Fluctuations in the Background for Jets:

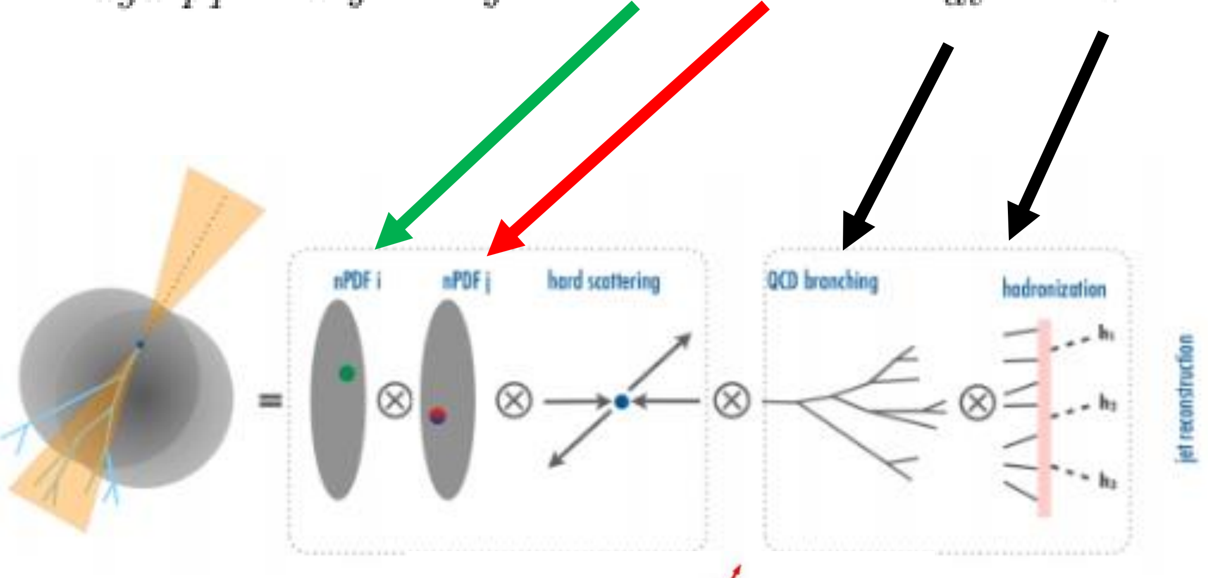
Model Studies, Mitigation, Machine Learning and More

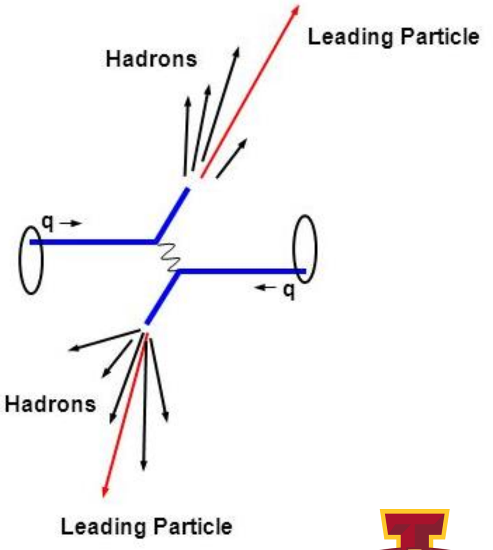
Charles Hughes
Iowa State University
2/07/2023

What Is a Jet ?

- In a pp/AA collisions, 2 incoming partons scatter with large momentum transfer
- Scattered partons fragment and hadronize
- This forms back-to-back, collimated sprays of particles – jets

$$\frac{d^3\sigma^h}{dy\,d^2p_T} = \frac{1}{\pi}\int dx_a \int dx_b f_a^A(x_a) f_b^B(x_b) \frac{d\sigma_{ab\to cX}}{d\hat{t}} \frac{D_c^h(z)}{z}$$

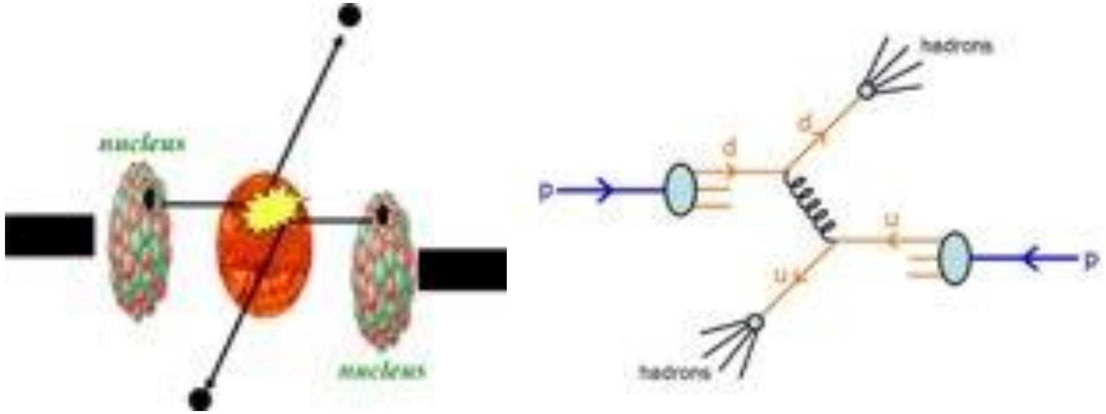




Why Study Jets ?

- Internally generated probe of QGP
- Jets lose energy in medium (AA) relative to vacuum (pp)
- Energy loss is path length dependent - giving us information about Quark Gluon Plasma (QGP) properties

Absolute Energy Loss (medium vs. vacuum)



Relative Energy Loss (more mediums vs. less medium)

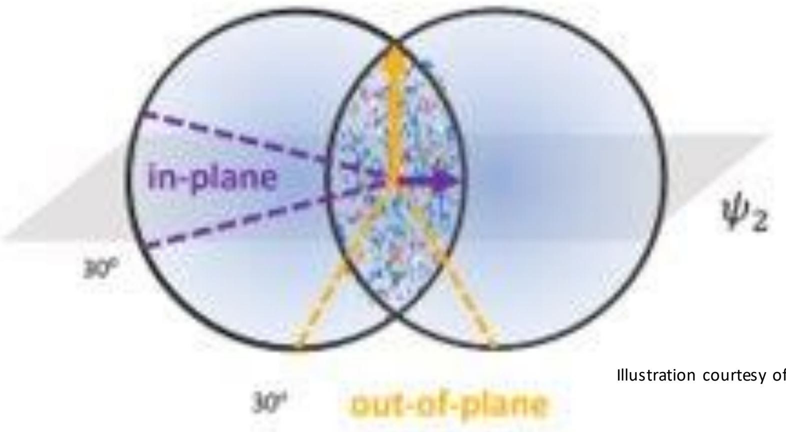
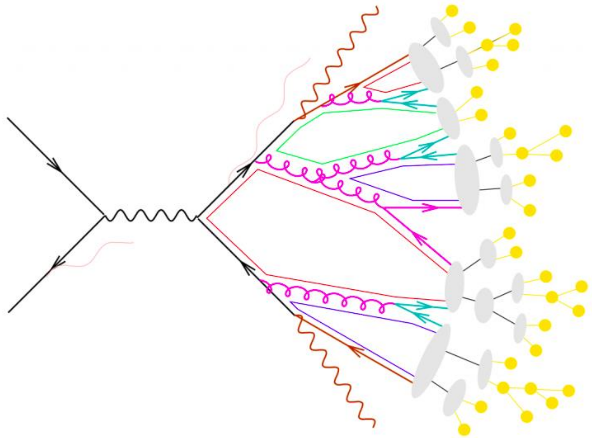


Illustration courtesy of Caitilin Beattie

Reconstructing Jets



- Jets are ambiguous objects – even at generator level

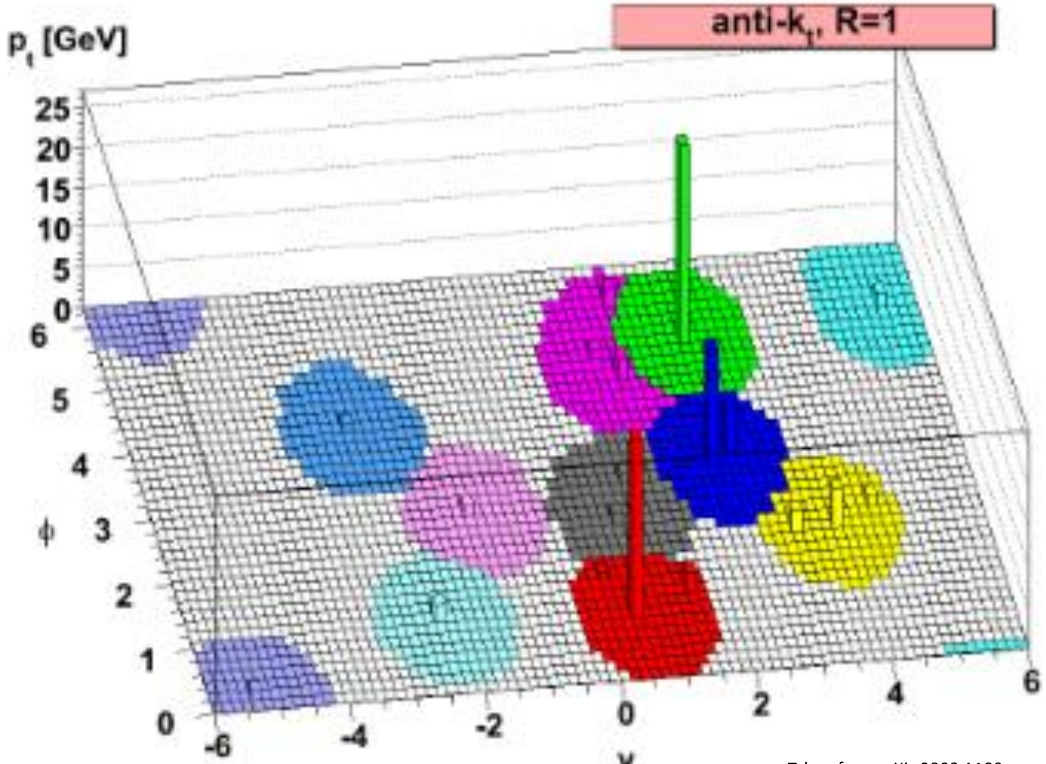
- Experiments rely on **reconstruction algorithms** (decide which particles go in the jet) and **recombination schemes** (decide how to calculate jet properties from particle properties)

- Example – anti-k_T algorithm w/ boost invariant p_T scheme

$$d_{ij} = \min(k_{ti}^{2p}, k_{tj}^{2p}) \frac{\Delta_{ij}^2}{R^2} ,$$

$$d_{iB} = k_{ti}^{2p} ,$$

p= -1



Taken from arXiv:0802.1189

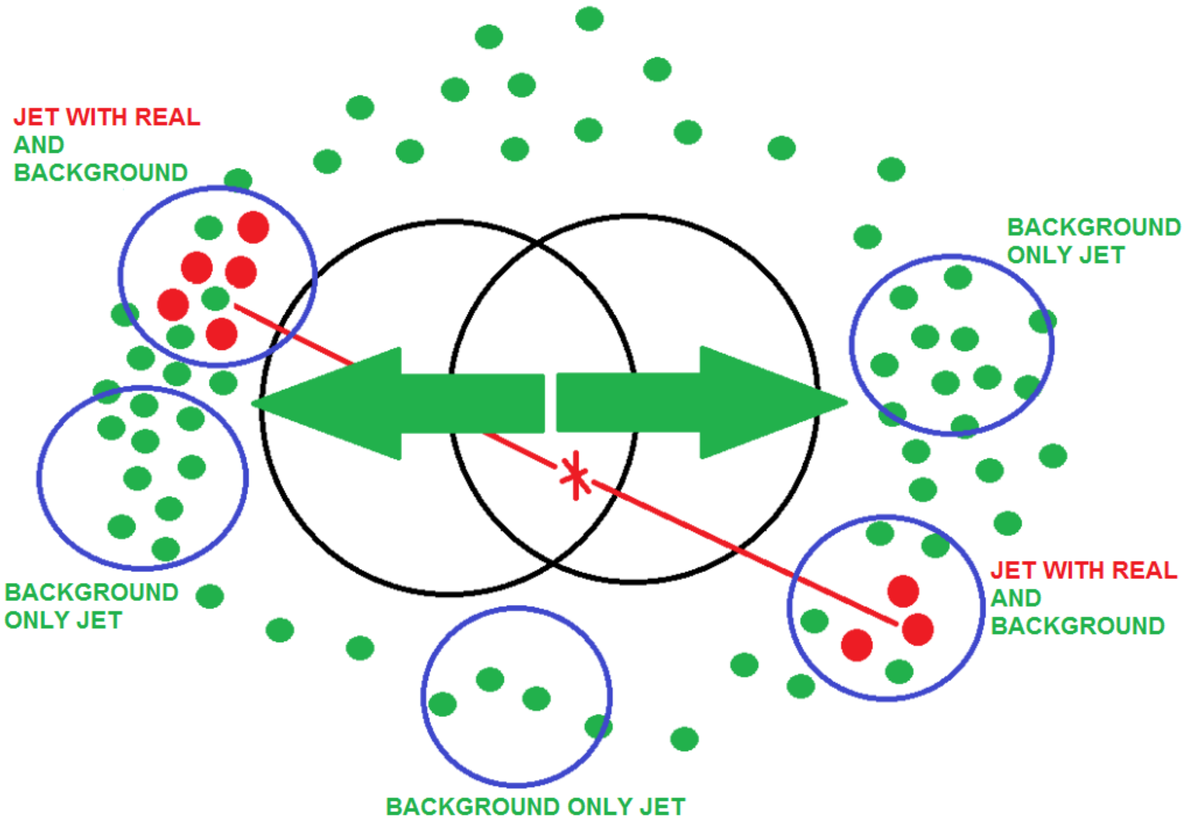
The Problem of Jet Background

- Simplified picture
  - **Signal/Real particles** from hard scatterings
  - **Background particles** from soft processes

- **Background fluctuates in η, ϕ, event-to-event**

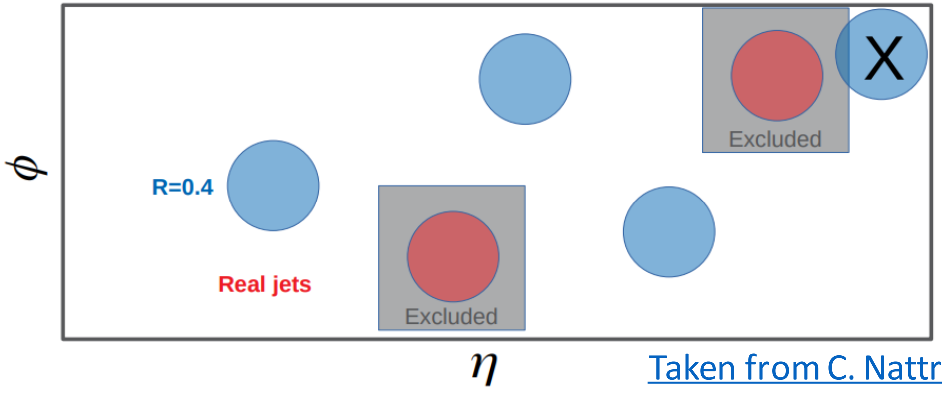
- **Jets** with **combinatorial background**

- **Jets composed of entirely combinatorial background**

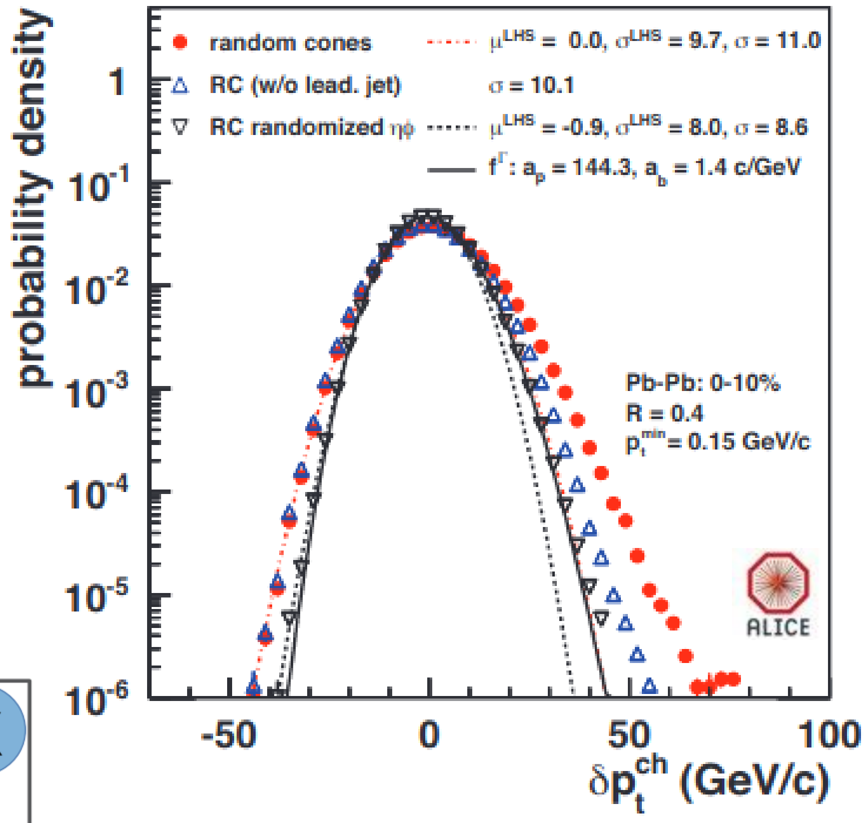


Measuring Background Fluctuations

- Following the analysis in *JHEP 03 (2012) 053*
  - Estimate rho
    - k_T jet finder
    - ρ = median(p_T/A)
  - Estimate jet background fluctuations
    - Draw random cone
    - $\delta p_T = p_T^{cone} - \rho * A^{cone}$
    - CalcuLate $\sigma(\delta p_T)$ - next slide



[Taken from C. Nattrass]



Taken from *JHEP 03 (2012) 053*

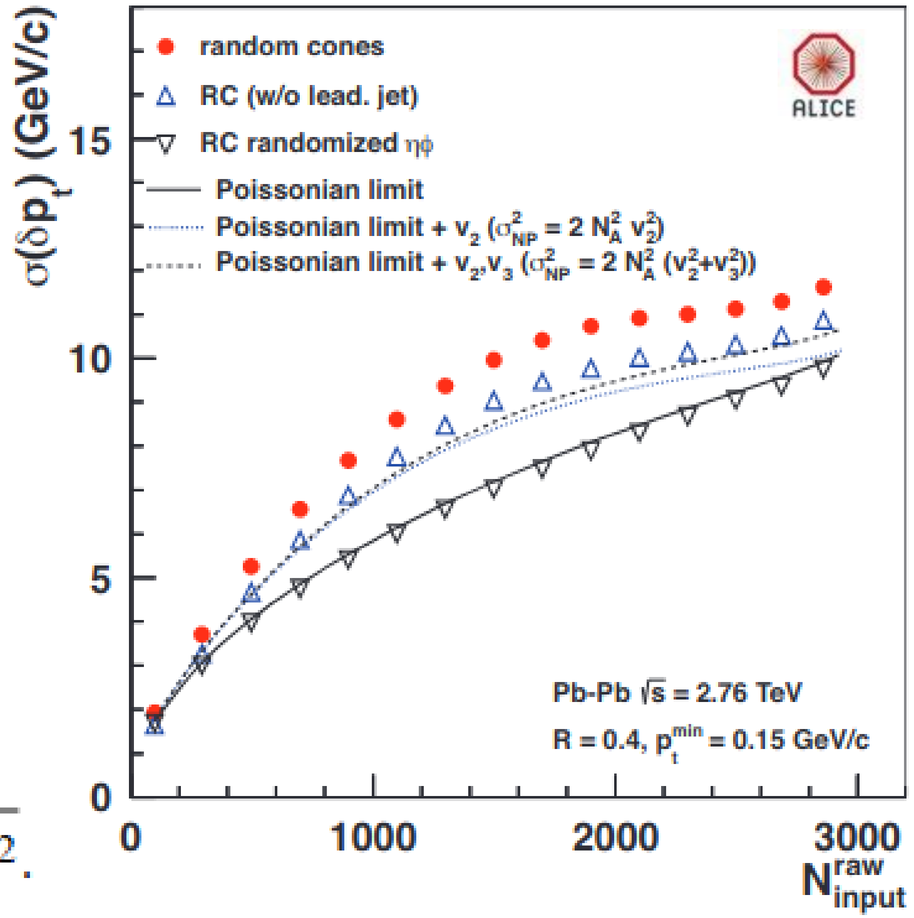
Characterizing Background Fluctuations

- Following the analysis in *JHEP 03 (2012) 053*
  - Calculate $\sigma(\delta p_T)$
  - Compare to model as in ([Tannenbaum et. al.]())
    - Assumes single particle p_T spectrum is gamma distribution, then:

N_A = # of particles in cone

Assuming no flow $$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3 $$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$



Characterizing Background Fluctuations

- Following the analysis in *JHEP 03 (2012) 053*
  - Calculate $\sigma(\delta pT)$
  - Compare to model as in (Phys.Lett.B 498 (2001) 29-34)
    - Assumes single particle pT spec

Width of N-fold convolution of gamma pT spectrum term

Poissonian fluctuations term

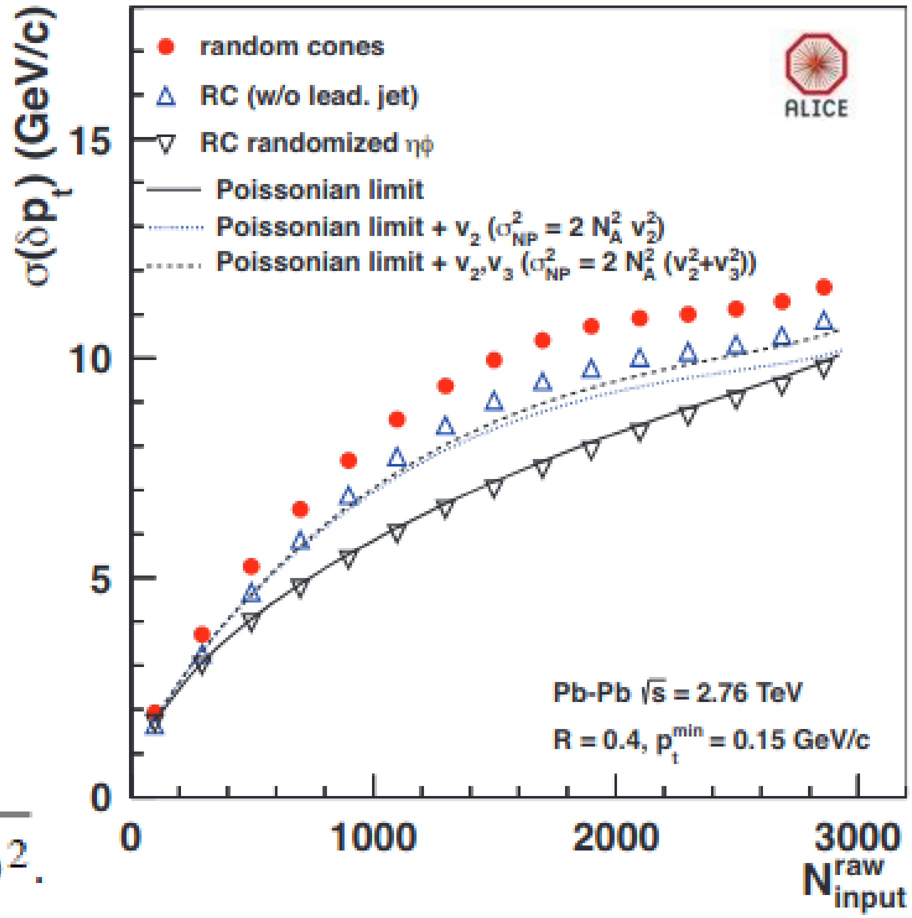
N_A = # of particles in cone

Assuming no flow

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$



Characterizing Background Fluctuations

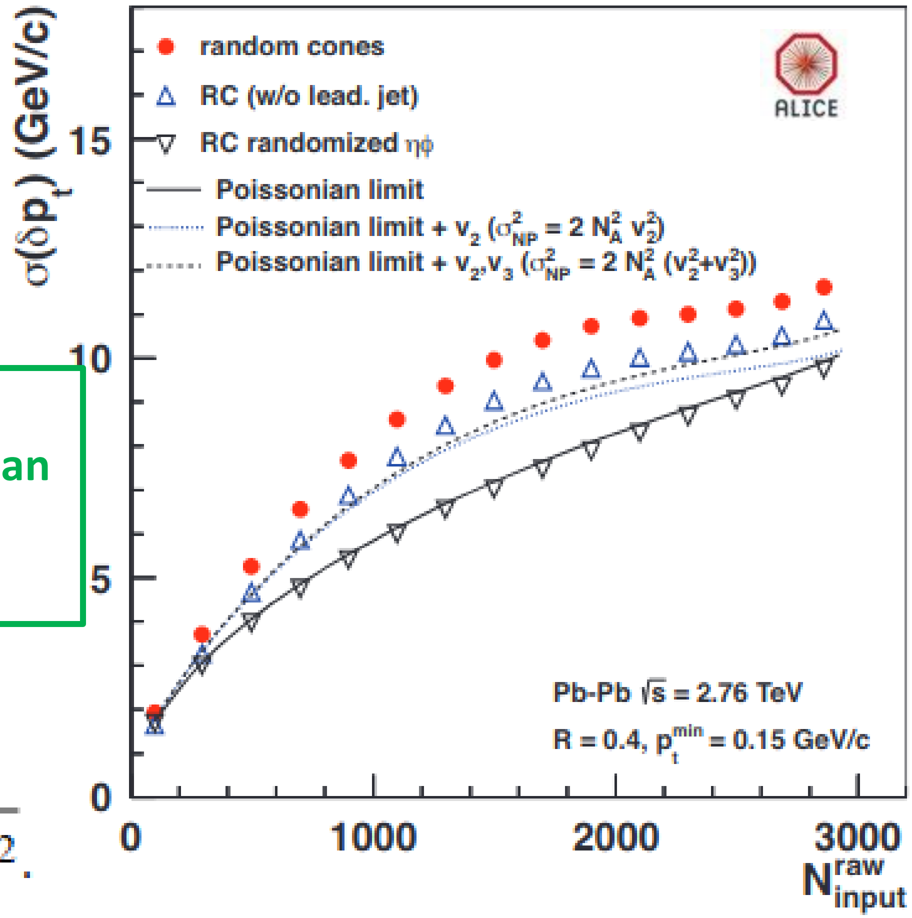
- Following the analysis in *JHEP 03 (2012) 053*
  - Calculate $\sigma(\delta pT)$
  - Compare to model as in (Phys.Lett.B 498 (2001) 29-34)
    - Assumes single particle pT spectrum is gamma distribution, then:

N_A = # of particles in cone

Non-Poissonian term (v2/v3)

Assuming no flow $$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3 $$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$



$$\sigma^2_{NP}(N_A) \approx 2v_2^2 N_A^2 \quad \text{OR} \quad (\sigma^2_{NP}(N_A) \approx 2N_A^2(v_2^2 + v_3^2))$$

Background Fluctuations - Model Studies

- Following the analysis in *JHEP 03 (2012) 053*

- Some questions to ask:

  •What can we learn from this background characterization in simple model studies ?

  •What implications does this have for background mitigation in jet observables in data ?

"Model studies of fluctuations in the background for jets in heavy ion collisions"

Phys. Rev. C **106**, 044915 – Published 31 October 2022

Charles Hughes, Antonio Carlos Oliveira da Silva, and Christine Nattrass

Charles Hughes

Antonio Da Silva

Christine Nattrass

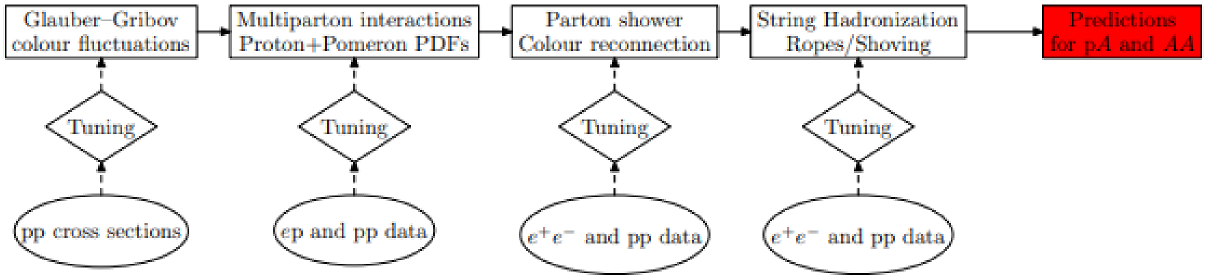
Background Fluctuations - Model Studies

- What can we learn from this background characterization in simple model studies ?

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

- Looking at 2 models

    - Angantyr Pythia - arXiv:1806.10820
        - MPI/Diffractive Excitation



    - TennGen - (github)
        - Next Slide

Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

- TennGen:

What TennGen is for:

a) Computationally cheap way to generate particles with realist pT spectrum and flow as in heavy ion collisions (and NO OTHER correlations)

b) Understanding how a realistic heavy ion background affects jet finders/jet observables

c) Development of background subtraction/mitigation techniques

d) Seeing how analysis depends on background with/without v_1, v_2, v_3, etc...

Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

- TennGen:

What TennGen is for:

a) Computationally cheap way to generate particles with realist pT spectrum and flow as in heavy ion collisions (and NO OTHER correlations)

b) Understanding how a realistic heavy ion background affects jet finders/jet observables

c) Development of background subtraction/mitigation techniques

d) Seeing how analysis depends on background with/without v_1, v_2, v_3, etc...

What TennGen is NOT for:

a) NOT a replacement for physics based generators (HIJING, AMPT, JEWEL, etc...)

b) NOT for jet/background physics interaction model studies (e.g. quenching/energy loss/back-reaction)

c) NOT for testing hydro/flow models

Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

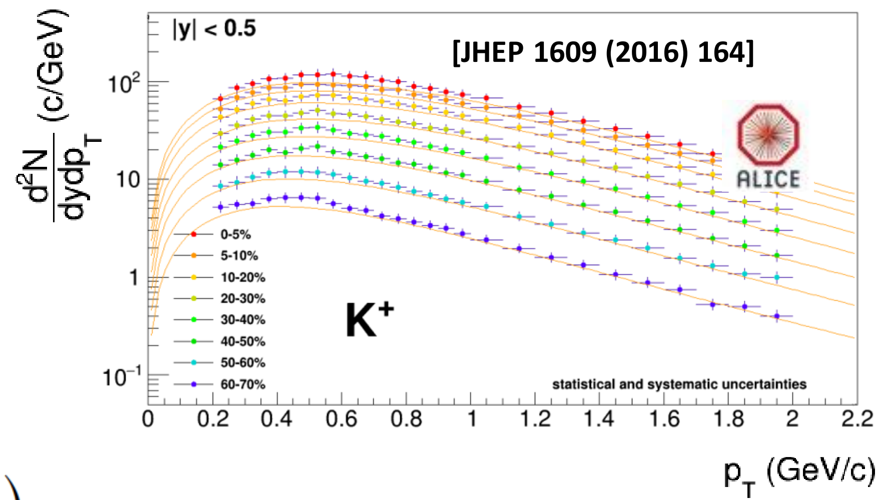
- TennGen:
  - Particle generator meant to simulate $\pi^{+/-/0}$, $K^{+/-}$, p, $\bar{p}$ in 2.76 TeV PbPb collisions (0-5% : 40-50 %)
  - Particle p_T according to fits of data to Boltzmann-Gibbs Blast Wave

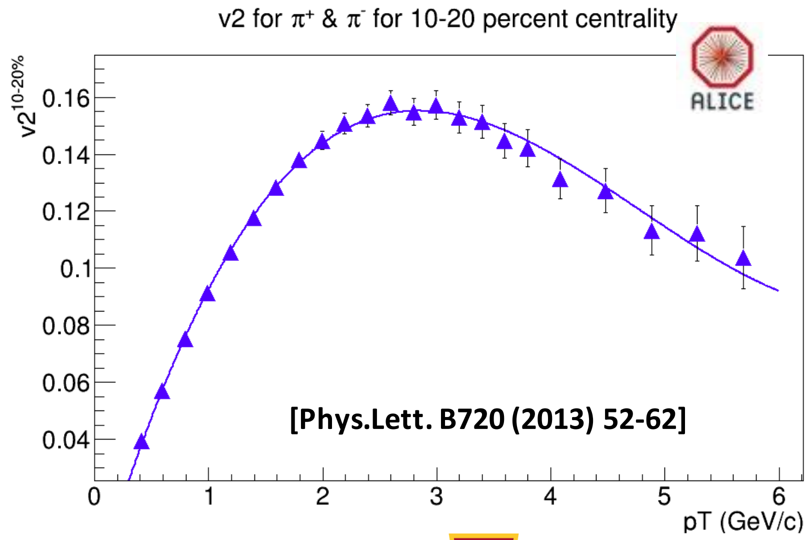
$$\frac{d^2N}{dp_T dy} = N p_T \int_0^1 r' dr' \left(\sqrt{m^2 + p_T^2}\right) \times I_0\left(\frac{p_T \sinh\left[\tanh^{-1}\left(\beta_s r'^n\right)\right]}{T_{\text{kin.}}}\right) \times K_1\left(\frac{\sqrt{m^2+p_T^2}\cosh\left[\tanh^{-1}\left(\beta_s r'^n\right)\right]}{T_{\text{kin.}}}\right)$$

  - $v_n(p_T)$ from polynomial fits to data (v_1 : v_5)
  - Particle ϕ from Fourier Sum

$$\frac{dN}{d\phi} = \frac{N_0}{2\pi}\left(1 + \sum_{n=1}^{5} 2v_n \cos[n(\phi - \Psi_n)]\right)$$

  - Particle η Uniform (mid-rapidity approximation)

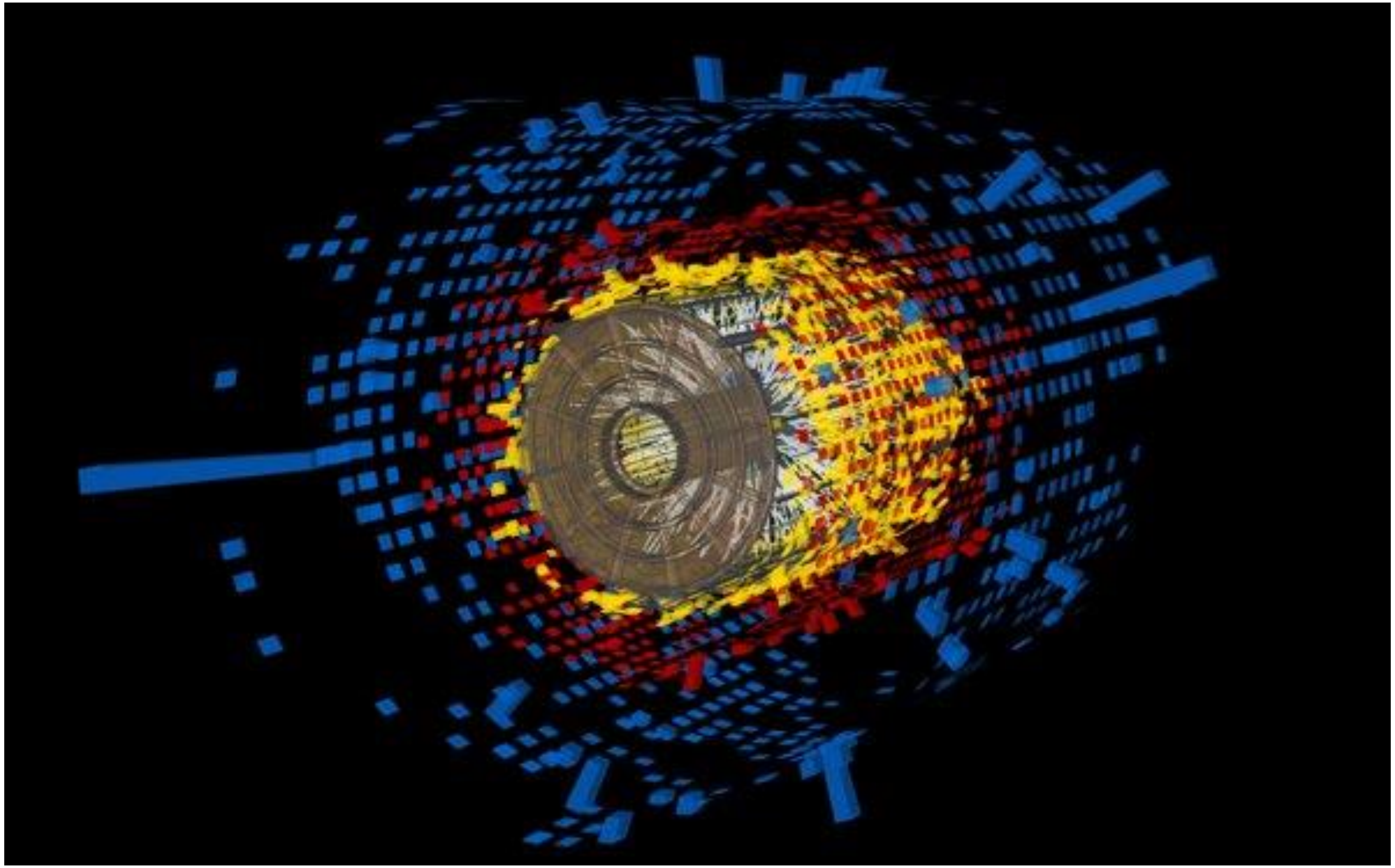




Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
[Phys. Rev. C 106, 044915]

- TennGen: 40-50 % 2.76 TeV PbPb event display for sPHENIX

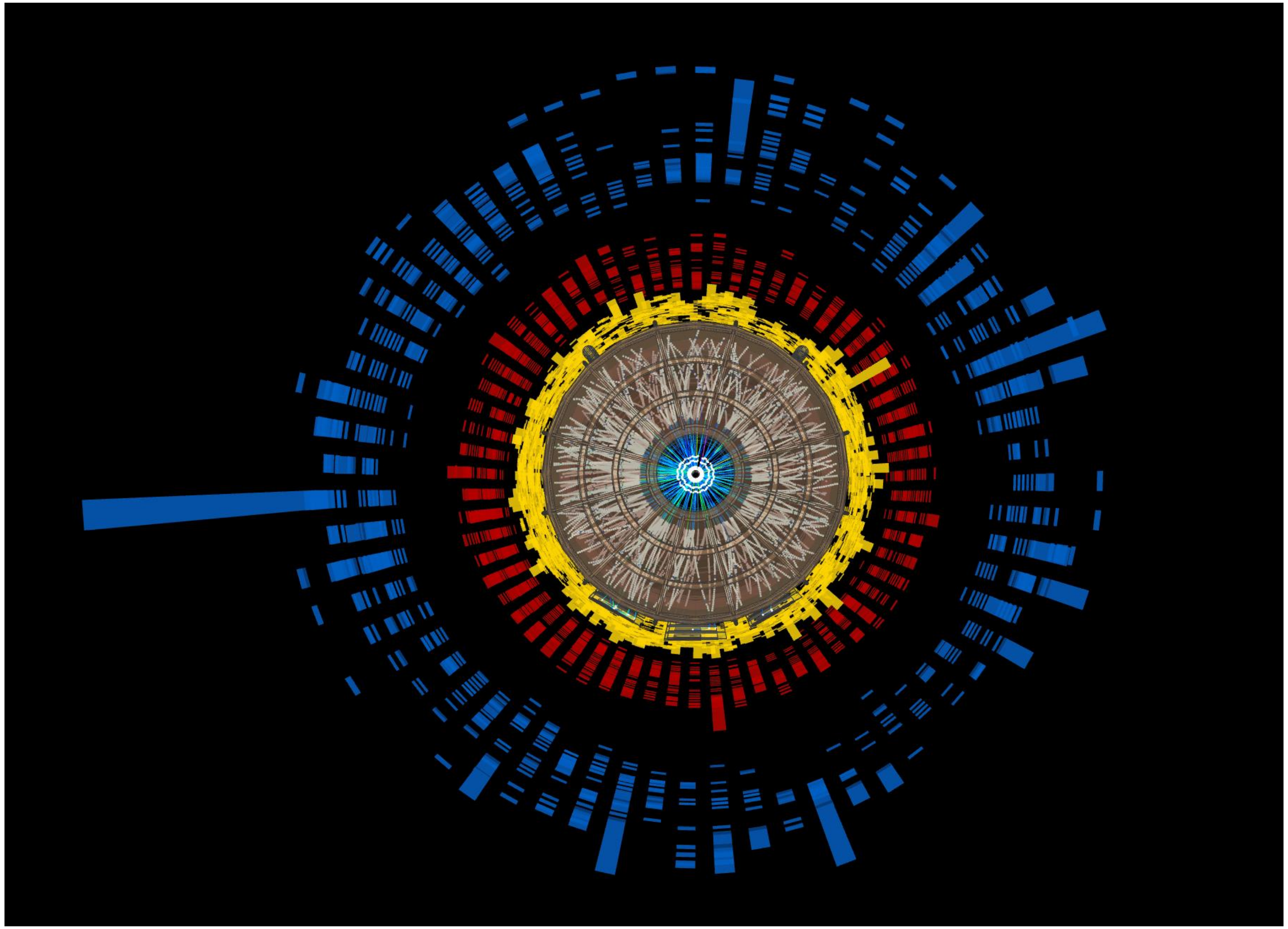



Ejiro Umaka

Background Fluctuations - TennGen

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

- TennGen: 40-50 % 2.76 TeV PbPb event display for sPHENIX



Ejiro Umaka

Background Fluctuations - Model Studies

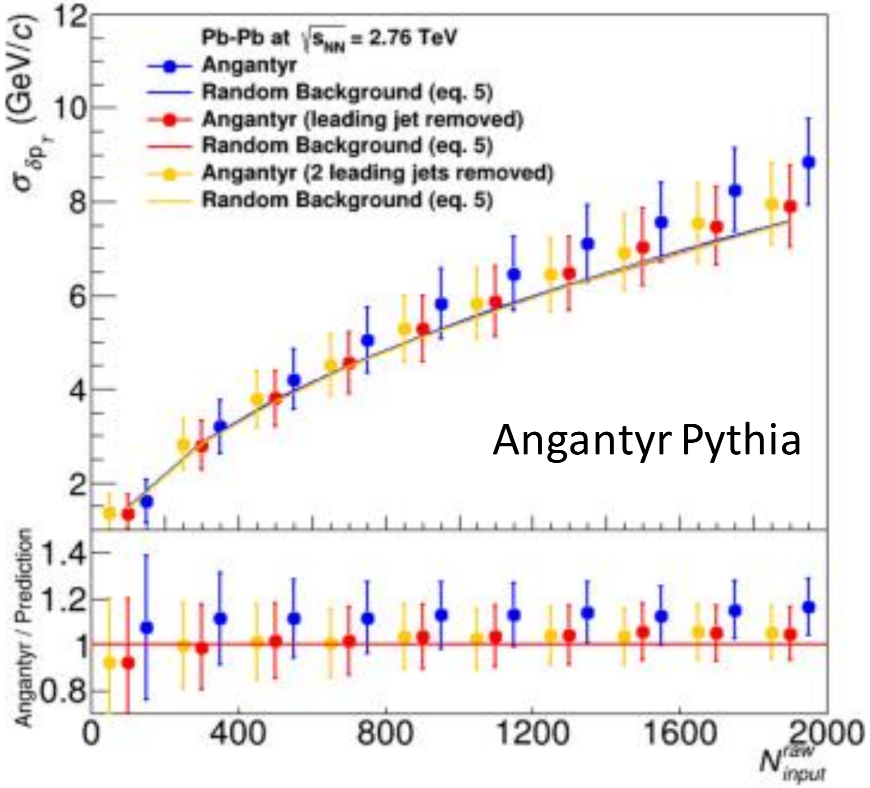
- What can we learn from this background characterization in simple model studies ?

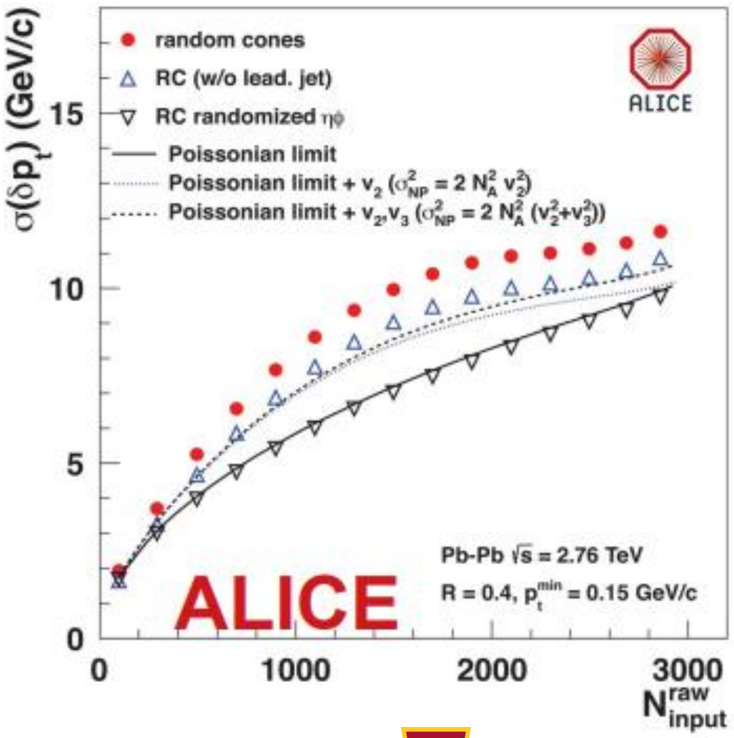
Hughes, da Silva, Nattrass
[Phys. Rev. C 106, 044915]

Abelev et. al. On behalf of ALICE
[JHEP 03 (2012) 053]

Eq. 5

$$\sigma(\delta p_{\mathrm{t}}) = \sqrt{N_{\mathrm{A}} \cdot \sigma^2(p_{\mathrm{t}}) + N_{\mathrm{A}} \cdot \langle p_{\mathrm{t}} \rangle^2}.$$





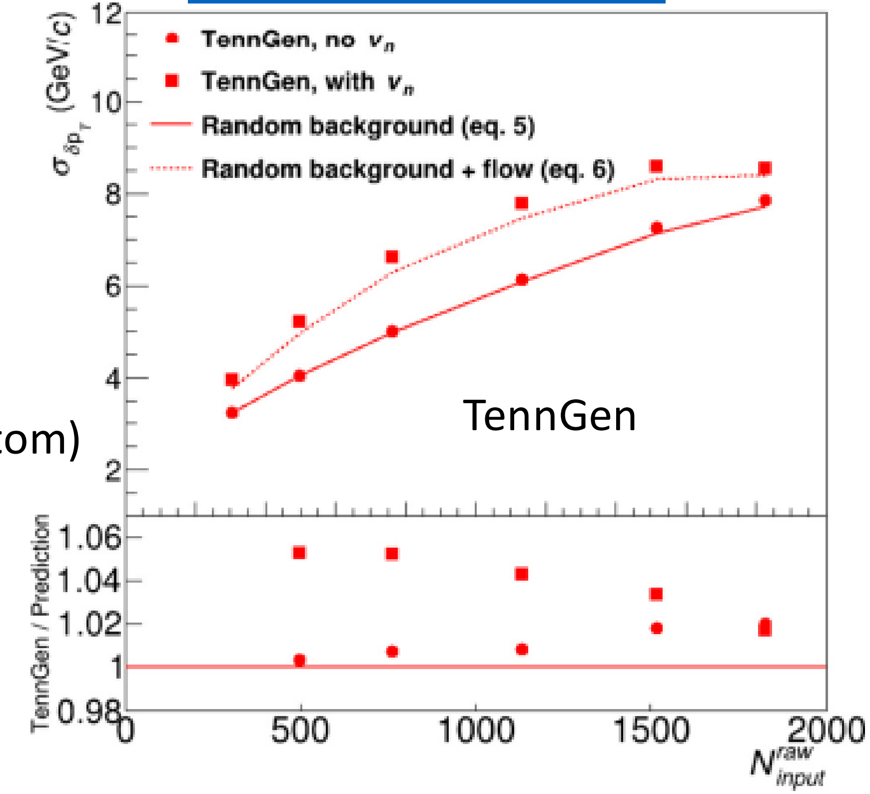
Background Fluctuations - Model Studies

- What can we learn from this background characterization in simple model studies ?

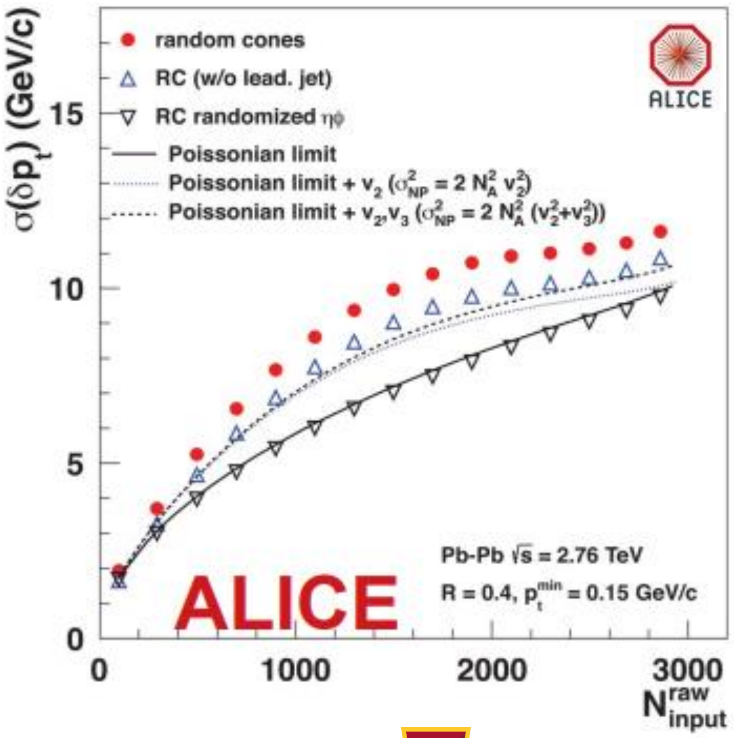
(with flow -top)

(without flow – bottom)

Hughes, da Silva, Nattrass
[Phys. Rev. C 106, 044915]



Abelev et. al. On behalf of ALICE
[JHEP 03 (2012) 053]



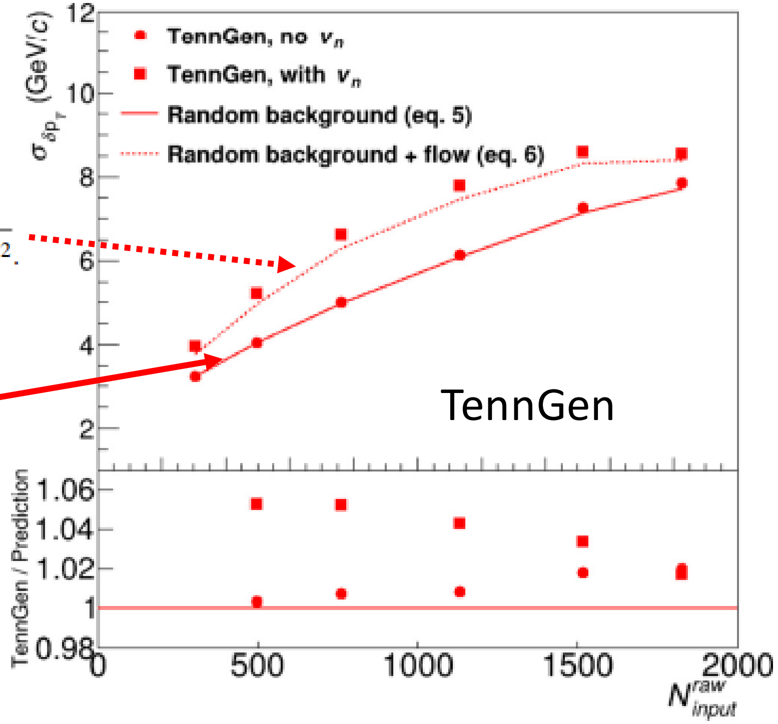
Background Fluctuations - Model Studies

- What can we learn from this background characterization in simple model studies ?

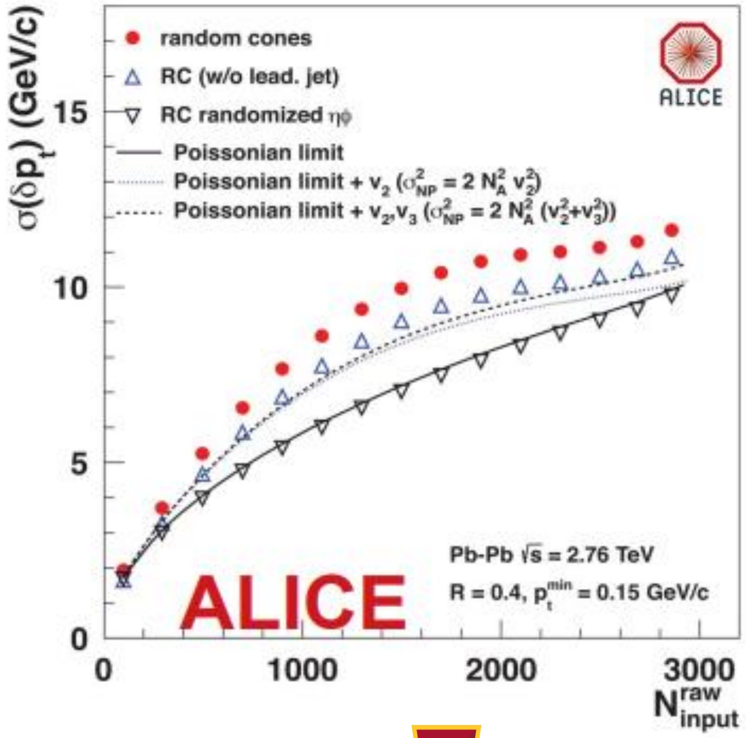
Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

Eq. 6 $\sigma(\delta p_{\mathrm{t}}) = \sqrt{N_{\mathrm{A}} \cdot \sigma^2(p_{\mathrm{t}}) + (N_{\mathrm{A}} + \sigma^2_{\mathrm{NP}}(N_{\mathrm{A}})) \cdot \langle p_{\mathrm{t}} \rangle^2}.$

Eq. 5 $\sigma(\delta p_{\mathrm{t}}) = \sqrt{N_{\mathrm{A}} \cdot \sigma^2(p_{\mathrm{t}}) + N_{\mathrm{A}} \cdot \langle p_{\mathrm{t}} \rangle^2}.$



Abelev et. al. On behalf of ALICE
JHEP 03 (2012) 053



Background Fluctuations - Model Studies

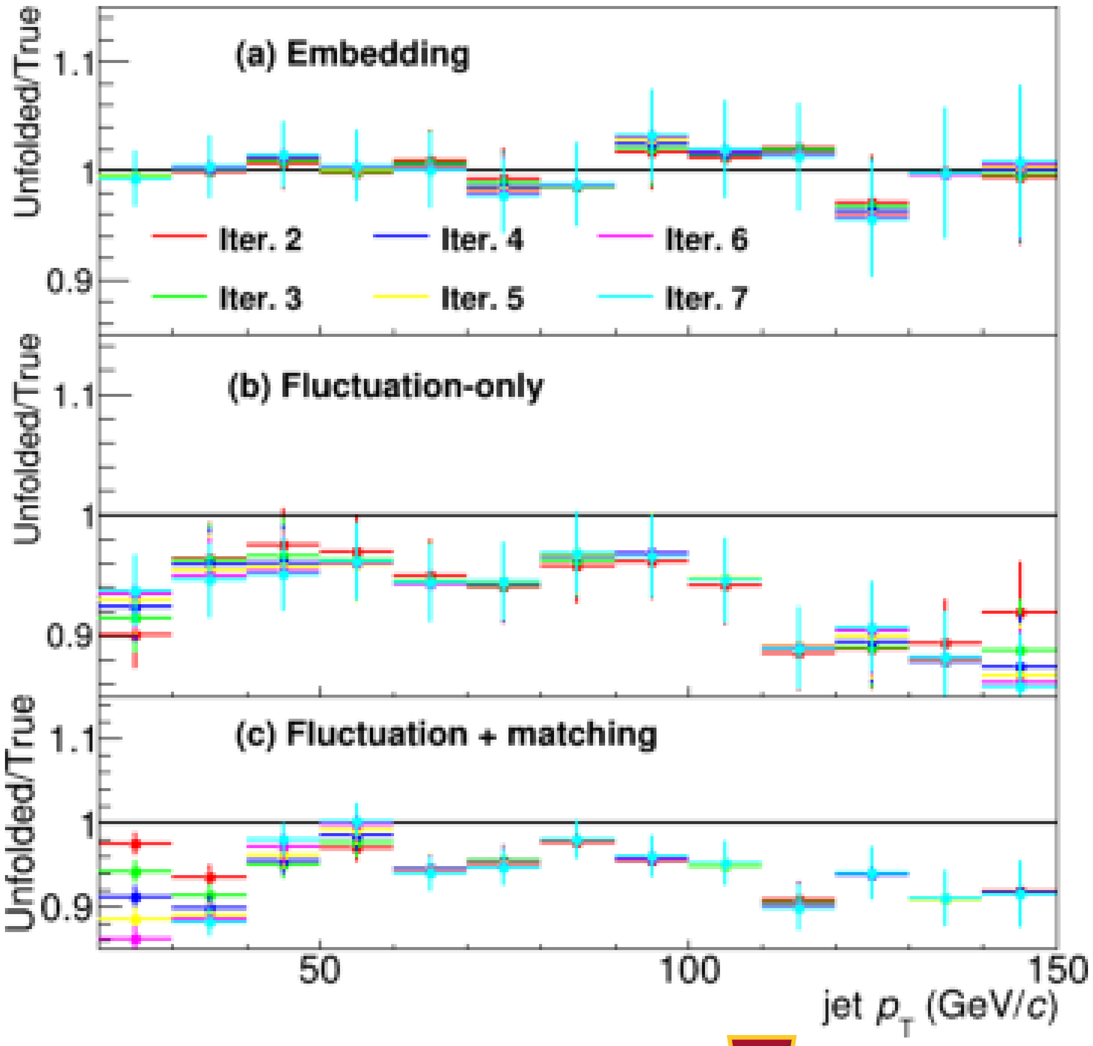
- What can we learn from this background characterization in simple model studies ?
  - Fluctuations widths dominated by Poissonian number fluctuations (expected as seen in data)
  - However, models such as Angantyr can differ by up to 13 % (no flow) and Tenngen up to 6 % (flow)
  - The fluctuations in models do indeed depend on the choice of thermal spectrum etc... – details seem to be ~ 10 % effect
  - **What can be done ?**

Hughes, da Silva, Nattrass
Phys. Rev. C 106, 044915

Background Fluctuations - Model Studies - Unfolding

Hughes, da Silva, Nattrass [Phys. Rev. C 106, 044915]

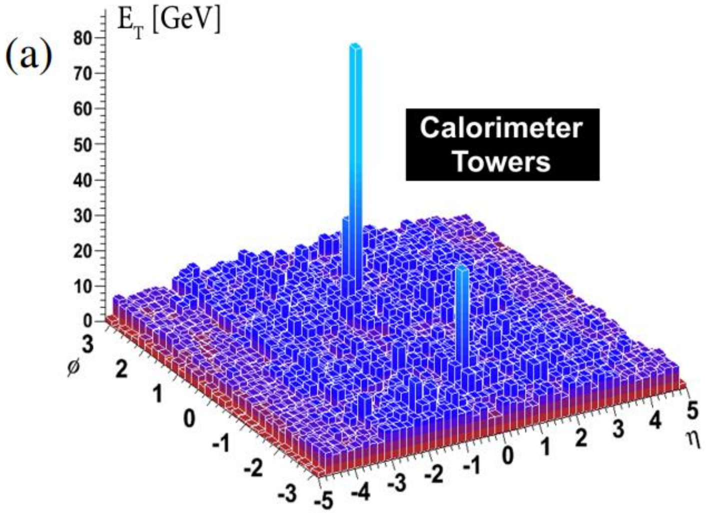
- Fluctuations in models are sensitive to details of model. What can be done ?
- Must unfold in a model just as done with data (because models have background !!!)
- Closure best when using an embedding technique (Pythia pp embedded in Pythia Angantyr PbPb)



Background Fluctuations - Subtraction

- Unfolding in a model to deal with background fluctuations
- What about subtracting the background in data ?
- Many techniques exist but one standard in the area-based subtraction method
  - $P_{T,\,jet}^{corr.} = p_{T,\,jet}^{raw} - \rho * A_{jet}$
- However, fluctuations in pT remain after
  - (std. Dev. ~ 20 GeV for R = 0.4)

Taken from arXiv:1702.07231



Background Fluctuations - Subtraction

- Fluctuations remain after area-based background subtraction technique
- But maybe we can devise a technique to correct for fluctuations ?

Background Fluctuations - Machine Learning

- Fluctuations remain after area-based background subtraction technique

- But maybe we can devise a technique to correct for fluctuations ?

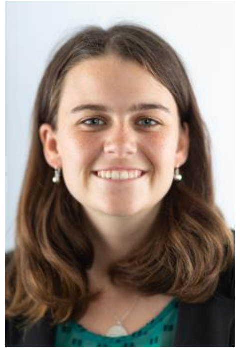
  - Yep – It's Machine Learning

  - Technique from Haake and Loizides:
  "Machine-learning-based jet momentum reconstruction in heavy-ion collisions"
  Phys. Rev. C **99**, 064904

  - Improves upon Area Based method (reduces fluctuations)

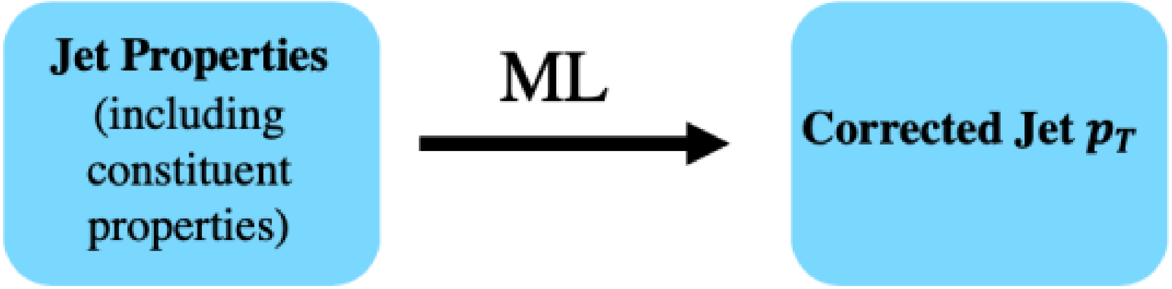
Background Fluctuations - Machine Learning

- Technique from Haake and Loizides: *"Machine-learning-based jet momentum reconstruction in heavy-ion collisions"* Phys. Rev. C **99**, 064904

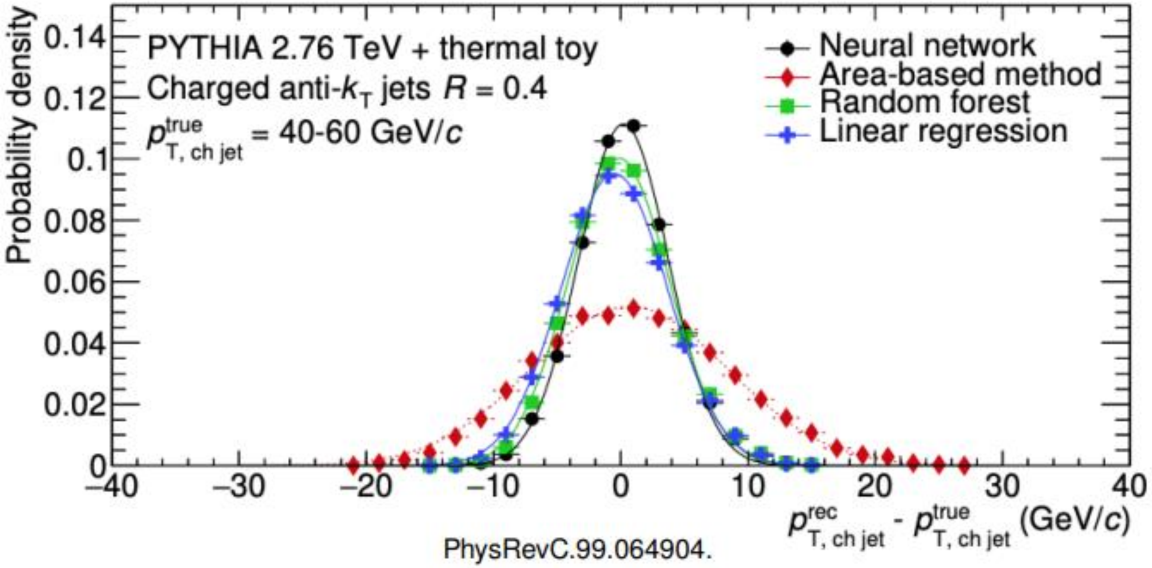


Hannah Bossi

- Further developed by Hannah Bossi and used in by her in ALICE Jet RAA measurements (arXiv:2208.14492)



Taken from H. Bossi

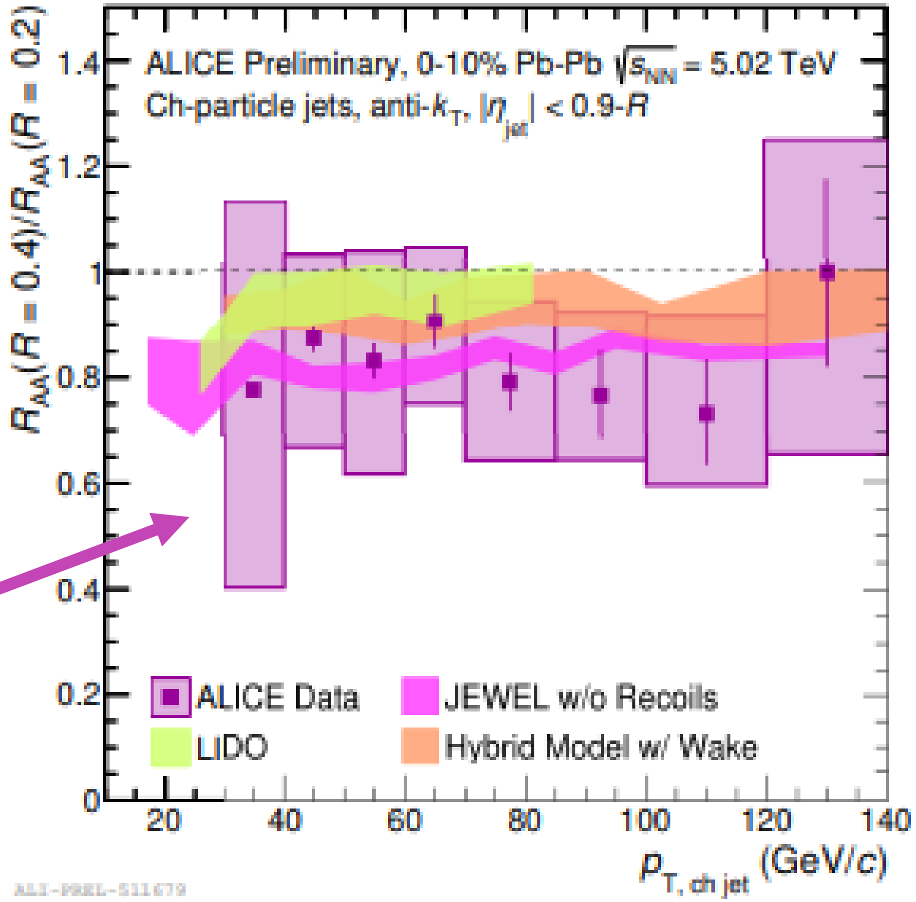


Background Fluctuations - Machine Learning

- Technique from Haake and Loizides: *"Machine-learning-based jet momentum reconstruction in heavy-ion collisions"* Phys. Rev. C **99**, 064904

- What does this technique buy you ?

- "*R-dependence of inclusive jet suppression and groomed jet splittings in heavy-ion collisions with ALICE*" arXiv:2208.14492v1
  - Unfolding still necessary
  - BUT
  - Reduced fluctuations
  - Lower momentum (down to p_T^{jet} = 30 GeV)



Taken from H. Bossi

Background Fluctuations - Machine Learning

- Following the analysis in *Phys. Rev. C 99, 064904*
- Some questions to ask:
    - Can we improve on these results ?
    - Can we interpret the ML methods to improve p_T resolution ?

"Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"

5th ML4Jets Workshop – 11/01/22 - 11/04/22

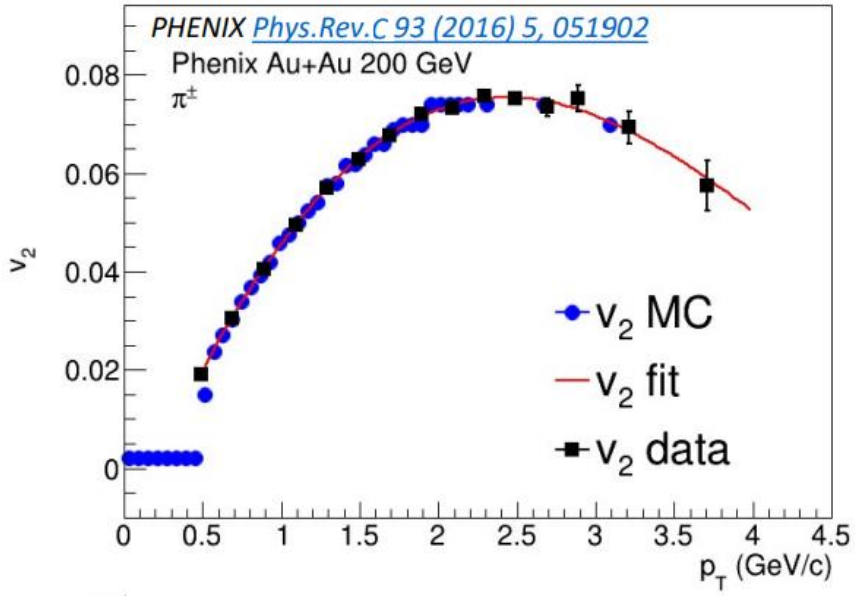
Tanner Mengel

Tanner Mengel

Background Fluctuations - Machine Learning

- Can we improve on results in *Phys. Rev. C 99, 064904* ?
- Add a more complex generator that includes flow (TennGen – fit to PHENIX and STAR data) for background
- Add in PYTHIA 8 for signal
  - Use p_T hard bins (1M events / bin)
- PYTHIA 8 is truth, we want to predict $p_{T,\,jet}^{pythia}$
- TennGen fit to RHIC data is background

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]()



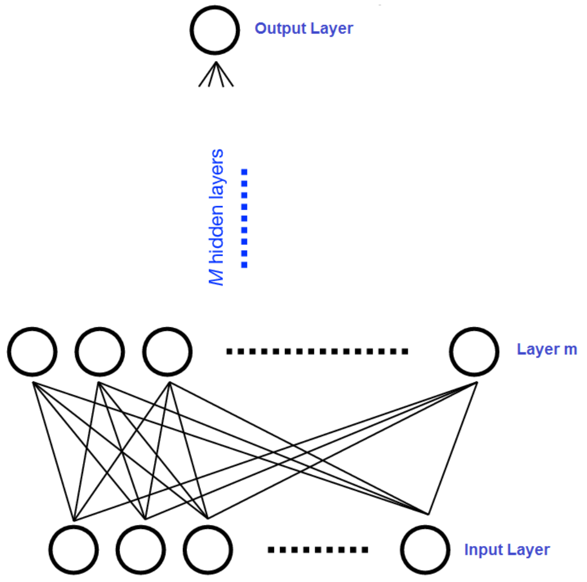
Taken from T. Mengel

Background Fluctuations - Machine Learning

- Can we improve on results in *Phys. Rev. C 99, 064904 ?* - 2 methods

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

Deep Neural Net (DNN)
Architecture: [N, 100, 100, 50, 1]
Features: [11]



Taken from [ResearchGate]

Physics Inspired (Multiplicity)

$P_T^{corr.} = p_T^{uncorr.} - \text{rho}(N^{jet}_{constit.} - \langle N^{jet}_{pythia\ constit.} \rangle)$

Inspired by:

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Assuming no flow

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + (N_A + \sigma^2_{NP}(N_A)) \cdot \langle p_t \rangle^2}.$$

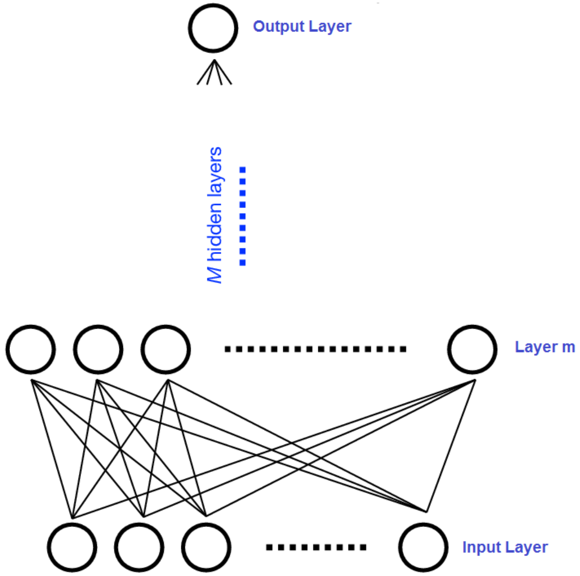
Accounts for v2/v3

Background Fluctuations - Machine Learning

- Can we improve on results in *Phys. Rev. C 99, 064904 ?* - 2 methods

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

Deep Neural Net (DNN)
Architecture: [N, 100, 100, 50, 1]
Features: [11]



Taken from ResearchGate

Physics Inspired (Multiplicity)
$P_T^{corr.} = p_T^{uncorr.}$ - rho($N^{jet}_{constit.}$ - <N^{jet}_{pythia}>)

Inspired by:

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + N_A \cdot \langle p_t \rangle^2}.$$

Assuming no flow

$$\sigma(\delta p_t) = \sqrt{N_A \cdot \sigma^2(p_t) + \left(N_A + \sigma^2_{NP}(N_A)\right) \cdot \langle p_t \rangle^2}.$$

Accounts for v2/v3

Background Fluctuations - Machine Learning

Mengel
"Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"

- Can we improve on results in *Phys. Rev. C 99, 064904* ?
  - 2 methods

Deep Neural Net (DNN)
Architecture: [N, 100, 100, 50, 1]
Features: [1]

Physics Inspired (Multiplicity)
$P_T^{corr.} = p_T^{uncorr.} - rho(N^{jet}_{constit.} - \langle N^{jet}_{pythia} \rangle)$

- DNN is powerful but black box

- Want to understand it better ?

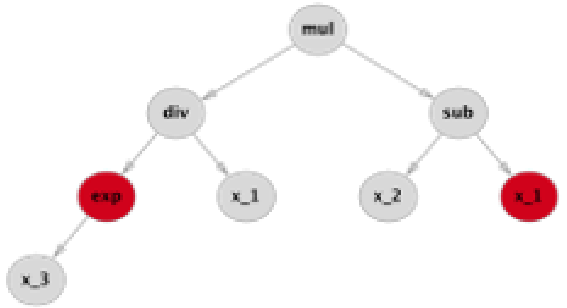
- Symbolic regression

Background Fluctuations - Machine Learning

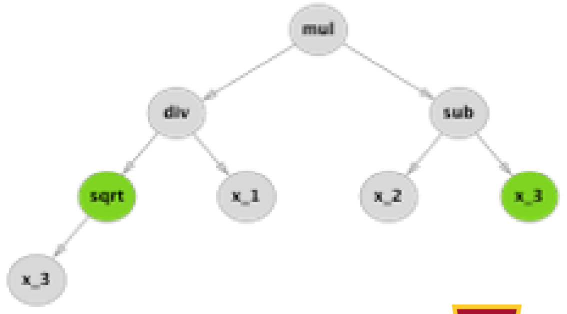
Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

- Symbolic Regression
  - Genetic programming where "traits" = operators
  - Each iteration creates new population with traits from each parent.
  - Highest performing offspring selected
- Plan
  - Train DNN for jet p_T regression
  - Fit input space to DNN prediction using Symbolic Regression implementation in PySR

Parent



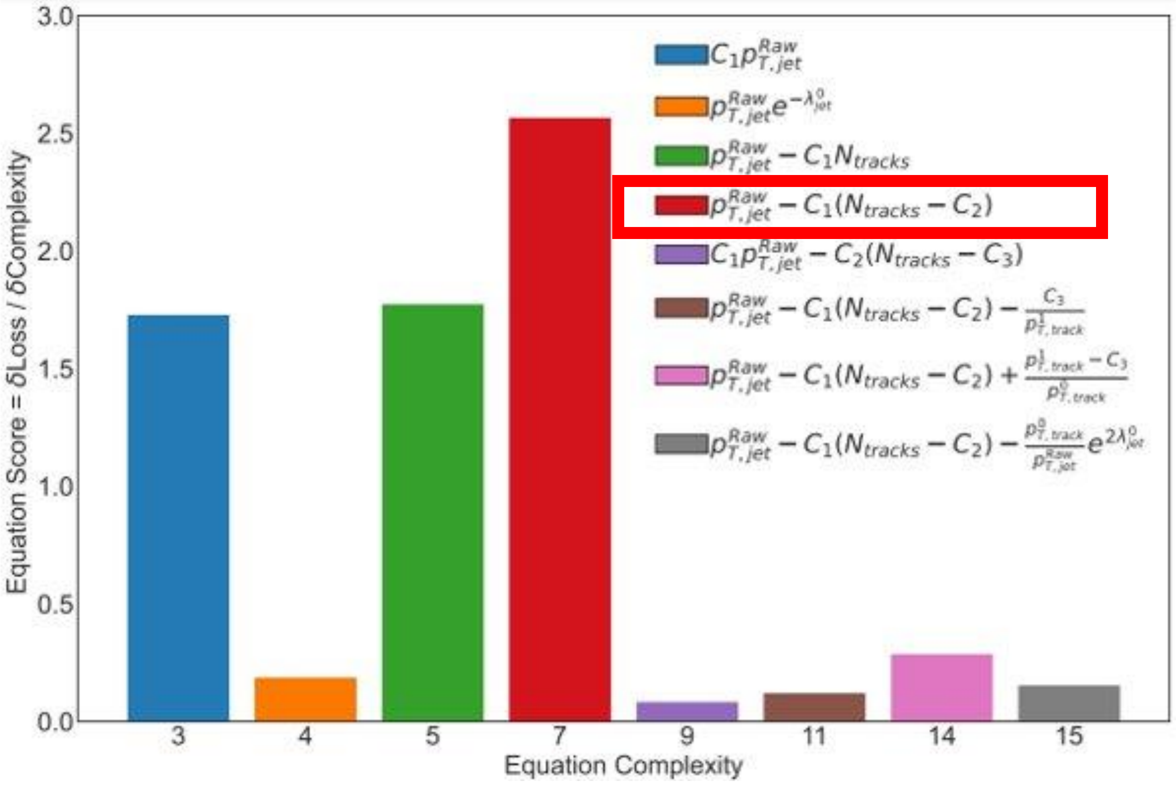
Offspring



Background Fluctuations - Machine Learning

- Highest Scoring looks like multiplicity method !!!

Mengel
"Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"

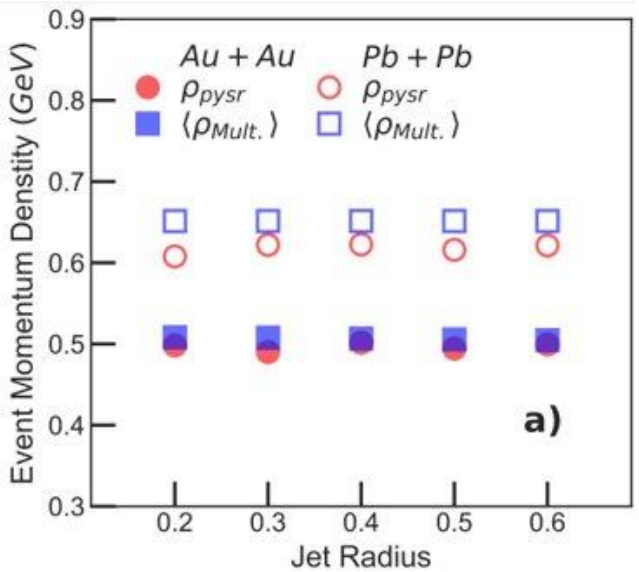


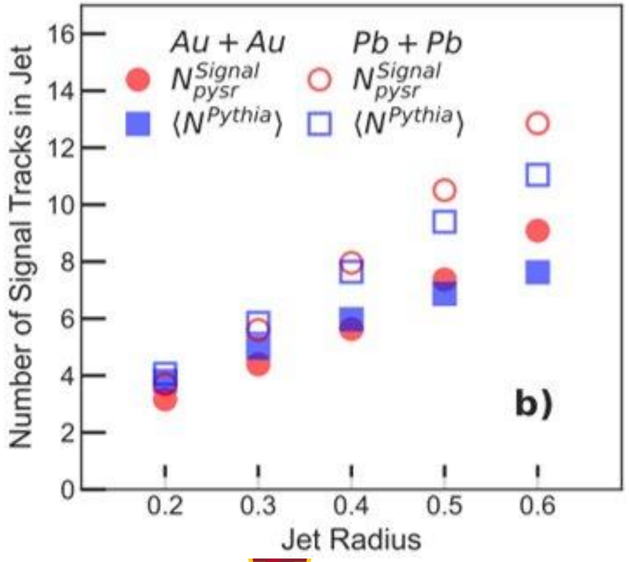
Physics Inspired (Multiplicity)

$P_T^{corr.} = p_T^{uncorr.} - rho(N^{jet}_{constit.} - \langle N^{jet}_{pythia\ constit.} \rangle)$

$C_1 \rightarrow \rho$

$C_2 \rightarrow \langle N_{pythia} \rangle$

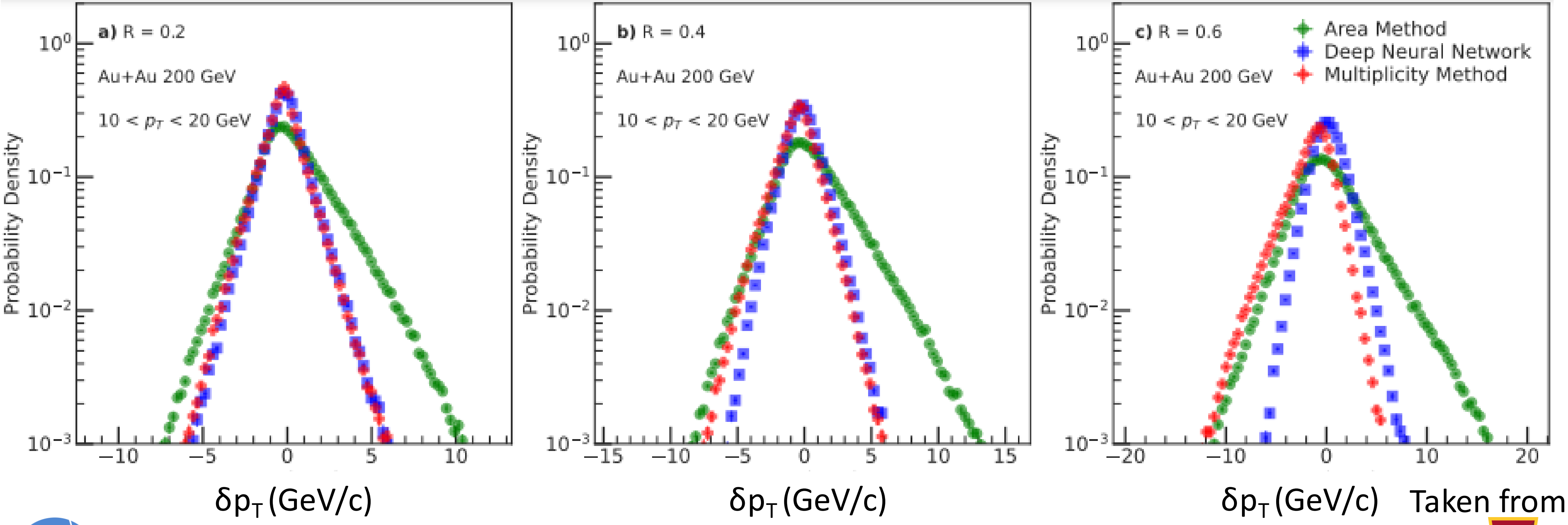




Background Fluctuations - Machine Learning - Results

- Compare performance across methods
- Measure $\delta p_T = p_T^{predicted} - p_T^{pythia}$
- Extract variance (next slide)

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]

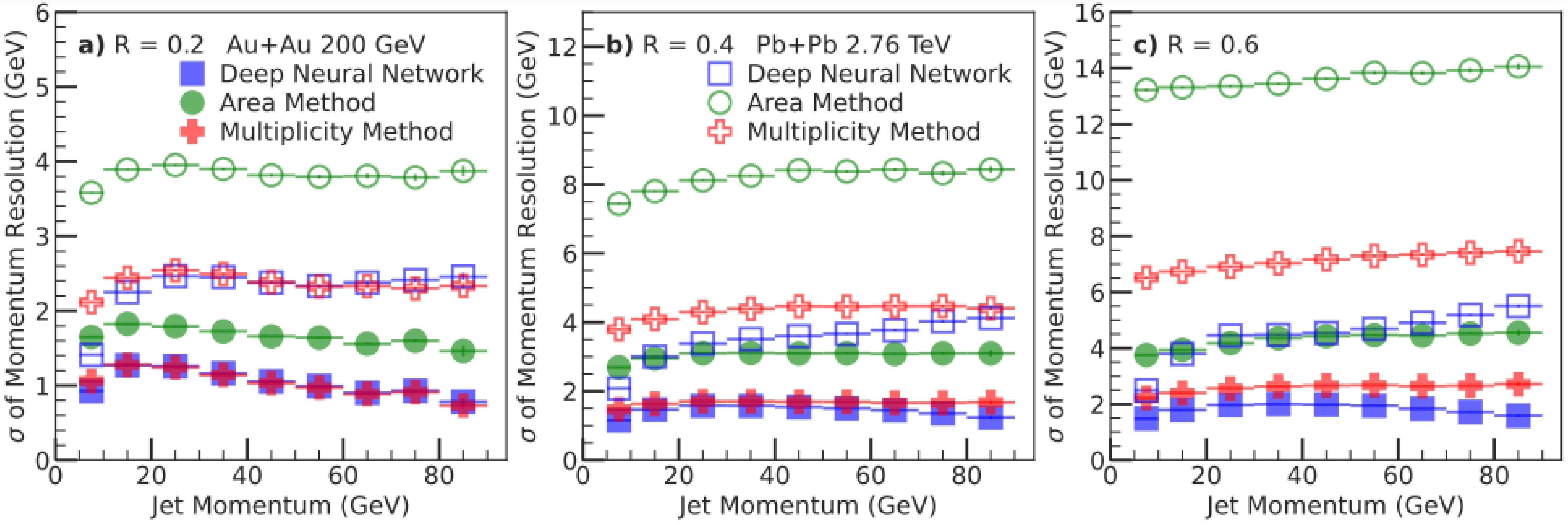


Taken from T. Mengel

Background Fluctuations - Machine Learning

- Extract variance of δpT distribution

Mengel
["Using Machine Learning to Improve our Understanding of the Jet Background in A+A Collisions"]



Taken from [T. Mengel]

Background Fluctuations - Machine Learning

Area based method performs the worst (highest width)

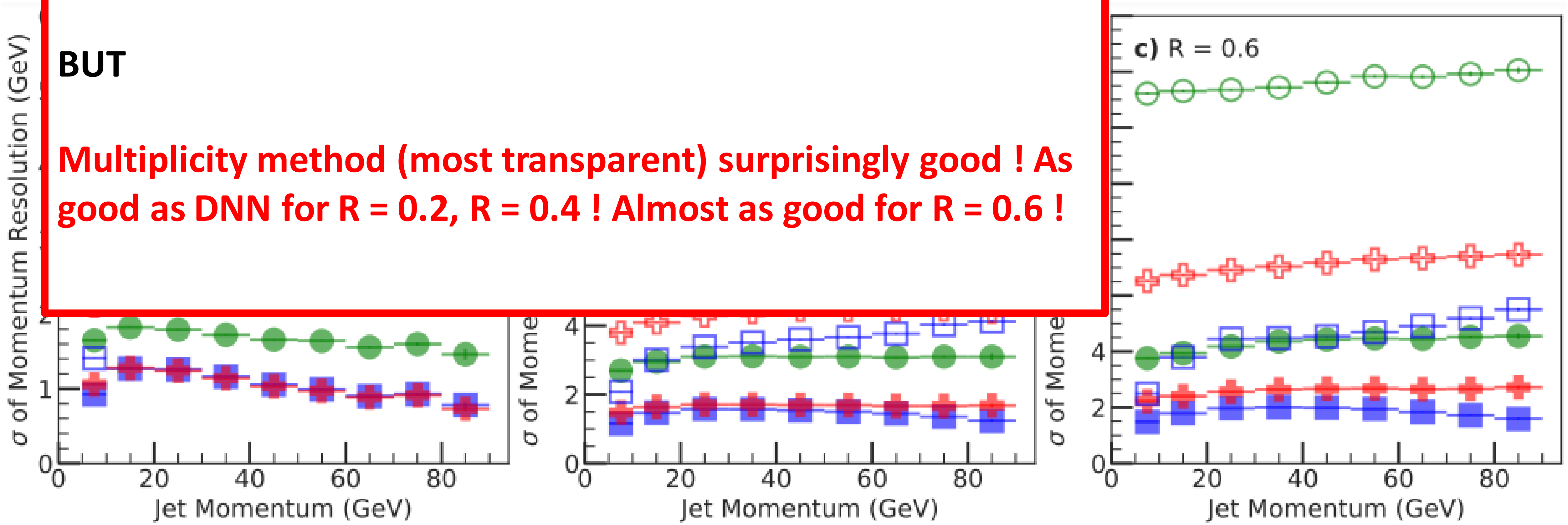
DNN performs best in almost all cases

BUT

Multiplicity method (most transparent) surprisingly good ! As good as DNN for R = 0.2, R = 0.4 ! Almost as good for R = 0.6 !

engel

Using Machine Learning to Improve our nderstanding of the Jet Background in +A Collisions"



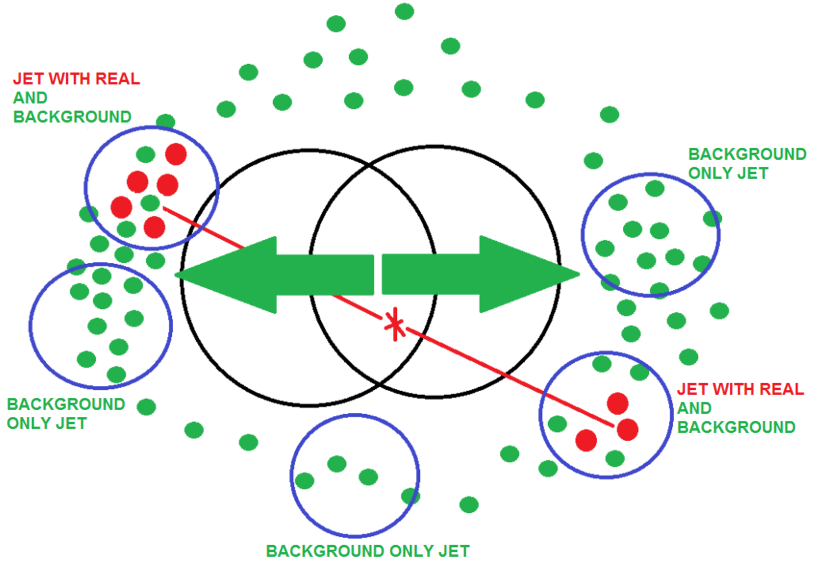
Taken from T. Mengel

Background Fluctuations - Mitigation

- We see that background subtraction can be heavily informed by machine learning methods where model studies play a crucial role

- We learned that these methods can often point to a simpler/more transparent background subtraction method

- What about the problem of combinatorial jets? Can we use model studies to look at mitigating those?



Background Fluctuations - Mitigation

- What about the problem of combinatorial jets? Can we use model studies to look at mitigating those?
- Look at the effect of cuts on removing combinatorial jets.
- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

["Separating signal from combinatorial jets in a high background environment"]

arXiv:2301.09148v2 – (also submitted to PRC)

Patrick Steffanic et. al.

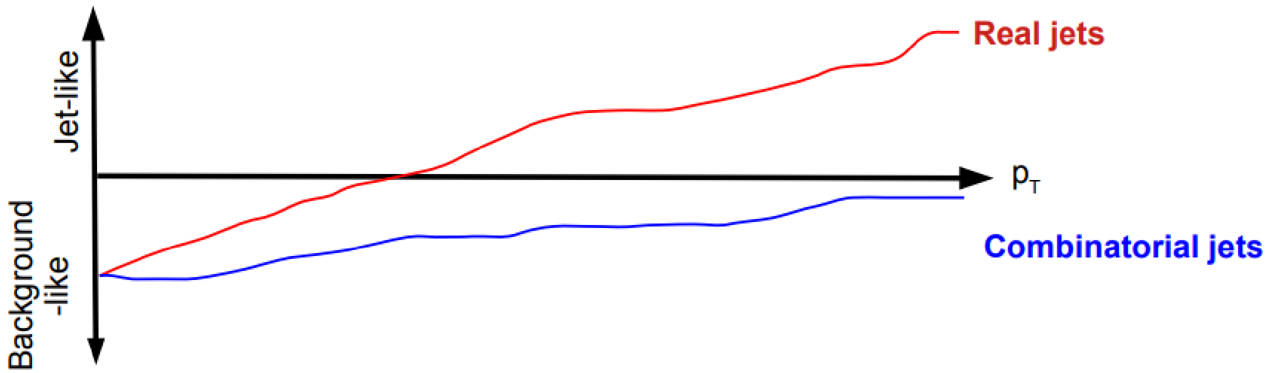
Patrick Steffanic

Background Fluctuations - Mitigation

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

Steffanic et. al.
arXiv:2301.09148v2

- Machine Learning (Random Forests)
- Pythia 6 signal
- TennGen background



Taken from C. Nattrass

Background Fluctuations - Mitigation

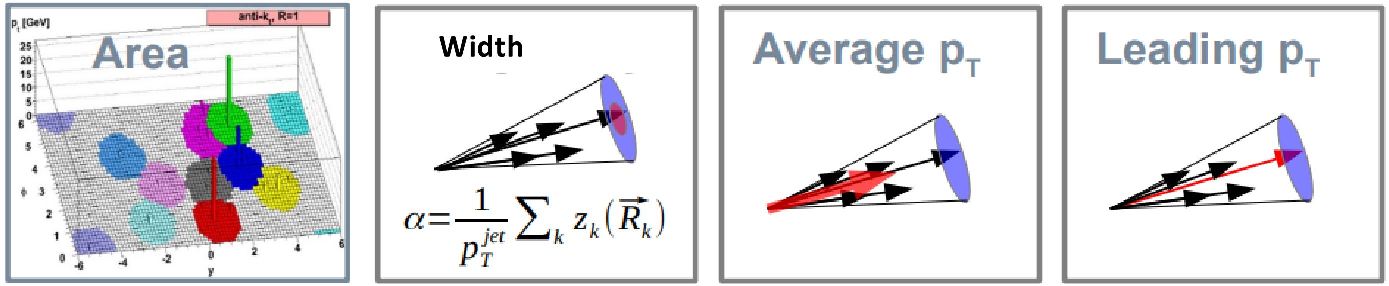
- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

Steffanic et. al.
arXiv:2301.09148v2

- Combinatorial jets: $p_T^{pythia} < 2\pi R^2$ GeV

- Signal jets: $p_T^{pythia} > 0.8 * p_T^{hard\ min.}$ GeV

- Observables:
  - Area: $N_g \langle A_g \rangle$
  - Jet Width: $\Sigma z_i(\Delta R_{i,\ jet})/p_T^{jet}$
  - Leading hadron p_T
  - Mean constituent pT: $\langle p_{T,\ constit.} \rangle$



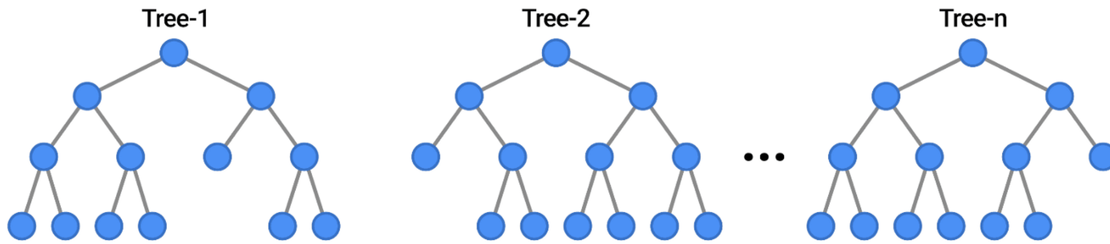
Taken from C. Nattrass

Background Fluctuations - Mitigation

Steffanic et. al.
arXiv:2301.09148v2

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?
- Random forest Ensemble Oracle Method
  - Train random forest
  - Apply single decision tree to predictions of random forest
  - Extract top level node of decision tree
    - This is the cut !

EXAMPLES



Taken from Tensorflow Blog

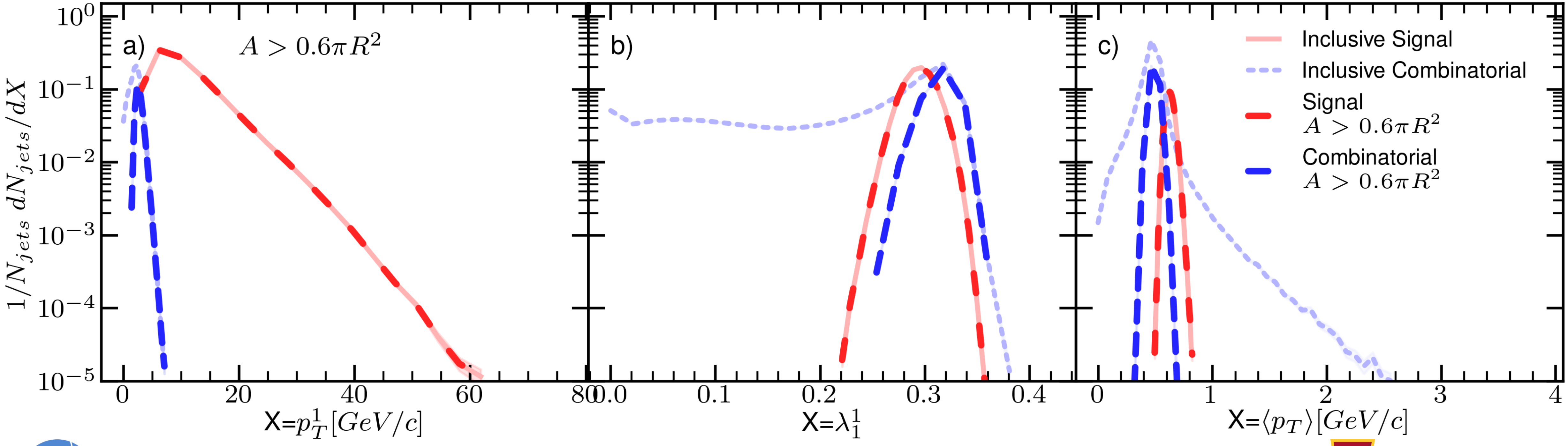
Parameter name	This study	Default
n_estimators	200	100
max_depth	3	None
min_samples_leaf	100	1
min_weight_fraction_leaf	0.1	0.0
max_samples	0.9	1.0
random_state	42	None

Background Fluctuations - Mitigation

Steffanic et. al.
arXiv:2301.09148v2

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- Baseline "canonical cut" – $A > 0.6\pi R^2$

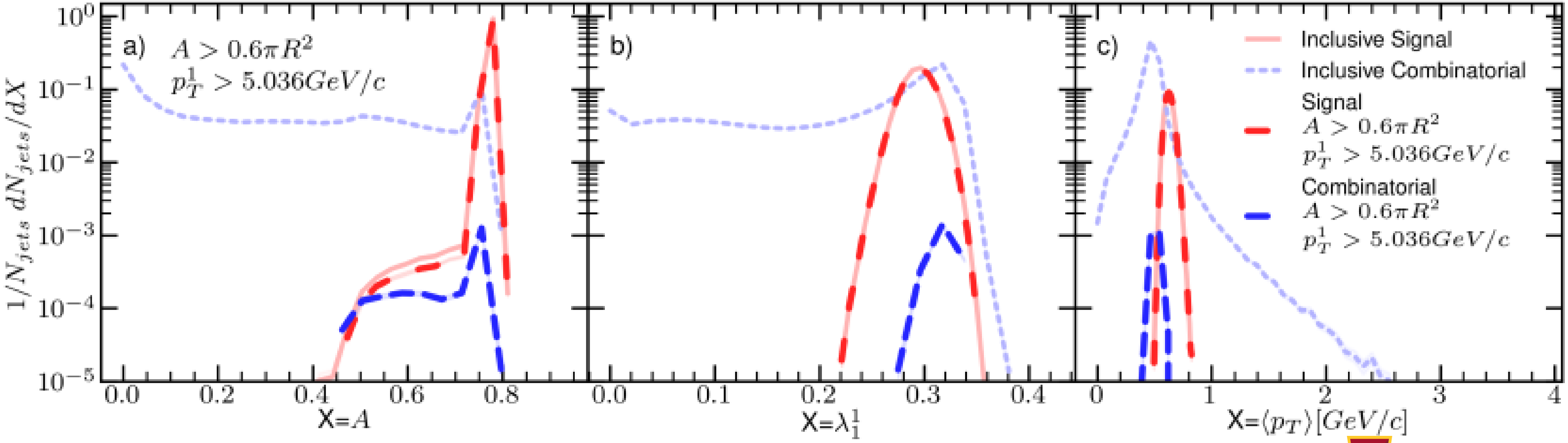


Background Fluctuations - Mitigation

Steffanic et. al.
arXiv:2301.09148v2

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- From Random forest – $p_T^{\text{lead. trk.}} > 5.036$ GeV/c (+ $A > 0.6\pi R^2$)



Background Fluctuations - Mitigation

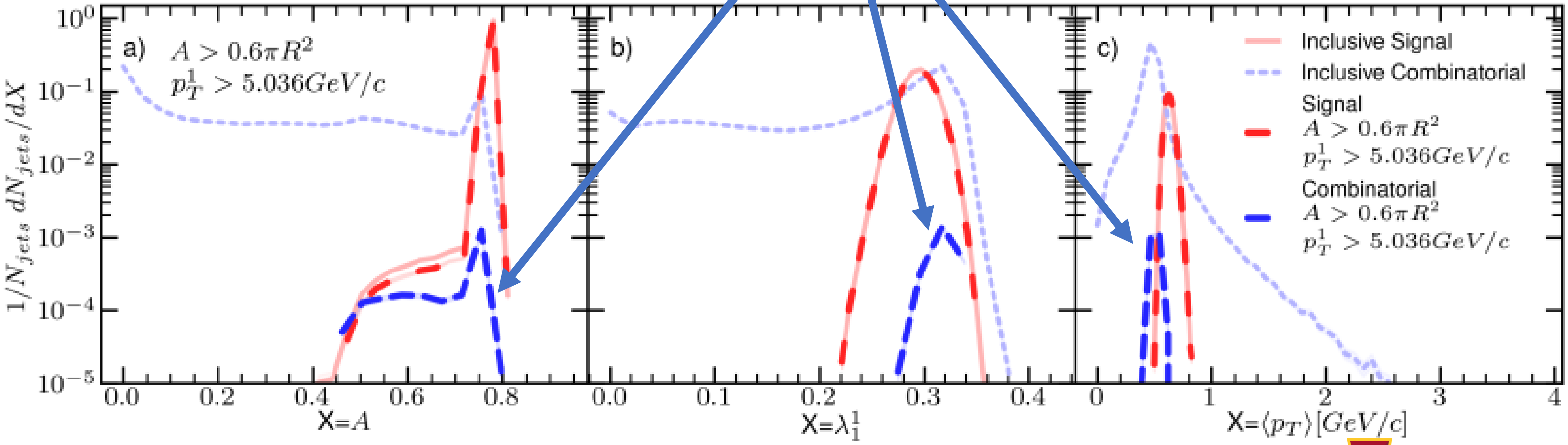
Steffanic et. al.

v2

These cuts work well but ALWAYS leave a population of combinatorial jets that look like signal jets. The addition of the leading hadron p_T cut removes a lot of combinatorial jets compared to area cut alone.

- Can we come up with a set of cuts to remove combinatorial jets from signal (hard scattering origin) jets ?

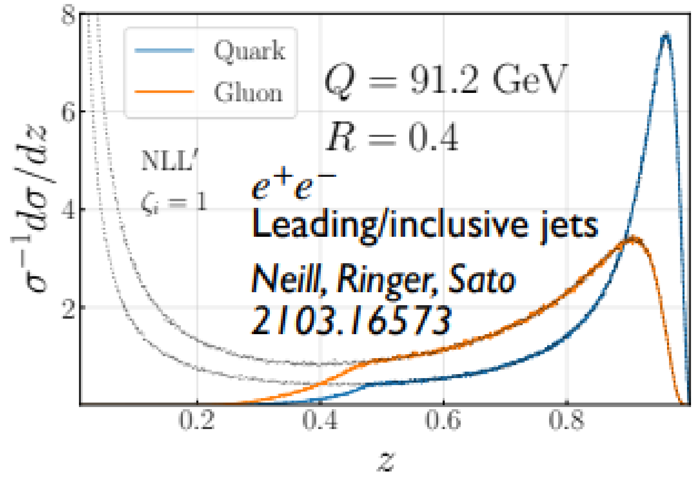
- From Random forest – $p_T^{\text{lead. trk.}} > 5.036$ GeV/c



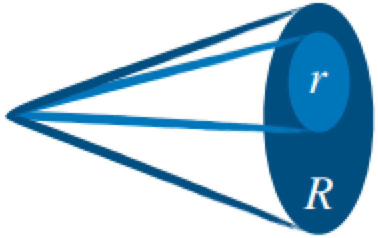
Background Fluctuations - Mitigation

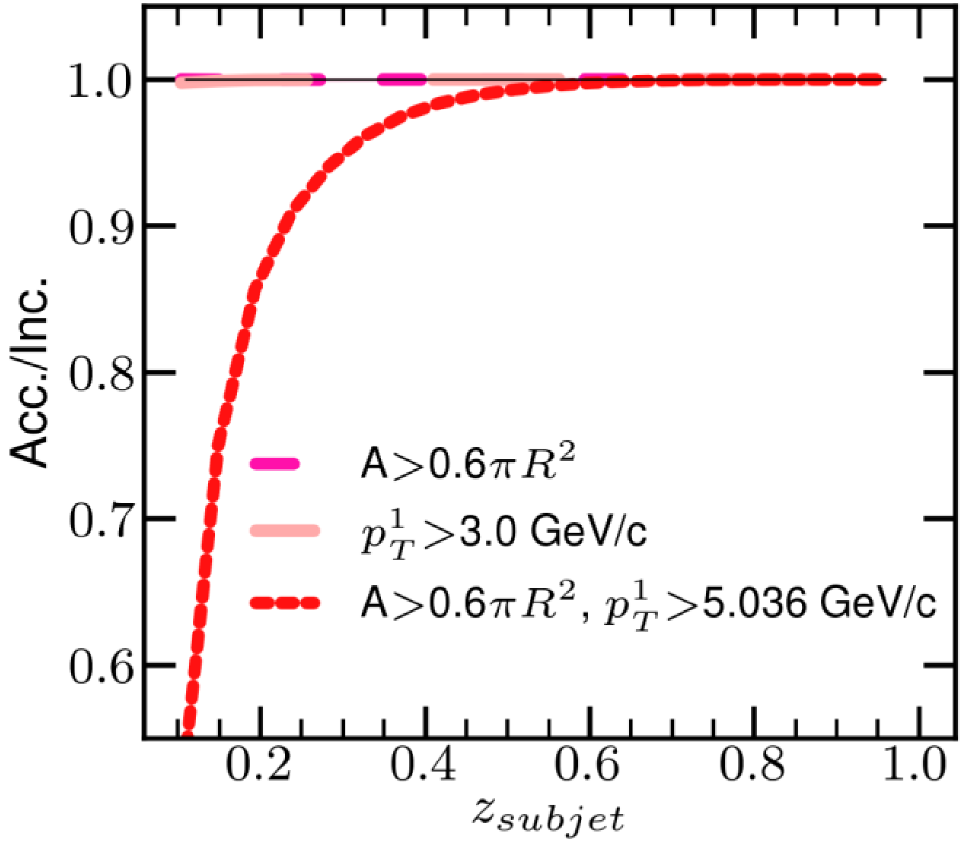
Steffanic et. al.
arXiv:2301.09148v2

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?
- From Random forest – $p_T^{lead.\ trk.} > 5.036$ GeV/c
- This cut does induce a bias towards quark-like jets



$$z_{subjet} = \frac{p_T^{lead.\ subjet}}{p_T^{jet}}$$





Background Fluctuations - Conclusions/Takeaways

- The details of the background matter – *even in a model*

Background Fluctuations - Conclusions/Takeaways

- The details of the background matter – *even in a model*

- The best way to detail with background in models is to treat it *exactly like the data* (e.g. unfold the data, unfold the MC)

Background Fluctuations - Conclusions/Takeaways

- The details of the background matter – *even in a model*

- The best way to detail with background in models is to treat it *exactly like the data* (e.g. unfold the data, unfold the MC)

- Machine learning background subtraction – powerful but obscure

Background Fluctuations - Conclusions/Takeaways

- The details of the background matter – *even in a model*
- The best way to detail with background in models is to treat it *exactly like the data* (e.g. unfold the data, unfold the MC)
- Machine learning background subtraction – powerful but obscure
- Multiplicity method (and similar methods) more transparent and nearly as effective in our studies (and can be informed by ML)

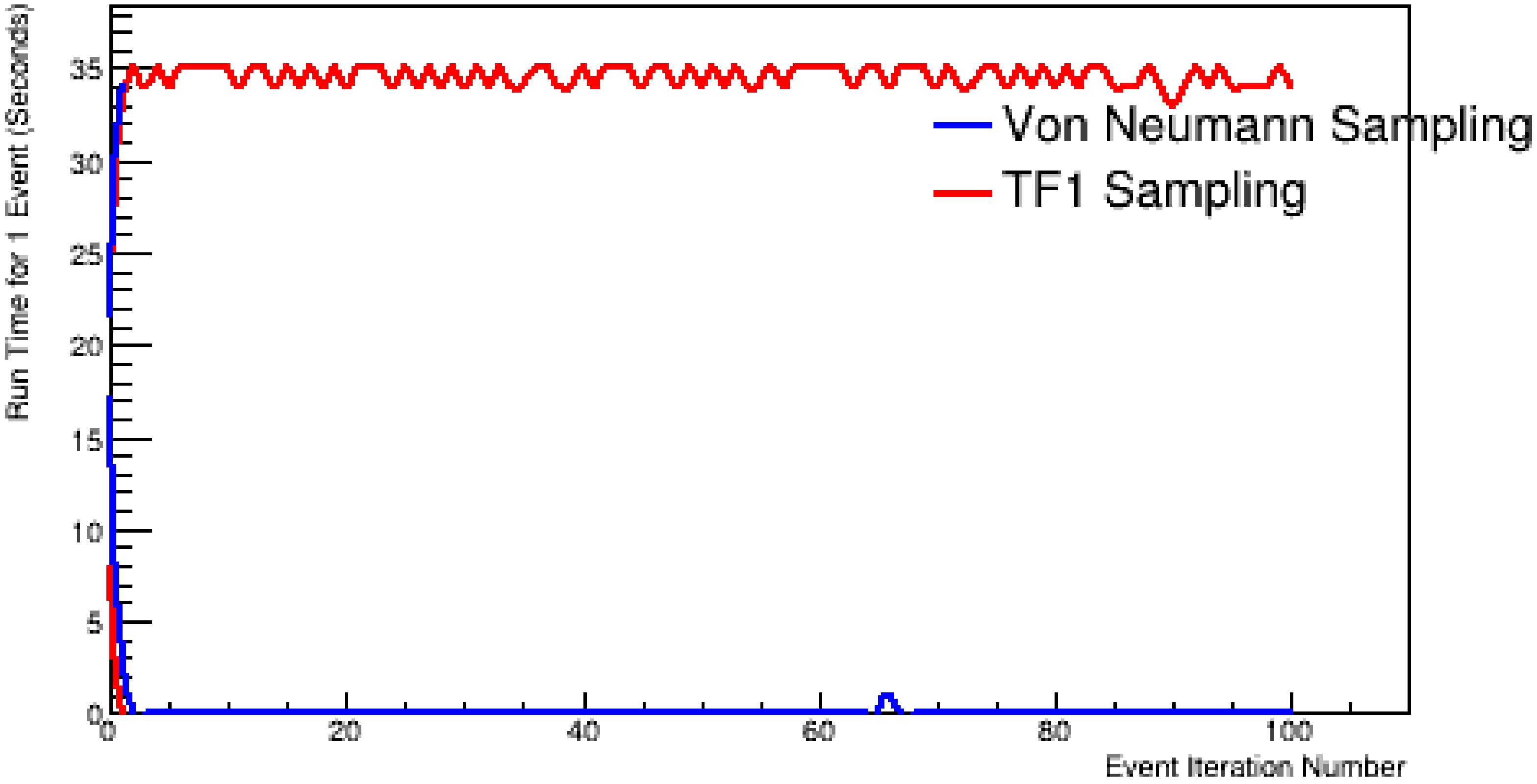
Background Fluctuations - Conclusions/Takeaways

- The details of the background matter – *even in a model*
- The best way to detail with background in models is to treat it *exactly like the data* (e.g. unfold the data, unfold the MC)
- Machine learning background subtraction – powerful but obscure
- Multiplicity method (and similar methods) more transparent and nearly as effective in our studies (and can be informed by ML)
- Combinatorial jets are a tough problem – no silver bullet
    - Cuts always leave some combinatorial jets that look like signal jets – may bias jets

Backup

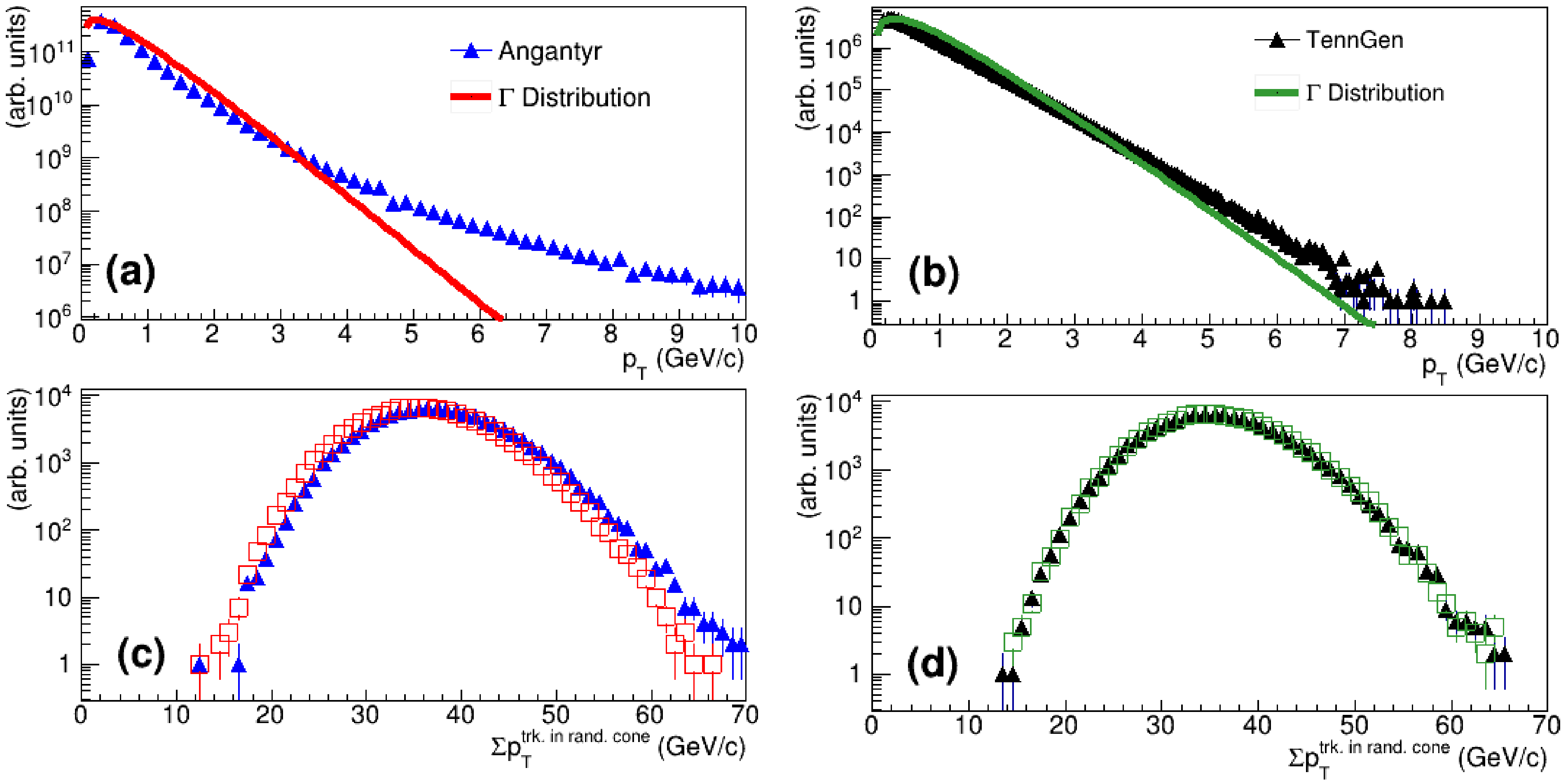
Backup

- TennGen v1 (PbPb only)
- TennGen v2 (AuAu)



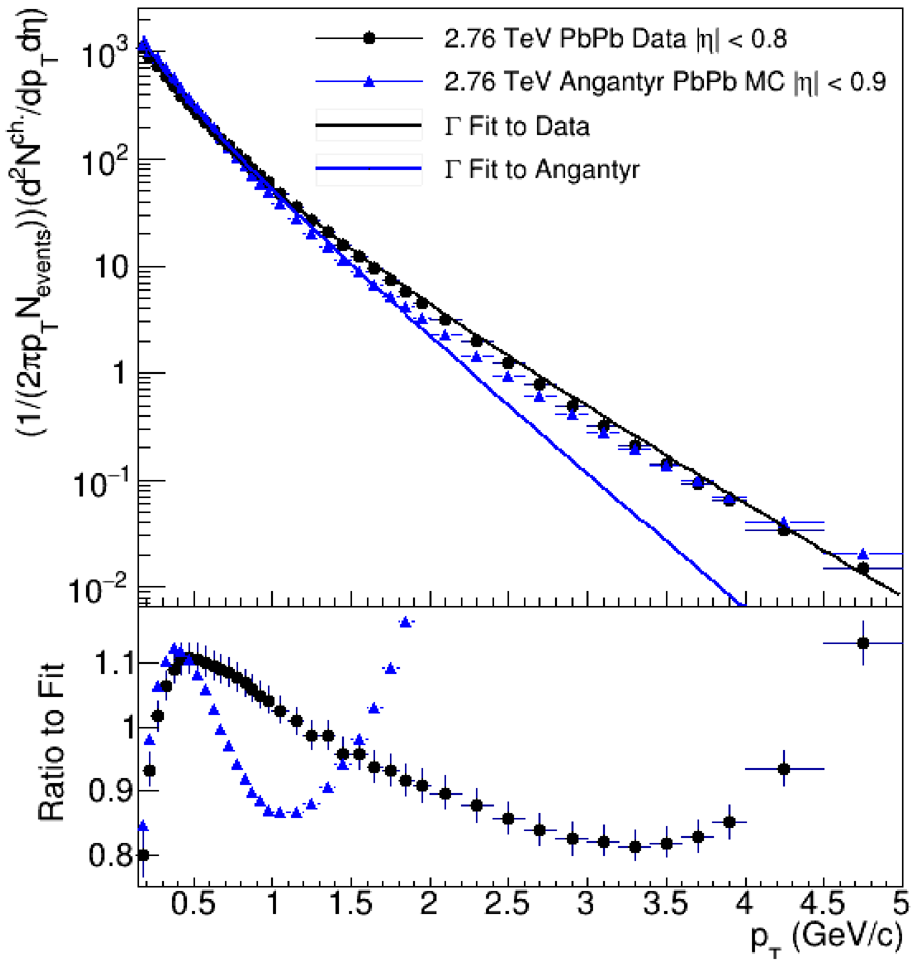
Backup

- Spectra shapes compared to Gamma Distribution



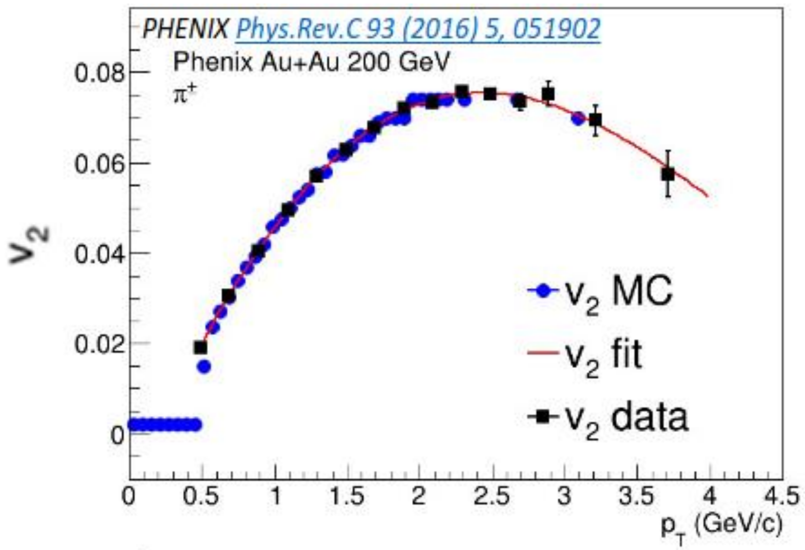
Backup

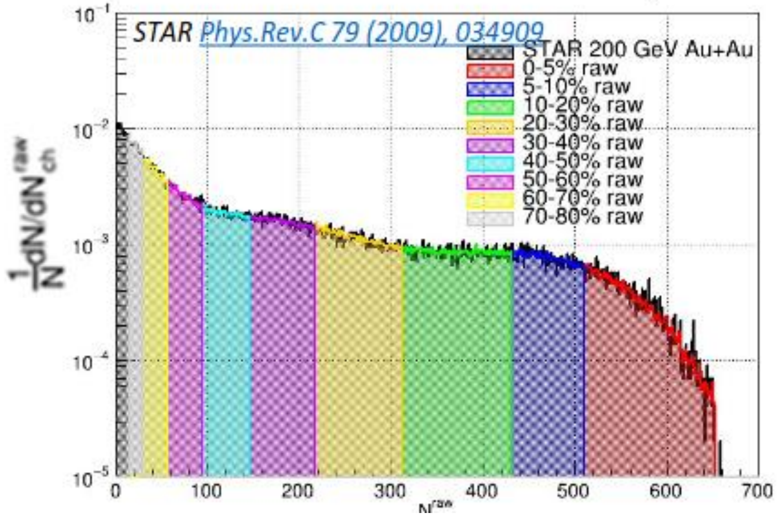
- Spectra shape compared to Gamma Distribution (Angantyr Only)



Backup

- PYTHIA8 (Signal):
  - 25 Million (1 million per p_T hard bin) p+p events at 200 GeV, Tune 14
- TennGen (Background): *C. Hughes et al* Phys. Rev. C 106 (2022), 044915
  - **Multiplicity**: Sampled from corrected N_{ch} distribution *STAR* Phys.Rev.C 79 (2009), 034909
  - $\boldsymbol{p_T}$: Identified particle p_T spectrum fit with Boltzmann-Gibbs Blast wave *PHENIX* Phys.Rev.C 88 (2013) 2, 024906
  - $\boldsymbol{\phi}$: Identified particle flow harmonics (v_2, v_3, v_4) *PHENIX* Phys.Rev.C 93 (2016) 5, 051902
  - $\boldsymbol{\eta}$: Uniform distribution |η| < 1.1
- Merge PYTHIA8 charged particles with TennGen Au+Au 200 GeV background
- Find anti-k_T jets
  - Only save jets with $p_T^{PYTHIA} > 5.0$ GeV
  - ~30 Million jets per dataset
- Take p_T^{PYTHIA} to be truth value
  - Train-Test split: 20/80%

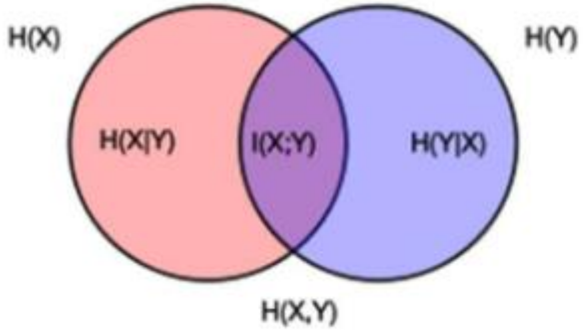


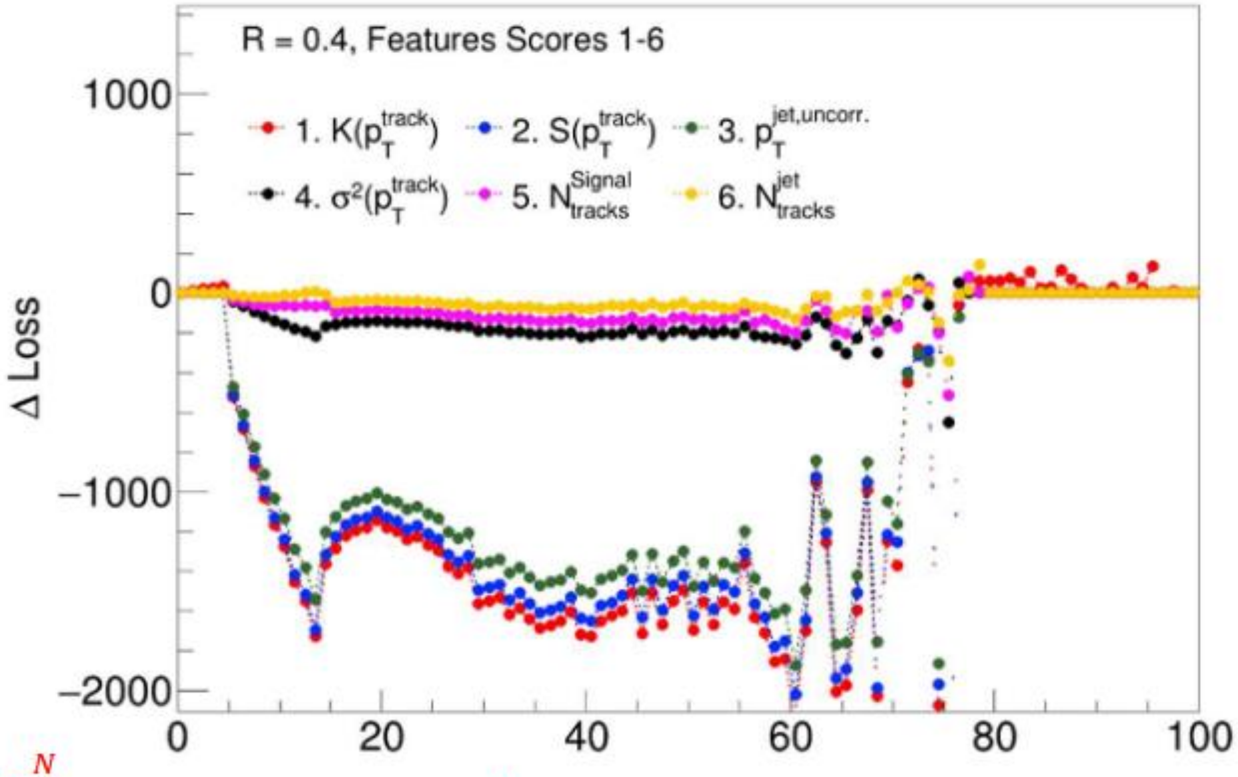


Backup

Feature space optimization

- Mutual Information: I(X;Y) = H(X,Y) – H(X|Y) –H(Y|X)
- Permutation Scoring: Randomly permutes feature to see change in Cost (mean squared error) evaluation





$$MSE = \frac{1}{N}\sum_{i}^{N}\left(p_{T,i}^{PYTHIA} - p_{T,i}^{Predicted}\right)^2$$

Backup

Genetic programming where ‘traits’ = operators.

Each iteration creates new population with traits from each parent. Highest performing offspring are selected.

Proof of concept (Area correction):

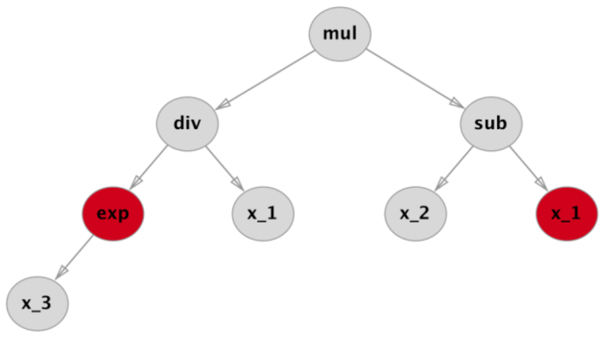
Target = area_based_prediction

Inputs = jet_pt_raw , jet_area, median_pt_over_area

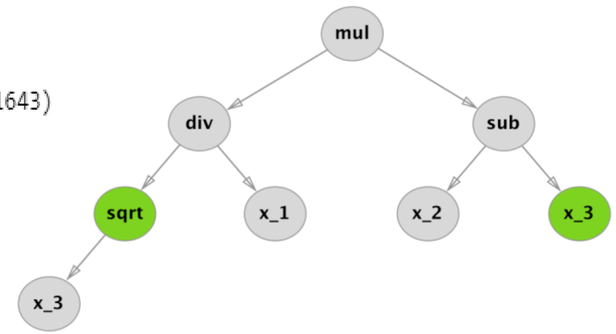
Complexity	Loss	Score	Equation
1	1.708e+02	4.292e-01	jet_pt_raw
3	5.111e+01	6.032e-01	(jet_pt_raw * 0.7568252)
4	8.534e+00	1.790e+00	(jet_pt_raw - (jet_area * median_pt_over_area))
5	8.412e+00	1.440e-02	(jet_pt_raw + (median_pt_over_area * -0.4965229))
9	8.237e+00	5.256e-03	(((jet_area - median_pt_over_area) * 0.47800952) - (jet_pt_raw * -0.9821865))
15	8.115e+00	2.473e-03	((((jet_pt_raw - ((median_pt_over_area * 0.4997403) + 1.1289326)) - -1.282973) * 0.96711856) - -0.95751643)

```
PySRRegressor.equations_ = [
        pick     score                                          equation  \
    0         0.000000                                        jet_pt_raw
    1         0.603197                          (jet_pt_raw * 0.7568252)
    2  >>>>   1.789895   (jet_pt_raw - (jet_area * median_pt_over_area))
```

Parent

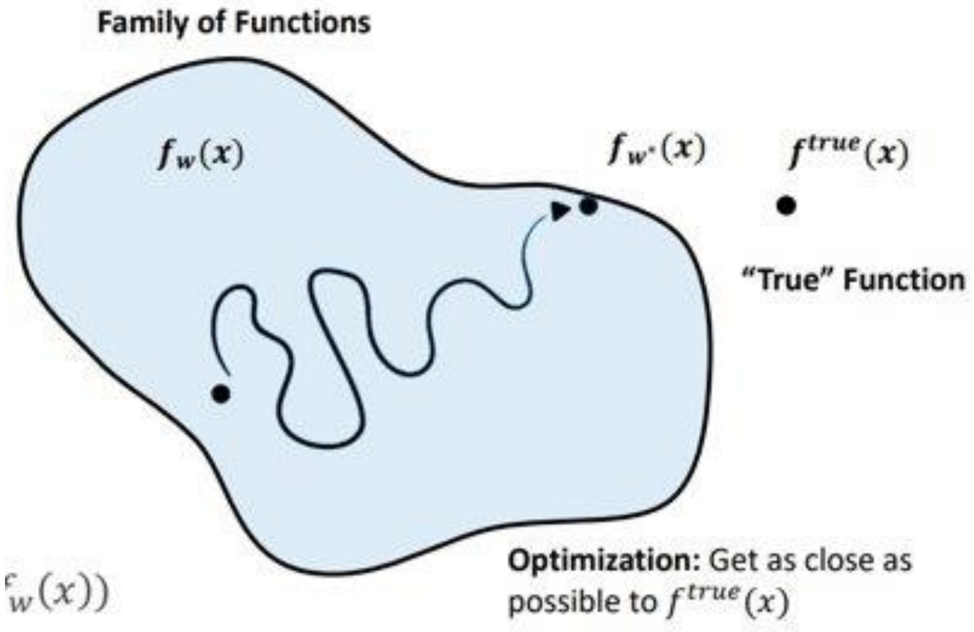


Offspring



Backup

- *SR models are good if defined arithmetic expression exists that maps inputs to outputs f(x) = y*
- *DNN models are good at factorizing and approximating multivariate mappings: f(x|theta) = yhat ~ y*
- *Plan*
  1) *Train DNN on jet pT regression*
  2) *Fit input space to DNN prediction using PySR*



Backup

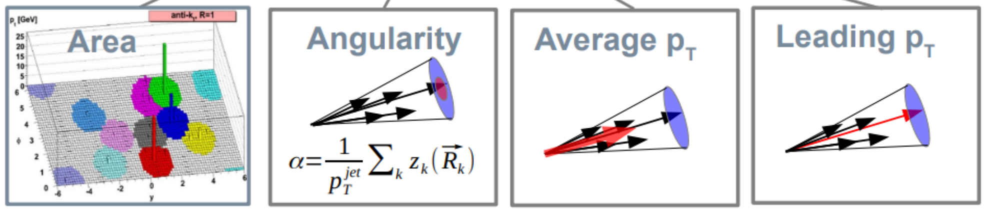
- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- Silhouette Measure:

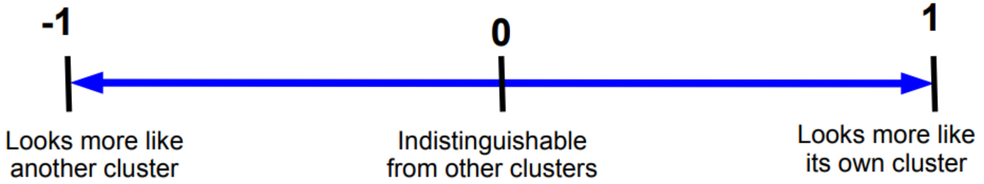
  - $a_i = \langle d_{i,j} \rangle_{j \neq I}$
    - (avg. distance b/w jet candidate and others in **its own cluster**)
  - $b_i = \langle d_{i,j} \rangle$
    - (avg. Distance b/w jet candidate and others **in other clusters**)
  - $s_i = (b_i - a_i)/(\max[b_i,a_i])$

Steffanic
"Separating Signal from Combinatorial Jets in a High Background Environment" (arXiv entry)

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



- Silhouette value $s_i=\frac{b_i-a_i}{max[b_i,a_i]}$



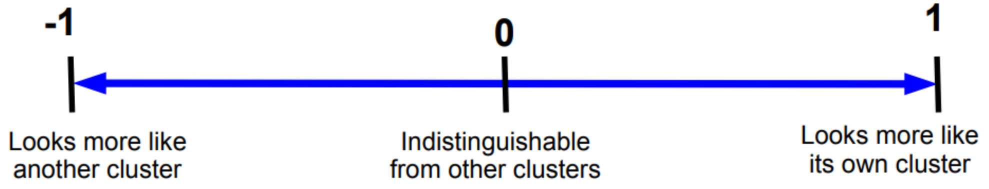
Taken from C. Nattrass

Backup

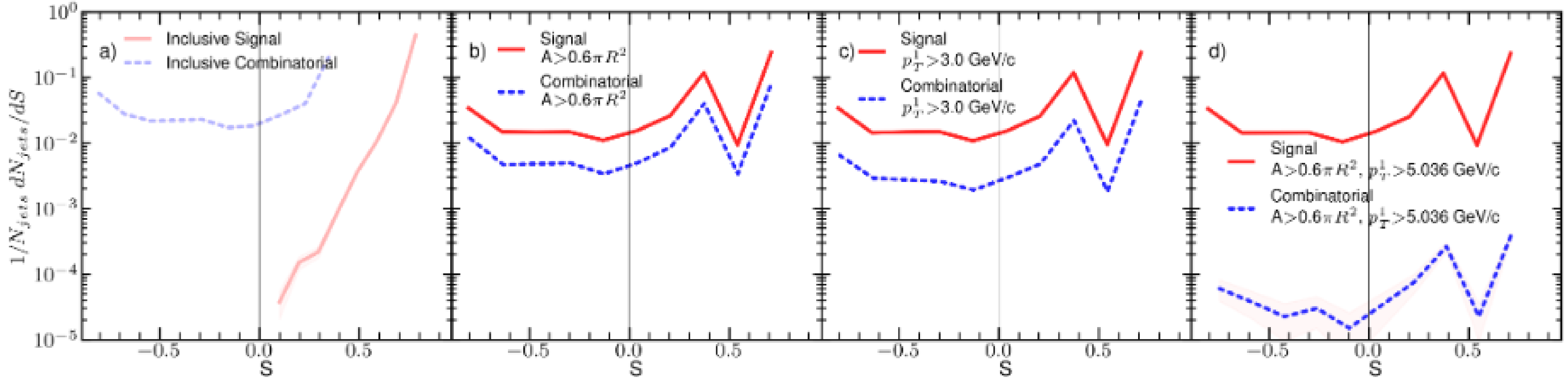
- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

- Silhouette value

$$s_i = \frac{b_i - a_i}{max[b_i, a_i]}$$



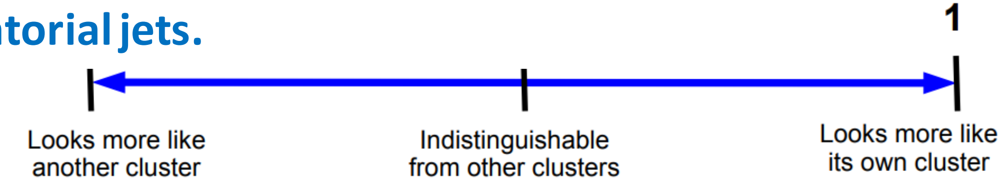
Taken from C. Nattrass



Backup

- Can we come up with a set of cuts to better remove combinatorial jets from signal (hard-scattering origin) jets ?

These cuts work well but ALWAYS leave a population of combinatorial jets that look like signal jets. The addition of the leading hadron pT cut removes a lot of combinatorial jets.



Taken from C. Nattrass