MULTIPLE COMPARISON OF SIMULATIONS WITH DATA OBTAINED IN EAS EXPERIMENTS AND BY X-RAY EMULSION METHODS

G. Kh. Gulfayan, A. M. Dunaevskii, M. Z. Zazyan,
and A. A. Chilingaryan

Izvestiya Akademii Nauk SSSR. Seriya Fizicheskaya,
Vol. 50, No. 11, pp. 2146-2148, 1986

UDC 539.171.017

The paper discusses the choice of strong-interaction model to give the most accurate description of the measurements and the determination of the chemical composition of the primary flux.

1. Current experiments on cosmic rays involve various methods of obtaining data, which have to be compared with computer experiments based on models for the transmission of radiation through the atmosphere and detectors. Monte Carlo simulation is a unique way of comparing hypotheses in complicated indirect experiments. Computer models provide information on possible states, and realization sets in models (training sets) represent different hypotheses.

In a training set, one varies all the physical quantities related to the individual sets of features permissible within a given state. The learning sets reflect the stochastic character of cosmic-ray interactions, including the numerous channels for strong electromagnetic interactions, as well as the random character of the phenomena used in the recording instruments [1].

Statistical procedures are used in analyzing measurements and computer data, which is an essential aspect of simulation and requires a more general approach than the theory of mathematical statistics, in which the model and the additional assumptions are considered as given. Here we present an approach in which we avoid arbitrary assumptions difficult to check about the form of the probability densities and merely assume that the characteristic features of any measurable phenomenon in some way reflect the structure of the learning set. This means that hypotheses form regular structures within a given research [illegible], which can be interpreted by means of pattern-recognition theory [2-4].

2. The mathematical characteristics and processing methods in each particular case are dependent on the a priori information. Computer models provide for nonparametric a priori information specification, so our task is to set up a consistent system of nonparametric procedures for handling the following processing and interpretation tasks [5].

1) A physical problem involves choosing the theoretical model best describing the experimental data, including the identification of elementary particles and interaction processes and distinguishing events of a particular type.

2) The statistical problem is that of quantitative comparison for multidimensional distributions representing learning sets, with classification from mixtures of distributions, learning with "a teacher", and local probability density estimation.

This list allows for a set of particular solutions. It is not our purpose to examine these in detail, but we give some features of the methods used: a) Bayes decision rules [6], b) adaptive probability density estimation [7] by the method of K nearest neighbors KNN, and c) the Bayes risk as a measure of similarity between training sets [8].

3. Here we examine the limits to the application of these methods and evaluate the expected accuracy in determining certain physical quantities. The experimental data are replaced by model ones. One knows the true data assignment (the results from computer experiments with controllable parameters), so one can readily identify the errors arising in processing.

© 1986 by Allerton Press, Inc.

Table 1

Results of Recovering the Proportions of γ-h Families Generated by Iron Nuclei

Apparatus	Features	$M4 + Fe_{min}$		$M6 + Fe_{max}$	
		P_H	$\hat{P}_H$	P_H	$\hat{P}_H$
γ-block	$\Sigma E_\gamma, \bar{R}_\gamma, a_\gamma$	0,05	0,029±0,049	0,05	0,042±0,038
		0,07	0,063±0,048	0,07	0,082±0,037
		0,1	0,094±0,045	0,1	0,1±0,036
γ-block + h-block	$\bar{R}_\gamma, a_h, E_h$	0,05	0,035±0,030	0,05	0,037±0,024
		0,07	0,062±0,029	0,07	0,059±0,024
		0,1	0,087±0,028	0,1	0,1±0,023
γ-block + shower part	$\Sigma E_\gamma, \bar{R}_\gamma, a_\gamma, E_0$	0,05	0,001±0,031	0,05	0,036±0,025
		0,07	0,029±0,030	0,07	0,058±0,025
		0,1	0,055±0,029	0,1	0,084±0,024

The models used include the pure scaling M6 and models with increasing cross section M4, Fe_{min}, and Fe_{max}. The initial features are the total energy of the γ-rays in a famil the multiplicity, the mean distance of the γ-rays from the center of gravity, the azimuth asymmetry, and the sphericity. These characteristics were calculated for three different threshold conditions. We also used the analogous characteristics calculated from the hadrons, the initial energy, and so on. In all, we considered 42 features [9,10].

To identify one of the alternative models, it is necessary to select combinations of the initial features providing reliable differences. We used the Bayes risk R^B, which is the probability of making an error in classifying pairwise-compared training sets by mean of an optimal decision rule, which is used to characterize the closeness between mul dimensional distributions. This probability is close to 50% if the sets are identical, but it tends to zero if they differ greatly. A preliminary study was made on the individual features not only by standard methods (by calculating P, the significance criteria for the difference of the means in two sets by means of the Students t test, the Wilcoxon test, and the Kolmogorov-Smirnov test), and also by calculating the Bayes risk. All the methods gave similar results. Also, it was shown that this method of estimating R^B on sets of different sizes from a single population was unbiased and independent [11].

We now consider distinguishing γ-h families generated by iron nuclei [11]. As training sets we took events generated from the M4 and M6 models, which contained only light nuclei, and events generated from the Fe_{min} and Fe_{max} models, which contained only events from primary iron nuclei. We classified the events generated by these models (not appearing in the training set) with various set proportions P_H of events from iron nuclei. Tha choice means that we assume identical distributions for the experimental events and the events in the training set, i.e., that we have chosen the strong-interaction model correc ly. Table 1 shows that the recovered proportion $\hat{P}_H$ is fairly close to the true one. However, if we transfer to the more realistic case of inexact knowledge of the strong-interaction model (this case is simulated by the training and check sets being generated by different models), then one cannot recover P_H with acceptable accuracy. Therefore, one should either use models whose characteristics are closer to the experimental data than the M4 and M6 or else find a combination of features insensitive to the strong-inter ic model but at the same time sensitive to the type of primary nucleus.

4. These methods enable one to perform a quantitative comparison of the model and experimental multidimensional distributions for the characteristics of families and EAS, which may enable one to enrich the distinct events with families from heavy nuclei and subsequently to analyze the $Fe-^{14}N$ interaction at energies $\geq 10^4$ TeV.

REFERENCES

1. A. A. Chilingaryan, Aspects of Nuclear Science and Engineering, Series Technical Physics and Experiment [in Russian], issue 7, p. 59, 1981.

2. U. Grenander, Lectures on Pattern Theory, ch. 3 [Russian translation], Mir, Mos-
, p. 249, 1984.
3. S. A. Auvazyan, I. S. Enyukov, and L. D. Meshalkin, Applied Statistics [in Rus-
an], Finansy i Statistika, Moscow, p. 470, 1983.
4. K. Fukunaza, Introduction to the Statistical Theory of Pattern Recognition [Rus-
an translation], Nauka, Moscow, pp. 57, 176, 188, 267, 1979.
5. N. Z. Akopov, S. Kh. Arutyunyan, S. Kh. Gulfayan, et al., Preprint ErFI, no. 819
9)-84, Erevan, 1985.
6. A. A. Chilingaryan, Proceedings of the Symposium on Transition Radiation from
gh-Energy Particles [in Russian], Erevan, p. 478, 1984.
7. A. A. Chilingaryan, Preprint ErFI, no. 818(45)-85, Erevan, 1985.
8. S. Kh. Gulfayan and A. A. Chilingaryan, in: Control Problems [in Russian], issue
, Izd, Int. Mat. Kib..AN Lit. SSR, Vilnius, p. 66, 1985.
9. A. M. Duraevskii and A. V. Uryson, Preprint FIAN, no. 150, Moscow, 1975.
10. A. M. Duraevskii, Yu. A. Emel'yanov, S. V. Pashkov, et al., Trudy FIAN, vol. 154,
142, 1984.
11. Kh. Gulfayan, A. M. Duraevskii, M. Z. Zazyan, and A. A. Chilingaryan, Preprint
AN, no. 307, Moscow, 1986.
12. A. A. Chilingaryan, Ph. D. Thesis, ErFI, Erevan, 1984.

Erevan Physics Institute

Lebedev Physics Institute,
Academy of Sciences of the USSR