Preprint YERPHI-1209(86)-89

ԵՐԵՎԱՆԻ ՖԻԶԻԿԱՅԻ ԻՆՍՏԻՏՈՒՏ
ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ
YEREVAN PHYSICS INSTITUTE

A.A.CHILINGARIAN, G.Z.ZAZIAN

ON THE POSSIBILITY OF INVESTIGATION OF THE ENERGY SPECTRA OF PCR PROTONS AND NUCLEI IN THE ENERGY RANGE FROM 10^{15} TO 10^{17}eV USING EAS DATA

ЦНИИатоминформ
ЕРЕВАН - 1989

Նախնատիպ ԵրՖԻ-1209(86)-89

Հ.Զ.ԶԱԶՅԱՆ, Ա.Ա.ՉԻԼԻՆԳԱՐՅԱՆ

ՍԿԶԲՆԱԿԱՆ ՏԻԵԶԵՐԱԿԱՆ ՃԱՌԱԳԱՅԹՄԱՆ ՊՐՈՏՈՆՆԵՐԻ ԵՎ ՄԻՋՈՒԿՆԵՐԻ ԷՆԵՐԳԵՏԻԿ ՍՊԵԿՏՐՆԵՐԻ ՈՒՍՈՒՄՆԱՍԻՐՄԱՆ ՀՆԱՐԱՎՈՐՈՒԹՅԱՆ ՄԱՍԻՆ $10^{15}-10^{17}$ ԷՎ ԷՆԵՐԳԻԱՆԵՐԻ ՏԻՐՈՒՅԹՈՒՄ՝ ԸՍՏ ԼՄՀ ՏՎՅԱԼՆԵՐԻ

Նկարագրված է ոչ պարամետրիկ ռեգրեսի վրա հիմնված սկզբնական մասնիկի էներգիայի գնահատման նոր բազմաչափ եղանակ: Առաջարկվող եղանակը ԼՄՀ բազմաչափ դասակարգման հետ համատեղ թույլ կտա՝ ըստ ԼՄՀ էլեկտրոնա-ֆոտոնային և մյուոնային բաղադրիչների բնութագրերի որոշել պրոտոնների և միջուկների ընկնող «փնջերի» պարամետրերը: Այդ մասնիկների թիրախի հետ փոխազդեցության արդյունքների գրանցումը և հետազոտումը թույլ կտա ուսումնասիրել PA և AA փոխազդեցությունը $10^{15}-10^{17}$ էՎ էներգիաների տիրույթում:

Երևանի ֆիզիկայի ինստիտուտ
Երևան 1989

Preprint YERPHI-1209(86)-89

A.A. CHILINGARIAN, G.Z. ZAZIAN

ON THE POSSIBILITY OF INVESTIGATION OF THE ENERGY SPECTRA OF PCR PROTONS AND NUCLEI IN THE ENERGY RANGE FROM 10^{15} TO 10^{17}eV USING EAS DATA

A new multivariate method of incident particle energy estimation based on the nonparametric regression method is described. The method proposed, together with the multivariate EAS classification one, allows to determine the parameters of incident proton and nuclear "beams". Detection and investigation of the products of interaction of these particles with the atmosphere (target) will allow to study PA and AA interactions at energies from 10^{15} to 10^{17}eV.

Yerevan Physics Institute
Yerevan 1989

Препринт ЕФИ-1209(86)-1989

Г.Э. ЗАЗЯН, А.А. ЧИЛИНГАРЯН

О ВОЗМОЖНОСТИ ИССЛЕДОВАНИЯ ЭНЕРГЕТИЧЕСКИХ СПЕКТРОВ ПРОТОНОВ И ЯДЕР ПКИ В ОБЛАСТИ ЭНЕРГИЙ 10^{15}-10^{17}эВ ПО ДАННЫМ ШАЛ

В настоящей работе описан новый многомерный способ оценки энергии первичной частицы, основанный на непараметрической регрессии. Предлагаемый метод, совместно с методом многомерной классификации ШАЛ, позволяет по характеристикам электронно-фотонного и мюонного компонентов ШАЛ определять параметры падающих "пучков" протонов и ядер. Регистрация и исследование продуктов взаимодействия этих частиц с атмосферой (мишенью) сделает возможным изучение PA и AA взаимодействий при энергиях 10^{15}-10^{17}эВ.

Ереванский физический институт
Ереван 1990

1. Introduction

The primary cosmic ray (PCR) energy spectrum has been investigated up to energy 10^{15}eV in satellite and balloon experiments. The measurements based on detection of EAS mainly fit to the satellite and balloon data in the energy range up to 10^{15}eV [1]. The integral spectrum index changes from 1.6 to 2 in the energy range $5\cdot10^{14}$-10^{16}eV – the so-called spectrum "knee".

A possible step to description of spectrum breaking is the study of the energy spectra of separate groups of nuclei and protons in the region of the knee, since by them one can judge about validity of different models of the origin and propagation of CR. The proton energy spectra are studied up to 10^{15}eV by the JACEE collaboration [2] and in satellite experiments – up to 10^{14}eV [3]. Selecting the proton showers via the presence of high-energy hadrons in the calorimeter (E_h/E_0~0.25), the energy spectrum of protons with energy 10^{15}eV has been obtained in Ref.[4]. This method is based on the fact that the incident particle energy dissipation is more intense in the cascades initiated by incident nuclei. But due to large fluctuations in the portion of energy transferred to hadrons at a fixed initial energy, such selection may considerably reduce the number of proton showers. In Ref.[5] the fractions of different groups of nuclei in PCR were estimated and the energy spectra of the corresponding nuclei in the energy range 10^{15}–10^{16}eV were obtained by the method of solving the inverse problem. The results of Refs.[4,5] mainly coincide with the direct experiment data.

In this paper we shall show how the classification method of simultaneous analysis of simular and experimental data developed in Ref.[6], together with the new multivariate method of estimation of incident particle energy, allows to recover the energy characteristics of fluxes of different nuclei incident on the atmosphere.

2. EAS Classification

The detailed description of the multivariate method of EAS classification is given in Ref.[7]. The main parameters of the model used are also given. Table 1 can serve as an illustration of the results obtained. The diagonal elements of this matrix show the probability for a correct classification of four groups of nuclei (P & α, CNO, H, VH), the nondiagonal elements - the probability of possible errors. It is seen from Table 1 that the protons and iron group nuclei (A=50–56) are classified reliably (with an efficiency of ~70–80%), which gives reason to hope for a reliable separation of these events from the total PCR flux.

Note, that the main difference of the classification method from the ones used earlier for solution of the inverse problem [3] is that the object of analysis are not the alternative distributions, but each experimental event. It means that we determine not only the fraction of some nuclei group in the primary flux, but also belonging of each event (together with the corresponding error) to a certain group.

The used Bayes decision rules based on the adaptive Parzen estimation of multivariate probability density provide a complete account of a priori information obtained by simulation and, what is more, allow to choose an optimal set of features by which the events are classified. Not dwelling upon this method (see Refs.[6–9]), in the following section we shall briefly describe the nonparametric regression method used also to estimate the hadron energy according to X-ray emulsion chamber data [9].

3. The Nonparametric Regression Method

As ground for estimation of the primary particle energy serves the fact of its correlation with the EAS parameters measured. Table 2 presents the coefficients of the primary particle energy correlation with various shower parameters. It is seen that though the total number of electrons in a shower (N_e) is the main parameter used to estimate the primary energy, the characteristics of the muon component of EAS correlate with E_0 somewhat better. That is why our purpose was to investigate the possibility of improvement of the accuracy of estimation of the primary particle energy via the characteristics of the electron-photon and the muon components of EAS.

First, some words about general formulation of the regression problem (we'll mainly follow Ref.[10]). Suppose, a flux of particles is sporadically and independently incident on the atmosphere in accordance with some spectrum f(E). Then these particles, undergoing random collisions and interactions with air atom nuclei, initiate an extensive air shower, the parameters of which are registered by the experimental setup, i.e. each value of E is put into coincidence with some random vector of measurements, $\vec{X}$, according to some conditional probability density $P(\vec{X}/E)$.

The peculiarity of solution of the regression problem in the cosmic-ray physics is the fact that neither the true spectrum $f(E)$ nor the conditional density $P(\vec{X}/E)$ are known in the general case, but there is a training sequence $\{E_i, X_i\}$, $i=1,M_{TS}$ (obtained by simulation) and it is required to "recover" the regression $E=E(X)$ by this sequence (M_{TS} is the number of events in the training sample).

In the absence of systematic errors $M\{X\}_E=X(E)$ (the mathematical expectation of random vector measurement at a fixed independent variable (energy) is equal to the regression function value in that point) this problem is reduced to one of minimization of the average risk

$$I(\alpha) = \int (E-F(\vec{X},\alpha))^2 \, P(X,E) \, dX \, dE \,, \qquad (1)$$

where $F(X,\alpha)$ is some functional family depending on the parameter α, $P(X,E)\sim P(X)P(X/E)$ is the probability density function. If there is available a priori information about the form of probability function and the chosen functional family $F(\vec{X},\alpha)$ is not too complex, then the regression problem can be solved by the least mean squares or the maximal likelihood standard methods.

Due to the complicated stochastic picture of particles and nuclei passing through the atmosphere and the detectors, we have not to expect a standard probability interpretation of all random processes, that is why we have chosen a method based on a nonparametric way of treatment of a priori information, which does not impose any structure and totally uses the information carried by TS.

The method is based on the obvious fact that the events close to some metric (usually the Mahalanobis metric [11] is used) in the feature space have similar energy – the compactness hypothesis. The method based on consideration of the "nearest neighbours" is first analyzed in Ref.[12]. In this work it was shown that when the number of the nearest neighbours, K, and the total number of events in TS, M, tend to infinity so that $K/M\to 0$, then the risk of the procedure tends to the minimum achievable Bayes risk and even the use of one neighbour increases the risk only twice as compared to the Bayes risk. The uniform consistency of the following estimate is shown in Ref.[13]:

$$\hat{E}(X) = \sum_{i=1}^{k} C_i E_{[i]}(X), \qquad \sum_{i=1}^{k} C_i = 1, \qquad (2)$$

where $E_{[i]}(X)$ is the value of the independent variable (energy) of the i-th nearest neighbour of the event $\vec{X}$ in the feature space.

The weight coefficients C_i are optimized by TS so that some quality function, e.g., the mean-square error (MSE) of estimation, is minimized,

$$MSE = \sqrt{\sum_{i=1}^{m} (E_i-\hat{E}_{(i)}(X))^2/M}, \qquad (3)$$

where the index (i) means that the i-th event, the energy of which is estimated, is temporarily removed from TS (leave-one-out test). Despite the fact that the nonparametric procedures are optimal under unlimited sampling, for the case of finite samples there are practically no theoretical and practical recommendations on the choice of the method parameters (e.g., the number of nearest neighbours). That is why we apply the estimate adaptation ideology to the regression analysis, which was developed for multivariate nonparametric estimation of density function [14]. In this approach there are simultaneously calculated several estimates corresponding to different method parameters. The median of the ordered sequence is taken as final estimate.

4. Results of Calculations

To estimate the accuracy of primary particle energy estimation by the method of nonparametric regression, there were generated showers with initial energy $E_0 > 500$TeV. The preprocessing of showers was carried out by the data handling algorithms used in the Tien-Shan experiment [15]. The detector-induced fluctuations were taken into account when determining the characteristics of the electron-photon and the muon components. After preprocessing of showers, part of them were used as TS and another - as "experimental" data. Table 3 presents the results of estimation of the energy of "pseudo-experimental" events in various ranges of N_e initiated by incident protons and nuclei with A>24. The relative mean-square errors are presented ibid. Figs.1 and 2 show the histograms of relative mean-square deviations (RMSD), for proton and nuclear events, calculated via:

$$RMSE_i = (E_i - \hat{E}_i)/E_i, \quad i = 1,M \qquad (6)$$

where E_i is the true energy of the event estimated and $\hat{E}_i$ is its nonparametric estimate. The features used are N_e and N_μ, the number of the kernel widths used is 5 - from 0.1 to 10. It is seen from these figures that in case of events initiated by

incident nuclei, the mean-square error of estimation is ~2.5 times smaller than that for proton-induced events, which is due to a stronger correlation of EAS parameters with the initial energy in case of nuclear events as compared to the the proton-initiated events (see Table 2).

Figs.3 and 4 show the "true" integral energy spectra of protons and nuclei with A>24 (the true spectrum corresponds to 100% of correct classification of protons and nuclei and to zero error in determination of primary particle energy). The "experimental" spectra obtained as a result of classification of events initiated by the technique presented in Ref.[7] and then - by nonparametric estimation of energy, are presented ibid. As is seen from these figures, there is a satisfactory agreement between the "true" and estimated energy spectra. The energy of events is somewhat overestimated for the "proton" events (the measured EAS characteristics of misclassified nuclei are attributed to protons with high energy) and are underestimated in case of selected events attributed to heavy nuclei (vice versa). The presence of such distortions leads to some change in the index of the integral energy spectrum of incident protons and nuclei. To obtain a quantitative estimate of the degree of distortion of the index of the integral energy spectrum of protons and heavy nuclei, the corresponding distributions were approximated by power-law, with the help of the minimization program "FUMILI" from the CERN program-library. For true proton events the spectrum is approximated by:

$$I_p^{tr}(E>500) = (4.75 \pm 0.036)E^{-1.761 \pm 0.013}$$

for selected proton events it is approximated by:

$$I_p^{est}(E>500)=(4.59 \pm 0.061)E^{-1.702 \pm 0.021}.$$

For the events initiated by heavy nuclei it is approximated by

$$I_A^{tr}(E>500)=(4.56 \pm 0.075)E^{-1.691 \pm 0.030}$$

$$I_A^{est}(E>500)=(4.84 \pm 0.099)E^{-1.770 \pm 0.037}$$

respectively.

It follows from these formulae that the relative error at determination of the integral energy spectrum index is ~3% for incident protons and ~5% for incident nuclei with $A>24$.

5. Conclusion

The method proposed for studying the energy spectra of PCR in the energy range from 10^{15} to 10^{17}eV via EAS data is based on:

a) high reliability of multivariate classification of EAS (efficiency of identification of protons and nuclei in a PCR flux is ~70-80%);

b) high accuracy of primary particle energy determination by the nonparametric regression method (relative mean-square error is 10-25%).

By the characteristics of the electron-photon and the muon components of EAS we can determine the parameters of proton and nuclear "beams" incident on the interface with the atmosphere.

Detection and investigation of the products of interaction of these particles with the atmosphere (target) will allow to study PA and AA interactions at energies 10^{15}-10^{17}eV.

The authors thank A.M.Dunaevsky for discussions and E.A.Mamidjanian for a stimulating interest in this work.

Table 1

The Bayes Error Matrix

$1\cdot10^5<N_e<2\cdot10^5$

	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 2

The coefficients of correlation of the characteristics of the electron-photon and the muon components of EAS with initial energy $1\cdot10^5<N_e<2\cdot10^5$

	N_e	$N_\mu(E_\mu>5\text{GeV})$	S	$N_\mu(E_\mu>200\text{GeV})$	$\sum E_\mu(E_\mu>200\text{GeV})$
P	0.355	0.730	0.33	0.678	0.584
Fe	0.495	0.953	0.23	0.899	0.892

Table 3

Mean-square errors of estimation of the energy of protons and nuclei with A>24

$\langle N_e \rangle / 10^5$	0.66±0.196	1.404±0.281	2.716±0.534	9.758±1.079
$\langle N_\mu \rangle / 10^3$	2.74±0.258	3.839±0.335	5.881±0.667	14.545±1.01
N^P_{TS}	864	913	484	402
N^P_{exp}	439	465	256	216
$RMSE^P$(%)	20	24.3	25	25
N^A_{TS}	377	357	184	128
N^A_{exp}	184	225	102	73
$RMSE^A$(%)	10.1	10.6	9.7	10.6

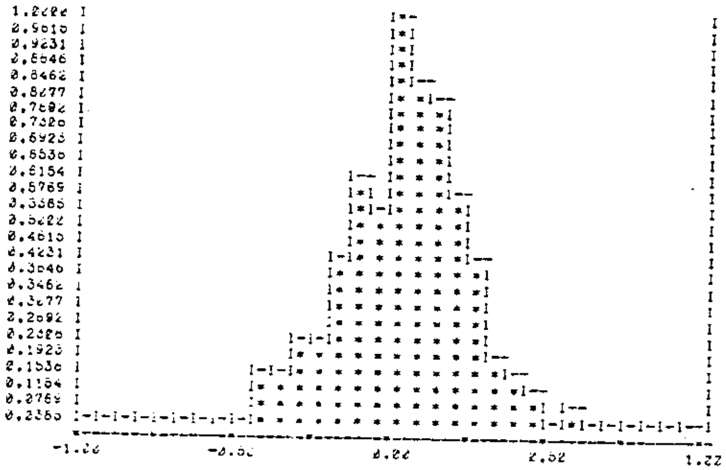

Fig.1

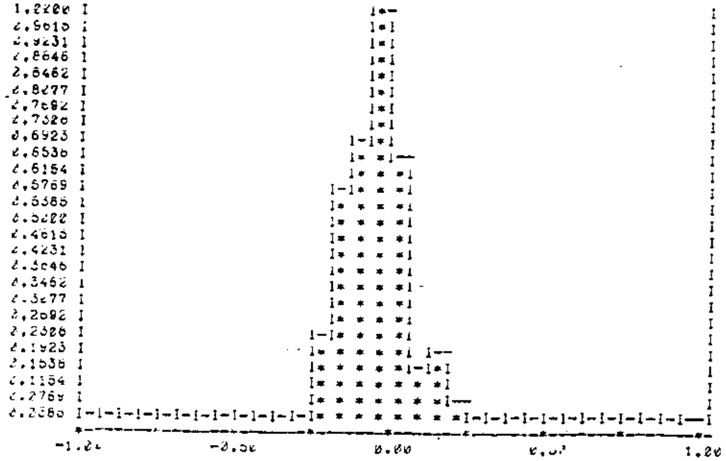

Fig.2

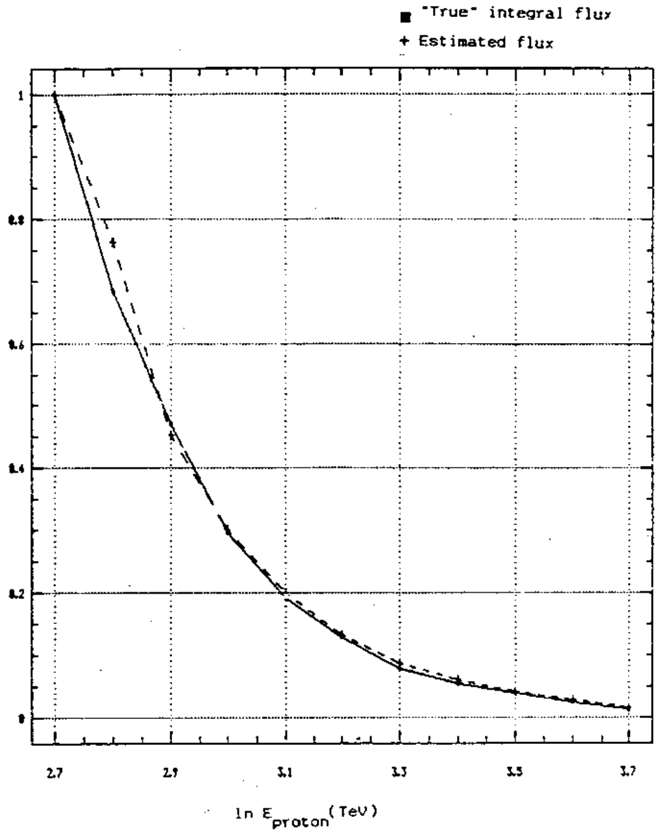
"True" integral flux
+ Estimated flux
1
0.8
0.6
0.4
0.2
0
2.7
2.9
3.1
3.3
3.5
3.7
ln E_{proton}(TeV)

Fig.3

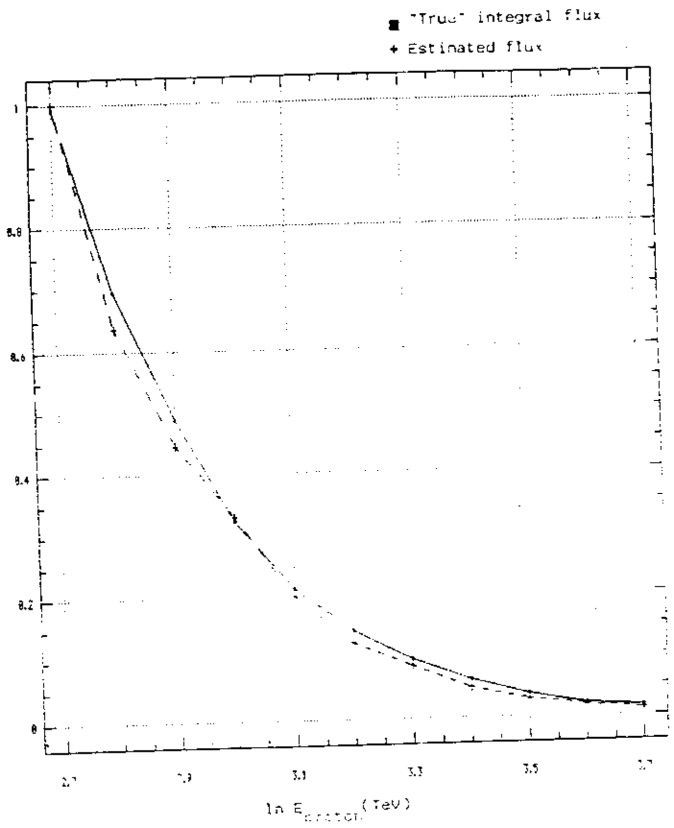
"True" integral flux
+ Estimated flux
1
0.8
0.6
0.4
0.2
0
2.7
2.9
3.1
3.3
3.5
3.7
ln E_{proton}(TeV)

Fig.4

Figure Captions

Fig.1 Distribution of the relative error of energy estimates of incident protons within $1\cdot10^5<N_e<2\cdot10^5$.

Fig.2 Distribution of the relative error of energy estimates of incident nuclei with $A>24$ within $1\cdot10^5<N_e<2\cdot10^5$.

References

1. Никольский С.И. Энергетический спектр и ядерный состав первичных космических лучей, Проблемы физики космических лучей, Москва, Наука, с.164-185, 1987.
2. JACEE collaboration. Energy spectra of primary protons and helium nuclei in the energy range 10^{12}-10^{15}eV from JACEE, Proc. 20th ICRC, vol.1, p.371, Moscow, 1986.
3. Ivanenko I.P. et al.. Spectrum and charge composition of cosmic rays with energy from 2 to 100 TeV, 5th Int. Symp. on Very High Energy Cosmic Ray Interaction, Lodz, Poland, p.197, 1988.
4. Chubenko A.P., Krutikova N.P. et al., Primary protons, inclusive cross-section in P-A interaction and their relation to energy spectra of hadrons in the atmosphere, Proc. of 20th ICRC, vol.1, p.373, Moscow, 1986.
5. Стаменов И.Н. Исследования состава первичного космического излучения с помощью ШАЛ, Автореферат дис. на соиск. учен. степени доктора физ.-мат. наук, Москва, ФИАН, 1981.
6. Чилингарян А.А. Анализ высокоэнергичных компонент космического излучения методами непараметрической статистики. Автореферат дис. на соиск. учен. степени кандидата физ.-мат. наук, Ереван, ЕФИ, 1984.
7. Заэян Г.З., Чилингарян А.А. Классификационный метод определения массового состава ПКИ при энергии 10^{15}-10^{17}эВ по данным ШАЛ. Препринт ЕФИ-1204(81), Ереван 1989.
8. Галфаян С.Х., Дунаевский А.М. и др. Многомерный анализ данных получаемых в экспериментах с РЭК и в ШАЛ, Препринт ФИАН, №332, Москва, 1987.

9. Зазян М.З., Чилингарян А.А. Непараметрический метод определения энергии адронов в экспериментах с РЭК, Препринт ЕФИ-1217(3), Ереван 1990.
10. Вапник В.Н. Восстановление зависимостей по эмпирическим данным, Москва, Наука, 1974.
11. Mahanalobis P.C. On the generalization distance in statistics, Proc. of Nat. Inst. of Science of India, 1963, vol.12, p.49-55.
12. Cover T.M. Estimation by the nearest neighbor rule, IEEE Trans. on Information theory, 1968, vol.IT-14, p.50-55.
13. Devroye A.P. The uniform convergence of Nearest Neighbor regression function estimator and their application in optimization, IEEE Tarns. on Information theory, 1978, vol.IT-24, p.142-151.
14. Chilingarian A.A. Statistical decisions under nonparametric a priori information, Comp. Phys. Comm., 1989, vol.54, p.381-390.
15. Никольская Н.М., Стаменов И.Н. Исследование алгоритма статистической обработки экспериментальных данных и характеристик управляющей системы, отбирающей ливни на Тянь-Шаньской комплексной установке ШАЛ. Препринт ФИАН, №125, Москва, 1975.

The manuscript was received October 15, 1989

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

Г.З. ЗАЗЯН, А.А. ЧИЛИНГАРЯН

О ВОЗМОЖНОСТИ ИССЛЕДОВАНИЯ ЭНЕРГЕТИЧЕСКИХ СПЕКТРОВ ПРОТОНОВ И ЯДЕР ПКИ В ОБЛАСТИ ЭНЕРГИЙ 10^{15}-10^{17}эВ ПО ДАННЫМ ШАЛ

(на английском языке, перевод Г.А. Папяна)

Редактор Л.П. Мукаян

Технический редактор А.С. Абрамян

Подписано в печать 26/XII-89 ВФ-11437 Формат 60x84x16
Офсетная печать. Уч.изд.л. 1.0 Тираж 299 экз. Ц.15 к.
Зак.тип. 1973 Индекс 3649

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

ИНДЕКС 3649

ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ