Development of an agnostic global PID tool for the ND280 near detector

Patrick Bates
sgpbates@liverpool.ac.uk
23/05/2025
Supervisors: Prof. Neil McCauley, Dr. Sam Jenkins

T2K experiment

- The T2K experiment is a long baseline neutrino experiment in Japan
- The experiment is composed of the three facilities:
  - Japan Proton Accelerator Research Complex (J-PARC)
  - The near detector suites – INGRID, ND280, and WAGASCI-BabyMIND, located 280 m downstream
  - The far detector – Super-Kamiokande (Super-K), located 295 km downstream
- The goals of the T2K experiment are:
  - Measure δ_{cp} by comparing oscillation of neutrinos and anti-neutrinos
  - Constraining the neutrino interaction cross section

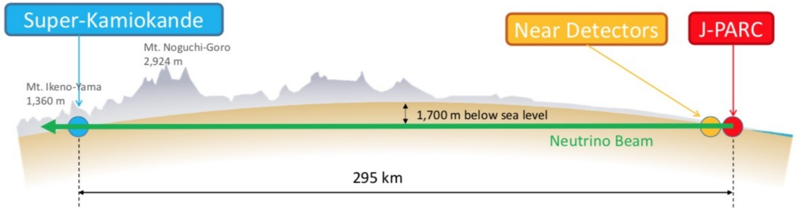


Diagram of the T2K experiment: J-PARC facility, Near detector suites, and Super-K

ND280 off-axis near detector

- ND280 is an off-axis near detector
  - Off-axis angle of 2.5º
  - Located 280m downstream of the beam source
- ND280 samples the neutrino beam before oscillation for measuring neutrino disappearance/appearance and constrains the neutrino-interaction cross section
- The old configuration was made up of the following sub-detectors:
  - Pi-Zero detector (P0D)
  - Time projection chambers (TPC)
  - Fine-grained detector (FGD)
  - Side muon range detector (SMRD)
  - Enclosed in an electromagnetic calorimeter (ECal)
  - Placed within a UA1 magnet
- ND280 near detector has undergone an upgrade:
  - Removal of the P0D
  - Installation of Super Fine Grained Detector (SFGD), High-Angle Time Projection Chamber (HATPC), and Time of Flight planes (TOF).

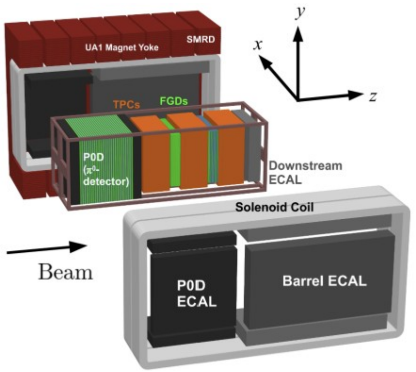


Diagram of the old ND280 near detector: composed of FGDs, TPC, SMRD, P0D, and ECAL

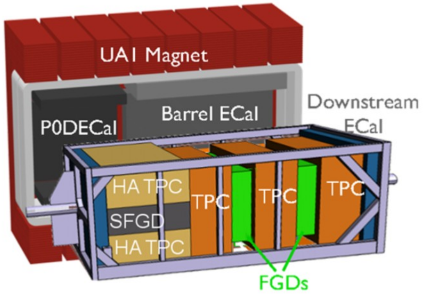
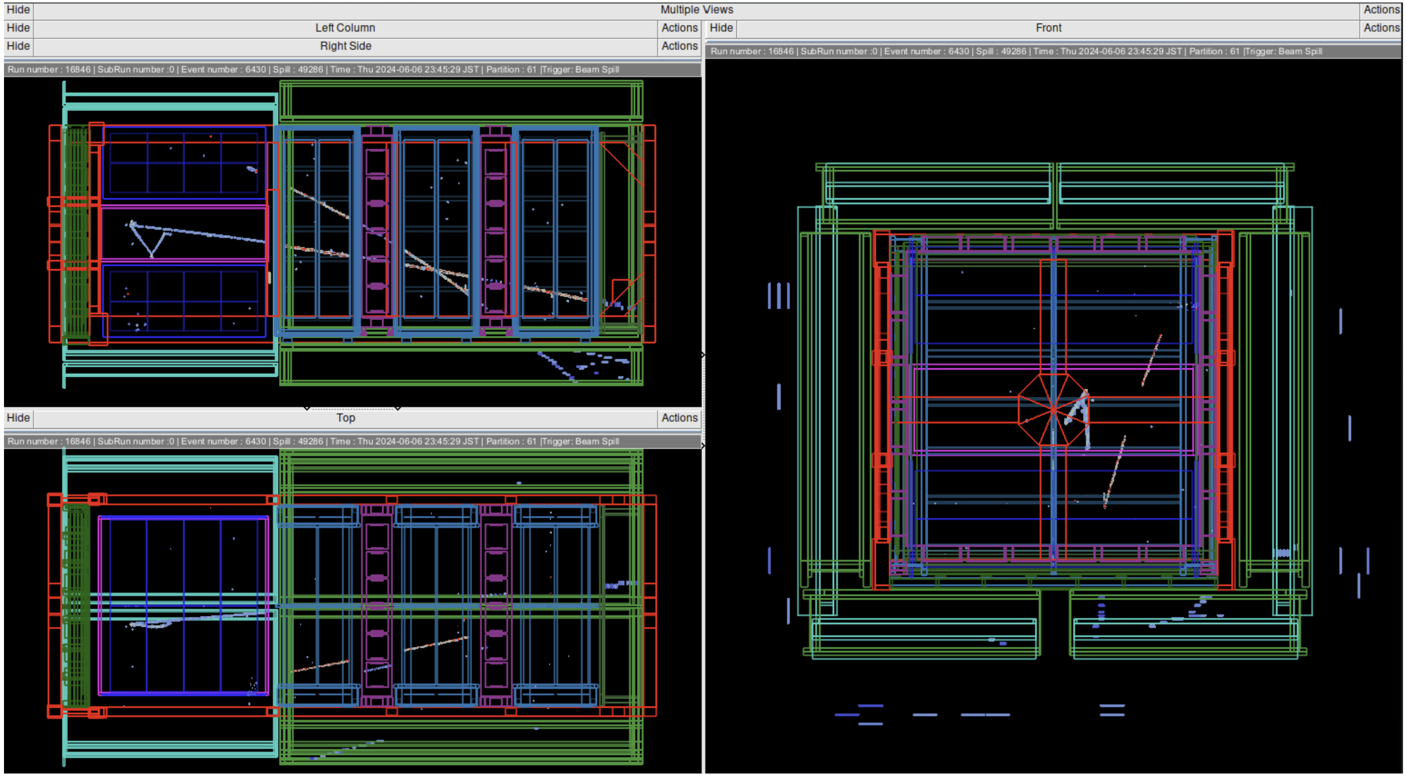


Diagram of the current ND280 near detector: composed of SFGD, FGDs, TPC, HATPC, TOF, SMRD, and ECAL

CC1pi selection (1/2)

- Charge current 1 pion (CC1pi) is an interaction topology with one pion in the final state
- CC1pi make up a significant portion of charged current events at T2K as a result they form a major background for charge current quasi-elastic (CCQE) samples which are essential for oscillation analysis
- Cross-section measurements require effective event selection algorithms
    - Uses a sequence of rectangular cuts on the reconstructed track + Done on a per sub-detector basis
- Anti-muon CC1π^- selection identify tracks with a μ^+ and a π^-, rejecting all others
- There is an issue of poor discrimination between muons and pions that behave like minimum ionizing particles (MIPs)
    - Pions belong to two topologies: MIPs and electromagnetic showers (EM showers)
- A μ^- and a π^+ can look like a μ^+ and a π^- which is termed the ‘wrong signed’ background

CC1pi selection (2/2)

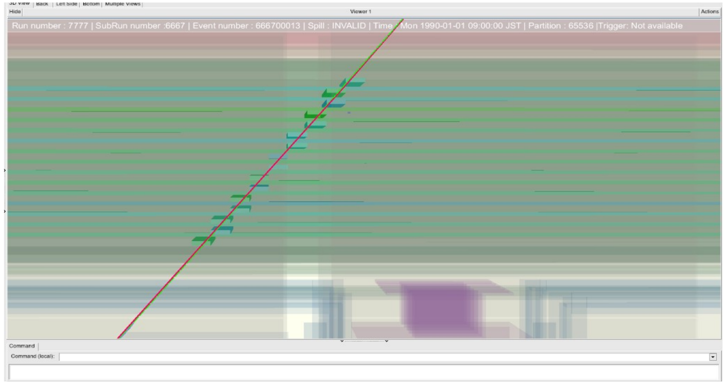


Global event display of the current ND280 configuration: Shows CC1pi events, with tracks originating in the SFGD

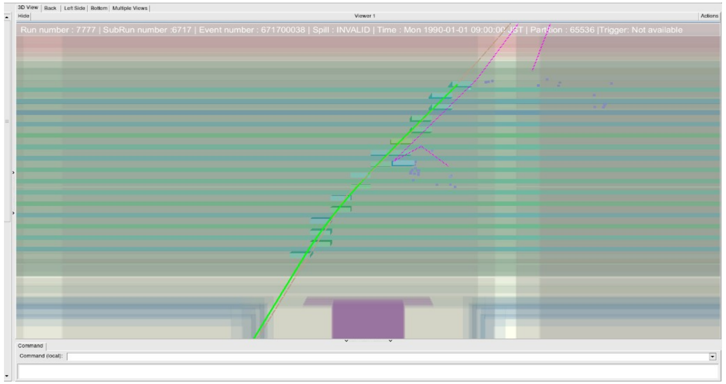
Development of a global agnostic PID tool

- The goal of my analysis is to develop a global agnostic particle identification (PID) tool to be used in the ND280 near detector
    - It has no assumptions on what selection it is being used in and uses information from all available sub-detectors in the ND280 detector
- This utilizes a boosted decision tree (BDT) using the XGBoost package
- I train and test my BDT using particle gun events:
    - All samples used so far are produced using the old ND280 detector configuration
    - Events start at FGD1
    - Momentum: uniformly distributed between limits of 150 MeV/c and 2000 MeV/c
    - Direction: uniformly distributed in a cone of opening angle 65º
    - 1×10^6 events per training and testing sample
    - Each sample is made up of μ^+, π^+, e^+, and p^+ subsamples - each subsamples being 2.5×10^5 events

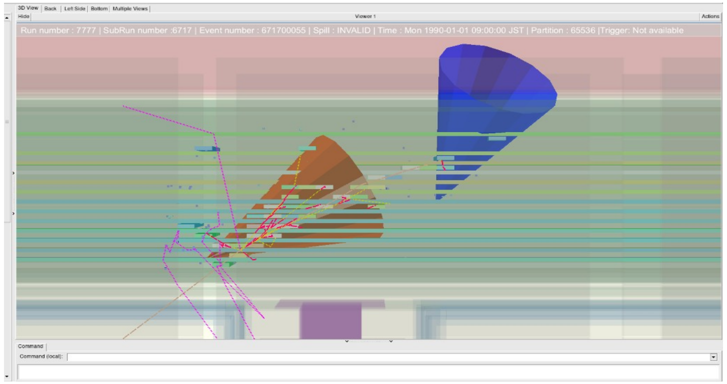
Different particle topologies seen in ND280



MIP like tracks passing through the barrel ECal produced by an anti-muon event



MIP like tracks passing through the barrel ECal produced by a positive pion event



Electromagnetic shower seen in barrel ECal produced by a positive pion event

BDT input variables (1/2)

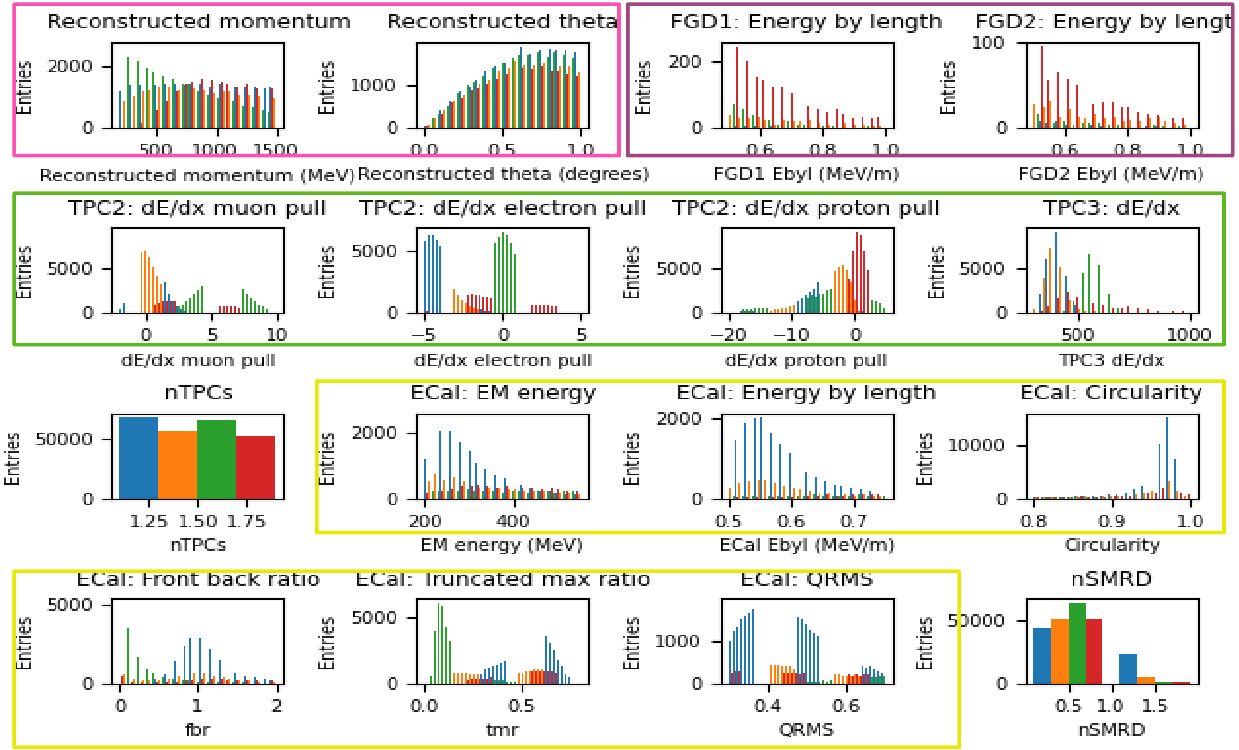
- 16 input variables were chosen to be used to train and test the BDT
  - They were chosen via N-1 study
- Issues with low momenta reconstruction
  - Drop-off in proton statistics at low momenta (around 500 MeV/c and below), due to detector acceptance thresholds
  - The momenta of positrons are skewed heavily towards lower values, due to brehmsstrahlung energy
- Re-weighting is applied to prevent direct dependency on momentum using the following equation:

$$W_{preco}(i)=1/N(i)$$

$W_{preco}(i)$ is the weight for events in reconstructed momentum bin i, and N(i) is the total number of events in that bin

BDT input variables (2/2)

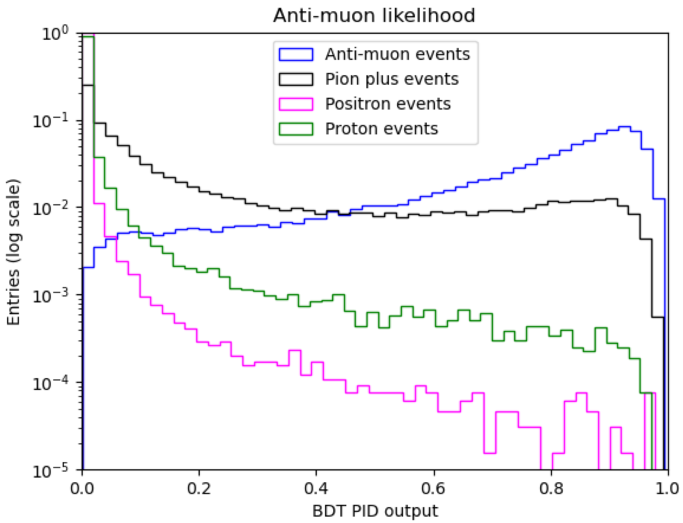
- Anti-muon
- Pion-plus
- Positron
- Proton



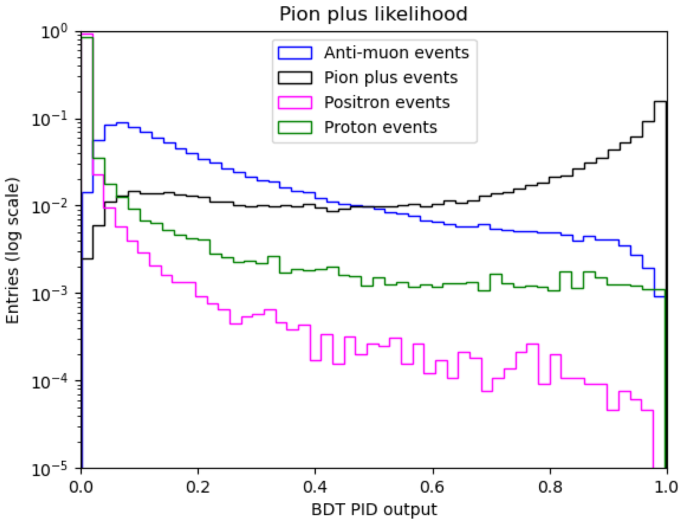
- FGD variables
- TPC variables
- ECal variables
- Global Track variables

BDT input variables: Reconstructed momentum, Reconstructed theta, FGD1 and FGD2 Energy by Length, TPCs particle pull, TPC3 dE/dx, nTPCS, ECAL EM energy, ECAL Energy by Length, ECAL circularity, ECAL Front back ratio, ECAL Truncated max ratio, ECAL QRMS and nSMRD

Results of BDT



Anti-muon likelihood: shows the likelihood of each particle type (μ^+, π^+, e^+, and p^+) being a muon



Positive pion likelihood: shows the likelihood of each particle type (μ^+, π^+, e^+, and p^+) being a pion

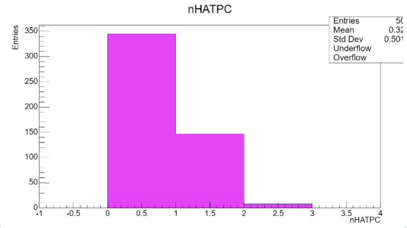
Note: data is pgun from old ND280 configuration

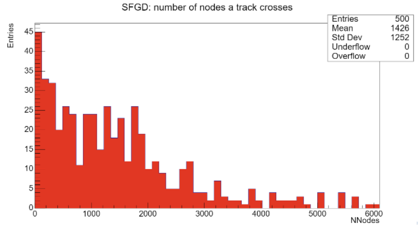
Developing the BDT for the current ND280 configuration (1/2)

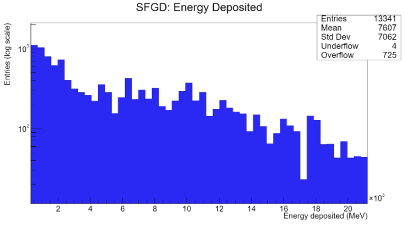
- Develop a selection that uses the current configuration of the ND280 near detector
  - This is underway – lots of debugging
  - This will provide new variables to train and test the BDT with such as dE/dx in the SFGD, track length in the SFGD, TOF inputs, etc.
  - Follow with N-1 study
- Look into alternative machine learning algorithms
  - Current developing a script for random forests
- Test the effectiveness of my PID tool I will compare the performance of current CC1pi selection on its own against the same selection using my BDT

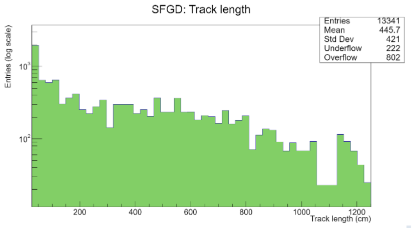
Developing the BDT for the current ND280 configuration (2/2)

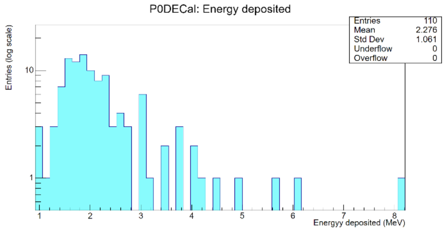
Anti-muon like events



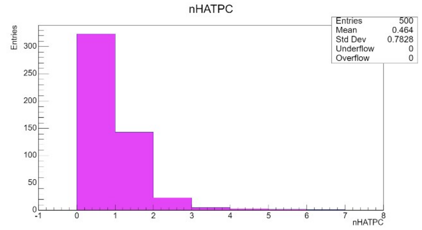


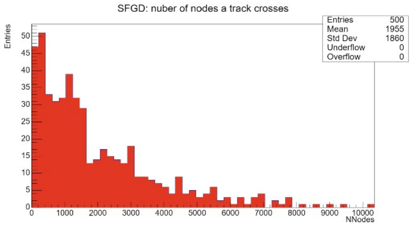


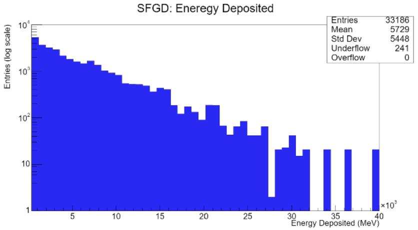


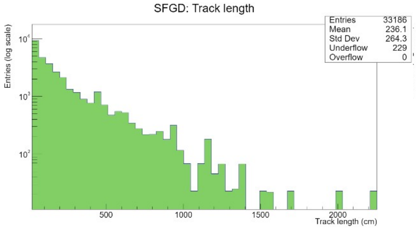


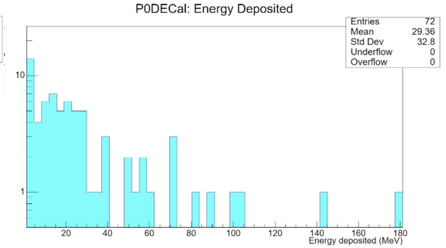
Positive pion like events











Number of HATPCs a track passes through

The number of nodes a track passes through while inside the SFGD

The energy deposited by a track as it passes through through the SFGD

The length of a track as it passes through the SFGD

The energy deposited by a track as it passes through the P0DECal

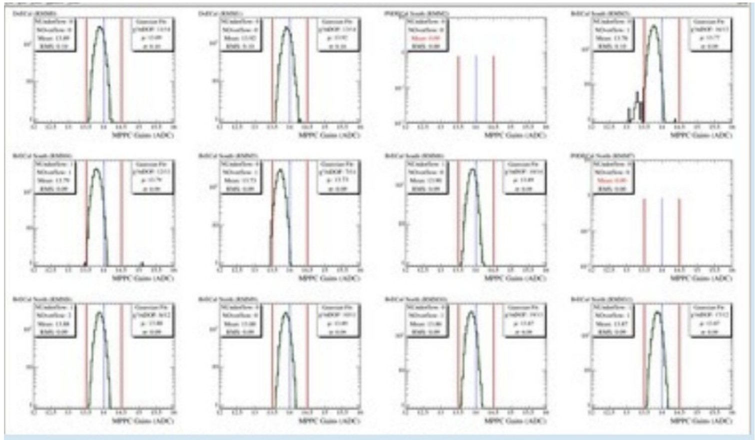
Note:

- No TOF nor dE/dx due to software issues
- Incomplete list – more variables will be used

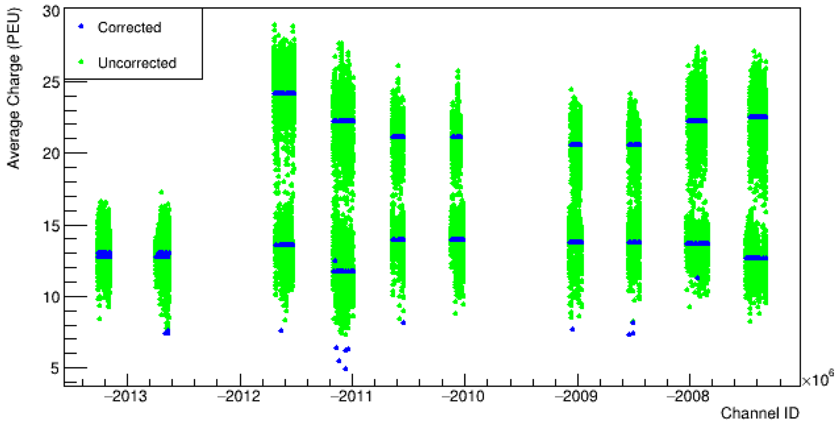
UNIVERSITY OF LIVERPOOL

Calibration work

- During my LTA (and when available to visit Japan) I am TTD expert which involves monitoring the SMRD, Ecal and us-ECal and performing calibration of the detectors during maintenance days of a run
- For the ecal bar-to-bar the bars of scintillating material are calibrated using cosmic muons
  - Done on a per run basis



One set of plots, MPPC gains for Ecal, that show the performance of the TripT detectors



bar-2-bar correction run13

Thank you for listening

Questions?

Back-up slides

Conclusion

- I have a BDT built that can be trained on ND280 variables which will be used in my agnostic global PID tool
- Results in current configuration seem promising
- I am developing a CC1pi selection that uses the current ND280 configuration
- I know which variables I want to use from the current configuration – not a complete list yet
- Looking into alternative machine learning algorithms

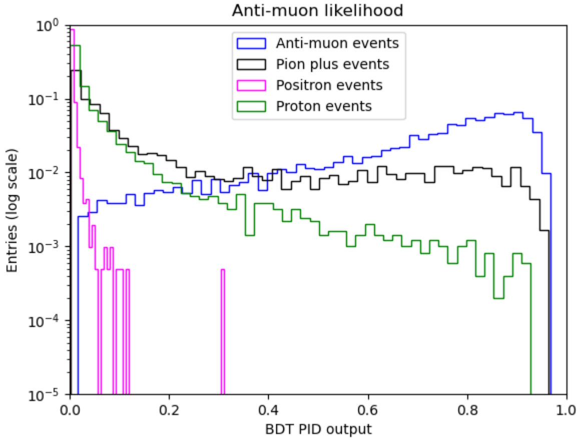
Development of a global agnostic PID tool (3/)

- Hyper-parameters are variables that define how a BDT builds, controlling the max depth, the number of trees, and pruning methods to combat over complexity
  - Overcomplexity causes a dependency on the training set resulting in poor performance
- The BDTs hyper-parameters was optimized used Hyper-opt library, which utilizes a bayesian method to find the optimal value of hyper-parameters which reduces the loss function
  - Define a range for each hyper-parameter used, which creates a parameter space which is searched
- It was found that the optimization values are degenerate – no one optimized set of value but a combination of value which gives similar performance within the hyper-parameter space
- Tests done with hyper-parameters emphasises the importance for appropriate values to ensure good performance

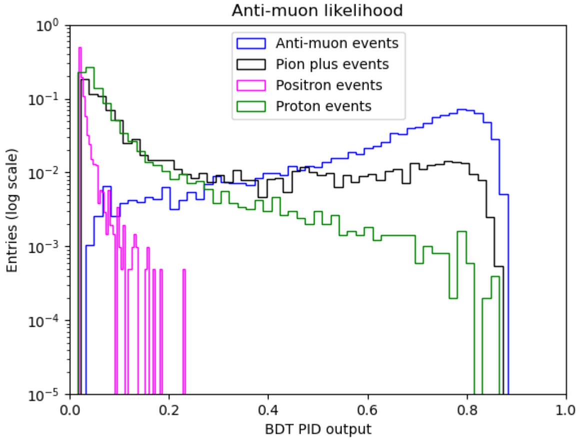
Investigating removal of hyper-parameter from optimization search (1/2)

- The cause of the max likelihood stopping around 0.8 is due to Xgboost using default values
- Carried out a study that removed different hyper-parameters from my parameter space search which makes XGBoost use default values
  - Removed max_delta_step
  - Removed colsample_bytree
  - Removed min_child_weight
- This study is not finished as the script takes some time to finish

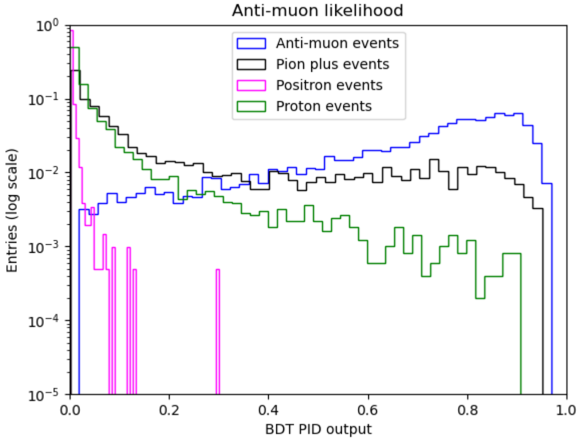
Investigating removal of hyper-parameter from optimization search (2/2)



Anti-muon likelihood with no max_delta_step



Anti-muon likelihood with no col_sample_bytree



Anti-muon likelihood with no min_child_weight

Note: data is pgun from old ND280 configuration