PROCEEDINGS OF SCIENCE

A Top-Down Approach to the Muon Puzzle: Validation of the method using the new EPOS LHC-R and QGSJET-III hadronic interaction models

Kevin Almeida Cheminant,[a,b,*] **Nataliia Borodai,**[c] **Ralph Engel,**[d,e] **Dariusz Góra,**[c] **Tanguy Pierog,**[e] **Jan Pękala,**[c] **Markus Roth,**[e] **Michael Unger,**[e] **Darko Veberič**[e] **and Henryk Wilczyński**[c]

[a] *NIKHEF, Science Park, Amsterdam, The Netherlands*
[b] *IMAPP, Radboud University, Nijmegen, The Netherlands*
[c] *Institute of Nuclear Physics PAN, Krakow, Poland*
[d] *Karlsruhe Institute of Technology, ETP, Karlsruhe, Germany*
[e] *Karlsruhe Institute of Technology, IAP, Karlsruhe, Germany*

E-mail: kchemina@nikhef.nl

The muon content predicted by hadronic interaction models falls short of describing the data from multiple air shower experiments. This discrepancy, known as the Muon Puzzle, poses significant challenges for mass composition studies and limits our understanding of the origins and acceleration mechanisms of ultra-high-energy cosmic rays. The recent releases of the EPOS LHC-R and QGSJET-III models provide a new opportunity to investigate this divergence using a top-down approach to air shower simulations. This strategy consists of constraining the electromagnetic component of a simulated air shower by matching its longitudinal profile to that of an observed air shower. Consequently, any inconsistency found between the simulated and observed signal in ground particle detectors must originate from a mismatch in the muon content. In the present work, the top-down analysis is tested on a mock dataset that includes air showers simulated with EPOS LHC-R at around 10 EeV and reconstructed using the Pierre Auger Observatory framework. The top-down simulations are performed using QGSJET-III, considering proton, helium, oxygen and iron nuclei as primary particles. The quality of the method is assessed by comparing the true difference in muon content between the two models to that derived from the top-down simulations. Finally, a maximum likelihood estimation accounting for composition is performed to determine the overall hadronic rescaling required to adjust QGSJET-III in order to match the muon content of EPOS LHC-R.

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

*Speaker

© Copyright owned by the author(s) under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

PoS(ICRC2025)173

1. Introduction

Monte Carlo simulations of extensive air showers are essential for interpreting data from ultra-high-energy cosmic-ray (UHECR) experiments — whether in the analysis of arrival directions [1], the search for neutral messengers [2, 3], or the study of mass composition [4, 5]. Hadronic interaction models, which describe the interactions and the energy losses of secondary particles in the atmosphere, are a fundamental cog of the simulation machinery. Over the years, several models have emerged, using different approaches such as Regge–Gribov theory (QGSJET-II [6]), minijet phenomenology (SIBYLL [7]), or parton-based multiple scattering with collective effects (EPOS-LHC [8]). Despite these efforts, the muon content predicted by these models remains insufficient to describe experimental data across different energies, zenith angles, and experiments [9, 10]. To better understand the origin of this discrepancy, we propose a method inspired by the Pierre Auger Collaboration's study of vertical air showers observed in hybrid mode, i.e. simultaneously measured by the surface detector array and by fluorescence telescopes measuring their longitudinal profiles [11]. The method consists in selecting simulated air shower simulations whose longitudinal profiles reproduce those of the hybrid data, thereby constraining the electromagnetic component. This *top-down* approach enables a direct comparison of the total signals in the ground particle detectors, where any remaining difference between data and simulation can only be attributed to the muonic component. In [11], a general muon rescaling factor was derived through a maximum-likelihood estimation, assuming either a pure proton composition or a mixed composition based on the depth of maximum shower development X_{max} measurements. In [12], this approach was tested with more modern hadronic interaction models and air-shower simulation codes, and applied to a mock dataset of air showers simulated with the muon-enhanced Sibyll⋆ model [13], confirming the robustness of the method. In this work, we present a similar analysis aimed at calculating mass-dependent rescaling factors by exploiting several mass-sensitive observables. This approach provides insight into how the muon deficit depends on the primary mass at a given energy, and into the strength of the muon signal as a mass discriminator observable. This paper is structured as follows. In Section 2, we detail the *top-down* simulations chain, followed in Section 3 by a description of the mock dataset and the simulation parameters used to test the method. Section 4 presents the derivation of the mass-dependent muon rescaling factors, and we conclude with a discussion of the results and possible follow-ups to improve the overall analysis.

2. The top-down simulation method

For a given shower with energy E_0 and direction (θ, ϕ), the first step in the top-down simulation chain consists in finding a simulated shower with a similar longitudinal profile, assuming the same energy and direction within uncertainties. To this end, a primary mass is chosen and several thousands of one-dimensional simulations are generated to model only about the longitudinal development of the shower and thus provides a substantial reduction in computing time. Each simulated shower is reconstructed using fluorescence telescopes software, and reconstructed observables such as X_{max}, the calorimetric energy E_{cal}, and the energy deposited at X_{max} are compared with those of the target shower. If all observables agree within their uncertainties, the simulated shower is retained and a χ^2 is calculated by comparing the Gaisser–Hillas fits and measured profiles of

PoS(ICRC2025)173

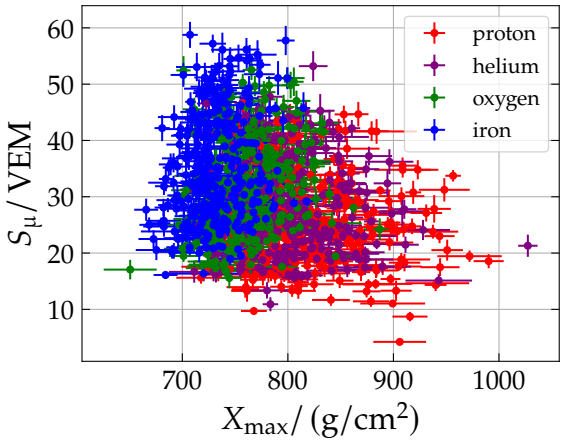

Figure 1: Muon signal at 1000 m from the shower core S_μ, as a function of the depth of maximum shower development X_{max} for the EPOS LHC-R mockup dataset. The primaries found in this dataset are represented by different colors.

both showers. The parameters of the simulated shower yielding the smallest χ^2 are then carried forward to the next step of the simulation chain, where a full Monte-Carlo simulation is performed to obtain the particle distribution at ground. At this stage, the shower is reconstructed using both fluorescence telescopes and surface particle detector software, enabling a direct comparison of the ground signals between the target and the simulated showers. Finally, the entire procedure is repeated, selecting a different primary mass for the simulations.

3. Mock dataset and simulation results

In this work, top-down simulations are performed on a mock dataset of air showers generated with the recently released EPOS LHC-R hadronic interaction model [15]. This model incorporates collective effects of hadronization processes and shows improved agreement with experimental data, including the number of muons observed in air showers. The showers in the mock dataset are simulated with CORSIKA 7.8010 [16] and reconstructed with the Pierre Auger Observatory's 1500 m surface array and the fluorescence telescopes software. These showers have energies between $10^{18.8}$ and $10^{19.2}$ eV and zenith angle below 60°. The mock dataset contains 1280 showers, evenly split among proton, helium, oxygen and iron primaries, selected with quality cuts that ensure well-reconstructed longitudinal profiles. At the Pierre Auger Observatory, the energy estimator of the surface array is the total signal S_{tot} at 1000 m from the shower core, expressed in *Vertical Equivalent Muon* (VEM). This signal corresponds to the sum of the electromagnetic and muonic signals, S_{EM} and S_μ, respectively. Figure 1 shows the distributions of two mass-discriminating observables, X_{max} and S_μ, for all considered primary masses, illustrating that, on average, heavier primaries produce shallower showers with larger muon signals.

Every shower in the mock dataset is then processed through the *top-down* simulation chain described in the previous section. For this purpose, we use the same CORSIKA version as was employed to generate the mock dataset, but select the new QGSJET-III hadronic interaction model [17] to perform the simulations. Throughout the entire chain, the thinning level is set to 10^{-6} and

PoS(ICRC2025)173

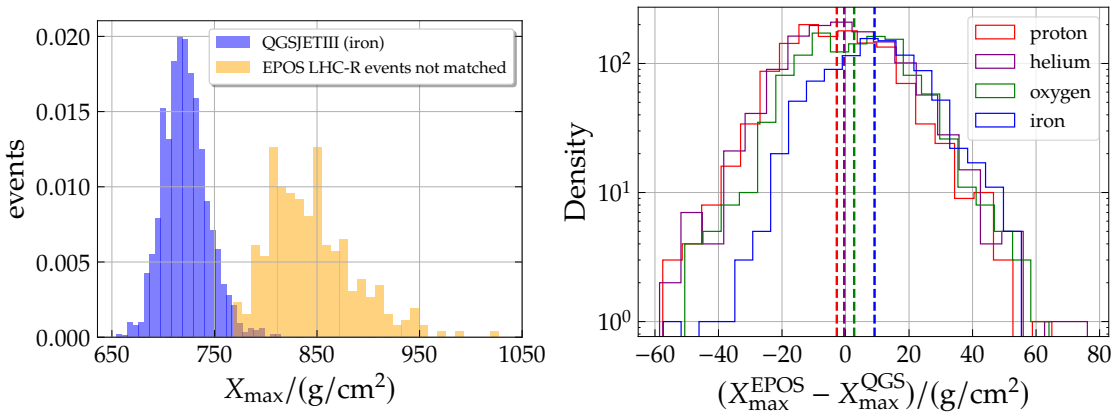

Figure 2: *Left*: X_{max} distribution of iron showers for QGSJET-III hadronic interaction model, around 10^{19} eV (blue) and X_{max} distribution of EPOS LHC-R mock dataset showers that could not be matched with iron showers (orange), following the matching procedure described in Section 2. *Right*: Difference in reconstructed X_{max} between the showers in the mock dataset and that of the best profiles that were matched to them, considering different simulated primary masses. The mean of each distribution is represented by a colored dashed vertical lines.

the energy cuts to 0.05, 0.01, 0.001, and 0.001 GeV, for hadrons, muons, electrons and photons, respectively. These values are low enough to preserve the accuracy of the shower development while still high enough to significantly reduce the computing time.

As each of the 1280 showers in the mock dataset is run through the first step of the *top-down* simulation chain for each considered primaries, we may attempt to match deep proton showers profiles with those of shallower iron showers. However, showers that are too deep to be reproduced by iron showers are not matched with that primary, and conversely, very shallow showers are not matched with proton primaries. This idea is illustrated in the left panel of Figure 2, which shows a typical X_{max} distribution of iron showers for the QGSJET-III model around 10^{19} eV, together with the X_{max} values of the showers in the mock dataset that could not be matched with iron shower profiles. Nevertheless, every shower profile in the mock dataset is reproduced by at least one of the four simulated primaries. In the right panel of Figure 2, we show the difference in reconstructed X_{max} between the EPOS LHC-R showers of the mock dataset and that of the best matching profiles found with QGSJET-III, for all considered primaries. While in some cases, this difference might be as large as 80 g/cm^2, the reconstructed X_{max} agree within uncertainties due to the selection criteria applied during that stage. The larger bias observed for simulated iron primaries arises from the attempt to match a mock dataset containing a large fraction of deep showers with iron showers, which also exhibit smaller shower-to-shower fluctuations than those initiated by lighter primaries. As a result, the differences between the X_{max} values of the mock dataset showers and those of the matched iron showers are larger and a positive bias is observed, as highlighted by the blue dashed line. The reversed situation is also observed for simulated proton primaries, although the larger shower-to-shower fluctuations of proton showers help to significantly reduce the bias.

The full Monte Carlo simulation of the best-matching CONEX shower provides the S_{tot} observable, which can then be compared with that of the corresponding shower in the mock dataset. This comparison is shown in Figure 3. In the left panel, the ratio of S_{tot} between the mock dataset

PoS(ICRC2025)173

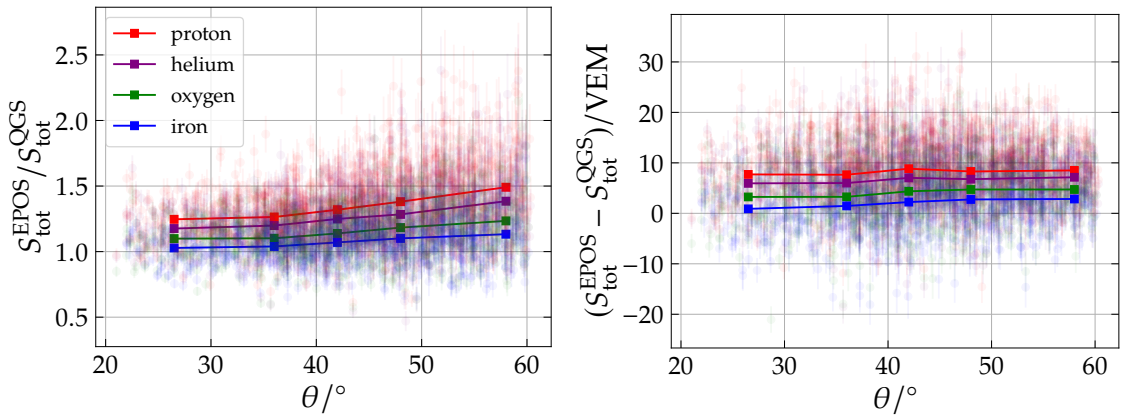

Figure 3: Ratio (left) and difference (right) between the total signal at 1000 meters from the shower core $S_{\rm tot}$ of the mock dataset showers (EPOS) and that of the matching showers (QGS), for all four simulated primaries.

showers and their matching counterparts, considering all four primaries, is seen to increase with zenith angle θ. This trend indicates that the muon signal in the mock dataset grows faster with θ than in the simulations, while the electromagnetic component does not, confirming a muon deficit in QGSJET-III simulations relative to EPOS-LHCR. The right panel shows the difference in $S_{\rm tot}$. Since the top-down approach ensures that the electromagnetic component of QGSJET-III showers matches that of the corresponding mock dataset showers, this difference directly reflects the discrepancy in the muon signals. Since this discrepancy appears to be roughly constant as a function of the zenith angle, it indicates a zenith-independent muon excess at ground (the angular dependence of both hadronic interaction models appears to be similar).

4. Muon rescaling factors

The final output of the *top-down* simulation chain consists of the total and muon signals, $S_{\rm tot}^{\rm QGS}$ and $S_{\mu}^{\rm QGS}$, at 1000 m from the shower core in the QGSJET-III simulations. Since the mock dataset is based on EPOS LHC-R simulations, we also have direct access to the muon signal of the corresponding showers. This allows us to construct the distributions of the muon signal ratio defined as:

$$r_{\mu,i} = \frac{S_{\mu}^{\rm EPOS}}{S_{\mu,i}^{\rm QGS}}, \tag{1}$$

for a simulated primary i — that is, the rescaling factor required for QGSJET-III simulations to reproduce, on average, the EPOS LHC-R muon signal expectation under the assumption of a pure composition of primary i. The $r_{\mu,i}$ distributions for proton and iron primaries are shown by the full histograms in the left panel of Figure 4, and their means with corresponding errors are reported in the first row of Table 1. As expected, the largest rescaling factor is obtained for a pure proton composition, decreasing with increasing primary mass. In practice, for Phase I of the Pierre Auger Observatory data acquisition, the muon signal of vertical showers around 10 EeV is not directly measured. Nevertheless, the *top-down* method allows the computation of the muon rescaling factor

PoS(ICRC2025)173

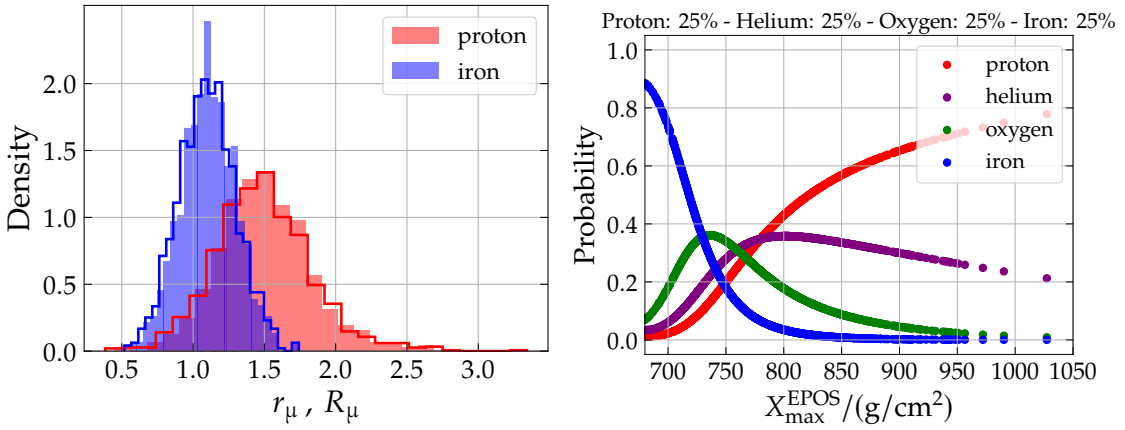

Figure 4: *Left*: Distributions of r_μ (filled histograms) and R_μ (solid-line histograms) for proton and iron primaries. *Right*: Probability for mock dataset showers to originate from a given primary, based on their X_{max} values.

without relying on this observable, as described in [12]:

$$R_{\mu,i} = 1 + \frac{S_{tot}^{EPOS} - S_{tot,i}^{QGS}}{S_{\mu,i}^{QGS}}. \tag{2}$$

If the *top-down* method works as expected, the $r_{\mu,i}$ and $R_{\mu,i}$ distributions should be broadly consistent, as illustrated in the left panel of Figure 4 for proton and iron primaries.
Lastly, a maximum-likelihood estimation can also be used to calculate the rescaling factors. Following the method described in [11], we maximize the likelihood function $\prod_j \mathcal{L}_{i,j}$, for a given primary i, over all events j in the mock dataset, where

$$\mathcal{L}_{i,j} = \frac{1}{\sqrt{2\pi}\sigma_{i,j}} \exp\left[-\frac{1}{2}\left(\frac{S_{tot,j}^{EPOS} - \left(S_{EM,j}^{QGS} + \hat{r}_{\mu,i} S_{\mu,i,j}^{QGS}\right)}{\sigma_{i,j}}\right)^2\right], \tag{3}$$

and where $\sigma_{i,j}$ denotes the uncertainties of the different signal components, added in quadrature. The resulting $\hat{r}_{\mu,i}$ values are reported in the third row of Table 1, showing excellent agreement with the two other methods.

	proton	helium	oxygen	iron
$\langle r_\mu \rangle$	1.50 ± 0.01	1.37 ± 0.01	1.22 ± 0.01	1.11 ± 0.01
$\langle R_\mu \rangle$	1.49 ± 0.01	1.36 ± 0.01	1.20 ± 0.01	1.10 ± 0.01
$\hat{r}_\mu$	1.48 ± 0.01	1.35 ± 0.01	1.19 ± 0.01	1.10 ± 0.01
$\hat{r}_\mu(f_i, X_{max}^{EPOS}, S_\mu^{EPOS})$	1.34 ± 0.04	1.28 ± 0.03	1.23 ± 0.03	1.17 ± 0.02
$r_{\mu,i}^{true}$	1.28 ± 0.02	1.25 ± 0.01	1.26 ± 0.01	1.27 ± 0.01

Table 1: Summary of the muon rescaling factor computations obtained with the different methods described in the text.

PoS(ICRC2025)173

The previous muon rescaling factors were obtained under the assumption of pure compositions. However, since the mock dataset was constructed with equal fractions ($f_i = 25\%$) of the four primary masses, the proton rescaling is overestimated while the iron one is underestimated. Consequently, we incorporate mass-discriminating observables into the likelihood function so that events in the mock dataset of a given primary i contribute with higher weight when calculating the rescaling factor for that primary. This is implemented by multiplying Equation 3 with a probability term that explicitly depends on the $X_{\max}$ and S_{μ} values of the showers in the mock dataset, such that

$$\mathcal{L}'_{i,j}(f_i, X^{\text{EPOS}}_{\max,j}, S^{\text{EPOS}}_{\mu,j}) = p_i(f_i, X^{\text{EPOS}}_{\max,j}, S^{\text{EPOS}}_{\mu,j})\mathcal{L}_{i,j}, \tag{4}$$

where $p_i(f_i, X^{\text{EPOS}}_{\max,j}, S^{\text{EPOS}}_{\mu,j})$ represents the convolution of the probabilities that a EPOS LHC-R shower j in the mock dataset is produced by a primary i, given its $X^{\text{EPOS}}_{\max,j}$ and $S^{\text{EPOS}}_{\mu,j}$ measurements (see right panel of Figure 4 for the probability computation based on $X_{\max}$). Here, we assume that the $X_{\max}$ and S_{μ} distributions in QGSJET-III are described by a Gumbel and a log-normal function, respectively. The results are reported in the fourth row of Table 1, showing that the inclusion of this probability term shifts the muon rescaling factors, yielding a lower value for protons and a higher value for iron primaries, as expected. In order to compare these results to the true muon signal ratios of EPOS LHC-R and QGSJET-III, we must now consider only the *top-down* simulations whose primary masses are identical to that of the mock dataset shower they are matching such that Equation 1 becomes

$$r^{\text{true}}_{\mu,i} = \frac{S^{\text{EPOS}}_{\mu,i}}{S^{\text{QGS}}_{\mu,i}}. \tag{5}$$

The means of these distributions and their corresponding errors are reported in the last row of Table 1. The muon rescaling factors obtained from the maximum-likelihood estimation using $X_{\max}$ and S_{μ} measurements are consistent with the true rescaling factors within uncertainties, except for iron primaries, where the value appears to be underestimated. A possible remedy is to apply a cut on the $X_{\max}$ distribution of the mock dataset to isolate its lower tail. By imposing a cut at 700 g/cm^2, we allow for a 5% contamination from other primaries. Repeating the maximum-likelihood estimation on this subsample yields a muon rescaling factor of 1.25 ± 0.05 for iron primaries, in much better agreement with the true value. A summary plot is shown in Figure 5.

5. Conclusion

We have introduced a new method to calculate mass-dependent muon rescaling factors in order to quantify the discrepancy between (mock) data and hadronic interaction models. By employing a maximum-likelihood estimation and incorporating mass-discriminating observables, as well as cuts on the $X_{\max}$ distribution, we demonstrated that it is possible to recover the true rescaling factors between EPOS LHC-R and QGSJETIII. A limitation of the present study is that the S_{μ} observable is not easily accessible in real hybrid observations of vertical air showers. Future work will therefore focus on identifying additional mass-sensitive observables, such as universality features, the rise time of individual detector signals, or machine-learning approaches applied to station traces. Furthermore, the method can be extended to estimate the Heitler–Matthews β coefficient of air showers [18] and to explore other energy intervals. Ultimately, the next step is

PoS(ICRC2025)173

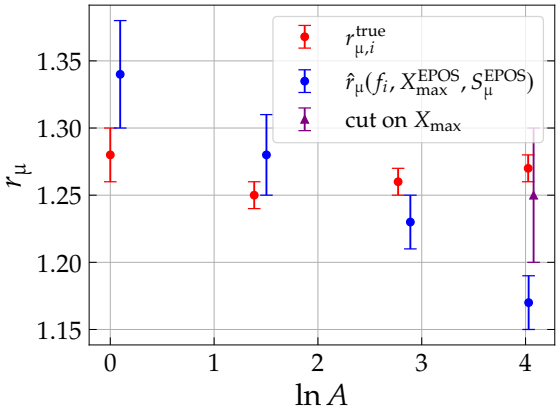

Figure 5: True (red) and computed (blue) muon rescaling factors for different primary masses. The purple data point corresponds to the muon rescaling factor of iron after applying a cut on the X_{max} distribution of the mock dataset at 700 g/cm^2, as described in the text.

to apply this *top-down* simulation method directly to real hybrid events, thereby testing its full potential to improve our understanding of the muon content in extensive air showers.

Acknowledgements

The authors are grateful for the support and feedback of the Pierre Auger Collaboration, which has made this work possible. N.B., D.G., J.P. and H.W. are supported by the Polish National Science Centre under grant 2020/39/B/ST9/01398 and 2022/45/B/ST9/02163, and the Ministry of Science and Higher Education, grant No. 2022/WK/12.

References

[1] Pierre Auger and Telescope Array Collaborations, *EPJ Web of Conferences* **283** (2023) 03002.
[2] Pierre Auger Collaboration, *Physics Review D* **110** (2024) 062005
[3] Telescope Array Collaboration, *MNRAS* **492(3)** (2020) 3984
[4] Pierre Auger Collaboration, *POS(ICRC2023)* **365** (2023).
[5] Telescope Array Collaboration, *Physics Review D* **110** (2024) 022006
[6] S. Ostapchenko, *Nuclear Physics B* **151(1)** (2006) 143
[7] F. Riehn, *et al.*, *Physics Review D* **102** (2019) 063002
[8] T. Pierog, *et al.*, *Physics Review C* **92** (2019) 034906
[9] H. Dembinski, *et al.*, *EPJ Web Conf.* **210** (2019) 02004
[10] D. Soldin, *et al.*, *POS(ICRC2021)* **349** (2021).
[11] Pierre Auger Collaboration, *Physics Review Letters* **117** (2016) 192001.
[12] K. Almeida Cheminant, *et al.*, *POS(UHECR2024)* **063** (2024).
[13] F. Riehn, A. Fedynitch, and R. Engel, *Astroparticle Physics* **160** (2024) 102964
[14] T. Bergmann, *et al.*, *Astroparticle Physics* **26(6)** (2007) 420
[15] T. Pierog and K. Werner, *POS(ICRC2025)* **358** (2025)
[16] D. Heck, *et al.*, *Report FZKA 6019* (1998)
[17] S. Ostapchenko, *Physics Review D* **109** (2024) 094019
[18] J. Matthews, *Astroparticle Physics* **22(7-8)** (2005) 387

PoS(ICRC2025)173