Pruning - Comments and Explanations

A dramatic contrast (to my thoroughly biased eye) between the CMS and ATLAS jet grooming/tagging analyses as reported at the recent Boosted Boson Workshop (CERN, 3/25/14) is that essentially **all** CMS analyses use jet pruning while essentially **no** ATLAS analysis does!

Why is that? No one at the Workshop asked this question, but someone should have!

The following notes attempt to address this difference between the two collaborations.

Pruning Refs: 0912.0033, 0903.5081

A Brief Review of Pruning

- Like other groomers, given a jet (identified by some generic jet algorithm like AkT, kT or C/A) pruning attempts to remove from the jets those constituents that are unlikely to be ``associated'' with the jet or at least carry no significant/useful information.

- In particular, we want the mass of the resulting pruned jet to be small if we start with an every-day QCD jet, and near the particle mass if we start with a jet containing the decay products of a heavy particle.

- Pruning will remove the uncorrelated contributions from UE and PU that make significant contributions to the jet mass.

Basic Idea of Pruning -

- Prune (remove) those constituents of the original jet that are:
  soft
  large angle

- These soft, large angle constituents are (statistically) less likely to be correlated with the energetic constituents in the jet and yet can still make measureable contributions to the mass

- Soft, small angle constituents can also be uncorrelated, but make a small contribution to the mass

- Most configurations that arise from actual heavy particle decay will not tend to be pruned (not all, but most).

Pruning in Action

- Given the list of constituents in a jet, remerge using the kT or C/A algorithm

- At each potential merging step, j+k $\rightarrow$ l, check for
  soft - $p_k/p_l < z_{cut}$ ($p_k < p_j$)
  large angle - $\Delta R_{jk} > R_{cut} * (2m_{jet}/p_{jet})$,
  where $2m_{jet}/p_{jet}$ is angular scale set by jet itself

- If **both** cuts are satisfied, prune (remove) constituent k and proceed

- Larger z_{cut} and smaller R_{cut} values correspond to more aggressive pruning

- The level of pruning tends to be determined by the **LESS** aggressive of the two parameters (since we must satisfy **both** cuts)

Default Parameters

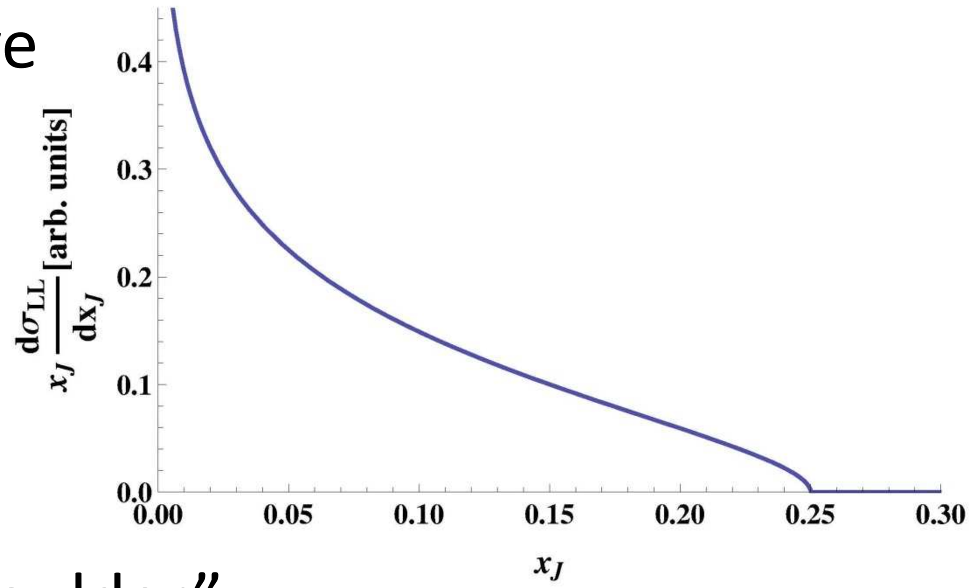
- The original studies (0912.0033) suggested
  R_{cut} = 0.5 (kT & C/A) $\Rightarrow$ $\Delta R > m_{jet}/p_{jet}$
  z_{cut} = 0.1 (C/A)
  z_{cut} = 0.15 (kT, since nearby soft constituents are merged early and are no longer as soft)

- Also, to ensure that decay products of “signal” particle “fit” in jet (size R) and are rarely pruned,
  require $m_{particle}/p_{jet}/R$ be less than 0.5

  as illustrated in the next 2 slides

Naïve NLO 2-body analysis

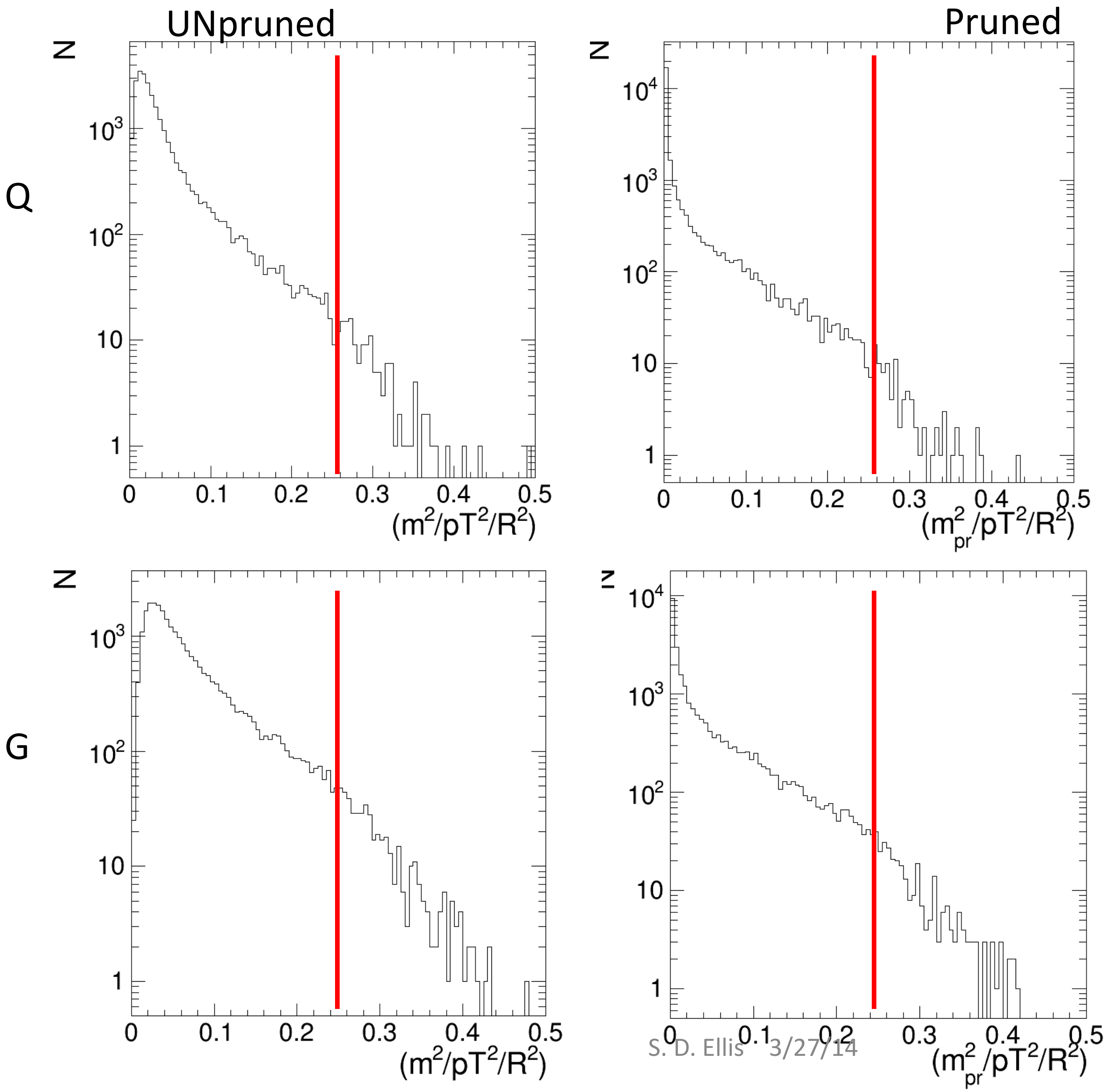
- Distribution vanishes above $x_J=(m_{particle}/p_{jet}/R)^2=0.25$ (from 0912.0033)



- With more complete showering the distribution goes past 0.25 but the "shoulder" is rapidly falling there

- QCD jet distributions from Nhan shown on next slide

Recall plots from Nhan



Distributions cut-off rapidly above $m^2/pT^2/R^2 = 0.25$

Cut-off (a little) more quickly after pruning

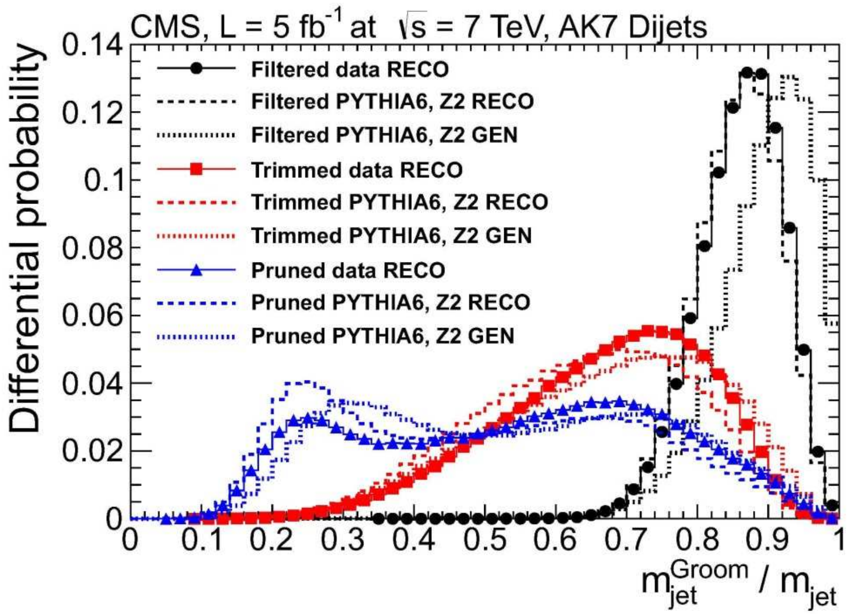
Pruning at CMS, e.g., 1303.4811

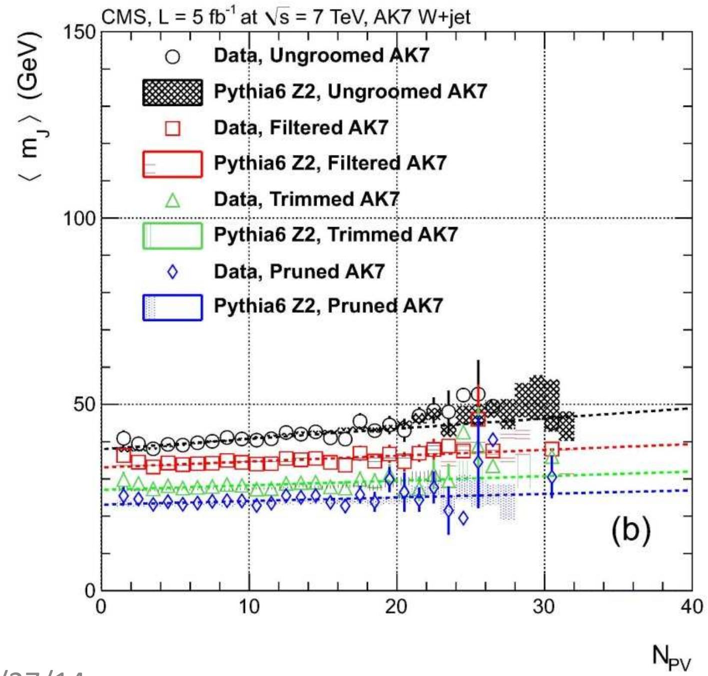
- Prune with C/A using parameters

  z_{cut} = 0.1 (default)

  R_{cut} = 0.25, $\Delta R > 0.5\, m_{jet}/p_{jet}$ (aggressive)

- Pruning is most aggressive groomer studied -



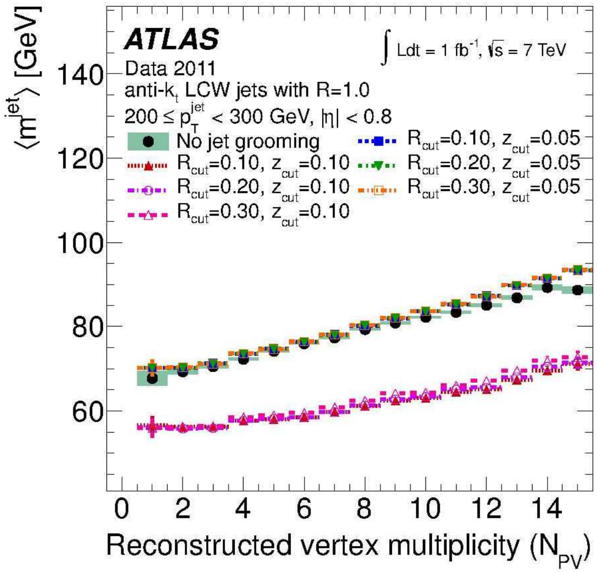


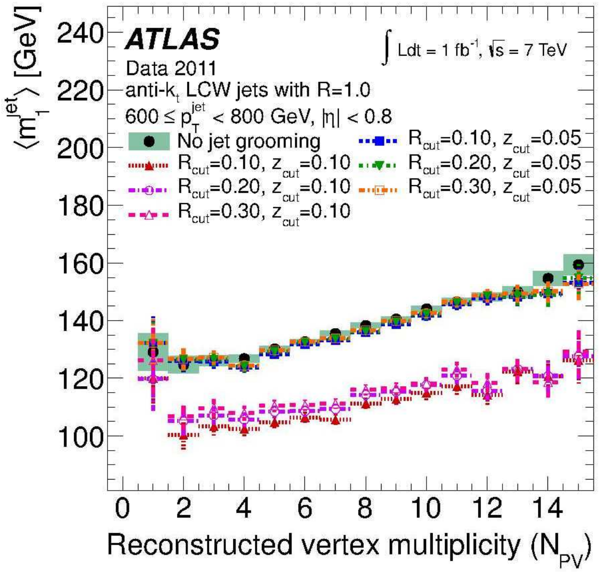
Pruning at ATLAS – 1306.4945

- Prune with kT using parameters

  z_{cut} = 0.1, 0.05 (less aggressive than default = 0.15)

  R_{cut} = 0.1, 0.2, 0.3, $\Delta R > R_{cut}\,(2m_{jet}/p_{jet})$ (more aggressive than default)





- Conclude pruning is **NOT** most aggressive groomer – **AS EXPECTED** due to parameter choices

At BOOST 2013 Clear Difference

- CMS reports pruning is aggressive groomer (when using aggressive (R_{cut}) and default (z_{cut}) parameters,

  and continues to use pruning

- ATLAS reports pruning is **NOT** aggressive groomer using aggressive (R_{cut}) and **NOT** aggressive (z_{cut}) parameter values (controlled by latter)

  and does not use pruning (in studies)

Latest ATLAS results – ATL-PHYS-PUB-2014-004 (3/26/14)

- Study 5 combinations of algorithm + groomer, :
  AK10 + trim, C/A12 + BDRS, C/A12 + BDRS-A, C/A8 + C/A prune (0.1,0.5 = default), C/A8 + kT prune (0.1,0.5 less aggressive than default as above)

  in 3 pT bins using 7 kinematic variables for W tagger

- First do ROC study of algorithm + groomer: define groomed jet mass window to keep 68% of signal (W) and check QCD fake rate (MC data, check that found jets match truth jets almost all of time)

ATL-PHYS-PUB-2014-004

- Results (Table 1) - ε_{QCD} = fake rate, $\mathcal{P}^{+\mathcal{U}}_{-\mathcal{L}}$= mass window

Jet collection	$200 < p_{\mathrm{T}}^{turth} < 350$ GeV		$350 < p_{\mathrm{T}}^{truth} < 500$ GeV		$500 < p_{\mathrm{T}}^{truth} < 1000$ GeV	
	$\mathcal{P}^{+U}_{-L}$ [GeV]	ϵ_{QCD}	$\mathcal{P}^{+U}_{-L}$ [GeV]	ϵ_{QCD}	$\mathcal{P}^{+U}_{-L}$ [GeV]	ϵ_{QCD}
Trimmed	82^{+10}_{-18}	$13.6 \pm 0.1\%$	80^{+8}_{-8}	$9.9 \pm 0.2\%$	82^{+10}_{-10}	$8.4 \pm 0.5\%$
BDRS	78^{+14}_{-16}	$14.8 \pm 0.1\%$	80^{+6}_{-18}	$7.8 \pm 0.2\%$	76^{+6}_{-14}	$6.7 \pm 0.5\%$
BDRS-A	78^{+12}_{-16}	$23.2 \pm 0.1\%$	82^{+8}_{-14}	$15.9 \pm 0.3\%$	80^{+10}_{-14}	$10.0 \pm 0.6\%$
C/A-pruned	78^{+22}_{-40}	$28.9 \pm 0.1\%$	78^{+8}_{-12}	$8.0 \pm 0.2\%$	78^{+6}_{-14}	$6.5 \pm 0.5\%$
k_t-pruned	84^{+22}_{-42}	$40.7 \pm 0.2\%$	82^{+16}_{-14}	$16.9 \pm 0.3\%$	82^{+18}_{-14}	$16.4 \pm 0.7\%$

kT-pruned performs less well (larger ε_{QCD}) than C/A pruned, as expected

In largest 2 pT bins C/A pruning is (effectively) tied for most aggressive groomer, smallest ε_{QCD} (more like CMS)

What happens in the lowest pT bin?

Recall that

$\mathcal{P}^{+\mathcal{U}}{}_{-\mathcal{L}}$= mass window

Jet collection	$200 < p_T^{turth} < 350$ GeV		$350 < p_T^{truth} < 500$ GeV		$500 < p_T^{truth} < 1000$ GeV	
	$\mathcal{P}_{-L}^{+U}$ [GeV]	ϵ_{QCD}	$\mathcal{P}_{-L}^{+U}$ [GeV]	ϵ_{QCD}	$\mathcal{P}_{-L}^{+U}$ [GeV]	ϵ_{QCD}
Trimmed	82^{+10}_{-18}	$13.6 \pm 0.1\%$	80^{+8}_{-8}	$9.9 \pm 0.2\%$	82^{+10}_{-10}	$8.4 \pm 0.5\%$
BDRS	78^{+14}_{-16}	$14.8 \pm 0.1\%$	80^{+6}_{-18}	$7.8 \pm 0.2\%$	76^{+6}_{-14}	$6.7 \pm 0.5\%$
BDRS-A	78^{+12}_{-16}	$23.2 \pm 0.1\%$	82^{+8}_{-14}	$15.9 \pm 0.3\%$	80^{+10}_{-14}	$10.0 \pm 0.6\%$
C/A-pruned	78^{+22}_{-40}	$28.9 \pm 0.1\%$	78^{+8}_{-12}	$8.0 \pm 0.2\%$	78^{+6}_{-14}	$6.5 \pm 0.5\%$
k_t-pruned	84^{+22}_{-42}	$40.7 \pm 0.2\%$	82^{+16}_{-14}	$16.9 \pm 0.3\%$	82^{+18}_{-14}	$16.4 \pm 0.7\%$

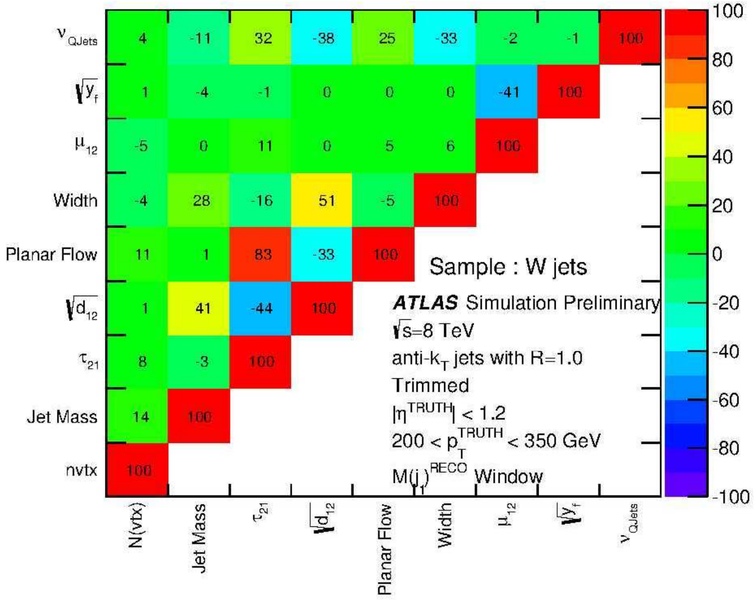
Pruned mass window for lowest pT bin is **TWICE** the size of the other mass windows, essentially **DOUBLES** the fake rate!

Why is the large mass window needed?

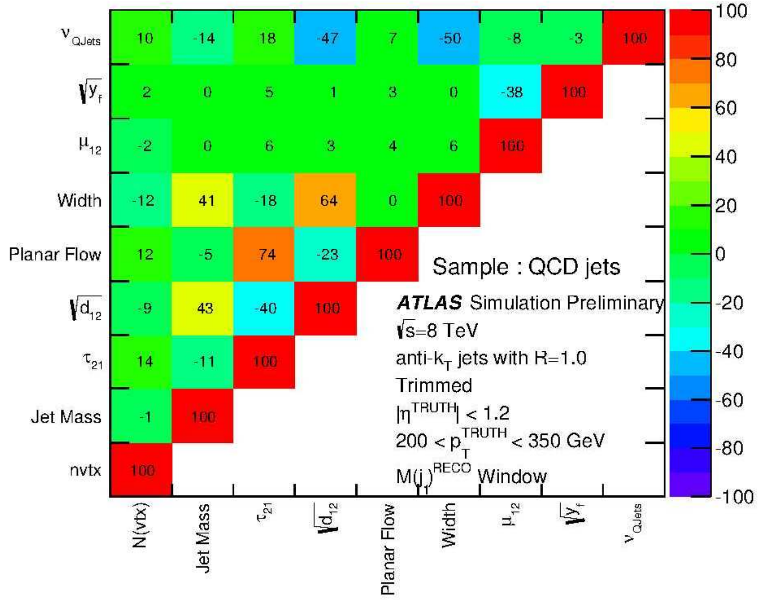
I think this is because these low pT (low boost) W's are difficult to fit in a small R (0.8) jet – see slides 5-7, i.e., we are on the low efficiency edge of the shoulder – should plot the distribution versus $(m_{particle}/p_{jet}/R)^2$ to check!

Upside of analysis - Correlations:

- Of interest to our BOOST 2013 studies is that ATLAS attempts to indicate the correlations between pairs of kinematic variables, $\rho = \mathrm{cov}(x,y)/[\sigma_x\sigma_y]$, in plots like –



(a) *W* jets, trimmed.

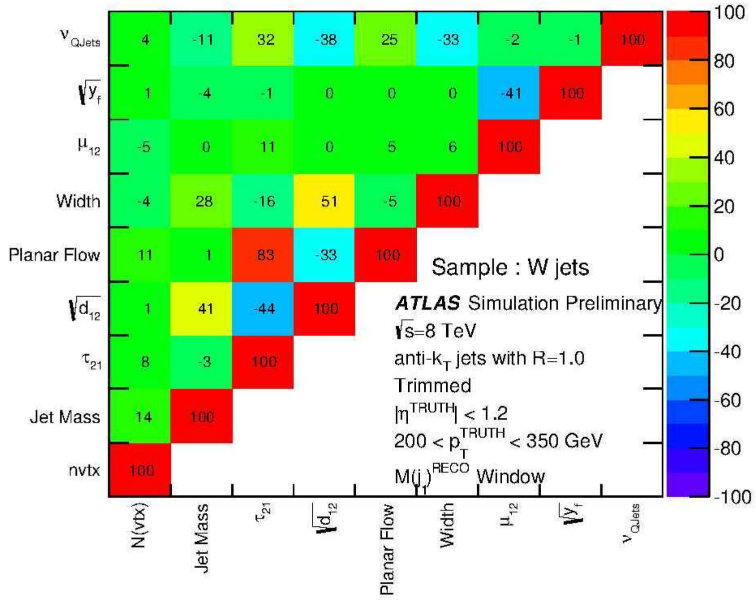


(b) QCD jets, trimmed.

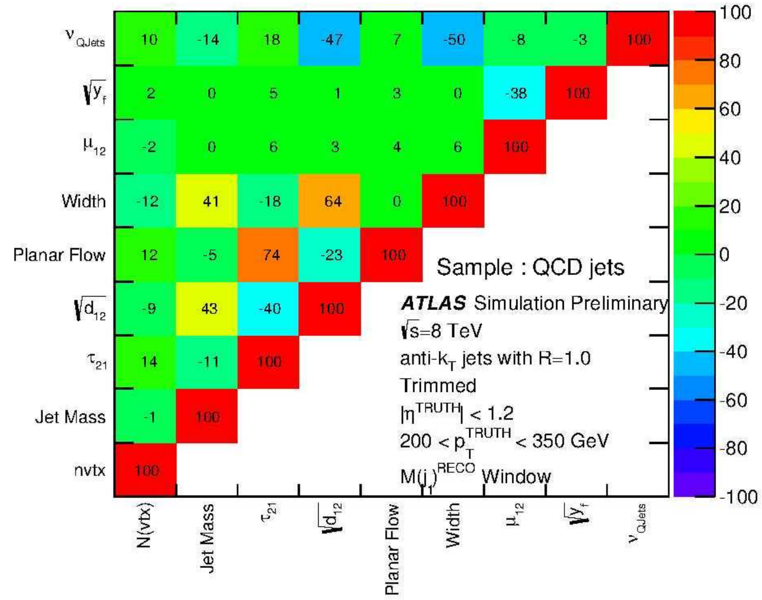
- Is this what we should think in terms of?

Correlations 2:

- Unfortunately the figures shown in ATL-PHYS-PUB-2014-004 are for the lowest pT bin, where the different R values can behave quite differently (see above)



(a) *W* jets, trimmed.



(b) QCD jets, trimmed.

Conclusions:

- Pruning is observed to be performant at ATLAS when appropriate parameter values are chosen, i.e., consistent with CMS

- Analysis of lowest pT bin needs to be clarified (boost too small for small R?)

- ATLAS study has kinematic variable correlation information, but currently difficult to interpret due to multiple “knobs” being turned at once (e.g., vary algorithm AND R values)