https://indico.bnl.gov/event/31398/contributions/119324/attachments/67866/116660/JSTGupdate_1_28_26_JustinBennett_GammaJets.pdf

γ-jet Calibration

Justin Bennett

01.28.26

Overview

Project code

Analysis Sample

- **MC** → Run-28 PYTHIA8 γ+jet (PhotonJet10 + PhotonJet20 – merging steps here), p+p 200 GeV, no pileup

Discussion Points

- **Analysis definitions** → reco/truth event selection and matching framework
- **Core observation** → identify where the low-x_J excess shows up using truth-tagging comparisons
- **Mechanism diagnosis** → quantify and classify the recoil-jet failure modes that generate the excess
- **Action items** → stress-test selections and lay out the path to data + in-situ calibration

Analysis Notes

- **First JSTG presentation from 1.07.26 here**
- **Using** PhotonClusterBuilder for SS variables, and isolation calculation used for photon selection
- **JES calibration** from JetCalib
- **Jet $p_T \geqq 5$ GeV and $\Delta\varphi$(leading γ, leading jet) $\geqq \pi/2$**
- **E_T^γ = [10, 12, 14, 16, 18, 20, 22, 24, 26, 35] GeV** → match PPG12

γ+Jet Event Definition and Matching: Reco, Truth, and Reco↔Truth

Reco-level (data/sim)

Pick event-leading photon that is Tight $\wedge$ Isolated (PPG12 cuts used verbatim)

- E_T^{γ} (PPG12) = [10, 12, 14, 16, 18, 20, 22, 24, 26, 35] GeV
- Details on iso criteria *here*
- Details on SS criteria *here*

Reco-jet match ($R \in \{0.2, 0.4\}$)

- $p_T^{\text{jet, reco}} \geq$ 5 GeV,
- $|\eta_{jet}^{reco}| < 1.1 - R$,
- $|\Delta\varphi(\gamma_{reco}, jet_{reco})| \geq \pi/2$

Defines the reco-selected γ+jet

Truth-level (sim)

Truth signal photon (γ_{truth}):

→ For *all PDG=22* truth particles

- $|\eta_{\gamma}^{truth}| < 0.7$
- Filter for direct or fragmentation γ's via HepMC ancestry (details)
- $E_T^{\text{iso, truth}} < 4$ GeV (details)

Truth recoil jet combination:

- $\mathbf{p_T^{\text{jet, truth}} \geq 5}$ **GeV**
- $\mathbf{|\eta_{jet}^{truth}| < 1.1 - R}$
- $\mathbf{|\Delta\varphi(\gamma_{truth}, jet_{truth}) \geq \pi/2}$

Defines the truth-selected γ+jet

Truth-Reco matching (sim)

$\gamma_{truth} \leftrightarrow \gamma_{reco}$

- $\mathbf{\Delta R(\gamma_{truth}, \gamma_{reco}) < 0.05}$
- CaloRawClusterEval "best match" condition (γ_{truth} is top energy contributor to γ_{reco})

For a jets passing truth/reco level away-side topology:

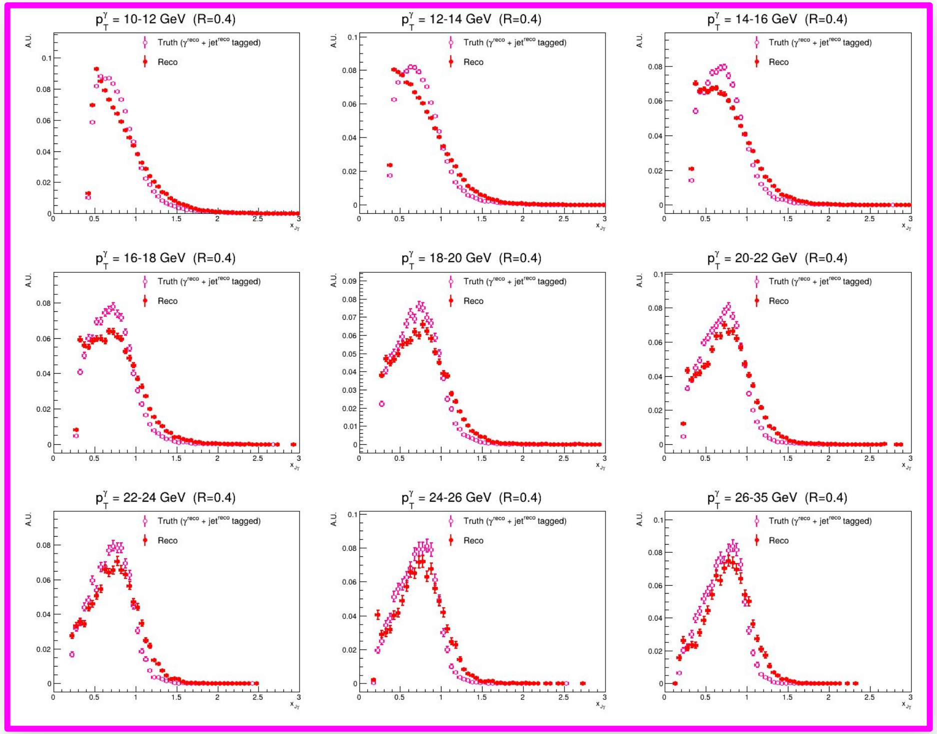
- Accept matched pair iff $\Delta R(jet_{reco}, jet_{truth}) < 0.3$

Defines the matched-pair subset used for truth-tagging

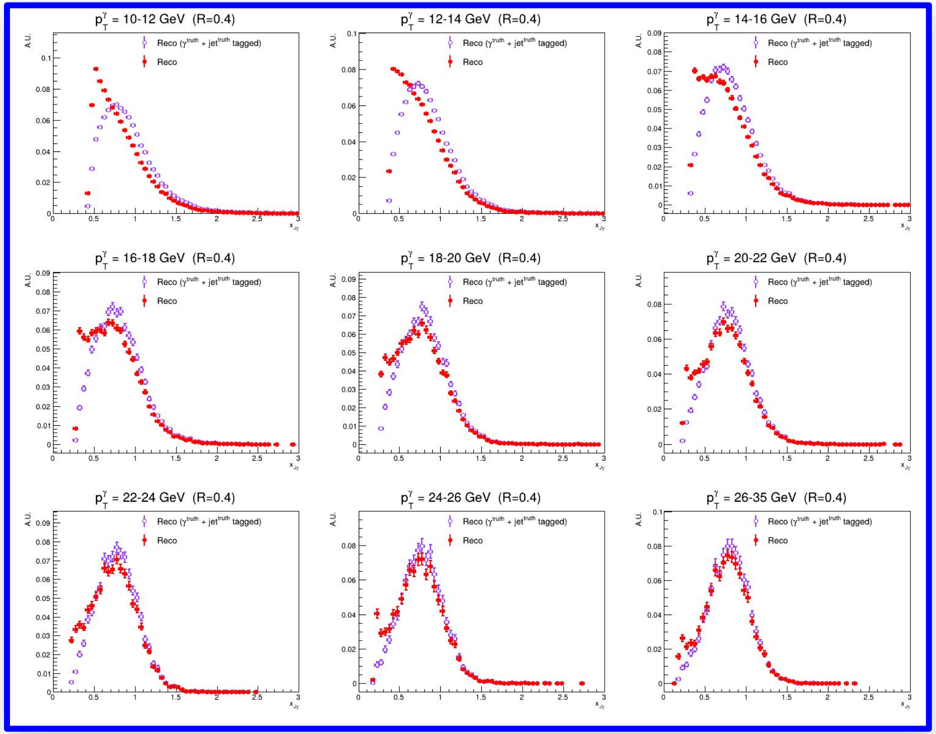
Reco–Truth Cross-Conditioning: Conditioned $x_{J\gamma}$ vs Reco Baseline

Compare truth (reco) $x_{J\gamma}$ in truth (reco) selected samples → *same* reco baseline in both panels

Truth x_J (conditioned) vs **Reco x_J** (baseline)



Reco x_J (conditioned) vs **Reco x_J** (baseline)



Pink = Truth-in-Reco ($x_{J\gamma}^{truth}$ in the *reco-selected* sample)

Blue = Reco-in-Truth ($x_{J\gamma}^{reco}$ in the *truth-selected* sample)

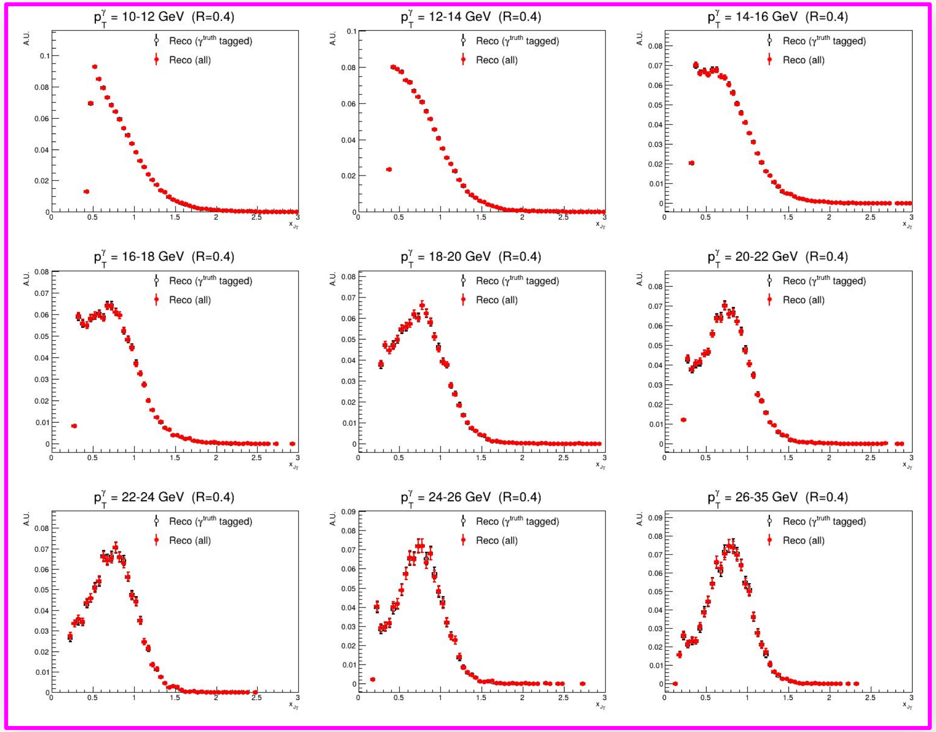
Truth Tagging: Photon-Only vs Photon+Jet (Localizes the Low-x_J Excess)

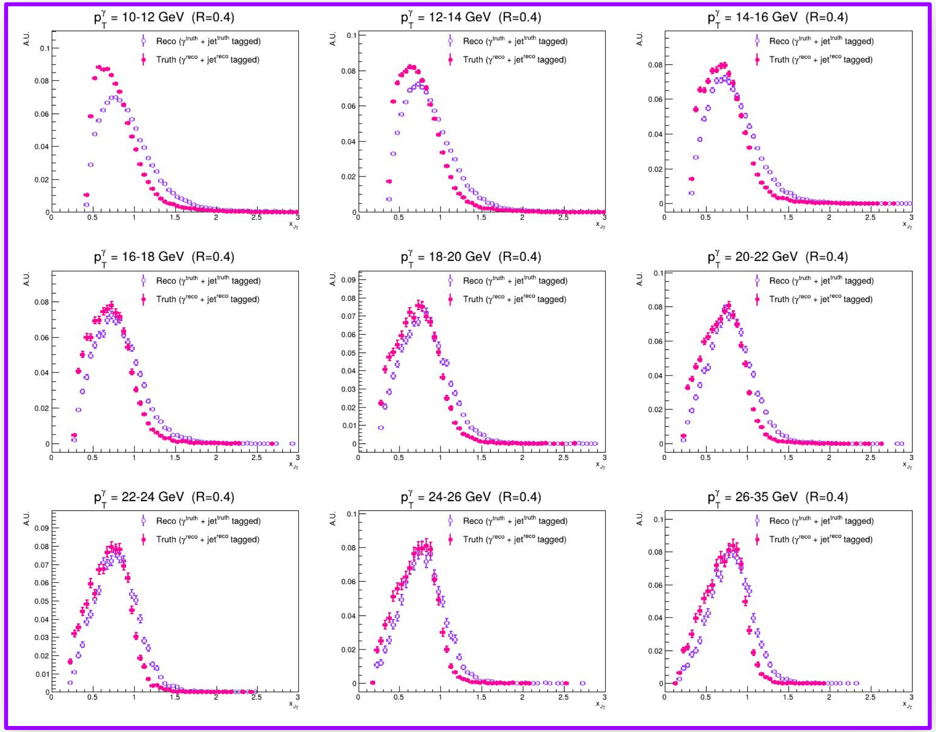
Photon-only tag

reco x_J unconditional (**red**) vs reco x_J, truth-γ tagged (**black**)

Photon+jet tag

truth reference (**red**) vs reco xJ, **truth-γ+jet** tagged (**blue**)

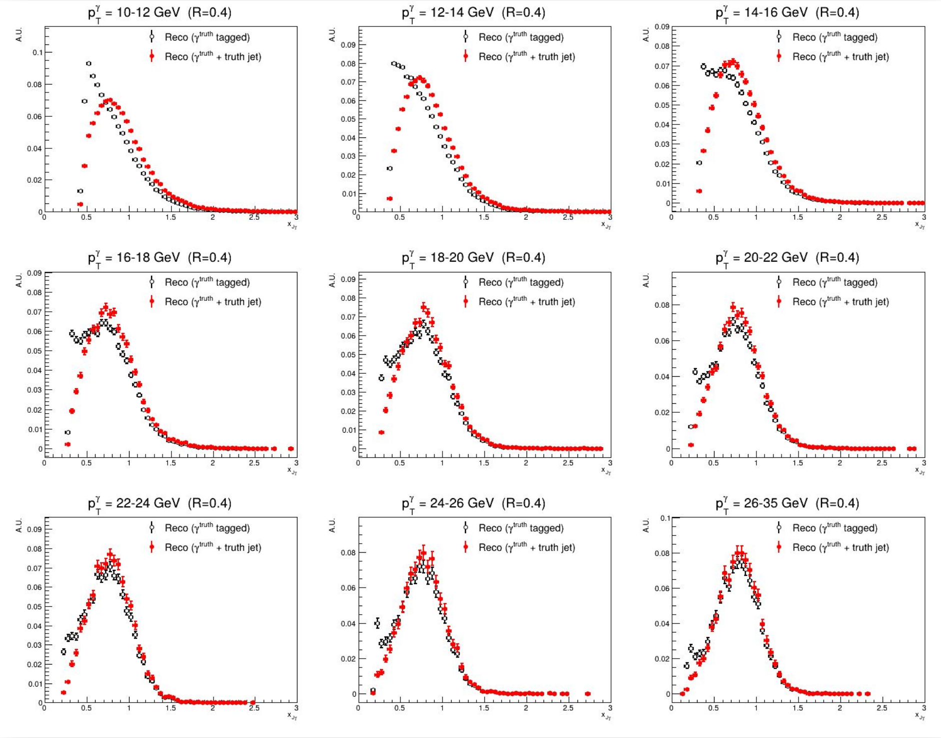




Observation → *truth shape matches reco* **when conditioning on photon+jet**, unconditioned reco matches reco with γ^{truth} match

Forcing Correct Recoil Removes Low $x_{J\gamma}$ Noise

Reco-only – *conditioned* on **truth photon vs photon + recoil jet**



Only the truth-jet condition removes the low-x_J population → the problem is recoil-jet identity (not γ selection)

Lead Recoil-Jet Truth-Match Efficiency vs $p_T^{\gamma,\,truth}$

$\varepsilon_{lead}(p_T^{\gamma,\,truth})$ = ***P***(the *selected reco recoil* jet **matches** the *truth-leading away-side recoil jet* | truth recoil jet exists)

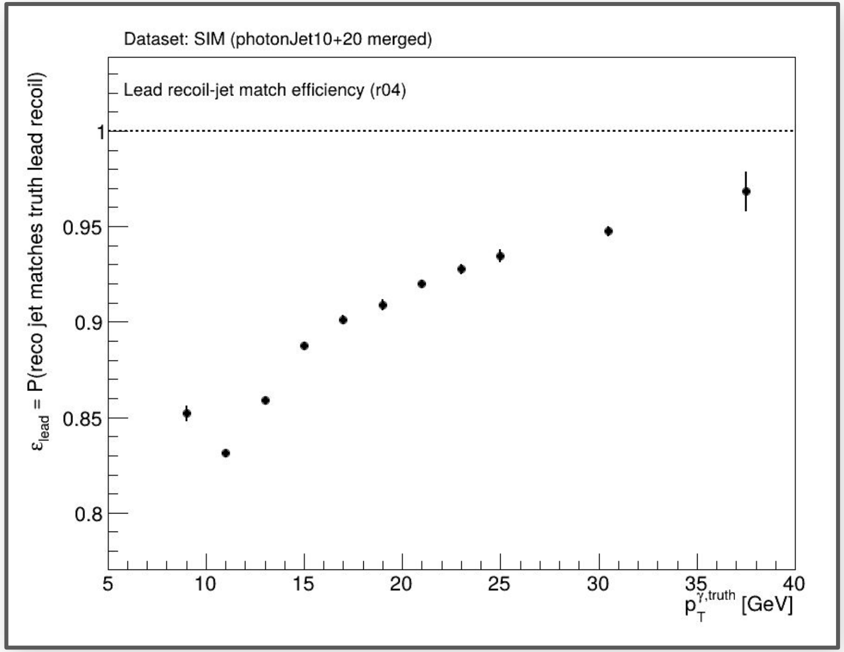
What this measures → how often the selected *reco* recoil jet is the true leading away-side recoil jet

- **Denominator** → truth-leading away-side recoil jet exists
- **Numerator** → Selected reco recoil jet matches it ($\Delta R < 0.3$)

Failure = denominator fills but numerator doesn't (we found truth recoil but not reco) → **set two categories:**

A) **MissA** → truth-leading recoil jet is *reconstructed* (matched to some reco jet) *but* we selected the wrong reco jet

B) **MissB** → the truth-leading recoil jet has no reco match at all → true reconstruction/matching failure.



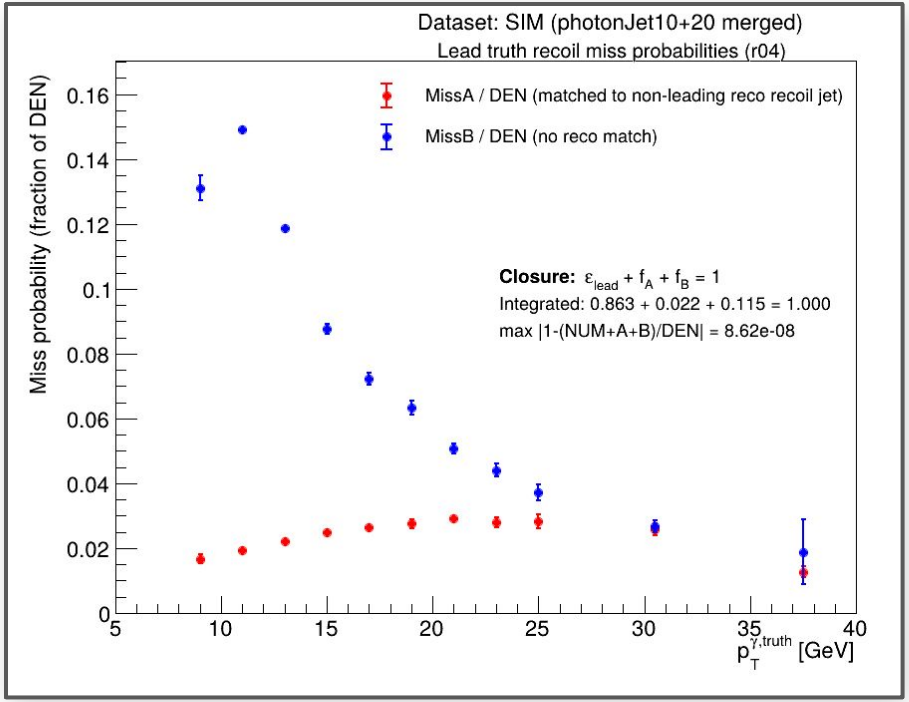
Understanding Failure Categories

Start from efficiency definition → DEN (denominator) = events where a truth-leading away-side recoil jet exists

- **When denominator fills** *but numerator doesn't* — we failed to identify the true recoil jet

1. **MissA/DEN** → *wrong-jet selection*
   Truth-leading recoil jet *is reconstructed* (matches *some* reco jet) but we *did not select it* as the analysis recoil jet

2. **MissB/Den** → *true reconstruction/matching miss*
   Truth-leading recoil jet has *no reco match at all*



At low p_T → *dominant failure* is **MissB** (no reco match) while wrong-jet selection (**MissA**) stays a small, roughly constant few-percent effect

Closure = MissA and MissB are a complete, mutually exclusive breakdown of all efficiency failures. *(more info)*

Leading Recoil Diagnostics – $p_T^{jet, reco}$ vs $p_T^{jet, truth}$

For events with a *truth-leading away-side* recoil jet → **plot <$p_T^{jet, reco}$>** of the *analysis-selected recoil* jet vs p_T of the *truth recoil jet* — split into **matched (eff numerator)**/**wrong-jet (MissA)**/**no-match cases (missB)**

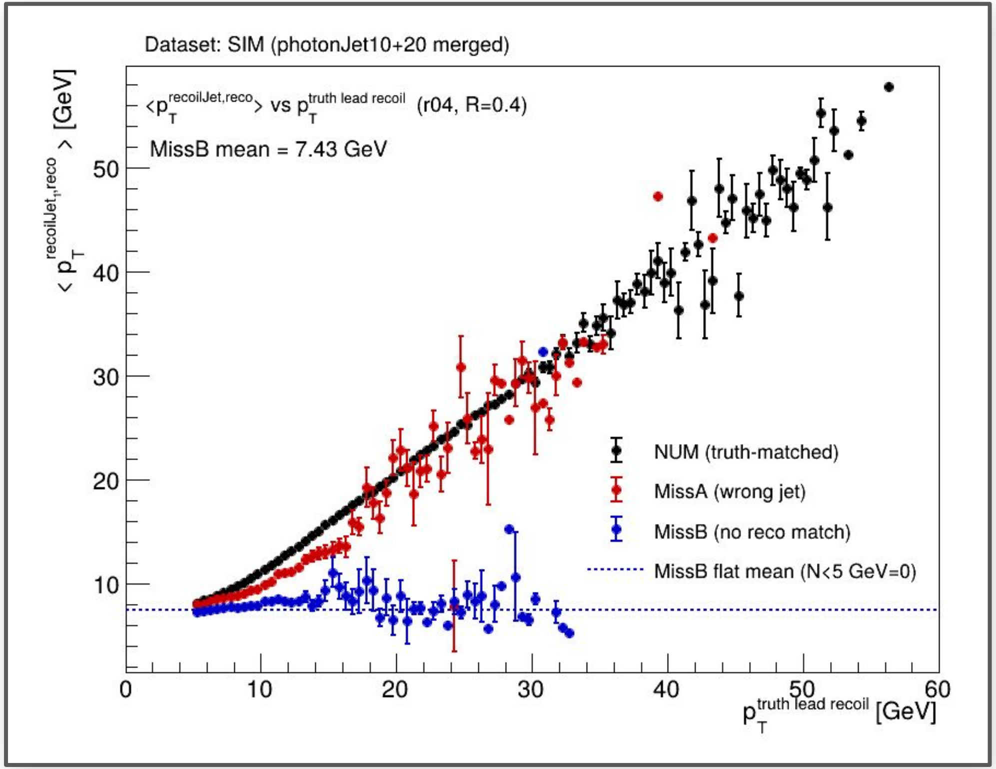
Black (matched) → analysis-selected recoil jet matches truth-leading recoil jet ($\Delta R < 0.3$)

Red (wrong jet) → truth-leading recoil jet is reconstructed (has a reco match) — but a different reco jet is the one the analysis selects

Blue (no match) → truth-leading recoil jet has no reco match — the *analysis-selected recoil* jet is a (soft) substitute

MissB creates 'soft substitute recoils'

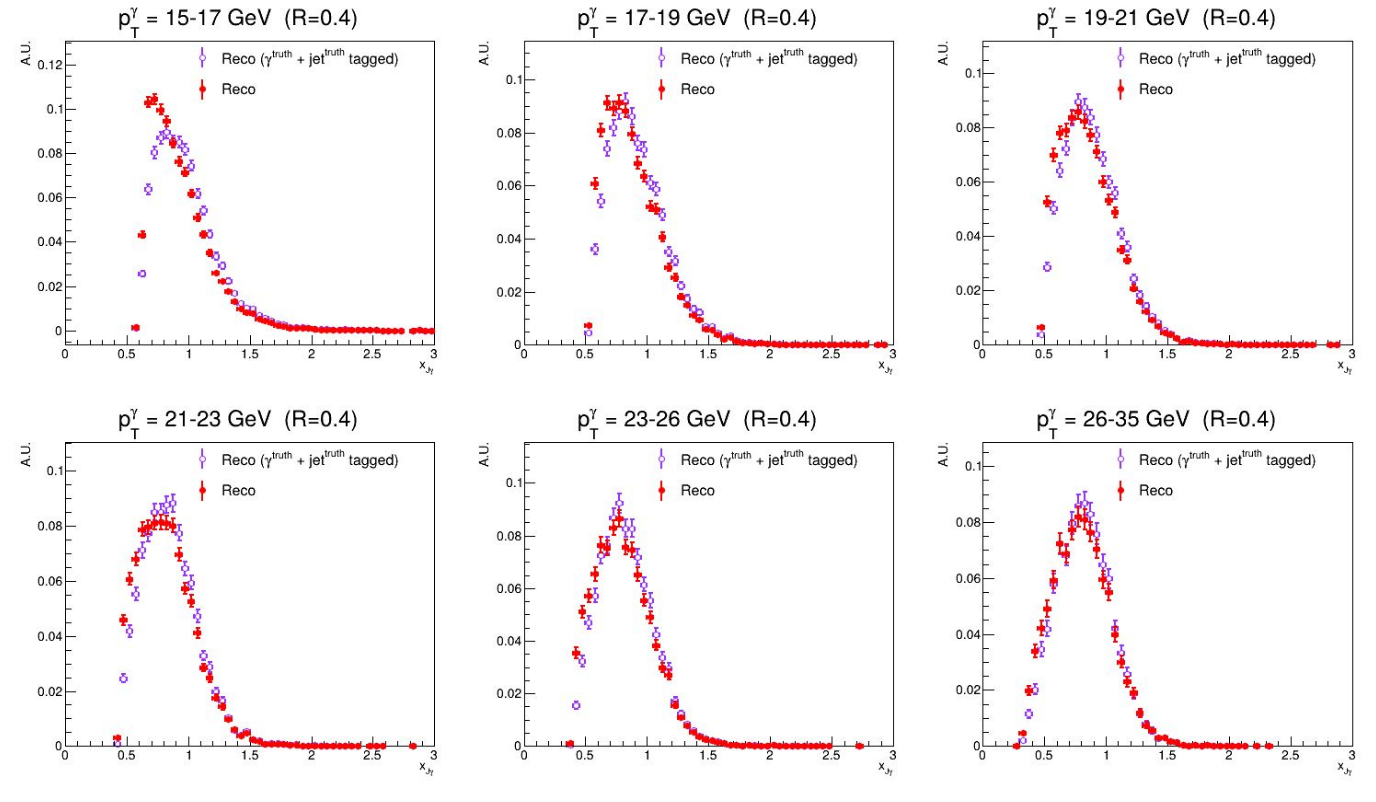
- When the true recoil jet is missing in reco – analysis *still picks an* away-side jet – but it saturates near a low p_T floor (~7 GeV) rather than tracking the true recoil – *driving low x_J pileup*



Changing Binning & Jet p_T Cut to Remove Soft Substitutes

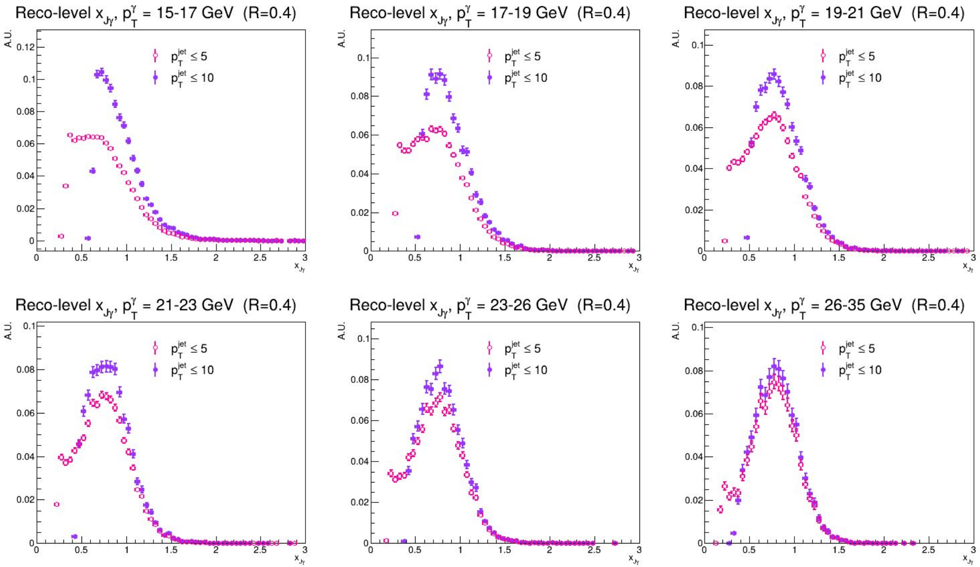
With soft substitute recoils piling up at ~7 GeV → can bump up jet p_T floor to 10 GeV and shift photon binning wrt this

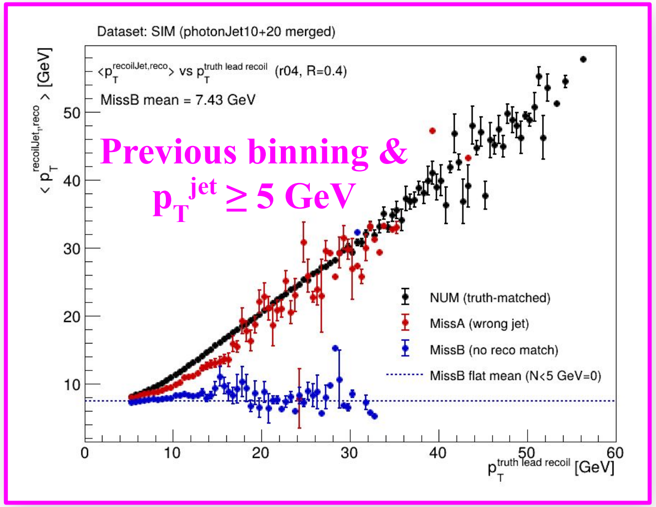
- **p_T^{γ} = 15-17, 17-19, 19-21, 21-23, 23-26, 26-35 GeV**

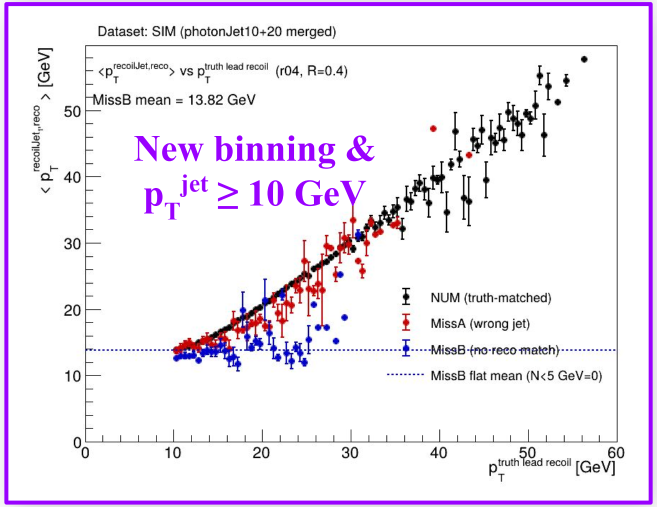


Soft Substitute Jets Still Occur

- **Jet $p_{T,\,min}$** ↑ → fewer soft substitute recoils (MissB) → reduced low-x_J excess
- **Mitigates the symptom** but doesn't fully remove the mechanism
- **Smaller R** → *larger* effect – expected if driven by under-measuring/splitting of truth recoil (backup: radii overlay)





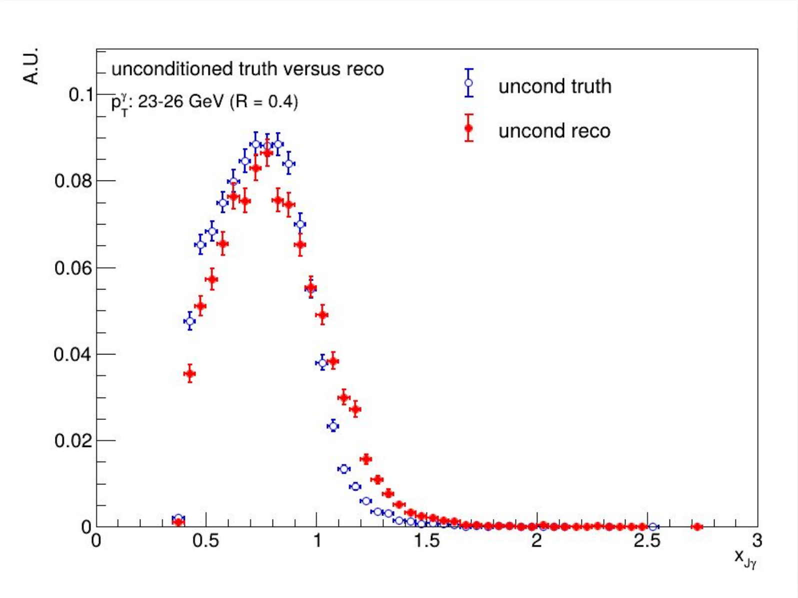


Next Steps

1. Inspect MissA vs MissB (and matched) event displays to see what the selected recoil jets look like in η–φ

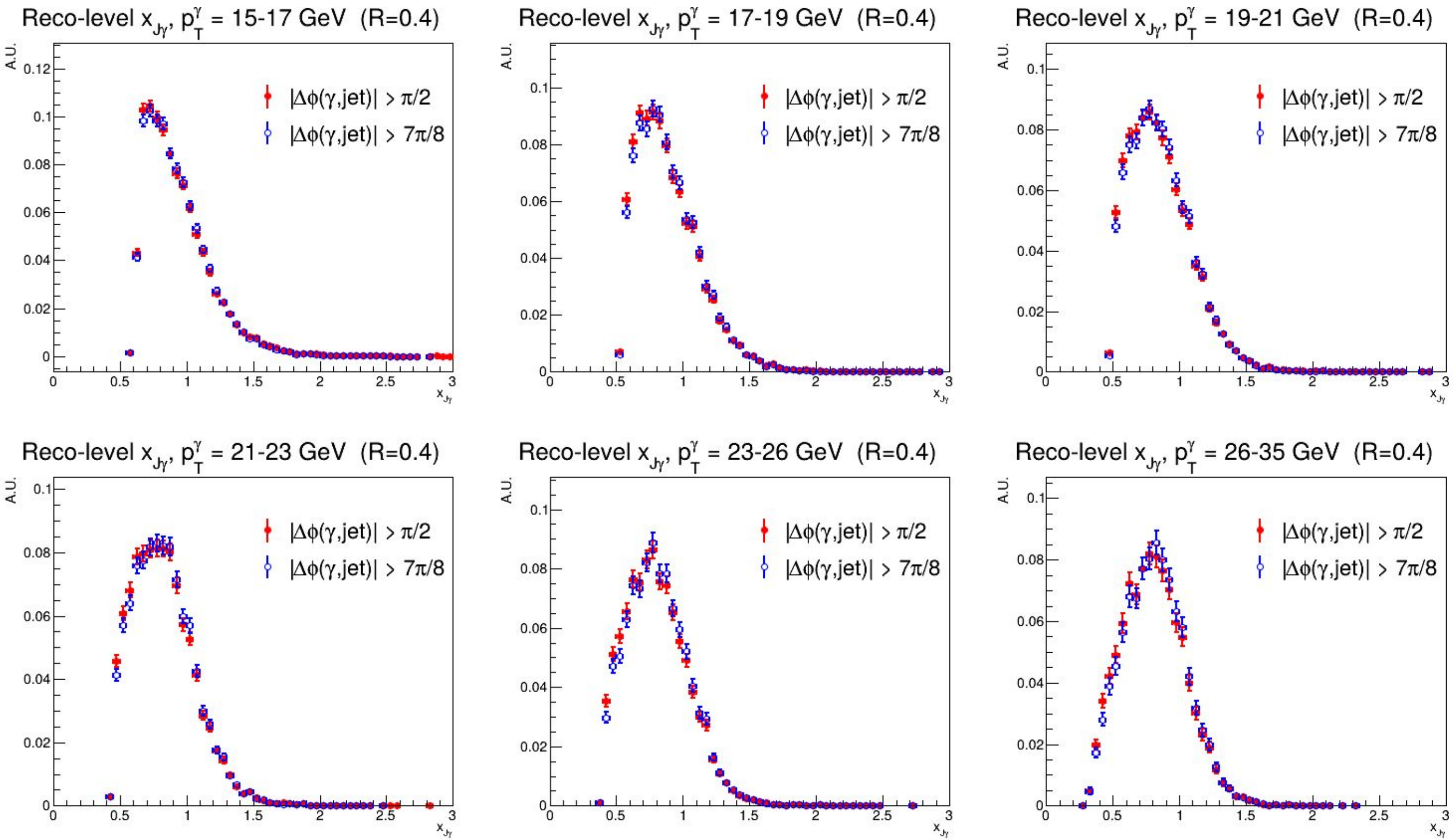
2. Measure reco xJ in pp data and compare directly to the SIM baseline

3. Proceed to in-situ calibration on data using the updated reco x_J distributions (Gaussian-fit per p_T bin → extract mean vs p_T)



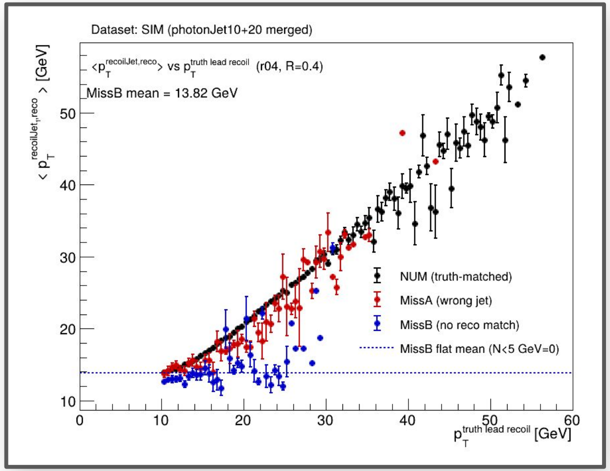
BACKUP

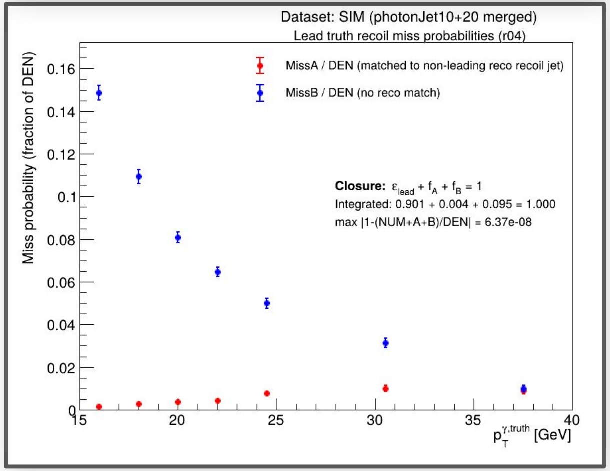
Comparing Back-To-Back Cuts (1)



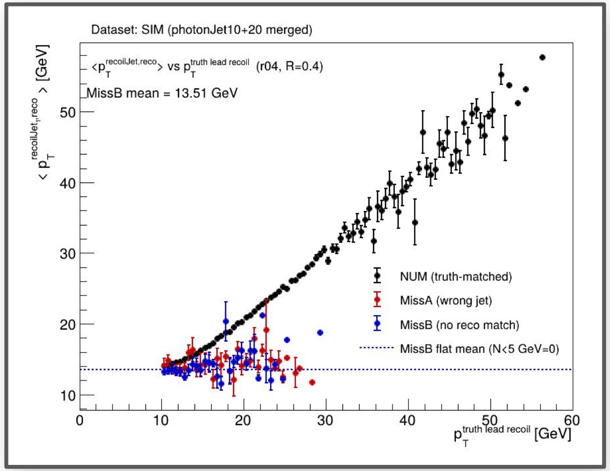
Variation of Back-to-Back Cut (2)

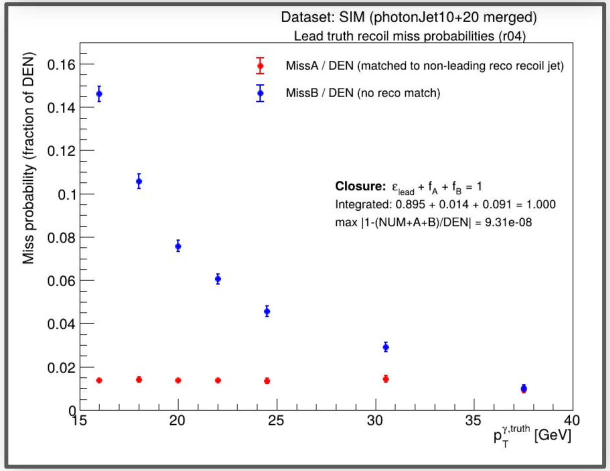
Back-to-back cut π/2



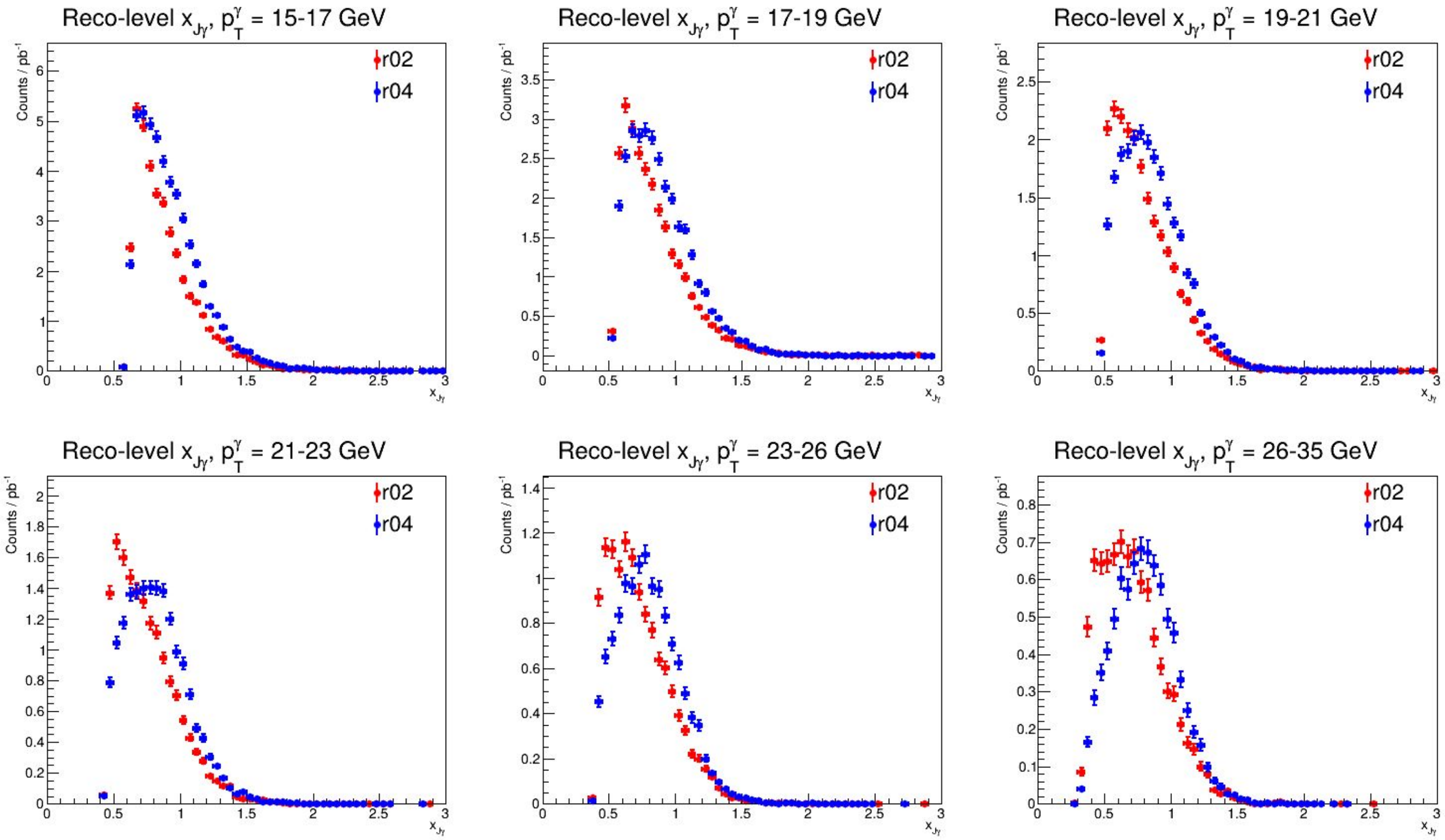


Back-to-back cut 7π/8



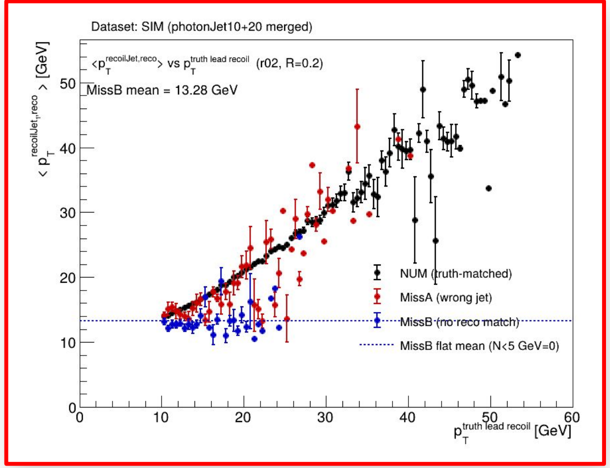


Comparing Radii Variants – R = 0.2 & R = 0.4 (1)

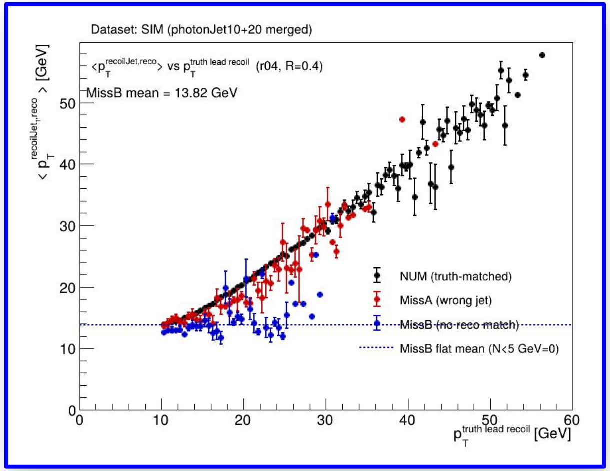


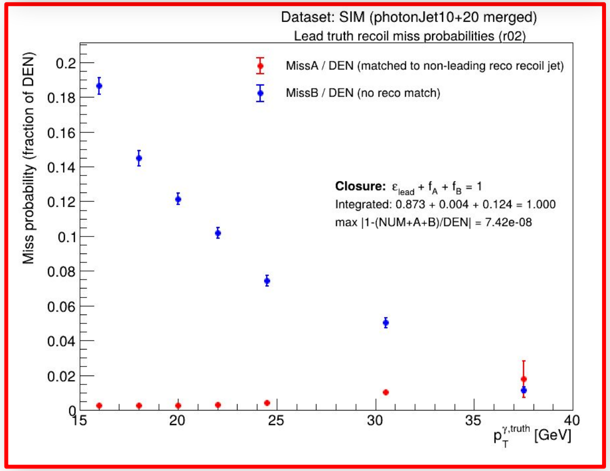
Comparing Radii Variants – R = 0.2 & R = 0.4 (2)

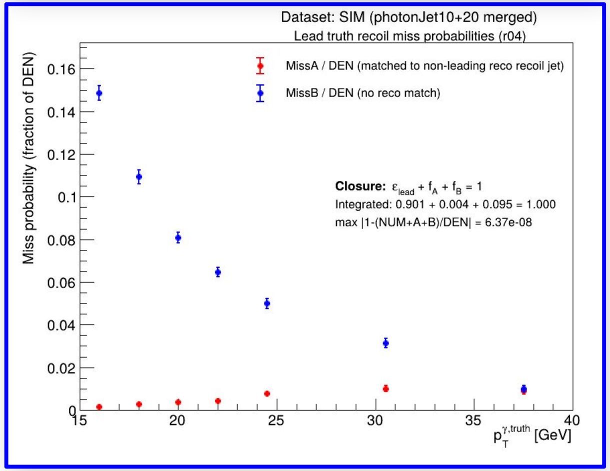
R = 0.2



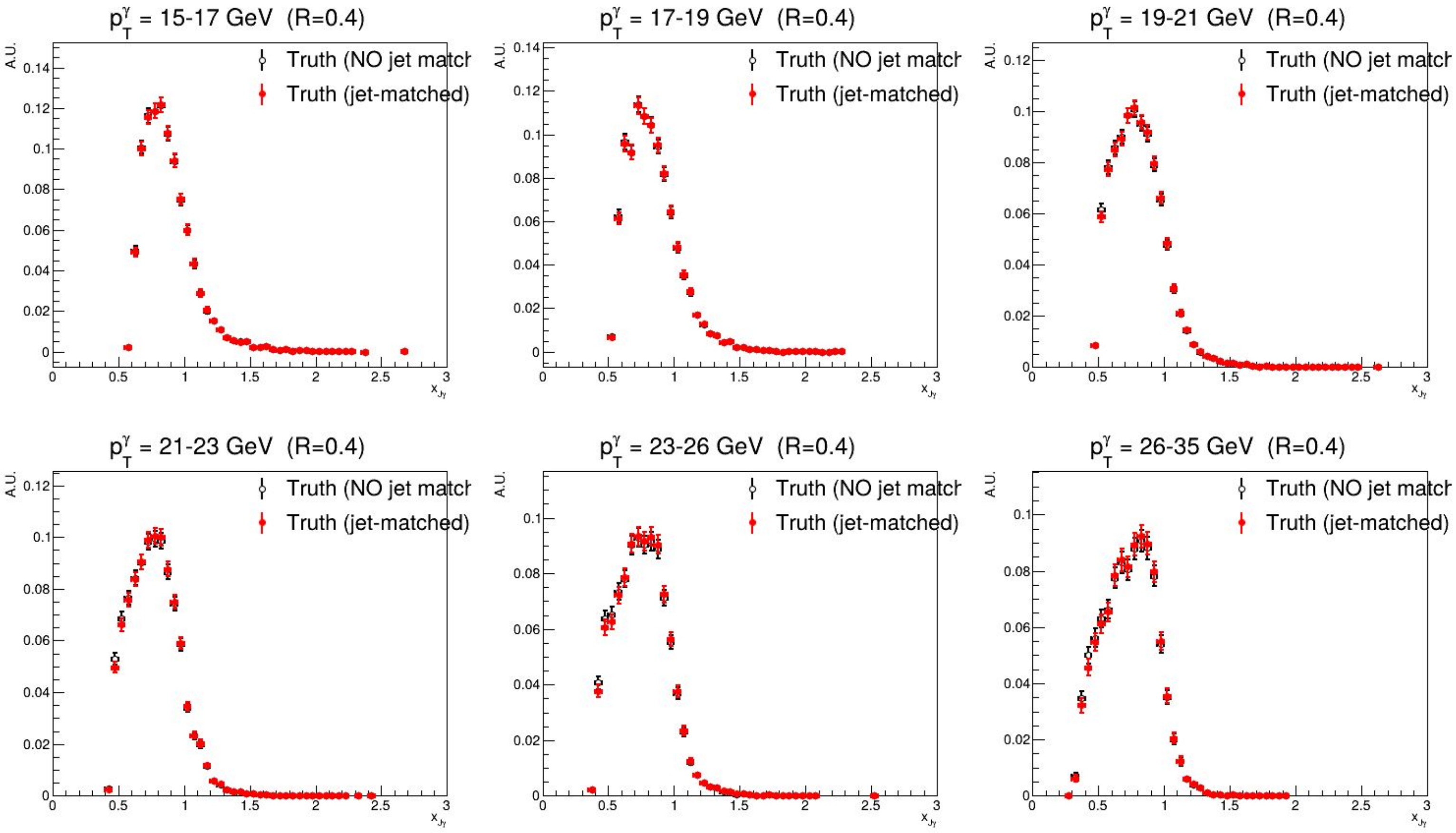
R = 0.4







Truth Conditioned on Reco Photon vs Reco Photon + Reco Recoil



Probability of a Valid Recoil Jet vs p_T^{γ}

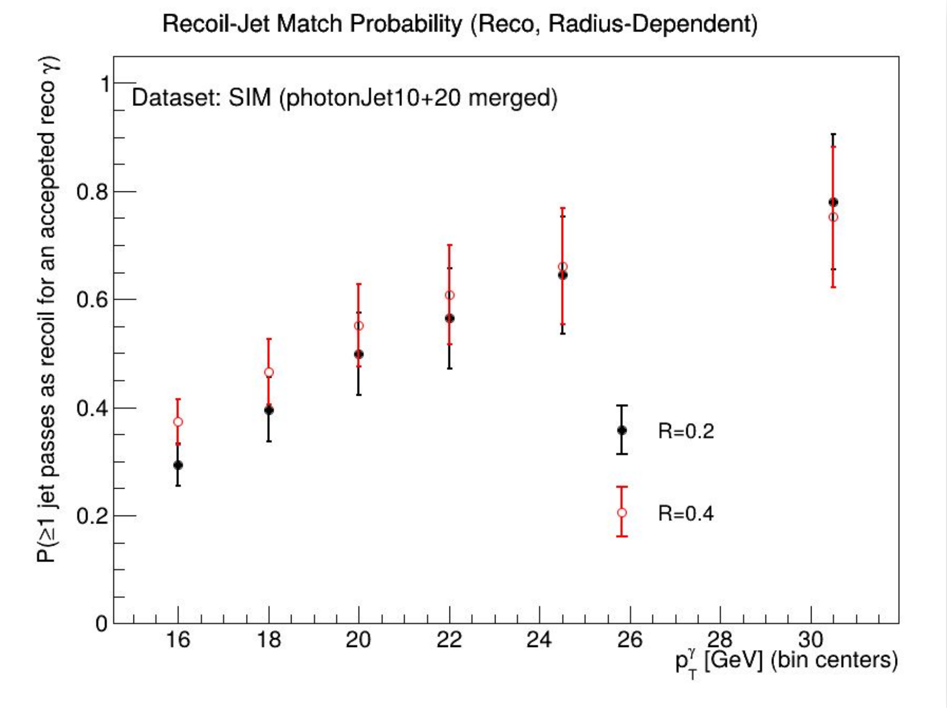
Plot built from a histogram filled from TH2F: $x \rightarrow p_T^{\gamma,\,reco}$ & $y \rightarrow$ *categorial status code* (1 - 4) counting events in each category:

1) No jet passes $p_T^{min} < 5$ GeV

2) Jet fails fiducial $|\eta| < (1.1 - R)$

3) Jets not back-to-back with γ

4) Valid recoil jet found

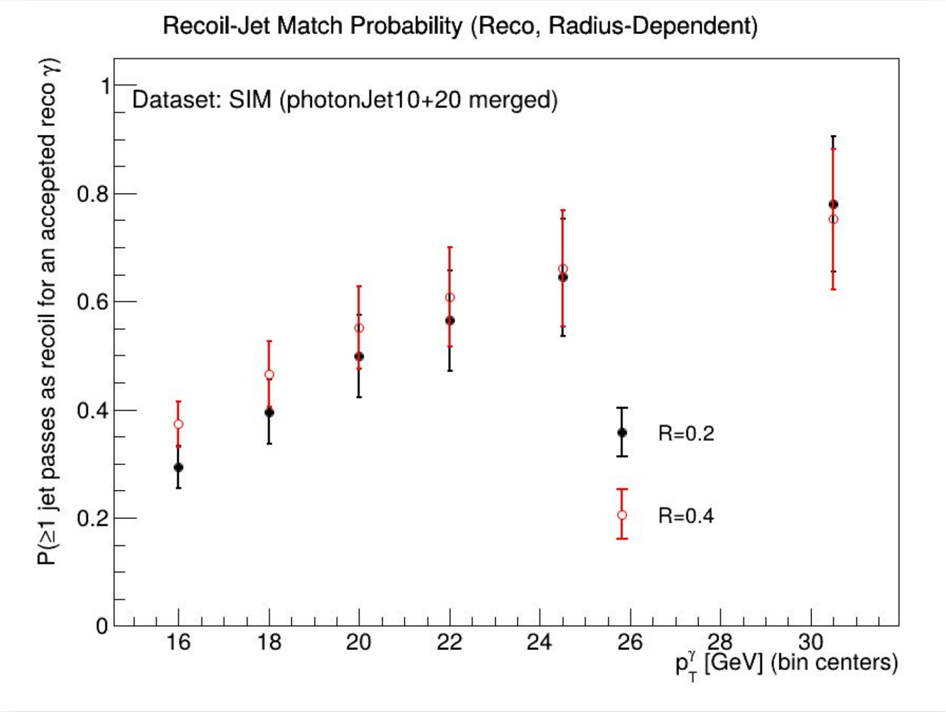


Matched fraction = $N_{matched}/N_{total}$

- **Not** a detector/reco efficiency
- Each point conditioned on leading reco iso $\wedge$ tight photon in a given $p_T^{\gamma,\,reco}$ bin
- **y values** → fraction of those events that contain ≥1 recoil jet passing all jet selection cuts
- Stat errors → binomial fraction (matched vs total leading-photon trials).

Recoil-Jet Reconstruction Success vs Photon p_T^{γ}

1) No jet passes $p_T^{min} < 5$ GeV
2) Jet fails fiducial $|\eta| < (1.1 - R)$
3) Jets not back-to-back with γ
4) Valid recoil jet found



Matched fraction = $N_{matched}/N_{total}$

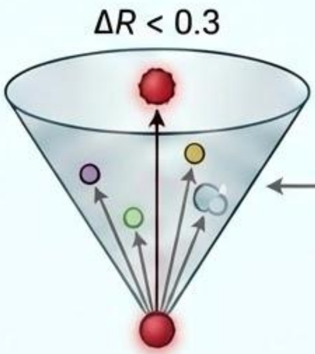
- Once we have a leading photon – the probability of finding ≥1 recoil jet passing $p_T/\eta/\Delta\varphi$ increaseas from ~O(50%) to 80-90% at high p_T
  - Recoil topology becomes more reliably reconstructable as hard scale increases
- Separation b/w R = 0.2 and 0.4 quantifies how often the recoil jet survives selection for each cone size
- Categorical breakdown can be used to gauge wheter losses dominated by p_T threshold, acceptance, or back-to-back topology cut

Truth vs Reconstructed Isolation Energies

Instead of using one fixed iso threshold → derived by PPG12 via fitted function of E_T^{γ}

- Allows prompt isolated photon efficiency to stay roughly *constant* vs E_T^{γ}

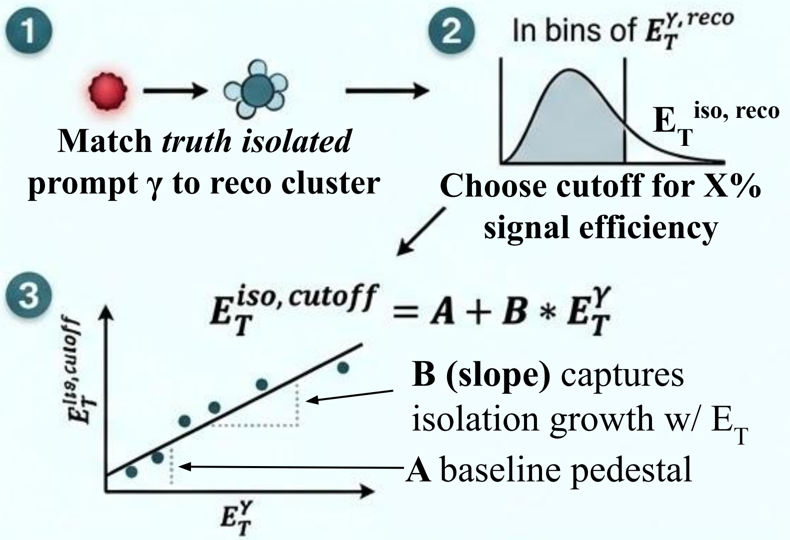
Truth-level Isolation (RecoilJets)



ΣE_T of all *final-state particles* inside **ΔR < 0.3 cone** & subtract photon contribution

Isolated if $E_T^{\text{iso, truth}} < 4.0$ GeV

Reco-level Sliding Isolation (PPG12 IAN)



Sliding Isolation Window

Current pipeline – uses PPG12 studies verbatim for reco isolation and data isolation definition at 90% efficiency cutoff

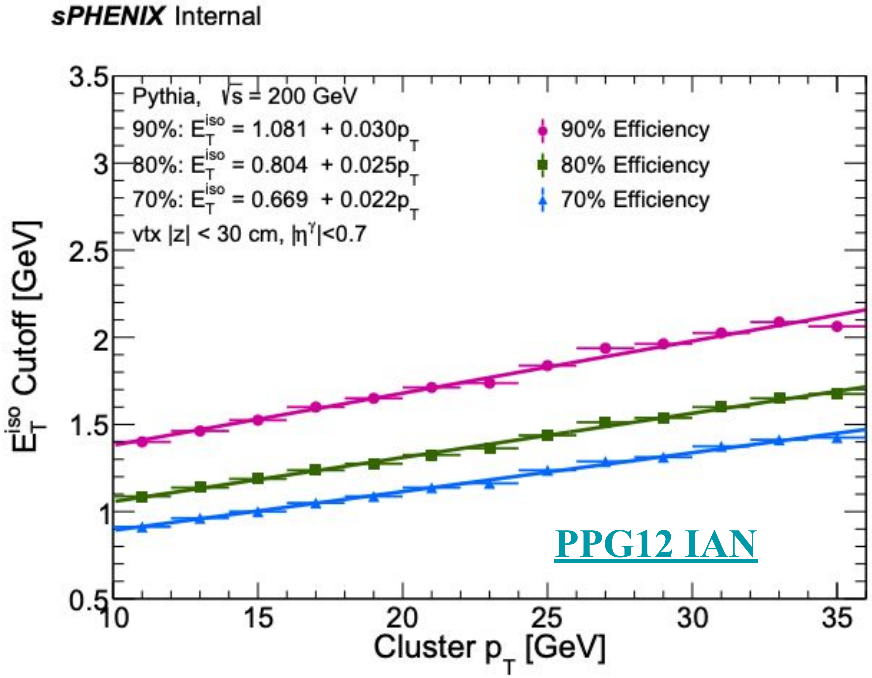
Is isolated:

$$E_T^{iso} < 1.08128 + 0.0299107 * E_T^{\gamma,\, reco}$$

Non isolated:

$$E_T^{iso} > \mathbf{1 +} 1.08128 + 0.0299107 * E_T^{\gamma,\, reco}$$

Everything between the “+1” gap is excluded from ABCD regions



PPG12 IAN

Photon Selection – SS Cuts and ABCD Regions

Pre-selection (loose **γ-ID quality-gate w/ $|\eta^{\gamma}| < 0.7$**): $\boxed{\frac{E_{11}}{E_{33}} < 0.98, \quad 0.6 < \mathrm{et1} < 1.0, \quad 0.8 < \frac{E_{32}}{E_{35}} < 1.0, \quad w_{\eta}^{\mathrm{cogX}} < 0.6}$

Tight γ-ID (5 SS variables → compact, EM-like showers)

(1 & 2) Seed-excluded widths (around CoG, exclude the seed tower) → $\boxed{0 < w_a^{\mathrm{cogX}} < 0.15 + 0.006\,E_T^{\gamma} \quad \text{for } a \in \{\eta, \phi\}}$

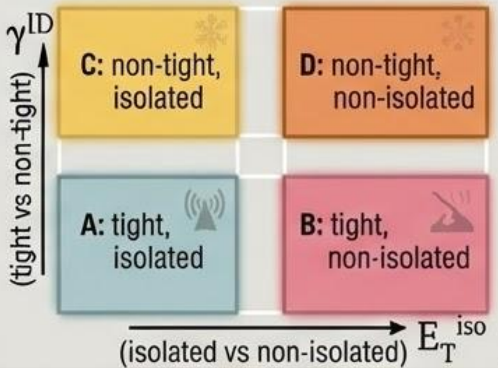
(3) Core-to-cluster compactness (odd×odd window sums) → $\boxed{0.4 < \frac{E_{11}}{E_{33}} < 0.98}$

(4) 2×2 core energy fraction (et1) → $\boxed{0.9 < \mathrm{et1} < 1.0}$ $\boxed{\mathrm{et1} = \frac{E_1 + E_2 + E_3 + E_4}{E_{\mathrm{tot}}}}$

(5) Elongation veto (strip ratio) → $\boxed{0.92 < \frac{E_{32}}{E_{35}} < 1.0}$

ABCD Regions

- **Purpose:** estimate the signal purity of your tight & isolated photon sample
- **Axes** → γ-ID (tight vs non-tight, w/ 5 SS variables), and isolated vs non-isolated
  - **Non-tight** → fails any two of five tight SS requirements
  - **Non-isolated** → $E_T^{\mathrm{iso}} > 1 + 1.08128 + 0.0299107 \cdot E_T^{\mathrm{reco}}$ (+1 gap > iso def)



Weighted Merging PhotonJet10 + PhotonJet20 GeV Sim Samples

Goal → Combine two MC slices into one physics-normalized SIM file.

PhotonJet10 Input

Event Count
$N_{10} = 3{,}557{,}412$
Cross Section
$\sigma_{10} = 6692.7611$ pb

PhotonJet20 Input

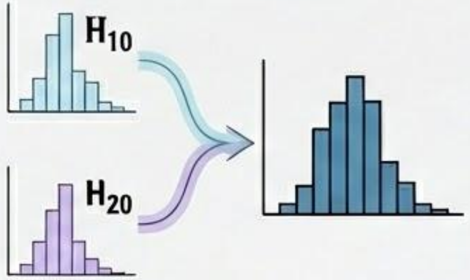
Event Count
$N_{20} = 3{,}555{,}769$
Cross Section
$\sigma_{20} = 105.79868$ pb

Weight Calculation

$w_k = \frac{\sigma_k}{N_k}$ [pb/event]

w_{10} for 10 GeV
↓
w_{20} for 20 GeV

Histogram Merging



$H_{merged}(b) = w_{10}H_{10}(b) + w_{20}H_{20}(b)$

Analysis runs reading from the single merged SIM file.

Result → Merged histograms are in cross-section units (pb/bin), not raw counts.

$$L_k = \frac{N_k}{\sigma_k}\ (\mathrm{pb}^{-1})$$

Each slice **k** simulated an "integrated luminosity".

$L_{10} = N_{10}/\sigma_{10}$

$L_{20} = N_{20}/\sigma_{20}$

Total Effective Luminosity (z < 30 cm):

$L_{eff} = L_{10} + L_{20} = 34140\ \mathrm{pb}^{-1}$

Tracing HepMC Ancestry – Classifying Direct vs Fragmentation

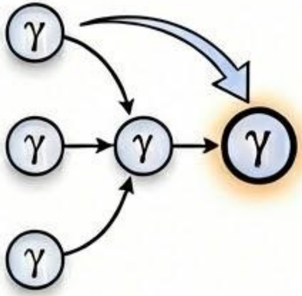
(1) Start with truth photon candidate

PDG = 22, status = 1

$|\eta^{\gamma,\,\mathrm{truth}}| < 0.7$

Ensure valid 'production_vertex()' otherwise reject

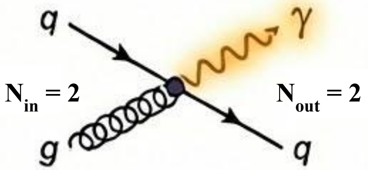
(2) HepMC walkback (collapse bookkeeping)



Skip generator copies.
Trace back to first nontrivial emission vertex

(3) Classify at 'true origin' vertex

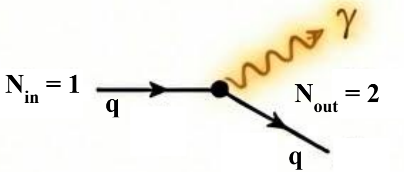
(a) Direct (hard-scatter) [class=1]



All legs |PDG| < 22 (partons/γ only)

Final Vertex

(b) Fragmentation [class=2]



Emitter continues + radiates γ
(e.g. q → q + γ)

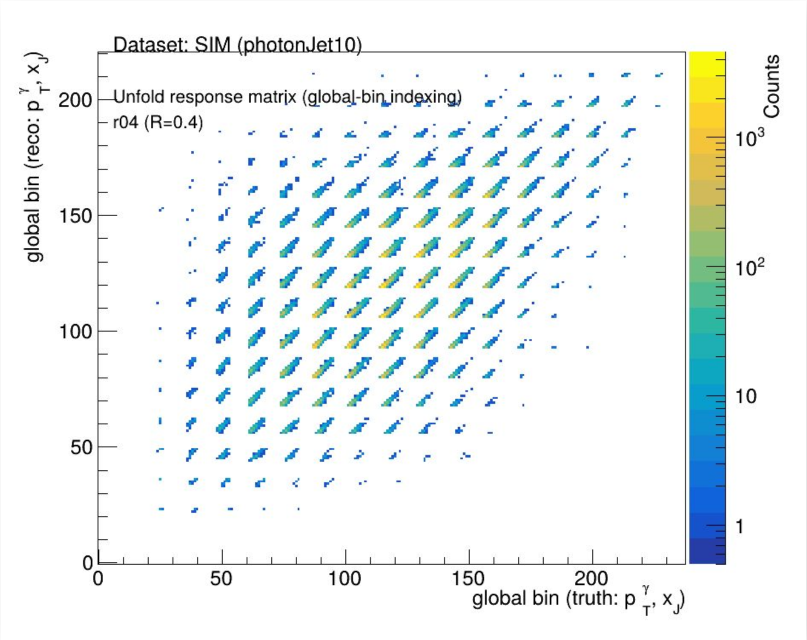
(4) Final selection

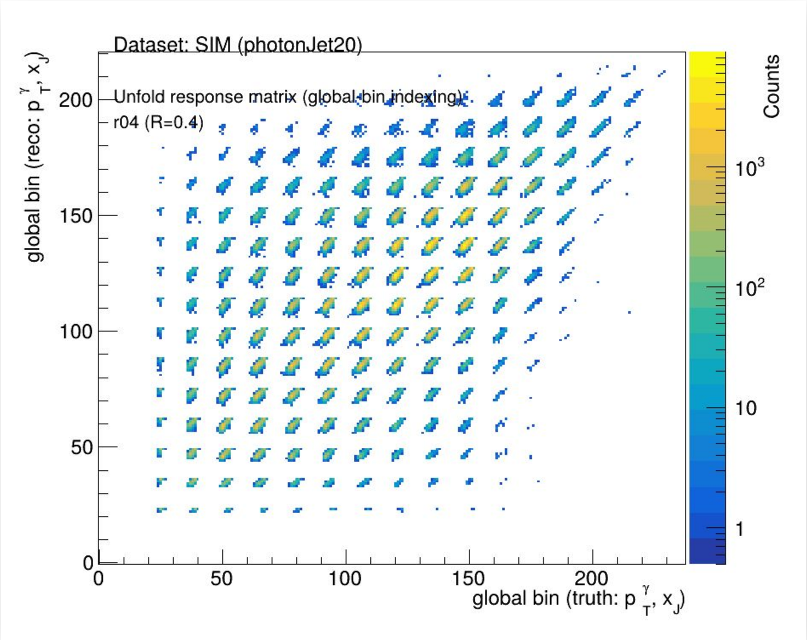
Prompt truth γ = accept only class 1 or 2

Otherwise reject

Prepared Response Matrix for R = 0.2 & R = 0.4 Jets

- 2D response matrix for unfolding (p_T^{γ}, $x_{J\gamma}$) shown for photonJet10 and 20 for R = 0.4 jets
- Each pixel is the number of events that migrate from a truth global-bin (x-axis) to a reco global-bin (y-axis), with log-Z counts (yellow = high occupancy).
- $x_{J\gamma}$ is a ratio, but smearing acts on the two measured energies (pTγ and pTjet) in a correlated way. Building the response in the joint observable space preserves these correlations





Consistency Check – Miss Decomposition Closure

Denominator (DEN) → set of *all trials* in a given $p_T^{\gamma,\,truth}$ bin — **DEN(i)** = "truth-leading recoil jet exists" events in bin *i*

- For **every** DEN event – code makes one and only one outcome choice:

  1) Success → it filles NUM(i)

  2) Failure → does *not* fill NUM — failure classified into:

  MissA(i) → reco match exists, but picked wrong jet, or

  Miss B → no reco match exists at all

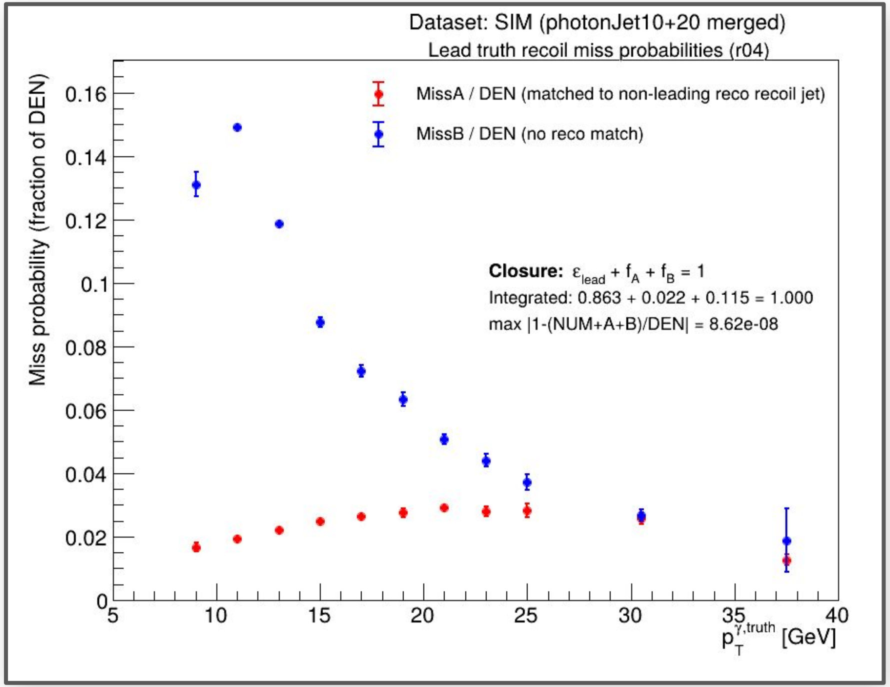
- **So** → per event (within a fixed bin *i*):

$$\underbrace{1}_{\text{this DEN event}} = \underbrace{\mathbf{1}_{\text{NUM}}}_{\text{success}} + \underbrace{\mathbf{1}_{\text{MissA}}}_{\text{failure type A}} + \underbrace{\mathbf{1}_{\text{MissB}}}_{\text{failure type B}}$$

- Summing over **DEN events** gives:

$$\boxed{\text{DEN}(i) = \text{NUM}(i) + \text{MissA}(i) + \text{MissB}(i)}$$

- Let → $\varepsilon(i) = \dfrac{\text{NUM}(i)}{\text{DEN}(i)}, \quad f_A(i) = \dfrac{\text{MissA}(i)}{\text{DEN}(i)}, \quad f_B(i) = \dfrac{\text{MissB}(i)}{\text{DEN}(i)}$
  - So should have → $\varepsilon(i) + f_A(i) + f_B(i) = 1$



B/c closure satisfies this (integrated sum ~ 1 and max bin wise residual << 1) – **MissA**/**MissB** **fractions** are a *complete*, *mutually exclusive* breakdown of the efficiency failures

Back-to-Back Fraction vs $p_T^{\gamma,\ truth}$

Hist construction –

1) Start from the class-split TH2's:

H_C(x = truth p_T^{γ}, y = |Δϕ|) for NUM/MissA/MissB

2) Count "all" and "back-to-back" events per truth photon bin → for each p_T^{γ} bin *i*:

- $N_{all}(i)$ → sum over all |Δϕ| ∈ [0, π] bins
- $N_{BB}(i)$ → sum only over $|\Delta\phi| > \Delta\phi_0 = 2.8$ rad

3) For ratio with propagated uncertainties

$f_{BB}(i) = N_{BB}(i)/N_{all}(i)$

What it tells us → how often the selected recoil jet is truly "tight back-to-back" in each p_T bin for:

- NUM(black) → correctly matched recoil jets
- MissA(red) → wrong-jet selection evts
- MissB → "no reco match" (substitute jet) evts

For each $p_T^{\gamma,\ truth}$ bin – and each class C ∈ {NUM, MissA, MissB} plot:

$$f_{BB}^{(C)}(i;\Delta\phi_0) = \frac{N_C\left(|\Delta\phi(\gamma^{truth}, j^{reco})| > \Delta\phi_0\right)}{N_C\left(\text{all } |\Delta\phi|\right)} \quad \text{with} \quad \Delta\phi_0 = 2.8 \text{ rad}$$