Composition Studies with the Telescope Array surface detector

Mikhail Kuznetsov[a], Maxim Piskunov[a], Grigory Rubtsov[a], Sergey Troitsky[a], Yana Zhezher[*a] for the Telescope Array collaboration

[a]*Institute for Nuclear Research of the Russian Academy of Sciences, 60th October Anniversary st. 7a, Moscow 117312, Russia*

E-mail: zhezher.yana@physics.msu.ru

The results on ultra-high energy cosmic rays' chemical composition based on the data from the Telescope Array surface detector are reported. The Telescope Array (TA) is an experiment, located in Utah, USA, designed for observation of extensive air showers from the ultra-high energy cosmic rays. TA surface detector (SD) array consists of 507 detector stations, placed in a square grid with 1.2 km spacing with w total area of approximately 700 km^2. Each station has two layers of 1.2 cm thick plastic scintillator of 3 m^2 area. This talk is focused on the analysis of the mass composition of primary particles based on the 9 years data of the TA surface detector. The method employs the Boosted Decision Trees (BDT) technique for the analysis of the multiple composition-sensitive shower parameters. The BDT classifier is trained with the two Monte-Carlo training sets: proton, which is considered as background events, and iron, considered as signal events. The classifiers results in a single variable ξ for data and Monte-Carlo test sets. The data to Monte-Carlo comparison results in an average atomic mass of UHECR for energy range $10^{18.0} - 10^{20.0}$ eV. The comparison with TA hybrid composition results and the other experiments is presented.

---

[*]Speaker.

1. Introduction

Ultra-high-energy cosmic rays (UHECRs) are particles and nuclei of energies more than 10^{18} eV entering the Earth atmosphere. While UHECRs are registered for many years, their origin remains a puzzle for physicists. Knowledge of the UHECR mass composition is crucial for understanding the source mechanism and propagation of cosmic rays.

UHECRs can't be observed directly due to it's low flux. Particles with corresponding energies interact with the atmosphere, causing an extensive cascade of the secondary particles – so-called extensive air shower (EAS). Two large-scale EAS facilities: Pierre Auger Observatory [2] at Southern hemisphere and Telescope Array [1] at Northern hemisphere operate in the hybrid mode, measuring both the particle flux on the ground with the surface detectors and the emitted fluorescence light with the fluorescence telescopes.

The Telescope Array is an experiment designed for observation of extensive air showers from high energy cosmic rays, located in Utah, USA. The Telescope Array surface detector (SD) array consists of 507 detector units, placed in a square grid with 1.2 km spacing with total area covered approximately 700 km^2. Each detector has two layers of 1.2 cm thick plastic scintillator of 3 m^2 area each.

In this work, a method of composition studies with the Telescope Array surface detector (SD) data is suggested. A common way to obtain UHECR composition is the calculation of the $\langle X_{max} \rangle$, the average value of the atmospheric depth where the shower development reaches its maximum, which requires operation of fluorescence telescopes. The main advantage of SD data usage is that surface detectors obtain much more data than fluorescense telescopes, which can operate only on clear moonless nights (10 % duty cycle). Surface detectors, on the contrary, operate full duty cycle.

2. Method

2.1 Multivariate analysis

Multivatiate analysis is a common name for a variety of statistical techniques, used to analyze data described by more than one variable. Given a set of observables MVA transforms it into a single variable, usually called ξ, which then allows to apply conventional one-dimensional techniques.

In case of EAS data, during the reconstruction procedure a number of observables is obtained, some of which are known to be composition-sensitive. After applying the MVA method ξ disitribution for data may be compared with the corresponding distributions for Monte-Carlo (MC) event sets and as a result, average atomic mass $\langle \ln A \rangle$ as a function of energy is derived.

2.2 Boosted Decision Trees

The Boosted Decision Trees (BDT) technique is used to build a p-Fe classifier based on multiple observables [3]. The BDT classifier works as follows:

- For each variable a splitting value with best separation is found. This value divides the full range of the values of the variable into two ranges, which are called branches. It will be mostly signal in one branch, and mostly background in another. The classifier is trained using the proton MC and iron MC training sets as background and signal respectively;

2

- Then the algorithm is repeated recursively on each branch;

- The decision tree will iterate until the stopping criterion is reached (for example, number of events in a branch). The terminal node is called a leaf.

The concept of boosting allows one to create a good classifier using a number of weak ones. In this work, "adaptive boosting", or AdaBoost, was implemented [4]. In AdaBoost, a weak learner is run multiple times on the training data, and each event is weighted by how incorrectly it was classified. An improved tree with reweighted events may be built now, and as a result, averaging over all trees allows to create a better classifier.

The classifier gives a single variable ξ for each event from the data and from the MC sets. This variable ξ is the main parameter which is subsequently used to distinguish between primaries.

In this work, root::TMVA package [5] is used as a stable implementation. The classifier converts the set of observables for an event to a number $\xi \in [-1:1]$: 1 – pure signal (*Fe*), -1 – pure background (*p*).

2.3 Composition-sensitive variables

A set of thirteen composition-sensitive variables is used:

1. Linsley front curvature parameter, *a*.

   Joint fit of lateral distribution function (LDF) and shower front is performed with 7 free parameters: x_{core}, y_{core}, θ, ϕ, S_{800}, t_0, a [6]:

$$t_0(r) = t_0 + t_{plane} + a \times 0.67 \ (1 + r/R_L)^{1.5} \, LDF(r)^{-0.5},$$

$$S(r) = S_{800} \times LDF(r),$$

$$LDF(r) = f(r)/f(800 \text{ m}),$$

$$f(r) = \left(\frac{r}{R_m}\right)^{-1.2} \left(1 + \frac{r}{R_m}\right)^{-(\eta-1.2)} \left(1 + \frac{r^2}{R_1^2}\right)^{-0.6},$$

$$R_m = 90.0 \text{ m}, \ R_1 = 1000 \text{ m}, \ R_L = 30 \text{ m},$$

$$\eta = 3.97 - 1.79(\sec(\theta) - 1),$$

   where t_{plane} is a shower plane delay, S_{800} is a scintillator signal density at 800 m core distance and *a* is the Linsley front curvature parameter.

2. Area-over-peak (AoP) of the signal at 1200 m and 3. AoP slope parameter [7].

   Given a time resolved signal from a surface station, one may calculate it's peak value and area, which are both well-measured and not much affected by fluctuations.

   AoP(*r*) is fitted with a linear fit:

$$AoP(r) = \alpha - \beta \, (r/r_0 - 1.0),$$

   where $r_0 = 1200 \text{ m}$, α is *AoP*(*r*) value at 1200 m and β is it's slope.

4. Number of detectors hit.

5. Number of detectors excluded from the fit of the shower front by the reconstruction procedure [8].

6. $\chi^2/d.o.f.$

7. $S_b = \sum S_i \times r^b$ parameter [9]

$$S_b = \sum_{i=1}^{N} \left[S_i \times \left(\frac{r_i}{r_0} \right)^b \right],$$

where S_i is the signal of i-th detector, r_i is the distance from the shower core to this station in meters and $r_0 = 1000$ m – reference distance. The two S_b for $b = 3$ and $b = 4.5$ are used in the analysis.

8. The sum of the signals of all the detectors of the event.

9. Asymmetry of the signal at the upper and lower layers of detectors.

10. Total number of peaks within all FADC traces.

11. Number of peaks for the detector with the largest signal.

12. Number of peaks present in the upper layer and not in the lower.

13. Number of peaks present in the lower layer and not in the upper.

In addition, zenith angle θ and S_{800} are included in the data set.

2.4 ξ parameter conversion to $\langle \ln A \rangle$

After applying the BDT method, ξ parameter distribution is derived for proton and iron MC and for the data in each energy bin. The range between $\ln A = 0$ (proton) and $\ln A = 4.02$ (iron) is divided into 40 equal parts and at every point a "mixture" of protons and iron (e.g. 5 % p and 95 % Fe) is created.

The Kolmogorov-Smirnov (KS) distance between ξ parameter distribution of the each "mixture" and data is performed, and the case with the smallest KS-distance is chosen, thus allowing to determine $\langle \ln A \rangle$ at a particular energy bin.

$\langle \ln A \rangle$ is estimated as:

$$\langle \ln A \rangle = \varepsilon_p \times \ln (1) + (1 - \varepsilon_p) \times \ln (56),$$

where ε_p is a fraction of protons in the mixture.

3. Data set

In this work, 9-year data collected by the Telescope Array surface detector is used (2008-05-11 — 2017-05-11). Following cuts are applied:

1. quality cuts used for spectral analysis

2. 7 or more detectors triggered

3. $E > 10^{18}$ eV

After the cuts, the data set contains 18077 events.

The analysis uses two Monte-Carlo event sets modeled for the primary *p* and *Fe* with COR-SIKA and QGSJETII-03 hadronic model. The dethinning procedure is used to reproduce the shower fluctuations while using THINNING procedure to reduce required computing resources. MC sets are split into two equal parts: (I) for training the classifier, (II) for MVA estimator calculation [10].

4. Results

On the Figure 1 the average atomic mass $\langle \ln A \rangle$ is shown in comparison with the Telescope array hybrid results [11]. On the Figure 2 the average atomic mass $\langle \ln A \rangle$ is shown in comparison with the Pierre Auger Observatory surface detector results, based on muon X_{max} and risetime asymmetry calculations, respectively [12].

The mass composition obtained is qualitatively consistent with both the Telescope Array fluorescence detector and the Pierre Auger Observatory results. Within the sensitivity of the method, no energy dependence may be seen in the chemical composition. If one assumes steady composition and the QGSJETII-03 hadronic model, the average atomic mass of primary particles corresponds to $\langle \ln A \rangle = 1.8 \pm 0.1$.
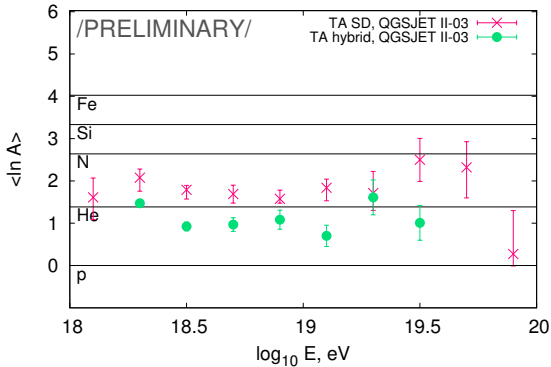
Figure 1: Average atomic mass $\langle \ln A \rangle$ in comparison with the Telescope Array hybrid results [11].
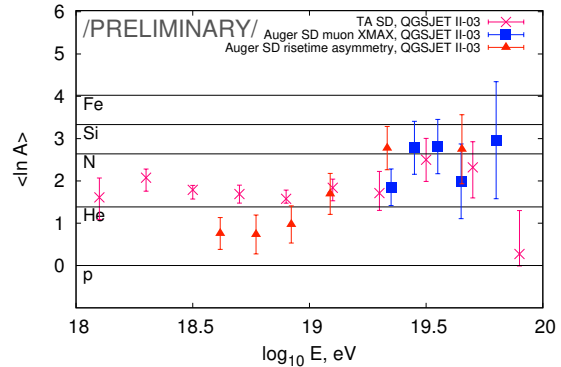
Figure 2: Average atomic mass $\langle \ln A \rangle$ in comparison with the Pierre Auger Observatory surface detector results, based on muon X_{max} and risetime asymmetry calculations [12].

Acknowledgment

References

[1] Telescope Array Collaboration, Prog.Theor.Phys.Suppl. **151** (2003) 206-210.

[2] Pierre Auger Observatory Collaboration, Nucl.Instrum.Meth. **A798** (2015) 172.

[3] L. Breiman et al., "Classification and Regression Trees" Wadsworth International Group (1984).

[4] Y. Freund, R. E. Schapire, Journal of JSAI, **14(5)** (1999) 771-780.

[5] J. Therhaag, PoS ICHEP **2010** (2010) 510.

[6] Telescope Array Collaboration, Phys.Rev.D **88**, no. 11, 112005 (2013).

[7] Pierre Auger Collaboration, Phys.Rev.Lett. **100** (2008) 211101.

[8] Telescope Array Collaboration, ApJL **768** (2013) L1.

[9] G. Ros, A. D. Supanitsky, G. A. Medina-Tanco et al. Astropart.Phys. **47** (2013) 10.

[10] B. T. Stokes et al., Astropart. Phys. **35** (2012) 759.

[11] W. Hanlon for the Telescope Array Collaboration, UHECR'16.

[12] Pierre Auger Collaboration, Contributions to the 32nd International Cosmic Ray Conference (2011).