On the statistics of the mean value determination of the depth of maximum for extensive air showers

A. D. Supanitsky\textsuperscript{1}, G. Medina-Tanco\textsuperscript{2}

\textsuperscript{1} Instituto de Astronomía y Física del Espacio (IAFE), CONICET-UBA, Argentina.
\textsuperscript{2} Departamento de Física de Altas Energías, Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de México, A. P. 70-543, 04510, México, D. F., México.

supanitsky@iafe.uba.ar

Abstract: The nature of the primary cosmic ray at the highest energies is still a matter of debate. The latest results obtained by the Pierre Auger Observatory, concerning the measurement of the mean value and the fluctuations of the atmospheric depth at which the showers reach the maximum development, $X_{\text{max}}$, are inconsistent with the ones obtained by the HiRes Collaboration. From comparison with air shower simulations it can be seen that, while the Auger data may be interpreted as a gradual transition to heavy nuclei for energies larger than $\sim 2 \times 10^{18}$ eV, the HiRes data are consistent with a composition dominated by protons. Recently, the Telescope Array experiment reported the measurement of the mean value of $X_{\text{max}}$ as a function of primary energy. The Telescope Array data is consistent with a composition dominated by protons, which is in agreement with the HiRes results. It has been suggested that a possible explanation of the deviation of the mean value of $X_{\text{max}}$ from the proton expectation observed by Auger could be explained by a statistical bias originated as a consequence of the asymmetry of the $X_{\text{max}}$ distribution combined with the decrease of the number of events as a function of primary energy. We show in this work that considering a better description of the $X_{\text{max}}$ distribution the suggested bias becomes negligible and then we conclude that the deviation of the Auger data from the proton expectation cannot be explained by this statistical effect.

Keywords: Cosmic Rays, Composition Determination, Shower Maximum

1 Introduction

One of the most important limitations for the determination of the composition of the cosmic rays comes from the lack of knowledge of the hadronic interactions at the highest energies. Composition studies are based on the comparison of experimental data with Monte Carlo simulations of atmospheric cosmic ray showers, which makes use of hadronic interaction models which extrapolate the available low energy accelerator data to the energies of the cosmic rays.

One of the most sensitive parameters to the mass of the primary cosmic ray is the atmospheric depth at which the showers reach their maximum development. Lighter primaries generate showers that are more penetrating, producing larger values of $X_{\text{max}}$. Also, the fluctuations of this parameter are smaller for heavier nuclei. The Pierre Auger Observatory and the HiRes experiment are able to observe directly the longitudinal development of the showers by means of fluorescence telescopes. Therefore, in both experiments, the $X_{\text{max}}$ parameter of each observed shower can be reconstructed from the data taken by the telescopes.

The mean value and the standard deviation of $X_{\text{max}}$, as a function of primary energy, obtained by Auger [3, 4] and HiRes [6] appear to be inconsistent. From the comparison with simulations, the Auger data suggest a transition to heavier nuclei starting at energies of order of $2 \times 10^{18}$ eV, whereas, the HiRes data are consistent with protons in the same energy range. Note that, although the reduced statistics, the $X_{\text{max}}$ data from the Telescope Array experiment is consistent with the HiRes results [5].

In Ref. [1] a new parameter, the difference between the mean value and the standard deviation of $X_{\text{max}}$, was introduced in order to reconcile the Auger and HiRes results. This new parameter has the advantage of being much less sensitive to the first interaction point than the mean value and the standard deviation separately. From a comparison of the experimental values of this parameter, obtained by Auger and HiRes, with simulated data, they infer that the composition of the cosmic rays is dominated by protons. They say that the energy dependence of the distribution of $X_{\text{max}}$, observed by Auger, seems to be caused by an unexpected change in the depth of the first interaction point, which can be explained by a rapid increase of the cross section and/or increase of the inelasticity. Both possibilities require an abrupt onset of new physics in this energy range, which makes them questionable. They also suggest that the deviation of the distribution of $X_{\text{max}}$ from the proton expectation, present in the Auger data, could be originated in the statistical techniques used to analyze the data. In particular, they suggest that the deviation of the mean value of $X_{\text{max}}$ from the proton expectation could be explained by a bias originated from the exponential nature of the $X_{\text{max}}$ distribution and the decreasing number of events as a function of primary energy.

In this work we show that, considering a better description of the $X_{\text{max}}$ distribution, the bias in the determination of the mean value of $X_{\text{max}}$ become more than one order of magnitude smaller than the one obtained for the exponential distribution. We find that the value of the bias in the last energy bin (the one with the smallest number of events) of the Auger data, published in Ref. [4], is $\lesssim 1$ g cm$^{-2}$, which is one order of magnitude smaller than the systematic uncertainties on the determination of the mean value of $X_{\text{max}}$ estimated in Ref. [4].
2 Numerical approach

Following Refs. [1, 2] let us consider the parameter,

$$\xi(N) = 1 - \frac{\text{mode}[X^N_{\text{max}}]}{\langle X_{\text{max}} \rangle},$$

(1)

where $\langle X_{\text{max}} \rangle$ is the mean value of the $X_{\text{max}}$ distribution,

$$\bar{X}^N_{\text{max}} = \frac{1}{N} \sum_{i=1}^{N} X_{\text{max}}^i,$$

(2)

is the sample mean corresponding to samples of size $N$ and $\text{mode}[X^N_{\text{max}}]$ is the value of $\bar{X}^N_{\text{max}}$ that occurs most frequently, i.e. the maximum of the distribution function of $X^N_{\text{max}}$. Therefore, the bias on the determination of $X_{\text{max}}$ appears when a particular realization of the sample mean is equal to the mode of the sample mean distribution function. Note that the sample mean (Eq. (2)) is an unbiased estimator of the mean of the exponential distribution, i.e. $E[\bar{X}^N_{\text{max}}] = \langle X_{\text{max}} \rangle$. In Ref. [1] it is shown that approximating the $X_{\text{max}}$ distribution by an Exponential function $\tilde{\xi}_E(N)$ results inversely proportional to the sample size, $\tilde{\xi}_E(N) = 1/N$. In order to better describe the distribution of $X_{\text{max}}$ two different types of functions are considered. They are chosen in such a way that the distribution of $\bar{X}^N_{\text{max}}$ can be obtained, at least, in a semi-analytical way. The first function considered is a shifted-Gamma distribution [7],

$$P_G(X_{\text{max}}) = \begin{cases} \frac{(X_{\text{max}} - X_0)^{k-1}}{\Gamma(k) \tau_X} \exp \left( - \frac{X_{\text{max}} - X_0}{\tau_X} \right), & X_{\text{max}} \geq X_0, \\ 0, & X_{\text{max}} < X_0, \end{cases}$$

(3)

where $k = 5$ and the other two parameters can be obtained from the mean value and the standard deviation of $X_{\text{max}}$,

$$X_0 = \langle X_{\text{max}} \rangle - k \tau_X,$$

(4)

$$\tau_X = \frac{\sigma[X_{\text{max}}]}{\sqrt{k}}.$$  

(5)

The second function under consideration is the convolution between an exponential function and a Gaussian (Exp-Gauss),

$$P_{E\!G}(X_{\text{max}}) = \frac{1}{\sqrt{2\pi}\beta} \int_{-\infty}^{\bar{X}^N_{\text{max}}} du \exp \left( - \frac{X_{\text{max}} - u}{\lambda} \right) \times$$

$$\times \exp \left( - \frac{(u - \alpha)^2}{2\beta^2} \right) = \frac{1}{\sqrt{2\lambda}} \exp \left( - \frac{X_{\text{max}} - \alpha}{\sqrt{2\lambda}} \right) \times$$

$$\times \text{Erfc} \left( \frac{\beta \sqrt{2}}{\sqrt{2\lambda}} - \frac{X_{\text{max}} - \alpha}{\sqrt{2\lambda}} \right),$$

(6)

where $\alpha$, $\beta$ and $\lambda$ are fitting parameters and

$$\text{Erfc}(z) = 1 - \frac{2}{\sqrt{\pi}} \int_{0}^{\infty} dt \exp \left( - t^2 / 2 \right).$$

(7)

A library of simulated showers was generated by using the program CONEX (\texttt{v2r2.3}) [8]. Monochromatic samples of $10^4$ proton showers were generated from $\log(E/eV) = 18$ to $\log(E/eV) = 19.5$ in steps of $\Delta \log(E/eV) = 0.1$. The arrival directions of the showers follow an isotropic distribution, such that the zenith angle is in the interval $[0^\circ, 60^\circ]$. The hadronic interaction models considered are QGSJET-II [9] and EPOS 1.99 [10].

The mean value and the standard deviation (needed for the description of the $X_{\text{max}}$ distribution using the shifted-Gamma function) were fitted with a quadratic function and a linear function of $\log(E)$, respectively (see Ref. [2]).

The distribution functions of $\bar{X}^N_{\text{max}}$ for every energy and hadronic interaction model considered, were fitted with the Exp-Gauss function, Eq. (6). The parameters $\alpha$, $\beta$ and $\lambda$ were fitted with linear functions of $\log(E)$, in order to obtain the Exp-Gauss representation of the $X_{\text{max}}$ distribution for every value of energy in the interval $[10^{18}, 10^{19.5}]$ eV (see Ref. [2] for details).

Figure 1 shows the distributions of $\bar{X}^N_{\text{max}}$, obtained by using CONEX with QGSJET-II, for $\log(E/eV) = 18.5$ and $\log(E/eV) = 19.5$. Red solid lines correspond to the fits of the simulated data with the Exp-Gauss function. The blue dashed lines correspond to the shifted-Gamma function, Eq. (3), for which the parameters $X_0$ and $\tau_X$ are obtained by using the expressions of $\langle X_{\text{max}} \rangle$ and $\sigma[X_{\text{max}}]$ from Ref. [2] to calculate $X_0$ and $\tau_X$ from Eqs. (4) and (5), respectively. From the figure it can be seen that the Exp-Gauss function is a better fit to the simulated data than the shifted-Gamma function. It can also be seen that the tail to larger values of $X_{\text{max}}$ is slightly overestimated by the Exp-Gauss distribution and underestimated by the shifted-Gamma function. Therefore, the distribution function of the universe (samples with $N \to \infty$) should fall between this two functions.

The distribution of $\bar{X}^N_{\text{max}}$ can be calculated by means of the characteristic function, which is defined as the expectation value of $\text{exp}(it \langle X_{\text{max}} \rangle)$, i.e. $\phi_{X_{\text{max}}} (t) = E[\text{exp}(it \langle X_{\text{max}} \rangle)]$. It is straightforward to show that the characteristic function of $\bar{X}^N_{\text{max}}$ is given by $\phi_{\bar{X}^N_{\text{max}}} (t) = (\phi_{X_{\text{max}}}(t/N))^N$ [11].

The characteristic function of the shifted-Gamma distribution is $\phi_{G_{\text{max}}}^S(t) = \exp(iX_0 t) (1 - it \tau_X)^{-k}$ and then the characteristic function of $\bar{X}^N_{\text{max}}$ is given by $\phi_{\bar{X}^N_{\text{max}}}^S(t) = \exp(iX_0 t) (1 - it \tau_X/N)^{-kN}$, which corresponds also to a shifted-Gamma distribution. Therefore, the distribution function of $\bar{X}^N_{\text{max}}$ is given by,  

$$P_G(\bar{X}^N_{\text{max}}) = \begin{cases} \frac{(X_{\text{max}} - X_0)^{k-1}}{\Gamma(k) \tau_X N} \exp \left( - \frac{X_{\text{max}} - X_0}{\tau_X N} \right), & \bar{X}^N_{\text{max}} \geq X_0 \\ 0, & \bar{X}^N_{\text{max}} < X_0 \end{cases}$$

(8)

By using Eq. (8) it is easy to show that,

$$\xi_G(N) = \frac{\sigma[X_{\text{max}}]}{\sqrt{k} \langle X_{\text{max}} \rangle} \frac{1}{N}.$$

(9)

In this case, $\xi$ is also proportional to $1/N$ but it is suppressed by the ratio between the standard deviation and the mean value of $X_{\text{max}}$. The blue solid line on the top panel of Fig. 2 corresponds to $\xi_G$ as a function of $N$ for $\log(E/eV) = 19.5$. From the figure, it can be seen that $\xi_G$ is more than one order of magnitude smaller than the function $1/N$.

The distribution function of $X_{\text{max}}$ is affected by the presence of fluctuations introduced by the detectors. The distribution function of $\bar{X}^N_{\text{max}}$, including a Gaussian uncertainty on the determination of $X_{\text{max}}$ is given by,

$$p_{G_{\text{RD}}}(\bar{X}^N_{\text{max}}) = \frac{\sqrt{N}}{\sqrt{2\pi}\sigma_{\text{Rec}}} \int_0^{\infty} dx \bar{P}_G(x) \exp \left( - \frac{(\bar{X}^N_{\text{max}} - x)^2}{2\sigma_{\text{Rec}}^2 / N} \right),$$

(10)
where $\sigma_{\text{Rec}}$ is the standard deviation of such uncertainty. The mode of this distribution is calculated numerically. Dashed and dashed-dotted lines on the top panel of Fig. 2 correspond to parameter $\xi_{EG}(N)$ obtained for $\sigma_{\text{Rec}} = 20$ g cm$^{-2}$ and $\sigma_{\text{Rec}} = 40$ g cm$^{-2}$, respectively. When a symmetric uncertainty on the determination of $X_{\text{max}}$ is included, the parameter $\xi$ becomes still smaller and decreases for increasing values of the uncertainty. This is due to the fact that $\xi$ is larger for asymmetric distributions, like the exponential, and the convolution of the pure $X_{\text{max}}$ distribution with a Gaussian is more symmetric than the original one.

The characteristic function of the Exp-Gauss distribution is the product of the characteristic function of the exponential distribution, $(1-i\lambda t)^{-1}$, with the one corresponding to a Gaussian, $\exp(iat-\beta^2t^2/2)$. Then, the characteristic function of $X_{\text{max}}$ is then given by,

$$\phi_{X_{\text{max}}}^{EG}(t) = \left(1-i\frac{\lambda}{N}t\right)^{-N} \exp\left(iat-\frac{\beta^2t^2}{2}\right),$$

which corresponds to the convolution of a Gamma distribution with a Gaussian,

$$P_{EG}(X_{\text{max}}) = \frac{N^{N+1/2}}{\sqrt{2\pi}\beta\lambda^N\Gamma(N)} \int_{-\infty}^{\infty} du \left(\frac{X_{\text{max}}}{u} - 1\right)^{N-1} \exp\left(-\frac{X_{\text{max}}^N - u}{\lambda/N}\right) \exp\left(-\frac{(u-\alpha)^2}{2\beta^2/N}\right).$$

Last integral is calculated numerically in order to obtain the mode of the resultant distribution. The solid red line in the bottom panel of Fig. 2 shows $\xi_{EG}$ as a function of the sample size for $\log(E/eV) = 19.5$. Note that $\xi_{EG}$ is smaller than $\xi_{G}$, this is due to the more extended tail to larger values of the Exp-Gauss distribution compared with the corresponding one to the shifted-Gamma distribution. In any case, $\xi_{EG}$ is still about one order of magnitude smaller than $1/N$. As for the case of the Gamma distribution, dashed and dashed-dotted red lines correspond to $\sigma_{\text{Rec}} = 20$ g cm$^{-2}$ and $\sigma_{\text{Rec}} = 40$ g cm$^{-2}$, respectively. In this case the effect of the uncertainty on the determination of $X_{\text{max}}$ is included in $P_{\text{Rec}}$, just by replacing the parameter $\beta$ by $\beta = \sqrt{\beta^2 + \sigma_{\text{Rec}}^2}$. As expected, the curves that include the uncertainty on the determination of $X_{\text{max}}$ fall below the one corresponding to the ideal case.

![Fig. 1](image1.png) ![Fig. 2](image2.png)
Fig. 3: $\xi$ (left panel) and $\Delta X_{\text{max}}$ (right panel) as a function of $\log(E/\text{eV})$, for the statistics of the Auger data of Ref. [4]. Solid lines correspond to QGSJET-II and dashed lines correspond EPOS 1.99.

which there is no uncertainty on the determination of $X_{\text{max}}$ (which gives larger values of $\xi$, as shown before). It can be seen that the values of $\xi$, obtained by using the Exp-Gauss distribution and the shifted-Gamma distribution, are more than one order of magnitude smaller than the corresponding one for the exponential distribution, in the whole energy range and for both hadronic interaction models considered. As in the previous calculation, $\xi_{\text{EG}}$ results are smaller than $\xi_{\text{EG}}$. In fact, the $\xi$ curve corresponding to the true distribution of $X_{\text{max}}$ should fall between the curves corresponding to the Exp-Gauss and the shifted-Gamma representation of the $X_{\text{max}}$ distribution.

The right panel of Fig. 3 shows the parameter $\Delta X_{\text{max}} = \langle X_{\text{max}} \rangle - \langle X_{\text{max}} \rangle$ which gives the grammage of the shift suffered by $\langle X_{\text{max}} \rangle$ if $\langle X_{\text{max}} \rangle$ takes the value of the mode of its distribution. It can be seen, that for the last energy bin, the one with 47 events, $\Delta X_{\text{max}}$ is $\lesssim 1\, \text{g cm}^{-2}$, which is much smaller than the systematic uncertainties on the determination of $\langle X_{\text{max}} \rangle$ estimated in Ref. [4].

Note that $\xi_{\text{EG}}$ calculated by using EPOS 1.99 is larger than the corresponding one for QGSJET-II, this is due to the fact that the $X_{\text{max}}$ distributions obtained with EPOS 1.99 are more asymmetric (increase faster, coming from small values of $X_{\text{max}}$, and have a more extended tail) than the corresponding ones to QGSJET-II.

3 Conclusions

In this work we studied in detail the statistical bias in the determination of the mean value of $X_{\text{max}}$, suggested in Ref. [1], as a possible explanation of the deviation of Auger data from the proton expectation. We used two different functions to fit the $X_{\text{max}}$ distribution obtained from simulations: (a) the convolution of an Exponential distribution with a Gaussian and (b) a shifted-Gamma distribution. We find that the bias obtained by using these two functions is more than one order of magnitude smaller than the corresponding one of the Exponential distribution, the one used in Ref. [1]. We find that the values of the bias, obtained for the convolution of the Exponential function with the Gaussian, are larger because it presents a more extended tail to larger values of $X_{\text{max}}$ than the shifted-Gamma distribution. We also find that the bias diminishes when a Gaussian (symmetric) uncertainty on the determination of $X_{\text{max}}$ is included.

We also calculated the expected bias, as a function of primary energy, using the actual number of events in each energy bin of the Auger data, published in Ref. [4], for both hadronic interaction models considered in this work, QGSJET-II and EPOS 1.99. We find that the largest value of the bias, corresponding to the bin with the smallest number of events, is smaller than $\sim 1\, \text{g cm}^{-2}$, about one order of magnitude smaller than the systematic errors on the determination of $\langle X_{\text{max}} \rangle$ estimated in Ref. [4].

Acknowledgment: A.D.S. is a member of the Carrera del Investigador Científico of CONICET, Argentina. The work of ADS is supported by CONICET PIP 114-201101-00360 and ANPCyT PICT-2011-2223, Argentina.

References