Ultra-high-energy cosmic rays mass composition studies with the Telescope Array Surface Detector data

M. Kuznetsov, M. Piskunov, G. Rubtsov, **Y. Zhezher** and S. Troitsky for the Telescope Array collaboration

Institute for Nuclear Research of the Russian Academy of Sciences, Moscow, Russia

May 31, 2018

Abstract

- Overview
- Method: Multivariate analysis based on Boosted decision trees
- Data set and results

UHECR $\gtrsim 10^{18}$ eV composition measurements

Experiment	detector	Observable
HiRes	fluorescence stereo	X_{MAX}
Pierre Auger	fluorescence + SD (hybrid)	X_{MAX}
Telescope Array	stereo	X_{MAX}
Telescope Array	hybrid	X_{MAX}
Yakutsk	muon	ρ_μ
Pierre Auger	SD	X^{μ}_{MAX}
Pierre Auger	SD	risetime asymmetry

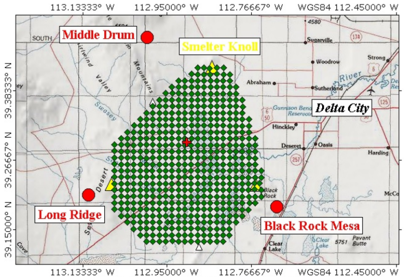
SD – surface detector

X_{MAX} *– depth of the shower maximum*

X^{μ}_{MAX} *– muon production depth*

risetime – time from 10% to 50% for the total integrated signal

Telescope Array Observatory



Map created with TOPO!® ©2002 National Geographic (www.nationalgeographic.com/topo)

Largest UHECR statistics in the Northern Hemisphere

- Utah, 2 hrs drive from Salt Lake City,
- 507 surface detectors, $S = 3\text{m}^2$, spacing 1.2 km
- 3 fluorescence detectors
- 10 years of operation

Method outline

1. Reconstruct every event, get the values of composition-sensitive observables.

2. Multivariate analysis: $(a, AoP, \ldots) \rightarrow \xi^i$. A set of observables is transformed into a single variable. The latter is used for composition analysis.

3. Compare distribution of ξ with Monte-Carlo modelling.

4. Result: average atomic mass $< \log A >$ as a function of energy.

List of relevant observables

1. Linsley front curvature parameter, a;
2. Area-over-peak (AoP) of the signal at 1200 m;

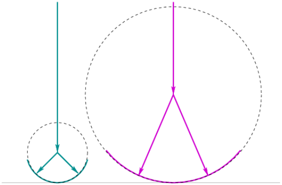
Pierre Auger Collaboration, Phys.Rev.Lett. 100 (2008) 211101

3. AoP slope parameter;
4. Number of detectors hit;
5. N. of detectors excluded from the fit of the shower front;
6. $\chi^2/d.o.f.$ of the LDF fit;
7. $S_b = \sum S_i \times r^b$ parameter for $b = 3$ and $b = 4.5$;

Ros, Supanitsky, Medina-Tanco et al. Astropart.Phys. 47 (2013) 10

8. The sum of signals of all detectors of the event;
9. Asymmetry of signal at upper and lower layers of detectors;
10. Total n. of peaks within all FADC traces;
11. N. of peaks for the detector with the largest signal;
12. N. of peaks present in the upper layer and not in lower;
13. N. of peaks present in the lower layer and not in upper;

Linsley front curvature parameter



Deeper shower maximum leads to more curved front.

Shower front is fit using the following function:

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

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

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

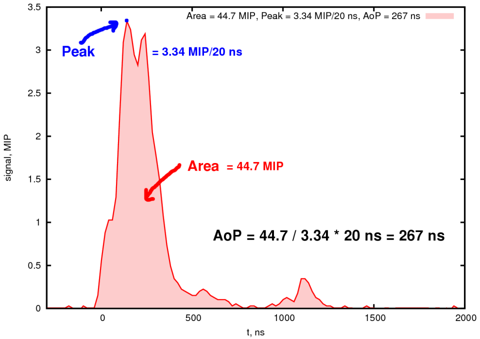
$$R_m = 90.0\text{ m},\ R_1 = 1000\text{ m},\ R_L = 30\text{ m}$$

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

t_{plane} – shower plane delay
a – Linsley front curvature parameter
LDF – lateral distribution function

Area-over-peak (AoP) and area-over-peak slope

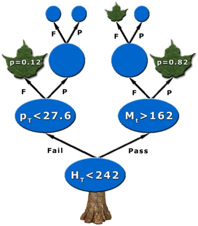
- Consider a surface station time-resolved signal



- Both peak and area are well-measured and not much affected by fluctuations

- $AoP(r)$ is fitted with a linear fit:
  - $AoP(r) = \alpha - \beta(r/r_0 - 1.0)$
  - $r_0 = 1200\,\text{m}$, α - value at 1200 m, β - slope

Boosted decision trees



Y. Coadou, ESIPAP'16

- For each variable, find splitting value with best separation between two branches: mostly signal in one branch, mostly background in another;
- Repeat algorithm recursively on each branch: take a new variable or reuse the former;
- Iterate until stopping criterion is reached (e.g. number of events in a branch). Terminal node = leaf;
- Boosting: create a good classifier using a number of weak ones (building a forest).

MVA BDT analysis

- The Boosted Decision Trees (BDT) technique is used to build *p*-*Fe* classifier based on multiple observables.

  Pierre Auger Collaboration, ApJ, 789, 160 (2014)

- BDT is trained with Monte-Carlo sets: *Fe* (Signal) and *p* (Background)

- BDT classifier is used to convert the set of observables for an event to a number $\xi \in [-1 : 1]$: 1 - pure signal (*Fe*), -1 - pure background (*p*).

- ξ is available for one-dimensional analysis.

Data and Monte-Carlo sets

- 9-year data collected by the TA surface detector:
  2008-05-11 — 2017-05-11

Cuts:

1. Events with 7 or more triggered counters
2. Events with zenith angle $\theta < 45°$.
3. Events with reconstructed core position of at least 1200 m away from the edge of the array.
4. Events with $\chi^2_G/d.o.f. < 4$ and $\chi^2_{LDF}/d.o.f. < 4$.
5. Events with geometry reconstructed with accuracy less than 5°.
6. Events with the fractional uncertainty of the S_{800} less than 25 %.
7. Events with $E > 10^{18}$ eV.

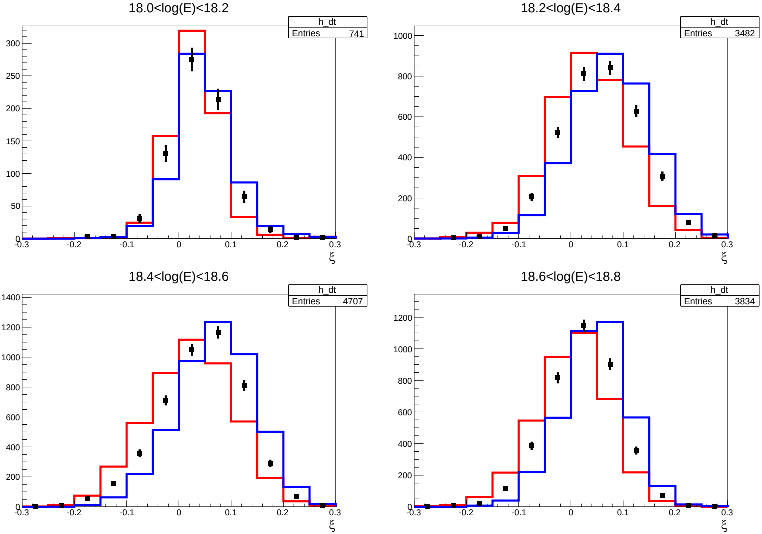
18077 events after cuts

Data and Monte-Carlo sets

p and **Fe** Monte-Carlo sets with QGSJETII-03

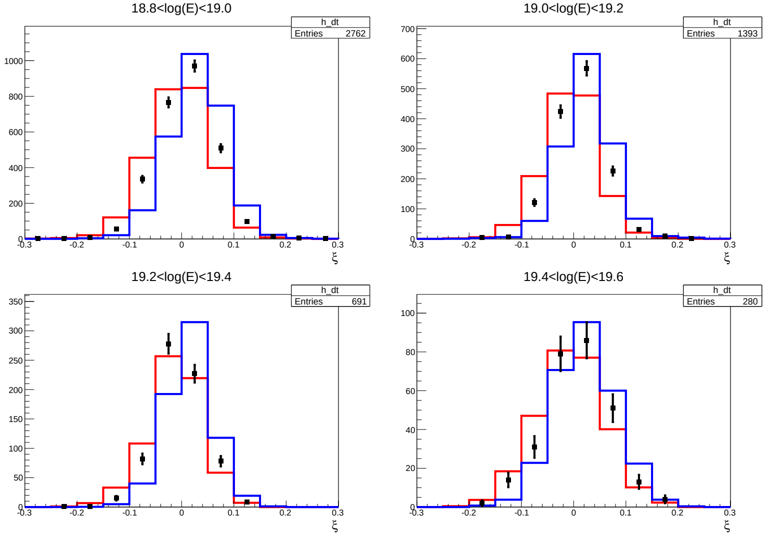
Note: MC sets are split into 3 equal parts: (I) for training the classifier, (II) for MVA estimator calculation, (III) for determination of systematical uncertanties.

Distribution of MVA estimator ξ



proton, iron, data

Distribution of MVA estimator ξ

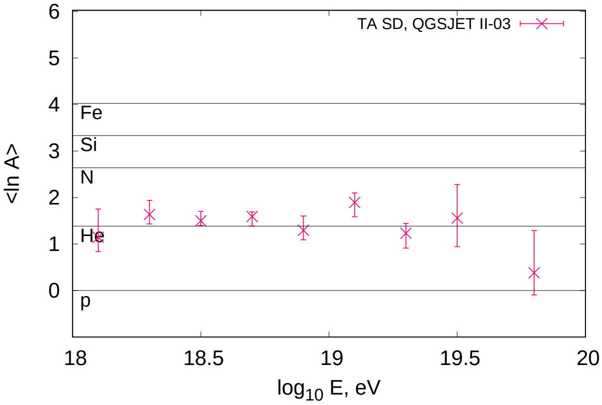


proton, iron, data

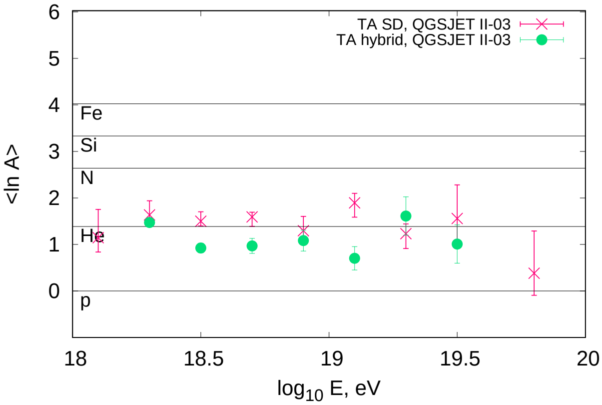
ξ parameter conversion to $\langle \ln A \rangle$

1. After applying the BDT method, ξ parameter distribution is derived for proton and iron MC and for the data in each energy bin.
2. The range between $\ln A = 0$ (proton) and $\ln A = 4.02$ (iron) is divided into 40 equal parts. At every point a “mixture” of protons and iron (e.g. 5 % p and 95 % Fe) is produced.
3. KS-distance between ξ parameter distribution of the each “mixture” and data is performed, and the case with the smallest KS-distance is chosen.
4. First approximation of average $\ln A$ is estimated as $\langle \ln A^{(1)} \rangle = \epsilon_p \times \ln(1) + (1 - \epsilon_p) \times \ln(56)$, where ϵ_p is a fraction of protons in the mixture.

Results: TA SD (MVA) composition

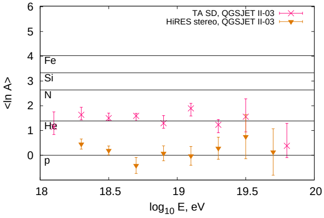


Results comparison: TA SD (MVA) vs TA hybrid

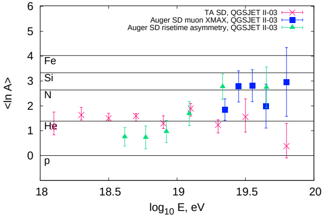


[TA] W.Hanlon, UHECR'16

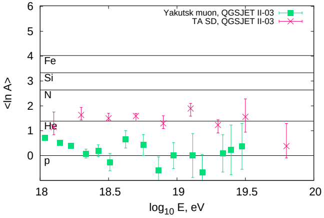
MVA result compared to other experiments



HiRes stereo, PRL, 2010



Pierre Auger Observatory X^{μ}_{MAX} and risetime asymmetry, ICRC'11



Yakutsk ρ_{μ} JPhysG, 2012

Conclusion

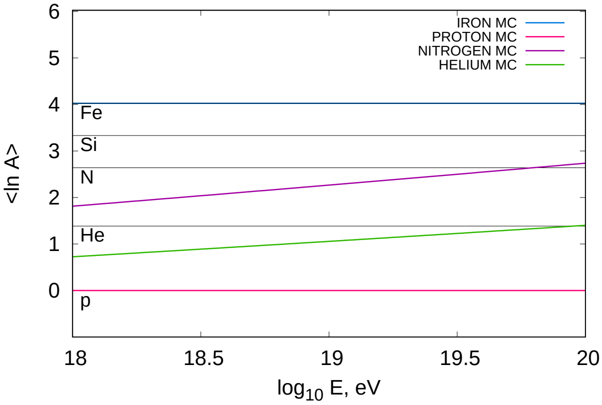
- The composition is qualitatively consistent with the TA hybrid results.
- It is also qualitatively consistent with the Auger SD results.
- The average atomic mass of primary particles corresponds to $\langle \ln A \rangle = 1.52 \pm 0.08(stat.) \pm 0.1(syst.)$.

Backup slides

Method verification

- The source of systematic uncertainties is the two-component assumption.
- The method is tested with He and N Monte-Carlo sets.

Method verification



Boosting (AdaBoost)

- Given a weak learner, run it multiple times on (reweighted) training data.
- On each iteration t: weight each training example by how incorrectly it was classified.
- New tree with reweighted events is built and optimized.
- Average over all trees to create a better classifier.

Bias corrections

Since now proton and iron MC points don't perfectly fit the straight lines $\ln A = 0$ and $\ln A = 4.02$, the data can be corrected assuming the MC points to be the endpoints of the segment $\ln A \in [0; 4.02]$ with the following linear function:

$$y_{cor} = \frac{y - y_p(x)}{y_{Fe}(x) - y_p(x)} \times \ln(56),$$

where $y_p(x)$ and $y_{Fe}(x)$ are linear approximations for the MC $\langle \ln A \rangle$ distributions.

The S_b parameter

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

S_i – signal of i-th detector
r_i – distance from the shower core to this station in meters
$r_0 = 1000$ m – reference distance
Best separation is for $b = 3$ & $b = 4.5$.

Ros, Supanitsky, Medina-Tanco et al. Astropart.Phys. 47 (2013) 10

Why primary composition is important?

- understand the acceleration mechanisms
- predict the flux of cosmogenic photons and neutrino
- probe the interaction cross-section at the highest energies
- precision tests of Lorentz-invariance