Principal components analysis of Cerenkov photon distributions from extensive air showers applied to GeV gamma–proton discrimination

J.L. Contreras a, E. Faleiro b,*, J.M.G. Gómez a, R.A. Molina c, L. Muñoz a, A. Relaño a, J. Retamosa a

a Departamento de Física Atómica, Molecular y Nuclear, Universidad Complutense de Madrid, 28040-Madrid, Spain
b Departamento de Física Aplicada, E.U.I.T. Industrial, Universidad Politécnica de Madrid, Ronda de Valencia 3, 28012-Madrid, Spain
c Max-Planck-Institut für Physik Komplexer Systeme, Nothnitzer Strasse 38, D-01187 Dresden, Germany

Received 6 February 2006; received in revised form 18 April 2006; accepted 27 April 2006
Available online 5 June 2006

Abstract

High-energy interactions of cosmic rays with Earth atmosphere and the resulting extensive air showers have been simulated by means of the CORSIKA Monte Carlo code. Statistical fluctuations of the two-dimensional secondary Cerenkov photon density distributions at ground level produced by GeV cosmic \( \gamma \) rays and protons are studied by means of principal components analysis (PCA). This provides a decreasing sequence of covariance matrix eigenvalues that may be fitted to an analytical expression allowing to distinguish among different primary cosmic rays. A very efficient discrimination method of gamma ray induced showers from proton simulated extensive air showers is proposed as a result of this analysis. A cutting parameter is calculated, and the efficiency of the cutting procedure for \( \gamma \)–proton separation is evaluated under experimental conditions.

© 2006 Elsevier B.V. All rights reserved.
PACS: 05.40.+j; 96.40.Pq; 98.70.Sa

Keywords: Principal components analysis; 1/f noise; Cosmic ray fluctuations; Cerenkov photons; \( \gamma \)-Hadron discrimination

1. Introduction

Cerenkov photons are mainly created in the atmosphere when the particles forming a extensive air shower (EAS) [1,2] reach velocities exceeding that of light in that media. In the last years detectors based on the study the Cerenkov component are becoming the preferred tool for the observation of EAS, due to the much higher number of Cerenkov photons compared to other detectable particles. Two main detection techniques comprise the field of Atmospheric Cerenkov Detectors: Cerenkov Telescopes and sampling arrays [3]. The first category corresponds to devices with imaging capabilities. In their simplest version they collect the light produced by the EAS at one single location, and focus it using a system of mirrors on to a camera composed of fast multipliers. The fast photograph of the shower development thus produced is consequently analyzed to yield information about the direction, energy and nature of the incident particle. Imaging Atmospheric Cerenkov Telescopes (IACT) are often grouped in arrays to provide stereoscopic images of showers, which greatly improve the determination of the EAS parameters [4].

The second type of Atmospheric Cerenkov detectors correspond to those composed of stations which sample...
the light front simultaneously at different places, but lacking imaging capabilities. Arrays composed of non-imaging Cerenkov detectors measure the lateral distribution of Cerenkov light at ground level.

Since cosmic gamma rays travel on straight lines through our Galaxy reaching the Earth atmosphere, they can be used to image sources of such high-energy particles if the isotropic background of cosmic hadron events can be reduced. This reduction constitutes a basic tool in high energy astronomy; therefore, several $\gamma$–hadron separation methods [5,6] are used to help to identify the meager portion of $\gamma$ initiated EAS among all the showers, and most of them are based in two stages. In a first stage a set of appropriate system variables for $\gamma$–proton separation are extracted from the EAS measurements. Then, in a second stage statistical tests are applied to measure the separation degree of two samples of independent distributions of these variables.

In particular for IACT, a good separation between gamma initiated EAS and those of hadronic origin is obtained by mean of the so-called Hillas parameters, derived from the second moments of the image formed in the camera. The use of the Hillas parameters, together with Bayesian methods and neural net analysis [7,8] have shown good $\gamma$–proton discrimination properties on Cerenkov photon images. Some results about $\gamma$–hadron separation using Cerenkov photon density fluctuations measured on non-imaging Cerenkov detector arrays [9–11] enhance the importance of using Cerenkov photons.

Other $\gamma$–hadron separation methods based on self-similar features of the secondary charged particle distributions have been proposed [12–14]. The $\gamma$–hadron separation within the framework of the PCA, or principal components analysis, has also been studied [15] from the secondary charged particle distributions, providing a rather high separation power.

In this paper we deal with the $\gamma$–hadron separation also within the framework of the PCA following the well established method of Ref. [15], but using the information supplied by the Cerenkov photon density distributions at detection level; specifically, we study the main properties of the covariance matrix eigenvalue spectrum and present an efficient method that provides a rather high separation power when it is applied to samples of different simulated EAS which have been generated according to experimental situations. The special features of secondary Cerenkov photons distributions justify itself a separated study of its $\gamma$–hadron discrimination capabilities. But, as it was pointed above, new experimental trends on cosmic ray detection on the Earth surface assign a main role to Cerenkov radiation, and thus separation methods based on Cerenkov photons distributions may become very important.

The organization of the rest of this paper is as follows: In Section 2 we describe the databases we have used and how the PCA is applied to them. Section 3 address to the characterization of the eigenvalue structure; it is found that the value of the parameters is different for electromagnetic showers and proton showers, and a cutting procedure is proposed for identification of the $\gamma$ ray showers. Finally, the conclusions are summarized in Section 4.

2. Simulated extensive air showers and principal components analysis

To simulate EAS we have used the version 6.20 of the CORSIKA Monte Carlo code [16]. The hadronic interactions at high energies have been described by a model based on the parton-based Gribov–Regge theory to simulate ultra-relativistic heavy ion collisions. We have initiated EAS from high energy cosmic $\gamma$ rays and protons with energies ranging from 10 to 500 GeV following a flat spectral law, the height of the first interaction varying at random according to the appropriate mean free path (some other details about the simulation can be found in the above reference). The observation level was set to 2200 m above sea level, corresponding to the altitude of the Roque de los Muchachos, where the HEGRA experiment is located. More than 300,000 primary protons and approximately 50,000 primary $\gamma$ rays constitute the global amount of simulated showers. As a simplification, all the showers have been generated vertically. The effect of a non-vertical shower axis may be incorporated to the analysis by geometrical reconstruction of the shower front from the arrival time structure at detection level. On the other hand, it is introduced an uniform random distribution of the core position on the detector area. In any case, the core must not be too far from the center of the detection area to avoid a dramatic lost of information. A useful criterion shall be established further.

An event profile consists of a bidimensional distribution of Cerenkov photons collected at detection level. In practice, bidimensional data fields $F(x_1, x_2)$ representing the secondary Cerenkov photons density distributions at the point $(x_1, x_2)$ from the core, taken as the center of a relative coordinate system, are considered; even more, in real data only discrete fields \{\(F(j_1,j_2)\)\} with \(j_1 = 1, \ldots, N_1\), \(j_2 = 1, \ldots, N_2\) of \(N_1 \times N_2\) data are available. Fig. 1 shows in the upper panels some examples of EAS Cerenkov photons distributions at the observation level; two events generated by a 402 GeV primary proton and a 87 GeV primary photon having between 40,000 and 50,000 Cerenkov photons and their first collision at different altitudes are displayed. The global shape of Cerenkov photon distributions has a decreasing exponential behavior [17], and just like the secondary charged particle distributions, only depends on the core distance; therefore, the polar coordinates are the most convenient representation system for the bidimensional cosmic ray events $F(r, \theta)$.

The statistical tool we are going to use for discriminating $\gamma$ rays from protons is the principal component analysis (PCA). This is a very useful tool widely used in many different frames, sharing all them the basic idea of reducing the dimension of the original data-set by projecting the raw data onto a few dominant eigenvectors with large
variance; thus, it may provide a simple description of the original data and a better insight on its underlying characteristics. PCA has been applied to data compressing and classification of spectral data of stars [18] and galaxies [19,20]. It was recently shown [15] that this tool provides a new, effective, simple c–proton separation method; we encourage the reader interested in details to consult the quoted reference.

The analysis of EAS Cerenkov photon distributions starts with the classification of primary cosmic rays. In simulated EAS we know the energy of the primary particle, but in real experiments a direct measure of the primary cosmic ray energy is not possible; therefore, it is more convenient to classify the events not in energy bins, but according to a quantity experimentally measurable. Here, we classify the events by the number of Cerenkov photons $N_s$ at ground level, since this quantity can be measured in experiments. We have considered a binning of amplitude $\Delta N_s = 10,000$ Cerenkov photons; Table 1 shows the distribution of the simulated events into several of such bins.

Once the original events have been classified according to their number of Cerenkov photons, we proceed to reduce the polar bidimensional profile to a unidimensional representation $q(h)$; as it is pointed in [15], this step is necessary to apply PCA. The one-dimensional representation is defined by

Here, $N_p$ is the number of primary cosmic rays whose number of Cerenkov photons arriving to the detection surface is within the bin $\Delta N_s$. $A_0$, $A_1$ and $A_2$ are the parameters of a fitting procedure to a second-order Chebyshev-series polynomial approach of the mean eigenvalue spectrum from PCA analysis.

### Table 1

<table>
<thead>
<tr>
<th>Primary</th>
<th>$N_p$</th>
<th>$\Delta N_s$</th>
<th>$A_0$</th>
<th>$A_1$</th>
<th>$A_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\gamma$</td>
<td>10,554</td>
<td>$\Delta N_s$: 10,000–20,000</td>
<td>2.601</td>
<td>-1.827</td>
<td>0.981</td>
</tr>
<tr>
<td></td>
<td>10,043</td>
<td>$\Delta N_s$: 20,000–30,000</td>
<td>2.618</td>
<td>-1.846</td>
<td>1.017</td>
</tr>
<tr>
<td></td>
<td>9,360</td>
<td>$\Delta N_s$: 30,000–40,000</td>
<td>2.628</td>
<td>-1.857</td>
<td>1.039</td>
</tr>
<tr>
<td></td>
<td>7,428</td>
<td>$\Delta N_s$: 40,000–50,000</td>
<td>2.647</td>
<td>-1.879</td>
<td>1.079</td>
</tr>
<tr>
<td></td>
<td>4,939</td>
<td>$\Delta N_s$: 50,000–60,000</td>
<td>2.671</td>
<td>-1.904</td>
<td>1.129</td>
</tr>
<tr>
<td></td>
<td>3,884</td>
<td>$\Delta N_s$: 60,000–70,000</td>
<td>2.690</td>
<td>-1.925</td>
<td>1.169</td>
</tr>
<tr>
<td></td>
<td>3,433</td>
<td>$\Delta N_s$: 70,000–80,000</td>
<td>2.693</td>
<td>-1.928</td>
<td>1.175</td>
</tr>
<tr>
<td></td>
<td>3,148</td>
<td>$\Delta N_s$: 80,000–90,000</td>
<td>2.698</td>
<td>-1.934</td>
<td>1.185</td>
</tr>
<tr>
<td></td>
<td>2,937</td>
<td>$\Delta N_s$: 90,000–100,000</td>
<td>2.718</td>
<td>-1.956</td>
<td>1.229</td>
</tr>
<tr>
<td>Proton</td>
<td>28,423</td>
<td>$\Delta N_s$: 10,000–20,000</td>
<td>3.157</td>
<td>-2.450</td>
<td>2.149</td>
</tr>
<tr>
<td></td>
<td>19,745</td>
<td>$\Delta N_s$: 20,000–30,000</td>
<td>3.222</td>
<td>-2.520</td>
<td>2.286</td>
</tr>
<tr>
<td></td>
<td>14,414</td>
<td>$\Delta N_s$: 30,000–40,000</td>
<td>3.241</td>
<td>-2.540</td>
<td>2.326</td>
</tr>
<tr>
<td></td>
<td>10,756</td>
<td>$\Delta N_s$: 40,000–50,000</td>
<td>3.254</td>
<td>-2.553</td>
<td>2.353</td>
</tr>
<tr>
<td></td>
<td>7,908</td>
<td>$\Delta N_s$: 50,000–60,000</td>
<td>3.266</td>
<td>-2.565</td>
<td>2.379</td>
</tr>
<tr>
<td></td>
<td>6,025</td>
<td>$\Delta N_s$: 60,000–70,000</td>
<td>3.277</td>
<td>-2.576</td>
<td>2.401</td>
</tr>
<tr>
<td></td>
<td>4,666</td>
<td>$\Delta N_s$: 70,000–80,000</td>
<td>3.293</td>
<td>-2.592</td>
<td>2.435</td>
</tr>
<tr>
<td></td>
<td>3,370</td>
<td>$\Delta N_s$: 80,000–90,000</td>
<td>3.301</td>
<td>-2.600</td>
<td>2.453</td>
</tr>
<tr>
<td></td>
<td>6,029</td>
<td>$\Delta N_s$: 90,000–100,000</td>
<td>3.309</td>
<td>-2.608</td>
<td>2.468</td>
</tr>
</tbody>
</table>

Fig. 1. Bidimensional Cerenkov photon density of simulated EAS initiated by a 402 GeV proton and a 87 GeV photon of 40,000–50,000 Cerenkov photons each, in a squared detection area of 256 m of side. Their unidimensional representations in terms of 256 sectors in the polar angle, ranging from 0 to $2\pi$, is done below when $r_{\text{max}}$ is chosen as $r_{\text{eff}}$. 
\[
\rho(\theta) = \int_0^{r_{\text{eff}}} \mathcal{d}r F(r, \theta),
\]

where \( r_{\text{eff}} \) stands for the effective radius from the core to reveal the main correlation features of \( \rho(\theta) \). When the core is completely centered on the detection area, \( r_{\text{eff}} \) may be taken as the maximum radius from the core where the Cerenkov photons may be detected and accepted for analyzing purposes; in this work \( r_{\text{max}} = 256 \) m, which is close to the radius of the HEGRA experiment at La Palma. Nevertheless, for \( \gamma \)-proton separation purposes it may be convenient to study the influence of this event parameter; this study shall be done in the next section before proposing the definitive separation procedure. Once \( \rho(\theta) \) has been adequately constructed, a binning of 256 sectors on the polar angle \( \theta \) leads us to characterize each EAS event as a series of 256 data length, i.e., each event is represented by a vector \( \psi_n \), where \( i = 1, 2, \ldots, 256 \). This binning is chosen to have a long series without loosing information about the fluctuations. In the lower panels of Fig. 1 the unidimensional representation of the upper panel events is displayed. Among other properties, the sequence may be considered as a wide-sense stationary series with a mean value representing the mean particle density for the \( r_{\text{eff}} \) chosen.

As it is pointed in [15], PCA requires an ensemble of series \( \psi_n \) to properly estimate their covariance matrix (see below to a precise definition of this quantity). This ensemble can be built cutting each single event in \( M \) non-overlapping pieces of length \( N \), where \( M \gg N \). Before selecting the size \( N \) of these non-overlapping pieces, we have to check that there are no correlation structures for \( \Delta \theta \geq N \), i.e., we are not losing relevant information. This verification can be performed by means the power spectrum analysis of the whole series \( \psi_n \). In Fig. 2 it is shown the mean power spectrum of protons and photons belonging to the \( \Delta N_4 \) database. They may be fitted to the model \( P(k) = a/k^b + c \), with constants \( a, b \) and \( c \), that make clear their underlying scaling properties [21,22]. According with this scaling behavior, such a series exhibit self-similarity features; thus, the absence of underlying characteristic structures allows us to consider a whole series decomposition in pieces of any length without losing relevant information. To better estimate the covariance matrix, we have selected a cut of 16 non-overlapping pieces of 16 data each.

The covariance matrix can be defined as follows:

\[
C_{ij} = \frac{1}{M} \sum_{k=1}^{M} \frac{(\rho_{ik} - \mu_i)(\rho_{jk} - \mu_j)}{\sigma_i \sigma_j},
\]

where the first index \( i \) runs over the components of the \( k \)th vector of the ensemble, while the index \( k \) runs over the \( i \)th component of the different vector of the ensemble. The quantities \( \mu_i \) and \( \sigma_i \) are defined as follows:

\[
\mu_i = \frac{1}{M} \sum_{k=1}^{M} \rho_{ik}, \quad \sigma_i^2 = \frac{1}{M} \sum_{k=1}^{M} (\rho_{ik} - \mu_i)^2.
\]

Once \( C \) is evaluated, the eigenvalue equation \( C \Psi_n = \lambda_n \Psi_n \) is solved according to standard numerical recipes [23]. Fig. 3 shows the mean decreasing ordered eigenvalue spectrum \( \lambda_n \) as a function of the rank \( n \) from the proton and also photon samples belonging to the same \( \Delta N_4 \) bin. From Fig. 3 it is clear that protons series possesses a dominant eigenvalue carrying away a significant percent of the total variance; for \( \Delta N_4 \) proton databases the largest mean eigenvalue represents around the 80% of the total variance. On the opposite, in photon series the

---

**Fig. 2.** Comparison of mean power spectrum in a log–log scale for the \( \Delta N_4 \) samples of protons and photons. Rather similar shapes are obtained by using the other databases.

**Fig. 3.** Decreasing ordered eigenvalue spectrum corresponding to the \( \Delta N_4 \) databases. A 80% of the total variance is carried by the largest eigenvalue for the proton sample. The variance is smooth distributed among all the eigenvalues for the photon sample instead.
variance is smooth distributed from the largest to the shortest eigenvalue. Rather similar results are found in the rest of databases. Thus, proton series are more correlated than photon series and, therefore, protons and photons present neat differences in the mean behavior of the eigenvalue spectrum which constitute the start point of the identification method we shall present in the next section.

3. The identification method of primary cosmic \( \gamma \) rays

The qualitative differences for protons and photons shown in Fig. 3 compel us to use \( \lambda_n \) as a separation method. The simplest way to do it is by fitting the eigenvalue spectrum to a few-parameter function, so that the set of parameters identify the primary cosmic ray. In [15] a three-parameter polynomial was shown to be a good choice; we are going to apply the same method in this paper.

The procedure consists in analyzing the \( A_0, A_1, A_2 \) parameters as a result of a fitting procedure to

\[ \lambda_n = \frac{1}{2} A_0 T_0(n) + A_1 T_1(n) + A_2 T_2(n), \]

where \( \lambda_n \) is the decreasing ordered eigenvalue spectrum and \( T_j \) stands for the Chebyshev polynomial of the first kind of degree \( j \). Then, we use the value of the most adequate parameter to separate cosmic \( \gamma \) rays from primary protons.

Resulting parameters \( A_0, A_1 \) and \( A_2 \) are shown in Table 1.

The separation power of a parameter is frequently measured by the quality factor \( Q \) defined by [24,25]

\[ Q = \frac{\kappa_\gamma}{\sqrt{\kappa_p}}, \]

where \( \kappa_\gamma \) and \( \kappa_p \) are the fractions (efficiencies) of photons and protons kept by the algorithm, namely \( \kappa_\gamma = N_{\text{cut}}^\gamma / N_i \) and \( \kappa_p = N_{\text{cut}}^p / N_i \), where \( N_{\text{cut}} \) is the number of remaining particles after the cutting procedure. Thus, \( Q \) will be used along this work as a measure of the enrichment in primary photons of the original sample after the cutting procedure.

Fig. 4 shows the distribution of the \( A_0, A_1, A_2 \) parameters for \( \gamma \) and protons belonging to the \( \Delta N_4 \) database. It is clearly seen that there are qualitative differences for photons and protons in all the cases, but, at least in some of them, the shape of the distributions is quite similar. The best way to determine if any of these parameters can be a good choice to perform the separation algorithm consists in applying some statistical test to the distributions. We set a null hypothesis consisting in that the samples for protons and photons belong to the same population, and we calculate the \( P \)-values of Student’s \( t \)-test, Kolmogorov–Smirnov test and Mann–Whitney \( U \)-test. Larger \( P \)-values correspond to the smaller probability of accepting the null hypothesis, i.e., in such case we can conclude that samples for photons and protons are different. Table 2 shows the \( P \)-values of this three statistical tests for the \( \Delta N_4 \) bin; quite

![Fig. 4. Distribution of the values for parameters \( A_0, A_1 \) and \( A_2 \) from the \( \Delta N_4 \) database. The panels show that any of them may be used in separation algorithms, as it is confirmed by the results of the three statistical tests (see text).](image-url)
similar results are found from the other databases used in the present work. Larger P-values for at least two of the three tests select the parameter to be a good candidate for separation methods. The numerical results suggest that any of the three parameters is suitable to perform an effective separation. Moreover, it is found that all three parameters are linear correlated. Thus we have chosen the \( A_1 \) parameter since the P-values for this parameter are greater than those of the other two.

We concentrate our efforts on the \( A_1 \) parameter and study the influence of \( r_{\text{eff}} \) (the maximum radius from the core where the Cerenkov photons may be detected and accepted for the analysis and introduced in (1)) in order to improve the separation power. We proceed to evaluate the quality factor \( Q \), preserving a fixed amount of proton background, for a separation procedure according the \( A_1 \) parameter when \( r_{\text{eff}} \) varies from a few meters to the maximum available value \( r_{\text{eff}} = 256 \) m. This is done by binning the complete range of \( r_{\text{eff}} (0–256 \) m) into 32 equal bin units (b.u.). For a fixed value of \( r_{\text{eff}} \) the unidimensional

### Table 2

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( t )</th>
<th>( U )</th>
<th>( K )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( A_0 )</td>
<td>159.69</td>
<td>95.28</td>
<td>0.29</td>
</tr>
<tr>
<td>( A_1 )</td>
<td>172.85</td>
<td>95.58</td>
<td>0.30</td>
</tr>
<tr>
<td>( A_2 )</td>
<td>92.14</td>
<td>92.14</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Larger (positive or negative) P-values for at least two of the three tests, select the parameter to be a good candidate for separation methods.

---

**Fig. 5.** Quality factors \( Q \) of \( \gamma \)-proton discrimination, as a function of the value of \( r_{\text{eff}} \) evaluated for \( \Delta N_4 \), with odd \( s \), databases. Quality factor steeply grows from \( r_{\text{eff}} = 5 \) to 16 b.u. and then it smoothly increases up to \( r_{\text{max}} = 32 \) b.u. A bin unit is equivalent to 8.0 m.

**Fig. 6.** Distribution of the values for \( A_1 \) and \( Q \) from the \( \Delta N_4 \) database when 1000 realizations of a uniform random core distribution over the detection area are considered.
representation of each event is obtained according with (1) and the separation procedure is applied. Fig. 5 clearly shows that the quality factor steeply grows when \( r_{\text{eff}} \) ranges from 5 to 16 b.u. and then it smoothly increases up to \( r_{\text{max}} = 52 \) b.u. If a good separation power is set up from a quality factor \( Q > 2 \), then all the events randomly placed on detection area with \( r_{\text{eff}} < 100 \) m (\( \approx 12 \) b.u.) must be discarded. Thus, the separation procedure can be improved by taking in any case \( r_{\text{eff}} = 12 \) b.u. as the minimum of \( r_{\text{eff}} \) obtaining the so-called optimized databases.

With the optimized databases, the core of each event is placed randomly on the detection area and the \( r_{\text{eff}} \) is evaluated as the radius from the core of the shower front circular surface that remains inside the detection area; thus it is obtained the unidimensional events representation. It is clear that a realization of the uniform random distribution of all the cores belonging to a database on the detection area will give rise to a particular value of the quality factor \( Q \), and therefore to a value of the \( A_1 \) separation parameter when a fixed amount of proton background (\( \kappa_p = 5\% \)) is kept. This leads us to consider many realizations of the uniform random distribution of all the cores and to perform the analysis each time to get statistical distributions of the relevant magnitudes, namely \( Q \) and \( A_1 \). The mean value of such distributions must be considered as the representative value of the separation power and the separation parameter respectively, but the shape of the distributions is also important to ensure that every single event can be identified with a good value of \( Q \). Thus, considering 1000 realizations, the distributions of \( Q \) and \( A_1 \) for the \( \gamma \)-proton separation inside the \( \Delta N_4 \) database are shown in Fig. 6: the mean values and standard deviations are also displayed in the figure. It is clearly seen that in all the cases the value of \( Q \) stands near from its mean value – there are not dramatic differences from one realization to other. Consequently, the \( A_1 \) mean value is then proposed as a specific cut for this parameter. Table 3 shows for all the databases the mean values and their standard deviations for the magnitudes \( \kappa_\gamma \), \( \kappa_p \) and \( Q \) involved in the separation procedure when the cut is performed. As a final result, a \( \gamma \) enrichment measured by \( Q \approx 3.5 \) is obtained in any case; that represents a signal-background ratio rather suitable to developing of the High Energy Cosmic Gamma Ray Astronomy. As a reference, using wavefront sampling techniques, the differences in the fluctuations of Cerenkov photons distributions measured by some density parameters for primary energies ranging from 100 GeV to 2 TeV are studied [10,11]. The resulting quality factors (the best around 3.3) lead us to reject more than 90% of background protons.

4. Conclusions

Cerenkov light emitted by the charged particles moving through the atmosphere in EAS development is of great interest for Gamma Ray Astronomy. Although there are also other ways to detect air showers, the Cerenkov method is characterized by the lowest energy threshold now and let us to design a great variety of devices to perform an efficient detection of air showers. Since gamma induced air showers can only be used to image sources of such high-energy particles, the rejection of cosmic hadron induced showers is of crucial importance.

For this purpose, we have generated several Monte Carlo samples of extensive air showers initiated by protons and \( \gamma \) rays at energies between 10 and 500 GeV, and we have studied the fluctuations of the bidimensional Cerenkov photon density distributions at detection level.

The EAS events are classified in several bins \( \Delta N_4 \) according to the total number of Cerenkov photons \( N_s \) arriving to the detection area. From the polar coordinate system representation of each event \( P(r, \theta) \), we have obtained a stationary unidimensional series by integrating over the radial coordinate up to an effective radius \( r_{\text{eff}} \) which essentially depends on the core position on the detection area. We have also proof that the maximum value of \( r_{\text{eff}} \), which stands for including inside the series the secondary particles far from the core, can be adequately chosen to improve the \( \gamma \)-proton discrimination power of the proposed method.

The correlation structure of all the recorded EAS is studied by applying PCA. The shape of the decreasing ordered eigenvalue spectrum \( \lambda_n \) of the covariance matrix is shown to be an adequate choice for discrimination purposes. A cut in a single parameter \( A_1 \) preserves a fixed percent of the proton background and gives rise to an enrichment in photons events measured by a quality factor \( Q \), while a fraction \( \kappa_p \) of initial photons is kept. Provided that a random core distribution of each database is considered as an adequate approximation to the real situation, the separation procedure must be repeated for each database in order to obtain a statistical distribution of the main parameters, namely, \( Q \) and \( A_1 \). Table 3 shows the mean values of such distributions providing the final results of the analysis.

Although various shower characteristics like image shape, time profile or wavefront sampling technique, among others, have been used in the literature for \( \gamma \)-hadron separation providing suitable quality factors, we conclude that the study of the covariance matrix eigenvalue

---

**Table 3**

The set of all EAS events having a value of the \( A_1 \) parameter larger than the given value of \( A_1 \), contains a fraction \( \kappa_p \) of the protons and a fraction \( \kappa_\gamma \) of the \( \gamma \) rays.

<table>
<thead>
<tr>
<th>( \Delta N_4 )</th>
<th>( A_1 )</th>
<th>( Q )</th>
<th>( \kappa_\gamma (%) )</th>
<th>( \kappa_p (%) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( 10,000-20,000 )</td>
<td>( &gt;-2.29 )</td>
<td>2.01 ± 0.03</td>
<td>44.66 ± 0.04</td>
<td>4.92 ± 0.08</td>
</tr>
<tr>
<td>( 20,000-30,000 )</td>
<td>( &gt;-2.55 )</td>
<td>3.23 ± 0.03</td>
<td>70.23 ± 0.04</td>
<td>4.73 ± 0.01</td>
</tr>
<tr>
<td>( 30,000-40,000 )</td>
<td>( &gt;-2.67 )</td>
<td>3.52 ± 0.04</td>
<td>77.95 ± 0.03</td>
<td>4.91 ± 0.01</td>
</tr>
<tr>
<td>( 40,000-50,000 )</td>
<td>( &gt;-2.73 )</td>
<td>3.58 ± 0.05</td>
<td>80.06 ± 0.03</td>
<td>5.01 ± 0.02</td>
</tr>
<tr>
<td>( 50,000-60,000 )</td>
<td>( &gt;-2.78 )</td>
<td>3.60 ± 0.08</td>
<td>79.65 ± 0.04</td>
<td>4.89 ± 0.02</td>
</tr>
<tr>
<td>( 60,000-70,000 )</td>
<td>( &gt;-2.79 )</td>
<td>3.55 ± 0.08</td>
<td>77.18 ± 0.05</td>
<td>4.72 ± 0.02</td>
</tr>
<tr>
<td>( 70,000-80,000 )</td>
<td>( &gt;-2.81 )</td>
<td>3.58 ± 0.07</td>
<td>80.83 ± 0.05</td>
<td>5.10 ± 0.02</td>
</tr>
<tr>
<td>( 80,000-90,000 )</td>
<td>( &gt;-2.83 )</td>
<td>3.57 ± 0.09</td>
<td>80.12 ± 0.05</td>
<td>5.04 ± 0.03</td>
</tr>
<tr>
<td>( 90,000-100,000 )</td>
<td>( &gt;-2.87 )</td>
<td>3.61 ± 0.07</td>
<td>79.92 ± 0.05</td>
<td>4.91 ± 0.02</td>
</tr>
</tbody>
</table>

The quality factor \( Q \) corresponding to the given value of \( A_1 \) is also shown.
spectrum of Cerenkov photon distributions from EAS provides an effective simple $\gamma$–proton separation method, which can be used in a real-time selection procedure and also combined with other methods currently used to identify high energy $\gamma$ rays arriving from space. The same method has also been shown to be an useful tool to discriminate $\gamma$ rays from protons in the secondary charged particle distributions at ground level [15], providing similar values of the quality factor $Q$. Therefore, we can conclude that this method has a very huge applicability in primary cosmic rays identification, independently of the kind of secondary particle distribution considered at ground level.

Acknowledgement

This work is supported in part by Spanish Government grants for the research projects BFM2003-04147-C02 and FTN2003-08337-C04-04.

References