PROCEEDINGS OF SCIENCE

ONLINE ICRC 2021
THE ASTROPARTICLE PHYSICS CONFERENCE
Berlin | Germany
37th International Cosmic Ray Conference
12–23 July 2021

Vetoing the high energy showers in the GRAPES-3 experiment whose cores lie outside the array

M. Chakraborty,[a,*] S. Ahmad,[c] A. Chandra,[c] S.R. Dugad,[a] U.D. Goswami,[l] S.K. Gupta,[a] B. Hariharan,[a] Y. Hayashi,[b] P. Jagadeesan,[a] A. Jain,[a] P. Jain,[d] S. Kawakami,[b] H. Kojima,[e] S. Mahapatra,[i] P.K. Mohanty,[a] R. Moharana,[j] Y. Muraki,[g] P.K. Nayak,[a] T. Nonaka,[h] A. Oshima,[e] B.P. Pant,[j] D. Pattanaik,[a,i] G.S. Pradhan,[k] P.S. Rakshe,[a] M. Rameez,[a] K. Ramesh,[a] L.V. Reddy,[a] R. Sahoo,[k] R. Scaria,[k] S. Shibata,[e] J. Soni,[d] K. Tanaka,[f] F. Varsi[d] and M. Zuberi[a] [The GRAPES-3 Collaboration]

[a] *Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400005, India*
[b] *Graduate School of Science, Osaka City University, Osaka 558-8585, Japan*
[c] *Aligarh Muslim University, Aligarh 202002, India*
[d] *Indian Institute of Technology Kanpur, Kanpur 208016, India*
[e] *College of Engineering, Chubu University, Kasugai, Aichi 487-8501, Japan*
[f] *Graduate School of Information Sciences, Hiroshima City University, Hiroshima 731-3194, Japan*
[g] *Institute for Space-Earth Environmental Research, Nagoya University, Nagoya 464-8601, Japan*
[h] *Institute for Cosmic Ray Research, Tokyo University, Kashiwa, Chiba 277-8582, Japan*
[i] *Utkal University, Bhubaneshwar 751004, India*
[j] *Indian Institute of Technology Jodhpur, Jodhpur 342037, India*
[k] *Indian Institute of Technology Indore, Indore 453552, India*
[l] *Dibrugarh University, Dibrugarh 786004, India*

E-mail: 10medha.riya@gmail.com, pkm@tifr.res.in

The GRAPES-3 experiment located in Ooty consists of an array of 400 plastic scintillator detectors spread over an area of 25000m^2 and a large area (560 m^2) muon telescope. Every day, the array records about 3 million showers induced by the interaction of primary cosmic rays in the atmosphere. One of the primary objectives of the experiment is to measure the energy spectrum and composition of the cosmic rays in the TeV-PeV energy range. However, some of the detected showers have cores outside the array. This fraction increases with energy due to the higher lateral spread of shower particles at higher energies. Identifying these events is thus crucial for accurate measurement of the cosmic ray energy spectrum. This work will describe simple cut based as well as machine learning based strategies for identifying and excluding such events and their impact on the cosmic ray energy spectrum as measured by the Bayesian unfolding technique.

**** 37th International Cosmic Ray Conference (ICRC2021), ****
**** 12-23 July 2021 ****
**** Berlin, Germany - Online ****

*Presenter

PoS(ICRC2021)394

1. Introduction

The acceleration and propagation mechanism of cosmic rays is a century-old mystery and one of the ways to probe it is by the study of cosmic ray (CR) energy spectrum. However, there are systematic effects hampering the precise measurement of the energy spectrum. Figure 1 shows the unfolded spectrum energy spectrum (unfolded by the Bayesian unfolding method using the ROOUNFOLD package [1]) deviating from the input energy spectrum generated from simulation. This is because high energy showers landing hundreds of metres away from the array also trigger the array, and such showers often have their reconstructed cores inside the array due to improper reconstruction. This work demonstrates methods to remove such contaminated showers by using simple cut based and machine learning strategies, and the subsequent improvements in energy spectrum measurements caused by these methods.

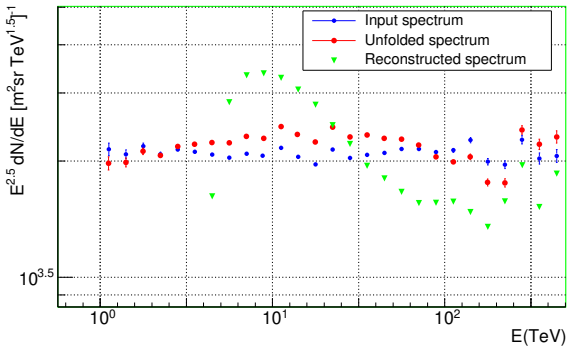


Figure 1: The unfolded spectrum shows some deviation from the expected true spectrum of CORSIKA simulated showers with spectral index -2.5, the showers were thrown in a circular radius of 150 m

2. GRAPES-3 experiment

The GRAPES-3 (Gamma Ray Astronomy at PeV Energies Phase-3) experiment is located at Ooty (11.4°N, 76.7°E, 2200 m a.s.l.), India. The GRAPES-3 extensive air shower (EAS) array consists of 400 plastic scintillator detectors. Each of these detectors records particle densities and relative arrival times of particles in an air shower [2]. The scintillator array covers an area of 25000 m^2. GRAPES-3 uses two level trigger, level-0 trigger is a simple 3-line coincidence in 100 ns time window and level-1 trigger requires at least 10 detectors hit in 1 µs time window. The scintillator detectors are arranged in hexagonal geometry, with an inter-detector separation of 8 m. GRAPES-3 also has 560 m^2 tracking muon detector consists of 3712 proportional counters (PRCs) [3]. A schematic of GRAPES-3 with the fiducial area (14560 m^2) is shown in Figure 2.

3. MC simulations

Simulated air shower data is produced by using CORSIKA (version 7.69) simulation package. 5×10^8 air showers with proton primaries and spectral index -2.5, having energies from 1 TeV to 10 PeV were simulated using hadronic interaction generators SIBYLL-FLUKA. The showers are

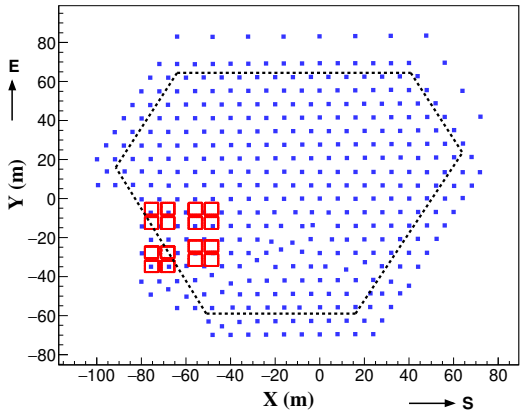


Figure 2: Schematic of GRAPES-3 air shower array, the area enclosed by the black dashed line marks the fiducial area

thrown within a circular area of radius 'r' such that the trigger fraction is less than 1% outside this circle. The value of 'r' is within 300-500 m for showers having energies of a few hundred TeV, and within 500-800 m for showers with energies above a PeV. The detector response of these simulated showers is then calculated using GEANT-4, followed by shower reconstruction similar to data in order to obtain the shower parameters.

4. Reconstruction of shower

The particle densities recorded in the detectors are fitted by the well known Nishimura-Kamata-Greisen (NKG) function to obtain the shower parameters as described in [4], namely, the shower size (N_e), age (s) and the shower core (X_c, Y_c), as shown in Figure 3.

$$\rho_i = \frac{N_e}{2\pi r_m^2} \frac{\Gamma(4.5-s)}{\Gamma(s)\Gamma(4.5-2s)} (\frac{r_i}{r_m})(1+\frac{r_i}{r_m})^{s-4.5} \tag{1}$$

where ρ_i is the expected particle density at i^{th} detector, r_i is the distance of i^{th} detector from shower core. r_m is the Moliere radius which is 103 m at Ooty.

The NKG fit is performed by negative log-likelihood minimisation. For an expected density of ρ_i, the probability p_i of detecting n_i particles in the i-th detector can be expressed by a Poisson distribution,

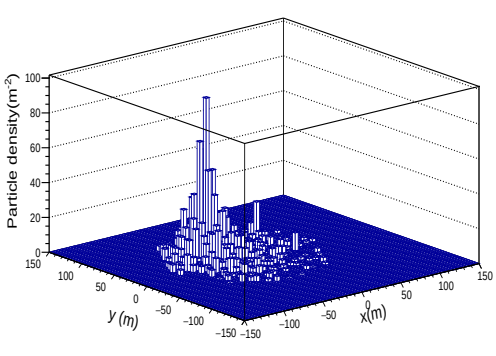
$$p_i = \frac{(\rho_i A_i cos\theta)^{n_i}}{n_i!} e^{-\rho_i A_i cos\theta} \tag{2}$$

where A_i is the detector area which is 1 m^2.

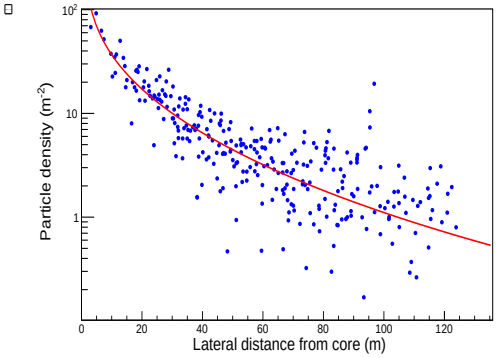
$$L = \prod_i p_i \tag{3}$$

Calculation of products is computationally intensive. Hence, the product is converted to a summation by taking its natural logarithm.

$$ln(L) = \sum_i (n_i \times ln(\rho_i cos\theta) - \rho_i cos\theta - ln(n_i!)) \tag{4}$$

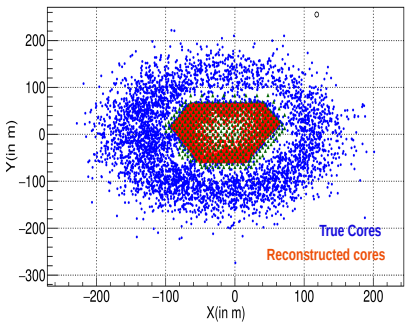


(a) The particle densities recorded in an air shower

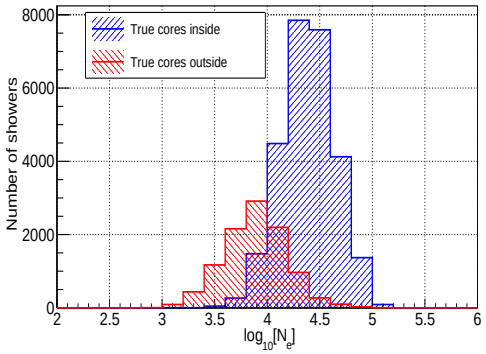


(b) The NKG fit of the lateral density profile of the shower

Figure 3: The shower profile (left) and its lateral distribution fitted by NKG function (right). The core is around the region where the highest particle densities are recorded.



(a) The true cores and their respective reconstructed cores, that contribute to contamination



(b) The shower size distribution for showers with true cores outside and inside

Figure 4: Showers with true cores outside but mis-reconstructed cores inside having energies within 100 TeV $\leq E <$ 158 TeV (left), and the shower size distributions for showers with the same energy are shown for well and mis-reconstructed showers (right). The mis-reconstructed showers have lower shower size.

$-ln(L)$ is minimised in order to fit and obtain the shower parameters as described in [5].

5. Selection criteria

The following shower selection criteria are applied,

- Showers with successful NKG fit
- Zenith angle is restricted to 25°
- Showers with reconstructed cores within the fiducial area.

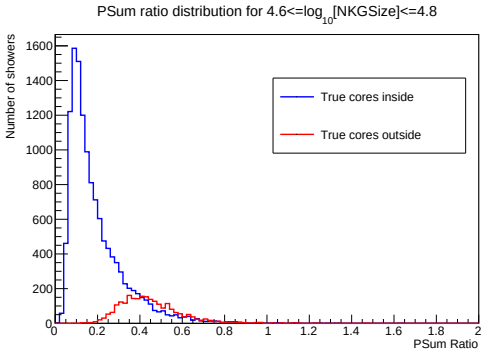
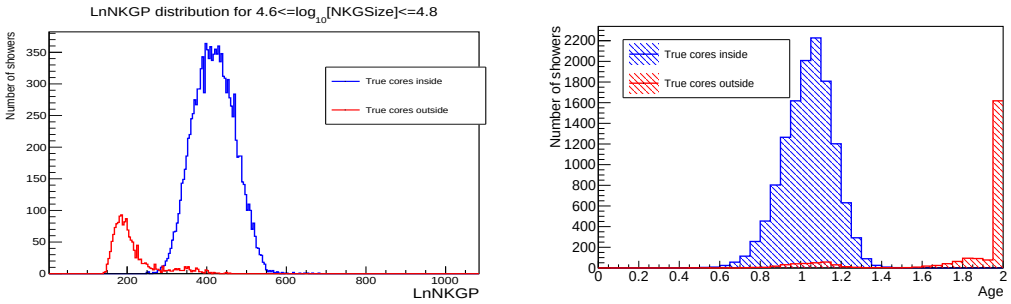


Figure 5: Particle sum ratio for showers with true cores outside is higher



(a) LnNKGP (b) Age

Figure 6: LnNKGP and Age distributions for shower size $4.6 \leq log_{10}[N_e] < 4.8$, both the variables show a good distinction between well and mis-reconstructed showers

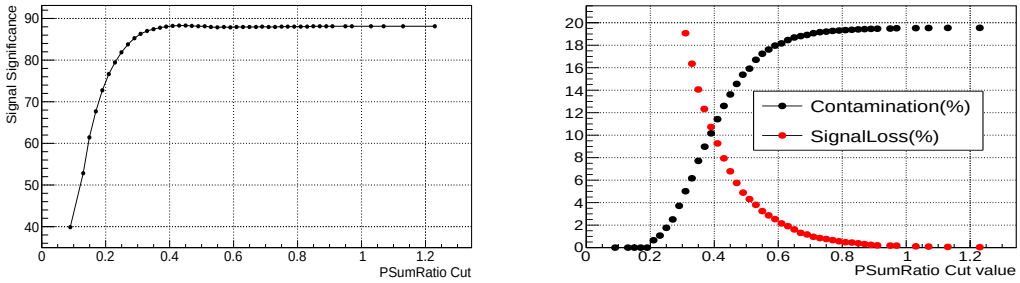
6. Analysis

The showers that have their true cores outside, but reconstructed cores inside (Figure 4(a)) have lower shower size due to a partial shower front lying within the array, and hence they are interpreted as lower energy showers by the array as shown in Figure 4(b). Variables that can distinguish well-reconstructed and mis-reconstructed showers are selected and cuts are developed as a function of shower size. The variables selected are as follows:

- Particle sum ratio (PSumRatio): It is defined as the ratio between the total particle density collected outside the fiducial area and the total particle density collected within the fiducial area. This ratio is higher when the true cores lie outside the array as shown in Figure 5.

- Quality of NKG fit: The mis-reconstructed showers have a poor NKG reconstruction. Thus, the best value of the log-likelihood function used for NKG fit as described in section 4, denoted by LnNKGP in Figure 6(a) serves as a distinguishing variable.

- Age: Age is mostly very high in the case of improper core reconstruction as shown in Figure 6(b).



(a) Significance for PSumRatio

(b) Contamination and signal loss

Figure 7: Decision of cut for PSumRatio for $4.6 \leq log_{10}[N_e] < 4.8$, cut was applied at 0.46. This repeated for all other variables in all other shower size bins

- Age error: The error in age parameter is usually greater than 0.5 in the case of mis-reconstruction.

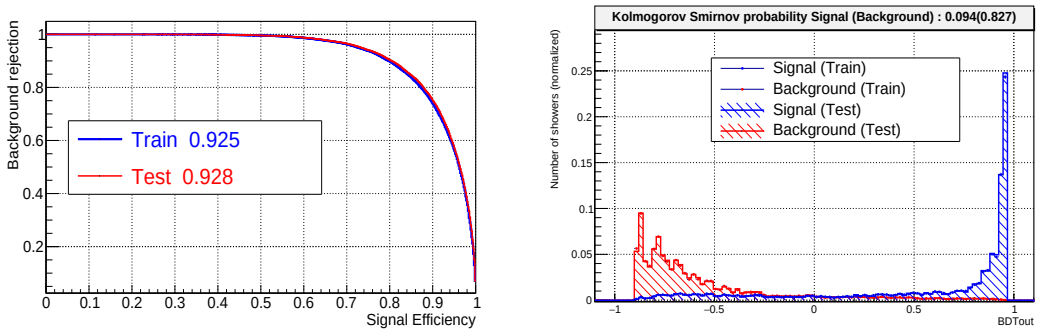
- χ^2: χ^2 of the planar fit performed on the relative arrival times of the air shower in order to obtain the shower direction.

6.1 Manual cuts

The guiding properties to decide cuts on the above mentioned variables are the signal significance, the fraction of mis-reconstructed showers reduced, and well reconstructed showers lost due to a particular cut value. Cuts are chosen at the value where the signal significance, given by $S/\sqrt{S+B}$, attains its maximum. Here, S is the number of showers having both true and reconstructed cores inside the array and the B is the number of showers having their true cores outside but reconstructed cores within the array. The % reduction in background and the % signal loss, after applying the chosen cut, are then evaluated to test the performance of the cut. If the % decrease in background is less than or comparable to the % signal loss, then the cut is loosened or not used. This is repeated for all other shower size bins using all the variables. As an example, in Figure 7, the decision of cut for PSumRatio for shower size $4.6 \leq log_{10}[N_e] < 4.8$ has been shown.

6.2 Machine learning

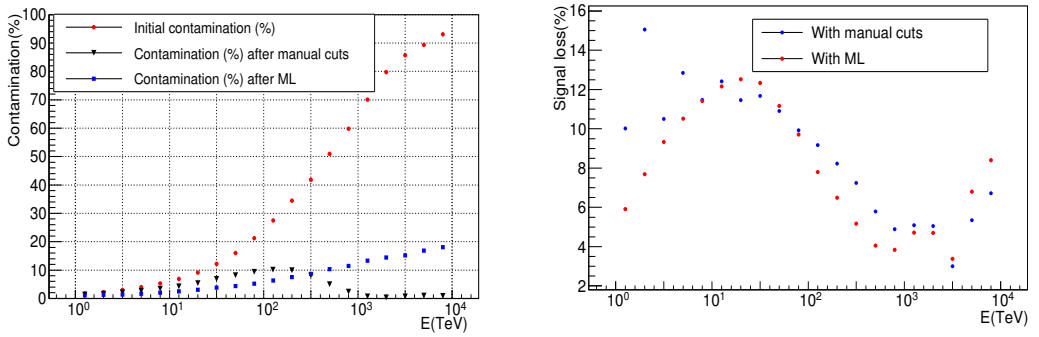
A gradient boosted decision tree (BDT) from the TMVA package [6] is used to separate the mis-reconstructed showers from well reconstructed showers. The simulated dataset is split into two statistically independent parts, one is used to train and the other is used to test the performance of the BDT. The hyperparameters and input variables were optimized to ensure that the integral of receiver-operator-characteristic curves (ROC) for train dataset is close to 1. In order to avoid possible overtraining, we ensure that the Kolmogorov-Smirnov (KS) probability between train and test distributions of the BDT output variable is above 0.05 . We also compare the performance of the BDT by plotting the ROC for the train and test datasets, as shown in Figure 8.



(a) The area under ROC curve matches well for train and test datasets

(b) The results of KS test between test and train samples, for both signal and background datasets

Figure 8: ROC and KS test results of BDT output variable for $4.6 \leq log_{10}[N_e] < 4.8$



(a) Contamination

(b) Signal loss

Figure 9: Contamination and signal loss (in %) after applying the cuts with both the methods (Preliminary)

7. Results and discussion

After applying the cuts, we calculate the contamination reduction and the resulting signal loss with true energy. The preliminary results are shown in Figure 9. The contamination increases with increase in energy due to the lateral spread of the shower but is largely reduced with both the methods with a corresponding signal loss with 20%. At higher energies, the manual cuts are seen to perform better and its impact on energy spectrum measurements was studied. After applying the manual cuts, the deviation in the unfolded energy spectrum from the input spectrum was largely reduced as shown in Figure 10. At 50 TeV, the initial deviation before applying any cuts is around 15% from the input spectrum which is reduced to around 3% after applying manual cuts. Further analysis is in progress to improve the results of this analysis in order to perform precise measurements of the CR energy spectrum.

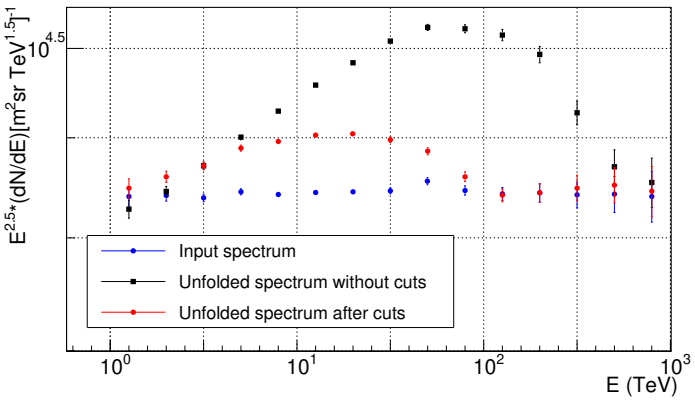


Figure 10: Energy spectrum before and after applying the cuts (Preliminary). The deviations are seen to reduce significantly

8. Acknowledgement

We are grateful to D.B. Arjunan, G.P. Francis, V. Jeyakumar, S. Kingston, K. Manjunath, S. Murugapandian, B. Rajesh, K. Ramadass, V. Santoshkumar, M.S. Shareef, C. Shobana, R. Sureshkumar for their role in the efficient running of the experiment. We would also like to thank Prof. Gagan Mohanty for his lectures on machine learning to GRAPES-3 members and Rahul Tiwary (TIFR) for his help with BDT.

References

[1] https://hepunx.rl.ac.uk/ adye/software/unfold/RooUnfold.html

[2] S.K. Gupta et al., Nucl. Instr. Meth. A 540 (2005) 311

[3] Y. Hayashi et al., Nucl. Instr. Meth. Phys. A 545 (2005) 643

[4] H. Tanaka et al 2012 J. Phys. G: Nucl. Part. Phys. 39 025201

[5] B. HariHaran et. al., Exp Astron 50, 185–198 (2020)

[6] http://tmva.sourceforge.net/docu/TMVAUsersGuide.pdf