PROCEEDINGS OF SCIENCE

Measuring the Proton-Proton Interaction Cross Section with Hybrid Data of the Pierre Auger Observatory

Olena Tkachenko[a,*] **for the Pierre Auger Collaboration**[b]

[a] *Institute of Physics of the Czech Academy of Sciences, Na Slovance 1999/2, 182 00 Prague, Czech Republic*

[b] *Observatorio Pierre Auger, Av. San Martín Norte 304, 5613 Malargüe, Argentina*
Full author list: https://www.auger.org/archive/authors_2024_11.html

E-mail: spokespersons@auger.org

The depth of the maximum of an air shower development, X_{max}, as observed with fluorescence telescopes, is among the most sensitive observables for studying the interaction characteristics and primary composition of ultra-high-energy cosmic rays. However, precise measurement of the interaction cross section remains challenging, as standard analyses often rely on assumptions about the composition, which are closely tied to the validity of specific hadronic interaction models. In this work, we discuss a method for the simultaneous estimation of the proton-proton interaction cross section and primary mass composition, addressing the limitations of separate measurements. The inclusion of the X_{max} scale into the fit further accounts for systematic uncertainties in the data and theoretical uncertainties in particle production. The performance of the method is evaluated using simulations that include detector responses under realistic conditions and with a particular focus on assessing the systematic uncertainties of the fit.

7th International Symposium on Ultra High Energy Cosmic Rays (UHECR2024)
17-21 November 2024
Malargüe, Mendoza, Argentina

*Speaker

ISSN 1824-8039. Published by SISSA Medialab.
https://pos.sissa.it/

1. Introduction

The atmospheric depth at which the number of particles in an extensive air shower reaches its maximum, $X_{\max}$, is a fundamental observable for probing the cosmic ray mass composition and the properties of hadronic interactions during the formation of extensive air showers in the atmosphere [1]. The shower maximum contains information about the initiating nuclei because it scales with the logarithm of the primary particle mass. At the same time, $X_{\max}$ is influenced by the depth of the first atmospheric interaction, which is directly related to the interaction cross section of the primary particle with the air. At ultra-high energies, however, neither the mass composition nor the interaction cross sections are precisely measured, and their estimation often depends on assumptions about one to infer the other. The interpretation of cosmic ray data at energies well beyond the reach of the ground-based accelerators thus requires extrapolations from collider measurements, which are subject to considerable uncertainties and may not fully account for the entire phase space of interaction properties. Hadronic interaction models can be significantly improved by the insights gained from studying the interaction properties of ultra-high-energy cosmic rays, as these studies provide crucial information for refining the models [2]. Validating these models through air shower simulations often requires making assumptions about other properties of the cosmic rays. For example, cross section measurements for proton-air or proton-proton interactions, which are the focus of this study, are typically made under the assumption of a tail in the Xmax distribution dominated by a specific mass, such as protons. Any potential helium contamination is considered a systematic uncertainty, with the helium fraction typically constrained to 25% [3, 4]. At the same time, predictions of the mass composition are derived using these interaction models and thus depend on the accurate estimation of the interaction properties, including the cross sections. The simultaneous estimation of the mass composition and the proton-proton interaction cross section [5] addresses the drawbacks associated with separate analyses, leading to a more reliable determination of both quantities of interest. In this paper, we discuss the improvements to the method, evaluate the potential uncertainties, and assess the performance of the method using simulations that consider the physical processes in extensive air showers and the detector responses of the Pierre Auger Observatory.

2. Algorithm for Simultaneous Fitting

To simultaneously estimate the mass composition and the proton-proton interaction cross section, we fit the full $X_{\max}$ distribution using model predictions obtained from the simulations with a modified interaction cross sections. The dependence of the shape of the $X_{\max}$ distribution on the proton-proton interaction cross section is obtained by adopting the approach outlined in [5], with the rescaling factor f_{19}^{pp} defined similarly to the original proton-air cross section analysis from the tail of the $X_{\max}$ distribution [3]. The energy threshold for the cross section modifications is set to 10^{17} eV. The proton-proton cross section modifications have been implemented in the Sibyll 2.3d generator [6] within the CONEX [7] air shower program. Since it is a semi-superposition model, the same resampling algorithms can be applied to showers initiated by any nuclei [8], where the energy scaling is treated as E_0/A, with E_0 being the total energy of the projectile nucleus and A being the atomic mass number.

Typically, estimating the mass composition by fitting the data to templates constructed with hadronic interaction models assumes that the prediction of the $X_{\max}$ scale from these models is accurate. However, it is well known that current models have difficulty accurately predicting air shower observations. Moreover, the uncertainties in the estimated $X_{\max}$ scale are larger than about a third of the difference between the mean values for proton and iron nuclei [19, 20]. Therefore, in addition to fitting the mass composition and cross section, we extend the fit to account for potential shifts in the $X_{\max}$ scale relative to the value predicted by the Sibyll 2.3d model. As a result, the fit is performed over a wide range of discretely varying values of the rescaling factor, f_{19}^{pp}, and shifts in the $X_{\max}$ scale, $\delta X_{\max}$, resulting in a scan over these parameters along with the best-fit mass composition estimate for each pair of values. We assess the goodness of fit for each combination of $\delta X_{\max}$ and f_{19}^{pp} by calculating the deviance, defined by the logarithm of the Poissonian likelihood of the fit, which is approximately χ^2-distributed. Assuming that both the factor f_{19}^{pp} and the shift $\delta X_{\max}$ are constant across energies, we analyze all energies together by summing the corresponding χ^2 values to estimate the best-fit combinations of the fitted variables that evolve consistently across energies. By comparing the resulting χ^2 values, we identify the combination of f_{19}^{pp} and $\delta X_{\max}$ that provides the best-fit quality (lowest χ^2) and determine the corresponding mass composition. In this way, we obtain the best-fit estimate of the proton-proton cross section rescaling factor f_{19}^{pp}, the shift in the $X_{\max}$ scale, and the cosmic ray mass composition at ultra-high energies.

3. Sources of Systematic Uncertainty

In addition to the statistical uncertainties inherent to the fit, it is essential to evaluate the systematic uncertainties, which account for the potential impact of various factors on the results. While for individual mass composition measurements, the primary source of the systematic uncertainty is the $X_{\max}$ scale itself [9], and, as mentioned above, for cross section measurements, the main contribution comes from the presence of helium in the tail of the $X_{\max}$ distribution, in the simultaneous estimation the contribution from these uncertainties is significantly reduced and other factors become more relevant. These include the energy scale, detector effects, energy-dependent $X_{\max}$ systematics, the underlying mass composition, elasticity and multiplicity. Their respective contributions to the total systematic uncertainty are shown in Table 1.

The uncertainties associated with the detector were evaluated by generating 100 data realizations from the Monte Carlo simulations of the Pierre Auger Observatory [10], assuming the composition inferred from the experimental data [11]. The effects of uncertainties in the extrapolations of elasticity and multiplicity to the highest energies were evaluated by modifying the simulations with rescaling factors, following the same approach to that applied for the cross section modifications [13], and then re-fitting them with model predictions derived from simulations with modified cross sections. The energy threshold was set to 10^{14} eV, with the rescaling factors ranging from 0.6 to 1.5 for elasticity and from 0.6 to 1.7 for multiplicity [14]. The reported uncertainty corresponds to the values of the estimated parameters obtained for the largest deviations in both elasticity and multiplicity from the nominal Sibyll 2.3d predictions. These two factors together represent the most significant source of systematic uncertainty in the estimated proton-proton cross section, potentially varying it by up to 20%. To account for the uncertainty in the energy scale estimation, we adopt the 14% estimate from [15] and conduct the fit with the energy scale shifted up

Origin	Impact on σ_{pp}, %	Impact on δX_{max}, g/cm^2
Energy scale	3.1	(+6 / -4)
Detector effects	(+7 / -12)	±1
E-dependent X_{max} syst.	±2	±7
Composition	(+3 / -7)	+5
Elasticity	(+15 / -17)	(+1 / -3)
Multiplicity	+9	(+1 / -8)

Table 1: The sources of systematic uncertainty in the rescaling factor and the shift of the X_{max} scale in the simultaneous mass composition and proton-proton cross section estimation.

and down by this amount. The systematic uncertainty from the detector acceptance and resolution parameterizations was considered by separately combining the uncertainties for each parameter, leading to the extreme acceptance values and the upper and lower 1-σ bounds on the resolution. Lastly, we assess the impact of energy-dependent X_{max} systematics by generating a sequence of random shifts for each energy. From the fits to simulated data with these shifts being applied, we then get the distributions of the f_{19}^{pp}, δX_{max}, and mass composition. The uncertainties are determined from the standard deviations of these distributions.

4. Method performance

It has been shown previously [5] that for a mixture of H and He nuclei simultaneous estimation of mass composition and cross section improves the bias in the cross section compared to the standard measurement obtained from the tail of the X_{max} distribution, regardless of any He contamination in the simulations. To better understand the behavior of the fit in different composition cases, we vary the relative particle fractions in different two-component configurations, including cases where no H nuclei are present in the composition. The fit performance for different composition scenarios as a function of the number of events evaluated for a single energy bin is shown in Figure 1. It can be seen that changes in the composition have minimal effect on the overall fit performance, with different compositions resulting in the near-perfect recovery of the input values and nearly identical results for sufficiently large sample sizes (greater than 1000 events). For smaller sample sizes, with less than 200 events, some composition scenarios, particularly those with a lower hydrogen fraction, can lead to a bias in the rescaling factor of up to 20%. Since we estimate a single rescaling factor over the entire energy range, any potential bias from the limited number of events at the highest energies is mitigated by the more reliable estimates at lower energies.

As another test, we evaluate the performance of the method under the assumption that the data is described by a different interaction model, where the X_{max} scale and the X_{max} distribution shape deviating from those anticipated by the Sibyll 2.3d model. For this, we fit templates from Sibyll 2.3d simulations with a modified cross section to simulations generated with the QGSJET-II.4 [16] and Sibyll 2.1 [17] models, assuming the observed mass composition. We do not test our analysis with EPOS-LHC [18] simulations, as this model treats nuclear fragmentation differently, making

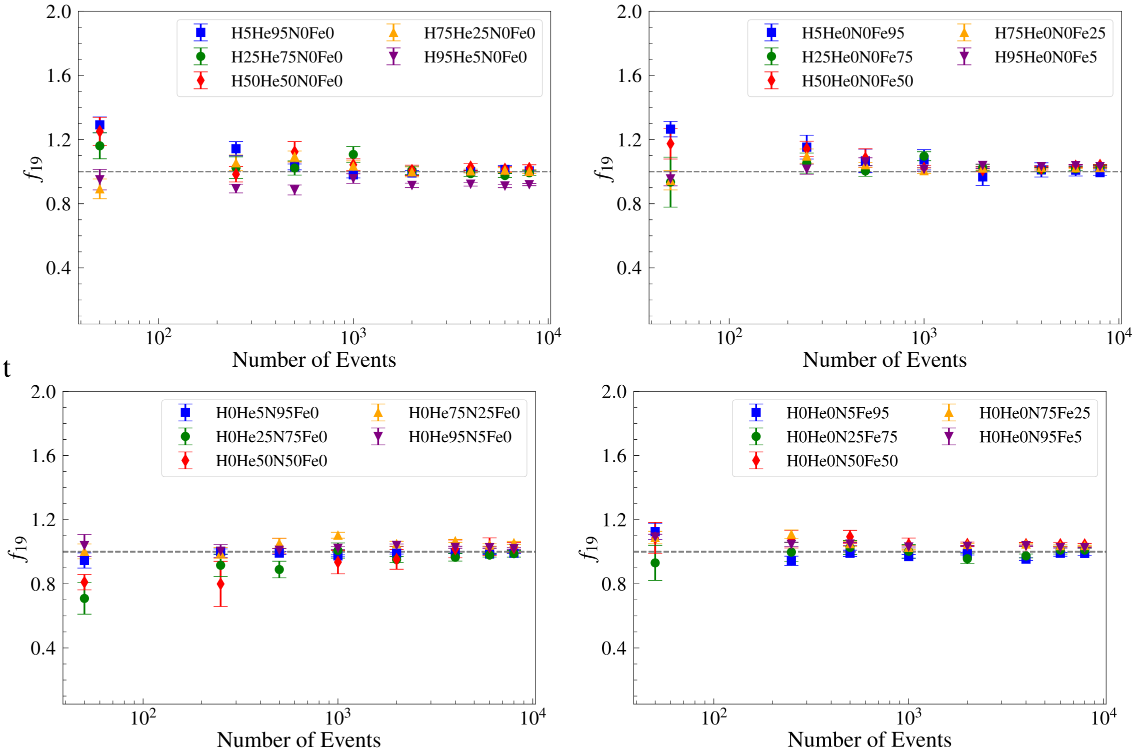

Figure 1: Estimated biases in the fitted rescaling factor as a function of the assumed composition and the number of events in the simulated data.

it challenging to accurately reproduce the $X_{\max}$ distribution shape predicted by EPOS-LHC, even with the modified Sibyll 2.3d model. The proton-proton cross section estimated from the fit to the QGSJET-II.4 and Sibyll 2.1 simulations is compared with the corresponding model values, along with the fitted compositions, in Fig. 2. The comparison also includes results from fits where the $X_{\max}$ scale was fixed at its nominal value. As can be seen, the cross section obtained from the fit to the QGSJET-II.4 model is very close to the actual QGSJET-II.4 value in the fit with unconstrained shift in the $X_{\max}$ scale, while fixing the $X_{\max}$ scale results in cross sections that are up to 10% larger than those in the simulated data at the highest energies. Since the Sibyll 2.1 model is not tuned to the latest LHC data, the energy evolution of the proton-proton cross section cannot be accurately reproduced. We are able to recover the cross section around 10^{19} eV, which is by definition the reference energy for the rescaling parameter, when the $X_{\max}$ scale is allowed to vary. In contrast, when the scale is fixed, the fitted cross section also significantly exceeds the data. The estimated composition also shows good agreement with the reference values, especially at the lowest energies considered, with the difference between the fitted and input fractions not exceeding 5%. Above the ankle, larger deviations are observed for nitrogen and helium. Similar to the cross section results, the consistency improves when the $X_{\max}$ scale is allowed to vary instead of being fixed to its nominal value. This shows that including the $X_{\max}$ scale shift as a fitting parameter improves the performance of the fit, allowing a more accurate determination of both the cross section and the mass composition, especially when there is a discrepancy between the $X_{\max}$ scale in the data and the one predicted by the Sibyll 2.3d interaction model. The estimated shifts in the $X_{\max}$ scale are

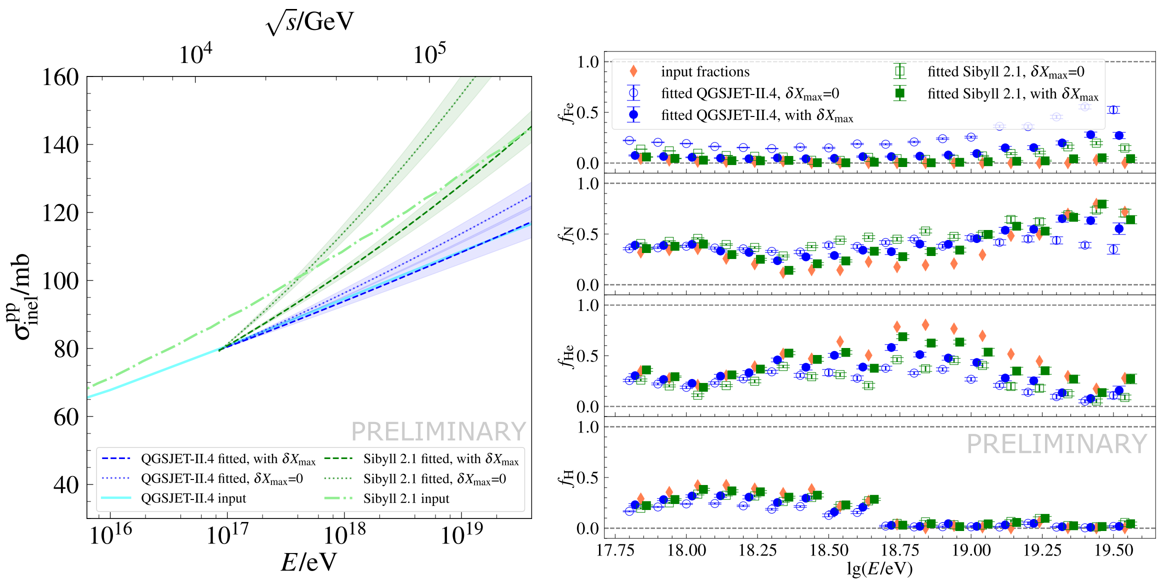

Figure 2: The results of the fit to simulations generated with the QGSJET-II.4 and Sibyll 2.1 interaction models, assuming a mass composition derived from Pierre Auger Observatory data. Left: Proton-proton interaction cross section, with the shaded region representing the statistical uncertainty. Right: Fitted mass composition. The results are shown for two cases: one where the $X_{\max}$ scale is free and one where it is fixed to the nominal value from Sibyll 2.3d.

approximately within ±5 g/cm² of the average differences between the predictions of Sibyll 2.3d and those of QGSJET-II.4 and Sibyll 2.1, respectively.

To assess the expected performance of the fit under scenario similar to the observed data, we applied the fit to Monte Carlo simulations that included the detector response and assumed the mass composition derived from the Pierre Auger Observatory data. The fit was performed on 100 sets of simulated data. We varied the rescaling factor f_{19}^{pp} between 0.2 and 2.0 in steps of 0.1, and the shift $\delta X_{\max}$ between -30 g/cm^2 and 15 g/cm^2 in steps of 1 g/cm^2.

The results of this scan for a single realization of the simulated data, indicating a well-defined minimum in the two-dimensional χ^2 distribution near the unmodified rescaling factor value are shown in Figure 3. The estimated shift in the $X_{\max}$ scale indicates a slight bias of -5 g/cm^2. In Figure 4, the separate evolutions of the cross section and mass composition with energy are demonstrated. The results are compared to the cross section derived from the fit to the tail of the $X_{\max}$ distribution, as well as to the composition-only fit, and include an estimate of the systematic uncertainties. As can be seen, both the fitted mass composition and the cross section closely match the input values, providing the same quality of performance as the standard analyses. Unlike the

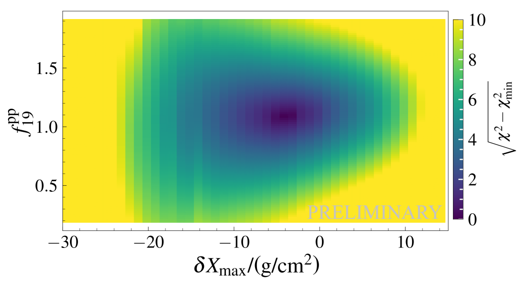

Figure 3: The χ^2 contour of the fit. The x-axis displays a scan over the shift in the $X_{\max}$ scale, while the y-axis covers a scan over the cross section rescaling factor. The color scale indicates the difference between the χ^2 value for each combination of parameters and the minimum χ^2 across the entire parameters space.

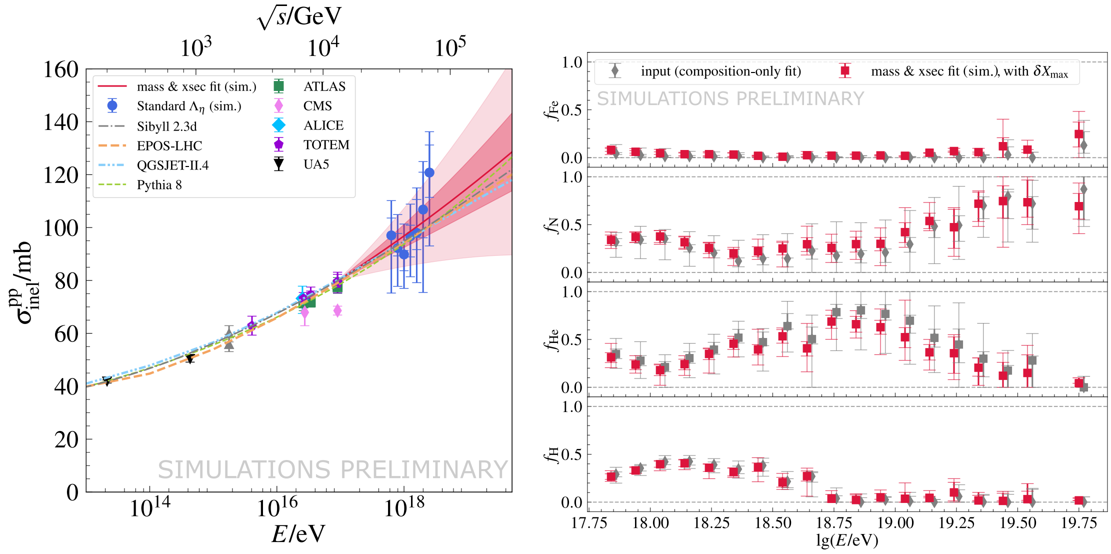

Figure 4: The results of the fit to simulations generated with the Sibyll 2.3d interaction model, assuming a mass composition derived from Pierre Auger Observatory data. Left: The proton-proton interaction cross section. For the simultaneous mass composition and cross section fit, the darker shaded region corresponds to the statistical uncertainty, while the lighter shaded region shows total uncertainty, including systematics. For the standard measurement from the tail of the $X_{\max}$ distribution, the inner cap indicates the statistical uncertainty, and the outer cap shows the total uncertainty, including systematics. Right: Fitted mass composition compared to input results from the composition-only fit with Sibyll 2.3d, with both statistical and total uncertainties given.

standard cross section measurement, which limits the fit to energies below $10^{18.4}$ eV to ensure that the relative helium fraction remains below 25%, the simultaneous estimation approach allows the fit to be extended to significantly higher energies. In addition, we observe a reduction in both the statistical and systematic uncertainties of the estimated parameters. This is because the primary source of uncertainty in the standard cross section analysis, the bias introduced by the presence of helium nuclei in the mass composition, is significantly reduced by the simultaneous fitting of the composition in the same analysis. Similarly, the uncertainty in the fitted composition is reduced by including the $X_{\max}$ scale shift in the fit, which is the dominant source of systematic uncertainty in separate analyses.

5. Conclusions

We explored the potential of simultaneously estimating the mass composition and proton-proton interaction cross section as a novel approach to measuring both quantities without depending on the underlying assumptions about either required in standard independent analyses. In addition, the $X_{\max}$ scale was allowed to vary to account for both systematic uncertainties in the data and theoretical uncertainties related to particle production in air showers. The performance of the proposed method was evaluated, and no significant bias in the measured variables was found, regardless of the assumed composition mix. Allowing the $X_{\max}$ to remain unconstrained improved performance, especially if we assume that its value in the data differs significantly from the predictions of the Sibyll 2.3d model. Additionally, the simultaneous estimation of the proton-proton interaction cross section, mass composition, and the $X_{\max}$ scale improves the uncertainty on the fitted variables. The total uncertainty on the proton-proton cross section, while smaller than in separate estimations, still accounts for a wide range of cross section modifications, including logarithmic energy evolution and nearly constant behavior across the considered energy range. Future work towards improving the method will focus on achieving better precision in assessing the uncertainties of the estimated variables and exploring possible changes in the functional form used for rescaling the interaction cross section. The method will then be applied to the $X_{\max}$ distributions as observed by the Fluorescence Detector of the Pierre Auger Observatory.

Acknowledgments

This work was supported by the Czech Academy of Sciences (LQ100102401) and the Ministry of Education, Youth and Sports of the Czech Republic (FORTE CZ.02.01.01/00/22_008/0004632, LM2023032).

References

[1] T. Fitoussi [for the Pierre Auger Collaboration], in these proceedings.

[2] J. Vicha [for the Pierre Auger Collaboration], in these proceedings.

[3] P. Abreu *et al.* [Pierre Auger Collaboration], *Phys. Rev. Lett.* **109** (2012) 062002.

[4] R. Ulrich [for the Pierre Auger Collaboration], *PoS ICRC2015* (2015) 401.

[5] O. Tkachenko *et al.*, *PoS ICRC2021* (2021) 438.

[6] F. Riehn *et al.*, *Phys. Rev. D* **102** (2020) 063002.

[7] T. Bergmann *et al.*, *Astropart. Phys.* **26** (2007) 420.

[8] R. J. Glauber, *Phys. Rev.* **100** (1955) 242.

[9] A. Aab *et al.* [Pierre Auger Collaboration], *Phys. Rev. D* **90** (2014) 122006.

[10] E. Santos [for the Pierre Auger Collaboration], *PoS ICRC2021* (2021) 232.

[11] O. Tkachenko [for the Pierre Auger Collaboration], *PoS ICRC2023* (2023) 438.

[12] A. Aab *et al.* [Pierre Auger Collaboration], *Phys. Rev. D* **90** (2014) 122005.

[13] R. Ulrich, R. Engel, and M. Unger, *Phys. Rev. D* **83** (2011) 054026.

[14] J. Ebr *et al.*, *Ukr. J. Phys.* **11** (2024) 786.

[15] V. Verzi [for the Pierre Auger Collaboration], *Proc. 33th ICRC 2013* [arXiv:1307.5059].

[16] S. Ostapchenko, *Phys. Rev. D* **83** (2011) 27.

[17] E.-J. Ahn *et al.*, *Phys. Rev. D* **80** (2009) 17.

[18] T. Pierog *et al.*, *Phys. Rev. C* **92** (2015) 15.

[19] A. Aab *et al.* [Pierre Auger Collaboration], *Phys. Rev. D* **109** (2024) 102001.

[20] R. Abbasi and G. Thomson, *JPS Conf. Proc.* **19** (2018) 011015.