Preprint YERPHI-1210(87)-89

ԵՐԵՎԱՆԻ ՖԻԶԻԿԱՅԻ ԻՆՍՏԻՏՈՒՏ
ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ
YEREVAN PHYSICS INSTITUTE

A.A. CHILINGARIAN, G.Z.ZAZIAN

A CLASSIFICATION METHOD OF DETERMINATION OF THE MASS COMPOSITION OF PRIMARY COSMIC RAYS IN THE ENERGY RANGE $E_o>10^{15}$ eV

ЦНИИатоминформ
ЕРЕВАН - 1989

Նախնատիպ ԵՖԻ-1210(87)-89

Հ.Զ.ԶԱԶՅԱՆ, Ա.Ա.ՉԻԼԻՆԳԱՐՅԱՆ

ՍԿԶԲՆԱԿԱՆ ՏԻԵԶԵՐԱԿԱՆ ՃԱՌԱԳԱՅԹՄԱՆ (ՍՏՃ) ԶԱՆԳՎԱԾԱՅԻՆ ԲԱՂԱԴՐՈՒԹՅԱՆ ՈՐՈՇՄԱՆ ԴԱՍԱԿԱՐԳԻՉ ԵՂԱՆԱԿԸ $E_0 > 10^{15}$ ԷՎ ԷՆԵՐԳԻԱՆԵՐԻ ՏԻՐՈՒՅԹՈՒՄ

Ներկայացված է մի եղանակ, որը հնարավորություն է տալիս որոշել սկզբնական տիեզերական ճառագայթման զանգվածային բաղադրությունը մոդելային և փորձարարական տվյալների համատեղ վերլուծման ճանապարհով: Եղանակի կարևորագույն մասն է՝ բազմաչափ բաշխումների քանակական համեմատությունը և ոչ-պարամետրիկ վիճակագրության եղանակների կիրառումը՝ հավանականության խտության գնահատման համար՝ փոփոխականների բազմաչափ տարածությունում: Առաջարկված եղանակի ստուգման համար խաղարկվել են Մոնտե-Կառլո եղանակով $E_0 > 500$ՏէՎ էներգիայով դեպքերը: Կուտակված հեղեղների գրանցումը կատարվել է 700 գ/սմ^2 խորությունում: Մոդելային դեպքերի նախնական մշակումը կատարվել է ըստ ալգորիթմների, որոնք կիրառվել են փորձարարական տվյալների մշակման ժամանակ: Հաշվի են առնվել գրանցող սարք մտնող ԼՄՀ չափվող բնութագրերի աղավաղումները: Եղանակը թույլ է տալիս նույնացնել սկզբնական պրոտոններից և երկաթի միջուկներից առաջացած փորձարարական դեպքերը 70-80% արդյունավետությամբ, և որոշել ՍՏՃ զանգվածային բաղադրությունը $10^{15}-10^{17}$էՎ տիրույթում:

Երևանի ֆիզիկայի ինստիտուտ

Երևան 1989

Preprint YERPHI-1210(87)-89

A.A. CHILINGARIAN, G.Z. ZAZIAN

A CLASSIFICATION METHOD OF DETERMINATION OF THE MASS COMPOSITION OF PRIMARY COSMIC RAYS IN THE ENERGY RANGE $E_0 > 10^{15}$ eV

A method allowing to determine the mass composition of primary cosmic radiation by means of simultaneous analysis of model and experimental data is presented in this paper. The most important part of this work is the quantitative comparison of multivariate distributions and the use of methods of nonparametric statistics for probability density estimation in a multivariate space of features. To check the method offered, events with $E_0 > 500$TeV were generated by the Monte-Carlo method. The showers generated were preliminarily processed by algorithms used in experimental data handling. The apparatus-induced distortions of the measured EAS characteristics have been taken into account. The method allows to select experimental event initiated by primary protons and iron nuclei with an efficiency of ~70-80% and to determine the mass composition of primary cosmic rays (PCR) at $10^{15}-10^{17}$eV.

Yerevan Physics Institute

Yerevan 1989

Препринт ЕФИ-1210(87)-89

Г. Э. ЗАЗЯН , А. А. ЧИЛИНГАРЯН

КЛАССИФИКАЦИОННЫЙ МЕТОД ОПРЕДЕЛЕНИЯ МАССОВОГО СОСТАВА ПКИ В ОБЛАСТИ ЭНЕРГИИ $E_0 > 10^{15}$ эВ

В настоящей работе представлен метод ,позволяющий определить массовый состав первичного космического излучения путем совместного анализа модельных и экспериментальных данных. Важнейшей частью метода является количественное сравнение многомерных распределений и использование методов непараметрической статистики для оценки плотности вероятности в многомерном пространстве признаков. Для проверки предлагаемого метода были сгенерированы события с энергией $E_0 > 500$ ТэВ методом Монте-Карло. Регистрация сгенерированных ливней проводилась на глубине 700г/см^2. Предварительная обработка модельных событий проводилась по алгоритмам ,применяемым при обработке экспериментальных данных. Учитывались искажения измеряемых характеристик ШАЛ,вносимые регистрирующей аппаратурой. Метод позволяет отобрать экспериментальные события, инициированные первичными протонами и ядрами железа с эффективностью порядка 70-80%, и определить массовый состав ПКИ при энергиях 10^{15}–10^{17}эВ

Ереванский физический институт

Ереван 1989

1. Introduction

The ambiguity of interpretation of the results of experiments with cosmic rays is connected with both significant gaps in our knowledge of the characteristics of hadron-nuclear interactions at superaccelerator energies and indefiniteness of the PCR composition.The extra difficulties are due to indirect experiments and hence, due to the use of Monte-Carlosimulations of development and detection of different components of nuclear electromagnetic cascades.

To research into hadron-nuclear interactions in CR, one should know the type of cascade-initiating incident particle. Besides, the investigation of the mass composition of PCR is of a particular interest in connection with the problem of the CR origin.

At present the data available on the mass composition of PCR in the energy range 10^{15}–10^{17}eV are obtained by detecting and investigating the different components of EAS and γ-families detected by X-ray emulsion chambers (X-REC). And if the first data set states a "normal composition" - extrapolation of PCR composition (40% protons and 20% iron nuclei) measured by direct methods in the energy range 10^{11}–10^{14}eV [1,5], then the data on γ-family fluxes testify to a decrease of the protons fraction in PCR at $E_0 > 10^{16}$eV down to 15-20% and hence - to dominance of iron nuclei [2].

This contradiction, inexplicable yet, may be due to different experimental data handling methods. Besides, the γ-family characteristics are more sensitive to the method of strong interactions than the EAS ones, hence, variation of model parameters can change the estimate of the proton and iron nuclei fractions. The problem of relative dominance of iron nuclei is very important, because the interpretation of the experimental data obtained in UHE CR is based on the mass composition of PCR [3].

In this paper an approach is presented, which allows to determine the mass composition of PCR by means of simultaneous analysis of model and experimental data. The most important part of the method is the quantitative comparison of multivariate distributions and use of a nonparametric technique to estimate the probability density in a multidimensional feature space. As compared to the earlier used methods of inverse problem solution, with the help of which the mass composition of PCR was first determined in the energy range $E_o > 10^{15}$eV with sufficient quantitative certainty [6], in the proposed method the object of analysis is each particular event (a point in the multivariate space of EAS parameters) rather than alternative distributions of model and experimental data. That is why, along with the averaged characteristics, the belonging of each event to a certain class is determined.

This approach was first used to estimate the upper limit of the iron nuclei fraction according to the γ-family characteristics [5]. As opposed to Ref.[5], where events were classified into two classes ($A > 50$ and $A < 50$), now it is possible to classify events into an unlimited number of classes.

Events with $E_o > 500$TeV have been simulated to check the method proposed. The showers were registered at a depth of 700g/cm^2. The model data were preliminarily handled according

to algorithms used in the data handling at the Tien-Shan station [6]. The finite resolution of the installation measuring EAS characteristics has been taken into account.

2. Simulation of a Nuclear-Electromagnetic Cascade in the Atmosphere

The incident particle energy was drawn according to the energy spectrum given in Ref.[7]. A normal CR composition was simulated (40% protons and 20% iron nuclei). Five groups of nuclei were considered: protons (p), α-particles (α), nuclei with A=7-16(CNO), A=24-27(H) and A=50-56(VH). Protons and α-particles further were unified into one group.

Strong interaction were simulated by algorithms which reproduce the quark-gluon string model (QGSM) spectra for hA interactions [8]. The algorithm used allows to simulate nucleon, pion, kaon and Λ-hyperon interactions with N^{14} at $0.03 < E < 10^6$TeV. Production of $N\bar{N}$ pairs, π, k, Λ, η was taken into account. In the frame of the Regge theory there were simulated processes of single and double diffraction as well as inelastic recharging of $\pi^{\pm} \to \pi^o$, $p \to n$.

The incident nucleus fragmentation was taken into account when simulating AN^{14} interactions. The hN^{14} interaction cross sections were approximated as:

$$\sigma^{pN^{14}}_{prod} = 295+23.94 lg(E)+3.55 lg^2(E) ,$$
$$\sigma^{\pi N^{14}}_{prod} = 226+24.51 lg(E)+2.31 lg^2(E) , \qquad (2.1)$$
$$\sigma^{kN^{14}}_{prod} = 178+27.02 lg(E)+1.81 lg^2(E) ,$$

where E is incident particle energy in the laboratory system of coordinates (E is in TeV). Energy dependence of the mean multiplicity of charged particles in hN^{14} interactions was approximated as:

$$\langle N\rangle_{ch}=0.817\ln^2E+3.127\ln E+10.27 \quad (E \text{ is in TeV}) \qquad (2.2)$$

The number of secondaries was drawn according to KNO distribution. It was taken that the mean transverse momentum of secondaries increased with the energy according to

$$\langle P\rangle_t=0.36(1+0.023\ln(E/0.1)) \quad (E \text{ is in TeV}) \qquad (2.3)$$

For different secondaries $\langle P_t\rangle$ and the shape of distribution over P_t differed and corresponded to the existing accelerator data.

Electromagnetic interaction were simulated as in Ref.[9]. Pair production, bremsstrahlung and multiple Coulomb scattering were taken into account. At the same time, it was assumed that transversal development of electron-photon cascades is due to only multiple Coulomb scattering. To calculate the average EAS characteristics, we have used the approximated formulae obtained in Ref.[10].

3. Comparison of EAS Single Characteristics and Choice of Optimal Features

To choose features most sensitive to the PCR composition, there were compared the single characteristics of EAS initiated by primary protons and iron nuclei. The following EAS characteristics have been considered: the total number of

electrons N_e, the total number of muons with E_μ>5GeV, the EAS age parameter S, the total number, energy, mean energy, average distance to EAS cores and dispersion of spatial and energy distributions of muons with E_μ>200GeV and hadrons with E_h>200GeV in EAS and the linear regression coefficients of spatial and energy distributions of muons and hadrons (E=C1+C2R).

The results of comparison of single EAS characteristics are summarized in Figs.1-4. A quantitative comparison of various features is presented in Table 1, where the P values of statistical tests of comparison of samples from univariate distributions as well as the Bhattacharya distance between the samples are given. The Student, Kolmogorov, Mann-Whitney tests have been used. It follows from these data that most appropriate feature to determine EAS composition are the high-energy muon characteristics. The hadron component characteristics are less sensitive to the primary particle type (the higher the P-values of the test, the stronger the difference between the corresponding distributions). In this paper we have used EAS characteristics only (N_e,N_μ (E>5GeV), S). Though the sensitivity of N_e and S to the primary particle type is low, however, due to different degree of correlation between N_e,N_μ and S for events initiated by primary protons and iron nuclei, as is seen from Table 2 and 3, the use of all the three EAS characteristics essentially improves the event classification reliability. The choice of these characteristics is also owing to their relatively low sensitivity to strong interaction characteristics, which allows to hope for obtaining a model- independent inference on the PCR mass composition.

4. Classification of the Distribution Mixture

Let us consider the stochastic mechanism $(A,\mathfrak{D})$ which generates the observation $\vec{v}$ in a multivariate feature space ($\vec{v}$ is a d-dimensional vector of values measured in experiment, d is dimensionality of the feature space). The basic event space A is a combination of events from different primary nuclei. We know no law of nature like $(A,\mathfrak{D})$, that is why, to determine a probability metric on A, the total Monte-Carlo simulation of the phenomenon under investigation is performed, including experimental data registration and handling.

The set of d-dimensional $\vec{U}$ vectors obtained in simulations is the simular analog of the experimentally measured values of $\vec{V}$. But, as opposed to experimental data, it is known to which of the alternative classes each of the events belongs. These "labeled" events include a priori information about dynamics of the process under investigation, which is given in a nonparametric form, as finite size samples. The sequence $\{U_i t_j\}$, where $i=1,M_{TS}$, $j=1,L$, t_j is class index, we usually call a training series or sample (TS) which is also denoted by $(A,\bar{\mathfrak{D}})$.

Since both, physical processes of particle production and those of registration are stochastic and the information about the phenomena under investigation is smeared out, the data analysis is uncertain in the sense that one need not wait for event separation into compact nonoverlapping groups corresponding to different primary nuclei. The only thing we can require when classifying experimental data by v ious primary nuclei is to minimize the losses due to incorrect classification to some degree and to ensure use of a priori information completely. Such a procedure is the Bayes decision

rule with nonparametric estimation of the multivariate probability density function, which, when using a simple loss function (the loss is zero in case of correct classification and is the same at any other error), takes the form:

$$\hat{A} = \eta(v,A,\tilde{\mathfrak{D}}) = \mathrm{argmax}\{\hat{P}(A_i/v)\} \quad , \; i=1,L \qquad (4.1)$$

where $\hat{P}(A_i/v) \sim P_i \hat{P}(v/A_i)$ are a posterior densities, $\hat{P}(v/A_i)$ are conditional densities which are estimated by TS $(A,\tilde{\mathfrak{D}})$ using one of many nonparametric methods available [11], L is the number of groups of nuclei.

Initial (a priori) values of P_i are taken equal. The monograph [10] is devoted to the interplay of a priori and experimental information in fraction estimation problems. Here we shall not go into discussion of competence of the choice of a uniform a priori distribution, but only mention that at such a choice the a posterior probability and hence, the results of classification will be totally defined by experimental information, which seems reasonable to us in the given physical task.

To estimate conditional densities, we used Parzen's method with automatic kernel width adaptation. In this method some probability density values are calculated which correspond to different values of method parameters. Then the sequence obtained is ordered and the median of this sequence is chosen as final estimate. Depending on the value of the probability density in the vicinity of $\vec{V}$, due to stabilizing properties of the median, each time we'll choose an estimate with a width most fitting for that region [13]. The probability density is estimated by:

$$P(V/A_i)=1/(2\pi^{n/2}h^n)\sum_{j=1}^{M_i} e^{-r_j^2/h^n} w_j \ , i=1,L \qquad (4.2)$$

where n is the feature space dimensionality, M_i is the number of vectors of the i-th TS class, r_j is distance to the j-th neighbour in the Mahalanobis metric:

$$r_j=(V-U_j)^T R^{-1}(V-U_j) \ , \qquad (4.3)$$

where R is a sampling covariance matrix, w_j are the events weights, h is the kernel width.

The classification methods, like all the statistical ones, include the procedure quality test as a necessary element. This stage beside all the others is also necessary for determination of the primary composition. The most natural procedure quality estimate is error probability which depends on both the degree of overlapping of alternative multivariate distributions and the decision rule being used (Bayes decision rules provide minimum error probability (EP) as compared to any other one):

$$EP \equiv R_M = E\{\theta[\eta(V,A,\mathfrak{D})]\} \ , \qquad (4.4)$$

where

$$\theta[\eta(V,A,\mathfrak{D})] = \begin{cases} 0, & \text{at correct classification} \\ 1, & \text{otherwise} \end{cases}$$

and E stands for mathematical expectation. The expectation is taken over all possible samples of volume M and over the whole d-dimensional space of measured values.

Since we do not exactly know to what class the experimental vectors belong, the estimate of R_M we obtain via TS:

$$\hat{R}_M=1/M_{TS}\sum_{i=1}^{M_{TS}}\theta\{t_j,\eta(U_i,A,\tilde{\mathfrak{D}})\} \qquad (4.5)$$

i.e we classify the $\{U_i\}$ TS and check correctness of classification over the index of the class t_j, j=1,L. However, as numerous investigations have shown (e.g.,[15]), this estimate is systematically biased and hence, a cross-validation estimation is preferable

$$R_M^e = 1/M_{TS}\sum_{i=1}^{M_{TS}}\theta\{t_j,\eta(U_i,A,\tilde{\mathfrak{D}}_{(i)})\} \ , \qquad (4,6)$$

where $A,\tilde{\mathfrak{D}}_{(i)}$ is a TS with a removed i-th element, which is classified. This estimate is unbiased and has an essentially smaller r.m.s. deviation. The advantage of R_M^e is especially notable when the feature space has a higher dimensionality [15].

Note, that we have the possibility to estimate the errors probability of various types by imposing to classification various TS classes - $\{U_i,t_j\}$, j=1,L, L is the number of classes.

By R_{ij} we denote the probability for classification of the i-th class events as belonging to the j-th class (misclassification).

Now let us estimate the a posterior fraction of various kernel types in the incident flux.

It is known [16] that the best estimate of a posterior fraction (in case of uniform a priori information and absence of classification errors) is the empirical fraction

$$P_i^e = M_i/M_{tot} \ , \qquad (4.\ )$$

where M_i is the number of events classified as initiated by the kernel group A_i, M_{tot} is the total number of events registered during experiment. It can be shown (see [17], where a formula for the case with L=2 is derived) that with account of classification errors the fraction of various kernels can be obtained as the solution of the following set of linear equations:

$$(1-\sum_{j\neq i}^{L} R_{ij})\hat{P}_i + \sum_{k\neq i}^{L} \hat{P}_k R_{ki} = P_i^e \ , \ i=1, L \ . \quad (4.8)$$

In the first sum summation goes over j, in the second -over k. All estimates of R_{ij} and P_i^e are obtained over one and the same TS using the same decision rules.

The accuracy of estimates is defined by the TS size and number of experimental data as well as by the value of the classification errors, which present the "quality" of discrimination in the chosen feature subset. Note, that the set (4.8) is a poorly defined system and at large values of classification errors the solutions of the set are unpredictable and hence, the choice of a feature combination providing a high percentage (≥60%) of correct classification is a necessary preliminary stage.

5. The Bootstrap Procedure of Fraction Estimation

As we have shown in the previous section, to estimate the fraction of various kernels in an incident flux of cosmic radiation, beside classification of an experimental sample by a TS, it is also necessary to calculate any misclassification coefficients, R_{ij}. The error in determination of the fraction

of various kernels is a function of the errors both from classification and in determination of R_{ij}.

The possibility to decrease the bias and variance of misclassification rates estimates was discussed in Ref.[17], where it was mentioned that it is possible to improve the accuracy of R_{ij} estimates, if the TS size is large enough to separate the TS into independent subsamples.

Unfortunately, time consumption per model event generation increases abruptly with energy and we have not to expect much model information in the energy range $E>10^{15}$eV.

Thus, the problem of an efficient use of the information contained in simular results is as never actual for cosmic-ray and accelerator physics, since the classical sampling models do not allow to extract the whole information carried by a sample.

The methods of sample control during handling are widely used in the last few years. One of these is the leave-one-out-for-a-time test considered in the previous section, which allows to decrease the sample bias. A more efficient procedure actively developing in both applied and theoretical respects in the last decade is the bootstrap which lies in replication of the initial sample very many times by means of random sampling with replacement. The thus obtained conditionally independent bootstrap-replicas in many respects stand for independent samples from the general population (under the condition of sufficiently large size of the initial sample). In fact, the bootstrap substitutes the unknown general population by a single sample, i.e. the ideology described in the section 4 of this paper is followed. The theoretical base of the bootstrap method is the analog of the central limit theorem (CLT) proved in Ref.[18]:

$$P\left\{\sqrt{B}\,(\mu_B-\mu_M) < t\,S_M \middle| x_1,\ldots x_M\right\} \to \Phi(t), \qquad (5.1)$$

when $M,B\to\infty$, $x_1,\ldots,x_M$ are independent, identically distributed (IID) random quantities, $\Phi(t)$ is a normal (Gaussian) distribution, μ_M and S_M are sample estimates of the first and the second moments, $\mu_B=\sum_{j=1}^{B}\mu_j^*/B$, $\mu_j^*=\sum_{i=1}^{M}x_i^{(j)}/M$ is the j-th bootstrap replica's mean. And what is more, analogies between sampling and the bootstrap are valid also for many other statistics. Referring our correspondents to the Ref.[19], we shortly summarize the main idea of the new procedure: a new procedure – the bootstrap-moments (denoted by E_*,σ_*) are introduced, which in many cases substitute the statistical momenta calculated according to a distribution function (in most cases of interest it is unknown).

Owing to the fact that the bootstrap is very important for high-energy physics, and to investigate its possibilities for finite samples and a limited number of bootstrap replicas we have carried out an investigation with the purpose to calculate the bootstrap expectation $(\mu_j^*-\mu_M)$ – a CLT test, and calculation of bootstrap expectations of the standard deviation of the mean IID of random variables – $\delta_*^2=\sigma_*^2/M$. To do this we used samples from the standard, normal distribution N(0,1); the sample size varied between 25 and 1000, the number of bootstrap replicas in a series was from 10 to 2000. The mean was calculated for each bootstrap replica and for each bootstrap series – the bootstrap-estimate of the mean standard deviation – δ_*.

A round of recalculations including 100 series of the same size has been carried out using different initial samples; the obtained data were averaged and the mean-square deviations were

calculated. The results of investigations, which are present in Table 4, state the validity of CLT and consistency of using the bootstrap expectation. Although the mathematical theorems were proved for the asymptotic cases $M,B\to\infty$, even with small sample sizes and small number of bootstrap replicas ($M,B\sim50$), the obtained estimations fit to the expected theoretical ones.

There are two ways of distribution mixture coefficient estimation: i) to obtain the bootstrap estimate of the misclassification coefficients R^e_{ij}, then classify and estimate the fraction or ii) carry out fraction estimation over each bootstrap replica, then obtain the fraction and the standard deviation bootstrap expectation. The second way is preferable, because obtaining of the standard deviation in the first case is time-consuming. It is enough to say that the errors propagation formulae obtained by the REDUCE program occupy several standard sheets in case of classification into four classes.

In the end of this section let us formalize the bootstrap method of the distribution mixture coefficient estimation. Let us define the solution of the set (4.8) as:

$$\vec{P} \equiv \vec{P}\{\hat{P}_1,\ldots\hat{P}_l\} = f\{V,\tilde{\mathfrak{D}},\eta(V,A,\tilde{\mathfrak{D}})\}\ . \qquad (5.2)$$

This solution is a complex function of experimental data and the TS as well as the decision rule η being used. By several TS bootstrap replicas we calculate the bootstrap expectation and the bootstrap standard deviation of the mixture coefficients $\hat{P}_i$, which are used as estimates of the fraction of different kernel groups in the primary flux.

6. Results of Calculations

To test the method, the generated events were grouped in two. The first were used to create a TS and the second - as pseudo-experimental events. The EAS characteristics (N_e,N_μ,S) were used in the events classification, where events in different fixed intervals over N_e were selected. The TS consisted of four classes in accordance with the primary kernel type (p-protons, α-particles, CNO-kernels with A=7-16, H-with A=24-27 and VH-with A=50-56).

Table 5 presents the Bayes error matrix obtained as a result of a leave-one-out test over TS. The diagonal elements of this matrix show the probability for a correct events classification and the nondiagonal elements - the probability for misclassifications. It is seen from Table 5 that the correct classifications make about 70-80% (classification of "boundary" groups (protons and iron group nuclei) is essentially better than that of the intermediate groups). Note, that the accuracy of classification can be improved by selecting events at narrow zenith angles ϑ (ϑ varies between 0 and 45°).

Table 6 shows the recovered kernel fractions obtained by classification of model events for one interval over N_e. The errors presented are obtained by the bootstrap procedure. Fractions of kernel groups given in EAS simulations (true fractions) are presented ibid. As is seen from this table, the proposed method allows to determine the fraction of protons and iron nuclei in the incident flux with quite a good accuracy. To improve the accuracy of determination of the fraction of intermediate nuclei it is necessary to increase the size of TS.

In this work events obtained by the same model are used as control (pseudo-experimental) and training samples. During the experimental data handling the model adequacy test is a necessary stage. The difficulty lies in the fact, that the changes both in the strong interaction model and in the mass composition can lead to the same change of the observed values. To overcome this ambiguity one can use the "self-consistency" method developed in Ref.[5].

7. Conclusion

The proposed method allows to select experimental events initiated by incident protons and nuclei with an efficiency of ~70-80% and determine the mass composition of PCR at energies from 10^{15} to 10^{17}eV. The main advantages of the method proposed are:

i) its being a multivariate one, i.e. inclusion of additional EAS parameters in the analysis meet no difficulties;

ii) individual analysis (event by event) - each experimental event is an object of analysis - their belonging to a certain class and the error of statistical solution are determined;

iii) a priori chosen probability family is not imposed on data - the results of simulation are used directly during the process of statistical solutions.

We hope that the use of the proposed method when handling the experimental data obtained at complex arrays will allow to get an unambiguous information about the character of strong interactions at superaccelerator energies.

We are grateful to A.M. Dunaevsky and N. Stamenov for useful discussions and to E.A. Mamidjanian for stimulating interest in the work.

One of the authors (Z.G.Z) also thanks A.M. Dunaevsky for provision of EAS simulation algorithms.

Table 1

P-values of statistical tests of comparison of univariate distributions of different characteristics of EAS initiated by primary protons and iron group nuclei at $1\cdot10^5 < N_e < 2\cdot10^5$

	Stud	Kolm	M-U	R Bh
N_e	0.835	1.021	1.17	0.002
$N_\mu(E_\mu > 5$ GeV)	23.647	7.329	16.12	0.433
S	7.046	3.178	6.60	0.074
$N_\mu(E_\mu > 200$ GeV)	25.435	7.553	16.34	0.451
$\sum E_\mu(E_\mu > 200$ GeV)	15.217	5.582	12.45	0.189
$\langle E_\mu\rangle(E_\mu > 200$ GeV)	6.792	5.528	11.13	0.458
$\langle R_\mu\rangle(E_\mu > 200$ GeV)	18.341	6.533	14.36	0.295
$N_h(E_h > 200$ GeV)	4.717	2.400	4.88	0.248
$\sum E_h(E_h > 200$ GeV)	4.015	2.609	5.30	0.495
$\langle E_h\rangle(E_h > 200$ GeV)	3.503	1.658	3.44	0.128
$\langle R_h\rangle(E_h > 200$ GeV)	5.903	3.395	6.73	0.033

Table 2

The coefficients of correlation between the characteristics of the electron-photon and the muon components of EAS with $1\cdot10^5 < N_e < 2\cdot10^5$ from primary protons (the quantities marked with * correspond to muons with $E_\mu > 200$ GeV)

	N_e	N_μ	S	N_μ^*	$\sum E_\mu^*$	$\langle E_\mu^*\rangle$	$\langle R_\mu^*\rangle$
N_e	****	0.393	-0.132	0.145	0.108	0.01	-0.180
N_μ	0.393	*****	0.443	0.838	0.717	0.044	0.516
S	-0.132	0.443	*****	0.468	0.441	0.039	0.469
N_μ^*	0.145	0.838	0.468	*****	0.896	0.104	0.726
$\sum E_\mu^*$	0.108	0.717	0.411	0.896	*****	0.476	0.658
$\langle E_\mu^*\rangle$	0.010	0.044	0.039	0.104	0.476	*****	0.093
$\langle R_\mu^*\rangle$	-0.18	0.516	0.469	0.726	0.658	0.093	*****

Table 3

The coefficients of correlation between the characteristics of the electron-photon and the muon components of EAS with $1\cdot10^5 < N_e < 2\cdot10^5$ from primary iron nuclei (the quantities marked with * correspond to muons with $E_\mu > 200$ GeV)

	N_e	N_μ	S	N_μ^*	$\sum E_\mu^*$	$\langle E_\mu^* \rangle$	$\langle R_\mu^* \rangle$
N_e	****	0.555	-0.205	0.353	0.366	0.252	-0.026
N_μ	0.555	*****	0.195	0.830	0.820	0.375	0.345
S	-0.205	0.195	*****	0.238	0.224	0.075	0.224
N_μ^*	0.353	0.830	0.238	*****	0.978	0.391	0.633
$\sum E_\mu^*$	0.366	0.820	0.224	0.978	*****	0.556	0.611
$\langle E_\mu^* \rangle$	0.252	0.375	0.075	0.391	0.566	*****	0.213
$\langle R_\mu^* \rangle$	-0.026	0.345	0.224	0.633	0.611	0.213	*****

Table 4

Bootstrap expectations and bootstrap standard deviations of sampling statistics

	B	10	50	100	200
M=25	$E_*\{\mu_b-\mu_m\}$	-0.0152	0.0031	-0.0048	-0.0003
δ_{25}=0.2	$\sigma_*\{\mu_b-\mu_m\}$	0.0639	0.0251	0.0174	0.0160
	$E\{\delta_*\}$	0.1891	0.1974	0.1929	0.1977
	$\sigma\{\delta_*\}$	0.0560	0.0300	0.0031	0.0028
M=50	$E_*\{\mu_b-\mu_m\}$	-0.0024	-0.0023	0.0003	-0.0001
δ_{50}=0.1414	$\sigma_*\{\mu_b-\mu_m\}$	0.0402	0.0227	0.0149	0.0097
	$E\{\delta_*\}$	0.1481	0.1398	0.1396	0.1395
	$\sigma\{\delta_*\}$	0.0286	0.0182	0.0167	0.0154
M=100	$E_*\{\mu_b-\mu_m\}$	-0.0171	-0.0010	0.0004	-0.0008
δ_{100}=0.1	$\sigma_*\{\mu_b-\mu_m\}$	0.0323	0.0152	0.0101	0.0066
	$E\{\delta_*\}$	0.0897	0.0959	0.1000	0.0988
	$\sigma\{\delta_*\}$	0.0212	0.0107	0.0097	0.0086

(continued)

Table 4 (continuation)

	B	10	50	100	200
M=200 δ_{200}=0.0707	$_*\{\mu_b-\mu_m\}$	0.0038	-0.0017	0.0001	0.0000
	$_*\{\mu_b-\mu_m\}$	0.0231	0.0107	0.0082	0.0048
	$\{\delta_*\}$	0.0593	0.0692	0.0694	0.0700
	$\{\delta_*\}$	0.0154	0.0078	0.0063	0.0049
M=500 δ_{500}=0.0447	$E_*\{\mu_b-\mu_m\}$	-0.0018	0.0007	0.0004	0.0003
	$\sigma_*\{\mu_b-\mu_m\}$	0.0115	0.0072	0.0040	0.0032
	$E\ \{\delta_*\}$	0.043	0.0452	0.0442	0.0446
	$\sigma\ \{\delta_*\}$	0.0095	0.0043	0.0033	0.0024
M=1000 δ_{1000}=0.032	$E_*\{\mu_b-\mu_m\}$	0.0038	0.0001	0.0002	0.0003
	$\sigma_*\{\mu_b-\mu_m\}$	0.0079	0.0050	0.0030	0.0022
	$E\ \{\delta_*\}$	0.0322	0.0317	0.0316	0.0315
	$\sigma\ \{\delta_*\}$	0.0073	0.0033	0.0022	0.0017

Table 5

The Bayes error matrix obtained by the leave-one-out method, by TS within the range $1\cdot10^5<N_e<2\cdot10^5$

	P	CNO	H	VH
P	0.798	0.102	0.067	0.033
CNO	0.127	0.688	0.105	0.080
H	0.072	0.113	0.691	0.124
VH	0.034	0.090	0.150	0.726

Table 6

Recovered fractions of four groups of nuclei within $1\cdot10^5<N_e<2\cdot10^5$
(W_{in} is a "true" fraction, W_{out} - a recovered one)

	N_{TS}	W_{in}	$E_*\{W_{out}\}$	$\sigma_*\{W_{out}\}$
P	200	0.370	0.345	0.038
CNO	188	0.272	0.229	0.067
H	194	0.168	0.232	0.057
VH	163	0.189	0.194	0.019

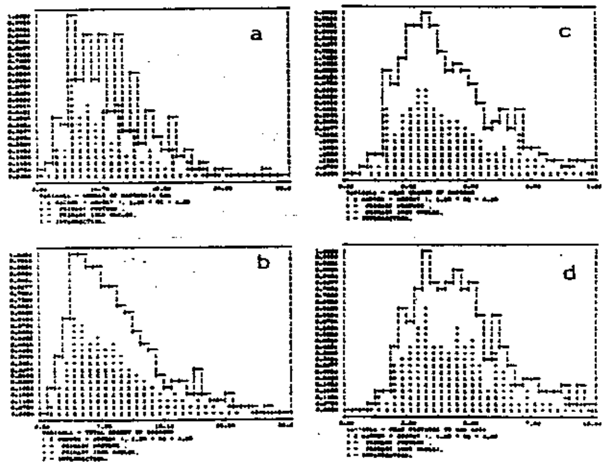

Fig.1

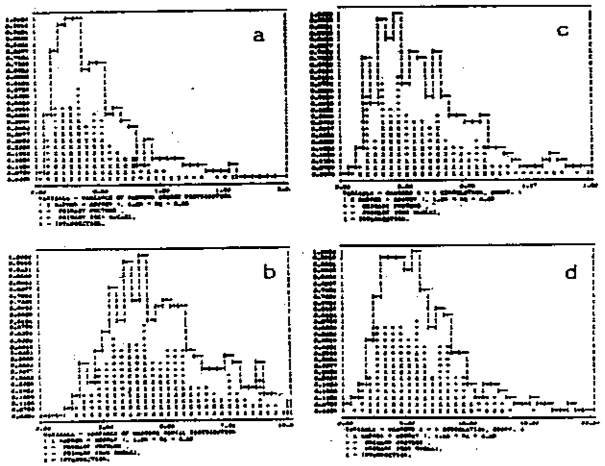

Fig.2

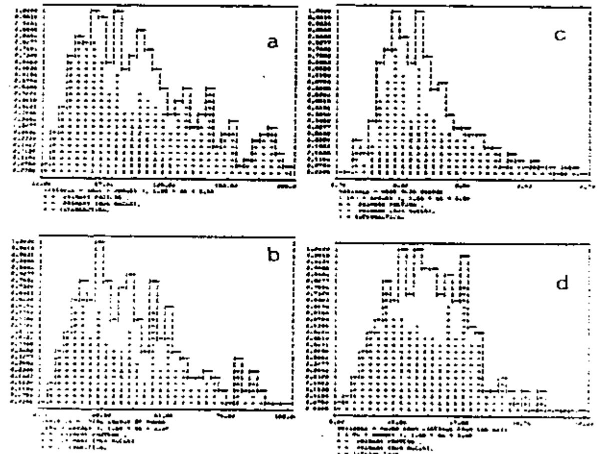

Fig.3

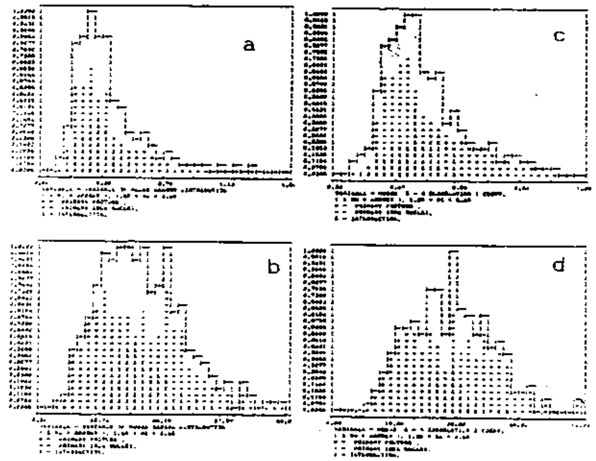

Fig.4

Figure Captions

Fig.1 Distribution of the characteristics of EAS hadrons with $E_h > 200GeV$.

* - primary protons, + - primary iron nuclei,

X - region of overlapping,

a) Total hadron number distribution,

b) Total hadron energy distribution,

c) The hadron mean energy distribution in EAS,

d) Distribution of the average distance of hadrons to EAS cores.

Fig.2 Distribution of the spatial and energy distributions dispersion and the coefficients of E-R correlations of hadrons with $E_h > 200$ GeV.

* - primary protons, + - primary iron nuclei,

X - region of overlapping.

Fig.3 Distribution of the characteristics of EAS muons with $E_\mu > 200GeV$.

* - primary protons, + - primary iron nuclei,

X - region of overlapping,

a) Total muon number distribution,

b) Total muon energy distribution,

c) The muon mean energy distribution in EAS,

d) Distribution of the average distance of muons to EAS cores.

Fig.4 Distribution of the spatial and energy distributions dispersion and the coefficients of E-R correlations of muons with $E_h > 200$ GeV.

* - primary protons, + - primary iron nuclei,

X - region of overlapping.

References

1. J.N.Stamenov, High Energy Cosmic Ray Mass Composition on the Basis of EAS Studies at Mountain Altitudes, 5th Int.Symp. on Very High Energy Cosmic Rays Interactions. Lodz , Poland, 1988.
2. Mt. Fuji Emulsion Chamber Collaboration, Apparent Decrease of Proton Fraction Around the "Knee" Observed in mt.Fuji Emulsion Chamber Experiment, Proc. 20th ICRC, v 1, p 382, Moscow, 1987.
3. J.Rich, D.L.Owen, M.Spiro, Experimental Particle Physics Without Accelerators, Phys. Reports, v.151, N 5&6, 1987; Momentum, Scaling Violation in hN^{14} Interactions and Composition of Primary Cosmic Rays at 10^{16}eV, 5th Int. Symp. on Very High Energy Cosmic Rays Interaction, p.143, Lodz, Poland, 1988.
4. И.Н.Стаменов, Исследование состава первичного космического излучения с помощью ШАЛ. Диссертация ФИАН, Москва, 1981.
5. С.Х.Галфаян, А.М.Дунаевскийи и др., Верхняя граница доли ядер железа в ПКИ при $E > 10^{16}$эВ. Препринт ФИАН N 75, Москва, 1989.
6. Н.М.Никольская, И.Н.Стаменов, Исследование алгоритма статистичестой обработки экспериментальных данных и характеристик управляющей системы, отбирающей ливни на Тянь-Шяньской комплексной установке ШАЛ. Препринт ФИАН, N 125, Москва 1975.
7. С.И.Никольский, Энергетический спектр и ядерный состав первичных космичкских лучей. Проблемы физики космических лучей, стр 164-185, Москва, "Наука", 1987.

8. Ю.М.Шабельский, Сечения и спектры вторичных частиц в столкновениях адронов с ядрами при высоких и сверхвысоких энергиях. Препринт ЛИЯФ N 1224, Ленинград, 1986.

9. А.М.Дунаевский, А.В.Урысон, Скейлинг, рост сечения и моделирование ядерно-электромагнитных каскадов в атмосфере. Препринт ФИАН N 150, Москва, 1975.

10. A.V.Plyasheshnikov, A.K.Konopelko, K.V.Vorobyev, The Three Dimensional Development of High Energy Electromagnetic Cascades in Atmosphere, Preprint FIAN N 92, 1988; Density Estimation, The John Hopkins University Press, Baltimore and London, 1978.

11. L.Devroy, L.Gyorfi, Nonparametric Density Estimation. The L1 View, John Wiley & Sons, New-York, 1985.

12. E.E.Leamer, Ad hoc Inference with Nonexperimental Data. John Wiley & Sons New-York-Chichester-Brisbane-Toronto , 1978.

13. A.A.Chilingarian, Statistical Decisions under Nonparametric a priori Information, Comp. Phys. Comm., vol.54, 381-390, 1988.

14. G.T.Toussaint, Bibliography of Misclassification, IEEE Trans. on Information, vol. IT-20, 472-478, 1974.

15. A.A.Chilingarian, S.Ch Galfayan, Calculation of Bayes Risk by KNN Method. Stat. Problem of Control, Vilnius, p.66-76, 1984.

16. J.D.Hey, An Introduction to Bayesian Statistical Inference, Martin Robertson, 1983.

17. С.Х.Галфаян, А.М.Дунаевский, и др. Многомерный анализ данных, получаемых в экспериментах с РЭК и в ШАЛ, препринт ФИАН N 332, 1986.

18. P.J.Bickel, D.A.Freedman, Some Asymptotic Theory for the Bootstrap, Ann.Stat. v.9, p.1198, 1981.

19. B.Efron, The Jackkife Bootstrap and other Resampling Plans, Society for Industrial and Applied Mathematics, Philadelphia, 1982.

The manuscript was received October 11, 1989

Г.Э. ЗАЗЯН, А.А. ЧИЛИНГАРЯН

КЛАССИФИКАЦИОННЫЙ МЕТОД ОПРЕДЕЛЕНИЯ МАССОВОГО СОСТАВА ПКИ В ОБЛАСТИ ЭНЕРГИЙ $E_0 > 10^{15}$ эВ

(на английском языке, перевод Г.А. Папяна)
Редактор Л.П. Мукаян
Технический редактор А.С. Абрамян

Подписано в печать 25/XII-89 БФ-11436	Формат 60x84x16
Офсетная печать. Уч.изд.л. 1,5	Тираж 299 экз. Ц. 22 к.
Зак.тип. 1972	Индекс 3649

Отпечатано в Ереванском физическом институте
Ереван-36, ул. Братьев Алиханян 2.

The address for requests:
Information Department
Yerevan Physics Institute
Alikhanian Brothers 2,
Yrevan, 375036
Armenia, USSR

ИНДЕКС 3649

ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ