Machine Learning Based Analysis of Non-fiducial Electrons

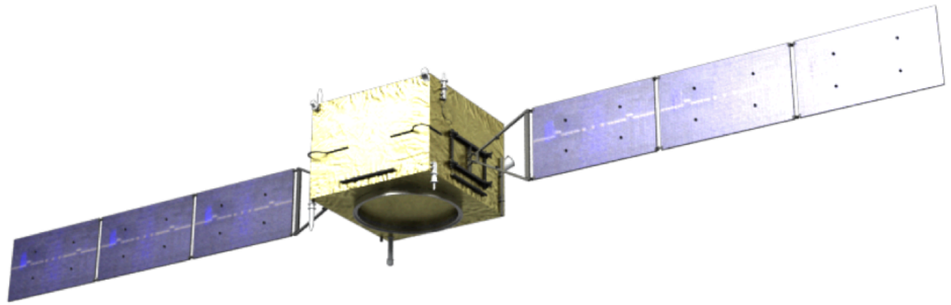
Mingyv Yan (阎明宇), Cong Zhao (赵聪) and Yifeng Wei (魏逸丰) *

University of Science and Technology of China

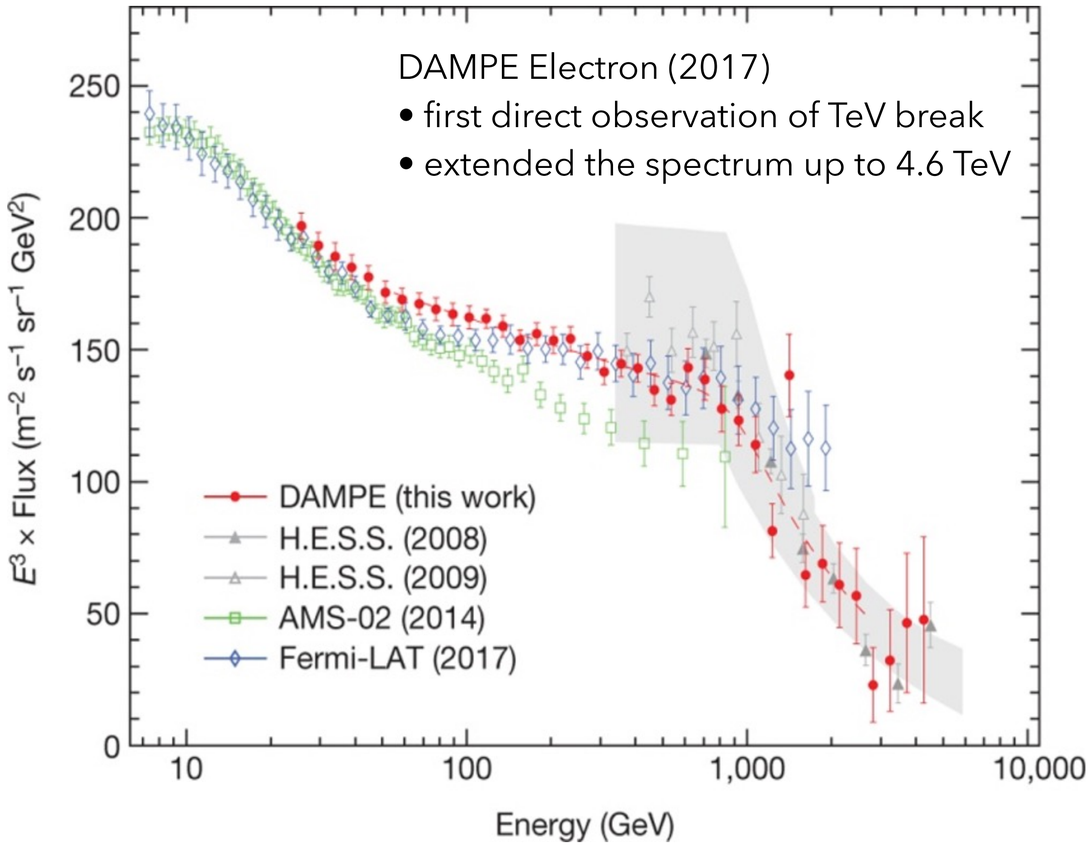
February 10th, 2026

Outline

- Motivation
- Method
- Analysis procedure
- Preliminary result of $e^+ + e^-$ flux
- Summary



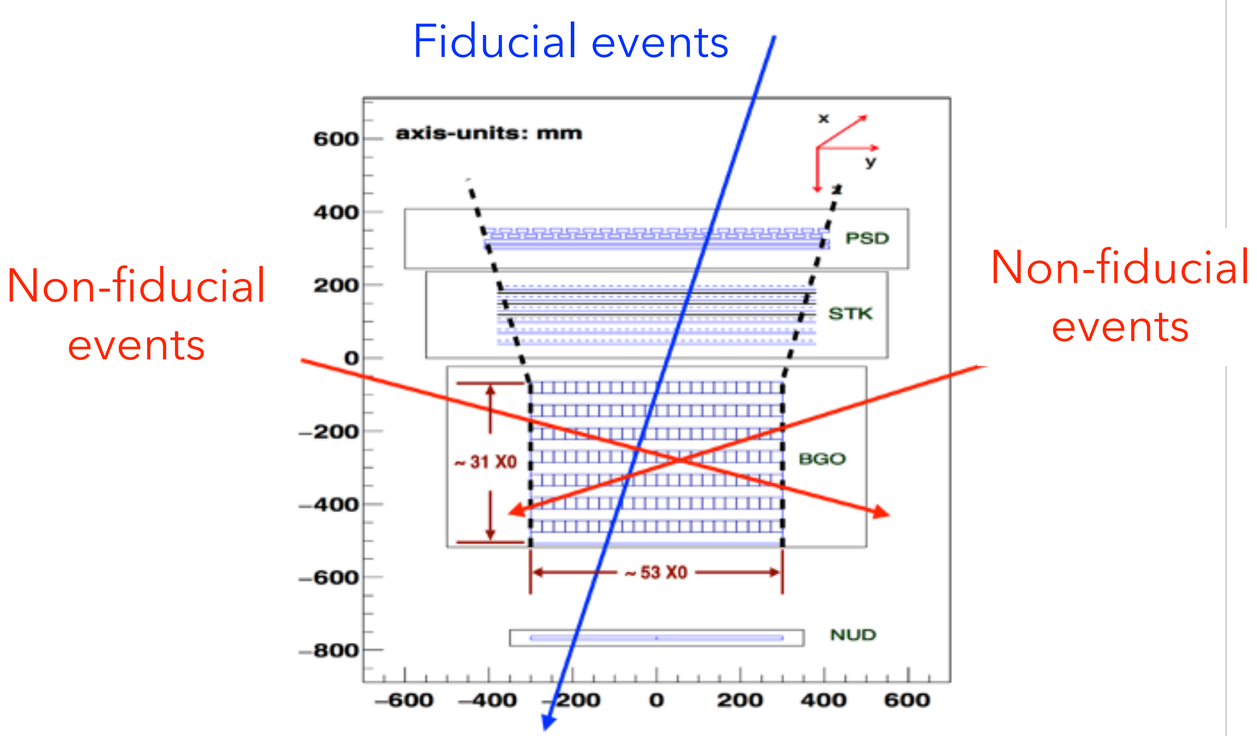
Motivation



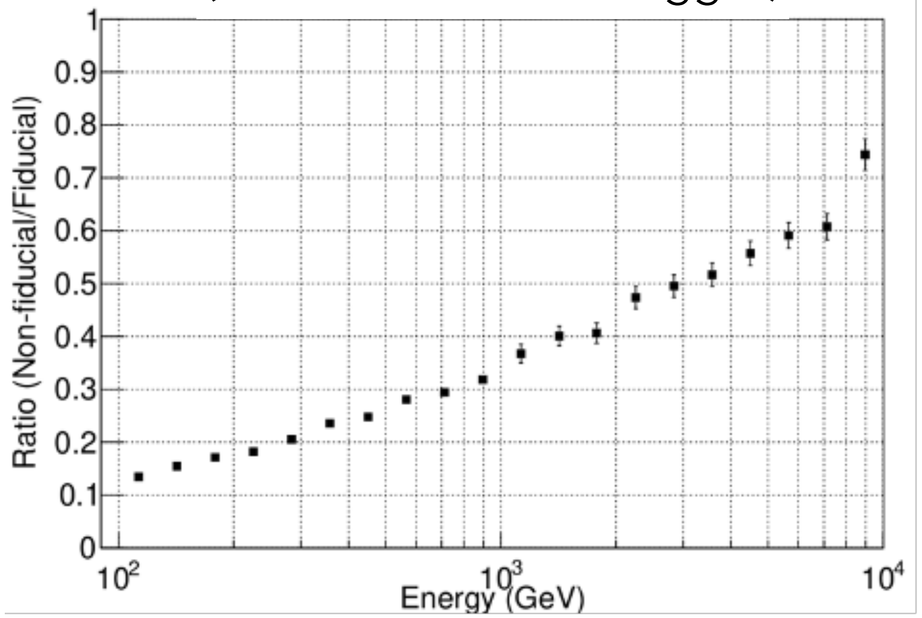
- In order to improve the e^++e^- measurement beyond TeV, several approaches can be pursued:
  - **Improve energy calibration and measurement**
  - **Increase statistics**
  - **Enhance the performance of e/p separation**

Method

- Non-fiducial definition in this work: events penetrate the calorimeter **from one lateral side to the opposite side**
  - Take into account additional events (compare to the fiducial analysis)
  - These events traverse over **~50 X_0**, improving e/p identification



Expected statistics gain
(ideal case with G3 trigger)



Challenge

- How to select lateral-penetrating events?
  - No track reconstruction → unable to infer hit geometry
- Possible solutions
  - Reconstruct the track using calorimeter information (PMO colleagues' work, with deep learning method)
  - Classifying lateral-penetrating events using **calorimeter features & XGBoost** (this work)

Data Sample

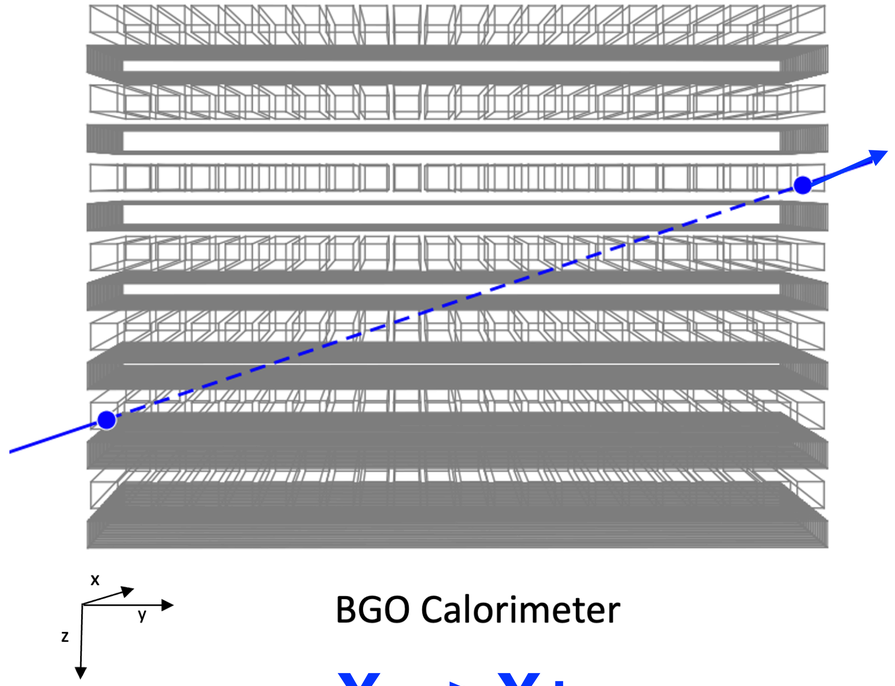
- Flight data:
  - 9 years (2016.01.01-2024.12.31)
  - Exposure time: 2.17×10^8s
- MC Sample:
  - Electron (upper): 1 - 10 TeV, QGSP_FTFP_BERT_HP, **99.5M**
  - Electron (lower): 1 - 10 TeV, QGSP_FTFP_BERT_HP, **50.3 M**
  - Proton (upper): 1 - 10 TeV, FTFP_BERT, **159.8M**
  - Proton (lower): 1 - 10 TeV, FTFP_BERT_HP, **137.4M**

Note:
- Upper hemisphere: standard scenario, zenith angle in [0°, 90°]
- Lower hemisphere:, zenith angle in [90°, 130°]
- Proton > 10 TeV: essential, but not included in this version

Analysis Procedure

- Input Flight data/MC sample
  - Pre-selection
  - XGBoost Classification (trained on mixed e/p sample)
    - ✓ **Classifier 1**: identification of lateral-penetrating events
    - ✓ **Classifier 2**: identification of incident direction (along the X-axis or Y-axis)
    - ✓ **Classifier 3**: identification of incident vertex (plus→minus or minus→plus)
  - e/p separation
  - Flux calculation

Training sample after pre-selection:
- electrons: 2.6M (upper) + 0.95M (lower)
- protons: 1.2M (upper) + 1.7M (lower)

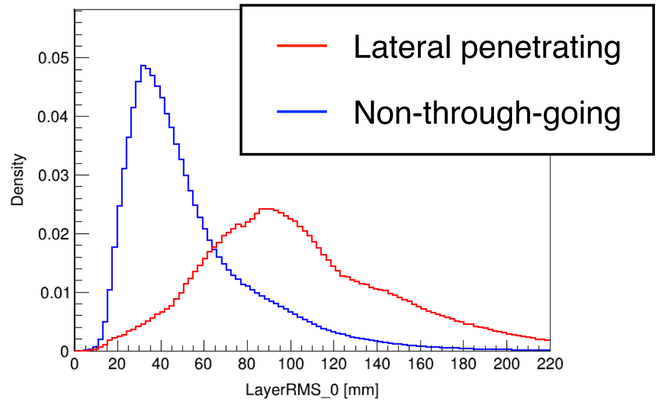


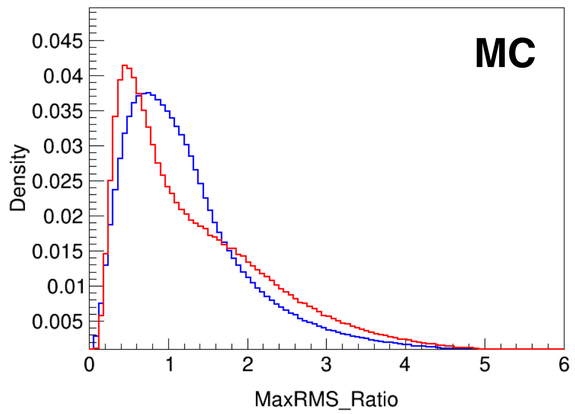
Pre-Selection

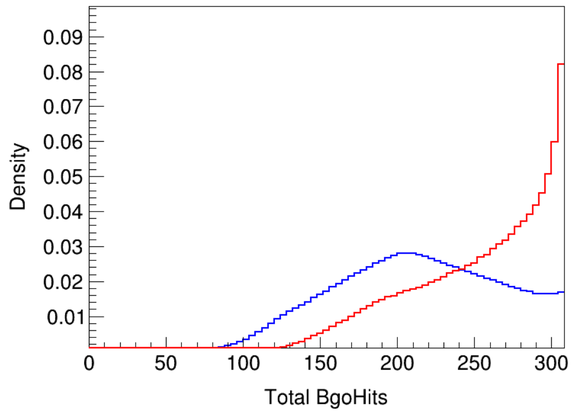
- Pass high energy trigger (G3)
- Total energy in [1, 10] TeV (focus on [1,10] TeV; extensible to lower/higher energies)
- STK veto: remove events with well-reconstructed STK tracks

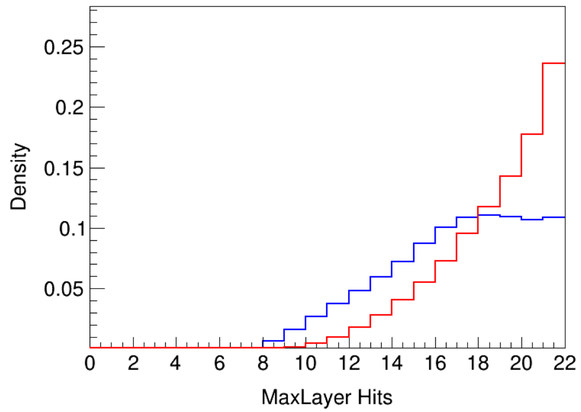
XGB Classifier 1

- **Classifier 1**: identification of lateral-penetrating events
- Typical feature profiles:

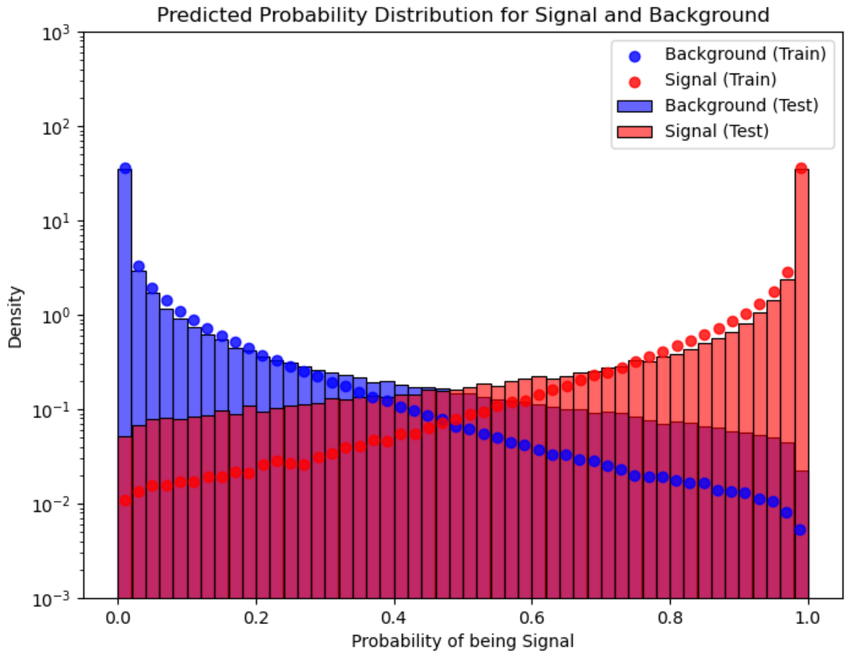






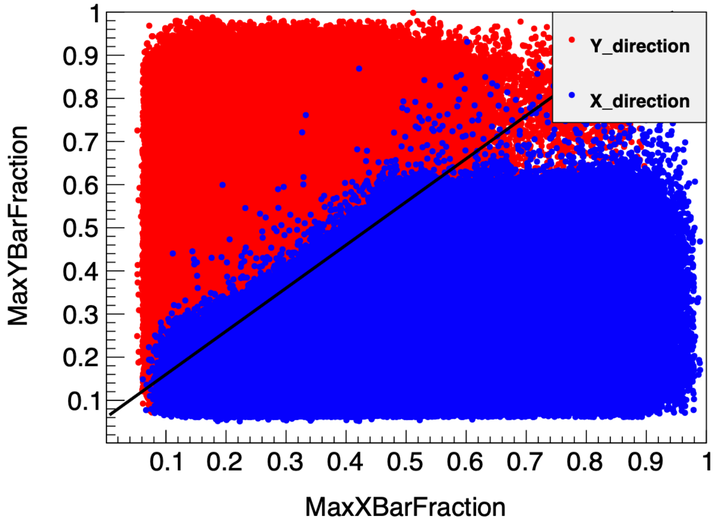


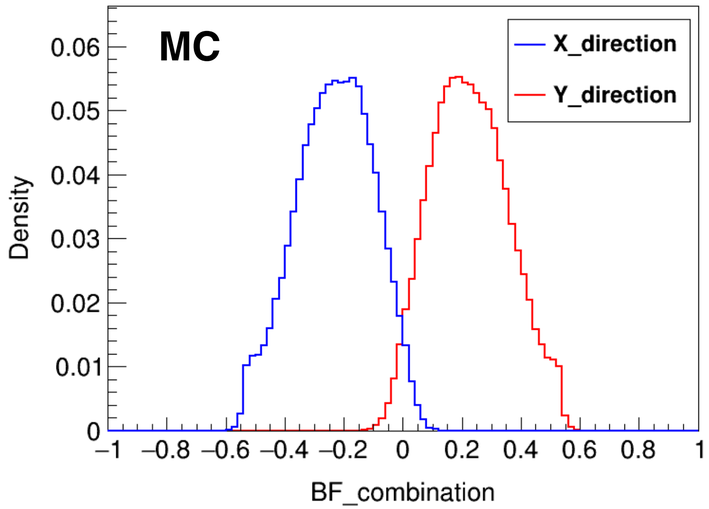
- Performance of Classifier 1:
  ~86.5% signal efficiency with
  ~6.5% misclassification



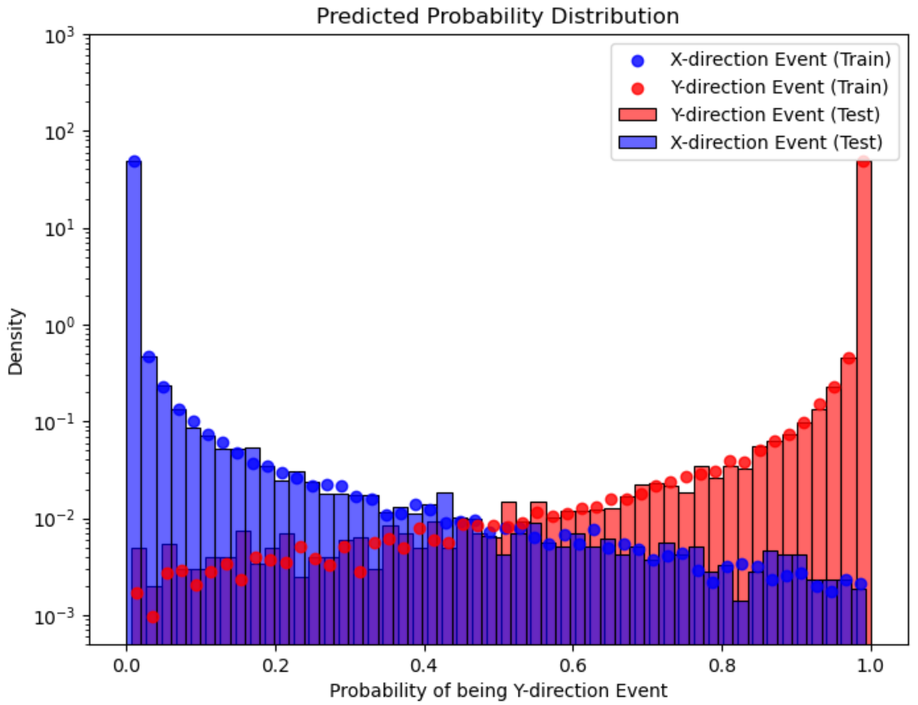
XGB Classifier 2

- **Classifier 2**: identification of incident direction
- Typical feature profiles:



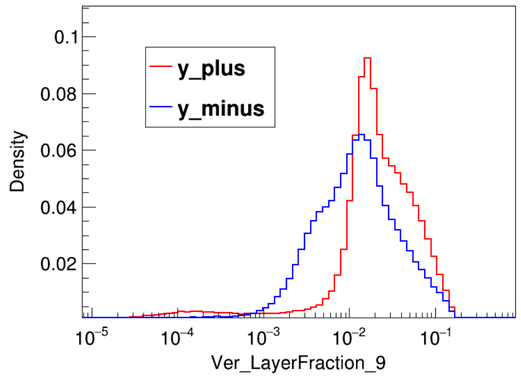


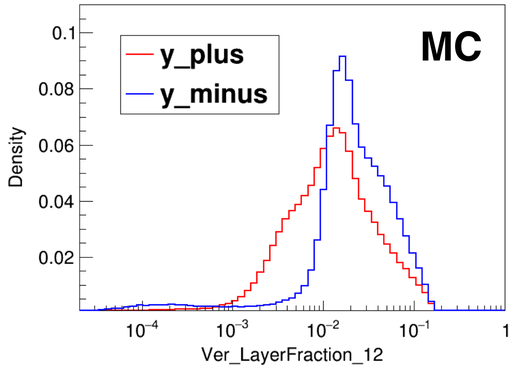
- Performance of Classifier 2:
  ~99.8% signal efficiency with
  ~0.2% misclassification

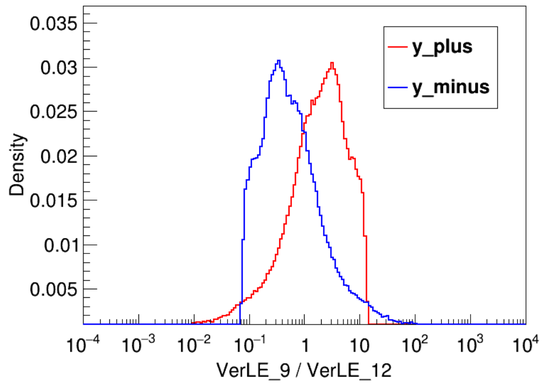


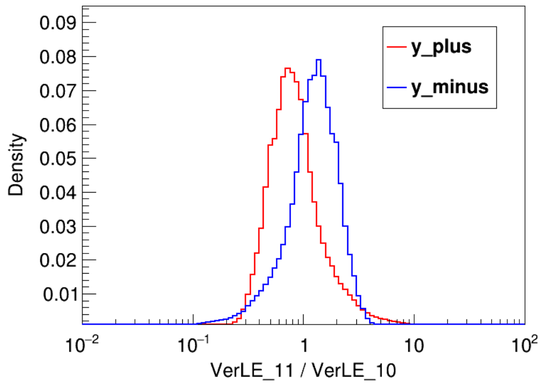
XGB Classifier 3

- **Classifier 3**: identification of incident vertex
- Typical feature profiles:

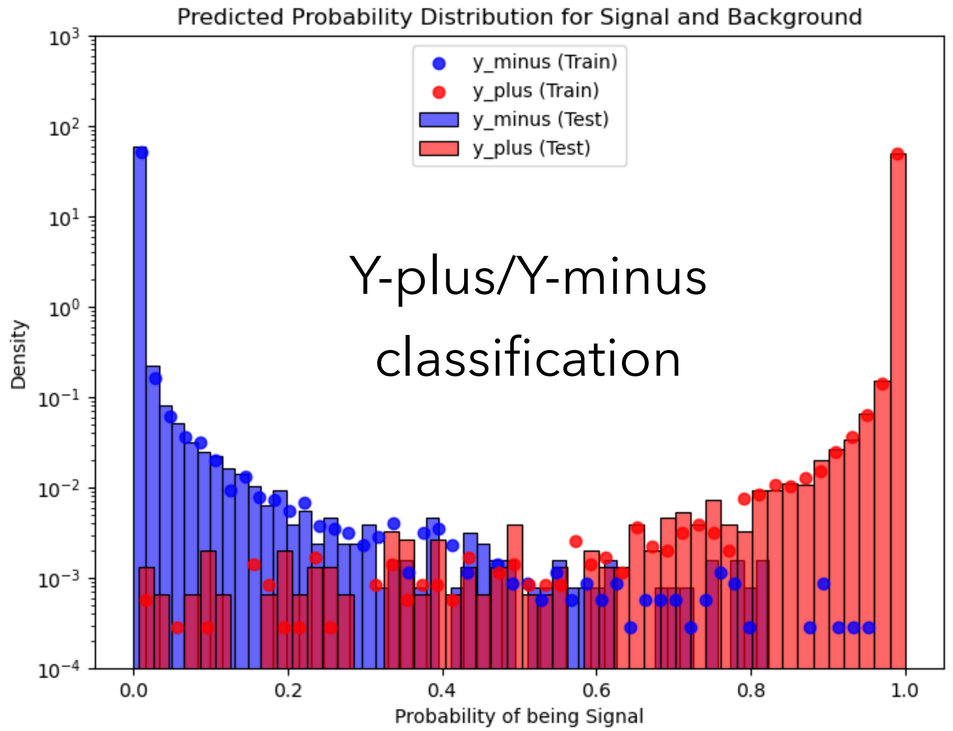






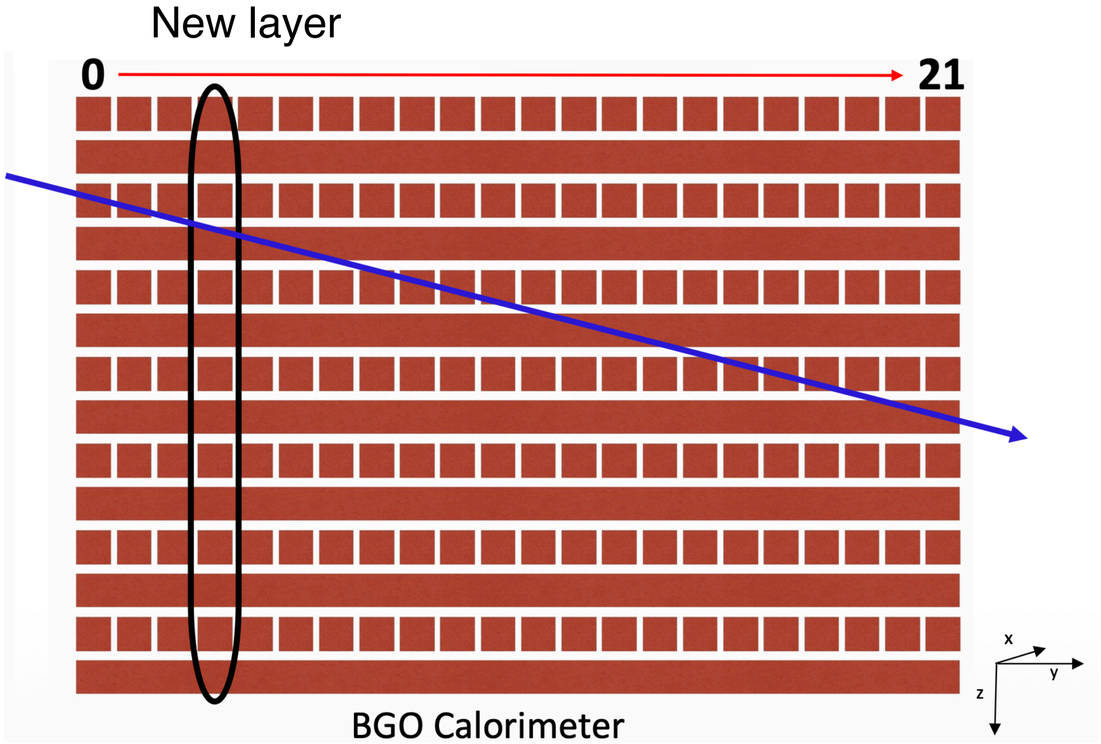


- Performance of Classifier 3: ~99.8% signal efficiency with ~0.2% misclassification



e/p Separation

- Following the e/p separation method in Nature (2017)
- New feature definition in the lateral-penetration coordinate frame
  - Define columns perpendicular to the dominant incident direction as new layers

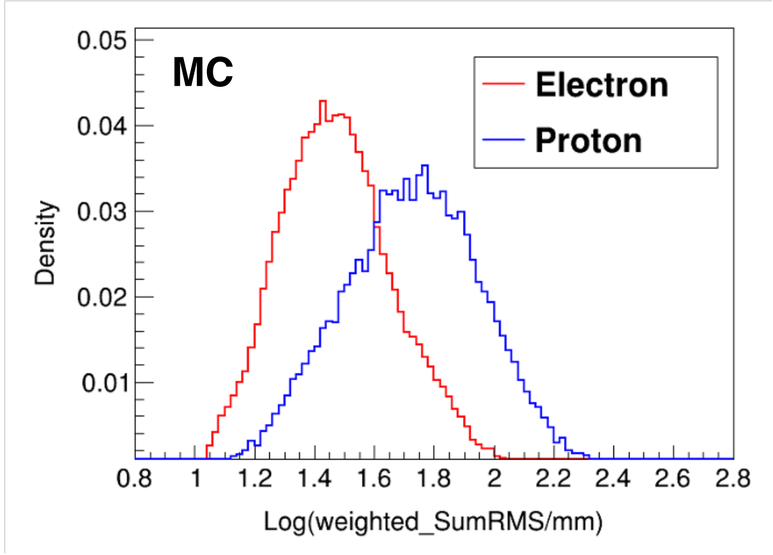


e/p Separation

- Following the e/p separation method in Nature (2017)
- New feature definition in the lateral-penetration coordinate frame
  - Define columns perpendicular to the dominant incident direction as new layers

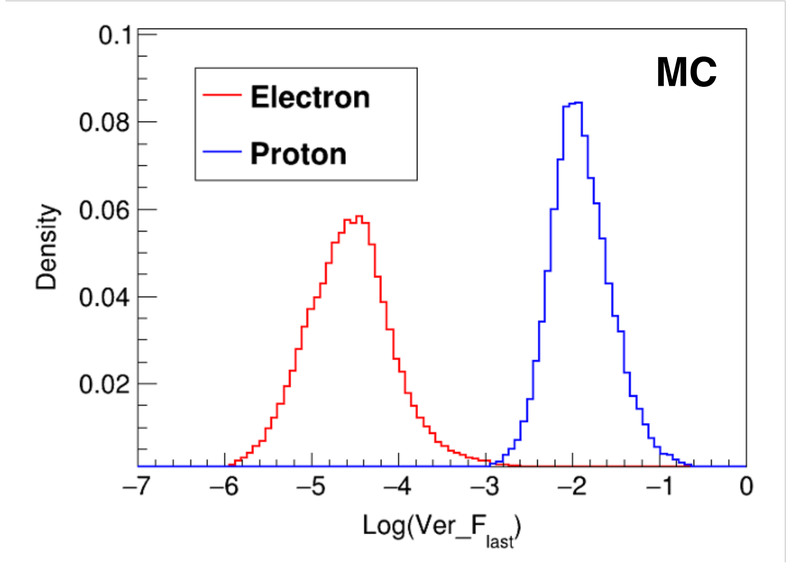
Features of shower **transverse** development

$$SumRMS = \sum_{i=0}^{6} \left(\frac{LayerEnergy_i}{MaxLayerEnergy} \times RMS_i \right)$$



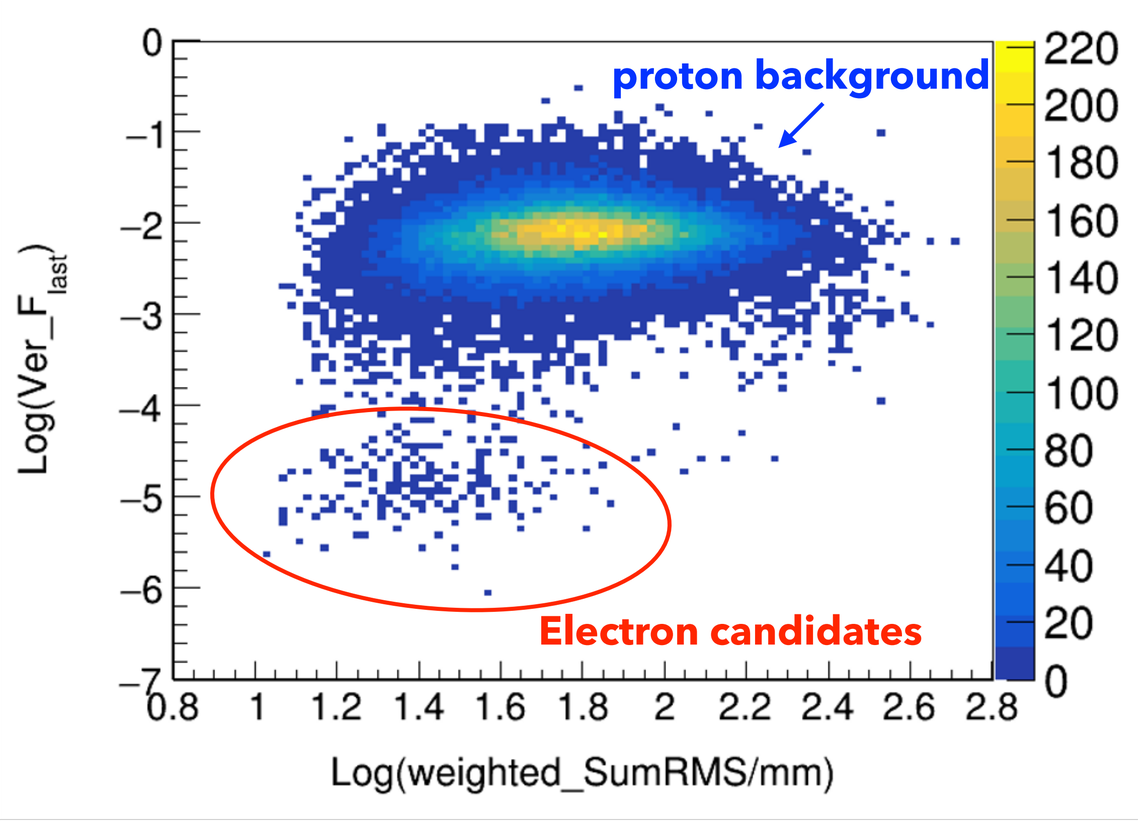
Features of shower **longitudinal** development

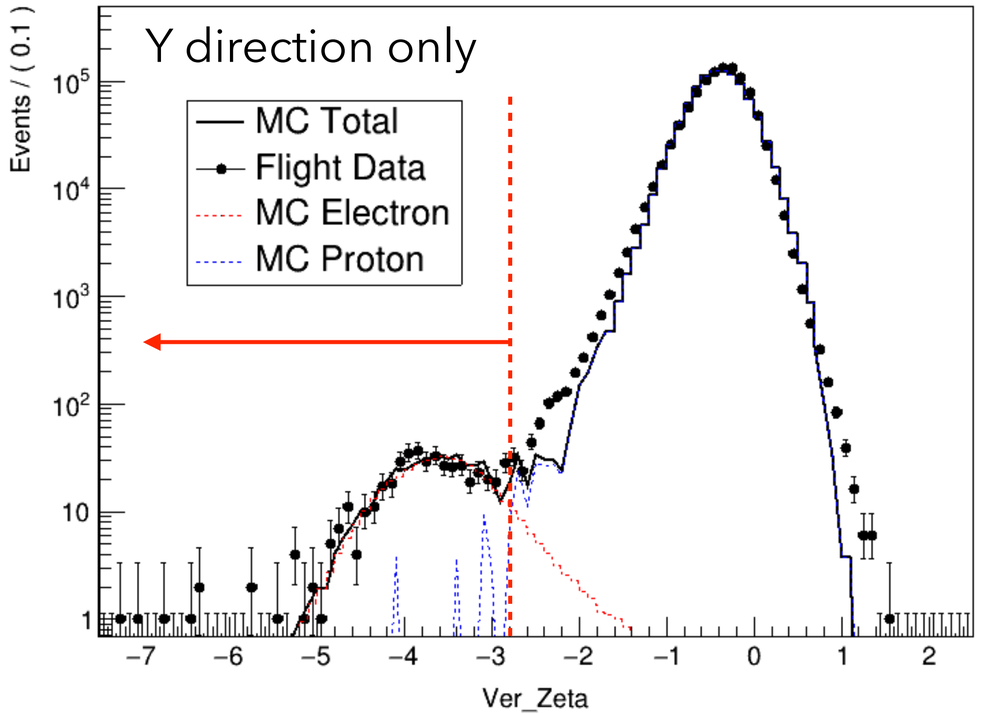
$$F_{last} = \frac{LayerEnergy_{21}}{TotalEnergy}$$



e/p Separation

- Apply the features of e/p separation on the flight data
- Define $Zeta' = \mathrm{Log}(F_{last}) + 0.8 \times \mathrm{Log}(SumRMS)$

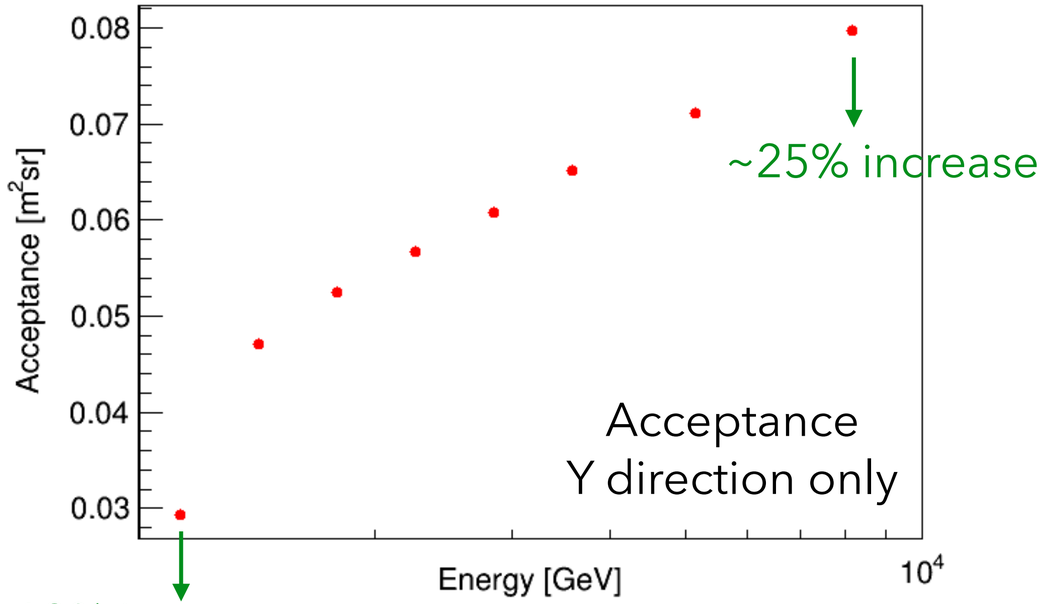




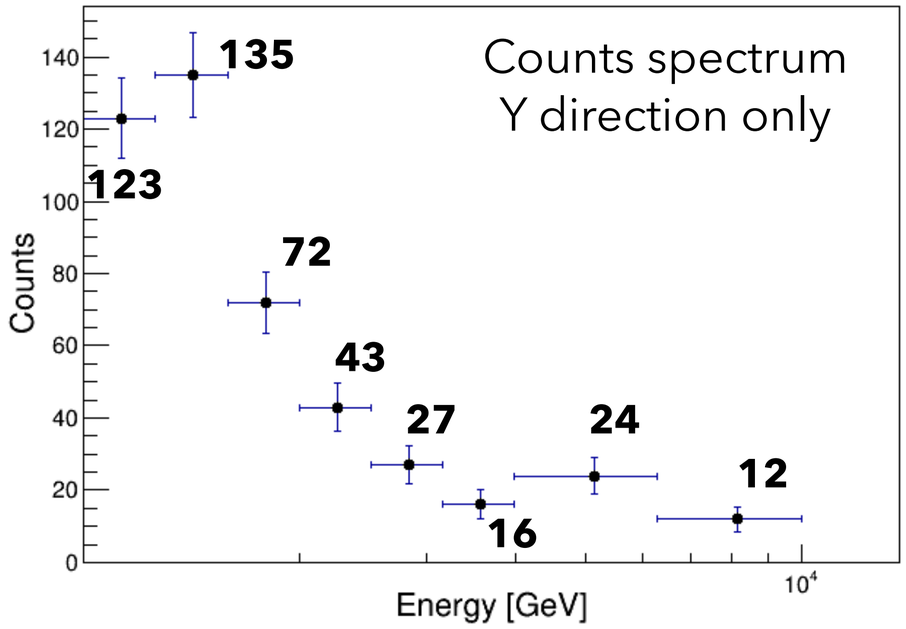
Selection cut: Zeta' < -2.8
Efficiency ~ 88.5%
Contamination ~ 4.6% (may be inaccurate)

Acceptance and Counts Spectrum

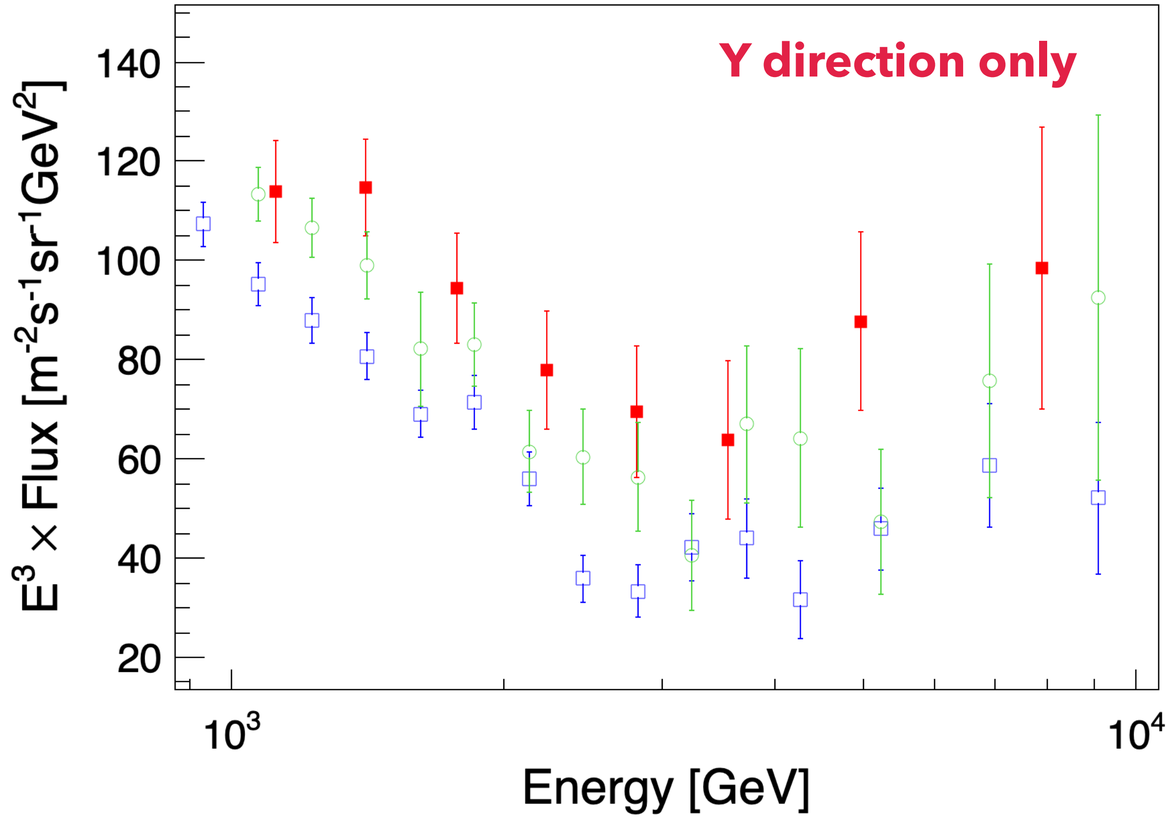
- 452 events above 1 TeV for Y direction, no background subtraction
- Comparable to 1.5 years of fiducial events in statistics
- Rough estimation: 9 years of lateral-penetrating events ≈ 3 years of fiducial events (in statistics)



Assuming fiducial acc ~0.3 m²sr



Preliminary Result



DAMPE non-Fidu, stand. Ereco, 9 years (this work)

DAMPE Fidu, new Ereco, 10 years (USTC)

DAMPE Fidu, stand. Ereco, 4.8 years (Geneva)

Note:

- Statistical uncertainties shown for this work
- Proton contamination is not well determined due to low statistic in MC proton sample
- XGB Classifiers will be retrained with extended MC proton sample above10 TeV
- XGB misclassifications are not taken into account

Summary & Next Plan

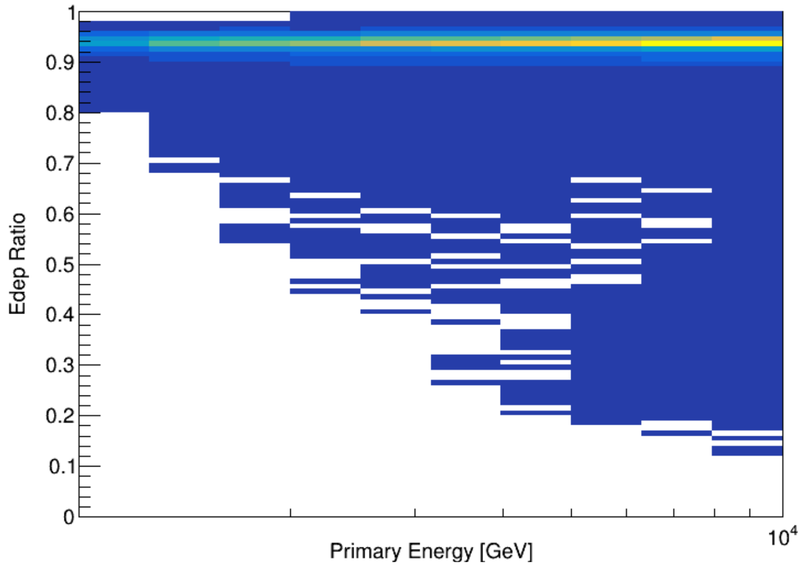
- A new method based on XGBoost has been developed for non-fiducial (lateral-penetrating) electrons
- Preliminary results agree with fiducial electron flux within uncertainties
- A rough estimation: including all lateral-penetrating events (both X and Y) boosts statistics by ~30%
- Next plan:
  - Generate additional MC samples
  - Investigate XGBoost misclassification, proton/nuclei contamination, and systematic uncertainties
  - Enhance electron/proton separation using ML techniques

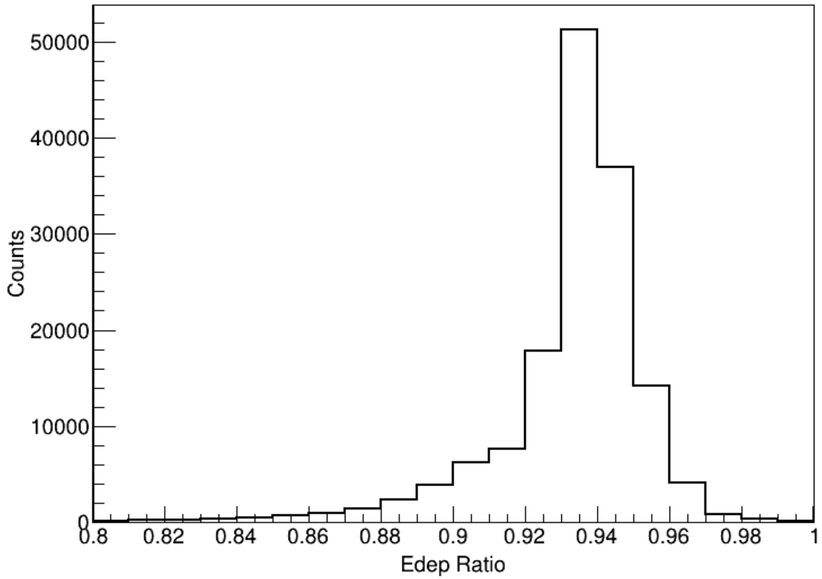
Thank you!

Backup

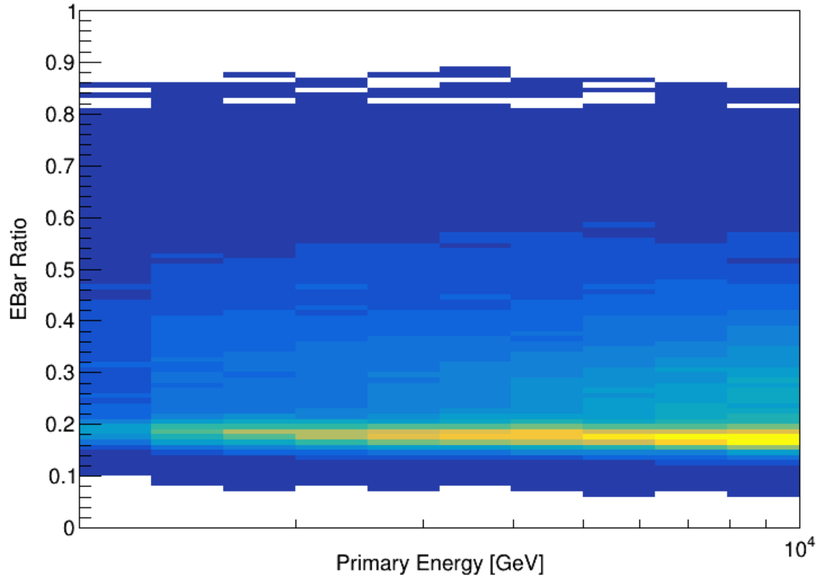
Energy Fraction

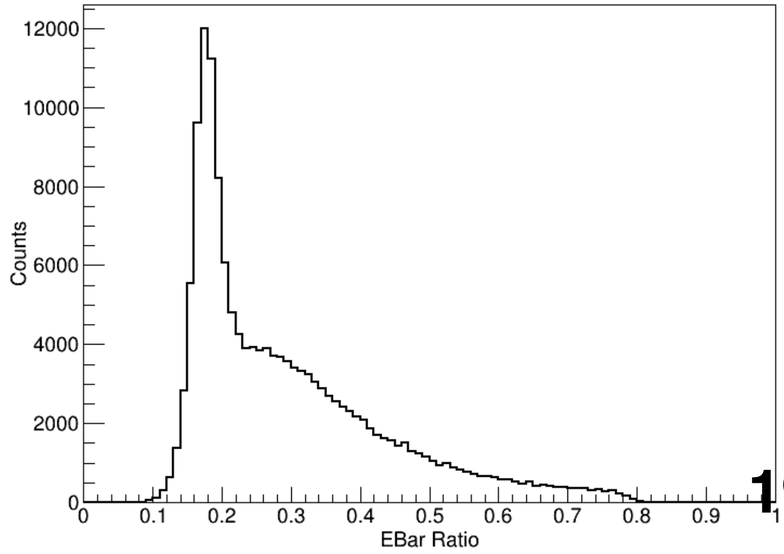
- BgoE/PrimE





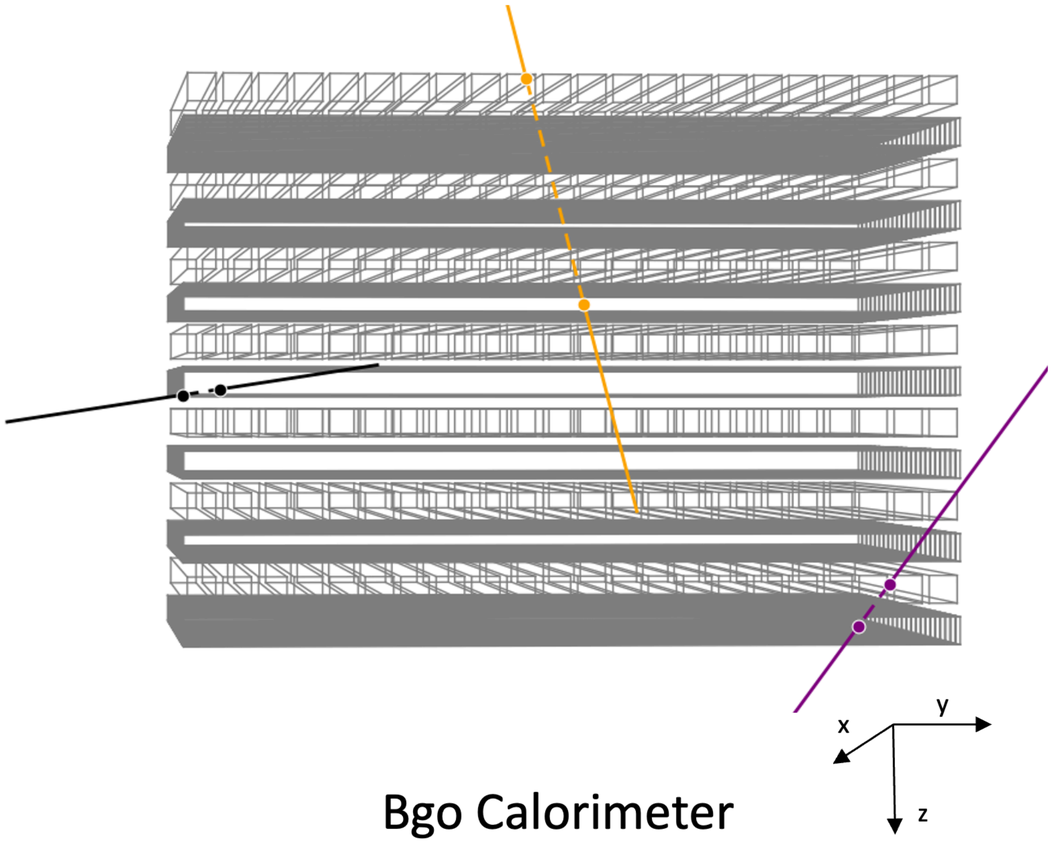
- MarBarE/ BgoE





Event Definition

- Misclassification types:
  - Else type 1: Top-to-Side Event
  - Else type 2: Side-to-Bottom Event
  - Else type 3: Side-to-Side Event



All Features Used

- **Variables used by training**

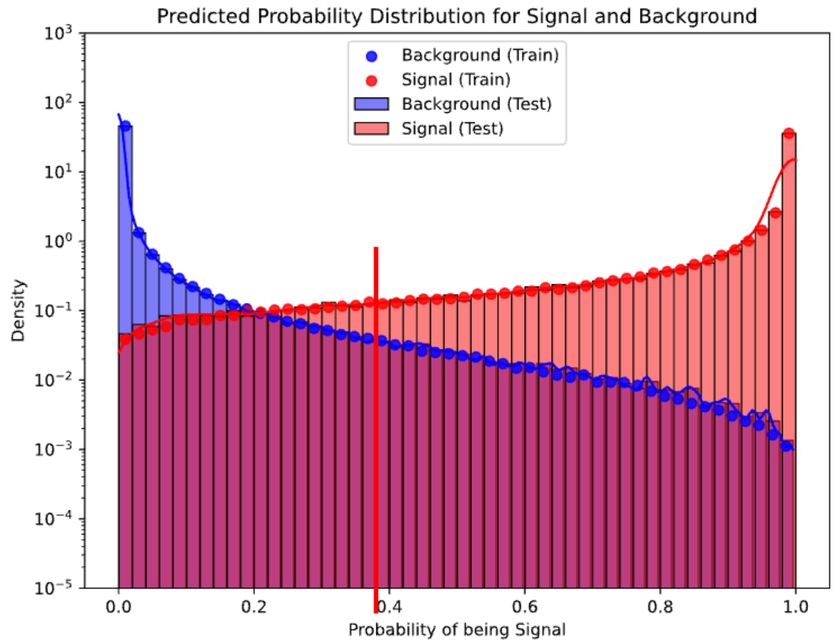
- RMS
- $LayerFraction$
- $FValue$: $FValue_i = RMS_i^2 \times LayerFraction_i$
- $nBarLayer$
- $Center\ of\ Gravity$: $CoG_i = \frac{\sum_j E_{i,j} \times x_{i,j}}{\sum_j E_{i,j}}$
- $Max\ X\ LayerFraction$
- $Max\ Y\ LayerFraction$
- $Max\ X\ BarFraction$: $MaxXBF = \frac{Max\ Bar\ Energy}{Layer\ Energy}$
- $Max\ Y\ BarFraction$: $MaxYBF = \frac{Max\ Bar\ Energy}{Layer\ Energy}$
- $Max\ RMS\ Area^2$: $Area^2 = (RMS_{max}^x \times RMS_{max}^y)^2$
- $TRMS$
- $LRMS$
- $RMS\text{ratio}$: $\frac{LRMS}{TRMS}$
- $Zeta$: $Zeta = \mathcal{F}_{last} \times (\sum_i RMS_i/mm)^4/(8\times10^6)$
- $Max\ RMS\ Ratio$: $\left(\frac{Max\ RMS_y}{Max\ RMS_x}\right)^2$
- $Total\ Hits$
- $Max\ Layer\ Hits$
- $Max\ Layer\ Hits\_1$
- $Max\ Layer\ Hits\_2$

Features Used in Classifier1

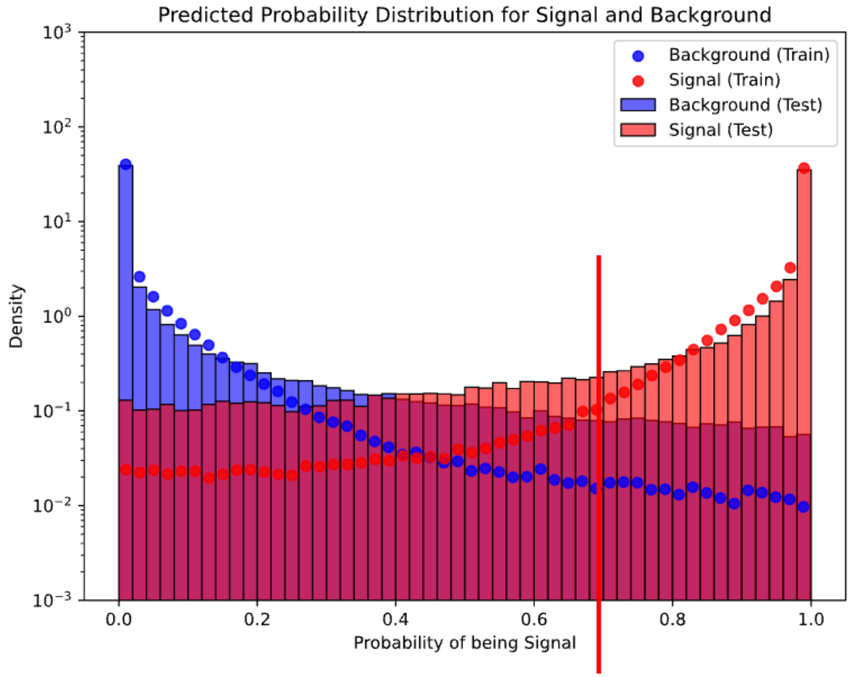
- **Top Else1&2 Eliminate**
  - $Layer\ RMS$: 0, 4~9
  - $FValue$: 0~13
  - $LayerFraction$: 0~13
  - $nBarLayer$: 0~13
  - $Center\ of\ Gravity$: 0~13
- **Else3 Eliminate**
  - $Layer\ RMS$: 0~13
  - $FValue$: 0~13
  - $LayerFraction$: 0~13
  - $nBarLayer$: 0~13
  - $Center\ of\ Gravity$: 0~13
- **X-direction & Y-direction Classification**
  - $Layer\ RMS$: 0~5
  - $FValue$: 1~5
  - $nBarLayer$: 0~7
  - $Center\ of\ Gravity$: 6, 7, 8
  - $BF\_conbination$

XGB Classifier 1

- **Background: Top, Else1, Else2:**
  - **The signal efficiency is ~ 96.5% at a contamination of 5%.**



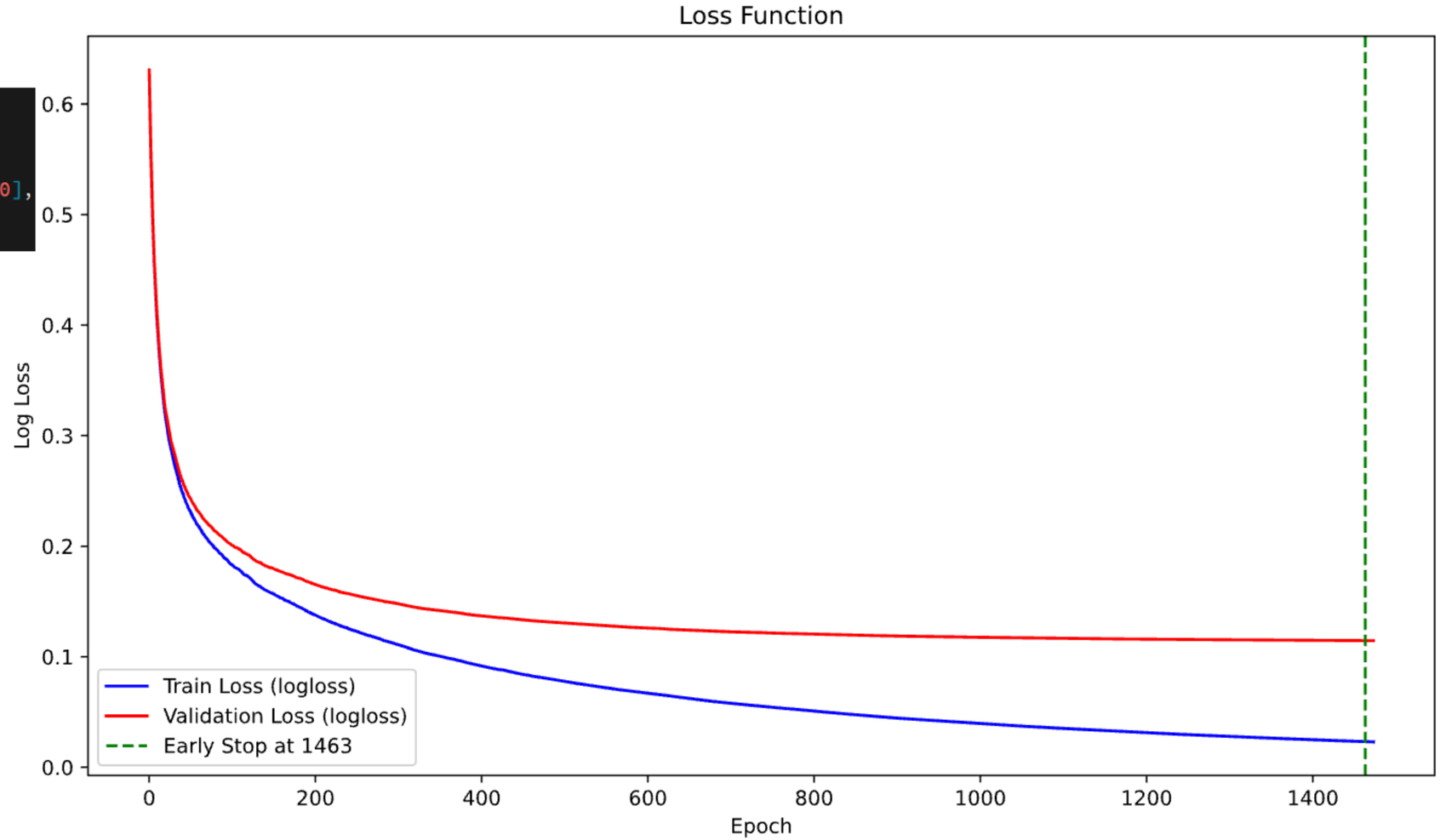
- **Background: Else3:**
  - **The Contamination is ~ 3.42% at a signal efficiency of 90%.**



XGB Classifier 1

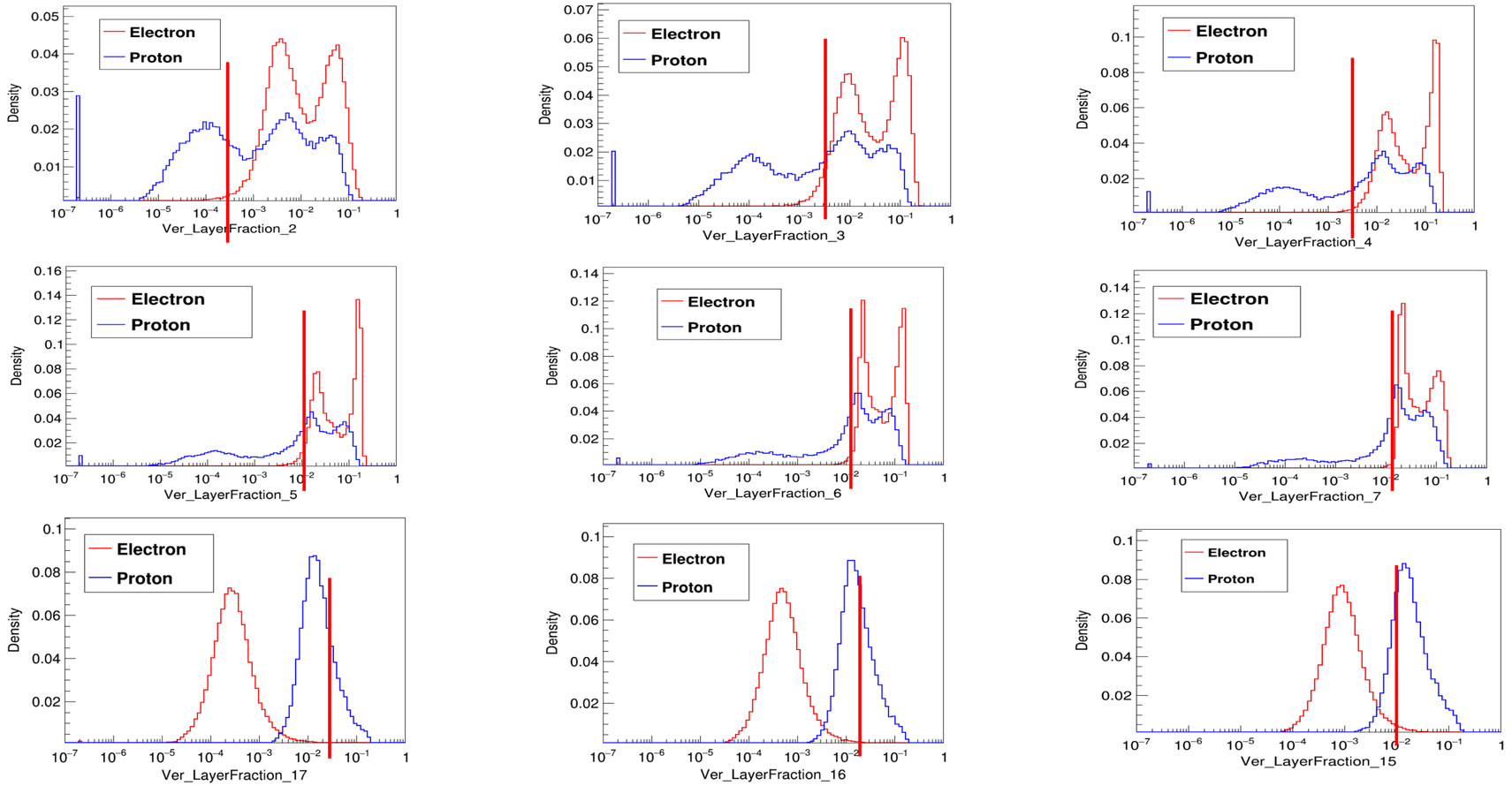
- For remove else3 events

```
param_grid = {
    'max_depth': [3,4,5,6,7,8],
    'learning_rate': [0.2,0.1,0.05],
    'n_estimators': [2000],
    'colsample_bytree': [0.7,0.8,0.9,1.0],
    'subsample': [0.7,0.8,0.9,1.0],
}
```



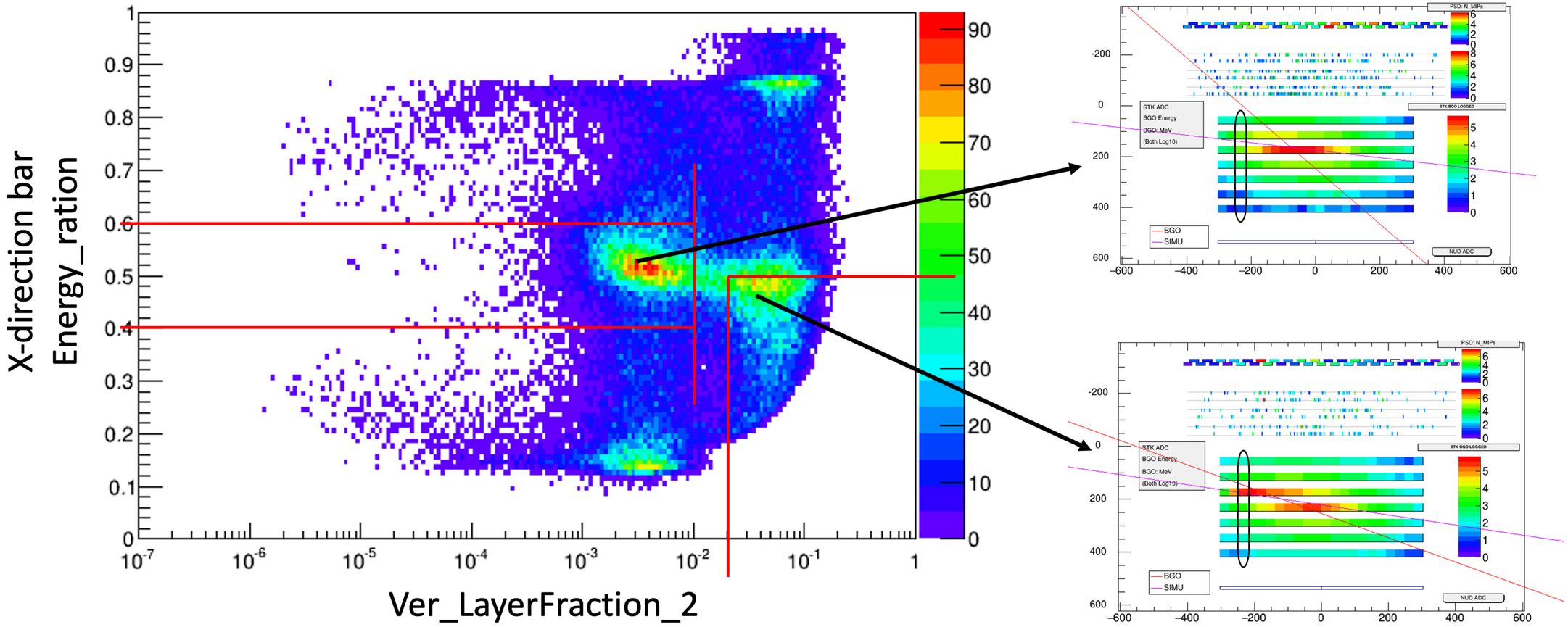
Features Study

- Multi-peak structures are found



Features Study

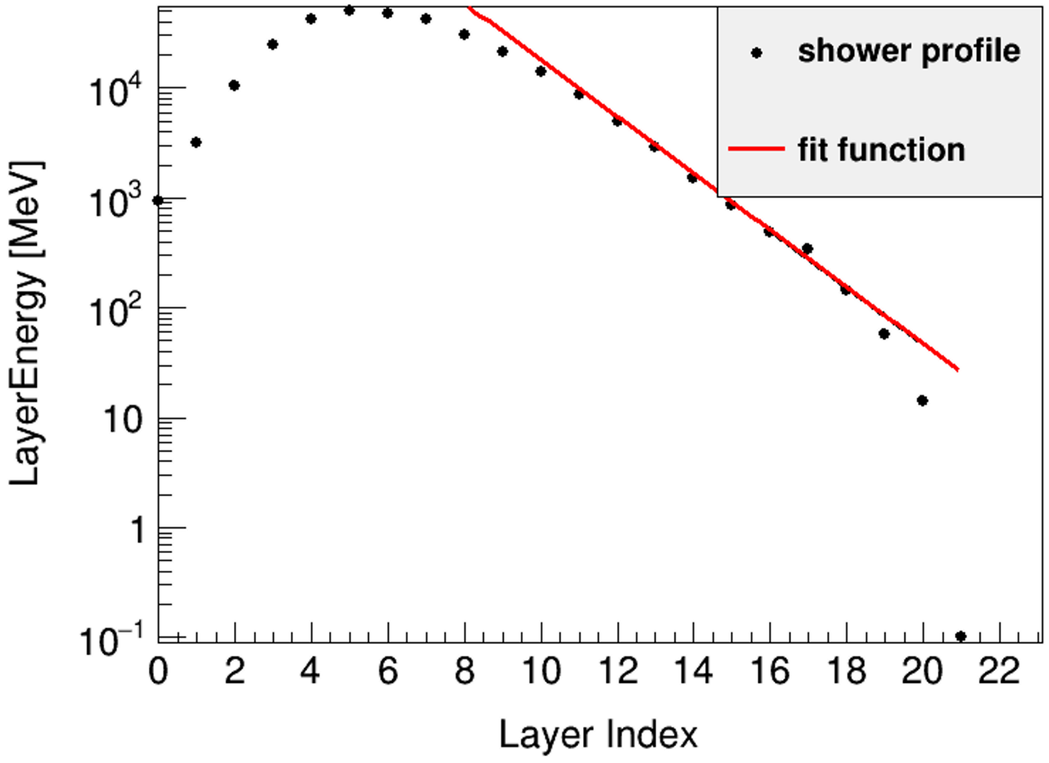
- Understanding of multi-peak structures

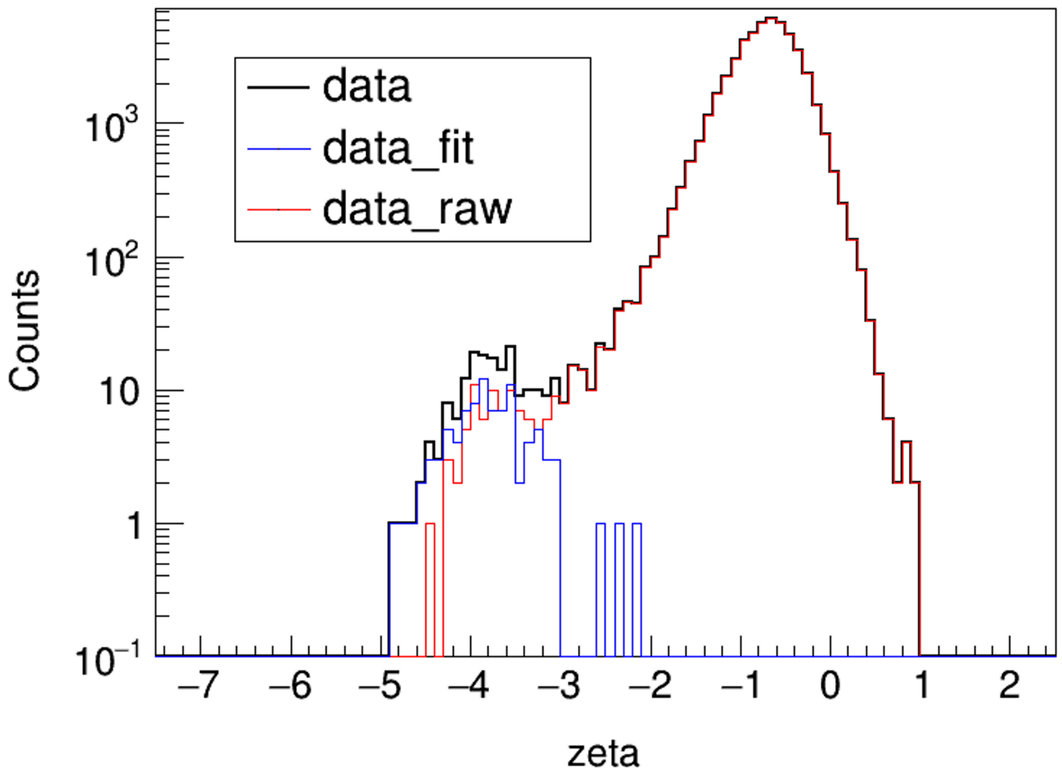


Features Study

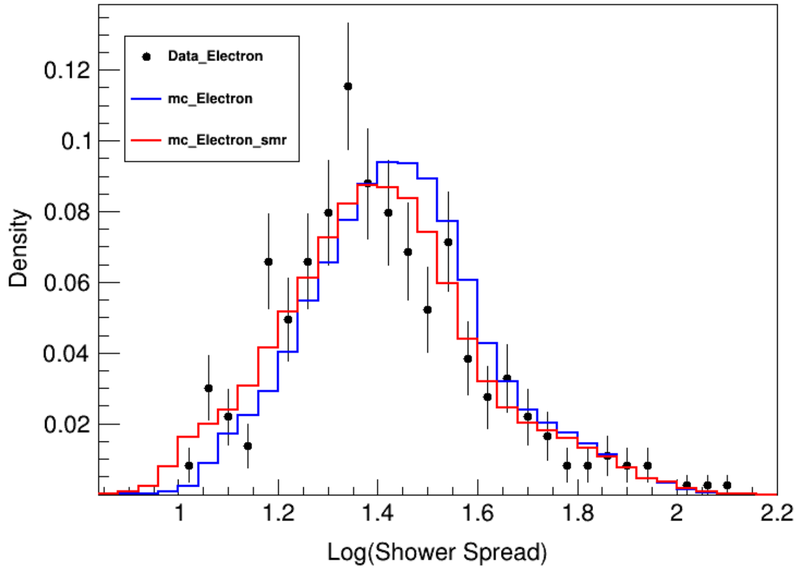
- Layer fraction extrapolation in the last layer

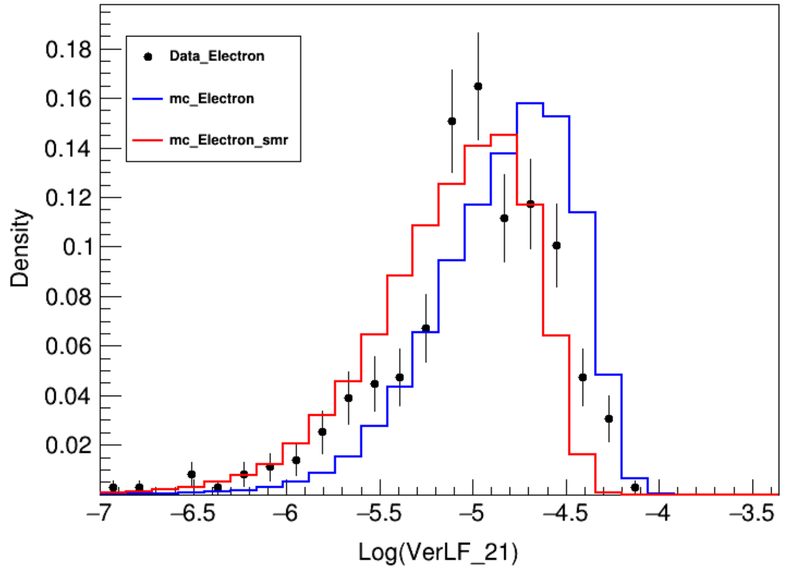
$$layer\ energy = a \cdot exp(b \cdot layer\_number + c)$$

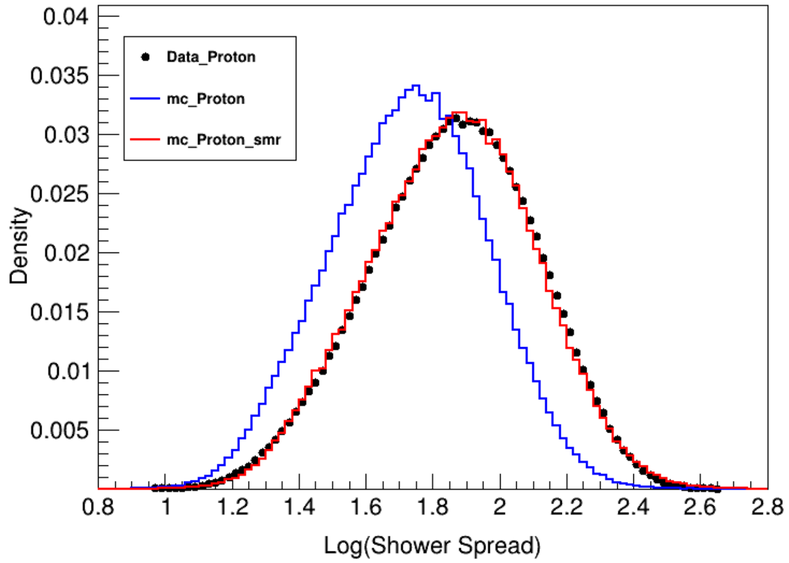


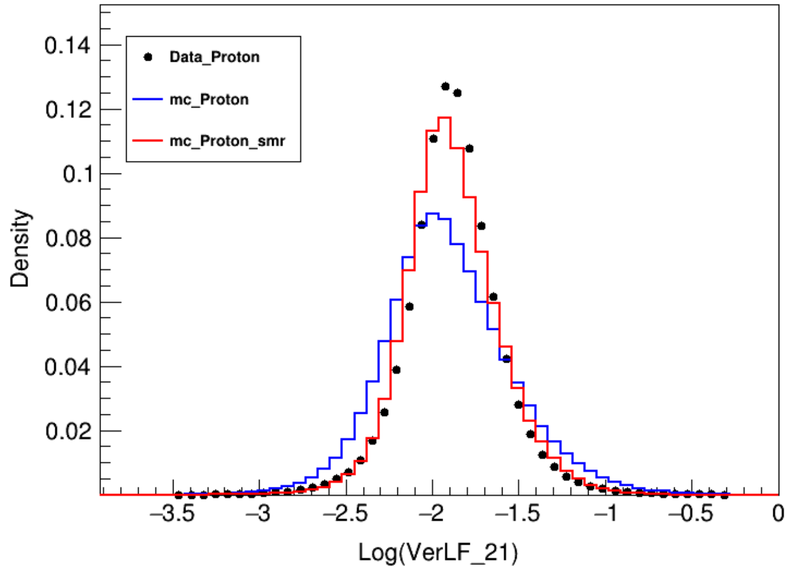


Features Smearing



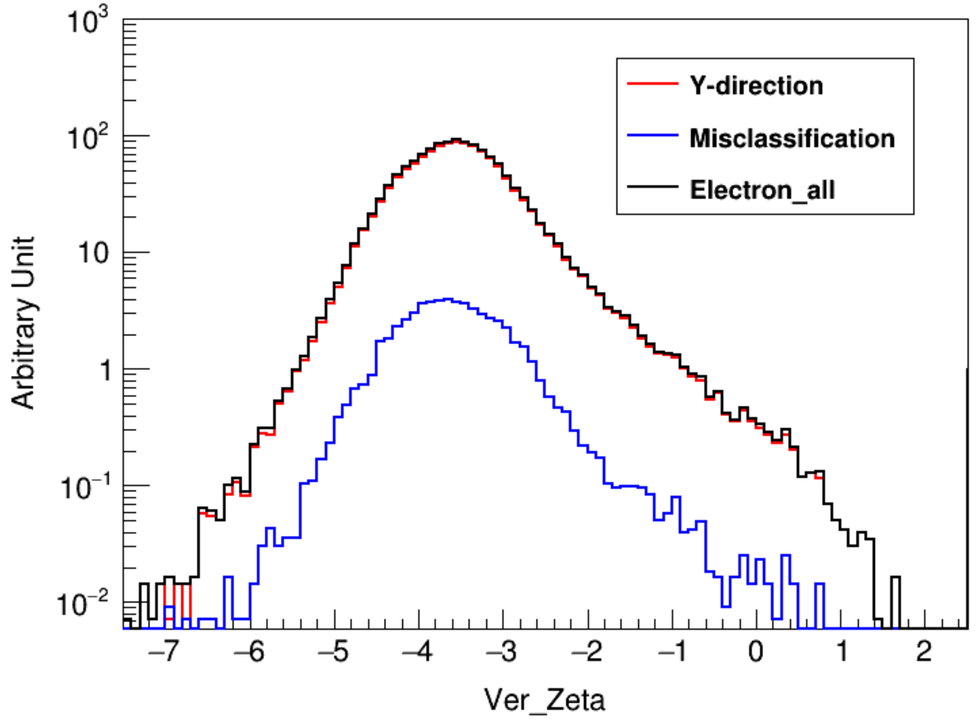




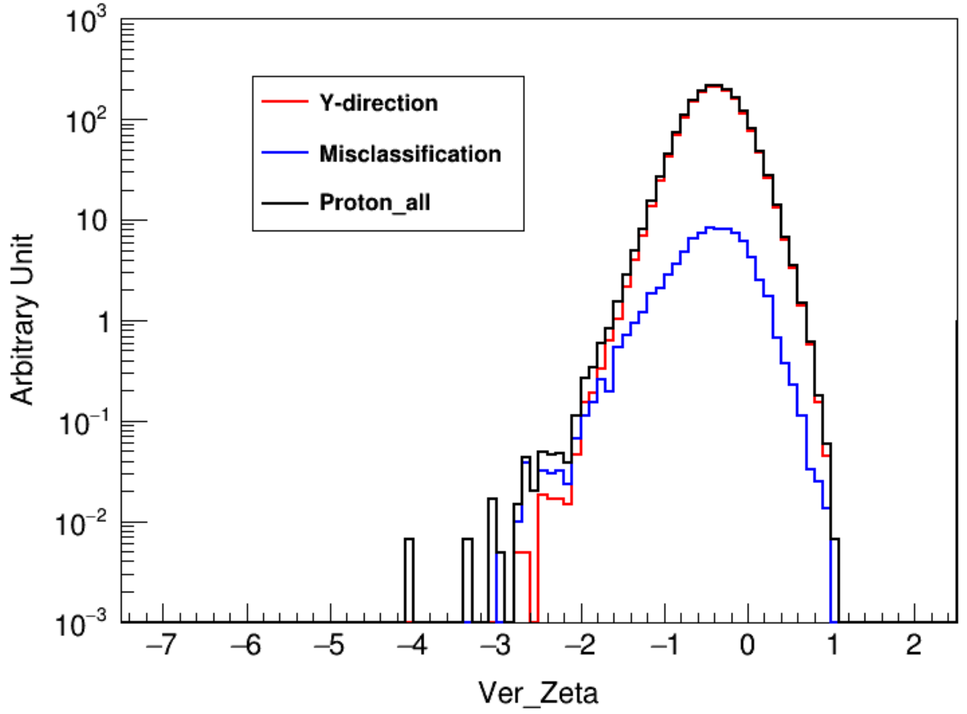


Misclassification Study

MC Electrons



MC Protons



e/p Separation