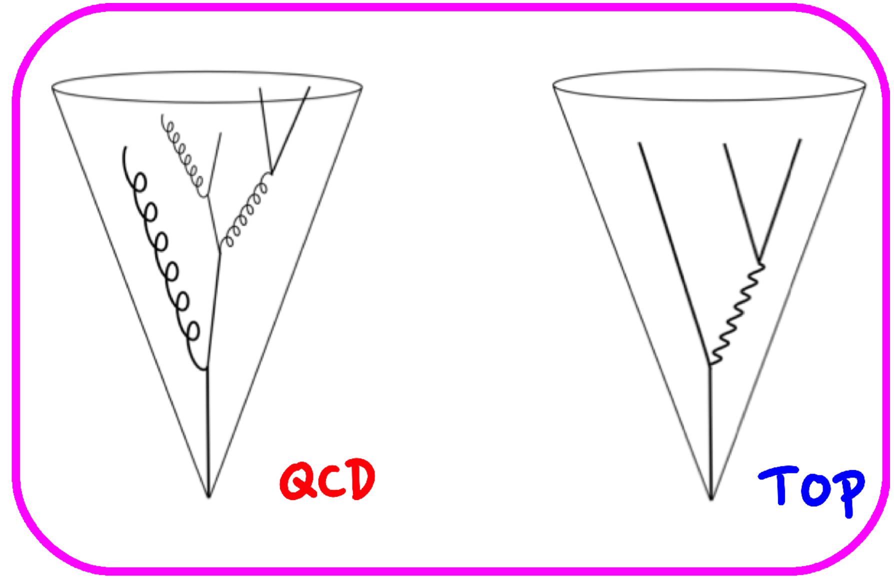
Jet Substructure at CMS

David Lopes Pegna (Princeton University-LPC FNAL)
On behalf of the CMS Collaboration

QCD@LHC 2012, Michigan State University
20 August 2012

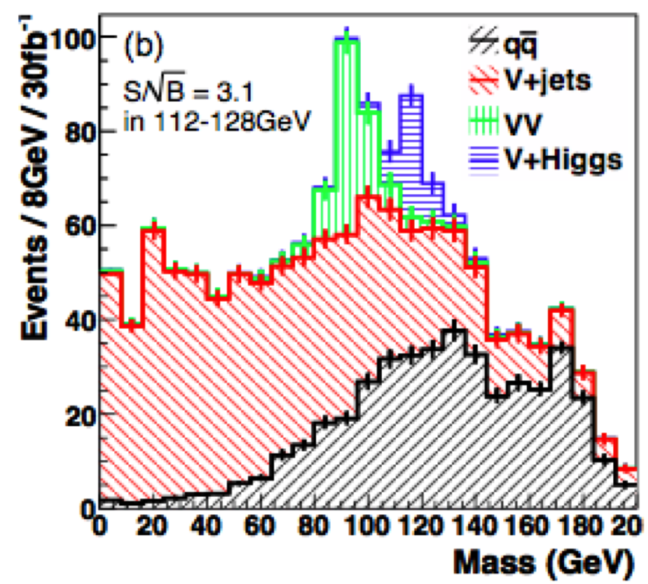
Jet Substructure at the LHC

- Many recent phenomenological works have highlighted the potential of jet substructure at the LHC
  → improve sensitivity of hadronic decays of boosted heavy particles, such as Higgs,W/Z and top
  → expected to be very robust toward Pile-Up, crucial for post LS1 LHC run and SLHC



- Requires understanding of QCD radiation inside jet, structure of constituent particles

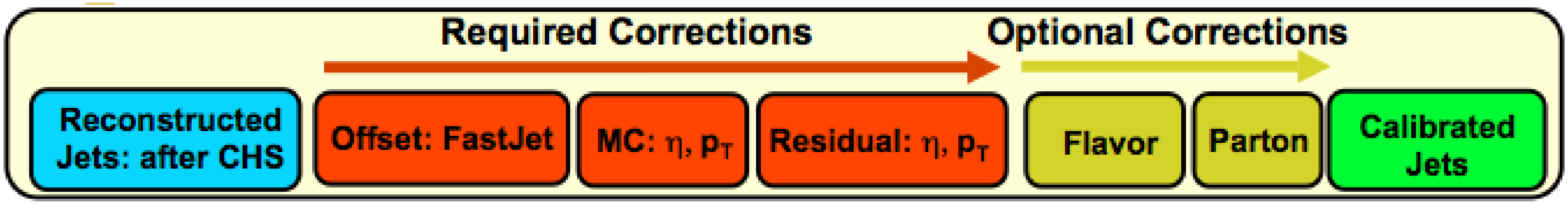
- Tools being used at CMS, broad program aimed at understanding these new approaches proceed in parallel



arXiv: 0802.2470

Jets in CMS

See also M.Voutilainen's talk



- Use charged hadron subtracted (CHS) Particle Flow jets

- Remove all charged hadrons not from the primary vertex
  Neutrals are removed via jet area ρ correction where ρ is energy density computed with the KT6 algorithm, "L1 fastjet" correction
  Active area used in all cases including groomed
  Non-linearities in η and p_T are corrected for, "Level 2/3"

- AK5, AK7 jets have dedicated corrections while other large radii jets use AK7 corrections

- Jet finding and grooming algorithms are all implemented via FastJet3

Grooming Algorithms

Jet grooming: get rid of softer components in a jet from UE or pileup and leave constituents from the hard scatter behind

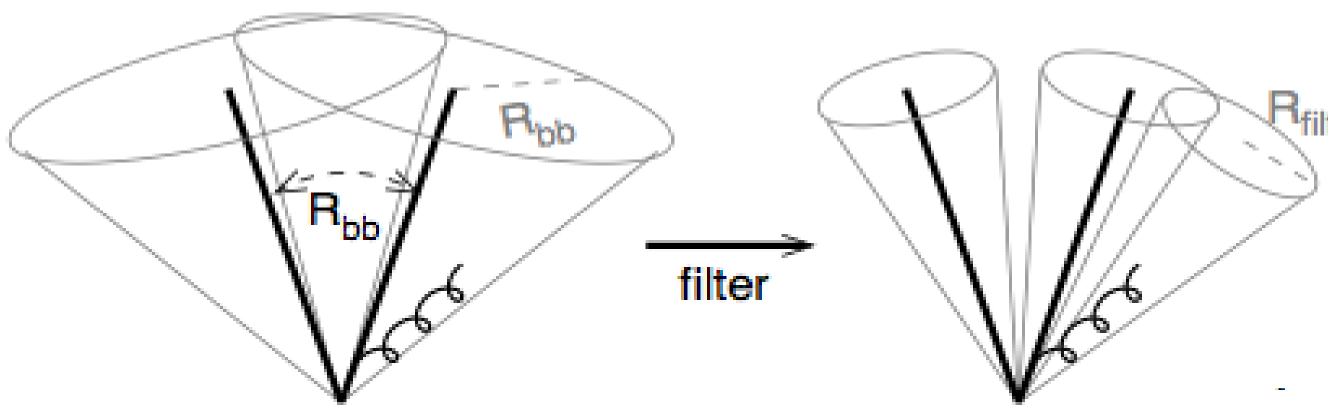
“Filtering” → http://arxiv.org/abs/0802.2470

(J. Butterworth, A. Davidson, M. Rubin, G. Salam, BDRS)

Re-clustering jet constituents with smaller radius, r_{filt}, keeping n_{filt} hardest sub-jets

→ default parameters: $r_{filt} = 0.3$, $n_{filt} = 3$

→ Was optimized for H→bb search using C/A jets...not applied to anti-kT jets!



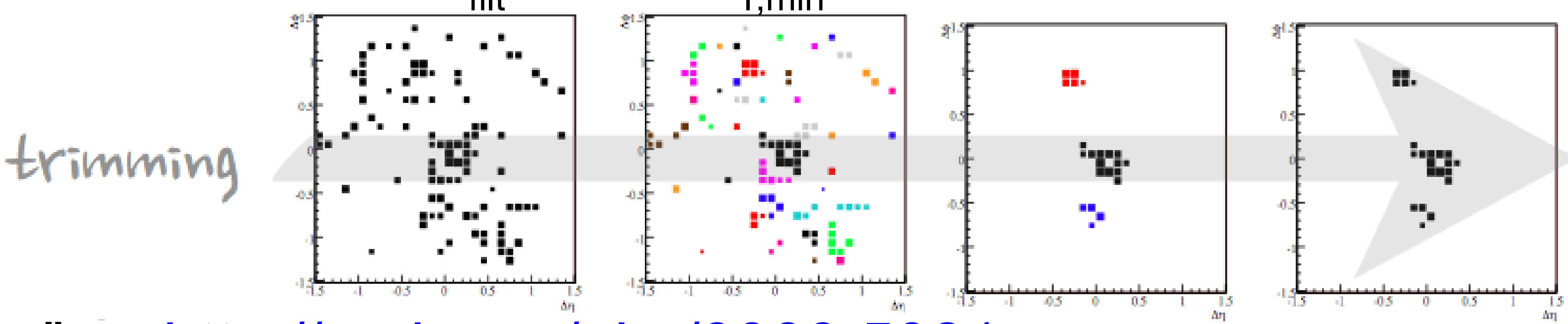
Grooming Algorithms

"Trimming" → http://arxiv.org/abs/0912.1342

(D.Krohn, J.Thaler, L.Wang)

Re-clustering (kT algo) with smaller radius r_{filt}, keeping subjects with a fraction $_{pT,min}$ of original jet p_T

→ default parameters: r_{filt} = 0.2, $p_{T,min}$=0.03



"Pruning" → http://arxiv.org/abs/0903.5081

(S.Ellis, C.Vermilion, J.Walsh)

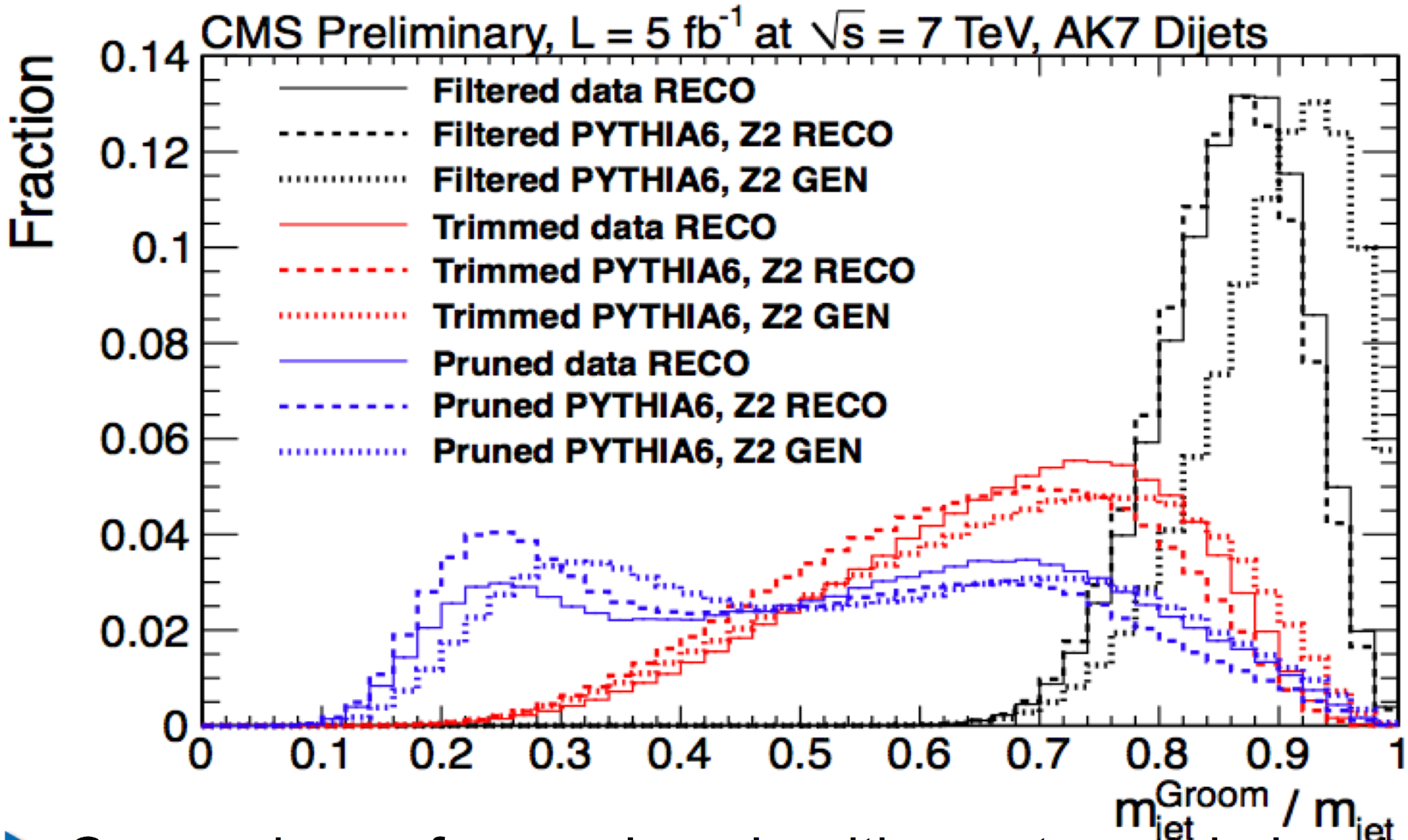
Re-clustering (CA algo) with while vetoing wide angle (R_{cut}) and softer (z_{cut}) constituents: Veto $d_{12} > R_{cut} \times 2m/p_T$; $z=\min(p_{T1},p_{T2})/p_T < z_{cut}$

→ default parameters: z_{cut} = 0.3, R_{cut} = 0.5

→ Doesn't recreate subjets but prunes at each point in jet reconstruction

Grooming Algorithms

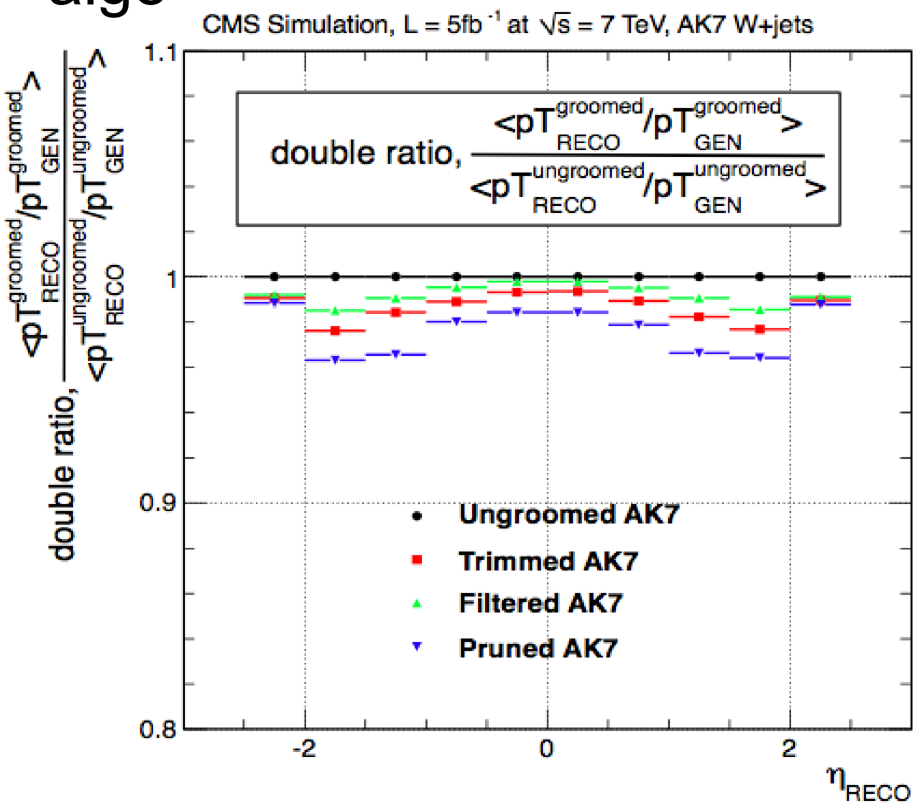
https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMP12019

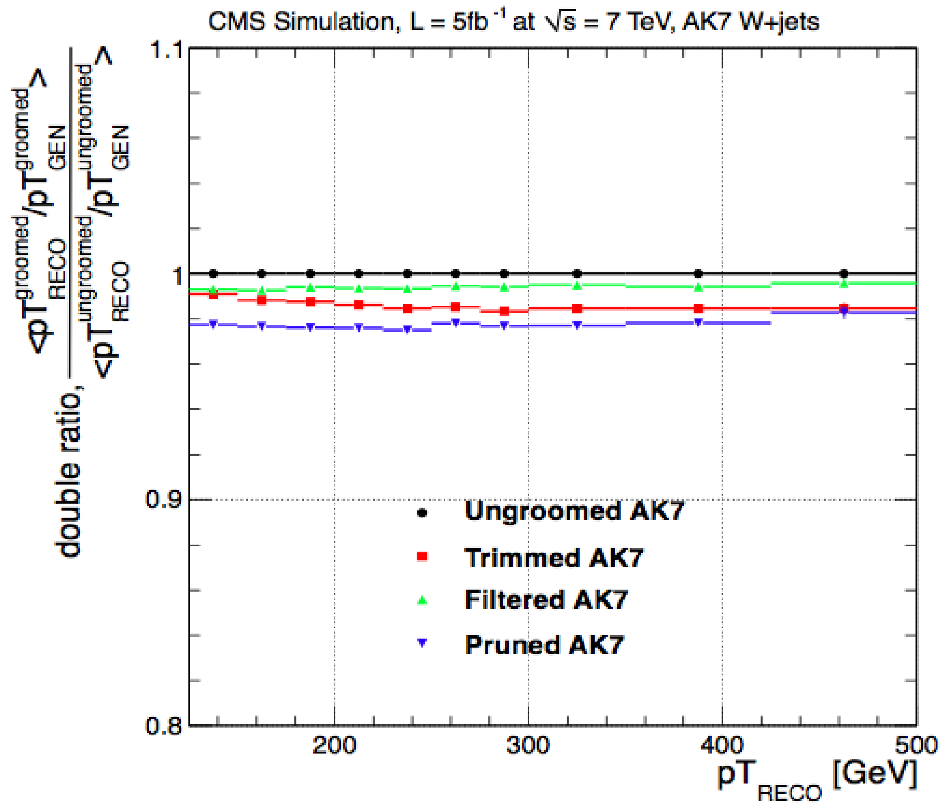


- Comparison of grooming algorithms at particle level (GEN), reconstructed simulation (RECO) and data
- Pruning is the most aggressive, filtering is the least aggressive

Jet Response: Scale

- Jet response for grooming algorithms, double ratio w.r.t. ungroomed algo

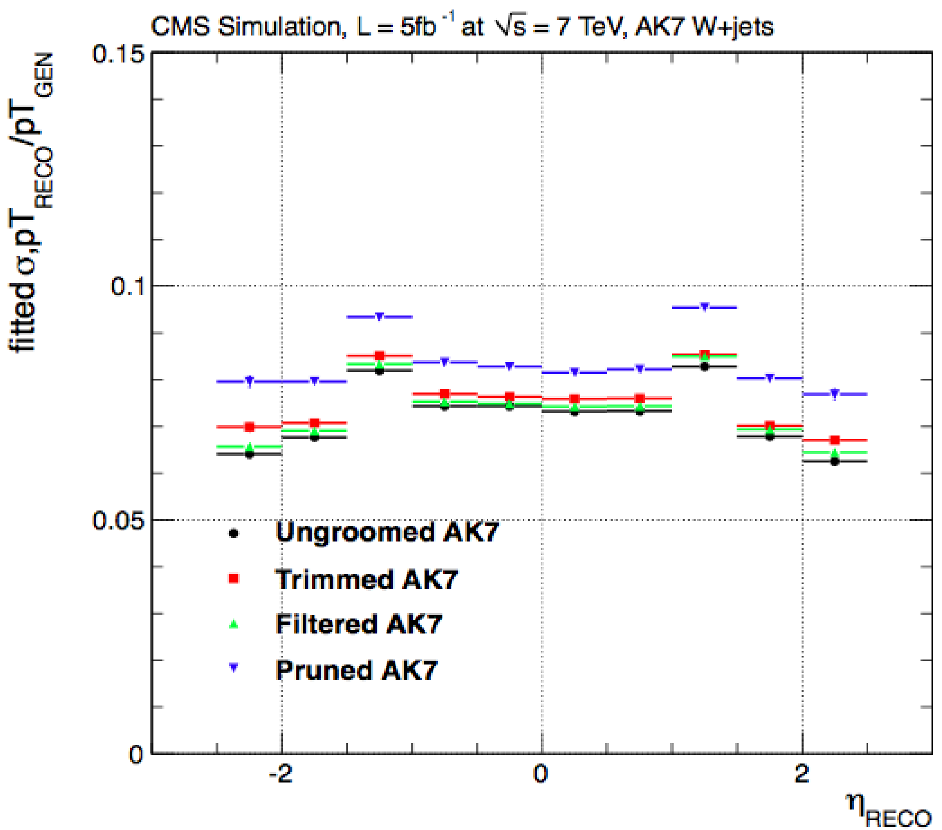


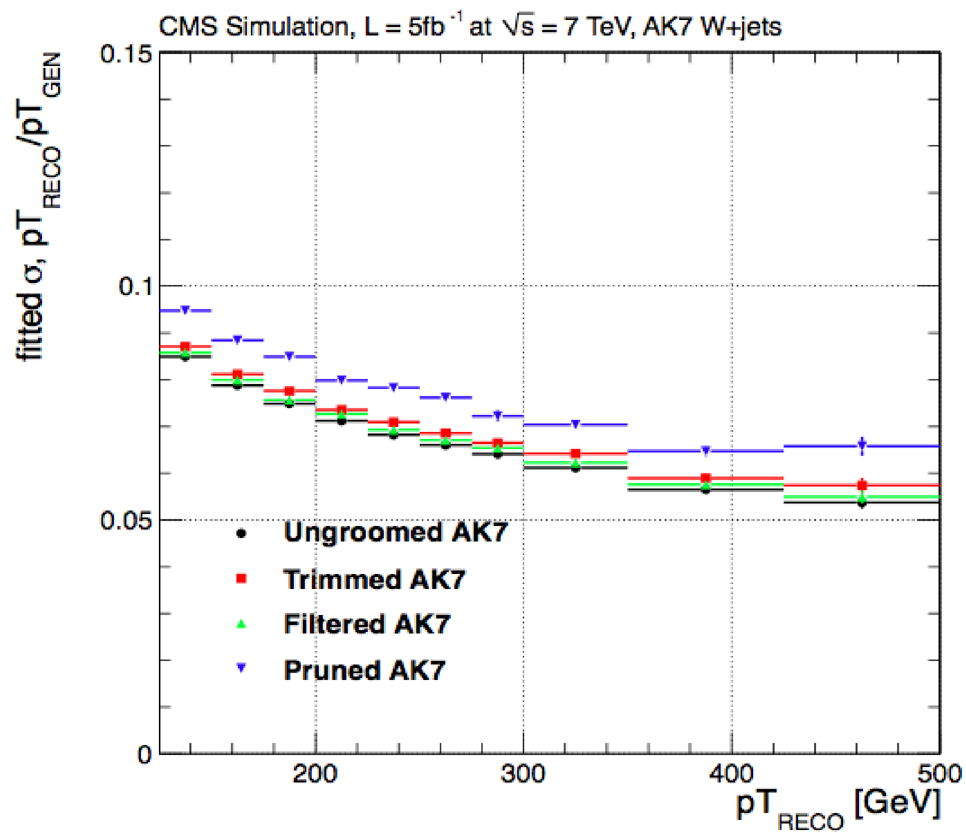


- Most aggressive grooming algorithm, pruned, still exhibits response within a few percent of ungroomed case
- No dedicated corrections applied! residual differences are encoded in response matrices

Jet Response: Resolution

- Jet p_T resolution for various grooming algorithms also shows good agreement to within a few percent.
  $\rightarrow$ Pruned jet p_T resolution degraded slightly.



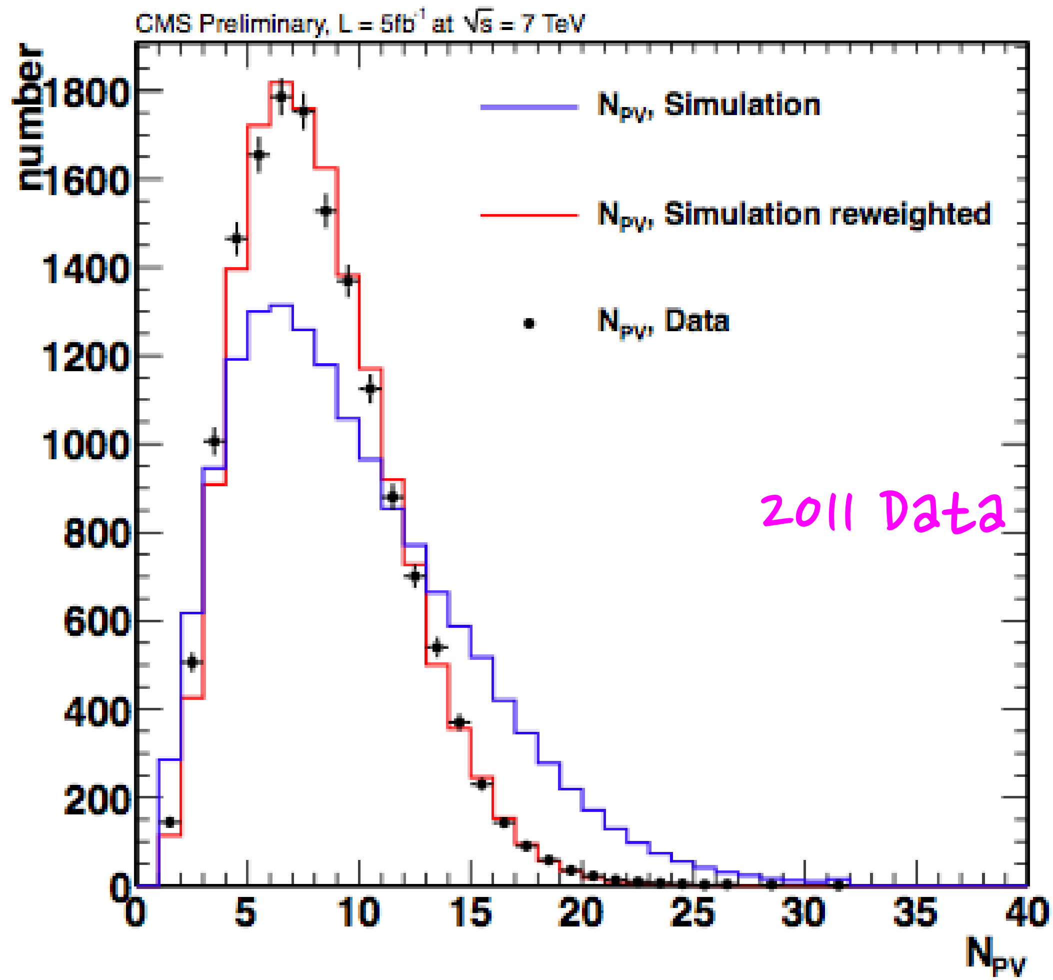


Performance vs Pile-Up

- It is of particular interest to understand the sensitivity of large size jets in presence of pileup

  → Grooming techniques may serve to mitigate pileup sensitivity by effectively reducing the jet area

  → Understand performance of mean jet mass as a function of number of primary vertices

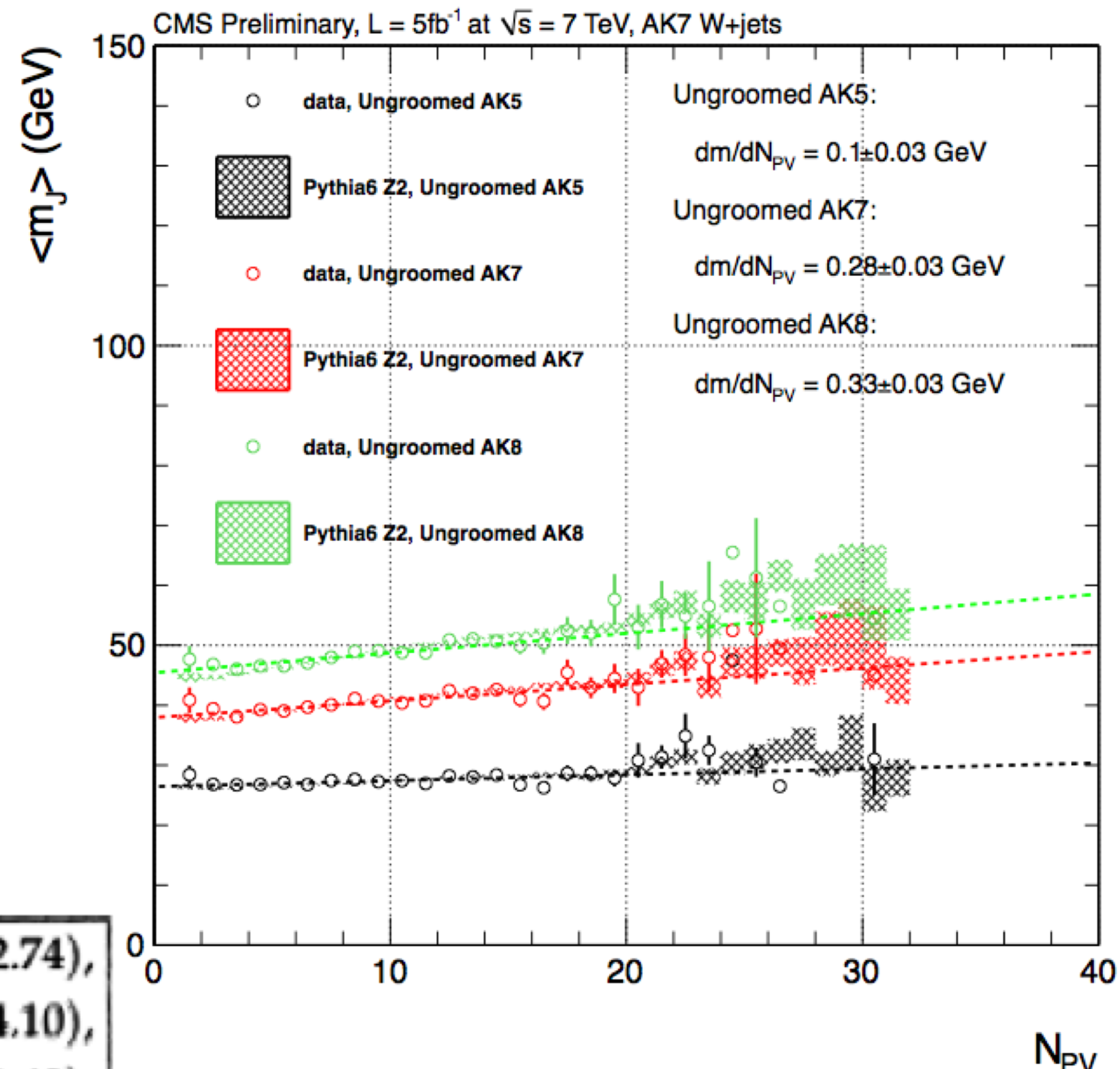


Performance vs Pile-Up

- Un-groomed jet mass is very sensitive to PU
- $\langle m_J \rangle$ increases linearly as a function of the number of primary vertices
- Effect becomes more pronounced as the jet size increases AK8 shows much worse effect than AK5

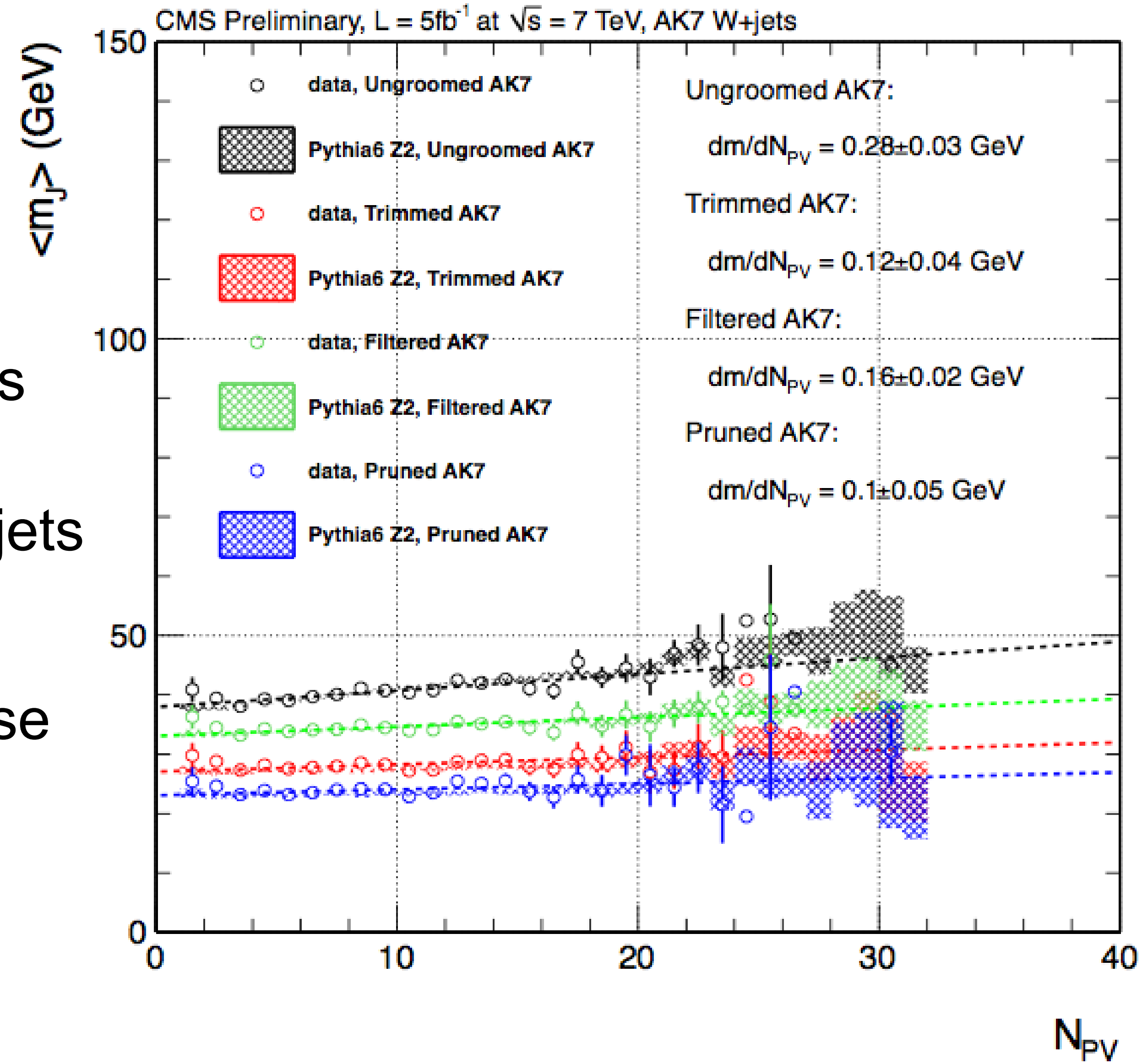
Observe the expected behavior that $\langle m_J \rangle$ typically scales as R^3

$$
\begin{aligned}
s_{0.7}/s_{0.5} &= 2.7 \pm 0.9 \qquad ((0.7/0.5)^3 = 2.74),\\
s_{0.8}/s_{0.5} &= 3.3 \pm 1.0 \qquad ((0.8/0.5)^3 = 4.10),\\
s_{0.8}/s_{0.7} &= 1.2 \pm 0.2 \qquad ((0.8/0.7)^3 = 1.49)
\end{aligned}
$$

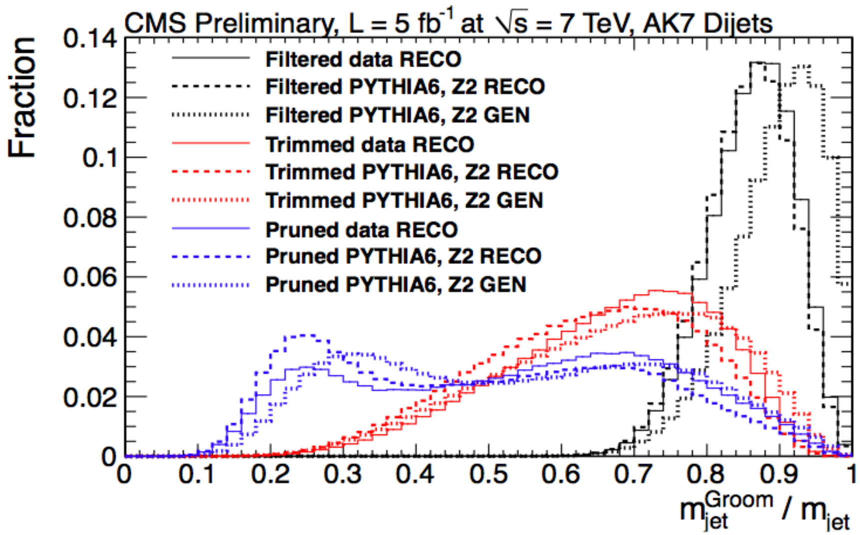


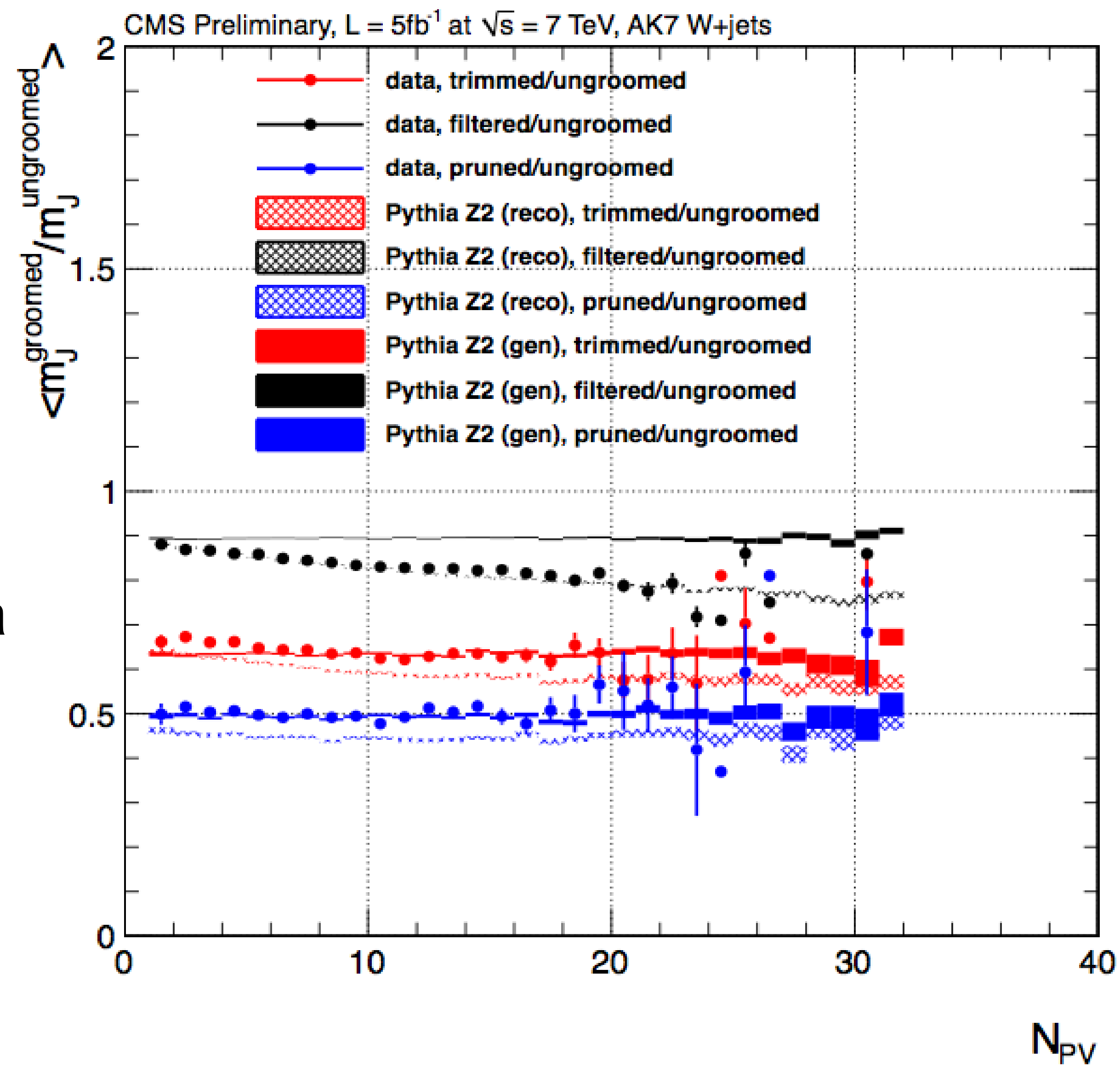
Performance vs Pile-Up

- Grooming techniques are less sensitive to PU
- $<m_J>$ vs N_{PV} slope becomes flatter
  → perfectly flat for pruned jets
- Very promising for future use of these techniques



Performance vs Pile-Up





- Grooming in simulation more aggressive than data
- PU-independent for the most aggressive grooming
- Dependence for filtering nicely reproduced in simulation

Analysis Benchmarks

- Jet mass among the most valuable observable: sensitive to energy scale and splashing contributions (PU,UE,ISR)

→ Provide an inclusive comparative study of jet mass for available jet grooming methods
→ Make a comparison of data to simulation as a validation of parton showering models
→ Provide detector-unfolded jet mass distributions as inputs to the theoretical community

Clustering Algorithm	Jet grooming Algorithm	Topology studied
Anti-kT (R=0.5)		
Anti-kT (R=0.7)	Trimmed/Filtered/Pruned	QCD, EWK
Anti-kT (R=0.8)		
Cambridge-Aachen (R=0.8)	Pruned	EWK
Cambridge-Aachen (R=1.2)	Filtered	EWK

Jet Mass Analysis

- Present cross-sections as double differential distributions:

$$PDF(m_J) = \frac{1}{\frac{d\sigma}{dp_T}} \times \frac{d^2\sigma}{dp_T dm_J}$$

- In the di-jet analysis, consider the average jet p_T and mass for the two leading jets in the event
  In the V+jet analysis, consider the leading jet p_T and mass

- Final distributions unfolded using iterative Bayesian method

pT binning

bin	‹pT› for di-jet pT for V+jet
1	125-150
2	150-220
3	220-300
4	300-450
5	450-500
6	500-600
7	600-800
8	800-1000
9	1000-1500

V+jets only (bins 1–2)
Di-jet only (bins 5–9)
Di-jet and V+jet (bins 3–4)

Data sample and selection

Analysis on 5 fb^{-1} from 2011 data, $\sqrt{s}$ = 7 TeV

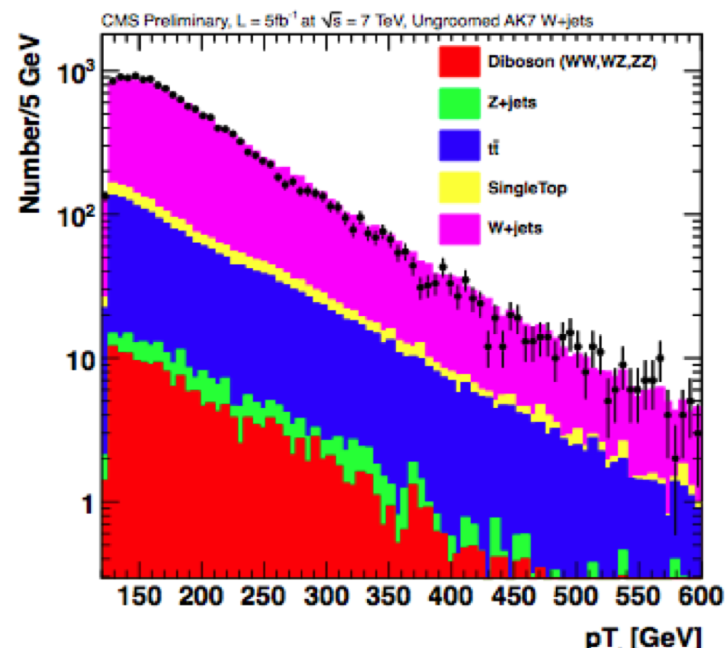
Single jet triggers with variable (pre-scaled) threshold
For V+Jet, single isolated lepton trigger

Simulation:

Dijets
QCD: Pythia6 Tune Z2/Pythia8 Tune 4c/ Herwig++ 23
V+jets
WW,WZ,ZZ: Pythia6 Tune Z2
ttbar and single top: MadGraph+Pythia6
W/Z+jets: MadGraph+Pythia6 Tune Z2/Herwig++ 23

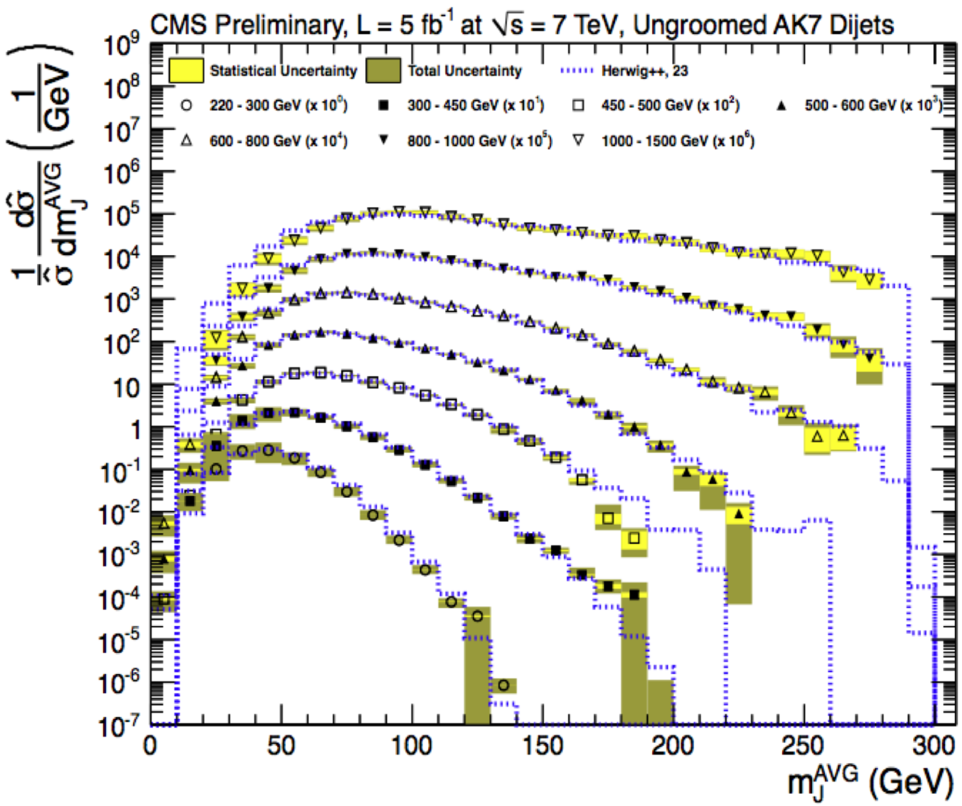
PDFs: CTEQ6L1

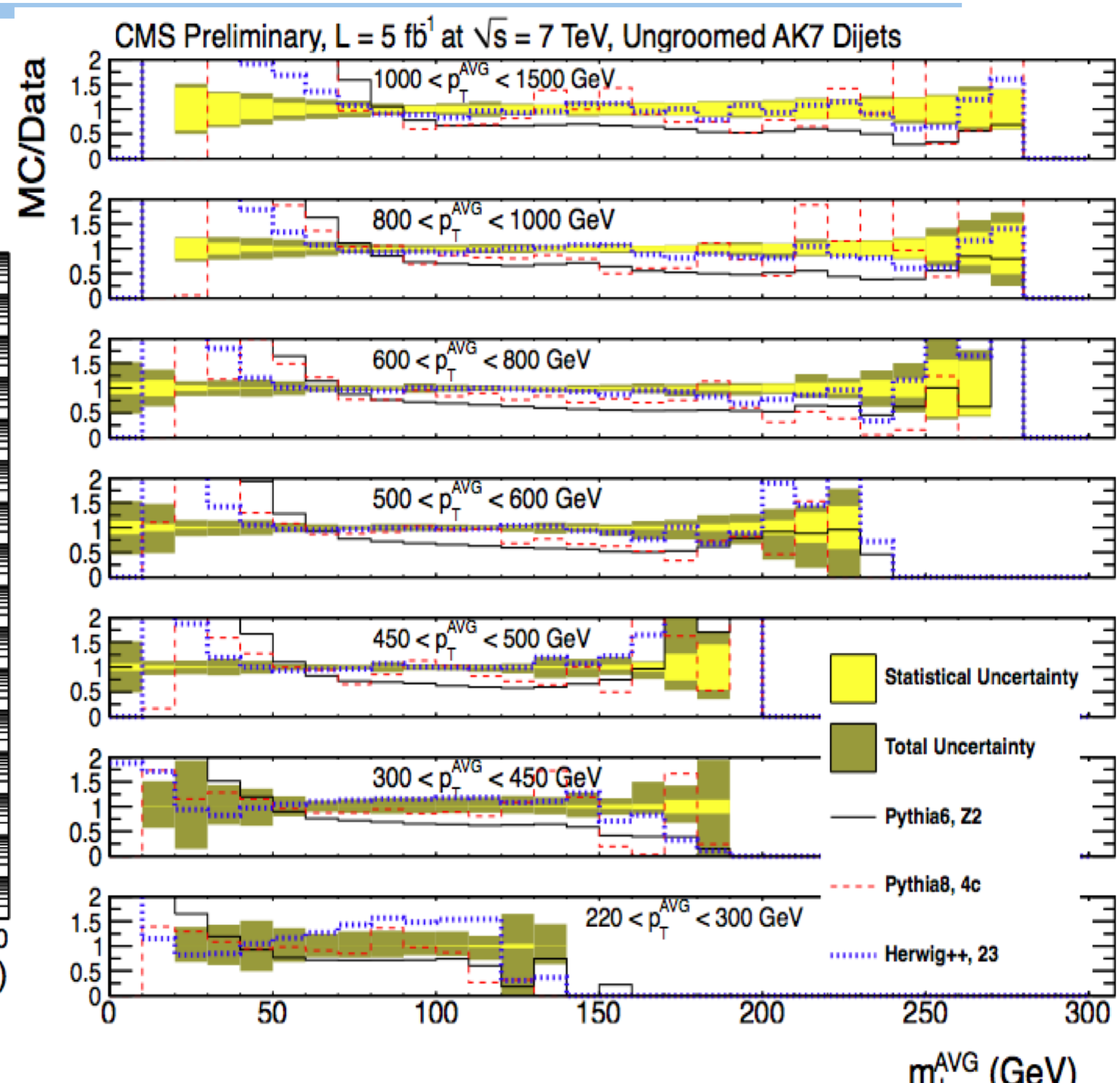
W(eν)	W(μν)	Z(ee)	Z(μμ)
$pT_W > 120$ GeV	$pT_W > 120$ GeV	$pT_Z > 120$ GeV	$pT_Z > 120$ GeV
$mT > 50$ GeV	$mT > 50$ GeV	$mZ = [80,100]$	$mZ = [80,100]$
MET > 50 GeV	MET > 50 GeV	MET < 50 GeV	MET < 50 GeV
$pT_e > 80$ GeV WP (ϵ = 70%) ($iso_{rel} < 0.05$)	$pT_\mu > 80$ GeV $iso_{rel} < 0.1$ μ quality cuts**	$pT_e > 20,20$ GeV WP (ϵ = 95%)	$pT_\mu > 30,30$ GeV $iso_{rel} < 0.1$ μ quality cuts**
$\|d\Phi_{V,j}\| > 2.0$, $\|dR_{lep,j}\| > 1.0$, $\|d\Phi_{MET,j}\| > 0.4$ [3]	$\|d\Phi_{V,j}\| > 2.0$, $\|dR_{lep,j}\| > 1.0$, $\|d\Phi_{V,j}\| > 0.4$ [3]	$\|d\Phi_{V,j}\| > 2.0$, $\|dR_{lep,j}\| > 1.0$	$\|d\Phi_{V,j}\| > 2.0$, $\|dR_{lep,j}\| > 1.0$
Jet selections			
pT(Jet) > 125 GeV, matched GEN jet (>99%), $\|\eta\| < 2.5$			



Unfolded Jet Mass, di-jet

Ungroomed AK7

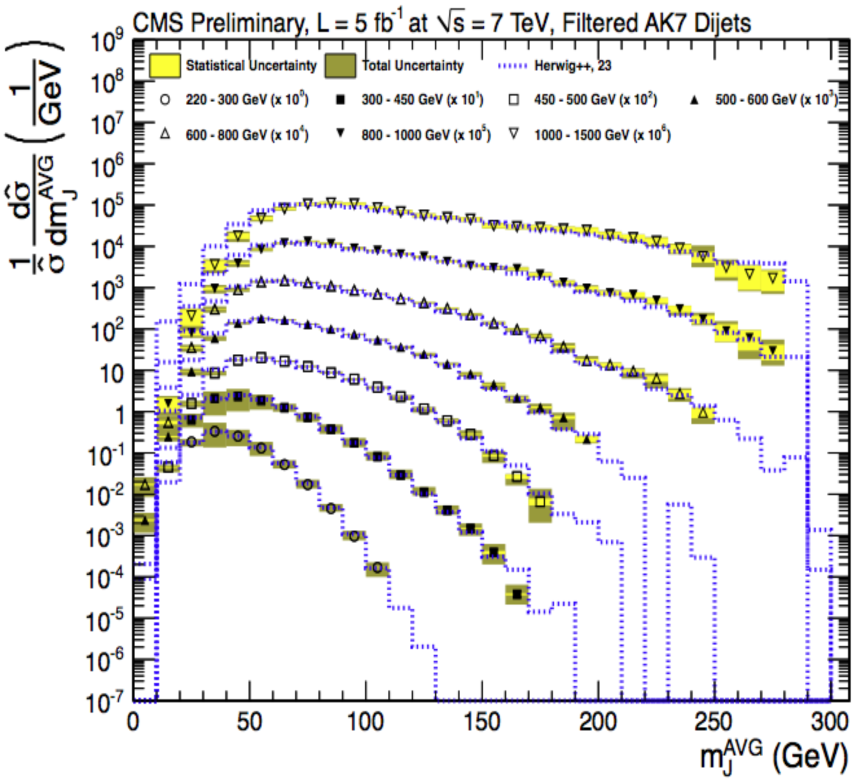


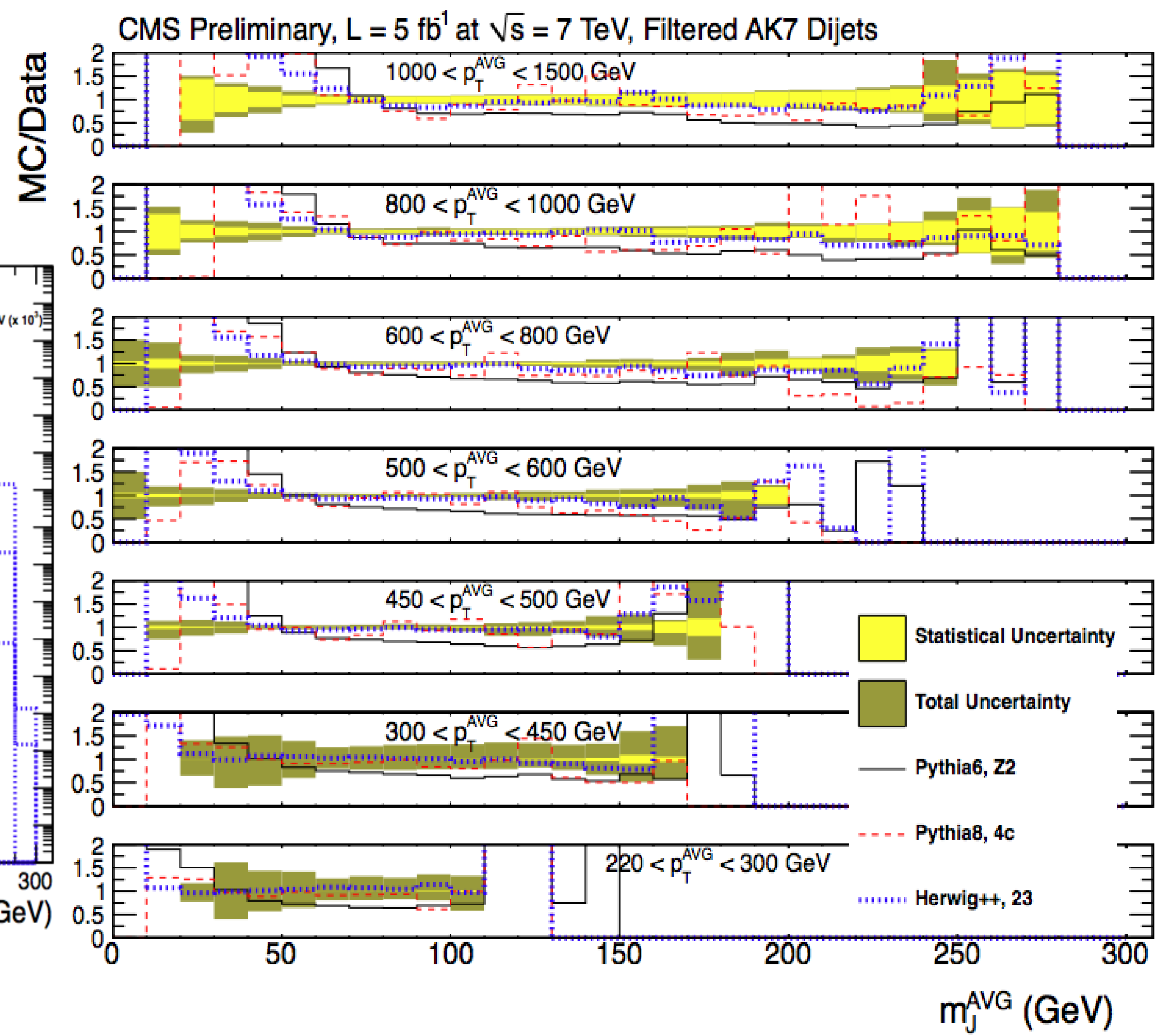


Final plots are produced in compact form with higher offset for larger pT bins. MC/Data ratios on the right

Unfolded Jet Mass, di-jet

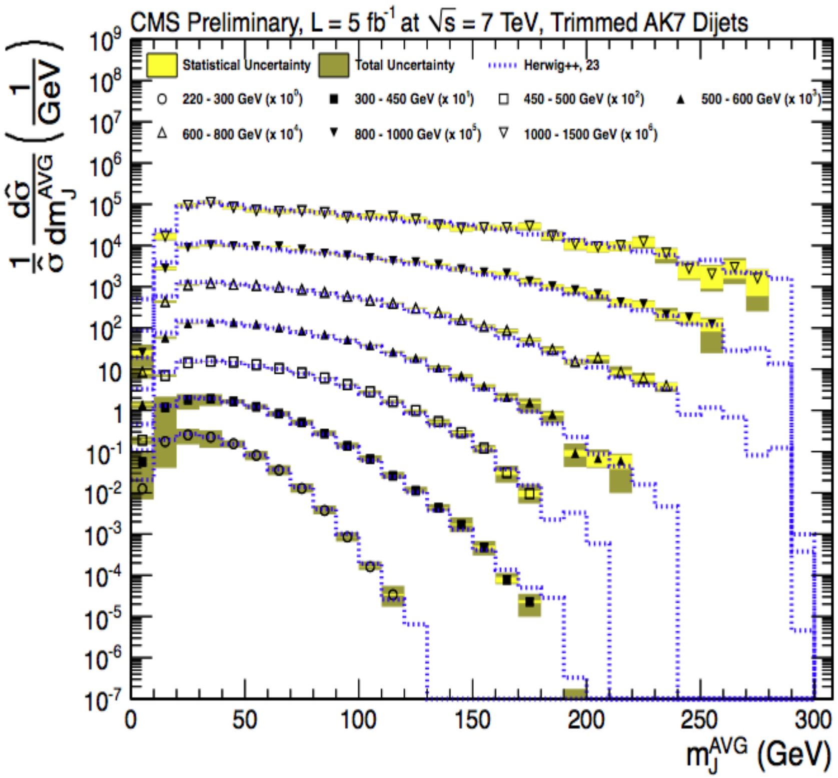
Filtered AK7

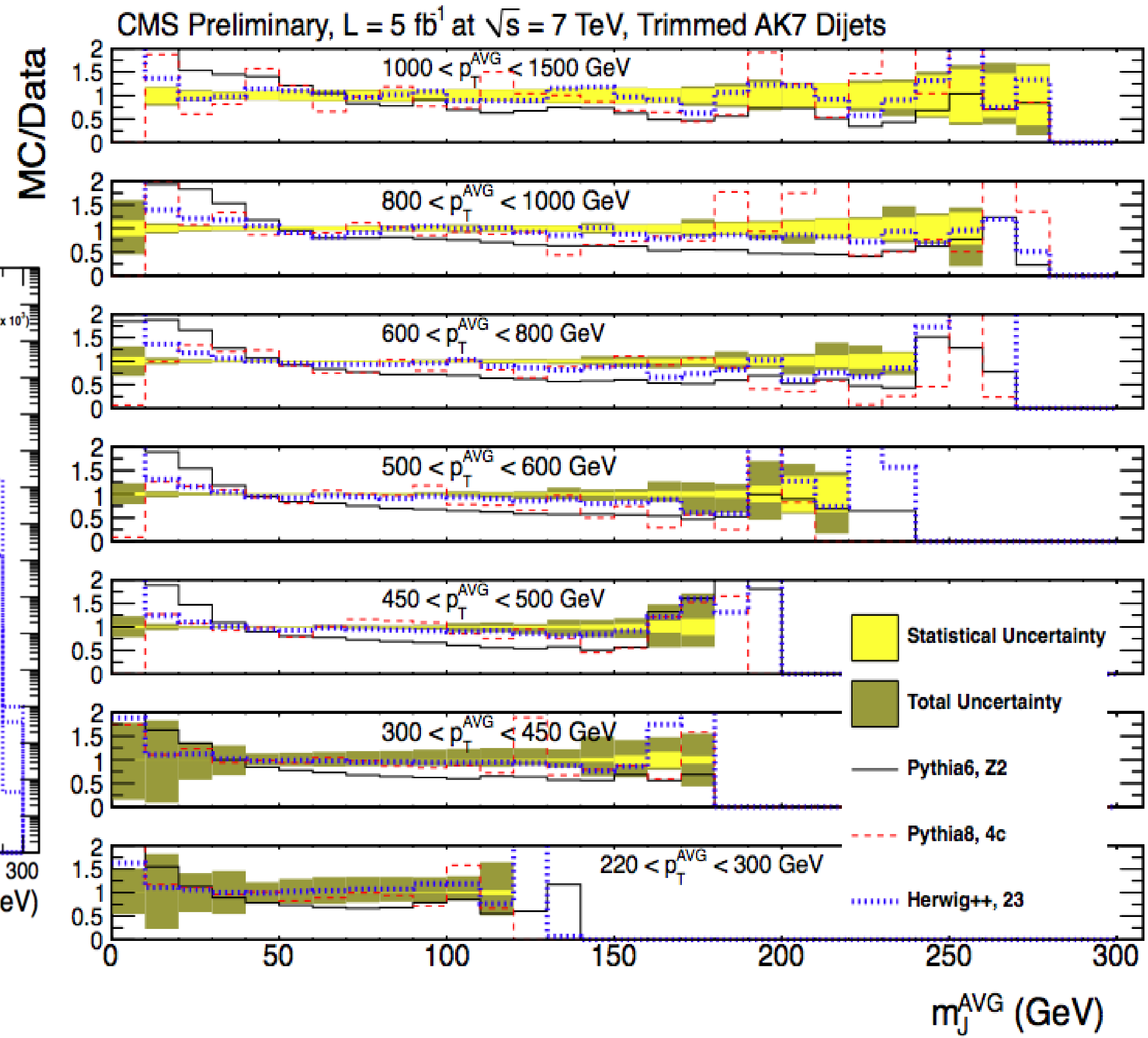




Unfolded Jet Mass, di-jet

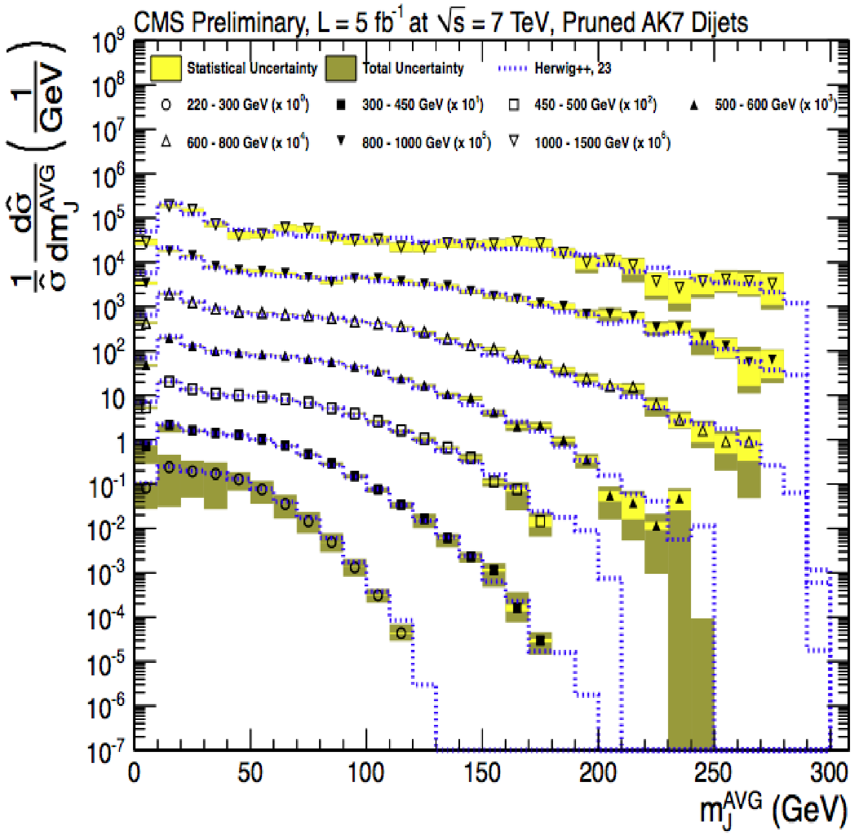
Trimmed AK7

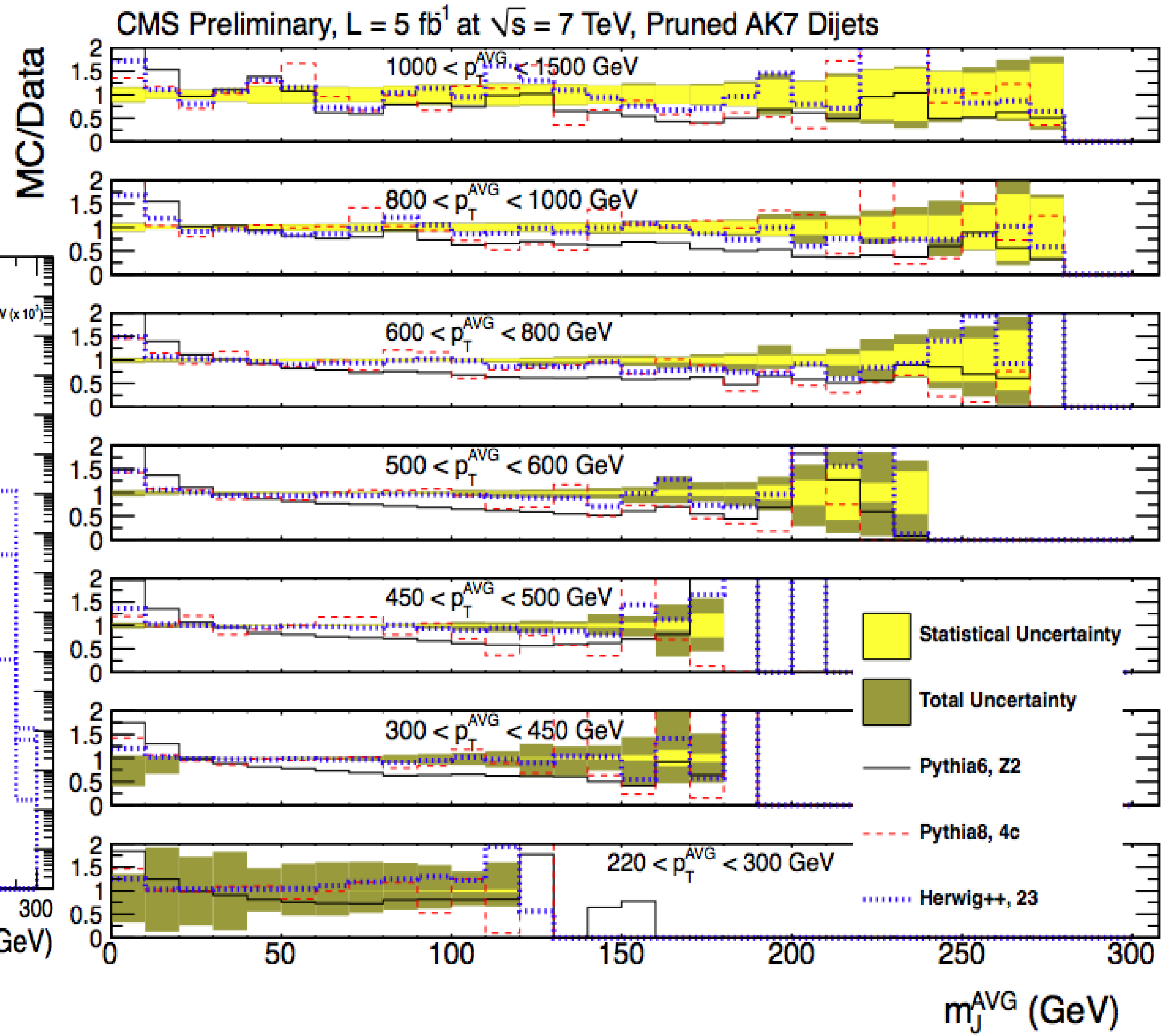




Unfolded Jet Mass, di-jet

Pruned AK7





Unfolded Jet Mass Analysis

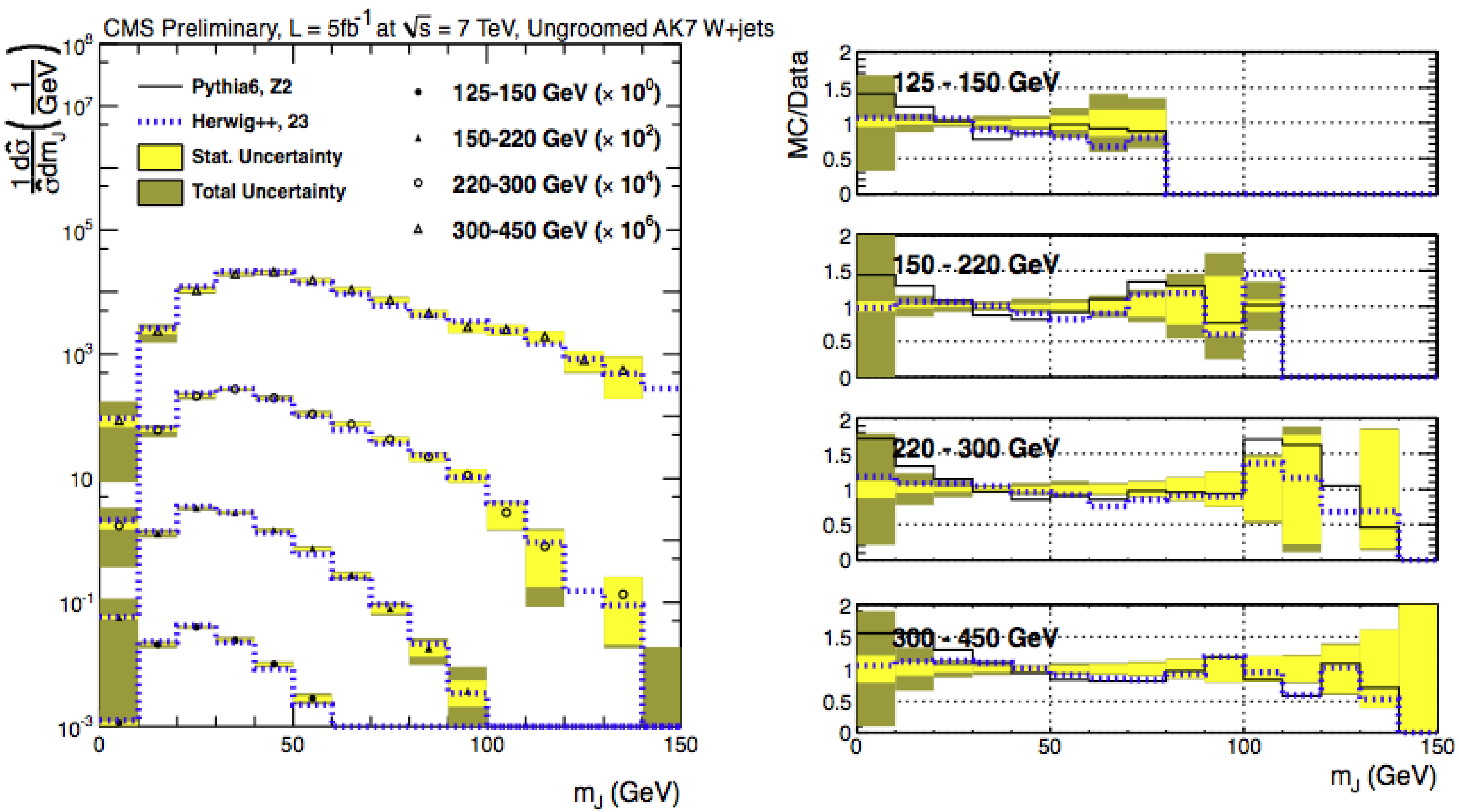
- ✔ Jet mass turnover region moves to the left with harder grooming while tails remain similar in all cases
  → important feature in new physics searches.

- ✔ MC/data agreement reasonable for both pythia and Herwig, slightly better for the latter especially for more aggressive grooming
  → excellent agreement with Herwig for $m_J > 20$ and $p_T > 300$ GeV

- ✔ Worse agreement for $m_J << p_T$
  → challenging theoretically, most sensitive region to UE/PU, MC underestimates showering

Unfolded Jet Mass, V+jet

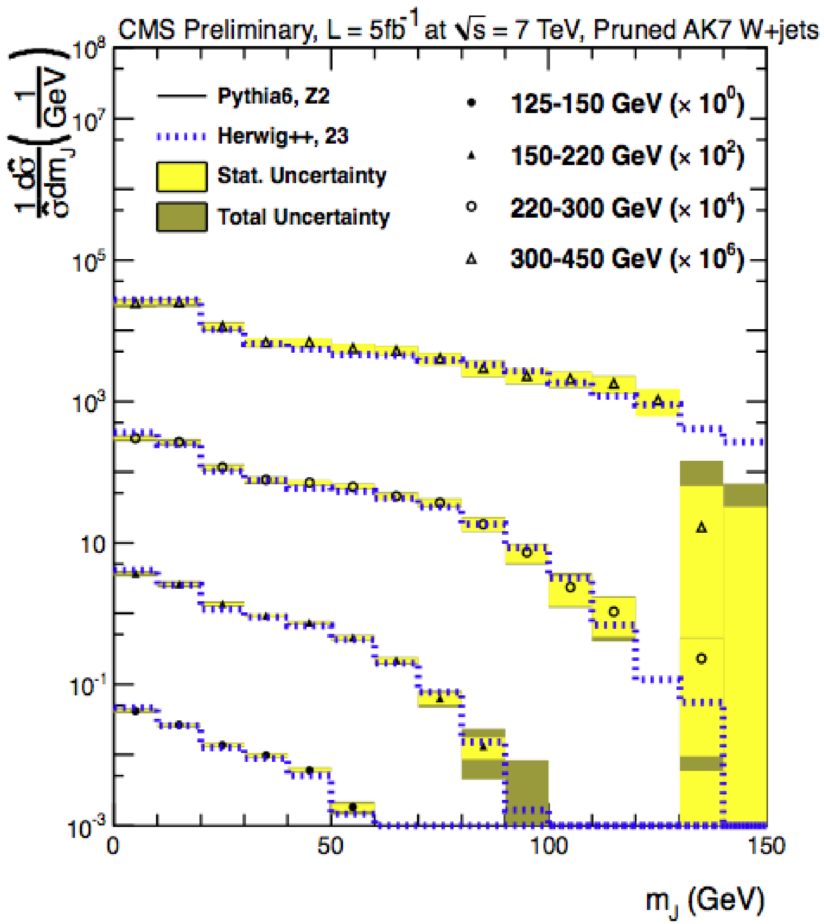
Ungroomed AK7

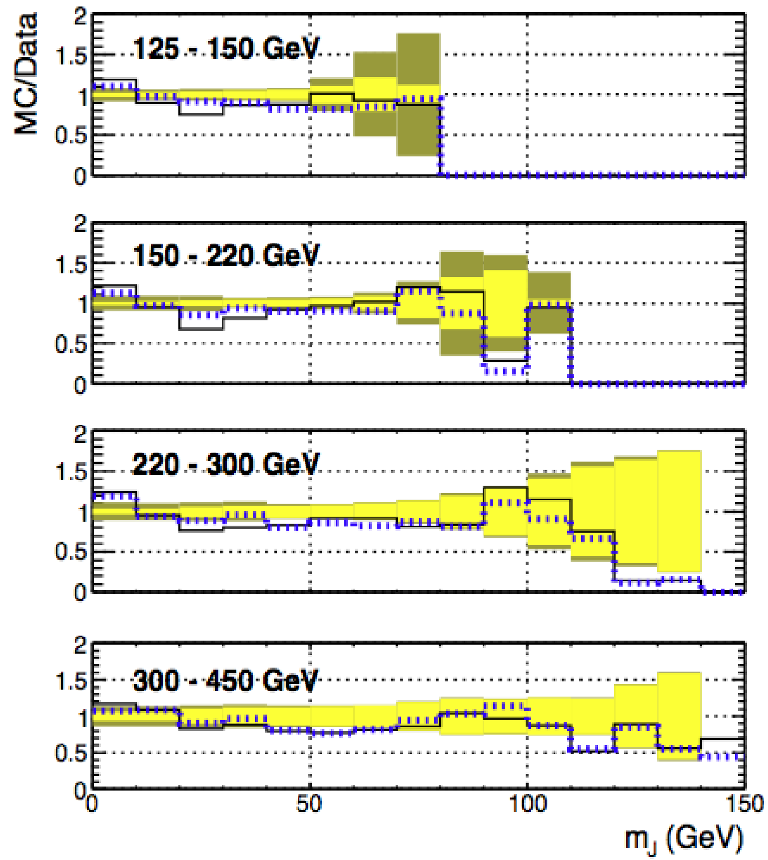


Similar qualitative behavior in V+jet as for di-jet, good agreement with both Madgraph-pythia and Herwig++

Unfolded Jet Mass, V+jet

Pruned AK7

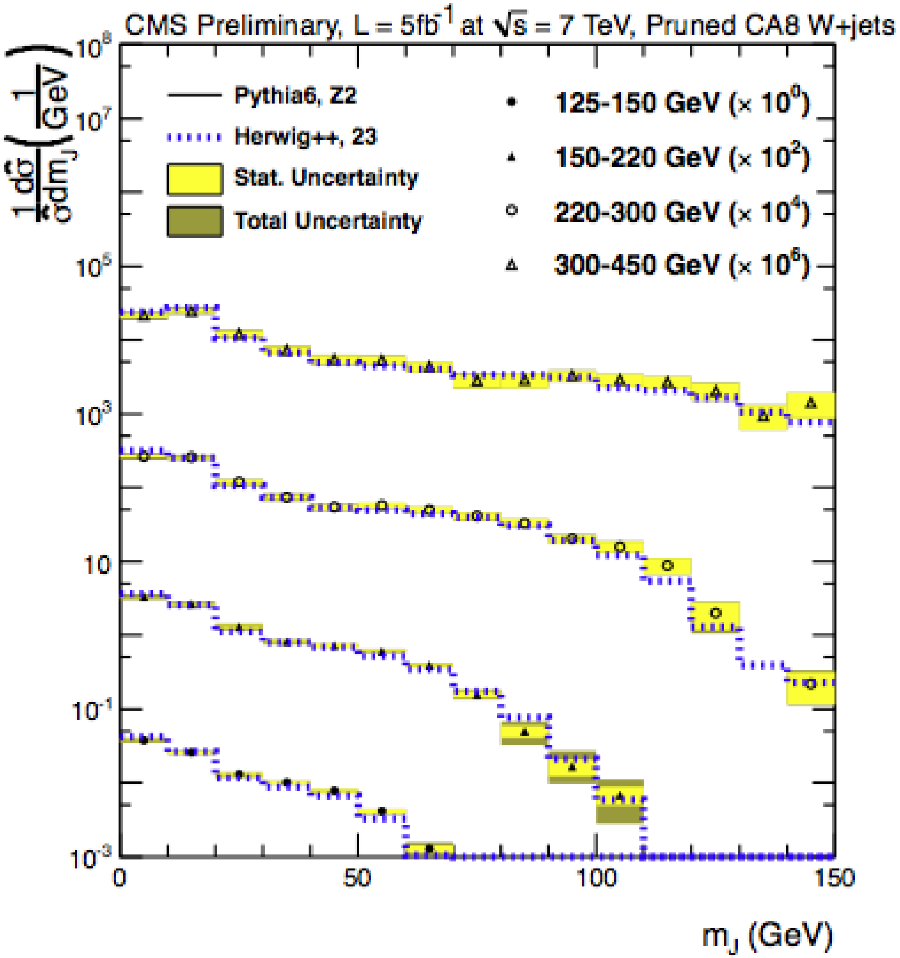


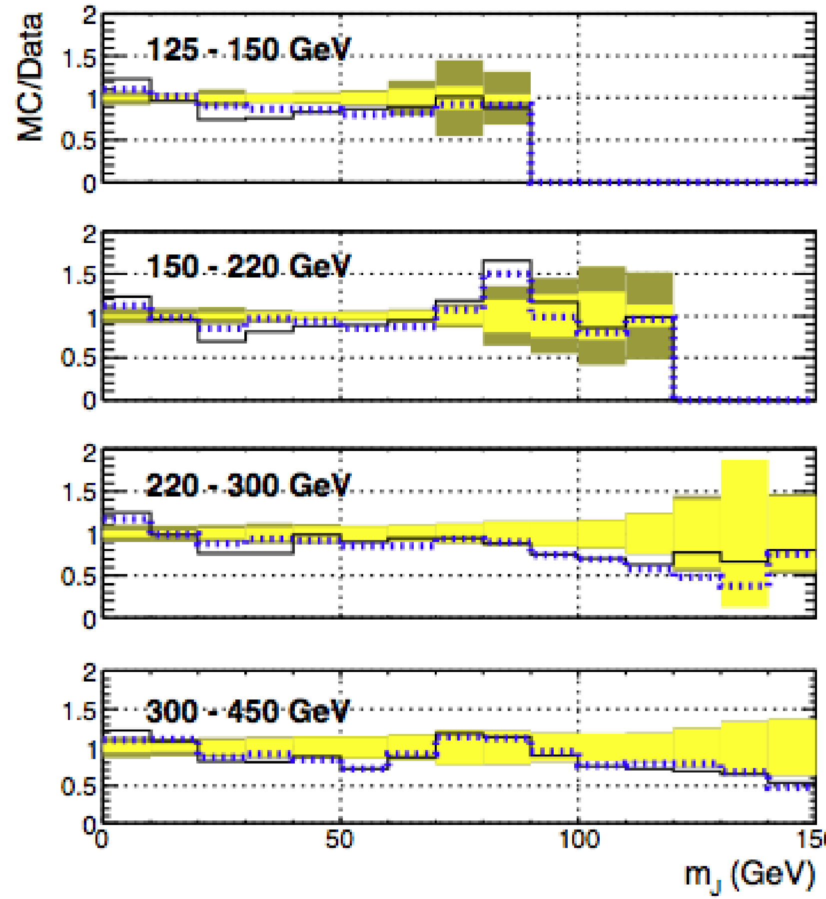


Similar qualitative behavior in V+jet as for di-jet, good agreement with both Madgraph-pythia and Herwig++

Unfolded Jet Mass, V+jet

Pruned CA8

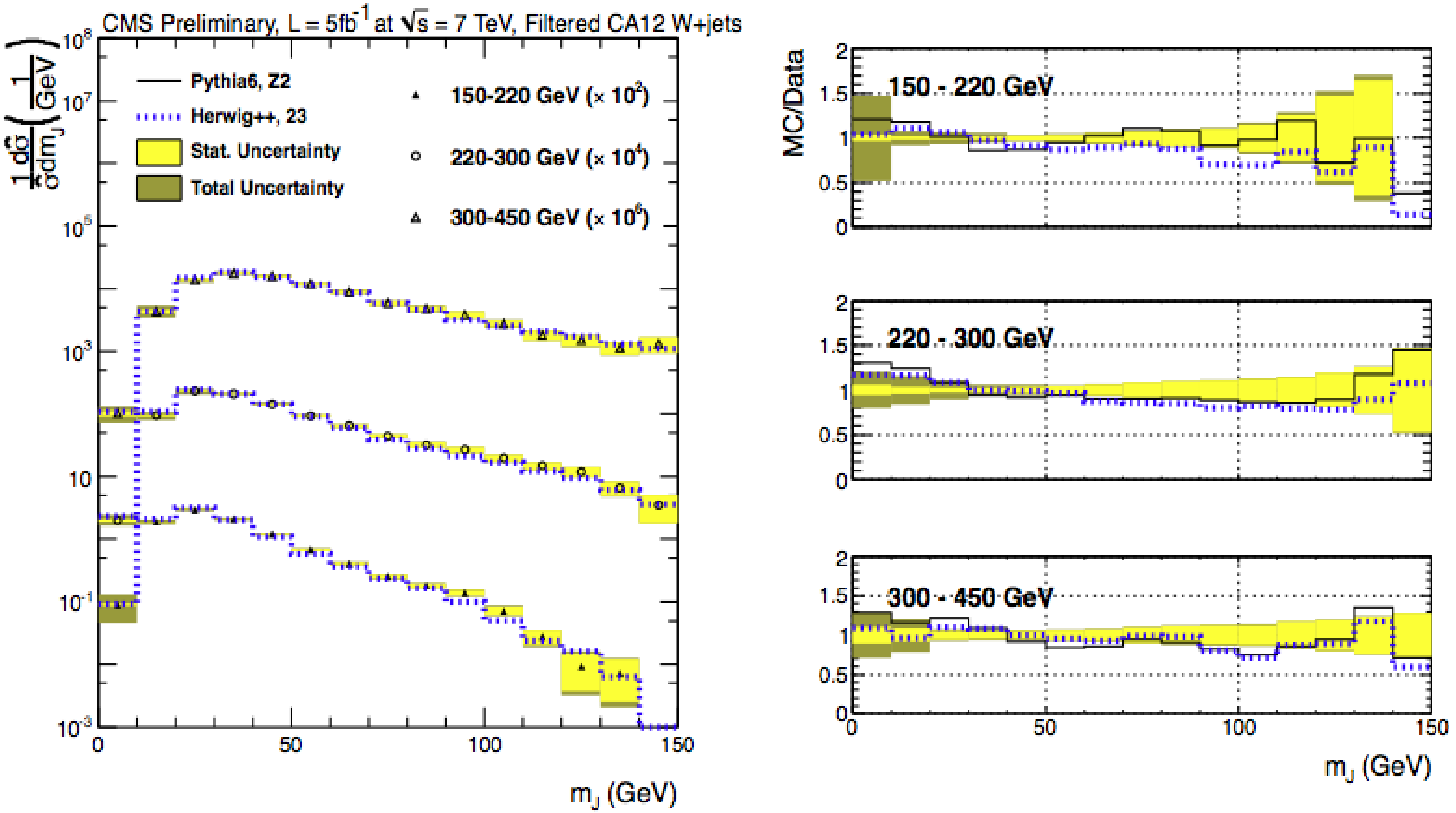




Also study Cambridge-Aachen groomed jet, relevant for ongoing CMS analyses

Unfolded Jet Mass, V+jet

Filtered CA12



First step toward use of BDRS in boosted Hbb analysis!

Systematic Uncertainties

Parton Shower theoretical uncertainty

→ difference between unfolding with two parton shower MC models: Pythia Z2 and Herwig++

Jet Energy Scale

→ Up and down variations as a function of η and p_T

→ Additional 1% uncertainty from scale from pruned W jets [1]

Jet Energy/Angular Resolution

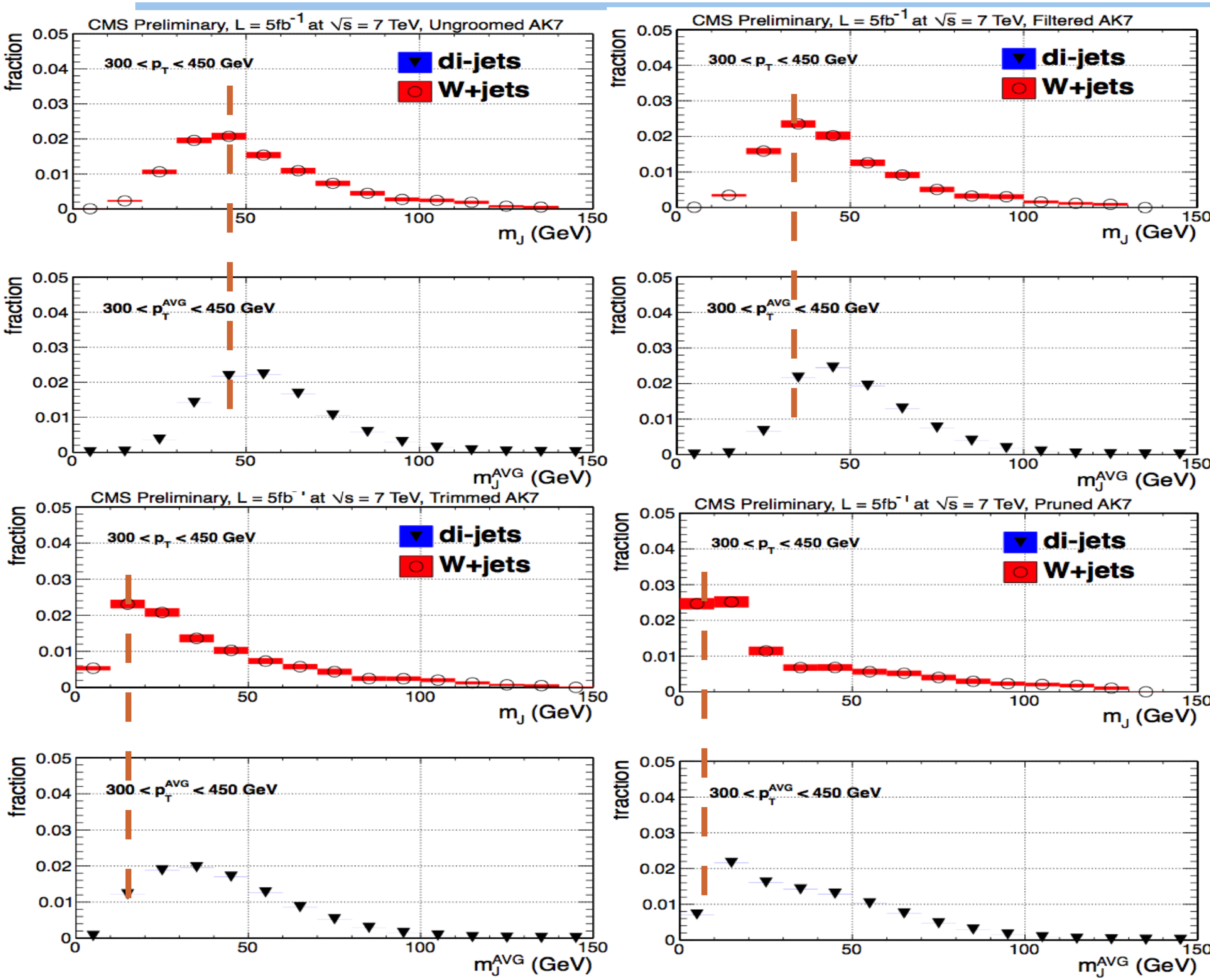
→ Vary the jet energy, η, and p_T resolution up and down by 10%

Pile-up

→ Vary the Min Bias cross-section by 8%

Bias corrections from unfolding procedure

Comparing di-jet and V+jet



Compare di-jet and V+jet distributions

pT bin: 300-450 GeV

Different jet parton content in di-jet and V+jet events reflected in jet mass: harder spectrum for gluon jets than quark one

Conclusions

- Grooming algorithms are a powerful tool at the LHC to increase sensitivity in heavy particle boosted searches
  → Very robust against PU, important for future LHC runs

- Ongoing program to validate the use of these tools at CMS
  Jet mass powerful observable, sensitive to scale, PU,UE and FSR
  → Good Data/MC agreement, simulation well models algorithm features

- More to come in the future:
  → Analysis of 2012 data with larger PU
  → Study of additional jet shape variables (subjettiness, etc.)
  → Apply to searches!

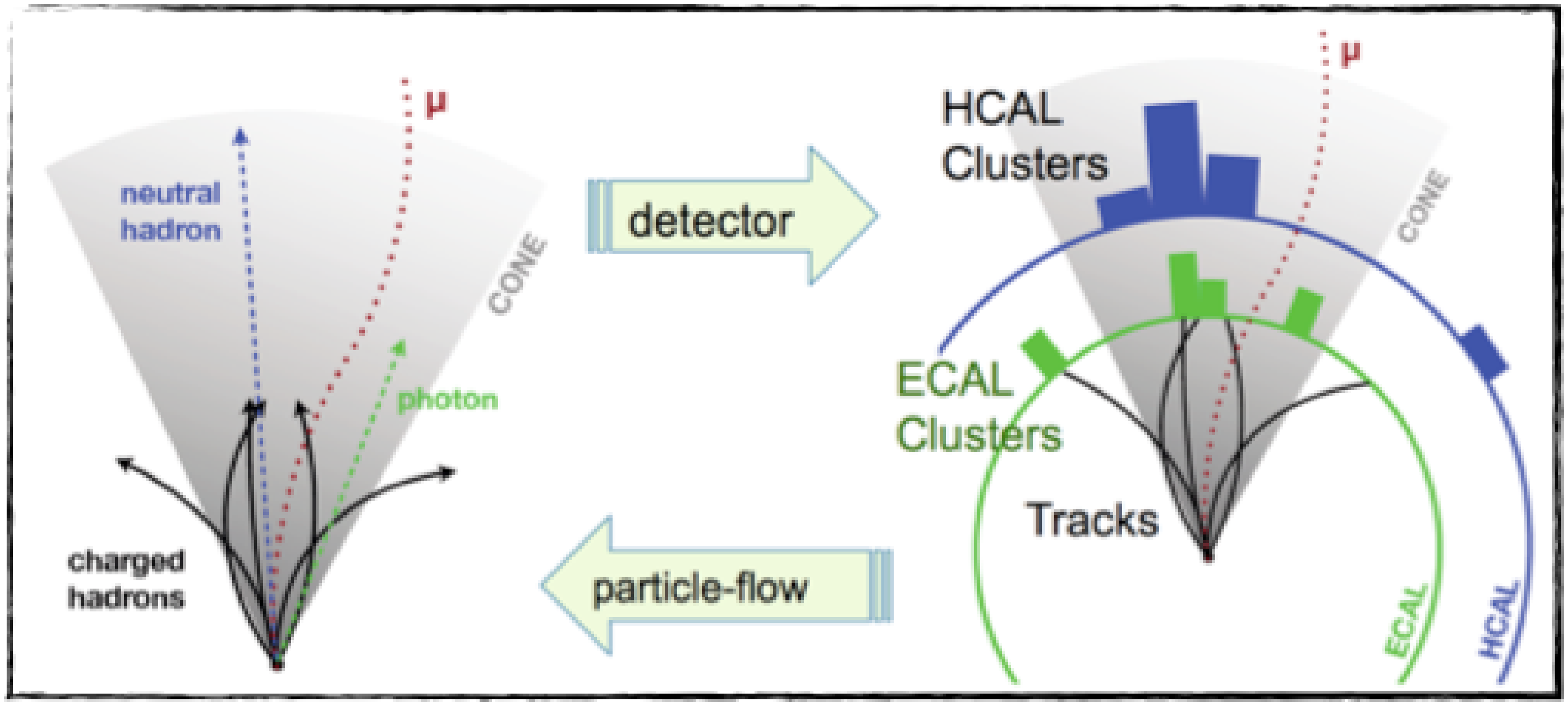
Backup Slides

Jet Scale Uncertainty

1. The uncertainty on the tracking is included in the PF cluster calibration, and hence, in aggregate, is included in the JES
2. We have the magnitude of the neutral components computed in data + MC, included in JES
3. Constituent fraction:
   - 3a. data/MC comparison of the W mass in data and madgraph+pythia6, which covers the "detector response part".
   - 3b. Spread of neutrals in data and MC, which we compute from the pythia/herwig comparison.

Particle Flow

Combines information from all CMS subsystems in an optimal way

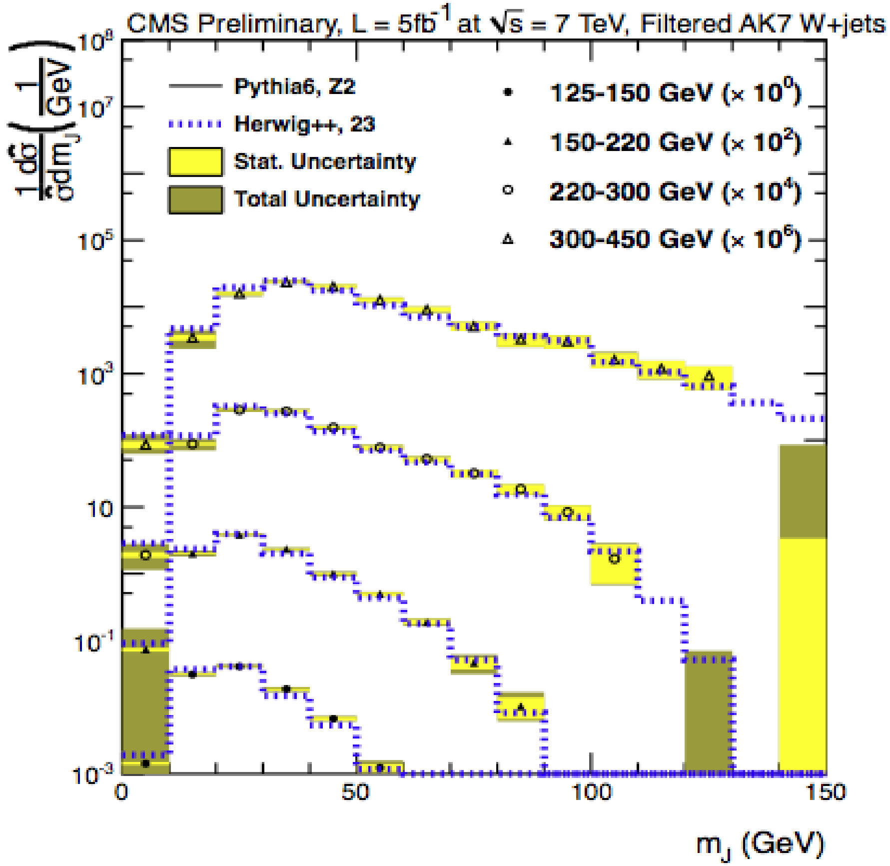


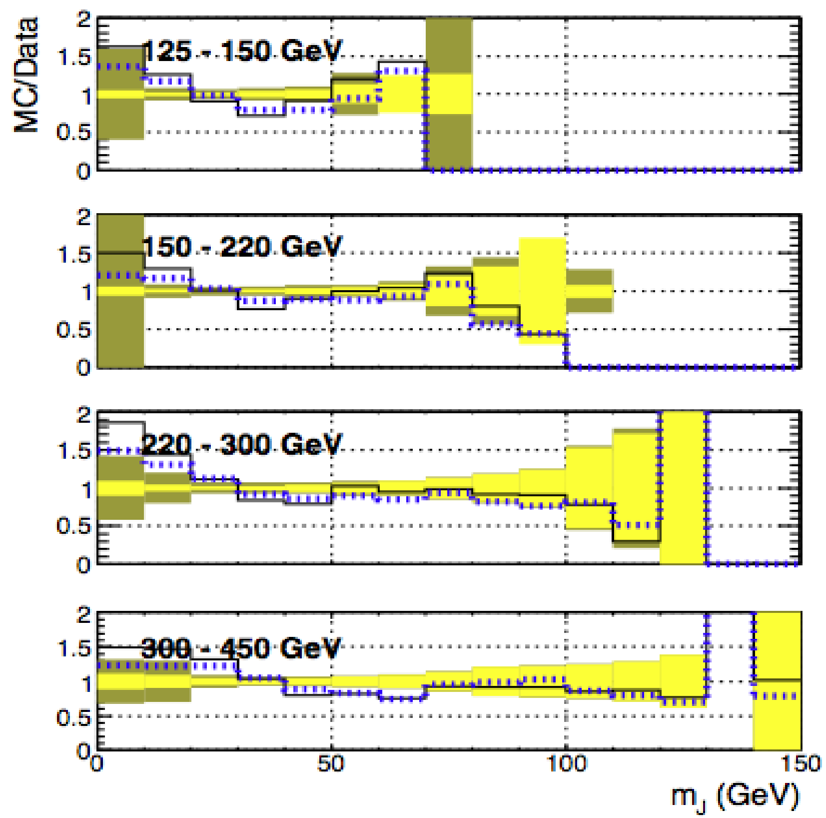
Algorithm returns a list of particles: e, μ, γ, charged hadrons, neutral hadrons

Collection of particles is used as inputs for jets, taus, MET, etc

Unfolded Jet Mass, V+jet

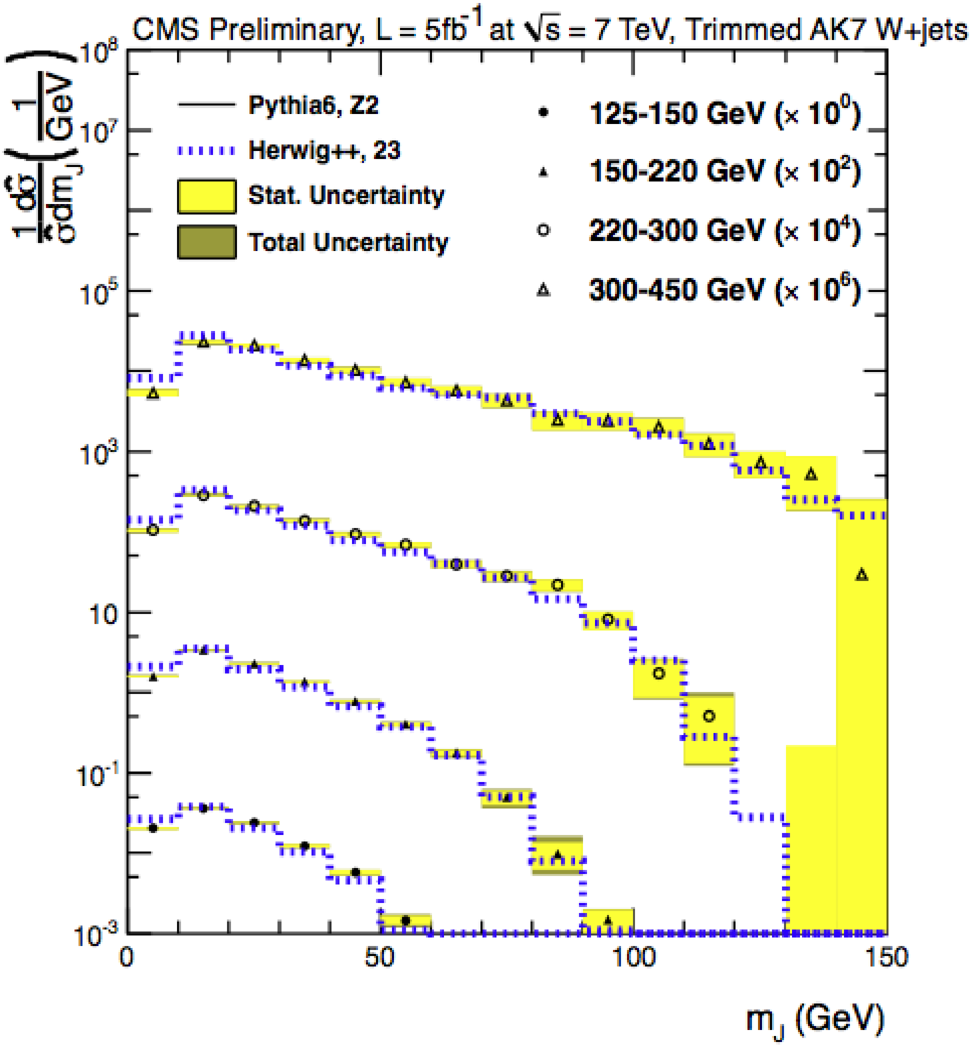
Filtered AK7

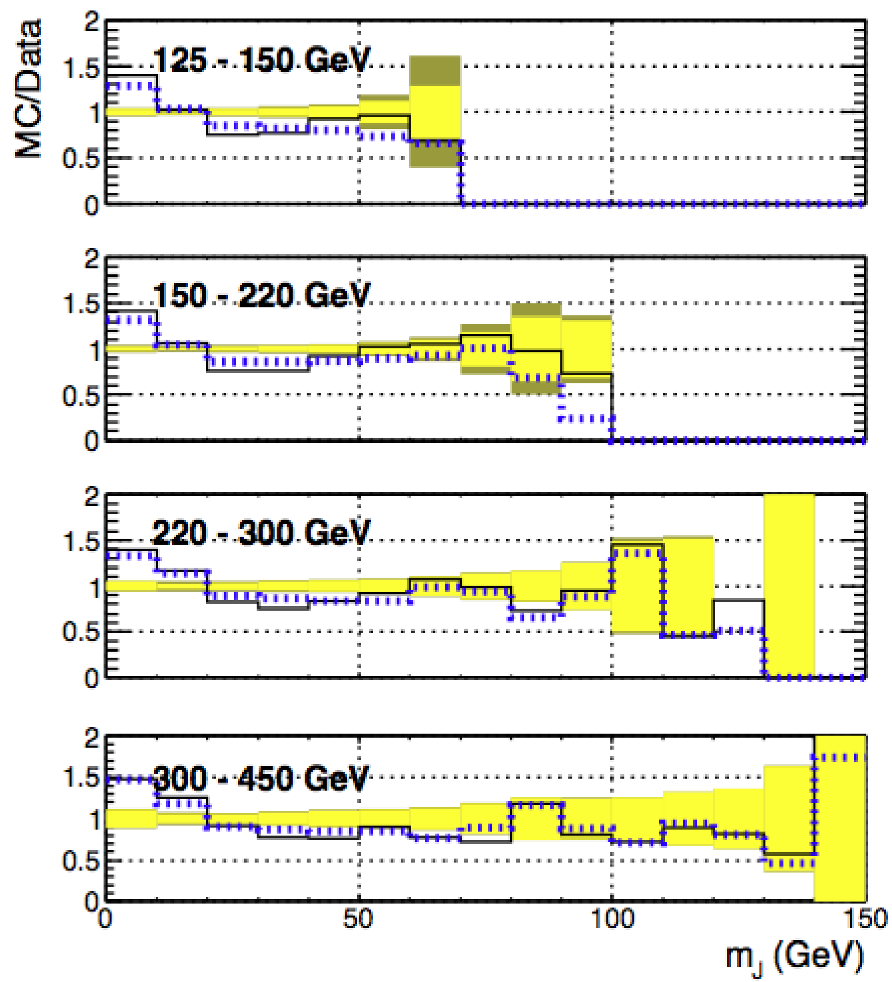




Unfolded Jet Mass, V+jet

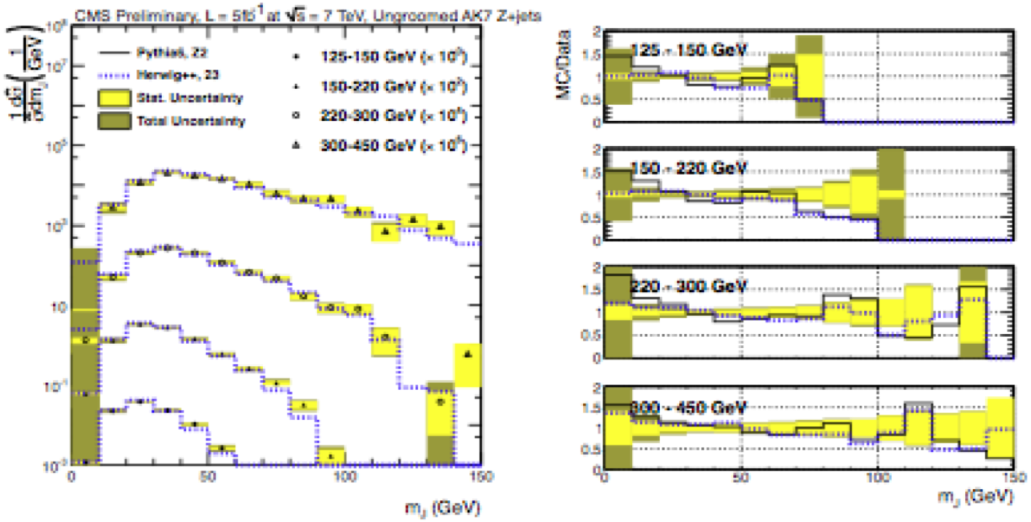
Trimmed AK7



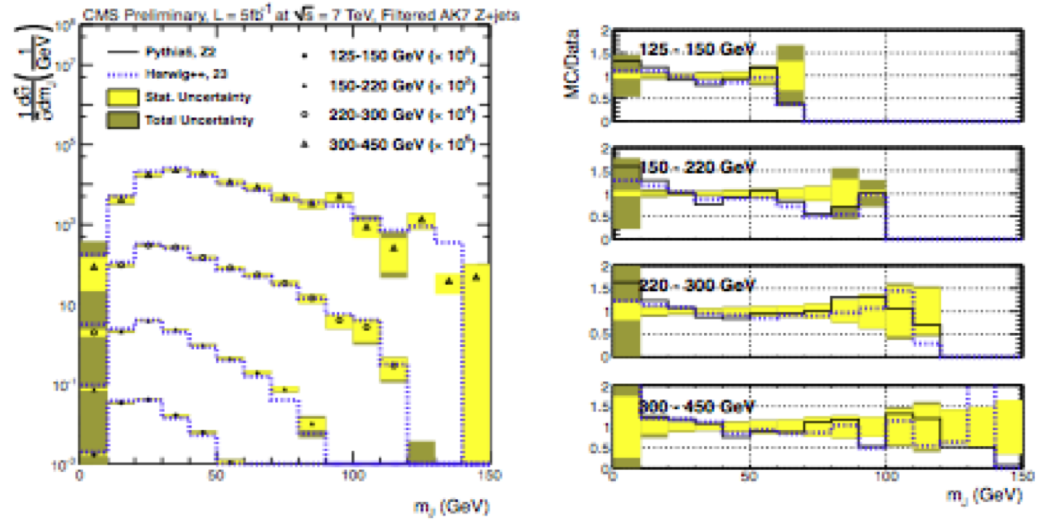


Unfolded Jet Mass, Z+jet

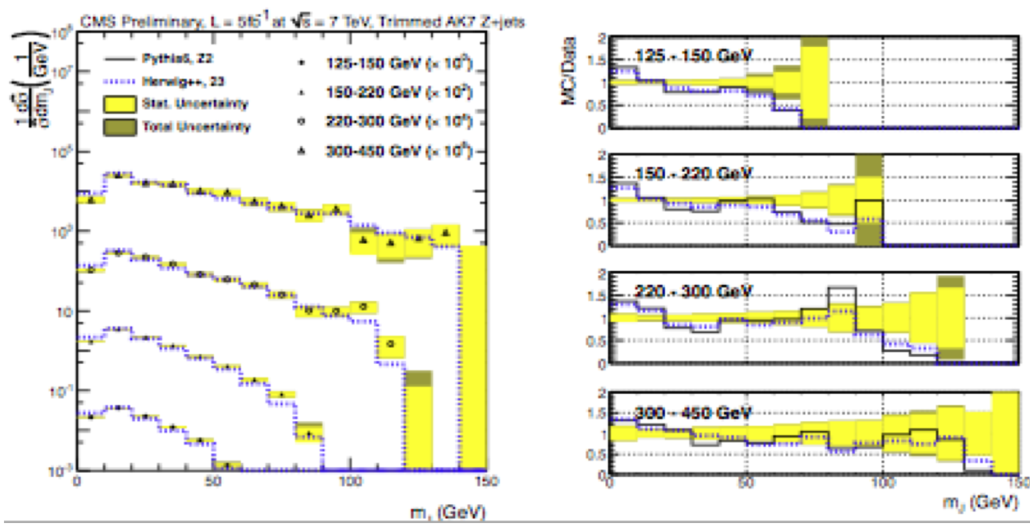
AK7



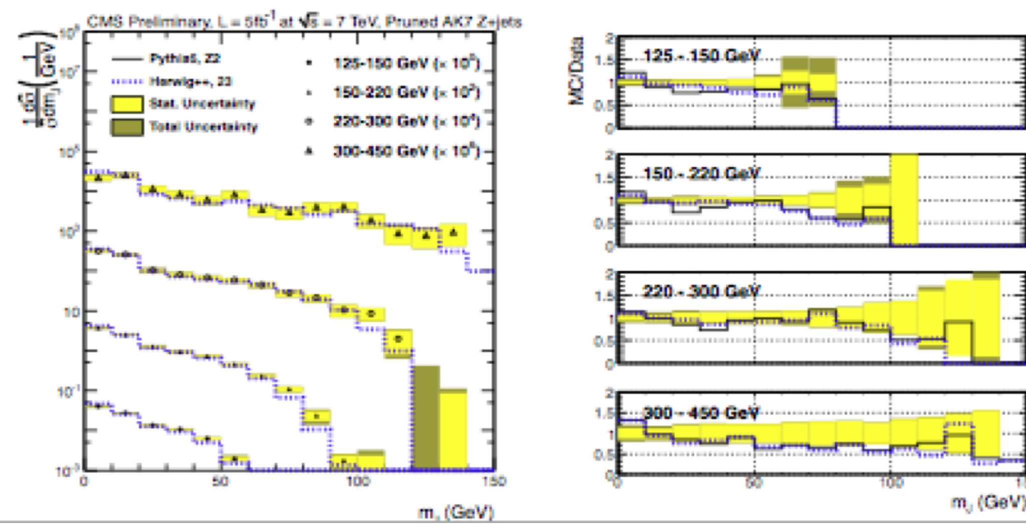
AK7, Filtered



AK7, Trimmed

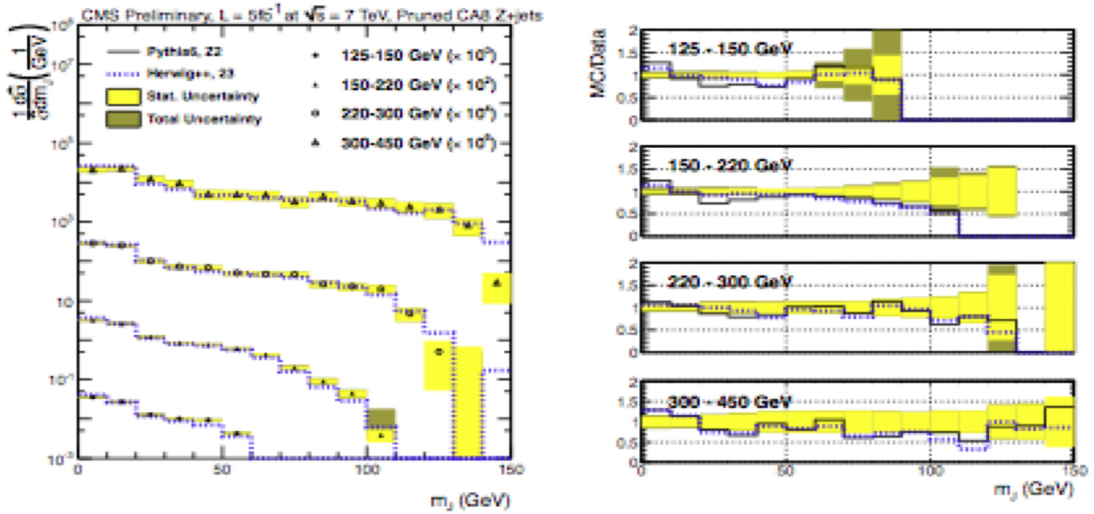


AK7, Pruned



Unfolded Jet Mass, Z+jet

CA8, PRUNED



CA12, FILTERED