DEEP UNDERGROUND NEUTRINO EXPERIMENT

WARWICK
THE UNIVERSITY OF WARWICK

Track/Shower Discrimination Refresh in Pandora for DUNE FD

28th October 2019

Mousam Rai

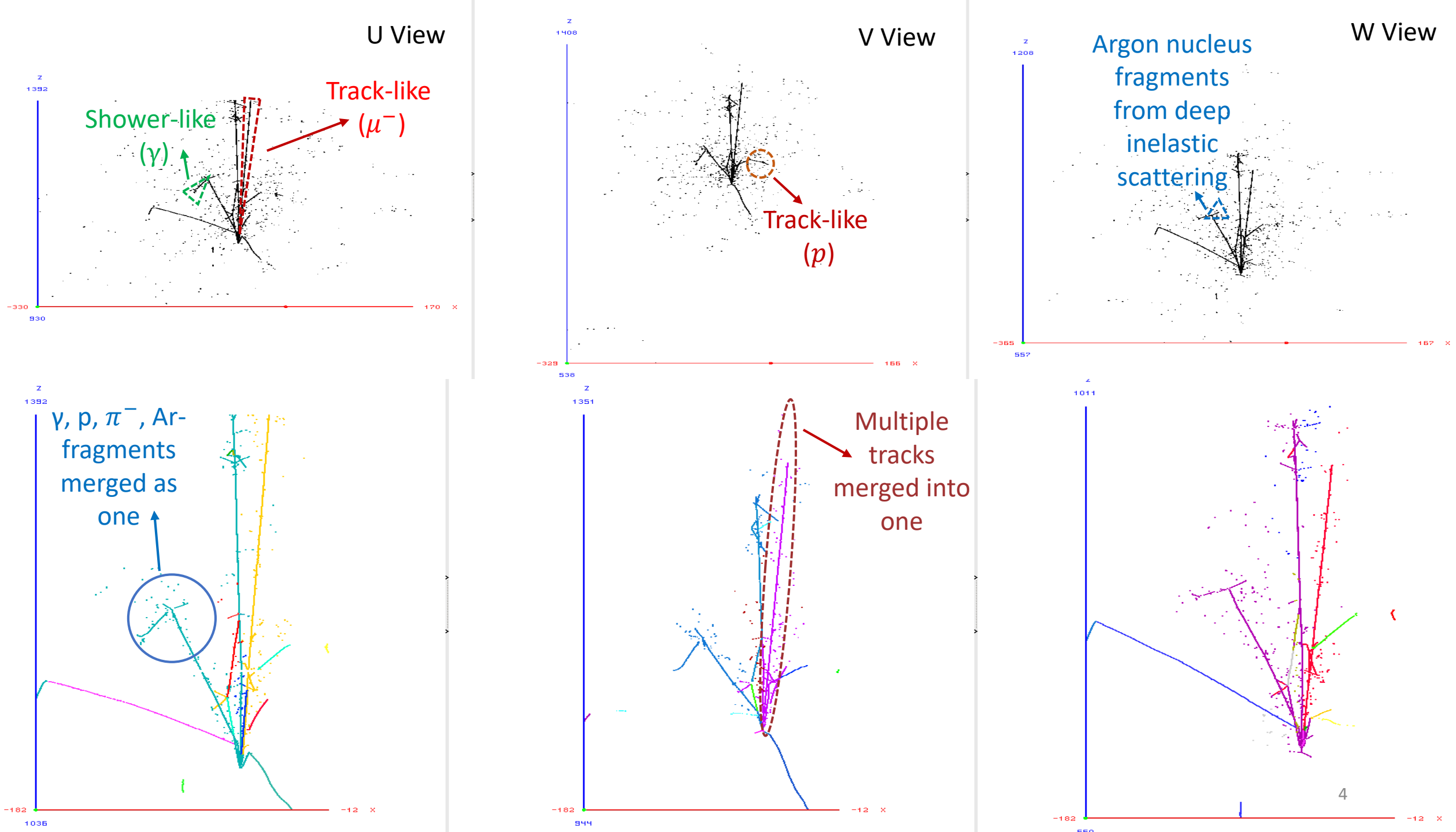
Supervisor: Prof John Marshall

Roadmap for this presentation

- The Problem
- MicroBooNE Variables
- Current Approach In Pandora
- BDT1, BDT2, and BDT3
- The End.

The Problem

Track vs Shower discrimination

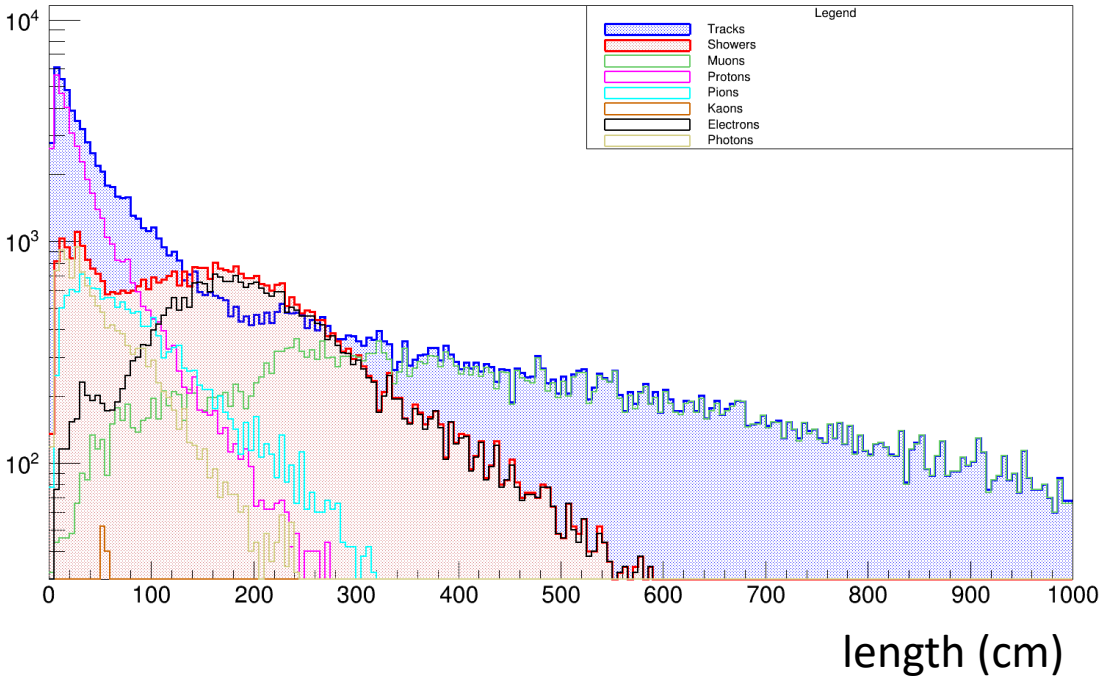

U View
Shower-like
(γ)
Track-like
(μ^-)
V View
Track-like
(p)
W View
Argon nucleus
fragments
from deep
inelastic
scattering
γ, p, π^-, Ar-
fragments
merged as
one
Multiple
tracks
merged into
one

MicroBooNE Variables

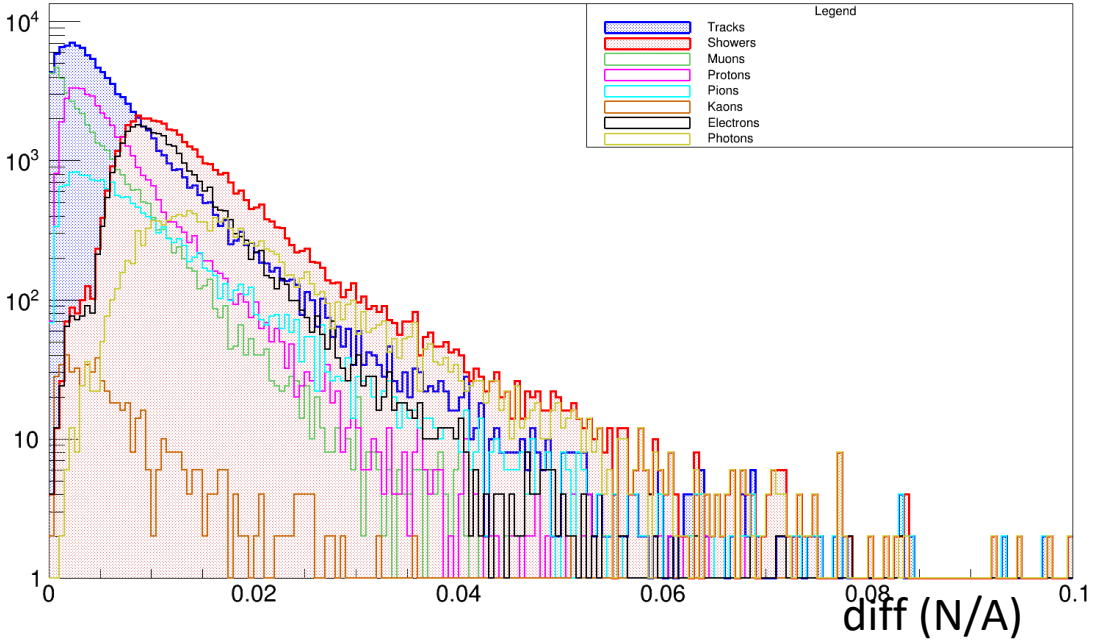
What are they?

Types of MicroBooNE Variables

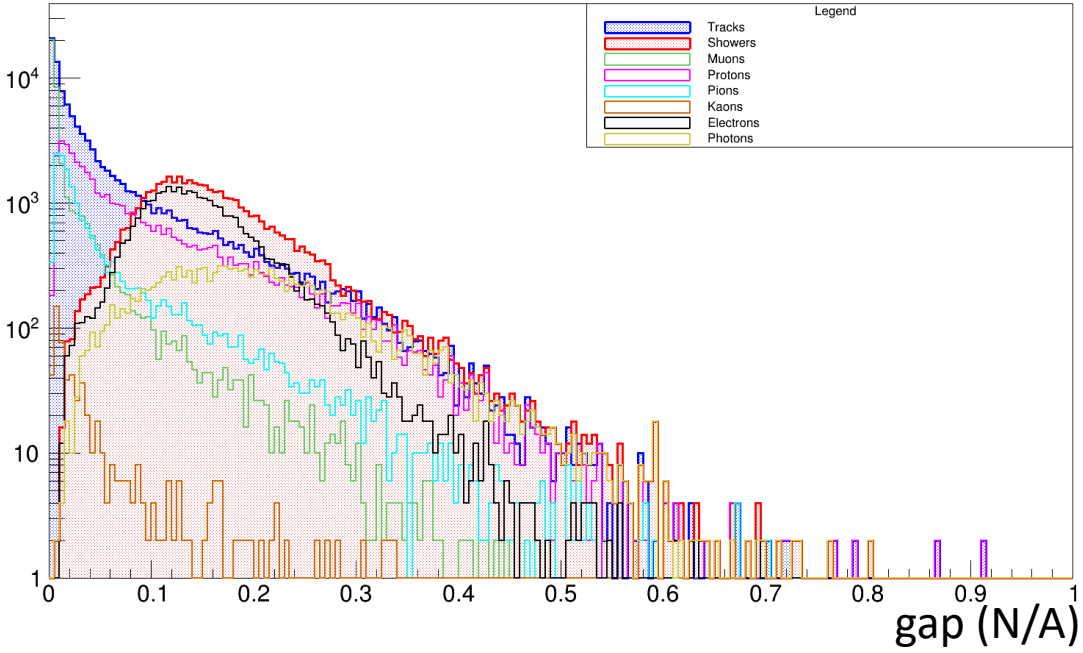
- 8 Topological Variables:
  - length
  - diff
  - gap
  - rms
  - vertexDistance
  - diffAngle
  - pca1
  - pca2
- 2 Calorimetric Variables:
  - charge1
  - charge2



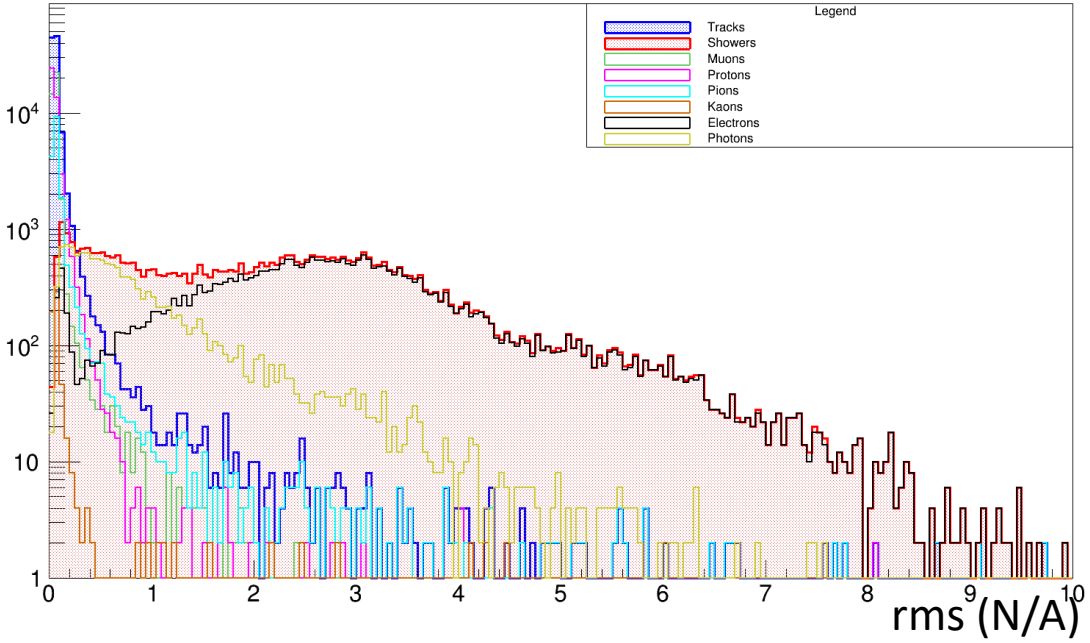
length – 3D length of the PFO



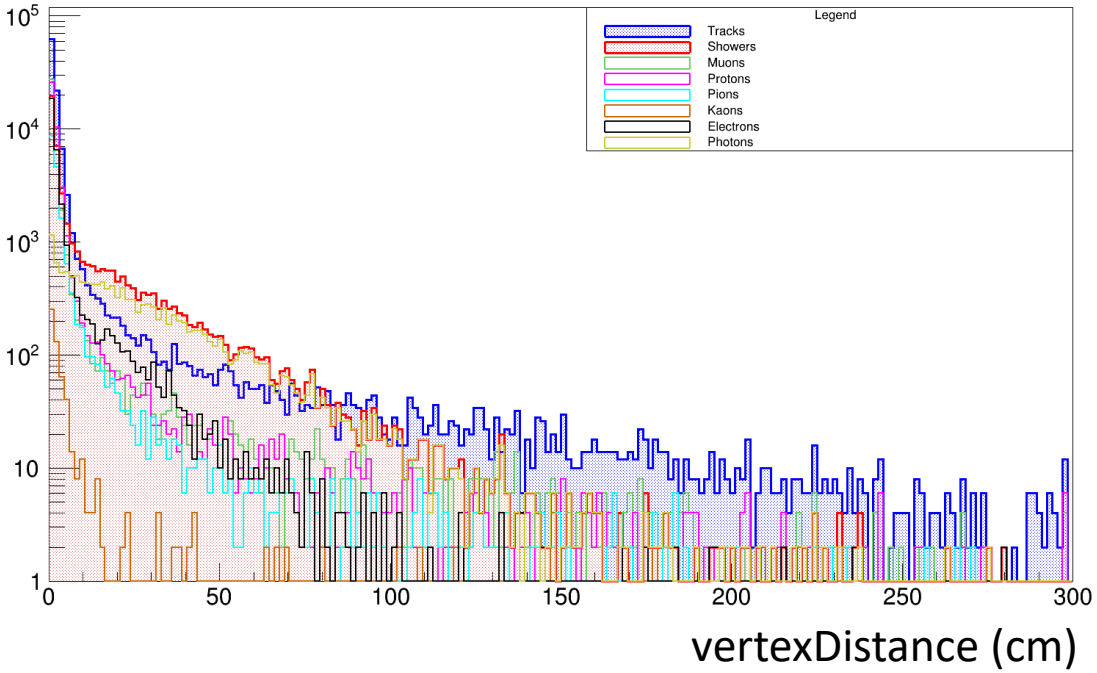
diff – Mean difference between the position of the hits and a straight line, divided by the straight line length



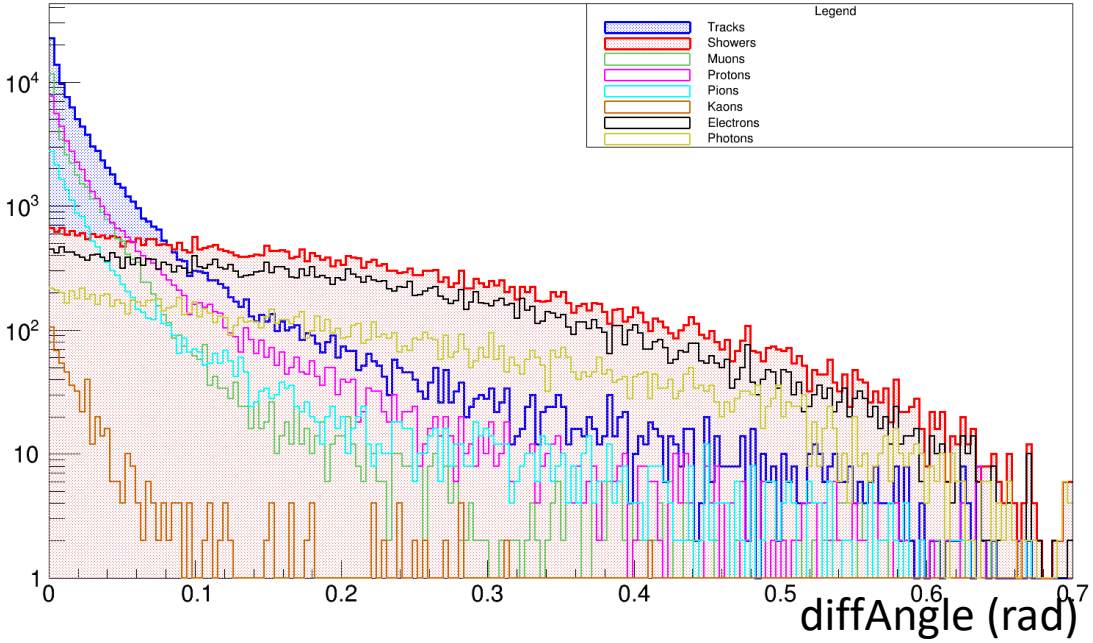
gap – Average max gap distance, divided by straight line length



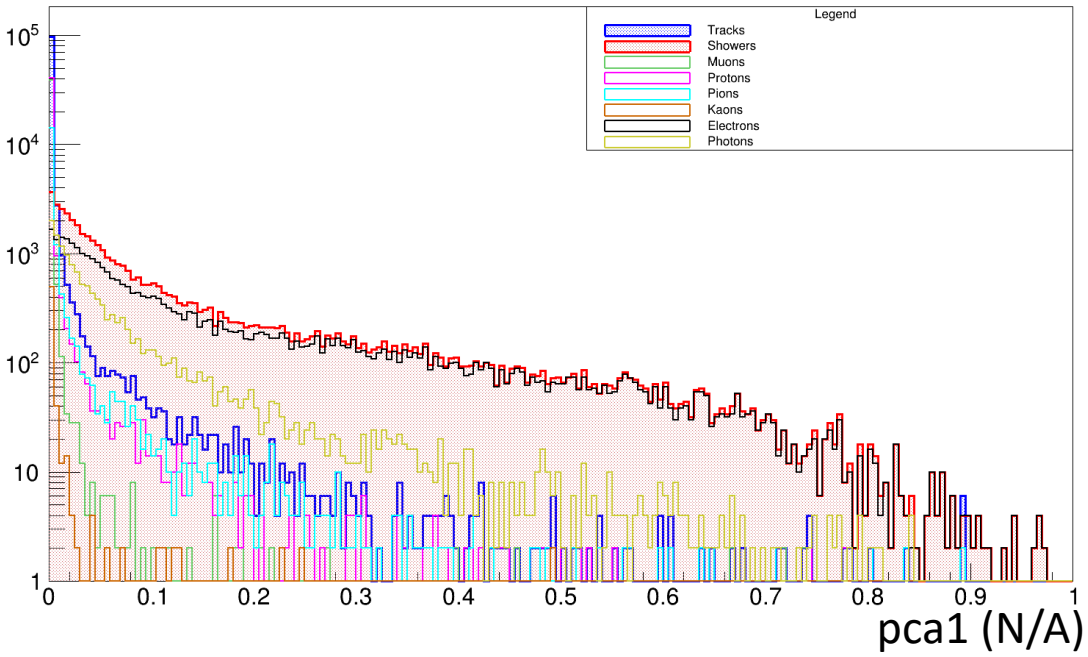
rms – Average root mean square of linear sliding fit, divided by straight line length



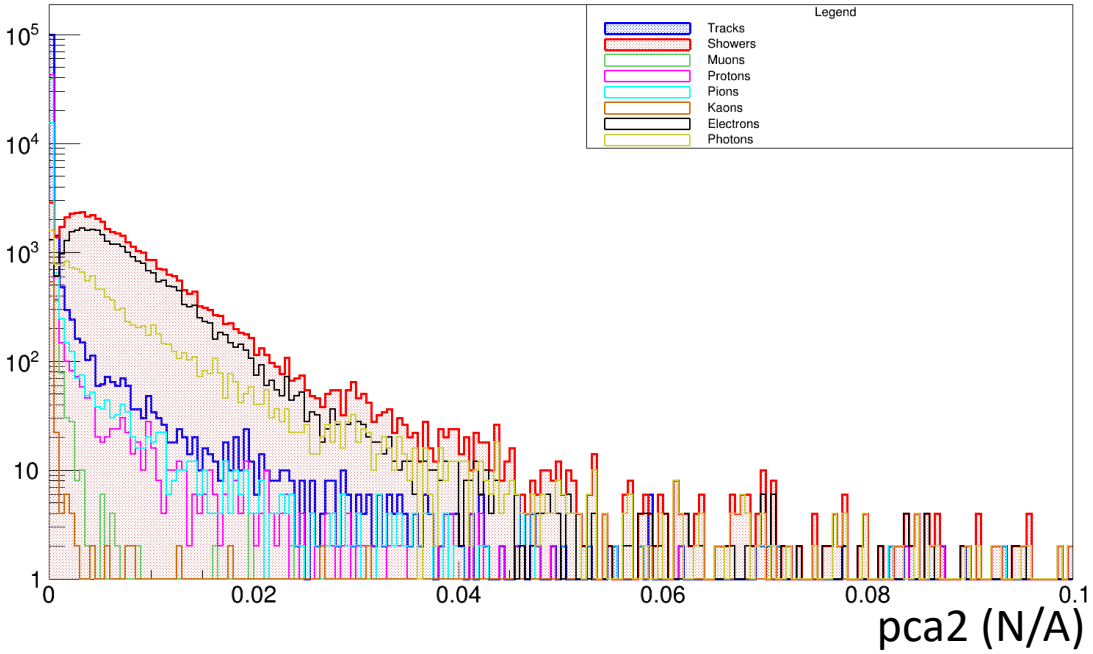
vertexDistance – Distance between the PFO vertex and the primary vertex



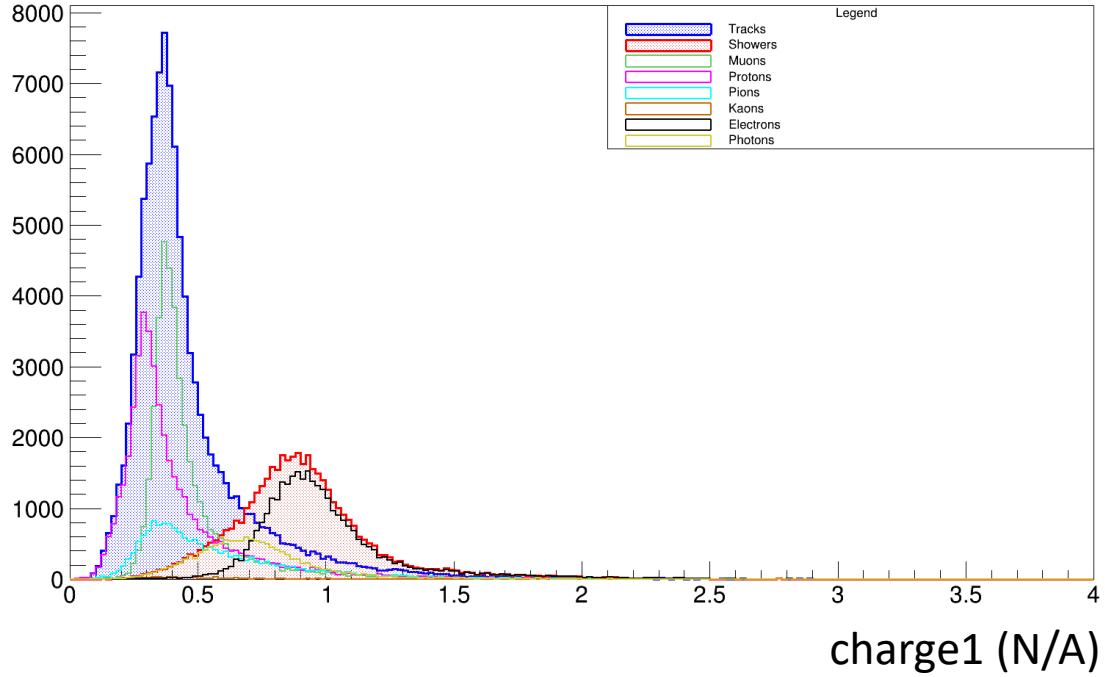
diffAngle – Difference between the opening and closing angles calculated over 50% of the pfo closest and furthest from the vertex.



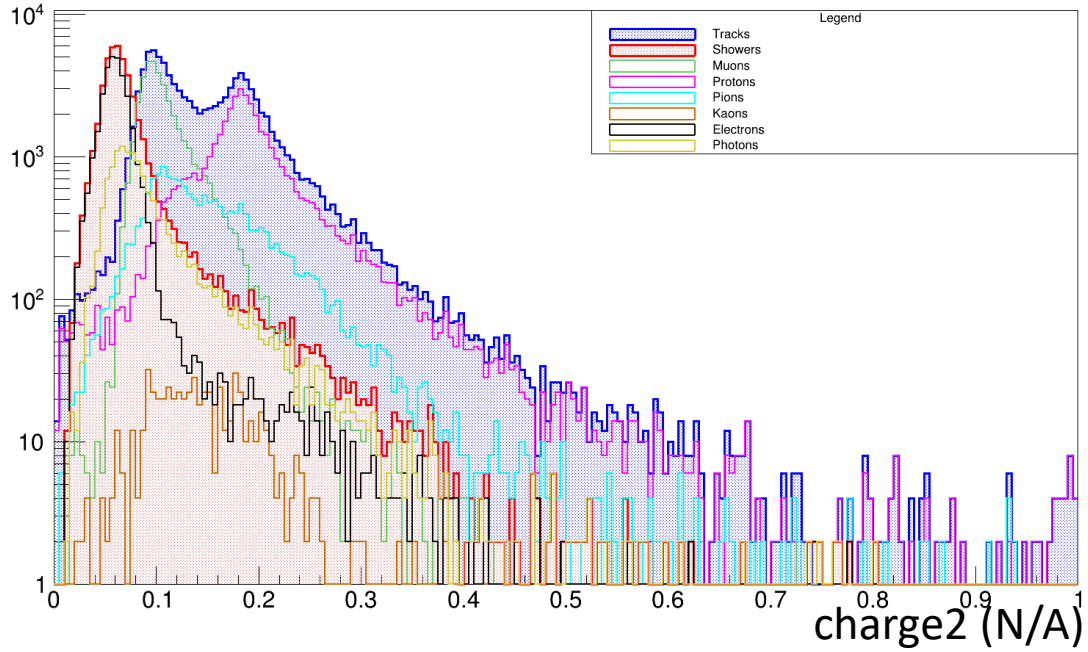
pca1 – Ratio between the second largest and the largest PCA eigenvalue



pca2 – Ratio between the third largest and the largest PCA eigenvalue



charge1 – Ratio between sigmaCharge $((charge - meanCharge)^2)$ and the mean charge in collection plane.



charge2 – Ratio of charge in the last 10% of the PFO and the mean charge in the collection plane

Current Approach In Pandora

Cut flow approach to track/shower characterisation

Cut-flow approach

- Does some basic cuts based on topological variables and then decide if a PFO is track-like or shower-like
- For anyone interested (PfoCharacterisationBaseAlgorithm.cc)

BDT1, BDT2, and BDT3

BDT setup

BDT1

- Only uBooNE variables
- Sensible fiducial cuts (20cm in x and y direction, and 200 cm in z direction)
- Completeness >= 0.1 && Purity >= 0.5
- 50% numu and 50% nue MCC11 samples
- Veto mischaracterised pfo *(will talk about it in the next slide)*
- After cuts -> 137,000 signals (track-like) and 52,000 backgrounds (shower-like)
- kBDT, Ntrees = 800, MinNodeSize=5%, MaxDepth=3, BoostType=AdaBoost, AdaBoostBeta=0.5, BaggedSampleFraction=0.6, SeparationType=GiniIndex, nCuts=20

BDT2

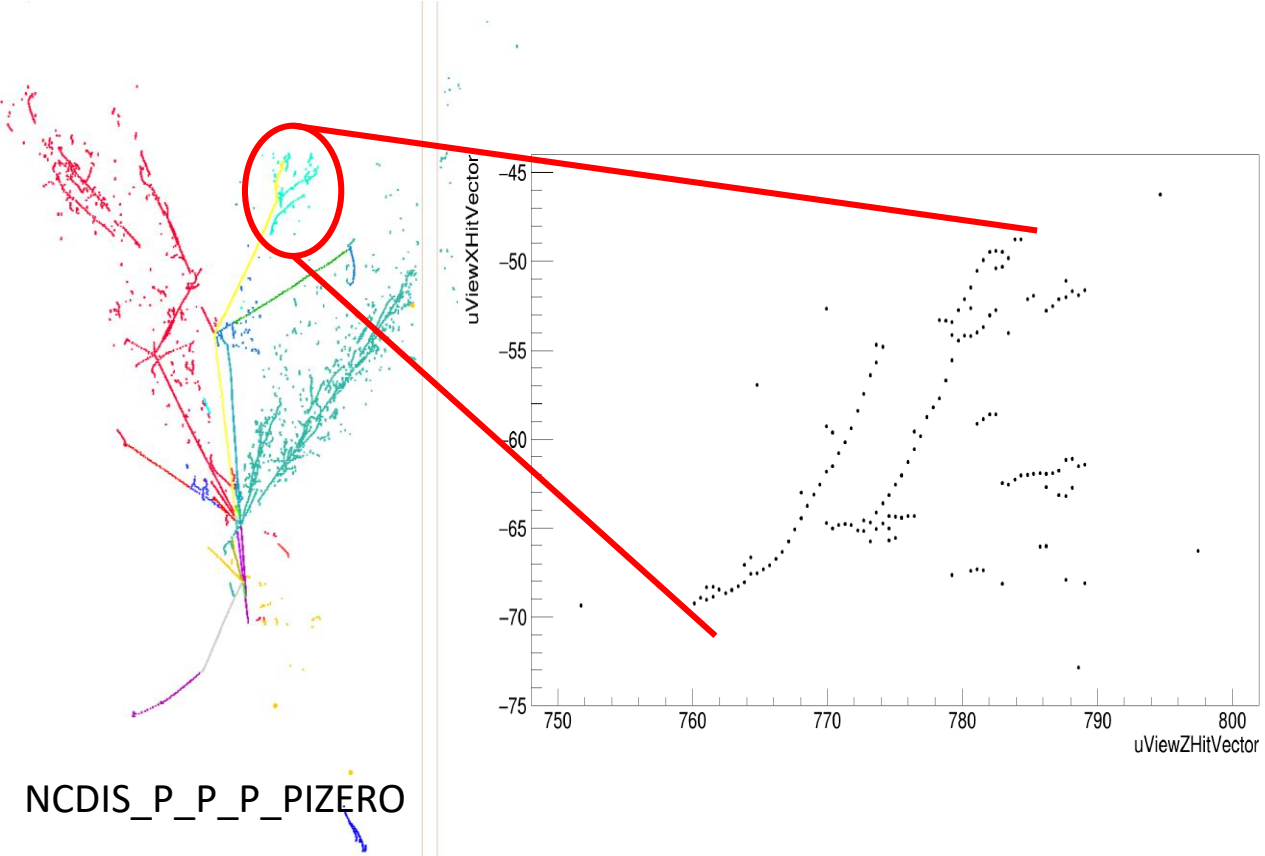
- Same as BDT1
- Completeness >= **0.8** && Purity >= **0.8**
- After cuts -> **52,000** signals (track-like) and **23,000** backgrounds (shower-like)

BDT3

- Same as BDT2
- uBooNE variables **+ hierarchy variables (will talk about it later as well)**

mischaracterisedPfo

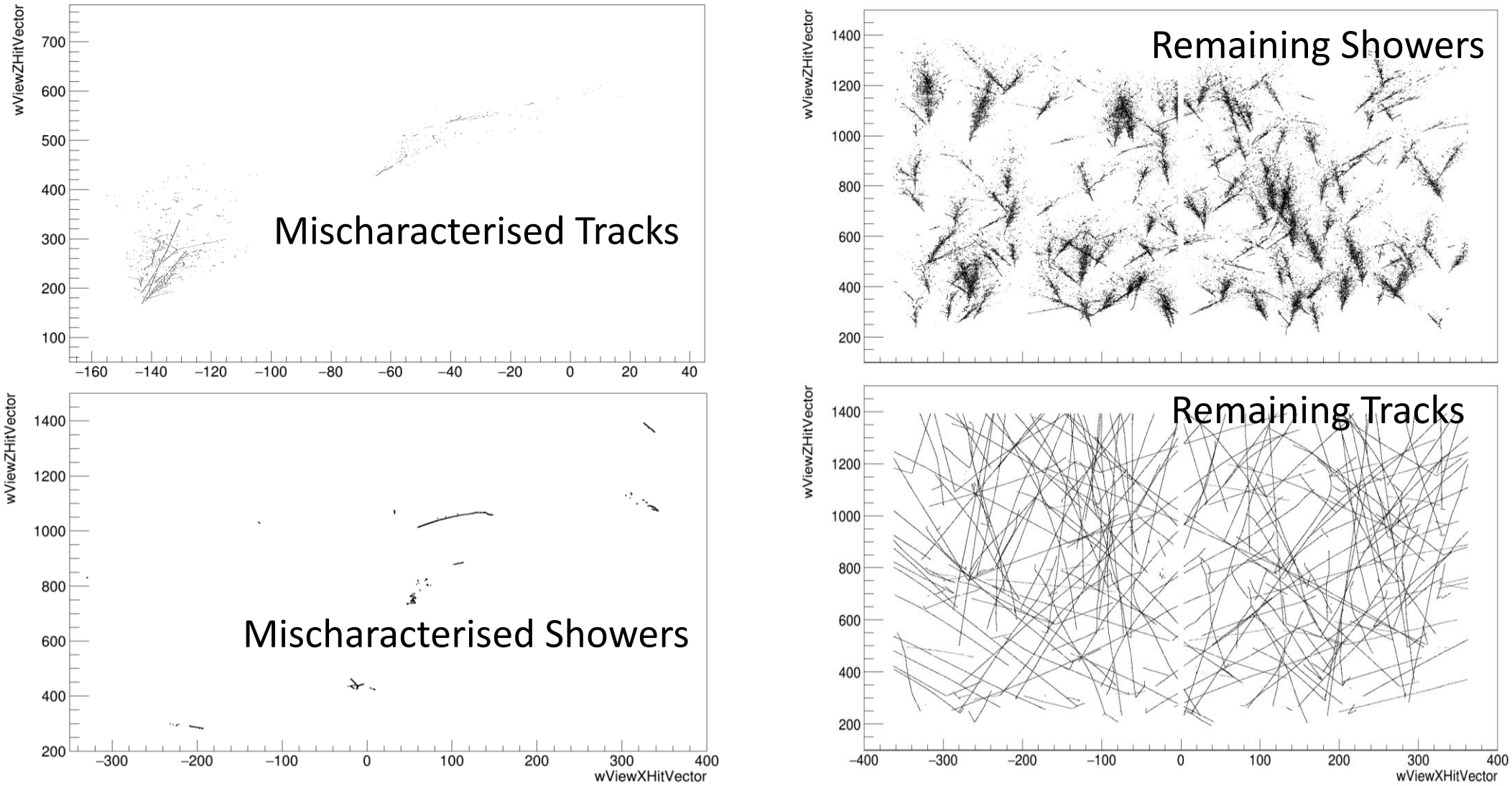
Motivation



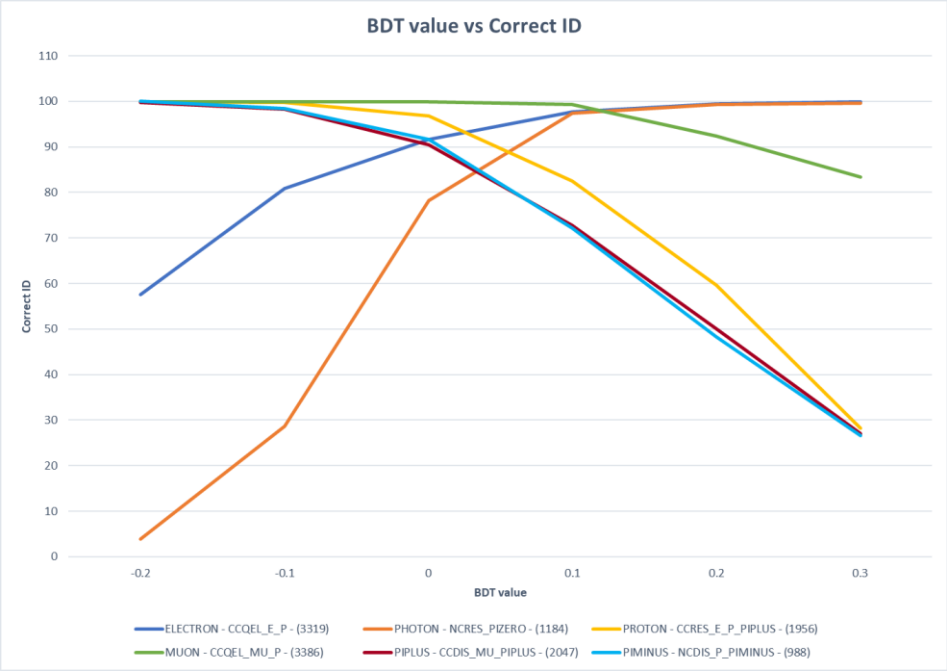
Description

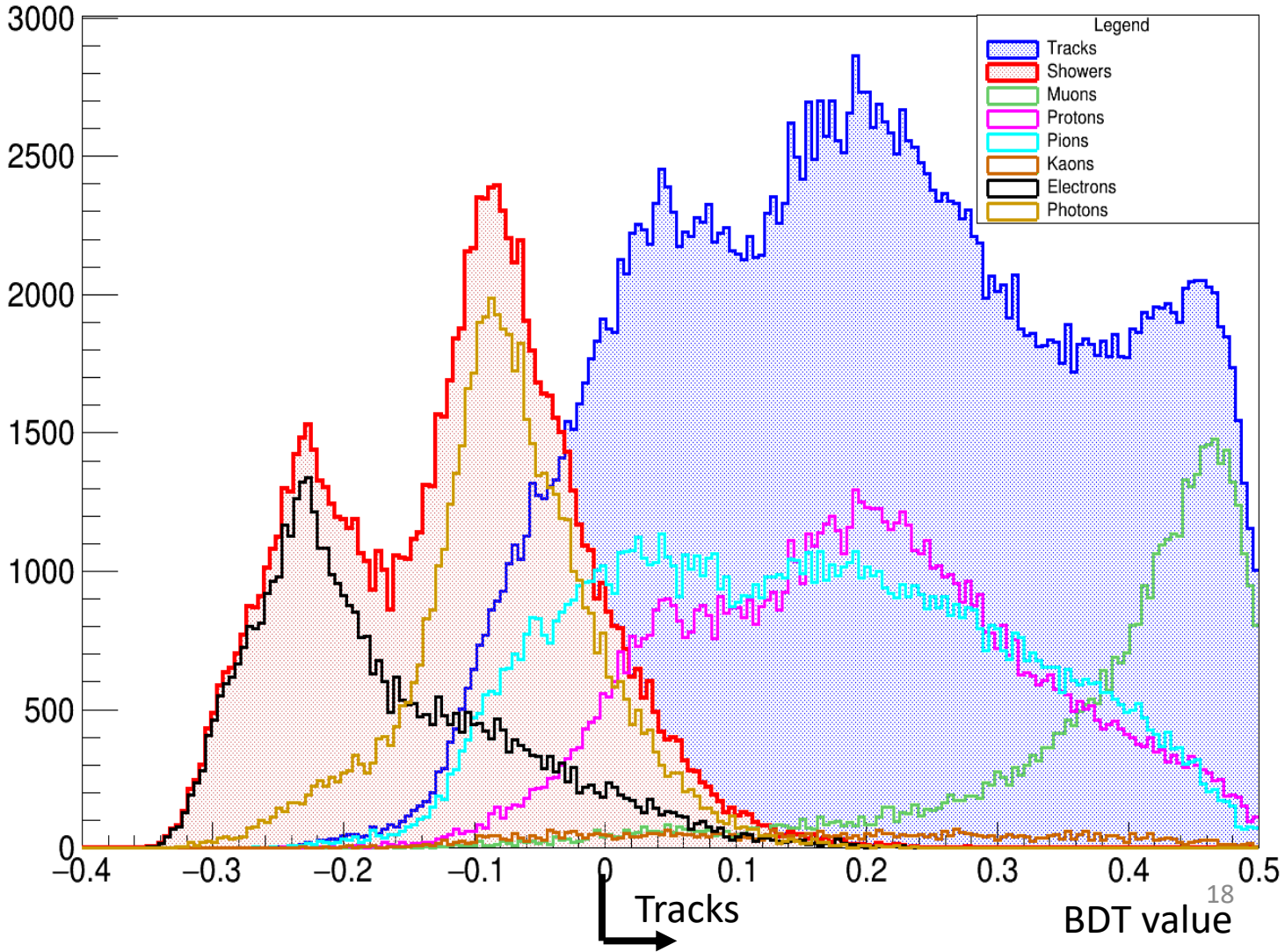
- For a given PFO, work out what MCParticle each hit maps to
- Workout showerProbability where,
  - $showerProbability = \frac{total\ shower-like\ hits\ in\ all\ view}{total\ hits\ in\ all\ view}$
- Flag a PFO as mischaracterised if showerProbability >= 0.5 but called as track-like or vice versa
- Use this flag to improve BDT training
- Essentially, don’t train on wrong topology

Examples of vetoed and remaining PFOs

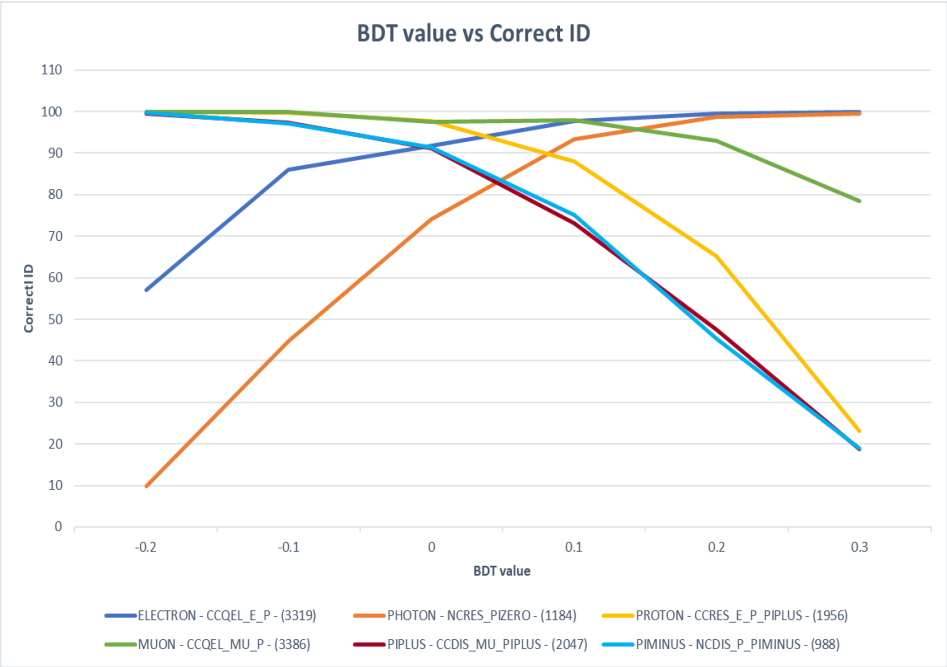


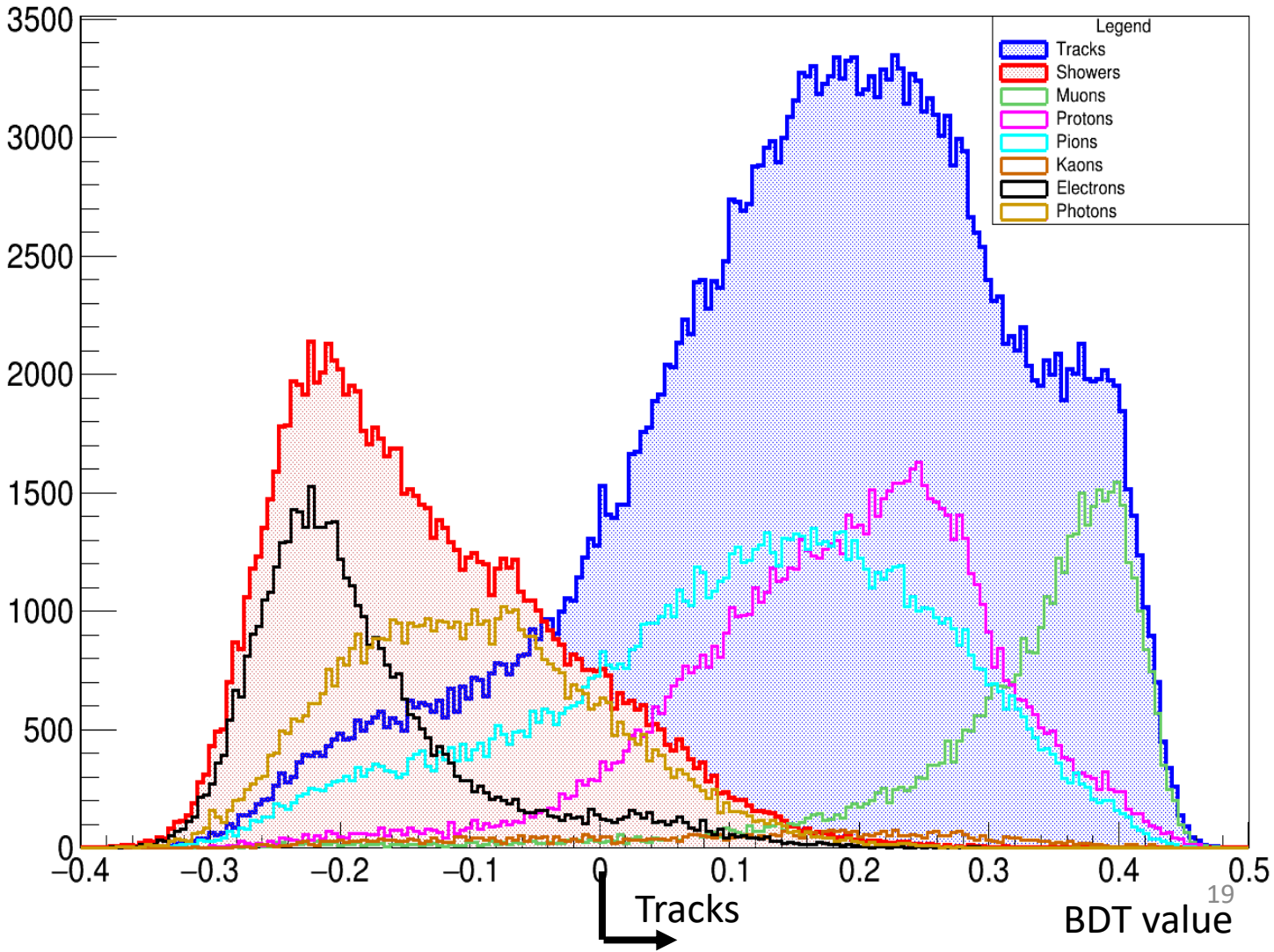
BDT value distribution for BDT1





BDT value distribution for BDT2



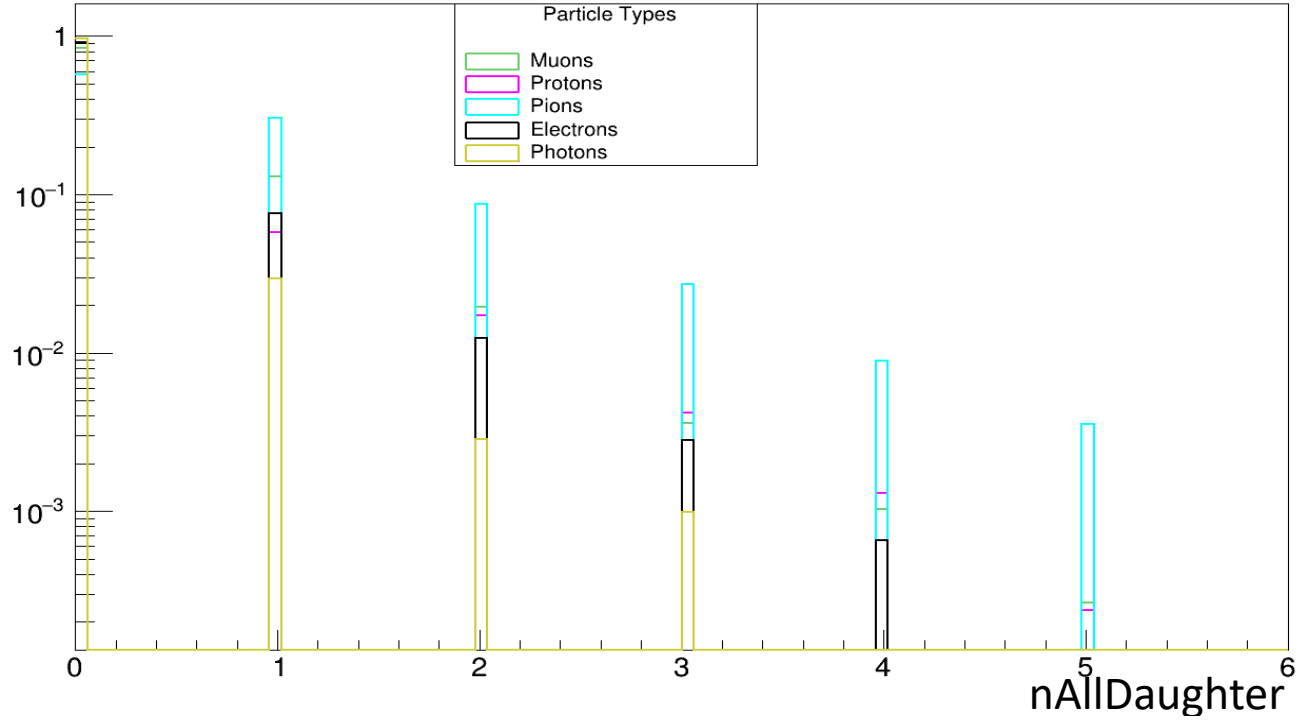


Hierarchy Variables

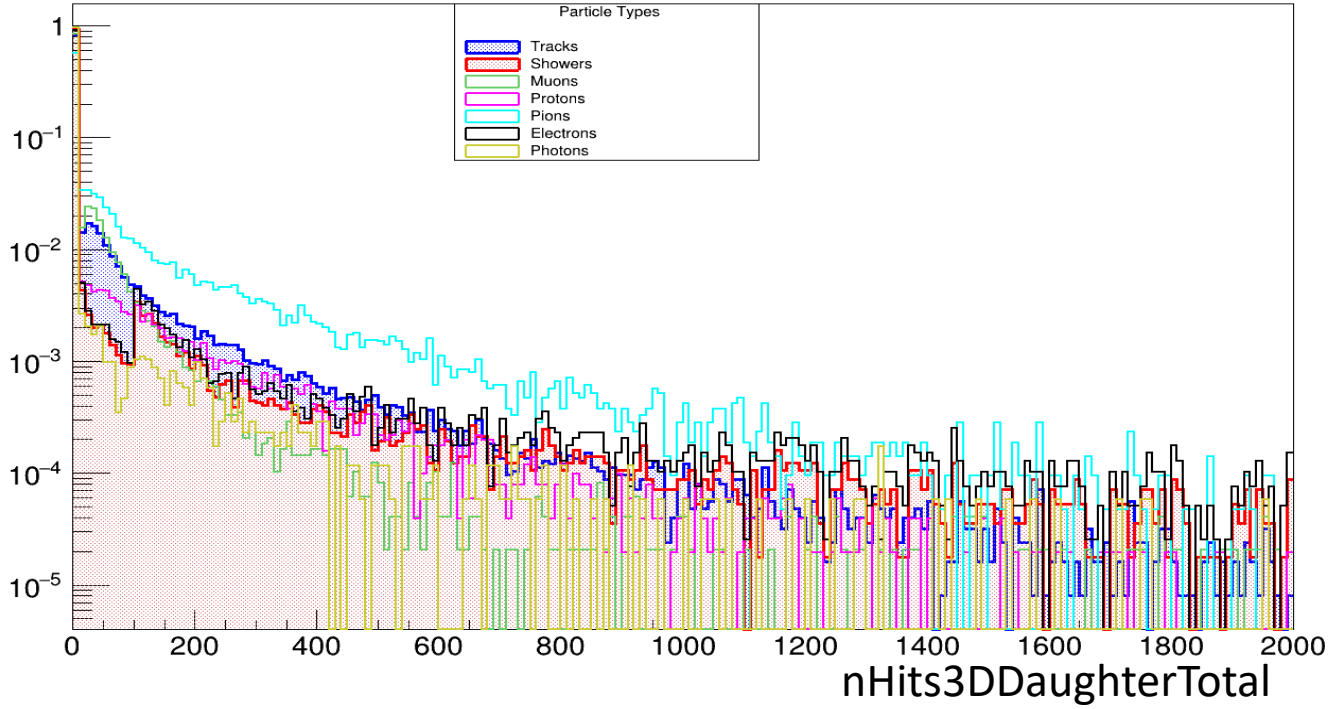
What are they?

Hierarchy variables

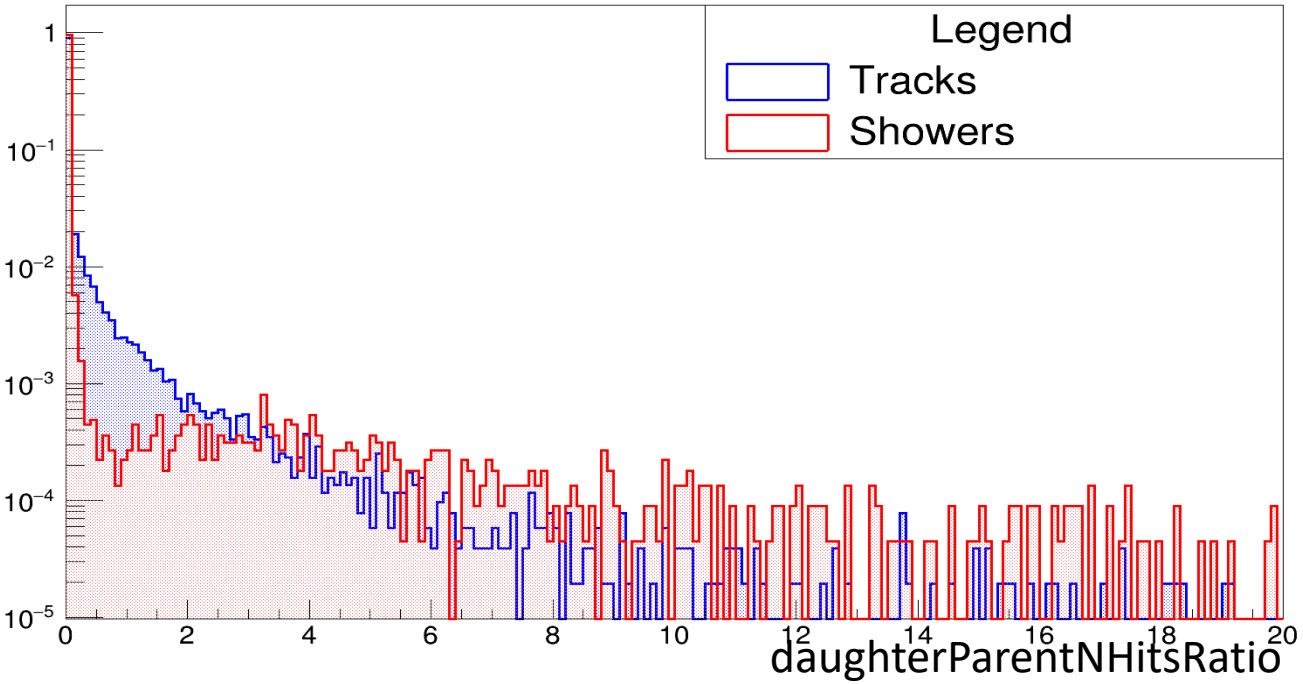
- Indicates hadronic activities
- 3 new variables
  - nAllDaughter
  - nHits3DDaughterTotal
  - daughterParentNhitsRatio
- Aim is to push protons and pions towards the track-like region to improve track/shower separation

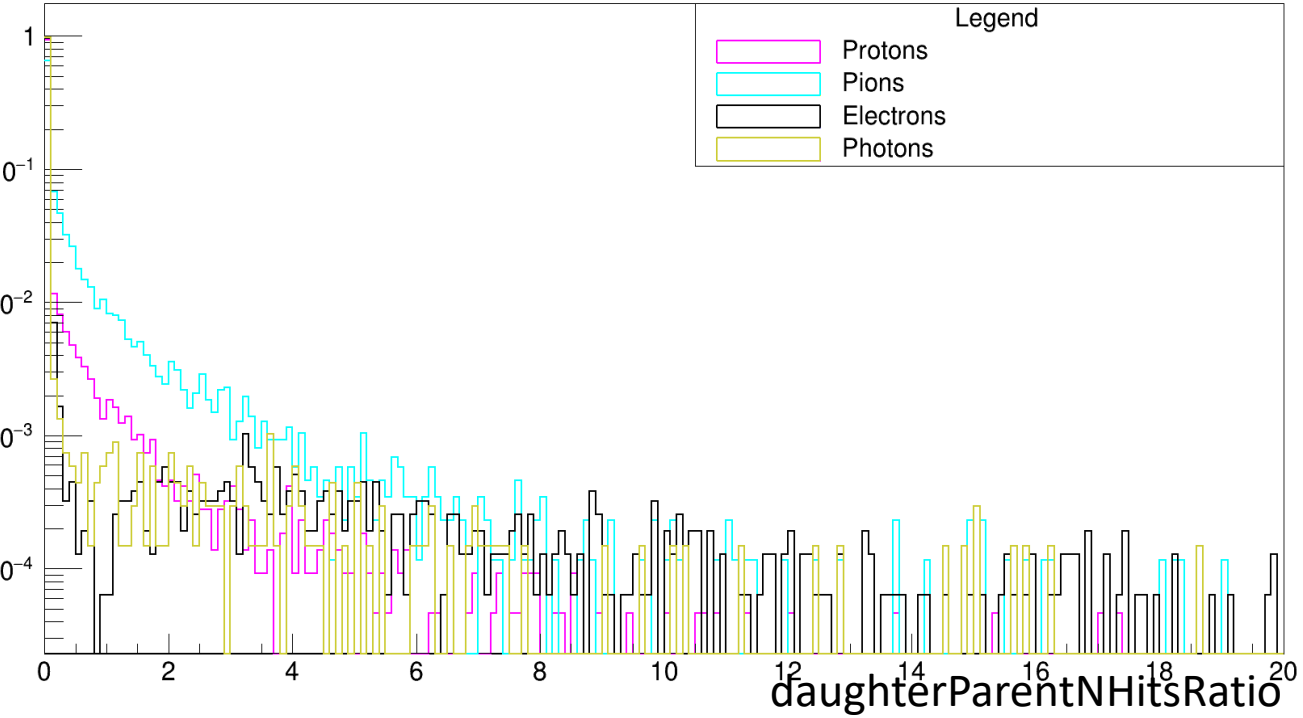


nAllDaughter – total number of all downstream daughter pfos



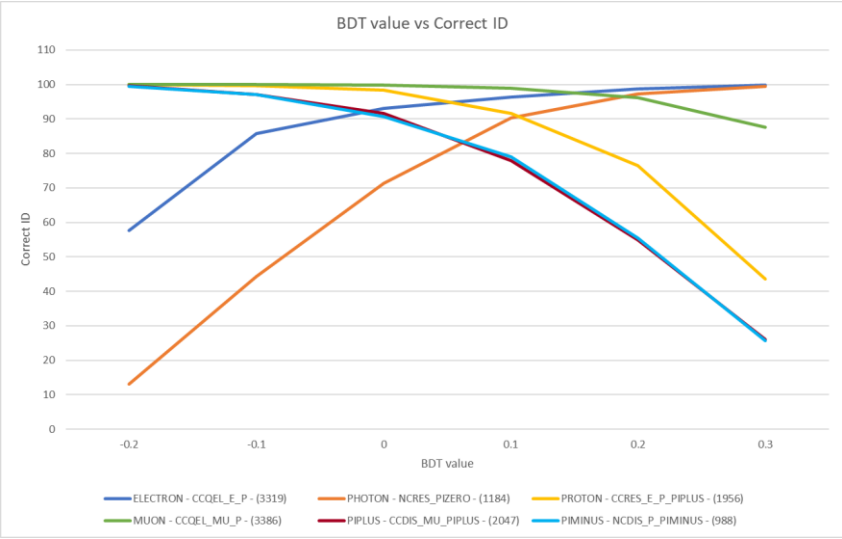
nHits3DDaughterTotal – total number of 3D hits in all downstream daughter pfos

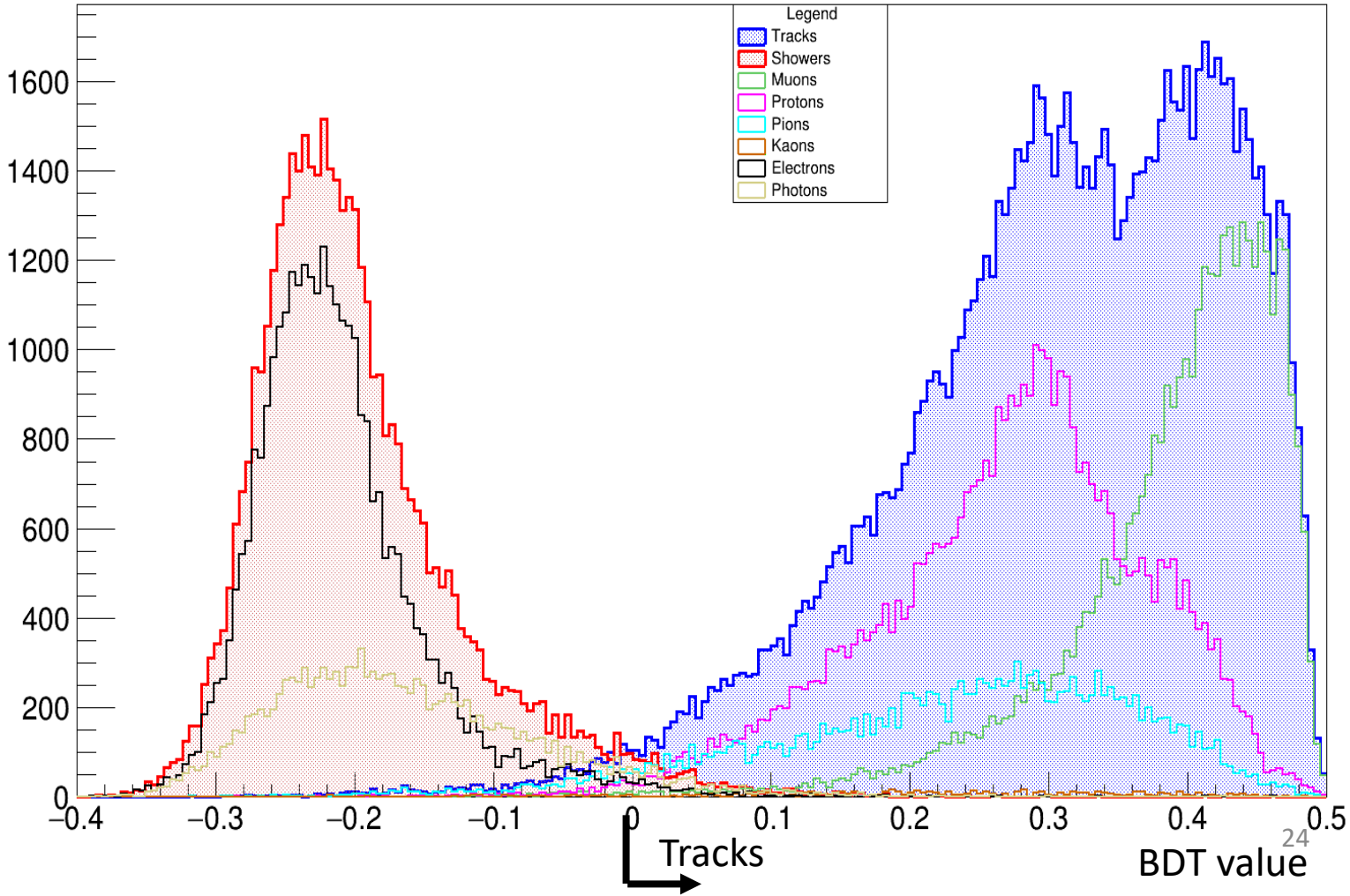




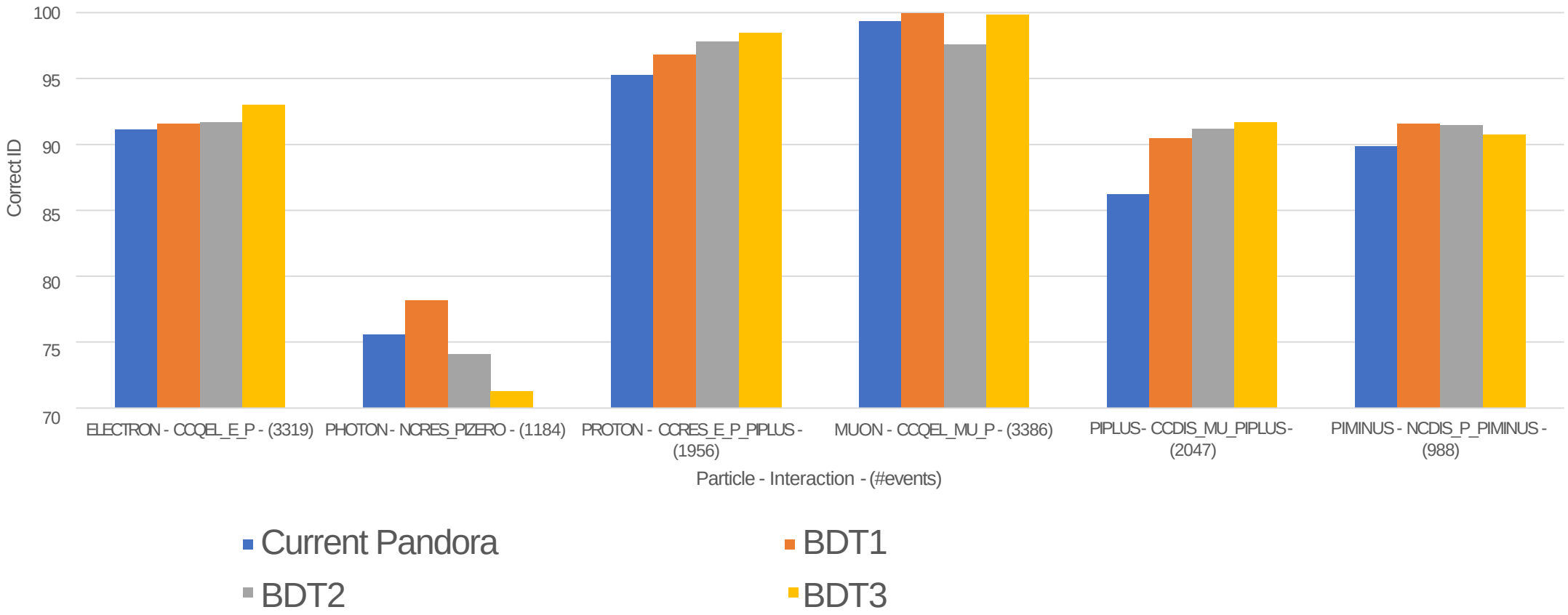
daughterParentNhitsRatio – 3D hits ratio between all downstream daughter pfos and parent pfo.

BDT value distribution for BDT3





Comparing Different BDTs



Validation Output - Electron

Current Pandora	BDT3
CCQEL_E -ELECTRON: nEvents: 3419, correctId 92.478%	CCQEL_E -ELECTRON: nEvents: 3419, correctId 92.205%
CCQEL_E_P -ELECTRON: nEvents: 3319, correctId 91.1111%	CCQEL_E_P -ELECTRON: nEvents: 3319, correctId 92.9938%
CCRES_E_P_PIPLUS -ELECTRON: nEvents: 1956, correctId 89.6534%	CCRES_E_P_PIPLUS -ELECTRON: nEvents: 1956, correctId 91.0189%
CCRES_E_P_PIZERO -ELECTRON: nEvents: 822, correctId 92.2094%	CCRES_E_P_PIZERO -ELECTRON: nEvents: 822, correctId 92.848%

Validation Output - Photon

Current Pandora	BDT3
CCRES_E_P_PIZERO -PHOTON1: nEvents: 822, correctId 77.4576Z	CCRES_E_P_PIZERO -PHOTON1: nEvents: 822, correctId 80.5085%
CCDIS_MU_P_PIZERO -PHOTON1: nEvents: 981, correctId 87.6457%	CCDIS_MU_P_PIZERO -PHOTON1: nEvents: 981, correctId 90.9091%
NCRES_PIZERO -PHOTON1: nEvents: 1184, correctId 75.5474%	NCRES_PIZERO -PHOTON1: nEvents: 1184, correctId 71.2591%
NCRES_P_PIZERO -PHOTON1: nEvents: 834, correctId 79.3829%	NCRES_P_PIZERO -PHOTON1: nEvents: 834, correctId 81.9074%

Validation Output - Muon

Current Pandora	BDT3
CCQEL_MU -MUON: nEvents: 3161, correctId 99.6145%	CCQEL_MU -MUON: nEvents: 3161, correctId 99.8394%
CCRES_MU_P -MUON: nEvents: 3386, correctId 99.3714%	CCQEL_MU_P -MUON: nEvents: 3386, correctId 99.8503%
CCRES_MU_P_PIPLUS -MUON: nEvents: 1875, correctId 98.75%	CCRES_MU_P_PIPLUS -MUON: nEvents: 1875, correctId 99.1848%
CCDIS_MU_PIPLUS -MUON: nEvents: 2047, correctId 97.6744%	CCDIS_MU_PIPLUS -MUON: nEvents: 2047, correctId 98.8878%

Validation Output - Proton

Current Pandora	BDT3
CCQEL_MU_P -PROTON1: nEvents: 3386, correctId 94.0557%	CCQEL_MU_P -PROTON1: nEvents: 3386, correctId 98.7307%
CCQEL_E_P -PROTON1: nEvents: 3319, correctId 94.2895%	CCQEL_E_P -PROTON1: nEvents: 3319, correctId 98.4903%
CCRES_E_P_PIPLUS -PROTON1: nEvents: 1956, correctId 95.2663%	CCRES_E_P_PIPLUS -PROTON1: nEvents: 1956, correctId 98.4615%
CCRES_MU_P_PIPLUS -PROTON1: nEvents: 1875, correctId 95.3071%	CCRES_MU_P_PIPLUS -PROTON1: nEvents: 1875, correctId 98.5516%

Validation Output - PiPlus

Current Pandora	BDT3
CCRES_MU_PIPLUS -PIPLUS: nEvents: 1867, correctId 83.4203%	CCRES_MU_PIPLUS -PIPLUS: nEvents: 1867, correctId 92.6377%
CCRES_E_P_PIPLUS -PIPLUS: nEvents: 1956, correctId 78.7336%	CCRES_E_P_PIPLUS -PIPLUS: nEvents: 1956, correctId 91.3979%
CCDIS_MU_PIPLUS -PIPLUS: nEvents: 2047, correctId 86.2016%	CCDIS_MU_PIPLUS -PIPLUS: nEvents: 2047, correctId 91.6796%
CCDIS_MU_P_PIPLUS -PIPLUS: nEvents: 825, correctId 85.3247%	CCDIS_MU_P_PIPLUS -PIPLUS: nEvents: 825, correctId 90.6494%

Validation Output - PiMinus

Current Pandora	BDT3
NCDIS_P_PIMINUS -PIMINUS: nEvents: 988, correctId 89.8361%	NCDIS_P_PIMINUS -PIMINUS: nEvents: 988, correctId 90.7104%
NCRES_P_PIMINUS -PIMINUS: nEvents: 574, correctId 88.417%	NCRES_P_PIMINUS -PIMINUS: nEvents: 574, correctId 95.1737%

What’s next?

- Replicate these results inside Pandora using SKLearn
- Do more tests
- Release
- Deep Learning approaches

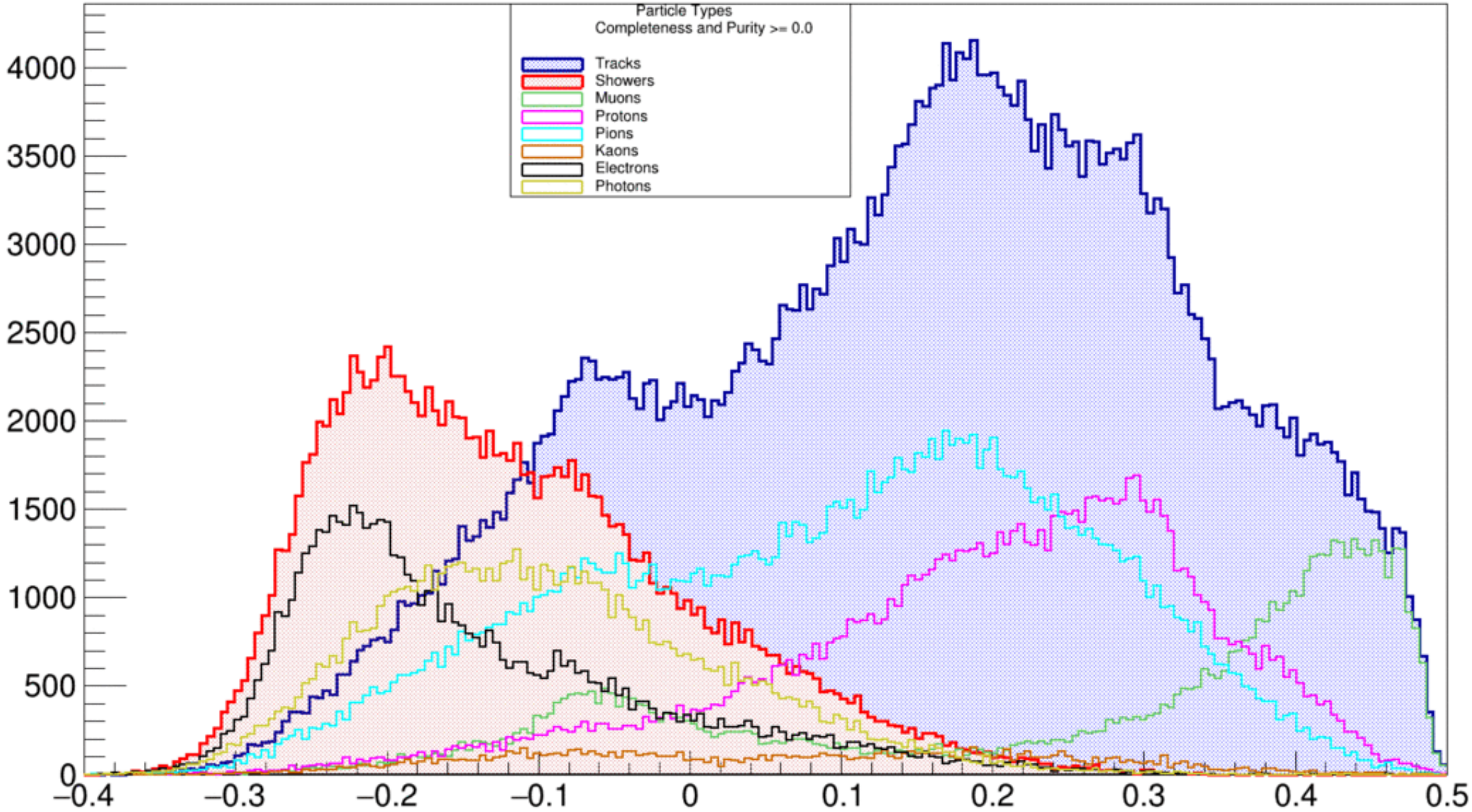
The End

Thank you. Any questions?

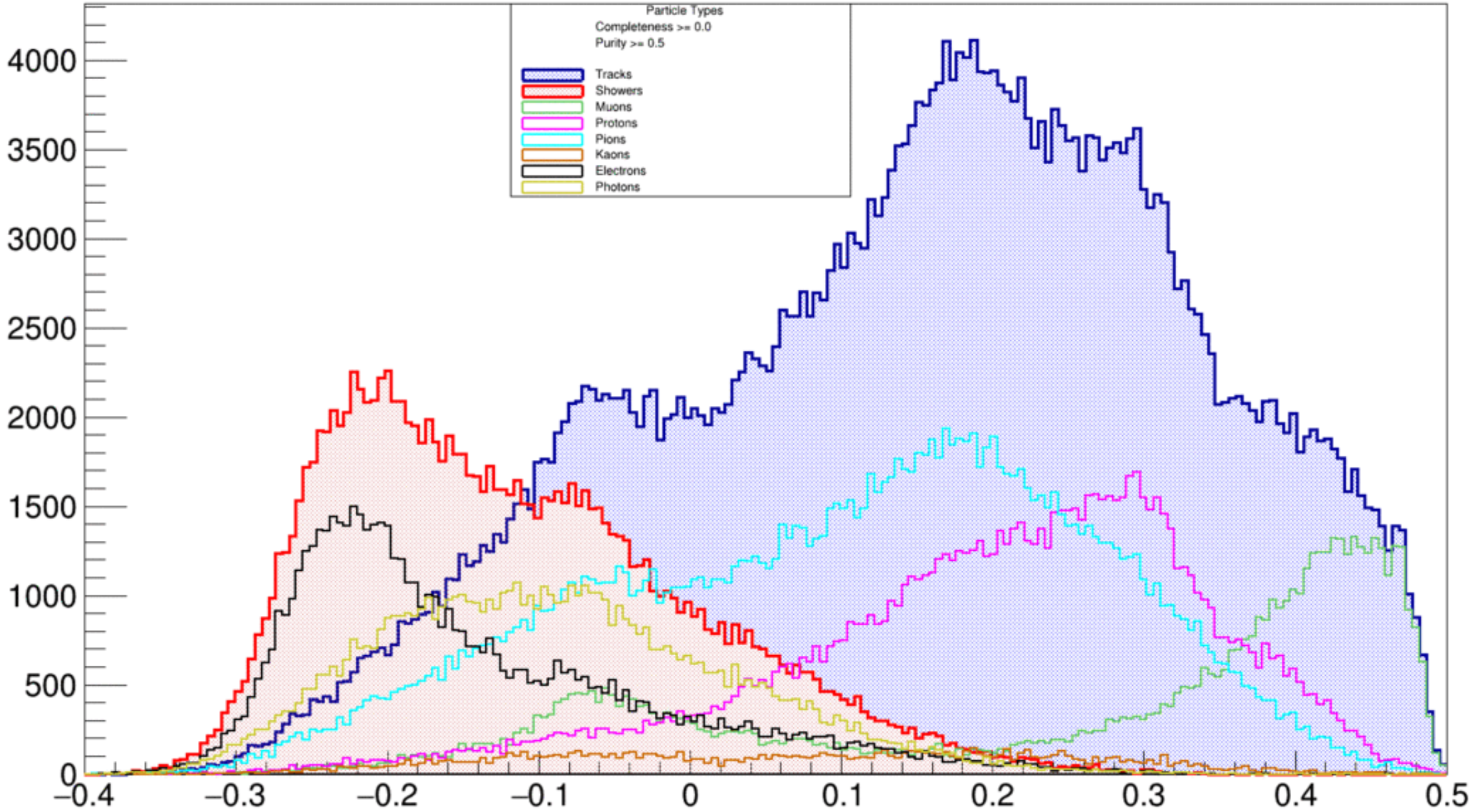
Backup Slides

Small Note On Completeness VS Purity

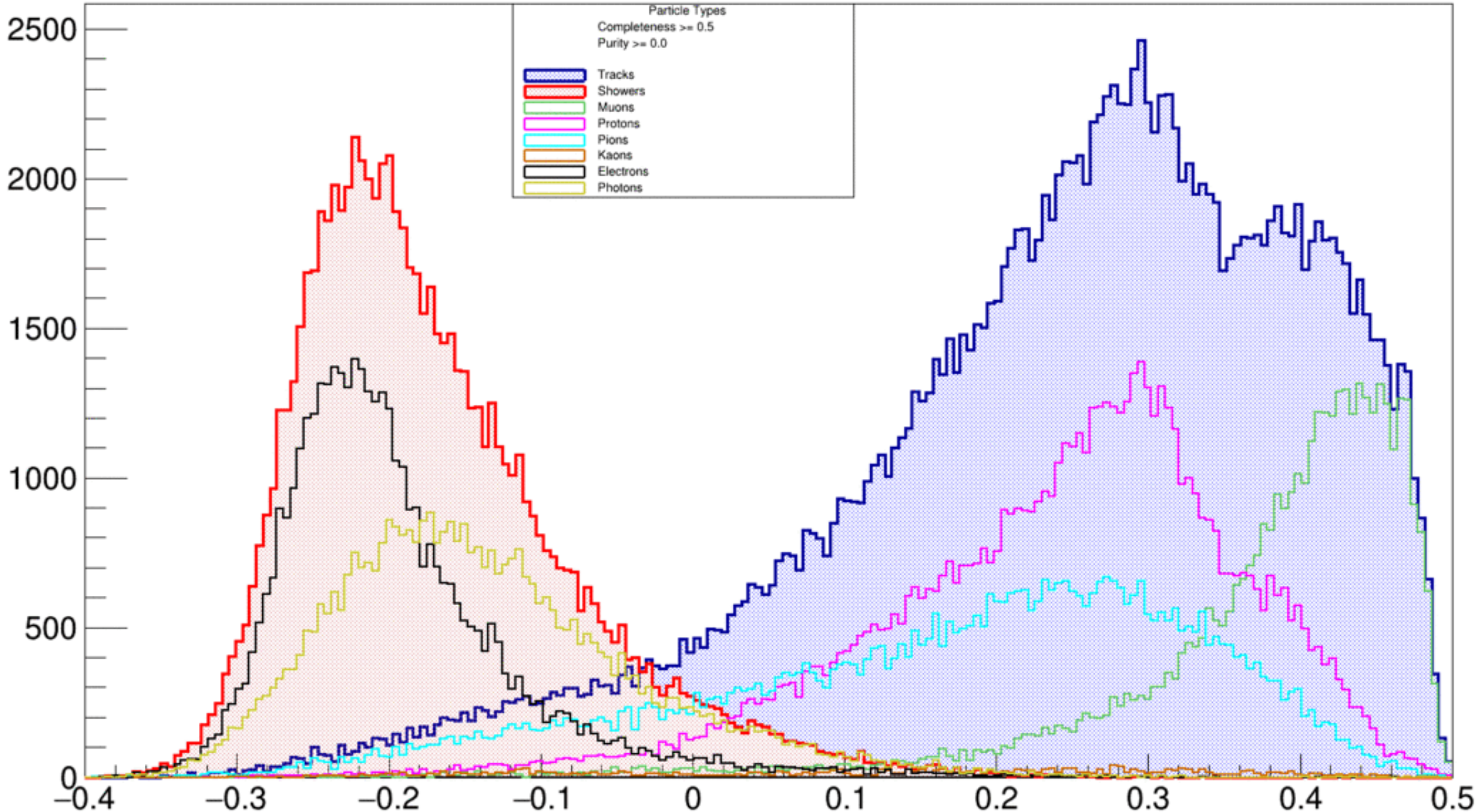
The GIF below cycles from Completeness && Purity >= 0.0 to 1.0.



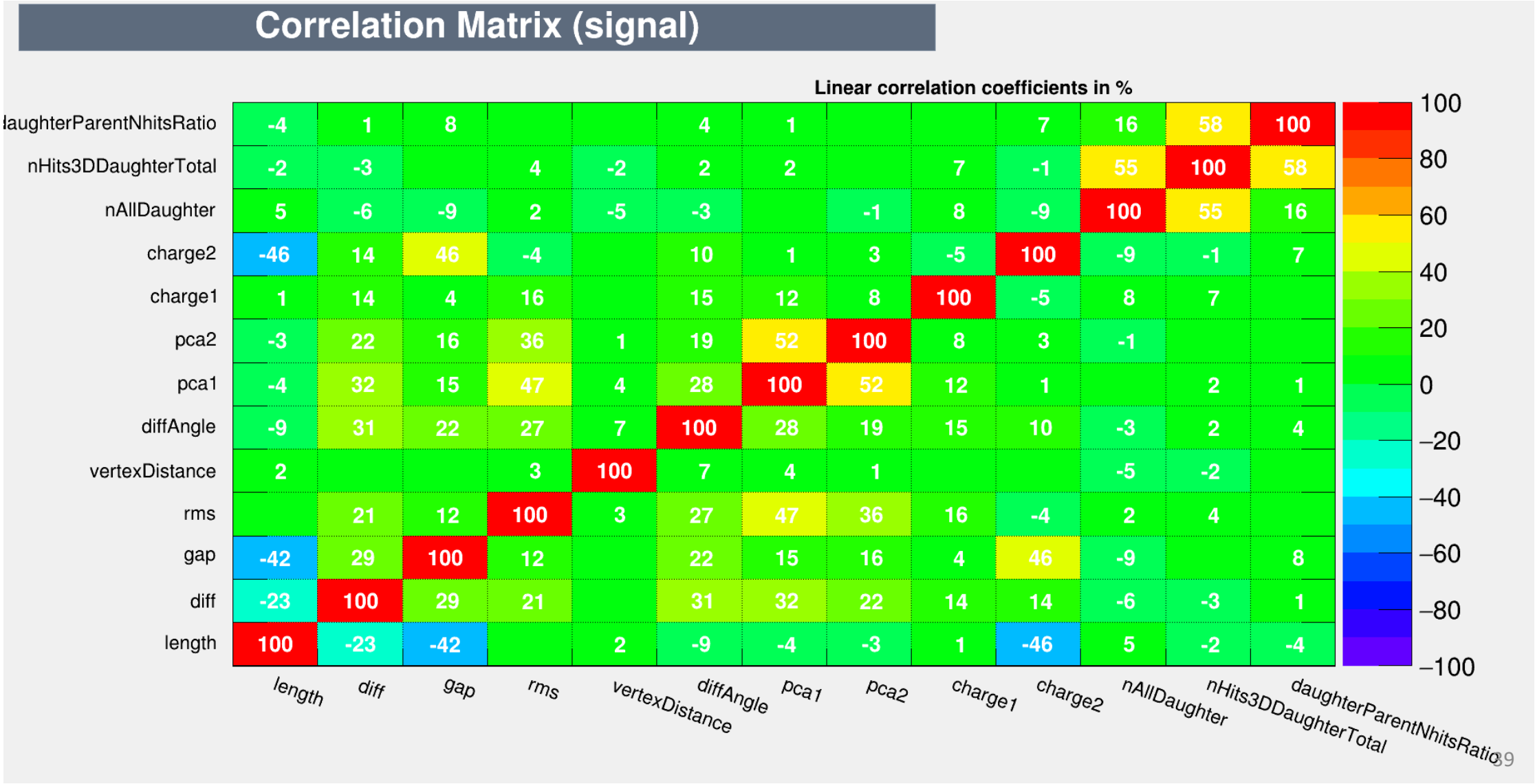
The GIF below cycles from Purity >= 0.5 && Completeness >= 0.0 to 1.0.



The GIF below cycles from Completeness >= 0.5 && Purity >= 0.0 to 1.0.

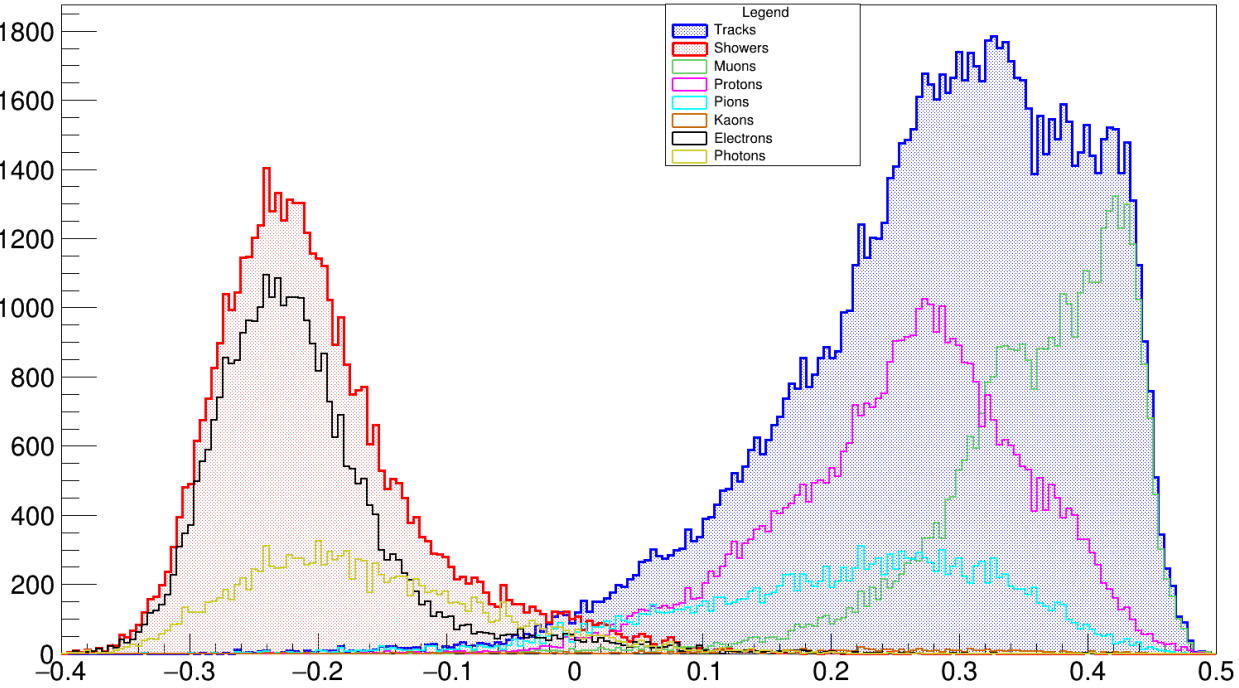


Correlation Matrix for 13 variables in BDT3

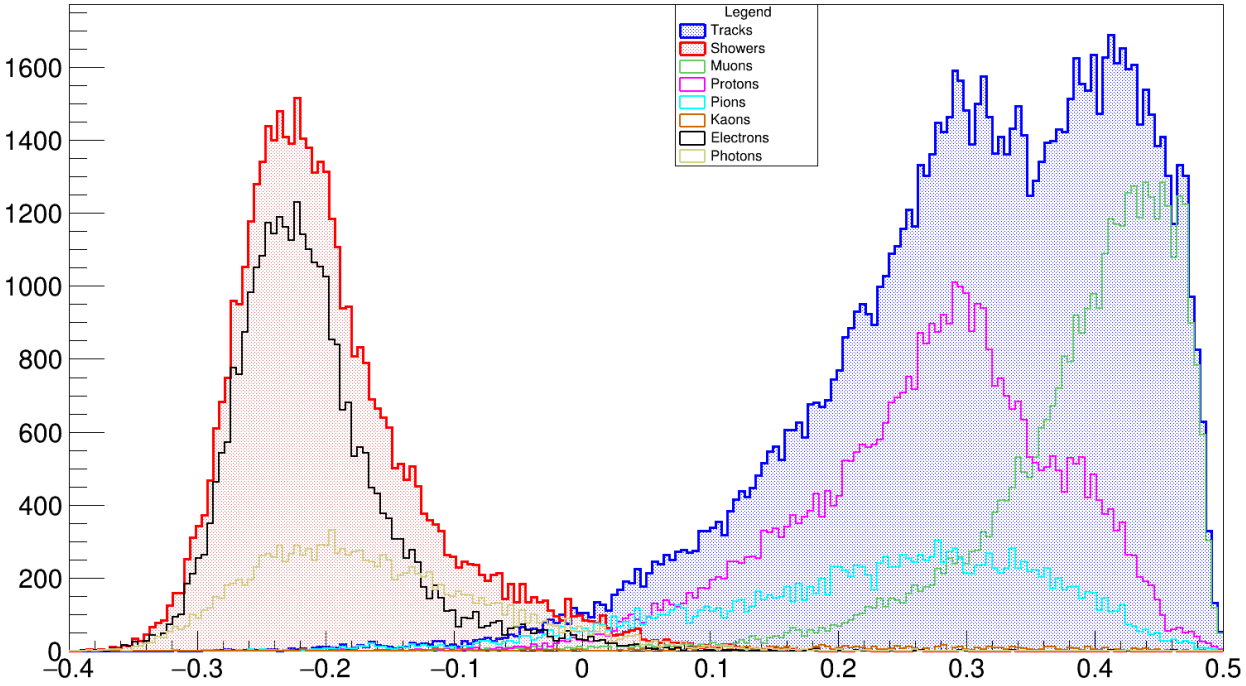


How removing nHits3DDaughterTotal affects Track/Shower ID

BDT3 w/o nHit3DDaughterTotal



BDT3



nHits3DDaughterTotal for Track-like particles