Proceedings of the 30th International Cosmic Ray Conference
Rogelio Caballero, Juan Carlos D'Olivo, Gustavo Medina-Tanco,
Lukas Nellen, Federico A. Sánchez, José F. Valdés-Galicia (eds.)
Universidad Nacional Autónoma de México,
Mexico City, Mexico, 2008

Vol. 4 (HE part 1), pages 531–534

30TH INTERNATIONAL COSMIC RAY CONFERENCE

ICRC'07
Mérida, México

New statistical parameters for mass composition studies

F. CATALANI[1],J.A. CHINELLATO[1], V. DE SOUZA[2], J. TAKAHASHI[1].
[1] *Instituto de Física Gleb Wataghin, Universidade Estadual de Campinas, Brasil*
[2] *Institut für Experimentelle Kernphysik, Universität Karlsruhe, 76021 Karlsruhe, Germany*
catalani@ifi.unicamp.br

Abstract: The determination of the mass composition of the ultra high energy cosmic rays is essential to many open questions in astroparticle physics. However the identification of the cosmic ray primary particle is a complex task due to several difficulties such as the large fluctuations in the shower development and the low number of experimental measurables. We present a proposal for composition studies applying multivariate analysis to make use of several parameters extracted from the longitudinal development of the showers to improve primary particle identification. Measurable features of the CR shower longitudinal profile such as the N_{max}, X_{max}, asymmetry, kurtosis, and skewness were combined using linear discriminant analysis (LDA). Studies were done using cosmic ray showers simulated by the CONEX code considering gamma, proton, helium, carbon and iron as primary particles.

Introduction

The mass composition of ultra high energy cosmic rays (UHECR) plays a fundamental role in the understanding of their origin, acceleration and propagation mechanisms. In the energy region of cosmic ray spectra around 10^{15}, known as the knee, the most probable source for these particles are the galactic supernova remnants (SNR). However, for higher energies there are no known sources inside our galaxy that could be capable of accelerating particles to such energies and thus a transition to extragalactic origin is expected.

The exact energy point where the transition from galactic to extragalactic component occurs and the chemical abundance are the two main parameters in the astrophysical models trying to describe this high energy region. Reference [1] summarizes the picture by exploring this two parameters and comparing them to the data measured by the HiRes experiment [2].

In fluorescence detectors, the composition of the cosmic ray is achieved by studying, on a statistical basis, the position in the atmosphere (X_{max}) where the shower has its maximum number of particles. In a previous paper [3] we have presented a method for chemical composition studies based on the application of a statistical method known as Linear Discriminant Analysis (LDA) to enhance the separation between proton and iron primary particle. We made use of several features of the longitudinal development of the CR shower such as the N_{max}, X_{max}, asymmetry, kurtosis, and skewness rather than using only the depth of the shower maximum (X_{max}). In the previous work, the method was tested using simulated showers of proton, iron and photons generated by the CORSIKA simulation code.

In this paper, we extend our studies testing the same method with showers generated by the CONEX code. We have also studied the separation capability of the method for other primary types. Helium and carbon primaries were included in this study. Besides, the proton, iron and gamma separation shown in the previous paper, LDA has shown a good identification power for carbon primaries which underlines its applicability due to the astrophysical significance of this particle type in acceleration models taking place in SNR.

Shower Longitudinal Profile

We simulated showers of photon, proton, helium, carbon and iron primaries with energy of 10^{18} eV

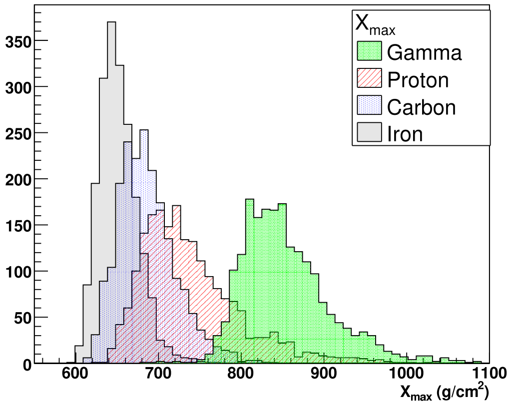

Figure 1: Distribution of the shower maximum depth X_{max} for simulated showers initiated by gamma, proton, carbon and iron showers

using the CONEX program [4]. CONEX is a hybrid Monte Carlo(MC) program that combines a MC treatment together with the solution of cascade equations. The hadronic interaction model used was QGSJETII [5] and the shower longitudinal development was sampled in steps of 5 g/cm^2 in slant depth. We have simulated and analyzed 2000 showers for each primary particle species arriving at 60° zenith angle.

Figure 1 shows the distribution of the depth of the shower maximum X_{max} for the different primary particles. From left to right, we have the distributions of iron, carbon, proton and gamma sequentially. It is clear that the X_{max} parameter shows some discrimination capability, specially between hadrons and gamma initiated showers. However its discrimination power is poor within different hadrons as can be seen by the small separation of the mean values and the large overlaps of the distributions.

We have included in our studies, CR showers initiated by helium nuclei, but they are not included in the plots due to the small separation between proton and helium showers distributions. To quantify the separation capability between two distributions, hence the discrimination between the different primary particles, we have chosen to use the merit factor (MF) statistical parameter that is defined as:

$$MF = \frac{\bar{A} - \bar{B}}{\sqrt{\sigma_A^2 + \sigma_B^2}}, \tag{1}$$

where $\bar{A}$ and $\bar{B}$ are the distributions averages, and σ_A and σ_B the respective standard deviations.

The distributions shown in figure 1 yields a separation merit factor of 1.4 between proton and gamma initiated showers, 0.7 between proton and carbon initiated showers and 1.3 between proton and iron initiated showers. These values are similar to the values obtained in [3], where we have used showers simulated by the CORSIKA code. Also, in the same work, we have presented a study that shows the dependence of the merit factor with the number of events and for different proton/iron relative abundance. For a distribution of 2000 events we have shown that the error in the merit factor is below 5% in despite of the proton/iron relative abundance.

The longitudinal charged particle profile were analyzed and the following composition sensitive parameters were determined:

$\underline{X_{max}}$:the atmospheric depth (g/cm^2) in which the shower has the maximum number of particles. It is the most used composition parameter and is calculated in any analysis procedure of fluorescence telescopes data. In order to avoid fitting particularities and not to bias our analysis by asymmetries of a particular function we have determined X_{max} by searching the bin with the greatest number of particles in the simulated profile. This would represent a maximum error in the X_{max} determination of 5 g/cm^2 which is the sampling step of the simulation code.

$\underline{N_{max}}$: the number of particles in the shower at X_{max}. This is also a standard parameter calculated in any fluorescence telescope analysis and it is directly proportional to the shower energy [6]. If the error in the energy reconstruction is large the inclusion of this parameter in the composition study would lead to a dependence with energy that is hard to disentangle. However, the fluorescence telescopes reconstruct the energy with an error of about 15% what we believe is a safe margin since variations in the energy of an EAS of this order do not affect the hypothesis about the chemical composition of the primary.

Asymmetry and Sigma: in order to measure the asymmetry of the distribution we have fit an asymmetric function to the longitudinal profile defined as:

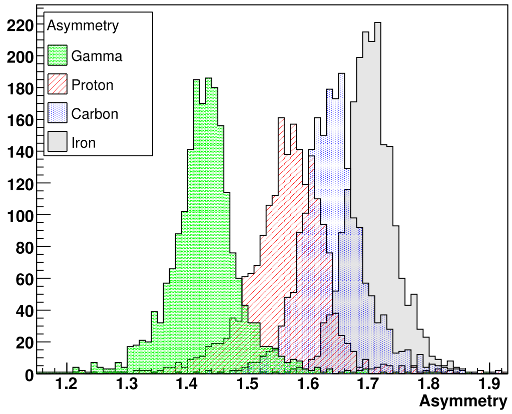

Figure 2: Distribution of the shower longitudinal profile asymmetry for simulated showers initiated by gamma, proton, carbon and iron showers

if ($X < X_{max}$)

$$N_{part} = N_{max} \exp \frac{-(X - X_{max})^2}{Sigma^2}$$

if ($X > X_{max}$)

$$N_{part} = N_{max} \exp \frac{-(X - X_{max})^2}{Asymmetry^2 * Sigma^2}$$

X_{max} and N_{max} are fixed in the fit to the values given by a direct scan of the simulated longitudinal profile. Asymmetry and sigma are the only two variables allowed to vary in the fit procedure. The asymmetry variable is a direct measure of the difference between the parts of the shower below and above X_{max}. Sigma gives a measure of the width of the shower.

Skewness: is the third moment of the distribution and is also a measurement of the asymmetry of the longitudinal distribution.

Kurtosis: is the fourth moment of the distribution and is a combined measurement of the size of the peak and the tails.

Figure 2 shows the distribution of the asymmetry of the different shower profiles. With this parameter, we achieve a separation merit factor of 2.1 between proton and iron showers and 2.0 between proton and photon showers, which is a better separation then the X_{max} parameter.

Linear discriminant analysis

We have studied the separation capability of all the different shower profile parameters mentioned above, and have combined them using a statistical method for event discrimination known as Linear Discriminant Analysis (LDA) [7]. LDA is a statistical discrimination method used to find a function of linear combinations of variables that maximizes the separation between two or more classes of objects or events.

It accomplishes that by minimizing the error of event classification, assigning an event to the group of highest conditional probability (Bayes theorem). LDA calculates this probability theoretically assuming that each group has gaussian multivariate distributions and the covariance matrix is the same for all groups. After this assumptions LDA writes a score for each class that are a linear combination of the dataset. The maximum score classifies the event in the respective population. In this analysis, we have used all six parameters of the shower longitudinal profile to obtain a discriminant coefficient calculated using the proton and the iron populations. The same linear coefficients were used to calculate the discrimination parameters for all the showers including the ones initiated by helium, carbon and gamma.

Training datasets with 500 simulated showers for proton and iron showers were used to determine a set of discriminant coefficients. To discriminate the primaries of the simulated events, two linear discriminants $f1$ and $f2$ for each dataset points were calculated using the coefficient previously obtained. The discriminant of larger values indicates which population the new data point should be classified. We used the difference between the two LDA discriminants ($f1 - f2$) to obtain the best separation between the two populations. The same discriminant coefficients, calculated using the proton and iron showers were then applied to helium and carbon initiated showers. The final distribution of $f1 - f2$ parameters for all four different shower types are shown in figure 3. The method was not able to separate helium from proton showers and therefore the distributions for helium primaries are not shown.

We have calculated the separation merit factor between the different distributions. For proton and iron showers, the distributions in figure 3 yields a separation merit factor of 3.7, while the separation between proton and gamma initiated showers pro-

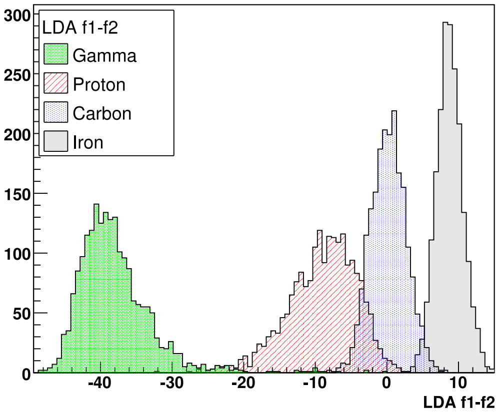

Figure 3: Distribution of LDA parameter $f1 - f2$ for gamma, proton, carbon and iron showers.

vides a merit factor of 5.3, and proton and carbon yields a merit factor of 1.8.

Conclusions

It is clear that the chemical composition of the cosmic ray spectra at the high energy region is essential for understanding the origin and propagation of particles at such high energies. With the increase of data available in this region of the CR spectra, from present experiments such as the Pierre Auger Observatory [8] and HiRes [2] and future experiments such as EUSO [9], operating fluorescence telescopes, it is important to develop new methods and techniques to improve chemical composition determination.

We studied different features of the cosmic ray shower longitudinal profile to determine a better set of parameters that can be used to improve chemical composition of the high energy cosmic ray spectra. To combine the separation capability of all the parameters the statistical method linear discrimination analysis was applied resulting in a new parameter that provided better separation efficiency between the different shower types. To quantify the separation between the different shower distributions, we have defined a merit factor parameter. For showers initiated by proton and iron, with energy of 10^{18} eV, we have achieved a separation merit factor of 3.7, that can be compared to the separation merit factor of 1.3 obtained by using only the shower X_{max} parameter. This result is different and better than the result we have obtained using the showers simulated by the CORSIKA code, in which we had obtained a separation of 2.6 between the proton and the iron. For proton and gamma initiated showers, we have achieved a separation merit factor of 5.7, that can be compared to the separation obtained using only the X_{max} parameter of 1.4. The studies were performed on complete showers simulated using the CONEX code. Results are very similar to the results obtained by simulated showers using the CORSIKA code, with a slightly better separation efficiency between the different CR shower types. Further studies, including truncated shower profiles simulating the limited range of view of real fluorescence detectors show that the separation capability decreases, but still yields a better separation when compared to using only the X_{max} parameter.

This work is partially supported by the Brazilian science funding agencies FAPESP and CNPq, to which we are grateful.

References

[1] A. D., et al., Astroparticle Physics 27 (2007) 61.
[2] A. R.U., et al., Astrophys. Journal 622 (2005) 910.
[3] C. F., et al., astro-ph 0703582.
[4] P. T., et al., astro-ph/0411260.
[5] O. S., Nuc. Phys. B (Proc. Sup) 151 (2006) 143.
[6] V. de Souza, et al., Phys. Rev. D 73 (043001).
[7] J. R.A., W. D. W., ”Applied Multivariate Statistical Analysis”, Prentice Hall.
[8] A. J., et al., Astropart. Phys. 27 (2007) 155.
[9] A. M. et al, Nuclear Instruments and Methods A 567 (2006) 107.