

PRINCIPAL COMPONENTS ANALYSIS OF EXTENSIVE AIR SHOWERS APPLIED TO THE IDENTIFICATION OF COSMIC TeV GAMMA-RAYS

E. FALEIRO¹

Departamento de Física Aplicada, EUIT Industrial, Universidad Politécnica de Madrid, 28012-Madrid, Spain

AND

J. M. G. GÓMEZ, L. MUÑOZ, A. RELAÑO, AND J. RETAMOSA

Departamento de Física Atómica, Molecular, y Nuclear, Universidad Complutense de Madrid, 28040-Madrid, Spain

Received 2004 April 27; accepted 2004 June 28

ABSTRACT

We apply a principal components analysis (PCA) to the secondary particle density distributions at ground level produced by cosmic γ -rays and protons. For this purpose, high-energy interactions of cosmic rays with Earth's atmosphere and the resulting extensive air showers have been simulated by means of the CORSIKA Monte Carlo code. We show that a PCA of the two-dimensional particle density fluctuations provides a decreasing sequence of covariance matrix eigenvalues that have typical features of a polynomial law, which are different for different primary cosmic rays. This property is applied to the separation of electromagnetic showers from proton simulated extensive air showers, and it is proposed as a new discrimination method that can be used experimentally for γ -proton separation. A cutting parameter related to the polynomial behavior of the decreasing sequence of covariance matrix eigenvalues is calculated, and the efficiency of the cutting procedure for γ -proton separation is evaluated.

Subject headings: cosmic rays — gamma rays: observations

1. INTRODUCTION

Extensive air showers (EASs) result from the interaction of high-energy particles and nuclei arriving from space, collectively called cosmic rays, with the high-altitude Earth atmosphere. Successive interactions of the primary particles and their secondaries give rise to a multiplicative cascade of elementary particles. If the energy of the incident cosmic ray is high enough, a considerable number of particles may reach the Earth's surface, where they are detected; at this point the secondaries are mostly e^+ , e^- , γ , and μ . A comprehensive description of EAS physics can be found in Sokolsky (1989) and Gaisser (1990).

The most frequent detection technique is based on an idea first used by Pierre Auger in the late thirties; it consists of a network of particle detectors (scintillators, water Cerenkov tanks, muon calorimeters) installed on the ground. The properties of the primary cosmic ray can be deduced from the lateral distribution of the secondaries in a cross section of the EAS. The parameters of such a ground array (altitude, surface area, spacing between the detector stations) must be adapted to the energy range aimed for.

A detector using a combination of two (or more) of these techniques (e.g., Cerenkov telescopes with a ground array) is called a hybrid. Such is the case of the Pierre Auger Cosmic Ray Observatory (Camin 2004), probably the largest presently operating air shower array now tracking high-energy cosmic rays. The Pierre Auger array detects particles falling anywhere over an area the size of Rhode Island state. This detection area of more than 3000 square kilometers lies just east of the Andes Mountains in Mendoza Province (Argentina).

Another example of hybrid detector is the HEGRA experiment (Karle et al. 1995; Aharonian et al. 2002), which is

located on the Roque de los Muchachos Observatory, in the Canary Island La Palma (Spain) at 2200 m above sea level. The scintillator array, of 4×10^4 m², has a threshold of about 40 TeV at 50% of acceptance. Other elements are an array of Geiger towers, a set of imaging Cerenkov telescopes (MAGIC), and a matrix of wide-angle air Cerenkov detectors.

The analysis of the EAS parameters supplies information on the nature, energy, and direction of the original cosmic ray, and its interactions in the atmosphere. It is thus known that at energies greater than 1 TeV, the great majority of the cosmic rays are nuclei of atomic elements, mostly protons, but also helium, oxygen and iron among others, with a small (less than 1%) proportion of electrons and photons. As is well known, charged particles and nuclei traveling through space constantly change their direction because of space magnetic fields; on the other hand, γ -rays always point to the source where they were produced, and thus they can be used in high-energy astronomy.

In real experiments it may be quite difficult to distinguish an electromagnetic shower (generated by a primary γ -ray) from a hadronic shower (generated by a primary proton or any other atomic nucleus). This separation constitutes a basic tool in high-energy astronomy when sources of high-energy γ -rays are sought. Therefore, several γ -hadron separation methods (Fegan 1997; Atkins et al. 2003) are used to help to identify the meager proportion of γ -initiated EASs among all the showers; most of them are based on the selection of appropriate system variables that are best for γ -proton separation. Some statistical tests have been designed to measure the separation degree of two samples of independent distributions; Student's t -test, the Kolmogorov-Smirnov test, or the Mann-Whitney U -test are examples of such statistical estimators. Bayesian methods and neural net analysis (Edwards et al. 1990; Chilingaryan 1995) are examples of special discrimination methods successfully applied to γ -hadron separation from Cerenkov photon images.

Recently, some rather simple and efficient γ -hadron separation methods based on the properties of the two-dimensional

¹ Corresponding author. Present address: Ronda de Valencia 3, 28012-Madrid, Spain; efaleiro@fais.upm.es.

power spectrum and fractal features of the secondary particle distribution at the detection level have been proposed (Faleiro & Contreras 1998; Faleiro et al. 2003, 2004). They take advantage of the power spectrum and higher order moment scaling properties of the particle density fluctuations to propose a separation parameter related to them.

In this paper we deal with the γ -hadron separation within the framework of the principal components analysis (PCA); specifically, we study the main properties of the covariance matrix eigenvalue spectrum and present a new simple and efficient method that provides a rather high separation power when it is applied to samples of different simulated EASs that have been generated close to experimental situations. The organization of the rest of this paper is as follows. In § 2 we present the PCA method and its adaptation to our data set characteristics. In § 3 we describe the databases we have used and how the autocovariance matrix is constructed and diagonalized. Section 4 addresses 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 γ -ray showers. Finally, the conclusions are summarized in § 5.

2. PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis is a popular data analysis method. One of the motivations for using PCA in practice is to reduce the dimensionality 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 previously been applied to data compressing and classification of spectral data of stars (Bailer-Jones et al. 1997) and galaxies (Glazebrook et al. 1998; Ronen et al. 1999). Furthermore, quantities measured from the spectra, such as broadband colors or equivalent width of lines, have been analyzed with this useful tool. From a most theoretical point of view and in accordance with our motivations, PCA may be applied to data representing generic signal samples with the scope of analyzing its correlation structure. In particular, PCA has been applied to the analysis of $1/f^\alpha$ noises (Gao et al. 2003), and it is conjectured that the decreasing ordered eigenvalue spectrum from PCA of a fractional Brownian motion process with parameter H decays as a power law $\lambda_n \sim n^{-(2H+1)}$.

From a general point of view, PCA deals with a set of M vectors each of $N \leq M$ components, $\{v_{ik}\}_{i=1, k=1}^{N, M}$, 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 vectors of the ensemble. The M vectors may represent M different realizations of a generic stationary or not stationary stochastic process, or a set of M objects each with N characteristic parameters. If we define

$$\mu_i = \frac{1}{M} \sum_{k=1}^M v_{ik}, \quad \sigma_i^2 = \frac{1}{M} \sum_{k=1}^M (v_{ik} - \mu_i)^2, \quad (1)$$

as the mean and the variance of the random variable associated with the i th component of the N -dimensional vector, the rescaled covariance matrix associated with the M vectors of size N can be defined as

$$C_{ij} = \frac{1}{M} \frac{\sum_{k=1}^M (v_{ik} - \mu_i)(v_{jk} - \mu_j)}{\sigma_i \sigma_j}. \quad (2)$$

In the following, we refer to it as the \mathbf{C} matrix. As $C_{ij} = C_{ji}$, and they are all real numbers, \mathbf{C} is a real symmetric matrix and thus the N eigenvalues are also real numbers. Moreover, with the normalization involved in equation (2), diagonal principal elements represent the single contribution $C_{ii} = 1$ of the i th component of the vector to the total variance $\sigma_i = N$.

Once \mathbf{C} is evaluated, the eigenvalue equation $\mathbf{C}\Psi_k = \lambda_k \Psi_k$ must be solved to obtain the N eigenvalues λ_k and the N eigenvectors Ψ_k of \mathbf{C} . In a standard PCA of the initial vector ensemble, only a small number of the N eigenvectors are relevant to describe the variability observed in the original data set; thus, we reduce dimensionality by projecting the initial vector ensemble onto the relevant eigenvectors. However, some initial data ensembles exhibit a decreasing ordered eigenvalue spectrum $\lambda_n \sim f(n)$ for the larger λ_n eigenvalues, which may characterize the data ensemble, especially when the function $f(n)$ depends on a few parameters. In this case, the eigenvalue spectrum from PCA is sufficient to classify the original data.

In this work we shall deal with two-dimensional data fields $F(x_1, x_2)$ representing the secondary particle density distributions at the point (x_1, x_2) from the core, taken as the center of a relative coordinate system. In real data, a discrete field $\{F(j_1, j_2) | j_1 = 1 \dots N_1, j_2 = 1 \dots N_2\}$ of $N_1 N_2$ data is available. As shall be detailed in the next section, from the two-dimensional discrete field $F(j_1, j_2)$, a one-dimensional representation composed of a large stationary series is obtained. In order to characterize each initial event, this one-dimensional whole series is then cut into M nonoverlapping pieces of N data points each, with $M \geq N$, which ensures an adequate estimator of the event covariance matrix. Generically, the length N is chosen according to some properties of the whole series such as some characteristic length of the system or the presence of structures of a typical size. However, when there is no characteristic length in the large stationary series, which stands for a sort of self-similarity of the whole series, then the length N of the pieces is of little significance since any piece possesses the main correlation properties of the whole.

3. COSMIC-RAY SIMULATION AND THE SINGLE-EVENT EIGENVALUE SPECTRUM

To simulate EASs we have used the version 6.019 of the CORSIKA Monte Carlo code (Heck & Knapp 2002). The hadronic interactions at high energies have been described by a model based on the parton-based Gribov-Regge theory to simulate ultrarelativistic heavy ion collisions. We have initiated EASs with high-energy γ -rays and protons with energies ranging from 20 to 70 TeV in steps of 10 TeV, 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. For each primary type at each energy, the number of events generated ranges from 1000 to 3000. More than 10,000 primary protons and approximately 5000 primary γ -rays constitute the global amount of simulated showers. As a simplification, all the showers have been generated vertically. The effect of a nonvertical 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, a uniform random distribution of the core position on the detector area is introduced, and also a random error in the reconstruction of the core location has been taken into account.

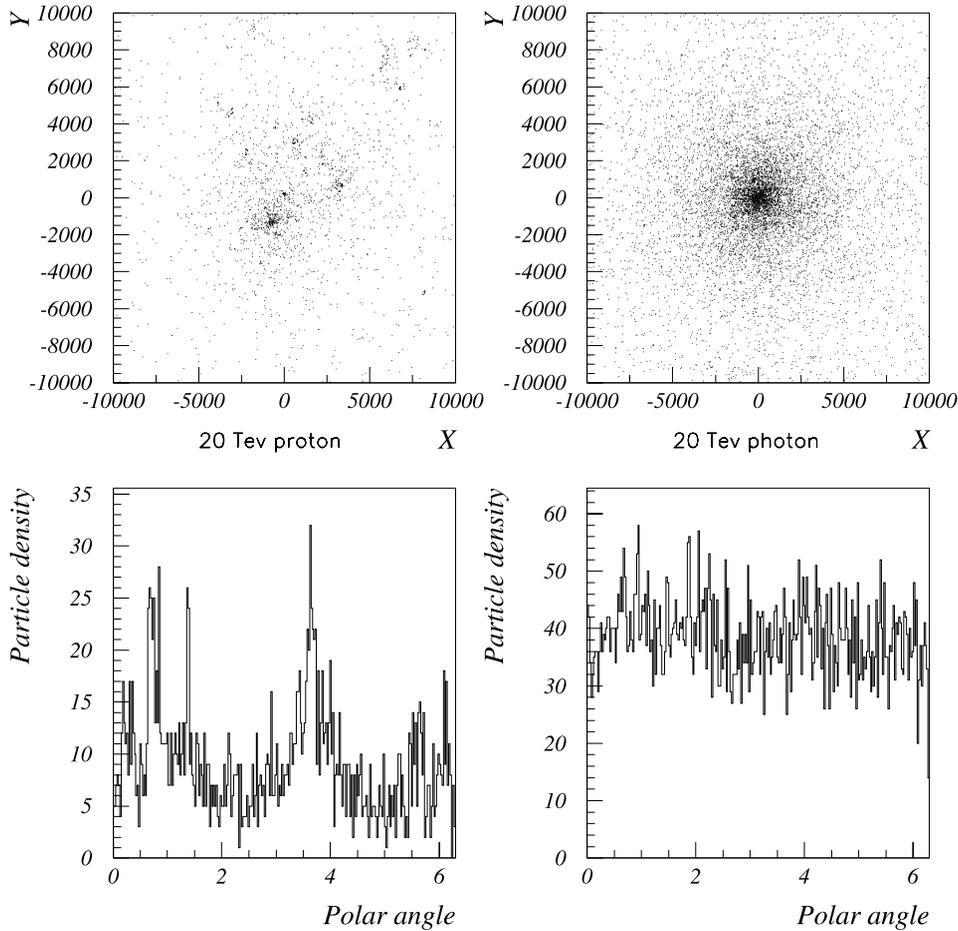


FIG. 1.—Two-dimensional e^+ and e^- particle density of simulated EASs initiated by a proton and a photon of 20 TeV each, in a squared detection area of 100 m of side. Their one-dimensional representations in terms of 256 sectors in the polar angle, ranging from 0 to 2π , is done below.

In any case, the core must not be too far from the center of the detection area to avoid a dramatic loss of information.

An event profile consists of a two-dimensional distribution of survival secondary e^+ and e^- collected at detection level. Figure 1 shows in the upper panels some examples of EAS particle distributions at the observation level; two events generated by a 20 TeV primary photon and a proton of the same energy having their first collision at different altitudes are displayed. These particle distributions can be considered as the sum of a Nishimura-Kamata-Greisen (NKG) (Greisen 1956; Kamata & Nishimura 1958) distribution centered at the core of the event and a mean zero fluctuating term. The analysis of this last component requires it to be separated from the NKG component to obtain a two-dimensional fluctuating field. Among other properties, the NKG distribution only depends on the core distance; therefore, the polar coordinates are the most convenient representation system for the two-dimensional cosmic-ray events $F(r, \theta)$.

The analysis of EAS particle distributions starts with the classification of primary cosmic rays. In simulated EASs 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 that is experimentally measurable. Here we classify the events by the number of secondary particles N_s at ground level, since this quantity can be measured in experiments. We have considered a binning of amplitude $\Delta N_s = 10,000$ secondary particles; Table 1 shows

the distribution of the simulated events into four of such bins. It is noted that the ΔN_0 bin starts with events containing more than 5000 secondary particles, since statistical calculations over events containing fewer particles give rise to biased estimators owing to an undersampling effect.

Once the original events have been classified according to their number of secondary particles, we proceed to reduce the polar two-dimensional profile to a one-dimensional representation $\rho(\theta)$ defined by

$$\rho(\theta) = \int_0^{r_{\text{eff}}} dr F(r, \theta), \quad (3)$$

where r_{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, it may be taken as the maximum radius from the core, where the secondary particles may be detected and accepted for analyzing purposes; in this work $r_{\text{max}} = 100$ m, which is close to the radius of the HEGRA experiment at La Palma. Nevertheless, for γ -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 θ leads us to characterize each EAS event as a series of 256 data length. This binning is chosen to have a long series without losing information about the fluctuations. In the lower panels of Figure 1 the one-dimensional representation of

TABLE 1
MAIN FEATURES OF THE EAS MONTE CARLO SIMULATION

Primary	N_p	ΔN_s	A_0	A_1	A_2
γ	640	ΔN_0 : 5000–10000	2.372	–1.573	0.502
	1080	ΔN_1 : 10000–20000	2.377	–1.577	0.511
	648	ΔN_2 : 20000–30000	2.384	–1.584	0.526
	333	ΔN_3 : 30000–40000	2.393	–1.596	0.544
Proton	1922	ΔN_0 : 5000–10000	2.709	–1.962	1.209
	1596	ΔN_1 : 10000–20000	2.737	–1.989	1.268
	523	ΔN_2 : 20000–30000	2.721	–1.971	1.234
	203	ΔN_3 : 30000–40000	2.726	–1.974	1.244

NOTES.—Main features of the EAS Monte Carlo simulation for several types of primaries with energies ranging from 20 to 70 TeV. Here N_p is the number of primary cosmic rays whose number of secondary particles arriving to the detection surface is within the bin ΔN_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.

the upper panel 20 TeV events is displayed. Among other properties, the series is stationary with a mean value representing the mean particle density for the r_{eff} chosen.

As it was pointed at the end of the previous section, to ensure an adequate estimator of the event covariance matrix, the one-dimensional series representing a single event must be cut into M nonoverlapping pieces of length N . To select the size N of these nonoverlapping pieces, we proceed to investigate the correlation structure of the whole series by means of its power spectrum. In Figure 2 we show the mean power spectrum of protons and photons belonging to the ΔN_0 database. In accordance with the previously reported scaling properties of such a series (Faleiro & Gómez 1999, 2001), the absence of underlying characteristic structures allows us to consider a whole series decomposition in 16 nonoverlapping pieces of 16 data points each. The covariance matrix is then constructed, as equation (2) indicates, and diagonalized according to stan-

dard numerical recipes (Press et al. 1992). Figure 3 shows the mean decreasing ordered eigenvalue spectrum λ_n as a function of the rank n from the proton and also photon samples belonging to the same ΔN_2 bin. From Figure 3 it is clear that proton series possess a dominant eigenvalue carrying away a significant percentage of the total variance; for ΔN_2 proton databases the largest mean eigenvalue represents around the 40% of the total variance. On the other hand, in photon series the variance is smoothly 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 starting point of the identification method we will present in the next section. This is quantitatively confirmed by fitting the mean eigenvalue spectrum to a three-parameter Chebyshev-series polynomial approach, the details

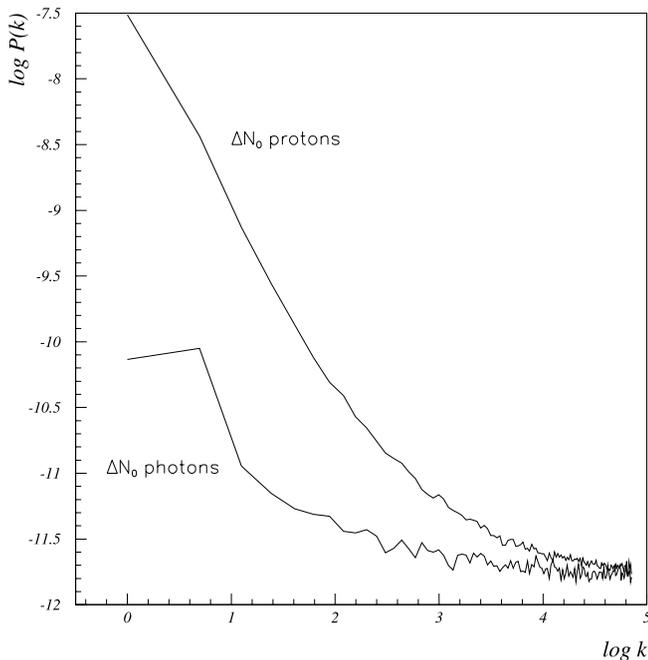


FIG. 2.—Comparison of mean power spectrum in a log-log scale for the ΔN_0 samples of protons and photons. They may be fitted to the model $P(k) = a/k^b + c$, with constants a , b , and c that make clear their underlying scaling properties (Faleiro & Gómez 1999, 2001).

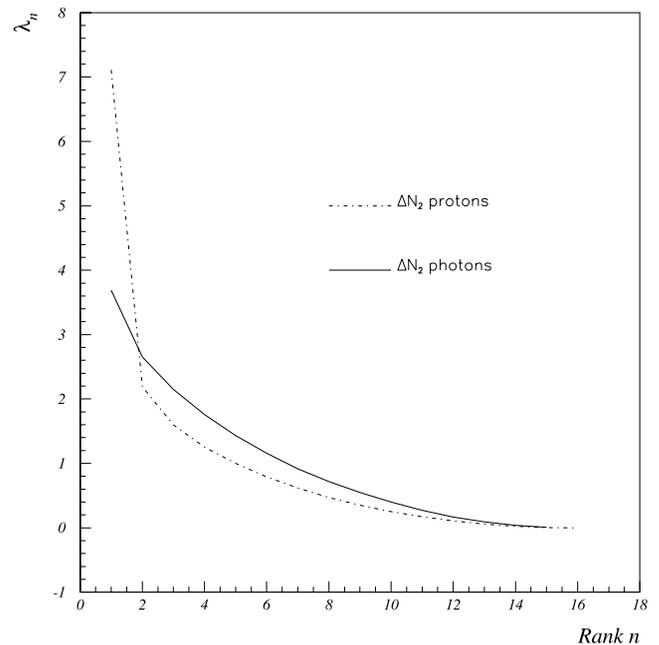


FIG. 3.—Decreasing ordered eigenvalue spectrum corresponding to the ΔN_2 databases. Forty percent of the total variance is carried by the largest eigenvalue for the proton sample. The variance is smoothly distributed among all the eigenvalues for the photon sample instead.

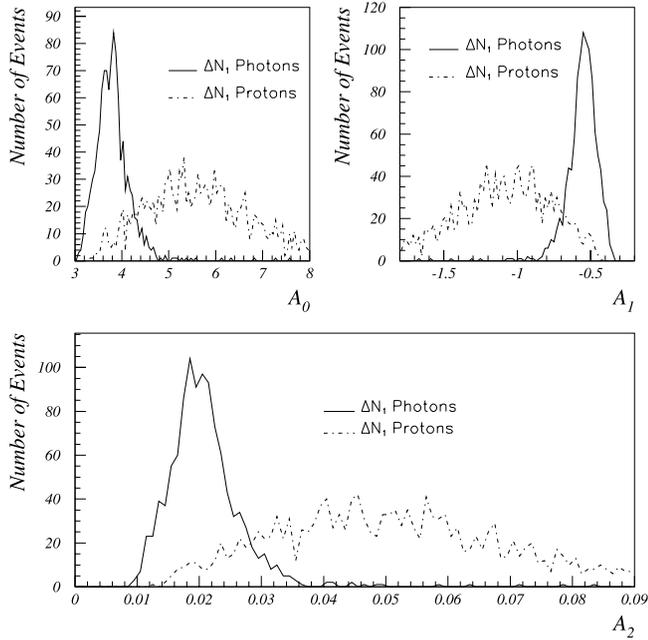


FIG. 4.—Distribution of the values for parameters A_0 , A_1 , and A_2 from the ΔN_1 database. The panels show that any of them may be used in separation algorithms, as confirmed by the results of the three statistical tests (see text).

of which will be given in the next section. Parameter values A_0 , A_1 , and A_2 are shown in Table 1.

4. THE γ IDENTIFICATION METHOD

Provided that the mean eigenvalue spectrum exhibits different features for different primary cosmic rays, we proceed to reduce the eigenvalue spectrum structure of each event to a set of parameters by fitting the eigenvalue spectrum to some adequate functional form. We have tried several multiparameter dependent functions, and we have found that with three or more parameters a rather similar separation power is obtained; thus, we restrict the model to three-parameter functions. The most significant among them are a second-order polynomial form and a second-order Chebyshev-series approximation. We have chosen the least-squares polynomial approximation by Chebyshev series method since it provides a slightly better separation power.

The procedure consists of analyzing the A_0 , A_1 , and A_2 parameters as a result of a fitting procedure to

$$\lambda_n = \frac{1}{2}A_0T_0(n) + A_1T_1(n) + A_2T_2(n), \quad (4)$$

where λ_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 γ -rays from primary protons.

The separation power of a parameter is frequently measured by the quality factor Q defined by (Schafer et al. 2001; Bock et al. 2001)

$$Q = \frac{\kappa_\gamma}{\sqrt{\kappa_p}}, \quad (5)$$

where κ_γ and κ_p are the fractions (efficiencies) of photons and protons kept by the algorithm, namely, $\kappa_\gamma = N_\gamma^{\text{cut}}/N_\gamma$ and $\kappa_p = N_p^{\text{cut}}/N_p$, where N^{cut} is the number of remaining particles after the cutting procedure. Some authors (Bock et al. 2001)

TABLE 2
P-VALUES

Parameter	t	U	K
A_0	42.33	-28.76	0.826
A_1	42.66	-28.43	0.824
A_2	42.34	-28.76	0.825

NOTES.— P -values of the null hypothesis about distinction of ΔN_1 proton and photon samples calculated from the Student's t -test, the Mann-Whitney U -test, and the Kolmogorov-Smirnov test statistics. 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.

recommend that Q could be replaced as a quality factor by the statistical significance

$$\sigma = \frac{S}{\sqrt{2B + S}}, \quad (6)$$

where S and B are, respectively, the number of γ and proton events in the sample retained after cuts. Nevertheless, it is easy to check that σ depends on the size of the chosen samples. Thus, Q will be used in this work as a measure of the enrichment in primary photons of the original sample after the cutting procedure.

Figure 4 shows the distribution of the A_0 , A_1 , and A_2 parameters for γ and protons belonging to the ΔN_1 database. The P -values of Student's t -test, the Kolmogorov-Smirnov test, and the Mann-Whitney U -test of these parameters are calculated for all γ and proton samples. The P -values are quantitative characteristics of the measure of distinction between distributions (Zacks 1977). Larger P -values correspond to the smaller probability of accepting the null hypothesis, which states that the samples belong to the same population. Table 2 shows the P -values of these three statistical tests for the ΔN_1 bin; quite 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 linearly correlated. Thus, we have chosen the A_2 parameter since it provides a slightly better separation power.

We concentrate our efforts on the A_2 parameter and study the influence of r_{eff} (the maximum radius from the core where the secondary particles may be detected and accepted for the analysis and introduced in eq. [3]) 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 in accordance with the A_2 parameter when r_{eff} varies from a few meters to the maximum available value $r_{\text{eff}} = 100$ m. This is done by binning the complete range of r_{eff} (0–100 m) into 32 equal bin units (b.u.). For a fixed value of r_{eff} the one-dimensional representation of each event is obtained in accordance with equation (3) and the separation procedure is applied. Figure 5 clearly shows that the quality factor reaches a maximum value around $r_{\text{eff}} = 16$ b.u. (50 m) for all the databases. With the exception of ΔN_3 , whose Q decays from $r_{\text{eff}} = 16$ b.u., the quality factor is approximately constant for the rest of databases. Thus, the separation procedure can be improved by taking in any case $r_{\text{eff}} = 16$ b.u. as the maximum of r_{eff} , obtaining the so-called optimized databases.

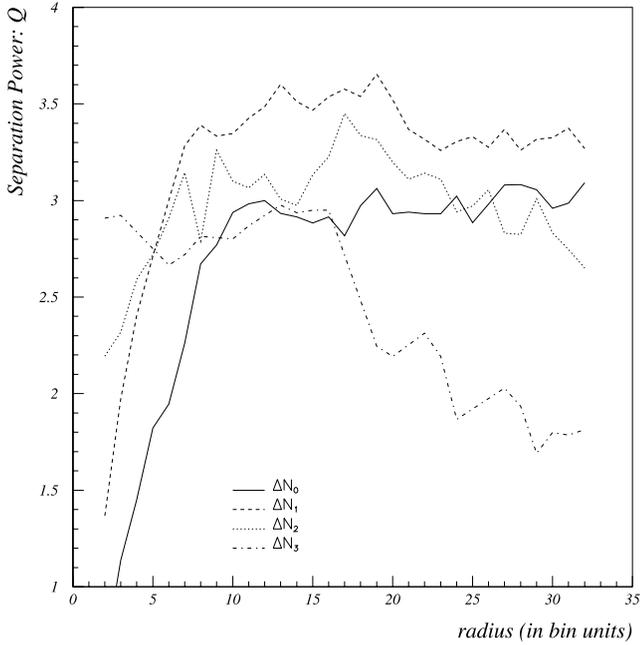


FIG. 5.—Quality factors Q of γ -proton discrimination, as a function of the value of r_{eff} . With the exception of the ΔN_3 database, all the rest reach a stationary value for Q from $r_{\text{eff}} \approx 16$ (b.u.). A bin unit is equivalent to 3.125 m.

With the optimized databases, the core of each event is placed randomly on the detection area and the r_{eff} is evaluated as the radius from the core of the shower front circular surface that remains inside the detection area; thus, the one-dimensional events representation is obtained. 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_2 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_2 . The mean values of such distributions must be considered as the representative value of the separation power and the separation parameter, respectively. Thus, considering 1000 realizations, the distributions of Q and A_2 for the γ -proton separation inside the ΔN_1 database are shown in Figure 6; the mean values are also displayed in the figure. The A_2 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 κ_γ , κ_p , and Q involved in the separation procedure when the cut is performed. As a final result, a γ enrichment measured by $Q \approx 3$ is obtained in any case; that represents a signal-to-background ratio rather suitable to the development of high-energy cosmic gamma ray astronomy.

5. CONCLUSIONS

We have generated several Monte Carlo samples of extensive air showers initiated by protons and γ -rays at energies between 20 and 70 TeV, and we have studied the fluctuations of the two-dimensional secondary charged particle density at detection level. The EAS events are classified in several bins ΔN_s according to the total number of secondary particles N_s arriving to the detection area. From the polar coordinate system representation of each event $F(r, \theta)$, we have obtained a stationary one-dimensional series by integrating over the radial

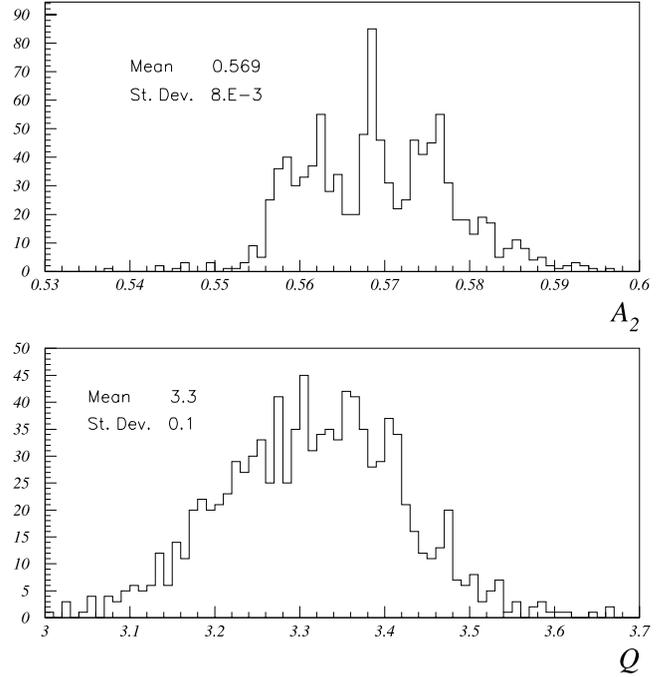


FIG. 6.—Distribution of the values for A_2 and Q from the ΔN_1 database when 1000 realizations of a uniform random core distribution over the detection area are considered.

coordinate up to an effective radius r_{eff} , which essentially depends on the core position on the detection area. We also have proof that the maximum value of r_{eff} , which stands for including inside the series the secondary particles far from the core, can be adequately chosen to improve the γ -proton discrimination power of the proposed method. The γ -proton discrimination is important for high-energy astronomy, since only electromagnetic EASs provide information on the direction of the source of the cosmic rays.

The correlation structure of all the recorded EASs is studied by means of the rescaled covariance matrix \mathbf{C} , which is then diagonalized and its eigenvalue spectrum is analyzed. This constitutes the basis of the PCA method, which is used to determine the optimal variables to reduce the dimensionality of a set of initial data. With the aim of classifying each EAS event according to the primary cosmic ray that gives rise to it, the one-dimensional series representing each event is cut into non-overlapping pieces and the decreasing ordered eigenvalue spectrum λ_n of the associated \mathbf{C} matrix is evaluated and parameterized using a second-order Chebyshev polynomial approach. As some statistical tests reveal, the value of the A_2 coefficient becomes the most adequate parameter for discrimination purposes. The enrichment in primary cosmic γ -rays is measured by the quality factor Q associated with a cut in some separation parameter associated with a system feature. Thus, a cut in A_2 that preserves a fixed percentage of the proton background gives rise to an enrichment in photon events measured by the quality factor Q , while a fraction κ_γ of initial photons is kept. Provided that a random core distribution of each database is considered 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_2 . Table 3 shows the mean values of such distributions providing the final results of the analysis.

We conclude that the study of the covariance matrix eigenvalue spectrum of extensive air showers, a useful tool included

TABLE 3
EAS EVENTS

ΔN_s	A_2	Q	κ_γ (%)	κ_p (%)
ΔN_0 : 5000–10000.....	<0.509	2.7 ± 0.4	59.84 ± 0.04	5.09 ± 0.01
ΔN_1 : 10000–20000.....	<0.569	3.3 ± 0.1	74.53 ± 0.03	5.07 ± 0.03
ΔN_2 : 20000–30000.....	<0.575	3.1 ± 0.2	68.73 ± 0.04	5.03 ± 0.02
ΔN_3 : 30000–40000.....	<0.566	2.7 ± 0.3	61.68 ± 0.05	5.15 ± 0.03

NOTES.—The set of all EAS events having a value of the A_2 parameter shorter than the given value of A_2 contains a fraction κ_p of the protons and a fraction κ_γ of the γ -rays. The quality factor Q corresponding to the given value of A_2 is also shown.

inside the PCA method, provides a new, effective simple γ -proton separation method that can be used in a real-time selection procedure and also combined with other methods currently used to identify high-energy γ -rays arriving from space. The main purpose of this paper is to point out the existence of a structure in the rescaled covariance matrix associated with each EAS event that can be used to discriminate γ -rays from protons and to quantify to some extent its discrimination power; but further work is needed to explore the influence of factors such as

the effect of a limited detection area covered by real detectors and the control of detector device noise. Finally, the role played by other different primary types constitutes another interesting matter in the application to real experiments.

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

REFERENCES

- Aharonian, F. A., et al. 2002, *A&A*, 390, 39
 Atkins, R., et al. 2003, *ApJ*, 595, 803
 Bailer-Jones, C. A. L., Irwin, M., Gilmore, G., & von Hippel, T. 1997, *MNRAS*, 292, 157
 Bock, R. K., et al. 2001, *Nucl. Instrum. Methods Phys. Res.*, 516, 511
 Camin, D. V. 2004, *Nucl. Instrum. Methods Phys. Res.*, 518, 172
 Chilingaryan, A. A. 1995, *Pattern Recognition Lett.*, 16, 333
 Edwards, W., Lindman, H., & Savage, L. J. 1990, in *Robustness of Bayesian Analyses*, ed. J. B. Kadane (North Holland: Elsevier)
 Faleiro, E., & Contreras, J. L. 1998, *J. Phys.*, G24, 1795
 Faleiro, E., & Gómez, J. M. G. 1999, *Europhys. Lett.*, 45, 437
 ———. 2001, *Fluctuation & Noise Lett.*, 1, L117
 Faleiro, E., Gómez, J. M. G., & Relaño, A. 2003, *Astropart. Phys.*, 19, 617
 Faleiro, E., Gómez, J. M. G., Relaño, A., & Retamosa, J. 2004, *Astropart. Phys.*, in press
 Fegan, D. J. 1997, *J. Phys. G*, 23, 1013
 Gaisser, T. K. 1990, *Cosmic Rays and Particle Physics* (Cambridge: Cambridge Univ. Press)
 Gao, J. B., Cao, Y., & Lee, J.-M. 2003, *Phys. Lett. A*, 314, 392
 Glazebrook, K., Offer, A. R., & Deeley, K. 1998, *ApJ*, 492, 98
 Greisen, K. 1956, *Prog. Cosmic Ray Phys.*, 3, 3
 Heck, D., & Knapp, J. 2002, *Extensive Air Shower Simulation with CORSIKA* (Karlsruhe: Inst. Kernphys.)
 Kamata, K., & Nishimura, J. 1958, *Prog. Theor. Phys. Suppl.*, 6, 93
 Karle, A., et al. 1995, *Astropart. Phys.*, 4, 1
 Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical Recipes in FORTRAN* (Cambridge: Cambridge Univ. Press)
 Ronen, S., Aragon-Salamanca, A., & Lahav, O. 1999, *MNRAS*, 303, 284
 Schafer, B. M., et al. 2001, *Nucl. Instrum. Methods Phys. Res.*, 465, 394
 Sokolsky, P. 1989, *Introduction to Ultra-high Energy Cosmic Ray Physics* (New York: Addison Wesley)
 Zacks, S. 1977, *The Theory of Statistical Inference* (New York: Wiley)