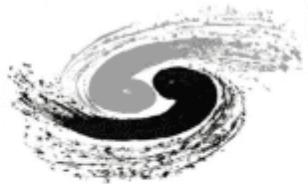


中国科学院高能物理研究所
Institute of High Energy Physics
Chinese Academy of Sciences

PID using XGBoost

Geliang Liu

Apr. 9th, 2025

Introduction

Build XGBoost model to distinguish 5 flavors of charged particles:

- e, μ, π, K, p

One individual XGBoost in each bin of momentum and θ.

- Samples: single e, μ, π, K, p particle gun samples, uniformly distributed in p = 1-10 or 10-80 GeV, θ = 8-172°, individually for positive and negative charge.
- /cms/user/liugeliang/CEPC/202503/Production/ParticleGun

12 Variables used:

- Tracker: TOF, dN/dx
- Calorimeters: E_{ECAL}/p, E_{HCAL}/p, l_{HCAL}, $R_{HCAL}^{modiere}$, $N_{hadClus}$ New!
- Muon detector: ΔR(trk, hit) of the three closest hits, in different superlayers; ΔR(trk, hit) of the closest hit in the last two layers; N_{muHit}

Output:

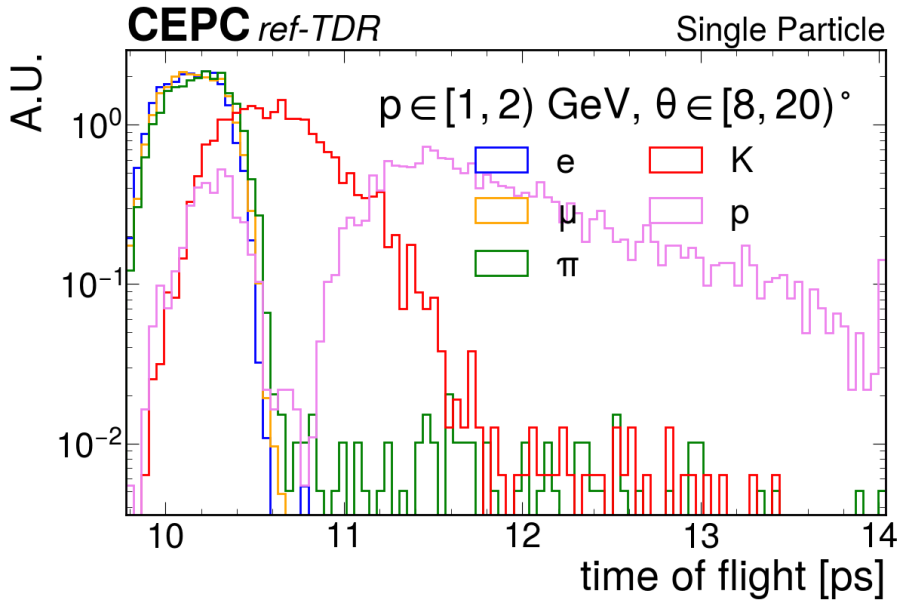
- Probability of the PFO to be each flavor: $prob_f$

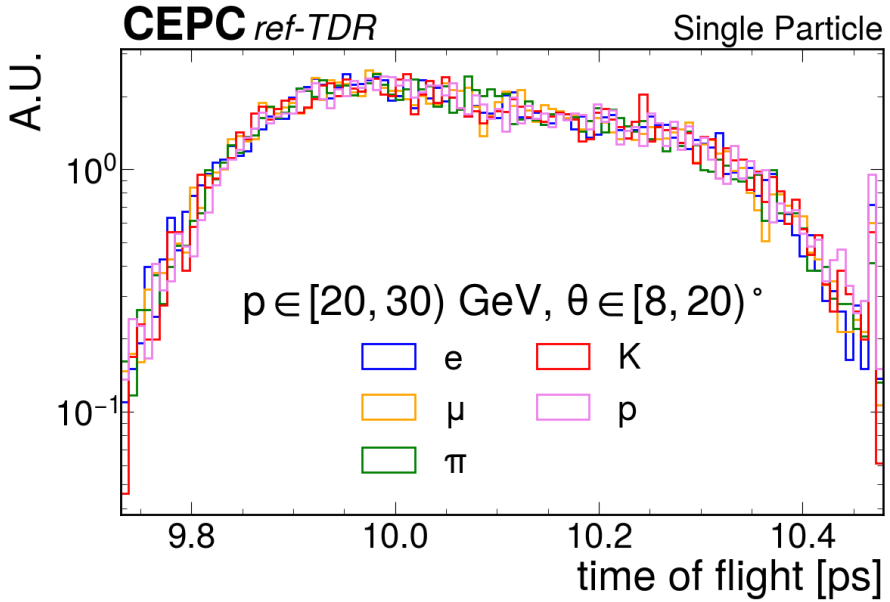
Variables

TOF

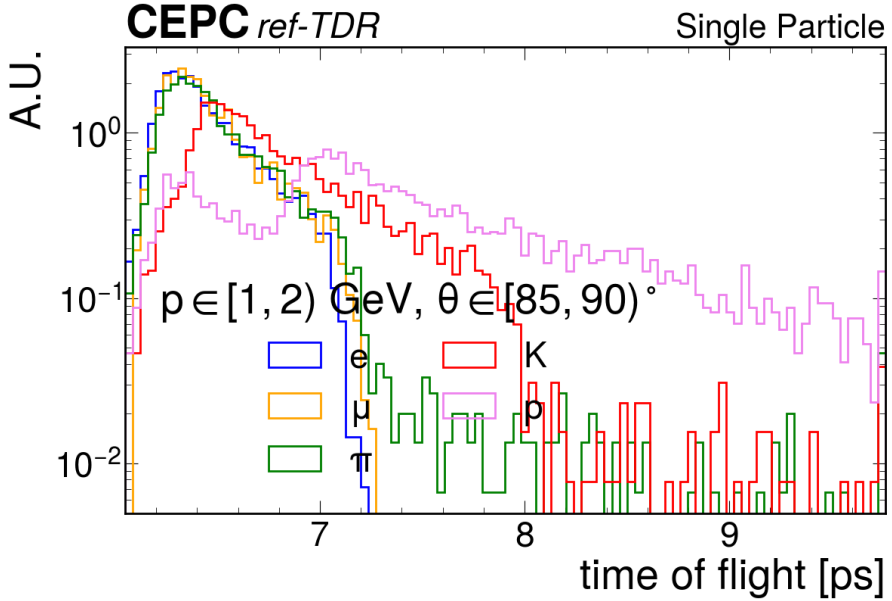
low p

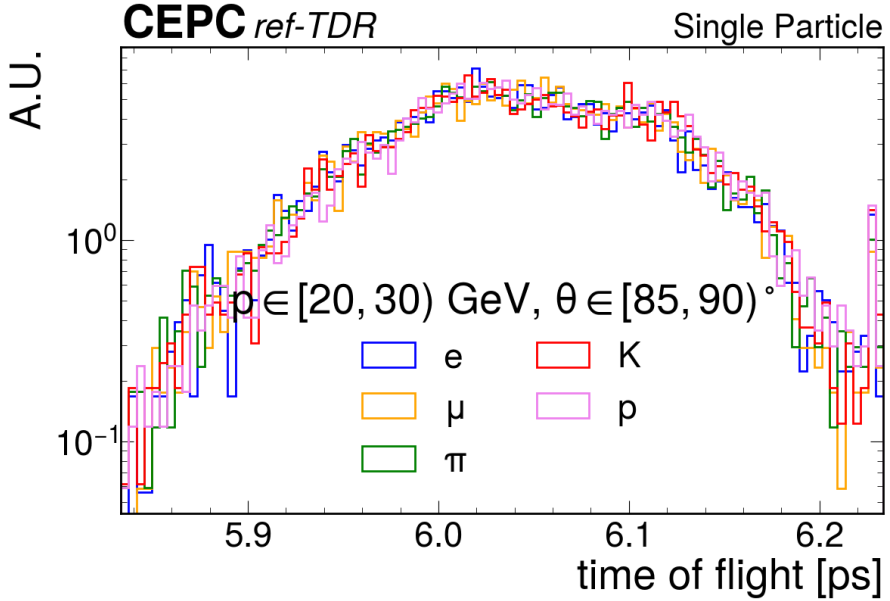
high p





endcap





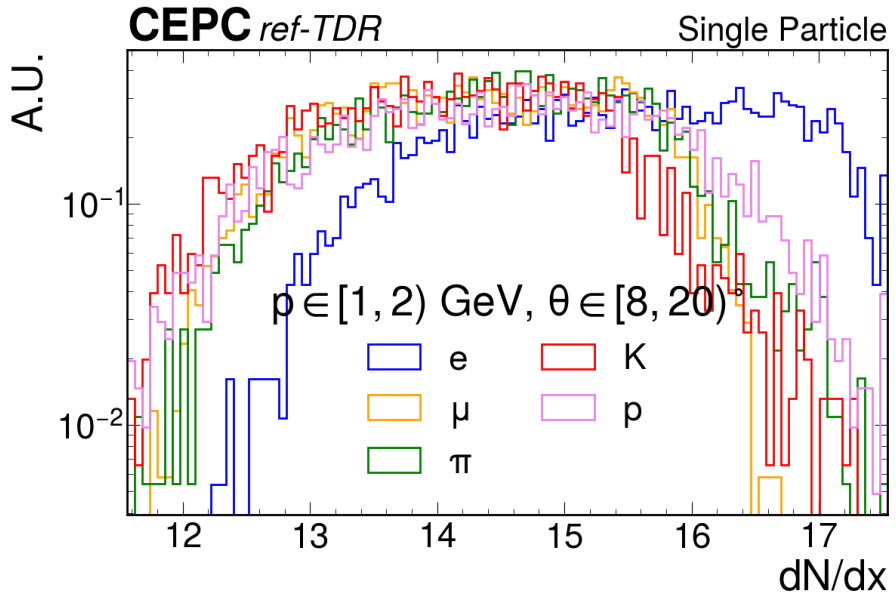
barrel

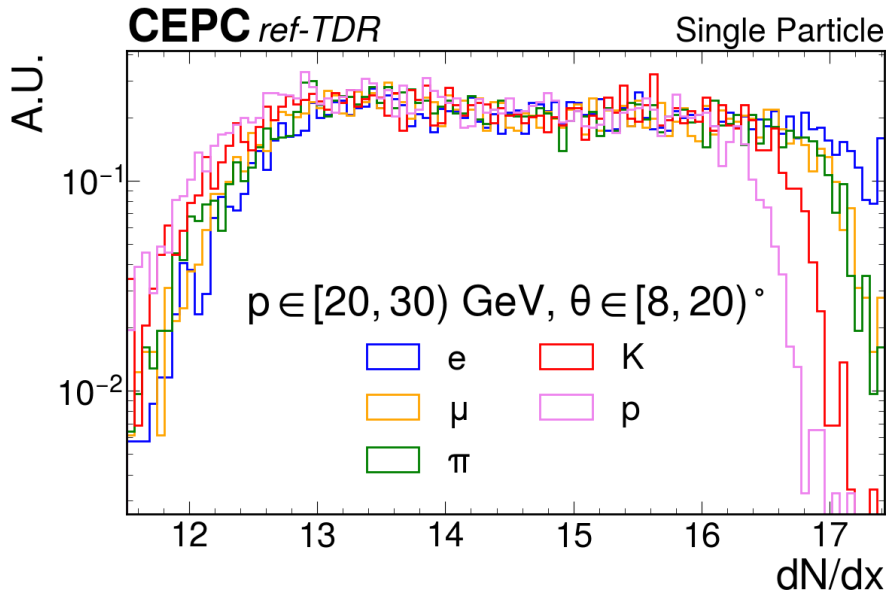
Variables

dN/dx

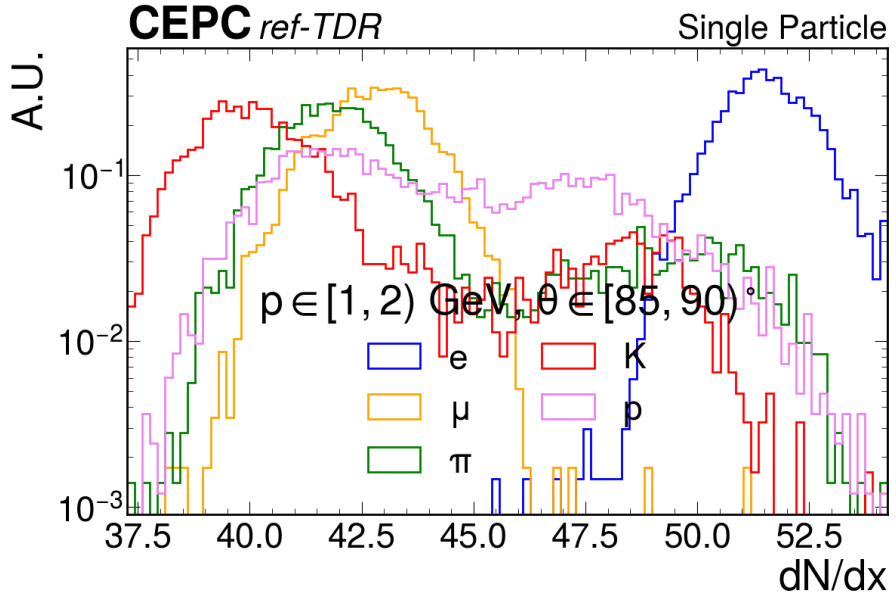
low p

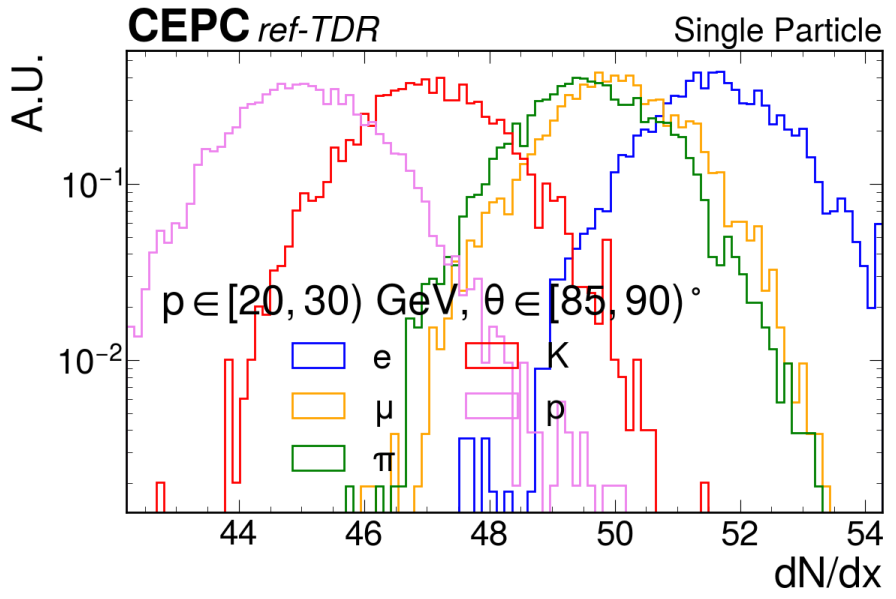
high p





endcap





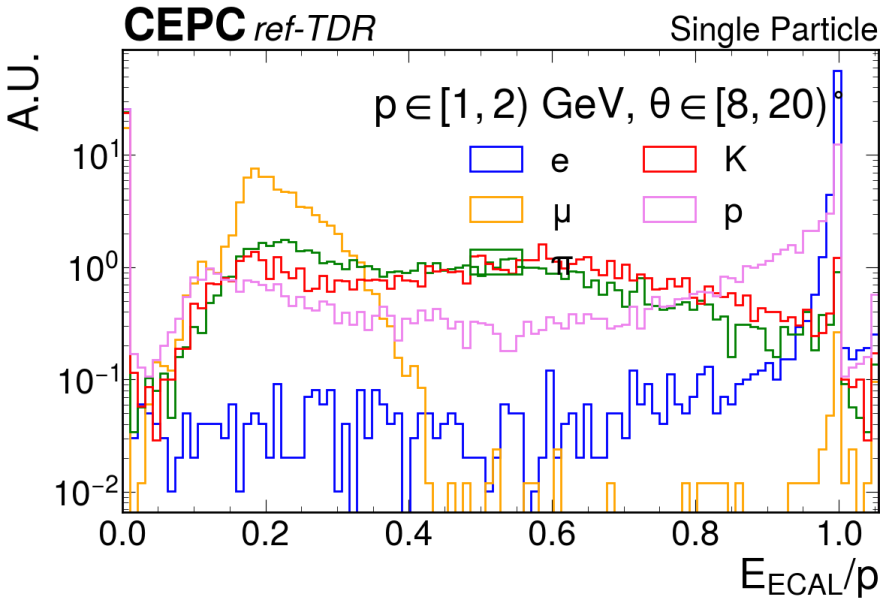
barrel

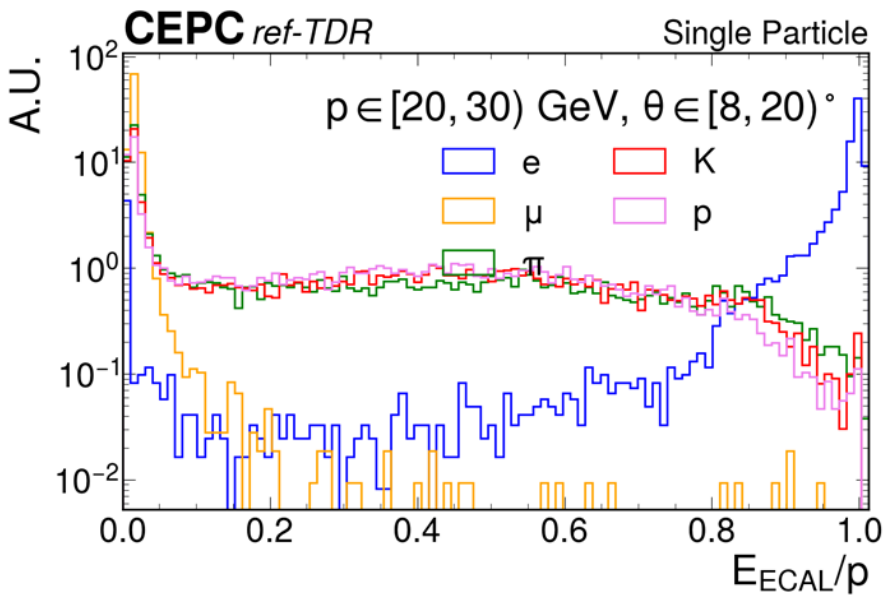
Variables

E_{ECAL}/p

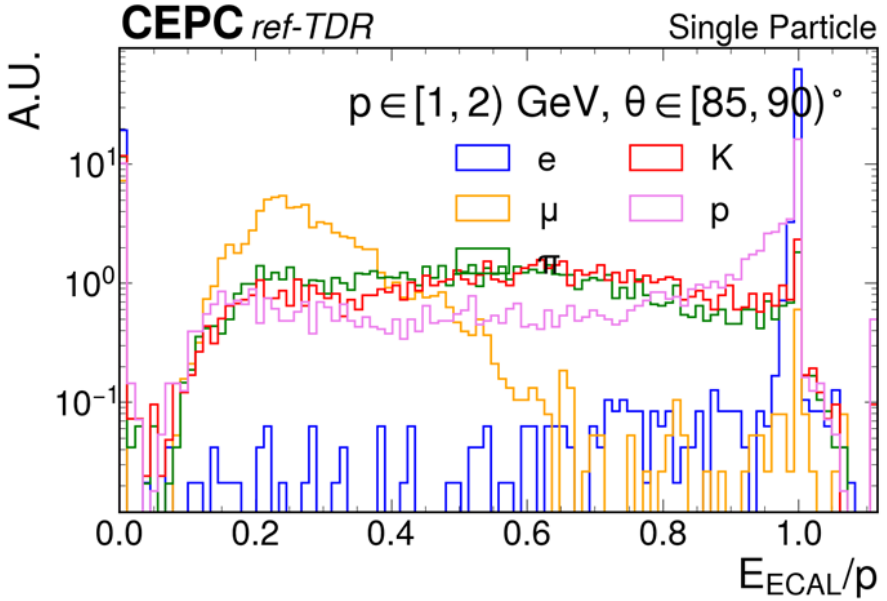
low p

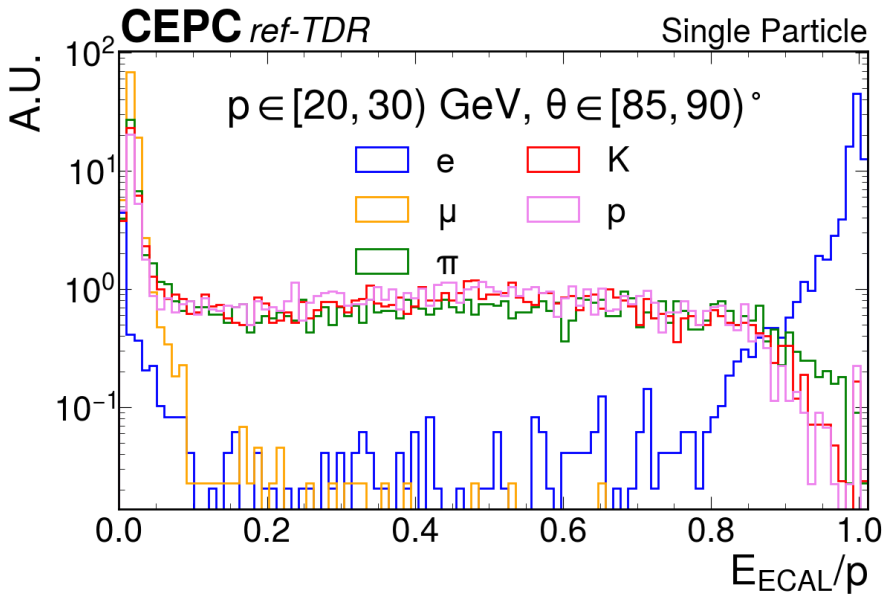
high p





endcap





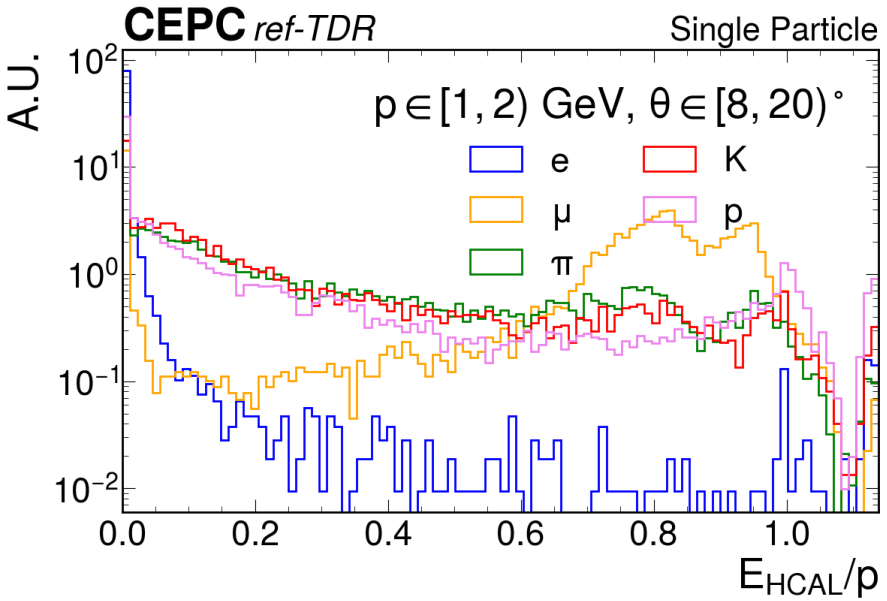
barrel

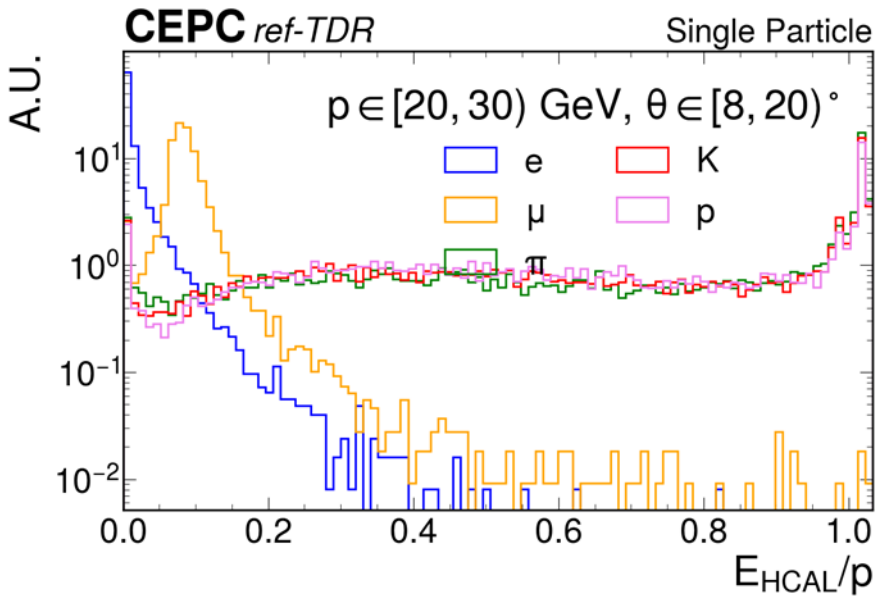
Variables

E_{HCAL}/p

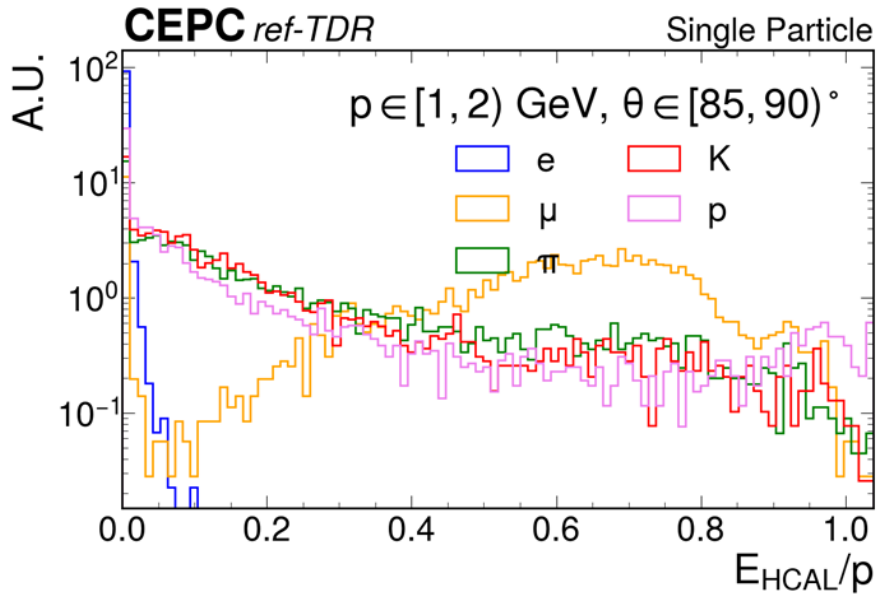
low p

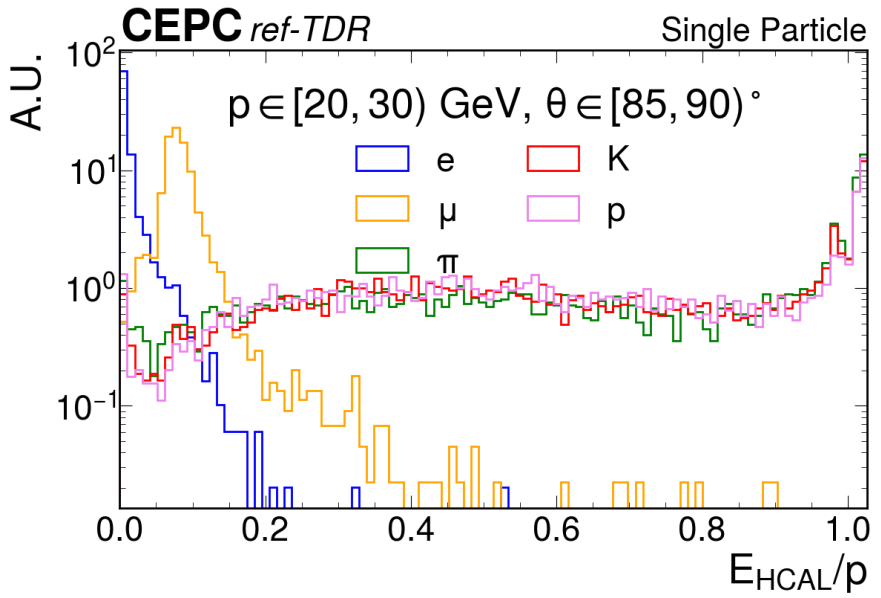
high p





endcap





barrel

Variables

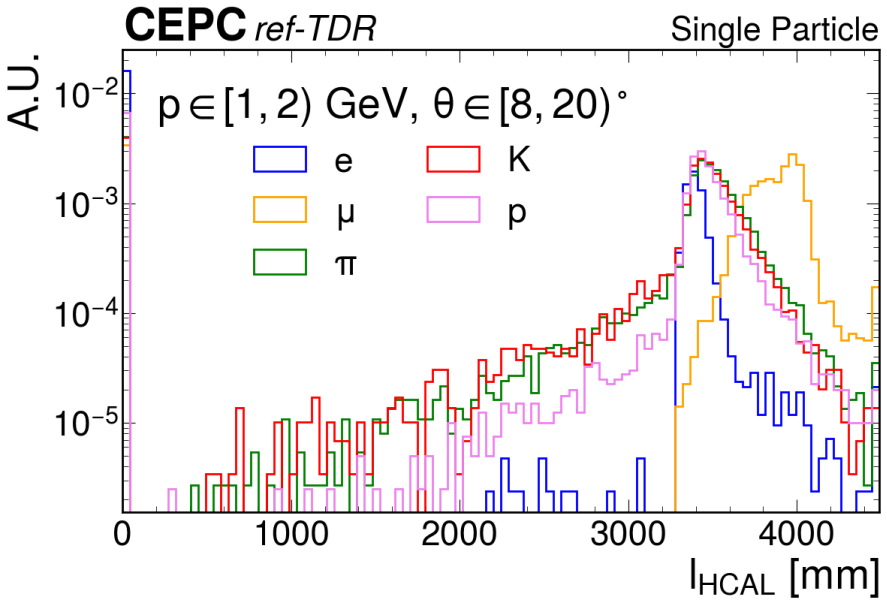
l_{HCAL}

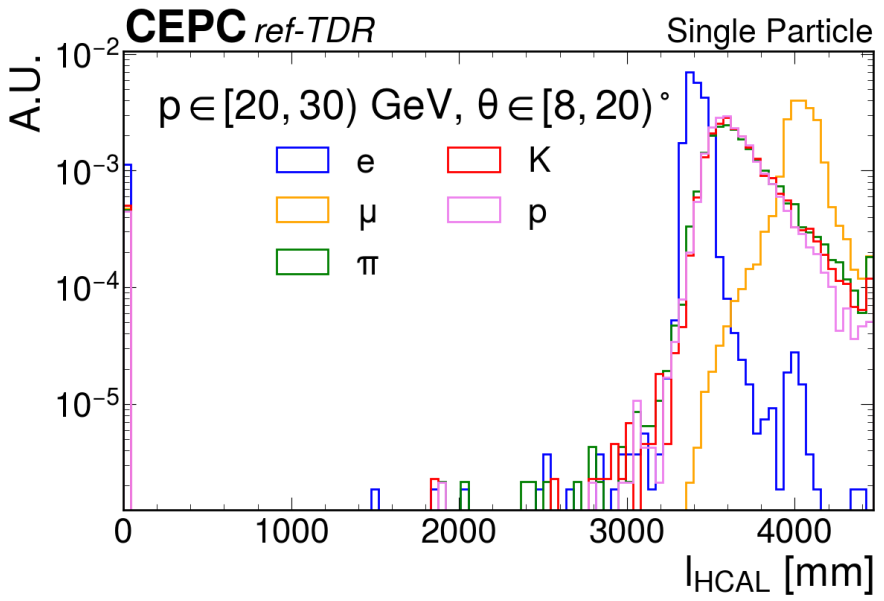
low p

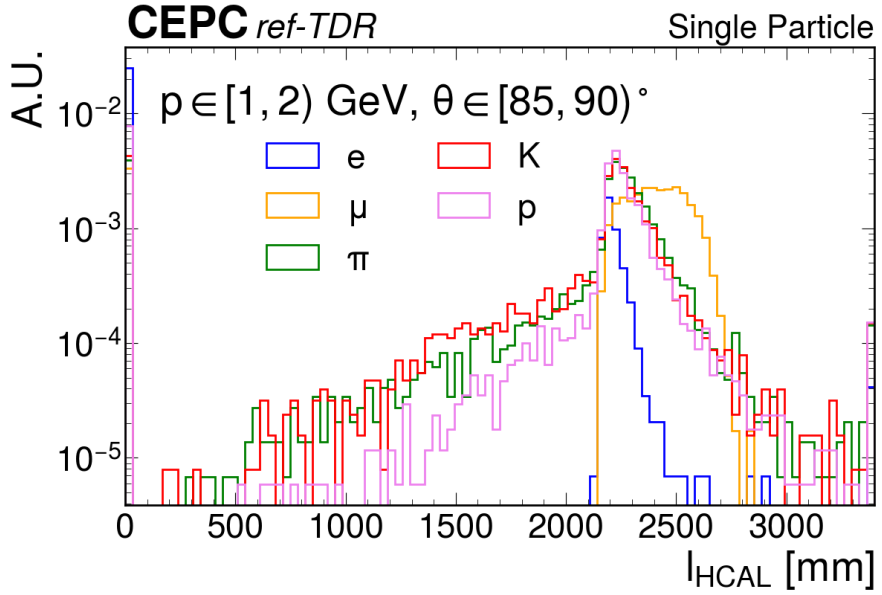
high p

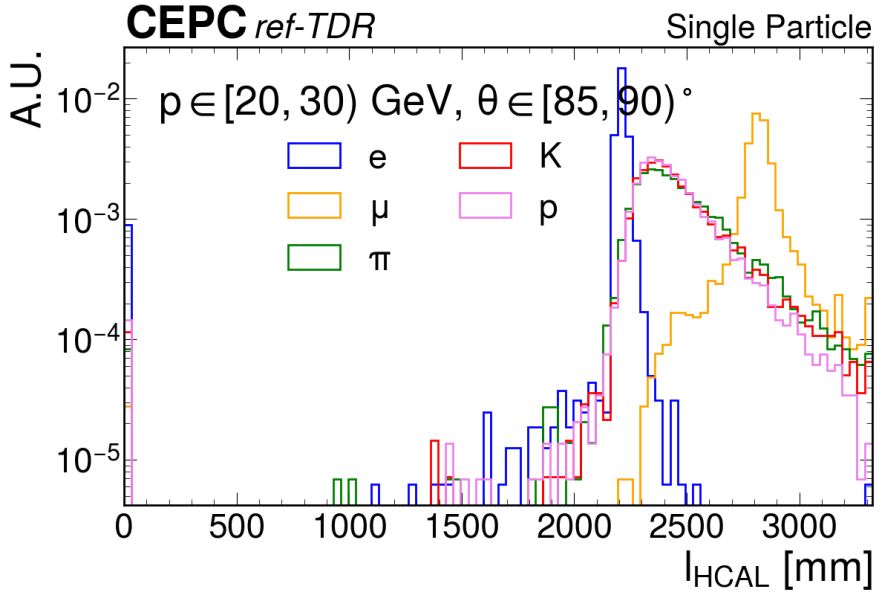
endcap

barrel







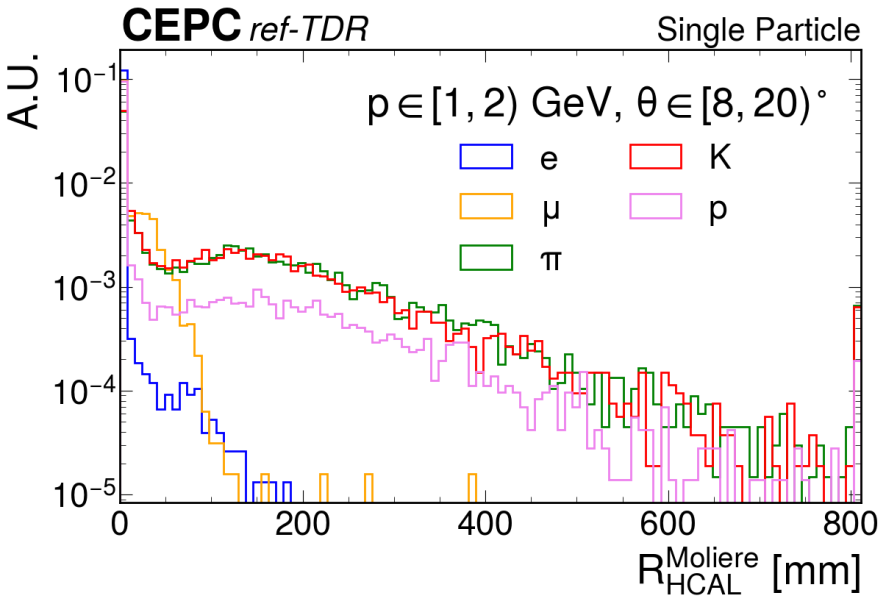


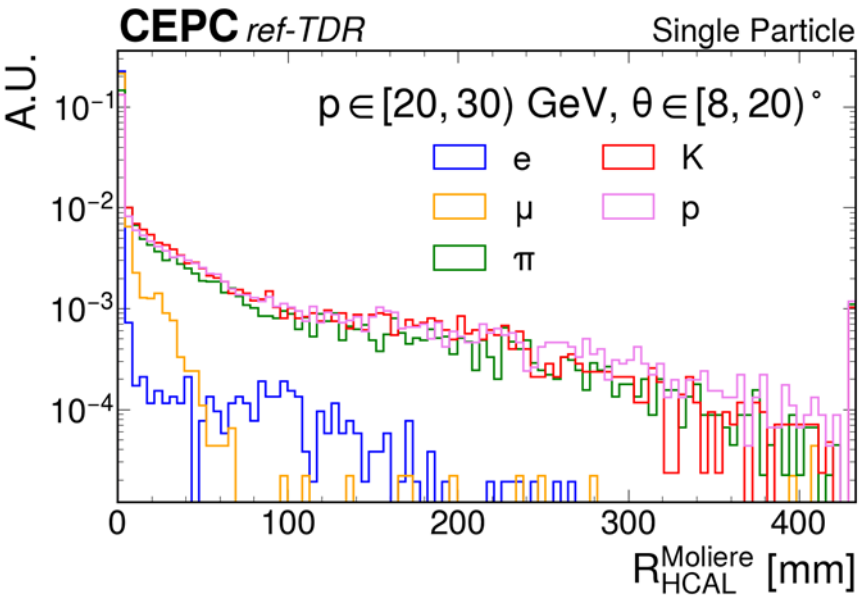
Variables

$R^{moliere}_{HCAL}$

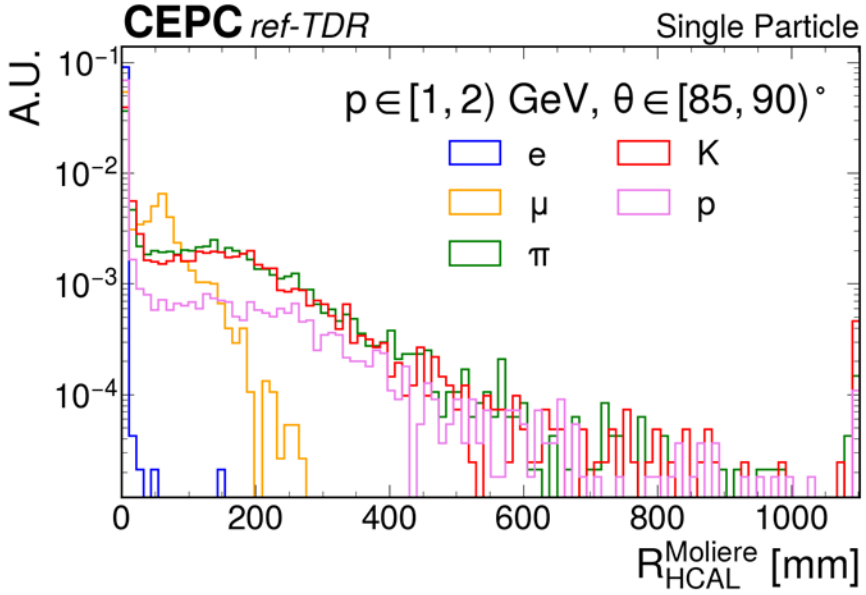
low p

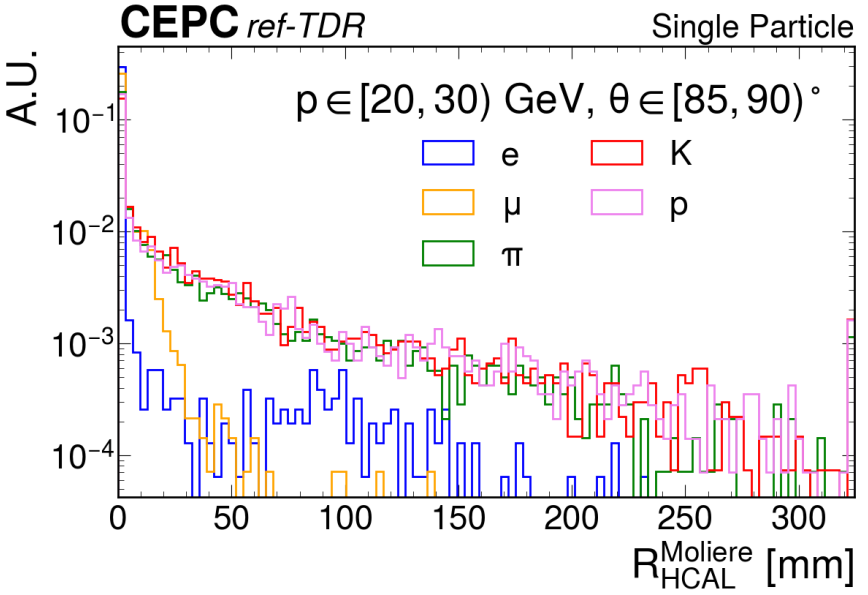
high p





endcap





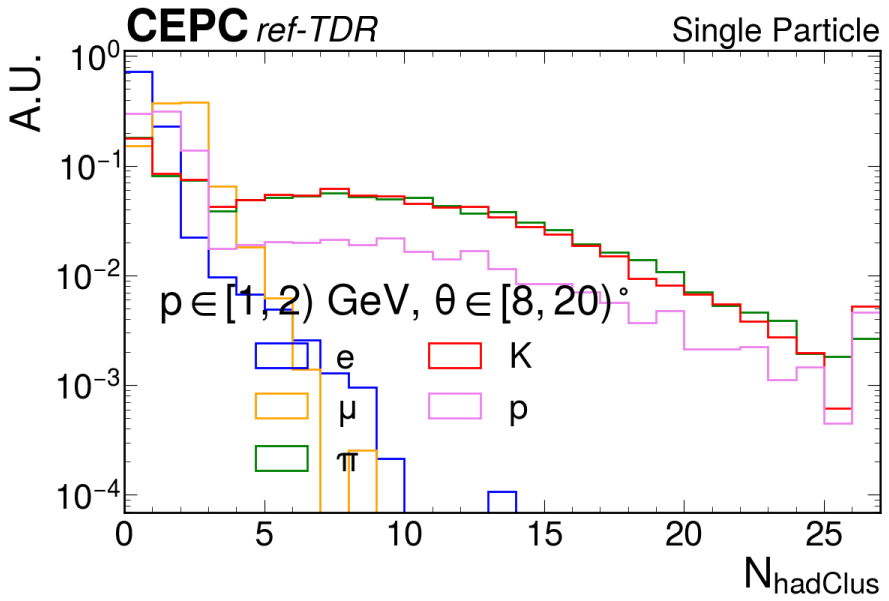
barrel

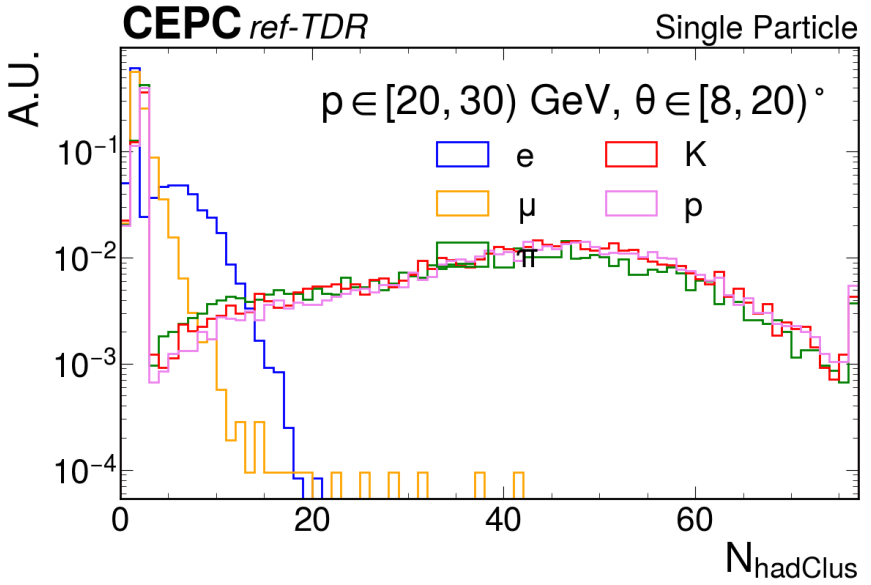
Variables

$N_{hadClus}$

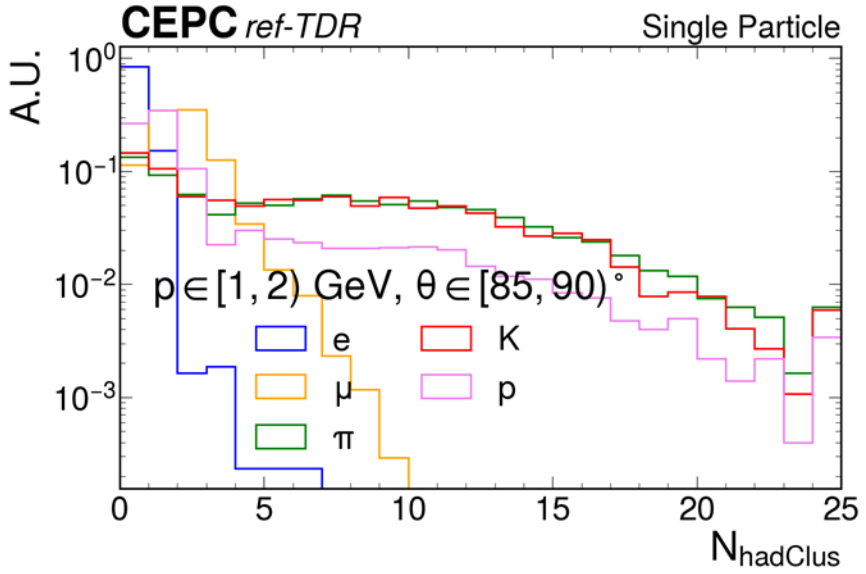
low p

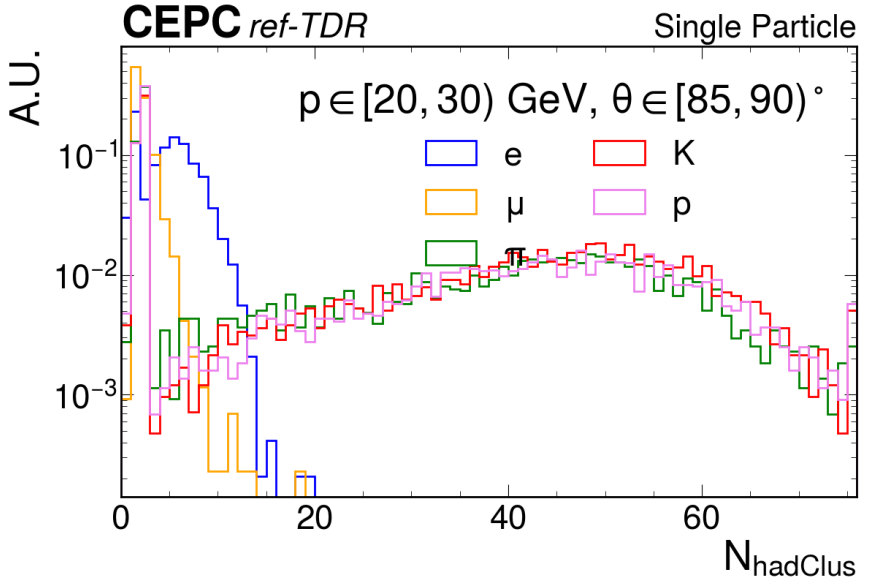
high p





endcap





barrel

Variables

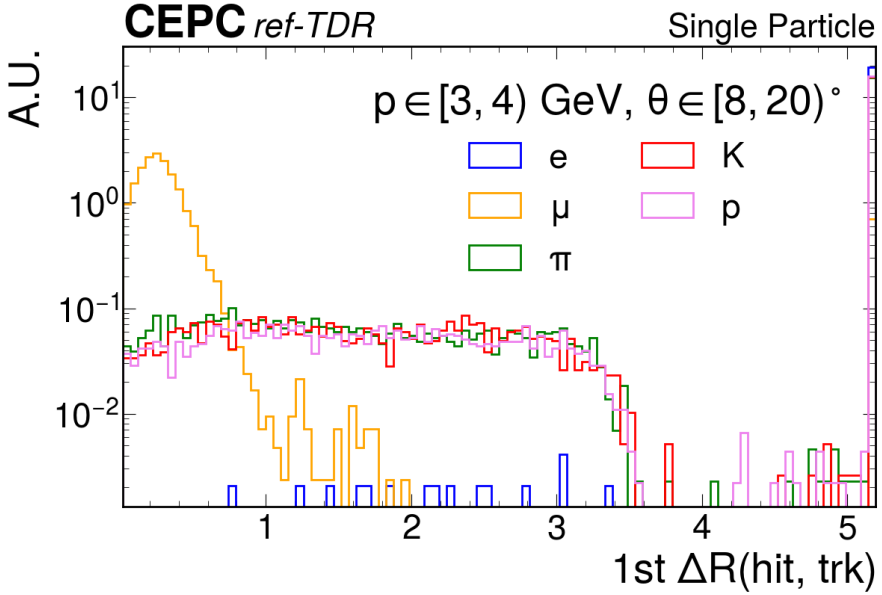
min ΔR

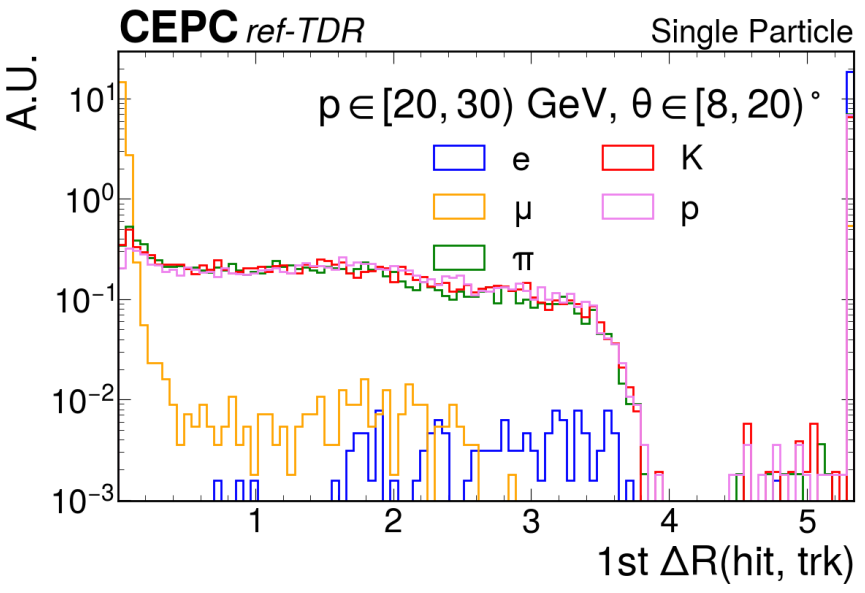
low p

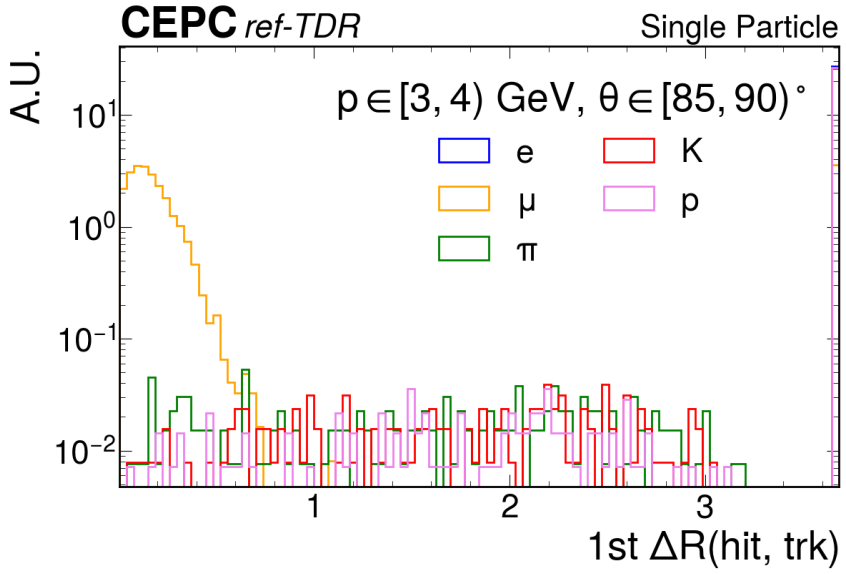
high p

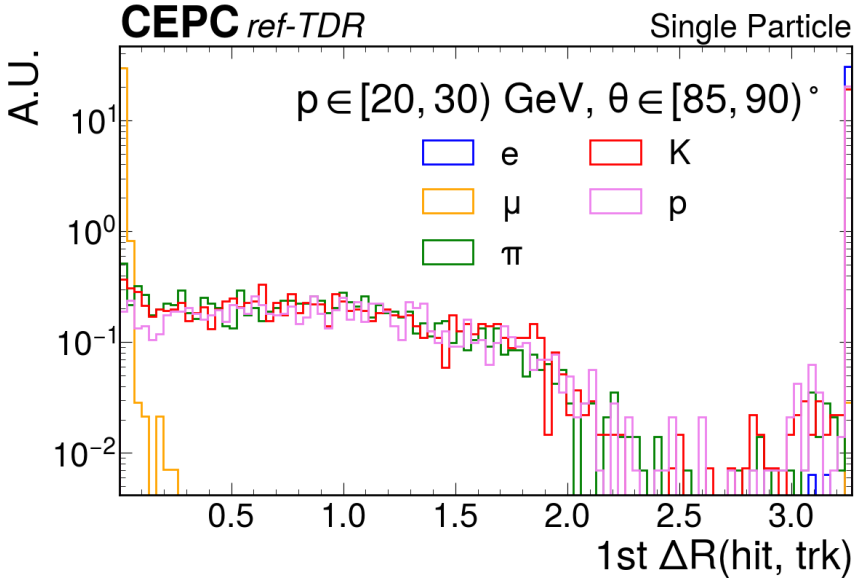
endcap

barrel







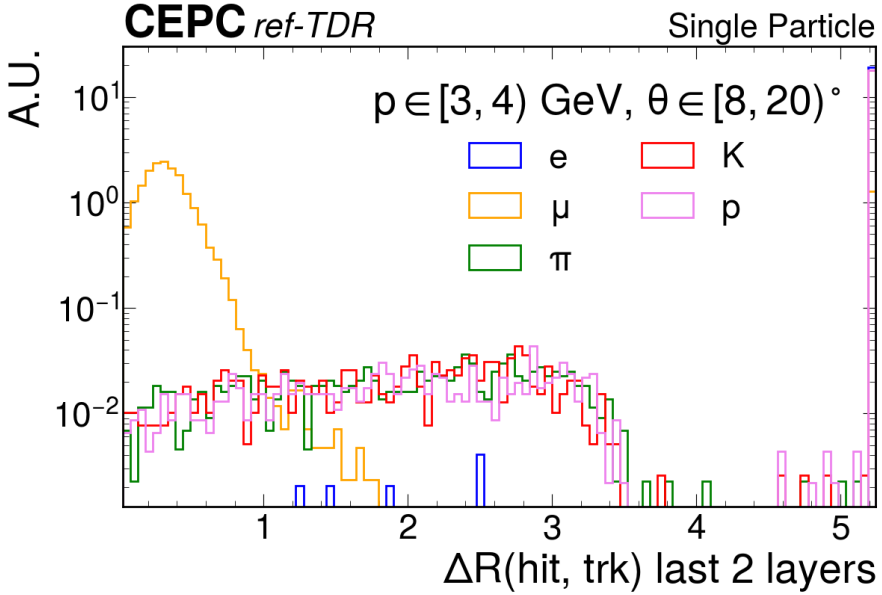


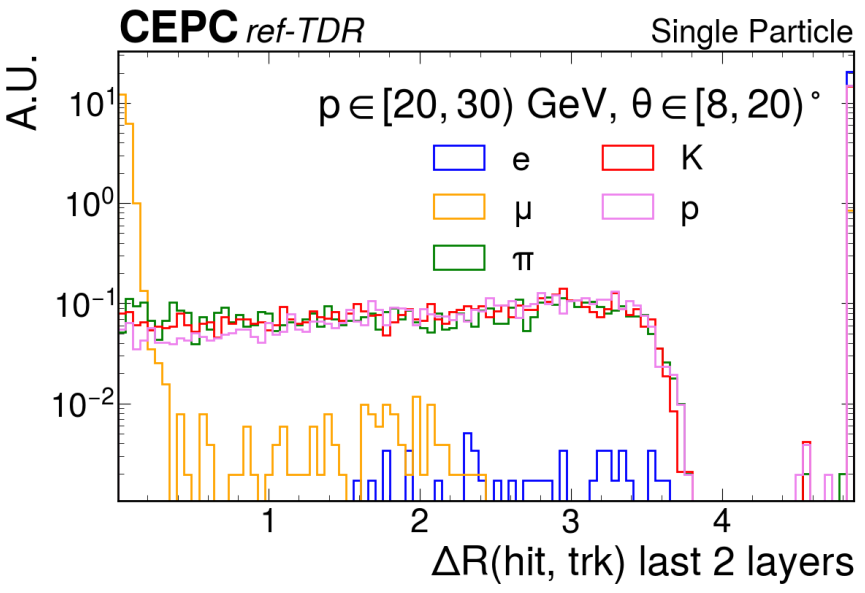
Variables

min ΔR in the last two superlayers

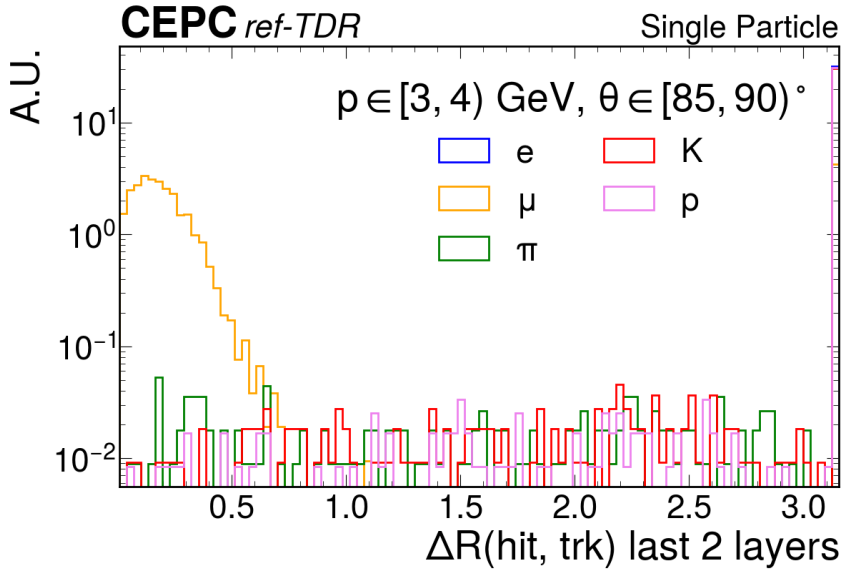
low p

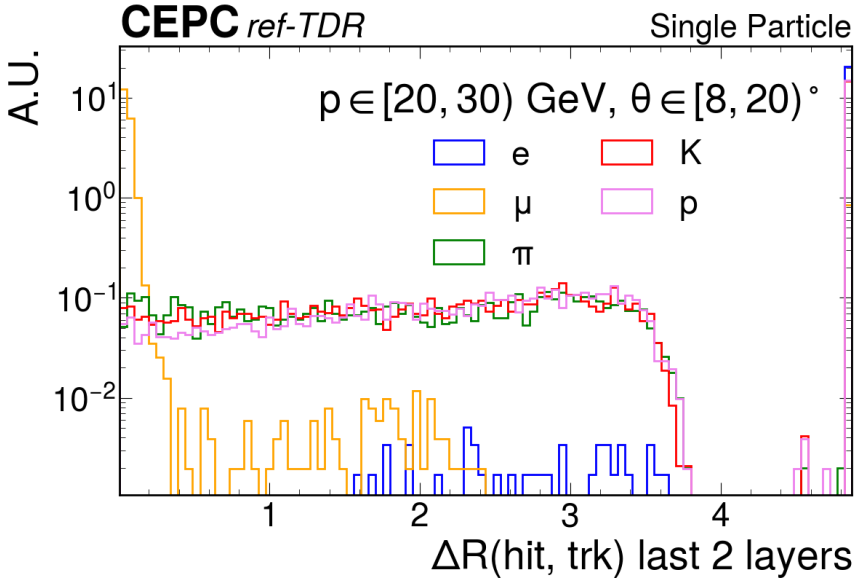
high p





endcap





barrel

Variables

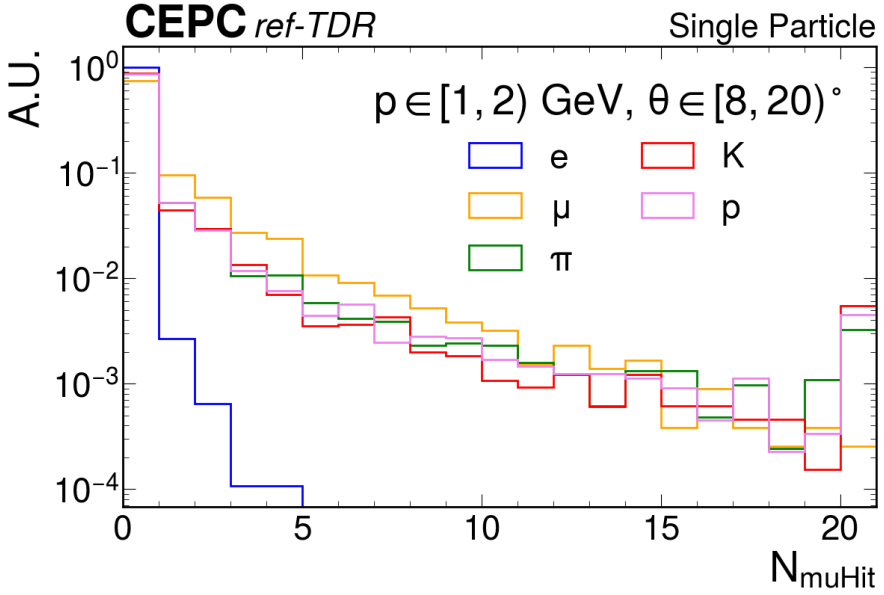
N_{muHit}

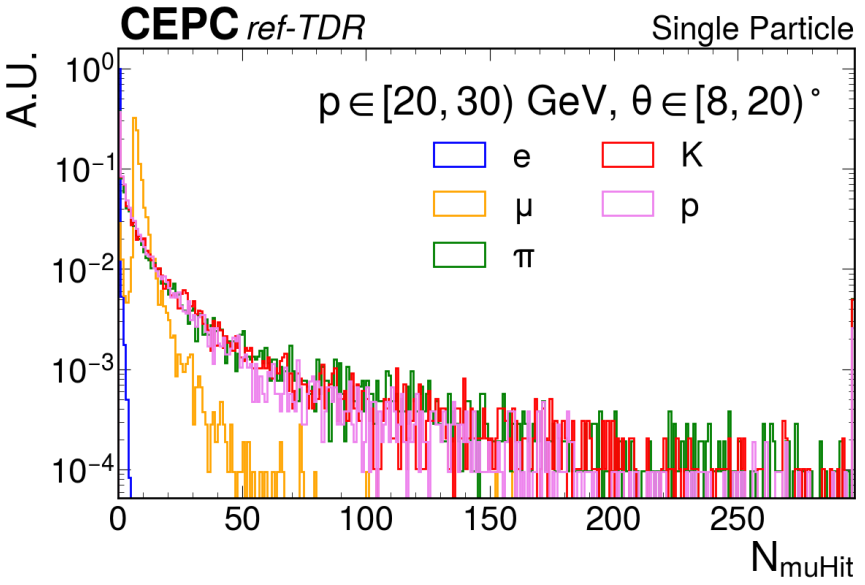
low p

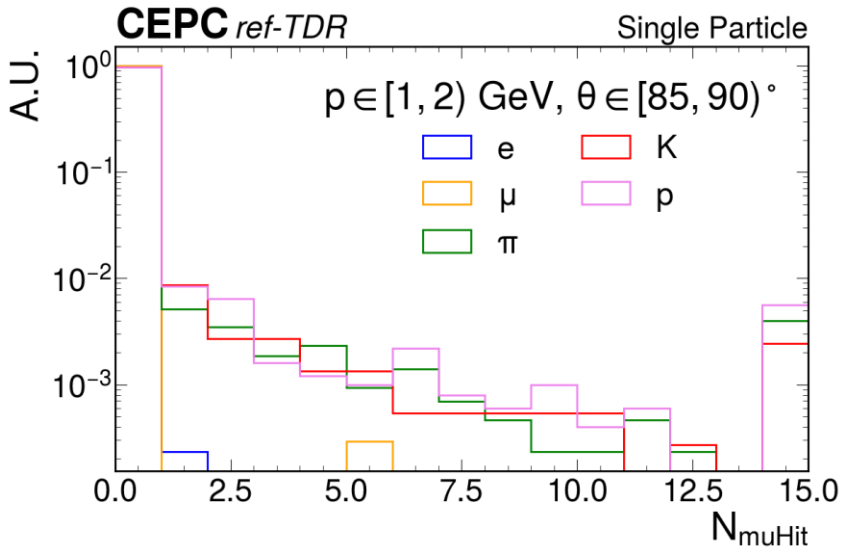
high p

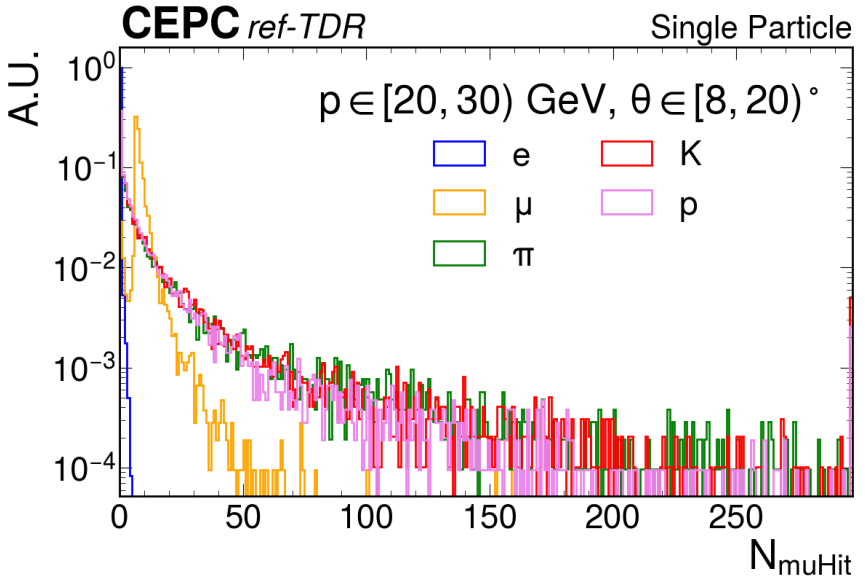
endcap

barrel

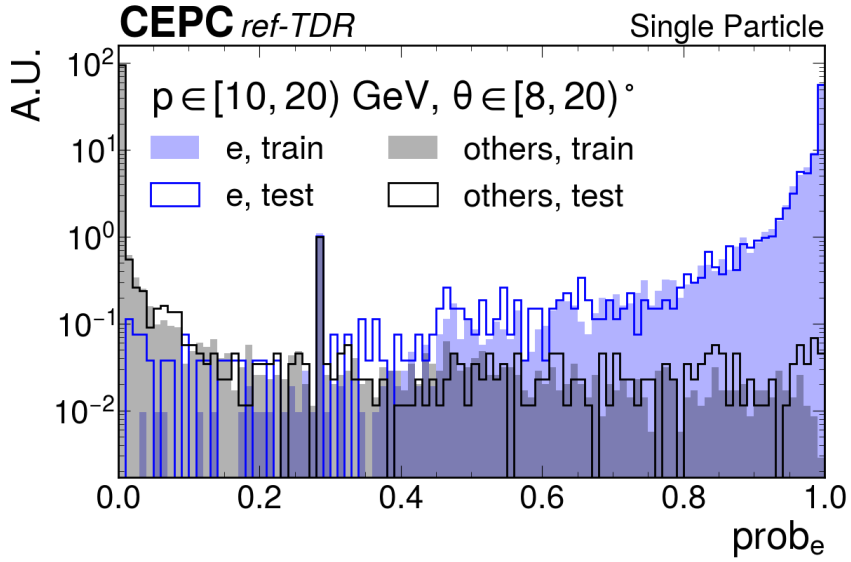


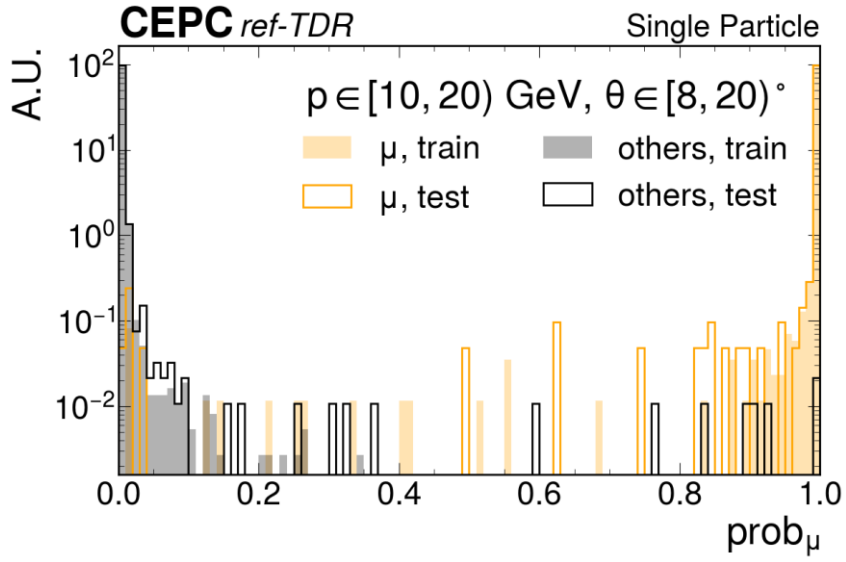


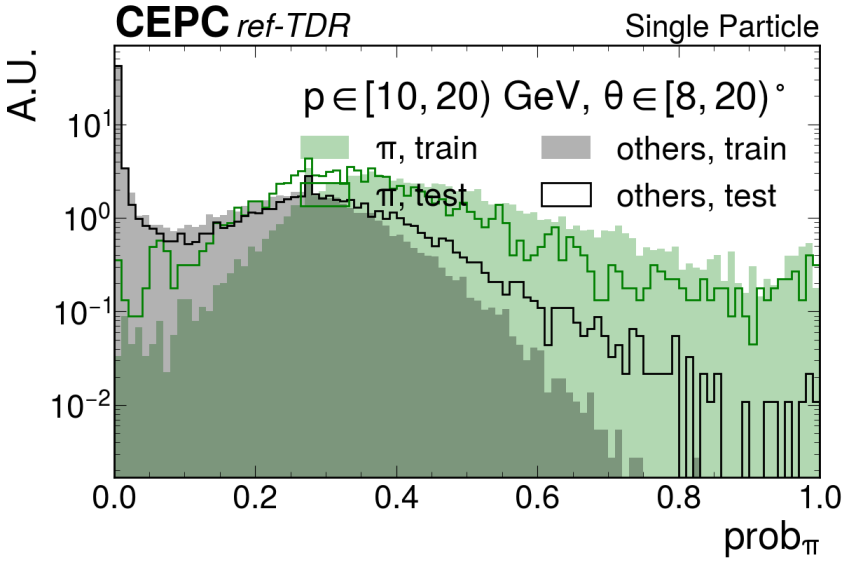


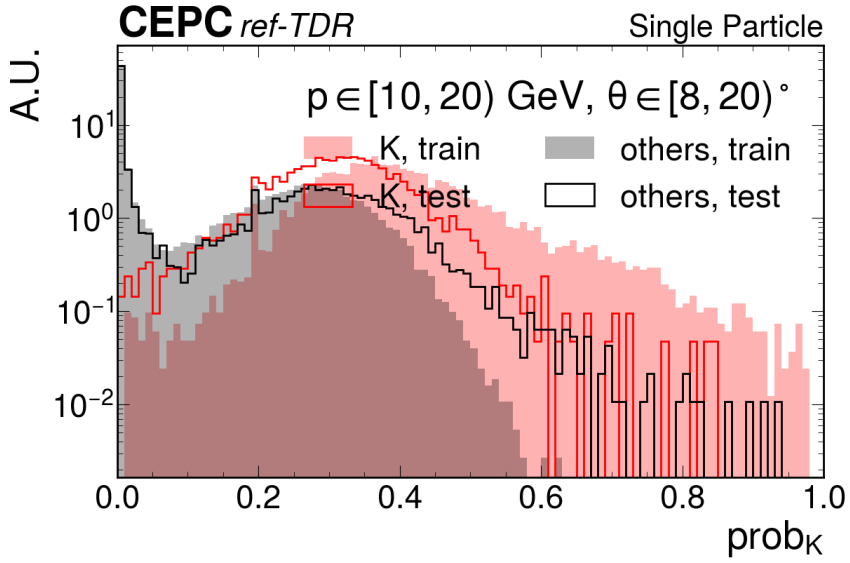


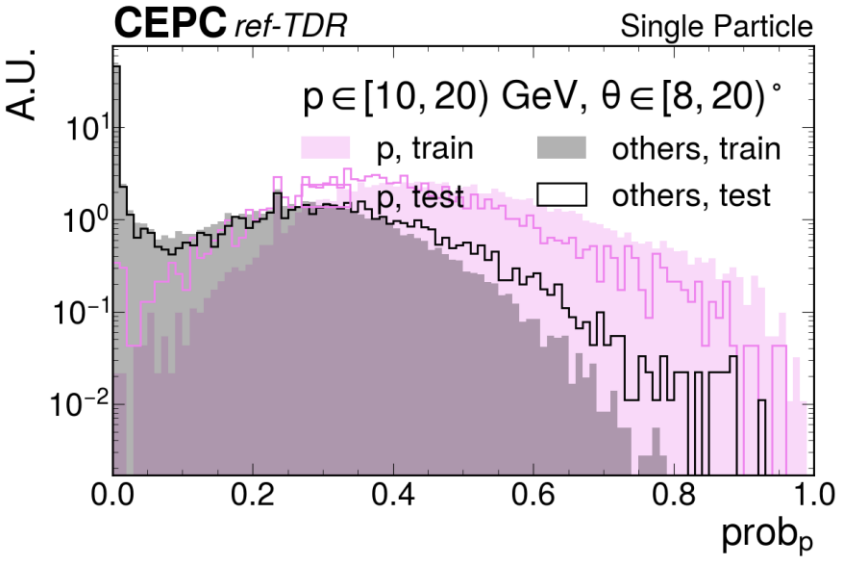
Train v.s. test



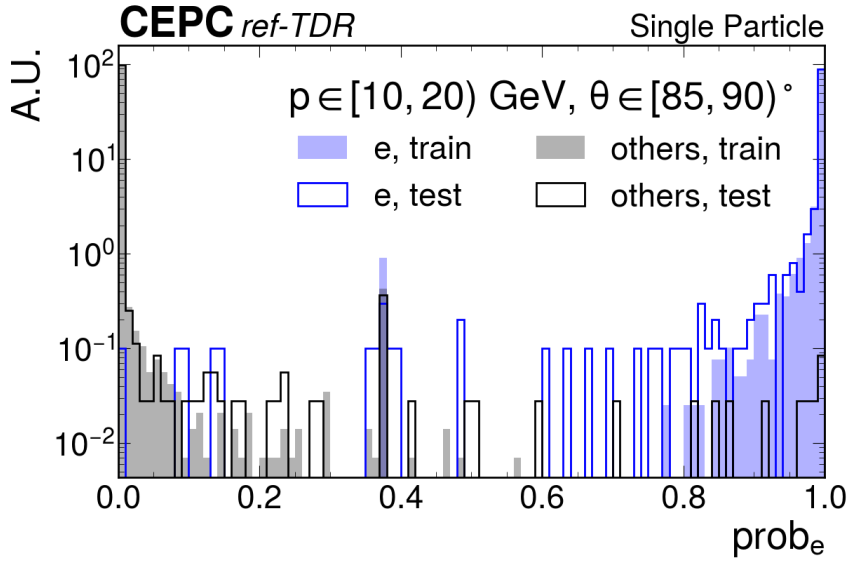


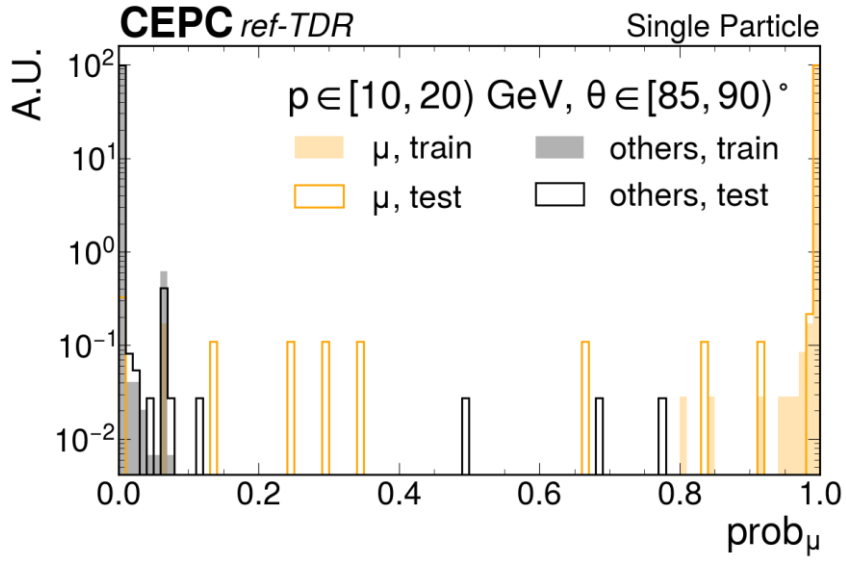


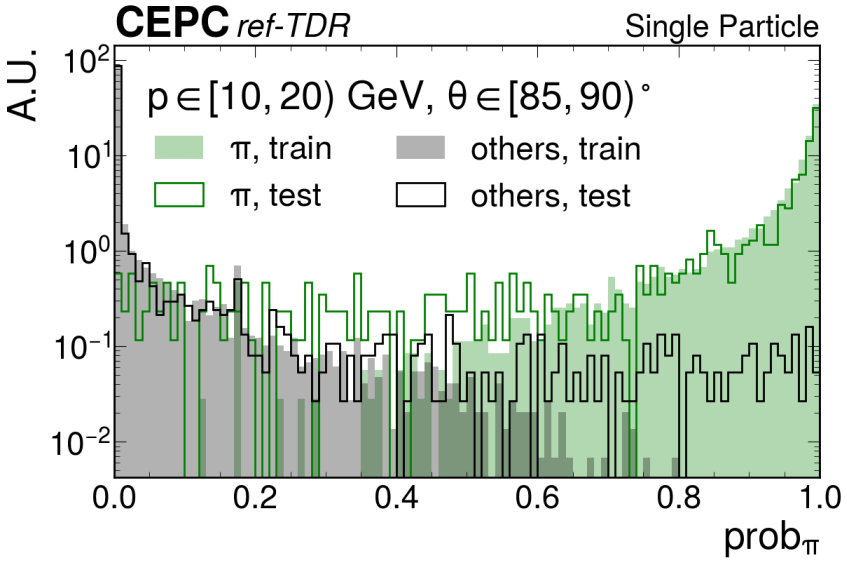


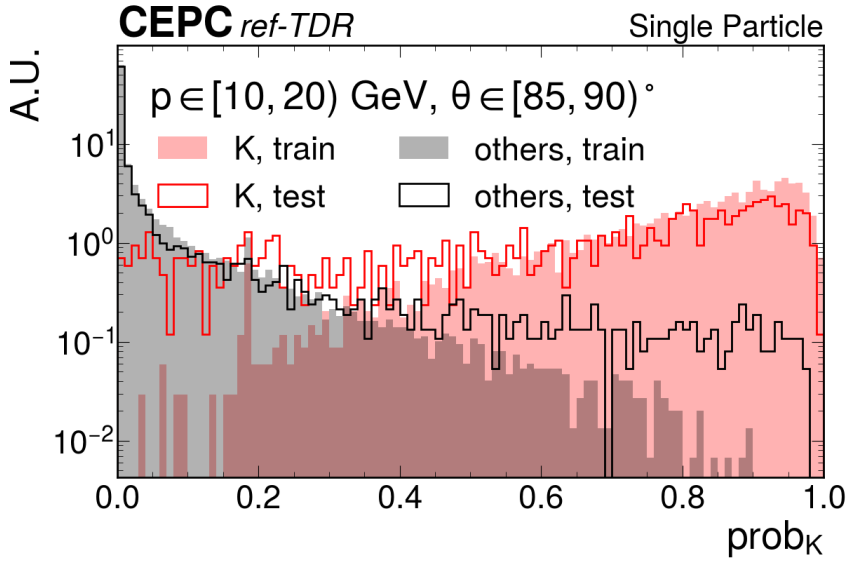


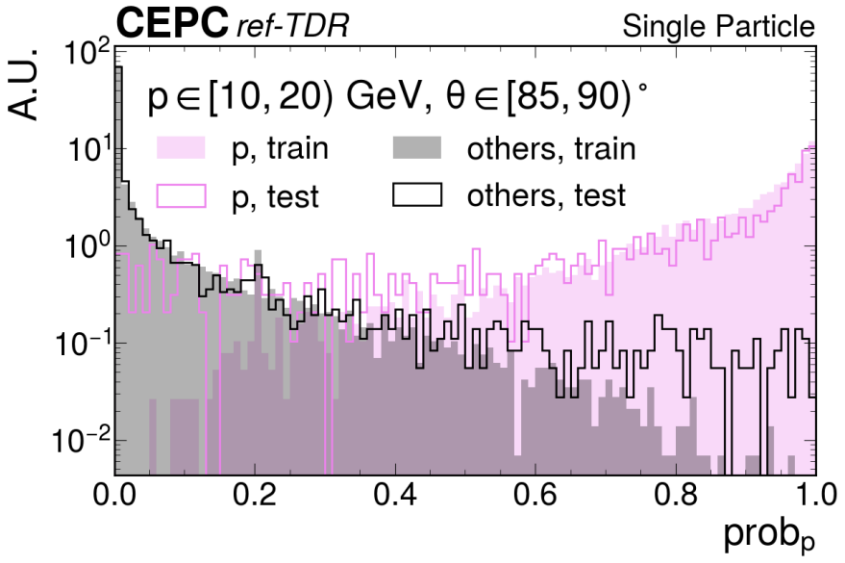
Train v.s. test





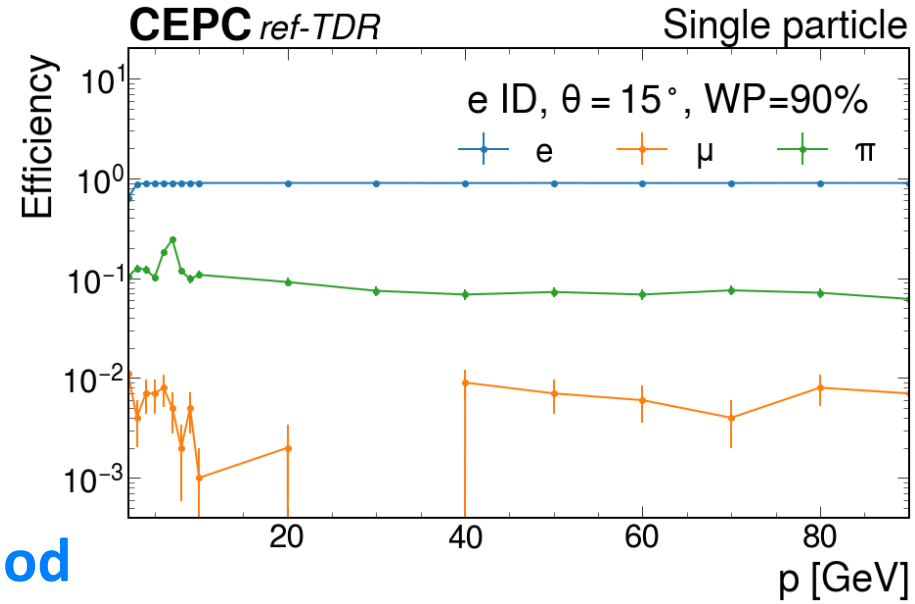


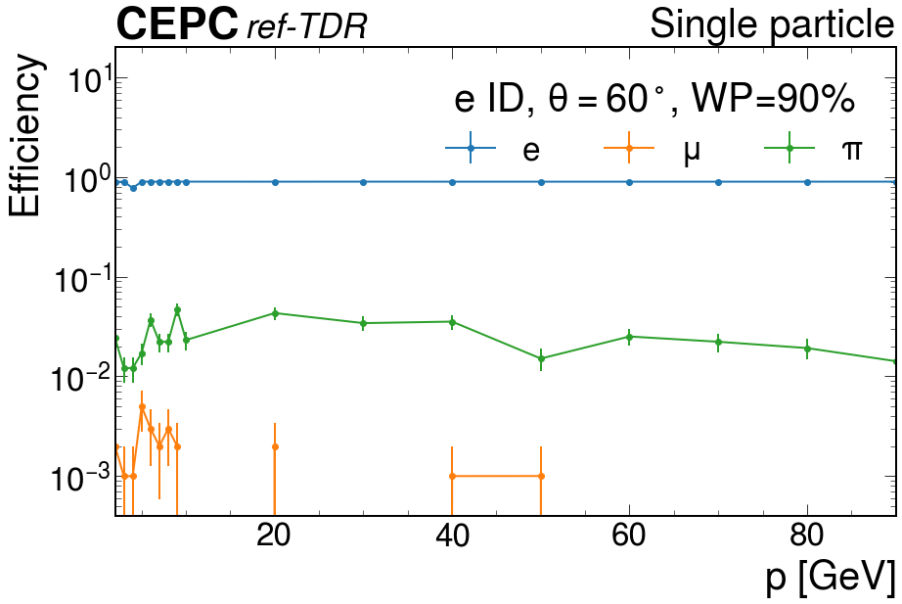




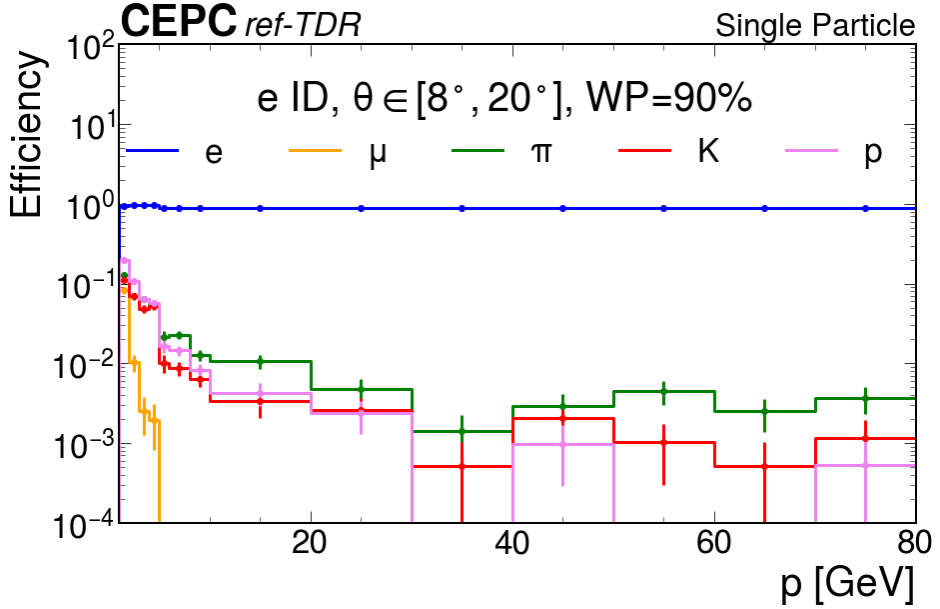
Efficiency: electron ID 90% WP

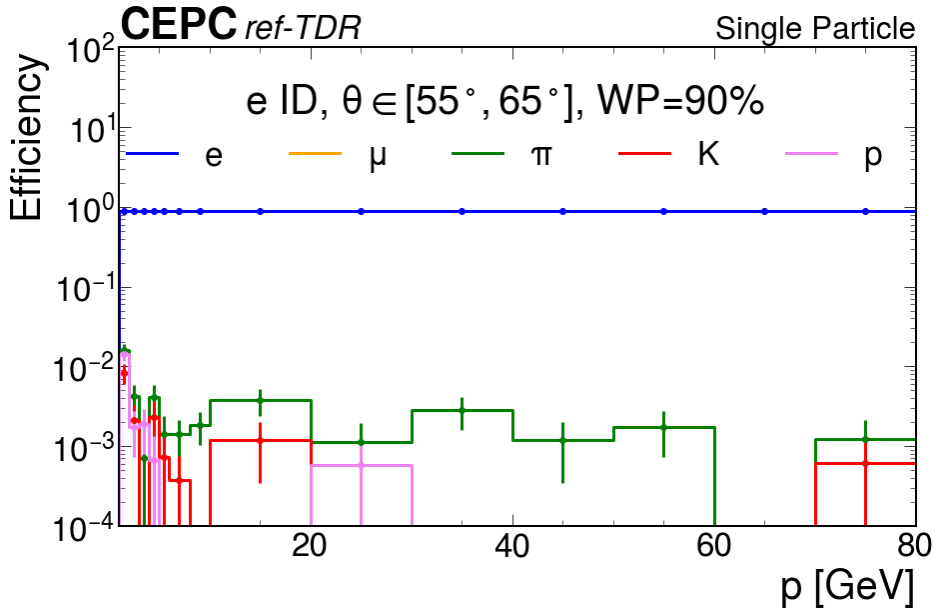
χ2 method





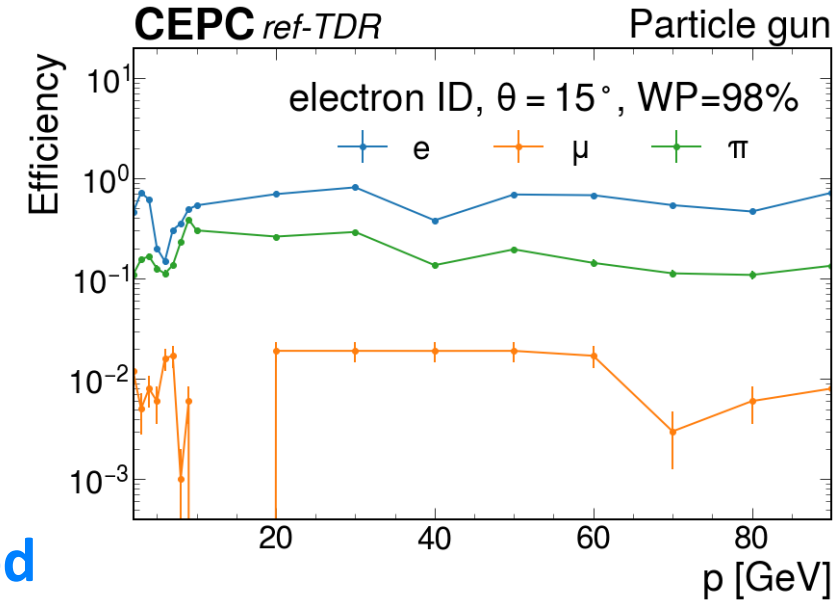
XGBoost

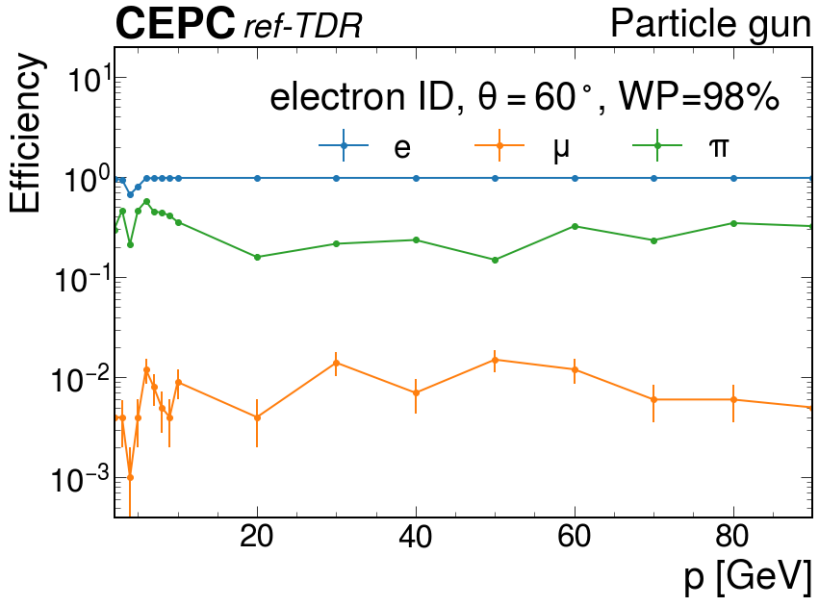




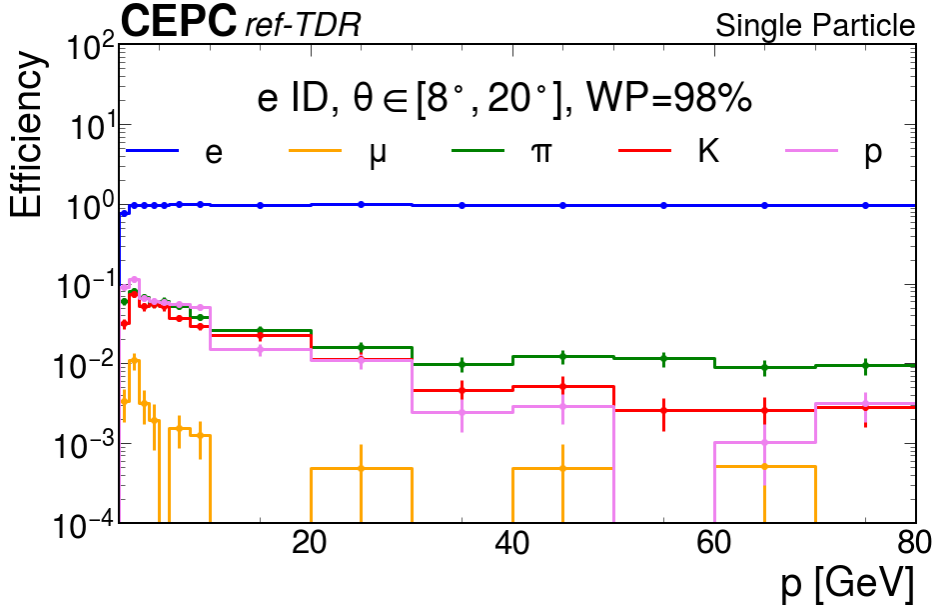
Efficiency: electron ID 98% WP

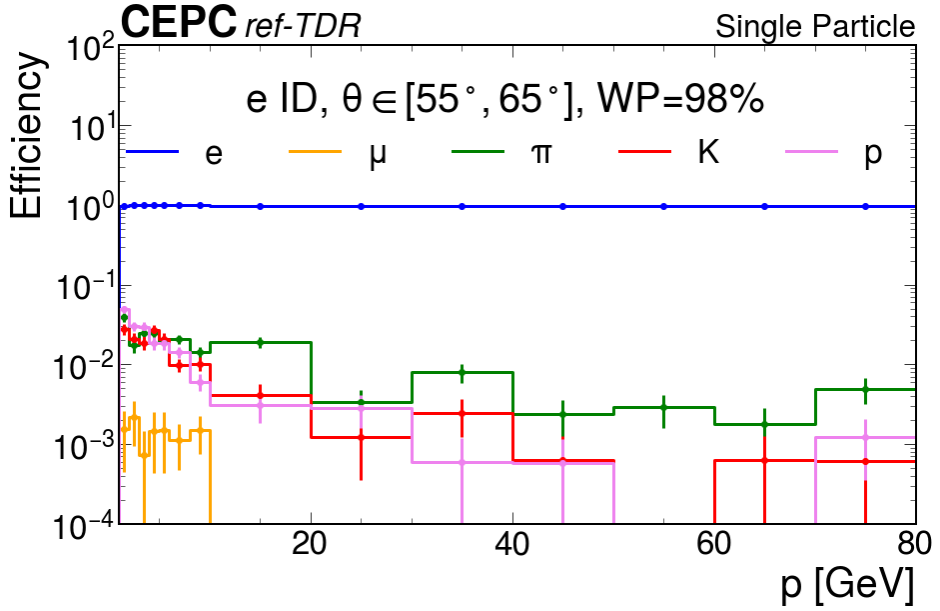
χ2 method





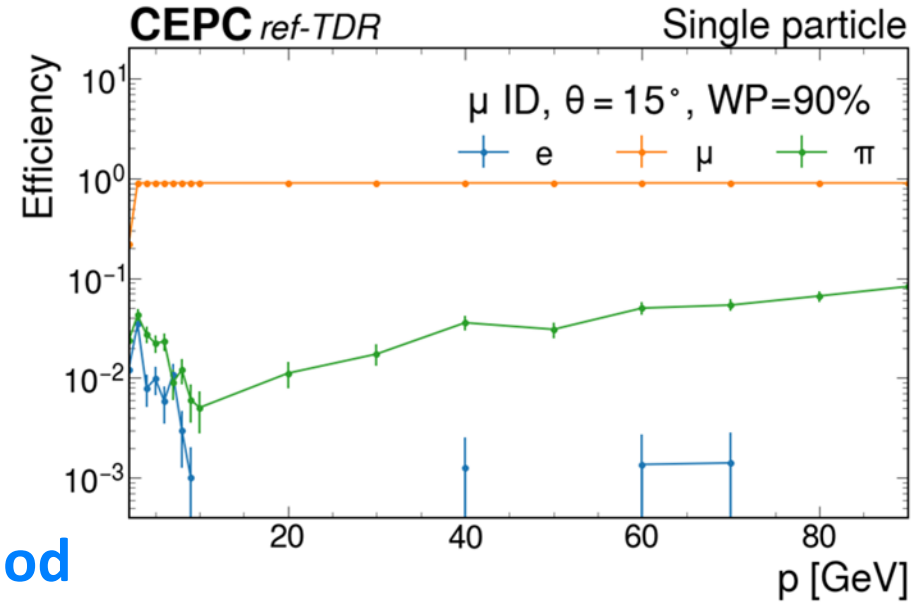
XGBoost

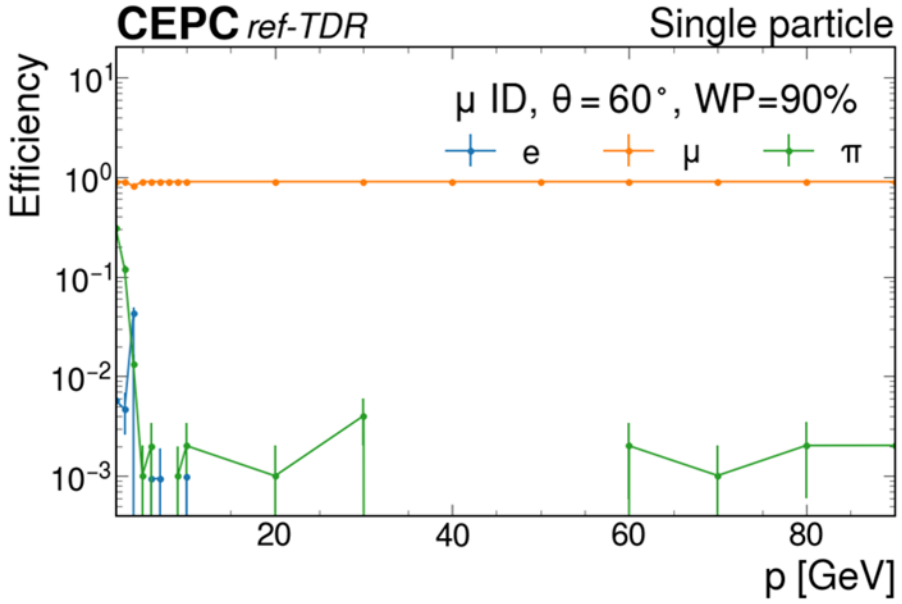




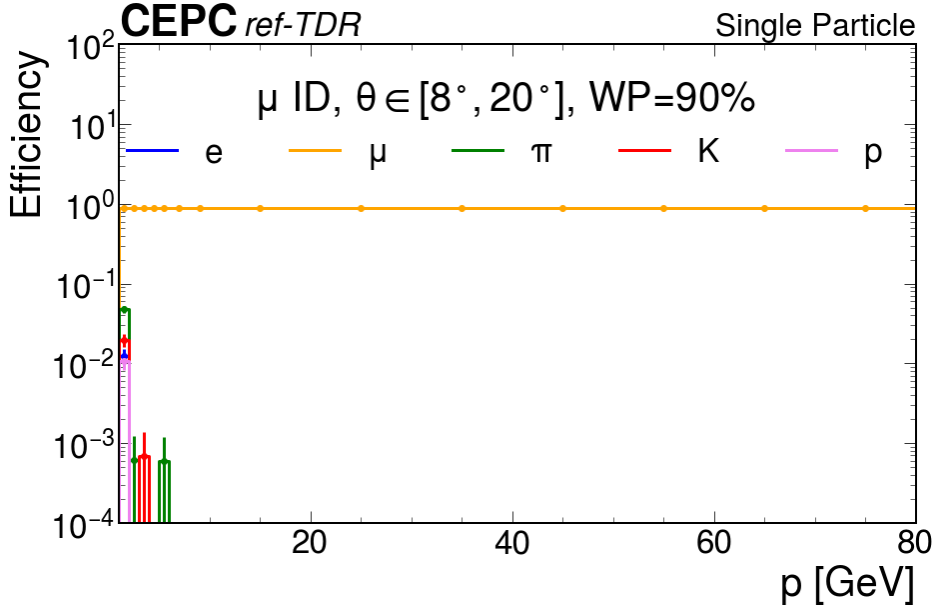
Efficiency: muon ID 90%

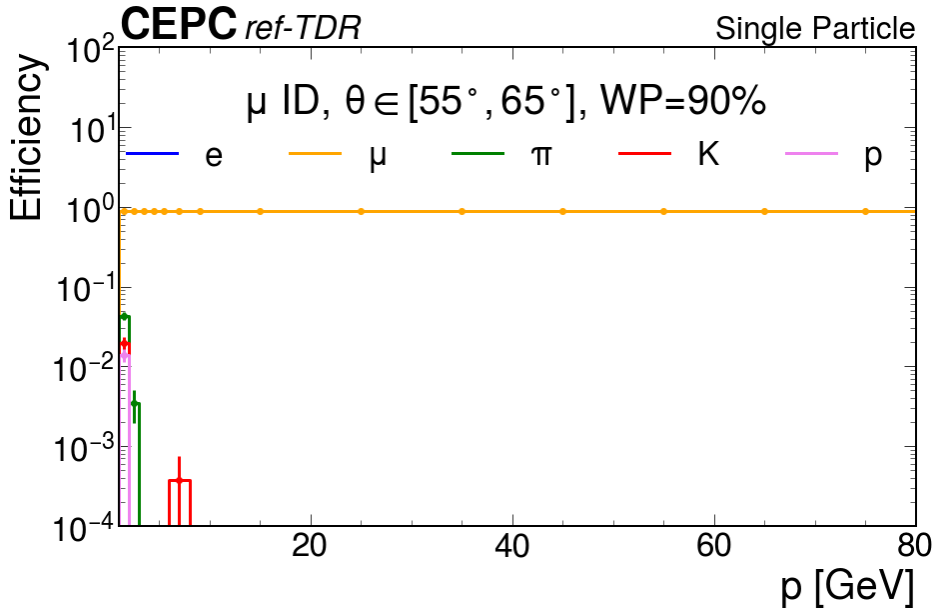
χ2 method





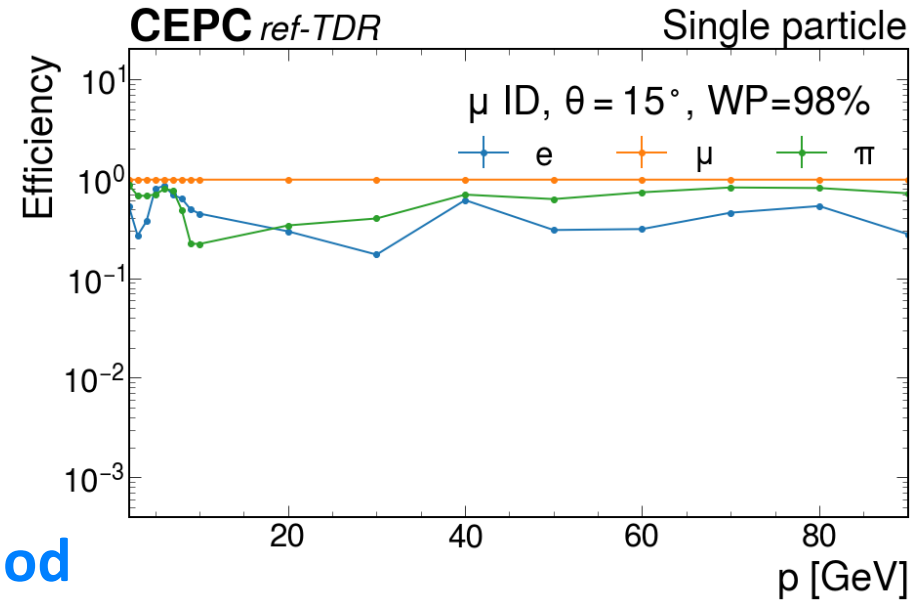
XGBoost

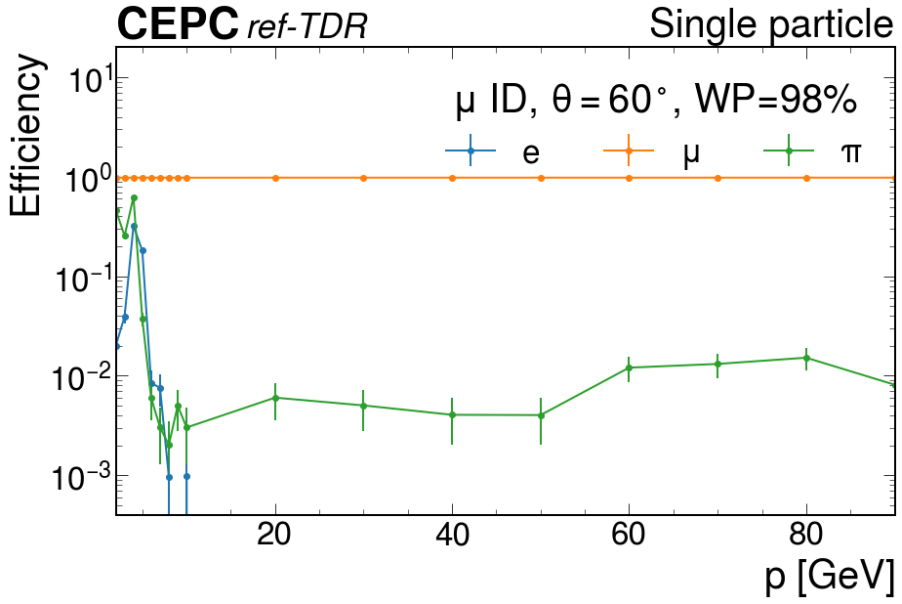




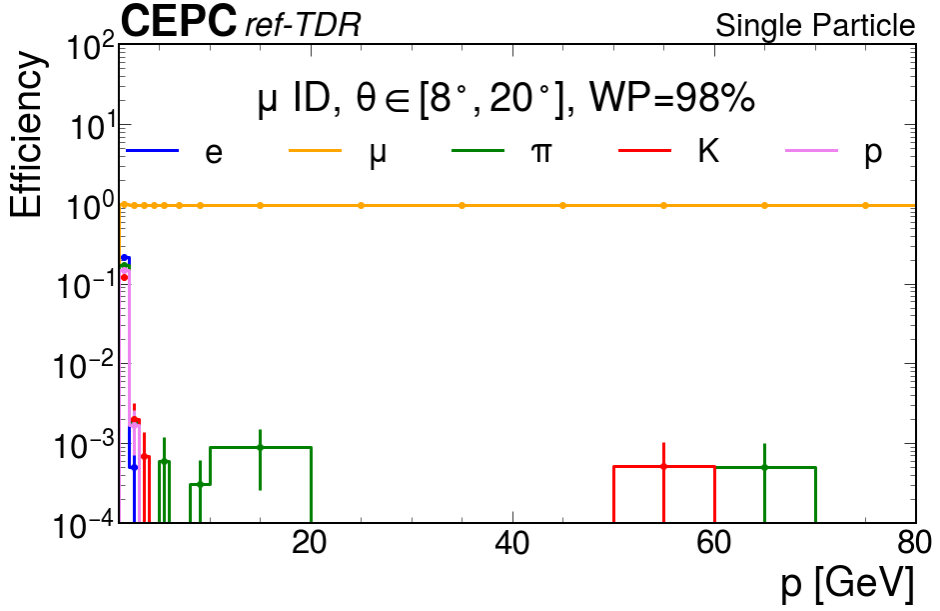
Efficiency: muon ID 98%

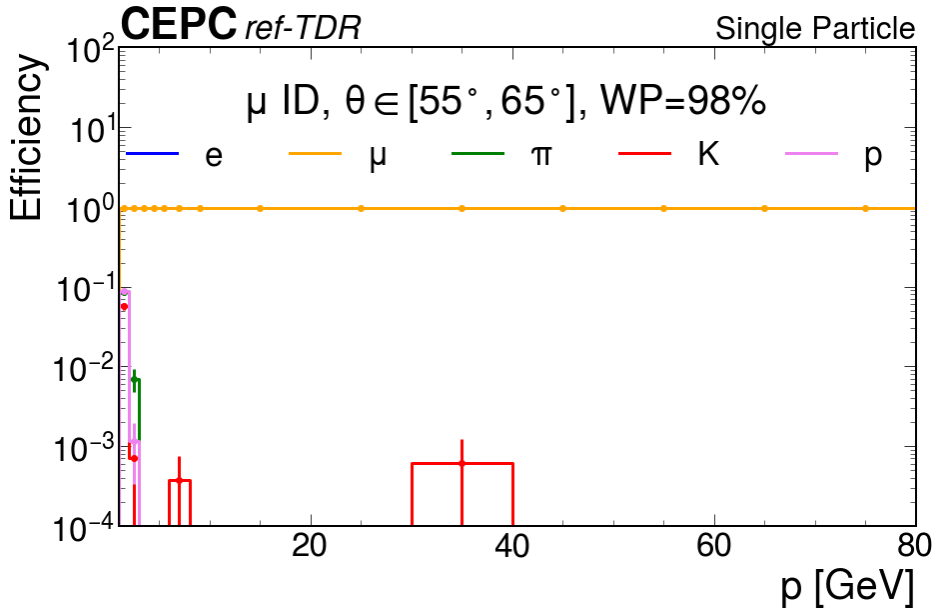
χ2 method





XGBoost

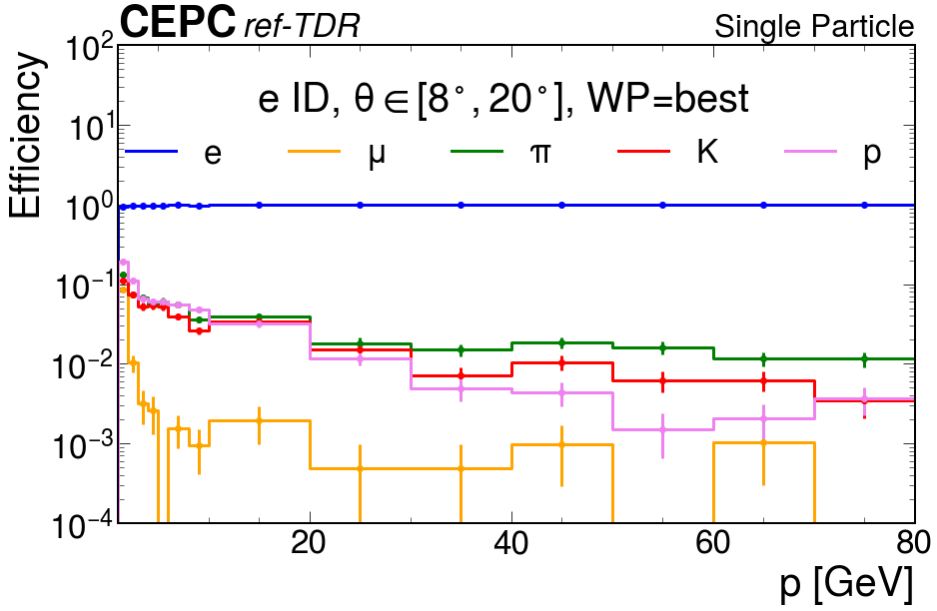


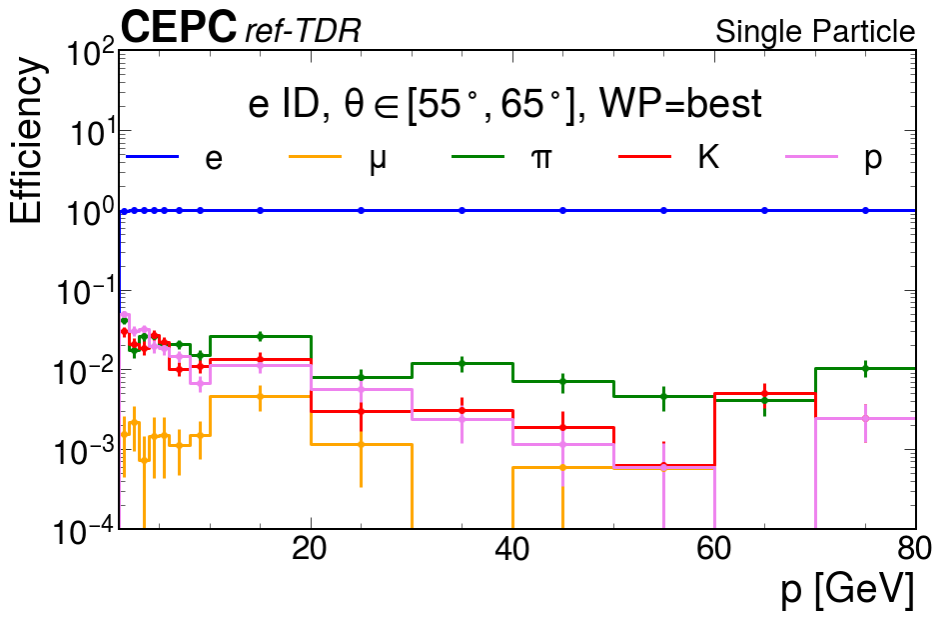


Efficiency: “best” WP

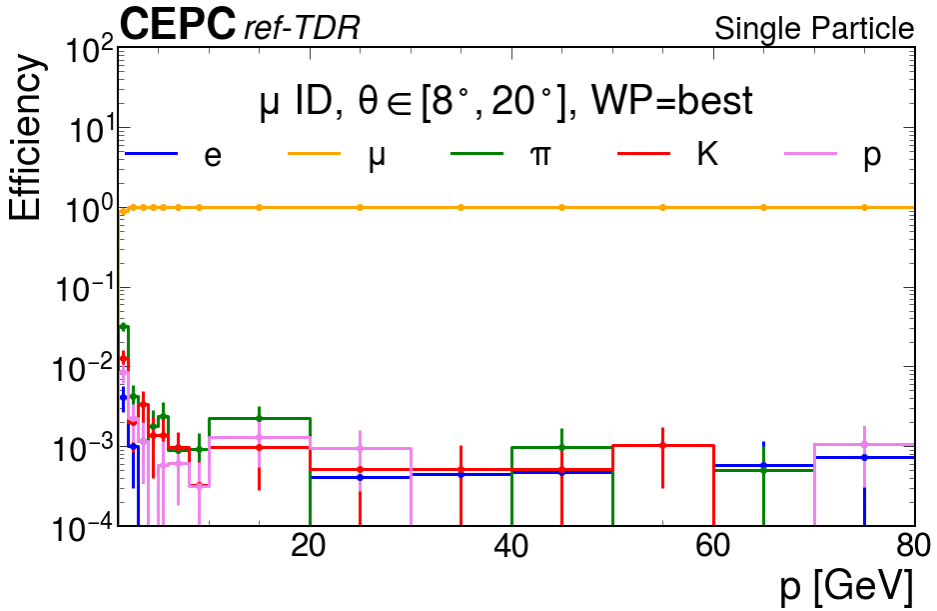
Decision not made on cuts on prob, but compare 5 probs to get the highest one.

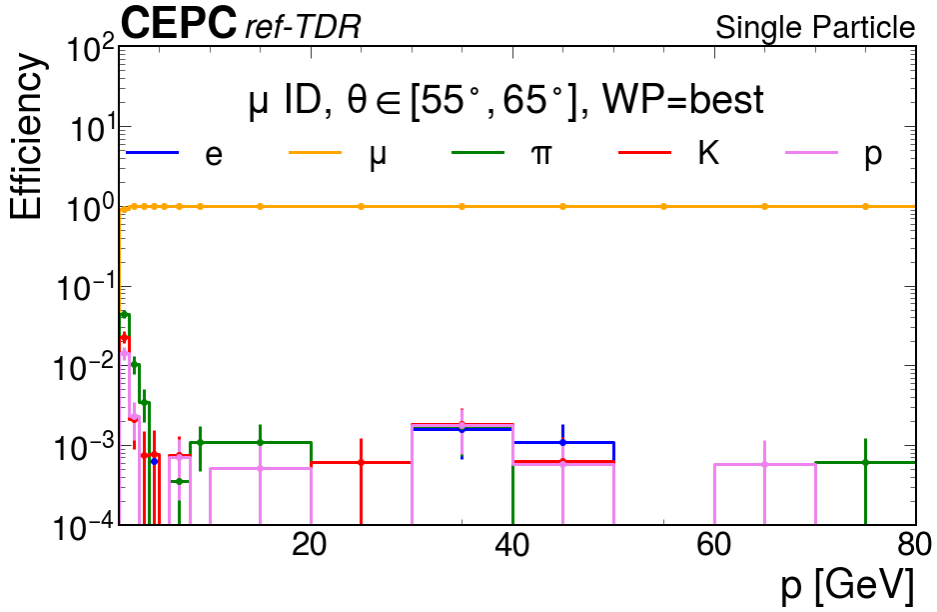
electron ID





muon ID

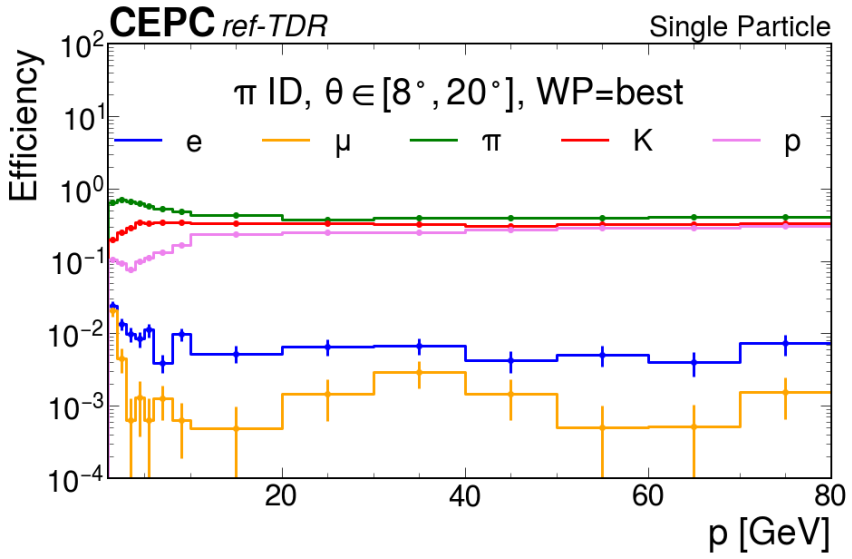


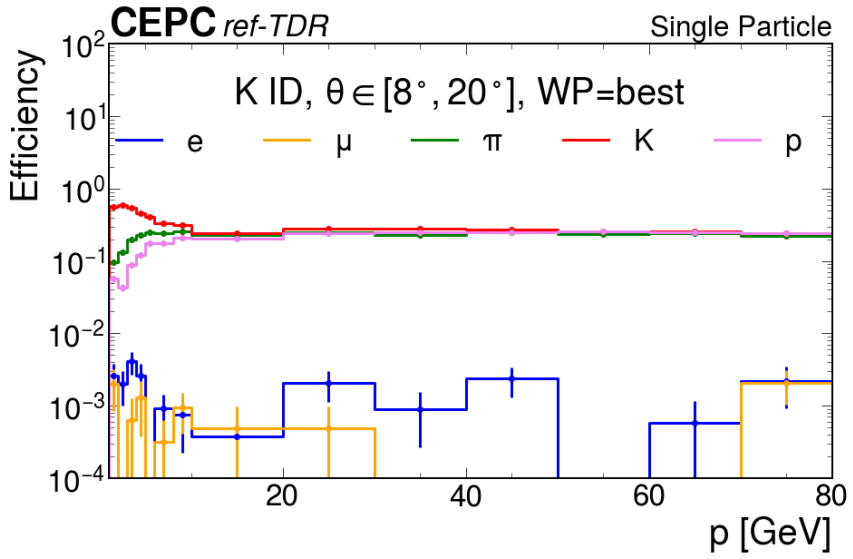


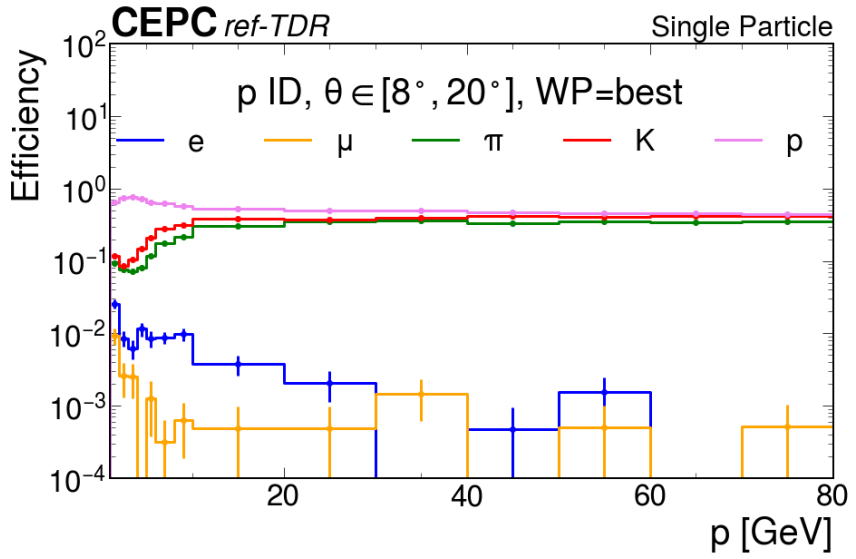
Reasons of improvements

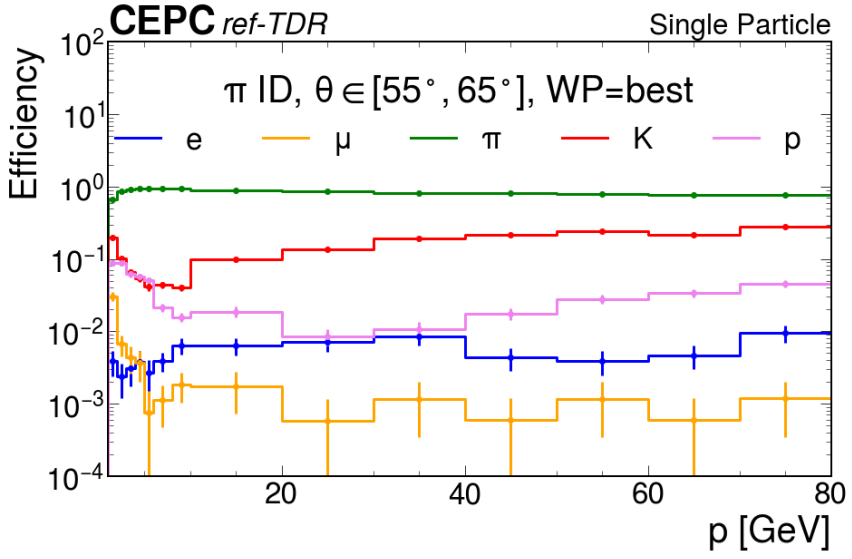
- More variables used
- Separate particle gun samples
- Better considerations of the shape of each variable than a simple χ^2
- Better consideration of correlations among different variables

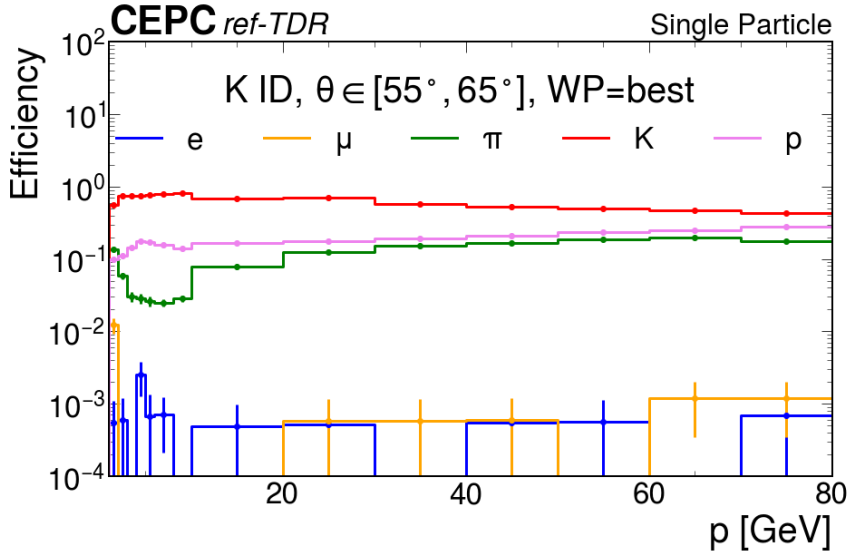
Efficiency: charged hadron ID

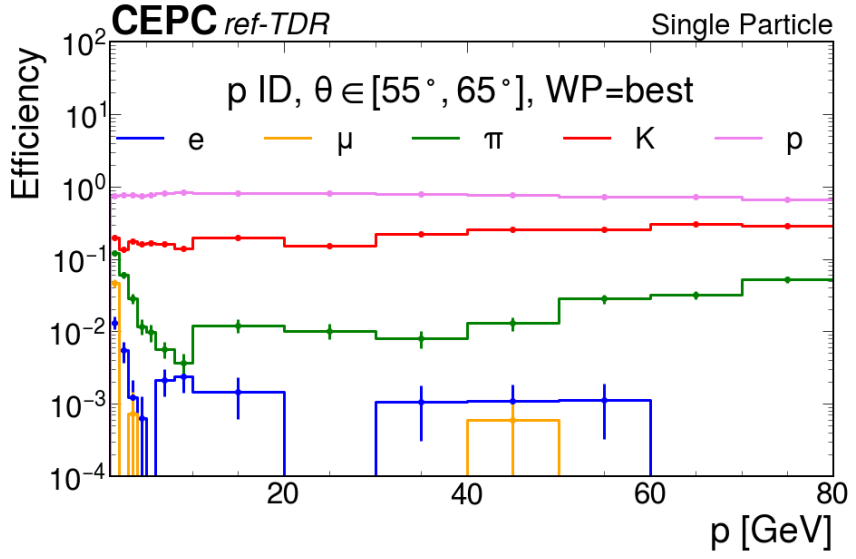






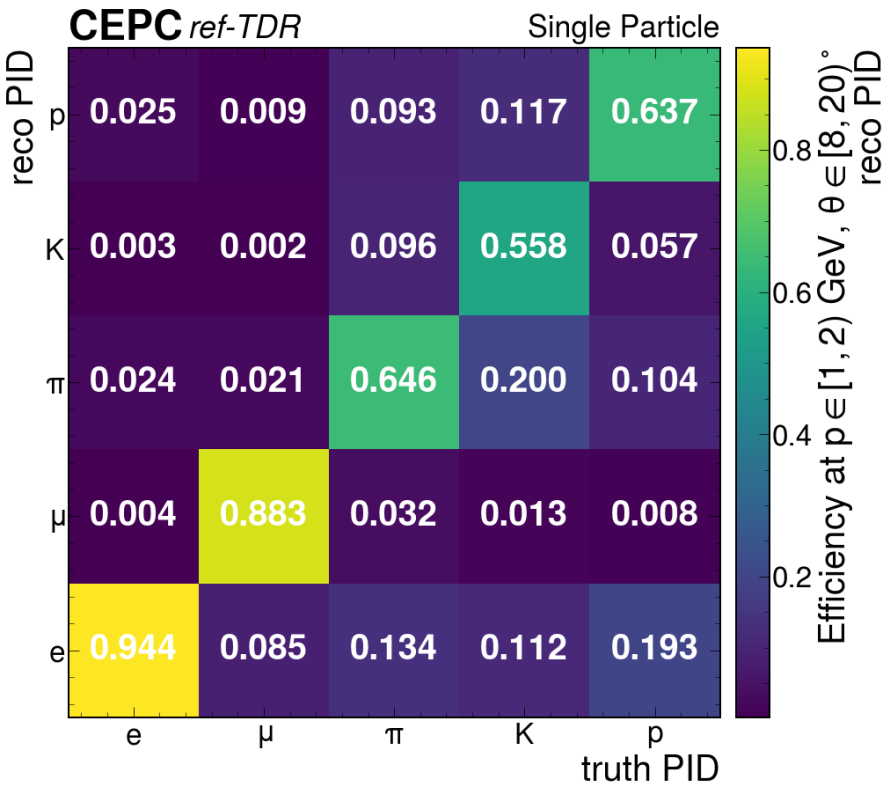




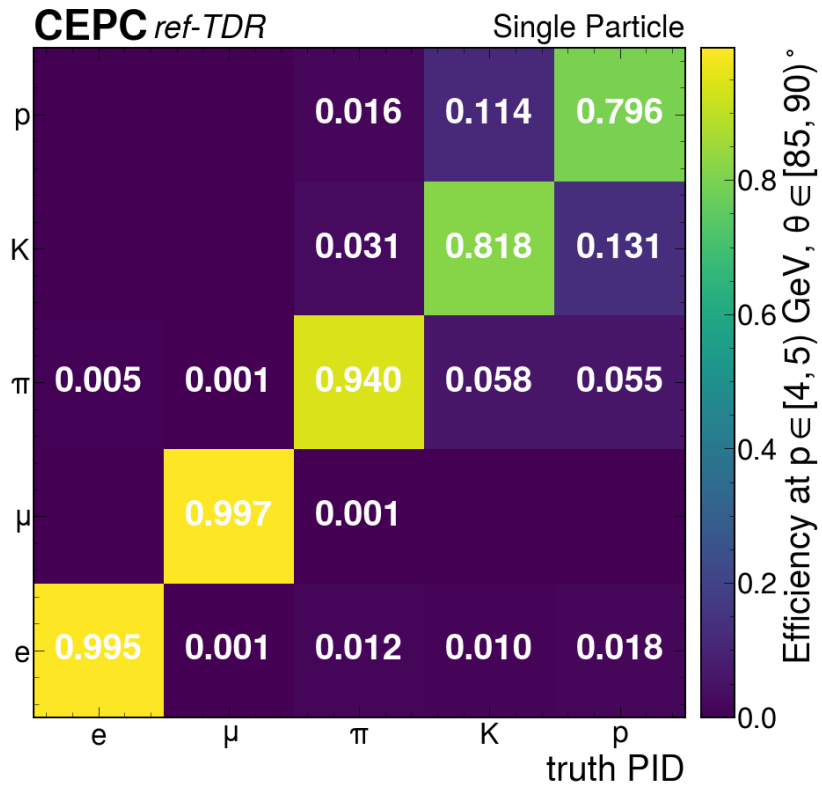


Confusion matrix

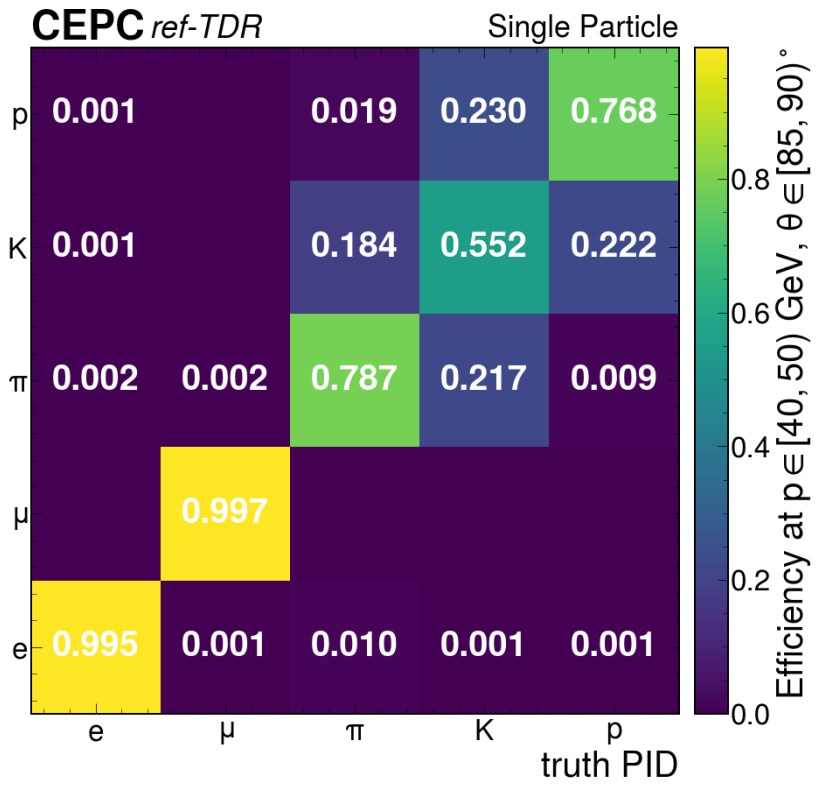
$\mathbf{p \in [1,2)GeV}$
$\mathbf{\theta \in [8,20)^\circ}$



$\mathbf{p \in [4,5)GeV}$
$\mathbf{\theta \in [85,90)^\circ}$



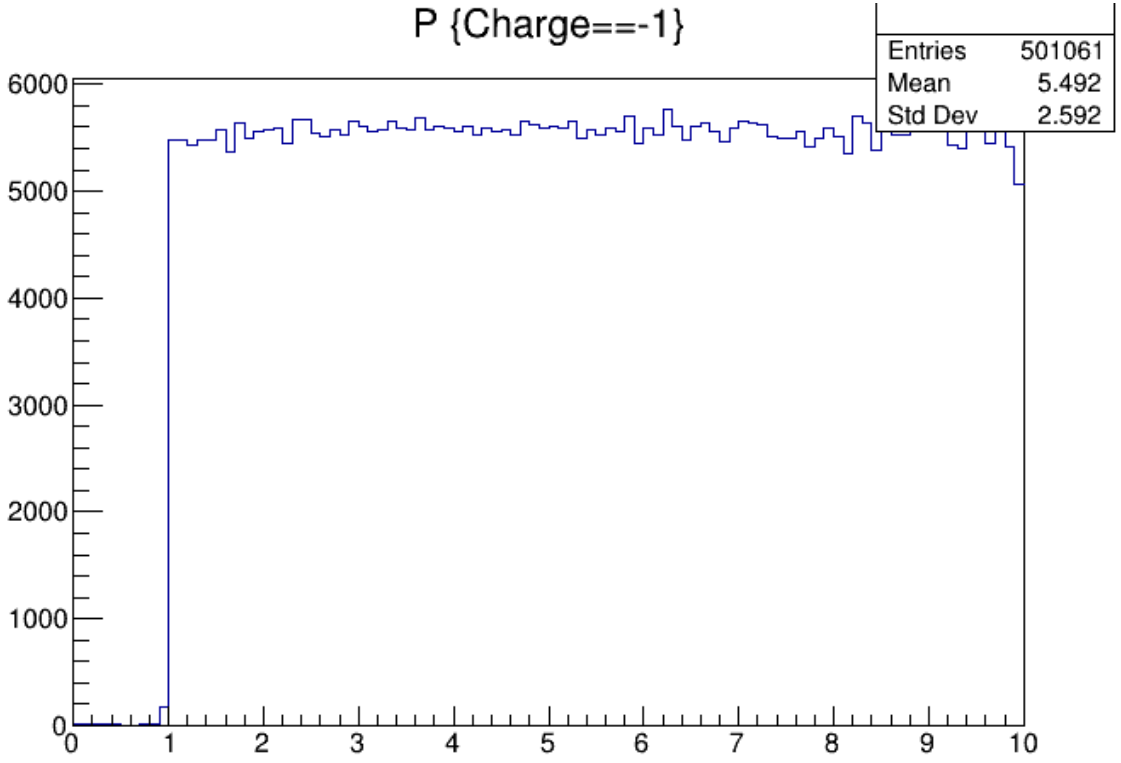
$\mathbf{p \in [40,50)GeV}$
$\mathbf{\theta \in [85,90)^\circ}$

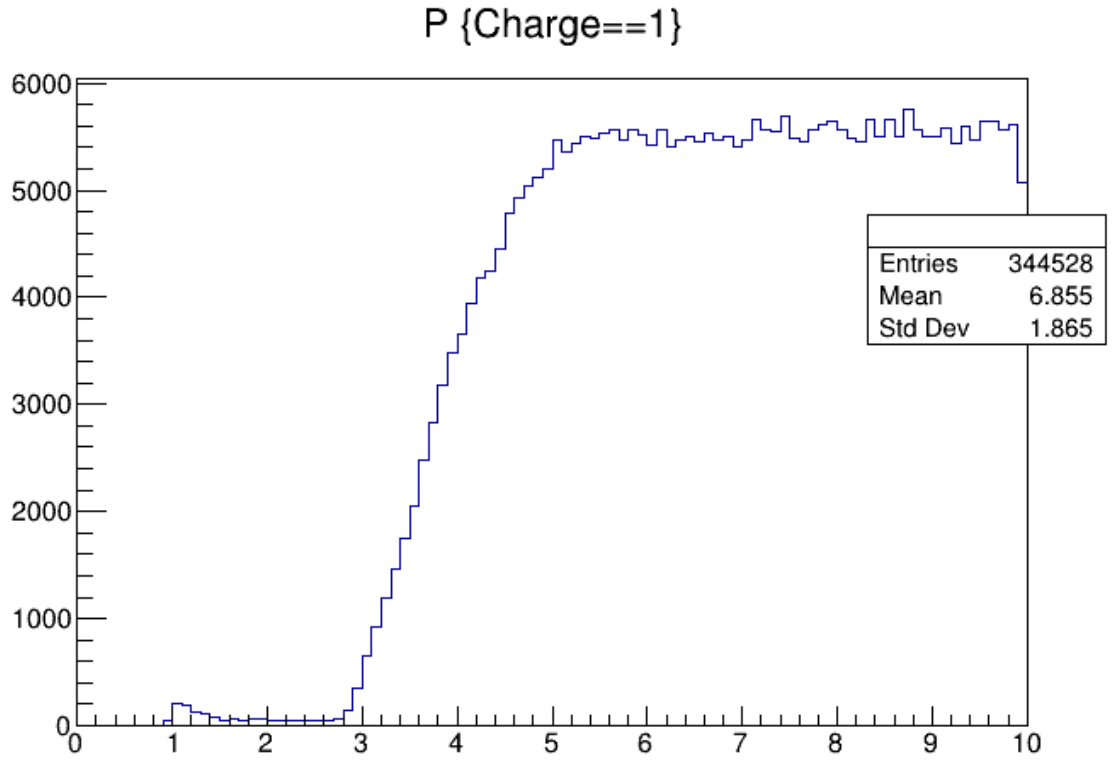


Do we need to consider different charges?

Something weird with muons

- Reconstruction efficiency drops at p<5 GeV

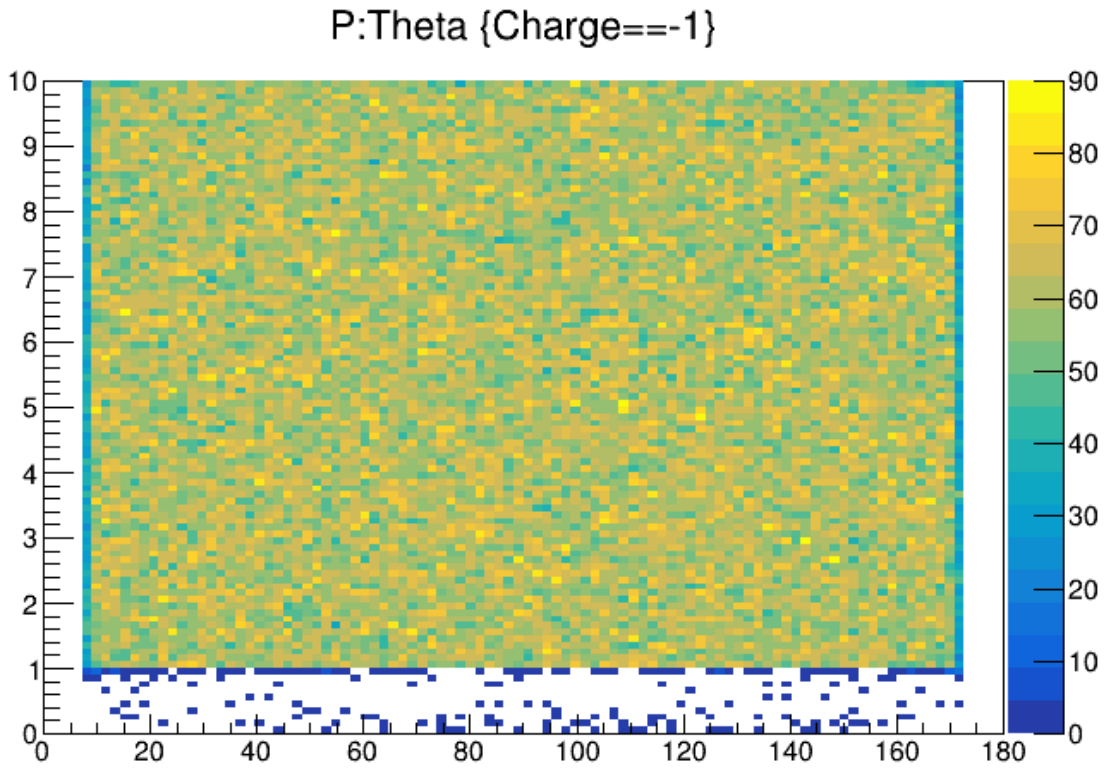


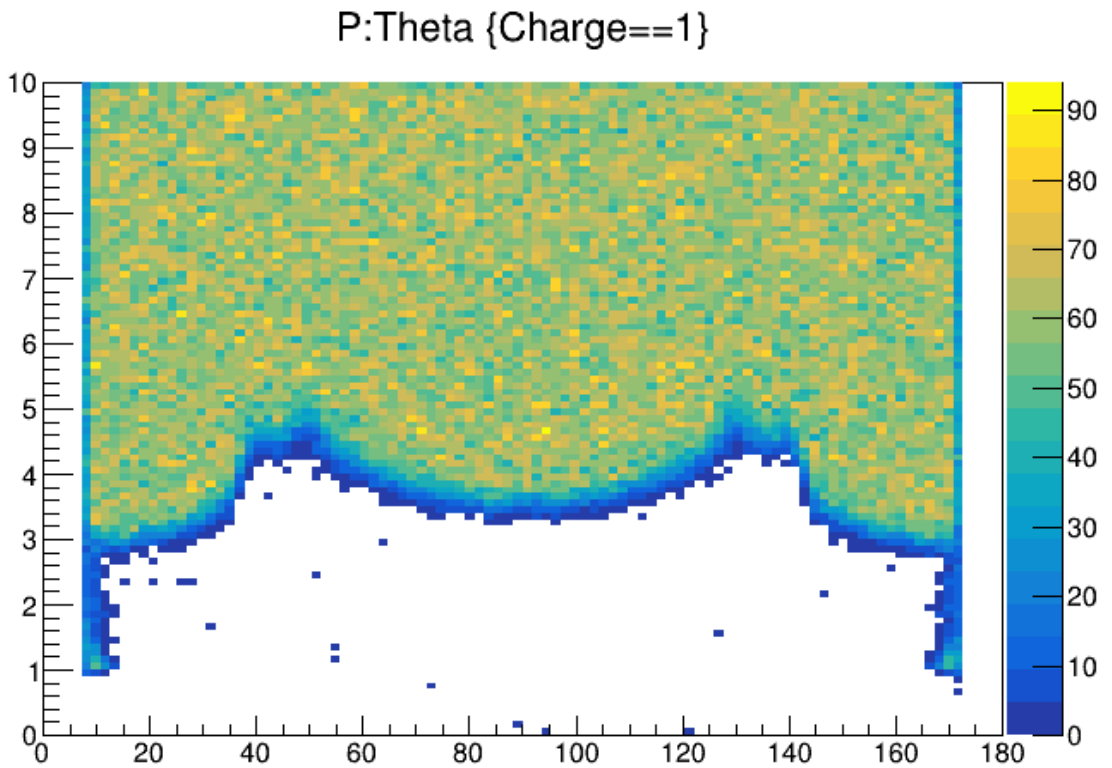


Do we need to consider different charges?

Something weird with muons

- Reconstruction efficiency drops at p<5 GeV
- Mostly around barrel regions

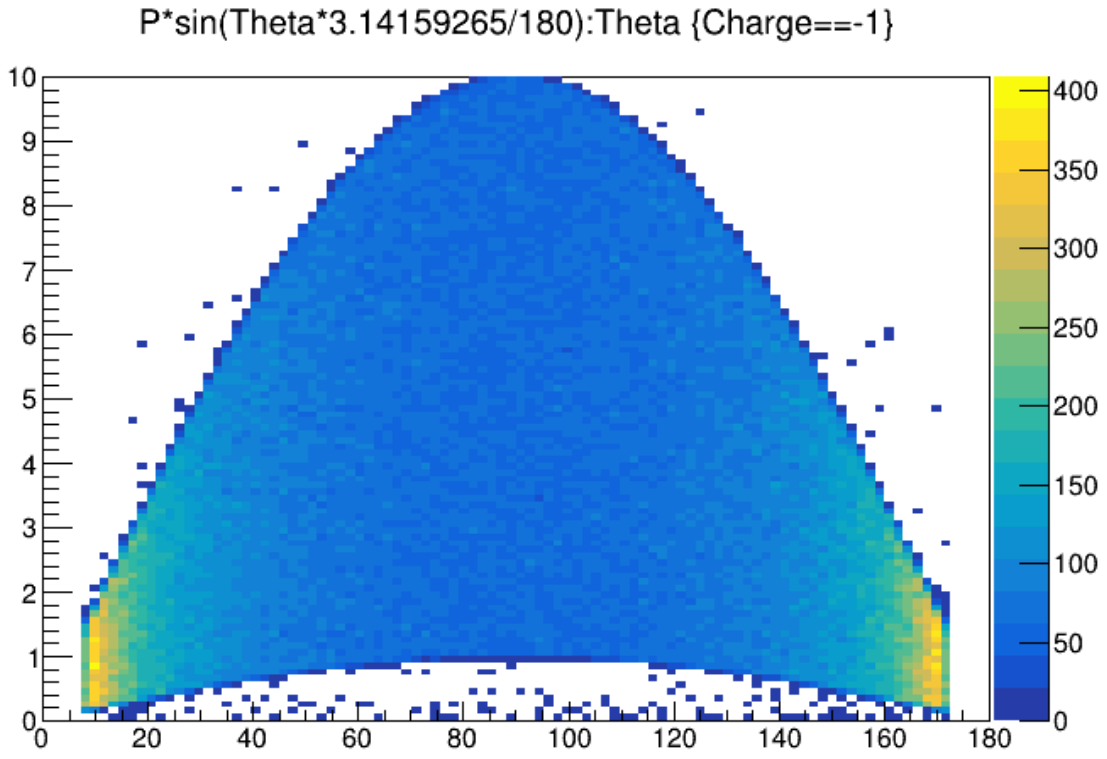


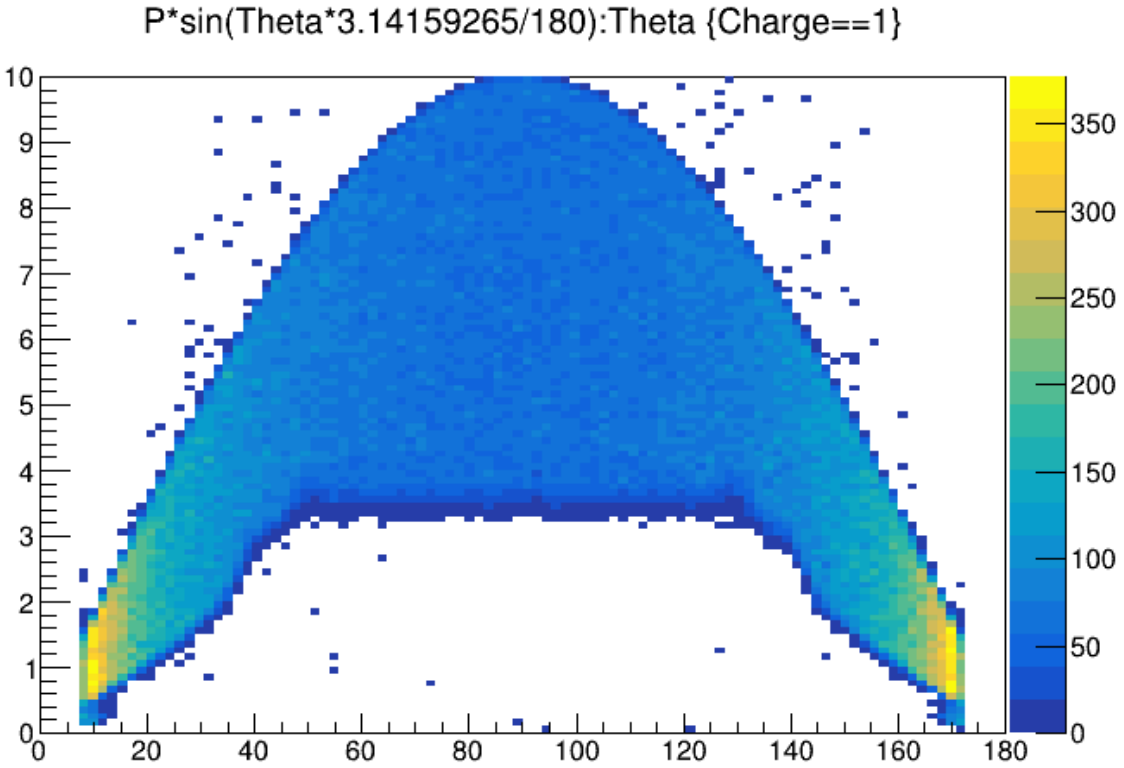


Do we need to consider different charges?

Something weird with muons

- Reconstruction efficiency drops at p<5 GeV
- Mostly around barrel regions



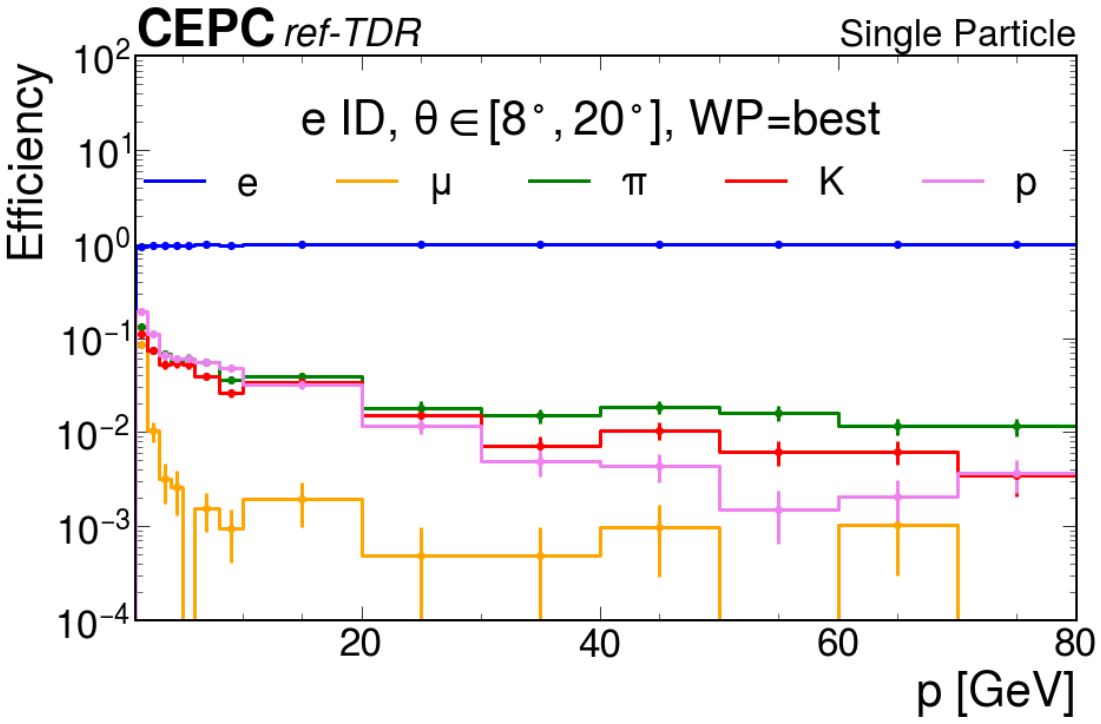


Do we need to consider different charges?

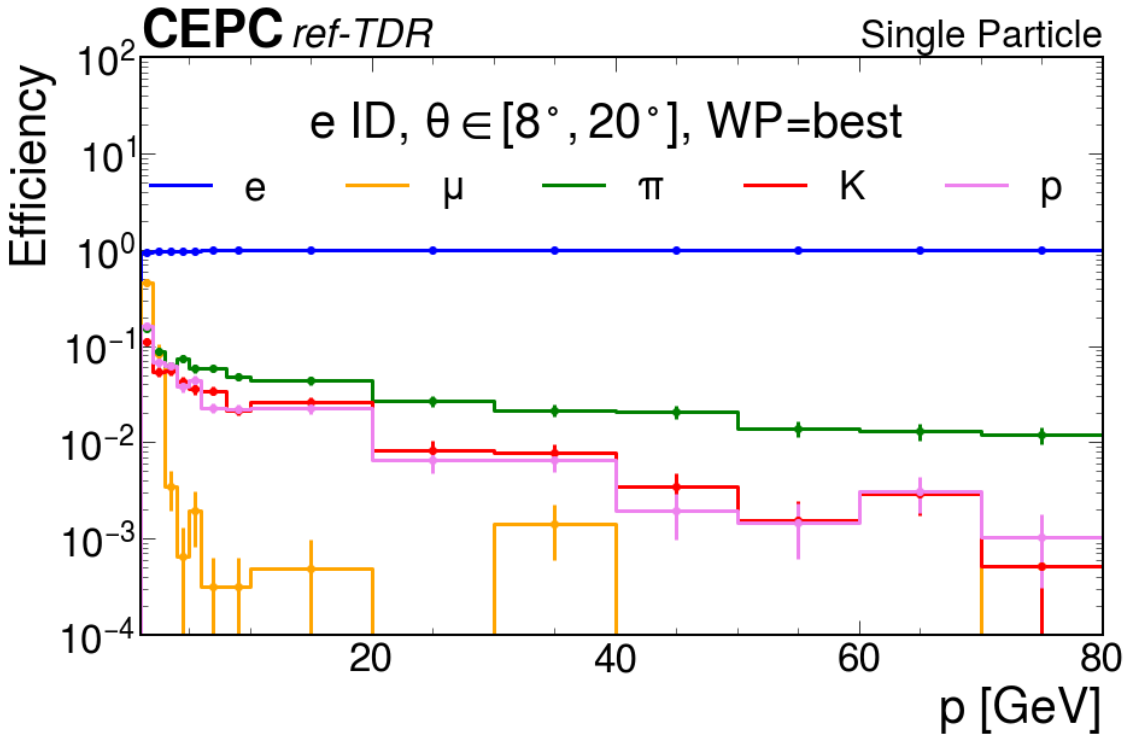
Results shown so far are with negative charges

- Results with positive charges are also produced.

Negative



Positive

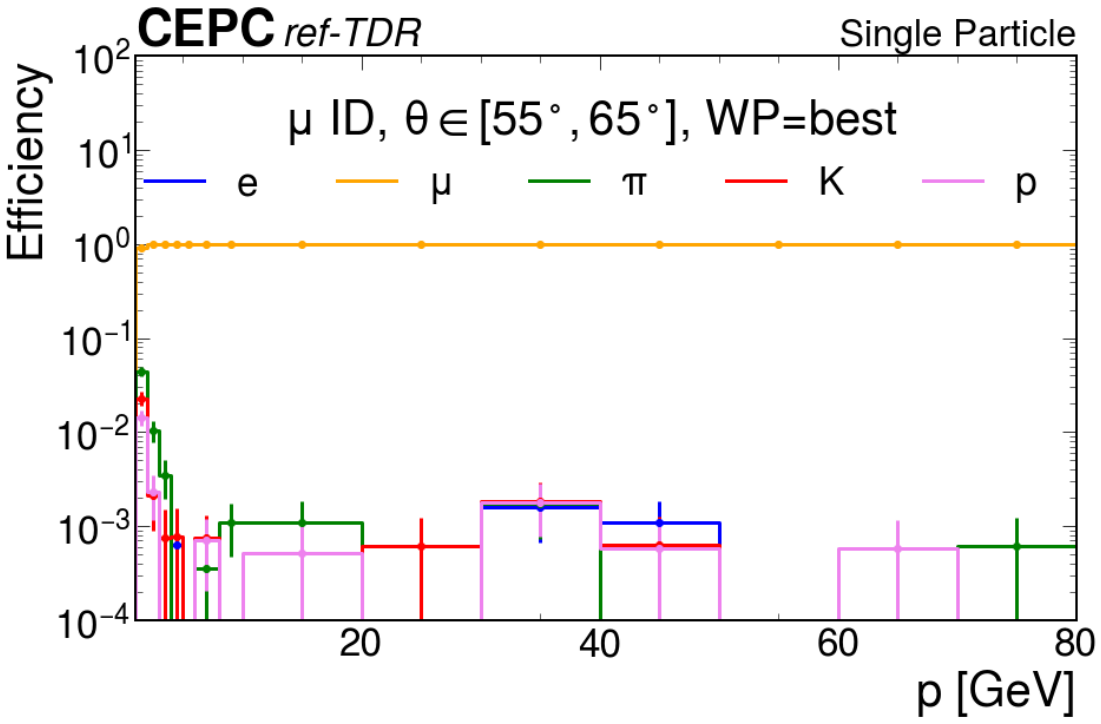


Do we need to consider different charges?

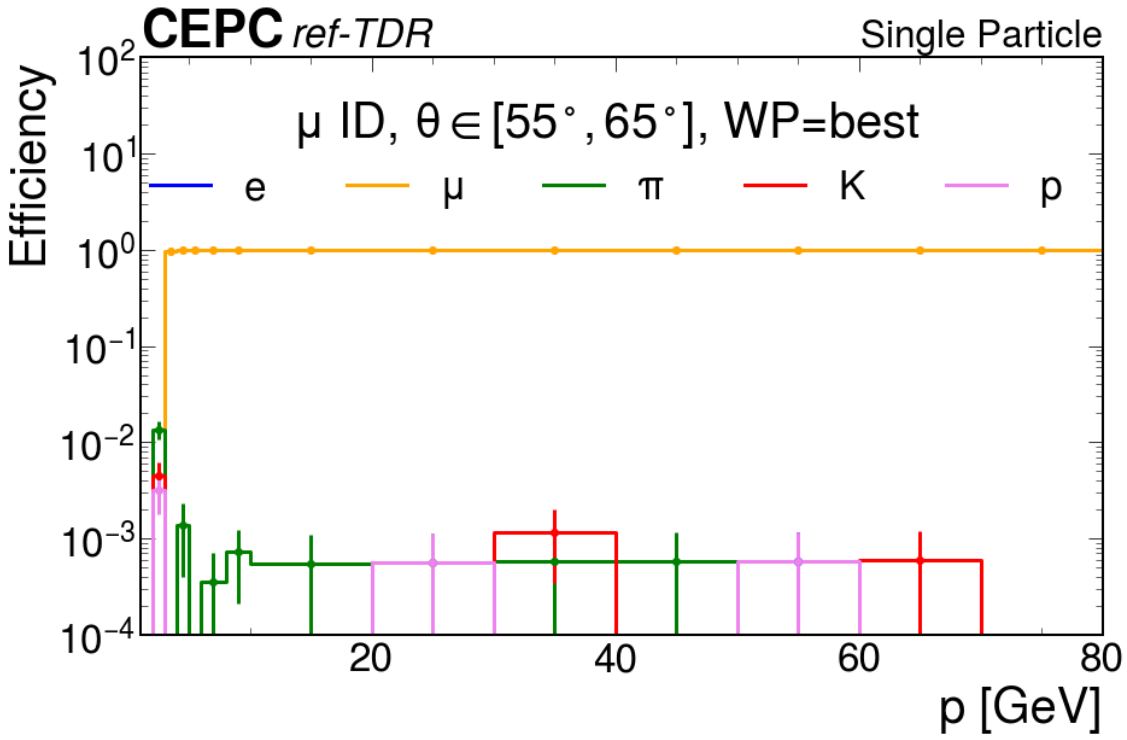
Results shown so far are with negative charges

- Results with positive charges are also produced.

Negative



Positive



What to do next

- **Test the new PID under ZH inclusive samples**
  - Should be done by today
- **Do we want this BDT method for charged hadron ID?**
  - Probably need to go below 1 GeV
- **Implement to CEPCSW**
- **Write the contents in TDR**