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

Optimization of the proton background rejection in the measurement of the electron flux at high energies with CALET on the International Space Station

Sandro Gonzi[1,2,3], **Eugenio Berti**[2,3] and **Lorenzo Pacini**[2,3]

for the **CALET collaboration**

The **CALorimetric Electron Telescope (CALET)**, operating aboard the **International Space Station (ISS)** since October 2015, is an experiment dedicated to high-energy astroparticle physics. In this contribution the results of a study conducted on different **multivariate analysis techniques** in order to optimize the **proton rejection** at **high energies** in the measurement of the **electrons and positrons (all-electron)** flux are discussed.

Physics motivations

A precise measurement of the cosmic-ray electron and positron spectrum in the high-energy TeV region could provide:

- a unique probe of *nearby cosmic accelerators* [1];
- an indication on the origin of the *observed increase of the positron fraction* over 10 GeV [2,3].

In both cases, it is expected that the all-electron spectrum would exhibit some peculiar spectral features.

CALET detector

CALET detector [4] employs a calorimeter with a field of view of $\sim 45°$ from zenith, a geometrical factor of ~ 1040 cm^2 sr and a total depth of ~ 30 radiation-length X_0 for particles at normal incidence. It consists of:

- **Charge Detector (CHD)**: a pair of plastic scintillator hodoscopes arranged in two orthogonal layers, in order to identify the charge of the incident particle;
- **Imaging Calorimeter (IMC)**: a sampling calorimeter made of alternated thin layers of Tungsten absorber and scintillating fibers read-out individually;
- **Total AbSorption Calorimeter (TASC)**: a tightly packed lead-tungstate (PWO) hodoscope, capable of almost complete containment of the TeV-electromagnetic showers.

This design leads to excellent detector performances: an electromagnetic shower energy resolution of $\sim 2\%$ above 20 GeV and a protons rejection factor of $\sim 10^5$.

Event display

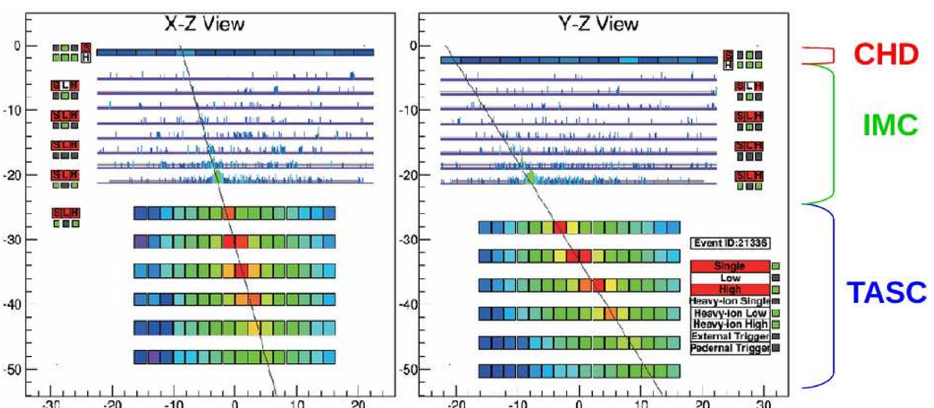

Figure 1: electron (or positron) event candidate [5] showing energy deposit in each detector channel in the X - Z and Y - Z views (reconstructed energy of 3.05 TeV and energy deposit sum of 2.89 TeV).

Analysis strategy

Monte Carlo (MC) simulations of electrons and protons, performed with the EPICS [6] framework, are used to evaluate efficiencies and background contamination.

A group of **pre-selections** [7] is applied to obtain a well reconstructed sample of electron candidates, removing contamination from events outside acceptance and particles with charge $Z > 1$.

Two different **proton rejection selections** [7] are applied, depending on the energy, to further suppress the contaminating proton background:

- a simple two-parameter cut (**K-cut**), used *below 500 GeV*;
- a **multivariate algorithm**, used *above 500 GeV*.

In both cases the **electrons identification efficiency** has been fixed to 80%.

The residual **proton contamination** is subtracted from the final measurement of the all-electron flux.

Multivariate analysis (MVA) techniques

The multivariate algorithms tested are the standard ones developed in the ROOT-integrated [8] environment Toolkit for Multivariate Analysis (TMVA) [9].

All multivariate techniques in TMVA make use of **training** events, for which the output is known, to determine the mapping function that describes a decision boundary (classification).

Selected methods belong to the Boosted Decision Trees (BDT), Artificial Neural Network (ANN) and Deep Learning (DL) classes:

- BTD: changes in the number of trees in the forest (t) and in the maximum depth of the decision tree allowed (d);
- ANN: MultiLayer Perceptrons (MLP) and Multi Layer Perceptrons Bayesian Neural Network (MLPBNN) tested algorithms. Changes in the number of training cycles (n) and in the specification of hidden layer architecture (h);
- DL: Deep Neural Network (DNN) tested algorithm. Changes in the number of hidden layers (h) and in the number of neurons of each layer (n). The network activation function has been fixed (ss: *SIGMOID* for all the neurons).

Comparison of MVA methods

TMVA estimators have been built with a sample of 13 variables related to the shower development in the CALET detector [7].

MC samples of electrons (signal) and protons (background) after the pre-selection have been splitted into training and test samples (same number in each energy bin) with a random seed.

The Receiver Operating Characteristic (ROC) diagram (background rejection versus signal efficiency) has been built.

Best performances (for each class):

- BDT ($t = 1000$, $d = 5$);
- BDT ($t = 200$, $d = 10$);
- MLP ($n = 600$, $h = 10$);
- MLPBNN ($n = 60$, $h = 9$);
- DNN ($n = 20$, $h = 1$, ss).

BDT is the best performing method.

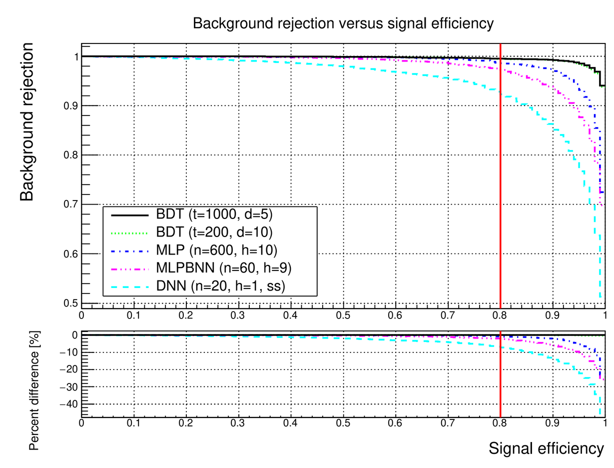

Figure 2: ROC curves for $E \in [2899, 4594]$ GeV (80% signal efficiency highlighted).

Stability and performances

- **Stability**: BDT remains the best algorithm while changing the energy or the seed in the training-test splitting. Other methods have an underperformance of at least 2% (80% signal efficiency).
- **Performances**: the **BDT** $(t = 1000, d = 5)$ method (selected as the reference one) shows stable performances while changing the energy or the training-test splitting.

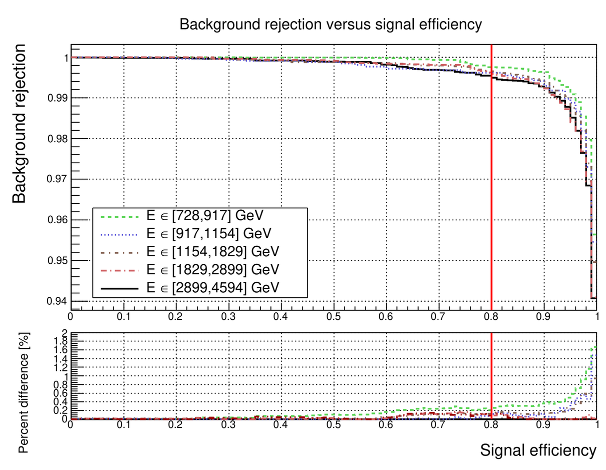

Figure 3: performances as a function of the energy in the training-test sample splitted with seed 1.

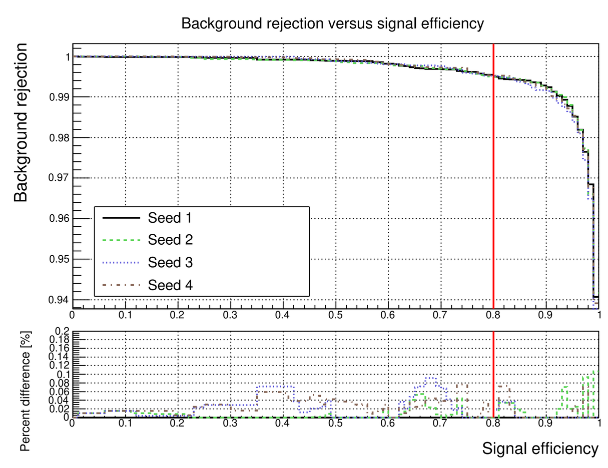

Figure 4: performances as a function of the training-test splitting, with $E \in [2899, 4594]$ GeV.

MVA in the all-electron flux measurement

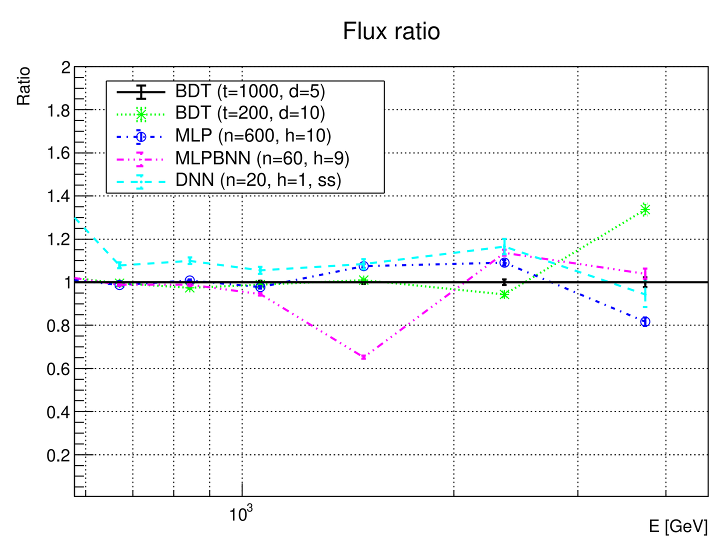

Figure 5: flux ratio between the selected TMVA methods and the BDT ($t = 1000$ $d = 5$) one, selected as reference (only statistical errors are shown).

MVA rejection algorithms have been applied on the CALET all-electron flux measurement:

- 2637 days of flight data (high-energy shower trigger) in the full detector acceptance processed with the standard procedure [5];
- proton contamination in the final electrons sample for BDT ($t = 1000$, $d = 5$) is in general less than 15% above 500 GeV;
- fluxes obtained by changing the MVA methods are stable considering the large contamination obtained using non-BDT algorithms.

The **BDT** $(t = 1000, d = 5)$ algorithm, selected for the **CALET all-electron analysis**, is the one with the best performances.

References:

[1] T. Kobayashi *et al.*, *Astrophys. J.* **601**, 340 (2004)

[2] O. Adriani *et al.*, *Nature (London)* **458**, 607 (2009)

[3] L. Accardo *et al.*, *Phys. Rev. Lett.* **113**, 121101 (2014)

[4] S. Torii, P. S. Marrocchesi *et al.*, *Adv. Space Res.* **64**, 2531 (2019)

[5] O. Adriani *et al.*, *Phys. Rev. Lett.* **120**, 261102 (2018)

[6] K. Kasahara, in *Proceedings of 24th ICRC*, **1**, 399 (1995)

[7] E. Berti *et al.*, in *Proc. Sci.*, **ICRC2021**, 065 (2021)

[8] R. Brun *et al.*, *Nucl. Instrum. Methods A*, **389**, 81 (1997)

[9] A. Hoecker *et al.*, *arXiv:physics/0703039* [physics.data-an] (2009)

Acknowledgements:

We gratefully acknowledge JAXA's contributions to the development of CALET and to the operations onboard the International Space Station. We also express our sincere gratitude to ASI and NASA for their support of the CALET project. The CALET effort in Italy is supported by ASI under Agreement No. 2013- 018-R.0 and its amendments.

(1) University of Florence, Department of Physics and Astronomy, Sesto Fiorentino, Italy

(2) National Institute for Nuclear Physics INFN, Division of Florence, Sesto Fiorentino, Italy

(3) National Research Council CNR, Institute of Applied Physics IFAC, Sesto Fiorentino, Italy