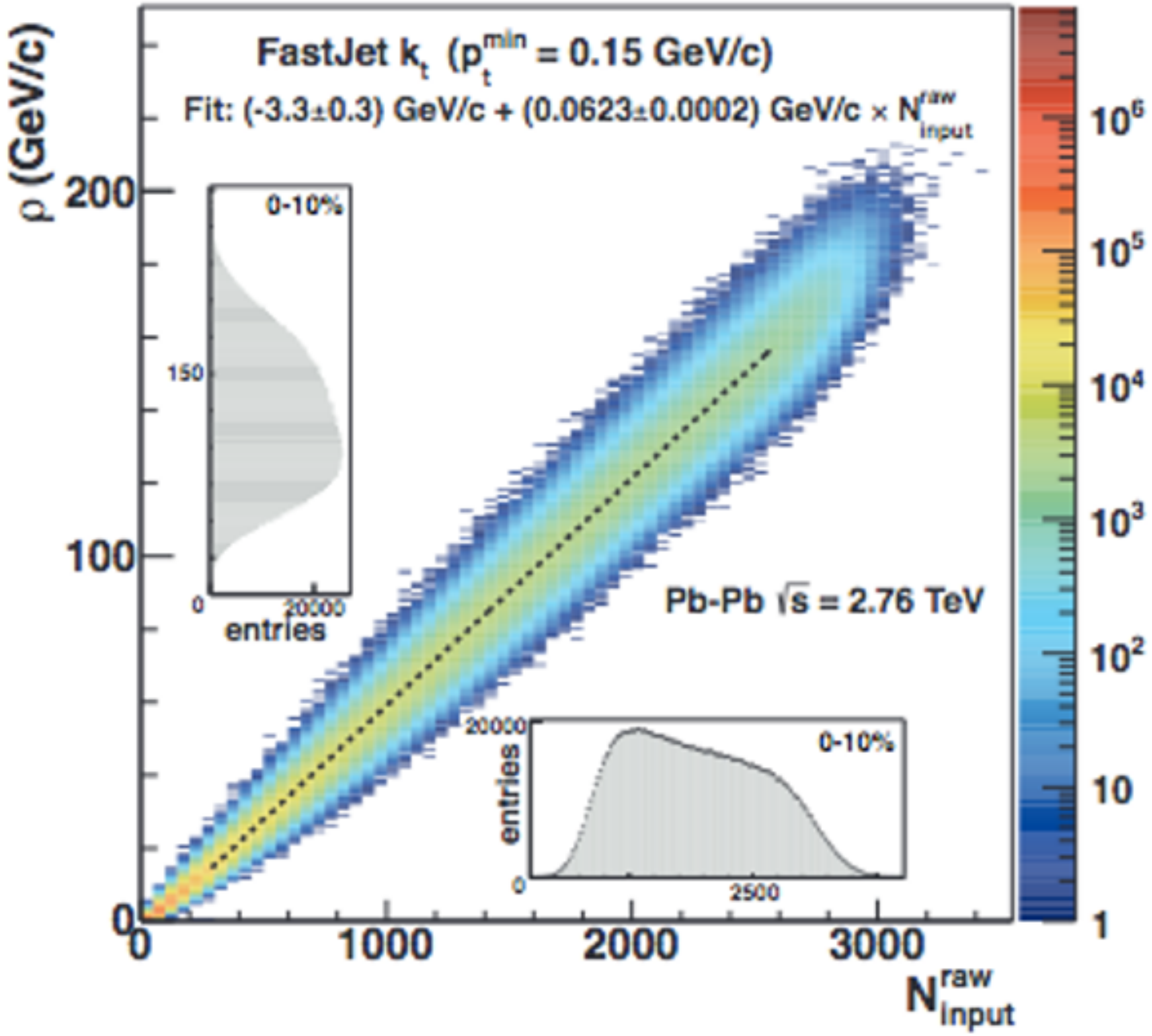
European Research Council
Established by the European Commission

SAPIENZA
UNIVERSITÀ DI ROMA

Lesson 4
from the raw to a fully corrected jet measurement

Leticia Cunqueiro Mendez

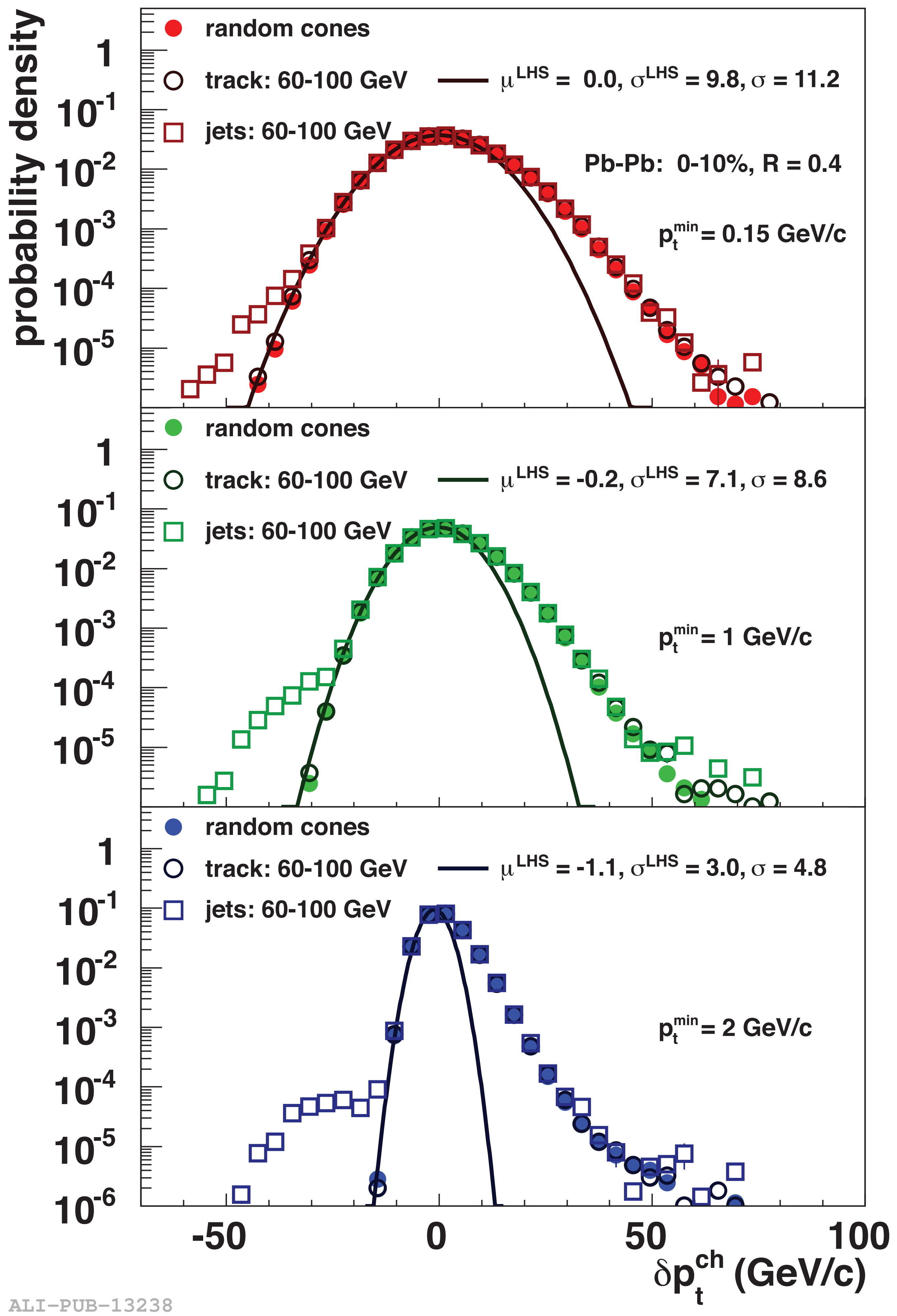
Measurement of the underlying event



Underlying event $\rho \approx 150$ GeV in central collisions

Remember UE contamination ∝ jet area, 150 * pi *R^2 **~75 GeV in a jet of R=0.4**

Background response: region-to-region fluctuations



[ALICE, JHEP03(2012)053]

$$\delta p_{\mathrm{T}} = p_{\mathrm{T,jet}}^{\mathrm{reco}} - \rho \cdot A_{\mathrm{jet}} - p_{\mathrm{T}}^{\mathrm{part,embed}}$$

We embed different probes into Pb-Pb events and estimate the background response through δp_{T}

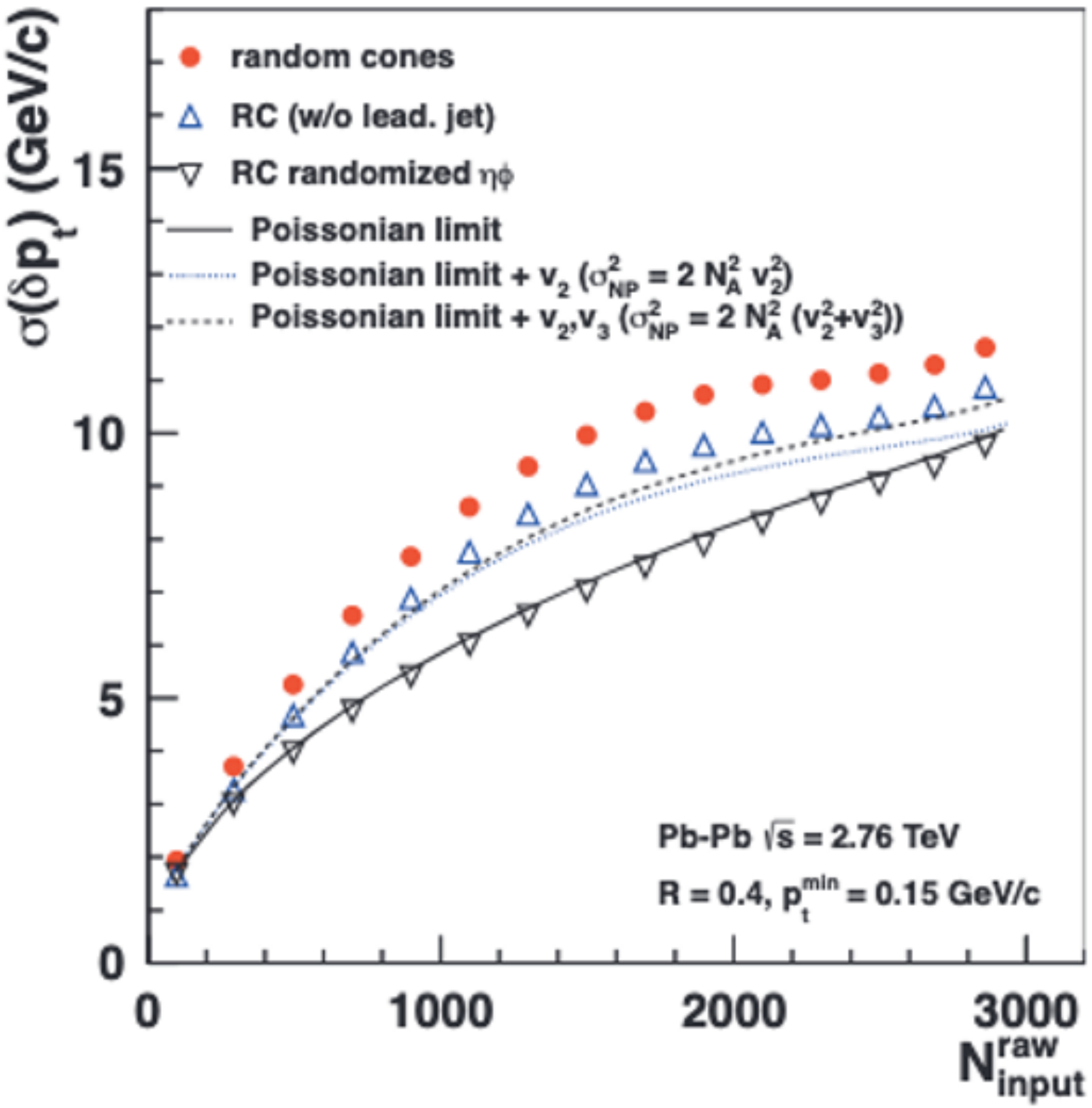
-Small dependence on the probe fragmentation pattern

-Small **back reaction** effects in the tails of the response due to jet splitting and jet merging

-Minimum constituent p_{T} cut-off reduces fluctuations

Note that different experiments have different effective pt cutoffs

Components of the background fluctuations



Randomized event or poissonian limit:
the number of particles is kept but their
η, ϕ coordinates are randomly redistributed

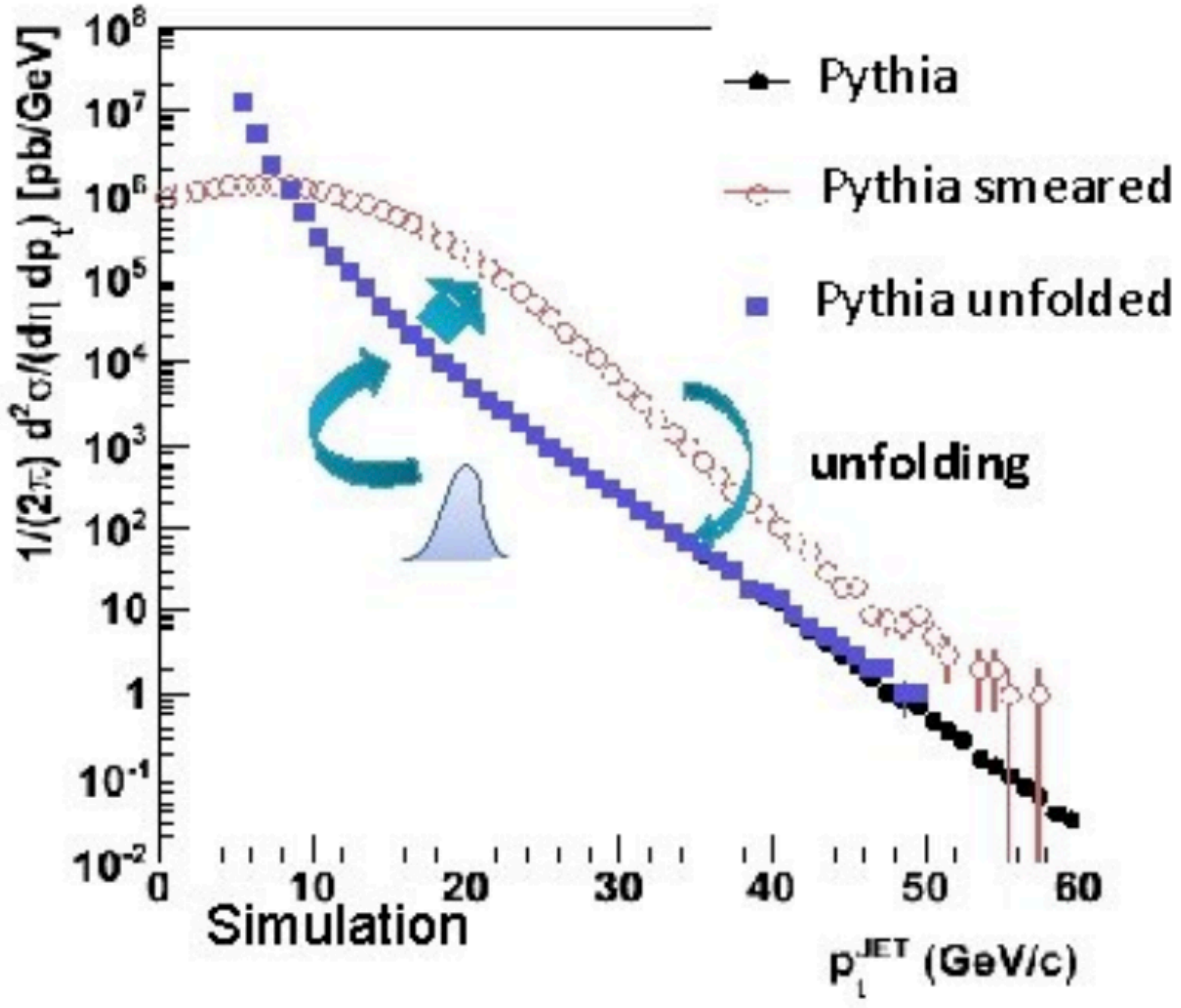
When azimuthal anisotropies are simulated in the event,
naturally fluctuations grow

What does a δp_T of 11 GeV mean?

$$\frac{dN^{\mathrm{Meas}}}{dp_T} = \frac{dN^{\mathrm{True}}}{dp_T} \otimes f^{\mathrm{Resol}}(\delta\rho)$$

δP_T distribution:
'smearing' of jet spectrum
due to background fluctuations

Large effect on yields
Need to unfold

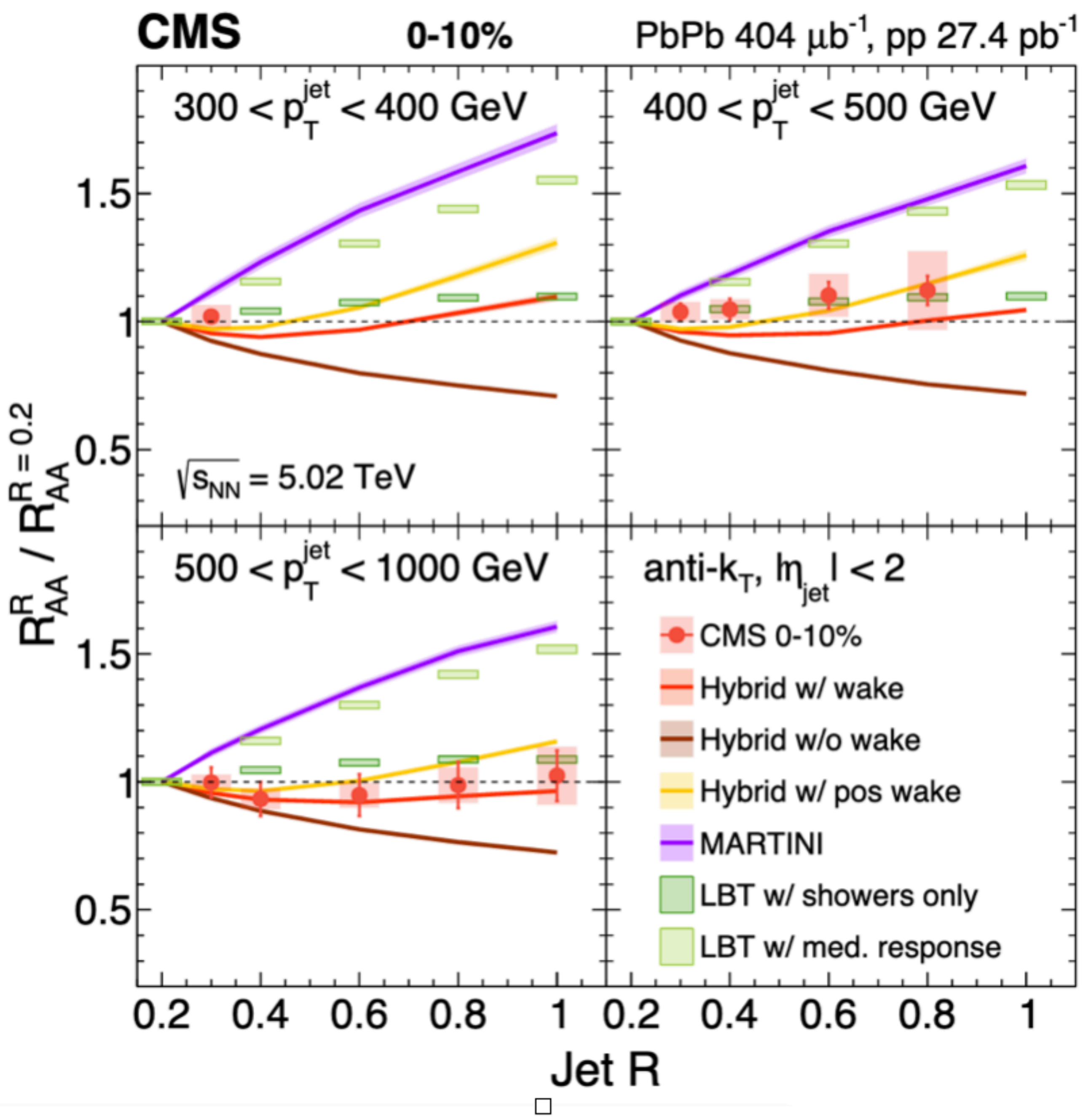


Note: the impact of the smearing depends on the underlying distribution, if you have a flat distribution for instance, smearing does not modify the truth
With a power-law, smearing goes in one direction preferentially

Fake jets

It also means that it is likely to reconstruct jets that are not correlated to any signal
Those purely bkg jets are called fake jets
Fake jets need to be suppressed before unfolding, cause unfolding conserves counts.
This is done typically:

a) by considering very high p_T jets (recall CMS results for large-R jets in lesson 2.)

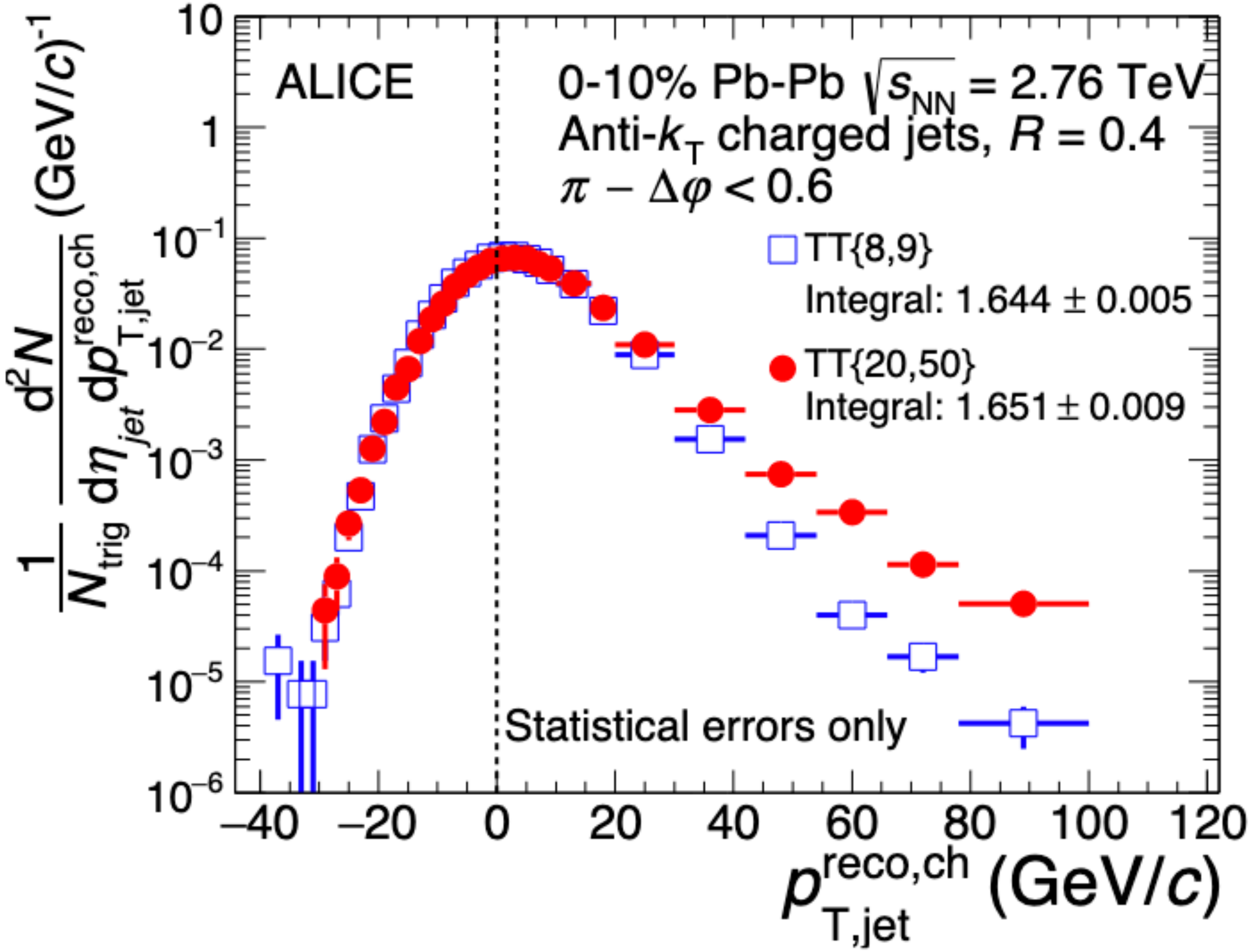


R=1 jets are only reported in the very high p_T bin

Faket jets

$$\Delta_{\mathrm{recoil}} = \frac{1}{N_{\mathrm{trig}}} \frac{\mathrm{d}^2 N_{\mathrm{jet}}}{\mathrm{d}p^{\mathrm{ch}}_{\mathrm{T,jet}}\,\mathrm{d}\eta_{\mathrm{jet}}}\Bigg|_{TT_{\mathrm{sig}}} - c_{\mathrm{ref}} \cdot \frac{1}{N_{\mathrm{trig}}} \frac{\mathrm{d}^2 N_{\mathrm{jet}}}{\mathrm{d}p^{\mathrm{ch}}_{\mathrm{T,jet}}\,\mathrm{d}\eta_{\mathrm{jet}}}\Bigg|_{TT_{\mathrm{ref}}}$$

b) by applying data-driven coincidence measurements

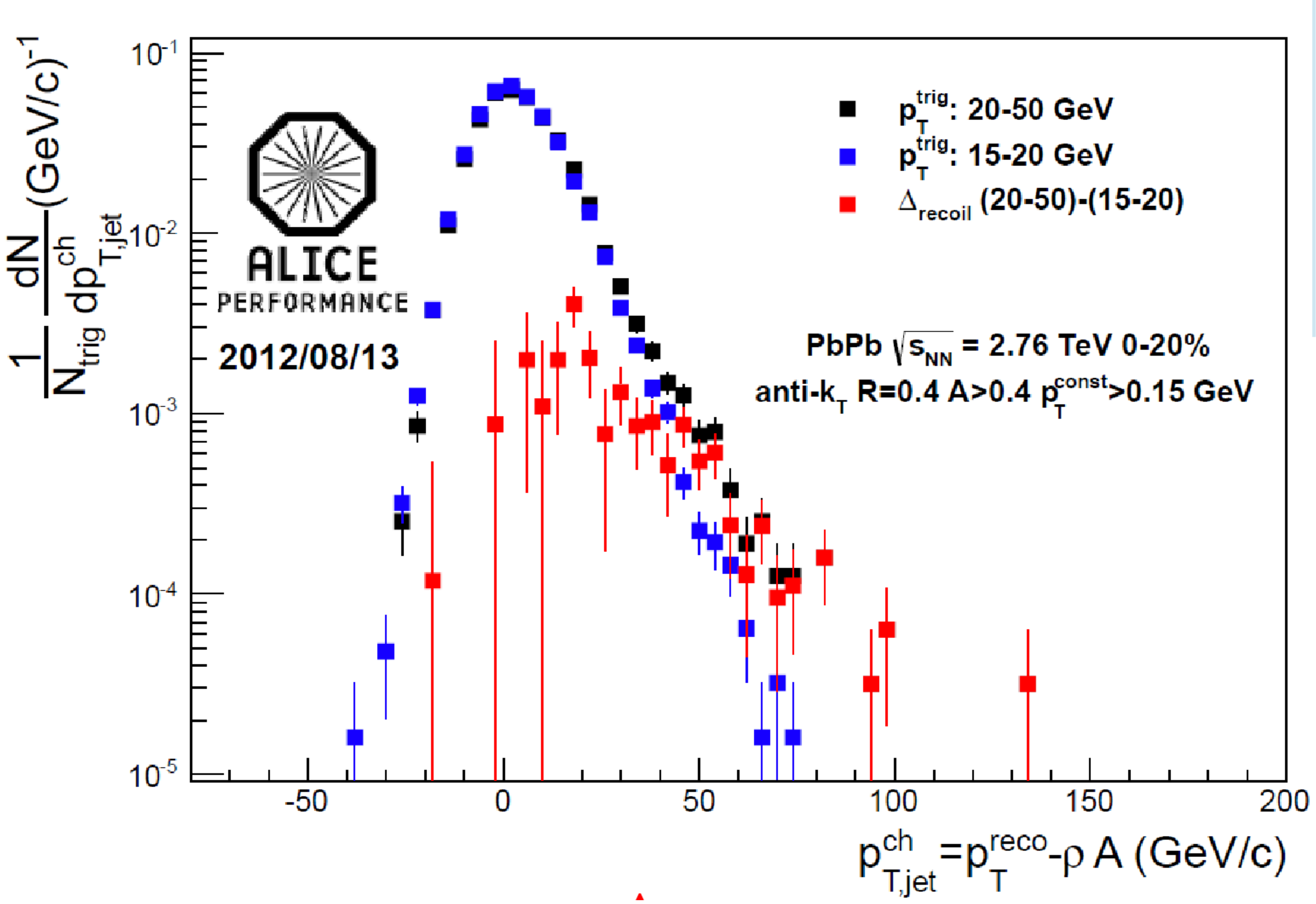


Faket jets

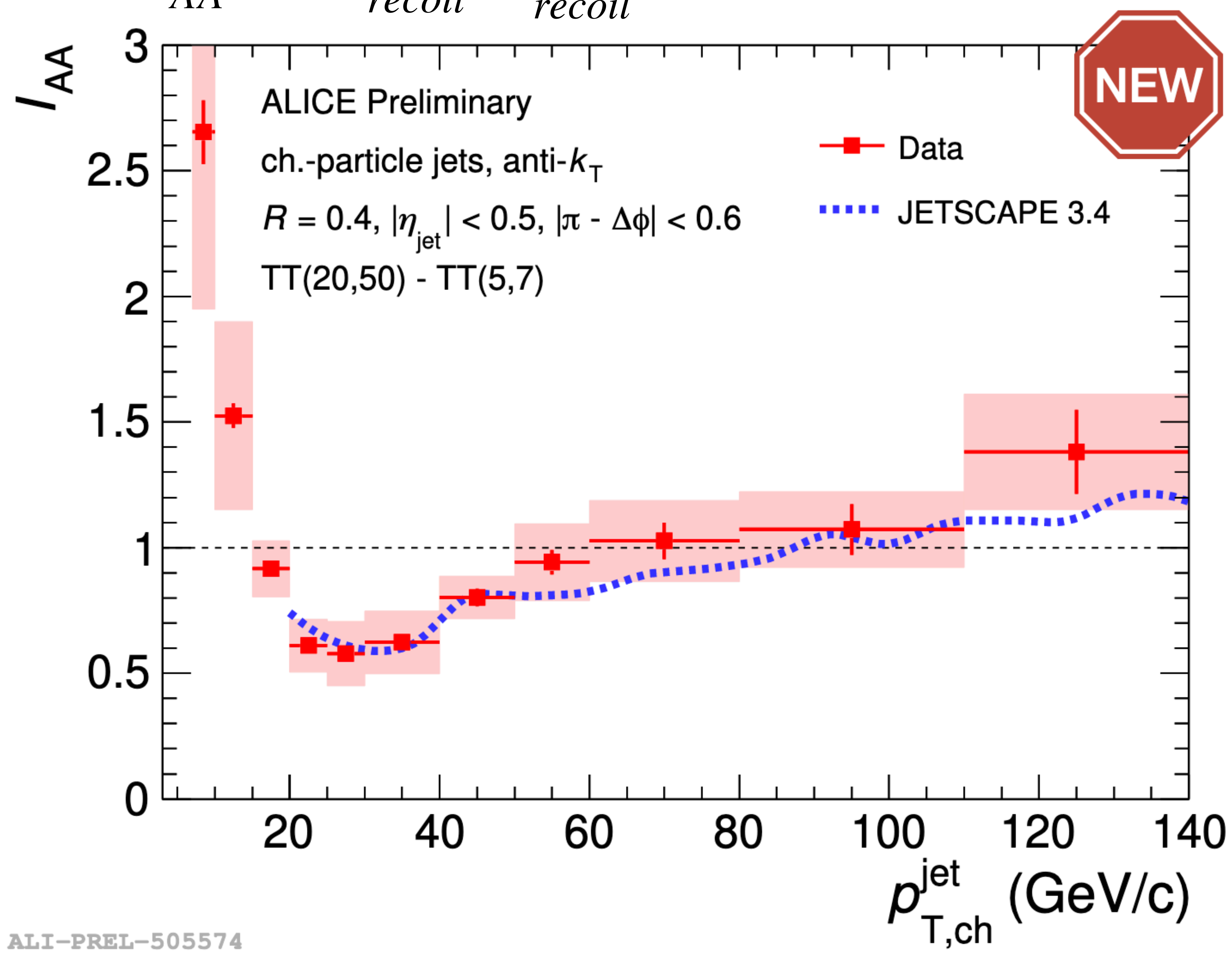
$$\Delta_{\mathrm{recoil}} = \frac{1}{N_{\mathrm{trig}}}\frac{\mathrm{d}^2N_{\mathrm{jet}}}{\mathrm{d}p^{\mathrm{ch}}_{\mathrm{T,jet}}\mathrm{d}\eta_{\mathrm{jet}}}\bigg|_{TT_{\mathrm{sig}}} - c_{\mathrm{ref}}\cdot\frac{1}{N_{\mathrm{trig}}}\frac{\mathrm{d}^2N_{\mathrm{jet}}}{\mathrm{d}p^{\mathrm{ch}}_{\mathrm{T,jet}}\mathrm{d}\eta_{\mathrm{jet}}}\bigg|_{TT_{\mathrm{ref}}}$$

b) by applying data-driven coincidence measurements

Δ_{recoil} is a fake-free distribution that can be unfolded down to zero transverse momentum



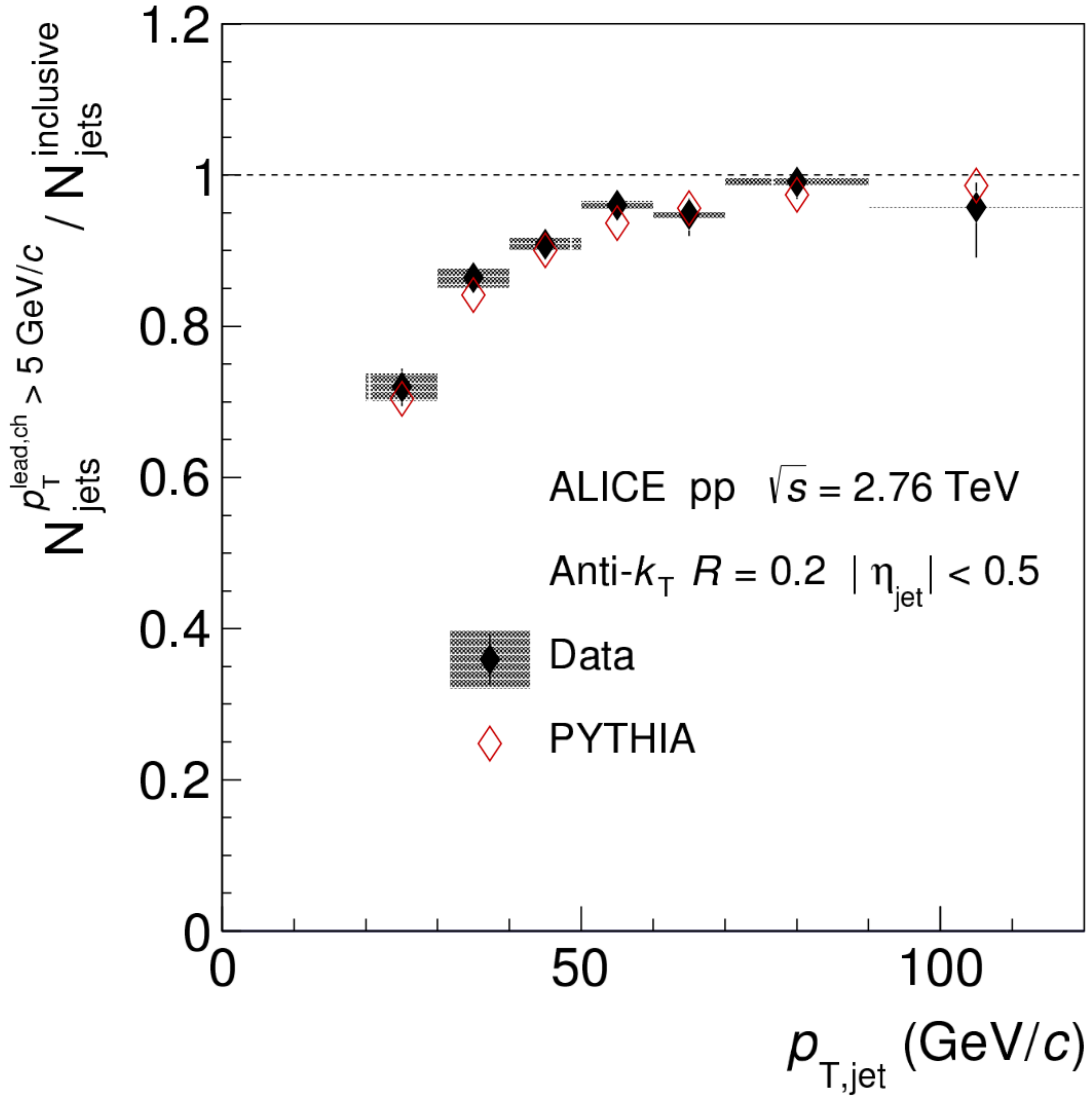
$$I_{AA} = \Delta^{PbPb}_{recoil}/\Delta^{pp}_{recoil}$$



ALI-PREL-505574

Fake jets

c) by requiring hard fragmentation (whatch out, it introduces a potential bias!)



Require that your jets have a constituent with p_T>5 GeV

This suppresses fakes but also potentially softer quenched jets!

Fake jets

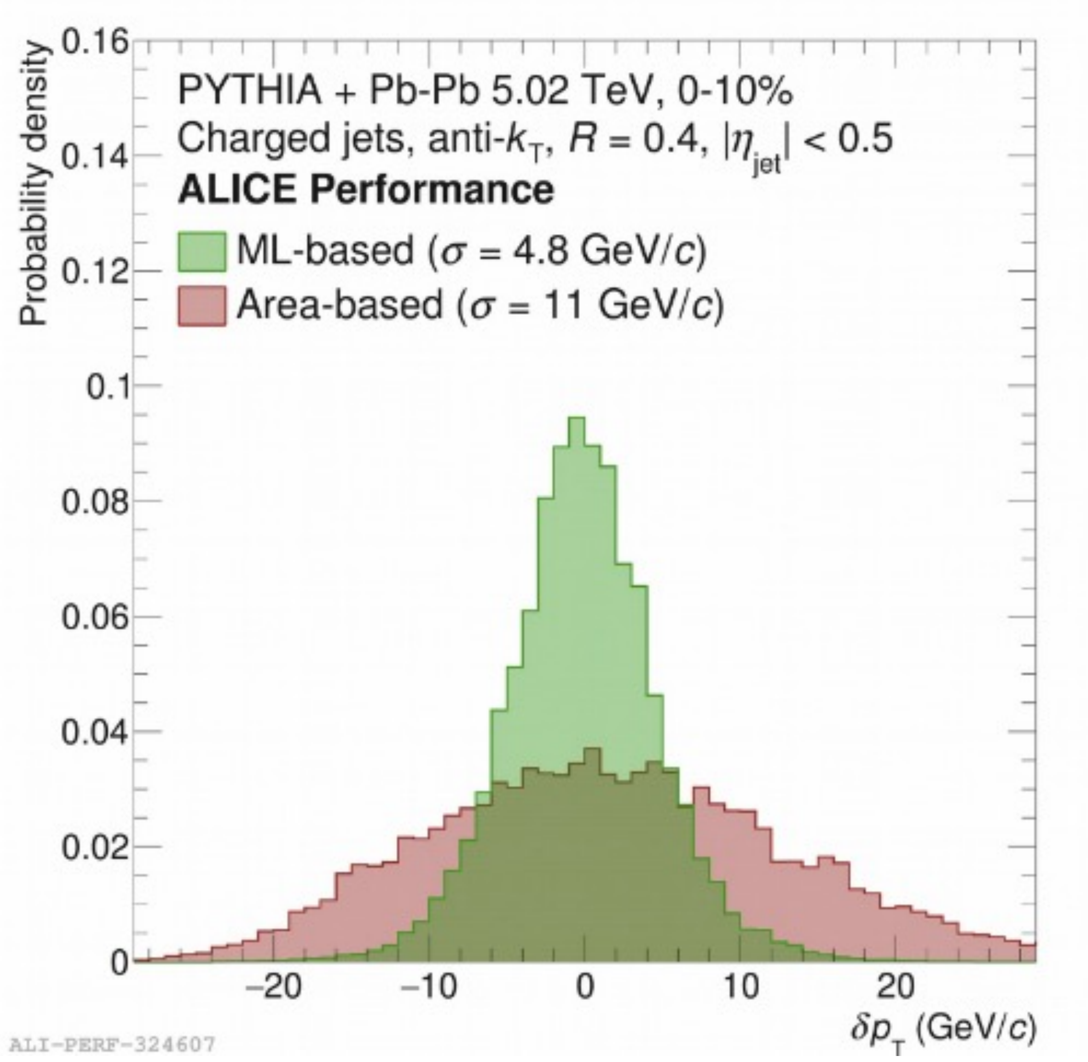
d) by reducing bkg fluctuations using for instance ML techniques

Area-based is an unbiased method, knows nothing about the jet structure

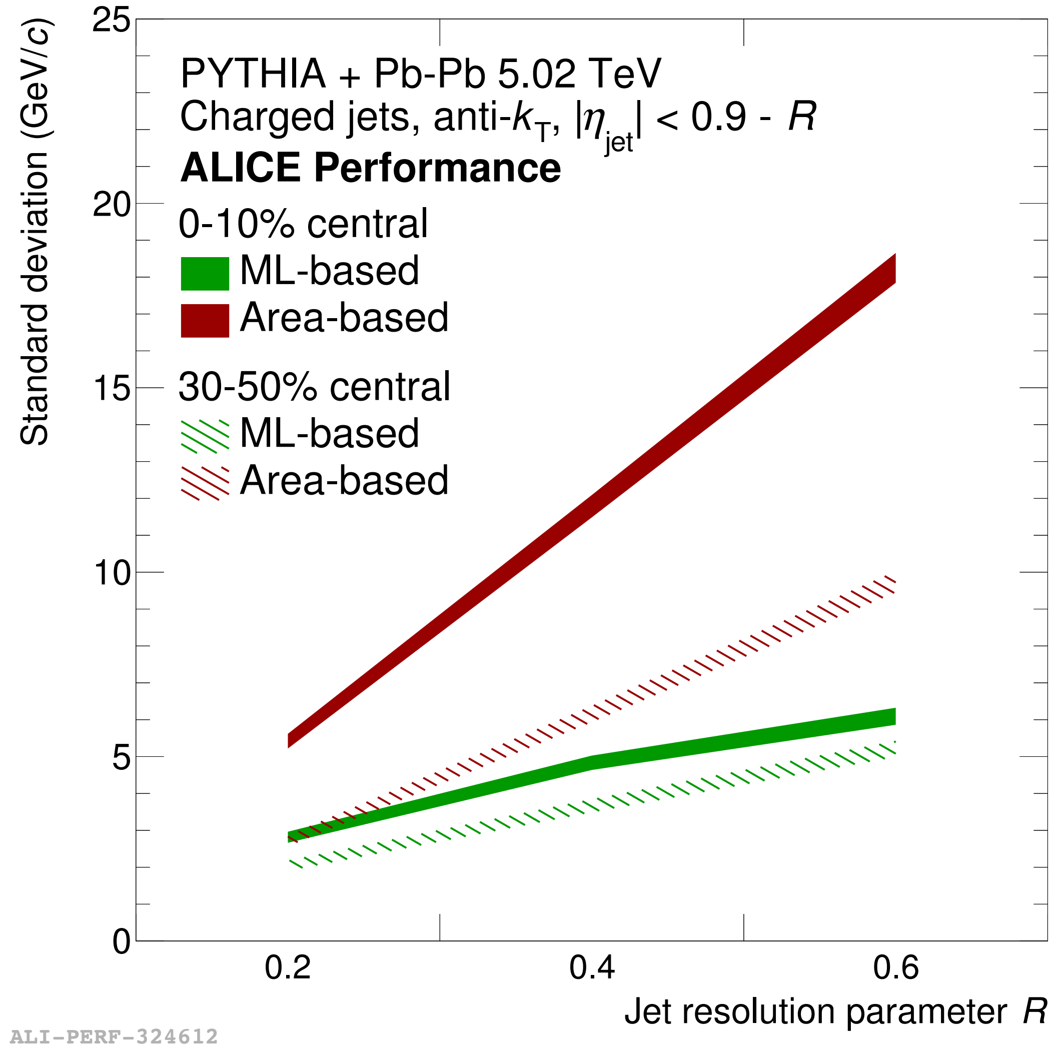
$p_{Tsub}=p_{Traw} - \rho \text{Area}$

New ML approach: a regression model to subtract background. Input parameters: area-based subtracted jet p_T, angularity, number of constituents, p_T of the 8 leading jet constituents

Background fluctuations



Width of the fluctuations vs jet *R*



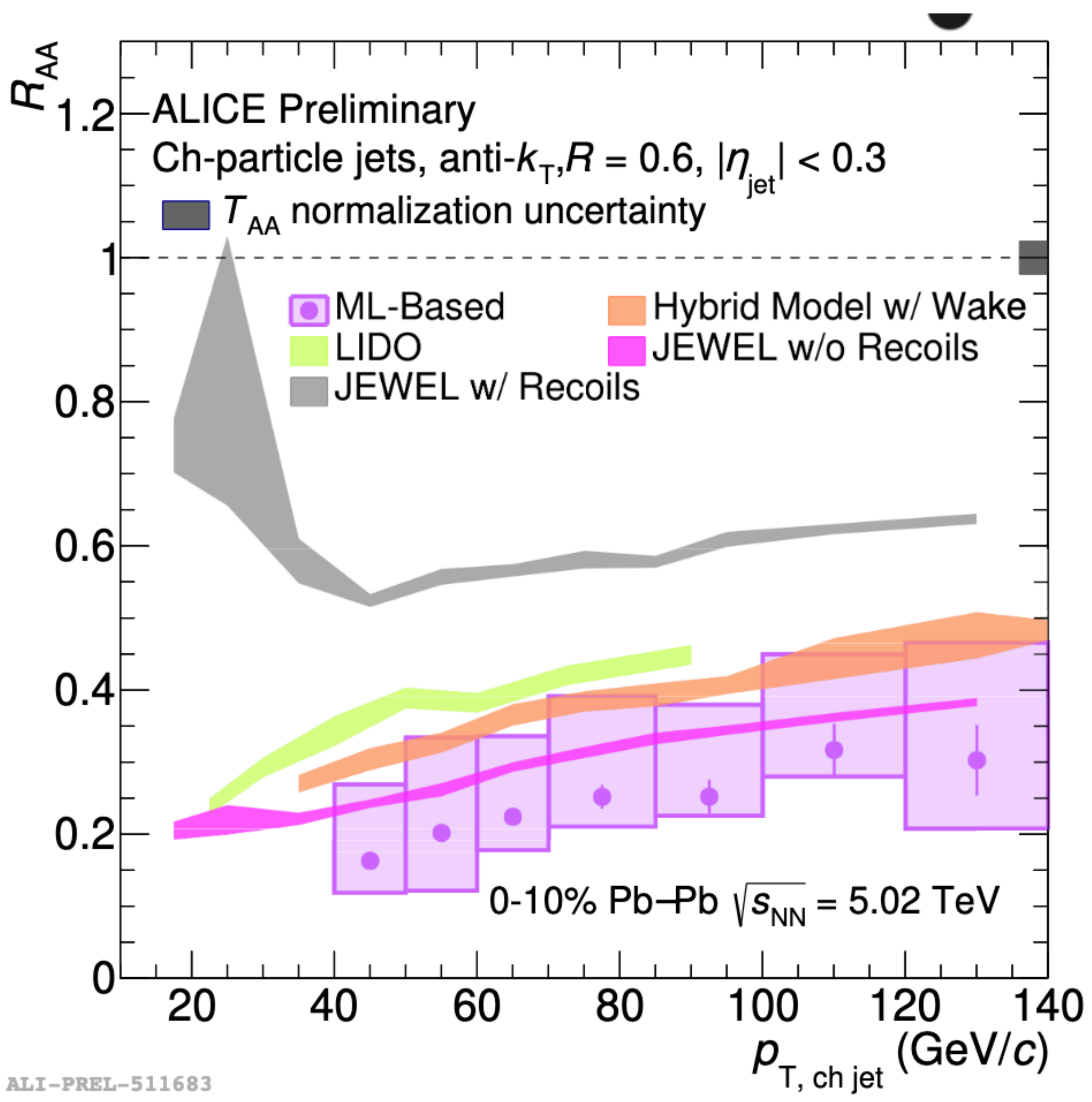
Method paper: *Haake, Loizides Phys.Rev. C99 (2019) no.6, 064904*

Background fluctuations suppressed:
- Combinatorial jets reduced
- Smaller unfolding correction
- Extend to low jet p_T and large R

Caveat: background subtraction might become model dependent

Fake jets

d) by reducing bkg fluctuations using for instance ML techniques

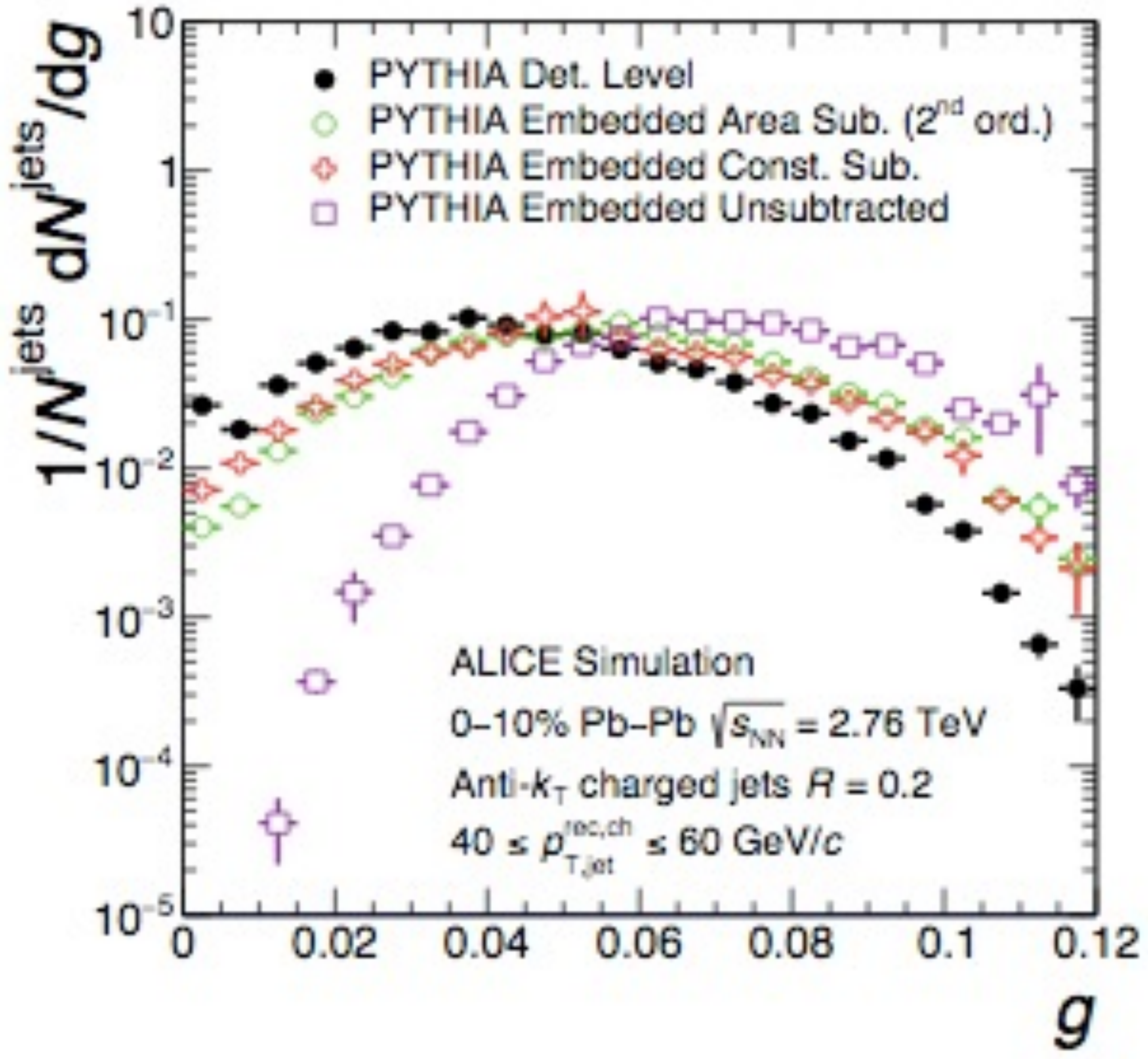


(recall lesson 2. and ALICE results for large-R jets at unprecedently low p_T)

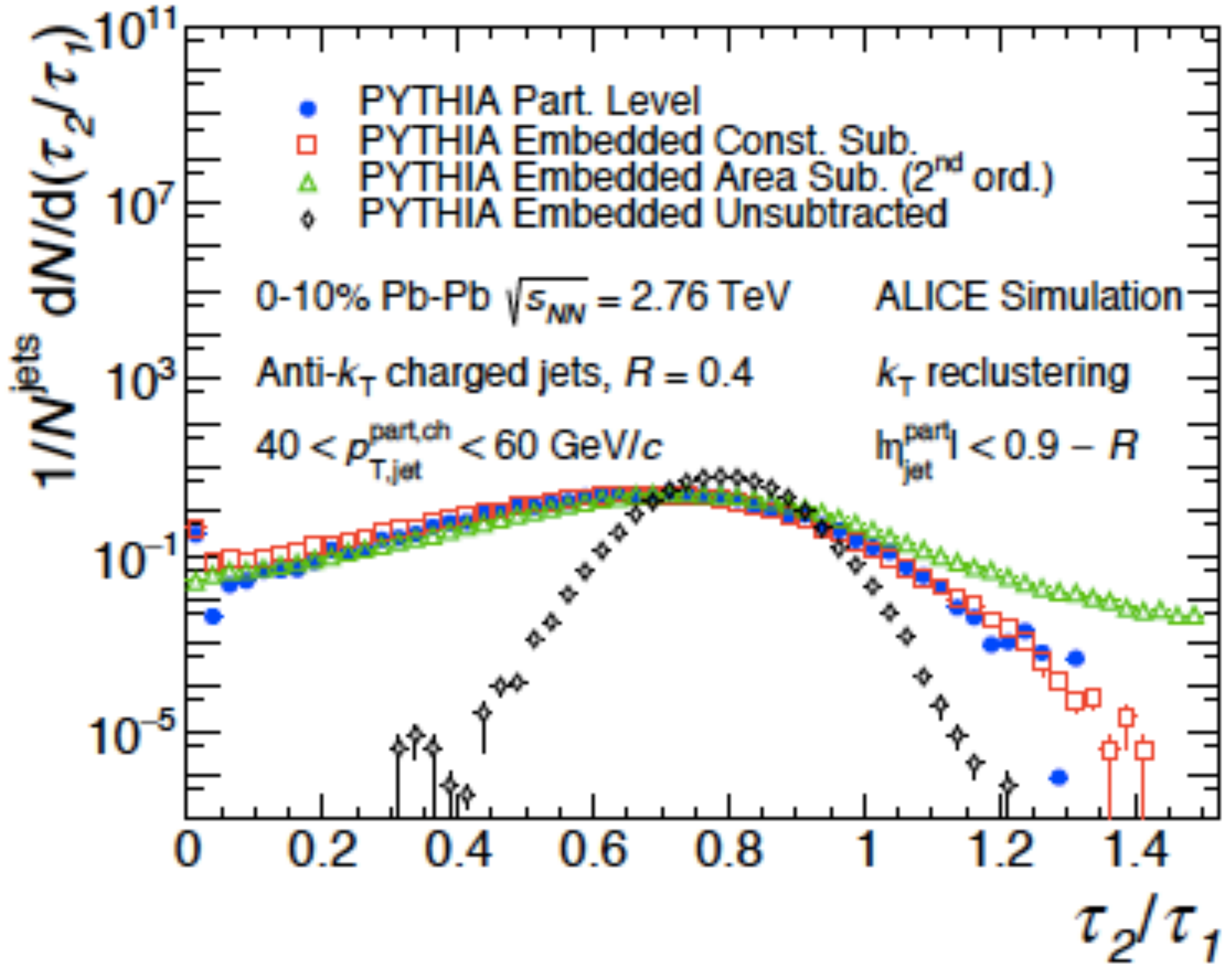
Uncertainty dominated by systematic uncertainty on training model dependency

Background at the level of jet substructure

Subtraction Performance angularity



Subtraction performance 2-subjettiness

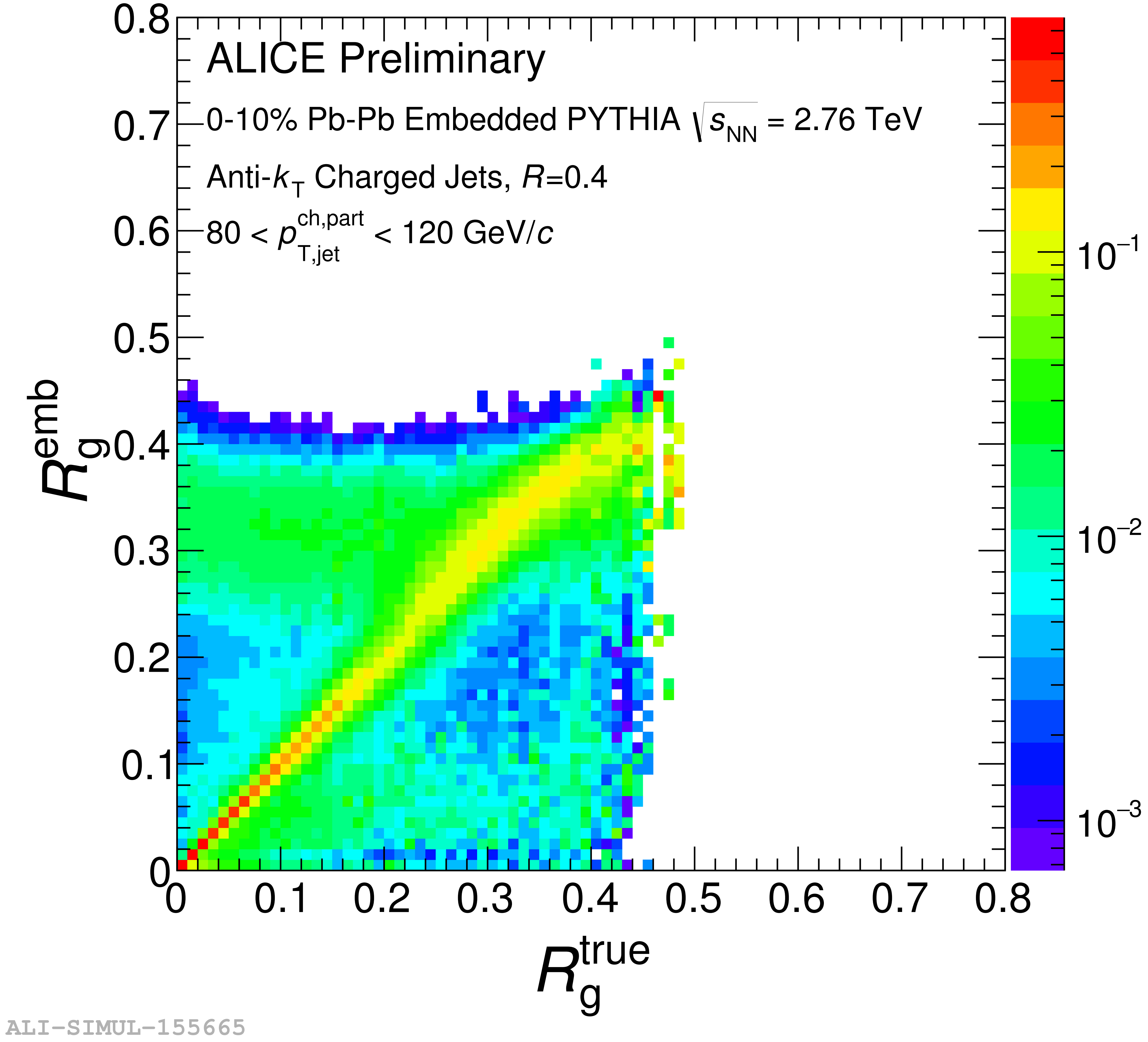


Pedestal background contamination subtracted using Area-derivatives [1] and Constituent Subtraction [2] methods.
Residual differences between the probe and the subtracted embedded probe are unfolded

[1] Soyez et al, Phys.Rev.Lett. 110 (2013) no.16, 162001

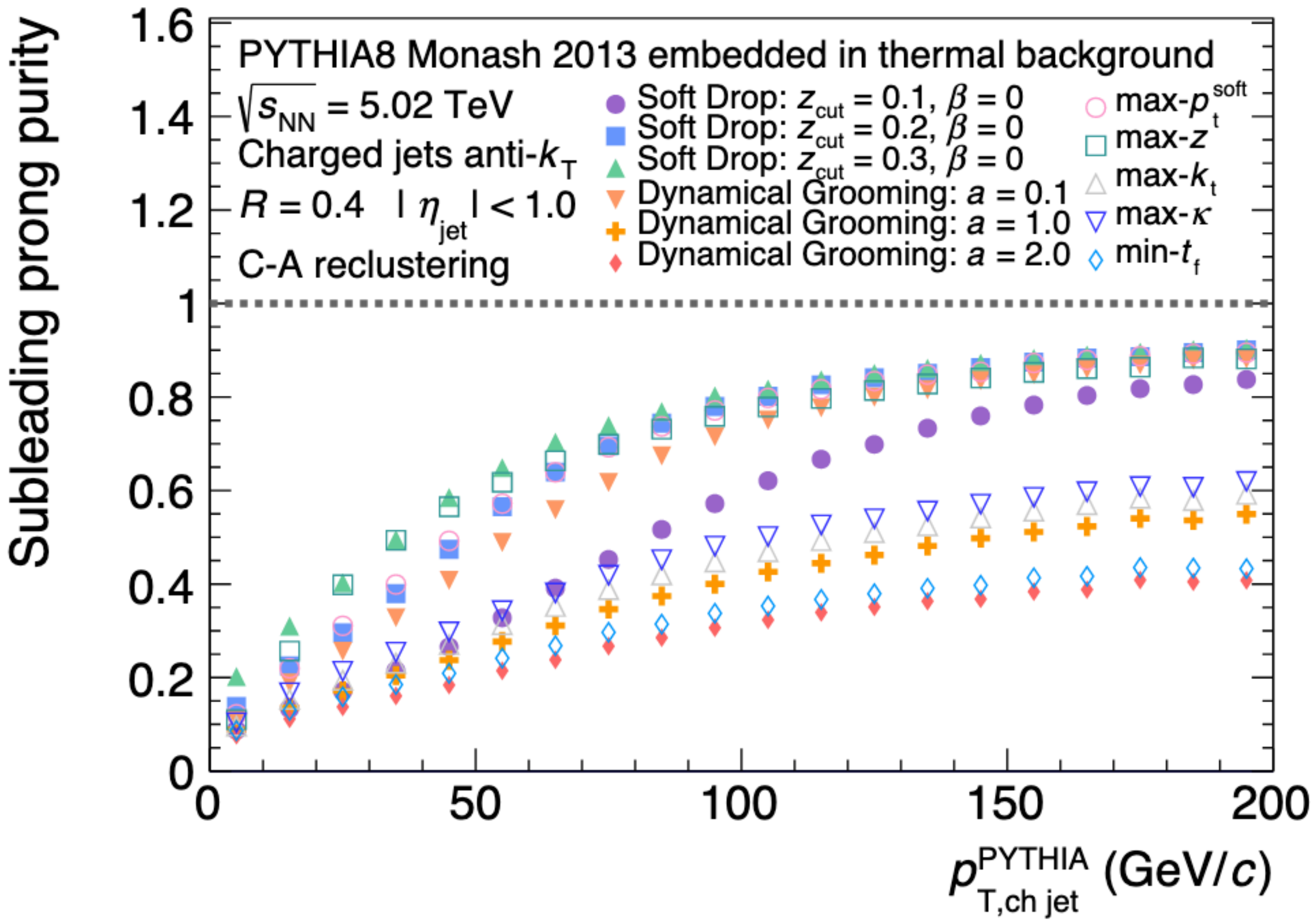
[2] Berta et al, JHEP 1908 (2019) 175

The groomed substructure response



Off-diagonalities in the reponse due to fake prongs
Render the problem not unfoldable

Fake splittings

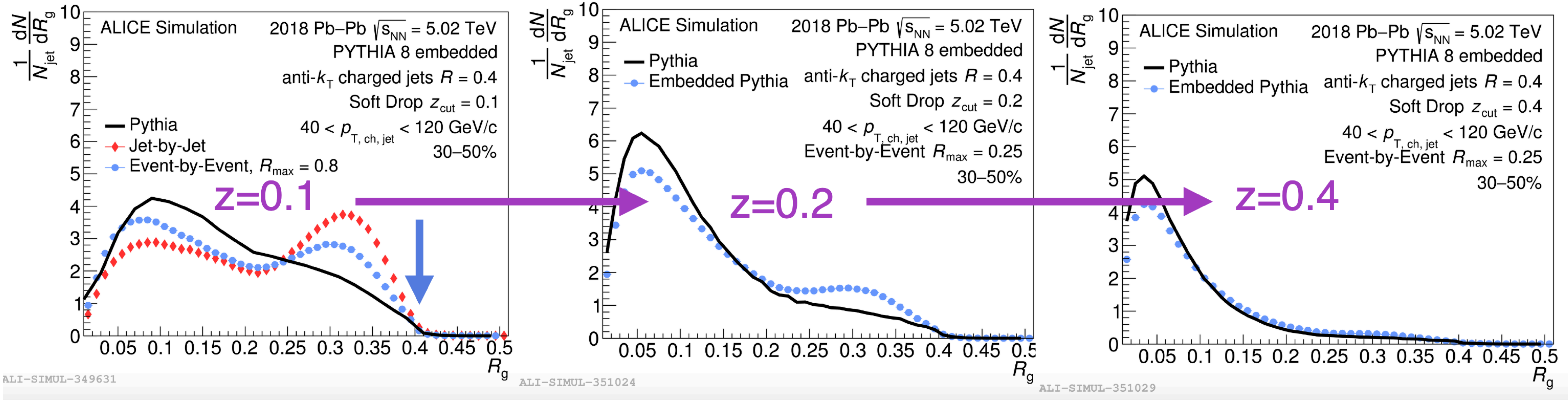


Subleading prong purity is the fraction of reconstructed subleading prongs that are connected to the true subleading prong

See big improvement when varying zcut from 0.1 to 0.2 (harder prongs, more difficult that it's just bkg)

See worse performance for dynamical grooming

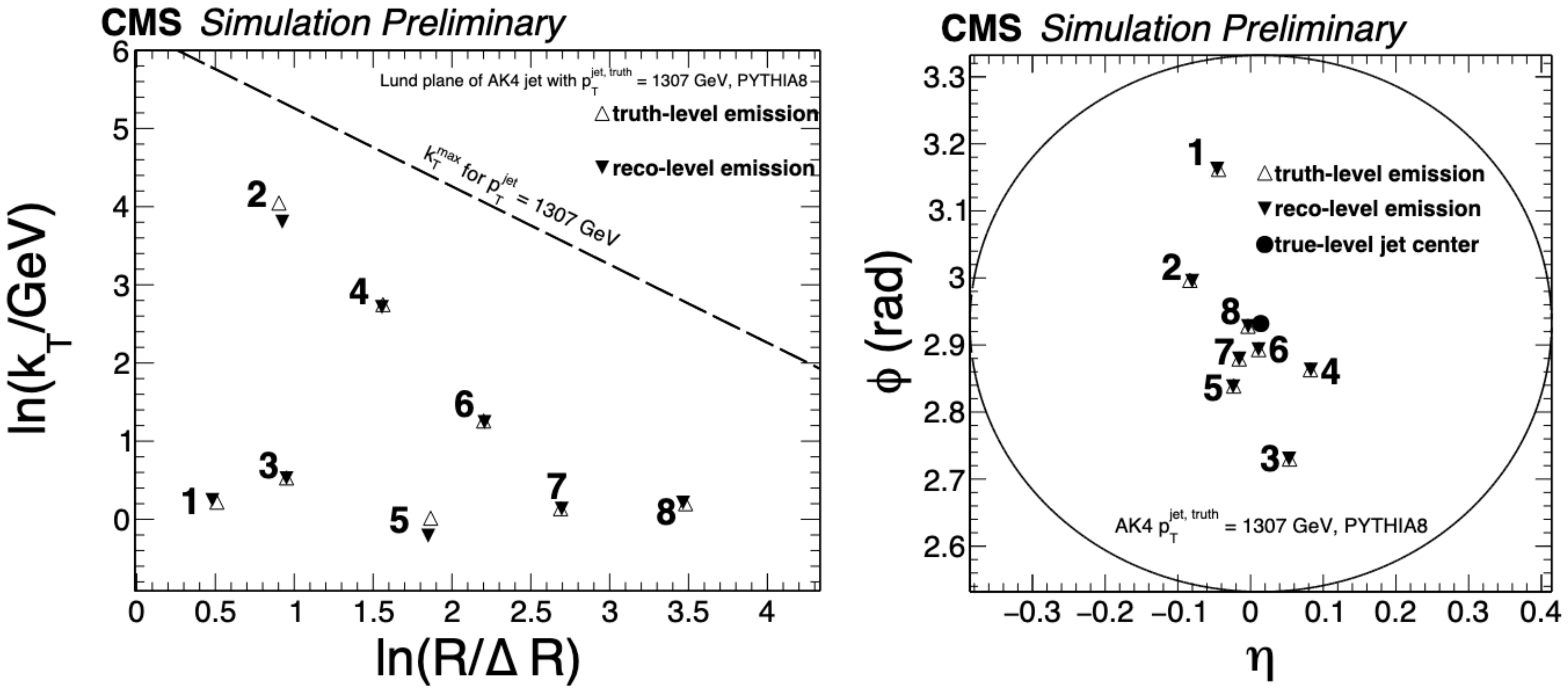
Experimental strategies in HIC



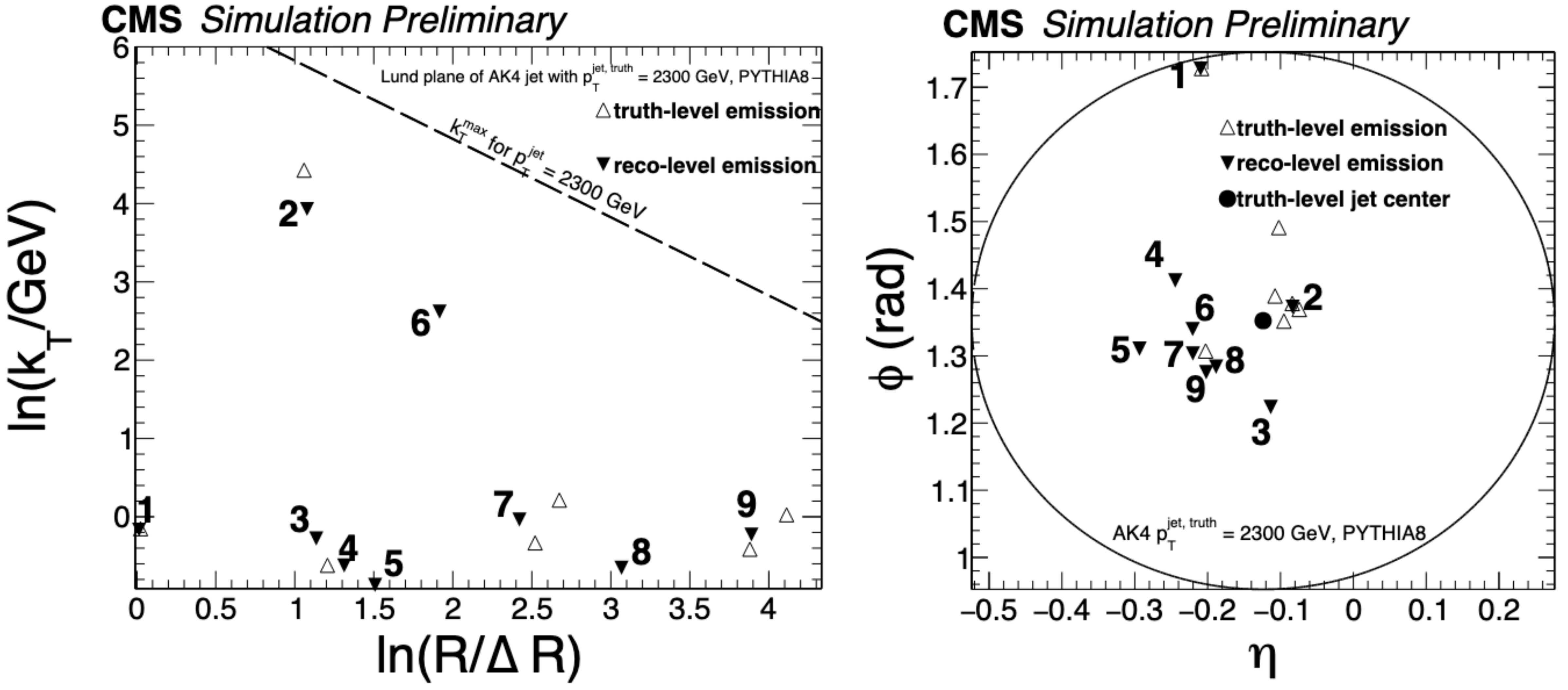
More aggressive grooming cuts reduce combinatorial prongs (but also naturally worsens statistics)

See *Laura Havener's talk at LHCP'20*

Fake splittings and missmatches in pp



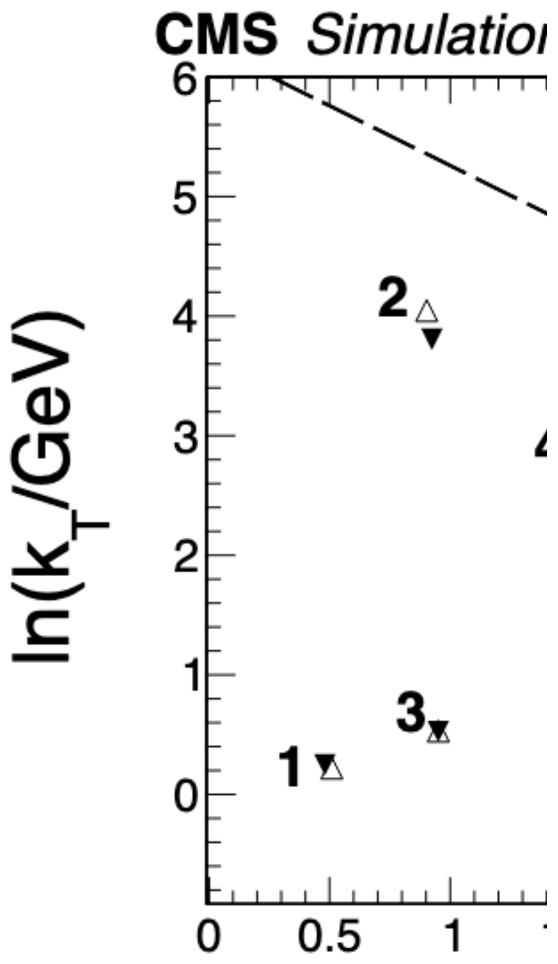
good mapping!

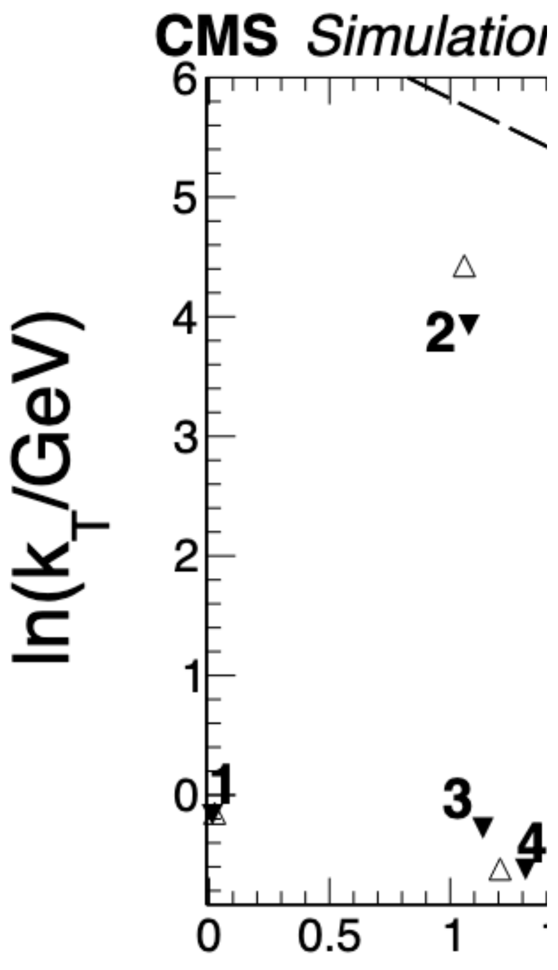


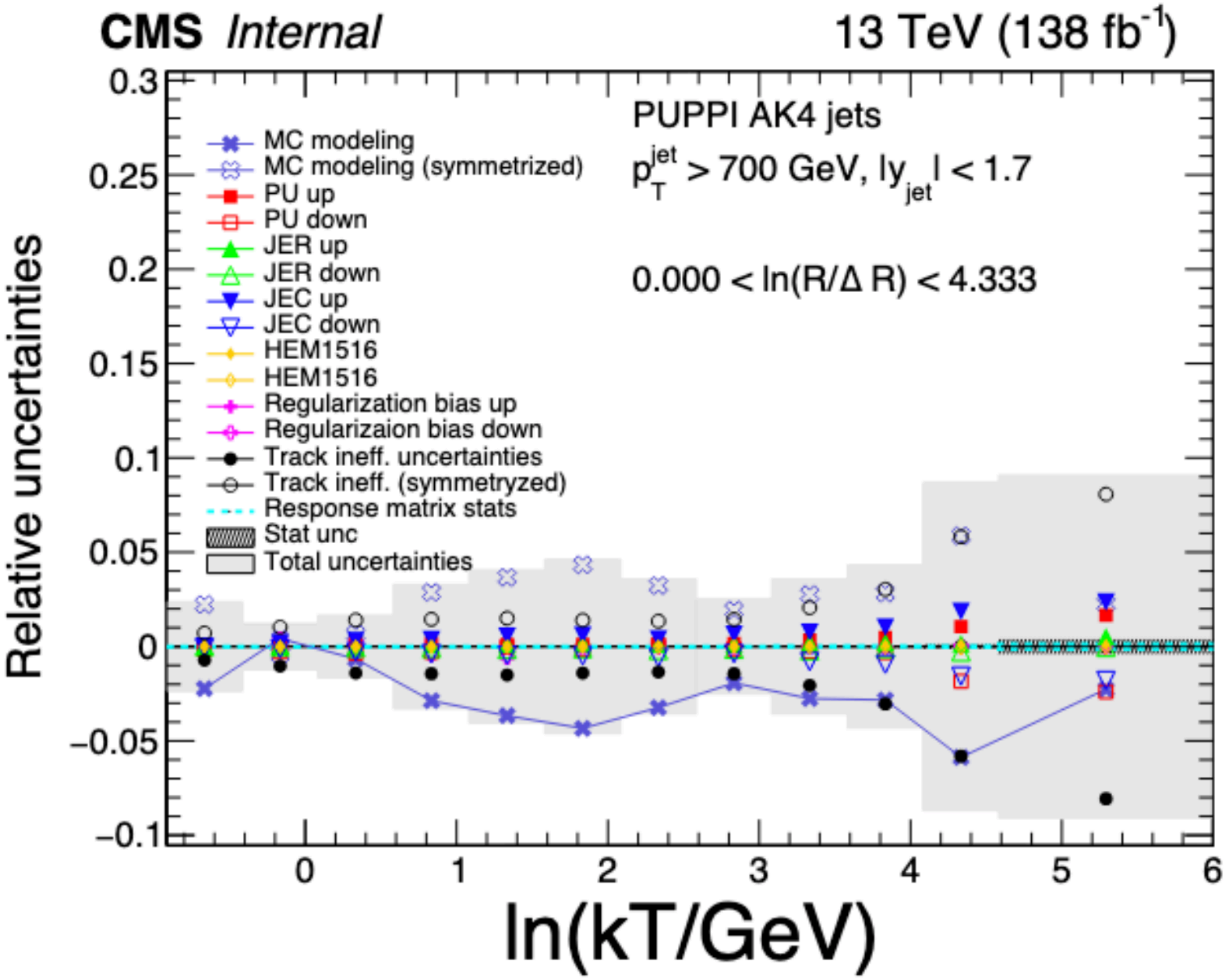
bad mapping!
all gone wrong since the begining
due to high-kT prongs being very symmetric in z
->easy to swap lead and subleaing prong at det-level
due to track.eff loses or pileup

from prelminary results of Lund plane in CMS by Cristian Baldenegro

Fake splittings and missmatches in pp



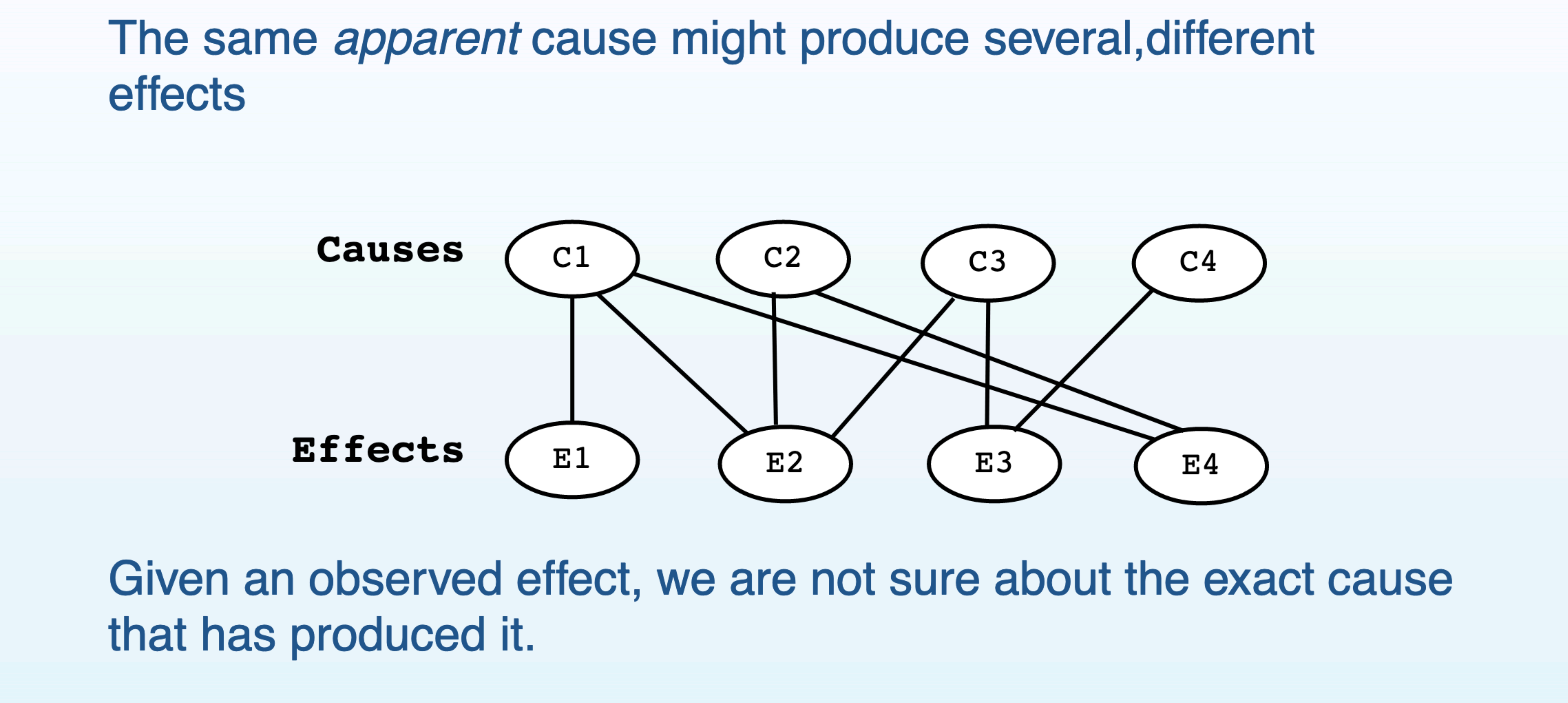




and this has consequences in terms of large sys. uncert. due to tracking ineff. uncertainty at high k_T, which is in principle the cleanest region for theory comparisons

from prelminary results of Lund plane in CMS by Cristian Baldenegro

The unfolding



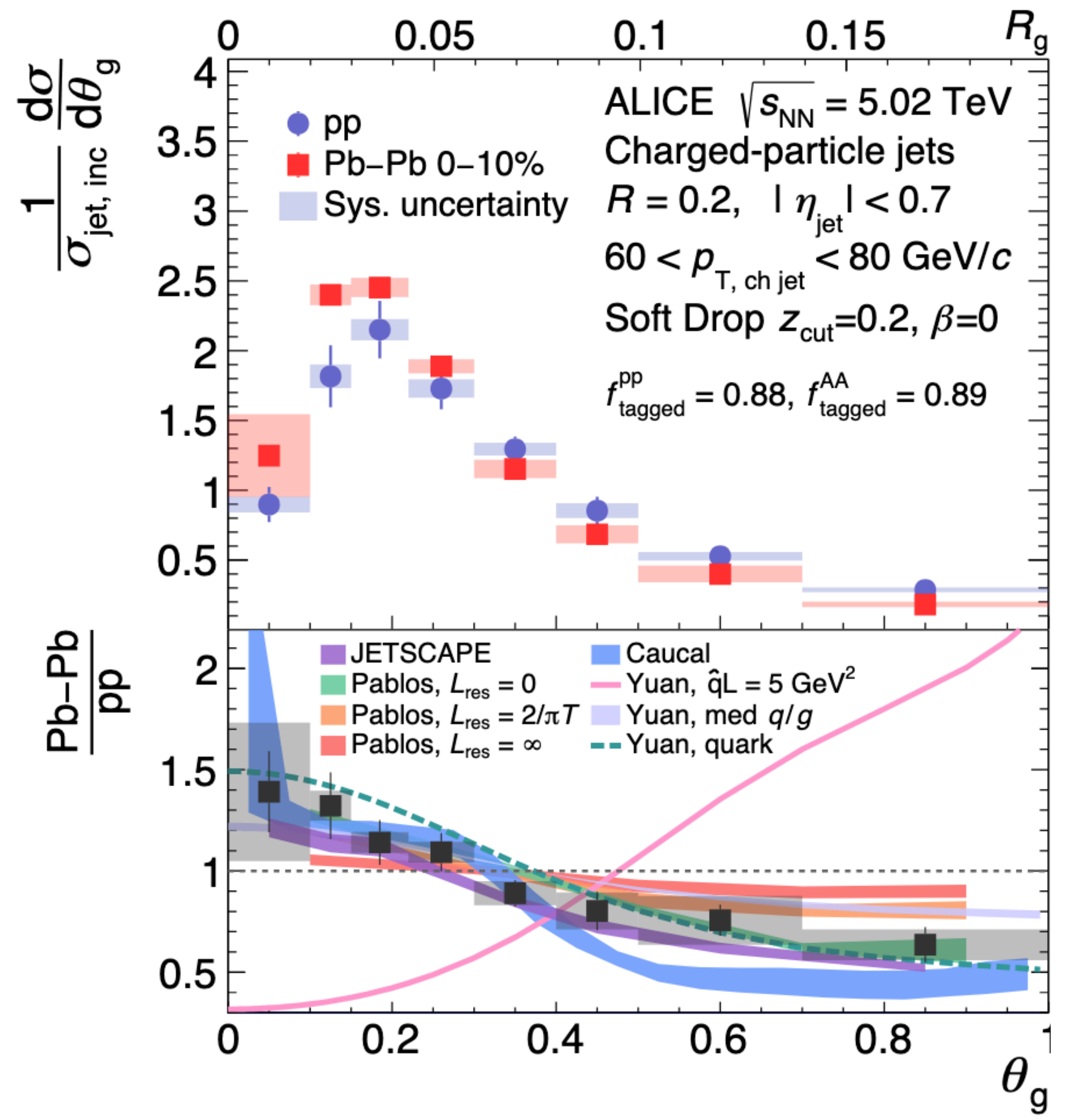
The same *apparent* cause might produce several,different effects

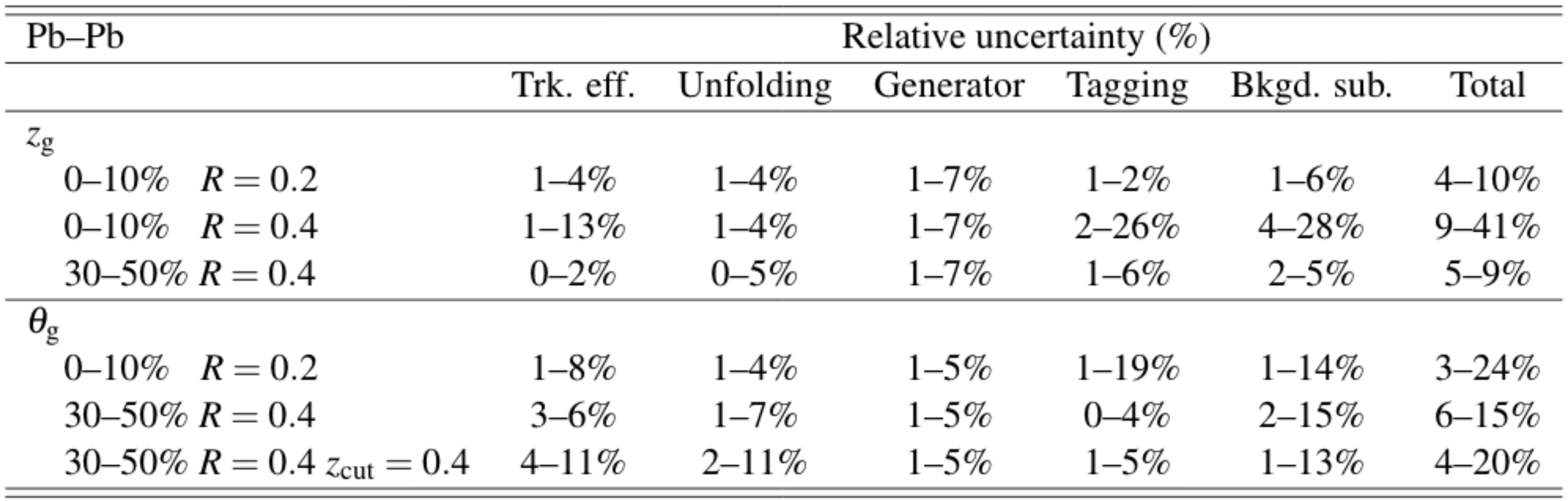
Given an observed effect, we are not sure about the exact cause that has produced it.

P(i->j) can be estimated by MC (response matrixes I’ve just shown you)
Measured_i =Sum_j P(j->i)x Truth_j ->this is forward folding, clean numerical procedure
Truth has to be guessed, several methods available, have a look to the Bayesian method which is commonly used:

Bayesian method: https://www.roma1.infn.it/~dagos/unf2_hh.pdf

The final results with systematic uncertainties





Pb–Pb	Relative uncertainty (%)					
	Trk. eff.	Unfolding	Generator	Tagging	Bkgd. sub.	Total
z_g						
0–10% $R = 0.2$	1–4%	1–4%	1–7%	1–2%	1–6%	4–10%
0–10% $R = 0.4$	1–13%	1–4%	1–7%	2–26%	4–28%	9–41%
30–50% $R = 0.4$	0–2%	0–5%	1–7%	1–6%	2–5%	5–9%
θ_g						
0–10% $R = 0.2$	1–8%	1–4%	1–5%	1–19%	1–14%	3–24%
30–50% $R = 0.4$	3–6%	1–7%	1–5%	0–4%	2–15%	6–15%
30–50% $R = 0.4$ $z_{\rm cut} = 0.4$	4–11%	2–11%	1–5%	1–5%	1–13%	4–20%

Table 1: Summary of systematic uncertainties on the Pb–Pb measurements. The ranges correspond to the minimum and maximum systematic uncertainties obtained. All values correspond to $z_{\rm cut} = 0.2$ unless otherwise noted.

See dominant effects of tracking efficiencies and model dependency