Machine learning approach to the background reduction in singly charged cosmic-ray isotope measurements with AMS-02

Bueno, E. F.; Barão, F.; Vecchi, M.

Published in:
Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment

DOI:
10.1016/j.nima.2023.168644

Publication date:
2023

Citation for published version (APA):
Bueno, E. F., Barão, F., & Vecchi, M. (2023). Machine learning approach to the background reduction in singly charged cosmic-ray isotope measurements with AMS-02. *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *1056*, Article 168644. https://doi.org/10.1016/j.nima.2023.168644

Nuclear Instruments and Methods in Physics Research A 1056 (2023) 168644

Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

journal homepage: www.elsevier.com/locate/nima

Full Length Article

Machine learning approach to the background reduction in singly charged cosmic-ray isotope measurements with AMS-02

E.F. Bueno [a], F. Barão [b,c], M. Vecchi [a,*]

[a] Kapteyn Astronomical Institute, University of Groningen, Landleven 12, 9747 AD, Groningen, The Netherlands
[b] Laboratório de Instrumentação e Física Experimental de Partículas (LIP), 1649-003 Lisboa, Portugal
[c] Departamento de Física , Instituto Superior Técnico - IST, Universidade de Lisboa - UL, Avenida Rovisco Pais 1, 1049-001 Lisboa, Portugal

ARTICLE INFO

Keywords:
AMS-02
Machine learning
Cosmic-rays
Isotopes

ABSTRACT

Studying the isotopic composition of single-charge cosmic rays (CRs) provides essential data to investigate the CR propagation processes in our Galaxy. While current measurements are rare above 4 GeV/nucleon, the Alpha Magnetic Spectrometer (AMS-02) is able to measure the isotopic fluxes up to 10 GeV/n by combining the momentum measured by the silicon tracker with the precise measurements of the velocity provided by its Ring Imaging Cherenkov Detector (RICH). The correct measurement of the particles' velocity is essential for identifying isotopes through their mass. This is particularly challenging for single-charge particles due to the low number of photons they produce in their Cherenkov rings, which makes the reconstruction easily disrupted by noise. Hence, identifying the sources and cleaning the sample from the background is essential for ensuring the quality of the rings. In this paper, we propose a novel approach to track the events whose mass is misidentified due to interactions inside the AMS-02 detector. Based on the actual location of these interactions, we propose a novel strategy to mitigate the background effectively and with high efficiency, which includes using cut-based selection criteria and a multivariate estimator based on the signals detected by the RICH.

1. Introduction

Cosmic Rays (CRs) spectrum is dominated by positive, singly-charged nuclei [1]. Protons are the most significant component, but among the other particles, there is a small fraction of deuterons, the other stable isotope of hydrogen. CRs are usually classified as primary when they are produced and accelerated at the sources or secondary when their production happens due to the interaction of primary CRs with the interstellar medium (ISM). The deuteron component in CR, in particular, is expected to be mostly of secondary origin: deuterons are produced in the first step of the proton–proton chain in stars, only to be consumed in the following step [2], implying that the primary deuteron component is negligible. The secondary deuteron component comes mainly from the iterations of p, ^{3}He and ^{4}He with the ISM. Hence, studying the deuteron flux and the corresponding secondary-to-primary ratios, such as deuteron-to-proton and deuteron-to-helium-4, is essential for understanding the propagation processes in our Galaxy and the properties of the ISM itself [3].

Several magnetic spectrometers such as PAMELA, IMAX and CAPRICE have measured the isotopic composition of singly-charged particles in CRs [4–6]. However, the data above a few GeV/n are rare due to the technical challenges involved in this measurement. In this kind of detector, isotopes are separated through their mass, which is calculated by combining the rigidity (momentum per unit charge, $R = pc/Ze$) and the velocity $\beta = v/c$, through the equation

$$m = \frac{RZe}{\beta\gamma}\,, \tag{1}$$

where Ze is the magnitude of the charge and γ is the Lorentz factor. The mass resolution is given by:

$$\left(\frac{\Delta m}{m}\right)^2 = \left(\frac{\Delta R}{R}\right)^2 + \gamma^4\left(\frac{\Delta\beta}{\beta}\right)^2\,. \tag{2}$$

The dependence on the fourth power of the Lorentz factor indicates that the mass resolution increases rapidly as $\beta \to 1$. Hence, the correct velocity reconstruction is of great importance for mass reconstruction and isotope identification [7].

The Alpha Magnetic Spectrometer (AMS-02), a CR detector installed and collecting data aboard the International Space Station since May 2011 [8], can measure the deuteron flux in uncharted energy ranges due to the precise measurements of the velocity and rigidity provided

[*] Corresponding author.
E-mail addresses: e.ferronato.bueno@rug.nl (E.F. Bueno), barao@lip.pt (F. Barão), m.vecchi@rug.nl (M. Vecchi).

https://doi.org/10.1016/j.nima.2023.168644
Received 9 January 2023; Received in revised form 27 July 2023; Accepted 23 August 2023
Available online 25 August 2023

by its subdetectors. AMS-02 is composed of several subsystems: a silicon tracker of nine layers disposed from the top to bottom of the detector, that together with the permanent magnet of 0.15 T is responsible for measuring the magnitude and the sign of the charge, as well the rigidity of the particles; the Transition Radiation Detector (TRD), used to discriminate between leptons and hadrons; the Time of Flight (TOF), with two pairs of scintillators above (upper TOF) and below (lower TOF) the magnet, is responsible for measuring the velocity and the magnitude of the charge of the particles, as well as being the main trigger of the experiment; the Ring Imaging Cherenkov Detector, which measures the velocity and the charge of the particles; the Anti-Coincidence Counter (ACC), used to reject particles with high-incidence angle; and the Electromagnetic Calorimeter (ECAL), located right below the Ring Imaging Cherenkov Detector (RICH), responsible for measuring the energy of the particle and allowing for the discrimination between leptons and hadrons.

The RICH is essential for measuring the fluxes of hydrogen isotopes at higher energies, as it enables their identification up to 10 GeV/n [9]. However, it is important to stress that this measurement is very challenging for the RICH due to the nature of the Cherenkov effect. As the intensity of the signal is proportional to Z^2 [10], singly-charged isotopes have a faint signal, easily disrupting the velocity reconstruction by background. A more detailed description of the RICH and the relevant physical phenomena is given in the following section.

2. The AMS-02 RICH detector

The RICH is located below the lower TOF and is used to measure the velocity of the incoming particles with high precision [11,12]. The RICH detector (Fig. 1) has a truncated conical shape with a 60 cm top radius, a 67 cm bottom radius, and an expansion height of 47 cm. It is composed of a double radiator plane, an expansion volume, and a photo-detection plane. The double radiator plane consists of a central radiator formed by 16 tiles of sodium fluoride (NaF) with dimensions $8.5 \times 8.5 \times 0.5$ cm^3 and a refraction index of 1.33, surrounded by 92 tiles of silica aerogel with dimensions $11.5 \times 11.5 \times 2.5$ cm^3 and a refraction index of 1.05. The detection plane is equipped with an array of 680 Photo Multipliers Tubes (PMT) assembled in eight grids: four with a rectangular shape and four with a triangular shape. The PMT model used is the 4×4 multi-anode Hamamatsu R7600-00-M16, characterized by the fast and uniform response, low operational voltage, and low sensitivity to external magnetic fields. The Cherenkov photons are driven to the photocathode by an array of 10,880 solid pyramidal light guides made of Diakon LG-703 with different shapes to maximize the photon collection efficiency. Therefore, a detection cell is made of a photomultiplier coupled to 16 light guide pipes glued together, making a continuous detection surface of 34×34 mm^2. The detection plane presents a final spatial granularity of 8.5×8.5 mm^2 and gaps between PMTs of 3 mm. To reduce lateral losses of about 30% of the radiated Cherenkov photons, the expansion volume is surrounded by a highly reflective mirror. It has a conical shape and is made of three 120° segments, meeting roughness specifications of better than 150 nm, with a reflectivity exceeding 90% at $\lambda = 420$ nm. Dedicated calibration runs of all the AMS-02 subsystems are taken every time the detector crosses the Equator, on average, every 45 min.

Cherenkov emission occurs with an aperture angle $\cos\theta_c = \frac{1}{\beta n}$. Therefore, the particle's velocity threshold for NaF and aerogel is, respectively, $\beta = 0.752$ and $\beta = 0.952$. These velocity values translate to kinetic energy per nucleon thresholds of 0.5 GeV/n and 2.2 GeV/n, respectively, for aerogel and NaF [13]. The NaF measures the velocity of the particles with $Z = 1$ and $\beta = 1$ with resolution $\Delta\beta/\beta \approx 0.35\%$. The aerogel measures the velocity with resolution $\Delta\beta/\beta \approx 0.12\%$ in the same conditions [9].

The number of Cherenkov photons produced by a particle of charge Ze per wavelength, λ, and per distance, x, is given by the equation [10]:

$$\frac{d^2N}{d\lambda dx} = \frac{2\pi}{\lambda^2}\alpha Z^2 \sin^2\theta_C\,, \tag{3}$$

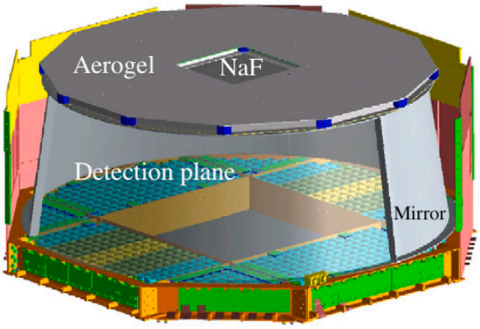

Fig. 1. The Ring Imaging Cherenkov Detector of AMS-02. The double radiator plane is at the top, with a central region of NaF surrounded by the aerogel. The detection plane is located below, with the hole to minimize the interactions before the particles reach the calorimeter. The expansion volume is surrounded by a highly reflective mirror used to reduce lateral losses. See text for discussion. Adapted from Ref. [14].

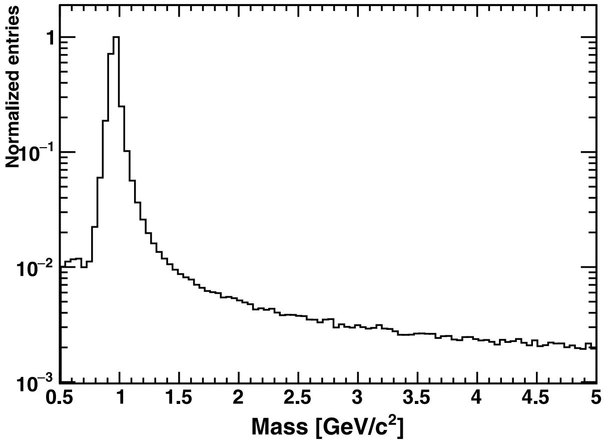

Fig. 2. Mass of simulated proton events in the RICH-NaF range, calculated with the generated rigidity and the reconstructed velocity. See text for discussion.

where α is the fine-structure constant, and θ_C the Cherenkov angle. Since the number of photons is proportional to Z^2, particles with $Z \geq 2$ are greatly benefited, having more intense, and thus easier to reconstruct, rings. As for singly-charged particles, there is no such increase of photons, and the rings are faint. The number of photon hits in a ring depends considerably on the impact point on the radiator and the particle inclination. Fully contained rings, only possible for a small fraction of near-threshold or very inclined events, have a multiplicity (for singly charged particles) of around four hits for NaF and six hits for aerogel. Moreover, secondary particle interactions (for example, delta-ray production) occurring in the region between the lower tracker and the RICH can induce a slight change in the particle direction and an invisible bias in the detected ring hits with respect to the reconstructed tracker track. In addition, aerogel is a source of photon scattering, providing a significant fraction of ring-uncorrelated hits (more than 40%). Given that at least three hits are required to reconstruct a Cherenkov ring, background hits near the ring or a misidentified particle direction from the tracker can bias the velocity reconstruction. This effect can be seen in Fig. 2, where the mass distribution calculated with the generated rigidity, and the measured velocity is shown for simulated proton events in the NaF. Since the generated rigidity is used, the long mass tails are caused by the velocity reconstruction. Although these tails are about 100 smaller than the proton mass peak, protons are about 100 times more abundant than deuterons in CRs. Hence, the proton mass tails constitute an essential source of background for deuteron identification that must be understood and mitigated.

This paper describes the study of such background, starting from tracking its sources via simulated proton events. Once the background sources are identified, selection criteria to remove interactions based on variables of different subdetectors are proposed, and their performance is assessed. Finally, a multivariate method based on the information available from the RICH ring reconstruction is presented.

AMS-02 data collected between May 2011 and May 2021 were used for this work. Singly-charged events were selected using the charge measured by the TOF and the silicon tracker. To ensure the quality of the reconstructed rigidity, the χ^2/NDF of the track fitting was required to be below 10 in both x (non-bending) and y (bending) coordinates. In addition, to ensure that the particles have a good velocity reconstruction in the TOF, only events with hits in the four scintillator planes were accepted. Both RICH radiators were used to measure the velocity in complementary velocity ranges: the NaF was used for events with $0.75 < \beta < 0.99$, and the aerogel in the interval $0.95 < \beta < 0.997$.

Monte Carlo (MC) simulations of protons traversing AMS-02 were also used. These simulations were produced by the AMS collaboration using a dedicated software based on the GEANT4 package [15]. The software simulates the propagation and interactions of a particle inside AMS-02, producing detector responses that are then used to reconstruct the event properties.

3. Tracking interactions inside AMS-02

Simulated events allow for tracing the propagation of the particles inside AMS-02, indicating the regions of the detector where inelastic interactions occurred. Fig. 3 shows the mass distributions of events with inelastic interactions in different locations inside AMS-02 for the NaF and in the aerogel velocity ranges. Panel (a) shows events that had interactions in the inner tracker (layers 2 to 8, inside the magnet) and its support structures; panel (b) shows events that interacted in the lower TOF and its support structure; events displayed in panel (c) interacted between the RICH radiator plane and the layer 9 (L9) of the tracker (between the RICH detection plane and the ECAL); while events displayed in panel (d) had no inelastic interactions. Since particles interacting in the upper part of the detector were entirely removed by the selection criteria applied to the simulations, the corresponding distributions for these events are not shown. The figure indicates that the inelastic interactions significantly shape the mass distribution, particularly the high-mass tail region. This is due to the fact that the additional particles produced during the interactions in the lower part of the detector (from the inner tracker to below) generate spurious hits in the RICH PMT plane, either by producing additional Cherenkov photons as they cross the radiators or by generating signals when they traverse the PMTs themselves.

The NaF and aerogel show different features in some of the distributions. In panels (a) and (b), the interactions lead to a more considerable increase of events reconstructed with low mass in the aerogel compared to the NaF. Due to the fact that the refractive index of the aerogel is smaller than that of the NaF, the particles radiating through aerogel will generate Cherenkov cones with smaller aperture than those radiating through the NaF. This difference is especially relevant for events close to the Cherenkov threshold, where a few spurious hits can cause larger rings to be wrongly reconstructed, whose velocity is consequently overestimated, and the mass (proportional to the inverse of the velocity) is thus underestimated. In addition, particles interacting between the RICH radiator plane and the L9 do not contribute as much to the tails in the aerogel as they do in the NaF. This indicates that in the aerogel, many of the interactions do not take place in the radiator plane itself but rather in other locations: the mirror, especially for particles with high incidence angle, where they do not affect the reconstruction, or in the PMT plane itself, where they produce an intense signal which is disregarded by the velocity reconstruction algorithm. Radiator interactions producing additional Cherenkov hits have a larger impact on the NaF reconstruction than on the aerogel due to the combined effect of having a lower number of hits associated with the rings and most of the interactions taking place in the radiator. For these reasons, mass reconstruction biases are more noticeable on NaF than on aerogel.

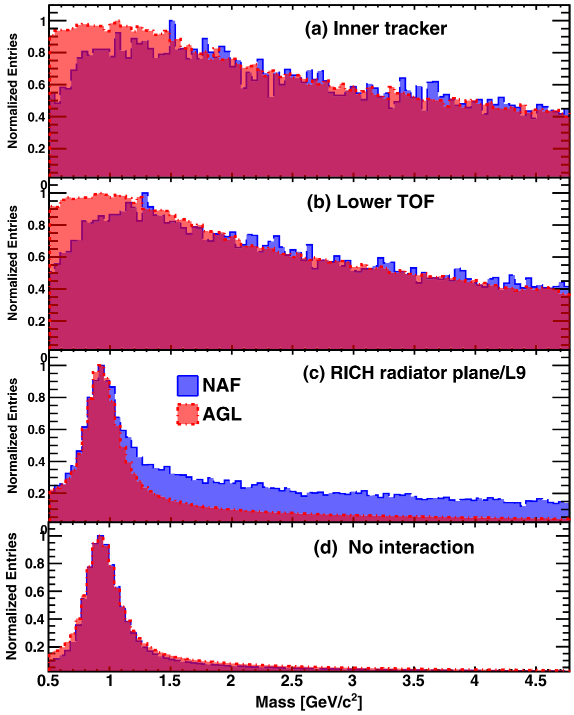

Fig. 3. Mass distributions for events in the whole NaF (blue, solid lines) and aerogel (red, dot-dashed lines) ranges with inelastic interactions inside AMS-02. Panel (a) shows events with interactions in the inner tracker; (b) with interactions in lower TOF; (c) with interactions between the RICH radiator plane and the top of the ECAL; and (d) events without any interaction. See text for discussion.

4. Variables for background mitigation

Once the sources of interactions relevant to the RICH reconstruction background were found, strategies for mitigating them were developed. First, we started by looking at variables from detectors other than the RICH, namely the tracker and the TOF, that could clean the sample. This was done by looking at the distribution of these variables in experimental data. Since the mass resolution is about 10% [7], events were considered signal-like if they had a reconstructed mass within 2σ from the proton mass, that is, $0.75 < m < 1.12$ GeV/c^2; and were considered background-like if the mass was above 4σ the triton mass, that is $m > 4$ GeV/c^2. The distributions of each variable were compared for both classes of events, allowing for the creation of selection criteria.

4.1. Rigidity matching

The silicon tracker comprises nine independent layers used to measure the rigidity of the particles [8]. The first layer is located on top of the TRD, the second is above the magnet, and L9 is between the RICH and the ECAL, and it covers the ECAL acceptance. Comparing the rigidity measured in different layers allows for accurate background tagging: if no inelastic interactions occur, the rigidity should be the same for any combination of layers apart from resolution effects.

The interactions happening in the lower TOF and RICH radiator plane can be reduced by checking the matching between the rigidity calculated with the inner tracker and the inner tracker plus the L9. This is expressed by the quantity $\rho_{I,I+L9}$ defined as:

$$\rho_{I,I+L9} = \frac{|R_I - R_{I+L9}|}{R_{I+L9}}, \tag{4}$$

where R_I is the rigidity from the inner tracker and R_{I+L9} from the inner tracker plus the L9. Larger deviations from 0 indicate an inelastic interaction between the inner tracker and the L9, the region where

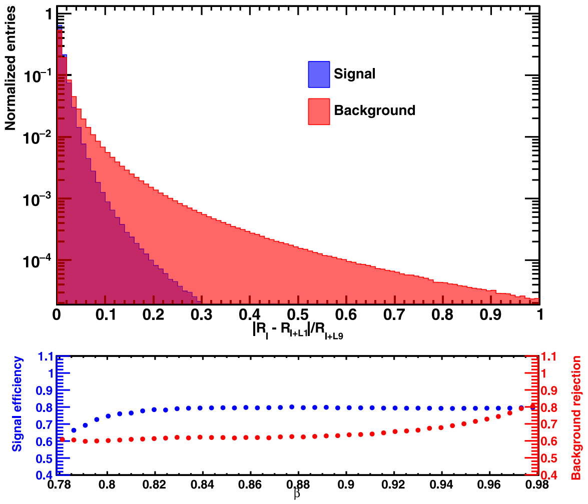

Fig. 4. The top panel shows the distribution of $\rho_{I,I+L9}$ for signal (blue) and background (red) events in the NaF. The bottom panel shows the signal efficiency (in blue) and the background rejection (in red) for the selection criterion $\rho_{I,I+L9} < 0.05$, as a function of velocity. See text for discussion.

lower TOF and the RICH radiator plane are located. The top panel of Fig. 4 shows the normalized distributions of this variable for events in the signal and background regions. Background events have a broader distribution, more likely than the signal at $\rho_{I,I+L9} > 0.05$. The bottom panel of Fig. 4 shows the efficiency (in blue) and the background rejection (in red) for events with $\rho_{I,I+L9} < 0.05$ as a function of the velocity. The signal efficiency ε_{sig} is computed by comparing the number of events that pass specific selection criteria, to the total number of events detected by the RICH. On the other hand, the background rejection is given by $1 - \varepsilon_{bkg}$ where ε_{bkg} is the efficiency of the same selection criterion but is applied to the sample of background events. Requiring events with $\rho_{I,I+L9} < 0.05$ can significantly reduce the background while keeping the signal efficiency above 75% in almost the entire velocity range. As this variable requires the measurement of the rigidity using L9, it cannot be used in the aerogel range as it would greatly reduce the acceptance due to the geometry of the events.

The interactions in the upper part of the detector (inner tracker and above) can be removed by comparing the rigidity measured in the inner tracker and the inner tracker plus layer 1 (above the TRD), that is, $\rho_{I,I+L1}$. The top panel of Fig. 5 shows the distributions of this variable for signal and background events. Similarly to the case of $\rho_{I,I+L9}$, the distribution of background suggests the criterion $\rho_{I,I+L1} < 0.05$ for selecting signal-like events. The bottom panel shows this criterion's signal efficiency and background rejection for the aerogel range.

The results show that this variable can remove over 60% of the background while keeping the efficiency above 90%.

4.2. Lower TOF charge

In addition to the tracker, the signals collected by the TOF can be used to tag and remove some noise from interactions. Inelastic interactions in the lower TOF lead to increased energy loss in the scintillator paddles. Since the measured charge increases with the intensity of the signal, it is expected that the charge of background-like events is larger than one. The top panel of Fig. 6 shows the distribution of the lower TOF charge, Z_{LTOF} for background and signal events in the aerogel. A slight shift in the peak and a large tail in the background distribution are seen. The bottom panel of the same figure shows the signal efficiency and background rejection of only allowing events with $Z_{LTOF} < 1.3$. The efficiency of this selection is above 90% in the entire velocity range, while the background rejection increases steadily from 60 to 90%.

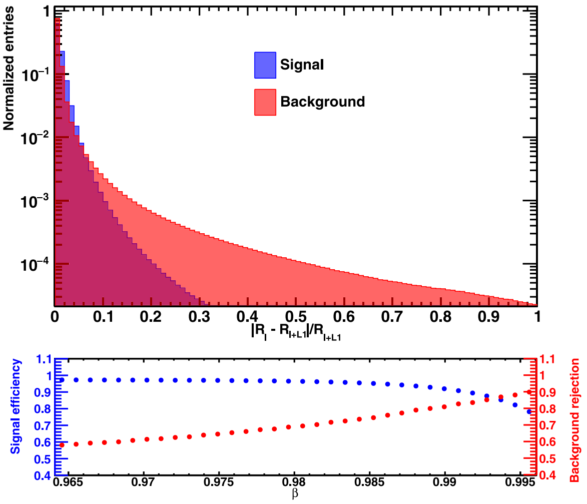

Fig. 5. The top panel shows the distribution of $\rho_{I,I+L1}$ for signal (blue) and background (red) events in the aerogel. The bottom panel shows the signal efficiency in blue and the background rejection in red for the selection criterion $\rho_{I,I+L1} < 0.05$, as a function of velocity. See text for discussion.

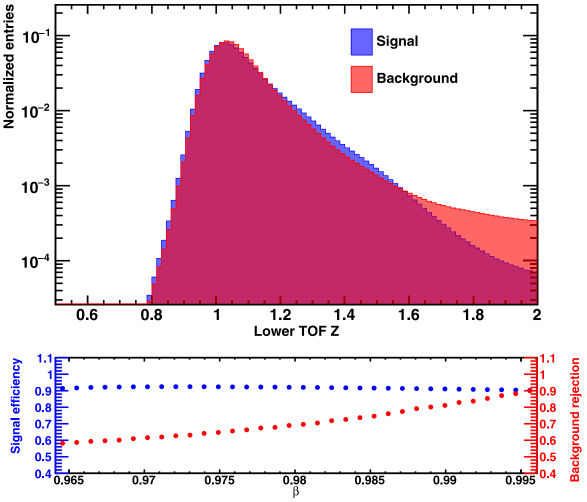

Fig. 6. The top panel shows the distribution of the lower TOF charge for signal (blue) and background (red) events in the aerogel. The bottom panel shows the signal efficiency in blue and the background rejection in red for the selection criterion $Z_{LTOF} < 1.3$. See text for discussion.

The results of applying the selection criteria based on the tracker and TOF variables are shown for both NaF and aerogel in Fig. 7. In both cases, it is clear that the selection reduces the high-mass tails considerably, indicating that the proposed variables help to remove events with interactions in the relevant locations. The additional cut on $\rho_{I,I+L9}$ in the NaF range makes the background rejection higher.

So far, the signals detected by the RICH detector have not been used to mitigate the background. The next section will describe the multivariate method used to incorporate the information from the RICH ring reconstruction in the event selection.

5. RICH boosted decision tree

In modern high-energy physics, machine-learning methods have become commonplace [16]. They are very popular due to their robustness and reliability [17] and are rapidly being applied, adapted, and developed for nuclear physics [18], particle physics [19], astroparticle physics [20,21] and cosmology [22,23]. For the particular purpose of

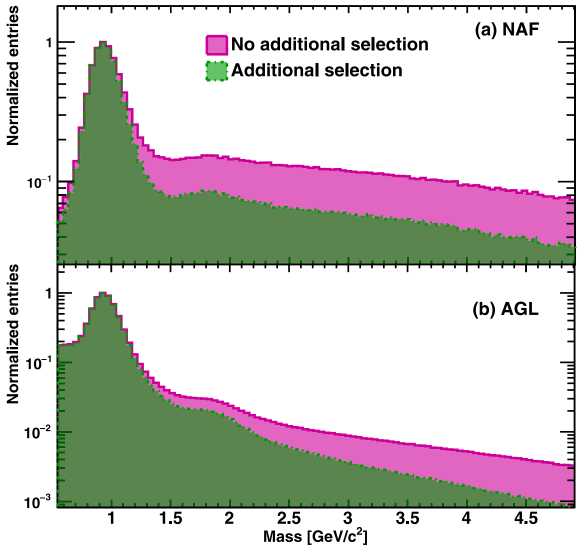

Fig. 7. Mass distributions of events in the NaF (panel (a)) and aerogel (panel (b)) for events without the additional cuts (pink, solid lines) and with the additional cuts proposed (green, dot-dashed line). See text for discussion.

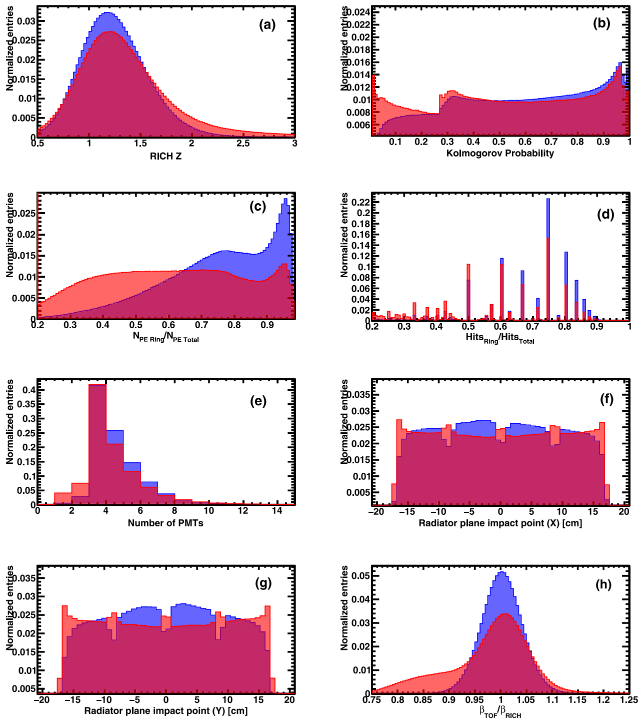

Fig. 8. Distributions of RICH variables for signal (blue) and background (red) events: (a) RICH charge; (b) ring Kolmogorov probability; (c) $N_{PE}^{Ring}/N_{PE}^{Total}$; (d) $N_{Hits}^{Ring}/N_{Hits}^{Total}$; (e) number of PMTs; (f) Coordinate X of radiator impact point; (g) Coordinate Y of radiator impact point; (h) β_{TOF}/β_{RICH}. See text for discussion.

particle identification, boosted decision trees (BDTs) [17,24] has been used since the pioneering work in the MiniBOONE experiment [25]. BDTs have been widely used also within the AMS-02 Collaboration to identify electrons and positrons and reject protons, based on the 3D reconstruction of the shower shape provided by the ECAL [26–29]. BDTs are well suited for classification problems, such as identifying events with good and bad velocity reconstruction in the RICH, as different variables related to the ring reconstruction can be used. In this section, we present the construction of the estimator and an assessment of its performance.

5.1. Construction

Taking into account their different acceptance and overlapping velocity ranges, two distinct BDTs have been trained: one for the NaF and the other for the aerogel. In order to avoid systematic uncertainties coming from differences between data and simulations, the training sample was a subset of the experimental data being analyzed. The criteria described in the previous section were applied to all the data, and the events were classified as signal or background, following the same criteria. The used variables are listed below, with a short description and motivation for each of them.

1. **RICH charge**: In the RICH, the charge Z is defined as $Z = \sqrt{N_{pe}/N_{exp}}$, where N_{pe} is the number of measured photoelectrons and N_{exp} is the number of expected photoelectrons for a $Z = 1$ particle, given its trajectory inside the radiator and its velocity. Events with higher charges are more likely to be the background.
2. **Kolmogorov probability**: A Kolmogorov–Smirnov hypothesis test [30] is performed to ensure that the azimuthal distribution of the Cherenkov photons is compatible with a uniform distribution (consistent with a ring-like shape). In this case, this standard hypothesis test compares the expected with the observed cumulative distribution of the azimuthal angle. The maximal discrepancy between both cumulative provides the Kolmogorov probability. In events with no compatibility between the two distributions, the test yields a low value of the Kolmogorov probability.
3. **Fraction of photoelectrons in the ring**: a fraction of the emitted Cherenkov photons is deflected due to interactions inside the radiator (Rayleigh scattering, absorption) and optical effects (refraction and reflection) [31]. Hence, only a fraction of the detected photons are used for the reconstruction. Thus, events with a larger fraction of detected photons being used in the ring reconstruction are more likely to be well-reconstructed.
4. **Fraction of used hits**: similar to the previous variable, but consists of the raw number of hits, neglecting the fluctuations in the gain of the PMTs.
5. **Number of PMTs**: number of PMTs used in the ring reconstruction. Events with few PMTs and a large number of photoelectrons indicate noisy events.
6. **Radiator impact point (X, Y)**: particles impacting the radiator in the borders between radiator tiles and on the edges of the radiators tend to yield fewer Cherenkov photons detectable by the PMT plane. Hence they are more prone to poor reconstruction.
7. **RICH/TOF velocity matching**: The velocity of the particle should be the same in both detectors. Deviations indicate misreconstruction in at least one of them.

Fig. 8 shows the distributions of these variables for signal and background in the NaF range. It is possible to note that all variables have some capacity to distinguish both classes of events, making them suitable for building such an estimator.

The BDT was trained using the TMVA package [32], using the default settings for this type of classifier. Fig. 9 shows the distributions of the RICH BDT for signal and background in both the NaF and aerogel ranges. It is possible to notice that the separation between signal and background is more significant in the NaF than in the aerogel. The rings in aerogel are smaller and have a higher number of photons, making the reconstruction less susceptible to noise. This indicates that a more significant fraction of the high-mass events in the aerogel come from tails in the rigidity distribution, which cannot be removed by a BDT based on RICH variables.

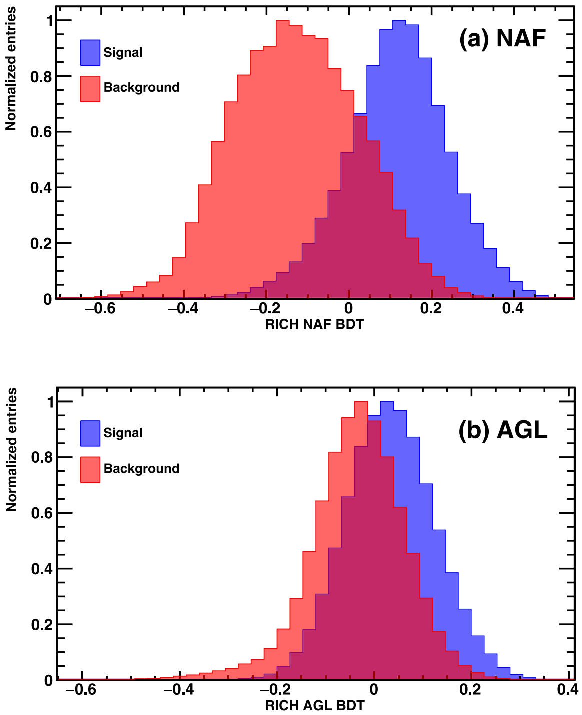

Fig. 9. Output distributions of the BDT classifier for signal (blue) and background (red) events in the NaF (a) and aerogel (b). See text for discussion.

5.2. Performance

Following the optimization strategy implemented in TMVA [32], we identified the selection criteria which maximize the significance of the signal, defined as:

$$S = \varepsilon_{sig} N_{sig} / \sqrt{\varepsilon_{sig} N_{sig} + \varepsilon_{bkg} N_{bkg}}, \qquad (5)$$

where ε_{sig} and ε_{bkg} are the signal and background efficiencies, respectively. N_{sig} and N_{bkg} correspond to the number of signal and background events before the cuts. This procedure yielded the criteria $\mathrm{BDT}_{NaF} > 0.05$ and $\mathrm{BDT}_{aerogel} > 0$. Panels (a) and (b) of Fig. 10 show the signal efficiency, ε_{sig}, and the background rejection, $1 - \varepsilon_{bkg}$, for these criteria in the NaF and aerogel ranges, respectively.

The clearer separation between signal and background in the classifier distributions for the NaF region shown in Fig. 9 is reflected here as higher signal efficiency and background rejection than the aerogel. Still, in both radiators, the efficiencies are above 60%, except at the beginning of each velocity range, where the efficiency drop is related to the Cherenkov emission threshold. Conversely, the background rejection remains above 90% in almost the entire velocity range for both radiators, decreasing at the end. This decrease in background rejection follows the same principle that explains the better performance of the BDT in the NaF. As the velocity increases, the rings have more photons, so they are less disturbed by noise in the detection plane. Hence a larger fraction of the high-mass tails come from the rigidity reconstruction and thus cannot be removed by the RICH BDT.

The result of applying the proposed criteria in the data is shown as the blue distributions in panels (a) and (b) of Fig. 11. The proton and deuteron mass peaks are sharper due to the background rejection, which eases the identification of the isotopes. In addition to cleaning the events with high mass, it is possible to notice a reduction of the low-mass background in both radiators. Although the estimator was not trained in such events, they are also caused by interactions in the detector, which lead to a misreconstruction of the velocity in the RICH. Hence, they are also removed by the BDT classifier.

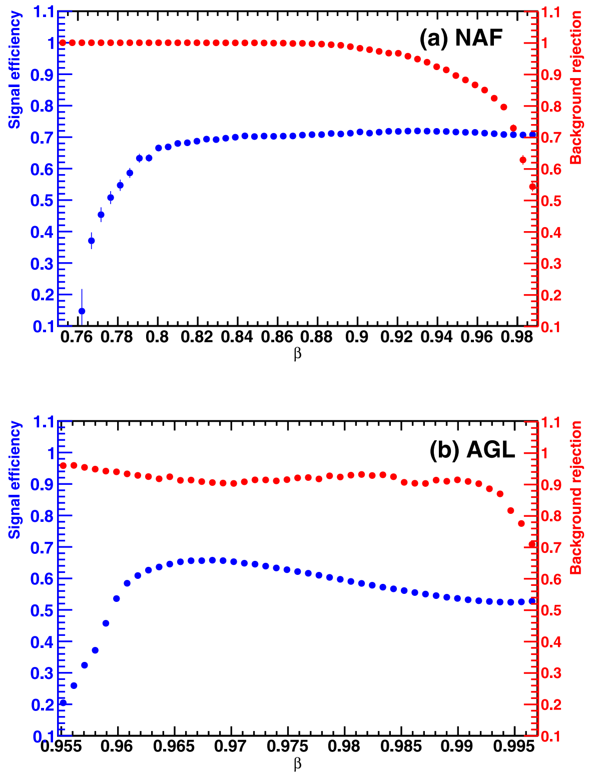

Fig. 10. Signal efficiency (blue) and background rejection (red) for events with $\mathrm{BDT}_{NaF} > 0.05$ and $\mathrm{BDT}_{aerogel} > 0$ in NaF and aerogel, in panels (a) and (b), respectively, as a function of velocity. See text for discussion.

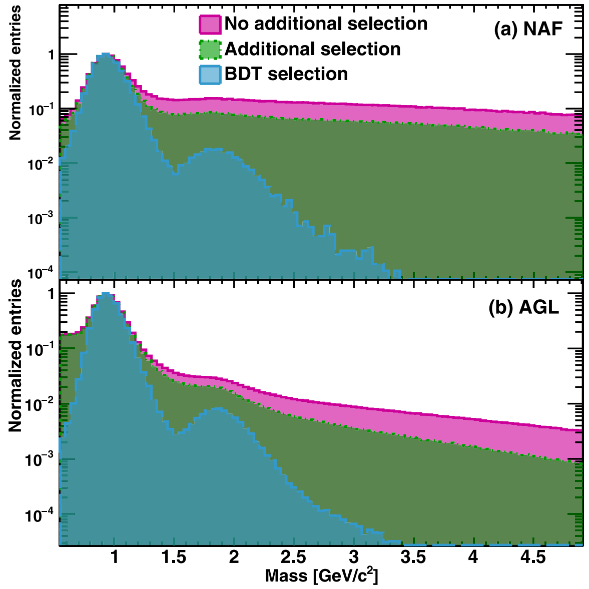

Fig. 11. Mass distributions of events in the NaF (panel (a)) and aerogel (panel (b)) for events without the additional cuts (solid pink lines); with the additional cuts in tracker and TOF variables (green dot-dashed line); and with the BDT selection (blue dashed line). See text for discussion.

6. Conclusion

Expanding the energy range of the measurements of the fluxes of $Z = 1$ isotopes in CRs will provide an essential asset in studying the propagation processes in our Galaxy. This is challenging since deuterons are one hundred times less abundant than protons in CRs.

The AMS-02 experiment can identify these isotopes up to 10 GeV/n thanks to the precise measurement of the velocity provided by its Cherenkov detector, thus filling the energy gap left by the current measurements. Nevertheless, the velocity reconstruction of single-charged particles is challenging, primarily due to the faint signal induced by single-charged particles traversing the detector.

In this article, we have presented a study of the background that affects the velocity reconstruction in the RICH, leading to high-mass tails that impact the deuteron identification. By tracking the propagation of simulated proton events inside AMS-02, we could pinpoint the locations where inelastic interactions lead to the misreconstruction of the velocity and, thus, the mass. They occur in the lower part of the detector, especially the lower TOF and its support structure, and in the RICH radiator plane itself. Motivated by the locations where these interactions occur, we proposed using variables from the silicon tracker and TOF that were shown to have a signal efficiency and background rejection above 80 and 60%, respectively. In addition, taking advantage of the information provided by the RICH ring reconstruction, a Boosted Decision Tree estimator was trained using the signals collected in each RICH radiator. Based on the output distributions of the classifier for both signal and background, selection criteria were proposed in the NaF and aerogel. The BDT has shown the capacity of removing over 90% of the background over the full velocity ranges of both RICH radiators while keeping the efficiency close to 70% in the NaF and 60% in the aerogel. By combining the information from the tracker, TOF and the RICH BDT, the data sample was thoroughly cleaned, revealing the deuteron mass peak among the proton tail, enabling the identification of such particles.

CRediT authorship contribution statement

E.F. Bueno: Conceptualization, Methodology, Software, Data curation, Formal analysis, Visualization, Writing – original draft. **F. Barão:** Conceptualization, Methodology, Software, Writing – review & editing. **M. Vecchi:** Conceptualization, Methodology, Writing – review & editing, Funding acquisition, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgments

This publication is part of the project “Statistical methods applied to cosmic ray anti-deuteron searches with the AMS-02 experiment” with project number 040.11.723 of the research programme Bezoekersbeurs 2019 VW, which is financed by the Dutch Research Council (NWO). We are grateful to Portuguese FCT for the financial support through grant CERN/FIS-PAR/0007/2021.

References

[1] T.K. Gaisser, R. Engel, E. Resconi, Cosmic Rays and Particle Physics, second ed., Cambridge University Press, http://dx.doi.org/10.1017/CBO9781139192194.

[2] E.G. Adelberger, et al., Solar Fusion Cross Sections. II. The pp Chain and CNO Cycles, Vol. 83, American Physical Society, pp. 195–245, http://dx.doi.org/10.1103/RevModPhys.83.195, URL https://link.aps.org/doi/10.1103/RevModPhys.83.195.

[3] B. Coste, et al., Constraining Galactic Cosmic-Ray Parameters with $Z \leq 2$ Nuclei, Vol. 539, p. A88, http://dx.doi.org/10.1051/0004-6361/201117927.

[4] O. Adriani, G.C. Barbarino, G.A. Bazilevskaya, R. Bellotti, M. Boezio, E.A. Bogomolov, M. Bongi, V. Bonvicini, S. Bottai, A. Bruno, F. Cafagna, D. Campana, P. Carlson, M. Casolino, G. Castellini, C.D. Donato, C.D. Santis, N.D. Simone, V.D. Felice, V. Formato, A.M. Galper, A.V. Karelin, S.V. Koldashov, S. Koldobskiy, S.Y. Krutkov, A.N. Kvashnin, A. Leonov, V. Malakhov, L. Marcelli, M. Martucci, A.G. Mayorov, W. Menn, M. Mergè, V.V. Mikhailov, E. Mocchiutti, A. Monaco, N. Mori, R. Munini, G. Osteria, F. Palma, B. Panico, P. Papini, M. Pearce, P. Picozza, M. Ricci, S.B. Ricciarini, R. Sarkar, V. Scotti, M. Simon, R. Sparvoli, P. Spillantini, Y.I. Stozhkov, A. Vacchi, E. Vannuccini, G. Vasilyev, S.A. Voronov, Y.T. Yurkin, G. Zampa, N. Zampa, Measurements of Cosmic-Ray Hydrogen and Helium Isotopes with the Pamela Experiment, Vol. 818, No. 1, The American Astronomical Society, p. 68, http://dx.doi.org/10.3847/0004-637X/818/1/68.

[5] G.A. de Nolfo, L.M. Barbier, E.R. Christian, A.J. Davis, R.L. Golden, M. Hof, K.E. Krombel, A.W. Labrador, W. Menn, R.A. Mewaldt, J.W. Mitchell, J.F. Ormes, I.L. Rasmussen, O. Reimer, S.M. Schindler, M. Simon, S.J. Stochaj, R.E. Streitmatter, W.R. Webber, A Measurement of Cosmic Ray Deuterium from 0.5–2.9 Gev/nucleon, Vol. 528, No. 1, pp. 425–428, http://dx.doi.org/10.1063/1.1324352, URL https://aip.scitation.org/doi/abs/10.1063/1.1324352.

[6] P. Papini, et al., High-Energy Deuteron Measurement with the CAPRICE98 Experiment, Vol. 615, pp. 259–274, http://dx.doi.org/10.1086/424027.

[7] E. Bueno, F. Barão, M. Vecchi, A Parametric Approach for the Identification of Single-Charged Isotopes with AMS-02, 1031, 166564, http://dx.doi.org/10.1016/j.nima.2022.166564, URL https://www.sciencedirect.com/science/article/pii/S016890022200167X.

[8] M. Aguilar, et al., The Alpha Magnetic Spectrometer (AMS) on the International Space Station: Part II — Results from the First Seven Years, Vol. 894, pp. 1–116, http://dx.doi.org/10.1016/j.physrep.2020.09.003, The Alpha Magnetic Spectrometer (AMS) on the International Space Station: Part II - Results from the First Seven Years. URL https://www.sciencedirect.com/science/article/pii/S0370157320303434.

[9] F. Giovacchini, J. Casaus, A. Oliva, The AMS-02 RICH Detector: Status and Physics Results, Vol. 952, Elsevier BV, 161797, http://dx.doi.org/10.1016/j.nima.2019.01.024.

[10] J.D. Jackson, Classical Electrodynamics, Wiley.

[11] L. Arruda, F. Barao, P. Goncalves, R. Pereira, The ring imaging cherenkov detector of the AMS experiment: test beam results with a prototype, Nucl. Phys. B Proc. Suppl. 172 (2007) 32–35, http://dx.doi.org/10.1016/j.nuclphysbps.2007.07.025, arXiv:0801.4484.

[12] F. Giovacchini, AMS RICH Collaboration Collaboration, Space application: The AMS RICH, Nucl. Instrum. Methods A 970 (2020) 163657, http://dx.doi.org/10.1016/j.nima.2020.163657.

[13] M. Aguilar-Benitez, et al., In-beam aerogel light yield characterization for the AMS RICH detector, Nucl. Instrum. Methods A 614 (2010) 237–249, http://dx.doi.org/10.1016/j.nima.2009.12.027.

[14] R. Pereira, The AMS-02 RICH Detector: Performance During Ground-Based Data Taking at CERN, Vol. 639, No. 1, pp. 37–41, http://dx.doi.org/10.1016/j.nima.2010.09.036, Proceedings of the Seventh International Workshop on Ring Imaging Cherenkov Detectors. URL https://www.sciencedirect.com/science/article/pii/S0168900210020292.

[15] J. Allison, et al., Recent Developments in Geant4, Vol. 835, pp. 186–225, http://dx.doi.org/10.1016/j.nima.2016.06.125, URL https://www.sciencedirect.com/science/article/pii/S0168900216306957.

[16] K. Albertsson, et al., Machine Learning in High Energy Physics Community White Paper, Vol. 1085, No. 2, IOP Publishing, 022008, http://dx.doi.org/10.1088/1742-6596/1085/2/022008.

[17] D. Bourilkov, Machine and Deep Learning Applications in Particle Physics, Vol. 34, No. 35, 1930019, http://dx.doi.org/10.1142/S0217751X19300199.

[18] M. Carnini, A. Pastore, Trees and Forests in Nuclear Physics, Vol. 47, No. 8, 082001, http://dx.doi.org/10.1088/1361-6471/ab92e3, arXiv:2002.10290.

[19] B. Liu, I. Laktineh, Q. Shen, G. Garillot, J. Guo, F. Lagarde, X. Wang, Y. Zhu, H. Yang, Particle Identification Using Boosted Decision Trees in the Semi-Digital Hadronic Calorimeter, Vol. 15, No. 5, p. C05022, http://dx.doi.org/10.1088/1748-0221/15/05/C05022.

[20] M. Krause, E. Pueschel, G. Maier, Improved γ/Hadron Separation for the Detection of Faint γ-Ray Sources Using Boosted Decision Trees, Vol. 89, pp. 1–9, http://dx.doi.org/10.1016/j.astropartphys.2017.01.004, arXiv:1701.06928.

[21] S. Ohm, C. van Eldik, K. Egberts, γ/Hadron Separation in Very-High-Energy γ-Ray Astronomy Using a Multivariate Analysis Method, Vol. 31, No. 5, pp. 383–391, http://dx.doi.org/10.1016/j.astropartphys.2009.04.001, arXiv:0904.1136.

[22] B. Ostdiek, A. Diaz Rivero, C. Dvorkin, Image Segmentation for Analyzing Galaxy-Galaxy Strong Lensing Systems, Vol. 657, 657, p. L14, http://dx.doi.org/10.1051/0004-6361/202142030, arXiv:2009.06663.

[23] J. Brehmer, S. Mishra-Sharma, J. Hermans, G. Louppe, K. Cranmer, Mining for Dark Matter Substructure: Inferring Subhalo Population Properties from Strong Lenses with Machine Learning, Vol. 886, No. 1, p. 49, http://dx.doi.org/10.3847/1538-4357/ab4c41, arXiv:1909.02005.

[24] Y. Coadou, Boosted decision trees, in: Artificial Intelligence for High Energy Physics, World Scientific, pp. 9–58, http://dx.doi.org/10.1142/9789811234033_0002.

[25] H.-J. Yang, B.P. Roe, J. Zhu, Studies of Boosted Decision Trees For MiniBooNE Particle Identification, Vol. 555, No. 1, pp. 370–385, http://dx.doi.org/10.1016/j.nima.2005.09.022, URL https://www.sciencedirect.com/science/article/pii/S0168900205018322.

[26] M. Aguilar, et al., First Result from the Alpha Magnetic Spectrometer on the International Space Station: Precision Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–350 GeV, Vol. 110, 141102, http://dx.doi.org/10.1103/PhysRevLett.110.141102.

[27] M. Aguilar, et al., Precision Measurement of the $(e^{+} + e^{-})$ Flux in Primary Cosmic Rays from 0.5 GeV to 1 TeV with the Alpha Magnetic Spectrometer on the International Space Station, Vol. 113, 221102, http://dx.doi.org/10.1103/PhysRevLett.113.221102.

[28] L. Accardo, et al., High Statistics Measurement of the Positron Fraction in Primary Cosmic Rays of 0.5–500 GeV with the Alpha Magnetic Spectrometer on the International Space Station, Vol. 113, 121101, http://dx.doi.org/10.1103/PhysRevLett.113.121101.

[29] M. Graziani, Electron/Proton Separation and Analysis Techniques Used in the AMS-02 $(E^{+} + e^{-})$ Flux Measurement, pp. 2351–2353, http://dx.doi.org/10.1016/j.nuclphysbps.2015.09.388, 37th International Conference on High Energy Physics (ICHEP). URL https://www.sciencedirect.com/science/article/pii/S2405601415008779.

[30] W.T. Eadie, D. Drijard, F.E. James, Statistical Methods in Experimental Physics, North-Holland, Amsterdam, 1971.

[31] F. Barao, et al., The AMS-RICH velocity and charge reconstruction, in: 30th International Cosmic Ray Conference, Vol. 2, pp. 457–460, arXiv:0709.2154.

[32] A. Hocker, et al., TMVA - Toolkit for Multivariate Data Analysis with ROOT: Users guide. TMVA - Toolkit for Multivariate Data Analysis, CERN, TMVA-v4 Users Guide: 135 pages, 19 figures, numerous code examples and references. URL https://cds.cern.ch/record/1019880.