PROCEEDINGS OF SCIENCE

Parametrization of X_{max} distributions in the ultra-high energy regime

Luan Arbeletche*
Instituto de Física de São Carlos, Universidade de São Paulo, Brasil
E-mail: luan.arbeletche@ifsc.usp.br

Vitor de Souza
Instituto de Física de São Carlos, Universidade de São Paulo, Brasil
E-mail: vitor@ifsc.usp.br

Distributions of depth of shower maximum (X_{max}) are parametrized and results are compared to previous studies. Large samples of extensive air showers were simulated using the CONEX simulation package in the energy range 10^{17-20} eV for four primary masses: proton, carbon silicon and iron. Three functions are studied to parametrize the distributions of X_{max}: exponentially modified Gaussian (EMG), generalized Gumbel (GMB) and log-normal (LOG). Results allow for a direct comparison between the proposed functional forms in terms of the Akaike information theory and this comparison suggests that GMB function should be used in the description of X_{max} data. LOG distribution also provides reasonable fits for low-mass primaries while the EMG function should be discarded. A parametrization of each function for each used hadronic interaction model is discussed and a comparison to previous parametrizations is presented.

36th International Cosmic Ray Conference -ICRC2019-
July 24th - August 1st, 2019
Madison, WI, U.S.A.

*Speaker.

PoS(ICRC2019)174

1. Introduction

Determination of the mass composition of ultra-high energy cosmic rays (UHECRs) is of paramount importance in the current astrophysics scenario. Constraints on models for cosmic ray acceleration and propagation rely on the knowledge of the mass distribution of these particles. Typically, experiments such as the Pierre Auger Observatory and the Telescope Array Observatory infer the composition of primary particles by the indirect measurement of extensive air showers' observables. The standard variable used in such analyses is the atmospheric depth at which a shower reaches its maximum number of charged particles ($X_{\max}$) in terms of its first and second moments [1, 2, 3]. Recent developments in detection techniques, however, have allowed the Pierre Auger observatory to estimate the relative abundances of primary particles by directly fitting observed $X_{\max}$ data with expected distributions [4, 5]. A combined analysis of the UHECR flux and their composition in view of Auger $X_{\max}$ data also allowed for a study of UHECR sources [6]. Novel techniques for mass estimation [7, 8] and characterization of UHECR sources [9] also depend on the understanding of the $X_{\max}$ distributions.

Although the physics of extensive air showers is well understood in terms of particle interaction processes, the complexity of such stochastic systems and the lack of a description of the bulk of particle interaction processes do not allow for an analytical description of the so-called fluctuation problem [10]. As a consequence, distributions of important variables such as $X_{\max}$ have no known functional form. This leads to the current experimental scenario where shower variables must be interpreted by comparison to prescriptions of Monte Carlo simulations of the particle cascades. Thus, an approach to the description of the fluctuations of shower variables from a theoretical point of view is only possible by the parametrization of simulated quantities.

This contribution summarizes the results on a recent parametrization of $X_{\max}$ distributions for showers with energies between 10^{17} eV and 10^{20} eV [11] using three distinct functional forms. The simulation procedure and the proposed functions are described in section 2. Section 3 presents the fit results and compare the proposed functions. In section 4 a parametrization of $X_{\max}$ distributions in terms of primary energy and mass is discussed. Finally, section 5 provides a summary of this contribution.

2. Description of $X_{\max}$ distributions

Simulations of large samples of extensive air showers were performed within the hybrid approach of the CONEX package [12]. The set of primary particles includes proton, carbon, and silicon with energies spanning the interval 10^{17} eV to 10^{20} eV in steps of 1 in $\log_{10}(E_0/\mathrm{eV})$. Uncertainties in the modelling of hadronic interactions are taken into account by using three post-LHC event generators, namely EPOS-LHC [13], SIBYLL2.3C [14] and QGSJETII.04 [15]. The number of shower simulated for each primary, energy, and model combination is 10^6. Profiles are sampled in steps of 10 g/cm^2 and $X_{\max}$ is extracted from each profile by means of a fit to second-degree polynomial around the point of maximum number of charged particles. Anomalous profiles [16] are excluded from this analysis as for those the depth of maximum cannot be defined without ambiguity.

Three distinct functions are proposed to fit the simulated $X_{\max}$ distributions. The first studied functional form, already proposed in [17], stems from the assumption that a primary particle enters the atmosphere and interacts at a depth X_{first}, whose distribution follows an exponential law with decay parameter λ, and the distance in g/cm^2 from X_{first} to $X_{\max}$ follows a normal distribution with mean μ and variance σ^2. The resulting function is an exponentially modified Gaussian distribution (EMG), that reads

$$f_{EMG}(x) = \frac{1}{2\lambda}\exp\left(-\frac{x-\mu}{\lambda}+\frac{\sigma^2}{2\lambda^2}\right)\mathrm{erfc}\left(\frac{\mu-x+\sigma^2/\lambda}{\sqrt{2}\sigma}\right), \tag{2.1}$$

where $\mathrm{erfc}(x)$ is the complementary error function.

The second function is the generalized Gumbel distribution (GMB), proposed in [18] to describe $X_{\max}$ distributions. It follows from the consideration that $X_{\max}$ has properties that relate to the statistics of extremes, the field in which the GMB distribution arise. In fact, the GMB appears as the correct distribution of the asymptotic sum of exponential variables with increasing amplitudes [19], providing an interpretation of $X_{\max}$ as a sum of multiple interaction lengths. The GMB distribution is written as

$$f_{GMB}(x) = \frac{1}{\sigma}\frac{\lambda^\lambda}{\Gamma(\lambda)}\exp\left\{-\lambda\left[\frac{x-\mu}{\sigma}+\exp\left(-\frac{x-\mu}{\sigma}\right)\right]\right\}. \tag{2.2}$$

Finally, a third function is proposed: the shifted log-normal distribution (LOG). Although no appealing direct relation of the LOG distribution to shower physics is possible, it is shown that it provides a reasonable fit to $X_{\max}$ simulated data.

$$f_{LOG}(x) = \begin{cases} 0, & \text{if } x \le m \\ \frac{1}{\sqrt{2\pi}\sigma}\frac{1}{x-m}\exp\left\{-\frac{[\ln(x-m)-\mu]^2}{2\sigma^2}\right\}, & \text{if } x > m. \end{cases} \tag{2.3}$$

Each of the three presented functions have three free parameters that are adjusted to fit the simulated $X_{\max}$ distributions for each combination of primary energy, mass and hadronic interaction model. An unbinned likelihood scheme available in the ROOT framework [21, 20] is employed to obtain the parameters that provide the best fit to data.

3. Results

In Fig. 1 some examples of fitted distributions are shown for the case of SIBYLL2.3C at 10^{20} eV. For the purpose of illustrating the simulated distributions, $X_{\max}$ data from CONEX were binned in intervals of 10 g/cm^2 and those are shown as black dots with the corresponding statistical uncertainty as bars. The small plots below each frame represent the deviations (pull) of each function with respect to the data points. Primaries are indicated in the top-right corner of each plot. This figure represents well the fact that the EMG distribution is not able to adjust itself to prescribed $X_{\max}$ distributions as it overestimates the number of events for both small and large $X_{\max}$ values. GMB and LOG distributions, on the other hand, are competitive in the case of SIBYLL2.3C at 10^{20} eV. Both these two distributions fit $X_{\max}$ data reasonably well.

A statistical method is necessary in order to select the best among the set of proposed models to describe $X_{\max}$ data. In the present case of unbinned likelihood fits for models that are not nested,

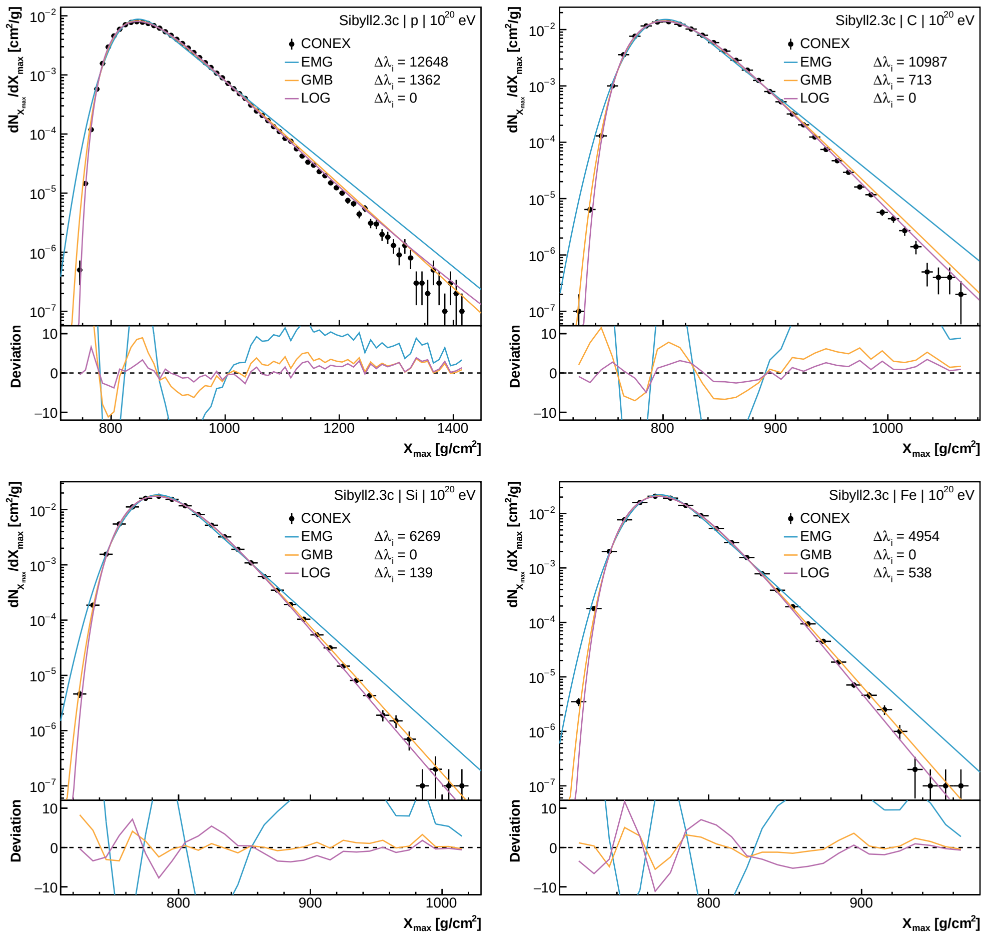

Figure 1: Examples of fits of $X_{\max}$ distributions. The primary particle is indicated at the top-right corner of each plot. Fit functions are shown as colored solid lines, while the simulated $X_{\max}$ distribution is shown as circular dots. The bottom panels show the deviation of each fitted function to the simulated point, defined as the difference between the function and the point divided by the statistical uncertainty of the point. Only results for SIBYLL2.3C are shown in this example. From [11].

the Wilks' theorem [22] do not hold and one can not rely on the traditional likelihood-ratio test and select models by computing p-values. Instead, based on the Akaike information theory [23], a direct comparison of the maximized likelihood functions is presented. The process is as follows: the model with the largest likelihood value is taken as a reference and the logarithm of the likelihood function is computed for this model (λ_{max}); the log-likelihood differences ($\Delta\lambda_i$) of the i-th to the reference model is computed; these differences, which coincide with the so-called Akaike differences, give an estimate of how many information is lost by exchange the best model (the one with λ_{max}) by the i-th model.

In table 1 the values of $\Delta\lambda_i$ obtained in each fit are presented. Note that a model with a value of zero in any case means that this is the best model for this case and also that small values of $\Delta\lambda_i$ do not provide an argument for discarding a model in favor of others. In the case of proton and carbon primaries, as can be seen in Table 1, the LOG distribution provides the best description of simulated

$X_{\max}$ data for all energies and hadronic interaction models, except in the case of EPOS-LHC with carbon primaries. In this last case, the GMB provides a better fit.

For silicon and iron simulations, on the other hand, the GMB distribution is preferred in almost all cases except for EPOS-LHC - silicon - 10^{20} eV and QGSJETII.04 - iron - 10^{20} eV. Note that the large values of $\Delta\lambda_i$ for the EMG distribution in almost all cases suggests that this function should be disregarded in $X_{\max}$ analyses. Between the LOG and the GMB distributions, it is conservative to say that differences between both descriptions are only marginal.

QGSJETII.04																
Primary	Proton				Carbon				Silicon				Iron			
$\log(E_0/\mathrm{eV})$	17	18	19	20	17	18	19	20	17	18	19	20	17	18	19	20
EMG	10113	11209	12226	12830	6636	6099	5181	5160	4213	3743	3447	3151	4920	5251	4875	4872
GMB	675	1044	1285	1397	131	105	32	104	0	0	0	0	0	0	0	19
LOG	0	0	0	0	0	0	0	0	402	381	384	330	202	79	81	0
EPOS-LHC																
EMG	8932	10507	13115	14264	4325	3884	3728	3027	2156	1236	1315	0	1742	1066	1571	1563
GMB	28	573	1425	1865	0	0	0	0	0	0	0	475	0	0	0	0
LOG	0	0	0	0	232	293	262	272	526	629	643	1222	681	754	781	802
SIBYLL2.3C																
EMG	9319	10117	11619	12648	11851	11493	11277	10987	6492	6637	6559	6269	6542	6282	5655	4954
GMB	420	666	1103	1362	914	805	760	713	0	0	0	0	0	0	0	0
LOG	0	0	0	0	0	0	0	0	247	182	123	139	326	379	495	538

Table 1: Relative log-likehood values ($\Delta\lambda_i$) of the fit of the unbinned $X_{\max}$ distributions for the three hadronic interaction models and primary particle energy ranging from 10^{17} to 10^{20} eV. From [11].

4. Parametrization in terms of primary energy and mass

Each function presented in section 2 and used to describe simulated $X_{\max}$ distributions have three parameters. The values of these parameters were obtained for combinations of four primary masses and four primary energies. A functional form is proposed to describe the dependency of these parameters with primary energy and mass:

$$\theta(E_0, A) = a(A) + b(A)\log_{10} E_0 + c(A)(\log_{10} E_0)^2\,, \tag{4.1}$$

where

$$\begin{aligned} a(A) &= a_0 + a_1 \log_{10} A + a_2(\log_{10} A)^2\,, \\ b(A) &= b_0 + b_1 \log_{10} A + b_2(\log_{10} A)^2\,, \\ c(A) &= c_0 + c_1 \log_{10} A + c_2(\log_{10} A)^2\,. \end{aligned} \tag{4.2}$$

For the values of a_i, b_i and c_i and their statistical uncertainty the reader is referred to [11]. This parametrization allow to describe $X_{\max}$ distributions within the energy range 10^{17-20} eV for any primary mass in the range 1-56. The uncertainty in the prescription of $X_{\max}$ from the use of 4.1 and 4.2 can be quantified by evaluating the differences between the first and second moments of parametrized distributions (par) to the simulated ones (MC). These differences are shown in

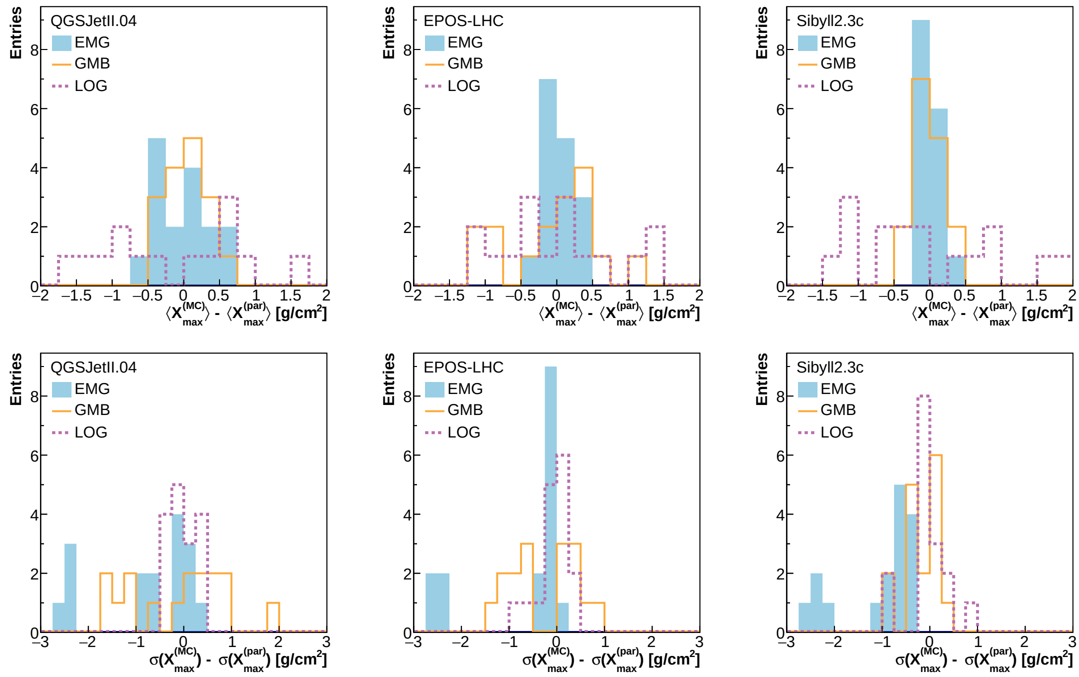

Figure 2: Error on the first moment (upper plots) and second moment (lower plots) between the parametrized distributions (par) and the simulated (MC) $X_{\max}$ distributions.

figure 2. It is seem that the difference between first and second moments between simulated and parametrized distributions are smaller than 2 g/cm^2 and 3 g/cm^2, respectively.

Similar parametrizations of the functions EMG and GMB were already performed into references [17] and [18] with much smaller sets of CONEX simulations (10^3 for each energy - mass - hadronic model bin). The effect of using large samples of simulated showers together with the unbinned maximum likelihood approach of this work reflects into a more reliable parametrization of $X_{\max}$ distributions. This effect can be seen in figure 3, where the parametrizations of GMB and LOG functions are compared to those of [17] and [18] at the energy of 10^{19} eV for proton (left) and iron (right) primaries for the case of QGSJETII.04 simulations.

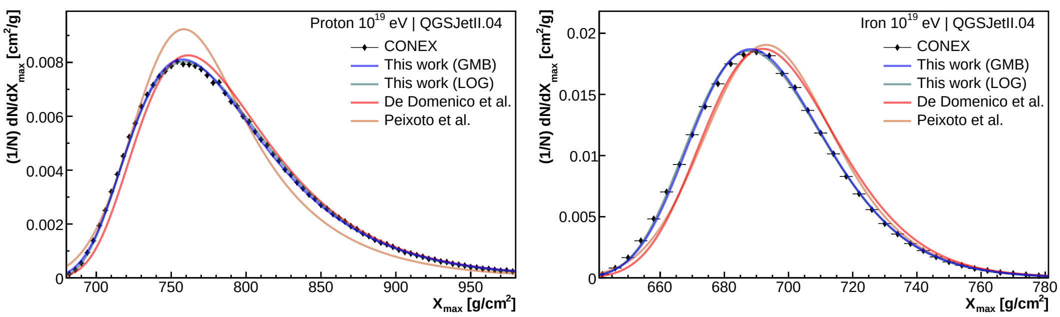

Figure 3: Comparison of parametrization using equations 4.1 and 4.2 with parameters of [11] to those of Peixoto et al. [17] and De Domenico et al. [18]. The case of proton (iron) primaries is shown in the left (right) plot for an energy of 10^{19} eV and simulations with QGSJETII.04. From [11].

The net effect of choosing one of the other available parametrizations instead of the ones presented here can be quantified by the differences on first and second moments between distinct parametrizations. Note that a direct comparison of the parametrizations in [17] and [18] with the present ones is only feasible when the same hadronic model is used. This comparison is shown in figure 4 for the cases of EPOS-LHC and QGSJETII.04 for the same functional forms used in these references for proton and iron primaries. Deviations of $\langle X_{\max}\rangle$ from previous parametrizations can be as large as 20 g/cm^2. With respect to RMS($X_{\max}$), these deviations can reach 12 g/cm^2.

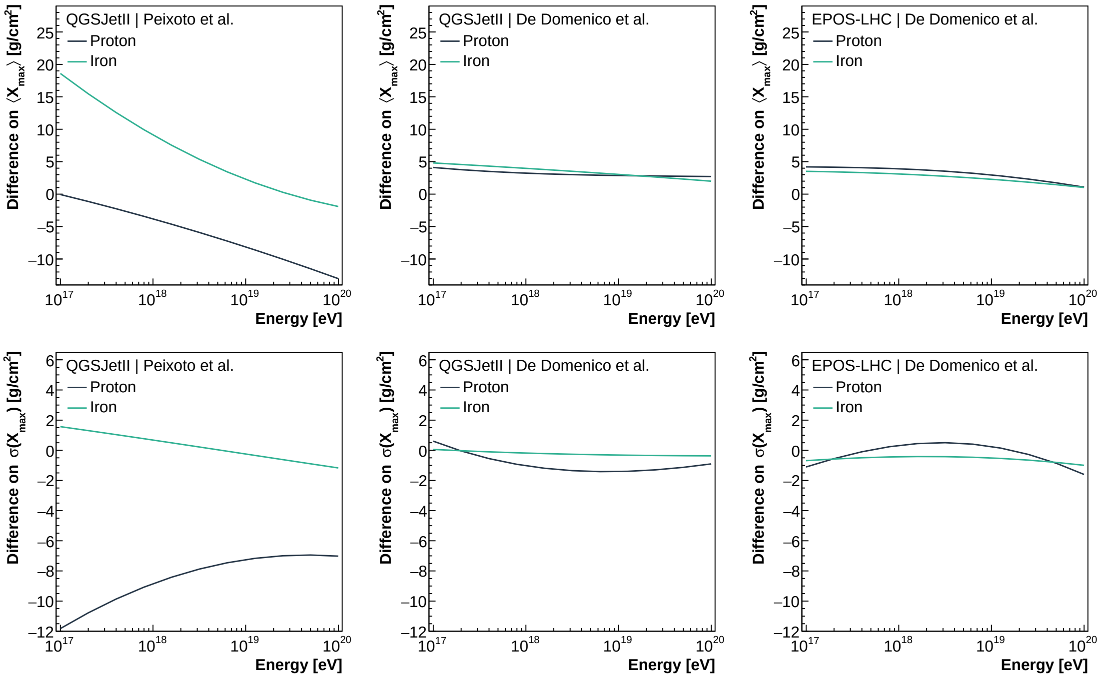

Figure 4: Comparison between first (upper plots) and second (lower plots) moments of parametrized $X_{\max}$ distributions with parametrizations from [17] (left) and [18] (middle and right). The hadronic interaction model is indicated with the boxes. From [11].

5. Summary

Determination of UHECR composition in current observatories depends on the knowledge of the functional form of expected $X_{\max}$ distributions. A parametrization of these distributions was studied by means of three distinct functions, which were compared in terms of their relative likelihoods. It was shown that the EMG function, proposed in [17], results in the overall worst description of simulated distributions. The LOG distribution, proposed here, provides a competitive description of simulated $X_{\max}$ data along with the GMB distribution, proposed in [18]. The use of the GMB distribution is suggested here as the best alternative as it provides the best likelihood in most studied cases.

All three functional forms, having three free parameters each, were parametrized as a function of the primary cosmic ray's energy and mass. Comparison of these parametrizations to the simulated data sets reveal that the maximum expected deviation in $\langle X_{\max}\rangle$ and RMS($X_{\max}$) is of 2 g/cm^2 and 3 g/cm^2, respectively.

Acknowledgements

We acknowledge the support from Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) through grant 2015/15897-1. We also acknowledge the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing HPC resources of the SDumont supercomputer, which have contributed to the research results reported within this paper (sdumont.lncc.br). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

References

[1] K. H. Kampert and M. Unger, Astropart. Phys. **35**, 660 (2012).

[2] A. Aab *et al.* [Pierre Auger Collaboration], Phys. Rev. D **90**, no. 12, 122005 (2014).

[3] R. U. Abbasi *et al.* [Telescope Array Collaboration], Astrophys. J. **858**, no. 2, 76 (2018).

[4] A. Aab *et al.* [Pierre Auger Collaboration], Phys. Rev. D **90**, no. 12, 122006 (2014).

[5] J. Bellido [Pierre Auger Collaboration], PoS ICRC **2017**, 506 (2018).

[6] A. Aab *et al.* [Pierre Auger Collaboration], JCAP **1704**, no. 04, 038 (2017), Erratum: [JCAP **1803**, no. 03, E02 (2018)].

[7] S. Blaess, J. A. Bellido and B. R. Dawson, arXiv:1803.02520 [astro-ph.HE].

[8] A. Yushkov, M. Risse, M. Werner and J. Krieg, Astropart. Phys. **85**, 29 (2016).

[9] M. S. Muzio, M. Unger and G. R. Farrar, arXiv:1906.06233 [astro-ph.HE].

[10] B. Rossi and K. Greisen, Rev. Mod. Phys. **13**, 240 (1941).

[11] L. Arbeletche and V. de Souza, arXiv:1903.03174 [astro-ph.HE].

[12] T. Bergmann *et al.* Astropart. Phys. **26**, 420-432 (2007).

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

[14] A. Fedynitch, F. Riehn, R. Engel, T. K. Gaisser and T. Stanev, arXiv:1806.04140 [hep-ph].

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

[16] C. Baus, R. Engel, T. Pierog, R. Ulrich and M. Unger, Proc. of 32nd ICRC.

[17] C. J. Todero Peixoto, V. de Souza and J. A. Bellido, Astropart. Phys. **47**, 18 (2013).

[18] M. De Domenico, M. Settimo, S. Riggi and E. Bertin, JCAP **1307**, 050 (2013).

[19] T. Antal, M. Droz, G. Györgyi and Z. Rácz, Phys. Rev. Lett. **87**, no. 24, 240601, (2001).

[20] F. James and M. Roos, Comput. Phys. Commun. **10**, 343 (1975).

[21] R. Brun and F. Rademakers, Nucl. Instrum. Meth. A **389**, 81 (1997).

[22] S. S. Wilks, Annals Math. Statist. **9**, no. 1, 60 (1938).

[23] H. Akaike. Information theory as an extension of the maximum likelihood principle. Pages 267-281 in B. N. Petrov, and F. Csaki, (eds.) Second International Symposium on Information Theory. Akademiai Kiado, Budapest (1973).