A tale of two jets

Christine Nattrass, University of Tennessee, Knoxville

Acknowledgements


Antonio Da Silva


Patrick Steffanic


Charles Hughes

What a theorist needs to know about background

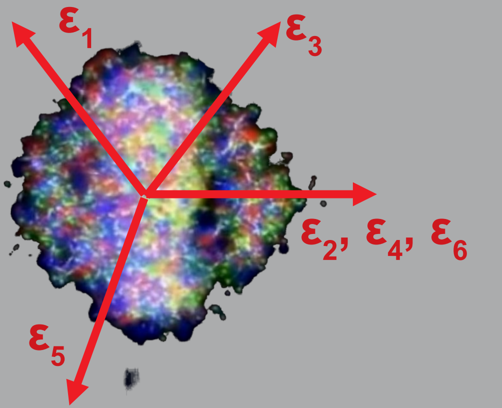
- You have background too!
- The distinction between signal and background is somewhat arbitrary
- Experimental background subtraction techniques may lead to non-trivial bias
- The gold standard is treating the model exactly like the data

Background is not just an experimental problem

arXiv:2005.02320

TennGen background generator

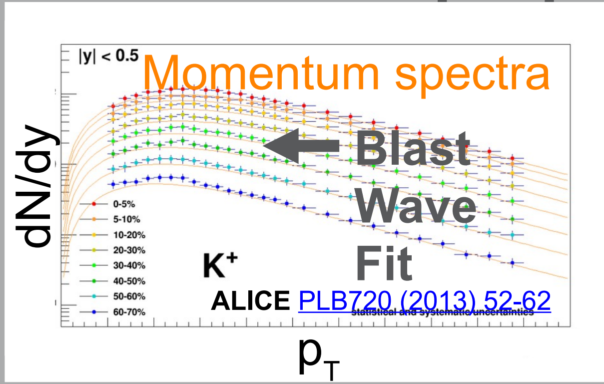
Event properties



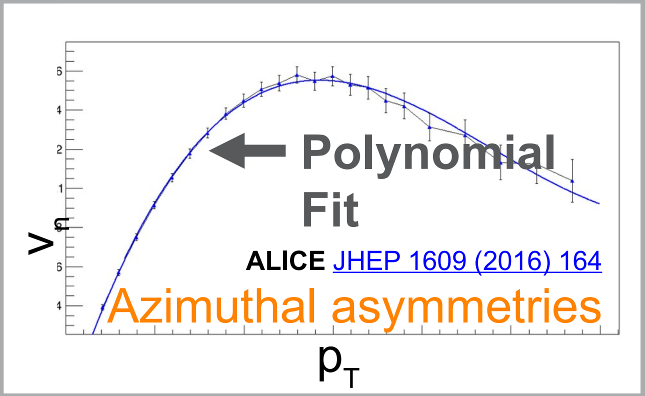
- Even event planes fixed at Ψ=0
- Odd planes at random φ
- Multiplies from ALICE [PRC88 (2013) 044910]

No jets! No resonances
Emulates hydro correlations

Track properties



→ Random p_T

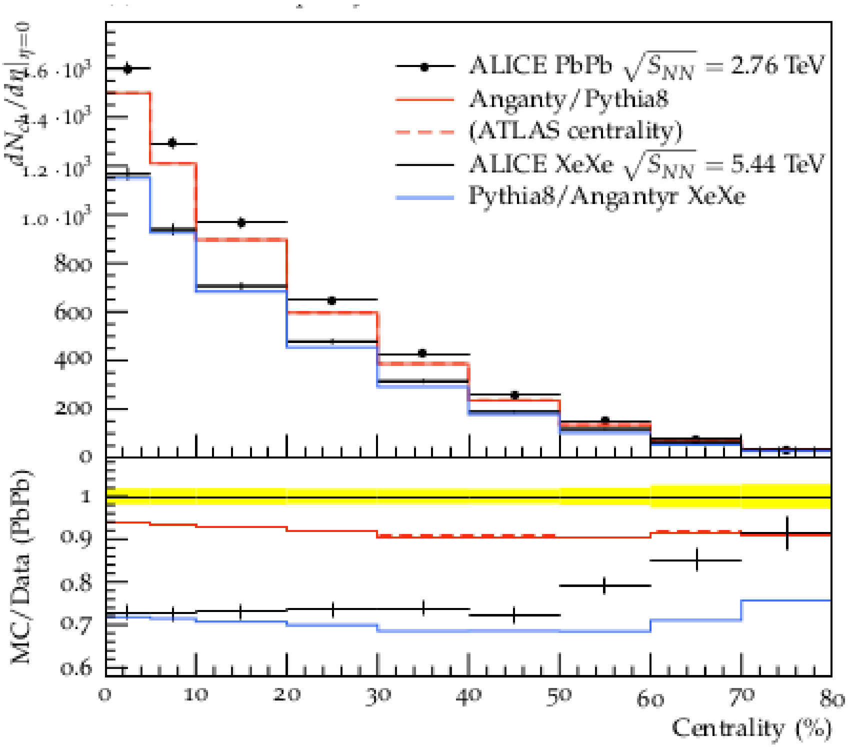


→ v_n → Random φ

PYTHIA Angantyr

JHEP (2018) 2018: 134

- Based on PYTHIA 8
  Sjöstrand, Mrenna & Skands,
  JHEP05 (2006) 026
  Comput. Phys. Comm. 178 (2008) 852.
- Based on Fritiof & wounded nucleons
- N-N collisions w/fluctuating radii → fluctuating σ



Lots of jets! And resonances!
No hydrodynamics, no jet quenching

Area-based background subtraction

Cacciari & Salam, [PLB659:119–126,2008]

Particles, clusters

↓

k_T algorithm

$$k_T = p_T, \Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$$

- For all *i,j* calculate:

$$d_{ij} = min(p_{T,i}^2, p_{T,j}^2)\,\Delta R_{ij}^2$$

$$d_{iB} = p_{T,i}$$

- Combine smallest d_{ij}.

  If d_{iB} smallest, $d_{iB} \rightarrow$ jet

Repeat until no particles left

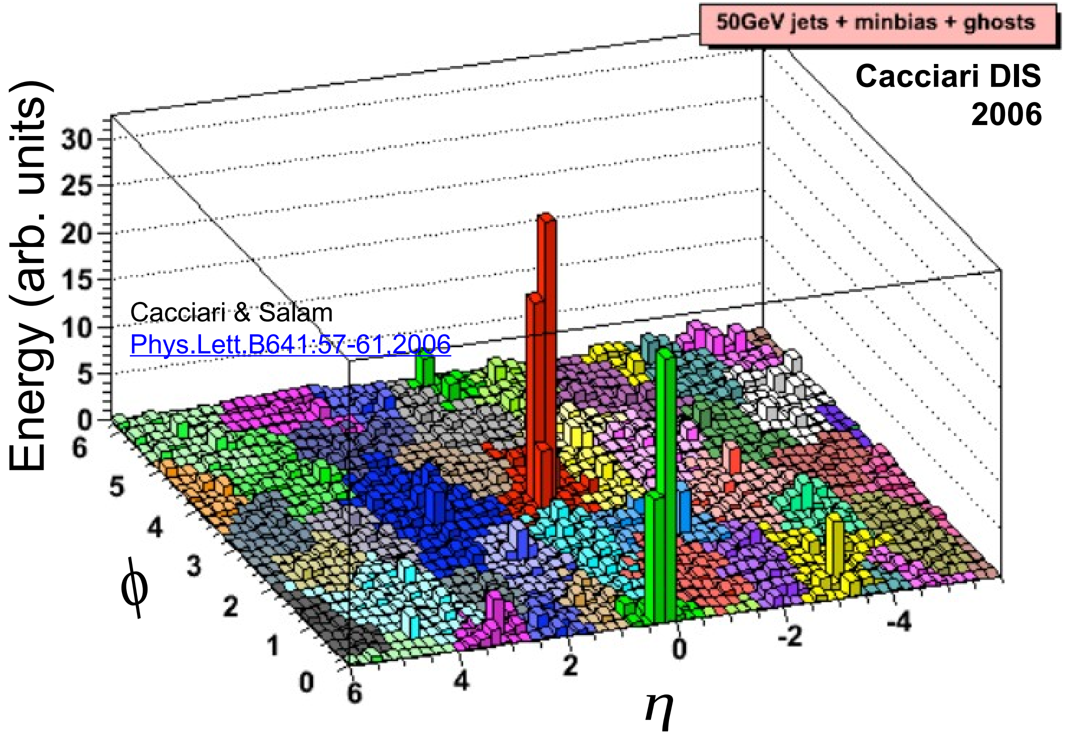
↓

Jet candidates

↓

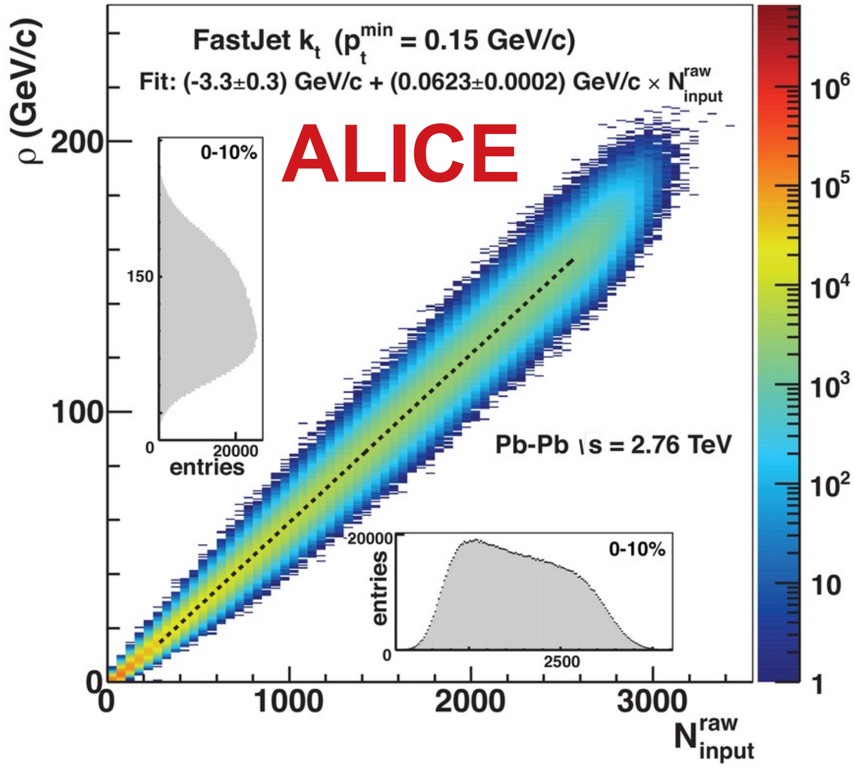
Median $\rho = p_T/A$

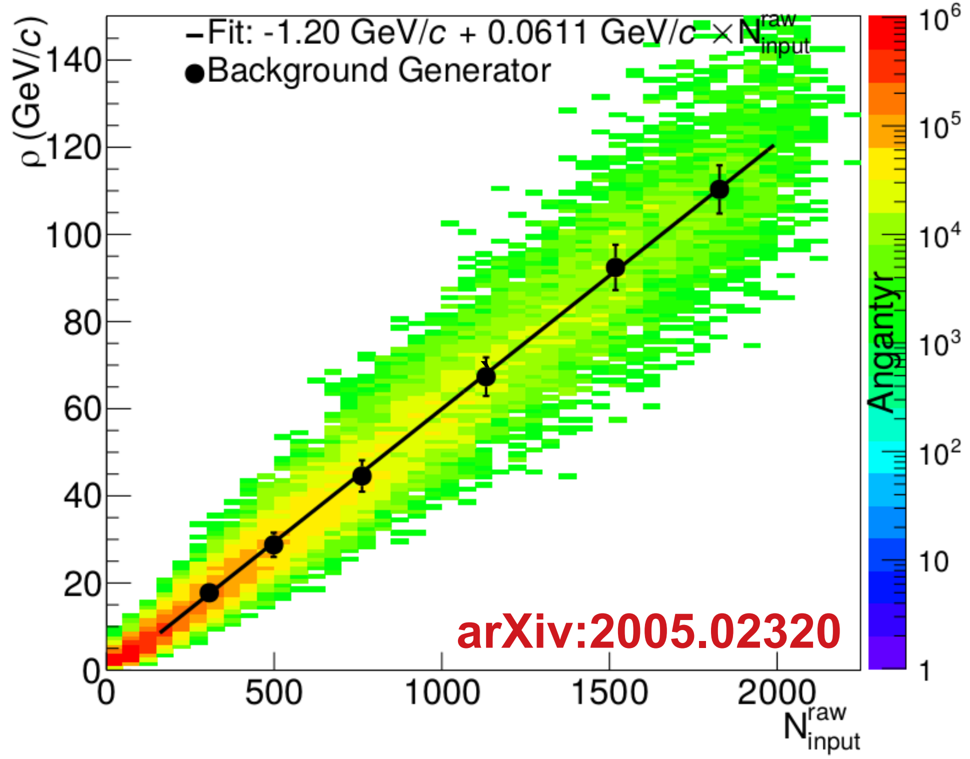
$$p_T^{jet} = p_T^{reco} - \rho_{median} A^{jet}$$



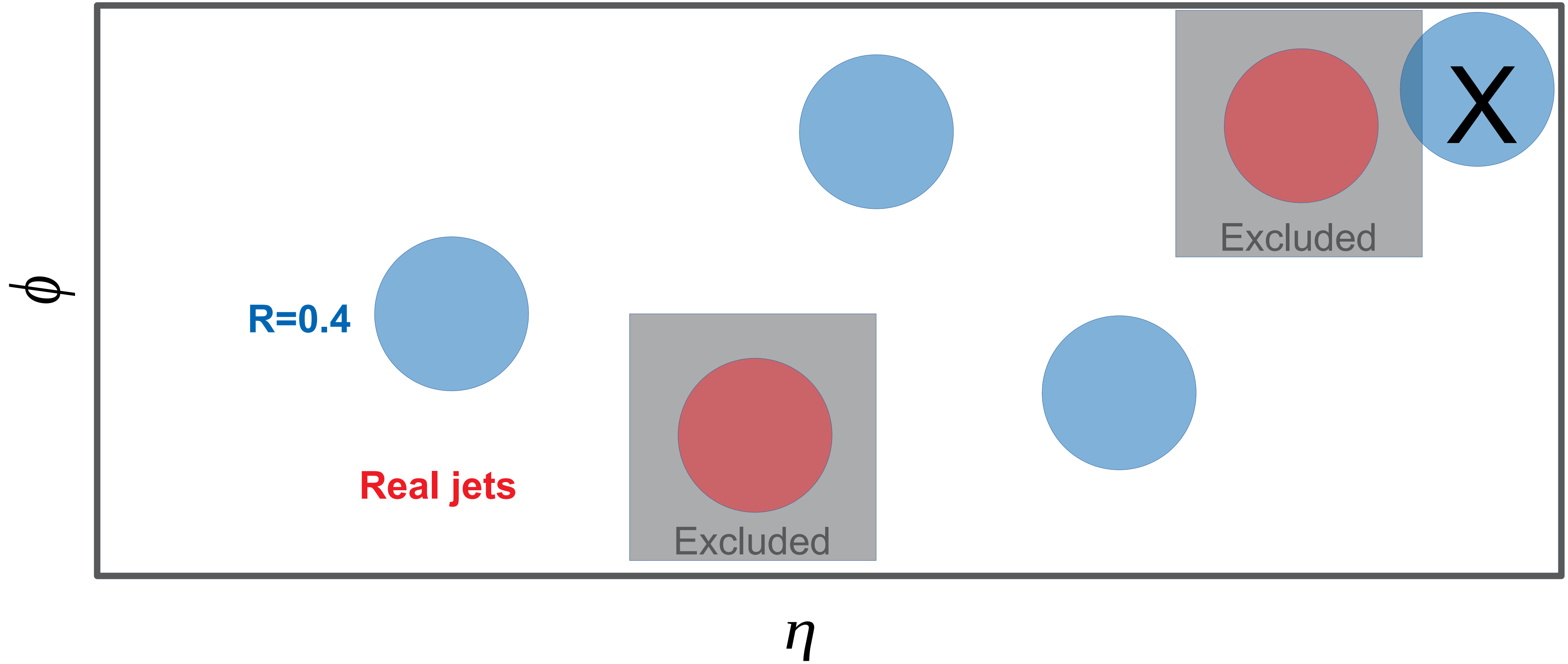
Cacciari & Salam
[Phys.Lett.B641:57-61,2006]

Background density ρ





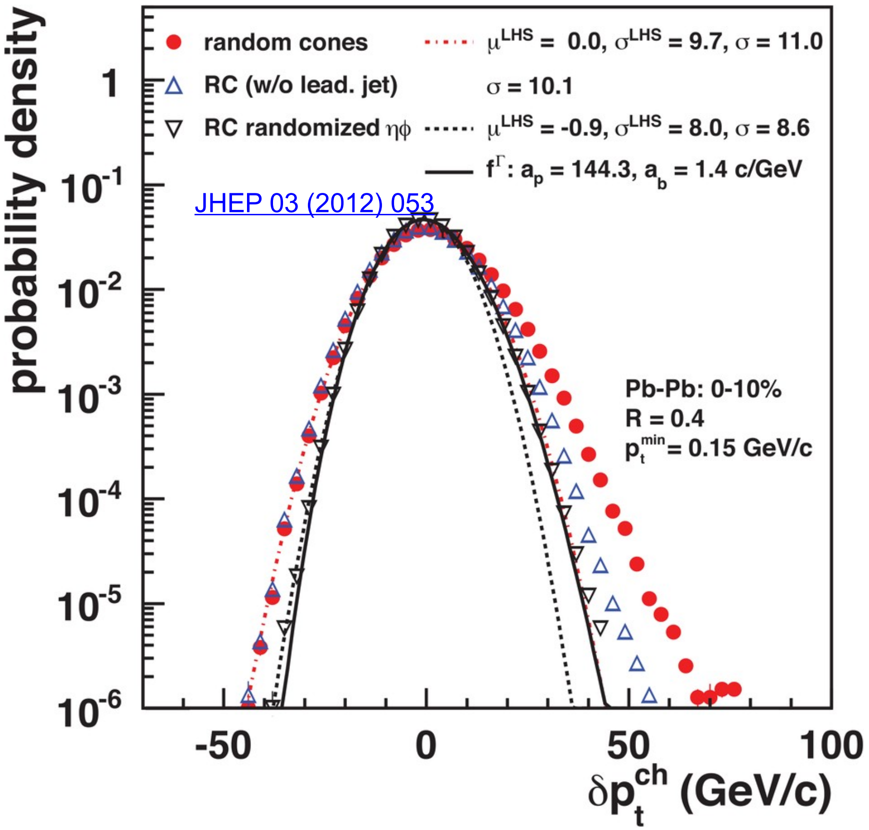
Random cones



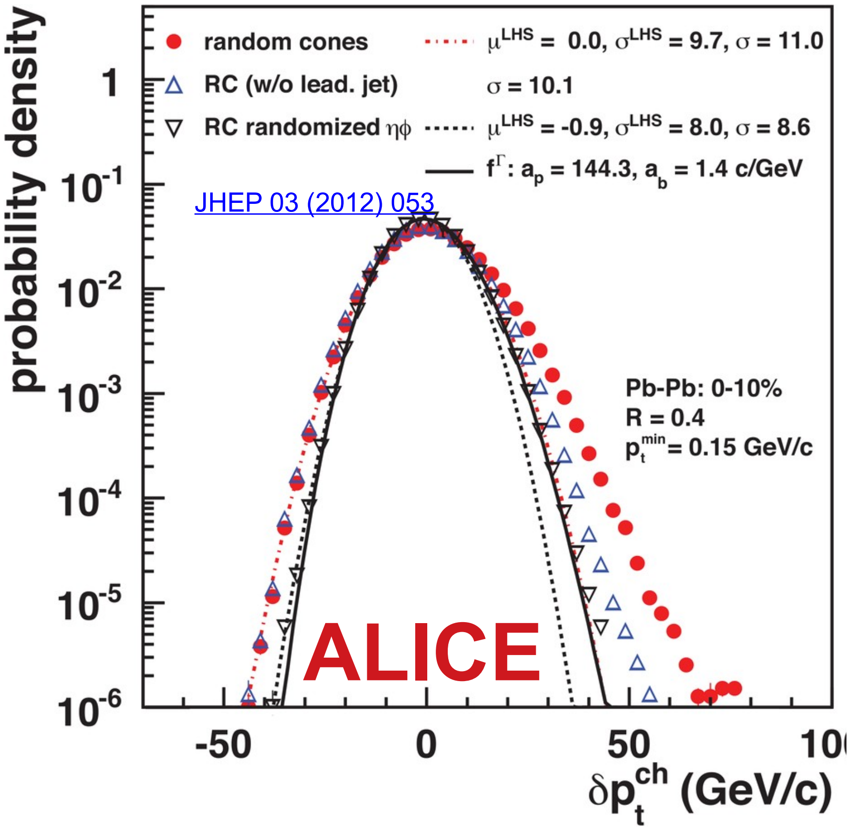
Random cones in ALICE

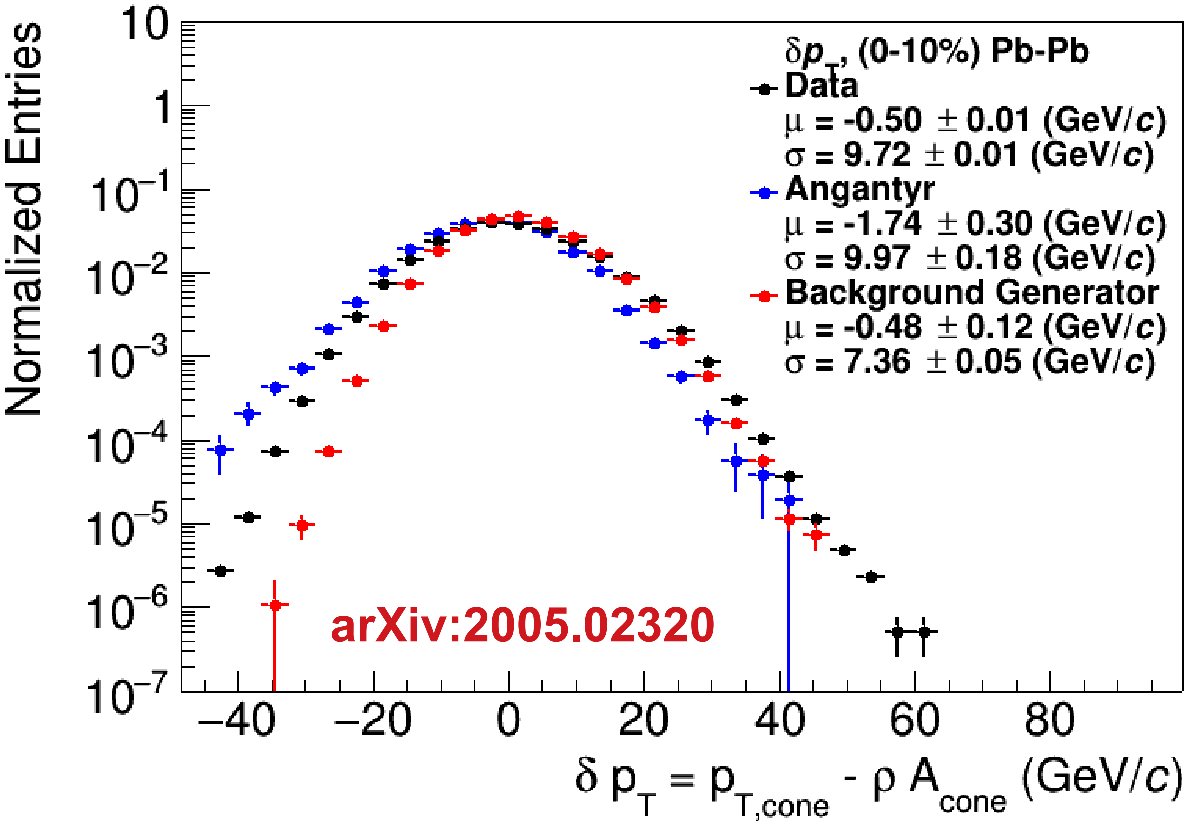
- Estimate ρ
  - k_T jet finder→jet candidates
  - ρ = Median(p_T/A)
- Draw Random cone

$$\delta p_T = p_T^{reco} - \rho A$$



Random cones





Shape of width of the distribution

Single particle spectra

$$f_\Gamma(p_T, p, b) = \frac{b}{\Gamma(p)}(b\,p_T)^{p-1} e^{-bx}$$

$$\frac{dN}{dy} \propto f_\Gamma(p_T, 2, b) = b^2 p_T e^{-k\,p_T}$$

$$\mu_{p_T} = \frac{p}{b}, \sigma_{p_T} = \frac{\sqrt{p}}{b}$$

Tannenbaum, PLB(498),1–2,Pg.29-34(2001)

Σp_T of N particles→N-fold convolution:

$$f_N(p_T, p, b) = f_\Gamma(p_T, Np, b) \qquad \frac{dpT^{total}}{dy} \propto f_N(p_T, Np, b)$$

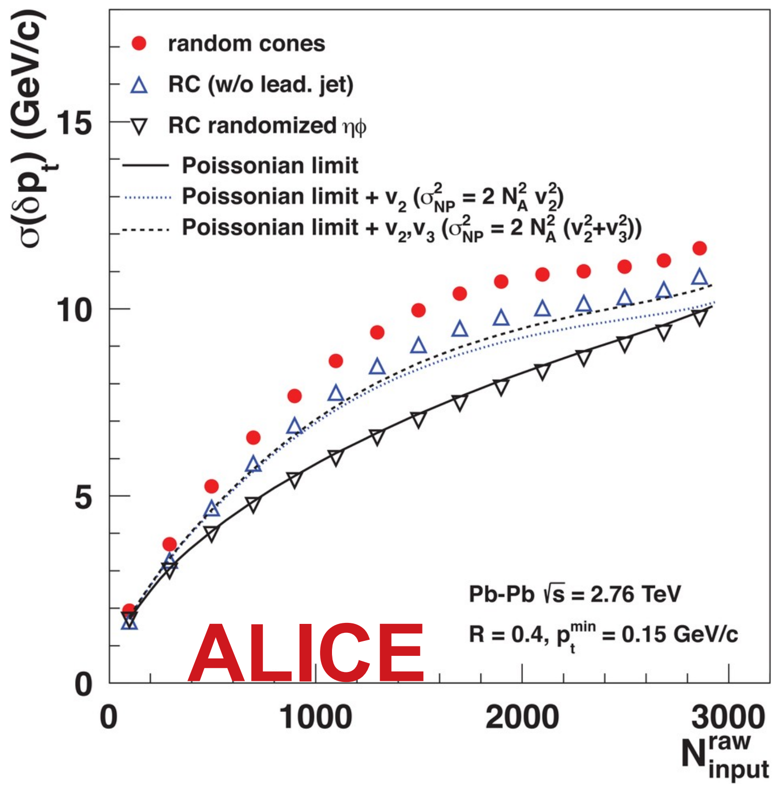
$$N = \frac{N_{total}}{A_{total}} \pi R^2 \qquad \mu_{total} = \frac{Np}{b} = N\,\mu_{p_T}, \sigma_{total} = \frac{\sqrt{Np}}{b} = \sqrt{N}\,\sigma_{p_T}$$

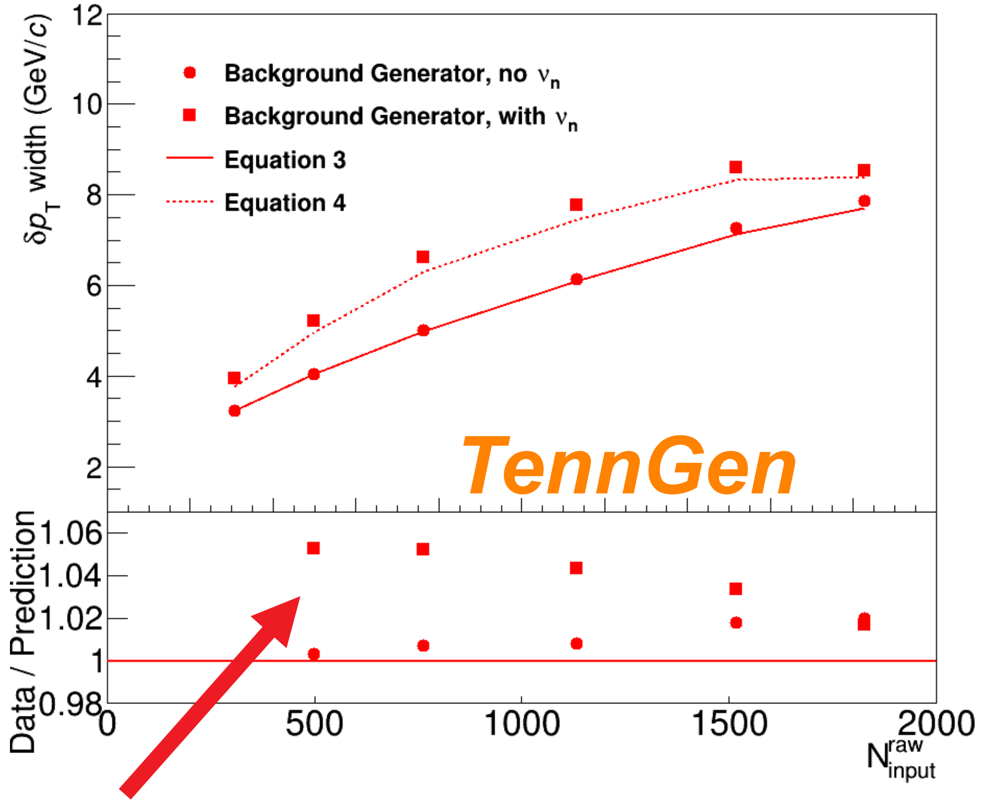
Add Poissonian fluctuations in N: $\sigma_{total} = \sqrt{N\,\sigma_{p_T}^2 + N\,\mu_{p_T}^2}$

Add non-Poissonian fluctuations in N due to flow

$$\sigma_{total} = \sqrt{N\,\sigma_{p_T}^2 + \left(N + 2\sum v_n^2\right)\mu_{p_T}^2}$$

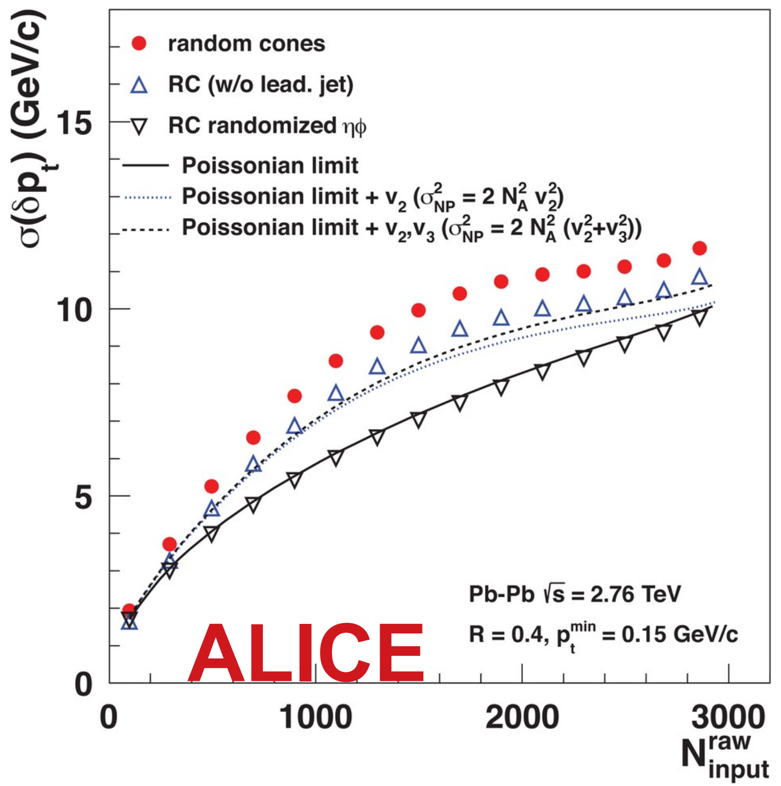
Width vs multiplicity

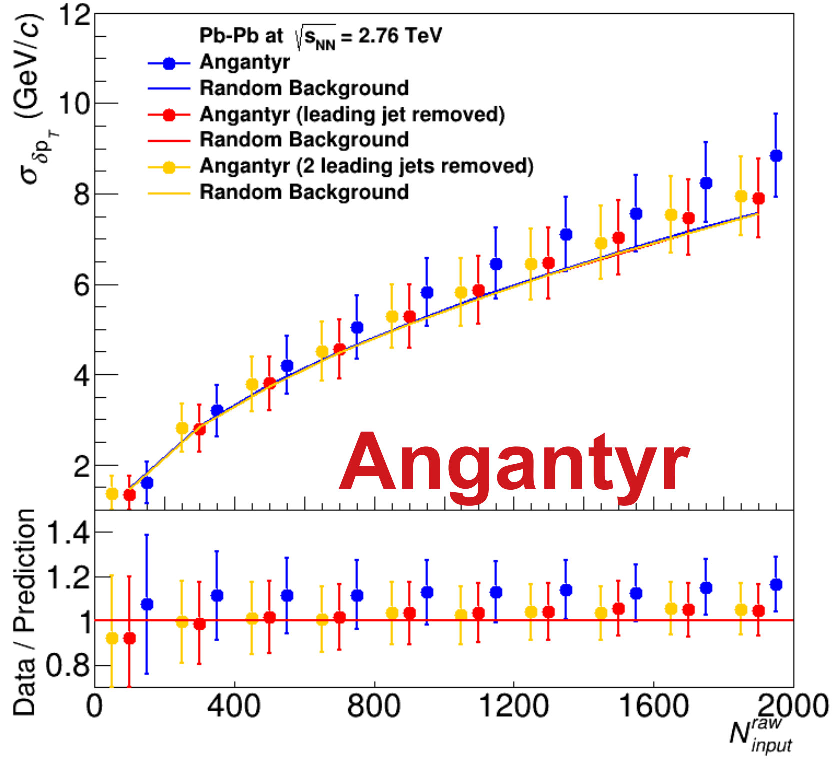




Small deviations

Width vs multiplicity





Shape of width of the distribution

Single particle spectra

$$f_\Gamma(p_T, p, b) = \frac{b}{\Gamma(p)}(b\,p_T)^{p-1} e^{-bx}$$

$$\frac{dN}{dy} \propto f_\Gamma(p_T, 2, b) = b^2 p_T e^{-k\,p_T}$$

$$\mu_{p_T} = \frac{p}{b}, \sigma_{p_T} = \frac{\sqrt{p}}{b}$$

Tannenbaum, PLB(498),1–2,Pg.29-34(2001)

Assumes shape

Σp_T of N particles→N-fold convolution:

$$f_N(p_T, p, b) = f_\Gamma(p_T, Np, b) \qquad \frac{dpT^{total}}{dy} \propto f_N(p_T, Np, b)$$

$$N = \frac{N_{total}}{A_{total}}\pi R^2 \qquad \mu_{total} = \frac{Np}{b} = N\mu_{p_T}, \sigma_{total} = \frac{\sqrt{Np}}{b} = \sqrt{N}\sigma_{p_T}$$

Add Poissonian fluctuations in N: $\sigma_{total} = \sqrt{N\sigma_{p_T}^2 + N\mu_{p_T}^2}$

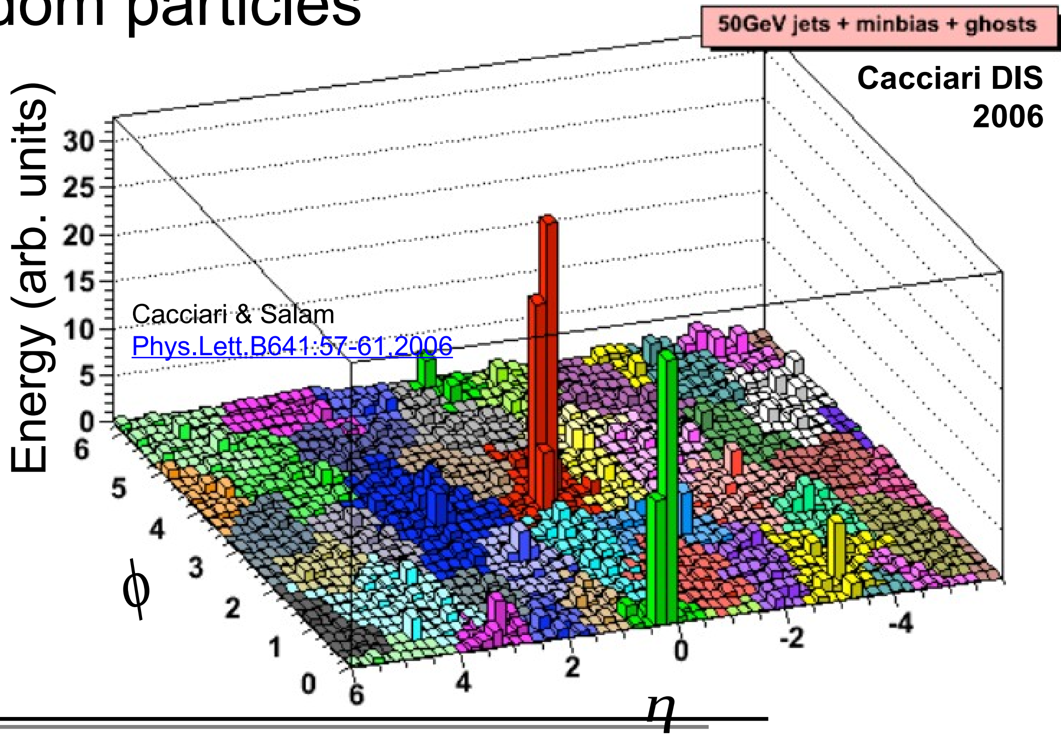
Add non-Poissonian fluctuations in N due to flow

$$\sigma_{total} = \sqrt{N\sigma_{p_T}^2 + \left(N + 2\sum v_n^2\right)\mu_{p_T}^2}$$

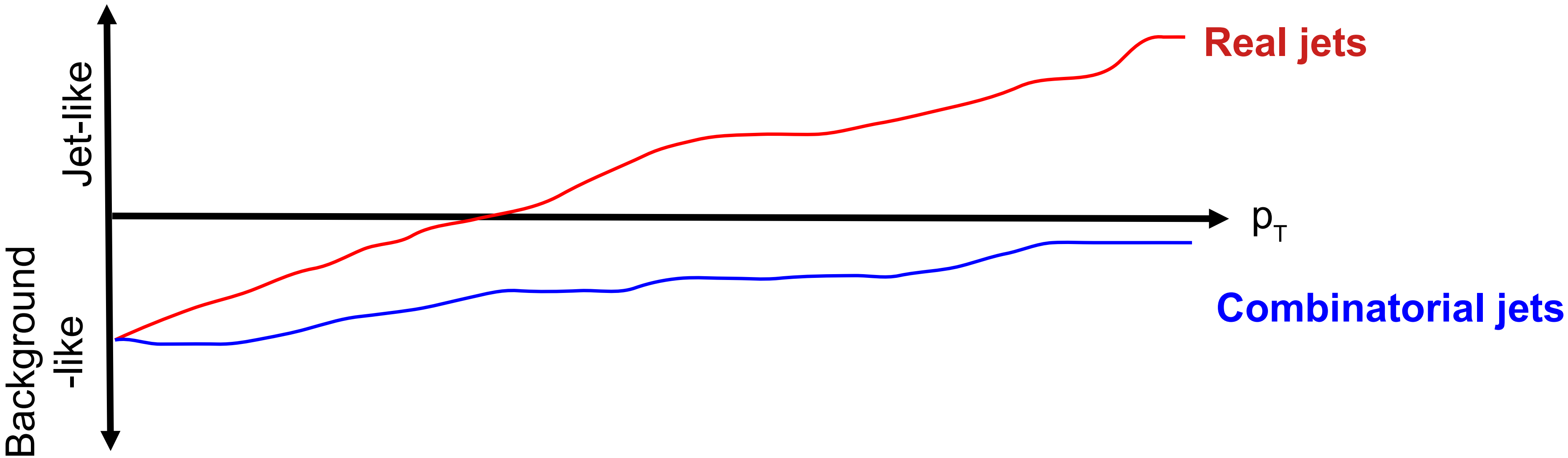
Assumes uncorrelated number fluctuations

Mini-summary

- Jet finders put all input clusters, tracks in a jet candidate
- Background is *dominated* by random particles
  - But ~5% effects from non-Poissonian fluctuations
- Models have background too!
  - Sensitive to multiplicity, implementation of flow

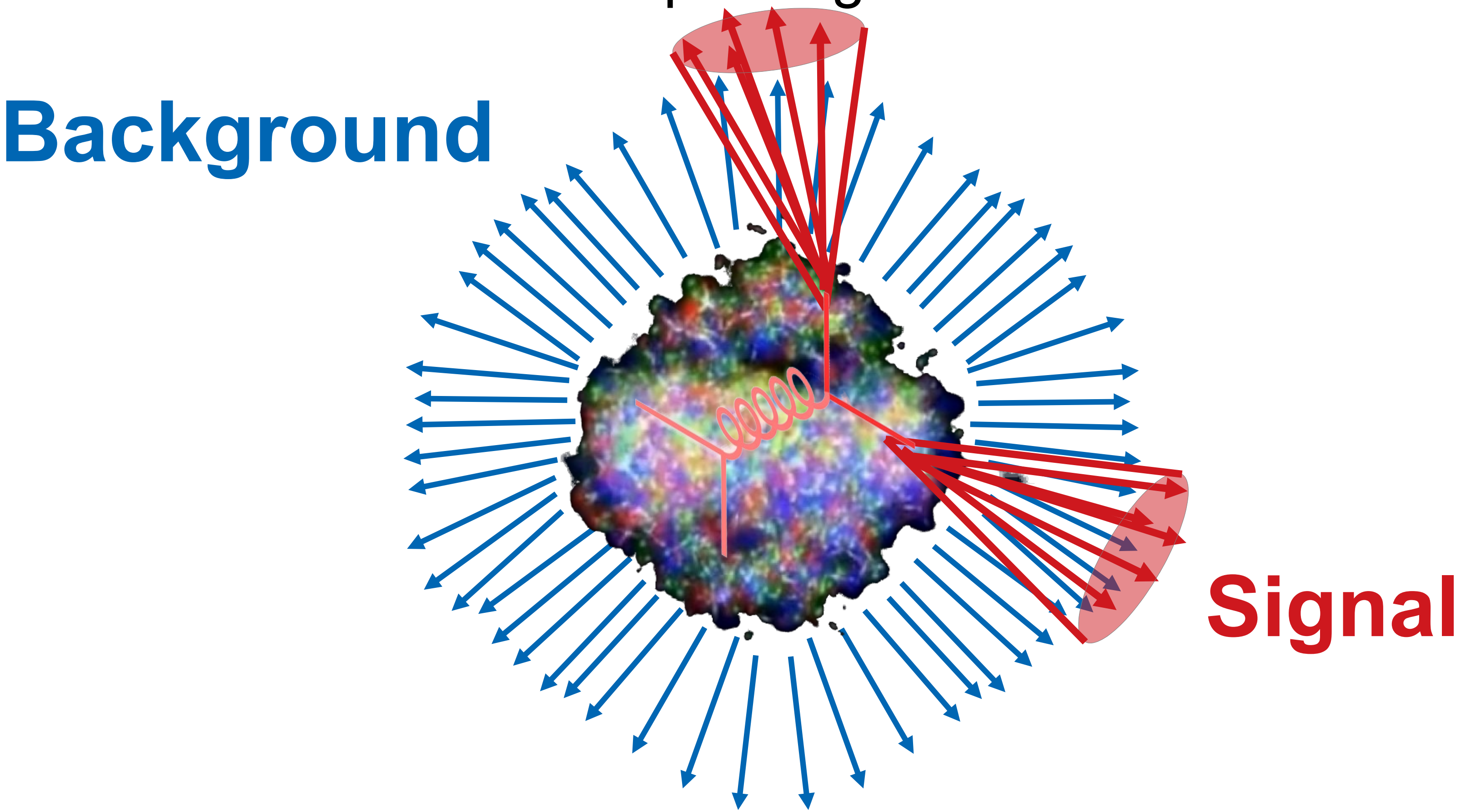


Signal and background overlap



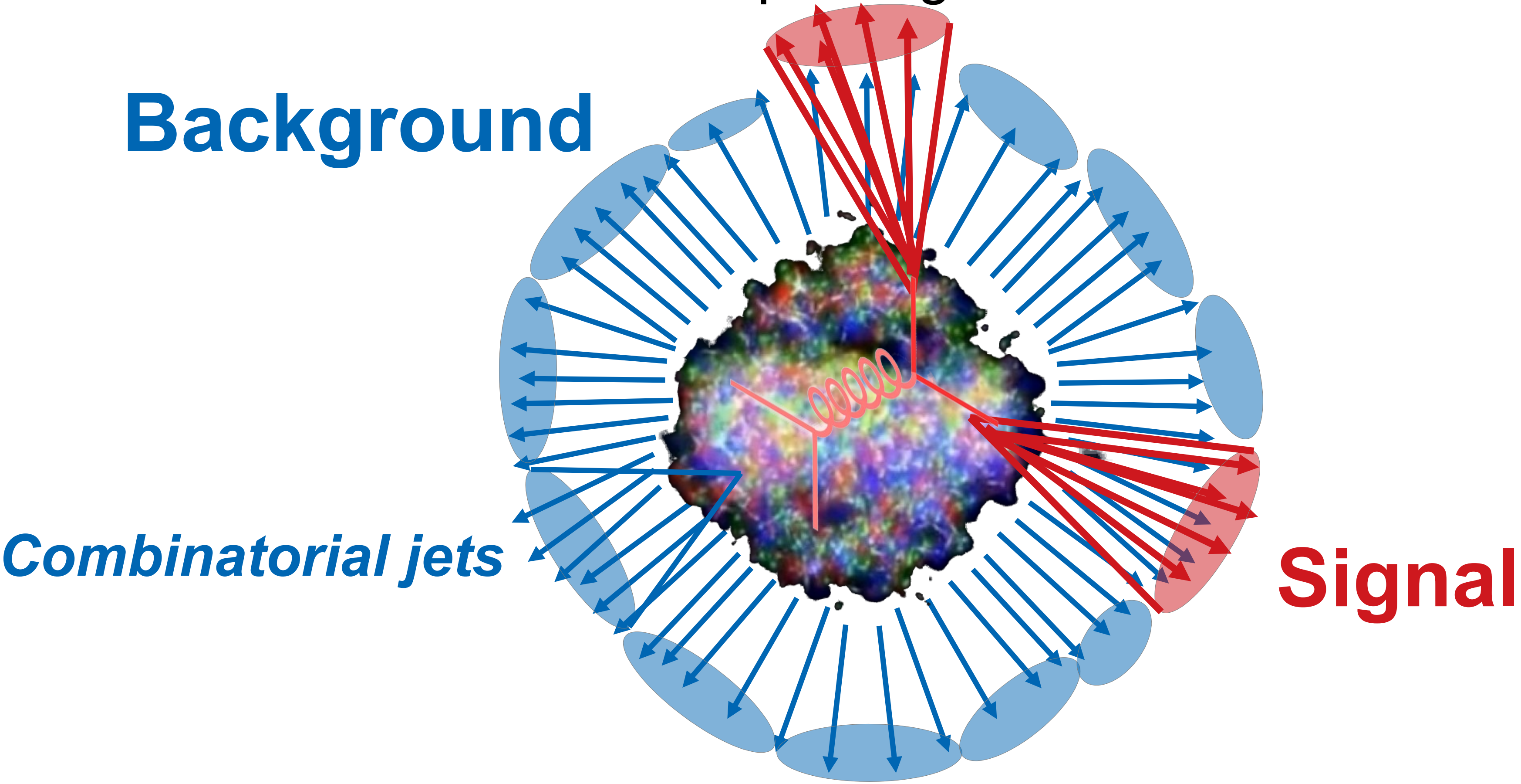
Signal vs Background:

The standard paradigm



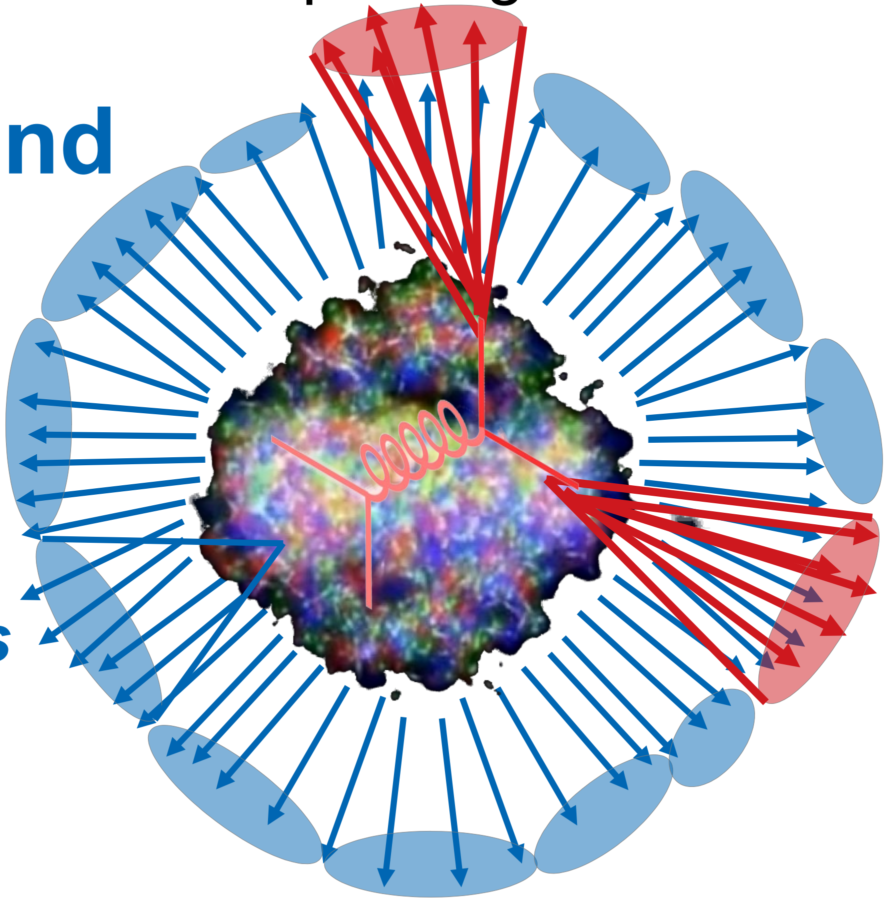
Signal vs Background:

The standard paradigm



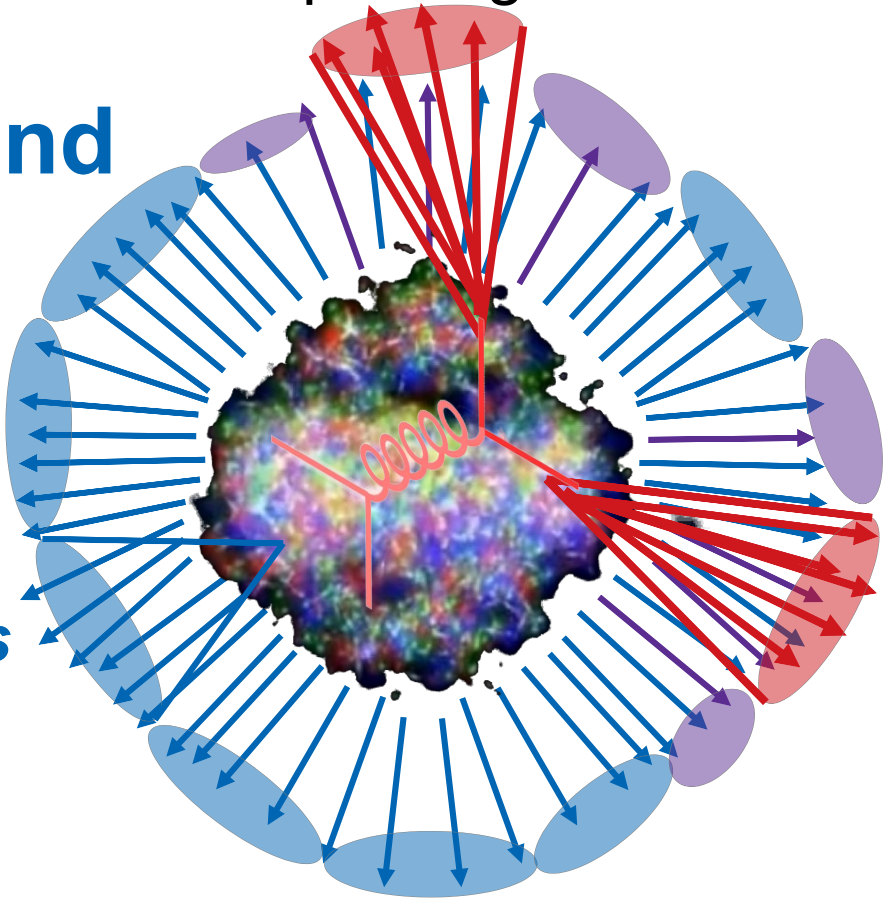
Signal vs Background:

The standard paradigm



Signal vs Background:

The standard paradigm



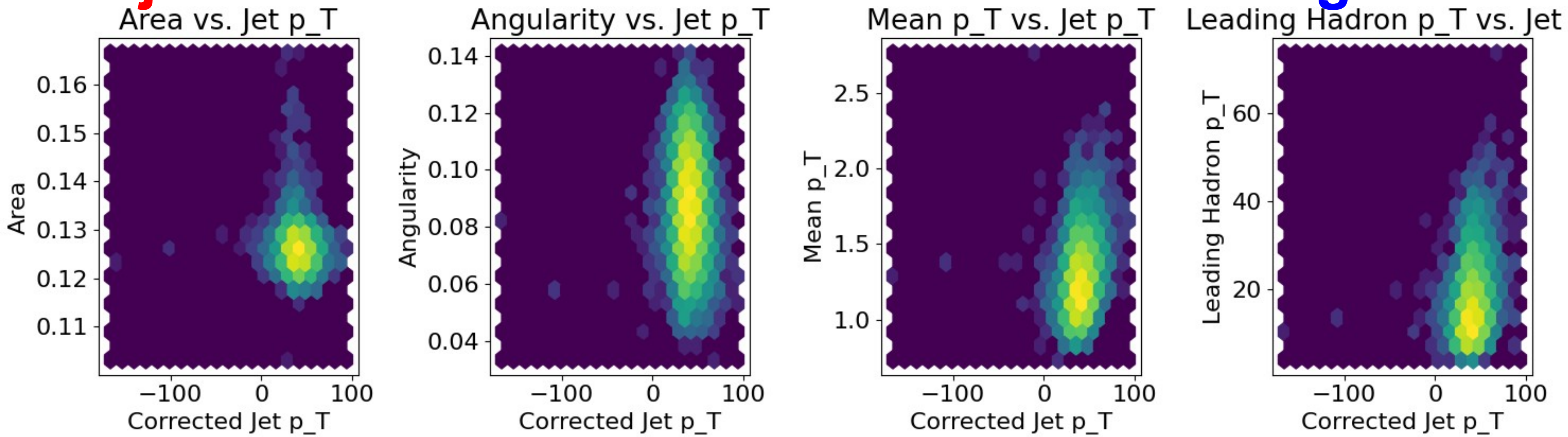
Technique

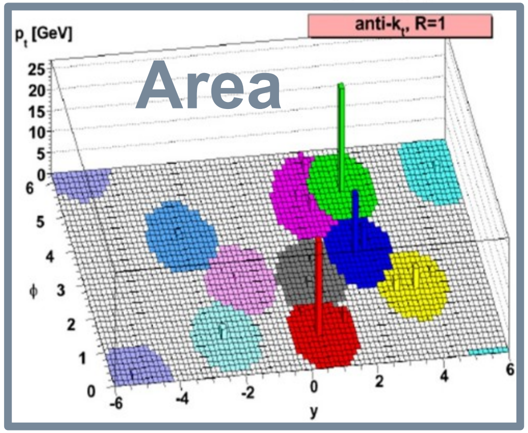
- Anti-k_T jet finder, $|\eta_{jet}|<0.5$
- **Combinatorial jets:** Only contain TennGen particles
- **Real jets:** Add a PYTHIA pp event. Real jets contain >80% of p_{Thard}^{min}

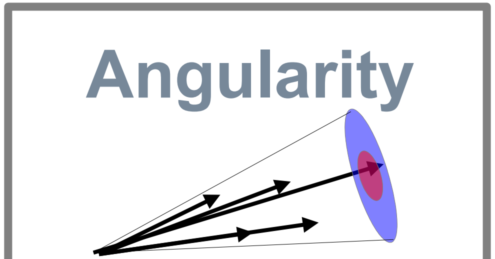
Real jets

R = 0.2, p_T hard min = 40

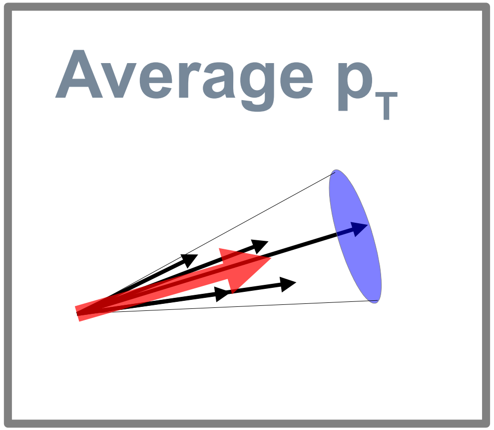
Log z scale

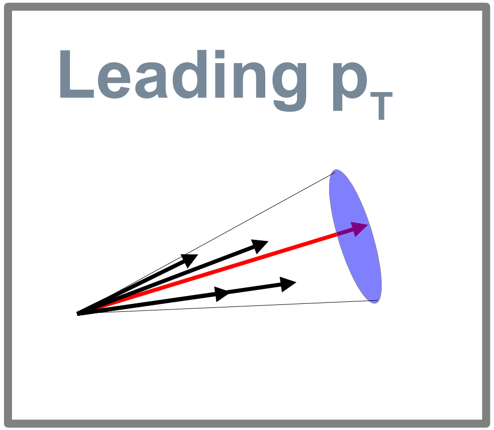






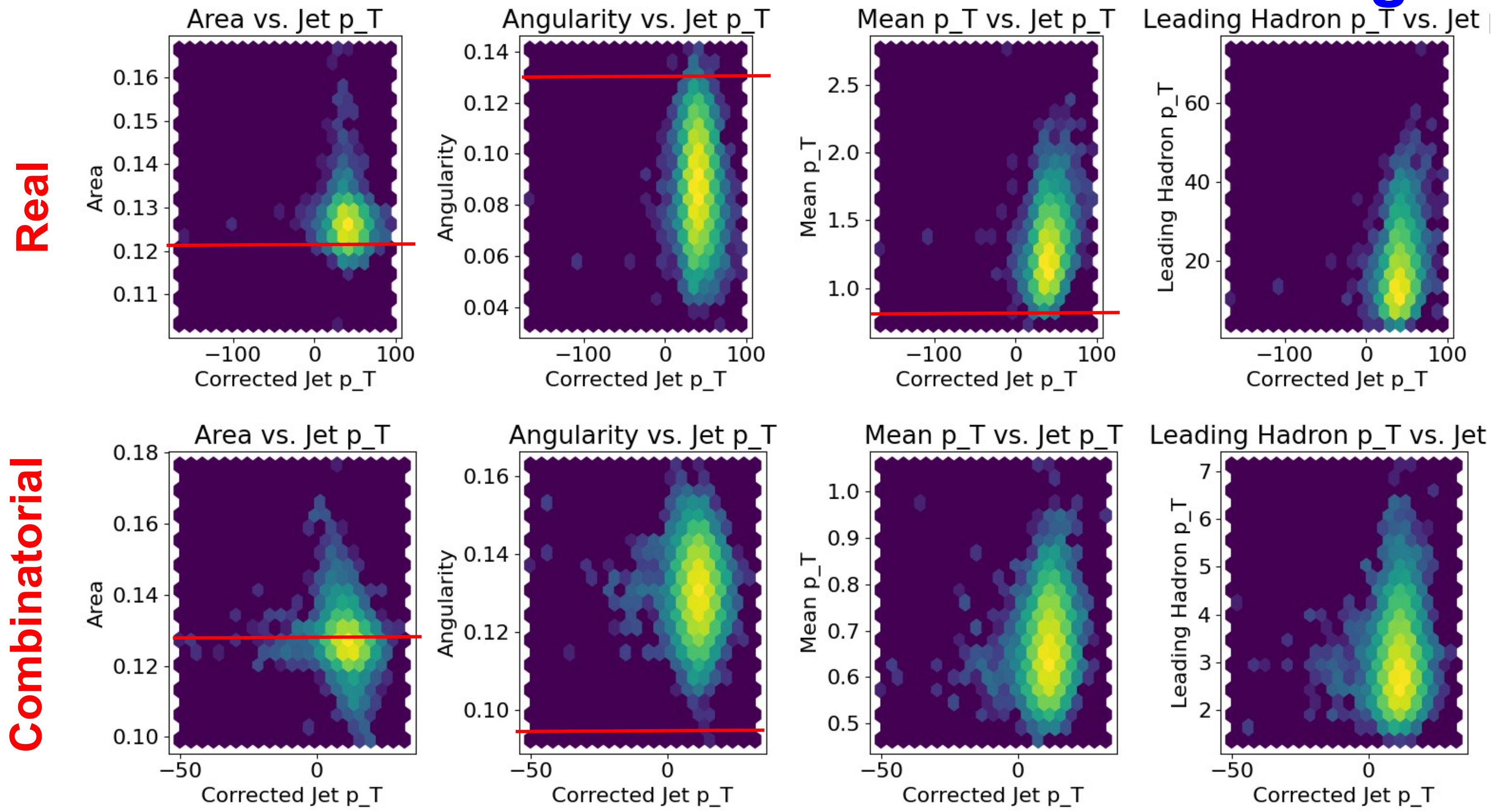
$$\alpha = \frac{1}{p_T^{jet}} \sum z_k \left(\vec{R}_k \right)$$





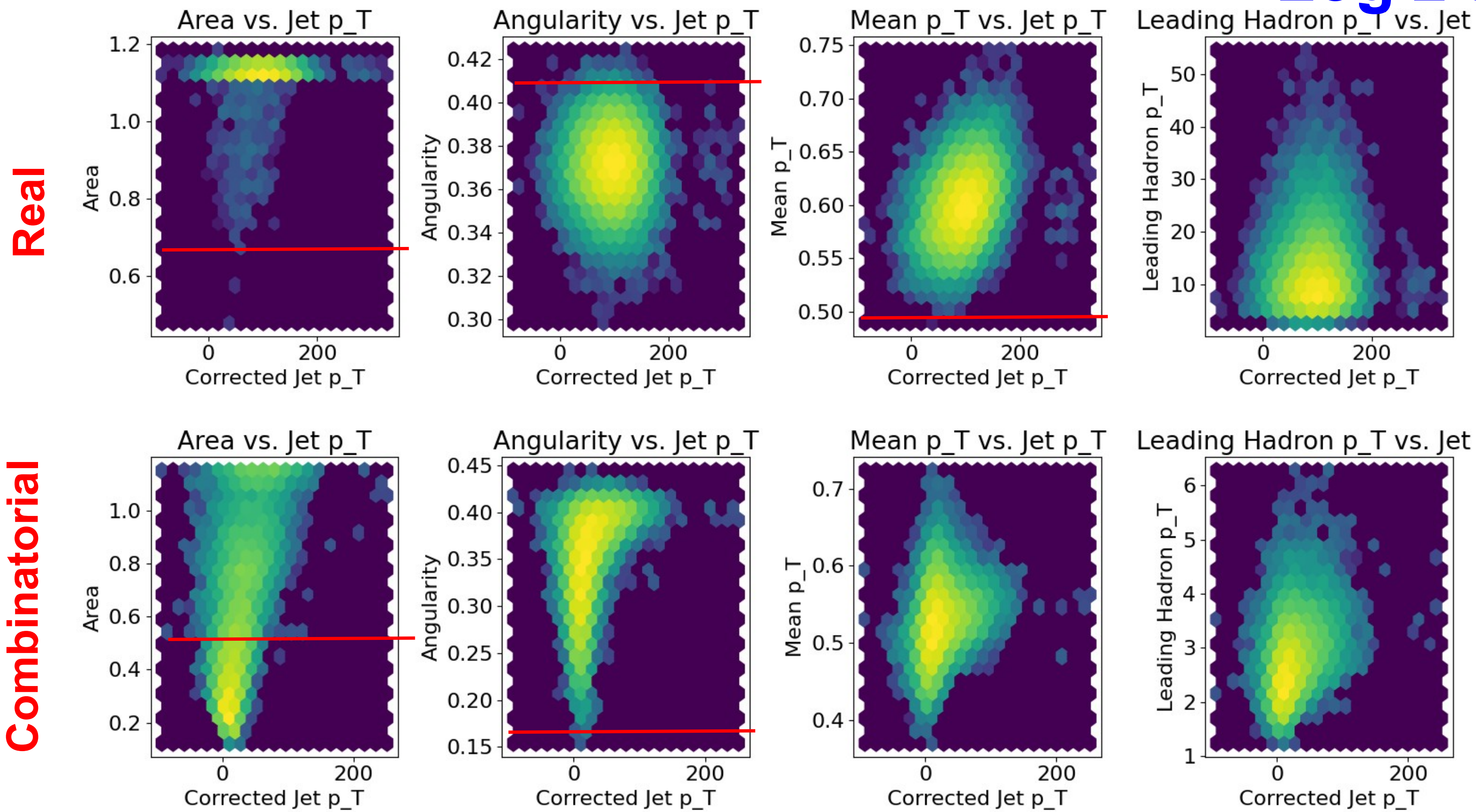
$p_{T hard}$>40 GeV/c, R=0.2

Log z scale



p_{Thard}>40 GeV/*c*, R=0.6

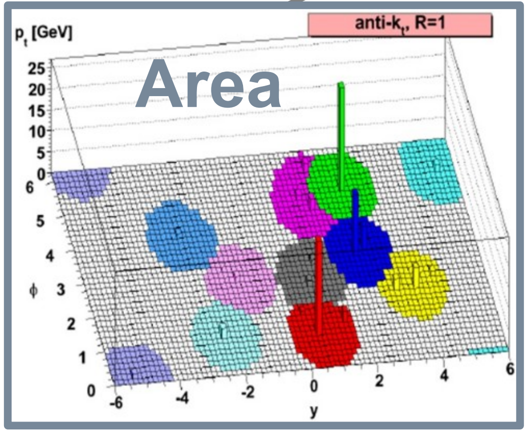
Log z scale

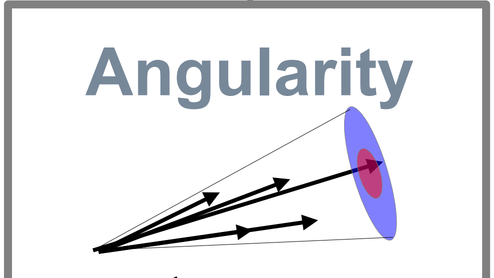


Silhouette Values

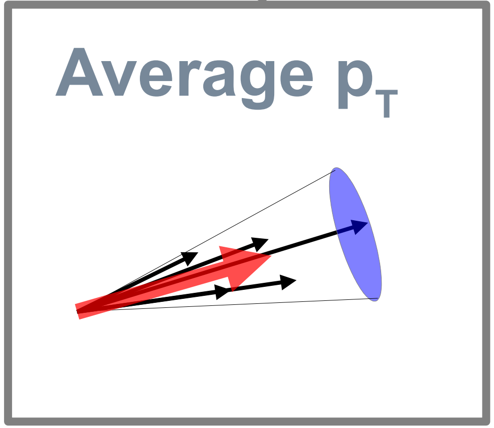
- Define a distance between two jet candidates to determine how similar they are

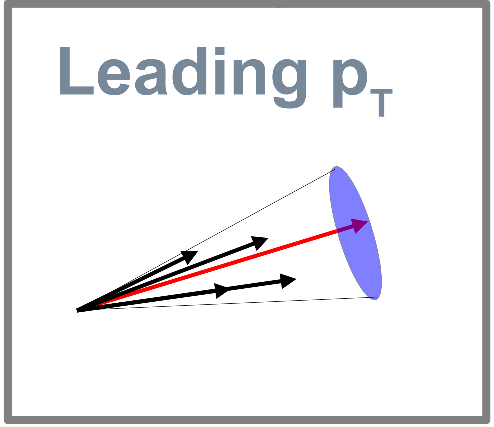
$$d_{i,j}=\sqrt{\left(\frac{A_i-A_j}{A^{max}-A^{min}}\right)^2+\left(\frac{\alpha_i-\alpha_j}{\alpha^{max}-\alpha^{min}}\right)^2+\left(\frac{\langle p_T\rangle_i-\langle p_T\rangle_j}{\langle p_T\rangle^{max}-\langle p_T\rangle^{min}}\right)^2+\left(\frac{p_{T,i}^L-p_{T,j}^L}{p_T^{L,max}-p_T^{L,min}}\right)^2}$$





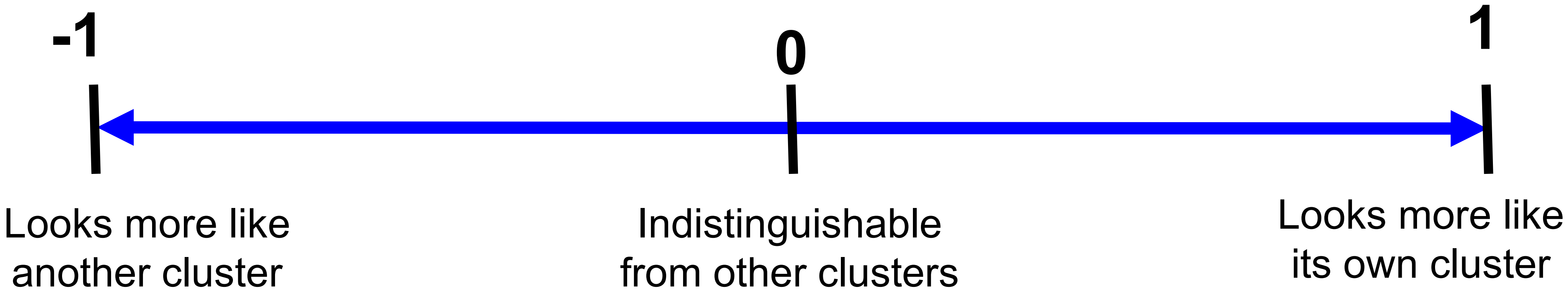
$$\alpha=\frac{1}{p_T^{jet}}\sum z_k\left(\vec{R}_k\right)$$





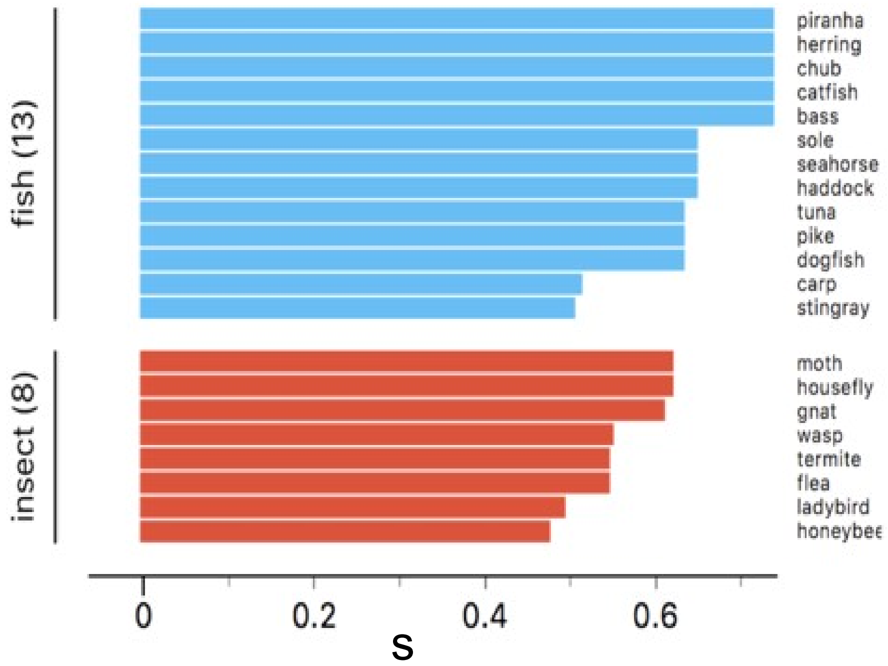
Silhouette Values

- Average distance between a jet candidate and other jet candidates in its cluster (signal or background) $a_i = \langle d_{i,j} \rangle_{j \neq i}$
- Average distance between jet candidate and jet candidates in the other cluster $b_i = \langle d_{i,j} \rangle$
- Silhouette value $$s_i = \frac{b_i - a_i}{max[b_i, a_i]}$$

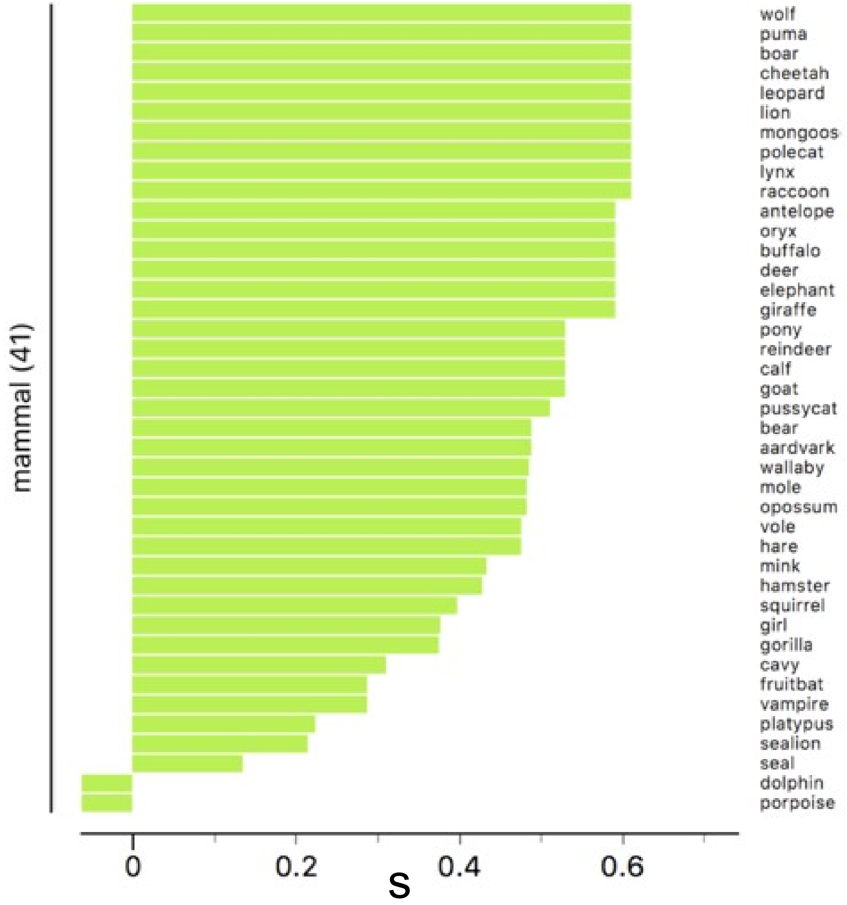


Silhouette values

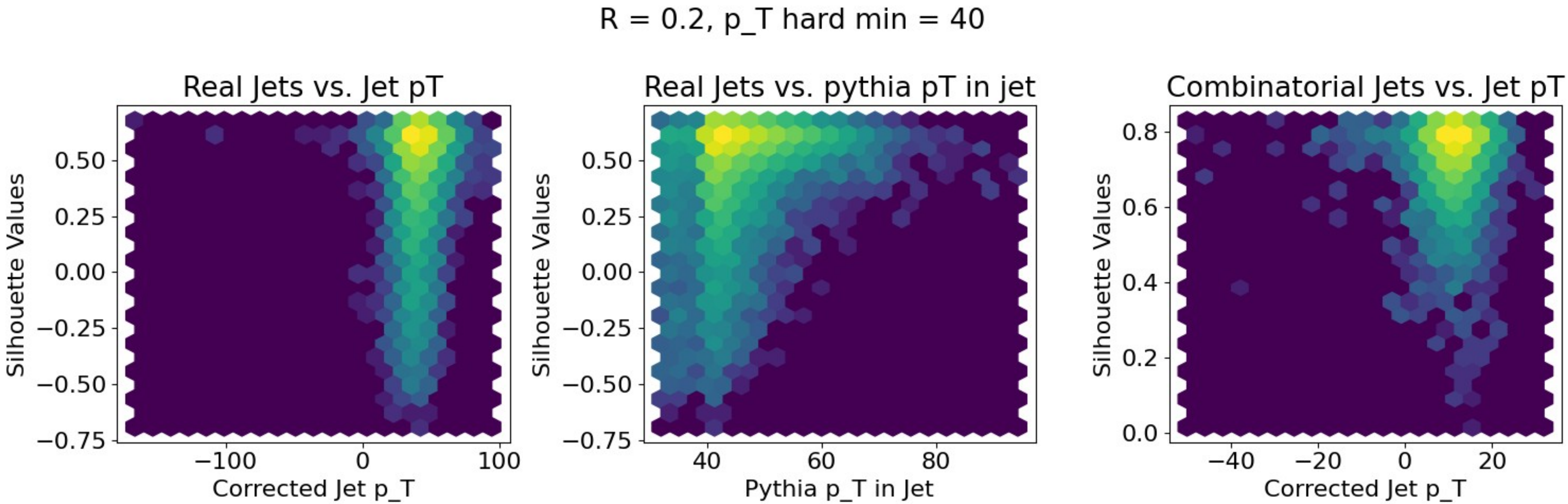
[Example from Wikipedia](#)



Silhouette scores from three types of animals rendered by Orange data mining suite.



Silhouette values

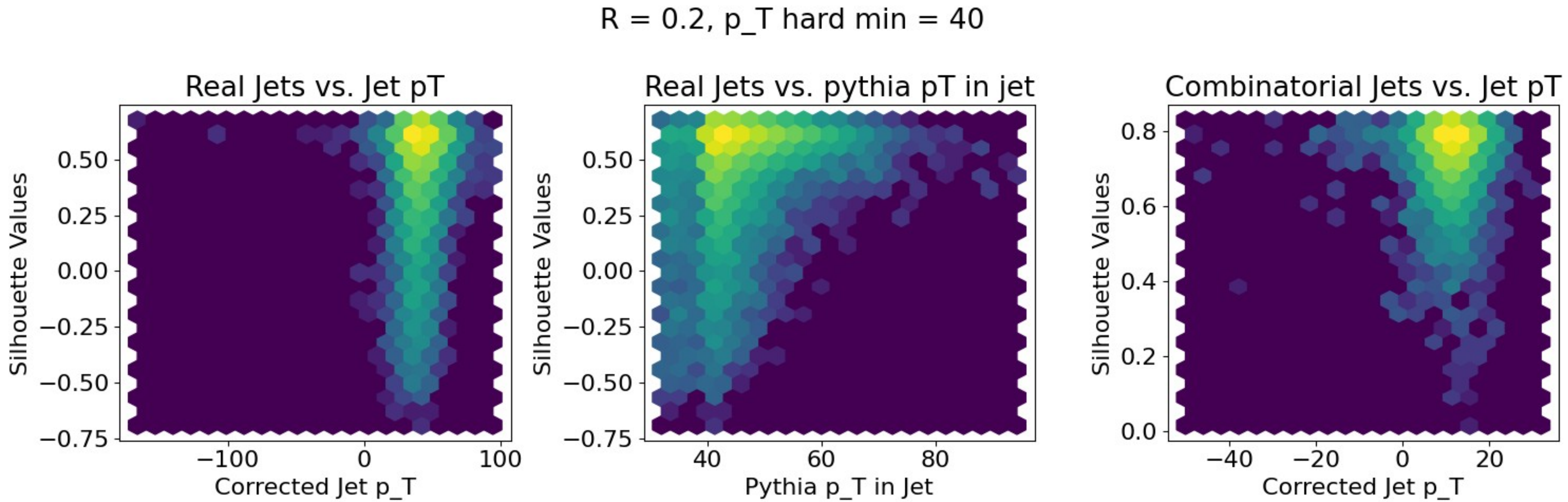


s<0: look more like background

Real jets look more real if PYTHIA p_T is higher

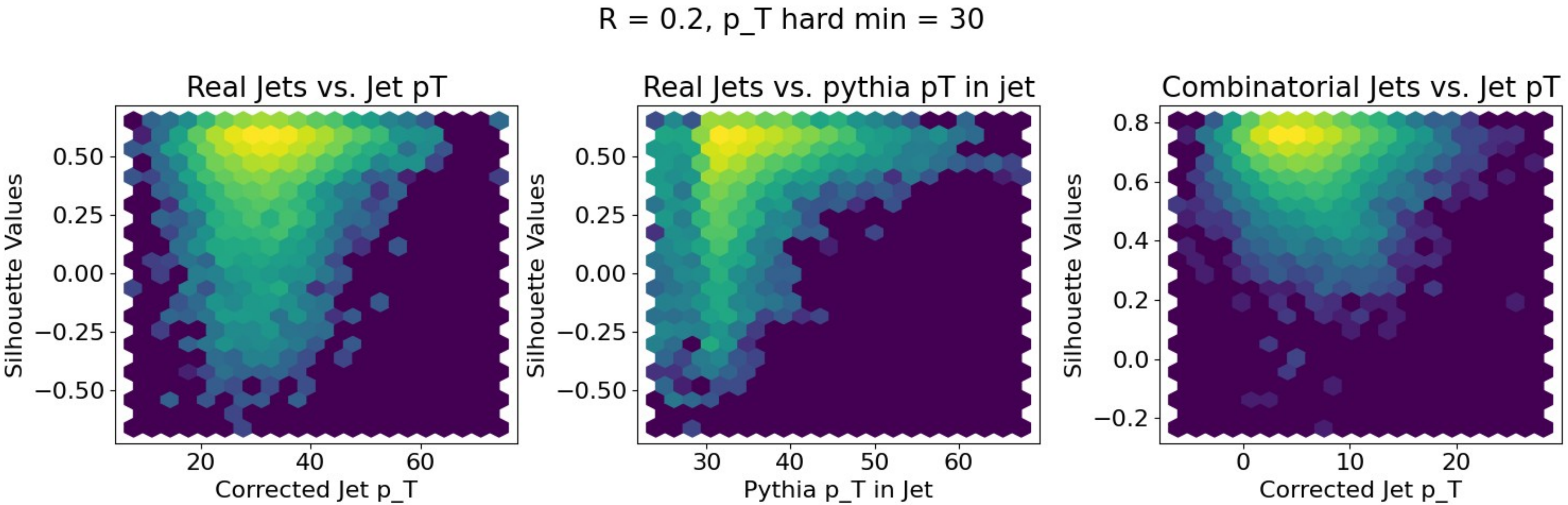
s~0: look similar to signal

Silhouette values – decreasing p_T



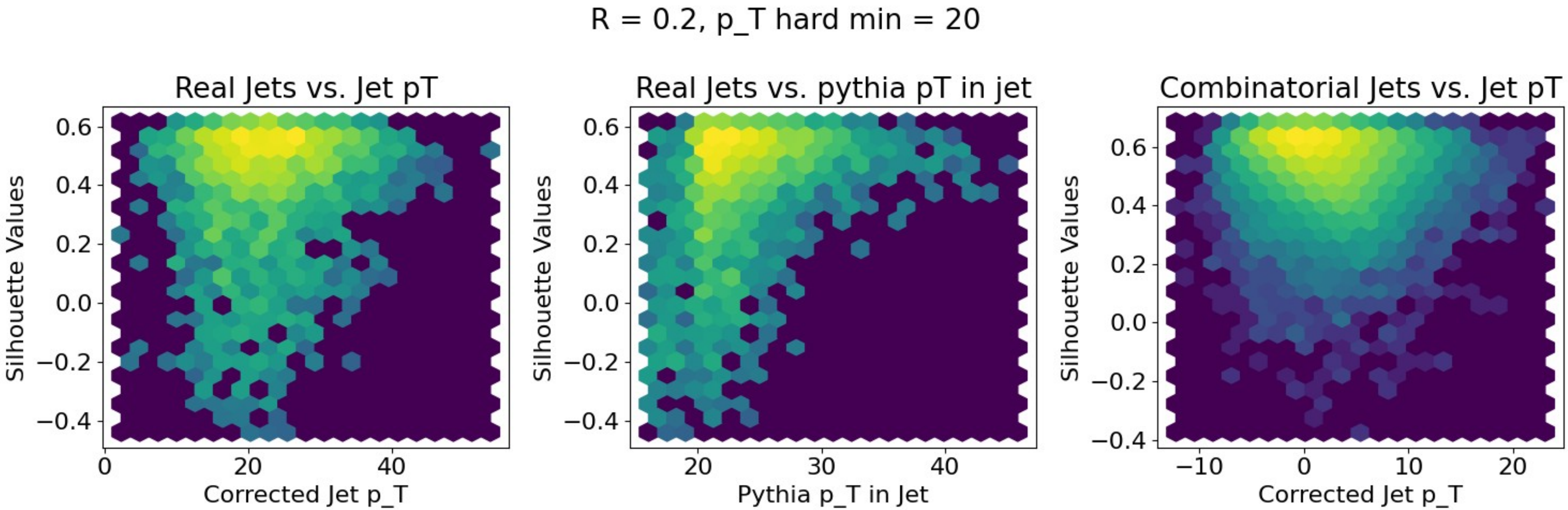
Log z scale

Silhouette values – decreasing p_T



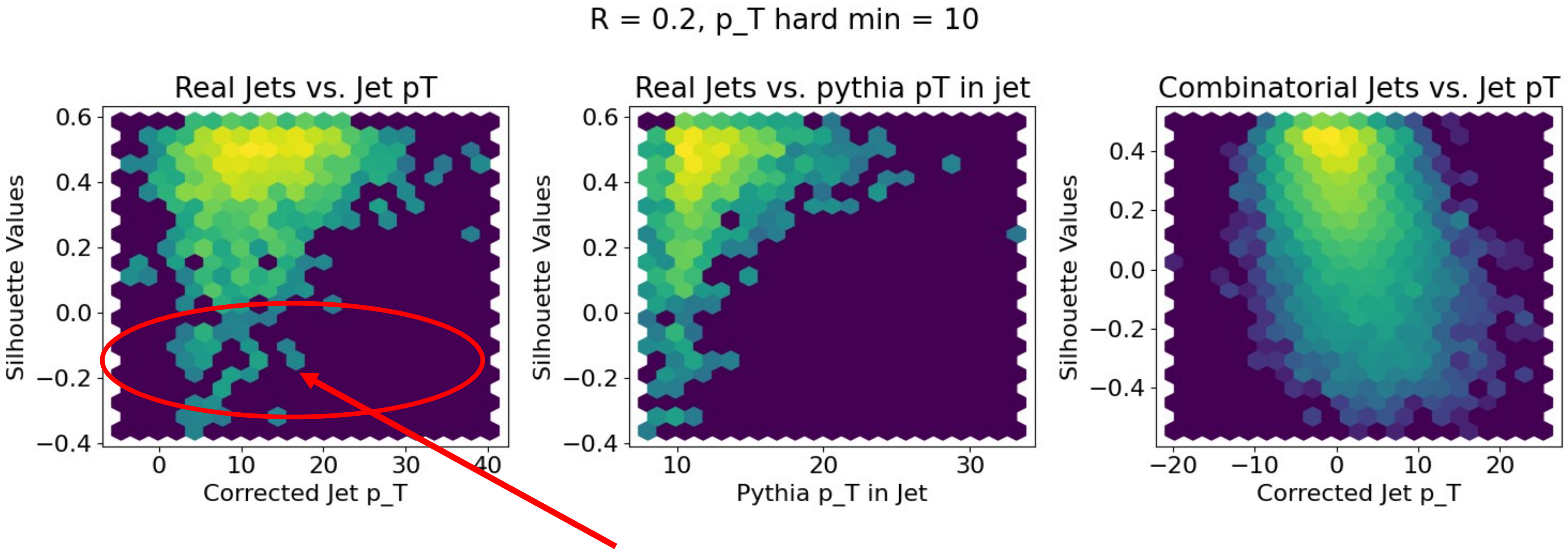
Log z scale

Silhouette values – decreasing p_T



Log z scale

Silhouette values – decreasing p_T



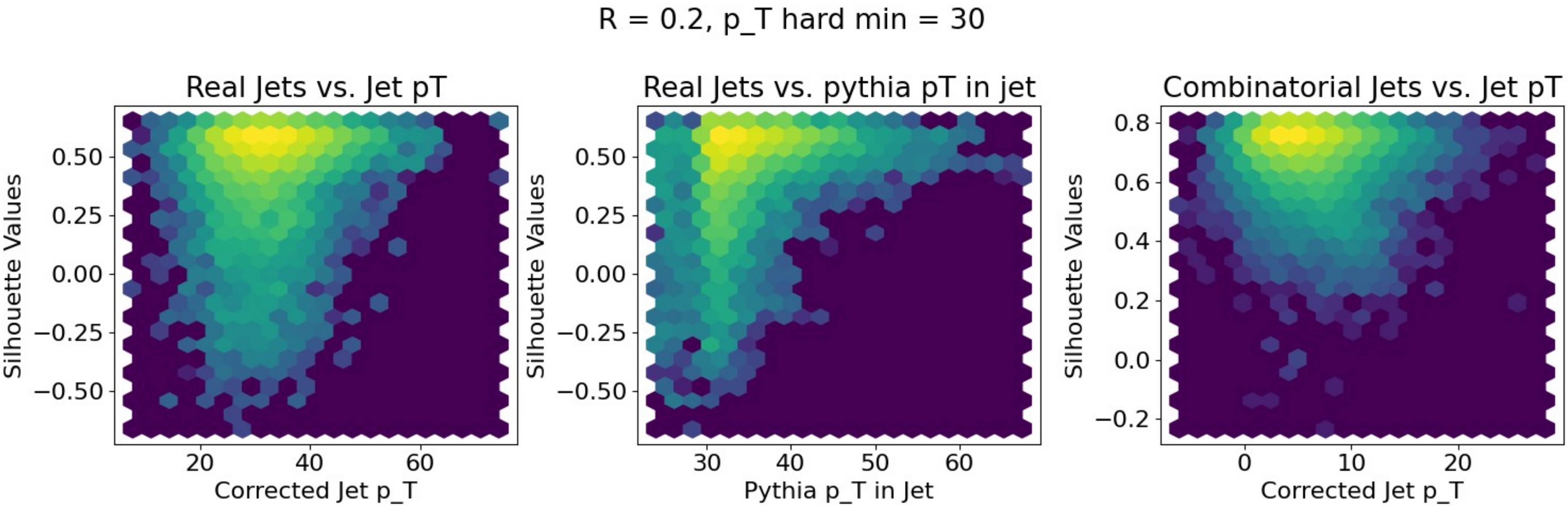
Real jets look more like combinatorial jets

These aren't random jets!

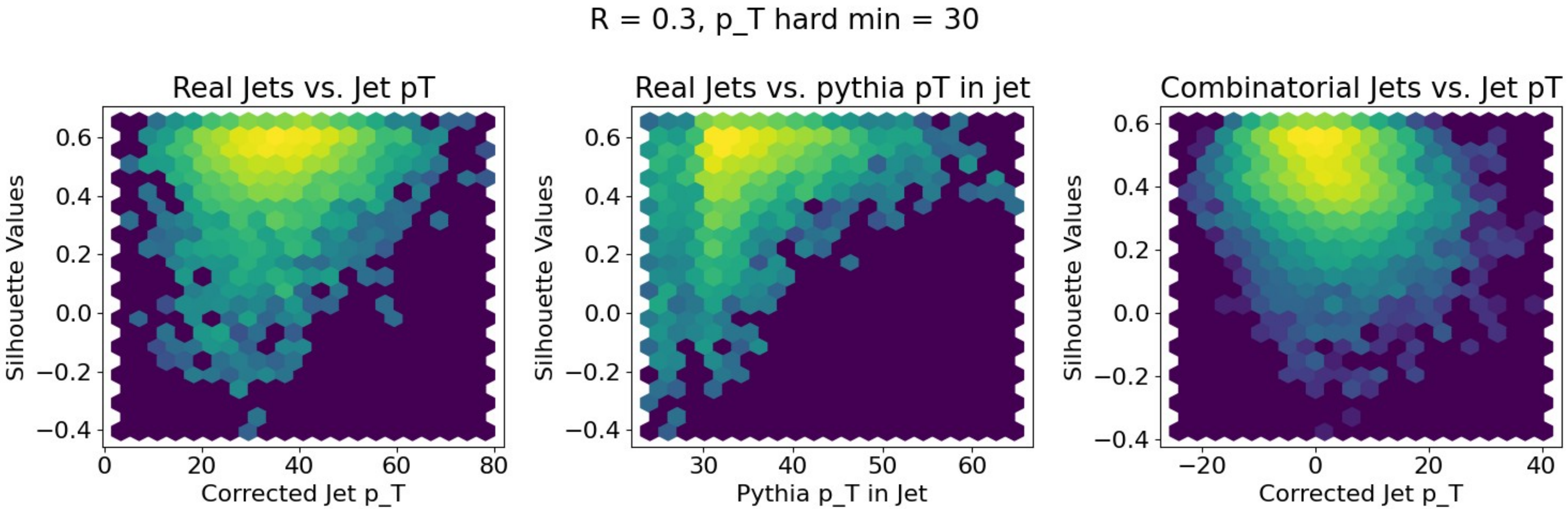
Combinatorial jets look more like real jets

Log z scale

Silhouette values – increasing R

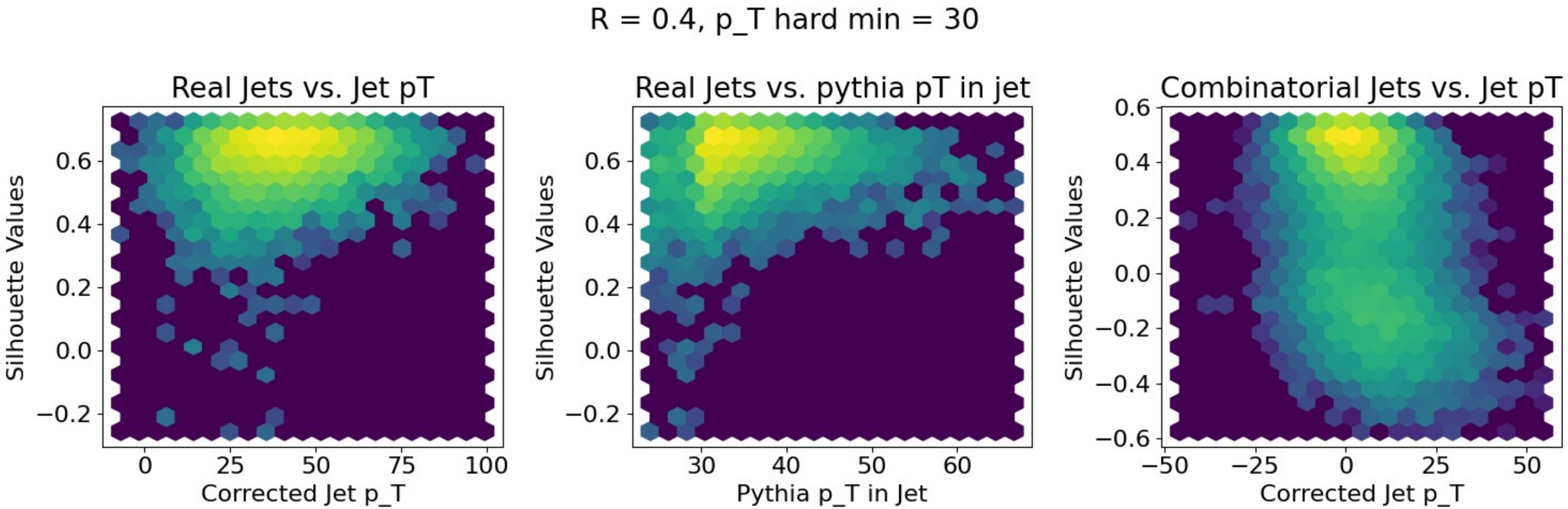


Silhouette values – increasing R



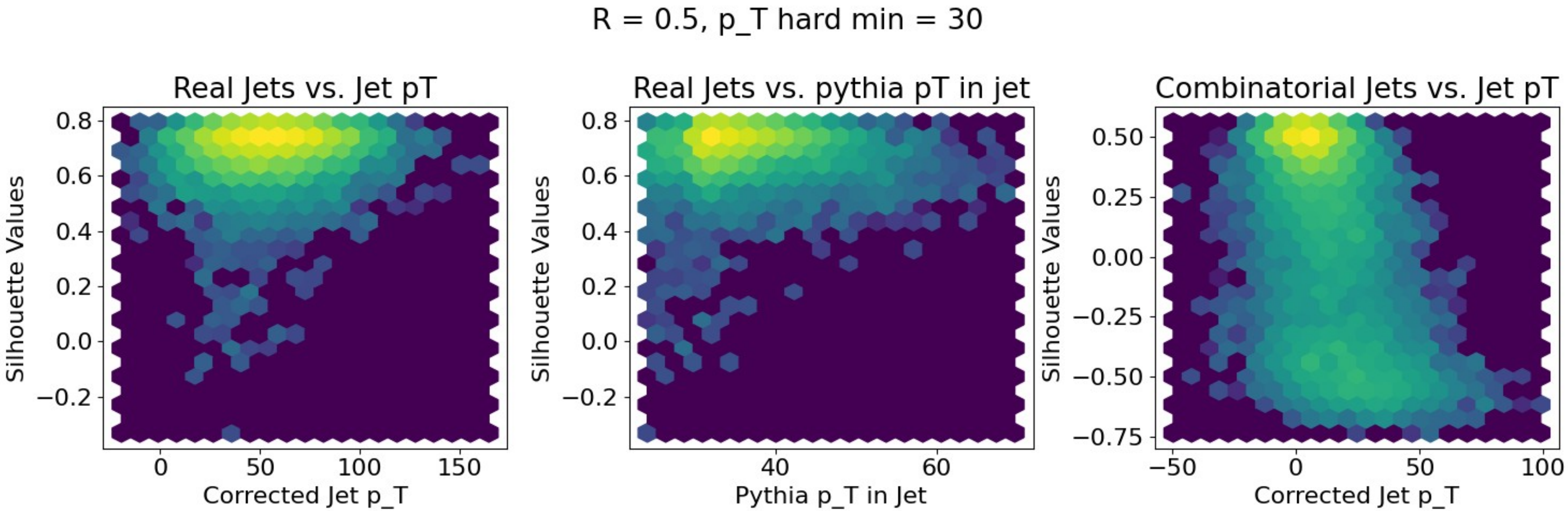
Log z scale

Silhouette values – increasing R



Log z scale

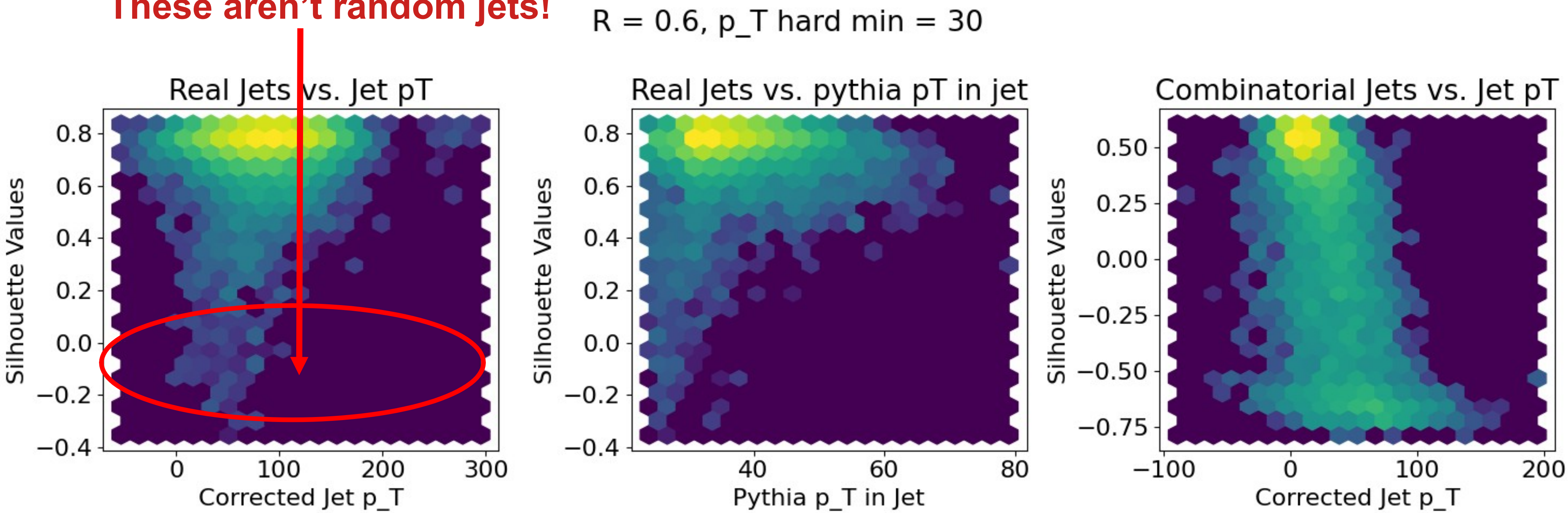
Silhouette values – increasing R



Log z scale

Silhouette values – increasing R

These aren't random jets!



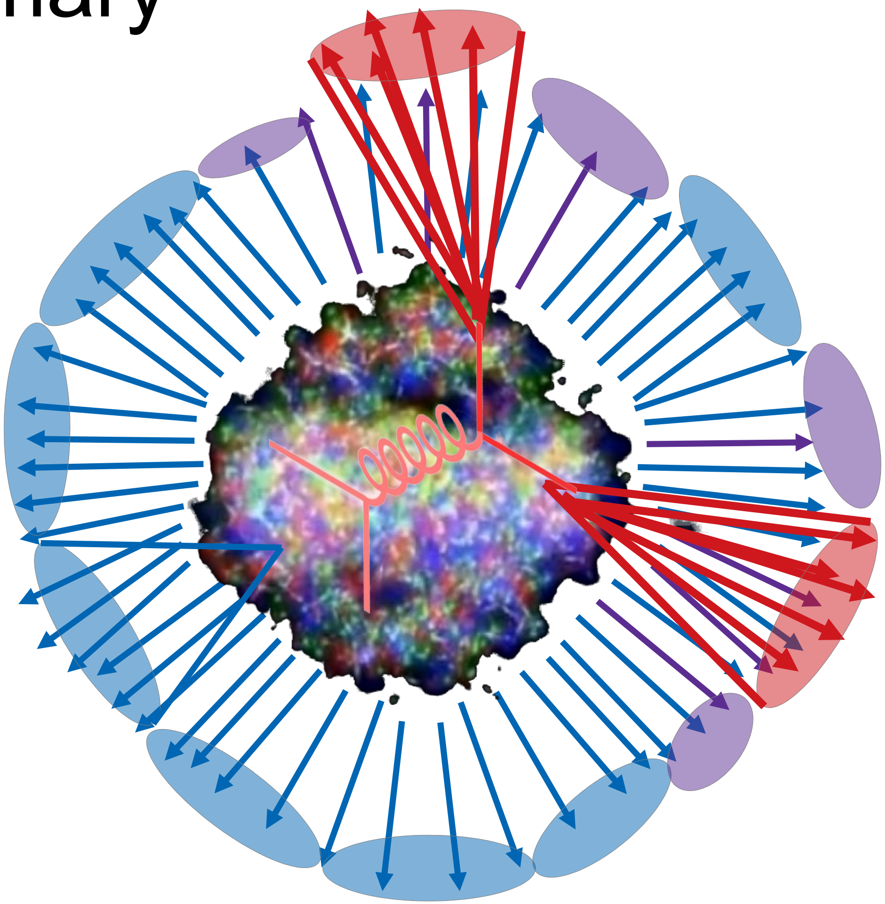
Real jets look more like real jets

Tail in distribution of real jets gets smaller

Combinatorial jets look more like real jets

Mini-summary

- “Signal” and “background” have different properties, but...
- Always overlap somewhat
- Any procedure to remove “background” will also cut signal

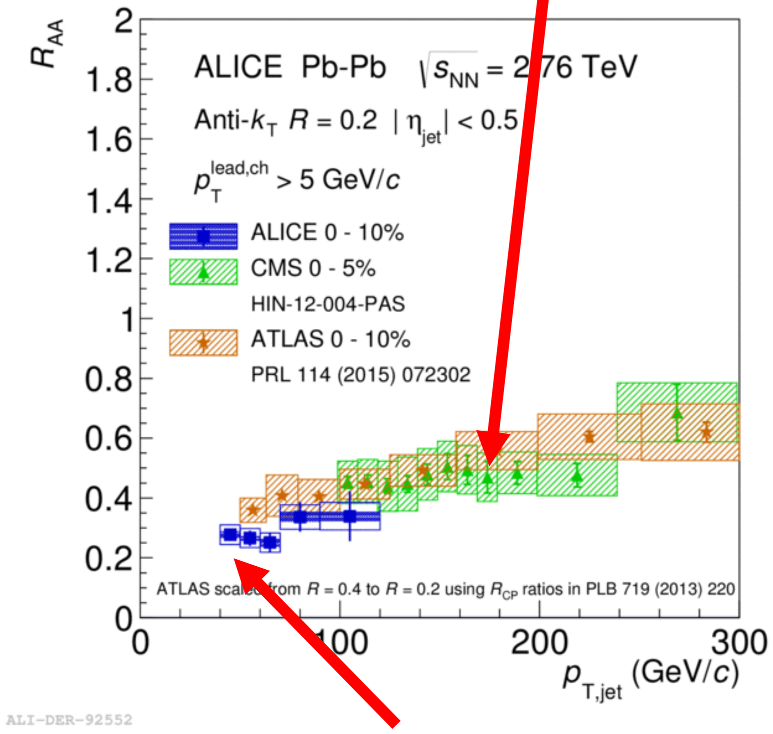


How to compare to models

Iterative procedure

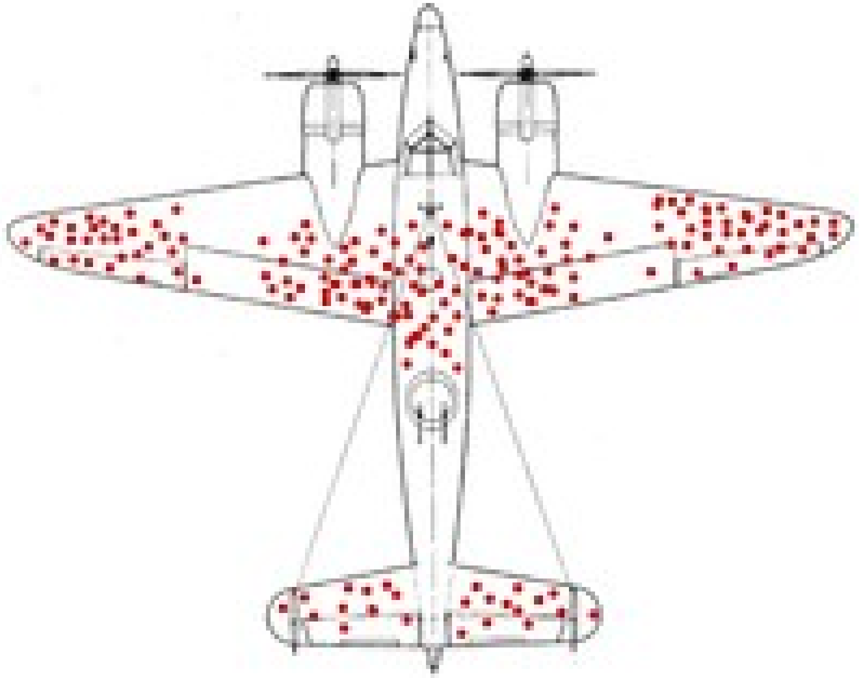
- Used by ATLAS & CMS
- ATLAS
  - **Calorimeter jets:** Reconstruct jets with R=0.2. v_2 modulated <Bkgd> estimated by energy in calorimeters excluding jets with at least one tower with
    $E_{tower} > \langle E_{tower} \rangle$
    Track jets: Use tracks with p_T>4 GeV/c
  - Calorimeter jets from above with E>25 GeV and track jets with p_T>10 GeV/c used to estimate background again.
  - Calorimeter tracks matching one track with p_T>7 GeV/c or containing a high energy cluster E >7 GeV are used for analysis down to E_{jet} = 20 GeV

Constituent biases don't matter that much up here



But they do matter down here!

Survivor bias



- WWII Example: holes planes returning indicate where it's *safer* to get hit
- We're looking at the jets which *remain*

Bias

- **Experimental background subtraction methods: complex, make assumptions, apply biases**
- **Survivor bias: Modified jets probably look more like the medium**
- **Quark/Gluon bias:**
  - **Quark jets are narrower, have fewer tracks, fragment harder [Z Phys C 68, 179-201 (1995), Z Phys C 70, 179-196 (1996),]**
  - **Gluon jets reconstructed with k_T algorithm have more particles than jets reconstructed with anti-k_T algorithm [Phys. Rev. D 45, 1448 (1992)]**
  - **Gluon jets fragment into more baryons [EPJC 8, 241-254, 1998]**
- **Fragmentation bias: Experimental measurements explicitly select jets with hard fragments**

Snowmass Accord: Apply the same algorithm to data and your model. Then the measurement and the calculation are the same.

Rivet: Apply the same algorithm to data and your model. Then the measurement and the calculation are the same.

What is Rivet?

AMPT ANGANTYR HIJING PYTHIA JETSCAPE

Monte Carlo Model

PHSD JEWEL EPOS smash

↓

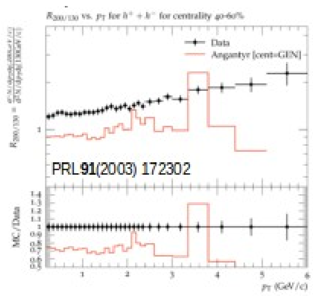
HepMC

↓

HEPData → Rivet

↓

Comparison to data



Why use Rivet?

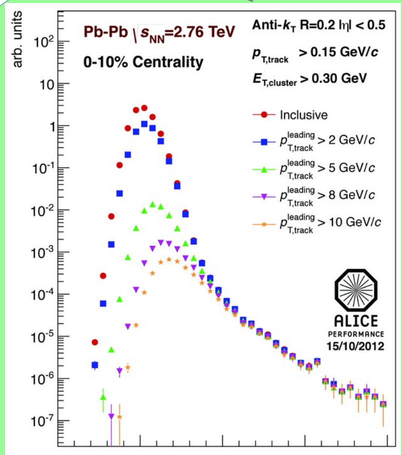
- Facilitates comparisons between Monte Carlos and data
- It’s not that hard
- It preserves analysis details

Analysis steps

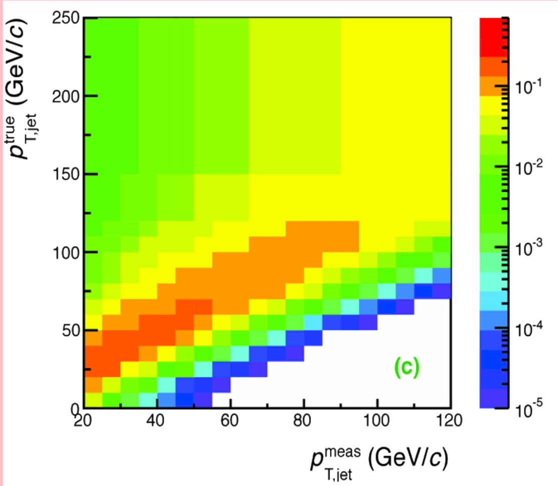
Tracks → Jet finding algorithm

Clusters → Jet finding algorithm

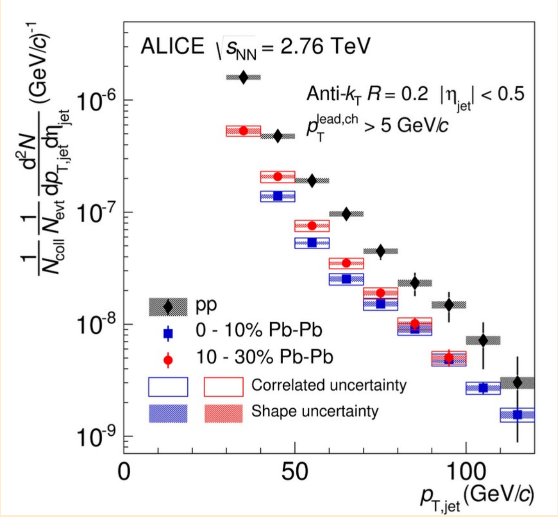
Jet finding algorithm → Jet candidates → Background subtraction



Jet spectrum smeared by energy resolution, background fluctuations



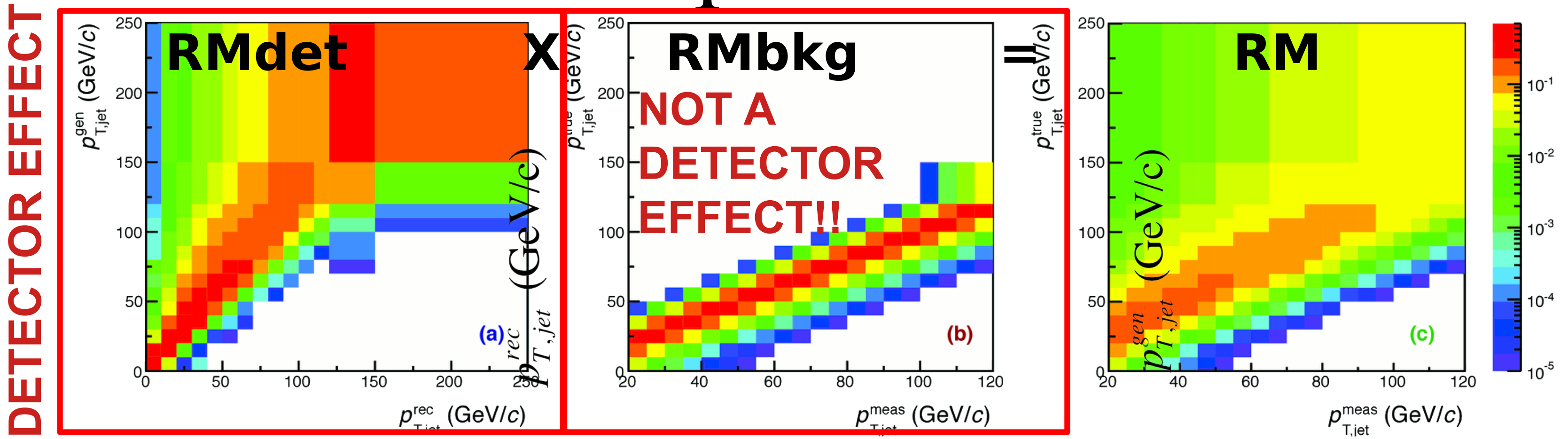
Unfolding – corrects for single track reco ε, E resolution, background fluctuations

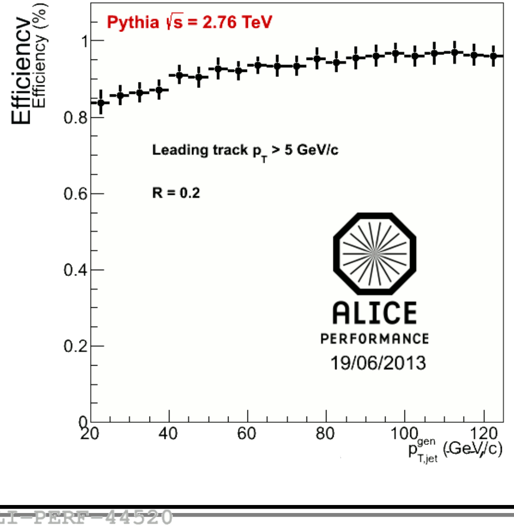


Corrected spectra

<footer>

Jets in ALICE: Response Matrix Construction





Anti-k_T R=0.2

$p_{T,track}$ > 0.15 GeV/c

$E_{T,cluster}$ > 0.30 GeV

$p^{leading}_{T,track}$ > 5 GeV/c

(a) RM_{det} Detector response matrix

(b) RM_{bkg} Background fluctuation matrix

(c) $RM_{tot} = RM_{bkg} \times RM_{det}$

Pb-Pb $\sqrt{s_{NN}}$=2.76 TeV

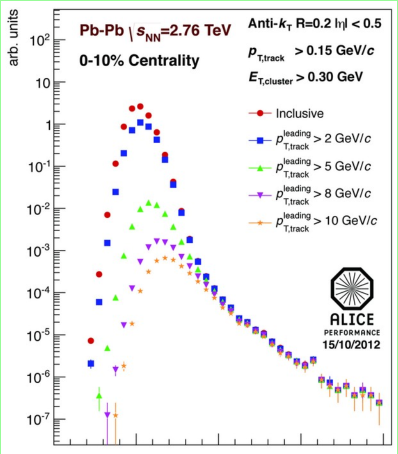
0-10% Centrality

ALICE
PERFORMANCE
15/10/2012

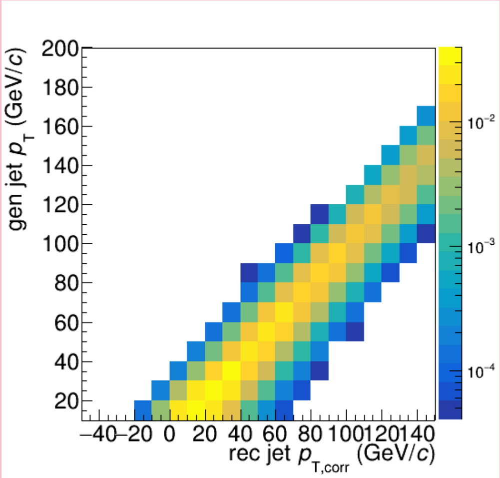
RM_{bkg} and RM_{det} are approximately factorizable

Analysis steps: Full Monte Carlo

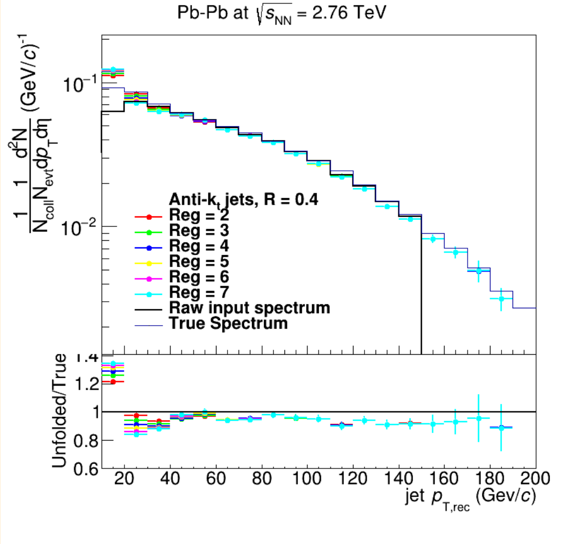
Particles → Jet finding algorithm → Jet candidates → Background subtraction



Jet spectrum smeared by energy resolution

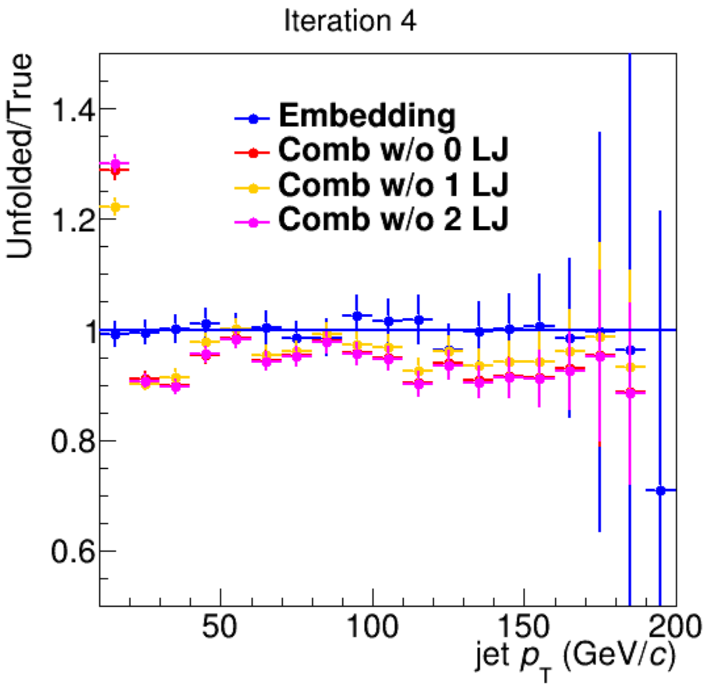


Unfolding – corrects for background fluctuations and other effects

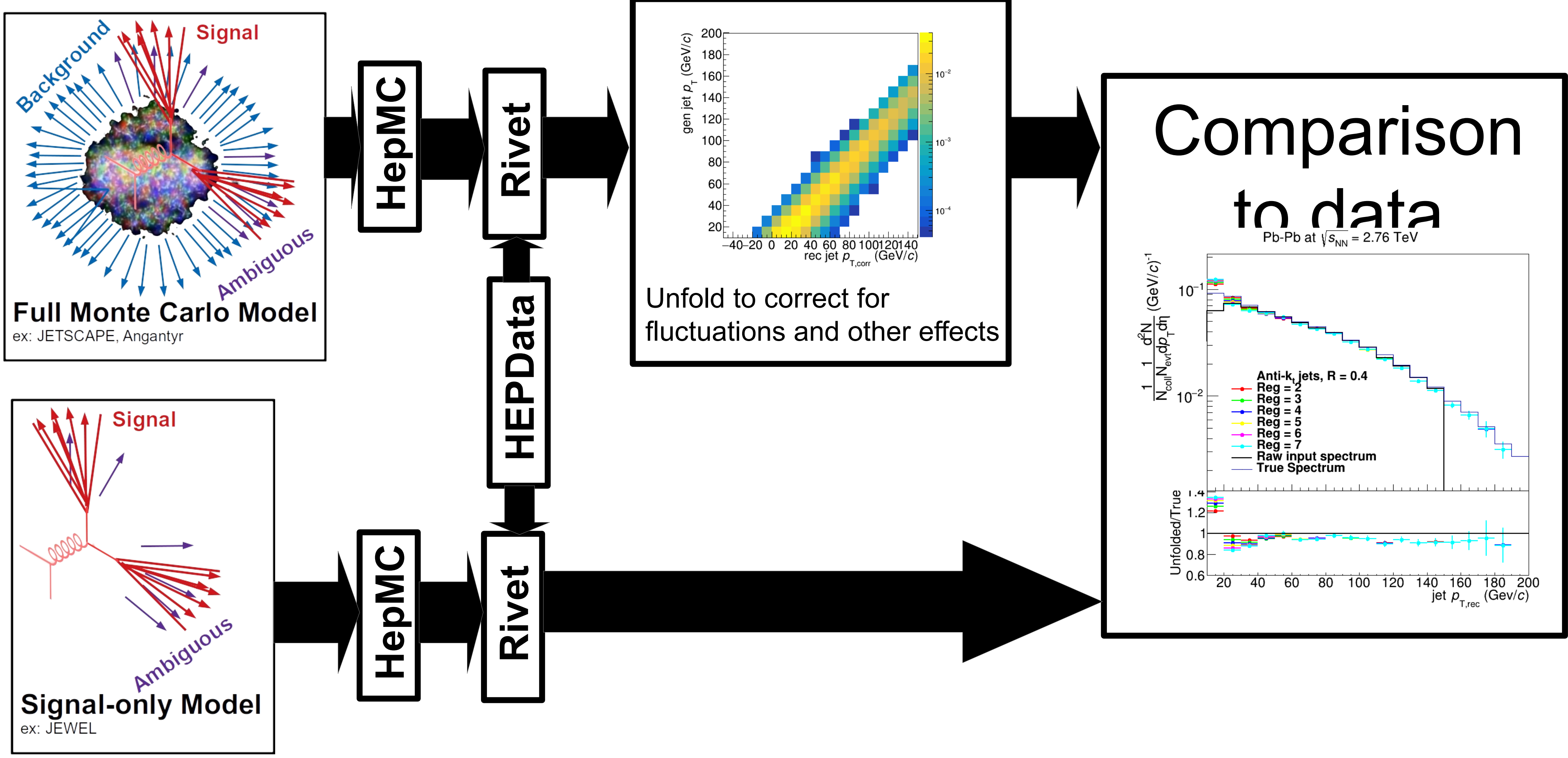


Corrected spectra

Closure



- Methods
  - Use δp_T method to measure width of fluctuations with varying numbers of leading jets (LJ) discarded
  - Embed PYTHIA pp event into PYTHIA heavy ion event
  - The PYTHIA pp event is “true”
- Only embedding leads to full closure


Background
Signal
Ambiguous
Full Monte Carlo Model
ex: JETSCAPE, Angantyr
Signal
Ambiguous
Signal-only Model
ex: JEWEL
HepMC
Rivet
HEPData
HepMC
Rivet
gen jet p_T (GeV/c)
rec jet $p_{T,corr}$ (GeV/c)
Unfold to correct for
fluctuations and other effects
Comparison to data
Pb-Pb at $\sqrt{s_{NN}}$ = 2.76 TeV
Anti-k_t jets, R = 0.4
Reg = 2
Reg = 3
Reg = 4
Reg = 5
Reg = 6
Reg = 7
Raw input spectrum
True Spectrum
Unfolded/True
jet $p_{T,rec}$ (Gev/c)

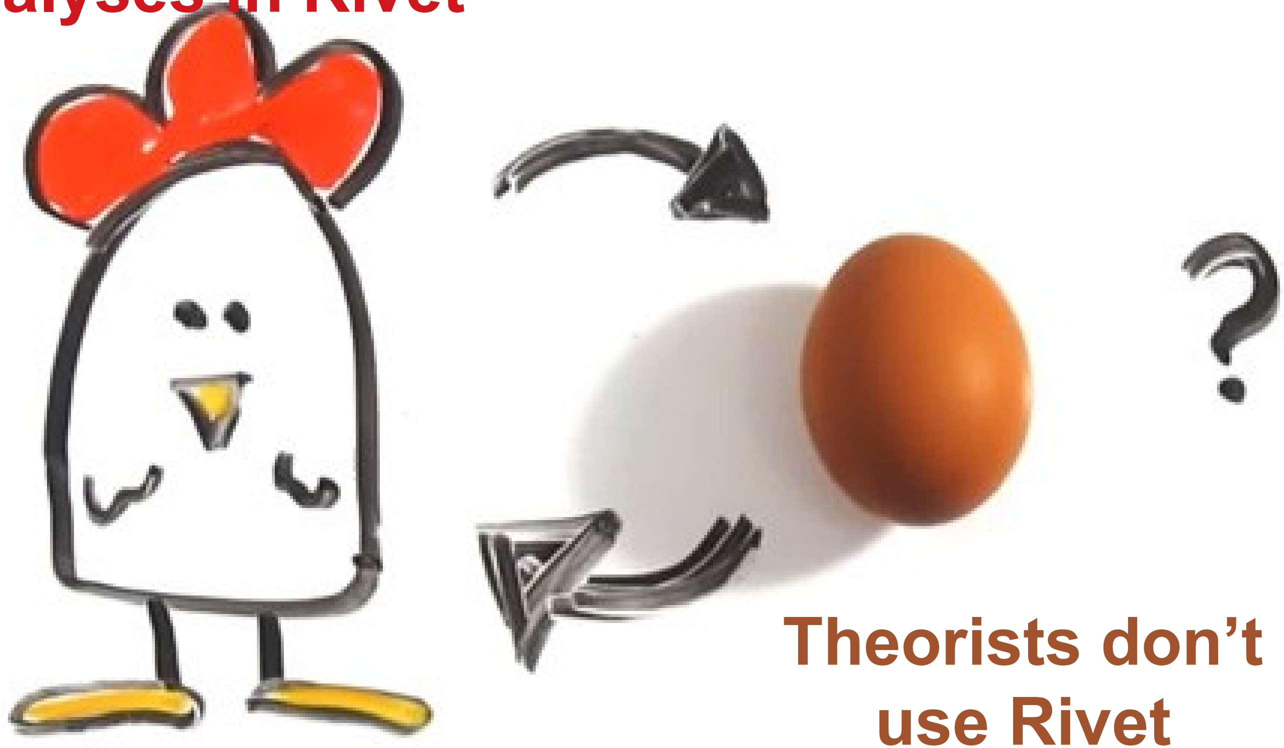
Conclusions

- "Background" is not just an experimental problem!
- "Signal" and "background" jets overlap → impossible to suppress background without biasing jets
- Gold standard is to use Rivet
  - ˜ But it requires treating the model *exactly* like data
  - ˜ A number of issues specific to jets need to be discussed in the field

    Recorded tutorials from Rivetizing Heavy Ion Collisions at RHIC

Backup: undergraduates

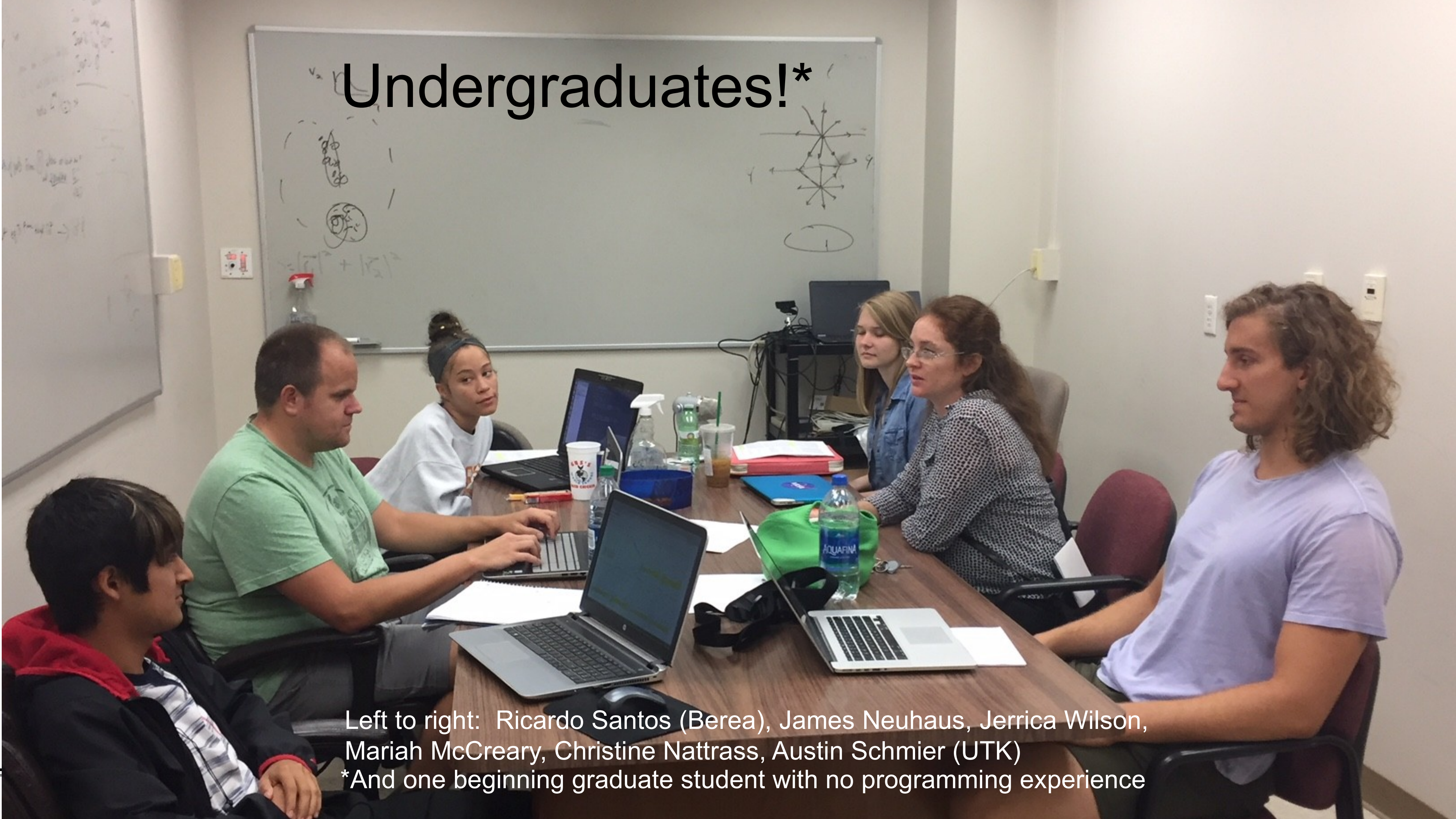
Few heavy ion analyses in Rivet



Theorists don't use Rivet

http://iterated-reality.com/en/2015/03/17/the-chicken-or-the-egg-causality-dilemma-solved-by-unity-consciousness/

Undergraduates!*



Left to right: Ricardo Santos (Berea), James Neuhaus, Jerrica Wilson, Mariah McCreary, Christine Nattrass, Austin Schmier (UTK)

*And one beginning graduate student with no programming experience

Course-based undergraduate research experience

Ask me if you want more info!

CBE—Life Sciences Education, Vol. 15, No. 2 | Articles   Free Ac

Early Engagement in Course-Based Research Increases Graduation Rates and Completion of Science, Engineering, and Mathematics Degrees

Stacia E. Rodenbusch, Paul R. Hernandez, Sarah L. Simmons, and Erin L. Dolan

Jennifer Knight, Monitoring Editor:

Published Online: 13 Oct 2017 | https://doi.org/10.1187/cbe.16-03-0117

Sections View Article   Tools Shai

Abstract

National efforts to transform undergraduate biology education call for research experiences to be an integral component learning for all students. Course-based undergraduate research experiences, or CUREs, have been championed for engagi students in research at a scale that is not possible through apprenticeships in faculty research laboratories. Yet there are f if any studies that examine the long-term effects of participating in CUREs on desired student outcomes, such as graduatir from college and completing a science, technology, engineering, and mathematics (STEM) major. One CURE program, the Freshman Research Initiative (FRI), has engaged thousands of first-year undergraduates over the past decade. Using propensity score–matching to control for student-level differences, we tested the effect of participating in FRI on students' probability of graduating with a STEM degree, probability of graduating within 6 yr, and grade point average (GPA) at graduation. Students who completed all three semesters of FRI were significantly more likely than their non-FRI peers to earn a STEM degree and graduate within 6 yr. FRI had no significant effect on students' GPAs at graduation. The effects were similar for diverse students. These results provide the most robust and best-controlled evidence to date to support calls for early involvement of undergraduates in research.

Phys 494 – Course-based Undergraduate Research Experience in Relativistic Heavy Ion Physics

Instructor:
Dr. Christine Nattrass
Office: SERF 609
Phone: 974-6211
Email: christine.nattrass@utk.edu
Office hours: TBA

Teaching assistant: N/A

Class time & Location: TR 12:40-1:55 SERF 210

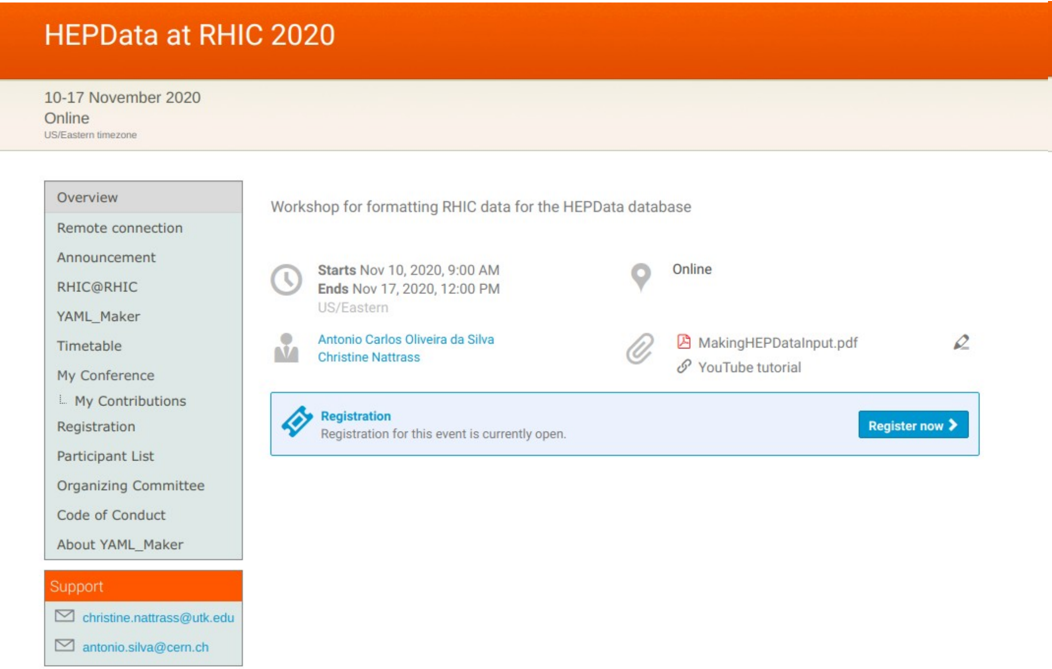
Course Description:
This course will incorporate undergraduates into a research project in high energy nuclear physics in a course setting. Each student will be responsible for implementing a heavy ion analysis in the program RIVET so that it can be used by the JETSCAPE collaboration to make comparisons between Monte Carlo models and data. Each student's project will be incorporated into a public software repository so that it is available to the field and, if possible, it will be validated by the relevant experiment and incorporated into the official RIVET software.

3 semesters	All Rivet students
15 students	22 students
8 women	11 women
3 minorities	7 minorities
3 non-traditional	4 non-traditional

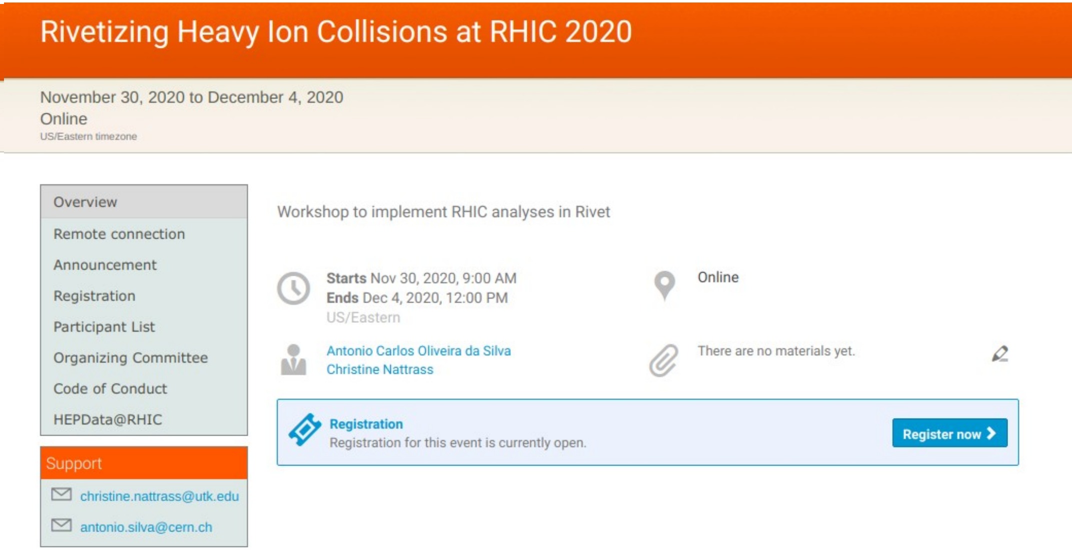
Learn Rivet yourself!

Or send your students & postdocs!

https://indico.bnl.gov/event/8843/



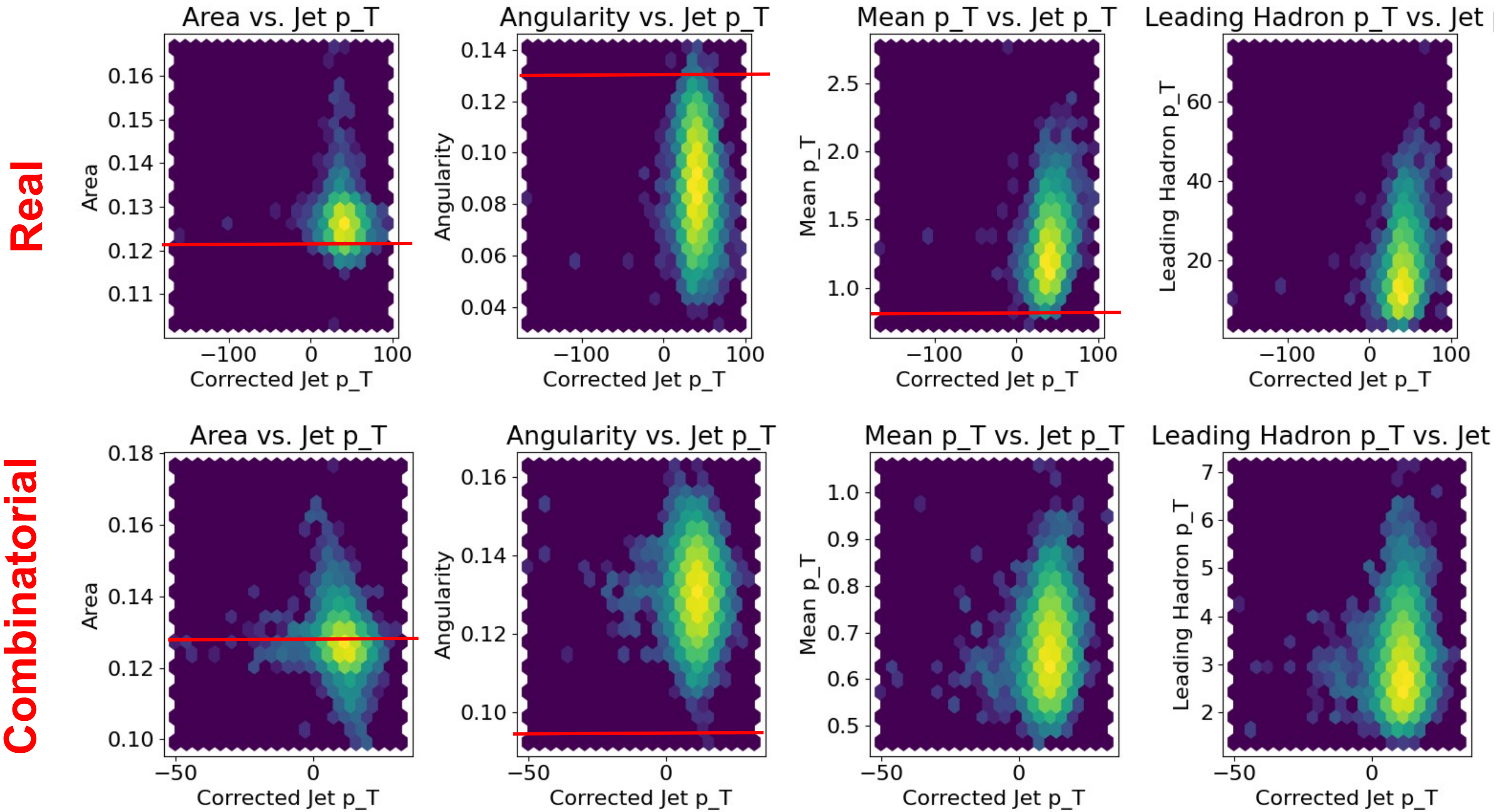
https://indico.bnl.gov/event/8840



Backup: jet properties

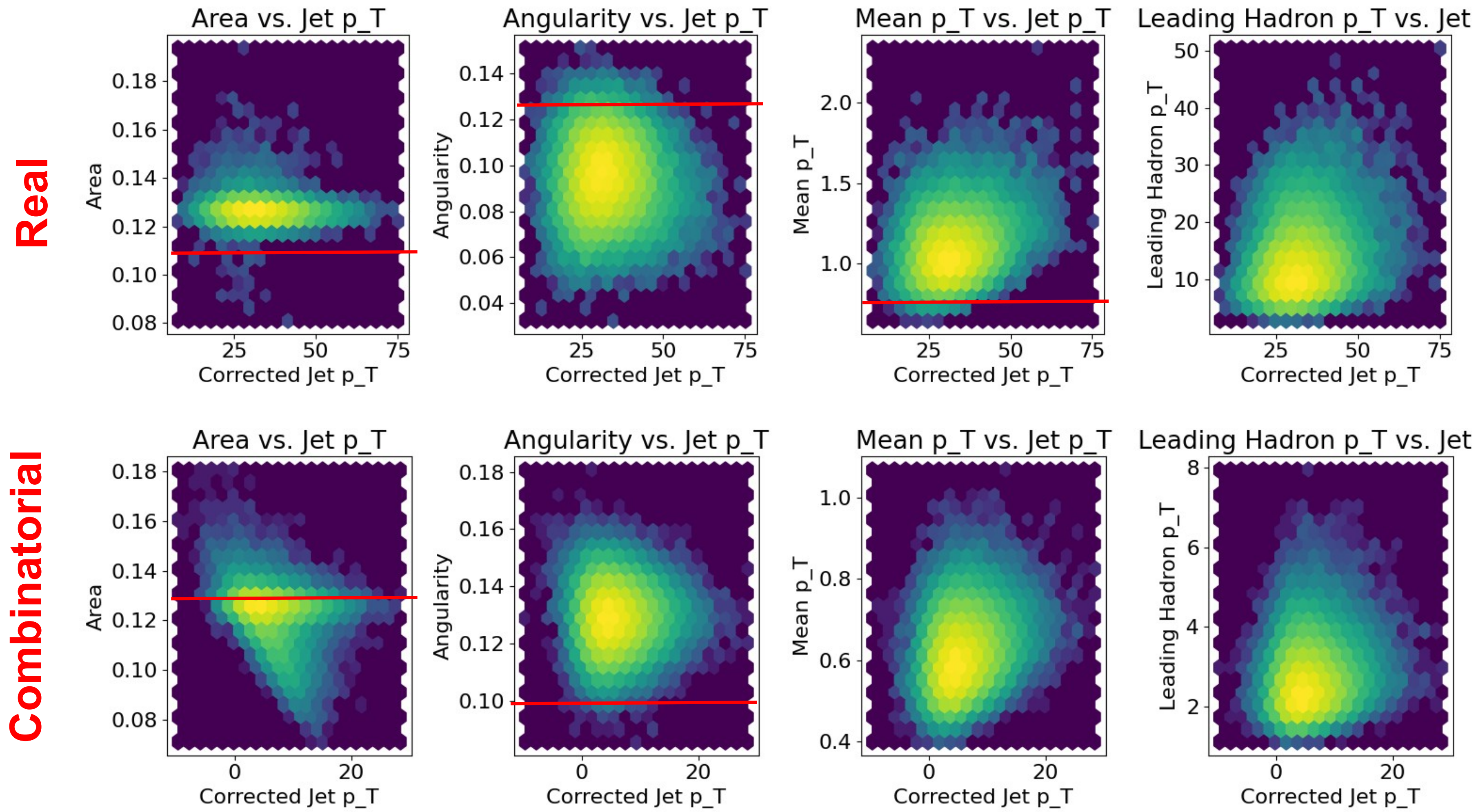
$p_{T hard}$>40 GeV/*c*, R=0.2

Log z scale



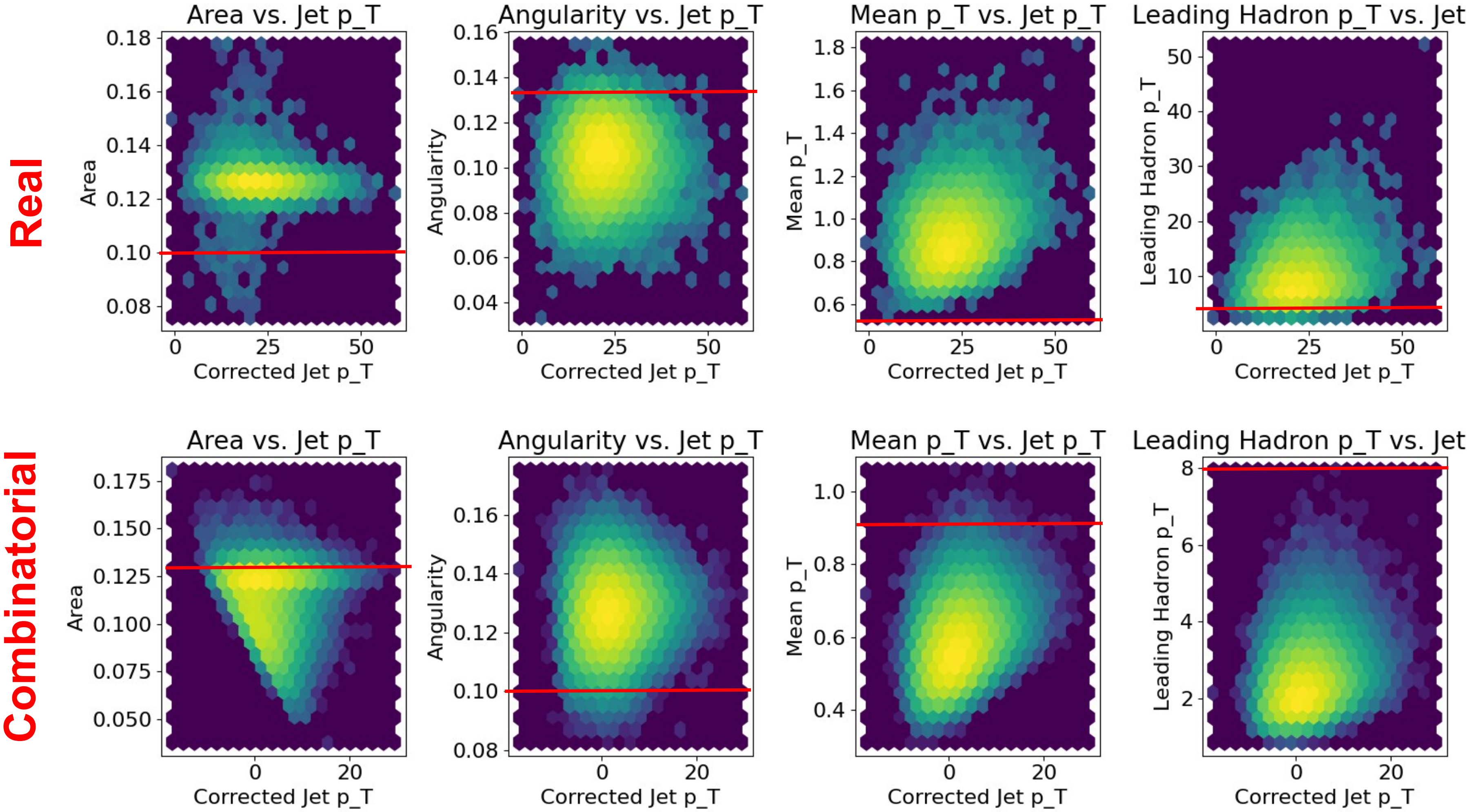
$p_{T hard}$>30 GeV/*c*, R=0.2

Log z scale



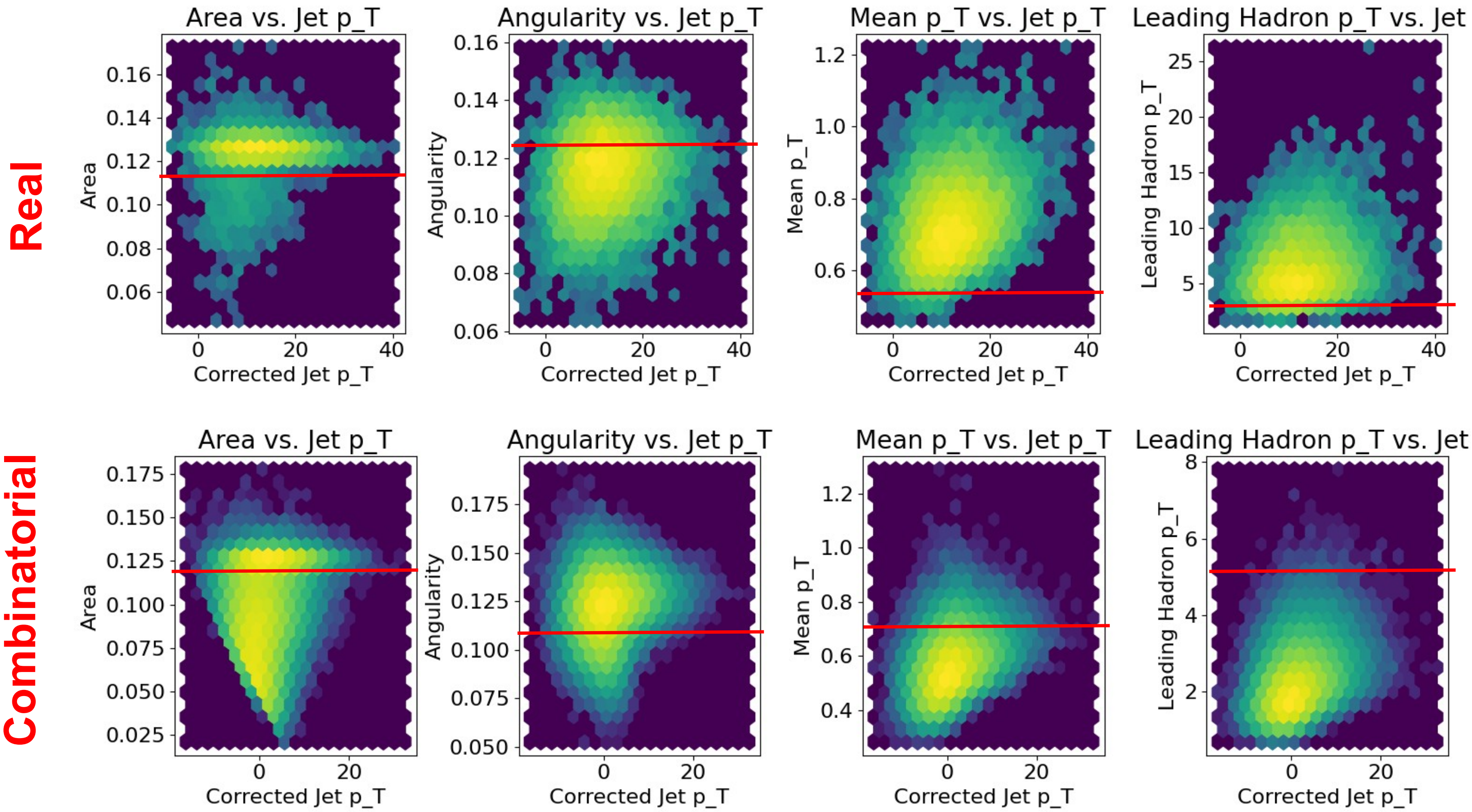
$p_{T,hard}$>20 GeV/c, R=0.2

Log z scale



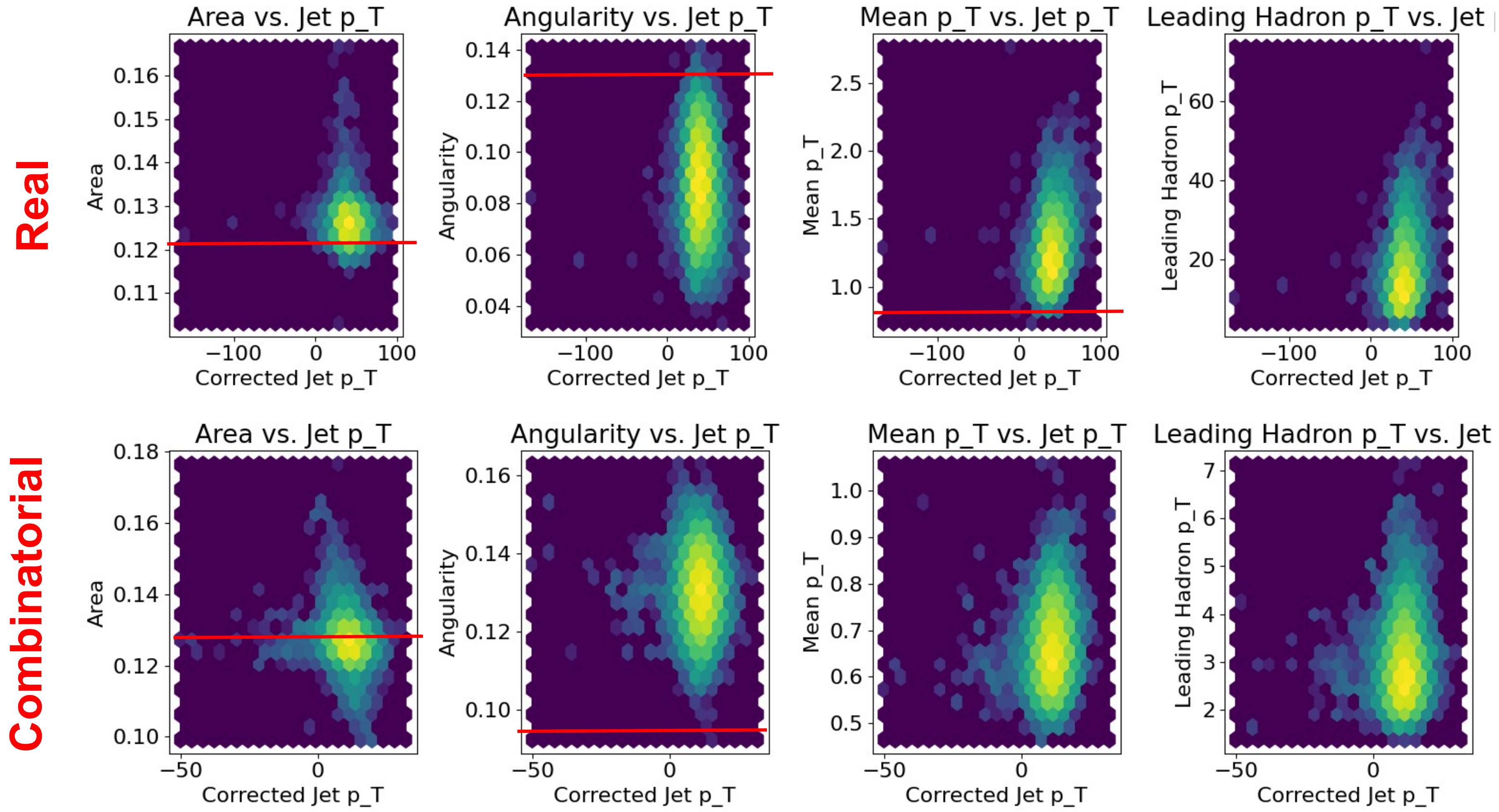
$p_{T\,hard}$>10 GeV/c, R=0.2

Log z scale



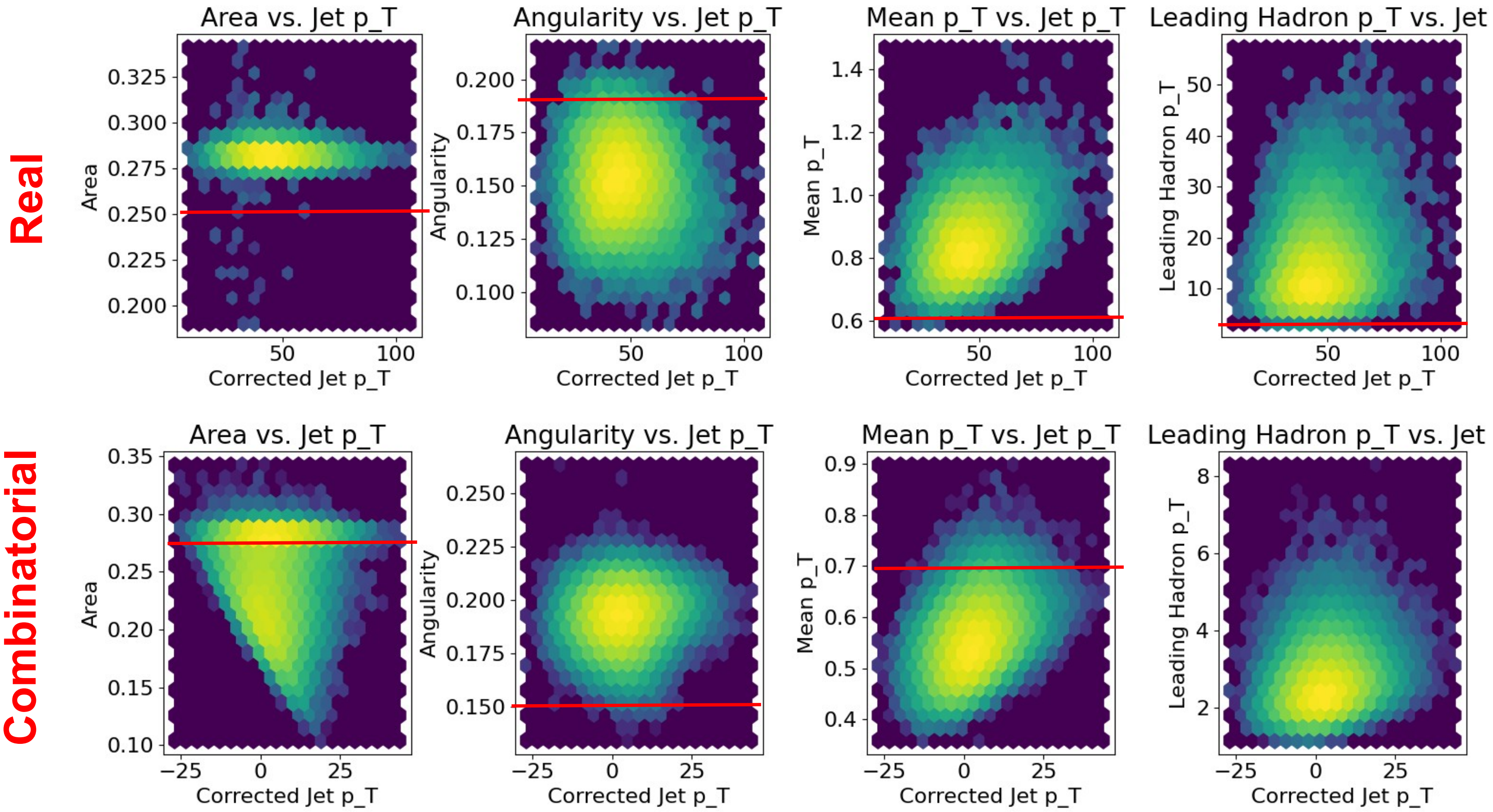
$p_{T hard}$>40 GeV/*c*, R=0.2

Log z scale



p_{Thard}>40 GeV/*c*, R=0.3

Log z scale

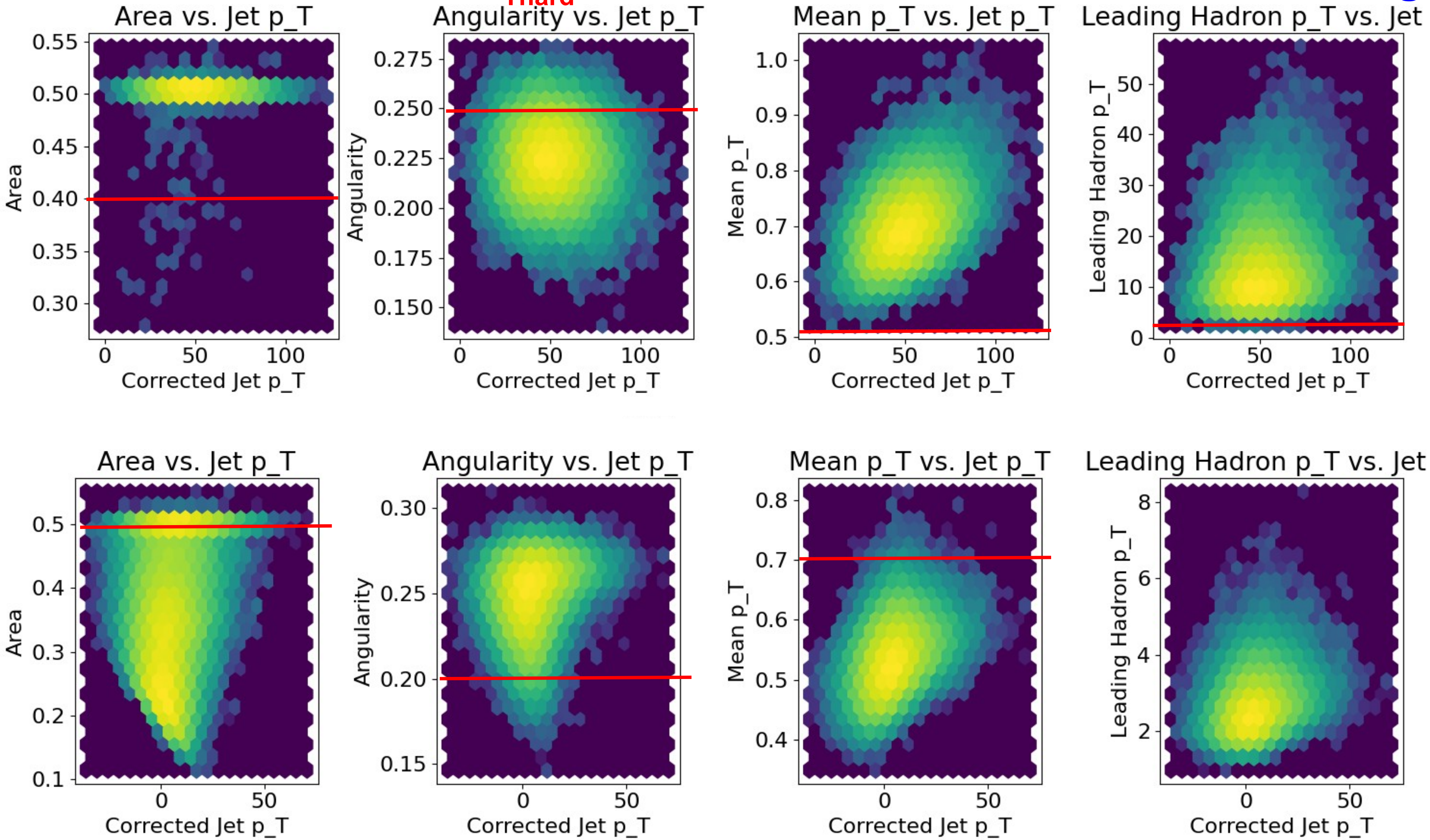


p_{Thard}>40 GeV/*c*, R=0.4

Log z scale

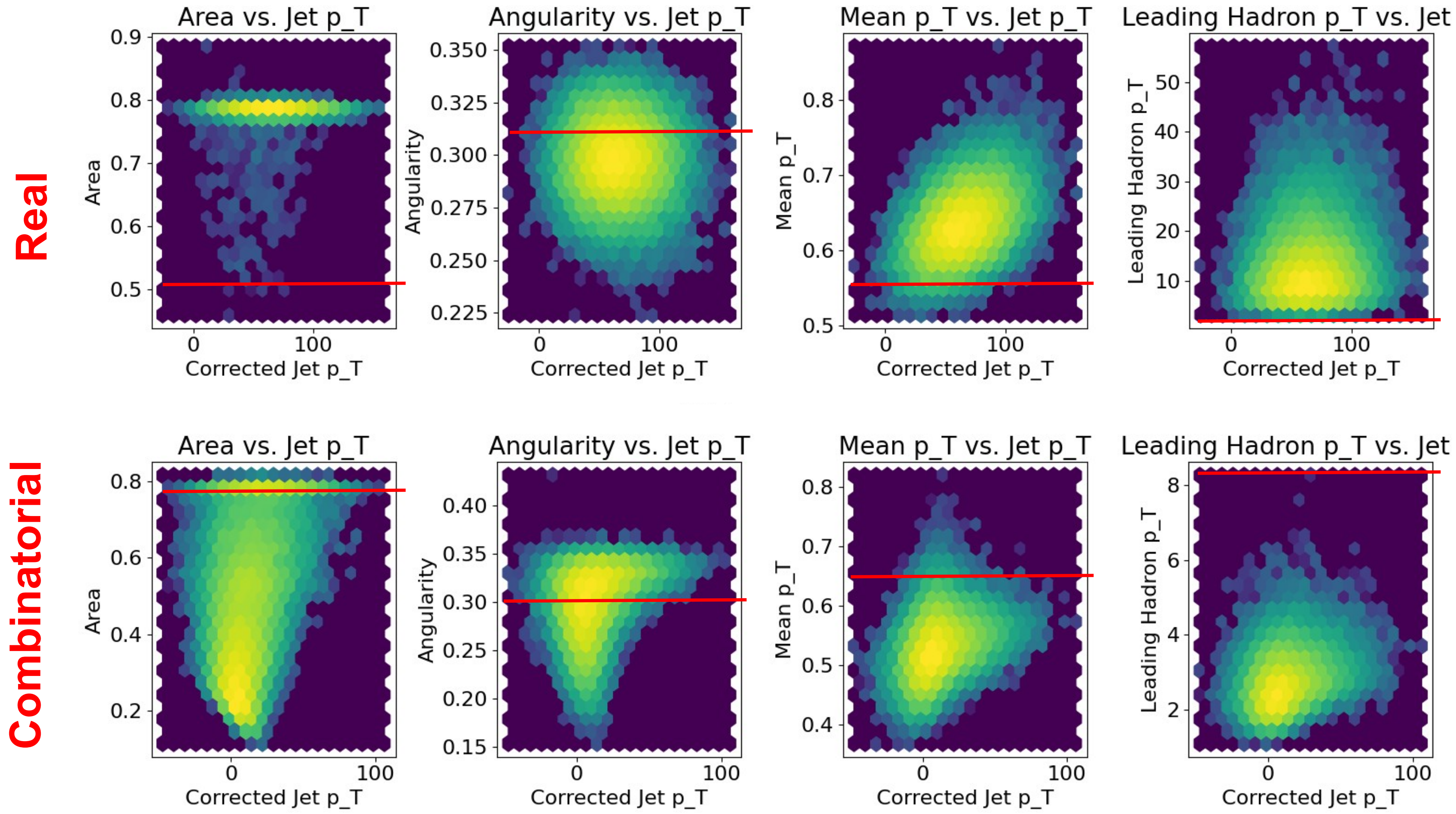
Real

Combinatorial



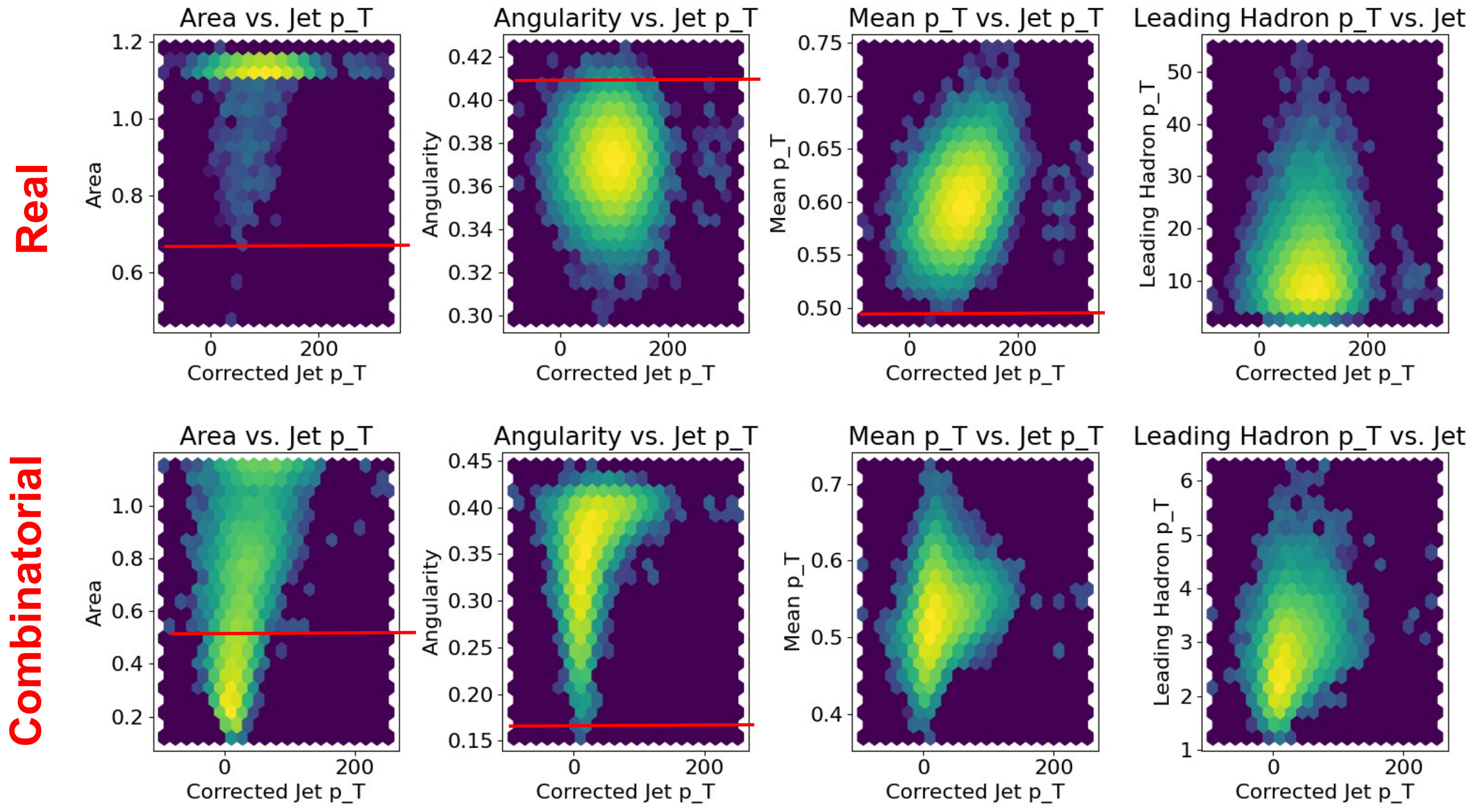
p_{Thard}>40 GeV/*c*, R=0.5

Log z scale



$p_{T hard}$>40 GeV/c, R=0.6

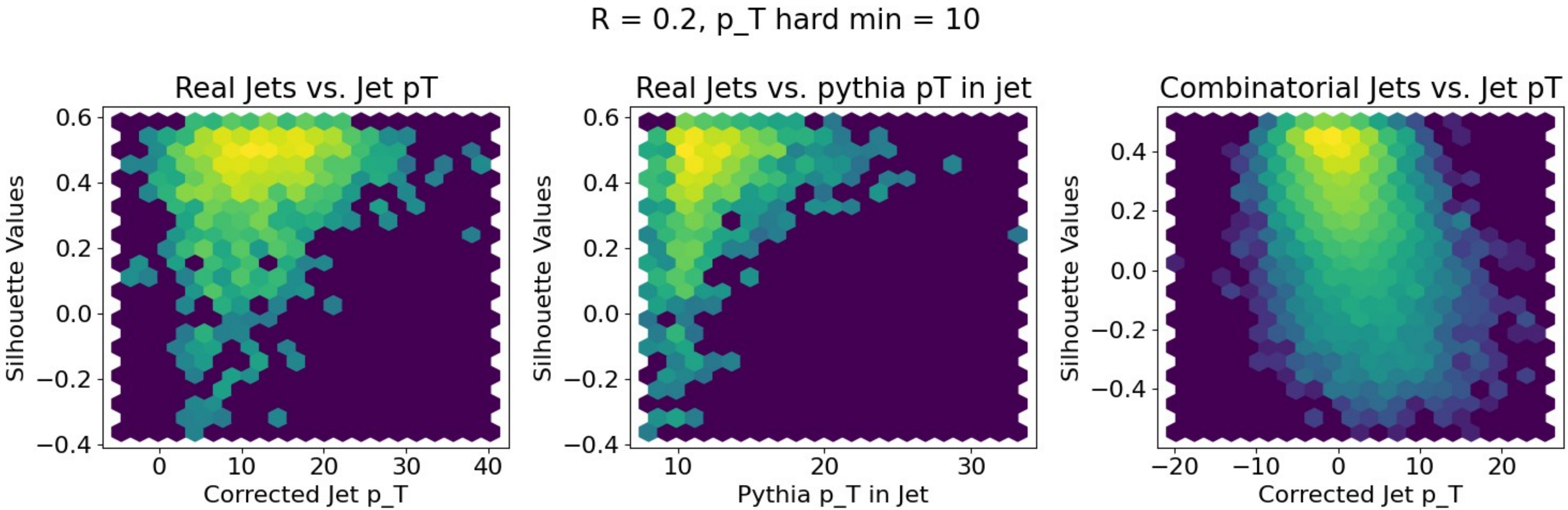
Log z scale



Backup: silhouette scores

Log z scale

Silhouette values – decreasing p_T

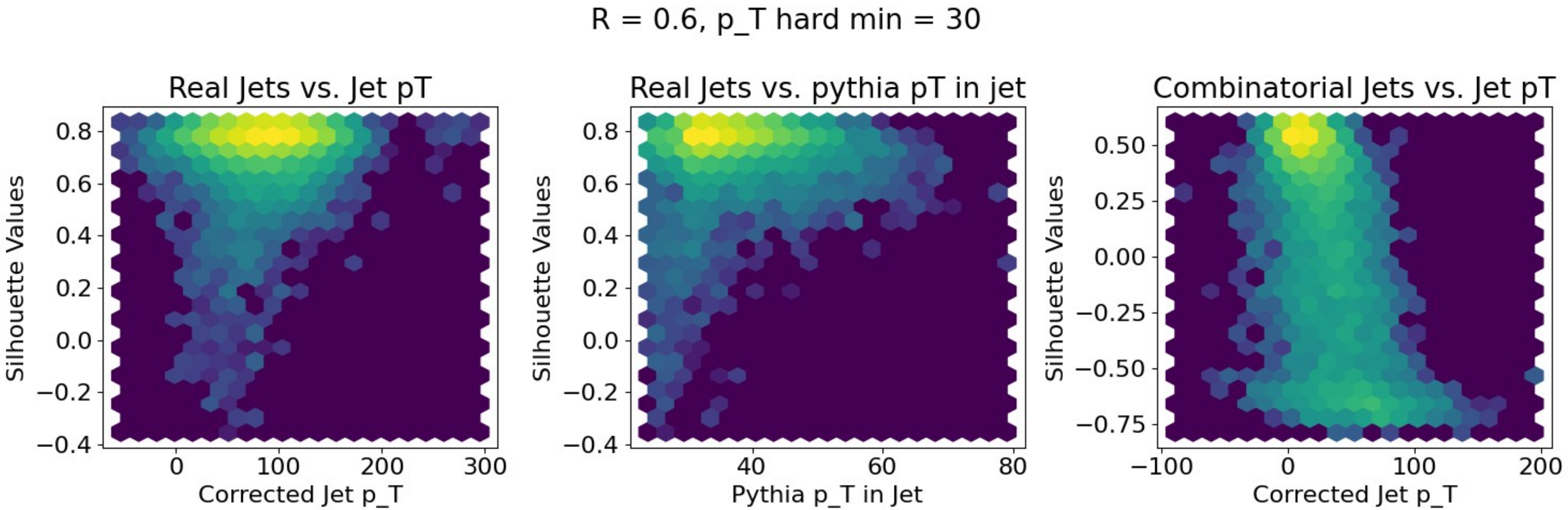


Real jets look more like combinatorial jets

Combinatorial jets look more like real jets

Log z scale

Silhouette values – increasing R



Real jets look more like real jets

Tail in distribution of real jets gets smaller

Combinatorial jets look more like real jets

Backup: jet definition

Jets in principle

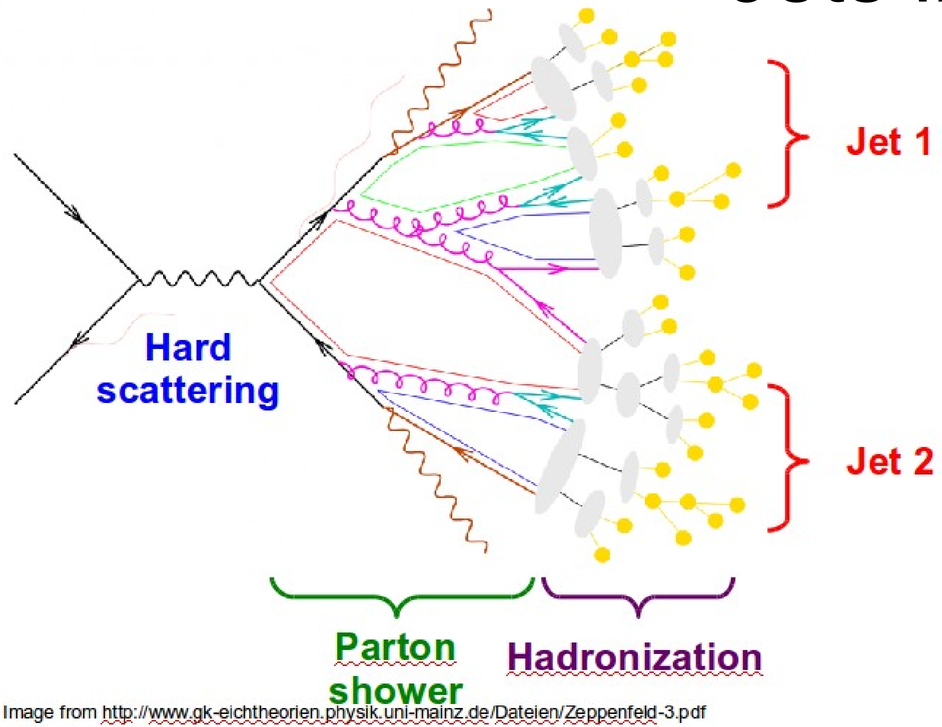
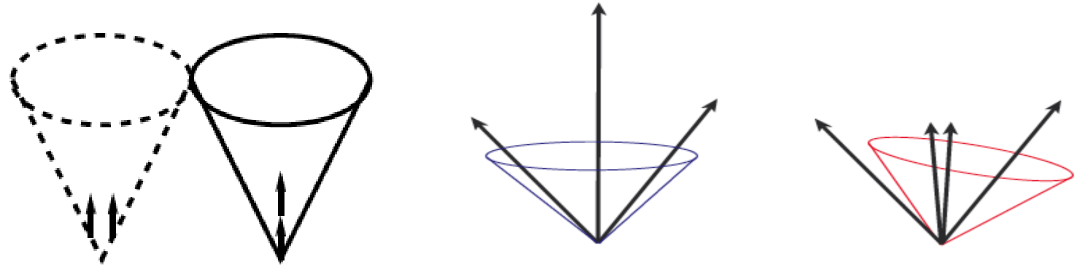
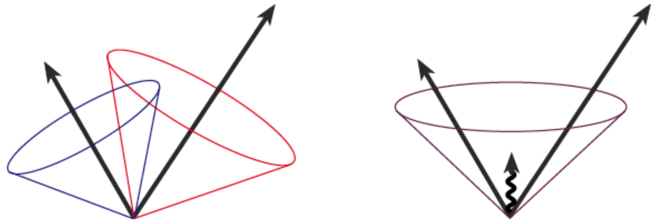


Image from http://www.gk-eichtheorien.physik.uni-mainz.de/Dateien/Zeppenfeld-3.pdf

- Jet measures **partons**
- Hadronic degrees of freedom are integrated out
- Algorithms are infrared and colinear safe





- **OK**
- **BAD**: 2 jets are merged in one

k_T jet finding algorithm

Particles, clusters

k_T algorithm

$$k_T = p_T, \Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$$

- For all *i,j* calculate:

$$d_{ij} = min\left(p_{T,i}^2, p_{T,j}^2\right)\frac{\Delta R_{ij}^2}{R^2}$$

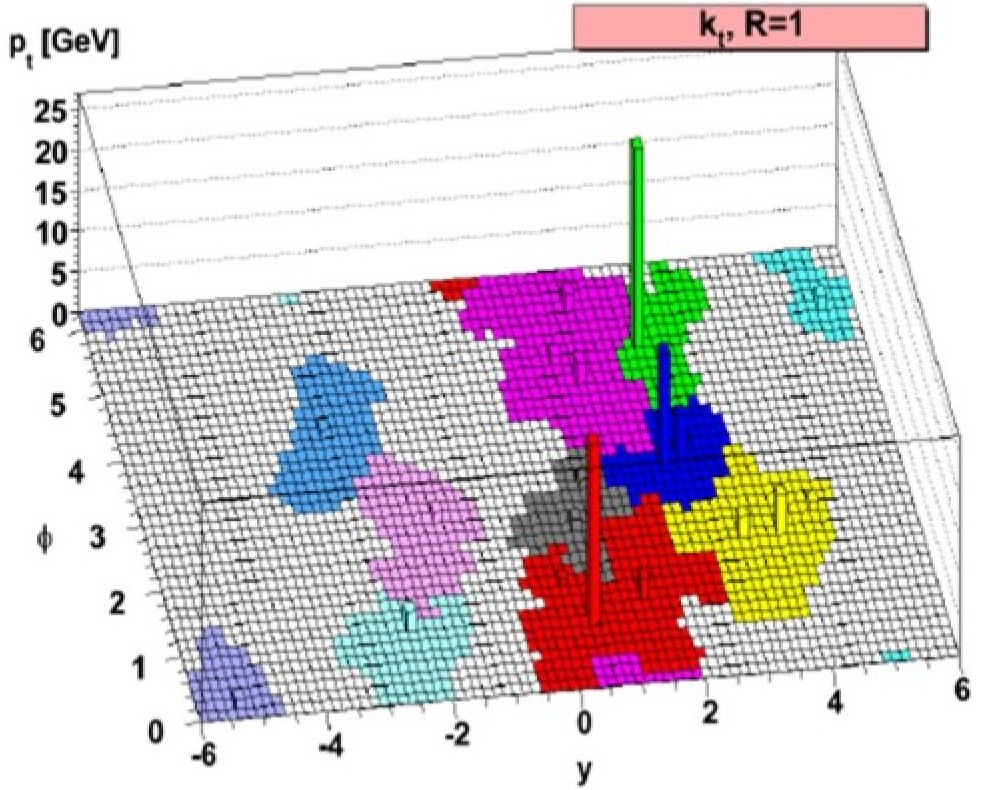
$$d_{iB} = p_{T,i}^2$$

- Combine smallest d_{ij}.

If d_{iB} smallest, $d_{iB} \rightarrow$ jet

Repeat until no particles left

Jet candidates



anti-k_T jet finding algorithm

Particles, clusters

k_T algorithm

$$k_T = p_T, \Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$$

- For all *i,j* calculate:

$$d_{ij} = min\left(p_{T,i}^{-2}, p_{T,j}^{-2}\right)\frac{\Delta R_{ij}^2}{R^2}$$

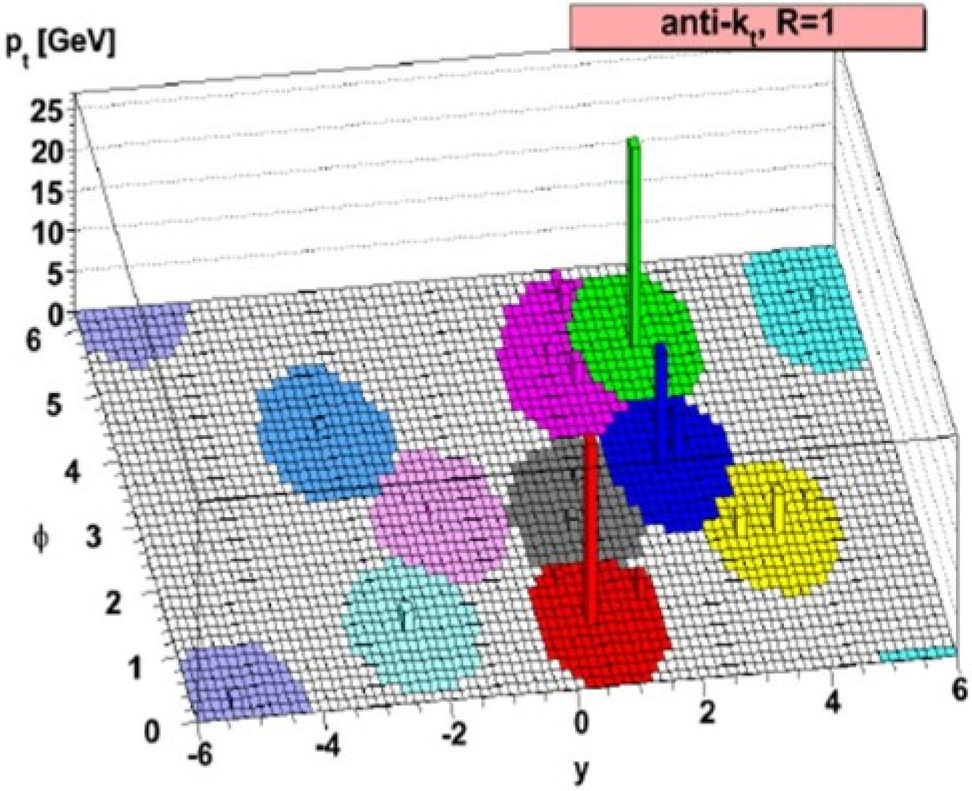
$$d_{iB} = p_{T,i}^{-2}$$

- Combine smallest d_{ij}.

  If d_{iB} smallest, $d_{iB} \rightarrow$ jet

Repeat until no particles left

Jet candidates



Cambridge/Aachen jet finding algorithm

Particles, clusters

k_T algorithm

$$k_T = p_T, \Delta R_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$$

- For all *i,j* calculate:

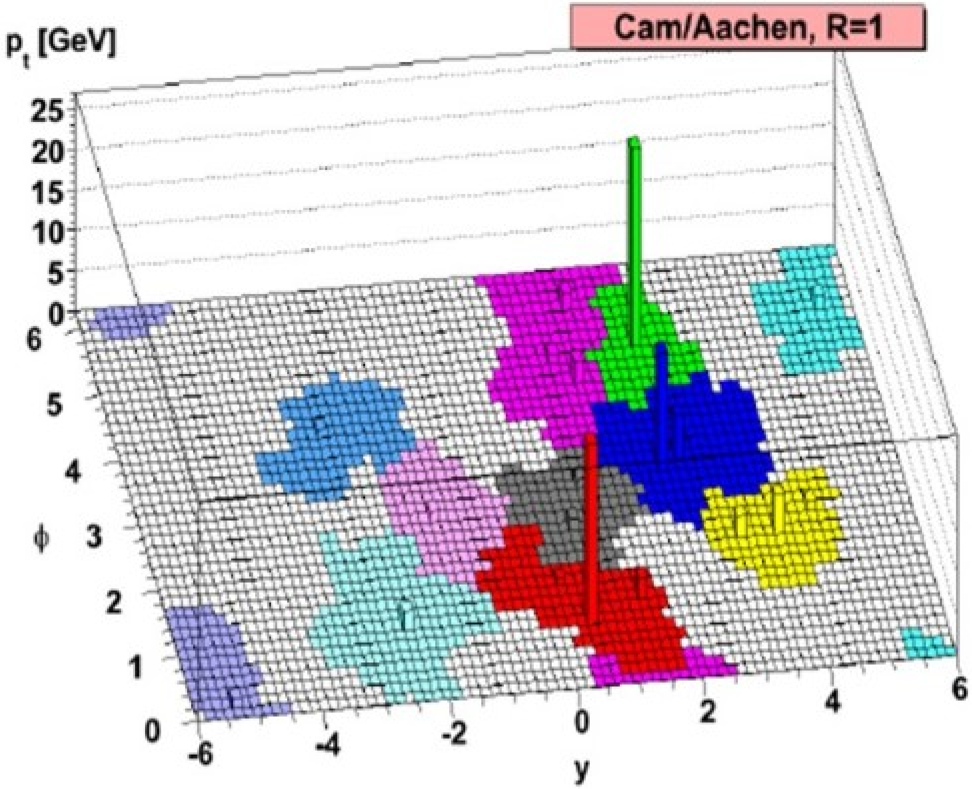
$$d_{iB} = 1 \qquad d_{ij} = \frac{\Delta R_{ij}^2}{R^2}$$

- Combine smallest d_{ij}.

  If d_{iB} smallest, $d_{iB} \rightarrow$ jet

Repeat until no particles left

Jet candidates



Backup: misc

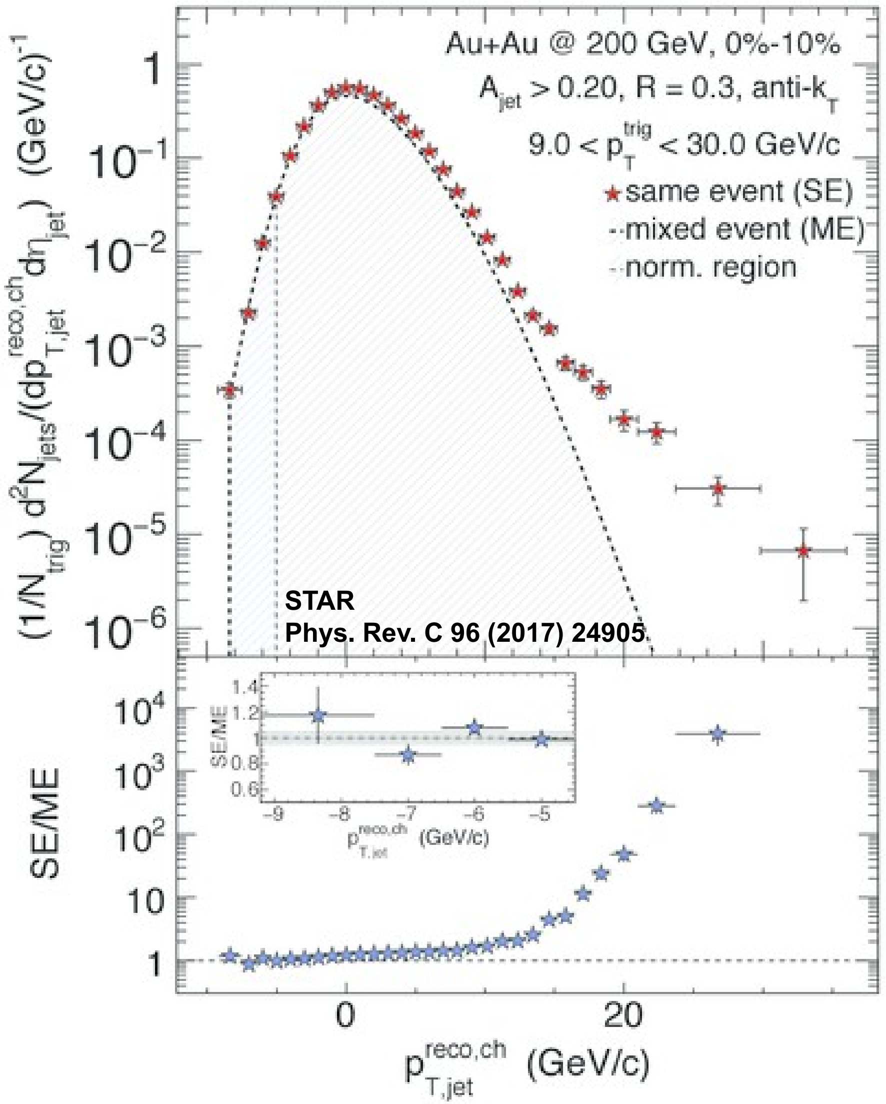
Unfolding

$$\vec{v} = R\vec{\mu} + \vec{\beta}$$

$$v_i = \sum_{j=1}^{M}(R_{ij}\mu_j) + \beta_i$$

- $\vec{\mu}$: the "true" histogram
- $\vec{v}$: the actual data we measure
- $\vec{\beta}$: background
- R : the response matrix

May correct for "missing" jets!



Mixed events

- Gets background up to a normalization factor
- Good agreement with the data… but 20% discrepancies still within uncertainties
- In measurement with background suppressed (h-jet correlations)
- Did not see such agreement at the LHC for jet spectra

Mini-summary

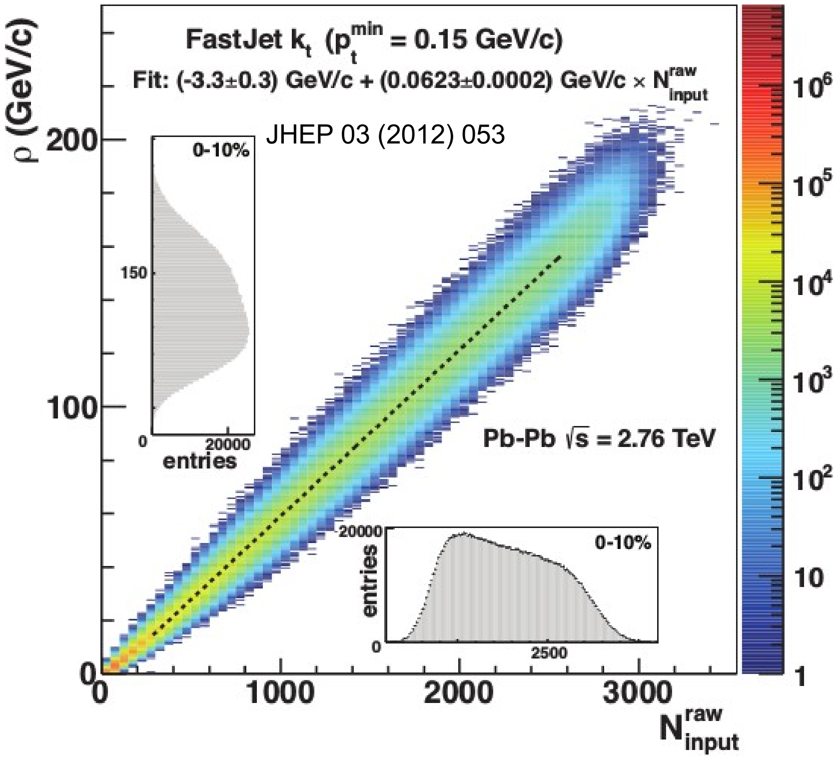
- Experimental techniques can bias measurement in subtle ways
  - Background subtraction
  - Kinematic cuts
  - Choice of jet finder, R
  - Centrality determination
  - Technique for finding reaction plane
- Larger influence at low momentum
- Safest to do the same analysis on data and model
  - But unfolding is necessary in a full Monte Carlo model!

Experimental techniques for background

Focus on smaller angles

- Pros
  - Background is smaller
  - Background fluctuations smaller
- Cons:
  - Modifications expected at higher R
  - Biases sample towards quarks

Aside: "quark" and "gluon" jet only defined at leading order.

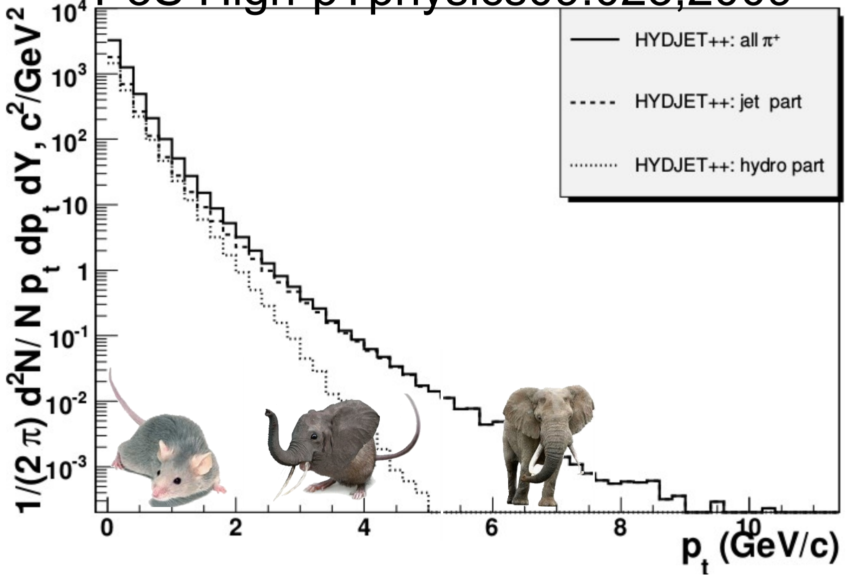


Focus on high p_T

- Pros:
  - Reduces combinatorial background
- Cons:
  - Cuts signal where we expect modifications
  - Could bias towards partons which have not interacted
  - Biases sample towards quark jets

"Quark" and "gluon" jets only defined at leading order!

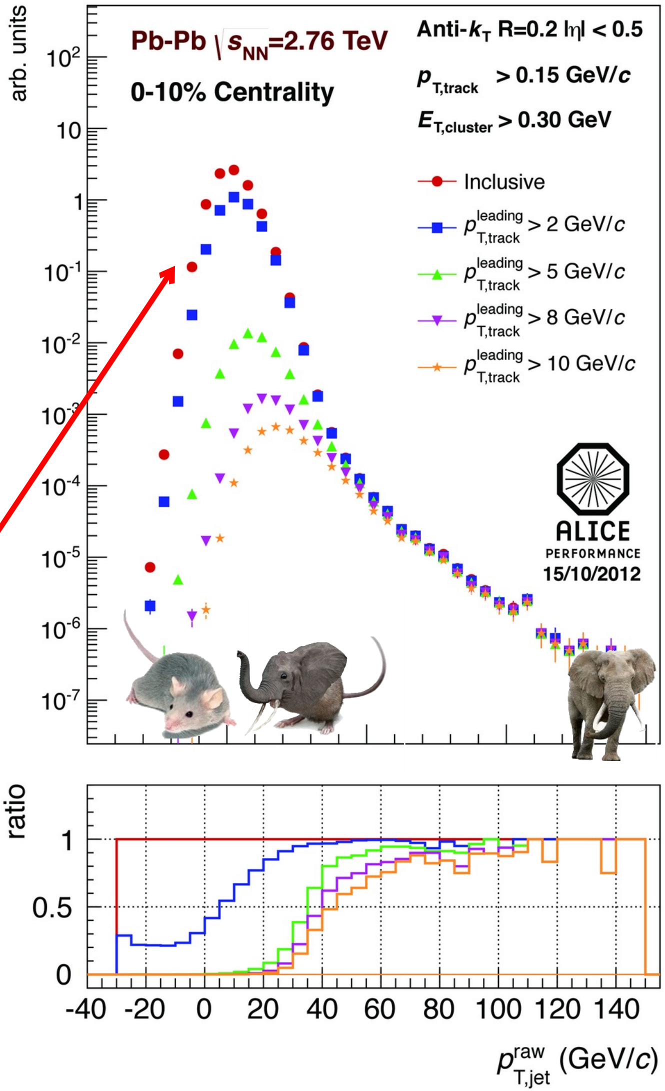
PoS High-pTphysics09:023,2009



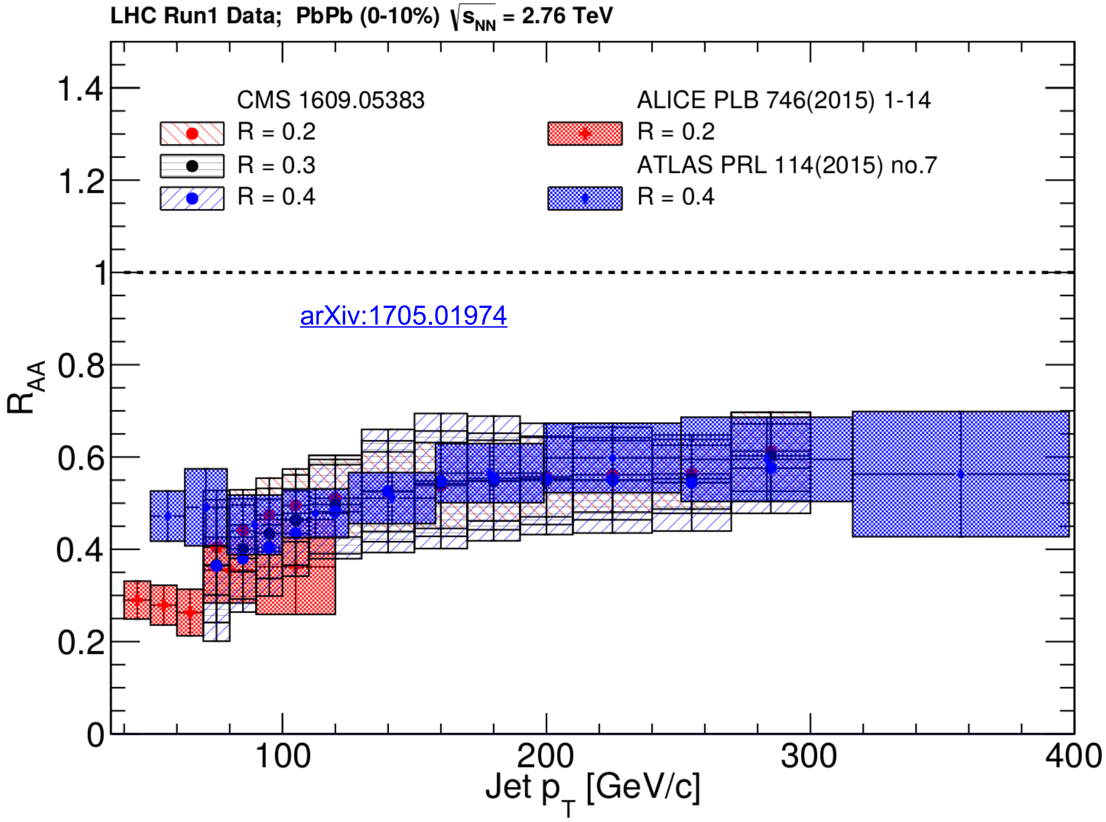
Area-based subtraction

- ALICE/STAR
- Require leading track p_T > 5 GeV/c
  - Suppresses combinatorial “jets”
  - Biases fragmentation
- No threshold on constituents
- Limited to small R – unstable unfolding

Combinatorial jets

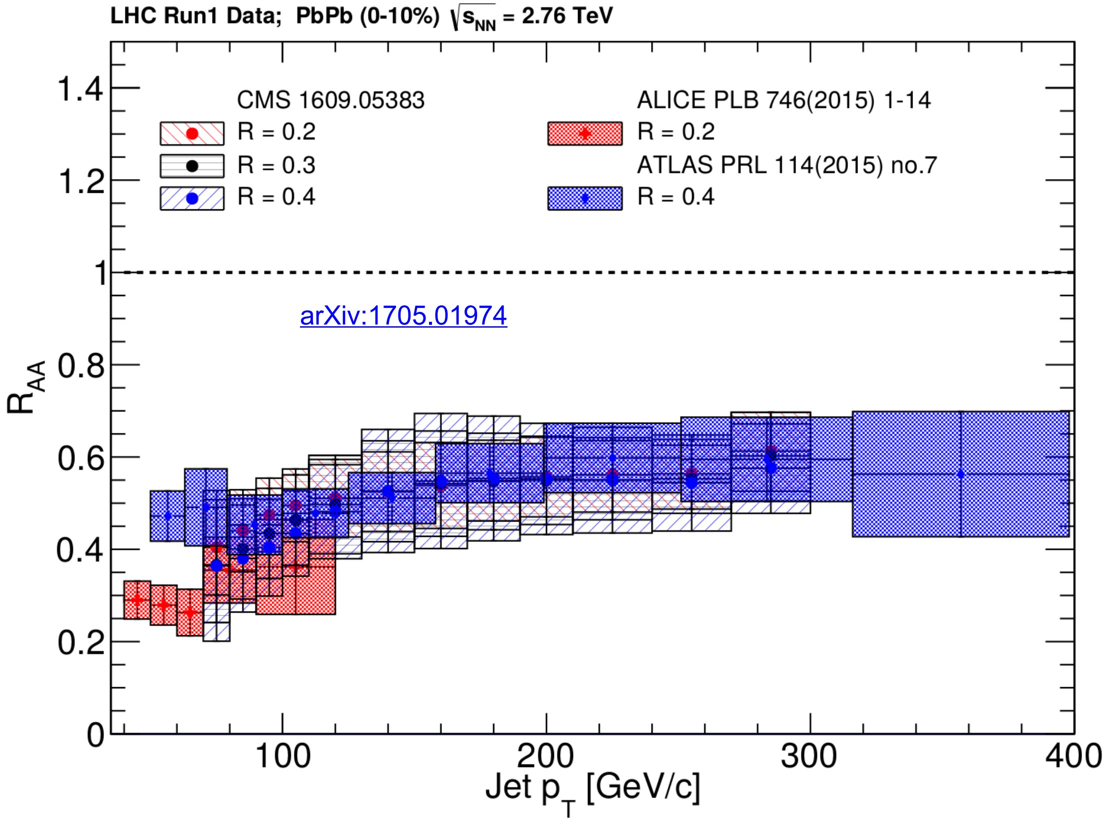


Jet R_{AA}



LHC Run1 Data; PbPb (0-10%) $\sqrt{s_{NN}}$ = 2.76 TeV
CMS 1609.05383
R = 0.2
R = 0.3
R = 0.4
ALICE PLB 746(2015) 1-14
R = 0.2
ATLAS PRL 114(2015) no.7
R = 0.4
arXiv:1705.01974
R_{AA}
0
0.2
0.4
0.6
0.8
1
1.2
1.4
100
200
300
400
Jet p_T [GeV/c]

Jet R_{AA}



Tension between ATLAS & ALICE/CMS

Mini-summary

- Most studies do one or more of the following:
  - Explicitly apply a (non-purturbative) bias
  - *Implicitly* apply a (non-purturbative) bias
  - Focus on small R
  - Focus on high pT
- May also → survivor bias
- Background subtraction should be part of definition of algorithm