Search for the Standard Model Higgs boson in H$\rightarrow$WW$\rightarrow\ell\nu jj$

Jake Anderson
Fermilab

On behalf of the H$\rightarrow$WW$\rightarrow\ell\nu jj$ analysis team

The team

Nural Akchurin[1], **Jake Anderson[2]**, Andrea Benaglia[3], Andrew Beretvas[2], Jeffrey Berryhill[2], Pushpa Bhat[2], Sarah Boutle[4], Chris Clarke[5], Fabio Colombo[3], Daniele del Re[9], Leonardo Di Matteo[3], Phil Dudero[1], Ricardo Eusebi[6], Pietro Govoni[13], Dan Green[2], Joey Goodell[4], Robert Harr[5], Pratima Jindal[7], Gordon Kaussen[11], Kristina Krylova[5], Kevin Lannon[8], Sung-Won Lee[1], Shuai Liu[12],Wuming Luo[8], Andrea Massironi[3], Kellen McGee[5], Kalanand Mishra[2], Chris Neu[4], Ilya Osipenkov[6], Alexx Perloff[6], Luca Pernie[9], Sasha Sakharov[5], Kevin Siehl[5], Jason Slaunwhite[8], Andre Sznajder[10], Nhan V. Tran[2], John Wood[4], Fan Yang[2], and Francisco Yumiceva[2]

[1] Texas Tech University, Lubbock, Texas, USA
[2] Fermi National Accelerator Laboratory, Batavia, Illinois, USA
[3] Milano-Bicocca University and INFN, Milan, Italy
[4] University of Virginia, Charlottesville, Virginia, USA
[5] Wayne State University, Detroit, Michigan, USA
[6] Texas A&M University, College Station, Texas, USA
[7] University of Nebraska at Lincoln, Nebraska, USA
[8] University of Notre Dame, Notre Dame, Indiana, USA
[9] University of Rome, La Sapienza, Rome, Italy
[10] Universidade do Estado do Rio de Janeiro (UERJ), Brazil
[11] University of Hamburg, Hamburg, Germany
[12] University of Peking, China
[13] CERN

Acknowledgements

- We sincerely thank our ARC members for their time and effort getting our analysis to this point.
  - Radia Redjimi Boulahouache
  - Vitaliano Ciulli
  - Gunther Dissertori
  - Marcello Maggi
- We would also like to thank our conveners for their support.

The analysis

- Cadi
  - http://cms.cern.ch/iCMS/analysisadmin/cadi?ancode=HIG-12-003
- Hypernews
  - https://hypernews.cern.ch/HyperNews/CMS/get/HIG-12-003.html
- twiki
  - https://twiki.cern.ch/twiki/bin/view/Main/TempWWlvjj
- Analysis notes
  - **AN-2011/110**, AN-2012/008, AN-2012/021, AN-2012/024, AN-2012/029
- PAS
  - http://cms.cern.ch/iCMS/analysisadmin/versions?analysis=HIG-12-003

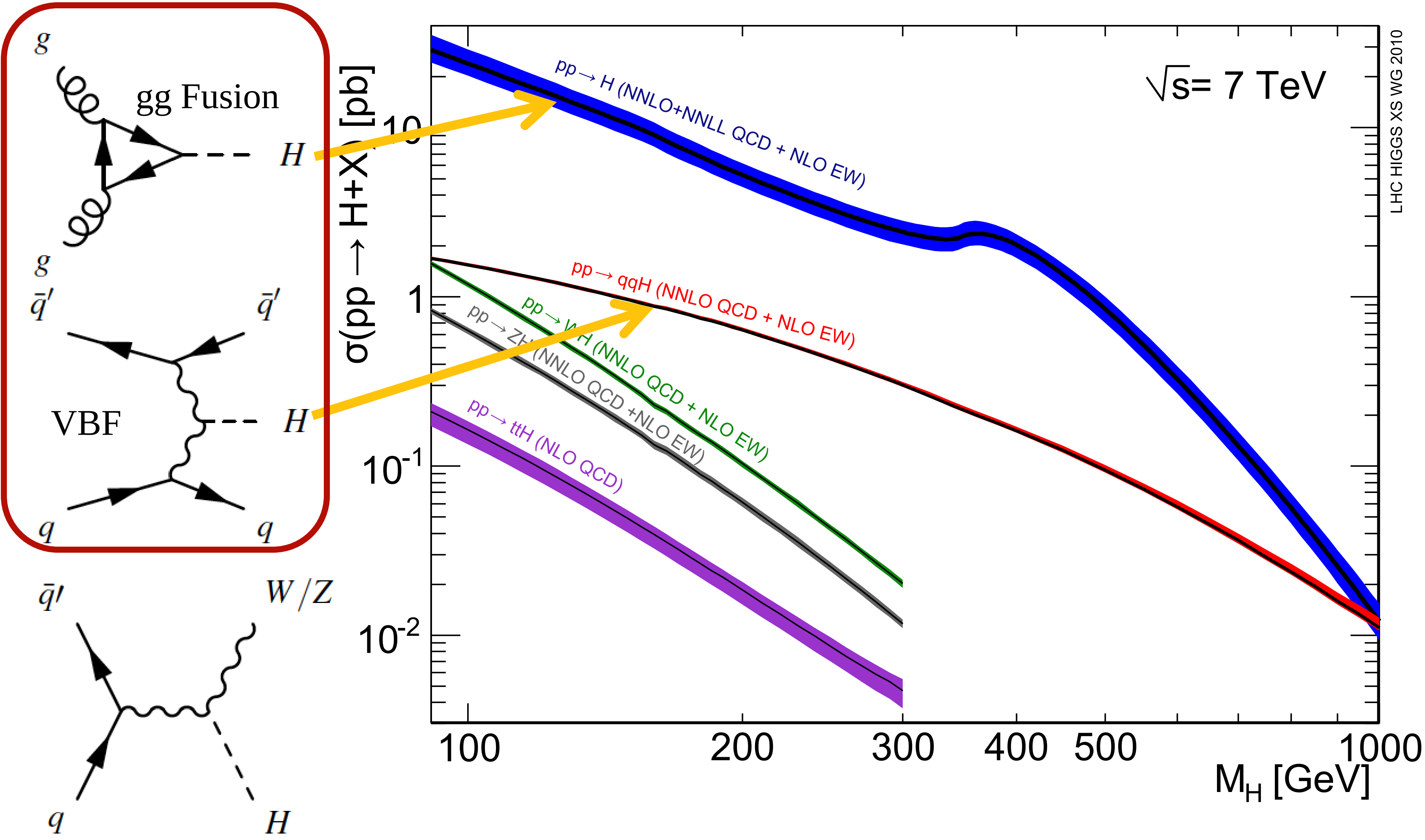
Outline

- Motivation, production and decay
- Dataset and simulation
- Pre-selection
- Two parallel and independent analyses

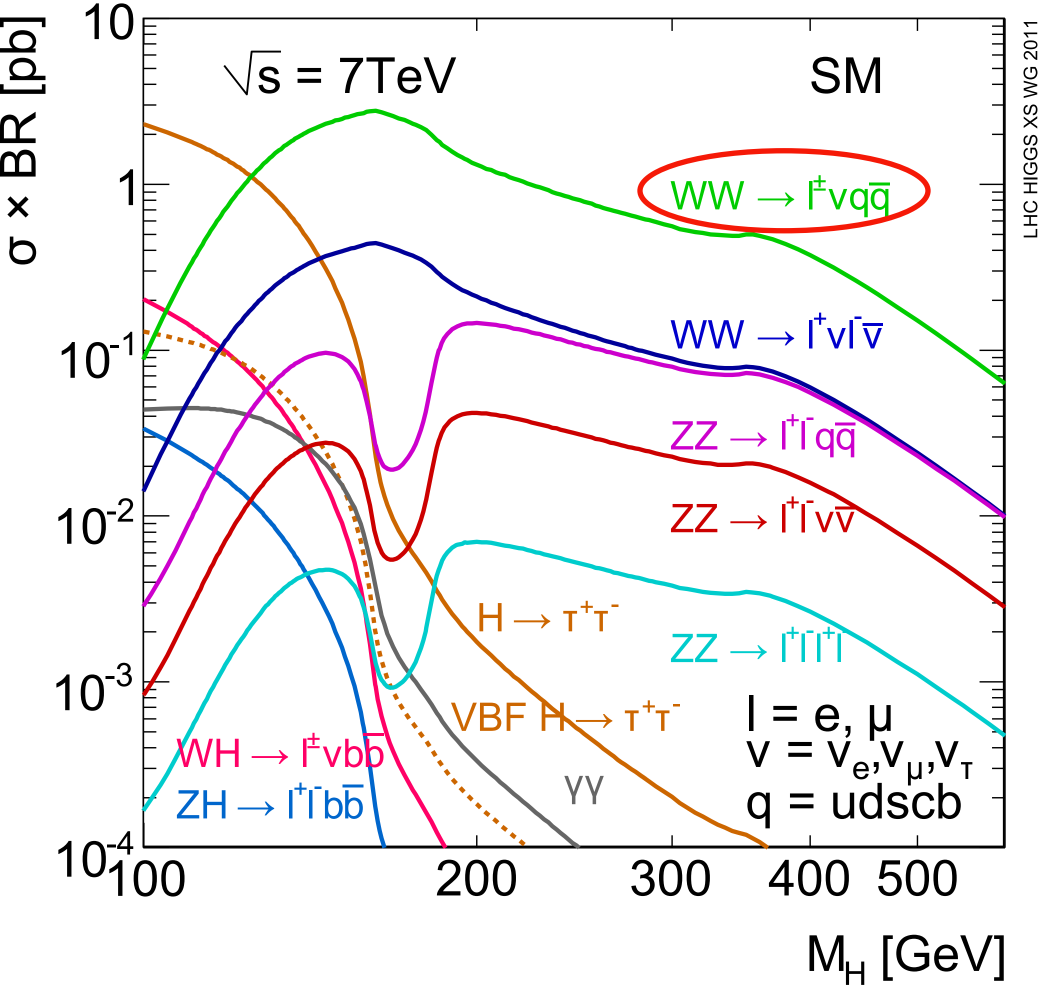
- Conservative fit-based analysis
  - selection
  - modeling
  - systematic errors
  - limits

- Advanced MVA-based analysis
  - MVA description and optimization
  - analysis flow
  - backgrounds
  - m_{jj} fit
  - data driven W+jets shape
  - systematic errors
  - limits

Higgs production



Higgs decays



- H → WW → ℓνjj does a lot of heavy lifting.
  - largest BR × σ over most of the mass range.
  - Using a W mass constraint, the decay is sufficiently reconstructed to produce a mass peak.
- Principal drawback is the large W+jet background
  - We employ data-driven techniques to understand and control this process.

Data samples

dataset name	run range	int. lumi
/SingleElectron/Run2011A-May10ReReco-v1/AOD /SingleMu/Run2011A-May10ReReco-v1/AOD	160404-153869	211 pb^{-1}
/SingleElectron/Run2011A-PromptReco-v4/AOD /SingleMu/Run2011A-PromptReco-v4/AOD	165088-167913	930 pb^{-1}
/SingleElectron/Run2011A-05Aug2011-v1/AOD /SingleMu/Run2011A-05Aug2011-v1/AOD	170249-172619	368 pb^{-1}
/SingleElectron/Run2011A-PromptReco-v6/AOD /SingleMu/Run2011A-PromptReco-v6/AOD	172620-173692	659 pb^{-1}
/SingleElectron/Run2011B-PromptReco-v1/AOD /SingleMu/Run2011B-PromptReco-v1/AOD	175832-180252	2512 pb^{-1}
	160404-180252	**5.0 fb^{-1}**

- Certified “**golden**” **JSON** used

MC samples

dataset name	x-sec (pb)	eq. lumi (fb^{-1})
/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	**31314**	**2.5**
/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	3048	11.3
/TTJets_TuneZ2_7TeV-madgraph-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	163	22.7
/T_TuneZ2_s-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	3.19	81.5
/T_TuneZ2_t-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	41.9	93.0
/T_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	7.87	103
/Tbar_TuneZ2_s-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	1.44	95.8
/Tbar_TuneZ2_t-channel_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	2.26	85.9
/Tbar_TuneZ2_tW-channel-DR_7TeV-powheg-tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	7.87	41.1
/WW_TuneZ2_7TeV_pythia6_tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	47.0	89.9
/WZ_TuneZ2_7TeV_pythia6_tauola/Fall11-PU_S6_START42_V14B-v1/AODSIM	18.2	234
/GluGluToHToWWToLNuQQ_M-XXX_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B-v1/AODSIM	depends on Higgs mass	a lot
/GluGluToHToWWToTauNuQQ_M-XXX_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B-v1/AODSIM		
/VBF_HToWWToLNuQQ_M-XXX_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B-v1/AODSIM		
/VBF_HToWWToTauNuQQ_M-XXX_7TeV-powheg-pythia6/Fall11-PU_S6_START42_V14B-v1/AODSIM		

- All MC samples corrected for to account for different pile-up conditions in data
  - 3D pile-up re-weighting technique applied

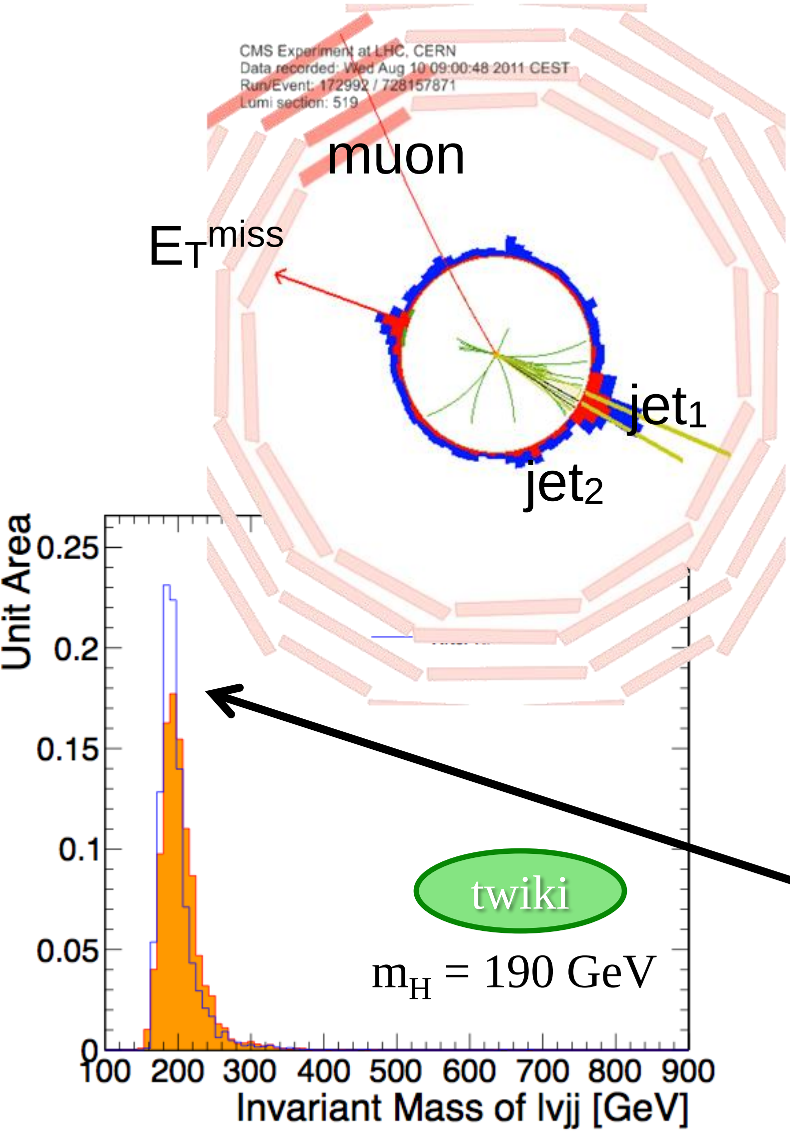
Trigger strategy

- Lepton as well as PFMT **HLT efficiencies** for electron events are **corrected** to account for data/MC differences by means of “Tag and Probe” with Z events
- Same technique is used to measure **reconstruction/identification scale factors**

electron trigger paths	**run range**
HLT_Ele27_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v*	160404-153869 (May10ReReco-v1)
HLT_Ele32_CaloIdVT_CaloIsoT_TrkIdT_TrkIsoT_v* HLT_Ele25_WP80_PFMT40_v* HLT_Ele27_WP80_PFMT50_v*	165088-167913 (PromptReco-v4)
HLT_Ele32_WP70_PFMT50_v*	170249-172619 (05Aug2011-v1)
HLT_Ele27_WP80_PFMT50_v*	172620-173629 (PromptReco-v6)
HLT_Ele27_WP80_PFMT50_v*	175832-180252 (PromptReco-v1)

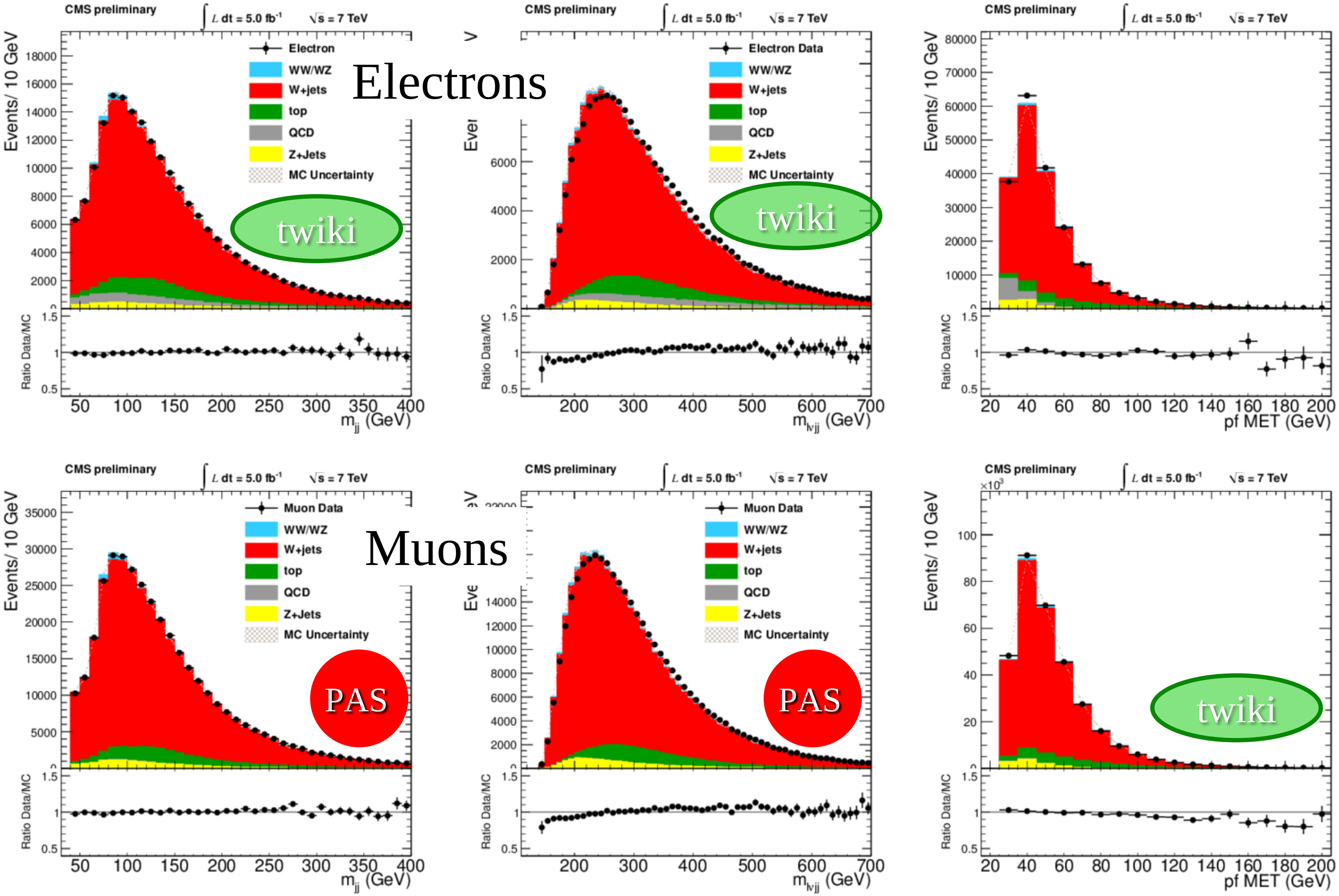
muon trigger paths	**run range**
HLT_IsoMu17_v* OR HLT_Mu30_v*	160404-153869 (May10ReReco-v1)
HLT_IsoMu17_v* OR HLT_Mu30_v*	165088-167913 (PromptReco-v4)
HLT_IsoMu17_v* OR HLT_IsoMu20_v* OR HLT_IsoMu24_v*	170249-172619 (05Aug2011-v1)
HLT_IsoMu20_v* OR HLT_IsoMu24_eta2p1_v*	172620-173692 (PromptReco-v6)
HLT_IsoMu24_eta2p1_v* OR HLT_IsoMu30_eta2p1_v*	175832-180252 (PromptReco-v1)

Pre-selection cuts



- One isolated, high-p_T lepton
  - $p_T > 35\ (25)$ GeV for electrons (muons)
- High E_T^{miss} from 1 neutrino
  - $E_T^{miss} > 30\ (25)$ GeV for electrons (muons)
  - $m_T(\text{lepton}+E_T^{miss}) > 50$ GeV
- Two high pT jets with $m_{jj} \sim 80$ GeV
  - AK5PFJets with CHS / FastJet corrections
  - $p_T > 30$ GeV, $|\eta| < 2.4$
  - $\Delta R(\text{jet-lepton}) > 0.3$
  - $N_{\text{extra-jets}} = 0,1$
- neutrino p_Z from m_W constraint
- We do a kinematic fit on lepton,E_T^{miss}, hadronic W to improve Higgs mass resolution and to remove the correlation between m_{WW} from m_{jj}.

Control plots - m_{jj} / $m_{l\nu jj}$ / E_T^{miss}



Electrons

twiki

twiki

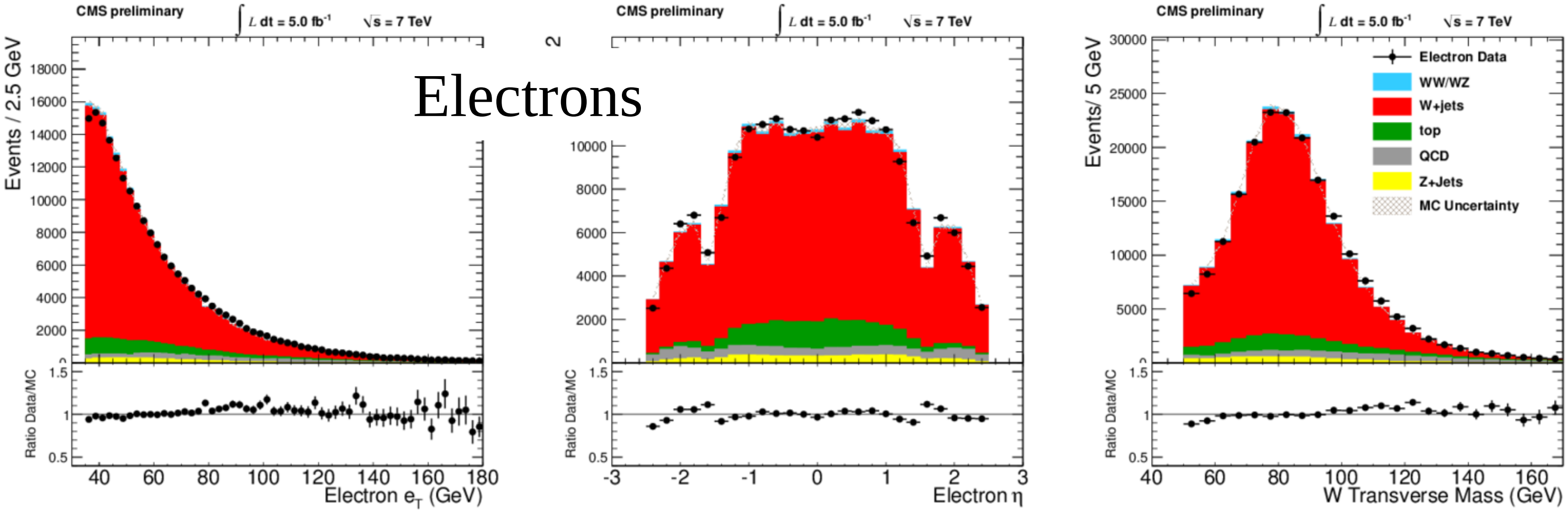
Muons

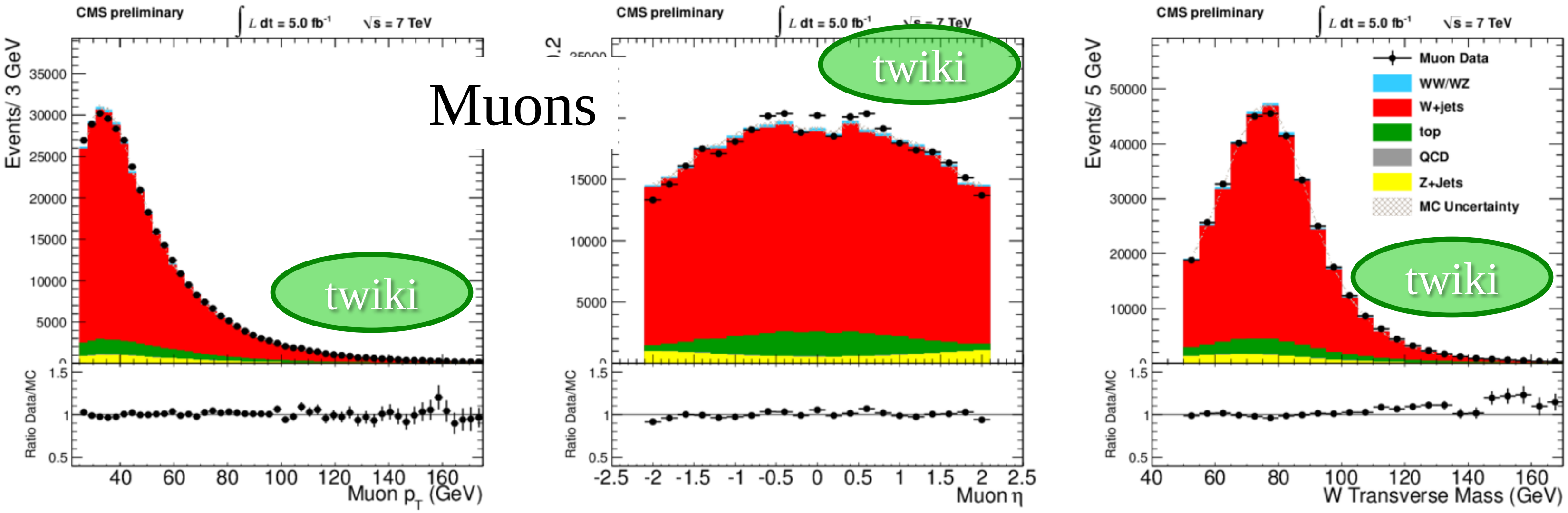
PAS

PAS

twiki

Control plots - lepton


Electrons

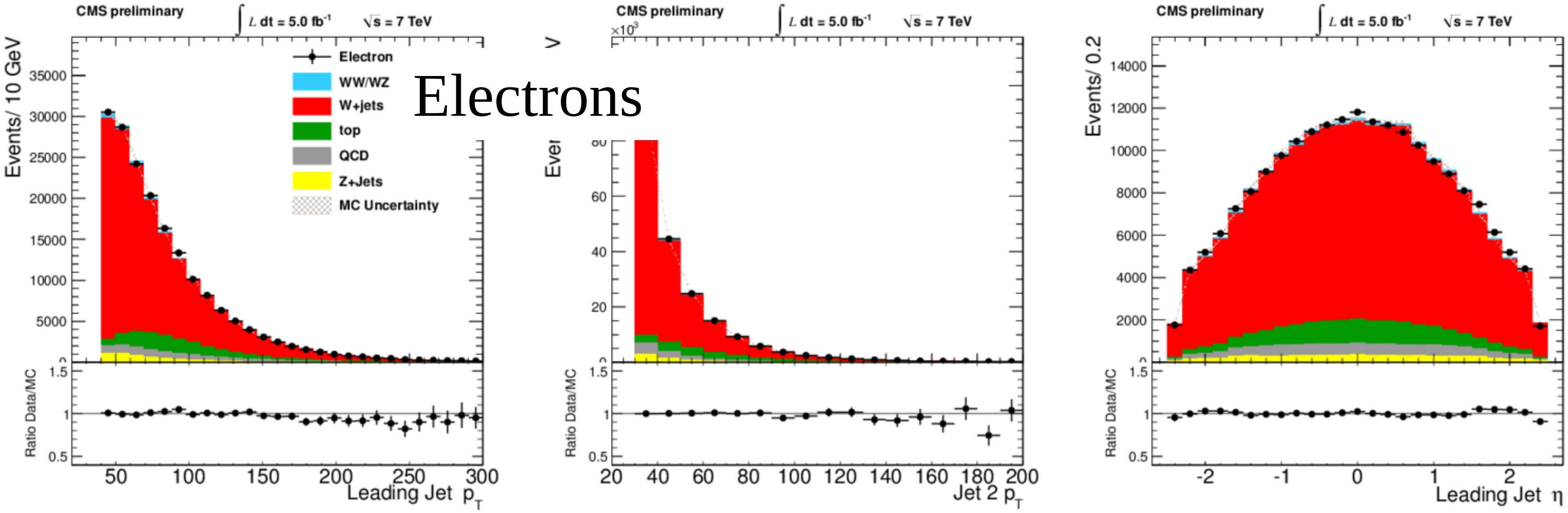

Muons

twiki

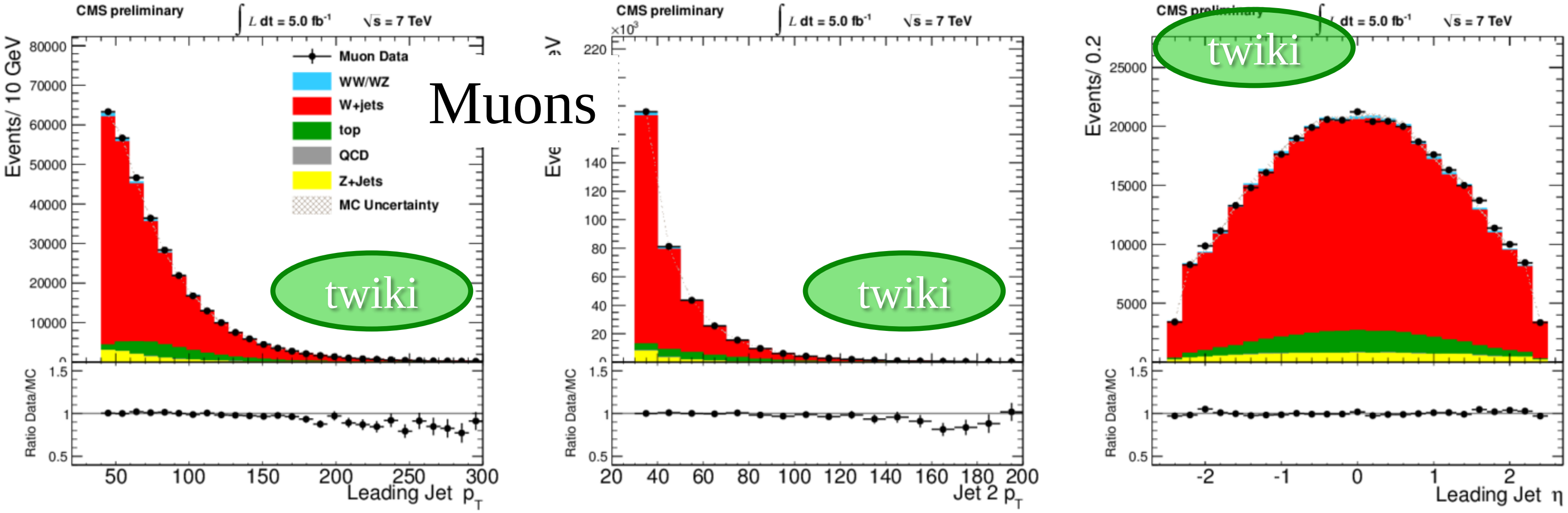
twiki

twiki

Control plots - jets



Electrons



Muons

twiki

twiki

twiki

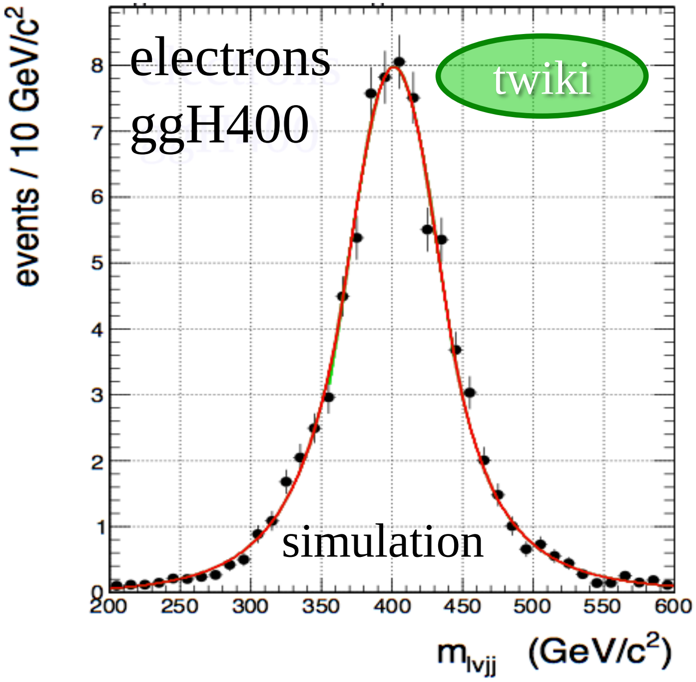
Fit-based limit

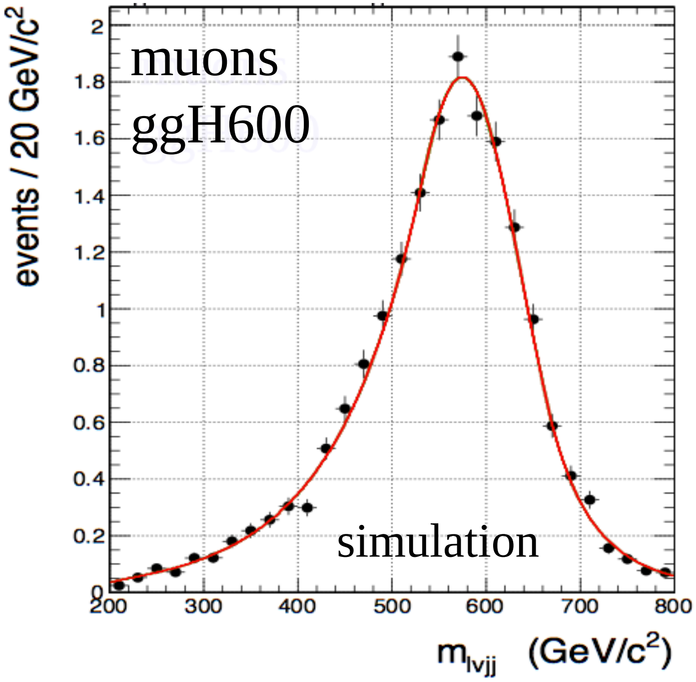
The fit strategy

- Extract a limit on SM Higgs x-section just after common pre-selections
    - lots of events survive: ~20k (~30k) events in the electron (muon) channel
    - the signal shape is modeled by fitting the MC shapes (double crystal ball)
    - the background shape is modeled by a suitable analytic function (attenuated power law)
- Limit extracted performing shape analysis on $m_{l\nu jj}$ spectrum
    - data are fitted within the combination tool with a B+μS hypothesis to calculate the test statistics
    - only Higgs mass points within 250 and 600 GeV are considered
- Only few additional cuts applied on top of pre-selections to enhance S/B:
    - centrality cuts: $|\eta_{lepton}| < 1.5$, $|\eta_{jj}| < 3.$, $|\Delta\eta_{jj}| < 1.5$
    - $m_T > 50$ GeV relaxed to 30 GeV for muons
    - veto b-tags
    - cut on had. W resonance: $(65 < m_{jj} < 95)$ GeV

Signal modeling

- Signal shape in the limit extraction is modeled with a smooth function
  - Gaussian core with two power laws to describe the leading/trailing edge (double crystal-ball)
  - fitted to the signal MC samples for each mass



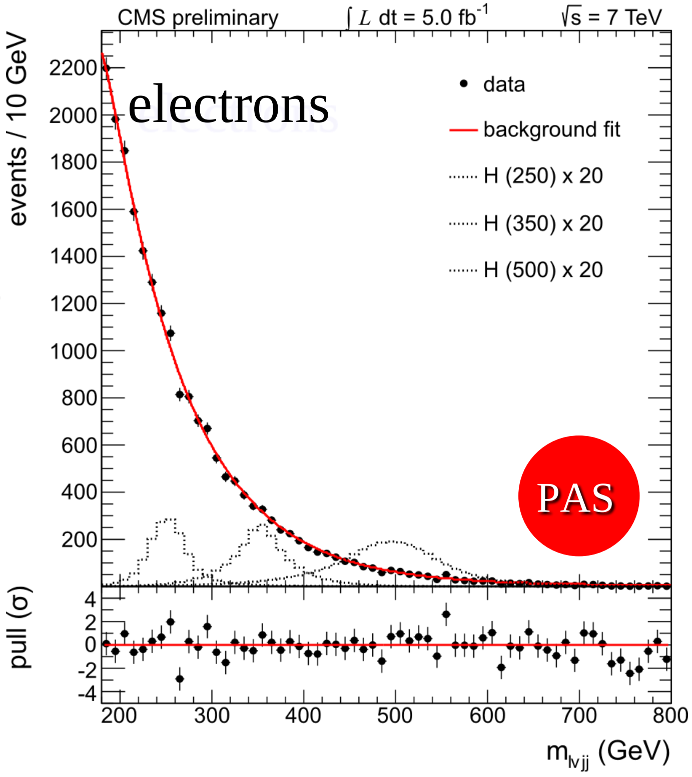


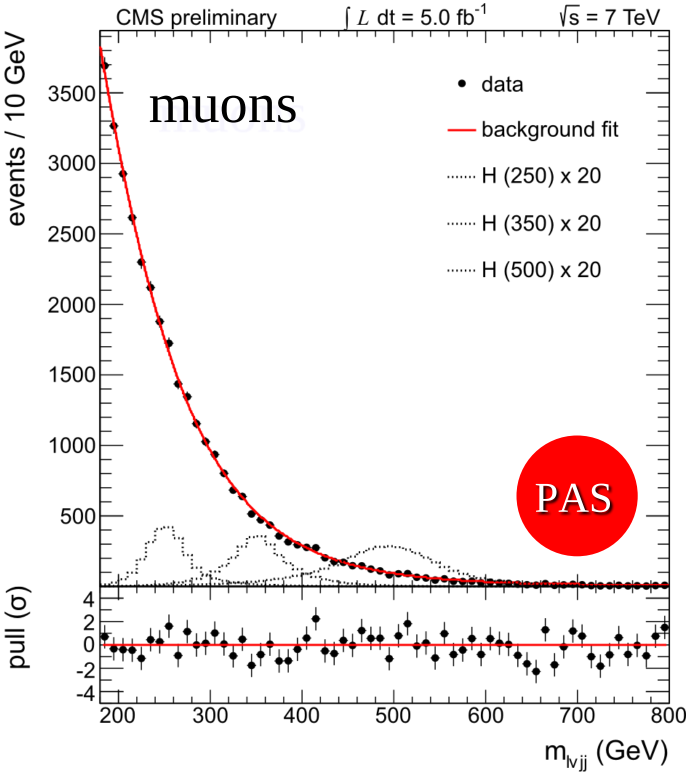
Background modeling

- An empirical attenuated power law shape is used to model the background $$\mathcal{F}_{\mathrm{B}} = \frac{1}{1+e^{-(x-\mu)/\tau}} \cdot \left(\frac{500+a}{x+a}\right)^{n}$$
  - μ, τ, a and n determined from data

fit range:
180-800 GeV/c^2

bin size:
5 GeV/c^2





Systematics uncertainties

background model choice(*)	2-40%
inclusive cross-section	15-20%
gluon-fusion higgs line shape	10-30%
acceptance uncertainty due to pdfs	1-2%
luminosity	2.2%
jet energy scale and E_T^{miss}	2-3%
lepton efficiencies	1-2%
lepton trigger efficiencies	<1%
pile-up	<1%
$PFMH_T$ trigger efficiency (electron channel)	1-2%
b-tagging	< 1%

(*) more on following slide

Background model systematics

- We need to determine the systematic uncertainty related to choice of the background shape
- Our study:
  - We have a set of functions that describe the MC equally well.
  - We consider the differences with respect to the nominal functional form as the uncertainty on our choice.
- The background uncertainty is converted into an uncertainty on the signal.

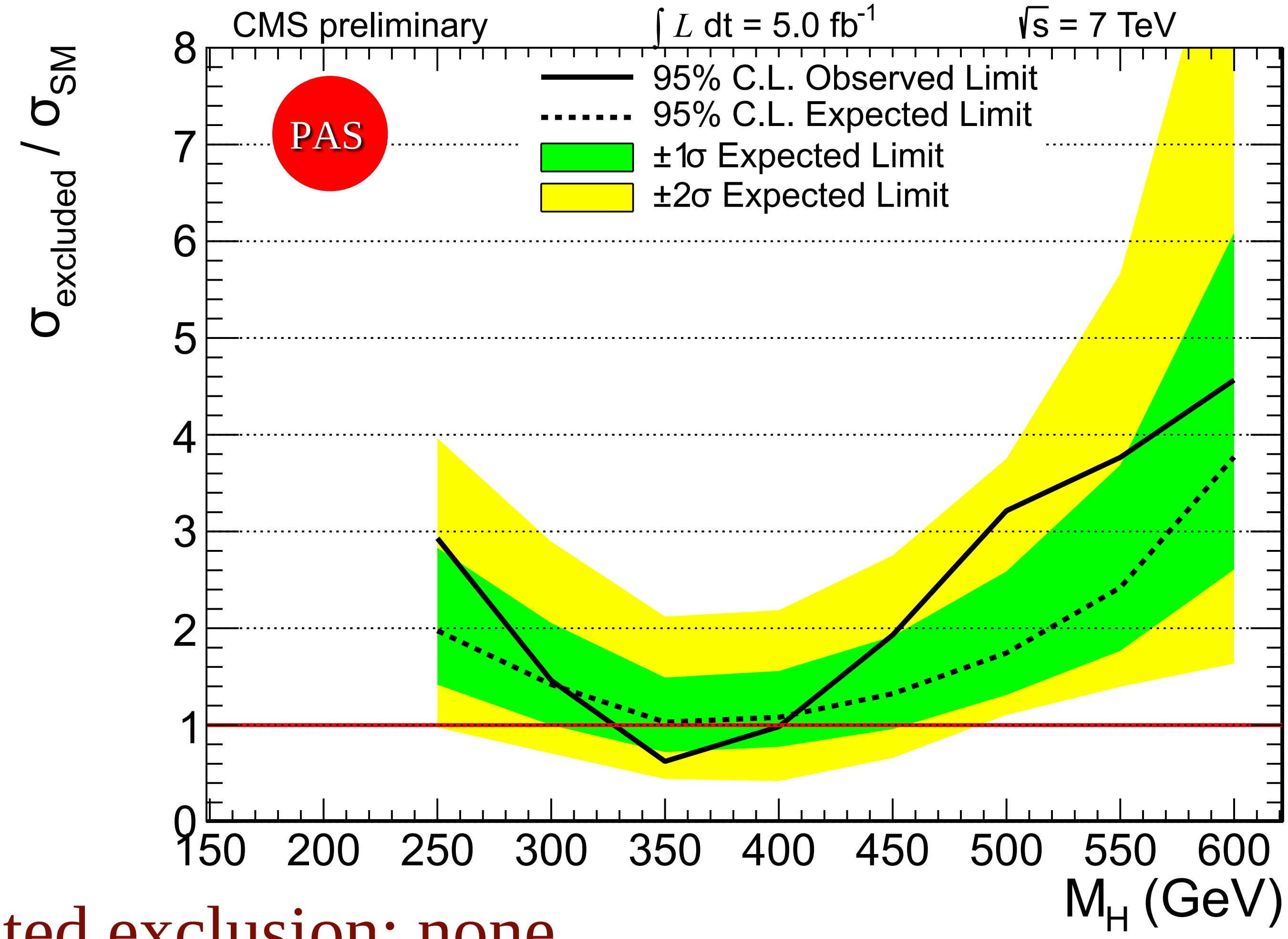
$$\mathcal{F}_{\mathrm{B}} = \frac{1}{1+e^{-(x-\mu)/\tau}} \cdot \left(\frac{500+a}{x+a}\right)^{n}$$

$$\mathcal{F}_{\mathrm{B,\ alt}} = \frac{1}{1+e^{-(x-\mu)/\tau}} \cdot \left(e^{ax^3+bx^2+cx}\right)$$

$$\mathcal{F}_{\mathrm{B,\ alt}} = \frac{1}{1+e^{-(x-\mu)/\tau}} \cdot \left(e^{-\lambda_1 x} + N_2 e^{-\lambda_2 x}\right)$$

	electrons	**muons**
m_H	**unc.**	**unc.**
250	13%	3.0%
300	17%	4.2%
350	14%	2.6%
400	4.5%	4.6%
450	4.3%	3.2%
500	14%	2.1%
550	24%	10%
600	38%	18%

The limit



- Expected exclusion: none
- Observed exclusion: 325 – 400 GeV

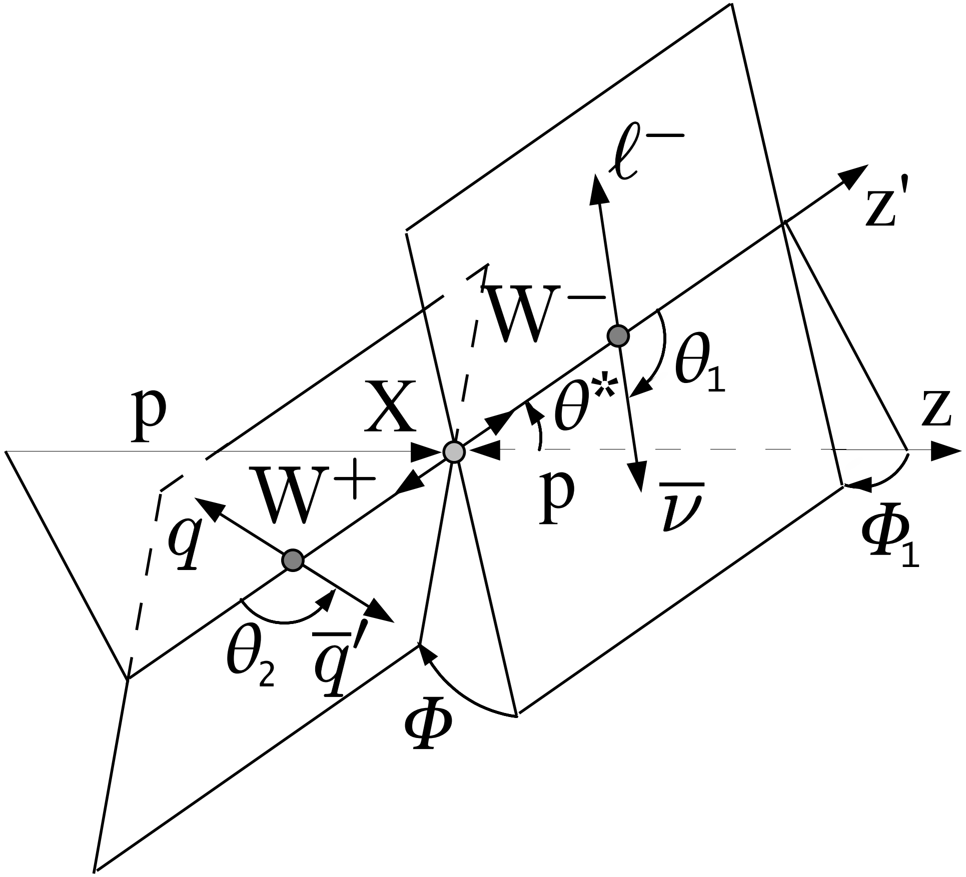
MVA analysis

Strategy

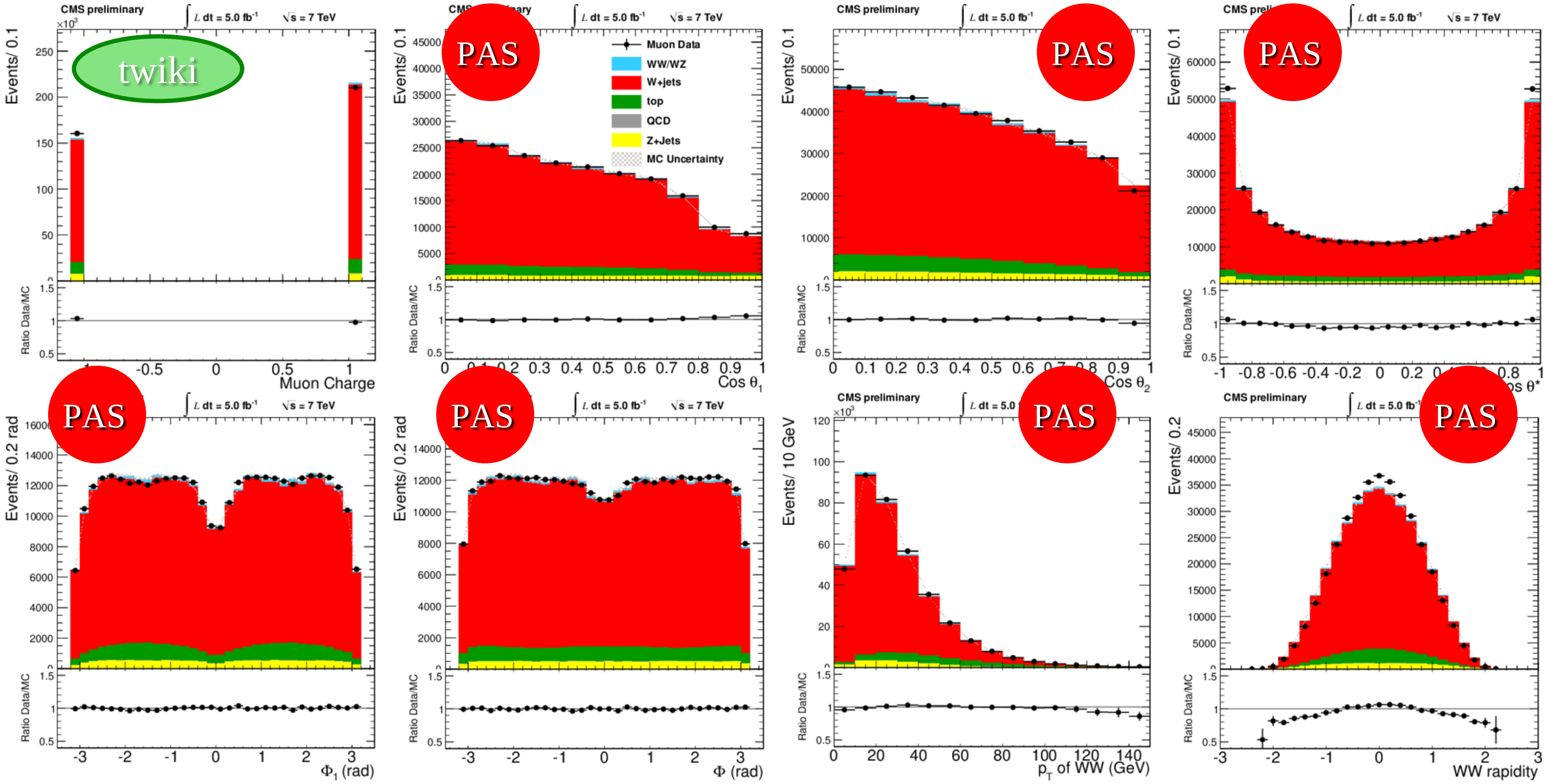
- Goals
  - Improve the sensitivity by significantly reducing the background and its inherent uncertainty.
  - Extend the mass reach of the analysis down as close to the kinematic edge as possible.
- We exploit the kinematic information in the event to discriminate against W+jets production.
  - Use TMVA package to build a **likelihood** discriminator
  - A separate discriminator for each of 12 mass points × 2 lepton flavors × 2 jet bins = 48 discriminators

MVA

- The Higgs production and decay kinematics are described by {m_{WW}, m_{jj}, θ_1, θ_2, θ^*, ϕ, ϕ_1}.
    - m_{WW} is used to set the limit.
    - m_{jj} is used to normalize the backgrounds.
- We include the lepton sign as the background is charge asymmetric.
- We include $p_T(WW)$ and y_{WW} for additional discrimination.
- All variables have negligibly small correlations.

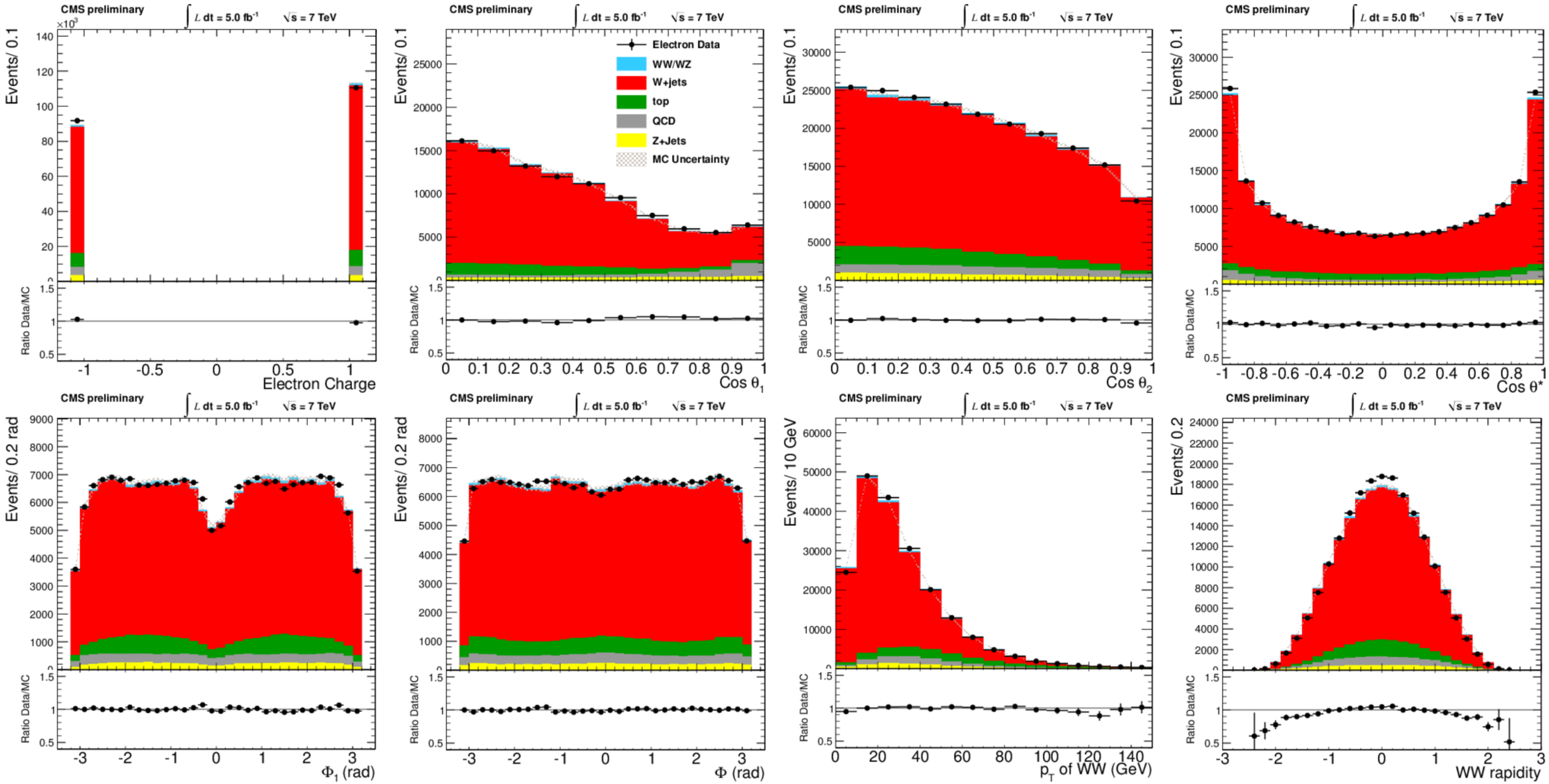


MVA inputs (muons)



Muon 2- and 3-jets

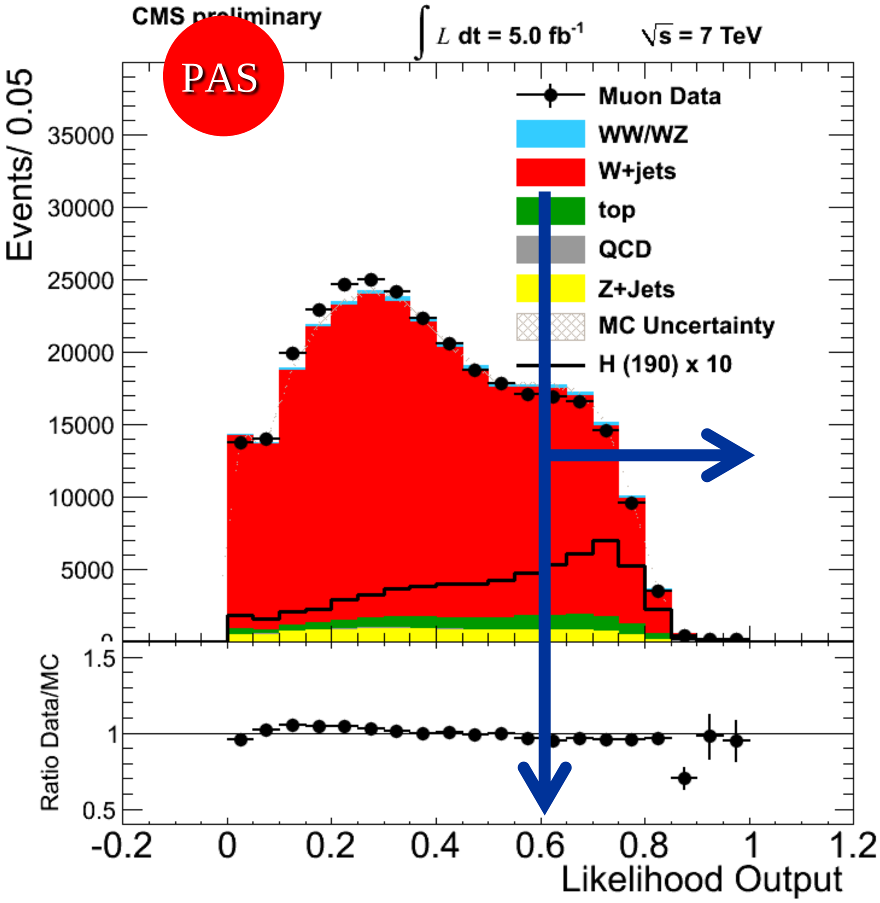
MVA inputs (electrons)



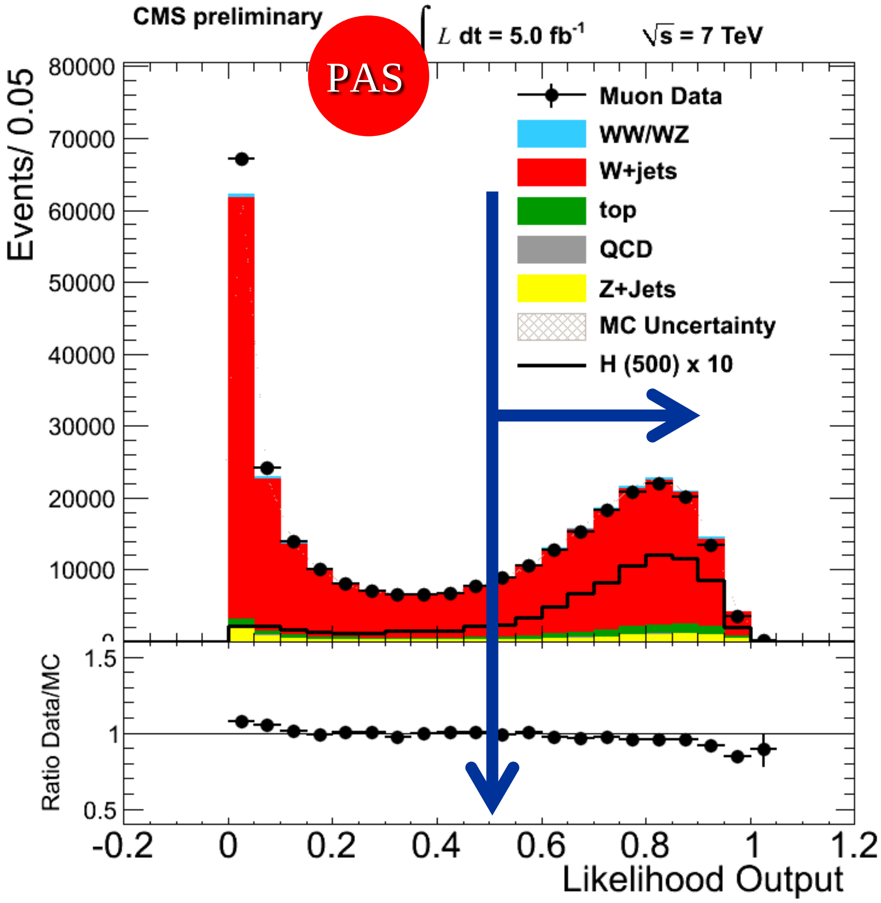
Electron 2- and 3-jets

MVA output

muon, 2-jet, m_H = 190 GeV



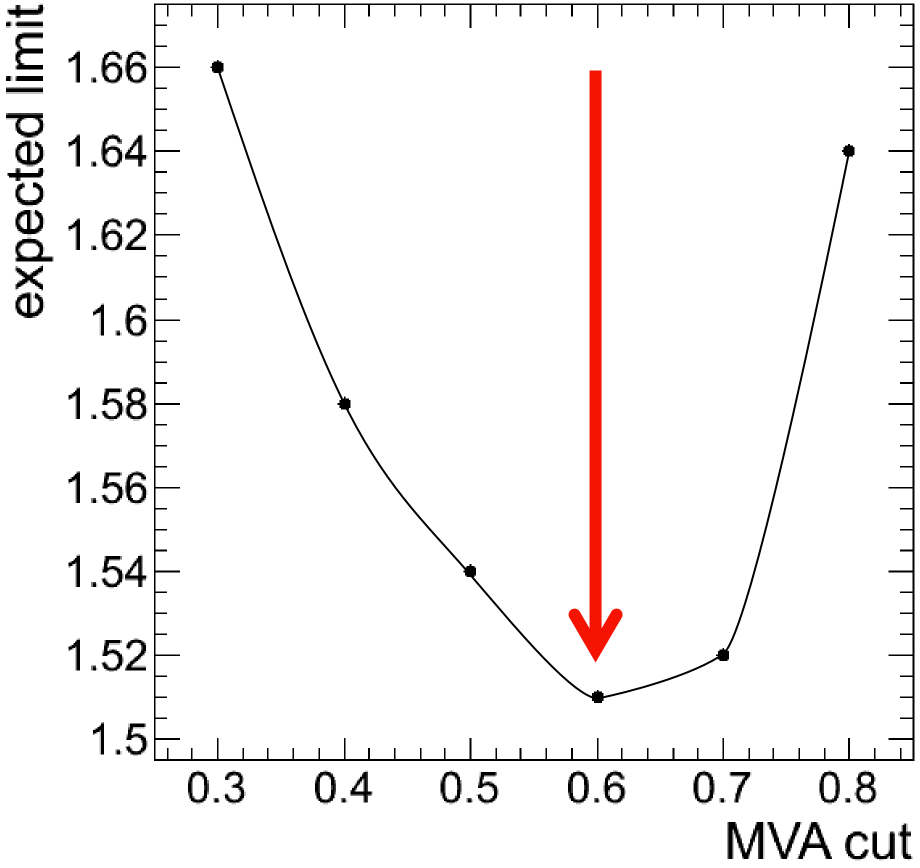
muon, 2-jet, m_H = 500 GeV



MVA cut optimization

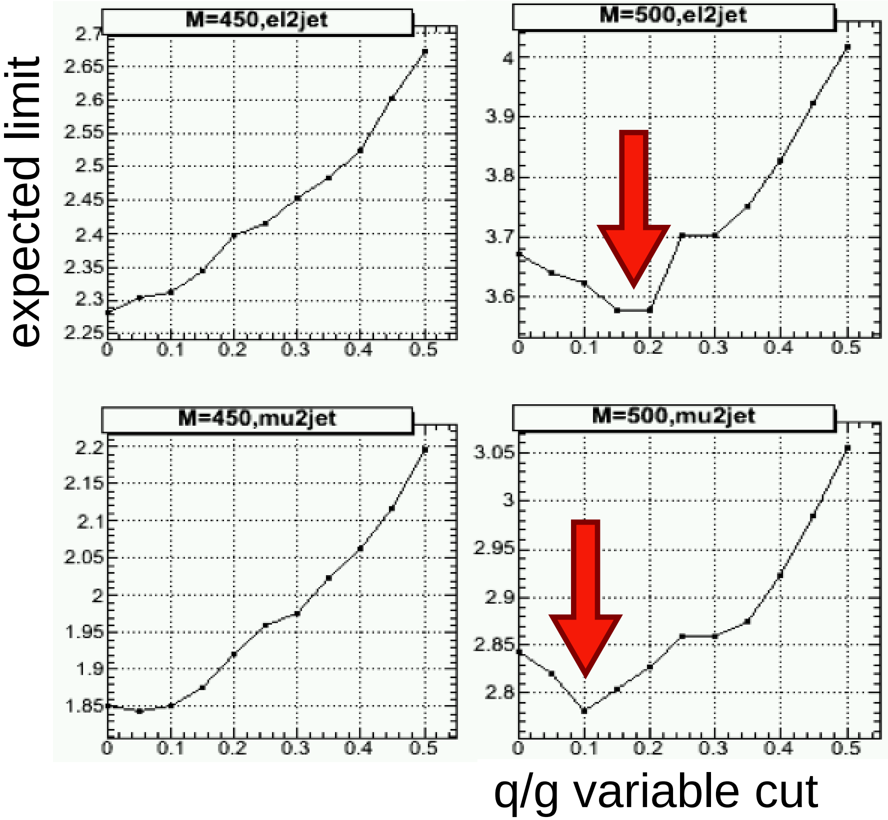
- We select the MVA cut value based on running the full asymptotic limit setting machinery and using the expected limit.
- Once the optimal cut is selected for each of the 48 analysis points they can be combined using the standard Higgs combination package.

muon, 2-jets, m_H = 350 GeV



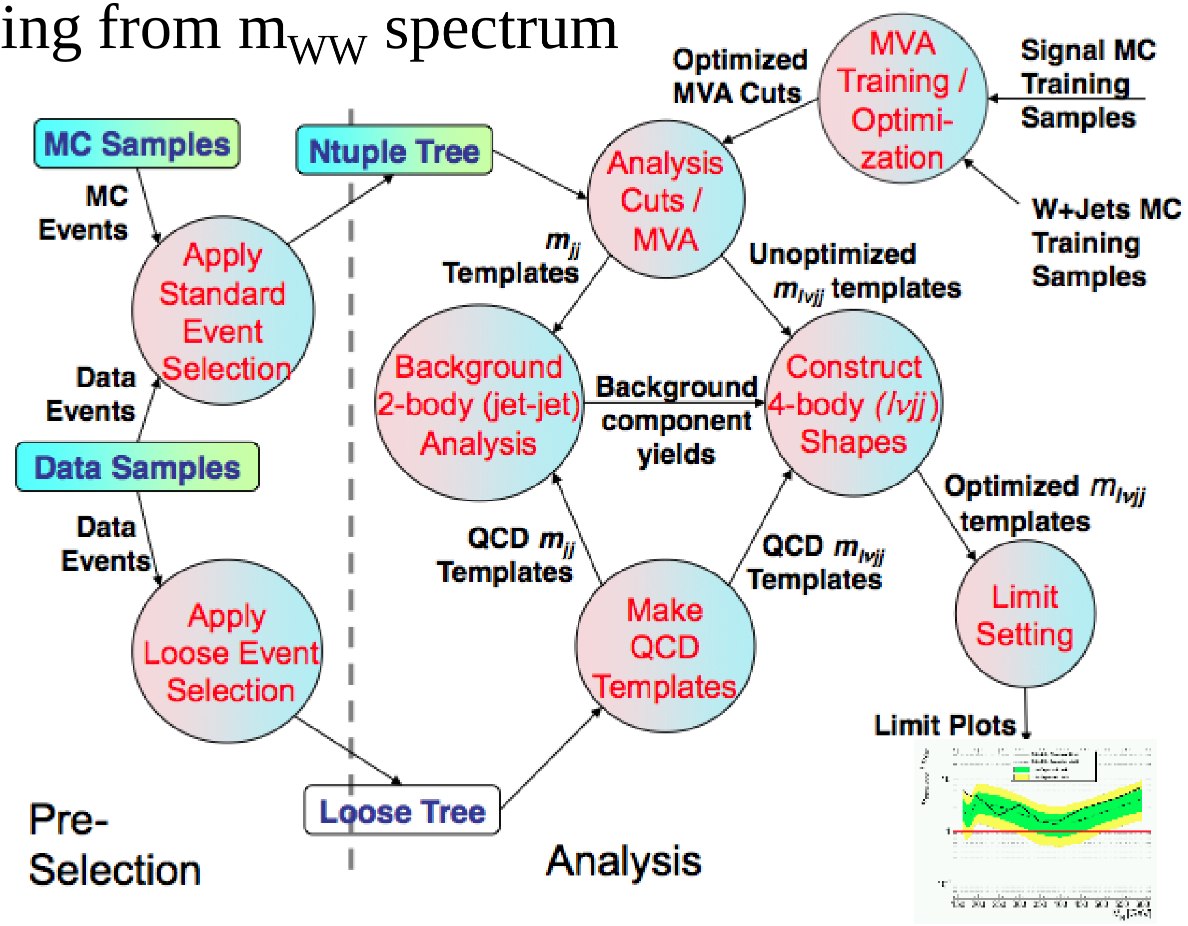
Additional high mass cut

- In addition to the MVA cut we look at separating gluon-initiated jets from quark-initiated jets using a quark-gluon likelihood discriminator.
- Here again we optimize this cut based on the final expected limit.
- We don't find this cut to be advantageous for Higgs masses below 500 GeV.



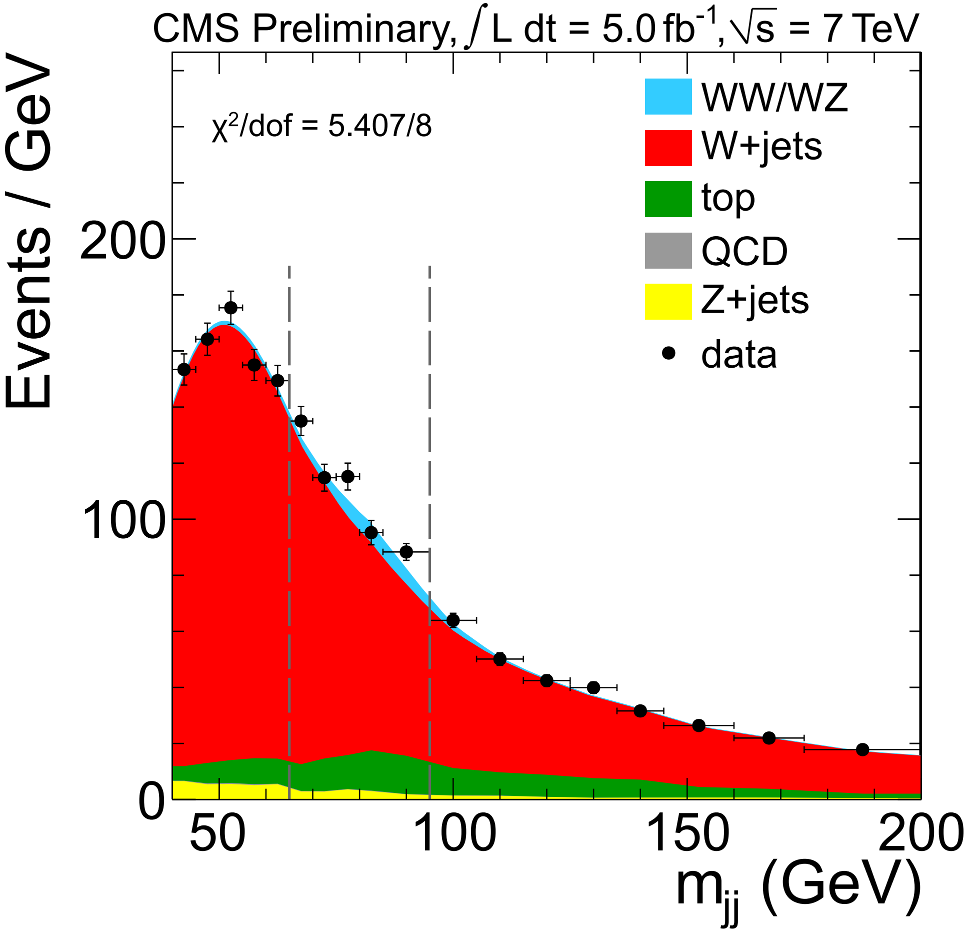
Fitting and limit setting

- The analysis has two major parts
  - Fit to the m_{jj} spectrum to determine the backgrounds
  - limit setting from m_{WW} spectrum

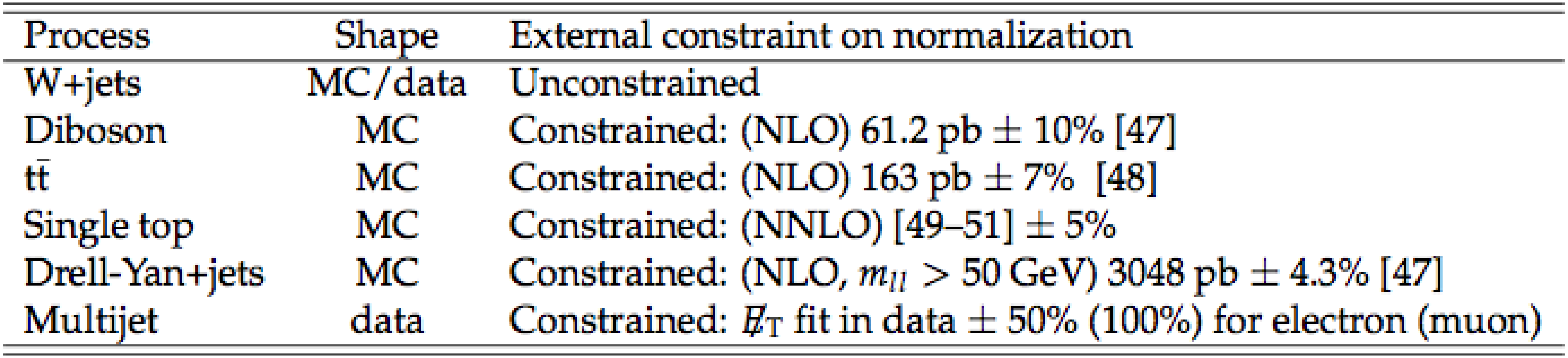


The fit procedure

- We determine the background composition in a 1D, unbinned, maximum likelihood fit to the di-jet invariant mass spectrum.
- The background shapes are taken from:
  - MC for all minor backgrounds (not for W+jets and QCD)
  - data-driven approach for QCD (next slides)
  - analytic description or MC for W+jets (following slides)



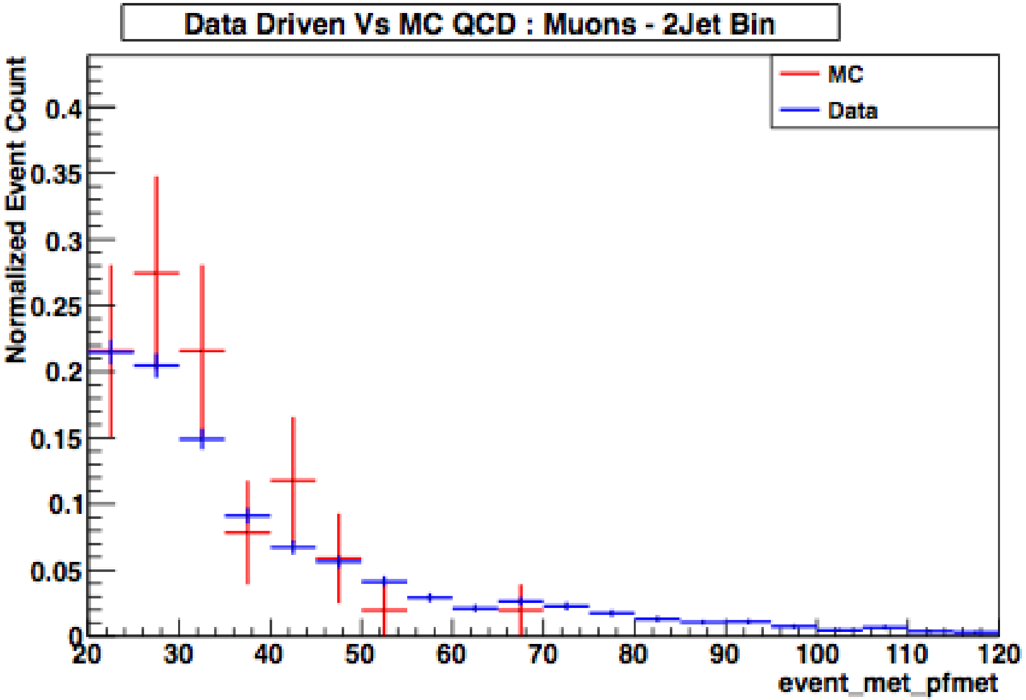
- The background yields in the fit are:
  - Constrained to the best MC cross-section including error for all minor backgrounds.
  - QCD constrained from the data.
  - W+jets is unconstrained.



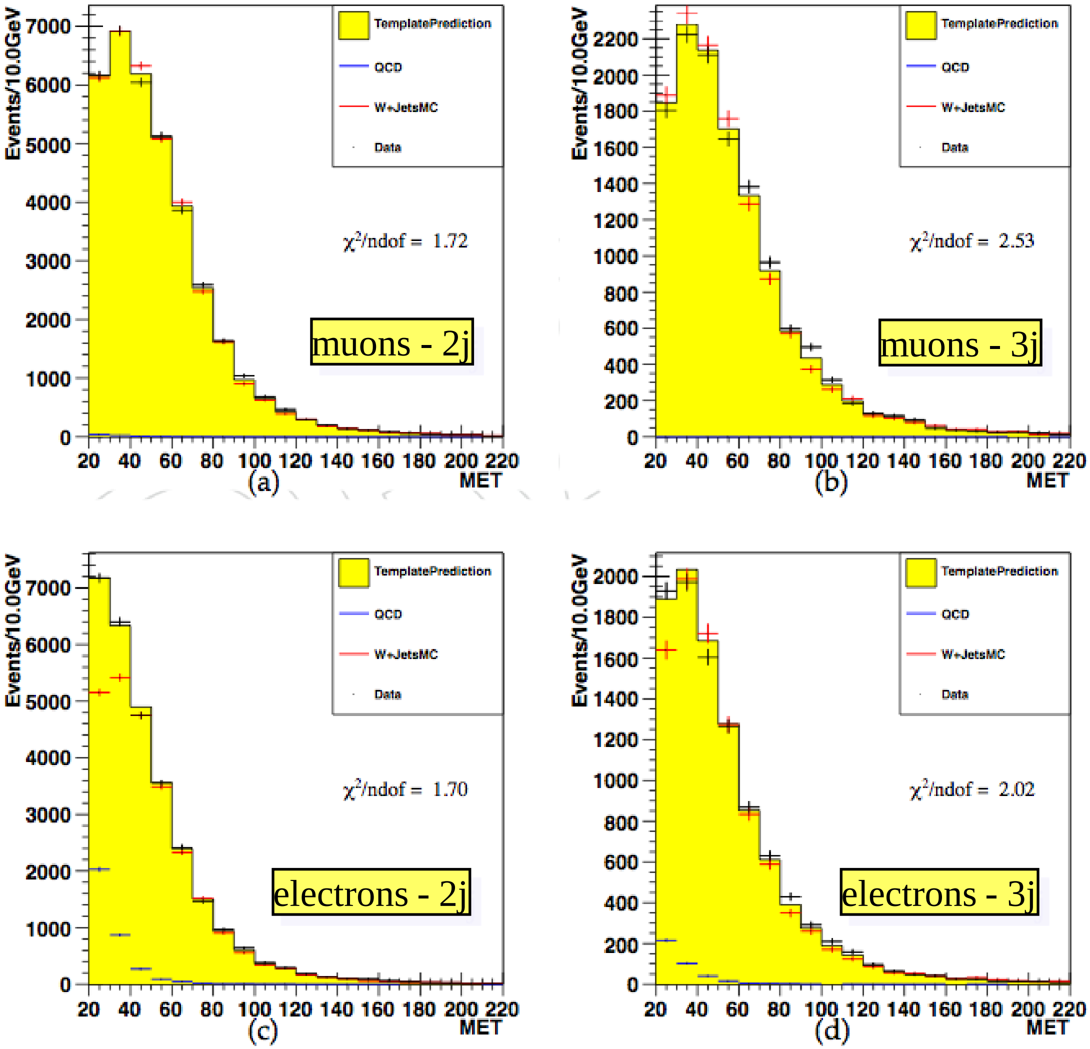
Process	Shape	External constraint on normalization
W+jets	MC/data	Unconstrained
Diboson	MC	Constrained: (NLO) 61.2 pb ± 10% [47]
$t\bar{t}$	MC	Constrained: (NLO) 163 pb ± 7% [48]
Single top	MC	Constrained: (NNLO) [49–51] ± 5%
Drell-Yan+jets	MC	Constrained: (NLO, $m_{ll} > 50$ GeV) 3048 pb ± 4.3% [47]
Multijet	data	Constrained: $\not\!\!E_T$ fit in data ± 50% (100%) for electron (muon)

Data driven QCD estimation

- We derive the QCD shape and normalization from the data.
  - invert the isolation requirements
  - relax ID requirements
  - relax the MET cut
- We can fit the MET distribution in data to get the normalization of the QCD contribution after accounting for differences due to the MET cut.
- The shape is also taken from this data as the MC is statistics starved.



QCD E_T^{miss} fit results



QCD fractions in data

	2 jets	3 jets
electron	6.2 ± 0.4%	2.1 ± 0.7%
muon	0.2 ± 0.4%	0.0 ± 0.4%

statistical fit errors only!

Uncertainty on QCD normalization

	2 jets	3 jets
electron	50%	50%
muon	>100%	>100%

W+jets background

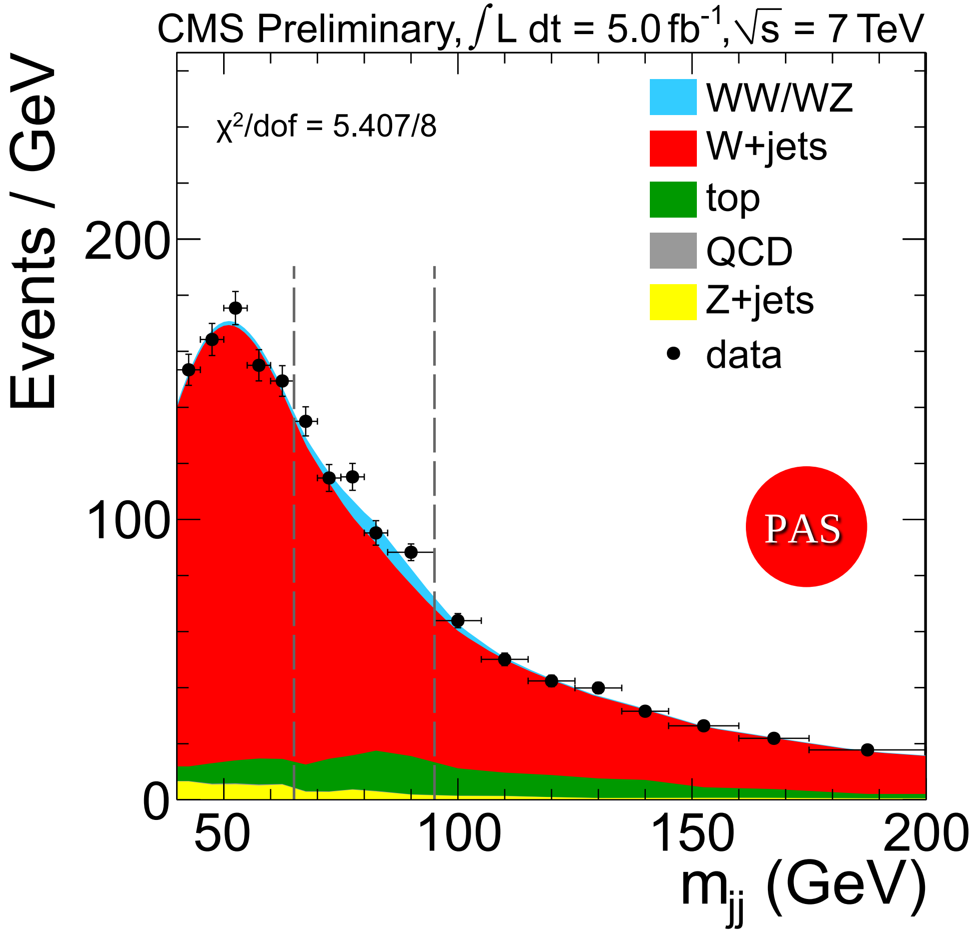
- The W+jets process is the dominant background.
- For $m_H \leq 180$ GeV, MC is used as the shape template, because statistics in this case are plentiful.
- For higher masses MC statistics are much lower so we take an analytic approach.

- The shapes are inspired by MC but ultimately the parameters of the functions are determined from the fit to the data.

$$
\begin{aligned}
\mathcal{F}_{\mathrm{W+jets}} &= \mathrm{erf}(m_{jj}; m_0, \sigma) \\
&\times \left[(m_{jj})^{-\alpha - \beta \ln(m_{jj}/\sqrt{s})} \right] \\
\mathcal{F}^{\mathrm{low\ mass}}_{\mathrm{W+jets,\ 2\text{-}jets}} &= \mathrm{erf}(m_{jj}; m_0, \sigma) \\
&\times (m_{jj})^{-\alpha} \times e^{m_{jj}\tau} \\
\mathcal{F}^{\mathrm{low\ mass}}_{\mathrm{W+jets,\ 3\text{-}jets}} &= (m_{jj})^{-\alpha - \beta \ln(m_{jj}/\sqrt{s})} \\
&\times e^{m_{jj}\tau}
\end{aligned}
$$

The background composition

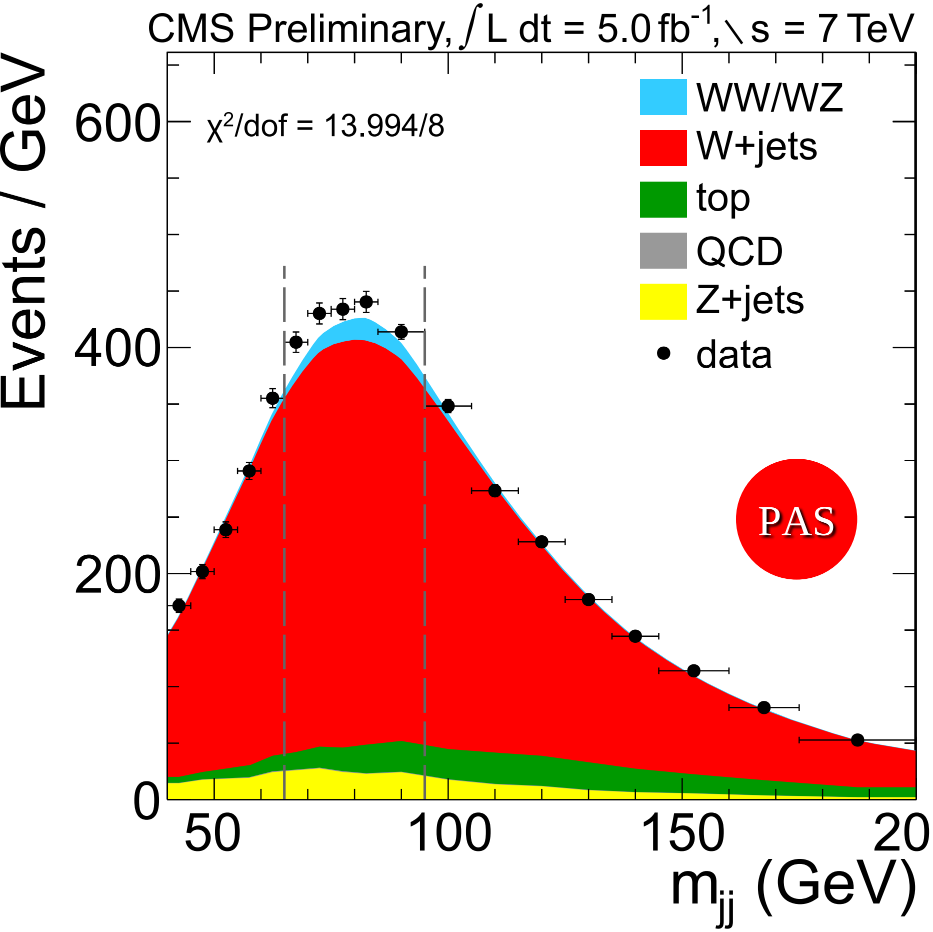
- **We exclude the region from 65 – 95 GeV excluded from the fit.**
- We see good agreement between the fitted composition and the data.
- The normalizations are extrapolated into the signal region of the fit and passed to the next stage for limit setting.



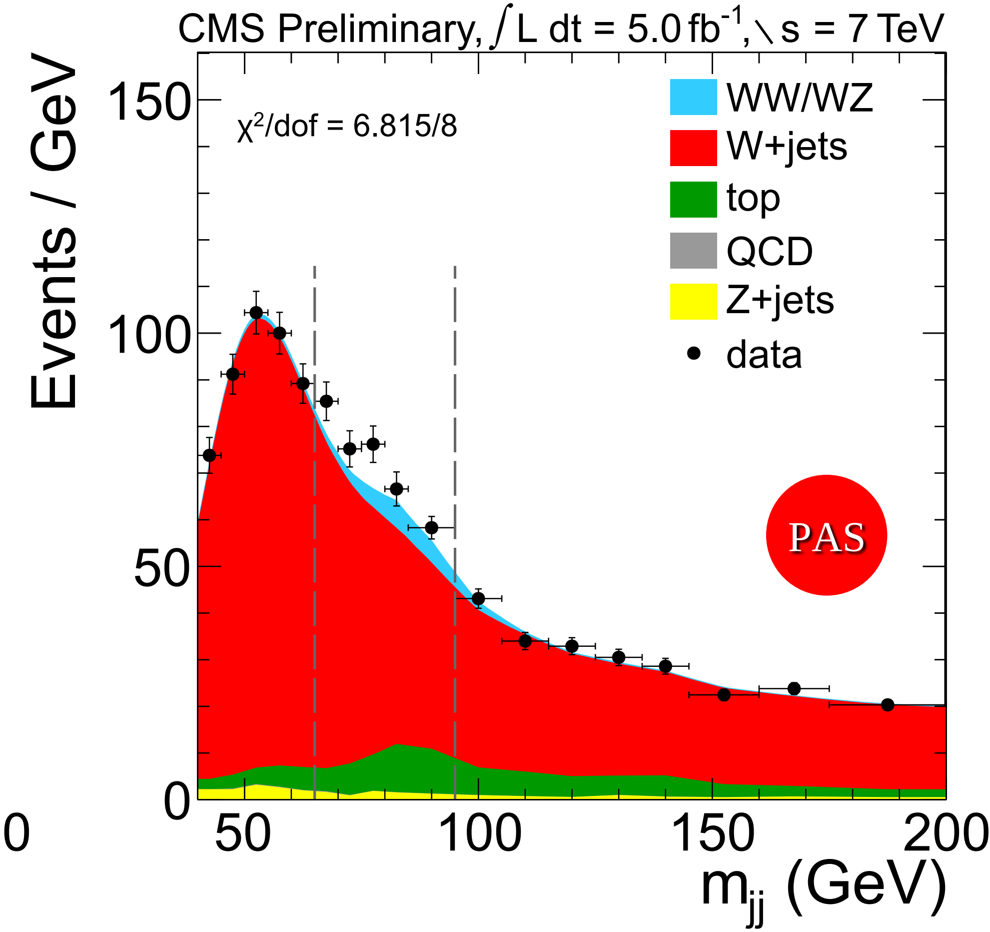
muon, 2-jet, m_H = 350 GeV

Additional m_{jj} examples

muon, 2-jets, m_H = 190 GeV



muon, 2-jets, m_H = 500 GeV



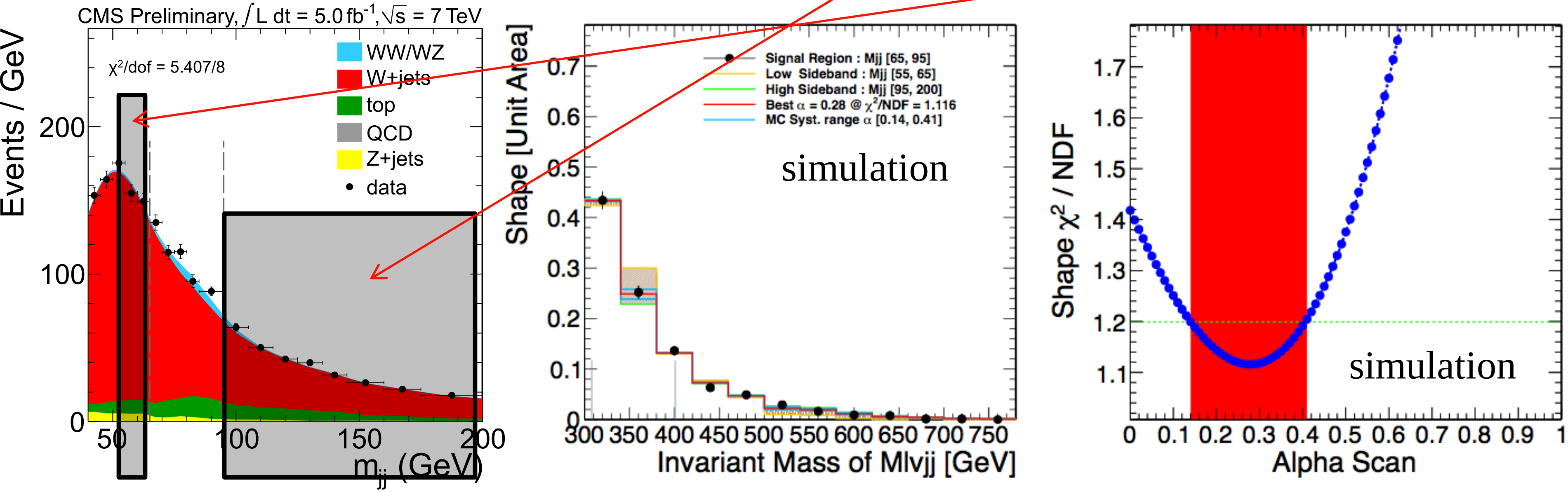
Limit setting with the $\ell\nu jj$ spectrum

- The fit to the m_{jj} spectrum determines the relative normalization of the backgrounds.
- The background components are stacked up and compared with the data with the additional selection $(65 < m_{jj} < 95)$ GeV.
- The shapes of the minor backgrounds are taken from MC.
  - Again, QCD is taken from the data-driven sample.
  - The W+jets shape is constructed from the m_{jj} sidebands.

The $\ell\nu jj$ W+jets shape

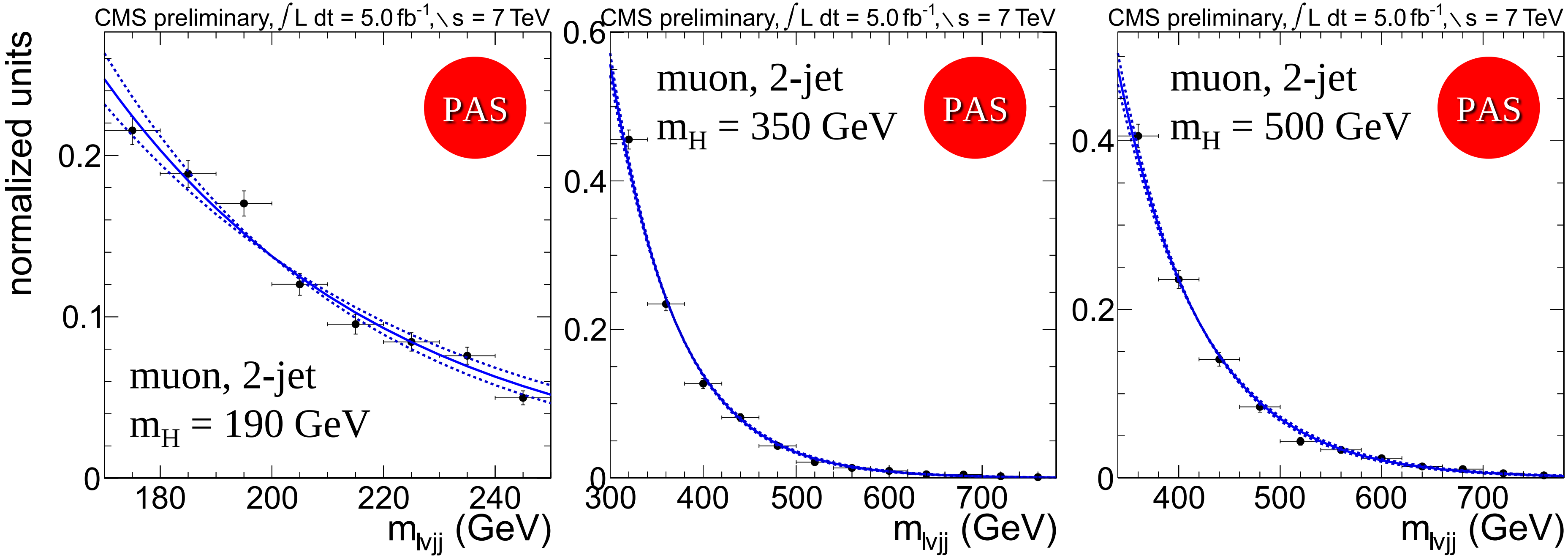
- We use the background-subtracted sidebands to derive a data driven W+jets shape.
    - separate α for each of the 48 channels
    - α determined from MC
    - uncertainty on alpha from MC used as systematic error.

$$\mathcal{F}^{\text{W+jets}}_{\ell\nu jj} = (1-\alpha)\mathcal{F}_{\text{HSB}} + \alpha\mathcal{F}_{\text{LSB}}$$

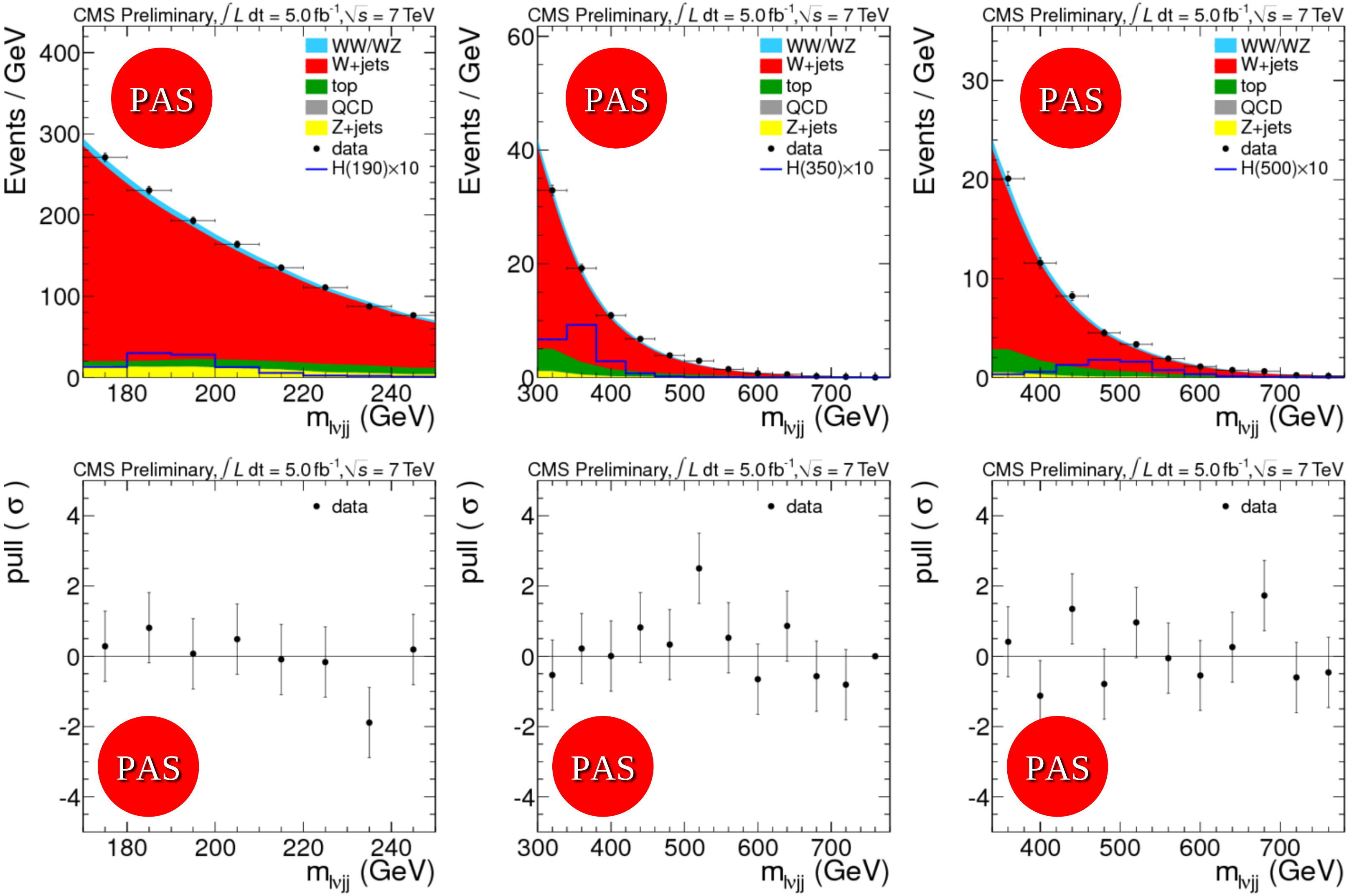


The $\ell\nu jj$ W+jets shape (II)

- The background-subtracted, alpha-combined sideband shape from data is smoothed using an exponential function.
  - The statistical uncertainty of the smoothing is combined with the uncertainty due to α and used as a systematic error.
  - The dotted lines are the total shape systematic envelope.



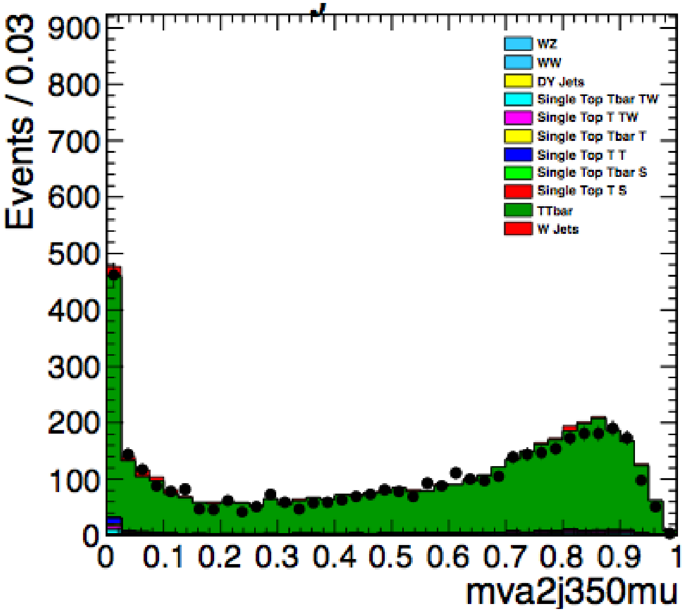
Final distributions



Signal systematic uncertainties

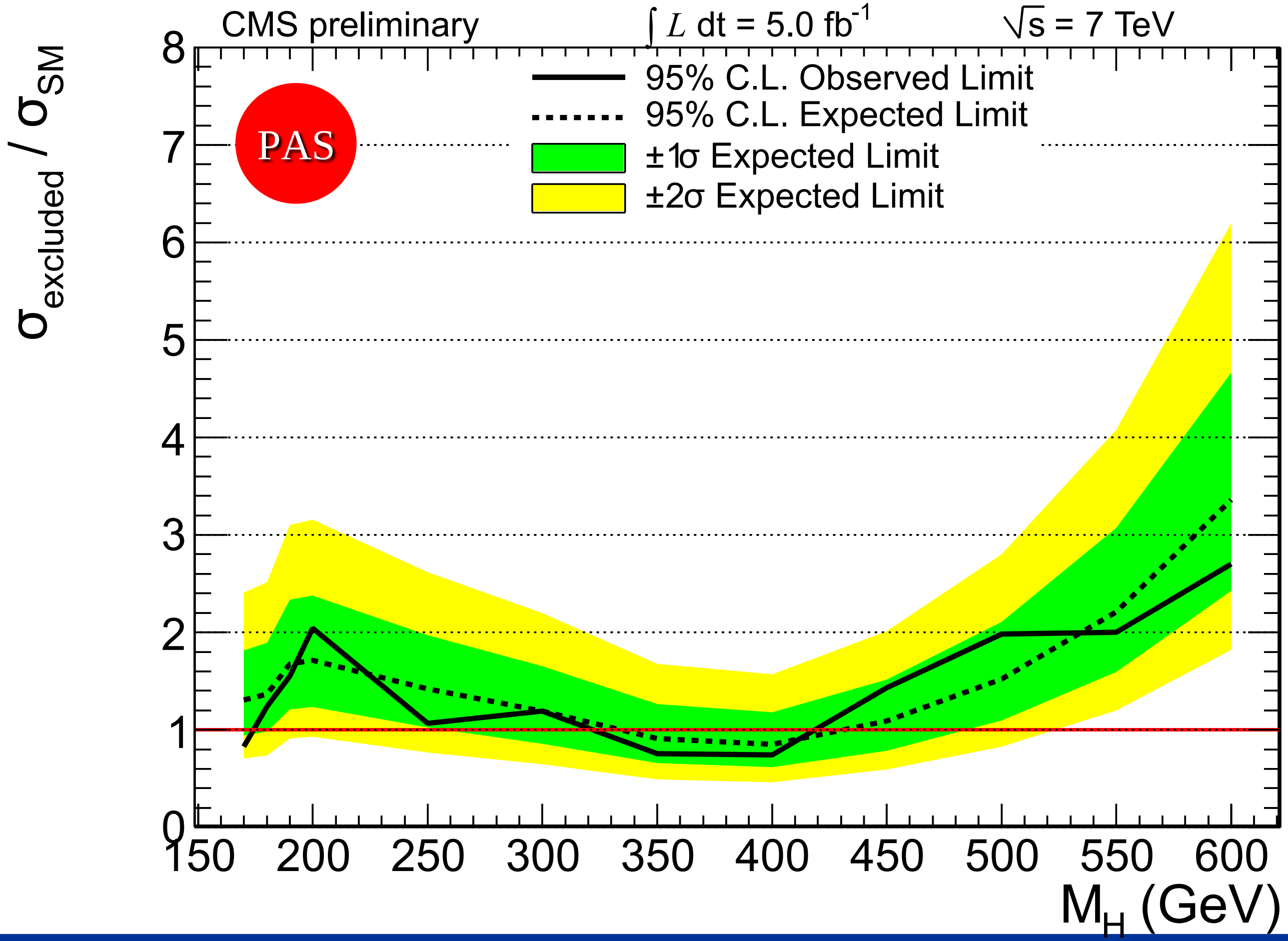
Source	**uncertainty**
Higgs line shape	0 – 30%
Signal cross-section	15 – 20%
Likelihood selection	7% (13% for 500 GeV and up)
Luminosity	2.2%
Jet energy scale, resolution and MET	< 1%
Theory (PDFs)	1 – 2%
Lepton trigger efficiency	1%
Lepton selection efficiency	1 – 2%
Pile-up	< 1%

- The likelihood selection systematic uncertainty is evaluated using a pure top sample as the difference between data and MC.



The final limit

- expected exclusion: 330 – 440 GeV
- observed exclusion: 320 – 420 GeV

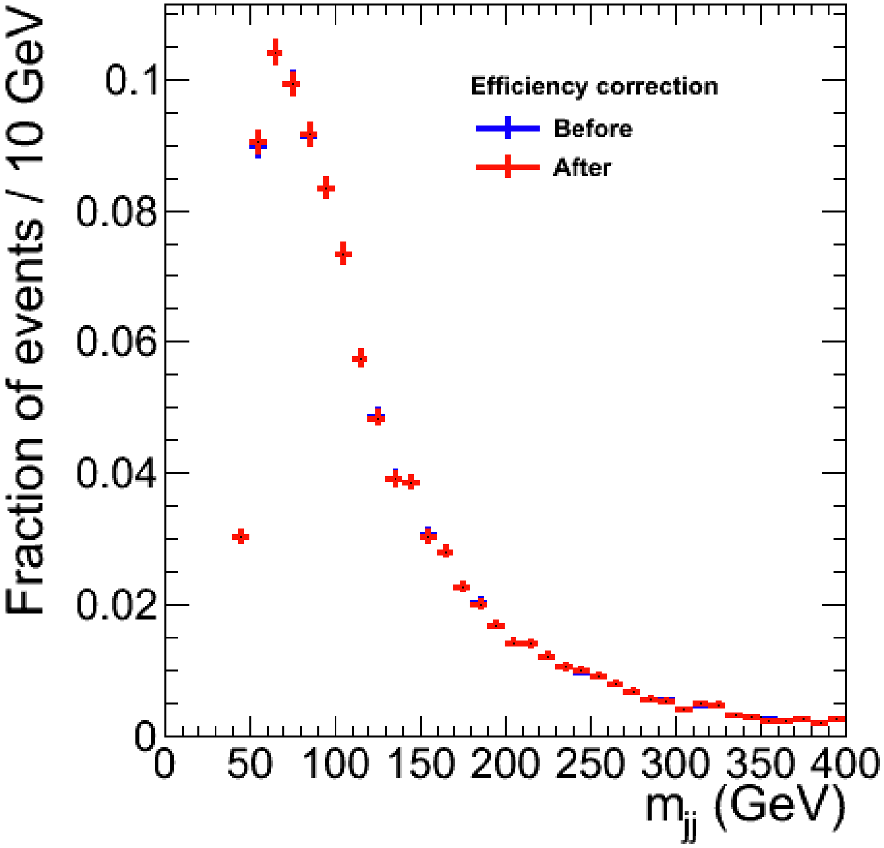


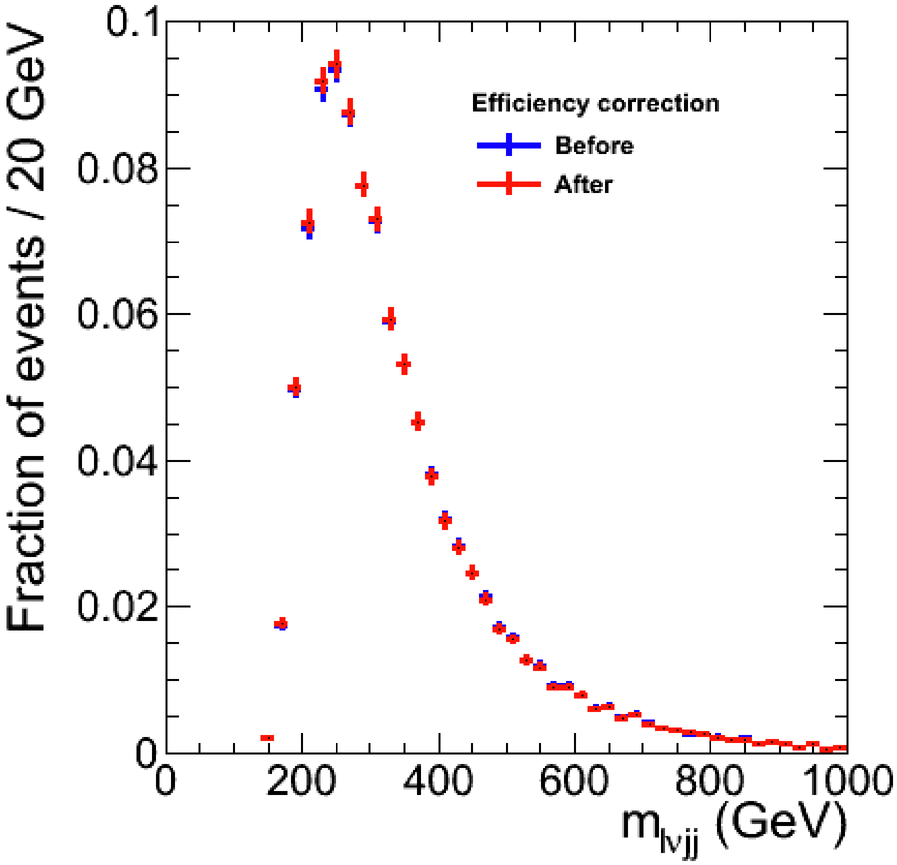
Conclusions

- We have set a world leading limit in $H \rightarrow WW \rightarrow \ell\nu jj$ decays.
    - We exclude the Standard Model Higgs boson in the mass range 320 – 420 GeV.
    - We employ data driven estimates for our principal background.
    - The fit-based analysis provides a basic robust limit.
    - We have used a multivariate discriminator to control the backgrounds and improve sensitivity.
    - The general features of both analyses mutually reinforce their conclusions.
    - We are poised to take advantage of the added statistics in 2012 to increase the mass reach of this important analysis.

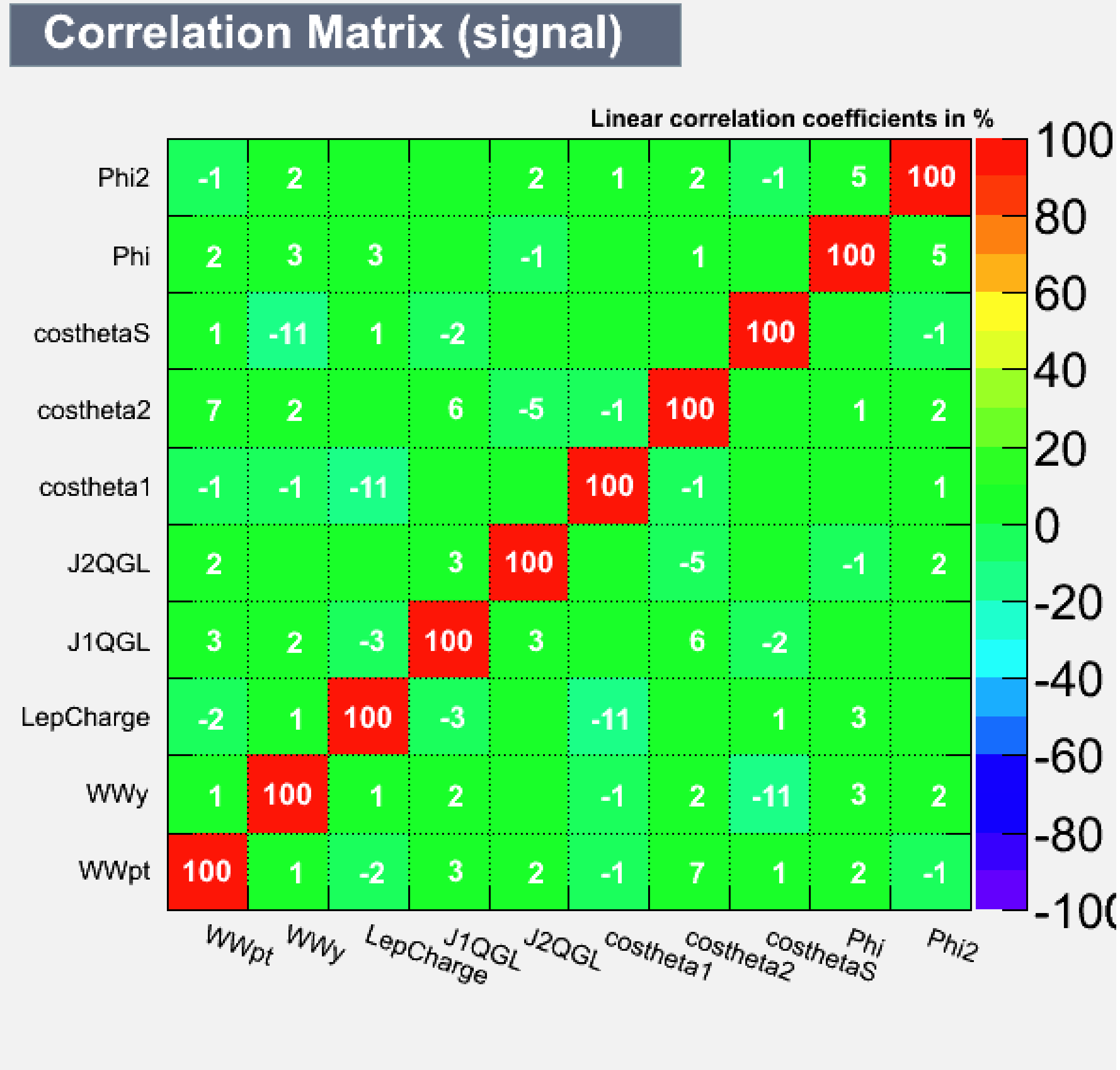
Backup

Trigger effect on key distributions





MVA correlations

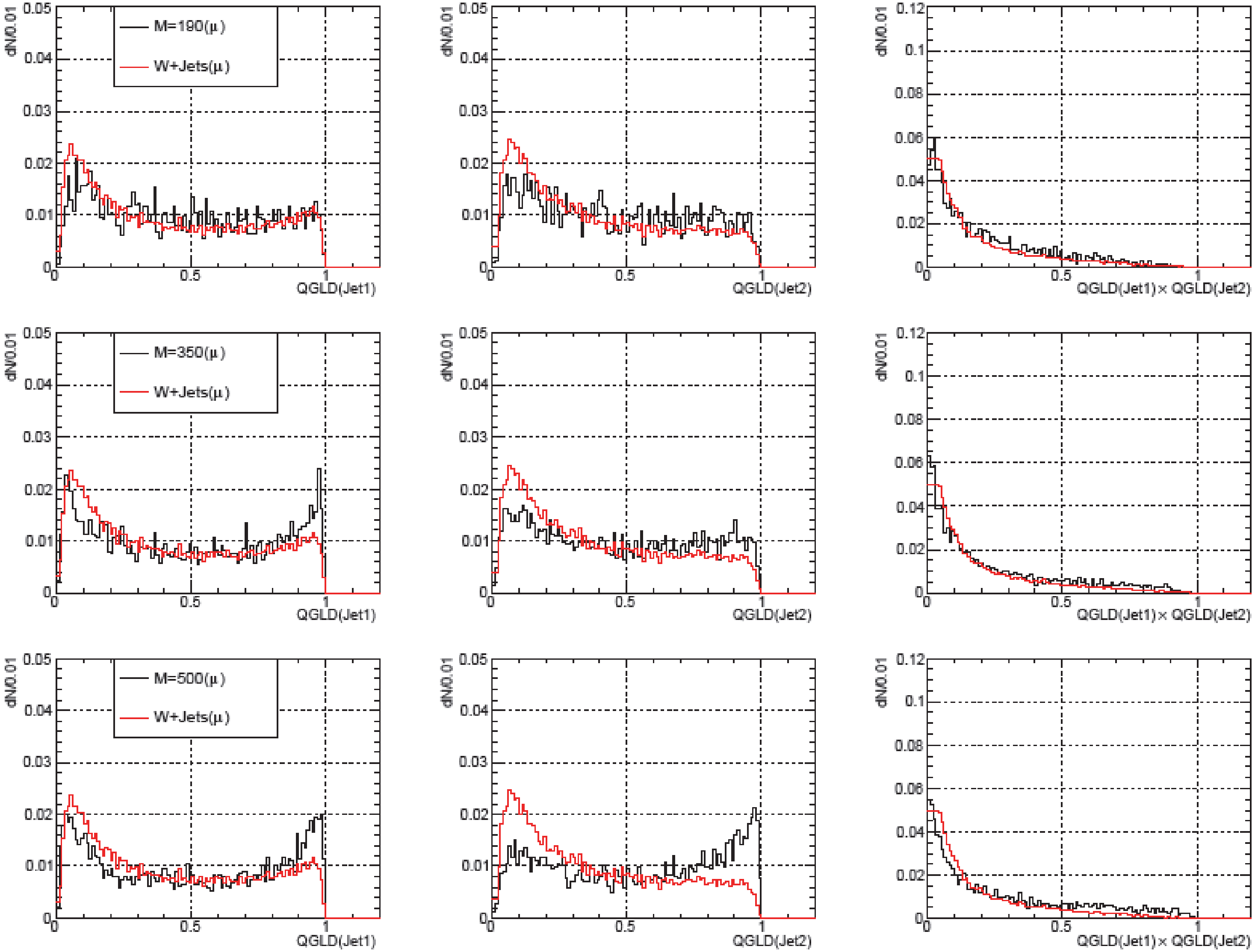


Correlation Matrix (signal)

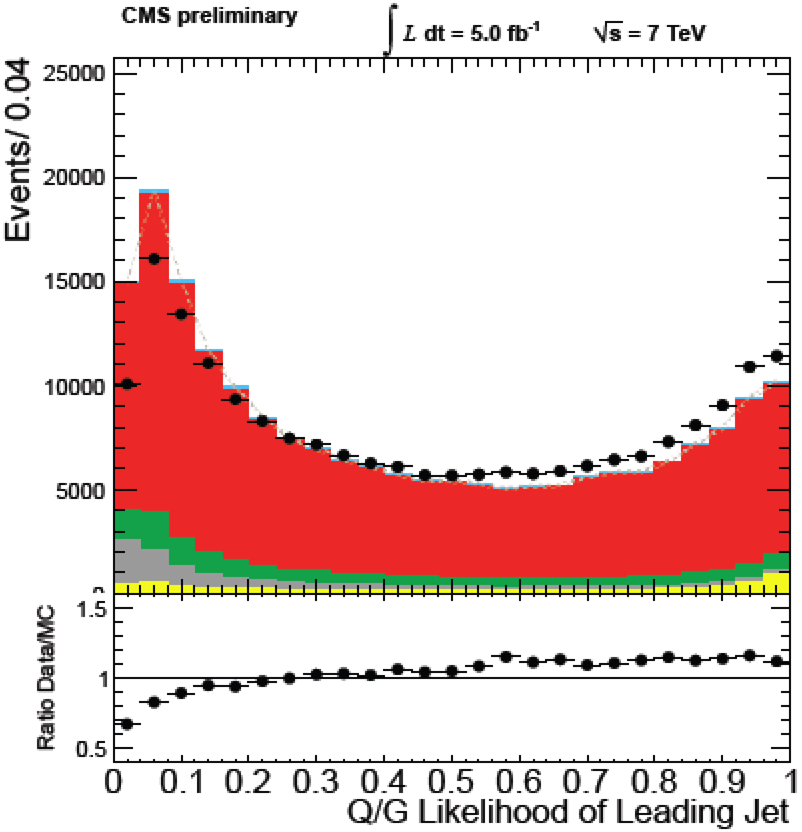
Linear correlation coefficients in %

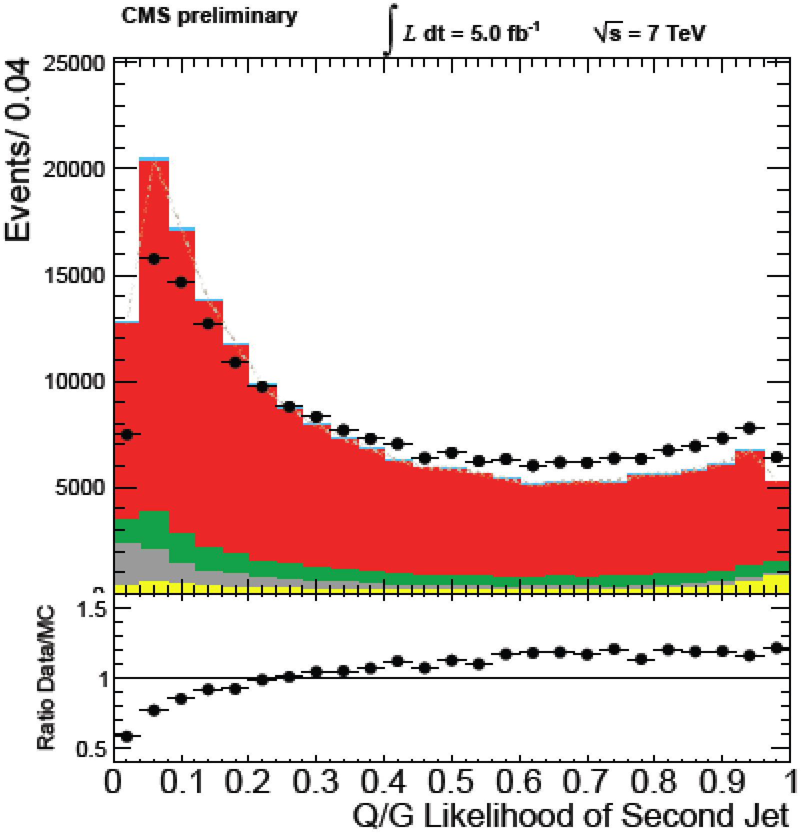
	WWpt	WWy	LepCharge	J1QGL	J2QGL	costheta1	costheta2	costhetaS	Phi	Phi2
Phi2	-1	2			2	1	2	-1	5	100
Phi	2	3	3		-1		1		100	5
costhetaS	1	-11	1	-2				100		-1
costheta2	7	2		6	-5	-1	100		1	2
costheta1	-1	-1	-11			100	-1			1
J2QGL	2			3	100		-5		-1	2
J1QGL	3	2	-3	100	3		6	-2		
LepCharge	-2	1	100	-3		-11		1	3	
WWy	1	100	1	2		-1	2	-11	3	2
WWpt	100	1	-2	3	2	-1	7	1	2	-1

q/g likelihood



q/g likelihood data / MC



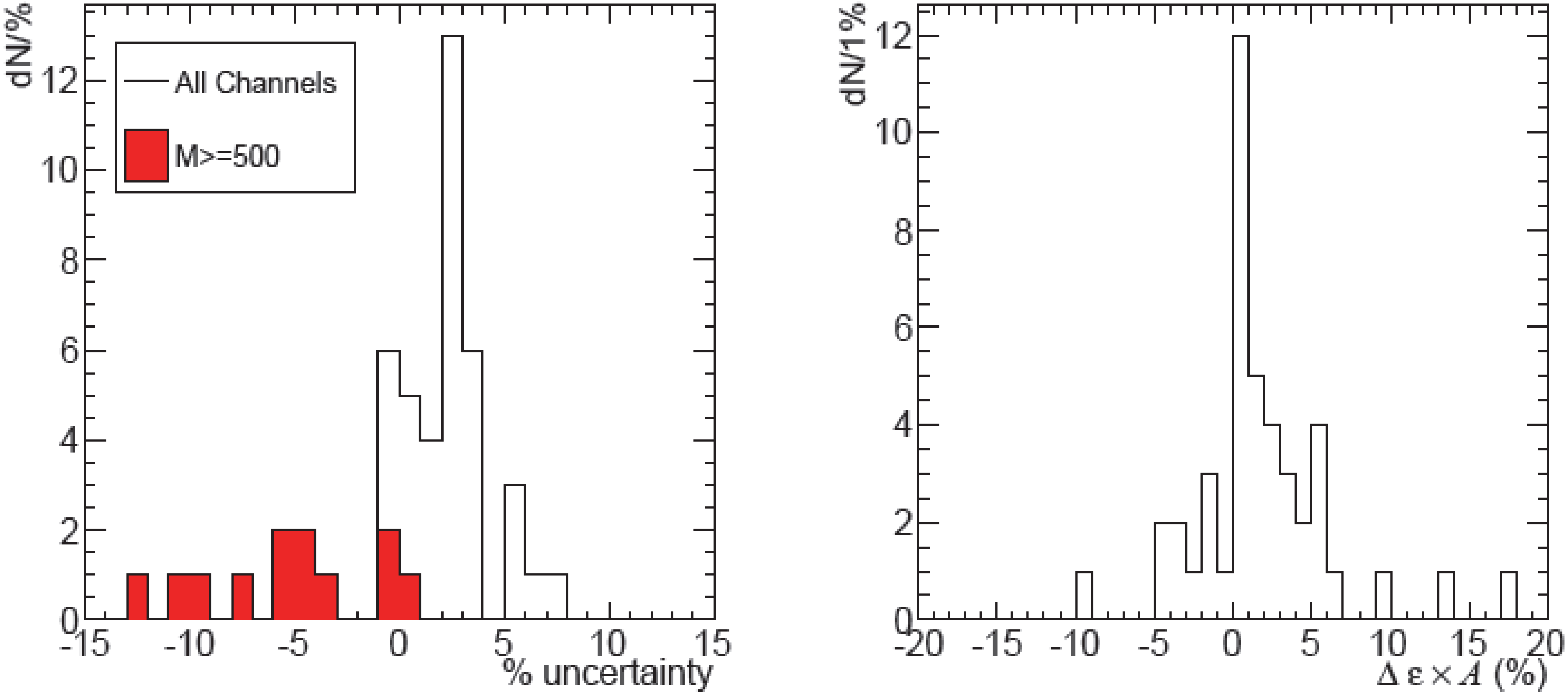


Initial and final yields

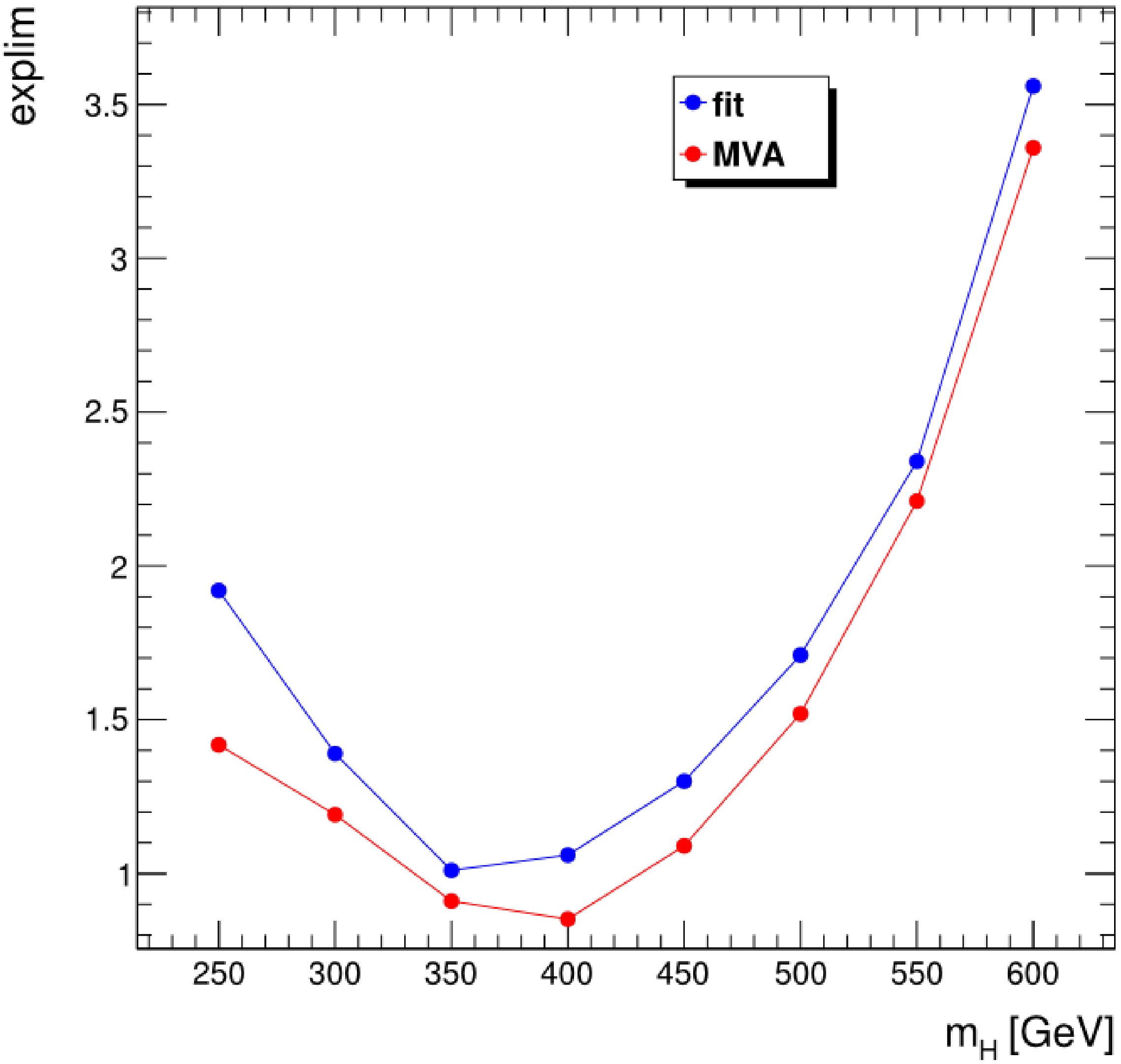
muon, 2-jets, $m_H = 350$

Parameter	Initial	Final
nDiboson	269.99	269.93 +/- 9.18
nQCD	17.258	16.769 +/- 44.6
nSingleTop	200.77	200.82 +/- 10
nTTbar	908.48	908.81 +/- 63.5
nWjets	10282	8828.5 +/- 167
nZjets	304.96	305.01 +/- 13.1

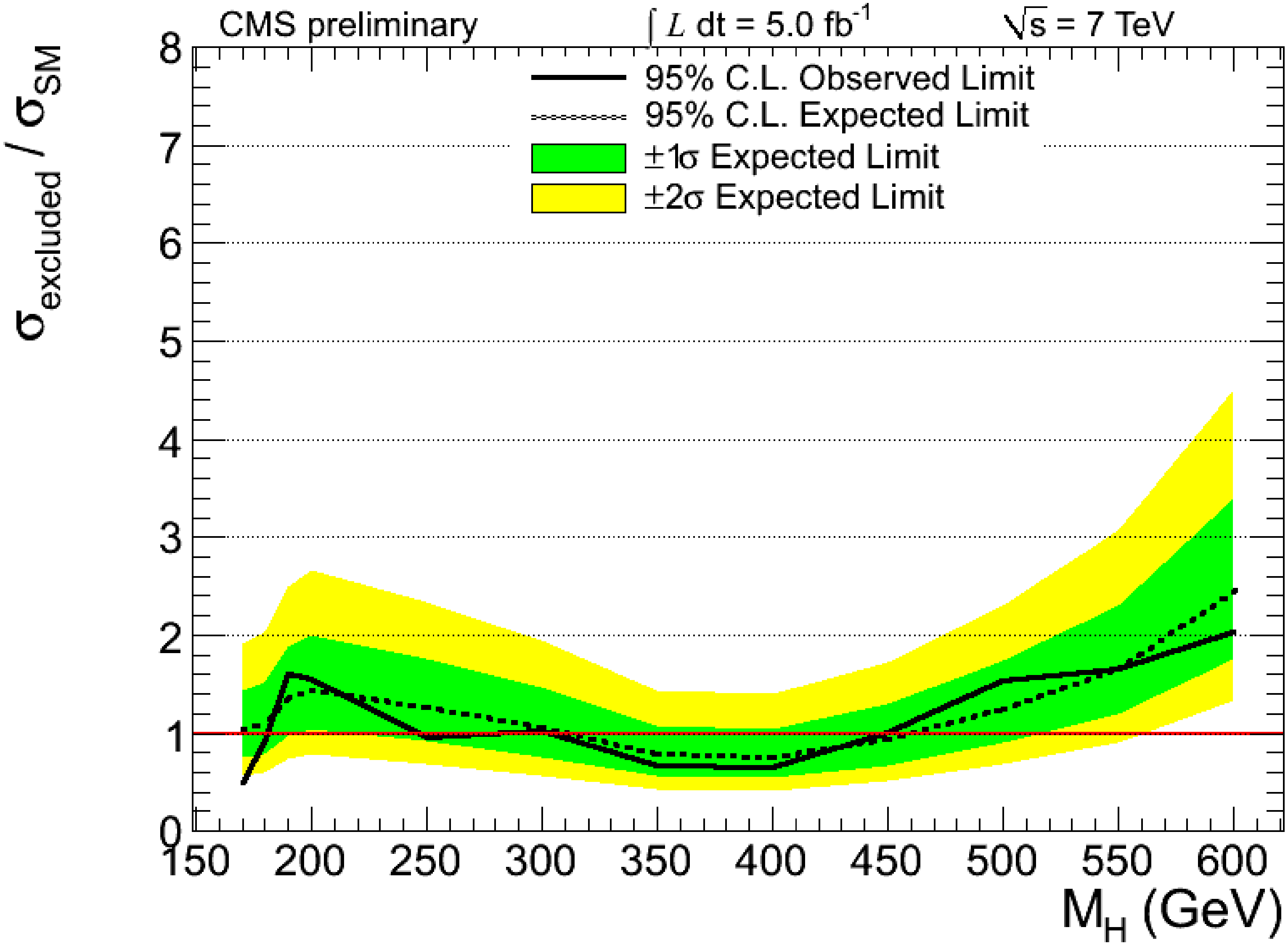
Likelihood selection efficiency



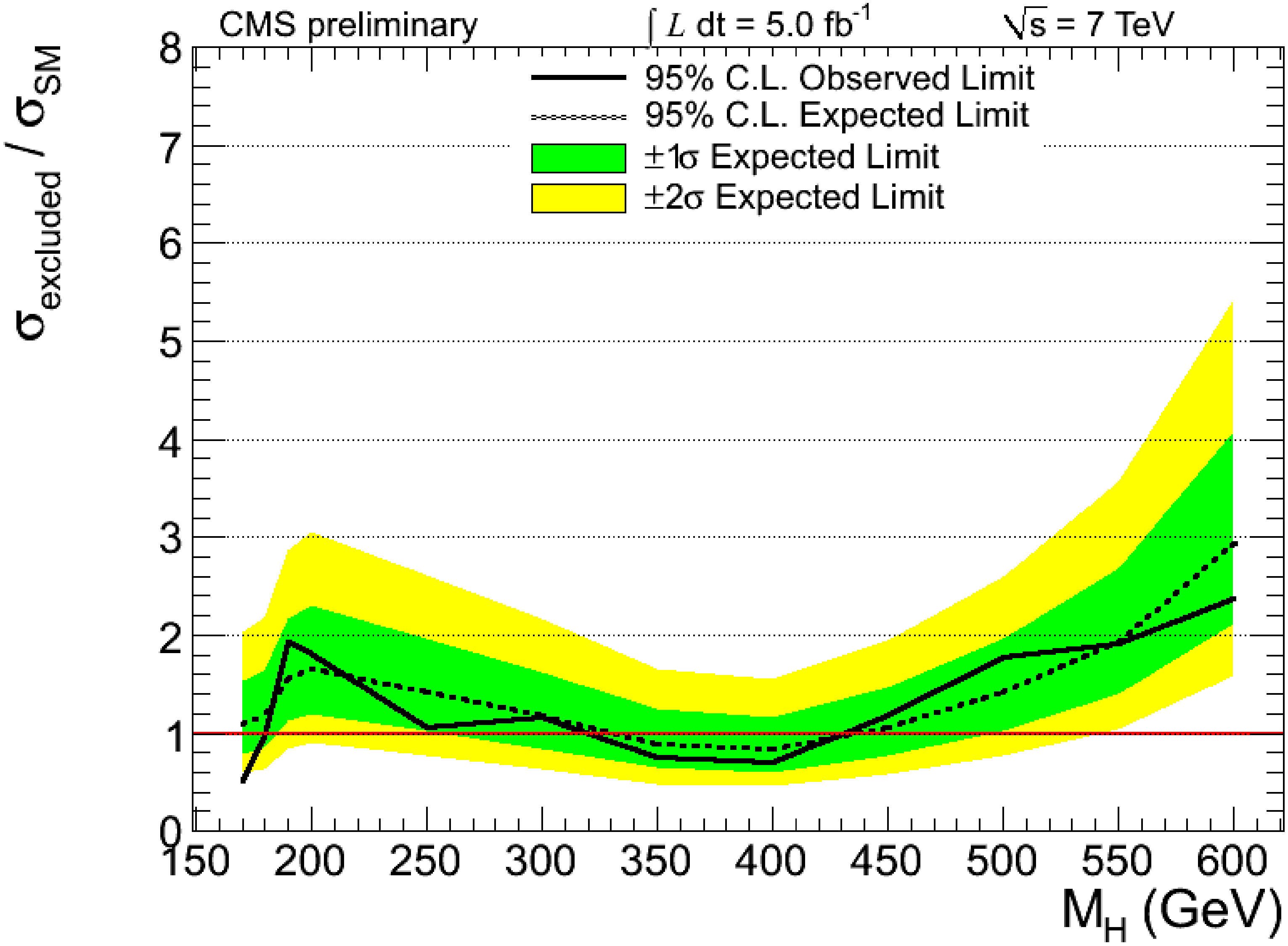
Fit vs. MVA


explim
3.5
3
2.5
2
1.5
1
250
300
350
400
450
500
550
600
m_H [GeV]
● fit
● MVA

No systematic limit



No shape systematic limit



all systematic limit