PU mitigation at CMS

Viola Sordini , IPNLyon

On behalf of the CMS collaboration – JetMET group

Vienna JetMET Workshop – August 2014

PU at the LHC and CMS

PU is one of the hardest challenges for LHC Run II (~40 additional PU events expected)

PU additional activity in the event :
- overlaid over the products of the hard scattering process
- additional PU jets

Handles against PU :
- Detector signal reconstruction improvements
- Exploit all the possible discriminating variables in reconstructed events

A lot of new ideas and techniques thanks to Run I data, still room for improvement.

In this talk :
- Focus on jets
- Several methods (old and new) tested and compared
- Performances for Run II (dedicated simulation)
- Crosschecks on Run I data

Outline

Performances of jets reconstruction

- Charged Hadron Subtraction (CHS)

- PU JetID [JME-13-005]

- Grooming techniques [JME-14-001]
  - Trimming
  - Pruning
  - Soft Drop/Modified Mass Drop

- Other advanced techniques [JME-14-001]
  - Constituents subtraction
  - Cleansing
  - PUPPI

Charged Hadron Subtraction

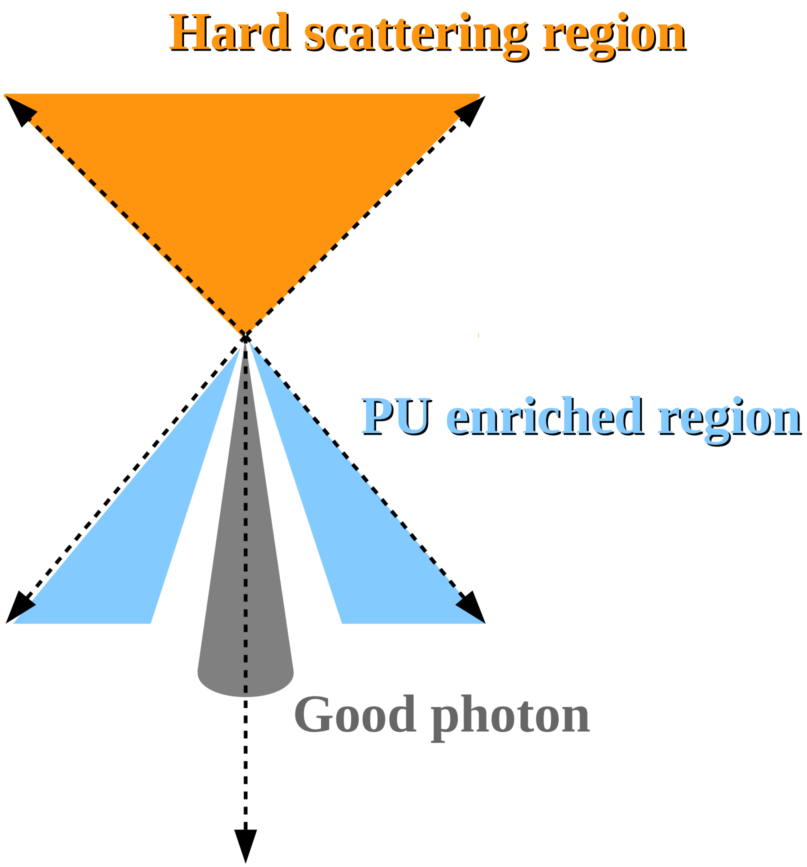
Identify all (charged) particles in the event not coming from the PV (only in tracker acceptance, $|\eta|<2.5$)
Remove them from further clustering in physics objects

Standard for RunI analyses

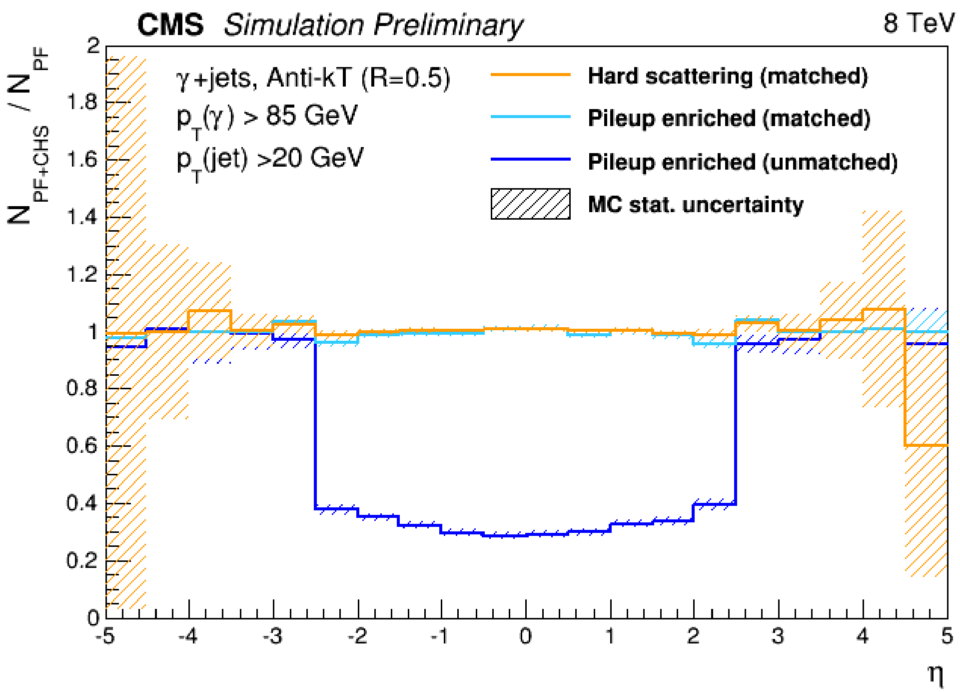
Tested on γ+jet 8TeV events simulation (PYTHIA6) and 2012 data (<PU>~20)

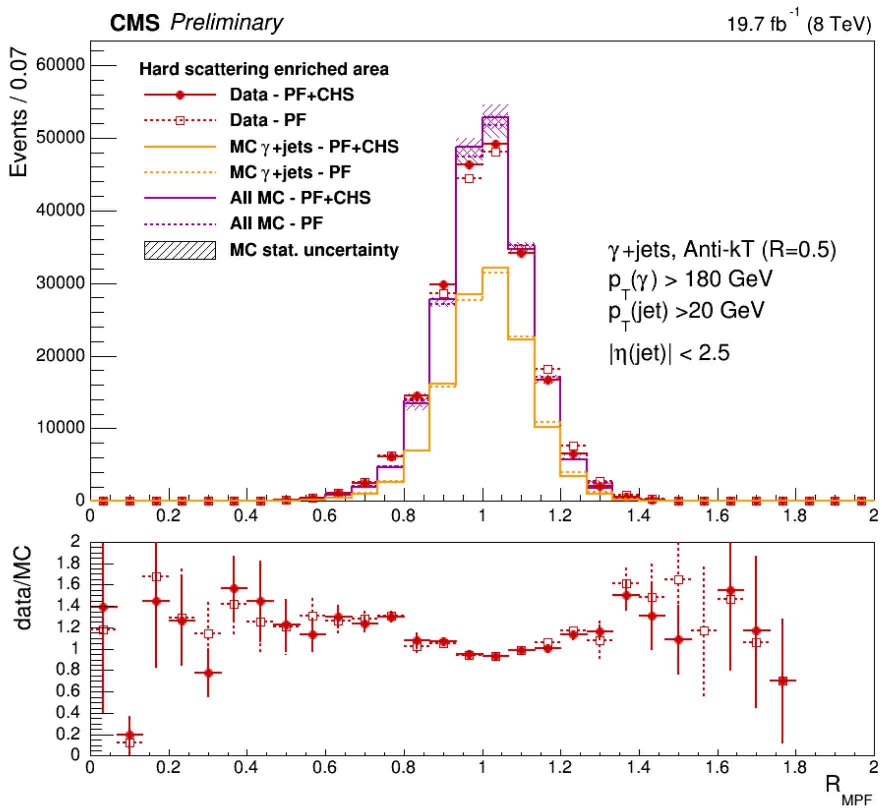
MC definition :
Jet is «**good**» (not PU) if **matched** to a generated jet (p_T>10, ΔR<0.25)
➔ Unmatched jets PU enriched

Event selection :
Hard scattering region ($|\Delta\varphi(\text{jet},\gamma)|>3$)
PU enriched region ($|\Delta\varphi(\text{jet},\gamma)|<1$)



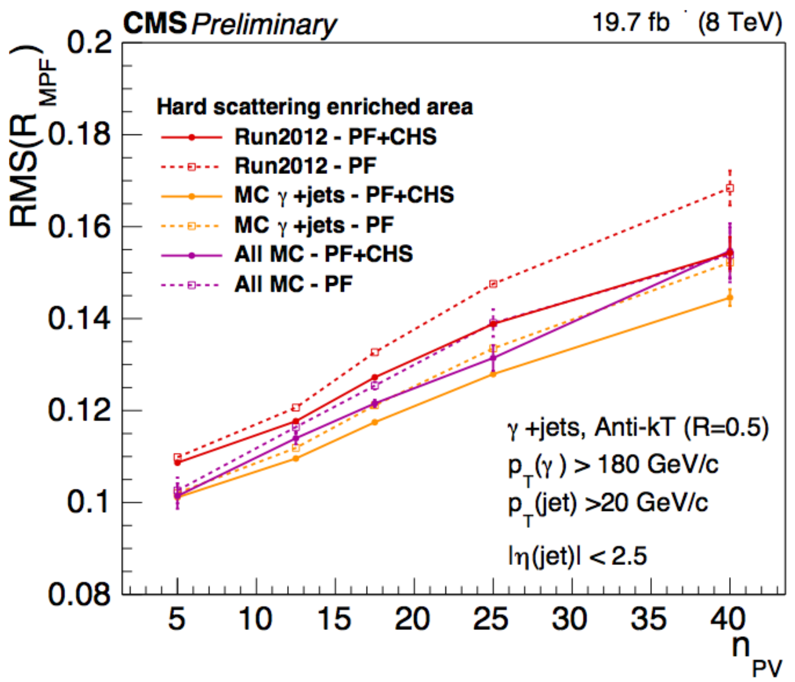
Charged Hadron Subtraction

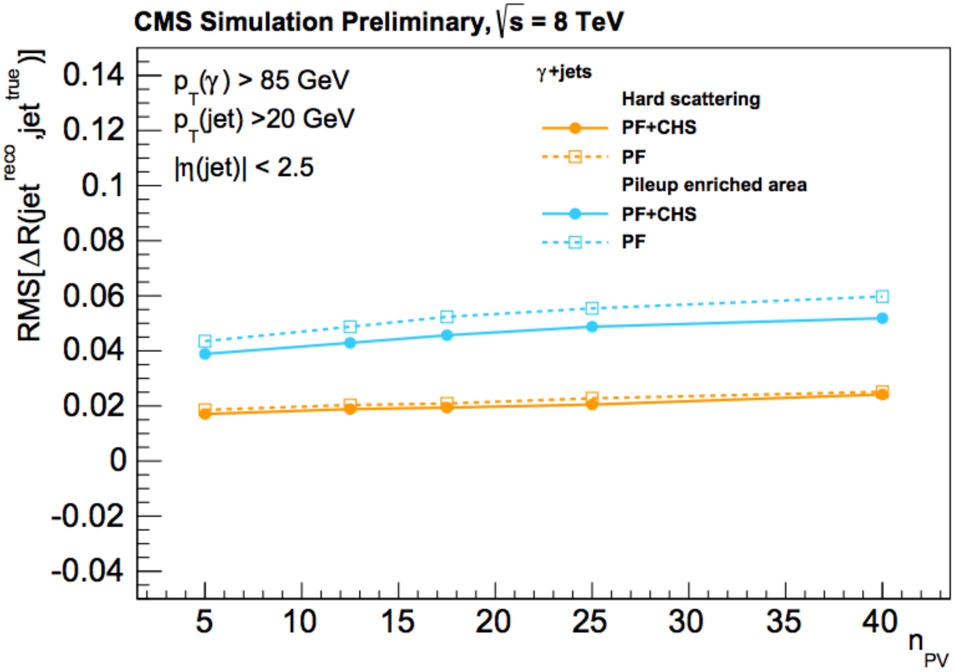




- x3 PU rate reduction (|η|<2.5)
- Almost no effect on hard scattering high p_T jets
- Reduces the rates of purely PU jets at low pt

Charged Hadron Subtraction





- Reduces effect of PU on real jets, improving p_T and angular resolution
- Caveat : PU enriched area different p_T spectrum
- Small bias effect at the tracker edge affect mostly low p_T jets

PU JetID

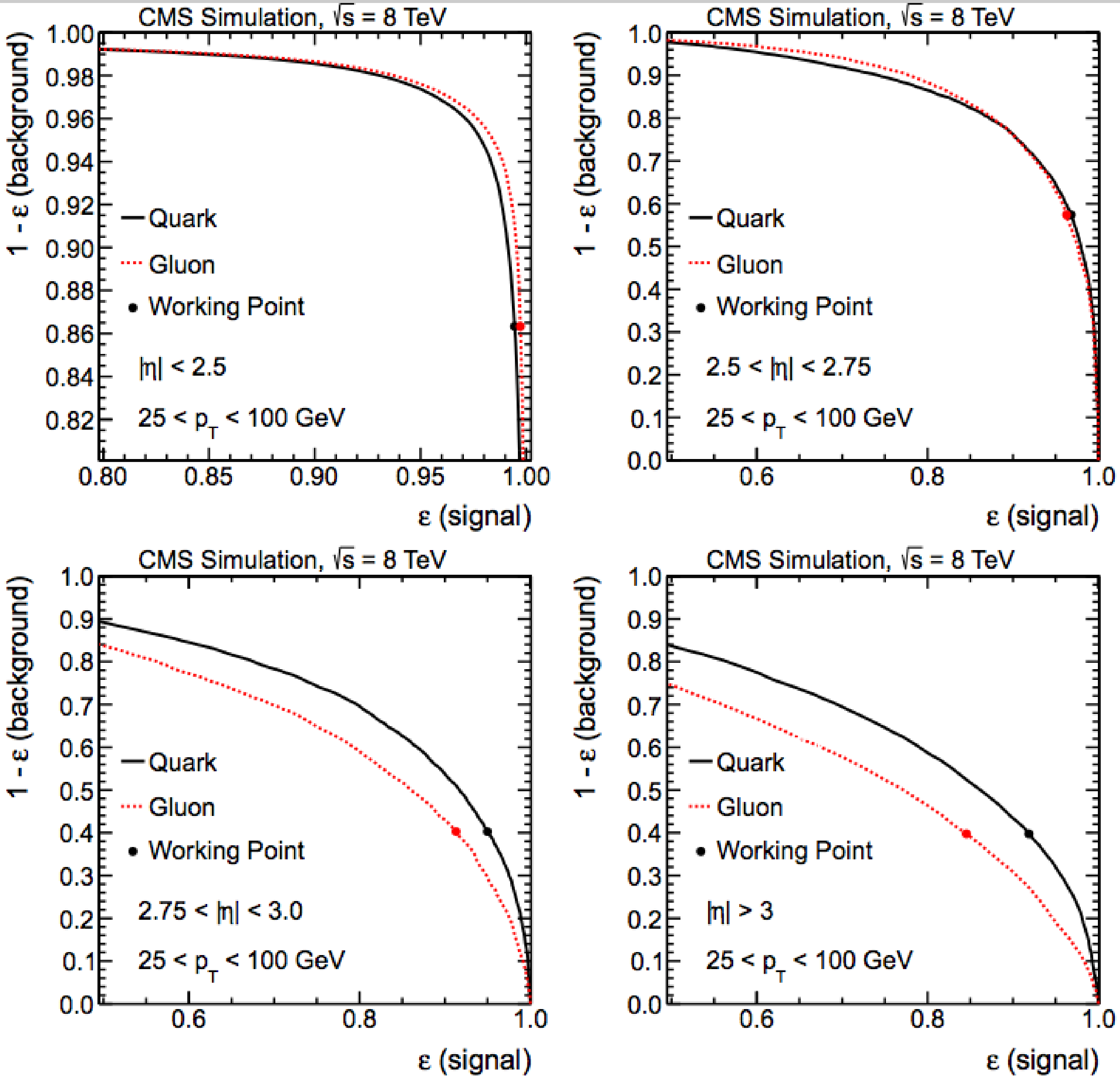
PU JetID tags the jets entirely coming from PU :
• charged constituents not pointing to the PV – highly discriminant, but only available in the tracker acceptance
• consituents more diffuse – extend the discrimination power to the whole detector

12 variables combined in a BDT :
• 4 vertexing related
• 8 shape related

Jet is **good** (not PU) if **matched** to a generator level jet (p_T>8 GeV , ΔR<0.25)

Performances tested on Z → μμ events

PUJetID : performances



Central region:
signal eff ~99%
bkg rej 90-95% (30<pT<50)
85% (20<pT<30)

Endcap
Signal eff 95%
Bkg rej 70%(60%)

Fwd :
Sig eff 90%(80%)
Bkg rej 60%(40%)

Gluons :
- Higher multiplicities
- Wider, more uniform energy spread

Grooming

Systematic removal of jet constituents
Typically used to distinguish fat heavy jets from qcd ones
Reduces PU dependence of jet mass
Studied on fat jets (R=0.8)

In this study :

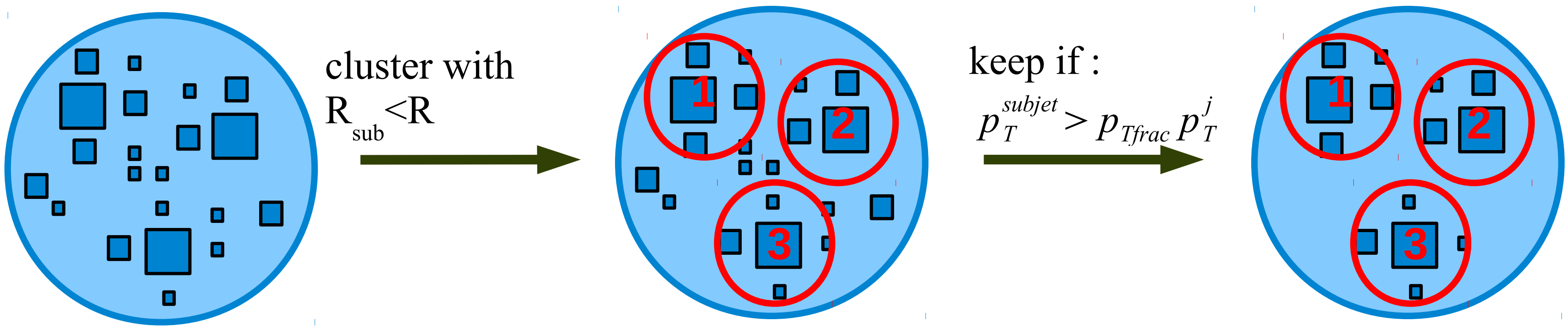
- Trimming
- Pruning
- Soft drop/Modified mass drop tagger

Evaluate the performances on simulation (13 TeV, <PU>=40, 50ns)

- Multijet (background) and RS graviton → WW (signal) (PYTHIA8)
- Criteria for comparison :
  - Stability wrt PU
  - Jet mass reconstruction response and resolution

Check the data/simulation agreement on 8 TeV collisions

Trimming

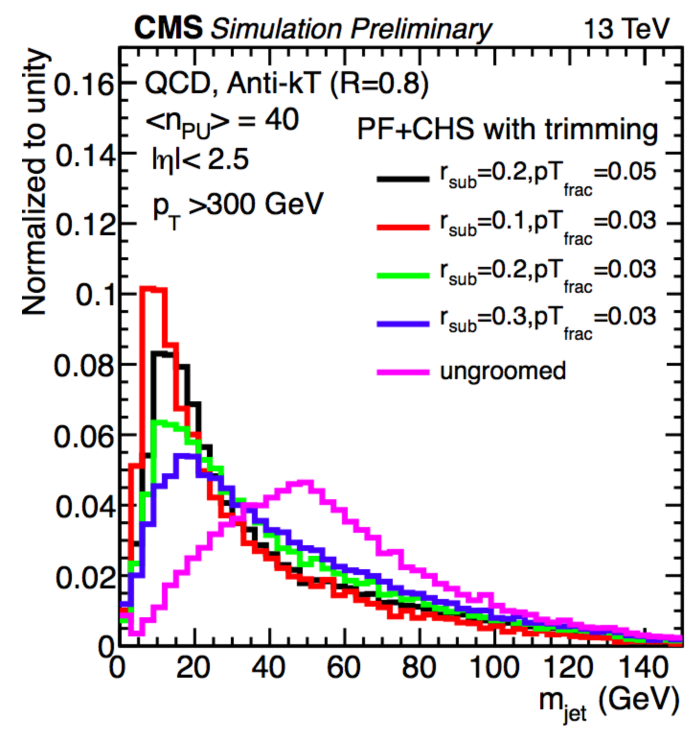


Keeps subjets over a dynamic pT threshold

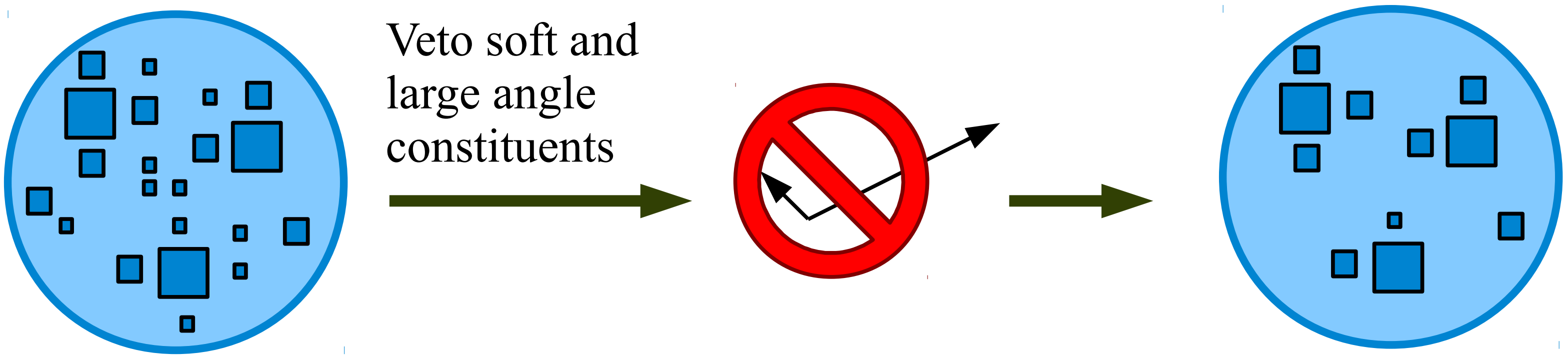
- Reclusters constituents with anti-kT into subjects (R_{sub})
- Keep the constituents if :

$$p_T^{subjet} > p_{Tfrac}\, p_T^{j}$$

Parameters :
R_{sub}, f_{cut}



Pruning

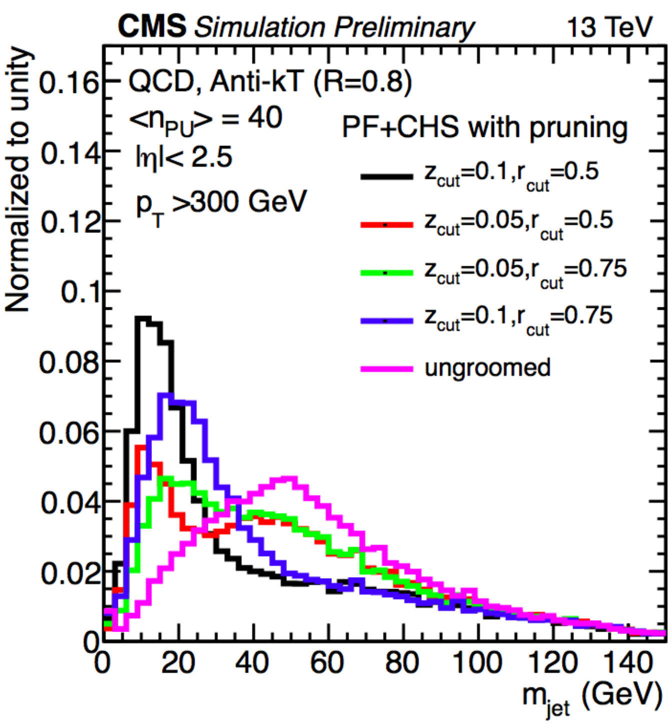


- Reclusters the constituents with CA
- At each step, the softer of the two particles

i and j is removed if :

$$\begin{cases} z_{ij} = \dfrac{min\{p_T^i, p_T^j\}}{p_T^i + p_T^j} < z_{cut} \\ \Delta R_{ij} > D_{cut} = \dfrac{2\, r_{cut}\, m^{jet}}{p_T^{jet}} \end{cases}$$

Parameters :
z_{cut}, r_{cut}



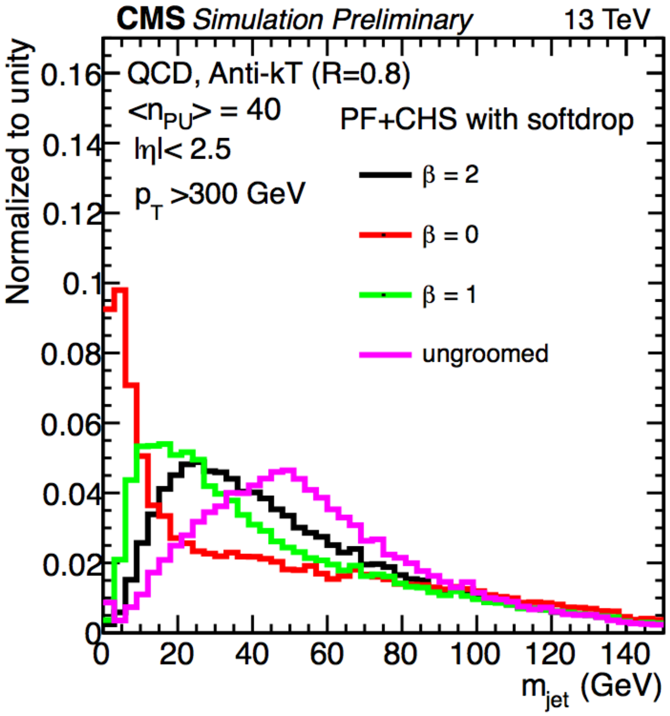
Soft drop/MMDT

- Jet is clustered with CA algorithm with distance R
- It is then declustered and, at each step, subjets j1 and j2 are defined.
- If the **condition** :

$$\frac{min\{p_T^{j1}, p_T^{j2}\}}{p_T^{j1}+p_T^{j2}} > z_{cut}\left(\frac{\Delta R_{12}}{R}\right)^{\beta}$$

➔ is **met** : the declustering of j1 and j2 continues
➔ is **NOT met** : only the leading pT subjet is kept for further declustering

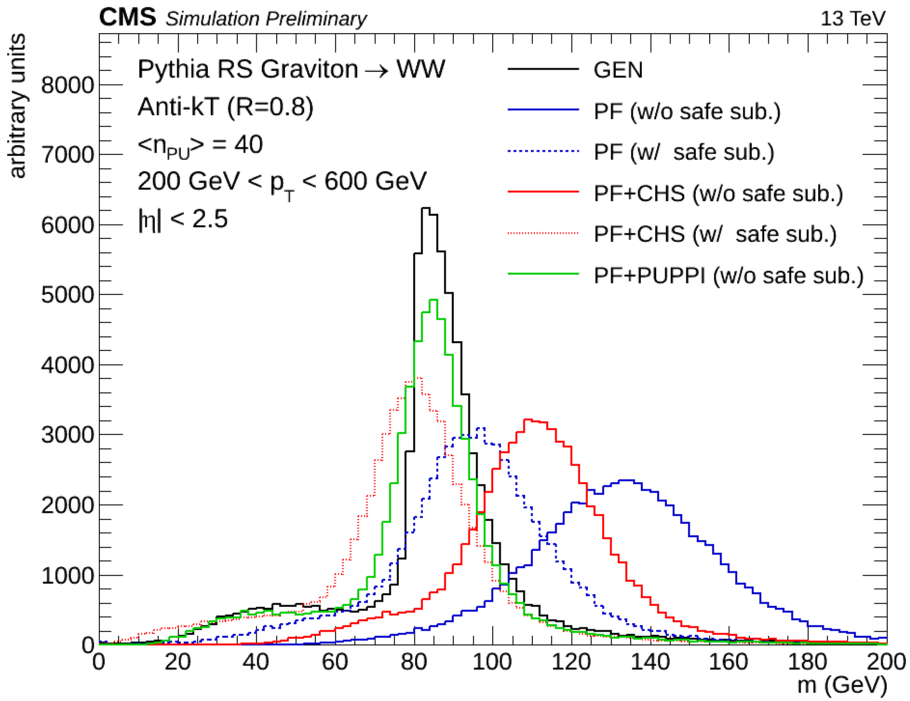
Parameters :
z_{cut}, β



Grooming : samples and selection

Several grooming algorithms considered, with different parameters (more or less aggressive) :

grooming algorithm	parameter(s)
Pruning	$z_{cut} = 0.1, r_{cut} = 0.5$ $z_{cut} = 0.05, r_{cut} = 0.5$ $z_{cut} = 0.1, r_{cut} = 0.75$ $z_{cut} = 0.05, r_{cut} = 0.75$
Trimming	$r_{sub} = 0.2, p_{T\ frac} = 0.05$ $r_{sub} = 0.2, p_{T\ frac} = 0.03$ $r_{sub} = 0.1, p_{T\ frac} = 0.03$ $r_{sub} = 0.3, p_{T\ frac} = 0.03$
Soft drop/MMDT	$z_{cut} = 0.1, \beta = 0$ $z_{cut} = 0.1, \beta = 1$ $z_{cut} = 0.1, \beta = 2$



Performances evaluated on simulation
- Multijet (background) and RS graviton → WW (signal)
- Dijet topology, leading jet pT>300 GeV, |eta|<2.5
- Using PF jets with and without CHS

All groomed jets are corrected for PU using a 4-vector safe subtraction

$$p^{\mu}_{sub} = p^{\mu} - \rho A^{\mu} - \rho_m A^{\mu}_m$$

Grooming : samples and selection

For the different algorithms, we will look at mainly two variables :

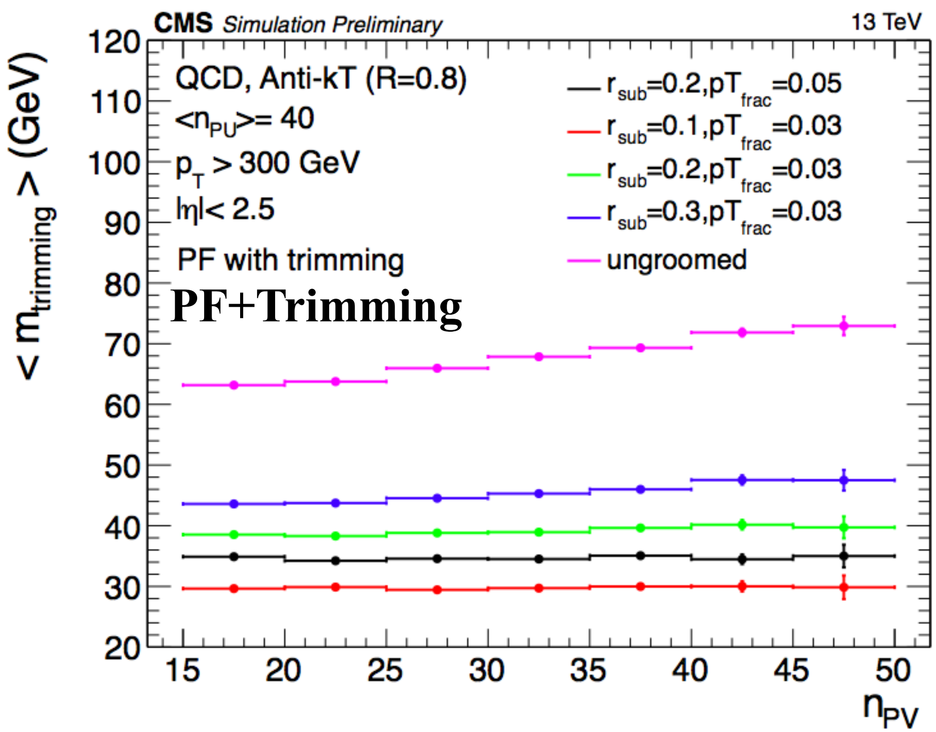
➔ Average reconstructed jet mass for **QCD jets** as a function of nPV, to monitor the stability VS PU

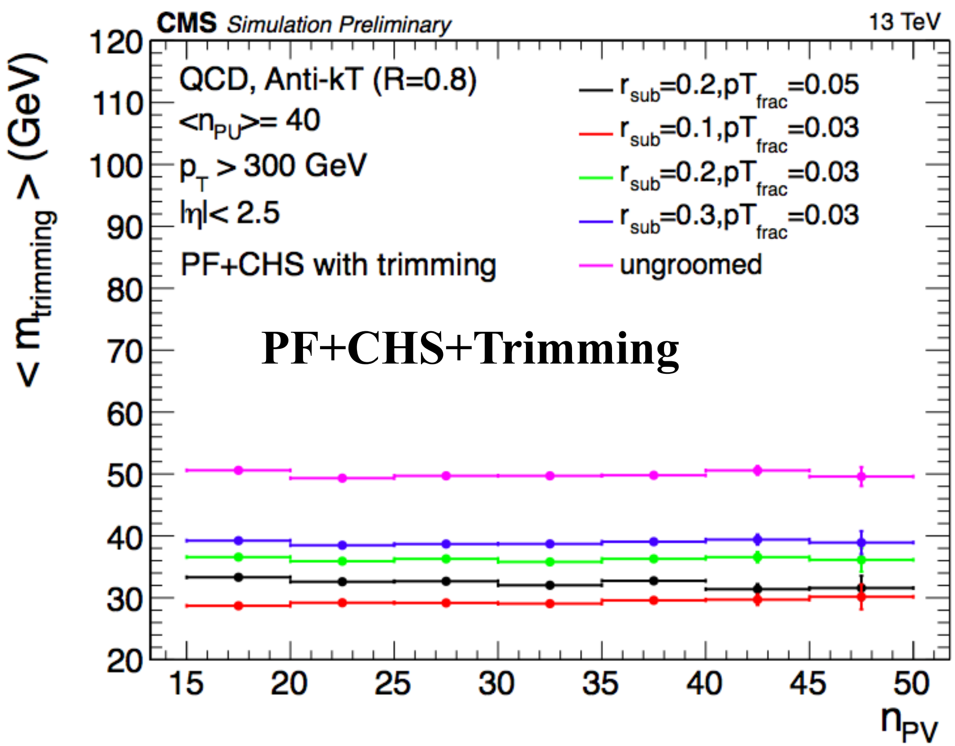
➔ W peak mass resolution for **W jets** from RS graviton → WW sample :

$m_{RECO}-m_{GEN}$

Both **RMS** (sensitive to the whole distribution, including tails) and **σ from a Gaussian fit** (sensitive to the bulk of the distribution).
Also monitored VS nPV.

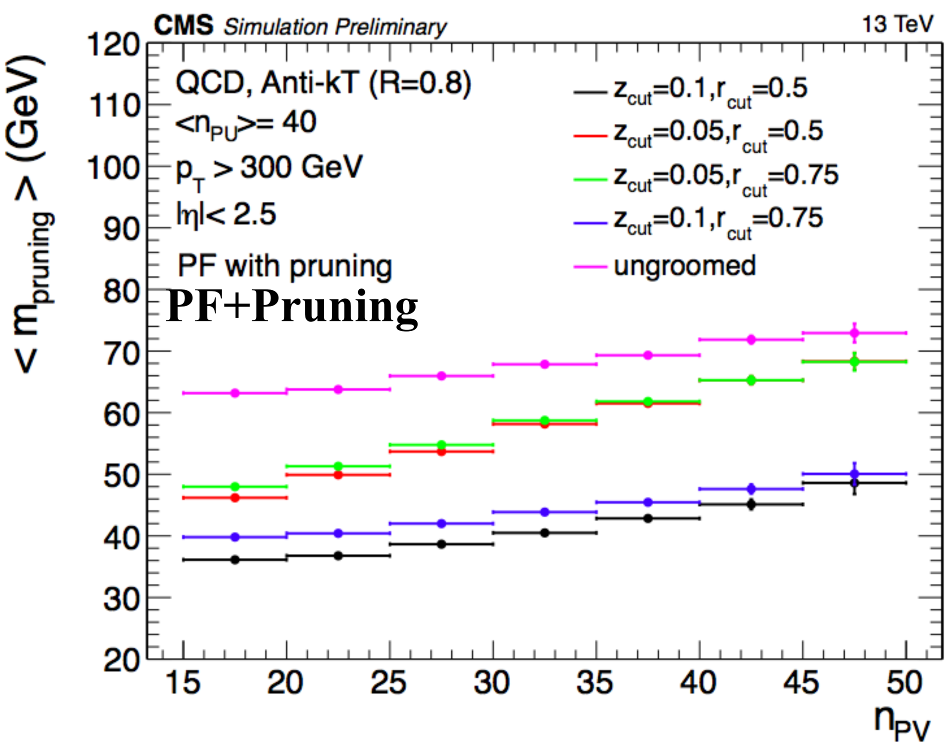
Trimming: stability VS PU

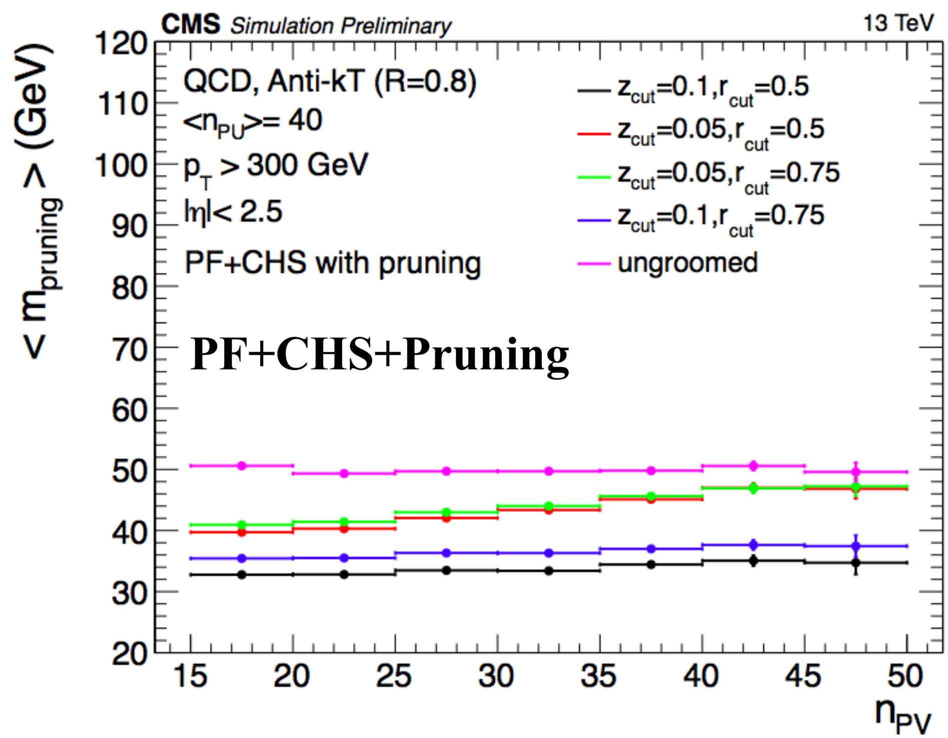




Sensible improvement wrt ungroomed
Average mass is quite stable wrt to PU
Not a big difference with and without CHS (slight improvement)

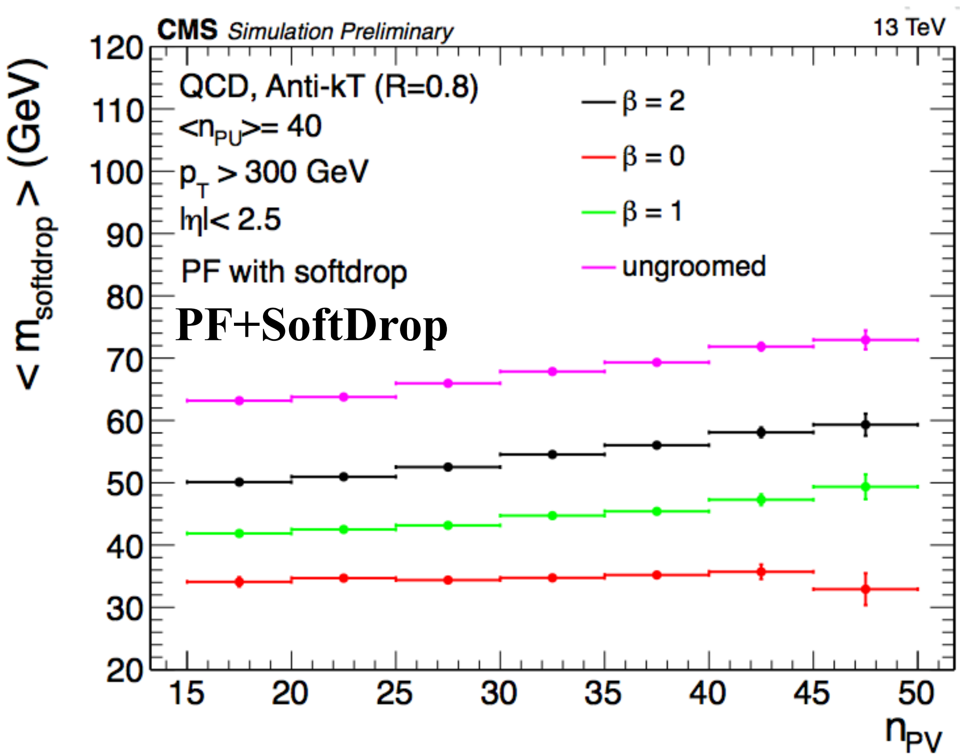
Pruning: stability VS PU

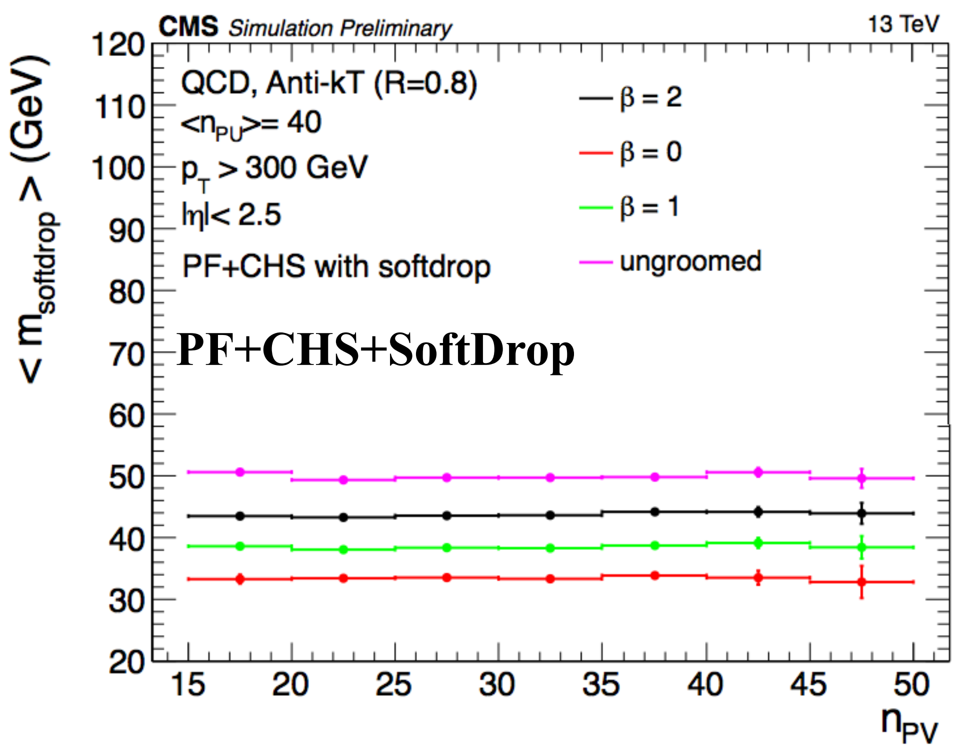




Residual dependence on PU for the average mass
Visible improvement in stability with the use of CHS

Soft Drop: stability VS PU

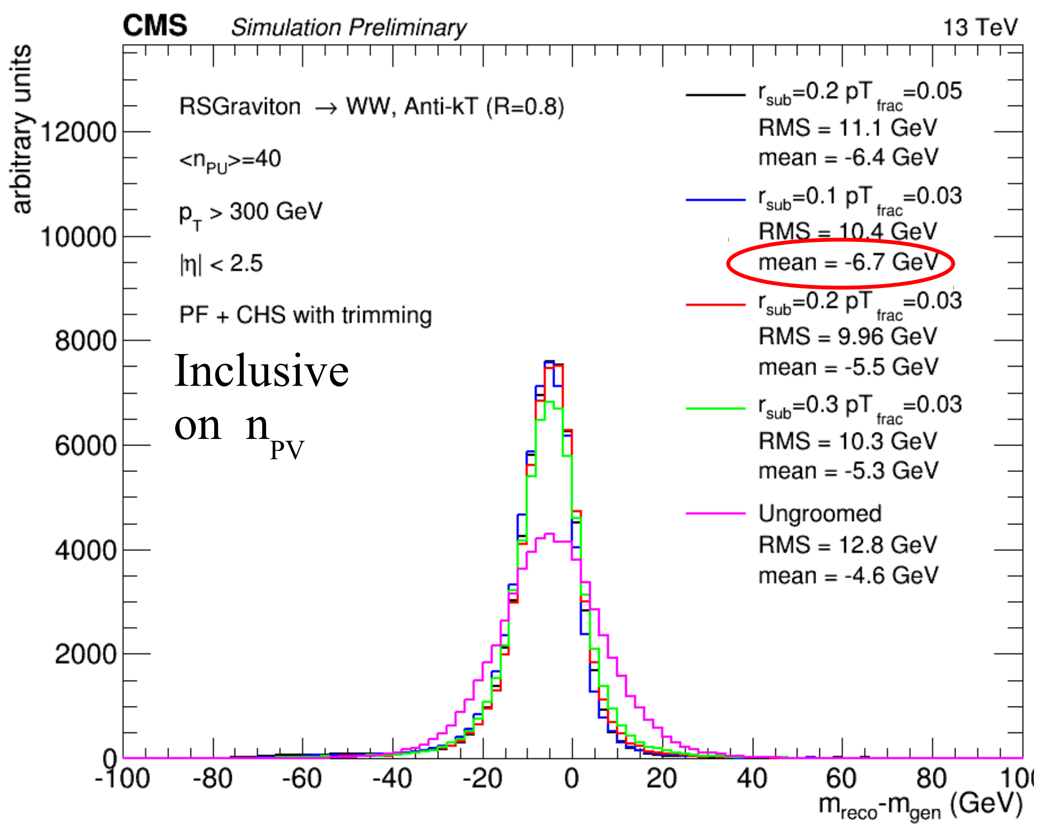


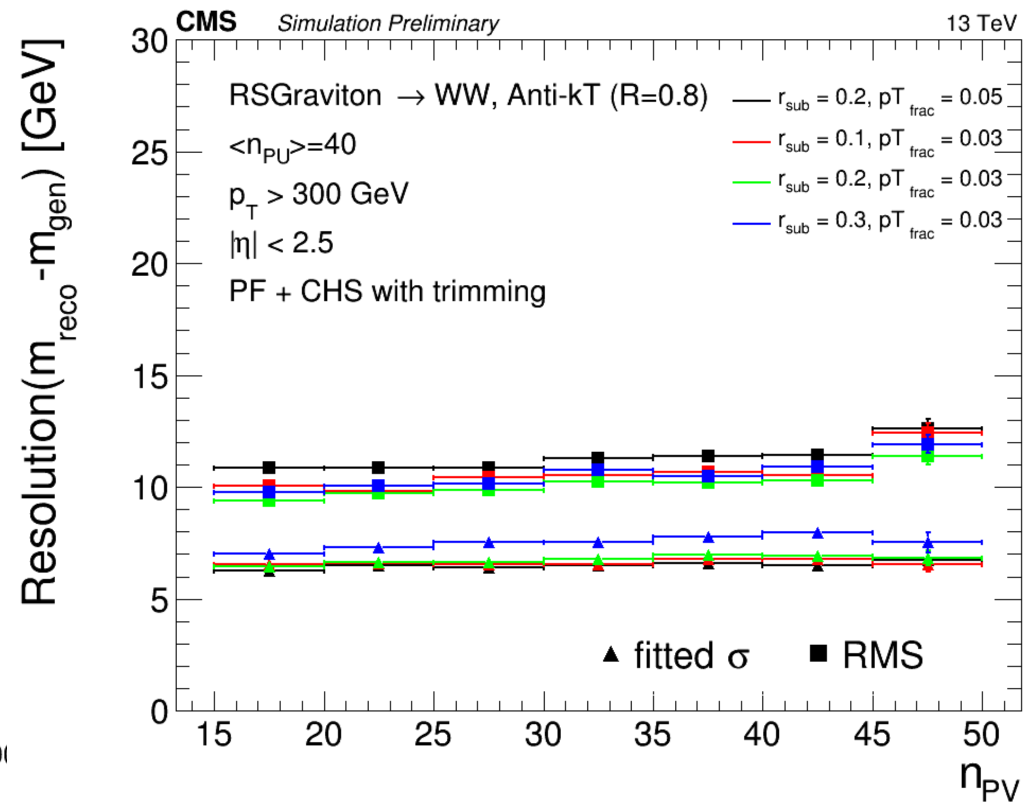


Average mass is quite stable wrt to PU
Not a huge difference with and without CHS, sill an improvement is clearly visible

Trimming : mass resolution

W jet mass resolution (m_{RECO}-m_{GEN}) on a RS graviton → WW simulated sample.
Gaussian fit σ : bulk of the distribution
RMS : tails

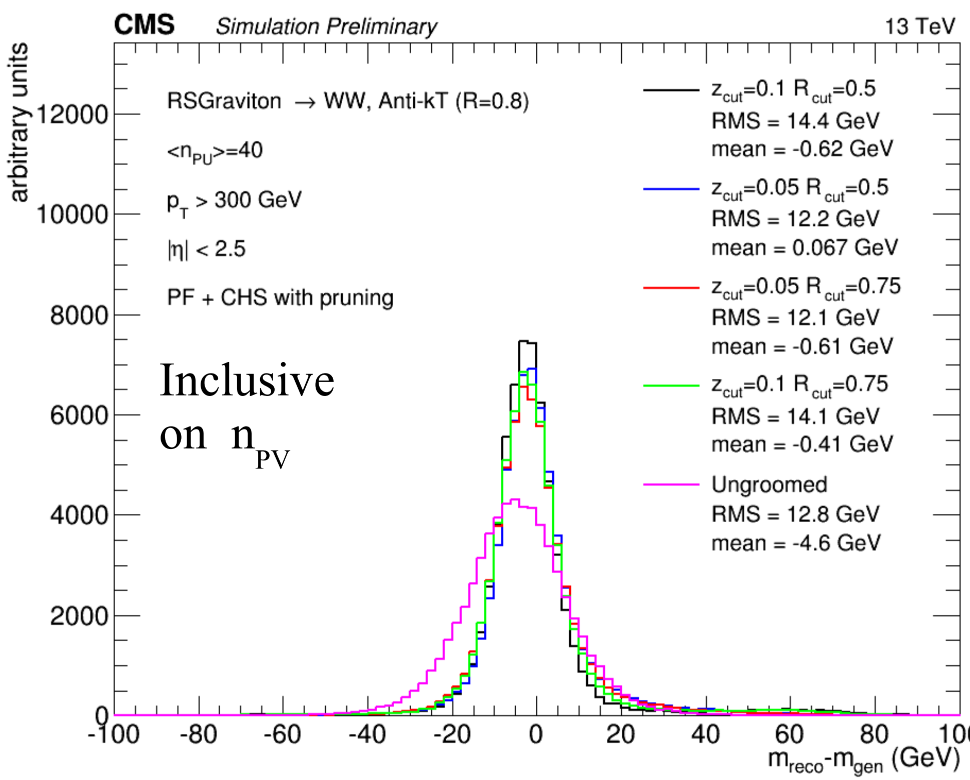


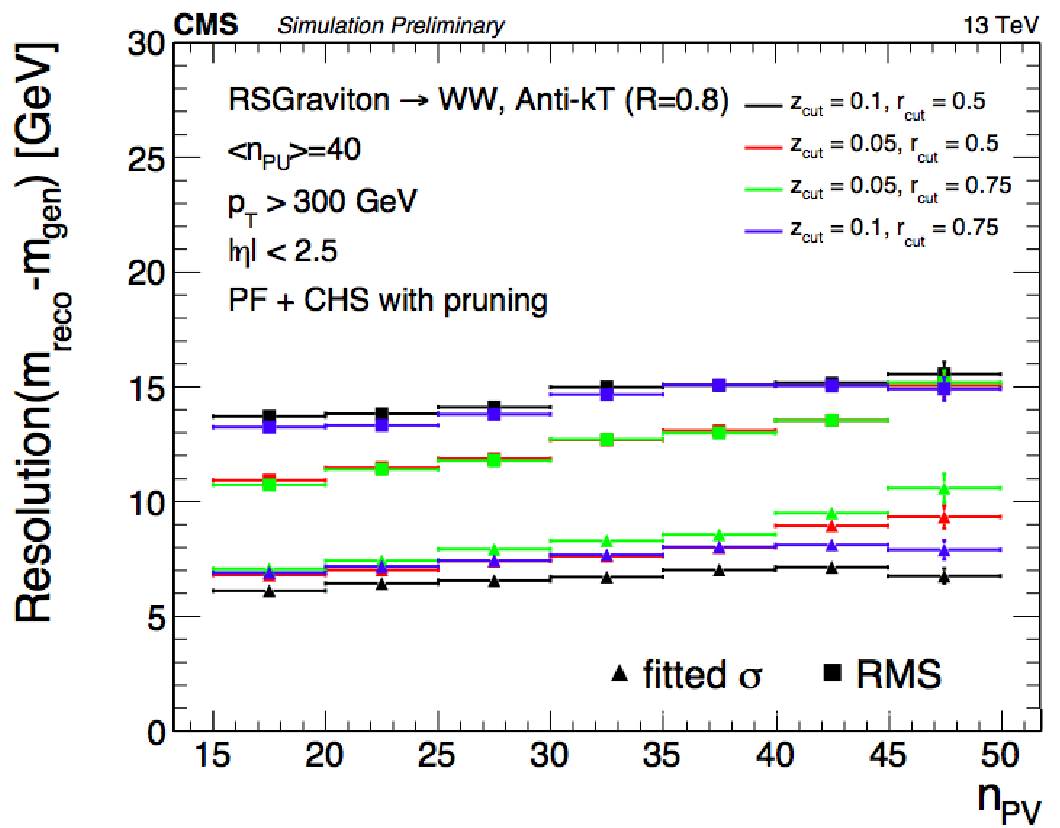


Good mass resolution (6-8 GeV in the bulk), quite stable vs PU
Residual offset in the mass mean

Pruning: mass resolution

W jet mass resolution (m_{RECO}-m_{GEN}) on a RS graviton → WW simulated sample.
Gaussian fit σ : bulk of the distribution
RMS : tails

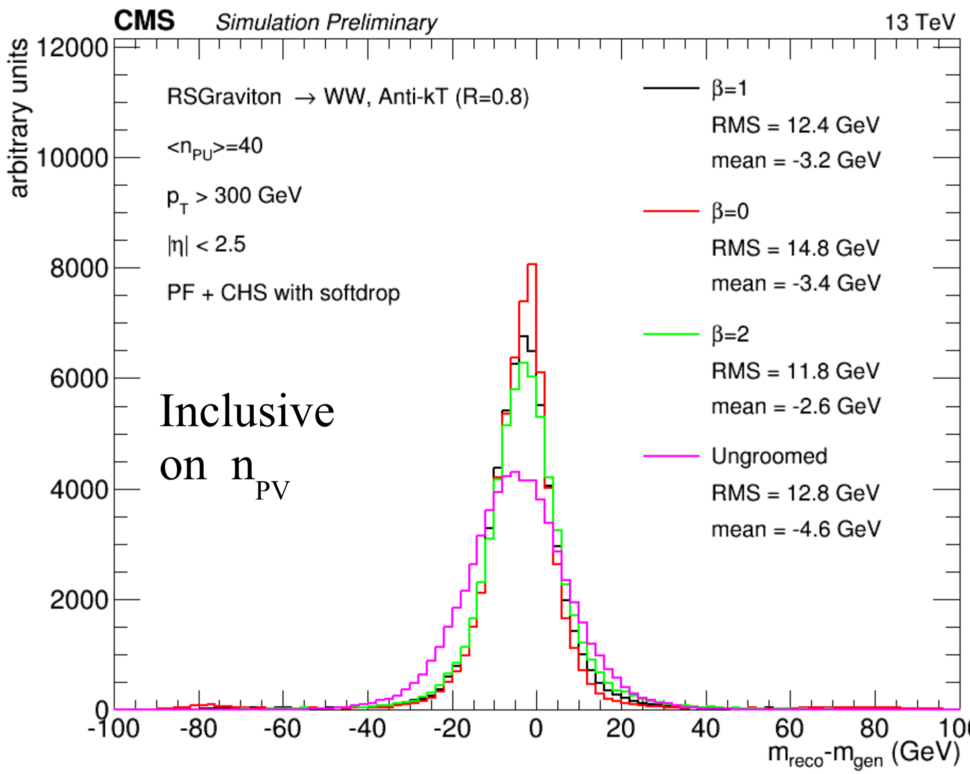


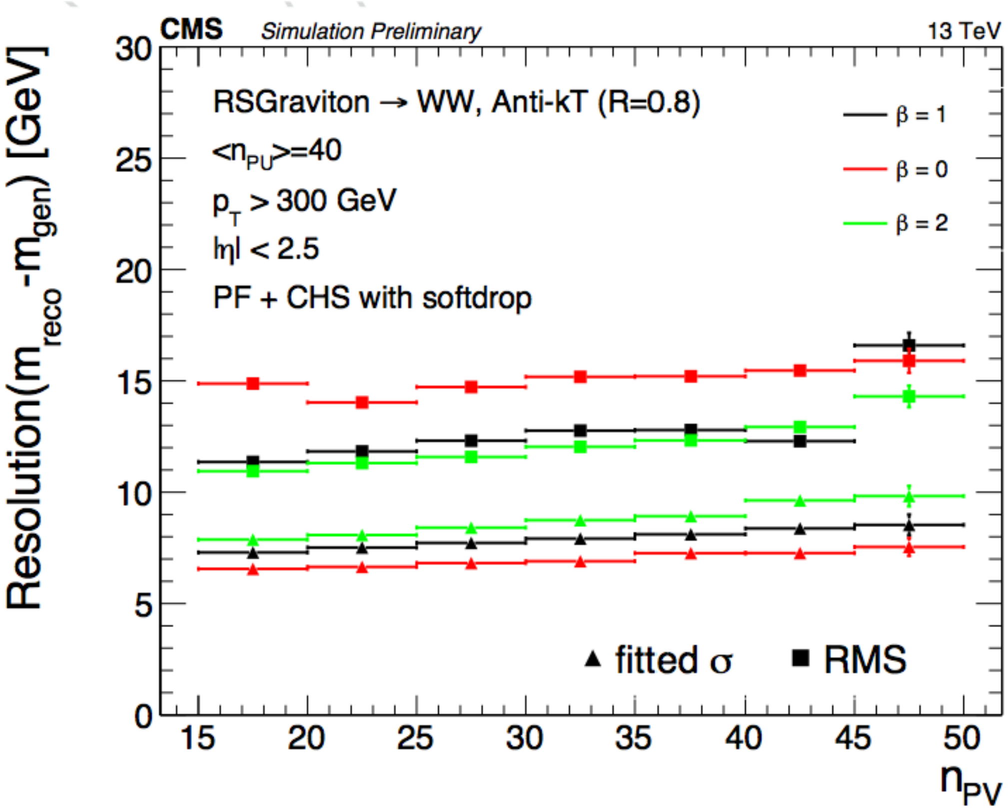


Mass resolution (6-11 GeV in the bulk) shows a dependence on PU
Visible presence of tails
Caveat : 4-vector safe subtraction performed on the final pruned jet

Soft drop: mass resolution

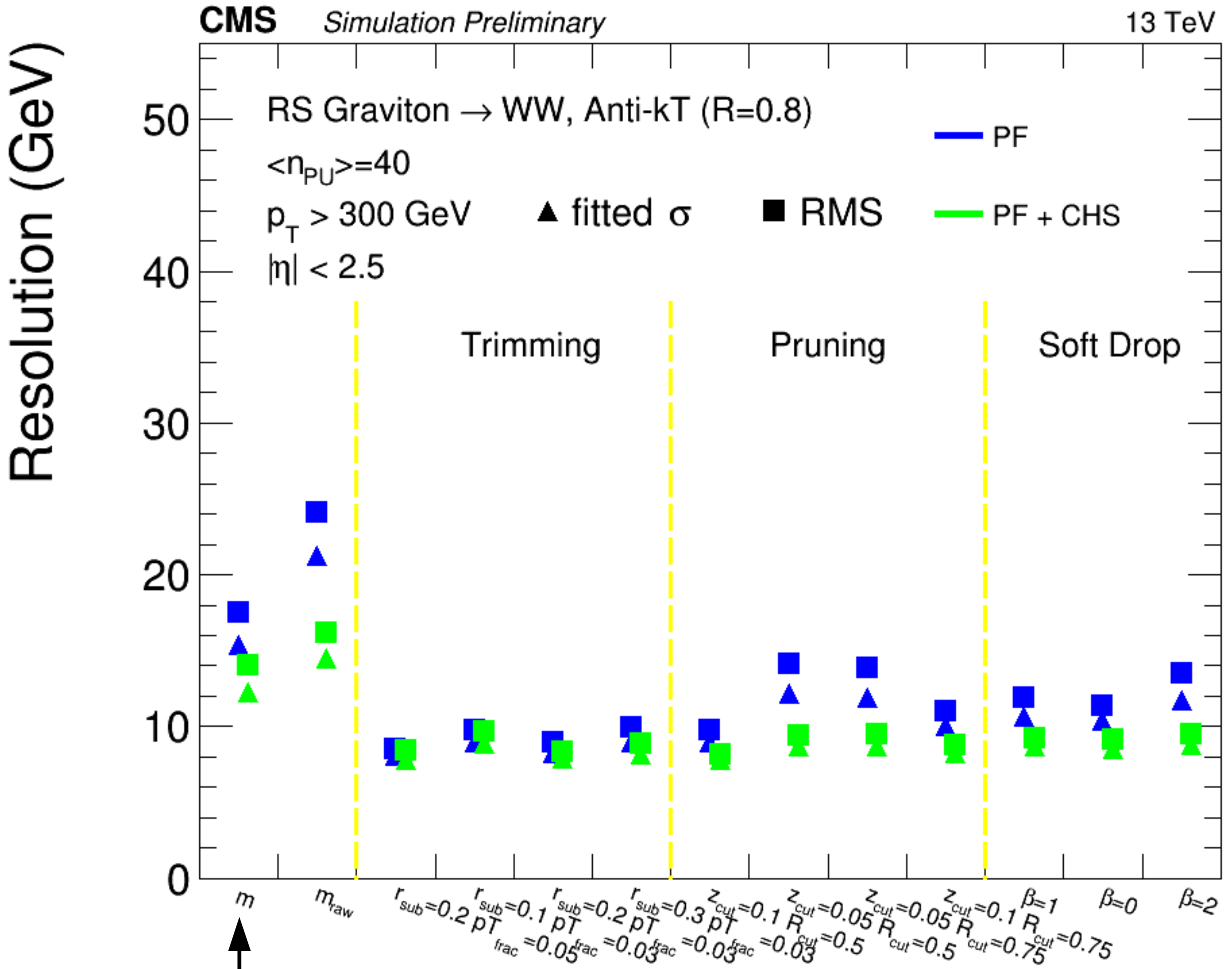
W jet mass resolution (m_{RECO}-m_{GEN}) on a RS graviton → WW simulated sample.
Gaussian fit σ : bulk of the distribution
RMS : tails





Good mass resolution (7-10 GeV in the bulk), residual dependence on PU
Residual presence of tails

Grooming – jet mass resolution



With 4-vector
safe subtraction

W jet mass resolution comparison

- Value of σ from fit in ±RMS range
- RMS truncated in ±3σ range

- PF+CHS+**Grooming**

➔ **improves resolution** wrt to PF (6-11 GeV in the bulk, depending on the algo and parameters)

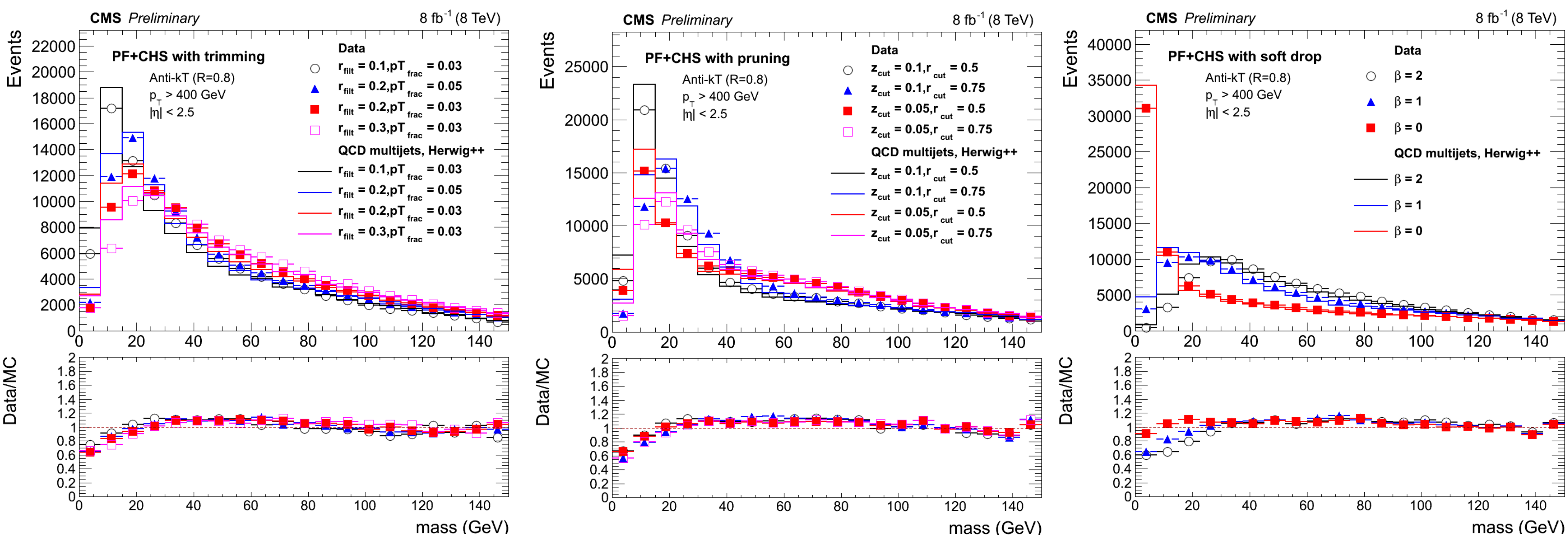
➔ **improves stability VS PU**

Grooming – data/MC comparisons

Use of 8 TeV data (8fb^{-1} from late RunI, <PU>~22) to evaluate how well our simulation describes the data for goomed jets

Basic dijet selection to target a region interesting for resonance searches

→ at least one jet with p_T>400 GeV, m_{jj}>900 GeV, $|\Delta\eta_{jj}|$<1.2



Overall reasonable agreement

Largest disagreement (up to ~40%) low jet mass region - non-perturbative effects hard to simulate

Other PU mitigation techniques

We explore additional PU mitigation techniques :

- Constituents subtraction [arXiv:1403.3108]
- Cleansing [arXiv:1309.4777]
- PUPPI [arXiv:1407.6013]

First explorative look, algorithms can still be tuned/optimised !

Evaluate the performances on simulation (13 TeV, <PU>=40, 50ns)

- Use anti-kT jets, R=0.8, 4-vector corrected
- Leading jet (p_T in [200, 600]) in Multijet and RS graviton → WW
- Using PF jets with and without CHS
- Look at mass, N-subjettiness

In addition, only for PUPPI

- Look at groomed mass
- Check the data/simulation agreement on 8 TeV collisions

Constituents subtraction - Jet cleansing

Constituents subtraction [arXiv:1403.3108] :
Natural extension of area based subtraction to jet constituents (PF particles).
Applied on top of PF+CHS

Jet Cleansing [arXiv:1309.4777]:
Uses vertex information to determine the charged PU contribution, then uses jet composition to evaluate the neutral PU contribution
Decompose jet into subjets (here R_{subjet}=0.2)
For each subjets inputs to the cleansing method are p_T^{tot}, $p_T^{C,PU}$, $p_T^{C,LV}$

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

Different schemes for γ_0 and γ_1, here we use the linear cleansing (γ_0 is constant and can be determined from MinBias data, γ_1 depends on γ_0)

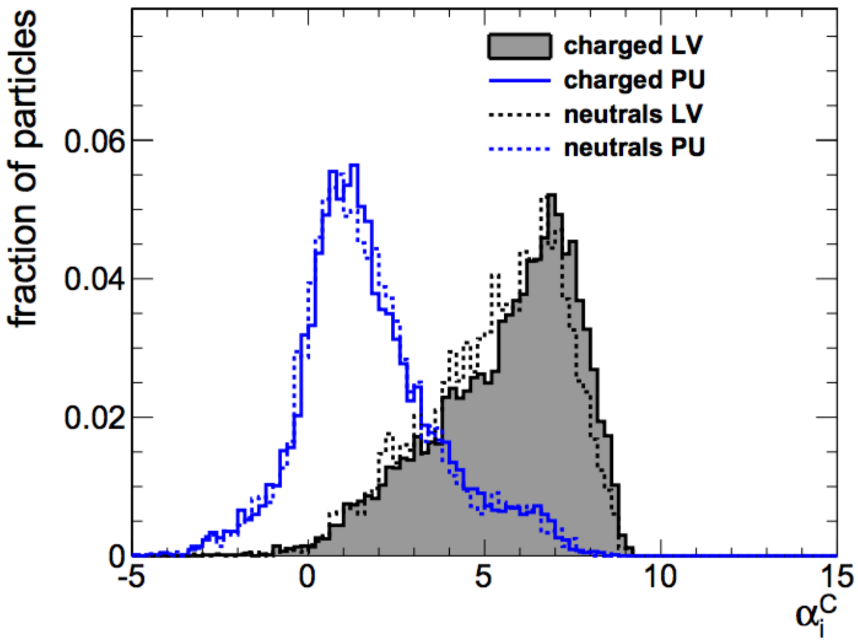
Pileup Per Particle Identification

Operates on the inputs to jet clustering (here PF candidate particles) [arXiv:1407.6013]

- A discriminating variable α is defined :

$$\alpha_i = \log \sum_{j \in Ch,PV} \left(\frac{p_{T,j}}{\Delta R_{ij}} \right)^2 \Theta(R_0 - \Delta R_{ij}) \quad \text{for } |\eta| < 2.5$$

$$\begin{cases} \alpha_i = \log \sum \frac{p_{\mathrm{T,j}}}{\Delta R_{ij}} \Theta(R_0 - \Delta R_{ij}) \\ \alpha_i = \log \sum p_{T,j} \Theta(R_0 - \Delta R_{ij}) \end{cases} \quad \text{for } |\eta| \geq 2.5$$



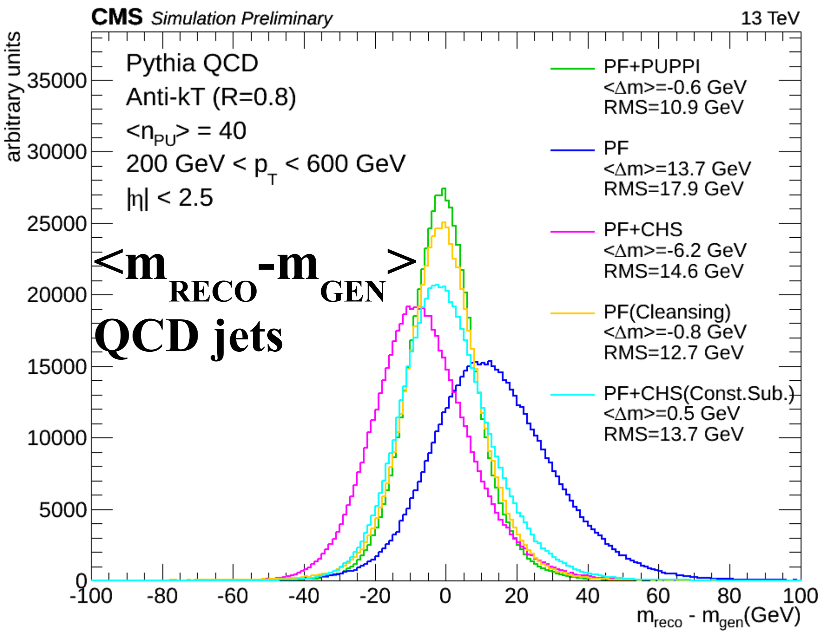
- The distribution for charged PU particles is used as template for the distribution for all PU particles

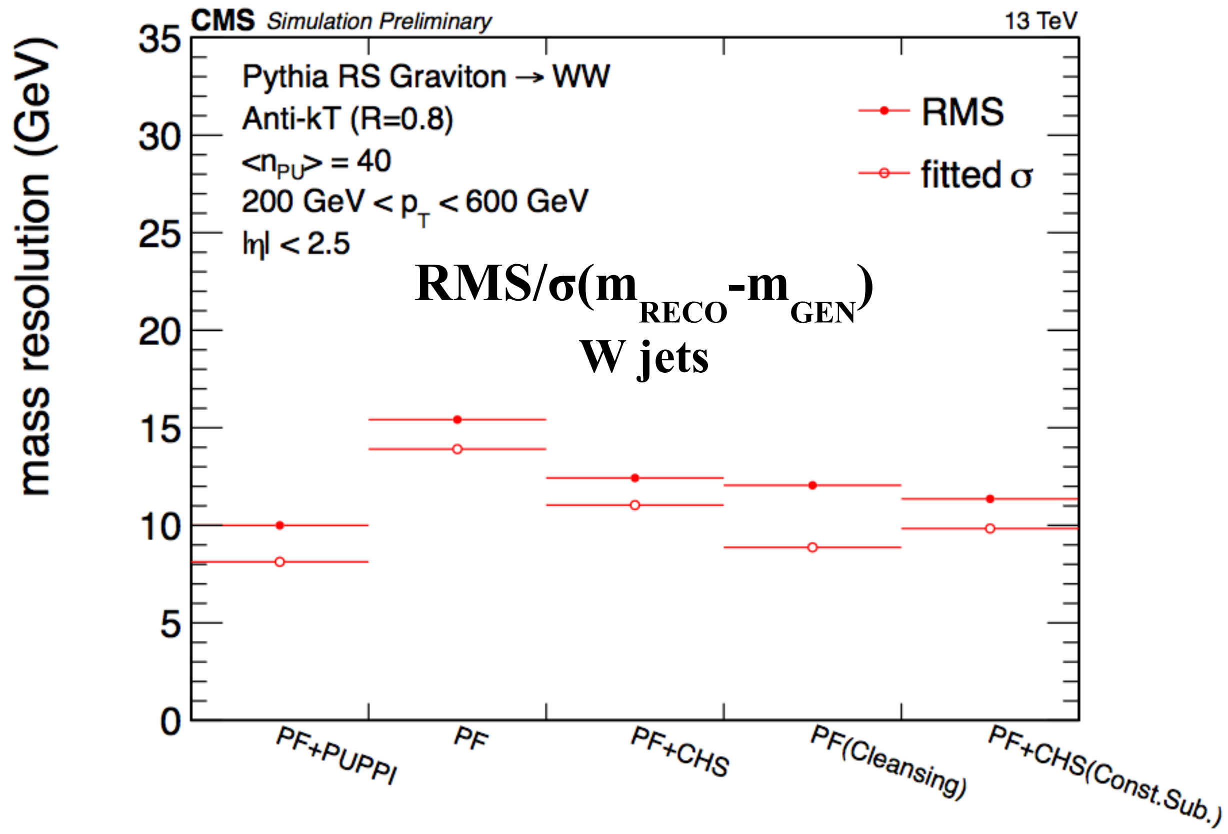
- For each neutral particle, a χ^2 variable is constructed (for $|\eta| \geq 2.5$, sum the two χ^2)

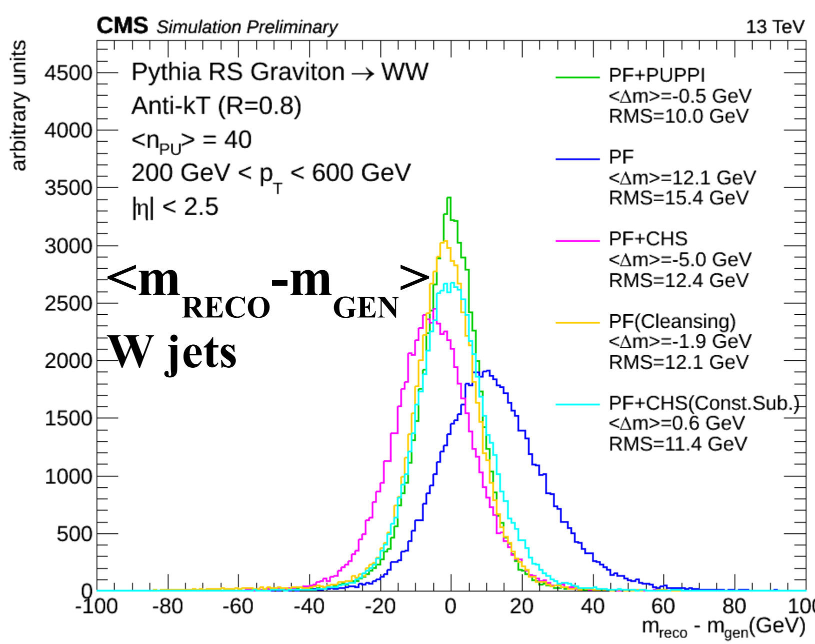
$$\chi_i^2 = \frac{(\alpha_i - \bar{\alpha}_{PU})^2}{RMS_{PU}^2}$$

- The probability of the particle to come from the Leading Vertex is calculated
- The particle 4-momentum is reweighted by this probability

Performances of mass reconstruction



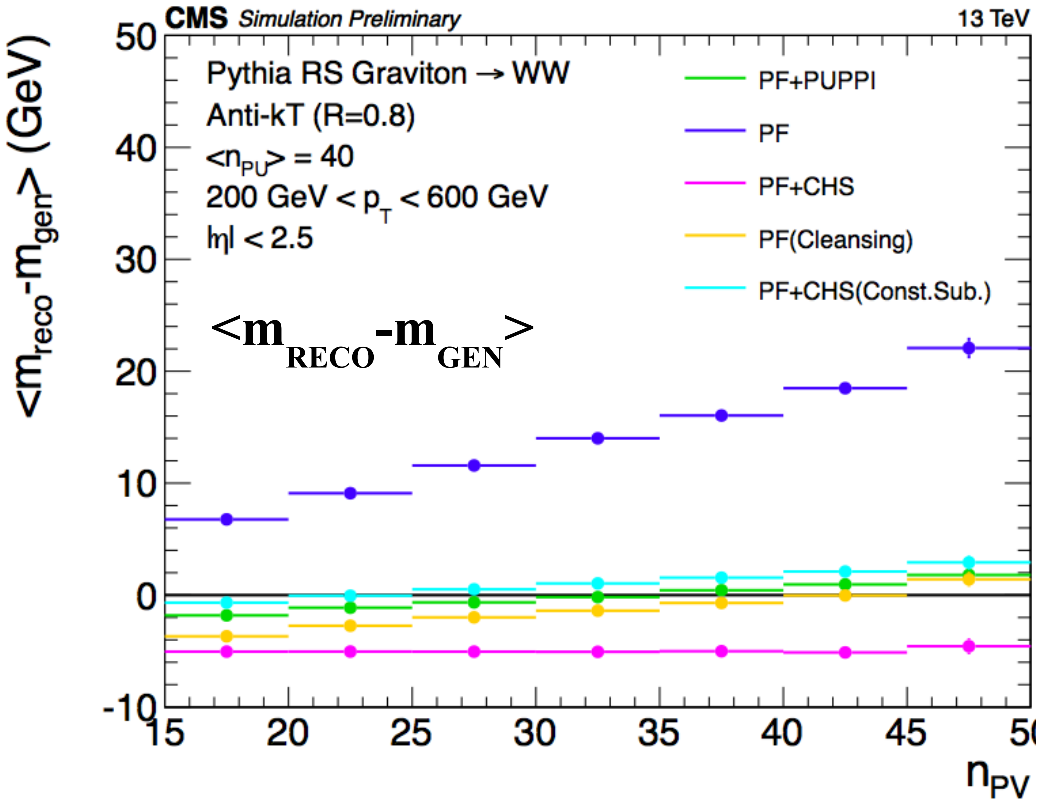


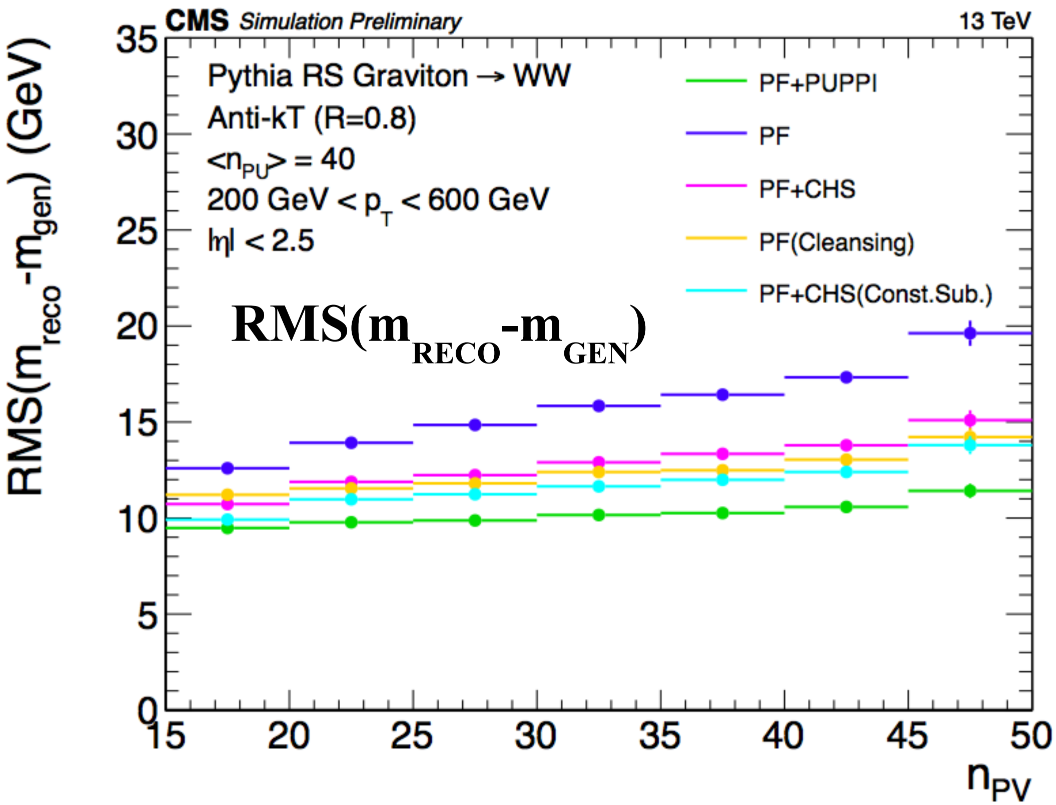


- Constituents subtraction improves the offset wrt PF+CHS
- Cleansing has good bulk resolution, but some residual tails
- Best resolution with PUPPI

Mass stability VS PU

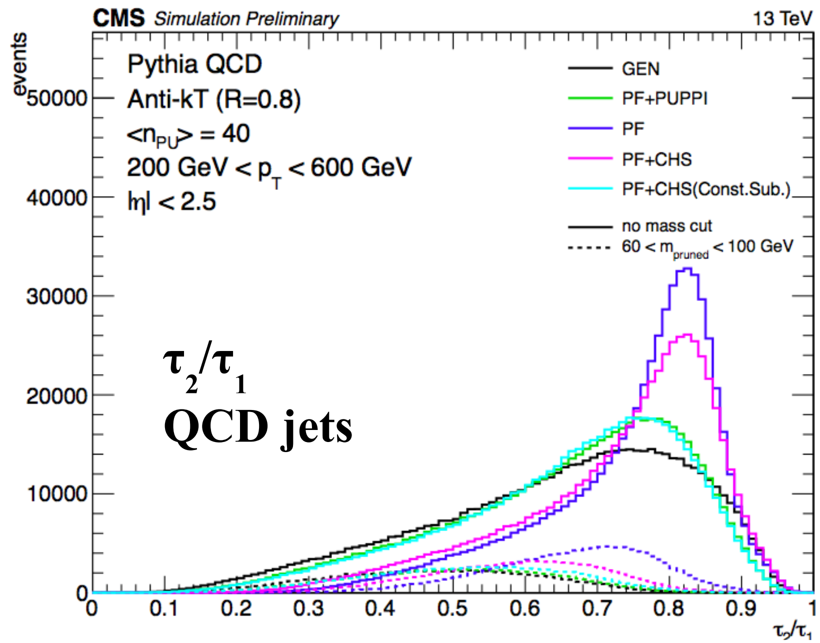
Jet mass response and resolution (W jets) stability VS n_{PV}

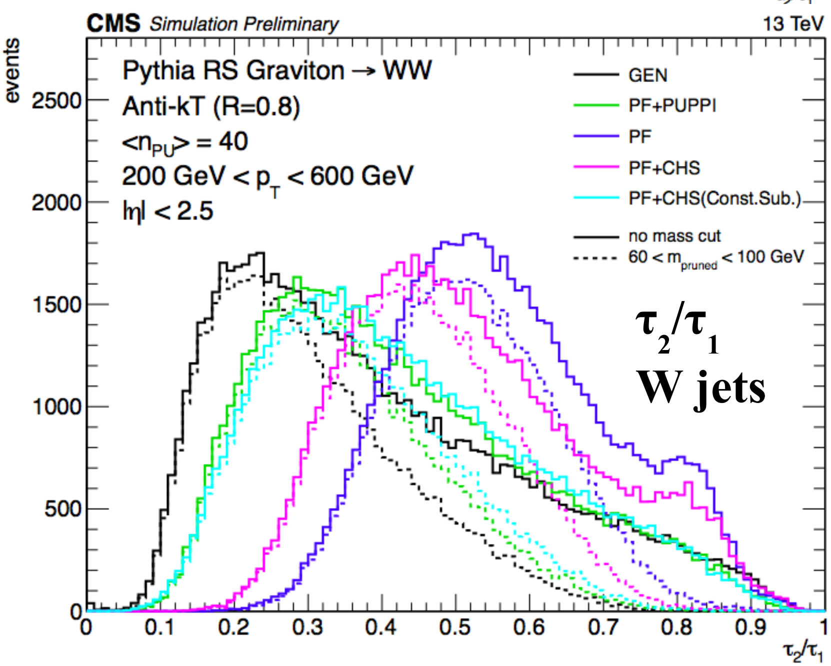




- PF+CHS stable mass response, but largest offset
- Constituens subtraction : better rensponse and resolution, still residual dependence on PU
- Cleansing : dependence on PU similar to constituents subtraction, but slighlty worse in offset and resolution
- PUPPI (tuned to get ~unity response): best resolution and reduced PU dependence

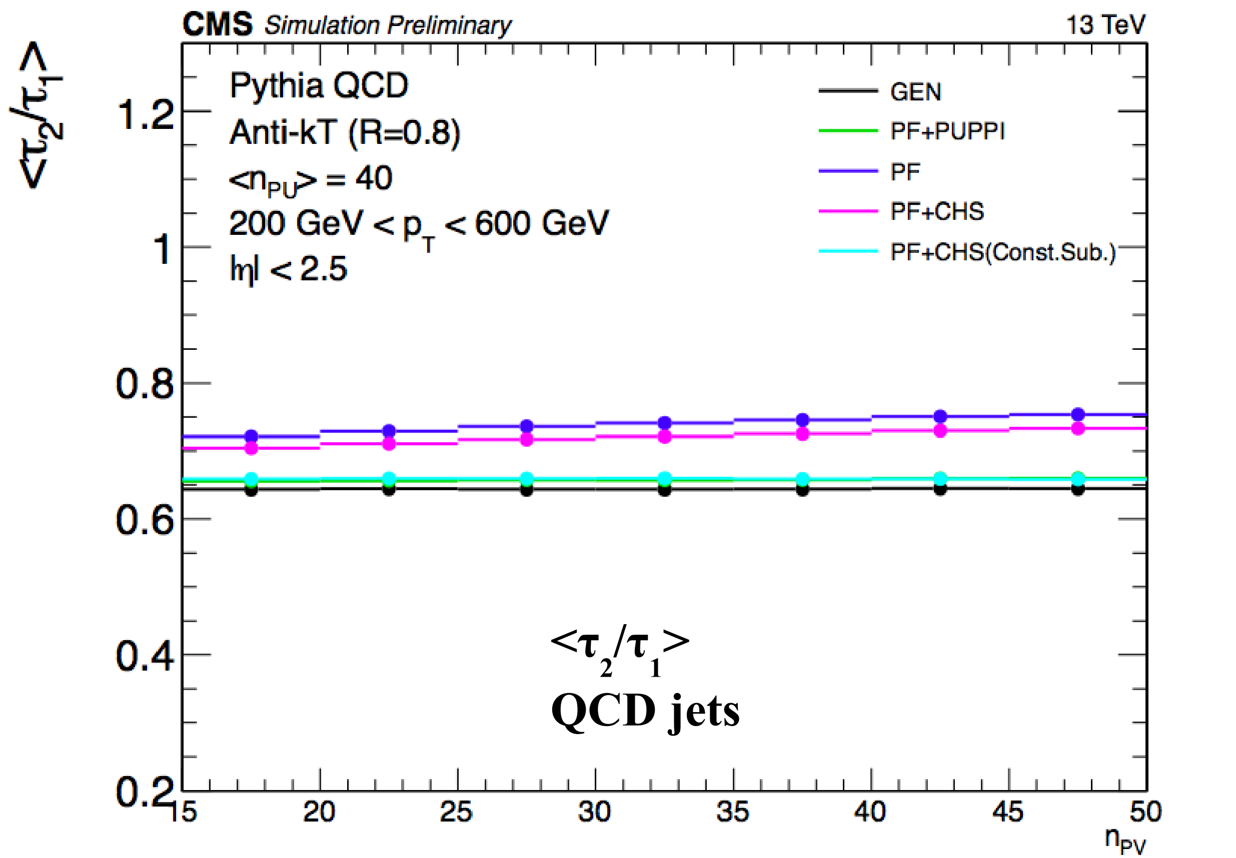
Performances of N-subjettiness



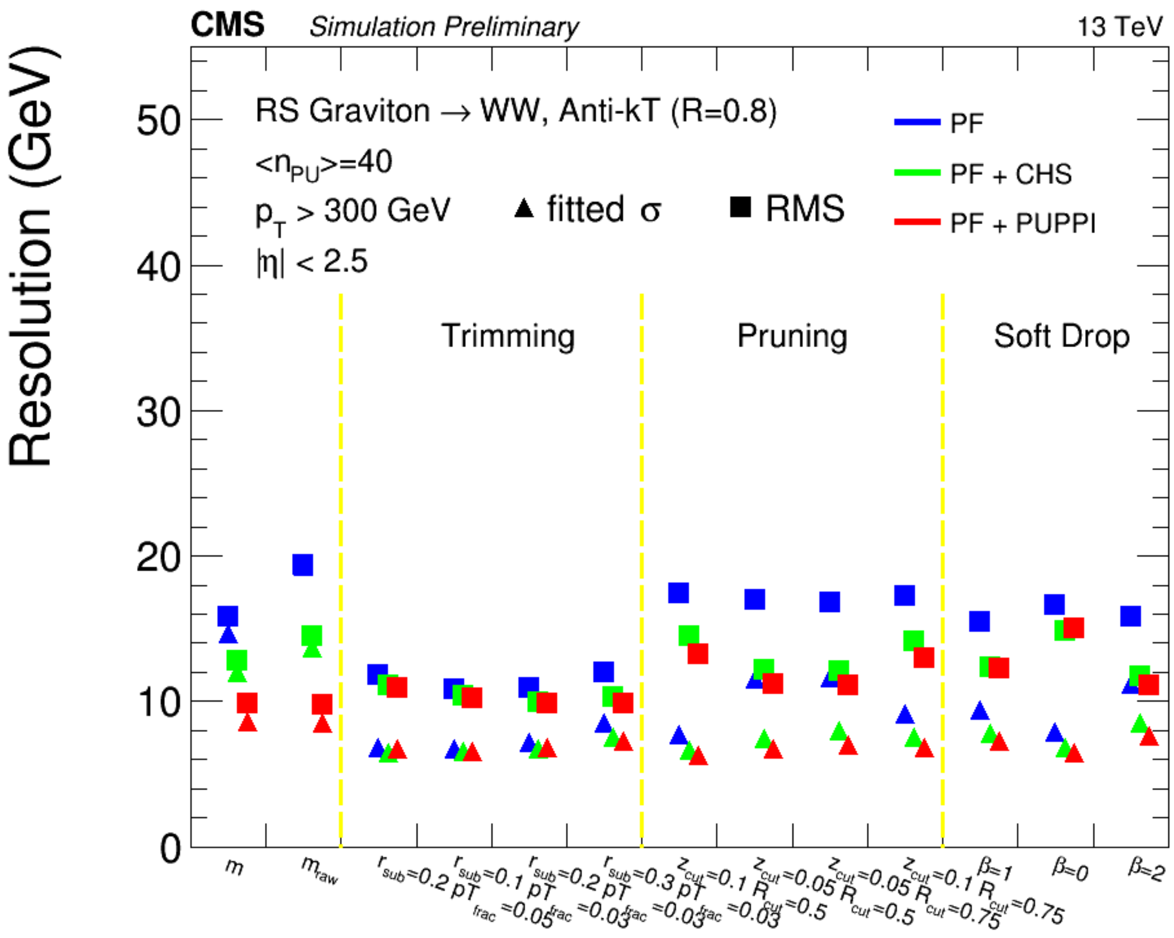


Best performances for PUPPI and constituents subtraction :

- Effective for reconstructing jet shape variables (not necessarily increasing the signal-background discrimination)
- Good stability VS PU



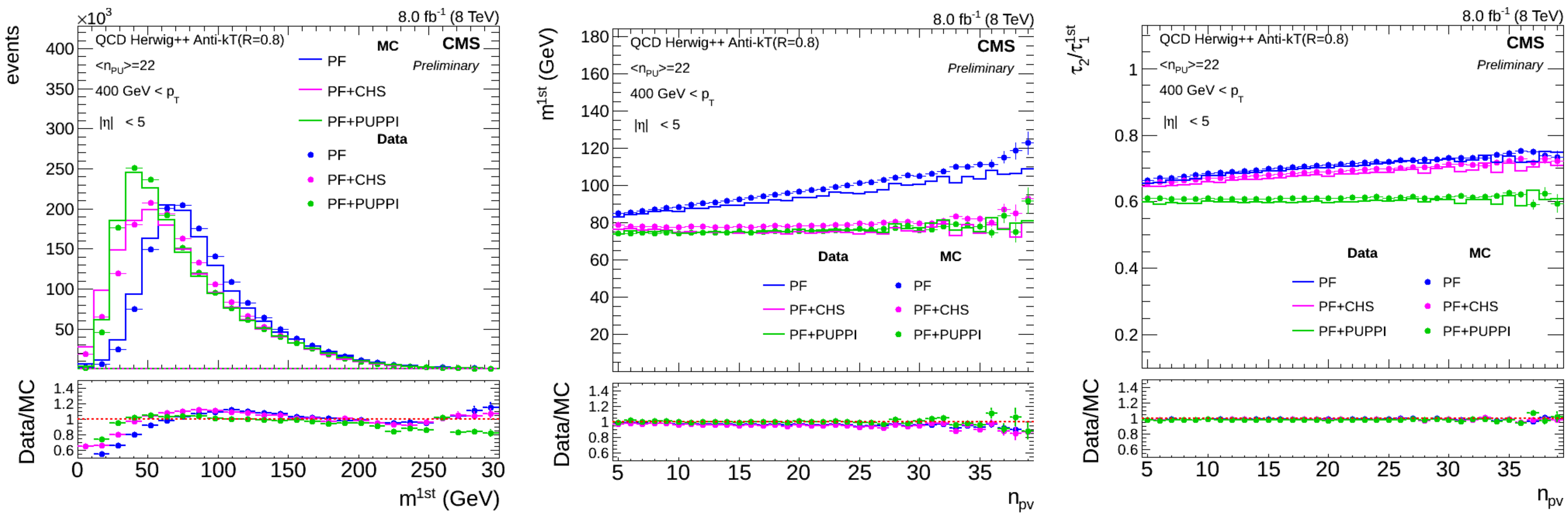
PUPPI + grooming



- PUPPI with and without 4-vector safe subtraction very similar : proof of PUPPI PU removal
- PUPPI : visible improvement for groomings that are per-particle
- PUPPI+grooming can introduce larger tails wrt to PUPPI alone (still best resolution)

PUPPI data-MC comparisons

Use of 8 TeV data (8fb^{-1} from late RunI, <PU>~22) and multijet simulation (Herwig++)
Basic dijet selection to target a region interesting for resonance searches
→ at least one jet with p_T>400 GeV, m_{jj}>900 GeV, $|\Delta\eta_{jj}|$<1.2



Overall reasonable agreement

Conclusions

- PU is one of the hardest challenges for the upcoming LHC Run II

- Exploit the physics content, combe all topological and kinematic properties of the different processes :
  - Several methods have been developed and tested thanks to RunI data
  - Important improvements in performances

A new effort is ongoing for Run II

- Many new ideas on the market, some of them still to be tested (e.g. soft killer)

- Several of them tested with CMS simulation (and crosschecked on 8 TeV data)
  - Use charged particles vertexing combined with shape information
  - Going smaller : subjets, grooming techniques
  - Going even smaller : look directly at the jet constituents

→ All is tunable, best algorithm depends on the analysis case
→ Many promising perspectives