The Exponentially Modified Gaussian Function as a Tool for Deconvolution of Astroparticle Physics Data

Livingstone OCHILO[1]
[1]JOOUST, P.O. Box 210, Bondo, 40601, Kenya
Tel: +254 0706 550 243, Email: livingstone.ochilo@gmail.com

Abstract:
In the period 2004 – 2012, the Pierre Auger Observatory has recorded more than two million of ultra-high energy cosmic rays. In seeking to interpret the data recorded for the events, it is necessary to simulate the interaction of primary cosmic rays with the atmosphere. One of the software that is available for this kind of simulation is CONEX. In this study, CONEX is used to simulate various compositions of primary cosmic rays, whose interactions with the atmosphere result in air showers, with a distribution of depths of shower maximum (Xmax), which is treated as the true distribution. Smearing this distribution with a known σ gives the "measured" distribution. By using the Exponentially Modified Gaussian (EMG) function, we have obtained deconvoluted distribution which is generally in good agreement with the original distribution.
Keywords: Ultra-High Energy Cosmic rays, air shower, Exponentially Modified Gaussian.

1. Introduction

As is typical of astroparticle physics experiments, the Pierre Auger observatory in Argentina has collected a large amount of data on the most energetic cosmic rays, referred to as Ultra-High Energy Cosmic Rays (UHECR) from 2004 to 2012 (Aab, *et al.*, 2014). The total number of events and the quality cuts that have been applied to them is given in Table 1.

Table 1: Summary of the number of events remaining after the application of event selection criteria to the Auger data (Aab, *et al.*, 2014). The selection efficiency in each case is calculated relative to the previous cut.

Cut	Events	Efficiency (%)
Pre-selection		
air shower candidates	2573713	-
hardware status	1920584	74.6
aerosols	1569645	81.7
hybrid geometry	564324	35.9
profile reconstruction	539960	95.6
clouds	432312	80.1
$E > 10^{17.8}$ eV	111194	25.7
quality and fiducial selection		
P(hybrid)	105749	95.1
Xmax observed	73361	69.4
quality cuts	58305	79.5
fiducial field of view	21125	36.2
profile cuts	19947	94.4

Part of the data that has been collected include the energy of the cosmic rays impinging the observatory as a function of the atmospheric depth at which the resulting cascade of particles, called an air shower, reaches its maximum in terms of both the number of particles present and the energy deposited by the particles. The mean of such depths is normally symbolized as <Xmax>. One of the fundamental physics questions that the Pierre Auger experiment seeks to unravel is the composition of cosmic rays arriving at the Earth. This can potentially be done if it can be possible to associate an observed depth of shower maximum and energy with a given particle. However, the depth of a shower maximum depends not only on the energy of the particle, but also its mass. By simulating a shower similar to the one detected by a detector, it is possible to estimate the composition of the primary particles which produced the shower in the first place.

The software CONEX (Bergmann, *et al*., 2007) and the hadronic interaction model EPOS-LHC (T. Pierog, *et al*.) have been developed and expanded over the years for the simulation of air showers. Using CONEX code, with EPOS-LHC as the interaction model, air showers composed of different primary cosmic ray particles at different energies can be simulated. The development of the shower can then be studied, including the final particles reaching the ground.

The exponentially modified gaussian function (EMG) is defined as (Grushka, 1972):

$$f(x;\mu,\sigma,\lambda,\beta)=\beta\frac{\lambda}{2}e^{\frac{\lambda}{2}(2\mu+\lambda\sigma^2-2x)}\operatorname{erfc}\left(\frac{\mu+\lambda\sigma^2-x}{\sqrt{2}\sigma}\right) \quad (1)$$

where erfc is the complimentary error function defined as erfc(x) = 1 − erf(x) and λ is the reciprocal of the standard deviation. EMG is a convolution of the normal and exponential probability density functions. If a reasonably good fit of the EMG is obtained on Xmax distribution (that in addition can be smeared to mimic the effect of the detector resolution), then the variance of the underlying normal distribution and hence its standard deviation may be obtained from the relation

$$\sigma^2_{EMG}=\sigma^2+1/\lambda^2 \quad (2)$$

The mean of the EMG distribution is given by

$$x_{EMG}=\mu+1/\lambda \quad (3)$$

For a large sample of data, it is expected that the smearing of the sample should not make its mean to deviate from that of the parent distribution, hence $\acute{x}$ in Eqn. (3) should remain the same for the sample as for the parent distribution.

2. The Problem

Analysis of the variation of the mean depth shower maximum <Xmax> with energy has already revealed a trend whereby the composition of the primary cosmic rays initially gets lighter with increasing energy, E, upto lg(E/eV) = 18.26 before beginning to get heavier. The contribution of individual nuclei to the overall composition is still uncertain. Based on different astrophysical models of acceleration of cosmic rays, a truncation of the Pierre Auger <Xmax> data into "light" and "heavy" components has been proposed, in order to shed more light on the contribution of protons and heavier nuclei respectively. However, this truncation causes an introduction of a bias in the characteristics of two subsets of data thus generated, due to detector resolution. This study aims at estimating this bias.

3. Research Objectives

1. To simulate primary cosmic rays whose depth of shower maximum correspond to that of measured data.
2. To smear the simulated distributions using standard deviations that correspond to the estimated resolutions of the detectors used in collecting experimental data.
3. To use the EMG function to deconvolute the smeared distribution so as to obtain the true distribution of Xmax.
4. To compare the characteristics of the simulated "true" distribution with the measured one.

4. Literature Review

Since the first observation of an UHECR event in 1962, their chemical composition has not been definitively established. As a result of the very low flux of UHECRs, they cannot be detected directly, but instead detectors which are spread out over a large area of the surface of the earth such as the Pierre Auger detector, must be used (Nagano & Watson, 2000). At energies above 10^{14} eV, it is not possible to measure the abundance of individual elements in the cosmic ray spectrum directly. However, the mean mass of the cosmic rays at a given energy can be estimated by analyzing for example the mean atmospheric depth at which the resultant air shower initiated by the primary cosmic rays reaches its maximum development (Linsley, 1963; Kampert & Unger, 2012). Determination of the individual masses that give the average measured is currently the subject of ongoing research. It is however possible to deduce a change in mass with increasing energy of the primary nuclei, but only on a statistical basis (Kampert & Unger, 2012). Based on simulation studies, an estimate of the fractions of nuclei present in the air showers detected by the Pierre Auger Observatory has been obtained(Aab, *et al*., 2014). Although this study concluded that an assumption of only two primary cosmic ray nuclei could not describe the measured data, it was a good starting point to include more primary particles in future studies. The use of the EMG function can likewise be extended to assumptions of more primary nuclei.

5. Methodology

A variety of mixed primary compositions of cosmic rays made of two nuclei: p-He, p-Fe and p-O were simulated using the CONEX code, with EPOS-LHC as the hadronic interaction model. To begin the procedure of EMG analysis, the smeared distribution was fitted with an EMG for various p-He, p-Fe, and p-O compositions, and hence λ_{fit} , σ_{fit} and μ_{fit} were obtained. In the meantime, the true mean of the light and heavy subsets were obtained from two separate histograms, filled for light and heavy ^{events} respectively without any smearing. The standard deviation of the EMG from Equation (2) is given by

$$\sigma_{fit}^{EMG} = \sqrt{\sigma_{fit}^{2} + 1/\lambda_{fit}^{2}} \qquad (4)$$

Then

$$\sigma_{true}^{EMG} = \sqrt{\square} \qquad (5)$$

where σ_{res} is the smearing due to resolution, and the standard deviation of the unsmeared underlying Gaussian is given by

$$\sigma_{true} = \sqrt{\square} \qquad (6)$$

The fit parameters σ_{true} , λ_{fit} and μ_{fit} were then substituted in the EMG, and the function integrated to obtain the deconvoluted means of the 'light' and 'heavy' subsets of the data; the two subsets being delimited by X_{max}^{cut} (see Figure 1).

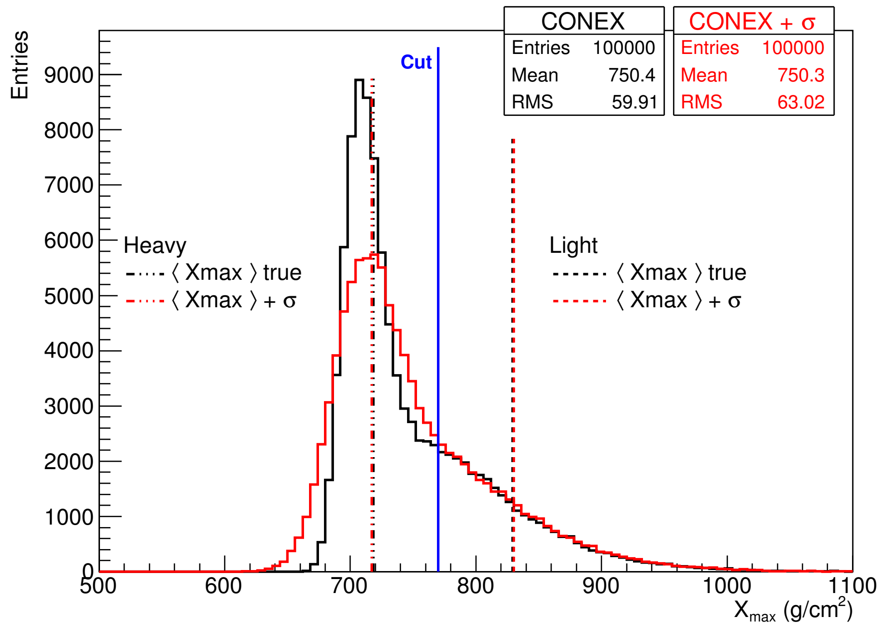

Figure 1: Simulated composition of 40% proton-60% oxygen mixture. A smear of the distribution with $\sigma_{res} = 20$ g/cm^2 is shown in red. The mean of the light and heavy subsets of data before and after smearing are represented by the vertical dashed lines.

The mean of Xmax for the light and heavy components are given respectively by Equations (7) and (8),

$$\langle X_{max}^{light} \rangle = \frac{\int_{X_{max}^{cut}}^{+\infty} X_{max} f(X_{max}; \mu_{fit}, \sigma_{true}, \lambda_{fit}) dX_{max}}{\beta_{light}} \quad (7)$$

$$\langle X_{max}^{heavy} \rangle = \frac{\int_{-\infty}^{X_{max}^{cut}} X_{max} f(X_{max}; \mu_{fit}, \sigma_{true}, \lambda_{fit}) dX_{max}}{\beta_{heavy}} \quad \langle X_{max}^{heavy} \rangle = \frac{\int_{-\infty}^{X_{max}^{cut}} X_{max} f(X_{max}; \mu_{fit}, \sigma_{true}, \lambda_{fit}) dX_{max}}{\beta_{heavy}}$$

(8)

where β_{light} and β_{heavy} β_{heavy} are the normalization factors given respectively by

$$\beta_{light} = \int_{X_{max}^{cut}}^{+\infty} f(X_{max}; \mu_{fit}, \sigma_{true}, \lambda_{fit}) dX_{max} \quad (9)$$

and

$$\beta_{heavy} = \int_{-\infty}^{X_{max}^{cut}} f\left(X_{max}; \mu_{fit}, \sigma_{true}, \lambda_{fit}\right) dX_{max} \qquad (10)$$

By subtracting the mean in the light and heavy subsets of the true distribution from the corresponding mean in the deconvoluted distribution, an estimate of the bias is obtained. All the graphs were plotted using the ROOT software, which is designed to handle large amounts of data, and is based on C++.

6. Findings and discussion

A summary of the results of this procedure, showing a sample of the fits to selected compositions together with the fit parameters in each case, is contained in Figure (2), where $\sigma_{res} = 50$ g/cm^2. Superposed plots of the actual true distributions give one an idea of how well this procedure unfolds the smeared distribution. It was observed that generally, when the primary distribution contains two distinct peaks, the smeared distribution does not reflect this, i.e. it has only one peak. In such cases, the true distribution obtained by the use of the EMG differs significantly from the actual true distribution. A look at the quality of fit represented by the values of χ^2/Ndf for different compositions shown in Figure (3) suggests that compositions containing only iron and proton generally have the lowest quality fits. This is especially so when the width of smear is 20 g/cm^2 . However, the procedure worked generally well, resulting in a bias in truncated <Xmax> close to zero as expected. A summary of the biases in <Xmax> for the resolutions σ_{res} = 20 and $\sigma_{res} = 50$ g/cm^2 are presented in Figure (4).

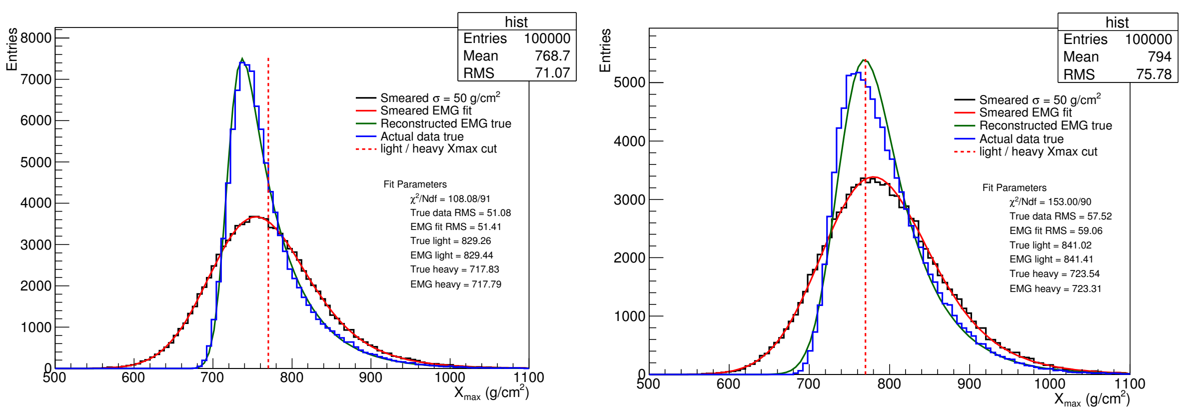

Figure 2: Smeared ($\sigma_{res} = 20$ g/cm^2) simulated 40%p-60%O and 80%p-20%O mixtures fitted with an EMG. The reconstructed true distributions as predicted by the EMG are also shown. The unit of <Xmax> shown on the r.h.s. of every plot is in g/cm^2 .

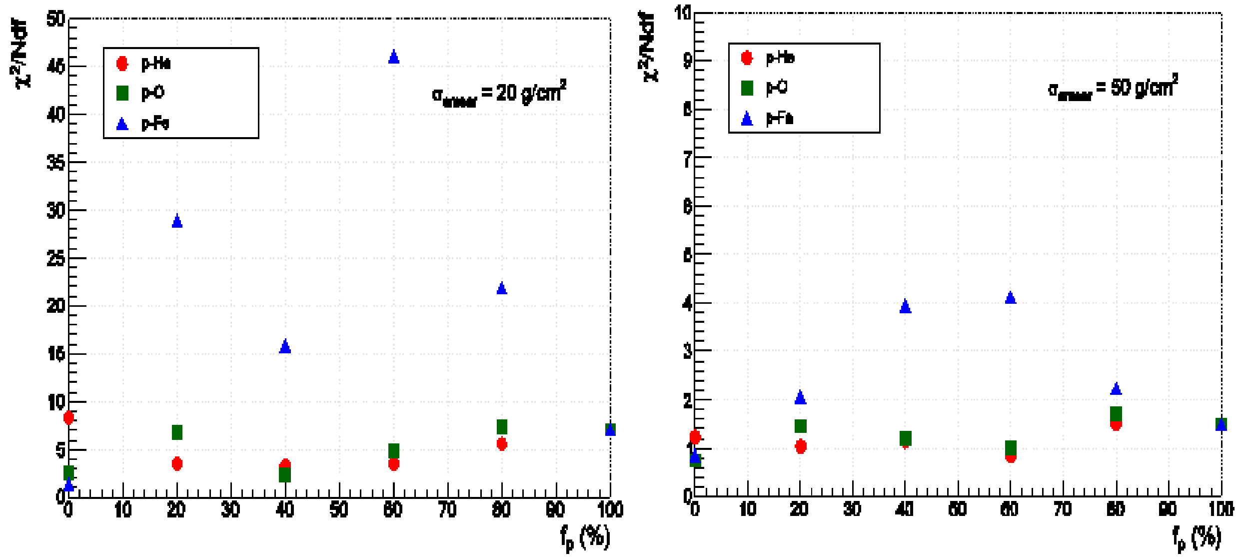

Figure 3: Evolution of the quality of EMG fit with proton fraction for p-He, p-O and p-Fe mixtures for widths of smear σ_{res} = 20 g/cm^2 (left panel) and σ res =50 g/cm^2 (right panel). The horizontal axis shows the percentage of proton in each mixture.

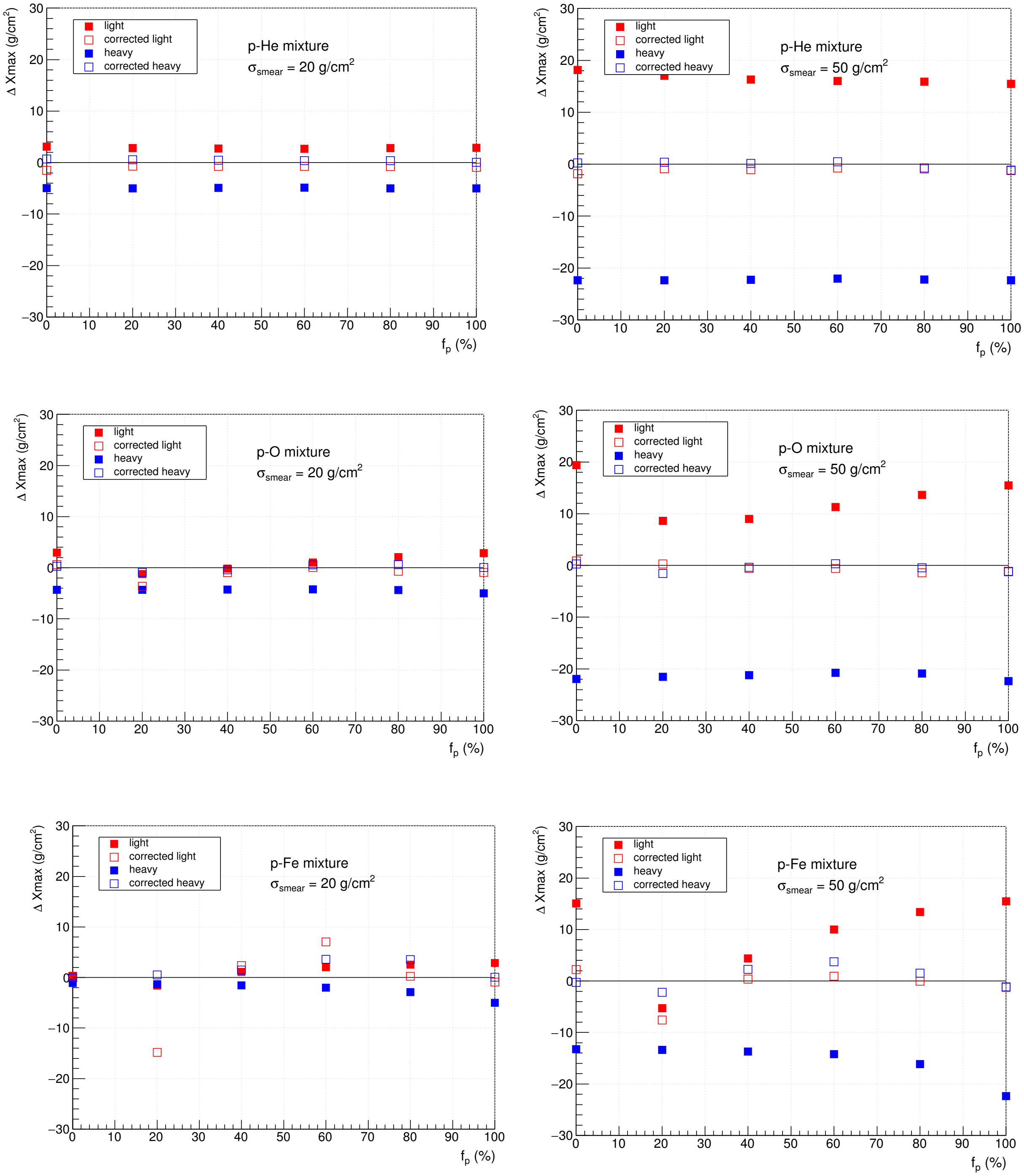

Figure 4: Bias in the light and heavy <Xmax> due to resolution effects (closed symbols) and results for the corrected <X max> values (open symbols).

7. Recommendations and areas for further research

We have investigated the possibility of using the exponentially modified gaussian function to unfold smeared simulated data produced by the CONEX code, with EPOS-LHC as the hadronic interaction model. Using a mixture of only two primary particles of p-He, p-O or p-Fe, we observe that the EMG fits the "measured" data well and hence we are able to obtain the true standard deviation and thus the true distribution with a good accuracy in cases of pure composition, compositions with σ_{smear} = 50 g/cm^2 or compositions containing only protons and helium. In most of the cases where the primary contains proton and iron, information gets lost during the smearing process, such that it is difficult to correlate the smeared data with the true data and hence unfolding is not easy. A further study is necessary to find a way around this. In order to represent the more realistic situation where many primary particles are involved, it would be of interest to simulate a composition containing more nuclei.

Acknowledgement

The author gratefully acknowledges the input of Alexey Yushkov and Markus Risse of the University of Siegen in the writing of this paper.

References

Aab, A. *et al.* (2014). Depth of maximum of air-shower profiles at the Pierre Auger observatory: measurements at energies above $10^{17.8}$ eV, *Physical Review D*

Aab, A. *et al.* (2014). Depth of maximum of air-shower profiles at the Pierre Auger observatory II. Composition implications. *Physical Review D* 90, 122006.

Bergmann, T., Engel, R., *et al*. (2007). One-dimensional hybrid approach to extensive air shower simulation. *Astropart. Phys*., 26:420.

Grushka, E. (1972). Characterization of exponentially modified gaussian peaks in chromatography. *Anal. Chem*., 44(11):1733–1738.

Kampert, K-H & Unger, M. (2012). Measurements of the cosmic ray composition with air shower experiments, *Astropart. Phys*., 35:660–678.

Linsley, J. (1963). The cosmic ray spectrum above 10^{19} eV at Volcano ranch and Haverah park, *8th ICRC, Jaipur, volume 77*.

Nagano, M. & Watson, A. A. (2000). Observations and implications of the ultrahigh-energy cosmic rays. *Rev. Mod. Phys*. 72(3):689–732.

Pierog, T., et al. (2015). EPOS-LHC: test of collective hadronization with data measured atthe CERN large hadron collider. *Phys. Rev. C*, 92:034906.