H→WW→lνjj analysis: update using 4.7 fb^{-1} data

Nural Akchurin[1], Jake Anderson[2], Pushpa Bhat[2], Jeffrey Berryhill[2], Phil Dudero[1], Ricardo Eusebi[3], Dan Green[2], Pratima Jindal[4], Sung-Won Lee[1], Kalanand Mishra[2], Ilya Osipenkov[3], Alexx Perloff[3], Andre Sznajder[5], Nhan V. Tran[2], Fan Yang[2], and Francisco Yumiceva[2]

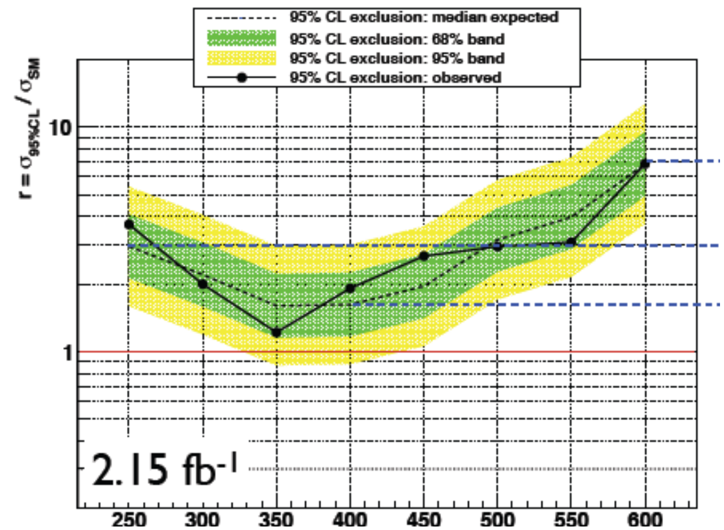
[1] Texas Tech University, Lubbock, Texas, USA
[2] Fermi National Accelerator Laboratory, Batavia, Illinois, USA
[3] Texas A&M University, College Station, Texas, USA
[4] University of Nebraska at Lincoln, Nebraska, USA
[5] Universidade do Estado do Rio de Janeiro (UERJ), Brazil

On behalf of H→WW (lνjj) working group

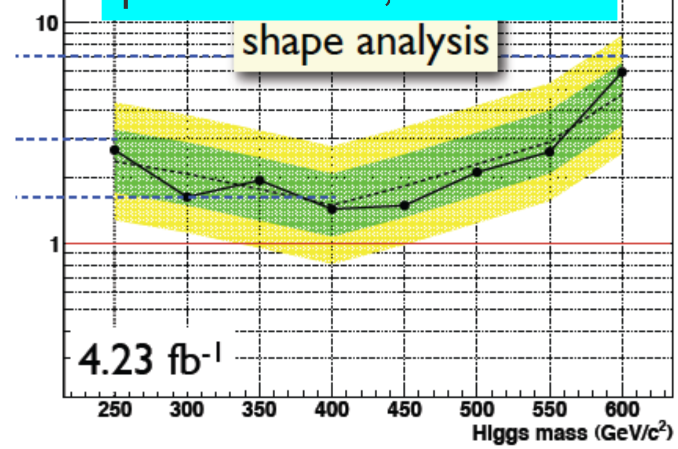
(January 12, 2012)

Recall where we were in December

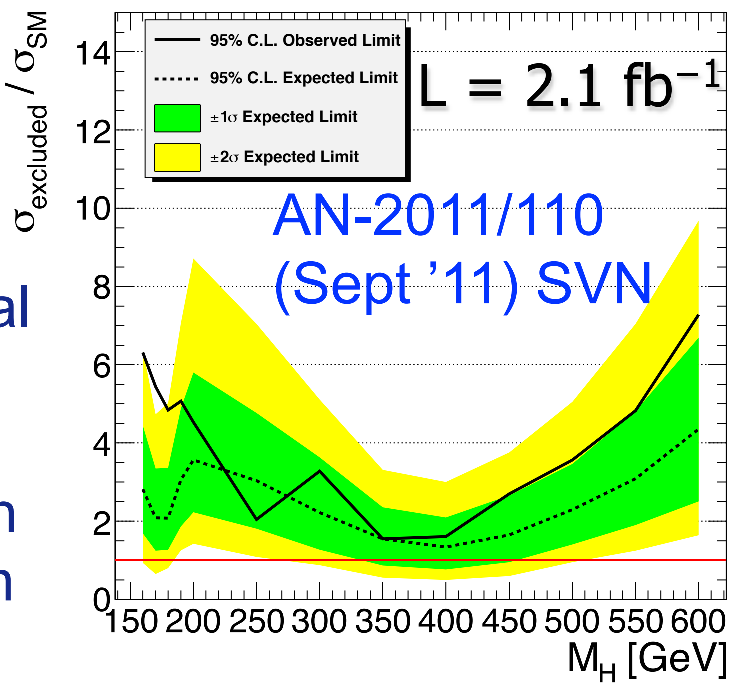
Using pre-selection cuts we were dominated by syst error ➜ exclusion limit from full 4.7 fb^{-1} data ≈ same as 2.1 fb^{-1}



From Andrea Benaglia presentation, Dec 14



For technical details see Fan Yang's presentation in this forum on Nov 17.



AN-2011/110 (Sept '11) SVN

Clearly we needed to improve S/B ...

Dan Green presented a detailed rationale, including several illustrated examples, during the last CMS week. His slides are at

https://indico.cern.ch/conferenceDisplay.py?confId=157054

Conclusion in December: Tightening the cuts would help, but individual cuts are highly inefficient .....

To improve the analysis qualitatively we have

- ☑ Optimized a simple likelihood (for each M_H) to reduce syst
- ☑ Finalized optimized cuts for each M_H, nJet, lepton category
- ☑ Computed trigger efficiency and data/MC efficiency SFs
- ☑ Recomputed systematics and exclusion limits

Optimization of likelihood

CMS AN AN-12-008

Multivariate optimization and background estimation for the Standard Model Higgs boson search in H→WW→ $\ell\nu jj$ decay

Apply cut to the likelihood output. Use m_{WW} to set limit.

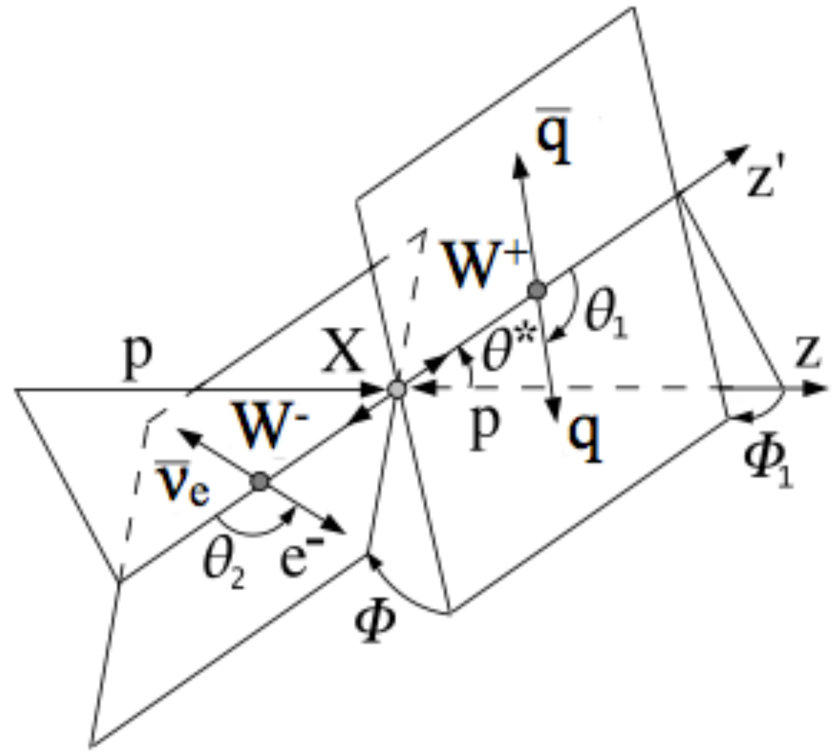
Abstract

We present a study of the multi-variate optimization for the Higgs boson search in the H →WW→ $\ell\nu jj$ final state in the gluon fusion production mechanism. Using a complete set of mostly un-correlated variables we optimize each Higgs mass points separately to distinguish between the Higgs signal and the dominant W+jets background. We are able to achieve S/B ~ 7% at the most optimal mass value M_H = 400 GeV in 2-jets events. This optimization corresponds to Higgs signal efficiency of about 32% and background rejection rate of about 95%. We also develop a data-driven technique to derive 4-body $m_{\ell\nu jj}$ invariant mass shape for W+jets background. For this we use events in the upper and lower sidebands of the m_{jj} distribution to extract the 4-body mass shape for events in the signal region 65 GeV $< m_{jj} <$ 95 GeV. Using this data-driven technique we are able to keep the systematic uncertainties (dominated by background shape systematics) to within 4%. This analysis note complements the documentation of the main analysis in AN-11/110.

Understanding event kinematics

- ✦Use full kinematics of event (matrix element) to discriminate between signal and background

- ✦Kinematics [at LO] described by
  $\{m_{WW}, m_{jj}, \theta_1, \theta_2, \theta^*, \Phi, \Phi_2\}$
  - $m_{l\nu}$ fixed by kinematic fit, one DOF lost ($p_{Z,\nu} \rightarrow m_{l\nu}$)



θ_1, θ_2, Φ: helicity angles
θ^*, Φ_2: production angles
θ_2 range = $[0,\pi]$ (cannot distinguish q-qbar)

There are five angular degrees of freedom

Use a complete set of observables to optimize

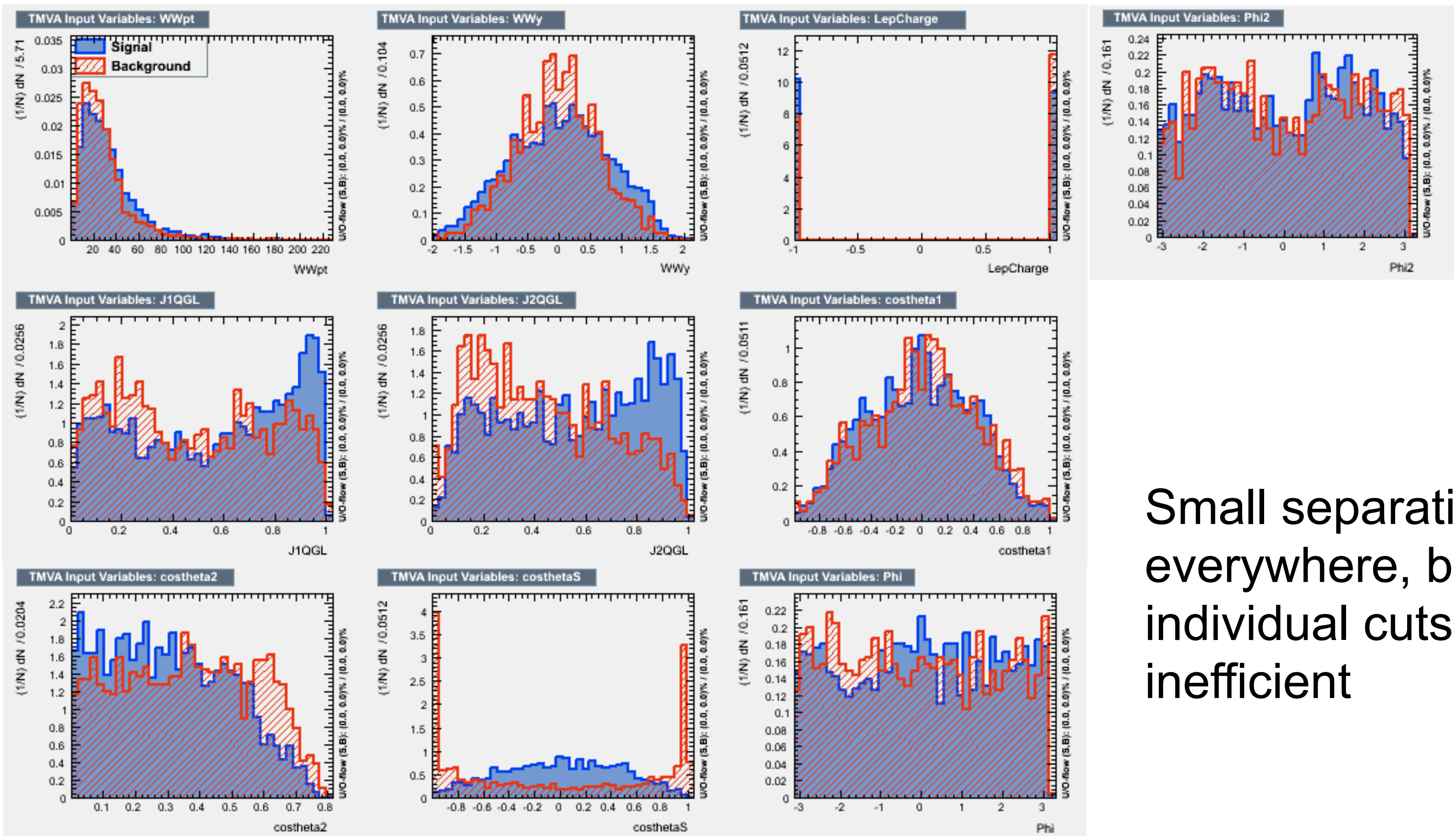
- Set of observables
  - $\{\theta_1, \theta_2, \theta^*, \Phi, \Phi_2\}$ + {qg (j1), qg (j2)} + $\{pT_{WW}, y_{WW}\}$ + {lepton charge}

mostly uncorrelated

- Exclude m_{WW}, m_{jj} from set of observables
  - m_{WW} for limit computation
  - m_{jj} for sideband extrapolation
- Separate MVA output in 4 categories for each mass point: 170-200 GeV (every 10), 250-600 GeV (every 50)
  - $WW \to \mu\nu jj$ in 2- and 3-jet category
  - $WW \to e\nu jj$ in 2- and 3-jet category

Inputs, example

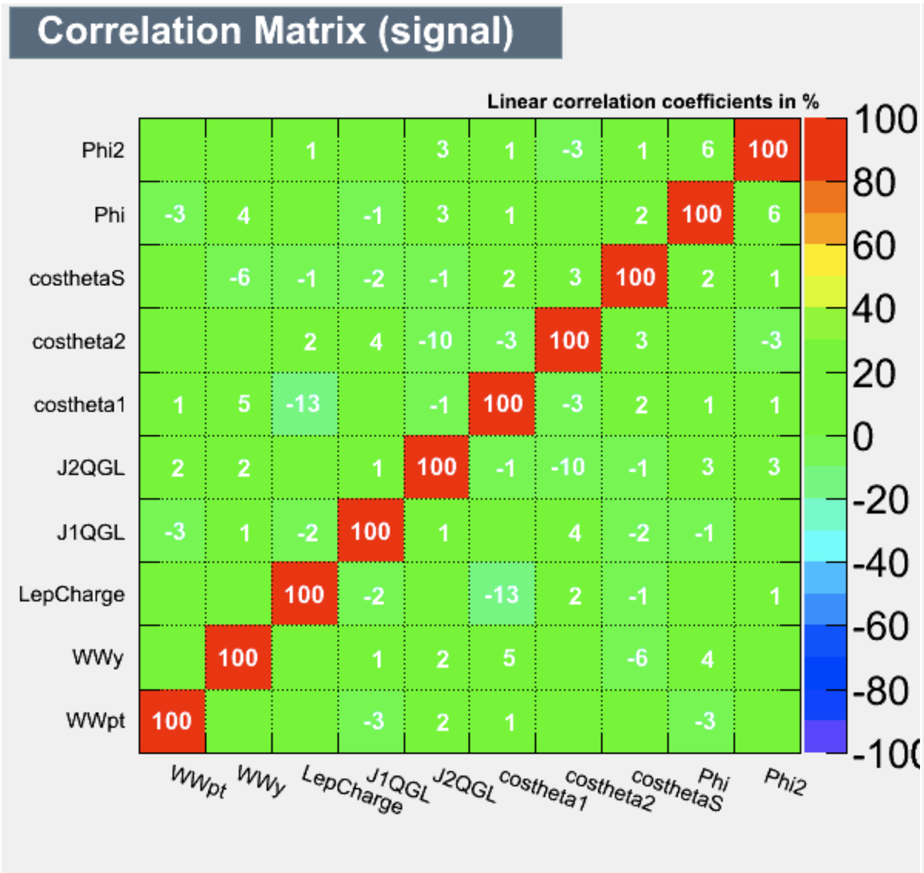
mH = 400 GeV: muon channel, 2-jets category

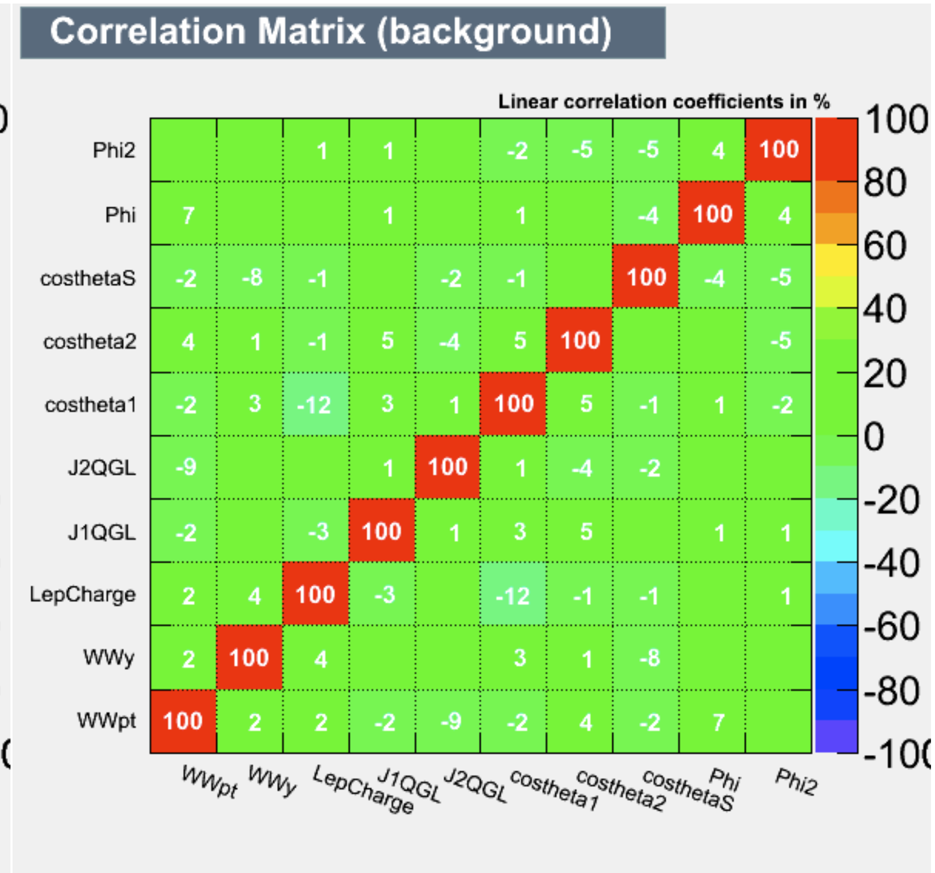


Small separation everywhere, but individual cuts inefficient

Correlations, example

mH = 400 GeV: muon channel, 2-jets category





Likelihood optimization, an example

Optimize the quantity:

$$\frac{S}{\sqrt{B + (\delta_{\text{sys-background}} \times B)^2}}$$

Muon, 2-jets and 3-jets categories

Mass Point	Win GeV	Cut	$\frac{S}{\sqrt{\cdots}}$	ϵ (%)	RJ-Wjets (%)	RJ-Tot (%)	N Sig	N Bkg	α C	α L	α H
2j170mu	[165, 180]	0.63	0.340	27.15	85.27	84.99	42.11	2802.47	0.13	0.10	0.15
2j180mu	[170, 200]	0.76	0.344	16.69	95.06	94.83	26.33	1627.43	0.00	0.00	0.04
2j190mu	[185, 210]	0.82	0.340	14.48	97.16	96.95	12.60	666.05	0.01	0.00	0.07
2j200mu	[190, 220]	0.82	0.317	21.27	95.47	95.03	18.37	1169.95	0.00	0.00	0.12
2j250mu	[226, 287]	0.87	0.442	18.30	97.14	96.59	18.06	756.28	0.05	0.00	0.13
2j300mu	[265, 347]	0.85	0.562	24.51	96.01	95.26	21.58	697.19	0.17	0.14	0.21
2j350mu	[308, 401]	0.87	0.925	26.59	96.08	95.19	24.65	423.20	0.15	0.11	0.20
2j400mu	[346, 457]	0.85	1.018	32.59	95.40	94.24	23.39	341.34	0.22	0.18	0.26
2j450mu	[381, 512]	0.85	0.841	33.17	95.79	94.39	14.96	230.95	0.21	0.16	0.25
2j500mu	[415, 568]	0.76	0.664	52.99	90.52	88.39	15.20	339.33	0.20	0.17	0.22
2j550mu	[440, 617]	0.89	0.529	26.80	98.29	97.13	4.53	66.31	0.12	0.06	0.17
2j600mu	[462, 663]	0.89	0.396	29.24	98.44	97.28	2.95	51.44	0.12	0.07	0.17
3j170mu	[165, 190]	0.59	0.305	33.85	84.00	82.96	8.12	423.62	0.24	0.19	0.29
3j180mu	[170, 210]	0.60	0.220	38.44	79.60	77.78	9.78	840.62	0.27	0.21	0.33
3j190mu	[185, 220]	0.59	0.217	41.04	83.33	80.82	7.32	588.08	0.29	0.22	0.37
3j200mu	[190, 230]	0.57	0.177	52.19	77.19	73.22	8.16	884.43	0.27	0.22	0.33
3j250mu	[216, 300]	0.75	0.243	40.61	89.25	84.03	10.11	773.20	0.39	0.30	0.48
3j300mu	[241, 355]	0.78	0.300	34.70	92.28	87.52	9.93	571.61	0.44	0.30	0.56
3j350mu	[269, 407]	0.79	0.412	36.71	91.58	86.22	12.69	519.09	0.15	0.05	0.28
3j400mu	[300, 465]	0.77	0.497	38.69	92.65	86.41	12.75	400.52	0.17	0.09	0.25
3j450mu	[332, 518]	0.76	0.421	42.99	92.62	86.32	9.05	309.79	0.42	0.29	0.54
3j500mu	[362, 569]	0.72	0.357	48.75	92.24	84.49	6.97	266.79	0.43	0.35	0.51
3j550mu	[398, 616]	0.68	0.245	55.76	88.94	80.87	4.43	236.92	0.24	0.15	0.35
3j600mu	[419, 660]	0.78	0.178	48.80	93.46	85.78	2.40	147.09	0.36	0.28	0.44

Details presented in H→WW →lνjj WG meeting last week by Fan Yang

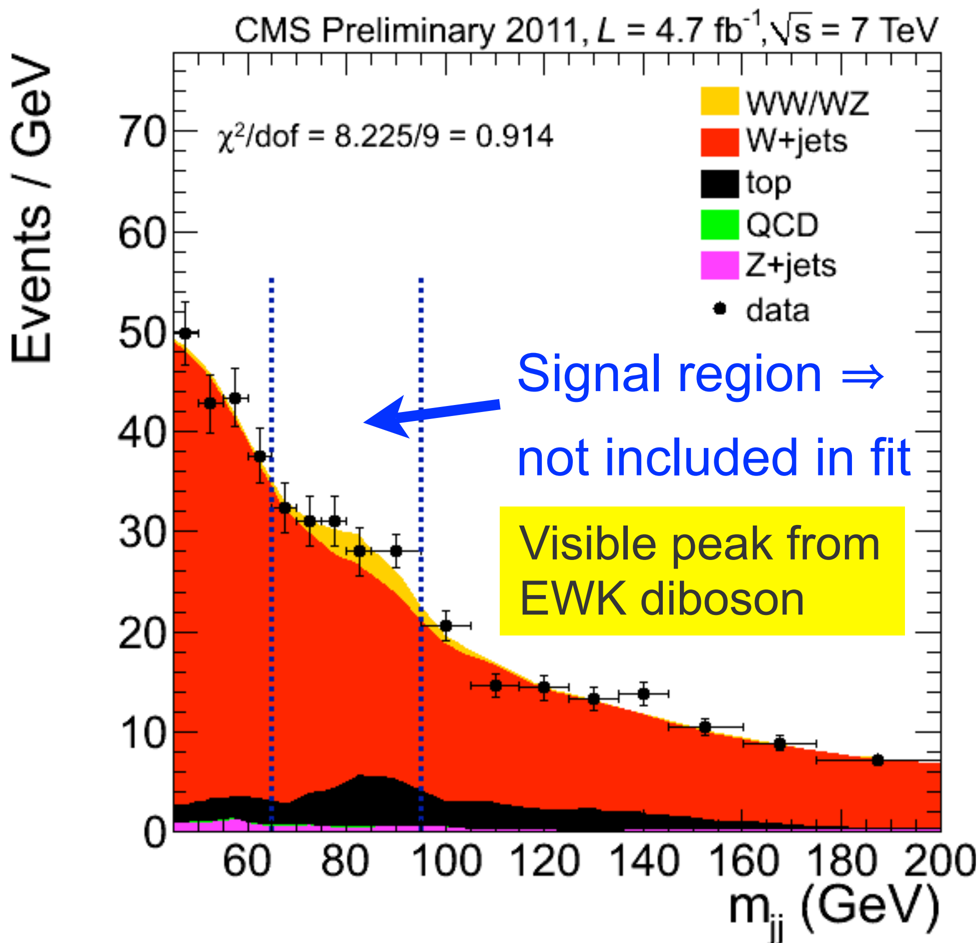
← **An example:**

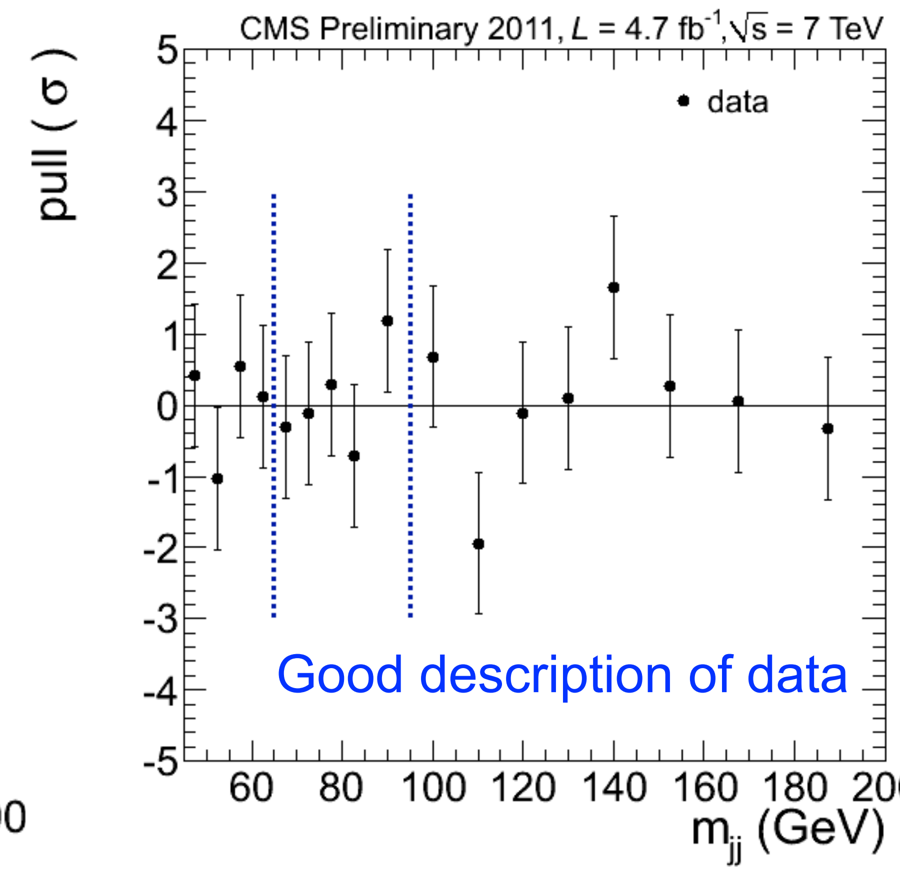
Higgs Signal = 23 ± 20% syst
Background = 341 ± 4% syst
Significance = 23 /√(341+ 14) = 1.2

Illustrates why you need a good control on the B error, why you need a larger S/B, and why you need to MEASURE the B error to understand the systematics.

Example: m_{jj} fit used to get bkg normalization

Muons, 2-jets category, M_H = 600 GeV

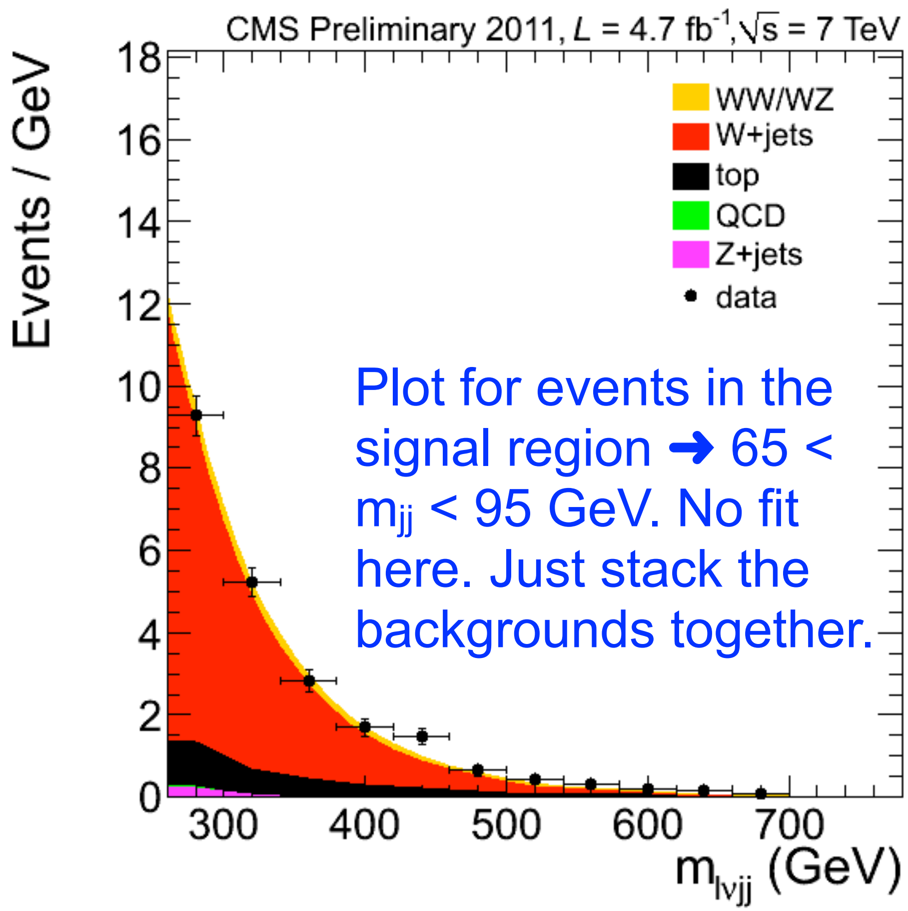


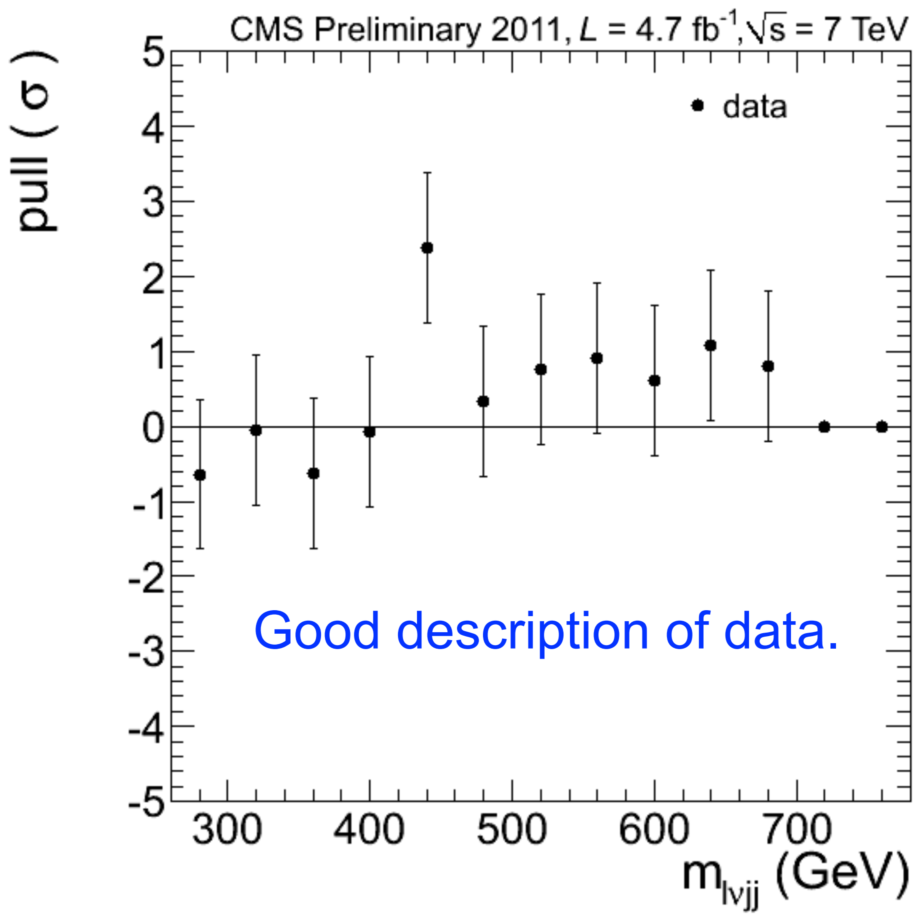


Efficiency corrections applied: see my presentation in H→WW (lvjj) working meeting on December 7.

Example: m_{WW} distribution used for limit setting

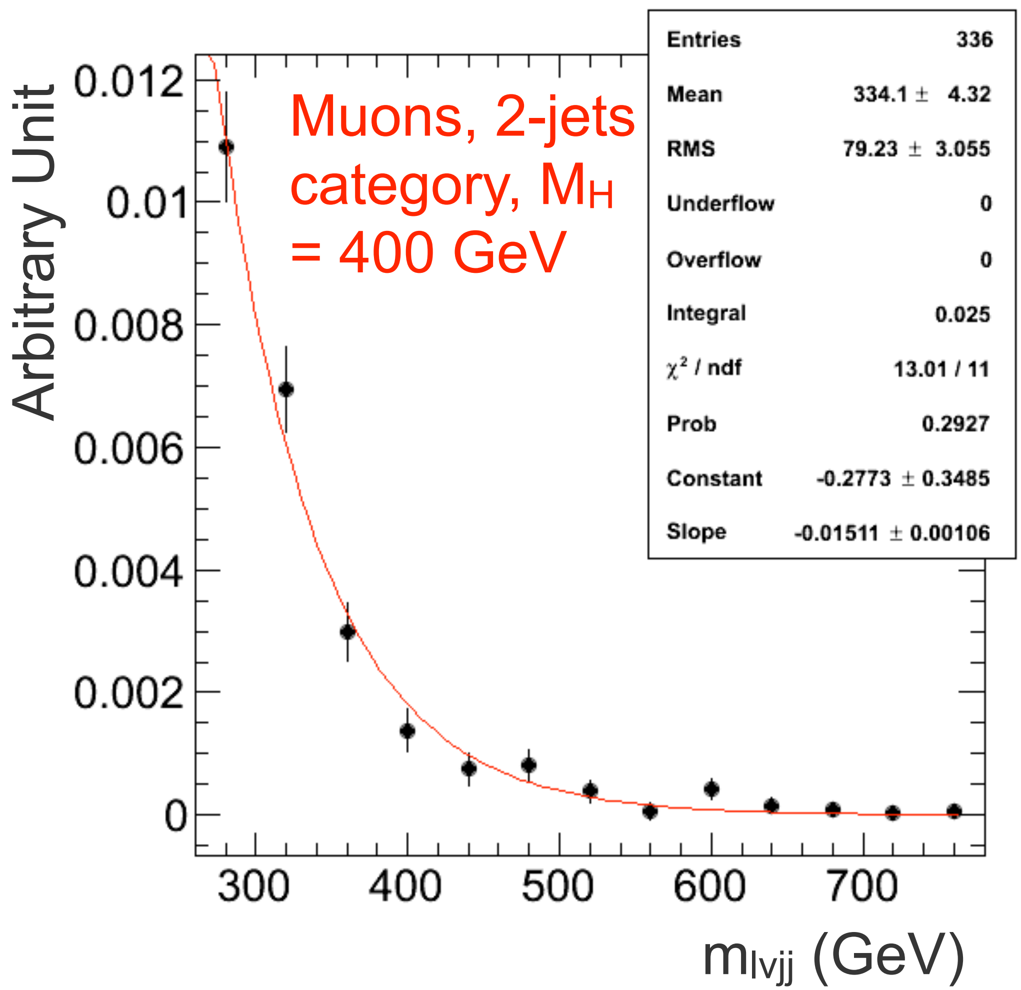
Muons, 2-jets category, M_H = 600 GeV





Normalization of each component came from m_{jj} fit of the previous slide. W+jets shape in the above plot is derived from data sidebands. Efficiency applied.

Improvement in W+jets shape from data



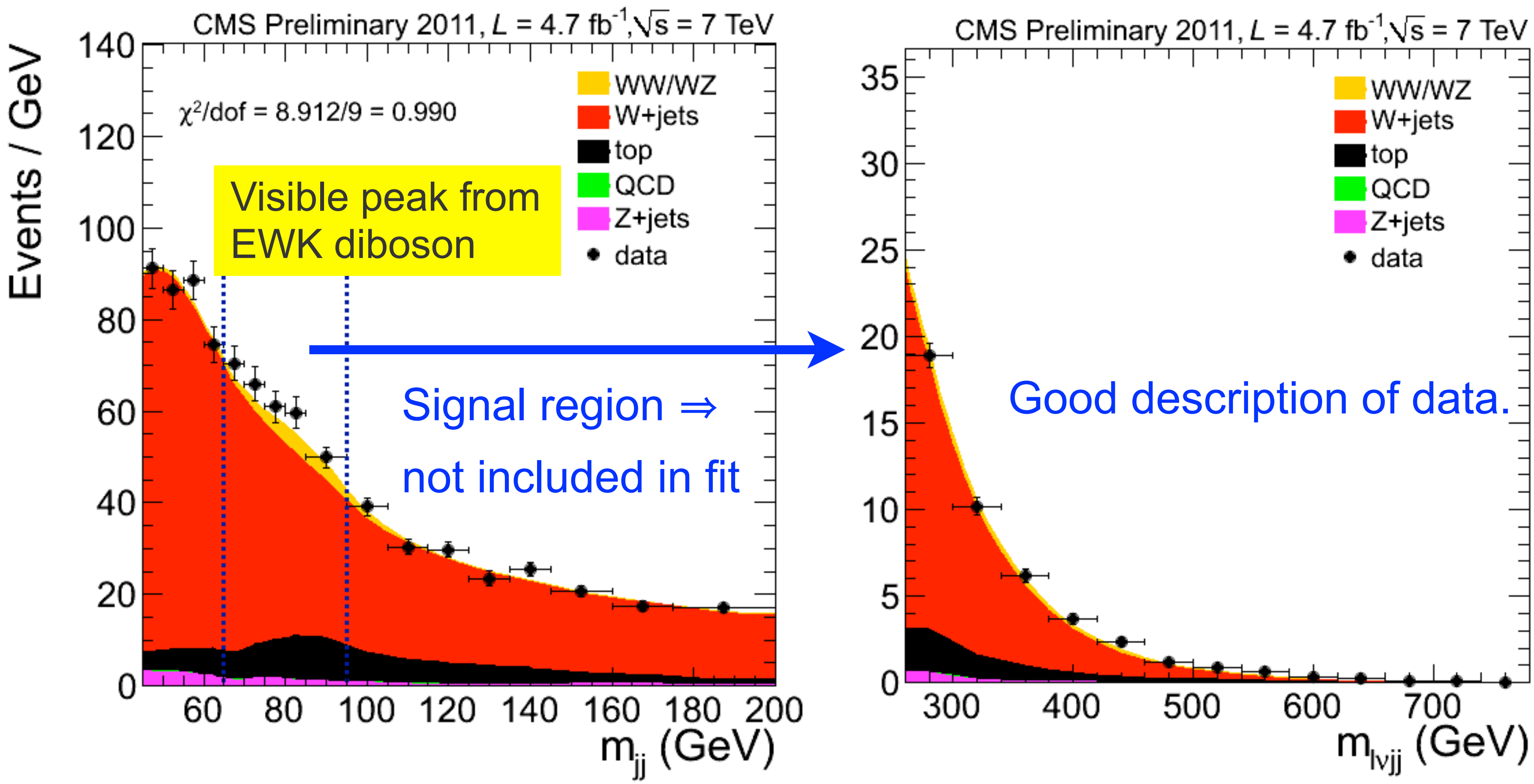
Still derive W+jets shape from data sidebands as before:

$$\alpha \,.\, SB_{Low} + (1-\alpha) \,.\, SB_{High}$$

But, now parametrize this shape (using an exponential) to get smooth curve. Uncertainty in the parameter determines the shape syst. Check coverage by assuming extreme values of α: [0, 1].

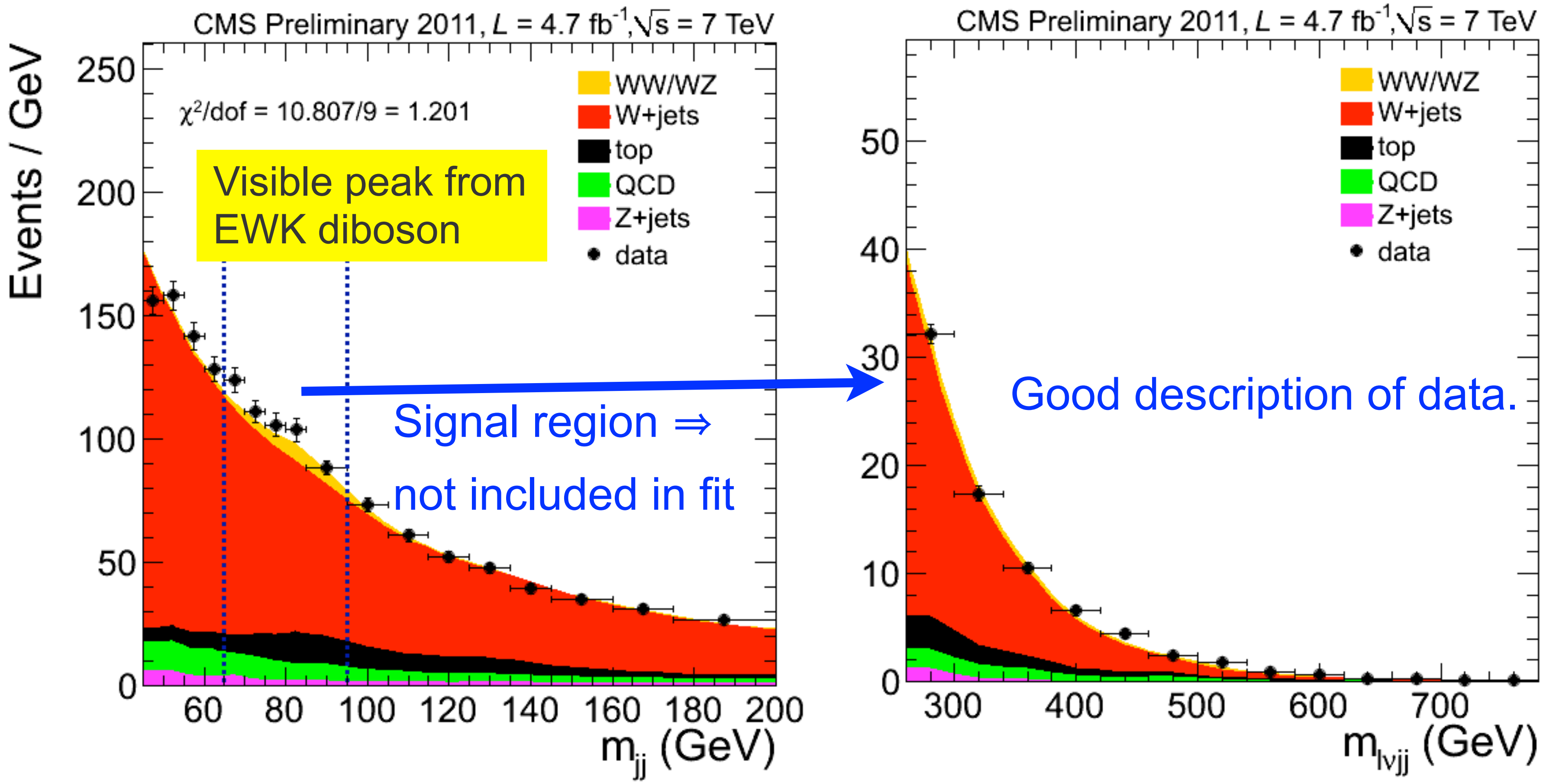
Second example: another mass point

Muons, 2-jets category, M_H = 400 GeV



Third example: different channel/ mass point

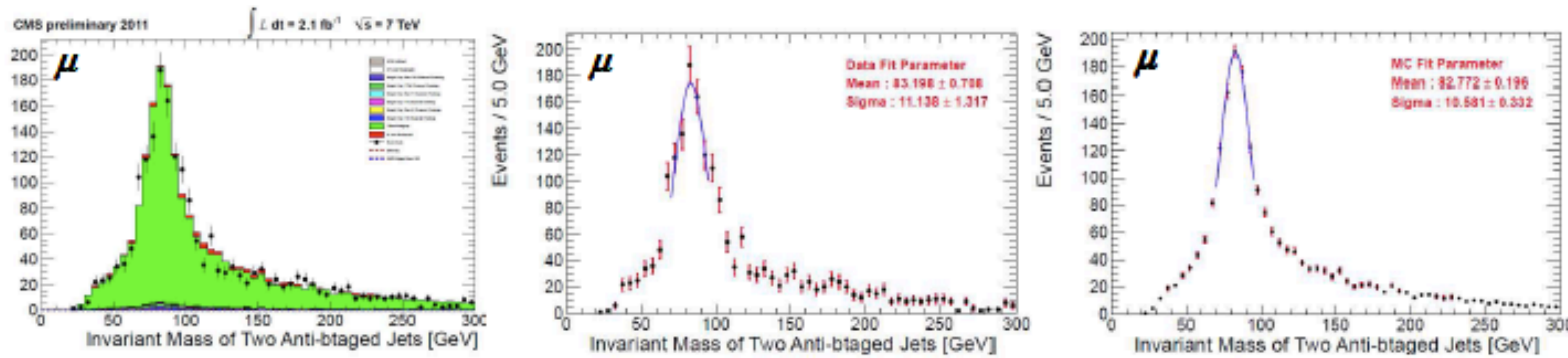
Electrons, 2-jets category, M_H = 500 GeV



An example of other systematics: JEC/JER

- **Compare to the (almost) pure top control sample:**
  - **Exactly four jets - two b-tagged and two anti-btagged**
  - **Use the anti-btagged jets to reconstruct the hadronic W**
  - **Compare the fits of the sample vs MC**

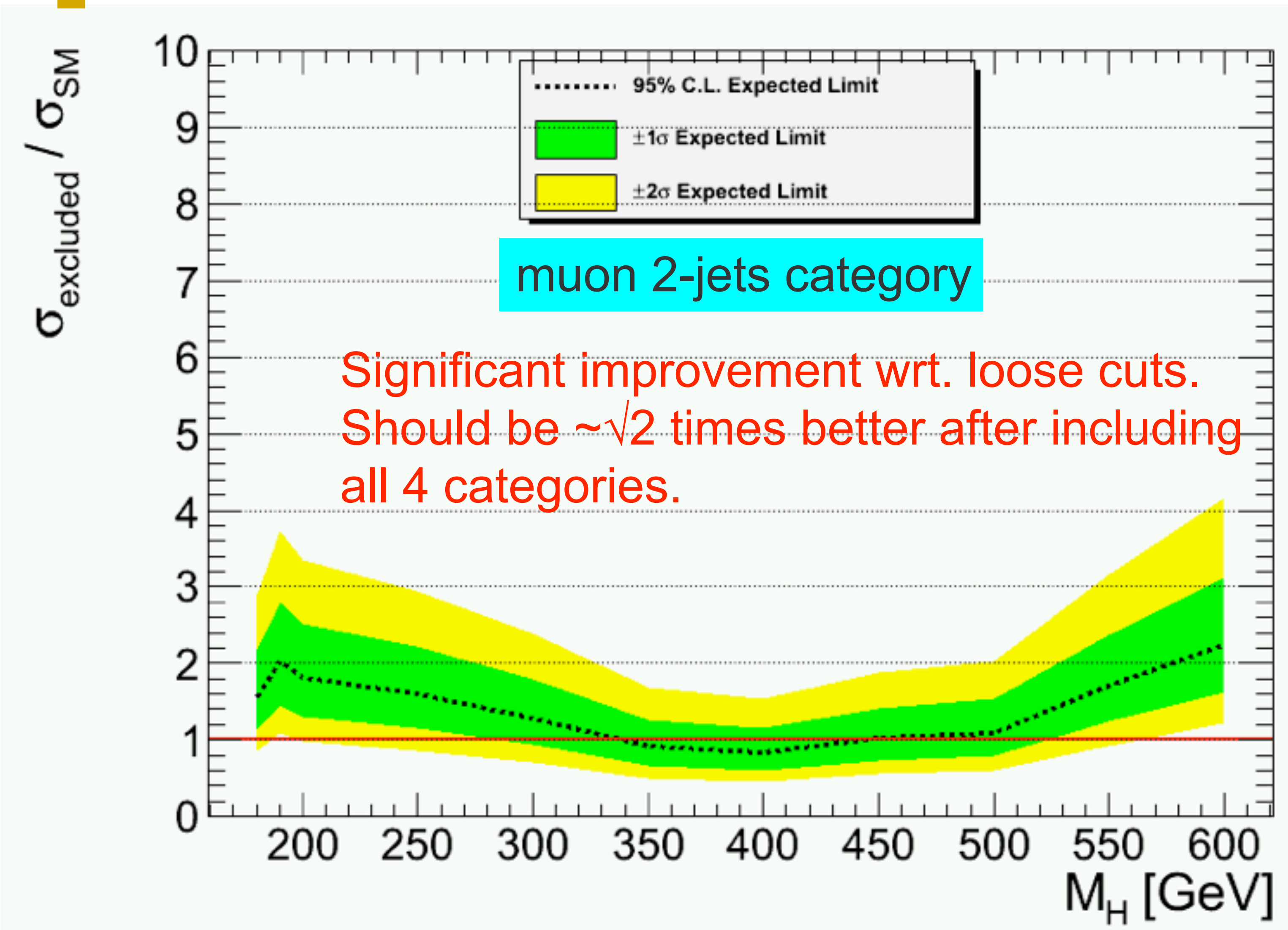
Use hadronic W from top events



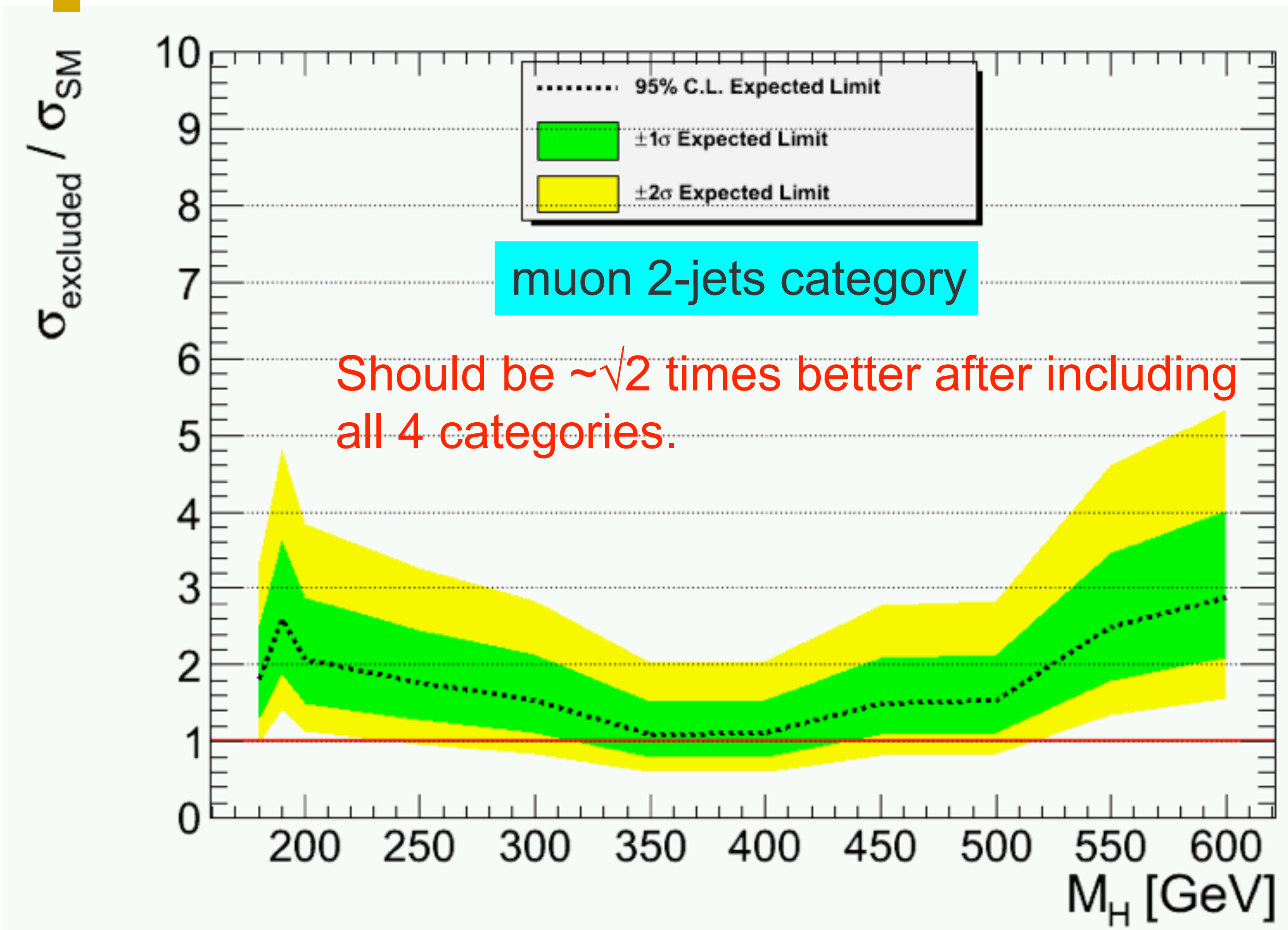
- **The difference in JES is propagated to our templates and makes a negligible impact**

JEC and JER in data and MC agree within 0.2%

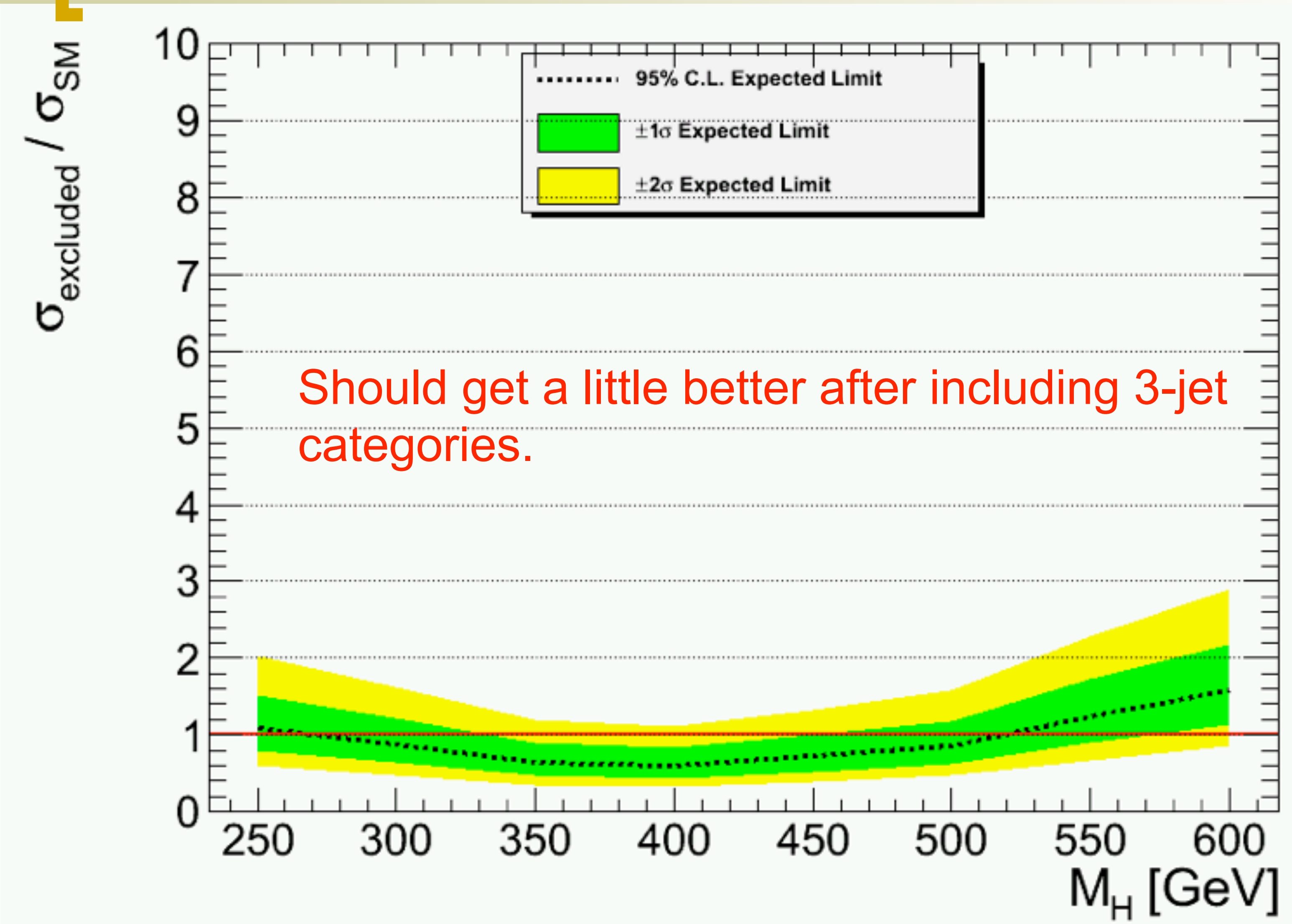
Statistics-only limit: muon 2-jets only



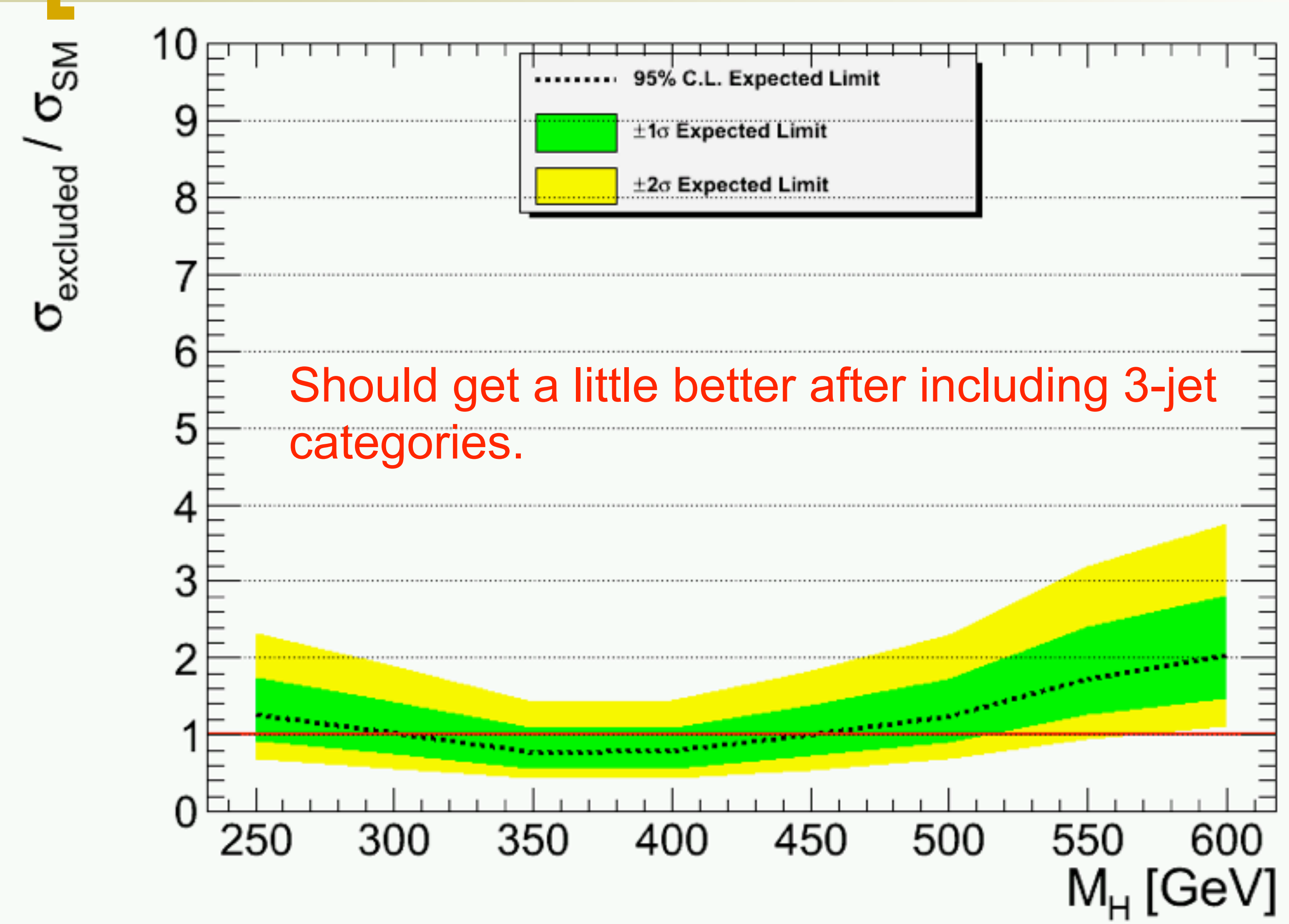
Limit after full syst: muon 2-jets only



Statistics-only limit: mu+ele 2-jets only



Limit after full syst: mu+ele 2-jets only



Should get a little better after including 3-jet categories.

Summary

- This time we do not want to miss the Moriond deadline. We were stuck so far due to large systematic uncertainties. Focus sharply on reducing the syst error.
  - Optimized a simple likelihood for each M_H
  - Use dijet mass (m_{jj}) to derive normalization as before
  - Improved data-driven estimation of W+jets shape for m_{WW}

- Reduced syst ➜ improved exclusion limits
  - have preliminary result for muon channel 2-jets category
  - work ongoing to complete analysis of other three categories

- Working on documentation in parallel
- Will reproduce all results using the agreed-upon PAT-tuples with other groups, hopefully in a week's time scale.

backup slides