

A multivariate study of mass composition for simulated showers at the Auger South Observatory.

G. A. Medina-Tanco and S. J. Sciutto

Instituto Astronomico e Geofisico, Universidade de São Paulo, Brazil

Departamento de Física, Universidad Nacional de La Plata, C. C. 67, 1900 La Plata, Argentina

Abstract. The output parameters from the ground array of the Auger South observatory, were simulated for the typical instrumental and environmental conditions at its Malargüe site using the code sample-sim. Extensive air showers started by photons, protons and iron nuclei at the top of the atmosphere were used as triggers. The study utilized the air shower simulation code Aires with both QGSJet and Sibyll hadronic interaction models. A total of 1850 showers were used to produce more than 35,000 different ground events. We report here on the results of a multivariate analysis approach to the development of new primary composition diagnostics.

1 Introduction

The experimental detection of ultra high energy cosmic rays ($E > 10^{20}$ eV) poses some of the most exciting problems in modern astrophysics. Up to now no astrophysical objects are known that could accelerate charged particles to such energies. If the sources are located on cosmological distances, then it would be expected that the Cosmic rays arriving to the Earth will loose energy after interacting with the cosmic microwave background, until reaching a threshold energy of about 6×10^{19} eV. This energy would therefore mark a sharp end of the Cosmic Ray spectrum. No such sharp end is seen by experiment so far. If the sources are nearby, then an anisotropic distribution of arrival directions is expected because in this case the directions of arrival would point to the sources.

Alternative explanations of the existence of the Ultra high energy Cosmic Rays have been developed by theorists over the last few years: New particles, new physics or exotic phenomena, such as decaying topological defects, or the violation of Lorentz invariance.

To effectively check any of these "classic" or alternative theories it is necessary to measure with adequate statistics

the highest energy Cosmic Rays. It is necessary to accurately determine the form of the spectrum, the distribution of arrival directions over the whole sky, and the identity of the particles.

The Auger Observatory (Auger Collaboration, 1997) has the aim of collecting enough experimental data to give appropriate answers to those questions. It consists in two detectors of 3000 km² each, positioned on the Southern and Northern hemispheres. Each detector will be capable of measuring the properties of the showers generated by the ultra high energy cosmic rays. An array of surface detectors (SD) will measure the characteristics of the shower particles reaching ground level, while a fluorescence detector will measure the light emitted after the interaction of the shower particles with the atmosphere.

The development of extensive air showers (EAS), as characterized by lateral distribution, curvature of the shock front, rising time, pulse shape, total number of photoelectrons, etc., carry information regarding the direction, energy and identity of the incoming primary. However, while direction and energy can be estimated rather easily from ground array data (e.g. Billoir (2000)), the definition of a convenient and efficient diagnostic for primary identity discrimination remains a challenging issue.

In particular, besides some punctual indications against UHE photons as primaries Bird et al. (1995); Halzen et al. (1995); Nagano et al. (1999), only one comprehensive study limiting the photon flux above 10^{19} eV has been published Ave et al. (2000) up to now, and it is based on an analysis of inclined showers at Haverah Park (zenith angles $> 60^\circ$). The separation between light (protons) and heavier (Fe nuclei) hadrons is still much more difficult.

In this paper we present preliminary results of an ongoing effort to develop primary identification diagnostics with the aid of multivariate techniques. A pragmatic approach is taken to the practical problem of statistically determining the identity of the primaries starting EAS at the top of the atmosphere with the ground array of the Auger observatory as the specific target.

2 Numerical method and discussion

A large sample of showers for primary photons, protons and iron nuclei is generated with the AIRES code and, transformed into ground array events of a model Auger observatory, used to trigger the surface detectors, simulated with the sample-sim code.

The AIRES system is a set of programs to produce simulations of air showers, and to analyze the corresponding data. All the relevant particles and interactions are taken into account during the simulations, and a number of observables are measured and recorded, among them, the longitudinal and lateral profiles of the showers, the arrival time distributions, and detailed lists of particles reaching ground that can be further processed by detector simulation programs. The AIRES system is explained in detail elsewhere (Sciutto, 2001, 1999).

The showers processed in this work were generated with the AIRES system, and consist in a series of 1831 proton, gamma, and iron showers, with energies in the range $10^{17.5}$ eV to $10^{20.5}$ eV, and zenith angles in the range 0 to 60 degrees. Each shower is reused 20 times at different location in the array, and so the final number of available events is 36620. The hadronic models used are QGSJET (Kalmykov et. al., 1997) and Sibyll (Fletcher et. al., 1994).

The surface detectors have been simulated using the "sample-sim" SD simulation program (?).

The directly observable output for each event, which include the number and spatial distribution of triggered tanks and the time profile of the signal at each station, together with more easily reconstructed quantities (e.g., energy and zenith angle) are used to define different sets of parameters. Each set of parameters constitutes an n-dimensional orthogonal space which is later studied using principal component analysis (PCA) in search for primary separation.

The PCA method simply performs a rotation in the n-dimensional space to a new orthogonal coordinate system whose unit vectors are the eigenvectors of the system. These new axes have a special meaning, since their associated eigenvalues are a measure of the dispersion of the data along each axis. Thus, the principal eigenvector has the largest associated eigenvalue, and therefore the largest dispersion, or information content, of the sample; the second eigenvector has the second largest dispersion and so on. Typically, one can quantify the amount of information associated with a subset of axes, and can even expect to uncover the true dimensionality of the system if this has been overestimated.

One advantage of the PCA method is that, involving only rotations, the new axes are only linear combinations of the original magnitudes.

As an illustrative example, let's take a parameter space defined arbitrarily by:

- (1) a (sort of) curvature estimator,

$$P_1 = \left[\frac{\langle T_{0,ext} \rangle - \langle T_{0,int} \rangle}{\langle r_{ext} \rangle - \langle r_{int} \rangle} \right] \times \sin \theta \quad (1)$$

where the subscripts "ext" and "int" refer to stations that are farther away and nearer the shower axis than the median distance r_c of the triggered stations, and r_{ext} and r_{int} are the average distances inside each region.

- (2) the third largest total number vertical equivalent muon,

$$P_2 = [(vem)_{total}]_{3rd} \quad (2)$$

- (3) pulse shape/rising time (average),

$$P_3 = \left\langle \frac{T_{50}}{T_{10} + T_{50}} \right\rangle \quad (3)$$

where T_i are the fluence times for 10% and 50% of the total fluence at a given station.

- (4) pulse shape/rising time (3rd largest value),

$$P_4 = (T_{10} + T_{50} + T_{90})_{3rd} \quad (4)$$

- (5) (sort of) lateral distribution,

$$P_5 = \left[\frac{(vem)_{total}}{P_4} \right]_{5th} / \left[\frac{(vem)_{total}}{P_4} \right]_{3rd} \quad (5)$$

- (6) median of the station distances to the axis of the shower, $P_6 = r_c$,

- (7) rising time (3rd largest value),

$$P_7 = \left(\frac{T_{10} - T_0}{T_{90} - T_0} \right)_{3rd} \quad (6)$$

- (8) primary energy $P_8 = E$,

- (9) zenith angle $P_9 = \theta$,

- (10) number of triggered stations $P_{10} = N_{stat}$,

All these parameters are later normalized so that their dynamical ranges are in the interval $(-1, 1)$.

When a PCA analysis is performed in this parameter space, it is found that the first 4 eigenvectors are responsible for $\sim 80\%$ of the variance (or information content) of the system. The 7th eigenvector is responsible for only $\sim 6\%$ of the variance.

The best separation between nuclei and photons is obtained for the projection onto the plane defined by the first and seventh eigenvectors (see figure 1). The thick line, $EV_7 = -48.89 \times (EV_1 + 0.007)^2 + 0.011$, leaves only 0.8% of the nuclei in the region of photons and 12% of the photons in the region corresponding to nuclei. Therefore, the probability of misidentifying a photon is 2.7% and the probability of misidentifying a nucleus is 3.8%.

Once the photons have been separated, the same process can be applied to nuclei alone. However, as was stated before, this is a much more complicated problem as shown in

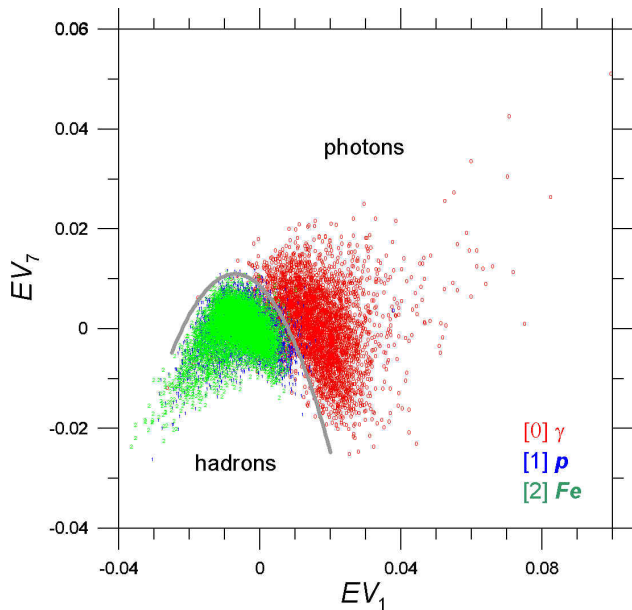


Fig. 1. PCA results on the illustrative parameter space. The best separation between nuclei and photons is obtained for the projection on the plane defined by the first and seventh eigenvectors. The thick line misclassifies 3.8% of the nuclei as photons and 2.7% of the photons as nuclei.

figure 2. The optimization of of a diagnostic method in this case is still ongoing work.

Despite these complications, multivariate analysis techniques look promising and, together with detailed numerical simulations of EAS and detectors, can be a powerful tool to identify primary CR.

Acknowledgements. This work is partially supported by the Brazilian agencies FAPESP and CNPq .

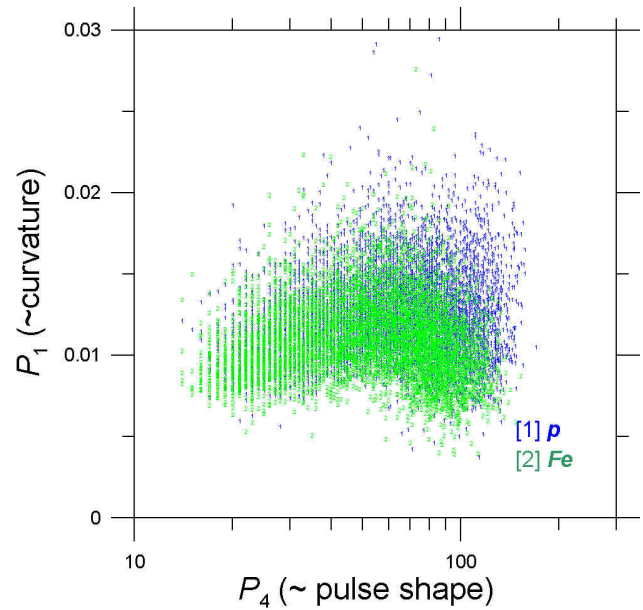


Fig. 2. Projection onto the P_1 – P_4 plane of the sample points, once the photon events have been extracted, showing the difficulty involved in the separation of light and heavy nuclei.

References

- Auger Collaboration, Auger Project Design Report, FNAL (1997)
- Ave M., Hinton J. A., Vázquez R. A., Watson A. A., Zas E., *Phys.Rev.Lett.* 85 (2000) 2244.
- Billoir P., 2000, GAP note 25.
- Bird D. et al., *Astrophys. J.*, **441**, 144 (1995).
- Fletcher, R. T., *et. al.*, *Phys. Rev. D*, **50**, 5710 1994.
- Halzen F., Vazquez R. A., Stanev T., Vankov H. P., *Astroparticle Phys.*, **3**, 151 (1995).
- Kalmykov, N. *et. al.*, *Nucl. Phys. B*, **52B** 17, 1997.
- Nagano M., Heck D., Shinozaki K., Inoue N., Knapp J., *Astroparticle Phys.* (to be published), astro-ph/9912222.
- Sciutto, S. J., these proceedings.
- Sciutto, S. J., preprint astro-ph/9911331, 1999.