CMS results on boosted objects & jet substructure

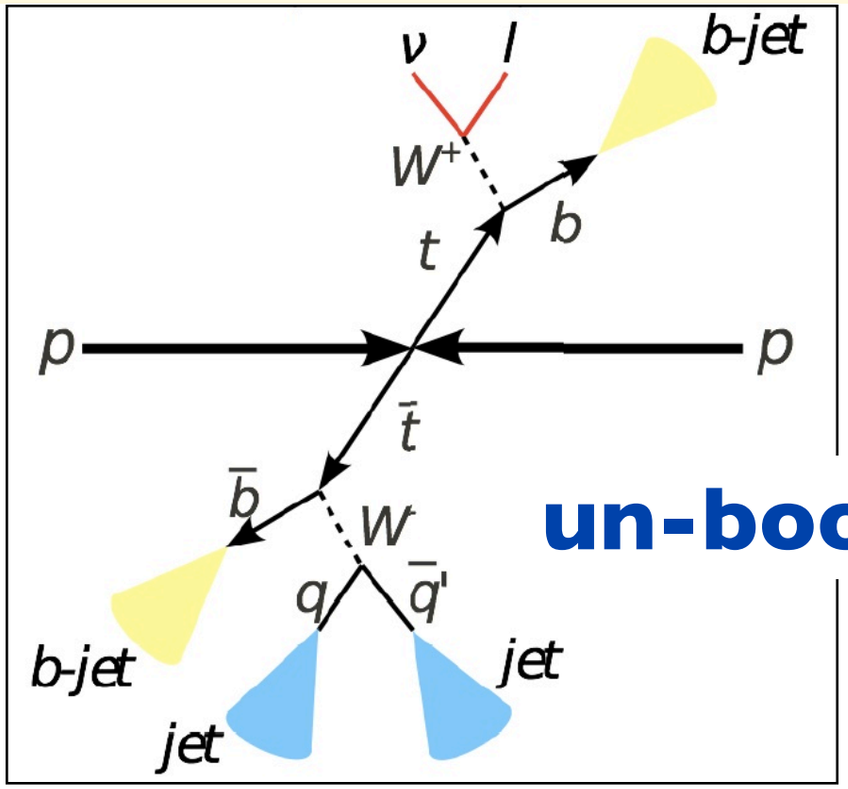
Alexander Schmidt (University of Hamburg)
on behalf of the CMS Collaboration

HEP 2013
Stockholm
18-24 July 2013

Introduction

domain of 'classical' analysis methods:

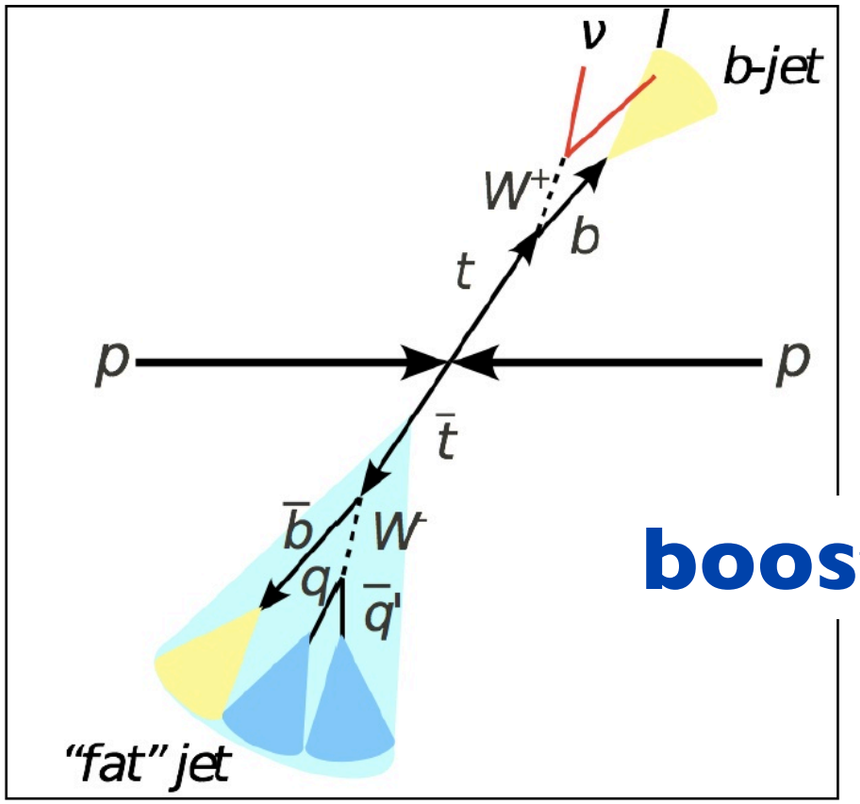
- isolated leptons (e.g. relevant for trigger)
- multiple, well separated 'small' jets



un-boosted

boosted regime:

- classical methods fail
- dedicated tools to resolve jet-substructure
- top-tagging, H-tagging, W-tagging, ...



boosted

W tagging

early efforts in 2010 [CMS PAS JME-10-013]:

- top- and W-tagging algorithms commissioned in CMS with 36pb^{-1}

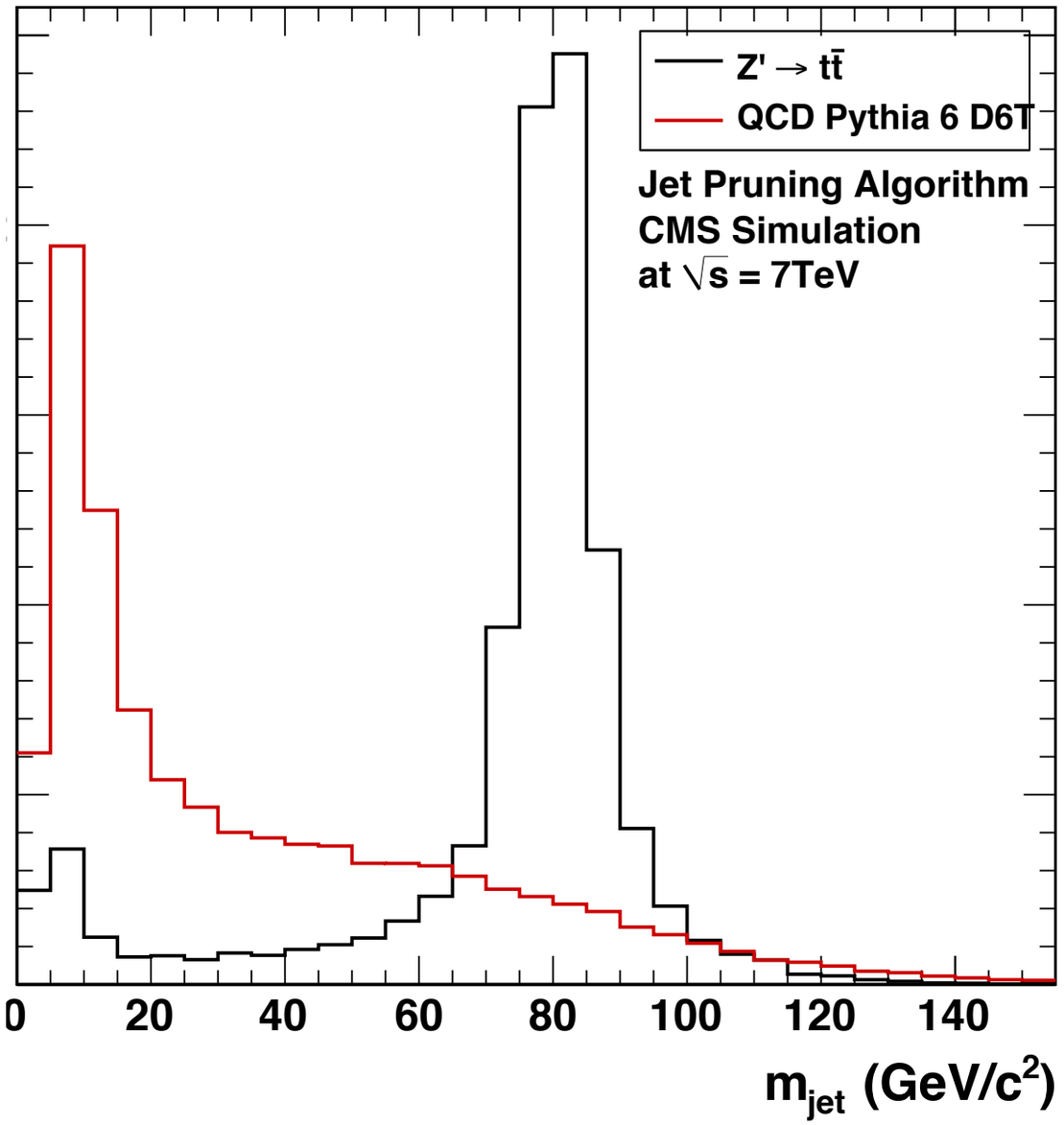
W-tagging [Phys Rev D80 (2009) 051501]:

- starting with Cambridge-Aachen (CA) jets with distance parameter 0.8
- **pruning** removes soft clusters by reclustering the jet with requirements
- at each step of merging clusters *i* and *j* into a single cluster *p*:

$$z_{ij} \equiv \frac{min(p_{T,i}, p_{T,j})}{p_{T,p}} > z_{cut}$$

$$\Delta R_{ij} < D_{cut} = \alpha \times \frac{m_J}{p_T^J}$$

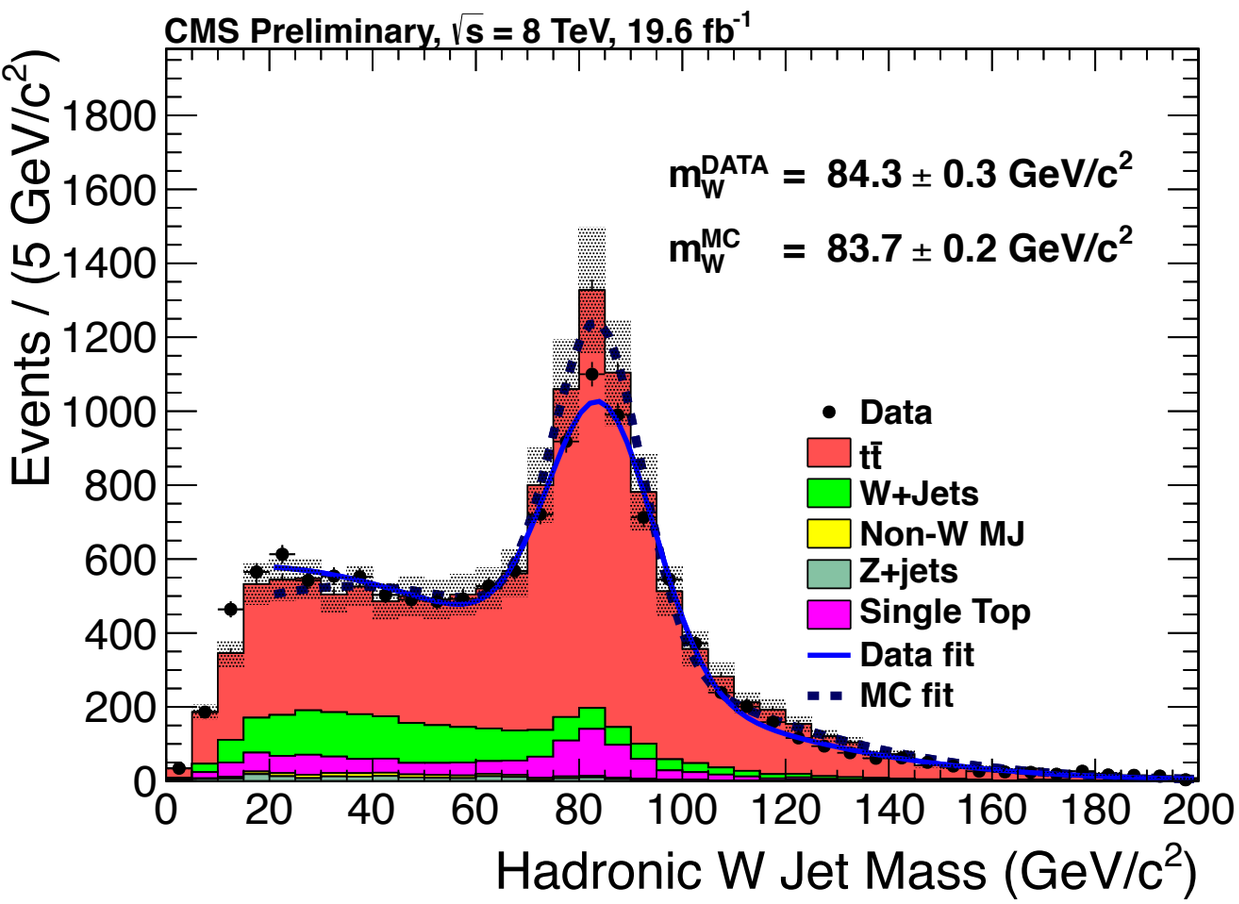
- where z_{cut} = 0.1 and α=0.5
- identify W jets by requiring 2 pruned subjets, m_{jet} in [60, 100], mass drop $m_1/m_j < 0.4$



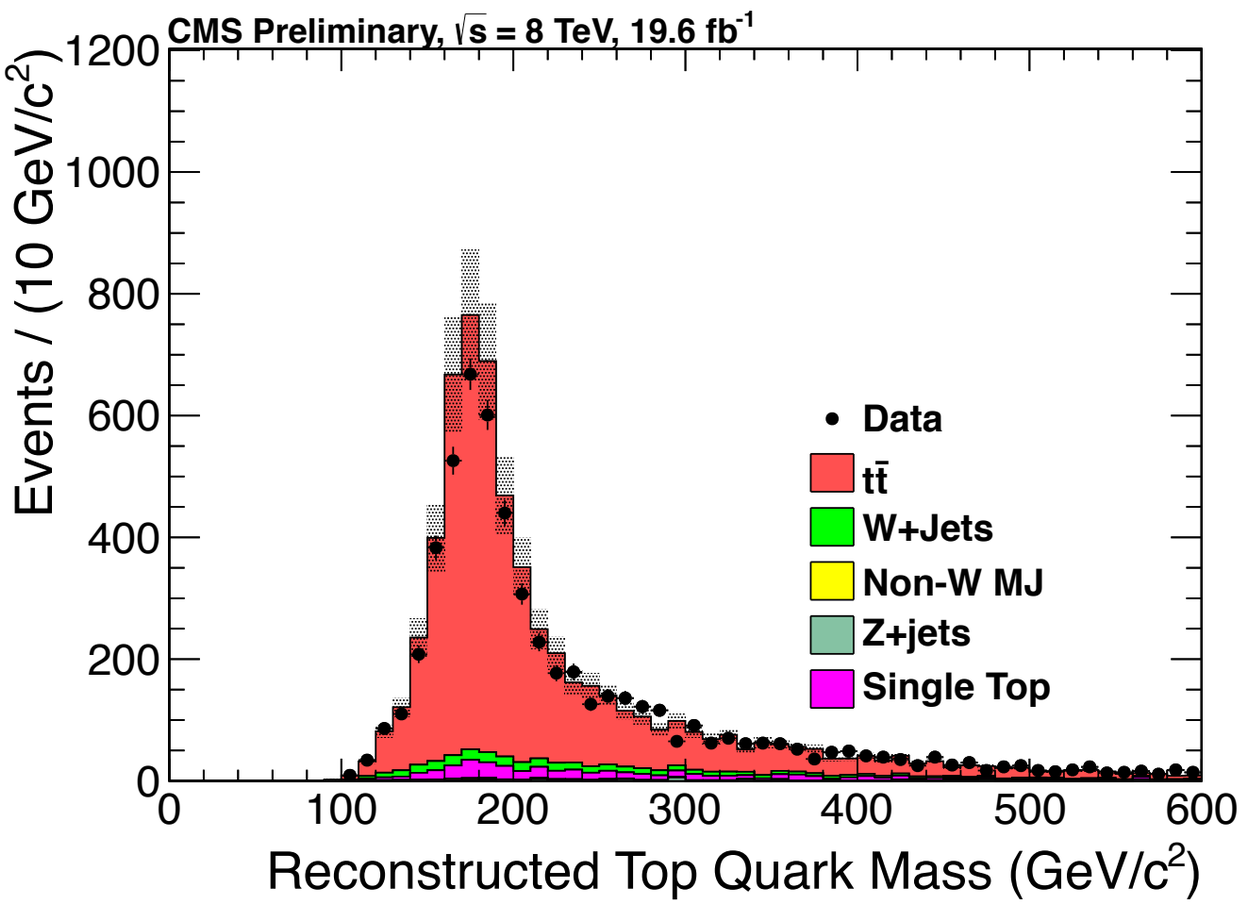
W tagging: in data

high purity top-quark selection with a high-pt muon + jet:

W-tagged jet mass:



W-jet combined with a b-jet:



→data agree well with simulation

top-tagging

adapted JHU top-tagging algorithm

[PRL 101 (2008) 142001]:

- CA jets with distance parameter 0.8
- reverse clustering sequence, examine clusters pairwise
- continue until separated clusters found

$$\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} > 0.4 - A \times p_{\mathrm{T}}^{C}$$

where p_T^C is the parent cluster

- check momentum fraction criterion

$$p_{\mathrm{T}}^{\mathrm{cluster}} > \delta_{\mathrm{p}} \times p_{\mathrm{T}}^{\mathrm{hardjet}}$$ for

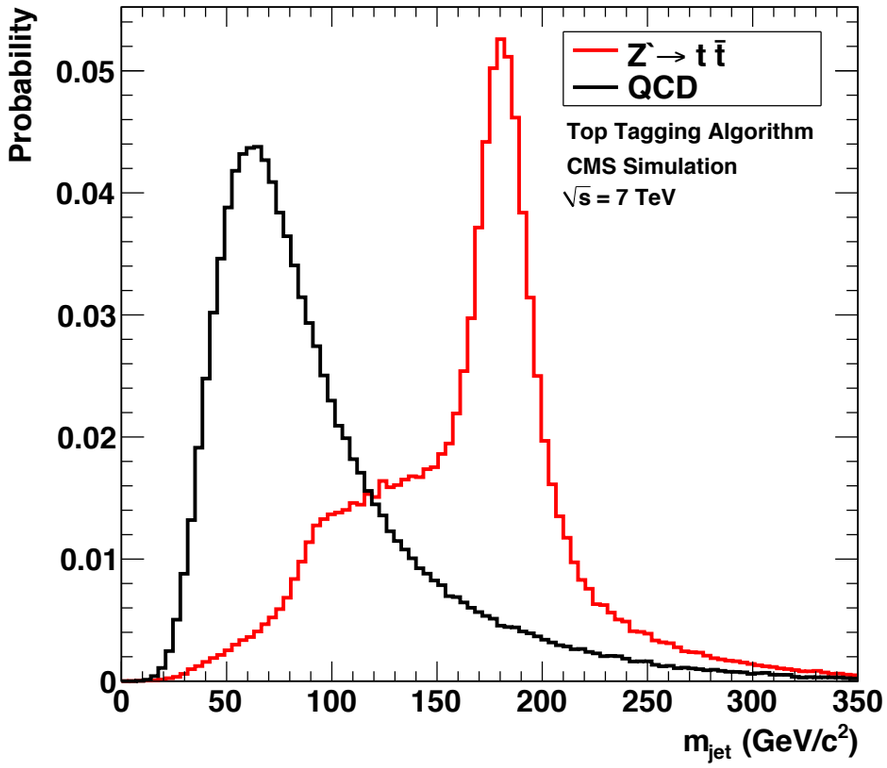
to ignore low p_T subclusters

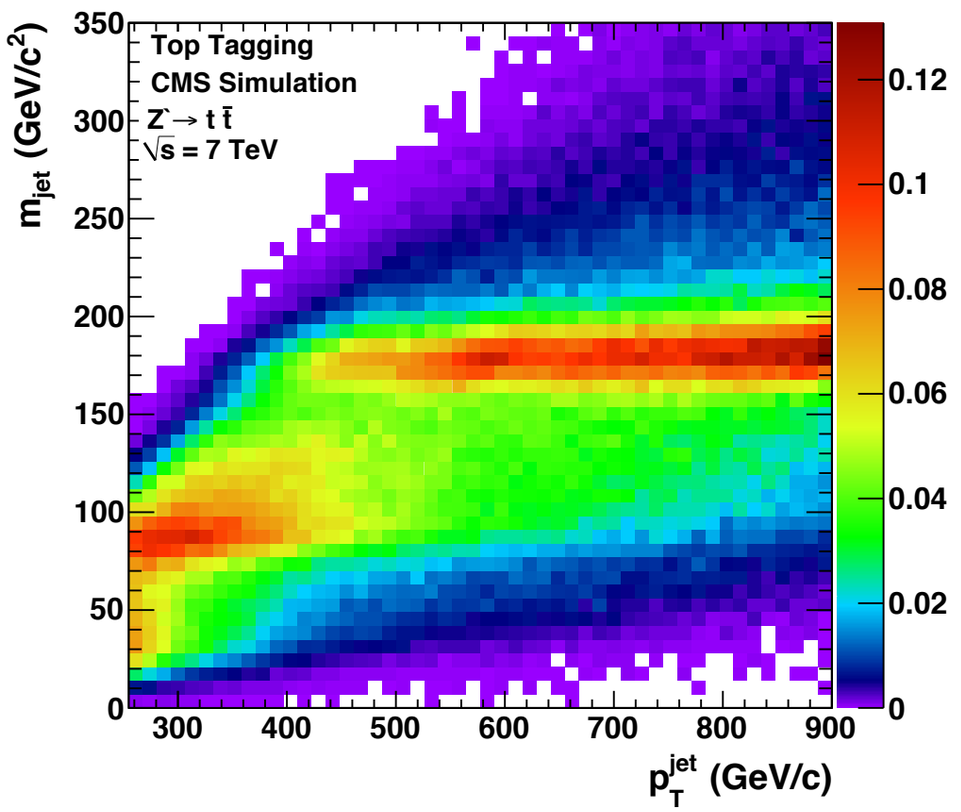
- variables used for top-tagging:

$$140 < m_{\mathrm{jet}} < 250 GeV/c^2$$

$$N_{\mathrm{subjets}} \geq 3$$

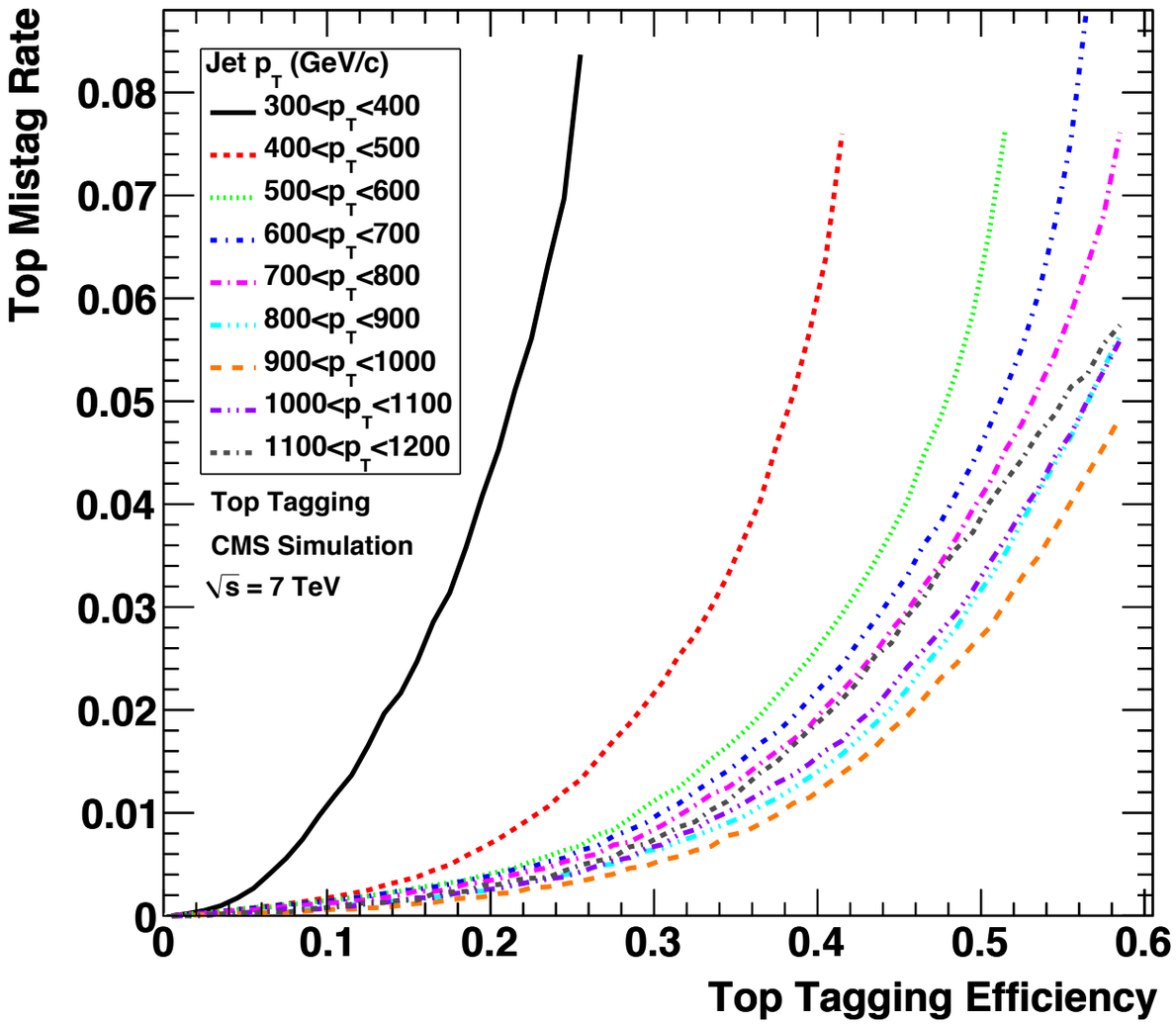
$$m_{\mathrm{min}} > 50 GeV/c^2$$ min pairwise mass





efficiencies and mistag

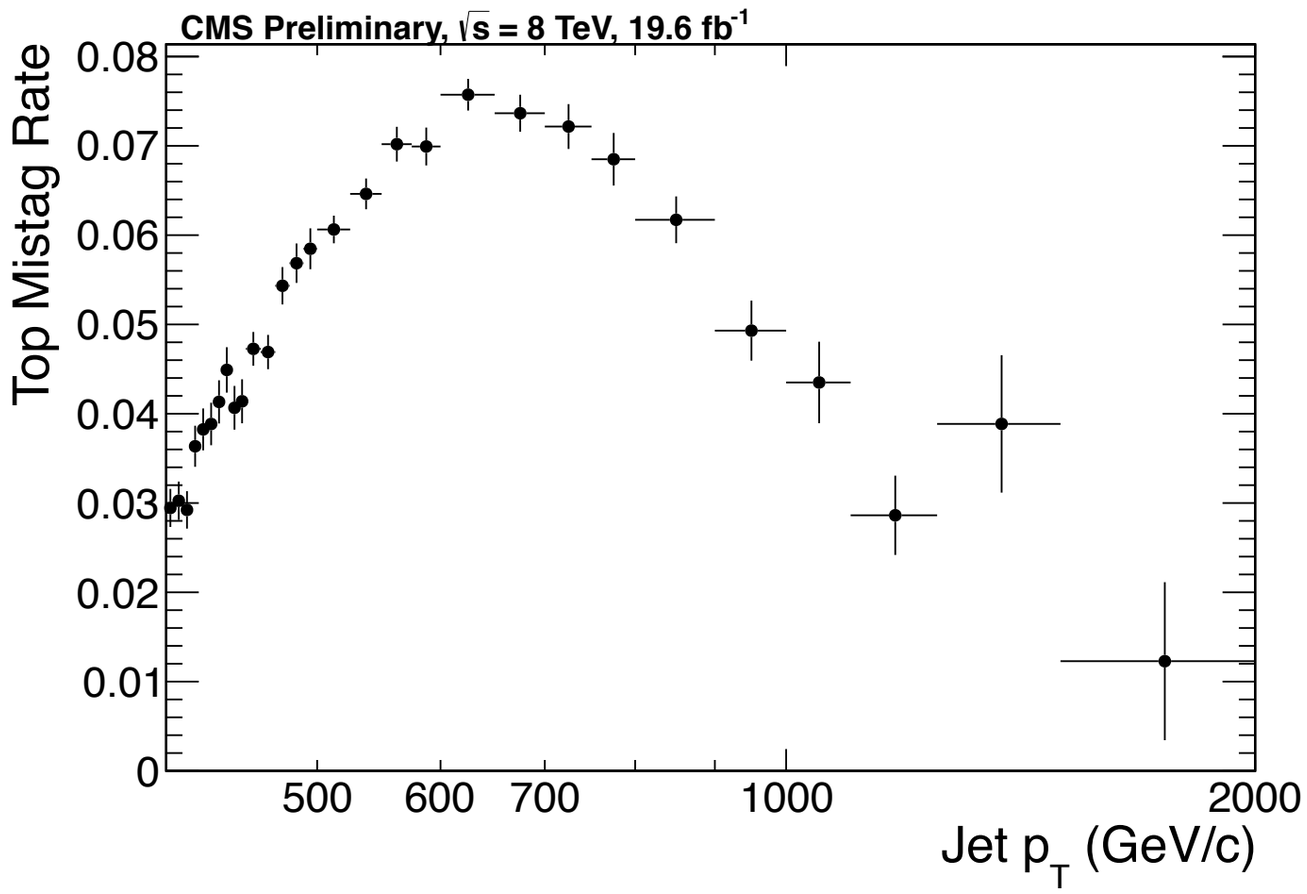
top-mistag (QCD) vs. efficiency in simulation



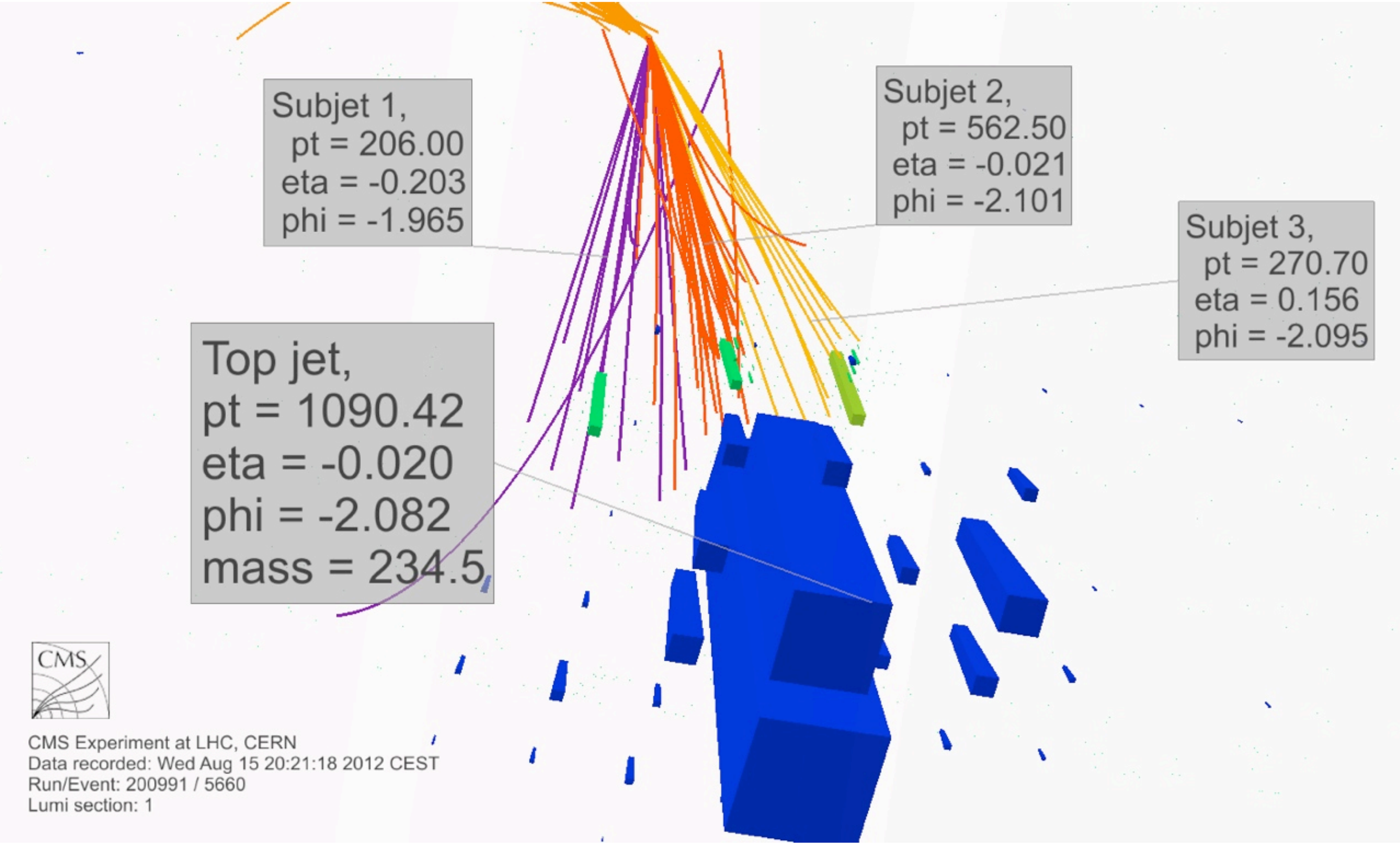
data/MC scale factor
for efficiency = 0.93 ± 0.04

mistag from data using top-depleted sample:

- two high-p_T jets, p_T>400GeV
- invert min pairwise requirement of top-tag algorithm for one jet
- if the other jet is top-tagged it is mistagged

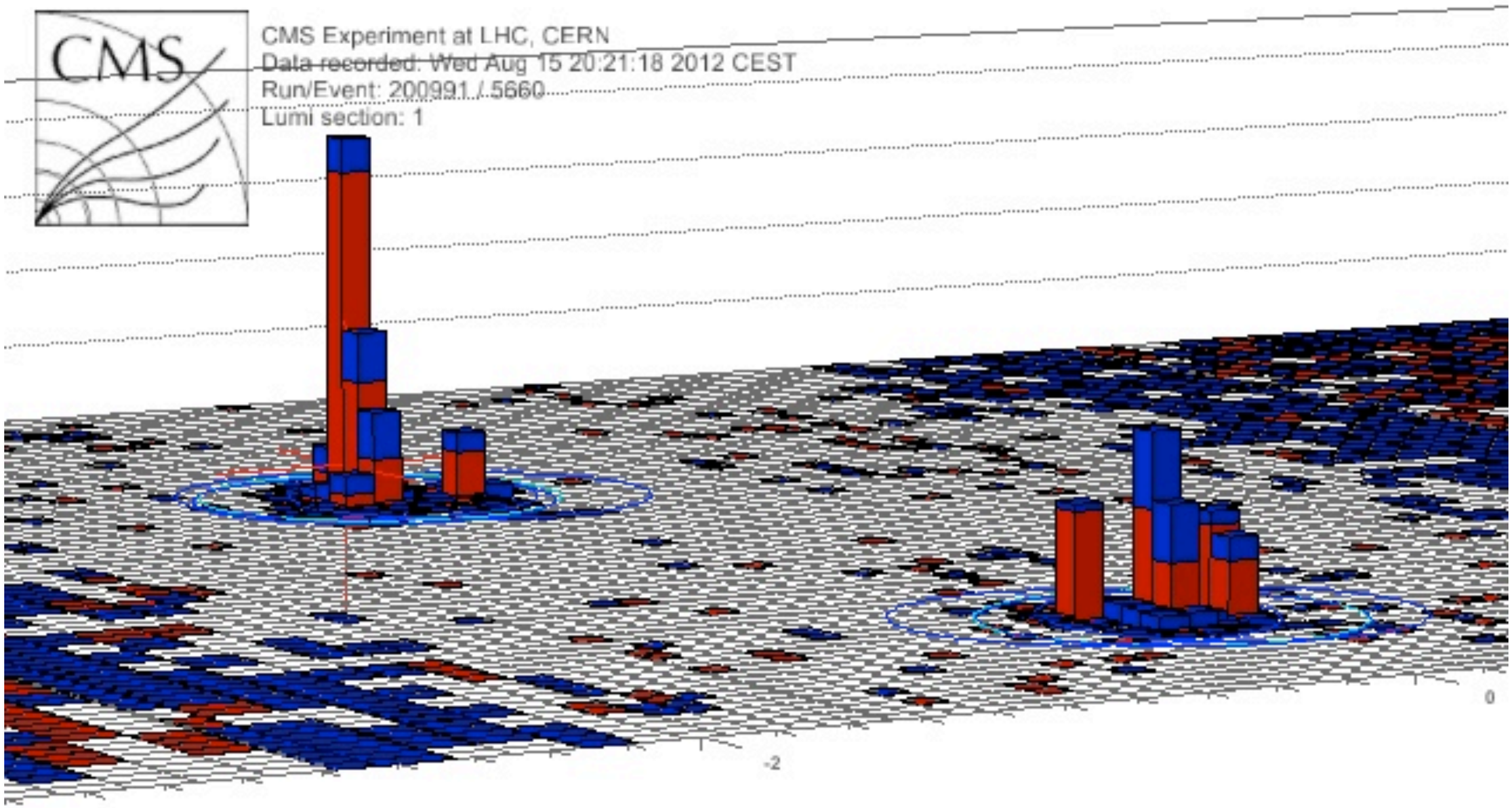


event displays



event displays

two back-to-back top-tagged jets



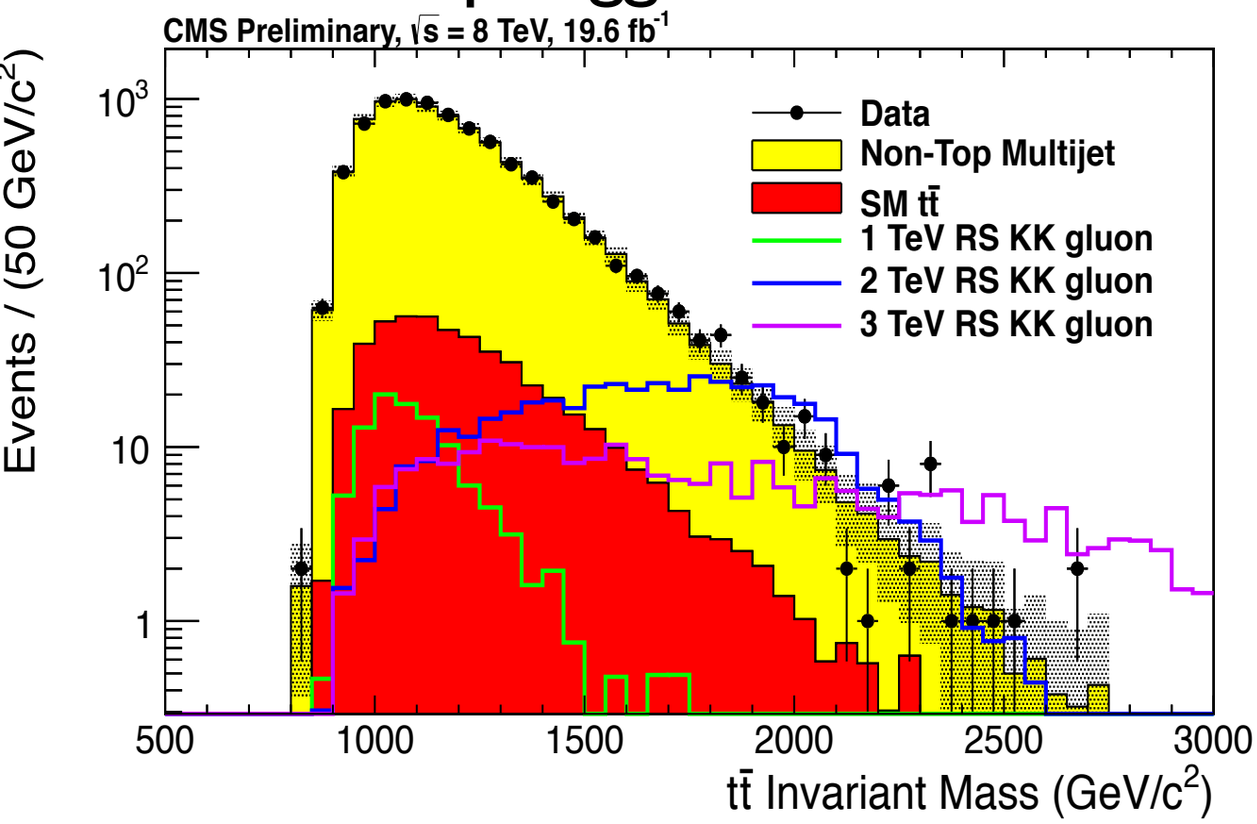
application in analysis

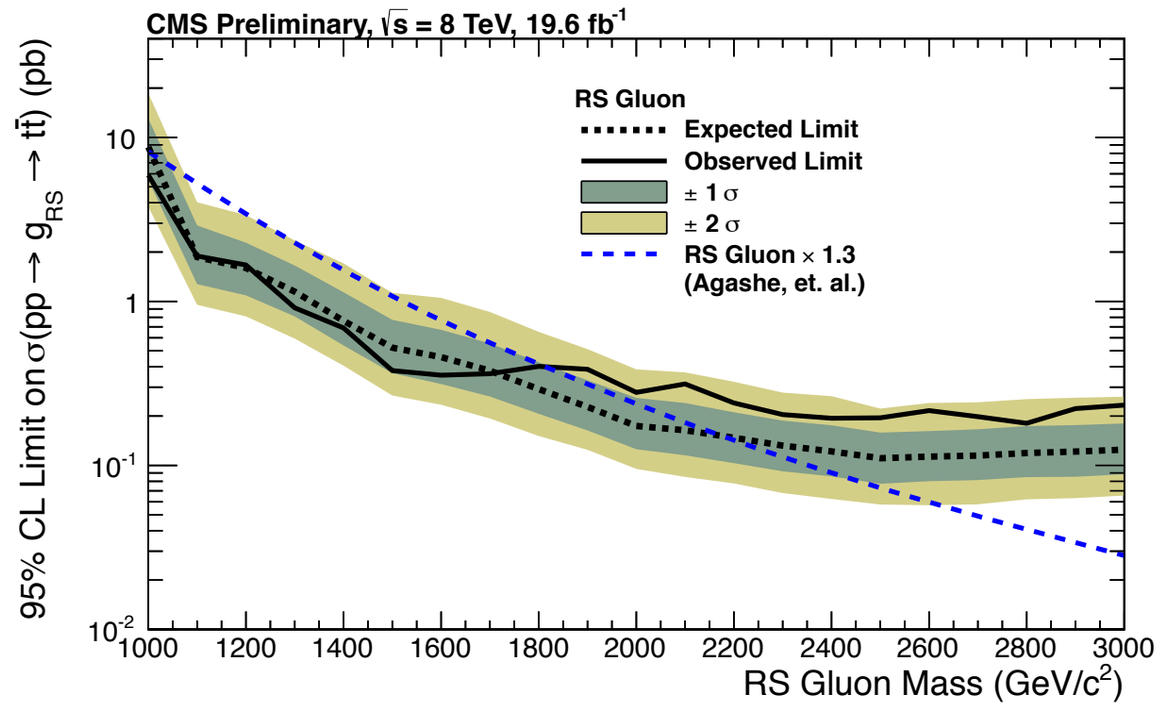
flagship for boosted search for new physics: $t\bar{t}$ resonances

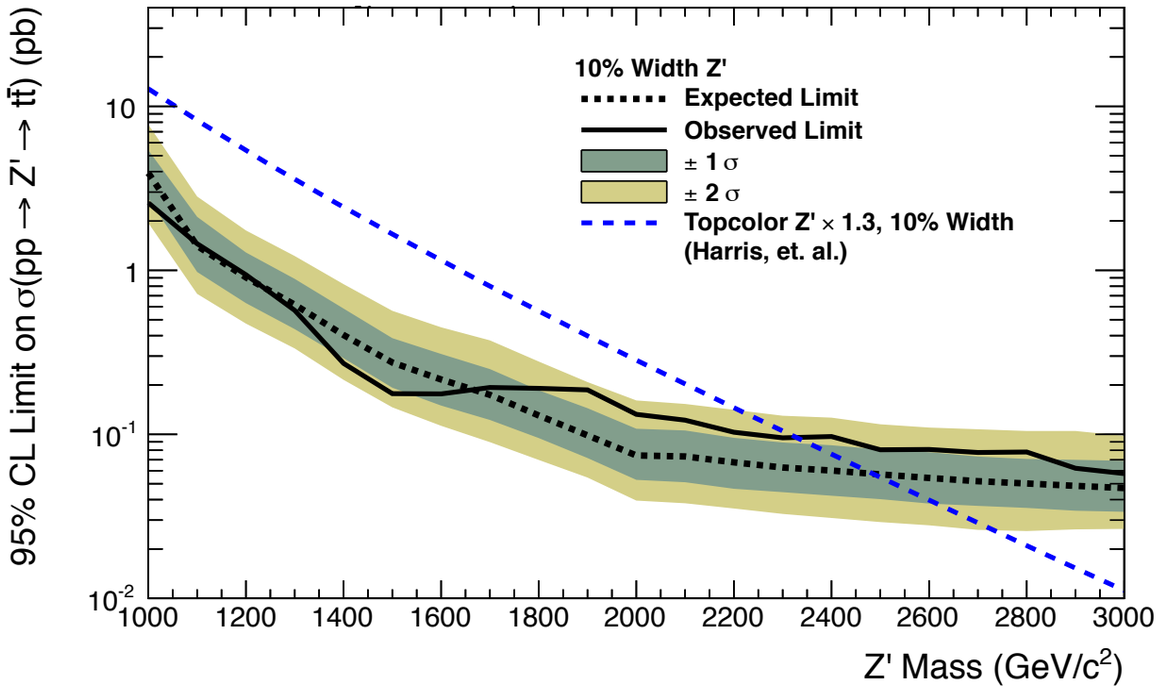
- various models: RS KK gluon, narrow ($\Gamma/m=0.01$) and wide ($\Gamma/m=0.1$) Z'
  [CMS PAS B2G-12-005]

all hadronic final state:

- two back-to-back jets
- both top-tagged



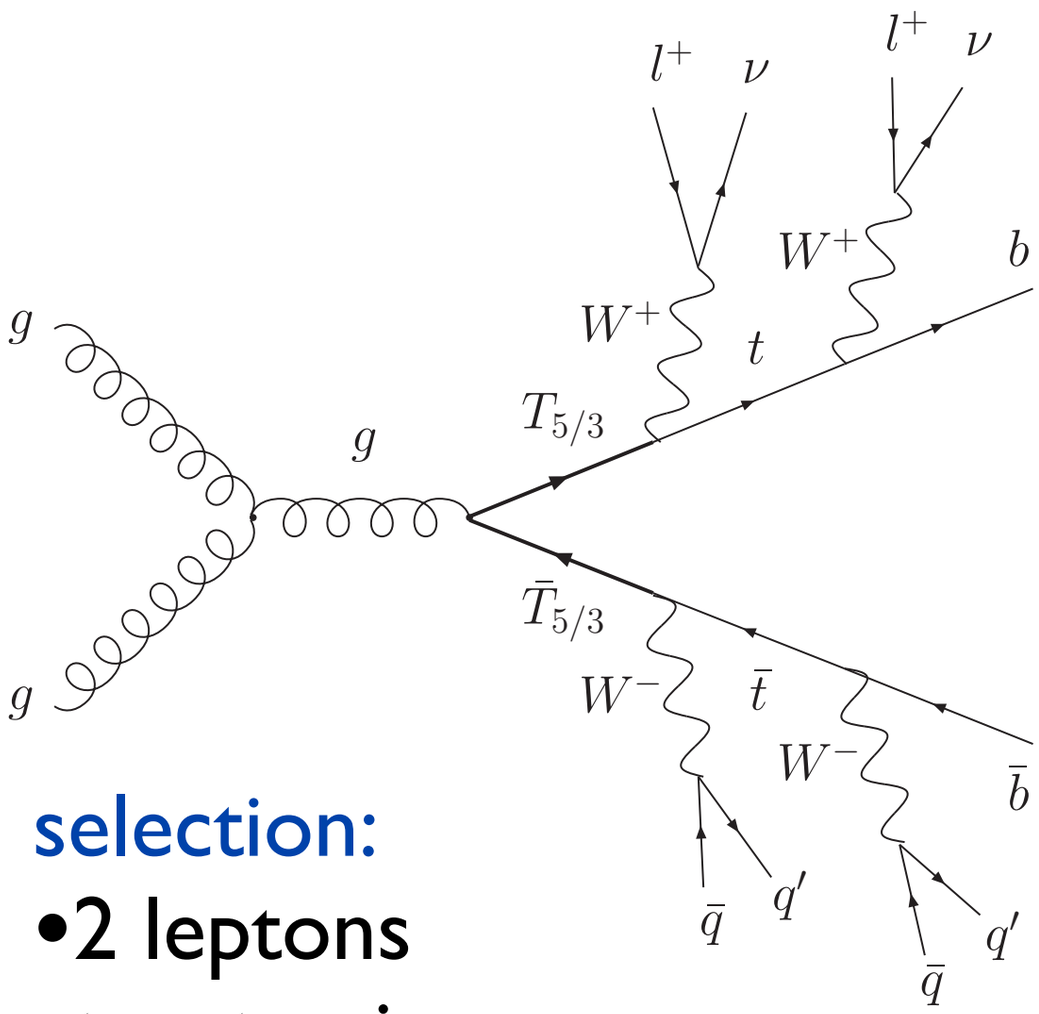




more details: R. Kogler, today 14:30, session Higgs and new Physics

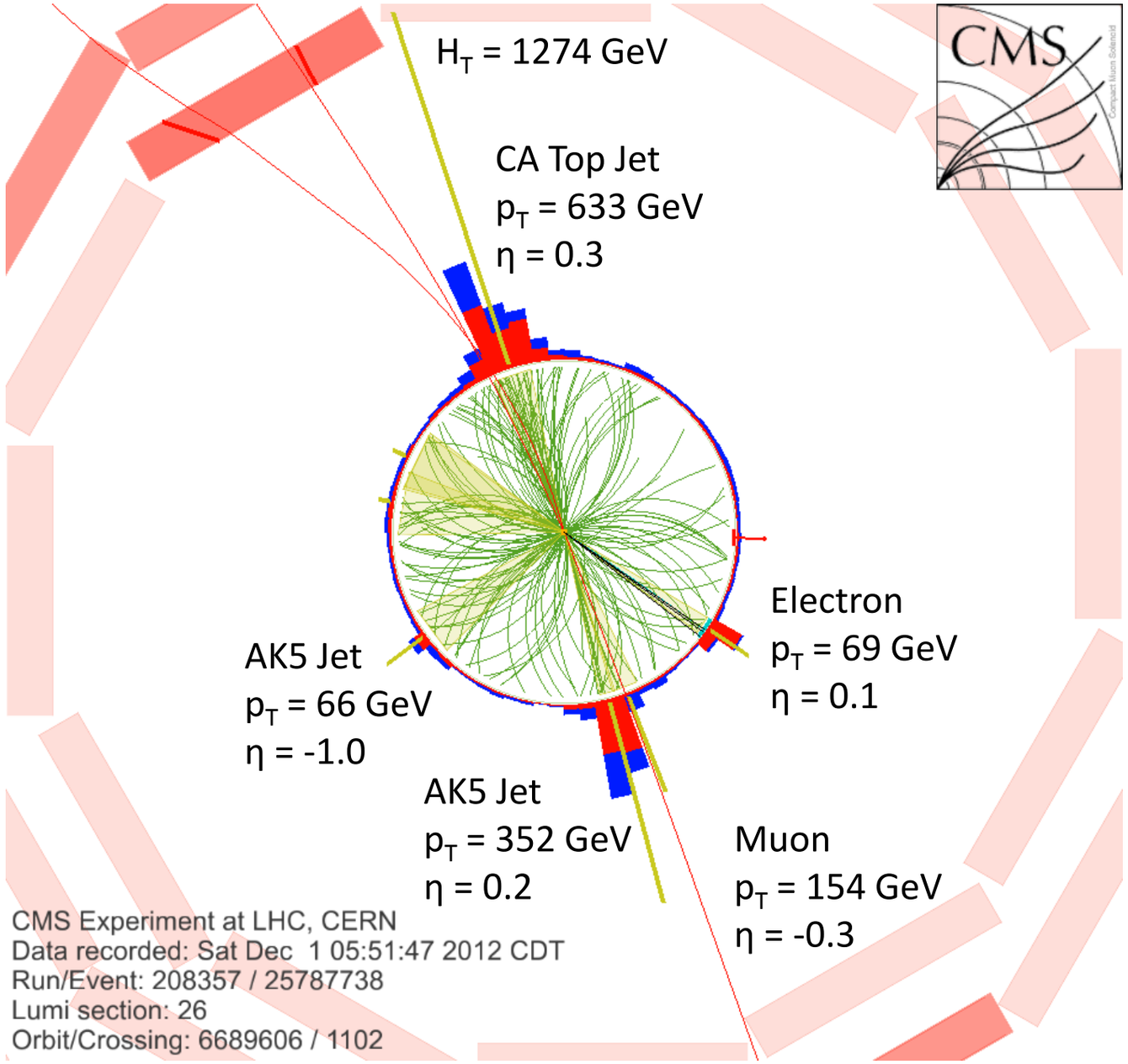
other search analyses

substructure techniques applied in several BSM searches, example: top partners with charge 5e/3 [CMS PAS B2G-12-012]

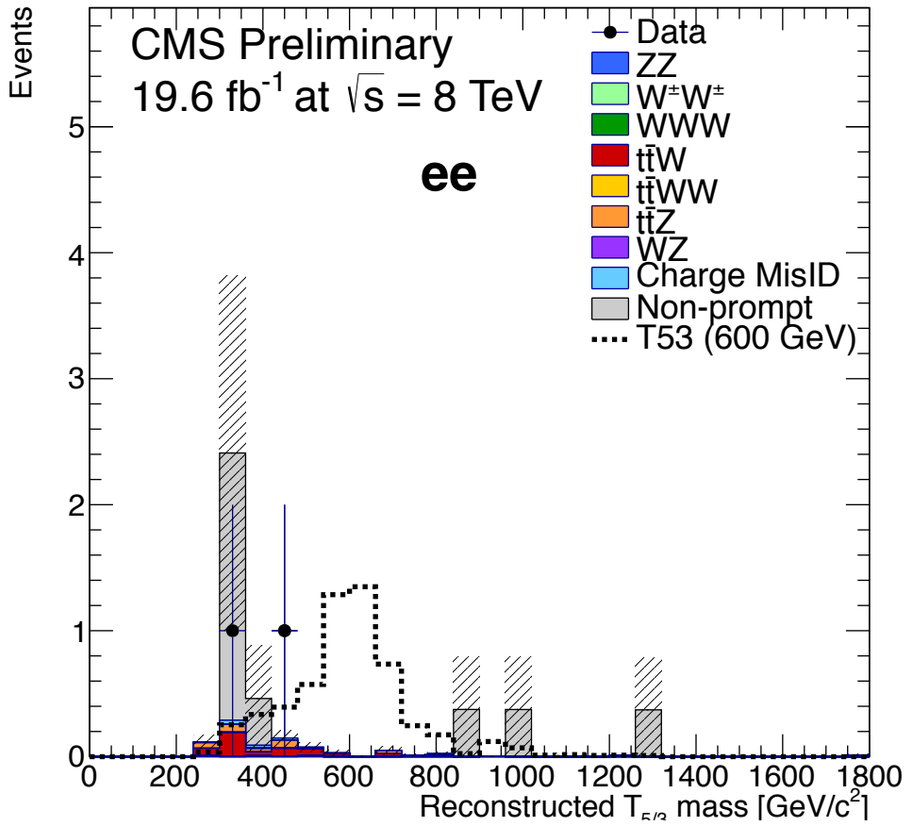


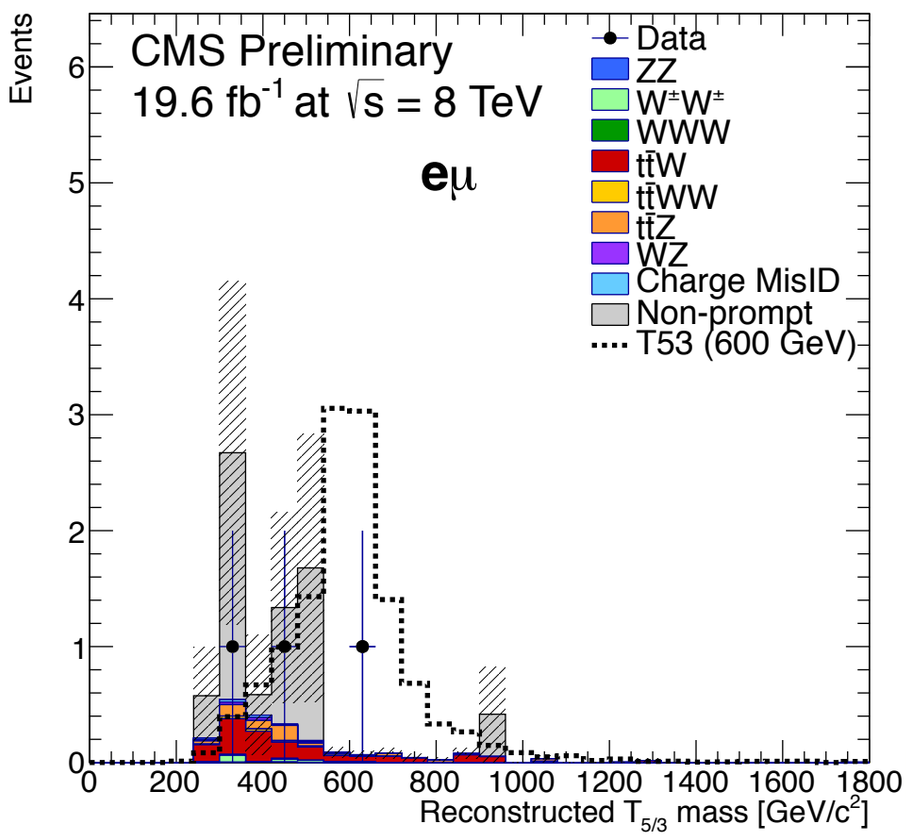
selection:

- 2 leptons
- top-tagging
- W-tagging

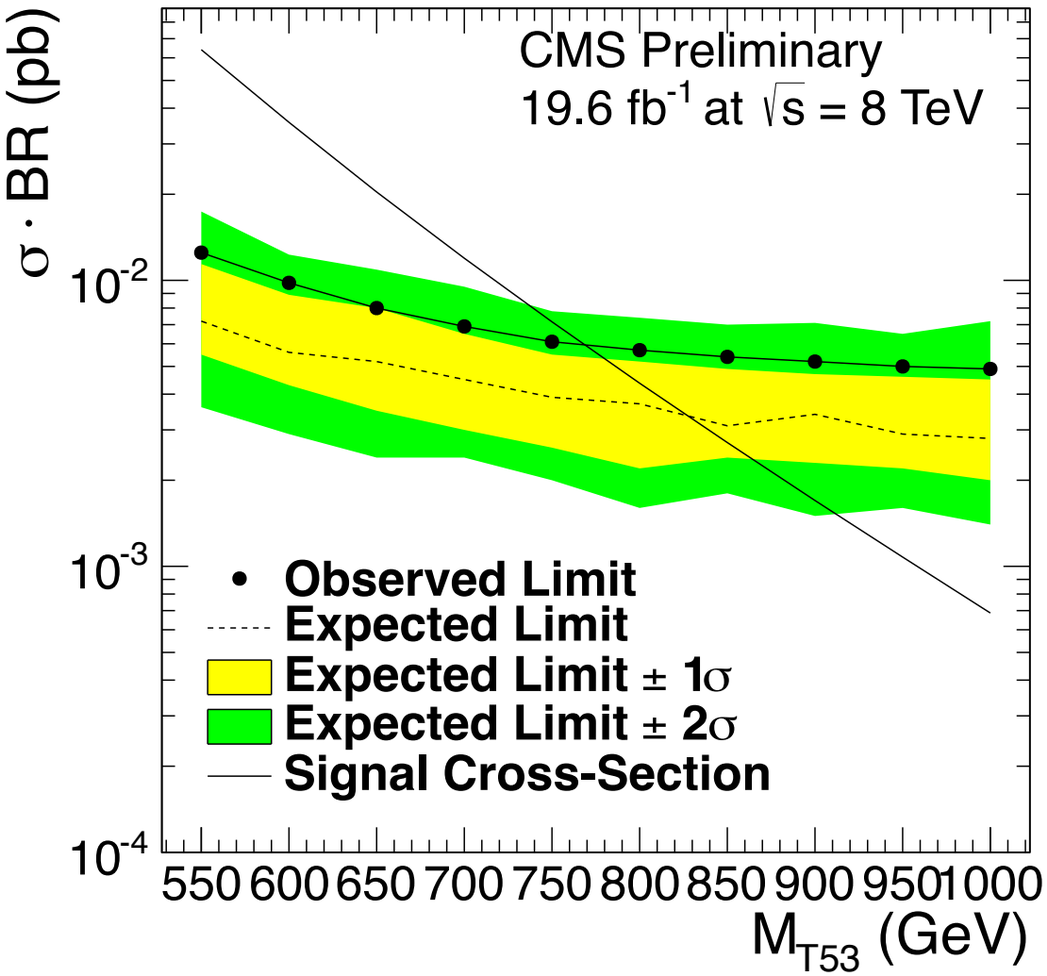


results from $T_{5e/3}$ search





- limit calculated from event yield in various channels
- data-derived backgrounds



- limit: expected 830, observed 770 GeV

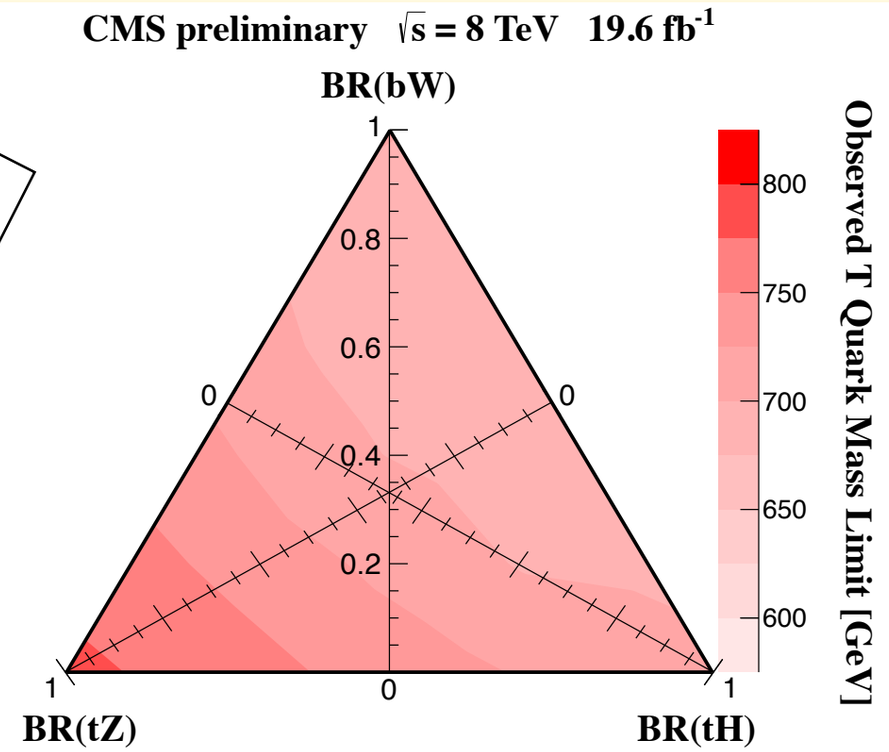
more details: D. Majumder, today 15:00, session Higgs and new Physics

and more

NEW:

W tagging applied in search for vector-like T' quarks [CMS PAS B2G-12-015]:

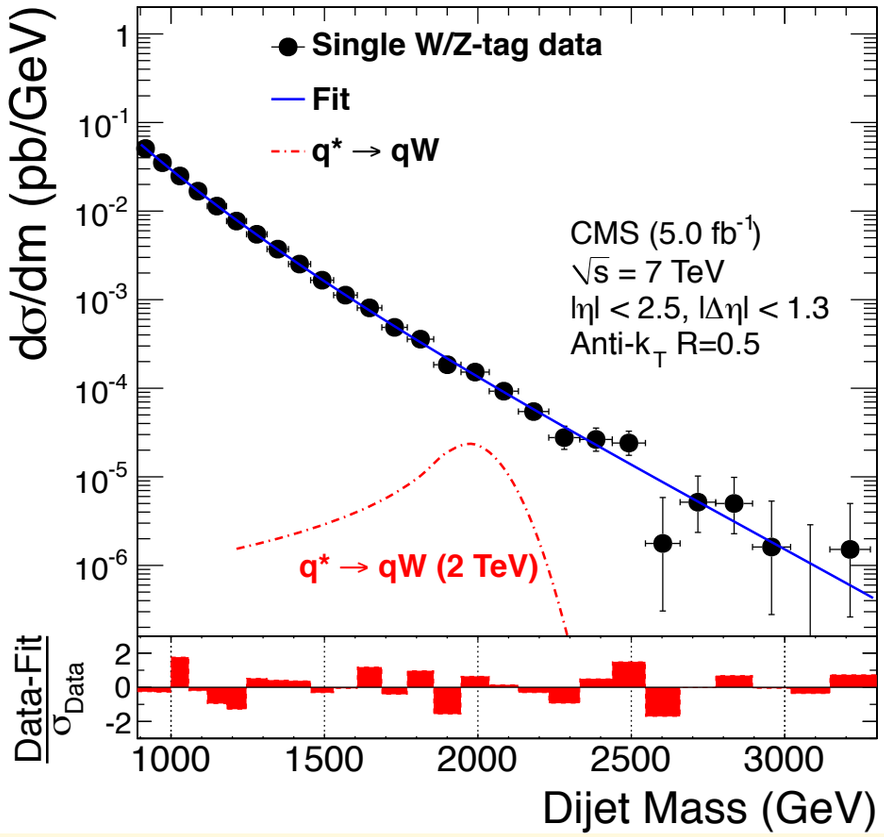
- final states with one or more leptons
- soon to come: all-hadronic final state



search for resonances in W/Z tagged dijet mass spectrum

[PLB 5 (2013) 040]:

- resonances decaying into qW, qZ, WW, WZ, ZZ, each decaying hadronic
- limits: 2.38 (2.15) TeV for qW (qZ)



jet mass

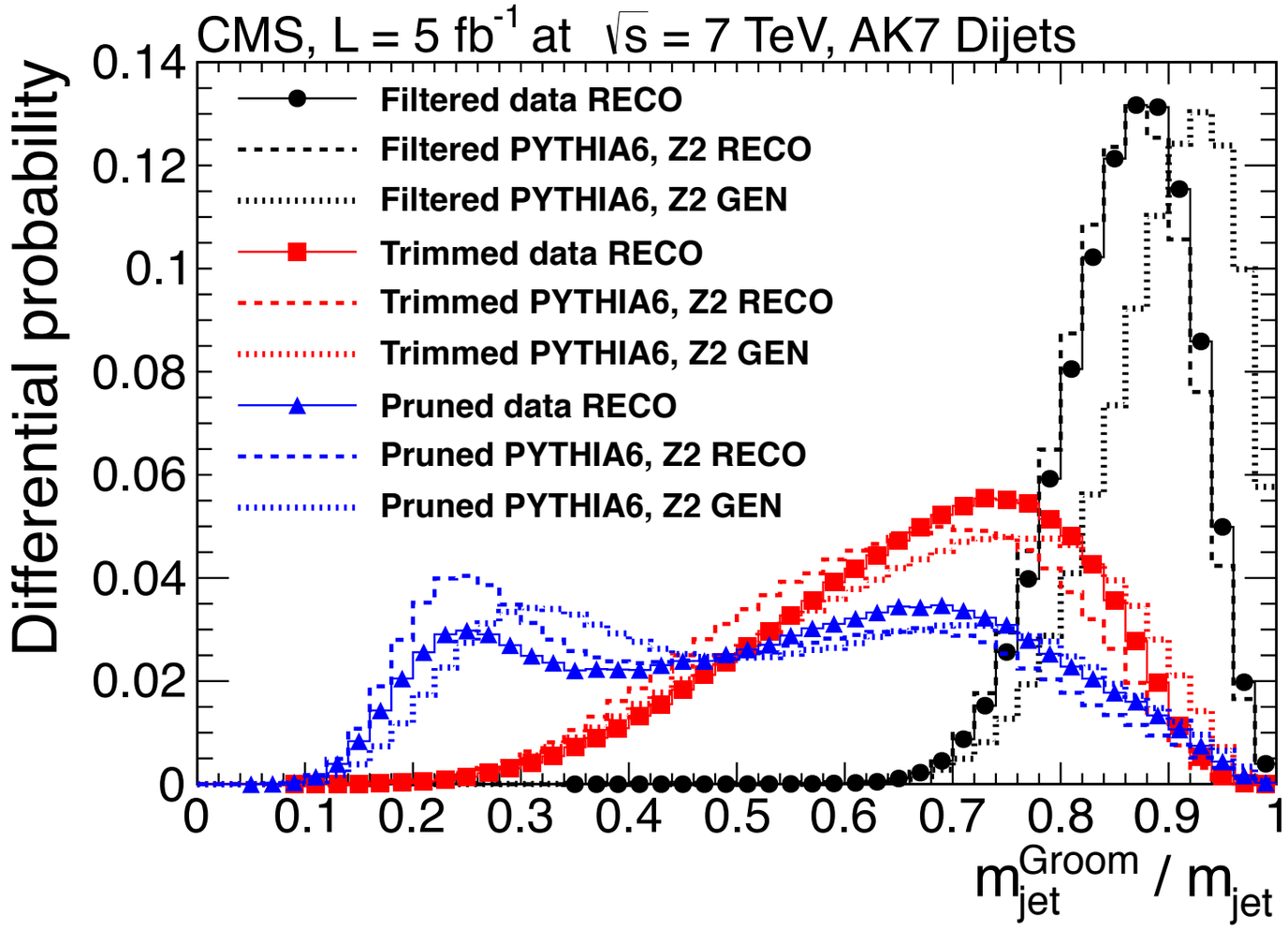
study of jet mass as an SM phenomenon [JHEP 05 (2013) 090]:

- verify grooming techniques (reducing UE, PU, soft radiation,...)
- verify modelling of internal jet structure by MC simulations

grooming methods:

- **pruning:** explained in slide 3
- **mass-drop/filtering:**
  - find symmetric splitting by reclustering with CA3
  - new jet defined by 4-vector sum of three hardest subjets
- **trimming:**
  - recluster with k_T algorithm with R=0.2
  - accept only subjets with $p_{Tsub} > f_{cut}\,\lambda_{hard}$ where λ_{hard} is chosen equal to p_{Tjet}

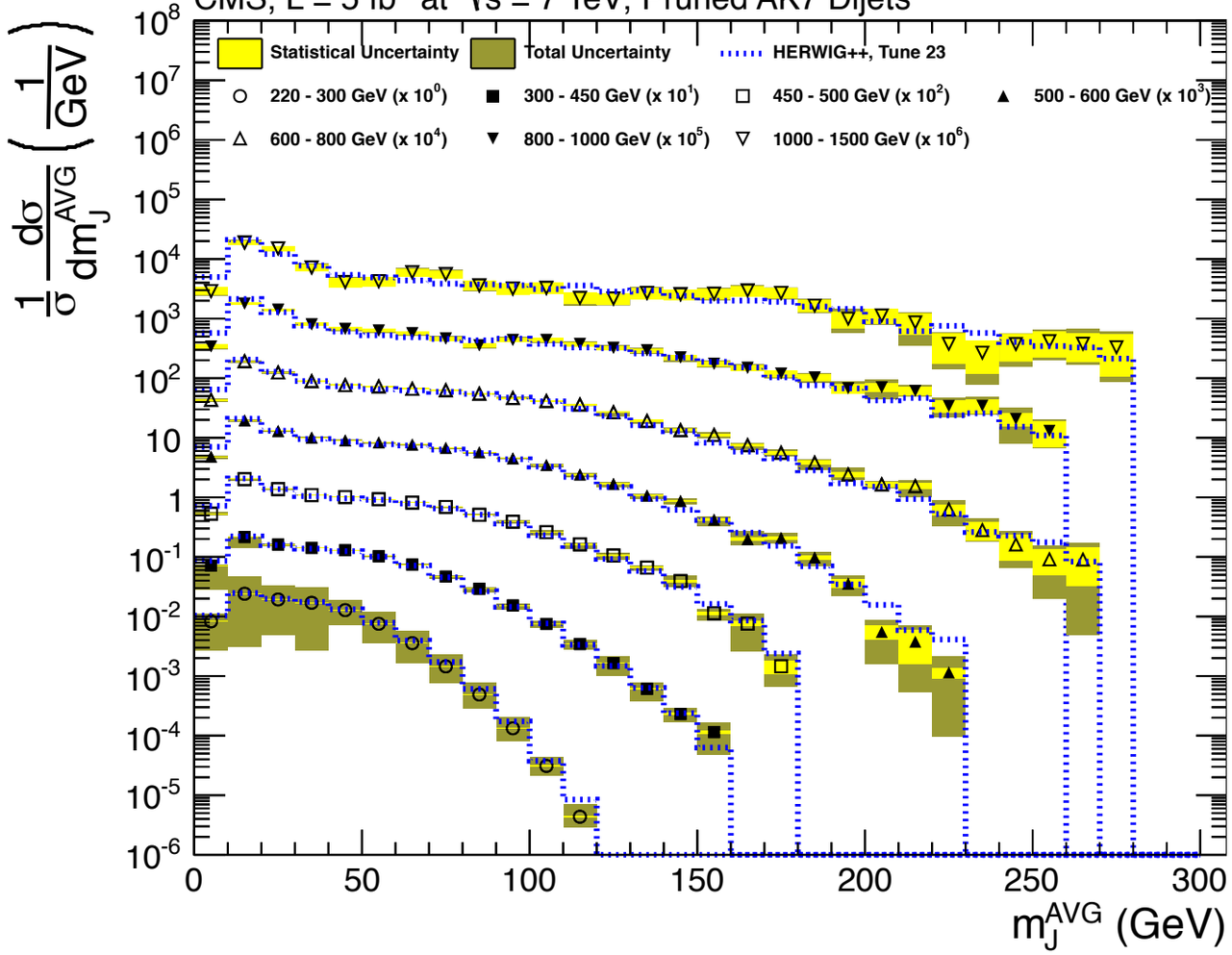
groomed AK7 jet mass compared to PYTHIA in a dijet sample



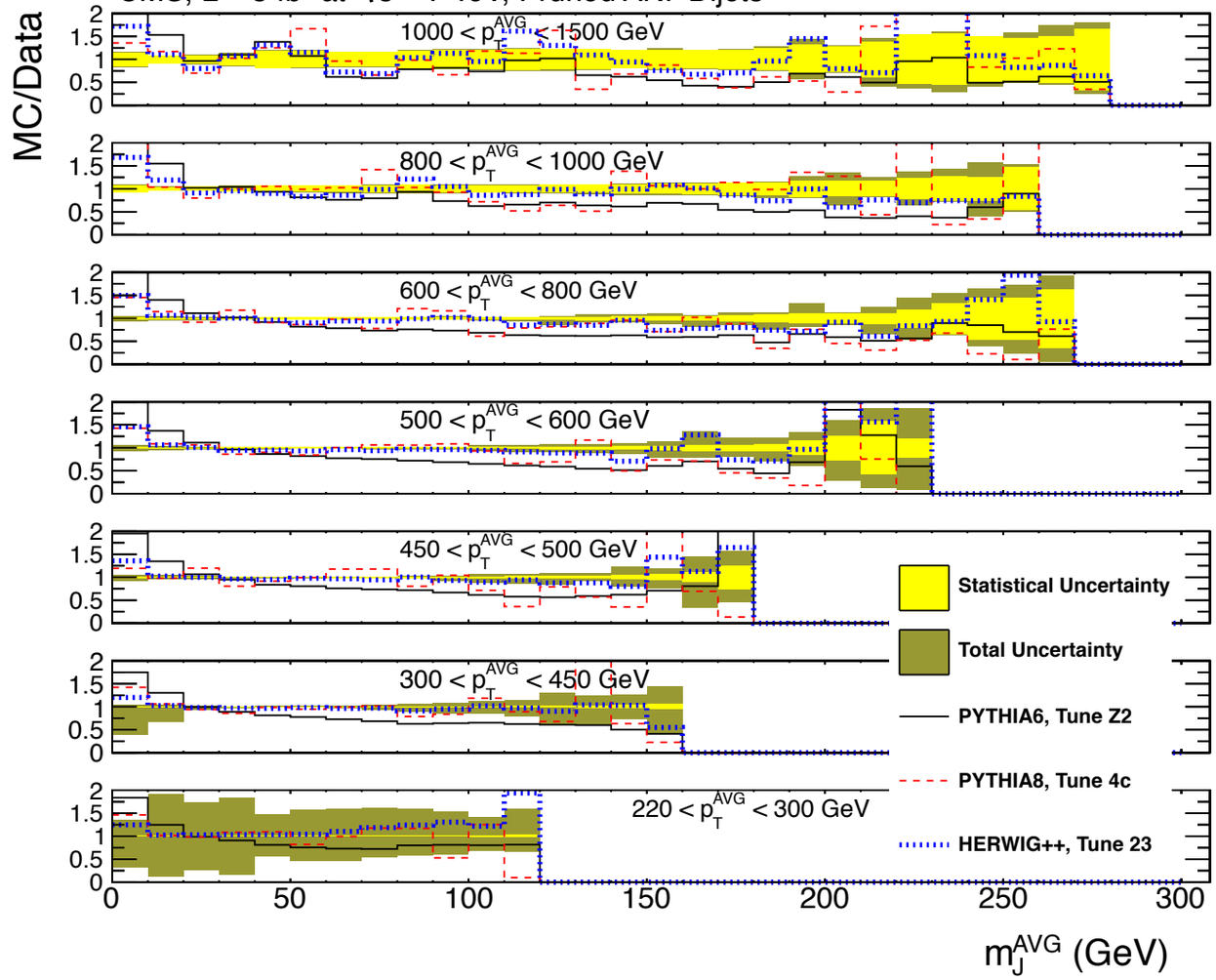
jet mass

results for pruned AK7 dijets

average jet mass



ratio simulation/data



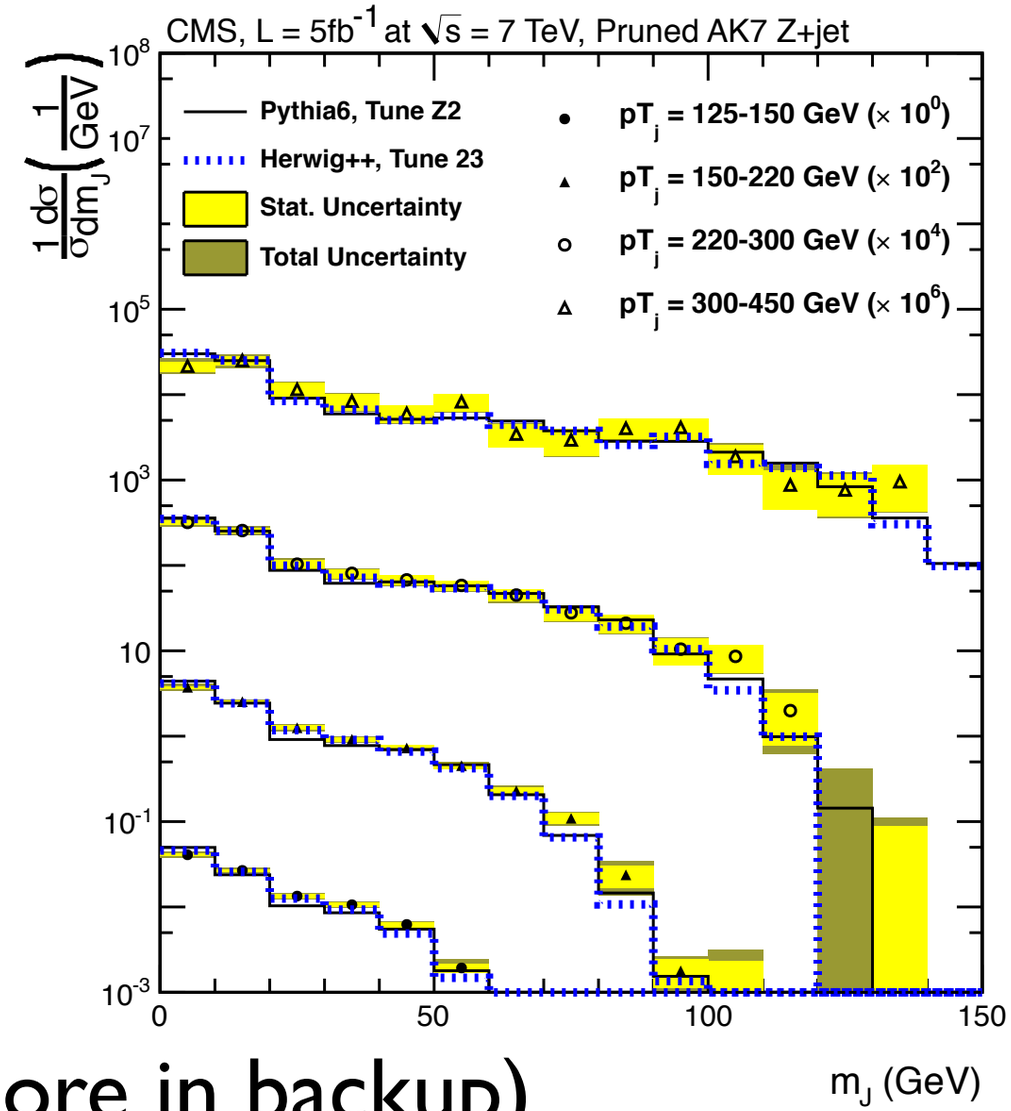
many more distributions (ungroomed, other grooming algorithms etc...) in backup slides

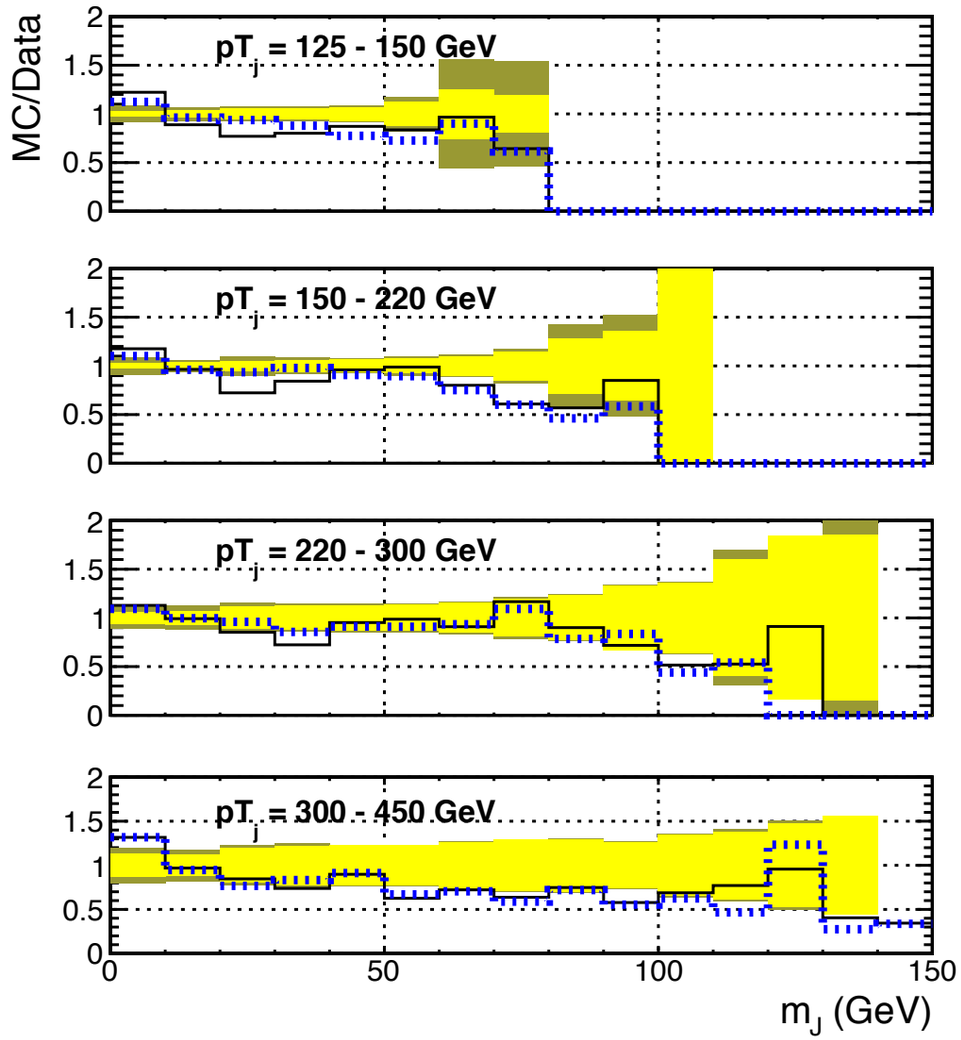
jet mass

results also obtained for V+jets events:

- V+jets preferably initiated from quarks
- multijets preferably from gluons

} complementary parton flavor content



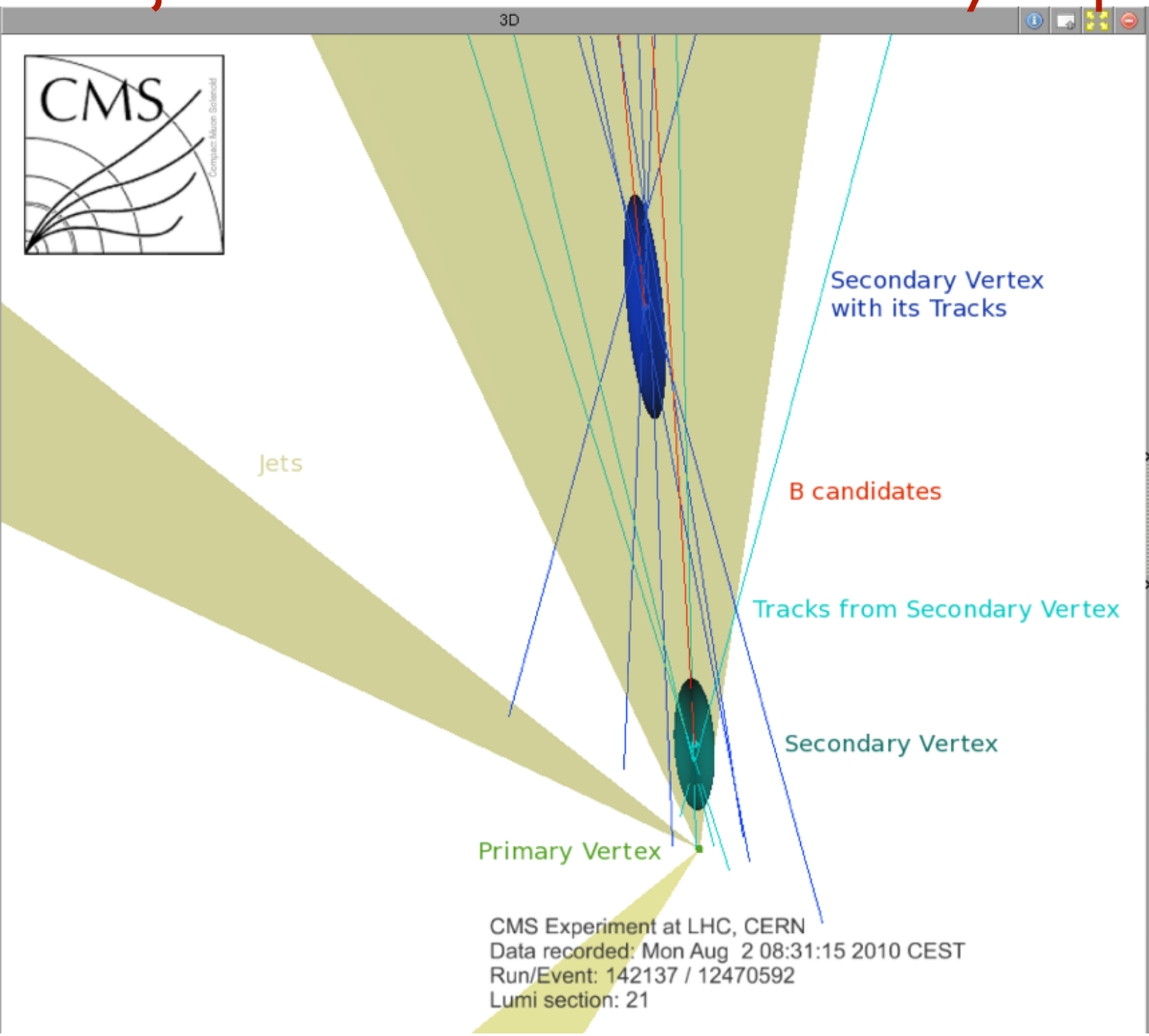


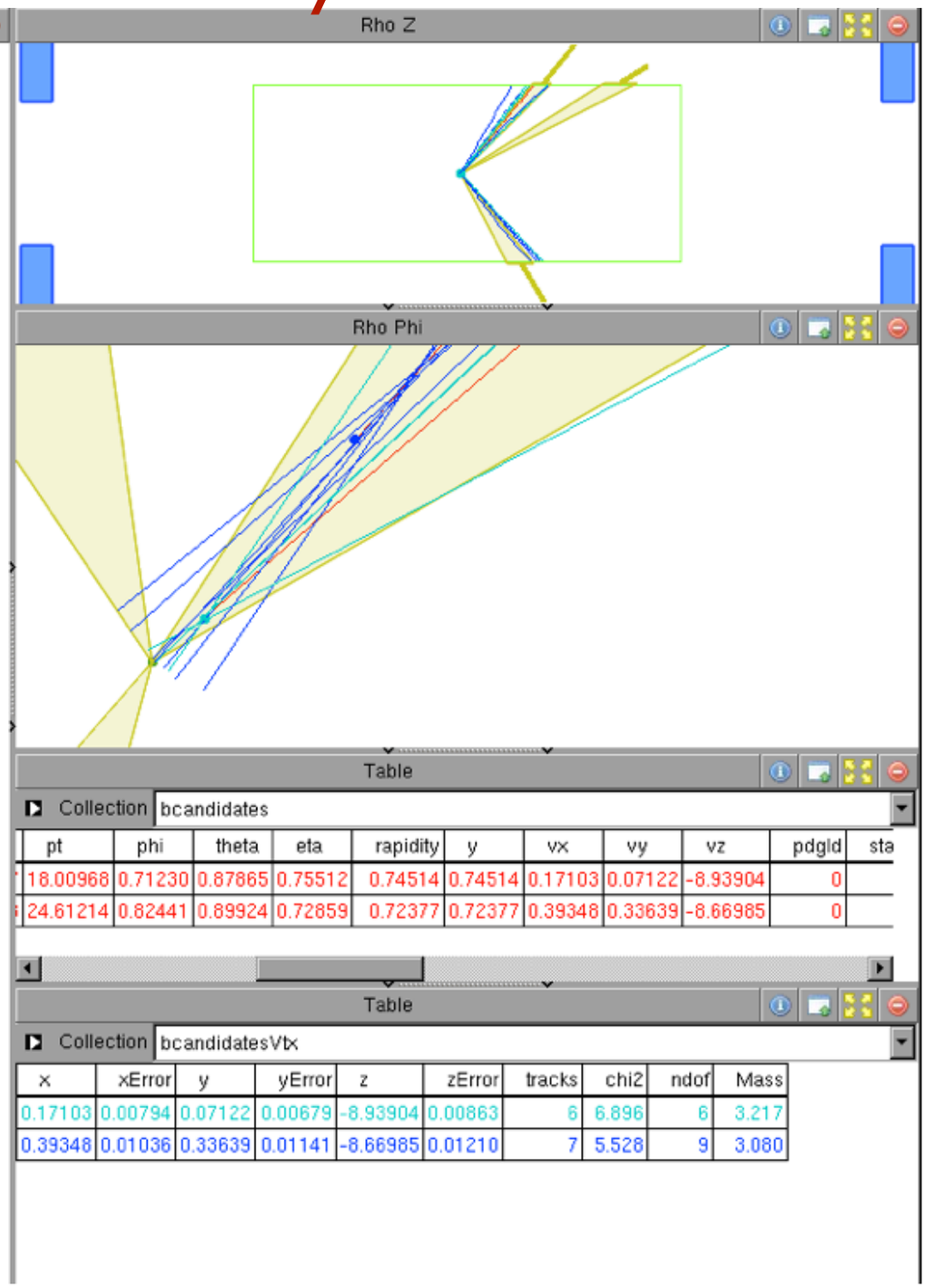
(more in backup)

another dimension

two collinear b-jets with two reconstructed collinear vertices:

➡jet substructure does not really help to identify this



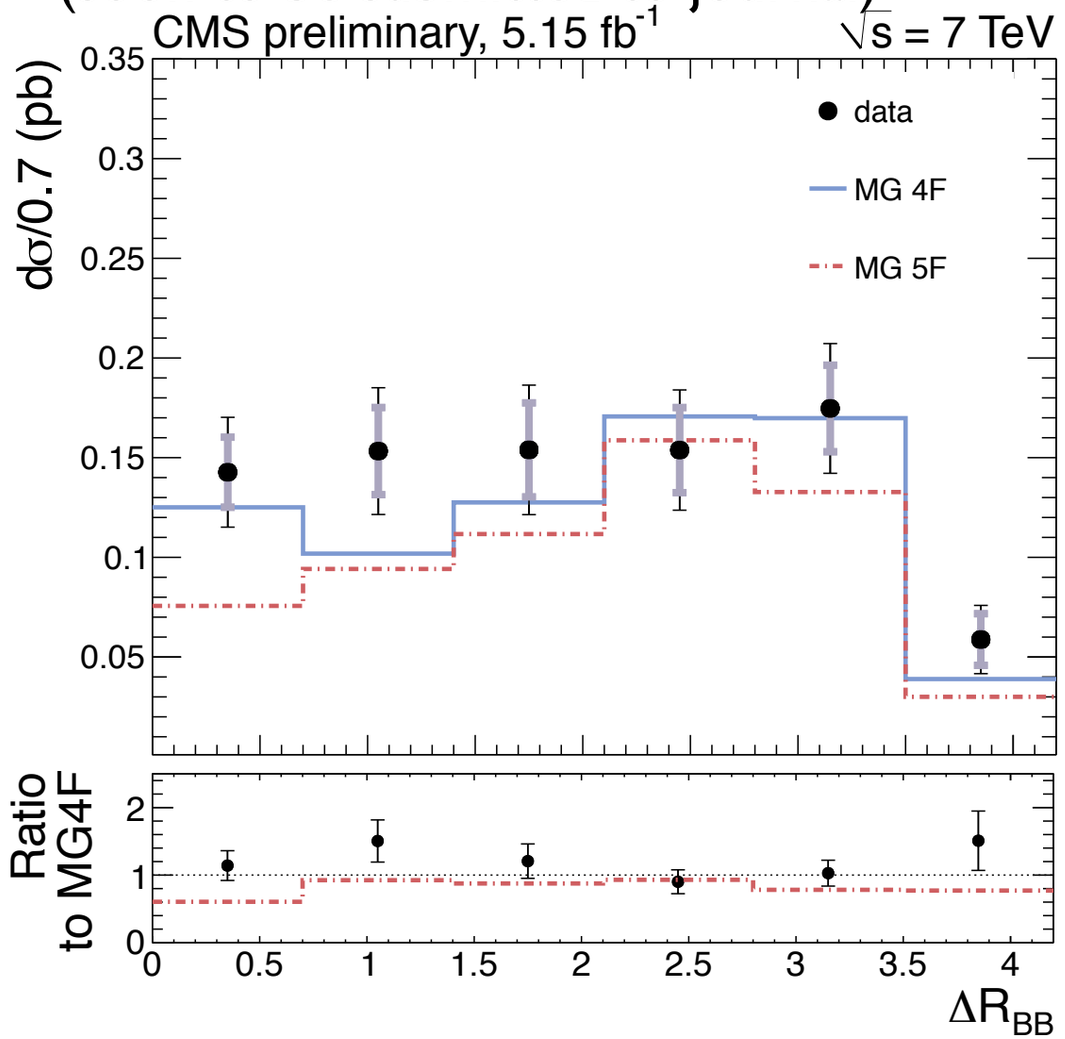


another dimension

- CMS pioneered an interesting new tool:
- inclusive vertex reconstruction:
  - ➡seeded from tracks with high impact parameter **(instead of jets)**
  - ➡no jets required
  - ➡angular resolution is amazing: $\sigma(\Delta R)=0.02$
  - ➡**used to probe b-(jet) production at small angles**

applied in analyses:

- $b\bar{b}$ angular correlations [JHEP 03 (2011) 136]
- $b\bar{b}$ angular correlations in $Z+b\bar{b}$ events [CMS PAS EWK-11-015] (soon to be submitted to journal)



Summary / Outlook

- substructure techniques widely used in CMS
- reasonable description by simulation

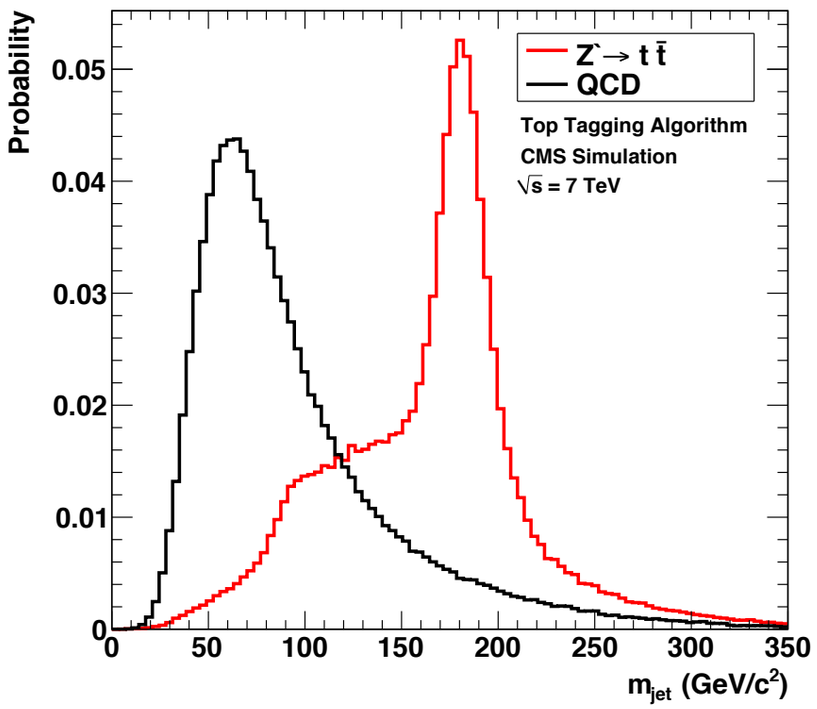
- applied not only in typical $Z' \rightarrow t\bar{t}$ search

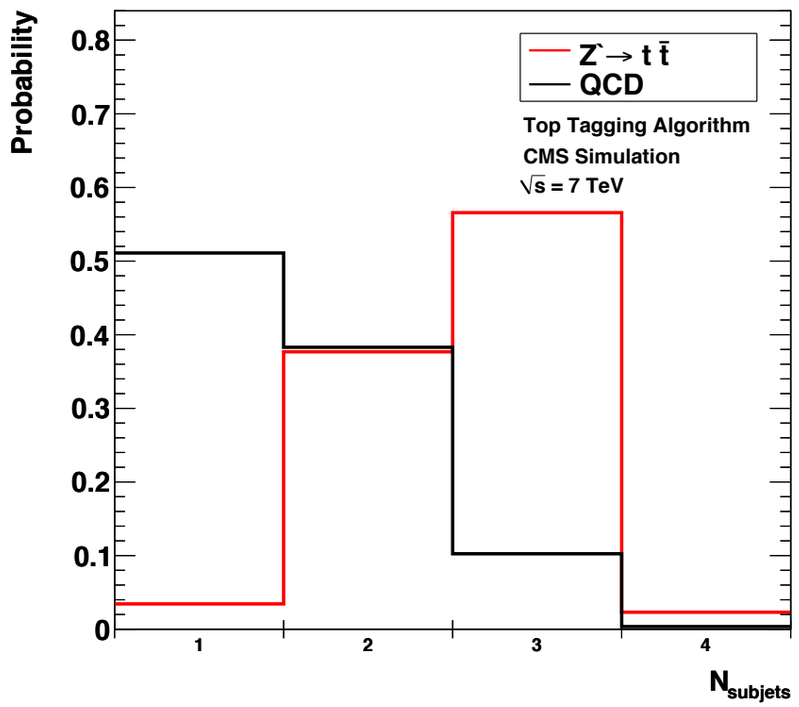
- lot's of ongoing activity right now

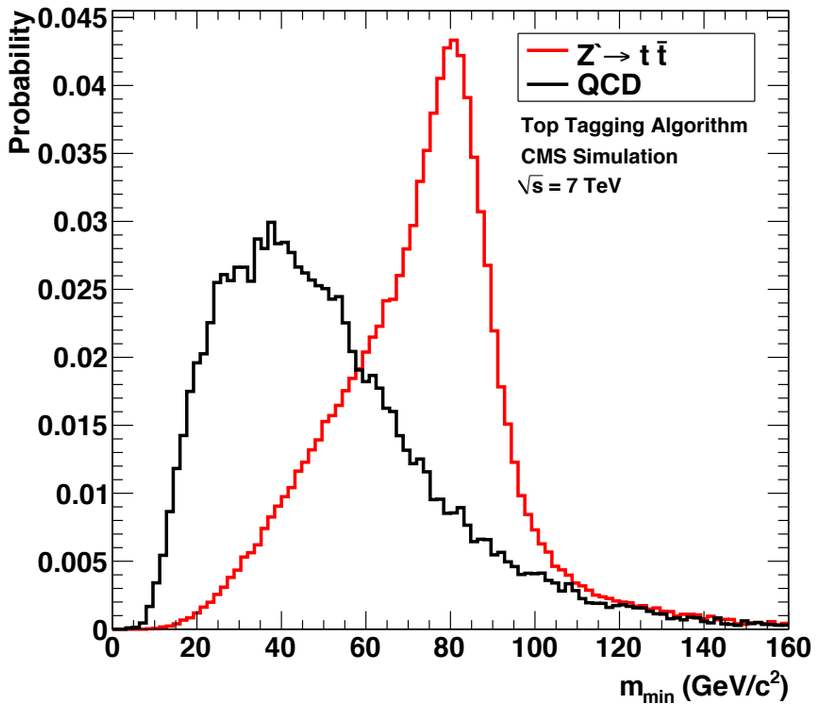
- expect major new results to be released for BOOST 2013 (e.g. b-tagging in subjets, more top-tagging algorithms)

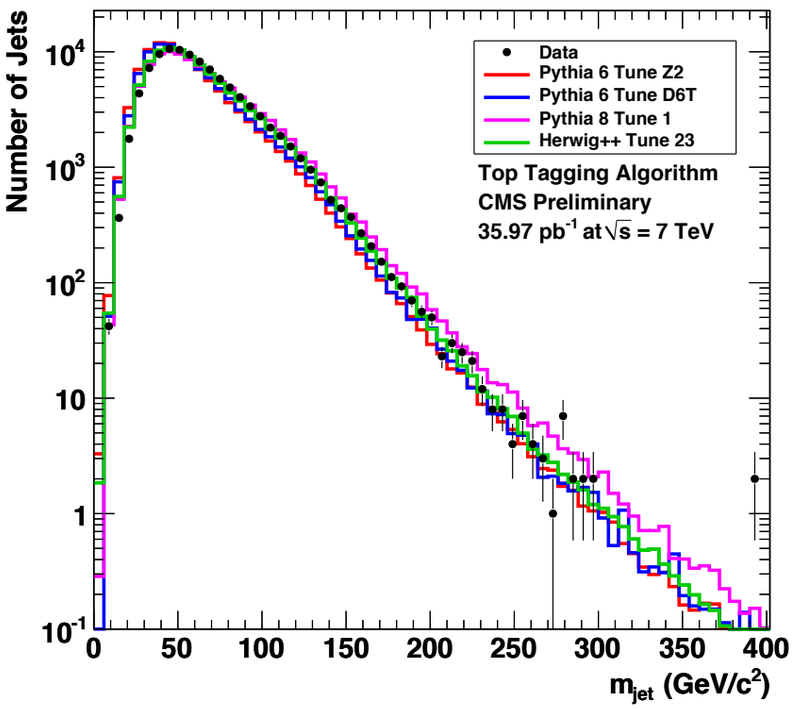
backup slides

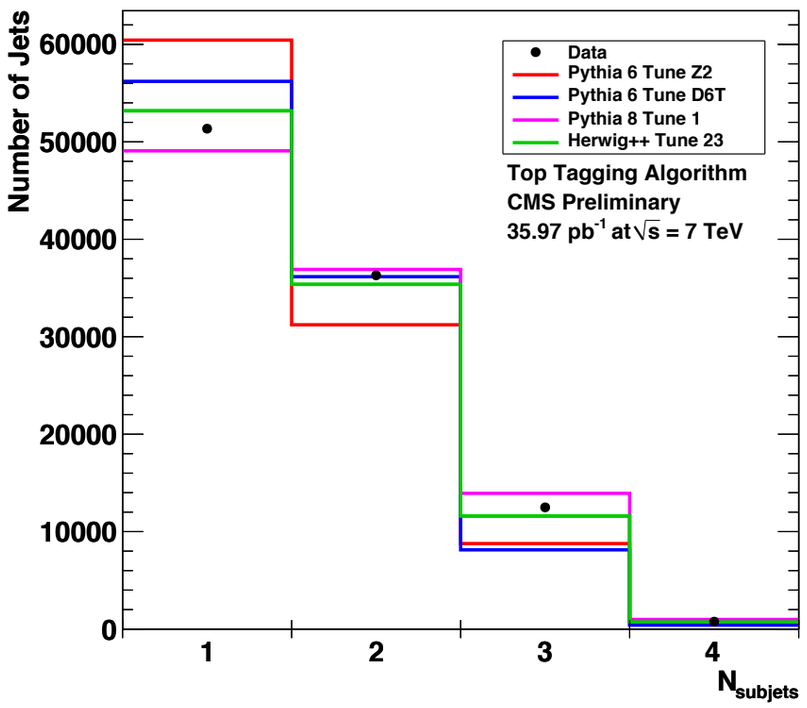
commissioning of top-tagging

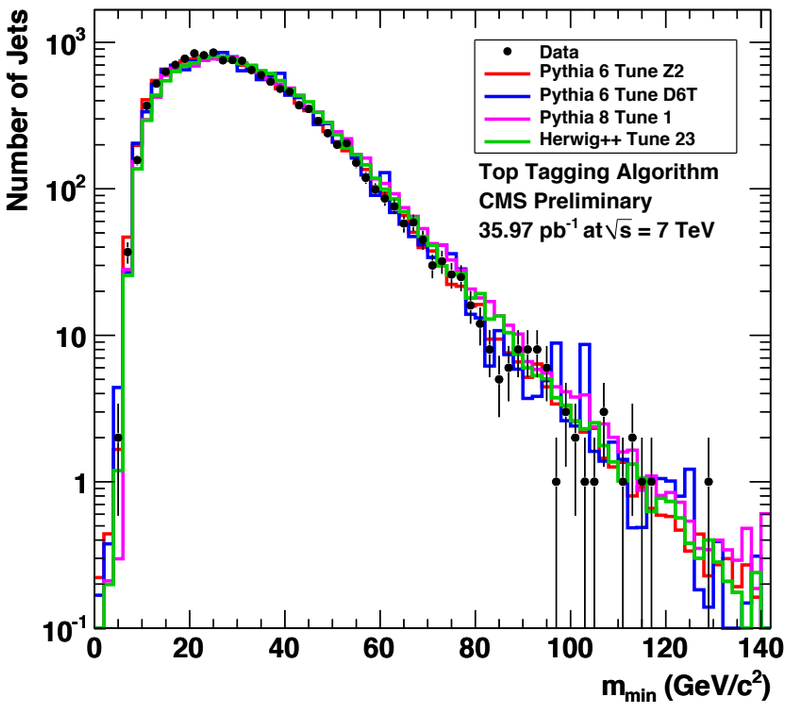




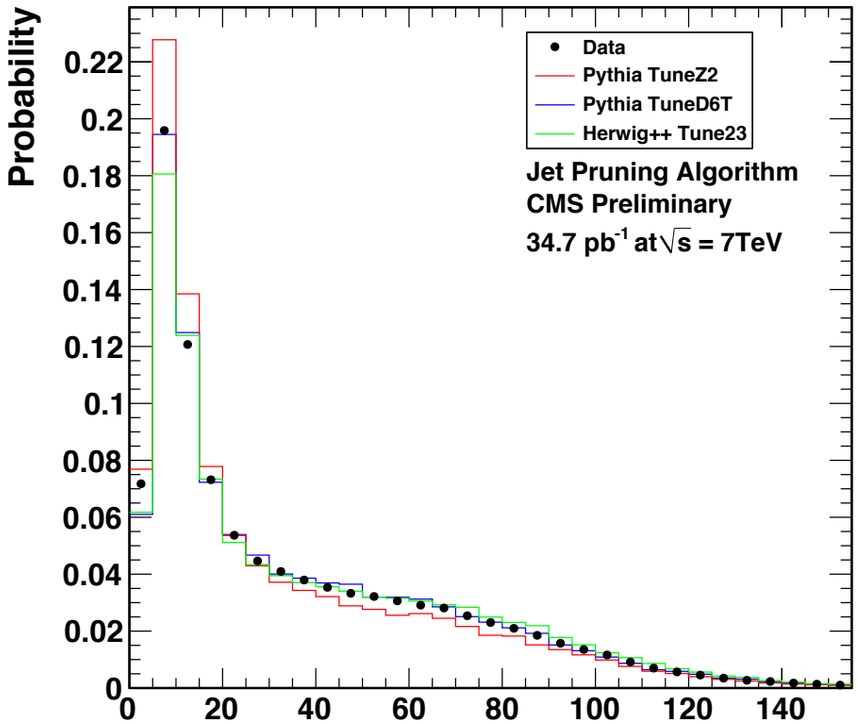


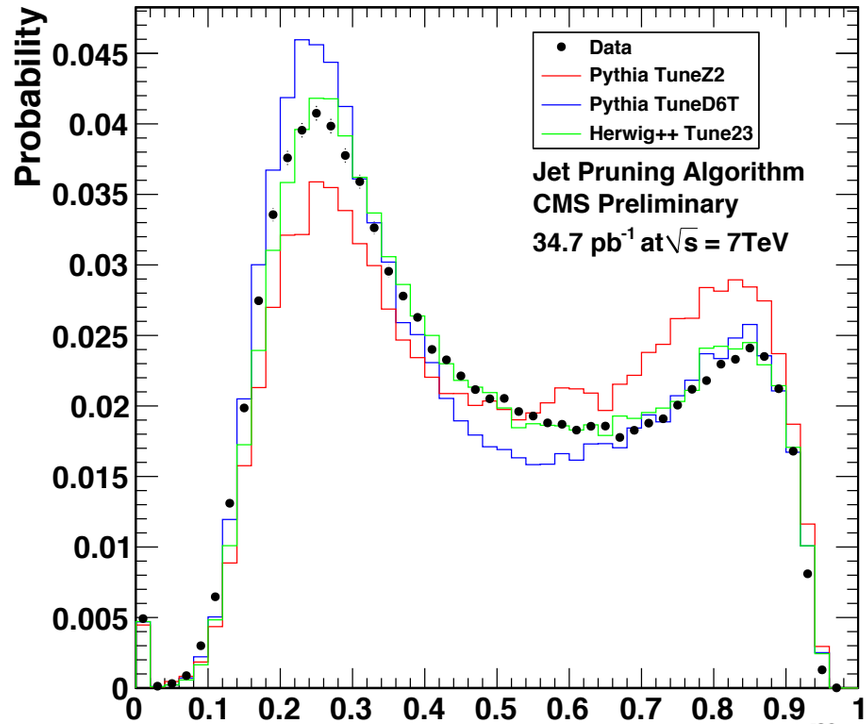


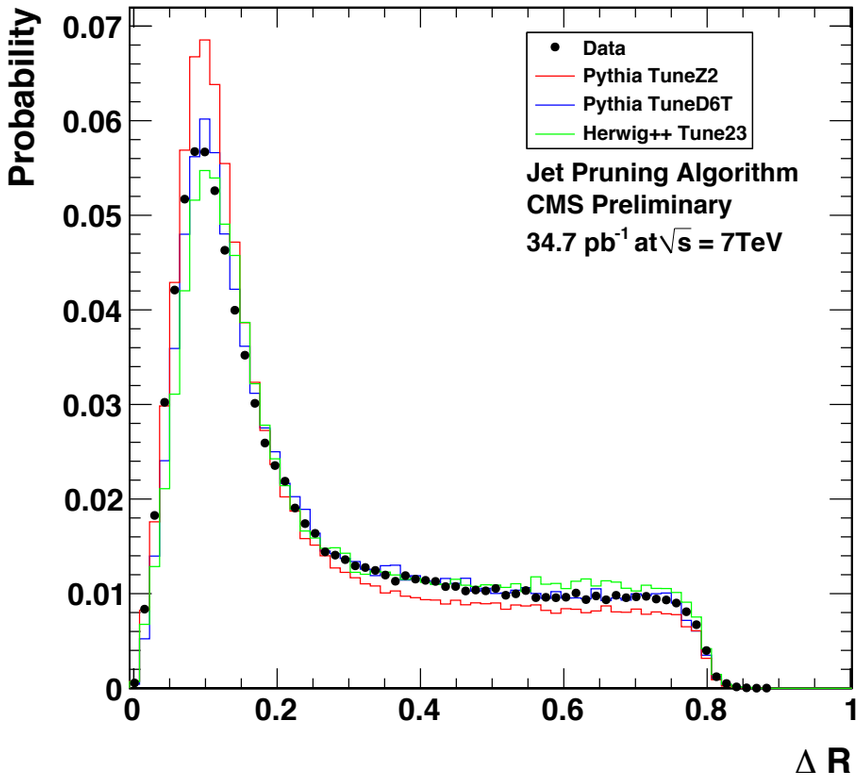


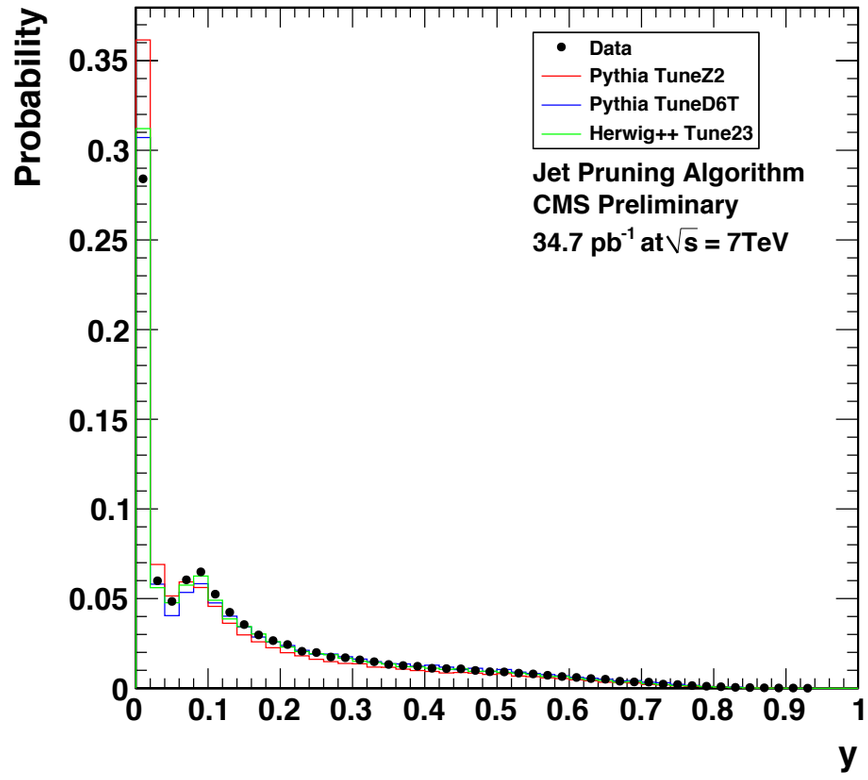


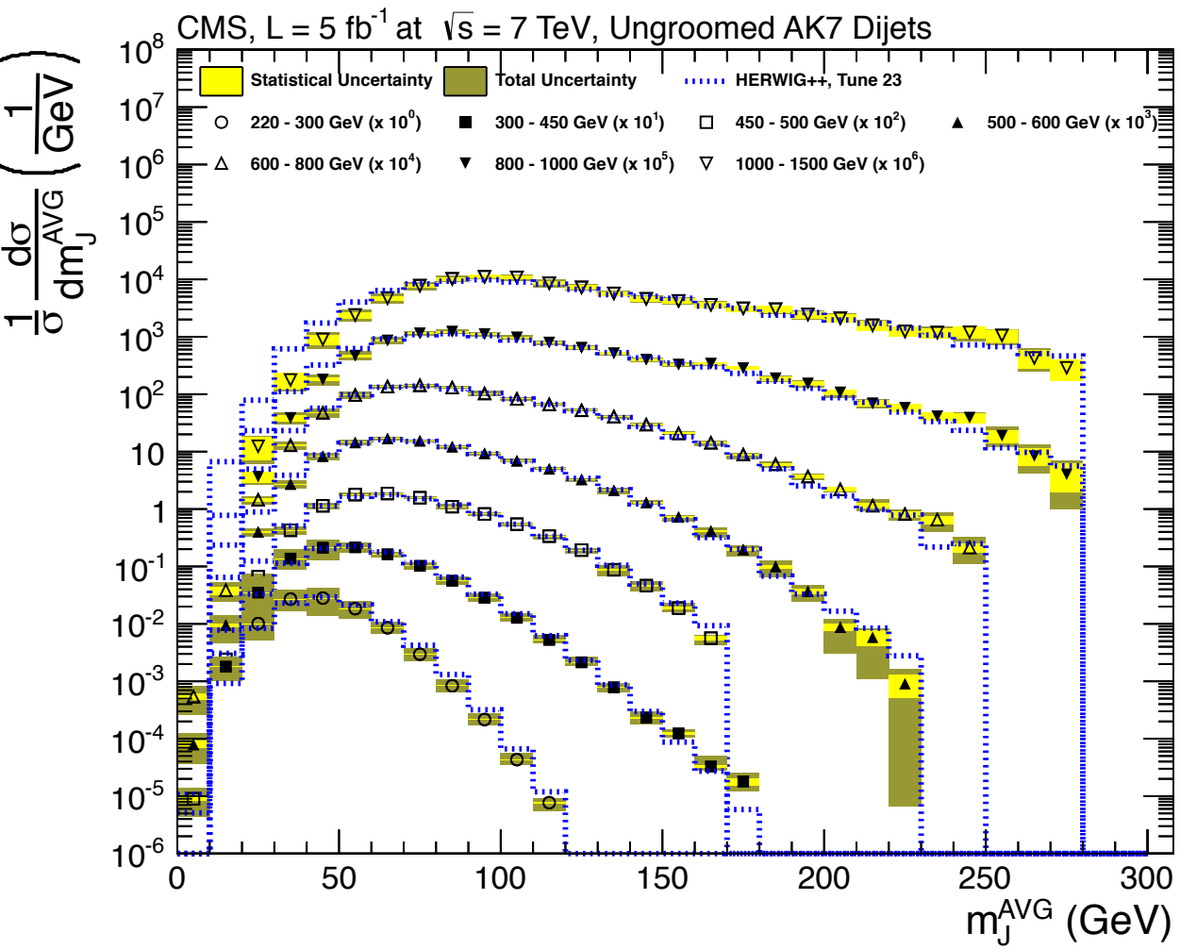
commissioning of W-tagging

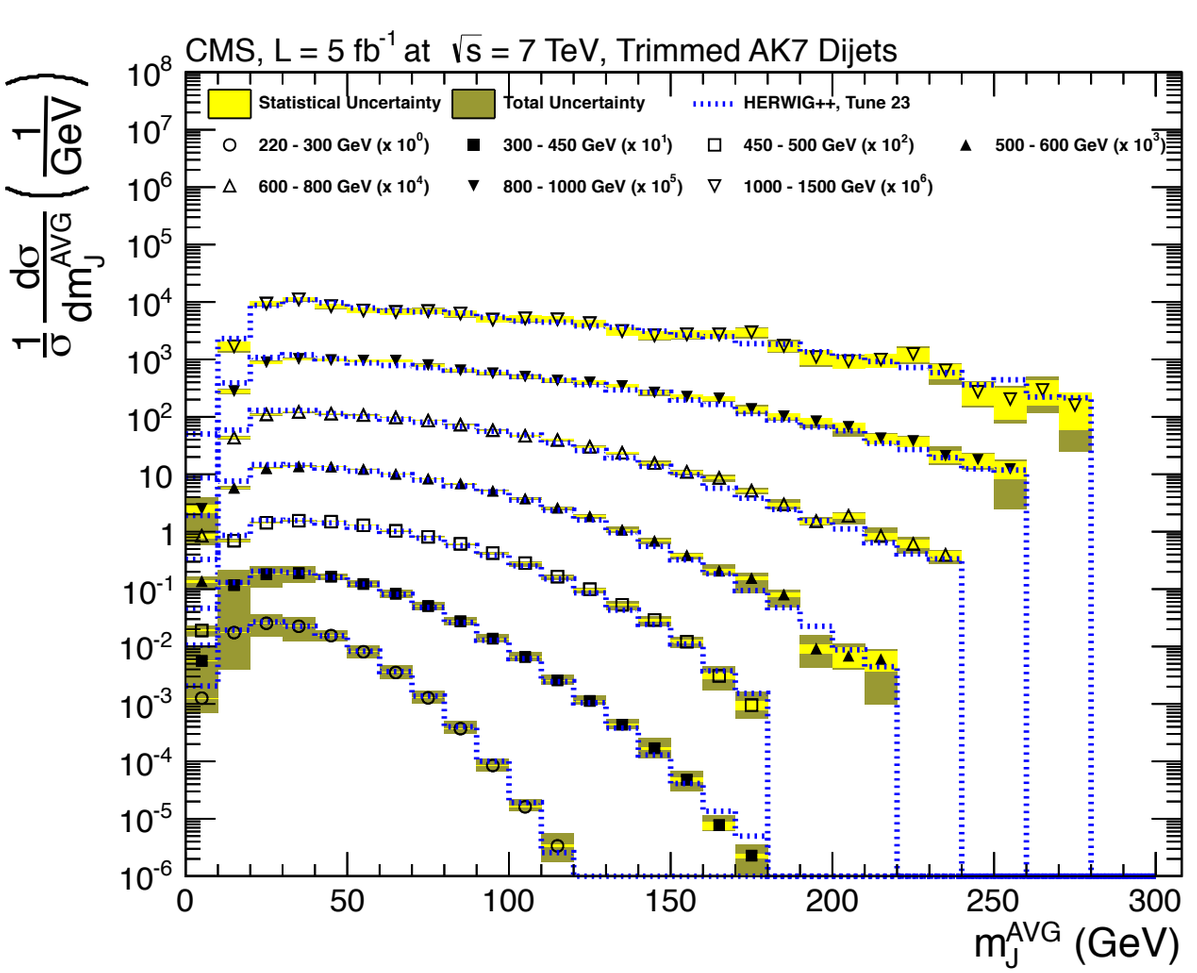


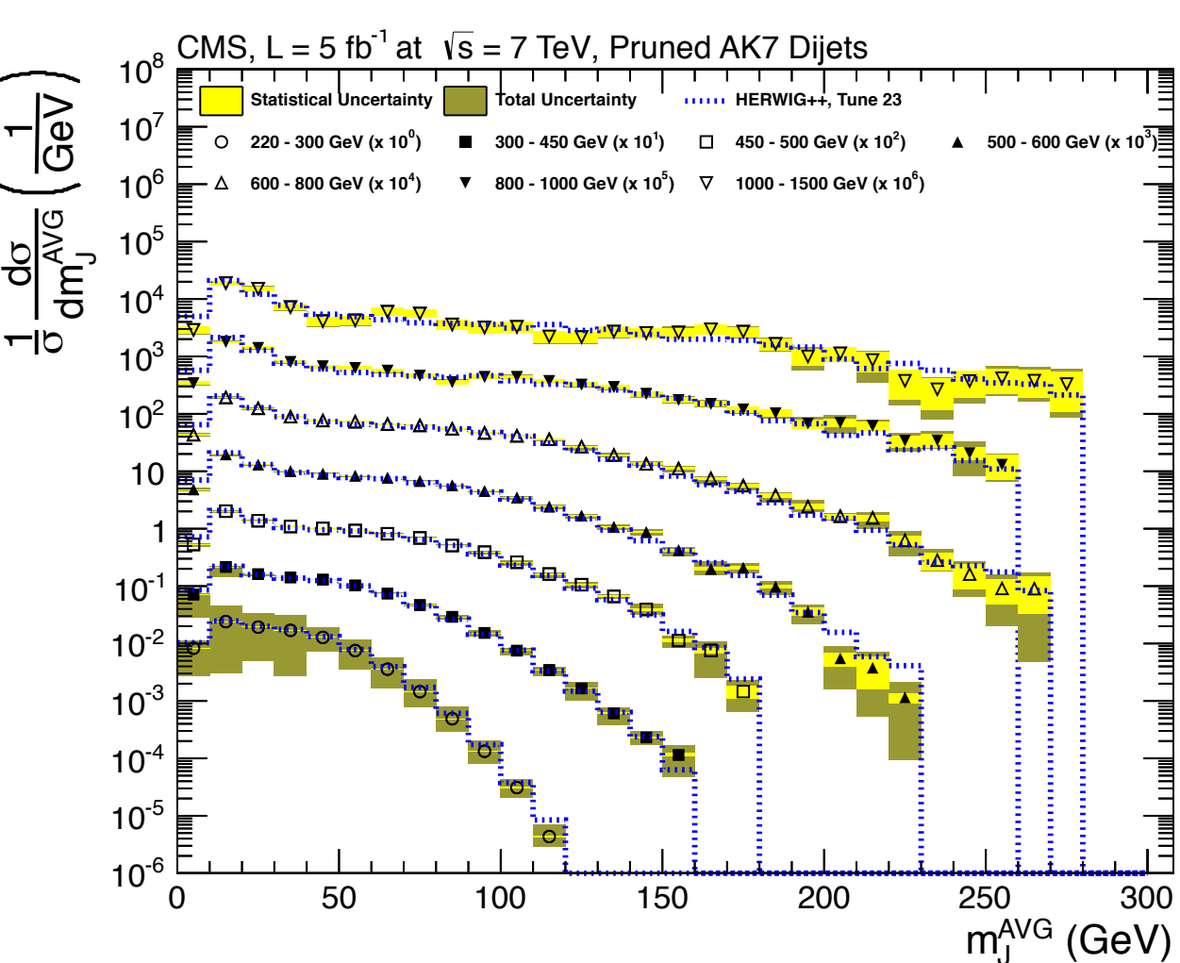


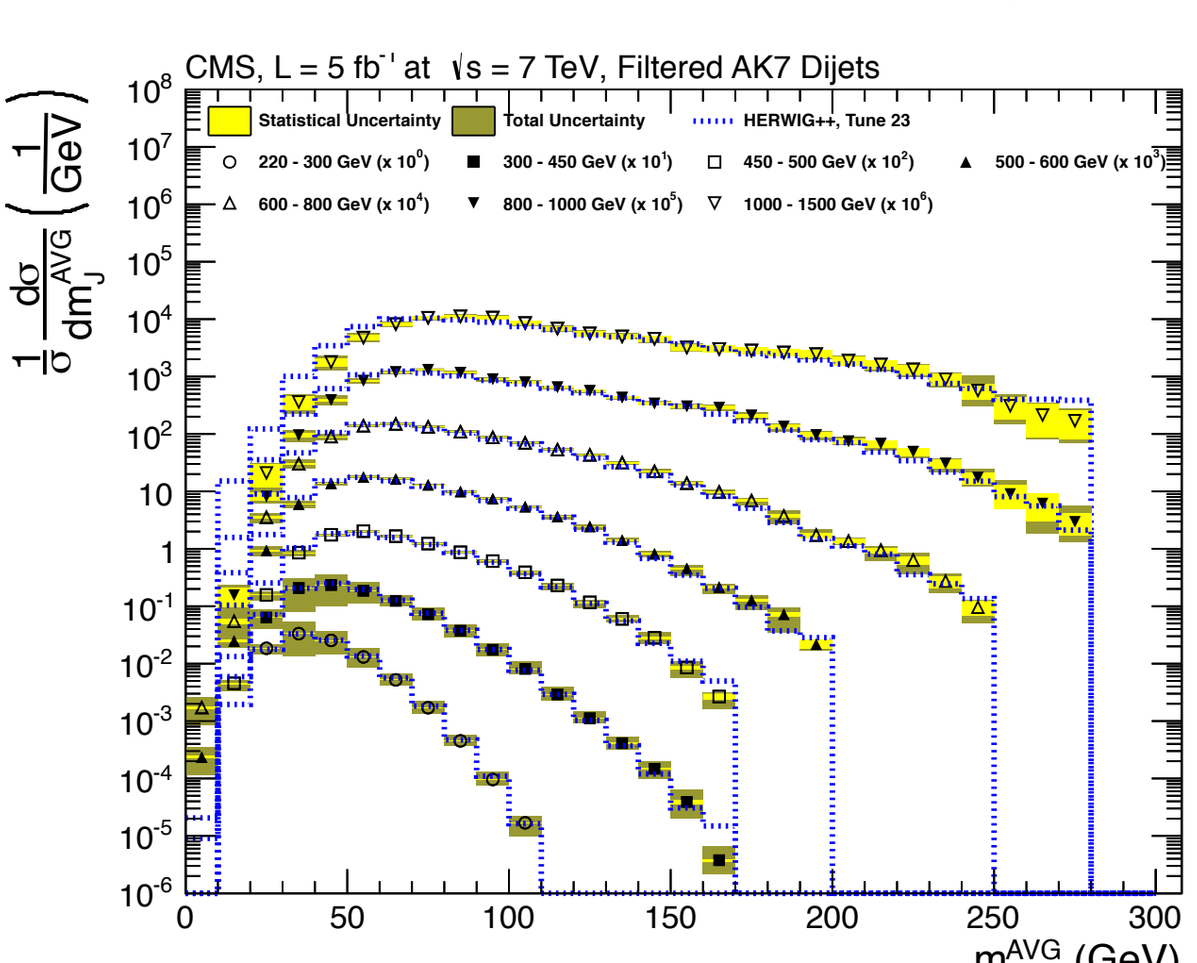


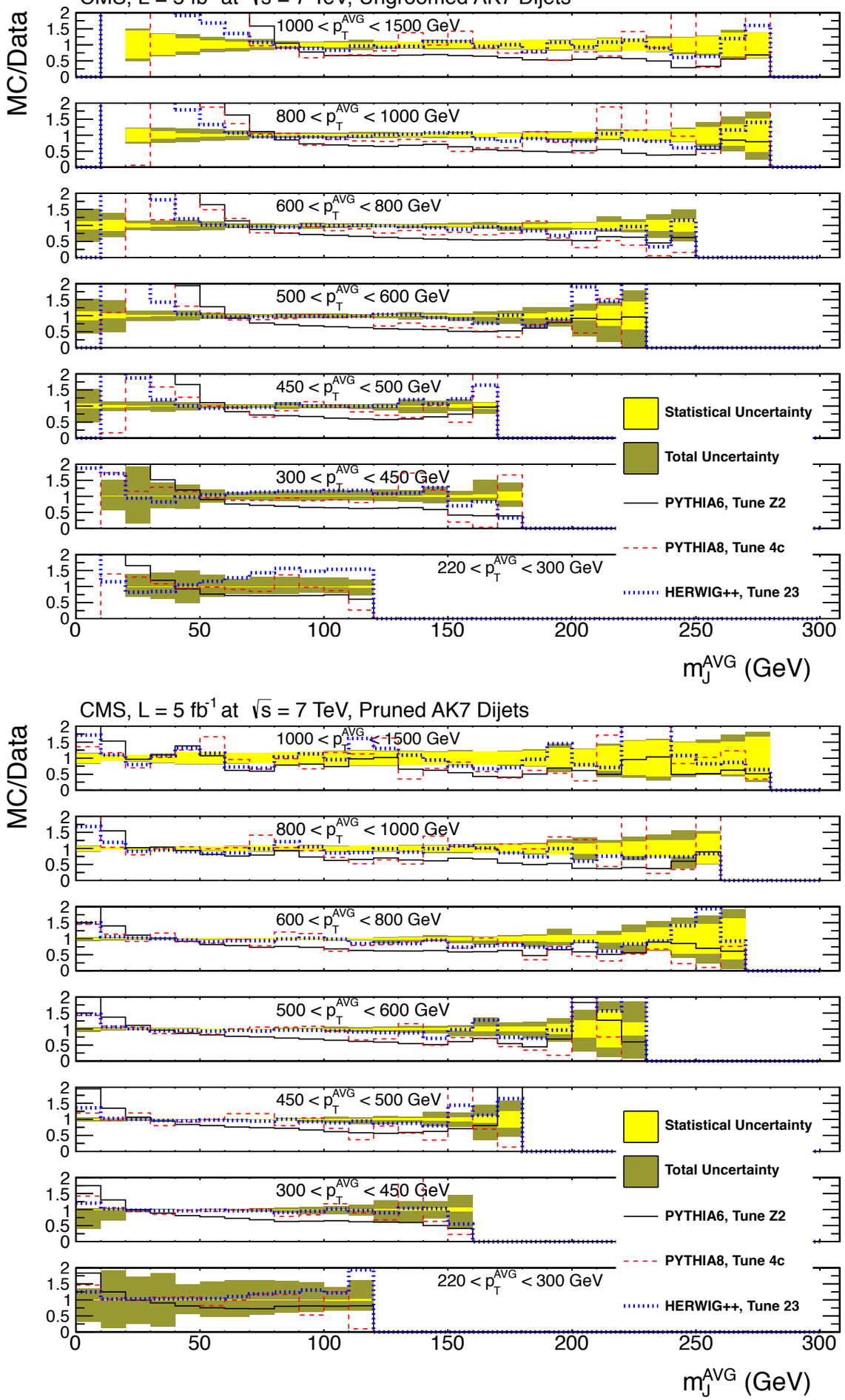


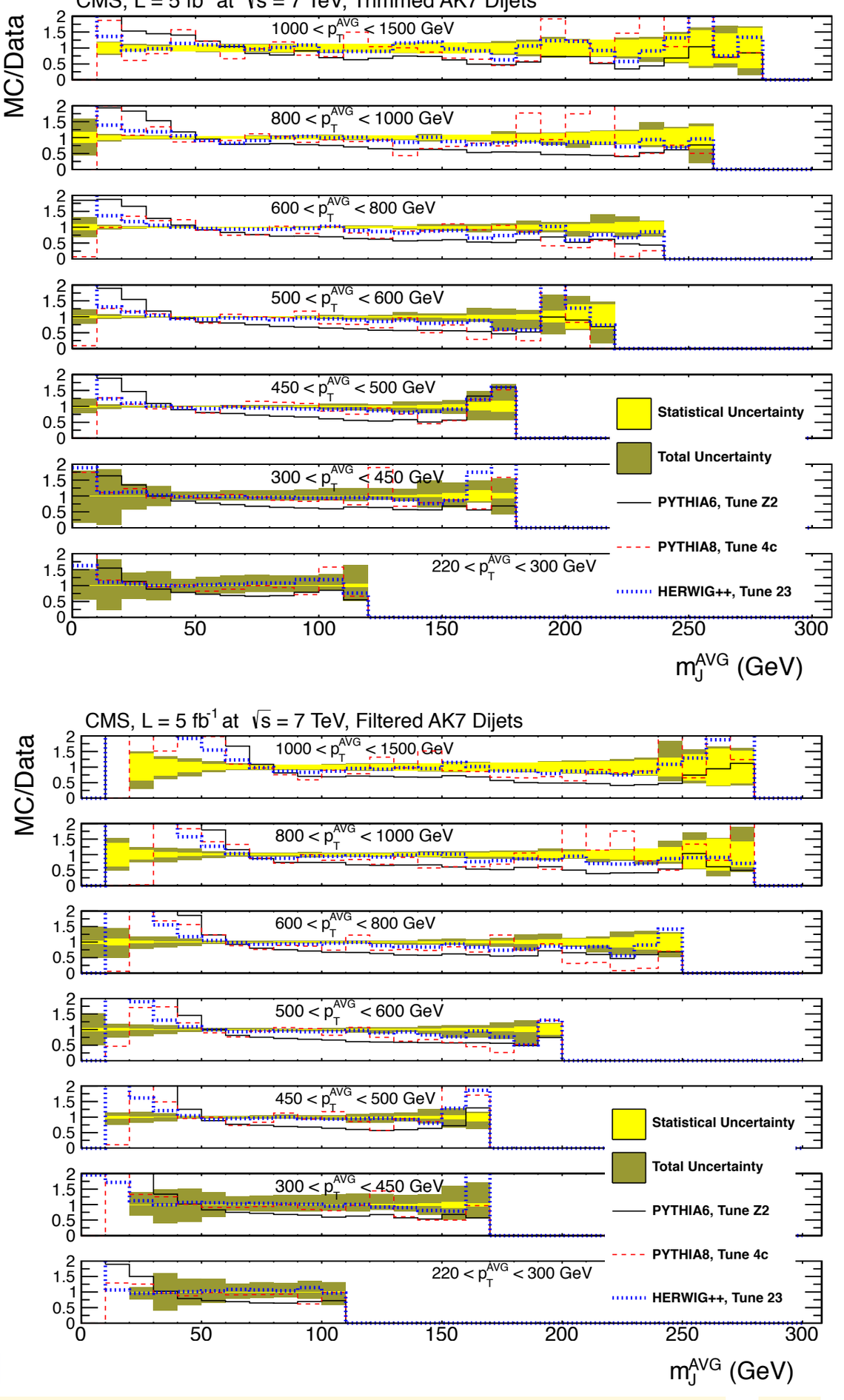

CMS, L = 5 fb-1 at √s = 7 TeV, Ungroomed AK7 Dijets
Statistical Uncertainty
Total Uncertainty
HERWIG++, Tune 23
220 - 300 GeV (x 10^0)
300 - 450 GeV (x 10^1)
450 - 500 GeV (x 10^2)
500 - 600 GeV (x 10^3)
600 - 800 GeV (x 10^4)
800 - 1000 GeV (x 10^5)
1000 - 1500 GeV (x 10^6)
1/σ dσ/dm_J^AVG (1/GeV)
m_J^AVG (GeV)


CMS, L = 5 fb-1 at √s = 7 TeV, Trimmed AK7 Dijets
Statistical Uncertainty
Total Uncertainty
HERWIG++, Tune 23
220 - 300 GeV (x 10^0)
300 - 450 GeV (x 10^1)
450 - 500 GeV (x 10^2)
500 - 600 GeV (x 10^3)
600 - 800 GeV (x 10^4)
800 - 1000 GeV (x 10^5)
1000 - 1500 GeV (x 10^6)
1/σ dσ/dm_J^AVG (1/GeV)
m_J^AVG (GeV)


CMS, L = 5 fb-1 at √s = 7 TeV, Pruned AK7 Dijets
Statistical Uncertainty
Total Uncertainty
HERWIG++, Tune 23
220 - 300 GeV (x 10^0)
300 - 450 GeV (x 10^1)
450 - 500 GeV (x 10^2)
500 - 600 GeV (x 10^3)
600 - 800 GeV (x 10^4)
800 - 1000 GeV (x 10^5)
1000 - 1500 GeV (x 10^6)
1/σ dσ/dm_J^AVG (1/GeV)
m_J^AVG (GeV)


CMS, L = 5 fb-1 at √s = 7 TeV, Filtered AK7 Dijets
Statistical Uncertainty
Total Uncertainty
HERWIG++, Tune 23
220 - 300 GeV (x 10^0)
300 - 450 GeV (x 10^1)
450 - 500 GeV (x 10^2)
500 - 600 GeV (x 10^3)
600 - 800 GeV (x 10^4)
800 - 1000 GeV (x 10^5)
1000 - 1500 GeV (x 10^6)
1/σ dσ/dm_J^AVG (1/GeV)
m_J^AVG (GeV)

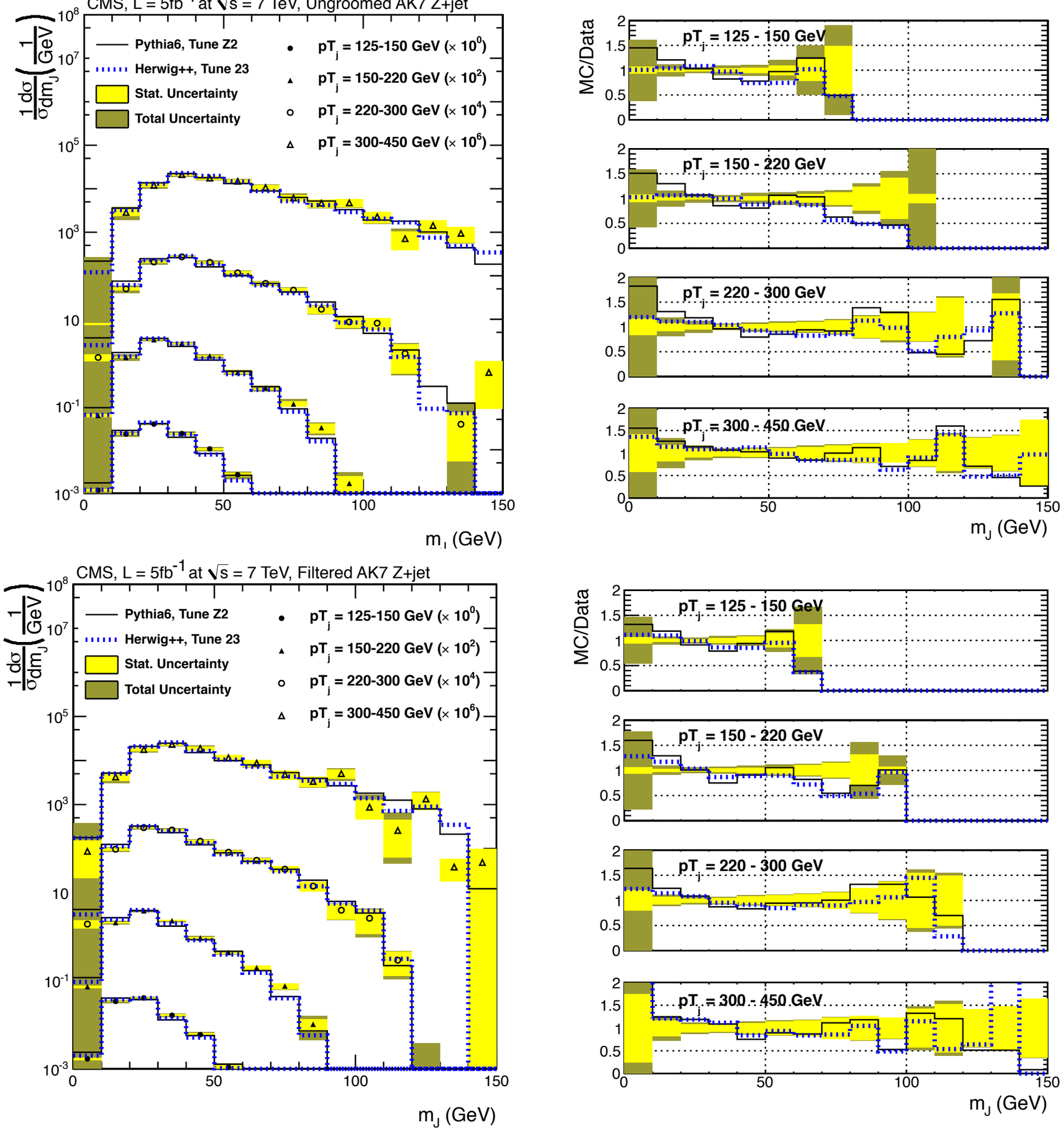

CMS, L = 5 fb-1 at √s = 7 TeV, Ungroomed AK7 Dijets
MC/Data
1000 < pTAVG < 1500 GeV
800 < pTAVG < 1000 GeV
600 < pTAVG < 800 GeV
500 < pTAVG < 600 GeV
450 < pTAVG < 500 GeV
300 < pTAVG < 450 GeV
220 < pTAVG < 300 GeV
Statistical Uncertainty
Total Uncertainty
PYTHIA6, Tune Z2
PYTHIA8, Tune 4c
HERWIG++, Tune 23
mJAVG (GeV)


CMS, L = 5 fb-1 at √s = 7 TeV, Trimmed AK7 Dijets
MC/Data
1000 < pTAVG < 1500 GeV
800 < pTAVG < 1000 GeV
600 < pTAVG < 800 GeV
500 < pTAVG < 600 GeV
450 < pTAVG < 500 GeV
300 < pTAVG < 450 GeV
220 < pTAVG < 300 GeV
Statistical Uncertainty
Total Uncertainty
PYTHIA6, Tune Z2
PYTHIA8, Tune 4c
HERWIG++, Tune 23
mJAVG (GeV)

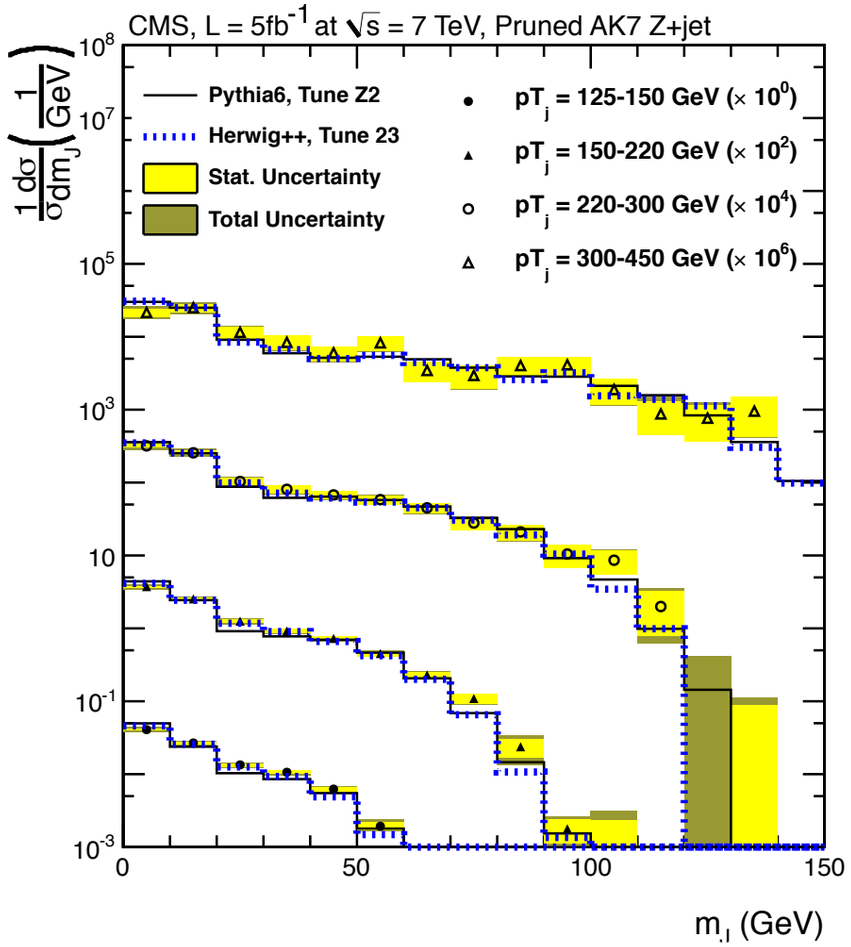
CMS, L = 5 fb-1 at √s = 7 TeV, Pruned AK7 Dijets
MC/Data
1000 < pTAVG < 1500 GeV
800 < pTAVG < 1000 GeV
600 < pTAVG < 800 GeV
500 < pTAVG < 600 GeV
450 < pTAVG < 500 GeV
300 < pTAVG < 450 GeV
220 < pTAVG < 300 GeV
Statistical Uncertainty
Total Uncertainty
PYTHIA6, Tune Z2
PYTHIA8, Tune 4c
HERWIG++, Tune 23
mJAVG (GeV)

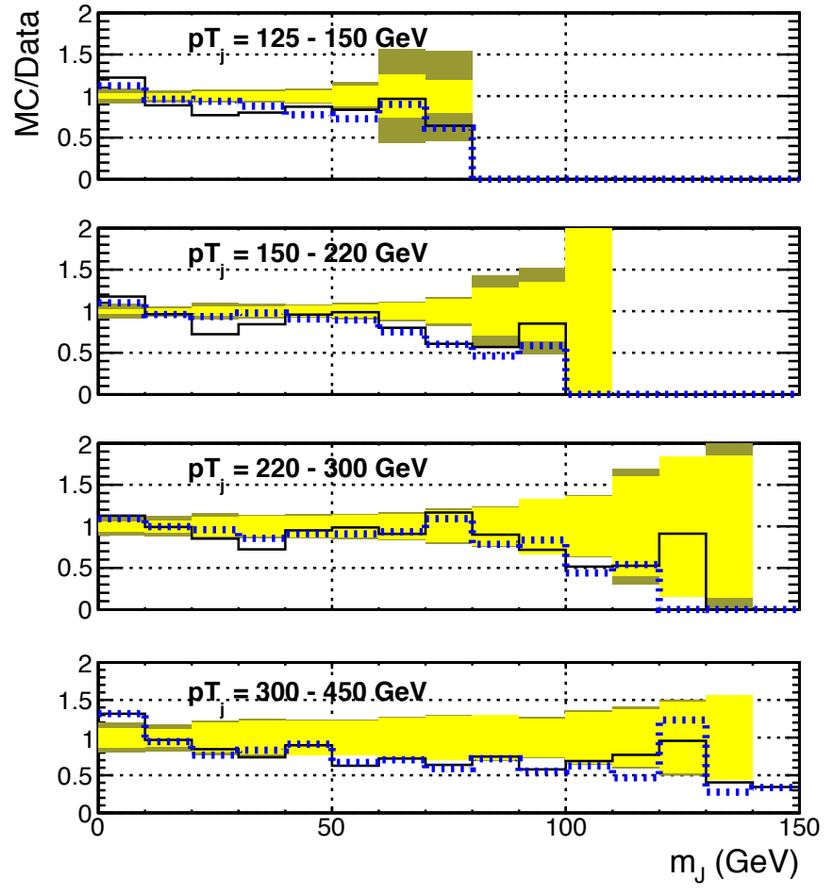
CMS, L = 5 fb-1 at √s = 7 TeV, Filtered AK7 Dijets
MC/Data
1000 < pTAVG < 1500 GeV
800 < pTAVG < 1000 GeV
600 < pTAVG < 800 GeV
500 < pTAVG < 600 GeV
450 < pTAVG < 500 GeV
300 < pTAVG < 450 GeV
220 < pTAVG < 300 GeV
Statistical Uncertainty
Total Uncertainty
PYTHIA6, Tune Z2
PYTHIA8, Tune 4c
HERWIG++, Tune 23
mJAVG (GeV)

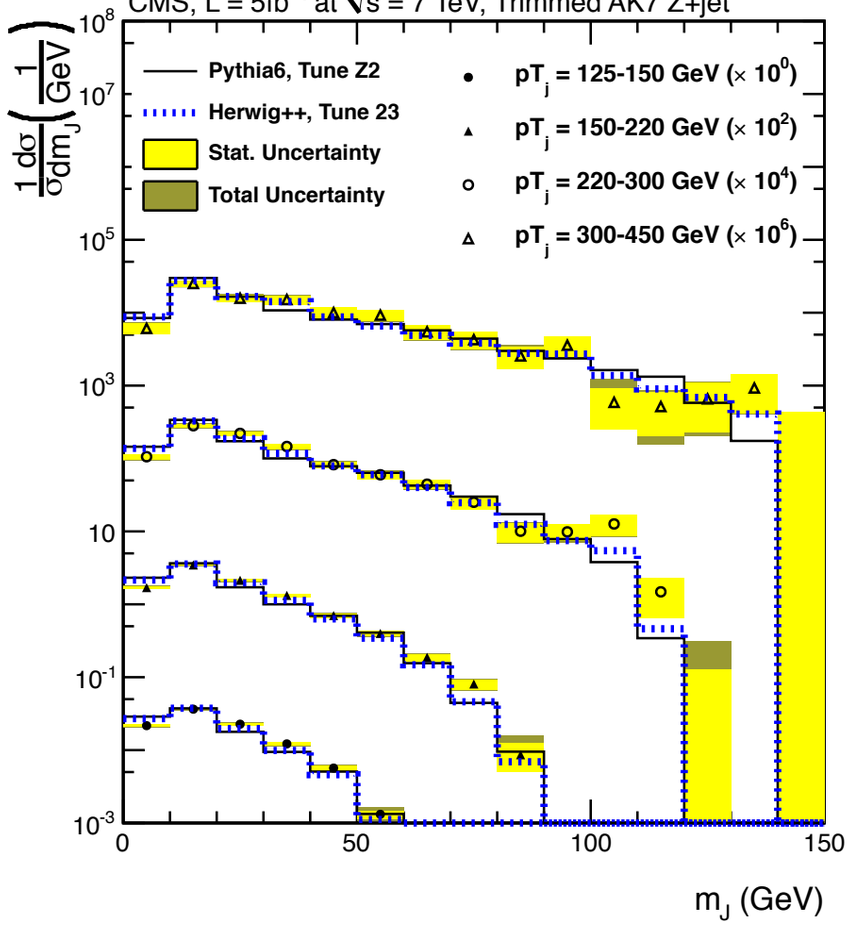
JHEP 201

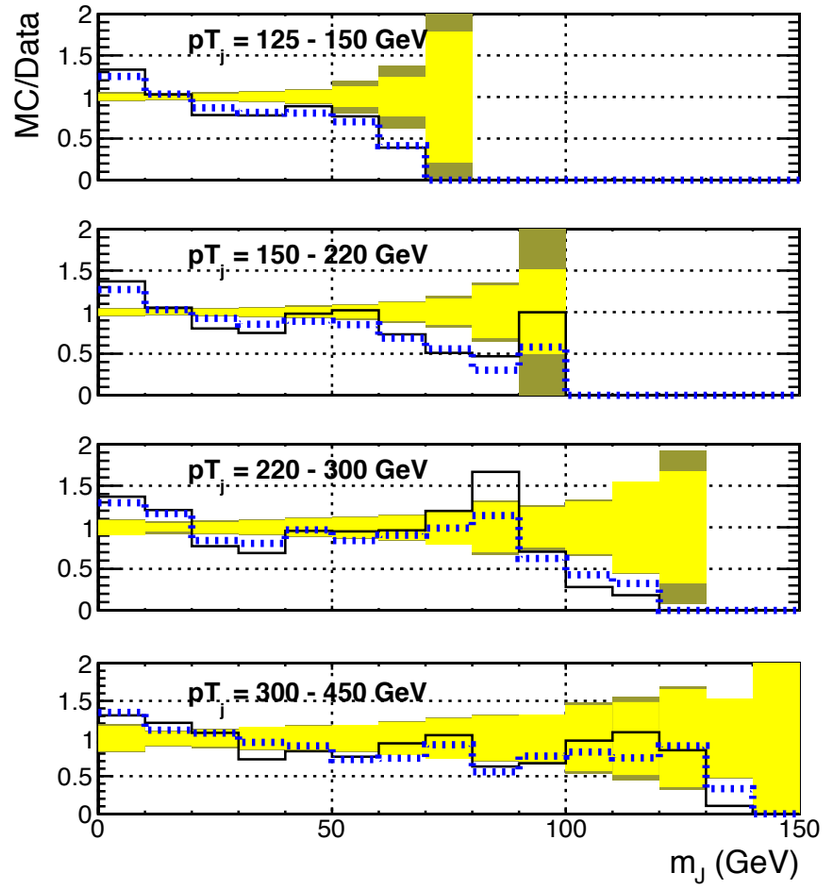

CMS, L = 5fb^{-1} at $\sqrt{s}$ = 7 TeV, Ungroomed AK7 Z+jet
$\frac{1}{\sigma}\frac{d\sigma}{dm_J}\left(\frac{1}{GeV}\right)$
Pythia6, Tune Z2
Herwig++, Tune 23
Stat. Uncertainty
Total Uncertainty
pT$_j$ = 125-150 GeV (× 10^0)
pT$_j$ = 150-220 GeV (× 10^2)
pT$_j$ = 220-300 GeV (× 10^4)
pT$_j$ = 300-450 GeV (× 10^6)
m$_J$ (GeV)
MC/Data
pT$_j$ = 125 - 150 GeV
pT$_j$ = 150 - 220 GeV
pT$_j$ = 220 - 300 GeV
pT$_j$ = 300 - 450 GeV
m$_J$ (GeV)

CMS, L = 5fb^{-1} at $\sqrt{s}$ = 7 TeV, Filtered AK7 Z+jet
$\frac{1}{\sigma}\frac{d\sigma}{dm_J}\left(\frac{1}{GeV}\right)$
Pythia6, Tune Z2
Herwig++, Tune 23
Stat. Uncertainty
Total Uncertainty
pT$_j$ = 125-150 GeV (× 10^0)
pT$_j$ = 150-220 GeV (× 10^2)
pT$_j$ = 220-300 GeV (× 10^4)
pT$_j$ = 300-450 GeV (× 10^6)
m$_J$ (GeV)
MC/Data
pT$_j$ = 125 - 150 GeV
pT$_j$ = 150 - 220 GeV
pT$_j$ = 220 - 300 GeV
pT$_j$ = 300 - 450 GeV
m$_J$ (GeV)


CMS, L = 5fb^{-1} at $\sqrt{s}$ = 7 TeV, Pruned AK7 Z+jet
$\frac{1}{\sigma}\frac{d\sigma}{dm_J}\left(\frac{1}{GeV}\right)$
Pythia6, Tune Z2
Herwig++, Tune 23
Stat. Uncertainty
Total Uncertainty
pT$_j$ = 125-150 GeV (× 10^0)
pT$_j$ = 150-220 GeV (× 10^2)
pT$_j$ = 220-300 GeV (× 10^4)
pT$_j$ = 300-450 GeV (× 10^6)
m$_J$ (GeV)


MC/Data
pT$_j$ = 125 - 150 GeV
pT$_j$ = 150 - 220 GeV
pT$_j$ = 220 - 300 GeV
pT$_j$ = 300 - 450 GeV
m$_J$ (GeV)


CMS, L = 5fb^{-1} at $\sqrt{s}$ = 7 TeV, Trimmed AK7 Z+jet
$\frac{1}{\sigma}\frac{d\sigma}{dm_J}\left(\frac{1}{GeV}\right)$
Pythia6, Tune Z2
Herwig++, Tune 23
Stat. Uncertainty
Total Uncertainty
pT$_j$ = 125-150 GeV (× 10^0)
pT$_j$ = 150-220 GeV (× 10^2)
pT$_j$ = 220-300 GeV (× 10^4)
pT$_j$ = 300-450 GeV (× 10^6)
m$_J$ (GeV)


MC/Data
pT$_j$ = 125 - 150 GeV
pT$_j$ = 150 - 220 GeV
pT$_j$ = 220 - 300 GeV
pT$_j$ = 300 - 450 GeV
m$_J$ (GeV)