LIP-STUDENTS-20-10

A proposal for a graphic analysis of public data from the fluorescence detector of the Pierre Auger Observatory

Miguel Pereira[1,a] and Jorge Gouveia[2,b]

[1] *Instituto Universitário, Lisboa, Portugal*
[2] *Universidade da Madeira, Lisboa, Portugal*

Project supervisor: S. Andringa

October 2020

Abstract. The Pierre Auger Observatory makes 10% of the data collected available for dissemination and education activities. The collaboration team of this observatory produced teaching instruments for teachers and students of Physical Sciences using the surface detector as an instrument. We aimed to continue this work with the detector data based on the telescopes. The VISPA platform based on the phython programming language was the framework chosen to proceed with the analysis. Here, we therefore, present some results, taking into account its compatibility with the results already achieved by the collaboration. The analysis is made through the work with numbers, units, prefixes and graphical representation.

Keywords: Pierre Auger Observatory, high energy cosmic rays, air fluorescence detectors, VISPA

1 Introduction

1.1 Observatory Pierre Auger

Pierre Auger Observatory, located at an average altitude of about 1400 m, in the "Pampa Amarilla", Argentina, has, currently, a system consisting of two detectors which allow the detection of events simultaneously (referred to as a hybrid system), namely the surface detector (SD) and the fluorescence detectors (FD). We can see in figure 1 a top view of the Observatory [1]. A SD is composed by 1660 Cherenkov placed in water and arranged in a triangular shape spaced 1500 m apart, covering a total area of about 3000 km^2 .

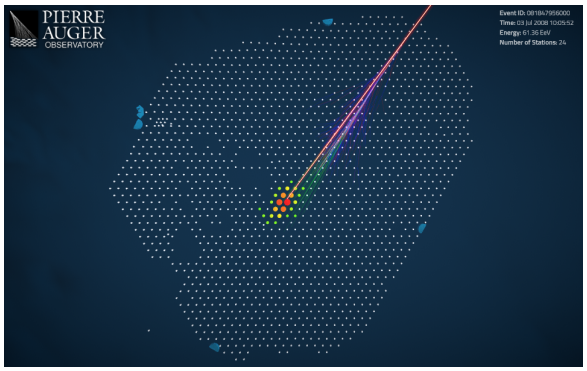

Figure 1. Top view of the Pierre Auger Observatory. The four positions of the telescopes are marked in blue, half circle. Telescopes' field of view sees events covered by surface detectors. Surface detection stations are represented by white dots.

On the other hand, the FD detectors are positioned in 4 locations (mentioned posteriorlly as 4 eyes): 1) Leones; 2) Los Morados; 3) Loma Amarilla; 4) Coihueco. Each of them with 6 optical telescopes and 440 pixels each, covering 6x30º in azimuth and 30º in zenith angle (from 2º of the horizontal plane to 32º in the sky), making up to the total of 24 telescopes, which allows to cover the area where the SD detectors are installed [2]. Telescopes are used to detect ultraviolet light in the atmosphere, so they only work during the night and on moonless nights (or during periods of the night when the moon is not visible). Atmospheric conditions, smoke or other phenomena that affect the interaction of ultraviolet light with the atmosphere, are a constraint to the final selection of data.

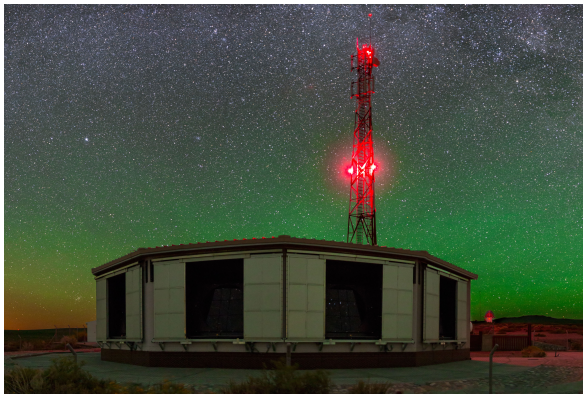

Figure 2. Telescope Los Morados, photo by Steven Saffi, University of Adelaide, Australia

Therefore, the weather stations record these constraint events in order to take then into account afterwords, in the selection of the data. In figure 2 we can see one of the fluorescence detectors in operation named Los Morados.

This document presents the work of an analysis conducted with 2 samples of the data retrieved from the FD of the Pierre Auger Observatory: sample one, a file with 3151 events and sample two, a folder with 300 files (one for each event). The analysis is presented in terms of their

[a] e-mail: mdmpereira1@gmail.com
[b] e-mail: jorge.mariano@staff.uma.pt

graphic bi-dimensional presentation – histograms and dispersion graphics, here therefore to be referred as graphics. These samples are compiled with only 10 % of all the data retrieved by the SD (i.e., the collaboration data), following similar reconstruction. Moreover, this article exemplifies the work done previously by the observatory team to prepare the "public data" analysis from the FD of the Pierre Auger Observatory. It will allow to have concrete data available, made exclusively of partial samples looking to widen the discussion at Pierre Auger Observatory in such a way as to verify their numeric content and physic significance. It can also give a didactic perspective to be explored by other interested Scientists. For the data of the first sample, about 34 identifiers and qualifiers (flags) were analyzed in order to verify the relationship between them, in particular those used in the total energy spectrum (TotalEnergy). Furthermore, we also characterized and analyzed the energy through their labels with flags: FlagSpectrumEye, FlagCalibEye and FlagXmaxEye - corresponding to their use in different analyses. So, the FlagSpectrumEye is responsible for building the energy spectrum, the FlagCalibEye is responsible for calibrating the SD detector in energy and last, the FlagXmaxEye is responsible for distinguishing protons from heavy nuclei. The analysis in calorimetric energy (CalEnergy) and total energy (TotalEnergy) allowed for the evaluation of the calibration factor. A selection of depth Xmax was carried out in terms of a bin width of $\Delta log10(E/eV) = 0.5$, which allowed to reproduce the data and check the compatibility between public and collaboration data. Using the data from the second sample, a cascade profile analysis was conducted in terms of three variables: Xmax, dEdXmax and UspL. A Gaussian adjustment was made and the calorimetric energy of an event was calculated. Finally, we evaluated the the results of this parameterization and compared it with the collaboration data in terms of the number of standard deviations.

1.2 Tools

The tool used to proceed with the graphic construction and analysis was the internet connect platform VISPA[3], using the python language (.py). The Pierre Auger collaboration team made available all the data used and organized them in text files. The data are organized in two different ways: the first data set is organized in a single file and the second data is organized in a folder, both with distinct reading modes. The data of the first analysis are compiled in a file with a total of 3151 events, 35 of which were detected by two eyes[1]. In another folder[2] 300 files are found, corresponding to 300 events, 3 of which are seen by two eyes.

[1] Sample one with file name "FD.10.txt"
[2] Sample two with folder name "2010_Dados"

2 First tests of the public data of the Fluorescence Detector

2.1 Energy

One of the questions to be clarified is the compatibility of the energy seen by each eye. The figure 3 shows the histogram of the cosmic ray energy seen by the different eyes, concerning the first sample of data. It shows the total number of events in 2010 of each eye as a function of the logarithm of the total energy. We can see that the energy limits are the same for all the four eyes. In addition, the greater the energy, the less events are seen and so the greater is the statistical error[3].

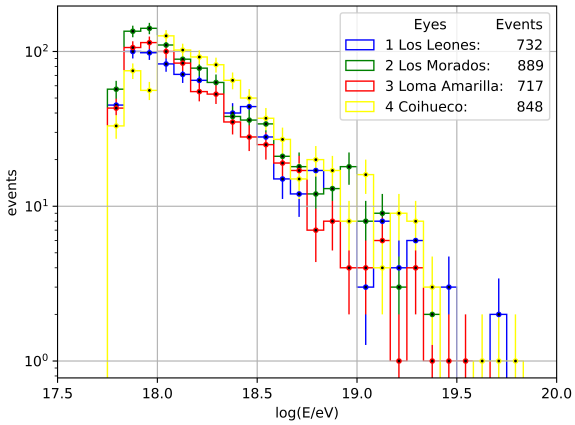

Figure 3. Energetic histogram seen by each eye

For a better comparison between the 4 eyes, it is important to take into account the table 1 that states the linear regression between the logarithm of events log10 (N/event) as a function of the average energy centered on the bin <log10(E/eV)>, allowing us to calculate the slope (a1) and the y-intercept (a0)[4]. In order to have a more accurate analysis and to see the events in the center of the bins, a cut was made in the energy range between Log10(E/eV) = 18.0 and log10(E/eV) = 19.5. A program with the same functionalities as the Excel - LibreOffice Calc - was used to perform a best fit line, taking into account the y error [5].

Table 1. Result of the linear fit, the slope a1±Δa1, y-intercept a0±Δa0, the standard deviation of the slope and standard deviation of the y-intercept respectively

Eye	a1±Δa1	a0±Δa0
1. Los Leones	-1.06±0.06	21.00±1.15
2. Los Moratos	-1.12±0.06	22.22±1.10
3. Loma Amarilla	-1.29±0.07	25.24±1.31
4. Coihueco	-1.07±0.05	21.31±1.00

In this way it was possible to calculate the statistical error of each parameter $\Delta a1$ and $\Delta a0$. Table 1 shows the

[3] Program "EnergiaT_q1.py"
[4] log10(N/event) = a1 <log10(E/eV)> + a0
[5] $\Delta log10(N/event) = \sqrt{N}/(\ln 10\ N)$

results of the constants obtained for the four eyes. An analysis of the slopes shows that eyes 1, 2 and 4 are compatible. For example, the deviation factor between eye 1 and 2 is relatively small $(-1.12 + 1.06)/0.06 \sim 1$. On the contrary, the deviation between eye 1 and 3 has major implications for the expected energy $(-1.06 + 1.29)/0.06 \sim 4$. This is where the difficult aspects lay regarding the data analysis with few events. Another example is for log10(E/eV)=19.0, where the number of events is around the units. It results in big statistical error and with that comes the difficulty in the comparison. These results can not be directly compared among different eyes or with the official results due to the fact that we do not have the exposure – ie, the time that each eye has been operating under good atmospheric conditions. In the FD, the effective exposure is also different for each energy, since the light is absorbed as it travels from the shower to the eye, and far away showers may fail to pass the minimum photon detection threshold. Therefore, this partially explains the difference of the results in statistics, and may impact the slope. For the future, a new opportunity that could be useful to improve the statistical results is the access to the operating time parameters which would allow to normalize the data in order to obtain the flux and finally, represent the energy spectrum.

2.2 Flags

In a sample of 3151 events, the selection of three flags was made - FlagSpectrumEye, FlagCalibEye, FlagXmaxEye shown in figure 4. The flag with a greater statistic is the FlagXmaxEye depth with 3093 events, followed by the 1380 events of FlagSpectrumEye and finally, the FlagCalibEye calibration with 437 events. The energies selected for the spectrum start from around 18.0 EeV, while for energy calibration were chosen higher energies that vary around 18.5 EeV. [6]

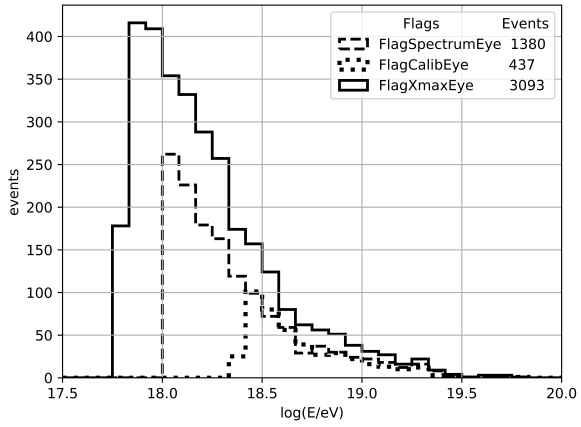

Figure 4. Histogram of the number of events according to the log of the selected energy flags associated to the eyes.

2.3 Conversion factor between total and colorimetric energy

Figure 5 shows a linear dependence between the total energy and calorimetric energy, the slope of the graphic is used to estimate the conversion. The FD measures the calorimetric energy – deposited in the atmosphere. However, part of it is carried by muons and neutrinos which are invisible to the FD, so they must be corrected. Analyzing the graphic (b), we have a central value around 1.17, and by that we can see that 17% of energy is carried by muons and neutrinos. The python program we created was conceived to generate the graphic (a) and the histogram (b)[7].

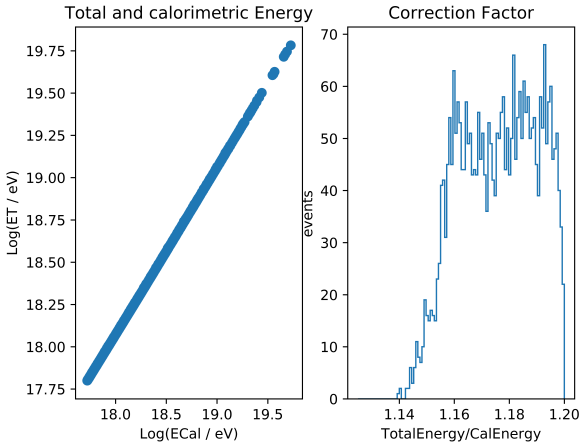

Figure 5. Relationship between total energy and colorimetric energy: (a) linear relationship between total energy as a function of colorimetric energy; (b) Distribution of the ET / ECal ratio.

2.4 Evolution of Xmax with energy

The selection of the maximum depth for the Xmax allows us to extract information from the primary particle mass [4]. Hence, Xmax is used to distinguish protons and heavy nuclei. The histogram of figure 6 allows to observe that the higher the energy of the particle the larger is the depth (Xmax). A selection criterion was adopted in terms of bins. The counts start for energies log10 (E/eV)> 17.8 and end for energies log10(E/eV)> 19.0. In fact, we can consider that the bins have a width of 0.5.

[6] Program “EnergiaT_q2.py”

[7] Program "EnergiaT_q5_scatter_hist.py"

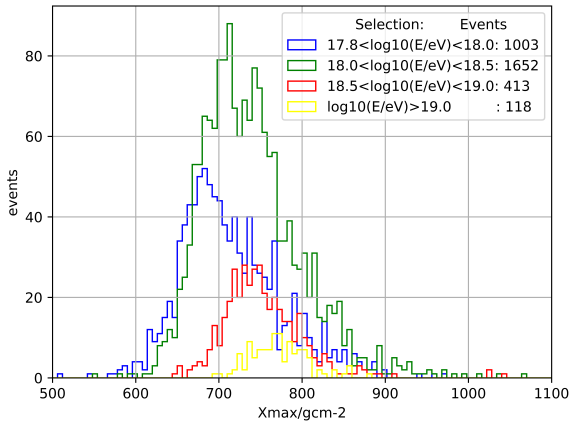

Figure 6. Number of events as a function of depth for various energy intervals

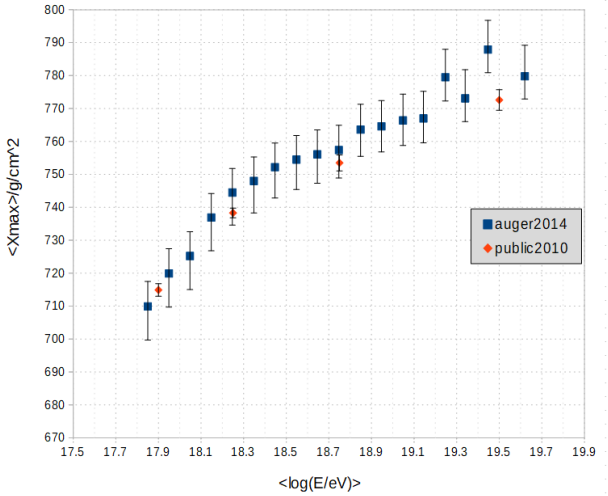

Figure 7. Xmax data from the Pierre Auger Observatory as published in A. Aab et al. [Pierre Auger Collaboration] and public data 2010.

The higher depth is verified for energies larger than log10(E/eV) > 19.0, where the statistics drastically decreases, to 188 events. This is in accordance with current knowledge[8].

2.4.1 Comparison with the values of collaboration

To compare the different results, it is important to know that the Xmax is the primary observable to separate protons from heavier nuclei, as already stated previously. Also, the collaboration results are usually shown as the evolution of the mean and dispersion as a function of the energy. In our analysis, we obtained a plot for the 4 energy bins: <log10(E)>, <Xmax>, sigma [5]. To calculate the average value of the logarithm <log10(E)> , the average value was estimated, adding the extreme values of the bins and dividing by two. The <Xmax> is, then, the mean value of the histogram and sigma the standard deviation of this histogram.

3 Evolution of the cascade in the atmosphere dE/dXmax

Folder “2010_Dados" is composed by a sample of 300 events, organized in terms of a file per sample, making a total of 300 files with all the (10% of selected) data from the year 2010. Each file contains the physical parameters measured at each moment. For example, atmospheric depth - "atmDepthProf" -, energy deposited at each depth - "energyDepositProf"-, error of the energy deposit in depth -"energyDepositProfErr"-, etc. Several profiles were analyzed but, we will only present a representative profile, associated with the event 100086876100. Analyzing the graphic of the figure 8 we can identify the max atmospheric depth (Xmax) and the corresponding energy deposition (dEdXmax). Besides this, we can also identify the width parameter, universal shower profile (L=length) (UspL). On that account, we present here what can be a didactic exploration of the attempt to answer the question: "How does the energy deposition vary along the atmosphere?". The program[9] we created through the VISPA platform generates the graphic of figure 8. This direct reading describes an increase in deposited energy in the [0, 2.5] PeV/gcm^{-2} interval for depths in the [400, 800] gcm^{-2} interval. Differently, in the next interval]800, 1100] gcm^{-2}, the energy deposit decreases. Also, an approximately symmetric distribution is formed with maximum deposition of energy (dEdXmax) of around 2.5 PeV/gcm^{-2} corresponding to a medium depth value of 800 gcm^{-2} (Xmax) .

[8] Program "EnergiaT_q6_Xmax.py"

[9] Program “segundo_read_one_file.py”

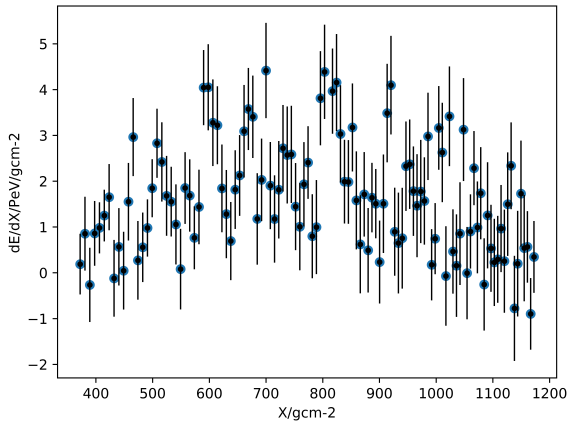

Figure 8. Energy deposited as a function of depth. Event number 100086876100.

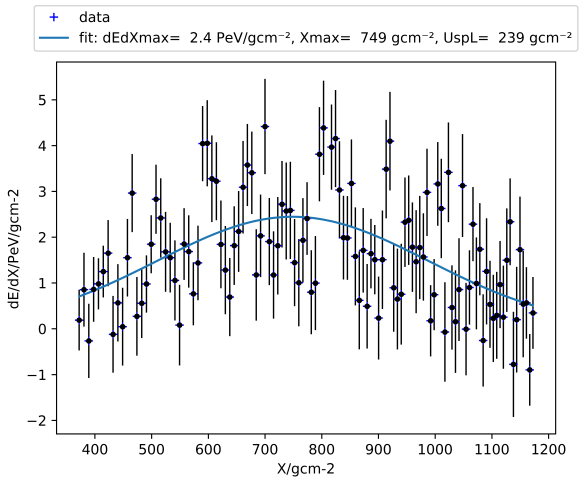

Figure 9. Energy profile deposited as a function of profundity. Adjusting a Gaussian function to experimental values. Event number 100086876100

The same program prints the data of these parameters calculated by the collaboration team allowing that way an auxiliary in the analysis. With the urge to improve the quantification of the parameters, we implemented and created a program, allowing us to make a gaussian adjustment (fit) to the experimental values[10]. Moreover, in figure 9 we have the fitted line and, in the top of the graph, the resulting adjusted parameters. First, We designed a program dedicated to the calculation of the dEdX (normalization), Xmax (mean), UspL (width) gaussian parameters and ECal[11] from the gaussian integral. Secondly, we indicated an event, concerning the type of data processing carried out. Here we show one example, but we tested that the results are similarly good for a sample of fourteen, selected at random. In addition to the graph, we have the output of the quantities dEdX, Xmax, UspL and ECal and the respective uncertainties . For the three quantities of the energy deposition profile (dEdX, Xmax, UspL), the diagonal matrix was considered, the standard deviations was extracted from the diagonal elements of the error correlation matrix. For calorimetric energy (ECal), the numerical integration extended from negative to positive infinite allows to obtain an expression Where the area of integration is calorimetric energy $ECal = dE/dXmax\ UspL\sqrt{2\pi}$. This expression was used to calculate the error through the rules of the propagation of the statistical error $\Delta ECal$[12].

[10]Program "compare_Auger_public_FD_2010.py"
[11]ECal = CalEnergy"
[12]

$$\Delta ECal = \sqrt{(\Delta dEdXmax/dEdXmax)^2 + (\Delta UspL/UspL)^2}ECal$$

With the parameters that the program generates, we present the results available to compare: 1) direct reading of the graph in figure 9; 2) Gaussian adjustment to experimental values; 3) Collaboration calculations.

Table 2. Parameters calculated by the collaboration and the authors of this work with event number 100086876100. S: 1. Direct measurement; 2. Fitting; 3. Collaboration

S	dE/dXmax PeV/gcm^{-2}	Xmax gcm^{-2}	UspL gcm^{-2}	ECal EeV
1.	2.5	800	200	-
2.	2.45±0.18	749±21	239±25	1.47±0.19
3.	2.48±0.12	719±36	230±11	1.44±0.11

We wrote another program in which the results of the Gaussian adjustment to the experimental values were generated (line 2. of the table 2) and the reading of the quantities calculated by the collaboration (line 3. of the table 2). To understand better this analysis, an exemplification with the event (100086876100) will be carried out. Then we made an interpretation of the results, i.e., consisting in a realization of a calculation of the relative deviation of values (1) and (2) in relation to (3). The relative deviation of the deposited energy, dE/dXmax, in direct reading (1) relative to the one of the collaboration (3) is: $(2.50 - 2.48)/0.12 \sim 0.17$. Repeating the same calculation to other magnitudes, we obtained, for the Xmax, $(800 - 719)/36 \sim 2$, a deviation of 2.25 because of a direct reading (800 gcm^{-2}) and a collaboration value of (719 36) gcm^{-2}. From a didactic point of view it can be argued that the value is reasonable $\sim$ 2 standard deviations. Note that this is not so bad for an estimate by eye! On the other hand, the gaussian width, UspL, is more difficult to access. By doing a visual inspection, we can approach it as a width at half height (200 gcm^{-2}), which corresponds to $(200 - 230)/11 \sim 0.83$ standard deviations. In a second approach, with more rigorous numerical analysis, we compare the results of fitting (2) with

those of the collaboration (3.). For the dE/dXmax parameter we have about three tenths of the standard deviation $(2.45-2.48)/0.12 \sim -0.3$, which is consistent. For Xmax, $(749-719)/36 \sim 0.83$ about one standard deviation, which means a good compatibility, and for the UspL we have $(239-230)/11 \sim 0.82$, about one a standard deviation, meaning a good agreement. Regarding ECal calorimetric energy, $(1.47-1.44)/0.11 \sim 0.3$, we have a deviation of about three tenths of the standard deviation, which tells us that it has a good compatibility. Thus, a fitting was carried out that corresponds to the calorimetric energy (ECal). With this we can compare the values of Fitting (2) and Collaborations (3). The result of the gaussian fit is again in agreement with that from the collaboration within about one half standard deviation $(1.47-1.43)/0.10 \sim 0.4$. Note that the collaboration uses a more complex function with an extra parameter – to which the sensitivity is even smaller. In order to detect deviations in relation to collaboration, we gathered in a histogram the result of fourteen similar analyzes described above.

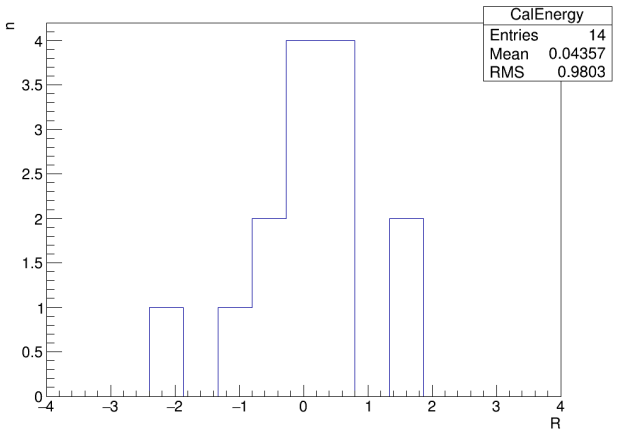

Figure 10. Ratio of the difference in CalEnergy (ECal) of the collaboration results presented in this report (fitting), to the uncertainty given by the collaboration.

We can see on the figure 10, the histogram related to the R ratio R=(collaboration-fitting)/(error collaboration), of the calorimetric energies, in which the gap between the two results is relatively small (0.04). After that, we resumed the remaining results in terms of the following parameters: mean value and RMS. The remaining results are shown in table 3.

Table 3. Results of histograms of N = 14 events.

	R - Mean	RMS
dEdXmax	-0.03	1.23
Xmax	-0.98	0.85
UspL	0.34	1.58
ECal	0.04	0.98

After analyzing both histogram and the table we can reach to some conclusions. For the dEdXmax, the distribution was significantly symmetrical with a deviation of -0.03 from zero. Furthermore, for the Xmax, our results were systematically above those of the collaboration team, representing about 1 RMS. This result has to do with the Xmax profile, because it isn't a gaussian, and it is not the best option for a more rigorous treatment. It is also necessary to consider the asymmetry parameter (UspR) that we did not take into account due to the complexity of its implementation and the lack of time to do so. This also explains the observations on the parameter of the profile width, UspL: while it didn't deviate significantly in a systematic way, it has a very large statistical variation, with an RMS of 1.58.

4 Results and Conclusions

As a conclusion, we can reflect that although we worked with real data, the low statistic was a challenge in the search for scientific results. The real challenge of rediscovering the results made it possible to carry out a set of partial results, such as the construction of energy histograms, selection of flagged events - FlagSpectrumEye, FlagCalibEye and FlagXmaxEye flags - to build the energy spectra and calibrate the SD detector, as well as, calculating the calorimetric energy respectively. Using the values of the collaboration we reached a result of Xmax in function of the <log10(E/eV)> compatible, within the margin of error. Analysis of the profile with three parameters, dEdX, Xmax, UspL, allowed us to obtain results with 2 standard deviations and with frequency less than 3 standard deviations. Finally, the calculation of calorimetric energy ECal proved to be reproducible regarding the comparison between the collaboration team results.

Acknowledgements

We would like to thank to the summer school program - LIP Internship Program 2020. This initiative allowed to learn directly from many scientists specialized in the particle physics - in particular cosmic rays. It allowed also to work on the VISPA platform, learn the python language and propelled the writing of the report.

References

[1] M.n. Event display 3D developed by LIP Laboratório de Instrumentação e Física Experimental de Partículas, *Event display 3D*, http://wminho.lip.pt/AugerVisualizer/v2/ (2019), [Online]

[2] P.A. Collaboration, Nucl. Instrum. Methods A798, 172 (2015)

[3] VISPA, *Visual Physics Analysis*, https://vispa.physik.rwth-aachen.de/server/ (2002), [Online]

[4] L.P. et al. For the Pierre Auger Collaboration, arXiv:1107.4804 (2011)

[5] A.A. et al., Phys. Rev. D 90 (2014)