NONPARAMETRIC METHODS OF DATA ANALYSIS IN COSMIC RAY ASTROPHYSICS

AN APPLIED THEORY OF MONTE CARLO STATISTICAL INFERENCE

Ashot Chilingarian

Yerevan 2004, edited in 2020

A new methodology for solving the inverse problems for the physical inference from multivariate measurements made by large astroparticle physics experiments is described; examples of implementations are presented, physical inferences discussed. From the distributions of secondary electrons, positrons gamma rays, muons, and hadrons we should recover the primary particle type, energy, and for neutral particles, also coming direction. Due to numerous random fluctuations and methodical errors, as well as, uncertainties in the simulations (the cosmic ray energies surpass the collider energies by several orders of magnitude), this problem is extremely difficult to solve. We describe a scheme for solving such problems, that comprises multiple solving of direct problem of cosmic rays (simulations with known parameters of the primary hadron) and implementation of Bayesian and Neural Net models for final statistical inference on the type and energy of each particle, and as well on the "correct" model from a set of alternative explanations.
Thus, the book describes a unified methodology of the data analysis in the framework of Neural Network statistical models and big data concept, using advanced nonparametric techniques in order to be able to address reliably the most difficult and most important problem of high energy astroparticle physics data analysis - event by event analysis of CR interactions, determination of the type and energy of each primary nucleus, initiated cascade process in the atmosphere and detected by surface detectors (WHIPPLE and MAGIC imaging air telescope, MAKET ANI, KASCADE surface arrays).
As well will be presented new approaches for big data analysis from the LHC experiments.
Several methods and techniques were proposed, developed, and implemented in Neural Information technologies to control the learning from examples, to gain the high efficiency and speed of Neural classification and estimation procedures, to estimate the method performance, and to stabilize the obtained results.
Procedures invoking the Cross-Validation, Evolutionary training, Hardware Neural accelerator were implemented, in order to obtain stable, fast, and reliable results on the general nature of the investigated phenomenon, rather than the particular (maybe spurious) solution of the problem overfitted to the data on hand.
The handling of EAS simulated data proves, that the proposed methodology allows determining with ~ 70% efficiency the type of the primary and to estimate its energy with ~ 25% relative error, in this way allowing first time to obtain energetic spectra of three species of Cosmic Ray flux. Also, the world-best results were obtained implementing our methods for background rejection in very high-energy astrophysics with IASTs.
The developed methods are universal and can be used for high-energy physics data analysis as well, as for other real-life applications. All methods are integrated with the ANI (analysis and nonparametric inference) package that will be available to readers of the book (user manual, description of input and output information).

Keywords: Neural Networks, Big data, Machine learning, Nonparametric methods, High Energy Cosmic Ray Physics, Monte Carlo Statistical Inference, Applied Programming, Software-Hardware Combined System Development, Trigger Applications, Multivariate Data Analysis, Learning Algorithms.

CONTENTS

CONTENTS 4

LIST OF FIGURES 9

LIST OF TABLES 14

INTRODUCTION 17

CHAPTER 1. MONTE-CARLO STATISTICAL INFERENCE 19

1.1. NONPARAMETRIC INFERENCE 19

1.2. PARAMETRIC CLASSIFICATION 20

1.3. NONPARAMETRIC CLASSIFICATION 22

1.3.1. BAYESIAN PARADIGM 24

1.3.2. BAYESIAN DECISION RULES 25

1.3.3. NONPARAMETRIC PROBABILITY DENSITY ESTIMATORS 26

1.3.4. BAYES ERROR ESTIMATION 27

CHAPTER 2. UPPER BOUNDARY OF IRON NUCLEI IN PRIMARY COSMIC RAYS AT E > 10^{16} EV 30

CHAPTER 3. STATISTICAL TECHNIQUES IN BACKGROUND REJECTION FOR ATMOSPHERIC CHERENKOV TELESCOPE 35

3.1. ATMOSPHERIC CHERENKOV TELESCOPES OPENING NEW WINDOW TO UNIVERSE 35

3.2. IMAGING TECHNIQUE 36

3.3. BACKGROUND REJECTION STRATEGIES 38

3.4. A PRIORI METHODS OF BACKGROUND REJECTION 42

3.5. ESTIMATES OF MAGIC ACT BACKGROUND REJECTION RATES 47

3.6. CONCLUSIONS 51

CHAPTER 4. THE NEURAL INFORMATION TECHNIQUES 52

4.1. FEED-FORWARD NEURAL NETWORKS 52

4.2. NN LEARNING PROBLEM 54

4.3. SETTING OF THE LEARNING PROBLEM 55

4.3.1. NEURAL ESTIMATION (LEARNING REGRESSION FUNCTION) 55

4.4. TRAINING ALGORITHMS 57

4.5. EVOLUTIONARY ALGORITHMS 59

4.6. COMMITTEE OF NETWORKS 60

4.6.1. ENSEMBLES OF NETWORKS 60

4.6.2. THE COMMITTEE PROCEDURES 61

4.6.3. RESOLVING OF THE MIXTURE OF ANALYTIC MODELS 62

4.7. ESTIMATION OF THE GENERALIZATION ERROR IN NEURAL CLASSIFICATION TO MULTIPLE CATEGORIES 65

4.7.1. NET TRAINING AND GENERALIZATION ERRORS 65

4.7.2. THE CROSS-VALIDATION PROCEDURE FOR FINAL PREDICTION ERROR (FPE) ESTIMATION 67

4.7.3. ESTIMATES OF FPE AS AN NET ARCHITECTURE SELECTION TOOL 69

4.8. COMPARISON OF THE DIFFERENT ALGORITHMS OF NN TRAINING FOR THE PROBLEM OF CLASSIFICATION 70

4.8.1. RANDOM SEARCH LEARNING ALGORITHM 71

4.8.2. DATA AND PROCEDURES 72

4.8.3. COMPARISON RESULTS 73

4.9. INVESTIGATION OF INTERPOLATION POSSIBILITIES BY NEURAL NETWORKS 77

CHAPTER 5. INVESTIGATION OF EAS CHARACTERISTICS 79

5.1. THE KASCADE EXPERIMENT 79

5.2. THE SIMULATION PROCEDURE 79

5.3. VALIDATION OF MODELS 81

5.3.1. COMPARISON OF THE SINGLE EAS VARIABLES 81

5.3.2. CORRELATION ANALYSIS 82

5.3.3. PROBABILITY DISTANCES 83

5.3.4. QGS AND VENUS COMPARISON 84

5.4. CLASSIFICATION OF EXPERIMENTAL SAMPLE 85

5.5. THE HOMOGENEITY OF THE KASCADE DATA 85

5.6. THE KASCADE CLASSIFICATION MATRICES 86

5.6.1. COLORED NUCLEAR MAPS 88

CHAPTER 6. EAS DATA ANALYSIS 90

6.1. ENERGY SPECTRA AND MASS COMPOSITION OF PRIMARY COSMIC RAY FLUX, MEASURED ON SHOWER-BY-SHOWER BASIS 90

6.1.1. COSMIC RAY (CR) SOURCES AND ACCELERATION MECHANISMS 90

6.1.2. PRIMARY ENERGY ESTIMATION 91

6.1.3. PRIMARY MASS DETERMINATION 92

6.2. NEURAL ESTIMATION OF MASS AND ENERGY 94

6.3. THE "PURIFICATION" PROCEDURE 96

6.3.1. EXPERIMENTS WITH PURE (MONONUCLEAR) CR 100

6.4. CONCLUDING REMARKS 103

CHAPTER 7. JOINT ANALYSIS OF THE ANI AND KASCADE EXPERIMENT 104

7.1. NUCLEAR - ELECTRON CASCADE IN THE ATMOSPHERE 104

7.2. ANI-KASCADE EXPERIMENTAL SPECTRA 104

7.3. PROCEDURES FOR INFERENCE OF THE ATTENUATION LENGTH FROM SIZE SPECTRA 107

7.3.1. ATTENUATION OF THE INTENSITY OF FIXED N_e : ABSORPTION LENGTH 107

7.3.2. CONSTANT INTENSITY CUT (CIC) 108

7.3.3. WHICH PRIMARY ENERGIES WE SELECTED BY CIC METHOD? 109

7.3.4. THE RELATION BETWEEN THE ABSORPTION AND ATTENUATION LENGTH 109

7.3.5. ESTIMATE OF THE INELASTIC CROSS SECTION 110

7.4. ESTIMATION OF THE ATTENUATION LENGTH 110

7.5. CONCLUSION 114

CHAPTER 8. CHARACTERISTICS OF THE LIGHT AND HEAVY MASS GROUP COSMIC RAY FLUX OBTAINED BY MAKET-ANI DETECTOR 117

8.1. COSMIC RAYS IN THE VICINITY OF EARTH 117

8.2. COSMIC RAY ACCELERATION IN SUPERNOVAE EXPLOSIONS AND PROPAGATION IN THE INTERSTELLAR MEDIUM 120

8.3. THE MAKET-ANI EXPERIMENT 121

8.4. SELECTION OF EAS PARAMETERS FOR CLASSIFICATION AND ESTIMATION 123

8.5. DATA CLASSIFICATION INTO LIGHT AND HEAVY GROUPS OF NUCLEI, PURIFICATION OF SELECTED GROUPS OF NUCLEI 128

8.6. ESTIMATION OF THE PRIMARY ENERGY OF DIFFERENT GROUPS OF NUCLEI 132

8.7. ENERGY SPECTRA 133

8.8. SEARCH OF POINT SOURCE OF COSMIC RAYS 138

CHAPTER 9. “SPECTROSCOPY” OF THE CR SOURCE 145

9.1. WHAT WE CAN LEARN FROM SOLAR ACCELERATORS 150

9.2. DISCUSSION AND CONCLUSIONS 152

CHAPTER 10. SPACE WEATHER FORECAST 155

10.1. INTRODUCTION **155**

10.2. PARTICLE ACCELERATION AT THE SUN **155**

10.3. THE ARAGATS SPACE ENVIRONMENT CENTER **156**

10.4. CALCULATED ARRIVAL TIMES OF THE RELATIVISTIC SOLAR IONS AT 1 AU AND ENHANCEMENTS REGISTERED BY ASEC MONITORS **159**

10.5. DISCUSSION **161**

10.6. CONCLUSION **163**

ANNEX A. MULTISTART RANDOM SEARCH WITH EARLY STOP AS TOOL FOR SELECTION OF SETS OF DIFFERENTLY EXPRESSED GENES **165**

ANNEX B. NOTATIONS **176**

BIBLIOGRAPHY **177**

LIST OF FIGURES

Figure 1. Monte-Carlo Simulation Procedure 21

Figure 2. The scheme of events attribution to classes: If $Z > Z_0$ an event is attributed to the H class, otherwise to L. 31

Figure 3. Definition of the "Hillas parameters" for hadron and gamma image 37

Figure 4. Comparisons of different image parameter subsets 49

Figure 5. Comparison of the shape and orientation parameter distributions for MAGIC and Whipple ACT 50

Figure 6. Two-dimensional scatter plots of MAGIC and Whipple nontrivial correlations 50

Figure 7. Contributions of different networks to final estimate (3 nets) 61

Figure 8. Contributions of different networks to final estimate (5 nets) 61

Figure 9. Distribution of regression function and input parameters (mixture of models). 63

Figure 10. The regression function estimation by different networks trained on different models: 1 - exponential, 2 - Gaussian, 3 - power 64

Figure 11. Mean and median committee results 64

Figure 12. Training and generalization errors 66

Figure 13. Classification boundaries for over-trained network 66

Figure 14. The training process of the NN with architecture 2:2:1, TS consists of 300 events (30 validation events) 67

Figure 15. The training process of the NN with architecture 2:6:1, TS consists of 600 (above part) and 1500 (bellow part) events (VS - 60(150) events). 68

Figure 16. The dependence of the FPE and training errors on the number of hidden nodes in NN for different sample sizes. 69

Figure 17. The dependence of the FPE and training errors on the number of hidden nodes in NN for different sample sizes (the case of very large training samples). 70

Figure 18. The net training results using random numbers from Cauchy population and uniform random numbers. 74

Figure 19. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (small steps). 74

Figure 20. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (small steps). 74

Figure 21. The dependence of final performance from learning step size. 75

Figure 22. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (medium steps) 76

Figure 23. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (medium steps). 76

Figure 24. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (large steps) 76

Figure 25. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (large steps) 76

Figure 26. The distribution of primary energy in simulated sample. 77

Figure 27. The history of NN training for the estimation task. 77

Figure 28. The NN regression using events with individual weights, estimated energy versus true energy 78
Figure 29. Lay-out of the KASCADE experiment..... 79
Figure 30. Features distribution for proton and iron 81
Figure 31. Proton and iron events distribution in 3 dimensional space of features 82
Figure 32. 3-way map, array information. Green points represent oxygen MC data. 89
Figure 33. The accuracy of the energy estimation. Used observables: N_e , N_μ^{tr} and s 92
Figure 34. The accuracy of the energy estimation. Used observables: N_e and N_μ^{tr} 92
Figure 35. N_e versus E_0 for simulated and experimental proton and iron events (the primary energy is estimated by neural regression method)/ used observables: N_e and N_μ^{tr} 93
Figure 36. N_μ^{tr} versus E_0 for simulated and experimental proton and iron events (the primary energy is estimated by neural regression method)/ used observables: N_e and N_μ^{tr} 93
Figure 37. Shower age s versus E_0 for simulated and experimental events (the primary energy is estimated by neural regression method). Used observables: N_e and N_μ^{tr} 93
Figure 38. Relative abundance of different groups of nuclei as a function of energy. Used observables N_e , N_μ^{tr} and s. 95
Figure 39. Differential energy spectra of three nuclear groups and all particle spectrum. Used observables: N_e , N_μ^{tr} and s… 95
Figure 40. Relative abundance of different groups of nuclei as a function of energy. Used observables: N_e , N_μ^{tr} and s..... 95
Figure 41. Differential energy spectra of three nuclear groups and all particle spectrum. Used observables: N_e , N_μ^{tr} and s..... 96
Figure 42. The cumulative abundances of different groups of nuclei..... 96
Figure 43. Event selection efficiency us purity for proton and iron events (obtained by classification of the control samples). Used observables: N_e and N_μ^{tr} 97
Figure 44. Feature distributions for classified and purified proton and iron samples..... 97
Figure 45. Feature distributions for classified and purified proton and iron samples..... 97
Figure 46 The two-dimensional distribution of the NN input parameters 98
Figure 47 The distribution of age (not used for NN input) parameter for purified samples 99
Figure 48 s vs N_e for removed proton and iron events..... 99
Figure 49 The first interaction height (Z) distribution for purified and removed proton and iron events..... 100
Figure 50 The Energy dependence of the Hadron Energy for the Proton and Iron Primaries 101
Figure 51 The Energy dependence of the Total Energy for the Proton and Iron Primaries..... 101
Figure 52 The Energy dependence of the Number of Hadrons for the Proton and Iron Primaries..... 102
Figure 53 The Energy dependence of the Number of Muons for the Proton and Iron Primaries. 102

Figure 54 Differential size spectra for different zenith angles ranges observed with MAKET ANI array, compared with spectra reported by the KASCADE (Glasstetter, 1998) and the EAS TOP (Aglietta et al. 1999) collaborations. 105

Figure 55 N_e cascade in the observed range of the atmospheric slant depth..... 106

Figure 56. Integral size spectra for different zenith angles ranges observed with MAKET ANI array, compared with spectra reported by the KASCADE (Glasstetter, 1998): illustration of the procedures for absorption and attenuation length estimates. 107

Figure 57. Indices of EAS size differential and integral spectra in comparison. 107

Figure 58.(a-d) Fluctuations of energy for different integral intensities and zenith angles 108

Figure 59. Attenuation length obtained by joint analysis of the MAKET ANI and KASCADE data….. 111

Figure 60 Attenuation Length dependence on Spectra Intensity (Primary Energy) 113

Figure 61 Attenuation Length dependence on the Shower Size N_e 113

Figure 62 MAKET-ANI detector setup 122

Figure 63 Shower Age distribution for simulated light and heavy primaries with (right) and without (left) incorporating of detector response. 127

Figure 64. Scatter plot Shower Age versus Shower Size for simulated light and heavy primaries with (right) and without (left) incorporating of detector response. 127

Figure 65. Output of the Neural Network (NN) trained to distinguish light and heavy nuclei..... 131

Figure 66. Purity – Efficiency plots obtained by shifting the NN decision boundaries..... 132

Figure 67. The relative errors of energy estimates fro 10 energy intervals of light and heavy groups. 133

Figure 68. Summary of the all particle spectra from 18 experiments 134

Figure 69. Energy spectra of light and heavy nuclei obtained by neural classification and energy estimation. EAS characteristics used : Shower Size and Shape (Age). 135

Figure 70. Energy spectra of light and heavy nuclei obtained by neural classification and energy estimation. The same as in Figure 69 but obtained with purified light and heavy data samples. Purification intervals [0.,0.3) and (0.7,1.] 136

Figure 71. Relative abundances of light and heavy nuclear groups (purified EAS). 136

Figure 72. Mean logarithmic mass dependence on primary energy (world data survey) (Allesandro, 2001); (Fowler et al., 2001); (Roth et al., 2001), (Ulrich et al., 2001); (Vardanyan et al., 1999)..... 137

Figure 73. Reconstructed energy spectra from size spectra in different energy intervals 138

Figure 74. Light group spectra (world survey) (Arqueros, et al., 2000), (Vardanyan et al., 1999) 138

Figure 75. Distribution of number of events in each of 120 RA bins for declination band of 12.5° – 15.5°..... 139

Figure 76. Signal significance test with full equatorial coverage with 2400, 3° x 3° bins; Ne > 10^6. 140

Figure 77 Random scanning star sky, directions to RA bins demonstrating significant excess, σ > 5..... 143

Figure 78 Random scanning star sky, directions to RA bins demonstrating significant excess, σ > 5..... 144

Figure 79 Random scanning Monogem Ring, directions to RA bins demonstrating significant excess, σ > 4 144

Figure 80 Random scanning Monogem Ring, directions to RA bins demonstrating significant excess, σ > 5. 145

Figure 81 CR Source within Monogem Ring.(MR as seen in the ROSAT all-sky survey in the 0.25- 0.75KeV) 145

Figure 82 Monogem Ring Coverage, center of signal bin 750α+14δ, size 3° x 3°. 145

Figure 83 The 2-way neural classification of EAS registered from declinations 12.5 – 15.5º 146

Figure 84 The 2-way neural classification of EAS registered from SS direction: 750α+14δ 146

Figure 85 Energy Spectra of light and heavy mass groups coming from declinations 12.5 – 15.5º comparing with SS direction: 750α+14δ, shower size Ne > 10^5 147

Figure 86 Energy Spectra of light and heavy mass groups coming from declinations 12.5 – 15.5º comparing with SS direction: 750α+14δ, shower size Ne > 10^6 147

Figure 87 Distribution of the azimuthally accuracies of Maket-Ani detector, declination band 12.5 – 15.5º 149

Figure 88 Distribution of the zenith accuracies of Maket-Ani detector, declination band 12.5 – 15.5º 149

Figure 89 Distribution of the azimuthally accuracies of Maket-Ani detector in signal bin, center of signal bin 750α+14δ, size 3° x 3°. 150

Figure 90 Distribution of the zenith accuracies of Maket-Ani detector in signal bin, center of signal bin 750α+14δ, size 3° x 3°. declination band 12.5 – 15.5º. ..... .150

Figure 91 The energy spectra of protons and Fe ions registered by space-born detectors during SEP event of 23-rd solar cycle. 152

Figure 92 Nor Amberd Multidirectional Muon Monitor installed around the 8-NM-64 158

Figure 93 Linear correlation of ion arrival times to 1 AU versus their inverse speed. The filled symbols correspond to the expected arrival times of relativistic ions. The proton event is shifted down by 3 hours for better visualization of the regression lines. 160

Figure 94 The 15 April 2001 GLE. Estimated arrival times of relativistic ions superimposed on the ANM 1 minute time series. Horizontal dashed lines indicate 1, 2 and 3% enhancements relative to mean count rate. 160

Figure 95 Comparison of gene selection with early and late stopping of the multi-start random search algorithm. First row: histogram of the value of the 'last best iteration' in the $N_{cycle} i_s$ searches. Second row: histogram of the estimated Mahalanobis distances for the $N_{cycle} i_s$ selected sets. Third row: frequency of inclusion of the differentially expressed genes (1-20) in one of the selected sets. One simulated data set. 172

Figure 96 ROC curves for various values of N_{ner} controlling the stopping time. Based on 10 simulated data sets; the error bars show the standard errors of the point estimates. 173

Figure 97 Comparison of the search procedure used for same and different tissues. First row: histogram of the value of the 'last best iteration' in the $N_{cycle} i_s$ searches. Second row: Histogram of the estimated Mahalanobis distances for the $N_{cycle} i_s$ suboptimal sets. HCT - HT data 173

Figure 98 Comparison of the frequency of inclusion in the selected locally optimal set in the searchprocedure used for same tissue with (gray) or without (white) the six control *Drosophila* genes on the two channels and for different tissues (black). HCT - HT data 175

LIST OF TABLES

Table 1. The comparison of the mean values of parameters calculated from different simulations with corresponding experimental values.....32

Table 2. The fractions of families produced by Fe nuclei for the two chemical composition N (normal) and F (Fe enriched) for the various family parameters.....33

Table 3. The upper limits of the Fe intensity and of the fraction of Fe in primary cosmic rays for the two chemical compositions N (normal) and F (Fe enriched). 33

Table 4. Comparison of the different background rejection methods 41

Table 5. Root mean square errors of 3 different model networks and committee 64

Table 6. *Performances of different learning algorithms* 73

Table 7. EAS features detected by KASCADE experiment 80

Table 8. P-values of statistical tests for proton and iron for different models: t Student,Nμ* D - Kolmogorov-Smirnov, U - Mann-Whitney 83

Table 9. Correlation matrix for QGS data.....83

Table 10. Correlation matrix for Venus data.....84

Table 11. The best feature subsets according to the Bhattacharya distance.....84

Table 12. Comparison of exp. data with VENUS and QGS models.....85

Table 13. One dimensional tests for models and experiment.....85

Table 14. Exp. data homogeneity test, features used: N_{μ}^{-CD}, E_h^{sum} 85

Table 15. Exp. data homogeneity test, features used: N_e, N_{μ}^{tr}, N_{μ}^{CD}, E_h^{sum} 85

Table 16. Exp. data homogeneity test, features used: N_e and N_{μ}^{tr}.....86

Table 17. One dimensional tests of exp. data: *t* - Student, *D* - Kolmogorov-Smirnov, *U* - Mann- Whitnay 86

Table 18. Array data, features used N_e, N_{μ}^{tr}.....86

Table 19. KASCADE data, features used 87

Table 20. 3-way classification by N_e, N_{μ}^{tr}.....87

Table 21. 3-way classification by $N_e, N_{\mu}^{tr}, N_{\mu}^{CD}, E_h^{sum}$ 87

Table 22. 2-way classification by N_e, N_{μ}^{tr}.....87

Table 23. 2-way classification by $N_e, N_{\mu}^{tr}, N_{\mu}^{CD}, E_h^{sum}$ 87

Table 24. Separability index for KASCADE.....88

Table 25. Correlations between primary mass and energy with EAS observables.....93

Table 26. Purity of classified events. Used observables: N_e, N_{μ}^{tr} and s 94

Table 27. Purity of classified events.Used observables: N_e and N_{μ}^{tr} 94

Table 28. The mean values of different parameters of purified and removed proton and iron samples 98

Table 29. The probability values of different tests for initial and purified proton and iron samples (t-Student, D- Kolmogorov-Smirnov, U- Mann- Whitnay) 99

Table 30. The classification results of pure proton subsample into initial and pure proton classes (features used: N_e, N_{μ} and age) 100
Table 31. The Summary of the Measured Differential Flux and Knee Region Parameters, Zenith Angles Range of (0° — 45°), Altitude 3200m (intensities are given in $m^{-2} s^{-1} sr^{-1}$ unit) 105
Table 32. Attenuation lengths for the data from the MAKET ANI and KASCADE installations estimated by the CIC method from differential and integral size spectra 110
Table 33. Attenuation lengths for the data from the MAKET ANI and KASCADE installations estimated by the recalculation from the absorption length for differential and integral size spectra..... 111
Table 34. EAS parameters for p and Fe primaries..... 124
Table 35. One-dimensional tests of p-Fe samples (t- Student, ... 124
Table 36. One-dimensional tests of p-He samples (t- Student,..... 124
Table 37. One-dimensional tests of Si-Fe samples (t- Student,..... 125
Table 38. Correlation matrix for proton sample without detector response 125
Table 39. Correlation matrix for iron sample without detector response 126
Table 40. Correlation matrix for proton sample with detector response..... 126
Table 41. Correlation matrix for iron sample with detector response..... 126
Table 42. Neural classification into two classes using H+He and Si+Fe events without and with detector response 128
Table 43. Bayesian classification into two classes using H+He and Si+Fe events without and with detector response 128
Table 44. Efficiency of the neural classification of EAS initiated by different primaries into two mass groups 128
Table 45. Purity of the classification of different nuclei in light and heavy groups. 129
Table 46. Efficiency of the neural classification of EAS initiated by different primaries into two mass groups (purification intervals [0.,0.3) and (0.7,1.]). 130
Table 47. Purity of the classification of different nuclei in light and heavy mass groups (purification [0.,0.3) and (0.7,1.]). 130
Table 48. Dependence of the “signal” on shower Size cut 141
Table 49. CR source Localization around the center of the “signal” bin..... 142
Table 50. Parameters of the ASEC Monitors at Mt. Aragats in Armenia..... 159
Table 51. Characteristics of the GLE detected by ACE/SIS, GOES-8 and ASEC detectors..... 161

INTRODUCTION

Problems of Data Analysis in Cosmic Ray Physics Experiments.

The observational evidence for the existence of very high energy Cosmic Ray (CR) point sources has reanimated the substantial interest in CR physics and initiated construction of very large experimental complexes. Arrays of particle detectors covering a large area are measuring different characteristics of numerous secondary products of the Primary CR (PCR) interactions with the atmosphere. Only a simultaneous measurement of a large number of independent characteristics in each individual Extensive Air Shower (EAS) can yield reliable information to reconstruct the primary particle origin and its energetic characteristics as well as the peculiarities of strong interaction at the top of atmosphere.

The ambiguity of interpretation of the results of experiments with cosmic rays is connected with both significant gaps in our knowledge of the parameters of hadron-nuclear interactions at super accelerator energies and indefiniteness of the primary cosmic ray composition, as well as with strong fluctuations of all shower characteristics. The extra difficulties are due to indirect experiments and hence, due to the use of Monte – Carlo simulations of development and detection of different components of nuclear electromagnetic cascades.

To make the conclusions about the investigated physical phenomenon more reliable and significant, it is necessary to develop a unified theory of statistical inference, based on nonparametric models, in which various statistical approaches (density estimation, Bayesian decision making, error rate estimation, feature extraction, sample control during handling, neural net models, etc.) would be used.

The most important part of the present approach is the quantitative comparison of multivariate distributions and use of a nonparametric technique to estimate the probability density in the multidimensional feature space. As compared to the earlier used methods of inverse problem solution, in this work the object of analysis is each particular event (a point in the multivariate space of measured 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 (statistical decision problem).

This approach was first used to estimate the upper limit of the iron nuclei fraction according to the γ – family characteristics, registered by PAMIR experiment (Chilingarian et al., 1987). Our results confirm the normal composition hypothesis and reject the hypothesis of dominance of iron nuclei in PCR at $E > 10$ PeV.

The handling of EAS simulation data proves that the methodology allows us:

- to determine with ~ 70% efficiency the type of the primary and to estimate its energy with an accuracy of ~ 25% (Aglietta et al. 1999);
- to investigate the possibility of selecting mononuclear CR beams from experimental data using characteristics of the electron – photon and the muon components of EAS, and further comparison of such beams with alternative simulations. These results give us hope that the study of the parameters of p — A and A — A interactions in the energy range 1-10 PeV is possible.

Statistical Inference in Cosmic Ray Physics.

The most difficult and most important part of high energy physics data analysis is the comparison of competitive hypotheses and decision making on the nature of the investigated physical phenomenon.

In the cosmic ray physics the main technique of statistical inference connected with the problem of determination of initial physical parameters (such as mass composition and energy spectrum of PCR, strong interaction parameters, flux of very high energy 7-quanta from point sources, etc.), - is the indirect problem solution with detailed simulation of the cosmic ray traversal through the atmosphere and the experimental installation with a following comparison of the multivariate simulation and experimental data. Actually, an algorithm is constructed, which describes EAS development and registration of its different components on the observation level, which is based on a certain model of the process investigated, i.e. the set of the parameters that characterize the PCR flux and interaction of hadrons and nuclei and γ-quanta with the air nuclei.

By simulations with different models and comparing the experimental and model data, a class of models is selected, which describe the experimental data satisfactorily.

For almost all problems of inference, the crucial question is: whether the fitted probability family is in fact consistent with data? Usually parametric models are chosen for their statistical tractability, rather than for their appropriateness to the real process being studied. Of course, any statistical inference is conditioned on the model used, and, if the model is oversimplified, so that essential details are ever omitted, or improperly defined, at best only qualitative conclusions may be done. Now, in cosmic ray and accelerator physics very sophisticated models are used, completely mimicked a stochastic mechanism whereby data are generated. Such models are defined on a more fundamental level than parametric models, and provide us with a wide range of outcomes from identical input variable sets, so-called, "labeled", or "training" samples (TS). These sets of events with known membership represent the general - nonparametric mode of a priori information. Though simulation in data analysis in high energy physics is widely used, we can aware of a very few systematic investigations of theoretical aspects about how data may be compared with their simulated counterpart.

What we need is a well defined technique, that one can call Monte-Carlo Statistical Inference (MCSI). The presented approach considers the classification and hypothesis testing problems in the framework of Bayesian paradigm and the main steps of the unified data analysis methodology areas following.

CHAPTER 1.
MONTE-CARLO STATISTICAL INFERENCE

1.1 NONPARAMETRIC INFERENCE

The scientific method is characterized by data classification, the study of their interrelations and relations to past experience, accumulated in various theories and hypotheses. Usually, it is impossible either to prove or to refuse hypotheses by deductive method. The challenge is to draw sensible conclusions from noisy, discrepant information.

The main aspect of statistics is the collection and interpretation of data, the interpretative aspect being the one that is now regarded as the essence of the subject (Lederman, 1984). The fundamental idea of statistics is that useful information can be obtained from individual small bits of data. An inductive method leads to empirical statements that may be connected with theoretical ones by means of rational inductive conclusion rules (Hajek et al., 1979).

The most natural and most general framework in which to formulate solutions to the physical inference in CR physics is the statistical one, which recognizes the probability nature both of the physical processes of propagation of cosmic radiation through the atmosphere and the detectors, and - of the form in which data analysis results should be expressed.

However, it is very important to provide the scientist with objective criterion by which to judge the claims of hypotheses (models) under investigation (problem solving strategy). By model we mean a complete probability statement of what currently supposed to be known a priori about the mode of generation of data and of uncertainty about the parameters (Box, 1984).

If this statement consists in the existence of an analytic distribution family, (like Poisson or Gaussian), appropriate to the problem in hand, we have prescribed parametric model. For such parametric models a well-known concept of statistical inference consists in obtaining estimates of its parameters and verifying the validity of a chosen family (Eadie et al., 1971). In the CR physics the main technique of physical analysis is the Monte-Carlo Statistical Inference (MCSI), the detailed simulation of the CR traversal through the atmosphere and the experimental installation with a following comparison of the multivariate simulation and experimental data. Actually, an algorithm is constructed, which describes the EAS development and registration of its different components on the observation level, which is based on a certain family of models of the physical processes investigated.

MCSI is a process. It takes requirements specification (basic physics, experimental techniques, data analysis techniques), it generates families of models to meet this spec- ifications and it synthesizes a priori knowledge and experimental results to create new knowledge.

Complexity of the MCSI is determined by its multifunctionality, adaptability and flexibility Herein lies MCSI flexibility. It allows the input vectors to be formed directly from initial measurements or from reconstructed EAS parameters.

MCSI incorporates and uses such advanced nonparametric methods as Nonparametric Boundary Analysis, Adaptive Multivariate Density Estimation, Fractal Dimensionality Analysis, Artificial Neural Networks, etc.

The detailed proposals to use MCSI in Astroparticle Physics experiments were performed by A.Chilingarian and H.Zazyan in early 90-s (Chilingarian et al., 1991) for planned ANI experiment (ANI Collaboration, 1992). The main idea of the project was to use advanced multivariate statistical methods (Chilingarian, 1989) to perform event-by-event (or shower-by-

shower) analysis of EAS data and for the first time present CR physics experimental data with precision and reliability comparable with accelerator physics experiments (Chilingarian 1991A).

1.2 PARAMETRIC CLASSIFICATION

The classification problem in parametric case (Newman-Pearson test) is traditionally described in terms of null and alternative hypothesis, critical and acceptance regions and level of significance (Zacks, 1977). The "best" critical region (the region of rejection of the null hypothesis) is constructed by means of the Likelihood Ratio (LR):

$$LR(v) = \frac{p(v/\psi_1)}{p(v/\psi_2)} \qquad (1.1)$$

each of two classes is defined by values of ψ_i - the parameter of a prechosen analytic probability density function, v is a multivariate observation vector (point in multidimensional feature space) $p(v/\psi_1)$, (v/ψ_2) - are conditional probability density functions describing distinct, mutually exclusive (non overlapping) and full $p(v/\psi_1)+ p(v/\psi_2)=1$ statements (null and alternate hypothesis).

The threshold value reflects the costs of consequences of statistical decision. Usually one selects this value to keep on some constant minimal level error for one class, while maintaining to minimize the error of the other class.

For the L class case the $p(\psi)$ - will be chosen as a "true" class

$$\psi = argmax_{\psi_i} p(v/\psi_i), \quad i = l, \ldots L. \qquad (1.2)$$

If ψ takes infinite number of values from some metric space Ψ then we deal with an estimation problem and the Maximal Likelihood Estimate (MLE) is asymptotically unbiased and effective

$$\psi_{MLE} = \text{argmax}_{\psi} \sum_{i=1}^{M} \ln f(v_i/\psi), \psi \subset \Psi. \qquad (1.3)$$

where $\{v_i\}, i = 1, M$ are the experimental events. The parametric estimation uses whole experimental sample set, instead of only one event in the classification problem, with the benefit of solving regression problem (parameter estimation) for all possible experimental situations.

The analytical function $f(v/\psi_{MLE}) \equiv f(v)$ can be used for energy estimation, of course if the shape of particular functional family $f(\bullet)$ is known.

Although the results of analysis using parametric statistics usually are easy to present and understandable, it is very important to remember that any inferential conclusion based on parametric technique is not exactly valid unless every assumption is satisfied.

If these assumptions cannot be substantiated, or are discarded, or are not even known to the investigator, then the inference may be less reliable than a judicious opinion, or even arbitrary guess (Edwards et al, 1984).

The parametric methods superimpose very restrictive assumptions on the nature of the population from which the sample is drawn. For example, the assumption of a normal distribution implies a continuous, symmetric, bell shaped distribution with infinite domain and a

specific mathematical function. And statistical inference is exact for these sampling distributions only and may not even be close to the obtained one, if the population assumption comes to be incorrect.

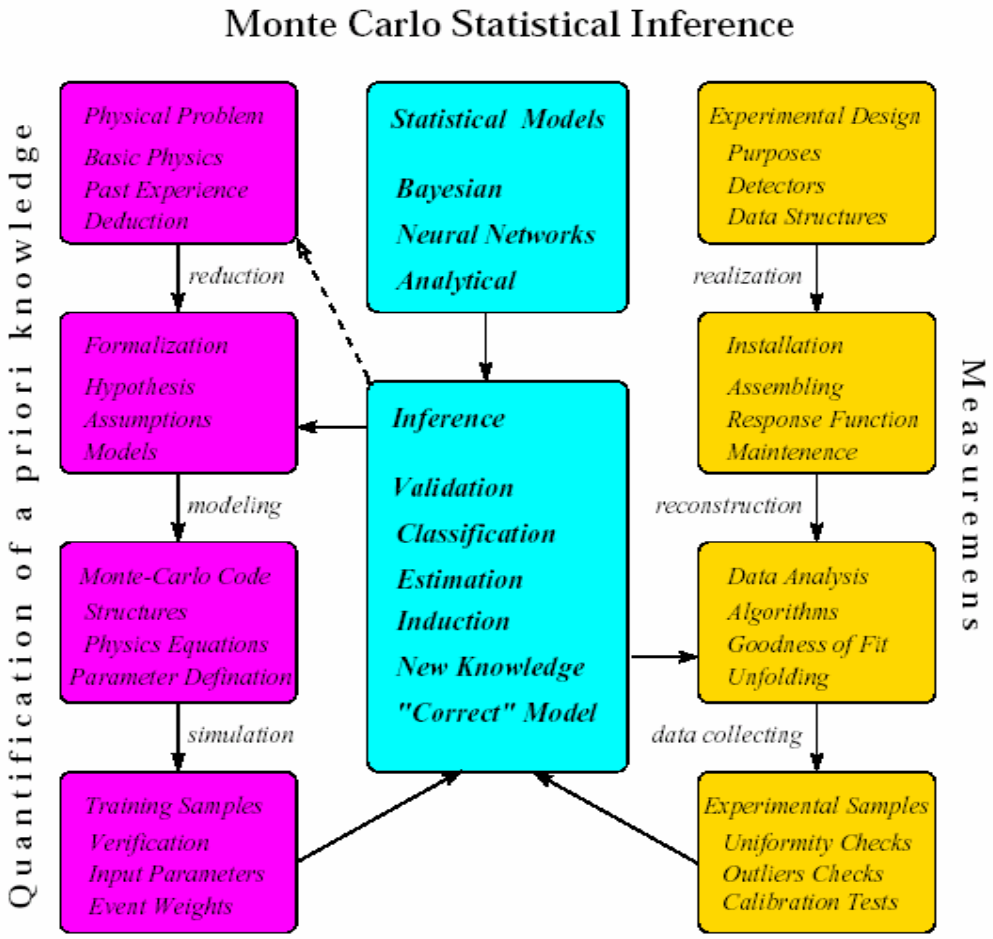


Figure 1. Monte-Carlo Simulation Procedure

1.3 NONPARAMETRIC CLASSIFICATION

Usually, for experimental physics data analysis, the Likelihood Function cannot be written explicitly, and we deal with implicit, nonparametric models, for which no parametric form of underlying distribution is known, or can be assumed.

Nonparametric methods use much less stringent assumptions about population than those made in parametric statistics. Usually the underlying population distribution is assumed to be continuous only. Of course this assumption is rather mild comparing with the very specific assumptions made in parametric case.

Let us consider the stochastic mechanism (A, P) which generates the observations v in a multivariate feature space - v, **v** is a d-dimensional vector of EAS parameters measured experimentally. We assume that observations are random and can be described by some conditional probability density function depending on the primary particle type. The feature space v covers possible acceptable values of EAS parameters including cuts on age and Ne parameter, etc. . .

The basic states space A consists of alternative models or classes (the alternative strong interaction models, or - different primary nuclei). The appropriate statistical model to describe

this situation is the probability mixture model:

$$p(v) = \sum_{k=1}^{L} P_k p(v/A_k) \tag{1.4}$$

And the main problem in EAS physics is to determine the proportions (frequencies) P_k of events in each category A_k.

We don't know also the full statistical description (conditional probability density functions $p(v/A_k)$of how nature produces EAS from incident particles, nor the possibility to use particle beams outside the atmosphere to calibrate the installations, that is why, to determine the mutual probability measure on the direct product of A and v spaces the total Monte-Carlo simulation of the EAS development in the atmosphere and in detectors is performed, including experimental data registration and handling for alternative primary particles and possible strong interaction models in a wide energy range.

The problem is how to assign the probability measure in the primary particle parameters space I (L-dimensional metric space). Usually the following parameters are used as inputs for a Monte-Carlo simulation program:

- the primary type;
- the primary energy;
- the angle of incident;
- the strong interaction mode (one of the several strong interaction models) .

As will be shown further, the initial probability measure on I will be modified by experimental information in accordance with Bayesian paradigm. Experimental evidence will change possibly incorrect assigned initial probability measure, therefore, we can take uniform measure for most of simulation parameters. Of course, we have to outline the physical restriction and define the bounded subspace of I, from which we will take randomly uniform (or according to some established already in previous experiments distribution function) $(t_i, i = 1, M)$, where M is the number of simulation trials) values.

For the classification task among $M(t_i)$ values there will be many same values-class "labels". The primary particle classes will be restricted by 5 groups, including all primaries from proton to iron. Therefore, some of one-dimensional projections of the I space will be discreet.

The set of corresponding d-dimensional $u_i, i = I, M$ vectors obtained from simulations is an

analog of the experimentally measured values of $v_i, i = 1, M_{exp}$ where M_{exp} is the number of detected events. But, as opposed to experimental data, it is exactly known which primary particle was used in simulations.

These labeled events include a priori information about dynamics of the EAS development and registration with inherent fluctuations. All statistical variability of events belonging to the definite class is given in a nonparametric form, in form of simulation trials.

The sequence (u_i, t_j), where $i = 1, M_j$, $j = 1, L$, t — is the class index, is generated by a detailed Monte-Carlo simulation program and consists of L classes each containing M_j

simulation trials.

This "controlled" stochastic mechanism we denote by $(A, \tilde{P})$ and name training sample (TS). The training sample is the basis of all statistical procedures in applied Bayesian and neural approaches. Usually we denote a TS by A_k or explicitly by the primary group - *p, O, ... ,Fe*.

The corresponding distribution mixture model takes the form:

$$\hat{p}(v) = \sum_{k=1}^{L} \hat{P}_k \hat{p}(v/A_k) \tag{1.5}$$

Of course this substitution of unknown conditional densities $p(v/A_k)$ by their "simulation" analog $\hat{p}(v/A_k)$, estimated by means of training sample $\{u_i, t_j\}$, is only valid if the used model is adequate. And validation of the model remains the most crucial and yet unsolved problem for EAS data analysis.

Of course, for reliable estimation of conditional densities we'll need significant amount of training trials to cover all intrinsic variations of measurable EAS parameters and completely represent all categories (primary nuclei).

Since both physical processes of particle production and those of registration are stochastic, only by careful measurement of probabilities we can gain an understanding of the EAS phenomena. We can't expect simple solutions, as multidimensional distributions of EAS parameters overlap significantly and any decision on primary particle type and its energy will contain uncertainty.

The only thing we can require when classifying a distribution mixture 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.

1.3.1 BAYESIAN PARADIGM

The Bayesian approach of the statistical inference is a modification of the opinions of consistent experts (a priori knowledge) in the light of new evidence and the Bayes theorem specifies how such modification should be made.

Moreover, as we believe, Bayesian a posteriori measures are only trustworthy and sensible measures of how the uncertainty about the phenomenon under investigation should be modified after new experimental data are achieved (Berger, 1984).

The Bayesian approach formalizes the account of all the losses connected with probable misclassification and utilizes all the differences of alternative classes (Lindley, 1978). The decision problem in the Bayesian approach is simply described in terms of the following probability measures defined on the metric spaces:

- The space of possible states of nature $A \equiv (p, \alpha, O, N, Fe)$ - groups of primary nuclei;
- The space of possible statistical decisions $\tilde{A} \equiv (\tilde{p}, \tilde{\alpha}, \tilde{O}, \tilde{N}, \tilde{F}e)$where $\tilde{p}, \ldots \tilde{F}e$ are the decisions that the examined event is caused by a primary proton, or. . . , iron nuclei;
- Cost (losse) measure $c_{A\tilde{A}}$, or $c_{A_iA_j}$, or in simple notion C_{ij}. This measure is defined on

the direct product of nature states and decision spaces ($A \otimes \tilde{A}$). All losses, connected with definite statistical decision A) are equal to

$$c_{ij} = \sum_{j=1}^{L} c_{ij,}\, \mathrm{i,j} = 1, \mathrm{L} \tag{1.6}$$

At correct classification of primary particles into "proton" and "iron" classes the losses are equal to zero:

$$c_{Fe\tilde{F}e} = c_{p\tilde{p}} = 0 \tag{1.7}$$

or for problem of background rejection in TeV γ-ray astronomy

$$c_{\gamma\tilde{\gamma}} = c_{h\tilde{h}} = 0 \tag{1.8}$$

If we misclassify the signal event, we decrease the efficiency of γ -event registration. If we attribute hadronic images to γ -ray ones, we increase the background contamination. As we expect a significant excess of background against signal, we are interested in a strong background rejection. So, it is therefore reasonable to take the non symmetric loss function for this case

$$c_{\gamma\tilde{h}} = 0.9, c_{h\tilde{\gamma}} = 0.1. \tag{1.9}$$

For elemental composition studies one can take uniform a priori losses function

$$C_p = C_\alpha = C_O = C_N = C_{Fe} = 0.2 \tag{1.10}$$

- Event (measurement, feature) space V - a set of measurable characteristics of EAS, Cherenkov image parameters etc.
- The prior measure $P_A \equiv (P_p , P_{Fe} ...)$.
- Conditional densities (Likelihood functions):

$$\{\hat{p}(v/p), \{\hat{p}(v/\alpha), \{\hat{p}(v/O), \ldots\}. \tag{1.11}$$

These density functions are estimated by means of training samples obtained in simulation trials with different primaries.

Multivariate probability density estimation is a fundamental problem in data analysis, pattern recognition and artificial intelligence. The estimation of the conditional density on the basis of the collection of simulations is also a key problem in high energy CR physics.

1.3.2 BAYESIAN DECISION RULES

The Nonparametric Bayesian decision rule takes the form

$$\tilde{A} = \eta(v, A, \tilde{P}) = \mathrm{argmax}_i\{C_i \hat{p}(A_i/v)\}, \mathrm{i} = 1, \ldots, \mathrm{L}. \tag{1.12}$$

where Ci is the losses connected with $\tilde{A}$ decision, $\hat{p}(A_i/v)$ are the nonparametric estimates of the a posteriori density, connected with conditional ones by Bayes theorem:

$$\tilde{p}(A_i/v) = \frac{\hat{P}_i \hat{p}(v/A_i)}{\hat{p}(v)}. \qquad (1.13)$$

And finally, substituting the a posteriori densities by the conditional ones we get the Bayesian decision rule in the form

$$\tilde{A} = \text{argmax}_i\{C_i P_i \hat{p}(v/A_i)\}, \text{i} = 1, \ldots, \text{L}. \qquad (1.14)$$

As one can easily see from above formulae, the Bayesian statistical decision is dependent on multiplicator C_i P_i therefore we can't separate the influence of losses (cost) measure and prior measure on decision made. Changes in losses can be compensated by changes in prior to keep constant the Bayesian decisions.

The robust Bayesian inference claims that after considering repeated evidence, the initial used prior distribution can't influence the *a posteriori* distribution heavily (Berger, 1984).

So, the choice of prior distribution isn't of critical importance for fraction estimation, because of the very big volumes of experimental data overwhelming the initial prior knowledge.

For the investigation of the influence of the chosen value of the a priori losses on the classification results, the statistical decision are made simultaneously for different alternative variants of a priori losses. Examining the, so called, "influence curves" obtained with different losses one can select the preferable regime of estimator operation. For example, it is possible to select the desired ratio of background rejection and signal detection efficiency.

A provision is required to avoid the statistical decision if all classes are very far from experimental events (outliers problem). If

$$\hat{p}(v/A_i) < ST for all i = 1, \ldots L, \qquad (1.15)$$

then the decision will not be made. *ST* is so called, Strangeness criteria, usually set to very small number; conditional densities are estimated by the $TS(A, \tilde{P})$ using one of many nonparametric methods available, *L* is the number of classes.

The Nonparametric Likelihood Ratio for classes A_1 , A_2 and experimental event **v** can be represented as

$$LR(v) = \frac{\hat{p}(v/A_1)}{\hat{p}(v/A_2)} \qquad (1.16)$$

Usually for comparison purposes we'll use the sampling mean of Log Likelihood ratio. The nonparametric Log-likelihood function for *k -th* class has the form:

$$L_k = \sum_{i=1}^{M} \ln \hat{p}(v_i/A_k), \qquad \text{k=1,L,} \qquad (1.17)$$

1.3.3 NONPARAMETRIC PROBABILITY DENSITY ESTIMATORS

To estimate conditional densities, we used Parzen and KNN methods (Devroye et al., 1985), (Rosenblatt, 1957), (Parzen, 1962), (Fix et al., 1951), (Lofsgaarden et al, 1966), (Mahalonobis, 1936), (Tapia et al., 1978), (Fukunaga et al., 1987), (Rabiner et al., 1974) with automatic parameter (kernel width - for Parzen estimate, and number of neighbor - for KNN estimate) adaptation (Chilingarian et al., 1984).

Several probability density values corresponding to different values of parameters are calculated simultaneously. Then the obtain sequence is ordered and the median of this sequence is chosen as final estimate (so called L-estimate). Depending on the intrinsic probability density in the vicinity of point v, where the density is estimated, due to stabilizing properties of the median, each time the best estimate will be chosen [38].

The Parzen kernel probability density is estimated by:

$$\hat{p}(v/A_i) = \frac{|\Sigma_i|^{-0.5}}{(2\pi)^{d/2}h^d}\sum_{j=1}^{M_i} e^{-r_j^2/2h^2}\omega_{j,}$$

$$i = 1,\ldots,L, \sum_{j=1}^{M_i}\omega_j = 1 \qquad (1.18)$$

where *d* is the feature space dimensionality, M_i is the number of events in the *i*— th TS, r_j is the distance from experimental event **v** to the *j*— th event of the TS in the Mahalanobis metric:

$$r_j^2 = (v-u_j)^T\sum\nolimits_i^{-1}(v-u_j) \qquad (1.19)$$

where $\sum_i$ is the sampling covariance matrix of the class to which $\mathbf{u}_j$ belongs, ω_j are the event weights, *h* is the kernel width (parameter controlling the degree of the "smoothness" of an estimate) .

The K nearest neighbors estimate (equal weights only are accepted) takes the form

$$\hat{p}(v/A_i) = \frac{k-1}{M_i V_k(v)} \qquad (1.20)$$

where Vk(v) is the volume of a (d-dimensional hypersphere containing the k nearest neighbors to the experimental event v:

$$V_k(v) = V_d|\Sigma_i|^{1/2}r_k^d, V_d = \frac{\pi^{d/2}}{\Gamma(\mathrm{d}/2+1)}, \qquad i = 1, L, \qquad (1.21)$$

where r_k is the distance to the *k*-th nearest neighbor of *v*, *Ã(.)* is the Gamma function. $|\sum_i|$ is the determinant of the covariance matrix of the class to which the *K*-th neighbor belongs.

1.3.4 BAYES ERROR ESTIMATION

The classification methods, like all the other statistical methods, include a procedure quality test as a necessary element. This stage is also necessary for the determination of the frequencies of the probability mixture (1.5).

The most natural measure for quality test is the error probability which depends on both the degree of overlapping of alternative multivariate distributions and the decision rule being used (Bayes decision rule provides minimal error as compared to any other decision rule using the same features):

$$R^B = E\{\theta[\eta(v, A, P)]\} = \int \theta p(v) dv \tag{1.22}$$

Where

$$\theta[\eta(v, A, P)] \begin{cases} 0, \text{for correct classification} \\ 1, \text{other wise} \end{cases} \tag{1.23}$$

where $\eta(v, A, P)$ is the decision defined by (1.12).

The mathematical expectation is taken over the whole *d*—dimensional feature space V. In other words the Bayes error is a measure of the overlapping of alternative distributions in the feature space V, e.g. the expected proportion of the "incorrect" classification. Since we do not know to which class experimental vectors belong, we obtain an estimate of R^B via the TS:

$$\hat{R}^B = E\{\frac{1}{M_{TS}} \sum_{i=1}^{M_{TS}} \theta[\eta(u_i, A, \tilde{P})]\} \tag{1.24}$$

i.e. we classify the $\{u_i\}$, i = 1, M_{TS} and check the correctness of the classification over the index of the class t_j , $j = 1,L$. The expectation is taken over all possible samples of volume M_{TS}.

However, as numerous investigations have shown (e.g. [94]), this estimate is systematically biased and hence, a one-leave-out-for-a-time estimate is preferable

$$\hat{R}^e = \frac{1}{M_{TS}} \sum_{i=1}^{M_{TS}} \theta\{\eta(u_i, A, \tilde{P}_{(i)})\} \tag{1.25}$$

where $\left(A, \tilde{P}_{(i)}\right)$ is a TS with a removed i-th element, which is classified and then "returned" to the sample. This estimate is unbiased and has an essentially smaller m.s. deviation compared with other estimators [89]. The advantage of R^e is especially notable when the feature space has a high dimensionality.

Note, that we have the possibility to estimate the error probability of various types by classifying various TS classes - $\{\mathbf{u}_i, t_j\}$, j = 1,L.

By R_{ij}^{e} (or simply R_{ij}) we denote the probability of classifying the *i*-th class events as belonging to the *j*—th class (misclassification). By *RH* the "true" classification probability will be denoted. For EAS classification according to 5 primary groups, each element of the "classification matrix" have to be determined, using the Bayes risk estimate (1.25):

$$\begin{pmatrix} R_{p\to p} & R_{p\to\alpha} & R_{p\to 0} & R_{p\to si} & R_{p\to fe} \\ R_{\alpha\to p} & R_{\alpha\to\alpha} & R_{\alpha\to 0} & R_{\alpha\to si} & R_{\alpha\to fe} \\ R_{0\to p} & R_{0\to\alpha} & R_{0\to 0} & R_{0\to si} & R_{0\to fe} \\ R_{si\to p} & R_{si\to\alpha} & R_{si\to 0} & R_{si\to si} & R_{si\to fe} \\ R_{fe\to p} & R_{fe\to\alpha} & R_{fe\to 0} & R_{fe\to si} & R_{fe\to fe} \end{pmatrix}$$

This matrix presents accumulate α priori knowledge on the possibility of data classification into 5 categories. If all diagonal elements are greater than 0.6 (and therefore - the sum of all non-diagonal elements in each line is less than 0.4), you can expect unambitious results of fraction estimation after reconstruction procedures explained in next section.

The overall index reflecting the "goodness" *G* of features used is the following index of separability:

$$G = \left(\prod_{i=1}^{L} R_{ii} \right)^{1/L} \qquad (1.26)$$

This averaged product of diagonal elements represents the "mean" probability of true classification into one of L categories. This index, of course, is directly connected with Bayes error.

1.3.5 FRACTION ESTIMATION

Now let us estimate the *a posteriori* fraction of various classes in the distribution mixture. The best estimate of the a posterior fraction (Hey, 1983) (in case of a uniform *a priori* information and absence of systematic errors) is the empirical fraction

$$P_i^e = \frac{M_i}{M_{tot}} \qquad (1.27)$$

where, M_i is the number of events classified by the Bayesian decision rule (1.12) as belonging to the class A_i, M_{tot} is the total number of events. With account of classification errors the corrected fraction (p roportion) can be obtained as the solution of the following set of linear equations:

$$\sum_{k=1}^{L} \hat{P}_k R_{ki} = P_i^e, \qquad i = ,1, \dots L. \qquad (1.28)$$

where $\hat{P}_k$ is the estimate of the proportions P_k of the distribution mixture (1.4).

The accuracy of the estimates is defined by the TS size and number of control data as well as by the value of the Bayes risk, which represents the "quality" of discrimination with the chosen feature subset.

Note, that the set (1.28) 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 ($\geq 60\%$) of correct classification is a necessary preliminary stage.

For classification in two categories (for example "heavy" -Fe, and "light" - p nuclei) the system of two equations can be easily solved explicitly:

$$P_{Fe} = \frac{P_{Fe}^{e} - R_{p,Fe}}{1 - R_{p,Fe} - R_{Fe,p}}, \quad \mathrm{P_p} = 1 - P_{Fe}. \tag{1.29}$$

CHAPTER 2.
UPPER BOUNDARY OF IRON NUCLEI IN PRIMARY COSMIC RAYS AT $E > 10^{16}$ EV

Historically the multivariate nonparametric analysis methods were applied first to PAMIR collaboration emulsion chamber data (Pamir, 1973) with Lebedev institute group. Results obtained by new data analysis methodology were summarized by Nina Roinishvili in topical review (Roinishvili,1995). An emulsion chamber presents a sandwich of solid material interlined with x-ray films covered on both sides by sensitive layers. An electromagnetic cascade induced by electrons, positrons or γ-quanta is registered by the x-ray film. After development of the x-ray films the shower particles are observed by the naked eye as dark spots a few hundred μk in radius. The density of a dark spot is proportional to the energy of the initial particle and the relative displacement of two spots on the two sensitive layers of film allows measurement of the angle of incidence.

A layer of light material, a few interaction lengths thick, is located in the emulsion chamber between x-rays for registration of charged pions and nucleons. In this layer a hadron produces γ-quanta, mainly from π°, which are then registered in x-ray films. A dark spot from a hadron has the same size (~ 100 μk) as a single γ-quantum and its darkness is proportional to the energy converted by the hadron to the electromagnetic component. The energy threshold in emulsion chambers is about 2-4 TeV.

A nuclear-electromagnetic cascade in the atmosphere induced by a primary cosmic-ray particle is registered by the emulsion chamber as a group of "genetically connected" spots- all of which have the same incident angles, so called, γ-families. The typical size of a family is about 15 cm, though its size range is very wide, from 1 cm to several meters. The energies of the γ-quanta and hadrons as well as their distance from the family centre are measured. This is the raw information for a further analysis of the families.

The analysis of data obtained by means of emulsion chambers is especially complicated due to two circumstances:

- The observed spots results from well developed nuclear electromagnetic cascades in the atmosphere and it is very difficult to reconstruct the primary particle and its interaction phenomenology.
- The interaction which is studied is characterized by many unknown parameters, each of which influence cascade development.

Thus the problem is highly indirect. The solution could be only approached by the simulation of the cascade in the atmosphere. The simulations are based either on "physical" models prompted by a theory and tuned on the experimental data at lower energies or on special "simplified" models, intended to investigate the influence of certain parameters on cascade development. It is well known that the most difficult obstacle to cosmic-ray data interpretations is the yet unknown chemical composition of primary cosmic rays at the energies above 10^{14} eV.

An ambiguity in the data interpretation arises in connection with this uncertainty: most manifestations of a heavy nucleus interaction with an air nucleus are identical to that of a nucleon interaction with large energy dissipation. Due to this the longstanding argument of whether there is heavy nucleus dominance or scaling violation in the fragmentation region has

not yet been solved.

The statistical decision approach for multidimensional information has been elaborated over the past few years for the Pamir Collaboration data (Chilingarian, 1983), (Chilingarian et al., 1985). The aim was to divide experimental γ-families into those caused by light nuclei (class

L) and those caused by heavy nuclei (class H). To classify some events into these two classes L and H by means of a parameter Z (or a set of them—multidimensional information), so- called training samples based on certain models are used. The chosen parameter (or parameters) has to satisfy the following requirements: the Z distributions are essentially different in the two classes (L and H) and the experimental Z distribution is well described by their mixture.

The method is illustrated for the one-parameter case in the Figure 2. The distributions of the parameter Z for families induced by light nuclei (L) and heavy nuclei (H) obtained by a training sample for some models are shown in the upper part of the Figure 2. The conditional "experimental" distribution is in the lower part of the figure. The real experimental event is attributed to class H (heavy nuclei) if Z > Z0, which corresponds to

$$PH(Z) > PL(Z) \qquad (2.1)$$

Here P(Z) is the probability of obtaining the given value Z for the given classes L or H. In contrast the family is attributed to the class L (light nuclei) if $Z \leq Z_0$ (Chilingarian et al., 1987A).

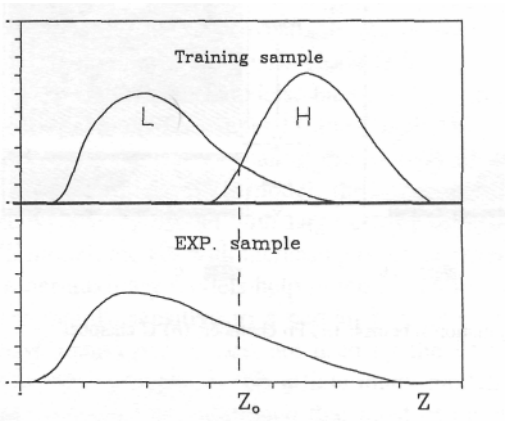


Figure 2. The scheme of events attribution to classes: If $Z > Z_0$ an event is attributed to the H class, otherwise to L.

Naturally all events produced by light nuclei with $Z > Z_0$ would be incorrectly labeled. Their fraction is denoted by RL. Similarly RH is the fraction of events produced by heavy nuclei which are mistakenly attributed to the class L ($Z \leq Z_0$). The training samples help to obtain RL and RH. They allow the experimental fraction of events attributed to class H (Fe in our case), (P^*_{Fe}), to be corrected by means of the misclassification rates RL and RH, obtained by training sample. Of course, the simulation model used for generation of simulation should be adequate.

$$P_{Fe} = \frac{P^*_{Fe} - R_L}{1 - R_L - R_H}, \qquad (2.2)$$

Here P_{Fe} is the corrected value for the fraction of families belonging to the H(Fe) class, assumed to be produced by heavy nuclei .For the problem under consideration training samples were simulated by the MJEp⊥ (Dunaevsky, 1986) model which includes an increasing cross section, scaling violation in the fragmentation region of hadron-air nuclei collisions (the inclusive

cross section at x = 0.3 is 2.2 times less at 1016 eV than at 1013 eV), charge-exchange and the production of semi-hard jets. This model describes most of the properties of the experimental families rather well. By means of training samples four classification parameters were chosen ⟨ER⟩, n′, b′ and Σ ′ER, where E and R are the energy and radius of the given γ-quantum, n' is their multiplicity and b' a parameter of azimuthally asymmetry of a γ-family. The ⟨⟩ brackets stand for the family averaging, the prime′ denotes that only most energetic quanta with fractional energy f = E/ΣE > 0.04 are taken into account. (ΣE is the total energy of a γ-family).

Two alternative chemical compositions of primary cosmic rays (PCR) were investigated: N, normal (42% p, 17% Fe at 10^{16} eV); and F, enriched by Fe (11% p, 63% Fe at the same E0). This means that two samples of families for two models were simulated with the same

MJEp⊥ model. We simulate separately γ-families according to the N and F chemical compositions. Then we select from simulations only families generated by the light nuclei, so called L events; by the heavy nuclei – H events; the case when we select all families we'll denote as M class. Therefore, we have 6 possible combining chemical composition and used and event selection method. For example, N-L model will denote families generated by the "normal" composition model and selection of the "light" primaries. Our goal now is to investigate which of theses 6 combinations provides parameters whose means are close to the experimentally measured ones.

Table 1 (from (Chilingarian et al., 1987A)) presents these comparisons for the families with total energy of 100-400 TeV at an energy registration threshold of 4 TeV and R ≤ 15 cm. It is apparent that M sets (especially N-M), i.e. the mixture of L and H sets, are compatible with the experimental data for all parameters, while L and H sets (classes) predict quite different values for the chosen parameters.

Table 1. The comparison of the mean values of parameters calculated from different simulations with corresponding experimental values.

	⟨ER⟩ (G eV cm)	n′	b′	′ER, (GeV cm)
Experiment	22.4 ± 0.9	10.4 ±0.2	0.211 ±0.009	260 ± 10
Model				
N-M	23.7 ± 0.6	9.9 ±0.1	0.218 ±0.007	273 ±7
F-M	26.2 ±0.9	10.0 ±0.1	0.227 ± 0.008	306 ±13
N-L	23.4	9.9	0.217	268
N-H	44.0	11.4	0.320	590
F-L	23.8	9.8	0.219	271
F-H	44.0	11.4	0.319	593

Simultaneously we can read from the Table 1 that chosen parameters are systematically different for the H and L classes. The, difference is smallest in the case of n′ variable and is more than 2 standard deviations for the ⟨Σ ′ER⟩ parameter.

The results of classification of the Pamir experimental data are described in (Denisova et al., 1987). Training samples for N (normal composition) and F (heavy nuclei enriched) models were used to determine the parameters of the decision rule (similar to Z_0). Different combinations of variables were used for the classification. Obtained decision rules were applied to classify the

Pamir families into L and Fe classes. Fractions of families produced by Fe nuclei for the chemical compositions N and F obtained in this way are given in Table 2. The maximal values (mean + one standard deviation) of P_{Fe} are shown in parentheses.

Table 2. The fractions of families produced by Fe nuclei for the two chemical composition N (normal) and F (Fe enriched) for the various family parameters.

Parameters	N-composition	F-composition
⟨ER⟩, b′	0.01 ± 0.04(0.05)	0.02 ± 0.04 (0.06)
b′,Σ′ER	-0.01 ± 0.04 (0.04)	0.02 ± 0.05 (0.07)
n′,Σ′ER	-0.001 ± 0.04(0.04)	0.05 ±0.05 (0.10)
n′, b′	-0.07 ± 0.09 (0.09)	-0.02 ± 0.09 (0.09)

The logic of the analysis of the results posted in the Table 2, are described in details in (Chilingarian et al., 1987). As we want to reject hypothesis that almost all primaries after "knee" are iron nuclei, we'll try to estimate proportion of irons in worst for it conditions, i.e., taking variables giving maximal value of irons among all used combinations, add one standard deviation for actual proportion, etc. The negative values of proportion, of course, have no meaning. There only signaling that "true" proportion is very near to zero and application of equation (2.1) introduce corrections shifting proportion to the negative region. When encountered such case, we assume that proportion is simply zero. Also to make our estimate of "iron families" proportion in some sense "model-independent" we use different combinations of family parameters and, as we already mention, take the "worst" case – maximal proportion. As can be seen, the maximal values for P_{Fe} are 0.09 and 0.10 for N and F compositions correspondingly.

Now the problem is to reproduce the fraction of iron nuclei in PCR, which is not trivial as families produced by light and heavy nuclei have different registration efficiencies. Again, selecting the "worst" conditions in estimating solid angle and registration efficiency (see for details (Chilingarian et al., 1987A)) we come to results posted in the Table 3.

Table 3. The upper limits of the Fe intensity and of the fraction of Fe in primary cosmic rays for the two chemical compositions N (normal) and F (Fe enriched).

Model	Max IFe (> 10^{16} eV) (m^{-2} s^{-1} $ster^{-1}$)	Max PFe (> 10^{16} eV)
N	5.9 x 10^{-9}	0.22
F	7.9 x 10^{-9}	0.29

The conclusion from the foregoing analysis is that the estimated value for the Fe fraction in the PCRs, 0.29, is in direct contradiction with that introduced in the Fe-enriched F-model (P_{Fe} (>10^{16} eV) = 0.63), while the fraction 0.22 for the N model is in reasonable agreement with the normal chemical composition.

CHAPTER 3.
STATISTICAL TECHNIQUES IN BACKGROUND REJECTION FOR ATMOSPHERIC CHERENKOV TELESCOPE

3.1 ATMOSPHERIC CHERENKOV TELESCOPES OPENING NEW WINDOW TO UNIVERSE

Most energetic processes in Universe are accompanied by acceleration of the elementary particles and ions to enormously big energies. Up to ~30% of total energy of the supernova explosion is believed to be taken by the high energy particles. Among these particles, high energy gamma-quanta born in the interactions of the high energy hadrons are of most interest, because, due to absence of charge, gammas didn't deflect in the magnetic fields filling the Universe. Therefore, detecting of the high energy gamma-rays will give possibility to locate sources of high energy particles acceleration and will help to solve long standing problem of cosmic rays origin, as well as open new window for detecting celestial objects in high energy particle fluxes.

The space-born gamma-ray detectors detect photons upward to energies in excess of 10 GeV. Compton Gamma-Ray Observatory, operated near 10 years detects hundreds of gamma-ray point sources. However we didn't expect in near future growing of the energy range of satellite detectors, due to very weak fluxes, typically proportional $\sim E^{-\gamma}$ where γ values are ~ 3. Thus, due to limitation in mass of satellite detectors a practical upper bound, to be reached by GLAST satellite to be launched in 2006 is 30-50 GeV.

Fortunately, ground-based techniques permit gamma-ray astronomy to be extended to much higher energies. Primary particles, interacting with atmosphere nuclei are generating cascades of secondary particles. These particles can reach mountain altitudes and be detected by the big surface arrays. Cascades of charged particles interacting with atmosphere nucleus also are generating Cherenkov photons (in the narrow cone around primary particle direction) and fluorescence photons (distributed much more wide).

These photons during moonless nights could be collected by the big optical reflectors and registered by the optical sensors located in their phocal plane.

So, using the atmosphere as a target, multiplying primary gamma millions times and collecting these low energy Cherenkov photons by reflectors with huge collecting power we can detect very weak flux from exotic objects in Universe. The main obstacle of the Cherenkov Atmospheric Telescopes (ACT) is presence of the overwhelming background of cosmic ray flux, though falling also as $E^{-\gamma}$, (γ= 2.7- 3) far exceeds the expected flux of gamma-rays. Hence, for implementing the ACTs crucial was to develop technique for distinguishing gamma-rays from cosmic-ray ions. Such technique was developed by the Whipple collaboration by the revolutionary changing of the 37 photomultiplier sensor pixels (5 cm diameter phototubes, spaced 0.5° apart; the full field of view was 3.5°). of big 10 m. optical reflector located on mt. Hopkins in Arizona by the "imaging" 109 phototube sensor (typical spacing of 0.25°) (Cawley et al, 1990).

3.2 IMAGING TECHNIQUE

Originally imaging was considered as a means of increasing the angular resolution of the ACT. Only later was the possibility of using this technique to identify the nature of primary considered (Cronin et al.).

Gamma and hadron initiated showers are distinct due to different kind of interactions: gamma showers are purely electromagnetic, generating mostly gammas and electrons before dying deep in the atmosphere; in hadron showers (mostly protons at energies up to 10 TeV) only approximately third part of energy is converted from the beginning to electron-photon cascade, two thirds of energy is converted to charged pion energies and undertake further strong interactions in the atmosphere, to dye finally again in electromagnetic shower.

Therefore, due to broad angular distribution of high energy pions taking significant part of primary energy and due to hard interacting core of shower including penetrating muons, the hadron showers are broader and longer comparing with more compact electromagnetic showers.

And, of course, there is pure geometrical distinction. ACT axes are directed towards the location place of the investigated source, keeping source(star or galaxy) at center of the field of view, therefore, the gamma-rays are incident parallel to the optical axes. Hadronic showers, initiated from highly isotropic cosmic ray flux, are not similarly constrained. The orientation of a shower's image in the field of view of the telescope depends at its angle of incidence with respect to orientation of the telescope axes. Showers incident parallel to the optical axes will have images pointed toward the center of the field of view. Of course, cosmic rays, traveling randomly from the same direction will have also the same orientation (but not shape).

Thus, both orientation of the image and its shape could be used for the background suppression. And because of the ability of Wipple Cherenkov telescope to discriminate between gamma-ray – and hadron-initiated air showers, the camera archived a flux sensitivity an order of magnitude better than conventional nonimaging telescopes (Reynolds et al., 1993).

The initial method for enumerated the image was proposed by Michael Hillas ((Hillas, 1985), see Figure 3), so called "Hillas" parameters till now are widely used in ACT's of next generation for the "purification" of the initial image samples. Best ellipse was fitted to the calibrated (details on the phototube tuning, "pedestal" extraction and other experimental technique one can find in (Cawley et al, 1995) phototube counts, giving a best mean square estimate of ellipse axes size and relative orientation, as well as distance till center of field of view, where investigated point source is projected (see Figure 3). This rather simple parameterization, giving separately shape and orientation of the image, prove to be very useful, implementing in last decade to many ACT's of next generation.

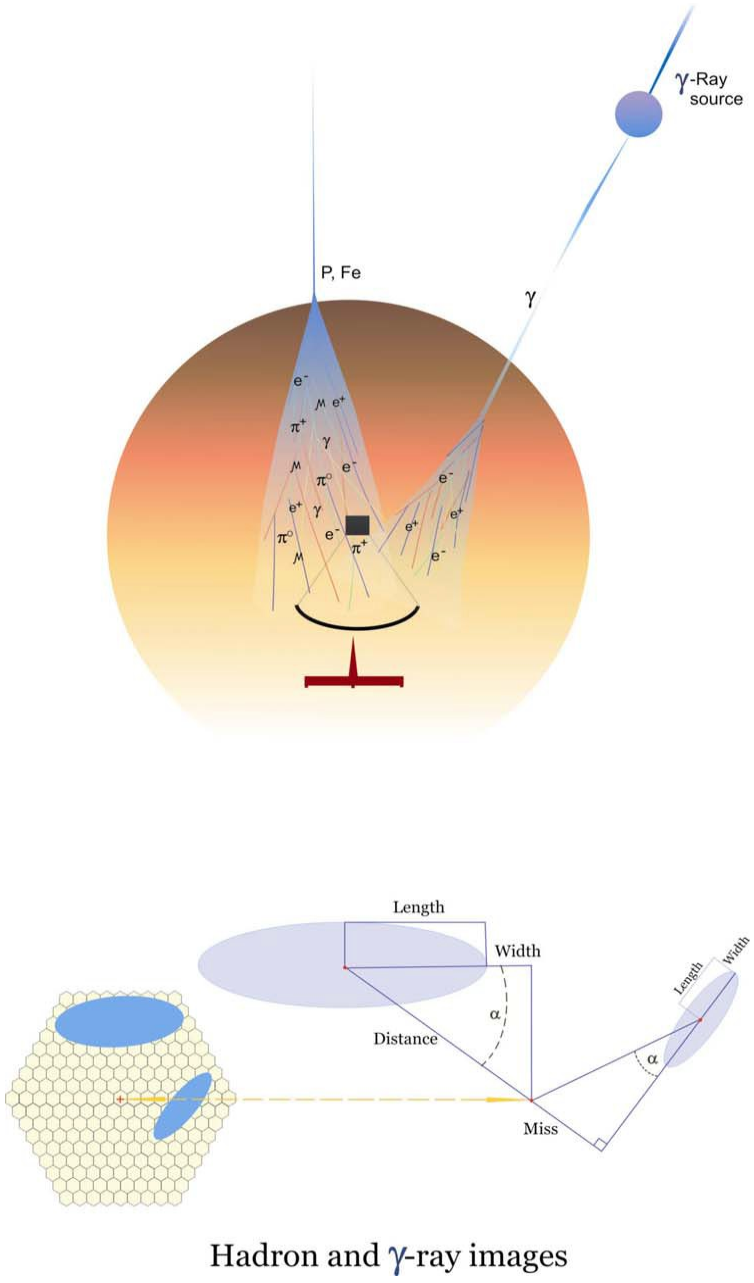


Figure 3. Definition of the "Hillas parameters" for hadron and gamma image

3.3 BACKGROUND REJECTION STRATEGIES

Proceeding from the estimated image parameters different approaches to construct the image classification procedure could be developed.

These strategies could be divided into big categories:

1. *a-priori* strategy, proceeding from the simulation of the gamma and hadron-initiated showers and it's propagation in the atmosphere. For each shower the Cherenkov light image is obtained and parameterized. The distribution functions of the "pseudo experimental" image parameters corresponding to both gamma and hadron primaries are given in nonparametric form as subsets of simulation trials corresponding to the particular primary with definite angles of incidence, energy and, of course, inherent instrumental accuracies and hard-ware solutions chosen for construction of ACT and trigger.

a-posteriori strategy, proceeding from already registered, so called,

- "ON" sample- obtained with telescope axes oriented in direction of the putative gamma-ray source and
- "OFF" – sample, obtained by pointed telescope axes in direction of the same celestial co-ordinates, but after the source already leave the destination.
- If the field of view of the telescope is enough large it is possible to take "ON" and "OFF" scans simultaneously, selecting within field of view samples pointed to source and to "empty" space.

Having 2 samples, one containing signal, another – not, it is possible to search for, so called, "signal domain", one or multidimensional boundary outlining the space in parametric space where signal concentration is relatively higher comparing with noise, those enlarging the signal-to-noise ratio..

Empirical distribution of the hadron showers could be obtained with high significance by detecting "pure" background, pointing telescope axes to the region where no potential sources are expected. The distribution of gamma-image parameters are obtain with Monte Carlo programs simulated shower development in the atmosphere, transport of the Cherenkov photons till optical reflector, and then simulating of trigger conditions, experimental noise and other sources of distortions. Modern simulation programs like CORSIKA (Heck et al., 1998) are sophisticated enough to count for such physical effects as change of density of atmospheric profile and influence of the magnetic field of Earth on the shower particles.

Nevertheless, we are not aware of any successful attempts to parameterize multivariate shower parameter distributions using any analytic form of the distribution function.

Just first comparisons of simulated showers obtained by the codes developed for the Whipple 109 phototube camera reveal big differences in shape and orientation of the gamma and hadron images. First the main activity of Whipple collaboration was directed to invention of the single "best parameter", combining shape and orientation differences.

Among numerous more or less successful trials the AZWIDTH parameter (Hillas, 1985) proves to be best, rejected the background (of course, from the model samples) down to few tenths of percent, at a few hundred GeV (over a field of view of 3°), whilst still accepting more than 50% of gamma showers.

First attempt to develop statistical theory of background rejection was done in (Aharonian et al. 1990),(Aharonian et al., 1991). The statistical theory includes:

- Selection of the optimal subset of parameters for discrimination purposes;
- Introducing of the Bayesian decision rules;
- Introducing of the P-values of statistical tests as measures of the "closeness" of the parameter distributions;
- Correlation analysis revealing best pairs to be used in discrimination;
- Estimation of the Bayes risk (probability of misclassification) as measure of closeness of multivariate distributions;
- Parzen and K-Nearest-Neighbor nonparametric density estimation modes with adaptation of the method according to random sample.

Overall scheme of statistical inference from simulated samples, including method of selecting "best" subset of image parameters was checked by simulated samples performed with

code developed by A.Plyasheshnikov. (Plyasheshnikov et al., 1985).

The combination including image shape parameters and AZWIDTH parameter (there was not yet included in simulation "pure" orientation parameter - "ALPHA") was pointed as most perspective.

Historically, in the end of 80-ths, different discriminants were intensively applied to the first unambiguous detection of Crab nebula by the Whipple collaboration, contained in the famous 1988-1989 data files, consisting of 65 ON-OFF pairs, more than 1 million images (Vacanti et al., 1989).

To prove existence of the source, one looks for an abundance ($N_{on} - N_{off}$) of events coming from the direction of a possible source (N_{on}) as compared with the control measurement, when pure background is registered (N_{off}). As the expected fluxes are very weak (the signal to background ratio is less than ~ 0.01), one should always answer the following question: is the detectedabundance a real signal or only a background fluctuation? The measure (level) of statistical significance used in gamma-ray astronomy is the so-called criterion size (σ) (Zhang et al., 1990):

$$\sigma = \frac{N_{on} - N_{off}}{\sqrt{N_{on} + N_{off}}} \qquad (3.1)$$

The greater is σ the lesser is the probability that the detected excess is due to a background fluctuation. Development of new detector hardware and new data handling methods aim to enlarge the value of σ. After selecting the "gamma-like" events from raw data (both from the ON and OFF samples), the criterion takes the form:

$$\sigma^* = \frac{N^*_{on} - N^*_{off}}{\sqrt{N^*_{on} + N^*_{off}}} \qquad (3.2)$$

where N^*_{on}, N^*_{off} are the of events surviving image selection cuts.

The initial value (measured by the raw data) of σ for the 65 ON-OFF pairs equals to 4.8 (Table 4). After applying the AZWIDTH cut, selecting "small" shape events well aligned to the center of point of view of telescope, it was possible to enlarge σ up to 20. Implementing the multivariate "Wedge" cut (WIDTH, LENTH and AZWIDTH, using the correlation between AZWIDTH and WIDTH, significantly different for the gamma and hadron images) (Chilingarian et al., 1990), (Chilingarian et al., 1991B) the σ was enlarged to 27 (more than was possible with any single parameter).

Following improvement of the image classification technique was done by Michael Punch (Punch et al., 1991), (Punch, 1994), introducing, so called, "supercuts", remaining till now the basic strategy of the image selection. Instead of AZWIDTH, new parameter ALPHA, specifying the orientation of the major axis of the image relative to the source position was introduced, as having more obvious interpretation than AZWIDTH. Also DIST parameter specifying the position of the image relative to the telescope center of view was added to the LENGTH, WIDTH and ALPHA combination. The σ value after applying the supercuts – rectangular 4-dimension gamma-domain – reaches value of 34.

Along with multidimensional approach, the Neural Network technique also was proposed for the image classification (Chilingarian 1991A). Neural Network is universal generalizing device providing mapping of the input multidimensional information, i.e. image parameters, to the class assignments --see for example, (Bishop, 1995). To provide correct classification

Network should be "trained" on the samples with known membership. The gamma and hadron images obtained from simulations are used for Network training. The total number of misclassifications (so called "classification score") is the "goal" function – to be minimized in the training process. After training the Network is used for classification of the experimental data, or for the determination of Network performance - for classifying simulations not used for training. However, Neural classifiers trained on simulations, as well as Bayesian classifiers fails to outperform the simple techniques of multidimensional cuts, tuned on the experimental data itself. Those, the *a-priory* technique couldn't compete with *a-posteriori* technique for experimental images analysis. The main reason of the weakness of *a-priory* approach is inherent oversimplification of the simulation and difficulties to account for the all kinds of experimental errors. Therefore, *a-priory* analysis is used mostly for the outlining expected values of the multidimensional cuts and for the investigation of possible nontrivial correlations. Also *a-priory* analysis is very important in telescope design studies, in trigger and analysis methodologies selection.

Therefore we decided to use Neural Networks as the *a-posteriori* analysis device. Instead of using the "classification score" as goal function of Net training (utilizing the information on the imagebelonging to the gamma or hadron classes) we maximize directly the σ value, obtained by the executing experimental ON-OFF pairs, and not pure gamma and hadron images from simulation. We use a simple feed forward Network with number of nodes 4:: 5:: 1 (Chilingarian, 1994) to select a more realistic nonlinear shape for the gamma domain. During the iterations each particular ON& OFF event was classified according to following decision rule:

$$OUT(\vec{x})\begin{cases} < C^* \rightarrow \vec{x} \text{ is classifed as signal} \\ \geq C^* \rightarrow \overrightarrow{x} \text{is classifed as background} \end{cases} \tag{3.3}$$

where OUT(x) is the output node response for an image parameters subset x. Choosing an appropriate boundary value of net output (changing from 0 to 1) we determine, so called, decision point C^* to define the classification procedure: an event with an output greater than or equal to the decision point is attributed to the background class, while all the other events areassigned to the gamma class.

By moving the decision point C^* along the (0, 1) interval we can change the relation between the errors of the first and second kind (the position of the decision point is the neural analog of the lossfunction in the Bayesian approach). This decision rule is a Bayesian decision rule; therefore the output signal of a properly trained Neural Net is an estimate of the a posteriori probability density (Ruck et al., 1990). The Network training is performed using one of stochastic training scenarios implementing in the ANI package (Chilingarian, 1998). By making random changes in the multidimensional space of the Network weights we are searching the optimal weightconfigurations, outlining optimal shape of the gamma domain.

After executing off all images a new σ value is calculated each time and if its value enhanced comparing with previous value, the changes in the Network weights survive, if not, changes are subtracted and new random step is implemented. After several hundred thousands of iterations the more complicated gamma-domain shape is outlined and the σ value was enlarged up to 35.8 (Chilingarian, 1993), (Chilingarian, 1995), (Chilingarian et al., 1994).

The comparison of different background suppression methods is shown in Table 1, where DIFF $=N_{on}^* - N_{off}^*$ is the estimate of the signal, DIFF/ N_{off}^* is the estimate of the signal-to-noise ratio,

N^*_{on}/N^*_{off} is the estimate of background suppression by the technique used.

Table 4. Comparison of the different background rejection methods

	N^*_{on}	N^*_{off}	σ	DIFF	DIFF/N^*_{off}	N^*_{off}/ N_{off}
Raw	506255	501408	4,8	4847	0.01	
Azwidth	14622	11389	20.4	3233	0.28	0.0227
Wedge cut	6017	3381	27.2	2636	0.78	0.0067
Supercut	4452	1766	34.3	2686	1.52	0.0035
Neural 4::5::1	6278	2858	35.8	3420	1.20	0.0057

Concluding this section we can also state that proposed Michael Hillas simple second moments imaging proves to be very relevant to the experimental accuracy of the first imaging telescopes. And introducing of the new methods--see for example (Danaher, 1993) and topical review of David Fegan, (Fegan, 1997)--couldn't introduce any significant improvement to the results summarized in Table 1. Also adding of the additional image parameters --see (Badran et al., 1997)--didn't enhance the background rejection to the levels significantly differing from the ones presented in Table 4.

Now, when next generation of ACTs (HESS, MAGIC) is about to start collect data, the challenge is to investigate if the same data analysis methodology have to be change to the new one, exploiting in more details the information from much bigger optical sensors.

3.4 A PRIORI METHODS OF BACKGROUND REJECTION

As was mentioned in the previous section, *a priori* methods of background rejection although are not so effective as a-posteriori ones, are very useful for the detector design purposes. For each new telescope it is necessary to investigate appropriate data analysis methodologies for selecting one providing required level of background rejection and signal acceptance efficiency.

Usually such studies are preceding the start of telescope operation and intensive simulations are required for the understanding telescope performance and preparing necessary software for data analysis and physical inference. Having enough simulated samples of "pure" signal and background on hand we can apply one from the several possible decision rules (say we reject image if its WIDTH > 0.2°) and calculate both kinds of the misclassifications: loss in the signal detection efficiency and contamination of the selected "signal" sub-sample by the noise. Plotting these pairs we obtain the operational characteristic of the decision rule used, reflecting its power and possible tradeoff between both kinds of errors. Of course, as it is well known from statistical decision theory, it is impossible to reduce simultaneously both kinds of errors. Therefore, this *decision quality diagram* is measure of the adequateness of the decision rule chosen (by decision rule we understand both the algorithm and its particular realization by code and chosen subset of image parameters). Comparing curves obtained with different classification strategies, but one and the same data we can outline the most appropriate one.

If it is difficult to compare the curves, we can fix signal accepting efficiency at say ~50% level, and compare corresponding contaminations of selected "signal" sub-sample by hadronic images

In such studies, trying numerous statistical classification methods with many tunable parameters, always a question arises: are we utilizing all distinctive information completely, are

there any chancres to improve the quality of classification?

To answer this question we have to adopt some general scheme of the statistical decisions and give rigorous formulation of the decision optimality criterion.

Let's assume first that we know the probability function $f_h(\boldsymbol{v})$ defined in each multidimensional point $\boldsymbol{v}$ of a subspace $\boldsymbol{V}$ of the metric space of image parameters $\boldsymbol{R^n}$ (say, 3 dimensional space combining shape and orientation parameters – *WIDTH, LENGTH and ALPHA*). $\boldsymbol{V}$ subspace is comprised from "realistic" values of the image parameters, in correspondence with expected distribution of image parameters as measured by particular detector setup. The probability function $f_h(.)$ is quantifying the information on belonging of

particular image with parameters ($\boldsymbol{v_i} \equiv WIDTH_i, LENGHT\ and\ ALPHA_i$), $\boldsymbol{v_i} \in \boldsymbol{V}$, to background (hadron) class. The symmetric probability function $f_\gamma(.)$ quantifying belonging to alternative signal (gamma) class also is defined in each point of $\boldsymbol{V}$ and is connected to $f_h(.)$, by

$$f_h(v) + f_\gamma(v) = 1, v \in V \qquad (3.4)$$

We can assume, that in principle the $f(.)$ functions could be obtained by extremely large cycle of simulation (not possible with existent CPU power) with very detailed Monte Carlo (MC) code (not available yet) mimicking traversal of all particles and ions through atmosphere and transporting Cherenkov photons till phototubes. The input parameters for these simulations should be chosen from a subspace $\boldsymbol{T^n}$ of the metric space $\boldsymbol{R^n}$ of all parameters influencing the experimental measurement. $\boldsymbol{T^n}$ consists of gamma quanta (or hadron) energy, angles of incidence, conditions of Earth magnetic field, night sky illumination, possible local light conditions, many of experimental facility parameters.

Or, we can assume the imaginary calibration experiment with orbiting accelerator sending different type of particles in direction of ACT.

This concept is reminding the "stable candle "concept, sometimes referring to Crab Nebula highly stable gamma-ray source.

Resuming, we have on hand 2-way classification problem. When dealing with classification, with some decision on the belonging of the measurement to the one or another class we need not only probabilistic description of the alternative classes, but also an algorithm (decision rule) performing classification. We are interesting in finding the best solutions; therefore we need also some measure (loss function) counting for consequences of any decision committed by the adopted decision rule. Bayesian paradigm consists in introducing of several spaces and probabilistic measures on it.

- The space of the "basic states - states on Nature" - Signal – Background, Gamma – Hadron, we denote this space as A. In our case A space consists of 2 members: γ and h, we'll denote these basic states as A_h and A_γ.
- The "Outcomes space" – space of possible statistical decisions we denote as $\tilde{\boldsymbol{A}}$.
- The loss function C is defined on the direct product of "basic states" and "decisions spaces" $A \otimes \tilde{A}$

The Bayesian decision rules minimize the losses due to incorrect classification:

$$\tilde{A} = \mathrm{argmax}_{ij}\{C_{ij} \cdot p(A_i/v)\}, i, j = h, \gamma \qquad (3.5)$$

where $\boldsymbol{v}$ is a multivariate measurement (image parameters) and

$$p(A_i/v), i = h, \gamma \qquad (3.6)$$

are estimates of *a-posteriori* probability densities, connected with conditional ones by Bayes theorem:

$$p(A_i/v) = \frac{P_i \cdot p(v/A_i)}{p(v)}, i = h, \gamma \qquad (3.7)$$

where

$$p(v) = P_h p(v/A_h) + P_\gamma p\left(v/A_\gamma\right) \qquad (3.8)$$

and P_i are a-priory probabilities.

Conditional probability densities are directly connected with the $f_h(\boldsymbol{v})$ probability function introduced above:

$$f_h(v) = \frac{C_h P_h p(v/A_h)}{C_h P_h p(v/A_h) + C_\gamma P_\gamma p\left(v/A_\gamma\right)} \qquad (3.9)$$

Substituting the equation of *a-posteriori* density (Bayes theorem, (3.7)) in the Bayesian decision rule we'll obtain new formulation as follows

$$\tilde{A} = argmax_{ij}\left\{C_{ij} \cdot P \cdot p(v/A_i)\right\}, i, j = h, \gamma \qquad (3.10)$$

The multiplier $C_{ij} \cdot P_i$ is affecting the decision rule as one entity; you can't separate the influence of *a-priory* probabilities from the effect of losses function on the performance of decision rule. And it is senseless to define and plug in separately both measures; therefore we introduce the notion of the "*a-priory* losses" in one term equal to $C_{ij} \cdot P_i$. Further we'll denote *a-priory* losses again as C_{ij}.

Bayesian decision rule maps a decision boundary (*db*) in $\boldsymbol{V}$ (point – in one-dimensional case, line – in 2 dimensional case and surface in three dimensional case) and the misclassification losses are determined by the probability "mass" lying on left or right from these boundaries. If we assume that losses are equal to zero for correct classification, $C_{h\gamma} = C_{\gamma h} = 0$, then minimal misclassification risk classifying the image with parameters $\boldsymbol{v}$, so called, Bayes risk is determined by

$$r^b(v) = \frac{argmax_i\{C_i \cdot p(v/A_i)\}}{p(v)}, i = h, \gamma \qquad (3.11)$$

To get more insight in misclassification rate let's consider one-dimensional case. In this case the decision boundary is shrinks to a point *db* where:

$$p(db/A_h)\}/p(db/A_\gamma) = C_\gamma/C_h \qquad (3.12)$$

$$f_h(db) = \frac{C_h}{C_h + C_\gamma} \qquad (3.13)$$

and we can obtain for mathematical expectation of Bayes risk $R^b = E(r(\boldsymbol{v}))$, following equation

$$R^b = l(v < db)\int_{db}^{\infty} p(v/A_h)\,dv + l(v \geq db)\int_{-\infty}^{db} p(v/A_\gamma)\,dv \qquad (3.14)$$

where l is indicator function

$$l(\mu) = 1 if\ \mu\ is\ true, 0\ other\ wise \qquad (3.15)$$

As we haven't possibility to obtain the full probabilistic description (conditional densities $p(v/A_h)$ and $p(v/A_\gamma)$) of classification problem from unbiased perfect simulation, or from calibration, we are obliged to use some Monte Carlo code for approaching the problem. We can generate several realizations of the simulation code with input parameters uniformly selected from $\boldsymbol{T^{n1}}$ subspace of $\boldsymbol{T^n}$ ($n_1 < n$). We have to select limited number of relevant of MC code input parameters (n_1) allowing us to generate in limited time simulation trials covering $\boldsymbol{V}$ subspace of the realistic image parameter values. We'll denote the image parameters obtained by this code by $\boldsymbol{u_i}$, $i=1, M$, M is number of simulation trials. The sample of simulation trials we'll name *training sample.*

The probabilistic measure induced by the M $\boldsymbol{u_i}$ vectors on $\boldsymbol{V}$ will differ from the "genuine". Instead of unique probabilistic description in ideal case – function $\boldsymbol{f(.)}$ or conditional densities $\boldsymbol{p(v/A_i)}$, we have to deal with random estimates of these functions obtained from limited samples $\boldsymbol{u_i}$, $i=1, M$. Therefore, these estimates of conditional densities will introducing additional (to Bayesian) misclassification errors. The decision rule, using these density estimates will not be any more Bayesian decision rule and error committed by this decision rule r will be greater comparing with Bayes error. For minimization of these errors and for approaching theoretical minimal Bayes risk we have to select relevant density estimation methodology. Following the analysis of Jerome Friedman (Friedman, 1996) we'll assume that the estimates of conditional densities $\hat{p}(v/A_\gamma)$, $\hat{p}(v/A_h)$ in each point $\boldsymbol{v} \in \boldsymbol{V}$ are random Gaussian variables with mean value E ($\hat{p}(v/A_i)$) and variance $\sigma(\hat{p}(v/A_i))$. The mathematical expectation of the random variable $\hat{p}(v/A_i)$ is taken according to different training samples of the same size M. Following analysis is easier to perform by transforming conditional densities to probability function $f_h(\boldsymbol{v})$, according to equation (3.14).

We can make same assumptions about estimates of probability function $\hat{f}_h(\cdot)$ as about estimates of conditional densities. Let's denote following paper of Friedman the distribution function and moments of estimates of $\hat{f}_h(\cdot)$ as $p(\hat{f}_h)$, $E(\hat{f}_h)$, $\sigma(\hat{f}_h)$

Then the additional error due to smearing of the could be written in the following form

$$r(v) = l(f_h(v) < f_h(db))\int_{f_h(db)}^{\infty} \hat{p}(\hat{f}_h(v))d\hat{f} + l(f_h(v) \geq (f_h(db))\int_{-\infty}^{f_h(db)} \hat{p}(\hat{f}_h(v))d\hat{f} \qquad (3.16)$$

where $f_h(v)$, $f_h(db)$ are "true" values of probability function in arbitrary $\boldsymbol{v}$ point and at decision boundary. The additional (to Bayes error) errors are determined by the distribution of the $\hat{f}_h(v)$ function around the true decision boundary $f_h(db)$. Integrating the tails of distribution separately for the greater and smaller values (comparing with true value of boundary value) we count for additional errors connecting with sampling and density estimation errors.

In case of symmetric losses function $C_{h\gamma} = C_{\gamma h} = 1$, probability function at boundary is equal to

½, but as we already discuss the hadronic background rejection problem for ACT is highly nonsymmetrical problem due to huge difference in gamma and hadron fluxes, therefore we'll didn't simplify equation by putting explicitly ½ value in equation (3.16).

After plugging in the assumed Gaussian shape of distribution in (3.16) we obtain (Friedman, 1996) following equation

$$r(v) = \widetilde{\Phi}[sign(f_h(v) - f_h(db))]\frac{E(f_h(v)) - f_h(db)}{\sigma(f_h(v))} \qquad (3.17)$$

$$\widetilde{\Phi}[z] = \frac{1}{2\pi}\int_z^{\infty} e^{-1/2x^2}\,dx \qquad (3.18)$$

The term

$$b(f_h(v), E(f_h(v)) = sign(f_h(v) - f_h(db))(E(\hat{f}_h(v)) - f_h(db)) \qquad (3.19)$$

is governing the behavior of the *r(v)* in the very different way comparing with density estimation errors. Estimating density we are optimizing some global measure, defined on all function support, i.e. integrated mean-square error L2, or integrated module of function estimation error – L1 (Devroye et al., 1985). In the case of the classification error only local region around the decision boundary is important and therefore the influence of function estimation method is also very different.

Examining the term (3.19) (J.Friedman name it "boundary bias", but we use a bit different equation), we can see that the most important is to have both $f_h(v), E(\hat{f}_h(v))$, on the same side (both greater or both smaller) concern boundary value $f_h(db)$. So, if our function estimate isn't too biased and we haven't many of images having the boundary value $f_h(db)$in between $\left[f_h(v), E(\hat{f}_h(v))\right]$of interval we can dramatically reduce the classification error by reducing the variance of probability function estimates $\sigma(\hat{f}_h)$ (see (3.17)).

If mathematical expectation of the probability function estimate is significantly biased concern true value, the situation can dramatically change and decreasing of the variance estimate willlead to increasing of the misclassification errors. Both goals (keeping bias not very large andreducing variance) could be achieved by using, so called, median estimate introduced in 1989 (Chilingarian, 1989).

Probability density function is estimated by training samples using Parzen's (Parzen, 1962) or K Nearest Neighbors (KNN) (Fukunaga, 1990) methods with automatic kernel width (r K parameter) adaptation. Instead of using one method parameter for all density support a several numbers of parameters (Parzen kernel widths h_k, or K_k neighbors) are implementing forobtaining a number of density estimates.

Then the final estimate is obtained from ordered sequence of these density estimates. The median of the ordered sequence is chosen as a final estimate. The probability density for each parameter h_k is estimated by

$$p(v/A_i) = \frac{(\det\sum)^{-1/2}}{2\pi^{d/2}h_d^k}\sum_j^{M_i} e^{-r_j^2/2h_k^2} \qquad (3.20)$$

one density estimate several density estimates are calculated (selected from some predetermined parameter set).

where d is the feature space dimensionality, M_i, is the training sample size, h_k is width of the Parzen kernel function, r_j is the distance to the *jth* neighbor in the Mahalanobis metric (Mahalonobis, 1936):

$$r_j = (v - u_j)^T \sum^{-1} (v - u_j) \qquad (3.21)$$

where Σ is a sampling covariance matrix of the class to which $\boldsymbol{u}$ belongs.

Our investigations of the median estimate prove both unbiasness and reducing variance leading toBayes risk estimates very close to theoretical values (Chilingarian, 1989), (Chilingarian et al., 1984). Therefore we apply the median density estimates for calculating the possibilities of background rejection using simulation for the MAGIC experiment (MAGIC A).

3.5 ESTIMATES OF MAGIC ACT BACKGROUND REJECTION RATES

Recently the first sets of MAGIC detector response simulations set became available at MAGIC home page (MAGIC A). The description of the parameters and data sets are also posted in the same WEB page. Image parameters were calculated by the Monte Carlo program, Corsika (Heck et al., 1998) adding simulation of the MAGIC setup and trigger conditions.

MAGIC telescope is located on the Canary island of La Palma and now is under assembling; it has the largest mirror surface of all existing gamma ray telescopes (234 m^2) and a camera with 577 pixels. For more technical details refer the MAGIC collaboration home page (MAGIC A).

After selecting best discriminants and discriminant pairs according to technique first developed in (Aharonian et al. 1990), (Aharonian et al., 1991) the same parameters as for Whipple telescope simulation prove to be the best, There are shape and orientation parameters WIDTH, LENGTH and ALPHA, also negligible improvement could be achieved adding the DIST and SIZE parameters (details are posted in the Yerevan MAGIC group report from July 2002, (MAGIC).

After investigating single parameters and parameter pairs we perform intensive study of different variable subset using the Bayesian techniques described in details in previous paragraph.

Also we investigated the source elevation influence on achieved classification results. It should me mentioned that in first trial when the elevation were changed in the interval (0-5°) the results was a factor 2 better (in units of the hadronic images contamination for the approximately fixed signal efficiency) than if we use broader elevations range (0-20°). Therefore, if we explicitly add the source elevation THETA in parameter set we again obtain same results (2-3% contamination keeping 50% efficiency) as described in our report from July 2002 (CRD, 2002).

The results of subset comparison are posted in the Figure 3. The most striking feature of the classification plots is, that the shape parameter are discriminated an order of magnitude worse comparing with Whipple telescope. To get insight in detected contradiction we compare one and 2-dimensional patterns of Whipple -using the same simulation as one used (Chilingarian et al., 1990) and MAGIC parameters. As the energy threshold of the Whipple telescope isapproximately 300 GeV, we select from MAGIC simulations also images corresponding to the 300 GeV primaries. In such way we only "improve" discrimination ability, because reject numerous low-energy hadronic images, overlapping with gamma images corresponding to muchgreater energies. From Figure 3 and Figure 4 it is apparent that although the image parameters are calculated using different units (millimeters for MAGIC and degrees for Whipple), that theoverlapping of hadronic

and gamma shape parameters are much more for the MAGIC simulation comparing the Whipple data. The most valuable discriminative feature of gamma images – compactness of the shape comparing with more long and wide hadron images is completely lost in MAGIC images. Presence of the overwhelming amount of small images in MAGIC hadronic simulation didn't allow to effectively use shape parameters for background rejection.

Before inferring that MAGIC camera isn't good enough for discrimination purposes it should be checked if routine image cleaning operations for huge MAGIC camera were done properly for the simulated images. In any case, as was mentioned in another Yerevan group report from October 1999 (CRD, 1999), very low thresholds and very big cameras of new generation of ACT should pose more firm restrictions on the image analysis software. The Hillas parameter approach, very well suited the first generation ACT's with considerably high threshold (>300 GeV) and moderate camera size, may be will not work for 30 GeV threshold and huge MAGIC camera.

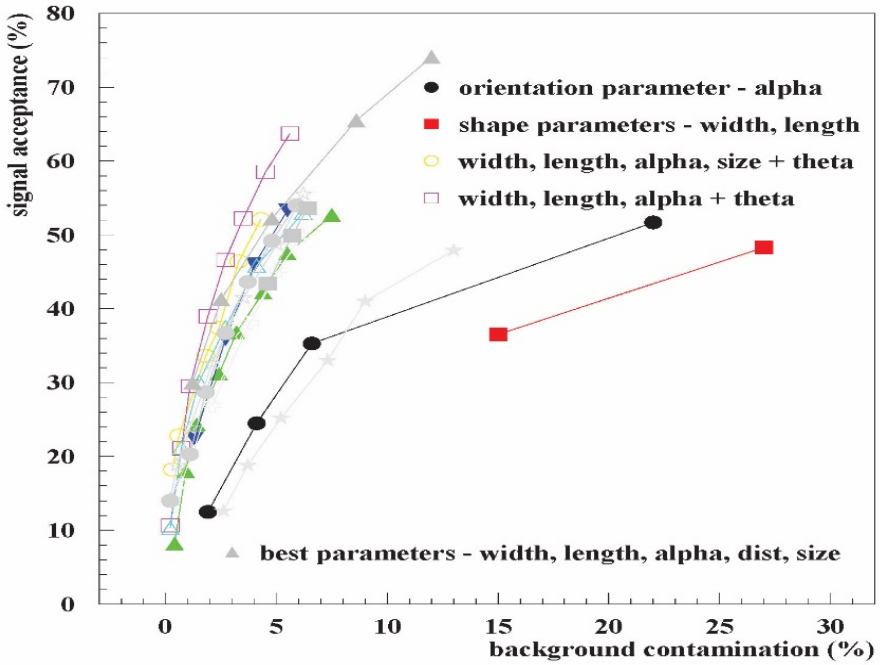


Figure 4. Comparisons of different image parameter subsets

In the October 1999 report first attempts were described to use all pixel information instead of parameterized first moments of the image. MAGIC images expected to be too detailed and discrepant to be reduced to first moments only. We'll lost valuable discriminative information washing out all "islands" and leaving only central spot.

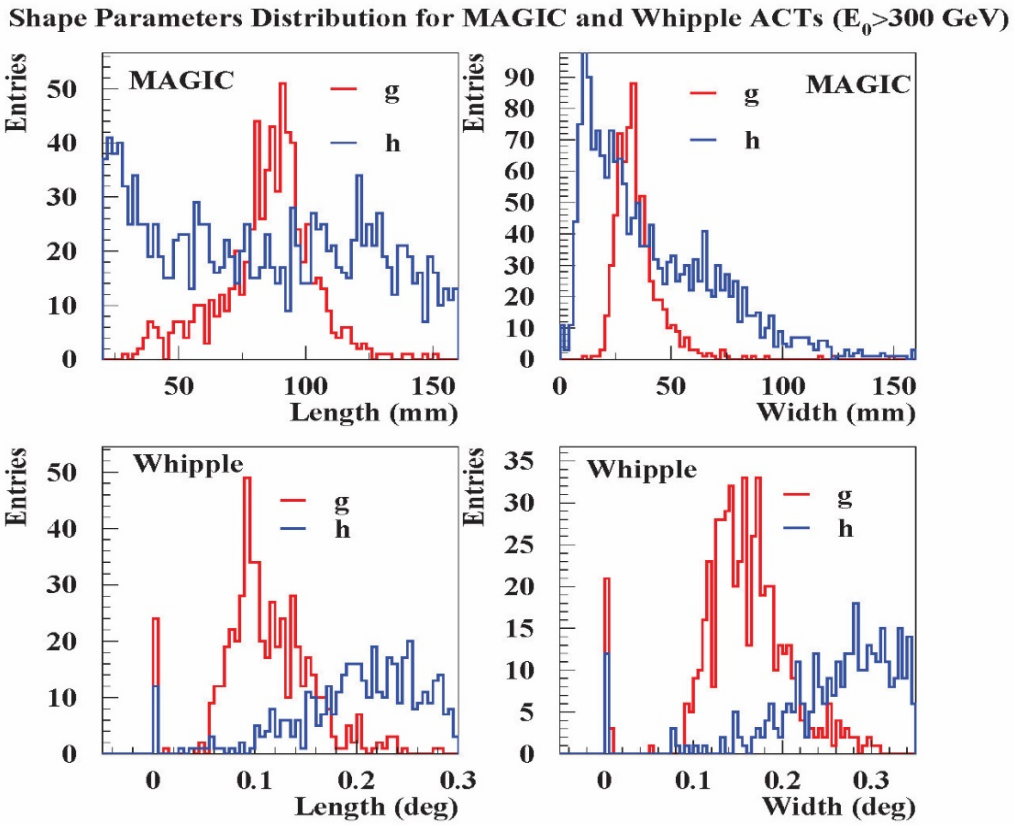


Figure 5. Comparison of the shape and orientation parameter distributions for MAGIC and Whipple ACT

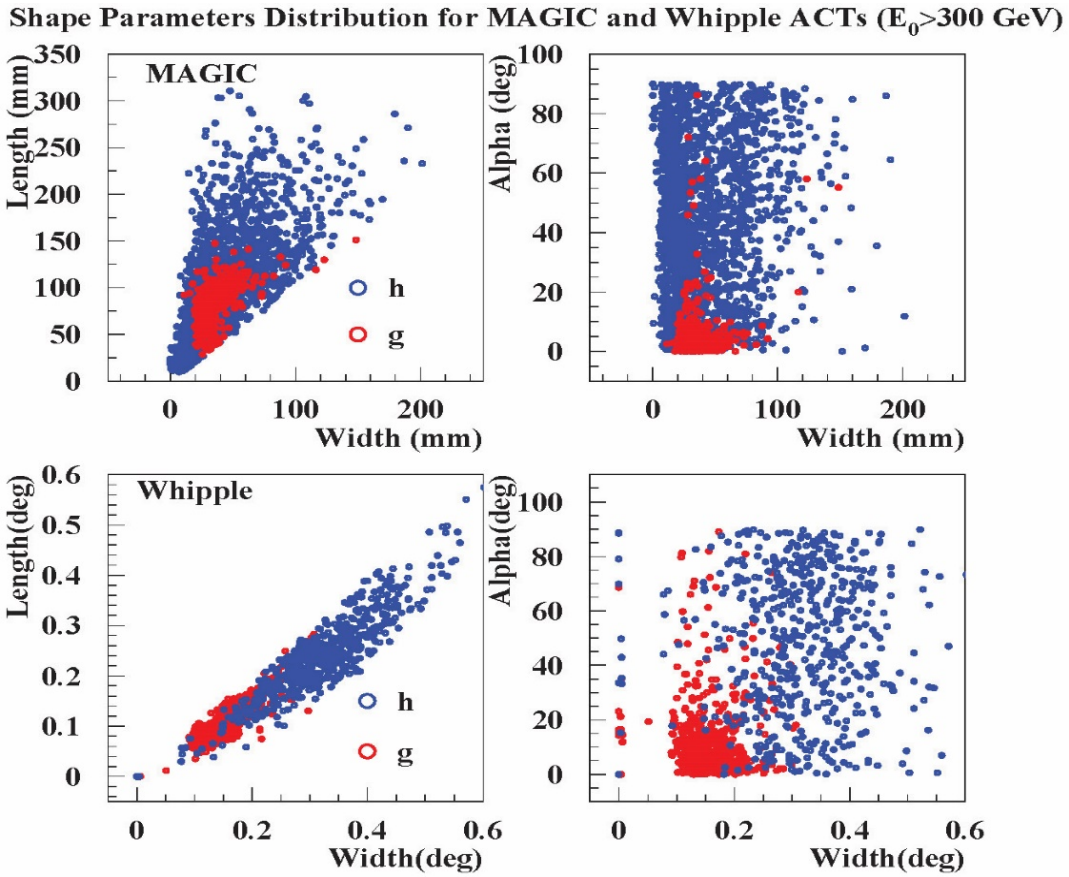


Figure 6. Two-dimensional scatter plots of MAGIC and Whipple nontrivial correlations

3.6 CONCLUSIONS

More than a decade experience with ACT background rejection techniques proved:

A-posteriori methods of background rejection are superior to any modification of the *a-priory* methods, despite what exotic method of classification is used. Neural Network technique can outline complicated shape of gamma-domain improving results obtained with super-cut method;

Among a-priory method Bayesian decision rules with nonparametric density estimation provide better results, especially when estimate variance reducing techniques are used;

It is not useful to add big list of parameters in subset for background rejection – there is no results reporting significant improvement comparing with "classical" combination shape & orientation parameters;

For the ACT's with huge matrices combined with low threshold "classical" methods using Hillas parameter aren't any more optimal, new methods utilizing the pixel information in all details should be developed and used for background rejection.

A few minor comments:

CHAPTER 4.
THE NEURAL INFORMATION TECHNIQUES

4.1 FEED-FORWARD NEURAL NETWORKS

Feed-Forward Neural Networks (FFNN) represent very simple structures composed of processing elements (nodes) and connections (weights). FFNN belongs to the general class of non-parametric methods that do not require any assumption about the parametric form of the statistical model they use.

The central issue of FFNN is implementation of the bounded mapping (Hecht-Nielsen, 1990):

$$f: v \subset R^n \rightarrow R^m \qquad (4.1)$$

from a bounded subset v of n dimensional Euclidean space to a bounded subspace $f[V]$ of m dimensional Euclidian space (usually $n > m$).

The special case of such mapping when m = 1, constitutes the classification problem. Usually it is possible to define a probabilistic measure on V and write down the probability mixture model in analogy with (1.4).

Of course, for real live problems it is impossible to define non-overlapping division of V corresponding to different categories, but using the examples of mapping action, a Network configuration can be tuned to minimize the misclassification errors near to minimal achievable Bayesian error (1.22).

The net architecture consists of L layers each has K nodes. The first layer consists of N elements that simply accept the components of input vector v and distribute them, without modification, to the all of the nodes of the second layer. The nodes of the second layer (synapses) calculate a weighted sum of all inputs and then transform it to some nonlinear (sigmoid) function. This output is distributed again to all nodes of the third layer, and so on till the output layer with M nodes is reached. The output of a FFNN can be used directly for classification and estimation purposes, as well as for the generation of the control signals.

Note, that FFNN can be regarded as a non-linear combination of several transformation matrices, with entries (denoted as weights) adjusted in the training phase by a least square minimization of an error function. The analytic form of the final transformation is very difficult to write down, due to sophisticated hierarchical multiplication of weights and nonlinearity's. Therefore we'll treat f function rather as an operator (algorithm), and not as analytic function, whose parameters have to be fitted.

The usage of more general structures of input-output relations is the main advantage of FFNN over classical regression analysis. The Neural Networks are free of dependency on linear superposition and orthogonal functions and provide a significant increase in ability to approximate a function. The theoretical proof of it is the Kolmogorov's theorem (Kolmogorov, 1957), concerning the ability of representation of an arbitrary continuous function with three layered Neural Network.

The basic computing element in a FFNN is a node (formal neuron). A general i-th node receives signals from the outputs O_i^j of the all neurons of the previous layer:

$$IN_i^{l+1} = T_i + \sum_{j=1}^{NODES(l)} W_{ij}^l \times O_j^l i = 1, NODES(l+1), l = 1, L-1 \qquad (4.2)$$

where the threshold T_i and connection strengths (weights) are W_{ij}^l parameters associated with the node i, I is the layer index, L is the total number of layers, $NODES$ (l) is the number of neurons in the l-th layer and O_i^j is the output of the j-th neuron in l-th layer. The index j corresponds to the higher layer and the index i to the next layer. The total number of searching net parameters (weights) equals:

$$NTOT = \sum_{l=2}^{L} NODES(l) + \sum_{l=2}^{L-1} NODES(l)\, NODES(l+1) \tag{4.3}$$

The output of the neuron is assumed to be a simple function of its input; usually it is formed by sigmoid function:

$$O_i^l = \frac{1}{(1+e^{-IN_i^l})}, i = 1, NODES(l), \quad l = 2, L \tag{4.4}$$

where IN_i^l is the input of the i-th neuron in the l-th layer.

The multidimensional feature vector entering the first layer is translated from input through hidden layers to the output nodes. Therefore FFNN provides the mapping of a complicated input signal to the class assignments.

For classification purposes this mapping takes a special form with aim to "shift" different classes of TS from each other as much as possible.

Therefore the "goal" output O^{goal} (k) for events of the k-th category could be chosen as follows:

$$O_k^{goal} = \frac{k-1}{K-1}, \; k = 1, K \tag{4.5}$$

where K is total number of classes. Of course, it is possible to define another set of "goal" values.

In the case of two classes, i.e. signal and background events, the "goal" outputs, as one can easily see, are equal to zero and one. The actual events classification is performed by comparing the obtained output value with the "goal" one. We expect, that the data flow passing through the trained net will be divided in two clusters concentrated in the opposite regions of the (0, 1) interval. Choosing an appropriate point in this interval (the so-called decision point c), controls the relation between two kinds of misclassification committed by decision rule), the classification procedure can be defined: an event with an output greater or equal than the decision point is attributed to the background class, while all the other events - to the signal class:

$$O(u)\begin{cases} < c \rightarrow u \text{ is classifed as signal} \\ \geq c \rightarrow u \text{ is classifed as background,} \end{cases} \tag{4.6}$$

where O(u) is the output node response for a particular experimental measurement u.

This decision rule is a Bayesian decision rule, therefore the output signal of a properly trained feed forward neural net is an estimate of the a posteriori probability density (Ruck et al., 1990) .

For the multi-way classification one can define a set of non-overlapping bounded intervals

in (0-1) for each category. This sequence of bounded non-overlapping sets $O_k, k = 1, K$, along with the chosen "goal" values (located within corresponding subset), will determine the mapping into the K class labels:

$$O(u) \subset O_k \rightarrow u \text{ belongs to k}^{\text{th}}\text{category.} \qquad (4.7)$$

The objective (error) function to be minimized is simply the discrepancy of apparent and target outputs over all training samples (so called classification score):

$$Q = \sum_{k=1}^{K} \sum_{j=1}^{M_k} \omega_k \left(O_k^j - O_k^{goal}\right)^2, \sum_{k=1}^{K} \omega_k = 1. \qquad (4.8)$$

where O_k^j is the actual output value for the j -th training event, belonging to the k -th class, and the O_k^{goal} is the target value for the k -th class output, where K is number of categories, and M_k is the number of examples for the k^{th} class.

The ω_k weight coefficients control the "contribution" of each particular class of TS to the overall error function. For the primary cosmic ray nuclei type identification by EAS observables, usually intermediate nucleus with masses between the lightest (proton) and heaviest with significant abundance (iron) are trained much worse compared with edge classes.

Among ANI training modes there are two possibilities of checking the classification accuracy of middle categories. First of all we can enlarge the category acceptance region O_{middle}, (*a posteriori* solution) (4.7), and second - the corresponding weight value in error function could be enhanced before starting net training (*a priori* solution) (4.8). This two possibilities provide flexible tools for obtaining desired balance between signal acceptance and background rejection, between purity and efficiency.

4.2 NN LEARNING PROBLEM

Currently there is no universally accepted theory of the predictive learning. Statistical learning theory, developed by V. Vapnik (Vapnik, 1979), based on theoretical analysis of Empirical Risk Minimization (ERM) is a theory for nonparametric dependency estimation with finite data. The Vapnik-Cervonenkis (VC) theory drives necessary and sufficient conditions for consistency of generalization from finite set of examples. The generalization ability of a learning algorithm depends both on the possibility to find the particular function describing the examples, and on the measure of the complexity of the used family of approximated functions. Classical notions of complexity (number of free parameters or degrees of freedom) fail to account for applications to the functional families with infinite number of members, like Neural Networks models. The VC theory, introducing the so-called VC dimension (capacity) as a measure of complexity, can deal and specify functional families ever with infinite number of members. Therefore the VC theory provides conceptual framework for the setting bounds for model complexity control. The issue of the model complexity control is of crucial importance for the practical application of learning algorithms. Of course, there is still much work needed to bridge the gap between theory and practical applications. However, such empirical approaches for selecting the optimal Network solution to avoid overfitting problem, as prediction risk

estimation and median networks committee could be addressed in rigorously defined mathematical scope. Neural models capture the statistics of processes directly from data vectors – collection of "pseudo-experimental" variables, corresponding to all significant variations of the model input parameters.

Neural methods are universal and can deal with very big input vectors. A common complaint about nonparametric techniques is the dependence of the results on the purity and finiteness of training sets (small training samples effects). However, due to the inherent robust characteristics of Neural Network (generalization ability), the results from neural analyses are relatively insensitive to modest impurities in the training sets.

4.3 SETTING OF THE LEARNING PROBLEM

The overall scheme of learning from examples can be defined as following (Vapnik, 1999):

1. random event generator, drawn independently from a fixed but unknown distribution mixture;
2. a supervisor (absolute decision rule) that returns an output vector for every input vector, according to a unknown, but also fixed conditional distribution function;
3. a learning machine (algorithm) capable to implement a number (may be infinite) of different approximation functions.

The problem of learning is that of choosing the appropriate set of functions, and then particular member of this family, which predicts the supervisor's response in the best way (optimal decision rule). The selection is based on the training set (sample), of independent and identically distributed observations presented to the supervisor.

Our fundamental assumption will be that we can generate examples of mapping $\boldsymbol{f}(V)$ by detailed Monte-Carlo simulation of multidimensional random variables $\boldsymbol{u}_i, i = 1, M$ in accordance with assigned probability measure on ¿ initial parameters space.

A primary advantage of mapping networks over classical statistical analysis methods is that the FFNNs have more general (algorithmic) functional forms than classical statistical methods can effectively deal with (Hecht-Nielsen, 1990). The FFNNs are free from depending on linear superpositions or orthogonal functions and can mimic sophisticated stochastic mechanism whereby the Nature generates the data. Therefore, in contrast with classical classification problem, we've to specify not the particular member of known analytic family of functions rather the non-parametric algorithm (classificator) which generalizes the unknown mapping rule, implementing learning strategy on the training sample. The classification learning strategy will be based on the fundamental notion of the generalization.

The strategy connected with the Prediction Risk estimation that reuse data and gives unbiased estimate ever for small sample sets, is generalization of one-leave-out-for-the-time estimate.

4.3.1 NEURAL ESTIMATION (LEARNING REGRESSION FUNCTION)

In the first section of chapter we considered the classification mode of the neural mapping (4.1). The recovering of the unknown functional dependence is another realization of neural mapping possibilities implemented by FFNN. This problem is of vital importance for EAS experiments for solving one of two interconnected problems - energy spectra and mass composition of the primary CR flux. And if the classification statistical model is appropriate for

mass composition studies, the learning of regression function is more appropriate for energy estimation.

Of course, we can define the mass estimation also as regression model and, visa-versa, the energy estimation - as classification of energy into preselected bins. Whereas, we have to keep in mind that this division is in some sense arbitrary.

Our fundamental assumption will be that we can generate examples of mapping $f(v)$ by detailed Monte-Carlo simulation of multidimensional random variables $\boldsymbol{u_i}$, $i = 1, M$, in accordance with assigned probability measure on I initial parameters space (1.3).

The bounded subspace of I, e.g. simulated primary energies, will be determined by the installation threshold and acceptance and flux intensity (the information on the steeply fallen energy spectra will also be incorporated in preparing of the simulation trials).

The most common drawback in FFNN performance is the limited number of training and test samples. Usually, in CR physics applications we can't simulate enough simulation trials, especially for most interesting ultra-high energies. And, therefore, we never can be sure that we use sufficient number of examples to learn a general problem and not the specific training data set.

As we are not sure that training samples used reflect all variability of physical processes, learning of specific selected examples too well is not desirable. What we need is to generalize from the used training set to entire problem.

Therefore, the strategy, checking the expected performance of FFNN during training is of crucial importance.

The strategy, proposed in (Barron, 1984) is connected with the prediction risk as performance measure. In general, particular FFNN model can be specified (indexed) by the λ parameter:

$$\lambda \subset \Lambda \equiv (V, G, W) \qquad (4.9)$$

where $V \subset v$ notes a chosen subset of variables from the set of all possible variables (measurements) *V;*
G is a selected architecture from the class of possible architectures *G*; and *W* is the set of net parameters (weights).

The prediction risk *P(λ)* is defined as expected net performance on a finite test set:

$$P(\lambda) \approx E\left\{\frac{1}{M}\sum_{i=1}^{M}\left(t_j^* - O_\lambda(u_j^*)\right)^2\right\} \qquad (4.10)$$

where $\left(t_j^*, u_j^*\right)$ weren't used in training, O_λ is the trained network output.

The strategy exists in the selection of the particular λ from the model space Λ, which minimizes an estimate of the prediction risk.

The procedure of the prediction risk estimation that reuse data and gives unbiased estimate ever for small sample sets, is connected with the generalization of one-leave-out-for-the-time estimate used in previous chapter for Bayes risk estimation (1.25). The k-fold cross- validation, introduced by Geisser (Geisser, 1975) and Wahba (Wahba, 1990), instead of leaving only one event, deletes larger subsets from training sample. Let the training sample $\left(\tilde{P} \equiv (t_j, u_j), j =\right.$

$1, M)$ be divided into k randomly selected disjoint subsets of the equal size $M_k = M / k$, denoted by $\tilde{p}$. And the $\tilde{P}_i$ will denote the training sample with deleted i — th subsample $\tilde{p}_i$.Then the cross-validation mean square error (MSE) for the selected subset $\tilde{p}_i$- is defined as:

$$MSE_{\tilde{P}_i}(\lambda) = \frac{1}{M_k} \sum_{(t_j,u_j)\subset\tilde{p}_i} \left(t_j - O_{\lambda,\tilde{P}_i}(u_j)\right)^2 \tag{4.11}$$

and

$$MSE(\lambda) = \frac{1}{k}\sum_{j=1}^{k} MSE_{\tilde{P}_i}(\lambda) \tag{4.12}$$

Typical choices of k are 5 and 10. An useful modification of cross-validation mean square error, penalizing complicated networks comprising many hidden units, is the Akaike's final prediction error (Akaki, 1970). For large enough training sets it takes following form:

$$P(\lambda) \equiv MSE(\lambda)\left(1 + 2\frac{NTOT}{M}\right) \tag{4.13}$$

where NTOT is total number of networks weights (4.3). Just this expression is recommended by authors of (Barron, 1984) as an estimate of prediction risk. Estimates of the prediction risk offer a sound basis for assessing the generalization performance of the model and can be used as a tool for architecture selection and constructing the stopping rule. Therefore, it is important to check the training results not with the "training error", but with the "generalization error", represented by the prediction risk.

Below we consider the scalar regression function case, but all the results obtained can be used also for the vector regression function learning. FFNN with appropriate defined error function was used for the simultaneously primary energy and mass estimation. The following function have to be minimized:

$$Q = \sum_{j=1}^{M} \omega_j\, g(t_j - O_j), \sum_{\mathrm{j}=1}^{\mathrm{M}} \omega_j = 1 \tag{4.14}$$

where O_i is the vector output of the FFNN last layer (no sigmoid function is implemented for the nodes in the output layer!) and $\boldsymbol{t}_i$ is the vector of parameters used in simulation (primary mass and energy of "pseudo-experimental" event), and ω_i highest energy events get higher weights). is the event weight (usually the highest energy events get higher weights).

A weighted quadratic metric is used as a measure of discrepancy of actual and "true" regression function values:

$$g(.) \equiv \alpha\left(\overline{mass}(u_j) - mass(u_j)\right)^2 + (1-\alpha)(\hat{E}(u_j) - E(u_j))^2 \tag{4.15}$$

The α coefficients are changing during training cycles to provide stable and reliable recovery of both energy and mass. By setting a value to 0 and 1 we have possibility to estimate first energy, and then, grouping events in energy bins - estimate the primary mass. And visa-versa, first estimating the mass, then estimate the energy spectra for different species of cosmic ray flux.

4.4 TRAINING ALGORITHMS

The only information to train network for nonlinear mapping is contained in a priory given pairs - $(t_i,u_i),i = 1,M$, where M is the number of training events.

During the minimization procedure the calculated differences between the actual network output and the desired output are used to adjust the weights.

The back-propagation (BP)(Rumelhart et al., 1986) algorithm of neural network training is one of the most important historical developments in neurocomputing. The simple rule (based on gradient descent) of weights updating after processing of one or more training examples in principle will lead to arbitrary small mean square error of function approximation. The family of BP algorithms is realized in numerous packages, with the JETNET package being most popular in HEP community (Petersen et al., 1993).

Minimizing of the eq. (4.8) with gradient descent is done in (Rumelhart et al., 1986):

$$W_{t+1} = W_t + \Delta W_t \qquad (4.16)$$

where

$$\Delta W_t = -\eta \frac{\partial Q_t}{\partial W} = -\eta \nabla Q_t \qquad (4.17)$$

the learning rate (step size) η is rather difficult to choose appropriately. A good choice depends on the shape of the error function. A small learning-rate will result in long convergence time on a flat error- function, whereas a large learning-rate will possibly lead to oscillations, preventing the error to fall below a certain value.

Another problem with gradient descent (Riedmiller et al., 1993) is the "contra-intuitive" influence of the partial derivative on the size of step. If the error-function is shallow, the derivative is quite small, resulting in a small step. On the other hand, in the presence of steep ravines in the energy landscape, where cautious steps should be taken, large derivatives lead to large steps, possibly taking the algorithm to a completely different region of weight space. An early idea, introduced to make learning more stable, was to add a momentum term

$$\Delta W_{t+1} = -\eta \frac{\partial Q}{\partial W} + \alpha \Delta W_t \qquad (4.18)$$

where momentum parameter a < 1 scales the influence of the previous step on the current one. The local minima can to a large extent be avoided by introducing random noise σ to thegradient descent updating rule of eq.(4.17). This is conveniently done by adding a properly normalized Gaussian noise term

$$\Delta W = -\delta \nabla Q + \sigma \qquad (4.19)$$

which we refer to as Langevin updating, or by using the more crude non-strict gradient descent procedure provided by the Manhattan (Peterson et al., 1989) updating rule:

$$\Delta W = -\eta \cdot \mathrm{sgn}\left[\frac{\partial Q}{\partial w}\right] \qquad (4.20)$$

It should be noted, that elimination of the harmful influence of the size of partial derivative on the step and replacing the step by a constant update value 8 forms a new BP family member the "Resilient back-propagation", or "Rprop" (Riedmiller et al., 1993A).

The basic principle of Rprop is the direct adaptation of the weight update values δ individually for each neuron (connection). Rprop modifies the size of step directly by introducing the concept of resilient update-values, independent on the value of the partial derivatives.

The theorems, proving the consistence and unbiasness of the BP training algorithm is the existence theorem only, it guarantees the ability of a multilayer network with the correct weights to unbiased approximate an arbitrary function, it does not demonstrate how to learn this weights using available training samples.

Real world problem solving with BP net training (as with any NN training algorithm) requires several input parameters whose values are to be assumed.

First of all it is the net architecture: number of hidden layers and number of nodes. It is important to realize that although the Kolmogorov's theorem proves that "one hidden layer is enough", in real-world problem solving, it is often essential to have two or three hidden layers. If we limit ourselves with single hidden layer an impractical huge number of a nodes would be required, whereas an adequate solution can be obtained with a tractable network size by using more than one hidden layer. Also the initial range of weights must be specified and size of the momentum term should be selected.

The choice of step size and the number of training epochs on learning before weight updating influence heavily the quality of obtained solutions. For instance the possibility of getting trapped in a local minimum is enhanced for a wrong net initialization. The problem of restricting the parameter space to a region which ensures convergence to the global minimum remains thus to a large extend unsolved (Sinkus, 1995).

The evolutionary algorithms, introduced in the next section seem to be more appropriate for the search of the "best" solutions in an extremely large spaces for the wide range of underlying error-function profiles.

4.5 EVOLUTIONARY ALGORITHMS

Genetic Algorithms (GA) and Evolutionary Programming (EP) are both search techniques based on an simulation of the evolutionary processes. The challenge is to find "good solutions" (chromosomes) in very large search spaces.

GA employ the successive reproduction among an assembly (pool) of fittest parents using genetic operations such as crossover, inversion, mutation and selection with predefined rules forming the next generation and the search stopping. Different m:n scenarios (m - number of parents, n - number of offsprings) can be realized.

EP can be treated as a special case of GA, namely 1:1 scenario.

The current best chromosome (parent) undergoes the zero-mean phenotypic Gaussian mutation (realized by the Random Search (RS) algorithm with return at an unsuccessful step). The solutions obtained are very much dependent on the initial population. It is not easy to decide if the GA can be started from a completely random initial pool (like EP), or it is worth to spent significant efforts to find "acceptable" solutions using more "simple" methods like EP or BP.

During the search most parameters had to be determined empirically. The relationship between tuned search parameters usually is not clear and the strategy of parameter selection turns out to be a nonlinear multivariate optimization problem itself. Also it is very difficult to understand how a certain solution is reached and explain why it is the "best" one.

The multidimensional error surface is very complicated and it is very hard to visualize it and

acquire any exact information about it. A basic fact is that there exist many global minima. This is because there are many weight permutations that yield exactly the same overall network input/output relation (Hecht-Nielsen, 1990). The same words could be said about the local minima. Beyond these facts, very little is known. The following questions still wait answering:

How large are the basins of attraction for each of different minima?

What is the shape of the boundary between attractive basins?

We perform the following strategy of the development of net training technique:

- train networks of different architectures with various methods and then use all information to integrate acquired knowledge in the final "best" solution;
- decrease as much as possible the number of free parameters of network;
- use procedures of atomization of parameter tuning;
- use prediction risk as net performance measure;
- develop methods of visualization of training procedures to explore the multidimensional error surface.

The neuron mode of ANI (Analysis and Nonparametric Inference, nonparametric statistical analysis package developed in Cosmic Ray Division of Yerevan Physics Institute) (Chilingarian, 1998) package was implemented for crossover modeling in net training. First, the number of particular net node is randomly chosen and then from an "infinite pool" of genes a new gene is selected and introduced at the same place in the net. If the error function decreased, the new chromosome (collection of genes) survived and a new random step is performed. If not - the selected gene is excluded, and another attempt is performed. If we remember that each gene (neuron, node) in the net outlines some part of nonlinear decision surface in multidimensional feature space, it will be easy by combining such piece-wise surfaces arrive to a better decision boundary separating different classes from each other.

This kind of net training seems to be much more effective than simple Random Search (RS) algorithms, also realized in ANI. The multi and single modes of ANI are designed for random search correspondingly in all net parameter space and - to make random change of also randomly chosen net parameter. Different net training scenarios combine different search modes with different search parameters.

For fast scanning of the net weights space a deterministic algorithm is implemented. The error function is calculated in each point of the multidimensional quasi-random sieve (Sobol, 1979) uniformly filling the JV-dimensional cube (N is the total number of NN weights). Positioning the sieve center at the previously found best point, and subsequently decreasing sieve size, we'll arrive to the "best" net.

The different strategies of RS algorithm implemented in ANI will be described in more details in section 4.8

4.6 COMMITTEE OF NETWORKS

4.6.1 ENSEMBLES OF NETWORKS

The alternative to genetic algorithms approach of integration of different obtained networks in one "best" solution, is the networks "committee" or "ensemble" approach.

This approach consists in using not a single network, but an ensemble of networks, each of which have been trained with different methods and, may be, with different data-sets. The basic idea is to classify a given input pattern by obtaining a classification (or estimation) from each member of committee and using an integration scheme make a final estimate with some collective decision strategy (Hansen et al., 1990). The most important is to understand the "committee" principle of integration of information acquired from different networks trained with various parameters.

It is common practice in the application of neural information technologies to train many different candidate networks and then to select the best, (on the basis of performance on an test

data set) and to use further only this network.

There are three disadvantages with such an approach. First of all, if we select the best network among many trained, all of the efforts involved in training the remaining networks are wasted.

Second, the network with best performance on training data set can be over-trained, and third, even if the over-training effects are controlled, the generalization error (of particular network from all available) on the test data might not be the smallest. Different networks trained on the same training data set but with different training scenarios contain different information on the general nature of considered phenomenon, combination of these networks can give more generalized representation of these data than the single "best" network.

Usually this procedure is called a committee of networks. In simplest case, when all networks are equally contributed, i.e. the final output is the average of all outputs of different networks, it is called a mean committee.

Another type of committee - the median committee takes as final estimate the median value of variation row of single outputs obtained with different networks.

Using the committee of networks approach and the simulated data we demonstrate that the reduction of estimation error is generally obtained even in case of small number of committee members. If there are several trained networks, to find the best one for a problem solution, it is always sensible to make a committee of that networks. Therefore future investigations will concern the synthesis of different networks trained with applying various scenarios and methods of global minimum reaching.

4.6.2 THE COMMITTEE PROCEDURES

There are many different possibilities to make a committee of networks. We'll consider two types of integration - the averaging and the median committees. The first procedure is a simple averaging of all neural networks outputs and taking this mean as a final estimate. In ideal case, when the distribution of error of each trained network has zero mean (unbiased training) and all errors of different networks are uncorrelated, it was shown that even for the simplest averaging committee can improve the overall performance: $E_{com} = E_{av}/L$ is, where E_{com} is the committee error, E_{av}- averaged error of all networks and L is the total number of different networks (Bishop, 1995). In practice, the reduction of estimation error is much smaller because the errors are highly correlated. However, it is easy to show that the committee process cannot produce an increase of the expected error $E_{com} \leq E_{av}$.

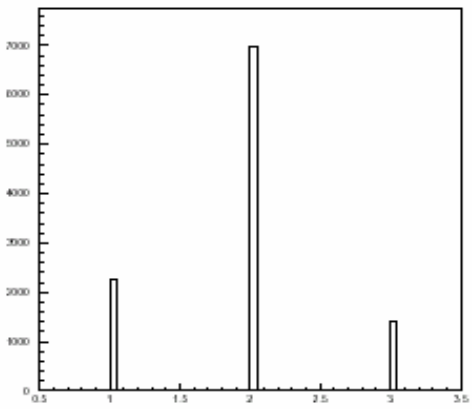


Figure 7. Contributions of different networks to final estimate (3 nets)

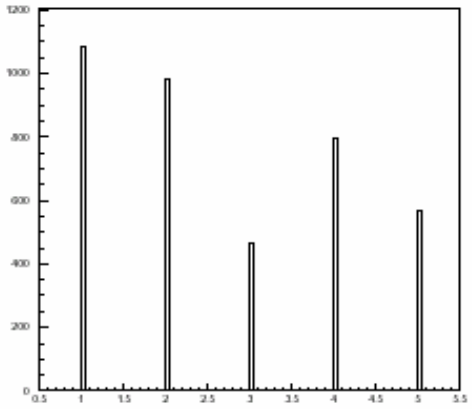


Figure 8. Contributions of different networks to final estimate (5 nets)

We might expect that some members of the average committee will make better predictions than

other members. We would therefore expect to be able to reduce the error still further if we give greater weight to some committee members than to others. But the better prediction on training sample does not mean better prediction on control sample, so we do not know to which members to give greater weights. To automate the optimal weighting of neural networks we use the ordered statistics technique and call that procedure as a median committee, in analogy to the nonparametric estimate of the adaptive multivariate probability density function, introduced in (Chilingarian, 1989).

This procedure can be explained as follows: trained networks are implemented for the independent (control) data set and then the output values of all different networks are ordered in increasing (or decreasing) sequence and the middle (median) value of this sequence is taken as an final estimate. If L is an odd number the median of variation row is selected, if L is an even number, the half sum of two middle outputs is taken.

In this case again only one network's output is selected, but for each particular event as a median will be selected each time different network from all available networks. So, all networks are implemented in estimation process, but each of them as many times as its output becomes the median of variation row of all outputs.

The median committee looks somehow like weighted mean committee, because each network has a different contribution to the generalized final output. The problem how to give the weights to networks is avoided, because the predictions are automatically weighted according to the median choosing. For each individual event the networks giving larger and smaller outputs (placed before and after the median value) are weighted with 0. And the proportion of "selected" networks, calculated over the control sample gives the overall weight (quality) of different networks (Figure 7, Figure 8).

Thus, the main attention will be paid to the median committee procedure.

4.6.3 RESOLVING OF THE MIXTURE OF ANALYTIC MODELS

For the demonstration of the advantage of committee estimator comparing with the single best network the combined regression function is used, which consist of three different analytic models. The regression function in each model is a certain distribution of means of the five dimensional Gaussian populations:

$$f(\vec{x}_j) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{\vec{x}_j - M_j}{\sigma}\right)^2} \tag{4.21}$$

where x_j is the input vector, $x = [x_1, x_2, x_3, x_4, x_5]$.

In the first model M_j is distributed in 0 — 5 interval according to exponential low, in second and third models it is distributed by Gaussian and power lows respectively:

$$M_j(1) = \ln\left(\frac{1}{e^{-b_2}} - RNDM_j(e^{-b_2} - e^{-b_1})\right) \tag{4.22}$$

$$M_j(2) = N(\mu, \sigma) \tag{4.23}$$

$$M_j(2) = N(\mu, \sigma) \tag{4.24}$$

where a = 2, b_2 = 5, b_1 = 0, μ = 2.5, σ = 1, j = 1, 10000

So, we have three different training data sets, each of them represents the general data set only partly. This is a common case, when we are forced to use a Monte-Carlo simulation of some complicated phenomenon, when no *a priory* information can be assumed to describe exactly the underlying multidimensional probability distributions. Three different networks trained by these three models are applied to estimate the true regression function of the models mixture, and the results are compared with the committee results.

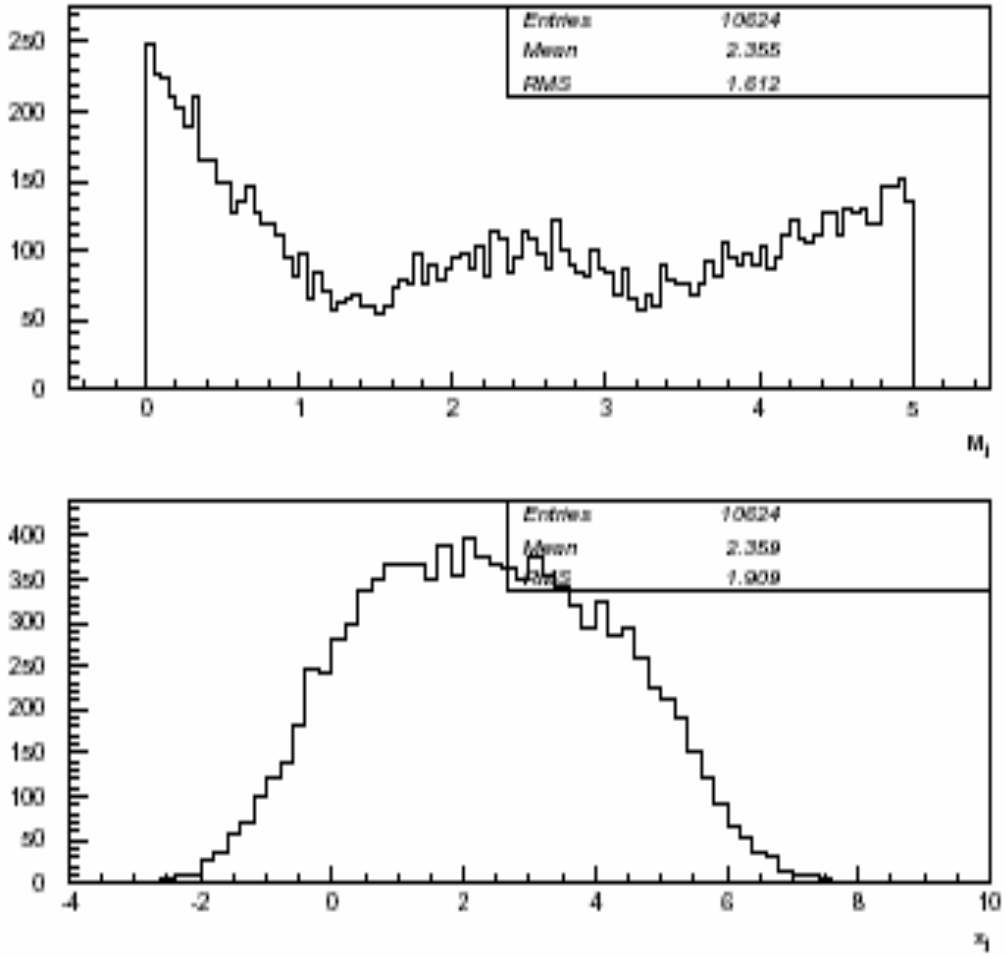


Figure 9. Distribution of regression function and input parameters (mixture of models).

The statistical model to be considered is the mixture of these three models: in 0 -1.7 - exponential distribution 1.7 — 3.3 - Gaussian distribution and 3.3 — 5 - power distribution [eq.(4.25)], as shown in Figure 9.

$$M_j = \begin{cases} \ln\left(\frac{1}{e^{-b2}} - RNDM_j(e^{-b2} - e^{-b1})\right); for 0 < M_j \le 1.7 \\ N(\mu, \sigma); for 1.7 < M_j \le 3.3 \\ RNDM_j^{\frac{1}{\alpha+1}}(b_2 - b_1) + b_1; for 3.3 < M_j \le 5 \end{cases} \quad (4.25)$$

On the Figure 10 the estimation results by three different networks are plotted. It is easy to

see that even for the best network there is significant bias of the estimates. Comparing these results with the ones depicted in Figure 11, where the estimation errors for different networks and committee are plotted, one can see that the committee procedure leads to unbiased estimation with higher accuracy.

From Table 5 one can see that the reduction of estimation error is about 10% and the median estimator gives better result than mean committee.

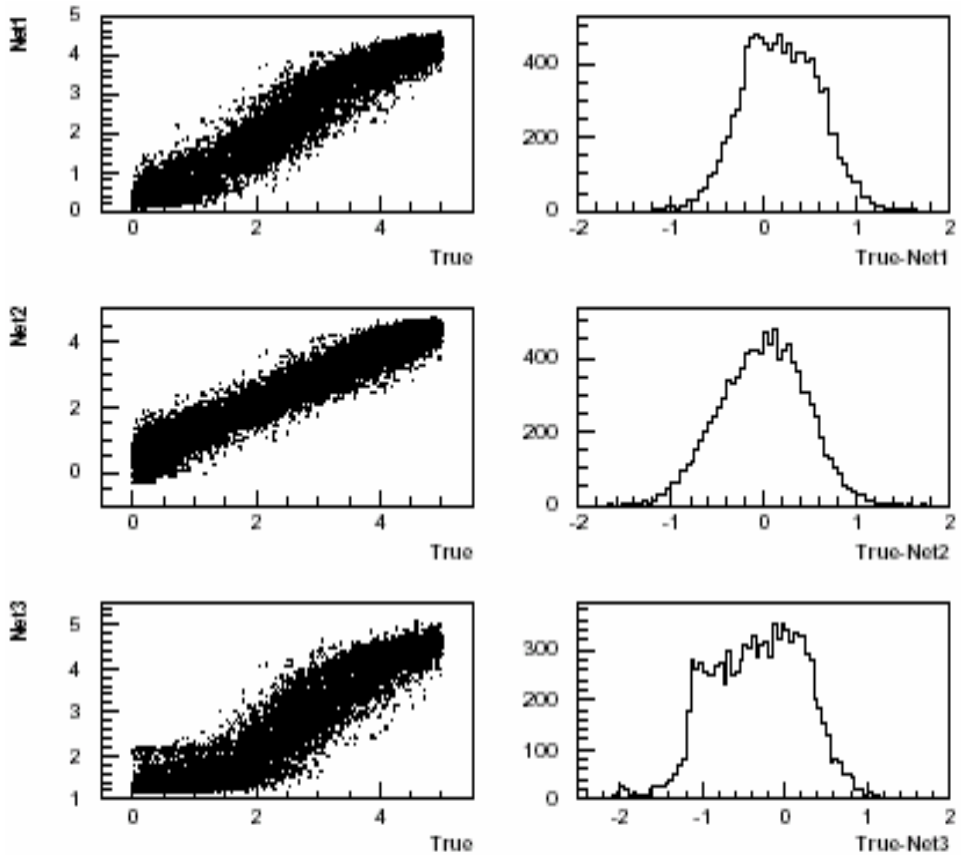


Figure 10. The regression function estimation by different networks trained on different models: 1 - exponential, 2 - Gaussian, 3 - power

Table 5. Root mean square errors of 3 different model networks and committee

	Net1	Net2	Net3	Median	Mean
RMSD	0.4536	0.4584	0.5688	0.4168	0.4455

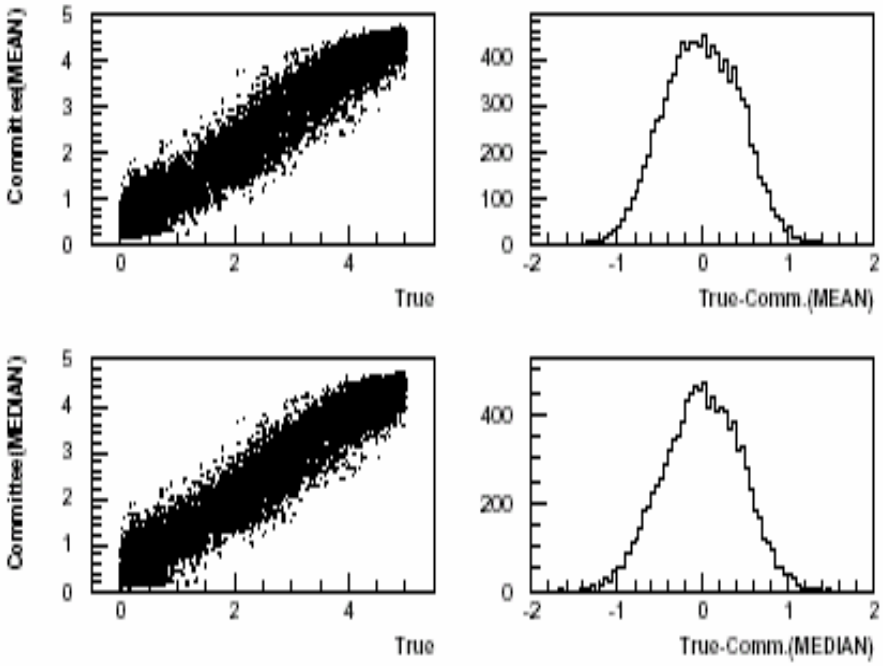


Figure 11. Mean and median committee results

4.7 ESTIMATION OF THE GENERALIZATION ERROR IN NEURAL CLASSIFICATION TO MULTIPLE CATEGORIES

4.7.1 NET TRAINING AND GENERALIZATION ERRORS

The most common drawback in Feed-Forward Neural Networks (FFNN) performance is the limited number of training and test samples. Usually, in many NN applications we do not have enough simulation trials, therefore, we never can be sure that we use sufficient number of examples to learn a general problem and not the specific training data set.

As we are not sure that training samples used reflect all variability of the considered phenomenon, learning of specific selected examples too well will effect the so called over-training, when the NN performance on training sample is much better than on the independent (test) sample. Therefore, it is important to check the training results not with the "training error", but with the "generalization error", represented by the prediction risk.

Estimates of the prediction risk offer a sound basis for assessing the generalization performance of the model and can be used as a tool for architecture selection and constructing of the stopping rule.

To demonstrate the over-training effects we have trained a rather complicated network (with 2::4::3::1 architecture) for the two-way classification of a few (100 events per class) events. The 2 dimensional samples from Gaussian populations N(0,l) and N(l,l) were used. The corresponding Bayes Risk equals RB = 0.2415.

After reaching Bayes risk value on 1000-th iteration, continuation of training leads to the "over-training". The following decrease of error function is due to the tuning of the decision boundaries and reflects only the peculiarities of the finite random sample, and not, required better approximation of the underlying probability density functions. Therefore over-training doesn't improve the "generalization" abilities of trained network.

This fact is demonstrated by the classification with obtained networks, of much bigger control sample (2000 events from each class). As it is seen from Figure 12 error function for

control sample (generalization error), didn't follow the decrease of training error after reaching the Bayesian limit. Moreover, it abruptly increase on the minimum of training error, proving that the over-training results in random net couplings, unable to approximate general rule.

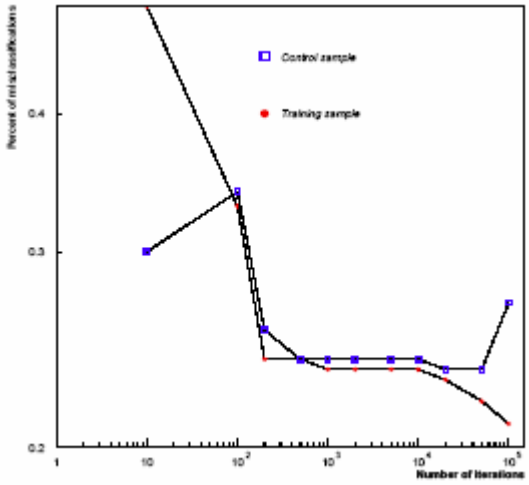


Figure 12. Training and generalization errors

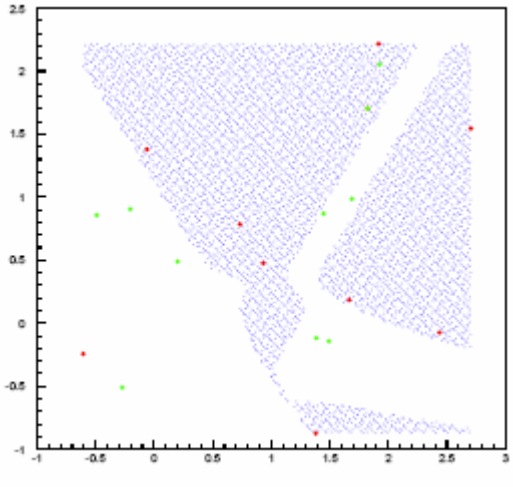


Figure 13. Classification boundaries for over-trained network

The demonstration of the embedded discriminative boundaries, obtained with over trained network is shown on the Figure 13. The net of the same configuration was trained with only 20 events (10 from each class, depicted with the different colors on the figure).

As one can see, only 3 events are misclassified, therefore the error function takes the value 1.15 (which is less than RB=0.24).

Continuation of the training will shift the decision boundaries to decrease the misclassification down to 0, but of course, the generalization capability of such network will be very poor.

4.7.2 THE CROSS-VALIDATION PROCEDURE FOR FINAL PREDICTION ERROR (FPE) ESTIMATION

Described in section 2.7 technique with appropriately defined error function was used for the training of NN with different architectures for the 3-way classification problem. The FPEs of these networks were estimated for different sample sizes. The aim was to estimate the generalization capabilities of different networks for different sample sizes and to select an "optimal" configuration of NN for the given training sample size. The considered problem was to classify the random events from two-dimensional Gaussian populations with means equal to 1, 2 and 3 for the first, second and third classes respectively:

$$f(\vec{x}_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{1}{2}\left(\frac{\vec{x}_j - \mu}{\sigma}\right)^2} \qquad (4.26)$$

where x_j is the input vector, $\vec{x} = [x_1, x_2]; j = 1...50000$, $\sigma = 1$; for all three classes.

Six different samples with different sizes were used to train the five different NN for the 3-way classification. Particularly, samples containing 100, 200, 500, 1000, 2000, and 5000 events per class were used. Five different NN architectures were constructed as follows: 2 input nodes and one output node for all configurations, and the number of hidden nodes varying from 2 to 10

nodes. The generalization capabilities (FPE) of NN with different architectures were estimated for these sample sizes and compared with the training error.

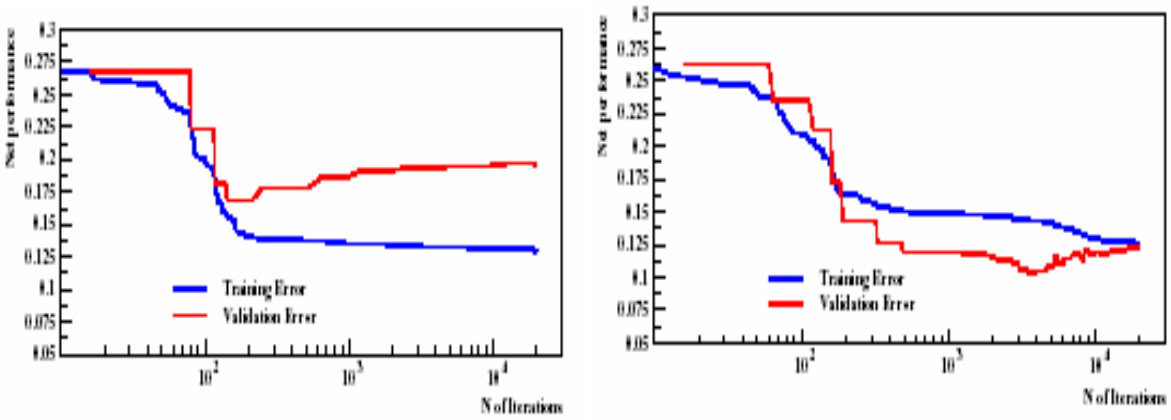


Figure 14. The training process of the NN with architecture 2:2:1, TS consists of 300 events (30 validation events)

For large enough training sets the calculation of the net performance by the one- leave-out- for-a-time method can be much time consuming, therefore we have used a 10 — fold cross-validation for each sample size. Particularly, we use 10% of events as a validation sample (VS) and remaining 90% as a TS. After the training process is finished (the maximum number of iterations is reached or some stopping criteria is applied), the removed 10% of events are returned to the TS and the next 10% of events are taken for the validation. So, each sample is reused for the training process 10 times until all events of the sample are subsequently removed and used for validation.

Note, that in each trial of the cross-validation the VS events are different. The MSE for each VS is calculated by the formula (4.11), and the expected MSE for the given sample is calculated by formula (4.12) for both training and validation samples. Then the formula (4.13)

is used to calculate the PR of NN on validation sample, this is used as the Final Prediction Error.

Figure 14 and Figure 15 display the training error evolution for different sample sizes. From Figure 14 one can see that when the TS size is very small the over-training occurs even with the smallest NN architecture (the training error is decreasing during further training, while the error on independent validation sample is reached some local minima and then is increasing). Therefore for very small TS one could not expect to obtain good generalization, because the net weights are easily tuned to classify rather well the small number of training patterns. In such cases, the NN should not be trained with many training epochs (presenting the training patterns to the NN many times). The question, when to stop the training process, is still open, it depends on the particular problem to be solved, sample size and NN architecture.

As it is easy to see from Figure 14, using the same sample and the same NN architecture, in different stages of the cross-validation procedure the over-training has occurred in completely different ways in terms of number of training iterations and training and validation errors.

One possibility is to stop the training when the validation error starts to increase (or this increase is larger than some threshold value), but this criteria is not desirable due to the random character of the overtraining effect. As it is demonstrated on the Figure 15, the random increase of the validation error is occurred rather often. The curve of the validation error has many local minima, but during further training the error is decreasing and the best point in multidimensional space of the NN weights is obtained at the end of training process.

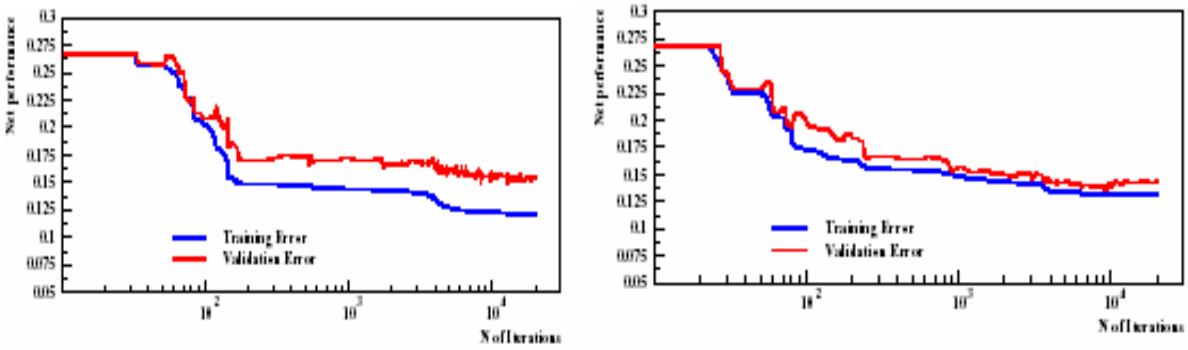


Figure 15. The training process of the NN with architecture 2:6:1, TS consists of 600 (above part) and 1500 (bellow part) events (VS - 60(150) events).

The acceptable technique for this purpose is not to apply any stopping criteria, but to apply the following procedure:

- after each successful iteration of the learning process the net error (MSE) is calculated for the test sample;
- if the test error value is less than the one obtained in previous step the NN weights obtained at the current training step are stored;
- else, the NN weights obtained at the previous step are memorized;
- at the end of the training process (when the maximum number of iterations is reached) the weights which gave a minimal error on the test sample are found and used as a final best weights for. NN.

4.7.3 ESTIMATES OF FPE AS AN NET ARCHITECTURE SELECTION TOOL

In previous section the procedure of the FPE estimation was described. This procedure gives an estimate of the FPE of different architectures using the same TS, but does not give any information about the error bounds of the estimated FPE. It is important to obtain not only unbiased estimate of the FPE but to estimate the errors of this estimation as well. With aim of this we have applied the cross-validation procedure for each of the 6 different samples 10 times, of course using 10 different (from the same general population) samples for each size.

So, having 10 different estimates of the FPE for each network architecture we calculate the mean and the standard deviation of FPE.

Figure 16 displays the dependence of the FPE on the used NN architecture for different sample sizes. As one can see from this figure for the smallest sample size (300 training events) the best NN configuration turns to be the net with 2 hidden nodes, because the training error is decreasing with increase of the number of hidden nodes, but the FPE is increasing dramatically. The remaining graphs on this figure show that when the sample size varies from 600 to 3000 the best configuration turns to be the net with 4 hidden nodes.

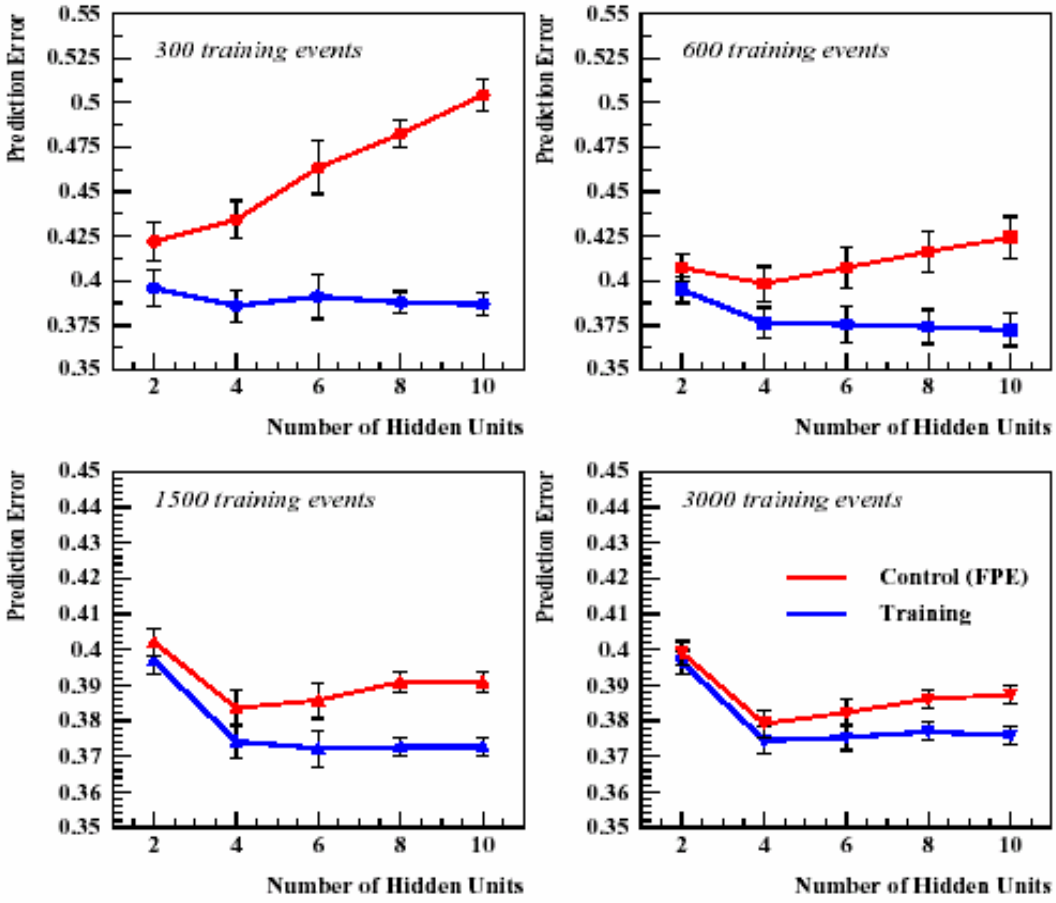


Figure 16. The dependence of the FPE and training errors on the number of hidden nodes in NN for different sample sizes.

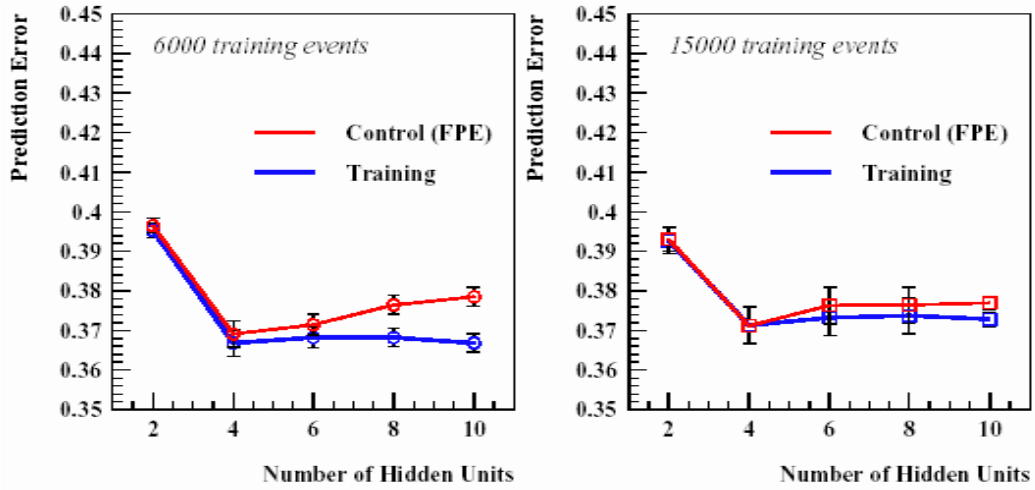


Figure 17. The dependence of the FPE and training errors on the number of hidden nodes in NN for different sample sizes (the case of very large training samples).

On Figure 17 the same dependence as on Figure 16 is plotted for large sample sizes. It is easy to see that the best architecture is again 2:4:1. It is also easy to recognize that the FPE and training error values became closer in contrast with the Figure 16. Although the best FPE is obtained using the net with 4 hidden units, from the right graph of this figure one can see that in case of very large training sample (15000 events) all networks (except the smallest) give very near FPE and within the error bounds all networks have the same performance.

4.8 COMPARISON OF THE DIFFERENT ALGORITHMS OF NN TRAINING FOR THE PROBLEM OF CLASSIFICATION

The Artificial Neural Networks represent a subset of the multivariate statistical analysis methods for function approximation and data classification, and they belong to the general class of non-parametric methods that do not make any assumption on the parametric form of the function to be modeled. These methods contain more free parameters, therefore they require more training data to achieve good generalization performance.

The Net training methods are of wide variety. Each of them is designed to outperform others in such a drawback of the learning algorithms as the speed versus quality.

The choice of the learning algorithm depends on many problem dependent factors, like the complexity of the problem, availability of the training set, real-time requirements, the cost of the losses (required quality of the solution). Hence it is not possible to give an unique decision on what strategy to follow. Nevertheless, the goal of each training algorithm is to achieve a good generalization performance relatively easy and fast (low resources requirements and efficiency of algorithm).

The ANI statistical analysis program package (Chilingarian, 1998) contains a special section of artificial neural network models. Several strategies are implemented for the NN training for estimation and classification tasks. The main algorithm for NN weights updating is the so called Random Search (RS) algorithm, the second one is the deterministic algorithm based on Sobol's pseudo-random numbers.

In the next section a short description of different strategies is presented. In section 4.8.2 the results of comparisons are discussed.

4.8.1 RANDOM SEARCH LEARNING ALGORITHM

The basics of NN learning problem and a description of NN regression and classification can be found in (Bishop, 1995), (Chilingarian, 1995), (Chilingarian et al., 1997A). Thus, we bring only a description of learning algorithms and strategies implemented in ANI package in short. The (RS) algorithm implements the following steps:

1. The initial values of NN weights are chosen randomly from Gaussian distribution with zero mean and small variance ($\mu = 0$; $\sigma \approx 10^{-2}$) ;
2. The random iteration step in multidimensional space of NN weights is performed from initial point to modify the NN weights.

Three submodes of random search are implemented:

- single mode - modifying the single random weight of NN;
- neuron mode - modifying all weights of randomly chosen neuron;
- multi mode - modifying all weights of neural network simultaneously*

The alternation of weights is performed according to following:

$$\vec{V}_i = \vec{V}_{i-1} + Q_{i-1}^{p} * (RNDM - 0.5) * \Delta v; i = 1, N_{iter} \qquad (4.27)$$

where $\vec{V}_i$ is the vector of NN weights obtained at i-th iteration step, Δv is step size and RNDM is a random number from [0-1] interval. The term Q_{i-1}^{p} introduces dependence of value of random step on the already achieved quality function and controls the degree of this

dependence on achieved quality function, p=0,1,2...

3. The objective (quality) function is calculating at each iteration step, by presenting all the training events to Neural Network:

$$Q_i = \frac{1}{M_{events}} \sum_k (OUT_k - TRUE_k)^{speed} * W_k$$
$$k = 1, M_{events}, \quad speed = 1,2 \qquad (4.28)$$

were OUT_k is the actual output of NN, $TRUE_k$ is the corresponding goal value of k-th input vector from training set, and W_k is event weight. If speed=1, Q gives information on the bias of estimation; i f *speed=2*, $\sqrt{Q}$ is the estimation error;

4. If $Q_i \leq Q_{i-1}$, then the vector $\vec{V}_i$ is kept as new weights of NN and next step is initializing from point $\vec{V}_i$ in multidimensional space of NN weights, otherwise-return to the previous point is implemented and a new random step is performed.

The deterministic algorithm works as follows:

1. a multidimensional hypercube with dimension $d=N_{weights}$ and side $A = \Delta v$ is filled in such a way that the projection to any axis gives a uniformly distributed points (number of points in hypercube is equal to the number of given iterations) - so called pseudorandom numbers [90];
2. each "pseudorandom" point in this hypercube represents a vector of the NN weights, the network with such a weights is applied to calculate the quality function on a training sample, and the point which provides the smallest value of the quality function is taken as a vector of NN weights.

Different search modes *(single, neuron, multi)* give a possibility to perform different training strategies, obtaining quite different points of multidimensional space of neural net weights, and allow to investigate symmetries of local minima in this space (due to very large dimensionality of NN weights space there are many symmetries (Hecht-Nielsen, 1990)).

Both algorithms have many free parameters like the number of hidden layers in NN, number of hidden neurons in these layers, objective function to be minimized random step size in *Random Search* algorithm and the size of the hypercube in deterministic algorithm, number of training epochs and number of "pseudorandom" points used in RS and deterministic algorithms respectively. So the comparative study of different learning strategies in terms of speed and quality is of great interest, despite that we can use all methods in one session of training in different combinations. It is also worth to compare the performance of these algorithms with well known and widely used algorithms of *Back Propagation* (BP) (Rumelhart et al., 1986) family. One of the optimized BP algorithms was chosen, known as a *Rprop* (Riedmiller et al., 1993A). This algorithm is known as one of the easiest in implementation and has only a few free parameters like the training rate (step size) and scale-up/down factors, which allow to use an individual learning rate for each iteration. It was reported (Schiffman et al., 1993) that the final results are not much sensitive to the scale factors, thus, the only essential free parameter remaining is the initial learning rate. In the same reference was reported that the *Rprop* outperforms all other back propagation learning algorithms both in speed and quality. The used program package was JetNet version 3.5, in which the *Rprorp* algorithm is implemented as well among many others.

*One of the two submodes of multi mode, the so called deterministic algorithm based on pseudorandom numbers, was used

4.8.2 DATA AND PROCEDURES

It is obvious that in order to compare the performance of the different learning algorithms, it is necessary to apply all of them to the same data and under the same conditions. The latest requirement is, of course, difficult to satisfy, due to the different parameters used by different algorithms, nevertheless it is possible and necessary to fix some free parameters same for all algorithms.

To compare the performance of different learning algorithms, the 3-way classification problem was investigated. The "toy-problem" is selected in some sense our main physical analysis goal - classification of the primary CR flux into three categories (Chilingarian et al., 1997), (Chilingarian et al., 1997A), (Chilingarian et al., 1999B), (Chilingarian et al., 1997), (Roth, 1999).

Or in case of atmospheric cherenkov techniques to the classification of images initiated by 7 quanta, hadrons or local muons. The neural classificators, particularly the *Random Search* algorithms were introduced and successfully applied for such kind of data analysis by A.A. Chilingarian since early 90's. In reference (Chilingarian, 1993) the superiority of neural methods for the 7 ray astrophysics data analysis to other methods was demonstrated. The above mentioned and also the following papers (Chilingarian, 1994), (Chilingarian, 1995), (Chilingarian et al., 1997B), (Aharonian et al. 1990), (Chilingarian, 1991), (Chilingarian, 1994) contain the description of techniques and application details for classification problem of different kind of HEP data. So, besides the very successful implementation of these strategies also their comparative study is of great interest.

Events from 2-dimensional Gaussian populations with means (1,1), (2,2) and (3,3) were classified using NN with the same configuration - 2:6:1 for all algorithms. 2000 events per class were used for the training, 20000 iterations were done to train each network. In order to obtain and unbiased estimate of the performance the *10-fold Cross-Validation* (CV) procedure was applied (see section 2.7.2 for procedure details). The 10-fold CV procedure was applied 10 times for each networks using independent data set each time, and the results were averaged over 100 runs. This is of vital importance in order to estimate an averaged reliable performance (prediction) of network instead of obtaining a particular (random) solution by making one learning session.

As a performance measure on control data set the geometrical mean per cent of the right classifications over 3 classes was used, and as a measure of speed the amount of the CPU time spent to make a 1000 iterations was used: $\sqrt[3]{R_{11}R_{22}R_{33}} \equiv R^{True}$ where R_{11} are the diagonal elements of the misclassification matrix. Of course, another more informative characteristic of the algorithm speed is the number of iterations after which the best point (minima in quality function) in the multidimensional space of NN weights is found, and then the algorithm is converged (saturation of the quality function). Nevertheless, such a benchmark could be used when the Bayes risk is reached within the given number of iterations for all NN to be compared.

Also the speedup of the algorithm convergence can be observed using different strategy within the given algorithm. An example of such case is the recommendation (Yao et al., 1999) to use the random numbers from Cauchy distribution (instead of the uniformly distributed random numbers) in formulae (4.27). The idea is, using random numbers from Cauchy distribution the probability of large random steps will be higher and it should yield to faster falling to minima of quality function. Such a comparison is also done, the neuron mode of RS

algorithm is tested with the usage of uniform random numbers and with random numbers from Cauchy distribution.

4.8.3 COMPARISON RESULTS

Table 6 displays the generalization performances of networks trained by different algorithms. The geometrical mean of the true classification rates over 3 classes is presented as a performance measure.

Table 6 . Performances of different learning algorithms

	RS(single mode)	RS (neuron mode)	RS(multi mode)	Rpro
R^T	66.5%	64.5%	61.4%	51.8%

The obtained performances are compared with results of Bayesian classification using Parzen's windows method for probability density estimation (Parzen, 1962). The mean per cent of the true classifications estimated by Bayesian method equals approximately to 67%. So, comparing the learning algorithms performances for this simple problem with linear decision boundaries with this value, it is easy to see that all modes of the RS algorithm give close results and the best is single mode.

The Rprop algorithm, as one can see, is the last in "quality row". This result was rather surprising, therefore the classification performance in two-way classification was checked. Both, RS and Rprop have shown very good results - ~ 85% mean per cent of true classifications, as was expected.

Figure 18 displays the learning process history for 20 NN trained with RS algorithm. Half of them were trained using uniform random numbers and the second half were trained using random numbers from Cauchy distribution. All parameters of the learning procedure and used data were the same, the neuron mode of RS strategies was used.

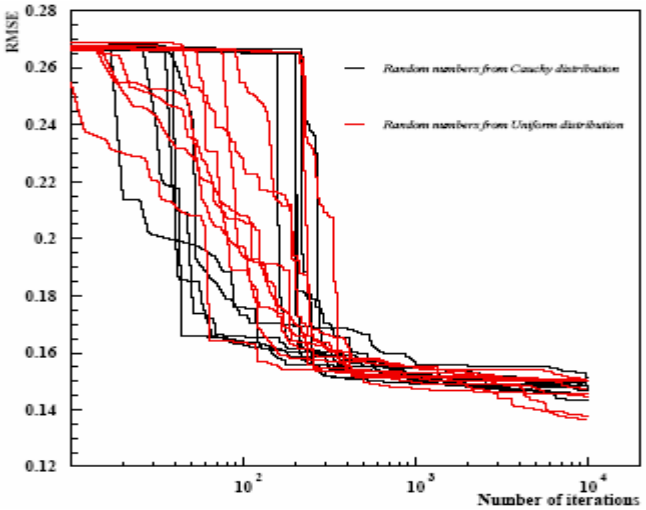


Figure 18. The net training results using random numbers from Cauchy population and uniform random numbers.

From this figure one can see that the better weights are found quicker when using random numbers from Cauchy distribution, but this is observed only at the first 100-200 iterations. In both cases the near-optimum is found nearly at 1000-th iteration, so, it is difficult to make any conclusion on advantage of using Cauchy random numbers for chosen problem. It is necessary to make investigations on different tasks and different kind of functions to be able to make

definite conclusion of superiority of networks trained using Cauchy random numbers.

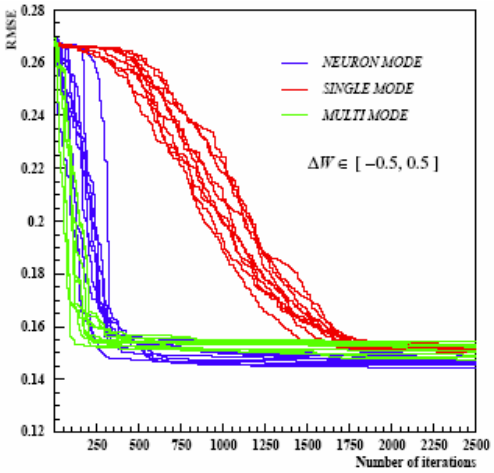


Figure 19. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (small steps).

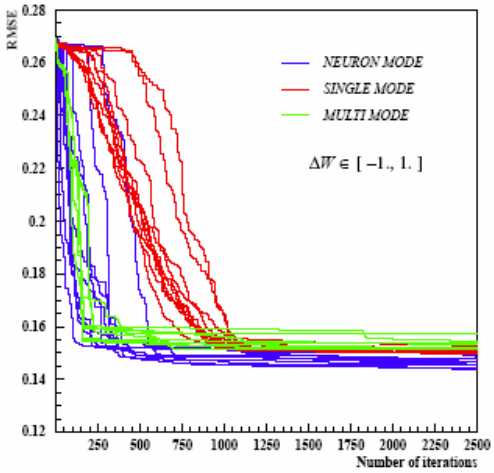


Figure 20. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (small steps).

More interesting is to investigate the performance dependence from the step size (learning rate). It is apparent that the performance of different modes of RS algorithm depends on this factor as opposite to *Rprop*. Thus, this dependence is investigated for three modes of RS algorithms in wide range of step sizes.

Figure 21 shows the dependence of the performance measure from step size for different training strategies. It is easy to see that the classification results of NN trained in single mode depend on step size slightly and the larger steps are preferable, since the Bayes optimum is reached. The other two modes have the opposite behavior. It is apparent from this figure that up to 16 step size the results of all three modes are very close, but in case of larger steps the neuron and multi modes show very poor performance.

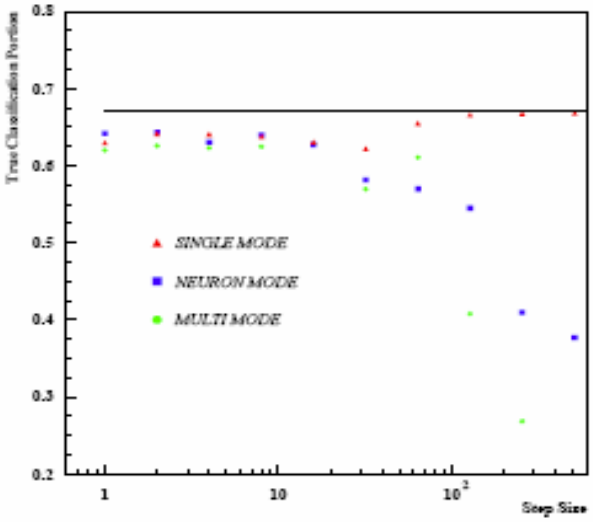


Figure 21. The dependence of final performance from learning step size.

The speed versus quality characteristic of these training strategies is also investigated. Figure 22, Figure 23, Figure 24, Figure 25, Figure 26, Figure 27 display the quality dependence from the number of iterations for different step sizes. As it is easy to see from first two figures the neuron and multi modes have a significant advantage as compared with single mode. The better points in multidimensional space of NN weights are found rather faster than in case of

single mode. When the step size becomes larger the situation changes, the number of iterations required to find a better points are reduced in single mode, and the performances of single and neuron modes are closer, while the multi mode get a bit worse (see Figure 22, Figure 23).In case of very large steps the situation dramatically changes for neuron and multi modes, the number of successful steps are very rare and for a long training time the near optimum solutions are not found. For the networks trained with multi mode this falling of the performance with increasing of the step size can be explained easily. Since this strategy is deterministic and the search procedure is performed in the multidimensional hypercube with side equal to the step size (see above), and the number of iterations (i.e. the number of predefined points in this hypercube) were fixed, in case of very large step sizes 10000 points in hypercube are very sparse and the probability to find a good point is not high.

What we can conclude is that, in case of small step sizes the Bayes optimum is practically always reached by all three modes and the neuron and multi modes are very fast (~ 250 iterations are required to fall down to satisfactory minimum in quality function) in finding of good point in multidimensional space of weights. Taking into account that the very large steps have no principal advantage over small ones, one can recognize that the neuron mode is the best in terms of quality versus speed ratio.

The networks trained with single mode show the excellent quality within the same number of iterations in case of large steps as well as in case of smaller steps. This strategy of the RS algorithm shows a rather stable performance for a wide range of the random step size.

However, the different strategies can be used in one session of training, by the following scenario: the learning can be started using multi mode to scan a very large space of weights, to find a better point and then to switch to the neuron mode and continue the search in space near to the already obtained point with small steps; finally the single mode can be used to tune the NN weights for the given problem solution with high verbosity, of course, if the problem is complex enough and the saturation of the quality function is not available yet (the " global" minimum is not found). The use of the CV procedure for the overtraining effects controlling is assumed by default here.

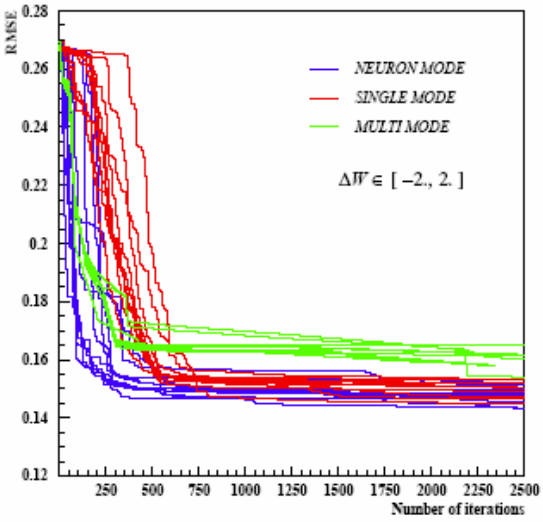


Figure 22. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (medium steps).

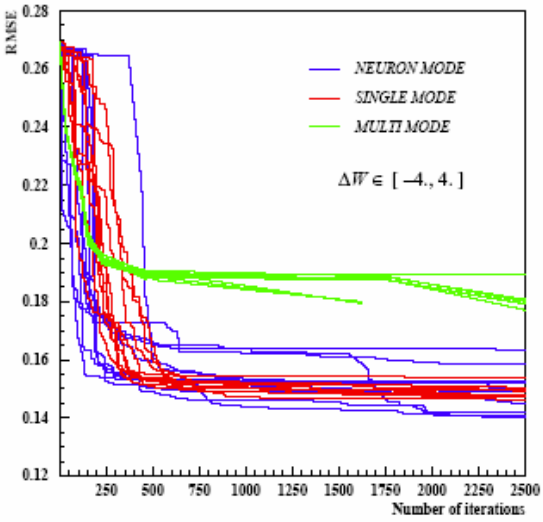


Figure 23. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (medium steps).

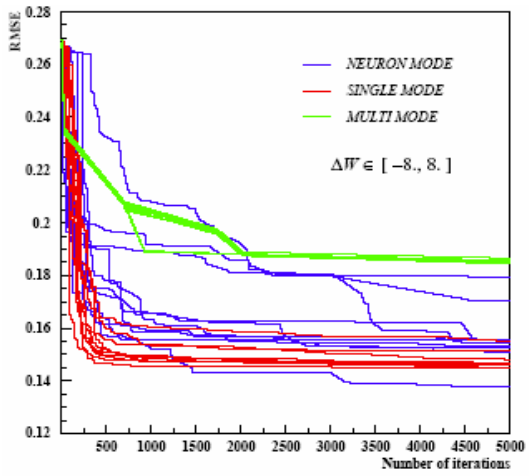


Figure 24. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (large steps)

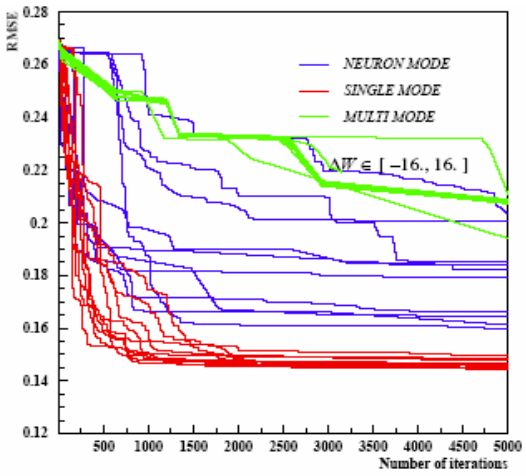


Figure 25. Quality function versus number of iterations in different CV stages for three modes of RS algorithm (large steps)

4.9 INVESTIGATION OF INTERPOLATION POSSIBILITIES BY NEURAL NETWORKS

In the regression tasks good interpolation of the training sample events (small training error) is necessary condition to be sure in accurate estimation of regressand in control (experimental) sample. As will be shown bellow, if the inherent Bayes error is small and training examples are sufficient, this is practically always reachable by neural networks. In problem of the primary energy estimation of CR particles we meet this requirements. What we need is, having finite simulation samples to teach the NN to generalize the functional dependence of the following form:

$$E_0 = f\left(N_e, N_\mu^{tr}, \dots\right) \qquad (4.29)$$

in a wide range of E_0 - energy of PCR particles (regressand).$N_e, N_\mu^{tr}, \dots$, are the EAS characteristics (inputs of the NN). The form of f is unknown and will be accumulated in NN weights during training process using simulation trials (training samples, where for each input vector the corresponding true value of the regressand is known).

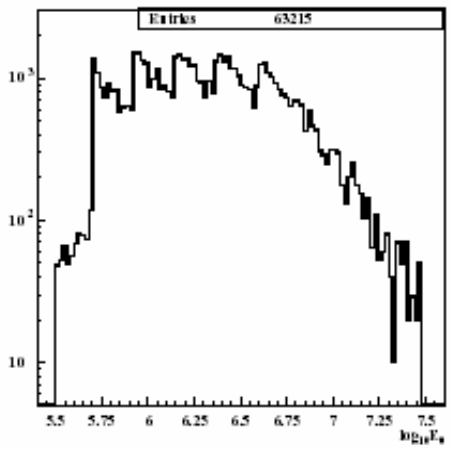


Figure 26. The distribution of primary energy in simulated sample.

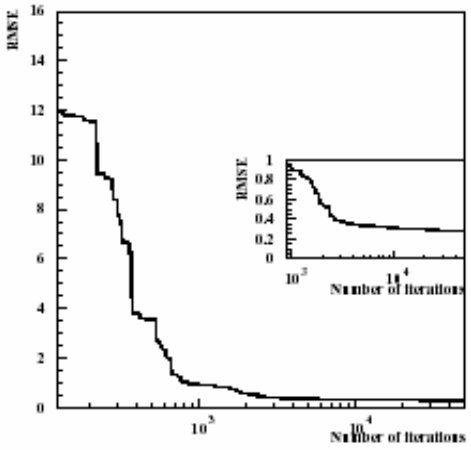


Figure 27. The history of NN training for the estimation task.

The problem of sufficient training examples arises from the difficulties to simulate large number of events in the whole energy interval of investigation. Usually we are able to handle much more low energy events than more interesting high energy events, since the simulation of high energy events is very hard and time consuming process as compared with lower energy events simulation.

The main problem of unbiased and accurate estimation of primary energy in a wide energy range is associated with the deficiency of simulation trials. The Figure 26 shows a common shape of the energy distribution in simulations. In order to overcome the problem of low statistics in high energy interval usually one takes the narrow energy bins and simulates events distributed uniformly (or by some low) in each bin.

The NN with 2x8x6x1 configuration was trained to estimate the primary energy using more than 60000 events shown in Figure 26. Relatively large NN was used since the TS size is large enough. The neuron mode of RS algorithm was used for the training, 50000 iterations were performed. The Figure 27 shows the learning history of the NN. It is easy to see the saturation of objective function (RMSE) after 10000 iterations. On the inset of this figure the same dependence is depicted in more details above 1000 iterations, indeed, the change of objective function is negligible for a long training time. From Figure 28 and Figure 23 one can see the estimation results. In spite of small overestimation for lowest energies, where a few number of events are available, the well agreement is observed between true and estimated values of energy in the low and middle energy region. The agreement is much more worse in the region of grater energies, although the spread is smaller, but there is a strong bias of estimation for the most interesting highest energies. This result can be explained by the deficiency of training events in high energy region. Indeed, the main contribution to the quality function (eq.(4.8)) minimizing in training process comes from a large number of low energy events, and the overall decrease of quality function exists while the discrepancy of apparent and true values of small number of high energy events remains relatively large.

The primary energy spectrum of CR can be described by power low in energy:

$$\frac{dI}{dE} = I_0 \cdot E^{-\gamma} \qquad (4.30)$$

where dI/DE is the differential flux in particles, Io is the absolute flux normalization and γ is the spectral index. So, in the simplest case when in simulated sample the energy is distributed according to this low, its very easy to give weights to each event according to the inverse low:

$$W_i = \frac{E_i^{\gamma}}{I_0}, \quad i=1,N_{events} \qquad (4.31)$$

But, as was mentioned above, the simulations are usually performed bin by bin in whole energy interval to accumulate more high energy events (Figure 26). In this case the weights were defined as follows:

$$W_{ij} = \frac{1}{N_b M_i} \qquad (4.32)$$

where $N_b$is the number of bins in which the distribution of energy is approximately uniform, $M_i; i = 1, N_b$is the number of events in i-th bin, $j = 1, M_i$. Figure 29 and Figure 25 display the estimation results using TS with weighted events. It is worth to mention that 10000 training iterations were performed to obtain this result. It is easy to see the well agreement of true and

estimated values in whole interval of energy distribution support. The overall bias of estimation does not exceed 5%. So, the introduction of individual event weights enhances the estimation power significantly and makes the NN training faster.

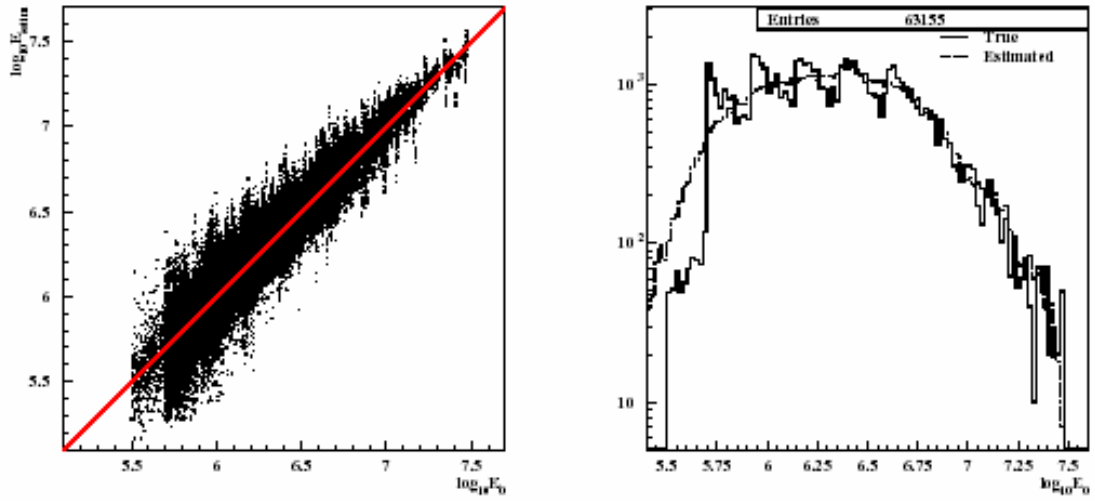


Figure 28. The NN regression using events with individual weights, estimated energy versus true energy.

CHAPTER 5.
INVESTIGATION OF EAS CHARACTERISTICS

5.1 THE KASCADE EXPERIMENT

The KASCADE (Klages et al., 1998) experiment consists of three main parts - the scintillator array, the central detector and the muon tunnel, and is able to measure large number of EAS characteristics for each individual event.

A scintillator array measures the electrons, photons and muons outside the core region of extensive air showers in 252 detector stations on a rectangular grid of 13 m spacing, hence forming an array of $200 \times 200 m^2$.

The key component of the central detector is a finely segmented hadron calorimeter. A $20 \times 16 m^2$ iron stack arranged in 9 horizontal planes measures the hadrons and their energy. Below 30 radiation length of absorber the central calorimeter contains a layer of 456 scintillation detectors acting as trigger for the calorimeter and measuring the arrival time of hadrons. Underneath the calorimeter two layers of multiwire proportional chambers measure muon tracks above an energy of 2 GeV with about 1.0 deg angular accuracy.

North of the central detector in a 50 m long tunnel muons above an threshold energy of 0.8 GeV are measured with the help of streamer tubes. On an area of coverage of $150 m^2$ they track muons with 0.5 deg accuracy. The schematic view of the KASCADE experiment is shown in Figure 29.

5.2 THE SIMULATION PROCEDURE

All statistical decisions and procedures are correct within the prechosen model. Thus a realistic simulation is the key problem of any physical inference in indirect experiments. Extensive air shower investigations are a classical example of such a situation. An adequate consideration of detector response and an identical reconstruction of experimental and simulated data are necessary steps of data analysis.

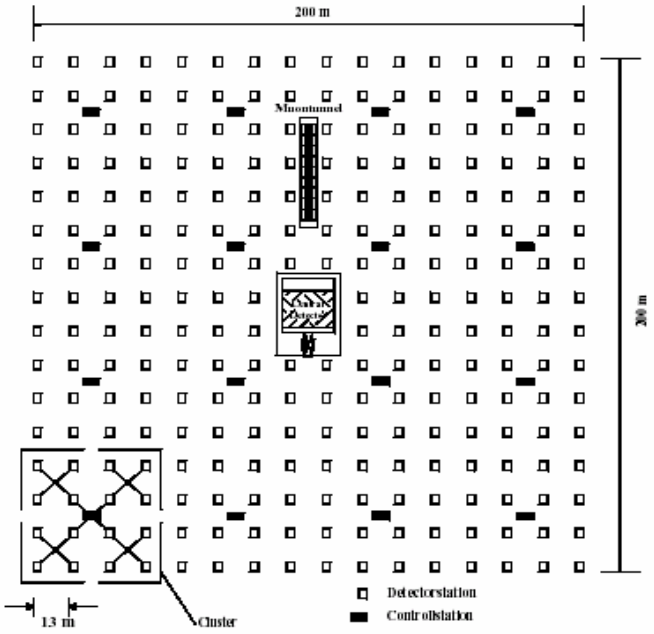


Figure 29. Lay-out of the KASCADE experiment

The simulation data base of the KASCADE experiment is fulfilling the above requirements, and we use specific EAS parameters, like the numbers of electrons/photons (N_e) and truncated number of muons (N_μ^{tr}) and the age (S_{89}) parameter as input for data analysis (Glasstetter, 1997), (Weber, 1997).

The physical meaning of these variables will not be discussed, we only mention that the procedures of their use are identical for experimental data and simulations. It is also very important to say that hypotheses about lateral distributions of muons and electrons at very small and large distances are not of influence.

The simulations of the EAS development in the atmosphere were done with the CORSIKA code (Heck et al., 1998) (VENUS (Werner, 1993) and QGSJest (Kalmykov.et al., 1997) models).

To estimate the bias due to finite sampling and reconstruction errors the detailed detector simulation on the basis of the GEANT (CERN, 1993) package was made taking into account all shower particles, absorbers and active materials, energy deposits, trigger conditions and efficiencies, as well as the electronics, digitization of pulse heights, times, etc.

In the second step the EAS parameters reconstruction programs 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 detector, and many other characteristics are calculated.

The parameters of simulated showers were reconstructed with the same programs as experimental ones. The measured EAS parameters by KASCADE are shown in the Table 7.

Table 7. EAS features detected by KASCADE experiment

N_e	Number of electrons in EAS
N_μ^{tr}	Truncated number of muons (number of muons in the range of 40 to 200m)
S_{89}	Shower age associated with a Molier radius 89m
N_μ^{CD}	Number of muons in Central Detector (CD)
N_h	Number of hadrons
E_h^{max}	The energy of most energetic hadrons
E_h^{sum}	Total energy of hadrons

As one can see from the Figure 30 the overlapping of shower parameters corresponding ever to more distinct classes (proton and iron initiating primaries) is rather big and one can't expect reliable classification of primaries according to the single EAS features.

In the multidimensional features space, as one can see from the Figure 31 the differences between proton and iron samples could be detected. Therefore, the detailed examination of all EAS characteristics and their correlations will allow to find a subset of features to be used for experimental data classification.

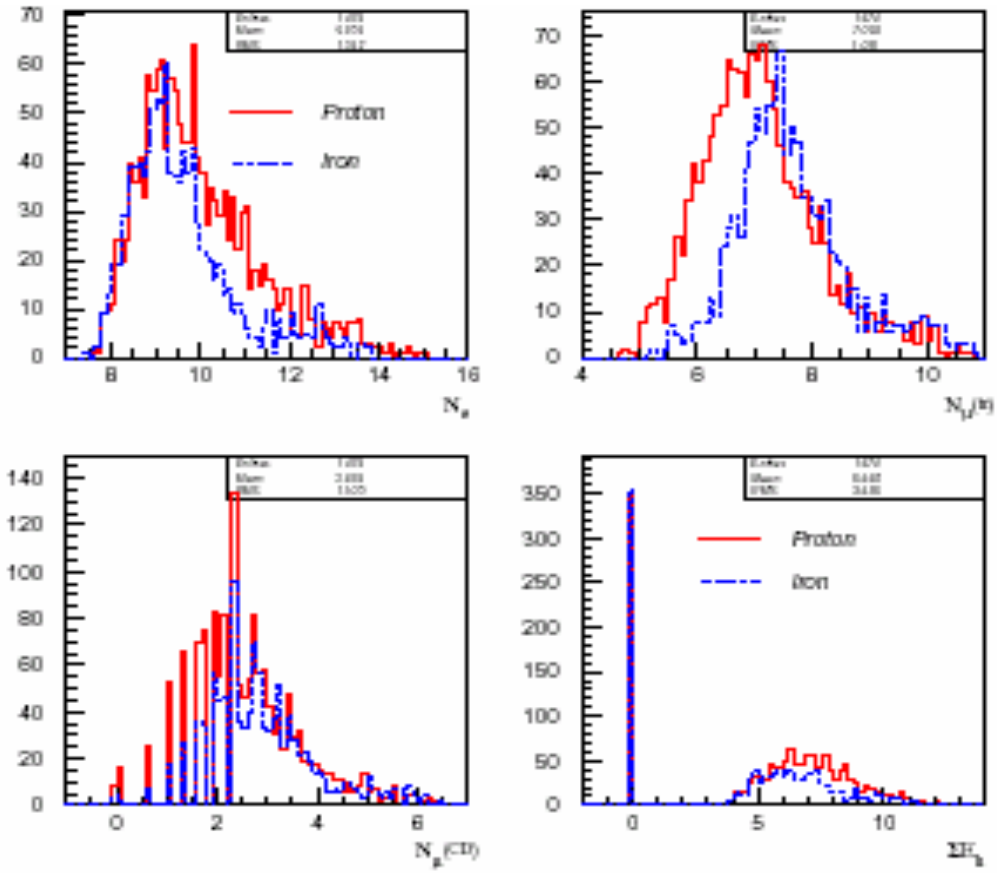


Figure 30. Features distribution for proton and iron

5.3 VALIDATION OF MODELS

5.3.1 COMPARISON OF THE SINGLE EAS VARIABLES

First of all we've to examine the variables to select primary mass discriminants and variables correlated with the primary energy. We use for this purposes simulated events, for which the "true" values of mass and energy are known. For all measurable EAS variables we calculate the P-values of following statistical tests:

- Student's t-test

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} \qquad (5.1)$$

where the μ_1, σ_1 and μ_2, σ_2 are the mean values and the standard deviations of the first and second class respectively.

- Kolmogorov-Smirnov .D-test

$$D = \sup remum_v |F_1(v) - F_2(v)|; \qquad (5.2)$$

were $F_1(v)$ and $F_2(v)$ are the cumulative probability function for first and second class (model) respectively.

- Mann-Whitney U-test

$$U = \frac{T_1}{M_1} - \frac{T_2}{M_2}; \qquad (5.3)$$

were the T_1 and M_1, M_2 and T_2 are the sum of ranks of events from first and second samples

respectively, are the number of events in samples. The rank is the number of particular event in ordered sequence of events (variation row).

From the Table 8 one can see that results of applying of three mentioned statistical tests are consistent with each other and the best primary discriminators are N_μ^{tr},age, N_e, N_μ^*. The results are model-independent, the same variables are chosen both by QGS and VENUS modes of CORSIKA simulations.

5.3.2 CORRELATION ANALYSIS

The correlation information is used for the choice of variables for energy estimation. From the Table 9 and Table 10 it is seen that again information on EAS muons and electrons is the most valuable for the estimation of the primary energy. Muon information is slightly better (correlation approaching 1) than electron information. The results are again model- independent.

The correlation analysis are used also to select the best pairs of variables for distinguishing between classes. The pairs correlated different within the same model, but for different primaries are selected.

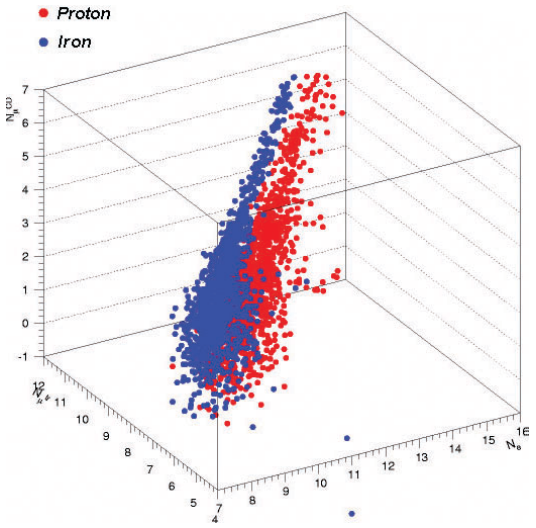


Figure 31. Proton and iron events distribution in 3 dimensional space offeatures

Table 8. P-values of statistical tests for proton and iron for different models: t Student, Nμ* D - Kolmogorov- Smirnov, U - Mann-Whitney

QGS	t	D	U	VENUS	t	D	U
N_e	3.177	2.747	4.996	N_e	2.869	3.161	5.778
N_μ^{tr}	12.601	6.026	12.723	N_μ^{tr}	10.274	4.282	9.403
S_{89}	17.160	7.489	17.294	S_{89}	20.415	8.314	19.473
N_μ^{CD}	7.207	3.452	7.132	N_μ^{CD}	5.650	2.031	3.872
N_h	0.673	1.647	2.811	N_h	2.265	3.335	5.848
E_h^{max}	5.564	3.066	6.144	E_h^{max}	3.612	2.402	4.458
E_h^{sum}	2.478	3.126	3.985	E_h^{sum}	3.457	3.140	6.174

It have to be mentioned, that for constructing both primary and energy estimators, better to use nonoverlapping complexes of features. It is not easy to do, because, muon and electron information is the best signature as well for primary mass, as for primary energy.

Table 9. Correlation matrix for QGS data

	Mass	E_0	N_e	N_μ^{tr}	S_{89}	N_μ^{CD}	N_h	E_h^{max}	E_h^{sum}
Mass	1.00	0.21	-0.03	0.27	0.32	0.15	-0.03	-0.11	-0.07
E_0	0.21	1.00	0.92	0.95	-0.25	0.94	0.78	0.53	0.73
N_e	-0.03	0.92	1.00	0.90	-0.43	0.93	0.85	0.62	0.81
N_μ^{tr}	0.27	0.95	0.90	1.00	-0.23	0.93	0.78	0.52	0.72
S_{89}	0.32	-0.25	-0.43	-0.23	1.00	-0.33	-0.39	-0.33	-0.38
N_μ^{CD}	0.15	0.94	0.93	0.93	-0.33	1.00	0.86	0.60	0.82
N_h	-0.03	0.78	0.85	0.78	-0.39	0.86	1.00	0.70	0.95
E_h^{max}	-0.11	0.53	0.62	0.52	-0.33	0.60	0.70	1.00	0.73
E_h^{sum}	-0.07	0.73	0.81	0.72	-0.38	0.82	0.95	0.73	1.00

5.3.3 PROBABILITY DISTANCES

Another important measure of the separability of two samples is the Bhattacharya distance, which takes the form:

$$Bhata = \frac{1}{8}(\mu_2 - \mu_1)^T \left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1} (\mu_2 - \mu_1) + \frac{1}{2}\ln \frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}} \tag{5.4}$$

where the μ_i and Σ_i are the expected vector and covariance matrix of i-th class. The first term of this equation is the *Mahalanobis distance* and the last term is the so called correlation distance.

$$R_{Mahal} = (\mu_2 - \mu_1)^T \left(\frac{\Sigma_1+\Sigma_2}{2}\right)^{-1} (\mu_2 - \mu_1) \tag{5.5}$$

$$R_{Corr} = \ln \frac{\left|\frac{\Sigma_1+\Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}} \tag{5.6}$$

The selection of the best subsets of EAS features according the Bhattacharya distance is presented in Table 11. The best method of feature subset selection is connected with Bayes risk (1.22) and Error function (4.8) estimation.

Table 10. Correlation matrix for Venus data

	Mass	E_0	N_e	N_μ^{tr}	S_{30}	N_μ^*	N_h	E_h^{max}	E_h^{sum}
Mass	1.00	0.18	-0.06	0.18	0.33	0.09	-0.05	-0.09	-0.07
E_0	0.18	1.00	0.91	0.95	-0.24	0.94	0.80	0.52	0.76
N_e	-0.06	0.91	1.00	0.90	-0.40	0.93	0.89	0.60	0.85
N_μ^{tr}	0.18	0.95	0.90	1.00	-0.24	0.92	0.81	0.51	0.76
S_{89}	0.33	-0.24	-0.40	-0.24	1.00	-0.32	-0.41	-0.28	-0.40
N_μ^{CD}	0.09	0.94	0.93	0.92	-0.32	1.00	0.86	0.57	0.83
N_h	-0.05	0.80	0.89	0.81	-0.41	0.86	1.00	0.65	0.95
E_h^{max}	-0.09	0.52	0.60	0.51	-0.28	0.57	0.65	1.00	0.68
E_h^{sum}	-0.07	0.76	0.85	0.76	-0.40	0.83	0.95	0.68	1.00

Table 11. The best feature subsets according to the Bhattacharya distance

	2 best		next best			worst	
QGS	Ne	Nμtr	NμCD	S89	sumEh	Nh	Ehmax
VENUS	Ne	Nμtr	S89	NμCD	sumEh	Nh	Ehmax

5.3.4 QGS AND VENUS COMPARISON

To compare different strong interaction models, one should have the same mass composition and energy distribution in experimental and simulated data, to avoid mass depended differences.

The mass composition of primary cosmic radiation in low energy region (below 10^{15} eV) is measured by direct methods and the following proportion of different nucleus is assumed to be true (Watson, 1997): H - 24%; He - 31%; 0 - 21%; Si - 12%; Fe - 12%.

To avoid energy spectrum based differences we choose rather narrow energy interval. The truncated muon interval (in logarithmic scale) $7.82 \leq N_{\mu}^{tr} \leq 9.21$ is corresponding to the $6 * 10^{14} \leq E_0 \leq 2 * 10^{15} eV$.

Thus, we construct simulation samples from VENUS and QGS models with this proportion of primaries and in the mentioned energy range. The selection was made by in N_{μ}^{tr} both, Monte Carlo and experimental data.

5.4 CLASSIFICATION OF EXPERIMENTAL SAMPLE

Table 12. Comparison of exp. data with VENUS and QGS models

	L	R_{Bhata}	R^e
QGS	1.2036 ±0.01	0.023 ±0.001	0.456 ±0.01
VENUS	1.1818 ±0.01	0.014 ±0.001	0.469 ±0.02

Both models are very close to experimental data (see Table 12, Table 13), but all tests give a bit preference to the QGS model.

Table 13. One dimensional tests for models and experiment

QGS	t	D	U	VENUS	t	D	U
N_e	0.916	2.312	0.901	Ne	3.369	3.004	4.368 4.609
N_{μ}^{tr}	3.199	1.450	2.653	N_{μ}^{tr}	2.632	4.787	

QGS - 0.55 ±0.02

VEN - 0.45 ± 0.02

5.5 THE HOMOGENEITY OF THE KASCADE DATA

2371 events of central calorimeter and array data and 450000 events of array data only were available for analysis. The events were selected within 15 — 20° zenith angle range (the simulations were done for these angles).

To check homogeneity of data we divide experimenta l data to 3 parts and make multiple comparisons with techniques described above. Also the negative Log likelihood function value L (1.17) and estimate of Bayesian error R^e (1.25) were calculated.

Table 14. Exp. data homogeneity test, features used: N_μ^{CD}, E_h^{sum}

	£	$R_{Mahal..}$	$R_{Bhata.}$	Rcorr.	R^e
1 class	2.911	0.011	0.007	0.024	0.479
2 class	2.803	0.029	0.021	0.069	0.466

The homogeneity check for array and calorimeter data one can find in the Table 14, Table 15, Table 16, Table 17. All tests demonstrate rather good agreement with each other and prove the homogeneity of experimental data samples.

Table 15. Exp. data homogeneity test, features used: N_e, N_μ^{tr},N_μ^{CD},E_h^{sum}

	£	$R_{Mahal.}$	$R_{Bhata.}$	$R_{corr.}$	R^e
I class	4.434	0.129	0.038	0.086	0.429
2 class	4.289	0.056	0.041	0.135	0.453

Table 16. Exp. data homogeneity test, features used: N_e, N_μ^{tr}

	£	$R_{Mahal.}$	$R_{Bhata.}$	$R_{corr.}$	R^e
I class	1.428	0.002	0.000	0.000	0.490
2 class	1.425	0.001	0.000	0.000	0.494

Table 17. One dimensional tests of exp. data: t - Student, D - Kolmogorov-Smirnov, U - Mann-Whitnay

	t	D	U
N_e	1.816	1.114	1.433
N_μ^{tr}	0.080	0.823	0.570
N_μ^{CD}	1.595	1.118	1.776
E_h^{sum}	1.418	0.407	1.358

5.6 THE KASCADE CLASSIFICATION MATRICES

The examination of classification matrix and its index (1.26) gives clues for understanding the discriminative power of different EAS measurables for composition estimation.

The value greater than of 0.6 are still allow for solving the system of equation (1.28). For the lower values the solutions didn't converge and fraction couldn't be reconstructed. Therefore we have to find appropriate variables, or reduce the number of classes. Only balance between expected classification errors and number of classes used, will allow to obtain reasonable and reliable estimates of the fraction.

As one can see from the Table 18 and Table 19, present status of a priori knowledge accumulated in M.C. models and represented in training samples, didn't support the attempts to make 5-way classification even for all available features.

The situation with 3-way classification is much better, as we need much less a priori information, comparing with classification into 5 nuclei groups. As we can see from Table 20 and Table 21, even array information only allows to resolve the distribution mixture (1.4).

The calorimeter information significantly increase the expected fraction reconstruction accuracy.

Table 18. Array data, features used N_e N_μ^{tr}

0.5148	0.2620	0.1024	0.0565	0.0643
0.3240	0.2979	0.1846	0.0885	0.1050
0.1185	0.1803	0.2340	0.1721	0.2951
0.0707	0.1064	0.1942	0.1958	0.4330
0.0445	0.0659	0.1362	0.1362	0.6527

Table 19. KASCADE data, features used $N, N_\mu^{tr}, N_\mu^{CD}, E_h^{sum}$

0.4785	0.3085	0.1323	0.0438	0.0368
0.2678	0.3863	0.1943	0.0766	0.0750
0.0558	0.1932	0.2903	0.2200	0.2407
0.0367	0.1247	0.2311	0.2506	0.3570
0.0368	0.0706	0.1544	0.1471	0.5912

Table 20. 3-way classification by N_e N_μ^{tr}

0.6831	0.2595	0.0574
0.2132	0.4849	0.3019
0.0919	0.3078	0.6003

Table 21 3-way classification by $N, N_\mu^{tr}, N_\mu^{CD}, E_h^{sum}$

0.7108	0.2419	0.0473
0.1777	0.5176	0.3048
0.0779	0.2559	0.6662

The 2-way classification in "heavy" and "light nuclei can be done with greater accuracy. See Table 22 and Table 23

Table 22. 2-way classification by N_e, N_μ^{tr}

| 0.863 0.137 |

| 0.088 0.912 |

Table 23. 2-way classification by $N, N_\mu^{tr}, N_\mu^{CD}, E_h^{sum}$

| 0.865 0.135 |

| 0.076 0.924 |

The information concern 5, 3 and 2 -way classifications for KASCADE different parts is summarized in Table 24, where the separability indexes arepresented.

Table 24 Separability index for KASCADE

	Index-	Index-3	Index-2
CD	0.154	0.462	0.708
ARRAY	0.341	0.584	0.887
ARRAY+CD	0.38	0.626	0.894

5.6.1 COLORED NUCLEAR MAPS

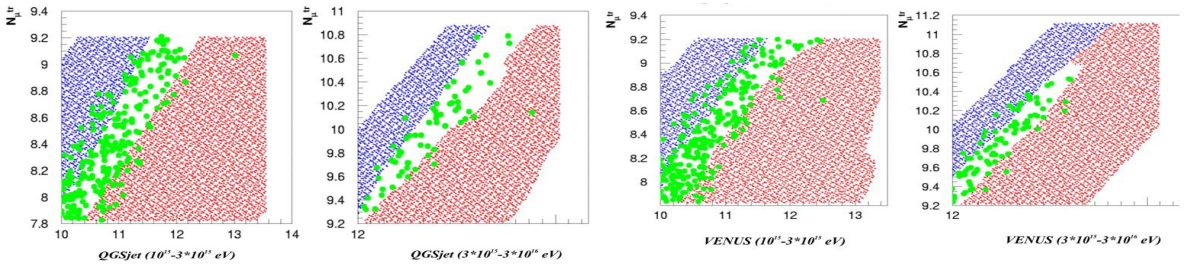


Figure 32. 3-way map, array information. Green points represent oxygen MC data.

It is of greatest importance to divide initial feature space according to different primaries. Each decision rule maps v_i events to one of 3 nuclei groups. Visa-versa, each nuclei group is mapped by decision rule (1.12) to the definite region of feature space V.

By examining of such "nuclear maps" we can make insight to the possibility of defining the type of particular nuclei and about expected misclassification to the other nuclear groups.

Overlaying the experimental data on the colored nonlinear "masks" we can visualize the Bayesian decision procedure.

The different masks, for various variables, two energy regions and 2 strong interaction models were used for multiple comparisons and for visualization of the decision boundaries.

On the Figure 3.4 3-way classification maps for array variables are presented. The oxygen simulations (green points) are superimposed on the "light" (red color), "heavy" (blue), and "intermediate (white) clusters.

The colored maps directly correspond to the R^e estimates and tables from the previous section.

CHAPTER 6.
EAS DATA ANALYSIS

6.1 ENERGY SPECTRA AND MASS COMPOSITION OF PRIMARY COSMIC RAY FLUX, MEASURED ON SHOWER-BY- SHOWER BASIS

6.1.1 COSMIC RAY (CR) SOURCES AND ACCELERATION MECHANISMS

There is widespread agreement that the bulk of high energy cosmic rays is accelerated by supernova shocks in the interstellar medium by a mechanism called 'diffusive shock acceleration' although direct observational evidence is still missing. There is also agreement, though, that this mechanism is not effective to the highest energies observed. The exact upper limit is not clear but is usually assumed to lie between 1 and 10 PeV. Therefore the determination of the exact spectral shape and of the primary chemical composition in this range are of prime importance. In addition, the all particle spectrum in this energy region shows a change of the slope of the overall power law which appears to describe the spectrum satisfactorily. This 'knee' of the spectrum which was first observed more than 40 years ago (Kulikov et al., 1959) is the only remarkable structure over a range of many decades.

Measurements of the energy spectra of the individual species of the cosmic ray flux could provide clues to the origin of this knee such as a rigidity cut off, inherent to Fermi acceleration, or peaks, revealing the influence of individual nearby supernovae.

Above a primary energy of a few hundred TeV the direct measurement of energy and mass of individual cosmic ray nuclei is unfeasible due to the drastic decrease of intensity with increasing energy. One has therefore to resort to the measurement of extensive air showers (EASs) which are produced when high energy cosmic ray particles enter into the atmosphere of the earth. To determine primary energy and mass from EAS observables has been tried since many years but has proved to be a very tough problem.

The idea to use advanced statistical techniques of multivariate analyses (Chilingarian, 1989) for isolating certain classes of EAS stems from an early proposal of A. Chilingarian and H. Zazyan (Chilingarian et al., 1991), (Chilingarian 1991A) to prepare quasi-mononuclear beams by mass discriminative analyses of event-by-event EAS observations, planned for the ANI experiment (ANI Collaboration, 1992), (Chilingarian, 1998). The realization of this proposal has become realistic by the recent results of the multi-detector experiment KASCADE (Kampert, et al., 1999) which provide an accurate experimental basis of event-by-event data of many EAS observables.

This approach appears to be very promising with the aspect of refined tests of current interaction models and to pave the way to a consistent description of the hadronic interaction at extremely high energies by experimental road signs. Still the results of the KASCADE experiment concerning the energy spectrum and mass composition of primary cosmic rays are considerably affected by a model error, estimated for the energy slope to be 10 times larger than the statistical uncertainty (Chilingarian et al., 1999A).

The present report introduces in recent results on energy dependence of mass composition and energy spectra of different species of the primary flux. Also first attempt is made to analyze

"mononuclear beams" interactions with atmosphere using KASCADE Central Detector (CD) information.

It is necessary to mention that results obtained within event-by-event analysis approach are

conditional on the particular strong interaction model used for simulation of the primary passage through the atmosphere. The deconvolution of the triple uncertainty (primary mass, primary energy and strong interaction model) and obtaining of "pure" mononuclear beams even within one prechosen model will provide hints to understand direction in which family of strong interaction models have to evolve to meet experimental consistence criteria.

6.1.2 PRIMARY ENERGY ESTIMATION

The multi-layered perceptron (MLP) algorithm is used to determine the mass composition and the energy spectrum of the primary cosmic rays (PCR) in the knee region. This method allows the possibility of primary energy estimation as well as primary mass classification into multiple categories. The basics of Bayesian and neural regression and classification techniques are described in chapters 1, 2 and in (Chilingarian, 1994), (Chilingarian, 1995), (Chilingarian et al., 1997), (Chilingarian et al., 1997A).

For the estimation and classification tasks the same EAS observables are used. These are observables of the electromagnetic and muonic components measured by the KASCADE field array detector installation:

- N_e number of electrons in EAS,
- the shower age parameter s both associated with a Moliere radius of 89m
- $N_\mu{}^{tr}$ truncated number of muons ($N_\mu^{tr} = 2\pi \int_{40m}^{200m} \rho_\mu(r) r dr$)

Restricting these observables, is justified by following reasons:

- It is assumed that the electromagnetic and muonic components of EAS are described by the MC models with sufficient accuracy (a partially insufficient knowledge about the hadronic component is illustrated in (Antoni et al., 1999), (Roth, 1999).
- Due to the larger statistical accuracy the uncertainties caused by strong EAS fluctuations are eliminated as compared with hadronic information of EAS.
- The KASCADE CD information can be used independently, after obtaining "mononuclear beams".

The energy estimation is performed using two sets of observables: N_e and N_μ^{tr} and N_e, and s. As one can see in the Figure 33, adding the observable s the accuracy of the energy estimation is enhanced. Indeed, the correlations of the N_e and N_μ^{tr} with primary energy are very strong (Chilingarian et al., 1999B), and further the s parameter is correlated with primary mass (see Table 25), therefore the usage of all 3 parameters, "fixing" in some sense the primary type, enlarges the overall accuracy of the energy estimation for all 3 groups of nuclei. Another important characteristic is the bias of estimator. Only small biases allow to reconstruct energy spectra and estimate slopes and "knee position" adequately. In Figure 33 and Figure 34 the relative error of the energy estimation is shown. For both sets of observables an almost unbiased estimation in the whole energy interval (except at the most lowest and highest energies) is apparent. Therefore we use a wider energy interval for simulated events to avoid over- and underestimation of primary energies at the boundaries.

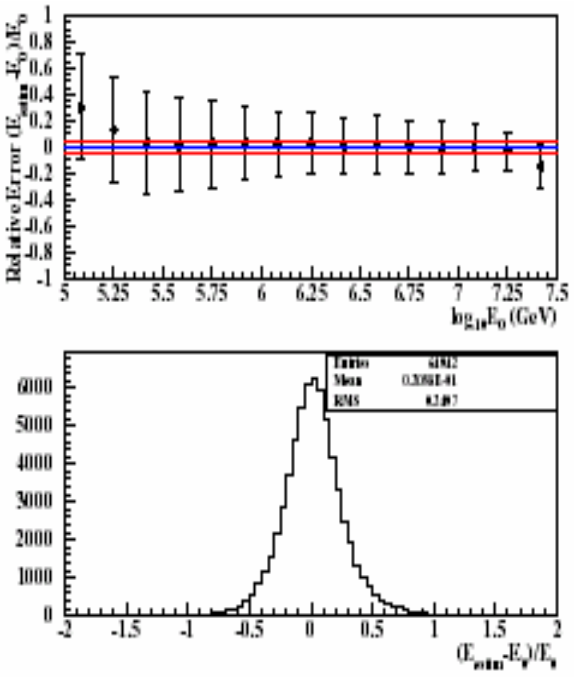


Figure 33. The accuracy of the energy estimation. Used observables: N_e, N_μ^{tr} and s.

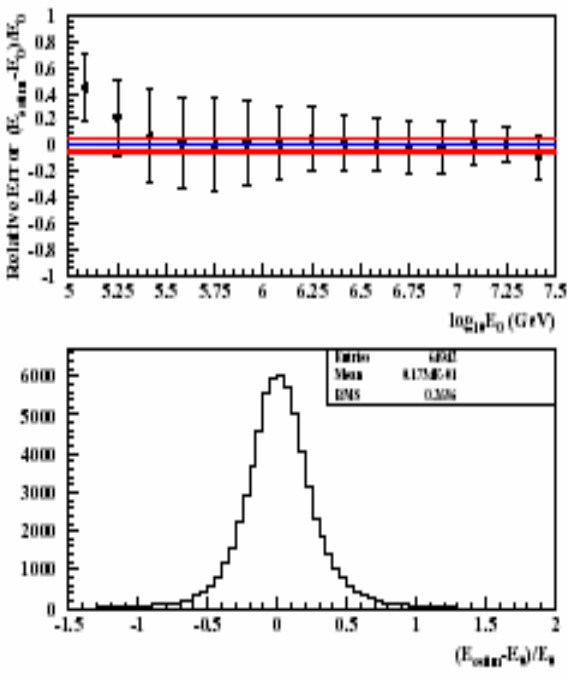


Figure 34. The accuracy of the energy estimation. Used observables: N_e and N_μ^{tr}

6.1.3 PRIMARY MASS DETERMINATION

After performing energy estimation, each EAS event is classified as induced by light (H,He), intermediate (CNO) or heavy (Si-Fe) nuclei (bellow we will refer these groups as "proton", "oxygen", and "iron"). The parameters of experimental events, classified as initiated from proton and iron nucleolus, are afterwards compared with those of simulated ones. The results are given in Figure 35 and Figure 36. A rather good agreement for simulated and experimental proton and iron induced events is obvious. In contrast, as one can see in Figure 37, the mean shower age of experimental data is shifted compared with the simulated data in whole energy interval, but still a clear difference between proton and iron events exist.

Although simulations do not describe correctly the observable s, nevertheless the s parameter is a good signature of the primary mass. However the systematic bias of the s parameter can be misleading. For example, a neural net (NN) trained by N, N^{tr} and s can result in a heavier mass composition, when applying it for experimental data classification. As seen from Figure 37, smaller s corresponds to lighter nucleus in both simulated and experimental data, but the absolute values differ significantly and thus, in mean, a proton initiated events with high probability could be classified as belonging to the iron type.

From Table 25 where the correlations between primary energy and measured EAS parameters are presented, one can see the negative correlation of s and primary energy. It means, that applying trained neural network for energy estimation, systematically larger values of s entered network comparing with ones used during the training. Therefore, the incorporation of the s parameter will lead to the systematic underestimation of higher energies and correspondingly to a bias of the reconstructed energy spectrum. For that reason, we use for energy estimation only N_e and N_μ EAS characteristics that adequate describe the experimental distributions.

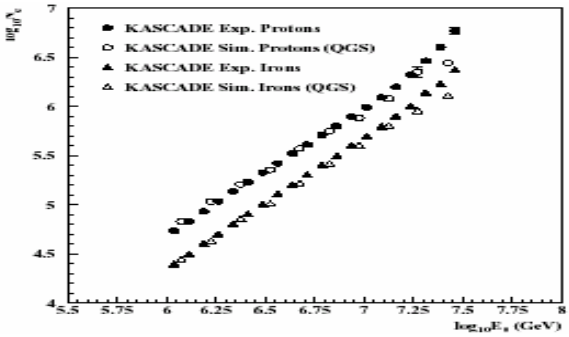


Figure 35. N_e versus E_0 for simulated and experimental proton and iron events (the primary energy is estimated by neural regression method)/ used observables: N_e and N_μ^{tr}

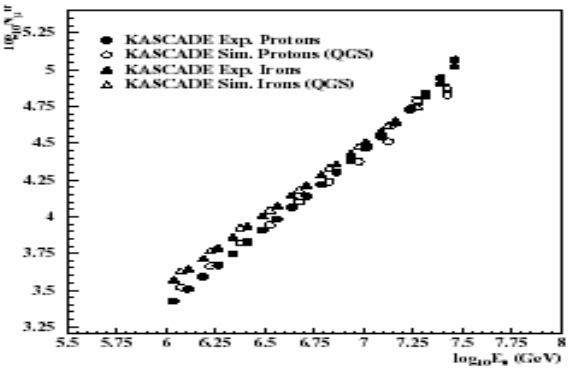


Figure 36. N_μ^{tr} versus E_0 for simulated and experimental proton and iron events (the primary energy is estimated by neural regression method). Used observables: N_e and N_μ^{tr}.

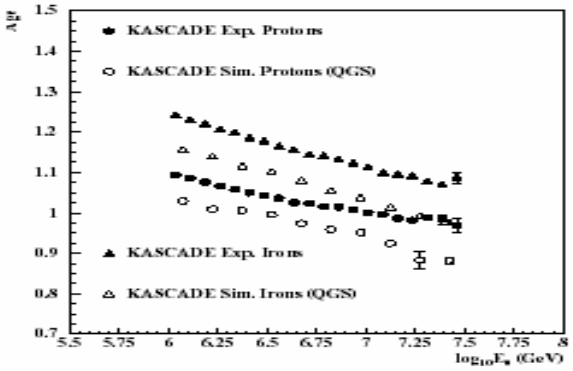


Figure 37. Shower age s versus E0 for simulated and experimental events (the primary energy is estimated by neural regression method). Used observables: N_e and N_μ^{tr}.

Table 25. Correlations between primary mass and energy with EAS observables

	A_0	E_0	N_e	N_μ^{tr}	s
A_0	-	0.00	-0.19	0.07	0.31
E_0	0.00	-	0.95	0.98	-0.47
N_e	-0.19	0.95	-	0.94	-0.58
N_μ^{tr}	0.07	0.98	0.94	-	-0.47
S	0.31	-0.47	-0.58	-0.47	-

6.2 NEURAL ESTIMATION OF MASS AND ENERGY

Demonstrated in previous sections unbiasness of the energy estimation, and rather high percent of the correct decisions in 3-way classifications (see Table 26 and Table 27) allow to use KASCADE array data for physical analysis such as the determination of energy dependence of the mass composition and energy spectra of three species of the primary CR flux in the energy range 10^{15} — $2 \cdot 10^{16}$eV (limitation to this very narrow interval is due to simulations available). Figure 38, Figure 39 display the relative abundance of three group of nuclei for different energies and the differential energy spectra respectively using N_e, N_μ^{tr} and s parameters. Figure 38 shows the energy dependence of the mass composition of the PCR flux for using all three observables: the relative abundance of light nuclei group is decreasing when approaching the

knee, while the behavior of intermediate and heave groups of nuclei is opposite. Figure 39 demonstrates the knee feature in the all-particle and light nuclei spectra, whereas to make a conclusion on the energy spectrum of the intermediate and heavy nuclei is difficult due to the relatively large proportion of misclassified events in these groups (see Table 26).

Table 26. Purity of classified events. Used observables: N_e and N_μ^{tr} and s

$P_{i\to j}$	j=p [%]	j=O [%]	j=Fe [%]
P	77	22	1
O	18	63	19
Fe	3	28	69

Table 27. Purity of classified events. Used observables: N_e and N_μ^{tr}

$P_{i\to j}$	j=p [%]	j=O [%]	j=Fe [%]
P	80	18	2
O	19	58	23
Fe	2	23	75

Figure 40 and Figure 41 present the relative abundances and the energy spectra of three group
of nuclei obtained by using N_e and parameters. It is easy to indicate the essential differences between these graphs and analogical Figure 38 and Figure 39, which were obtained using the *s* parameter additionally to N_e and N_μ^{tr} parameters. Although the same dependence of the relative abundances on the primary energy is observed, the relative proportion of heavy and intermediate nuclei is changed.

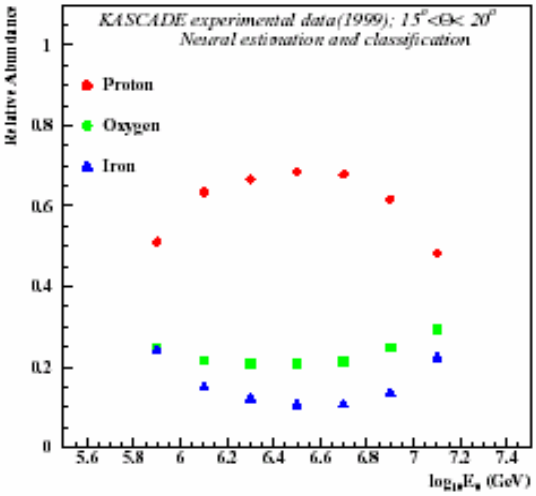


Figure 38. Relative abundance of different groups of nuclei as a function of energy. Used observables: N_e, N_μ^{tr} and s.

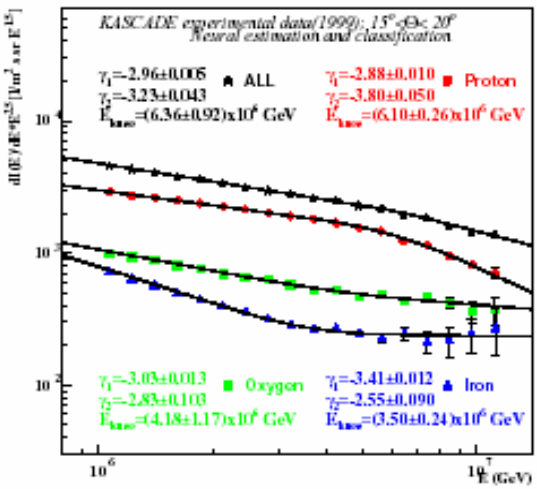


Figure 39. Differential energy spectra of three nuclear groups and all particle spectrum. Used observables: N_e, N_μ^{tr} and s.

In both cases the knee feature is clearly seen for the all-particle and light nuclei spectra, and the light nuclei spectra break is more sharp than for all-particle spectra. For the spectrum of intermediate nuclei group the difference of the slopes is negligible and there is no evidence allowing to stand the existence of a change in spectral indices.

An inverse knee observed in the spectrum of the heavy group of nuclei (most probably caused by not corrected spectra), is seen using both sets of observables, but the spectral indices

of these energy spectra and the knee positions differ. The main difference, which is impossible to neglect, is that removing the s parameter one obtains smaller absolute values of spectral indices below the knee. Additionally the knee positions are shifted to the lower energies for all-particle and light nuclei energy spectra, too. Fits of the energy spectra in both cases were done by the method described in (Sokhoyan, 1998).

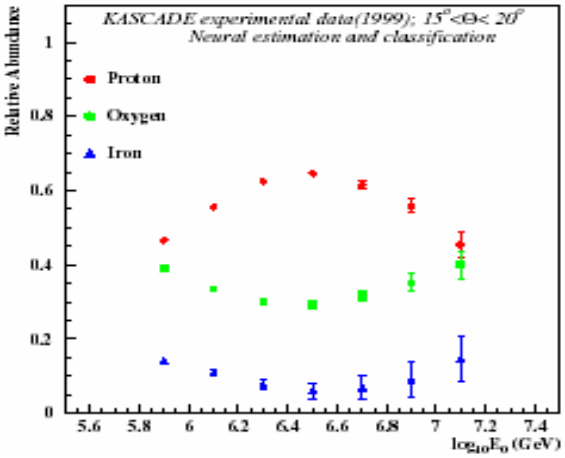


Figure 40. Relative abundance of different groups of nuclei as a function of energy. Used observables: N_e and N_μ^{tr}.

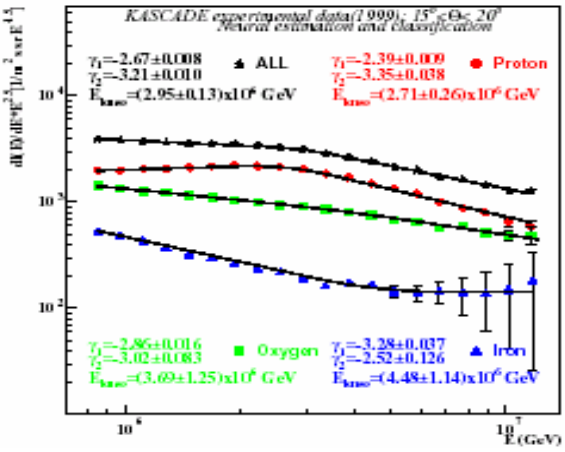


Figure 38. Differential energy spectra of three nuclear groups and all particle spectrum. Used observables: N_e and N_μ^{tr}.

Careful calculations of the misclassification rates and estimation of it's dependence on the primary energy will allow to make corrections to the energy spectra and to the relative abundances of different primary nuclei. Also the method and the model errors have to be taken into account in addition to the statistical errors, which are presented in Figure 39 and Figure 41.

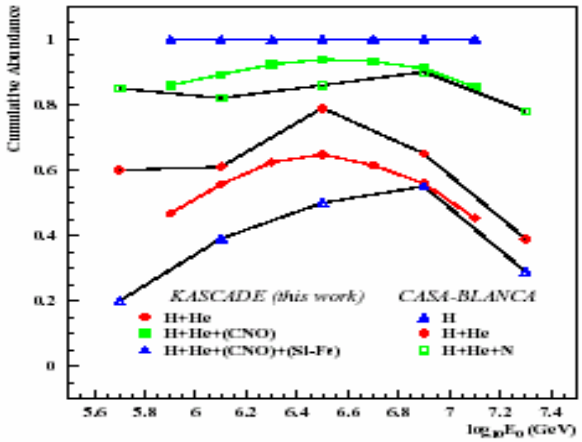


Figure 42. The cumulative abundances of different groups of nuclei

The results of the current investigation illustrate the possibility of obtaining the energy spectra of different species of PCR and to estimate the mass composition of PCR in the "knee" region using the advanced methodology of nonparametric statistical analysis (MCSI) of experimental data on event-by-event basis.

Figure 42 presents a comparison of the cumulative abundances of different groups of nuclei with recent results from CASA-BLANCA experiment (Fowler et al., 2000) using another experimental technique (measuring the lateral distribution of Cherenkov light in addition to the charged component of EAS) and completely different method of experimental data analysis. Similar trends and overall agreement in abundances of different nuclear groups are apparent.

6.3 THE "PURIFICATION" PROCEDURE

After estimating the misclassification rates, the possibility to select maximal pure* nuclear beams was investigated. The developed neural network technique allows a decrease in the contamination of misclassified events for each class of nuclei. Of course, efficiency** of classification is reduced at the same time. The purification was done in the following way: the NN performs a nonlinear mapping of EAS multidimensional characteristics to the real number interval [0, 1]. Particular class assignments for three way classification are subintervals [0. - 0.33], [0.33 - 0.66] and [0.66 -1] for first, second and third class respectively. The misclassification matrix for these decision intervals are given in Table 26 and Table 27.

If the NN is trained well enough to have generalization capabilities, NN output distributions for different classes are overlapping at subinterval boundaries. Therefore by the shrinking of the subintervals one can remove a large proportion of misclassified events, of course, simultaneously loosing some part of true classified events. From Figure 43, where the purity versus efficiency is plotted, one can see that the purity of proton and iron nuclei is larger than 90%, when the efficiency is still greater than 50%.

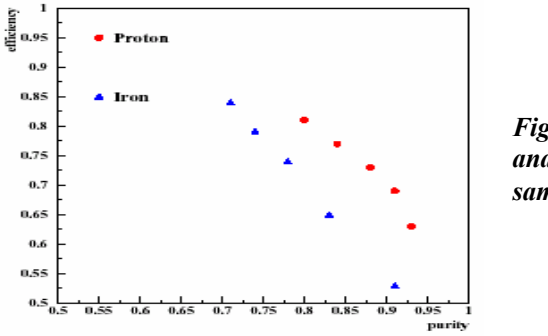


Figure 43 Event selection efficiency us purity for proton and iron events (obtained by classification of the control samples). Used observables: N_e and N_μ^{tr}.

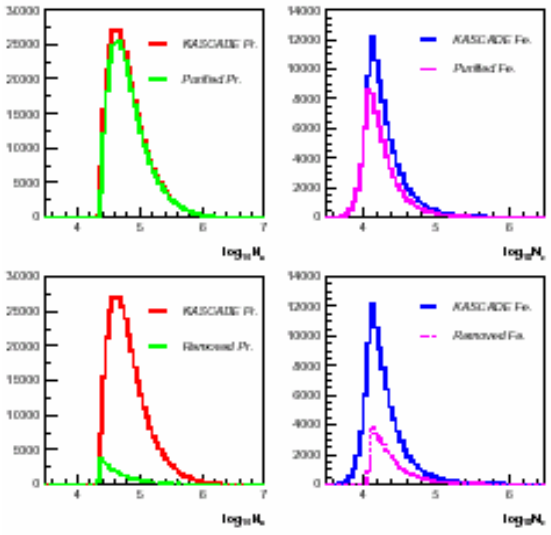


Figure 44. Feature distributions for classified and purified proton and iron

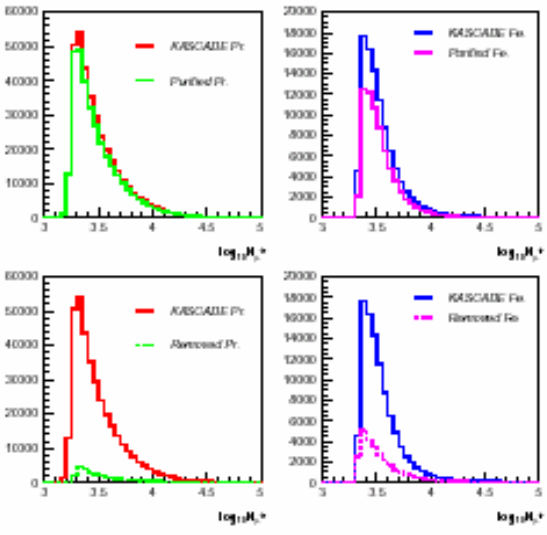


Figure 45. Feature distributions for classified and purified proton and iron samples

* purity: fraction of true classified events in actual number of events assigned to a given class
** efficiency: fraction of true classified events in total number of events of a given class

Purity estimates were obtained classifying 4000 control events (not used for the training) per class. For a given purity value the efficiency of proton events classification is always larger than the efficiency of iron events classification, thus, the purification of proton events turns out to be easier, than the purification of iron events. After preparing the "mononuclear" CR beams for investigation of the different nuclei interaction with the target (atmosphere) we have to check for possible biases introduced by the purification procedure. First of all we have to investigate how the shrinking of the decision interval for different classes (on the NN output) reflects on the one and two-dimensional distributions of EAS parameters (NN input). For this purpose we investigate and compare the "purified" and "rejected" EAS characteristics distributions as well.

On the Figure 44 and Figure 45 the distributions of EAS parameters of proton and iron classes are posted. In the above two pictures the "purified" distributions are compared with initial ones. The bottom pictures display distributions of the rejected (removed) events. The estimated purity of both proton and irons events is more than 90%. Initial purity of the proton events, as one can see from Figure 43 was 80%, and iron events -70% (most left points for both classes). It is easy to see that the events are removed over approximately the whole range of the features distribution support***. The shape and mean values of the distributions only slightly are changed after purification (see Table 28).

Table 28. The mean values of different parameters of purified and removed proton and iron samples

	μ^p_{age}	μ^f_{age}	μ^p_{Ne}	μ^f_{Ne}	$\mu^p_{N_\mu^{tr}}$	$\mu^f_{N_\mu^{tr}}$
ALL	1.066	1.231	4.821	4.295	3.489	3.543
Pure	1.061	1.236	4.834	4.255	3.486	3.542
Removed	1.134	1.218	4.671	4.396	3.529	3.547

Generally, the one dimensional distributions are not affected strongly by the mild cuts of NN output distribution tails, because the NN performs the nonlinear mapping of multidimensional input vector to one-dimensional output, therefore the cut applied to the NN output isn't linearly transferred to input parameters. More detailed investigation of Figure 44 allows to see that no events with smallest N_e are removed from iron sample as opposes to the proton events distribution. The opposite situation is observed for the N_μ^{tr} distribution, see Figure 45. This fact is more efficiently demonstrated on the Figure 46, which displays the N_e, N_μ^{tr} parameters two-dimensional distribution for the proton and iron events. It is obvious that we remove both proton and iron events from the boundary region, where misclassification is highly probable.

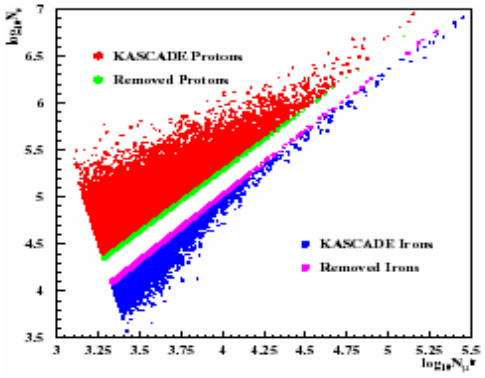


Figure 46. The two-dimensional distribution of the NN input parameters

*** Purification didn't reduced the number of selected "Proton" and "Iron" events significantly. Only 28004 events are removed from initial 357689 classified as proton, and 29325 from 100643 classified as iron

If the events would be removed uniformly over whole range of the features space, we would

still have overlapping (in the region of N_e-, N_μ^{tr} parameters space, where the probability of both classes is approximately equal), but the efficiency will be decreased. Figure 47 displays analogical distributions as in Figure 44 and Figure 45, but for the characteristic didn't used as NN input (EAS age parameter). As one can see the distribution is not affected by the shrinking the decision intervals on the NN output. Figure 48 demonstrates that events are removed over whole range of the distribution.

To check for possible systematic distortions introduced in EAS parameters by the purification procedure also the one dimensional statistical tests (for detailed description on statistical tests used see (Chilingarian, 1998) are performed for initial and purified samples.

The Table 29 displays the results of different tests for initial and purified proton and iron samples. The presented values are the probabilities of accepting the null hypothesis, that the samples are from one and the same population. If the tests give a small values of the probabilities one can reject the null hypothesis, e.g. there exists big difference between two examined samples.

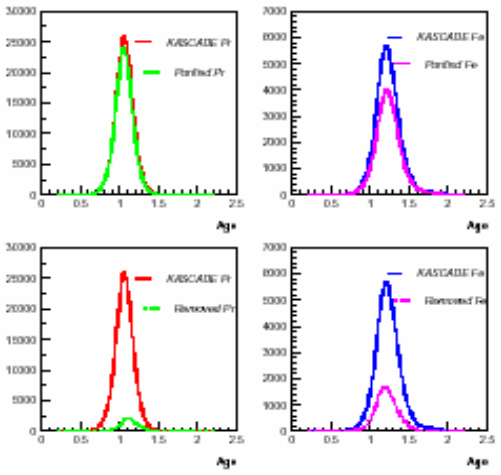


Figure 47. The distribution of age (not used for NN input) parameter for purified

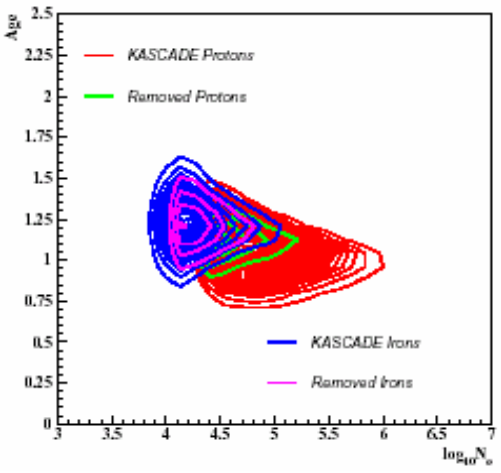


Figure 48. s vs N_e for removed proton and iron events

As one can see from Table 29 the test values are rather high and one can not reject the null hypothesis. So, the initial and purified proton and iron samples belong to the same population, which demonstrates that the purification did not introduced systematic distortions.

Table 29. The probability values of different tests for initial and purified proton and iron samples (t- Student, D- Kolmogorov-Smirnov, U- Mann- Whitnay)

	Initial and purified protons			Initial and purified irons		
	t	*D*	*U*	*t*	*D*	*U*
N_μ^{CD}	0.11	0.36	0.13	0.18	0.44	0.12
N_h	0.36	0.99	0.49	0.35	0.99	0.28
E_h	0.36	0.77	0.40	0.38	0.72	0.23
E_h^{max}	0.30	0.81	0.29	0.40	0.84	0.26
E_{tot}	0.38	0.89	0.46	0.36	0.74	0.25

The analogical results are obtained from multidimensional comparisons. Particularly, the *likelihood ratio, the log-likelihood function and Bayes risk* values (for more information see (Chilingarian, 1998)), for two-way classification of purified proton subsample into initial and pure proton classes, are presented in Table 30. These results support our thesis on soundness of the purification procedure.

Table 30. The classification results of pure proton subsample into initial and pure proton classes (features used N_e, and N_μ^{tr} age)

	Likelihood function	Classification score
ALL Pure	0.17 0.14	0.44 0.56
	Likelihood Ratio (ALL/Pure)	Bayes Risk
	0.92	0.47

All this procedures illustrate that no systematic uncertainties are introduced in to the features distributions while "purifying" the proton and iron samples, but there is still one parameter which may be affected by the cuts systematically. This parameter is the height of the first interaction (of the primary nucleus with atmosphere). Unfortunately this parameter can not be measured experimentally by surface array, therefore the distributions of this parameter for initial and purified proton and iron simples are presented for MC data (Figure 49). From this figure one can also see that no systematic cuts are introduced also to this parameter.

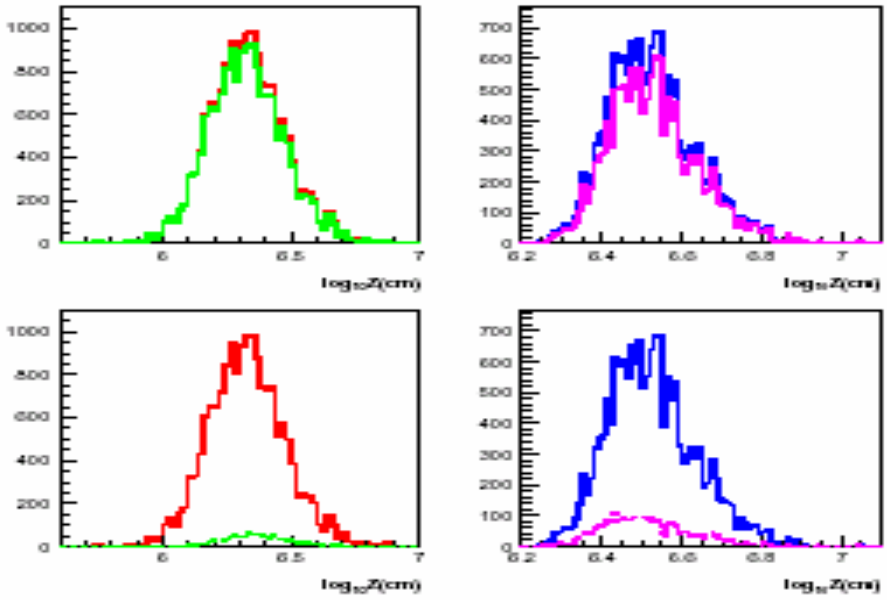


Figure 49. The first interaction height (Z) distribution for purified and removed proton and iron events

6.3.1 EXPERIMENTS WITH PURE (MONONUCLEAR) CR

In previous sections the possibility of making precise (~ 25% relative error) and unbiased estimation of primary energy, and accurate classification of primary particles into 3 categories in the range of 10^{15} — $2 \cdot 10^{16}$ eV was displayed. The characteristics of mononuclear CR beams obtained by the classification of the showers detected by KASCADE array along with the energy estimation accuracies were presented. High levels of purity of the proton and iron " beams" opens possibility of consideration high energy muons and hadrons

initiated by known primary with definite energy and detected by the KASCADE Central Detector (CD).

It is worth to mention that achieved purity of proton and iron beams (~ 90%) exceeds the estimate obtained in (Chilingarian et al., 1991) (70 — 80%). The relative error of primary energy estimation by the KASCADE array (~ 25%) almost coincide with expected one.

The comparisons of the QGSJET model (Kalmykov.et al., 1997) and KASCADE data are depicted in Figure 50, Figure 51, Figure 52, and Figure 53. The event selection procedure is equivalently done for the simulated and experimental data samples.

Due to the large statistical accuracy, the experimental distributions of the hadronic parameters of the showers originating from the primary protons demonstrate a rather smooth variation increasing with the energy. On the other hand the corresponding distributions, originating from the primary iron nuclei do less agree. However, in general, the overall dependences are in agreement with QGSJet simulations (Kalmykov.et al., 1997).

It is worth to note, that contamination of both proton and iron induced events by the intermediate nuclei has been ignored. The results in Figure 52-Figure 53 in contrast with (Antoni et al., 1999), present directly the energy dependence of EAS hadronic and high energy muonic content, those allowing physical inference and direct comparisons with alternative strong interaction models. In the present case the primary energy is determined event-by-event.

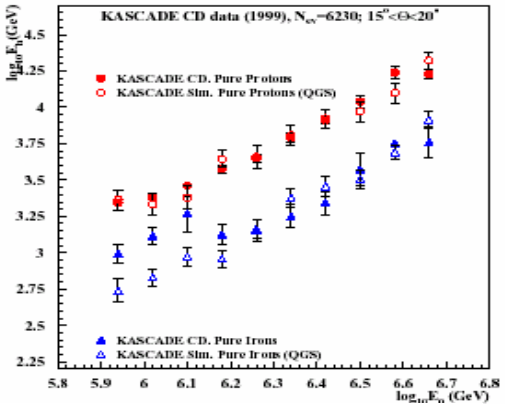


Figure 50. The Energy dependence of the Hadron Energy for the Proton and Iron Primaries

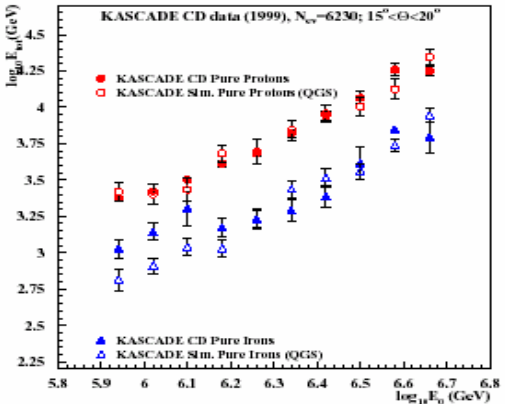


Figure 51. The Energy dependence of the Total Energy for the Proton and Iron Primaries

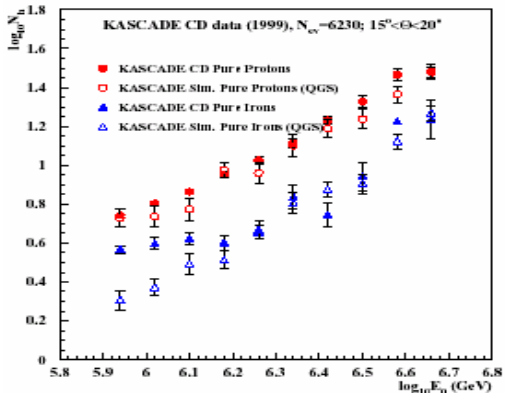


Figure 52. The Energy dependence of the Number of Hadrons for the Proton and Iron Primaries

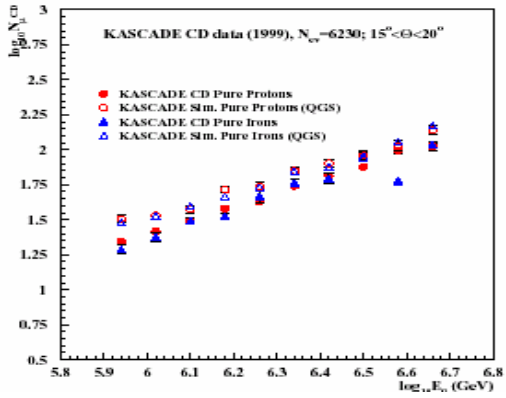


Figure 53. The Energy dependence of the Number of Muons for the Proton and Iron Primaries

If we take into account the limited efficiency ($\approx 90\%$) of the muon detecting facility, the agreement of the experimental data with the predictions is rather remarkable. It is interesting also to point on the approximately primary invariant form of the muon number energy dependence, as demonstrated on Figure 53.

The difference between figures presented in this paper and in (Vardanyan et al., 1999) is caused by use of different data samples and different selection criteria. In (Vardanyan et al., 1999) we use rather mild cuts requiring at least 1 hadron and 5 muons in KASCADE CD, and the number of events survived cuts were broadly distributed around CD, thus introducing uncertainties and biases in reconstruction of the shower core location, primary energy, true number of hadrons and number of hadrons registered by the CD.

6.4 CONCLUDING REMARKS

Modern arrays of particle detectors covering a large area are measuring different parameters of numerous secondary products of the primary cosmic ray interactions with the atmosphere. Only a simultaneous measurement of a large number of independent parameters in each individual Extensive Air Shower can yield reliable information to reconstruct the Primary Cosmic Radiation particle mass and its energy as well as the characteristics of strong interaction with atmosphere nuclei.

To make the conclusions about the investigated physical phenomenon reliable and significant, we develop a unified framework of statistical inference, based on nonparametric models, in which various nonparametric methods (Bayesian decisions, Neural Networks models, Feature extraction, etc.) are incorporated. Developed software are widely used for coherent solution of data analysis problems encountered in Astroparticle Physics experiments.

The current investigations are a first approach to obtain energy spectra of 3 species of primary flux and to use "mononuclear beams" for addressing the CR interaction problem on the event-by- event basis. The strong interaction model in general explains adequately almost all EAS parameters. If, as we demonstrate, surface array parameters are described with sufficient accuracy, the hadron information demonstrates slightly more discrepancy with model predictions, which could be caused by very poor statistics of CD events. The question of the significant disagreement of experimental and simulated events by the shower age parameter remains still open. Of course, better statistics for simulation and experimental data are required, other strong interaction models have to be tested as well. Nevertheless, we emphasize that the advocated approach is the only one which takes into account the shower fluctuations properly and is able to specify in a transparent way quantitative difference between model predictions and experimental data for different species of primary flux and various EAS parameters in wide primary energy range.

CHAPTER 7.
JOINT ANALYSIS OF THE ANI AND KASCADE EXPERIMENT

7.1 NUCLEAR - ELECTRON CASCADE IN THE ATMOSPHERE

The intensity of Extensive Air Showers (EAS) with fixed shower sizes Ne is assumed to decrease exponentially with increasing atmospheric depth of the observation level. This is considered to be due to the absorption of the particles of the EAS cascade following an exponential law

$$N_e(X) = N_e(X_0)\exp\left(-\frac{X-X_0}{\Lambda}\right), \quad \text{with } X \square X_0 \qquad (7.1)$$

X_0 is a definite initial atmospheric depth after the maximum of the longitudinal development where the number of (charged) particles is $N_e(X_0)$ and further decreasing exponentially, $N_e(X)$ is the
number of particles of the EAS at the slant depth $X[g \cdot cm^{-2}\}$. The quantity Λ controls the attenuation of particles of the individual cascade (Hayakawa, 1969) (size attenuation length).

The attenuation of the flux intensity of Extensive Air Showers is characterized by a related quantity λ_N (intensity attenuation length, absorption), which can be directly measured by cosmic rays detector arrays located at different atmospheric depths. Thus, measurements of the attenuation of the EAS intensity in the atmosphere are considered to be an interesting source of information about hadronic interactions, especially if extended to the ultrahigh energy region expected from the forthcoming LHC and TESLA accelerators. In addition due to the sensitivity of the cross sections to the mass of the primary, alterations of the attenuation length with the energy may be indicative for the variations of the mass composition. Measured results also imply tests of the energy dependence of the extrapolated cross sections used for Monte Carlo simulations. The investigations of the present paper are based on an EAS sample measured 1997- 1999 with the MAKET ANI array (Avakyan et al., 1993), (Hovsepyan, 1998) on Mt. Aragats station (Armenia) and registered for different angles-of-incidence in the zenith angle interval $\Theta = 0 - 45^{o}$. The data basis of the analysis can be enlarged by published data from KASCADE $\left(1046g \cdot cm^{-2}\right)$ (Klages et al., 1998) experiment.

We apply different procedures to deduce the attenuation. First we consider the degradation of the EAS flux with fixed shower size N_e with increasing zenith angle i.e. increasing atmospheric thickness of the shower development (characterized by the intensity attenuation length λ_N (Kristiansen et al., 1975)). Differently the technique of the constant intensity cut (GIG) (Nagano et al., 1984), (Gaisser, 1992) considers the intensity spectrum of EAS events and relates equal intensities observed at different atmospheric depths for obtaining cascade curves.

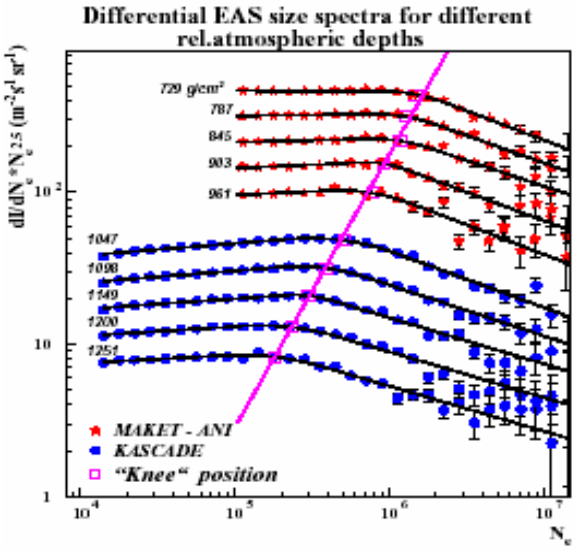
7.2 ANI-KASCADE EXPERIMENTAL SPECTRA

The experimental basis of the present investigations are measurements of shower size spectra in the knee region and their zenith-angle dependence performed with the MAKET ANI array of the Mt. Aragats Cosmic Ray Station (3200 m a.s.l.) in Armenia. Details of the measurements and the experimental procedures taking into account the detector response are

given elsewhere (Gharagyozyan, 1998), (Blokhin at al., 1999). For a detailed description of the knee region the traditional approximation with two different spectral indices below and above the knee-and consequent, defining the knee position as intersection of two lines in a logarithmic presentation, appears to be insufficient. Hence a more sophisticated method has been applied with parameterization of the slope of the spectra (see (Sokhoyan, 1998)). Table 31 contains the characteristics of the size spectra measured with the MAKET ANI installation, the changes of the slopes in the knee region (ΔN) . Spectral indices below (γ) and above (γ) the knee position are correspondent to the showers arriving in the zenith-angle range of $\Theta = 0 - 45^{\circ}$.

For the analysis of the zenith-angle dependence, the size spectra are determined in 5 angular bins of equal $\Delta \sec \Theta$ widths. The accuracy of the zenith angle determination is estimated to be about 1.5° (Gharagyozyan, 1998). A correction due to barometric pressure changes, which leads to small fluctuations of the atmospheric absorption, has not been made. Figure 54 displays the size spectra in graphical form. On the figure the KASCADE experiment (located at sea level) data (Glasstetter, 1998) are also plotted.

By the following fixed intensities of the size spectra (see horizontal dashed lines at Figure 56) the cascade curves (shower size dependence on the atmospheric depth for fixed primary energy) can be immediately reconstructed for several fixed flux intensities (primary energies) as shown in Figure 55. Note that the results in the range of the slant depth observed with the ANI array (near shower development maximum) deviate from the exponential decrease eq.(7.1). That is an important feature which can be revealed more clearly when combining spectra accurately measured on different altitudes. In the present paper we accept the model of exponential decrease of electron number in shower according to eq.(7.1). It is our interest to explore, if this assumption applied to the ANI and KASCADE data leads to consistent results.



Figurem 54. Differential size spectra for different zenith angles ranges observed with MAKET ANI array, compared with spectra reported by the KASCADE (Glasstetter, 1998) and the EAS TOP (Aglietta et al. 1999) collaborations Primaries

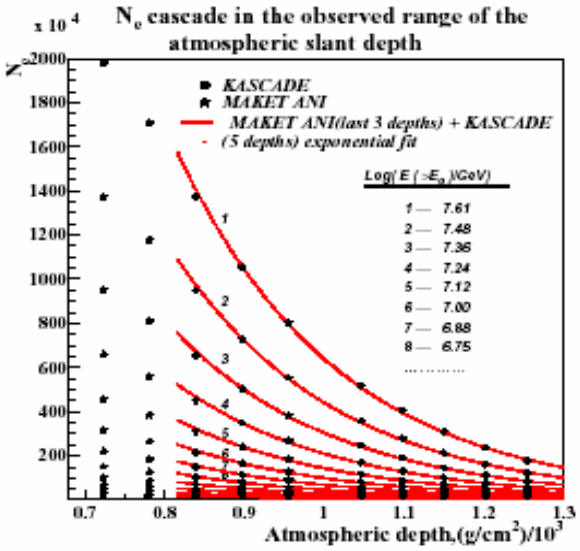


Figure 55. Ne cascade in the observed range of the atmospheric slant depth.

Table 31. The Summary of the Measured Differential Flux and Knee Region Parameters, Zenith

Angles Range of (0° — 45°), Altitude 3200m (intensities are given in $m^{-2}\,s^{-1}\,sr^{-1}$ unit)

γ_1	-2.50±0.002
γ_2	-2.87 ±0.020
$\Delta(N_{e_k})$	(8.13 ±0.037) •105 - (2.19 ±0.052) • 106
N_{e_k}	(1.37 ±0.04) •106
$I(N_{e_k})$	(2.14 ±0.11) •I0-13

7.3 PROCEDURES FOR INFERENCE OF THE ATTENUATION LENGTH FROM SIZE SPECTRA

We consider the differential and integral size spectra $I(N_e, X)$ and $I(>N_e, X)$, respectively. In addition to the basic assumption of exponential attenuation of N_e (eq.(7.1)) a power-law dependence of the size spectrum with the spectral index γ is adopted.

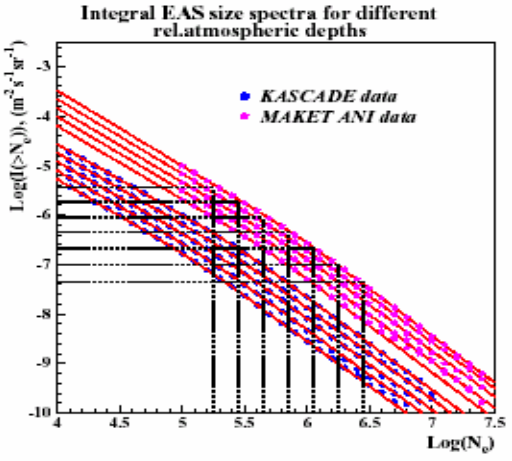


Figure 56. Integral size spectra for different zenith angles ranges observed with MAKET ANI array, compared with spectra reported by the KASCADE (Glasstetter, 1998): illustration of the procedures for absorption and attenuation length estimates.

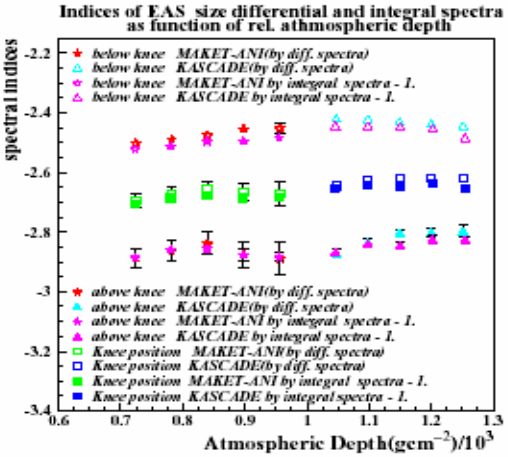


Figure 57. Indices of EAS size differential and integral spectra in comparison.

7.3.1 ATTENUATION OF THE INTENSITY OF FIXED N_e: ABSORPTION LENGTH

For different fixed values of shower size N_e, on different depths in the atmosphere (or/and different zenith angles of incidence), from measured spectra we obtain several values of corresponding intensities from the equivalent depths from 700 till 1280 g•cm^{-2} (see Figure 56). By fitting the depth dependence of the intensities by the straight line (in logarithmic scale) according to equation:

$$I(N_e, X) = I(N_e, X_0)\exp\left(-\frac{X-X_0}{\lambda_N}\right) \qquad (7.2)$$

we obtain the estimate of the absorption length λ_N. The absorption length can be estimated both by integral and differential spectra.

7.3.2 CONSTANT INTENSITY CUT (CIC)

The basic idea of this procedure is to compare the average size of showers which have the same rate (showers per $m^2 \cdot s \cdot sr$] in the different bins of the zenith angle of shower incidence and different slant depth, respectively (Nagano et al., 1984).

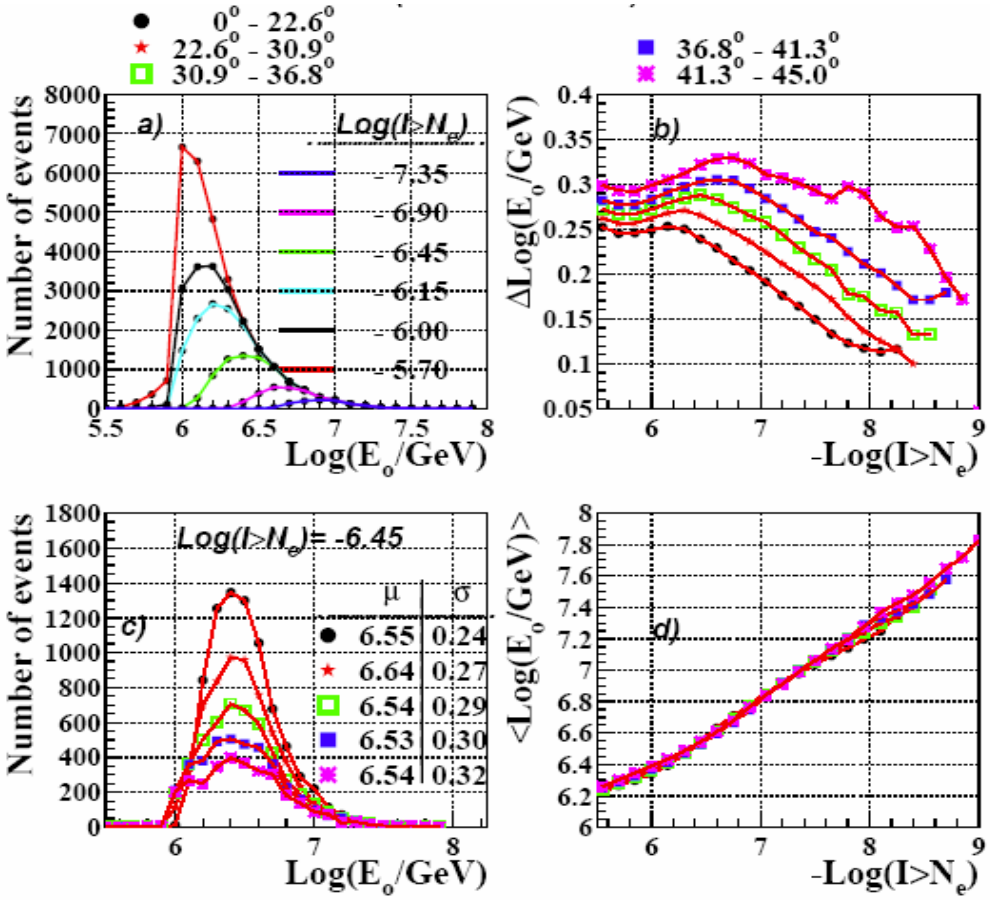


Figure 58. (a-d) Fluctuations of energy for different integral intensities and zenith angles

Considering two different depths in atmosphere X_1, X_2 $>X_0$ the expressions of differential intensities $I\,(N_e\,, X\,)$ has the form

$$N_e(X_1)^{-\gamma}\exp\left[-(\gamma-1)\frac{X_1-X_0}{\Lambda}\right] = N_e(X_2)^{-\gamma}\exp\left[-(\gamma-1)\frac{X_2-X_0}{\Lambda}\right] \tag{7.3}$$

With simple transformations we obtain:

$$\Lambda_{diff}(I) = \frac{\gamma-1}{\gamma}\frac{X_2-X_1}{\ln\left(\frac{N_e(X_1)}{N_e(X_2)}\right)} \tag{7.4}$$

The attenuation lengths, obtained by integral spectra do not depend explicitly on spectral index:

$$\Lambda_{\text{int}}(I) = \frac{X_2-X_1}{\ln\left(\frac{N_e(X_1)}{N_e(X_2)}\right)} \tag{7.5}$$

Practically the estimate of the attenuation length is obtained by fitting the Ne dependence on the depth in atmosphere by the straight line according to the equation (7.1). The sequence

of Ne values is obtained according to the fixed values of the flux intensity, selected from the interpolation curves of the differential or integral size spectra. The value of Λ_{diff}(I) is estimated by Selecting equal intensities (≈ primary energies) corresponding to different shower sizes N_e and different depths. The intensity values from 10-9 to $5.\bullet10^{-6}$ were used for CIC method.

7.3.3 WHICH PRIMARY ENERGIES WE SELECTED BY CIC METHOD?

The tacit assumption that by cutting on the same EAS size spectra intensities, we chose approximately constant primary energy was checked during pilot study with simulated samples. CORSIKA 562 code (Ostapchenko) was used for simulation of EAS traversed through atmosphere; MAKET ANI response function was taken into account (Blokhin at al., 1999). Mixture of the primary nucleus and partial flux intensities were chosen according to (Ter-Antonyan). Approximately 300,000 events were used to construct Integral spectra in 5 angular bins. Figure 58 a) demonstrates the distribution of the primary energies of events falling in the size spectra bins correspondent to different horizontal cuts of spectrum for the first zenith angles interval (0° - 22.6°) (see also 7.3). The variances of the energy distributions, correspondent to different intensity cuts are decreasing with increasing of energy.

Figure 58 c) depicts the "selected energies" distributions for fixed EAS flux intensity and different angles of incidence. The unbiasness of distributions mode proves that by "cutting" spectra we are choosing EAS events, corresponding to approximately same primary energies. Figure 58 b) demonstrates the relative errors of assigned energies. Increasing of errors with increasing of zenith angles is evident. On altitude 3200m. primary energy variance of "cutted" events didn't exceed 50%.

Figure 58 d) shows dependencies of spectra intensities on correspondent mean energies for five zenith angle intervals. The approximate linearity of E-I relation for all angular bins and majority of energies is observed, therefore we'll use primary energy estimate obtained from this relation.

7.3.4 THE RELATION BETWEEN THE ABSORPTION AND ATTENUATION LENGTH

We consider the quantity $I(N_e,X)dN_e$ - the number of EAS at the depth X which comprise N_e to $N_e + dN_e$ particles fallen in $(N_e\text{-}N_e+dN_e)$ interval:

$$I(N_e,X)dN_e \sim N_e^{-\gamma} exp\left[-(\gamma-1)\frac{X-X_0}{\Lambda}\right]dN_e \qquad (7.6)$$

With eq.5.2 we obtain:

$$\Lambda_{diff}(N_e) = (\gamma(N_e)-1)\lambda_N \qquad (7.7)$$

where, $\gamma(N_e)$ is the differential size spectra index (therefore, for λ estimation we have to usethe N_e intervals where γ is not changing dramatically). For the integral spectra:

$$\Lambda_{\text{int}}(N_e) = \gamma(N_e)\lambda_N \qquad (7.8)$$

where, $\gamma\ (N_e)$ is integral size spectra index.

For the evaluation of the inelastic cross section and for comparison of the three methods described above we propose to use the calculated values of the attenuation length Λ (instead of using absorption length λ_N). The attenuation of the number of particles in the individual cascade is more directly connected with the characteristics of the strong interaction and does not depend on the parameters of the cosmic ray flux incident on the atmosphere. In its turn the absorption length, i.e. the attenuation of the CR flux intensity, reflects also characteristics of the primary flux and is dependent on the change of the slope of the spectra.

7.3.5 ESTIMATE OF THE INELASTIC CROSS SECTION

The inelastic cross sections of the primary nuclei with atmosphere nuclei is related by (Nagano et al., 1984):

$$\sigma_{A-air}^{inel}(mbarn) = \frac{2.41 \cdot 10^4}{\lambda_A(g \cdot cm^{-2})} \qquad (7.9)$$

where A denotes the primary nuclei. The quantity λ_A is the interaction length of the A-nucleus in the atmosphere (note: in some publications the interaction length is denoted by λ_N, where N is primary nuclei, in contrast in this paper N is reserved for the shower size). The interaction length λ_A is related with the absorption length λ_N by

$$\lambda_A = K(E) \cdot \lambda_N \qquad (7.10)$$

The $K(E)$ coefficient reflects peculiarities of the strong interaction model used for simulation. The value of the parameter K has to be determined by simulations of the EAS development in the atmosphere and by detector response. Such studies require the development of procedures for the selection of EAS initiated by primaries of a definite type (see example in (Chilingarian et al., 1999), (Vardanyan et al., 1999)).

7.4 ESTIMATION OF THE ATTENUATION LENGTH

The mean values of the attenuation lengths obtained by various methods from data of the ANI and KASCADE installations, as well as for the joint ANI & KASCADE data by the differential Λ_{diff} and integral spectra Λ_{int} are compiled in the Table 32 and Table 33.

The discrepancies of alternative estimates of the attenuation length reflect the inherent uncertainties of the methods and the statistical errors, as well as the fluctuations of cascade development in the atmosphere. The energy dependence of the inelastic cross section and possible changes in mass composition also are reflected in the curves.

Table 32 Attenuation lengths for the data from the MAKET ANI and KASCADE installations estimated by the CIC method from differential and integral size spectra

Min.depth	*MAKET ANI*		*ANI + KASCADE*		*KASCADE*	
$_0, g \cdot cm^{-2}$	Λ_{dif}	Λ_{int}	Λ_{dif}	Λ_{int}	Λ_{dif}	Λ_{int}
700	250 ± 26	246 ± 16	211 ±6	206 ±9	—	—
758	242 ± 33	231 ±20	205 ±6	200 ±8	—	—
816	228 ± 50	221 ±29	200 ±7	195 ±8	—	—
1020	—	—	—	—	186	183

Table 33 Attenuation lengths for the data from the MAKET ANI and KASCADE installations estimated by the recalculation from the absorption length for differential and integral size spectra

Min.depth	*MAKET ANI*		*ANI + KASCADE*		*KASCADE*	
X_0,*g•cm* $^{-2}$	Λ_{dif}	Λ_{int}	Λ_{dif}	$h\,\Lambda_{int}$	Λ_{dif}	Λ_{int}
700	243 ± 20	254 ± 21	211 ±6	206 ±9	—	—
758	231 ±17	242 ± 16	205 ±6	200 ±8	—	—
816	221 ± 23	237 ± 24	200 ±7	195 ±8	—	—
1020	—	—	—	—	180	186 ±7

As we can see in Figure 55, the values corresponding to the minimal equivalent depths of used MAKET ANI data, deviate significantly from the exponential dependence. The observations reflect the flattening of the cascade curve just after the shower maximum, expected at the altitudes 500 - 600 g • cm'2. Therefore, due to these features the attenuation lengths calculated by MAKET ANI data appear to be significantly larger than those derived for the KASCADE data (Table 32, Table 33).

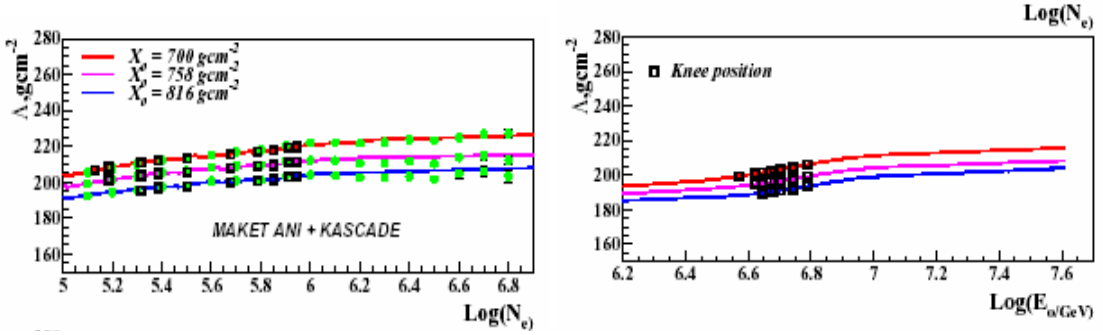


Figure 59. Attenuation length obtained by joint analysis of the MAKET ANI and KASCADE data.

Consequently, for the combined analysis of the KASCADE and ANI data we omitted the first and the second zenith angle bins of MAKET ANI and calculate the attenuation lengths by the remaining 9 (minimal equivalent depth 758 g • cm^{-2}) and 8 (minimal equivalent depth 816 g • cm^{-2}) angular bins. The dependences of estimated values of attenuation length on the shower size and flux intensity for different amount of the angular bins used, are displayed in Figure 60 and Figure 61. The attenuation length estimates obtained from the differential and integral spectra agree fairly well. The results of both GIG and recalculation from absorption length agree within the error bars. The results obtained by the "attenuation of knee position" are larger for MAKET ANI and KASCADE.

By taking the advantage of the precise measurements of the cascade curves by KASCADE and ANI detectors we fit joint data with one decay parameter for the first time (see Figure 59). The "knee" position, obtained by fitting of the size spectra, also are posted on the picture.

There is a concentration of the knee positions on the curve showing the dependence of the attenuation of the flux intensity (≈ primary energy). In its turn, the curve displaying the dependence of the attenuation length on the shower size demonstrates a rather large dispersion of the "knee positions". Interpretation and physical inference based on the obtained results will require detailed simulation of cascade development in atmosphere and detector response now underway.

Experimental studies of EAS characteristic like the depth of the shower maximum X_{max}, the elongation rate $dX_{max}/dlog_{10}E$ and the attenuation length Λ are of particular

importance, since they map rather directly basic features of the hadronic interaction. However, the interpretation of these quantities in terms of hadronic cross sections cannot bypass the necessity of detailed calculations of the shower development. Nevertheless, these type of EAS quantities, if compared with Monte Carlo simulation results, provide stringent tests of the interaction model used in simulations.

The recent results of various experimental installations are sufficiently accurate to enable relevant studies of this kind, and combining the data from arrays situated on different altitudes (like MAKET ANI and KASCADE) allows a large span in the atmospheric slant depth for reconstructing the longitudinal development of the EAS.

By use of methods to isolate different primary groups ("pure nuclear beams") of the size spectra (Vardanyan et al., 1999), (Gharagyozyan et al., 1999), these kind of interaction studies would get of extreme interest.

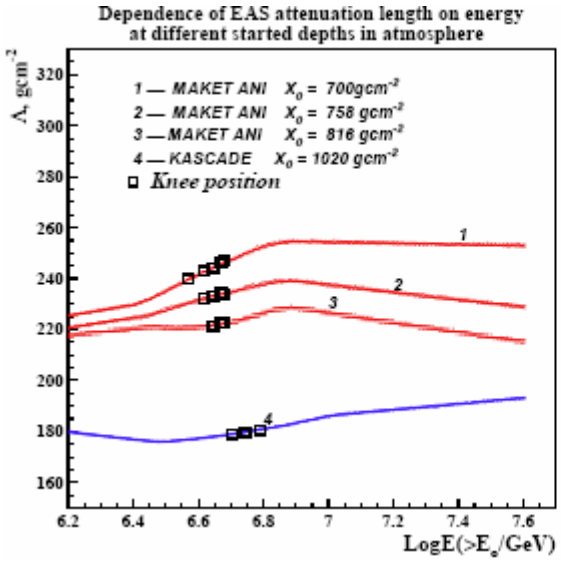


Figure 60. Attenuation Length dependence on Spectra Intensity (Primary Energy)

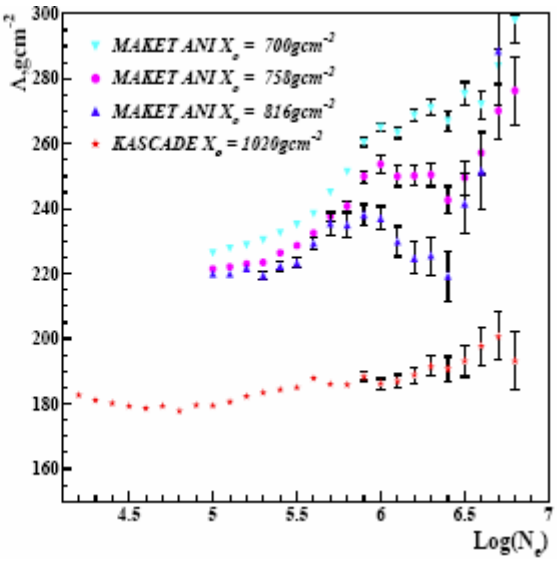


Figure 61. Attenuation Length dependence on the Shower Size N_e.

What ANI is intended to do.

The ANI package represents an unified methodology of multivariate data analysis and physical inference from the high energy cosmic ray physics experiments. The main procedures of ANI package are as follows:

- Bayes and Neural Network (NN) classification and decision making;
- Adaptive modes of class conditional probability density estimation;
- Calculation of the Bayes risk.
- Resolving of the distribution mixture statistic models.
- Determination of method errors using Bootstrap model.
- Library of Random Search methods of NN training.
- Interdependent system of various NN training scenarios.
- Best feature subsets selection and initial dimensionality reduction.
- Evolutionary algorithms of Clustering.
- Scanning of multivariate distributions to investigate and visualize embedded non-

trivial structures;

- Nonparametric (Bayes and NN) estimation of regression function;
- Cross validation of obtained estimates and Stopping rules in NN training.
- Committee principle of final decision.
- Continuous comparison of alternative nonparametric techniques.
- 2-dimensional Images Pattern Recognition.
- Incorporation of the hardware neural accelerator.

The developed methods were applied for the EAS data analysis. The data from world biggest KASCADE experiment were treated on event-by-event basis. The main physical results obtained from KASCADE data analysis are the following:

1. Classification of the PCR into three nuclear groups was performed, the relative abundances of different nuclear groups on the primary energy were obtained;

- Accuracy of the classification into light, intermediate and heavy groups of nuclei is - for light group ~80%, for intermediate group ≈60% and for heavy group ≈75%,
- Relative abundances in the "knee" region (2 —4 x 1015 eV) are as follows: light group (H,He) ~0.6 ± 0.02 , intermediate group (C, N, O) ≈ 0.32 ± 0.06 and heavy group (Si-Fe) ≈ 0.08 ± 0.04,
- Spectral exponent index of the "light" nuclei flux demonstrate slight increase before the "knee" and turn to decrease after the "knee", while there was no evidence for the decrease of the "intermediate" and "heavy" nuclei flux spectral indexes above the "knee" energies.

2. Estimation of energy of the PCR in the range of 10^{14} — 10^{16} eV is performed and for the first time the energy spectra of three spices of PCR are obtained;

   Relative error of the energy estimation is ≈25%,

   "Knee" feature is observed in all-particle and "light" nuclei spectra.

3. Using developed NN methodology the possibility to obtain pure CR beams and to make experiments with such beams was demonstrated:

   Achieved purity of the "light" and "heavy" nuclei is more than 90% while the efficiency of the selection is kept >50%,

   General agreement of dependencies in real data and QGSJet simulations is demonstrated for the light group of nuclei.

   It is also worth to mention achievements in designing combined hardware-software systems for accelerating various Neural Net applications and constructing fast "intelligent" trigger for world biggest MAGIC Atmospheric Cherenkov Telescope.

   New algorithms were realized in Fortran77 and C programming languages, the graphical interface for ANI was written using TCL 7.6 script and TK 4.2 toolkit, (for details on ANI and graphical user friendly interface see the Annual

   Report 1998) which are available on most of UNIX and LINUX platforms; ANI program package User Guide with tutorials was prepared according to the project partners request;

Tests and benchmarks with MiND PCI board of SAND/1 neurochips delivered by project collaborators - Institute for Electronics and Data Processing, Research Center Karlsruhe, were done;

New drivers for MiND PCI board for Linux OS was developed, optimized and tested;

Graphical user interface was written using gtk+ 1.2.8 to provide an easy access to the SAND kit and its drivers under Linux OS. Space Weather Forecast.

7.5 CONCLUSION

The most general framework in which to formulate solutions to physical inference problems in High Energy Cosmic Ray (CR) Physics experiments is a statistical one, which recognizes the probabilistic nature both of the physical processes of cosmic radiation propagation, and of the form in which the data analysis results should be expressed. To make the conclusions about investigated physical phenomenon more reliable and significant we have developed a unified theory of statistical inference, based on nonparametric models, in which various nonparametric approaches (including Neural Networks models) are implemented and compared. In this context in is necessary to mention that we consider the Neural information technology not as " black box", but as an extension of conventional nonparametric technique of statistical inference.

The ANI program package is the software realization of our concept and appropriate tool for the physical inference. During last 3 years ANI package was updated and intensively used for comparisons of different nonparametric techniques and for data analysis of CR physics experiments.

The challenge of Extensive Air Shower (EAS) investigations is to infer reliably from the observed data the development of cascade in atmosphere and its dependence from the mass and energy of primary particle. One of the important tasks in experiments (often underestimated) is to understand the performance and response of the detectors with sufficient accuracy for a reliable reconstruction of meaningful shower parameters, starting from measuring energy deposits and time signals in single detector elements.

It is specific for cosmic ray experiments that there is no possibility to calibrate the detector output by use of a suitable test beam of known quality like in accelerator experiments. In the theoretical analysis of the measured and reconstructed EAS parameters the major uncertainties of systematic character arise from the lack of our knowledge of the total cross sections and of details of particle production for high - energy hadronic interactions with small momentum transfer. The proton collider results have to be extrapolated over many orders of magnitude, to forward angle emission and to the case of complex nuclear projectiles and targets.

Standard Neural Network architectures, such as FFNN always have too large parameter space; therefore they are prone to overfilling. While the network seems to get better and better (the training, apparent error decreases), at some point during continues training it actually begins to get worse again (the control, generalization error increases). The standard cross-validation technique is checking the generalization error on the independent control sample and stop when generalization deteriorates. However, the real situation is a lot more complex. Real generalization curves almost always have more than one local minimum. Thus, it is impossible in general to tell from the beginning of the curve whether the global

minimum has already reached or not, i.e., whether an increase in the generalization error indicates real overfilling or is just intermittent.

The choice of the Neural Network learning algorithm depends on many problem dependent factors, like the complexity of the problem, availability of the training set, real-time requirements, the cost of the losses (required quality of the solution). Hence it is not possible to give an unique decision on what strategy to follow. Nevertheless, the goal of each training algorithm is to achieve a good generalization performance relatively easy and fast (low resources requirements and efficiency of algorithm).

Different modes implemented in ANI package (single, neuron, multi] give a possibility to perform different training strategies, obtaining quite different points of multidimensional space of neural net weights, and allow to investigate symmetries of nontrivial local minima in this space.

The low level trigger (LO) rate from planned Cherenkov Atmospheric Telescopes (ATC) of new Generation as MAGIC, is dominated by Night Sky Background (NSB) and the Photomultiplier (PMT) afterpulsing and at low thresholds trigger rate can reach 1 MHz for each channel. For the practical telescope operation the trigger rate must be in the region less than 100 Hz. Therefore, 1 MHZ rate has to be reduced down to the few of kHz and tenths of Hz by the higher level triggers (LI and L2).

LO can be derived from four-fold majority coincidence of all pixels. LI can be constructed using sophisticated coincidence chime, exploiting the topology of hits pattern requiring close-

patched configurations. To obtain short trigger decision times (≈100nsec) the Altera 10K family programmable logic devices can be used. LI trigger will provide reduction of the trigger rate down to several units of kHz. For further reduction we propose to use neurochip SAND as fast "intelligent" trigger. L2 Pattern Recognizing Trigger (PRT) can help to reject muon and hadron backgrounds which at present is only possible off-line.

For the first time we intended use whole pixel information as input of analysis in contrast with significant reduction to the second moments of pattern (Hillas parameters) used previously.

Trained combined hard/software networks will provide possibility to enlarge signal-to-noise ratio. For working with such huge inputs new network training algorithms and powerful training accelerators are implemented in ANI package.

The primary aim of EAS studies is to determine the mass composition, energy spectrum and anisotropy of arrival directions of the primaries. The intermediate-energy region (10^{15} — 10^{16}eV), where pronounced structures in the charged particle spectra are observed more than forty years ago, is of particular interest as the Supernova (SN) driven stochastic shock acceleration mechanism is expected to fail in this region. However, the direct evidence is lacking to support strongly this models.

Measurements of the energy spectra of the individual species of the Cosmic Ray flux can provide a detailed description of the structures in knee region, revealing the rigidity cut off, inherent to Fermi acceleration, or peaks, corresponding to the individual elements fluxes from nearest SN.

Obtained using CORCIKA data base and ANI training modes energy and primary estimators were applied for analysis of experimental data of ANI and KASCADE

experiments.

Combining data from two installations, located on 3200 m and on sea level, enlarged significantly amount of physical information on the propagation of cascades in the atmosphere. Very broad slant depth from 700 — 1300g/cm2 allows to investigate in details the cascade curves and estimate elongation and absorption lengths with accuracy, never obtained with single installation. New methodology of mutual analysis of the size spectra brings new interesting physical results and new step to solving long standing problem of the origin of the high energies cosmic rays.

Selection of the light "nuclei" performed by ANI MAKET installation and "light", "intermediate" and "heavy", performed by KASCADE installation allows at first time to investigate partial size spectra of primary cosmic ray radiation. The first conclusions from the comparative analysis of these data are confirming hypothesis of the Supernova Remnants as source of cosmic rays and rigidity dependent Fermi acceleration in supernova explosion shock waves.

CHAPTER 8. CHARACTERISTICS OF THE LIGHT AND HEAVY MASS GROUP COSMIC RAY FLUX OBTAINED BY MAKET-ANI DETECTOR

8.1 COSMIC RAYS IN THE VICINITY OF EARTH

Cosmic Ray (CR) flux incident on terrestrial atmosphere consist mostly of protons and heavier stripped nucleus accelerated at numerous galactic and extragalactic sites. Most exciting problem connected with cosmic rays is exploration of a particular astrophysical accelerating source. Due to the bending in galactic magnetic fields charged particles lost information about parent sites during long travel and arrive at Earth highly isotropic. Only stable neutral particles i.e. roentgen, gamma quanta and neutrinos travel directly from sources and reveal exotic celestial objects and violent processes of their production. Orbiting telescopes and spectrometers as well as ground based Atmospheric Cherenkov Telescopes (ACT) and neutrino detectors opened new windows to Universe detecting in unprecedented details spreading of heavy elements during supernovae explosions, ejection of the relativistic jets from black holes and many other phenomena known in last century only from science fiction.

New paradigm in astrophysics research consists in simultaneous detection of celestial objects in radio, optical, X and gamma rays. Variety of compatible measurements gives enough information for building realistic models of physical processes of supernovae explosion accompanied by the gamma-ray bursts, of accretion discs interactions with super dense objects and finally about the evolution of Universe itself. In this concern adding information on particles of highest energies will significantly enlarge information on most violent processes in Universe and on operation of the biggest particle accelerators in space.

Galactic cosmic rays can't map the objects where they burn, so, only integrated information from all sources are available from measurements of cosmic ray fluxes nearby Earth and on Earth surface. This information contains in the shape of the energy spectra of cosmic rays, in the mass composition and its energy dependence, and in anisotropy of CR arrival.

Space-born spectrometers at ACE space station, AMS detector on Space Shuttle as well as numerous balloon-born detectors precisely measure fluxes of different isotopes up to energies of 10 TeV.

Particle fluxes follows an overall power-law $I(E) \sim E^{\gamma}$ with spectral index of $\gamma \sim -2.7$, therefore because of very weak fluxes of the CR of heist energies and due to very strict restrictions on the spacecraft payload it is extremely difficult to get reliable information on particle fluxes above 10 TeV from space-born spectrometers and calorimeters. Although recent successes with long-lasting new technology balloon flights give hopes that precise information on particle spectra up to several hundreds of TeV will be available soon.

Recently, so called kinematical method (Adams et al., 2001) was proposed, using thin (about 10 g/cm^2) target and silicon coordinate and charge detectors for precise detection of the charge and emission angles of secondaries produced in an inelastic interaction of primary nuclei. The angle distribution of the particles produced in the target carry information about the energy of a primary particle. This technique does not require total release of energy like

ionization calorimeter and could be made very light. One-Ear flight on Space Station of

such device will provide data up to several PeV with 0.2 units of charge resolution.

No space experiment in the PeV region currently is funded and at least in current decade data will be assessable only from the Extended Air Showers (EAS) initiated by the “primary” ion triggering particle generation chain reaction in terrestrial atmosphere. A variety of physical processes during travel of the relativistic cloud of “secondary” particles to the earth surface gave rise to different experimental methods, aiming to reconstruct the particle type, trajectory and energy.

Signatures of the primary particles are microwave radio signals, fluorescent light, cherenkov light, electrons, muons, neutrons and hadrons reaching Earths surface, muons detected deep underground. Intensity and correlation matrix of each of combination of mentioned signals carry information on primary, but due to highly indirect nature of the experimentation, only some very robust characteristics of cosmic ray flux of PeV and higher energy primaries were unambiguously established till now. First of all it is all particle energy spectra, reconstructed from so-called size spectra measured by plastic or liquid scintillators (so called particle density detectors), distributed on the surface up to several tens and hundreds of thousands of square meters with detector coverage of very few percent or tenths of percent. Assuming definite shape of EAS electron lateral distribution function and measuring density of electrons on some rectangular or circular grid of distributed density detectors, and using standard minimization technique the overall number of electrons (shower size) could be determined.

Simultaneously measuring by the system of distributed “fast timing detectors” the time delay of arriving of particles from inclined “shower disc”, the zenith and azimuth angles of the shower core can be calculated (very good estimate of the primary particle angles of incidence on the terrestrial atmosphere).

Shower size is correlated with particle energy, but also with several unknown parameters such as particle type and height of first interaction. The functional form of size-energy dependence introduce additional uncertainty, because it is obtained from particular model of strong interaction of protons and ions with atmosphere nuclei, and at PeV energies there are no accelerator data to check this models. Different approximations of models fitted with manmade accelerator data at lower energies give significantly different results at higher energies.

Nevertheless, during last 50 Ears some important characteristics of spectra were established during intensive measurements with EAS surface detectors (for the list of detectors and their operational characteristics, see (Haungs et al., 2003). The most striking feature of spectra is approximately constant power index in energy range more than a decade. Power index slightly changed from value $\gamma \sim -2.7$ to value $\gamma \sim -3.0$ at 3-4 PeV (“knee” or suppression of spectra) and it is another important and well established feature. Some authors (Nikolsky et al., 2000), (Nikolsky et al., 2003) claim that this “knee” is feature only of size spectra reflecting some peculiarities of EAS propagation and interaction in the atmosphere and flux of cosmic rays incident on atmosphere can be described by constant power index and that the CR origin is of extragalactic nature. In paper (Stenkin, 2003), the “knee” is treated as consequence of the shower size reconstruction method only. The difference between pure electromagnetic showers and those having survived hadron "cores" is cause of the “knee”. Another, very interesting approach is connected with enigma of supernovae implosion and collapse. In (Plaga, 2002) the cannonball model of the supernovae

explosion (Dar et al., 1999) was proposed as source of the cosmic rays. The blobs of plasma with mass of Earth are ejected from poles of supernovae at nearly light-speed. The population of such plasmoids filling the Galactic halo is responsible for the acceleration of major part of hadronic cosmic rays with energies till another feature of all particle spectra, so called "ankle" occurred near 10^{18} eV.

In contrast, the "standard" model of CR acceleration name the Supernovae remnants (SNR) as a major source of CR. The detected non-thermal radio emission from Supernovae Remnants (SNR), which led to the natural assumption of the presence of accelerated electrons, made SNR the main candidate engine for particle acceleration (Koyama et al., 1995). Recent CHANDRA measurements of the X-ray distributions, obtained to very small spatial scale (Long et al., 2003) point on very large effective magnetic field of $100\mu G$ in the SN 1006. In the (Berezhko et al., 2003) authors conclude that such a large field could be generated only due the nonlinear interactions of the accelerated protons and stripped heavier nuclei with self- generated Alfven waves in a strong shock. Therefore, the SN 1006 data confirms acceleration

of the nuclear component at least till several units of 10^{14} eV. Gamma-ray pulsars usually located near the

SNR center are another candidate for the cosmic ray acceleration (Bednarek et al., 2002). As mentioned in (Bhadra, 2003) pulsar accelerated cosmic rays are expected to have a very flat spectra. Therefore, the impact of the nearest pulsar to energies higher than

10^{14} eV can be tremendous and can explain the fine structure of the energy spectrum, which may reflect acceleration of the specific groups of nuclei.

To investigate various scenarios of particle acceleration in SNR, we still have to use indirect information contained in Cosmic Ray spectra in the vicinity of Earth.

As Galaxy magnetic fields can't confine particles with such energies, the extragalactic origin of highest energies is widely accepted. MAKET ANI installation due to its modest size is effectively collected cores of EAS initiated by primaries with energies up to 10^{17} eV, therefore, we'll constraint our analysis by energy range $5 \cdot 10^{14}$ - 10^{17} eV - so called "knee" region.

Energy spectra of primary ions from Z=1 to Z=26 will provide valuable information on validity of standard model. Information from the EAS experiments didn't provide enough clues for such "spectroscopy" of the "knee region". Nevertheless, precise measurements of electron and muon content, and implementation by the KASCADE experiment of the CORCIKA simulation code (Heck et al., 1998) as we demonstrate in numerous papers (see for example (Chilingaryan et al., 1999E), (Antoni et al., 2003)), allows classifying primaries according to 3 classes: "light", "intermediate" and "heavy". Using nonparametric multivariate methodology of data analysis ((Chilingarian, 1989), references on development and application of methods contained in (Chilingarian et al., 2003B)) we solve the problem of event-by-event-analysis of EAS data (Chilingarian et al., 1991), using Bayesian and Neural Network information technologies (Chilingarian, 1994), (Bishop, 1995).

On each stage of analysis we estimate the value of information content of variables used for EAS classification and energy estimation and restrict complication of the physical inference according to this value. MAKET-ANI experiment is located at 3200 m. above sea level; the quality of reconstruction of EAS size and shape are good enough and we can use

for EAS classification shower size and shape parameters (so called shower age). The distinctive information contained in distributions of these 2 parameters allows us to classify EAS with high level of accuracy into 2 distinct groups: initiated by “light” or “”heavy” nucleolus. In KASCDE experiment (Antoni et al., 2003) where muon content of EAS is measured additionally, and we can classify showers into 3 categories adding also “intermediate” class. Enlarging number of nuclear classes for more detailed classification can decrease quality of physical inference, see for example recent discussion of KASCADE collaboration results in Astroparticle Physics (Schatz, 2002).

Proceeding from the MAKET-ANI data and including in the analysis also data from KASCADE experiment obtained with same data analysis methodology (Heck et al., 1998), (Chilingarian, 1998)), we’ll try proving or disapproving the standard model by our experimental evidence.

Another intriguing problem of CR origin is connected with way how the “ensemble” of SN maintains the CR flux in vicinity of Earth. The fine structure of spectra at knee, suggested hypothesis that one or several recent nearby SN are responsible for spectra structures in knee region (Erlykin et al., 1997), (Erlykin et al., 1998). Therefore, identification of such SN and measuring additional flux of particles from its direction will be another proves of the standard model.

The MAKET-ANI detector reveals significant excess of extensive air showers with arrival directions pointed to the Monogem ring, a supernova remnant located at a distance of ≈300 pc from the Sun with ≈100 kyr old radio pulsar PSR B0656_14 near the center (Thorsett et al., 2003).

8.2 COSMIC RAY ACCELERATION IN SUPERNOVAE EXPLOSIONS AND PROPAGATION IN THE INTERSTELLAR MEDIUM

The power of cosmic ray sources should be more than PCR ~ 10^{41} erg/sec to maintain estimated cosmic ray energy density. This number was obtained by multiplying the CR energy density in Galaxy ρ_{CR} ~ 10-12 erg/cm^3 by Galaxy volume VG ~ 10^{67} cm^3 and dividing by particle mean escape time from Galaxy τesc ~ 10^{14} sec.

Kinetic energy of supernova ejecta: WSNR ~ 10^{52} erg and frequency of SN explosions in Galaxy: freq ~ 20^{-1} Ear, lead to CR luminosity of same order of magnitude if we assume that a few tens of percent of ejecta kinetic energy is transformed to CR energy.

In addition detected non-thermal radio emission from SNR, leading to natural assumption of accelerated electrons, along with Fermi’s concept of particle magnetic interaction with macroscopic gas clouds, made SNR main candidate engine for particle acceleration.

Supernova explosions and expanded shocks forming supernovae remnants were and are the favorite candidates for the source of cosmic rays in energy range from 10^{10} to 10^{17} eV. Power law is rather satisfactory describing spectra from 10^{12} eV (far above solar modulation effects), till several units of 10^{18} eV, where Galaxy magnetic field of 3 μG can’t confine particles anymore. At low energies till 10^{14} eV, where direct measurements are available, spectra indices of protons, carbon, oxygen and iron are very close to each other and equal to ~-2.7. Same index is describing spectra of all particles from 5•10^{14} till 3-4•10^{15} eV (knee region), where power index changed to ~ -3, to again return to value -2.7 at 10^{17} eV (ankle).

The "classical explanation" of the changing behavior of spectra consist in existence of 3 distinct acceleration mechanisms: first, usually connected with SNR shock acceleration, fading in knee region, second, unknown, responsible for energies from knee to ankle, and extragalactic, after ankle.

Numerous papers are devoted to SNR-based acceleration. Obtained values of spectra at source obey power law with index $\gamma s \sim (-2.0 - -2.1)$. Models of particle acceleration in SNR can be compared with observations only if we accounts for the diffusion and escape of CR in Galaxy. Usually energy dependence of the escape time is taken also in form of power law – $\tau esc \sim E^{-\chi}$, and relation of spectra in source and detected spectra takes form: $E^{-\gamma} \sim E^{-(\gamma s+\chi)}$. Theoretical calculations of the diffusion coefficient are based on the assumption on the distribution of magnetic inhomogeneities in Galaxy. There are 2 main distributions: "Kolmogorov Spectrum", giving $\chi=0.33$ and "Kraichnan spectrum", giving $\chi=0.5$.

Measurements of the spectra of low energy isotopes ("radioactive clocks") gives another value of $\chi=0.6$. This value, seem to be in perfect accordance with observed spectra of $\sim E^{-2.7}$, but it refers only low energies available from satellite and balloon isotope spectrometers, additional measurements of isotope spectra at higher energies are needed.

Another bulk of evidence is concern the possibility of estimation of the energy dependence of so called "mean" mass. As we can't resolve Et the "all particle"spectrum, the attempts are made at least estimate the trend of changing "mean mass". The calculations of the average depth of shower maximum, made by the fluorescence and Cherenkov detectors signal on "lightening" of mean mass just before the knee, and transition to heavies above the knee.

This behavior could be explained by influence of one or several nearest SNR, giving additional surplus flux added to the smeared superposition of thousands Galaxy SNR.

If knee feature is due only to numerous distant sources the steepening of the spectra should be much smoother than detected. Attempts to find time-temporal coordinates of SNR explaining observed fine structure of spectra heavily depend on the adopted energy dependence of the diffusion coefficient. Authors of recent estimates of the possible location of the Single Supernovae (SS), (Erlykin et al., 2003) proceeding from the "anomalous" diffusion introduced in (Lagutin, 2001) derives to following boundaries: 300 – 350pc from the Sun and 90 – 100 kEar old. They also adopted energy dependence of diffusion coefficient with $\chi=0.5$. Very Long Baselene interferometric measurements of the 100 KEar old pulsar PSR656 + 14 (Brisken et al., 2003) locates pulsar in the center of the SNR called Monogem Ring at 300 pc distance from the Sun. And it was logical to assume that the Monogem Ring, the shell of debris from a supernova explosion, was the remnant of the blast that created the pulsar (Thorsett et al., 2003). The Monogem Ring was investigated as possible source of the high energy cosmic rays by MAKET-ANI data.

8.3 THE MAKET-ANI EXPERIMENT

The MAKET-ANI surface array (Chilingarian et al., 1999) consists of 92 particle density detectors formed from plastic scintillators with thickness of 5 cm. Twenty four of them have area 0.09 m^2 and 68 area $1m^2$. The central part consists of 73 scintillation detectors and is arranged in a rectangle of 85 x $65m^2$. Two peripheral points of a distance of 95m and 65 m. from the center of the installation consists of 15 and 4 scintillators respectively.

In order to estimate the zenith and azimuthally angles 19 detectors from 92 (with area $1m^2$)

are equipped with timing readout measuring the EAS front appearance with an accuracy of ~ 5 ns.

The photomultipliers (PM-49) are placed in light-tight iron boxes. Logarithmic amplitude- digital converters (ADC) and constant fraction discriminators (CFD) are assembled just above PM. The dynamic range of the registered particle number is ~ 5 x 103.

Two types of triggers are used:

1. The hardware trigger: at least 7 of 11 central density detectors are hit with more than 3 particles.
2. The timing trigger: at least 4 from 9 symmetrically arranged relatively to the center timing detectors are hit.

If the first two conditions are fulfilled in time window of 20 μsec the event is stored. The trigger and data readout systems are done in the CAMAC standard.

Monte-Carlo calculations shown that used trigger selects EAS with sizes Ne > 5•104 and cores located within the rectangle of 40 x 12m^2 around the geometrical center of the installation.

The uncertainties of the reconstruction of EAS parameters are as following: shower size ΔNe~ 10%, the shower shape (age) parameter - Δs ~ 0.06. The accuracies of EAS angles determination are: $\Delta\theta \sim 1.5°$ and $\Delta\varphi < 5°$.

In the period of 1998 - 2002 approximately 7,788,000 EASs were registered with effective registration time of about 24,000 hours. From these showers only ~ 963,000 events were selected for the spectra calculation. The selection criteria was to have more than 95% efficiency of registration, so we selected EAS core from more compact area around the geometrical center of MAKET detector, ensuring high efficiency of EAS registration.

The following cuts were applied for the events selection: Ne > 10^5, 0.3 < s < 1.7, -24m < X0 < 24m, -12m < Y0 < 12m , $\theta<45°$. The details on MAKET-ANI operation and investigations of the uniformity and accuracy of parameter reconstruction are summarized in (Hovsepyan et al., 2003).

During multiyear measurements, the detecting channels were continuously monitored. Data on background cosmic ray spectra was collected for each detector. The slope of the spectra was used for detector calibration. Slope of background spectra is very stable parameter didn't changing ever during very severe Forbush decreases, when the mean count rates can decrease up to 30% (Chilingarian, 2003). The detailed information of the MAKET-ANI detector operation in 1997-2003, various comparisons and uniformity checks are summarized in (Vardanyan et al., 2003).

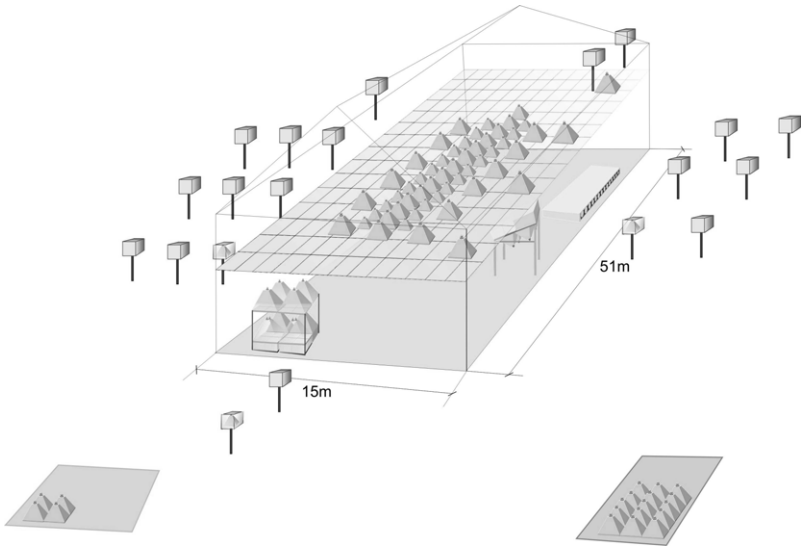


Figure 62. MAKET-ANI detector setup

8.4 SELECTION OF EAS PARAMETERS FOR CLASSIFICATION AND ESTIMATION

We are interesting in choosing combination of the EAS measured characteristics significantly differing for light and heavy initiated showers. The discriminative power of EAS characteristics were investigated using CORSIKA. (Heck et. al., 1998) and MAKET-ANI response simulation codes (Hovsepyan, 2002).

For comparison of EAS initiated by different primary ions a number of statistical methods were used, including one-dimensional statistical tests, correlation analysis and misclassification rates estimation by neural and Bayesian classificators.

Input parameters of simulation program included particle type, energy, angles of incidence, as well as geographical coordinates and altitude of MAKET-ANI detector. Energy and angular distributions were taken in a way allowing forming samples according to modern theoretical expectations. Due to stochastic nature of particle propagation through the atmosphere the output parameters of simulation programs are random variables. By the numerous simulations of cascades we obtain statistical characteristics of these random variables, posted in the Table 34. We also put in table the variable we can't directly measure – the primary energy E_0, for investigation its correlations with measured EAS parameters. The detection techniques superimpose additional random "noise" to measured entities, therefore it is extremely difficult to isolate "pure" states, i.e., showers generating by the particular primary ions. Therefore we'll precede from arbitrary assumption of 2-way division of all primary nucleolus, so called, light and "heavy" mass groups. As representatives of light group we'll take proton and He nucleus, for heavy group Si and Fe nucleus we'll be representatives. As we will see, the intrinsic differences of the light and heavy ion cascades in the atmosphere make distributions of EAS parameters different. We'll investigate is this difference enough for reliable 2-way classification and how the detector response smeared it. Due to isotropy of the primary flux, we can't expect any differences between angular parameters and coordinates of shower axes (X_0, Y_0). Integrated over whole energy range shower sizes of heavy and light nuclei initiated EAS also are very similar. Only parameter, showing significant difference is shower shape – age (s) parameter. Although the detector smear this difference (see Figure 63), difference remain significant, and as we'll see further the different correlations of this feature with shower size made pair (N_e , s) effective

both for classification and energy estimation.

In Table 35 we enumerate differences between alternative one-dimensional distributions by the so-called P-values of statistical tests (Aharonian et al. 1990). We use parametric Student t statistics, D – Kolmogorov-Smirnov tests statistics and Mahn-Whitney transmutation statistics U. Statistical tests are checking so called H_0 hypothesis consisted in the statement that 2 distinct examined samples (i.e. EAS generated by proton and iron nuclei) are coming from one and the same distribution. If P-value of test is big enough it is highly improbable that H_0 is true. And higher is P-value, with higher statistical significance we can reject H_0 and, therefore, we can use P-values for estimating "overlapping" of distributions – the greater value of test, smaller is overlapping. Overlapping is directly proportional to expected misclassification rate, and therefore we can measure "the discriminative power" of variables by the P-values. Greater P-value – greater discriminative power. Of course, we can make only relative estimates comparing different variables, or same variable before and after detector smearing, no precise estimates of misclassification rates could be made. For it we have well establish procedures of Bayesian risk estimation, but nevertheless P-values provide fast and statistically sound possibility of comparing EAS characteristics and influence of detector response function. From Figure 63 it is apparent that detector smearing significantly enlarges overlapping of distributions of age parameter, related to p and iron initiated showers. Corresponding P-values, as we can see from Table 35, decrease from 159 to 125. Also we can deduce from Table 35 that shower age parameter s has best discriminative power, shower size N_e , because of integration over broad energy range, is obviously less effective for distinguishing p and Fe-initiated showers. Other listed in table variables, obviously, has not discriminative power at all, because angular distributions and core co-ordinates should be highly independent from primary particle type. All three statistical tests give very coherent results for the EAS parameters when comparing proton and iron samples.

From Table 35 and Figure 63 we can make conclusion that it is extremely important to take into account detector response, visa verse the physical conclusions will be optimistically biased. The shape of the shower due to not perfect sampling of cascade electrons (MAKET-ANI detector coverage is less than 1%) is distorted during detection and reconstruction procedures, therefore this parameter loose portion of its "discriminative power".

Table 34. EAS parameters for p and Fe primaries

	Without detector response				*Detector response*			
	Mean		*MSD*		*Mean*		*MSD*	
	H	*Fe*	*H*	*Fe*	*H*	*Fe*	*H*	*Fe*
lnE_0	*13.35*	*14.35*	*1.21*	*1.15*	*13.34*	*14.35*	*1.19*	*1.15*
lnN_e	*12.08*	*12.36*	*1.26*	*1.32*	*12.15*	*12.41*	*1.20*	*1.29*
S	*0.97*	*1.15*	*0.10*	*0.10*	*0.94*	*1.10*	*0.15*	*0.14*
X_0	*-0.52*	*-0.61*	*18.70*	*19.2*	*-0.68*	*-0.69*	*18.64*	*19.24*
Y_0	*0.13*	*0.18*	*10.0*	*10.0*	*0.21*	*0.07*	*10.06*	*10.27*
$cos\theta$	*0.87*	*0.87*	*0.09*	*0.09*	*0.87*	*0.87*	*0.09*	*0.09*
φ	*3.13*	*3.14*	*1.82*	*1.82*	*3.14*	*3.12*	*1.82*	*1.82*

Table 35. One-dimensional tests of p-Fe samples (t- Student, D- Kolmogorov-Smirnov, U- Mann Whitney)

	without detector response		*with detector response*	

	t	D	U	t	D	U
lnN_e	20.87	9.88	22.97	20.49	8.97	21.41
S	158.94	56.69	125.03	104.63	41.31	95.37
X_0	0.43	1.35	0.29 0.44	0.06	1.38	0.31 1.27
Y_0	0.89	0.42 5.32	2.55	1.43	1.00 5.97	2.85
cos θ	5.36 0.29	0.89	0.29	6.06 1.16	1.11	1.17

Table 36. One-dimensional tests of p-He samples (t- Student, D- Kolmogorov-Smirnov, U- Mann-Whitney)

	without detector response			with detector		
	t	D	U	t	D	U
lnN_e	2.58	1.48	1.72	2.18	1.77	2.55
S	41.76	15.39	38.61	23.92	9.24	22.64
X_0	0.70	0.86	0.73	1.13	1.23	1.15
Y_0	0.32	0.52	0.31	0.33	1.06	0.38
cos θ	1.41	1.04	1.32	1.36	0.95	1.42

Table 37. One-dimensional tests of Si-Fe samples (t- Student, D- Kolmogorov-Smirnov, U- Mann-Whitney)

	without detector response			with detector		
	t	D	U	t	D	U
lnN_e	1.68	0.89	1.32	1.71	0.96	1.77
S	33.91	15.09	33.67	21.91	9.93	22.74
X_0	0.13	0.55	0.10	1.06	1.20	1.07
Y_0	0.01	0.43	0.01	0.01	0.99	0.19
cos θ	0.49	0.91	0.79	0.46	0.75	0.49

For the comparison we compare “similar” nuclei pairs: p and He as well as Si and Fe. Table 36and Table 37 demonstrate that P-values of statistical tests became very small, proving that distributions of the EAS parameters for corresponding nucleolus practically coincide.

Although ingle EAS characteristics of light and heavy nuclei also significantly overlap as we can see from Figure 63, we can use more than one parameter, utilizing difference of the correlations between parameters. The generalized distance between multidimensional distribution and experimental event (so called Mahalonobis distance) is dependent on the difference in correlations in alternative samples. In the Table 38, Table 39, Table 40 and Table 41Table 40 we post correlation matrices for proton and iron induced samples. Iron nuclei interacting high in atmosphere generate electron cascade significantly attenuating when reaching 3200 m. We enumerate shape of such “old” cascades by large values of age parameter. Therefore, small shower sizes (small N_e), is corresponding to large values of s parameter. This anticorrelation is much more pronounced for iron nuclei comparing with proton, due to higher probability of interaction of iron nuclei high in the atmosphere and faster attenuation of iron-induced cascades. Again the detector “smeared” this nuclear signature and realistic estimate of correlation difference in proton and iron samples, as we can see from Table 38 and Table 39 and Figure 64 becomes significantly smaller when the detector response function is taken into account. As we use both parameters and pair-wise correlations for multivariate classification, expected misclassification rates will be larger

comparing with "pure" cascade simulations. But, nevertheless, the difference in p and Fe classes remain significant and that gives hope that these 2 parameters will allow us to distinguish at least light and heavy primaries. Very large correlations of primary energy E_0 with shower size N_e didn't distorted by detector response and we'll use shower size and age for energy estimation.

Table 38. Correlation matrix for proton sample without detector response

	lnE_0	lnN_e	S	X_0	Y_0	$cos\theta$	φ
lnE_0	1.00	0.91	-0.16	-0.00	-0.01	-0.23	0.00
lnN_e	0.91	1.00	-0.47	-0.00	-0.01	0.04	0.00
S	-0.16	-0.47	1.00	-0.01	0.01	-0.40	0.02
X_0	-0.00	-0.00	-0.01	1.00	0.02	0.00	-0.01
Y_0	-0.01	-0.01	0.01	0.02	1.00	0.01	0.01
$cos\theta$	-0.23	0.04	-0.40	0.00	0.01	1.00	-0.01
φ	0.00	0.00	0.02	0.01	0.01	0.00	1.00

Table 39. Correlation matrix for iron sample without detector response

	lnE_0	lnN_e	S	X_0	Y_0	$COS\theta$	φ
lnE_0	1.00	0.93	-0.70	-0.01	-0.02	-0.32	0.00
lnN_e	0.93	1.00	-0.83	-0.01	-0.01	0.03	0.00
S	-0.70	-0.83	1.00	0.02	0.01	-0.19	0.03
X_0	-0.01	-0.01	0.02	1.00	0.00	0.00	0.01
Y_0	-0.01	-0.01	0.01	0.00	1.00	0.01	0.00
$Cos\theta$	-0.32	0.03	-0.19	0.00	0.01	1.00	-0.02
φ	-0.00	0.00	0.03	0.01	0.00	-0.02	1.00

Table 40. Correlation matrix for proton sample with detector response

	lnE	lnN	S	X_0	Y_0	$cos\theta$	φ
lnE_0	0	e	-0.06	-0.01	-0.01	-0.24	0.00
lnN_e	0.90	1.00	-0.28	-0.02	-0.01	0.05	0.01
S	-0.06	-0.28	1.00	-0.04	-0.04	-0.29	0.02
X_0	-0.01	-0.02	-0.04	1.00	0.01	0.01	0.00
Y_0	-0.01	-0.01	-0.04	0.01	1.00	0.00	0.00
$cos\theta$	-0.24	0.04	-0.29	0.01	0.00	1.00	-0.01
φ	0.00	0.00	0.02	0.00	0.00	-0.01	1.00

Table 41. Correlation matrix for iron sample with detector response

	lnE_0	lnN	S	X_0	Y_0	$cos\theta$	Φ
lnE_0	1.00	e	-0.38	0.0	-0.02	-0.32	-0.00
lnN_e	0.92	1.00	-0.45	-0.01	-0.03	0.03	0.00
S	-0.38	-0.45	1.00	0.05	0.02	-0.17	0.01
X_0	0.01	-0.01	0.05	1.00	-0.01	0.01	0.00

Y_0	-0.02	-0.02	0.01	-0.01	1.00	0.00	0.01
$cos\theta$	-0.32	0.03	-0.17	0.01	0.00	1.00	-0.01
φ	-0.00	0.00	0.01	0.00	0.01	-0.01	1.00

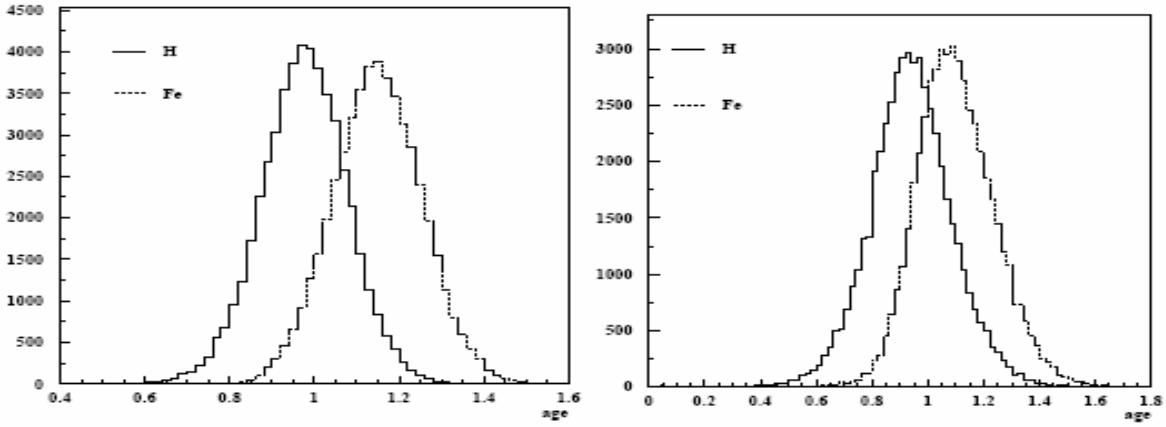


Figure 63. Shower Age distribution for simulated light and heavy primaries with (right) and without (left) incorporating of detector response.

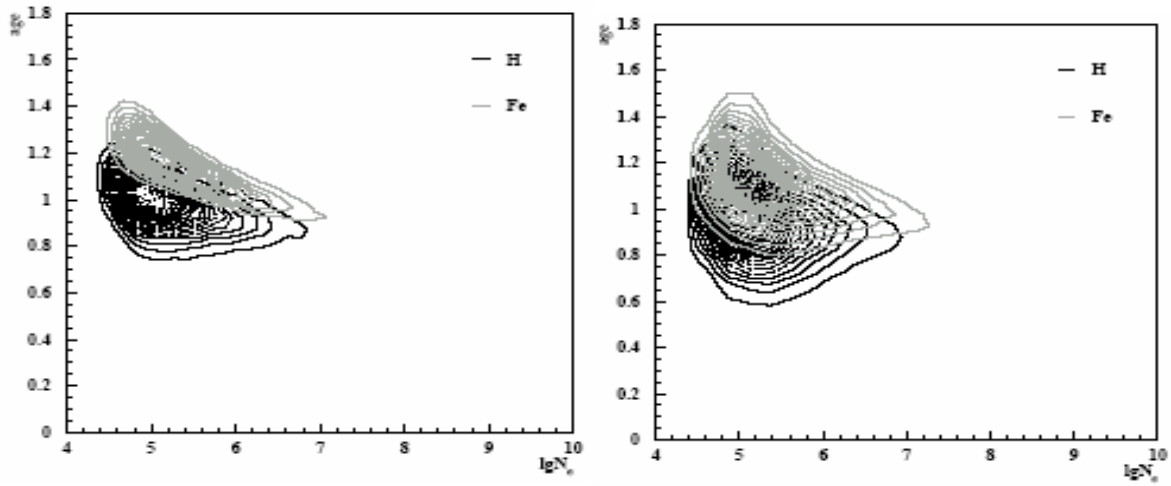


Figure 64. Scatter plot Shower Age versus Shower Size for simulated light and heavy primaries with (right) and without (left) incorporating of detector response.

The most direct estimates of the "discriminative power" of EAS characteristics are obtained by the classification of the samples obtained from EAS simulations. Overlapping of the 2- dimensional distributions apparent from Figure 64 is enumerated by the misclassification rates via Bayesian or neural network classification of EAS initiated from the alternative groups of nuclei. Using only EAS electron characteristics, we can't resolve nucleus with similar masses, as p and He, Fe and Si, therefore we joint these nucleus in groups naming them "light" and "heavy", those restricted ourselves to 2-way classification of the experimental data. Results on expected classification results posted in the Table 40 and Table 41 demonstrate, that although detector smearing significantly enlarges misclassification rates, Nevertheless >70% of correct classifications is very encouraging and N_e – s pair as measured by MAKET ANI detector provide enough information for the 2-way classification.

We also want to point on good agreement between results obtained by using 2 completely different methods of classification: Bayesian classification with nonparametric estimation of multivariate probability density function and Neural Network classification

using stochastic net training methodologies.

Table 42. Neural classification into two classes using H+He and Si+Fe events without and with detector response

	Without detector response		*With detector response Light Heavy*	
	Light	*Heavy*	*Light*	*Heavy*
Light	*0.925*	*0.075 0.045*	*0.720*	*0.280*
Heavy	*0.955*		*0.240*	*0.760*

Table 43. Bayesian classification into two classes using H+He and Si+Fe events without and with detector response

	Without detector response		*With detector response Light Heavy*	
	Light	*Heavy*	*Light*	*Heavy*
Light	*0.938*	*0.062*	*0.712*	*0.288*
Heavy	*0.043*	*0.957*	*0.237*	*0.763*

8.5 DATA CLASSIFICATION INTO LIGHT AND HEAVY GROUPS OF NUCLEI, PURIFICATION OF SELECTED GROUPS OF NUCLEI

According to results of the previous section we use 2 "training samples" of "light" and "heavy" nucleus initiated N_e-s pairs, generated by the CORSIKA code including MAKET-ANI response function. Before Neural classification of the MAKET-ANI data we investigate expected purity[1] and efficiency[2] of our data analysis procedures.

From Table 44 we can see that efficiency of classification, i.e. correct identification of nuclei from light and heavy groups is above 70%, the "intermediate" Oxygen nuclei are distributed approximately equally among 2 groups. To obtain purity estimates we assume, so called "normal" primary composition: 30% H, 24% He, 17% O, 17.5% Si and 11.5% Fe. Table 43 demonstrates that purity of light group is above 70%, purity of heavy group is below 50% with large contamination of the Oxygen and light nucleus.

Table 44. Efficiency of the neural classification of EAS initiated by different primaries into two mass groups

	Light	*Heavy*
H	*0.720*	*0.280*
He	*0.691*	*0.309*
O	*0.453*	*0.547*
Si	*0.352*	*0.648*
Fe	*0.240*	*0.760*

Table 45. Purity of the classification of different nuclei in light and heavy groups

	H	*He*	*O*	*Si*	*Fe*
Light	*0.407*	*0.298*	*0.137*	*0.111*	*0.047*

To enlarge purity of the heavy nuclei group we introduce purification procedure (first introduced in ISTC A116 project final report, 1999, see also, (Antoni et al., 2003)), enlarging purity of each nuclear group on the cost of decreasing the efficiency.

[1] purity: fraction of true classified events in an actual number of events assigned to a given class

[2] efficiency: fraction of true classified events in total number of events of a given class

The purification of the selected "light" and "heavy" groups was done by selecting the appropriate domain in the entire range of the network output. The feed-forward Neural Network (NN) performs a nonlinear mapping of the multidimensional characteristics of the EAS to the real number interval [0,1], called the output of the NN. Figure 1 shows the network output histogram. The network was trained to shift the "heavy" group to the right and the "light" group to the left of the histogram. The 0.5 point of the NN output is the so called decision point. The particular class assignments for the two-way classification are the subintervals [(0.0,0.5) and (0.5,1.0)] for the "light" and "heavy" class respectively. If the neural network is satisfactorily trained to have generalization capabilities, the output distributions for the different classes will overlap at the subinterval boundaries. Therefore, by shrinking the subintervals, i.e. moving the interval boundary to the left and right of the decision point 0.5, it is possible to remove a large proportion of the misclassified events. Of course, simultaneously we loose parts of the true-classified events, i.e., decrease the efficiency.

Thus, instead of one decision point in the middle of the NN output interval, we will have two "decision intervals" for accepting "light" and "heavy" nuclei, and a third interval in between where we reject the classification. Figure 1 demonstrates this "purification" procedure.

Figure 2 shows the results of the purification. The values next to the symbols indicate the selected decision interval used for obtaining particular purity-efficiency relation. For example, if we select [(0.0,0.3) and (0.7,1.0)] intervals for classification of the "light" and "heavy" nuclei, we obtain 96% purity and 56% efficiency for the "light" class; 78% purity and 55% efficiency for the "heavy" class. Therefore, we can enhance the purity of the light nuclei up to 95% and the purity of the heavy nuclei up to 80%, while still holding the efficiency above 50%. The purity and the efficiencies are obtained by classifying 35000 light (H,He) and 17000 heavy (Si,Fe) control events, which are not used for the training of the neural network.

The high purity for both classes is achieved, since the intermediate nuclei (simulated oxygen initiated EAS) were not demonstrated on the plot. More realistic purity and efficiency estimates are apparent from the Table 46 and Table 47 (we include also Oxygen nuclei).

As we can see from Table 47 the purity of the light group increases from 70 to 77% and for the heavy ones from 46 to 55%, we'll keep in mind that approximately 20% of heavy group are due showers initiated by O nuclei. The purification allows us significantly increase the purity of 2 alternative samples and we can, therefore, estimate energy spectra of light and heavy groups. Of course, first we describe the energy estimation procedures used.

Table 46. Efficiency of the neural classification of EAS initiated by different primaries into two mass groups (purification intervals [0.,0.3) and (0.7,1.]).

	Light	*Heavy*
H̆e	*0.475*	*0.135*
O	*0.252*	*0.303*
Si	*0.176*	*0.393*
Fe	*0.099*	*0.561*

Table 47. Purity of the classification of different nuclei in light and heavy mass groups (purification [0.,0.3) and (0.7,1.]).

	H	*He*	*O*	*Si*	*Fe*
Light	*0.459*	*0.310*	*0.115*	*0.084*	*0.032*
Heavy	*0.115*	*0.131*	*0.207*	*0.278*	*0.268*

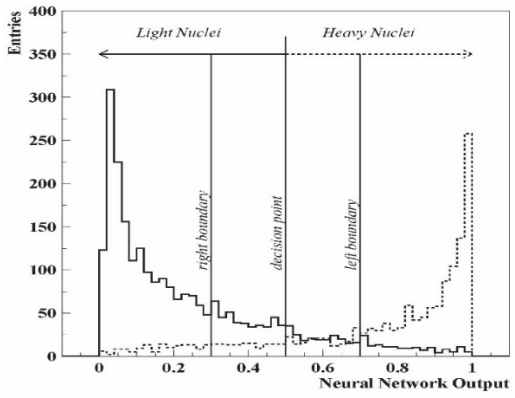


Figure 65. Output of the Neural Network (NN) trained to distinguish light and heavy nuclei.

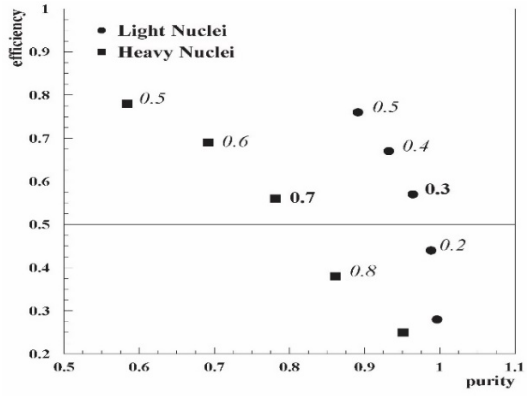


Figure 66. Purity – Efficiency plots obtained by shifting the NN decision boundaries

8.6 ESTIMATION OF THE PRIMARY ENERGY OF DIFFERENT GROUPS OF NUCLEI

Primary energy of each shower was obtained by neural network estimators separately for the light and heavy nuclei induced events, exploiting very large correlation of shower size N_e with primary energy and different correlations between primary energy and shower shape in light and heavy nuclei groups (see Table 36, Table 37, Table 38, Table 39). In Figure 67 relative errors of energy estimation for 10 energy intervals are posted. The bias of the energy estimation, displayed does not exceed the 5% for the light group (left) in the whole energy range except the lowest energies. For the heavy group of nuclei the estimation bias in the energy range of 10^{15} — 10^{16} eV is not larger than 5%, nevertheless, one can observe some overestimation for low and high energy regions. The energy resolution for heavy group of nuclei is significantly better (MSD ~ 20%) as compared to the light group of nuclei (MSD ~ 30%) due to the smaller fluctuations of heavy initiated EAS size and shape. Also, accuracy of the energy estimation is enhancing with enlarging primary energy.

By 2 horizontal lines around the 0-line the 5% error corridor is outlined. Error bars are correspondent to Mean Square Deviation (MSD).

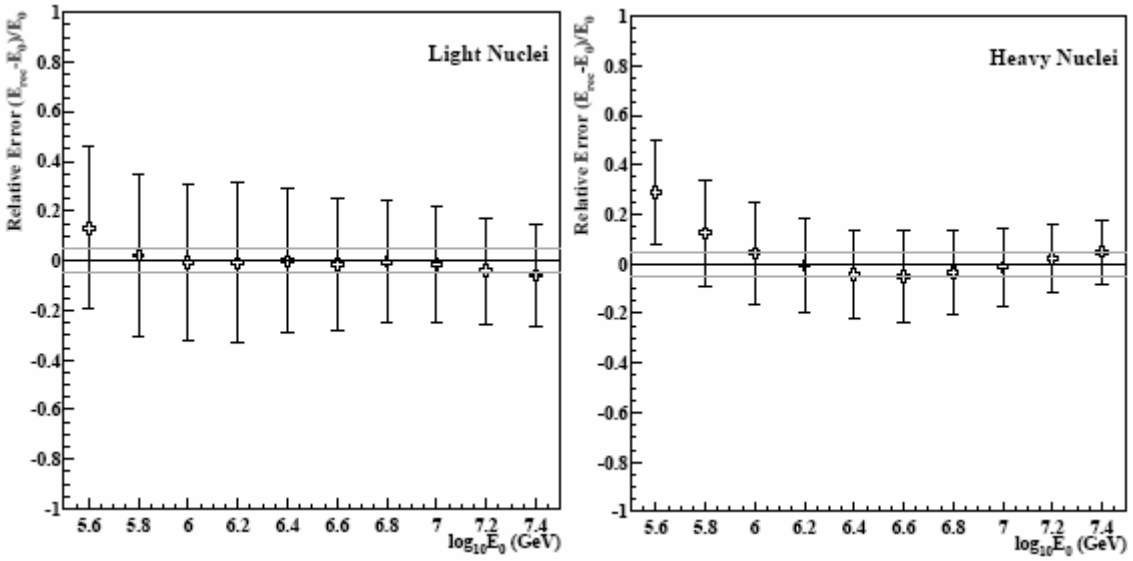


Figure 67. The relative errors of energy estimates fro 10 energy intervals of light and heavy groups.

8.7 ENERGY SPECTRA

Figure 68 adopted from (Haungs et al., 2003) presents energy spectrum measured by different detectors exploiting various experimental techniques and energy reconstruction methods.

Energy estimation for all experiments was done using Monte Carlo simulations with different numerical algorithms. Despite considerable differences in experimental techniques and different EAS components used for the energy estimation (showers shape and electron size parameters, muons, Cherenkov light) and differences in systematic errors (usually not reported in publications) almost all spectra are in rather good agreement if we assume energy estimation accuracy about 20%. Only at energies after the knee feature the spectra disagree, probably because of the saturation effects in the scintillators in some experiments.

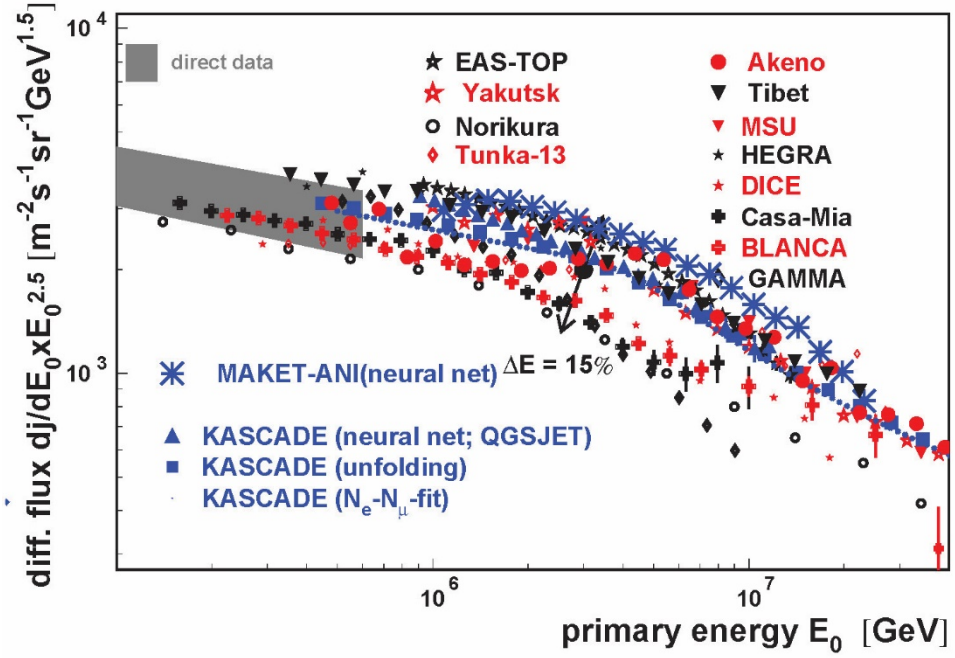


Figure 68. Summary of the all particle spectra from 18 experiments

All particle spectra and mean logarithmic mass in many cases presented as an outcome of

the EAS experiment aren't too informative. We never know which combination of primaries constitutes the mean and which groups of primaries are responsible for the knee. The best solution will be to separate different groups of nuclei and reconstruct energy spectra to detect spectral knees of different nuclei at different positions. This program was partly fulfilled with data from the MAKET-ANI experiment.

After checking for the purity and the efficiency each of the near 1 million showers registered by the MAKET-ANI installation in 1999-2002, with shower size greater than 10^5 , was classified according to the techniques described in (Chilingarian et al., 1991), (Antoni et al., 2003). The energy of the classified into 2 distinct classes showers was estimated for each group separately using again the CORSIKA simulations and neural estimation techniques. In Figures 3 we present the obtained energy spectra of the "light" and "heavy" mass groups. The spectrum of the "light" group shows a "knee" in the region of $3\text{-}4 \cdot 10^{15}$ eV. The "knee" feature is not observed for the spectrum of the "heavy" component at least till energies of 10^{16} eV. The number of «light» and «heavy» nuclei at $\sim 10^{15}$ eV is approximately equal and the number of "heavy" nuclei gets larger at energies greater than the "knee" energy.

The "purified" spectra shown in Figure 4, demonstrate the lower flux intensities for both classes of particles due to the lower efficiency. The "knee" position shifts to lower energies as we expect that after purification the proportion of protons is enlarged. In addition, the slope of the spectrum (spectral index) of the "purified" light component becomes steeper: $= -2.63$, compared to $= -2.54$ before purification. Both results are consistent with the rigidity dependent acceleration and, consequent fading of the proton flux at high energies. Another important feature of the obtained spectra is the very large difference between spectral indices before and after the "knee": $\Delta\gamma(light) = \gamma_2 - \gamma_1 \sim 0.9$. It is well known that the same parameter for the all-particle spectra is $\Delta\gamma\ (all - particle) \sim 0.3$, (Haungs et al., 2003).

Erlykin and Wolfendale, in their simulations, failed to reproduce the actual shape of the all-particle spectrum by averaging the proton and nuclei fluxes produced by nearly 50,000 distant supernovae in our Galaxy (Erlykin et al., 2001). Therefore, they propose that the nearby young supernova (< 500 pc and < 110*kyr*), is responsible for the approximately 60% of the detected cosmic ray flux in the vicinity of earth (Erlykin, Wolfendale, 2003). The very large difference of the spectral indices before and after the knee of the "light" component (~ 0.9) confirms Erlykin and Wolfendales proposal regarding the huge impact of the nearest supernova on the cosmic ray flux in the vicinity of earth. It suggests the necessity to make detailed calculations of the influence of the nearest supernova on the detected cosmic ray fluxes, i.e., to obtain the partial spectra of the nuclei accelerated by the single source (for candidate of such source see (Thorsett et al., 2003).

Figure 71 explicitly demonstrates the dependence of the relative abundances of purified light and heavy nuclei on primary energy. Primary composition is light dominant with a trend to change to a heavier composition at higher energies.

In Figure 72 the mean logarithmic mass versus the primary energy is plotted in comparisons with other experiments. Overall tendency of the lightening of the cosmic ray composition just before the knee, first mentioned in (Chilingarian et al., 1997), as well as "havening" after the knee is common feature of all of spectra. Although rather big discrepancy between mean values derived from different experiments is apparent, the plots of mean logarithmic masses of the KASCADE and MAKET-ANI experiments agree rather well. The same neural technique of energy estimation and mass classification was used in

both cases (Chilingarian, 1998); KASCADE results are obtained after three way classification in contrast to the two way classification performed for MAKET-ANI data.

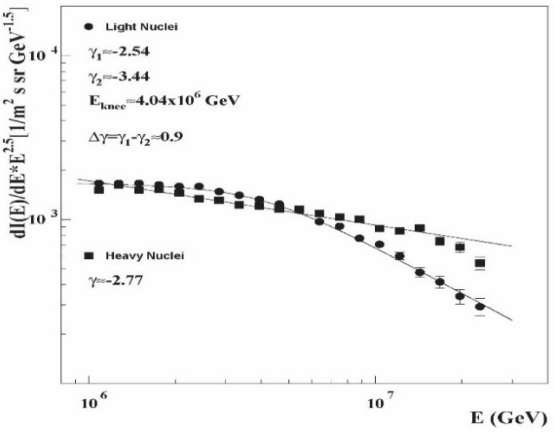


Figure 69. Energy spectra of light and heavy nuclei obtained by neural classification and energy estimation. EAS characteristics used: Shower Size and Shape (Age).

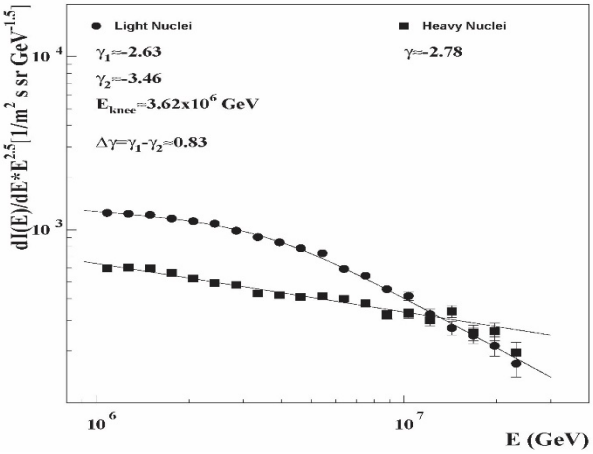


Figure 70. Energy spectra of light and heavy nuclei obtained by neural classification and energy estimation. The same as in Figure 69 but obtained with purified light and heavy data samples. Purification intervals [0.,0.3) and (0.7,1.].

In Figure 74 the energy spectra of heavy and light mass groups are posted. The coincidence of the energy spectra from the independent data obtained by different zenith angle intervals proves both good angular accuracy of MAKET-ANI detector and robustness of Neural Network classifier used. In Figure 74 we present the world data on the light mass group spectra. Again energy spectra of KASCADE **(Vardanyan et al., 1999)** and MAKET-ANI experiments using same Neural Network analysis methods are in good agreement in terms of intensities, shape of the spectra and spectral indices. HEGRA specter, obtained with completely different experimental methodic, also prove steepening of the light mass group spectra and shift of the knee position to the lower values of primary energy comparing with all-particle spectra.

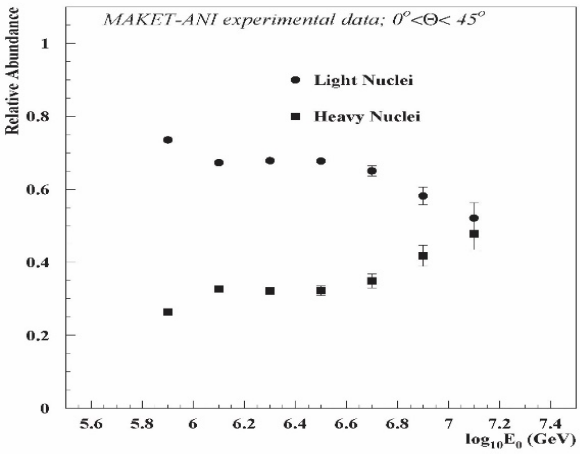


Figure 71. Relative abundances of light and heavy nuclear groups (purified EAS)

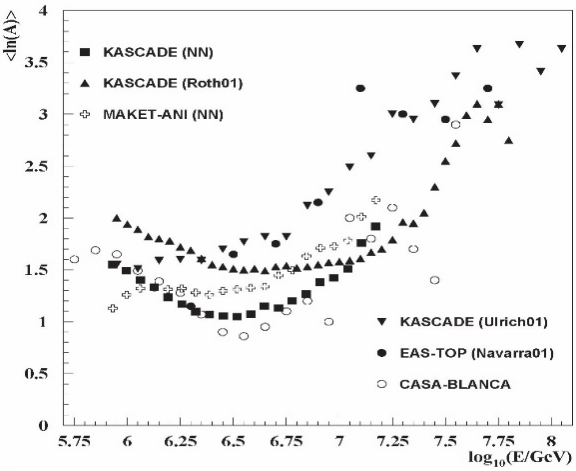


Figure 72 Mean logarithmic mass dependence on primary energy (world data survey) (Allesandro, 2001); (Fowler et al., 2001); (Roth et al., 2001), (Ulrich et al., 2001); (Vardanyan et al., 1999)

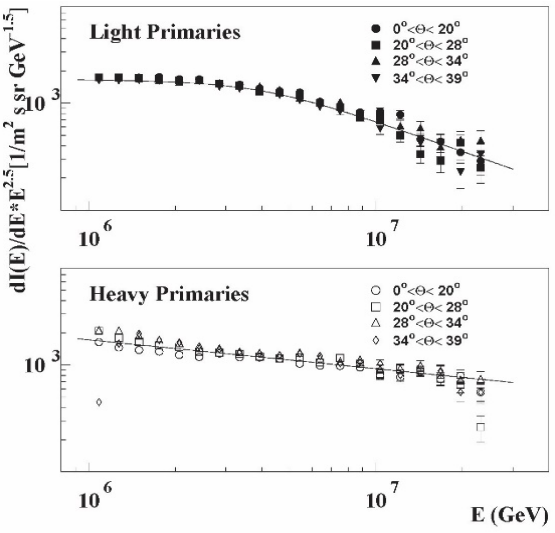


Figure 73. Reconstructed energy spectra from size spectra in different energy intervals

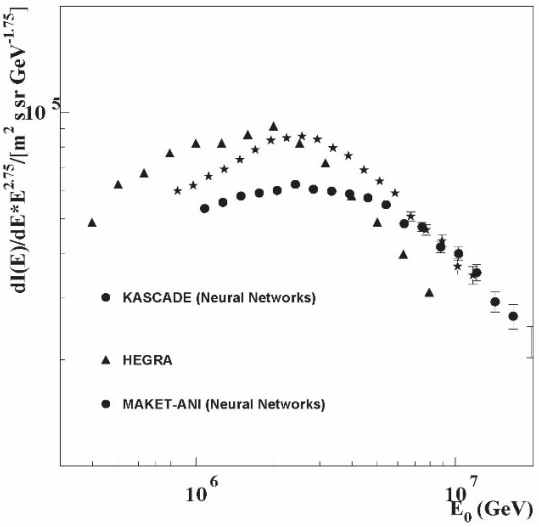


Figure 74. Light group spectra (world survey) (Arqueros, et al., 2000), (Vardanyan et al., 1999)

8.8 SEARCH OF POINT SOURCE OF COSMIC RAYS

The most exciting problem connected with cosmic rays is the exploration of a particular accelerating astrophysical source. Unfortunately, due to the bending in galactic magnetic fields, charged particles lose information about parent sites during the long travel and arrive to Earth highly isotropic. The Supernovae (SN) explosions are the most popular candidates for acceleration sites. The problem is in understanding how the Galactic "ensemble" of Supernovae maintain the CR flux in the vicinity of Earth. The fine structure of all-particle spectra at the knee suggests hypothesis that one or several recent nearby SN are responsible for the observed spectra structures (Erlykin et al., 1997), (Erlykin et al., 1998). Therefore, identification of such SN and measuring the flux of particles from its direction will be the best proof of the most popular model of hadron acceleration.

Very Long Baselene interferometric measurements of the ~100 KEar old pulsar PSR656 + 14 (Brisken et al., 2003) locates the pulsar near the center of the Supernovae Remnant (SNR) called Monogem Ring at ~300 parsec (pc) distance from the Sun. It was logical to assume that the Monogem Ring, the shell of debris from a supernova explosion, was the remnant of the blast that created the pulsar (Thorsett et al., 2003). The position and age of the SNR perfectly fits the Single Source (SS) model (Erlykin et al., 2003) and following the recommendation in (Thorsett et al., 2003) we "scanned" the Monogem Ring with high energy cosmic rays detected by the MAKET-ANI detector (Chilingarian et al., 1999A) at Mt Aragats in Armenia (N40°30′, E44°10′).

We chose high energy particles, not deflected significantly by the Galactic magnetic fields. More than 2,000,000 EAS detected by the MAKET-ANI experiment with size greater than $N_e > 10^5$ (primary energy > 3-4x10^{14} eV) were selected for the search of the CR point source.

Two-dimensional grids were generated in equatorial coordinates with the bin center tuned in the direction of the Monogem Ring center (circle of 9.2°). The best signal was obtained with bin center coordinates of (750+14) and bin size (3° x 3°). The selected

direction corresponds to the detector looking at the zenith coordinate of ~ 28°, where MAKET-ANI zenith angular accuracy is ~ 1.5°, the azimuth angle estimation accuracy is about 3°. Shower cores were collected from an area of 18 x 36 square meters around the rectangular central area of the detector. The shower age parameter was selected in the range of 0.3 – 1.7.

After analyzing more than 2,000,000 events with $N_e > 105$, we tested different locations of the source within the Monogen Ring using different cuts on the shower size. Results are summarized in Figure 75 and Figure 76 and in Table 48 and Table 49. From this analysis we determined the declination band where the candidate source is located to be (δj = 12.5° - 15.5°). In the bin distribution of Figure 75 we see a large peak corresponding to the Right Ascension (RA) bin of 7.4 – 7.6 hours.

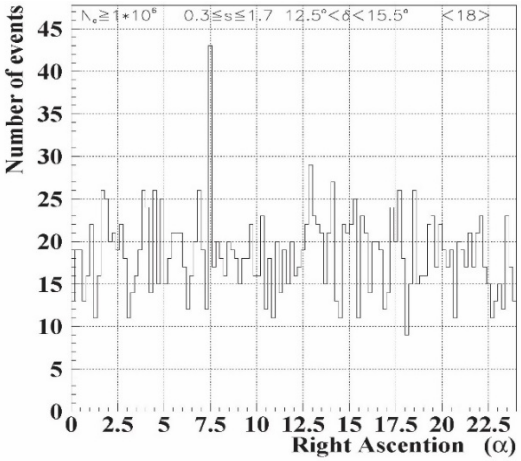


Figure 75. Distribution of number of events in each of 120 RA bins for declination band of 12.5° – 15.5°

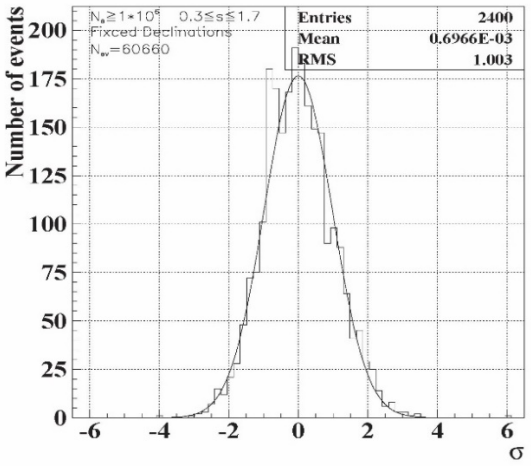


Figure 76. Signal significance test with full equatorial coverage with 2400, 3° x 3° bins; $N_e > 10^6$

For more details about signal dependence on shower size (particle energy) we calculate the number of events which fall in the "signal" bin for different N_e cut values. Table 48 demonstrates that the estimated number of signal events ($N_s \sim N_{750+14} - N_{background}$) remain approximately constant from $N_e > 5 \cdot 10^5$ up to at least $N_e = 10^6$ and fade rapidly after. The best significance value was obtained at $N_e > 10^6$.

We use modified Right Ascension Scan (RAS) method to confirm the existence of the CR point source. The background events were taken from the mean value of other RA bins in the same declination band (in our case 120 rectangular RA bins centered on 20 different declinations). The significance of the source was calculated by:

$$\sigma_{i,j} = \frac{N_{i,j} - \overline{N}_j}{\sqrt{\overline{N}}} \; i = 1, N_\alpha, 3j = N_{\delta 1}, N_{\delta 2}, 3 \qquad (8.1)$$

where $N_{i,j}$ is the number of events in the equatorial coordinates bin , N_α =360 is the range of RA; $N_{\delta 1}$ =6 is the first declination , $N_{\delta 2}$ =66 is the last declination, for a total of 20 declination "bands".

We are looking for Single Source candidates in the two-dimensional Δα x Δδ (3° x 3°)

grid, covering a 360° x 60° equatorial coordinate range with 2400 bins. More than 2 million showers detected by the MAKET-ANI detector with sizes starting from $N_e > 10^5$ were distributed among the 2400 angular bins.

We assume that for j-th declination belt number of events fallen in each of RA bins is random variable obey the Gaussian distribution with parameters $N(\overline{N}_j, \sqrt{\overline{N}_j})$: we calculated the RA bins average (over 120 bins) and used its square root as a measure of the background variance for this particular declination. To integrate information from all declination bands we perform normalization transformation, equation (8.1), and obtain joint distribution for all declination bins.

As usually in statistical hypothesis testing, the main hypothesis we want to check (named H_0) consists in the opposition to the hypotheses we are interested, i.e. we will check the hypothesis that arrival of the particles detected by the MAKET-ANI detector is isotropic ("no-signal" hypothesis) and, therefore, detected enhancement in the "signal bin" is simple random fluctuation of the isotropic distribution. As we hope to reveal signal we are interested in rejection of H_0 with the maximal possible confidence. Detecting large peak we'll estimate very low probability of H_0 being true, but, of course, it doesn't imply that the opposite hypothesis is automatically valid. As was mentioned by P.Astone and G.D'Agostini (Astone et al., 1999) behind logic of standard hypothesis testing is hidden a revised version of the classical proof by contradiction. "In standard dialectics, one assumes a hypothesis to be true, then looks for a logical consequence which is manifestly false, in order to reject the hypothesis. The 'slight' difference introduced in 'classical' statistical tests is that the false consequence is replaced by an improbable one".

If the experimental histogram will not differ significantly from test distribution we'll have no reason to reject H_0 and therefore we can't claim that our results are supporting the hypothesis that detected peak is statistical fluctuation only.

And if we'll be able reject H_0, we can accept with high level of confidence that detected enhancement is due to the additional cosmic rays from the Monogem Ring.

According to the logic described above we calculate the test statistics applying equation 1 to the experimentally detected showers and using the equatorial grid covering all directions seen by MAKET-ANI detector. As we can see from Figure 14 the shape of the cumulative distribution is very close to standard Gaussian distribution N(0,1), the χ^2 test value is ~1.5 per degree of freedom. Only one direction (pointed to the Monogem Ring) from the 2400 demonstrates significant deviation from N(0,1). Our model will consist in isotropic "background" and "signal" mixed with background in one of 2400 equatorial bins.

From the obtained value of $\sigma = 6.04$ for this particular "signal bin" we can calculate corresponding probability of obtaining this value under H_0 hypothesis to be $\sim 2400 \cdot 10^{-9} \cdot (1-10^{-9})^{2399} \sim 2 \cdot .10^{-6}$. The null hypothesis could be true only in 2 cases out of a million, therefore we have good reason to reject the null hypothesis and conclude that the MAKET-ANI detector detected high energy cosmic rays from the direction of the Monogem Ring.

We, also, investigated the high-energy events with shower sizes of $N_e > 10^5$, $5 \cdot 10^5$ $8 \cdot 10^5$,10^6, $2 \cdot 10^6$. All the distributions are very close to N(0,1); at a lower cut on the shower size ($N_e > 10^5$ corresponds to approximately $\sim 5 \cdot 10^{14}$ eV) the significance of the signal is minimal, considerably enlarging when we enlarge the shower size.

To check if our result depends on a particular grid, we randomly selected the declination

of the bin center and then calculated the events falling in the 120 RA bins in that band. We repeated this procedure 100 times. The distribution of 100 random grids did not deviate from a standard Gaussian. Therefore, we conclude that the signal significance did not depend on the particular grid used for binning of the events. Another test concerns the influence of the chosen bin size on the signal significance.

From Table 49, we can conclude that the (3° x 3°) bin size provides the best coverage of the signal domain. Enlarging the bin size leads to the reduction of the signal due to the enlarged fluctuations of "background", but number of signal events remains approximately constant. The best estimate of the number of signal events is the difference between number of events in the "signal bin" and the mean number of events in the considered declination band.

Table 48. Dependence of the "signal" on shower Size cut

Ne	*Number of events in the declination band δj = 12.5°- 15.5°*	*Mean number of events in RA of 3° bin (background)*	*Number of events in the "signal" bin, RA of 7.4 – 7.6 hours*	*Number of "signal" events with stat. errors*
$>10^5$	*73382*	*611*	*663*	*52±25*
$>5x10^5$	*7123*	*58*	*84*	*26±7*
$>8x10^5$	*3282*	*26*	*57*	*31±5*
$>10^6$	*2225*	*18*	*57*	*25±4*
$>2x10^6$	*573*	*4*	*13*	*9±2*

Obviously it is again random variable with variance controlled by the variance of the "background". The statistical errors in the Table 48 and Table 49 illustrate that signal events number obtained in "best confidence" bin and equals to 25 is consistent with both enlarging of the bins and lowering the shower size cut. We have not enough strong statistical arguments to reject this hypothesis, because 25 value is within approximately one σ intervals of considered cut values and enlarged bin sizes.

Note also very good scaling of mean bin rate with enlarging bin size (Table 49). Nevertheless, we didn't claim that 25 is the best estimate of the signal, for checking the statistical hypothesis on the best signal value we need to tune more precisely the shape of the signal domain using techniques described in (Chilingarian, 1994).

Table 49. CR source Localization around the center of the "signal" bin.

Bin size (α x δ)	*Number of events in the chosen declination band*	*Mean number of events in RA of 3° bin (background)*	*Number of events in the "signal" bin*	*Number of "signal" events with stat. errors*
1 x 1	*744*	*2*	*11*	*9±2*
2 x 2	*1468*	*7*	*22*	*15±4*
3 x 3	*2225*	*18*	*43*	*25±6*
4 x 4	*2952*	*32*	*48*	*16±8*
5 x 5	*3739*	*51*	*71*	*20±10*

We understand that for proving high energy cosmic ray source we need to do additional tests. In 1983 Stamm and Samorski using Kiel surface detector data publish papers

(Samorski et al., 1983), (Samorski et al., 1984) claiming detection of signal from direction of Syg X-3. The chance probability was estimated to be ~2 from 100,000. Further attempts to confirm cosmic ray source by more precise detectors CASA-MIA, HEGRA and others fall to find any significant enhancement. Despite that significance of detection of Monogem Ring is an order of magnitude higher (chance probability is 2 from million), we continue to performs different statistical tests to reveal any of systematic biases.

One of such tests is random permutation test. We want to check if fixing of the bin center, as we have done with Monogem Ring, can introduce bias in the significance test, i.e., if it is possible by selecting special directions in equatorial coordinates obtain randomly such large excesses in bin count rates. To check for it we perform special statistical test. The test consists in random choice of bin center (and corresponding declination band) and further count of number of the MAKET-ANI events fallen in the 2400 RA bins of the size (3° x 3°). We repeat this procedure 1000 times and for each random trial apply the normalization according to equation 1. For generated is such way 1000 rectangular grids we calculate the "sigma" values for each of 2400 originated (3° x 3°) bins and select the bin (direction) with maximal sigma. In Figure 77 we demonstrate scatter plot of the equatorial directions, corresponding to the $\sigma \geq 4$, and in Figure 78 appropriate directions correspondent to the $\sigma \geq$

And if several regions we see 4 σ excess (to be examined separately) only directions compatible with the Monagem ring demonstrate significant excess greater than 5σ.

We also use this "random search" technique for the detailed scanning of the signal region. In Figure 79 and Figure 80 we post scatter plots corresponding to the directions demonstrated significant excess, $\sigma \geq 4$, 5, but in this studies we restrict the random directions by the Monogem Ring vicinity, for detailed investigation of source under question.

As we can see from Figure 79 and Figure 80, bin centers have circular symmetry with center approximately coinciding with direction (750+14), the largest sigma value was obtained with equatorial coordinates (749+14.08). Therefore, we conclude that concentration of the EAS directions in the bin centered (750+14), see Figure 81 and Figure 82, isn't methodical effect. With chance probability 2 from million us really have additional flux from Monogem Ring.

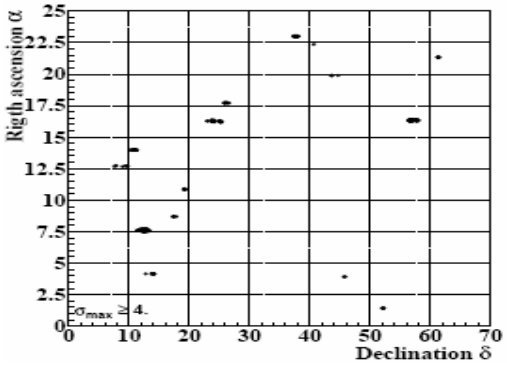


Figure 77. Random scanning star sky, directions to RA bins demonstrating significant excess, σ > 4.

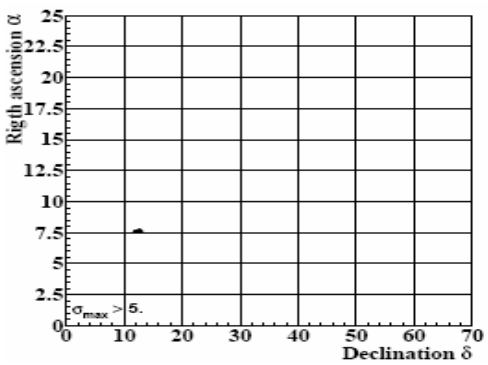


Figure 78. Random scanning star sky, directions to RA bins demonstrating significant excess, σ > 5.

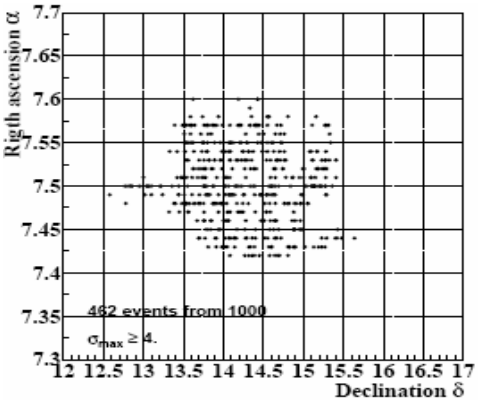


Figure 79 .Random scanning Monogem Ring, directions to RA bins demonstrating significant excess, σ > 4.

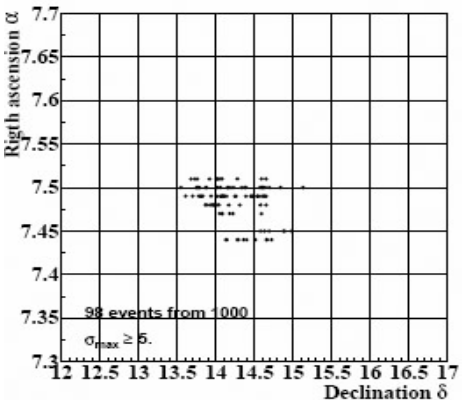


Figure 80. Random scanning Monogem Ring, directions to RA bins demonstrating significant excess, σ > 5.

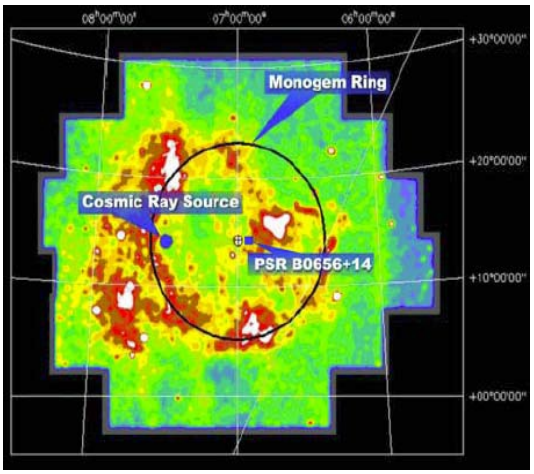


Figure 81. CR Source within Monogem Ring.(MR as seen in the ROSAT all-sky survey in the 0.25-0.75KeV)

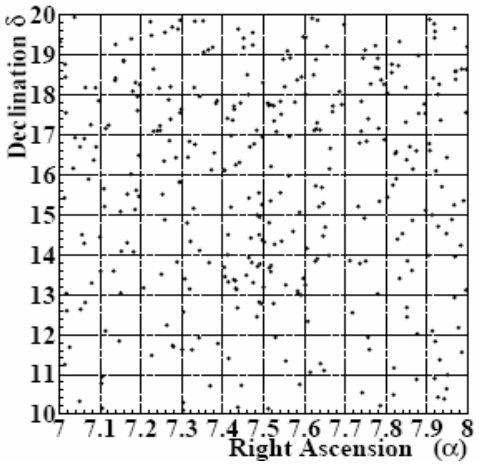


Figure 82. Monogem Ring Coverage, center of signal bin 750α+14δ, size 3° x 3°.

CHAPTER 9.
"SPECTROSCOPY" OF THE CR SOURCE

Next step of analysis of the CR point source is investigation of mass composition and energy spectra of the primaries coming from the direction of Monogem Ring and comparisons with the overall isotropic flux of CR of same energies. In the Figure 83 and Figure 84 the EAS 2- way classification into light and heavy mass groups is illustrated. In Figure 69 a) all 120 RA bins in declination band 12.5 – 15.5 degrees are posted, in Figure 84 the same is done for 43 events from the "signal" bin. The neural classification again was done with $N_e - s$ pair and selection criterion was $N_e > 106$. No significant difference is detected.

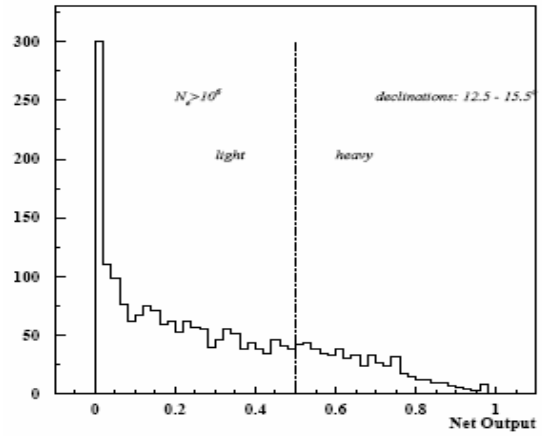


Figure 83. The 2-way neural classification of EAS registered from declinations 12.5 –15.5º

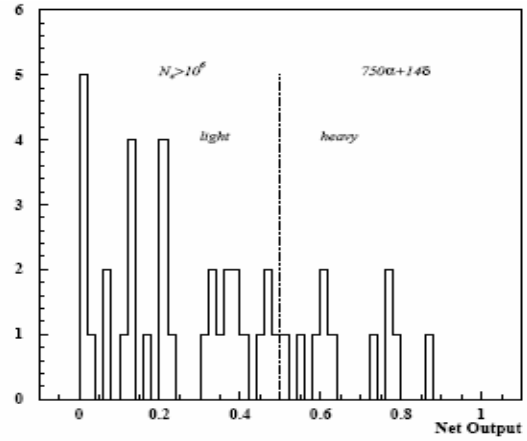


Figure 84. The 2-way neural classification of EAS registered from SS direction: 750α+14δ

In Figure 85 and Figure 63 the energy spectra of light and heavy mass groups separately for the all declination band and "source" bin are depicted.

The proportion of heavy nuclei is the same in the both samples and obviously is enlarging with selecting showers corresponding to primaries with higher energy. The heavy mass group spectra is shifted to the right relative to light mass group, because showers initiated by heavy nuclei are attenuated faster in the atmosphere and constant shower size will correspond to larger heavy nuclei group energies comparing with light group.

Information contained in Figure 83 and Figure 84 didn't allow us to make any definite conclusion about particular nuclei mass group coming from direction of SS, i.e. Oxygen nuclei as propose Erlykin and Wolfendale (Erlykin et al., 1997).

Another comparison between declination band means and signal bin means concern angular accuracies. In Figure 87 and Figure 88 we post the distribution of the azimuthal and zenith accuracies for the declination band 12.5 – 15.5º. In Figure 89, Figure 90 the same distributions for only signal bin are posted. We again see no significant difference in band and bin distributions.

Another important problem is transition from polar angular accuracies to the accuracies of the equatorial coordinates.

We denote quatorial coordinates (α,δ), z-axis is directed to North Pole and Althazimuth coordinates by (h,A), z-axis is directed to zenith, y-axis is directed to East, x-axis is directed to South.

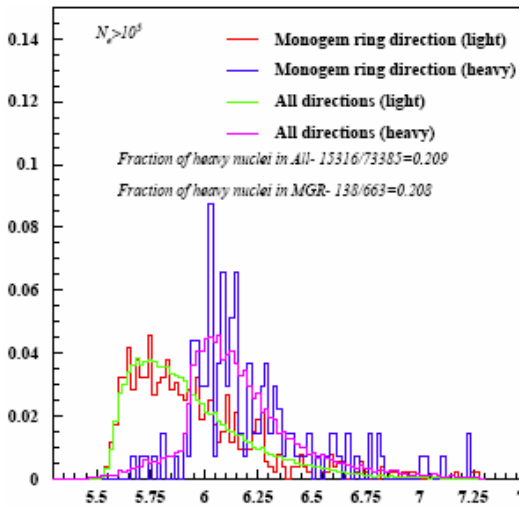


Figure 85. Energy Spectra of light and heavy mass groups coming from declinations 12.5 – 15.5° comparing with SS direction: 750α+14δ, shower size Ne >10^5

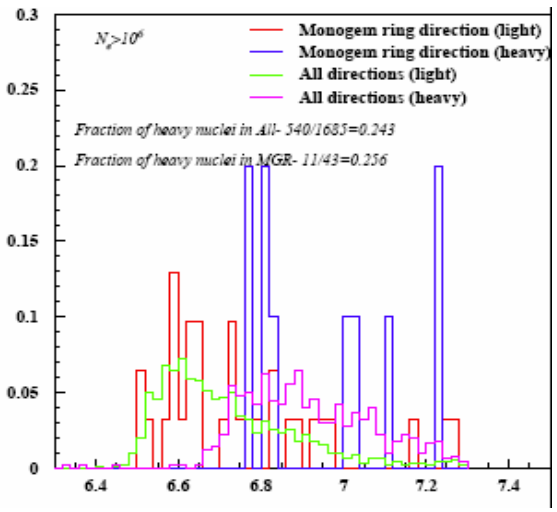


Figure 86. Energy Spectra of light and heavy mass groups coming from declinations 12.5 – 15.5° comparing with SS direction: 750α+14δ, shower size Ne >10^6

Relationship between 2 coordinate systems is as follows

$$\begin{aligned} sin\delta &= -cos\Phi coshcosA + sin\Phi sinh \\ cos\delta cos\alpha &= sin\Phi coshcosA + cos\Phi sinh \\ cos\delta sin\alpha &= coshsinA \end{aligned} \qquad (8.2)$$

where Φ is altitude of Polar Star above the horizon in altazimuth coordinates From 8.2 we obtain

$$\begin{aligned} \frac{\partial\delta}{\partial h} &= \frac{cos\Phi sinhcosA+sin\Phi cosh}{cos\delta} \\ \frac{\partial\delta}{\partial A} &= \frac{cos\Phi coshsinA}{cos\delta} \\ cos\delta &= \sqrt{(sin\Phi coshcosA + cos\Phi sinh)^2 + (coshsinA)^2} \end{aligned} \qquad (8.3)$$

and

$$\begin{aligned} \frac{\partial\delta}{\partial A} &= -\frac{sin^2\alpha(Asin\Phi+cos\Phi tghcosA)}{sin^2A} \\ \frac{\partial\delta}{\partial h} &= \frac{sin^2\alpha cos\Phi}{sinAcos^2h} \\ sin^2\alpha &= \frac{1}{1+(sin\Phi ctgA+cos\Phi tgh/sinA)^2} \end{aligned} \qquad (8.4)$$

Relationship between the small angles Δh, ΔA in altazimuth system and Δα, Δδ in equatorial system is the following

$$\begin{aligned} \Delta\alpha &= \frac{\partial\alpha}{\partial h}\Delta h + \frac{\partial\alpha}{\partial A}\Delta A \\ \Delta\delta &= \frac{\partial\delta}{\partial h}\Delta h + \frac{\partial\delta}{\partial A}\Delta A \end{aligned} \qquad (8.5)$$

For the Monogem Ring co-ordinates seen from the Mt. Aragats and for obtained

MAKET- ANI detector accuracies we obtain $\Delta\alpha = 1.3^{o}$□, $\Delta\delta = 1.2^{o}$, each more than 2 times less comparing with size of the bin of the equatorial co-ordinates grid, used for the source searching.

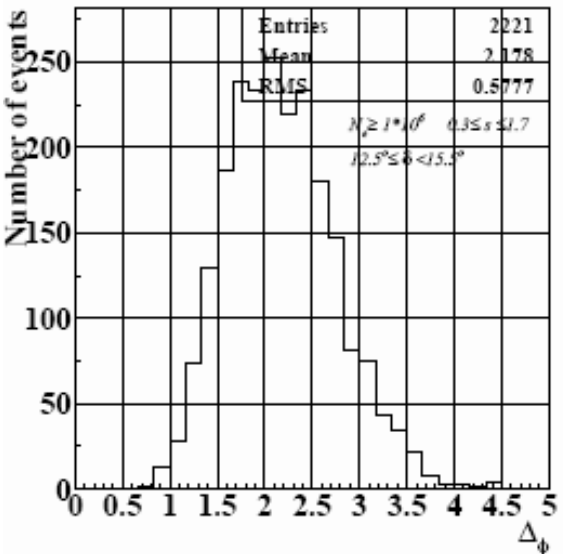


Figure 87. Distribution of the azimuthally accuracies of Maket-Ani detector, declination band 12.5 – 15.5º.

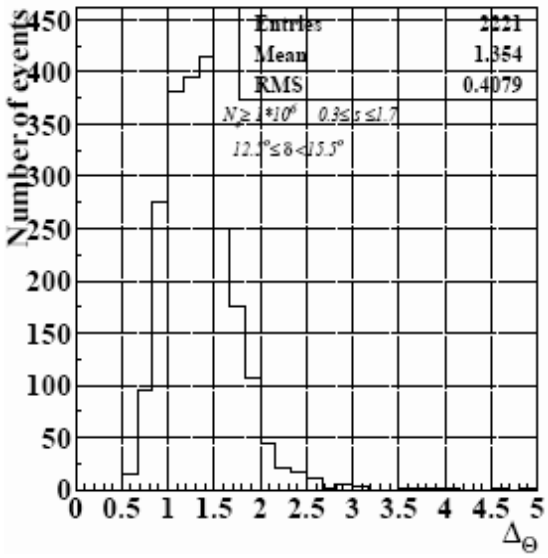


Figure 88. Distribution of the zenith accuracies of Maket-Ani detector, declination band 12.5 – 15.5º.

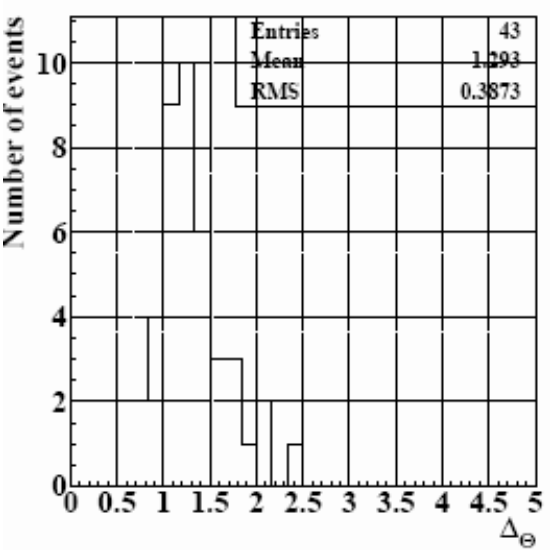


Figure 89. Distribution of the azimuthally accuracies of Maket-Ani detector in signal bin, center of signal bin 750α+14δ, size 3° x 3°.

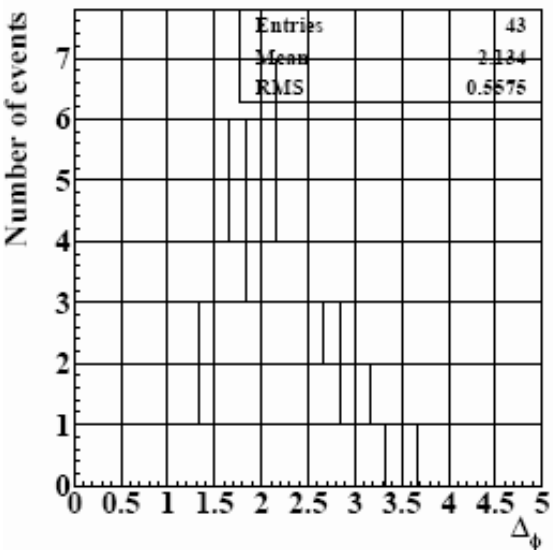


Figure 90. Distribution of the zenith accuracies of Maket-Ani detector in signal bin, center of signal bin 750α+14δ, size 3° x 3°. declination band 12.5 – 15.5º.

9.1 WHAT WE CAN LEARN FROM SOLAR ACCELERATORS

Starting from the 70s, with the launch of particle spectrometers, began the continuous monitoring of the low and medium energy cosmic rays in space. Time histories of the simultaneously detected X-rays, gamma-rays, electrons, and ions of different energy and charge, combined with the detection of the developing flares and Coronal Mass Ejections

(CME) using coronagraphs, helped to create a comprehensive picture of the major solar events, accelerating protons to high energies, so called, Solar Energetic Proton (SEP) events (Reames, 1999). SEP events include also highest energy ions, accompanying protons in giving rise to Ground Level Enhancements(GLE), additional fluxes of secondary cosmic rays, mostly neutrons and muons, detected by the world-wide network of the Neutron Monitors and Muon Telescopes. "New Instruments on WIND and ACE satellites operating during the 23-rd solar cycle, with geometry factors ~100 times larger than those of the previous cycle, have yielded unprecedented observations of temporal evolution in composition and spectra over a wide range of energies and species" (Tylka et al., 2001B).Multiwavelength measurements from very sensitive X-ray detectors, high resolution imaging coronagraphs and radiotelescopes now reveal the location and characteristics of the natural accelerators at the Sun and in the interplanetary space in much more details.

Impulsive flare events, are believed to accelerate electrons and ions in large structures originating in the magnetic reconfiguration regions. After discovery of the above-the-loop-top hard X-ray source (Masuda et al., 1994) with the Yohkoh/HXT (Kosugi et al., 1991) it became apparent that particles are accelerated by the dynamic electromagnetic forces during the reconfiguration of the magnetic fields (Ashwanden et al., 1996). The most probable acceleration mechanism is the stochastic acceleration, allowing detectable intensities of nonthermal X-ray radiation from locally trapped electrons. Direct hard X-ray detection, as well as application of the time-of-flight technique to the electrons traveling from acceleration site to the chromosphere reveals that the location of the acceleration region is 5,000 – 35,000 km above the top of the soft X-ray-bright flare loop (Ashwanden, 2002).

The natural assumption that positively charged protons and ions will be accelerated with the same mechanisms as the electrons is proven by the registration of the lined gamma radiation in coherence with hard X-ray radiation. The time sequence of the bremsstrahlung radiation peaks produced by accelerated electron beams, interlaced by the nuclear de-excitation lines produced by proton and ion bombarded chromosphere, clearly demonstrate, that ions and electrons are accelerated in the same region and nearly simultaneously.

The efficiency of the stochastic acceleration of ions via the mutual wave-particle interactions depends on the relation between the frequencies of the resonant waves (Alfven waves, magnetosonic waves, sound waves) and ion gyrofrequency. Alfven waves, if fast enough (~2000 km/sec) can accelerate 20 KeV protons up to GeV energies during time scales of 1-10 sec. (Barbosa, 1979), (Miller et al., 1990).

Gradual events are associated with CME development in corona and in interplanetary space. CME driven shock should be fast enough (> 500 km/sec, Reames, 1999) to produce SEP events. Shock acceleration is believed to be one of the major mechanisms in the Universe for accelerating particles to highest energies. Multiple traversals of shock are required for the acceleration of solar ions up to MeV energies. Ambient magnetic turbulence is not sufficient for scattering and trapping ions with such energies. Self-generated Alfven waves effectively scatter energetic ions, providing their trapping near the chock and, therefore, increasing their energy. Maximum attainable energy of accelerated ions is proportional to the rate of re- crosses of the shock; this rate in turn is proportional to the particle trapping time. "As trapping increases for particles of one rigidity, they are more likely to be accelerated to a higher rigidity, where they again stream out and produce resonant waves, etc" (Reames, 2000).

Numerical calculations and Monte Carlo simulation proves that solar protons could be accelerated up to energies of 100 GeV during propagation of CME in middle and high corona (Miller et al., 1990). The same authors examining the 1982 June 3 flare, mention that protons were accelerated within 16 seconds from 30 MeV to ~ 1GeV. Krucker and Lin, (Krucker et al., 2000), based on the data from WIND/SST instrument (Lin et al., 1995), conclude that protons at energies up to 6 MeV are injected simultaneously at heights ≤ 10 R☼. The maximum energy attainable by the shock acceleration depends on shock speed and height of

shock start in the corona. Shock waves as fast as ~1500 km sec^{-1} starting below ~5R☼ can accelerate ions up to 10-30 GeV (Tylka et al., 2001), (Tylka et al., 2001A). Study of the association between SEP events and CME (Gopalswamy et al., 2002) proves that CME interaction is important for high energy SEP production. For most of the SEP events detected in1997-2001 the primary, fast CME overtakes one or more slower CMEs within a heliocentric distance 0f ~ 20 R☼. The summary of the present knowledge on particle acceleration by various mechanisms at Sun and in the interplanetary space can be summarized as follows:

- Electron Accelerators also accelerate protons and heavier ions, acceleration sites are very close in space and time;
- Particle acceleration is much more effective when several shocks are present in the interplanetary space.
- The "magnetic bottle" structures formed by interacting shocks are major sites for reacceleration of particles primarily accelerated by "impulse" and "gradual" mechanisms;
- Maximal attainable energy of particle accelerators is proportional to the particle charge;
- Moving Shock is carrying bulk of particles;
- Maximal attainable energy of particle is changing from event to event and depends on total energy of solar blast, on speed of shock wave, and on time-temporal history of the solar flare. (Positions of the Spectral "knees" change from 10 MeV to several GeV).
- For detection of charged particles on the Earth, shock should intercept observers magnetic tube;
- The streaming limit controls the transport of particles;

The rigidity dependent maximal acceleration energy in solar Energetic Events (SEP) occurred during current 23rd solar cycle is apparent from Figure 91. We again, like for galactic cosmic rays see very sharp knee for the light nuclei group, namely protons, and no knee for heave nuclei group, namely iron. Of course, for much less energies, as Sun and CME driven shocks are much less energetic comparing with Supernova explosions and correspondent shocks.

The most famous, so called Bastille day SEP from 14 July 2000, as you can see from demonstrates remarkable exactness of the knee positions according to accelerated ion charge: proton knee is at ~20 MeV, He knee at ~40 MeV and Carbon knee (Carbon charge is equal to 5, remember that temperature at Sun atmosphere isn't too high to fully strip ions, like at Supernova explosion sites) at ~100 MeV (Tylka et al., 2001a).

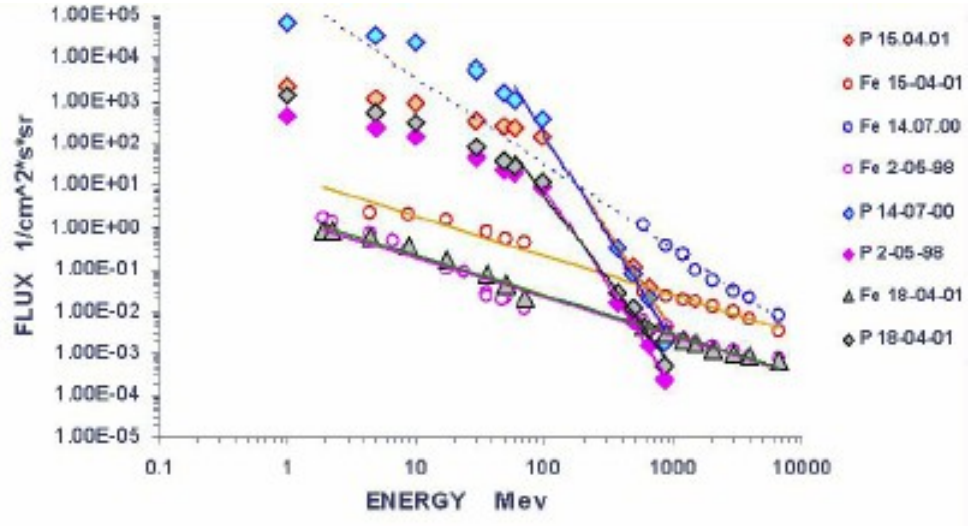


Figure 91. The energy spectra of protons and Fe ions registered by space-born detectors during SEP event of 23-rd solar cycle.

9.2 DISCUSSION AND CONCLUSIONS

Recent unprecedented detailed observations of the nonthermal X-ray radiation from SN 1006 made by CHANDRA (Long et al., 2003), point very definitely on SNR as a host of the hadron accelerators providing energy at least till several units of 10^{14} eV. Observations of the wind synchrotron nebulae around pulsars in vicinity of the SNR center reveals another accelerator site e.g. the termination shock at which the relativistic shock from pulsar wind is forced to join the slower expansion of the outer nebula (Gaensler, 2003),. Therefore we have at least 2 independent accelerators in the SNR working at near, but not coinciding energetic scales and differing also at particle production spectra.

The Single Source model of Erlykin and Wolfenfdale (Erlykin et al., 1997), (Erlykin et al., 1998), also get huge support by discovery of the nearest pulsar, located at the space-temporal distance in remarkable concordance with SS model expectations (Thorsett et al., 2003).

In the light of recent discoveries evidence from MAKET-ANI experiment confirms SNR and SS models of Cosmic Ray Origin.

Very sharp knee of the energy specter of the light mass group suggest accepting SS hypothesis, because it is highly improbable that Galaxy ensemble of distant Supernovae with variety of explosion energies, shock wave speeds, distances and explosion times will provide sharp knee feature. Instead we should expect rather smooth depletion of the light mass group flux.

Very sharp knee of light mass group and absence of knee in heavy mass group till at least 10 PeV also confirm hypothesis of rigidity dependent maximal energy of SNR accelerators.

The value of this maximal energy is dependent on the total energy of explosion, on speed of the shock wave, on density of gas nearby SNR, magnetic fields and its spectra distribution and many other, random circumstances connected with supernovae explosion and its development history. Therefore, we expect that each supernovae has its own maximal energy. Proceeding from the obtained light group spectra and remembering that we have contamination of heavier masses in it we can estimate maximal energy of Monogem Ring SNR ~10^{15} eV.

MAKET ANI and data allow us to summaries experimental evidence in the following statements:

The energy spectrum of the "heavy" mass group of cosmic rays shows no "knee" in the energy interval of 10^{15} - 10^{16} eV.

The estimated energy spectrum of the "light" mass group of nuclei shows a very sharp "knee" $\Delta\gamma \sim 0.9$ compared to $\Delta\gamma \sim 0.3$ for the all-particles energy spectra.

Proceeding from the experimental evidence we conclude that:

The SNR acceleration model is supported by the MAKET-ANI data on partial energy spectra;

_The nearest SNR, produces a significant portion of the high energy cosmic rays in the "knee" region.

- the mean logarithmic mass is increasing with primary energy.
- the obtained spectra of light and heavy ions and mass composition coincide well with analogical characteristics estimated by the KASCADE experiment using the same simulation code and data analysis methodology.
- the MAKET-ANI experiment detects significant excess of particles from the direction of the Monogem Ring with chance fluctuation probability of 2 per million.
- the maximal energy of surplus particles from the Monogem Ring is estimated to be $\sim 5 \cdot 10^{15}$ eV.
- the position of the CR source, 750α+14δ, is fully consistent with the SN shock propagation.
- the energy spectra and mass composition of the particles coming from the “source” direction didn’t differ significantly from isotropic background cosmic rays.

Proceeding from the experimental evidence we can conclude that:

- Nearest SNR – the Monogem Ring produce significant portion of high energy cosmic rays in the knee region;
- Maximal energy of the SNR is proportional to the particle charge: $E_{MAX} \sim ZE_{max}$, with $E_{max} \sim 10^{15}$ eV.
- Our conclusions are consistent with evidence we get recently about work of the solar accelerators. We detect same behavior of the “light” and “heavy” primaries spectra, and same rigidity dependent cutoffs in solar particle fluxes, of course, on the much less energy scales. Recently propose mechanisms of particle acceleration in the SN1006 (Berezhko et al., 2003) is fully consistent with mechanisms of solar particle acceleration by CME driven shocks, of course, again, at much less scales.
- The time history of the cosmic ray intensity (Schlickeiser, 2002), suggest 50% enhancement of CR flux integrated over last 400,000 Ears comparing with all available time record of 10^9 Ears. It is also consistent with nearby SS model.

The data collected by the MAKET-ANI detector in 1997-2002 is the property of the ANI collaboration We thank the ANI collaboration members for their fruitful cooperation over many years, and A.Haungs, H.Rebel and M.Roth for multiyear cooperation in development of multivariate methods of data analysis for KASCADE experiment and for kind permission to use Figure xy. This work was supported by Armenian government grant 1465, by the ISTC A216 grant, and by the INTAS IA-2000-01 grant.

CHAPTER 10.
SPACE WEATHER FORECAST

10.1 INTRODUCTION

Unpredictable bursts of *Solar Energetic Particles (SEP),* peaking in 11 year cycles are one of the major constraints on the operation of space systems and further technological utilization of near-Earth space (Tylka et al., 2001B). Some of these bursts produce fluxes of high energy particles which can be harmful to satellite electronics, the Space Station, its crew and to flights over the poles. In the 1999 report on space weather, the US National Security Space Architect finds that during the preceding 20 years about one or two satellites per year have suffered either total or partial mission loss due to space weather (SSB, 1999). Since our lives depend so heavily on the satellite based technologies, not to mention the value of protecting humans in space and in aircraft, it is becoming increasingly important to have an accurate and reliable forewarning about the arrival of these dangerous particles, so that mitigating action can be taken if necessary.

The use of large-area detectors which can only be accommodated at ground based stations is vital for measuring the low fluxes of high energy particles accelerated in the vicinity of the Sun due to Solar Flares (SF) and Coronal Mass Ejections (CME). The high energy particles from the most severe events which can cause damage arrive to Earth about a half hour earlier than the abundant "killer" medium energy particles, thus providing an opportunity to establish an early warning system to alert the client about the potential damage to satellites, the Space Station, space personnel, and flights scheduled over the poles (Dorman 1999). Taking into account that only very few of a great number of SF and CME produce dangerous ion fluxes, it is not only critical to alert clients about the arrival of the most severe radiation storms, but also to minimize the number of false alarms against events which are not severe enough to cause damage. We can accomplish both goals by using detectors on mountain altitudes and low latitudes to detect secondary fluxes generated by the few high-energy ions as they enter the Earth's atmosphere. Because the high energy ions are so few in number and because secondary particles are scattered and attenuated in the Earth's atmosphere, large-area detectors, located at high mountain altitudes are necessary to measure them. The information about primary ion type and energy is mostly smeared during its successive interactions with atmospheric nuclei, therefore, only coherent measurements of all secondary fluxes (neutrons, muons, and electrons), along with their correlations, can help to make unambiguous forecasts and estimate the energy spectra of the upcoming dangerous flux.

Lev Dorman has demonstrated in numerous papers that detecting at least two or three cosmic ray components at different altitudes and latitudes will make it possible not only to reconstruct the solar ion flux outside the Earth's atmosphere, but also to estimate the energy spectra of upcoming solar particle fluxes (Dorman et al., 1993), (Dorman et al., 1993A). Multidimensional statistical methods of analysis of the multivariate time series as well as timely delivery of the alert are also of utmost importance (Chilingaryan et al., 1999E).

10.2 PARTICLE ACCELERATION AT THE SUN

Galactic cosmic rays (mostly protons and fully stripped ions) incident on the atmosphere create fluxes of secondary particles reaching the Earth. The intensities of these fluxes, as measured at different geographical positions by various particle detectors, are characterized mostly by geophysical conditions of the site, by the particular phase of solar activity, and by the local time of day. Transient solar events also influence the count rates of particle monitors. In this report we will examine short

term enhancements of count rates in surface detectors (so called, Ground Level Events – GLE) correlated with the fluxes of relativistic solar ions incident on the Earth's atmosphere.

The first experimental evidence of the GLE was observed in 1942 with Ionization Chambers at Cheltenham, Maryland by Scott Forbush (Forbush, 1946). Correlating detected GLEs with large SF, Forbush concluded that the cause of the rise in the detector counts is the flux of charged particles, accelerated during large disturbances on the Sun and reaching Earth.

Established in the 50's, the world-wide network of Neutron Monitors (NM) provides more detailed data on GLE. Mutual analysis of data on GLE detected by NM located at different latitudes leads to the estimation of the energy spectra of the solar proton flux and it's time dependence, (Meyer et al., 1956).

In the 60's satellite detectors measured precise proton energy spectra for energies from 1 MeV up to 500 MeV and helped to establish correlations between increases of particle fluxes at 1 AU with abrupt decreases of surface detector count rates, called Forbush decreases (Fd), and with sudden commencement of the geomagnetic storms (Bryant et al., 1963).

Starting from the 70's, with the launch of particle spectrometers, began the continuous monitoring of the low and medium energy cosmic rays in space. Time histories of the simultaneously detected X- rays, gamma-rays, electrons, and ions of different energy and charge, combined with the detection of the developing flares and CME using coronagraphs, helped to create a comprehensive picture of the major solar events that include also highest energy ions giving rise to GLEs (Reames, 1999). "New Instruments on WIND and ACE satellites operating during the 23-rd solar cycle, with geometry factors ~100 times larger than those of the previous cycle, have yielded unprecedented observations of temporal evolution in composition and spectra over a wide range of energies and species" (Tylka et al., 2001B).

Multiwavelength measurements from very sensitive X-ray detectors, high resolution imaging coronagraphs and radiotelescopes now reveal the location and characteristics of the natural accelerators at the Sun and in the interplanetary space in much more details. Given this background information, two types of solar events which accelerate particles – impulsive and gradual - were categorized and described in numerous publications --see, for example, Miroshnichenko (Miroshnichenko, 2001).

10.3 THE ARAGATS SPACE ENVIRONMENT CENTER

The Aragats Space Environmental Center (ASEC, Chilingarian et al., 1999a) consists of two high altitude stations on Mt. Aragats in Armenia (Geographic coordinates: 40°30'N, 44°10'E. Cutoff rigidity: ~7.6 GV, altitude 3200 and 2000 m.). At these stations several

monitors continuously measure the intensity of the cosmic ray fluxes and send data to the Internet in real time (see Table 50, and (Chilingarian et al., 2003), for detailed description of ASEC monitors).

After 50 years of operational experience, neutron monitors with threshold values from ~1 GV (in the Polar Regions) to ~15 GV (Equatorial Regions) continue to be the best instrumentation for measuring intensity variations of cosmic rays (Moraal et al., 2000). In the 60's Carmichael developed a neutron monitor with statistical accuracy of 0.1% for hourly data in preparation for the Year of Quiet Sun (IQSY), (Carmichael, 1964). This type of neutron monitor is usually designated by the name X- NM-64 where X denotes the number of counters of the entire monitor. For more details and for a list of world-wide monitors see (Shea et al., 2000).

Two 18NM-64 neutron monitors are in operation at Nor-Amberd (2000 m. elevation), and at Aragats, (3200 m. elevation) research stations respectively called the Nor Amberd Neutron Monitor (NANM), and the Aragats Neutron Monitor (ANM). The monitors are equipped with interface cards, providing time integration of counts from 1 sec up to 1 minute. Real-time data from these monitors is available at URL http://crdlx5.yerphi.am .

The Solar Neutron Telescope (SNT-1) at the Aragats station is part of a world-wide network coordinated by the Solar-Terrestrial Laboratory of the Nagoya University (Matsubara et al., 1999), (Tsuchiya et al., 2001). An important advantage of the SNT over the NM is its possibility to estimate the energy of detected neutrons. The SNT consists of four 1 m2, 60 cm thick scintillation blocks with anti-coincidence shielding consisting of four plastic scintillators 5 cm thick and 1 m2 each which veto the near vertical charged flux. Incoming neutrons are converted to protons in nuclear interactions inside the thick scintillator target. The energy deposited due to ionization by recoil protons is measured by photomultipliers over the scintillators. The amplitude of the photomultiplier output signals is discriminated according to 4 threshold values, approximately corresponding to neutron energies of 50, 100, 150 and 200 MeV. The count rate of SNT-1, measuring large amounts of neutrons and inclined muons and electrons is sensitive to transient solar events. We use short term variations of the SNT count rate along with Aragats and Nor Amberd neutron monitors data to analyze the of GLE and Forbush decreases (Chilingarian et al., 2003). Data from SNT-1 is available online at URL http://crdlx5.yerphi.am.

One of the improvements to the Aragats monitoring facilities includes registration of the variations of the muon flux under different angles of incidence. The Nor-Amberd Muon Multidirectional Monitor NAMMM, shown in Figure 1 consists of two layers of plastic scintillators above and below one of the three sections of the Nor Amberd NM. The lead (Pb) filter of the NM absorbs electrons and low energy muons. The threshold energy of the detected muons is estimated to be 350 MeV. The NAMMM consists 6 up and 6 down scintillators, each having the area of 0.81 m2. The distance between layers is ~ 1 m., and the mean angular accuracy is ~ 25°. The data acquisition system of the NAMMM can register all coincidences of detector signals from the upper and lower layers, thus, enabling measurements of the arrival of the muons from different directions.

At the Aragats high altitude station two surface arrays, MAKET (Hovsepyan, 1998) and GAMMA (Garyaka et al., 2002) are in operation for the main purpose of detecting Extensive Air Showers (EAS) initiated by very high energy (E >5·1014 eV) ions and protons. The EAS installations are triggered 3-5 times per minute. The

plastic scintillator-photomultipliers assemblies symmetrically distributed over the ground surface are used to measure charged particle densities and arrival times. The total area of the detectors of GAMMA and MAKET installations is about 150 m2. The spacing between detectors varies from several meters to tens of meters.

In the underground hall originally constructed for the ANI Cosmic Ray experiment (Danilova et al., 1992) another one hundred fifty of the same type of detectors are located to measure the muon content of the EAS. The 6 m thick concrete blocks plus 7 m soil filter the electrons and the low energy muons. Thus, only muons with energies > 5GeV reach the detectors. The count rates of the charged components at mountain altitudes are ~ 420 counts/m^2/sec for >10 MeV electrons and ~ 50 counts/m2/sec for >5 GeV muons. These high count-rates combined with the large area of the electron and muon detectors on Mt. Aragats are very attractive for establishing a monitoring facility to investigate the correlations between short term variations of electron and muon count rates with the enhancing flux of solar ions incident on the Earth's atmosphere.

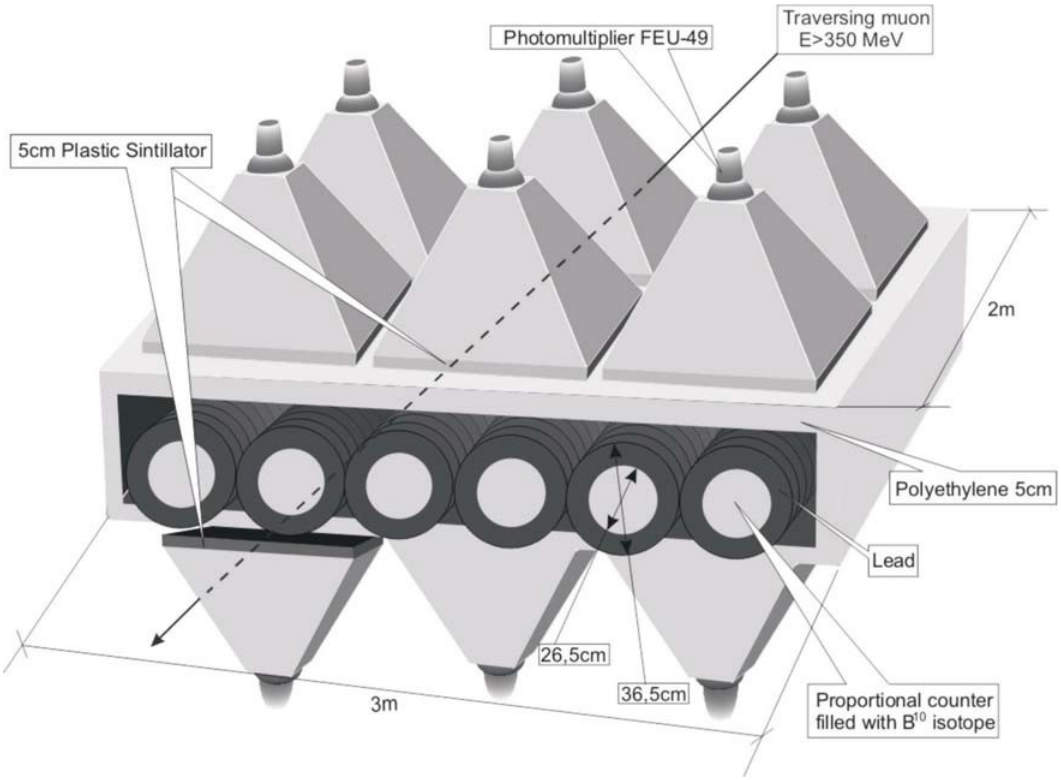


Figure 92. Nor Amberd Multidirectional Muon Monitor installed around the 8-NM-64

The count enhancements of the ASEC neutron monitors are integrated over all directions. In that case the question arises whether the signal enhancement is due to the solar particles or disturbance of the Earth's magnetic field inducing a temporary decrease of the local rigidity threshold--see, for example, (Kudela et al., 2001). The scattering of the high energy muons in the atmosphere is negligible, therefore by measuring the incident muon direction we can determine the direction of the coming solar and galactic ions. It will give us additional evidence on the registration of solar particles. The Aragats Multidirectional Muon Monitor (AMMM) consist of 15 m^2 scintillation detectors located on the top of the ANI concrete calorimeter and 72 m^2 array of same type detectors 24 m below. Using the coincidence technique we can monitor changing count rates from numerous space

directions. Detectors on the top are grouped in 3, while those in the underground hall are grouped in 8 to provide significant amount of coincidences. The geometry of the detector arrangement will allow us to detect directions from the vertical to 60º declination with accuracy of ~ 5º with very good statistics.

Measuring the intensity deficit of the galactic cosmic rays, it will be possible to determine the loss cone direction and perform "screening" of approaching magnetized plasma cloud. The world-wide network of the muon monitors covering as much as possible incident directions could be used for the early forecasting of the upcoming severe geomagnetic storm (Munakata et al., 2000). Along with the Moscow TEMP muon telescope (Borog et al., 2001) the AMMM could improve sky coverage when combined with present muon detector network.

The top layer of the AMMM can be used separately as the electron & low energy muon monitor with sensitivity of 0.15% for one minute counts. The short scale variations in the low energy charged particle fluxes are of major interest for correlation analysis with high energy muon and neutron data. The lower layer of the AMMM constitutes a very sensitive high energy muon monitor, robust to local atmospheric conditions due to the rather high energy threshold. Total count rate of the monitor is approximately ~200,000 per minute. Thus, the sensitivity of this monitor reaches record value of ~0.2% for one minute count rates, 3 times better compared to the Aragats NM. Simultaneous detection of variations in low energy charged particles, in neutron and high energy muon fluxes by ASEC monitors with characteristics given in Table 50, will provide new possibilities for investigating the transient solar events and will allow us to classify GLEs according to their origin and physical nature.

Table 50. Parameters of the ASEC Monitors at Mt. Aragats in Armenia.

Detector	*Altitude m*	*Surface m^2*	*Threshold(s) MeV*	*In operation since*	*Mean count rate (min^{-1})*
NANM (18NM64)	*2000*	*18*		*1996*	*2.5×10^4*
ANM (18NM64)	*3200*	*18*		*2000*	*6.2×10^4*
SNT-1	*3200*	*4*	*50,100,150,200*	*1998*	*$6.7 \times 10^{4*}$*
NAMMM – test operation	*2000*	*5 + 5*	*350*	*2002*	*$2.5 \times 10^{4**}$*
AMMM – test operation	*3200*	*15 + 72*	*5000*	*2002*	*$2 \times 10^{3**}$*

*Count rate for the first threshold; near vertical charged particles are excluded

Expected total coincidences rate for the near vertical muon flux

10.4 CALCULATED ARRIVAL TIMES OF THE RELATIVISTIC SOLAR IONS AT 1 AU AND ENHANCEMENTS REGISTERED BY ASEC MONITORS

The expected arrival times of the SEP relativistic ions at 1 AU are calculated by the technique proposed in (Lockwood et al., 1990), (Fluckiger, 1991) and successfully applied by Krucker and Lin (Krucker et al., 2000) for data analysis from the Spectroscopic Survey Telescopes on board of WIND satellite (Lin et al., 1995). We apply the same techniques to

estimate the relativistic particle arrival at location of the ASEC monitors. We use the arrival times and energies of the first ions registered by the Solar Isotope Spectrometer (SIS) on board of the ACE satellite (Stone et al., 1998), and protons, registered by GOES satellites (GOES, Internet). We assume that the first ions of all energies are accelerated in the one and the same spatial region and that interplanetary propagation of the high energy ions is essentially scatter-free--see (Kahler, 1994). It was demonstrated in Krucker and Lin (Krucker et al., 2000) that the path length of 6 MeV protons to 1AU is very close to the Parker spiral length. Therefore, the arrival times of the ions of different energies will be linearly correlated with the inverse of their speed.

We can extrapolate the velocity-time relationship to calculate the expected arrival times of the first relativistic ions that are energetic enough to enter the atmosphere at the Aragats geographical location and produce secondary fluxes that reach the ASEC monitors. We calculate the threshold values of ion velocities (unique to each ion) for which secondary particles will reach the Aragats monitors. Then we check the correctness of our assumptions by calculating the correlation coefficients of the linear regressions. The reconstructed regression lines for some of the events are shown in Figure 93. All correlations are greater than 0.96, thus justifying the validity of the proposed technique to estimate of the arrival times of the first ions and protons at the location of the ASEC monitors.

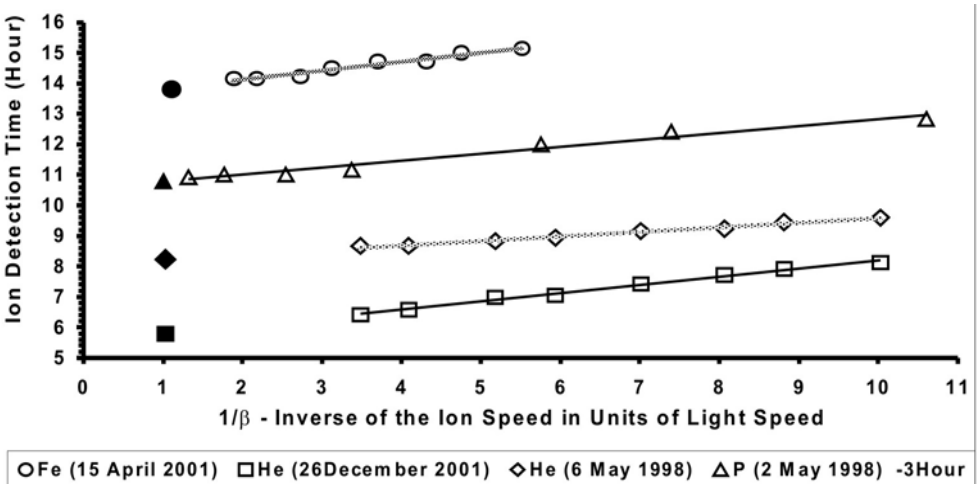


Figure 93. Linear correlation of ion arrival times to 1 AU versus their inverse speed. The filled symbols correspond to the expected arrival times of relativistic ions. The proton event is shifted down by 3 hours for better visualization of the regression lines.

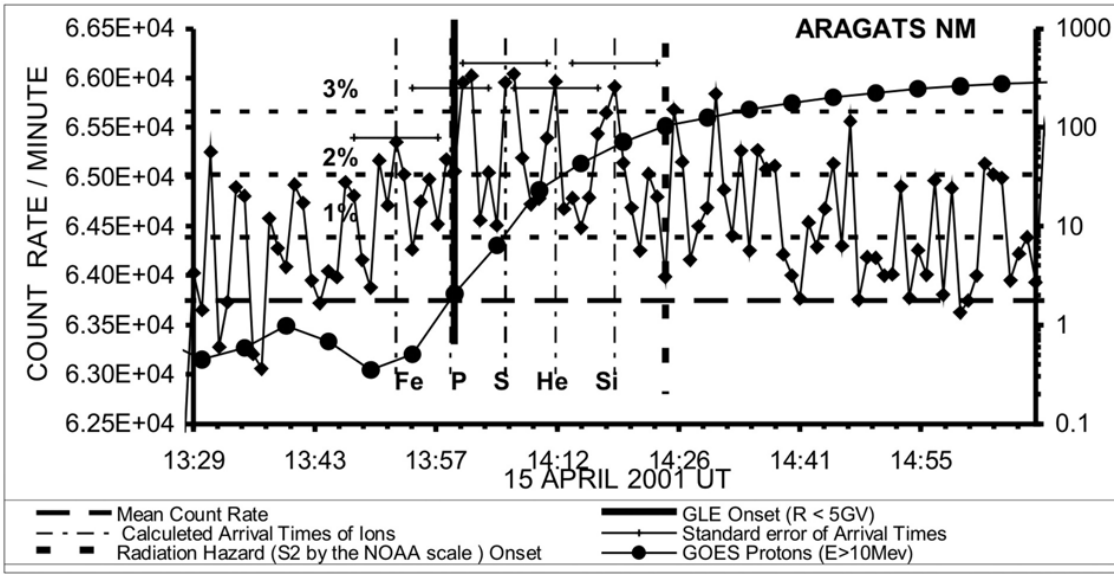


Figure 94. The 15 April 2001 GLE. Estimated arrival times of relativistic ions superimposed on the ANM 1 minute time series. Horizontal dashed lines indicate 1, 2 and 3% enhancements relative to mean count rate.

Figure 94 shows the April 15, 2001 GLE, registered by the ANM and expected arrival times of the various species of ions calculated as in Figure 93. The solid vertical line denotes the onset time of the GLE registered by the high latitude neutron monitors with rigidities R < 5 GeV. The vertical dashed line indicates the time of the S2 type radiation storm onset. The estimated expected arrival times of different species of ions (Fe, P, S, He, and Si) are denoted by the vertical dashed-dotted lines. Each expected arrival time is associated with an error bar indicated on the graph.

The overall pattern of the peaks seems to fit the expected arrival times of the various species rather well. For more firm inference on the possibility of "ion spectroscopy" with ground based monitors we also need to register fluxes of the muons and electrons in correlation with the neutrons, now possible and in process at the ASEC since the summer of 2002.

The characteristics of 4 GLE events are summarized in Table 51. Column 5 shows the calculated expected arrival time of the particular ion species at the location of the ASEC monitors. Column 6 shows the time of the first large peak in the Aragats NM time series indicating start of the GLE. Column 7 represents the significance of detected peaks in percents (relative enhancement) and in the number of standard deviations (relative standard deviation of the ANM, $\sigma \sim 0.7\%$). For all 4 events the significance of the peak is greater than 3σ, thus the probability that the peak is due to random fluctuation only is very small. The last column shows the time of onset of the S2 type radiation storm during each event represented in Figure 93.The S2 type radiation storm is characterized by count rates of >100counts/cm2/sec for >10 MeV protons. According to the NOAA Space Weather scales for Solar Radiation Storms (NOAA SW scales), the intensity of the CR flux during the S2 type radiation storm is high enough to cause considerable difficulties for satellite operation.

Table 51. Characteristics of the GLE detected by ACE/SIS, GOES-8 and ASEC detectors

1 Date and Order No. of GLE	*2 Flare Importance Category*	*3 First Ion Type*	*4 Correlation Coefficient of the Linear Regression*	*5 Calculated Arrival Time at 1AU in UT*	*6 GLE start at ASEC, in UT*	*7 Estimates of Signifi-cance%*	*8 I(Ep>10MeV >100/cm^2.s.srS2 Onset, UT**
02-05-1998 56	*X1.1*	*P*	*0.97*	*13:43 ± 8.2min*	*13:47*	*2.3 (3.2σ)*	*15:25*
06-05-1998 57	*X2.7*	*He*	*0.98*	*8:13 ± 4.6min*	*8:08*	*2.4 (3.4σ)*	*9:15*
15-04-2001 60	*X14.4*	*Fe*	*0.96*	*13:52 ± 6.5min*	*13:53*	*2.5 (3.6σ)*	*14:25*
26-12-2001 63	*M7.1*	*He*	*0.99*	*5:47 ± 5.4min*	*5:52*	*2.4 (3.4σ)*	*6:35*

* NOAA Space Weather Scale for Radiation Storms.

10.5 DISCUSSION

In all 4 cases of the GLE events described in the previous section, precursors of the S2 type radiation storms could be found in the one-minute time series of the ANM counts, which coincide, within the error bars, with the expected arrival times of the first ions accelerated at Sun. From Table 2 we can see that these "early" ions come about 32 to 67 minutes prior to the onset of the S2 type radiation storm. Of course, the actual forecasting will be much more difficult than our post-event analysis. Nevertheless, we want to emphasize the feasibility of Space Weather forecasting based on short-term enhancements of cosmic ray fluxes at mountain altitude. We also want to emphasize the possibility of enhancing the reliability of the prognosis significantly when using data from all of the ASEC monitors. We plan to use not only the information about abrupt enhancements in cosmic ray fluxes, but also information contained in the correlation matrix of the measurements of the neutron, muon and electron fluxes. The significance of the information from the different measured components of the particle fluxes and the estimation of the forecasting efficiency versus the frequency of false-alarms will be analyzed in detail after completing the computer simulations of the various aspects of the Aragats alert service. These include particle propagation in the atmosphere, detector response sensitivity, and alternative scenarios of alert triggering.

Here we want to discuss the overall possibilities of forecasting severe radiation storms by various experimental techniques. A variety of measurements of solar activity starting from radio-waves to gamma-rays can be used as precursors of severe radiation storms. Although numerous attempts of modeling the major SEP events have been conducted and this activity is getting more and more support (BU News, 2002), existing models are mostly phenomenological, because the basic knowledge about the consequence of energetic processes on the Sun is still not complete. Formal systems, such as Neural Network estimators, cannot provide reliable forecasting due to lack of training examples. Among the historically very severe SEPs, only the event from September 1989 was measured and documented well enough to be used for the neural network training.

Correlations between various characteristics of the solar energetic phenomena and spectra of accelerated ions and protons studied on smaller events in the absence of the analysis of data from severe events can be misleading when applying them to forecasting of the major event. Only the high enough fluxes of the highest energy particles measured by an ensemble of surface detectors in combination with data from space borne detectors can predict an upcoming severe radiation storm with any degree of reliability.

Below we present some advantages of using surface detectors at high altitudes and low latitudes in consort with the satellite-based sensors for the SEP forecasting

- Low latitude monitors with high cutoff rigidities have the advantage over high latitude monitors, because the enhanced signal-to-noise ratio minimizes the "false-alarm" rate increasing the possibility of early detection of low fluxes of highest energy ions. These ions can go undetected by the high latitude monitors, because their signal can be overwhelmed by the large background of galactic cosmic rays of lower energies at high latitude locations.
- The advantage of an alert service which utilizes data from the ground based detectors over services which use satellite based detectors only lies in the possibility of detecting ions of the highest energies. No instrument exists and nor have been envisioned for the flight that can measure high-energy ions in SEP events (Reames, 2000). Also it is worth to mention that high intensity radiation have the potential of blinding the very system that warns against the SEP events.
- ASEC monitors provide precise information on short-term variations of 3 species of cosmic ray flux (electrons, muons, and neutrons). Simultaneously detected abrupt changes of all 3 types of species count rates from the ASEC monitors will significantly improve reliability of forecasting.
- Directional information provided by the ASEC muon monitors will further enhance the signal- to-noise ratio and further improve the quality of forecasting.
- Along with count rates the correlation matrices of all ASEC measuring channels are calculated on-line as well, providing additional information on the nature and strength of the detected SEP.
- Measuring CR fluxes at 2 altitudes of 3200 m and 2000 m, gives additional information to estimate the spectra of SEP particles at highest energies and - the "knee" position of the spectra, most important for calculation of expected radiation doze (Reames, 2000).

The mentioned advantages of the SEP forecasting using surface detectors in addition to satellite based instruments, point to the necessity of improving the existing world-wide networks for measuring different species of cosmic ray fluxes along with the directional information. The detectors of the type presented in Figure 92 distributed along different Earth longitudes and latitudes, preferably at high altitudes, connected with space-born sensors via interplanetary Internet connections will highly improve existing warning services.

10.6 CONCLUSION

Our analysis, based on the SEPs unleashing S2 type radiation storms, demonstrates that the large area ground-based neutron monitors on Mt. Aragats in Armenia can detect the

early arriving highest energy particles released near the sun 30 to 70 minutes in advance of the arrival of the lower energy higher flux particles which can cause the damage to the satellites and humans in space. We have also designed detectors to measure directional information about the arriving radiation, which helps to enhance the signal to noise ratio. One of the very attractive features of the Aragats Space Environmental Center is its ideal geographical location - latitude and altitude - to maximize the signal to noise ratio. And finally, the variety of the types of monitors which allows us to measure all 3 species of cosmic radiation (electrons, muons, and neutrons) simultaneously is a very important tool to increase the reliability of accurate forecasting and decrease the number of false alarms. Information from the ASEC ground based detectors is an essential element to add to the information available from Satellites for engineering an accurate and reliable early warning forecast of the most severe irruptions on the sun directed towards earth.

The influence of solar radiation on humans and orbiting technological system has been summarized in the Public Documents of the ESA Space Weather Programme Studies as follows: *Energetic ions from SEP arriving to 1 AU can produce Single Event Effects (SEEs) in satellite electronics (single hard errors, single event upsets, latchups, burnouts, gate and dielectric ruptures). These effects are normally due to heavy ions, but particles as light as protons or neutrons can produce the same effects as heavy ions through nuclear reactions with silicon inside the electronics (in the future, due to increasing miniaturization, protons may be able to directly induce SEEs). These effects, as they are sporadic, are of major concern for space weather. Some of them are permanent, either directly (e.g., latchup, burnout) or indirectly (e.g., changes in permanent program memory).*

The radiation effects on human beings are similar to the effects on electronics. Dose effects affect all cells, especially those, which are not renewed or at least not rapidly renewed. Single energetic particles can also break the DNA chain in the cell nucleus, producing chromosome aberrations, translocations and tumor induction. They can induce also cell mutation that can have effects on the genetics. Energetic ions and electrons locally increase dark currents in detectors. This effect is clearly visible from imagers on board Earth or Sun observatories. Energetic ions and electrons also produce atom displacements in solar cells, decreasing the output power. Energetic electrons as they penetrate inside the spacecraft produce internal charging and electrostatic discharges. The effect of discharges can be direct destruction but normally they create electromagnetic pulses which produce signals interpreted as false commands by onboard computers (Horne, 2001), (Koskinen et al., 2001), (Koskinen, 2001A).

In 2001 Tylka stated the following: "SEP events are not predictable in any meaningful sense. We cannot give a reliable prediction of when such event will occur, nor can say, once an event has started, what its characteristics will be, even a few hours in advance" (Tylka et al., 2001B).

Nevertheless, we can optimistically look to the future. More and more of the measurements from world-wide network of ground-based neutron monitors are posted on the Internet in real time. Soft X-ray, hard X-ray, and gamma radiation detections by space borne detectors also are posted on the Internet in real-time. There are plans to put the real time Sun images from EIT and LASCO Instruments of the SOHO satellite on the Internet (Berghmans, 2001). The real time information about the CME, its magnitude, and the direction in which it is headed from LASCO and EIT satellites will provide additional

valuable information for forecasting of the danger from the energetic solar events. The joint multidimensional multidetector analysis of all relevant information from space- borne and ground-based detectors, including the full complement of large ground based detectors at ASEC, will minimize the number of false alarms and will maximize the reliability and timeliness of forecasting the arrival of the dangerous SEP.

ANNEX A. MULTISTART RANDOM SEARCH WITH EARLY STOP AS TOOL FOR SELECTION OF SETS OF DIFFERENTLY EXPRESSED GENES

Introduction

The abundance of biologically active molecules (genes) provide basis of cell behavior (operation). By regulating amount of genes cell activates it's functions. Therefore, there exist direct correlation between gene abundance and biological function of cell from particular tissue.

The spotted complementary DNA Microarray technique developed at Stanford university (Schena et al., 1995) provides opportunity to quantify gene abundance (or relative abundance) in cells under investigation. Usually these cells are taken from alternative tissues (cancer classification, time series after a given treatment). In such way researchers can screen for potential tumor predictors (markers) or drug targets (results of drug intervention).

Unfortunately both cell variability (intrinsic genomic instability of tumors) and measurement instability (bad reproducibility) can obscure actual difference between normal and pathological cells (or introduce a fake ones). Very big amount of genes (the dimensionality of the "feature space" is reaching tenths of thousands), relatively small number of measurements (usually very few dozens), complicated nature of biochemical pathways and numerous sources of experimental errors prevent rigorous diagnostic inference from microarray data and found correlations between genes and several diseases (usually cancers) still remain ambitious.

What we need is an integrated system of computation methods dealing with big amounts of multidimensional data with high level of statistical variability. Such system with user-friendly interface will help researchers to reveal hidden dependencies and lead gene analysis in a way to discover gene complexes that might be of use for diagnostic purposes or/and as therapeutic targets.

The core elements of such system are best feature selection methods and classification algorithms.

The selection of the feature subset (clustering) representing different concepts increase classifier efficiency and reveal nontrivial correlations of measurements with alternative classes (tissues), in this way leading path to discovery and prediction for different types of cancer, independent of previous biological knowledge.

The problem of cluster selection is well suited to formulation as an optimization problem. Given a set of two alternative classes of N-dimensional input patterns, the task of the EA is to find a subset of features in N1 (N1<<N) lower dimension space that maximizes optimization criteria.

The selected gene clusters are evaluated based upon maximal class separation or the classification accuracy as an evaluation criterion.

Unfortunately, two obstacles prevent straightforward implication of well-established

statistical techniques for dimensionality reduction. The first one is the rather large amount of genes participating in particular pathway (up to 50) and, the second one is the scarceness of "training samples", examples of measurements related to tissues with known diagnose.

The number of possible gene combinations is enormously large; therefore it is impossible to compare all gene subsets. On the other hand the size of training samples are highly unbalanced with dimensionality and presence of outliers makes estimates of covariance matrix unstable. Consequently all statistical methods based on covariance matrix, and its eigenvalues became unreliable.

To avoid mentioned above difficulties we develop the fast gene selection method operated in the feature spaces of not very large dimensionality (in which the covariance matrix estimation still make sense). We name this method "evolutionary bootstrap" in analogy with famous Efron's bootstrap method, or Multiple Random Search with Early Stop – MRSES method.

Both probabilistic distances (Mahalonobis and Bhattacharya distances), and misclassification probability can be used as optimization criteria.

The Bayes Risk (probability of misclassification) to our experience is the best quality function for multidimensional classification problems. But, very limited number of samples makes even the best estimate of Bayes risk – the "one-leave-out-for a-time" estimate discreet and usually after few iterations the misclassification rate didn't changed any more. Therefore we plan to use measures, exploiting the differences in estimates of covariance matrix of alternative classes. Among these estimates are Mahalanobis and Bhattachariy probabilistic distances.

The explanation of obtained gene cluster (biochemical pathway) remains explicit, and easy to identify and analyze and lead to a deeper understanding of the data. This capability is of primary importance in microarray data analysis, when the goal is – to isolate group of genes responsible for disease.

A key advantage proposed feature extraction technique is that it can produce gene subsets of rather large dimensionality operated with data of significantly lower dimension. In such way we overcome both obstacles of "course of dimensionality" and instability of covariance matrix and probability density estimates.

1. Description of the Multistart Random Search with Early Stop (MRSES) Agorithm

Random search algorithms are well suited for finding optima in complex combinatorial spaces **(Zhigljavsky, 1991)**. Here we propose a modification of the basic random search algorithm, the idea of which is to apply a local search procedure multiple times and then use all the selected genes for further analysis. As overfitting is an ever-present danger, we concentrate on short local searches to be able to examine *local maximum regions* and to prevent convergence to a unique global maximum. The method can be applied for a wide variety of parametric and non-parametric quality functions; our choice will be discussed in the next section. The basic structure of the multi-start random search method with early stopping (MRSES) is outlined by the following sequence of steps:

Step 0 Randomly select N_{cluat} genes from N_{all}, $N_{cluat} \ll N_{all}$. Evaluate the quality function.

Step 1 Generate a new evaluation point by swapping one or more randomly selected genes between the currently selected set and the rest of the genes.

Step 2 Evaluate the quality function for the new combination; if its value has decreased, then return to the previous combination.

Step 3 Repeat Steps 1 and 2 until the number of iterations reaches $N_{iter,}$ then save the best configuration.

Step 4 Repeat Steps 0—3 until the number of these cycles reaches N_{cycle}.

Step 5 Post-process the resultant N_{cycle} groups of N_{clust} genes each in N_{cycle} manner described below.

The value of the cluster size parameter N_{clust} is limited by the number of available training samples (microarray slides). The nature of this limitation depends on the particular choice of the quality function, but generally both parametric and non-parametric criteria are very sensitive to the scarceness of training samples in a high-dimensional feature space. In our colon cancer cell line application only 12 samples per tissue are available so in this study we used $N_{clust} = 5$, the largest value that did not cause the quality function estimates to become unstable (specifically, the critical element was the determinant of the covariance matrix). It is important to note that due to the multiple repetition of the search and the post-processing described below, the choice of N_{clust} does not limit the detectable number of differentially expressed genes, but rather the depth of the estimated interaction structure.

The number of iterations N_{iter} is crucial to control overfitting; its value should balance between being too small to provide enough time for finding truly differentially expressed genes and being too large. In the latter case one can expect the same maximum attained in many cycles of search because of overfitting. We investigate the role of the choice of in the search for differentially expressed genes in the simulation study below.

The variability of the algorithm start points for detecting differentially expressed gene clusters depends on the number of search cycles N_{cycle}. Therefore it should be as large as possible and is limited by CPU power only (we have used $N_{cycle} = 10,000$).

In the post-processing step the local optima are combined to provide a final, global solution. We proceed from the heuristic argument that strongly differentially expressed genes shouldappear in many of the local maxima, so that each gene is characterized by the frequency of its occurrence in the suboptimal set selected by each of the N_{cycle} cycles. A conservative estimate of the p-value corresponding to the observed frequency can be calculated. If a gene is not differentially expressed, then the probability that it will be in the selected subset by chance is expected to be equal to N_{clust} / N_{all} (and most likely smaller). As the number of repetitions N_{cycle}
is large, the final selection frequency of this gene is well approximated by a Poisson distribution with mean $N_{cycle}N_{iter} / N_{all.}$ Based on this null-distribution the corresponding (conservative) p-values for each gene can be calculated.

2. Optimization Criteria

Most popular parametric measures of the discrepancy between 2 multidimensional samples are based on sampling estimate of the 2 samples centroids, covariance matrices and it's determinants. The Mahalonobis distance R_M (Mahalonobis, 1936) and Correlation distance R_{corr} were used for the feature selection for imaging Cherenkov telescopes data analysis (Aharonian et al., 1991)

$$R_M^2 = (\bar{v} - \bar{u})^T \sum_{(v+u)/2}^{-1} (\bar{v} - \bar{u})$$
$$R_{corr} = \ln \frac{0.5|\sum_v + \sum_u|}{|\sum_v^{0.5} \sum_u^{0.5}|} \qquad (10.1)$$

Their combination so-called Bhatacharia distance $R_{Bha} = 1/8R_M + 1\ 2R_{corr}$ is equal to zero if the 2 samples completely overlap and – to infinity if they do not overlap at all. Through the Bhatacharia distance one can express the upper bound of the expected misclassification rate:

$$U_B = 1 - exp(-2\ R_{Bha})$$

The most natural nonparametric measure for measuring discrepancy between two multidimensional samples is the classification error probability, which depends on the degree of overlapping of alternative multivariate distributions. Of course we have to use for classification optimal Bayes decision rule.

The Nonparametric Bayesian decision rule has the form (Chilingarian, 1989)

$$\tilde{A} = \eta(v, A, \tilde{P}) = \text{argmax}_i \{C_i \hat{p}(A_i/v)\}, i = 1., \ldots, L, \qquad (10.2)$$

where c_i are the losses connected with $\tilde{A}$ decision, $\hat{p}(A_i/v)$ is the nonparametric estimate of the *a posteriori* density, connected with conditional ones by the Bayes theorem:

$$R_{Div} = \int \{p(v/w1) - p(v/w2)\} \ln \frac{p(v/w1)}{p(v/w2)} dv \qquad (10.3)$$

Finally, substituting the posteriori densities by the conditional ones we get the Bayesian decision rule in the form

$$R_{Ent} = -\int p(v/w_i) \ln p(v/w_i) dv \qquad (10.4)$$

To estimate conditional densities, we use Parzen kernel (Devroye et al., 1985), (Parzen, 1962) and K Nearest Neighbors (KNN) methods (Lofsgaarden et al, 1966), (Schena et al., 1995) with automatic adaptation of the method parameter (kernel width - for Parzen estimate, and number of neighbors - for KNN estimate) (Chilingarian et al., 1984). Several probability density estimates corresponding to different values of parameters are calculated simultaneously. Then the obtained sequence is ordered and the median of this sequence is chosen as a final estimate.

Depending on the intrinsic probability density in the vicinity of point v, where the density is estimated, due to stabilizing properties of the median, each time the best method parameter will be chosen (Efron, 1981).

The Parzen kernel probability density is estimated by:

(2π)d /2 hd

$$\hat{p}(v/A_i) = \frac{|\sum_i|^{-0.5}}{(2\pi)^{d/2} h^d} \sum_{j=1}^{M_i} e^{-r_j^2/2h^2\omega_j}, i = 1, \ldots, L, \sum_{j=1}^{M_i} \omega_j = 1 \qquad (10.5)$$

where d is the feature space dimensionality, M_i is the number of events in the *i-th* TS, r_j is the distance from experimental event v to the *j-th* event of the TS in the Mahalanobis metric

$$r_j^2 = (v - u_j)^T \sum_i^{-1} (v - u_j), \qquad (10.6)$$

where Σ_i is the sampling covariance matrix of the class to which $\boldsymbol{u}_j$ belongs, ω_j are the event weights, h is the kernel width (parameter controlling the degree of the "smoothness" of an estimate).

The K nearest neighbors estimate takes the form:

$$\hat{p}(v/A_i) = \frac{k-1}{M_i V_k(v)}, \qquad (10.7)$$

where V_k (**v**) is the volume of a d-dimensional hypersphere containing the k –1 nearest neighbors to the experimental event v,

$$V_k(v) = V_d |\sum_i|^{1/2} r_k^d, V_d = \frac{\pi^{d/2}}{\Gamma(d/2+1)}, \qquad (10.8)$$

where r_k is the distance to the k-th nearest neighbor of v, Γ(.) is the gamma function, $|\Sigma_i|$ is the determinant of the covariance matrix of the class to which the k-th neighbor belongs.

The Bayes risk takes the form:
where

$$\hat{R}^B = E\{\theta[\eta(v, A, P)]\} = \int \theta p(v)\, dv, \qquad (10.9)$$

0, for correct classification, 1, otherwise

$$\theta[\eta(v, A, P)] = \begin{cases} 0, \text{for correct classification} \\ 1, \text{other wise} \end{cases} \qquad (10.10)$$

Where the mathematical expectation is taken over the whole d-dimensional feature space V. In other words the Bayes error is a measure of the overlapping of alternative distributions in the feature space V, e.g. the expected proportion of the "incorrect" classification. Since we do not know to which class experimental vectors belong, we obtain an estimate of R^B via the TS:

$$\hat{R}^B = E\left\{\frac{1}{M_{TS}} \sum_{i=1}^{M_{TS}} \theta\left[\eta(u_i, A, \tilde{P}_{(i)})\right]\right\} \qquad (10.11)$$

i.e. we classify the $u_i, i = 1, M_{TS}$ and check the correctness of the classification over the index of the class $t_j, j = 1, L$. The expectation is taken over all possible samples of volume M_{TS}.

However, as numerous investigations have shown (Szabo et al., 2002), this estimate is systematically biased and hence, a one-leave-out-for-a-time estimate is preferable:

$$\hat{R}^e = \frac{1}{M_{TS}} \sum_{i=1}^{M_{TS}} \theta\{\eta(u_i, A, \tilde{P}_{(i)})\} \qquad (10.12)$$

where $A, \tilde{P}_{(i)}$ is a TS with a removed i_{th} element, which is classified and then "returned" to the sample. This estimate is unbiased and has an essentially smaller m.s. deviation compared with other estimators (Snappin et al., 1984). The advantage of R^e the feature space has a high dimensionality.
is especially notable when Bayes error, due to it's discrete character, couldn't be used effectively for small sample size case (usual for microarray applications). Therefore we use more complicated continuous measures, as Divergence (Rdiv) and Entropy (REnt) , also based on the nonparametric estimates of the multivariate probability density function (Chilingaryan et al., 2002)

$$R_{Div} = \int \{p(v/w1) - p(v/w2)\} \ln \frac{p(v/w1)}{p(v/w2)} dv \qquad (10.13)$$

$$R_{Ent} = -\int p(v/w_i) \ln p(v/w_i) dv \qquad (10.14)$$

3. Tuning of the MRSES on computer simulations

The main goal of the simulation study is to investigate under controlled conditions whether the multivariate search method really increases the power of detection of differentially expressed genes. We also want to assess the effect of changing the parameters of the proposed MRSES algorithm, with special attention to N_{iter}.

For testing of the new method first we simulate cDNA microarray data where the extent of differential expression of each gene is known. The simulation program CORRSIM developed in Utah univ. Cancer Institute (Szabo et al., 2002), allows to over or under express several selected genes and take into account both biological variability and instrumental errors.

In our studies we select particular number of genes (Kalmykov.et al., 1997), randomly assign to them expression level in coherent way (simulation of biochemical pathway).

Then we "forget" about superimposed structure and data enter MRSES program for multivariate analysis. Of course we perform also standard data preprocessing (*normal, log-normal or rank-based transformations*) and one-dimensional 2-samle statistical tests (*Student, Kolmogorov-Smirnov and Mann-Whitney*).

The MRSES algorithm was investigated and compared with traditional methods. In this study we were interested in understanding of the ways of integration of independent local random searches in global selection of the "interesting" gene combinations. Applying different stopping rules to random search algorithm (changing number of iterations N_{iter}), we force algorithm to converge to different local maximum (selecting different correlated genes) or finally reach one global maximum.

We simulated data sets and applied the search algorithm using the parameters $N_{clust}=5$, $N_{cycle}=10,000$ and several values of N_{iter}. In Figure 95 we compare some characteristics of the algorithm; the left side corresponds to $N_{iter}=1000$ and the right side to $N_{iter}=100,000$.

The top graphs display the histogram of the values of the «last good iteration» - the number of iterations after which no successful steps was encountered (that is when the final set was found). It is clear that 1000 iterations are not sufficient to reach the global maximum, on the other hand 100,000 iterations are more than enough for the random search to

converge. Themiddle pictures illustrate this in another way. In the case of early stopping the distribution of the Mahalanobis distances (the algorithm's quality function) corresponding to the suboptimal sets is unimodal with high variability. Therefore, we can conclude that the algorithm explored many different local maxima with a variety of corresponding values of the quality function. When the iterations are continued to reach 100,000, the distribution of the Mahalanobis distances achieved in the suboptimal sets becomes very discrete. In half ofthe cases the search reaches the global maximum on a unique combination of genes, one ofwhich was not preset as differentially expressed. So we do find the global maximum, but miss many local maxima and the corresponding differentially expressed genes from the predefined cluster. When we stop early at the 1000-th iteration none of the 10,000 cycles finds the global maximum, but a variety of genes are selected. This is demonstrated in the bottom pictures, where the frequencies of selection for the 20 genes belonging to the differentially expressed cluster are plotted. In our experiment, 17 from 20 genes pass the selection criteria (to have the frequency of occurrence higher than 0.5%) with the early stop, while only 10 genes are correctly selected when we force algorithm to attain the global maximum.

To further investigate the dependence of the power of the algorithm on the stopping time, we estimated the ROC curves corresponding to values of N_{iter} ranging from 100 to 10,000 based on 10 independently simulated data sets (other parameters were held constant: $N_{clust} = 5$, $N_{cycle} = 10000$). For each search we obtained a list of genes with associated frequencies of occurrence in the selected subset and compiled a final selection of differentially expressed genes by applying cutoff values ranging from 0.1% to 10%. Based on the null hypotheses of no differential expression, for each of these sets we estimated the type I error as the proportion among all non-differentially expressed genes of those genes that wereincluded in the selection and the type II error as the proportion among the genes in the differentially expressed cluster that were not included in the final selection. The resulting ROC curves are shown in Figure 96. As a reference a point representing the type I error and the power of the marginal t-test with 5% significance level is also plotted. From Figure 96 we can see that the choice of does indeed have a large influence on the performance of the algorithm and long searches are inferior to early stopping. Of course there is a limit on how early one should stop, because very short searches do not have a chance to reach any local maxima. Under our setup $N_{iter} = 500$ gives the best performance.

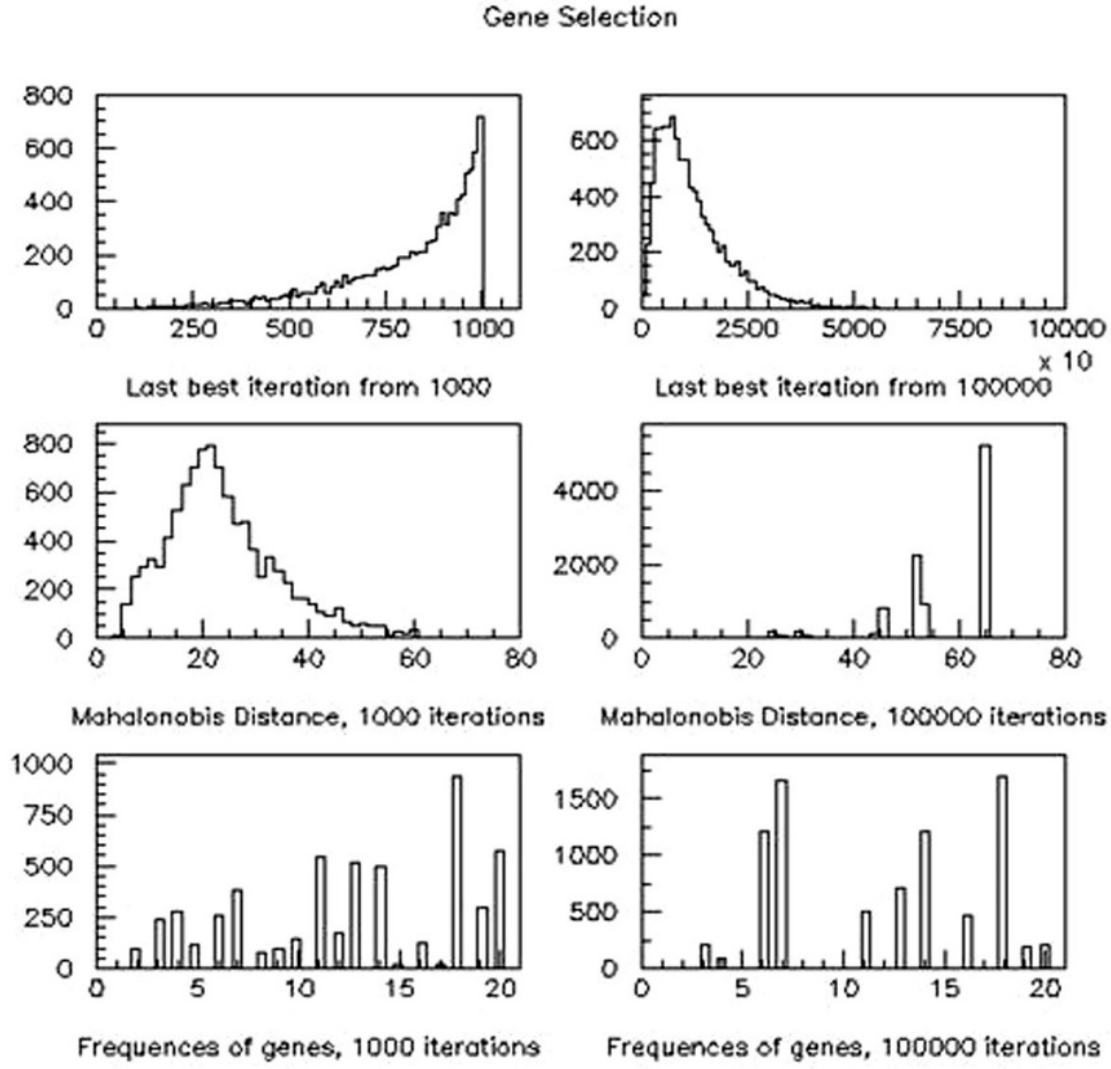


Figure 95. Comparison of gene selection with early and late stopping of the multi-start random search algorithm. First row: histogram of the value of the 'last best iteration' in the $N_{cycle}i_e$ searches. Second row: histogram of the estimated Mahalanobis distances for the $N_{cycle}i_e$ selected sets. Third row: frequency of inclusion of the differentially expressed genes (1-20) in one of the selected sets. One simulated data set.

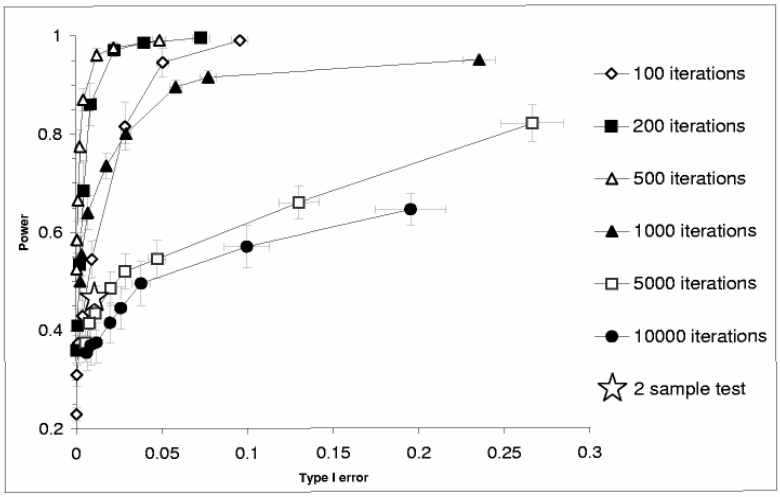


Figure 96. ROC curves for various values of N_{ner} controlling the stopping time. Based on 10 simulated data sets; the error bars show the standard errors of the point estimates.

4. Application for medical diagnostics: two colon cancer cell lines

We selected two commonly studied colon cancer cell lines for our analyses. HT29 cells represent advanced, highly aggressive colon tumors. They contain mutations in both the APC gene and p53 gene, two tumor suppressor genes that frequently mutate during colon tumorigenesis. As another cell type, we selected HCT116 cells. This cell line models less aggressive colon tumors and harbors functional p53 and APC. However, they show a deficiency of those genetic systems that are responsible for the repair of mismatched regions of DNA. To generate the data, three samples of each mRNA (1 /xg each) were labeled by production of first-strand cDNA in the presence of Cy3-dCTP (green) or Cy5-dCTP (red). Cy-3 was used to label HCT116 cells while Cy-5 was used for HT29 cells. Each comparison set was hybridized against two microarray slides (facing each other) containing 4608 minimally redundant cDNA's spotted in duplicate. As a quality control tool, six *Drosophila* genes are always added to the Cy-5 sample, so in a red vs. green comparison they have to be differentially expressed by design. This experiment resulted in a total of twelve measurements on each channels for each gene on the microarrays. While there is a nested dependence structure of the samples, we used them as independent replicates. From a separate experiment we also have data with ten HCT116 samples hybridized with Cy-5, that is on the red channel. The normal score adjustment (3.3) was used.

/ We have performed two comparisons: HCT116 vs. HT29 and HCT116 (green) vs. HCT116 (red), the latter after the exclusion of the six *Drosophila* control genes. Based on our simulation studies and the available sample size we chose 97 we compare the two search procedures. The left column corresponds to the comparison of the different cell lines, while the right column to the comparison of the same cell line on different channels. The histogram of the last best iteration is very similar for both cases and looks like the one obtained in our simulation studies where the global maximum was not obtained. Thus in both cases due to the early stopping the algorithm keeps exploring local maxima. However the distribution of the estimated Mahalanobis distance at these local maxima is very different: when different cell lines were compared the Mahalanobis distance $R4_{ah}$ based on the locally optimal subsets tended to be much larger, that is the separation of the two tissues was considerably better.

We also compared the two lists of genes ordered by decreasing frequency of occurrence inthe selected subset. The histogram of the first 115 genes on the list is shown in Figure 98 (white and black columns); the right tails of the histograms are very close to each other. As an additional reference we included the histogram generated by the HCT116 (green) vs.

HCT116 (red) comparison with the control genes included (grey columns). It is clear that some of the genes in the HCT116/HT29 comparison (black) are selected more often than expected under the null hypothesis of no difference between the two tissues (white). In the no-difference comparison (without control genes) there are just two genes that are selected in more than 3% of the cycles; if the control genes are included, this number increases to six and 4 out of the top 5 genes (# 1,2,3 and 5) are actually the *Drosophila* controls.

Based on the histograms we decided to use 1% as the cutoff for selecting differentially expressed genes, thus selecting 59 cDNA spots. We have compared this list with the top 59

genes selected by values of the t-statistic. Almost half of these genes (25 exactly) appear in both lists. However a striking feature of the selection using multivariate random search is its ability to include correlated genes. Some of the genes have several spots that correspond to them, thus their expression level at various spots is known to be correlated. Among the 59 genes selected by the multivariate method 13 have two, and 2 have three spots related to them. Four of these genes have all their replicates in the selected set including one of the genes with three replicates. By comparison, in the list based on the marginal t-statistic 17 genes have two or more replicates on the slide, and only one of them has all of its replicates selected. While one would hope to be able to find all the replicates of a gene, in practice localized errors can make it infeasible. Part of the problem is the unreliability of the gene identifications in our dataset, so some cDNA's labeled the same might actually correspond to different genes. Despite the caveat, this finding is encouraging and is in line with what we have seen in the simulation studies: genes with less pronounced differential expression can be selected if they are correlated with more strongly differentially expressed genes.

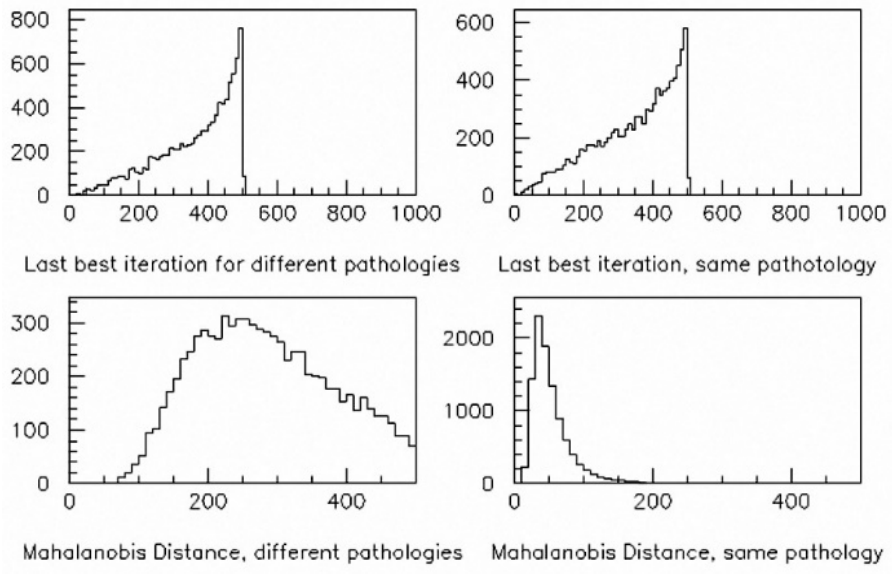


Figure 97. Comparison of the search procedure used for same and different tissues. First row: histogram of the value of the 'last best iteration' in the $N_{cycle}i_e$ searches. Second row: Histogram of the estimated Mahalanobis distances for the $N_{cycle}i_e$ suboptimal sets. HCT - HT data.

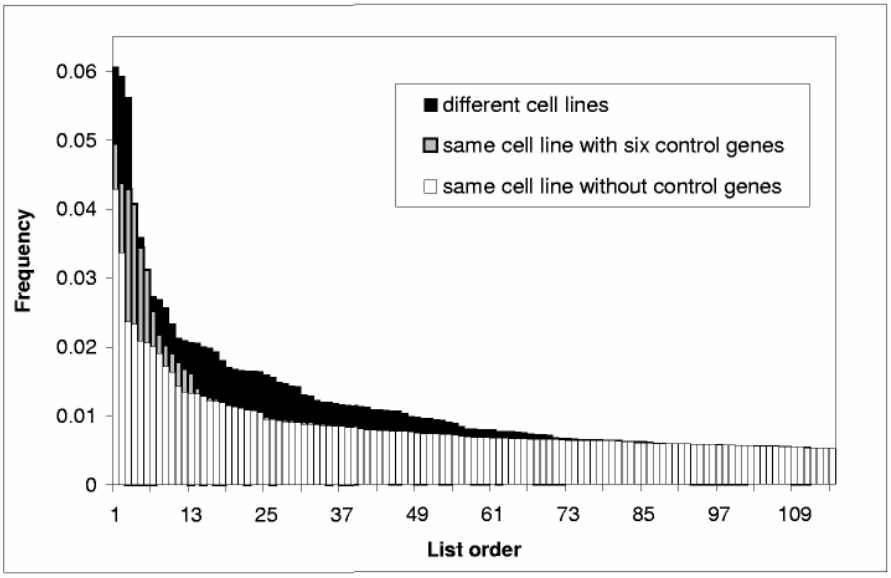


Figure 98. Comparison of the frequency of inclusion in the selected locally optimal set in the search procedure used for same tissue with (gray) or without (white) the six control Drosophila genes on the two channels and for different tissues (black). HCT - HT data.

ANNEX B. NOTATIONS

$N(\mu, \sigma)$	*Gaussian Population*
V	*Event (Measurement, Feature) Space*
A	*Basic States Space*
Λ	*Space of all FFNN Model*
G	*Space of possible FFNN Architectures*
W	*Space of FFNN weigts (Couplings, Connections)*
O_λ	*Trained FFNN Model, Net Output*
d	*Dimensionality*
L	*Number of classes*
M	*Number of experimental events*
v_i	*Vector of measured variables*
u_i	*Vector of simulated events*
(A, P)	*Stochastic mechanism which generates experimental data*
$(A, P-)$	*"Controlled" stochastic mechanism, simulation program*
PA	*Prior measure*
C -	*Losses (cost) measure*
$\hat{p}(v / Ai)$	*Conditional probability density function*
$p\hat{}(v / Ai)$	*Estimate of conditional density*
$p\hat{}(Ai / v)$	*Posterior density*

BIBLIOGRAPHY

1. Adams J., Bashindzhagyan G., et al.(2001), Advances in Space Research, 27, No. 4, pp. 829-833 (Adams et al., 2001)
2. Aglietta M. et al., UHE Cosmic Ray Event Reconstruction by the Electromagnetic Detector of EAS-TOP, NIM, A336, 310, 1993 (Aglietta et al., 1993).
3. Aglietta M. et al., The EAS Size Spectrum and the Cosmic Ray Energy Spectrum in the Region 1015 - 1015 eV., Astroparticle Phys., 10, 1, 1999 (Aglietta et al. 1999).
4. Aharonian F. A., Chilingaryan A. A., Plyasheshnikov A. K., Konopelko A. K., On the possibility of an improvement of background hadronic showers discrimination against gamma-ray coming from a discrete source by a multidimensional Cherenkov light analysis, in Proc. 21 ICRC, vol. 4, p. 246, Adelaide, 1990. (Aharonian et al. 1990).
5. Aharonian F.A., Chilingarian A.A., et. al., A multidimensional analysis of the Cherenkov images of air showers induced by very high energy y-rays and protons, NIM A302, 522, 1991; YerPhI preprint 1171 (48)-79, 1989. (Aharonian et al., 1991)
6. Akaki H., Statistical Predictor Identification, Ann.Inst.Statist.Math., 22, 203, 1970 (Akaki, 1970).
7. Allesandro B.—EAS-TOP Collaboration et al 2001 Proc. 27th Int. Cosmic Ray Conf. (Hamburg) vol 1, p 127 (Allesandro, 2001)
8. ANI Collaboration, The ANI experiment on the investigation of the interactions of hadrons and nuclei in energy range 1000 - 100000 TeV, NIM, A 323, 104, 1992 (ANI Collaboration, 1992).
9. Antoni T. et al., Test of high energy interaction models using the hadronic core of EAS, J. Phys. G, 25, 2161, 1999 (Antoni et al., 1999).
10. Antoni T., et al., (KASCADE Collaboration), (2003a), Astroparticle Physics 19, 715–728 (Antoni et al., 2003)
11. Antoni T., et al., (KASCADE Collaboration), (2003b), submitted to NIMA (Antoni et al., 2003A).
12. Arqueros F et al 2000 Astron. Astrophys. 359 682 (Arqueros, et al., 2000)
13. Ashwanden M.J., Wills, M.J., et al., (1996), ApJ 468, 398 (Ashwanden et al., 1996).
14. Ashwanden M.J. (2002), Space Science Reviews, 101, N 1,2 (Ashwanden, 2002).
15. Astone, P., & D'Agostini, G. 1999, preprint (hep-ex/9909047 (Astone et al., 1999)
16. Avakian V.V., Bazarov E.B., et. al., (1986), VANT, ser. Tech. Phys. Exp., 5(31), p.1. (Avakian et al., 1986)
17. Avakyan V. V. et al., The EAS Integral Spectra on mountain altitudes in the size interval 5 x 104 - 5 x 107, Yad. Fiz, 56, 183, 1993 [in Russian] (Avakyan et al., 1993).
18. Babayan V.Kh., Bostanjyan N., Chilingarian A.A., et al., Proc 27th ICRC, 9, 3541, Hamburg, 2001 (Babayan et al., 2001).
19. H.M.Badran and T.C.Weekes, Improvement of gamma-hadron discrimination at TeV energies using a new parameter, image Surface Brightness, 1997, Astroparticle Physics 7, 307-314 (Badran et al., 1997).

20. Barbosa D.D., (1979), AJ, 425,383 (Barbosa, 1979).
21. Barron A., Predicted Squared Error: a Criterion for Automatic Model Selection, in Self-Organizing Methods in Modeling, Marsel Dekker, New York, 1984 (Barron, 1984).
22. Bednarek W., Protheroe R.J. (2002), Astroparticle Physics 16, 397 (Bednarek et al., 2002)
23. Berezhko E.G., Ksenofontov L.T., Volk H.J., (2003), A&A, 412, 11 (Berezhko et al., 2003)
24. Berezhko E. G., Spectrum and chemical composition of CR's accelerated in SNR's, JETF, 82, 1, 1996 (Berezhko,1996).
25. Berger J. O., The Robust Bayesian Viewpoint, in Robustness of Bayesian analyses, Elsevier Science Publishers, 1984 (Berger, 1984).
26. Berghmans, D., ESA Space Science News, 2 December 2001, 8 (Berghmans, 2001).
27. Bhadra A. (2003) Proc. 28th Int. Cosmic Ray Conf. (Tsukuba), 303 (Bhadra, 2003)
28. Bierman P. L., The cosmic ray spectrum between 10 GeV and 3 EeV, A&A, 271, 649, 1999 (Bierman, 1999).
29. Bishop C. M.: Neural Networks for Pattern Recognition. Oxford Univ. Press, New-York, 1995 (Bishop, 1995)
30. Blokhin S. V., Romakhin V. A., Hovsepyan G. G., Calibration of Scintillation Detectors for the Aragats EAS Installation, in Proc. of the Workshop ANI 99, vol. FZK Internal report 6472, p. Ill, Nor-Amberd, 1999 (Blokhin at al., 1999).
31. Borog V.V., Dronov V.V., et al., Proceedings of Russian Academy of Science, Physics series, 65, 381, 2001 (Borog et al., 2001).
32. Boston Univ., News release, Sept. 19, 2002 (BU News, 2002).
33. Box G. E. P., The importance of practice in the development of statistics, Technometrics, 1984 (Box, 1984).
34. Brisken, W. F., Thorsett, S. E., Golden, A., & Goss, W. M. 2003, ApJ, 593, L89 (Brisken et al., 2003)
35. Bryant D.A., Cline T.C., Desai U.D., and McDonald F.B., (1963), Phys. Rev. Lett., 11, 144 (Bryant et al., 1963)
36. Cane H.V., Highlight talk, 27 ICRC, Hamburg, 2000, edited by R.Schlickeiser, p.311 (Cane, 2000).
37. Carmichael H., Cosmic Rays, IQSY Instruction manual, London, 1964 (Carmichael, 1964).
38. Cawley M.F. et al., Exp.Astonomy, 1, 173, 1990 (Cawley et al, 1990).
39. Cawley M.F., Weeks T.C., Instrumentation for Very High Energy Gamma-Ray Astronomy, Exp. Astron. 6, 3, 1995 (Cawley et al, 1995).
40. CERN, GEANT: CERN Program Library Long Writeups, w 5013 edn., 1993 (CERN, 1993).
41. Chilingarian A. A., in Proc. 18 ICRC, 5, 524, 1983 (Chilingarian, 1983).

42. Chilingarian A. A. and Galfayan S. Kh., Calculation of Bayes risk by KNN method, Stat. Problems of Control, 66, 66, 1984, Vilnius (Chilingarian et al., 1984).
43. Chilingarian A. A., Galfayan S. Kh., Zazian M. Z. and Dunaevsky A. M., In Proc. 19 ICRC, 6, 392, 1985 (Chilingarian et al., 1985).
44. Chilingarian A. A., Galfayan S.Kh., Dunaevski A.M., Zazyan M.Z. et al., The upper boundary of of iron nuclei fraction in PCR obtained from PAMIR data, in Proc. 20 ICRC, vol. 1, p. 386, Moscow, 1987 (Chilingarian et al., 1987).
45. Chilingarian A. A., Galfayan S. Kh., Zazyan M. Z., Denisova V. G., Dunaevsky A. M., Kunova S. A., and Slavatmsky. S.A., in Proc. 20 ICRC, 5, 312, 1987 (Chilingarian et al., 1987A)
46. Chilingarian A. A., Statistical decisions under nonparametric a priory information, Computer Physics Comunication, 54, 381-390, 1989 (Chilingarian, 1989).
47. Chilingarian, A.A. and M.F. Cawley, Multivariate analysis of Crab Nebula data, Note to Whipple collaboration from July 5, 1990 (Chilingarian et al., 1990).
48. Chilingarian A. A., Neural Net Classification of the gamma and proton images registered with atmospheric Cherenkov technique, random search learning in feed-forward networks, in Proc. 22 ICRC, vol. 1, p. 540, Dublin, 1991 (Chilingarian, 1991).
49. Chilingarian, A.A., Classification of the gamma and proton images with aid of mathematical models of Neural Networks, Proc. 22ICRC, 1, 483, Dublin, 1991 (Chilingarian 1991A).
50. Chilingarian A. A., Zazyan H. Z., Experiments with particle bundles in cosmic ray physics. Determination of strong interaction parameters by a pattern recognition method, J. of Nuclear Phys., 54, 128, 1991, [in Russian] (Chilingarian et al., 1991).
51. Chilingarian A. A., Zazyan H. Z., On the possibility of investigation of the mass composition and energy spectra of PCR in the energy range 1015 — 1017 eV using EAS data, IL Nuovo Cimento, 14C(6), 555, 1991 (Chilingarian et al., 1991A).
52. Chilingarian, A.A. and M.F. Cawley. Application of multivariate analysis to atmospheric Cherenkov imaging data from the Crab nebula. Proc. 22 ICRC, 1, 460-463, Dublin, 1991 (Chilingarian et al., 1991B).
53. Chilingarian A. A., On the methods of the enhancement of the reliability of the signal detection with Cherenkov Atmospheric techniques, in Izv. AN USSR, Phys., vol. 57, p. 186, 1993, [in Russian] (Chilingarian, 1993).
54. Chilingaryan A. A., Neural classification technique for background rejection in high energy physics experiments, Neorocomputing, 6, 497, 1994 (Chilingarian, 1994).
55. Chilingarian A. A., Cawley M. F., Optimizing the non-linear gamma-ray domain in VHE gamma-ray astronomy using neural-network classifier, in Proc. 24 ICRC, vol. 3, p. 742, Rome, 1994 (Chilingarian et al., 1994).
56. Chilingaryan A. A., Detection of weak signals against background using neural network classifiers, Pattern Recognition Letters, 16, 333, 1995 (Chilingarian, 1995).
57. Chilingarian A. A. for KASCADE Collaboration, Ter-Antonyan S. Vardanyan A. A., et al., How to infer the mass composition from EAS observations demonstrated with KASCADE data, in Proc. 25 ICRC, vol. 4, p. 105, Durban, 1997 (Chilingarian et al.,

1997).

58. Chilingarian A. A., Ter-Antonyan S., Vardanyan A. A., The comparison of Bayesian and Neural techniques in problem of classification to multiple categories, NIM, A 389, 230, 1997 (Chilingarian et al., 1997A).
59. Chilingarian A. A., Ter-Antonyan S., Vardanyan A. A., et al., On the nonparametric classification and regression methods for the multivariate EAS data analysis, Nuclear Phys. B, 52B, 237, 1997 (Chilingarian et al., 1997B).
60. Chilingaryan A. A., The non-linear signal domain selection using a new quality function in neural net training, Nuclear Instruments and Methods, A 389, 242, 1997 (Chilingarian, 1997C).
61. Chilingarian A. A., Ter-Antonyan S. V, Vardanyan A. A., Zazyan M. Z, The Reconstruction of the Energy Spectra and Primary Composition by EAS multivariate analysis, Izv AN, 61 (3), 1997 (Chilingarian et al., 1997D)
62. Chilingarian A. A., Analysis and Nonparametric Inference in High Energy Physics and Astroparticle Physics, 1998, program Package ANI, (User's Manual, unpublished) http: //crdlx5.yerphi.am/proj/ani (Chilingarian, 1998).
63. Chilingarian A. A., Roth M. and Vardanyan A. A. for the KASCADE collaboration, Towards Target-type Experiments with mononuclear Cosmic ray Beams, in Proc. of the Workshop ANI 99, vol. FZK Internal report 6472, p. 11, Nor-Amberd, 1999 (Chilingarian et al., 1999).
64. Chilingarian A. A., Roth M., Vardanyan A. A. et al., Estimation of the energy spectrum in the knee region by the KASCADE experiment, in Proc. 26th ICRC, vol. 1, p. 226, Salt Lake City, Utah, 1999 (Chilingarian et al., 1999A).
65. Chilingarian A. A., Roth M., Vardanyan A.A. for KASCADE collaboration, A non-parametric approach for determination of energy spectrum and mass composition of cosmic rays from EAS observables, Nuclear Phys. B, 75A, 302, 1999 (Chilingarian et al., 1999B).
66. Chilingarian A.A, Gevorgyan N. et al., Proc. 26th ICRC, 6, Salt-Lake-City, 460, 1999 (Chilingarian et al., 1999C)
67. Chilingarian A., Gharagyozyan G., et al.,(1999), in Proc. Of the Workshop ANI 99, Ed. Chilingarian , Haungs,.Rebel, Zzazyan, Nor Amberd, 1999, FZK preprint 6472 (Chilingarian et al., 1999D)
68. Chilingaryan A A et al., 1999 Proc. 24th Int. Cosmic Ray Conf. (Salt Lake City) vol 1, p 240 (Chilingaryan et al., 1999E)
69. A.Chilingaryan, N. Gevorgyan, A. Vardanyan, D. Jones and A. Szabo, 2002, Math. Biosciences, Vol. 176 (1) pp. 59-69 (Chilingaryan et al., 2002).

70. Chilingarian, 2003 (Chilingarian, 2003)
71. Chilingarian A.A., Arakelyan A., et al., (2003), accepted for publication in NIM A (Chilingarian et al., 2003).
72. Chilinagarian A.A., Babayan V.Kh. et al., accepted for publication in Advances in Space Research, 2003 (Chilinagarian et al., 2003A).
73. Chilingarian A.A,, Vardanyan A.A. 2003, Nuclear Instruments & Methods, 502/2-3, p 787 (Chilingarian et al., 2003B)
74. Chilingarian A.A., Martirosian G.S., Gharagyozyan G.(2003b), The Astrophysical Journal, 597:L129 (Chilingarian et al., 2003C)
75. Chupp E.L., Forrest D.J., et al., (1982), AJ letters, 263, L95 (Chupp et al., 1982).
76. Chupp E.L., Debrunner H.., et al., (1987), AJ, 318, 913 (Chupp et al., 1987).
77. CRD, The Current Status of Linux Driver of MiND PCI board and Results Performance Tests of SAND/1 neuro-chip, report, October 1999 (CRD, 1999)
78. CRD, Gamma/hadron separation estimates based on first MAGIC simulations, report, July 2002 (CRD, 2002)
79. Cronin J.W., Gibbs K.G., and Weeks T.C., The search for Discrete Astrophysical Sources of Energetic Gamma Radiation, Center for Astrophysics, Preprint series 3630 (Cronin et al.)
80. Danaher S., Fegan D.J. and Hagan J., Application of singular value decomposition (SVD) in high energy gamma-ray astronomy, Astroparticle Physics, 1, 357-368, 1993 (Danaher, 1993).
81. Danilova T.V., et al., (1992), NIM, A323, 104-107 (Danilova et al., 1992).
82. Dar A, Plaga R., (1999), A and A, 349, 259 (Dar et al., 1999)
83. DenisovaV. A., Dunaevsky A. M., Slavatinsky S. A., Chilingarian A. A., Galfayan S. Kh., Zazyan M.Z., 1987. In Proc. 20 ICRC, 1, 390 (Denisova et al., 1987).
84. Dorman, L.I., Venkatesan D., 1993, Space Science Rev., 64, 183 (Dorman et al., 1993)
85. Dorman L.I., Iucci N., Villoresi G., 1993, Astrophys. & Space Sci., 208, 55 (Dorman et al., 1993A)
86. Dorman, L. I., Proc. 26th ICRC, 6, 382, Salt-Lake-City, 1999 (Dorman 1999).
87. Dunaevsky A. M., 1986, In Proc. 4 Int. Symp. Very High Energy Cosmic Ray Interactions, Beijing, pp.1-103 (Dunaevsky, 1986).
88. Devroye L., Gyorfi L., Nonparametric density estimation. The LI view, Jown Wiley and Sons, New-York, 1985 (Devroye et al., 1985).
89. Drescher H. J. et al., NEXUS: The New Strong Interaction Model, 1999, preprint hep-ph/9903296 (Drescher et al., 1999)
90. Drury L. O'C., An introduction to the theory of diffusive shocks acceleration of energetic particles in tenuous plasmas, Rep. Prog. Phys., 46, 973, 1983 (Drury, 1983).
91. Eadie E. A., Drijard D., James F. E., Ross M. and Sadoulet B., Statistical methods in experimental physics, North-Holland, Amsterdam, 1971 (Eadie et al., 1971).
92. Edwards W., Lindman H., Savage L. J., Bayesian Statistical Inference, in Robustness

of Bayesian analyses, Elsevier Science Publishers, 1984 (Edwards et al, 1984).

93. Efron B., Nonparametric standard errors and confidence intervals, Canadian J. Statist., 9, 139, 1981 (Efron, 1981).
94. Erlykin A. D. Wolfendale A. W., A single source of cosmic rays in the range 1015 — 1016 eV, J. Phys. G, 23, 979, 1997 (Erlykin et al, 1997).
95. Fegan D.J., Gamma/hadron separation at TeV energies, Topical Review, Phys: G:Nucl. Part. Phys. 23, 1997, 1013-1060 (Fegan, 1997).
96. Finley J.P., Ogelman H., Kisilogy U., 1993, ApJ, 394, L21 (Finley et al., 1993).
97. Fischer T., Eppler W., Gemmeke H., Kock G., Becher T., The SAND Neurochip and its Embedding in the MiND System, [Techreport] (Fischer et al.).
98. Fischer T., Eppler W., Gemmeke H., Menchikov A., Novel Digital Neural Hardware For Trigger Applications IN Particle Physics, in Proc. of 2nd Workshop on Electronics for LHC Experiments, Balatonfured, Hungary, 1996 (Fischer et al., 1996).
99. Fix E., Hodges J. L., Discriminatory analysis. Nonparametric discrimination, Consistency Properties, Project 21-49-004, Report 4,USAF School of Aviation Medicine, 1951(Fix et al., 1951).
100. Fluckiger E.O., (1991) Solar Cosmic Rays, Nuclear Physics B, 22B, 1-20 (Fluckiger, 1991).
101. Forbush S.E., (1946), Tree Unusual Cosmic Ray Increases Possibly Due to Charge Particles from the Sun, Phys. Rev. 70, 771 (Forbush, 1946).
102. Fowler J. W., Forston L. F., Juni C.C.H., Kieda D. B., Ong R.A., Pryke C. L., Dommers P., A Measurement of the Cosmic Ray Spectrum and Composition at the Knee, 2000, [preprint astro-ph/00003190] (Fowler et al., 2000).
103. Fowler J W et al., 2001, Astropart. Phys. 15 49 (Fowler et al., 2001)
104. Friedman J.H., On Bias, Variance 0/1 – loss, and the Curse-of-Dimensionality, Department of Statistics, Stanford Univ., report, 15 May, 1996 (Friedman, 1996).
105. Fukunaga K., Himmels D., Bayes error estimation using Parzen and KNN procedures, IEEE Trans., PAMI-9, 634, 1987 (Fukunaga et al., 1987).
106. Fukunaga K.: Introduction to Statistical Pattern Recognition. Academic Press, Harcourt Brace Jovanovich Publishers, 1990 (Fukunaga, 1990).
107. Gaisser T. K., Cosmic rays and Particle Physics, Cambridge Univ. Press, Cambridge, 1992 (Gaisser, 1992).
108. Gaensler B.M., (2003), RevMexAA, 15, 234 (Gaensler, 2003)
109. Garyaka A.P., Martirosov R.M., et al., 2002, J.Phys.G, 28, 2317 (Garyaka et al., 2002).
110. Geisser S., The Predictive Sample Reuse Method with Applications, JASA, 70, 350, 1975 (Geisser, 1975).
111. Gharagyozyan G. V. for the ANI Collaboration, Investigation of the EAS Size Spectra in the Knee Region, in Proc. ANI 98 Workshop, vol. FZK Internal report 6215, p. 51, Nor-Amberd, 1998 (Gharagyozyan, 1998).
112. Gharagyozyan G. V. et al., The Dependence of the Age Parameter from EAS Size

and Zenith Angle of Incidence, in Proc. of the Workshop ANI 99, vol. FZK Internal report 6472, p. 53, Nor-Amberd, 1999 (Gharagyozyan et al., 1999).

113. Glasstetter R. for the KASCADE Collaboration, Electron and Muon Ratio Size Spectra of EAS below and above 'Knee', in Proc. 25th ICRC, vol. 6, p. 157, Durban, 1997 (Glasstetter, 1997).

114. Glasstetter R. for the KASCADE Collaboration, Electron, Muon and Hadron Size Spectra of EAS in the 'Knee' region, in Proc. 16th ECRS, p. 564, Alcala de Henares, Spain, 1998 (Glasstetter, 1998).

115. Glasstetter R. et al., Analysis of the muon/electron ratio in EAS, in Proc. 26 ICRC, vol. 1, p. 222, Salt Lake City, Utah, 1999 (Glasstetter et al., 1999).

116. Gemmeke H., W. Eppler, T. Fischer, A. Menchikov, S. Neusser, Neural Network Chip for Trigger Purposes in High Energy Physics, in NSS, 1996 (Gemmeke et al., 1996).

117. GOES, NOAA geostationary satellite, online: http://www.goes.noaa.gov/ (GOES)

118. Gopalswamy N., Yashiro S., et al, 2002, ApJ 572, L103 (Gopalswamy et al., 2002)

119. Hajek P., Havranek T., Mechanizing Hypothesis Formation, Springer Verlag, Heidelberg, 1979 (Hajek et al., 1979).

120. Hansen L. K., Salamon P., Neural Network Ensembles, IEEE trans, on PAMI, 12, 993, 1990 (Hansen et al., 1990).

121. Haungs A. et al., Cosmic ray energy spectrum around the knee by muon density measurements at KASCADE, in Proc. 26 ICRC, vol. 1, p. 218, Salt Lake City, Utah, 1999 (Haungs et al. 1999).

122. Haungs A., Rebel H., and Roth M., (2003), Rep. Prog. Phys. 66, 1145–1206 (Haungs et al., 2003).

123. Hayakawa S., Cosmic Rays Physics, vol. 22, Interscience Monographs and Texts in Physics and Astronomy, Wiley-Interscience, 1969 (Hayakawa, 1969).

124. Hecht-Nielsen R., Neurocomputing, Addison-Westly Publishing Company, Redding, 1990, MA (Hecht-Nielsen, 1990).

125. Heck D., et al., The Karlsruhe Extensive Air Shower Simulation Code CORSIKA, FZKA-Report, 6019, Forschungszentrum Karlsruhe, Germany, 1998 (Heck et al., 1998).

126. Hey J. D., An Introduction to Bayesian Statistical Inference, Martin Robertson, 1983 (Hey, 1983).

127. Hillas, A.M., Cherenkov light images of EAS produced by primary gamma rays and by nuclei. Proc. 19th Internal. Cosmic Ray Conf., USA, Vol. 3, 445-448, 1985 (Hillas, 1985)

128. Hillas A.M., Simple Multi-Parameter Cluster Analysis of Images, Note to Whipple Collaboration from Michael Hillas, 9 May, 1990 [Revised 12 May] (Hillas, 1990)

129. Hillas A.M., Differences between gamma-ray and hadronic showers, Proceedings of Heidelberg workshop, 1996 (Hillas, 1996).

130. Horne R.B., ESA Space Weather Programme Study, Alcatel Consortium Benefits of a Space Weather Programme, Alcatel Consortium, British Antarctic Survey, Cambridge, UK, 2001 (Horne, 2001).

131. Hovsepyan G. G. for the ANI Collaboration, The Measurement of the EAS Charged Particle Using a Logarithmic ADC, in Proc. ANI 98 Workshop, vol. FZK Internal report 6215, p. 45, Nor-Amberd, 1998 (Hovsepyan, 1998).

132. Hovsepyan G.G, (2002), ANI collaboration report N 2, online, http://crdlx5.yerphi.am/ani/ani_collab.html (Hovsepyan, 2002)

133. Hovsepyan et al., 2003 (Hovsepyan et al., 2003)

134. Kahler S., 1994, AJ, 428, 837 (Kahler, 1994).

135. Kahler S.W., McAllister A.H. and Cane H.V., 2000, AJ, 553, 1063 (Kahler et al., 2000).

136. Kalmykov N. N., Ostapchenko S. S., Pavlov A. I., Quark-Gluon-String model and EAS simulation problem at ultra-high energies, Nuclear Phys. B, 52B, 17, 1997 (Kalmykov.et al., 1997).

137. Kampert K-H. et al., The KASCADE experiment: Composition analysis and energy spectrum, in Proc. 26 ICRC, vol. 3, p. 159, Salt Lake City, Utah, 1999 (Kampert, et al., 1999).

138. Klages H. O., et al., The KASCADE Experiment, Nucl. Phys. B, 60B, 151, 1998 (Klages et al., 1998).

139. Klein K.-L, Chupp E.L., et al, 1999, Flare-associated energetic particles in the corona and at 1 AU, A&A, 348, 271 (Klein et al., 1999).

140. Klepach E. G., Ptuskin V. S. Zirakashvili V. N., Cosmic Ray Acceleration by Multiple Spherical Shocks, Astroparticle Physics, 2000, 13, 161-172 (Klepach et al., 2000)

141. Kolmogorov A. N., On the representation of continuous function of several variables by superposition of continuous functions of one variable and addition, Dokladi AN USSR, 115(5), 953, 1957 (Kolmogorov, 1957).

142. Koskinen H., Tanskanen E., et al., SPACE WEATHER EFFECTS CATALOGUE, ESWS-FMI-RP-0001 Issue 2.2 January 2, 2001 (Koskinen et al., 2001).

143. Koskinen H., Tanskanen E., et al., RATIONALE FOR A EUROPEAN SPACE WEATHER PROGRAMME ESWS-FMI-RP-0002, Issue 2.3 March 30, 2001 (Koskinen, 2001A).

144. Kosugi, T., Makishima, et.al., (1991), Solar Phys., 136, 17 (Kosugi et al., 1991).

145. Koyama K., Petre R., et.al.(1995), Nature 378, 1306 (Koyama et al., 1995)

146. Krennrich, F., R. Mirzoyan, et. al., Observation of VHE gamma emission from the Crab nebula with the prototype of the Hegra air Cherenkov telescope array, Proc. 23rd Internal. Cosmic Ray Conf., Calgary, Vol. 1, 251-254, 1993 (Krennrich et. al., 1993)

147. Kristiansen B. G., et al., Cosmic Rays of Superhigh Energies, Atomizdat, 1975, [in Russian] (Kristiansen et al., 1975).

148. Krucker S., Lin R.P., 2000, AJ, 542, L61 (Krucker et al., 2000).

149. Kudela K., Storini M., Proc. “SOLPA, Vico Equense, Italy, 289, 2001 (Kudela et al., 2001).

150. Kulikov G. V. Khristiansen G. B., On the EAS size spectra, JETF, 35, 635, 1959 (Kulikov et al., 1959).

151. Lagutin A.A., 2001, in Proc. 27-th ICRC, Hamburg, 2001, v. 5, p. 1900 (Lagutin, 2001).

152. Lang, M.G., C.W. Akerlof, M.F. Cawley, et. al., TeV Observation of the Crab nebula and other plerions in the epoch 1988-91. Proc. 22nd Internal. Cosmic Ray Conf., Dublin, 204-207, 1991 (Lang, 1991).

153. Lederman E., Handbook of Applied Mathematics, Statistics, John Wiley and Sons, New-York, 1984 (Lederman, 1984).

154. Lin R.P., et al., 1995, Sol.Phys., 71,125 (Lin et al., 1995).

155. Lindley D. V., Bayesian statistics, Soc. for Indust. and Appl. Math., Philadelphia, 1978 (Lindley, 1978).

156. Lockwood J.A., Debrunner H., Fluckiger E.O., 1990, JGR, 95,4187 (Lockwood et al., 1990).

157. Lofsgaarden D. O. and Quesenberry C. D., A nonparametric estimate of a multi-variate density function, Ann. Math. Stat., 36, 1049, 1966 (Lofsgaarden et al, 1966).

158. Long K.S., Reynolds S.P., Raymond J.C. et al., (2003) ApJ, 586,1162 (Long et al., 2003)

159. MAGIC group reports, CRD, Yerevan, online:
http://crdlx5.yerphi.am/proj/magic.html (MAGIC)

160. MAGIC, online, http://hegra1.mppmu.mpg.de/MAGIC/private/outsource/Anal-MAGI (MAGIC A)

161. Mahalonobis P. C., On the generalized distance in statistics, National Inst. of India, 2, 49, 1936 (Mahalonobis, 1936).

162. Martinez M. for the MAGIC Collaboration, The MAGIC Telescope Project, in Proc. 26th ICRC, vol. 5, p. 219, Salt Lake City, Utah, 1999 (Martinez, 1999).

163. Masuda S., Kosugi T., et al., 1994, Nature, 371, 495 (Masuda et al., 1994).

164. Matsubara, Y., and Muraki, Y., et al., Proc.of 26th ICRC, 6, 42, Salt-Lake- City, 1999 (Matsubara et al., 1999).

165. Melkumyan L., Hovsepyan G., et al. (2003), Ani Collaboration report N 6, YerPhI Preprint 1587/08 (Melkumyan et al., 2003)

166. Meyer P., Parker E.N., and Simpson J.A., 1956, Phys. Rev., 104, 768 (Meyer et al., 1956).

167. Milgrom M., Usov V., Gamma ray busters as sources of cosmic rays, Astropart. Phys., 4, 365, 1996 (Milgrom, 1996).

168. Miller J.A., Guessoum N., Ramaty R., 1990, AJ, 361, 701 (Miller et al., 1990).

169. Miroshnichenko L.I., Solar Cosmic Rays, Kluwer Academic Publishers, 2001 (Miroshnichenko, 2001).

170. Moraal H., Belov A., Clem J.M., (2000), Space Science Reviews 93: 285-303 (Moraal et al., 2000)

171. Munakata K., Bieber J.W. et al, Bartol Research Institute Preprint, BA-00-11, 2000 (Munakata et al., 2000)

172. Nagano M. et al., Energy Spectrum of Primary Cosmic Rays between 1014.5-1018 eV., Journal Phys. G: Nucl. Phys., 10, 235, 1984 (Nagano et al., 1984).

173. Nikolsky S.I., Romakhin V.A., 2000, Phys.Atom.Nucl., 63(10), 1799 (Nikolsky et al., 2000)

174. Nikolsy S.I., Sinitsyna V.G., in Proc. 28-th ICRC, Tsucuba, 2003, p.2007-2010 (Nikolsky et al., 2003).

175. NOAA SW scales, online, http://sel.noaa.gov/NOAAscales/ (NOAA SW scales)

176. Ostapchenko S., Private Communications (Ostapchenko).

177. Pamir Collaboration, 1973, In Proc. 14 ICRC, 7, 2365 (Pamir, 1973).

178. Parzen E., On estimation of a probability density function and mode, Ann. Math. Stat., 33, 1065, 1962 (Parzen, 1962).

179. Petersen C., Rognvaldsson T., Lonnblad L., JETNET 3.0 - A Versatile Artificial Neural Network Package, LU TP 93-29 7135/94, CERN-TH., 1993 (Petersen et al., 1993).

180. Peterson C., Hartman E., Exploration of the Mean Field Theory Learning Algorithm, Neural Networks, 2, 475, 1989 (Peterson et al., 1989).

181. Plaga R., (2002), New Astron. 7, 317-336 (Plaga, 2002)

182. Plyasheshnikov A.V.and Bignami G.F., Nuovo Cimento, 1985, 8c, 3 (Plyasheshnikov et al., 1985).

183. Punch, M., C.W. Akerlof, M.F. Cawley, et. al.. Supercuts: an improved method of selecting gamma-rays. Proc. 22nd Internal. Cosmic Ray Conf., Dublin, 1, 464- 467, 1991 (Punch et al., 1991).

184. Punch M., Image Analysis at Whipple and "Supercuts" Technique, in Towards a Major Atmospheric Cerenkov Detector 3, (1994), edited by T. Kifune. Unoversal Academy Press Inc. (Punch, 1994)

185. Rabiner L. R., Levinson E., Rozenberg A. E. and Wilpon J, G., Speaker - independent recognition of isolated words using clustering techniques, IEEE Trans, on Acoustics, Speech, Signal Processing, ASSP-27, 336, 1974 (Rabiner et al., 1974).

186. Rao M. V. S., Sreekantan B. V., Extensive Air Showers, World Scientific, Singapore, New-Jersey, London, Hongkong, 1998 (Rao et al., 1998).

187. Reames D.V., 1999, Space Sci. Rev., 90, 413 (Reames, 1999).

188. Reames D.V., SEPs: Space Weather Hazard in Interplanetary Space, in Space Weather, ed. By P.Song and H.J.Singer, Geophysical Monograph 125, AGU, Washington DC, 101, 2000 (Reames, 2000)

189. Reynolds P.T., C.W.Akerlof, et al., Survey of candidate gamma-ray sources at TeV energies using high resolution Cherenkov imaging system:1988-1991, AJ, 404:206-218, 1993 (Reynolds et al., 1993).

190. Riedmiller M. and Braun H., A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP algorithm, Proceedings of the IEEE Int. Conf. on NN, p. 586, 1993, editor H. Ruspini (Riedmiller et al., 1993).

191. Riedmiller M. and Braun H., A Direct Adaptive Method for further Backpropagation Learning: The RPROP Algorithm, in Proc. ICNN, San Francisco, 1993 (Riedmiller et al., 1993A).

192. Roinishvili N.N., 1995, J.Phys. G:Nucl. Part. Phys. 21, pp. 1497-436(Roinishvili,1995).

193. Rosenblatt M., Remarks on some nonparametric estimates of a density function, Ann. Math. Stat., 27, 832, 1957 (Rosenblatt, 1957).

194. Roth M., Analysis of air shower measurements of the KASCADE experiment by applying statistical classification methods for the determination of the energy and type of the primary cosmic radiation, FZKA 6262, Forschungszentrum Karlsruhe, IK3, Forschungszentrum Karlsruhe, Germany, 1999 (Roth, 1999).

195. Roth M., Chilingarian A. A., Vardanyan A. A. et al., A multivariate approach for the determination of the mass composition in the knee region, in Proc. 26th ICRC, vol. 1, p. 333, Salt Lake City, Utah, 1999 (Roth et al., 1999).

196. Roth M. et al., for KASCADE Collaboration (2001), Proc. 27th Int. Cosmic Ray Conf. (Hamburg) vol 1, p 88 (Roth et al., 2001)

197. Ruck D. W., Rogers K. S. et al, The Multilayer Perceptron as an Approximation to a Bayes Optimal Discriminant Function, IEEE Trans, on Neural Networks, 1, 296, 1990 (Ruck et al., 1990).

198. Rumelhart D. E., Hinton G., and Williams R., Learning Internal Representation by Error Propagation, no. 1 in Parallel Distributed Processing, MIT Press, Cambridge, 1986, MA (Rumelhart et al., 1986).

199. Samorski M, Stamm W., (1983) Astrophys. J. v.268, p.L17 (Samorski et al., 1983)

200. Samorski M, Stamm W., (1984) Astrophys. J. v.277, p.897 (Samorski et al., 1984)

201. Schatz G., (2002), Astropart. Phys. 17, 13 (Schatz, 2002).

202. Schlickeiser R., 2002, Cosmic Ray Astrophysics, Springer, Berlin, Heidelberg, New-York, Tokyao, p.69 (Schlickeiser, 2002).

203. Shea M.A., Smart D.F., 2000, Space Science Reviews 93: 229-262 (Shea et al., 2000).

204. Schena M., Shalon D., et al., 1995, Science 270, pp.467-470 (Schena et al., 1995)

205. Schiffman W., Joost M. and Werner R., Comparisons of Optimized Backpropagation Algorithms, in Proc. ESANN 93, Brussels, 1993 (Schiffman et al., 1993).

206. Sinkus R., A novel approach to error function minimization for Feedforward Neural Networks, NIM, A-361, 290, 1995 (Sinkus, 1995).

207. Snappin S. M., Knoke J. D., Classification error rate estimators evaluated by unconditional mean squared error, Technometrics, 26, 371, 1984 (Snappin et al., 1984).

208. Sobol I. M., On the systematic search in a hypercube, SIAM J. Number. Anal, 16, 790, 1979 (Sobol, 1979).

209. Sokhoyan S. H. for the ANI Collaboration, Investigation of the integral EAS size spectra in the knee region, in Proc. ANI 98 Workshop, vol. FZK Internal report 6215, p. 55, Nor-Amberd, 1998 (Sokhoyan, 1998).

210. Space Studies Board, Radiation and the International Space Station, National Academy Press, Wahington DC, 1999, page 7, 21 (SSB, 1999).

211. Stenkin Y., (2003), Mod. Phys. Lett. A, 18(18),1225-1234 (Stenkin, 2003)

212. Stone E.C., et al., 1998, Space Sci. Rev., 86, 357 (Stone et al., 1998).

213. Szabo A., K. Boucher, W.L. Carroll, L.B. Klebanov, A.D. Tsodikov, A.Y. Yakovlev, (2002) Math. Biosciences, Vol. 176 (1) pp. 71-98 (Szabo et al., 2002).

214. Tapia R. A., Thompson J. R., Nonparametric probability density estimation, The John Hopkins University Press, Baltimore and London, 1978 (Tapia et al., 1978).

215. Ter-Antonyan S. V., Private Communications (Ter-Antonyan).

216. Thorsett, S. E., Benjamin, R. A., Brisken, W. F., Golden, A., & Goss, W. M., 2003, ApJ, 592, L71 (Thorsett et al., 2003)

217. Toussaint G. T., Bibliography of misclassification, IEEE trans, on Information, p. 472, 1974 (Toussaint, 1974).

218. Tsuchiya H., Matsubara Y., et al.., in Proc. 27 ICRC, 8, 3040-3043, Hamburg 2001 (Tsuchiya et al., 2001)

219. Tylka A.J., Cohen C.M.S., et al, in Proc. 27 ICRC, Hamburg, 2001 (Tylka et al., 2001).

220. Tylka A.J., Dietrich W.F., Lopate C., Reames D.V., in Proc. 27 ICRC, Hamburg, 2001 (Tylka et al., 2001A).

221. Tylka A.J., AGU publication N 2000JA004028, 2001 (Tylka et al., 2001B).

222. Ulrich H et al—KASCADE Collaboration 2001 Proc. 27th Int. Cosmic Ray Conf. (Hamburg) vol 1, p 97 (Ulrich et al., 2001)

223. Vacanti, G., M.F. Cawley, et. al. (1991). Gamma-ray observations of the Crab nebula at TeV energies. Aslroph. J. 377, 467-475. Weekes T.C., Cawley M.F.,Feagan D.F., et al., Observation of TeV gamma-rays from the Crab Nebyla using the Atmospheric Cherenkov Imaging Technique, Ap J,342, 379, 1989 (Vacanti et al., 1989).

224. Vapnik V. N., Estimation of dependencies based on empiric data, Nauka, Moscow, Russia, 1979, [Russian] (Vapnik, 1979).

225. Vapnik V. N., An overview of statistical learning theory, IEEE Trans, on NN, 10(5), 988, 1999 (Vapnik, 1999).

226. Vardanyan A. A., Chilingarian A. A. and Roth M. for the KASCADE collaboration, On the Possibility of Selecting Pure Nuclear Beams from Measurements of the KASCADE Experiment, in Proc. of the Workshop ANI 99, vol. FZK Internal report 6472, p. 23, Nor-Amberd, 1999 (Vardanyan et al., 1999).

227. Vardanyan et al., 2003 (Vardanyan et al., 2003)

228. Volk H.J., Bierman P.L., Maximum energy of cosmic ray particles accelerated by supernova remnants shocks in stellar wind cavities, Astrophys. J., 148, 217, 1998 (Volk et al., 1998).

229. Volk H.J., Zirakashvili V.N., in Proc. 28-th ICRC, Tsucuba, 2003, p.2031- 2034 (Volk et al., 2003)

230. Wahba G., Spline Models for Observational Data, no. 59 in Regional Conf. Series in Appl. Math., SIAM press, Philadelphia, 1990 (Wahba, 1990).

231. Watson A. A., Charged Cosmic Rays above 1 TeV, in Proc. 25th ICRC, Durban, 1997 (Watson, 1997).

232. Weber J. H. for the KASCADE Collaboration, The Electron/Muon Ratio in EAS at and above 'Knee' Region, in Proc. 25th ICRC, vol. 6, p. 153, Durban, 1997 (Weber, 1997).

233. Weber J. et al., Analysis of electron and muon size spectra of EAS, in Proc. 26 ICRC, vol. 1, p. 341, Salt Lake City, Utah, 1999 (Weber et al., 1999)

234. Werner K., Strings, pomerons and the VENUS model of hadronic interactions at ultrarelativistic energies, Phys. Rep., 232, 87, 1993 (Werner, 1993).

235. Wilde D.J., Optimum Seeking Methods, Prentice-Hall, Inc., Englewood Clifes, N.J., 1964 (Wilde, 1964).

236. Yao X., Liu Y. Lin G., Evolutionary Programming Made Faster, IEEE Trans, on Evolutionary Computation, 3(2), 82, 1999 (Yao et al., 1999).

237. Erlykin A.D., Wolfendale E.W., (1997), J.Phys.G: , 23. 979 (Erlykin et al., 1997).

238. Erlykin A.D., Wolfendale E.W., (1998), Astropart. Phys. 8, 265 (Erlykin et al., 1998).

239. Erlykin, A. D., and Wolfendale E. W. (2001), J. Phys.G, 27, 941 (Erlykin et al., 2001)

240. Erlykin A.D., Wolfendale E.W., (2003), in Proc. 28-th ICRC, Tsucuba, 2003, p.2349-2052 (Erlykin et al., 2003).

241. Zacks S., The theory of statistical inference, John Wiley and Sons, New-York, 1977 (Zacks, 1977)

242. Zhang, S.N. and D. Ramsden (1990). Statistical data analysis for g-ray astronomy. Exp. Astronomy 1, 147 (Zhang et al., 1990).

243. Zhigljavsky A.A., Theory of global random search, vol. 65 of Mathematics and its Applications (Soviet Series) [Kluwer Academic Publishers Group, Dordrecht, 1991] (Zhigljavsky, 1991).