a case study of using substructure analysis in new physics searches

Nhan Tran
for Kalanand Mishra
Fermi National Accelerator Laboratory

QCD tools for LHC Physics
November 14, 2013

See talk by Jeremy Love general discussion of substructure tools being utilized in analysis

This talk:

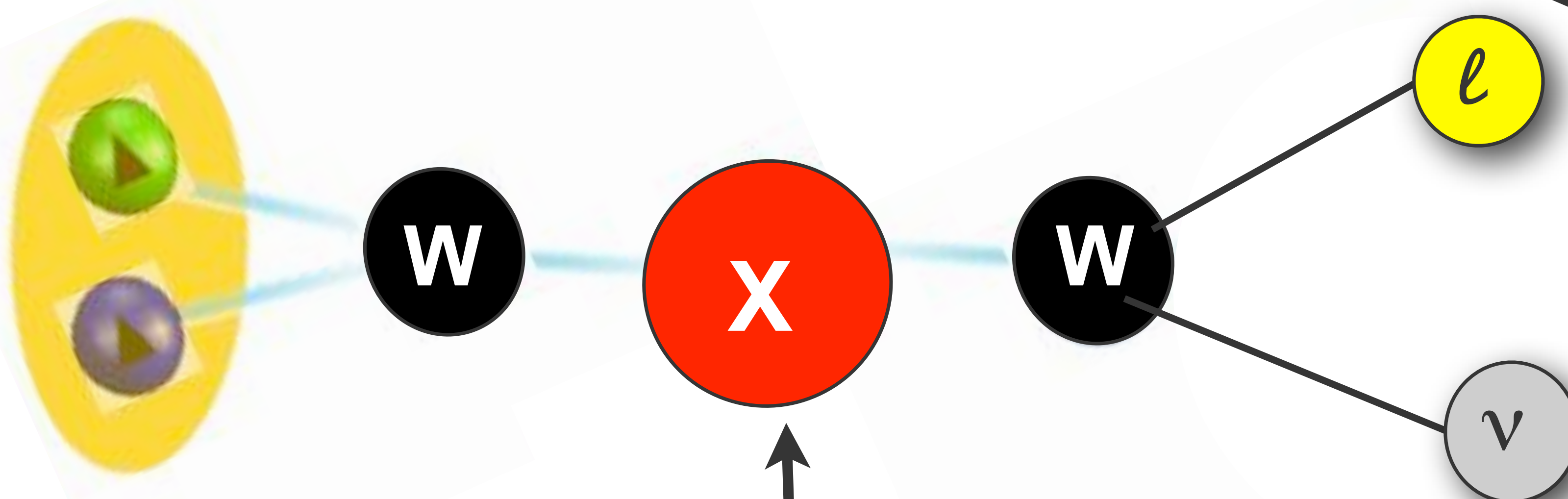
Take a specific example and discuss analysis issues related to jet substructure

- How can we ensure we understand our tagging efficiencies?
- What methods do we use to estimate our backgrounds?

Search for high mass W+V resonance in semi-leptonic final state

CMS PAS-HIG-13-008: heavy Higgs interpretation, 600-1000 GeV
CMS PAS-EXO-12-021: massive graviton interpretation, 0.8-2.5 TeV

For W p_T > 200 GeV, generally a merged jet.

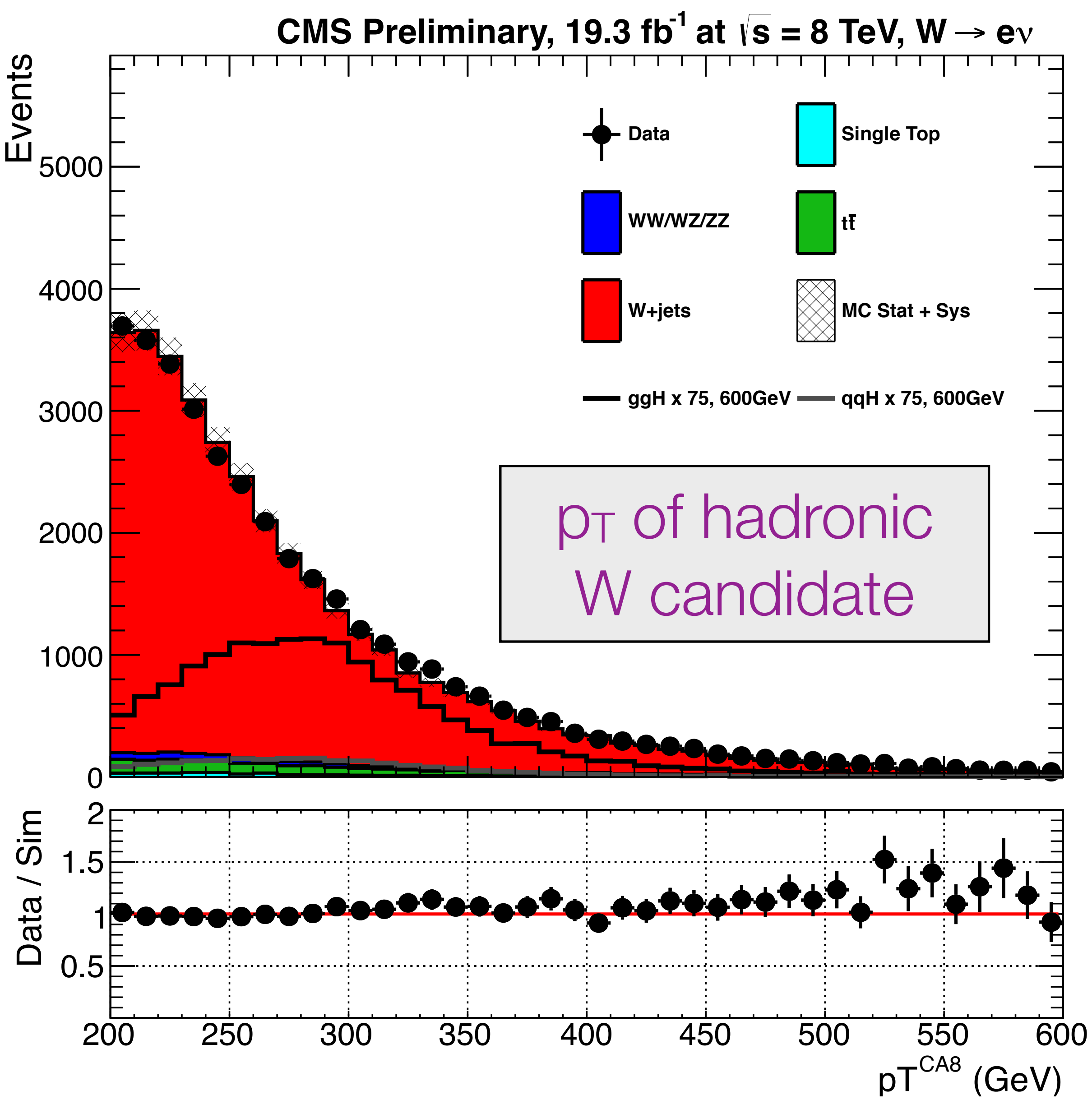


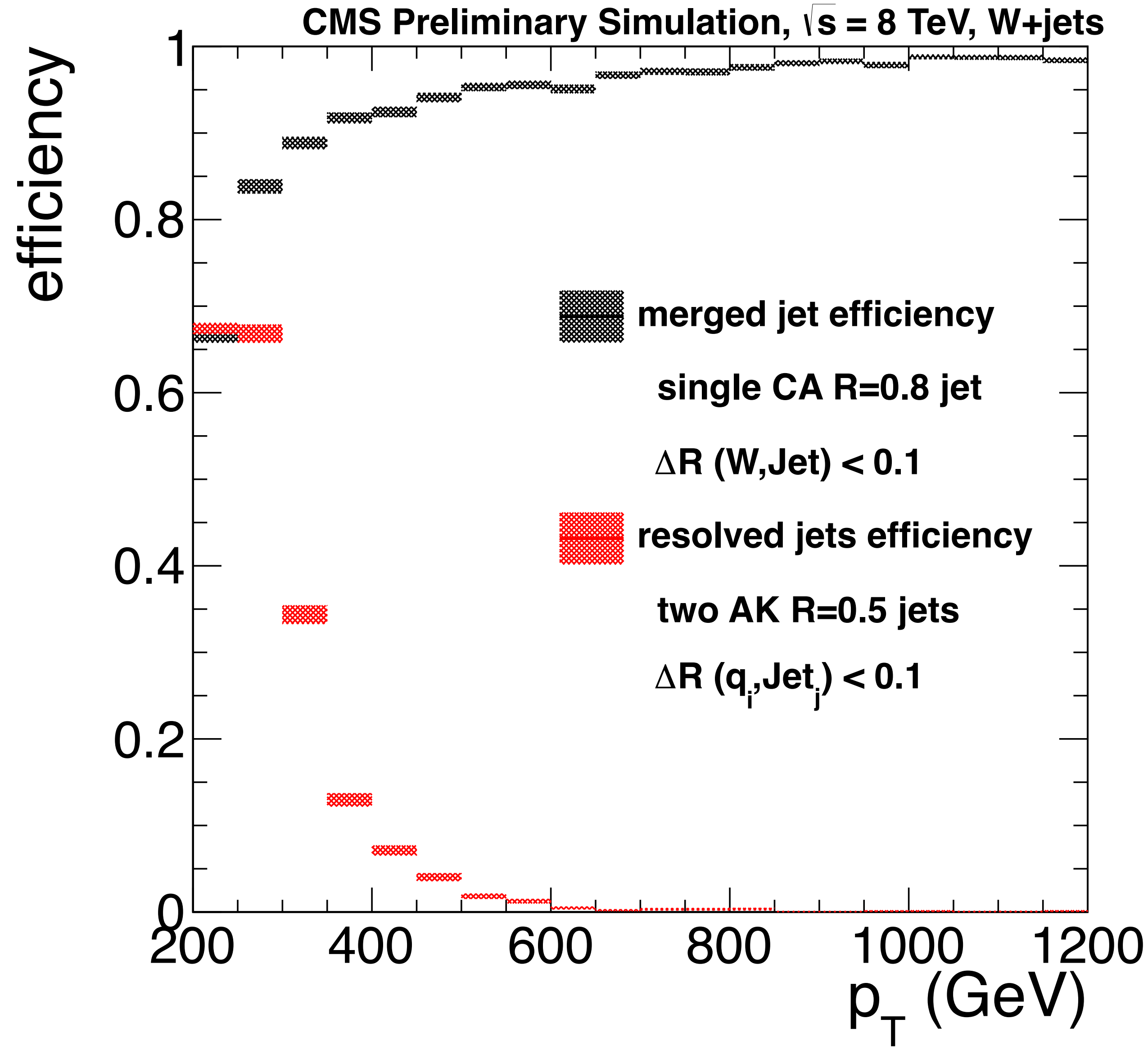
Used to trigger events

Use substructure tools to identify hadronic W (or Z)

High mass Higgs as bench mark for a generic WW resonance of M > 600 GeV

why substructure?





Boosted regime has higher mass reach and less backgrounds
Hadronic W's start to merge around 200 GeV

analysis features

Backgrounds:
W+jets (dominant)
ttbar,
single top,
WW/WZ

Signature:
WW → {l + MET} + fat jet

Use jet substructure techniques to identify W-jets, tagging efficiencies taken from top-enriched control regions

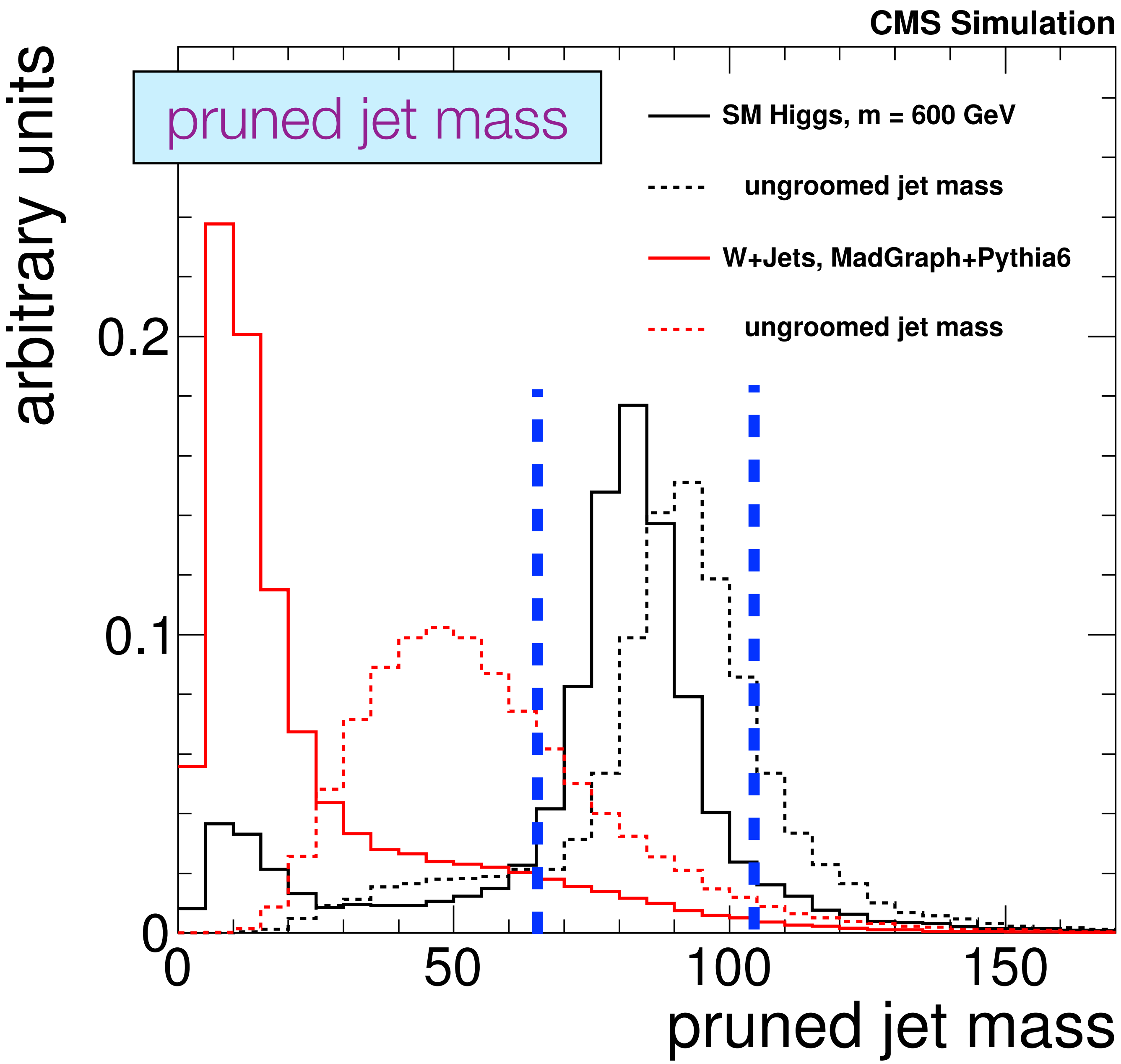
Kinematics:
highly boosted W(lv) back-to-back with fat jet

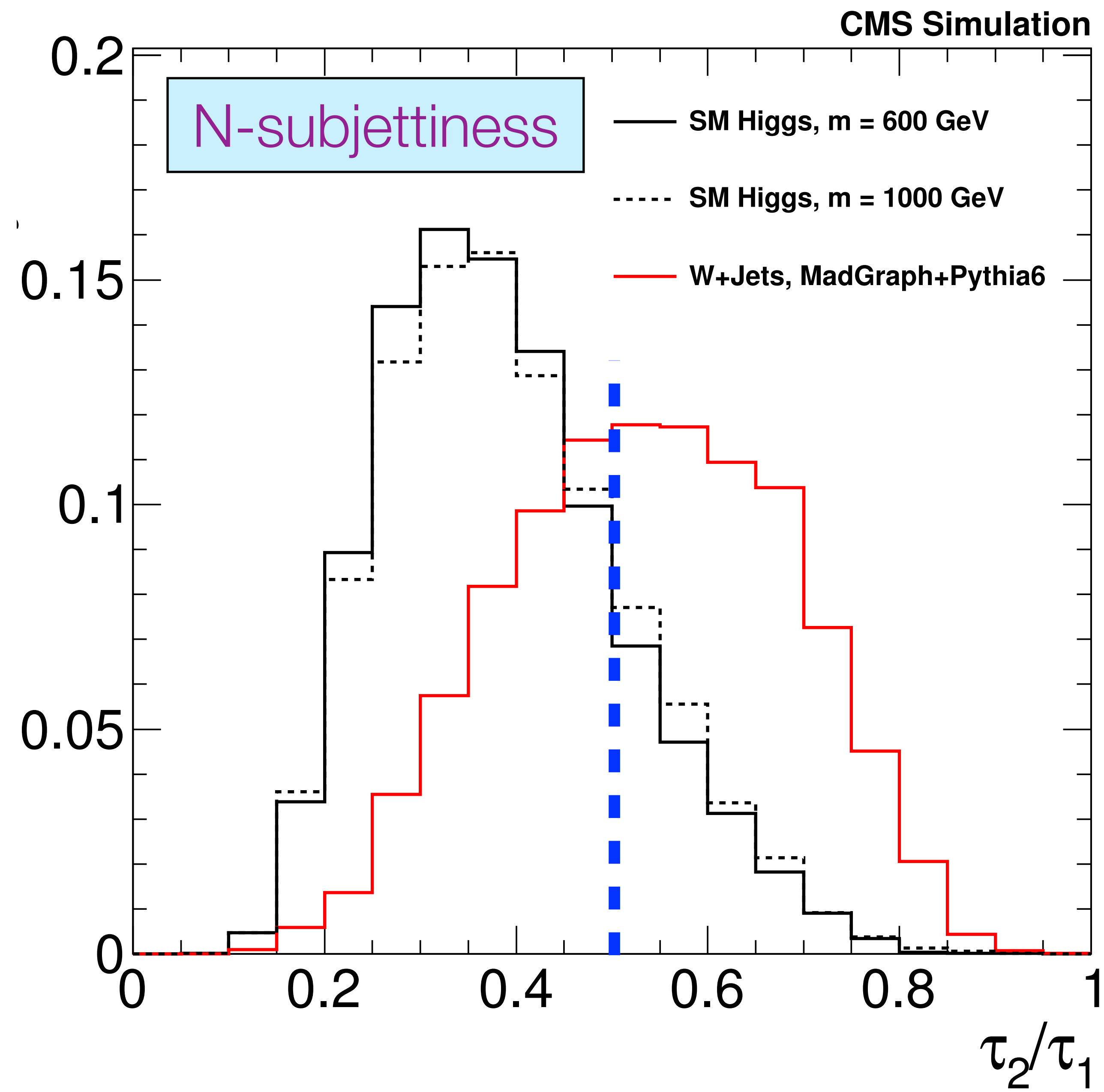
Discriminating observables:
pruned jet mass, m_J, and three-body mass, $m_{l\nu J}$

Background estimation:
data-driven background extraction via m_J sideband for W+jets and top-enriched control regions

Unbinned shape limits using $m_{l\nu J}$ shape
(B)SM Higgs and Graviton limits

W-tagging





N.B. the correlations between discriminating variables are important

comparisons of τ_2/τ_1 and other observables in backup

W-tagging

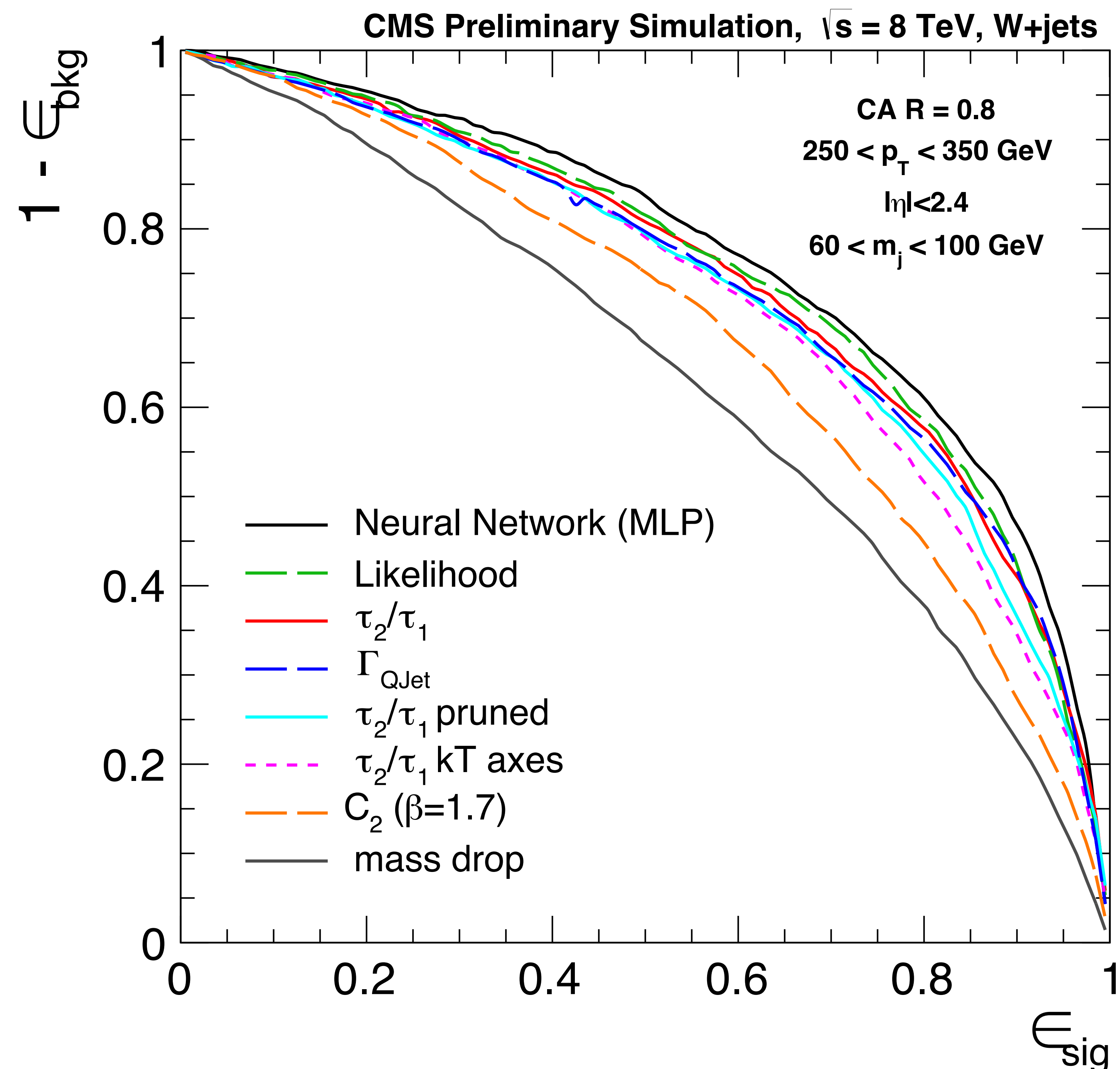
After pruned mass cut...

Compare the performance of several different observables

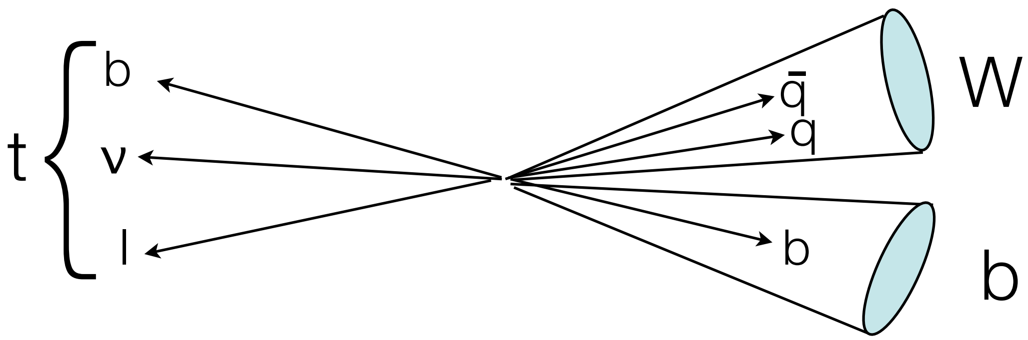
- Qjets and N-subjettiness are most powerful single variables

Try combining observables into an MVA discriminant; a small improvement

Set baseline tagger as pruned jet mass + τ_2/τ_1 cut

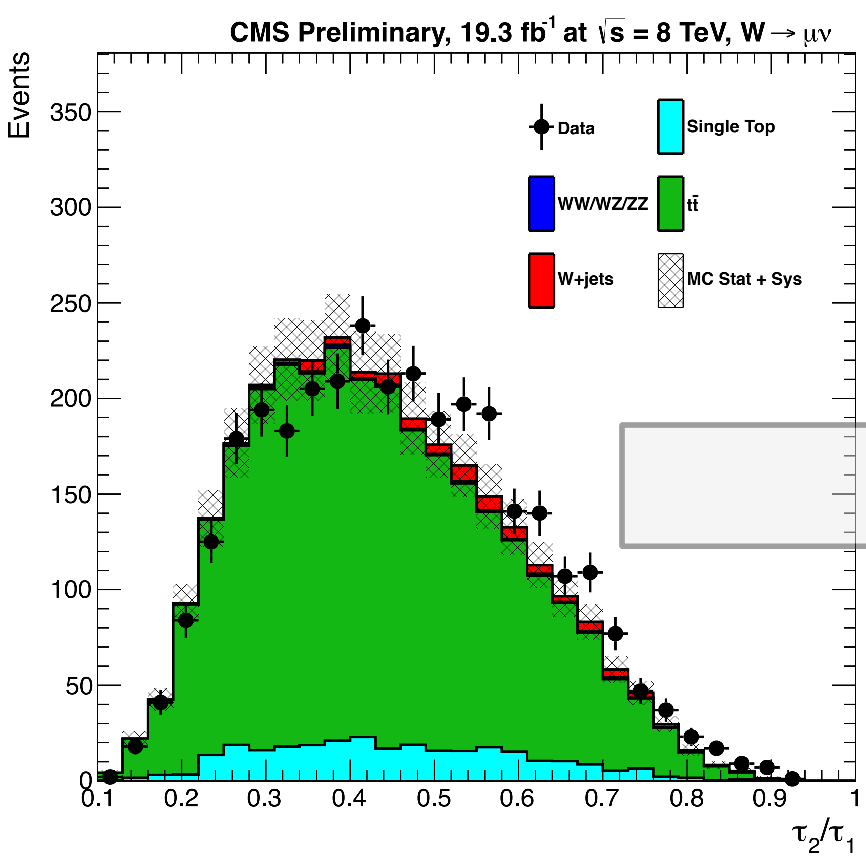


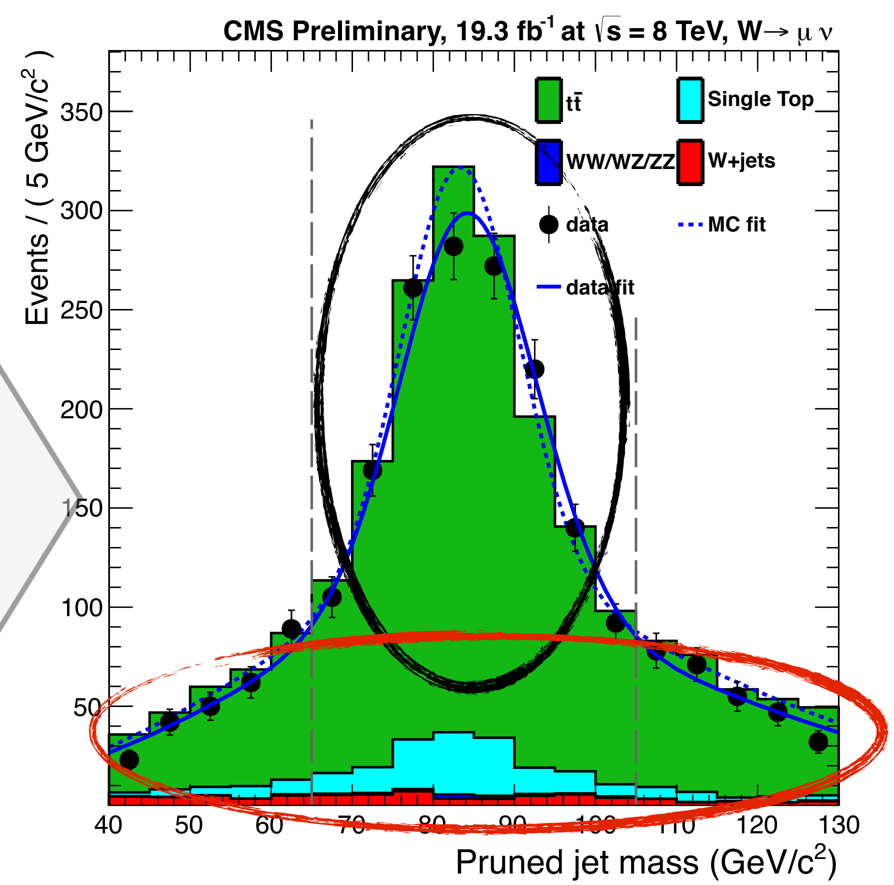
top-enriched control region



top-enriched control region: by inverting the b-tagging veto, we can get a very pure sample of semi-leptonic ttbar

1. extrapolate the ttbar background into the signal region
2. compute W-tagging efficiency, mass scale and resolution scale factors





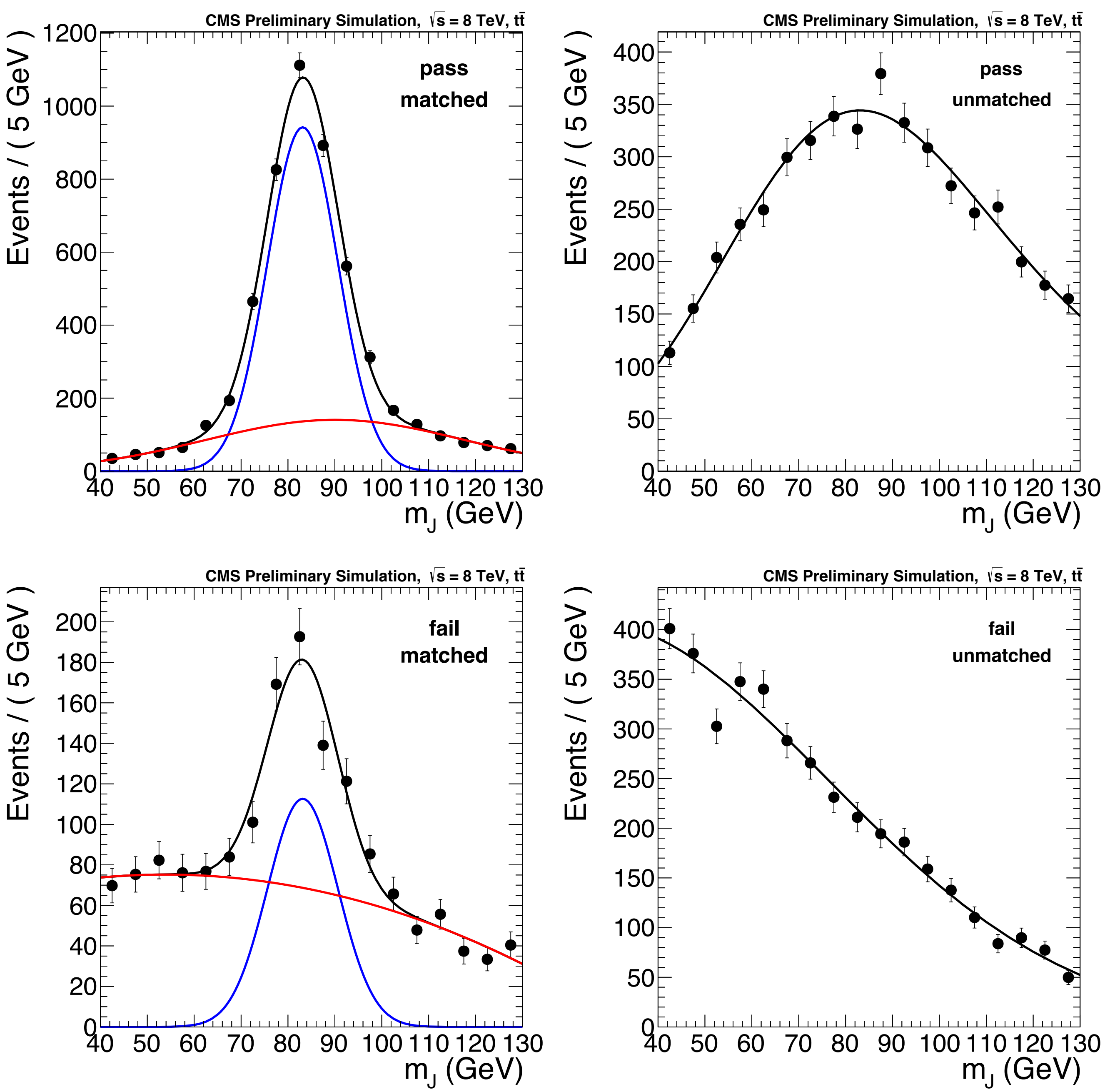
What is the efficiency of tagging a W, not from top decays?

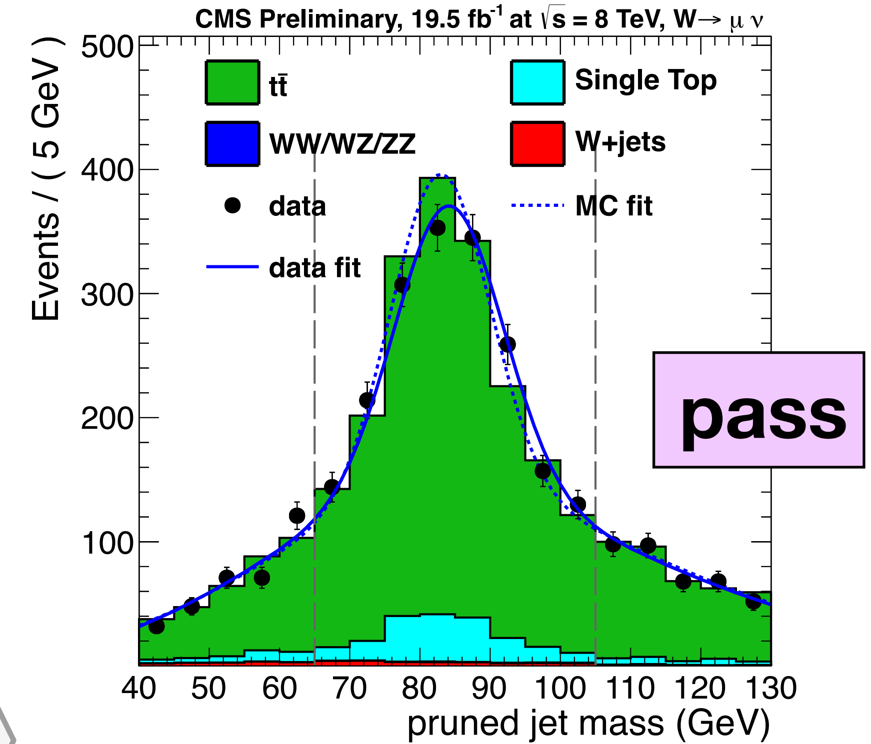
Compute the pure W-tagging efficiency from background subtracted fit.

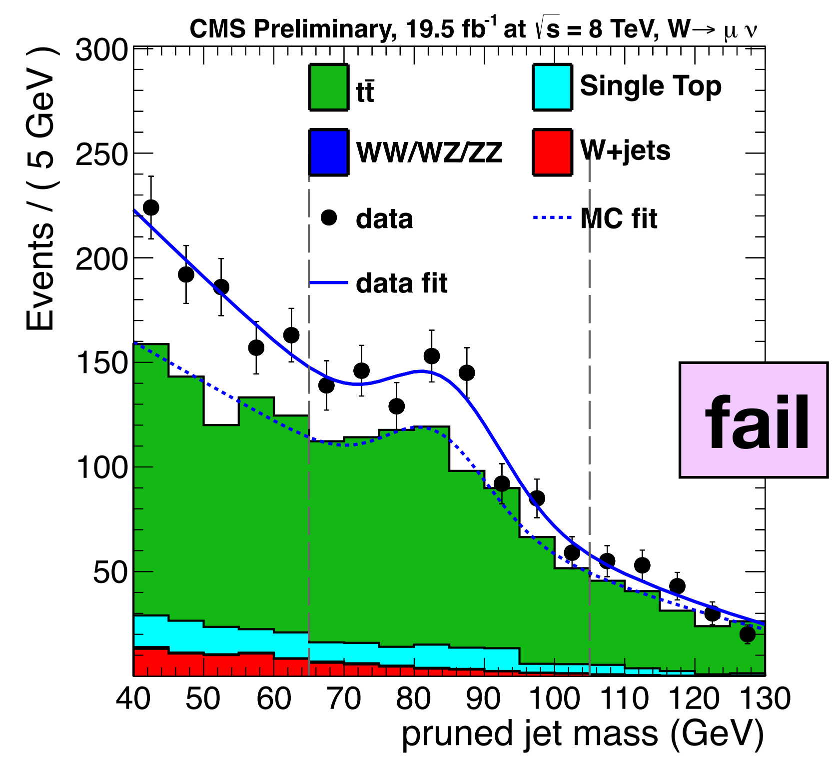
W-tagging efficiency, scale, resolution

Perform simultaneous fit of the 'pass' and 'fail' samples

Shapes inspired from MC, but floated in data

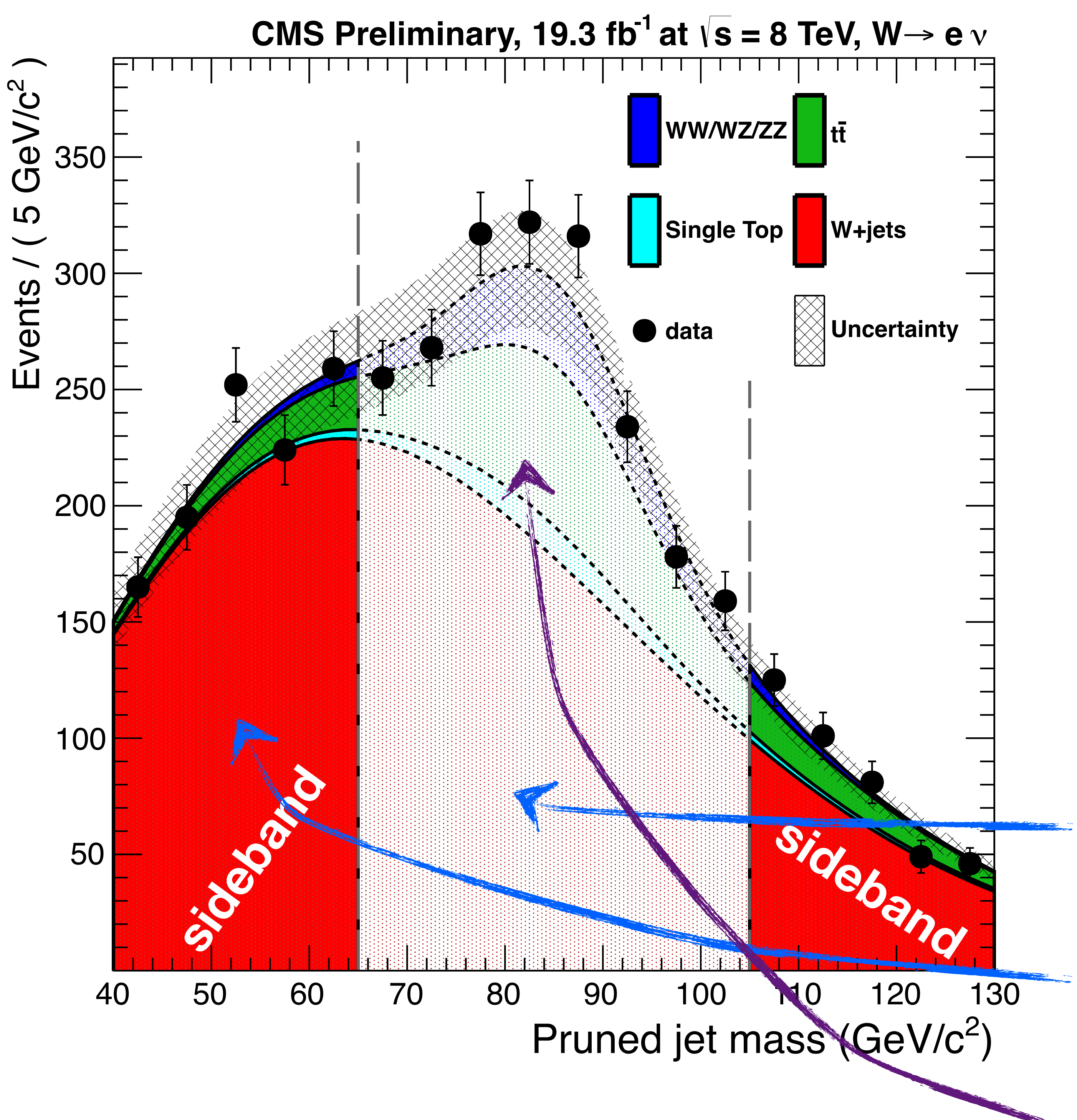


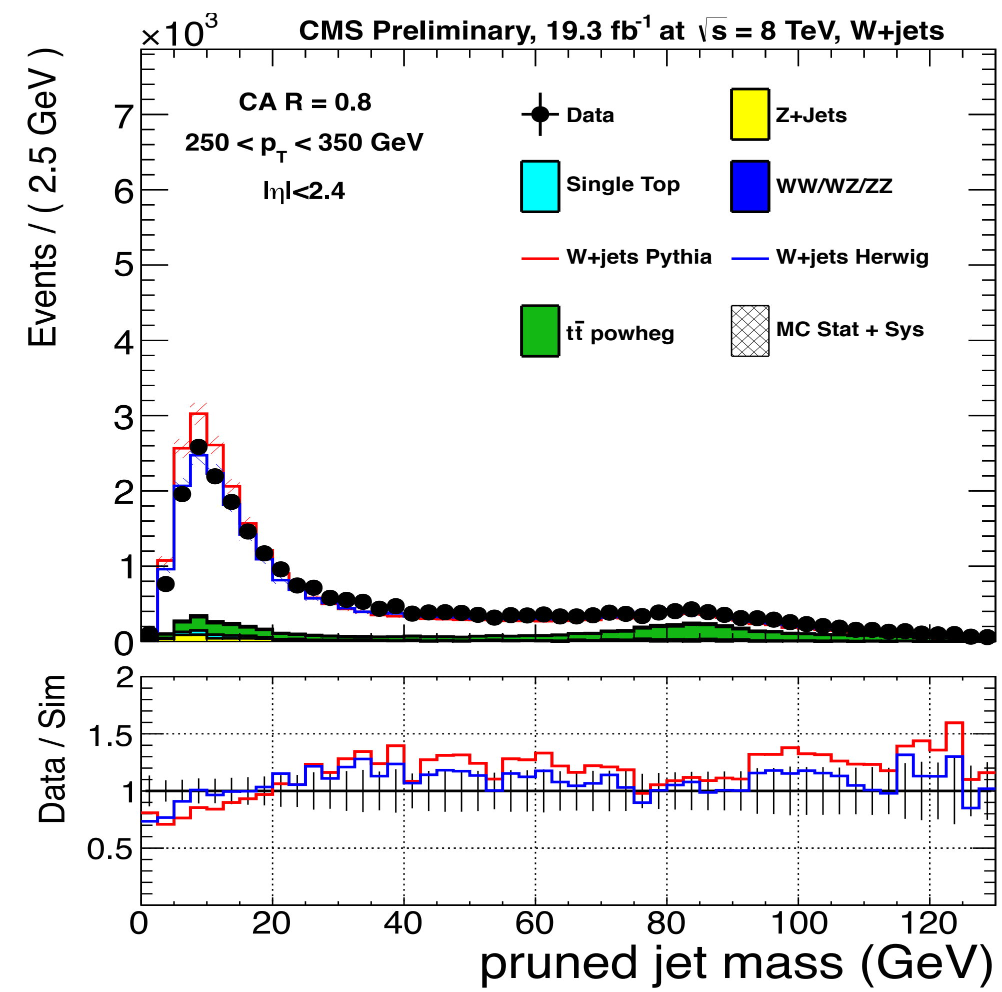




efficiency data/MC scale factor = 0.93±0.08

background estimation





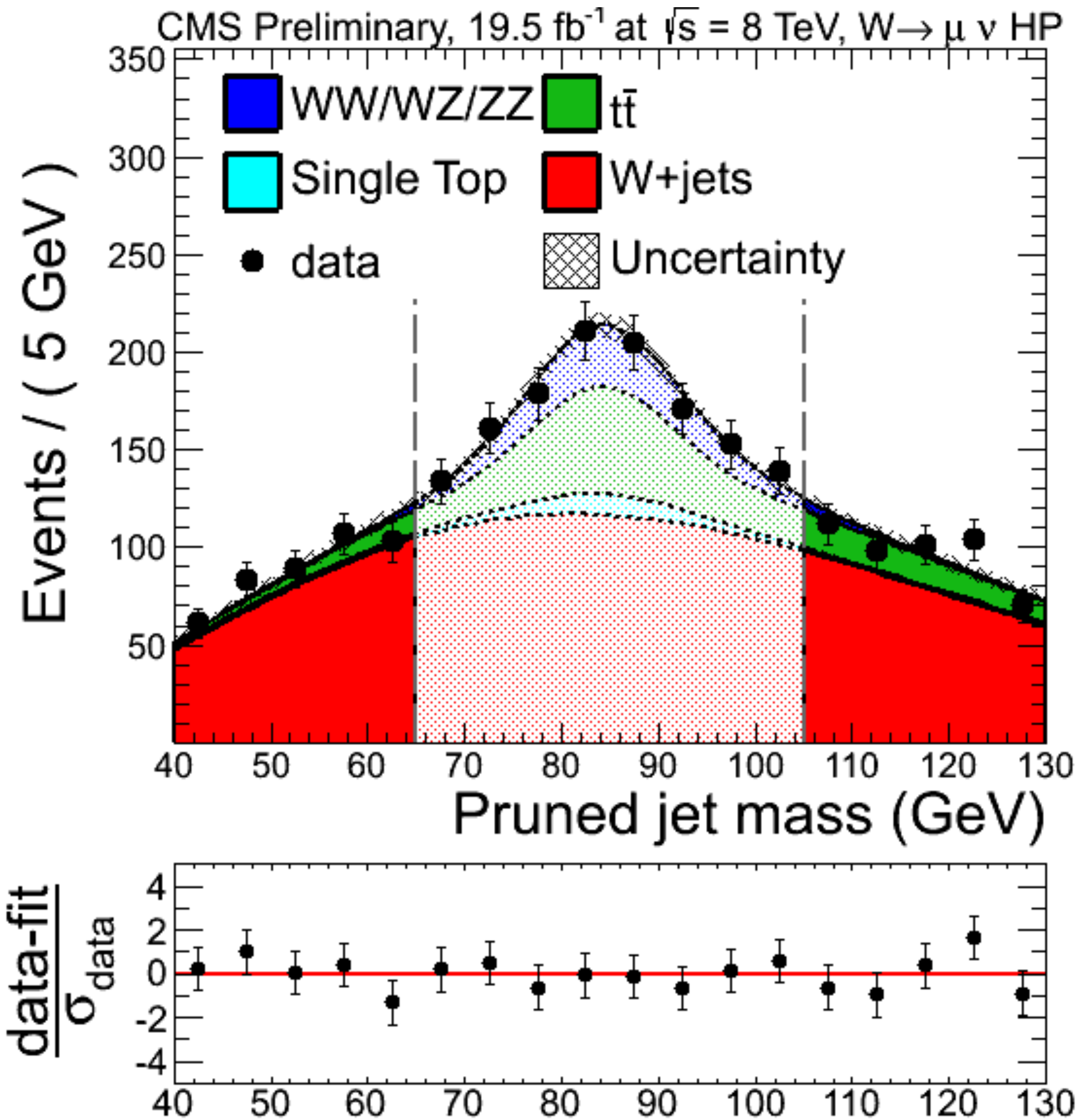
W+jets normalization estimated from sideband fit

W+jets $m_{l\nu J}$ shape extrapolated from data sidebands into signal region via MC transfer function

ttbar normalization taken from top-enriched control region; shape from MC constrained by variations on parton shower, matching and scale variations

other subleading backgrounds constrained from MC

aside: higher p_T



Same plot as previous slide, but for higher signal masses

ttbar background starts to disappear as tops become full merged, diboson contribution starts to become more significant

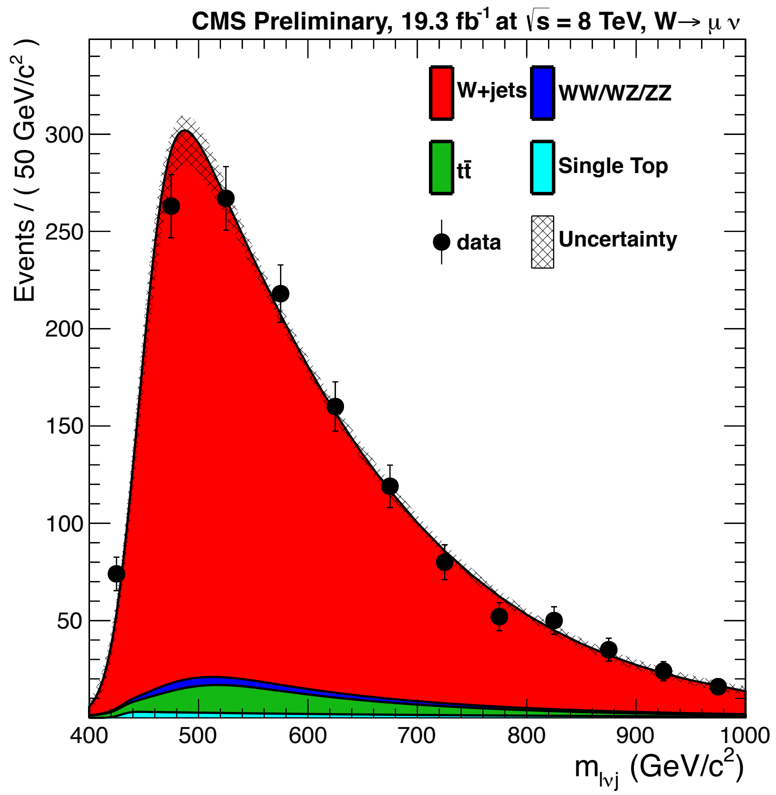
background estimation

$$F_{\text{data,SR}}(m_{l\nu j}) = \alpha_{\text{MC}}(m_{l\nu j}) \times F_{\text{data,SB}}(m_{l\nu j})$$

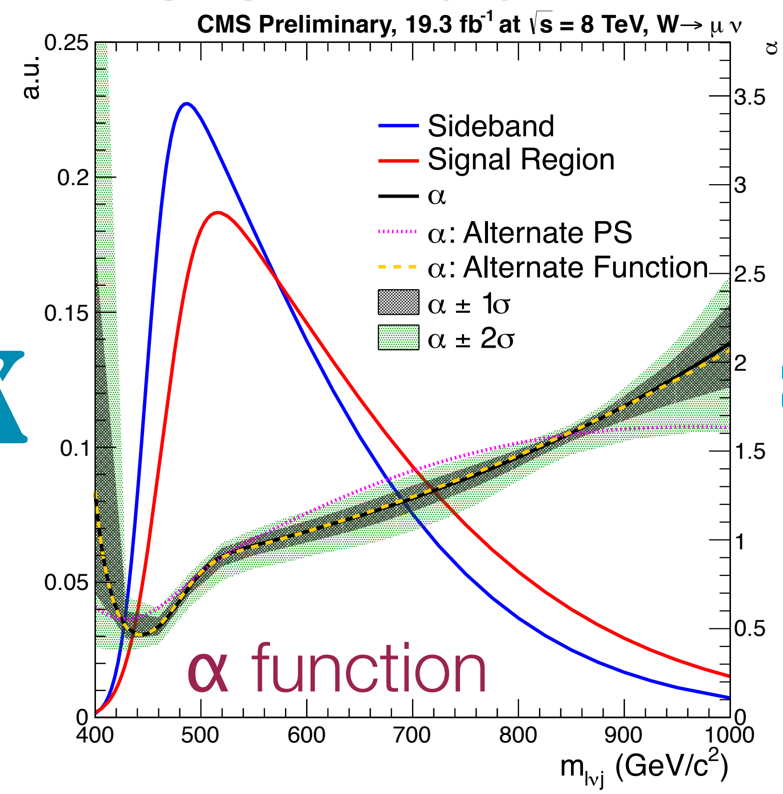
$$\alpha_{\text{MC}}(m_{l\nu j}) = \frac{F_{\text{MC,SR}}(m_{l\nu j})}{F_{\text{MC,SB}}(m_{l\nu j})}$$

Putting all together: Final $m_{l\nu J}$ distribution in signal region

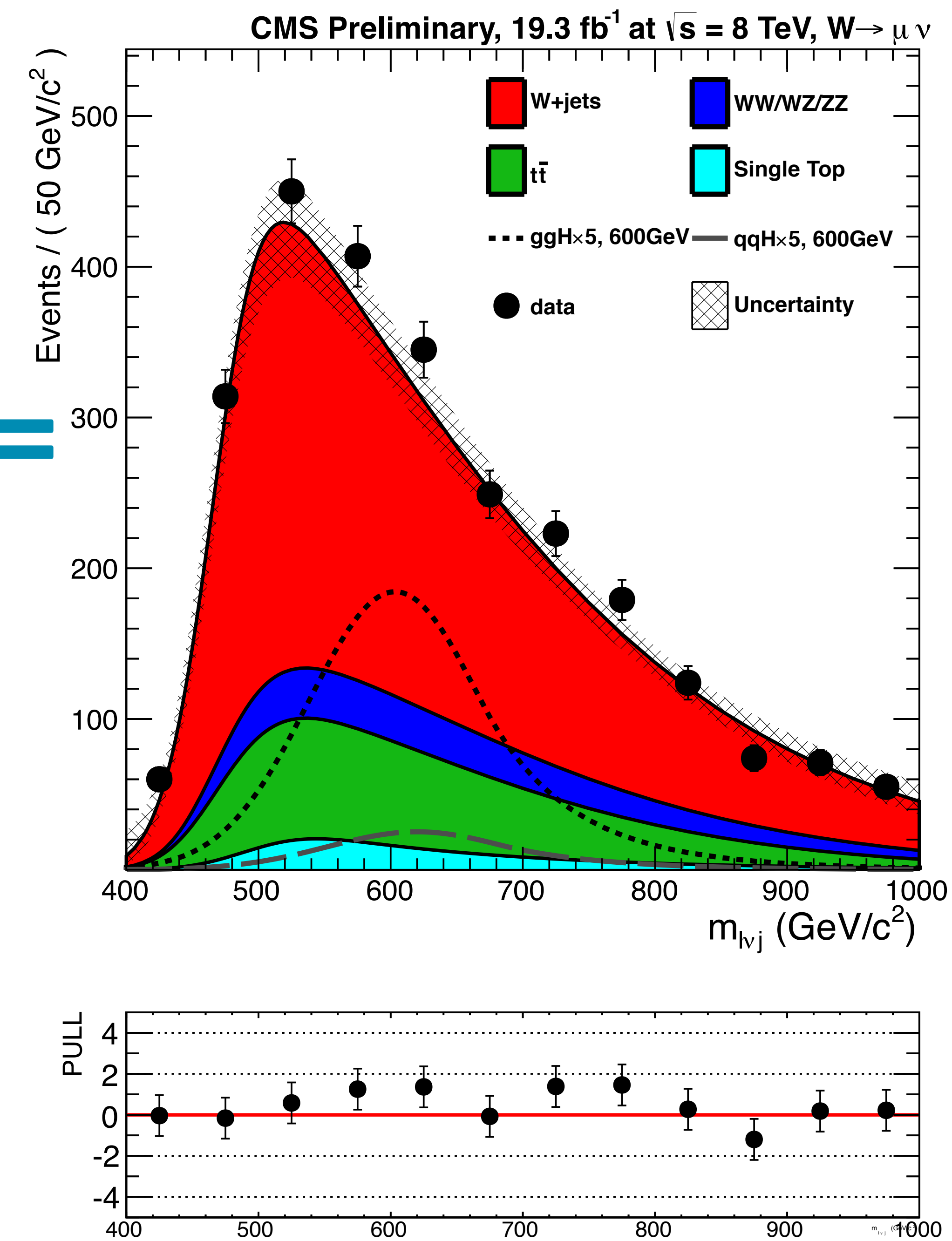
m_J sideband



X



=



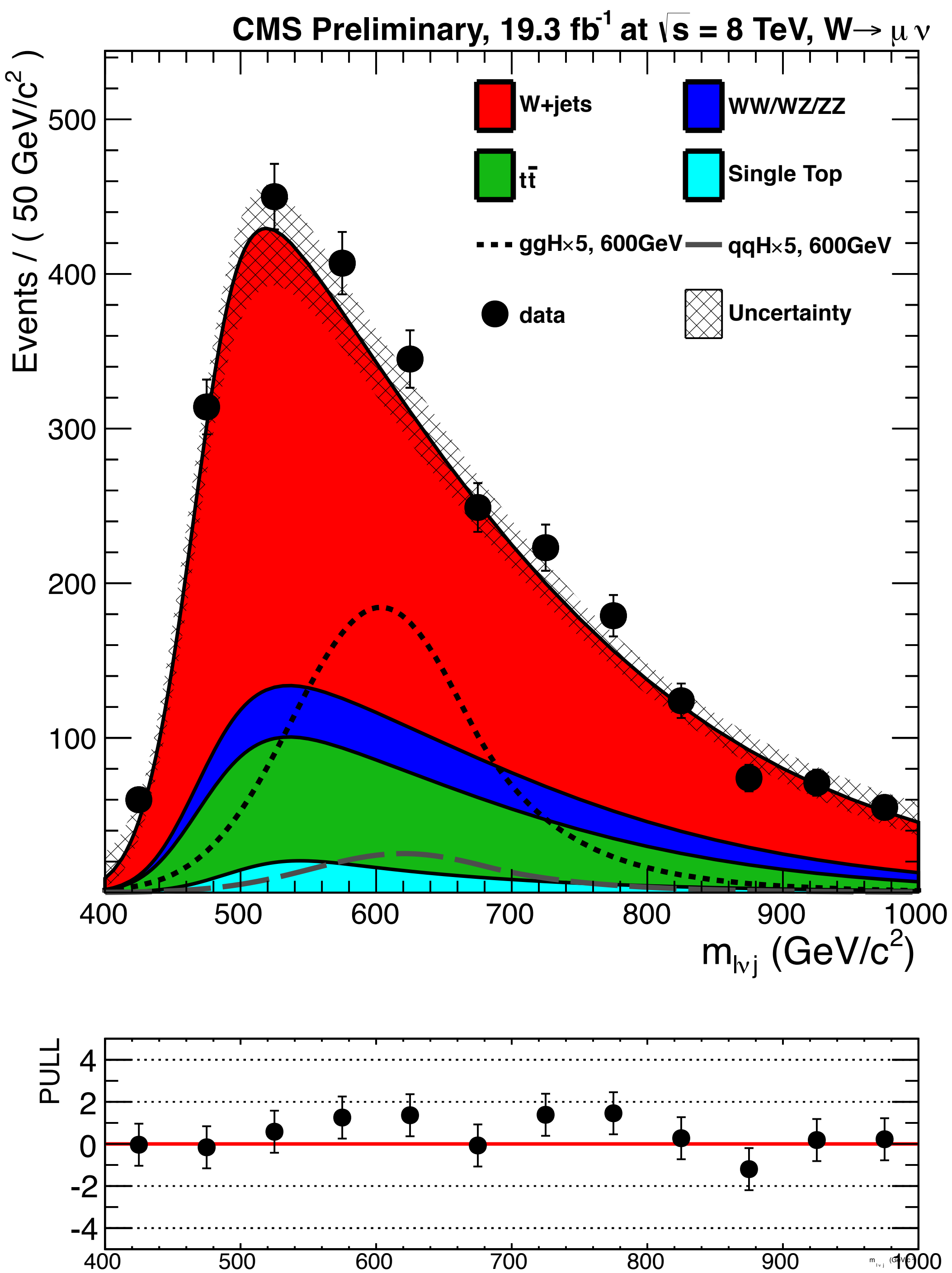
Uncertainties on W+jets shape/normalization from:

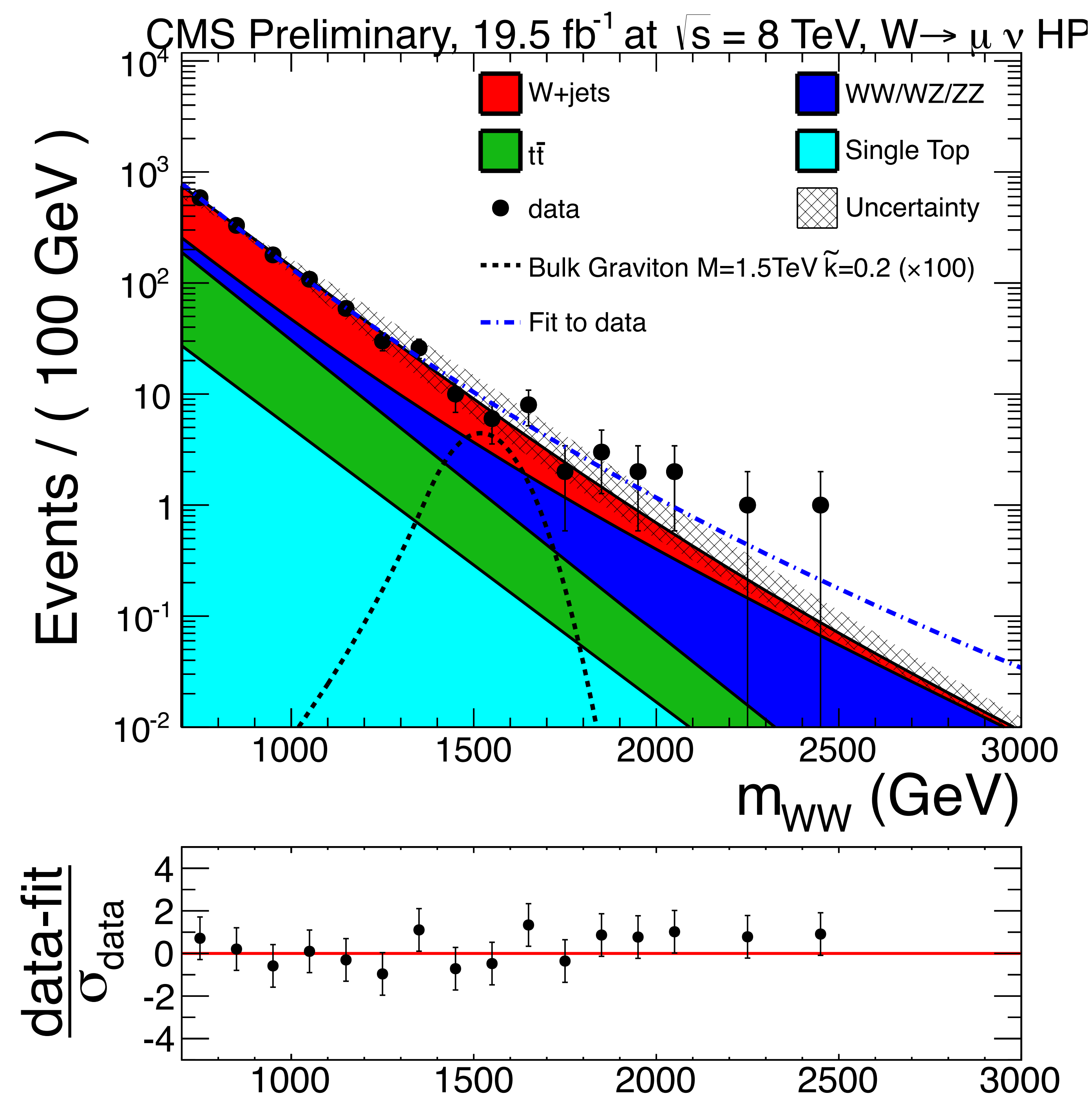
W+jets sideband fit

α function fit shape uncertainty

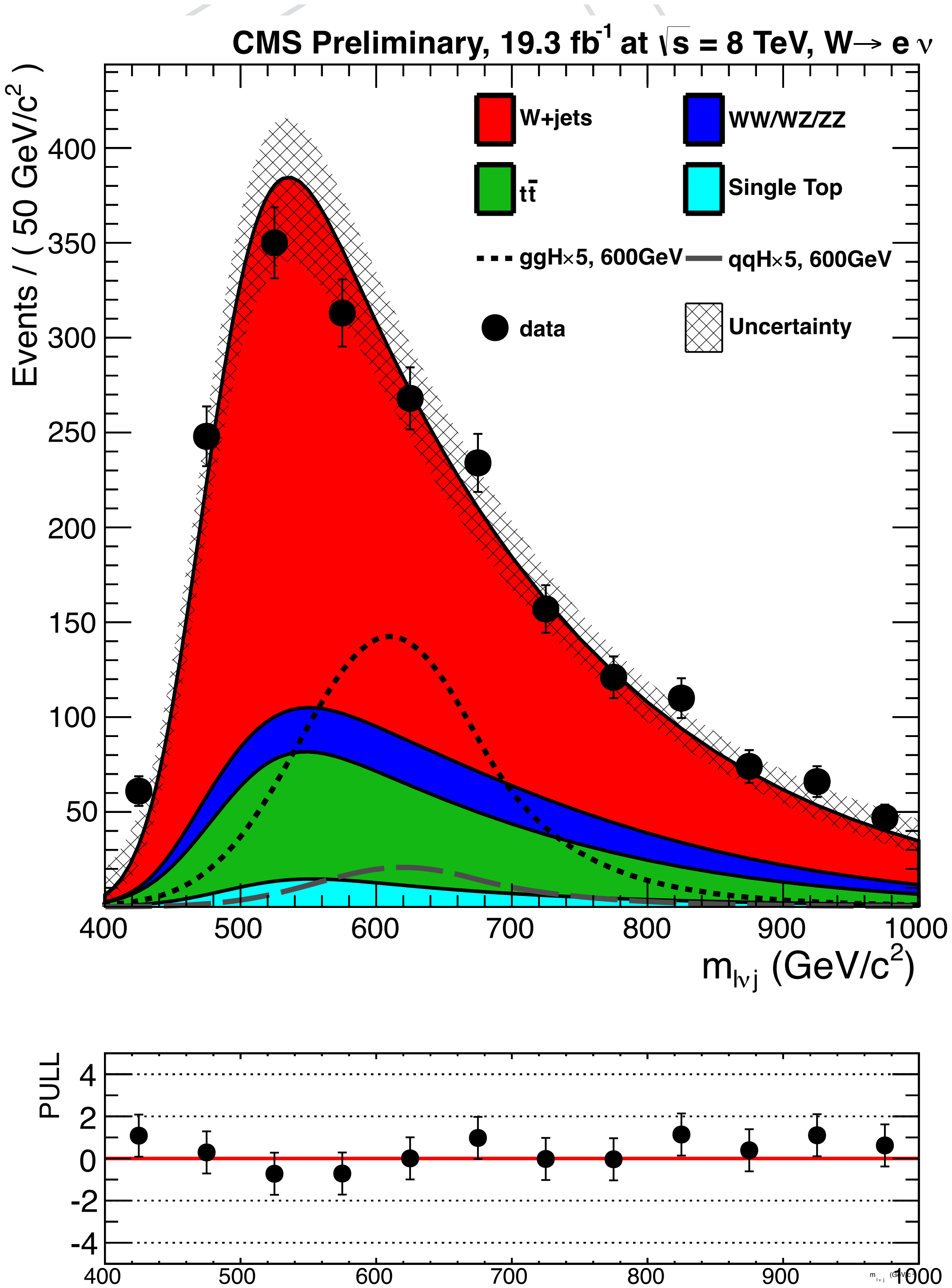
Shape uncertainty from alternate parton shower and alternate fitting functions

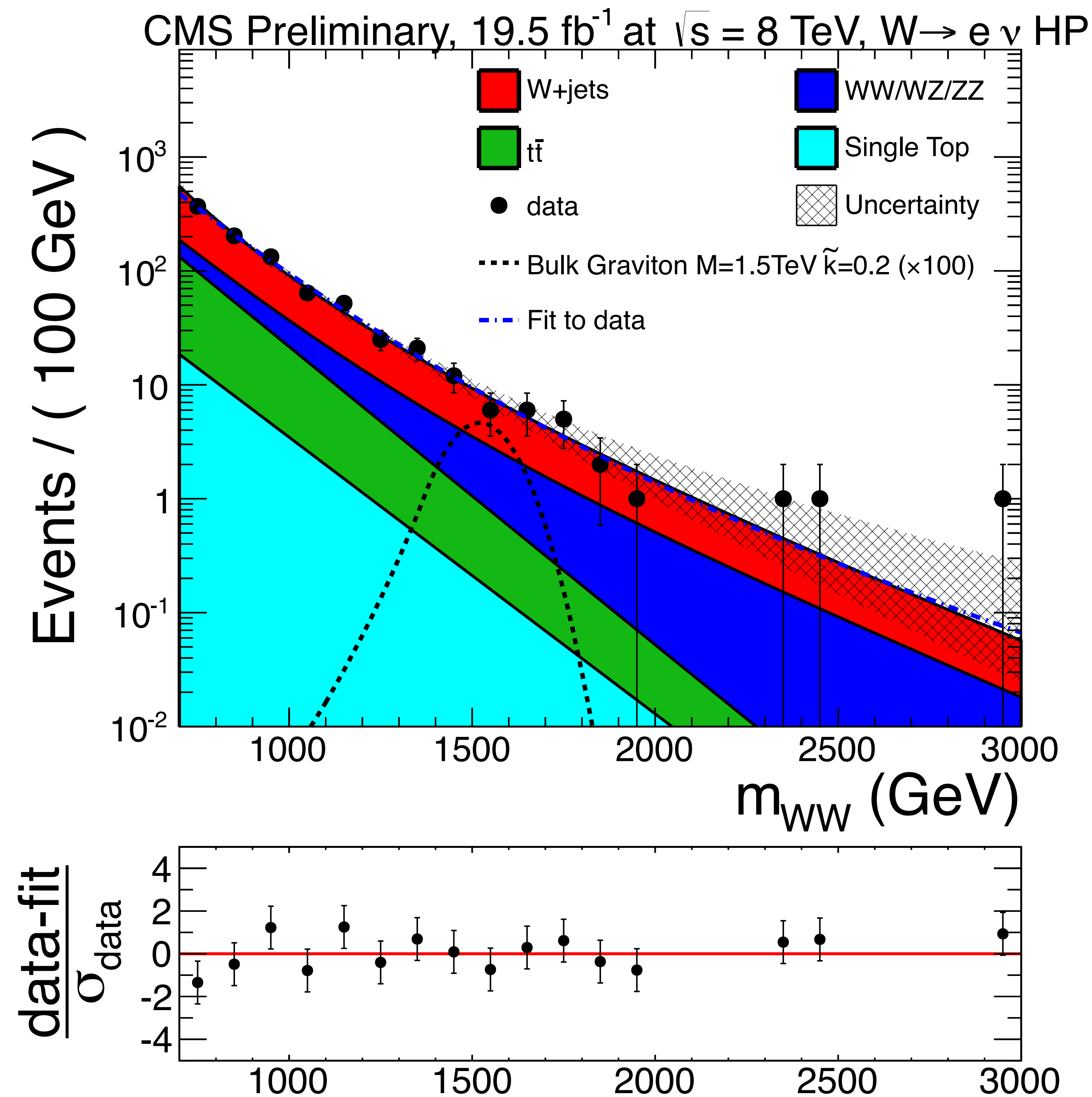
final m_{WW} distributions, μ channel

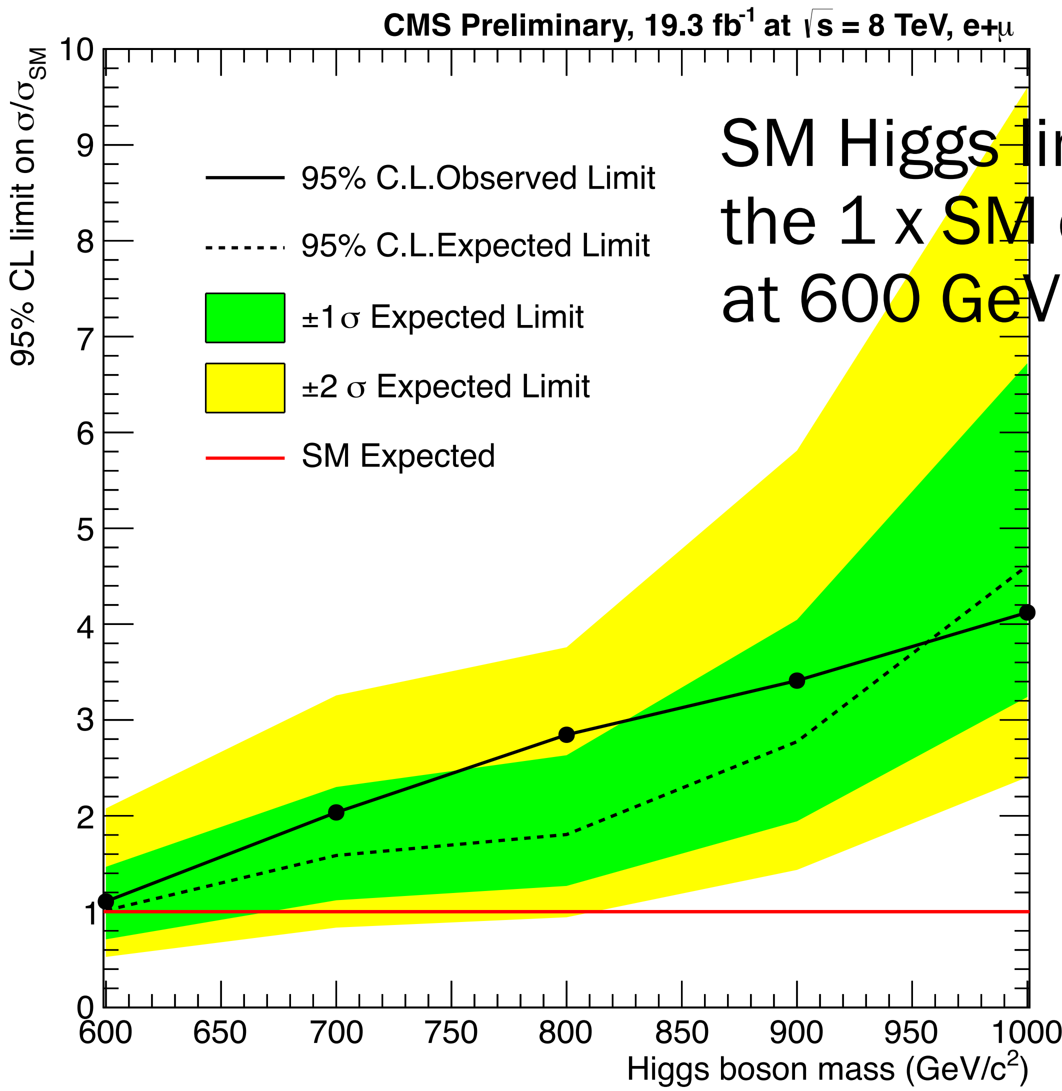




final m_{WW} distributions, e channel

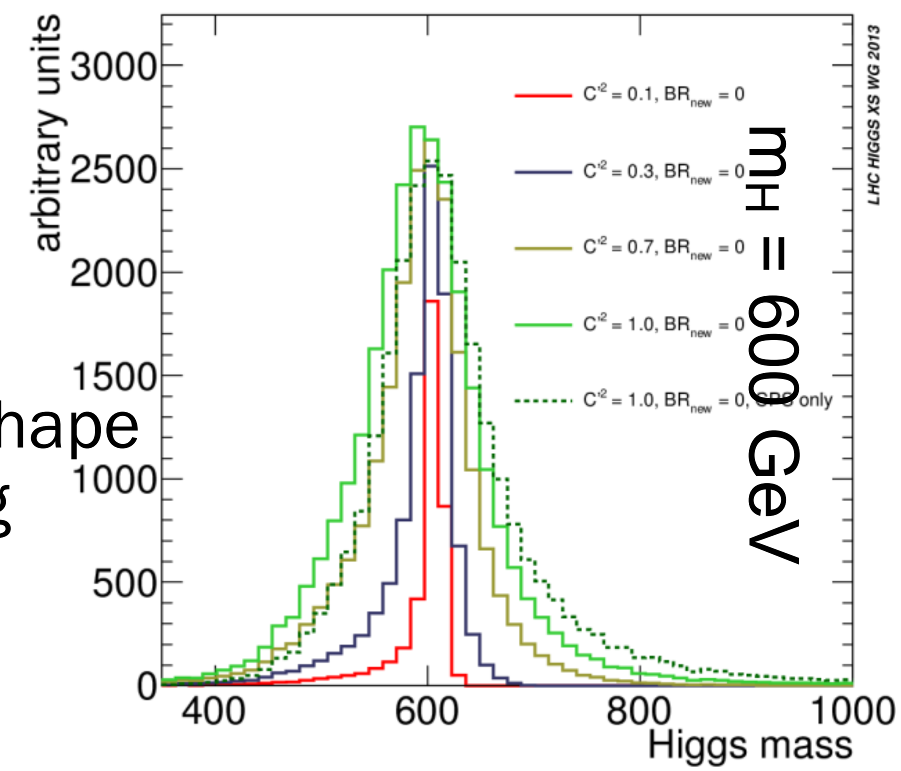




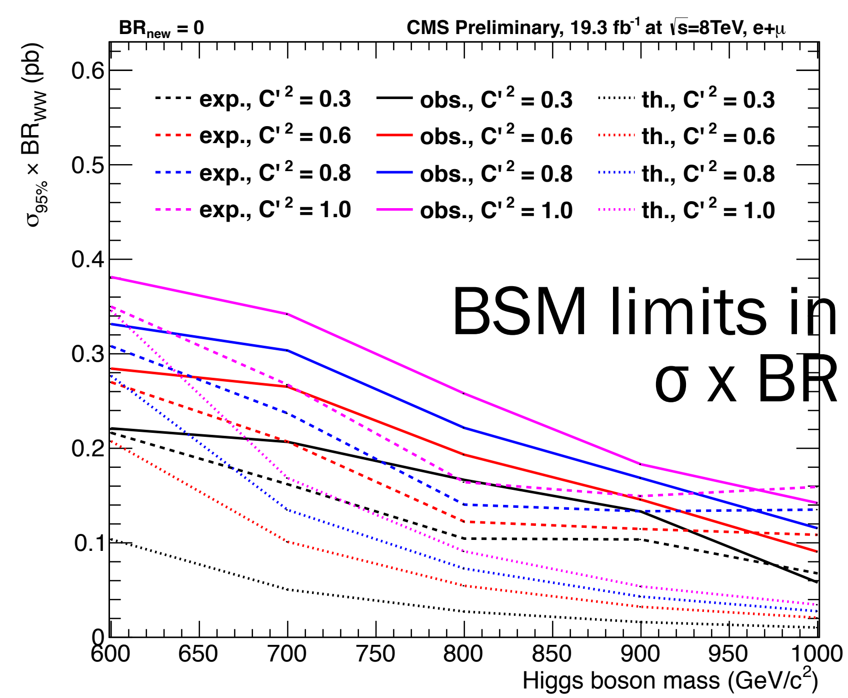


SM Higgs limits are just at the 1 x SM cross-section at 600 GeV

Higgs lineshape reweighting



m_H = 600 GeV



BSM limits in σ x BR

BSM models benchmark (LHC XS WG):

Higgs mixes with heavy EWK singlet

Higgs coupling modified -- $C^2 + C'^2 = 1$

Width and cross-section is modified accordingly

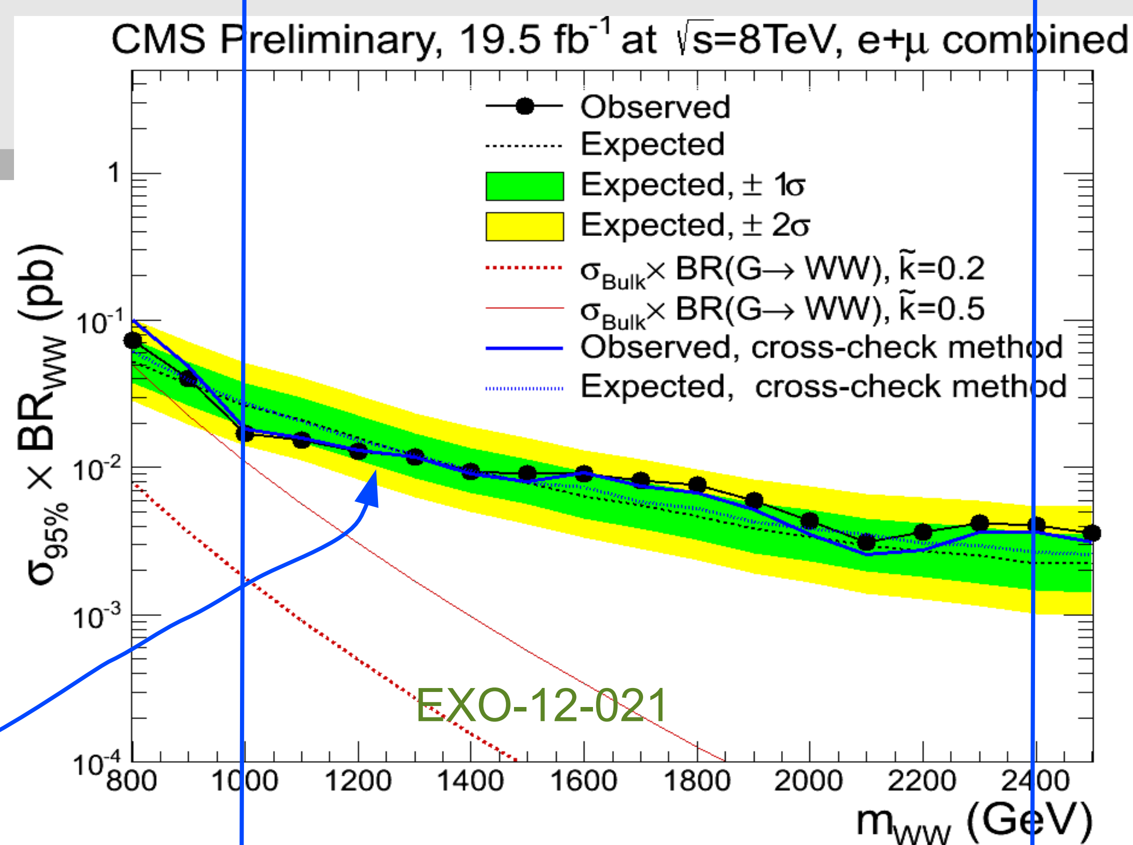
graviton interpretation

petar maksimovic

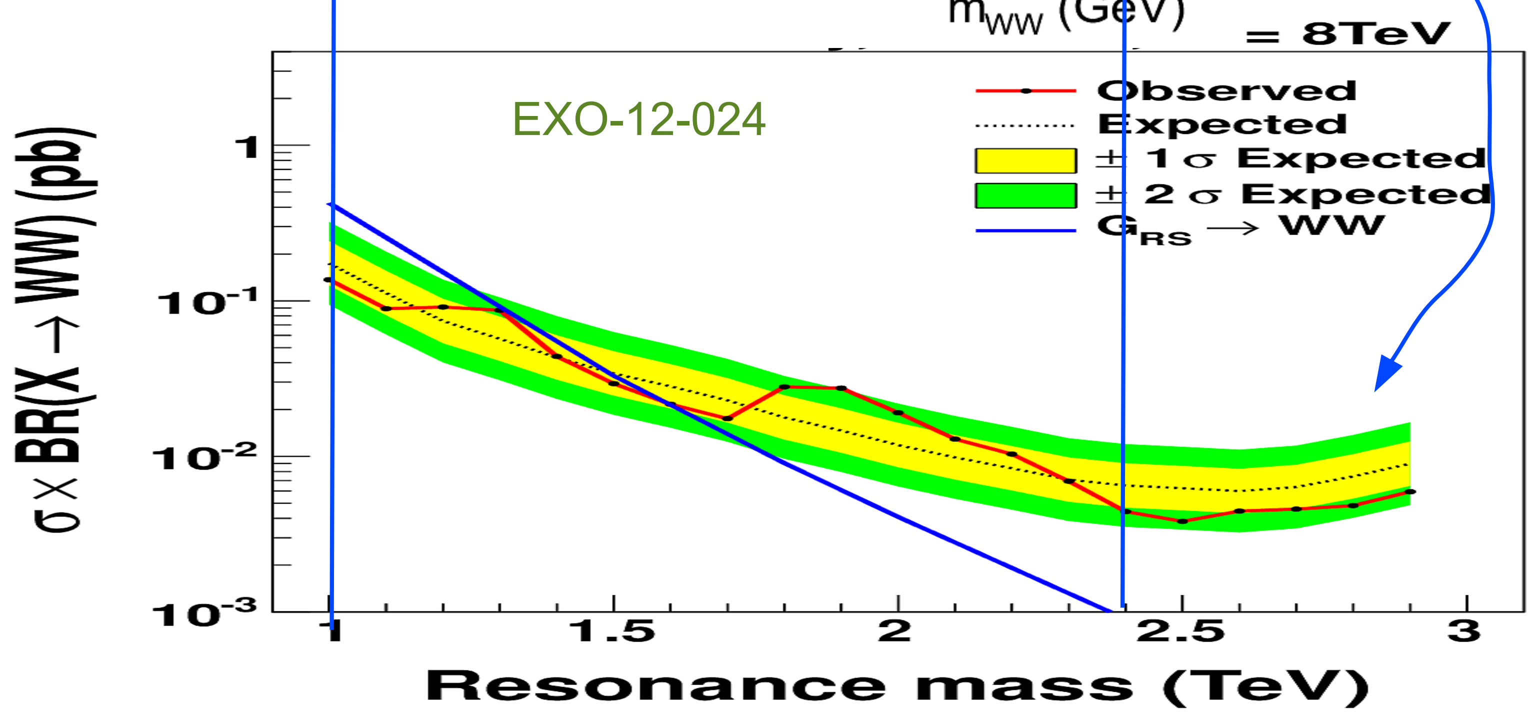
- Comparing two WW searches

(please ignore theory curves)

l+jets search does better for most of the mass range



Dijet search goes to higher masses



- Plan to combine the two analyses

Petar Maksimovic, Johns H

discussion points for the future

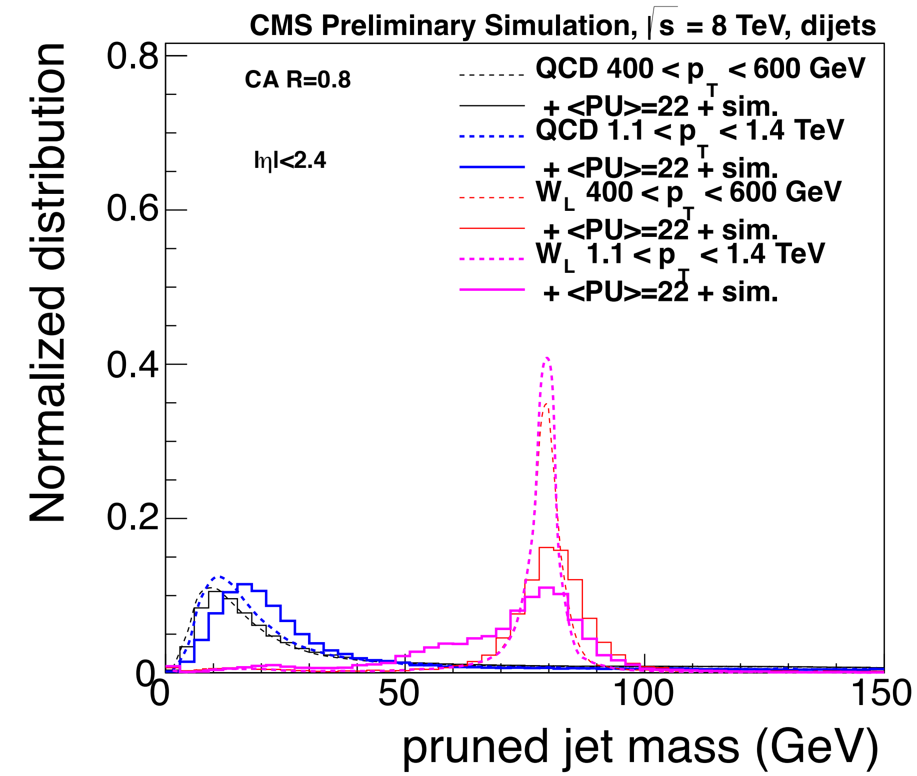
what happens at higher pT? methods begin to break down at W pT ~ 1.5 TeV

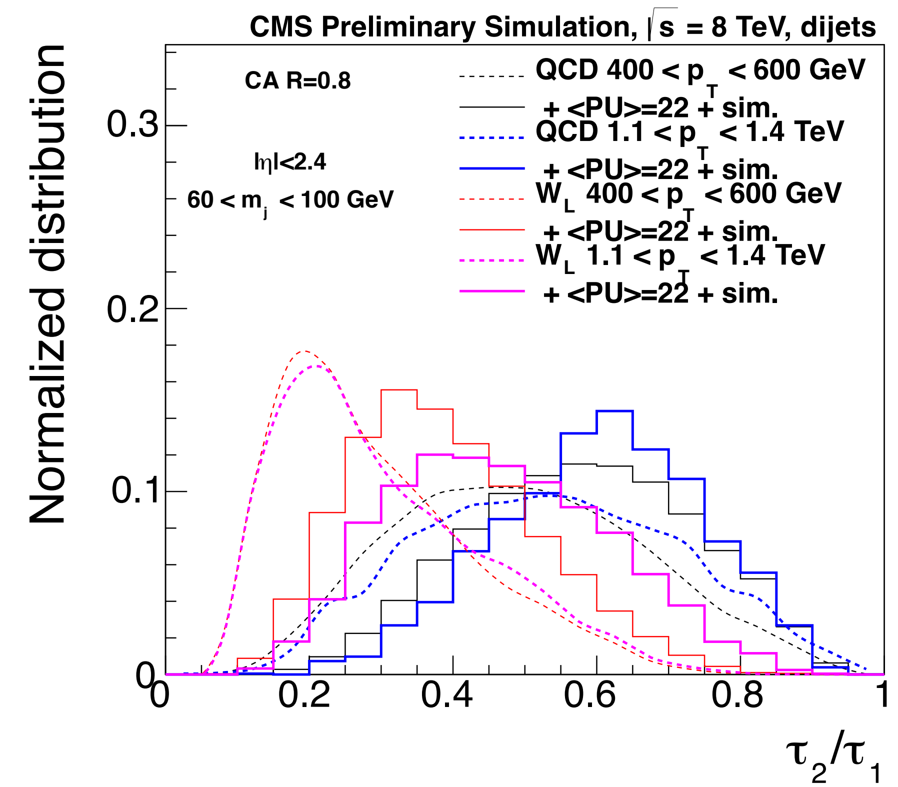
what happens at higher pTs? which variables are most robust to pileup?

how much would improved N$^{(x)}$LO MC and parton shower (tunes) reduced our systematic uncertainties?

for W's, how can we reduce systematics when extrapolating to where we have no standard candle? are there other calibrations samples?

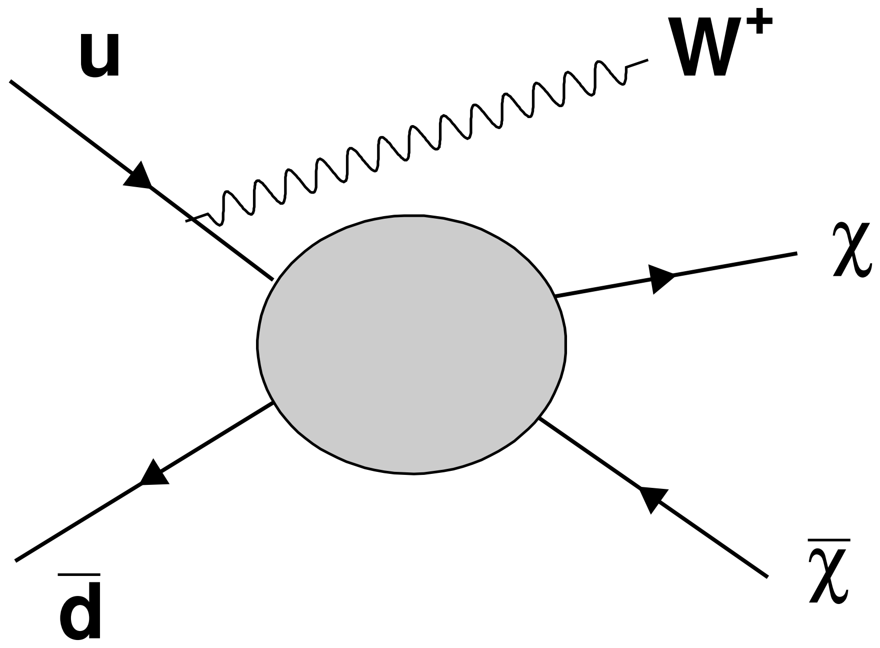
other kinematic variables? exploiting angular information?

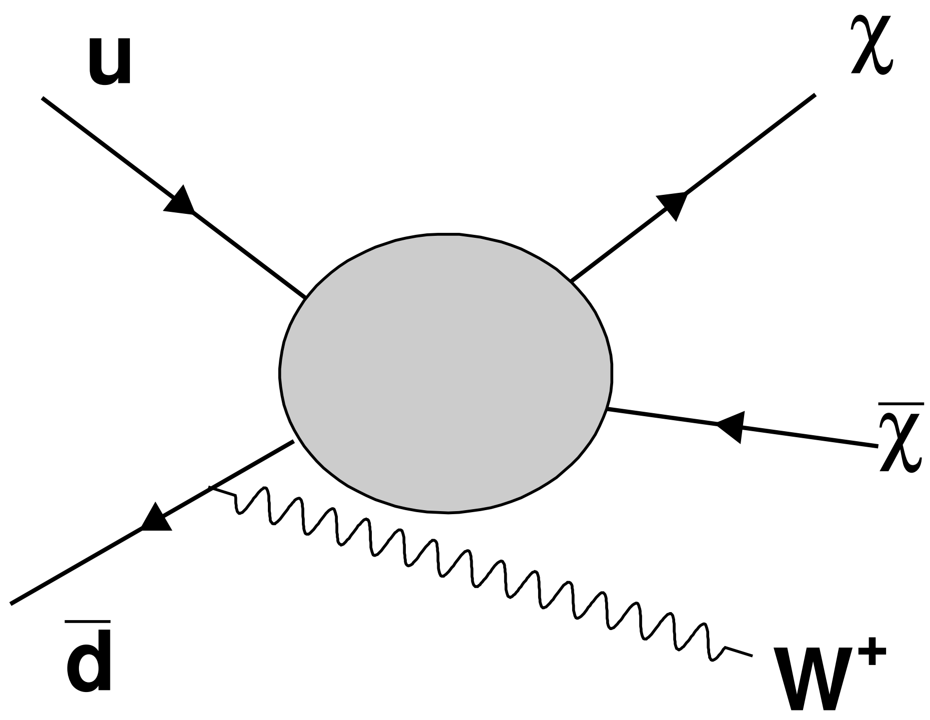




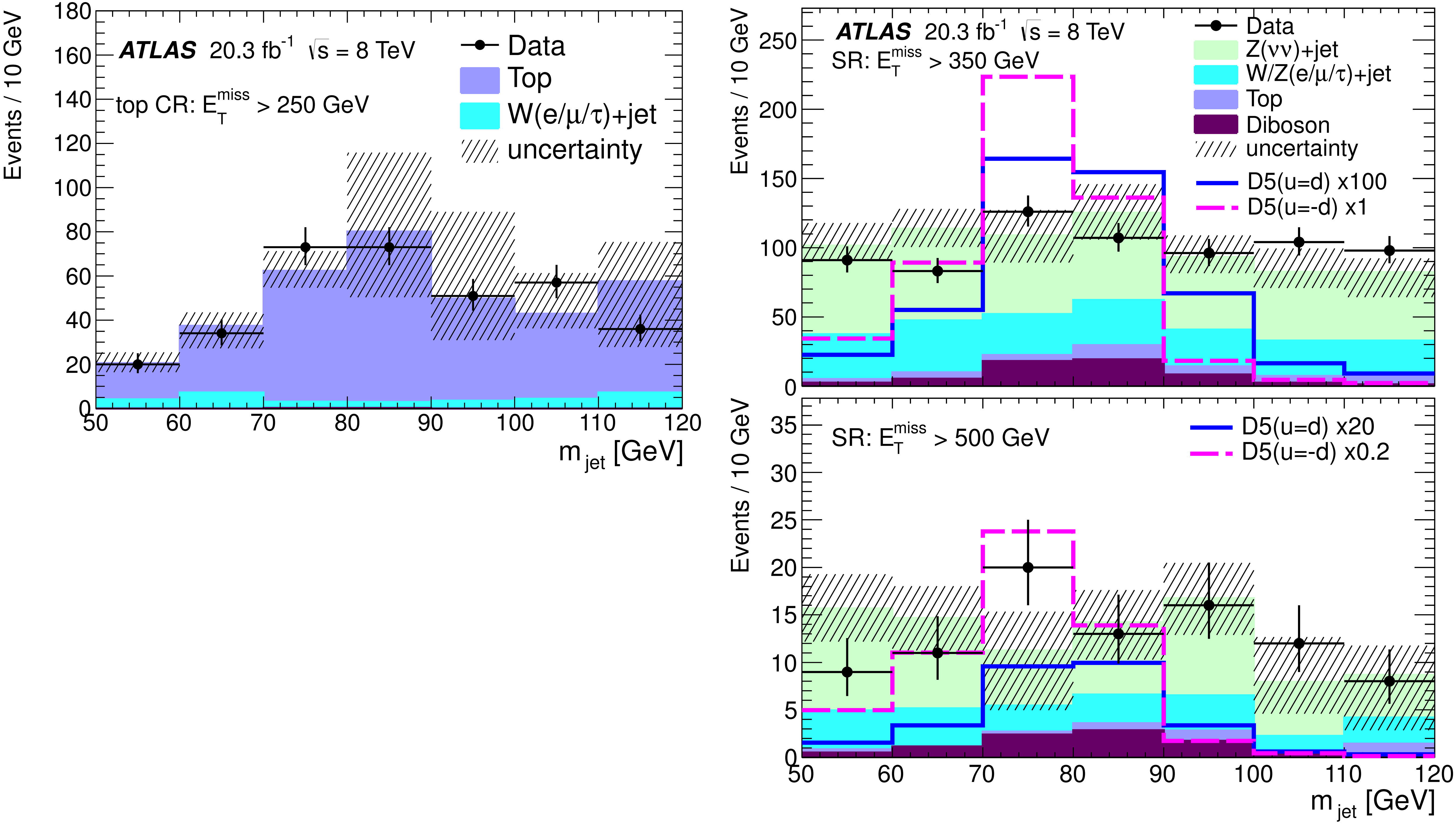
additional case study, mono-W

A variation of the above analysis



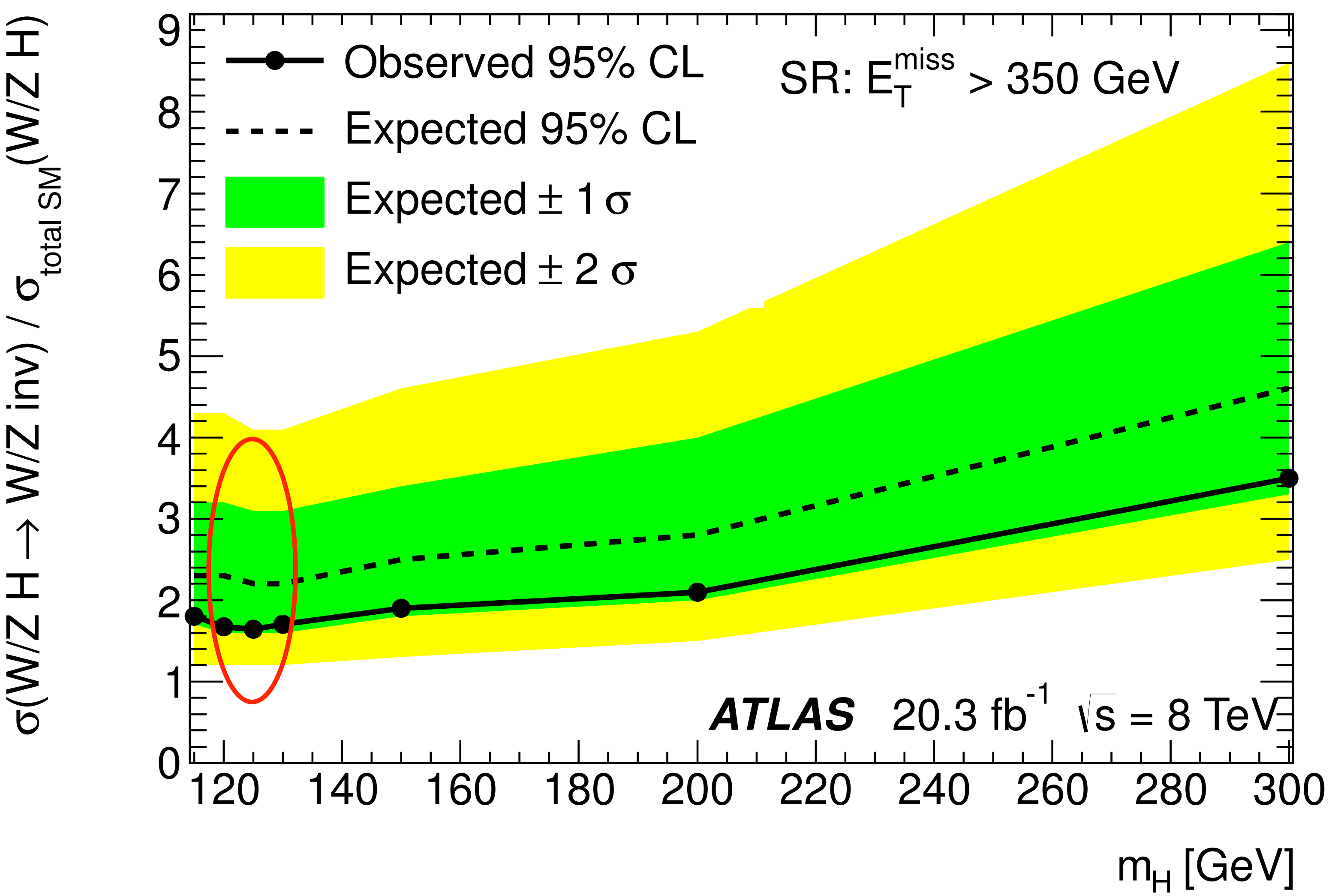


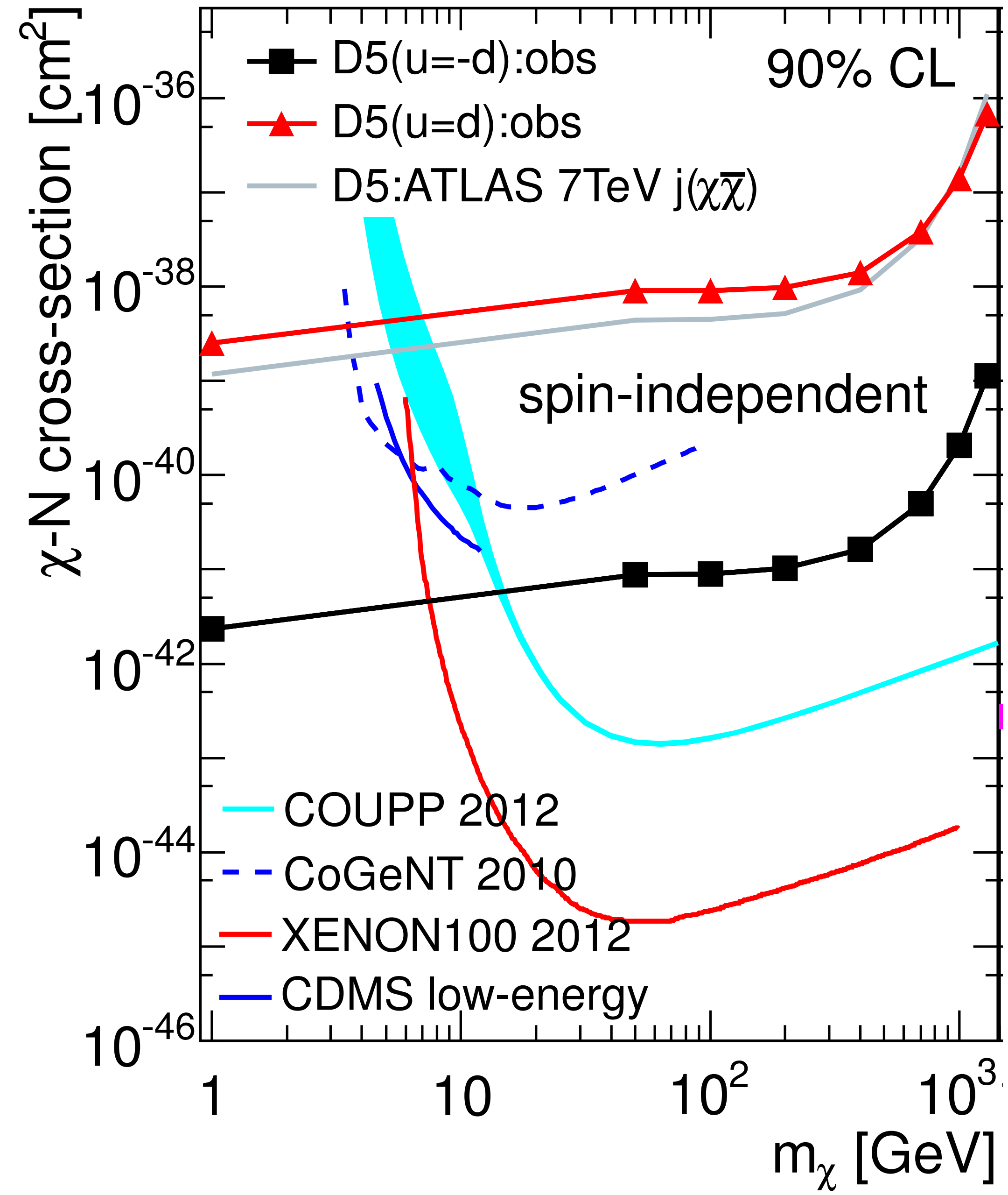
Again, m_J in the top CR and signal region



Probing Higgs invisible decay & dark matter

arXiv:1309.4017





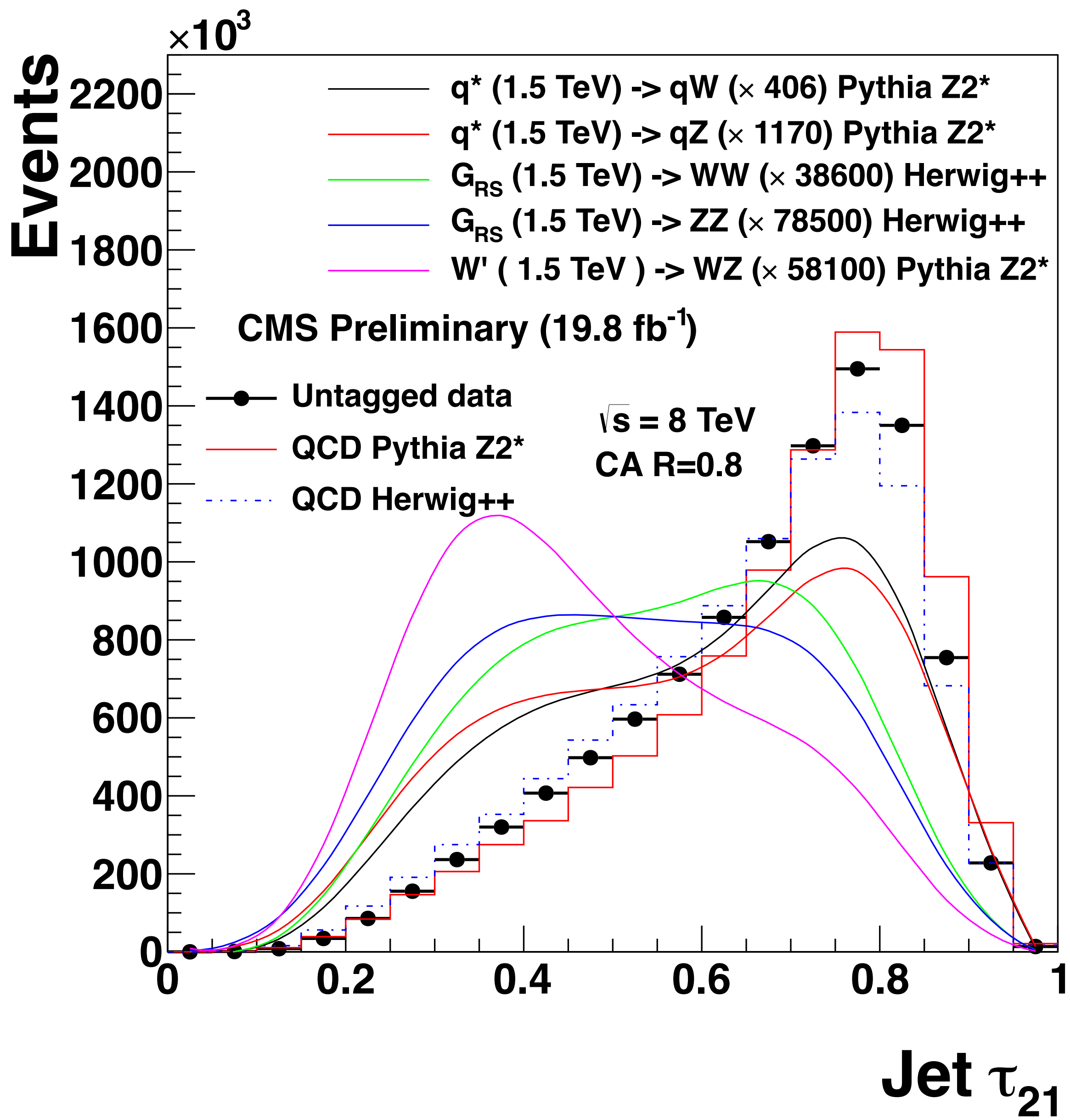
Started probing the Higgs invisible decay “directly” in the associated production mode.

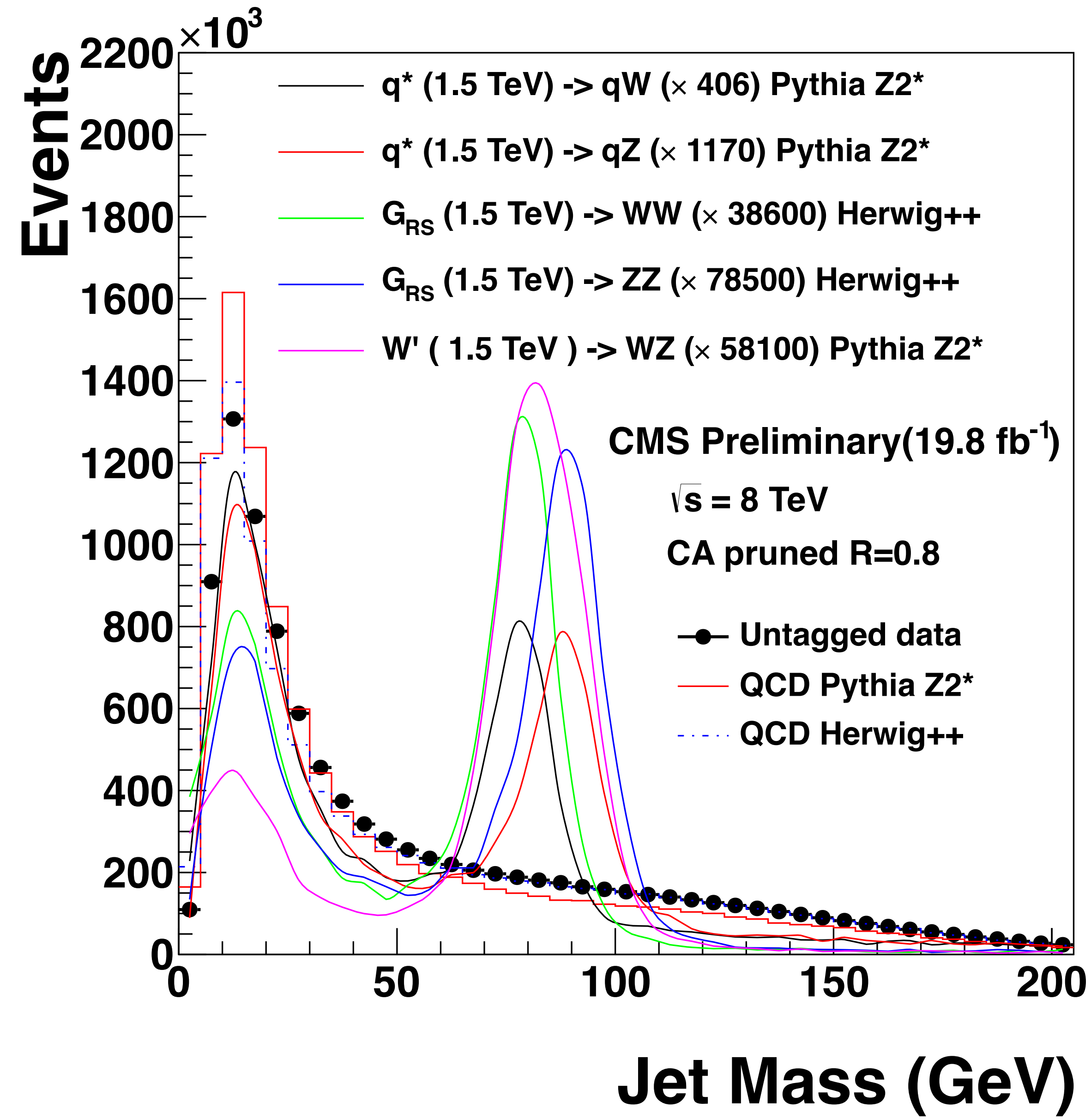
The best limit on WIMP DM mass 1−5 GeV. Comparable to mono-jet limits.

Just for completeness ....

CMS EXO-12-024

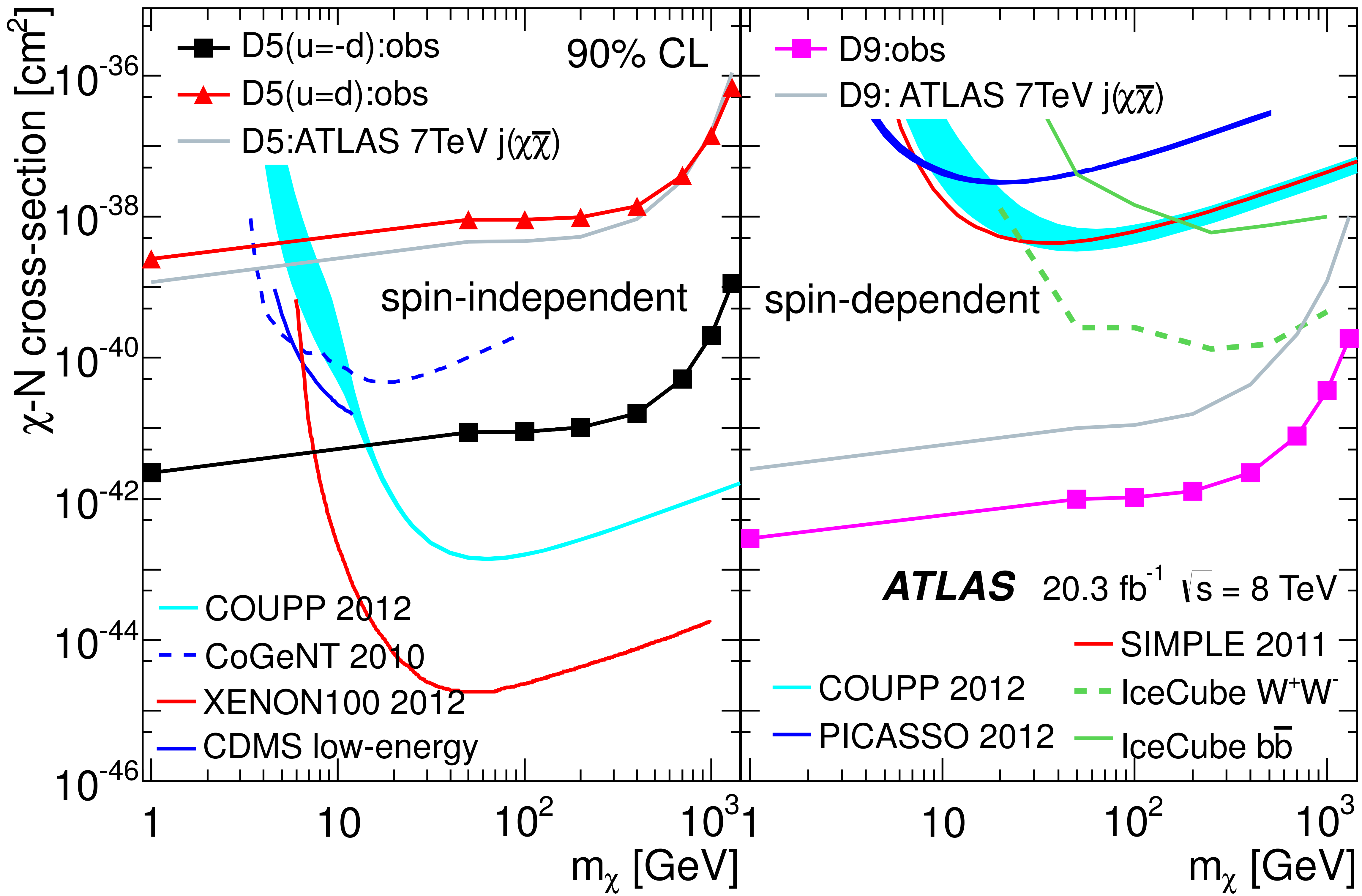
Events with 2 boosted jets. Hope to reco WW, WZ, ZZ.



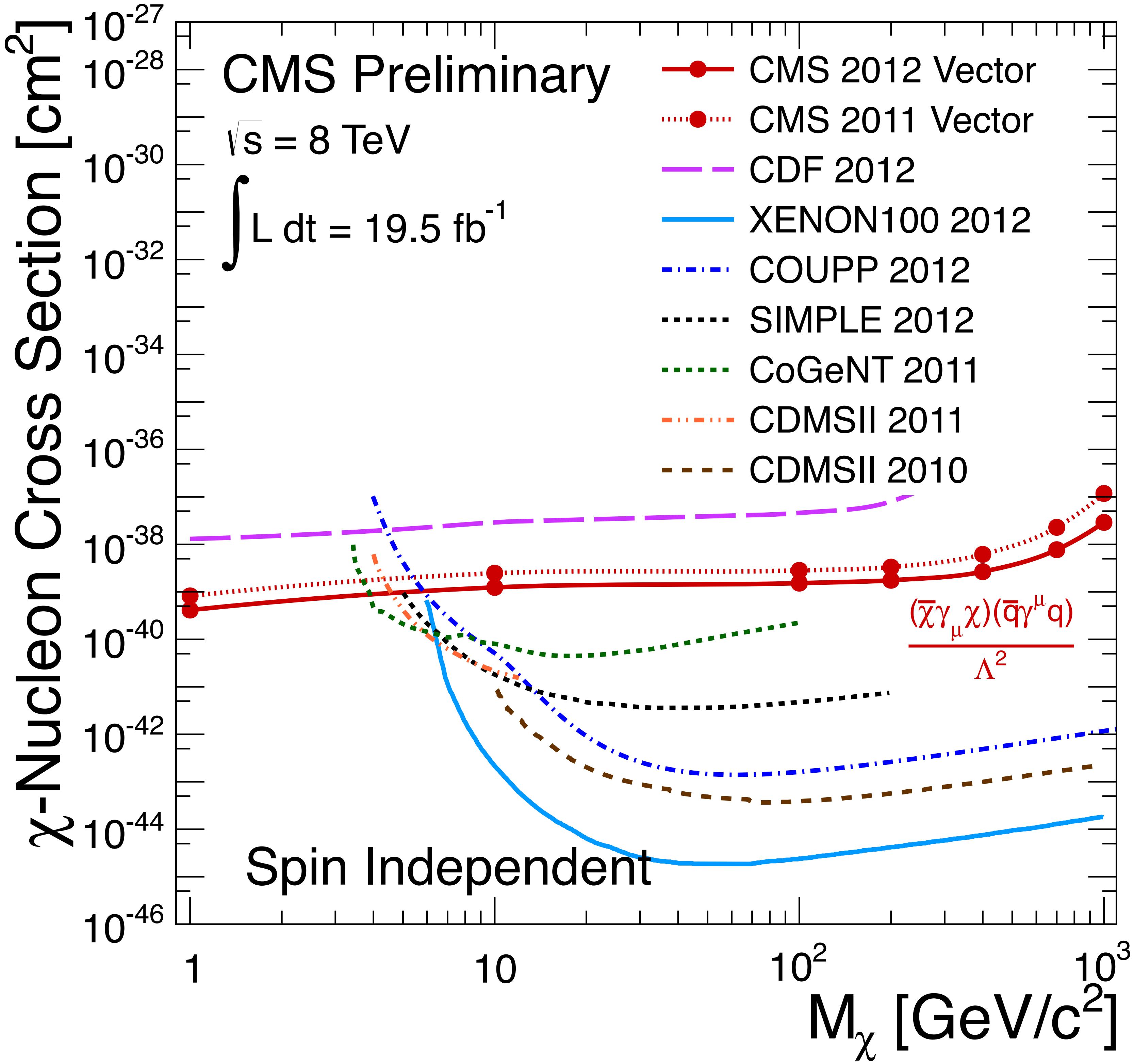


Swamped by QCD. Apply some W/Z tagging requirements on one/ both jets, then set limits on BSM using dijet invariant mass.

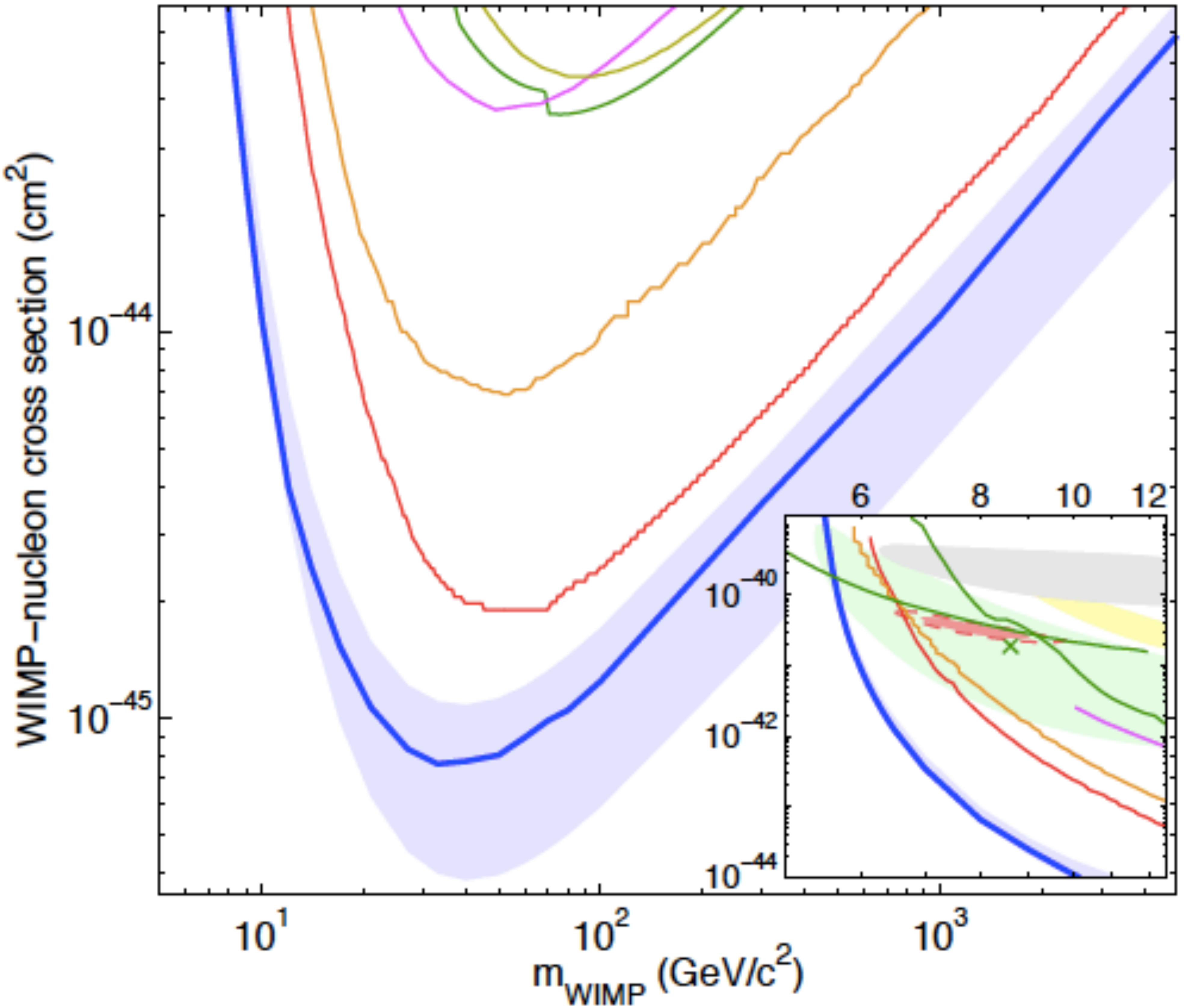
Best limits on low mass WIMP DM particle



Best limits on low mass WIMP DM particle



Latest result from LUX



additional

analysis details

Both leptonic and hadronic W p_T > 200 GeV

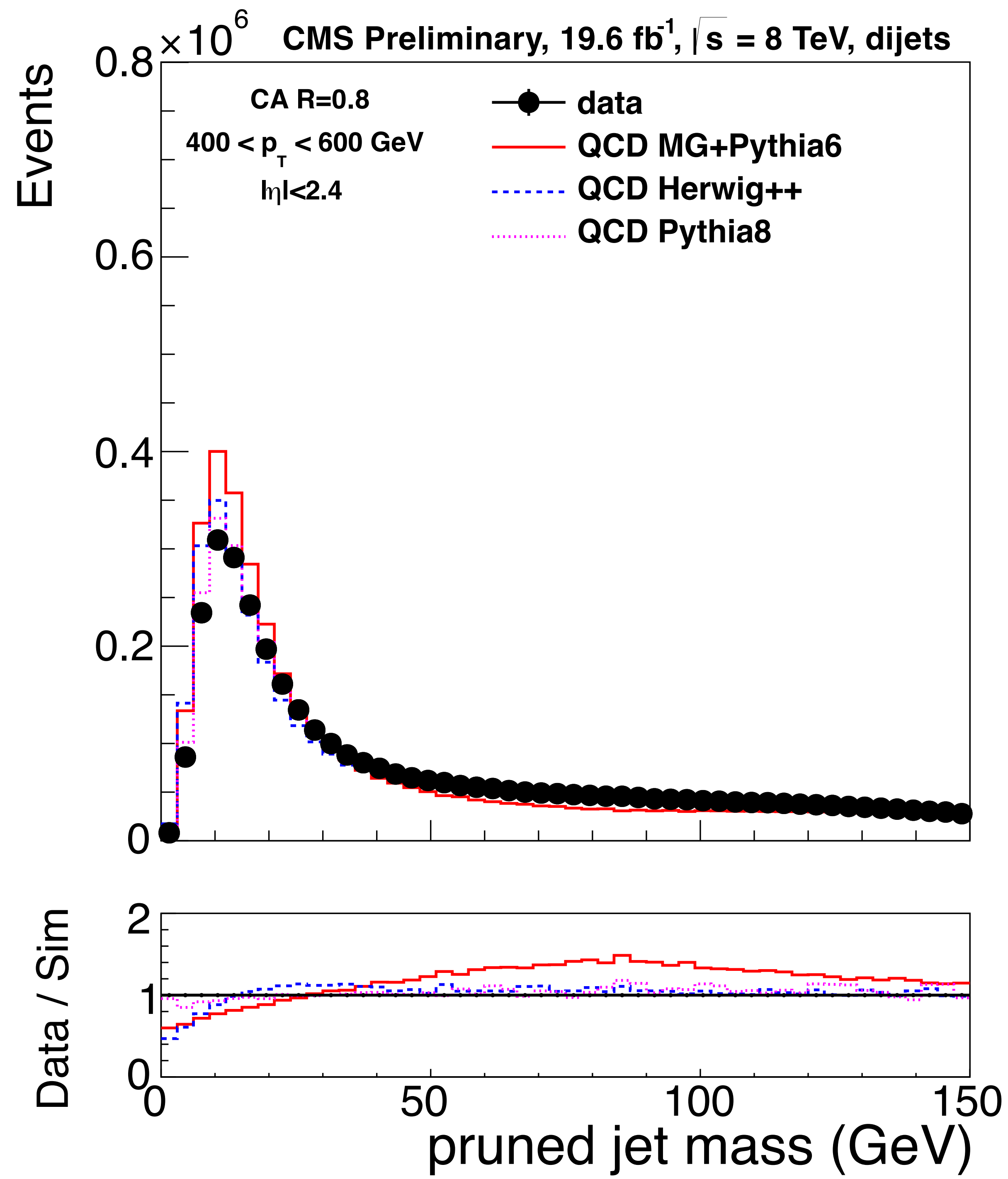
Exactly one lepton with p_T > 30/35 GeV for μ/e
Missing E_T > 50/70 GeV
At least one CA8 jet with p_T > 200 GeV
Veto presence of any b-tagged AK5 jets in the event

Additional cleaning cuts

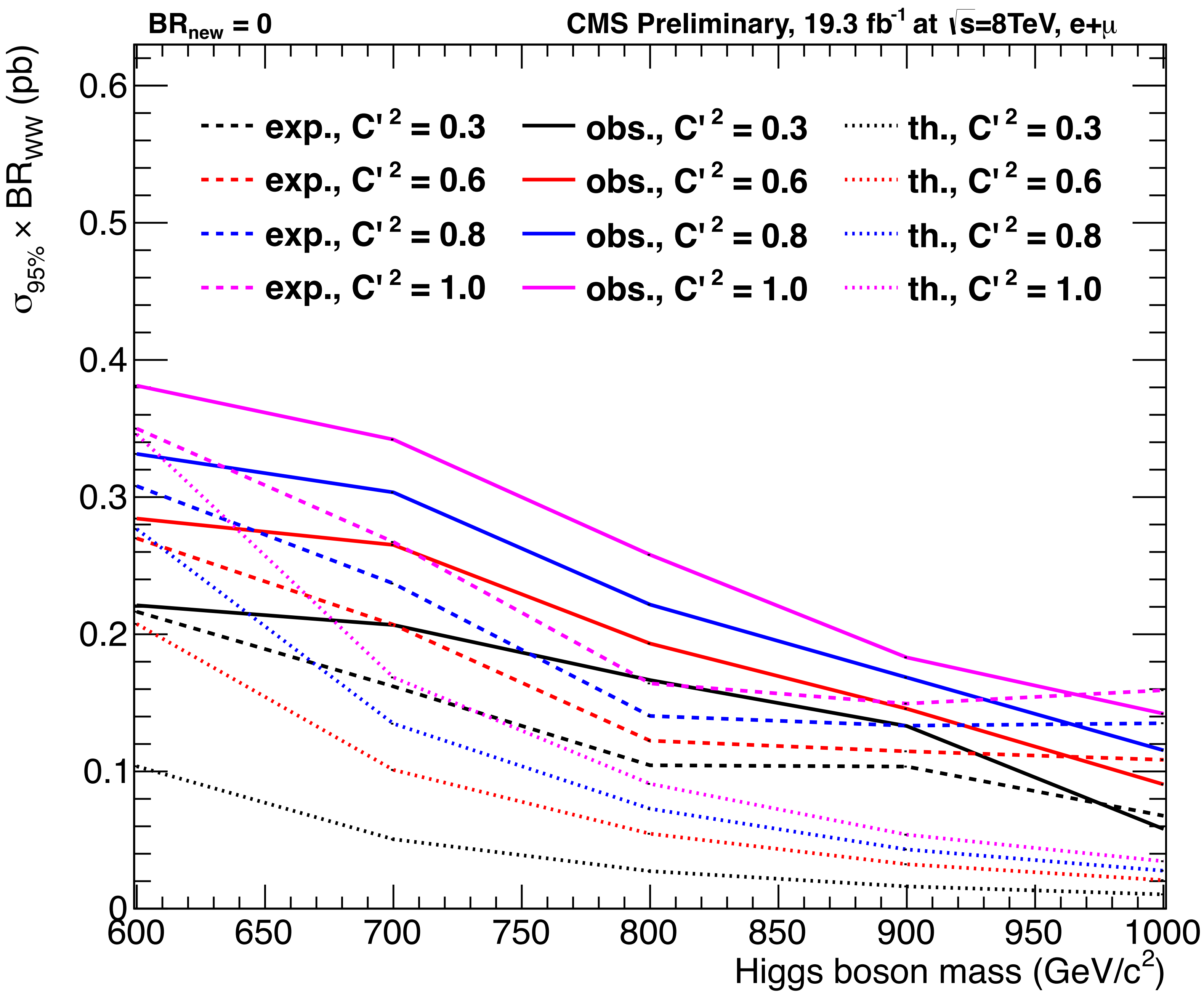
- $\Delta R_{l,j} > \pi/2$; distance between the lepton and the jet
- $\Delta\Phi_{\not{E}_T,j} > 2.0$; azimuthal separation between jet & MET
- $\Delta\Phi_{V,j} > 2.0$; azimuthal separation between the two Ws

background simulation
W+jets: MadGraph + Pythia6, Herwig++
ttbar: Powheg+Pythia6, MadGraph+Pythia6, MC@NLO+Herwig++
single top: Powheg+Pythia6
WW/WZ/ZZ: Pythia6

jet mass, dijets



A more generic BSM interpretation



In the BSM interpretation, search for an electroweak singlet scalar, where a heavy Higgs boson mixes with H(126)

$$C^2 + C'^2 = 1$$

The heavy Higgs cross cross-section and width are modified as

$$\mu' = C'^2(1 - \mathcal{BR}_{new})$$

$$\Gamma' = \Gamma_{SM} \times \frac{C'^2}{(1 - \mathcal{BR}_{new})}$$

W-tagging

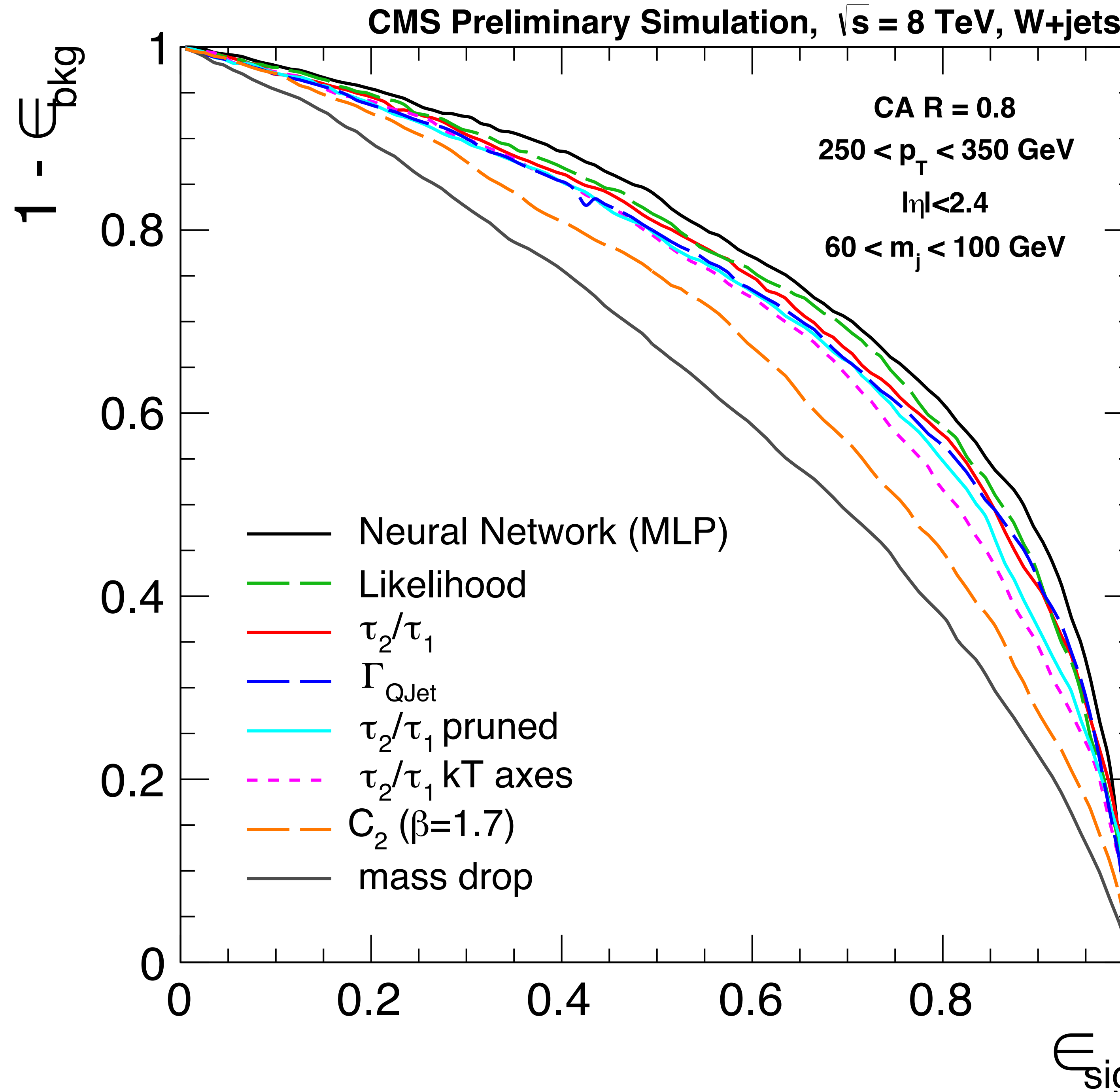
After pruned mass cut...

Compare the performance of several different observables

- Qjets and N-subjettiness are most powerful single variables

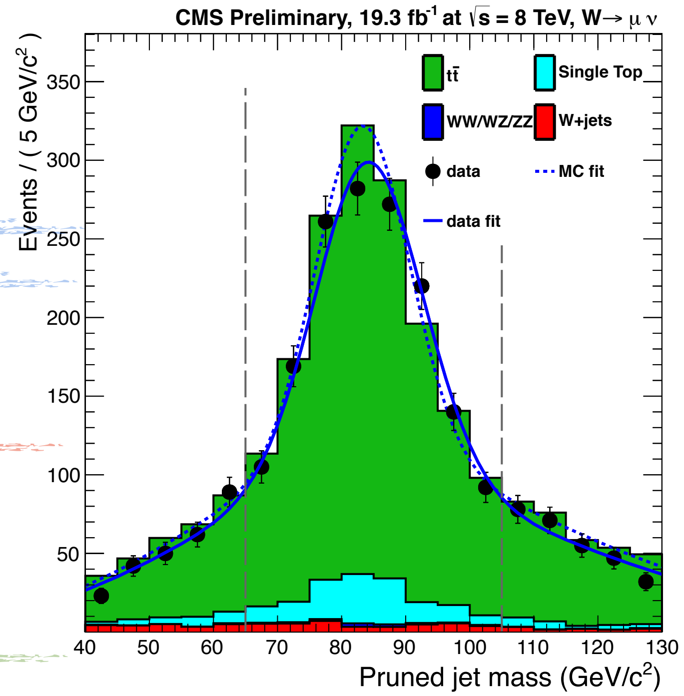
Try combining observables into an MVA discriminant; a small improvement

CMS baseline tagger is pruned jet mass + τ_2/τ_1 cut



more on systematics

Syst. uncertainty	sig, ggH	sig, VBF	W+jets	$t\bar{t}$	single t	WW/WZ
lumi	4.4%	4.4%	-	4.4%	4.4%	4.4%
Higgs QCD scale	6.5% †	1.3% †	-	-	-	-
Higgs PDF+α_s	12.1% †	5.9% †	-	-	-	-
Intf (sig/bkg)	10.0%	50.0%	-	-	-	-
Bkg cross-section	-	-	-	-	30.0%	30.0%
W+jets norm.	-	-	8%	-	-	-
W-tagging	10.0%	10.0%	-	-	-	10.0%
$t\bar{t}$ norm.	-	-	-	6.0%	6.0%	-
Jet mass/energy scale	2%	2%	-	2%	2%	2%
W+jets shape	-	-	see Sec. 6	-	-	-
b-tagging	2.5%	2.5%	-	-	2.5%	2.5%
Trigger (e & μ)	1%	1%	-	-	1%	1%
Selection Eff. (e & μ)	2%	2%	-	-	2%	2%

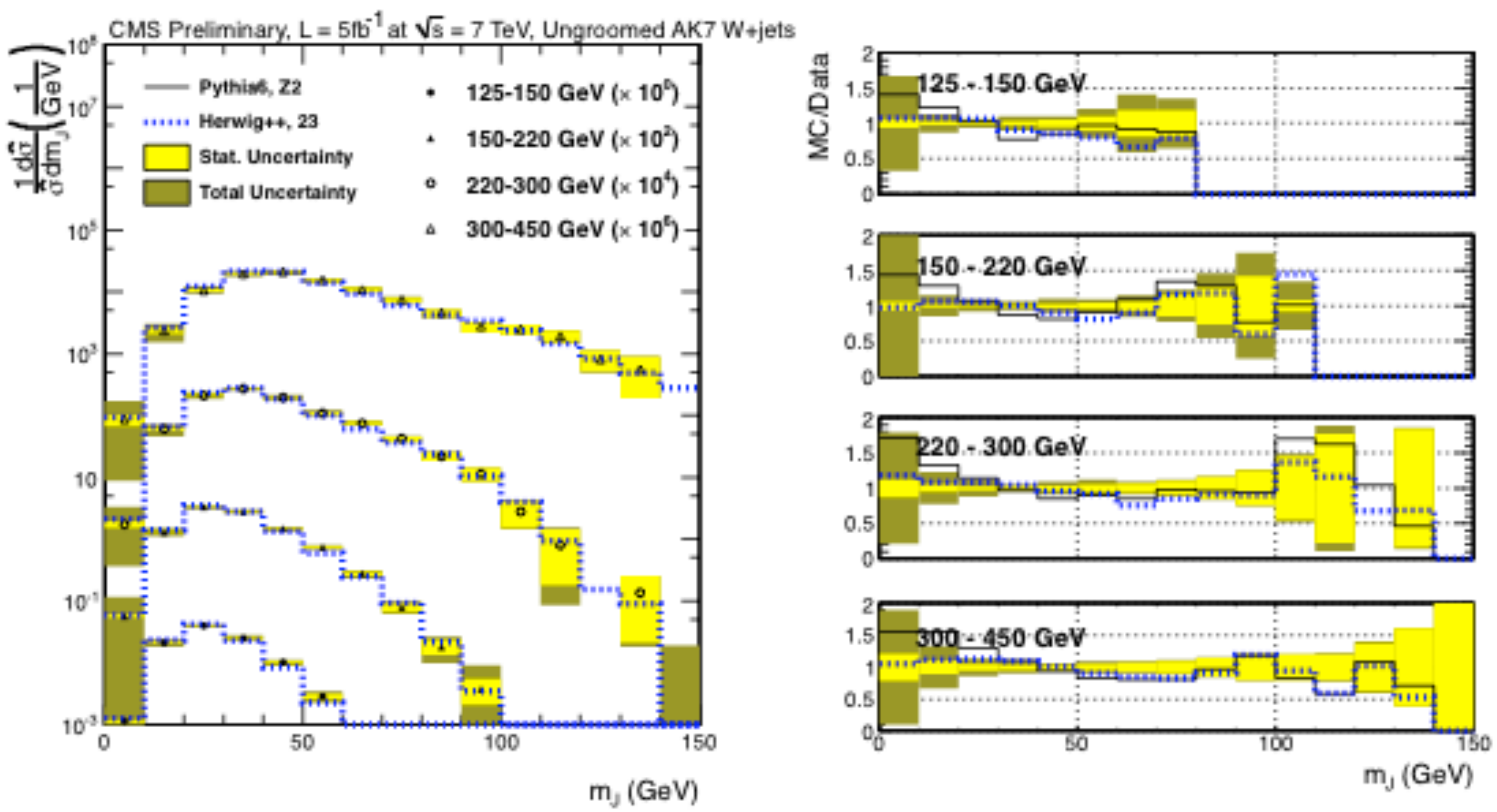


data:
<m> = 84.5 ± 0.4 GeV
σ = 8.7 ± 0.6 GeV
MC:
<m> = 83.4 ± 0.4 GeV
σ = 7.4 ± 0.4 GeV

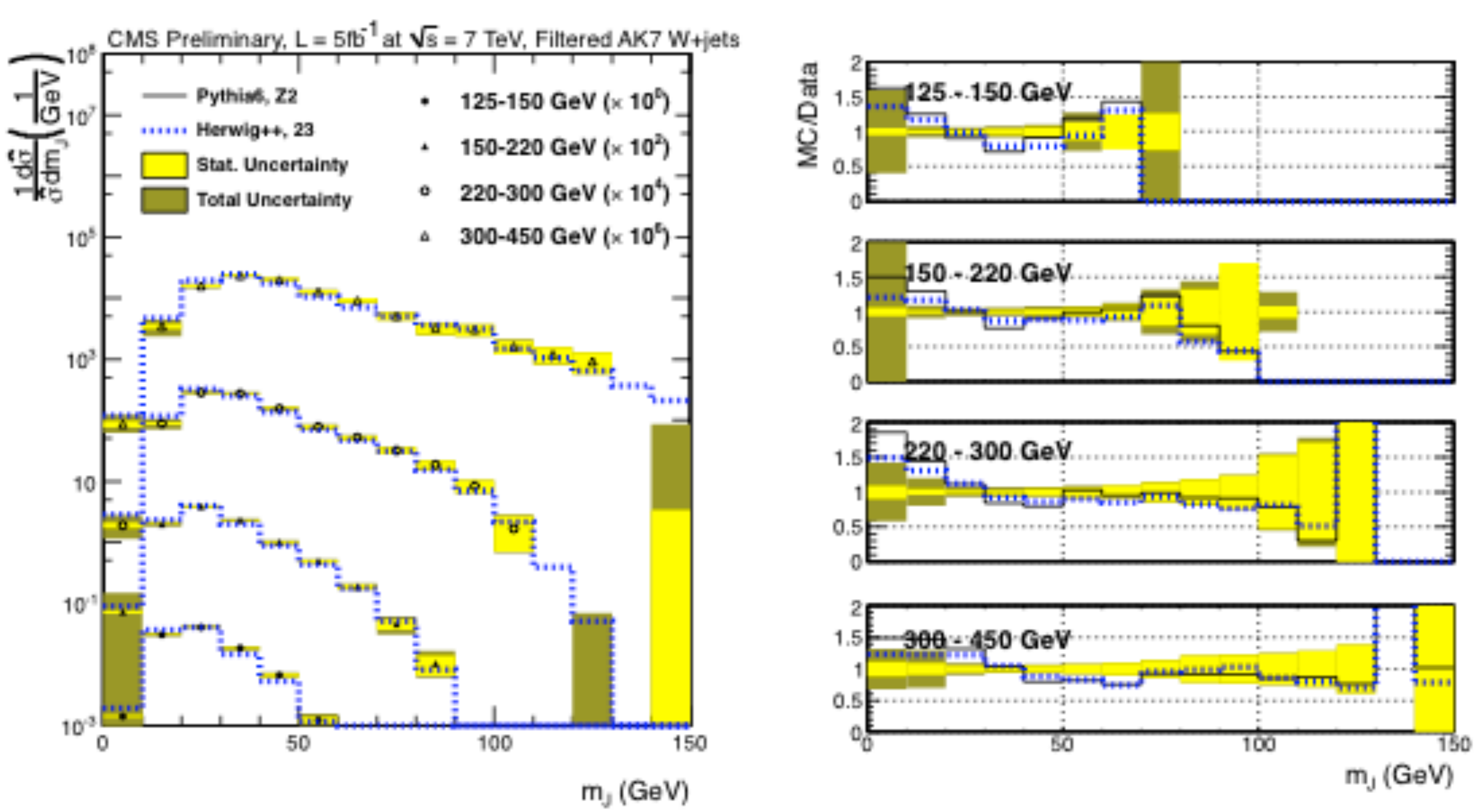
- W+jets shape is one of the larger systematics effects
- W-tagging scale factor estimated in the top-enriched control region to be 0.95 ± 0.10 (0.89 ± 0.10) for the μ (e) channel
- Signal uncertainties are dominated by theoretical uncertainties
  - PDF and α_s and interference effects -- standard within the LHCXSWG

Unfolded distributions, W+jet (AK7)

AK7, UNGR

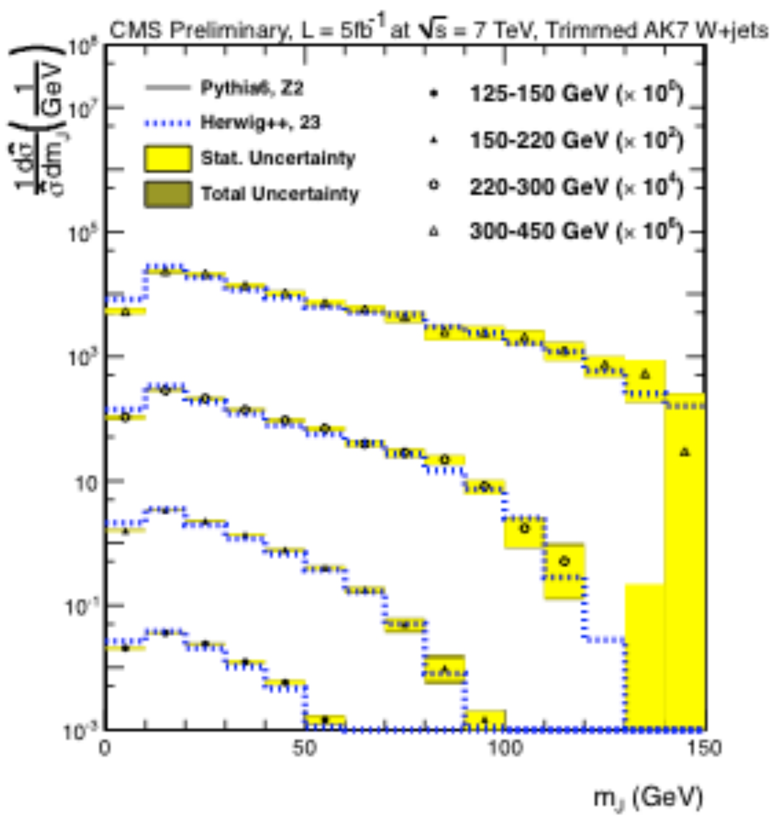


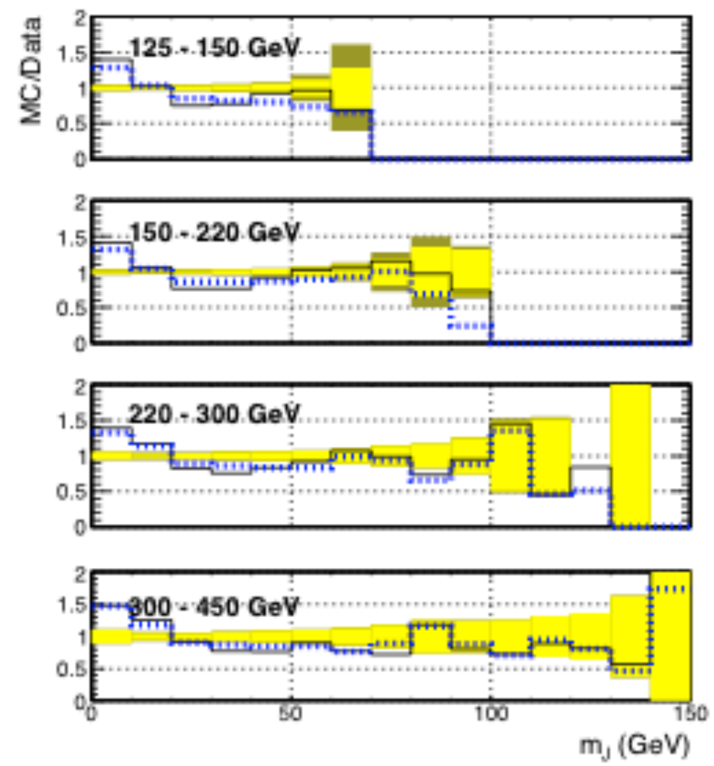
AK7, Filtered



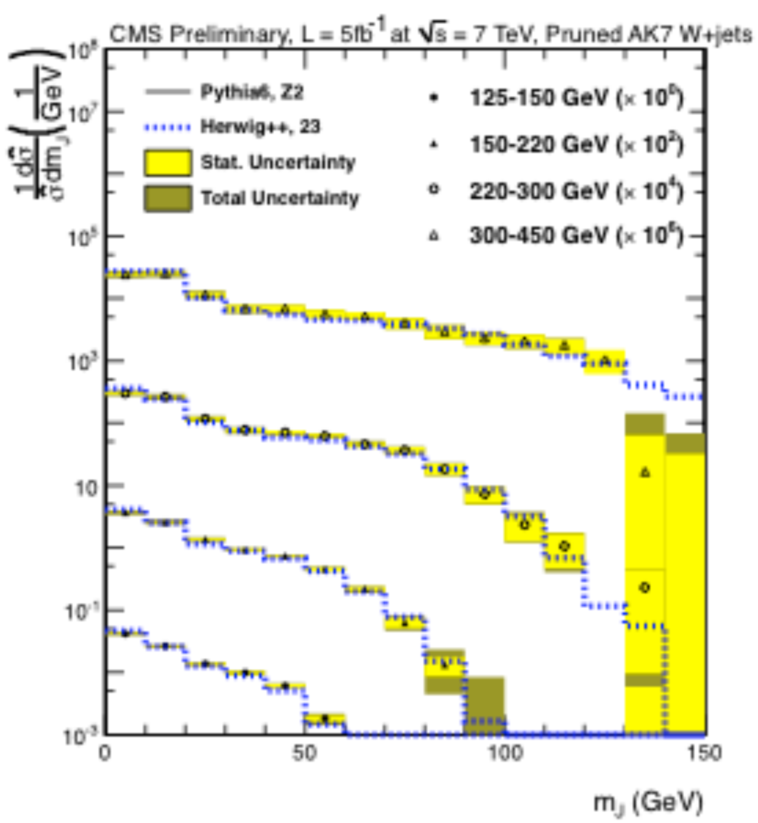
Unfolded distributions, W+jet (AK7)

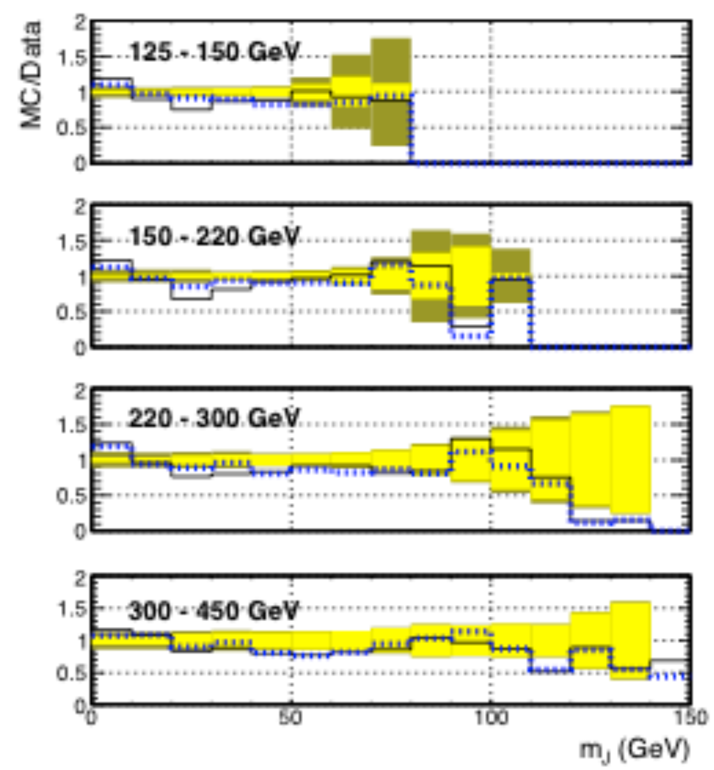
AK7, Trimmed



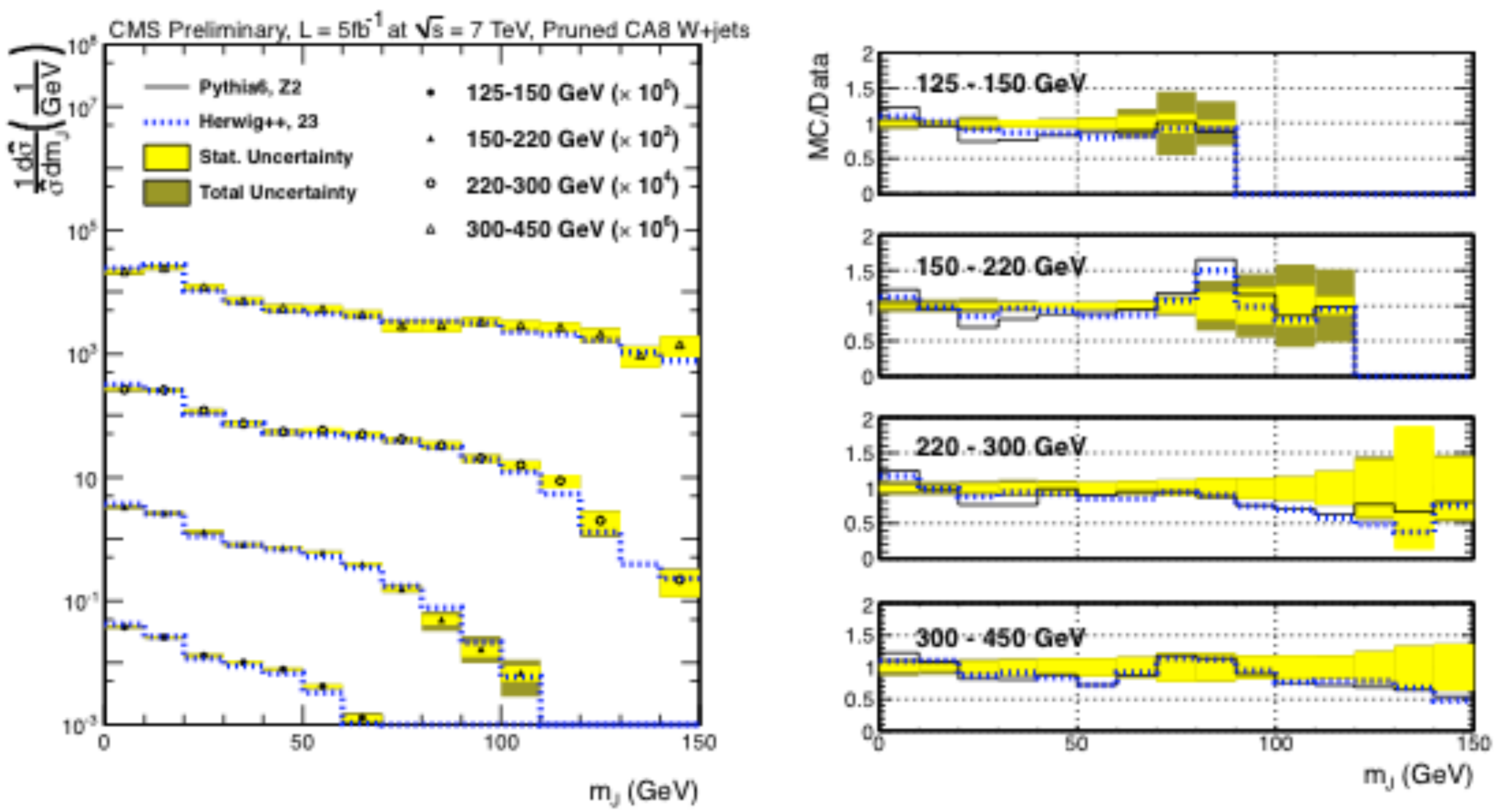


AK7, Pruned

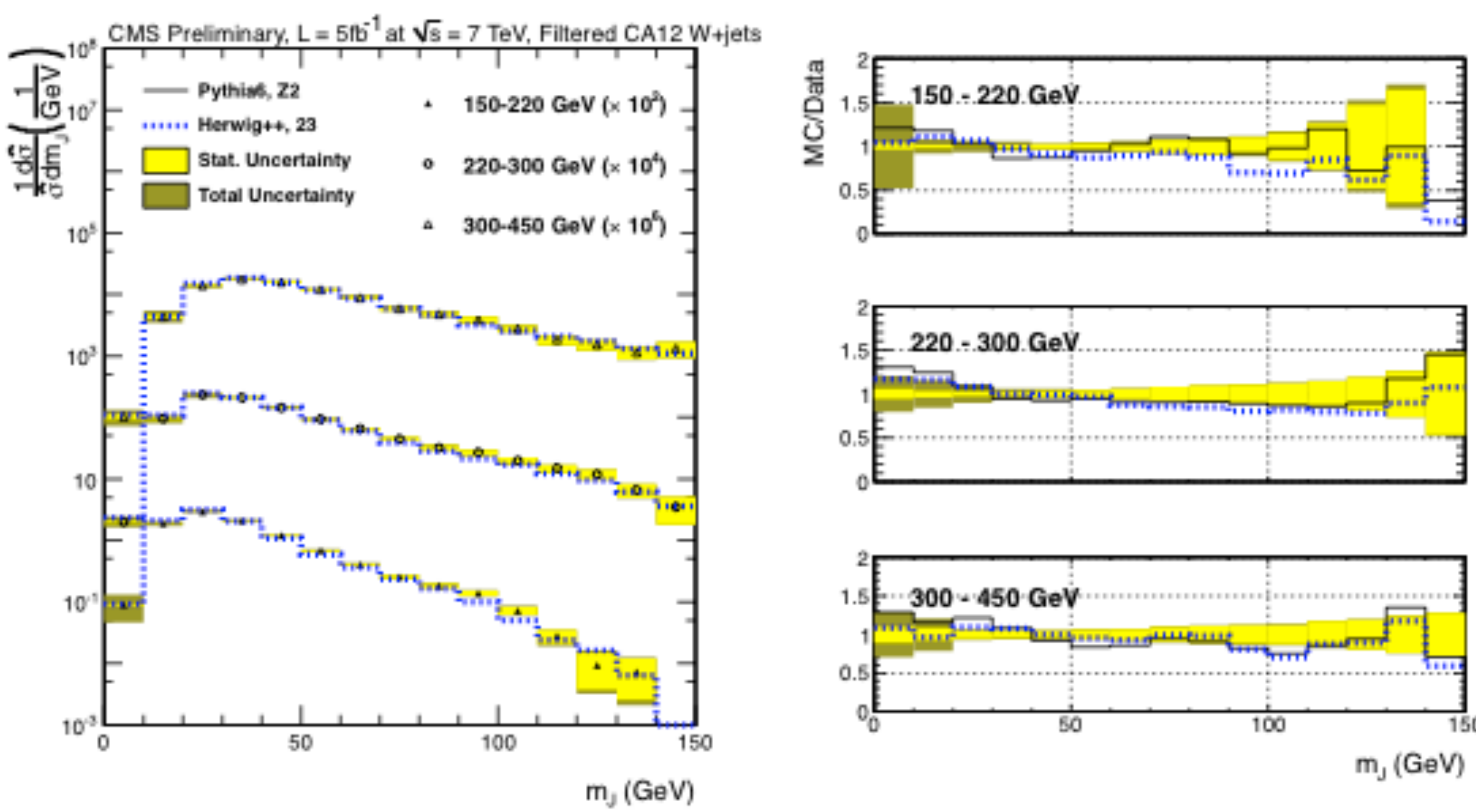




Unfolded distributions, W+jet: comparison



CA8, Pruned



CA12, Filtered