PROCEEDINGS OF SCIENCE

Carbon Flux with DAMPE Using Machine Learning Methods

Mikhail Stolpovskiy,[a,*] **Andrii Tykhonov,**[a] **Arshia Ruina,**[a] **Paul Coppin**[a] **and David Droz**[a] **on behalf of the DAMPE collaboration**

[a] *Department of Nuclear and Particle Physics, University of Geneva, CH-1211 Geneva, Switzerland*

E-mail: mikhail.stolpovskiy@unige.ch

DAMPE space-borne cosmic ray experiment has been collecting data since December 2015. Many high-impact results on the ion, electron and photon fluxes were obtained. This submission presents the carbon flux analysis with DAMPE using machine learning techniques. The readout electronics would saturate at energy deposits above several TeV in a single BGO bar of the DAMPE calorimeter. The total energy loss per event due to saturation can sometimes reach over a hundred TeV. We present a convolutional neural network model which can accurately recover the energy lost due to saturation and thus significantly increase the dynamic range of DAMPE. Another machine learning model combines the resolution of the hodoscopic BGO calorimeter and the high-resolution tracker of DAMPE to provide the best possible prediction of the direction of the incoming particle. This allows measuring charges at energies up to several hundred TeV. In this work, we present the application of these methods to carbon flux analysis.

38th International Cosmic Ray Conference (ICRC2023)
26 July - 3 August, 2023
Nagoya, Japan

ICRC2023
38th International Cosmic Ray Conference
The Astroparticle Physics Conference

*Speaker

PoS(ICRC2023)168

1. Introduction

Cosmic rays, high-energy particles originating from various astrophysical sources, hold significant importance in astroparticle physics. The DAMPE mission [1, 2], launched in December 2015, aims to explore these particles with exceptional precision and to shed light on fundamental astrophysical questions of origin, acceleration and propagation of the cosmic rays. Among the latest DAMPE results are the measurements of the proton [3] and helium [4] spectra, as well as the electron flux [6], boron-to-carbon flux ratio measurement [5] and many others [7–9]. DAMPE's key components include the Plastic Scintillator Detector (PSD) for charge identification [10, 11], the Silicon Tracker (STK) for precise tracking [12, 13], the BGO Imaging Calorimeter (BGO) for energy measurements and gamma-ray identification [14, 15], and the Neutron Detector (NUD) for providing additional electron/hadron discrimination [16]. DAMPE's exceptional capabilities, including its fine-segmented thick calorimeter and large acceptance, enable the detection of cosmic rays with remarkable energy resolution. Specifically, DAMPE achieves an energy resolution of $\sim$1% for electrons and gamma rays, and 30% for protons and ions.

To enhance measurement accuracy, DAMPE incorporates advanced machine learning techniques. A machine learning algorithm is developed to correct saturation effects observed in the BGO calorimeter [17]. This algorithm accurately identifies and corrects for saturation using data from the BGO calorimeter, providing more precise energy measurements and improving cosmic ray flux determination. Additionally, a machine learning-based track reconstruction algorithm optimizes track measurements in the STK [20]. Leveraging neural networks, this algorithm analyzes BGO information and STK sensor signals to estimate particle direction with improved spatial resolution, enhancing charge measurements and event selection for flux determination.

Integration of machine learning not only enhances measurement accuracy but also enables more efficient data analysis and processing, unlocking the full potential of DAMPE's data. The following sections of this paper will delve into the details of these machine learning methods and their impact on measuring the cosmic ray carbon flux.

2. BGO saturation correction

BGO saturation, which occurs when the energy deposit in a single bar of the BGO calorimeter exceeds a certain threshold, poses a challenge in accurately measuring the energy deposited by high-energy cosmic ray particles [17]. In case of an extreme energy deposit in a single bar the readout electronics cannot record the energy in that specific bar, leading to an underestimation of the total energy deposited. An example of an event monitor for a saturated event is shown on figure 1. Typically this happens at energy deposited in one bar above about 3 TeV if we consider the last layer of BGO, where the optical filter transparency are set for the precise measurements of the shower tail for better electron and gamma identification. In middle layers saturation appears starting at energies about 3 times larger.

To address this issue, a machine learning approach utilizing Convolutional Neural Networks (CNNs) trained on helium Monte Carlo (MC) simulations has been employed. The CNN-based BGO saturation correction consists of two distinct models. The first model aims to predict the energy lost due to saturation in the last layer of BGO. This model is only applied when the saturation occurs

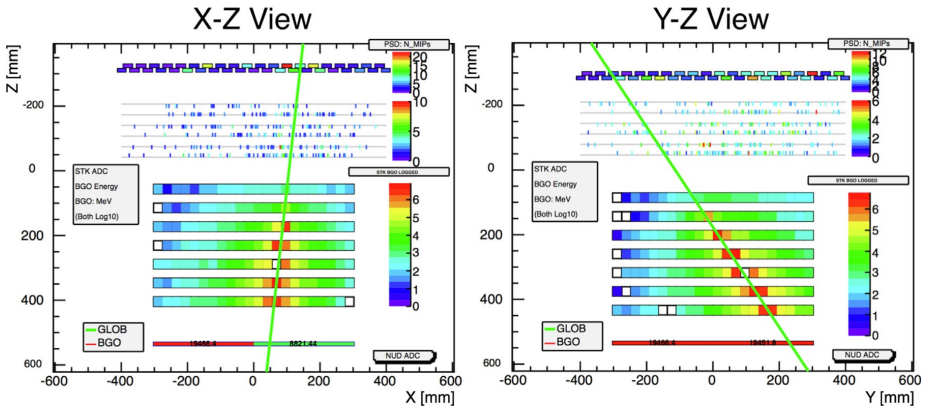

Figure 1: An example of saturated event. Views from two orthogonal directions are shown, corresponding to the coordinate system of DAMPE. The reconstructed direction of the incoming CR is shown with green lines. The sub-detectors are visible: on top the PSD detector, then STK tracker, BGO and NUD on bottom. Color shows the strength of the signal in each channel. The saturated bars are situated right on the shower axis and are shown as zero (white) bars. The off-axis zero bars are not saturated, the signal in them is below the noise level.

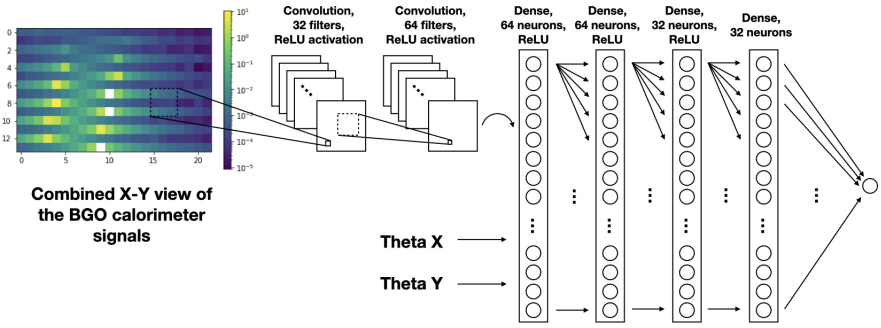

Figure 2: Architecture of the CNN models used in the saturation correction. The last layer applies linear activation.

in a single isolated bar in the last layer, meaning that the saturated bar is not adjacent to any other saturated bars. It is important to separate the last layer saturation to a distinct model, since the last layer saturation happens at lower energies and usually implies energy loss at about 2-4 TeV, while the saturation in the middle bars typically corresponds to ∼3 times larger energy loss. The architecture of the model is shown on figure 2. The model takes as an input the so-called combined view of BGO, where the XZ and YZ view are interlayed. As an additional input we provide the model with the reconstructed shower inclination on both axes (Theta X and Y on the figure 2). The model consists of two convolutional and four dense layers with ReLU activation between them [18]. The second model, with the same architecture, is designed to predict the energy loss per saturated bar in all other layers of the BGO calorimeter.

To account for the use of helium MC in our training, we take steps to unbias the resulting correction when applying it to carbon data. The corrected energy deposition contributes to the regularisation of the unfolding procedure [19], which aims to infer the true energy spectrum

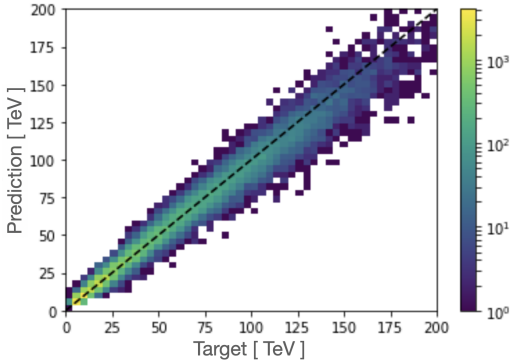

Figure 3: Two-dimentional distribution of reconstructed missing energy versus true missing energy. Diagonal line of *reconstructed* = *true* is shown for reference. Colors correspond to the number of entries in each bin.

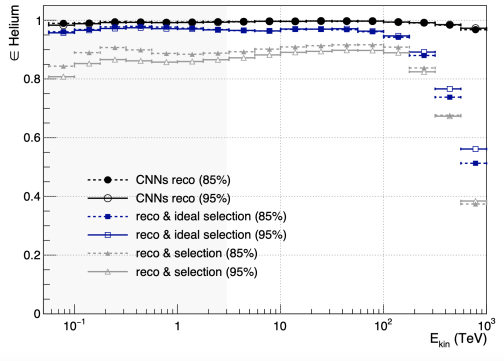

Figure 4: Efficiency of the track reconstruction and identification derived from helium MC as a function of the particle kinetic energy: (circles) the developed CNNs algorithm; (squares) standard track reconstruction with the ideal identification; (triangles) standard track reconstruction with the standard identification.

of cosmic ray carbon nuclei. Thus, this correction plays a vital role in accurately determining the carbon flux at the highest energies, where saturation effects are most pronounced. Figure 3 illustrates the performance of the correction on test sample of helium MC. From this figure one can see that in some cases saturation causes a massive energy loss up to several hundred of TeV, and the correction is able to recover it with high confidence.

3. Machine learning tracking

The limited precision of absolute charge identification poses a significant challenge in direct cosmic-ray detection, as it is closely tied to the accuracy of particle trajectory reconstruction. In addition to the BGO saturation correction, we recently introduced a machine learning technique for track reconstruction in the Silicon Tracker (STK) [20]. Unlike the standard tracking which in a nutshell requires the track to pass through the STK hits [21], the ML tracking approach provides a direction vector based on the full map of STK hits and seeded from the direction provided by the BGO calorimeter, allowing for more accurate determination of the primary particle's trajectory. The seeding BGO direction is estimated with a separate CNN model. The input for STK tracking model is the Hough images of STK [22] where the selected hits are transformed into the lines, see figure 5. Figure 4 illustrates the superior efficiency of the ML tracking approach compared to standard track reconstruction methods. The tracking techniques employed in our study are optimized and trained to accommodate various particles, including carbon, and not solely limited to proton or helium. While we present the tracking plot as an example for helium, it is important to note that the techniques extend to other particle species, such as carbon.

Furthermore, the ML tracking provides a valuable parameter known as the vertex variable. This variable is essentially an output of a classifier model and it takes on values between 0 and 1. A vertex value close to 1 indicates that the primary particle likely reached the first layer of the

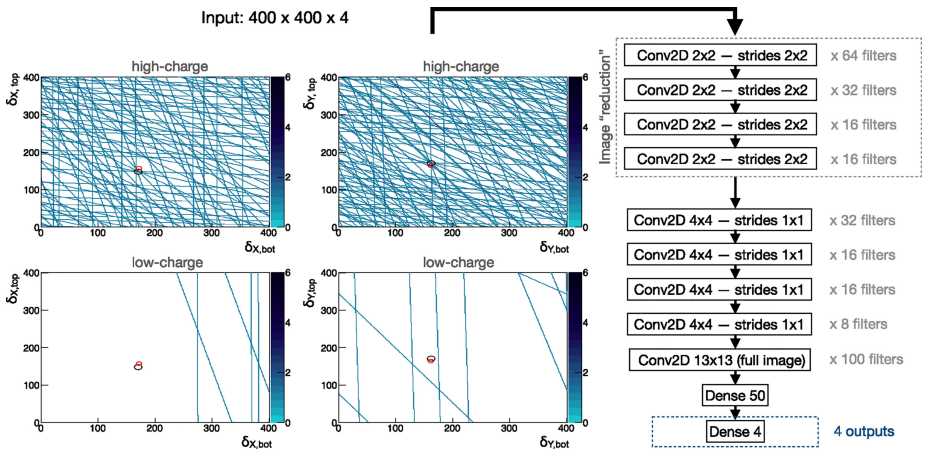

Figure 5: Hough image of a typical Helium event and the architecture of the tracker convolutional neural network. Big (black) and small (red) circles represent the true and the reconstructed trajectory of a primary particle, respectively. See details in [20]

STK without undergoing any inelastic interaction. Conversely, a vertex value close to 0 suggests that a hadronic shower initiated before reaching the STK. By applying a cut on the vertex value, researchers can select events with clearly defined charge in PSD, so that are well-suited for the analysis of ion fluxes.

4. Carbon flux measurement

The BGO correction plays a crucial role in enhancing the energy resolution, particularly at high energies, which is essential for accurately determining the flux. Simultaneously, the ML tracking algorithm aids in identifying carbon candidate events with minimal background contribution, ensuring a high acceptance rate. The synergy between these techniques enables us to measure the carbon flux with improved precision and reliability.

The present analysis is based on DAMPE data from December 2015 until the end of September 2022. The events are selected by requiring the selection by the high-energy trigger [23]. On top of it, the particle direction, reconstructed by the BGO detector, have to be well contained withing the detector volume. Furthermore, the events recorded while DAMPE flies by the South Atlantic Anomaly are excluded. These are the *preselection* criteria used in the current analysis.

Subsequently, the *vertex variable cut*, as discussed in the previous section, is implemented to exclusively select events with a trustworthy charge measurement. This cut significantly reduces the contamination of PSD by back-splashing particles, leading to a notable improvement in charge resolution. To ensure the robust charge measurement we require having at least 2 PSD hits crossed by the chosen ML track consistent between each other within 4 sigma of the carbon peak in PSD charge distribution (note that for a given straight track crossing PSD there must be at least 2 hits, 4 at most) – this is the *PSD N hits cut*. Then the charge for an event is defined as the average over the PSD hits consistent with the first hit.

The selection of carbon candidate events is achieved by applying a window *charge selection cut* on the charge, effectively isolating the desired carbon signal. To estimate the background

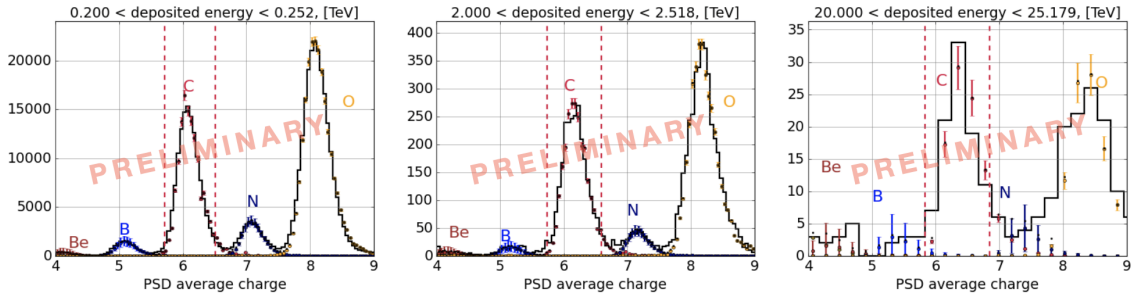

Figure 6: Charge distributions in three different bins of deposited energy (see picture titles). Black histogram shows the flight data events, and colored points show the templates from different ions, obtained with MC simulations. Carbon window cut is highlighted with two vertical crimson lines.

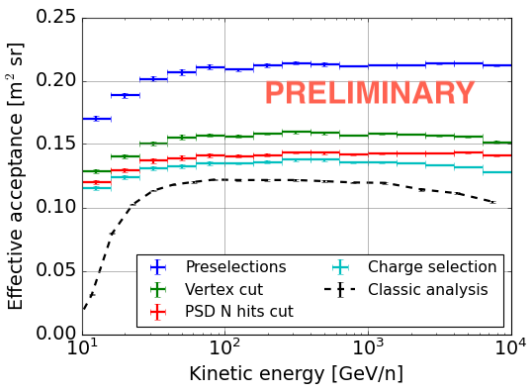

Figure 7: Effective acceptance of the carbon candidate selection. Refer to the accompanying text for detailed explanations of the various selection steps. The acceptance of classic analysis using the standard STK tracks is shown for comparison [24].

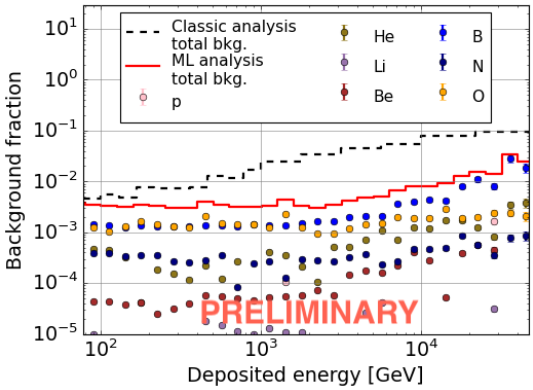

Figure 8: Fraction of background leaking into the carbon selection. Total background is shown with solid red line and doesn't exceed over few percent. The background estimated with the classical analysis is shown for comparison [24].

contribution from other ions, template fits are employed, allowing for a robust characterization of the non-carbon ion contributions in the data. Several examples of the charge distributions fitted with templates obtained from the MC simulations for different ions can be found in figure 6. The effective acceptance of the carbon selection, calculated on carbon MC, is shown in figure 7. The background, estimated by the template fits, is shown on figure 8

The dominant source of systematics in our analysis arises from the vertex cut. To estimate it, we select carbon events based on preselection criteria alone and calculate the efficiency of these events to pass the vertex cut in both flight data and carbon MC samples. However, the efficiency measured in flight data is influenced by the contribution of non-carbon background. To mitigate this bias, we subtract the background, estimated using template fits. We observe a discrepancy of up to 7% between the efficiency of the vertex cut in MC carbon and what is observed in flight data, see figure 9. We hypothesize that this significant uncertainty is, at least in part, a result of the inaccuracies in the hadronic model employed in our MC simulations. Further investigation and a more comprehensive study of this effect are necessary, see in particular [25, 26].

5. Conclusions

In conclusion, the utilization of machine learning (ML) techniques in the DAMPE experiment has proven to be instrumental in various aspects of cosmic ray analysis. The ML tracking approach enables improved track reconstruction, facilitating the identification of carbon candidate events with reduced background contribution and maintaining high acceptance rates. Additionally, the ML models applied for BGO saturation correction enhance the energy resolution, particularly at high energies. The successful application of ML techniques highlights their significance in advancing our understanding of cosmic rays, offering valuable insights into the carbon flux and its implications for astrophysics research.

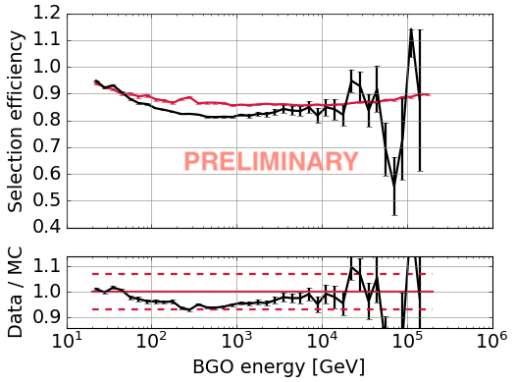

Figure 9: Top: efficiency of the vertex selection, measured on the flight data (black points) and carbon MC (crimson), as a function of deposited (BGO) energy. Bottom: ratio of efficiencies. The estimated vertex systematics is based on the deviation of this ratio from 1.

6. Acknowledgements

The DAMPE mission was funded by the strategic priority science and technology projects in space science of Chinese Academy of Sciences (CAS). In China, the data analysis was supported by the National Key Research and Development Program of China (No. 2022YFF0503302) and the National Natural Science Foundation of China (Nos. 12220101003, 11921003, 11903084, 12003076 and 12022503), the CAS Project for Young Scientists in Basic Research (No. YSBR061), the Youth Innovation Promotion Association of CAS, the Young Elite Scientists Sponsorship Program by CAST (No. YESS20220197), and the Program for Innovative Talents and Entrepreneur in Jiangsu. In Europe, the activities and data analysis are supported by the Swiss National Science Foundation (SNSF), Switzerland, the National Institute for Nuclear Physics (INFN), Italy, and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (No. 851103).

References

[1] J. Chang, Chin. J. Space Sci. 34 (5) (2014) 550–557.

[2] J. Chang, et al., Phys. 95 (2017) 6–24, http://dx.doi.org/10.1016/j.astropartphys. 2017.08.005.

[3] Q. An, et al., Sci. Adv. 5 (9) (2019) eaax3793, http://dx.doi.org/10.1126/sciadv.aax3793.

[4] Alemanno, F., et al. Physical Review Letters 126.20 (2021): 201102.

[5] Dampe Collaboration. Science Bulletin 67.21 (2022): 2162-2166.

[6] G. Ambrosi, et al., Nature 552 (2017) 63–66, http://dx.doi.org/10.1038/nature24475.

[7] Alemanno, F., et al. Science Bulletin 67.7 (2022): 679-684.

[8] DAMPE Collab. 2023 arXiv:2304.00137.

[9] Alemanno, F., et al. Physical Review D 106.6 (2022): 063026.

[10] Y. Yu, et al., Astropart. Phys. 94 (2017) 1–10, http://dx.doi.org/10.1016/j.astropartphys.2017.06.004, arXiv: 1703.00098.

[11] M. Ding, et al., Res. Astron. Astrophys. 19 (3) (2019) 047, http://dx.doi. org/10.1088/1674-4527/19/3/47, arXiv:1810.09901.

[12] P. Azzarello, et al., Nucl. Instrum. Methods A 831 (2016) 378–384, http://dx.doi.org/10.1016/j.nima.2016.02.077.

[13] A. Tykhonov, et al., Nucl. Instrum. Methods A 924 (2019) 309–315, http://dx.doi.org/10.1016/j.nima. 2018.06.036, arXiv:1806.10355.

[14] Y.-L. Zhang, et al., Chin. Phys. C 36 (2012) 71–73.

[15] Z. Zhang, et al., Nucl. Instrum. Methods A 836 (2016) 98–104, http://dx.doi.org/10.1016/j.nima.2016.08.015.

[16] Y.-Y. Huang, et al. Res. Astron. Astrophys. 20 (9) (2020) 153, http://dx.doi.org/10.1088/1674- 4527/20/9/153, arXiv:2005.07828.

[17] M. Stolpovskiy, et al. JInst 17.06 (2022): P06031.

[18] A. F. Agarap. arXiv:1803.08375 (2018).

[19] G. D'Agostini. Nucl. Instrum. Methods Phys. Res. A, 362.2-3 (1995): 487-498.

[20] Tykhonov, A., et al. Astroparticle Physics 146 (2023): 102795.

[21] Tykhonov A., et al. Nucl. Instrum. Meth., A893 (2018), pp. 43-56, 10.1016/j.nima.2018.02.105

[22] Hough P.V.C., Kowarski L. (Ed.), Conf. Proc. C, 590914 (1959), pp. 554-558

[23] Zhang, Yong-Qiang, et al. Research in Astronomy and Astrophysics 19.9 (2019): 123.

[24] Y. Wei, et al. ICRC2023 proceedings, ID: 5363383.

[25] Coppin P., et al. ICRC2023 proceedings, ID: 5362068.

[26] E. Xu, et al. ICRC2023 proceedings, ID: 5363264.