectra from the primary interaction was derived from existing high-energy hadronic interaction models. This universal framework enables the reconstruction of the main moments of the X_{max} distribution with biases significantly smaller than the spread of hadronic interaction model predictions and experimental uncertain-

ties. Moreover, it lays the foundation for data-driven constraints on the secondary hadron spectra from proton-air interactions, particularly in forward kinematic regions that remain poorly constrained by accelerator data. Given the mixed composition of the cosmic-ray flux, such constraints will require either applying the framework to proton-enriched EAS samples or generalising it to heavier primaries.

We also interpreted the physical meaning of the production variables $\zeta_{\rm had}$, $\zeta_{\rm EM}$, and $\alpha_{\rm had}$, and analysed the dependence of their main moments on the primary energy. Moreover, we demonstrated that their distributions can, in principle, be extracted from measurements in different pseudorapidity regions accessible to current particle detectors. The sensitivity to the hadronic interaction model increases with both the center-of-mass energy and the pseudorapidity range of the measurement. While the energy flow from various pseudorapidity intervals is strongly model-dependent and can be evaluated using current experimental data, we found that the event-by-event distributions of $\zeta_{\rm had}$ and $\zeta_{\rm EM}$ are most distinguishable in the far-forward region—the most relevant for EAS development.

The analysis presented here contributes to the growing body of EAS-data analysis aimed at extracting information on hadronic interactions from EAS observables, as developed in [53, 56, 106, 107, 124, 133]. In particular, combining constraints from accelerator-based measurements with those inferred from $X_{\max}$ distributions offers a promising avenue for reducing the long-standing tension between air-shower simulations and observational data.

5

MAPPING THE PRIMARY INTERACTION ONTO THE JOINT DISTRIBUTION OF N_μ AND $X_{\max}$

In this chapter, we use the multiparticle production variables defined in Chapter 4—namely, $\zeta_{\rm had}$, $\zeta_{\rm EM}$, and $\alpha_{\rm had}$—to characterise the different regions of the joint distribution $f(N_\mu, X_{\max})$. This is possible because the information encoded in the energy spectra of secondaries from the primary interaction is retained in the variability of N_μ and $X_{\max}$, via their strong causal links with α_1 and ξ, respectively. These connections enable a detailed description of the shape of $f(N_\mu, X_{\max})$ in terms of the primary interaction.

Building on the formalism introduced in Chapter 4, we develop a universal shower response model that maps prior distributions of $f(\alpha_1, \xi)$, pertaining to primary proton-air interactions, onto $f(N_\mu, X_{\max})$. We assess the precision of this mapping and demonstrate that it can be made independent of the hadronic interaction model. Additionally, we quantify the biases introduced when predicting the distribution of the relative number of muons using this probabilistic response model.

We further explore the applicability of this model-independent mapping by reproducing the full joint distribution $f(N_\mu, X_{\max})$ and the evolution of Λ_μ with $X_{\max}$ (see Chapter 3), solely from the primary interaction. The promising results indicate that the formalism developed for proton showers could be applied to study this evolution using cosmic-ray data, as the $X_{\max}$-dependence of Λ_μ is only weakly sensitive to heavier primaries in the cosmic-ray flux.

This chapter lays the foundation for constraining ultra-high-energy hadronic interactions, using the full information contained in measured distributions of two key EAS observables: N_μ and $X_{\max}$.

5.1 Primary-interaction interpretation of the different regions of $f(N_\mu, X_{\max})$

The connection between the distributions of $\zeta_{\rm had}$, $\zeta_{\rm EM}$ and $\alpha_{\rm had}$ and that of $\Delta X_{\max}$ follows directly from the link between this observable and ξ. Additionally, α_1, the estimator of the relative muon content of the shower [53] is also built from the energy spectrum of secondaries of the primary interaction, thus bearing a correlation with the mentioned multiparticle production variables. Moreover, given that both α_1 and ξ account for more than 50% of the variability of N_μ and $\Delta X_{\max}$, respectively, it is clear that the information

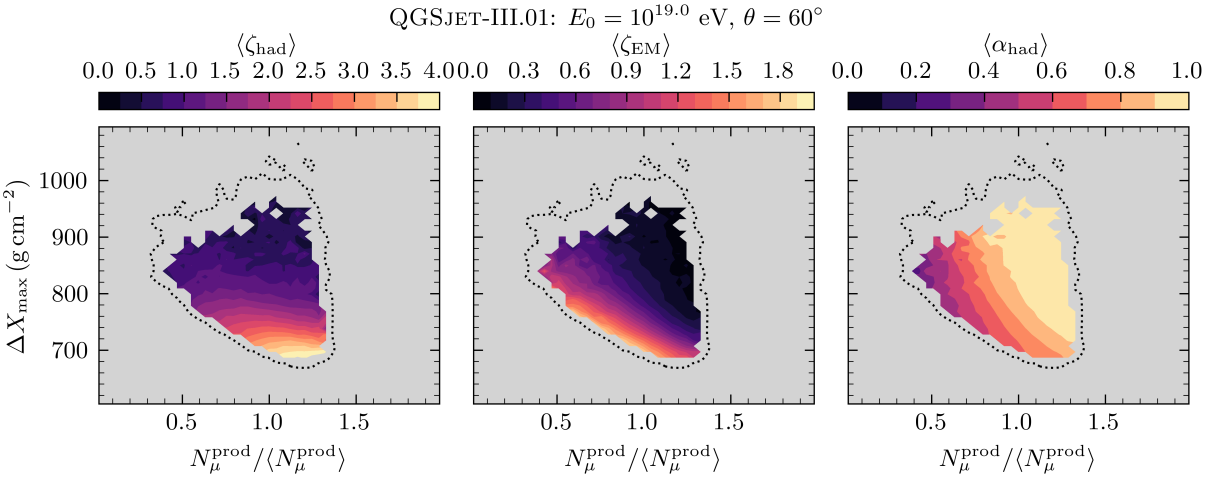

Figure 5.1: Contours of the multiparticle production variables of the primary interaction $\zeta_{\rm had}$ (left), $\zeta_{\rm EM}$ (middle) and $\alpha_{\rm had}$ (right) over the joint distribution $f(n_\mu^{\rm prod}, \Delta X_{\max})$. The dotted black contour represents the 3σ contour of the latter joint distribution. The figure was produced with libraries of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$ employing QGSJET-III.01.

about the primary interaction must be partially retained in $f(N_\mu, X_{\max})$. Hence, we aim to interpret the different regions of the latter joint distribution in light of the types of primary interactions characterised by different values of $\zeta_{\rm had}$, $\zeta_{\rm EM}$ and $\alpha_{\rm had} \simeq \alpha_1$. In particular, since the scale of the number of muons is dependent on the history of hadron production in every stage of the shower rather than to the primary interaction, we will work with $n_\mu^{\rm prod} \equiv N_\mu^{\rm prod} / \langle N_\mu^{\rm prod} \rangle$ and $n_\mu \equiv N_\mu / \langle N_\mu \rangle$.

Figure 5.1 shows the contours of $\zeta_{\rm had}$ (left), $\zeta_{\rm EM}$ (middle), and $\alpha_{\rm had}$ (right) over the joint distribution of $f(N_\mu^{\rm prod}, \Delta X_{\max})$. These contours are produced by weighting the showers in each bin of $(N_\mu^{\rm prod}, \Delta X_{\max})$ by the values of $\zeta_{\rm had}$, $\zeta_{\rm EM}$, and $\alpha_{\rm had}$ from the corresponding primary interactions. Note that we begin by interpreting $N_\mu^{\rm prod}$ and $\Delta X_{\max}$, rather than N_μ and $X_{\max}$, because the former observables are more directly connected with particle production in the primary interaction.

The gradient of $\zeta_{\rm had}$ is nearly vertical across the $N_\mu^{\rm prod}$–$\Delta X_{\max}$ plane and oriented towards shallower showers. As discussed in Section 4.4, higher values of $\zeta_{\rm had}$ are indicative of increased hadronic activity in the primary interaction. Such interactions are highly inelastic, with the differential energy flow into the hadronic sector peaking at lower pseudorapidities. The resulting high multiplicity of produced hadrons, each carrying a small fraction of the primary energy, leads to showers that develop rapidly. These showers reach $\Delta X_{\max}$ at shallower atmospheric depths and have an enriched muon content. In contrast, weakly inelastic interactions, or those dominated by energy transfer to neutral pions, are characterised by low or vanishing values of $\zeta_{\rm had}$. These interactions typically yield deeper showers, accompanied by increased fluctuations in their muon content.

The shallowest showers for a given number of produced muons tend to originate from primary interactions with high values of $\zeta_{\rm EM}$. Such values are characteristic of interactions producing a large number of neutral pions that share the primary energy relatively evenly. Depending on the total fraction of energy carried by these particles, namely $1 - \alpha_{\rm had}$, the muon content of the resulting shower can vary significantly, tracing the lower edge of

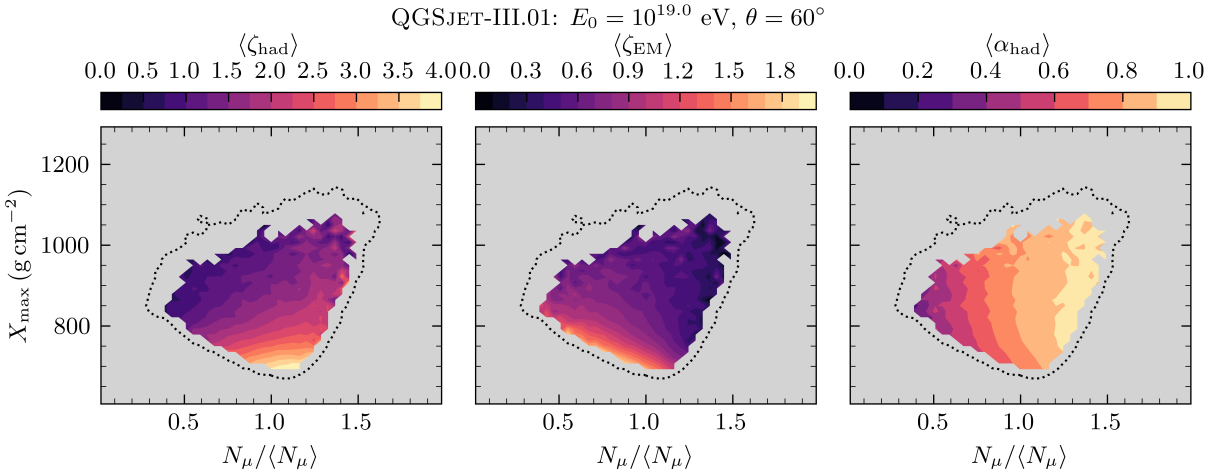

Figure 5.2: Contours of the multiparticle production variables of the primary interaction $\zeta_{\rm had}$ (left), $\zeta_{\rm EM}$ (middle) and $\alpha_{\rm had}$ (right) over the joint distribution $f(n_\mu, X_{\max})$. The dotted black contour represents the 3σ contour of the latter joint distribution. The figure was produced with libraries of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$ employing QGSJET-III.01.

the distribution of $f(N_\mu^{\rm prod}, \Delta X_{\max})$. Deeper showers typically arise from combinations of high $\zeta_{\rm EM}$ and low $\alpha_{\rm had}$, or from *quasi*-elastic primary interactions, $\zeta_{\rm EM} = 0$ and $\alpha_{\rm had} = 1$, which populate the upper-right corner of $f(N_\mu^{\rm prod}, \Delta X_{\max})$. The gradient of $\alpha_{\rm had}$ is more horizontal, pointing towards larger values of $N_\mu^{\rm prod}$. Showers with the highest muon content result from primary interactions with $\alpha_{\rm had} \simeq 0.8$ and the highest values of $\zeta_{\rm had}$. These correspond to interactions in which the primary energy is nearly equipartitioned among secondary hadrons, such that $\alpha_{\rm had}$ approaches the so-called charge ratio—the ratio of neutral pion multiplicity to the total multiplicity of secondary hadrons [55].

Importantly, the well-defined contours of the variables $\zeta_{\rm had}$, $\zeta_{\rm EM}$ and $\alpha_{\rm had}$ in the $N_\mu^{\rm prod} - \Delta X_{\max}$ plane evidence that information about the primary interaction is well preserved in these key shower observables, and is therefore accessible. In fact, this information is also retained in the joint distribution of the measurable shower observables N_μ and $X_{\max}$, as evidenced in Figure 5.2. These observables already account for the fluctuations of X_1, which are independent of the particle production in the primary interaction, as well as for the effects of muon attenuation and additional fluctuations arising from the truncation of the muon profile. In particular, these effects result in flatter gradients of the primary interaction production variables and cause slight rotations in their directions, due to the migration of similar primary interactions along the positive diagonal of the $N_\mu - X_{\max}$ plane.

5.2 Probabilistic mapping between the primary interaction and $f(N_\mu, X_{\max})$

To exploit the information about the primary interaction contained in $f(N_\mu^{\rm prod}, \Delta X_{\max})$, and ultimately in $f(N_\mu, X_{\max})$, we extend the formalism introduced in Equation (4.22). Recalling that α_1 is the primary interaction predictor of the relative muon content, $n_\mu^{\rm prod}$,

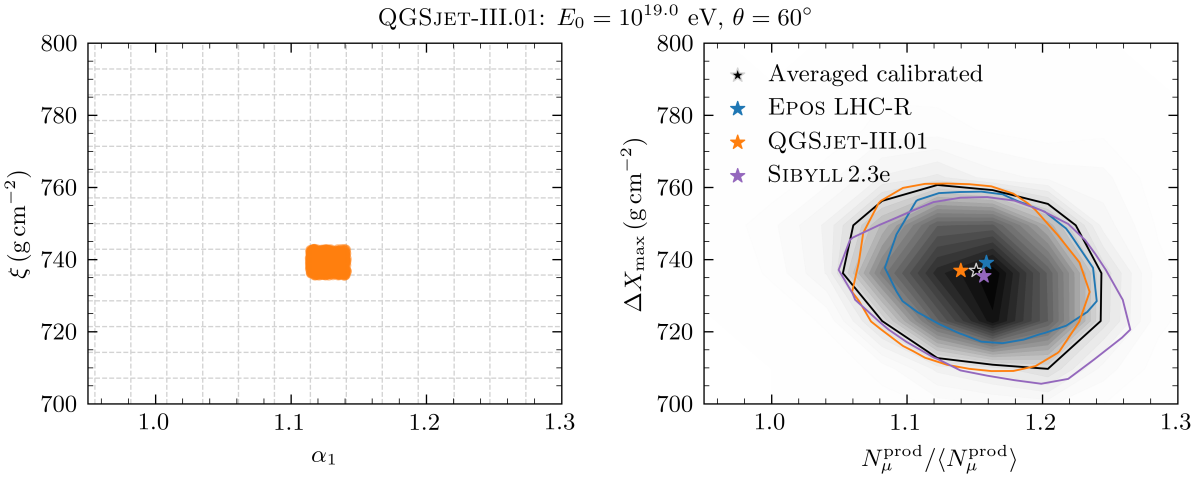

Figure 5.3: Schematic example of the shower response function of the different hadronic interaction models to a narrow bin in the input space (α_1, ξ). The events in the input space are provided by QGSJET-III.01. The grey contours in the output space $N_\mu^{\text{prod}} - \Delta X_{\max}$ depict the average response function of the shower after the calibration shown in Figure 4.18. The 1σ contours of the response function of the different models is depicted by the coloured lines. The centroids of the response functions are indicated by the starred markers.

we define the model-independent shower response $f(n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$, which is the kernel of the integral transformation

$$f(n_\mu^{\text{prod}}, \Delta X_{\max}) = \iint f(n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)\, f(\alpha_1, \xi)\, \mathrm{d}\alpha_1\, \mathrm{d}\xi\,, \tag{5.1}$$

describing the probabilistic mapping of the priors $f(\alpha_1, \xi)$ onto the joint distribution of n_μ^{prod} and $\Delta X_{\max}$. Note that describing the scale of the produced number of muons would require a complete account of the muon yield from later generations of the shower, which is beyond the scope of this work.

Using the shorthand notation in which the PDF of a random variable is denoted by the variable itself, we rewrite Equation (5.1) as

$$(n_\mu^{\text{prod}}, \Delta X_{\max}) = (n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle) \otimes (\alpha_1, \xi). \tag{5.2}$$

The shower response $(n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ is modelled as described in Section 4.3.1, with the distinction that the Gaussian Kernel Density Estimation (KDE) of the point set $(n_\mu^{\text{prod}}, \Delta X_{\max})$ is now performed within narrow bins of the $\alpha_1 - \xi$ space, as illustrated in Figure 5.3. The $\alpha_1 - \xi$ space will sometimes be referred to as the input space, or the domain of the integral transformation in Equation (5.2). Conversely, the $n_\mu^{\text{prod}} - \Delta X_{\max}$ and $n_\mu - X_{\max}$ spaces will be referred to as output spaces. As before, the average kernel, or shower response, is obtained by re-centring the response functions from each model $M \in \{\text{EPOS LHC-R}, \text{QGSJET-III.01}, \text{SIBYLL2.3e}\}$ and calibrating the overall $\Delta X_{\max}$ scale to $\langle\xi\rangle$, as shown in Figure 4.18.

To accurately describe the steeply falling edges of the joint distribution $f(n_\mu^{\text{prod}}, \Delta X_{\max})$, the bandwidth parameter of the kernel used in the Gaussian KDE is initially estimated using Silverman's rule and subsequently scaled by a factor that increases linearly with the

average $\Delta X_{\max}$ of the showers corresponding to each bin in (α_1, ξ). This scaling factor is optimised, for each model M, by minimising a χ^2 statistic comparing the true distribution $f(n_\mu^{\text{prod}}, \Delta X_{\max})$ with the distribution predicted by applying the model-specific kernel $f(n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1, \xi, M)$ to the corresponding prior $f(\alpha_1, \xi)$. The optimal bandwidth factor is then averaged over all hadronic interaction models and applied to the universal shower response $(n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$. The average bandwidth factor, f_b, is given by

$$f_b = 0.01 \times \left(\frac{\Delta X_{\max}}{\lambda_r} \right). \tag{5.3}$$

The same procedure is applied in the description of $f(n_\mu, X_{\max})$ in terms of (α_1, ξ):

$$(n_\mu, X_{\max}) = (n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle) \otimes (\alpha_1, \xi). \tag{5.4}$$

In the case of extensive air shower measurements enabling the reconstruction of the muon content at ground and the depth of maximum, n_μ^{rec} and $X_{\max}^{\text{rec}}$, the experimental resolution can be readily incorporated into the formalism. Letting $(n_\mu^{\text{rec}}, X_{\max}^{\text{rec}} \,|\, n_\mu, X_{\max}) \equiv \mathcal{R}(n_\mu^{\text{rec}}, X_{\max}^{\text{rec}} \,|\, n_\mu, X_{\max})$ denote the experimental response function, we can write:

$$(N_\mu^{\text{rec}}, X_{\max}^{\text{rec}}) = (N_\mu^{\text{rec}}, X_{\max}^{\text{rec}} \,|\, N_\mu, X_{\max}) \otimes (N_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle) \otimes (\alpha_1, \xi). \tag{5.5}$$

This forward-folding approach, in which the shower response to the primary interaction and the experimental response to the true shower observables are applied sequentially, can be used to fit the prior PDF $f(\alpha_1, \xi)$ that best describes the observed data. Although this procedure relies on an accurate modelling of the experimental response function, it is more robust than unfolding methods, which may produce unphysical results. A dedicated study implementing this iterative procedure is left for future work.

5.2.1 Visualizing the shower response to $f(\alpha_1, \xi)$

The application of the model defined in Equation (5.2) can be visualised by examining the shower response corresponding to narrow bins in the $\alpha_1 - \xi$ space, i.e., the input space or domain of the integral transformation. For a given bin in (α_1, ξ) indexed by ij, the corresponding universal response function is obtained by setting $f(\alpha_1, \xi) = \delta(\alpha_1 - \alpha_1^i)\,\delta(\xi - \xi^j)$ in Equation (5.2). This procedure is illustrated for six representative bins of $f(\alpha_1, \xi)$ in Figure 5.4. The left panel of the figure shows the full prior distribution $f(\alpha_1, \xi)$, while the middle and right panels depict the 1σ (dotted grey) and 2σ (solid grey) contours of the overall distribution $f(n_\mu^{\text{prod}}, \Delta X_{\max})$. The density contours of the shower response functions associated with each numbered bin in the input space are displayed in shades of orange. The centroids of each response function are marked with black stars.

The variables α_1 and ξ are moderately anti-correlated, with Pearson correlation coefficients ranging from $\rho = -0.52$ to $\rho = -0.35$, depending on the hadronic interaction model under consideration. This mild anti-correlation stems from the energy balance between the hadronic and electromagnetic sectors of the primary interaction. If neutral pions carry a large fraction of the primary energy, then α_1 tends to be small. Substituting this limit into Equation (4.14), we obtain, for $E_0 = 10^{19}$ eV, the approximation $\xi \simeq 917 - 37\,\zeta_{\text{EM}}$. Since the coefficients of α_{had} and ζ_{had} in the definition of ξ are negative and these production variables are non-negative, it follows that, for fixed ζ_{EM}, the condi-

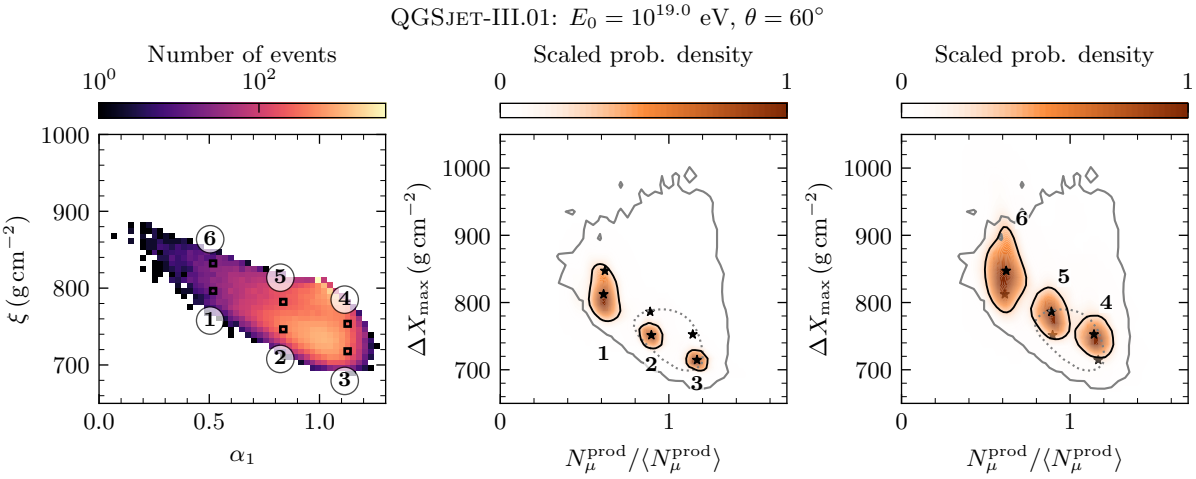

Figure 5.4: Left panel: joint distribution of α_1 and ξ, with highlighted and numbered bins centred at six different values of (α_1, ξ). Middle and right panels: contours of the universal shower response function $f(n_\mu^{\rm prod}, \Delta X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ applied to each of the highlighted bins in the left panel. The centroids of each response function are indicated by black starred markers, and their 1σ contours are depicted in black. The 1σ contour and the outline of $f(n_\mu^{\rm prod}, \Delta X_{\max})$ are represented by dotted and solid grey lines, respectively. The figure was produced using CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$, employing QGSJET-III.01.

tion $\alpha_{\rm had} = \zeta_{\rm had} = 0$ maximises ξ. This maximisation, combined with the minimisation of α_1, leads to an anti-correlation between these two primary interaction observables.

The detailed structure of the joint distribution $f(\alpha_1, \xi)$ arises from the energy spectra of the secondary particles produced in the primary interaction, as well as from the kinematic boundaries between the production variable pairs $(\zeta_{\rm had}, \alpha_{\rm had})$, $(\zeta_{\rm EM}, \alpha_{\rm had})$, and $(\zeta_{\rm had}, \zeta_{\rm EM})$. The precise influence of changes in the secondary energy spectra on the shape of $f(\alpha_1, \xi)$ remains a topic for further investigation. Finally, all *quasi*-elastic primary interactions are confined to a single bin in the $\alpha_1 - \xi$ plane, at position $\alpha_1 = 1$ and $\xi \approx 815\,\mathrm{g\,cm^{-2}}$.

Crucially, disjoint regions in the $\alpha_1 - \xi$ plane map onto narrow, largely non-overlapping regions in the $N_\mu^{\rm prod} - \Delta X_{\max}$ space. Moreover, the relative positions of selected bins in the input space (α_1, ξ) are preserved in the output space $(n_\mu^{\rm prod}, \Delta X_{\max})$. If this mapping behaviour holds for all pairs (α_1, ξ), excluding those associated with *quasi*-elastic interactions, then the following conclusions can be drawn:

1. The observed anti-correlation between $n_\mu^{\rm prod}$ and $\Delta X_{\max}$ is induced by the shape of the distribution $f(\alpha_1, \xi)$. Hence, information about the primary interaction is encoded in and recoverable from the distribution $f(n_\mu^{\rm prod}, \Delta X_{\max})$.

2. The universal shower response to a given type of primary interaction—characterised by specific values of (α_1, ξ)—is significantly narrower than the full distribution $f(n_\mu^{\rm prod}, \Delta X_{\max})$.

These two statements will be confirmed in Section 5.2.2, and constitute central results of this work.

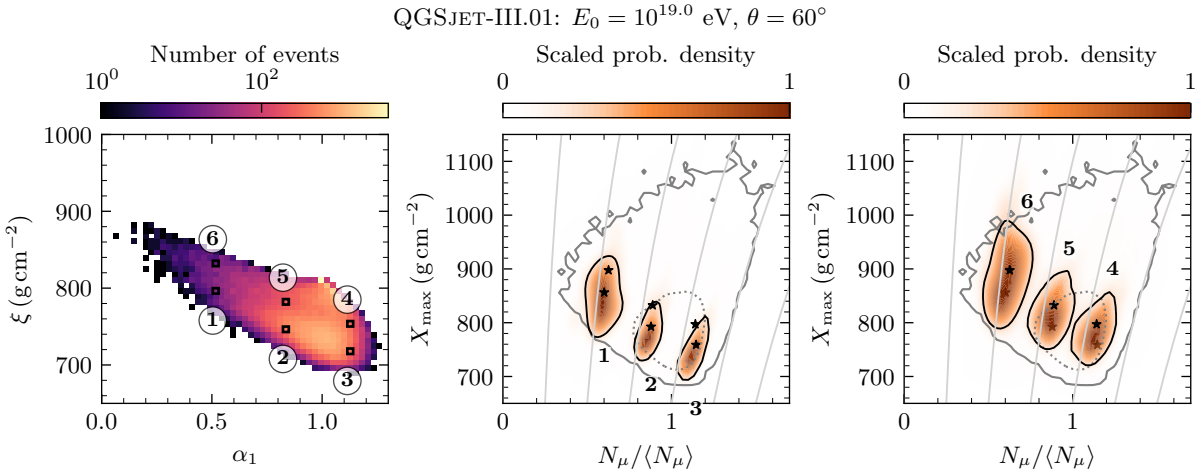

Figure 5.5: Left panel: joint distribution of α_1 and ξ, with highlighted and numbered bins centred at six different values of (α_1, ξ). Middle and right panels: contours of the universal shower response function $f(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ applied to each of the highlighted bins in the left panel. The centroids of each response function are indicated by black starred markers, and their 1σ contours are depicted in black. The 1σ contour and the outline of $f(n_\mu, X_{\max})$ are represented by dotted and solid grey lines, respectively. The light grey curvilinear grid represents the contours of the number of produced muons computed using Equation (2.18). The figure was produced with libraries of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$ employing QGSJET-III.01.

Lastly, note that reaching the deepest values of $\Delta X_{\max}$ requires accounting for the fluctuations of the interaction depth of the leading primary interaction in highly diffractive primary interactions. This leads to a vertical, exponential smearing of the response functions $f(n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$, which is absent in the joint distribution $f(\alpha_1, \xi)$.

The mapping between $f(\alpha_1, \xi)$ and the measurable shower observables n_μ and $X_{\max}$ is illustrated in Figure 5.5, for the same six input-space bins. The curvilinear light grey lines in the middle and right panels of the figure correspond to contours of the number of produced muons. These are computed by setting $n_\mu^{\text{prod}} = c$ in Equation (2.18) for several values of c, and using the relation $\mathcal{X}_{\max}^{\mu} = X_{\max} - X_c$, where $X_c = 220\,\text{g cm}^{-2}$ [108].

The shower response function $f(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ is elongated along the contours of constant number of produced muons, due to the exponential fluctuations of X_1 and their positive correlation with the attenuation of the muon flux. Nevertheless, the kernel retains sensitivity to different types of primary interactions, each sampling distinct regions of the distribution $f(\alpha_1, \xi)$. In particular, the kernel encoding the shower response remains narrow compared to the overall distribution $f(n_\mu, X_{\max})$.

5.2.2 Precision of the mapping $(\alpha_1, \xi) \to (N_\mu, X_{\max})$

We now assess the precision of the shower response functions $(n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ and $(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ across all bins of the distribution $f(\alpha_1, \xi)$. For each bin indexed by ij in the input space, centred at (α_1^i, ξ^j), we compute the area enclosed by the 1σ contour of $f(n_\mu^{\text{prod}}, \Delta X_{\max} \,|\, \alpha_1^i, \xi^j)$, denoted $\mathcal{A}_{1\sigma}^{ij}$. To a good approximation, this area is

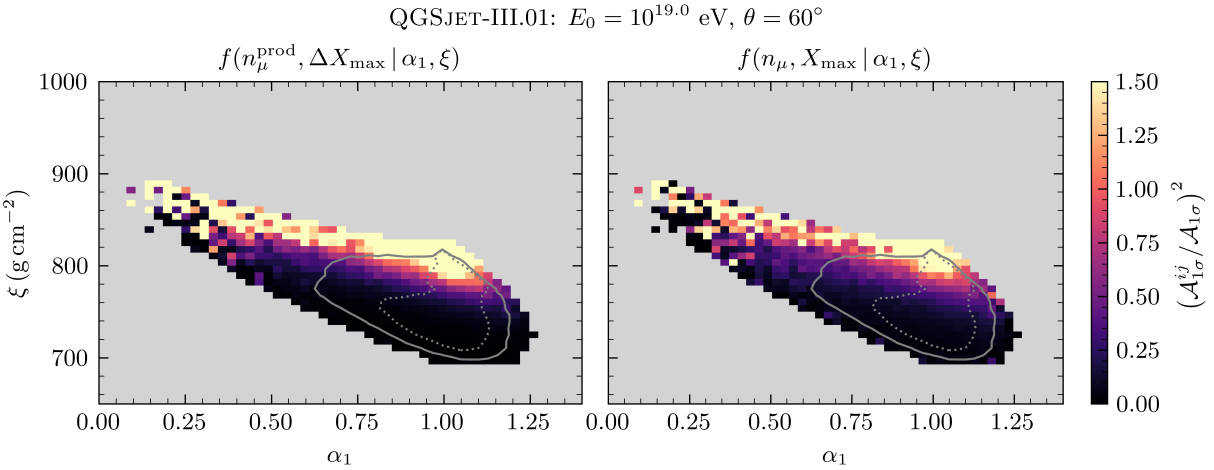

Figure 5.6: Square of the ratio of the 1σ area of the shower response functions $(n_\mu^{\rm prod}, \Delta X_{\max} \,|\, \alpha_1^i, \xi^j, \langle\xi\rangle)$ (left) and $(n_\mu, X_{\max} \,|\, \alpha_1^i, \xi^j, \langle\xi\rangle)$ (right) to the overall 1σ area of the distributions $f(n_\mu^{\rm prod}, \Delta X_{\max})$ and $f(n_\mu, X_{\max})$, respectively. This ratio is calculated for all interactions in narrow bins of α_1 and ξ indexed by ij covering the joint distribution $f(\alpha_1, \xi)$, provided by QGSJET-III.01. The dotted and solid grey lines are, respectively, the 1σ and 2σ contours of $f(\alpha_1, \xi)$.

given by $\mathcal{A}_{1\sigma}^{ij} \propto \sqrt{\det\{\Sigma_{ij}\}}$ [1], where $\det\{\Sigma_{ij}\}$ is the determinant of the covariance matrix Σ_{ij}, computed from vectors of random variables $(n_\mu^{\rm prod}, \Delta X_{\max})$ sampled from the shower response function $f(n_\mu^{\rm prod}, \Delta X_{\max} \,|\, \alpha_1^i, \xi^j, \langle\xi\rangle)$.

The squared ratio of $\mathcal{A}_{1\sigma}^{ij}$ [2] to the 1σ area of the overall distribution $f(n_\mu^{\rm prod}, \Delta X_{\max})$, denoted $\mathcal{A}_{1\sigma}$, is shown in the left panel of Figure 5.6. The right panel of the same figure shows the same analysis applied to the kernel $f(n_\mu, X_{\max} \,|\, \alpha_1^i, \xi^j, \langle\xi\rangle)$. A ratio equal to unity indicates that the shower response functions are independent of the values of (α_1, ξ) in that particular bin of the input space. Ratios larger than unity are mainly due to the inaccurate modelling of the response function when only a small number of events is present in the corresponding bin of the input space.

The majority of bins in the (α_1, ξ) space yield a very narrow shower response in terms of the explored observables, relative to their overall distributions. Thus, for the bulk of primary interactions, the stochasticity associated with hadron production and propagation in later stages of the shower plays a subdominant role in shaping the joint distributions $f(n_\mu^{\rm prod}, \Delta X_{\max})$ and $f(n_\mu, X_{\max})$. Values of the ratio $\mathcal{A}_{1\sigma}^{ij}/\mathcal{A}_{1\sigma}$ close to or greater than unity correspond to bins in the input space populated by highly elastic primary interactions, either due to *quasi*-elastic scattering of the incident proton or to diffractive interactions where a highly energetic neutral pion is produced. These interactions tend to populate the upper edge of the joint distribution $f(\alpha_1, \xi)$, and produce very little variation in α_1 and ξ, while the corresponding shower response, driven by subsequent interactions,

[1] This is trivial to show for a diagonal covariance matrix, for which $\det\{\Sigma_{ij}\} = \sqrt{\sigma^2(n_\mu^{\rm prod})\sigma^2(\Delta X_{\max})} = \sigma(n_\mu^{\rm prod})\sigma(\Delta X_{\max})$, which is indeed proportional to the area of an ellipse with minor and major axis of sizes $\sigma(n_\mu^{\rm prod})$ and $\sigma(\Delta X_{\max})$

[2] In analogy with the one-dimensional case, it is more appropriate to compare the variance of the estimator of the primary interaction with that of the residuals when defining the variance of the air-shower observable, since variances, unlike standard deviations, are additive.

Model	$\mathcal{A}_{1\sigma}(\alpha_1,\xi)/\mathcal{A}_{1\sigma}$	$\mathcal{A}_{1\sigma}(n_\mu^{\rm prod}-\alpha_1,\Delta X_{\max}-\xi)/\mathcal{A}_{1\sigma}$
EPOS LHC-R	0.63	0.48
QGSJET -III.01	0.71	0.44
SIBYLL2.3e	0.50	0.49

Table 5.1: Relative contributions of $f(\alpha_1,\xi)$ and $f(n_\mu^{\rm prod}-\alpha_1,\Delta X_{\max}-\xi)$ to the joint distribution of $f(n_\mu^{\rm prod},\Delta X_{\max})$. The notation $\mathcal{A}_{1\sigma}(\cdot,\cdot)$ refers to the area enclosed by the 1σ contour of a given distribution $f(\cdot,\cdot)$, while $\mathcal{A}_{1\sigma}$ by itself refers to the area enclosed by the 1σ contour of $f(n_\mu^{\rm prod},\Delta X_{\max})$.

may vary appreciably.

The overall influence of the stochasticity of the shower evolution on the correlated variability of $(n_\mu^{\rm prod},\Delta X_{\max})$ can be assessed by computing the area enclosed by the 1σ contour of the joint distribution of residuals, $\mathcal{A}_{1\sigma}(n_\mu^{\rm prod}-\alpha_1,\Delta X_{\max}-\xi)$, and comparing it to the area covered by the 1σ region of $f(\alpha_1,\xi)$, denoted $\mathcal{A}_{1\sigma}(\alpha_1,\xi)$. Explicitly, the overall response function is obtained by marginalizing the kernel of the integral transformation $(\alpha_1,\xi)\to(n_\mu^{\rm prod},\Delta X_{\max})$ as

$$f(n_\mu^{\rm prod}-\alpha_1,\Delta X_{\max}-\xi)=\iint f(n_\mu^{\rm prod}-\alpha_1,\Delta X_{\max}-\xi\,|\,\alpha_1,\xi)f(\alpha_1,\xi)\;\mathrm{d}\alpha_1\,\mathrm{d}\xi\,. \quad (5.6)$$

The ratios $\mathcal{A}_{1\sigma}(n_\mu^{\rm prod}-\alpha_1,\Delta X_{\max}-\xi)/\mathcal{A}_{1\sigma}$ and $\mathcal{A}_{1\sigma}(\alpha_1,\xi)/\mathcal{A}_{1\sigma}$ are shown in Table 5.1 for all hadronic interaction models.

Irrespective of the hadronic interaction model, the correlated variability of (α_1,ξ) accounts for the majority of the variability observed in $(n_\mu^{\rm prod},\Delta X_{\max})$. Although dominant, the relative contribution of the stochasticity of (α_1,ξ) to the fluctuations in $(n_\mu^{\rm prod},\Delta X_{\max})$ varies depending on the hadronic interaction model.

To investigate the directions of greatest variability in the distributions of (α_1,ξ), $(n_\mu^{\rm prod}-\alpha_1,\Delta X_{\max}-\xi)$, and $(n_\mu^{\rm prod},\Delta X_{\max})$, we extract their respective covariance matrices and use them to generate the corresponding 1σ elliptical contours, as shown in Figure 5.7. These contours have been re-centred to facilitate a direct comparison.

The major axis of the 1σ contour of $f(\alpha_1,\xi)$ is oriented from the top left to the bottom right, reflecting the anti-correlation between α_1 and ξ, irrespective of the hadronic interaction model. The variance along this direction drives the variance of $(n_\mu^{\rm prod},\Delta X_{\max})$ along the negative diagonal. By contrast, the major axis of the distribution $(n_\mu-\alpha_1,\Delta X_{\max}-\xi)$ is nearly vertical, contributing with a vertical subdominant smearing to $f(n_\mu^{\rm prod},\Delta X_{\max})$. Therefore, fluctuations in the later stages of the shower, beyond the first interaction, play a subdominant role in shaping the overall joint distribution $f(n_\mu^{\rm prod},\Delta X_{\max})$—particularly along its negative diagonal.

Including the fluctuations of the depth of the first interaction, we compare the 1σ contours of the distributions $f(\alpha_1,\xi)$, $f(n_\mu-\alpha_1,X_{\max}-\xi)$, and $f(n_\mu,X_{\max})$ in Figure 5.8.

The exponential fluctuations of X_1, together with the positively correlated muon attenuation, are responsible for tilting the contours of the marginal response function $f(n_\mu-\alpha_1,X_{\max}-\xi)$ towards the positive diagonal. This introduces a positively correlated variability in $(n_\mu,X_{\max})$, while their anti-correlation continues to be driven by the anti-correlation between α_1 and ξ. Naturally, the variability of X_1 degrades the causal

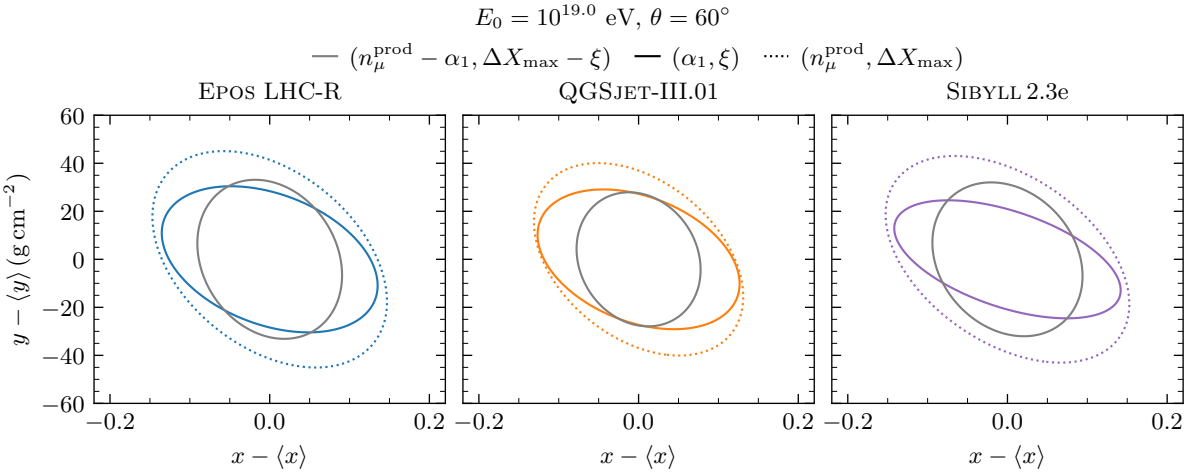

Figure 5.7: Elliptical 1σ contours computed from the covariance matrices of the joint distributions of (α_1, ξ) (solid colored), $(n_\mu^{\text{prod}} - \alpha_1, \Delta X_{\max} - \xi)$ (solid grey) and $(n_\mu^{\text{prod}}, \Delta X_{\max})$ (dotted coloured). The figure was produced with libraries of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$ employing EPOS LHC-R (left), QGSJET-III.01 (middle) and SIBYLL2.3e (right).

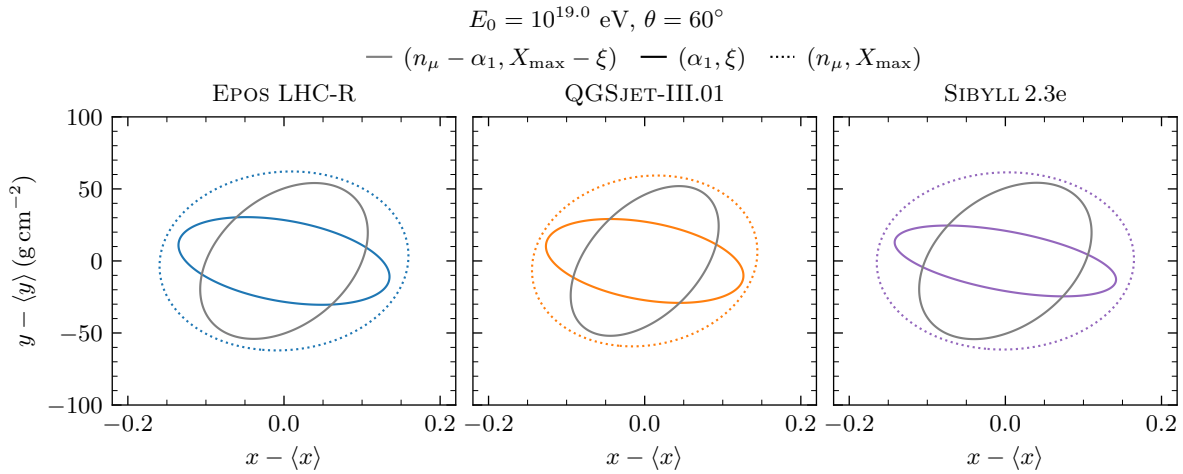

Figure 5.8: Elliptical 1σ contours computed from the covariance matrices of the joint distributions of (α_1, ξ) (solid colored), $(n_\mu - \alpha_1, X_{\max} - \xi)$ (solid grey) and $(n_\mu, X_{\max})$ (dotted coloured). The figure was produced with libraries of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$ employing EPOS LHC-R (left), QGSJET-III.01 (middle) and SIBYLL2.3e (right).

relation between ξ and $X_{\max}$. However, the shape of the distribution $f(N_\mu, X_{\max})$ retains the imprint of the primary interaction along its negative diagonal. Importantly, the different hadronic interaction models predict similar distributions of X_1, owing to their comparable predictions for the proton-air cross-section. Since the inclusion of these fluctuations has a similar impact across all models, they can be safely incorporated into the universal kernel, as was already evident from Figure 4.22, and as will become clearer in the following sections.

The strong causal link between (α_1, ξ) and the shower observables not only leads to reduced variability for similar types of primary interactions, but also determines the location

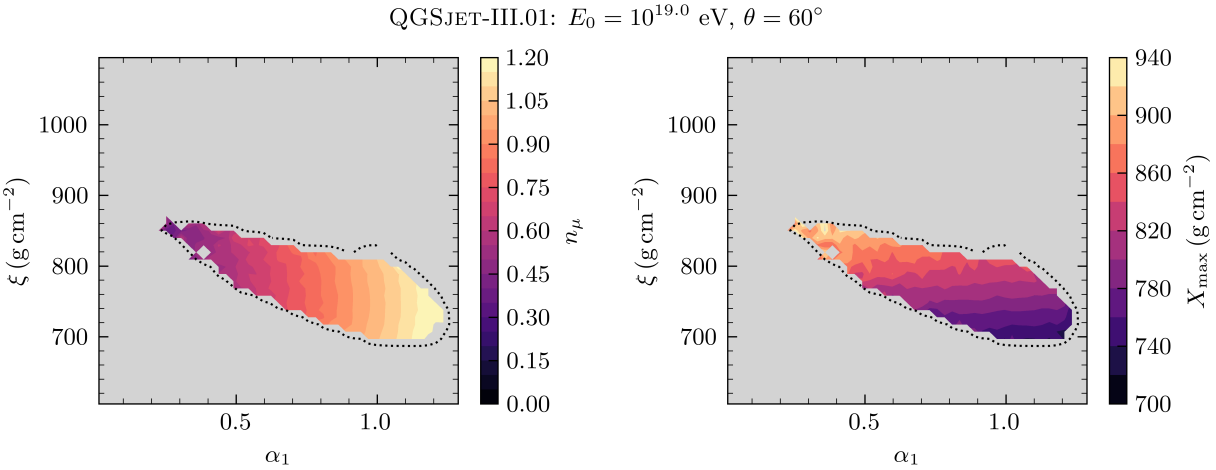

Figure 5.9: Contours of n_μ (left) and $X_{\max}$ (right) over the joint distribution of $f(\alpha_1, \xi)$. The dotted black contour represents the boundary of the latter joint distribution. The figure was produced with libraries of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$ employing QGSJET-III.01.

of the peak of the shower response with respect to the overall distribution $f(n_\mu, X_{\max})$. This was already evident in Figure 5.5, where the relative positions of specific bins in the (α_1, ξ) space are preserved in the output space. We confirm this behaviour across all bins in the input space by weighting each primary interaction in a given bin ij in the $\alpha_1 - \xi$ plane by the values of n_μ and $X_{\max}$ of the corresponding shower. This is shown in Figure 5.9.

The approximately horizontal gradient of n_μ across the distribution of $f(\alpha_1, \xi)$ indicates that, for each fixed value of ξ, the value of this shower observable is primarily determined by α_1. Similarly, despite the fluctuations in X_1, it is evident that the value of $X_{\max}$ is largely determined by ξ, for each fixed value of α_1.

5.2.3 Prediction of $f(N_\mu, X_{\max})$ from the primary interaction

Having established the precision and accuracy of the universal shower response $(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$, and shown that it can be reliably used to predict the distribution of $X_{\max}$ (see Chapter 4), we now demonstrate that it also enables the prediction of the distribution of n_μ. This distribution is obtained by marginalising Equation (5.4) as

$$f(n_\mu) = \int \left[\iint f(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle) f(\alpha_1, \xi) \, \mathrm{d}\alpha_1 \, \mathrm{d}\xi \right] \mathrm{d}X_{\max} \,. \tag{5.7}$$

Figure 5.10 shows the predicted distribution of n_μ, obtained by applying Equation (5.7) to the prior distributions $f(\alpha_1, \xi)$ provided by different hadronic interaction models.

The agreement between the true and predicted distributions $f(n_\mu)$ is excellent across the entire range of n_μ. In particular, the tail of n_μ towards low values, which is highly model-dependent and sensitive to the differential inclusive cross-section for the production of neutral pions in proton-air interactions [106], is accurately reproduced for all hadronic interaction models. The accuracy of the model in describing this tail, populated by muon-depleted showers, is assessed by fitting the low-end of $f(\ln n_\mu)$ to an exponential function

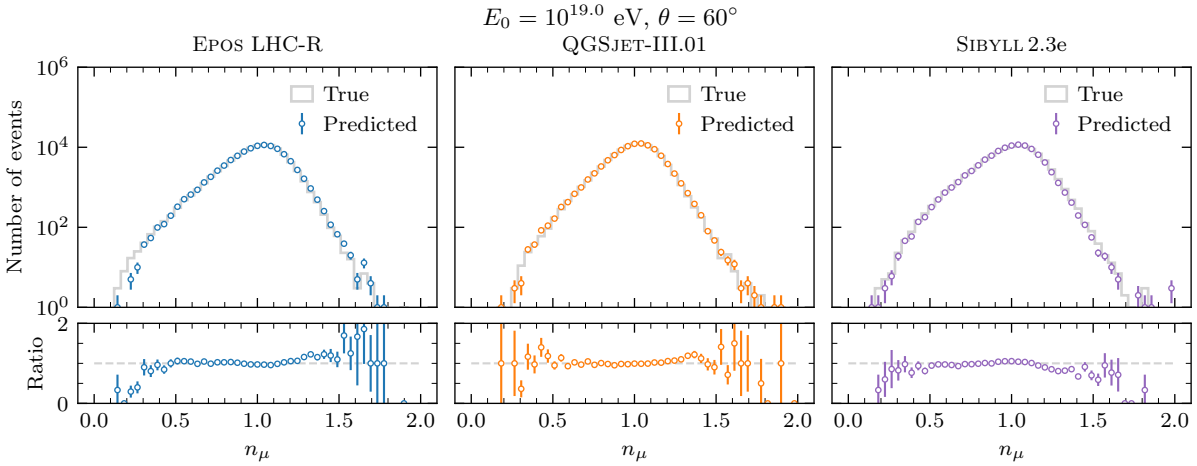

Figure 5.10: Upper panels: Distributions of N_μ predicted by applying the universal shower response model $(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ to prior distributions of ξ produced with the hadronic interaction models EPOS LHC-R (left), QGSJET-III.01 (middle) and SIBYLL2.3e (right). The underlying grey histograms are true distributions of $f(n_\mu)$. Lower panels: ratio of the predicted to the true distributions $f(n_\mu)$. This figure was produced with the library of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$.

of the form $y \propto \mathrm{e}^{\ln n_\mu / \Lambda_\mu}$, and extracting the corresponding slope parameter Λ_μ. As in Chapter 3, the tail is defined as all values of n_μ below the 10 % quantile of $f(\ln n_\mu)$. The fit is performed by maximising an unbinned likelihood using the exponential function as a probability density function, evaluated at each value of $\ln n_\mu$.

The bias in Λ_μ induced by the application of the universal shower response model is shown in the right panel of Figure 5.11. The left and middle panels of the figure show the corresponding biases in the first and second moments of $f(n_\mu)$, respectively. In the panels for $\sigma(n_\mu)$ and Λ_μ, the shaded bands represent the spread in the predictions across the different hadronic interaction models. For $\langle n_\mu \rangle$, since the value is normalised to 1 for all models, the shaded band represents a conservative estimate of the experimental uncertainty in the determination of the relative muon scale, taken to be 10 % [95].

The prediction of $\langle n_\mu \rangle$ exhibits a maximum bias of 2 % across all hadronic interaction models. This bias is taken as a systematic uncertainty inherent to the proposed probabilistic model. Moreover, the values of $\sigma(n_\mu)$ are systematically underestimated for all models, indicating that the modelling of the shower response can be further improved. In particular, this could be achieved by oversampling the regions of the distribution of $f(n_\mu)$ that are more poorly described by the universal kernel. Nevertheless, the spread of the bias in $\sigma(n_\mu)$ across models remains much smaller than the inter-model differences in this observable. Importantly, the systematic uncertainty in Λ_μ is less than 30 % of the spread of Λ_μ predicted across the different hadronic interaction models. This confirms that the two-dimensional shower response model can be reliably used to study the impact of variations in the primary interaction on the distribution $f(n_\mu)$. As with $X_{\max}$, discussed in Section 4.3.2, this framework provides the basis for data-driven constraints on hadronic interactions using the distribution of n_μ.

More significantly, since Λ_μ is weakly sensitive to the presence of heavier primaries in the cosmic ray flux, the developed formalism could be directly applied to propagate

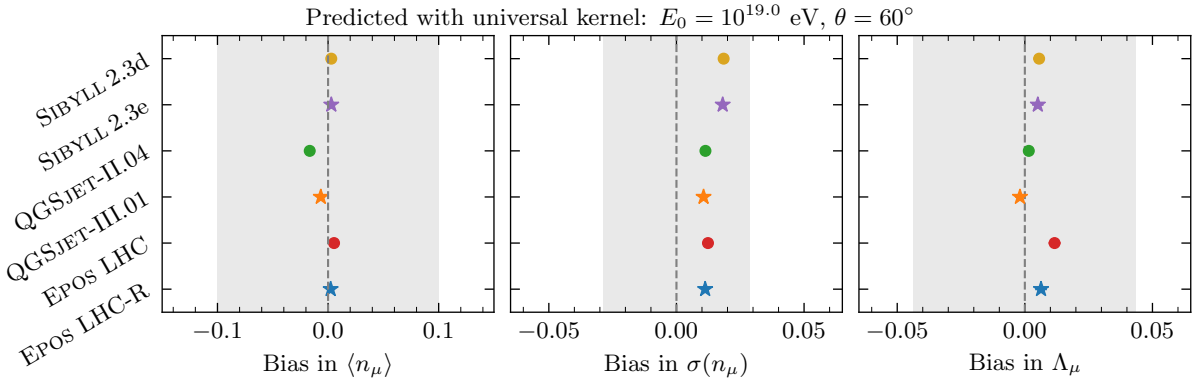

Figure 5.11: Bias in the main features of the distribution of n_μ: average (left panel), standard deviation (middle panel) and slope of the exponential tail of $f(\ln n_\mu)$ towards low-n_μ, Λ_μ (right panel), due to the prediction of $f(n_\mu)$ using the universal shower response $(n_\mu, X_{\max} \,|\, \xi, \langle\xi\rangle, \langle\xi\rangle)$ applied to different priors $f(\alpha_1, \xi)$ provided by several hadronic interaction models. This figure was produced with the library of CONEX simulations of proton-induced showers with $E_0 = 10^{19}$ eV and $\theta = 60°$.

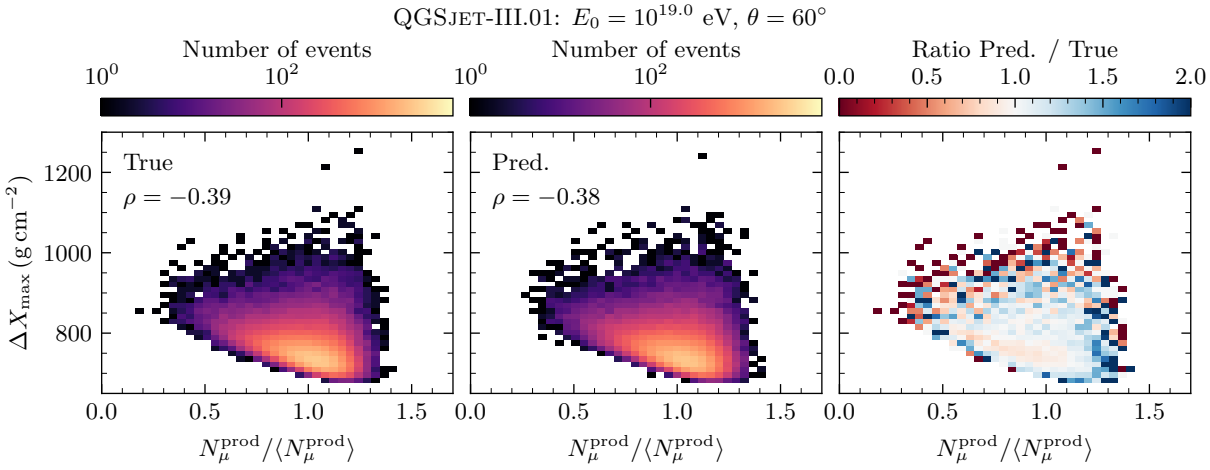

Figure 5.12: Left panel: true distribution $f(n_\mu^{\text{prod}}, \Delta X_{\max})$. Middle panel: predicted joint distribution of n_μ^{prod} and $\Delta X_{\max}$ via the application of the universal shower response function (see Equation (5.2)) to the prior distribution $f(\alpha_1, \xi)$ provided by QGSJET-III.01. Right panel: ratio between the predicted and true joint distribution $f(n_\mu^{\text{prod}}, \Delta X_{\max})$.

changes in the spectrum of neutral pions produced in primary proton-air interactions until agreement with a measured value of Λ_μ is reached. These changes would lead to a consistent change in $X_{\max}$, also predicted by our two-dimensional framework, that could act as a further constraint on the spectrum of neutral pions. This particular application will be investigated elsewhere.

Naturally, our formalism enables the prediction of the full joint distribution $f(n_\mu, X_{\max})$, rather than just the marginal distributions. Figure 5.12 shows the true joint distribution $f(n_\mu^{\text{prod}}, \Delta X_{\max})$ and the one predicted using Equation (5.2), for a prior distribution $f(\alpha_1, \xi)$ provided by the hadronic interaction model QGSJET-III.01.

The agreement between the true and predicted distribution $f(n_\mu^{\text{prod}}, \Delta X_{\max})$ is remarkable, as evidenced by the near-unity ratio across the N_μ^{prod}–$\Delta X_{\max}$ plane. We recall that

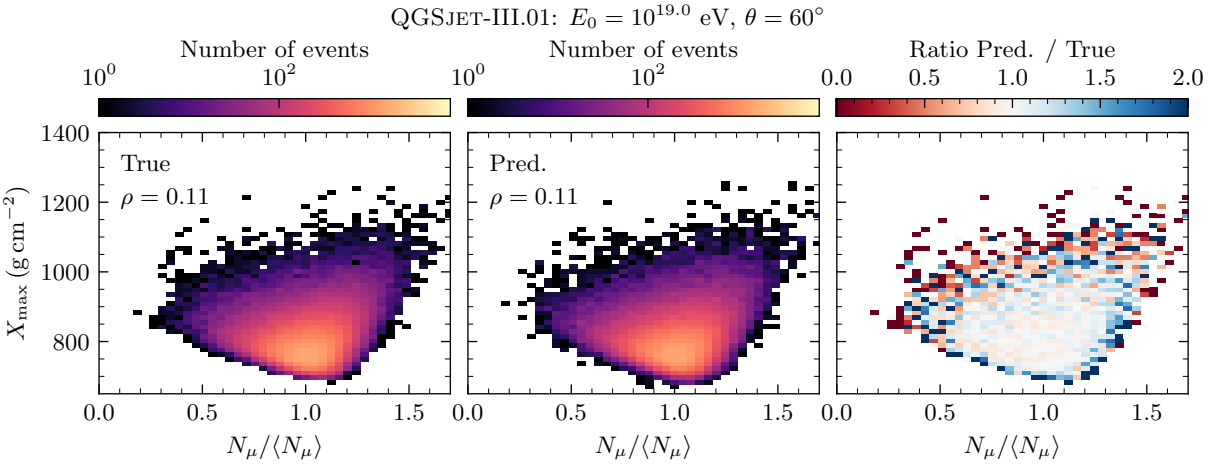

Figure 5.13: Left panel: true distribution $f(n_\mu, X_{\max})$. Middle panel: predicted joint distribution of n_μ and $X_{\max}$ via the application of the universal shower response function (see Equation (5.2)) to the prior distribution $f(\alpha_1, \xi)$ provided by QGSJET-III.01. Right panel: ratio between the predicted and true joint distribution $f(n_\mu, X_{\max})$.

this prediction relies on a universal shower response model, highlighting that modifications in the shape of the joint distribution of the shower observables are driven mostly by changes in the primary interaction, together with their average propagation through subsequent shower generations. The disagreement between the true and predicted distributions in the shallow $\Delta X_{\max}$-large N_μ region indicates that the bandwidth parameter of the kernel density estimation is too wide to account for the steep cut-off of the true distribution $f(n_\mu^{\text{prod}}, \Delta X_{\max})$. On the other hand, the disagreement in the deep $\Delta X_{\max}$ region is mostly due to undersampling.

Similarly, the agreement between the true distribution $f(n_\mu, X_{\max})$ and its prediction using Equation (5.4) is evident in Figure 5.13. This agreement arises in part from the similarity of the fluctuations in X_1 and their effect on the attenuation of the muon flux across different hadronic interaction models. Such similarity ensures that differences in $f(n_\mu, X_{\max})$ between models are primarily driven by differences in $f(n_\mu^{\text{prod}}, \Delta X_{\max})$, which in turn reflect the models' distinct treatments of the primary interaction. This further reinforces the applicability of the presented formalism for constraining the properties of the primary interaction based on measurements of the full joint distribution $f(n_\mu, X_{\max})$.

Given the agreement between the true and predicted distributions of $f(N_\mu, X_{\max})$, we now assess the model's ability to describe the evolution of Λ_μ with $X_{\max}$, as introduced in Chapter 3. This evolution is of particular interest due to its robustness against the presence of heavier primaries in an ensemble of cosmic-ray-induced extensive air showers.

Figure 5.14 shows the true and predicted distributions of $\ln n_\mu$ for the bins in $X_{\max}$ defined in Section 3.1, using the hadronic interaction model QGSJET-III.01. The predicted distributions $f(\ln n_\mu \,|\, X_{\max})$ are obtained in two ways: first, using the Gaussian KDE of the shower response specific to the selected hadronic interaction model; and second, using the universal response model described in Equation (5.4). This enables a separate assessment of the effects of approximating the shower response with a continuous function and of adopting a model-averaged approach.

Whether the model-specific response function of the shower — referred to as the

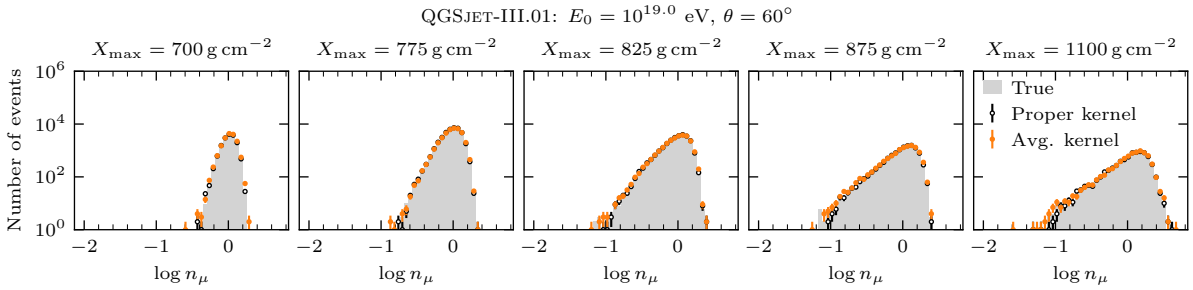

Figure 5.14: Distribution of $\ln n_\mu$ per bin in $X_{\max}$. Each panel corresponds to a different bin, whose center is indicated above each panel and increases from left to right. The true distributions $f(\ln n_\mu \,|\, X_{\max})$ are represented by the grey histograms. The distributions predicted using the model's proper shower response modelled with a Gaussian Kernel Density Estimation are represented by the black empty markers. The distributions predicted using the universal shower response $(n_\mu, X_{\max} \,|\, \xi, \langle\xi\rangle\,, \langle\xi\rangle)$ are represented by the orange markers. The priors $f(\alpha_1, \xi)$ are provided by QGSJET-III.01.

proper kernel — $f(n_\mu, X_{\max} \,|\, \alpha_1, \xi, M)$ for M = QGSJET-III.01, or the universal kernel $f(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ is used, there is very good agreement between the true and predicted distributions $f(n_\mu \,|\, X_{\max})$. In particular, a simple χ^2 test yields values of reduced χ^2/dof ranging between 1.1 and 2 when using the proper response function, and values between 1.3 and 11 when using the universal kernel. The poorer agreement of the latter may indicate residual differences between models in later shower generations, which can slightly affect the shape of $f(n_\mu, X_{\max})$, or that the averaging and/or calibration of the kernel could be further refined.

Moreover, the agreement between the true and predicted distributions worsens for shallower showers, where the distribution of $\ln n_\mu$ becomes narrower and thus more sensitive to the bandwidth parameter of the Gaussian kernel density estimation of the shower response. This suggests that the modelling of the kernels could be improved, possibly by employing alternative approaches beyond Gaussian KDE. Nonetheless, these limitations have little impact on the estimation of the main moments of the marginal distributions.

In Figure 5.15, we compare the true evolution of Λ_μ with $X_{\max}$ to the predictions obtained using either the proper kernel (left panel) or the universal kernel (right panel), applied to priors $f(\alpha_1, \xi)$ generated by different hadronic interaction models. The lower panels of the same figure show the relative bias between the true Λ_μ and its prediction $\widehat{\Lambda}_\mu$, computed as $1 - \widehat{\Lambda}_\mu/\Lambda_\mu$.

Using the proper kernel yields very similar biases between the true and predicted values of Λ_μ across hadronic interaction models. Furthermore, the absolute value of the relative bias tends to increase with $X_{\max}$, although it remains below 10% for all models. The underestimation of Λ_μ in the deepest $X_{\max}$ bin indicates that the predicted distribution $f(\ln n_\mu \,|\, 1100\,\mathrm{g\,cm^{-2}})$ exhibits a steeper low-end tail than the true one. This suggests that the bias may stem from an undersampling of the distribution $f(n_\mu, X_{\max})$ for deep, muon-depleted showers. Importantly, the modelling of the proper kernel preserves the inter-model spread in Λ_μ.

Using the universal kernel leads to a reduction in the separation between hadronic interaction models based on Λ_μ, suggesting that deeper shower generations might play a small but non-negligible role in shaping the tail of $\ln n_\mu$ for deep showers. Moreover,

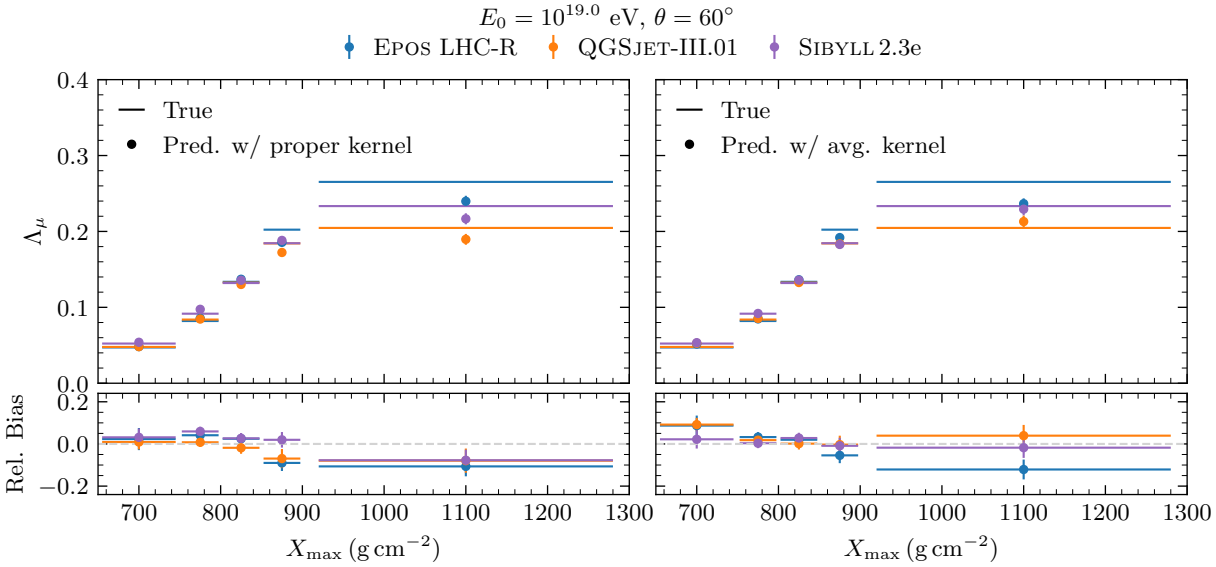

Figure 5.15: Upper panels: evolution of Λ_μ with $X_{\max}$ predicted by applying either the proper (left) or the universal probabilistic model of the shower response (right) to prior distributions $f(\alpha_1, \xi)$ provided by different hadronic interaction models. The predicted values of Λ_μ are presented by the circular markers, and the true values by the steps. Lower panels: relative bias in the prediction of Λ_μ.

the biases in Λ_μ become more dependent on the hadronic interaction model, though they remain within $\pm 15\%$ across all $X_{\max}$ bins. Nevertheless, the evolution of Λ_μ with $X_{\max}$ is well reproduced for all models, with only modest systematic uncertainties arising from the shower response modelling algorithm and the kernel averaging process. This is promising, since the evolution of Λ_μ with $X_{\max}$ is primarily governed by proton showers—the primary particle for which this mapping was developed. Therefore, the application of this framework to constrain hadronic physics at ultra-high energies via measurements of Λ_μ appears feasible in the near term.

5.3 Summary and Discussion

In this chapter, we demonstrated that the joint distribution $f(n_\mu, X_{\max})$ retains valuable information about the primary proton-air interaction, which can be interpreted in terms of the multiparticle production variables ζ_{had}, ζ_{EM}, and α_{had}.

In particular, we showed that the energy spectra of secondaries emerging from the primary interaction can be mapped onto $f(n_\mu, X_{\max})$ through the primary interaction variables $\alpha_1 \simeq \alpha_{\text{had}}$ and ξ, using a universal model of the probabilistic response function of the shower. This framework enables the direct prediction of the full distribution $f(n_\mu, X_{\max})$ from any physically consistent prior distribution $f(\alpha_1, \xi)$. Using this approach, we established the following key results:

1. The anti-correlation between α_1 and ξ, primarily induced by the energetic balance between the hadronic and electromagnetic sectors of the primary interaction, drives an anti-correlation between n_μ and $X_{\max}$.

2. The role of stochasticity in the shower development beyond the primary interaction is subdominant in shaping the joint distribution $f(n_\mu, X_{\max})$.

3. The main features of the distribution of the relative number of muons, n_μ, can be predicted with biases significantly smaller than the spread across hadronic interaction models and well within current experimental uncertainties. This includes an accurate prediction of Λ_μ based solely on the properties of the primary interaction. Notably, Λ_μ is robust to the presence of heavier primaries in the cosmic-ray flux, enabling data-driven constraints on the inclusive production cross section of neutral pions.

4. The evolution of Λ_μ with $X_{\max}$ can be reproduced with promising, though not yet optimal, precision.

Crucially, the developed and validated framework lays the foundation for using the full distribution $f(n_\mu, X_{\max})$ to constrain hadronic interactions in regions of phase space that remain inaccessible to current accelerator experiments.

The application of this framework to extensive air shower data is foreseen along two complementary directions. First, it can be applied to regions of the distribution $f(n_\mu, X_{\max})$ that are minimally contaminated by showers from heavier primaries. In particular, the universal model of the shower response already captures most of the $X_{\max}$ dependence of Λ_μ, which is largely insensitive to the mass composition of the cosmic ray flux. This enables constraints on the production cross-section of neutral pions as a function of the hadronic activity of the primary interaction.

Second, the framework can be extended to describe air-shower interactions for heavier primaries, thereby allowing for a global fit to the full measured distribution $f(n_\mu, X_{\max})$.

6

MEASURING MULTIPARTICLE PRODUCTION VARIABLES OF THE PRIMARY INTERACTION USING AUGER DATA

In this chapter, we present a set of linear relations linking the average values of the primary interaction variables ζ_{had}, ζ_{EM}, and α_{had}—both among themselves and with $\langle X_{\mathrm{max}} \rangle$—across different hadronic interaction models. We also study the dependence of these relations on the primary energy.

Moreover, we show that, in the absence of exotic physics, the primary-mass-independent shifts in $\langle X_{\mathrm{max}} \rangle$ reported in [102] enable the measurement of the average values of several primary interaction variables. This constitutes the first direct determination of such properties, providing constraints on the energy spectra of hadrons produced in ultra-high-energy proton-air interactions in regions of phase space poorly explored by accelerator experiments.

Finally, we demonstrate that this result is consistent with the observed excess of muons in data compared to simulations, illustrating the constraining power of the framework developed in Chapter 5.

6.1 Linear relations between primary interaction variables

In Section 4.4, we studied the shower-to-shower correlations between the production variables ζ_{had}, ζ_{EM} and α_{had}. We now investigate how their mean values change under consistent modifications of the energy spectra of secondary particles, while preserving fundamental conservation laws such as energy conservation and the conservation of key quantum numbers: baryon number, lepton number, isospin, and charge symmetries, among others. These consistent modifications are explored by examining the predictions of various hadronic interaction models, as shown in Figure 6.1. The figure also includes linear fits of the form $y = mx + b$ to the relations between the different pairs of variables $\langle \zeta_{\mathrm{had}} \rangle$, $\langle \zeta_{\mathrm{EM}} \rangle$, and $\langle \alpha_{\mathrm{had}} \rangle$.

It is clear that consistent modifications of the secondary energy spectra lead to correlated changes in the mean values of the multiparticle production variables, aligning, with little dispersion, along the linear fit curves. In other words, the derivatives $\mathrm{d}\langle \zeta_{\mathrm{had}} \rangle / \mathrm{d}\langle \alpha_{\mathrm{had}} \rangle$, $\mathrm{d}\langle \zeta_{\mathrm{EM}} \rangle / \mathrm{d}\langle \alpha_{\mathrm{had}} \rangle$ and $\mathrm{d}\langle \zeta_{\mathrm{had}} \rangle / \mathrm{d}\langle \zeta_{\mathrm{EM}} \rangle$ are approximately constant under consistent modi-

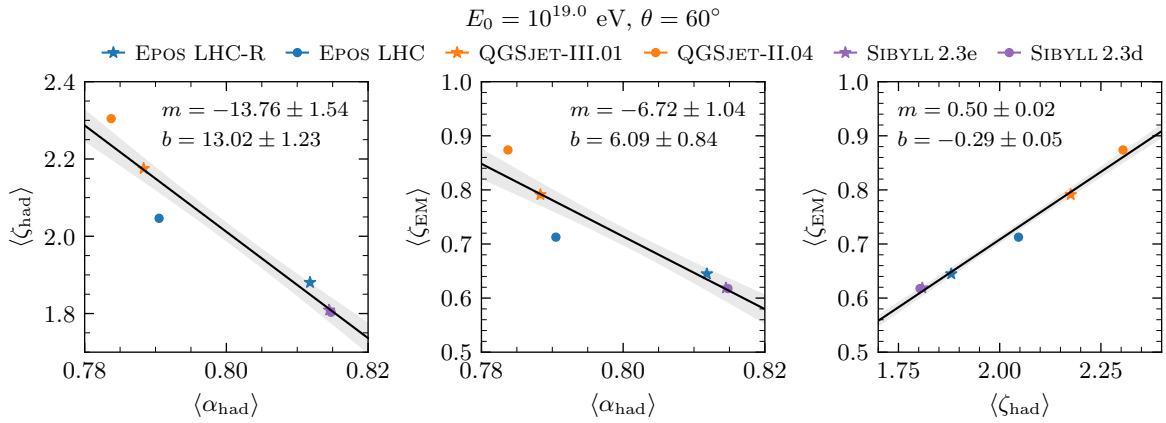

Figure 6.1: Linear relations between $\langle\zeta_{\rm had}\rangle$, $\langle\zeta_{\rm EM}\rangle$ and $\langle\alpha_{\rm had}\rangle$. The predictions of each hadronic interaction model are shown by the coloured markers, with the most recent versions highlighted by starred markers. The black lines represent linear fits to the model predictions, with uncertainty bands shown in light grey.

fications of the primary interaction. Note that both QGSJET-II.04 and EPOS LHC have a larger dispersion about the calibration curve than the more modern hadronic interaction models. Investigating the particular mechanisms responsible for this is beyond the scope of the thesis.

This does not imply that these variables fail to provide independent information about the energy spectrum of secondary hadrons. Rather, their mean values do not vary independently across models, reflecting the common underlying physical mechanisms implemented in the different hadronic interaction models. In the absence of exotic physics, it is therefore expected that the true mean values of $\zeta_{\rm had}$, $\zeta_{\rm EM}$ and $\alpha_{\rm had}$ lie along the calibration curves shown in Figure 6.1. These relations can thus be interpreted as a form of universal scaling law. We have also verified that similar relations hold at lower energies. This fact is essential in applying the framework developed in this thesis to data spanning a range of primary energies.

A natural next step would be to implement controlled, physically consistent modifications of the primary interaction itself, to understand which specific changes are responsible for the observed universality. This study is deferred to future work.

6.2 Energy dependence of the calibration between $X_{\max}$ and ξ

In order to apply the universal scaling law between $\langle X_{\max}\rangle$ and $\langle\xi\rangle$ (see Figure 4.21) in data-driven studies of hadronic interactions, we must first quantify its dependence on the primary energy. To this end, we consider primary energies E_0 ranging from 10^{17} to 10^{19} eV, in steps of $\Delta\log_{10}(E_0/\text{eV}) = 0.5$. Figure 6.2 shows the pairs of values $(\langle\xi_{19}\rangle, \langle X_{19}\rangle)$ predicted by a selection of recent hadronic interaction models, along with linear fits for three different primary energies. The fits have the form $\langle X_{19}\rangle = m\,\langle\xi_{19}\rangle + b$.

At lower primary energies, the predictions of the different hadronic interaction models tend to converge. Additionally, the slope m of the linear relation between $\langle\xi\rangle$ and $\langle X_{\max}\rangle$ shows a mild, albeit statistically insignificant increase as energy decreases, as seen in the

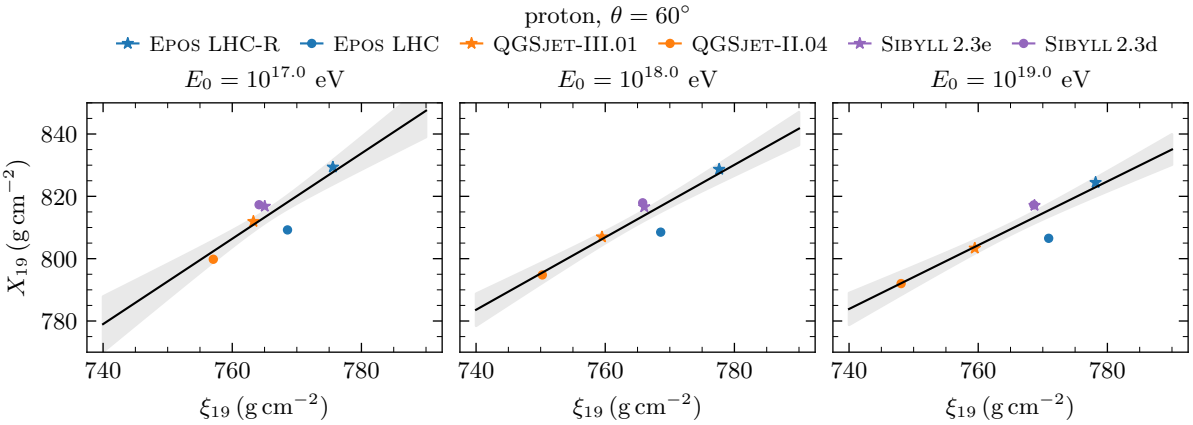

Figure 6.2: Linear relations between $\langle \xi_{19} \rangle$ and $\langle X_{19} \rangle$. The predictions of each hadronic interaction model are shown by coloured markers, with the most recent versions indicated by starred markers. The black lines represent linear fits to the model predictions, with uncertainty bands shown in light grey. Each panel corresponds to a different primary energy, increasing from left to right.

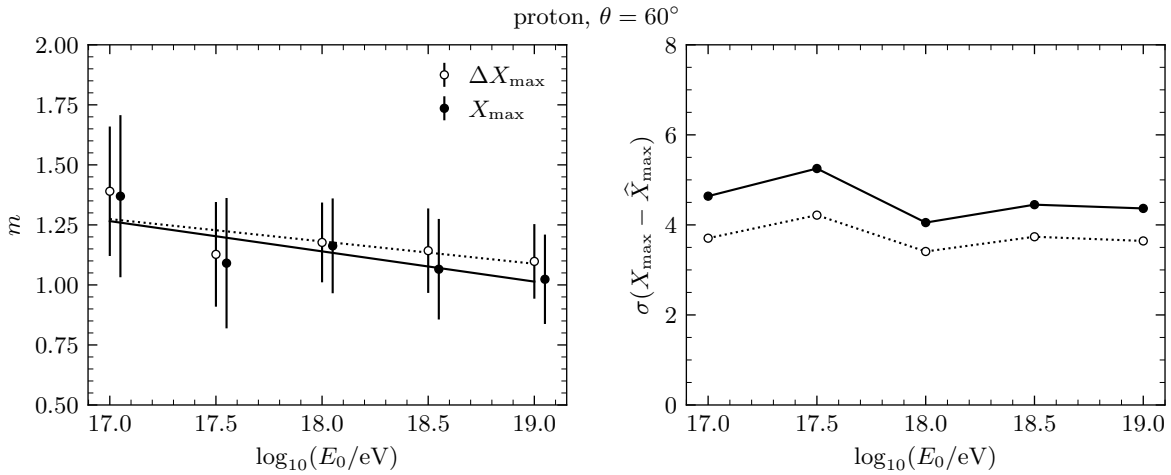

Figure 6.3: Left: Energy dependence of the slope of the $\langle \xi \rangle$–$\langle X_{\max} \rangle$ calibration (solid markers) and of the $\langle \xi \rangle$–$\langle \Delta X_{\max} \rangle$ calibration (open markers). Right: Energy dependence of the dispersion of the hadronic interaction models around the respective calibration curves.

left panel of Figure 6.3. This behaviour suggests that the predictions for ξ may converge faster than those for $\langle X_{\max} \rangle$. A similar trend is observed in the calibration between ξ and $\Delta X_{\max}$, also shown in Figure 6.3, indicating that this effect is unlikely to be caused by the energy evolution of the proton–air cross-section.

The right panel of Figure 6.3 shows the standard deviation of the residuals between the true model predictions of $X_{\max}$ and the estimated value $\langle \widehat{X}_{\max} \rangle = m \langle \xi \rangle + b$, and analogously for $\Delta X_{\max}$. This quantifies the dispersion of the model predictions around the calibration curves and reveals that it remains approximately constant across the considered energy range.

These results not only confirm the strong causal connection between ξ and $X_{\max}$, but also highlight that hadronic interaction models implement the energy dependence of hadron production in a similar, almost universal way. The persistence of this universality

at lower energies further supports its validity at higher energies, where primary cosmic-ray interactions induce lower-energy secondary interactions in the cascade. A break in this universality at lower energies should therefore lead to its break at higher energies.

Finally, this energy independence ensures that the shower response model introduced in Chapter 5 can be reliably generalized to a broad range of primary energies. This feature is essential for applying the framework to EAS data, especially since typical ultra-high-energy cosmic ray observatories probe a span of about three orders of magnitude in energy.

6.3 Application to Auger data

Recently, the Pierre Auger Collaboration reported that the best fit to the joint distribution of $X_{\max}$ and ground particle signals requires a mass-independent shift in $X_{\max}$ and a rescaling of the muon content [102].

6.3.1 Measuring the new hadronic interaction variables

Using the universal relation between $\langle\xi\rangle$ and $\langle X_{\max}\rangle$, designed for proton-initiated showers, we can assess how much of the shifts in the latter arise from differences in secondary particle production during the first proton–air interaction. Recall that this universal behaviour was observed with a collection of hadronic interaction models. This strategy ensures that the explored parameter space reflects physically motivated hadronic scenarios. Complementary approaches, such as those in [55, 146], explore a broader phase space by varying interaction parameters freely, allowing for the investigation of more extreme or unconstrained configurations. In contrast, our method focuses on configurations derived directly from established hadronic interaction models, providing a controlled comparison rooted in consistently modelled physics.

Given that the $\sigma_{p-\text{air}}$ measured by Auger agrees with model predictions within uncertainties [56], we can safely extrapolate the linear relation between $\langle X_{\max}\rangle$ and $\langle\xi\rangle$ (see Figure 4.21) to convert the shifted $\langle X_{\max}\rangle$ values into corresponding $\langle\xi\rangle$ measurements. Figure 6.4 illustrates this procedure: the empty markers near the y-axis indicate the shifted $\langle X_{\max}\rangle$ values, while those near the x-axis show the corresponding $\langle\xi\rangle$ values, obtained for each hadronic interaction model. Statistical uncertainties are indicated by brackets, and systematic ones by error bars, with $\langle X_{\max}\rangle$ uncertainties quoted from [102]. The uncertainties in the estimated values of $\langle\xi\rangle$ result from the propagation of the statistical and systematic uncertainties in $\langle X_{\max}\rangle$, accounting for the covariance matrix of the linear fit parameters. To determine a final value of $\langle\xi\rangle$ favoured by the data, we compute the average of the shifted values obtained for each hadronic interaction model. The uncertainty of these values is obtained by summing in quadrature the individual statistical and systematic uncertainties over the hadronic interaction models, along with the inter-model spread. This spread mainly reflects the fact that different hadronic interaction models, even when assuming a fixed primary composition, predict distinct higher moments of the joint distribution of the ground signal and $X_{\max}$. As the method employed in [102] only allows for freedom in the first moment of $X_{\max}$, residual discrepancies in higher moments across models may introduce biases in the shifted $X_{\max}$ values. While such biases are encompassed within the quoted systematic uncertainties, they may still influence the central

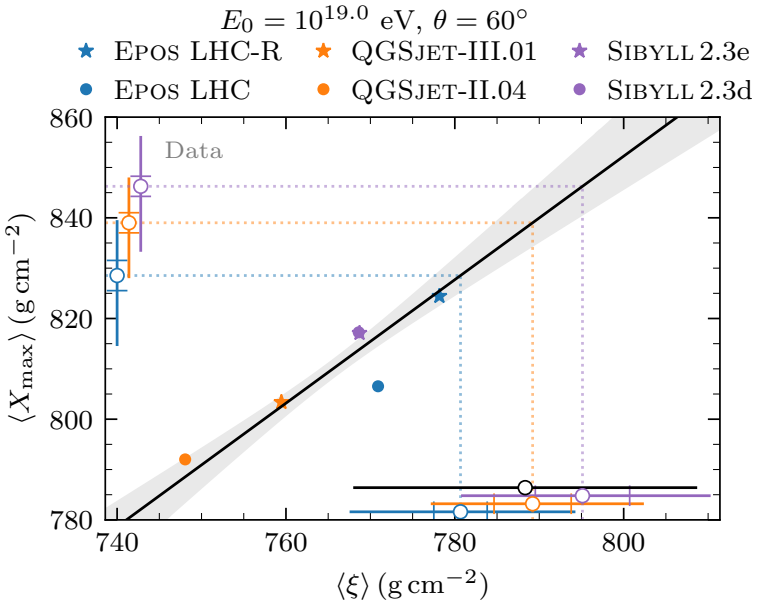

Figure 6.4: $\langle \xi \rangle$ vs. $\langle X_{\max} \rangle$ for different high-energy hadronic interaction models. The grey band represents the standard deviation of the models relative to the regression curve (solid black). Empty markers ("Data") are the shifted $\langle X_{\max} \rangle$ values from [102] for EPOS LHC, QGSJET-II.04, and SIBYLL2.3d, together with their projection to the x-axis and their average (black marker).

value of the shifts. This final result is represented by the black marker.

Despite the large uncertainty, the final $\langle \xi \rangle$ value excludes both the QGSJET-II.04 and QGSJET-III.01 models. This is in agreement with the known unphysical values of variance of $\ln A$ derived from $X_{\max}$ data interpreted with QGSJET-II.04 [76].

Given the connection between $X_{\max}$ and the multiparticle production variables $\langle \zeta_{\text{had}} \rangle$, $\langle \zeta_{\text{EM}} \rangle$, $\langle \alpha_{\text{had}} \rangle$, and $\langle \zeta_{\text{EM}}/(1-\alpha_{\text{had}}) \rangle$, as well as the universality of the $\langle \xi \rangle$–$\langle X_{\max} \rangle$ relation, it is natural to study how $X_{\max}$ scales with these production variables.

Figure 6.5 shows the dependence of $\langle X_{\max} \rangle$ on $\langle \alpha_{\text{had}} \rangle$, $\langle \zeta_{\text{had}} \rangle$, $\langle \zeta_{\text{EM}} \rangle$, and $\omega / \langle \zeta_{\text{had}} \rangle$ (see Equation (4.13) for the definition of ω). The figure also shows linear fits to the model predictions. Note, once again, that the mentioned dependence is examined via physically consistent changes in the primary interaction, sampled by recent hadronic models.

Regardless of the chosen variable, a universal linear relation exists between its mean value and $\langle X_{\max} \rangle$. The standard deviations of the dispersion of the models about these calibration curves range from 2.5 g cm^{-2} and 4.2 g cm^{-2}, thus allowing the projection of the shifted $\langle X_{\max} \rangle$ values onto each production variable. Table 6.1 reports these projections for EPOS LHC, QGSJET-II.04, and SIBYLL2.3d, along with the nominal values from more recent models EPOS LHC-R, QGSJET-III.01, and SIBYLL2.3e.

Only SIBYLL2.3d/SIBYLL2.3e and EPOS LHC-R are compatible with all production variables preferred by the data, despite the large uncertainties. The smaller $\langle \zeta_{\text{had}} \rangle$ and $\langle \zeta_{\text{EM}} \rangle$ suggest harder energy spectra in the primary interaction, possibly due to more frequent diffractive p–air interactions, which reduce hadronic activity. This scenario is consistent with the observed increase in $\langle \alpha_{\text{had}} \rangle$. This increase is, in turn, consistent with the enhanced muon scale found in data, as detailed in the next section.

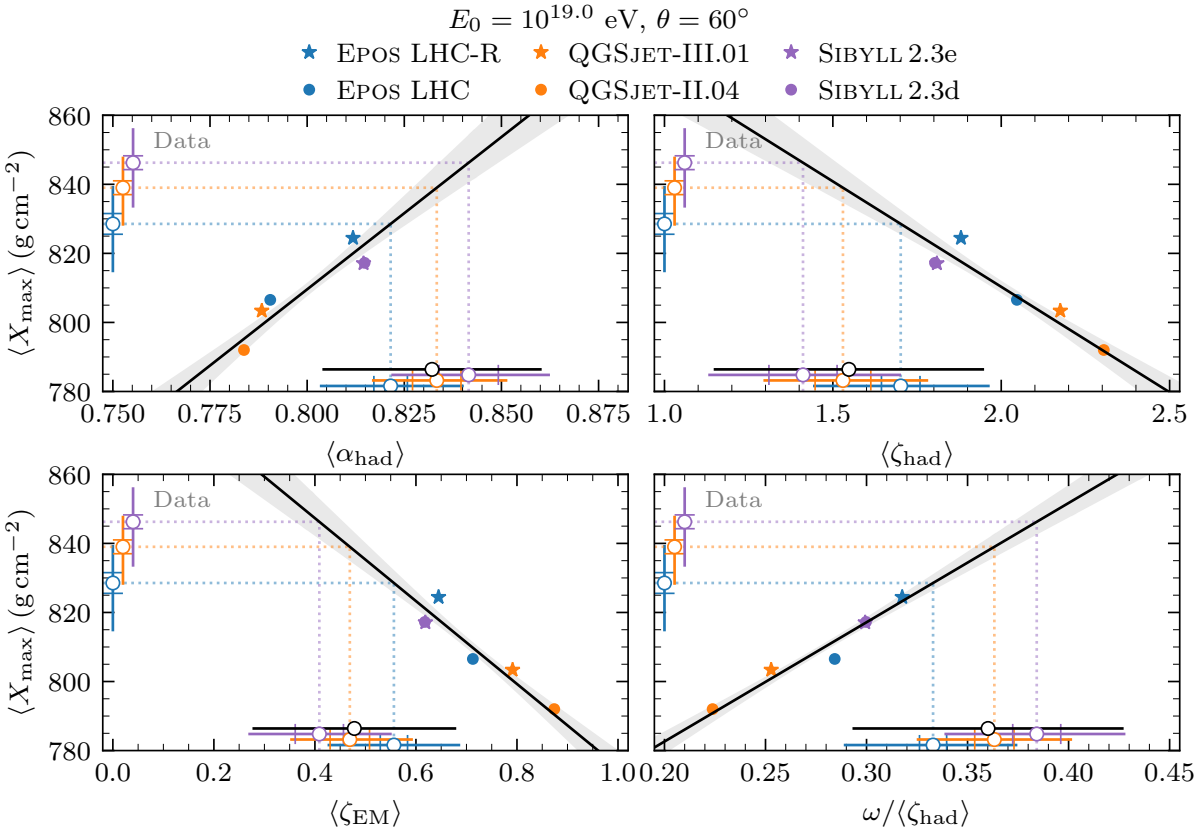

Figure 6.5: $\langle X_{\max} \rangle$ vs. the new multiparticle production variables: $\langle \alpha_{\rm had} \rangle$ (top left), $\langle \zeta_{\rm had} \rangle$ (top right), $\langle \zeta_{\rm EM} \rangle$ (bottom left), and $\omega / \langle \zeta_{\rm had} \rangle$ (bottom right), for different high-energy hadronic interaction models. The grey band represents the standard deviation of the models relative to the regression curve (solid black). Empty markers ("Data") are the shifted $\langle X_{\max} \rangle$ values from [102], together with their projection to the x-axis and their average (black marker).

Consistent re-scaling of the muon content

To reproduce the required shifts in the $X_{\max}$ scale, the mean value of $\langle \alpha_{\rm had} \rangle$ must increase by 4%, 7%, and 5% for the hadronic interaction models EPOS LHC, QGSJET-II.04, and SIBYLL2.3d, respectively. This increase transfers a larger fraction of the primary energy to the hadronic core of the shower, leading to an enhanced muon yield, as described by Equation (1.44). We now estimate how a change in $\langle \alpha_{\rm had} \rangle$ propagates to the mean number of muons, $\langle N_\mu \rangle$. Assuming a shift of the form $\ln \langle \alpha_{\rm had} \rangle \to \ln \langle \alpha_{\rm had} \rangle + \delta \ln \langle \alpha_{\rm had} \rangle$, we propagate this modification to later shower generations g as $\ln \alpha_{{\rm had},g} \to \ln \alpha_{{\rm had},g} + f_g \times \delta \ln \langle \alpha_{\rm had} \rangle$, where f_g denotes a generation-dependent modification factor, equal to 1 for $g = 1$ and progressively decreasing with increasing g. This assumed decrease stems from the expected convergence of the hadronic interaction models at lower energies.

Using Equation (1.44), the relative change in the mean number of muons is given by

$$\delta \ln \langle N_\mu \rangle = g_{\rm mod} \times \delta \ln \langle \alpha_{\rm had} \rangle \,, \tag{6.1}$$

where $g_{\rm mod} = \sum f_g$ is the effective cumulative number of generations contributing to the change. That is, if the value of $\alpha_{{\rm had},g}$ were to change as much as that of $\alpha_{\rm had}$ for all generations, then $g_{\rm mod}$ would be the number of generations needed to be changed.

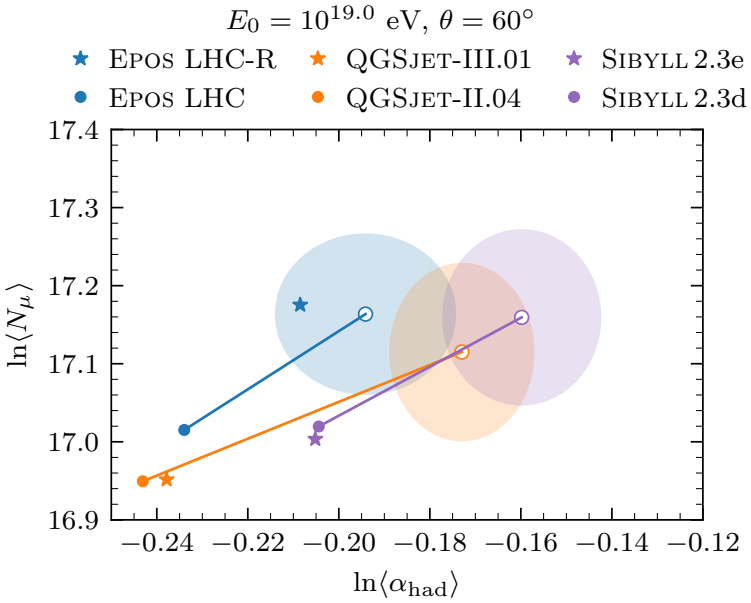

Figure 6.6: $\ln\langle\alpha_{\rm had}\rangle$ vs $\ln\langle N_\mu\rangle$ for different high-energy hadronic interaction models. The solid markers correspond to the default values for each hadronic interaction model and the empty markers to the shifted values determined from [102] and this work. Extracted from the solid lines, the cumulative number of generations needed to propagate the change in $\langle\alpha_{\rm had}\rangle$ to $\ln\langle N_\mu\rangle$ is: $g_{\rm mod} = \{3.7^{+5.1}_{-2.7}, 2.4^{+1.9}_{-1.6}, 3.1^{+3.7}_{-2.5}\}$ for EPOS LHC, QGSJET-II.04 and SIBYLL2.3d, respectively.

Equation (6.1) allows us to estimate the required $g_{\rm mod}$ to reproduce the observed muon rescaling factors of $16 \pm 1^{+20}_{-16}\%$, $18 \pm 1^{+15}_{-12}\%$, and $15 \pm 1^{+15}_{-11}\%$ for EPOS LHC, QGSJET-II.04, and SIBYLL2.3d, respectively, as reported in [102], for $\theta = 52°$. The relation between $\ln\langle\alpha_{\rm had}\rangle$ and $\ln\langle N_\mu\rangle$ for various hadronic interaction models, along with their shifted values and 1σ uncertainties, is shown in Figure 6.6.

The values obtained for the cumulative number of generations are $g_{\rm mod} = 3.7^{+5.1}_{-2.7}$ for EPOS LHC, $g_{\rm mod} = 2.4^{+1.9}_{-1.6}$ for QGSJET-II.04, and $g_{\rm mod} = 3.1^{+3.7}_{-2.5}$ for SIBYLL2.3d. These values are mutually compatible within uncertainties. Moreover, we note that EPOS LHC-R, the most recent version of EPOS LHC, naturally satisfies the required shifts in both $\alpha_{\rm had}$ and N_μ.

6.3.2 Traditional production variables

As mentioned in Section 4.2.2, there are other estimators of $\langle X_{\rm max}\rangle$ in the literature, such as that described in [119]

$$K = X_1 + \lambda_r \ln\left(\frac{E_0}{2\xi_c^e}\right) + \lambda_r \ln\left(\frac{\kappa_{\rm el}}{m_{\rm tot}}\right). \tag{6.2}$$

Despite the poorer shower-to-shower correlation between K and $X_{\rm max}$, when compared to ξ, its average value reproduces extremely well that of $X_{\rm max}$, as seen in Figure 6.7. In particular, the standard deviation of the residuals of the hadronic interaction models relative to the calibration $\langle K\rangle - \langle X_{\rm max}\rangle$ is solely $2\,{\rm g\,cm^{-2}}$, allowing $\langle K\rangle$ to be measured using the shifted values of $\langle X_{\rm max}\rangle$. Again, this conversion assumes the absence of exotic

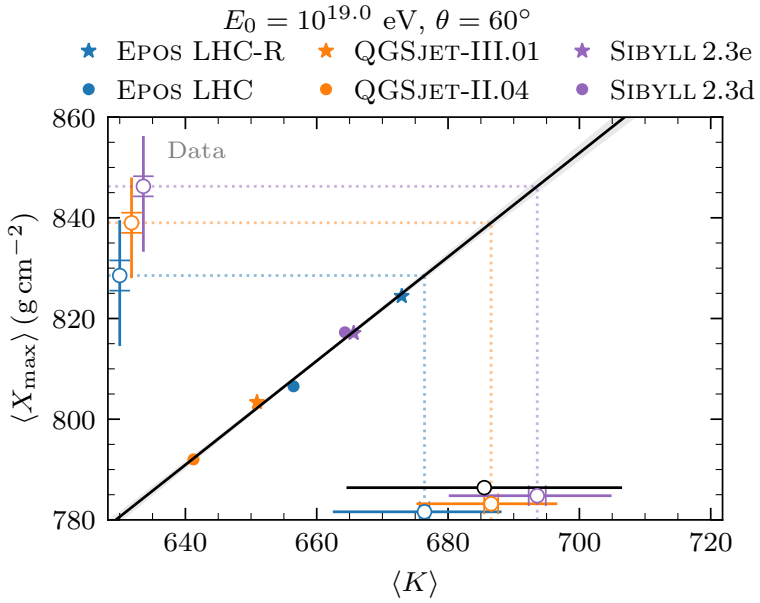

Figure 6.7: $\langle K \rangle$ vs $\langle X_{\max} \rangle$ for different high-energy hadronic interaction models. The grey band represents the standard deviation of the models relative to the regression curve (solid black). Empty markers ("Data") are the shifted values of $\langle X_{\max} \rangle$ determined in [102] for EPOS LHC, QGSJET-II.04 and SIBYLL2.3d together with their projection to the x-axis and their average (black marker).

physics at the highest energies and the primary mass independence of the shifts.

Similarly, there is a linear relation between $\langle \ln \kappa_{\rm el} \rangle$ and $\langle X_{\max} \rangle$ across hadronic interaction models with a dispersion of 4.4 g cm^{-2}. This is shown in Figure 6.8. This dispersion is compatible with that of the production variables included in this work, allowing for the measurement of $\langle \kappa_{\rm el} \rangle$ from the shifted values of $\langle X_{\max} \rangle$.

The dispersion of the true values of $\langle X_{\max} \rangle$ about the calibration $\langle \ln m_{\rm tot} \rangle - \langle X_{\max} \rangle$ has a standard deviation of 7.2 g cm^{-2}, which is considerably larger than the dispersions obtained for the $\langle K \rangle$–$\langle X_{\max} \rangle$ and $\langle \ln \kappa_{\rm el} \rangle$–$\langle X_{\max} \rangle$ correlations. For completeness, we therefore constrain $\langle \ln m_{\rm tot} \rangle$ by employing Equation (6.2) and using the relation $\langle \ln (\kappa_{\rm el}/m_{\rm tot}) \rangle = \langle \ln \kappa_{\rm el} \rangle - \langle \ln m_{\rm tot} \rangle$, which allows us to express the multiplicity in terms of the data values of $\langle K \rangle$ and $\langle \ln \kappa_{\rm el} \rangle$.

6.3.3 Final measurement

The final values of all the considered hadronic interaction variables are shown in Table 6.1.

This represents the first measurement of this kind using EAS data. It relies on a collection of universal relations between the scale of $X_{\max}$ and a collection of hadronic interaction variables constraining the spectrum of hadrons of the primary interaction. It also relies on the assumptions of mass-independence of the shifts in $\langle X_{\max} \rangle$ preferred by Auger data [147] and the absence of exotic physics phenomena appearing at the EeV-scale. Crucially, the increase in $\langle \alpha_{\rm had} \rangle$ is coherent with the observed muon excess with respect to shower simulations.

As a concluding remark, we note that this study only concerns the mean values of $\langle X_{\max} \rangle$, while the framework presented in Chapter 5 concerns all the information in the

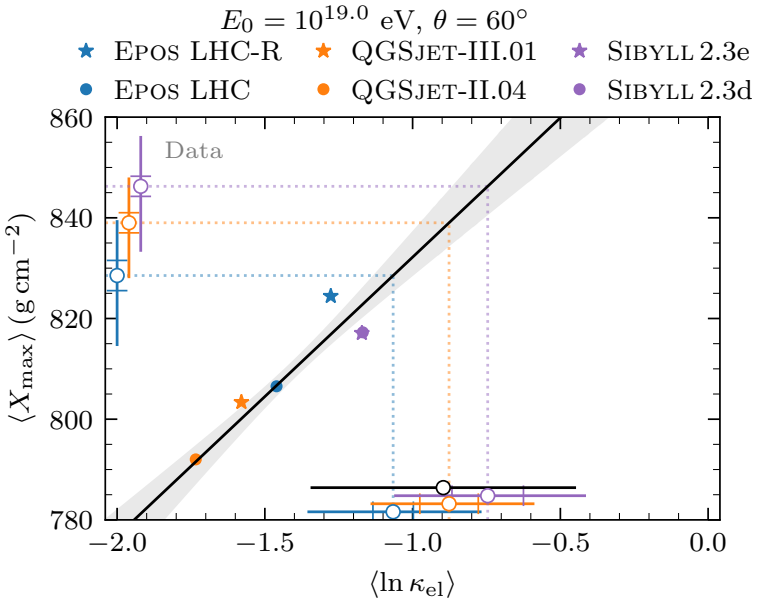

Figure 6.8: $\langle X_{\max} \rangle$ vs the multiparticle production variables $\langle \ln \kappa_{\rm el} \rangle$ for different high-energy hadronic interaction models. The grey band represents the standard deviation of the models relative to the regression curve (solid black). Empty markers ("Data") are the shifted values of $\langle X_{\max} \rangle$ determined in [102] for EPOS LHC, QGSJET-II.04 and SIBYLL2.3d together with their projection to the x-axis and their average (black marker).

Model	$\left\langle \ln \frac{\kappa_{\rm el}}{m_{\rm tot}} \right\rangle$	$\langle \ln \kappa_{\rm el} \rangle$	$\langle \ln m_{\rm tot} \rangle$	$\langle \alpha_{\rm had} \rangle$	$\langle \zeta_{\rm had} \rangle$	$\langle \zeta_{\rm EM} \rangle$	$\left\langle \frac{\zeta_{\rm EM}}{\alpha_{\rm EM}} \right\rangle$
EPOS-R	-6.61	-1.28	5.33	0.81	1.88	0.64	3.56
Q-III.01	-7.21	-1.58	5.63	0.79	2.18	0.79	3.85
SIB2.3e	-6.81	-1.17	5.64	0.81	1.81	0.62	3.95
EPOS	-7.08	-1.47	5.61	0.79	2.06	0.72	3.68
Q-II.04	-7.60	-1.73	5.87	0.78	2.30	0.87	3.99
SIB2.3d	-6.90	-1.17	5.72	0.82	1.81	0.62	3.97
Data	-6.3 ± 0.6	-0.9 ± 0.4	5.4 ± 0.3	0.83 ± 0.03	1.5 ± 0.4	0.5 ± 0.2	4 ± 1

Table 6.1: Numerical values of several multi-particle production variables of primary proton-air interactions derived from the shifts in $\langle X_{\max} \rangle$ derived in [102] for the hadronic interaction models EPOS LHC, QGSJET-II.04 and SIBYLL2.3d. The table also shows the default values of these variables, together with the averaged prediction from data.

joint distribution of $f(n_\mu, X_{\max})$. This evidences the constraining power of the collection of variables we developed through this work, and opens the possibility for an even more complete description of hadronic interactions based on EAS data.

6.4 Summary and Discussion

In this chapter, we have shown that there are universal relations between the average depth of the shower maximum, $\langle X_{\max} \rangle$, and key variables that characterize the first proton–air interaction. This universality was verified to be largely independent of the primary

energy and emerges because modern hadronic interaction models, despite differences in technical implementation—such as in the modelling of diffraction, hadronization, or energy partitioning—are rooted in the same fundamental physics. As a result, the models collectively sample a physically plausible phase space of hadronic interactions, enabling robust calibrations between multiparticle production observables and $\langle X_{\mathrm{max}} \rangle$.

We also showed that the shifts in $\langle X_{\mathrm{max}} \rangle$ reported by the Pierre Auger Collaboration can be mapped into changes in the secondary particle production during the primary interaction. In the absence of new physics, and assuming the primary-mass independence of those shifts, it is possible to decompose a shift in $\langle X_{\mathrm{max}} \rangle$ into shifts in $\langle m_{\mathrm{tot}} \rangle$ and $\langle \kappa_{\mathrm{el}} \rangle$, as well as $\langle \alpha_{\mathrm{had}} \rangle$, $\langle \zeta_{\mathrm{had}} \rangle$, and $\langle \zeta_{\mathrm{EM}} \rangle$ and $\omega / \langle \zeta_{\mathrm{had}} \rangle$ and consistently account for the observed re-scaling of $\langle N_{\mu} \rangle$. Given the modified value of $\langle \alpha_{\mathrm{had}} \rangle$ obtained from the shifts in $\langle X_{\mathrm{max}} \rangle$, we also computed the number of cumulative generations over which the fraction of hadronic energy must be changed, to account for the measured re-scaling of the number of muons.

In summary, our framework allowed the measurement of several multiparticle production variables of the primary interaction and constrained the modifications in the secondary energy spectra required to explain the observed shifts in both $\langle X_{\mathrm{max}} \rangle$ and $\langle N_{\mu} \rangle$. As such, we imposed multiple constraints on hadron production in proton-air interactions beyond the reach of current accelerators.

Conclusions

This thesis presented a comprehensive study of the information encoded in the joint distribution of the number of muons at ground level, N_μ, and the depth of the shower maximum, $X_{\max}$, with the goal of probing hadronic interactions in the ultra-high-energy regime beyond the reach of current accelerators. The work culminated in the development of a probabilistic framework that maps the energy spectra of hadrons produced in proton-air interactions onto the two-dimensional distribution of N_μ and $X_{\max}$. This model accurately reproduces the joint distribution for any physically consistent prior describing the primary interaction, achieving a precision far better than the differences between hadronic interaction models or typical experimental systematic uncertainties. In particular, expected values of multiparticle production variables derived from these spectra were estimated using data from the Pierre Auger Observatory, providing a coherent interpretation of the observations.

We first quantified the influence of cascade dynamics on the two-dimensional distribution of N_μ and $X_{\max}$, and isolated features arising from the stochastic nature of hadron production throughout the shower development. The latter is captured by fluctuations of the number of produced muons, which can be propagated to the ground level using a simple muon attenuation model. This model incorporates the spectral hardness at production, γ, and the position of the maximum of the muon production profile. Furthermore, we demonstrated that the increasing anti-correlation between N_μ and $X_{\max}$ with energy is driven solely by hadron production physics, and their joint distribution is significantly affected by the zenith angle due to the truncation of the muon profile. We further identified the regions of $f(N_\mu, X_{\max})$ dominated by proton primaries, and showed that the corresponding distributions for heavier nuclei cannot be reproduced through a simple superposition model, precluding a straightforward generalisation of the developed results.

We demonstrated that the shape of the muon number distribution in muon-depleted showers, as described by the parameter Λ_μ, evolves with $X_{\max}$ in a way that reflects the hadronic activity of the primary interaction. This evolution is highly sensitive to the properties of hadron production, enabling us to probe the inclusive cross-section for neutral pion production in proton-air interactions. We also showed that this measurement is experimentally feasible with current extensive-air-shower detectors, and is robust even in the presence of a mixed cosmic-ray composition.

Moreover, we extended existing semi-analytical models of air showers by introducing a novel set of multiparticle production variables: ζ_{had}, ζ_{EM}, and α_{had}. These were shown to characterise key features of the hadronic energy spectra in proton-air interactions, which can be directly measured in the phase-space covered by accelerator experiments. Furthermore, they were combined into a single variable, ξ, which captures most of the variability in $X_{\max}$ attributable to the primary proton-air interaction.

For proton-induced showers, we demonstrated that ξ and α_1, the primary interaction estimator of the relative muon scale, can be used to build a hadronic-interaction-model-independent probabilistic description of the shower response function, linking secondary hadron energy spectra to $f(N_\mu, X_{\max})$. Using Monte-Carlo-generated proton-air interac-

tions as priors, the framework predicts $f(N_\mu, X_{\max})$ within current systematic uncertainties and with sufficient precision to discriminate between hadronic interaction models. It also accurately reproduces the proton-dominated regions of $f(N_\mu, X_{\max})$, enabling a study of the $X_{\max}$-evolution of Λ_μ directly in connection with the primary interaction. A natural extension of this framework to heavier primaries, incorporating, for example, nuclear fragmentation processes, could allow precise hadronic interaction measurements using the full joint distribution $f(N_\mu, X_{\max})$.

Finally, we applied the framework to data published by the Pierre Auger Observatory, extracting for the first time measurements of the moments of the energy spectra of secondary hadrons in ultra-high-energy proton-air interactions. These results illustrate the feasibility and potential of inferring hadronic physics directly from air-shower observables, without relying solely on extrapolations from specific interaction models.

Bibliography

[1] R. Alves Batista et al. "Open Questions in Cosmic-Ray Research at Ultrahigh Energies". In: *Front. Astron. Space Sci.* 6 (2019), p. 23. DOI: 10.3389/fspas.2019.00023. arXiv: 1903.06714 [astro-ph.HE].

[2] A. Aab et al. "Observation of a Large-scale Anisotropy in the Arrival Directions of Cosmic Rays above 8×10^{18} eV". In: *Science* 357.6537 (2017), pp. 1266–1270. DOI: 10.1126/science.aan4338. arXiv: 1709.07321 [astro-ph.HE].

[3] A. Coleman et al. "Ultra high energy cosmic rays The intersection of the Cosmic and Energy Frontiers". In: *Astropart. Phys.* 149 (2023), p. 102819. DOI: 10.1016/j.astropartphys.2023.102819. arXiv: 2205.05845 [astro-ph.HE].

[4] P. Mészáros et al. "Multi-Messenger Astrophysics". In: *Nature Rev. Phys.* 1 (2019), pp. 585–599. DOI: 10.1038/s42254-019-0101-z. arXiv: 1906.10212 [astro-ph.HE].

[5] A. Condorelli and S. Petrera. "Modeling hadronic interactions in ultra-high-energy cosmic rays within astrophysical environments: A parametric approach". In: *Astropart. Phys.* 165 (2025). [Erratum: Astropart.Phys. 165, 103063 (2025)], p. 103047. DOI: 10.1016/j.astropartphys.2024.103047. arXiv: 2405.02658 [astro-ph.HE].

[6] J. Dörner et al. "Uncertainties in astrophysical gamma-ray and neutrino fluxes from proton-proton cross-sections in the GeV to PeV range". In: *JCAP* 04 (2025), p. 043. DOI: 10.1088/1475-7516/2025/04/043. arXiv: 2501.16967 [astro-ph.HE].

[7] V. F. Hess. "Über Beobachtungen der durchdringenden Strahlung bei sieben Freiballonfahrten". In: *Phys. Z.* 13 (1912), pp. 1084–1091.

[8] C. D. Anderson. "The Positive Electron". In: *Phys. Rev.* 43 (1933), pp. 491–494. DOI: 10.1103/PhysRev.43.491.

[9] S. H. Neddermeyer and C. D. Anderson. "Note on the Nature of Cosmic Ray Particles". In: *Phys. Rev.* 51 (1937), pp. 884–886. DOI: 10.1103/PhysRev.51.884.

[10] C. M. G. Lattes, G. P. S. Occhialini and C. F. Powell. "Observations on the Tracks of Slow Mesons in Photographic Emulsions. 1". In: *Nature* 160 (1947), pp. 453–456. DOI: 10.1038/160453a0.

[11] G. D. Rochester and C. C. Butler. "Evidence for the Existence of New Unstable Elementary Particles". In: *Nature* 160 (1947), pp. 855–857. DOI: 10.1038/160855a0.

[12] B. Rossi. "Misure sulla distribuzione angolare di intensità della radiazione penetrante all'Asmara". In: *Supplemento a La Ricerca Scientifica* 1 (1934), p. 579.

[13] P. Auger, R. Maze and J. Daudin. "Extension et pouvoir pénétrant des grandes gerbes cosmiques". In: *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 208 (1939), pp. 1641–1643.

[14] P. Auger et al. "Extensive cosmic ray showers". In: *Rev. Mod. Phys.* 11 (1939), pp. 288–291. DOI: 10.1103/RevModPhys.11.288.

[15] K.-H. Kampert, A. A. Watson and A. A. Watson. "Extensive Air Showers and Ultra High-Energy Cosmic Rays: A Historical Review". In: *Eur. Phys. J. H* 37 (2012), pp. 359–412. DOI: 10.1140/epjh/e2012-30013-x. arXiv: 1207.4827 [physics.hist-ph].

[16] J. Albrecht et al. "The Muon Puzzle in cosmic-ray induced air showers and its connection to the Large Hadron Collider". In: *Astrophys. Space Sci.* 367.3 (2022), p. 27. DOI: 10.1007/s10509-022-04054-5. arXiv: 2105.06148 [astro-ph.HE].

[17] 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: *The European Physical Journal C* 81.11 (2021), p. 966. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-021-09700-w.

[18] Z. Cao et al. "Measurements of All-Particle Energy Spectrum and Mean Logarithmic Mass of Cosmic Rays from 0.3 to 30 PeV with LHAASO-KM2A". In: *Phys. Rev. Lett.* 132.13 (2024), p. 131002. DOI: 10.1103/PhysRevLett.132.131002. arXiv: 2403.10010 [astro-ph.HE].

[19] Z. Cao et al. "First Identification and Precise Spectral Measurement of the Proton Component in the Cosmic-Ray 'Knee'". In: (May 2025). arXiv: 2505.14447 [astro-ph.HE].

[20] T. K. Gaisser, R. Engel and E. Resconi. *Cosmic Rays and Particle Physics: 2nd Edition*. Cambridge University Press, June 2016. ISBN: 978-0-521-01646-9.

[21] B. Peters. "Primary cosmic radiation and extensive air showers". In: *Nuovo Cim.* 22.4 (1961), pp. 800–819. DOI: 10.1007/bf02783106.

[22] Z. Cao et al. "Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources". In: *Nature* 594.7861 (2021), pp. 33–36. DOI: 10.1038/s41586-021-03498-z.

[23] V. Novotný. "Energy spectrum of cosmic rays measured using the Pierre Auger Observatory". In: *PoS* ICRC2021 (2021), p. 324. DOI: 10.22323/1.395.0324.

[24] W. D. Apel et al. "Kneelike structure in the spectrum of the heavy component of cosmic rays observed with KASCADE-Grande". In: *Phys. Rev. Lett.* 107 (2011), p. 171104. DOI: 10.1103/PhysRevLett.107.171104. arXiv: 1107.5885 [astro-ph.HE].

[25] S. Mollerach and E. Roulet. "A scenario for the Galactic cosmic rays between the knee and the second-knee". In: *JCAP* 03 (2019), p. 017. DOI: 10.1088/1475-7516/2019/03/017. arXiv: 1812.04026 [astro-ph.HE].

[26] A. Aab et al. "Features of the Energy Spectrum of Cosmic Rays above 2.5×10^{18} eV Using the Pierre Auger Observatory". In: *Phys. Rev. Lett.* 125 (12 2020), p. 121106. DOI: 10.1103/PhysRevLett.125.121106.

[27] A. A. Halim et al. "Constraining models for the origin of ultra-high-energy cosmic rays with a novel combined analysis of arrival directions, spectrum, and composition data measured at the Pierre Auger Observatory". In: *JCAP* 01 (2024), p. 022. DOI: 10.1088/1475-7516/2024/01/022. arXiv: 2305.16693 [astro-ph.HE].

[28] K. Greisen. "End to the Cosmic-Ray Spectrum?" In: *Phys. Rev. Lett.* 16 (17 1966), pp. 748–750. DOI: 10.1103/PhysRevLett.16.748. URL: https://link.aps.org/doi/10.1103/PhysRevLett.16.748.

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

[30] R. Aloisio. "Acceleration and propagation of ultra-high energy cosmic rays". In: *Progress of Theoretical and Experimental Physics* 2017.12 (Nov. 2017). 12A102. ISSN: 2050-3911. DOI: 10.1093/ptep/ptx115. eprint: https://academic.oup.com/ptep/article-pdf/2017/12/12A102/22075738/ptx115.pdf. URL: https://doi.org/10.1093/ptep/ptx115.

[31] A. Yushkov. "Mass Composition of Cosmic Rays with Energies above $10^{17.2}$ eV from the Hybrid Data of the Pierre Auger Observatory". In: *PoS* ICRC2019 (2020), p. 482. DOI: 10.22323/1.358.0482.

[32] A. M. Hillas. "The Origin of Ultrahigh-Energy Cosmic Rays". In: *Ann. Rev. Astron. Astrophys.* 22 (1984), pp. 425–444. DOI: 10.1146/annurev.aa.22.090184.002233.

[33] G. R. Farrar. "Binary Neutron Star Mergers as the Source of the Highest Energy Cosmic Rays". In: *Phys. Rev. Lett.* 134.8 (2025), p. 081003. DOI: 10.1103/PhysRevLett.134.081003. arXiv: 2405.12004 [astro-ph.HE].

[34] J. Arons. "Magnetars in the metagalaxy: an origin for ultrahigh-energy cosmic rays in the nearby universe". In: *Astrophys. J.* 589 (2003), pp. 871–892. DOI: 10.1086/374776. arXiv: astro-ph/0208444.

[35] P. Baerwald, M. Bustamante and W. Winter. "Are gamma-ray bursts the sources of ultra-high energy cosmic rays?" In: *Astropart. Phys.* 62 (2015), pp. 66–91. DOI: 10.1016/j.astropartphys.2014.07.007. arXiv: 1401.1820 [astro-ph.HE].

[36] T. Bister and G. R. Farrar. "Constraints on UHECR Sources and Extragalactic Magnetic Fields from Directional Anisotropies". In: *Astrophys. J.* 966.1 (2024), p. 71. DOI: 10.3847/1538-4357/ad2f3f. arXiv: 2312.02645 [astro-ph.HE].

[37] C. Isola, M. Lemoine and G. Sigl. "Centaurus A as the source of ultrahigh-energy cosmic rays?" In: *Phys. Rev. D* 65 (2002), p. 023004. DOI: 10.1103/PhysRevD.65.023004. arXiv: astro-ph/0104289.

[38] S. Mollerach and E. Roulet. "Case for Centaurus A as the main source of ultrahigh-energy cosmic rays". In: *Phys. Rev. D* 110.6 (2024), p. 063030. DOI: 10.1103/PhysRevD.110.063030. arXiv: 2406.19199 [astro-ph.HE].

[39] P. Abreu et al. "Arrival Directions of Cosmic Rays above 32 EeV from Phase One of the Pierre Auger Observatory". In: *Astrophys. J.* 935.2 (2022), p. 170. DOI: 10.3847/1538-4357/ac7d4e. arXiv: 2206.13492 [astro-ph.HE].

[40] S. Navas et al. "Review of particle physics". In: *Phys. Rev. D* 110.3 (2024), p. 030001. DOI: 10.1103/PhysRevD.110.030001.

[41] L. Cazon. "Probing High-Energy Hadronic Interactions with Extensive Air Showers". In: *PoS* ICRC2019 (2020), p. 005. DOI: 10.22323/1.358.0005. arXiv: 1909.02962 [hep-ex].

[42] J. Alvarez-Muñiz et al. "Hybrid simulations of extensive air showers". In: *Phys. Rev. D* 66 (3 2002), p. 033011. DOI: 10.1103/PhysRevD.66.033011. URL: https://link.aps.org/doi/10.1103/PhysRevD.66.033011.

[43] D. E. Groom, N. V. Mokhov and S. I. Striganov. "Muon stopping power and range tables 10-MeV to 100-TeV". In: *Atom. Data Nucl. Data Tabl.* 78 (2001), pp. 183–356. DOI: 10.1006/adnd.2001.0861.

[44] B. Rossi and K. Greisen. "Cosmic-ray theory". In: *Rev. Mod. Phys.* 13 (1941), pp. 240–309. DOI: 10.1103/RevModPhys.13.240.

[45] T. Pierog et al. "First results of fast one-dimensional hybrid simulation of EAS using CONEX". In: *Nucl. Phys. B Proc. Suppl.* 151 (2004). Ed. by P. K. F. Grieder, B. Pattison and L. K. Resvanis, pp. 159–162. DOI: 10.1016/j.nuclphysbps.2005.07.029. arXiv: astro-ph/0411260.

[46] T. Bergmann et al. "One-dimensional hybrid approach to extensive air shower simulation". In: *Astroparticle Physics* 26.6 (2007), pp. 420–432. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2006.08.005.

[47] A. Fedynitch et al. "Calculation of conventional and prompt lepton fluxes at very high energy". In: *EPJ Web Conf.* 99 (2015). Ed. by D. Berge et al., p. 08001. DOI: 10.1051/epjconf/20159908001. arXiv: 1503.00544 [hep-ph].

[48] B. Rossi. *High Energy Particles*. 1952.

[49] K. Greisen. "The Extensive Air Showers". In: *Progress in Cosmic Ray Physics* 3 (1956), pp. 1–141.

[50] W. Heitler. *The quantum theory of radiation*. Vol. 5. International Series of Monographs on Physics. Oxford: Oxford University Press, 1936.

[51] J. Matthews. "A Heitler model of extensive air showers". In: *Astropart. Phys.* 22 (2005), pp. 387–397. DOI: 10.1016/j.astropartphys.2004.09.003.

[52] J. Linsley and A. A. Watson. "Validity of scaling to 10**20-eV and high-energy cosmic ray composition". In: *Phys. Rev. Lett.* 46 (1981), pp. 459–463. DOI: 10.1103/PhysRevLett.46.459.

[53] 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: *Phys. Lett. B* 784 (2018), pp. 68–76. DOI: 10.1016/j.physletb.2018.07.026. arXiv: 1803.05699 [hep-ph].

[54] M. V. B. Junior et al. "Semi-analytical model of extensive air showers using branching processes". In: *Astropart. Phys.* 131 (2021), p. 102585. DOI: 10.1016/j.astropartphys.2021.102585.

[55] R. Ulrich, R. Engel and M. Unger. "Hadronic Multiparticle Production at Ultra-High Energies and Extensive Air Showers". In: *Phys. Rev.* D83 (2011), p. 054026. DOI: 10.1103/PhysRevD.83.054026.

[56] P. Abreu et al. "Measurement of the Proton-Air Cross Section at $\sqrt{s}$=57 TeV with the Pierre Auger Observatory". In: *Phys. Rev. Lett.* 109 (6 2012), p. 062002. DOI: 10.1103/PhysRevLett.109.062002.

[57] M. E. Peskin and D. V. Schroeder. *An Introduction to Quantum Field Theory*. Reading, USA: Addison-Wesley (1995) 842 p. Westview Press, 1995.

[58] B. Andersson et al. "Parton Fragmentation and String Dynamics". In: *Phys. Rept.* 97 (1983), pp. 31–145. DOI: 10.1016/0370-1573(83)90080-7.

[59] S. Chatrchyan et al. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[60] A. A. Alves Jr. et al. "The LHCb Detector at the LHC". In: *JINST* 3 (2008), S08005. DOI: 10.1088/1748-0221/3/08/S08005.

[61] O. Adriani et al. "The LHCf detector at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08006. DOI: 10.1088/1748-0221/3/08/S08006.

[62] T. Pierog and K. Werner. "EPOS LHC-R : up-to-date hadronic model for EAS simulations". In: *PoS* ICRC2023 (2023), p. 230. DOI: 10.22323/1.444.0230.

[63] S. Ostapchenko. "QGSJET-III model of high energy hadronic interactions: The formalism". In: *Phys. Rev. D* 109.3 (2024), p. 034002. DOI: 10.1103/PhysRevD.109.034002. arXiv: 2401.06202 [hep-ph].

[64] F. Riehn et al. "Hadronic interaction model SIBYLL 2.3d and extensive air showers". In: *Phys. Rev. D* 102 (6 2020), p. 063002. DOI: 10.1103/PhysRevD.102.063002.

[65] K. Werner. "Core-corona separation in ultra-relativistic heavy ion collisions". In: *Phys. Rev. Lett.* 98 (2007), p. 152301. DOI: 10.1103/PhysRevLett.98.152301. arXiv: 0704.1270 [nucl-th].

[66] S. Baur et al. "Core-corona effect in hadron collisions and muon production in air showers". In: *Phys. Rev. D* 107.9 (2023), p. 094031. DOI: 10.1103/PhysRevD.107.094031. arXiv: 1902.09265 [hep-ph].

[67] T. Pierog et al. "EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider". In: *Phys. Rev. C* 92.3 (2015), p. 034906. DOI: 10.1103/PhysRevC.92.034906. arXiv: 1306.0121 [hep-ph].

[68] T. K. Gaisser and F. Halzen. "Soft Hard Scattering in the TeV Range". In: *Phys. Rev. Lett.* 54 (1985), p. 1754. DOI: 10.1103/PhysRevLett.54.1754.

[69] A. Capella et al. "Dual parton model". In: *Phys. Rept.* 236 (1994), pp. 225–329. DOI: 10.1016/0370-1573(94)90064-7.

[70] R. J. Glauber and G. Matthiae. "High-energy scattering of protons by nuclei". In: *Nucl. Phys. B* 21 (1970), pp. 135–157. DOI: 10.1016/0550-3213(70)90511-0.

[71] A. Aab et al. "The Pierre Auger Cosmic Ray Observatory". In: *Nucl. Instrum. Meth. A* 798 (2015), pp. 172–213. DOI: 10.1016/j.nima.2015.06.058. arXiv: 1502.01323 [astro-ph.IM].

[72] M. A. Lawrence, R. J. O. Reid and A. A. Watson. "The Cosmic ray energy spectrum above 4x10**17-eV as measured by the Haverah Park array". In: *J. Phys. G* 17 (1991), pp. 733–757. DOI: 10.1088/0954-3899/17/5/019.

[73] N. Chiba et al. "Akeno giant air shower array (AGASA) covering 100-km**2 area". In: *Nucl. Instrum. Meth. A* 311 (1992), pp. 338–349. DOI: 10.1016/0168-9002(92)90882-5.

[74] H. Tokuno et al. "New air fluorescence detectors employed in the Telescope Array experiment". In: *Nucl. Instrum. Meth. A* 676 (2012), pp. 54–65. DOI: 10.1016/j.nima.2012.02.044. arXiv: 1201.0002 [astro-ph.IM].

[75] T. Abu-Zayyad et al. "The surface detector array of the Telescope Array experiment". In: *Nucl. Instrum. Meth. A* 689 (2013), pp. 87–97. DOI: 10.1016/j.nima.2012.05.079. arXiv: 1201.4964 [astro-ph.IM].

[76] A. Aab et al. "Depth of maximum of air-shower profiles at the Pierre Auger Observatory. I. Measurements at energies above $10^{17.8}$ eV". In: *Phys. Rev. D* 90 (12 2014), p. 122005. DOI: 10.1103/PhysRevD.90.122005.

[77] J. Abraham et al. "The Fluorescence Detector of the Pierre Auger Observatory". In: *Nucl. Instrum. Meth. A* 620 (2010), pp. 227–251. DOI: 10.1016/j.nima.2010.04.023. arXiv: 0907.4282 [astro-ph.IM].

[78] A. Abdul Halim et al. "A Catalog of the Highest-Energy Cosmic Rays Recorded during Phase I of Operation of the Pierre Auger Observatory". In: *Astrophys. J. Suppl.* 264.2 (2023), p. 50. DOI: 10.3847/1538-4365/aca537. arXiv: 2211.16020 [astro-ph.HE].

[79] T. K. Gaisser and A. M. Hillas. "Reliability of the Method of Constant Intensity Cuts for Reconstructing the Average Development of Vertical Showers". In: *International Cosmic Ray Conference*. Vol. 8. International Cosmic Ray Conference. Jan. 1977, p. 353.

[80] S. Andringa, R. Conceicao and M. Pimenta. "Mass composition and cross-section from the shape of cosmic ray shower longitudinal profiles". In: *Astropart. Phys.* 34 (2011), pp. 360–367. DOI: 10.1016/j.astropartphys.2010.10.002.

[81] A. Aab et al. "Data-driven estimation of the invisible energy of cosmic ray showers with the Pierre Auger Observatory". In: *Phys. Rev. D* 100 (8 2019), p. 082003. DOI: 10.1103/PhysRevD.100.082003. URL: https://link.aps.org/doi/10.1103/PhysRevD.100.082003.

[82] I. Allekotte et al. "The Surface Detector System of the Pierre Auger Observatory". In: *Nucl. Instrum. Meth. A* 586 (2008), pp. 409–420. DOI: 10.1016/j.nima.2007.12.016. arXiv: 0712.2832 [astro-ph].

[83] A. Abdul Halim et al. "The Pierre Auger Observatory open data". In: *Eur. Phys. J. C* 85.1 (2025), p. 70. DOI: 10.1140/epjc/s10052-024-13560-5. arXiv: 2309.16294 [astro-ph.HE].

[84] K. Kamata and J. Nishimura. "The Lateral and the Angular Structure Functions of Electron Showers". In: *Prog. Theor. Phys. Suppl.* 6 (1958), pp. 93–155. DOI: 10.1143/PTPS.6.93.

[85] J. Hersil et al. "Observations of Extensive Air Showers near the Maximum of Their Longitudinal Development". In: *Phys. Rev. Lett.* 6 (1961), pp. 22–23. DOI: 10.1103/PhysRevLett.6.22.

[86] M. Ave et al. "A generalized description of the signal size in extensive air shower detectors and its applications". In: *Astropart. Phys.* 87 (2017), pp. 23–39. DOI: 10.1016/j.astropartphys.2016.11.008.

[87] M. Stadelmaier et al. "Model of the response of surface detectors to extensive air showers based on shower universality". In: *Phys. Rev. D* 110.2 (2024), p. 023030. DOI: 10.1103/PhysRevD.110.023030. arXiv: 2405.03494 [hep-ph].

[88] A. Abdul Halim et al. "Inference of the Mass Composition of Cosmic Rays with energies from $\mathbf{10^{18.5}}$ to $\mathbf{10^{20}}$ eV using the Pierre Auger Observatory and Deep Learning". In: (June 2024). arXiv: 2406.06315 [astro-ph.HE].

[89] 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 (2021), P07016. DOI: 10.1088/1748-0221/16/07/P07016.

[90] S. Hahn. "Reconstruction of muon number of air showers with the surface detector of the Pierre Auger Observatory using neural networks". In: *PoS* ICRC2023 (2023), p. 318. DOI: 10.22323/1.444.0318.

[91] A. Castellina et al. "AugerPrime: the Pierre Auger Observatory Upgrade". In: *EPJ Web Conf.* 210 (2019), p. 06002. DOI: 10.1051/epjconf/201921006002.

[92] A. Abdul Halim et al. "Measurement of the Depth of Maximum of Air-Shower Profiles with energies between $\mathbf{10^{18.5}}$ and $\mathbf{10^{20}}$ eV using the Surface Detector of the Pierre Auger Observatory and Deep Learning". In: (June 2024). arXiv: 2406.06319 [astro-ph.HE].

[93] A. Aab et al. "Direct measurement of the muonic content of extensive air showers between $\mathbf{2 \times 10^{17}}$ and $\mathbf{2 \times 10^{18}}$ eV at the Pierre Auger Observatory". In: *The European Physical Journal C* 80.8 (2020), p. 751. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-020-8055-y. URL: https://doi.org/10.1140/epjc/s10052-020-8055-y.

[94] A. Aab et al. "Muons in air showers at the Pierre Auger Observatory: Mean number in highly inclined events". In: *Phys. Rev. D* 91 (3 2015), p. 032003. DOI: 10.1103/PhysRevD.91.032003.

[95] 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 (2021), p. 152002. DOI: 10.1103/PhysRevLett.126.152002. arXiv: 2102.07797 [hep-ex].

[96] S. Ostapchenko. "Monte Carlo treatment of hadronic interactions in enhanced Pomeron scheme: I. QGSJET-II model". In: *Phys. Rev. D* 83 (2011), p. 014018. DOI: 10.1103/PhysRevD.83.014018. arXiv: 1010.1869 [hep-ph].

[97] A. Abdul Halim et al. "The number of muons measured in hybrid events detected by the Pierre Auger Observatory". In: *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*. Vol. 444. 2023, p. 339. DOI: 10.22323/1.444.0339.

[98] H. P. Dembinski et al. "Data-driven model of the cosmic-ray flux and mass composition from 10 GeV to 10^{11} GeV". In: *PoS* ICRC2017 (2018), p. 533. DOI: 10.22323/1.301.0533. arXiv: 1711.11432 [astro-ph.HE].

[99] 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. DOI: 10.1103/PhysRevD.95.063005. arXiv: 1612.07328 [hep-ph].

[100] R. Mammen Abraham et al. "First Measurement of the Muon Neutrino Interaction Cross Section and Flux as a Function of Energy at the LHC with FASER". In: *Phys. Rev. Lett.* 134.21 (2025), p. 211801. DOI: 10.1103/PhysRevLett.134.211801. arXiv: 2412.03186 [hep-ex].

[101] A. Aab et al. "Muons in air showers at the Pierre Auger Observatory: Measurement of atmospheric production depth". In: *Phys. Rev. D* 90 (1 2014), p. 012012. DOI: 10.1103/PhysRevD.90.012012. URL: https://link.aps.org/doi/10.1103/PhysRevD.90.012012.

[102] A. Abdul Halim et al. "Testing hadronic-model predictions of depth of maximum of air-shower profiles and ground-particle signals using hybrid data of the Pierre Auger Observatory". In: *Phys. Rev. D* 109.10 (2024), p. 102001. DOI: 10.1103/PhysRevD.109.102001. arXiv: 2401.10740 [astro-ph.HE].

[103] E.-J. Ahn et al. "Cosmic ray interaction event generator SIBYLL 2.1". In: *Phys. Rev. D* 80 (2009), p. 094003. DOI: 10.1103/PhysRevD.80.094003. arXiv: 0906.4113 [hep-ph].

[104] R. M. Baltrusaitis et al. "Total Proton Proton Cross-Section at s**(1/2) = 30-TeV". In: *Phys. Rev. Lett.* 52 (1984), pp. 1380–1383. DOI: 10.1103/PhysRevLett.52.1380.

[105] R. U. Abbasi et al. "Measurement of the proton-air cross section with Telescope Array's Black Rock Mesa and Long Ridge fluorescence detectors, and surface array in hybrid mode". In: *Phys. Rev. D* 102.6 (2020), p. 062004. DOI: 10.1103/PhysRevD.102.062004. arXiv: 2006.05012 [astro-ph.HE].

[106] L. Cazon et al. "Constraining the energy spectrum of neutral pions in ultra-high-energy proton-air interactions". In: *Phys. Rev. D* 103.2 (2021), p. 022001. DOI: 10.1103/PhysRevD.103.022001. arXiv: 2006.11303 [astro-ph.HE].

[107] I. A. Goos, X. Bertou and T. Pierog. "Determination of high-energy hadronic interaction properties from observables of proton initiated extensive air showers". In: (Apr. 2023). arXiv: 2304.08007 [hep-ex].

[108] S. Andringa et al. "The muonic longitudinal shower profiles at production". In: *Astroparticle Physics* 35.12 (2012), pp. 821–827. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2012.03.010.

[109] L. Cazon, R. Conceição and F. Riehn. "Universality of the muon component of extensive air showers". In: *Journal of Cosmology and Astroparticle Physics* 2023.03 (2023), p. 022. DOI: 10.1088/1475-7516/2023/03/022.

[110] D. Heck et al. *CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers, Report FZKA 6019, Forschungszentrum Karlsruhe*. 1998.

[111] A. Ferrari et al. "FLUKA: A multi-particle transport code". In: *CERN-2005-010, SLAC-R-773, INFN-TC-05-11* (2005). DOI: 10.5170/CERN-2005-010.

[112] T.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. DOI: 10.1016/j.nds.2014.07.049.

[113] R. Brun and F. Rademakers. "ROOT — An object oriented data analysis framework". In: *Nucl. Instrum. Meth. A* 389.1-2 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: 10.1016/s0168-9002(97)00048-x.

[114] Wes McKinney. "Data Structures for Statistical Computing in Python". In: *Proceedings of the 9th Python in Science Conference*. Ed. by Stéfan van der Walt and Jarrod Millman. 2010, pp. 56 –61. DOI: 10.25080/Majora-92bf1922-00a.

[115] T. pandas development team. *pandas-dev/pandas: Pandas*. Version 2.0.3. Feb. 2020. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

[116] P. Virtanen et al. "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

[117] J. D. Hunter. "Matplotlib: A 2D graphics environment". In: *Computing in Science & Engineering* 9.3 (2007), pp. 90–95. DOI: 10.1109/MCSE.2007.55.

[118] P. Younk and M. Risse. "Sensitivity of the correlation between the depth of shower maximum and the muon shower size to the cosmic ray composition". In: *Astropart. Phys.* 35 (2012), pp. 807–812. DOI: 10.1016/j.astropartphys.2012.03.001. arXiv: 1203.3732 [astro-ph.HE].

[119] K.-H. Kampert and M. Unger. "Measurements of the Cosmic Ray Composition with Air Shower Experiments". In: *Astropart. Phys.* 35 (2012), pp. 660–678. DOI: 10.1016/j.astropartphys.2012.02.004. arXiv: 1201.0018 [astro-ph.HE].

[120] S. Ostapchenko and M. Bleicher. "Constraining pion interactions at very high energies by cosmic ray data". In: *Phys. Rev. D* 93.5 (2016), p. 051501. DOI: 10.1103/PhysRevD.93.051501. arXiv: 1601.06567 [hep-ph].

[121] H Dembinski. "Measurement of the flux of ultra high energy cosmic rays using data from very inclined air showers at the Pierre Auger Observatory". PhD thesis. Ph. D. thesis, RWTH Aachen University, 2009.

[122] I. A. Goos. "Determination of physical properties of high-energy hadronic interactions from the $X_{\max}$-N_{μ} anticorrelation". PhD thesis. Karlsruher Institut für Technologie (KIT), KIT-Bibliothek, Karlsruhe, KIT, Karlsruhe, Dept. Phys., 2023. DOI: 10.5445/ir/1000156068.

[123] L. Cazon et al. "A model for the transport of muons in extensive air showers". In: *Astroparticle Physics* 36.1 (2012), pp. 211–223. ISSN: 0927-6505. DOI: https://doi.org/10.1016/j.astropartphys.2012.05.017. URL: https://www.sciencedirect.com/science/article/pii/S0927650512001223.

[124] O. Tkachenko, for the Pierre Auger Collaboration. "Studies of the mass composition of cosmic rays and proton-proton interaction cross-sections at ultra-high energies with the Pierre Auger Observatory". In: *Proceedings of 38th International Cosmic Ray Conference — PoS(ICRC2023)*. Vol. 444. 2023, p. 438. DOI: 10.22323/1.444.0438.

[125] J. Engel et al. "Nucleus-nucleus collisions and interpretation of cosmic ray cascades". In: *Phys. Rev. D* 46 (1992), pp. 5013–5025. DOI: 10.1103/PhysRevD.46.5013.

[126] H. Akima. "A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures". In: *J. ACM* 17.4 (1970), 589–602. ISSN: 0004-5411. DOI: 10.1145/321607.321609. URL: https://doi.org/10.1145/321607.321609.

[127] S. Müller et al. "Impact of muon detection thresholds on the separability of primary cosmic rays". In: *Astropart. Phys.* 97 (2018), pp. 174–185. DOI: 10.1016/j.astropartphys.2017.11.005.

[128] M. G. Aartsen et al. "IceCube-Gen2: the window to the extreme Universe". In: *J. Phys. G* 48.6 (2021), p. 060501. DOI: 10.1088/1361-6471/abbd48. arXiv: 2008.04323 [astro-ph.HE].

[129] R. Abbasi et al. "Density of GeV muons in air showers measured with IceTop". In: *Phys. Rev. D* 106.3 (2022), p. 032010. DOI: 10.1103/PhysRevD.106.032010. arXiv: 2201.12635 [hep-ex].

[130] S. Verpoest. "Multiplicity of TeV muons in extensive air showers detected with IceTop and IceCube". In: *PoS* ICRC2023 (2023), p. 207. DOI: 10.22323/1.444.0207. arXiv: 2307.14689 [astro-ph.HE].

[131] M. G. Aartsen et al. "The IceCube Neutrino Observatory: Instrumentation and Online Systems". In: *JINST* 12.03 (2017). [Erratum: JINST 19, E05001 (2024)], P03012. DOI: 10.1088/1748-0221/12/03/P03012. arXiv: 1612.05093 [astro-ph.IM].

[132] S. Verpoest. "Muons in air showers with IceCube: muon density at ground and high-energy muon multiplicity". In: *PoS* UHECR2024 (2025), p. 035. DOI: 10.22323/1.484.0035.

[133] L. Cazon et al. "Proton-air interactions at ultra-high energies in muon-depleted air showers with different depths". In: *Phys. Lett. B* 859 (2024), p. 139115. DOI: 10.1016/j.physletb.2024.139115. arXiv: 2406.08620 [astro-ph.HE].

[134] 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 2016), p. 032007. DOI: 10.1103/PhysRevD.94.032007. URL: https://link.aps.org/doi/10.1103/PhysRevD.94.032007.

[135] O. Adriani et al. "Measurement of energy flow, cross section and average inelasticity of forward neutrons produced in $\sqrt{s} = 13$ TeV proton-proton collisions with the LHCf Arm2 detector". In: *JHEP* 07 (2020), p. 016. DOI: 10.1007/JHEP07(2020)016. arXiv: 2003.02192 [hep-ex].

[136] L. Cazon et al. "Probabilistic mapping between multiparticle production variables and the depth of maximum in proton-induced extensive air showers". In: *Accepted by Phys. Rev. D* (Apr. 2025). DOI: https://doi.org/10.1103/1664-pht6. arXiv: 2504.08610 [astro-ph.HE]. URL: https://arxiv.org/abs/2504.08610.

[137] K. Greisen. "Cosmic ray showers". In: *Ann. Rev. Nucl. Part. Sci.* 10 (1960), pp. 63–108. DOI: 10.1146/annurev.ns.10.120160.000431.

[138] S. Grimm, R. Engel and D. Veberic. "Heitler-Matthews model with leading-particle effect". In: *PoS* ICRC2017 (2018), p. 299. DOI: 10.22323/1.301.0299.

[139] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman & Hall/CRC monographs on statistics and applied probability. London: Chapman and Hall, 1986. URL: https://cds.cern.ch/record/1070306.

[140] A. Aab et al. "Evidence for a mixed mass composition at the 'ankle' in the cosmic-ray spectrum". In: *Phys. Lett. B* 762 (2016), pp. 288–295. DOI: 10.1016/j.physletb.2016.09.039. arXiv: 1609.08567 [astro-ph.HE].

[141] R. Engel, D. Heck and T. Pierog. "Extensive air showers and hadronic interactions at high energy". In: *Ann. Rev. Nucl. Part. Sci.* 61 (2011), pp. 467–489. DOI: 10.1146/annurev.nucl.012809.104544.

[142] Z. Citron et al. "Report from Working Group 5: Future physics opportunities for high-density QCD at the LHC with heavy-ion and proton beams". In: *CERN Yellow Rep. Monogr.* 7 (2019). Ed. by A. Dainese et al., pp. 1159–1410. DOI: 10.23731/CYRM-2019-007.1159. arXiv: 1812.06772 [hep-ph].

[143] J. L. Feng et al. "The Forward Physics Facility at the High-Luminosity LHC". In: *J. Phys. G* 50.3 (2023), p. 030501. DOI: 10.1088/1361-6471/ac865e. arXiv: 2203.05090 [hep-ex].

[144] F. Kling and L. J. Nevay. "Forward neutrino fluxes at the LHC". In: *Phys. Rev. D* 104.11 (2021), p. 113008. DOI: 10.1103/PhysRevD.104.113008. arXiv: 2105.08270 [hep-ph].

[145] A. Abada et al. "FCC Physics Opportunities: Future Circular Collider Conceptual Design Report Volume 1". In: *Eur. Phys. J. C* 79.6 (2019), p. 474. DOI: 10.1140/epjc/s10052-019-6904-3.

[146] J. Ebr et al. "Impact of modified characteristics of hadronic interactions on cosmic-ray observables for proton and nuclear primaries". In: *PoS* ICRC2023 (2023), p. 245. DOI: 10.22323/1.444.0245.

[147] J. Vícha. "Update on testing of air-shower modelling using combined data of the Pierre Auger Observatory and phenomenological consequences". In: *39th International Cosmic Ray Conference*. July 2025. arXiv: 2507.09802 [astro-ph.HE].

[148] L. B. Arbeletche and V. de Souza. "On the parametrization of the distributions of depth of shower maximum of ultra-high energy extensive air showers". In: *Astropart. Phys.* 116 (2020), p. 102389. DOI: 10.1016/j.astropartphys.2019.102389. arXiv: 1903.03174 [astro-ph.HE].

A

Tables with the main moments of the distribution of number of muons and depth of shower maximum

A.1 Tables for EPOS LHC-R

$\log_{10}\left(\frac{E_0}{\mathrm{eV}}\right)$	$\log_{10}\langle N_\mu\rangle$	$\frac{\sigma(N_\mu)}{\langle N_\mu\rangle}$	$\langle X_{\mathrm{max}}\rangle$ [g cm^{-2}]	$\sigma(X_{\mathrm{max}})$ [g cm^{-2}]	$\rho(N_\mu, X_{\mathrm{max}})$
17.0	5.6	0.18	713.37	73.68	0.29
17.5	6.1	0.18	742.82	72.09	0.27
18.0	6.5	0.17	770.66	68.47	0.21
18.5	7.0	0.16	796.83	63.85	0.14
19.0	7.5	0.16	824.44	62.10	0.06
19.5	7.9	0.15	852.96	60.93	0.01
20.0	8.4	0.14	879.45	60.35	0.01

Table A.1: Primary-energy dependence of the means and standard deviations of $\log_{10} N_\mu$ and X_{max}, together with the Pearson correlation coefficient $\rho(N_\mu, X_{\mathrm{max}})$, from 10^4 CONEX simulations of proton-induced showers at $\theta = 60°$ using the EPOS LHC-R hadronic interaction model.

$\log_{10}\left(\frac{E_0}{\mathrm{eV}}\right)$	$\log_{10}\langle N_\mu^{\mathrm{prod}}\rangle$	$\frac{\sigma(N_\mu^{\mathrm{prod}})}{\langle N_\mu^{\mathrm{prod}}\rangle}$	$\langle \Delta X_{\mathrm{max}}\rangle$ [g cm^{-2}]	$\sigma(\Delta X_{\mathrm{max}})$ [g cm^{-2}]	$\rho(N_\mu^{\mathrm{prod}}, \Delta X_{\mathrm{max}})$
17.0	6.1	0.16	660.06	52.03	-0.12
17.5	6.5	0.16	692.22	51.51	-0.14
18.0	7.0	0.15	722.34	48.22	-0.21
18.5	7.5	0.15	751.27	45.07	-0.31
19.0	7.9	0.15	781.06	45.06	-0.39
19.5	8.4	0.14	810.78	45.89	-0.47
20.0	8.9	0.13	839.75	46.81	-0.51

Table A.2: Primary-energy dependence of the means and standard deviations of $\log_{10} N_\mu^{\mathrm{prod}}$ and ΔX_{max}, together with the Pearson correlation coefficient $\rho(N_\mu^{\mathrm{prod}}, \Delta X_{\mathrm{max}})$, from 10^4 CONEX simulations of proton-induced showers at $\theta = 60°$ using the EPOS LHC-R hadronic interaction model.

A.2 Tables for QGSJET-III.01

$\log_{10}\left(\frac{E_0}{\mathrm{eV}}\right)$	$\log_{10}\langle N_\mu\rangle$	$\frac{\sigma(N_\mu)}{\langle N_\mu\rangle}$	$\langle X_{\max}\rangle$ [g cm^{-2}]	$\sigma(X_{\max})$ [g cm^{-2}]	$\rho(N_\mu, X_{\max})$
17.0	5.5	0.18	695.93	70.13	0.29
17.5	6.0	0.17	723.21	66.40	0.23
18.0	6.4	0.16	748.96	63.38	0.18
18.5	6.9	0.15	776.06	61.00	0.16
19.0	7.4	0.14	803.37	59.29	0.11
19.5	7.8	0.14	831.34	58.62	0.09
20.0	8.3	0.13	859.86	56.29	0.12

Table A.3: Primary-energy dependence of the means and standard deviations of $\log_{10} N_\mu$ and $X_{\max}$, together with the Pearson correlation coefficient $\rho(N_\mu, X_{\max})$, from 10^4 CONEX simulations of proton-induced showers at $\theta = 60°$ using the QGSJET-III.01 hadronic interaction model.

$\log_{10}\left(\frac{E_0}{\mathrm{eV}}\right)$	$\log_{10}\langle N_\mu^{\mathrm{prod}}\rangle$	$\frac{\sigma(N_\mu^{\mathrm{prod}})}{\langle N_\mu^{\mathrm{prod}}\rangle}$	$\langle \Delta X_{\max}\rangle$ [g cm^{-2}]	$\sigma(\Delta X_{\max})$ [g cm^{-2}]	$\rho(N_\mu^{\mathrm{prod}}, \Delta X_{\max})$
17.0	6.0	0.16	644.1	46.4	-0.10
17.5	6.5	0.15	672.6	44.4	-0.19
18.0	6.9	0.15	700.6	41.8	-0.27
18.5	7.4	0.14	730.0	40.8	-0.31
19.0	7.8	0.13	758.7	40.1	-0.39
19.5	8.3	0.12	787.9	40.1	-0.44
20.0	8.8	0.12	817.6	39.8	-0.46

Table A.4: Primary-energy dependence of the means and standard deviations of $\log_{10} N_\mu^{\mathrm{prod}}$ and $\Delta X_{\max}$, together with the Pearson correlation coefficient $\rho(N_\mu^{\mathrm{prod}}, \Delta X_{\max})$, from 10^4 CONEX simulations of proton-induced showers at $\theta = 60°$ using the QGSJET-III.01 hadronic interaction model.

A.3 Tables for SIBYLL2.3e

$\log_{10}\left(\frac{E_0}{\mathrm{eV}}\right)$	$\log_{10}\langle N_\mu\rangle$	$\frac{\sigma(N_\mu)}{\langle N_\mu\rangle}$	$\langle X_{\max}\rangle$ [g cm^{-2}]	$\sigma(X_{\max})$ [g cm^{-2}]	$\rho(N_\mu, X_{\max})$
17.0	5.5	0.19	700.80	72.54	0.24
17.5	6.0	0.19	730.81	68.50	0.17
18.0	6.5	0.18	758.64	65.31	0.10
18.5	6.9	0.17	788.90	63.27	0.05
19.0	7.4	0.16	817.09	61.49	0.02
19.5	7.9	0.16	845.83	59.09	0.01
20.0	8.3	0.15	875.61	59.00	0.03

Table A.5: Primary-energy dependence of the means and standard deviations of $\log_{10} N_\mu$ and $X_{\max}$, together with the Pearson correlation coefficient $\rho(N_\mu, X_{\max})$, from 10^4 CONEX simulations of proton-induced showers at $\theta = 60°$ using the SIBYLL2.3e hadronic interaction model.

$\log_{10}\left(\frac{E_0}{\mathrm{eV}}\right)$	$\log_{10}\langle N_\mu^{\mathrm{prod}}\rangle$	$\frac{\sigma(N_\mu^{\mathrm{prod}})}{\langle N_\mu^{\mathrm{prod}}\rangle}$	$\langle \Delta X_{\max}\rangle$ [g cm^{-2}]	$\sigma(\Delta X_{\max})$ [g cm^{-2}]	$\rho(N_\mu^{\mathrm{prod}}, \Delta X_{\max})$
17.0	6.0	0.17	647.1	48.4	-0.16
17.5	6.5	0.17	679.5	46.4	-0.24
18.0	7.0	0.16	710.9	45.0	-0.30
18.5	7.4	0.16	742.3	43.7	-0.39
19.0	7.9	0.15	772.5	43.0	-0.43
19.5	8.3	0.14	803.3	42.2	-0.49
20.0	8.8	0.13	833.8	43.9	-0.51

Table A.6: Primary-energy dependence of the means and standard deviations of $\log_{10} N_\mu^{\mathrm{prod}}$ and $\Delta X_{\max}$, together with the Pearson correlation coefficient $\rho(N_\mu^{\mathrm{prod}}, \Delta X_{\max})$, from 10^4 CONEX simulations of proton-induced showers at $\theta = 60°$ using the SIBYLL2.3e hadronic interaction model.

B

Additional derivations

B.1 Primary mass dependence of Pearson correlation between N_μ^{prod} and X_{max} using the superposition model

Here, we derive the primary-mass dependence of the correlation coefficient between N_μ^{prod} and X_{max} within the superposition principle. To do so, we start by expressing the Pearson correlation coefficient between $N_\mu^{\text{prod}}(E_0, A)$ and $X_{\text{max}}(E_0, A)$ in terms of the corresponding correlation for proton-induced extensive air showers. For clarity, we introduce the simplified notation: $n_A \equiv N_\mu^{\text{prod}}(E_0, A)$, $n_i \equiv N_{\mu,i}^{\text{prod}}(E_0/A, 1)$ (the muon number for proton-induced sub-showers), $x_A \equiv X_{\text{max}}(E_0, A)$, and $x_i \equiv X_{\text{max},i}(E_0/A, 1)$ (the depth of maximum for proton-induced sub-showers), so that Equations (2.22) and (2.23) can be written as

$$n_A = \sum_{i=1}^{A} n_i \quad \text{and} \quad x_A = \frac{1}{A}\sum_{j=1}^{A} x_j. \tag{B.1}$$

Moreover, the Pearson correlation coefficient between n_A and x_A is given by

$$\rho(n_A, x_A) = \frac{\text{Cov}[n_A, x_A]}{\sigma(n_A)\sigma(x_A)} = \frac{\langle n_A x_A\rangle - \langle n_A\rangle \langle x_A\rangle}{\sigma(n_A)\sigma(x_A)}. \tag{B.2}$$

In our simplified notation, we additionally know that

$$\langle n_A\rangle = A\langle n\rangle \;;\; \sigma(n_A) = \sqrt{A}\sigma(n) \quad \text{and} \quad \langle x_A\rangle = \langle x\rangle \;;\; \sigma(x_A) = \frac{\sigma(x)}{\sqrt{A}}, \tag{B.3}$$

from where it follows that

$$\rho(n_A, x_A) = \frac{\langle n_A x_A\rangle - A\langle n\rangle\langle x\rangle}{\sqrt{A}\sigma(n)\dfrac{\sigma(x)}{\sqrt{A}}} = \frac{\langle n_A x_A\rangle - A\langle n\rangle\langle x\rangle}{\sigma(n)\sigma(x)}. \tag{B.4}$$

Using once again Equations (2.22) and (2.23), we express $\langle n_A x_A \rangle$ as

$$\langle n_A x_A \rangle = \frac{1}{A} \left\langle \left(\sum_{i=1}^{A} n_i \right) \left(\sum_{j=1}^{A} x_j \right) \right\rangle = \frac{1}{A} \left[\sum_{i=1}^{A} \langle n_i x_i \rangle + \sum_{i=1}^{A} \sum_{j \neq i}^{A} \langle n_i x_j \rangle \right], \tag{B.5}$$

where we used the linearity of the mean value operator and separated terms with $i = j$ and $i \neq j$. Now, for $i \neq j$ the values of n_i and x_j are independent, while for $i = j$ we must account for the correlation between N_μ and $X_{\max}$. Since this correlation is independent of i, we have $\langle n_i x_i \rangle = \langle nx \rangle$ for all i, and hence

$$\langle n_A x_A \rangle = \frac{1}{A} \left[\sum_{i=1}^{A} \langle nx \rangle - \sum_{i=1}^{A} \sum_{i \neq j}^{A} \langle n \rangle \langle x \rangle \right] = \langle nx \rangle + (A-1) \langle n \rangle \langle x \rangle . \tag{B.6}$$

Therefore,

$$\rho(n_A, x_A) = \frac{\langle n_A x_A \rangle - \langle n_A \rangle \langle x_A \rangle}{\sigma(n_A)\sigma(x_A)} = \frac{\langle nx \rangle + (A-1) \langle n \rangle \langle x \rangle - A \langle n \rangle \langle x \rangle}{\sqrt{A}\sigma(n) \dfrac{\sigma(x)}{\sqrt{A}}} = \rho(n, x). \tag{B.7}$$

C

Additional validations

C.1 Validating the approximations taken in the derivation of $t_{\max}^{\mathrm{EM}}$

C.1.1 Description of the toy Monte Carlo procedure

In Equation 4.2, we presented an estimate for the depth of the shower maximum of the electromagnetic (EM) cascade triggered by the decay of neutral pions produced in the primary proton-air interaction. Recall that each neutral pion decays into two photons, each triggering EM sub-cascades whose longitudinal profiles can be described by Greisen functions. The mentioned estimation is based on three key assumptions:

- The dominant source of stochasticity in the EM shower arises from fluctuations in the energy fraction carried by each first-generation neutral pion. Consequently, we can neglect fluctuations in the energy sharing between the two photons resulting from each neutral pion decay, as well as fluctuations in the maximum particle content of each sub-cascade and the depth at which it is reached;

- The maximum number of EM particles produced by a photon scales linearly with its energy, while the depth of their shower maximum scales logarithmically with this parameter, as predicted by the Heitler model;

- The weighted average of the depths of the maxima of the individual sub-cascades provides a good estimate for the depth of the maximum of the combined EM profile.

To validate and quantify the accuracy of these assumptions, we designed a toy Monte Carlo simulation as follows:

1. For each proton-air interaction simulated with CONEX at a primary energy $E_0 = 10^{19}$ eV, we store the energy of each secondary neutral pion, $E_j = x_j E_0$.

2. For each neutral pion, a random fraction f_j is drawn from a uniform distribution in $[0, 1]$, representing the energy share of one of the decay photons. The second photon receives the remaining fraction, $1 - f_j$, by energy conservation. The uniformity of f_j arises from the kinematics of the two-body decay of the neutral pion.

3. Each photon, with energies $E_{1,j} = f_j x_j E_0$ and $E_{2,j} = (1 - f_j) x_j E_0$, respectively, is assigned a depth of maximum $X^{(1)}_{\mathrm{max},j}$ and $X^{(2)}_{\mathrm{max},j}$ drawn from Gumbel distributions parametrised as a function of the photon's energy [148]. The corresponding maximum particle contents $N^{(1)}_{\mathrm{max},j}$ and $N^{(2)}_{\mathrm{max},j}$ are sampled from Log-Gaussian distributions with energy-dependent mean and variance.

4. A Greisen longitudinal profile is simulated for each photon, using the corresponding X_{max} and N_{max} values.

5. All individual Greisen profiles, across all the $2m_{\mathrm{EM}}$ photons, are summed to construct the overall EM cascade profile.

6. The true depth of the shower maximum of the combined profile, $X^{(G)}_{\mathrm{max}}$, is determined as its maximiser.

Based on the resulting Greisen profiles, we compute approximations to $X^{(G)}_{\mathrm{max}}$ using three progressively stronger simplifications:

- $X^{(1)}_{\mathrm{max}}$: A weighted average of the depths of maximum of all photon-initiated subshowers, using their respective N_{max} values as weights. This estimator accounts for fluctuations in energy, N_{max} and X_{max}:

$$X^{(1)}_{\mathrm{max}} = \frac{\sum_{j=1}^{m_{\mathrm{EM}}} N^{(1)}_{\mathrm{max},j} X^{(1)}_{\mathrm{max},j} + N^{(2)}_{\mathrm{max},j} X^{(2)}_{\mathrm{max},j}}{\sum_{j=1}^{m_{\mathrm{EM}}} N^{(1)}_{\mathrm{max},j} + N^{(2)}_{\mathrm{max},j}} \tag{C.1}$$

- $X^{(2)}_{\mathrm{max}}$: Neglecting fluctuations in both N_{max} and X_{max}, assuming instead their Heitler scaling with energy. This yields:

$$X^{(2)}_{\mathrm{max}} = \ln\left(\frac{E_0}{\xi^e_c}\right) - \frac{\zeta_{\mathrm{EM}}}{\alpha_{\mathrm{EM}}} + \frac{1}{\alpha_{\mathrm{EM}}} \sum_{j=1}^{m_{\mathrm{EM}}} x_j \left[f_j \ln f_j + (1 - f_j) \ln(1 - f_j)\right] \tag{C.2}$$

- $X^{(3)}_{\mathrm{max}}$: This is the expression from Equation (4.2), which neglects fluctuations in the photon energy fractions altogether. It allows $X^{(2)}_{\mathrm{max}}$ to be rewritten as:

$$X^{(2)}_{\mathrm{max}} = X^{(3)}_{\mathrm{max}} + \ln 2 - \frac{1}{\alpha_{\mathrm{EM}}} \sum_{j=1}^{m_{\mathrm{EM}}} x_j \left[f_j \ln f_j + (1 - f_j) \ln(1 - f_j)\right] \tag{C.3}$$

C.1.2 Parametrizing the distributions of X_{max} and N_{max} for photon-induced showers

Using libraries of 10^4 CONEX showers induced by photons with primary energies $E_0 \in [10^{12}, 10^{18}]$ eV, in steps of $\Delta \log_{10}(E_0/\mathrm{eV}) = 1$, and simulated with SIBYLL2.3d, we obtain the distributions of the depth of shower maximum, $f(X_{\mathrm{max}} \,|\, \gamma)$, and of the logarithm of

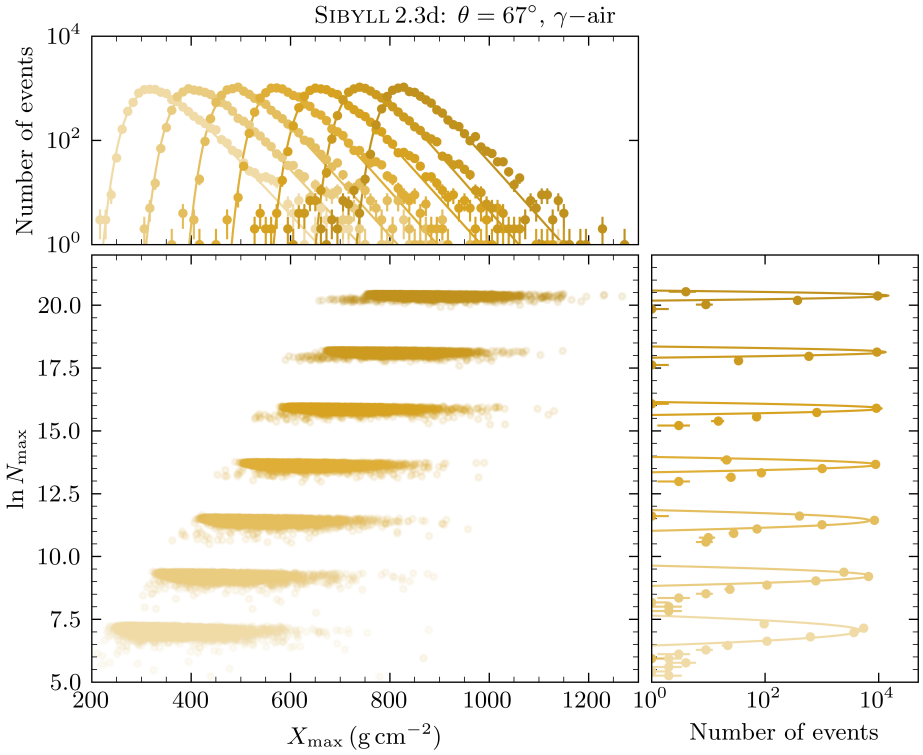

Figure C.1: Fitted distributions of $X_{\max}$ (upper panel) and $N_{\max}$ (right panel), respectively $f(X_{\max}\,|\,\gamma)$ and $f(N_{\max}\,|\,\gamma)$, for photon-induced showers, with different primary energies ranging from 10^{12} to 10^{18} eV in steps of $\Delta \log_{10}(E_0/\mathrm{eV}) = 1$. Middle panel: joint distributions $f(X_{\max}, N_{\max}\,|\,\gamma)$ for each primary energy. Higher primary energies are represented in darker shades of yellow. The true distributions are depicted by the solid markers and the fit function by the solid curves. The distributions $f(X_{\max}\,|\,\gamma)$ are fitted to Equation C.4, while all $f(N_{\max}\,|\,\gamma)$ are fitted to normal distributions. The libraries of Conex simulations of photon-induced showers are produced using the high-energy hadronic interaction model Sibyll2.3d.

the particle content at shower maximum, $f(\ln N_{\max}\,|\,\gamma)$, for each energy bin. These are shown in Figure C.1, which also displays their joint distribution, overlaid with a set of fit curves.

The unbinned log-likelihood fit to $f(X_{\max}\,|\,\gamma)$ is performed using the generalised Gumbel distribution:

$$f(X_{\max}\,|\,\gamma) = \frac{1}{\beta}\,\mathrm{e}^{-(z+\mathrm{e}^{-z})}, \quad \text{with} \quad z = \frac{X_{\max} - \mu}{\beta}, \tag{C.4}$$

where μ denotes the mode of the distribution, and β is a scale parameter related to its variance via $\sigma^2(X_{\max}) = \pi^2\beta^2/6$.

The same fitting procedure is applied to $f(\ln N_{\max}\,|\,\gamma)$, which is well described by a normal distribution.

While the Gumbel distribution provides a good description of the overall shape of $f(X_{\max}\,|\,\gamma)$, the Gaussian distribution does not fully capture the low tail of $f(\ln N_{\max}\,|\,\gamma)$. Nevertheless, the bulk of the latter distribution is well reproduced, which suffices for the intended application. Importantly, the observables $X_{\max}$ and $N_{\max}$ exhibit little correlation, justifying their independent sampling in the aforementioned toy Monte Carlo.

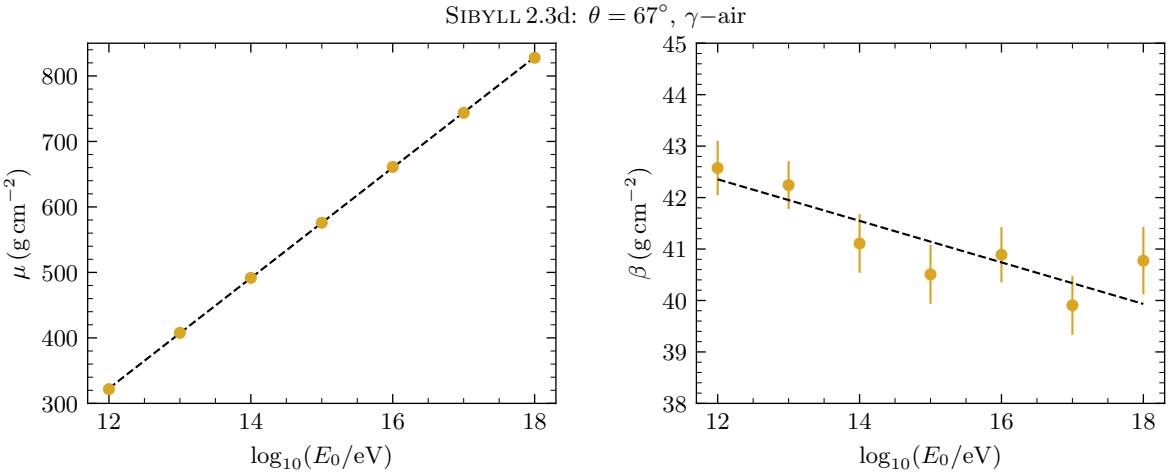

Figure C.2: Parameters of the Gumbel parametrisation of $f(X_{\max} \,|\, \gamma)$ as a function of the primary energy, and the corresponding linear fits. The values of the parameters for each shower energy are depicted in yellow, while the fit curves are depicted by dashed black lines.

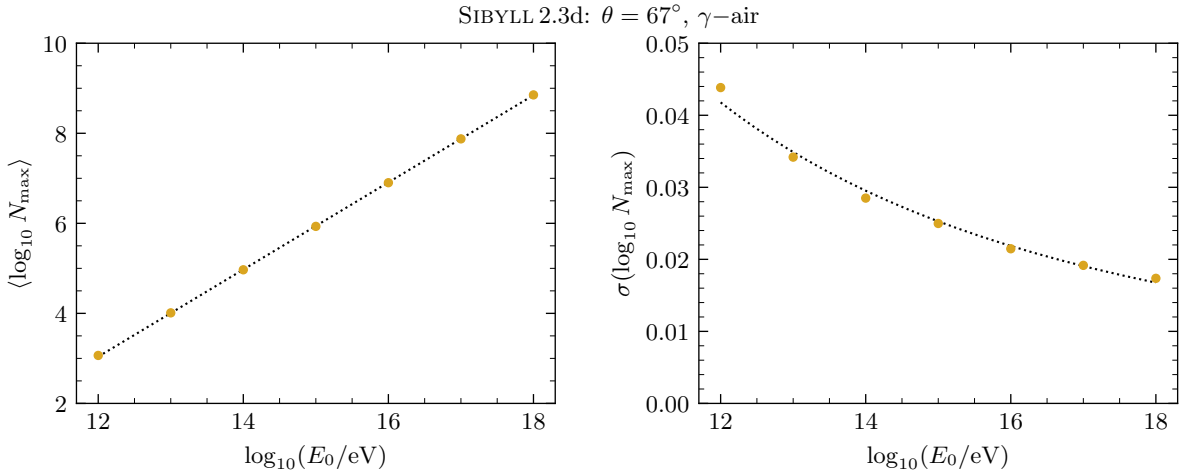

Figure C.3: Parameters of the Gaussian parametrisation of $f(N_{\max} \,|\, \gamma)$ as a function of the primary energy, and the corresponding power-law fits. The values of the parameters for each shower energy are depicted in yellow, while the fit curves are depicted by dashed black lines.

The energy evolution of the Gumbel parameters μ and β (see Equation (C.4)) is modelled with linear functions of $\log_{10}(E_0)$, while the energy dependences of the mean and standard deviation of $\ln N_{\max}$ are fit to power-law functions. The good quality of these parametrisations is illustrated in Figures C.2 and C.3.

C.1.3 Results

Having parametrised the energy evolution of the main moments of the distributions of $X_{\max}$ and $N_{\max}$ for photon-induced showers, we proceed to run the toy Monte Carlo described in C.1.1. For each simulated proton-air interaction, the value of $X_{\max}^{(G)}$—the depth of the maximum of the summed Greisen profiles—is computed and compared, on an event-by-event basis, with its approximations $X_{\max}^{(1)}$, $X_{\max}^{(2)}$, and $X_{\max}^{(3)}$. The neutral pion energy spectra used in this comparison are extracted from CONEX simulations of proton-

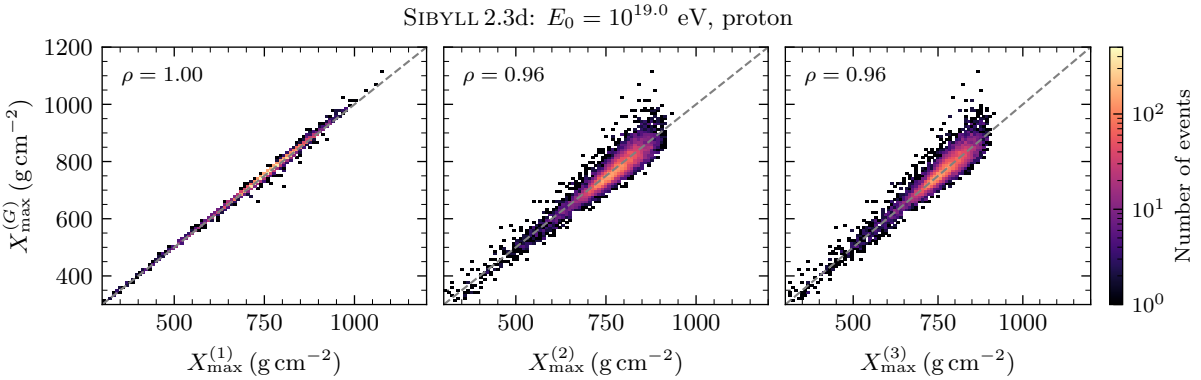

Figure C.4: Correlation plots between the true value of the depth of shower maximum extracted from the sum of Greisen longitudinal profiles, $X^{(G)}_{\max}$, and their estimations with different degrees of approximation: $X^{(1)}_{\max}$ (left), $X^{(2)}_{\max}$ (middle) and $X^{(3)}_{\max}$. The Pearson correlation coefficient, ρ, is indicated in the legend of each panel and the 1-1 line is depicted in dashed grey. This figure was produced employing the toy Monte Carlo described in Section C.1.1, using proton-air interactions simulated with Sibyll2.3d with $E_0 = 10^{19}$ eV.

air interactions at $E_0 = 10^{19}$ eV, using the hadronic interaction model Sibyll2.3d.

Figure C.4 shows the correlation between the true value $X^{(G)}_{\max}$ and each of its approximations, along with the corresponding Pearson correlation coefficients.

Irrespective of the chosen approximation, most of the correlation between the true value of $X_{\max}$ and its estimate is driven by the fluctuations of the fraction of energy retained in the electromagnetic sector of the primary proton-air interaction. Nevertheless, the depth of maximum of the sum of Greisen profiles is very accurately reproduced by a weighted average of the individual maxima, as implemented in $X^{(1)}_{\max}$, which accounts for the fluctuations of $N_{\max}$, $X_{\max}$, and the initial energy of each photon sub-shower.

Neglecting the fluctuations of $N_{\max}$ and $X_{\max}$, and replacing them with their Heitler-scaling predictions—as done in $X^{(2)}_{\max}$—still yields a robust estimator of $X^{(G)}_{\max}$, albeit with an expected increase in the dispersion. Remarkably, further simplifying the model by neglecting the fluctuations of the energy partition between the two photons from each neutral pion decay (i.e., using $X^{(3)}_{\max}$) does not significantly deteriorate the quality of the approximation. This observation is further illustrated in Figure C.5, which shows the distributions of the residuals $X^{(G)}_{\max} - X^{(i)}_{\max}$ for each of the approximations.

From the standard deviations of the residual distributions, we determine the resolution of each approximation: 4.5 g cm^{-2} for $X^{(1)}_{\max}$, 24.8 g cm^{-2} for $X^{(2)}_{\max}$, and 25.1 g cm^{-2} for $X^{(3)}_{\max}$. The latter value is particularly relevant, as it corresponds to the intrinsic resolution of the estimator of $X_{\max}$ from the EM component introduced in Equation (4.2).

These results demonstrate that computing a weighted average of the individual depths of maximum of the Greisen profiles—using the particle content at the maximum of each sub-shower as weights—yields a highly accurate estimate of the maximum of the combined electromagnetic cascade. The precision of this estimate degrades only when the intrinsic fluctuations of the photon-induced sub-showers are neglected.

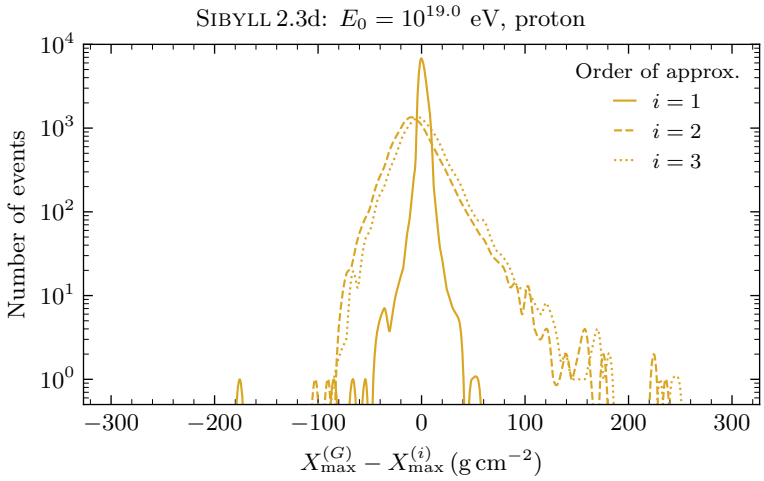

Figure C.5: Distributions of residuals of the shower-by-shower prediction of $X^{(G)}_{\max}$ using estimators with different assumptions: $X^{(1)}_{\max}$ (solid), $X^{(2)}_{\max}$ (dashed) and $X^{(3)}_{\max}$ (dotted). This figure was produced employing the toy Monte Carlo described in Section C.1.1, using proton-air interactions simulated with SIBYLL2.3d with $E_0 = 10^{19}$ eV.

D

RESUMO DA TESE

Cando os raios cósmicos de ultra alta enerxía (UHECR) colisionan con núcleos da atmosfera terrestre —a enerxías no centro de masas que superan nunha orde de magnitude as alcanzables polos colisionadores actuais construídos polo ser humano—, inician vastas cascadas de partículas secundarias coñecidas como Chuvias Atmosfericass Extensas ou *Extensive Air Showers* (EAS).

Estas cascadas evolucionan mediante a reinteracción e propagación de partículas secundarias que teñen a súa orixe en interaccións hadrónicas que non poden ser calculadas a partir dos principios fundamentais da Cromodinámica Cuántica. No seu lugar, estas interaccións descríbense mediante modelos fenomenolóxicos de interacción hadrónica que están pouco restrinxidos polos datos de aceleradores. A incerteza nas extrapolacións destes modelos dificulta a interpretación da composición en masa do fluxo de raios cósmicos, obstaculizando así os esforzos para identificar as fontes dos raios cósmicos de ultra alta enerxía.

Esta tese explora a relación entre a interacción primaria raio cósmico–atmosfera e a distribución conxunta de dous observables clave da cascada: a profundidade do máximo do desenvolvemento da ducha, $X_{\max}$, e o número de muóns ao nivel do chan, N_μ. O traballo culmina cunha estimación preliminar das características da produción hadrónica na interacción primaria, derivada dos datos publicados polo Observatorio Pierre Auger.

D.1 Raios cósmicos de ultraalta enerxía

Os raios cósmicos de ultraalta enerxía (UHECRs) son partículas cargadas que alcanzan a Terra con enerxías superiores a 10^{18} eV, superando en máis dunha orde de magnitude as enerxías no centro de masa alcanzadas no Gran Colisionador de Hadróns (LHC), o acelerador de partículas máis potente construido ata o momento. Os lugares de aceleración ou as fontes destes UHECRs seguen a ser unha cuestión central aberta na astropartículas. Aínda que a anisotropía das súas direccións de chegada suxire unha orixe extragaláctica, os mecanismos exactos responsables da súa aceleración, así como a natureza das súas fontes, aínda non foron identificados.

Unha das principais dificultades para atopar as fontes dos UHECRs é o seu fluxo extremadamente baixo: aproximadamente 1 partícula por quilómetro cadrado por século a $E_0 = 10^{19}$ eV. Polo tanto, estas partículas non poden ser observadas directamente mediante detectores en satélites ou globos estratosféricos. En cambio, a súa enerxía e

dirección de chegada deben medirse indirectamente mediante o estudo das cascadas de partículas secundarias que inducen na atmosfera. Estas chuvias mídense en observatorios de cascadas con gran exposición. Ademais, a composición de masa dos raios cósmicos de maior enerxía debe inferirse a partir da interpretación de observables sensibles á masa da choiva, empregando modelos fenomenolóxicos de interacción hadrónica. Estes modelos son necesarios para describir as interaccións hadrónicas que impulsan a chuvia, xa que as súas taxas non poden ser calculadas directamente a partir dos principios fundamentais da Cromodinámica Cuántica (QCD). Estas teñen que ser calculadas mediante extrapolacións de modelos de interacción hadrónica axustados a datos de aceleradores, que son escasos nas rexións cinemáticas relevantes para o desenvolvemento das cascadas atomosfericas. Como resultado, as diferentes suposicións dos modelos de interacción hadrónica permiten unha certa liberdade nas súas explicacións, o que xera grandes incertezas nos modelos para a interpretación da composición de masa dos UHECRs. Polo tanto, a súa difusión no medio intergaláctico, que inclúe desviacións dependentes da masa nos campos magnéticos galegos e extragalácticos, así como as súas interaccións dependentes da masa co po, o Fondo Cósmico de Microondas (CMB) e a Luz de Fondo Extragaláctica (EBL), non poden determinarse con precisión. Tampouco poden determinarse as súas interaccións no ambiente astrofísico das súas fontes. En resumo, a falta dunha determinación precisa da composición de masa dos UHECRs dificulta a identificación dos seus lugares de aceleración.

Máis aló do seu interese astrofísico, os UHECRs tamén ofrecen un laboratorio único para sondar as interaccións hadrónicas a enerxías e en rexións cinemáticas inaccesibles aos experimentos actuais en colisionadores. En particular, proporcionan acceso á rexión cara adiante do espazo de fases, onde reside a maior parte da enerxía e que está pouco restrinxida polos datos dos aceleradores. Así, os UHECRs serven tanto como mensaxeiros do universo de alta enerxía como sondas das interaccións fundamentais de partículas a enerxías moi superiores ás do LHC.

D.1.1 Fenomenoloxía das Cascadas Atmosféricas

Cando un UHECR entra na atmosfera terrestre, sofre unha colisión cun núcleo atmosférico —tipicamente nitróxeno ou osíxeno— iniciando unha complexa cascada de partículas secundarias. Esta primeira interacción produce un número de partículas con interacción hadrónica, como barións lixeiros, kaóns cargados, $K^{\pm}$, e pións cargadas, $\pi^{\pm}$, así como hadrõns que se desintegran rapidamente en partículas electromagnéticas, como as pións neutras mediante a desintegración $\pi^0 \to \gamma\gamma$. A reiteración e propagación dos hádrons xera unha cascada macroscópica coñecida como cascada atmosférica. O desenvolvemento da chuvia está gobernado pola física da produción e da desintegración das partículas, así como pola enerxía e masa do raio cósmico primario. Estes procesos complexos poden ser resoltos de forma determinista mediante un conxunto de ecuacións integrais-diferenciais coñecidas como Ecuacións da Cascada. Aínda que útiles, estas ecuacións non teñen en conta a estocasticidade da produción, transporte e desintegración dos hadróns, nin proporcionan un marco transparente para sondar a física das interaccións hadrónicas.

Dai xorden dúas compoñentes principais: a electromagnética e a hadrónica. A compoñente electromagnética está dominada por electróns, positróns e fotóns, e aliméntase principalmente do decaemento das pionas neutras producidas nas interaccións hadrón-

icas. A compoñente hadrónica consiste en barións e mesóns, particularmente pións e kaóns cargados, que interactúan posteriormente e producen muóns a través do seu decaemento. Mentres que a compoñente electromagnética determina a evolución do contido de partículas cargadas da chuvia en función da cantidade de atmosfera atravesada, o desenvolvemento lonxitudinal da chuvia, o decaemento do núcleo hadrónico na compoñente de muóns, aporta información valiosa sobre varias etapas da cascada. En particular, debido a que os muóns son leptóns altamente penetrantes, con longa vida media e que sofren baixas perdas de enerxía na atmosfera, poden medirse nos experimentos de cascadas atmosféricas.

Entre os observables clave reconstruídos a partir das chuvias están a profundidade do máximo da chuvia, $X_{\max}$, e o número de muóns ao nivel do chan, N_μ. A profundidade $X_{\max}$ corresponde á profundidade atmosférica na que o número de partículas da compoñente electromagnética alcanza o seu máximo e é sensible tanto á masa do primario como á produción de hadróns nas primeiras etapas da chuvia. Por exemplo, os primarios máis lixeiros, como os protóns, tenden a penetrar máis fondo e producir choivas con maiores fluctuacións en $X_{\max}$, mentres que os primarios máis pesados, como o ferro, interactúan máis arriba e producen distribucións máis estreitas.

O número de muóns, N_μ, é un sondaxe sensible da produción de partículas hadrónicas. A súa escala depende da historia da produción hadrónica mediante a fracción acumulada de enerxía que permanece na compoñente hadrónica en cada etapa da choiva. A evolución desta escala coa enerxía e a composición de masa do primario convértea nun potente discriminador entre primarios máis lixeiros e máis pesados. Non obstante, esta escala vese afectada considerablemente por pequenas incertezas na descrición das interaccións hadrónicas, debido ao crecemento multiplicativo da compoñente de muóns.

Con todo, a conexión precisa entre a masa do primario e os observables clave da chuvia mencionados debe derivarse de simulacións detalladas das cascadas atmosféricas. Estas simulacións de Monte Carlo tratan a interacción, transporte e decaemento das partículas secundarias producidas nas chuvias, empregando modelos de interacción hadrónica de alta e baixa enerxía. Estas simulacións úsanse despois para interpretar os datos das chuvias medidos por grandes experimentos de raios cósmicos, como o Observatorio Pierre Auger.

D.1.2 Experimentos de cascadas atmosféricas: o Observatorio Pierre Auger

O Observatorio Pierre Auger é a instalación máis grande e avanzada do mundo para o estudo dos raios cósmicos de ultraalta enerxía (UHECRs). Situado na provincia de Mendoza, en Arxentina, ocupa unha superficie de máis de 3000 km^2, o que permite detectar os raios cósmicos extremadamente raros que alcanzan a Terra con enerxías superiores a 10^{18} eV. Os datos de alta calidade e gran estatística recollidos polo observatorio son esenciais para discriminar entre modelos de fontes astrofísicas, comprender a propagación dos UHECRs polo espazo extragaláctico e interpretar características no espectro enerxético dos UHECRs, como a transición entre aceleración galáctica e extragaláctica, e a supresión do fluxo por riba de 5×10^{19} eV.

O deseño híbrido do Observatorio Pierre Auger combina dúas técnicas de detección complementarias. O desenvolvemento lonxitudinal das cascadas atmosféricas (EAS) mídese directamente mediante 27 telescopios de fluorescencia, chamados colectivamente

Detector de Fluorescencia (FD), situados en catro edificios na periferia da matriz do detector superficial. Estes telescopios observan a luz de fluorescencia emitida por moléculas de nitróxeno excitadas polas partículas cargadas da EAS ao atravesar a atmosfera, o que permite reconstruír a enerxía primaria e $X_{\max}$. A matriz do Detector Superficial (SD) consta de máis de 1660 detectores de auga-Cherenkov (WCDs) distribuídos en redes triangulares con separacións de 1.5 km (SD-1500), 750 m (SD-750) e 433 m (SD-433). As estacións do detector superficial amosan a área de impacto das cascadas ao nivel do chan, permitindo reconstruír a dirección de chegada, a función de distribución lateral da densidade de partículas da chuvia e a enerxía da chuvia. Xa sexa analizando chuvias moi inclinadas, empregando algoritmos de aprendizaxe automática ou usando marcos como a universalidade das cascadas atmosféricas, os datos recollidos polos WCDs poden usarse para estimar o contido de muóns das EAS.

Ademais, o Observatorio Pierre Auger está a realizar unha actualización coñecida como *AugerPrime*, que inclúe: un conxunto de detectores de muóns soterrados baixo a matriz SD-750, placas de centelleo enriba dos WCDs e antenas de radio, co obxectivo de mellorar a sensibilidade ao contido muónico da choiva e posibilitar estudos máis precisos da composición. Deste xeito, o Observatorio desempeña un papel central na constrición dos modelos de interacción hadrónica e na sondaxe da física fundamental das interaccións de partículas a enerxías extremas.

D.1.3 Descrición inconsistente das cascadas atmosféricas

Hai evidencias crecentes de que os modelos actuais de interacción hadrónica non describen de forma consistente as propiedades das cascadas. Un exemplo paradigmático disto é o chamado *Enigma dos Muóns*: unha subestimación sistemática e significativa da escala de muóns nas simulacións de cascadas, do orde do 30%, se a composición inferida a partir dos datos das choivas se interpreta con $X_{\max}$. Este déficit, que aparece en enerxías tan baixas como 10^{16} eV e medra coa enerxía, suxire que os modelos hadrónicos non conseguen describir aspectos clave da produción de partículas. A pesar da discrepancia entre os datos e as predicións dos modelos de interacción hadrónica para a escala de muóns, as fluctuacións medidas dunha chuvia a outra son compatibles coa composición derivada de $X_{\max}$. Isto indica que o enigma dos muóns xorde dun efecto acumulativo, máis que de fenómenos novos nas enerxías máis altas. Ademais, as composicións de masa inferidas a partir da profundidade máxima da distribución das profundidades de produción de muóns están en notable tensión cos resultados de $X_{\max}$.

Máis recentemente, a distribución conxunta de $X_{\max}$ e dos sinais medidos polos WCDs no Observatorio Auger axustouse a plantillas de Monte Carlo que permitían liberdade na escala de $X_{\max}$ e na normalización do sinal de muóns. Os resultados amosan que o mellor axuste aos datos de Auger conséguese ao profundizar o valor medio $\langle X_{\max} \rangle$ entre 22 e $47\,\mathrm{g\,cm^{-2}}$, dependendo do modelo de interacción hadrónica considerado, e aplicando un reescalado da compoñente de muóns entre o 15% e o 24%, segundo o modelo hadrónico e tamén o ángulo zenital da choiva. Estes resultados suxiren unha reinterpretación da composición de masa do fluxo de raios cósmicos: a masa primaria debe ser máis pesada do que se infire actualmente coas simulacións de cascadas, aliviando parcialmente o déficit de muóns. De xeito crucial, este último resultado combina medicións de diferentes observables da choiva, impoñendo restricións máis fortes ás interaccións hadrónicas nas

enerxías máis altas.

Por tanto, falta un enfoque fisicamente motivado que teña en conta toda a información da interacción primaria de forma consistente e toda a información da distribución completa de N_μ e $X_{\max}$. Isto permitiría restricións baseadas nos datos moi sólidas sobre as interaccións hadrónicas en réximes cinemáticos non sondados polos aceleradores fabricados polo ser humano. Este enfoque podería aliviar as tensións na descrición dos observables das cascadas e levar ao descubrimento de mecanismos físicos importantes que aparecen nas enerxías máis altas.

O estudo presentado neste traballo está baseado en deducións analíticas e en simulacións Monte Carlo xeradas polos simuladores de cascadas atmosféricas CONEX e CORSIKA. O primeiro combina simulacións Monte Carlo con solucións numéricas de ecuacións de cascadas, permitindo a produción eficiente de grandes bibliotecas de cascadas en pouco tempo, o cal é esencial para estudar en detalle distribucións bidimensionais. O segundo emprega unicamente métodos Monte Carlo e utilízase puntualmente para validación e variacións de parámetros específicos que non están dispoñibles en CONEX.

D.2 A distribución conxunta bidimensional do número de muóns e a profundidade do máximo da choiva

Para inferir as características das interaccións hadrónicas a partir da distribución conxunta bidimensional dos observables clave da choiva: o contido de muóns ao nivel do chan, N_μ, e a profundidade do máximo da choiva, $X_{\max}$, primeiro debemos desacoplar dúas fontes principais da súa variabilidade correlacionada: os efectos da dinámica da cascada e da xeometría, e a estocasticidade da produción de hádrons.

En traballos previos, as fluctuacións relativas de N_μ relacionáronse coas fluctuacións da fracción de enerxía que vai ás partículas con interacción hadrónica na interacción primaria: α_1. Ademais, dado o rápido desacoplamento da compoñente electromagnética tras a interacción primaria, as fluctuacións de $X_{\max}$ deben ser sensibles á estocasticidade da produción de partículas nas primeiras etapas da choiva. Por outra banda, $X_{\max}$ está impulsada pola cascada electromagnética, mentres que a produción de muóns procede do decaemento do núcleo hadrónico da choiva. Polo tanto, agardamos unha lixeira anticorrelación entre estes dous observables da choiva, de acordo co balance enerxético de cada sector de partículas na interacción primaria.

No Capítulo 2 estudamos en detalle a sensibilidade da distribución conxunta de N_μ e $X_{\max}$ en función da enerxía primaria, o ángulo zenital, a composición de masa do primario e o limiar de enerxía para a detección dos muóns. De feito, comprobamos que a física da produción de hádrons introduce unha anticorrelación entre N_μ e $X_{\max}$, mentres que a atenuación do fluxo de muóns dependente de $X_{\max}$ induce unha correlación positiva. Ademais, mostramos que a distribución conxunta de N_μ e $X_{\max}$ pode obterse a partir da distribución dos muóns producidos e de $X_{\max}$, moito máis ligada á produción hadrónica, mediante un sinxelo modelo semianalítico de atenuación de muóns. Este modelo ten en conta a dureza do espectro enerxético dos muóns na produción e a súa profundidade máxima de produción. Empregouse para demostrar que a crecente anticorrelación entre N_μ e $X_{\max}$ coa enerxía primaria procede directamente dos procesos hadrónicos.

Ademais, explicamos a dependencia co ángulo zenital da distribución bidimensional

de N_μ e $X_{\max}$ en termos das fluctuacións da truncación do perfil de muóns polo nivel do chan. Este importante efecto xeométrico induce, para choivas verticais, unha fonte adicional de anticorrelación entre os dous observables clave da choiva. De xeito crucial, atopamos que o principio de superposición non explica a dependencia coa masa primaria da correlación entre N_μ e $X_{\max}$, posiblemente debido a procesos de fragmentación nuclear nas interaccións núcleo-núcleo.

D.3 Sondaxe da sección eficaz inclusiva das pións neutras con chuviass con déficit de muóns

Se o espectro enerxético das pionas neutras producidas na interacción primaria é máis duro, entón aumenta a frecuencia de eventos con baixo contido de muóns, xa que a enerxía dispoñible para a produción de muóns diminúe. Ademais, as fluctuacións na interacción primaria son o principal motor das fluctuacións no número de muóns. Este feito pode empregarse para constrinxir a sección eficaz diferencial inclusiva para a produción de pionas neutras usando a distribución de N_μ.

No Capítulo 3 demostramos que a forma da distribución do número de muóns en choivas con déficit de muóns, descrita polo parámetro Λ_μ, evoluciona con $X_{\max}$ dunha maneira que reflicte a actividade hadrónica da interacción primaria. A alta actividade hadrónica está asociada con interaccións primarias altamente inelásticas nas que se producen moitas partículas secundarias, degradando de forma equitativa a enerxía primaria. Constatamos que a evolución mencionada é moi sensible ás propiedades da produción hadrónica. De xeito crucial, mostramos que a medida da evolución de Λ_μ coa $X_{\max}$ é experimentalmente factible cos detectores actuais de cascadas atmosféricas, e é robusta mesmo na presenza dunha composición mixta de raios cósmicos.

D.4 Marco para inferir propiedades das interaccións hadrónicas usando o contido de muóns e a profundidade do máximo das EAS

No Capítulo 4 derivamos un novo conxunto de variables —$\zeta_{\rm had}$, $\zeta_{\rm EM}$ e $\alpha_{\rm had}$— que caracterizan os espectros enerxéticos das partículas secundarias producidas na interacción primaria protón-atmosfera. Esta derivación xorde da generalización de modelos semianalíticos de cascadas. Mostramos que as variables mencionadas poden combinarse linealmente nun estimador de $X_{\max}$, ξ, que explica a maior parte da variabilidade deste observable da choiva que procede exclusivamente da interacción primaria. Este achado confirma que a maior parte das fluctuacións en $X_{\max}$ teñen a súa orixe na primeira interacción, contribuíndo o desenvolvemento secundario da chuvia só de forma subdominante. É importante destacar que demostramos que as distribucións de $\zeta_{\rm had}$, $\zeta_{\rm EM}$ e $\alpha_{\rm had}$ son moi sensibles á forma do espectro enerxético das secundarias da interacción primaria, e que poden ser constrinxidas de forma directa en detectores de partículas na rexión cara adiante en experimentos con aceleradores, establecendo un vínculo explícito entre as medicións en aceleradores e as das cascadas atmosféricas.

No capítulo 5, dada a forte conexión causal entre ξ e $X_{\max}$, e entre α_1 e N_μ, mostramos que a anticorrelación entre N_μ e $X_{\max}$ está motivada pola competencia enerxética entre os sectores hadrónico e electromagnético, codificada na anticorrelación entre α_1 e ξ. Os efectos estocásticos das etapas posteriores da choiva son subdominantes na configuración da distribución conxunta de N_μ e $X_{\max}$.

Motivados por esta forte conexión, derivamos o resultado central desta tese: un marco probabilístico que modela a resposta da choiva á interacción inicial mediante un núcleo universal. Este núcleo conecta calquera distribución previa fisicamente consistente de α_1 e ξ coa distribución conxunta observable de N_μ e $X_{\max}$, permitindo predicila con precisión significativamente superior á dispersión entre modelos de interacción hadrónica e dentro das incertezas experimentais actuais. De xeito crucial, demostramos que características sensibles principalmente ás choivas inducidas por protóns, para as que se construíu o marco, como Λ_μ, poden predicirse con alta precisión só a partir das propiedades da interacción primaria. Isto ofrece unha ferramenta robusta para constrinxir a sección eficaz inclusiva de produción de pionas neutras en interaccións protón-atmosfera.

Este formalismo universal non só explica a distribución de $f(n_\mu, X_{\max})$ nas choivas inducidas por protóns, senón que tamén establece as bases para a súa extensión a primarios máis pesados. As rexións da distribución conxunta dominadas por protóns xa están ben capturadas, o que suxire que mostras puras en composición poderían proporcionar restricións directas sobre as interaccións hadrónicas. Unha futura generalización do modelo a núcleos permitiría axustes globais á distribución completa, posibilitando inferir propiedades hadrónicas a partir dos datos das EAS. Deste xeito, o marco representa un paso significativo cara á resolución das discrepancias entre simulacións e observacións.

D.5 Medición de variables de interacción hadrónica usando datos de Auger

No capítulo final da tese, o Capítulo 6, establecemos relacións universais entre a profundidade media do máximo da choiva, $\langle X_{\max} \rangle$, e variables clave da produción multipartícula que caracterizan a primeira interacción protón-aire. Estas relacións son en grande medida independentes da enerxía primaria e mantéñense entre os modelos modernos de interacción hadrónica que, a pesar das súas diferenzas, amosan unha rexión físicamente plausible do espazo de fases baseada na mesma física subxacente. Esta universalidade permite calibracións robustas que vinculan observables como $\langle \alpha_{\text{had}} \rangle$, $\langle \zeta_{\text{had}} \rangle$ e $\langle \zeta_{\text{EM}} \rangle$ con $\langle X_{\max} \rangle$.

Deste xeito, aproveitamos os desprazamentos medidos en $\langle X_{\max} \rangle$ preferidos polos datos de Auger para estimar valores medios dun conxunto de variables de produción multipartícula, incluíndo $\langle \alpha_{\text{had}} \rangle$, $\langle \zeta_{\text{had}} \rangle$ e $\langle \zeta_{\text{EM}} \rangle$. Este estudo baséase na independencia coa masa destes desprazamentos, no feito de que a sección eficaz protón-atmosfera medida é compatible coas predicións dos diferentes modelos de interacción hadrónica, e na suposición de que non existen escenarios físicos exóticos que aparezan bruscamente a ultraaltas enerxías. De xeito crucial, atopamos que o aumento suxerido na fracción de enerxía hadrónica é consistente coa mellora necesaria no contido de muóns nas cascadas atmosféricas, ofrecendo unha explicación coherente tanto para o desprazamento en $X_{\max}$ como para o reescalado dos muóns.

Por tanto, o Capítulo 6 exemplifica como as variables empregadas no marco desen-

volvido nos Capítulos 4 e 5 poden usarse para constrinxir os espectros enerxéticos das partículas secundarias en interaccións protón-aire, ofrecendo unha sonda novedosa da produción hadrónica en rexións cinemáticas máis alá do alcance dos experimentos actuais de colisionadores.

E

PERMISSIONS FOR PUBLISHED CONTENT IN PEER REVIEWED JOURNALS

The permission for reference [136] is attached below:

American Physical Society
Reuse and Permissions License

30-Aug-2025

This license agreement between the American Physical Society ("APS") and Miguel Alexandre Martins ("You") consists of your license details and the terms and conditions provided by the American Physical Society and SciPris.

Licensed Content Information

License Number:	**RNP/25/AUG/096054**
License date:	30-Aug-2025
DOI:	10.1103/1664-pht6
Title:	Probabilistic mapping between multiparticle production variables and the depth of maximum in proton-induced extensive air showers
Author:	Lorenzo Cazon et al.
Publication:	Physical Review D
Publisher:	American Physical Society
Cost:	USD $ 0.00

Request Details

Does your reuse require significant modifications: No

Specify intended distribution locations:	Worldwide
Reuse Category:	Reuse in a thesis/dissertation
Requestor Type:	Author of requested content
Items for Reuse:	Figures/Tables
Number of Figure/Tables:	18
Figure/Tables Details:	Figures 4.1, 4.7, 4.8, 4.9, 4.11, 4.13, 4.14, 4.18, 4.19, 4.22, 4.25, 4.27, 4.30, 4.33, 4.34, 4.35 Tables 4.4 and 4.6
Format for Reuse:	Electronic

Information about New Publication:

University/Publisher:	Universidade de Santiago de Compostela
Title of dissertation/thesis:	Inferring hadronic interaction properties from the 2D-distribution of muon content and depth of shower maximum of extensive air showers
Author(s):	Miguel Alexandre Jesus da Silva Martins
Expected completion date:	Oct. 2025

License Requestor Information

Name:	Miguel Alexandre Martins
Affiliation:	Individual
Email Id:	miguelalexandre.jesusdasilva@usc.es
Country:	Spain

American Physical Society
Reuse and Permissions License

TERMS AND CONDITIONS

The American Physical Society (APS) is pleased to grant the Requestor of this license a non-exclusive, non-transferable permission, limited to Electronic format, provided all criteria outlined below are followed.

1. You must also obtain permission from at least one of the lead authors for each separate work, if you haven't done so already. The author's name and affiliation can be found on the first page of the published Article.
2. For electronic format permissions, Requestor agrees to provide a hyperlink from the reprinted APS material using the source material's DOI on the web page where the work appears. The hyperlink should use the standard DOI resolution URL, http://dx.doi.org/{DOI}. The hyperlink may be embedded in the copyright credit line.
3. For print format permissions, Requestor agrees to print the required copyright credit line on the first page where the material appears: "Reprinted (abstract/excerpt/figure) with permission from [(FULL REFERENCE CITATION) as follows: Author's Names, APS Journal Title, Volume Number, Page Number and Year of Publication.] Copyright (YEAR) by the American Physical Society."
4. Permission granted in this license is for a one-time use and does not include permission for any future editions, updates, databases, formats or other matters. Permission must be sought for any additional use.
5. Use of the material does not and must not imply any endorsement by APS.
6. APS does not imply, purport or intend to grant permission to reuse materials to which it does not hold copyright. It is the requestor's sole responsibility to ensure the licensed material is original to APS and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover or table does not indicate it was reprinted by APS with permission from another source.
7. The permission granted herein is personal to the Requestor for the use specified and is not transferable or assignable without express written permission of APS. This license may not be amended except in writing by APS.
8. You may not alter, edit or modify the material in any manner.
9. You may translate the materials only when translation rights have been granted.
10. APS is not responsible for any errors or omissions due to translation.
11. You may not use the material for promotional, sales, advertising or marketing purposes.
12. The foregoing license shall not take effect unless and until APS or its agent, Aptara, receives payment in full in accordance with Aptara Billing and Payment Terms and Conditions, which are incorporated herein by reference.
13. Should the terms of this license be violated at any time, APS or Aptara may revoke the license with no refund to you and seek relief to the fullest extent of the laws of the USA. Official written notice will be made using the contact information provided with the permission request. Failure to receive such notice will not nullify revocation of the permission.
14. APS reserves all rights not specifically granted herein.
15. This document, including the Aptara Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

F

FIGURES AND TABLES

List of Figures

List of Tables

Ultra-high-energy cosmic rays triggering Extensive Air Showers in the Earth's atmosphere provide a unique opportunity to probe hadronic interactions. This thesis develops a probabilistic framework to infer properties of these interactions in regions of the kinematic phase space poorly constrained by accelerator data, using the two-dimensional distribution of key shower observables: the shower muon content and the depth of the shower maximum. The precision achieved in the prediction of this distribution for any physically consistent prior for the primary proton-air interaction allows for hadronic-interaction-model discrimination, within typical experimental uncertainties. Constraints on the energy spectra of secondaries from the primary interaction are derived using data from the Pierre Auger Observatory.