

Gamma/hadron discrimination at PeV energies through the azimuthal fluctuations of air shower footprint at the ground

**A. Bakalová,^a R. Conceição,^{b,c,*} L. Gibilisco,^{b,c} V. Novotný,^{a,d} M. Pimenta,^{b,c}
B. Tomé^{b,c} and J. Vícha^a**

^a*Institute of Physics of the Czech Academy of Sciences, Prague, Czech Republic*

^b*LIP - Laboratório de Instrumentação e Física Experimental de Partículas, Lisbon, Portugal*

^c*Departamento de Física, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal*

^d*Charles University, Faculty of Mathematics and Physics, Institute of Particle and Nuclear Physics, Prague, Czech Republic*

E-mail: ruben@lip.pt

The identification of PeV gamma rays over the cosmic ray background requires an excellent discrimination capability, usually achieved through the measurement of the shower muon content. However, the deconvolution of secondary muons from the electromagnetic shower component is a difficult task that substantially increases the cost of the experiment. In this work, we propose a novel approach based on the analysis of the azimuthal fluctuations of the shower footprint at the ground measured by a low-density array of water Cherenkov detector units. Using a large statistical simulation sample, we demonstrate that it is possible to reach a gamma/hadron discrimination as high as the one achieved by counting muons.

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



*Speaker

1. Introduction

The recent detection of gamma rays with energies up to the PeV range [1, 2] has provided a new perspective into the extreme energy Universe. However, due to the low fluxes and the presence of high background charged cosmic rays, the detection of such gamma rays is only feasible at ground-based gamma-ray observatories with large surface areas (approximately equal to or greater than 1 km^2). These observatories must be able to effectively distinguish between hadron showers and gamma showers.

At high energies (above tens of TeV), it becomes possible to achieve high background rejection factors by analyzing the distribution of particles at the ground with respect to the distance from the shower core. This analysis involves identifying the presence of energetic sub-showers, examining the steepness, compactness, or bumpiness of the Lateral Distribution Function (LDF) [3–5], or, in the case of Cherenkov telescopes, observing differences in the longitudinal development of the shower [6, 7]. Additionally, when available, one of the most effective discriminators is the measurement of the number of muons reaching the ground [8–10]. Muon measurements can be accomplished by shielding the detectors with materials such as earth (e.g., [11–13]), water (e.g., [14], [15]), concrete, or other inert substances. Alternatively, muons can be detected by analyzing the differences in time and/or intensity of the collected signals in detectors equipped with multiple light sensors (e.g., [16, 17]). Ultimately, a global rejection factor should be achieved on the order of or greater than 10^4 .

In this work, we investigate the azimuthal non-uniformity of particle distributions at the ground by introducing a novel variable, C_k [18]. This quantity is computed in successive circular rings centered at the shower core, with a radius of r_k in the shower's transverse plane. By analyzing the C_k distribution as a function of r_k , we define a new discriminating variable between gamma rays and hadrons, denoted as LCm . This variable corresponds to the value of the logarithm of the C_k distribution at a given r_k .

2. The C_k variable

The new variable C_k , introduced in [18], is defined for each radial ring k as:

$$C_k = \frac{2}{n_k(n_k - 1)} \frac{1}{\langle S_k \rangle} \sum_{i=1}^{n_k-1} \sum_{j=i+1}^{n_k} (S_{ik} - S_{jk})^2, \quad (1)$$

Here, n_k is the number of active stations in ring k , $\langle S_k \rangle$ is the mean signal in the stations of ring k , and S_{ik} and S_{jk} represent the signals observed at stations i and j in ring k , respectively. The term $\frac{2}{n_k(n_k-1)}$ is the inverse of the number of two-combinations for n_k stations, $\binom{n_k}{2}$. In this study, each circular ring k is centered around the shower core position and has a width of 10 m.

Essentially, the variable C_k represents the average sum of the squared differences between the collected signals in any pair of two stations within ring k , normalized by the mean signal observed in a single station within the same ring. This normalization is intended to minimize potential correlations between the dependence of C_k on r_k and the lateral distribution function.

For perfectly uniform azimuthal distributions, C_k is designed to be zero, while otherwise, it should be greater than zero. Due to the presence of hadronic sub-showers, the ground pattern of

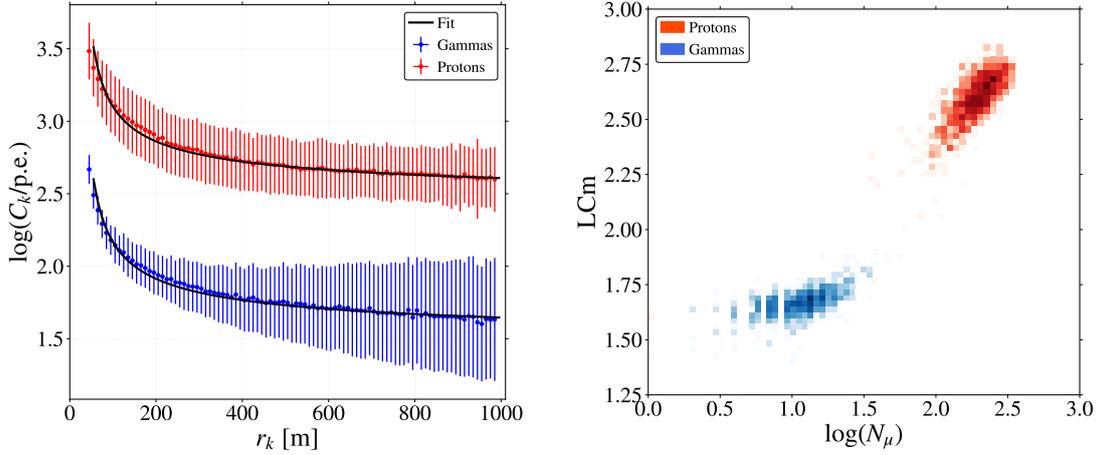


Figure 1: (left) Mean value of $\log(C_k)$ as a function of r_k for gamma showers with a primary energy $\in [100; 160]$ TeV (blue, lower points) and for proton showers (red, upper points) with similar energies at the ground considering an array $FF = 100\%$. The errors bars are the RMS of the $\log(C_k)$ distributions in each r_k bin. The full lines represents the best fit using the parametrisation expressed in equation 2. (right) LCm vs $\log(N_\mu)$ distributions for gamma showers with a primary energy of about 100 TeV (blueish histogram, to the left) and for proton showers (reddish histogram, to the right) with similar energies at ground, considering an array with $FF = 0.12$ and a radius of 1000 m and placing the shower cores at the centre of the array.

a proton shower is more complex than that of a gamma shower, which is predominantly governed by electromagnetic interactions. Therefore, C_k is expected to be higher for proton-induced showers compared to gamma showers of equivalent energies at the ground.

Such behavior is observed for r_k values greater than 40 m in Figure 1 (see reference [18] for details on the simulation). The data points of this figure are the mean values of the distributions of $\log(C_k)$ plotted as a function of r_k for gamma showers (represented by blue points) and proton showers (represented by red points). Both gamma and proton primaries have an energy of around 100 TeV, and the shower is sampled at the ground with water-Cherenkov detector units covering 12% of the array area (fill factor, $FF = 12\%$). The energy range for gamma-induced showers is 100 TeV to 160 TeV, while proton-induced showers were chosen from a broader energy range such that both have similar energies at the ground.

The overall behavior of the $\log(C_k)$ distributions in the region $r_k > 40$ m can be effectively described by the following parametrization (shown as solid lines in Figure 1):

$$\log(C_k) = a + \frac{b}{\log\left(\frac{r_k}{40\text{m}}\right) + 1}, \quad (2)$$

where a and b are real parameters that vary for each event. In the limit as r_k approaches 40 m, $\log(C_{k40}) = a + b$, while in the limit as r_k goes to infinity, $\log(C_{k \rightarrow \infty})$ approaches a .

The values a and b for each event are determined by fitting the corresponding $\log(C_k)$ distributions using the above parametrization. The root mean square (RMS) of the $\log(C_k)$ distributions in each r_k bin is used as the error for the fitting procedure. These RMS values are also displayed in Figure 1.

The overall quality of the fit is good, with a reduced chi-square value of approximately $\chi^2/n.d.f. \sim 1$. A very small fraction of events (less than 1%) exhibit $\chi^2/n.d.f. > 2$, which corresponds to a tail in the distribution. The analysis of these tail events is beyond the scope of the present article, and thus these events were excluded from further consideration. However, it is worth noting that this type of analysis has the potential to identify events characterized by extreme fluctuations in the shower development, such as the so-called double-bang events.

In Figure 1 (right), the correlation between the LCm distribution and the number of muons (N_μ) hitting the array stations located at a distance greater than 40 m from the shower core is presented for the 100 TeV energy bin. The blue points represent gamma showers, while the red points represent proton-induced showers. A uniform array with a fill factor of 12% and a radius of 1000 m was considered, with the shower cores placed at the center of the array.

A clear and nearly linear correlation between LCm and $\log(N_\mu)$ is observed for $N_\mu \gtrsim 15$. Hence, based on the available statistics, N_μ and LCm yield equivalent background rejection factors for efficient gamma event identification at an energy of 100 TeV. Notably, it was observed that if the muon information is discarded, the separation with LCm is still possible. This non-trivial statement indicates that the LCm variable is indeed sensitive to the showers' sub-cluster structure.

It is also important to note that while these distributions are illustrated for an energy bin, the same correlation exists when analyzing showers with fixed energy. Such confirms that the correlation arises from the intrinsic features of the showers rather than the primary energy alone.

3. Simulations to explore the LCm and N_μ tails

To statistically improve the considerations outlined in the previous section, a sample of 10^6 proton-induced showers with energies ranging from 1 to 2 PeV was generated using CORSIKA (version 7.7410) [19]. The simulation employed the UrQMD [20, 21] and QGSJet II-04 [22] hadronic interaction models for low and high-energy interactions, respectively. The showers were simulated with a fixed zenith angle of 20° relative to the vertical direction, while the azimuth angle was randomly selected from a uniform distribution. The secondary particles of the showers were collected at an altitude of 4700 m above sea level¹.

Following the approach described in reference [18], a two-dimensional histogram was created with cells having an area of approximately 12 m^2 to emulate a ground-based detector array with a fill factor of one (FF=1). Smaller fill factors were achieved by applying regular patterns as masks to the two-dimensional histogram. A bijective mapping between cells and water-Cherenkov Detector (WCD) stations was established. Therefore, the total signal in each station was obtained by summing the expected signals from particles hitting the corresponding histogram cell. The signal deposited by particles in a particular cell was calculated using parameterizations derived from a dedicated Geant4 simulation of the water-Cherenkov detector employed in this study [17]. The parameterizations were developed for muons, electrons and protons, representing the electromagnetic and hadronic shower components. The parameterizations accounted for the mean signal and the fluctuations arising from stochastic particle interactions, light collection, and the muon track lengths within the station.

¹This altitude was chosen as a reference height for the R&D studies conducted by the Southern Wide-field Gamma-ray Observatory [23].

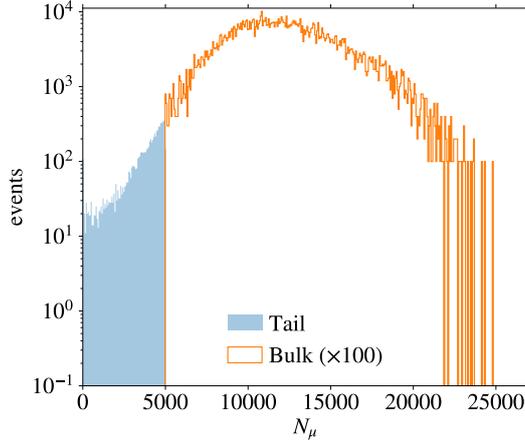


Figure 2: Distribution of the number of muons at the ground in the proton EAS: the blue filled bins correspond to the proton muon-depleted sample; the bins with orange contours are the proton reduced set, multiplying the mean number of the events in each bin by one hundred (the inverse of the sampling factor).

Additionally, a set of 10^3 gamma-induced showers was simulated under the same conditions as the proton showers, except for the fixed energy. The energy of the gamma showers was set to 1.6 PeV, which was determined to be the mean energy at which proton and gamma showers produce similar signal footprints at the ground. It should be noted that the purpose of this study is to provide a reference for comparing LCm with N_μ , rather than establishing absolute rejection factors.

From the original set of proton showers, which was too large to handle easily, two subsets were extracted. The first subset, referred to as the proton muon-depleted set (*tail*), consisted of all shower events below a fixed muon threshold scale. The second subset, known as the proton reduced set (*bulk*), included approximately one-hundredth of the events that were not selected for the first set, chosen randomly. The threshold for this decision was set at $N_\mu = 5000$, where N_μ represents the number of muons within one square kilometer. This threshold value, determined from a smaller shower sample of the order of 10^4 showers, ensures the selection of the 1% of showers with the lowest number of muons.

The purpose of the first subset is to preserve all proton events that are more likely to be identified as gamma candidates if the primary gamma/hadron discriminator is based on the number of muons detected at the ground. The second proton subset is utilized to reconstruct the complete shape of any distribution of interest. As an example, Figure 2 presents the distribution of the number of muons at the ground for the proton showers, combining both subsets. The bin-to-bin fluctuations in size reflect the statistics of the corresponding samples.

In this study, the experimental proxy for N_μ is the total signal recorded by the WCDs resulting from the passage of muons, denoted as S_μ . It is assumed that S_μ can be determined without any uncertainties other than those related to signal and track length fluctuations mentioned previously.

Figure 3 illustrates the cumulative distributions of the S_μ (left) and LCm (right) variables, assuming a detector array with a fill factor of 12.5%. The values corresponding to 90% gamma shower selection efficiency in each of these cumulative distributions are $S_\mu = 4.29 \times 10^{-4}$ and $LCm = 1.39 \times 10^{-4}$, respectively. Consequently, LCm exhibits a lower residual background of protons for selecting gamma showers, approximately three times lower than S_μ . The same study was conducted assuming a sparser array with FF= 1.4%. The proton selection efficiencies in this case were $S_\mu = 9.33 \times 10^{-4}$ and $LCm = 6.10 \times 10^{-4}$, indicating that LCm performs slightly better

as a discriminator, achieving approximately 50% improvement compared to S_μ .

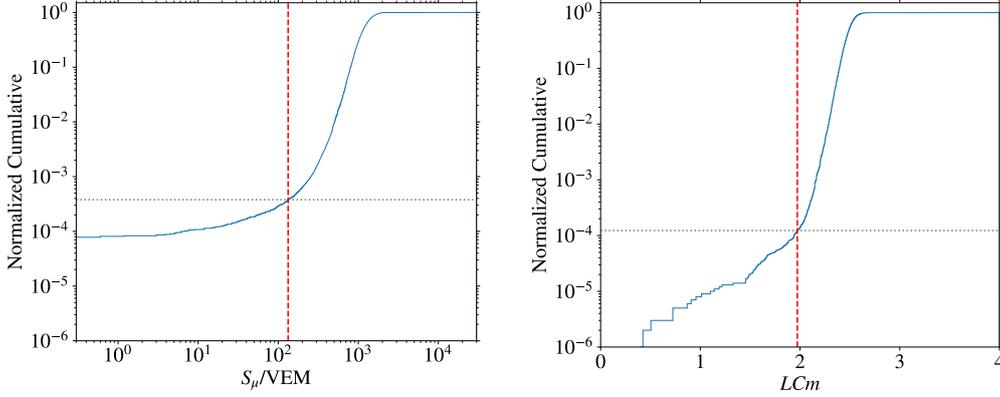


Figure 3: Cumulative distributions for the S_μ (left) and LCm (right) distribution for events in the reference proton set (proton tail + proton bulk renormalized to the total number of showers simulated). The red (dashed) lines define the values of S_μ and LCm for which the gamma set has a selection efficiency of 90%.

Figure 4 displays the observed correlation between LCm and S_μ for the analyzed samples, assuming a detector array with a fill factor of $FF= 12.5\%$. In the plot, shower events with $S_\mu = 0$ are situated at the far left, while low-quality events for extracting LCm are positioned at the top. The criterion for the latter decision was based on requiring that the azimuthal fluctuations of the radial profile, constructed using radial bins of 30 m, had more than two degrees of freedom for fitting LCm . It is evident from this plot that LCm and S_μ exhibit a strong correlation, even in the low- S_μ tails. Furthermore, the gray lines depicting the 90% efficiency of gamma shower selection demonstrate that LCm is more effective in identifying background showers compared to S_μ , consistent with the cumulative studies depicted in Figure 3.

4. Discussion and conclusions

The number of muons produced in a high-energy hadron-induced shower reaching the ground at a high altitude is approximately an order of magnitude greater than that produced in a gamma-induced shower with the same reconstructed energy. As a result, N_μ serves as an excellent discriminator between gamma and hadron showers, achieving rejection levels on the order of 10^{-4} at PeV energies. However, at these energies and altitudes, the number of electromagnetic component particles, such as EAS photons and electrons, reaching the ground is many orders of magnitude higher than the number of accompanying muons.

Directly counting muons requires the use of shielded detectors with inert materials such as earth, water, concrete, and iron. While effective, this strategy is considerably expensive to implement in large observatories spanning several square kilometers.

Alternatively, the LCm variable, which quantifies the azimuthal non-uniformity in the shower's ground pattern, has been found to exhibit a high correlation, in terms of mean values, with N_μ [18]. Furthermore, LCm can be implemented easily and at a lower cost.

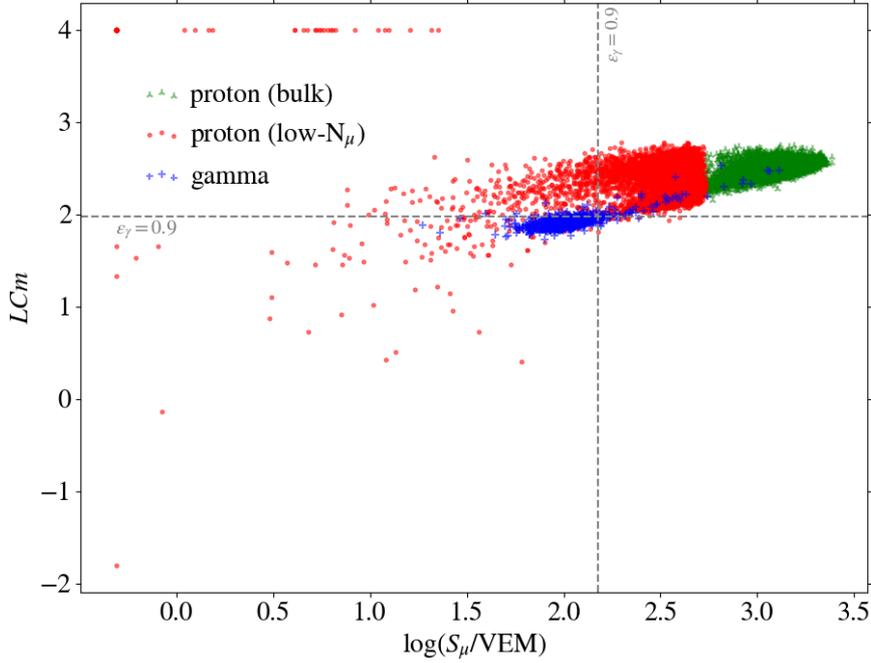


Figure 4: Correlation between $\log(S_\mu)$ and LCm for the muon-depleted (red), proton bulk (green) and gamma (blue) events. The dashed grey lines indicate the cuts on N_μ and LCm to select 90% for the gamma showers. The discrimination quantities were computed assuming a detector array with a fill factor of 12.5%.

In this study, a simulation strategy was developed to analyze the rare muon-depleted shower events, which constitute the primary background source for PeV gamma showers. It was demonstrated that both the S_μ and LCm variables maintain a strong correlation, leading to equivalent rejection levels for both variables. This conclusion holds true across various array fill factors, ranging from 50% down to 1.4%.

The findings presented in this study offer a unique opportunity to construct a cost-effective gamma-ray observatory based on water Cherenkov detectors, capable of covering a wide energy range from hundreds of GeV to tens of PeV. In this optimized array, the fill factor and the number of PMTs in each WCD station would be higher in the inner regions and gradually decrease towards the outer regions, allowing for the sampling of the shower's calorimetric footprint.

Acknowledgements

This work has been financed by national funds through FCT - Fundação para a Ciência e a Tecnologia, I.P., under project PTDC/FIS-PAR/4300/2020.

References

- [1] M. Amenomori et al. *Phys. Rev. Lett.* **126** (2021) 141101.
- [2] Cao, Zhen and et. al *Nature* **594** (May, 2021) 33–36.

- [3] K. Greisen *Annual Review of Nuclear Science* **10** (1960), no. 1 63–108.
- [4] Abeysekara and et. al *The Astrophysical Journal* **843** (Jun, 2017) 39.
- [5] P. Assis et al. *Astropart. Phys.* **99** (2018) 34–42, [[arXiv:1607.0305](https://arxiv.org/abs/1607.0305)]. [Erratum: *Astropart.Phys.* 101, 36–36 (2018)].
- [6] A. M. Hillas in *19th Intern. Cosmic Ray Conf-Vol. 3*, no. OG-9.5-3, 1985.
- [7] T. C. Weekes et al. *Astrophys. J.* **342** (1989) 379–395.
- [8] P. R. Blake and W. F. Nash *Journal of Physics G: Nuclear and Particle Physics* **21** (jan, 1995) 129–143.
- [9] N. Hayashida and et. al a *Journal of Physics G: Nuclear and Particle Physics* **21** (aug, 1995) 1101–1119.
- [10] W. Apel and et. al *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **620** (2010), no. 2 202–216.
- [11] A. Borione and et. al *Nuclear Instruments and Methods in Physics Research A* **346** (July, 1994) 329–352.
- [12] A. Aab and et. al *Journal of Instrumentation* **16** (apr, 2021) P04003.
- [13] **LHAASO** Collaboration, X. Zuo et al. *Nucl. Instrum. Meth. A* **789** (2015) 143–149.
- [14] P. Abreu et al. *Eur. Phys. J. C* **78** (2018), no. 4 333, [[arXiv:1712.0768](https://arxiv.org/abs/1712.0768)].
- [15] A. Letessier-Selvon, P. Billoir, M. Blanco, I. C. Mariş, and M. Settimo *Nucl. Instrum. Meth. A* **767** (2014) 41–49, [[arXiv:1405.5699](https://arxiv.org/abs/1405.5699)].
- [16] B. S. González and et. al *Journal of Physics: Conference Series* **1603** (sep, 2020) 012024.
- [17] P. Assis et al. *Eur. Phys. J. C* **82** (2022), no. 10 899, [[arXiv:2203.0878](https://arxiv.org/abs/2203.0878)].
- [18] R. Conceição, L. Gibilisco, M. Pimenta, and B. Tomé *JCAP* **10** (2022) 086, [[arXiv:2204.1233](https://arxiv.org/abs/2204.1233)].
- [19] D. Heck, J. Knapp, J. Capdevielle, G. Schatz, and T. Thouw *Report FZKA* **6019** (1998).
- [20] S. A. Bass et al. *Prog. Part. Nucl. Phys.* **41** (1998) 255–369, [[nucl-th/9803035](https://arxiv.org/abs/nuc1-th/9803035)].
- [21] M. Bleicher et al. *J. Phys. G* **25** (1999) 1859–1896, [[hep-ph/9909407](https://arxiv.org/abs/hep-ph/9909407)].
- [22] S. Ostapchenko *Phys. Rev. D* **83** (2011) 014018, [[arXiv:1010.1869](https://arxiv.org/abs/1010.1869)].
- [23] H. Schoorlemmer, R. Conceição, and A. J. Smith *PoS ICRC2021* (2021) 903.