PROCEEDINGS OF SCIENCE

A Universal Mapping Between Proton-Air Interaction Variables and Air Shower Observables: Depth of Shower Maximum and Muon Content

Lorenzo Cazon,[a] **Ruben Conceição,**[b,c] **Miguel Alexandre Martins**[a,*] **and Felix Riehn**[d]

[a] *Instituto Galego de Física de Altas Enerxías (IGFAE), Rúa de Xoaquín Díaz de Rábago, s/n, Campus Vida, Universidade de Santiago de Compostela, 15705, Santiago de Compostela, Galicia, Spain*

[b] *Departamento de Física, Instituto Superior Técnico (IST), Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal*

[c] *Laboratório de Instrumentação e Física Experimental de Partículas (LIP) - Lisbon, Av. Prof. Gama Pinto 2, 1649-003 Lisbon, Portugal*

[d] *Technische Universität Dortmund, August-Schmidt-Straße 4, 44221 Dortmund, Germany*

E-mail: miguelalexandre.jesusdasilva@usc.es

We introduce a set of new multiparticle production variables derived from the energy spectrum of secondary hadrons in ultra-high-energy proton-air interactions. The distributions of these variables can be measured within the phase space accessible to particle detectors in accelerator experiments, and are highly dependent on the hadronic interaction model. Furthermore, we demonstrate a precise, hadronic-model-independent mapping between these variables and the joint distribution of the depth of shower maxima and the number of muons in extensive air showers. This enables the use of air shower measurements to constrain hadron production in kinematic regimes beyond the reach of human-made colliders.

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland

*Speaker

PoS(ICRC2025)328

1. Introduction

Ultra-high-energy cosmic rays (UHECRs) interact with atmospheric nuclei at center-of-mass energies above 50 TeV, initiating Extensive Air Showers (EAS) governed by hadronic interactions at low momentum transfer. Their modelling relies on phenomenological extrapolations from accelerator data into the far-forward kinematic region. The lack of collider data and differing assumptions in hadronic interaction models leads to inconsistent interpretations of EAS observables [1–4].

Two key EAS observables are the depth of the shower maximum, $X_{\max}$, and the number of muons at ground, N_μ. Measurements of the $X_{\max}$ distribution enabled estimates of the proton-air cross-section at $\sqrt{s} > 57$ TeV [5, 6], consistent with model predictions. In contrast, the average muon number, $\langle N_\mu \rangle$—which reflects the history of hadron production—is systematically underestimated in simulations. This *Muon Puzzle* [2] depends on the composition inferred from $X_{\max}$, though data from the Pierre Auger Observatory support a reinterpretation of the mass composition of the cosmic ray flux at EeV energies [3]. Moreover, relative fluctuations in N_μ, mainly determined by the primary interaction [7], are well described by models [8], suggesting the muon deficit arises from a cumulative effect rather than exotic physics. These tensions may be alleviated by establishing a model-independent link between shower observables and hadronic physics, enabling data-driven constraints on hadronic interactions.

Some proposed constraints include: the inclusive π^0 production cross-section in proton-air collisions [9], via the shape of the N_μ distribution in muon-depleted showers; and the evolution of this cross-section with the hadronic activity of the primary interaction through a scan in $X_{\max}$ [10].

This contribution extends the formalism introduced in [11] to link the joint distribution of N_μ and $X_{\max}$ to the energy spectrum of secondaries of the primary interaction. Section 2 introduces new multiparticle production variables derived from secondary energy spectra, whose linear combination ξ explains ~50% of the variability in $\Delta X_{\max} \equiv X_{\max} - X_1$, where X_1 is the depth of the first interaction point. Section 3 interprets the joint N_μ–$X_{\max}$ distribution in terms of these variables, and Section 4 presents a universal, model-independent probabilistic mapping from secondary spectra to this joint distribution. This framework makes it possible to predict the N_μ and $X_{\max}$ distributions within current experimental and EAS-modelling uncertainties, thereby enabling data-driven tests of hadronic physics beyond the reach of human-made accelerators.

2. A new set of multiparticle production variables

In [7], the energy spectrum of hadronically interacting secondaries—i.e., excluding γ, $e^\pm$, π^0, and η mesons—of the primary proton-air interaction was connected to the shower-to-shower fluctuations of the relative muon number, $n_\mu \equiv N_\mu / \langle N_\mu \rangle$, via the multiparticle production variable

$$\alpha_1 = \sum_{i=1}^{m_{\text{had}}} x_i^{0.93}, \tag{1}$$

where x_i is the fraction of the primary energy carried by hadronically interacting secondary i in the lab. frame, and m_{had} is their multiplicity. The variance of α_1 accounts for ~ 70% of that of n_μ.

The stochastic nature of the primary interaction also affects the distribution of $X_{\max}$ via fluctuations in $\Delta X_{\max} = X_{\max} - X_1$, thereby encoding information about the energy spectrum of secondaries into this key observable.

2.1 Combining production variables into a primary interaction estimator of $X_{\max}$

As derived in [11], for proton primaries with $E_0 = 10^{19}$ eV, the distribution of $\Delta X_{\max}$ is linked to the energy spectrum of secondaries from the primary interaction via the production variable

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

a linear combination of the elementary multiparticle production variables

$$\zeta_{\rm had} = -\sum_{i=1}^{m_{\rm had}} x_i \ln x_i, \quad \zeta_{\rm EM} = -\sum_{j=1}^{m_{\rm EM}} x_j \ln x_j, \quad \alpha_{\rm had} = \sum_{i=1}^{m_{\rm had}} x_i, \tag{3}$$

where x_j denotes the fractions of the primary energy, in the laboratory frame, carried by each of the $m_{\rm EM}$ electromagnetic secondaries. The variable $\alpha_{\rm had}$ is the fraction of energy carried by hadronically interacting particles and quantifies the energy available for muon production. The variable $\zeta_{\rm had}$ reflects the hadronic activity of the primary interaction: high values indicate inelastic interactions with many secondary particles evenly carrying the primary energy, while low values correspond to *quasi*-elastic collisions. The values of $\zeta_{\rm EM}$ are interpreted analogously for the electromagnetic sector. The distributions of these variables can be measured within the kinematic phase-space covered by state-of-the-art particle detectors in accelerator experiments and are highly sensitive to the shape of the energy spectrum of hadrons [11].

The correlation between ξ and $\Delta X_{\max}$ was verified using 10^5 CONEX simulations of proton-induced showers at $E_0 = 10^{19}$ eV and $\theta = 60°$, employing the models EPOS LHC-R [12], QGSJET-III.01 [13], and SIBYLL2.3e [14]. Particles with $E > 0.005\,E_0$ are tracked individually; below this threshold, the longitudinal profiles of particles are obtained by numerically solving cascade equations. The ground level is set to 1 400 m a.s.l. ($X_{\rm gr} = 880\,\mathrm{g\,cm^{-2}}$), matching the altitude of the Pierre Auger Observatory [15]. The observable N_μ denotes the muons above 1 GeV counted at ground, and $X_{\max}$ is obtained from a Gaisser–Hillas fit to the profile of all charged particles.

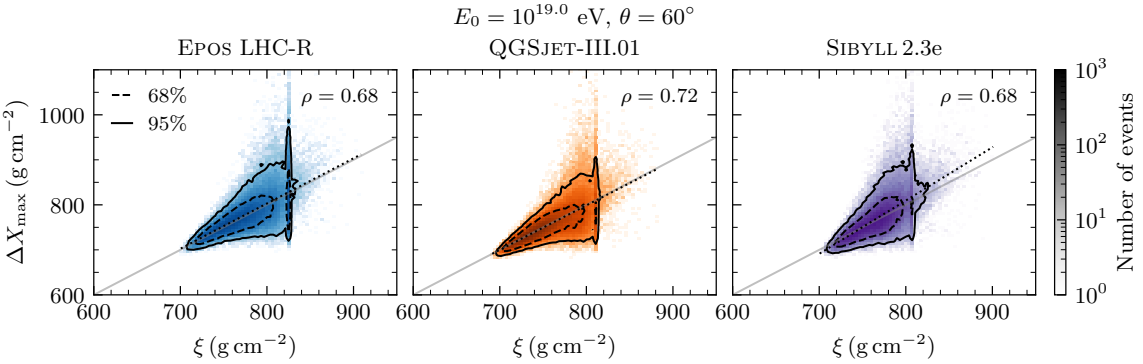

Figure 1: Joint distribution of ξ and $\Delta X_{\max}$ for proton showers with $E_0 = 10^{19}$ eV and $\theta = 60°$, simulated with different hadronic interaction models

Figure 1 shows a strong correlation between ξ and $\Delta X_{\max}$, quantified by Pearson correlation coefficients ranging from 0.68 to 0.72 and thus largely independent of the hadronic model. The variance in ξ explains $\sim 50\%$ of the variance in $\Delta X_{\max}$. When all information from the primary

interaction is retained (i.e., without approximations in ξ), this rises to $\sim 65\%$, implying that $\sim 80\%$ of the relevant information is preserved in the linear combination ξ.

Since the electromagnetic and hadronic cascades decouple rapidly, the scale of $\langle X_{\max} \rangle$ is mostly determined by the first few interactions of an EAS [16]. Thus, $\langle \xi \rangle$ should correlate closely with $\langle X_{\max} \rangle$ through $\langle \Delta X_{\max} \rangle$. This is confirmed across several modern hadronic models, including EPOS LHC [17], QGSJET-II.04 [18], and SIBYLL2.3d [14], as shown in Figure 2. A linear relation, $\langle X_{\max} \rangle = m \ \langle \xi \rangle + b$, is fitted to model predictions.

The universality of the calibration from $\langle \xi \rangle$ to $\langle X_{\max} \rangle$ reflects: the agreement among models for the proton–air cross-section (within 3 g cm^{-2} at $E_0 = 10^{19}$ eV); and a consistent propagation of energy spectrum differences in the primary interaction, captured by $\langle \xi \rangle$, through subsequent interactions, producing proportional values of $\langle X_{\max} \rangle$.

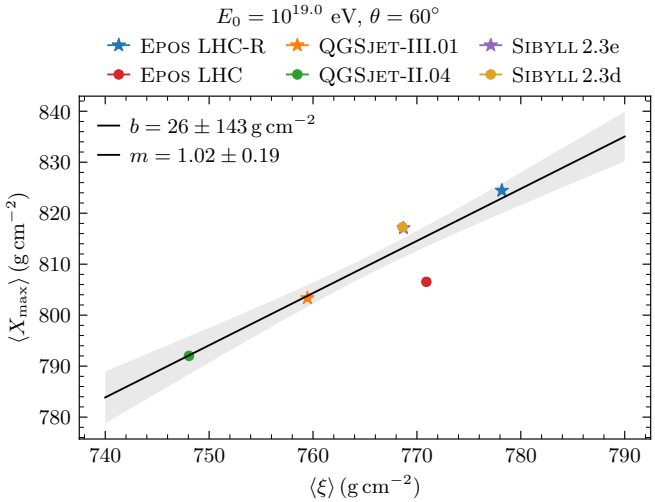

Figure 2: Calibration of $\langle X_{\max} \rangle$ as a linear function of $\langle \xi \rangle$: $\langle X_{\max} \rangle = m \ \langle \xi \rangle + b$, obtained by fitting the predictions of the current hadronic interaction models.

3. Interpreting the regions of the joint distribution of N_μ and $X_{\max}$

Given the strong connection between $\alpha_1 \approx \alpha_{\text{had}}$ and N_μ, and between ξ and $X_{\max}$, we can analyse the joint distribution $p(N_\mu, X_{\max})$ in terms of the multiparticle production variables ζ_{had}, ζ_{EM}, and α_{had}, as shown in Figure 3.

Despite fluctuations in later shower generations and those of X_1, distinct regions in the N_μ–$X_{\max}$ plane correspond to different triads $(\zeta_{\text{had}}, \zeta_{\text{EM}}, \alpha_{\text{had}})$, characteristic of specific types of primary interactions. High ζ_{had} values—typical of highly inelastic interactions with many secondaries—lead to shallow, muon-rich showers, as the primary energy is degraded rapidly. In contrast, low ζ_{had} indicates more elastic events, yielding deeper showers. Notably, the gradient of ζ_{had} is nearly vertical and uniform across the N_μ–$X_{\max}$ plane, enabling a clear separation of interaction types.

Below $N_\mu / \langle N_\mu \rangle < 1$, the shallowest showers for a given muon count are associated with large ζ_{EM}, corresponding to interactions with many neutral pions distributing the primary energy into the electromagnetic channel. This again leads to shallower showers, that get deeper with decreasing α_{had}. This latter observable, governs the muon yield, irrespective of $X_{\max}$.

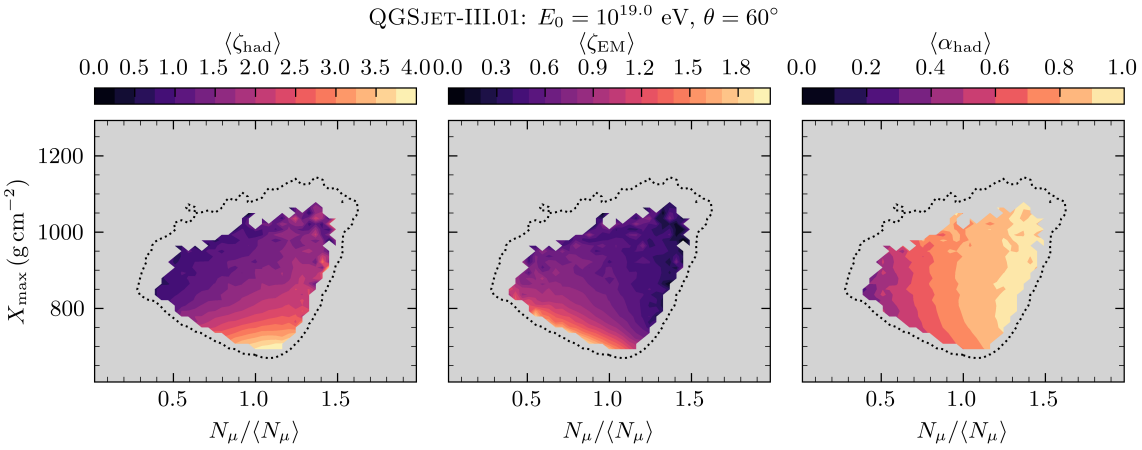

Figure 3: Contours of $\zeta_{\rm had}$ (left), $\zeta_{\rm EM}$ (middle) and $\alpha_{\rm had}$ (right) over the joint distribution $p(n_\mu, X_{\max})$. The outline of the latter distribution is represented by a dotted black contour.

Overall, the structure of the N_μ–$X_{\max}$ plane retains clear imprints of the primary interaction, enabling a direct, intuitive connection between its physics and the mentioned shower observables.

4. Mapping primary interaction variables onto N_μ and $X_{\max}$

We now present a framework to recover the information about the primary interaction retained in the joint distribution of N_μ and $X_{\max}$.

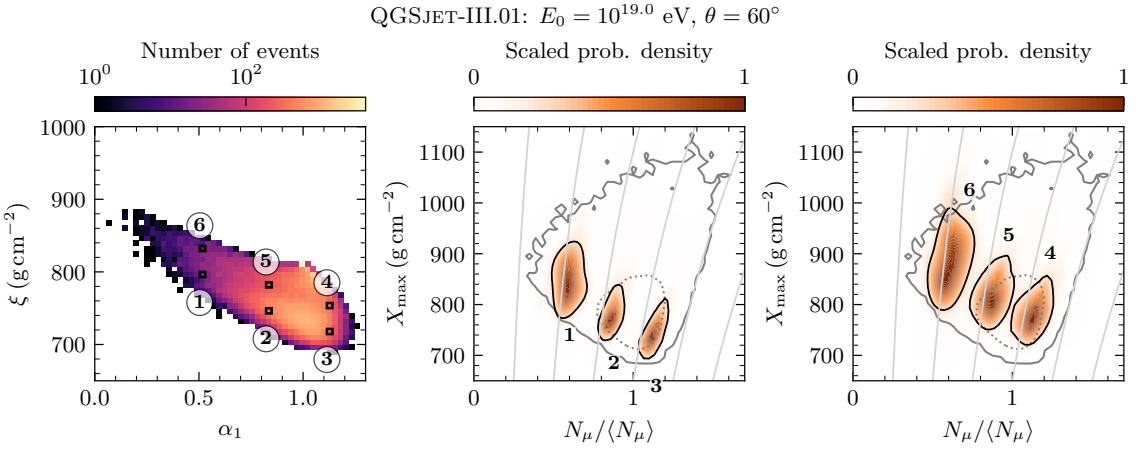

Figure 4: Examples of the probabilistic shower response $p(n_\mu, X_{\max} \mid \alpha_1, \xi, \langle\xi\rangle)$ for 6 narrow bins of the joint distribution $p(\alpha_1, \xi)$ provided by CONEX simulations of proton-induced showers employing QGSJET -III.01. The grey curvilinear grid lines are the contours of a constant number of produced muons.

For hadronic interaction model M, the energy spectra of secondaries from the primary interaction—governing the shape of $p(\alpha_1, \xi)$—is mapped to $p(n_\mu, X_{\max})$ via the integral transformation

$$p(n_\mu, X_{\max}) = \iint p(n_\mu, X_{\max} \mid \alpha_1, \xi, M)\, p(\alpha_1, \xi)\, \mathrm{d}\alpha_1\, \mathrm{d}\xi\,, \tag{4}$$

where $p(n_\mu, X_{\max} \mid \alpha_1, \xi, M)$ is the shower response to a given pair (α_1, ξ), i.e., the kernel of the integral transformation. Since the primary interaction sets most of the shape of $p(n_\mu, X_{\max})$, the kernel is narrow. To test whether the kernel's hadronic model dependence can be removed, we average it across models and calibrate the mean depth $\langle X_{\max} \rangle$ using its linear relation with $\langle \xi \rangle$ (see Figure 2). This ensures changes in the primary interaction also affect the response of the shower, as observed for all hadronic interaction models. The result is a universal kernel $p(n_\mu, X_{\max} \mid \alpha_1, \xi, \langle \xi \rangle)$, valid in the absence of exotic physics. The presented reasoning is further elaborated in [11].

Examples of $p(n_\mu, X_{\max} \mid \alpha_1, \xi, \langle \xi \rangle)$ for six bins in $p(\alpha_1, \xi)$ are shown in Figure 4, along with the prior $p(\alpha_1, \xi)$ and the outline of $p(n_\mu, X_{\max})$. Light contours in the target space indicate a constant muon number.

The joint distribution $p(\alpha_1, \xi)$ shows a correlation mainly driven by the energy balance between hadronic and electromagnetic sectors. Distinct regions of $p(\alpha_1, \xi)$ are mapped to areas of $p(n_\mu, X_{\max})$ with little overlap, preserving the imprint of the primary interaction. The elongation of the kernel arises from exponential fluctuations in X_1, uncorrelated with particle production in the primary interaction, yet positively correlated with muon attenuation.

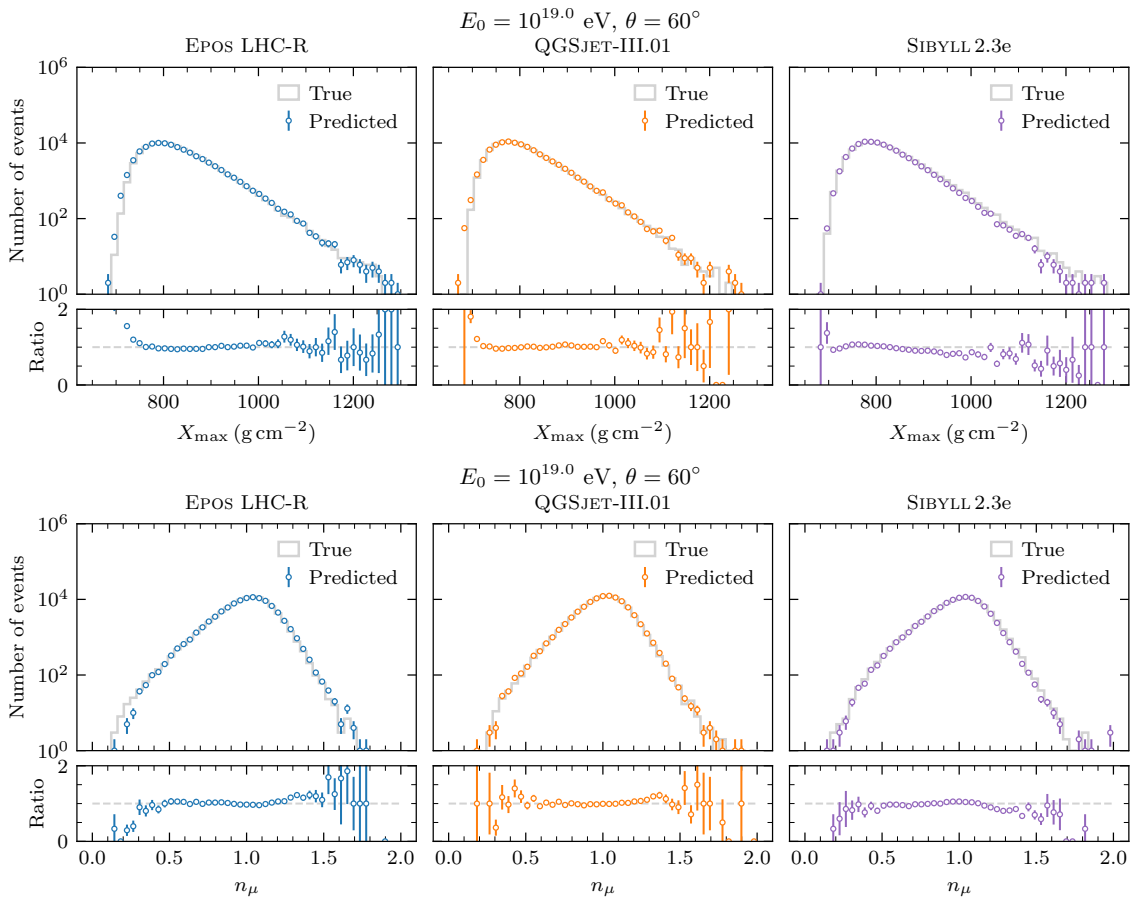

Figure 5: True (grey steps) and predicted distributions (colored markers) of $X_{\max}$ (upper panels) and of n_μ (lower panels) obtained from applying the universal kernel $p(n_\mu, X_{\max} \mid \alpha_1, \xi, \langle \xi \rangle)$ to the priors $p(\alpha_1, \xi)$ provided by different hadronic interaction models.

Using priors $p(\alpha_1, \xi)$ for three models convolved with the universal kernel $p(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$, we obtain the predicted $X_{\max}$ distribution, which is compared to the true $p(X_{\max})$ in the upper panels of Figure 5.

The agreement between the true and predicted distributions of $X_{\max}$ is remarkable. In particular, the upper panels of Figure 6 show biases in the predicted main moments of $p(X_{\max})$: mean (left), standard deviation (middle), and the slope of the deep tail Λ_η [5]. Shaded bands show the spread across hadronic interaction models. Biases in $\langle X_{\max}\rangle$ are below $10\,\mathrm{g\,cm^{-2}}$, less than 30% of the model spread. Biases in $\sigma(X_{\max})$ and Λ_η are below $4\,\mathrm{g\,cm^{-2}}$, also smaller than model spread and well below the experimental uncertainties achieved with fluorescence detectors [19].

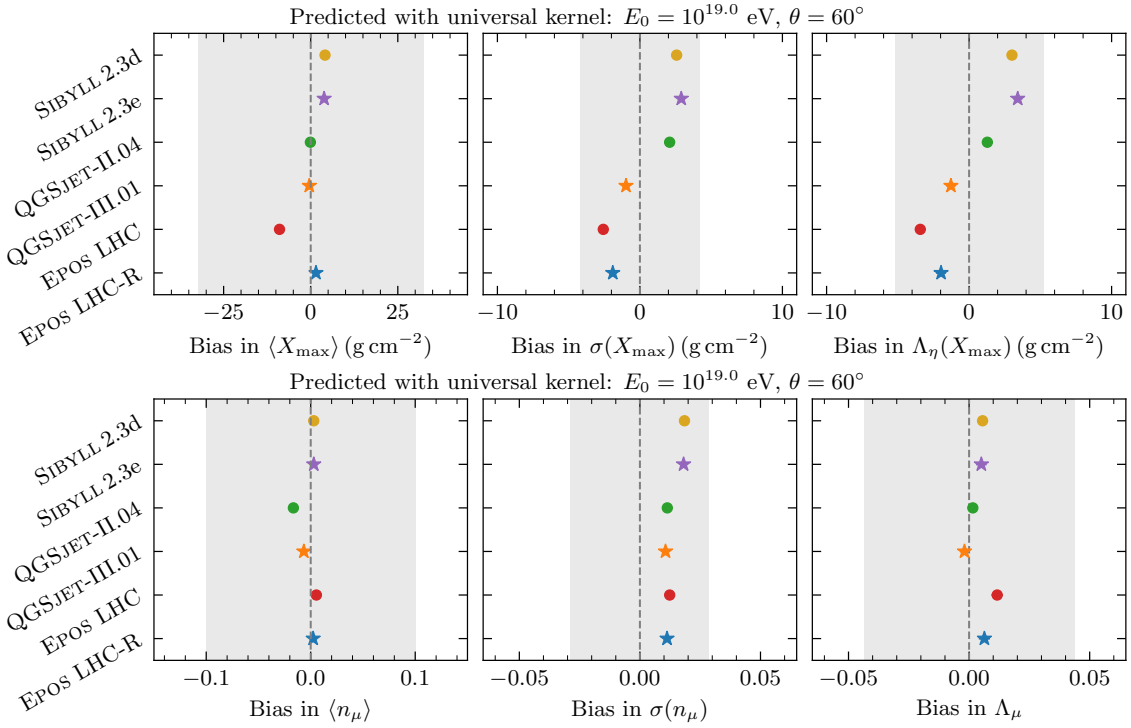

Figure 6: Upper panels: Biases in the main moments of the distribution of $X_{\max}$ predicted with the universal kernel. Lower panels: Bias assessment for the main moments of the distribution of n_μ. The priors $p(\alpha_1, \xi)$ are provided by different hadronic interaction models. Shaded bands show the spread across models, except for $\langle n_\mu\rangle$, for which they represent typical experimental uncertainties. Variables are defined in the main text.

Similar considerations apply to the n_μ distribution (lower panels of Figures 5 and 6). For $\langle n_\mu\rangle$, the bias is well under the typical 10% experimental systematic uncertainty [8]. The slope of the tail of $p(n_\mu)$ towards low values, Λ_μ, which constrains π^0 production in proton-air interactions [9], is also accurately reproduced. This observable is largely insensitive to primary mass and differs significantly between models, supporting model-independent constraints on forward pion production using the framework addressed in this contribution.

In summary, the universal kernel $p(n_\mu, X_{\max} \,|\, \alpha_1, \xi, \langle\xi\rangle)$ enables precise, model-independent reconstruction of n_μ and $X_{\max}$ distributions, enabling data-driven constraints on primary interactions outside the phase covered by particle colliders. As the kernel is currently valid for proton showers, either the formalism is extended to heavier primaries, or those constraints must be derived from

proton-enriched regions of $p(n_\mu, X_{\max})$.

5. Conclusions

This contribution presents an extension of the formalism developed in [11]. It introduces a set of multiparticle production variables—$\zeta_{\rm had}$, $\zeta_{\rm EM}$, $\alpha_{\rm had}$, and α_1—which capture the information contained in the energy spectrum of secondaries from the primary interaction and enable its mapping onto the joint distribution of n_μ and $X_{\max}$.

This probabilistic mapping is independent of the hadronic interaction model considered and, when applied to different prior distributions of $p(\xi, \alpha_1)$, enables the reconstruction of the N_μ and $X_{\max}$ distributions. The achieved biases in the main moments of n_μ and $X_{\max}$ are below current experimental and/or hadronic-model systematic uncertainties, therefore enabling the probing of hadronic physics beyond the reach of human-made accelerators, using EAS data.

References

[1] Pierre Auger Collaboration collaboration, *Phys. Rev. D* **91** (2015) 032003.

[2] J. Albrecht et al., *Astrophys. Space Sci.* **367** (2022) 27 [2105.06148].

[3] Pierre Auger Collaboration collaboration, *Phys. Rev. D* **109** (2024) 102001 [2401.10740].

[4] IceCube collaboration, *PoS* **UHECR2024** (2025) 035.

[5] Pierre Auger Collaboration collaboration, *Phys. Rev. Lett.* **109** (2012) 062002.

[6] R.U. Abbasi et al., *Phys. Rev. D* **102** (2020) 062004 [2006.05012].

[7] L. Cazon, R. Conceição and F. Riehn, *Phys. Lett. B* **784** (2018) 68 [1803.05699].

[8] Pierre Auger collaboration, *Phys. Rev. Lett.* **126** (2021) 152002 [2102.07797].

[9] L. Cazon, R. Conceição, M.A. Martins and F. Riehn, *Phys. Rev. D* **103** (2021) 022001 [2006.11303].

[10] L. Cazon, R. Conceição, M.A. Martins and F. Riehn, *Phys. Lett. B* **859** (2024) 139115 [2406.08620].

[11] L. Cazon, R. Conceição, M.A. Martins and F. Riehn, 2504.08610.

[12] T. Pierog and K. Werner, *PoS* **ICRC2023** (2023) 230.

[13] S. Ostapchenko, *Phys. Rev. D* **109** (2024) 034002 [2401.06202].

[14] F. Riehn, R. Engel, A. Fedynitch, T.K. Gaisser and T. Stanev, *Phys. Rev. D* **102** (2020) 063002.

[15] Pierre Auger Collaboration collaboration, *Nucl. Instrum. Meth. A* **798** (2015) 172 [1502.01323].

[16] S. Ostapchenko and M. Bleicher, *Phys. Rev. D* **93** (2016) 051501 [1601.06567].

[17] T. Pierog, I. Karpenko, J.M. Katzy, E. Yatsenko and K. Werner, *Phys. Rev. C* **92** (2015) 034906 [1306.0121].

[18] S. Ostapchenko, *Phys. Rev. D* **83** (2011) 014018 [1010.1869].

[19] Pierre Auger Collaboration collaboration, *Phys. Rev. D* **90** (2014) 122005.