Current Status of e-ID based on BDT Algorithm

Hai-Jun Yang
University of Michigan
(with X. Li, T. Dai, A. Wilson, B. Zhou)

BNL Analysis Jamboree
March 19, 2009

Goals

- ATLAS default electron-ID (IsEM) has relatively low efficiency (~70%) and high jet faked electron rate based on our diboson and Higgs studies. The goal of e-ID algorithm development based on Boosted Decision Trees (BDT) are:

➔To improve the e-ID selection efficiency

➔To reduce faked electron rate from dijet samples (previous works)

➔To reduce faked electron rate from γ jet samples (current works)

Electron ID Studies with BDT

Select electrons in two steps

1) Pre-selection: an EM cluster matching a track

2) Apply electron ID based on pre-selected samples with different e-ID algorithms (IsEM, Likelihood, AdaBoost and **EBoost**).

BDT e-ID development at U. Michigan (Rel. v12)

- H. Yang's talk at US-ATLAS Jamboree on Sept. 10, 2008

  http://indico.cern.ch/conferenceDisplay.py?confId=38991

BDT e-ID (EBoost) based on Rel. v13

- H. Yang's talk at ATLAS performance and physics workshop at CERN on Oct. 2, 2008

  http://indico.cern.ch/conferenceDisplay.py?confId=39296

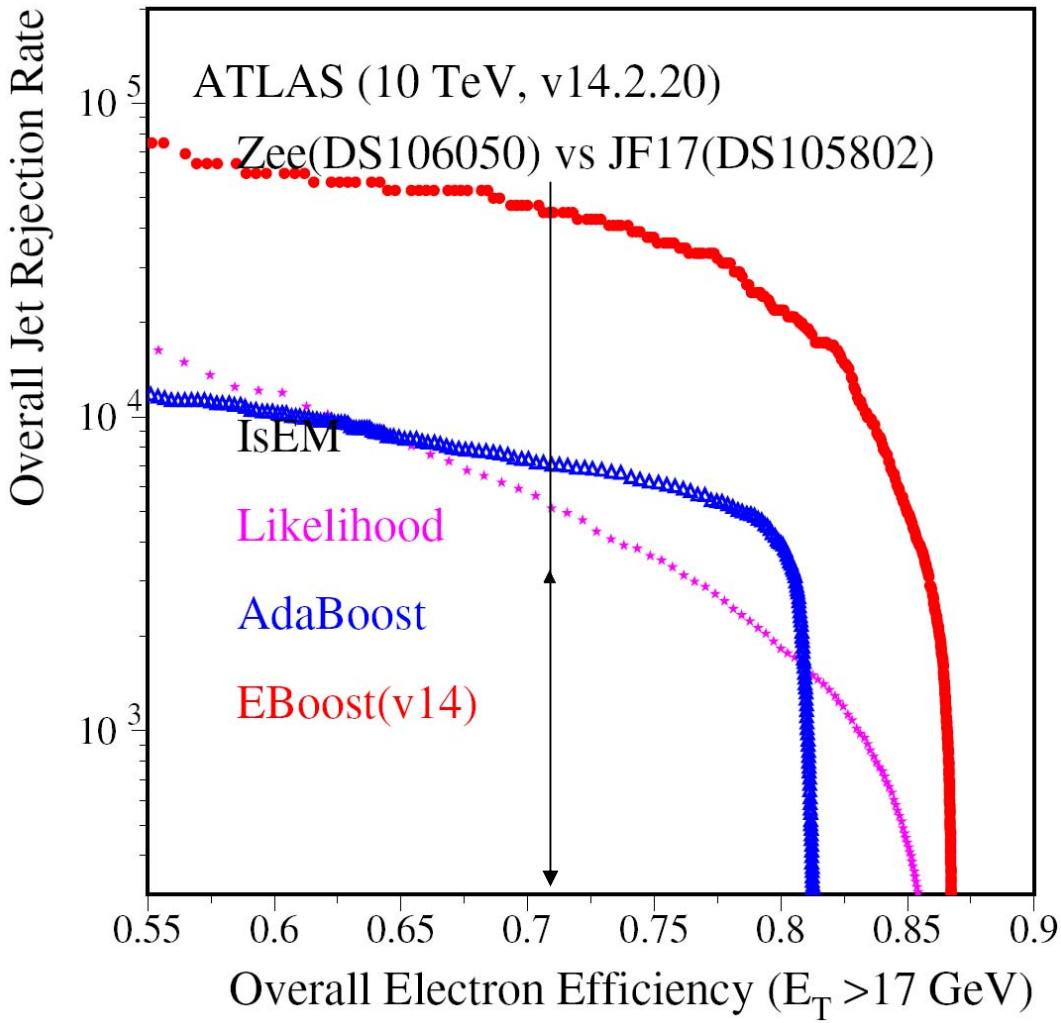
Implementation of EBoost in EgammaRec (Rel. v14)

- H. Yang's talk at ATLAS egamma conference on Dec. 17, 2008

  http://indico.cern.ch/conferenceDisplay.py?confId=43117

Comparison of e-ID Algorithms (v14)

➔ **A technical note about BDT e-ID will be submitted shortly**



➔IsEM (tight)
Efficiency = 70.9%
jet rejection rate=3092

➔Likelihood
Efficiency = 71%
jet rejection rate=5200

➔AdaBoost
Efficiency = 71%
jet rejection rate=7059

➔EBoost
Efficiency = 71%
jet rejection rate=47830

Current Works on e-ID background from γ-jet and D3PD Validation

➔ Background study using γ-jet Samples

- H. Yang et.al., "First Look at γ-jet samples using BDT e-ID algorithm", ATLAS Egamma Phone Conference, 03/09/2009

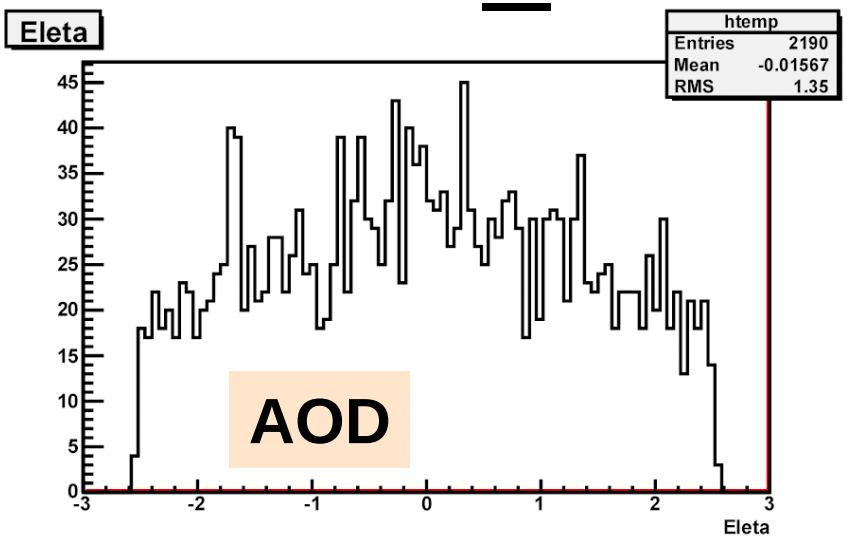
➔ 1st Version D3PD Validation

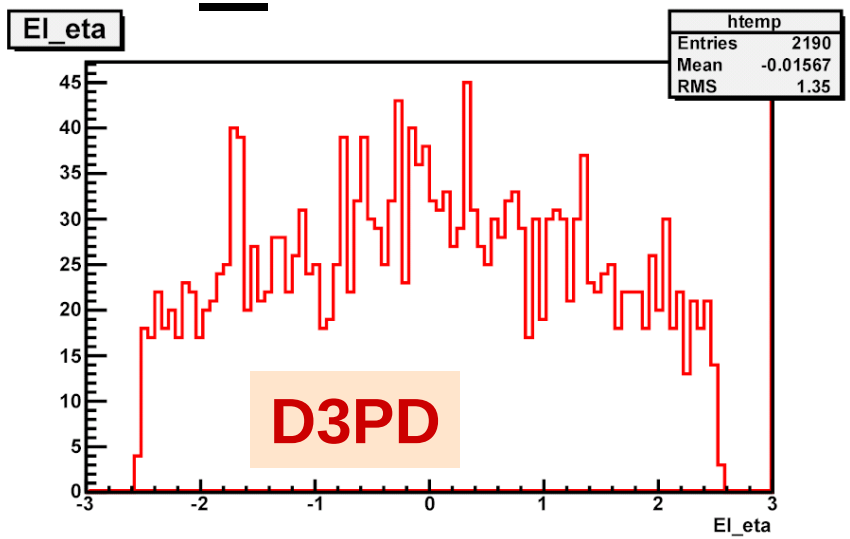
- H. Yang, "Validation of D3PD using e-ID related variables", US-ATLAS Egamma Meeting, 01/29/2009
- H. Yang, "Update on validation of D3PD using e-ID related variables", US-ATLAS Egamma Meeting, 02/26/2009
- H. Yang, "Update of D3PD validation using e-ID related variables", US-ATLAS Egamma Meeting, 03/12/2009

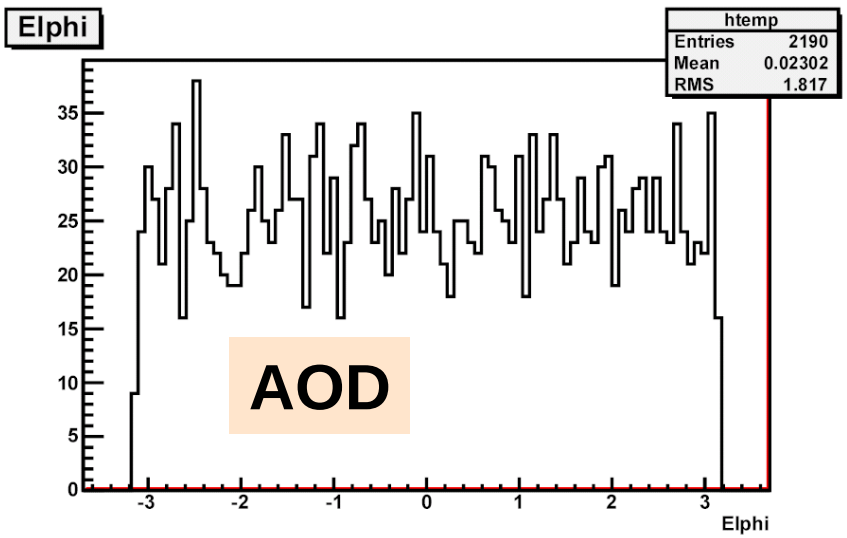
2nd Version D3PD Validation using Zee Samples

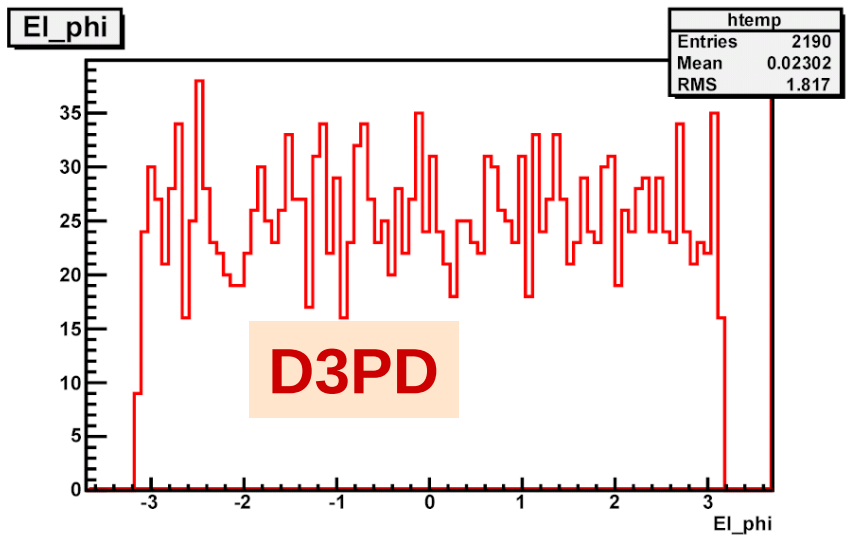
- D3PD samples (5 files)
  - egamma_PythiaZee_1Lepton.recon.v142203.[1-5].root
- Five identical AOD files for validation
  - mc08.106050.PythiaZee_1Lepton.recon.AOD.e347_s462_r541_tid028675/AOD.028675._0780[2-6].pool.root.1
  - Use DoElectron.cxx to extract electron related variables.

El_Eta and El_Phi

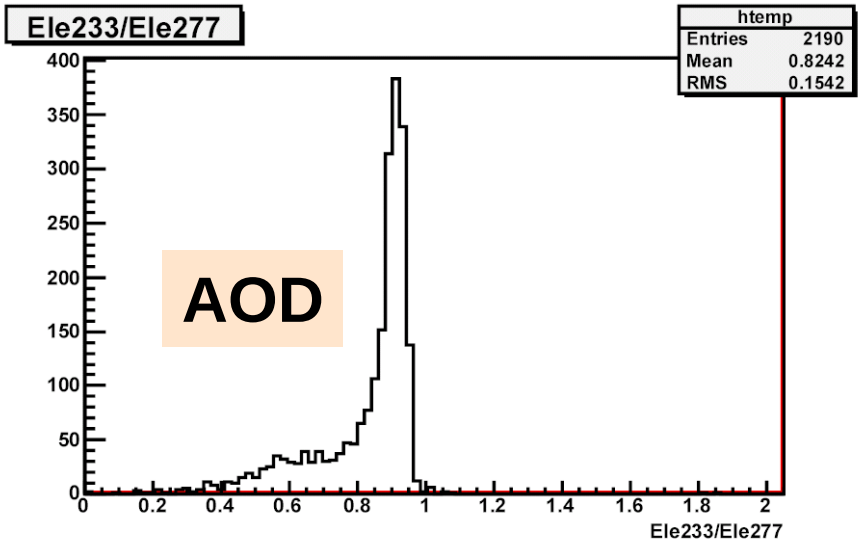


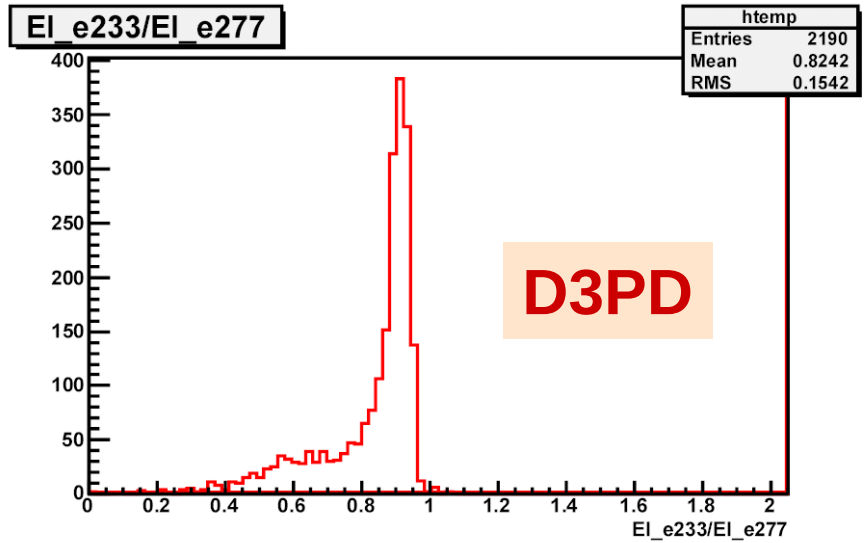


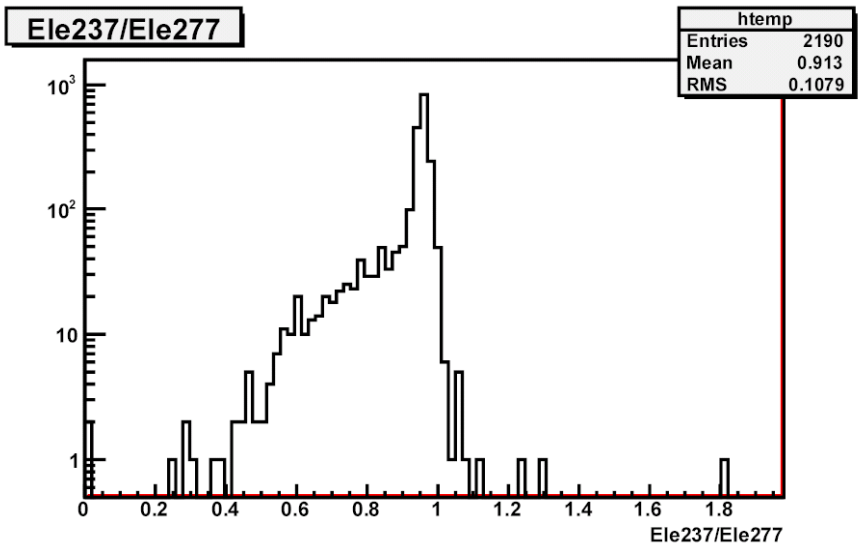


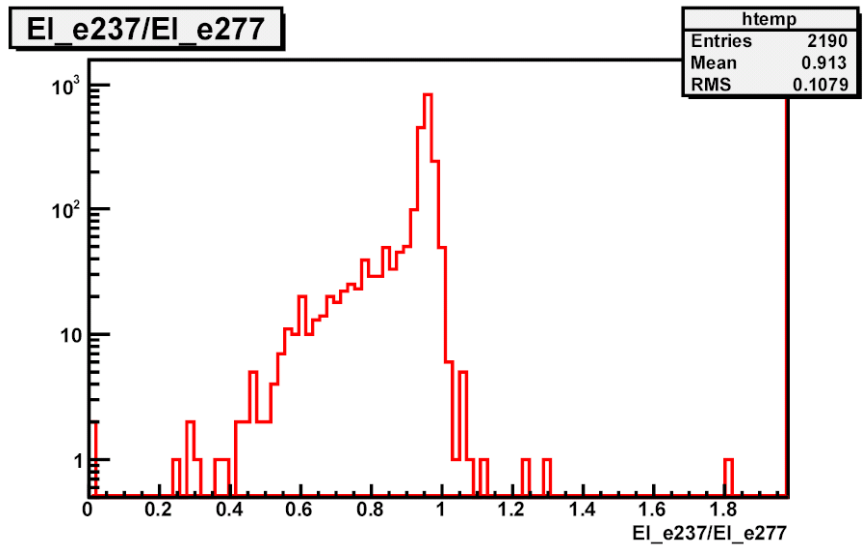


E233/E277 and E237/E277

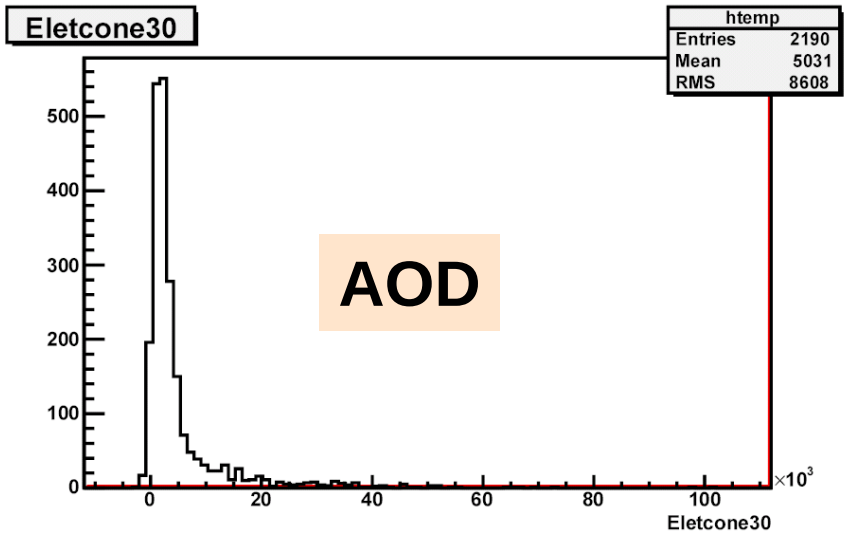


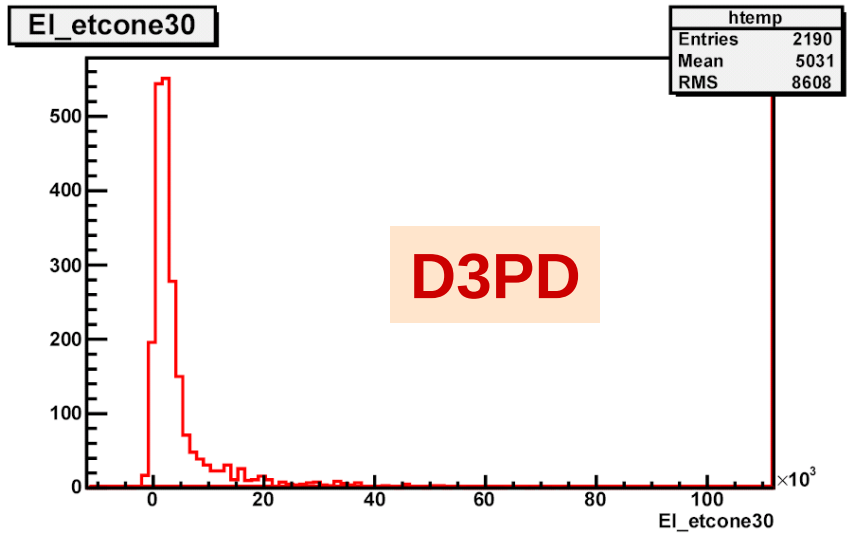


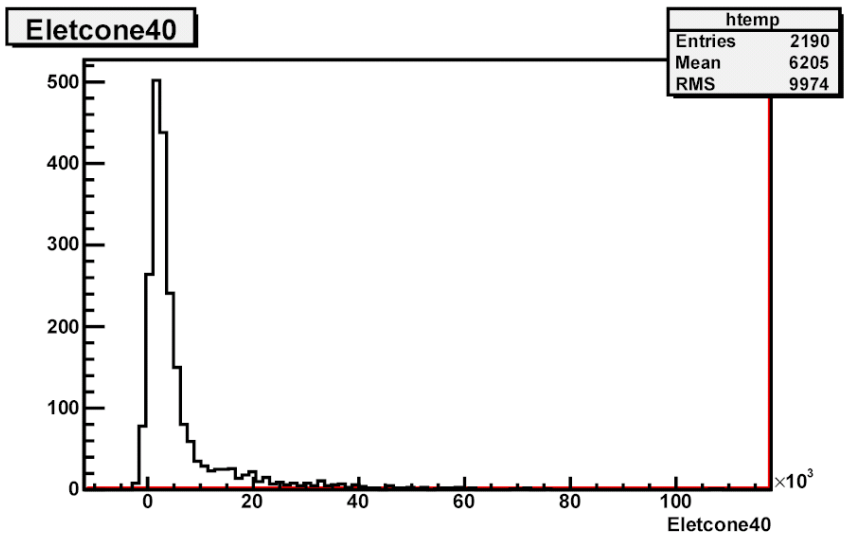


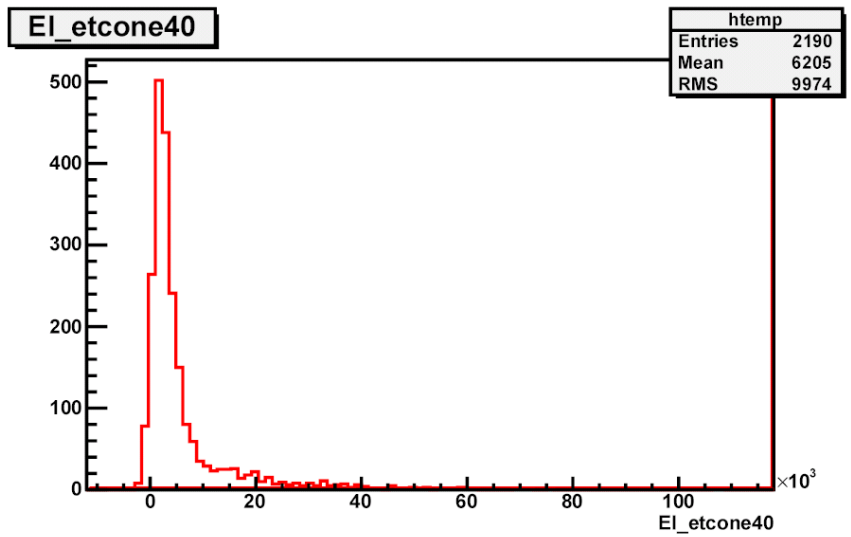


EtCone30 and EtCone40

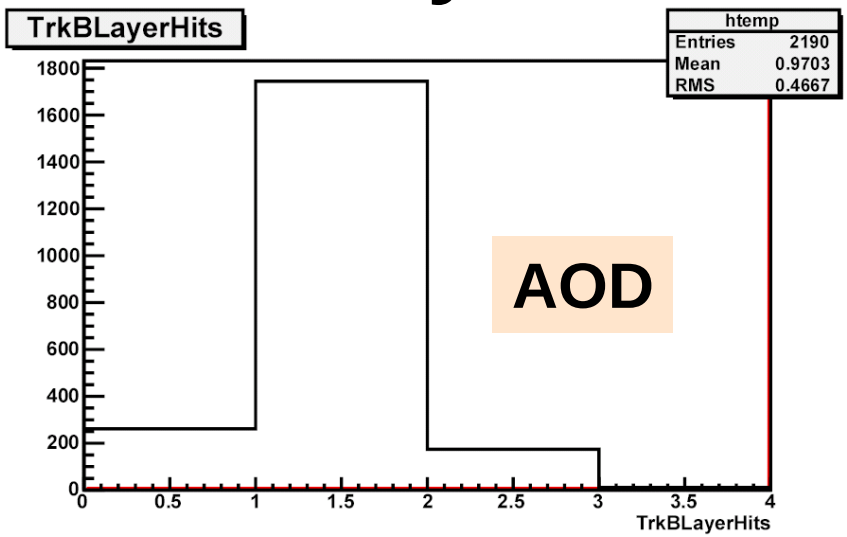


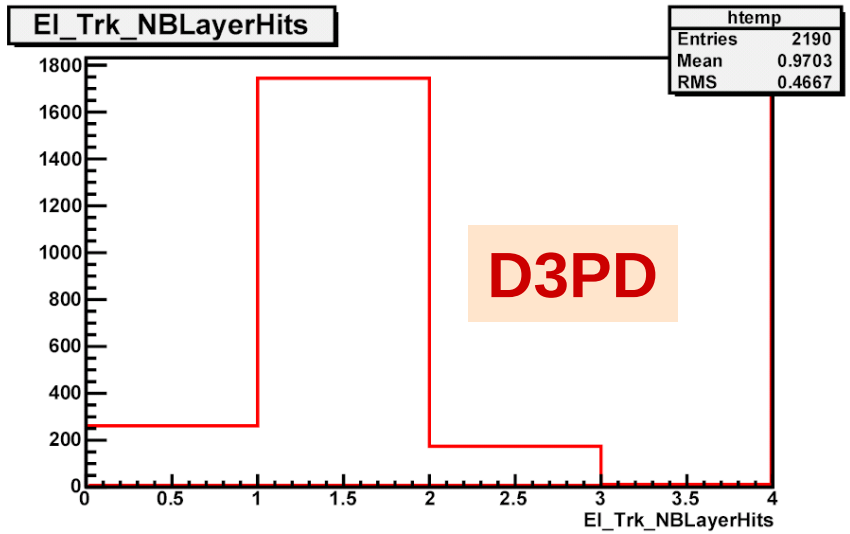


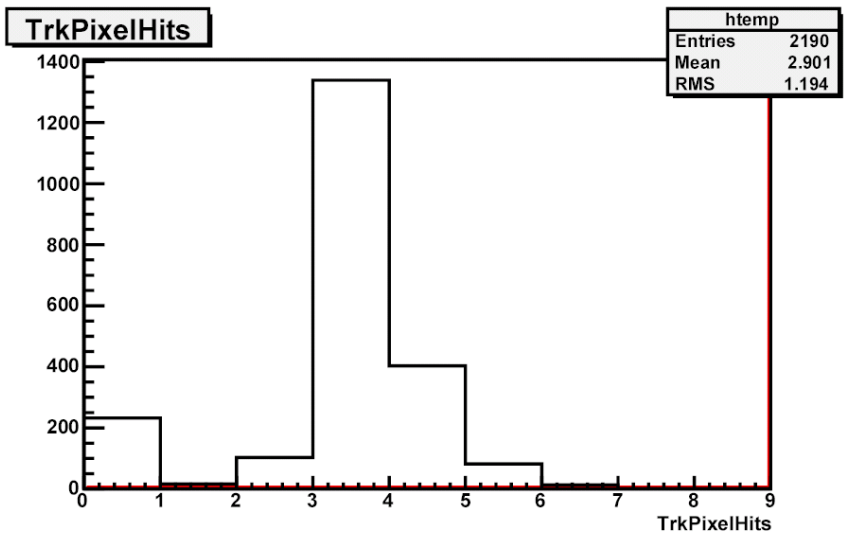


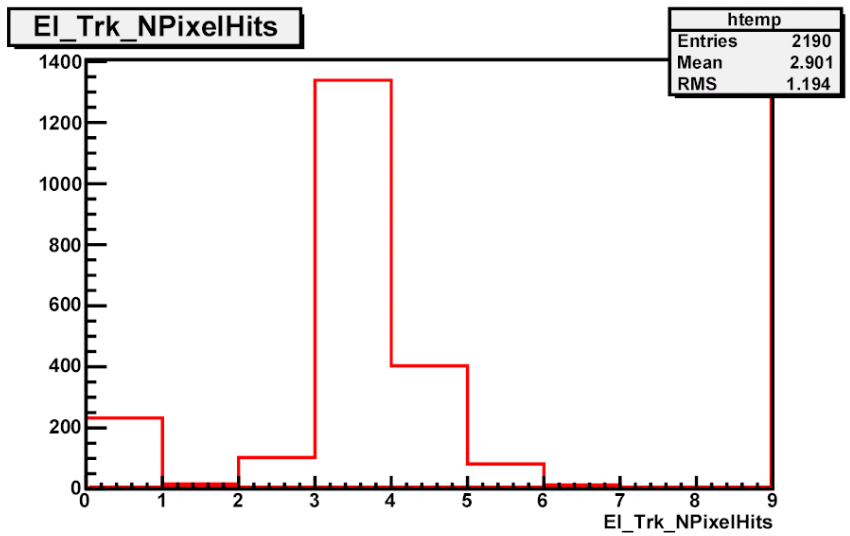


BLayerHits and Pixel Hits

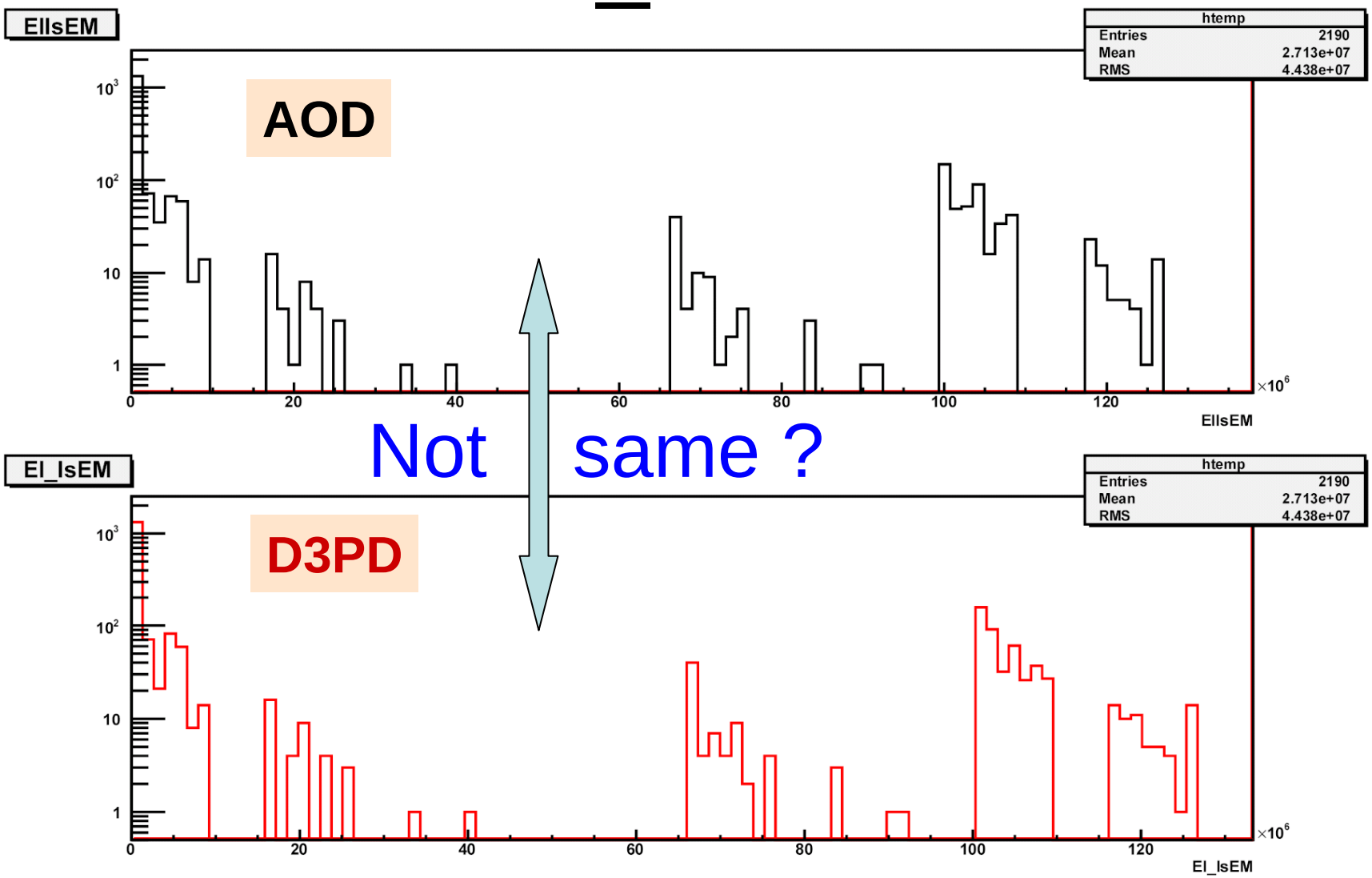




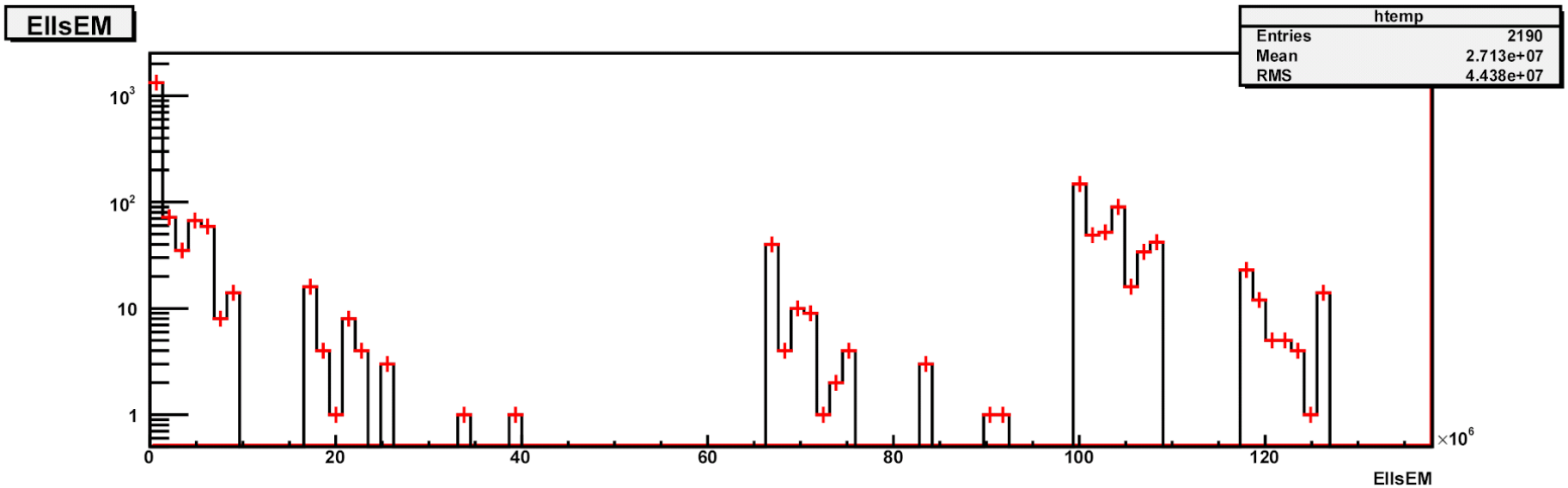




EL_lsEM ?


ElIsEM
htemp
Entries 2190
Mean 2.713e+07
RMS 4.438e+07
AOD
Not same ?
El_lsEM
htemp
Entries 2190
Mean 2.713e+07
RMS 4.438e+07
D3PD
ElIsEM
El_lsEM

EL_IsEM



AOD (black) and D3PD (red crosses) are identical if they are compared in same plot !

γ-jet: e-ID Background

- To estimate the γ jet and dijet rejection rates using various e-ID algorithms (IsEM, Likelihood and BDT).

- MC samples for test include:
  - DS108087, γ jet
  - DS105802, JF17 dijet
  - DS106050, Z→ee signal

MC γ jet samples for test (Pre-selection)

samples / N_electron	DS108087 γ jet	DS105802 JF17 dijets	DS106050 Z→ee
N_events	127887	237950	83690
N_candidate Et>17 GeV, |η|<2.5	194046	896818	108550
N_candidate (precuts) With EM/Track match	20441	20994	94153
Rejection/Efficiency after precuts	Rejection 9.5	Rejection 42.7	Acceptance 86.7%

Results of IsEM

- Et > 17GeV, Tight cuts

  - Efficiency (Z→ee) = 70.9%

  - Rejection (γjet) = 426(±4.7%)

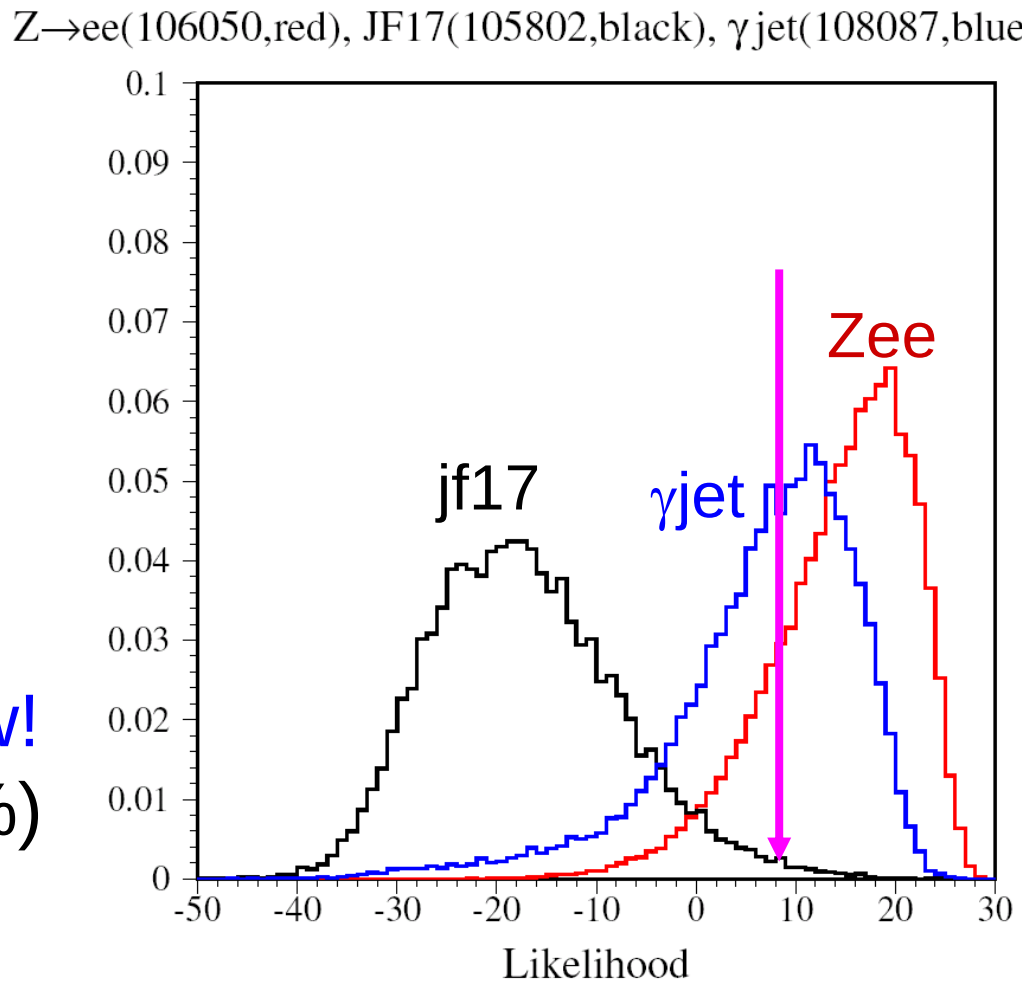
  - Rejection (jf17) = 3092(±5.9%)

e-ID (Likelihood)

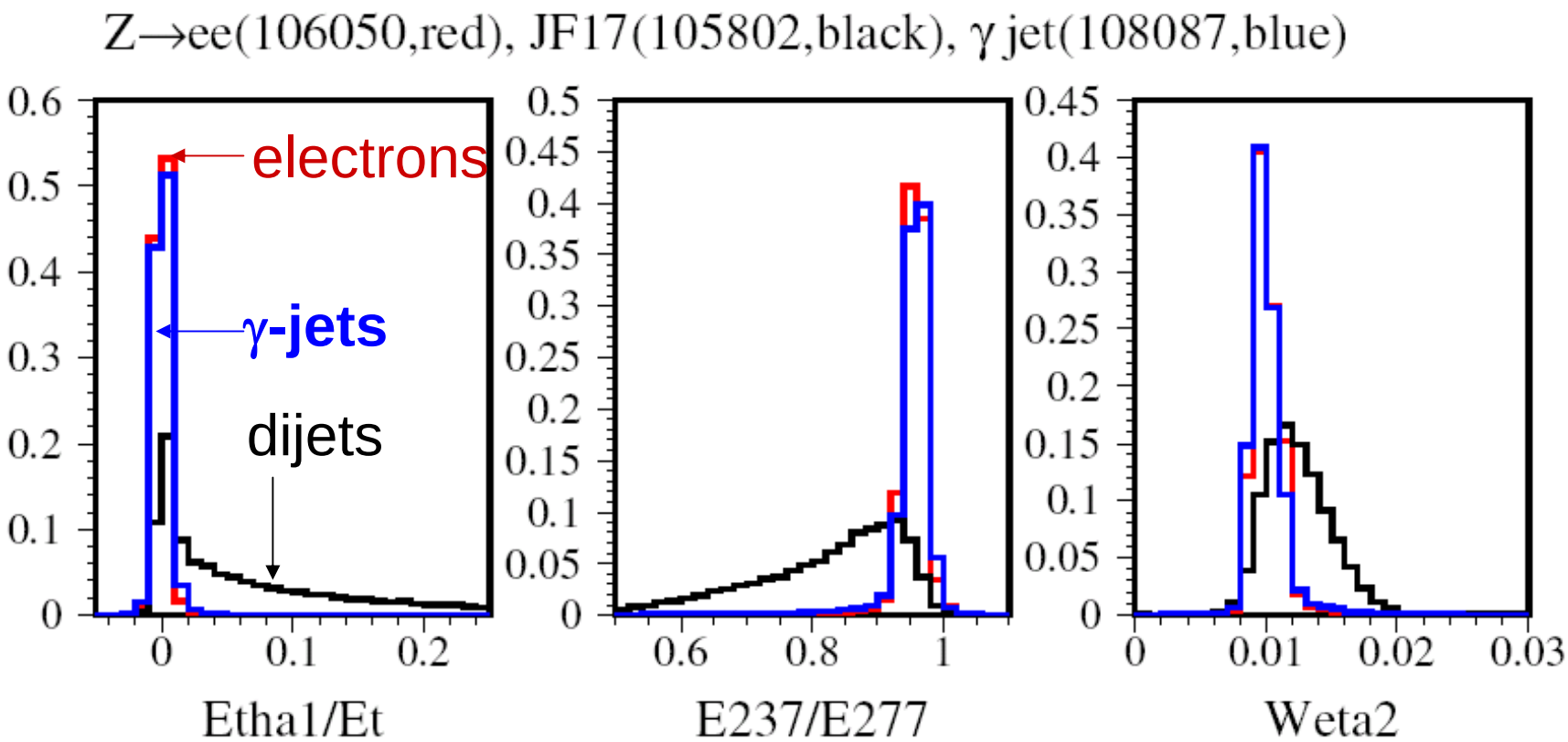
log(ElectronWt/BgWt)

Et > 17GeV

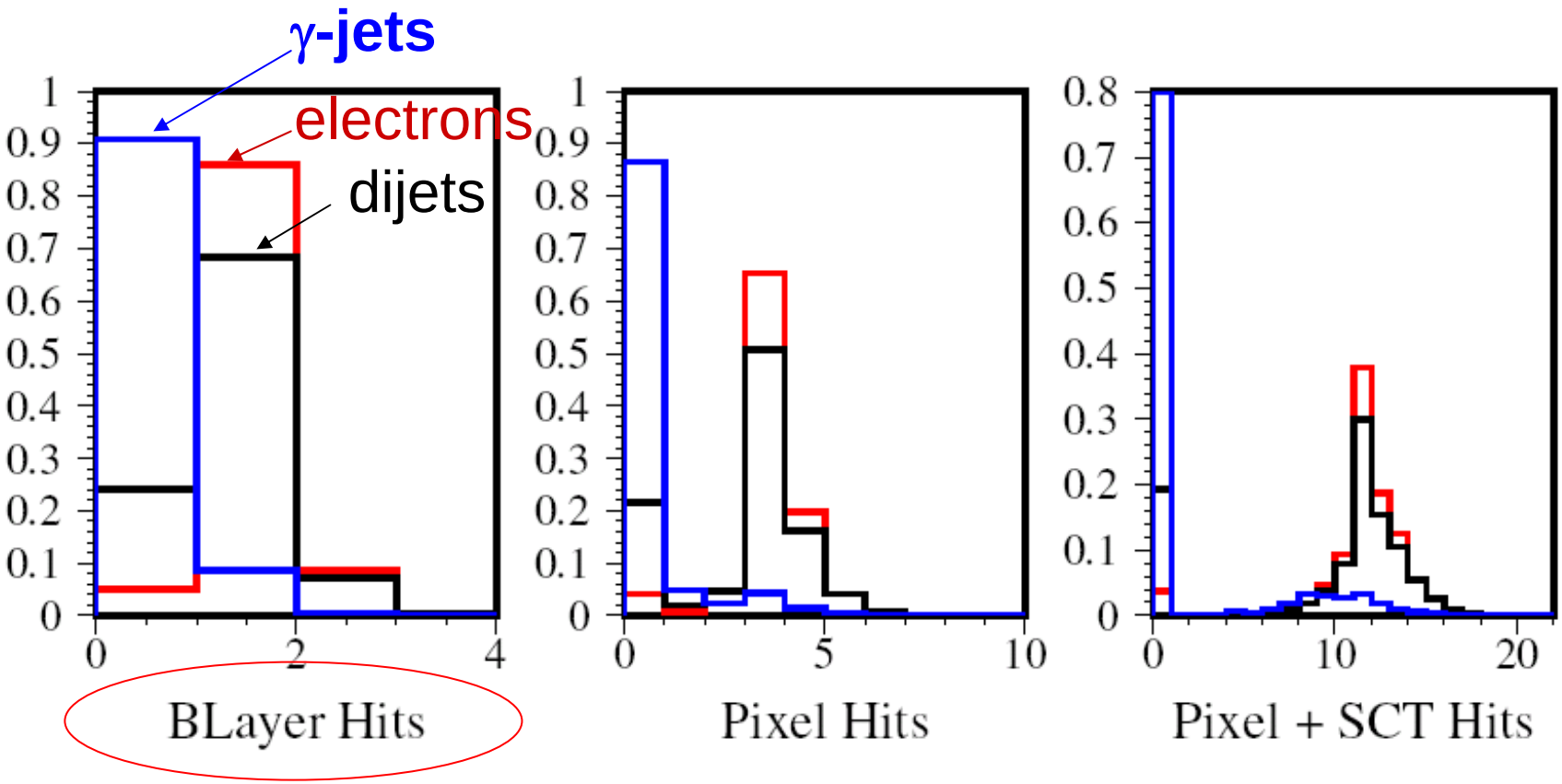
- Efficiency = 71%
- Rej(γjet)=20(±1%) low!
- Rej(jf17)=5200(±7.6%)



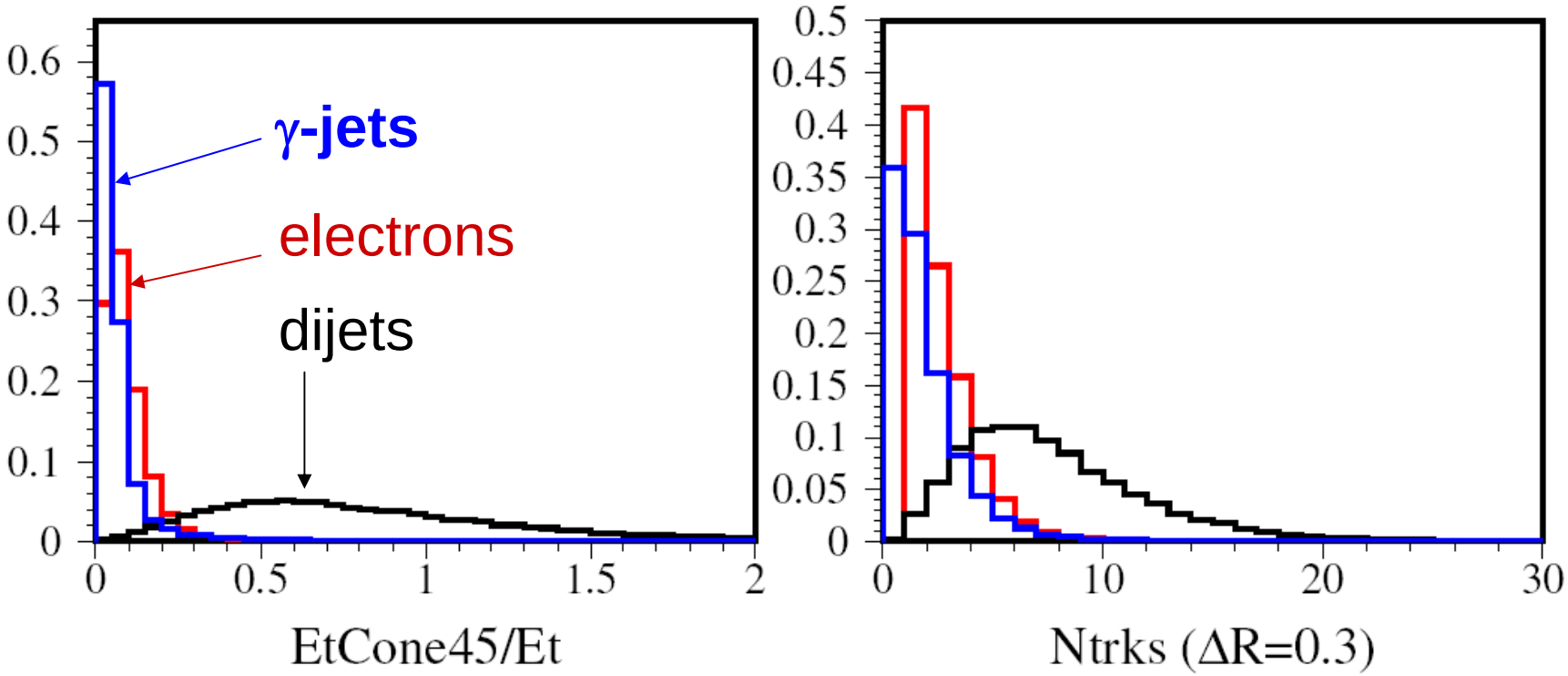
Input Variables for BDT



Input Variables for BDT



Input Variables for BDT



BDT e-ID trained using dijet as background (list of top 10 powerful variables)

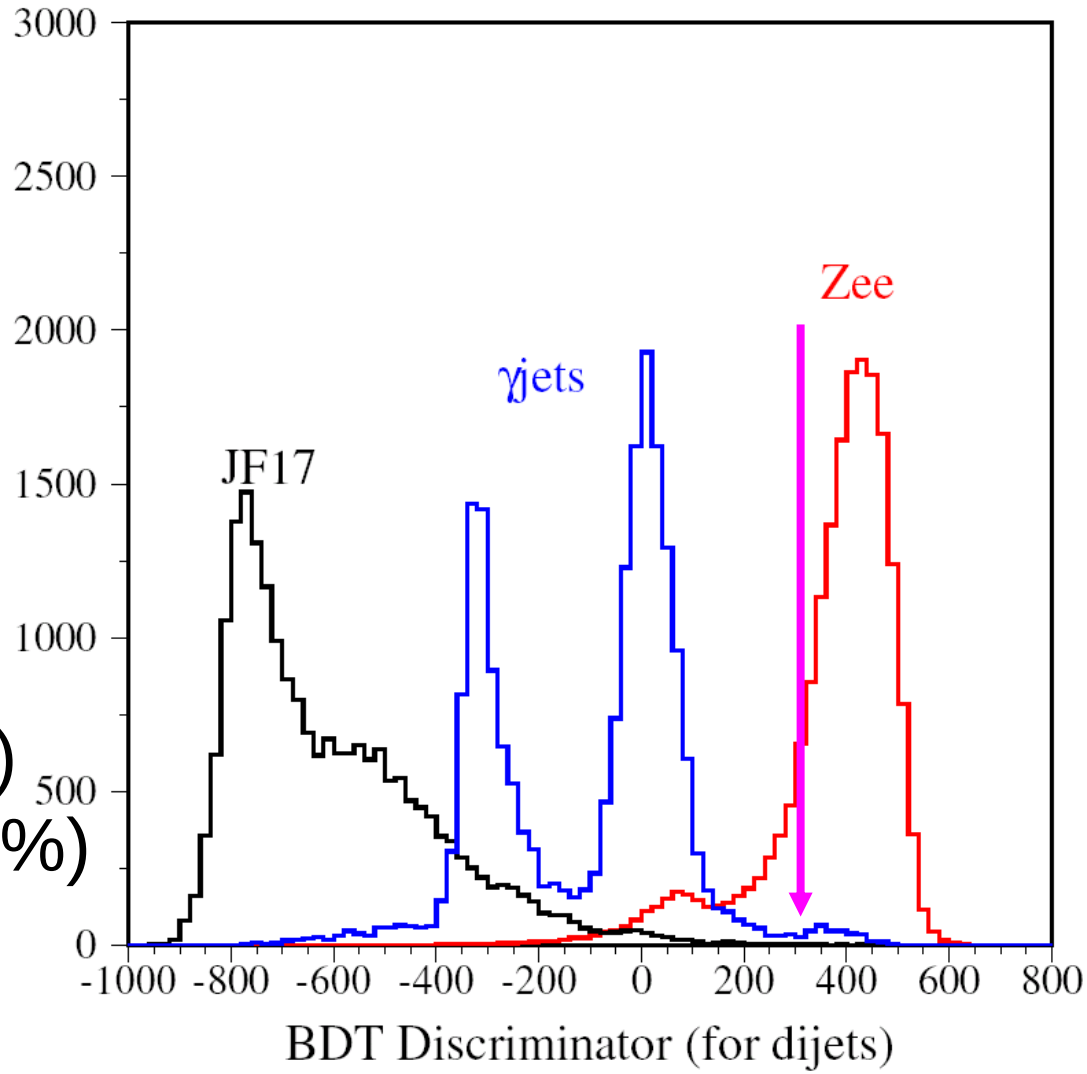
Rank	Input variable	Gini index
1	Etcone45 / Et	46.08%
2	E2tsts1-Emins1(Emax2-Emin in LAr.1st)	8.60%
3	No. of TRT hits / No. of B-layer hits	6.68%
4	deta1 between track and EM cluster	5.21%
5	Number of pixel hits	4.48%
6	F1(frac. of E deposited in LAr. 1st samp)	4.32%
7	Ethad1/Et (E leakage in hcal. 1st samp)	3.94%
8	E237 / E277	3.41%
9	Eta of inner track	2.33%
10	Number of B-layer hits	2.19%

BDT e-ID trained using dijets as background

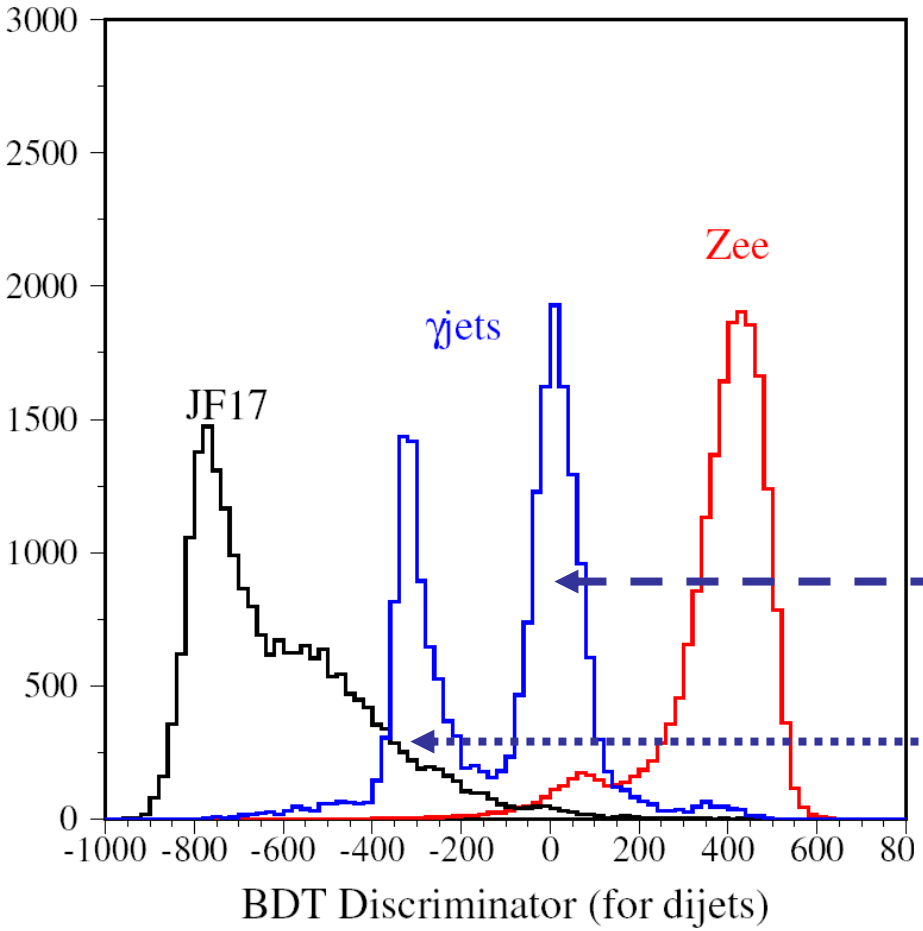
e-ID (BDT_dijet)

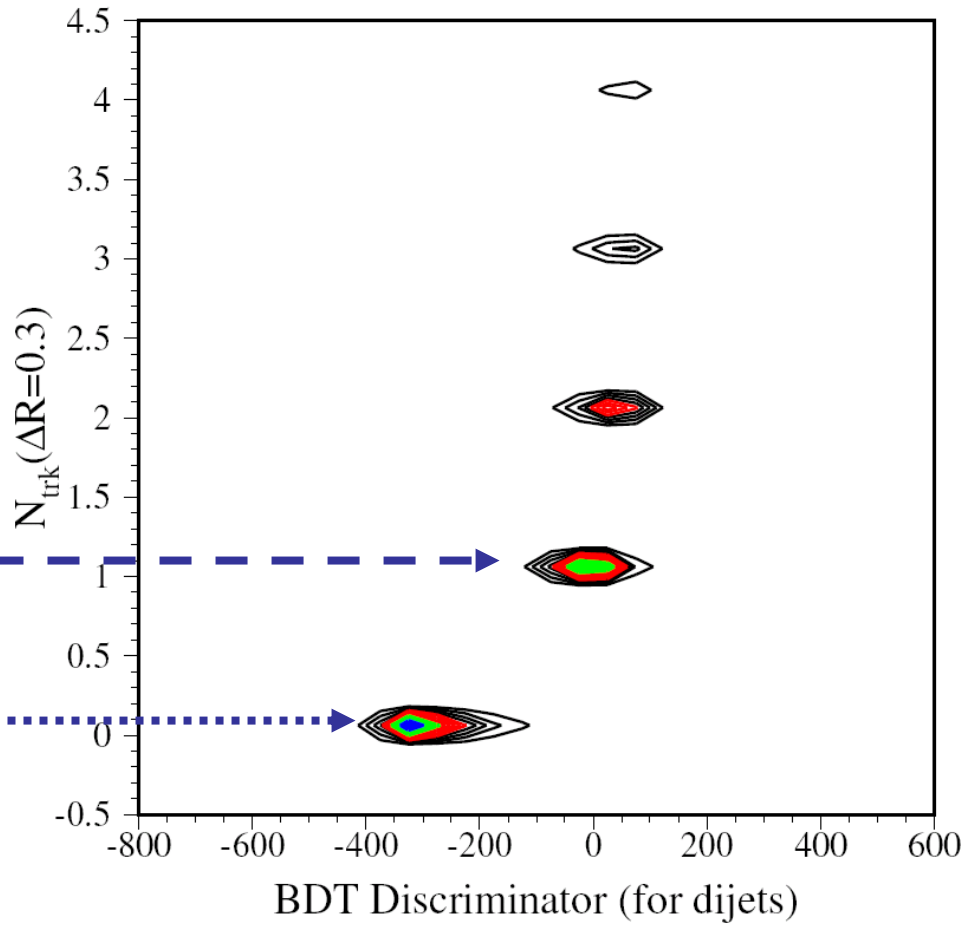
Et > 17GeV

- Efficiency = 71%
- Rej(γjet)=591(±5.5%)
- Rej(jf17)=47830(±23%)

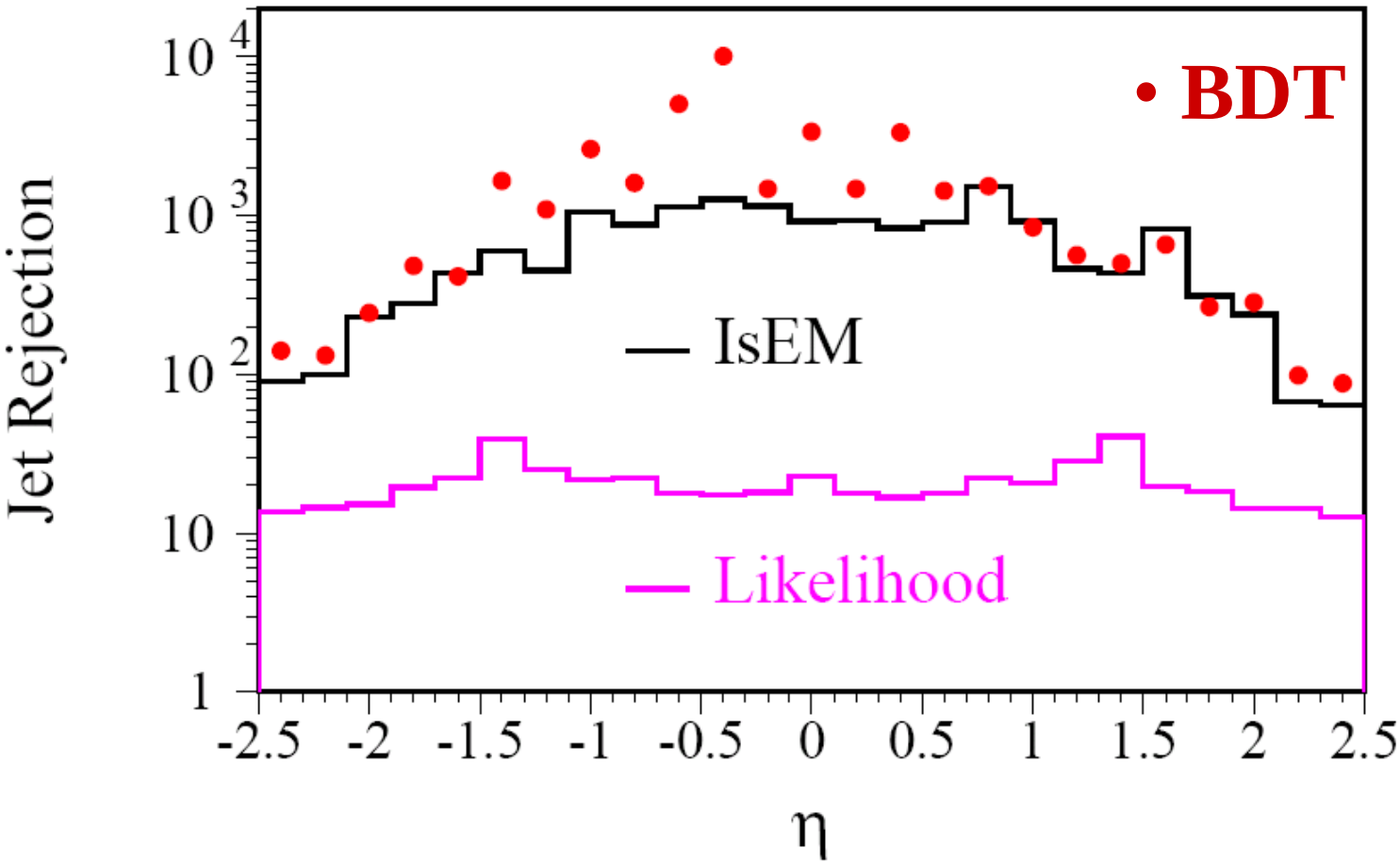


BDT peaks in γ jet samples? Correlations

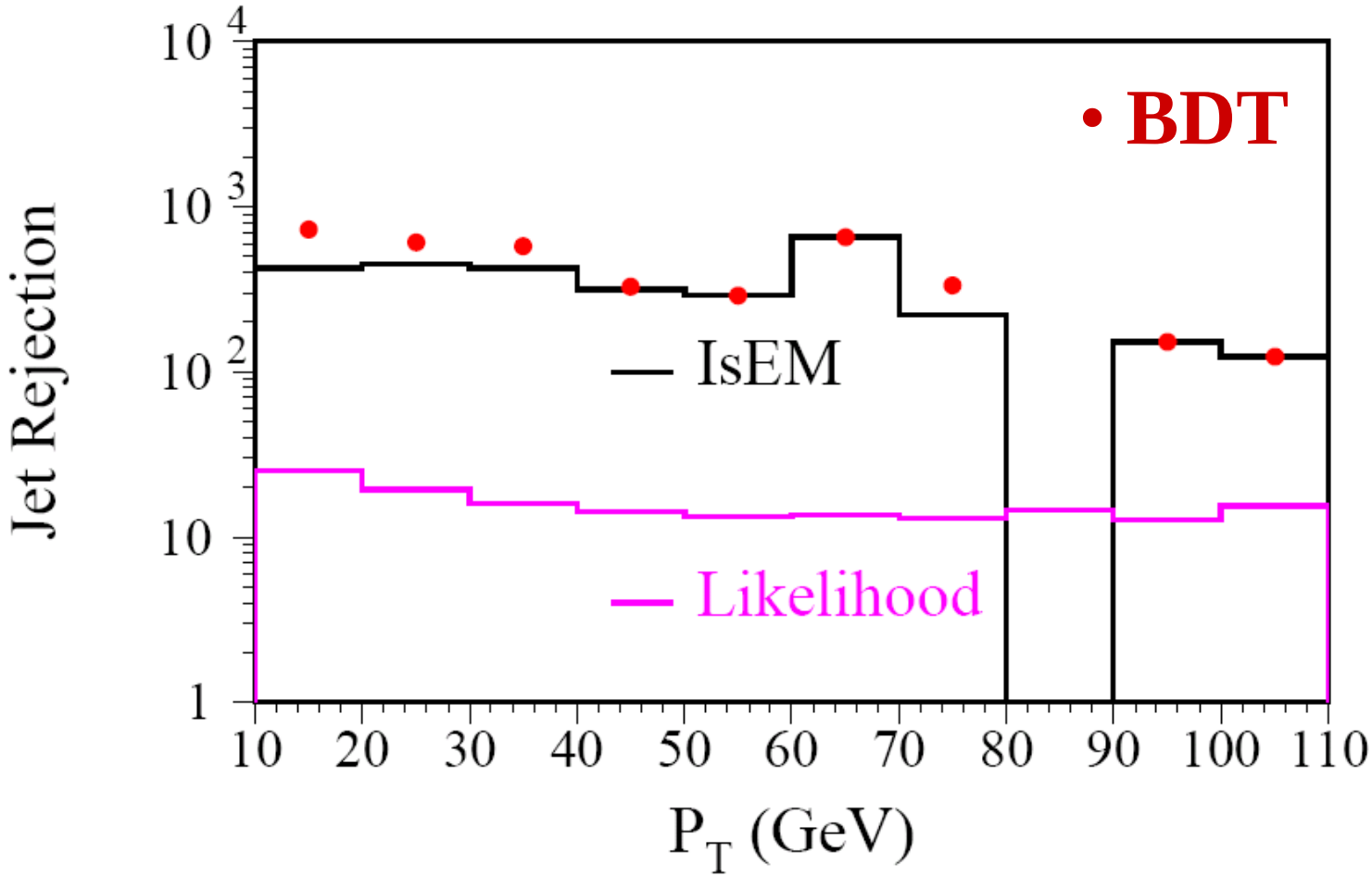




Comparison of IsEM, Likelihood and BDT (trained using dijets)



Comparison of IsEM, Likelihood and BDT (trained using dijets)



BDT e-ID trained using γ jets as background (list of top 10 powerful variables)

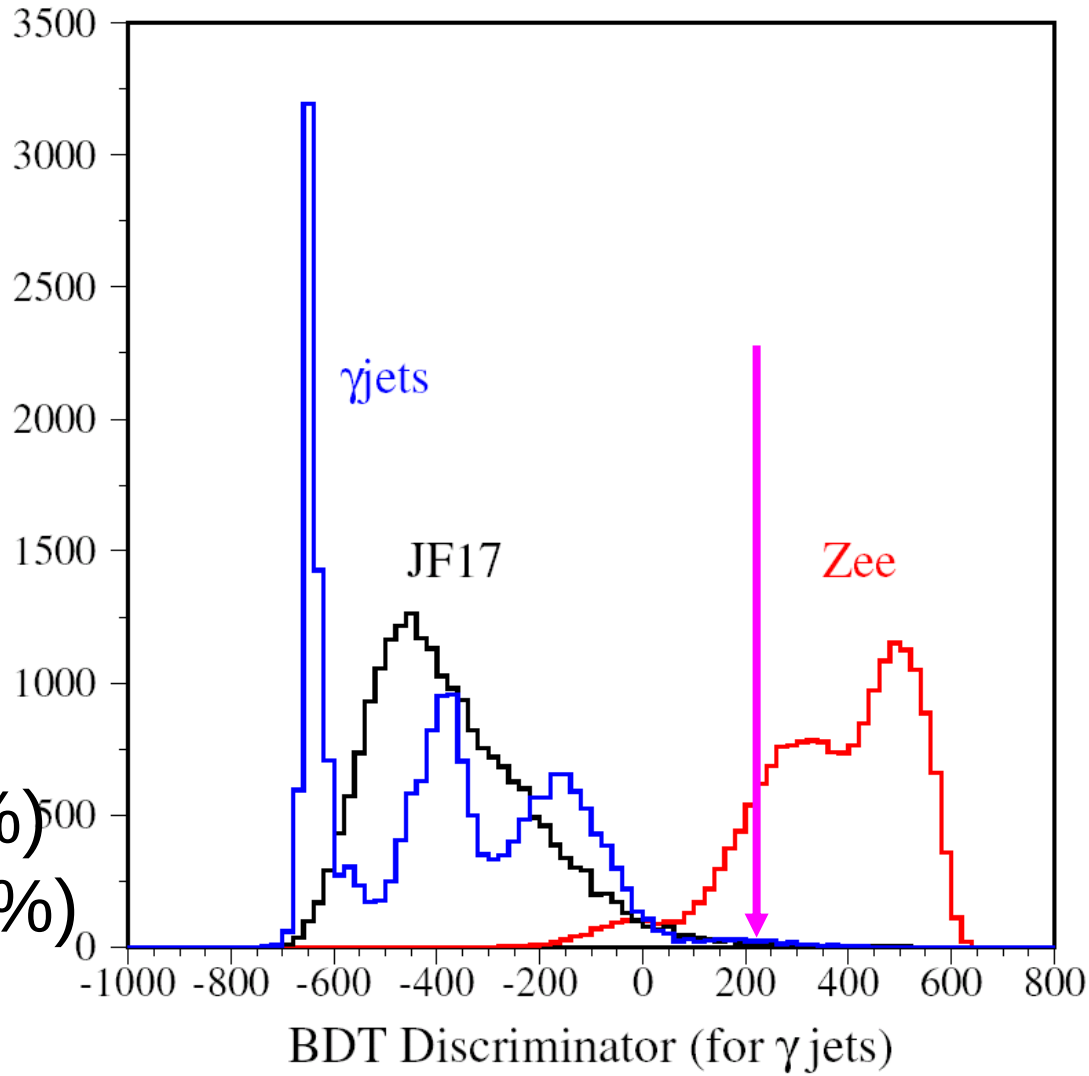
Rank	Input variable	Gini index
1	Number of B-layer hits	21.49%
2	Ntrk ($\Delta R=0.3$)	17.94%
3	$\sum$Pt ($\Delta R=0.3$)	11.41%
4	Number of pixel hits	11.17%
5	E233 / E277	5.79%
6	E237 / E277	4.77%
7	No. of TRT hits / No. of B-layer hits	4.52%
8	deta1 between track and EM cluster	4.49%
9	Etcone45 / Et	4.20%
10	F1(frac. of E deposited in LAr. 1st samp)	1.97%

BDT e-ID trained using γ jets as background

e-ID (BDT_γjet)

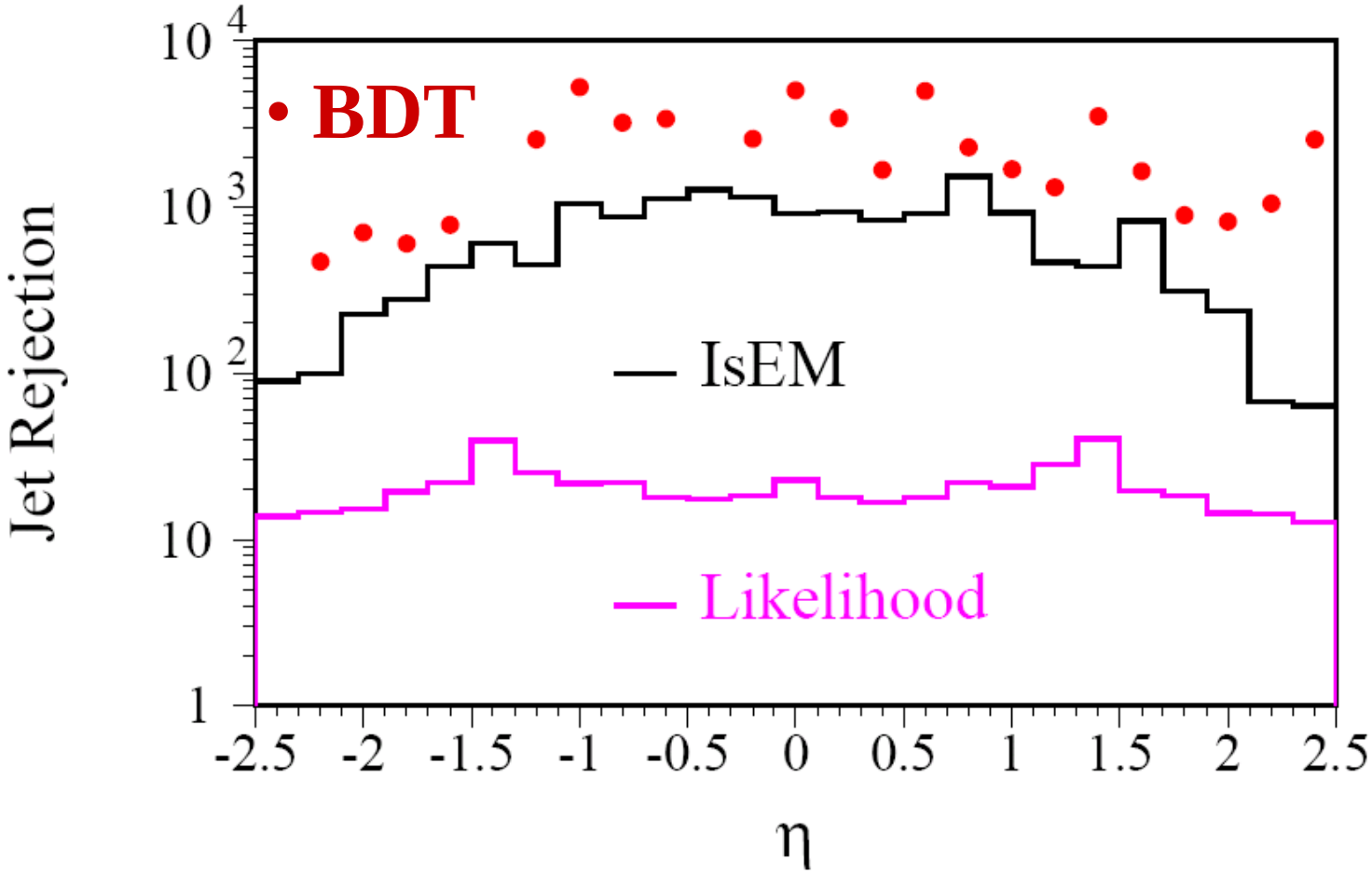
Et > 17GeV

- Efficiency = 71%
- Rej(γjet)=1788(±9.6%)
- Rej(jf17)=19081(±15%)



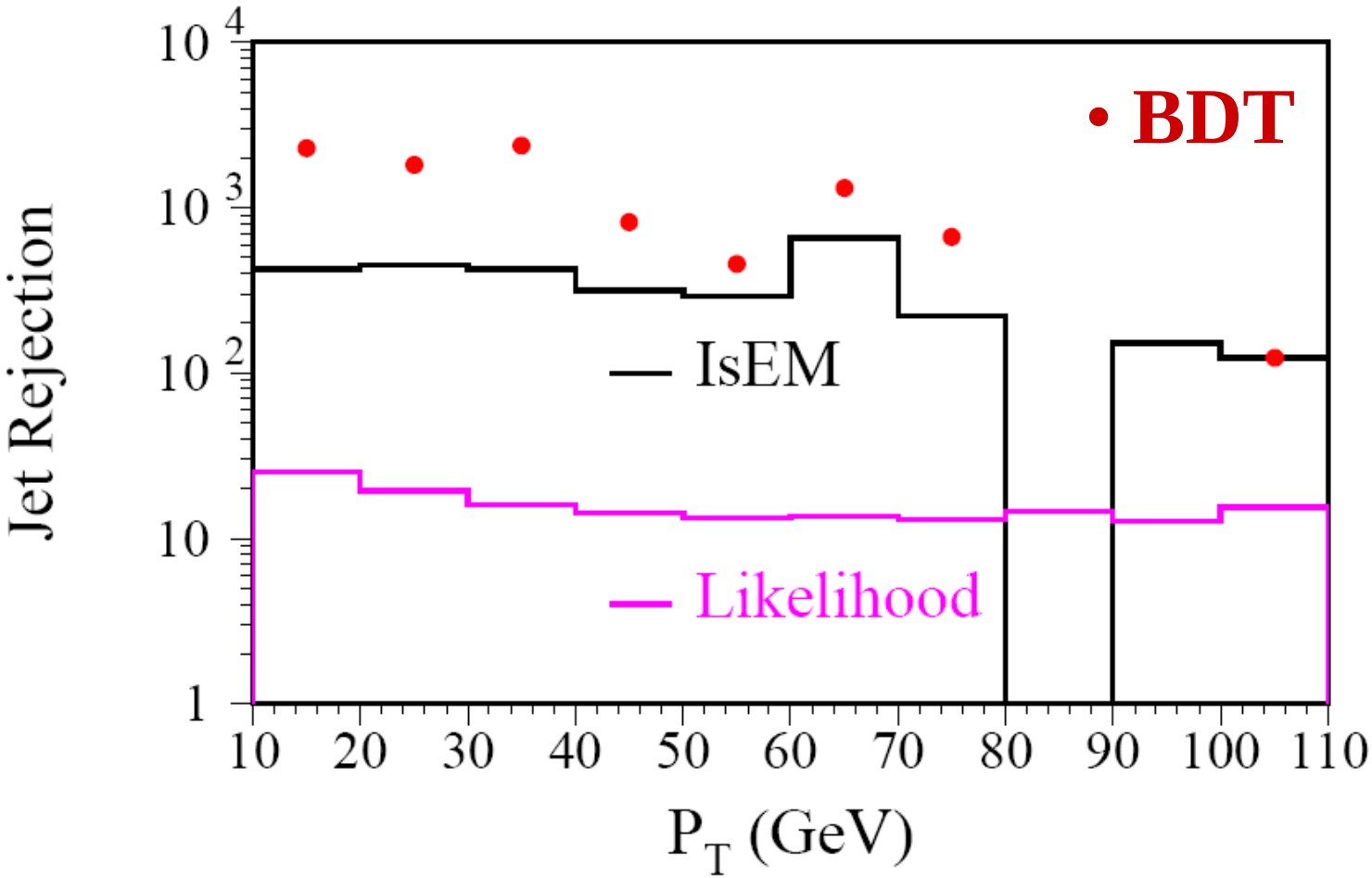
Performance Comparison

BDT e-ID trained using γ jets as background



Performance Comparison

BDT e-ID trained using γ jets as background

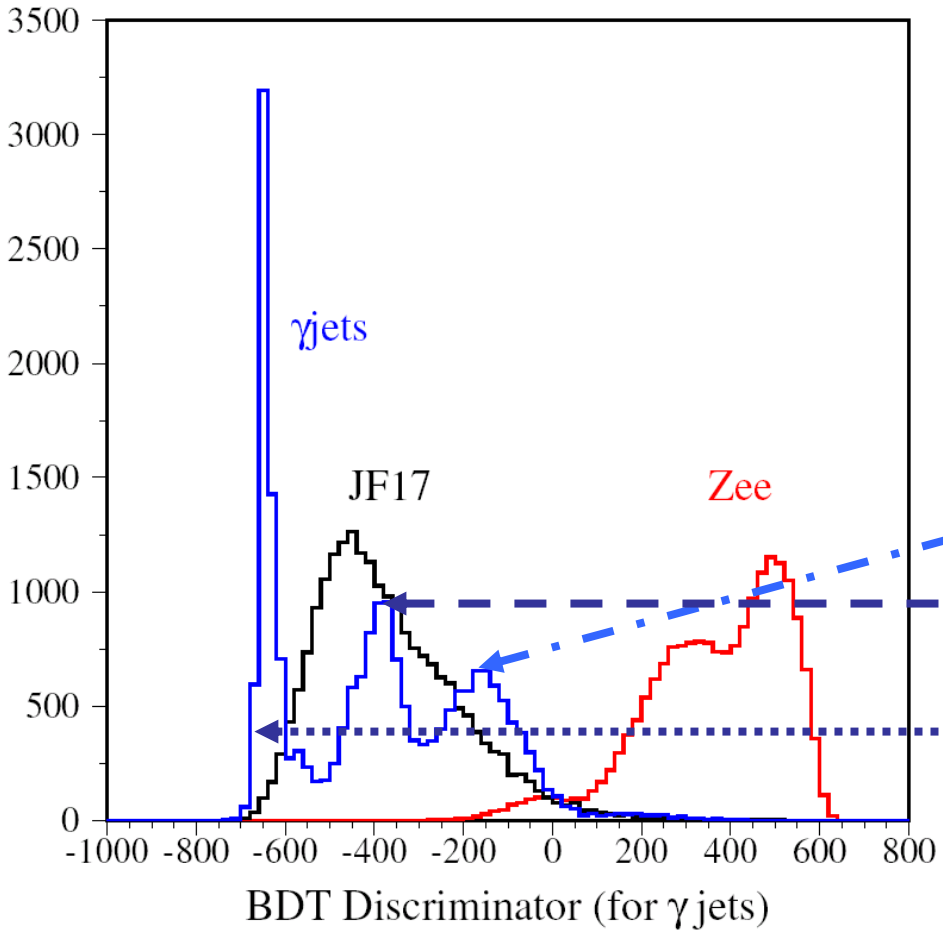


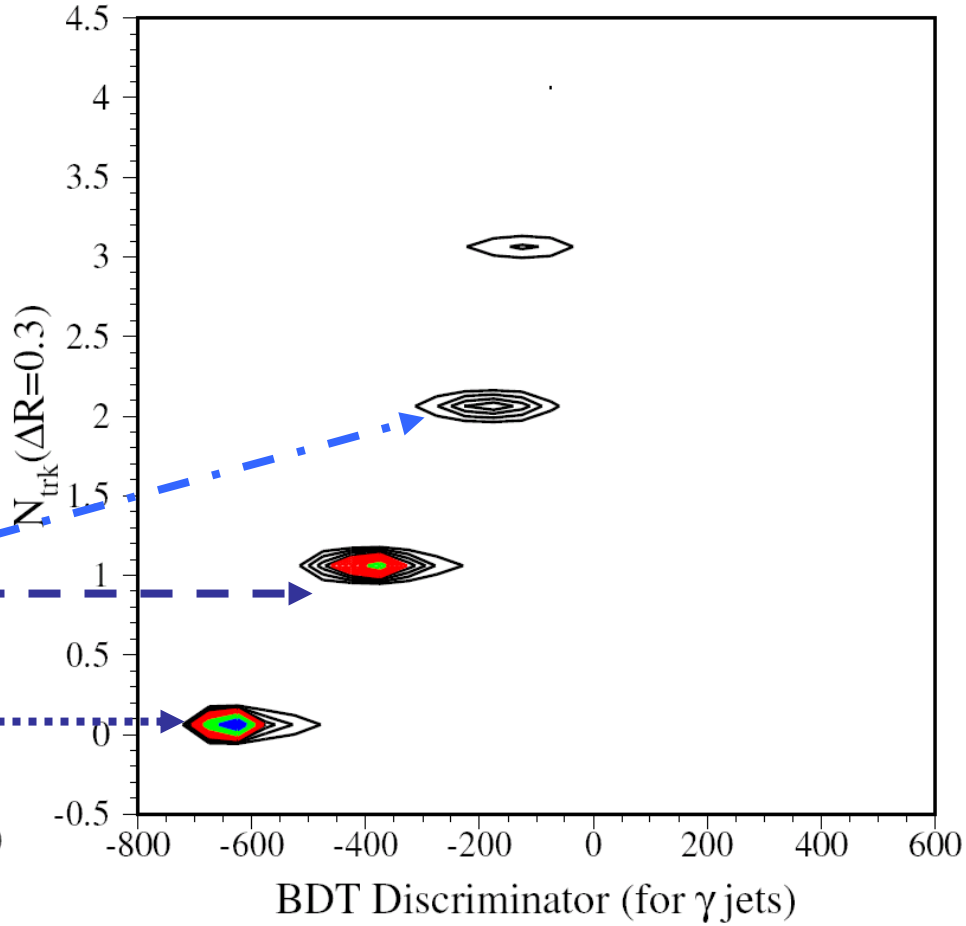
Summary

e-IDs \ samples	DS108087 γJet Rejection	DS105802 JF17 dijets Rejection	DS106050 Z→ee Acceptance
IsEM (tight)	426(±4.7%)	3092(±5.9%)	71%(±0.4%)
Likelihood	20(±1.0%)	5200(±7.6%)	71%(±0.4%)
BDT(using dijets)	591(±5.5%) 426(±4.7%)	47830(±23%) 27176(±17%)	71%(±0.4%) 77%(±0.3%)
BDT(using γ jets)	1788(±9.6%)	19081(±15%)	71%(±0.4%)

Backup Slides

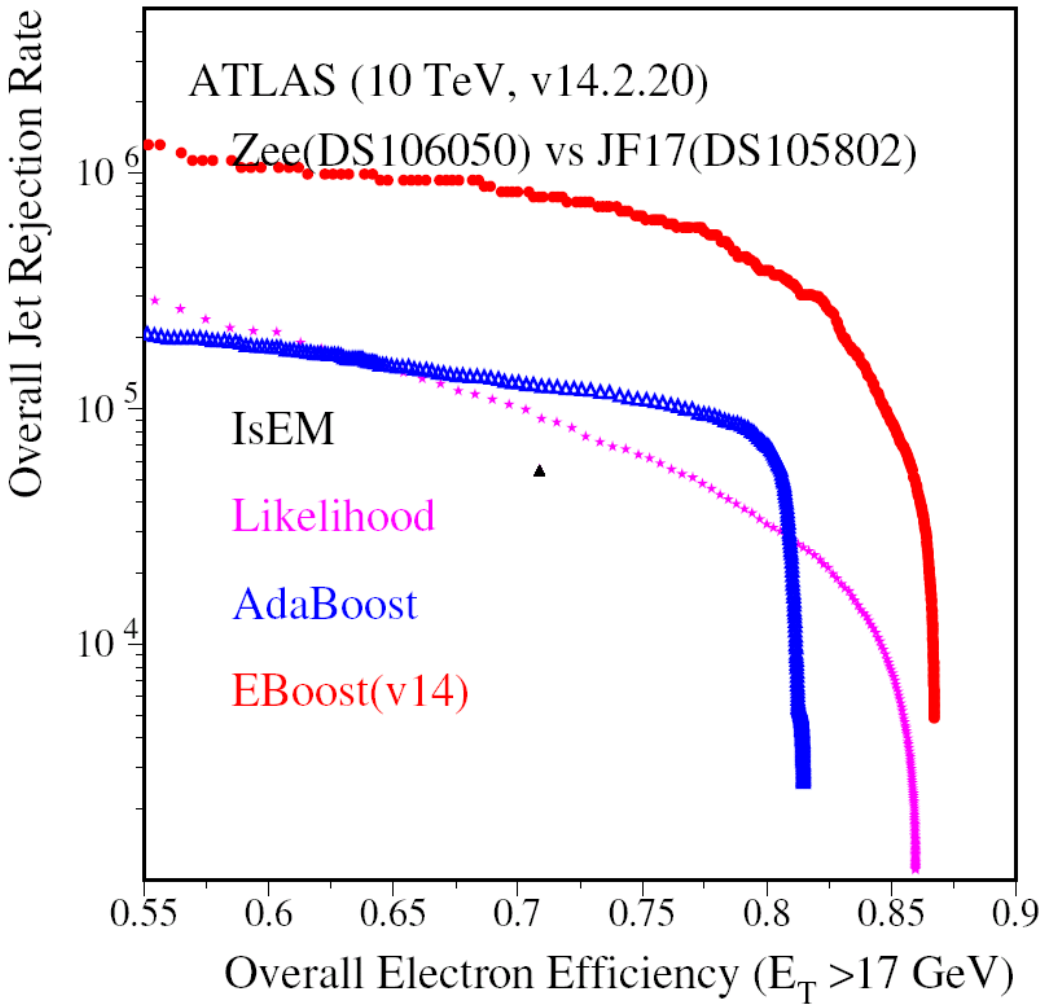
BDT peaks in γ jet samples?





Comparison of e-ID Algorithms (v14)

➔ **Jet rejection rate based on number of MC truth jets**



➔IsEM (tight)
Efficiency = 70.9%
jet rejection rate=5.5e4

➔Likelihood
Efficiency = 80%
jet rejection rate=3.2e4

➔AdaBoost
Efficiency = 80%
jet rejection rate=7.0e4

➔EBoost
Efficiency = 80%
jet rejection rate=3.9e5

Implementation of BDT Trees in Egamma Package and Test

- E-ID based on BDT has been implemented into egamma reconstruction package.
- We successfully run through the reconstruction package based on v14.2.22 and v14.5.0 to test the BDT e-ID.