Experimental Astronomy (2025) 59:13
https://doi.org/10.1007/s10686-025-09981-z

RESEARCH

γ/hadron discrimination by analysis of the muon lateral distribution and the ALPAQUITA array

M. Anzorena[1] · E. de la Fuente[2] · K. Fujita[1] · R. Garcia[1] · K. Goto[3] · Y. Hayashi[4] · K. Hibino[5] · N. Hotta[6] · G. Imaizumi[1] · A. Jimenez-Meza[7] · Y. Katayose[8] · C. Kato[9] · S. Kato[1] · T. Kawashima[1] · K. Kawata[1] · T. Koi[10] · H. Kojima[11] · T. Makishima[8] · Y. Masuda[9] · S. Matsuhashi[8] · M. Matsumoto[9] · R. Mayta[12,13] · P. Miranda[14] · A. Mizuno[1] · K. Munakata[9] · Y. Nakamura[1] · M. Nishizawa[15] · Y. Noguchi[8] · S. Ogio[1] · M. Ohnishi[1] · S. Okukawa[8] · A. Oshima[3,10] · M. Raljevic[14] · H. Rivera[14] · T. Saito[16] · T. Sako[1] · T.K. Sako[17] · T. Shibasaki[18] · S. Shibata[11] · A. Shiomi[18] · M. Subieta[14] · F. Sugimoto[1] · N. Tajima[19] · W. Takano[5] · M. Takita[1] · Y. Tameda[20] · K. Tanaka[21] · R. Ticona[14] · I. Toledano-Juarez[22] · H. Tsuchiya[23] · Y. Tsunesada[12,13] · S. Udo[5] · R. Usui[8] · G. Yamagishi[8] · K. Yamazaki[10] · Y. Yokoe[1]

Received: 9 August 2024 / Accepted: 21 January 2025

Abstract

A new method using the muon lateral distribution and an underground muon detector to achieve high discrimination power against hadrons is presented. The method is designed to be applied in the Andes Large-area PArticle detector for Cosmic-ray physics and Astronomy (ALPACA) experiment in Bolivia. This new observatory in the Southern hemisphere has the goal of detecting >100 TeV γ rays in search for the origins of Galactic cosmic rays. The method uses the weighted sum of the lateral distribution of the muons detected by underground detectors to separate between air showers initiated by cosmic rays and γ rays. We evaluate the performance of the method through Monte Carlo simulations with CORSIKA and Geant4 and apply the analysis to the prototype of ALPACA, ALPAQUITA. With the application of this method in ALPAQUITA, we achieve an improvement of about 18 % in the energy range from 60 to 100 TeV over the estimated sensitivity using only the total number of muons.

Keywords Sub-PeV γ-ray astronomy · Gamma/hadron discrimination · Muon lateral distribution

1 Introduction

Sub-PeV γ-ray astronomy (>100 TeV) is a growing and challenging field, which promises to shed light over the sources of Galactic Cosmic rays (CR) and their acceleration

---

Extended author information available on the last page of the article mechanism. Cosmic rays (protons and heavy nuclei) below 4 PeV are considered to be produced in our Galaxy by Supernova remnants (SNR) [1, 2]. They arrive at the top of Earth's atmosphere isotropically, deflected on their way from the source by galactic magnetic fields. Neutrinos and γ rays are produced by CRs interacting with the interstellar medium, yet preserving the direction of the parent particles. Nevertheless, neutrinos are very difficult to detect due to their extremely small cross section. Under such circumstances, photons become excellent galactic probes; as they provide source-pointing capabilities and are easily detected through their interaction with matter [3, 4].

γ rays at the 100 TeV are produced by the decay of neutral pions, which in turn come from PeV hadronic interactions with the interstellar medium. In this process, about 10 % of the parent particle's energy is transferred to the secondary photons. Therefore, in order to prove the Galactic origin of cosmic rays at the knee (around 4 TeV) we need to observe γ rays above 100 TeV. Observations by satellite of SNRs showed spectral signatures that could be originated by neutral pion decay, pointing to a hadronic origin in the emission [5, 6]. Moreover, more than 10 sources beyond 100 TeV have been identified in the Northern hemisphere by the Tibet ASγ, HAWC and LHAASO experiments [7–9]. Despite such results, there is still no conclusive evidence regarding this matter and no observation in this energy range has been carried out in the Southern hemisphere.

The first challenge to overcome in the detection of the sub-PeV extensive air showers is the development of experiments covering extensive areas (on the order of 10^5 m^2) at high altitude regions. The high altitude location is needed to sample the maximum development of the shower, which is close to 600g cm^{-2} for the 100 TeV γ rays. The second challenge to address is the discrimination against the dominant cosmic-ray background. Cosmic rays above 100 TeV have an integral flux of 20.64 m^{-2}d^{-1}, meanwhile the integral flux of γ rays from Crab Nebula is 2.84×10^{-6} m^{-2} d^{-1} [7]. Considering a background reduction by a factor of 40 by observing events only in an angular window, to obtain a 5σ detection significance in one year of observation, we require a background rejection factor of 10^{-3} or more. To achieve this background rejection level some experiments study the spatial and energy deposition characteristics using the information registered by a surface array [10, 11]. Conversely, measuring the number of muons in the shower through underground detectors is also a powerful discriminator [12, 13]. Using machine learning to exploit correlations in the observables to achieve the classification is recently becoming an important trend in γ-ray observation [14–16].

In this paper, we present the development of a new method to achieve γ/hadron discrimination by the analysis of the muon lateral distribution detected with a large area underground detector. We design this method with the purpose of applying it to a new γ-ray observatory in the Southern Hemisphere, the ALPACA experiment. The method combines information from the total number of muons and their lateral spread into a single variable, which may be also optimized to maximize the sensitivity to point sources. Besides this, we also evaluate the sensitivity of the detector using this method and compare it with the case when we use the total number of muons as discriminator. The paper is summarized as follows: Section 2 introduces the ALPACA and ALPAQUITA experiments; Section 3 presents a general discussion of the properties of the muon lateral distribution in the energy range of interest, done through a Monte Carlo study with CORSIKA. In Section 4 we apply the developed technique to the case of ALPAQUITA and show the improvement in the sensitivity. Finally, Section 5 presents the conclusion of our work.

2 The ALPACA experiment

The ALPACA experiment is an hybrid array, composed of a surface air shower (AS) array and underground water Cherenkov muon detectors (MD). The AS array will cover and area of 83000 m^2 and comprises 401 plastic scintillators of 1 m×1 m×0.05 m each. The AS detectors in the array are arranged in a grid pattern with separation of 15 m between elements. The scintillators are installed inside a pyramidal metal box which reflects the light to 2 photomultiplier tubes (PMT) installed at the bottom. On top of each box, a lead plate of 5 mm thickness is used to convert secondary γ rays in the air shower into $e^{\pm}$ pairs, effectively improving the angular resolution. By using the information from the particle density and relative timing registered by each PMT, we may reconstruct the energy and incoming direction of the primary particle.

The MD array of ALPACA is composed of four 900 m^2 underground pools. One MD is divided into 16 cells of 7.5 m×7.5 m, with one 50 cm (50 cm ∅) PMT installed on the ceiling of every cell. The PMTs collect the Cherenkov light produced by particles reaching the underground. Since the MD array is set up beneath a 2.0 m thick layer soil, the signal of the MD is produced by high energy muons (>1.2 GeV) from the air shower. Therefore, by observing in coincidence the air shower products with the AS and the MD array; high sensitivity to γ rays is achieved, as γ-ray induced showers are poor in muons in comparison to CR ones.

The ALPACA array is currently under construction in the Chacaltaya plateau (16°23’**S**, 68°8’**W**) in Bolivia. This location is an ideal place for a cosmic-ray observatory, due to the large and flat area and high altitude of 4740 m (572 $\mathrm{g\,cm}^{-2}$). In 2022, the prototype of ALPACA, ALPAQUITA, was completed and started observations. ALPAQUITA consists of 97 AS detectors (1/4 of the full ALPACA) over an area of 18 450 m^2. Figure 1 shows a schematic diagram of the ALPACA array, with ALPAQUITA located in the lower-left corner (shaded area). The construction of the underground MDs will start soon. A detailed Monte Carlo simulation study and performance study of ALPAQUITA, including one MD pool at the center of the array, is reported in [17]. According to this study, ALPAQUITA will be capable of detecting five γ-ray sources above 10 TeV with high statistical significance ($> 5\sigma$) in one year of observation.

3 Analysis of the Muon lateral distribution by MC simulation

In order to extract useful features from the Muon lateral distribution (MLD) to use it as discrimination method, we adapted a similar method to the one proposed in [18], as this avoids the analytical modeling of the MLD. For this matter we simulated the development of γ/CR ray induced air showers using CORSIKA 7.7410 [19]. In both

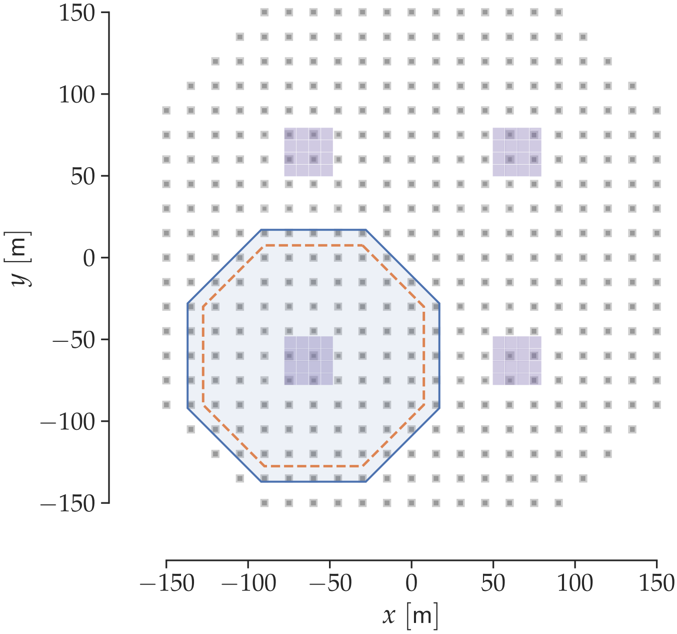

Fig. 1 Schematic diagram of the full ALPACA array with ALPAQUITA+MD indicated by the blue shaded area

cases, the events are generated in the energy range from 1 to 1000 TeV with an uniform angular distribution in the zenith and azimuth direction. The zenith angle is limited to a maximum of 60°. The primary energy distribution of the generated γ rays follows a power-law with spectral index of 2. A total of 10×10^6 events are simulated until they reach the altitude of the ALPACA experimental site. For cosmic-ray events we use the model for chemical composition and energy spectrum proposed in [20]. We use FLUKA [21] and EPOS-LHC [22] as low and high-energy hadronic interaction models. In this case, the total number of simulated events is 100×10^6.

From the output of the simulation we only select the muons which have $E_k \geq 1.0\,\mathrm{GeV}$ and lie inside an area of $1000{\times}1000\,\mathrm{m}$. After this, we divide the area into detection elements (DE) of $10\,\mathrm{m}{\times}10\,\mathrm{m}$ and count the number of muons N_μ inside each element. Particle density distributions are then constructed, separating the γ and cosmic-ray sets into 16 bins in the energy range (4 bins per decade) and 4 bins in the zenith angle range. The results from this analysis are shown in Fig. 2. The left panel shows the density distributions from γ-ray initiated showers in the energy range from 300 to 500 TeV and zenith angles between 15 to 30°, meanwhile the right panel shows the case for cosmic-ray showers. In both cases, the distributions are normalized by the total number of muons, and then to a maximum value of one. This is done with the purpose of examining only the spatial features of the MLD. The color scale in the Figure represents the density of the distributions: darker color equals higher density and lighter color equals low density.

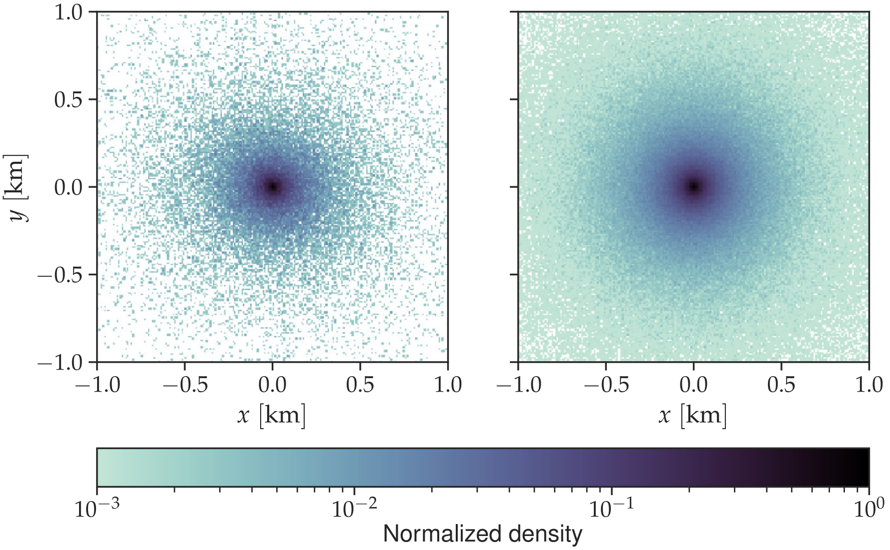

Fig. 2 Particle density maps observed at an altitude of 4740 m, for γ-ray initiated showers (left panel) and hadrons (right panel). The simulated air showers have energies in the range between 300 to 500 TeV and zenith angles between 15 to 30°

Taking into account both distributions in Fig. 2, it is clear that muons generated in the hadronic showers are distributed more broadly than in the γ-ray counterpart. To make the difference between the two distributions clearer, the normalized MLD is presented on Fig. 3. The blue line represents the MLD for γ-ray showers and the orange line is the case for CRs. The normalization procedure and energy/angular range presented are the same as the ones used on Fig. 2. From this we can see that the MLD from γ-ray initiated showers is steeper than the one for CRs, which agrees with the similar situation comparing MLDs from different mass number cosmic rays [23]. Consequently, by taking the sum of the number muons N_j at each DE j, weighted by their distance R_j from the shower axis, it is possible to distinguish between the two classes of showers. This procedure introduces a new observable m_k which can be written using 1:

$$m_k = \sum_j N_j \times \left(\frac{R_j}{R_n}\right)^k \tag{1}$$

where R_n is a normalization factor equal to 125 m and k is a free parameter used for the optimization of the classification process. In general, m_k is dependent on both the energy and zenith angle of the primary particles and therefore, the parameter k is useful to maximize the separation between primaries according to different conditions. As mentioned before, this method is similar to the one expressed in [18], which is in turn inspired on an observable developed for composition studies [24]. Lastly, is important to notice that when $k = 0$, Equation 1 is equivalent to the total number of muons in the shower ($\sum N_\mu$). We will use the analysis with $k = 0$ as our

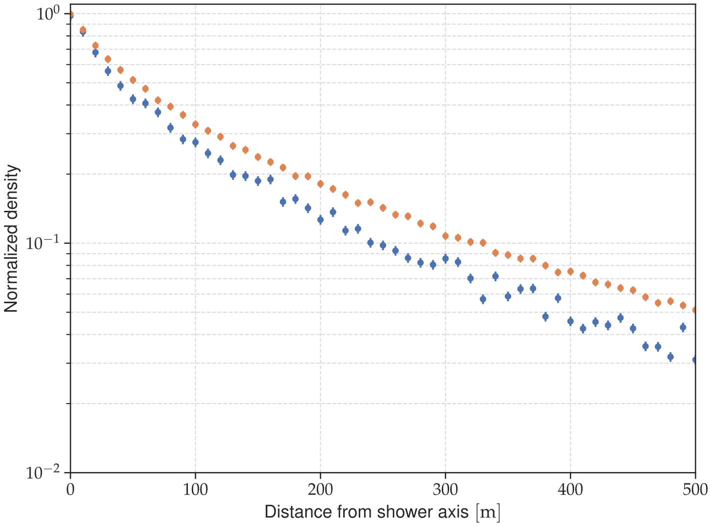

Fig. 3 Normalized muon lateral distribution at an altitude of 4740 m, for γ-ray initiated showers (blue) and hadrons (orange). The simulated air showers have energies in the range between 300 to 500 TeV and zenith angles between 15 to 30°

reference, since it constitutes the main method for γ/hadron discrimination in the ALPACA/ALPAQUITA experiment [17].

Next, we will assess the performance of the classification using m_k. We perform this analysis by means of the separation power η defined in Eq. 2:

$$\eta = \frac{\left| m_k^{\gamma,50} - m_k^{CR,50} \right|}{\sqrt{\sigma^2\left(m_k^{\gamma}\right) + \sigma^2\left(m_k^{CR}\right)}} \tag{2}$$

In this Equation, $m_k^{\gamma,50}$ ($m_k^{CR,50}$) represents the median of the m_k distribution and $\sigma\left(m_k^{\gamma}\right)$ ($\sigma\left(m_k^{CR}\right)$) is a modified standard deviation considering the asymmetry of the γ/CRs distributions [24]. The η parameter is capable of measuring the overlapping between two distributions and therefore it may be used to evaluate the improvement in the classification for different values of k. However, a higher separation power does not directly translate into a higher sensitivity, given that sensitivity is dependent of both class separation and the number of events on each class.

The normalized $m_{k=0}$ distribution for 100 TeV showers is shown on the left panel of Fig. 4. The filled histogram corresponds to γ-ray events, whereas CRs are represented by the solid line. The right panel of Fig. 4 shows the distribution of $m_{k=0}$ against the primary energy of the particles. The color scale represents the density of the distributions, that is: darker colors correspond to higher density regions and lighter represent lower density. The distribution depicted on purple is for γ-like showers

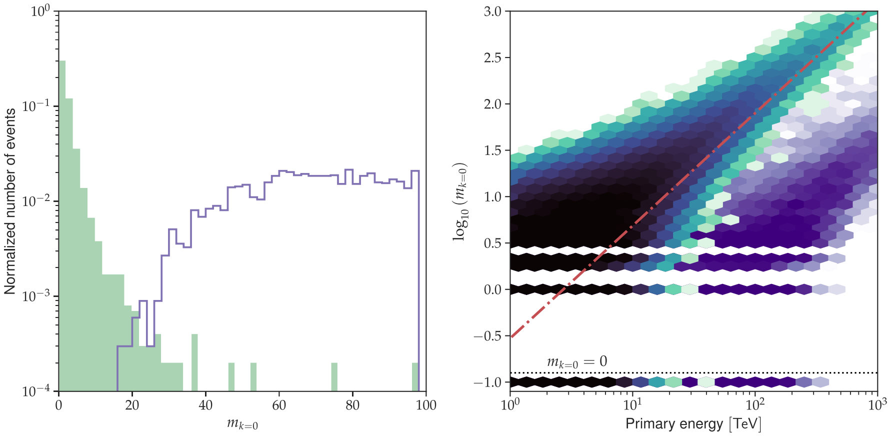

Fig. 4 Left panel: normalized distribution of $m_{k=0}$ for 100 TeV showers. The filled histogram corresponds to γ rays and the solid line to CRs. Right panel: distribution of $m_{k=0}$ (total number of muons) against primary energy for air showers induced by γ ray (right distribution) and CR (left distribution). The red dashed line represents the optimum decision criteria. Events accumulated at −1.0, are events that does not register any muon in the MC simulation

meanwhile the hadron-like showers are depicted in green-blue. The showers that do not produce any muon are accumulated in bin −1.0 on the Figure (below the dashed line). Considering both panels, we can see that CR initiated showers tend to produce a larger number of muons than γ-like showers, which in general are observed to be poor in muons. The red dashed line in the right panel optimum represents the decision boundary between the two groups (the details of the method to obtain the decision boundary are presented in Section 4).

After obtaining distributions for values of $k = 0$, 1 and 2, we estimate the separation power for each energy bin. The left panel in Fig. 5 shows the separation power as a function of primary energy for $m_{k=0}$. In general, the separation power increases with increasing energy, except for the region between 20 to 40 TeV. By inspecting each element on Eq. 2, we found that the behavior in this energy region is produced by the sudden change in the γ-ray distribution, which changes from being dominated by muon-less showers (located at bin −1.0 on the right panel of Fig. 4) to showers with a small number of muons. Furthermore, this characteristic is also present when using other values of k, although the location of the breaking point can be shifted to higher or lower energies. The ratio (performance ratio) between the separation power for $k = 1$ ($k = 2$) and $k = 0$ is shown on the right panel of the Fig. 5. By considering this results, we can conclude that the discrimination against the CR background is improved 2 or 4 times when we use the MLD for classification, in comparison with the case when only using the total number of muons. However, this should be interpreted as an ideal case, because several factors relating to the development of the air showers in the atmosphere and physical limitations of the detector impact the real separation power.

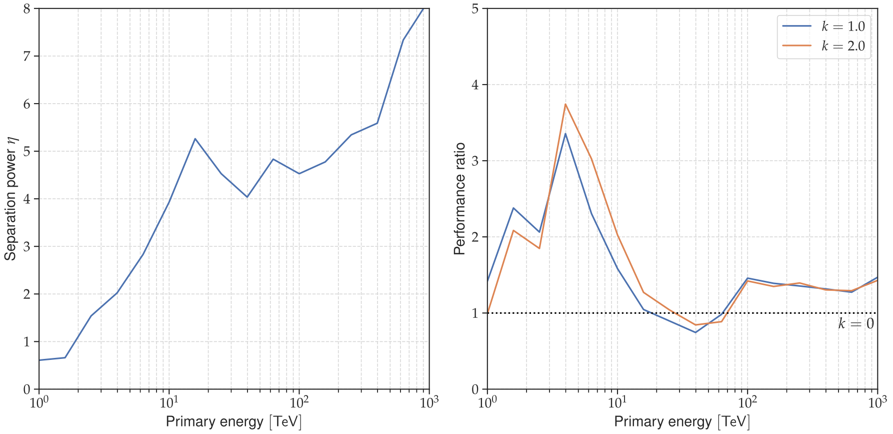

Fig. 5 Left panel: separation power as a function of energy for $k = 0$ (total number of muons). Right panel: ratio between MLD with different values of k and $k = 0$

4 Analysis of the Muon lateral distribution with ALPAQUITA+MD

4.1 MC simulation of ALPAQUITA+MD

In this section we will study the performance of the MLD method using the MC simulation of the ALPAQUITA detector plus one MD (ALPAQUITA+MD is shown inside the shaded area on Fig 1). The parameters included in this simulation are described in detail in [17]. Regarding the generation of air shower events with CORSIKA, the main difference with the analysis in Section 3 is that the events are simulated coming from a hypothetical point source along the path of RX J1713.7-3946 [25, 26]. Both CR and γ-ray events are generated in the energy range of 0.3 to 10 000 TeV; following the same trajectory in the sky, with a minimum zenith angle of 23.4°and maximum of 60.0°. In total 40×10^6 γ-ray and 900×10^6 CR events are simulated. The core of the showers is distributed randomly within a 300 m radius from the center of the AS detector.

The response of the detector to the secondary particles from the shower is simulated using Geant4 10.07p02 [27]. Using the MC we record the position, energy deposit and relative timing from the particles interacting with each of the AS detectors. To convert from energy deposits into particle densities ρ (particles/m^2) we use the single particle peak defined as 9.4 MeV. Considering this, a trigger event is defined as any event where 4 or more scintillators register ≥ 0.5 particles/m^2 within a time window of 600 ns. This definition is the same as the one used in the Tibet ASγ experiment [12].

An schematic diagram of the underground MD pool is shown in Fig. 6. Whenever an air shower triggers the AS array, the secondary particles that survive until reaching underground emit Cherenkov light in the MD. Then, light is reflected diffusely in the walls of the cells of the MD. In the MC simulation we count the number of photoelectrons (N_{pe}) that hit each 50 cm ∅ PMTs and also register their arrival time and position. The number of photoelectrons is sensitive to characteristics of the light

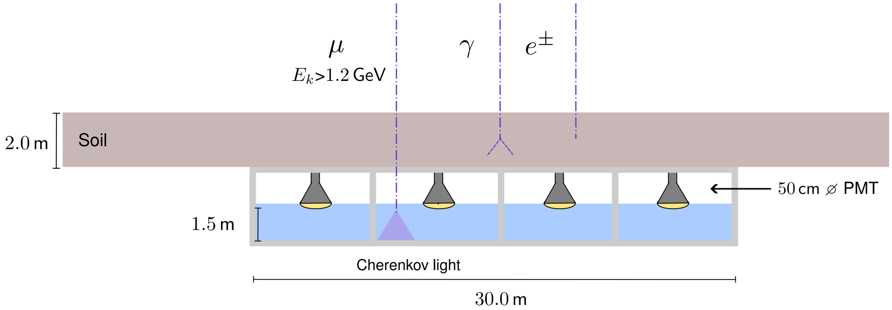

Fig. 6 Schematic diagram of one underground MD pool in the ALPACA experiment and principle of function

transport in the cells, the PMTs efficiency (quantum and light collection) and the geomagnetic field at the location, as well as saturation effects. All effects have been included in our simulation [12]. The background noise in the MD array is dominated by muons generated by lower energy cosmic rays. We include the effect of the accidental hits in the MD by adding to the N_{pe} signal a Poisson distribution with an expected value λ_{md} = 0.011 muons/cell. The threshold for the detection in a MD cell is 0.1 muons.

After simulating the response of ALPAQUITA+MD to the air showers with Geant4, we reconstruct the events to obtain the estimate arrival direction, core position, sum of particle densities from all detectors $\sum \rho$ in the AS array and the total number of photoelectrons $\sum N_{pe}$ in the MD. Figure 7 shows examples of a simulated γ ray (top panels) and CRs events (bottom panels). In both cases, the left panels show the response of the AS array and right panels show the underground MD. In the left panel the color scale represents the relative time of the particles in the shower, which is used in the reconstruction routine to estimate the arrival direction. The red arrow in the panel points to the estimated shower core and arrival direction. The size of the circles represents the particle density measured by each detector.

To estimate the energy of primary particles we use the correlation between primary energy and the particle densities detected with the AS array. For this purpose, $\sum \rho$ is defined as the sum of the density of detected particles ρ over all the hit detectors except for the detector that records the largest density. Considering the largest signal value is greatly affected by air shower development fluctuations, removing this value from the total density helps to improve the energy resolution. The conversion function between $\sum \rho$ and primary energy for gamma-ray events is presented in [17]. Using this definition we achieve an energy resolution of ±50 % in the 10 TeV energy range and ±20% for 100 TeV.

For the response of the MD (right panel Fig. 7) the color scale represents the number of muons (N_{μ}) in each cell. To obtain the number of muons from the values of N_{pe} we use the definition of single particle in the MD, which is defined as N_{SP} = 24.0 pe. In addition to the trigger condition previously mentioned, the criteria used in the reconstruction of the events is summarized as follows:

- At least 4 scintillator counters must register a density of 0.8 particles/m^2 or more.

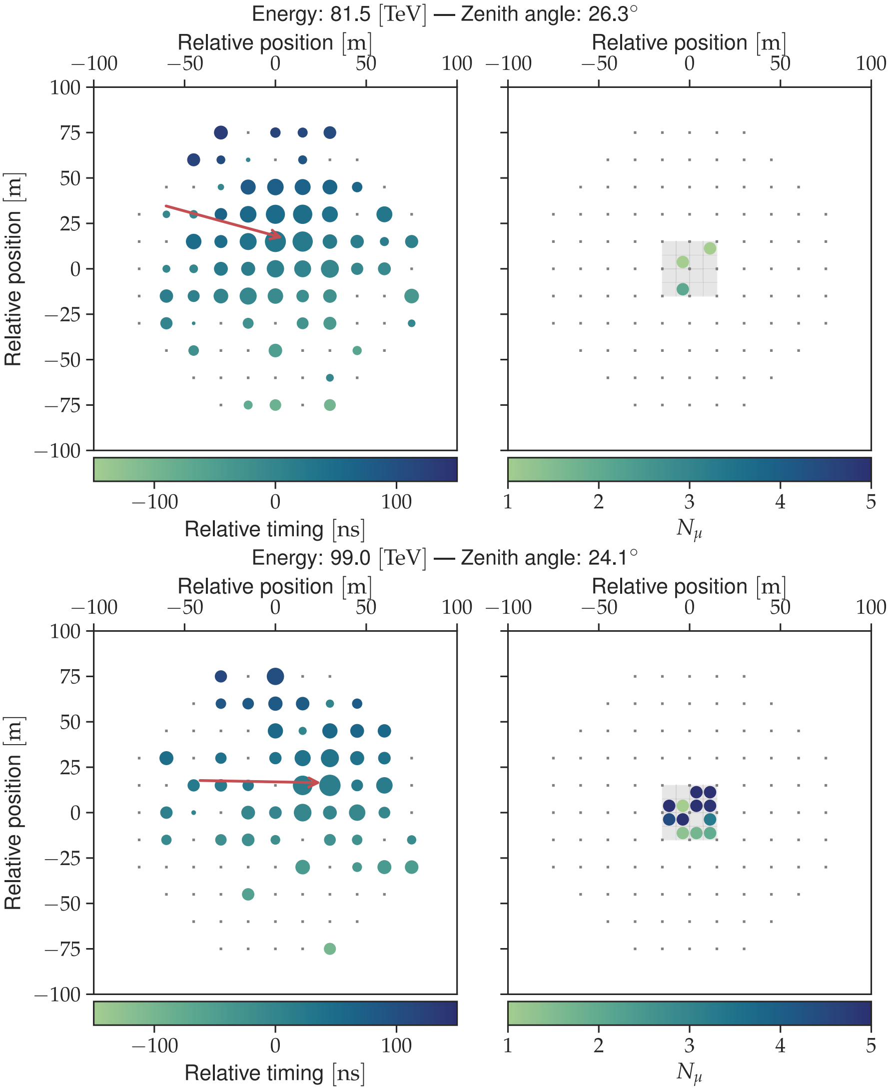

Fig. 7 Event shower maps from the MC simulation and their reconstructed parameters. Top panels correspond to a γ-ray event and the bottom panels to a CR one. The left panels show the response from the AS array. The size and color of each circle represent the particle density and the relative timing in each detector, respectively. The squares indicate positions of each detector. The arrow head and direction indicate the air shower core position and arrival direction, respectively. The right panels show the response of the MD and the color scale represents the N_μ in each cell

- Three out of the four detectors with the largest particle densities should be inside the inner area of the detector (area enclosed by the dashed line in Fig. 1).
- The residual error (χ) in reconstructing the arrival direction of the shower by means of the least-squares method must be less than 1 m.

Considering the case of point-source analysis, we imposed two further conditions: reconstructed zenith angles are less than 40°and the reconstructed direction lies inside a window of angular radius equal to $5.8^\circ \times \left(\sum \rho / \mathrm{m}^{-2}\right)^{-1/2}$. Furthermore, the angular radius is set to a maximum value of 1.5°when $\sum \rho < 15\,\mathrm{m}^{-2}$ and a minimum of 0.5° when $\sum \rho > 135\,\mathrm{m}^{-2}$.

Following the event selection we will look for the optimum classification method using the distributions of m_k against $\sum \rho$ for the two types of showers. Figure 8 shows an example for the case of $m_{k=0}$ (total number of muons). The distribution for the case of γ-ray showers is shown on the left panel of the Figure and the distribution for CR is shown on the right panel. In the upper horizontal axis on both panels, the equivalent γ-ray energy of the shower E_γ is presented. In this context, the equivalent γ-ray energy is the representative energy of the $\sum \rho$-bin, obtained by taking the weighted average of the logarithm of primary energies. The definition of the weight is the same as the one in Eq. 3, which takes into account the spectral index of the γ-ray source. Regarding Fig. 8, the weight factor is employed to change the energy distribution of the events from −2.0 (used in CORSIKA) to −2.62 (Crab-like source). From the distribution in the right panel we found, that for energies < 40 TeV, around 50 % of the γ-ray events do not produce signal in the MD ($\sum N_\mu = 0.01$ muons). Further inspection of these events shows that the majority comes from contamination of background muons.

Moreover, high energy electromagnetic components (EM) in the region close to the shower core may also penetrate the soil overburden and contaminate the signal in the pool. As EM components from the shower are distributed in space more compactly than muons, this effect can be lessen by excluding in the analysis of the MLD the signals close to the shower core. To estimate this effect, we record the signal generated by the EM component in the pool and then obtain the *punch-through* coefficient P_{em} by the ratio of the EM signal against the full signal (S_{em}/S_{tot}). The calculation of this parameter is done for all the events that pass the analysis conditions previously mentioned.

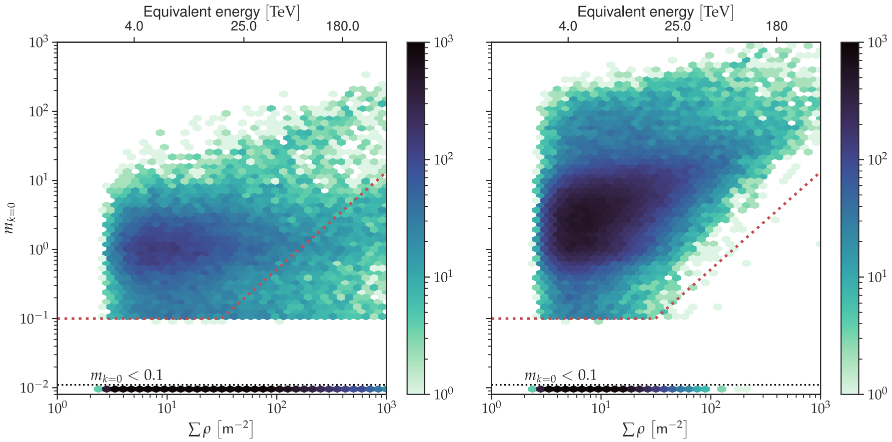

Fig. 8 Distribution of the $m_{k=0}$ ($\sum N_\mu$) against particle density $\sum \rho$ for different types of showers. Distribution in the left corresponds to γ-ray events and the one on the right is for CR. The red dashed line represents the optimum decision boundary. Events with $m_{k=0} < 0.1$ are accumulated at the bin 10^{-2}. The scale in the upper horizontal axis represents the equivalent γ-ray energy

The P_{em} coefficient as a function of the distance from the shower axis is shown on Fig. 9. The panel on the left corresponds to γ-ray showers with $25\,\mathrm{m}^{-2} < \sum\rho < 32\,\mathrm{m}^{-2}$ (γ-ray equivalent energy of 14 TeV) and the panel on the right of showers in the range $60\,\mathrm{m}^{-2} < \sum\rho < 100\,\mathrm{m}^{-2}$ (equivalent energy of 30 TeV). In both cases we can see that the effect of the contamination from the EM component is higher for distances from the shower axis $R < 20\,\mathrm{m}$ with a maximum value of 25%. As expected, the punch-through effect falls rapidly for distances far from the shower axis. The larger value at distance $R = 70\,\mathrm{m}$ (including error bar) comes from the physical size of ALPAQUITA and the limited statistics. Taking these results into account, we conclude that the contamination from the electromagnetic component is negligible and we did not exclude the events with core positions inside the area of the MD pool since this reduces the total number of analyzed events.

4.2 Point-source sensitivity using the muon lateral distribution

In order to estimate the sensitivity to point-sources, we first need to establish the optimal decision boundaries for each k. We perform this analysis by using the Q-factor (which may be interpreted as a signal noise ratio) defined as: $S/\sqrt{S+N}$, where S represents the expected number of γ-ray events in one year of observation and N represents expected number of CR events. Taking into account each energy bin, we search for the value of m_k which maximizes the Q-factor in that region and then define an optimum decision boundary between γ and hadron-like events. The determination of the decision boundaries using the Q-factor distributions first requires the normalization of the number of events in the simulation. For the case of γ-ray events, we suppose a hypothetical source following the path of RX J1713.7-3946, but with the spectral characteristics of Crab Nebula: spectral index of 2.62 and integral flux above 0.3 TeV: $F_{crab} = 1.228 \times 10^{-6}\,\mathrm{m}^{-2}\,s^{-1}$ [28]. On the other hand, for CR events we use

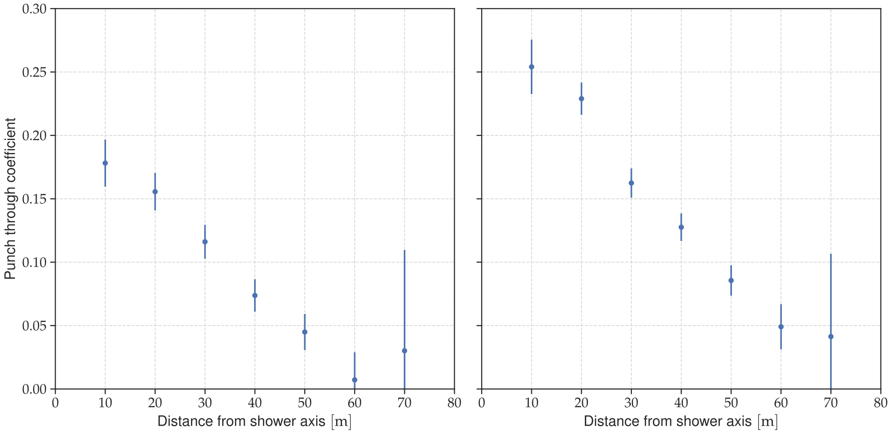

Fig. 9 Electromagnetic punch through coefficient for γ-ray initiated showers at different energy ranges. The left panel corresponds to showers with $25\,\mathrm{m}^{-2} < \sum\rho < 32\,\mathrm{m}^{-2} (E_\gamma = 14\,\mathrm{TeV})$. The right panel corresponds to showers in the range $60\,\mathrm{m}^{-2} < \sum\rho < 100\,\mathrm{m}^{-2}$ ($E_\gamma = 40\,\mathrm{TeV}$).

the model in [20] to obtain the all particle differential spectrum $\mathrm{d}J/\mathrm{d}E$. After integration, we obtain an integral flux above 0.3 TeV equal to $F_{CR} = 1.022\,\mathrm{m}^{-2}\,\mathrm{s}^{-1}\,\mathrm{sr}^{-1}$. Using the integral fluxes we normalize the total number of events $N_{\gamma,sim}$ ($N_{CR,sim}$) to the expected number of events during one year of observation N_{γ}^{1yr} (N_{CR}^{1yr}) using Eq. 3:

$$N_{\gamma}^{1yr}[j] = F_{crab} \cdot S_{sim} \cdot T_{1yr,\theta<60^{\circ}} \frac{1}{N_{\gamma,sim}} \sum_{k=1}^{N_{\gamma}[j]} w_k \tag{3}$$

where $S_{sim} = 300 \times 300 \times \pi\ \mathrm{m}^2$ is the covered area in the MC simulation, $T_{1yr,\theta<60}$ is the total time of observation of the source from the site within a zenith angle < 60∘, $N_{\gamma}[j]$ represents the total number of events in the j-th $\sum\rho$ bin and w_k is the weight factor used to take into account the change of spectral index in the analysis (from 2 to 2.62). The process for normalizing the CR does not require changing the energy spectrum, but to consider the isotropic characteristics of this source. Hence, we employ the definition of the weight factor following the procedure described in [17].

After normalization, we search for the value m_k^{cut}, which is the threshold value yielding a maximum Q-factor in a given $\sum\rho$ bin. Afterwards, in order to compensate the statistical fluctuations, we fit Eq. 4 to determine the optimal decision boundary:

$$\log_{10}(m_k^{cut}) = \begin{cases} m_0 & \sum\rho \leq \left(\sum\rho\right)_0 \\ b \cdot \log_{10}\left(\sum\rho/\left(\sum\rho\right)_0\right) + m_0 & \sum\rho > \left(\sum\rho\right)_0 \end{cases} \tag{4}$$

From this, we obtained the 3 parameters needed to establish the decision boundary: m_0, b and $\left(\sum\rho\right)_0$. Figure 10 shows the values of m_k^{cut} giving a maximum Q-factor for the case of $k = 0$ and the resulting optimum decision boundary (orange dashed line). It is important to stress that the general behavior of the optimum decision boundary is determined by both the characteristics of the m_k γ-ray and CR distributions on each $\sum\rho$ bin. However, the CR distribution has a positive gradient with respect to increasing energy (for all values of k). Therefore, the break point in the decision boundary ($\left(\sum\rho\right)_0$) and other discontinuities in the separation power are produced by the changes in the γ-ray distribution and are also dependent on the weight of the MLD.

Next we investigate the separation power, considering different values of k, ranging from 1.0 to 2.8 at intervals of 0.2. In order to assess the improvement over the separation power as we change k, we obtain the ratio (performance ratio) between the separation power for a given k, and $k = 0$. The separation power for the case $k = 0$ is shown on the left panel of Fig. 11. The performance ratio for all the range of values of k is shown on the right panel. The color scale in the right panel corresponds to the different values of k; from the lowest (dark color) to the highest one (light color). The equivalent γ-ray energy is presented in the upper horizontal axis on both panels. Overall, the separation power (left panel) increases with increasing $\sum\rho$, until reaching a maximum around $\sum\rho < 100\,\mathrm{m}^{-2}$. Similarly to the result presented in Fig. 5, the break point is originated by the γ-ray distribution changing from being dominated by muon-less showers. Considering the result regarding the performance ratio, we observe that a maximum separation power can be achieved with $k = 2.4$ in the range

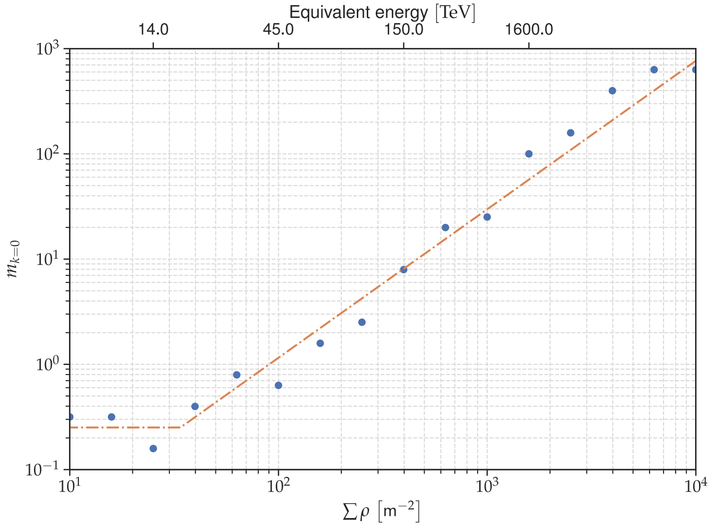

Fig. 10 Distribution of the values of m_k^{cut} in each $\sum\rho$ bin for the case of $k = 0$ (total number of muons). The solid line represents the optimum decision boundary obtained after fitting Eq. 4. The scale in the upper horizontal axis represents the equivalent γ-ray energy

$200\,\mathrm{m}^{-2} < \sum\rho < 400\,\mathrm{m}^{-2}$, equivalent to γ-ray energies from: 60 to 100 TeV. On the other hand, the apparent decrease in performance for values of $k > 1.6$ in the $\sum\rho < 50\,\mathrm{m}^{-2}$ region, is again produced by the break point in the γ-ray distributions (in this case by not being aligned between each other). We should regard both results (the maximum separation power and decrease in performance) first only qualitatively, since the real performance of the method is determined by the separation between classes and the number of events on each bin (signal noise ratio).

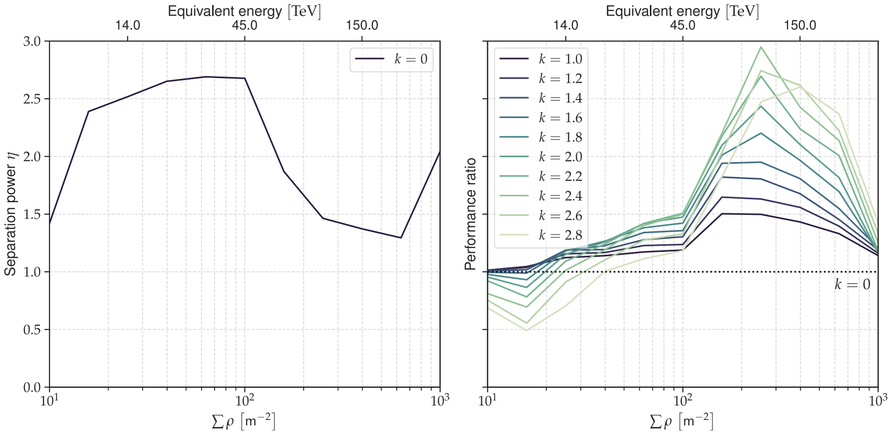

Fig. 11 Left panel: separation power as a function of $\sum\rho$ using the total number of muons ($k = 0$). Right panel: performance ratio as a function of $\sum\rho$ for different values of k. The color scale represents the values of k, ranging from $k = 1$ (dark) to $k = 2.8$ (light color) in intervals of 0.2. The scale in the upper horizontal axis in both panels represents the equivalent γ-ray energy

Finally, we estimate the effects of the increase in separation power over the sensitivity curve for ALPAQUITA+MD. The sensitivity analysis is done for a point source with 5σ detection significance or higher than 10 events (in the case where the number of background events is smaller than 1) in one year of observation. The result considering the case for $m_{k=0}$ is shown in Fig. 12. For this analysis, we have applied the optimum decision boundary and estimate the number of events γ-ray and CR surviving after the selection to achieve the 5σ sensitivity. The dotted line represents the integral flux of the hypothetical source, while the thick line represents the sensitivity curve. Besides, Figure 13 introduces the ratio of sensitivities ($S_k/S_{k=0}$) after varying the weight of the MLD. We made this comparison by obtaining the ratio of a sensitivity curve with a given value of k to the sensitivity with $k = 0$. The color scale represents the change in values of k, ranging from $k = 1$ (dark) to $k = 2.4$ (light color) at intervals of 0.2. Before interpreting the results on Fig. 13, we should emphasize the following characteristics. First, in contrast to the results regarding the separation power, a lower ratio $S_k/S_{k=0}$ in this case indicates better performance. Secondly, the apparent decrease in sensitivity above 100 TeV for all values of k is not an effect of the MLD analysis itself, but from reaching the background free region and the change of definition of the sensitivity (number of γ-ray events > 10). Hence, is clear that using the weight MLD we are able to enhance the sensitivity of ALPAQUITA+MD. In particular, by using a weight between 1.0 to 2.2 we achieve an improvement between 5 to 18% in the energy range from 10 to 100 TeV. Nevertheless, from the results in the Figure we should exclude the use of values $k > 1.8$, since they achieve worse performance. That means, despite these values having higher separation powers (taking into account the right panel of Fig. 11), the sensitivity for $k > 1.8$ is affected by the Q-factors in each bin and the shift in break point of the distribution. Consequently, the best sensitivity is achieved

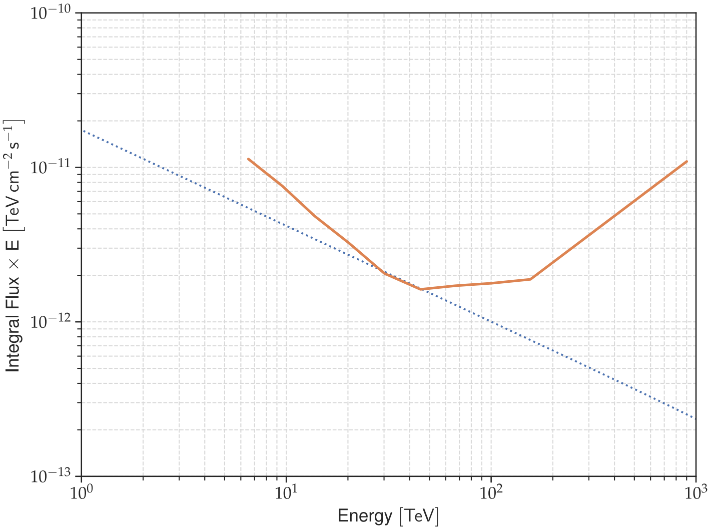

Fig. 12 Sensitivity curve for ALPAQUITA+MD considering a hypothetical Crab-like source in the Southern hemisphere. The dotted line represents the hypothetical source and the thick line represents the sensitivity curve

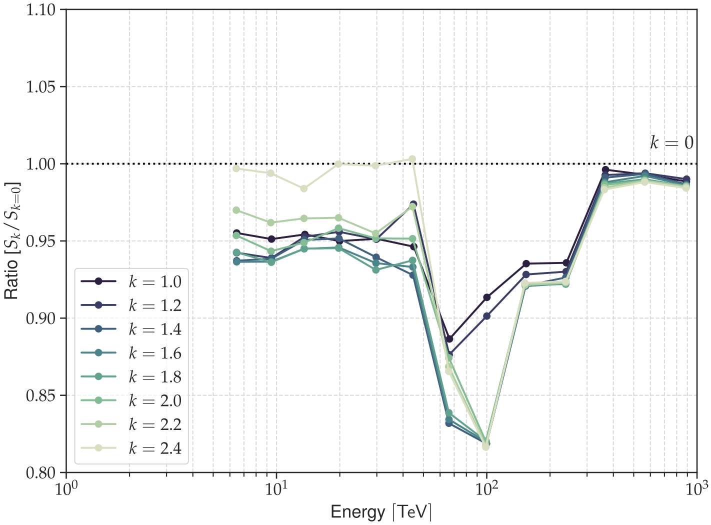

Fig. 13 Ratio of sensitivities $S_k/S_{k=0}$ for different values of k. The color scale represents the values of k, ranging from k = 1 (dark) to k = 3 (light color) in intervals of 0.2

with k = 1.6, which in turn could lead to the observation of 1 or 2 more γ-ray point sources in this energy range (taking into account Figure 17 in [17]).

5 Conclusion

The observation of γ with energies above 100 TeV will be crucial to establish the sources of cosmic rays in our Galaxy and understand its acceleration mechanisms. Recently, successful observations have been carried out in the Northern hemisphere by the Tibet ASγ, HAWC and LHAASO experiments. ALPACA is new experiment for detecting sub-PeV γ rays in the Southern Hemisphere for the first time. ALPACA is now under construction in Bolivia, and the prototype ALPAQUITA has been taking data since 2022.

To achieve large discrimination power against the dominant CR background the ALPACA experiment uses the same technique as Tibet ASγ, making the classification of air showers by detecting the muon component with an underground pool. The performance of ALPAQUITA+MD has been studied by MC simulation before and we concluded that it will be capable of detecting at least 5 γ-ray sources in one year of observations. In this paper we have developed a new analysis method, taking into account the design of ALPAQUITA+MD, and achieved an improvement over the separation power and sensitivity curves using the muon lateral distribution. From this we were able to establish the best parameter to use in the analysis as well as determine the optimal decision boundaries to made the classification. According to the calculation presented in this article, an improvement around 5 to 18% is expected when using the technique presented here, which could translate in the observation of one or two more γ-ray sources in one year of observation.

Furthermore, taking into account the study presented in Section 3, we conclude that better performance may be achieved by improving some factors imposed in the current analysis. One of these is the limited physical size of ALPAQUITA and the MD pool. We therefore expect that the sensitivity gain should be higher for the full array using the 4 MDs. Lastly, we should mention that using any machine learning techniques or other multivariate non-linear analysis, to include the spatial characteristics of the shower detected by the surface array could improve the discrimination against hadrons.

Acknowledgements The ALPACA project is supported by the Japan Society for the Promotion of Science (JSPS) through Grants-in-Aid for Scientific Research (A) 24H00220, Scientific Research (B) 19H01922, Scientific Research (B) 20H01920, Scientific Research (S) 20H05640, Scientific Research (B) 20H01234, Scientific Research (C) 22K03660, and Specially Promoted Research 22H04912, the LeoAtrox supercomputer located at the facilities of the Centro de Análisis de Datos (CADS), CGSAIT, Universidad de Guadalajara, México, and by the joint research program of the Institute for Cosmic Ray Research (ICRR), The University of Tokyo. Y. Katayose is also supported by JSPS Open Partnership joint Research projects F2018, F2019. K. Kawata is supported by the Toray Science Foundation. I. Toledano-Juarez acknowledges support from CONACyT, México; grant 754851.

Author Contributions All authors contributed to the study conception and design. Monte Carlo simulations and analysis were performed by M. Anzorena. The first draft of the manuscript was written by M. Anzorena and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding provided by The University of Tokyo.

Data Availability The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

References

1. Kobayakawa, K., Honda, H.S., Samura, T.: Acceleration by oblique shocks at supernova remnants and cosmic ray spectra around the knee region. Physical Review D. 66, 083004 (2002). https://doi.org/10.1103/PhysRevD.66.083004
2. Bell, A.R.: Cosmic ray acceleration. Astroparticle Physics. 43, 56–70 (2013). https://doi.org/10.1016/j.astropartphys.2012.05.022
3. Di Sciascio, G.: Ground-based Gamma-Ray Astronomy: an Introduction. Journal of Physics: Conference Series. 1263, 012003 (2019). https://doi.org/10.1088/1742-6596/1263/1/012003
4. Sitarek, J.: TeV Instrumentation: Current and Future. Galaxies **2022**(10), 21 (2022). https://doi.org/10.3390/galaxies10010021

5. Giuliani, A., et al.: Neutral pion emission from accelerated protons in Supernove Remnant W44. The Astrophysical Journal Letters 742, 30 (2011). https://doi.org/10.1088/2041-8205/742/2/L30
6. Ackermann, M., et al.: Detection of the Characteristic Pion-Decay Signature in Supernova Remnants. Science **339**(6121), 807–811 (2013). https://doi.org/10.1126/science.1231160
7. Amenomori, M., et al.: First Detection of Photons with Energy beyond 100 TeV from an Astrophysical Source. Physical Review Letters 123, 051101 (2019). https://doi.org/10.1103/PhysRevLett.123.051101
8. Albert, A., et al.: HAWC J2227+610 and Its Association with G106.3+2.7, a New Potential Galactic PeVatron. The Astrophysical Journal Letters 896, 29 (2020). https://doi.org/10.3847/2041-8213/ab96cc
9. Cao, Z., et al.: Ultrahigh-energy photons up to 1.4 petaelectronvolts from 12 γ-ray Galactic sources. Nature 594, 33–36 (2021). https://doi.org/10.1038/s41586-021-03498-z
10. Abeysekara, A.U., et al.: Observation of the Crab Nebula with the HAWC Gamma-Ray Observatory. The Astrophysical Journal 843, 39 (2017). https://doi.org/10.3847/1538-4357/aa7555
11. Conceição, R., et al.: Gamma/hadron discrimination at high energies through the azimuthal fluctuations of air shower particle distributions at the ground. J. Cosmol. Astropart. Phys. **2022**(10), 085 (2022). https://doi.org/10.1088/1475-7516/2022/10/086
12. Sako, T.K., et al.: Exploration of a 100 TeV gamma-ray northern sky using the Tibet air-shower array combined with an underground water-Cherenkov muon-detector array. Astroparticle Physics 32, 177–184 (2009). https://doi.org/10.1016/j.astropartphys.2009.07.006
13. Tian, Z., et al.: Study of the γ/p discrimination at ~100 TeV energy range with LHAASO experiment. Astroparticle Physics 99, 43–50 (2018). https://doi.org/10.1016/j.astropartphys.2018.02.009
14. Assunção, F., et al.: Automatic Design of Artificial Neural Networks for Gamma-Ray Detection. IEEE Access 7, 110531–110540 (2019). https://doi.org/10.1109/ACCESS.2019.2933947
15. Alfaro, R., et al.: Gamma/hadron separation with the HAWC observatory. Nuclear Instruments and Methods in Physics Research A 1039, 166984 (2022). https://doi.org/10.1016/j.nima.2022.166984
16. Okukawa, S., et al.: Neural networks for separation of cosmic gamma rays and hadronic cosmic rays in air shower observation with a large area surface detector array. Machine Learning: Science and Technology 5, 025016 (2024). https://doi.org/10.1088/2632-2153/ad3a33
17. Kato, S., et al.: Detectability of southern gamma-ray sources beyond 100 TeV with ALPAQUITA, the prototype experiment of ALPACA. Experimental Astronomy 52, 85–107 (2021). https://doi.org/10.1007/s10686-021-09796-8
18. González, N., et al.: A muon-based observable for a photon search at 30 to 300 TeV. Astroparticle Physics 114, 48–59 (2020). https://doi.org/10.1016/j.astropartphys.2019.06.005
19. Heck, D., et al.: CORSIKA: A Monte Carlo Code to Simulate Extensive Air Showers. Technical Report FZKA 6019, Forschungszentrum Karlsruhe (1998)
20. Shibata, M., et al.: Chemical composition and maximum energy of Galactic cosmic rays. The Astrophysical Journal 716, 1076–1083 (2010). https://doi.org/10.1088/0004-637X/716/2/1076
21. Battistoni, G., et al.: Overview of the FLUKA code. Annals of Nuclear Energy 82, 10–18 (2015). https://doi.org/10.1016/j.anucene.2014.11.007
22. Pierog, T., et al.: EPOS LHC: Test of collective hadronization with data measured at the CERN Large Hadron Collider. Physical Review C 92, 034906 (2015). https://doi.org/10.1103/PhysRevC.92.034906
23. Rastegarzadeh, G., Rafezi, L.: Energy, altitude, and mass dependence of steepness of the lateral distribution function of electrons and muons in extensive air showers. Nuclear Instruments and Methods in Physics Research A 763, 197–201 (2014). https://doi.org/10.1016/j.nima.2014.06.055
24. Ros, G., et al.: A new composition-sensitive parameter for ultra-high energy cosmic rays. Astropart. Phys. **35**(3), 140–151 (2011). https://doi.org/10.1016/j.astropartphys.2011.06.011
25. Muraishi, H., et al.: Evidence for TeV gamma-ray emission from the shell type SNR RX J1713.7-3946. Astronomy and Astrophysics 354, 57–61 (2000). https://doi.org/10.48550/arXiv.astro-ph/0001047
26. Abdo, A.A., et al.: Observations of the young supernova remnant RX J1713.7-3946 with the FermiLarge Area Telescope. The Astrophysical Journal 734, 28 (2011). https://doi.org/10.1088/0004-637X/734/1/28
27. Agostinelli, S., et al.: Geant4–a simulation toolkit. Nucl. Inst. Methods Phys. Res. A **06**(3), 250–303 (2003). https://doi.org/10.1016/S0168-9002(03)01368-8
28. Aharonian, F., et al.: The Crab Nebula and Pulsar between 500 GeV and 80 TeV: Observations with the HEGRA Stereoscopic Air Cerenkov Telescopes. The Astrophysical Journal 614, 897–913 (2004). https://doi.org/10.1086/423931

Authors and Affiliations

M. Anzorena[1] · E. de la Fuente[2] · K. Fujita[1] · R. Garcia[1] · K. Goto[3] · Y. Hayashi[4] · K. Hibino[5] · N. Hotta[6] · G. Imaizumi[1] · A. Jimenez-Meza[7] · Y. Katayose[8] · C. Kato[9] · S. Kato[1] · T. Kawashima[1] · K. Kawata[1] · T. Koi[10] · H. Kojima[11] · T. Makishima[8] · Y. Masuda[9] · S. Matsuhashi[8] · M. Matsumoto[9] · R. Mayta[12,13] · P. Miranda[14] · A. Mizuno[1] · K. Munakata[9] · Y. Nakamura[1] · M. Nishizawa[15] · Y. Noguchi[8] · S. Ogio[1] · M. Ohnishi[1] · S. Okukawa[8] · A. Oshima[3,10] · M. Raljevic[14] · H. Rivera[14] · T. Saito[16] · T. Sako[1] · T.K. Sako[17] · T. Shibasaki[18] · S. Shibata[11] · A. Shiomi[18] · M. Subieta[14] · F. Sugimoto[1] · N. Tajima[19] · W. Takano[5] · M. Takita[1] · Y. Tameda[20] · K. Tanaka[21] · R. Ticona[14] · I. Toledano-Juarez[22] · H. Tsuchiya[23] · Y. Tsunesada[12,13] · S. Udo[5] · R. Usui[8] · G. Yamagishi[8] · K. Yamazaki[10] · Y. Yokoe[1]

✉ M. Anzorena
anzorena@icrr.u-tokyo.ac.jp

[1] Institute for Cosmic Ray Research, University of Tokyo, Kashiwanoha, 277-8582 Kashiwa, Chiba, Japan

[2] Departamento de Física, CUCEI, Universidad de Guadalajara, Blvd. Marcelino García Barragán 1421, esq Calzada Olímpica, 731-3194 Guadalajara, Jalisco, Mexico

[3] Graduate School of Engineering, Chubu University, 487-8501 Kasugai, Aichi, Japan

[4] Department of Science and Technology, Shinshu University, 390-8621 Matsumoto, Nagano, Japan

[5] Faculty of Engineering, Kanagawa University, 221-8686 Yokohoma, Kanagawa, Japan

[6] Utsunomiya University, 321-8505 Utsunomiya, Tochigi, Japan

[7] Departamento de Tecnologías de la Información, CUCEA, Universidad de Guadalajara, Avenida Parres Arias,1012, Zapopan, Jalisco, Mexico

[8] Faculty of Engineering, Yokohama National University, 240-8501 Yokohama, Kanagawa, Japan

[9] Department of Physics, Shinshu University, 390-8621 Matsumoto, Nagano, Japan

[10] College of Engineering, Chubu University, 487-8501 Kasugai, Aichi, Japan

[11] Chubu Innovative Astronomical Observatory, Chubu University, 487-8501 Kasugai, Aichi, Japan

[12] Graduate School of Science, Osaka Metropolitan University, 558-8585 Osaka, Osaka, Japan

[13] Nambu Yoichiro Institute for Theoretical and Experimental Physics, Osaka Metropolitan University, 558-8585 Osaka, Osaka, Japan

[14] Instituto de Investigaciones Físicas, Universidad Mayor de San Andres, La Paz 8635, La Paz, Bolivia

[15] National Institute of Informatics, 101-8430 Chiyoda, Tokyo, Japan

[16] Tokyo Metropolitan College of Industrial Technology, 116-8523 Arakawa, Tokyo, Japan

[17] Department of Information and Electronics, Nagano Prefectural Institute of Technology, 386-1211 Ueda, Nagano, Japan

[18] College of Industrial Technology, Nihon University, 275-8576 Narashino, Chiba, Japan

[19] RIKEN, 351-0198 Wako, Saitama, Japan

[20] Faculty of Engineering, Osaka Electro-Communication University, 572-8530 Neyagawa, Osaka, Japan

[21] Hiroshima City University, 731-394 Hiroshima, Hiroshima, Japan

[22] Doctorado en Ciencias Físicas, CUCEI, Universidad de Guadalajara, Blvd. Marcelino García Barragán 1421, esq Calzada Olímpica, Guadalajara 44430, Jalisco, Mexico

[23] Japan Atomic Energy Agency, 319-1195 Tokai-mura, Ibaraki, Japan