PROCEEDINGS OF SCIENCE

Imprints of extragalactic propagation in the UHECR mass composition at the top of the atmosphere

C. Trimarelli,[a,b,c,*] **D. Boncioli,**[b,c] **S. Petrera**[d,c] **and F. Salamida**[b,c]

[a] *Département de Physique Nucléaire et Corpusculaire, Faculté de Sciences, Université de Genève Genève, Switzerland*

[b] *Università degli Studi dell'Aquila, Dipartimento di Scienze Fisiche e Chimiche, via Vetoio, L'Aquila, Italy*

[c] *INFN Laboratori Nazionali del Gran Sasso,Assergi (L'Aquila), Italy*

[d] *Gran Sasso Science Institute, L'Aquila, Italy*

E-mail: caterina.trimarelli@aquila.infn.it

The nuclear species responsible for the flux of the ultra-high-energy cosmic rays (UHECRs) at Earth can be retrieved from the distributions of the depth in the atmosphere at which the maximum number of particles in the extensive air showers is reached, i.e. $X_{\rm max}$. This is done by fitting model predictions of four or five mass groups (p, He, N, Si and Fe) to the measured $X_{\rm max}$ distributions. The derived mass-fractions-to-energy curves show that different nuclear species dominate different energy ranges. In this contribution, we investigate this finding by assuming a parametric model for each elemental spectrum and fitting, at the same time, the energy spectrum and the $X_{\rm max}$ distributions measured by the Pierre Auger Observatory without taking into account the extragalactic propagation in the fit procedure. We find that the peaks of the fractions-to-energy curves at Earth above $10^{17.8}$ eV exhibit a Lorentz-factor dependence that appears to be mainly driven by the UHECR spectral parameters describing the energy range above the "ankle". Despite the low maximum rigidity found in current astrophysical scenarios interpreting the UHECR data at the highest energies, our work confirms the relevance of the photo-hadronic interactions in shaping the observed cosmic-ray flux at Earth.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

ICRC2023
38th International Cosmic Ray Conference
The Astroparticle Physics Conference

*Speaker

PoS(ICRC2023)316

1. Introduction

The understanding of the composition of ultra-high-energy cosmic rays (UHECRs) constitutes a key point to progress in several aspects of astroparticle and high-energy physics. Different methods can be considered to investigate the nature of the primary cosmic ray interacting in the atmosphere [1]. Currently, the most reliable technique is the measurement of the depth at which the number of particles in air showers reaches its maximum, $X_{\rm max}$. This quantity can be observed with non-imaging Cherenkov detector, radio arrays and fluorescence telescopes. At the present time, only fluorescence detectors (FD) have reached enough precision to directly measure $X_{\rm max}$ at ultra-high energies. The measurements of the first two moments of the $X_{\rm max}$ distribution show that the composition of cosmic rays becomes lighter as the energy increases towards the ankle (the change of slope of the UHECR spectrum measured at about $10^{18.7}$ eV) and then becomes heavier again towards the highest energies [2].

To obtain a more quantitative estimate of the mass composition of UHECRs, templates of four or five mass groups (p, He, N, Si, Fe) to the $X_{\rm max}$ can be fitted. The best fit parameters will give the resulting set of mass fraction in each energy bin. In [3], a quantitative estimate of the mass composition of UHECRs is reported by fitting the measurements of the $X_{\rm max}$ distributions in each energy bin with Monte Carlo templates provided for a variety of hadronic interaction models. On the other hand, a similar procedure to retrieve the mass fractions can be done by fitting the measured distributions with well determined parametric model distributions, as reported in [4]. In the present work the fraction fit (FF) procedure [5] has been applied to the $X_{\rm max}$ distributions presented in [6]. The FF program is employed to find the best mixture of elemental model $X_{\rm max}$ distributions (p, He, N, Si, Fe) to the measured distribution, *at each log-energy bin*[1]. Therefore, the fraction set found at each energy is determined independently from the other ones. This fact reflects itself in a energy dependence not possessing any *continuity* requirement as expected from physics. In particular the unitarity condition, combined to the energy bin independence of the fit, can bring to some discontinuities at several energy bins in the mass-fraction-to-energy behavior, due for instance to the relative shifts of the elements because of the correlation of the fit parameters (independent mass fractions or cosines). The fit results indicate (open circles in Fig. 1) that the fractions of the nuclear species first increase and then decrease as a function of the energy, showing an ordering in terms of increasing charge or mass at Earth.

A physical continuity can be expected by exploiting the measurements of the energy spectrum at Earth, as done in [7], hereinafter called *GST* fit. In this work, as well as in the more recent analysis shown in [8], they use the magnetic rigidity as the relevant variable to interpret the data. This is justified since we are dealing with charged particles in environments where magnetic fields are present. Therefore, in the site where the particles are accelerated, considering a maximum energy that protons can achieve, if several nuclear species are present we might expect that protons will cutoff first, followed by other species (according to the Peters cycle [9]). On the other hand, one might expect the data to be sensitive to the effects of the cosmic-ray interactions. Cosmic-ray particles experience interactions with background photons, both in the source environment and in the extragalactic propagation, suffering energy and mass losses. Particularly important in the

[1]The fit results are given by Minuit package [11] from the maximum likelihood estimation method involving all the best fit parameters and the statistical information.

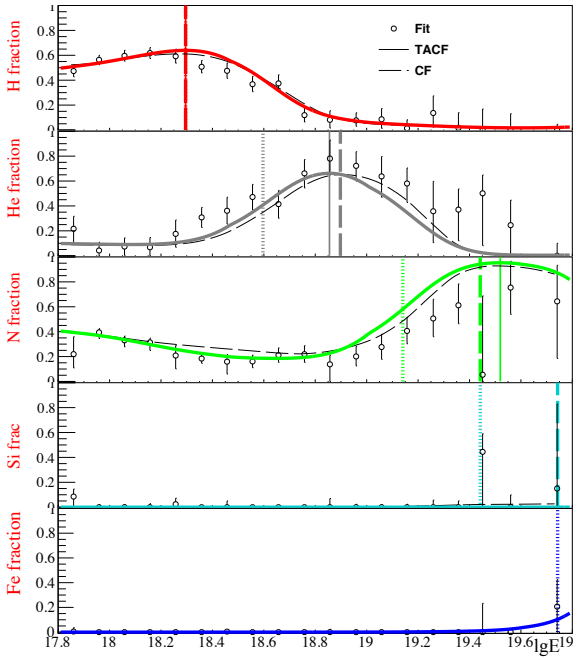

Figure 1: Outcome of the FF (open circles) with five mass groups masses (H, He, N, Si, Fe), showing the fitted fractions with their statistical errors, for the data set shown at [6], using EPOS-LHC model [12]; full (dashed) curves refer to TACF (CF) results. The energy values corresponding to the maximum of the FF curve are shown as vertical full lines. The energy values corresponding to A times (Z times) the energy value for the maximum of the proton fraction are shown as vertical dashed (dotted) lines. Colors correspond to the used nuclear species (H: red; He: grey; N: green; Si: cyan; Fe: blue).

energy range of interest for this study, nuclear photo-disintegration processes can be triggered. In such processes the Lorentz factor of the nucleus is conserved.

In this work, we pursue a combined fit of the energy spectrum and mass composition data at Earth in order to provide a continuous representation of the mass fraction at Earth as a function of the energy. We use the measurement of the UHECR energy spectrum obtained from events detected using the surface detectors (SD) of the Pierre Auger Observatory as published in [10], and the measurement of the $X_{\rm max}$ distributions as reported in [6]. Both the dependence on the rigidity and on the Lorentz factor, supported by the arguments reported above, will be tested.

2. The Top-of-Atmosphere Combined Fit

We assume a parametric model for each elemental spectrum *at the top of the atmosphere*, and fit the spectrum and mass composition, denoting this fit as TACF. To this aim, the fit procedure is the same as that used in the combined fit of spectrum and composition [13, 14] (hereafter referred to as CF), without the extragalactic propagation of the emitted cosmic rays.

Following [14], to fit the cosmic ray-spectrum in the the entire energy range (below and above the ankle) we assume two sets of parametric functions:

$$J_A^i(E) = J_{0A}^i \cdot \left(\frac{E}{E_0}\right)^{-\gamma^i} \cdot f_{\rm cut}(E, E_{\rm cut}^i(A)) \,, \tag{1}$$

	EPOS-LHC						Sibyll2.3d	
	(A1) broken exp		(A2) exp		(A3) sech		(A1) broken exp	
	LE	HE	LE	HE	LE	HE	LE	HE
γ	3.85	0.067	3.85	-0.11	3.85	0.012	3.72	0.52
$E_{\rm cut}(p)$	19.88	17.87	19.50	17.88	21.90	17.89	19.77	17.85
$D_{X_{\rm max}}/n_{\rm pts}$	542.2/329		539.4/329		539.9/329		619.1/329	
$D_J/n_{\rm pts}$	91.00/24		116.0/24		124.5/24		81.3/24	
$D_{\rm tot}/n_{\rm pts}$	633.2/353		655.4/353		664.4/353		700.4/353	

Table 1: Best fit parameters and deviances of top-of-atmosphere combined fit for three different A-dependent cutoff shapes: broken exponential, exponential and hyperbolic secant.

where $i = \text{LE}$ ($i = \text{HE}$) refers to the low-energy (high-energy) component and $J(E) = \sum_i \sum_A J^i_A(E) = J^{\rm LE}(E) + J^{\rm HE}(E)$ is the total spectrum. In Eq. (1), A has the values 1, 4, 14, 28, 56 and J^i_{0A} is the elemental spectrum for component i at fixed energy $E_0 = 10^{18}$ eV, $f_{\rm cut}$ is the cutoff function and $E^i_{\rm cut}(A)$ is the cutoff energy of component i for mass A.

The following models for the cutoff functions of the energy spectrum are taken into account in this analysis:

- A-dependent cutoff, with three cutoff shapes: broken exponential (**A1**), exponential (**A2**) and hyperbolic secant (**A3**)[2];
- same as previous models, with Z-dependent cutoff: broken exponential (**Z1**), exponential (**Z2**) and hyperbolic secant (**Z3**).

The goodness-of-fit is assessed with a generalized χ^2, (the deviance, D), defined as the negative log-likelihood ratio of a given model and the saturated model that perfectly describes the data. The total deviance consists on the sum of two terms defined as follows:

$$D_J = \sum_i \frac{(J^i_{\rm obs} - J^i_{\rm mod})^2}{\sigma_i^2}, \quad D_{X_{\rm max}} = 2 \cdot \sum_{ij} k^{i,j}_{\rm obs} \cdot \ln \frac{k^{i,j}_{\rm obs}}{n^i_{\rm obs} \cdot G^{i,j}_{\rm mod}} \tag{2}$$

D_J is related to the energy spectrum. Its likelihood is treated as the product of Gaussian distributions, where in each i-th energy bin $J^i_{\rm obs}$ is the observed flux, σ_i is its statistical uncertainty, and $J^i_{\rm mod}$ is the model prediction. $D_{X_{\rm max}}$ is a product multinomial distributions describing the likelihood for the $X_{\rm max}$ distributions where $k^{i,j}_{\rm obs}$ is the number of observed events in the i-th energy bin and in the j-th $X_{\rm max}$ bin, $n^i_{\rm obs}$ is the total number of observed events in the i-th energy bin and $G^{i,j}_{\rm mod}$ are the model predictions. Here we adopt a parametric model for the $X_{\rm max}$ distributions, which takes the form of a generalized Gumbel [15, 16] distribution, whose parameters are given in [14]. The binned model predictions $G^{i,j}_{\rm mod}$ are obtained from the relevant Gumbel distributions after correcting for the detector effects. The best fit parameter values are those with which the total deviance $D_{\rm tot} = D_J + D_{X_{\rm max}}$ reaches its minimum value using the Minuit package.

In Table 1, the best fit parameters and deviances for the A-dependent models are shown. The fitted spectrum and composition for the **A1** case are shown in Figure 2. On the top left side the

[2]The corresponding parametric functions are shown in [14]

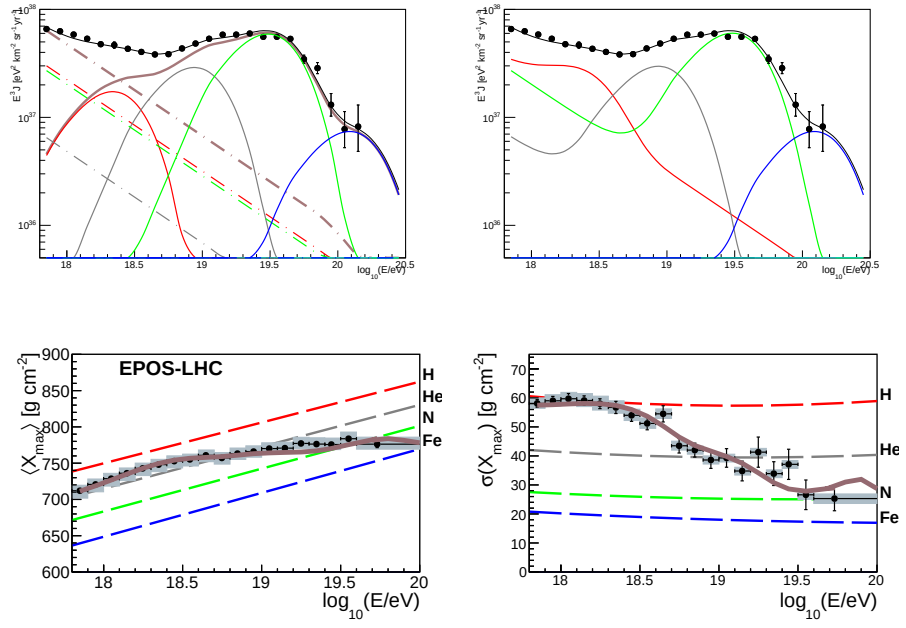

Figure 2: Energy spectra and composition at the top-of-atmosphere for the A-dependent broken exponential cutoff case, using EPOS-LHC [12]. The data from the Pierre Auger Observatory for the energy spectrum [10] and mass composition [6] are plotted as black circles. Top left side: The dash-dotted lines represent the low-energy component, while the solid lines represent the high-energy component. The brown lines represent the sum of each independent component, the black line is the sum of the two components. The colored lines show the elemental spectra (color codes as in Fig. 1). Top right side: The all-particle best fit function and the partial spectra for the different mass groups. Bottom: Average and standard deviation of the $X_{\rm max}$ distribution at each energy, as measured by Auger (black points) with systematic uncertainties (grey bands) for various pure compositions (dashed lines) and for the composition predicted by the best fit of the model (thick brown line).

measured energy spectrum is plotted along with the separated best fit functions; the partial spectra correspond to the contributions of the masses as defined in the caption. On the top right side we report the sum of the two components. On the bottom side we show the two first moments of the $X_{\rm max}$ distributions.

The following considerations can be done, independently of the cutoff function:

- the HE component consists of an observed spectrum with $\gamma \simeq 0$ and a low-energy cutoff;
- the LE component presents a soft energy spectrum, with a spectral index $\gamma = 3.85$. The proton maximum energy is large ($\log_{10}(E/\text{eV}) \geq 19.5$), so that this component penetrates into the HE region, but is subdominant above the ankle;
- the total reduced deviance is about 2.

In Table 1 a comparison between the fit obtained with EPOS-LHC and Sibyll2.3d [17] is also shown. The best fit parameters are mostly unchanged except for the spectral index above the ankle. At the LE the TACF result is mainly due to the fit of the mass fractions because of the absence of particular features in the measured spectrum. Therefore, the differences in the best fit parameters corresponding to the use of EPOS-LHC or Sibyll2.3d depend on the best combination

of nuclear species found to describe the mean and the width of the measured $X_{\rm max}$ distribution. At the HE the best fit obtained with Sibyll2.3d is characterized by a softer spectrum with respect to the case of EPOS-LHC. This can be explained by considering the fact that Sibyll2.3d predicts larger $\sigma(X_{\rm max})$ [18] that allows a larger superposition of elemental spectra (which can be obtained with soft spectra and/or high-energy cutoff).

Using Z-dependent cutoffs produces worse fits, with total deviances $D_{\rm tot}/n_{\rm pts} > 800/353$ for the three different cutoff shapes (broken exponential, exponential, hyperbolic secant). Both the A- and Z-dependent models, fitted using four mass groups below and above the ankle, gave similar results.

Once elemental spectra are obtained from the fit, the elemental fractions are

$$f_A(E|{\rm par}) = \frac{J_A(E|{\rm par})}{\sum_A J_A(E|{\rm par})}$$

where 'par' are the best fit parameters. Figure 1 shows the TACF results (solid line) superimposed to the points and errors obtained from the FF (open circles). The dashed line in Figure 1 refers to the CF results and will be discussed in the next section. The energy at which the fraction of hydrogen reaches its maximum $\log(E(f^H_{\rm max}))$ is found at $\log(E/{\rm eV}) = 18.3$. Vertical dashed (dotted) lines show the corresponding values of energy at which $E = A \cdot E(f^H_{\rm max})$ ($E = Z \cdot E(f^H_{\rm max})$).

The TACF lines reproduce the general behaviour of the FF data points. The initial trend is closer to fraction data points, whereas at higher energies helium (nitrogen) fraction points systematically exceed (underlie) the TACF line. To evaluate the level of agreement between the model lines and the fraction data points, we build an uncorrelated $\chi^2_{\rm unc}$ as

$$\chi^2_{\rm unc} = \sum_i \chi^2_i = \sum_i \sum_A \frac{(f^{\rm FF}_A - f^{\rm TACF}_A)^2_i}{\sigma^2_i(f^{\rm FF}_A)} \tag{3}$$

where $A = 1, 4, 14, 28, 56$ and i corresponds to the log-energy bin. The value found for $\chi^2_{\rm unc}$ is 65.2 for 65 degrees of freedom.

The differences between fractions obtained from FF and TACF, that in some case can be of the order of 10% or more, are less apparent considering that the $X_{\rm max}$ distribution are actually fitted and there is a certain degeneracy with respect to the actual values of the fractions. As an example, Fig. 3 shows the model $X_{\rm max}$ distributions (best fits) for the FF (solid lines) and for the TACF (dot-dashed lines) for the log-energy bin centered at 18.35. Even if the fraction sets are different, their sum results in total distributions very close to each other. This is the consequence of the correlations between fraction pairs that in some cases are sizeable. On the other hand the shape of the $X_{\rm max}$ distribution are primarily characterized by the first two $X_{\rm max}$ moments and, in turn, on $\langle \ln A \rangle$ and $\sigma_{\ln A}$. Therefore, what really matters is these two cumulative variables and not the single fractions. In Figures 4, the first two moments (the mean and the variance) of the log-mass distribution (obtained as described in [18]) are shown for the TACF A-dependent model (black curve) and Auger data (black full circles). It is visible that the model reproduces the data for the entire energy range standing within the statistical error bars.

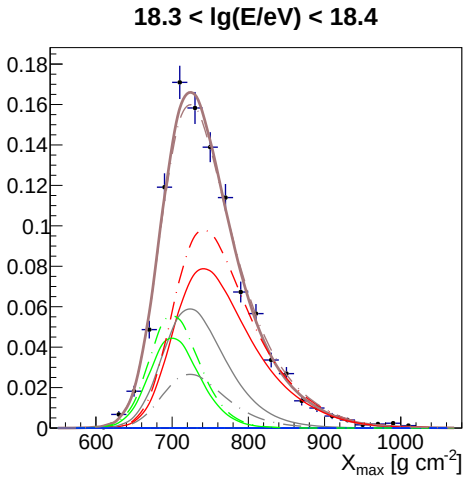

Figure 3: Model $X_{\rm max}$ distributions for the FF (solid lines) and for the TACF (dot-dashed lines) for the log-energy bin centered at $\log(E/{\rm eV}) = 18.35$. Crosses refer to data distributions.

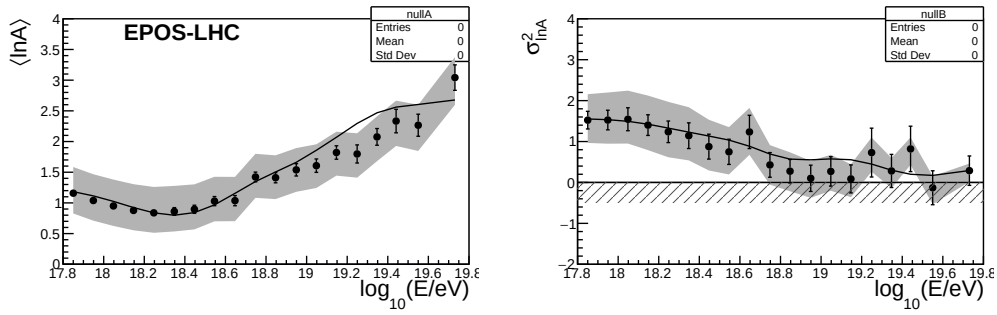

Figure 4: $\langle \ln A \rangle$ as a function of $\log_{10}(E/{\rm eV})$ obtained from Auger data are shown as full circles for EPOS-LHC. The shaded areas refer to systematic uncertainties obtained by summing in quadrature the systematic uncertainties on $\langle X_{\rm max} \rangle$ and $\sigma(X_{\rm max})$ data points and on the FD energy scale. The black curve corresponds to the TACF with the A-dependent broken exponential cutoff.

3. Discussion and astrophysical implications

In this contribution, the fit of the mass fractions as a function of the energy is discussed. The fraction set found as an outcome of the FF at each energy bin is independent from the other ones; on the contrary, a bin-to-bin relation should naturally exist. To explore this, we assumed a parametric model for each elemental spectrum at the top of the atmosphere and fitted the energy spectrum and the composition, excluding all propagation effects.

We tested both the Z-dependent and the A-dependent ordering in the cutoff, finding that the A-dependent case is best fitting the energy spectrum and composition at the top of the atmosphere, independently of the shape of the cutoff. This is reflected in the behavior of the mass fractions, as reported in figure 1. As soon as the energy corresponding to the maximum value of the proton fraction is identified, the corresponding energy multiplied by the atomic mass or charge can be computed. The maxima of the mass fractions for species heavier than hydrogen are found to

be close to A times the energy of the maximum for the hydrogen species. This finding can be interpreted as the imprint of the interactions in the extragalactic space, as well as in the environment of the sources, where the photo-disintegration plays a relevant role.

We observe that, while the power-law spectrum of the LE component in this analysis is similar to what found in the CF (which includes propagation) in the same energy range, the power-law of the HE component is softer than what found in the HE component of the CF (see Table 1 in [14]). This shows that the effect of the interactions is almost absent in the LE component while it shapes the spectrum and composition measured at Earth at HE. The interaction processes involving nuclei are in fact not efficient in the energy range below the ankle, where the adiabatic expansion of the universe becomes the most relevant source of energy losses but does not have effect in changing the shape of the spectrum. Similarly to what found in the CF, the heaviest mass of the LE component is the nitrogen. It is interesting to notice that the position of the maximum value of the functional form of the fractions is mainly determined by the peaked shape of the high-energy spectra. This result is also found when using a different hadronic interaction model, Sibyll2.3d.

We also notice that including the spectrum in the fit procedure allows to extrapolate to slightly higher energies with respect to considering only mass composition data. The Fe fraction found to be rising at the highest energies in the TACF is therefore only required by the fit of the high-energy spectrum at Earth. Larger statistics at these energies or the use of SD mass composition data will allow to better constrain the mass fractions at the highest energies and also to possibly confirm the ordering of the mass species up to Fe.

References

[1] K. H. Kampert and M. Unger, *Astropart. Phys.* **35** (2012) 660-678, [1201.0018].
[2] A. Aab *et al.* [Pierre Auger], *Phys. Rev. D***90** (2014) no.12, 122005, [1409.4809].
[3] A. Aab *et al.* [Pierre Auger], *Phys. Rev. D***90** (2014) no.12, 122006, [1409.5083].
[4] J. Bellido [Pierre Auger], PoS **ICRC2017** (2018), 506 doi:10.22323/1.301.0506
[5] C. Trimarelli, PhD Thesis, Unversity of L'Aquila (Italy), January 2023
[6] A. Aab *et al.* [Pierre Auger], [arXiv:1909.09073 [astro-ph.HE]].
[7] T. K. Gaisser, T. Stanev and S. Tilav, *Front. Phys. (Beijing)***8** (2013) 748-758, [1303.3565].
[8] H. P. Dembinski, R. Engel, A. Fedynitch, T. Gaisser, F. Riehn and T. Stanev, PoS **ICRC2017** (2018), 533 doi:10.22323/1.301.0533 [arXiv:1711.11432 [astro-ph.HE]].
[9] B. Peters, Il Nuovo Cimento, doi:10.1007/BF02783106.
[10] P. Abreu *et al.* [Pierre Auger], *Eur. Phys. J. C* **81** (2021) n.11, 966, [2109.13400].
[11] F. James and M. Winkler, MINUIT: Function Minimization and Error Analysis Reference Manual. In: (1998)
[12] T. Pierog, I. Karpenko, J. M. Katzy, E. Yatsenko and K. Werner, *Phys. Rev. C* **92** (2015) 034906, [1306.0121].
[13] A. Aab *et al.* [Pierre Auger], *JCAP***04** (2017) 038, [1612.07155].
[14] A. A. Halim *et al.* [Pierre Auger], *JCAP***05** (2023) 024, [2211.02857].
[15] E. Bertin, *Physical Review Letters* **95** (2005) no.17, 122005
[16] M. De Domenico, M. Settimo, S. Riggi and E. Bertin, *JCAP* **07** (2013), 050 [1305.2331].
[17] F. Riehn, R. Engel, A. Fedynitch, T. K. Gaisser and T. Stanev, *Phys. Rev. D***102** (2020) no.6, 063002,[1912.03300].
[18] P. Abreu *et al.* [Pierre Auger], *JCAP***02** (2013) 026, [1301.6637].