ELSEVIER

Nuclear Physics B (Proc. Suppl.) 52B (1997) 237–239

NUCLEAR PHYSICS B
PROCEEDINGS
SUPPLEMENTS

On the Nonparametric Classification and Regression Methods for Multivariate EAS Data Analysis

A. Chilingarian[a], S. Ter-Antonyan[a], A. Vardanyan[a] (ANI Collaboration) and
H.J. Gils[b], J. Knapp[b], H. Rebel[b], M. Roth[b] (KASCADE Collaboration)

[a]Cosmic Ray Division, Yerevan Physics Institute, Alikhanyan Brothers 2, Yerevan 36, Armenia

[b]Forschungszentrum und Universität Karlsruhe, Institut für Kernphysik, D-76021 Karlsruhe, Germany

Based on a special data analysis methodology developed for non-direct multivariate experiments, we present the expected accuracies of the KASCADE experiment on the elemental composition and primary energy estimation. The calculations were carried out with CORSIKA simulations using the ANI applied package routines. The detector responce also was simulated. The Neural Networks classification, the Bayesian Decision Making and the Nonparametric Regression approaches are used and compared.

1. Introduction

The main objects of Extensive Air Shower (EAS) observations are the possible sources of cosmic rays and the mechanisms of particle acceleration in the interstellar medium. The particular physical quantities to be measured are the energy spectra and elemental composition of cosmic radiation incident on Earth atmosphere. The direct cosmic ray measurements on board of satellites and balloons are well described by the supernovae diffusive shock acceleration mechanism [1]. However, for energies above 10^{14} eV Fermi acceleration becomes less effective [2] and the most well-known peculiarity of cosmic rays - the so-called all particles spectrum cutoff (knee) is detected in the region of $3 \times 10^{15} - 5 \times 10^{15}$ eV [3]. Several hypothesis are proposed for the explanation of the particle acceleration for such energies [4,5], but lack of precise and reliable results on the elemental composition around the "knee" - prevents the strict physical inference of the new acceleration mechanism and (maybe) new type of natural high energy particle accelerators.

Therefore, the problem of cosmic rays origin can only be solved if one succeeds in measuring the element composition of cosmic rays in the energy range of $10^{14} - 10^{16}$ eV. However, direct measurements in the mentioned energy region are impossible due to small fluxes, and mainly indirect methods based on registration of EAS are employed.

Because of the difficulties in posing and solving the inverse problems of cosmic ray physics, the primary composition has not yet been clearly determined, despite 40 years history of the experimental investigations [6-9]. Even the existence of the cutoff is put in question [10].

In contrast with previously used methods, where only the distributions (histograms) summarized over experiment and simulation respectively are compared, and mostly qualitative conclusions were obtained. The techniques proposed here provide the possibility to analyze EAS data on event-by-event basis and obtain quantitative results [11].

2. Mass discrimination of primaries

The only way for the EAS data interpretation is the detailed simulation of the primary particles strong interactions and cascade development in the atmosphere and detectors. The simulation trials including detector responce (training samples) obtained for different primaries and energies are used for physical inference. The training samples provide the nonparametric a priori information about phenomena under investigation. Correspondent to the type of a priori information, nonparametric statistical models are used in the

PII: S0920-5632(96)00894-8

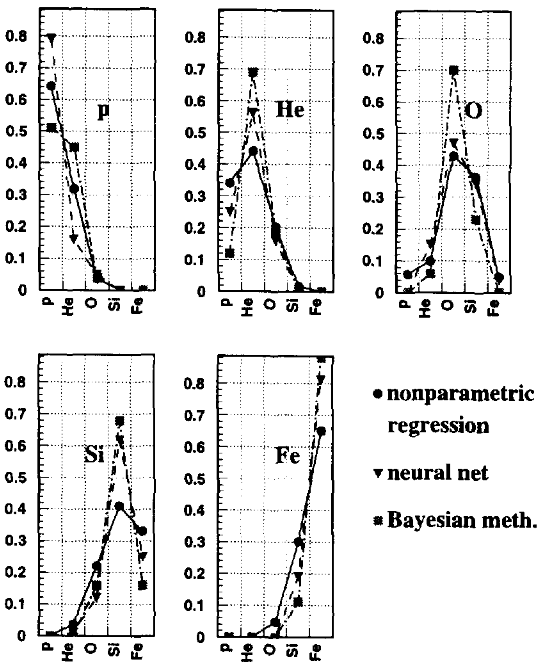

Figure 1. Primary nucleus classification by neural, Bayesian and regression methods (simulated data)

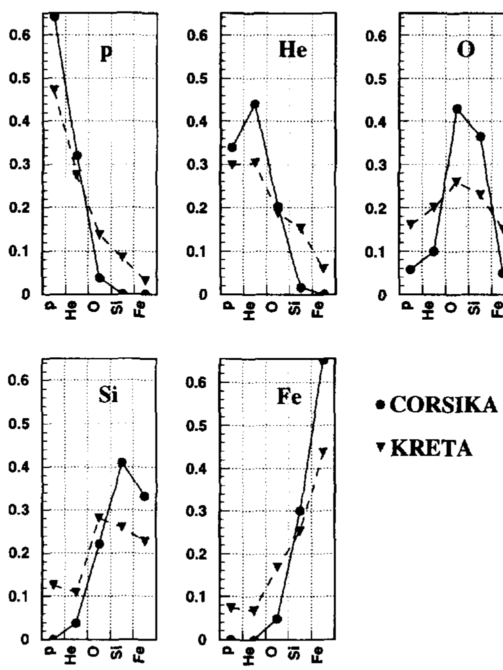

Figure 2. Primary nucleus classification by nonparametric regression method: comparison of pure simulation data and detector response

form of a stochastic mechanism, whereby the data is generated [12].

As local nonparametric models the Bayesian decision method [13] and regression method were applied to the CORSIKA [14] simulations of KASCADE [15] and ANI [12] installations. The local models use the neighborhood (in the multidimensional space of measured EAS parameters) information to decide on the particle and energy type. The Neural Networks are trained on the whole training sample, providing global solutions on the expenses of time consuming, complicated training strategies [16].

In figure 1 these methods are compared in the task of classifying multidimensional EAS data (number of electrons, muons and shower age were used) into 5 categories (proton, alpha, oxygen group, silicon group, iron group). Each of 5 control samples (1000 independent simulation trials) consists of "pure" particles of the mentioned groups and the proportions of their classifications to different nucleus groups are presented.

The results demonstrate rather good consistency of all nonparametric methods, proving relevance and power of nonparametric methodology. The Bayesian decisions are superior the regression methods, as in the later case only one nearest neighbor was used.

3. Detector response

The results presented in the previous section concerned the ideal case of knowing exact numbers of electrons and muons in the EAS. To estimate the bias due to finite sampling and reconstruction errors the detailed detector simulation on the basis of the GEANT package was made taking into account all shower particles, absorbers and active materials, energy deposits, times, trigger conditions and efficiencies, as well as the electronics, digitization of pulseheights, times, etc...

In the second step the reconstruction programs (KRETA) were applied. The EAS core position, arrival direction, electron- muon densities, electron and muon numbers from the array, hadron information, arrival time distributions in central the detector, and many other characteristics are calculated.

In figure 2 the classifications using pure simulation and corresponding detector response are compared. As expected, the misclassification rates increase for realistic experimental situation and resolving protons and alpha particles is very doubtful, but the overall results are satisfactory and the primary identification (especially for proton and iron) still remains reliable.

4. Primary energy estimation

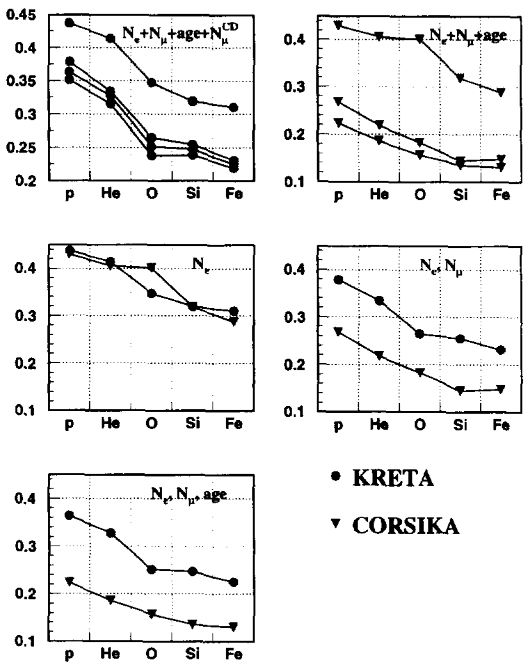

Figure 3. The accuracy of energy estimation using non-parametric regression method and different sets of observables

As one can see from figure 3 adding the additional EAS observables increase significantly the accuracy of energy estimation. The accuracies are different for different nucleus groups and, as expected, better for heavy nuclei. The introduction of detector responce deteriorate the accuracies, but the usage of additional parameters (as the number of muons in the central detector) can improve the situation.

5. Acknowledgments

This work is part of a scientific cooperaton project X131.2 between FZK, Karlsruhe, Germany and Yerevan Physics Institute, Armenia and was partly supported by the research grant N 94-694 of the Armenian government.

REFERENCES

1. Ichimura M et al., Phys. Rev. D48 (1993) 1949
2. Asakimori K. et al., Proc. 23rd ICRC 2 (1993) 21
3. Khristiansen G.B., Kulikov J.V., Solovjava V.I., Pis'ma V., JETF 18 (1958) 353
4. Bierman P.L., Cassinelly J.P., Astron. Astroph. 277 (1993) 691
5. Stanev T. et al., Astron. Astroph. 274 (1993) 902
6. The EAS-TOP Collaboration, Proc. 24th ICRC, 2 (1995) 664
7. Borione A. et al., Proc. 24th ICRC 2 (1995) 676
8. Bakatanov V.N. et al., Proc. 24th ICRC, 2 (1995) 720
9. Swordy S., Proc. 23rd ICRC, Rapporteur Talks, (1993) 243
10. Nikolsky S.I., Proc. 23rd ICRC 4 (1993) 243
11. Chilingarian A., Zazyan H., Nuovo Cimento 14C (1991) 555
12. Chilingarian A., Computer Physics Communications 54 (1989) 381
13. Aharonyan F. et al., NIM A302 (1991) 522
14. Capdevielle J.N. et al., Kernforschungszentrum Karlsruhe, Report KfK 4998 (1992)
15. see H.O. Klages, these proceedings
    P. Doll et al. (KASCADE Collaboration), Kernforschungszentrum Karlsruhe, Report KfK 4686 (1990)
16. Chilingarian A., Pattern Recognition Letters 16 (1995) 333