CEPC JOI

Kaili Zhang

IHEP

zhangkl@ihep.ac.cn

Machine Learnings on Jets

- P-CNN
  - https://scipost.org/10.21468/SciPostPhys.7.1.014
- Particle Flow Network
  - https://arxiv.org/abs/1810.05165
  - CEPC@Xiaotian : https://arxiv.org/abs/2410.04465v2
- LundNet
  - https://doi.org/10.1007/jhep03(2021)052
- ParticleNet
  - Arxiv:1902.08570
  - https://github.com/hqucms/ParticleNet

ParticleTransformer

- https://arxiv.org/abs/2202.03772
- https://github.com/jet-universe/particle_transformer
- Platforms: https://github.com/hqucms/weaver-core
- Application on CEPC: 2309.13231 , PRL 132, 221802 (2024)
- Tutorial on CEPC: https://github.com/ZHUYFgit/CEPC-Jet-Origin-Identification
- Inputs from CEPCsoft: /cefs/higgs/zhangkl/AI/datasets
- Inputs from LHC, JetClass: /cefs/higgs/zhangkl/AI/jetclass
- Require higgsgpu group. Request on https://ccsinfo.ihep.ac.cn/
- Follow the tutorial, build the env if you are interested.

ParticleTransformer @ CEPC

https://github.com/ZHUYFgit/CEPC-Jet-Origin-Identification

- Variable list in M11origin.cc
  - Under development to CEPCSW
  - Unit as one jet: 4 momentum, M11 id information……
- Train in Weaver: JetClass_full.yaml
- Submit jobs on IHEP: train_JetClass.sh
- Output: Pred.root: Label and score for each jets.
- Application: onnx format

Inputs for JOI

/cefs/higgs/zhangkl/CEPCSW/Analysis/JetOrigin/src

- Jet->Event;
- PFO->Component;
- Length: 200
- Label: M11

- Current training use truth PID information, in application reco PID will be used.

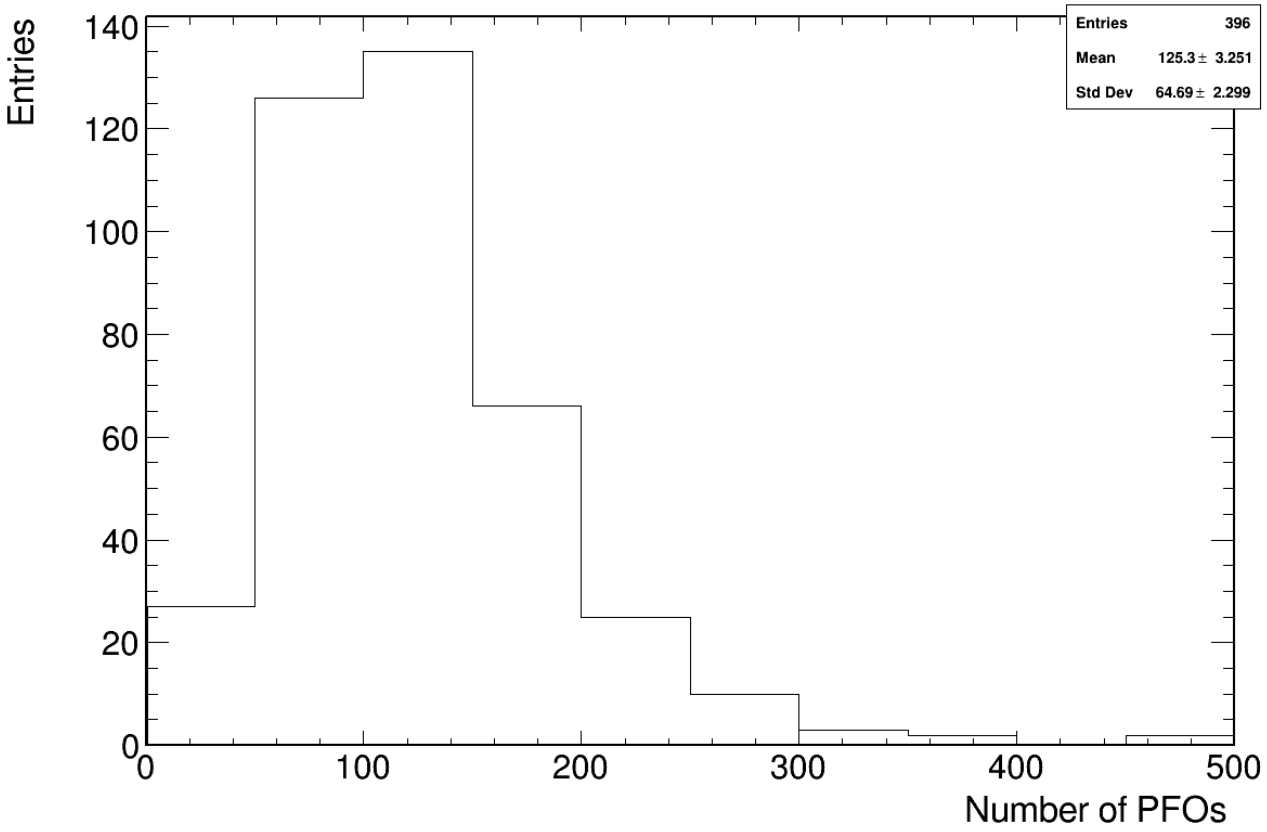
Type	Var	Comment
PFO point distance	$\Delta\phi(pfo, Jet)$	Delta Phi, pfo to jet
	$\Delta\eta(pfo, Jet)$	Delta Eta, pfo to jet
PFO Vector variable	(px, py, pz, E)	4 momentum of PFO
PFO feature variable	P_t^{PFO}, $\log\frac{P_t^{PFO}}{P_t^{jet}}$	Pfo pt and relative pt
	E_t^{PFO}, $\log\frac{E_t^{PFO}}{E_t^{jet}}$	Pfo E and relative E
	$\Delta R(pfo, Jet)$	Delta R, pfo to jet
	N_charge, N_chargeflip	Charge of PFO
	D0, Z0, D0err, Z0err	(if with track) impact parameters
	N_Ecluster, N_Hcluster	
	E_ecal, E_hcal	
	PID	Truth PID type

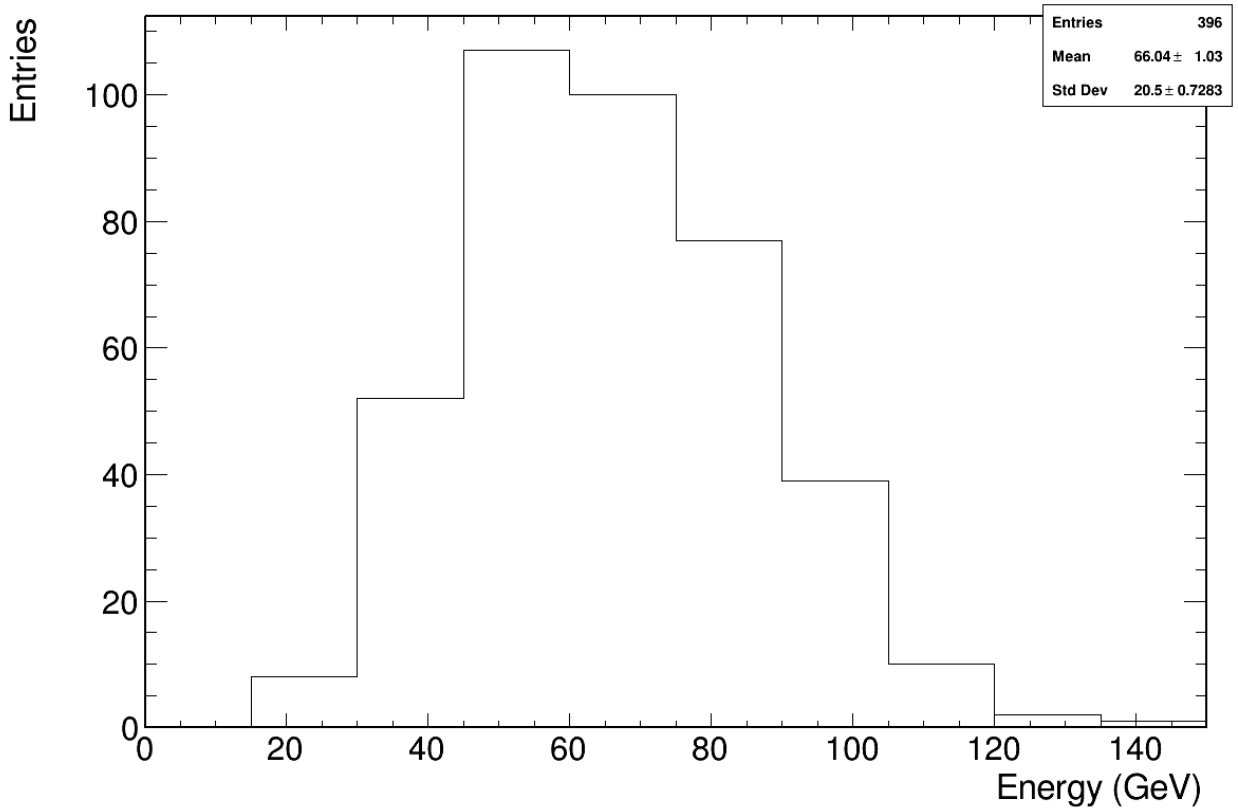
Variable convention

- Feature variable, Transformer prefer normal distribution with mean ~0, range (-1,1) with cut edge maximum(-5,5).
- (4-momentum vector variable not included)
- Normalization functions like Tanh() used.
-

Distribution for vvHbb

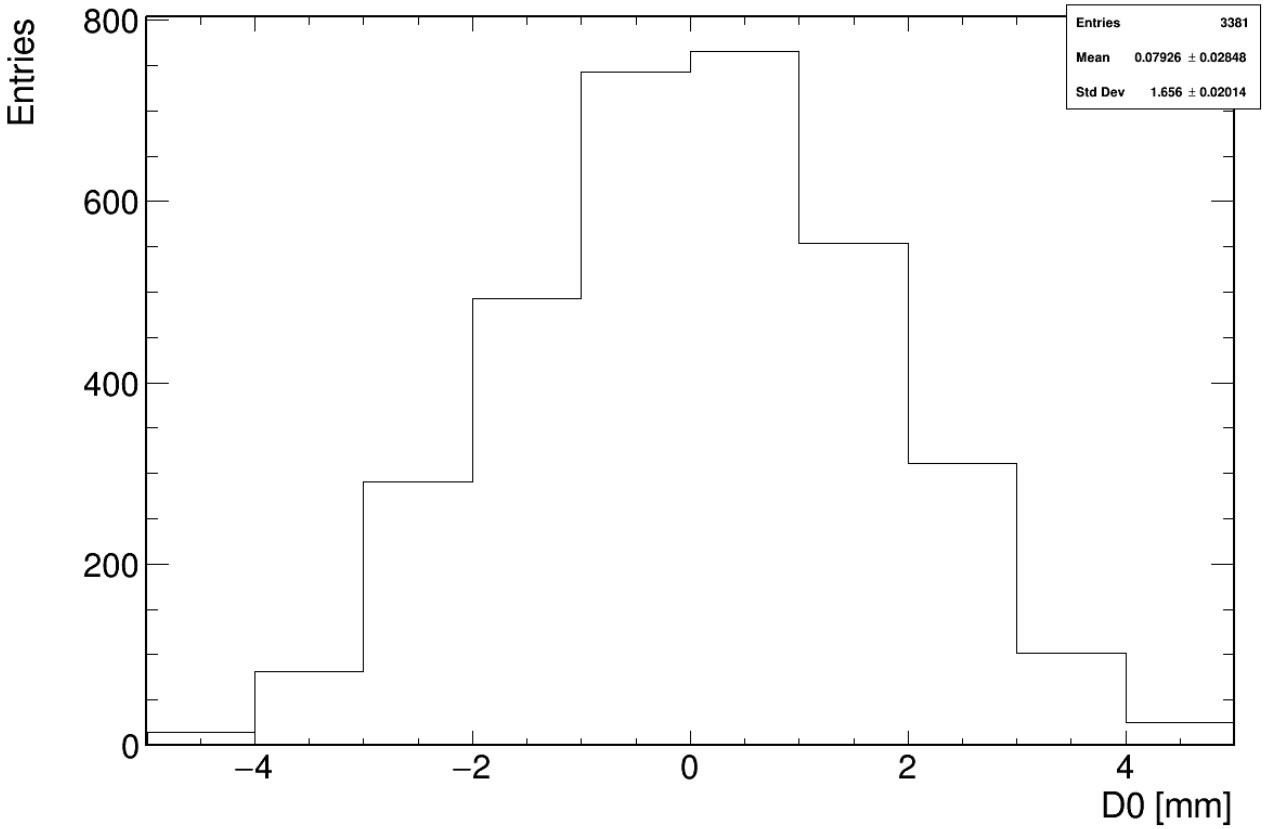
A typical bb dijet event n_PFO.
In average, one jet has ~66 GeV energy,
~60 PFOs, with ~15 PFOs, charged with tracks.

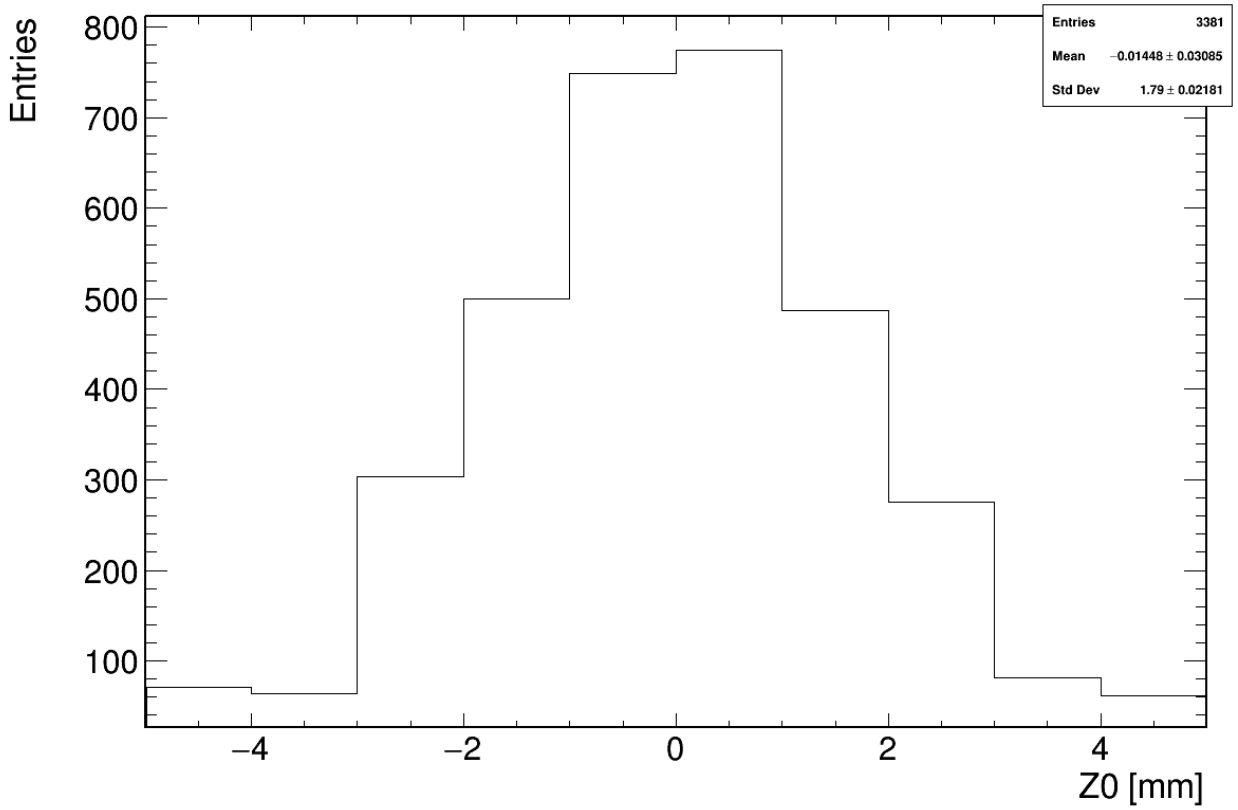




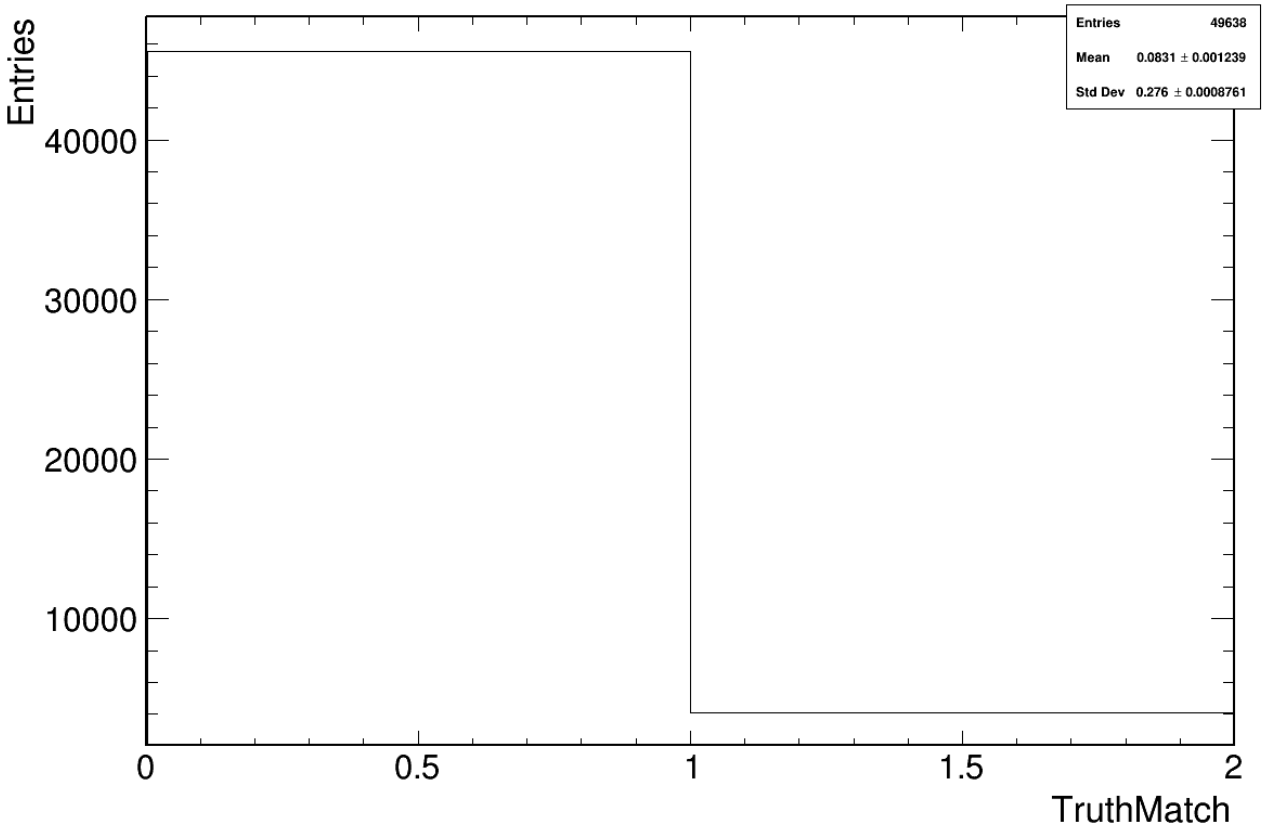
Impact parameter

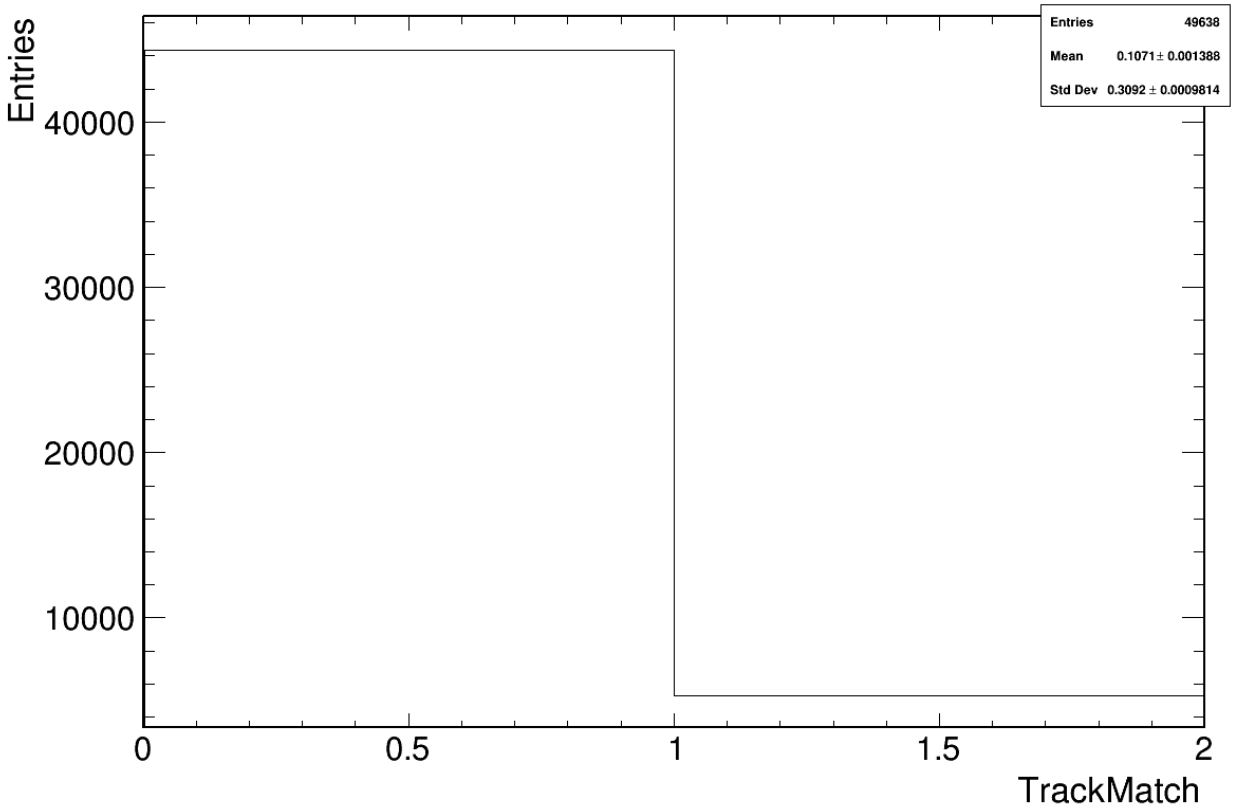
In the beginning, d0, z0 with mean value~(0um), sigma~(20um).
Use 5*tanh(D0/Z0*25).



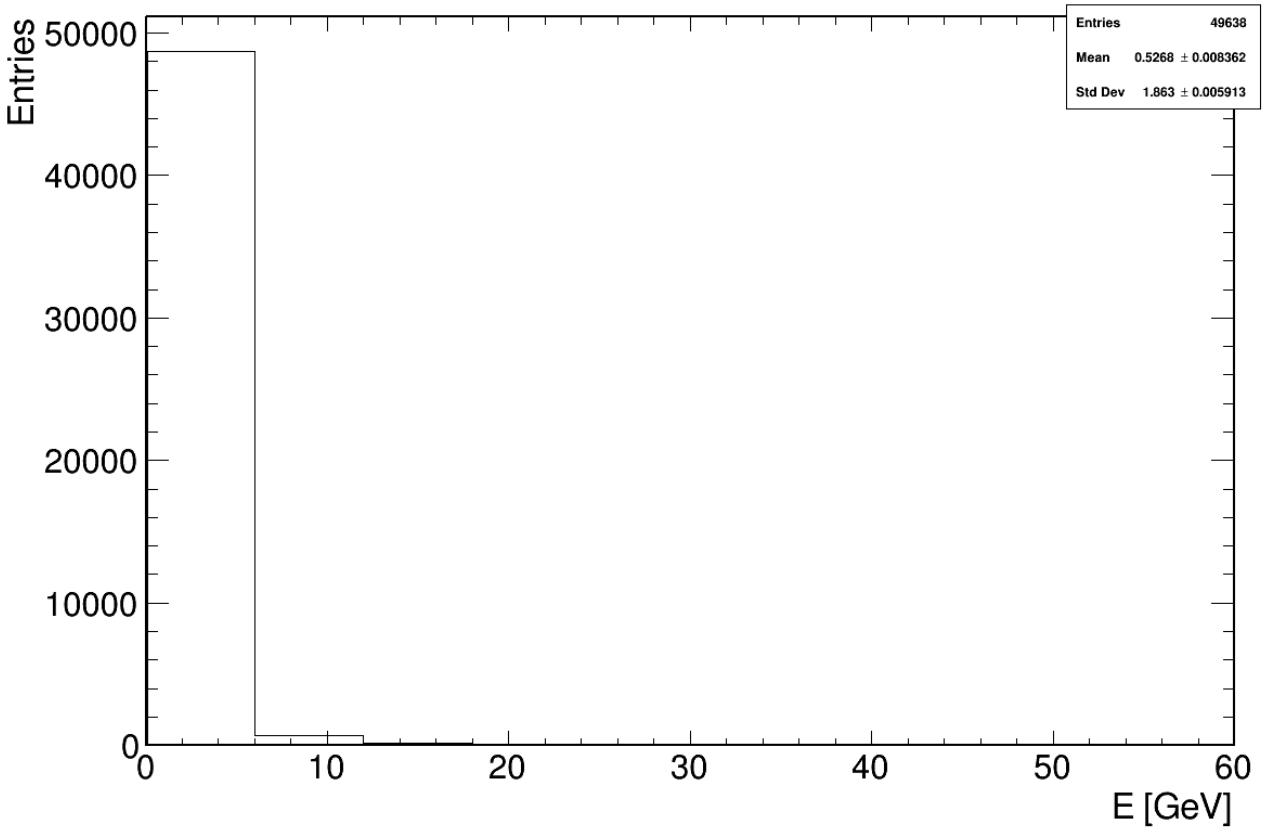


Track/Truth Match

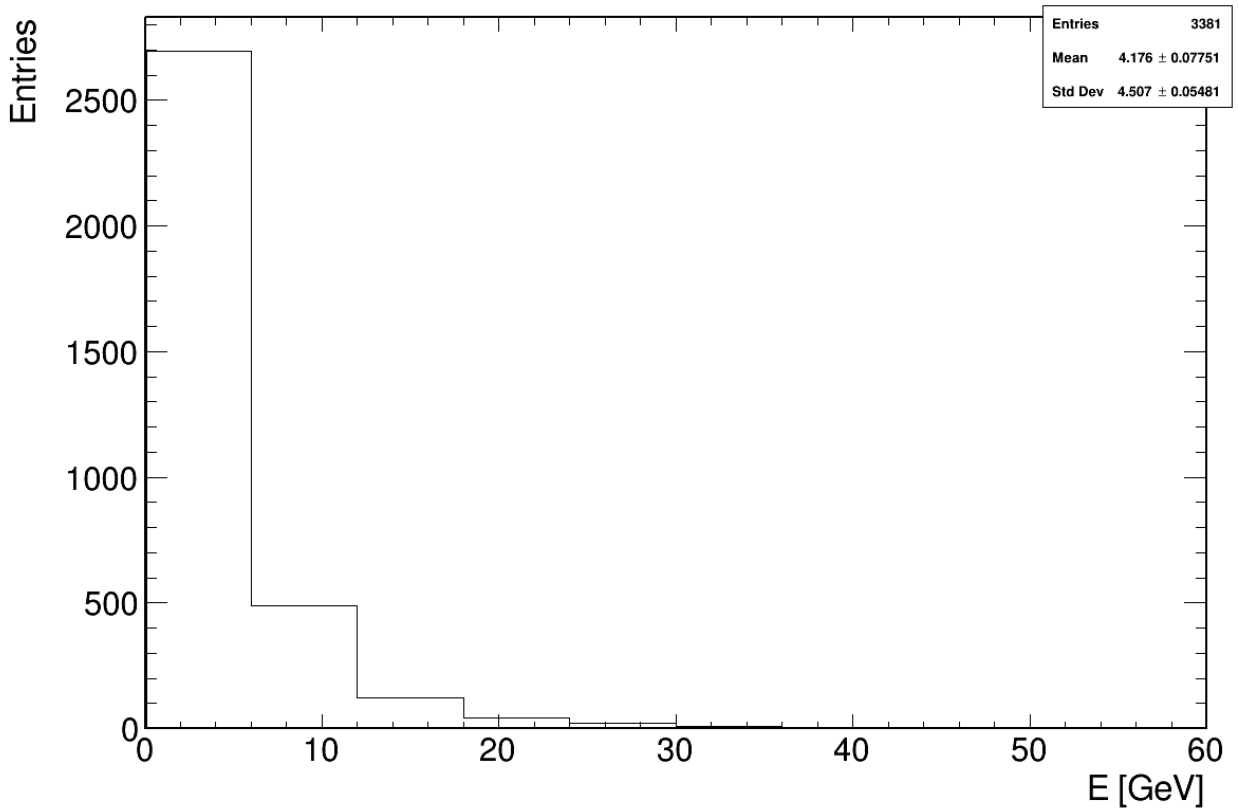




Energy per PFO

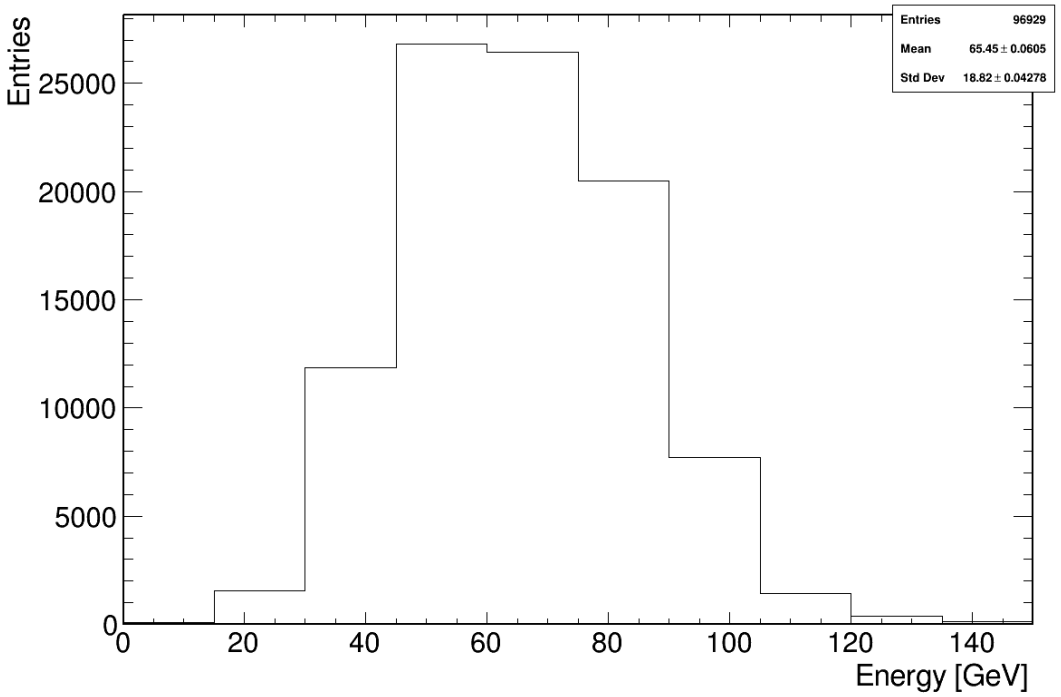


Energy per PFO with matched truth

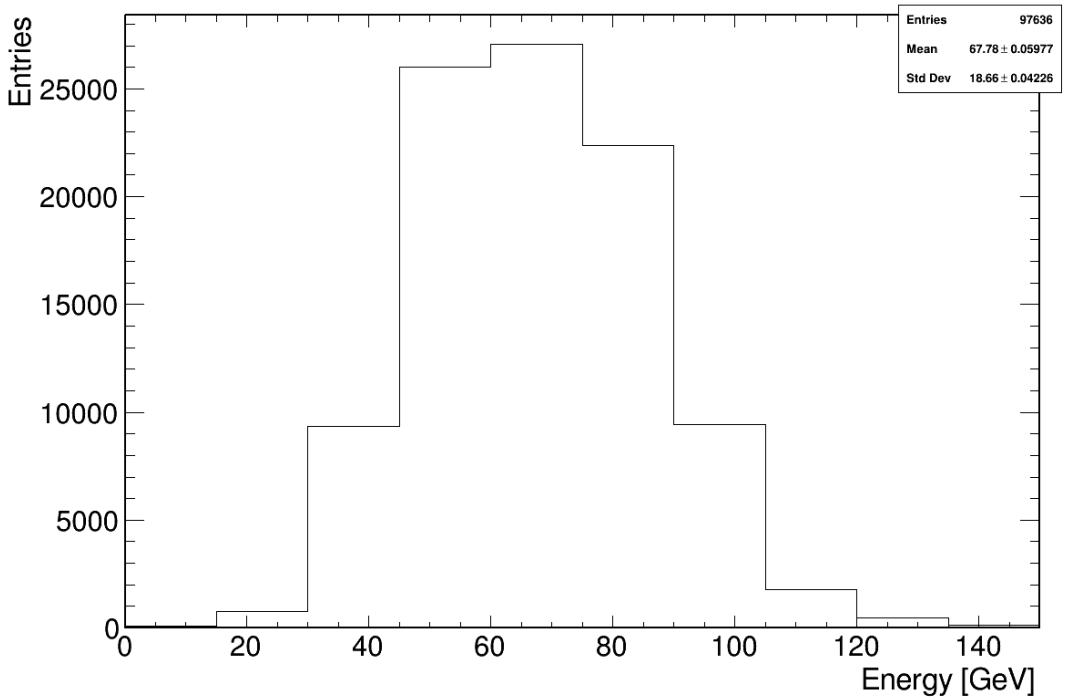


bb/cc jets

bb

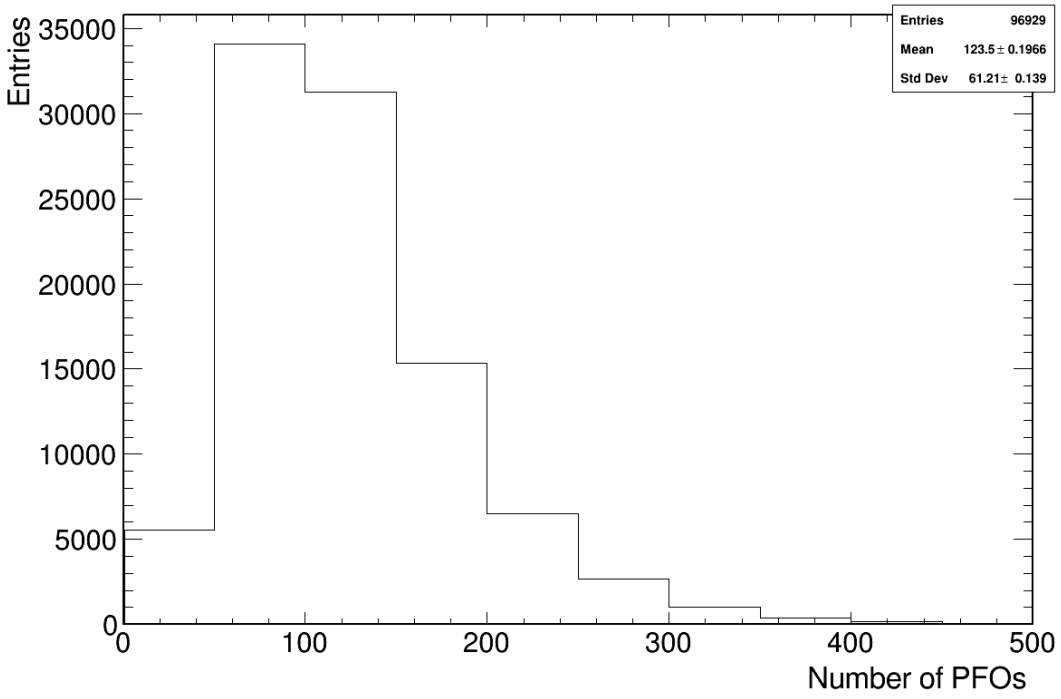


cc

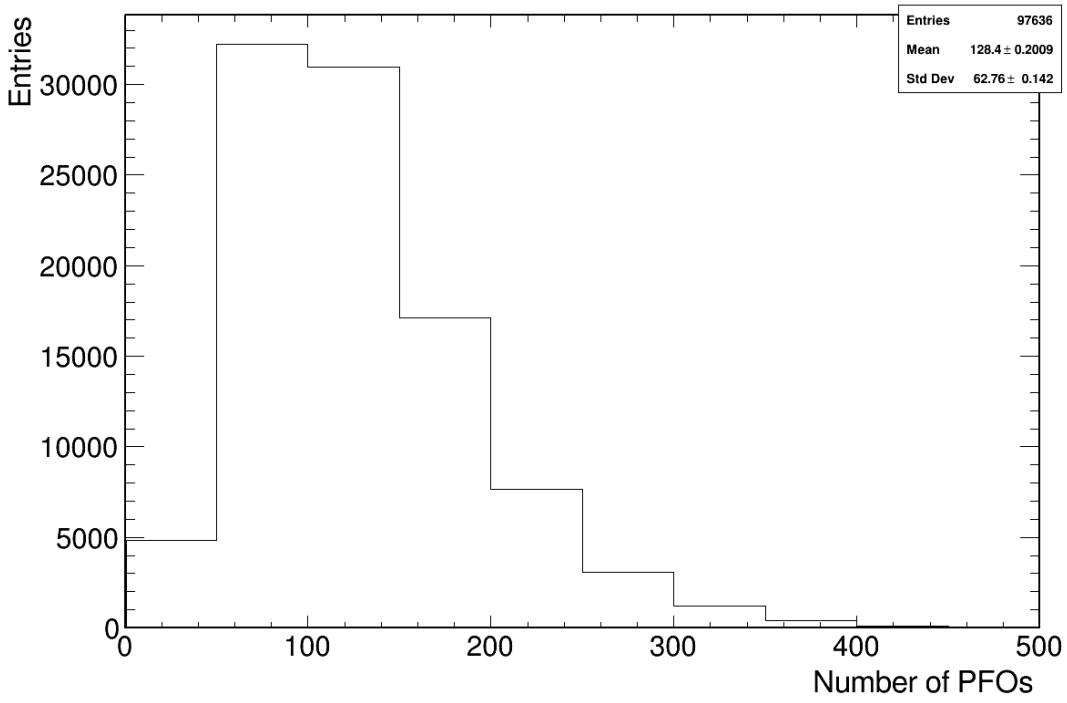


bb/cc jets

bb

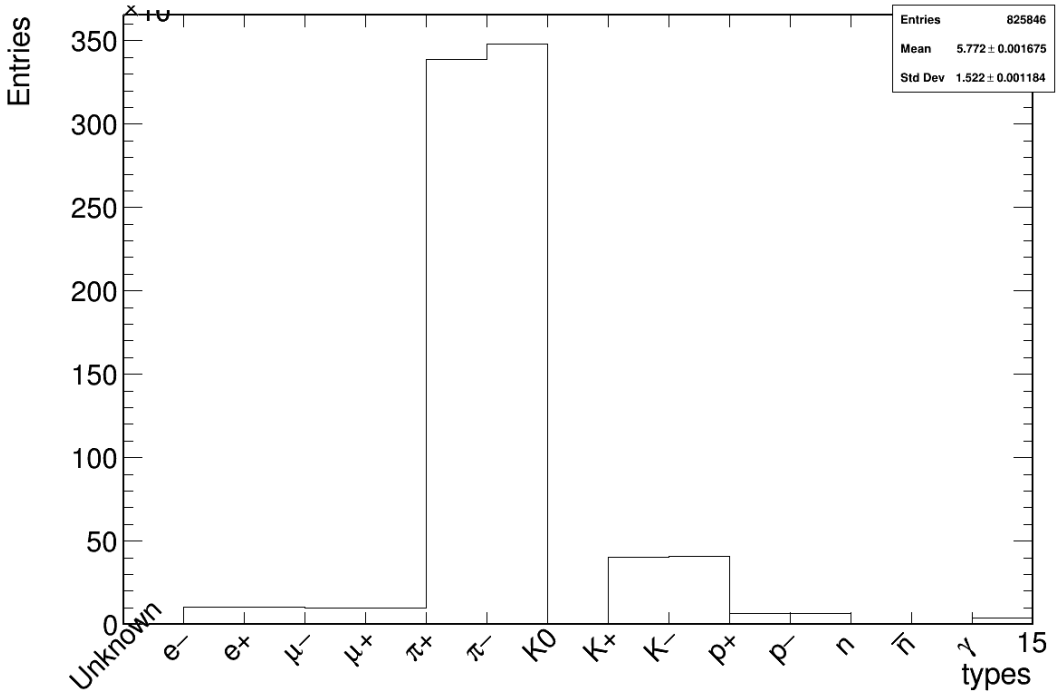


cc

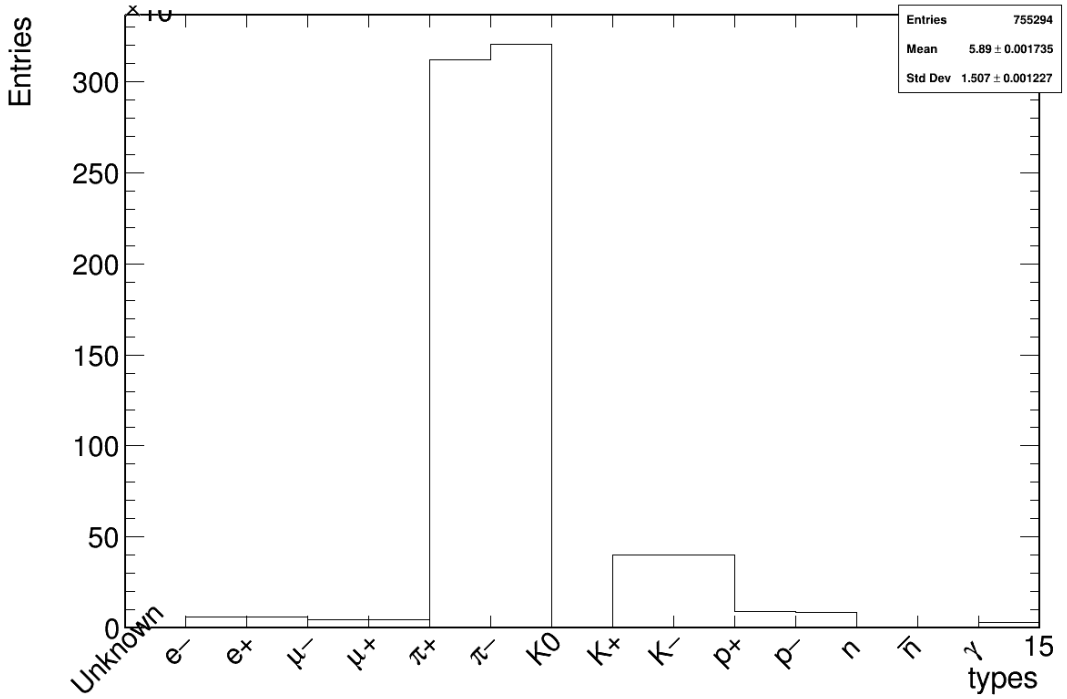


PFO types

bb

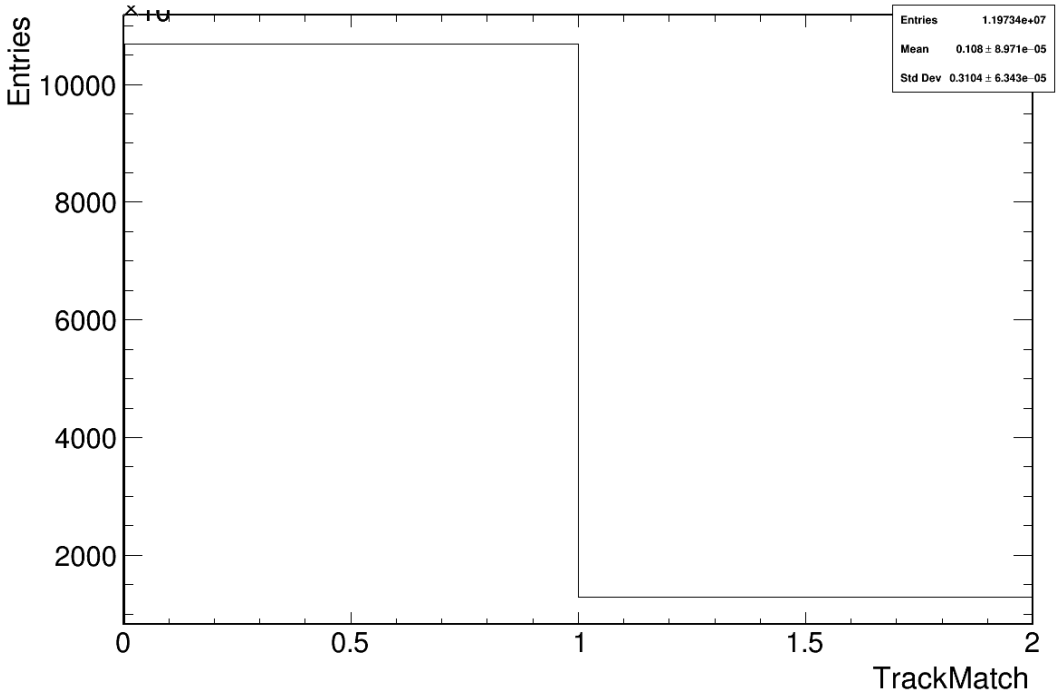


cc

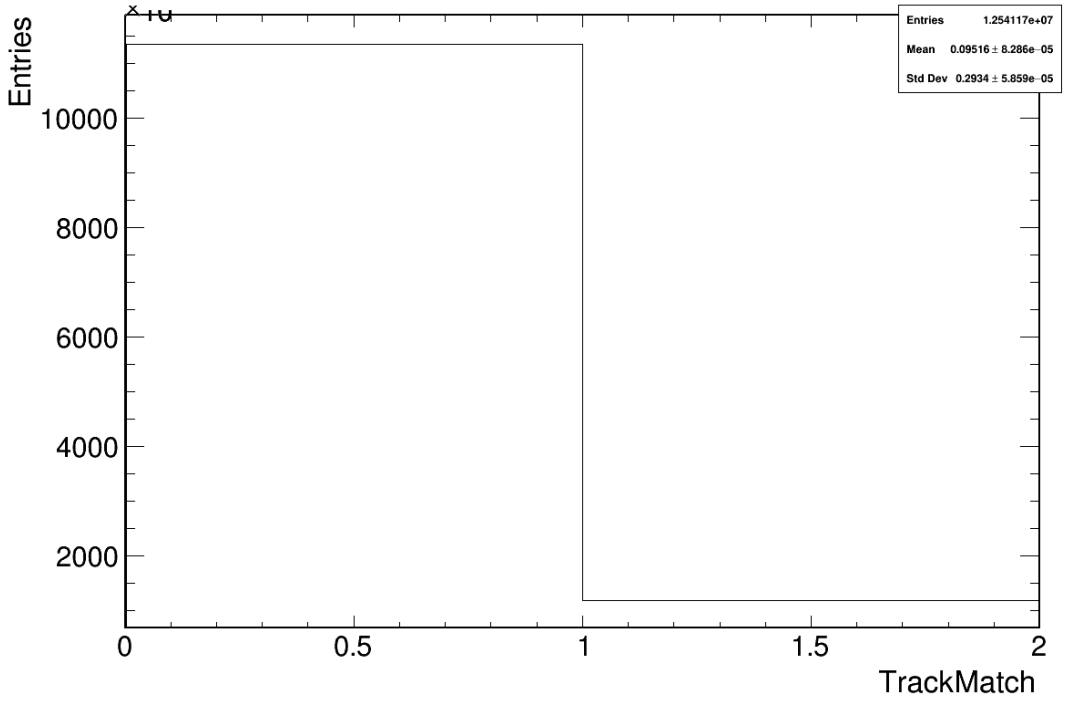


Trackmatching

bb

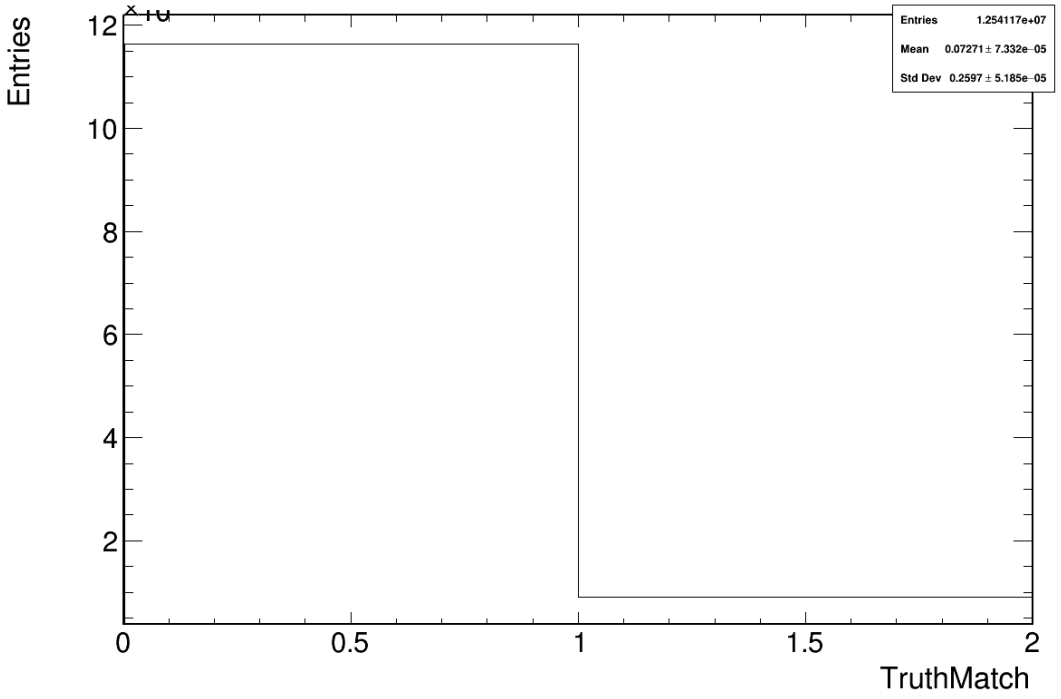


cc

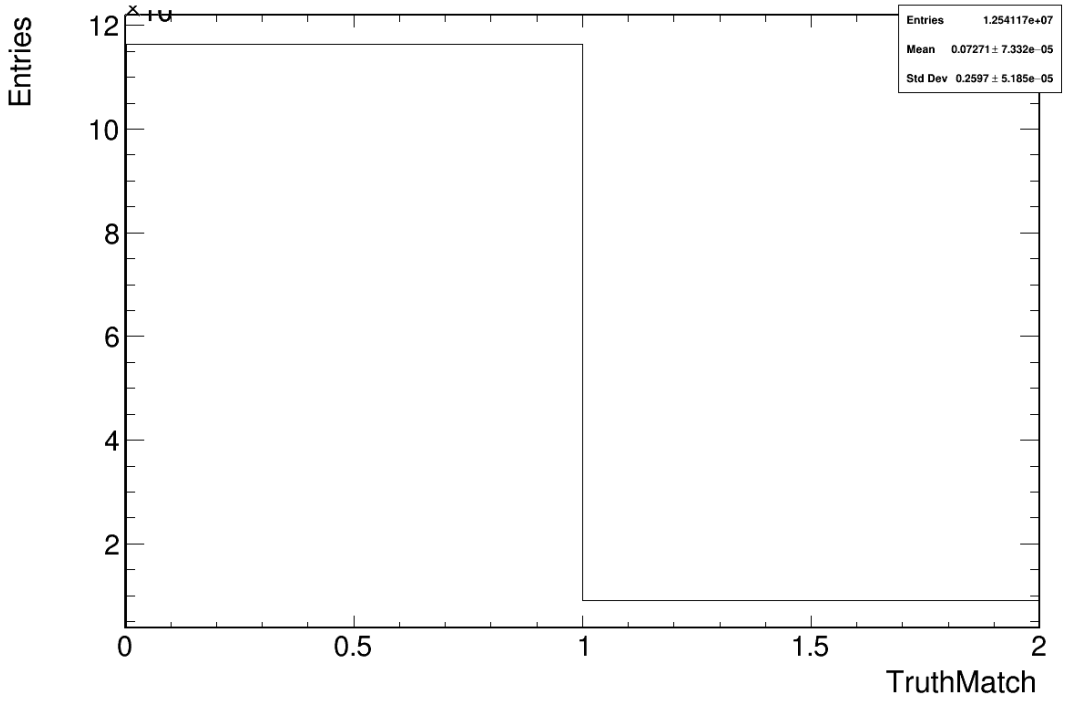


Truthmatching

bb

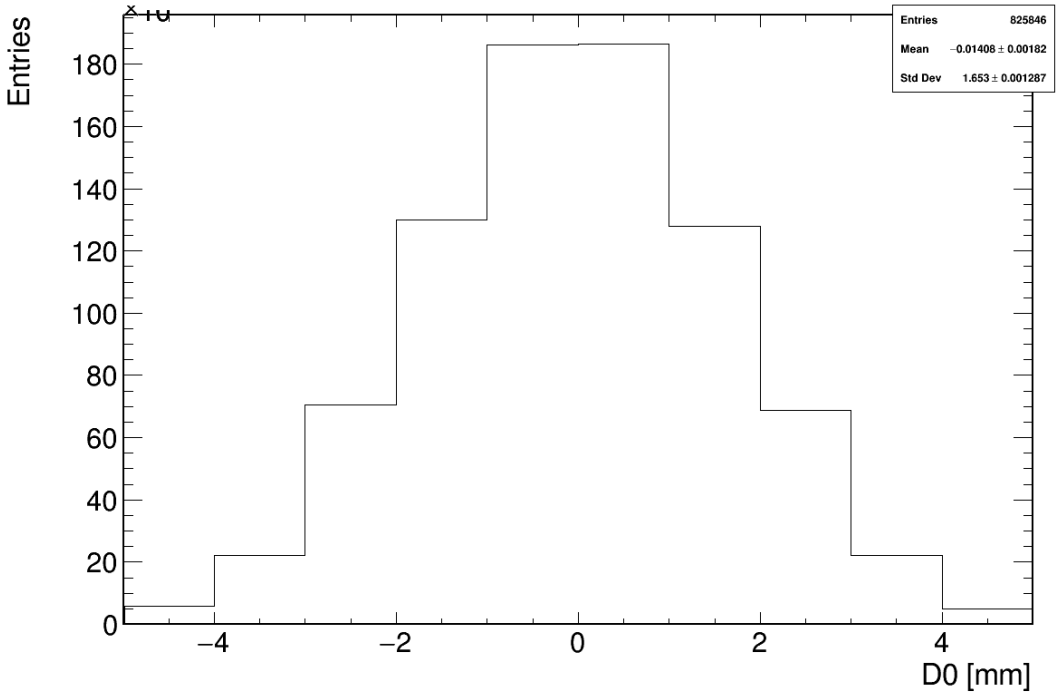


cc

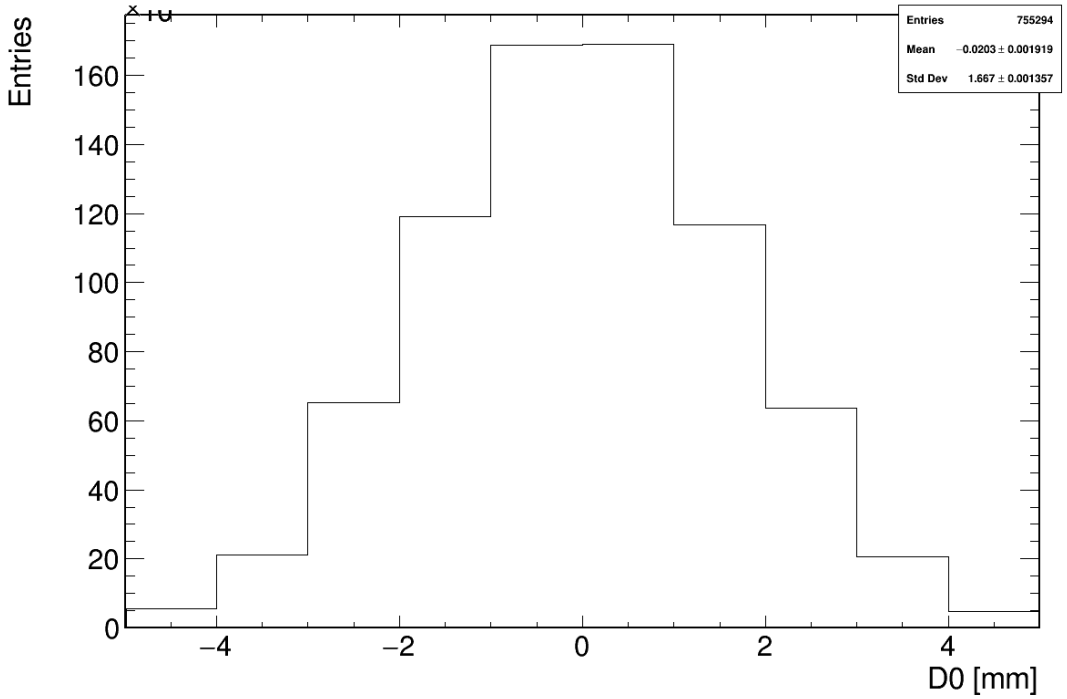


bb/cc d0

bb


Entries
Entries 825846
Mean −0.01408 ± 0.00182
Std Dev 1.653 ± 0.001287
0
20
40
60
80
100
120
140
160
180
−4
−2
0
2
4
D0 [mm]

cc


Entries
Entries 755294
Mean −0.0203 ± 0.001919
Std Dev 1.667 ± 0.001357
0
20
40
60
80
100
120
140
160
−4
−2
0
2
4
D0 [mm]