ЕФИ-8I9(46)-85

ԵՐԵՎԱՆԻ ՖԻԶԻԿԱՅԻ ԻՆՍՏԻՏՈՒՏ
ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ

N.Z. AKOPOV, Sv.Kh.ARUTUNIAN, A.A.CHILINGARIAN,
S.Kh.GALFAYAN, V.Kh.MATEVOSIAN, M.Z.ZAZIAN

~~THE~~ DESIGN PRINCIPLE AND STRUCTURE
OF "ANI" DATA CENTRE

ЦНИИатоминформ

ЕРЕВАН-1985

[illegible]

Н.Н.АКОПОВ, Св.Х.АРУТЮНЯН, С.П.МАЛЯН,
М.З.ЗАЗЯН, В.А.МАТЕВОСЯН и А.А.ОГАНЕСЯН

ОБЩИЕ ПРИНЦИПЫ И СТРУКТУРА ЦЕНТРА ДАННЫХ "АНИ"

Обосновываются основные принципы, лежащие в основе выбора методов и алгоритмов пакета прикладных программ статистического анализа "АНИ". Непараметрические процедуры позволяют создать эффективную методику совместного анализа результатов вычислительных экспериментов и данных экспериментов в космических лучах. Приводится описание реляционной базы данных, обеспечивающей унифицированное хранение информации, защиту, обновление и удаление данных, а также быстрый и удобный доступ к данным.

Ереванский физический институт

Ереван 1988

[illegible] prepared [illegible] experiment is intended for [illegible] [illegible] investigations in energy range up to 10^{16} eV [illegible] for the first time it is proposed to join the information [illegible] such installations as huge ionization calorimeter [illegible] emulsion chambers, high acceptance magnet spectr[illegible] arrays of hodoscopic counters.

Therefore it is very actual to create a consistent multiple statistic analysis system to process such an informa- tion, concerning various characteristics of different partic- les /2/. It is necessary to mention that indirect nature of CR investigations requires a series of particle traversal si- mulations through the atmosphere and experimental setups for obtained data and theory predictions comparison.

Thus, only the combined analysis of simulation results [illegible] experimental data enables one to draw conclusions on the nature of [illegible] interaction and the composition of the pri[illegible] [illegible] flux of [illegible] radiation at superaccelerator energies.

[illegible] of statistical procedures for combined ana- lysis is the problem of principle. It must be solved in the framework of an approach more general, than the theory of ma- thematical statistics, [illegible] additional

[illegible] mainly in the determination of the shape of the distribution function of motion of random particles, connected with experiment or simulation. The most obtain is the assumption of independently normally distributed random errors. However, as numerous investigations show, the reality is much complicated than any model, and deviations of processed data nature from proposed, lead to considerable decrease in statistical procedures effectiveness.

As mentioned in /3/ in assumption making we are guided by considerations, represented a mixture of experience and physical intuition. The danger is that, if adopted assumptions turn to be wrong, the analysis remaining valid in the framework of chosen model may have highly indirect connection to [illegible]ity.

This danger is overcome in the applied statistic - a technology in which the degree of adequetness of chosen model and reality is a crucial factor, that determines reliability of utilized data handling methods /4/. In applied statistic technology one tries to make more realistic assumptions. First of all is expressed in the refuse of known parametric shape of distribution function assumption, as the complicated analytical expressions can with high accuracy describe a given random set of data, but - worse the desired intrinsic dependence of physical values /5/.

Much weaker assumption about data lies in the fact, that the features of the phenomenon under consideration are somehow

[illegible]

[illegible] pattern recognition [illegible]

[illegible] statistics [illegible]

[illegible] [illegible] [illegible]

[illegible] [illegible] [illegible]

modal, unsymmetrical [illegible]

). As we see above the peculiarity of statistic[illegible] sions in cosmic ray physics consists in combined [illegible] of simulated and experimental data. The information on investigated physical problem contained in so called Training samples (TS) - a simulation results with controled input parameters. TS presented all possible variations of physical values due to individual differences, admittable in the given state limits.

Training samples reflected a stochastic nature of CR interactions, existence of many strong and electromagnetic interactions channels and probabilistic character of detector operation.

It is necessary to check that both the mathematical definition of problem as well as the data handling method depend on available a priory information type. Training samples provide us only nonparametric mode of a priory information. and so our aim is to create a self-consistent nonparametric statistic technology to carry out multidimensional data processing .

Based on pattern recognition concept a package of applied

programs for nonparametric multiple statistical analysis of CR data was developed. The use of package permits the computer realization of the following procedures:

1. Choice of theoretical model most adequately fitting experimental data.

2. Selection of events of definite type.

3. Identification of particles and interaction processes.

4. Search for cluster in multi-dimensional empirical distributions.

The principle features of package are:

a) Utilization of Bayesian decision rules /8/.

b) KNN (K-Nearest Neighbour) adaptive probability density estimation. /8/.

c) Use of Bayesian risk as a closeness measure of multidimensional distributions. /8/.

The package consists of independent routines, characterized by common procedure of data definition, similar programming technique and matched resources.

The routines of package are realized in FORTRAN-4 and some modules of service data handling programs are in BESM-6 Assembler.

The correctness and sound implementation of routiness were provided by an extensive testing on problems, that are believed to be typical of those encountered in a general environment.

4. The relational type data base organization provides the unification of information storage, dependable protection, renewal of data. The information is organized in the form of named data arrays (files), and the names contain the

valuable information to make possible the associative access to data.

The main attributes of relation structures are the name of file, the number of simulation realizations (number of strings number of features, and the weight. The weight is used to compile the training samples according to definite energetic spectrum and chemical composition of primary flux.

The service software includes rather simple and convenient diagnostic message, that practically prevents incorrect utilization of data and besides allows one to efficiently make changes in the structure of stored data. There are also security facilities against unsanctional access to the information. The provision for interactive mode of operation and exploratory plotting analysis was made.

The applied program package and the relational data base are intended for use in "ANI" data analysis and technique development centre, where the information from most important experiments in the field of cosmic ray physics is planned to gather.

The accumulation and systematization of experimental and simulation data permit one to carry out the comparison of obtained data with up to date theories, to predict the behavior of physical parameters at impossible so far energies, to perform experiments optimal design.

References

1. Danilova T.V., Dunaevsky A.M., [illegible]

   A project of experimental investigations of [illegible] energy range 10^3-10^5 GeV (experiment "ANI"), [illegible] of Scin. of Arm.SSR, vol.17, pp. [illegible], 198[illegible].

2. Chilingarian A.A. The development of statis[illegible] in cosmic ray physics 18 - ICRC, vol.5, p.[illegible] 17, [illegible] 1983.

3. Bickel P.J., Doksam K.A. Mathematical statistics Holden-Day. INC. San-Francisco-Dusseldorf, Johannesburg -London-Panama-Singapure-Sydney

4. Aivazyan S.A., Yanyakov I.S., Meshalkin L.D. Applied Statistics, Finansy i statistika, Moscow, 1983.

5. Vapnik V.N. Algorithms and programms of dependence reconstraction, Nauka, Moscow, 1984.

6. Granunder U. Regular Structures Springer-verlag New York-Heidelberg-Berlin, 1981.

7. Pearson E.S. Studies in the history of probability and statistics-Biometrica. 1965, vol.52, p.3 - 18.

8. Chilingarian A.A. On statistical methods of high energy particles identification. Proc. of Symp. on High Energy Particles Transitional Radiation, Yerevan 1984, pp. 47[illegible]-4[illegible].

The manuscript was received 3[illegible] May 1985

Н.З.АКОПОВ, Св.Х.АРУТЮНЯН, С.Х.ГАЛФАЯН, М.З.ЗАЗЯН,
В.Х.МАТЕВОСЯН, А.А.ЧИЛИНГАРЯН

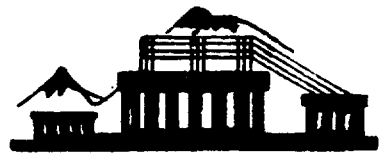
ОБЩИЕ ПРИНЦИПЫ И СТРУКТУРА ЦЕНТРА ДАННЫХ АНИ

Редактор Л.П.Мукаян
Технический редактор А.С.Абрамян

Подписано в печать II/X-85г. ВФ-09038 Формат 60x84/I6
Офсетная печать. Уч.изд.л. 0,5 Тираж 299 экз. Ц. 6 к.
Зак.тип.№ 437 Индекс 3624

Отпечатано в Ереванском физическом институте
Ереван 36, Маркаряна 2

индекс 3624



ЕРЕВАНСКИЙ ФИЗИЧЕСКИЙ ИНСТИТУТ