

## The MVA for pure proton and mixed H&He samples selection in hybrid measurement of cosmic ray showers with LHAASO experiment

---

**L.Q. Yin<sup>ab\*</sup>, S.S. Zhang<sup>a</sup>, Z. Cao<sup>ab</sup>, L.L. Ma<sup>a</sup>, M.J. Yang<sup>a</sup>, Y. Zhang<sup>a</sup>, Z.Y. You<sup>ab</sup>, Z.Z. Zong<sup>c</sup> and Tiina Suomijärvi<sup>d</sup> for the LHAASO Collaboration.**

<sup>a</sup> Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, CAS, Beijing, China.

<sup>b</sup> University of Chinese Academy of Science, Beijing, China.

<sup>c</sup> China Nuclear Power Engineering Co., Ltd., 100084, Beijing, China.

<sup>d</sup> Institut de Physique Nucléaire d'Orsay, IN2P3-CNRS, Université Paris-Sud, Université, Paris-Saclay, 91406 Orsay Cedex, France.

E-mail: [yinlq@ihep.ac.cn](mailto:yinlq@ihep.ac.cn)

Large High Altitude Air Shower Observatory is a composite cosmic ray and gamma ray observatory consisting of three sub-detector arrays: the kilometer square array (KM2A), the water Cherenkov detector array (WCDA) and the wide field of view Cherenkov telescope array (WFCTA). LHAASO is under construction now and the 1/4 LHAASO array is planned to be completed in this year.

In this work, samples of the light components are selected by the simulation of hybrid observation with 1/4 LHAASO array. Boosted Decision Trees (BDTG) in Toolkit for Multivariate Analysis (TMVA) is used for the selection of primary particles in cosmic rays. The apertures of pure proton and mixed proton and helium samples are  $900 \text{ m}^2 \cdot \text{Sr}$  and  $1800 \text{ m}^2 \cdot \text{Sr}$  respectively with high purity. Moreover, a prospect of proton and mixed H&He spectra with energies from 100TeV to 4PeV is also presented.

*36th International Cosmic Ray Conference -ICRC2019-  
July 24th - August 1st, 2019  
Madison, WI, U.S.A.*

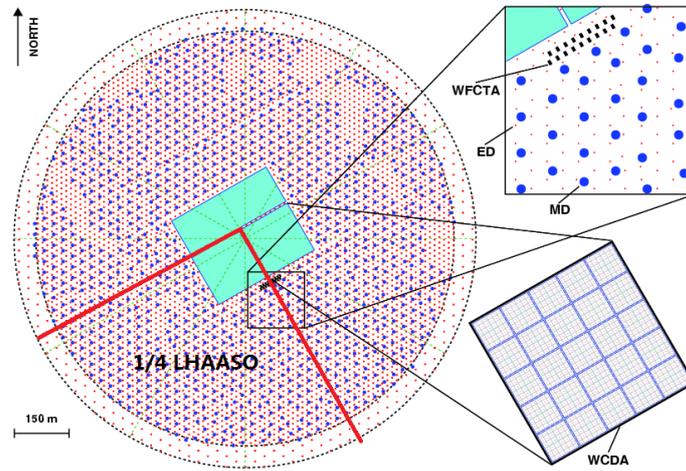
---

\*Speaker.

## 1. Introduction

The origin and acceleration of cosmic rays (CRs) have been a puzzle for more than 100 years. An unsolved problem in cosmic ray observation is the "knee" structure in the energy spectrum. Due to the lack of absolute energy scale and the way to identify the type of the primary particles of CRs, there is no consistent energy spectrum of individual species above knee region [1]. The Large High Altitude Air Shower Observatory (LHAASO) [2, 3], located in Haizhishan Daocheng, 4410 m a.s.l., Sichuan Province, China, will throw light on this traditional problem. LHAASO expects to obtain continuous energy spectra for each component of cosmic rays with energies from  $10^{14}$  eV to  $10^{18}$  eV by multi-segments and multi-parameters observation.

The detectors layout of LHAASO experiment is shown in Fig. 1. LHAASO consists of four types of detectors: electromagnetic detectors (KM2A-ED), muon detectors (KM2A-MD), water Cherenkov detectors (WCDA) and wide field of view (FOV) Cherenkov telescopes(WFCTA).



**Figure 1:** The layout of LHAASO detectors.

WCDA [4] consists of three water ponds: two have area of  $150m \times 150m$  and one has area of  $300m \times 110m$ . All of the ponds have the water depth of  $4.5m$ . Each pond is divided into small cells with an area of  $5m \times 5m$ , having two Photomultiplier Tubes (PMT) anchored on the bottom at the center of the cell. It detects the Cherenkov light produced in water by cascade processes of secondary particles in extensive air shower (EAS) [5].

WFCTA detects the Cherenkov photons in the air shower. Two telescope prototypes have been operated successfully at Yangbajing cosmic ray observatory in Tibet [6]. While eighteen telescopes in LHAASO experiment were updated. The optical sensor of the camera is Silicon photomultiplier (SiPM) [7], with 1024 pieces in total. The FOV of each pixel is  $0.5^\circ \times 0.5^\circ$ , and the total FOV is  $16^\circ \times 16^\circ$ .

The KM2A ED array [8] covers an area of  $1.3 km^2$  consisted of 5195 plastic scintillator detectors with a spacing of 15 meters. One ED detector has an effective area of  $1 m^2$ . MD array is an underground water Cherenkov detector array [9]. There are 1171 muon detectors also covering  $1 km^2$  area with a spacing of 30 meters. The area of one muon detector is  $36 m^2$ .

LHAASO is planned to obtain the consecutive measurement of energy spectra by four hybrid observation stages. First of all, WCDA can obtain absolute-energy-scale through moon shadow observation [10] up to 30 TeV. And it can be delivered to WFCTA by hybrid observation of cosmic rays with energies from 10 TeV to 100 TeV. The second stage is mainly for the knee observation of light nuclei components of cosmic rays with energies from 100 TeV to 10 PeV by the hybrid detection of WFCTA, WCDA, and KM2A. In the third stage, the layout of WFCTA will be changed to measure the knee of the heavy nuclei with energies from 10 PeV to 100 PeV [11]. WCDA will not be included in the hybrid detection due to saturation. The last stage is for the second knee study with energies from 100 PeV to 1 EeV, in which WFCTA will be operated in fluorescence observation mode [12].

This paper is devoted to the prospect of energy spectra observation of pure proton and H&He in the second stage through simulation. 1/4 LHAASO array will run for a few years to achieve the physical goal of light nuclei components spectra measurement. 1/4 array includes one of the ponds in WCDA, six Cherenkov telescopes, 1272 EDs and 300 MDs. The layout of the 1/4 array is shown in Fig. 1.

## 2. Simulation and Event Reconstructions

### 2.1 Simulation

The cascade processes of primary cosmic rays in the atmosphere are simulated by the CORSIKA [13] program with the version of 6990. The EGS4 model is chosen for the electromagnetic interaction. The QGSJET02 and GHEISHA models are chosen for the high and low energy hadronic processes respectively. Both the information of Cherenkov photons and secondary particle at the level of observatory are recorded to simulate the hybrid observation. Five components, proton, helium, CNO, MgAlSi, and iron are generated with energies from 10 TeV to 10 PeV according to a power law spectrum with an index of -2.7. The directions of the showers are sampled from  $24^\circ$  to  $38^\circ$  in zenith and from  $77^\circ$  to  $103^\circ$  in azimuth. The shower core position is evenly distributed in an area of  $260\text{ m} \times 260\text{ m}$ . The primary energy spectrum is normalized to the exponent of -2.7 from 10 TeV to 10 PeV, as shown in Fig. 2 (left). Different components are normalized to the same with proton.

### 2.2 Event Reconstructions

The simulation of the response of three LHAASO detector arrays for EAS is performed separately. Then the simulated results of different detectors are merged before reconstruction. According to the actual data acquisition, as long as the Cherenkov telescope is triggered, the events measured by other three detectors is reconstructed.

The shower geometric information is reconstructed by WCDA. The shower core position is given by NKG fitting [14] and the resolution is less than  $3m$ . Due to the lack of time response in fast simulation program [15], the arrival direction of the shower is obtained by a Gauss sampling, with angular resolution  $0.3^\circ$ .

Then the Cherenkov image measured by WFCTA is cleaned up and the Hillas parameters [16] are given, but it is  $R_p$  depended. Here  $R_p$  is the perpendicular distance between the telescope and

the shower axis. The total photoelectrons ( $N^{pe}$ ) in the cherenkov image is a good energy estimator. Before energy reconstruction,  $N^{pe}$  should be normalized to  $R_p = 0$  and  $\alpha = 0$ .  $\alpha$  is the space angle between the shower direction and the optical axis of the telescope. The normalized photoelectrons ( $N_0^{pe}$ ) can be expressed as:

$$N_0^{pe} = \lg(N^{pe}) + a \times (R_p/1m) + b \times \tan(\alpha) \quad (2.1)$$

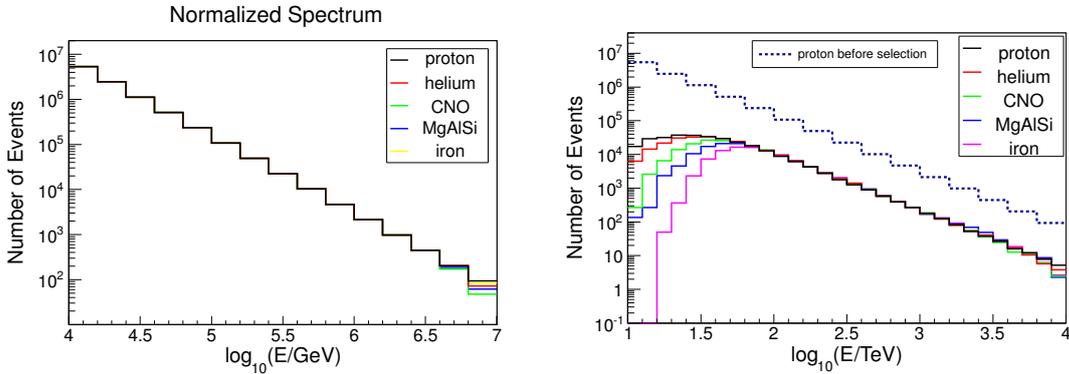
where the parameters  $a$  and  $b$  depend on the primary component of cosmic rays. For light component,  $a = 0.00916$  and  $b = 0.0182$ . The resolution of energy reconstruction is 20% for both pure proton and H&He after particle identification.

Finally, the lateral distribution of muons is fitted by an empirical formula Eq.( 2.2) through maximum likelihood fitting [5].

$$\rho(r, N_\mu) = k_G N_\mu \left(\frac{r}{r_G}\right)^{-k_1} \left(1 + \frac{r}{r_G}\right)^{-k_2} [m^{-2}] \quad (2.2)$$

where  $k_1 \cong 1.4$ ,  $k_2 \cong 1.0$ ,  $r_G \cong 220m$ . The muon content in the shower is strongly dependent on the primary composition.

Since full-coverage array can provide more accurate geometric information of air showers, the secondary particles are mainly measured by WCDA in the second hybrid detection stage. It means that the shower core position in this stage is outside of the ED array. Therefore, ED is not included in this simulation.



**Figure 2:** The energy distributions of all simulation events after normalization (left) and after hybrid event selection(right). Selection efficiency is about 15%.

### 2.3 Event Selection

There are three basic principles for events selection: firstly, the shower core position falls in WCDA; secondly, the Cherenkov image is complete; thirdly, the muon lateral distribution is well fitted. In detail, the reconstructed shower core position located in  $130m \times 130m$  around WCDA is selected. For a Cherenkov image, the number of fired pixels are more than 10; and the angular distance between the weighted center of the image and the center of the camera plane is less than  $6^\circ$ . For muon lateral distribution fitting, the fitted normalized muon size ( $k_G N_\mu$ ) should be more than  $10^{-7}$ .

The number of events in each energy bin after selection is shown in Fig. 2 (right). The dotted line represents the number of proton events involved in the simulation, which is consistent with the black line in Fig. 2 (left). The solid lines represent the number of events after selection, and different colors represent different components. It points out that the hybrid observation becomes nearly fully efficient above 100 TeV for all components.

### 3. Proton and Helium Event Selection

#### 3.1 Component Sensitive Parameters

Four types of detectors can detect electromagnetic particles, muons, Cherenkov photons in the EAS. The lateral and longitudinal distributions of EAS can be observed by hybrid observation. Thus many kinds of parameters can be used for primary nuclei identification.

It is known that the atmosphere depth of shower maximum,  $X_{max}$ , is a traditional mass sensitive parameter. The proton initiated showers develop to its maximum later than one of iron. Here,  $\Delta\theta$ , the exactly parameter used to reconstruct the  $X_{max}$ , is applied instead of the reconstructed  $X_{max}$ .  $\Delta\theta$  is the angular distance between the shower arriving direction and the gravity center of the Cherenkov image. But it can not be used directly to classify primary particles because of the  $R_p$  and shower energy ( $N_0^{pe}$ ) dependence. After normalization, the structure of the mass sensitive parameter  $P_X$  is as follows:  $P_X = \Delta\theta - 0.0103 \times R_p - 0.404 \times N_0^{pe}$ .

Moreover, the proton-induced showers exhibit an elongated elliptic shape [16]. And the ratio of length and width of the Cherenkov image is also a traditional and effective parameter:  $P_C = L/W - 0.0137 \times R_p + 0.239 \times N_0^{pe}$ .

The muon size  $N_\mu$  of a shower heavily depends on the atomic number ( $A$ ) of the primary particle:  $N_\mu^A/N_\mu^p \approx A^{(1-\eta)}$ , where  $\eta$  is approximately 0.9 [5] and  $p$  indicates the proton shower. Therefore, the muon size by fitting ( $N_\mu^F$ ), the total number of detected muons ( $N_\mu^M$ ) and the number of fired MDs (NMD) are significant variables to identify the primary particles. After geometric ( $R/p$ , the perpendicular distance between the center of MD array and the shower axis) and energy ( $N_0^{pe}$ ) correction, three component sensitive variables,  $P_{\mu 1}$ ,  $P_{\mu 2}$  and  $P_{\mu 3}$ , can be constructed.

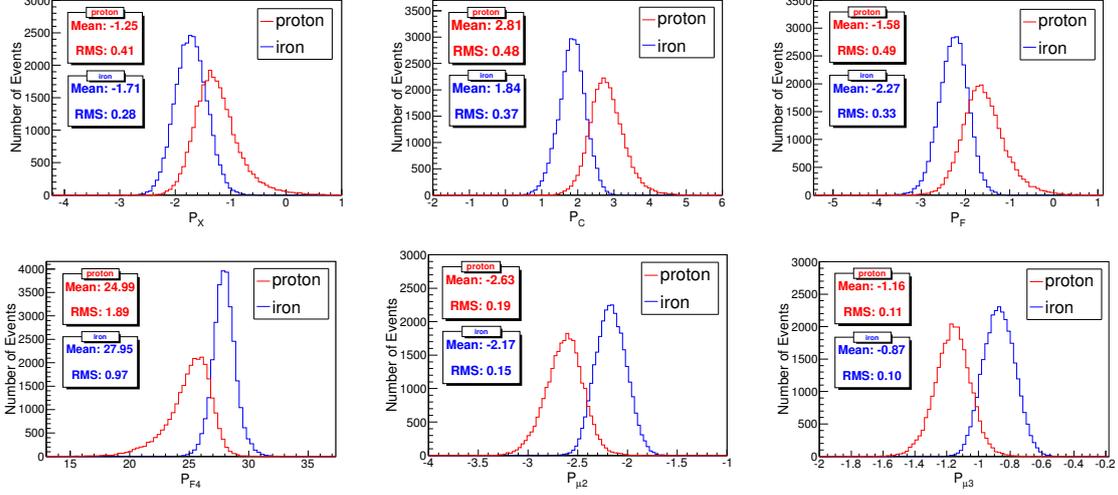
Furthermore, nearby the shower core axis, the particle density of iron-shower is much less than that of proton-shower. Hence the photoelectrons in the brightest cell ( $N_{max}$ ) measured by WCDA is sensitive to components. Because iron-shower is more extensive than proton-shower in the same observation level, the average lateral distribution ( $\langle ER \rangle$ ) and fluctuation ( $RMS$ ) of the iron-shower is larger.

$$\langle ER \rangle = \frac{\sum R_i \times Pe_i}{\sum Pe_i}, \quad RMS = \sqrt{\frac{\sum R_i^2 \times Pe_i}{\sum Pe_i} - \langle ER \rangle^2} \quad (3.1)$$

where  $R_i$  is the distance between the shower core position and the  $i$ th fired cell in WCDA.  $Pe_i$  is photoelectronic measured by the  $i$ th fired cell. After zenith ( $\theta_{rec}$ ) and energy ( $N_0^{pe}$ ) correction, component sensitive variables,  $P_F$ ,  $P_{F2}$  and  $P_{F4}$ , can be constructed.

Other parameters have also been studied, such as the total photoelectrons located 45 m away from the shower core in the WCDA ( $P_{F1}$ ) and etc. But the particle classification of these variables is weak. After tuning of parameters, six parameters,  $P_C$ ,  $P_X$ ,  $P_F$ ,  $P_{F4}$ ,  $P_{\mu 2}$  and  $P_{\mu 3}$ , are selected due

to the effective particle identification and weak correlation. The distributions of these parameters for proton and iron showers are shown in Fig. 3.



**Figure 3:** The distributions of six mass sensitive parameters,  $P_C$ ,  $P_X$ ,  $P_F$ ,  $P_{F4}$ ,  $P_{\mu 2}$  and  $P_{\mu 3}$  for proton (red line) and iron (blue) initiated showers.

### 3.2 TMVA analysis

Drawing on the experience of ARGO-YBJ & WFCT hybrid detection, two parameters are used for particle identification [17] in the event-by-event cut. However, too many variables are no longer suitable for particle identification through event-by-event cut because it is complicated to get the optimal cut value of each parameter. Therefore, the Boosted Decision Trees (BDTG) in Toolkit for Multivariate Analysis (TMVA) [18, 19] is taken.

The six variables mentioned above can be used as input variables. After training and testing, an output variable can be obtained, which can be used to select the target events. The classification of  $H$  and  $H&He$  from other mass components is carried out independently. The selected hybrid events are divided into two equal parts. One is used for TMVA training; the other is used as data.

The training results are shown in Fig. 4.  $H&He$  can be well separated from other components; However, separation of proton from other nuclei is barely satisfactory mainly due to the helium contamination.

The training results are applied to the other half of the data sample. After the cut described above, the aperture and contamination of hybrid observation are calculated. The aperture of pure proton is about  $900m^2Sr$  and the aperture of  $H&He$  is about  $1800m^2Sr$ , as shown in Fig. 5 (left). The contamination of proton sample is about 10% and the contamination of  $H&He$  sample is less than 3% according to the Hörandel model [20], as shown in Fig. 5 (right).

Considering the uncertainty due to the hadronic interaction models, there are two batches of data, QGSJET-FLUKA and EPOS-FLUKA, were used for error analysis. It is found that the aperture of  $H&He$  is consistent within  $\pm 5\%$ .

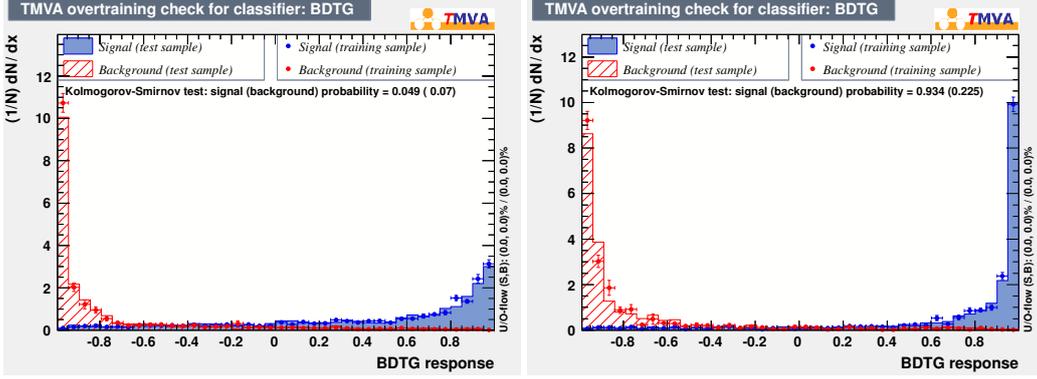


Figure 4: Training results of BDTG classifier for proton (left) and H&He (right).

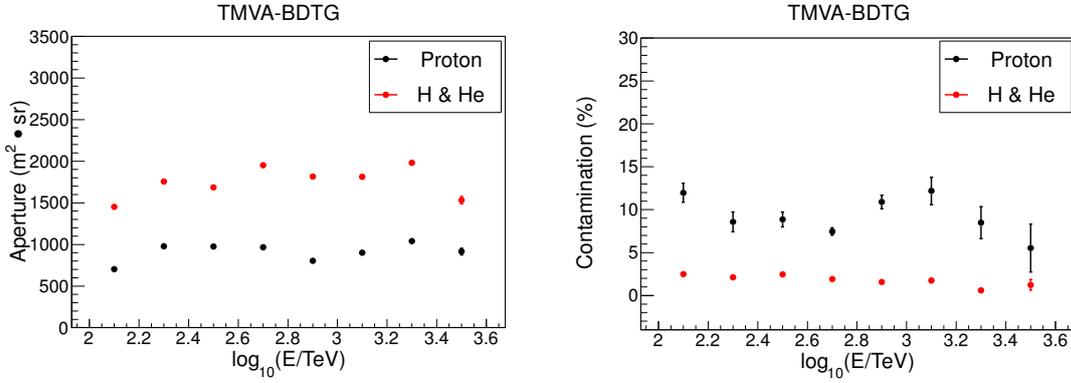


Figure 5: The final selected aperture (left) and contamination (right) for proton and H&He.

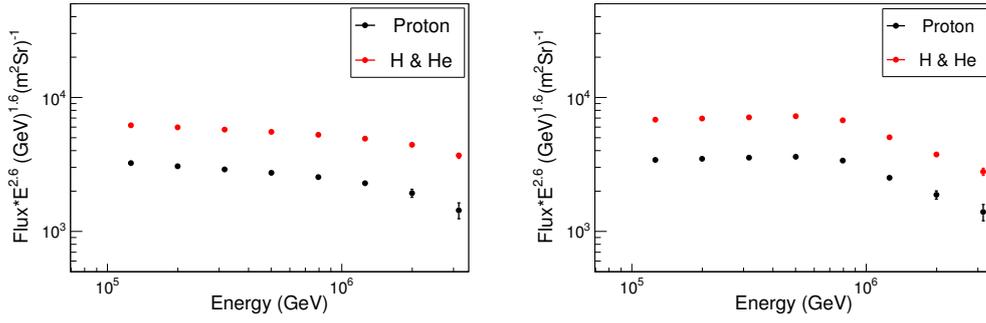
#### 4. Proton and Helium Spectrum Expectation of LHAASO

Based on the aperture described above, the cosmic ray spectra of proton and H&He are predicted according to the Hörandel model [20] and the ARGO-YBJ & WFCT model [21] which is an experimental model. The exposure time is set to one year with 10% duty cycle. Since the experimental model ARGO-YBJ & WFCT only shows the energy spectrum of H&He, the ratio of proton and helium is 1:1. The bending of the knee keeps unchanged.

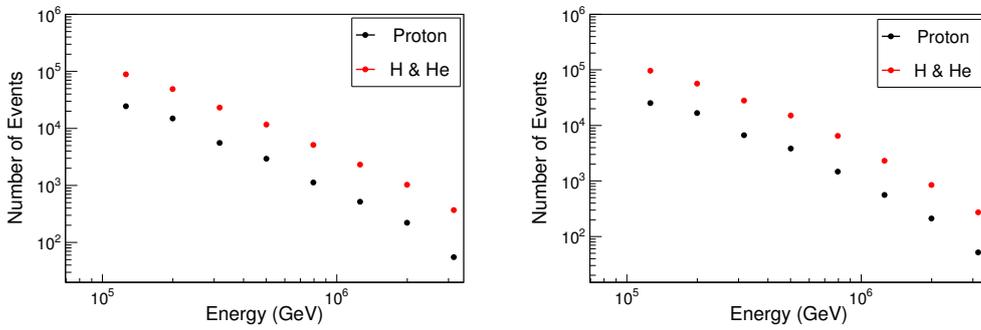
As shown in Fig. 6, the left panel shows the spectrum with the Hörandel model; and the right panel shows the spectrum with ARGO-YBJ & WFCT model. The black dots show proton and the red dots show *H&He*. The statistical error are very small with the aperture and effective time presented above. The corresponding event rate per year is shown in Fig. 7.

At 800 TeV, 5121/6461 H&He events and 1130/1476 proton events can be selected in one year according to the Hörandel model and the ARGO-YBJ & WFCT model respectively. At 2 PeV, 1023/849 H&He events and 222/210 proton events can be selected in one year according to the Hörandel model and the ARGO-YBJ & WFCT model respectively.

Therefore, if the knee of the cosmic ray light component is below 1 PeV, the 1/4 LHAASO array can give good measurements within one year. If the knee of light component is higher than 3 PeV, more observation time or more effective methods is needed to get enough statistics. Consid-



**Figure 6:** The prospect of spectra with the Hörandel model (left) and the ARGO-YBJ & WFCT model (right).



**Figure 7:** The statistics of one year with 10% duty cycle at the Hörandel model (left) and the ARGO-YBJ & WFCT model (right).

ering the application of SiPM, it allows observation in the moon night of WFCTA. The duty cycle of the hybrid observation can be extended. Moreover, during the construction of LHAASO, more Cherenkov telescopes can also increase the statistics effectively .

## 5. Summary

The simulation for the second stage of 1/4 LHAASO hybrid detection is operated. Three types of detectors, WFCTA, WCDA and MD, are included in this study. Parameters from each detector array are studied and tuned in detail. After removing the highly correlated parameters, six parameters are used as input of the BDTG classifier for TMVA machine training and testing. The results show that the classification of pure proton is weaker than that of H&He.

After cutting, high purity light component samples were selected. The contamination for pure proton and mixed proton and helium samples are 90% and 96% respectively. The apertures of pure proton and mixed proton and helium samples are  $900 \text{ m}^2 \cdot \text{Sr}$  and  $1800 \text{ m}^2 \cdot \text{Sr}$  respectively. The uncertainty of strong interaction models for proton and helium selection is less than  $\pm 5\%$ .

According to the energy spectrum expectation of the ARGO-YBJ & WFCT model, the spectrum of light component in cosmic rays will be measured accurately in a short time by 1/4 LHAASO array.

## 6. Acknowledgements

This work is supported by the National Key R&D Program of China (NO. 2018YFA0404201 and NO. 2018YFA0404202). This work is also supported by the Key Laboratory of Particle Astrophysics, Institute of High Energy Physics, CAS. Projects No. Y5113D005C, No. 11563004 and No. 11775248 of National Natural Science Foundation (NSFC) also provide support to this study.

## References

- [1] Johannes Blümer, Ralph Engel, Jörg R.Hörandel, *Progress in Particle and Nuclear Physics* 63(2009) 293-338.
- [2] Z. Cao (for LHAASO Coll.), *Chin.Phys. C*, 2010, (34)249-252.
- [3] Z. Cao (for LHAASO Coll.), 31th ICRC, 2009.
- [4] Q. An et al.(for LHAASO Coll.), *Nucl.Instrum.Meth. A*, 2013, 724: 12-19.
- [5] Peter K.F.Grieder, ISBN978-3-540-76940-8(2010).
- [6] S.S.Zhang et al., *Nucl.Instrum.Meth. A*, 2012, 629(1)57-65.
- [7] Baiyang Bi et al.(for LHAASO Coll.), *Nucl.Instrum.Meth. A*, 899 (2018) 94-100.
- [8] Zhao Jing et al.(for LHAASO Coll.). *Chin.Phys. C*, 2014, 38(3) 036002.
- [9] X. Zuo et al.(for LHAASO Coll.), *Nucl.Instrum.Meth. A*, 2015, 789: 143.
- [10] B. Bartoli et al.(for ARGO-YBJ Coll.), *Astrop.Phys.*, 90 (2017) 20-27.
- [11] L.L. Ma, 33th ICRC, 2013, p.1013.
- [12] J.L. Liu, 33th ICRC, 2013, p.0880.
- [13] D. Heck and T. Pierog, *Extensive Air Shower Simulation with CORSIKA: A User's Guide*, 2015.
- [14] Yu.A. Fomin et al., *Nuclear Physics B (Proc. Suppl.)* 175-176 (2008) 334-337.
- [15] Xiurong Li et al. (for LHAASO Coll.), 35th ICRC, 2017, p.381.
- [16] Hillas, A. 1985, 19th ICRC Vol.3, 445-448.
- [17] B. Bi, Z. Cao et al.(for LHAASO Coll.), 2016, *Frascati Physics Series Vol. 64*.
- [18] A. Hoecker, P. Speckmayer et al, *TMVA Users Guide*, 2013, arXiv:physics/0703039.
- [19] P Speckmayer, A Höcker et al., *Journal of Physics: Conference Series* 219 (2010) 032057.
- [20] Jörg R.Hörandel, *Astroparticle Physics*, 2003, 19(2): 193-220.
- [21] B. Bartoli et al.(for ARGO-YBJ and LHAASO Coll.), *Phys. Rev. D*, 2015, 92(9): 25-31.