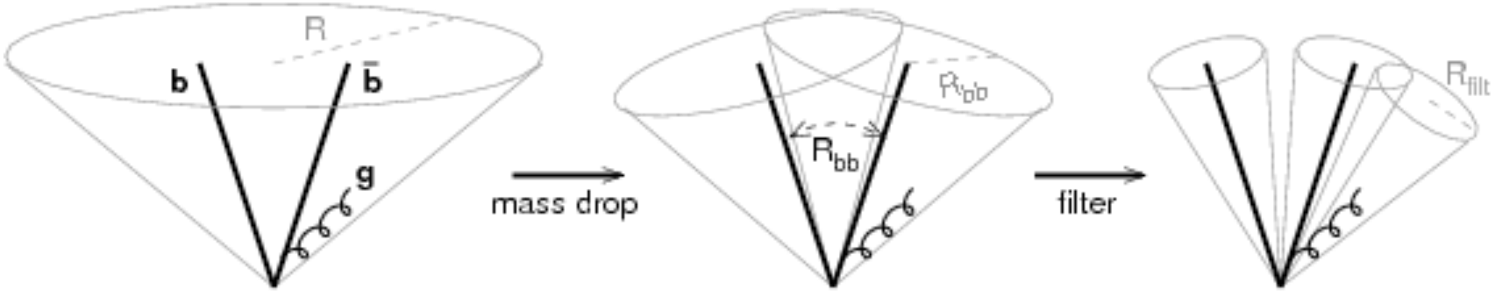
CMS Measurements of Jet Structure & Properties

Kalanand Mishra

Fermilab, CMS Collaboration



US ATLAS Hadronic Final State Forum, December 3, 2012

Outline

- Introduction and motivation

- Jet calibration in CMS
  - effect of pileup and pileup subtraction, uncertainties
  - calibration of groomed jets

- Jet mass distributions in di-jet and W/Z+jet events
  - comparisons of raw data and simulation

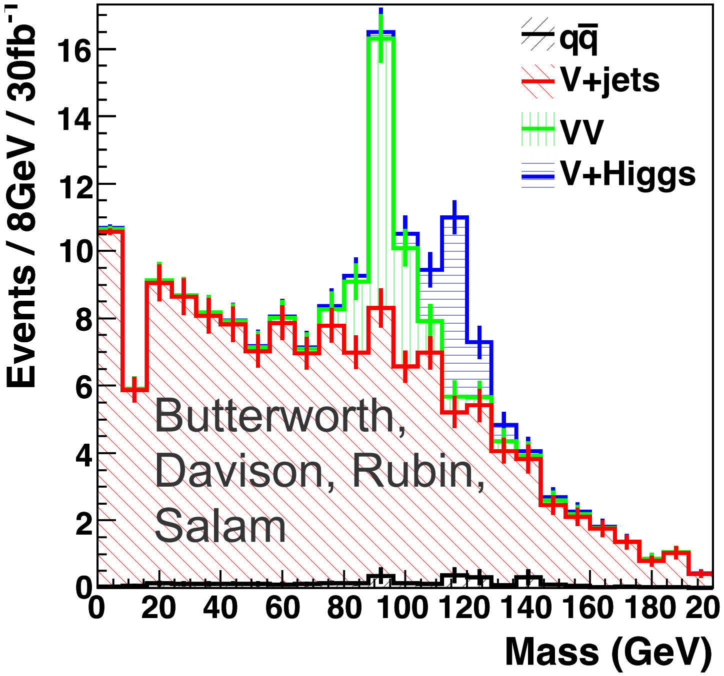
- Unfolded distributions
  - unfolding procedure and systematic uncertainties
  - comparison between di-jet and V+jet

- Summary

Jet mass and substructure

- Jet substructure can be used to improve sensitivity of hadronic decays of boosted heavy particles such as Higgs, W/Z, and top

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



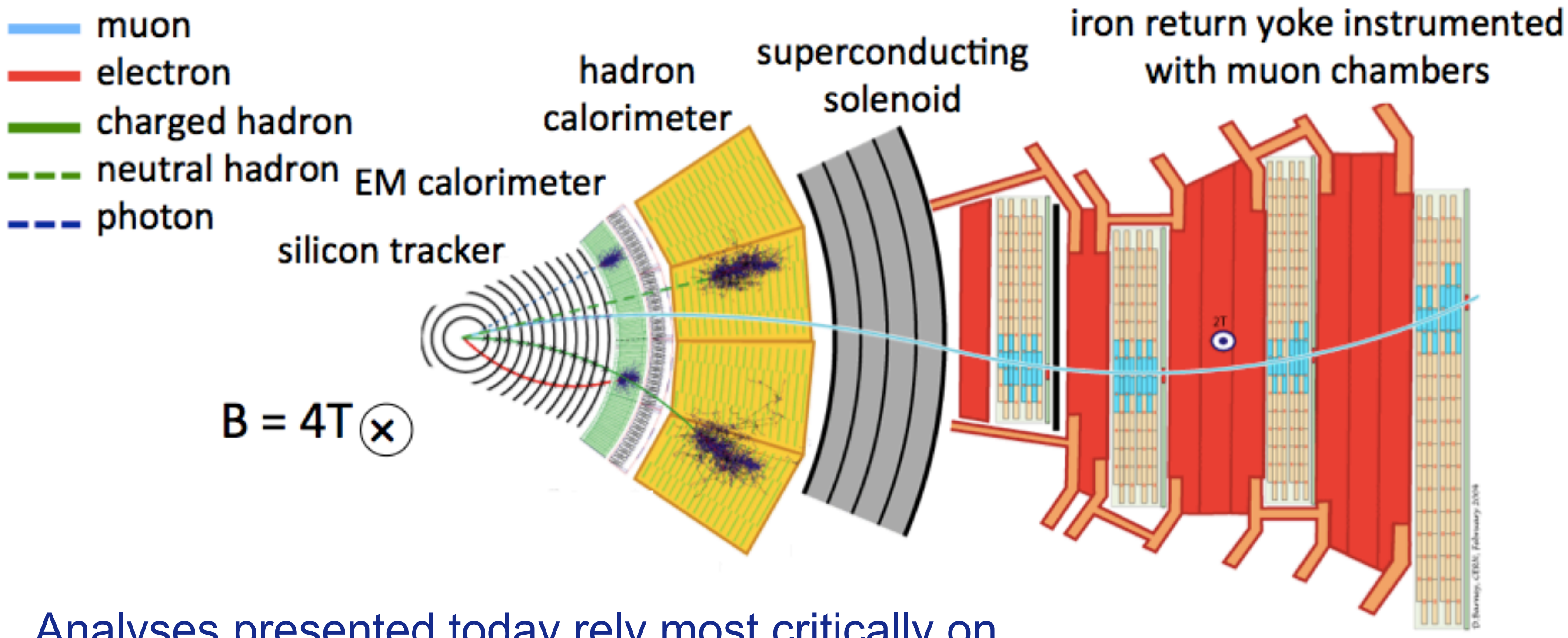
Butterworth, Davison, Rubin, Salam

arXiv: 0802.2470

- <u>Analyze jet mass properties with CMS 2011 data</u>
  - -inclusive & systematic comparison of jet grooming algorithms

- Use di-jet and V+jets events as complimentary probes of gluon- and quark-enriched environments, respectively

CMS detector



Analyses presented today rely most critically on

- **electrons**: tracks matched to clusters in EM calorimeter
- **muons**: minimum ionizing tracks, penetrate deep into muon system
- **jets / H_T**: constructed with combined tracking + calo info
- **MET**: constructed with combined tracking + calo info, hermetic detector

Jets in CMS

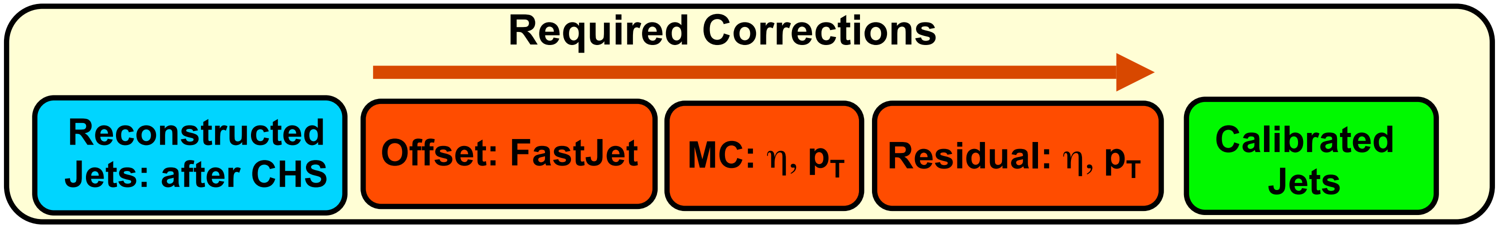
- Use charged hadron subtracted (CHS) Particle Flow jets
  - Remove all charged hadrons not from the primary vertex

- Neutrals are subtracted via jet area ρ correction where ρ is energy density computed with the KT6 algorithm
  - Active area used in all cases including groomed

- Non-linearities in η and p_T are corrected for using data & MC
  - AK5, AK7 jets have dedicated corrections while other large radii jets use AK7 corrections

- Jet finding & grooming algorithms implemented via FastJet3

Jet energy calibration: overview

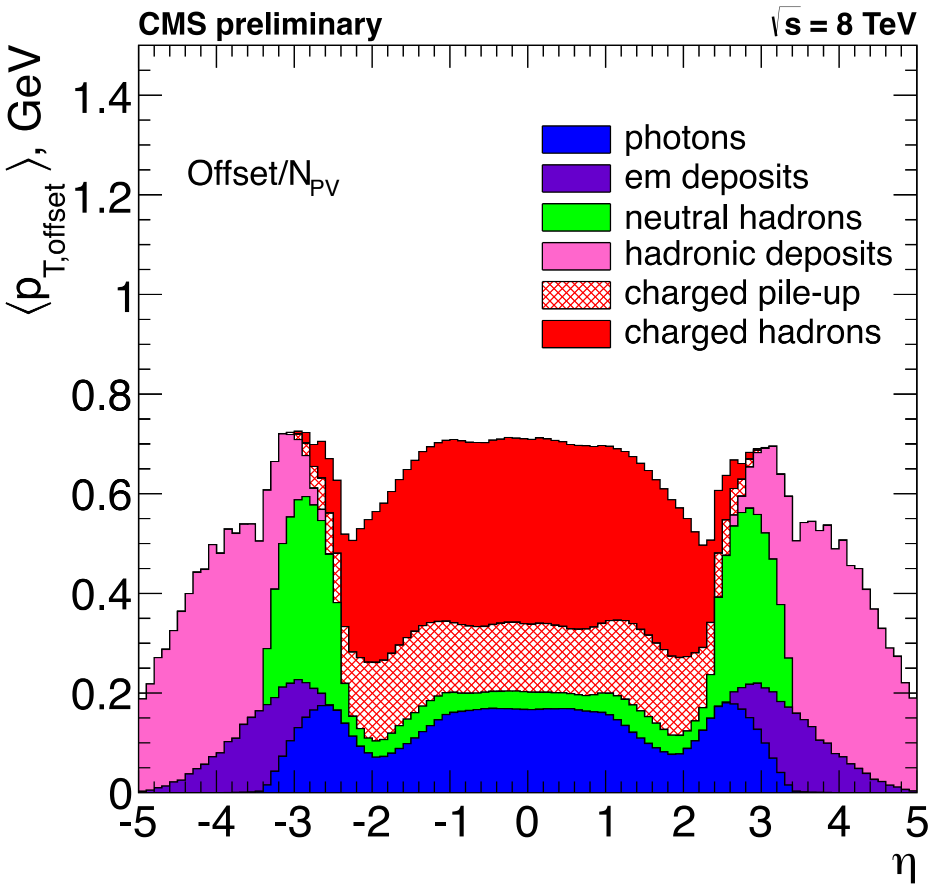


☑ Factorization facilitates the use of data-driven corrections

- Breaking the correction into pieces that are naturally measured in collider data:
  - **Offset**: pile-up and noise measured in zero-bias events.
  - **MC**: jet response vs. η, P_T using MC truth.
  - **Residual**: jet response vs. η, P_T using dijet balance and γ/Z+jet in data.

In CMS the most widely used jet is anti-k_T 0.5 (0.7 for QCD measurements). Jet substructure studies done with anti-k_T 0.5, 0.7, 0.8 with various grooming techniques.

Pileup contribution to jet energy



✦Pileup (PU) measured with Zero Bias data

- Most charged hadrons can be associated to pileup vertices and removed

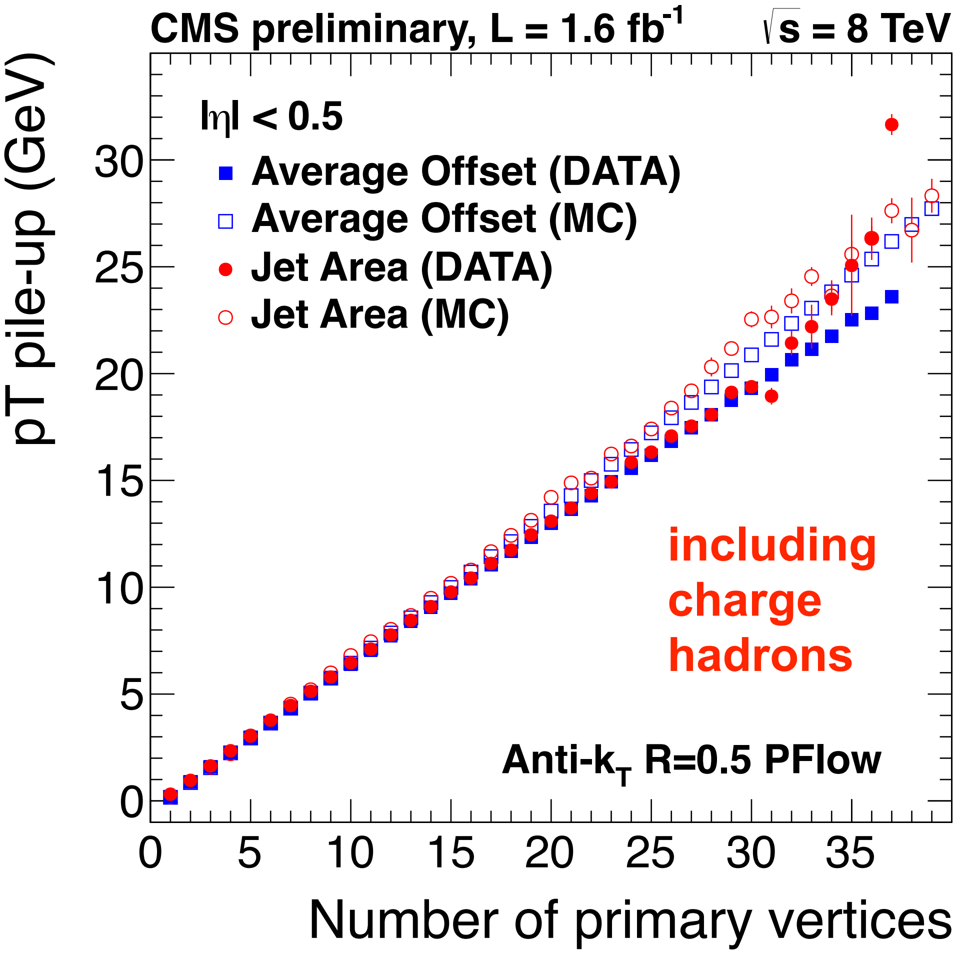
- Part that can be removed is labeled “charged hadrons”

- Part that remains as PU needs to be subtracted

  PU density x effective area

  (FastJet-ρ)

Pileup subtraction: $\rho_{FastJet}$ x Area



- PU density depends on the # of primary vertex in the event
  - Reweight pileup Poisson mean in MC to match data
  - Poisson mean determined from measured luminosity and MinBias cross section

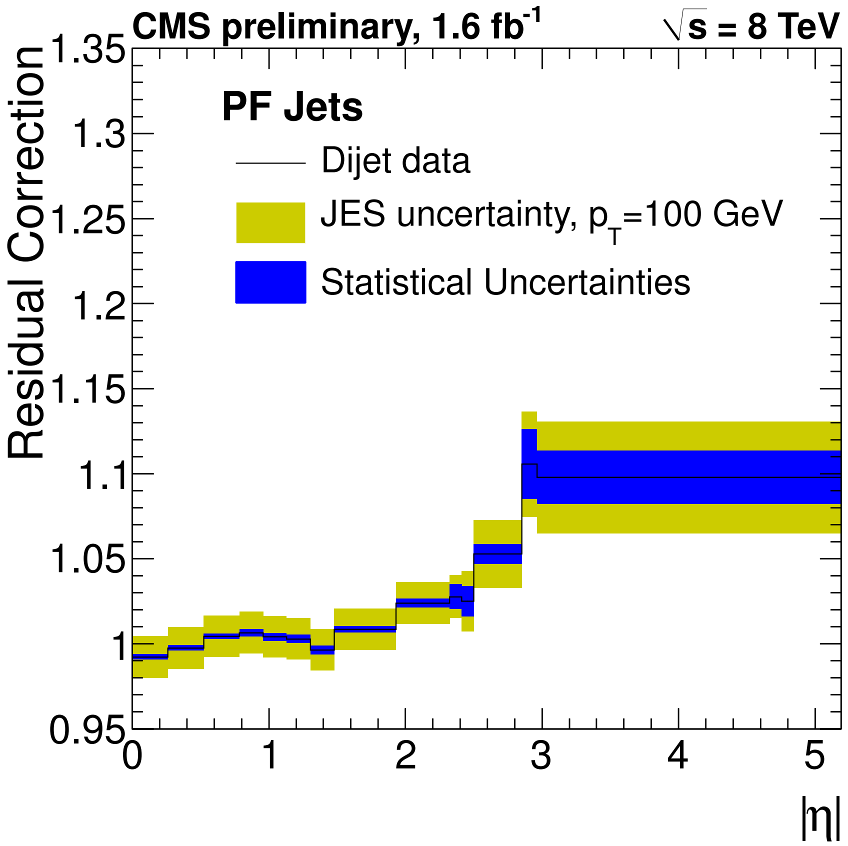
Residual correction in data: η & p_T dependence

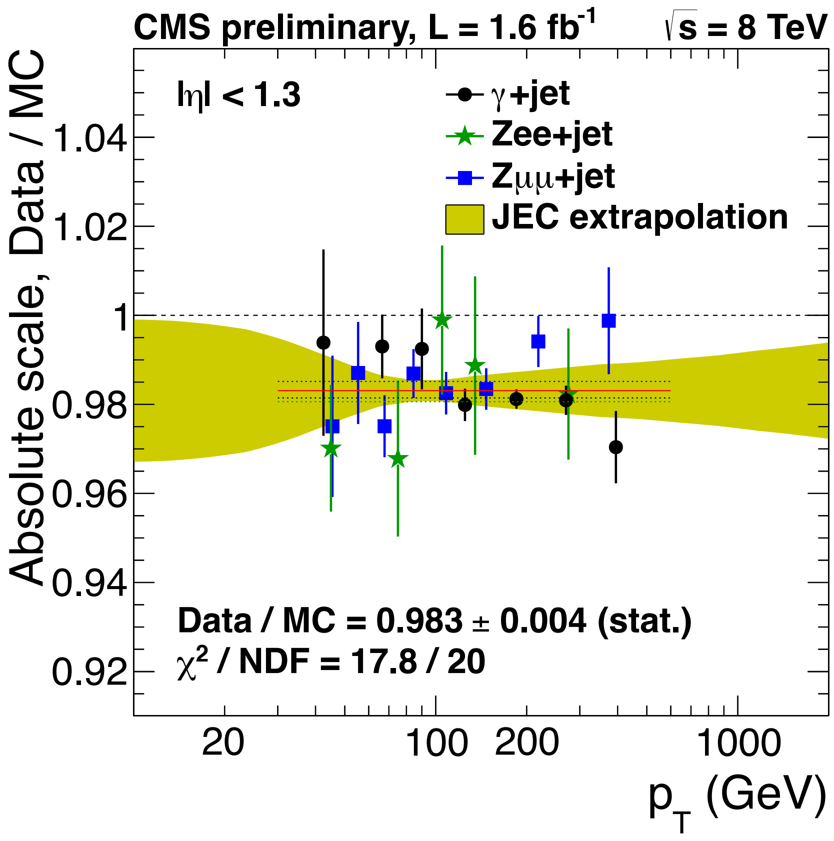
η dependence

- Calibrated using dijet events

No p_T dependence observed

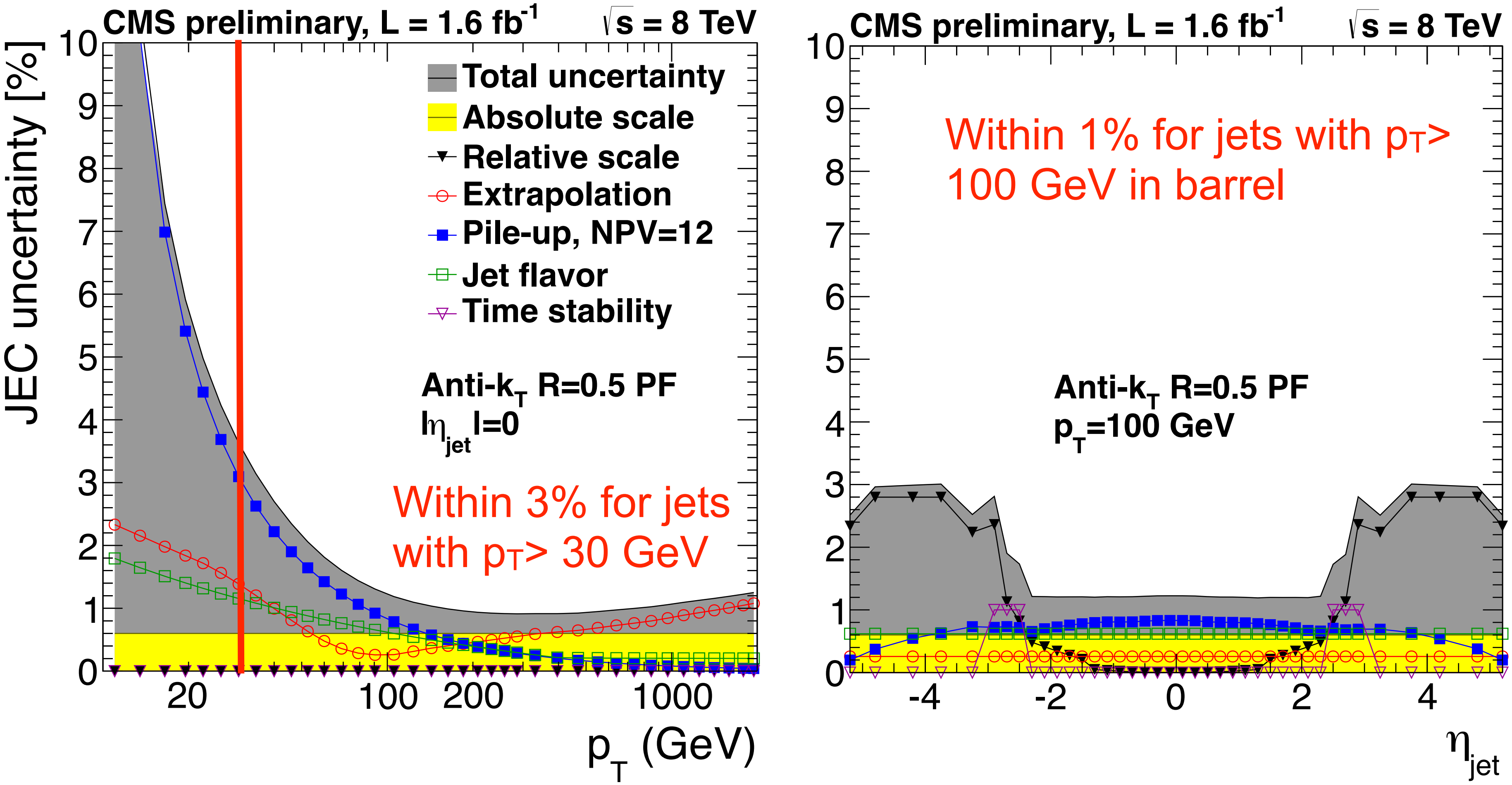
- Using Z/γ+jet events
- Use a single flat scale factor



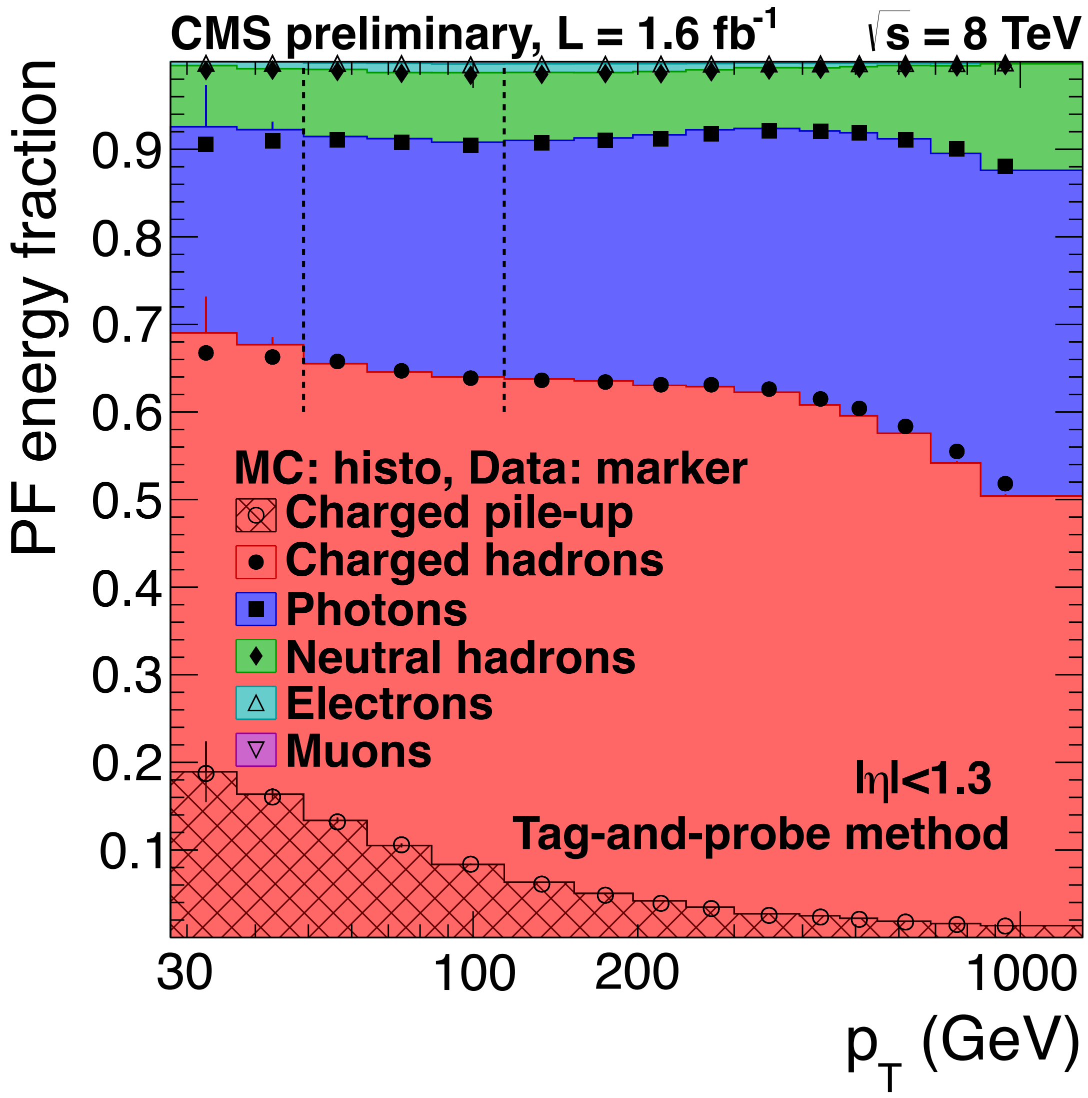


Correction uncertainties

Uncertainties in 2012 data comparable to 2010, 2011.



Jet composition: achieved a good understanding



Jet composition agrees well between Data & MC

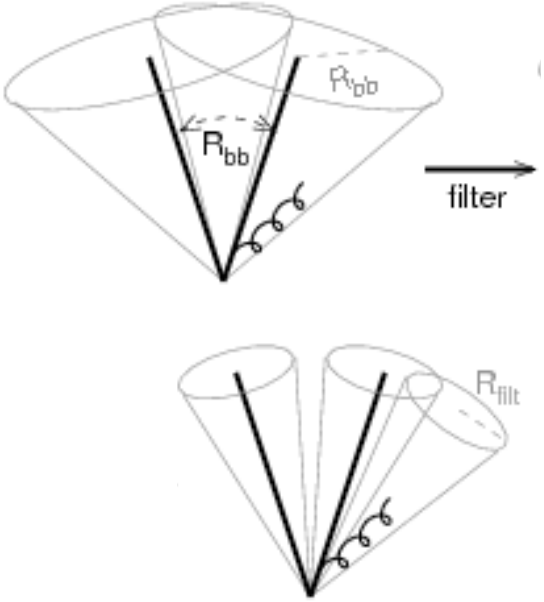
- consistent with small residual jet energy scale difference at the level of 1−2%.

Jet substructure algorithms I

✦Study jet mass properties with three grooming techniques

- Filtering: arxiv: 0802.2470 Butterworth, Davison, Rubin, Salam
- Trimming: arxiv: 0912.1342 D. Krohn, Jesse Thaler, LianTao Wang
- Pruning: arxiv: 0903.5081 Steve Ellis, Chris Vermilion, Jon Walsh
- This round of analysis uses default parameters from each of the references.

✦Filtering

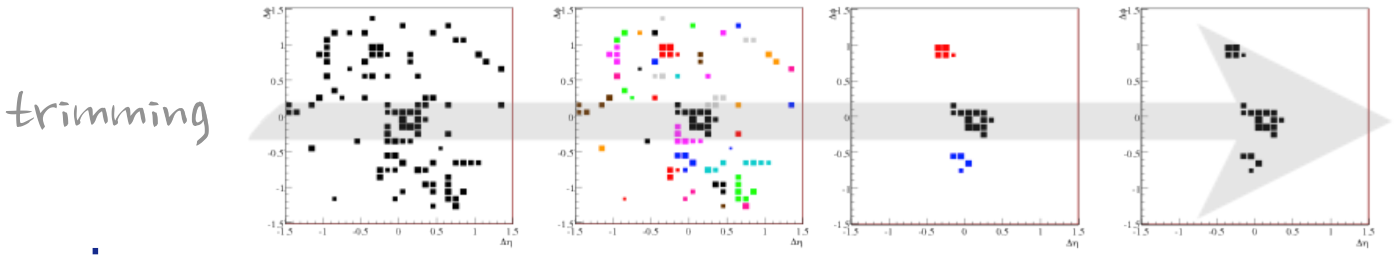


- reclustering jet constituents with smaller radius, R_{filt}, keeping n_{filt} hardest sub-jets
- default parameters: $R_{filt} = 0.3$, $n_{filt} = 3$
- sub-jet clustering algorithm: Cambridge-Aachen

Jet substructure algorithms II

✦Trimming

- reclustering with smaller radius, R_{filt}, keeping sub-jets with a fraction, $pT_{frac,min}$, of original jet p_T
- default parameters: $R_{filt} = 0.2$, $pT_{frac,min} = 0.03$
- sub-jet clustering algorithm: kT



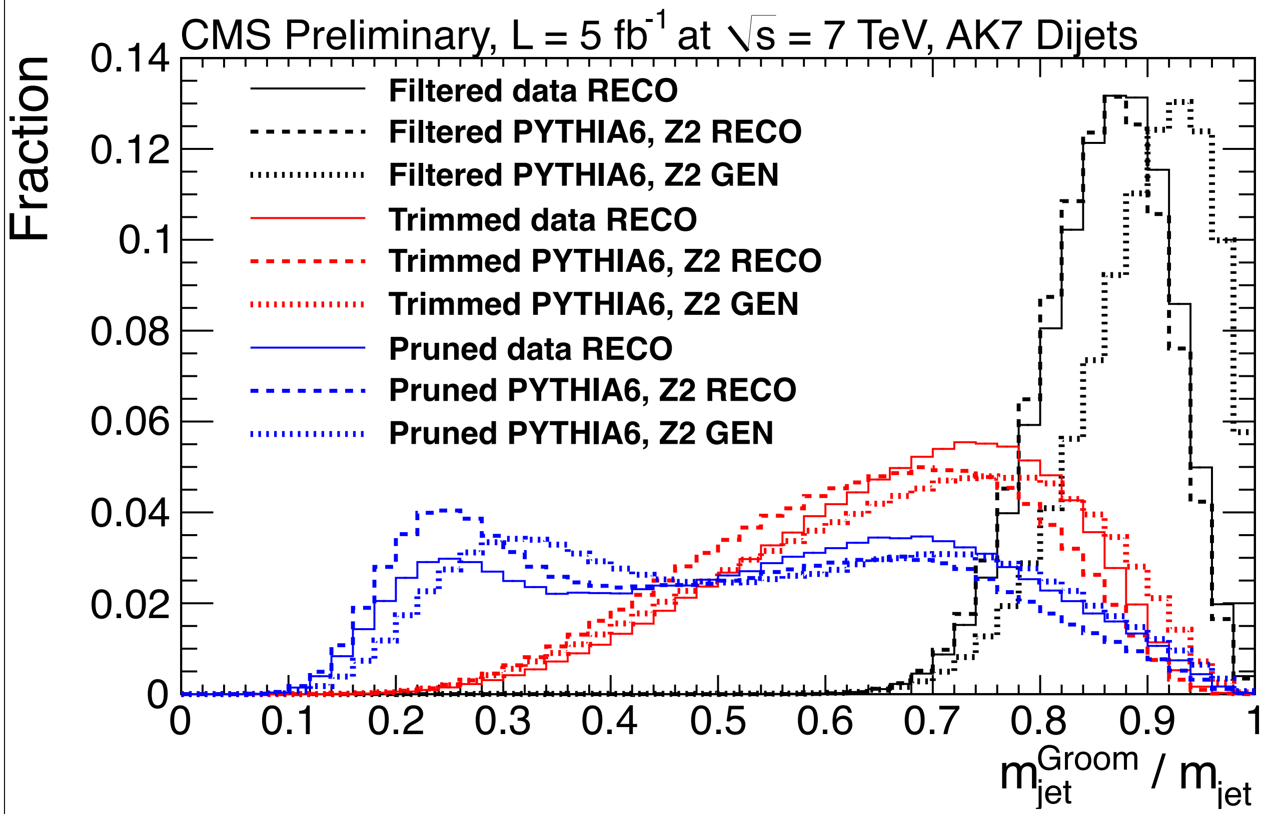
✦Pruning

- reclustering with sequential recombination algorithm, veto soft and large-angle recombinations between pseudojets i and j
  - veto: $\Delta R_{ij} > r_{cut} \times 2m/p_T$; $z = \min(pT_i, pT_j)/pT_{i+j} < z_{cut}$
- default parameters: $z_{cut} = 0.1$, default $r_{cut} = 0.5$
- sub-jet clustering algorithm: Cambridge-Aachen

Jet mass for groomed jets

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

Ratio of the groomed to ungroomed jet mass, encapsulates behavior of each grooming algorithm



Comparison of grooming algorithms at particle level (GEN), reconstructed simulation (RECO) level, and in data

Pruning is the most aggressive, filtering is the least aggressive

Use V+jet events to study jet properties

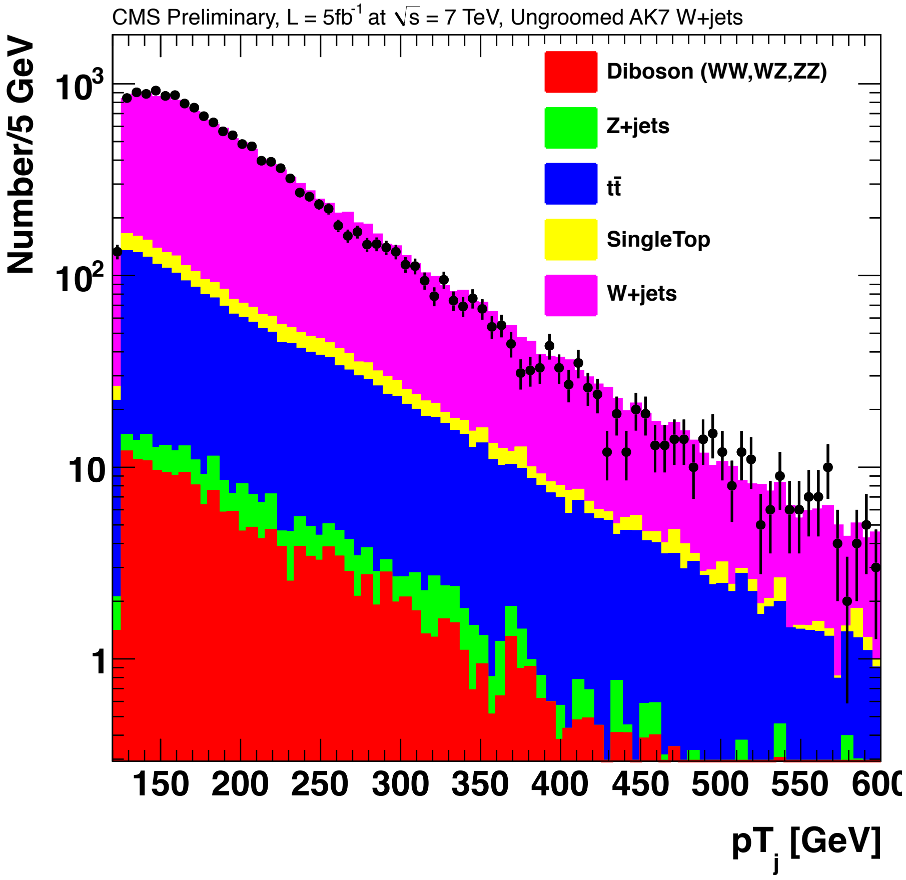
W(eν)	W(μν)	Z(ee)	Z(μμ)
$p_T > 120$ GeV			
$m_T > 50$ GeV, MET > 50 GeV		$m_Z = [80,100]$	
$pT_e > 80$ GeV WP ($\varepsilon = 70\%$) ($iso_{rel} < 0.05$)	$pT_\mu > 80$ GeV $iso_{rel} < 0.1$ μ quality cuts**	$pT_e > 20,20$ GeV WP ($\varepsilon = 95\%$)	$pT_\mu > 30,30$ GeV $iso_{rel} < 0.1$ μ quality cuts**
$\lvert d\Phi_{MET,j} \rvert > 0.4$ [3]	$\lvert d\Phi_{V,j} \rvert > 0.4$ [3]		
Jet selections			
$p_T > 125$ GeV, $\lvert\eta\rvert < 2.5$, $\lvert d\Phi_{V,j}\rvert > 2.0$, $\lvert dR_{lep,j}\rvert > 1.0$			

Kinematic selections are chosen to compliment trigger strategy in a specific boosted phase space.

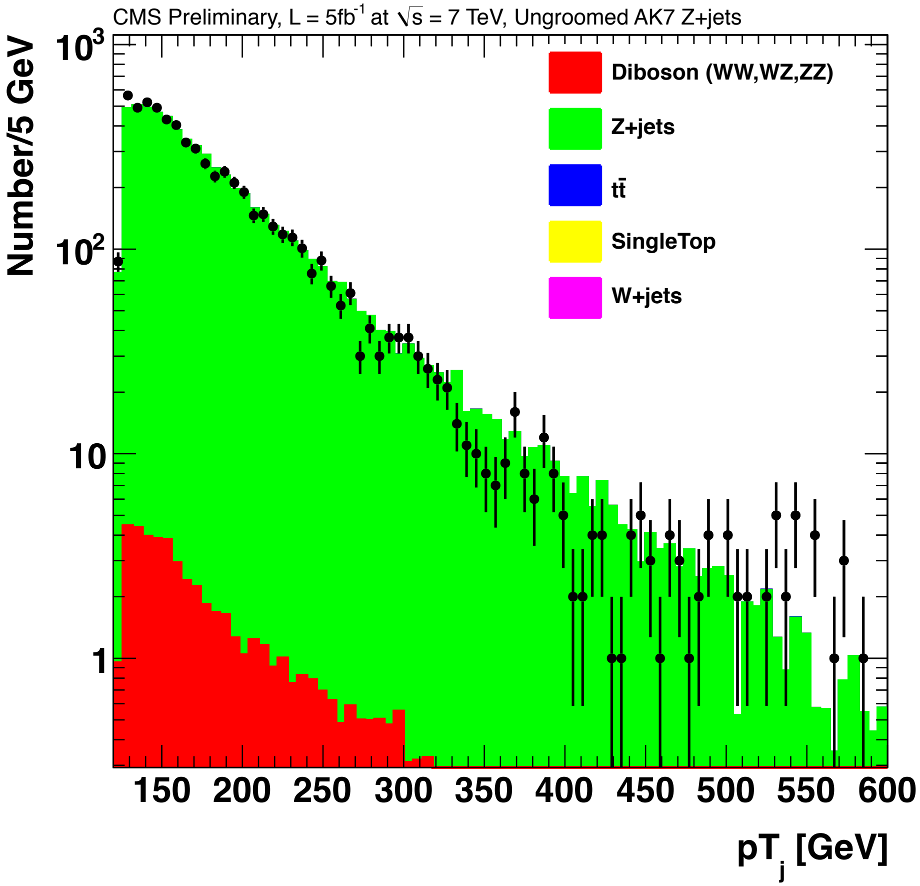
** nMatches > 0, χ^2/ndof < 10, nPixelHits > 0, nTrackerHits > 10, nMuonHits > 0

V+jet sample composition

W+jets

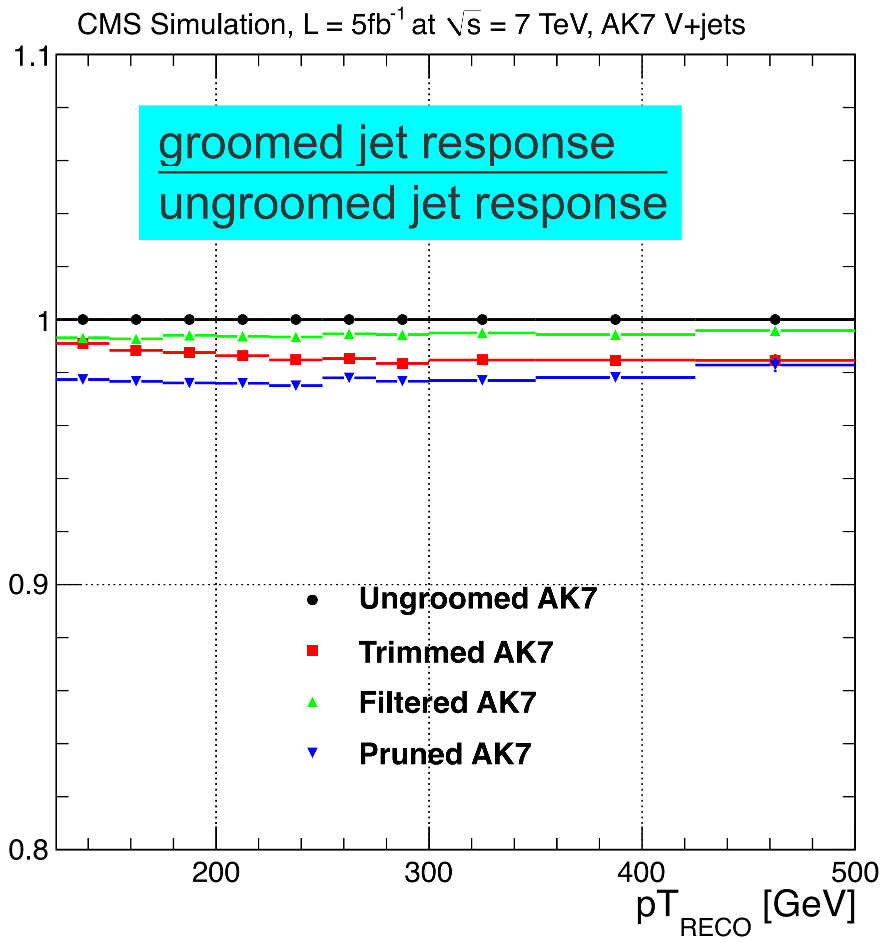


Z+jets

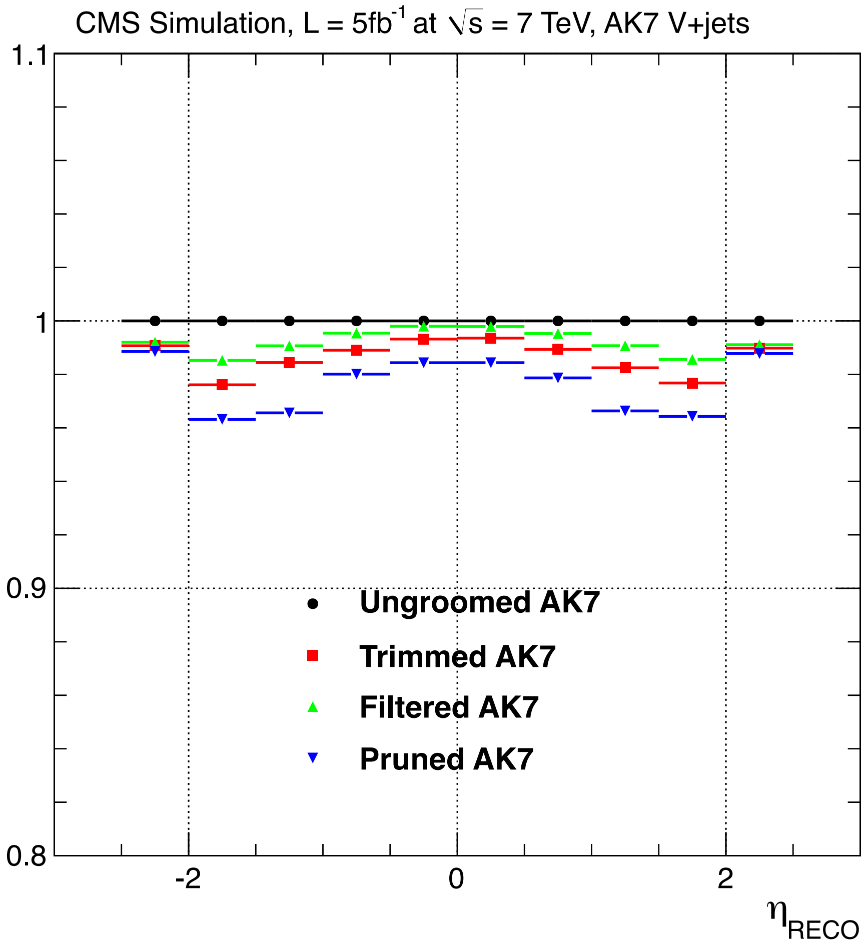


Jet p_T response for groomed jets

Groomed jet response within a few % of ungroomed case.



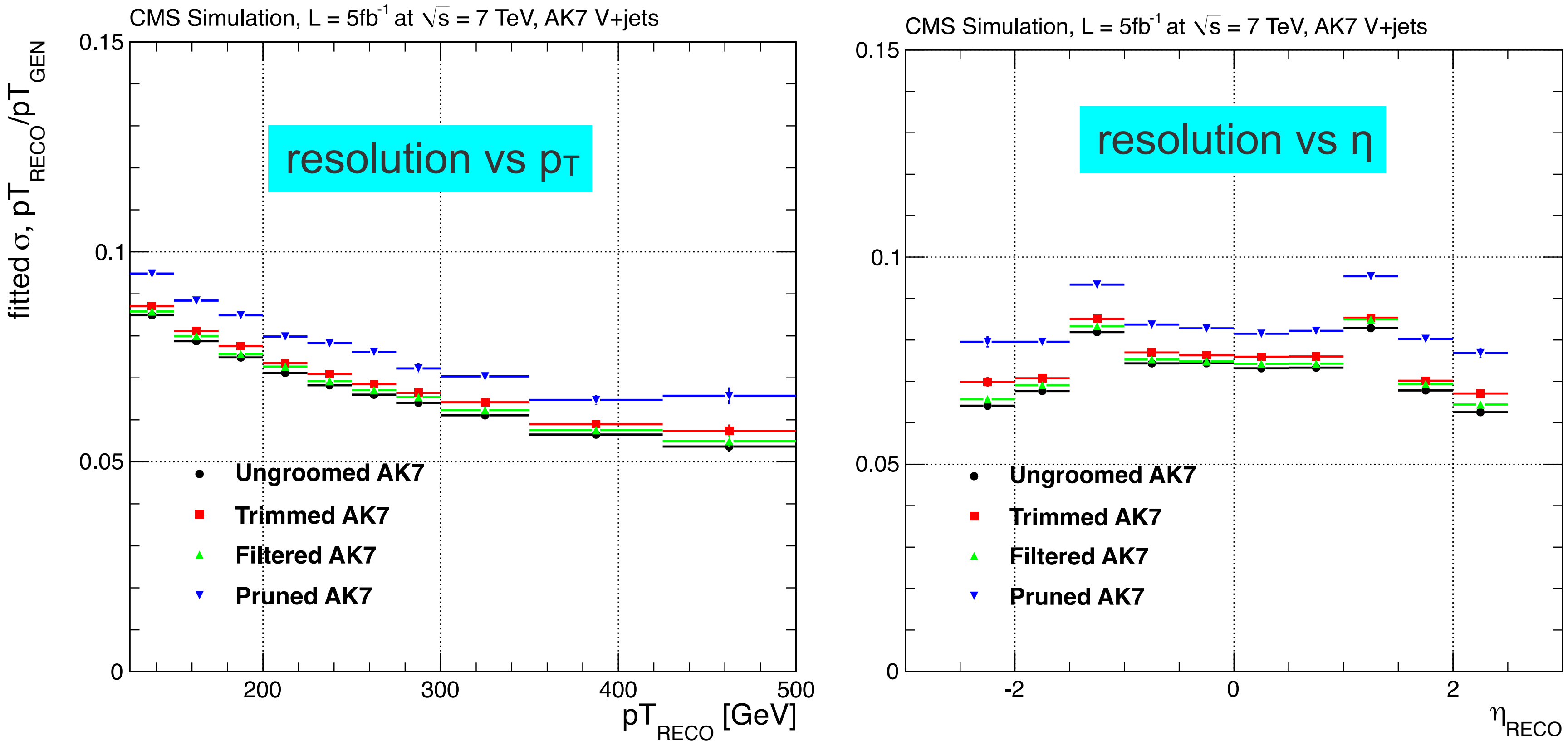
$$\frac{\text{groomed jet response}}{\text{ungroomed jet response}}$$



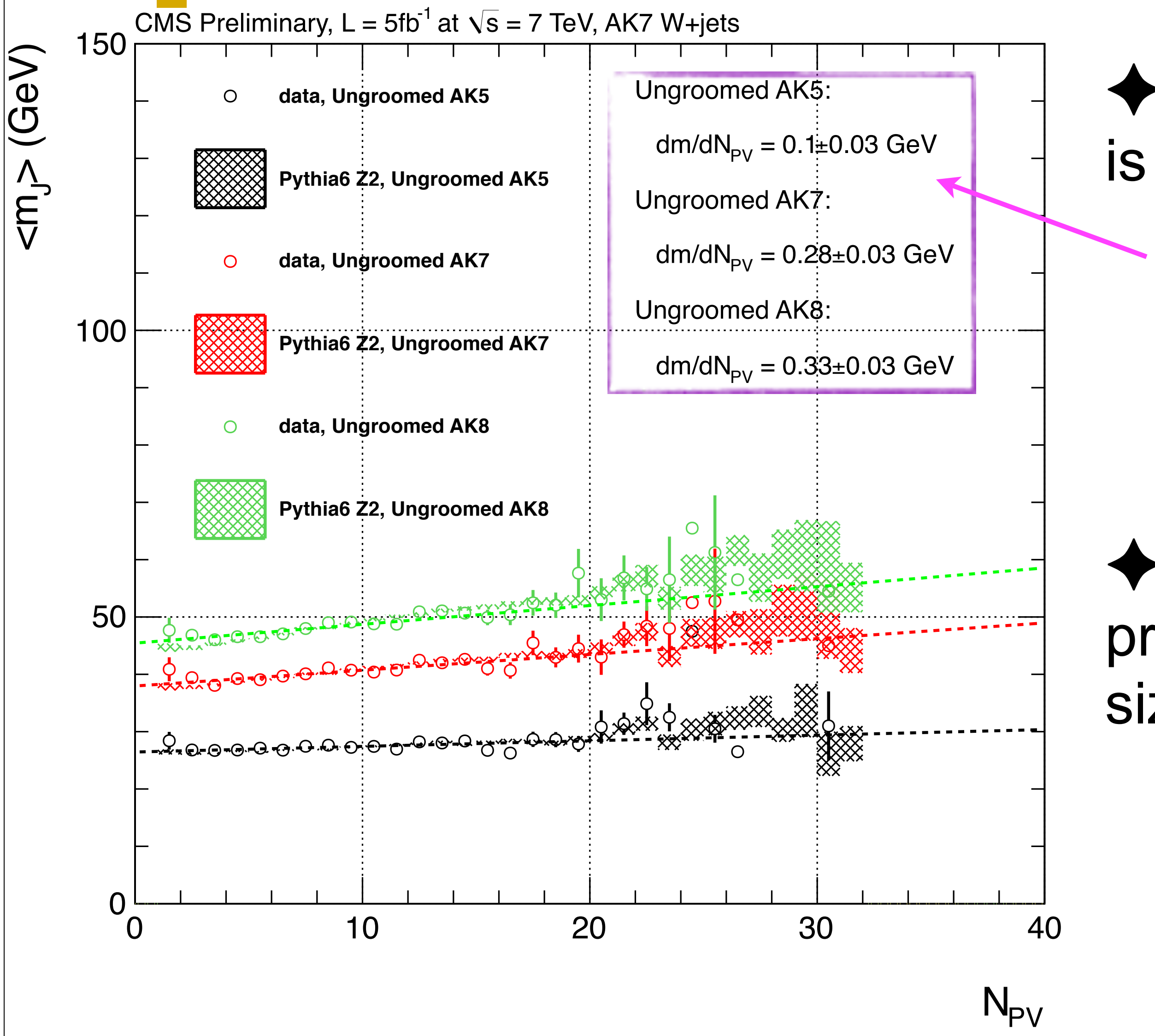
Jet p_T resolution for groomed jets

Jet p_T resolution for various grooming algorithms also shows good agreement to within a few percent.

- **Groomed jet p_T resolution degraded slightly.**



Performance versus pileup by jet size



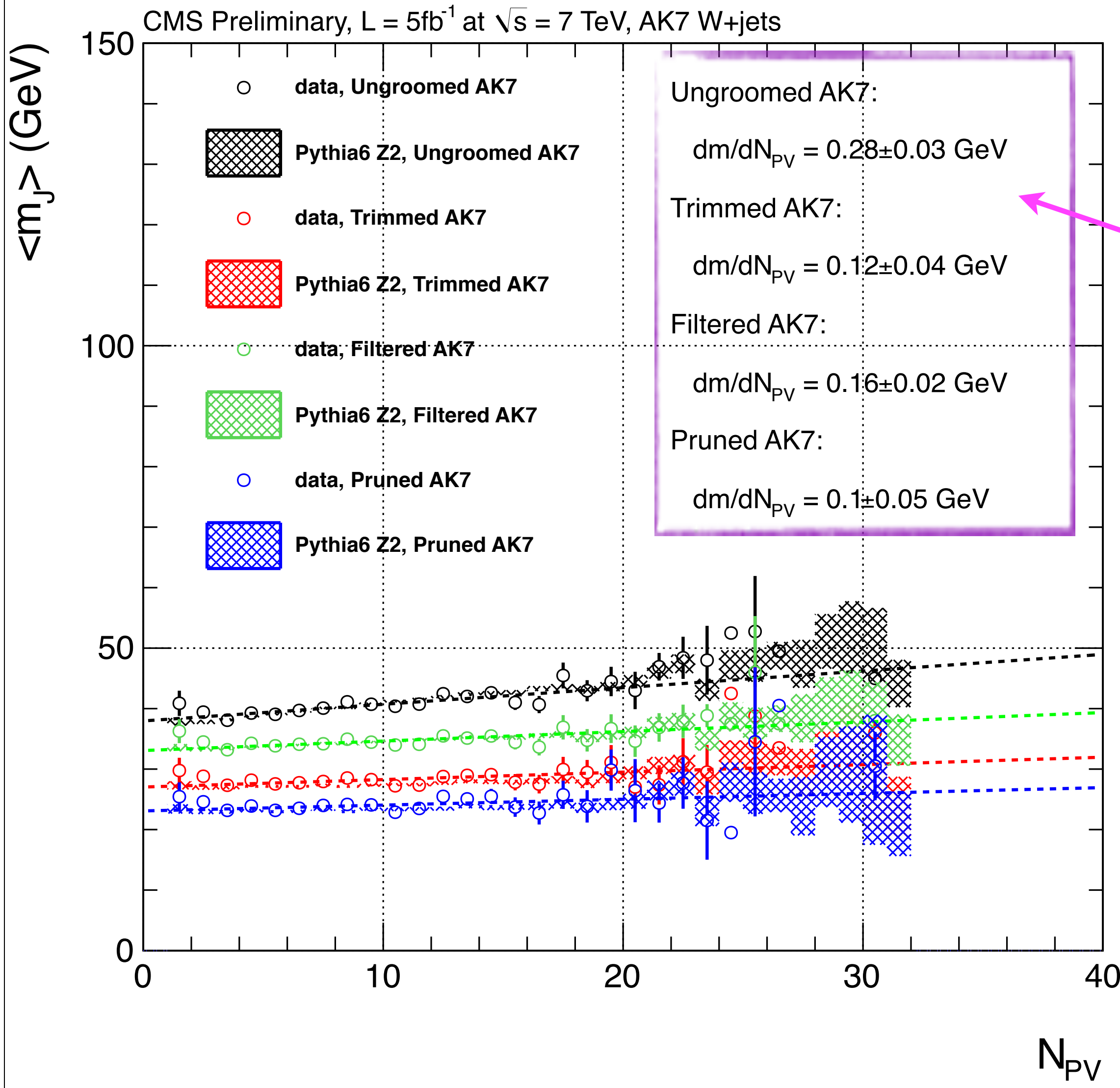
✦Ungroomed jet mass is very sensitive to PU

- <mJ> 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

Performance versus pileup for groomed jets



✦Grooming techniques are less sensitive to PU

- <m$_J$> vs NPV slope becomes flatter

✦Observe the expected behavior that <m$_J$> typically scales as R^3

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

Physics deliverable: jet mass comparisons

- For the grooming algorithm comparisons, focus on AK7 jets to make qualitative statements
- For V+jet plots, also present comparison of jet clustering algos and jet sizes
- To be insensitive to theoretical jet pT spectrum, we present results as normalized double differential distribution:

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

- For di-jet, use the average p_T and mass for the two leading jets in the event
- For V+jet, use p_T and mass of the leading jet

p_T 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
🟦 Di-jet only
🟧 Di-jet and V+jet

Unfolded distributions of jet mass

Unfold the m_J spectra to get detector-corrected distributions

$$\int f(\alpha)\hat{A}(\alpha,\beta)d\alpha = g(\beta)$$

Consider true distribution, f(α), convoluted with detector resolution A(α,β) to give you measured distribution, g(β)

For binned distributions, can be formulated as

$$Ax = b$$

where b is measured distribution, x is true distribution, A is a matrix containing detector effects

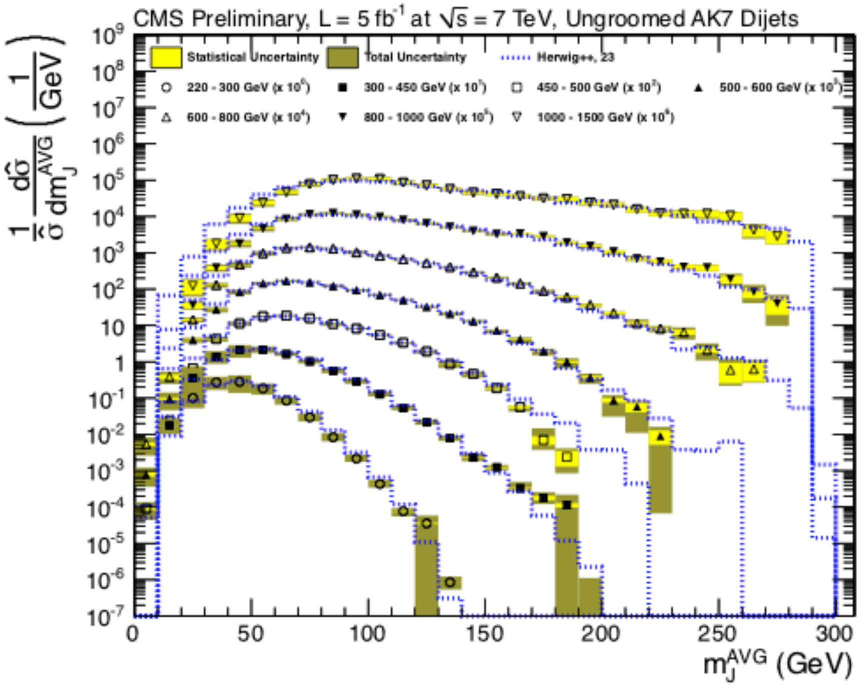
To get the measured distribution, invert matrix to get

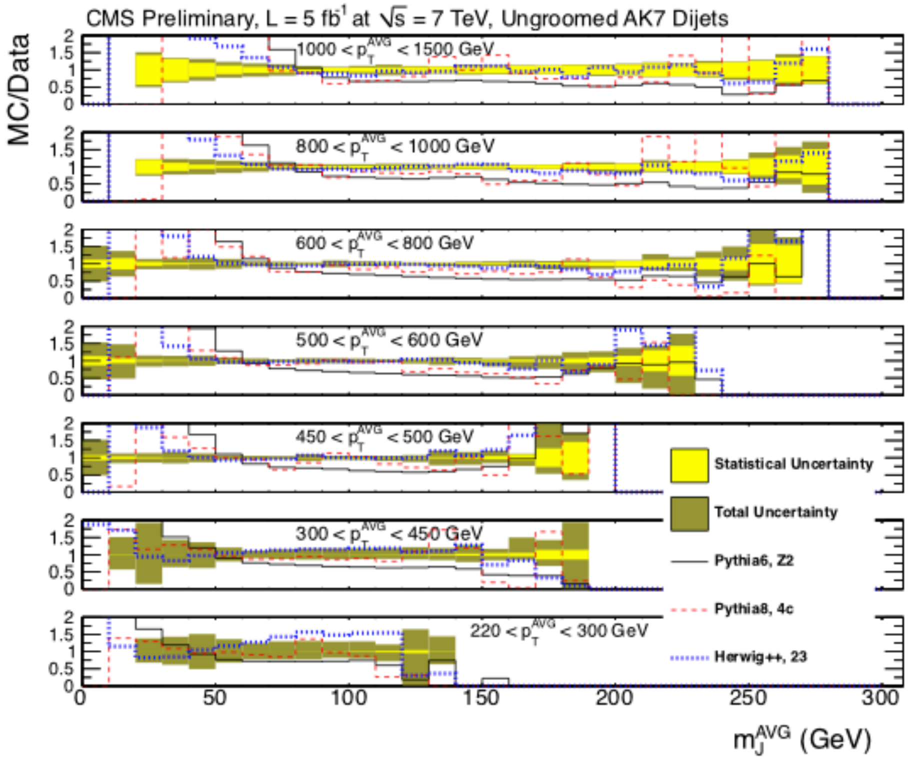
$$x = A^{-1}b$$

- For inversion of the matrix, use iterative Bayesian method to unfold to the true distributions.
- Combined electrons and muons in the case of V+jets using inclusive acceptance

Unfolded distributions, dijet ungroomed (AK7)

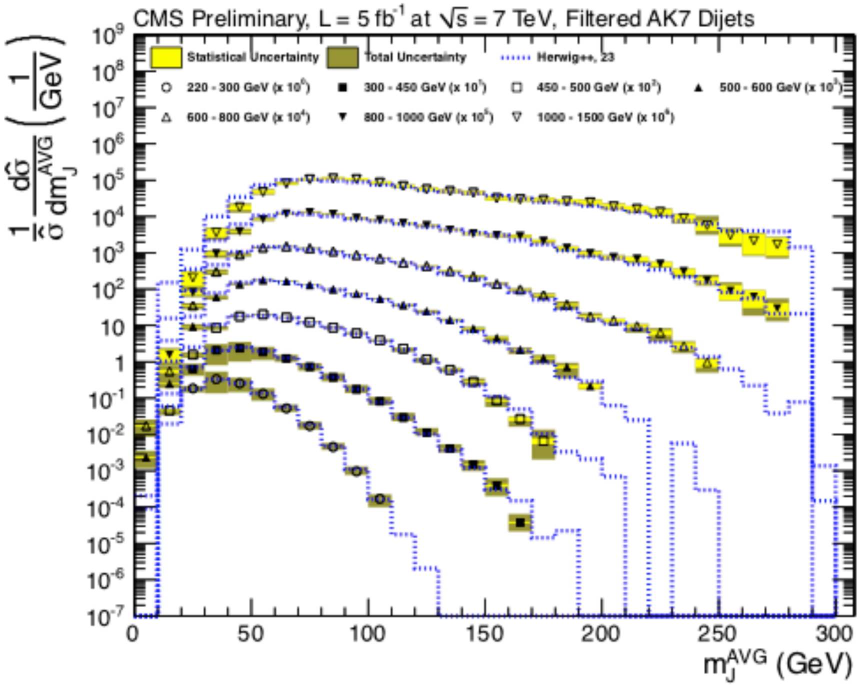
Plots are presented in increasing aggressiveness of grooming [Ungroomed, Filtered, Trimmed, Pruned] and best digested by flipping through.

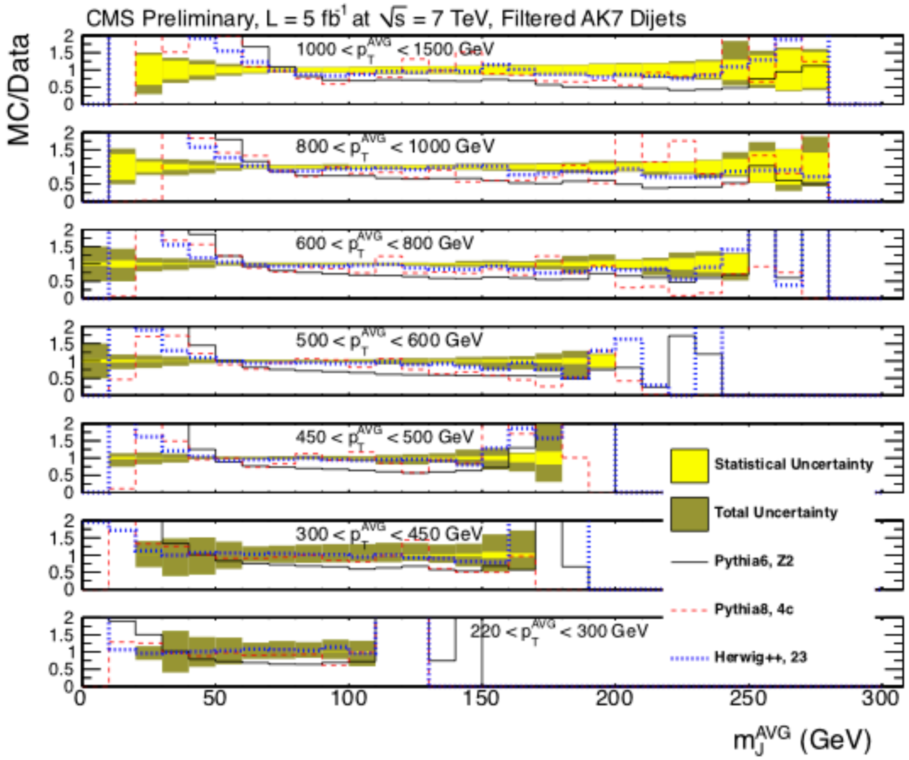




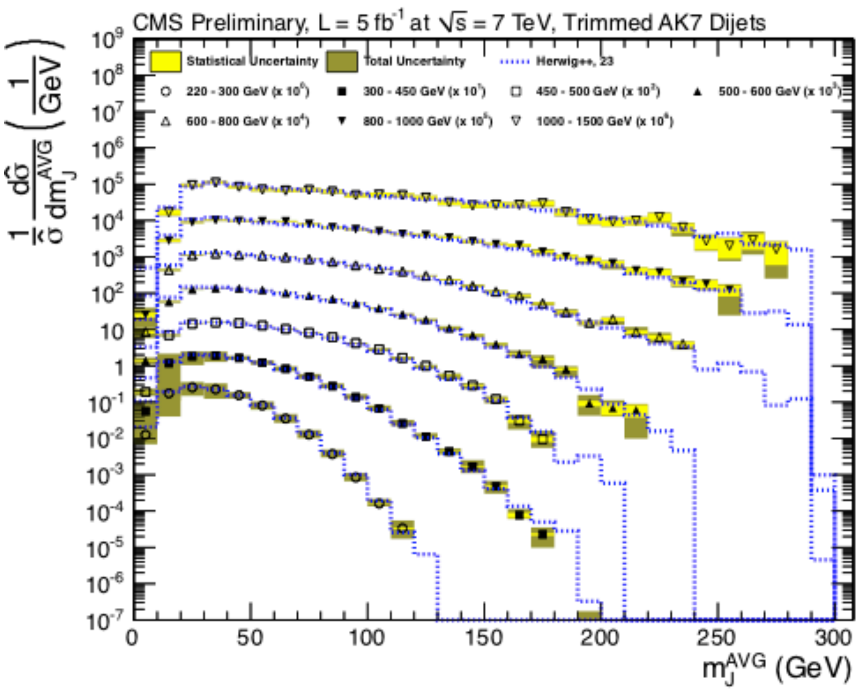
MC to data agreement is reasonable using both Pythia Z2 and Herwig++, though Herwig++ seems to have better agreement.

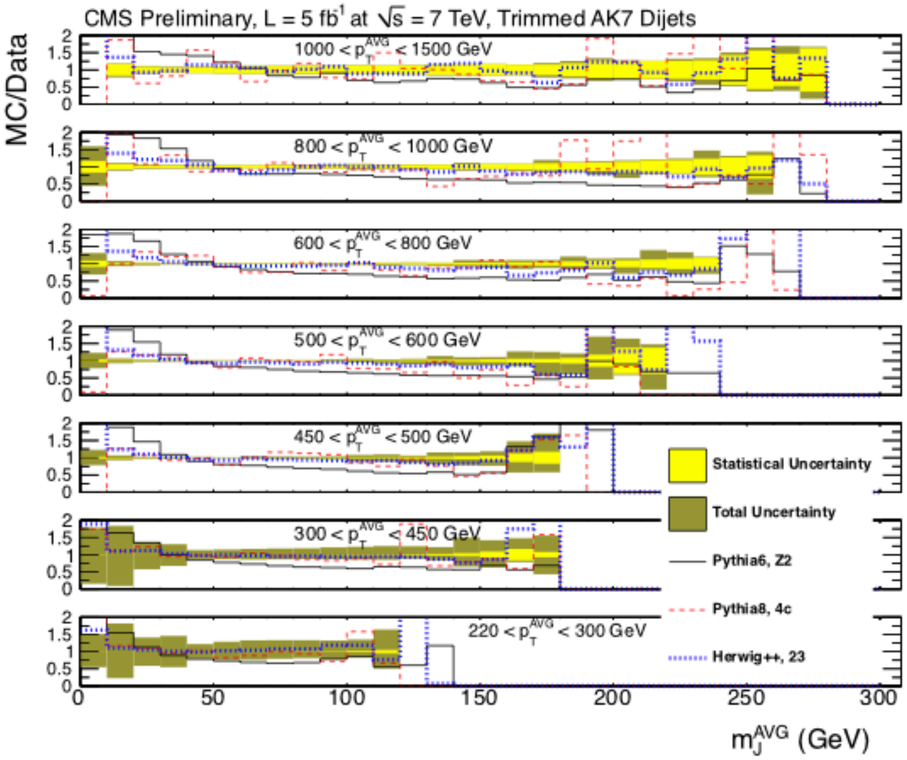
Unfolded distributions, dijet filtered (AK7)





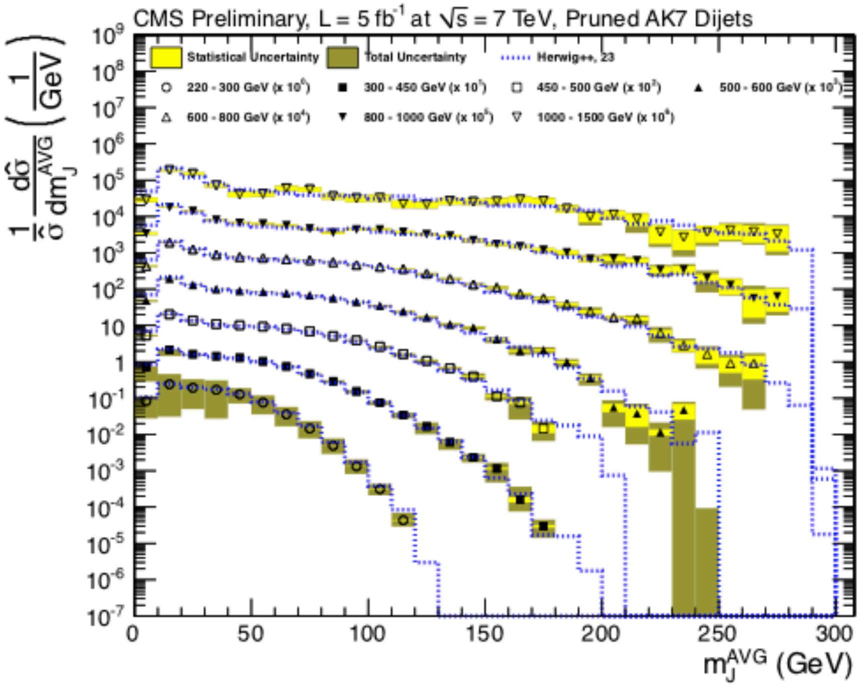
Unfolded distributions, dijet trimmed (AK7)

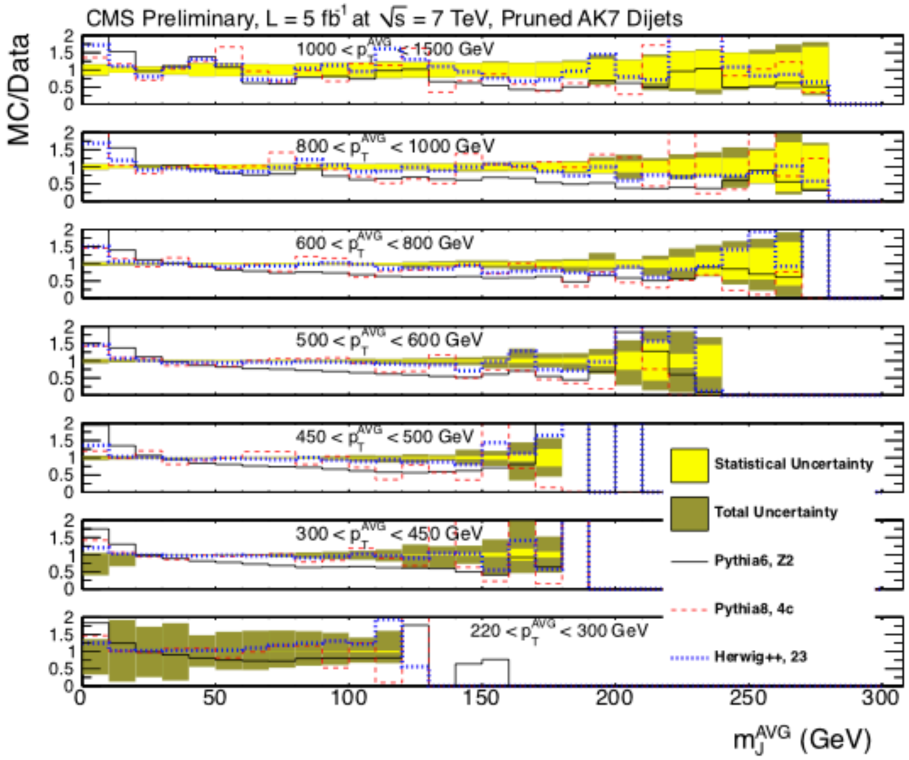




Unfolded distributions, dijet pruned (AK7)

As the grooming algorithm becomes more aggressive, Herwig++ agreement is better with data.





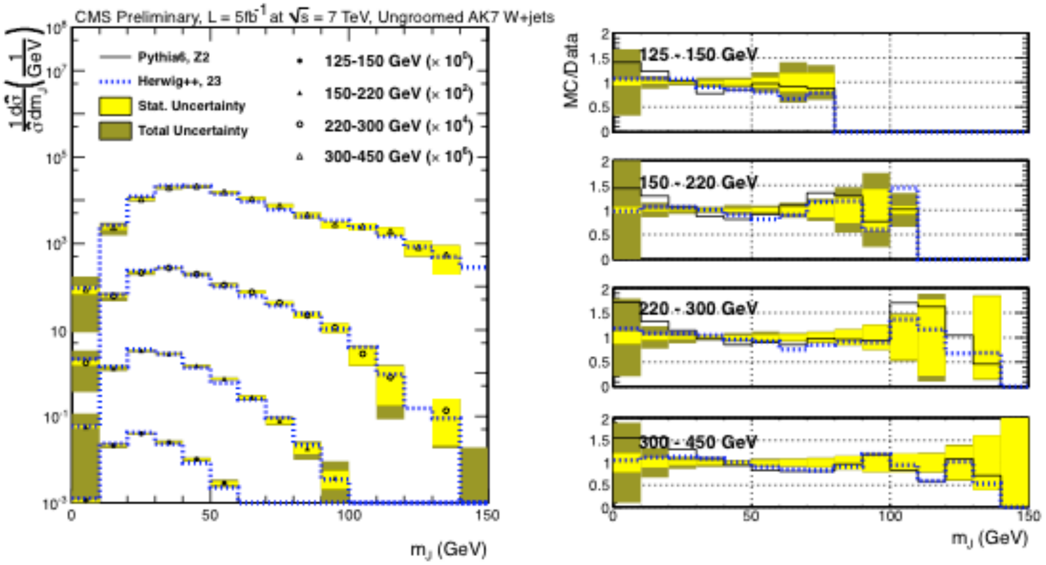
Turnover region moves to the left with harder grooming while tails remain similar in all cases − important feature in new physics searches.

Unfolded distributions, W+jet (AK7)

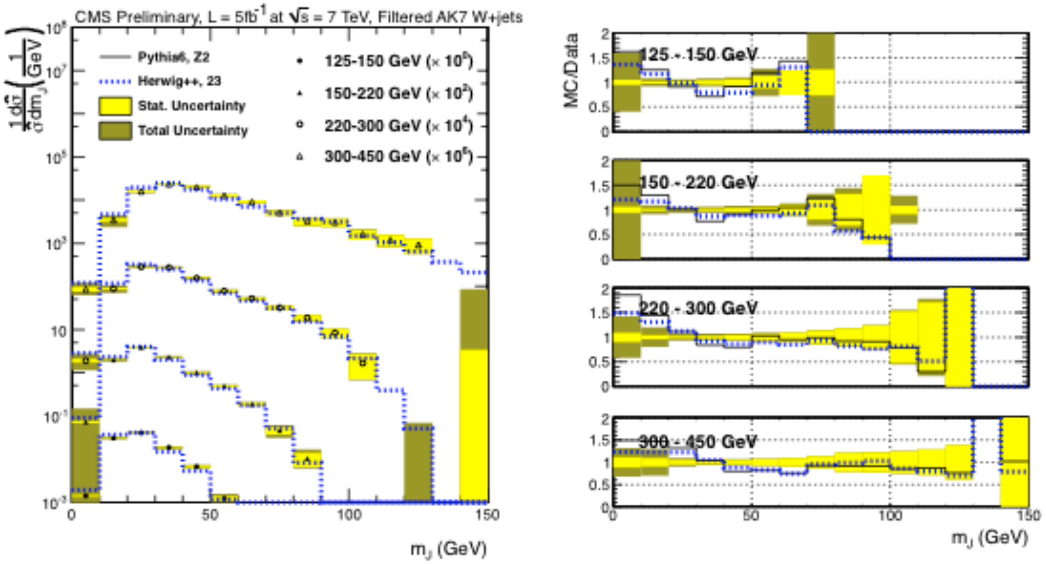
Qualitatively similar behavior to the dijet case.

N.B. Z+jet plots are in backup where qualitatively the same behavior is observed

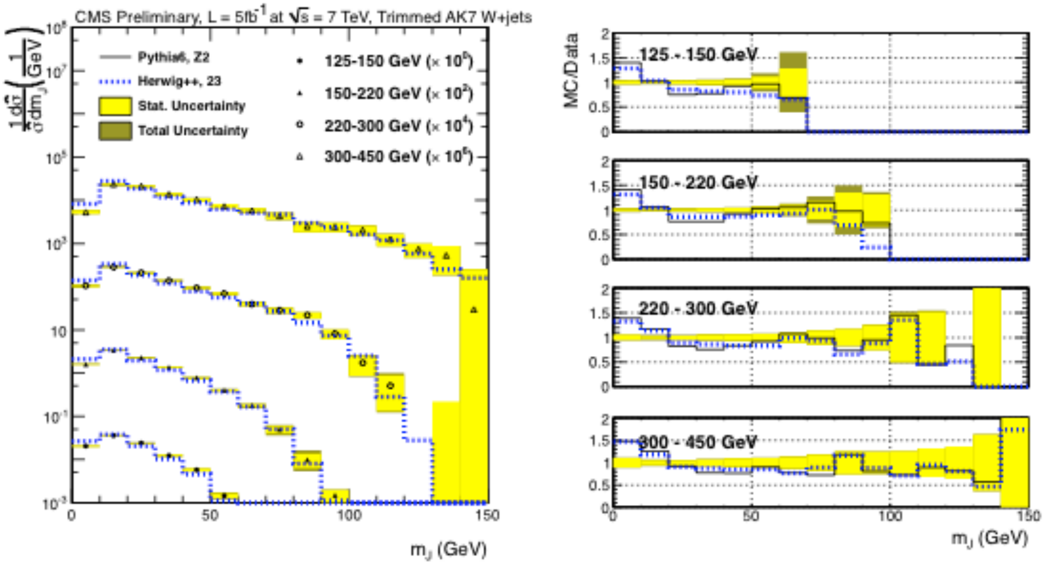
AK7, UNGR



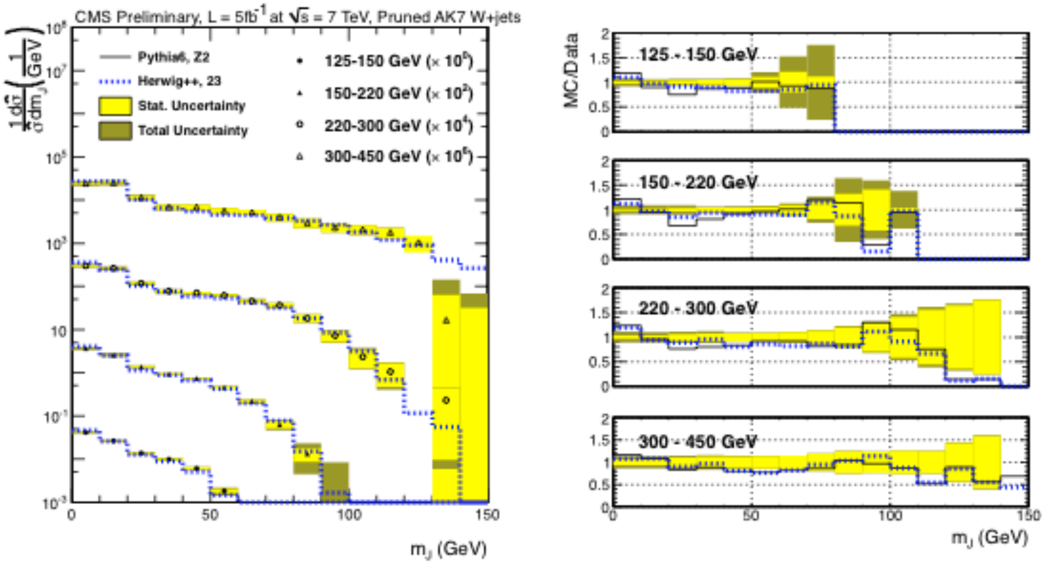
AK7, Filtered



Unfolded distributions, W+jet (AK7)

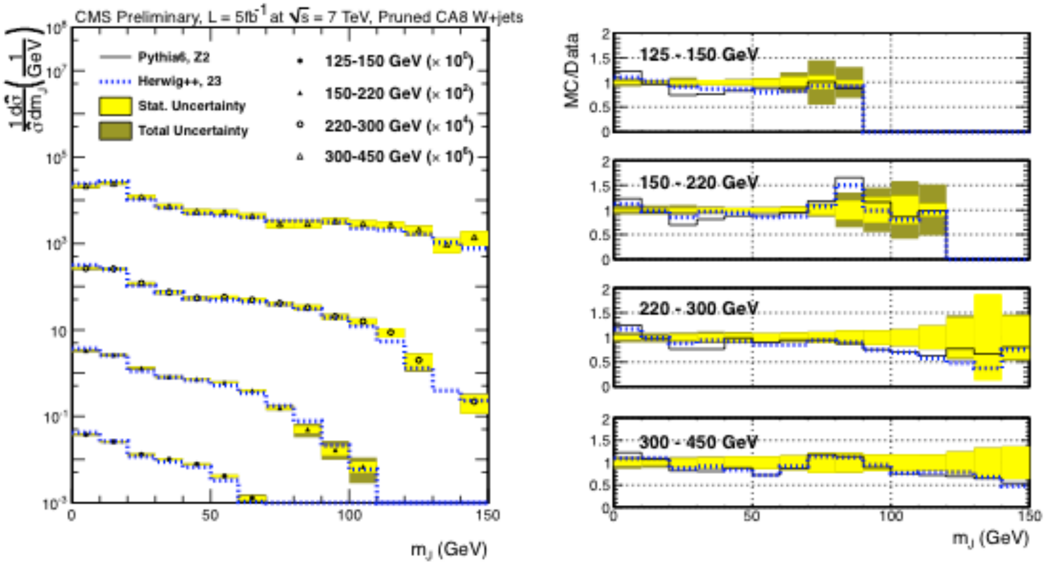


AK7, Trimmed

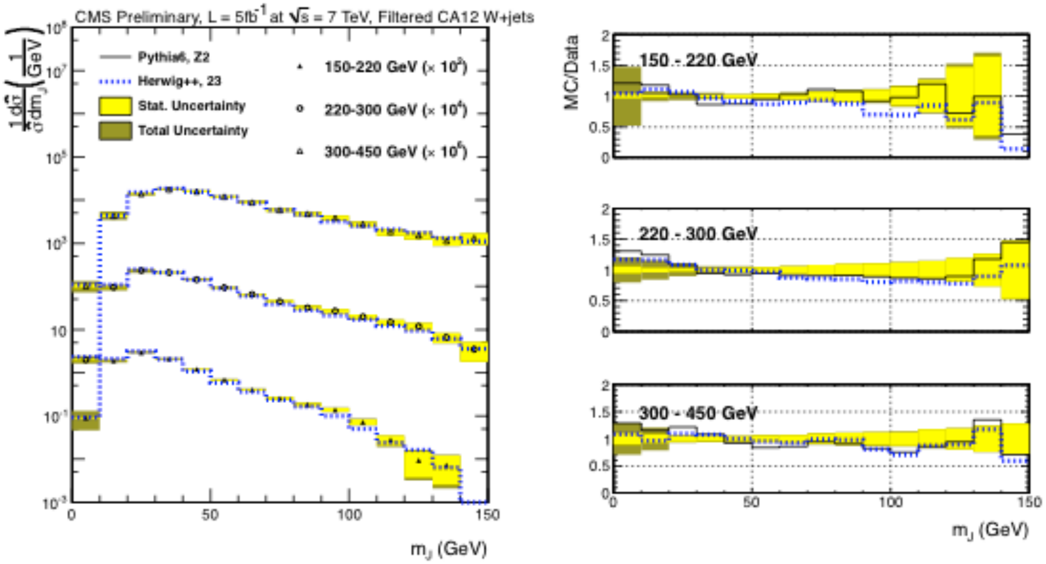


AK7, Pruned

Unfolded distributions, W+jet: comparison

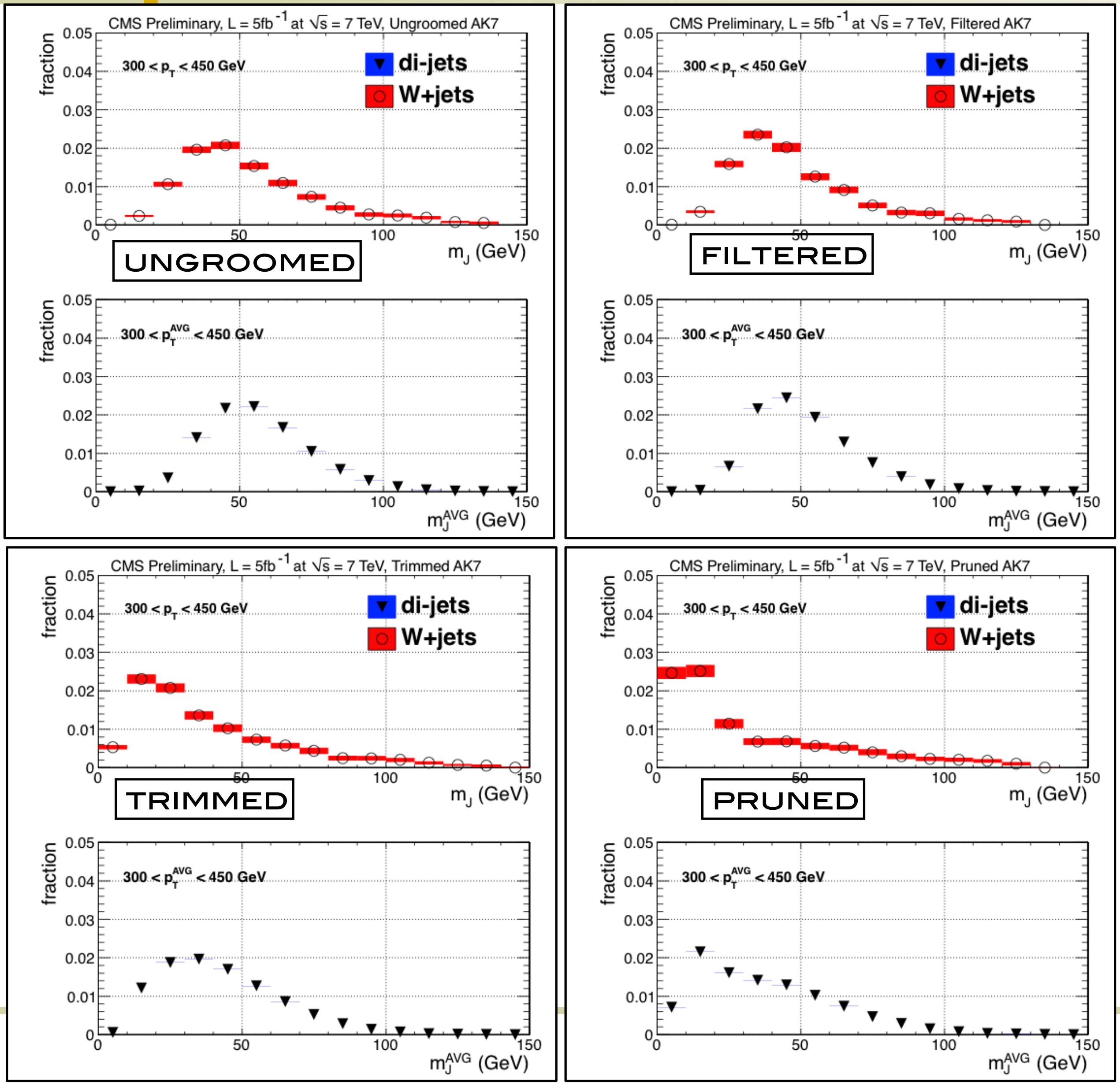


CA8, Pruned



CA12, Filtered

Comparing di-jet and V+jet



For visualization purposes, present di-jet and V+jet distributions with one above the other.

p_T bin:

300–450 GeV

N.B. not the exact same observables

m_J vs $\langle m_J \rangle$ in bins of pT, $\langle pT \rangle$

Shows the difference in jet mass for di-jet and V+jet processes where the former (latter) radiates more (less) due to larger composition of gluon (quark)-initiated jets

N.B. stat. err. only

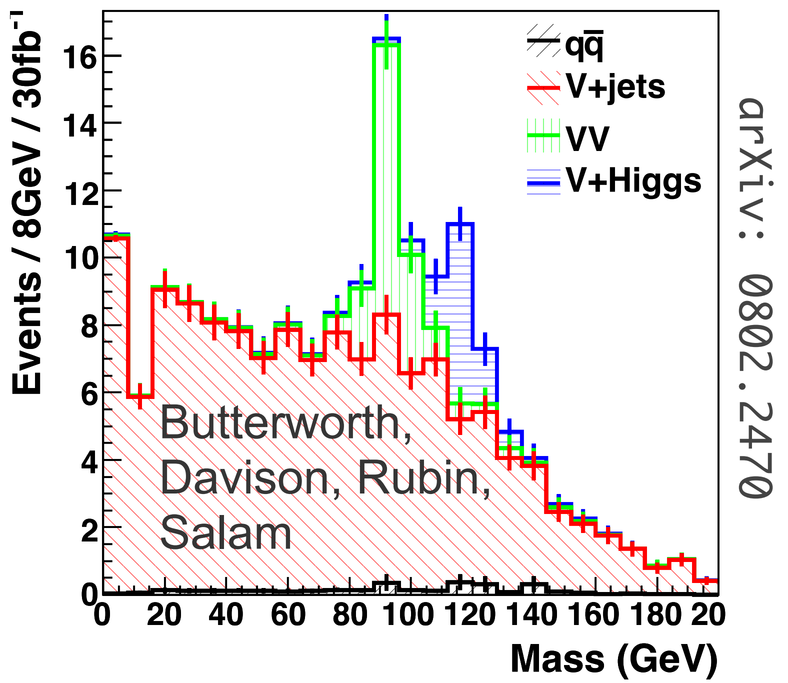
Systematic uncertainties

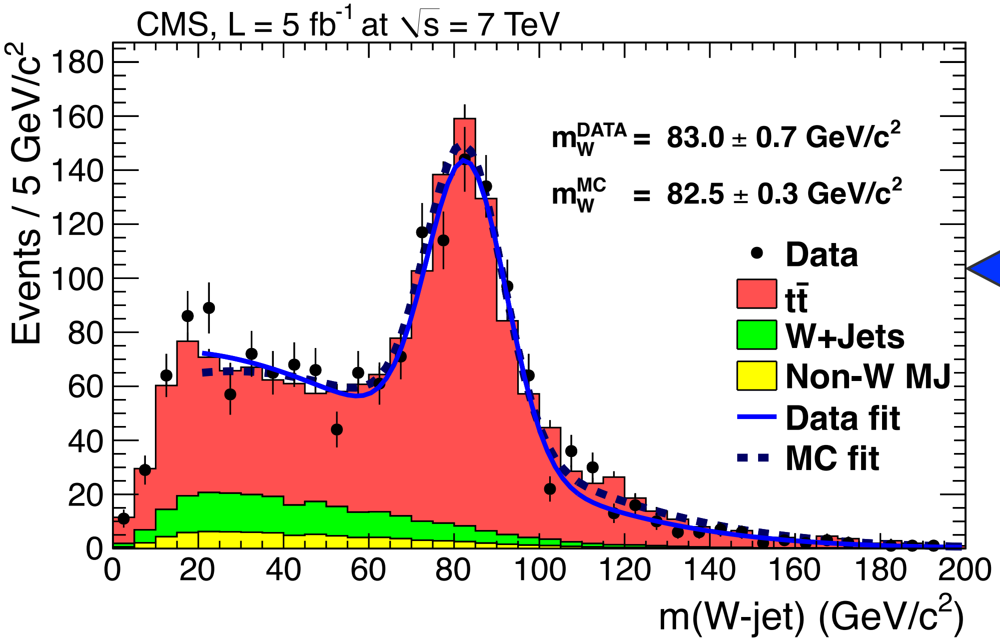
- Parton shower theoretical uncertainty
  - The difference between unfolding with two parton shower MC models: Pythia Z2 and Herwig++
- Jet Energy Scale (JES)
  - Up and down variations as a function of η and p_T
  - Additional 1% uncertainty from scale from pruned W jets (arXiv: 1204.2488, also see next slide)
- Jet Energy Resolution (JER)/Jet Angular Resolution(JAR)
  - Vary the jet energy, η, and pT resolution up and down by 10% − with nominal value of 10%
- Pile-up
  - Vary the Min Bias cross-section by 8%
- Bias corrections from unfolding procedure

A ground work for more interesting physics ....

Jet substructure can be used to improve sensitivity of hadronic decays of boosted heavy particles such as Higgs, W/Z, and top

This is what we aim to do





Started with hadronic W in boosted top events

http://cdsweb.cern.ch/record/1370237

See talk by Nhan Tran this afternoon for use of jet substructure in W/top/Higgs tagging and exotic searches.

Summary

☑ Jet calibration well understood for jets in tracker coverage
- at the level of a few %; groomed jets have sim. response
- effect of pileup well simulated, area subtraction working

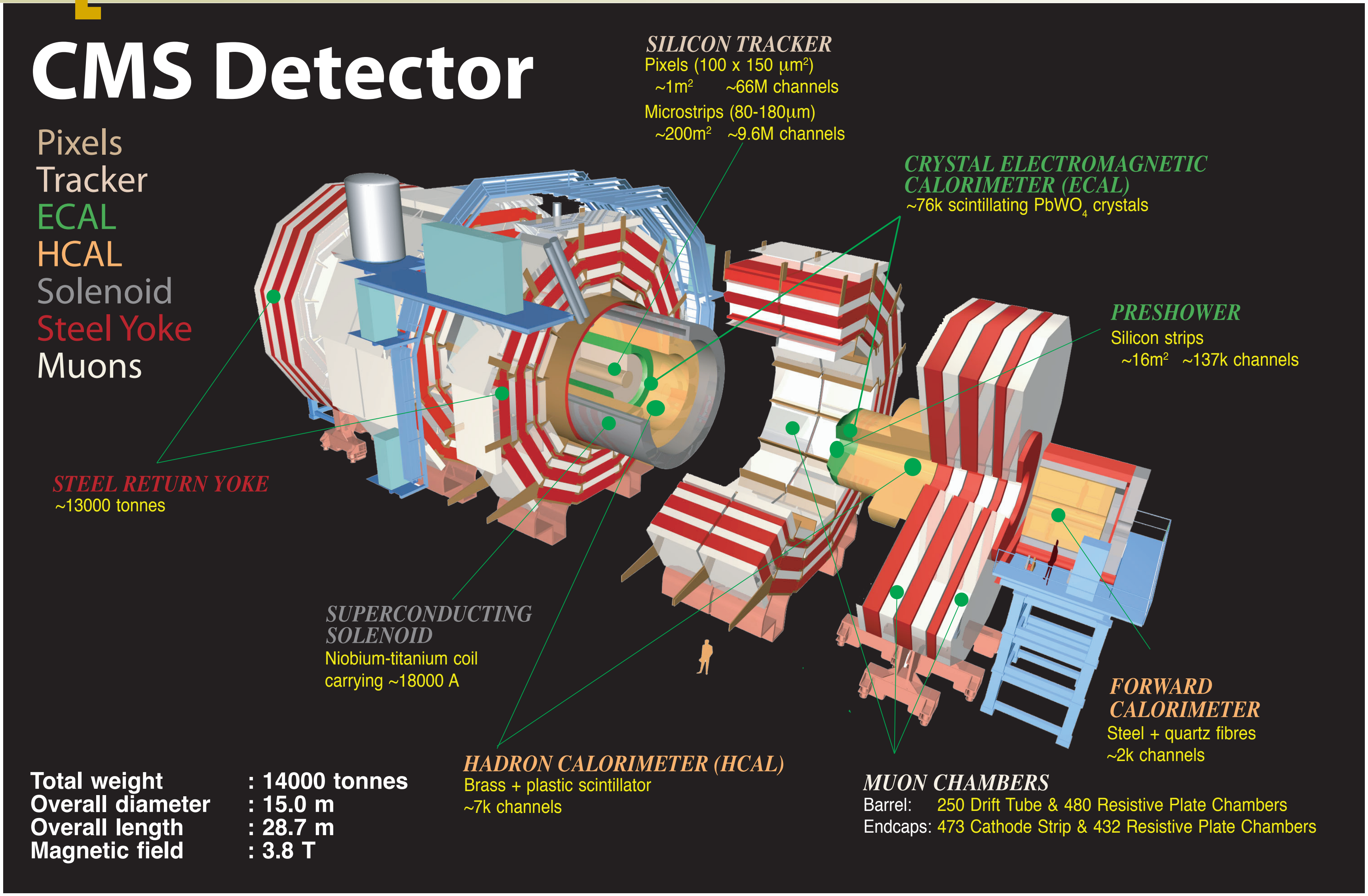
☑ First systematic study of jet grooming techniques on jet mass performed at CMS
- an important benchmark in understanding various algorithms for searches for new physics

☑ Comparison with Pythia Z2 and Herwig++ simulation gives reasonable agreement, Herwig++ appears to agree better
- performance of jet p_T and mass versus pileup shows that the grooming techniques lessen sensitivity to pileup

☑ Detector-unfolded distributions obtained for MC comparison

BACKUP SLIDES

Understanding CMS detector



CMS Detector

Pixels
Tracker
ECAL
HCAL
Solenoid
Steel Yoke
Muons

SILICON TRACKER
Pixels (100 x 150 μm^2)
~1m^2 ~66M channels
Microstrips (80-180μm)
~200m^2 ~9.6M channels

CRYSTAL ELECTROMAGNETIC CALORIMETER (ECAL)
~76k scintillating $PbWO_4$ crystals

PRESHOWER
Silicon strips
~16m^2 ~137k channels

STEEL RETURN YOKE
~13000 tonnes

SUPERCONDUCTING SOLENOID
Niobium-titanium coil carrying ~18000 A

FORWARD CALORIMETER
Steel + quartz fibres
~2k channels

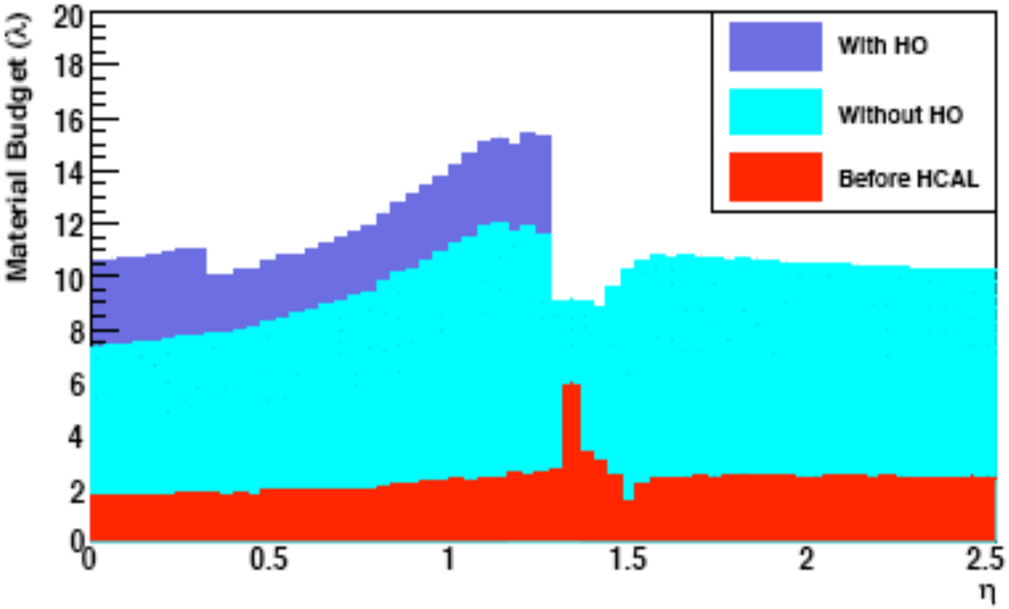
HADRON CALORIMETER (HCAL)
Brass + plastic scintillator
~7k channels

MUON CHAMBERS
Barrel: 250 Drift Tube & 480 Resistive Plate Chambers
Endcaps: 473 Cathode Strip & 432 Resistive Plate Chambers

Total weight	**: 14000 tonnes**
Overall diameter	**: 15.0 m**
Overall length	**: 28.7 m**
Magnetic field	**: 3.8 T**

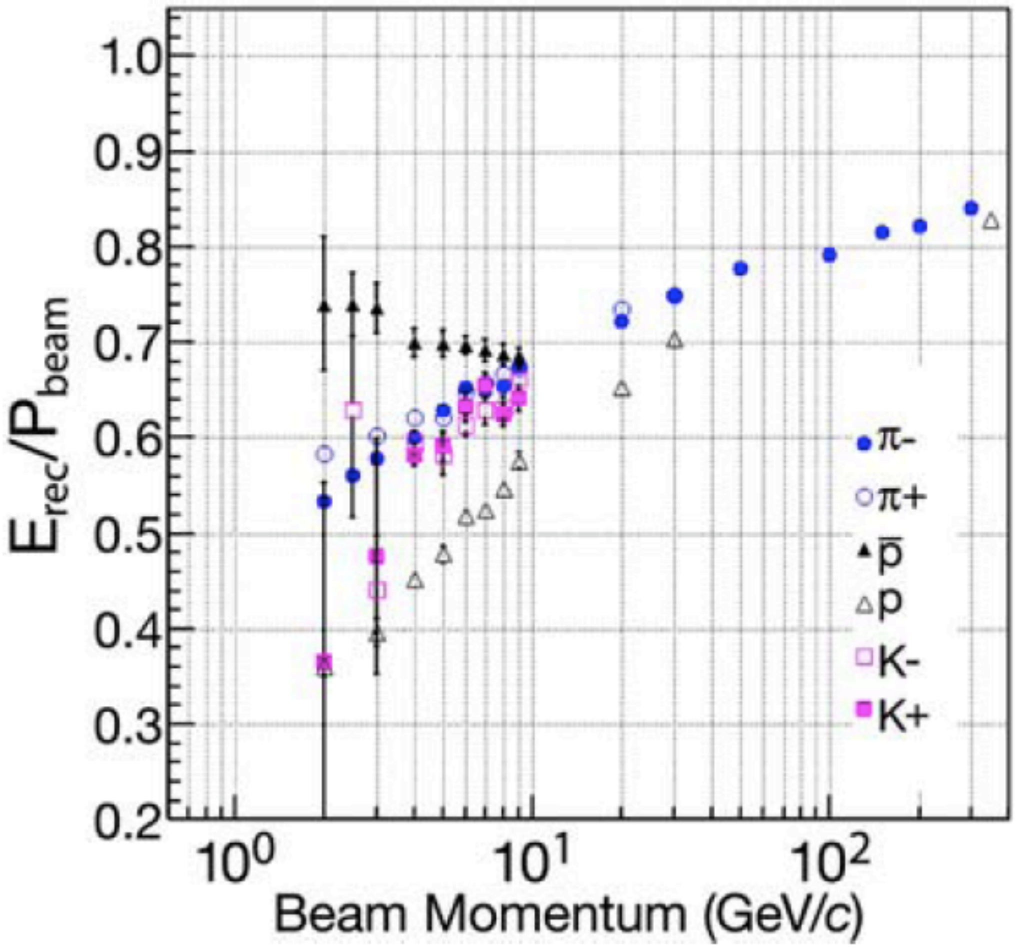
Material budget of the calorimeter

Thickness of HCAL in terms of interaction lengths



7-8 Interaction Lengths at η=0 with HCAL alone and is insufficient to fully contain the shower generated by pions above 100 GeV

Calorimeter response in test beam data

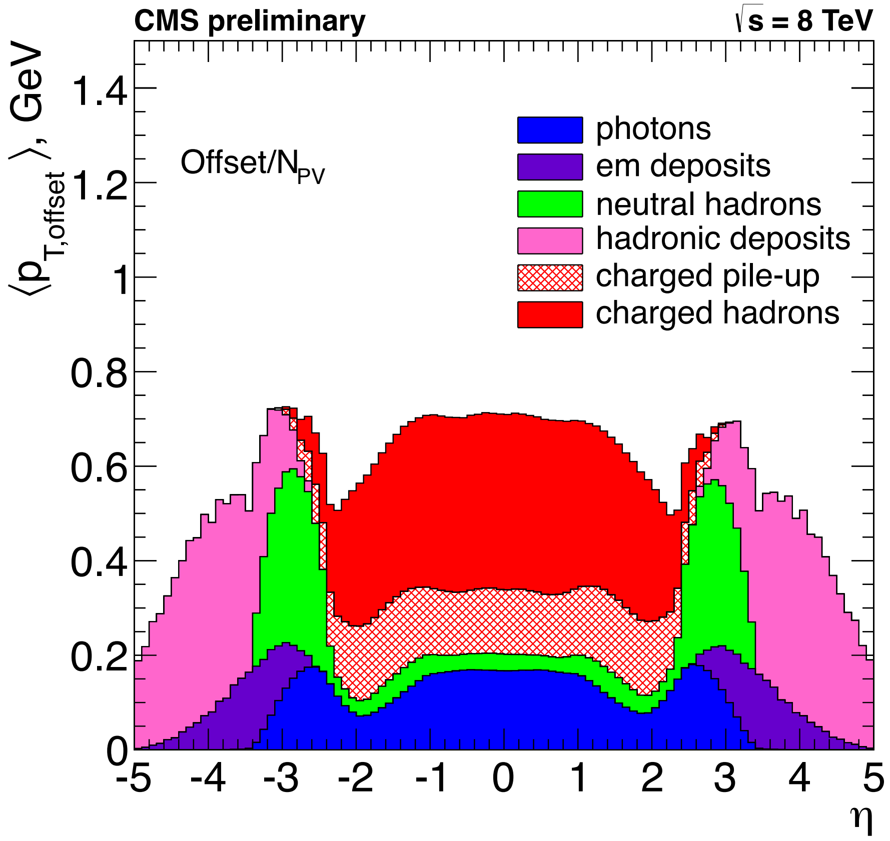


- The figure shows the combined response of EB+HB to different particles as a function of beam momentum.
  - -The response is normalized to 1 for electron.
  - -At 100 GeV, the pion response is 80% of electron.
  - -The proton response is always lower than pion.

- In collision data the response is lower than in test beam because of additional material in front of the calorimeter.
- The calorimeter response is clearly non-linear.

Pileup contribution to jet energy

- ✦Pileup (PU) measured with Zero Bias data and MC
  - Random cone allows to separate contribution per detector
  - Most charged hadrons can be associated to pileup vertices and removed



- Part that can be removed is labeled “charged hadrons”

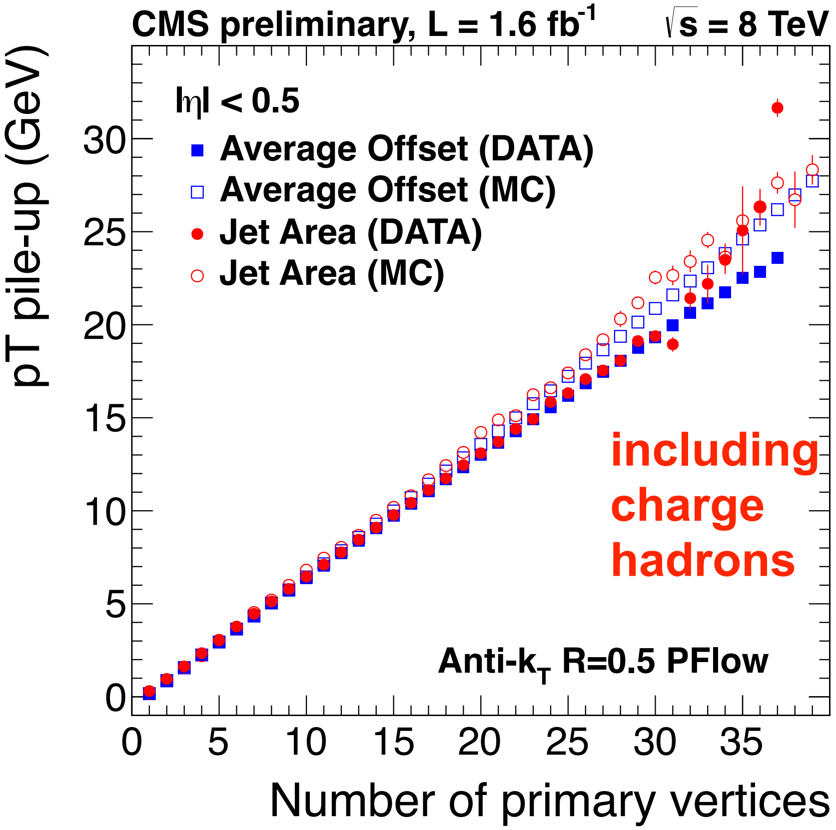
- Part that remains as PU after this needs to be subtracted

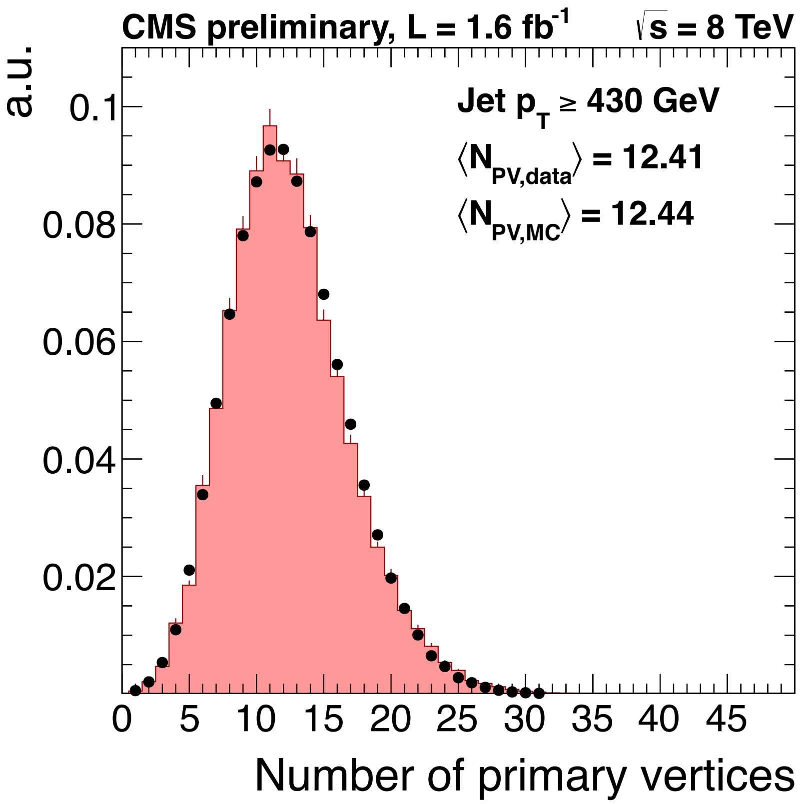
  PU density x Effective area
  (FastJet-ρ)

- PU density depends on the # of primary vertex in the event

Pileup correction: using NPV and [$\rho_{FastJet}$ x Area]

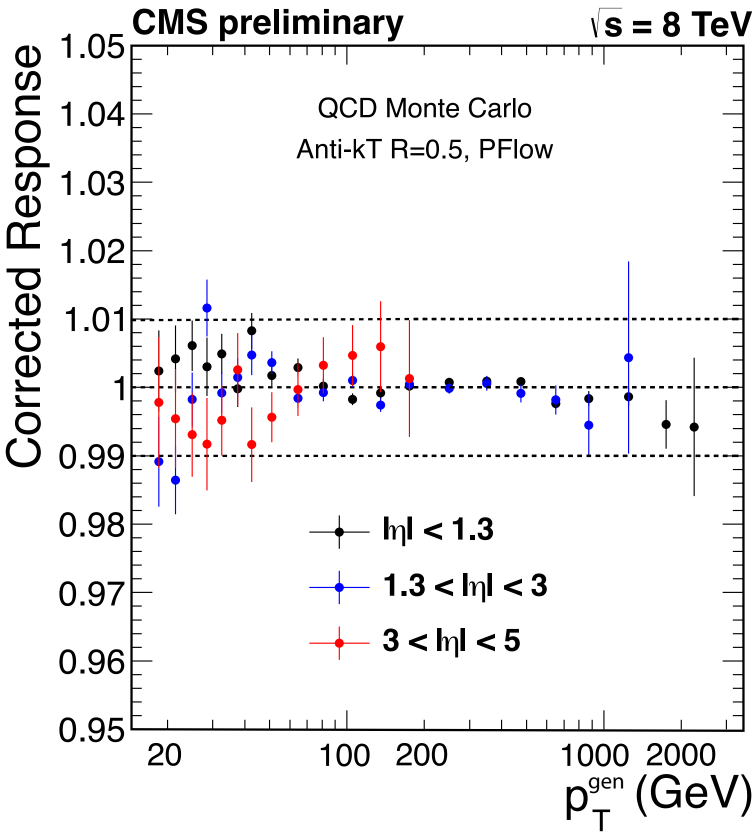
- Both NPV-based and FastJet-ρ-based corrections are in agreement
- Remaining Data/MC difference accounted for with separate PU corrections
  - Reweight pileup Poisson mean in MC to data. Poisson mean determined from measured luminosity and Minimum bias cross section.

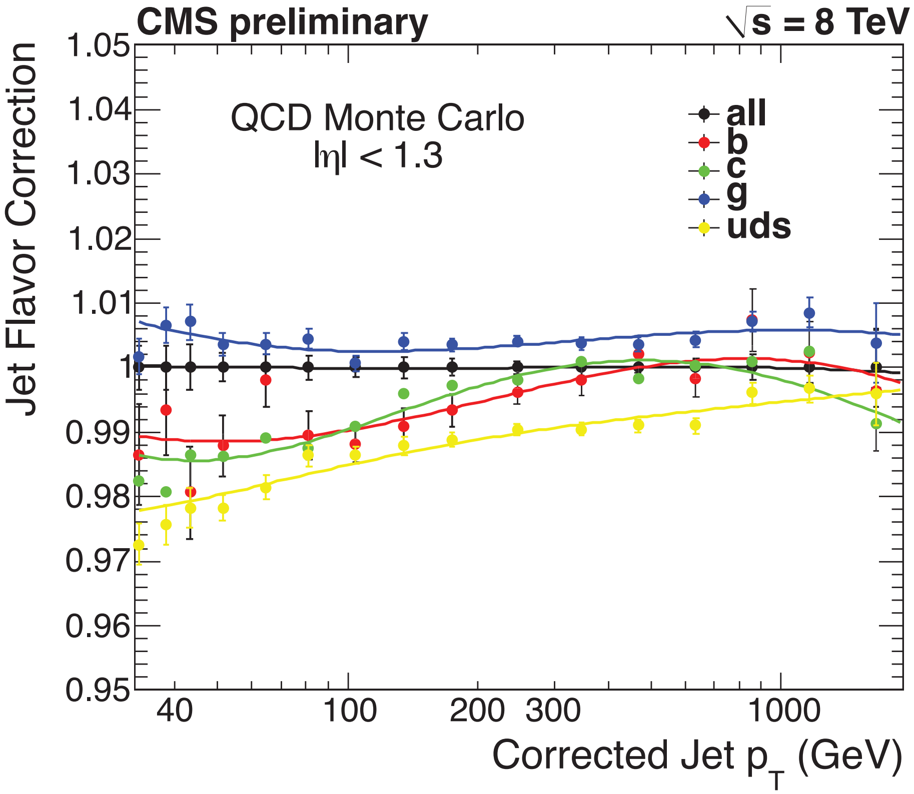




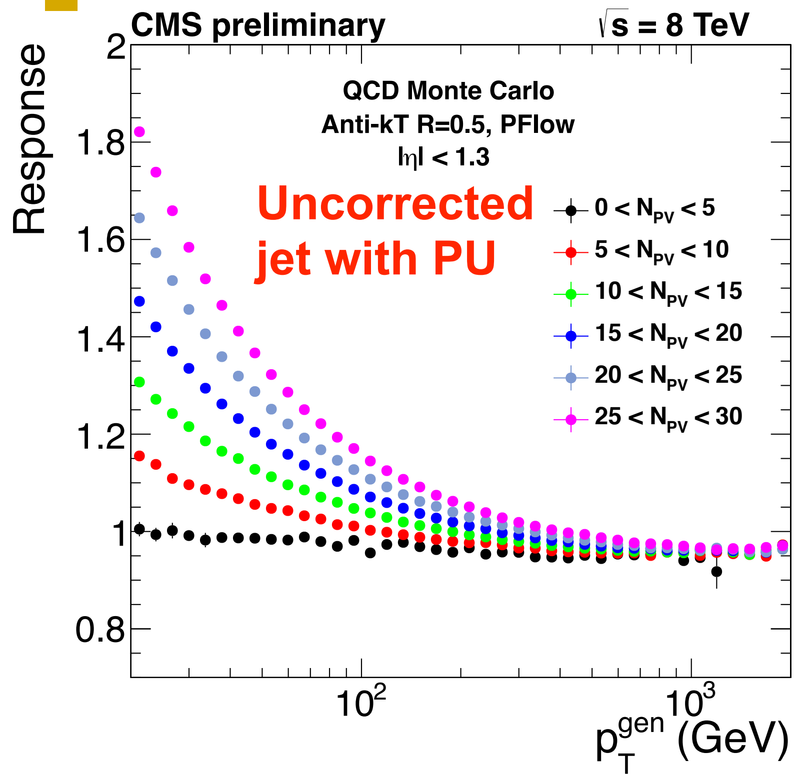
After PU offset: start with MC-based correction

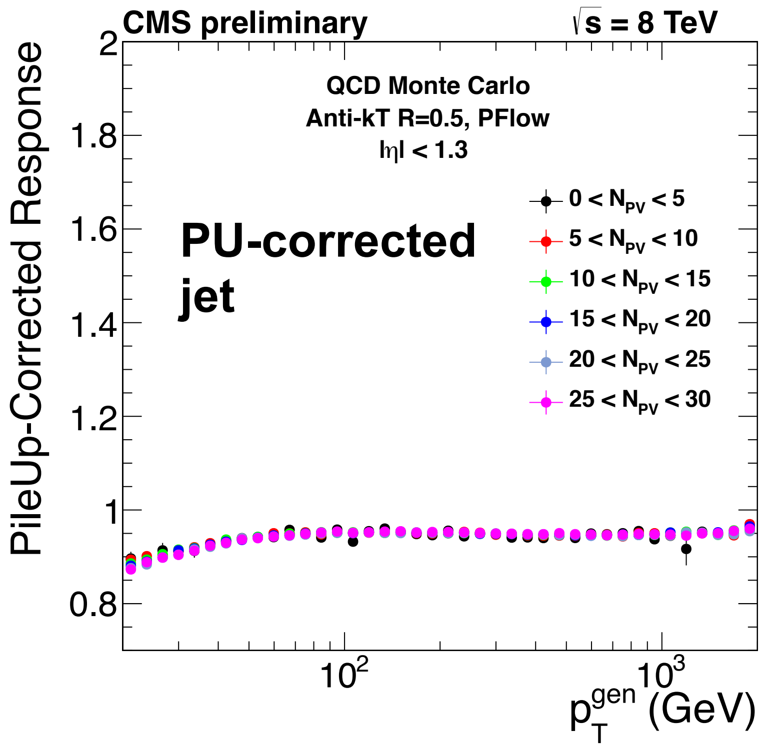
- ✦Eta and p_T corrections derived from QCD MC sample
  - Corrected response closes well in MC.
- ✦Particle flow minimizes flavor response differences
  - Maximum flavor difference within 3% in barrel for p_T > 30 GeV.

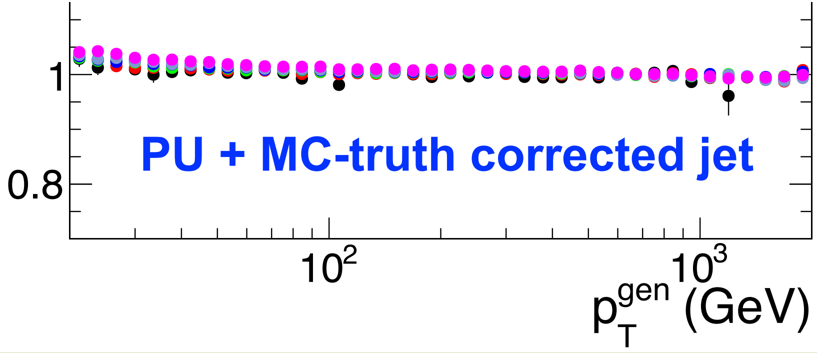




Combined pileup and MC truth effects





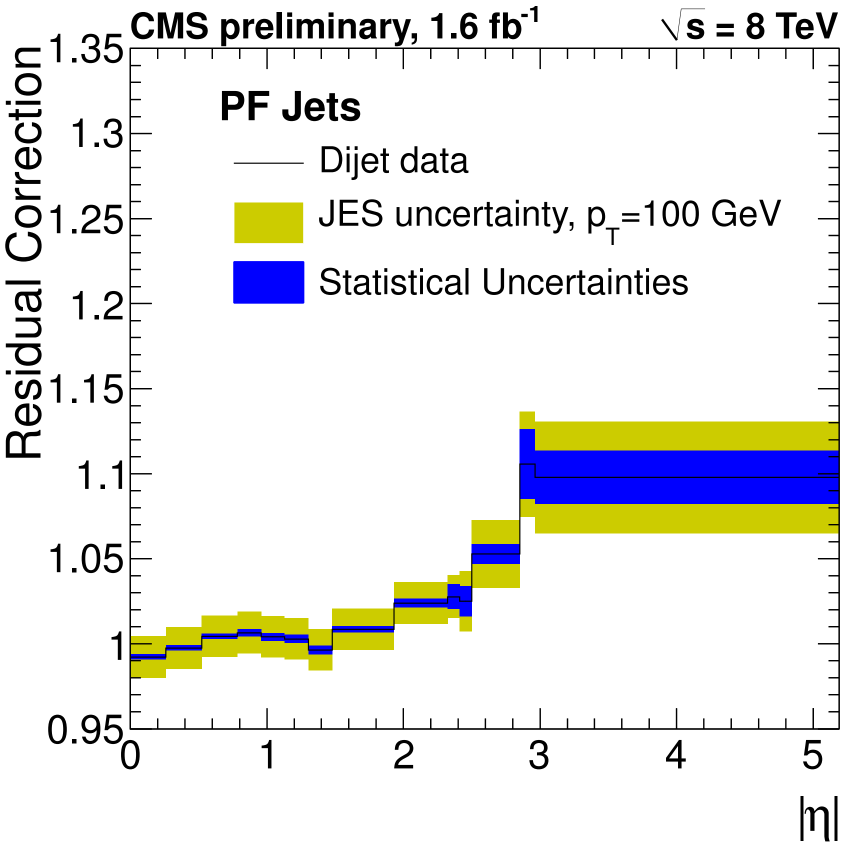


- ✦PU subtraction removes response dependance vs N_{PV}
- ✦MC truth correction brings the response back to one

Residual correction in data: η & p_T dependence

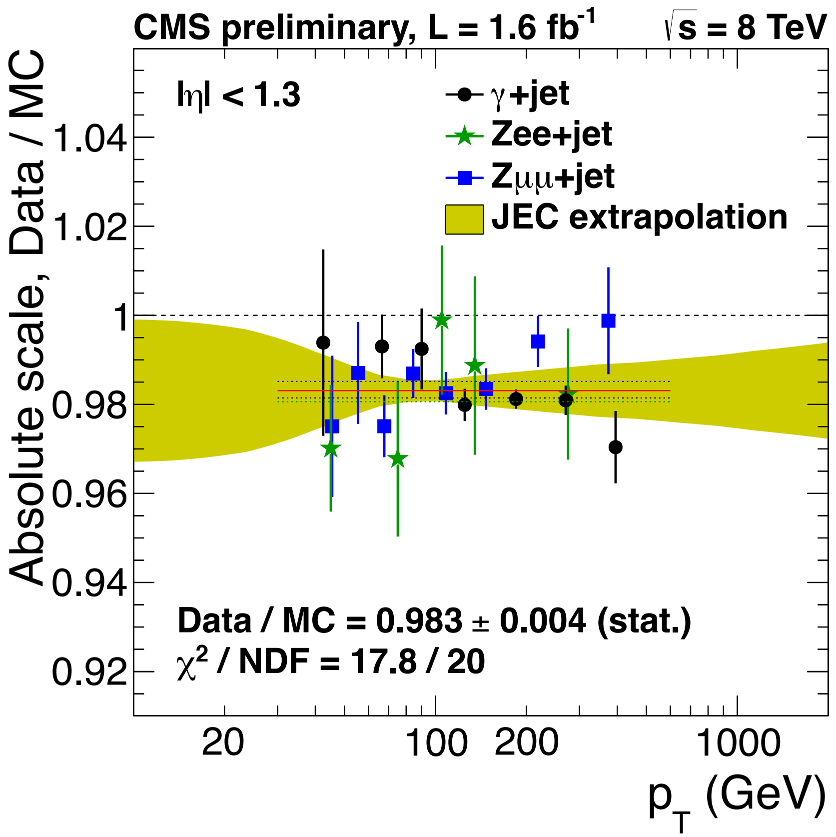
η dependence

- Using dijet events
- < 2.5% for jets in |η|<2.4
- HF modeling requires 5−10% correction

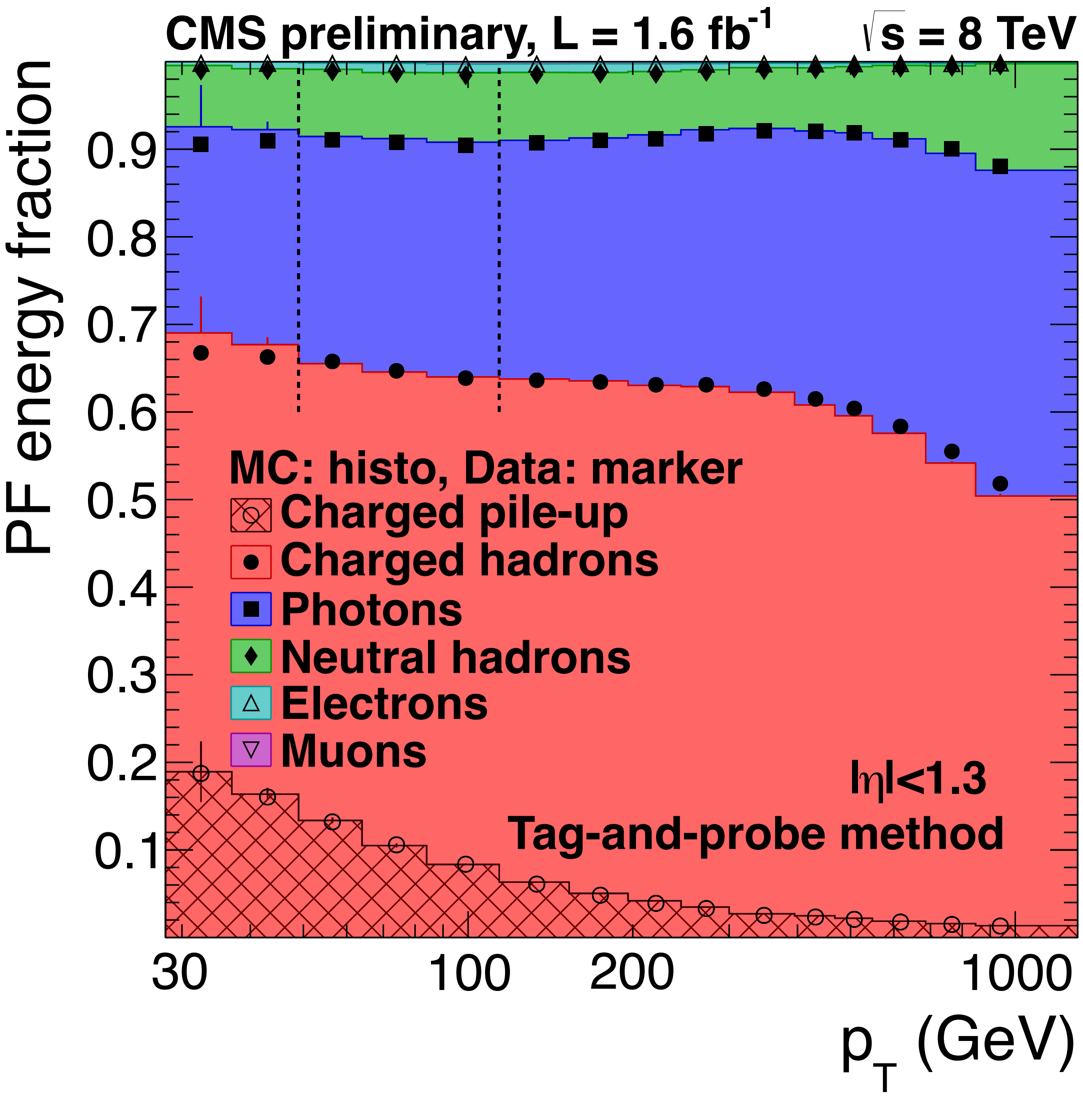


p_T dependence

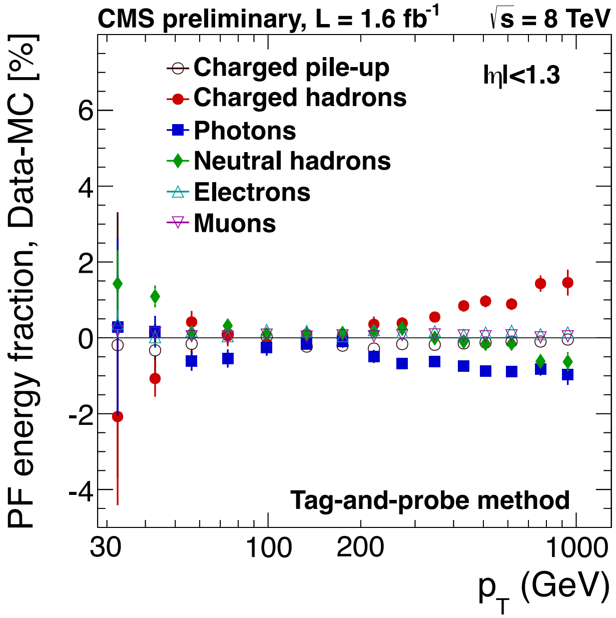
- Using Zμμ+jet, Zee+jet, γ+jet events
- No significant p_T dependance observed, so a single flat scale factor is used.



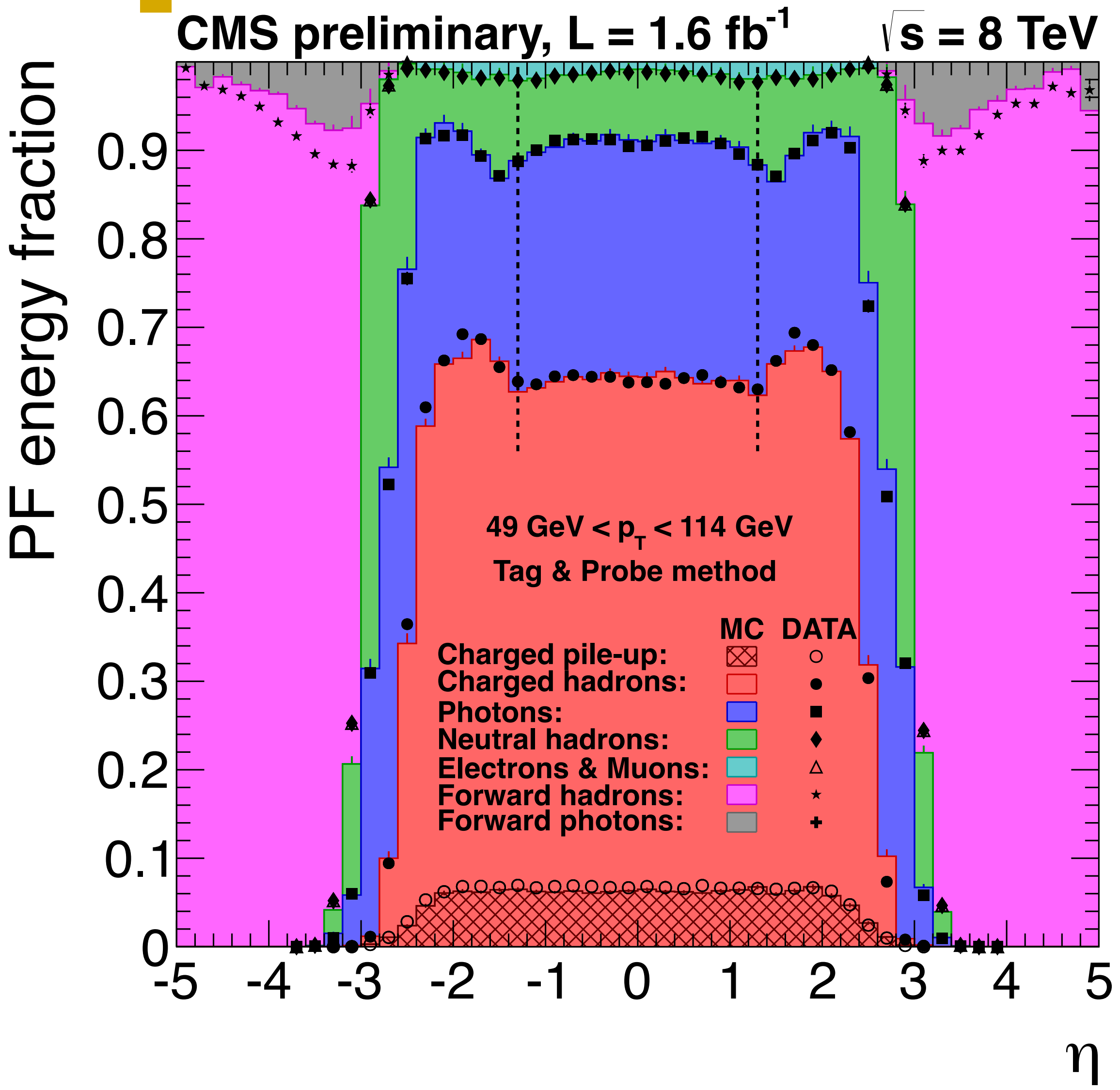
Jet composition vs p_T in barrel



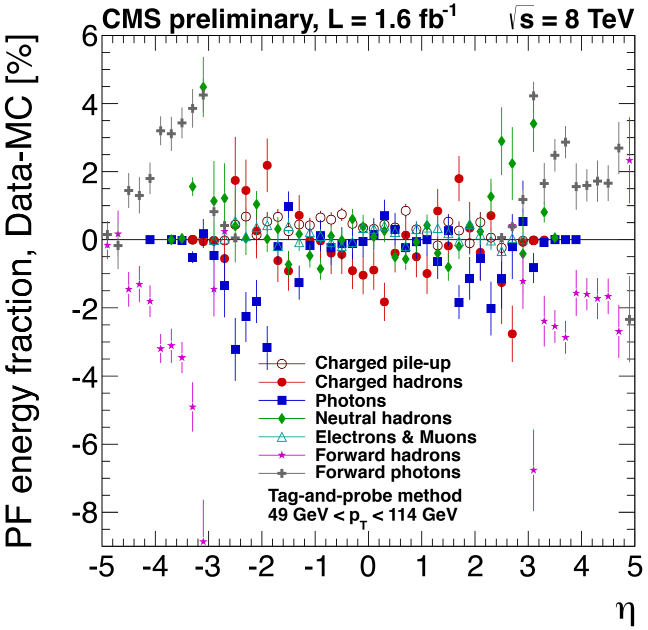
- ✦Jet composition agrees well between Data and MC
  - •consistent with small residual JEC at the 1-2% level.



Jet composition vs η



- ✦Jet composition shows increasing differences in the forward region
  - •consistent with JEC at 2-13% level.



Performance versus pileup

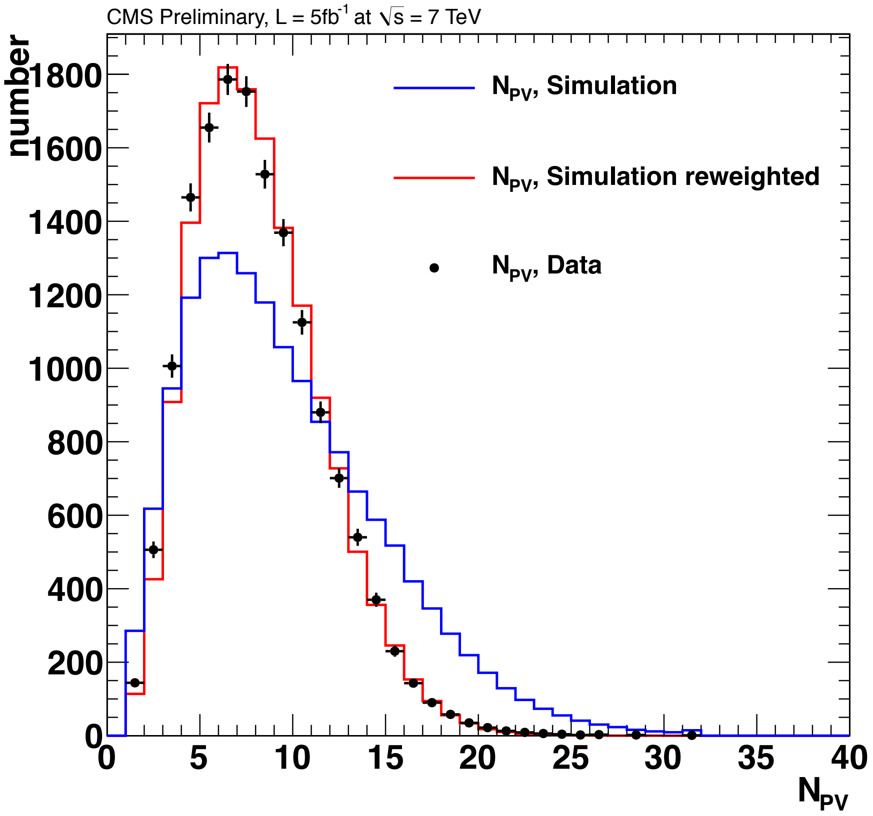
https://twiki.cern.ch/twiki/bin/view/CMSPublic/PhysicsResultsSMP12019

- 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

Pileup profile in 7 TeV data



Analysis benchmarks

- 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

Jets types under investigation

Clustering algorithms	Jet substructure algorithms
Anti-kT (R=0.5)	-
Anti-kT (R=0.7)	-/Trimmed/Filtered/Pruned
Anti-kT (R=0.8)	-
Cambridge-Aachen (R=0.8)	Pruned
Cambridge-Aachen (R=1.2)	Filtered

Study CA8 Pruned and CA12 Filtered due to their relevance in current CMS analyses

Simulation and samples

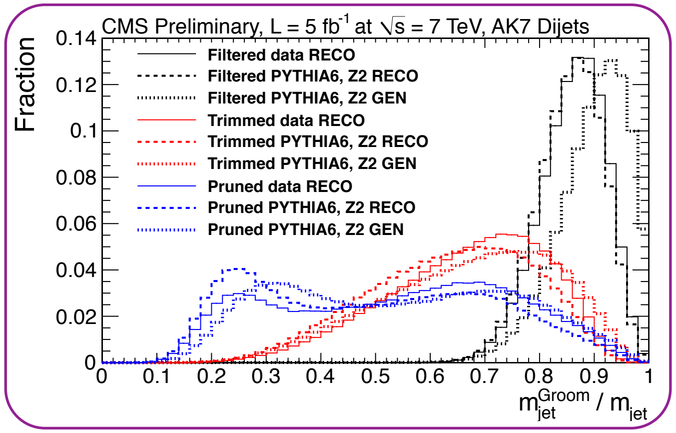
- Analysis with 5 fb^{-1} of 2011 data, $\sqrt{s} = 7$ TeV
- Simulated samples

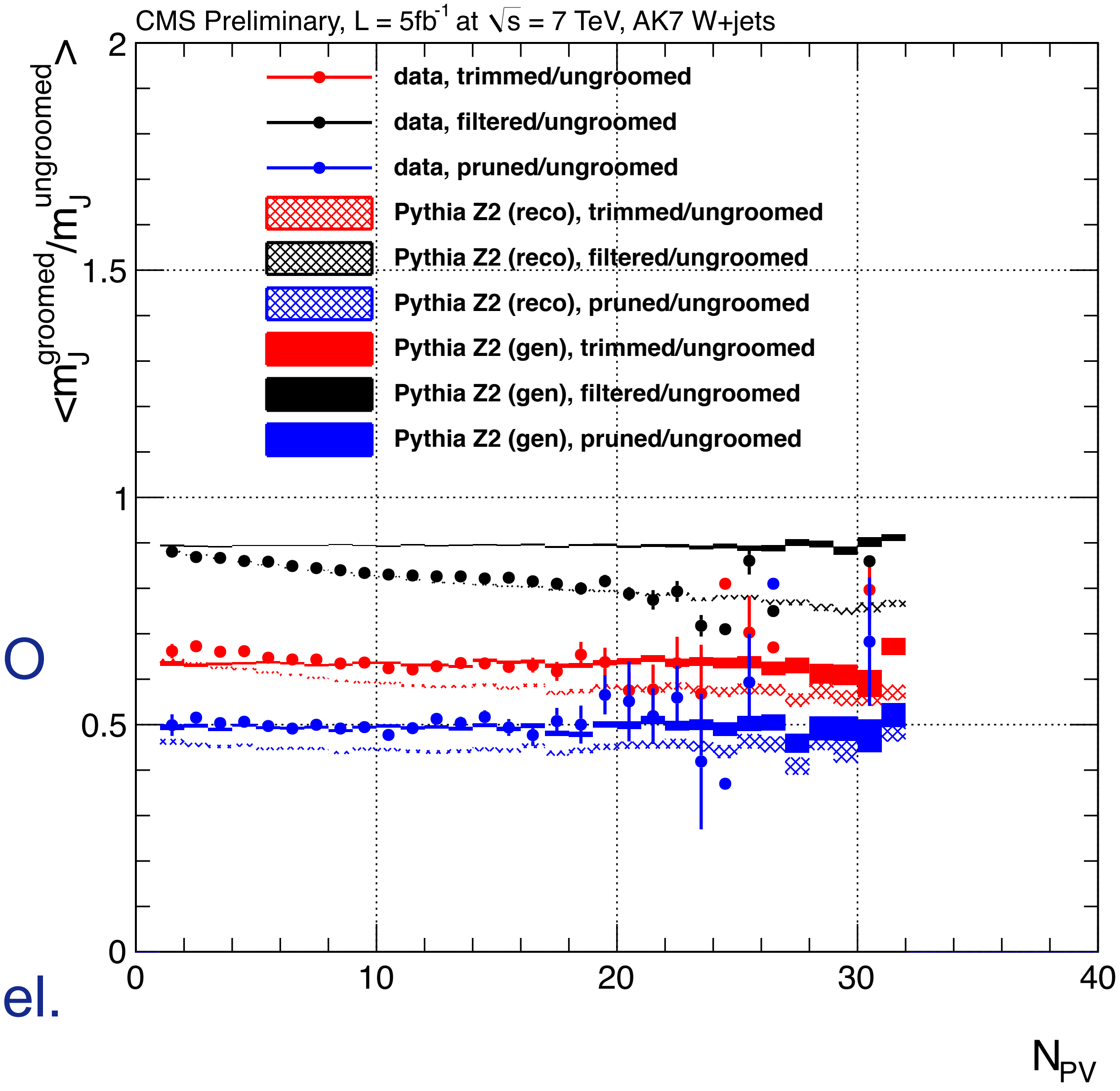
Dijets
QCD: Pythia6 Tune Z2/Pythia8 Tune 4c/ Herwig++ 23
V+jets
WW,WZ,ZZ: Pythia6 Tune Z2
ttbar and single top: MadGraph+Pythia6
W/Z+jets: MadGraph+Pythia6 Tune Z2/Herwig++ 23

- Motivation for jet pT bins come from trigger thresholds in the dijet analysis
- Lepton selections driven by increasing single electron trigger threshold throughout 2011

Performance versus pileup for groomed jets (II)

inclusive in NPV



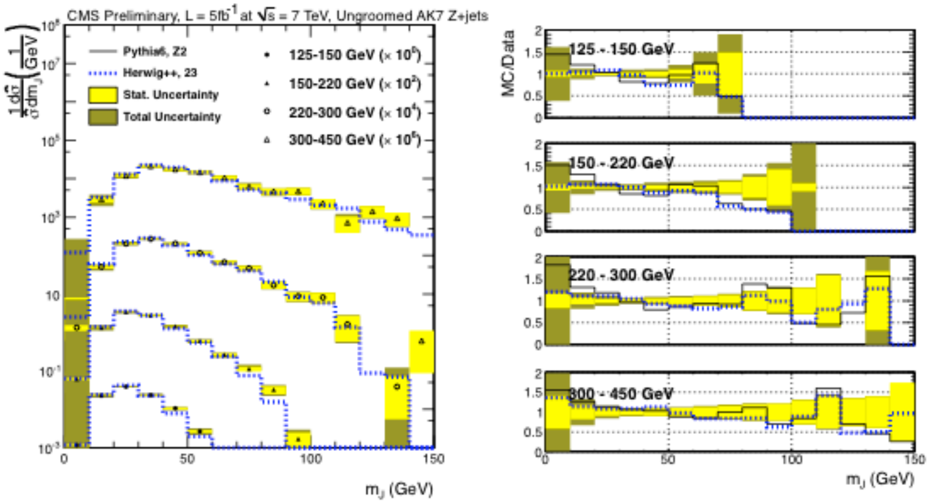


- Main differences between GEN Filtered jets and RECO Filtered jets from pileup.

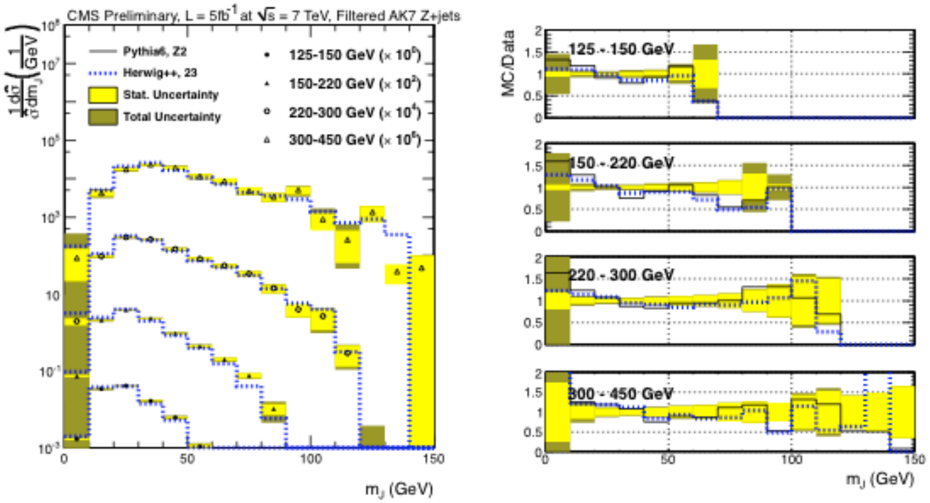
- Trimmed and Pruned jets differences are convoluted with parton showering model.

Unfolded distributions, Z+jet

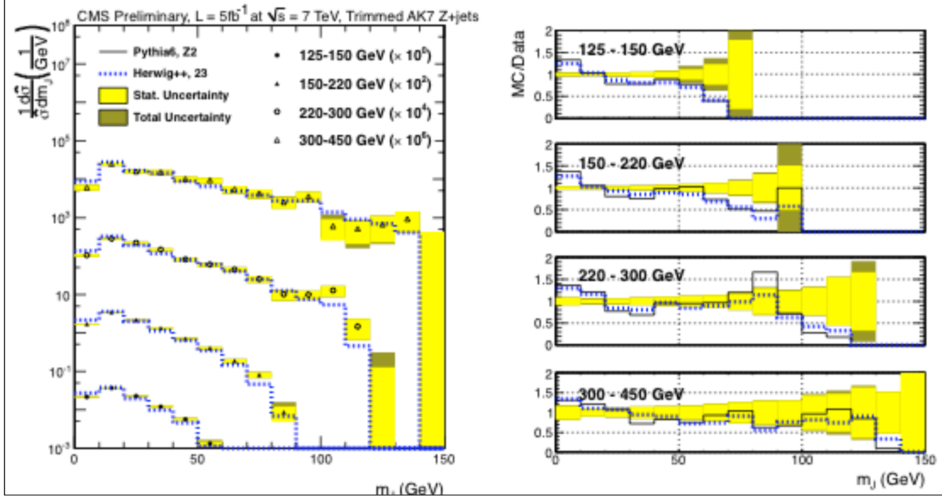
AK7



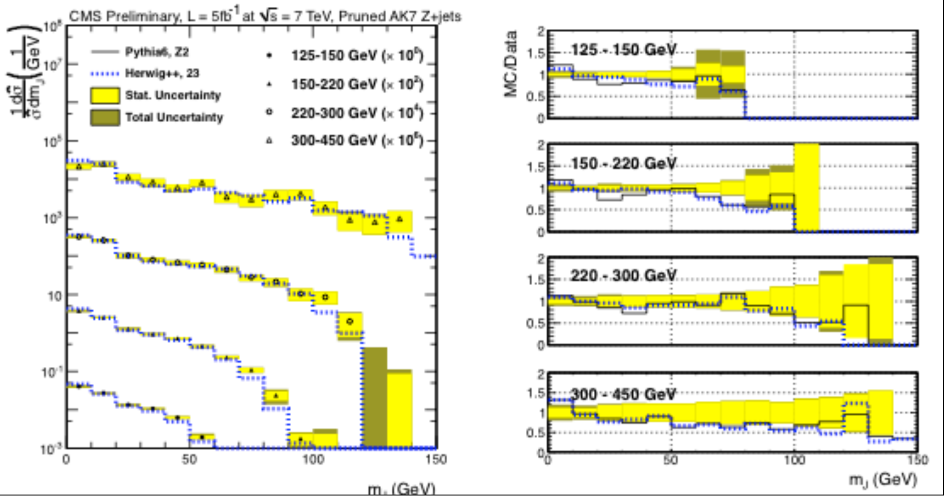
AK7, Filtered



AK7, Trimmed

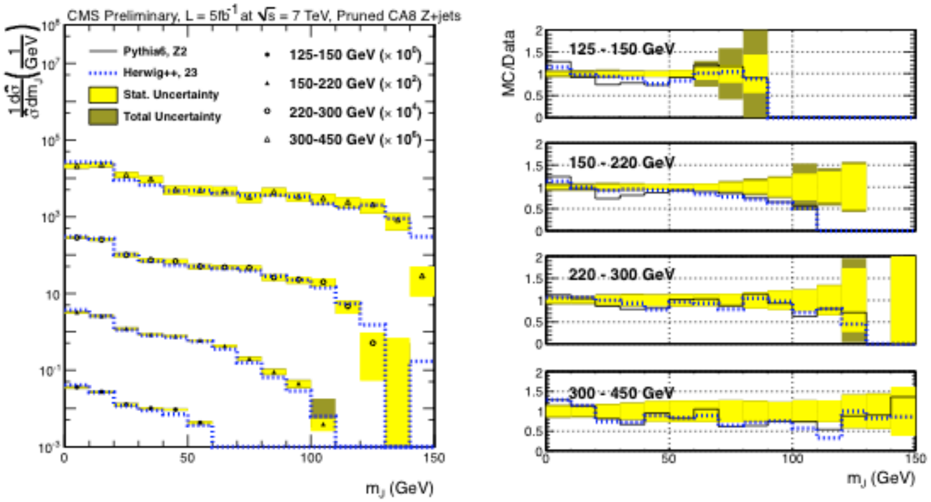


AK7, Pruned



Unfolded distributions, Z+jet

CA8, Pruned



CA12, Filtered