Topo cluster variables study

Fangyi Guo, IHEP CAS

fangyi.guo@cern.ch

中國科學院高能物理研究所
Institute of High Energy Physics Chinese Academy of Sciences

Previous review

Last report in PhotonID group

- Code for RadiativeZ framework has been finished.
- Shew the signal-bkg and data-MC comparison for topo-cluster variables.
- Feedback from meeting:

  Upload codes to gitlab. New merge request submitted.

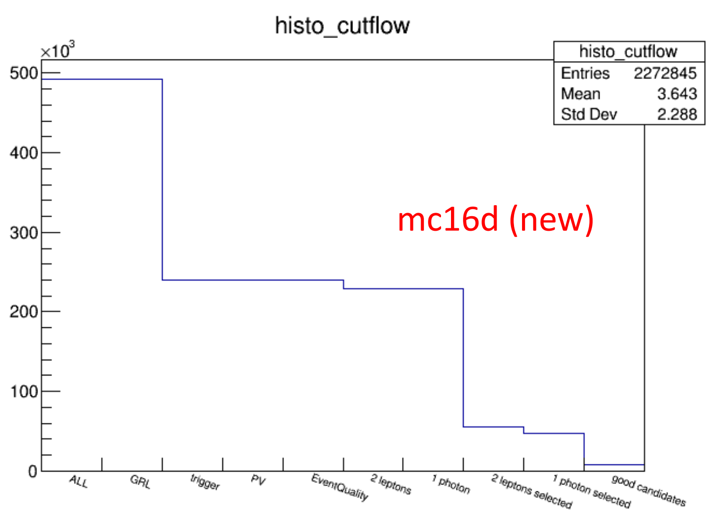
  Check the cutflow from RadiativeZ framework with other Zllg results.

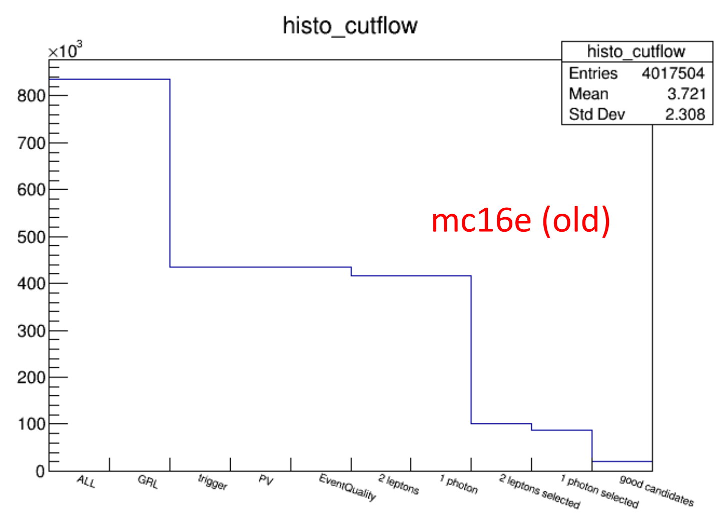
  Change the pre-selections. Some of them are unnecessary.

Cutflow cross check

Didn't find any official cutflow results. So I made one with official package and sample

- Code: RadiativeZ framework in gitlab.
- MC: mc16d $ee\gamma$, $pT_{\gamma} \in [35, 70] GeV$. mc16_13TeV.301899.Sherpa_CT10_eegammaPt35_70.deriv.DAOD_EGAM3.e3952_s3126_r10201_r10210_p3956





Signal-background comparison

Data and MC samples: EGAM3/4 derivation with topo-cluster vars, same as last time.

- Signal MC: mc16e Sherpa llg process, with pT_{γ}.
- Background MC: mc16e PowhegPythia Zee/Zmumu.

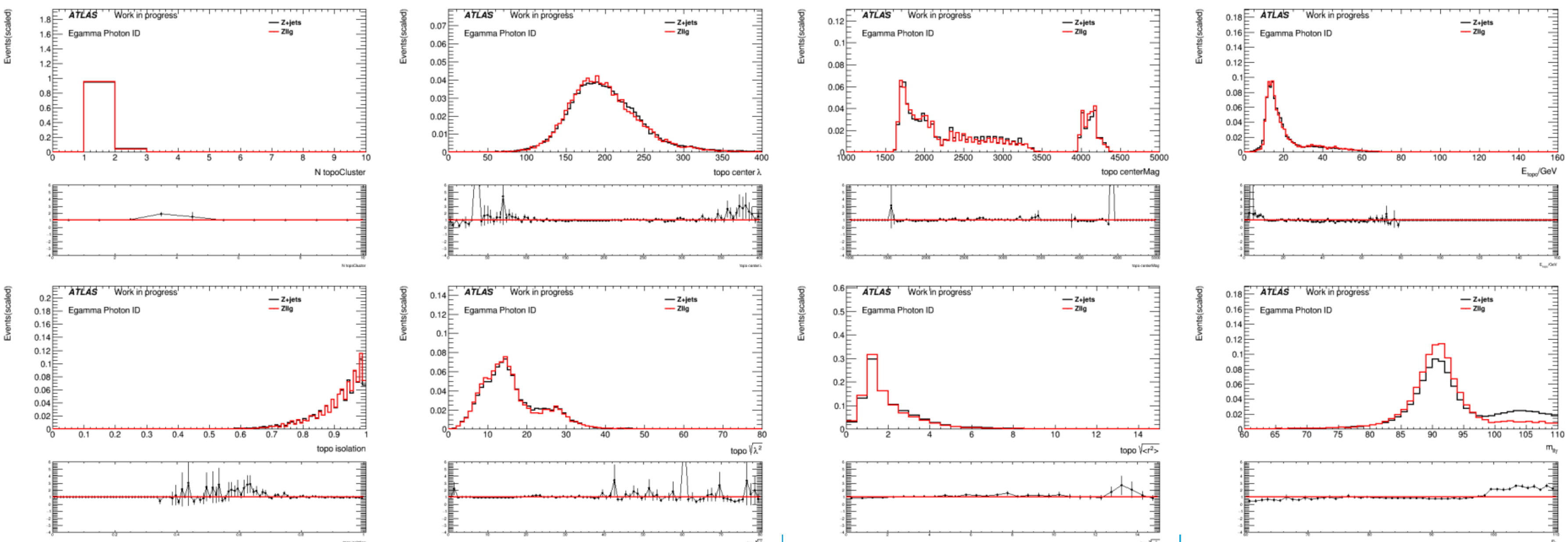
Pre-selection:

- Default selection in ZllgAnalysis framework
  - Add topo-cluster variables into framework (selected photon's first topo-cluster. Gitlab)
- GRL, trigger, Event Quality, 2l+1γ, good Zllg candidates.
- Tight photon ID & loose photon ISO.

Signal-background comparison

Zllg vs. Z+jets (real photon and fake photon)

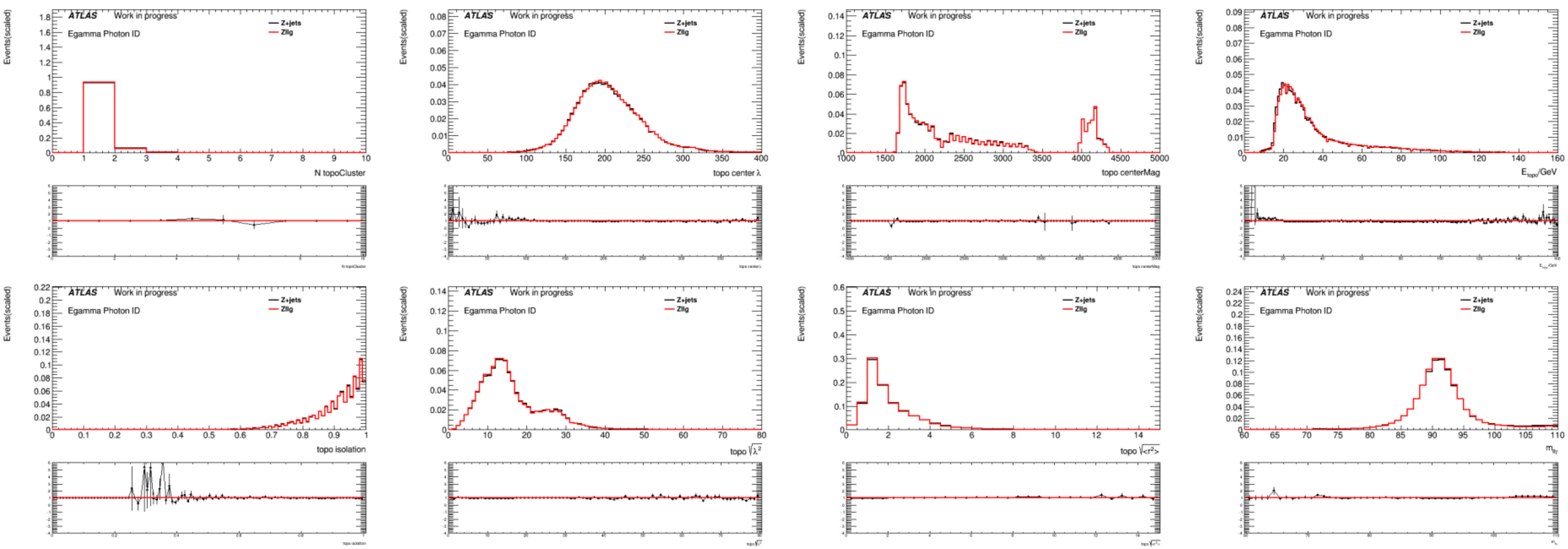
- Vars: E_{γ}, $N_{topoclus}$, E_{topo}, $\sqrt{\langle r^2 \rangle}$, $\sqrt{\langle \lambda^2 \rangle}$, center λ, centroid magnitude $\sqrt{x^2+y^2+z^2}$, isolation.
- ee+gamma, $7GeV < pT_{\gamma} < 15GeV$.



Signal-background comparison

Zllg vs. Z+jets (real photon and fake photon)

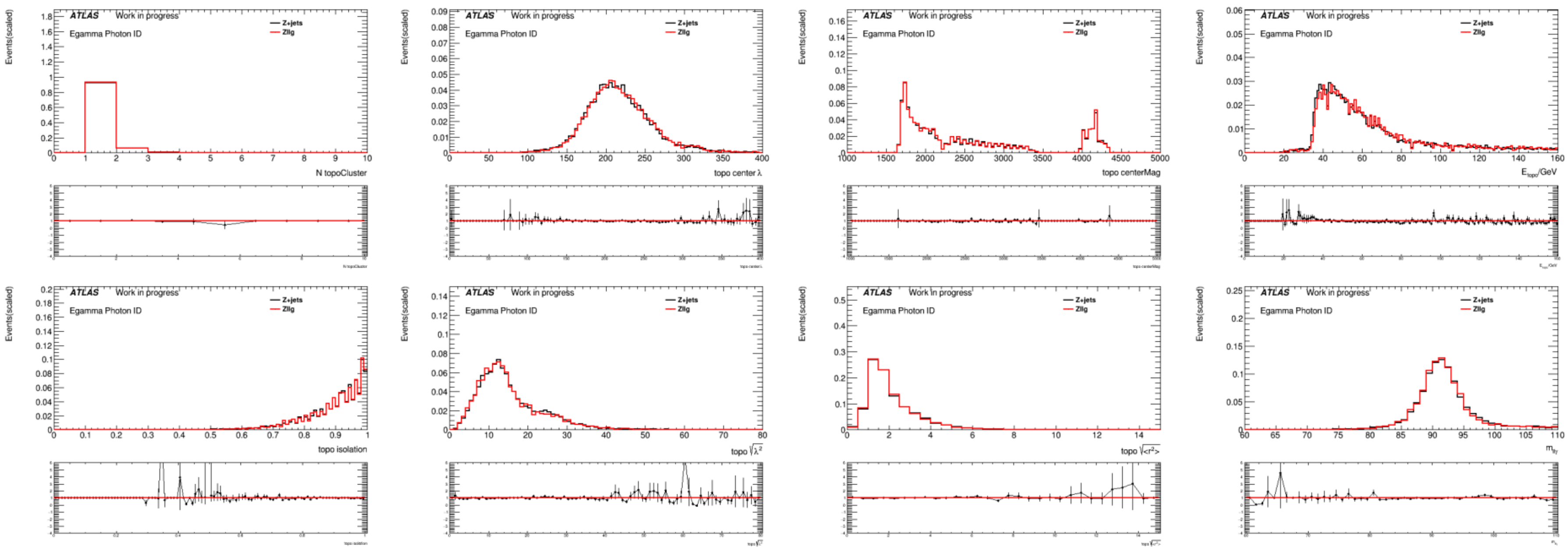
- Vars: E_γ, $N_{topoclus}$, E_{topo}, $\sqrt{\langle r^2 \rangle}$, $\sqrt{\langle \lambda^2 \rangle}$, center λ, centroid magnitude $\sqrt{x^2+y^2+z^2}$, isolation.
- ee+gamma, $15GeV < pT_\gamma < 35GeV$.



Signal-background comparison

Zllg vs. Z+jets (real photon and fake photon)

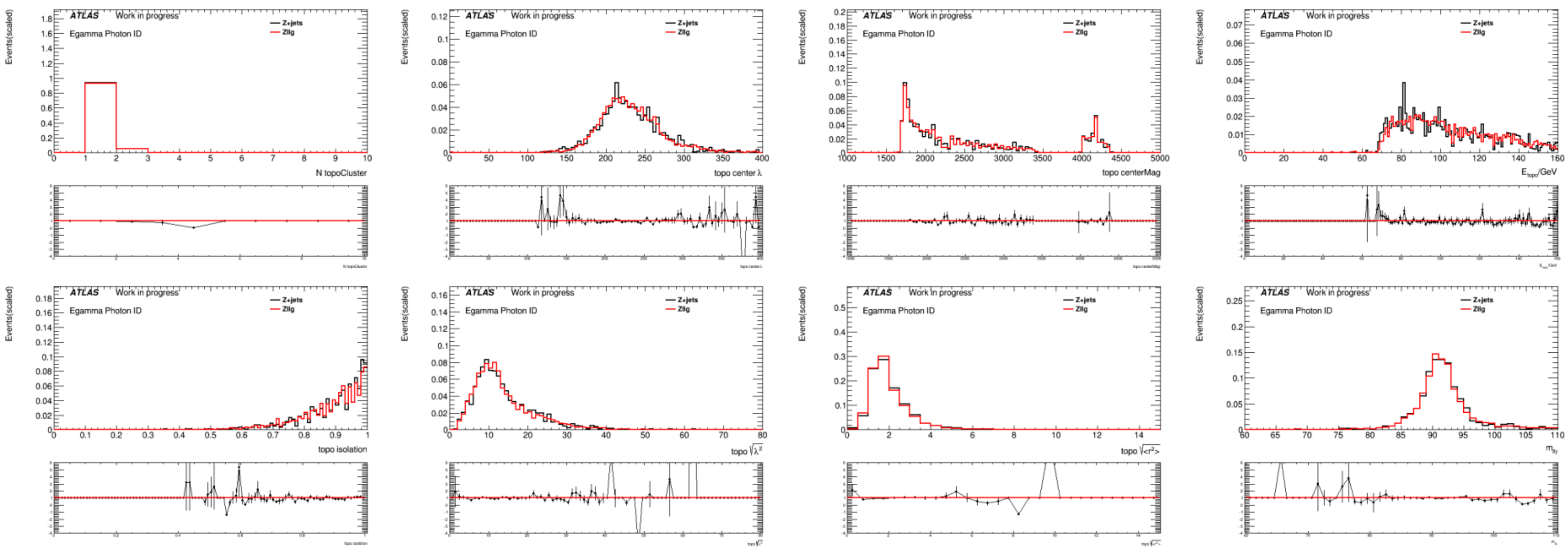
- Vars: E_γ, $N_{topoclus}$, E_{topo}, $\sqrt{\langle r^2 \rangle}$, $\sqrt{\langle \lambda^2 \rangle}$, center λ, centroid magnitude $\sqrt{x^2 + y^2 + z^2}$, isolation.
- ee+gamma, $35GeV < pT_\gamma < 70GeV$.



Signal-background comparison

Zllg vs. Z+jets (real photon and fake photon)

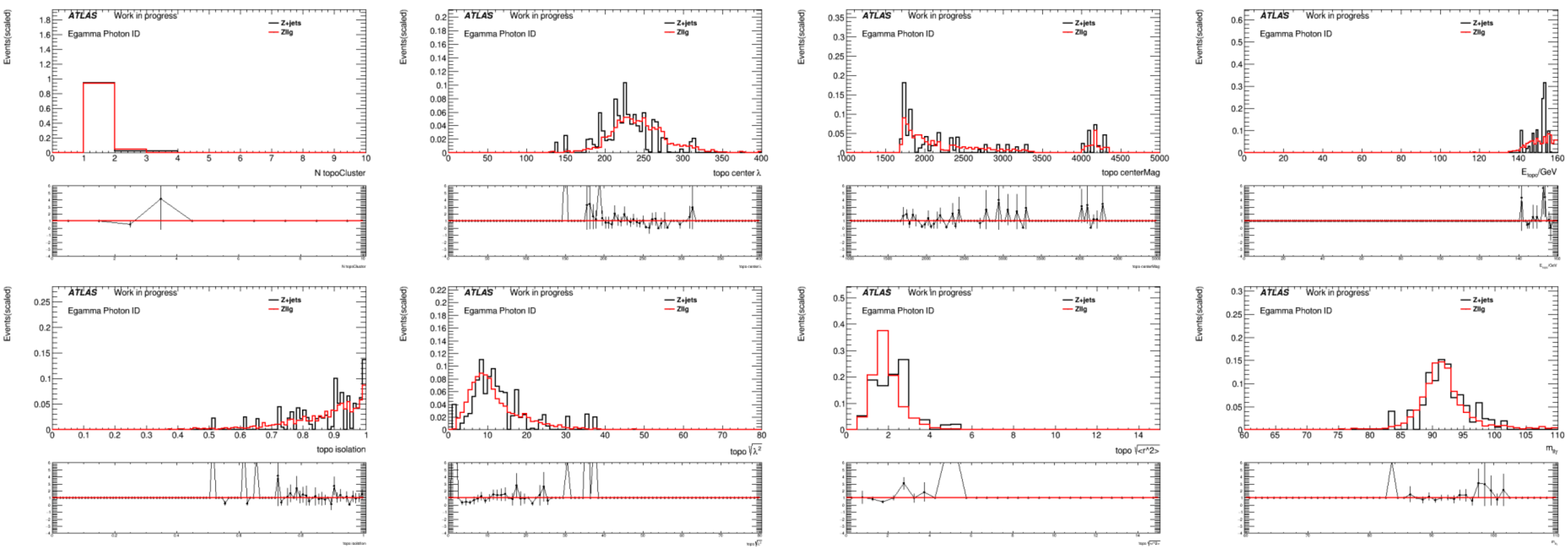
- Vars: E_γ, $N_{topoclus}$, E_{topo}, $\sqrt{\langle r^2 \rangle}$, $\sqrt{\langle \lambda^2 \rangle}$, center λ, centroid magnitude $\sqrt{x^2 + y^2 + z^2}$, isolation.
- ee+gamma, $70GeV < pT_\gamma < 140GeV$.



Signal-background comparison

Zllg vs. Z+jets (real photon and fake photon)

- Vars: E_γ, $N_{topoclus}$, E_{topo}, $\sqrt{\langle r^2 \rangle}$, $\sqrt{\langle \lambda^2 \rangle}$, center λ, centroid magnitude $\sqrt{x^2 + y^2 + z^2}$, isolation.
- ee+gamma, $pT_\gamma > 140GeV$.



Signal-background comparison

Separation power: $\langle S^2 \rangle = \frac{1}{2}\int \frac{(\hat{y}_s(y)-\hat{y}_b(y))^2}{\hat{y}_s(y)+\hat{y}_b(y)} dy = \frac{1}{2}\sum_{i=1}^{Nbins} \frac{(N_s(i)-N_b(i))^2}{N_s(i)+N_b(i)}$ (for histogram).

	[7, 15]	[15, 35]	[35, 70]	[70, 140]	[140, Ecms]
$N_{topoclus}$	0.000538	7.52E-05	0.000162	0.00084	0.010047
E_{topo}	0.004717	0.001908	0.00846	0.034274	0.466931
$\sqrt{\langle r^2 \rangle}$	0.002524	0.00029	0.000902	0.004866	0.107462
$\sqrt{\langle \lambda^2 \rangle}$	0.002256	0.000317	0.003505	0.013048	0.151165
center Mag	0.003665	0.00036	0.002055	0.015715	0.250503
center λ	0.002995	0.000358	0.003184	0.023818	0.184955
isolation	0.003163	0.000348	0.003165	0.01699	0.21177

Statistics for $pT_\gamma > 70$ MC is low, so large $\langle S^2 \rangle$ doesn't mean better separation power for high pT.

Conclusion

Cutflow cross check:

- No huge mismatch. Or if you have any recommended reference?

Signal-background separation power:

- No great power after tightID+looseISO.
- Loose some requirement, like Zllg cut?

Next step: need some suggestions.

- Consider other topo-cluster variables or second topo-cluster?
- Ask for more statistics and period 2015-2017? (now only 2018 used).
- Consider MVA?

Variables recorded

```
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_ENG_POS;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_FIRST_ENG_DENS;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_ISOLATION;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_N_BAD_CELLS;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_SECOND_LAMBDA;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_SECOND_R;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_AVG_LAR_Q;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_AVG_TILE_Q;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_BADLARQ_FRAC;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_CENTER_LAMBDA;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_CENTER_MAG;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_EM_PROBABILITY;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_ENG_BAD_CELLS;
vector<float>   *NewSwPhotonsAuxDyn_TOPOCLUS_ENG_FRAC_MAX;
```

Now only first topo-cluster is considered.

Other topo-cluster variables needs new derivation and jira tickets

backup

Cutflow cross check

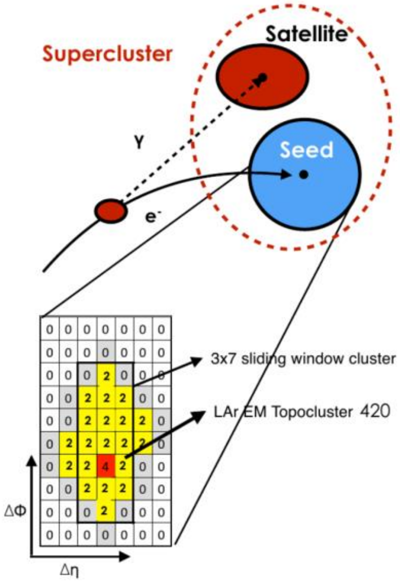
	mc16e (old)		mc16d (new)	
ALL	835000	/	492000	/
GRL	835000	100.0%	492000	100.0%
trigger	434906	52.1%	239987	48.8%
PV	434905	100.0%	239987	100.0%
EQ	434905	100.0%	239987	100.0%
2 leptons	416071	95.7%	229639	95.7%
1 photon	416040	100.0%	229630	100.0%
selected leptons	102110	24.5%	55069	24.0%
selected photon	87203	85.4%	46894	85.2%
Zllg candidate	21364	10.4%	7652	16.3%
total		2.6%		1.6%

Extracted from histo_cutflow in output file, good photon + good leptons + Zllg candidate requirement.

Introduction

Investigate photon topo-cluster performance, see if they could provide additional discrimination power in photonID.

- Format topo-cluster in LAr Ecal:
  - Find seed: cell significance > 4σ
  - Scanning neighbor cells: add significance > 2σ cell +neighbor
  - Merge 2 clusters if they grow into each other.



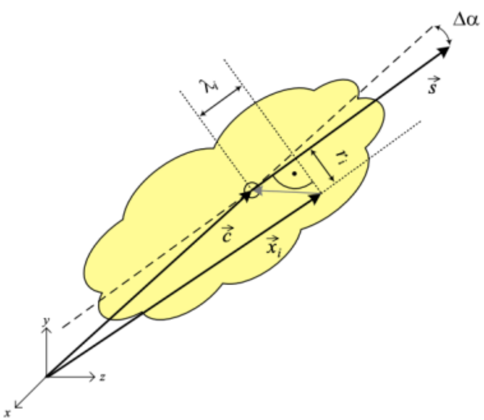
- Approach
  - Tyler has finished the comparison in single photon process.
  - Repeat it in Zllg process, for MC modelling.
  - Do MC/data comparison with Zllg MC and data.
  - Do signal/bkg comparison with Zllg and Z+jets.

[Tyler's report about topo-cluster variables](#)
[Data/MC in single photon](#)

Introduction

Cluster variables

- y_topoCluster0_Epos: Total positive Energy of this cluster.
- y_topoCluster0_secondR: Semi-major axis in width for the leading topo cluster associated to each photon.
- y_topoCluster0_secondLambda: Semi-major axis in depth for the leading topo cluster associated to each photon.
- y_topoCluster0_centerMag: Cluster centroid magnitude $\sqrt{x^2 + y^2 + z^2}$
- y_topoCluster0_centerLambda:Depth of leading topo cluster at its centroid.
- y_topoCluster0_isolation: Energy weighted fraction of non-clustered perimeter cells



$\vec{c}$ *centre of gravity of cluster, measured from the nominal vertex ($x = 0, y = 0, z = 0$) in ATLAS*

$\vec{x}_i$ *geometrical centre of a calorimeter cell in the cluster, measured from the nominal detector centre of ATLAS*

$\vec{s}$ *particle direction of flight (shower axis)*

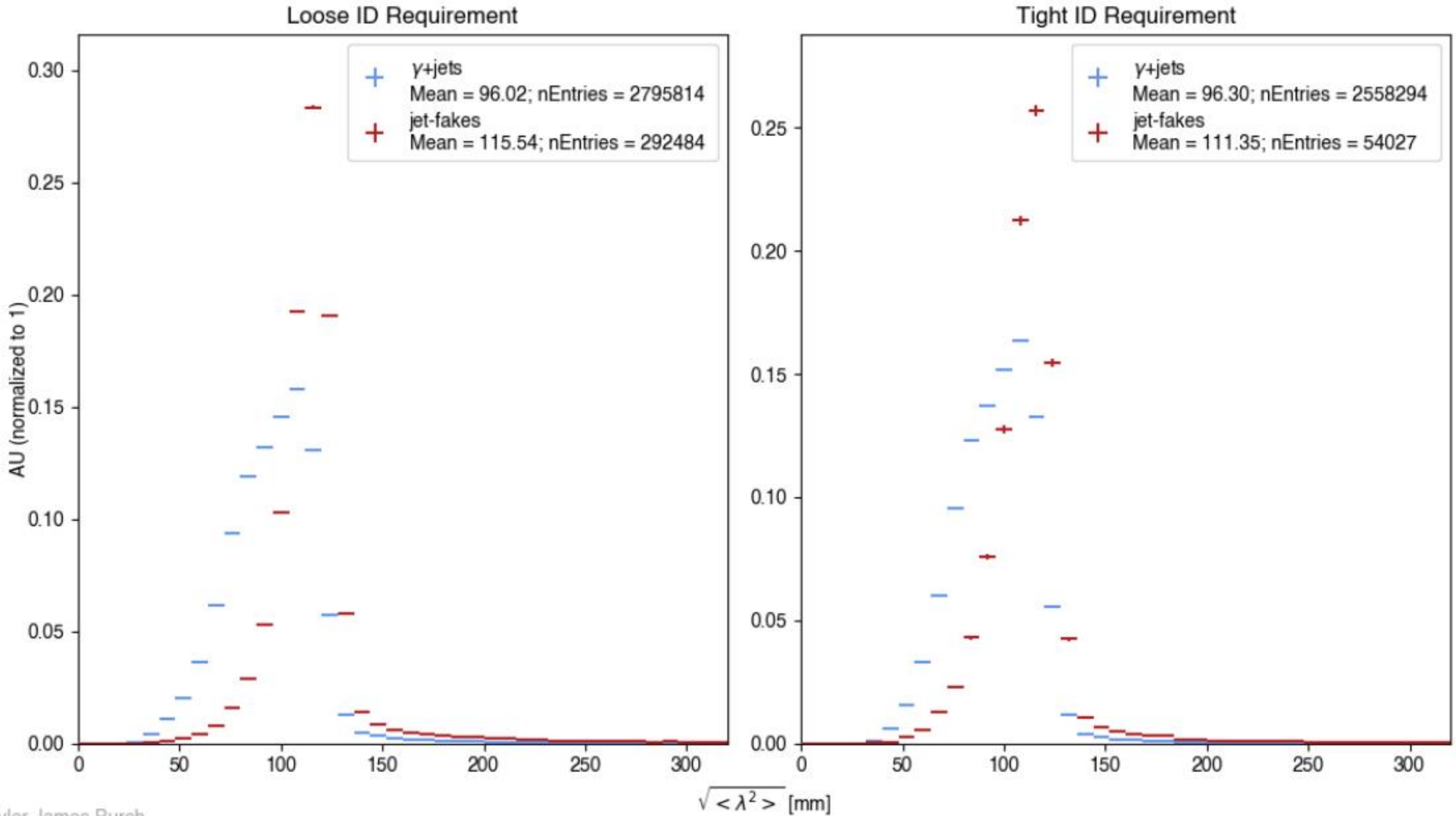
$\Delta\alpha$ *angular distance $\Delta\alpha = \angle(\vec{c}, \vec{s})$ between cluster centre of gravity and shower axis $\vec{s}$*

λ_i *distance of cell at $\vec{x}_i$ from the cluster centre of gravity measured along shower axis $\vec{s}$ ($\lambda_i < 0$ is possible)*

r_i *radial (shortest) distance of cell at $\vec{x}_i$ from shower axis $\vec{s}$ ($r_i \geq 0$)*

Reference note

Cluster Shape Information, λ



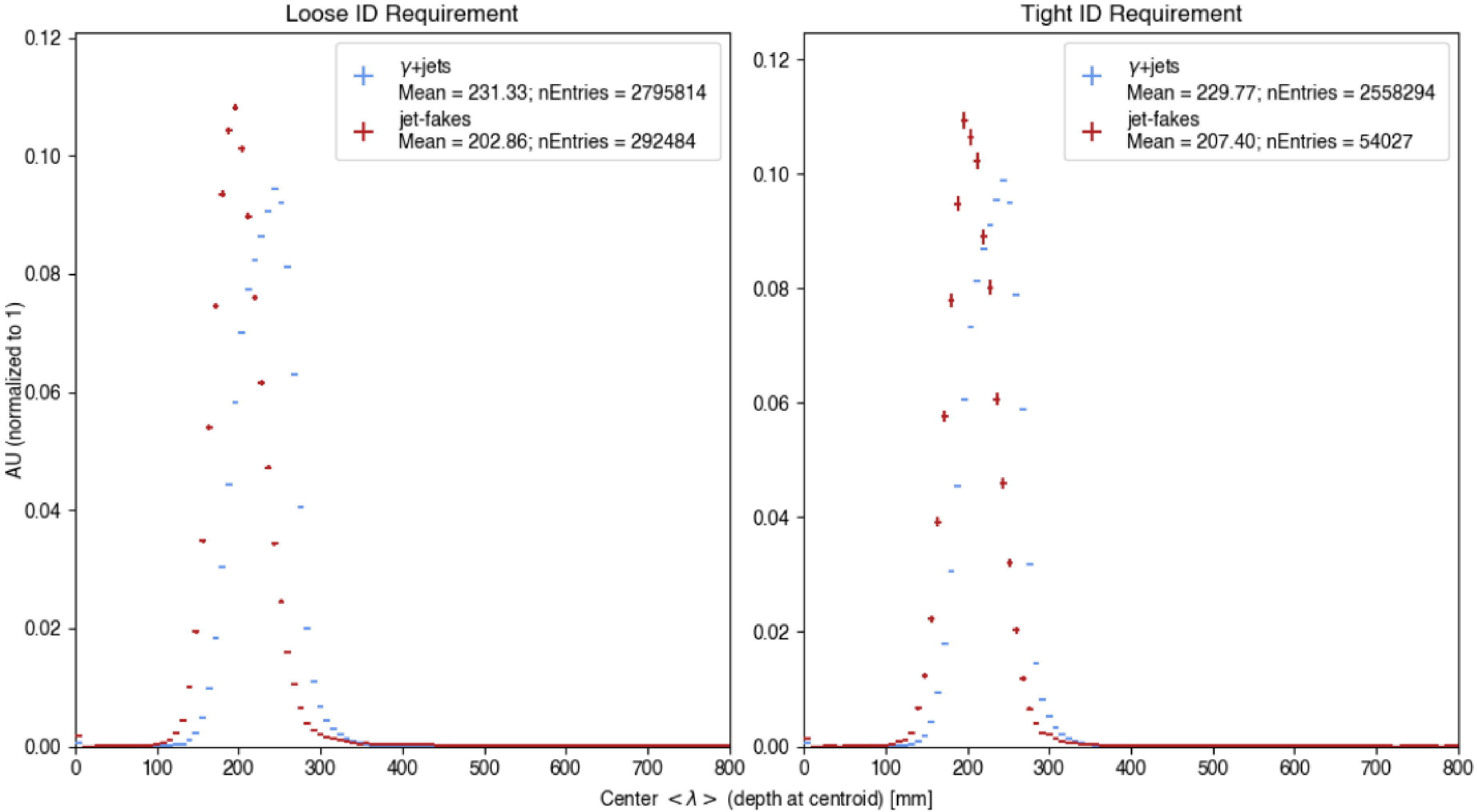
Tyler James Burch

Topo clusters associated with photons on average are less long in the calorimeter depth direction than jet-fakes

Argonne NATIONAL LABORATORY

Northern Illinois University

Cluster Shape Information, λ



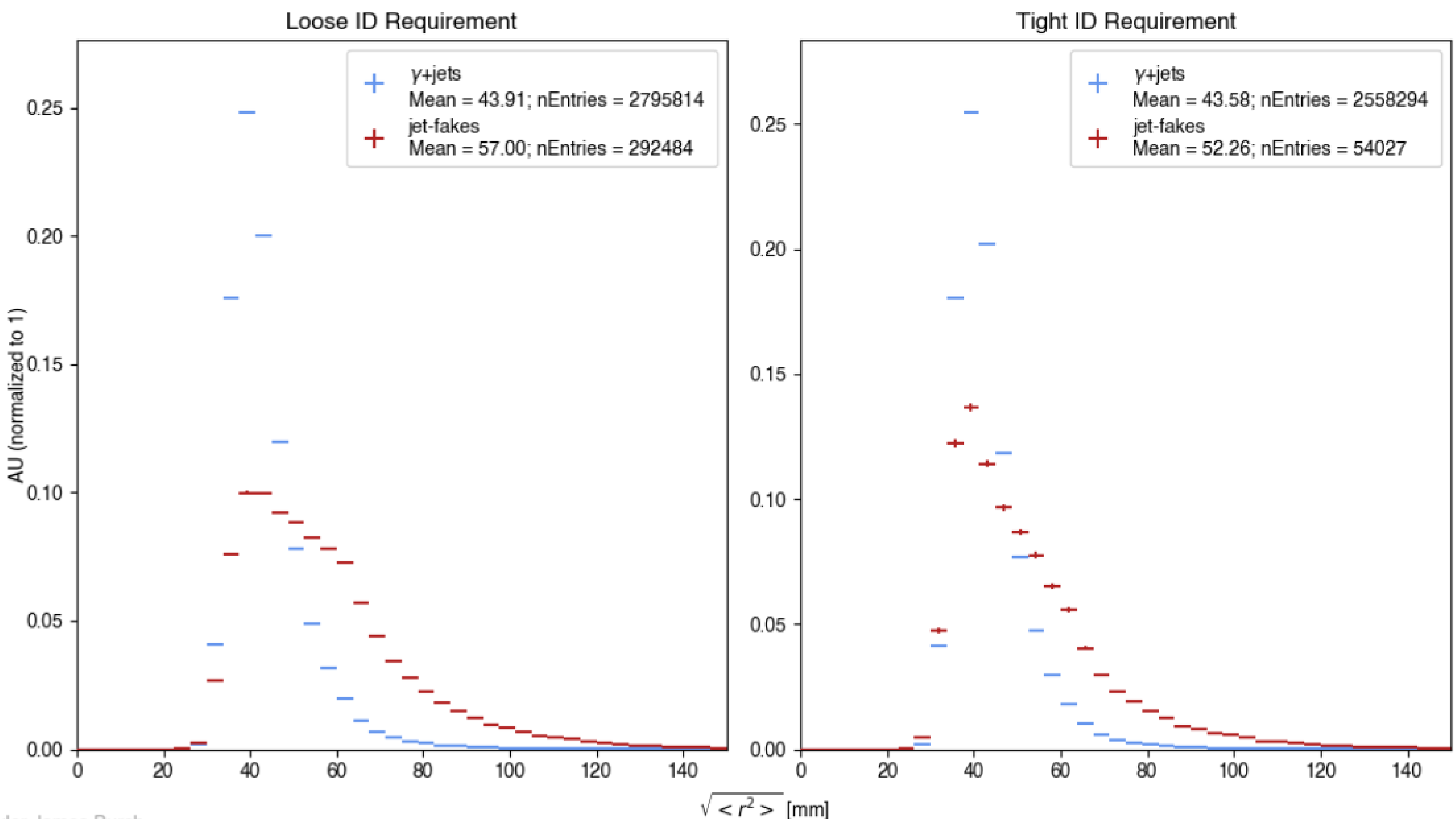
Tyler James Burch

Can also look at the depth at cluster centroid - photons are slightly deeper at the centroid than jet-fakes

Argonne NATIONAL LABORATORY

Northern Illinois University

Cluster Shape Information, R



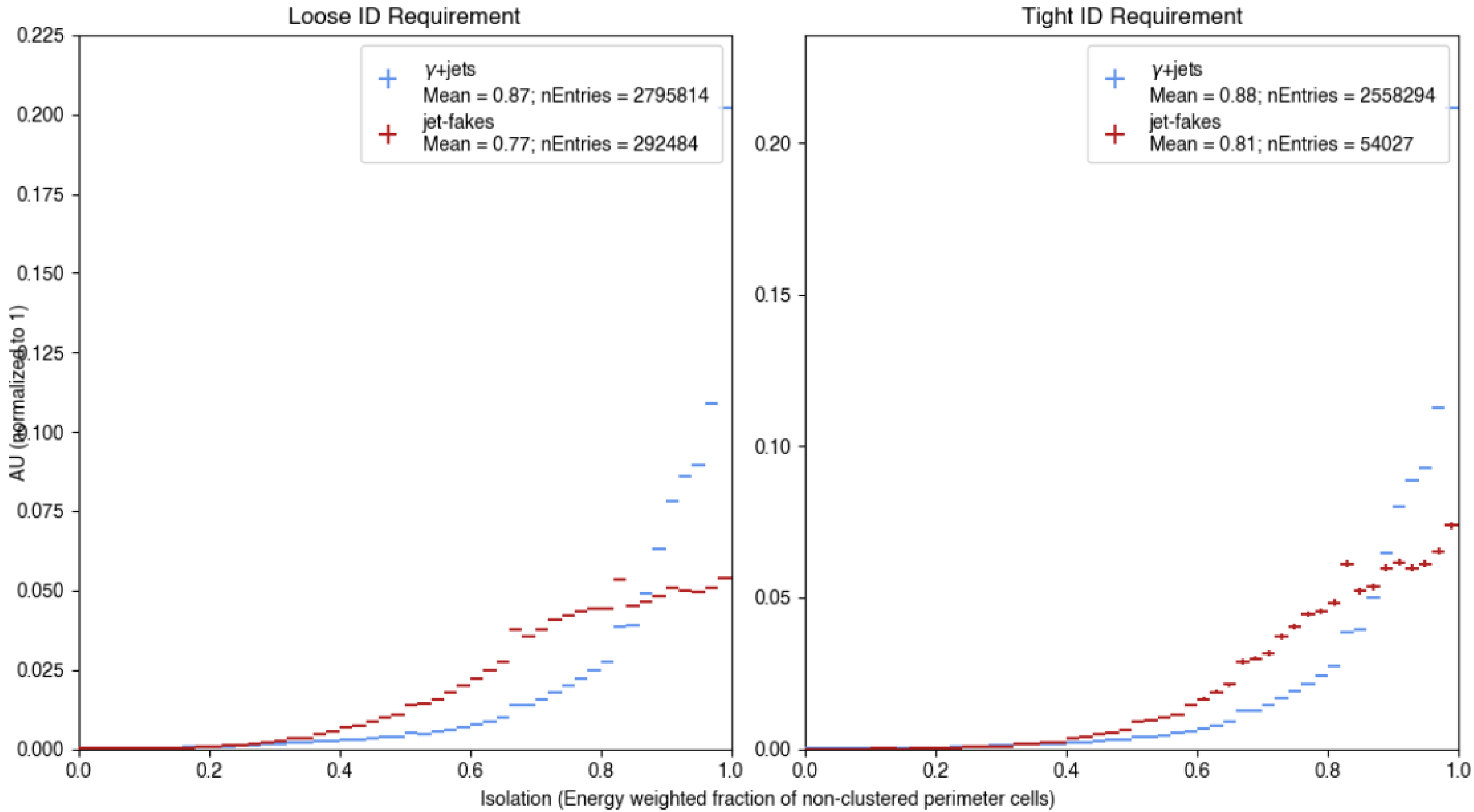
Tyler James Burch

Jet fakes are notably wider than photons. Difference is slightly less once applying tight ID requirement, but still some separation

Argonne NATIONAL LABORATORY

Northern Illinois University

Isolation



Tyler James Burch

Good separation observed… but should do further looking into correlation with calo/track isolation

Argonne NATIONAL LABORATORY

Northern Illinois University