JET CLEANSING

Mitigation of pileup effects at the LHC workshop

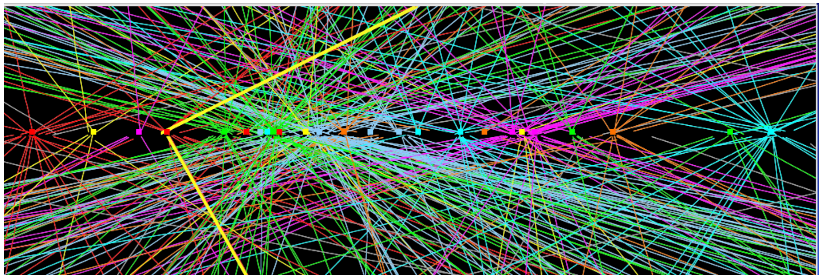
CERN

May 17, 2014

Matthew Schwartz
Harvard University

Based on work with David Krohn, Matthew Low, and Lian Tao Wang

Goals of pileup removal



1. Reconstruct **kinematic observables**
   - Easiest pileup removal technique: small R
   - Anti-k_T R=0.4 standard
     - about as small as we can go

2. Reconstruct **substructure observables**
   - Do boosted techniques work at high piluep?
     - Want R = 1.0 or larger
   - Can we maintain **correlations** between observables at high pileup
     - PU removal should be **observable-independent**

3. Maintain **search reach**
   - Remove PU from signal and background
   - How does PU removal work with jet grooming?

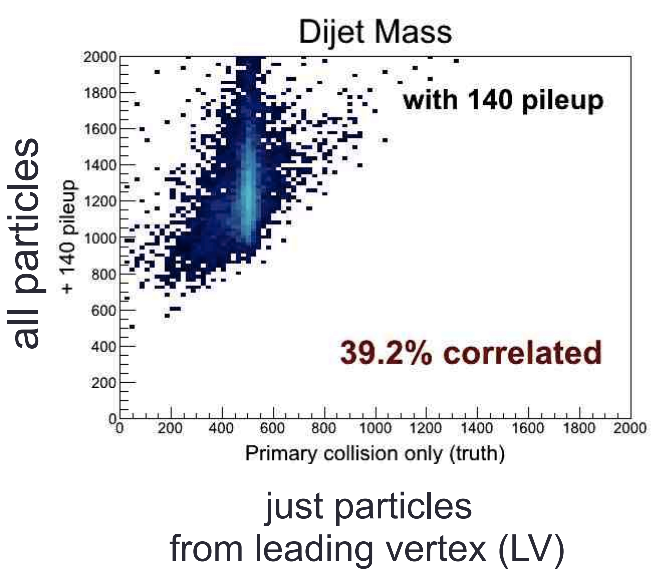
Measures of success

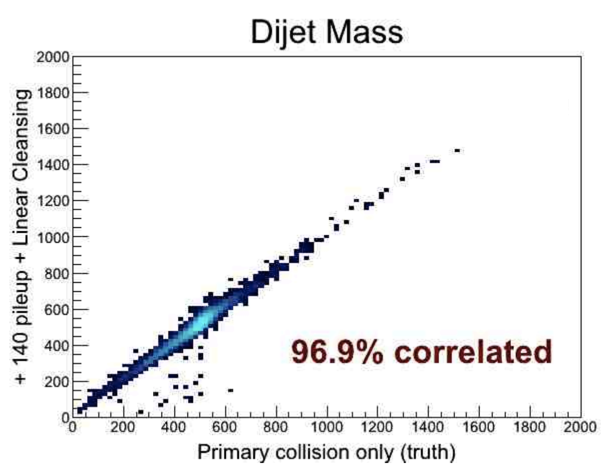
1. **S/B or S/√B**
   - Appropriate for search reach studies

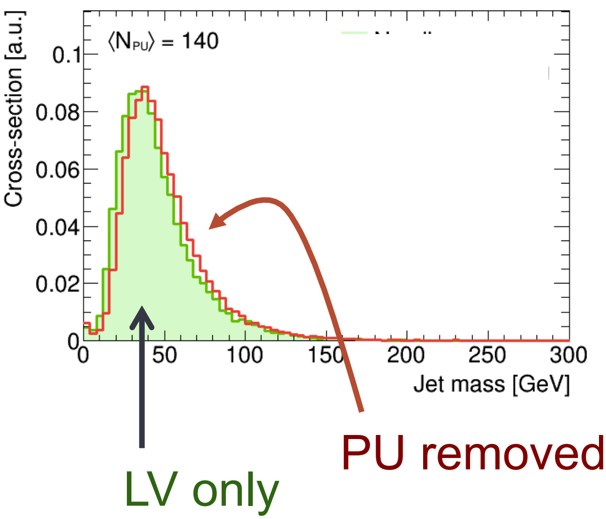
2. **Offset** (p_T^{cor} – p_T^{truth}) and **dispersion/resolution**
   - zero offset not essential -- can be corrected (as in ATLAS)

3. Pearson linear **correlation coefficient**
   - Standard statistical measure
   - Nice properties: constant offset = 100% correlated
   - Ideal combination of offset and dispersion.

4. Visualizations





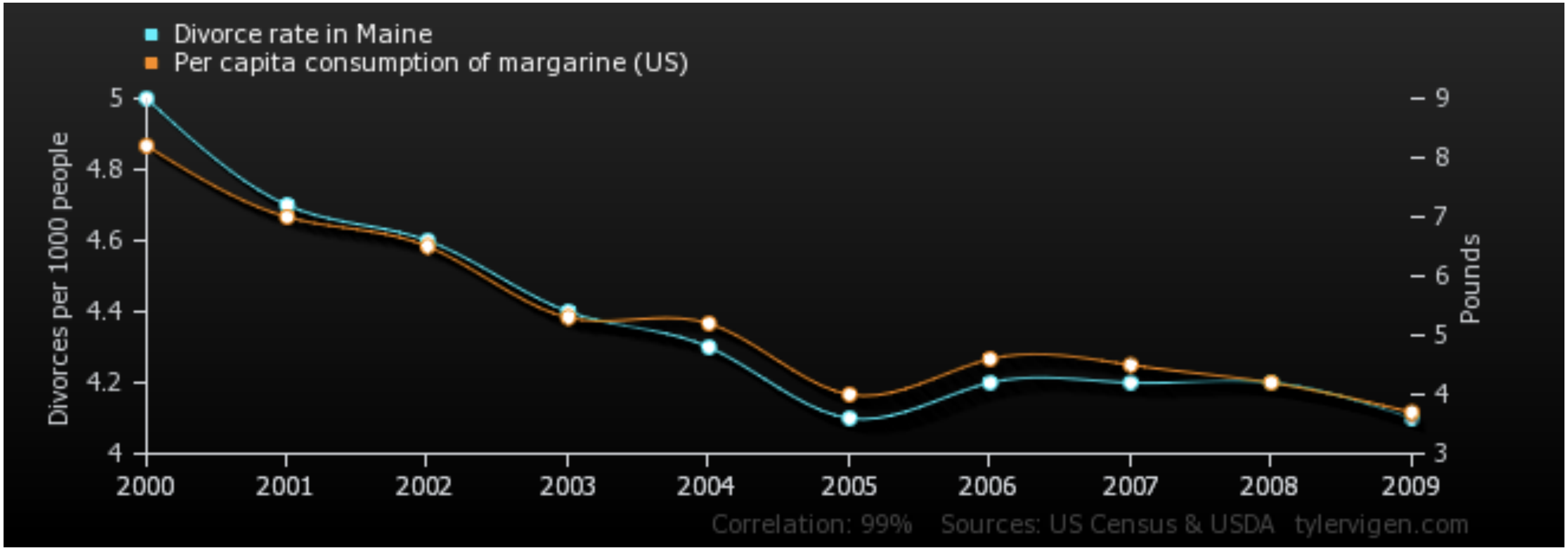


Correlation coefficients

Divorce rate in Maine

vs

Per capita consumption of margarine in the US



99.2% correlated

Area subtraction

Estimate amount of pileup contamination ρ in an event from control regions

$$p^\mu_{\text{corrected}} = p^\mu - \rho A^\mu$$

Use to correction jet 4-momentum

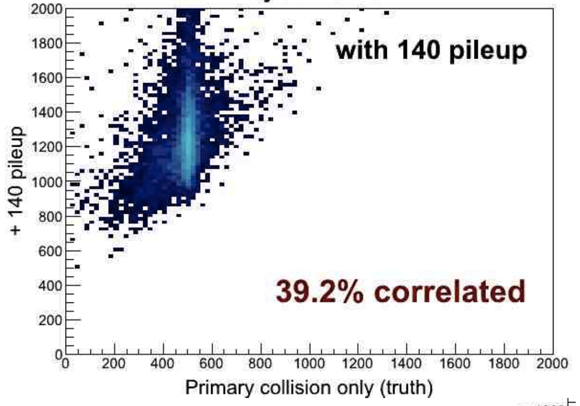
[Cacciari, Salam; arxiv:0707.1378]

$$V_{\text{corrected}} = V - \rho V^{[1]} + \frac{1}{2}\rho^2 V^{[2]} + \ldots$$

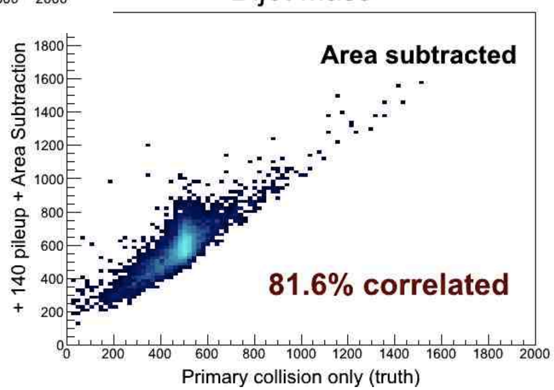
Use it to correct jet shapes (e.g. mass)

[Soyez, Salam, Kim, Dutta, Cacciari; arxiv:1211.2811]

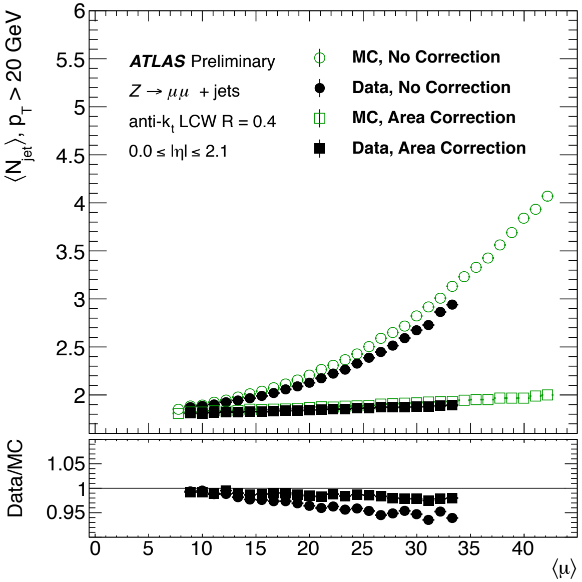
Dijet Mass



Dijet Mass

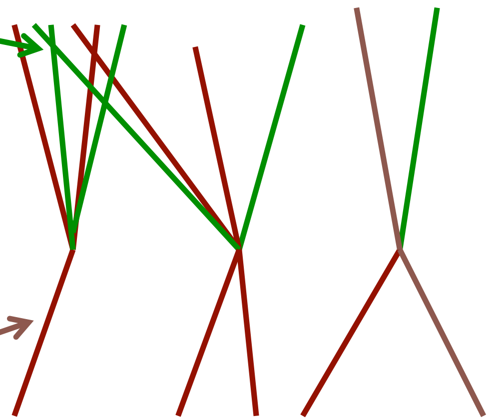


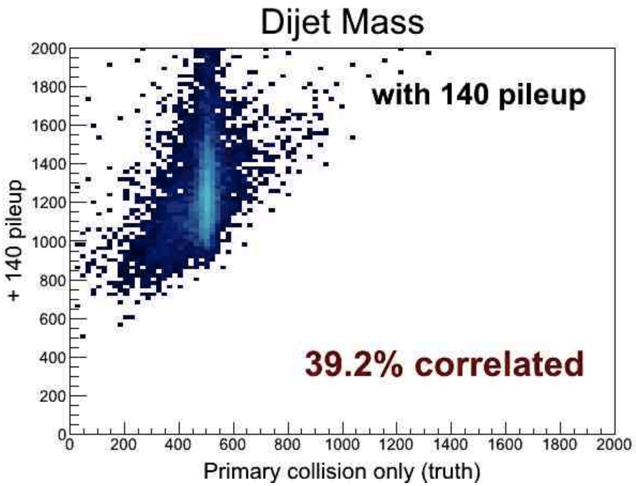
Works great on data



(a) Central Region

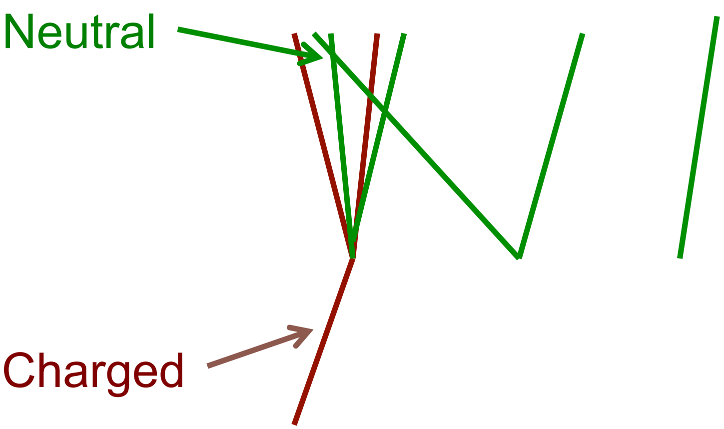
Tracking information

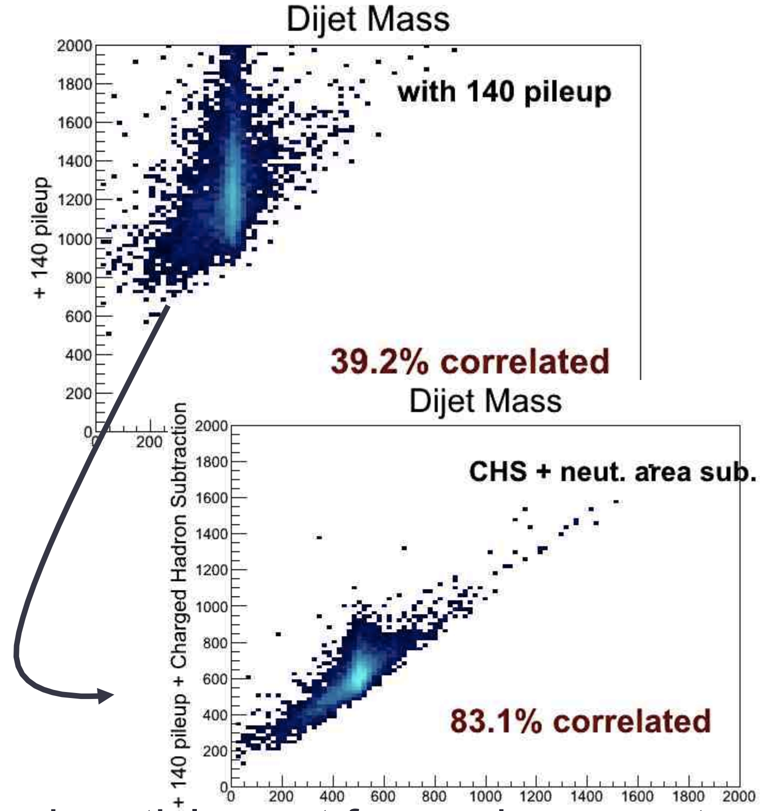




1. Charged Hadron Subtraction
   - Used by CMS (at least). Simply drop charged particles not from primary vertex
   - Since ~65% of particles are charged, this works well.
   - At very large pileup, the other 35% can really hurt

Tracking information





1. Charged Hadron Subtraction
   - Used by CMS (at least). Simply drop charged particles not form primary vertex
   - Since ~65% of particles are charged, this works well.
   - At very large pileup, the other 35% can really hurt

2. Jet vertex fraction = fraction of charged energy coming from pileup vertex
   - Cutting on JVF cleans event sample
   - Also does not scale well at large pileup

Can we do better?

Cleansing

We want to correct each jet in an **observable-independent** way

- Important for multivariate analysis
- Theory-independent

We want to combine tracking and calorimeter information optimally

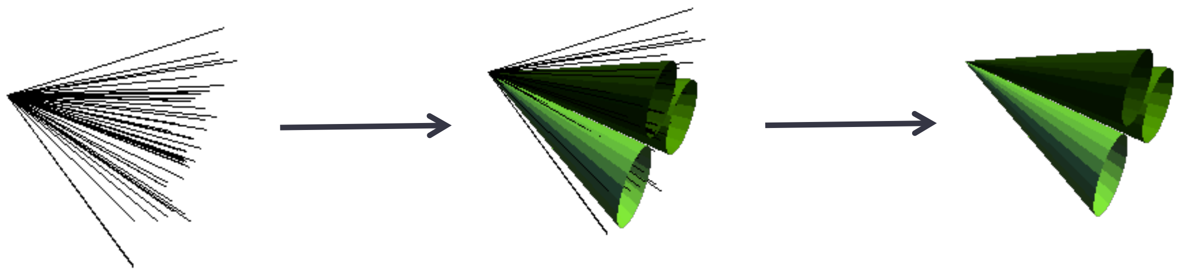
Two simple insights

1. Correcting subjets allows for observable-independence

2. Charged/neutral ratio is less variable in pileup than in jets.

Use subjets

- Natural infrared safe constituents are subjets
  - Find jets, then recluster at smaller R.
  - Can drop some soft subjets (e.g. filtering, trimming, pruning)



Let's try to correct each subjet:

- Observable independent
- Same correction should reproduce most infrared-safe observables
  - Jet 4-momentum
  - Jet shapes (mass, n-subjettiness, moments)

Inputs to cleansing

Let us assume we can measure

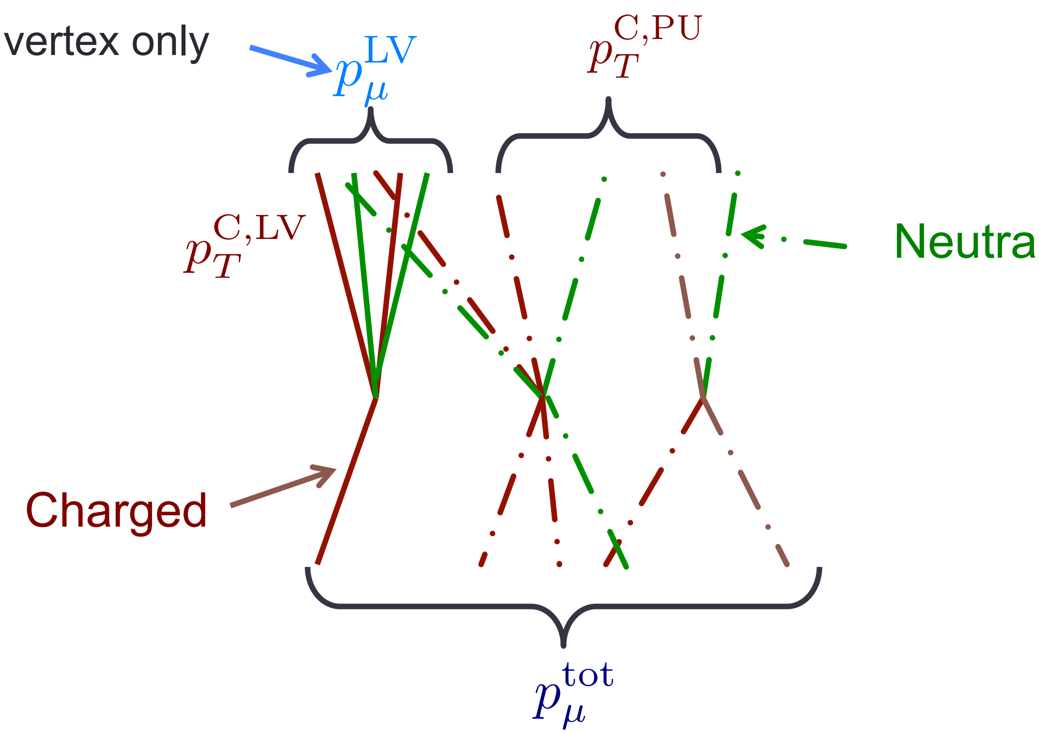
p_μ^{tot} = total momentum in a subjet (charge+neutral)

$p_T^{\text{C,LV}}$ = total charged energy in subjet coming from leading vertex

$p_T^{\text{C,PU}}$ = total charged energy in subjet coming from PU vertices

Goal: combine these to get

p_μ^{LV} = 4-momentum of subjet, leading vertex only



1. Neutral proportional to charged (NpC)

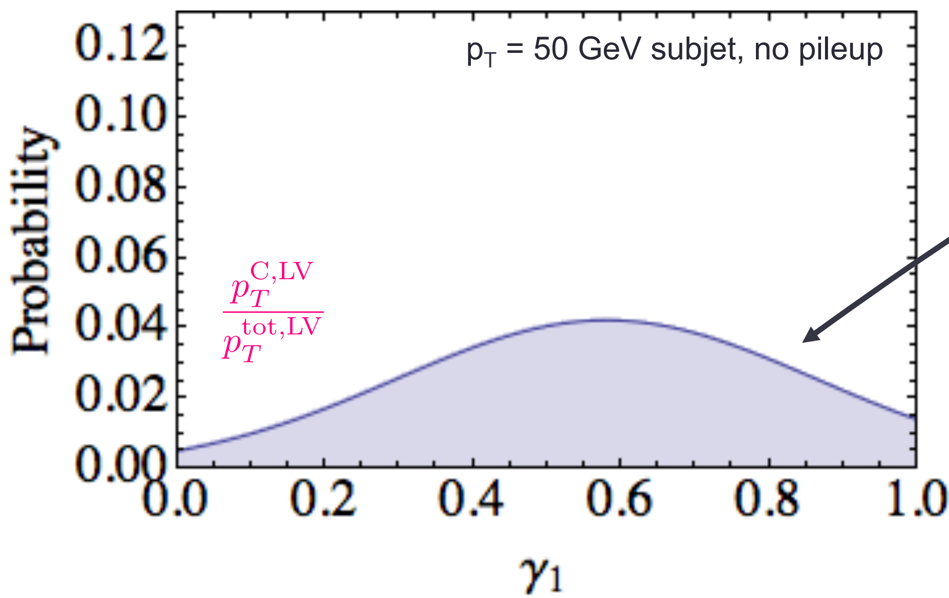
Assume $\gamma = \frac{\text{energy in charged particles}}{\text{total energy}}$ is constant $\approx \frac{2}{3}$ (isospin limit)

Correct with $p_\mu^{\text{jet,sub}} = p_\mu^{\text{jet}} - \frac{1}{\gamma_0} p_\mu^{\text{jet,chg-PU}}$

(or, after charged-hadron subtraction)

$$p_\mu^{\text{jet,sub}} = p_\mu^{\text{jet,CHS}} - \frac{1-\gamma_0}{\gamma_0 \epsilon} p_\mu^{\text{jet,rescaled-chg-PU}}$$

Cacciari, Salam, Soyez arXiv:1404.7353



Very broad distribution

- Energy in charged and neutral very uncorrelated
- Does not work very well

2. JVF cleansing

Assume charged/neutral ratio in each **subjet** is the same for PU or LV

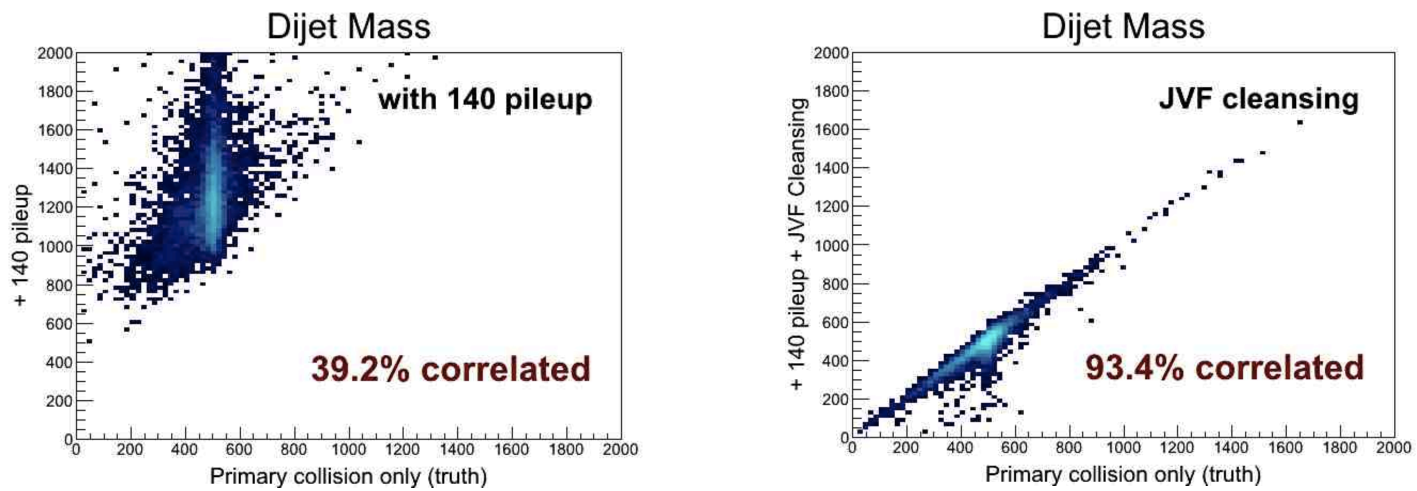
$$p_\mu^{\rm LV} = p_\mu^{\rm tot} \times \frac{p_T^{\rm C,LV}}{p_T^{\rm C,LV} + p_T^{\rm C,PU}}.$$

JVF

multiplication (not subtraction)

- can be significantly different

- Exact at zero PU
- Works very well at low pileup
- Degrades somewhat at higher pileup

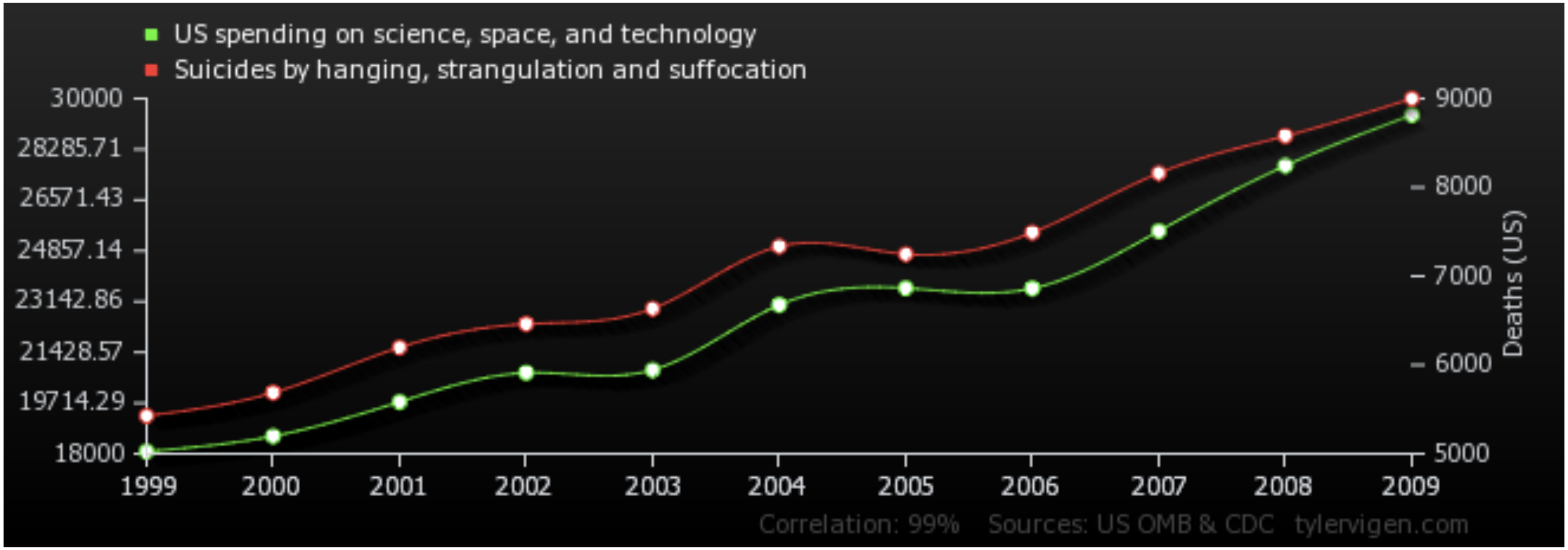


Correlation coefficients

US spending on science, space, and technology

vs

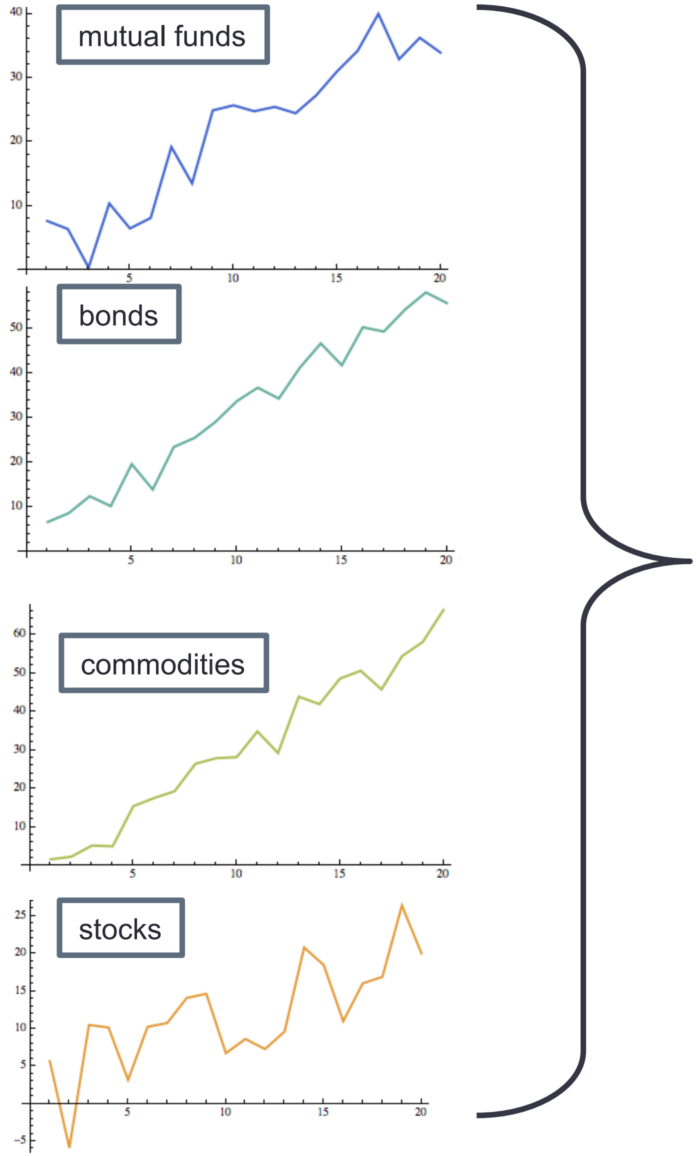
Suicides by hanging, strangulation and suffocation

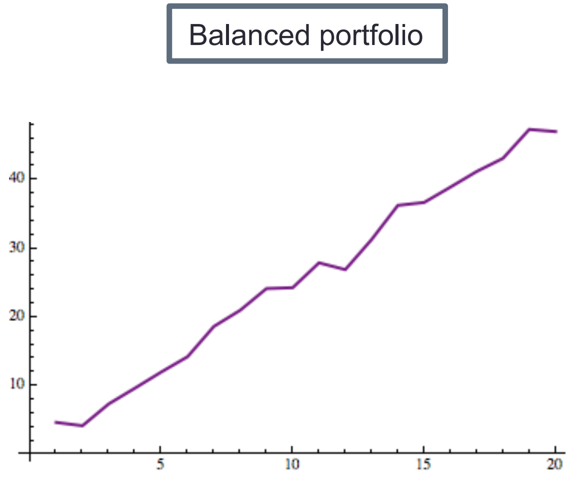


99.2% correlated

LINEAR CLEANSING

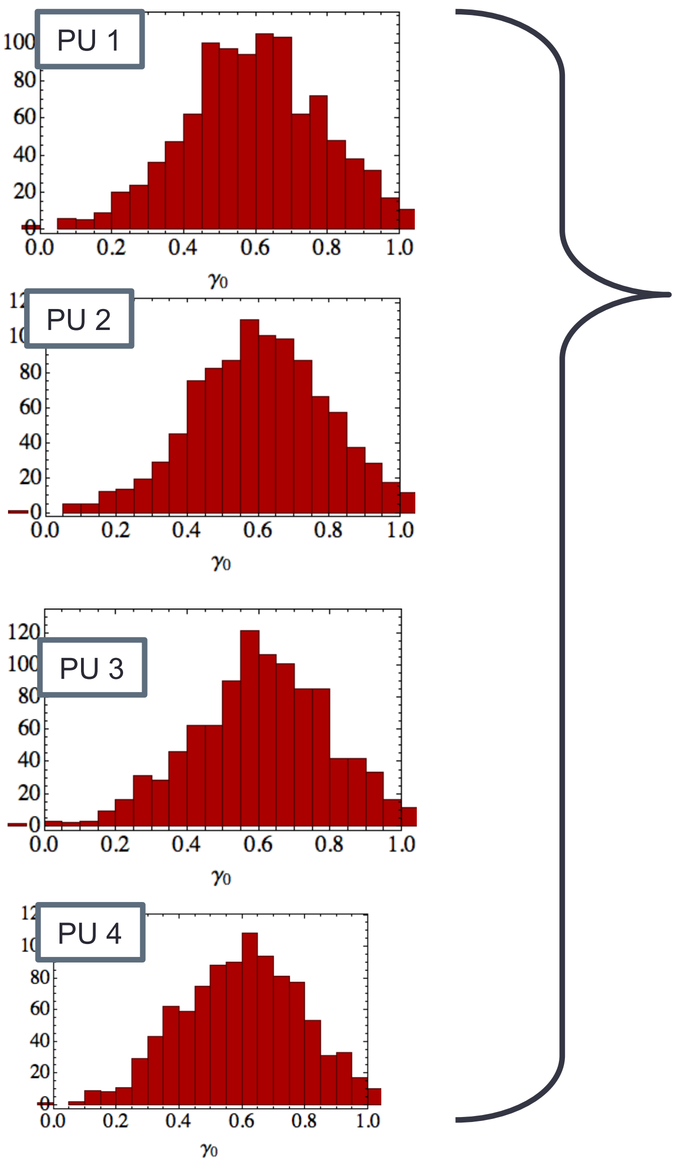
Diversify:



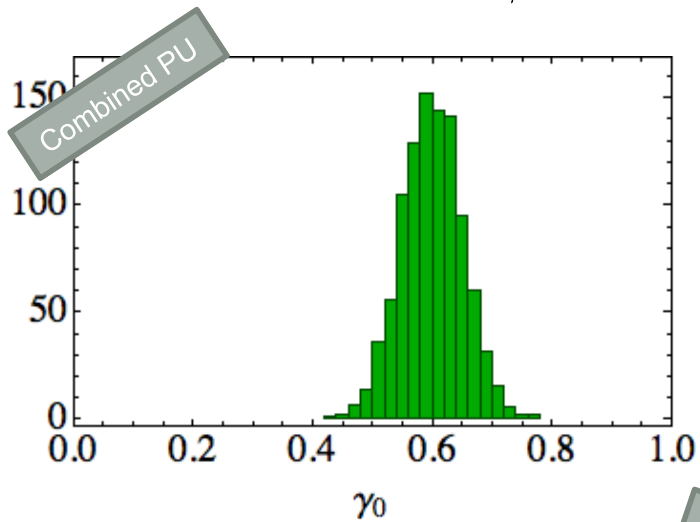


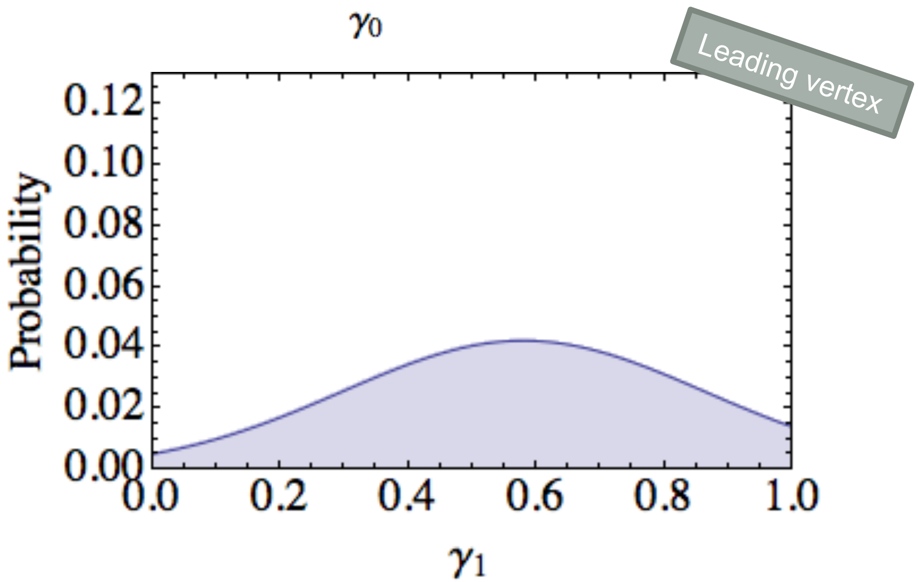
- Same gain, on average
- Less volatility

Pileup is stochastic: mathematica



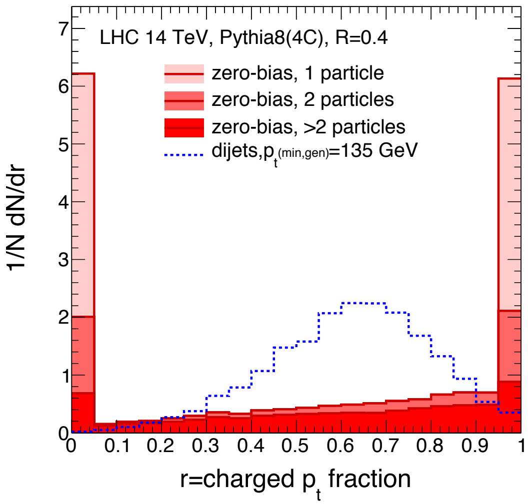
$$\gamma = \frac{\text{energy in charged particles}}{\text{total energy}}$$



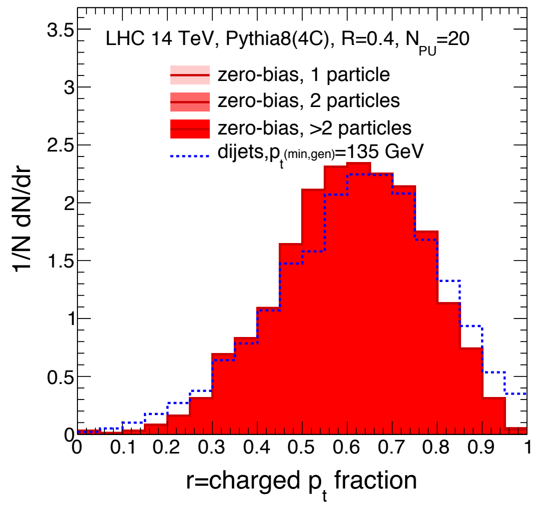


Piluep is stochastic: pythia

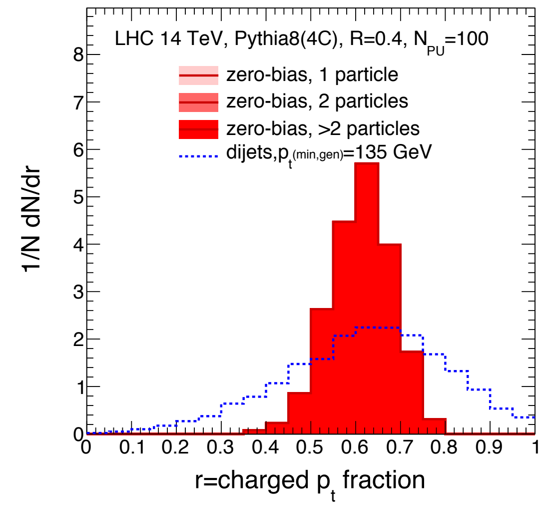
$N_{PU}=1$

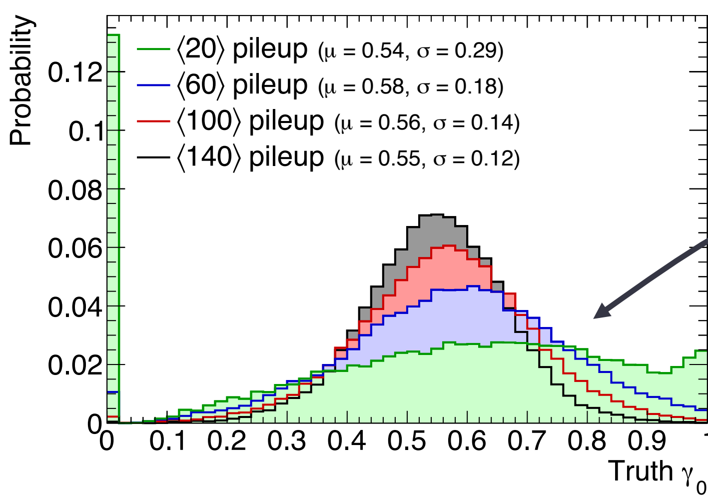


$N_{PU}=20$



$N_{PU}=100$





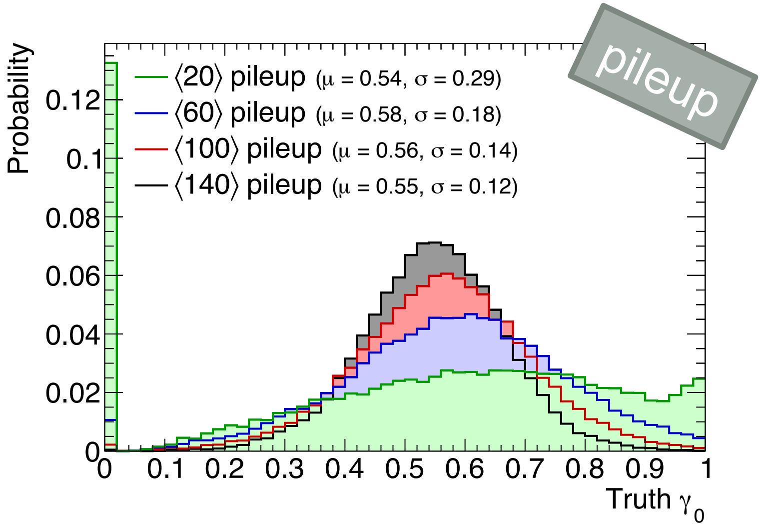
Sharper and sharper with more pileup

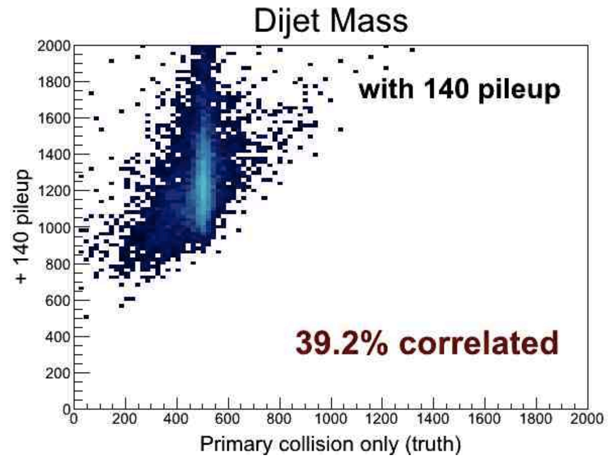
3. Linear cleansing

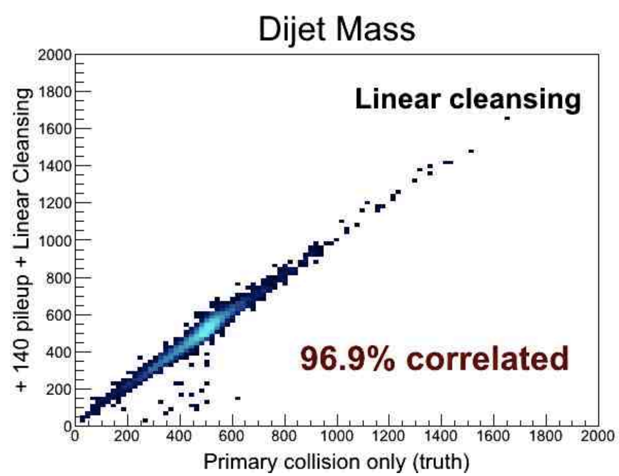
$$\gamma = \frac{\text{energy in charged particles}}{\text{total energy}}$$

Key insight: **Piluep is stochastic (Poisson)**

If you average a bunch of Gaussians, the width scales like $\frac{1}{\sqrt{N_{\mathrm{PU}}}}$







Linear cleansing

$$p_T^{\mathrm{C,PU}} \longrightarrow p_T^{\mathrm{tot,PU}}$$

$$p_\mu^{\mathrm{LV}} = p_\mu^{\mathrm{tot}} - p_T^{\mathrm{tot,PU}}$$

- Don't use $p_T^{\mathrm{C,LV}}$ at all
- Perfect in $N_{\mathrm{PU}} \to \infty$ limit

Correlations

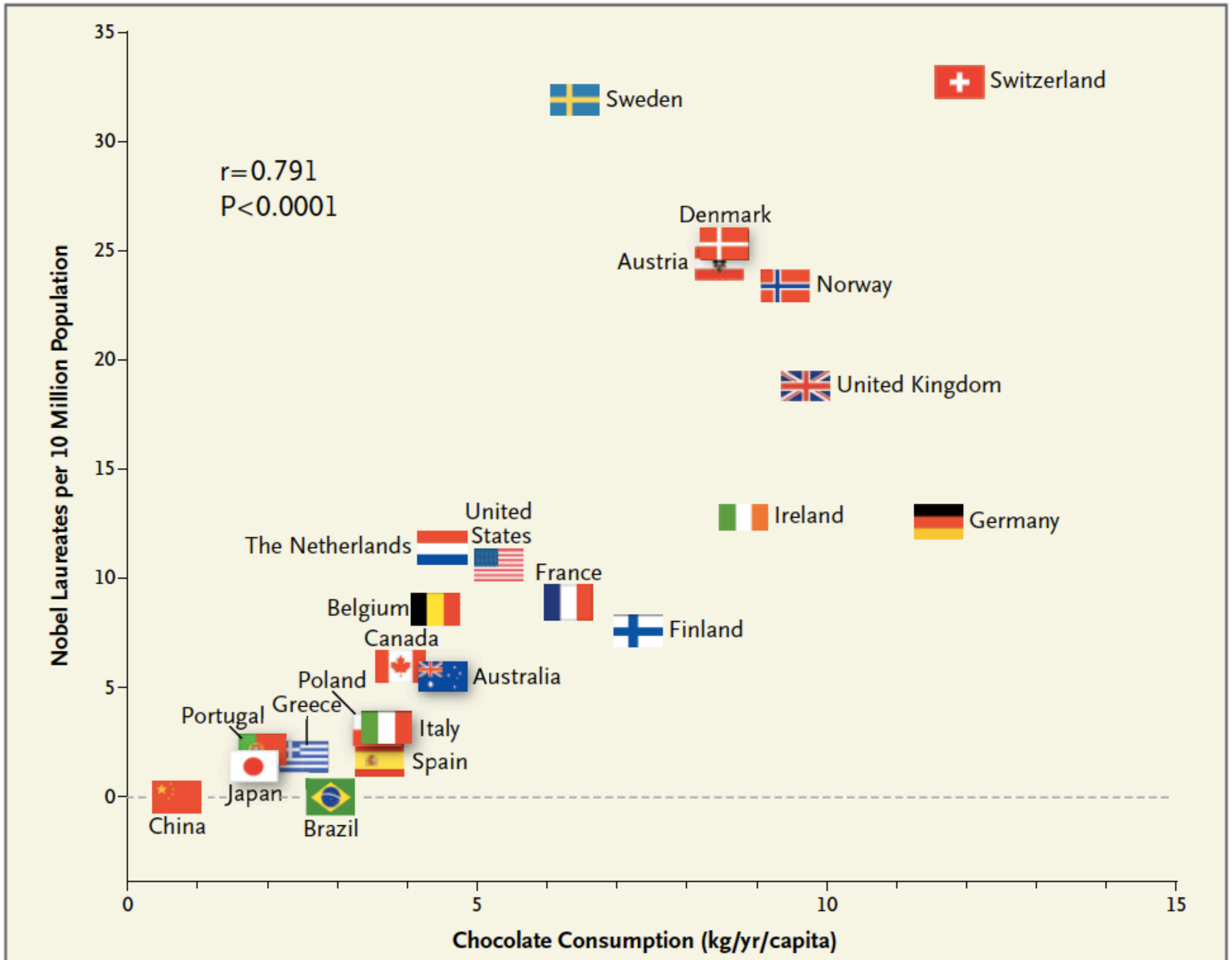


Figure 1. **Correlation between Countries' Annual Per Capita Chocolate Consumption and the Number of Nobel Laureates per 10 Million Population.**

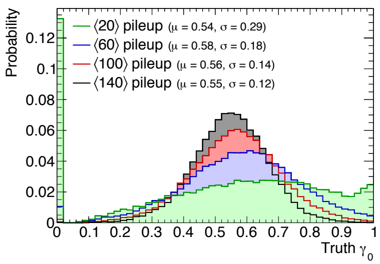
Gaussian cleansing

3 observables constrained by

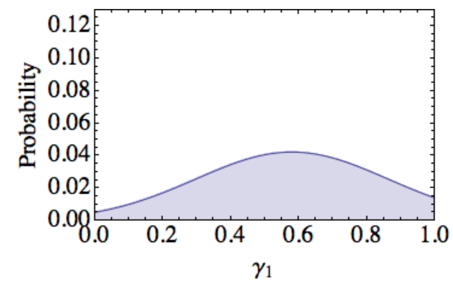
$$p_T^{\text{tot}} = \frac{p_T^{\text{C,PU}}}{\gamma_0} + \frac{p_T^{\text{C,LV}}}{\gamma_1}$$

$$\gamma_0 = \frac{\text{energy in charged particles for PU}}{\text{total energy for PU}}$$

$$\gamma_1 = \frac{\text{energy in charged particles for LV}}{\text{total energy for LV}}$$



$$\overline{\gamma_0} = 0.55$$
$$\sigma_0 = 0.15$$

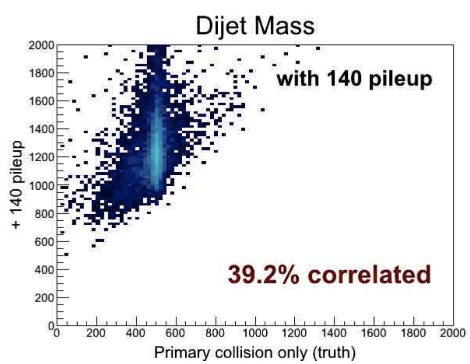


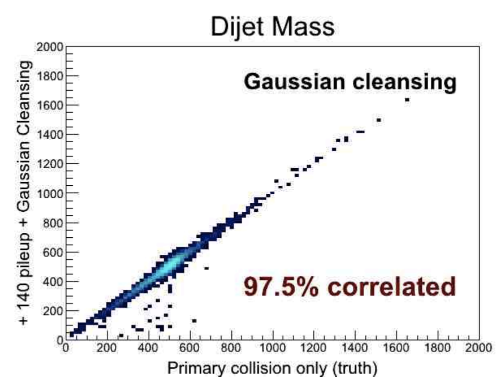
$$\overline{\gamma_1} = 0.62$$
$$\sigma_1 = 0.22$$

Approximate as Gaussian

$$P(\gamma_0, \gamma_1) \propto \exp\left[-\frac{1}{2}\sum_{i=0,1}\left(\frac{\gamma_i - \overline{\gamma_i}}{\sigma_i}\right)^2\right]$$

Maximize probability satisfying constraint



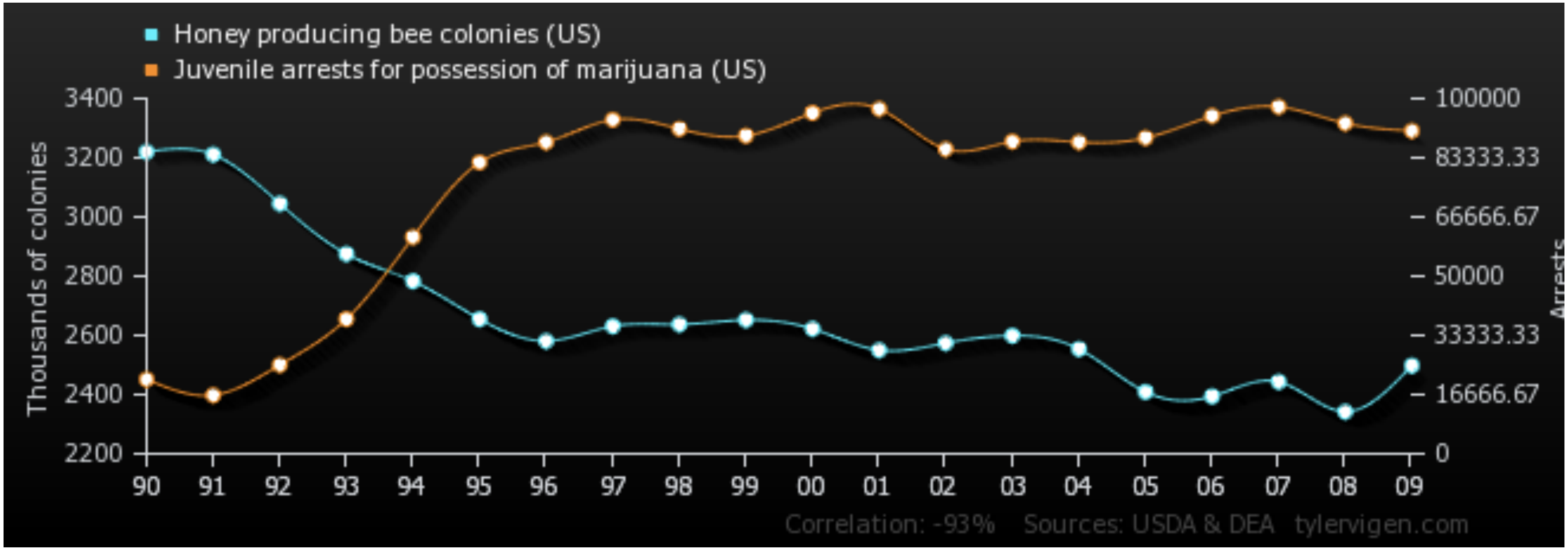


Correlations

Honey producing bee colonies

vs

Juvenile arrests for possession of marijuana

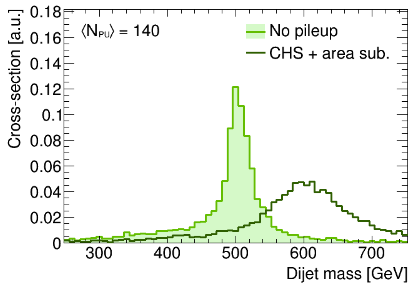


93.3% anti-correlated

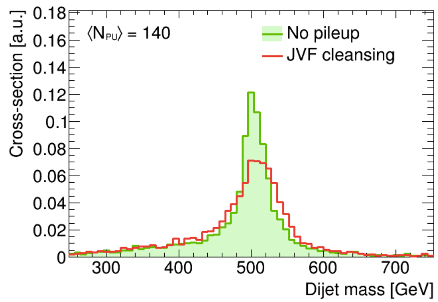
RESULTS AND COMPARISONS

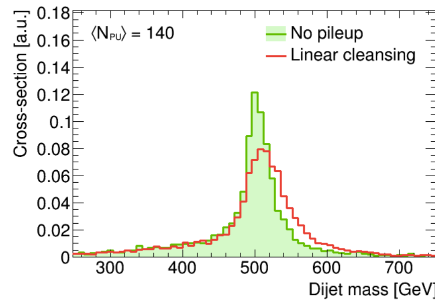
Many plots here: http://jets.physics.harvard.edu/Cleansing

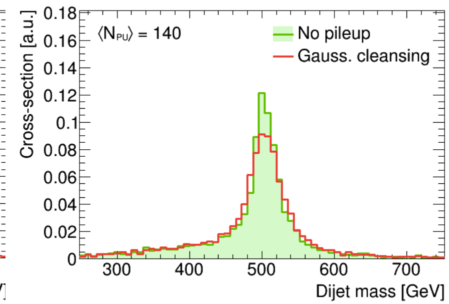
Dijet mass (kinematic variable)



CHS+area sub.





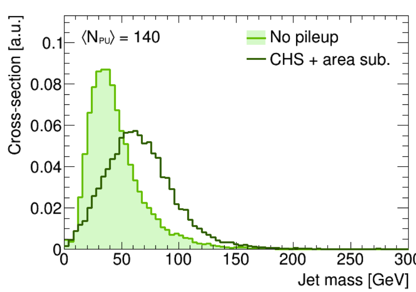


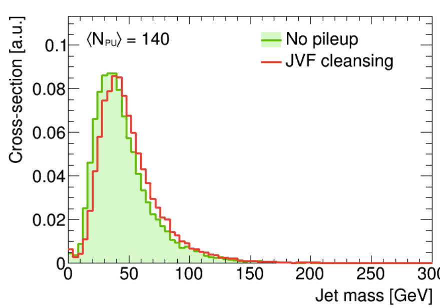
JVF cleansing

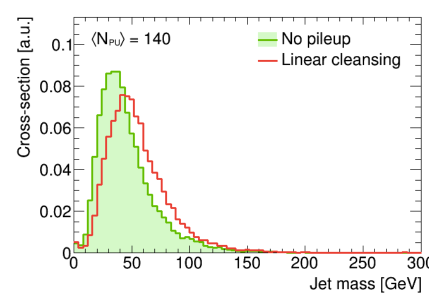
Linear cleansing

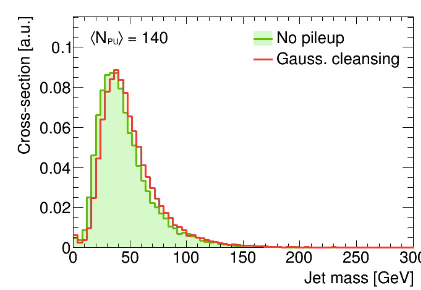
Gaussian cleansing

CHS+shape sub.



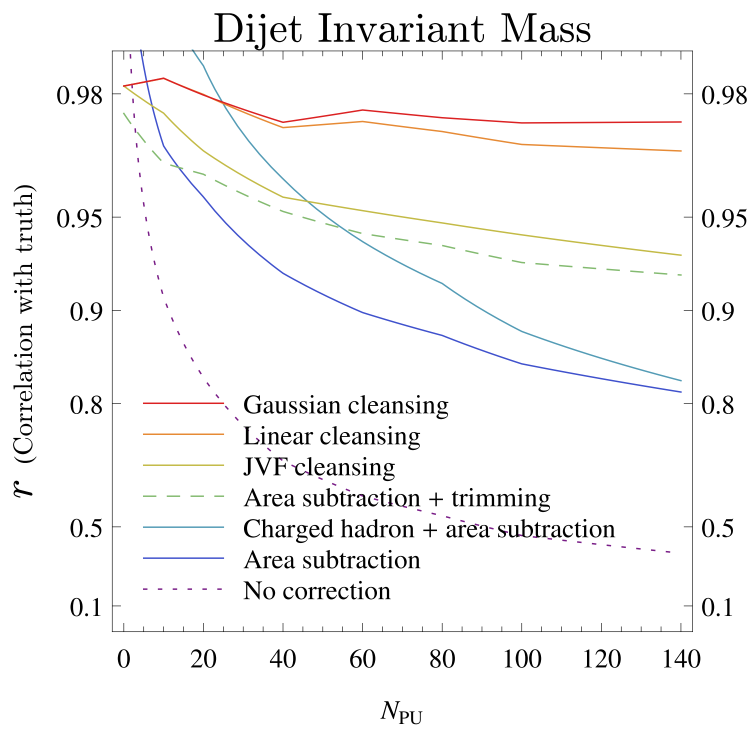


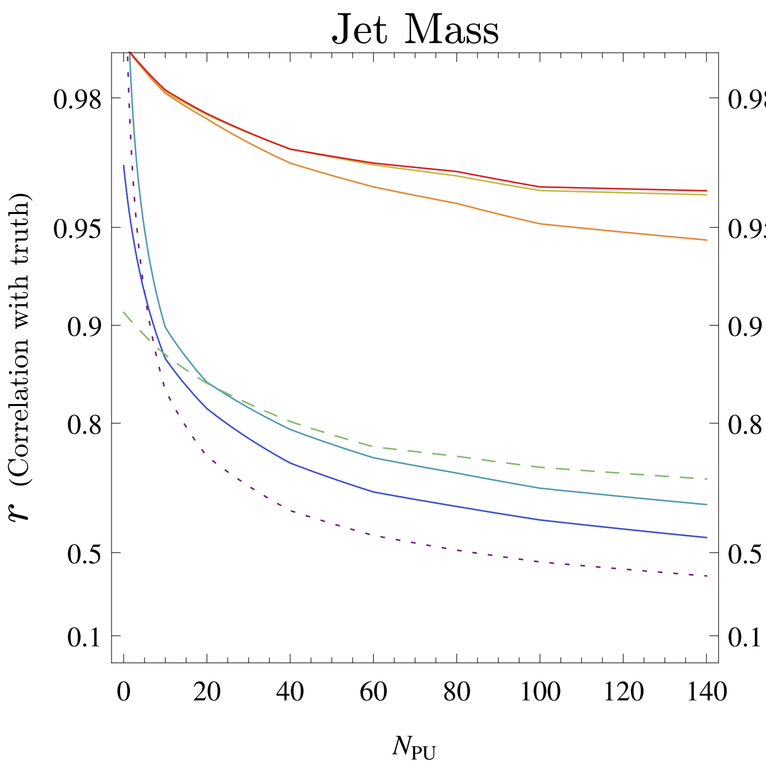




Jet mass (shape variable)

Correlation coefficients vs N_{PU}

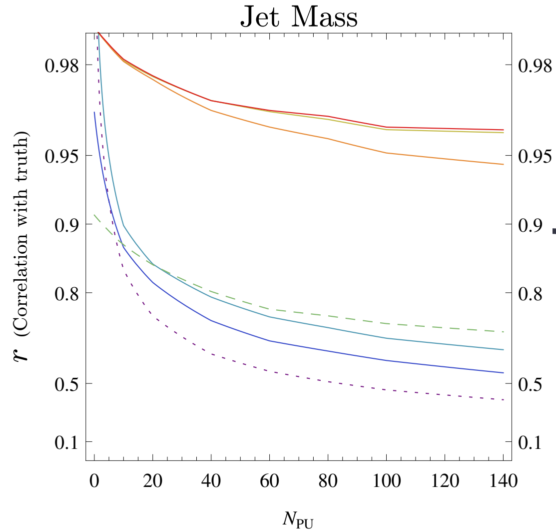




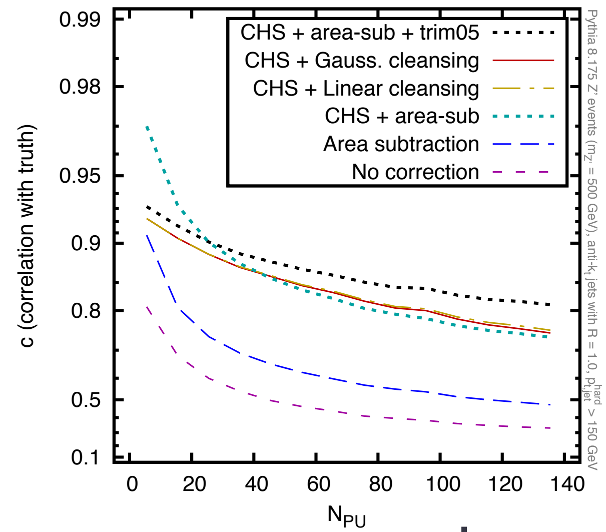
Look at Green and Red curves

Our original results (KLSW)

- Cleansing R_{sub}=0.3
- KLSW truth definition
- Fastjet area subtraction



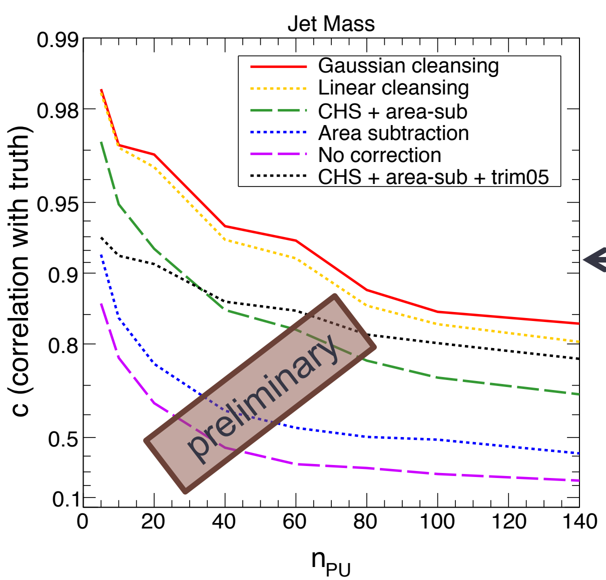
CSS arXiv:1404.7353

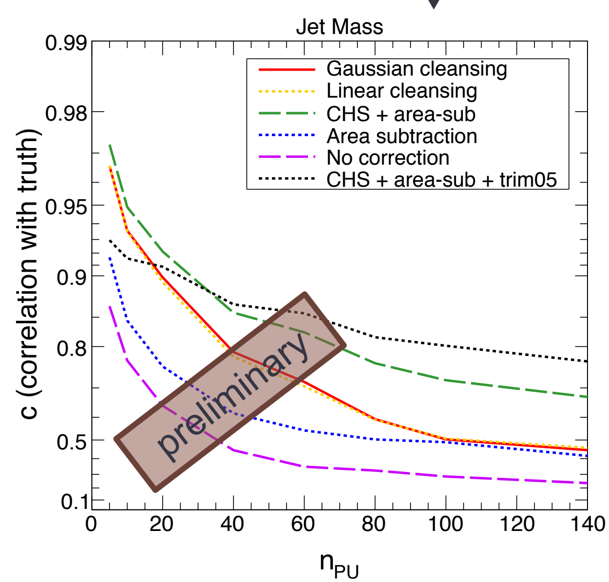


- Cleansing R_{sub}=1000
- CSS truth definition
- Safe area subtraction

Our results (new)

- Cleansing R_{sub}=0.3
- CSS truth definition
- Safe area subtraction





Our results (new)

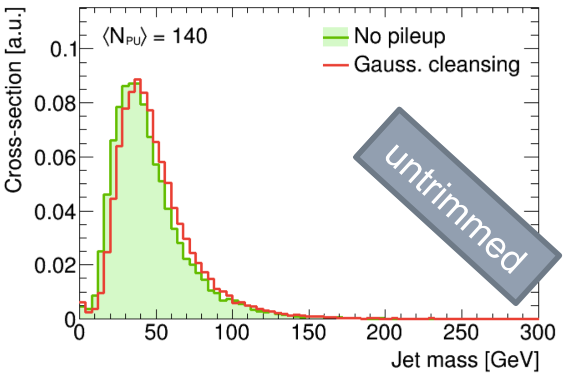
- Cleansing R_{sub}=1000
- CSS truth definition
- Safe area subtraction

Trimming

Different goals, different methods

- Just **remove pileup** (e.g. for precision QCD)

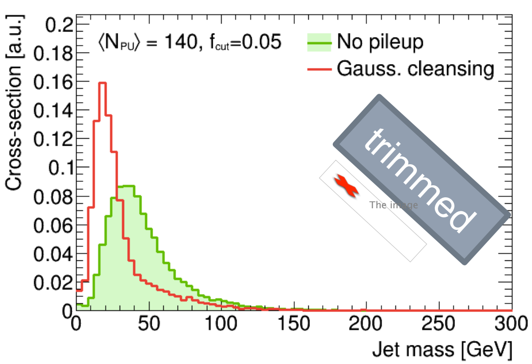
Pileup + grooming (e.g. to maximize search reach)



- Cleansing dovetails nicely with filtering and trimming

$$f = \frac{\text{cleansed subjet } p_T}{\text{cleansed jet } p_T}$$

Drop subjets with $f < f_{\text{cut}}$



Significance improvement				
Algorithm	$N_{\text{PU}} = 20$ plain	$N_{\text{PU}} = 20$ trimmed	$N_{\text{PU}} = 140$ plain	$N_{\text{PU}} = 140$ trimmed
CH + area Sub.	0.86	1.07	0.48	0.90
Area subtraction	0.87	1.00	0.45	0.85
JVF cleansing	0.93	1.06	0.82	0.81
Linear cleansing	0.94	1.08	0.78	1.00
Gaussian cleansing	0.95	1.07	0.91	0.98

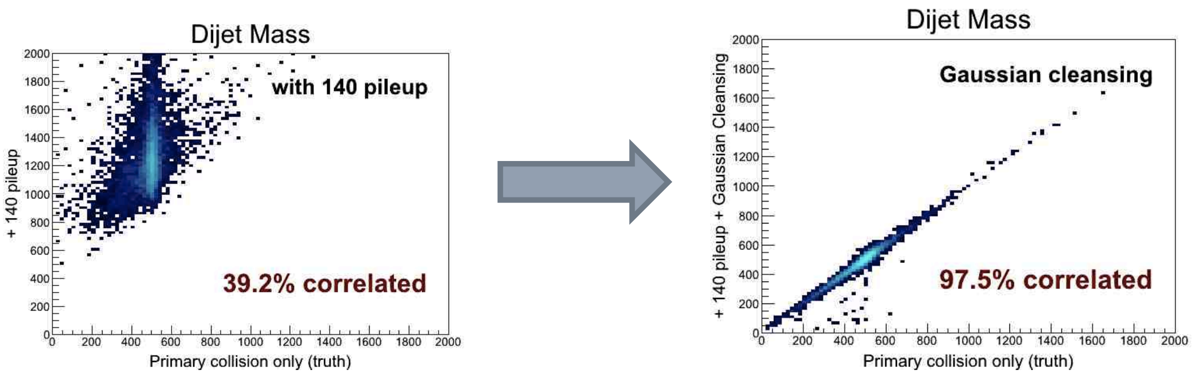
Conclusions

1. Pileup subtraction is naturally done at subjet level
   - Same corrections apply to kinematic variables and shape variables

2. How can we combine charged/neutural, leading/pileup to get subjet momenta?

$$p_T^{C,LV} \quad p_\mu^{tot} \quad p_T^{C,PU} \quad \Rightarrow \quad p_\mu^{LV}$$

a) JVF cleansing: at low PU, charge/neutral same in LV and LV+PU
b) Linear cleansing: at infinite pileup, charge/neutral in PU is exactly known
c) Gaussian cleansing: charge/neutral in LV and PU follow Gaussians



Final thoughts

For anti-k_T R=0.4, N_{PU}=40, pileup removal is basically solved

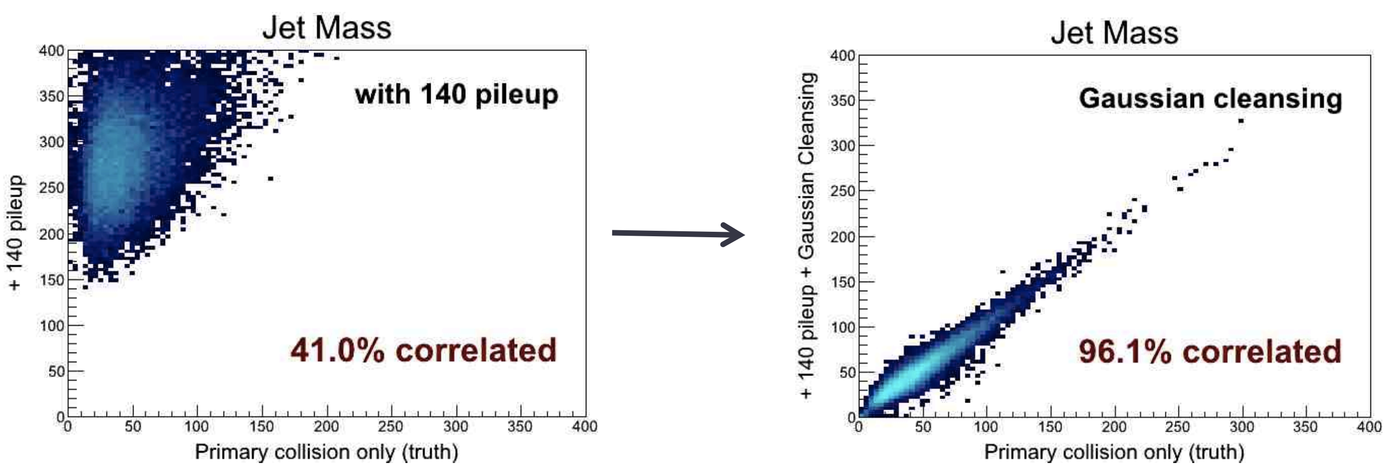
- Area subtraction + CHS + JVF offset works nearly perfectly
- Everything works
- Do not need subjets

Serious challenge:

Can jet substructure techniques be used at very high luminosity?

- Jet mass at R=1.0 is a good test
- N_{PU} = 140

Cleansing works well

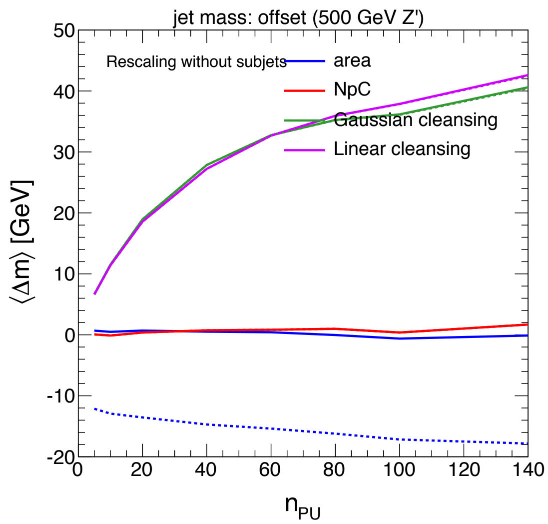


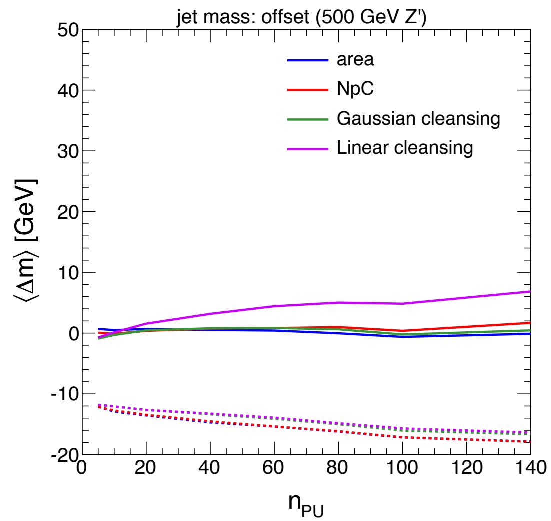
Still room for improvement

BACKUP SLIDES

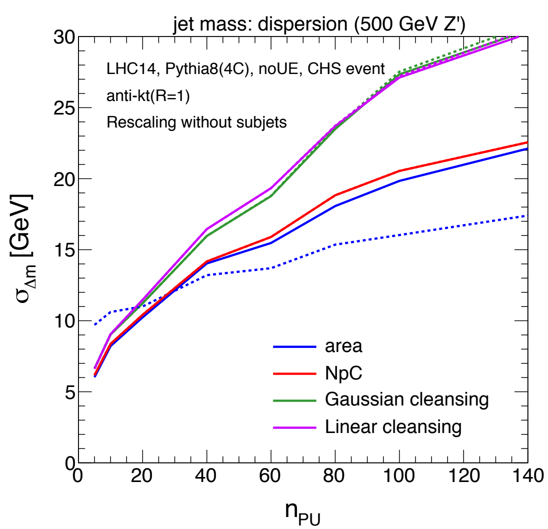
Offset and dispersion

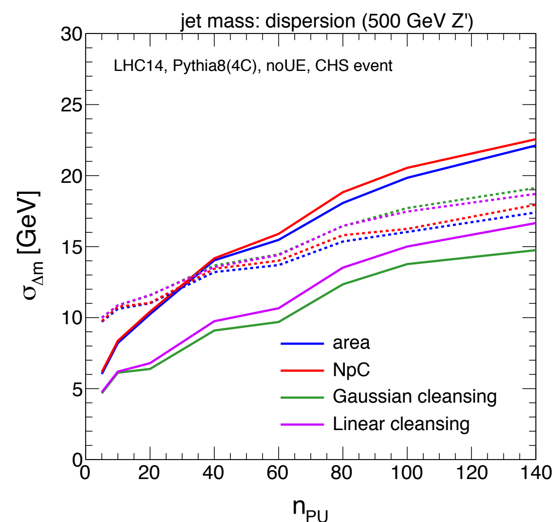
offset





dispersion





CSS (R_{sub} =1000)

Our results (KLSW)

Simulation details

- Resonance decaying to light quarks generated with MadGraph5 v1.5.8
- Pythia v8.176 tune 4C, used for pileup and showering.
- Amount of pileup drawn from Poisson.
- FSR, ISR, UE included.
- Charged particles with $p_T < 0.5$ GeV discarded.
- Remaining particles grouped in $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ calo cells.
- Only cells with $E > 1$ GeV are kept.
- Jets with $p_T > 150$ GeV and $|\eta| < 2.5$ after pileup are kept.