

YAC sensitivity for measuring the light-component spectrum of primary cosmic rays at the “knee” energies

L. M. Zhai^{a,b}, J. Huang^a, D. Chen^b, M. Shibata^c, Y. Katayose^c, Ying Zhang^a,
Xu Chen^a, X. B. Hu^{a,d}, Y. H. Lin^a and H. B. Jin^b

^a Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, Chinese Academy of Sciences, Beijing 100049, China

^b National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

^c Faculty of Engineering, Yokohama National University, Yokohama 240-8501, Japan

^d Department of Physics, Shandong University, Jinan 250100, China

A new air-shower core-detector array (YAC: Yangbajing Air-shower Core-detector array) has been developed to measure the primary cosmic-ray composition at the “knee” energies in Tibet, China, focusing mainly on the light components. The YAC-II consisting of 124 detectors has been constructed and operated at Yangbajing (4300 m a.s.l.) in Tibet since February 2014. The YAC-II has been placed near the center of the Tibet-III AS array and operates together with Tibet-III and a large underground water-Cherenkov muon-detector array (MD). In this paper, we have performed a detailed Monte Carlo simulation to check the sensitivity of YAC-II+Tibet-III+MD array to the light components in the primary cosmic rays around the knee energies, taking account of the observation conditions of actual YAC-II+Tibet-III+MD array. We have checked the sensitivity of YAC to the hadronic interaction models (QGSJET01c and SIBYLL2.1) and primary cosmic-ray composition models (“He-poor”, “He-rich” and “Gaisser-fit”) around the knee using the high-energy core events observed by the YAC-II+Tibet-III+MD array, and the capability of the measurement of the light-chemical components (proton, helium) with the new Tibet hybrid experiment is investigated. The simulation shows that the light-component spectrum estimated by our methods can well reproduce the input ones within 10% error, and there will be about 30% systematic errors mostly induced by the primary and interaction models used.

*The 34th International Cosmic Ray Conference,
30 July- 6 August, 2015
The Hague, The Netherlands*

1. Introduction

The “knee” of the all-particle energy spectrum is considered to be closely related to the origin, acceleration and propagation mechanism of cosmic rays. In order to explain the existence of the knee, many hypotheses and mechanisms have been proposed [1, 2]. Although all these approaches can well describe the knee structure, there are much discrepancies in the prediction of the individual components at the knee region. Therefore, precise measurements of the chemical composition around the knee, especially, measurements of the primary spectra of individual component till their knee, will be essentially important to understand the origin, acceleration and propagation mechanism of cosmic rays.

Until now, we may have no choice but to rely on ground-based air-shower (AS) measurements to study the primary chemical composition around the knee. The early relevant work was done by the Tibet old hybrid experiment of the emulsion chambers (ECs), the burst detectors (BDs) and the AS array (Tibet-II), however, the energy range of primary particles was limited to be above ~ 300 TeV because of the ECs’ high threshold energy [3], so was the low statistics of the high-energy core events. To improve this condition effectively, a new air-shower core detector named YAC-II (Yangbajing Air shower Core detector) has been developed so as to meet our requirements [4]. The YAC-II aims to observe the energy spectrum of proton and helium whose energy range will overlap with direct observations at lower energies such as CREAM, ATIC and TRACER, and Tibet-EC experiment at higher energies [5]. In this paper, we discuss the sensitivity and performance of YAC for observing light-component spectrum of primary particles through detailed Monte Carlo (MC) simulations based on the (YAC-II+Tibet-III+MD) hybrid experiment.

2. The experimental setup

The new Tibet hybrid experiment currently consists of three types of detector array, including the Tibet AS array (Tibet-III), the Yangbajing AS core-detector array (YAC-II) and an underground water-Cherenkov muon-detector array (MD), as shown in Fig.1. The Tibet-III consists of 789 detector units, with a covering area $\sim 50,000$ m²; the YAC-II consists of 124 detector units, with a covering area of ~ 500 m² and the five MD pools consists of 80 cells, with an area of ~ 4500 m². The (Tibet-III+YAC-II+MD) hybrid experiment has been constructed and operated at Yangbajing (4300 m above sea level; 606 g/cm²) in Tibet, China, since February 2014. This hybrid experiment is designed not only for observation of air showers of nuclear component origin, but also for observation of high-energy celestial gamma rays by using (MD+Tibet-III). The detail description about the hybrid experiment could be seen in paper [5].

The YAC-II array is constructed near the center of the Tibet-III (shown in Fig.1), and it has been operated simultaneously with Tibet-III and MD array. For an air shower event, the Tibet-III provides the arrival direction (θ, φ) and the air shower size (N_e) which are interrelated to primary energy, the YAC-II measures the high energy electromagnetic particles in the core region so as to obtain the characteristic parameters of air-shower cores, at the same time, the underground MDs record the high-energy muons above 1 GeV. When a YAC event is triggered, its accompanying air shower is simultaneously recorded. The matching between YAC, AS and MD events is made by their arrival time stamps.

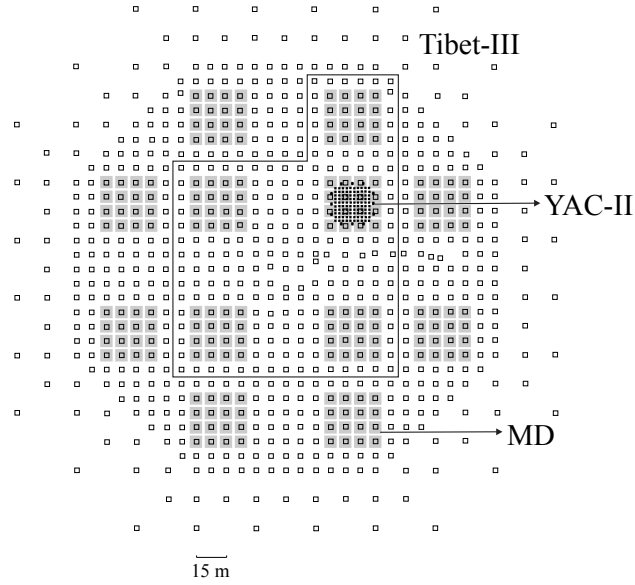


Figure 1: Schematic view of (Tibet-III+YAC-II+MD) array [5]. Open squares represent surface 0.5 m^2 scintillation detectors of the Tibet-III array. Filled gray squares represent the 12 planned pools, each of which consists of 16 cells with 52 m^2 , the five pools in the solid frame have been already constructed. Filled black squares represent 0.4 m^2 YAC detectors consisting of 3.5 cm thick lead layer and 1 cm thick scintillator.

3. Monte Carlo Simulations

We have carried out a full Monte Carlo (MC) simulation on the development of air showers in the atmosphere using the simulation code Corsika [6]. Two hadronic interaction models, including QGSJET01c and SIBYLL2.1, are used to generate the air-shower events in the atmosphere. Furthermore, we have checked other new hadronic interaction models including EPOS-LHC and QGSJETII-04 [4]. For the primary cosmic rays, we examined three composition models, namely, “He-poor”, “He-rich” and “Gaisser-fit” models, in order to evaluate the systematic errors attributable to primary composition models [4, 5]. The proton spectra of the former two models are fitted to the direct measurements at the low energy and consistent with the spectrum obtained from the Tibet-EC experiment at the high energy. The He spectrum of He-poor model coincides with the results from RUNJOB, but the He spectrum of He-rich model coincides with the results from JACEE, ATIC2 and CREAM. The Gaisser-fit model fits to a higher He model (almost same as our He-rich model) at the low energy range and to the KASCADE-QGSJET data at high energy range in which light components (P and He) dominate in the chemical composition. In all models mentioned above, each component is summed up so as to match with the all-particle spectrum with a sharp knee, which was obtained with the Tibet-III AS array [7]. The energy spectra of individual components (or mass groups) for three primary models are shown in Fig.2. It is seen that all the individual components of the three models in the low energy range (less than 100 TeV) are in good agreement with direct measurements while differ significantly at higher energy. The all-particle spectra of three models, however, coincide with each other and reproduce the sharp knee structure as well [7].

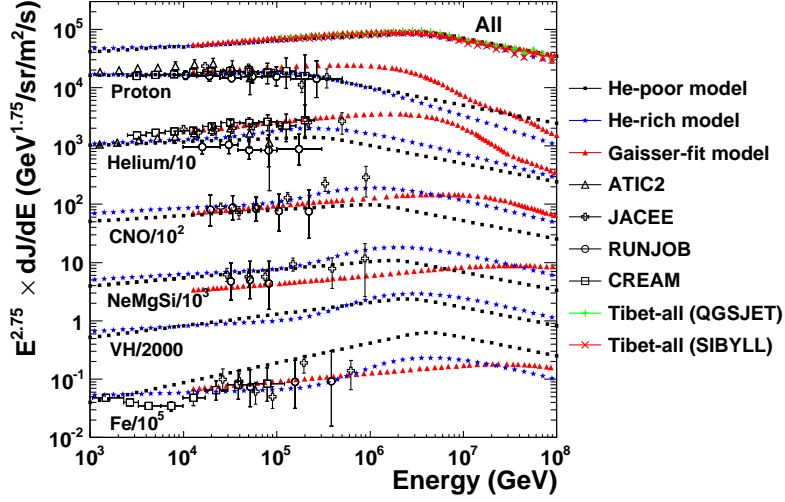


Figure 2: Primary cosmic-ray composition for He-poor, He-rich and Gaisser-fit models compared with those of direct measurements and the sum of all components (all-particle spectrum) compared with the results obtained by the Tibet-III experiment [4].

The detector responses to shower particles falling on the detectors of (YAC-II+Tibet-III+MD) array are calculated using the Geant4 [8], where the detector performance, trigger efficiency and effective area are adequately taken into account based on the experimental conditions. After reconstruction of air shower, we can obtain the following 8 parameters of the (Tibet-III+YAC-II+MD) array to characterize an air-shower core event: θ , N_e , N_{hit} , N_b^{top} , $\sum N_b$, $\langle R \rangle$, $\langle N_b R \rangle$ and N_μ , details about the reconstruction can be seen in paper [5]. Among the eight parameters, the first two parameters (θ , N_e) are reconstructed by Tibet-III, the middle five parameters are reconstructed by YAC-II and the last one (N_μ) is reconstructed by the MD array. We did the study of mass composition of primary cosmic rays based on those parameters.

In this simulation, two kinds of data sets are generated by Corsika. One is the data set I ($E \geq 1$ TeV), which is used to obtain the primary energy spectra of the low-energy range (50-300 TeV), and another is the data set II ($E \geq 50$ TeV), which is used to obtain the spectra of high-energy range (300-10,000 TeV). To select the high-energy core events and to improve the observation efficiencies, two data selection conditions are adopted for two different energy ranges. For low-energy range (50-300 TeV), the core-event selection condition is: $N_b \geq 100$, $N_{hit} \geq 4$ and $N_e \geq 30,000$. For high-energy range (300-10,000 TeV), we add $\sum N_b \geq 3000$ and $N_b^{top_{in}} > N_b^{top_{out}}$ to reject events falling far from the YAC-II array [5]. The final core-event selection condition for high-energy range is: $N_b \geq 100$, $N_{hit} \geq 4$, $N_e \geq 80,000$, $\sum N_b \geq 3000$ and $N_b^{top_{in}} > N_b^{top_{out}}$. The behavior of the effective $S\Omega$ of YAC-II array is shown in Fig.3 for proton and helium with two hadronic interaction models QGSJET01c and SIBYLL2.1 under two core-event selection conditions for data set I (Fig.3a) and data set II (Fig.3b). The effective $S\Omega$ depends weakly on the model used, but its difference is found to be smaller than 25% in our concerned energy range. The differences of observation efficiency by interaction models will be reflected in the absolute intensity of the primary energy spectra.

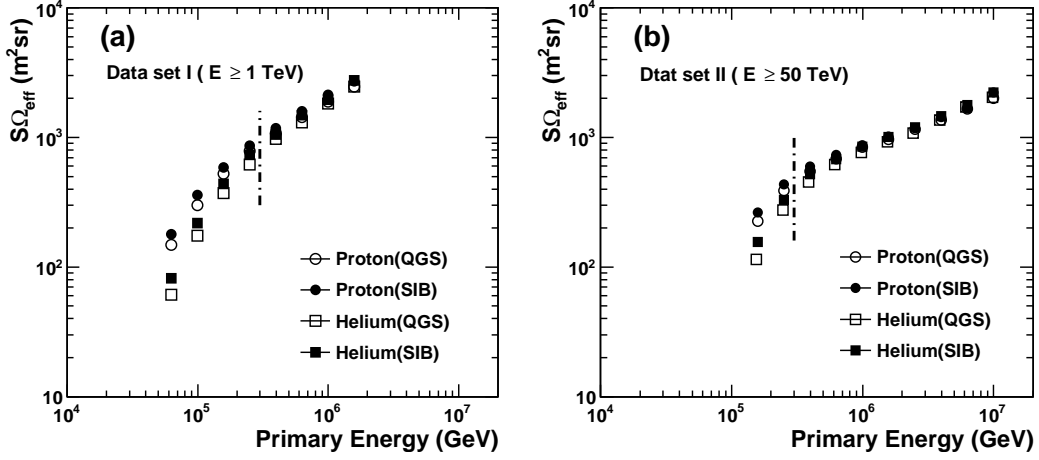


Figure 3: The effective $S\Omega$ of proton and helium under different hadronic interaction models ((a):data set I, and (b):data set II). We use the data points below 300 TeV in figure (a) and data points above 300 TeV in figure (b).

4. Results and Discussion

4.1 Selection of the primary Proton and Helium

The separation of the primary mass is realized with use of a feed-forward artificial neural network (ANN) method [9]. The following 8 parameters are input to the ANN with 40 hidden nodes and 1 output unit: (1) N_{hit} , (2) N_b^{top} , (3) $\sum N_b$, (4) $\langle R \rangle$, (5) $\langle N_b R \rangle$, (6) N_e , (7) θ , (8) N_μ . Firstly, we need to separate protons from other nuclei by training the network with a proton flag, and then separate proton+helium from other nuclei by training the network with a proton+helium flag. Thus, we can get the helium spectrum by subtracting the derived proton spectrum from the derived proton+helium spectrum. To train the ANN in separating protons from other nuclei, the input patterns for protons and other nuclei are set to 0 and 1, respectively, as well as for (proton+helium) and other nuclei. We then define a critical value of T_c to calculate the purity and selection efficiency of the selected Proton-like or (P+He)-like events.

Fig.4 shows the ANN output distribution for proton and (proton+helium) trained using the (QGSJET01c+He-poor) model. As seen in the Fig.4a, the events with $T_c \leq 0.2$ could be regarded as the Proton-like events, and the average selection purity and efficiency over whole energy range ($E \geq 50$ TeV) are 89% and 53%, respectively. We also accomplished the ANN training of (proton+helium), and the average selection purity and efficiency are 97%, 85%, respectively, using $T_c = 0.1$, as shown in Fig.4b. We found that ANN method works well to separate primary proton or (proton+helium) group from other nuclei almost independently of the used hadronic interaction models. The difference in the selection efficiency is at most 5% between different interaction models [5].

Fig.5-left shows the ANN output distribution for proton under different primary composition models (He-poor and Gaisser-fit). From this figure, we also found that the ANN test results trained by different composition models are almost the same. The slightly different shapes of the ANN

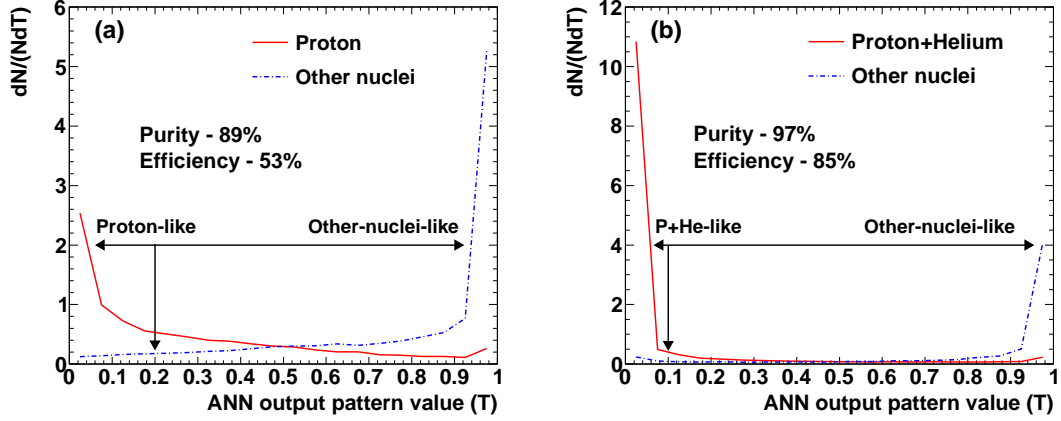


Figure 4: The ANN test results for proton (a) and (proton+helium) (b) based on QGSJET01c+He-poor model.

output distribution reflect the different composition models bringing little differences to the purity by less than 4% [5]. We also investigated possible methodical errors involved in the separation of primaries by using another machine learning method, i.e. the Random Forest (RF) method [10], which is widely used for the pattern recognition as well as ANN. We found that the two results under ANN and RF are almost the same leading to the estimation of the proton flux with less than 5% difference between them, as shown in Fig.5-right, however we found ANN reproduces the input spectrum better than RF in the high-energy range [5].

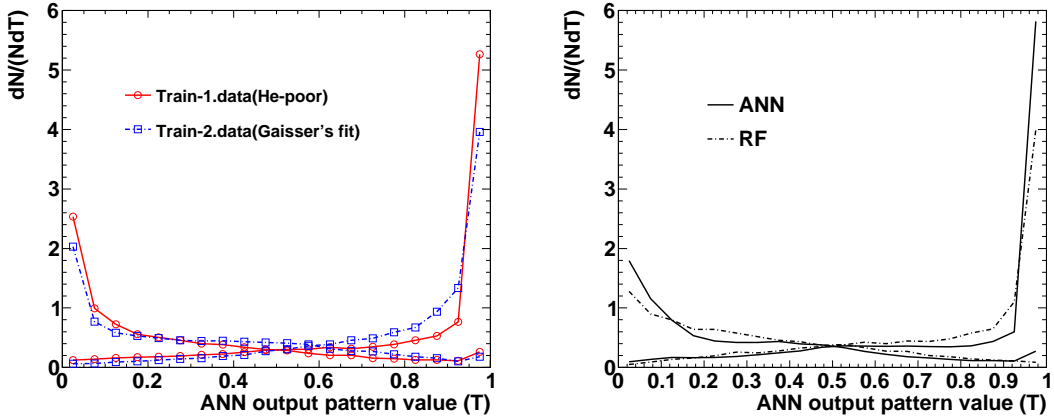


Figure 5: Left: The ANN test results for proton under different cosmic-ray composition models based on QGSJET01c hadronic interaction model. Right: The comparison of ANN and RF to separate protons based on SIBYLL2.1+He-poor model. The average purity and selection efficiency over whole energy range ($E \geq 50$ TeV) are 88%, 49% at $T_c = 0.2$ for ANN training, while 87%, 49% for RF training at $T_c = 0.25$.

4.2 Expected primary proton and helium spectrum

Using the ANN method, we select the Proton-like or (P+He)-like events from all the observed events. The primary energy E_0 of each selected events is then estimated using the AS size N_e ob-

tained by the Tibet-III. The relation between air shower size N_e and primary energy E_0 is expressed as: $E_0 = \alpha \times N_e^\beta$, where the parameters of α and β are estimated from AS events generated by the MC for the Tibet-III AS array, as shown in Fig.6-left. The energy resolution is estimated as about 12% at energies around 1 PeV [5]. We also checked the hadronic interaction model dependence in the correlation of N_e and E_0 , and we found that there is less than 3% difference for the determination of the primary energy based on different hadronic interaction models [5].

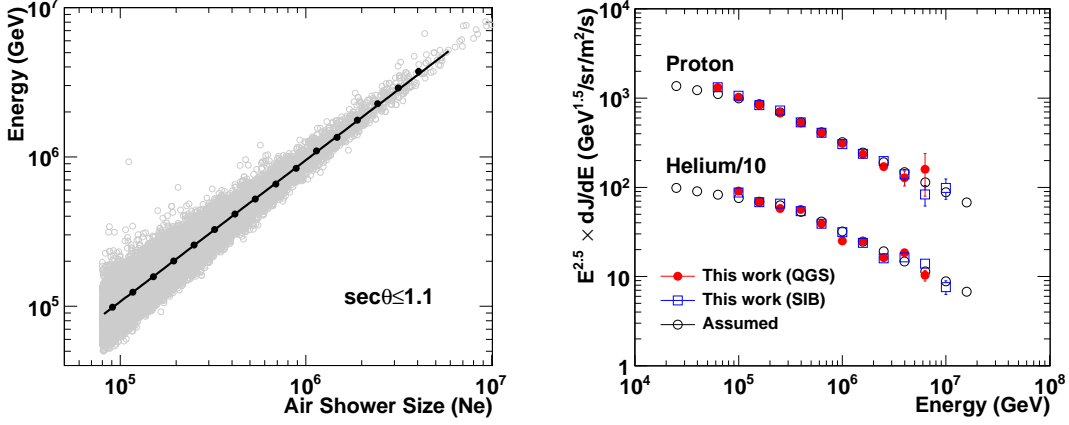


Figure 6: Left: Scatter plots of the primary energy E_0 and the estimated shower size N_e of (proton+helium)-like events based on SIBYLL2.1+He-poor model. Solid line shows the fitting result of $E_0 = 1.94 \times N_e^{0.948}$ GeV; Right: The final reconstructed energy spectra of primary protons and helium nuclei by ANN (This work) in comparison with the inputs to the simulation based on He-poor model.

Fig.6-right shows the last estimated primary energy spectra by adopting the data points below 300 TeV from data set I and data points above 300 TeV from data set II (results from two data sets are smoothly connected at 300 TeV within 5% error [5]), compared with the assumed ones under different hadronic interaction models, and we found that the estimated spectra derived by our methods can well reproduce the input ones within 10% error.

4.3 Systematic uncertainties

In this paper, we almost discuss all the possible systematic uncertainties in obtaining the flux of proton and helium as follows:

- (1) the hadronic interaction model dependence: in observation efficiency (less than 25%), in ANN method (less than 5%), in the correlation of N_e and E_0 (less than 3%);
- (2) the primary cosmic-ray composition model dependence in ANN method (less than 4%);
- (3) the different mathematical methods (ANN and RF) for nuclei-separation (less than 5%);
- (4) the reconstructed spectra using MC generated data and ANN method could reproduce the input ones within 10% error.

The total systematic errors are then estimated to be about 30% as the square root of quadratic sum of those systematic errors above, which are somewhat overestimated due to a weak correlation among those errors.

5. Summary

In this paper, we have performed a detailed Monte Carlo simulation to check the sensitivity of YAC-II+Tibet-III+MD array to the light components in the primary cosmic rays around the knee energies. We have checked the sensitivity of YAC-II to the hadronic interaction models and primary cosmic-ray composition models using the high-energy core events observed by the YAC-II+Tibet-III+MD array, and the capability of the measurement of the light-chemical components (proton, helium) with the new Tibet hybrid experiment is investigated. Our simulation result shows that the new hybrid experiment (YAC-II+Tibet-III+MD) is powerful enough to study the primary chemical compositions, in particular, to obtain the light-component spectra of the primary cosmic rays in 50-10,000 TeV energy range overlapping with the energy range of the direct observations and the ground based indirect observations around the knee where the break points of proton and helium are expected to be observed.

Acknowledgments

The authors would like to express their thanks to the members of the Tibet AS γ collaboration for the fruitful discussion. This work is supported by the Grants from the National Natural Science Foundation of China (11078002 and 11275212) and the Chinese Academy of Sciences (H9291450S3, 2013T2J0006) and the Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, CAS. The Knowledge Innovation Fund (H95451D0U2 and H8515530U1) of I-HEP, China also provide support to this study.

References

- [1] M. Shibata et al., *Chemical Composition and Maximum Energy of Galactic Cosmic Rays*, *Astrophys. J.* **716** (2010) 1076
- [2] J. R. Hörandel, *Models of the knee in the energy spectrum of cosmic rays*, *Astropart. Phys.* **21** (2004) 241 [astro-ph/0402356]
- [3] M. Amenomori et al., *Are protons still dominant at the knee of the cosmic-ray energy spectrum?*, *Phys. Lett. B* **632** (2006) 58 [astro-ph/0511469]
- [4] L. M. Zhai et al., *Sensitivity of YAC to measure the light-component spectrum of primary cosmic rays at the 'knee' energies*, *J. Phys. G: Nucl. Part. Phys.* **42** (2015) 045201
- [5] J. Huang et al., *Performance of the Tibet hybrid experiment (YAC-II+Tibet-III+MD) to measure the energy spectra of the light primary cosmic rays at energies 50-10000 TeV*, *Astropart. Phys.* **66** (2015) 18
- [6] Heck D. et al., 1998 *Forschungszentrum Karlsruhe Report FZKA 6019*
- [7] M. Amenomori et al., *The All-Particle Spectrum of Primary Cosmic Rays in the Wide Energy Range from 10^{14} to 10^{17} eV Observed with the Tibet-III Air-Shower Array*, *Astrophys. J.* **678** (2008) 1165
- [8] Agostinelli S. et al., *Geant4 - a simulation toolkit*, *Nucl. Instrum. Meth. Phys. Res. A* **506** (2003) 250
- [9] Peterson C. et al., 1994 *Comp. Phys. Comm.* **81** 185
- [10] L. Breiman and A. Cutler,
http://www.stat.berkeley.edu/~breiman/RandomForests/cc_papers.htm