Preliminary analysis using 350 pb^{-1} data

On behalf of H→WW (lvjj) working group

(June 6, 2011)

Data & selection used in this presentation

Acceptance

- Tight lepton selection from top PAG
- Exactly two jets with p_T > 25 GeV (using PF2PAT cleaning)
- pf MET > 25 GeV
- W transverse mass > 50 GeV

4 kinematic cuts to suppress W+jets:

With just a single cut the S/B is low enough that the fit runs into instability. With three additional cuts S/B ~1/5.

- $|\Delta\phi(W, W) - \pi| < 0.3$ (our original single cut)
- $\Delta\eta(j1, j2) < 1.8$
- $\Delta\phi(j1, \text{muon}) > 2.1$, $\Delta\phi(j1, \text{electron}) > 1.8$
- $\Delta\phi(j2, \text{muon}) > 2.3$, $\Delta\phi(j2, \text{electron}) > 2.0$

These cuts are not necessarily optimal or final. Alexx Perloff and KM are working on this.

See Alexx's talk

Processed ~ 350 pb^{-1} of data so far (340 pb^{-1} for electron, 360 pb^{-1} for muon). Still use 4.1.X MC. Big concern: MadGraph W+jets MC is only about 700^{-1}, observe same statistical jittering in MC as in data. Hard to get good template.

We take m_{jj} and m_{lvjj} shape from MC

Problem

We do not have large enough W+jets MC sample to make a good template. The MadGraph sample corresponds to 700 pb^{-1} which is only ~ 2 times larger than our data size. Once we process full 0.6 fb^{-1}, the MC and data will have about the same statistics. This creates large statistical jitter if one takes shape from a simple uniformly-binned histogram of MC events.

Current solution

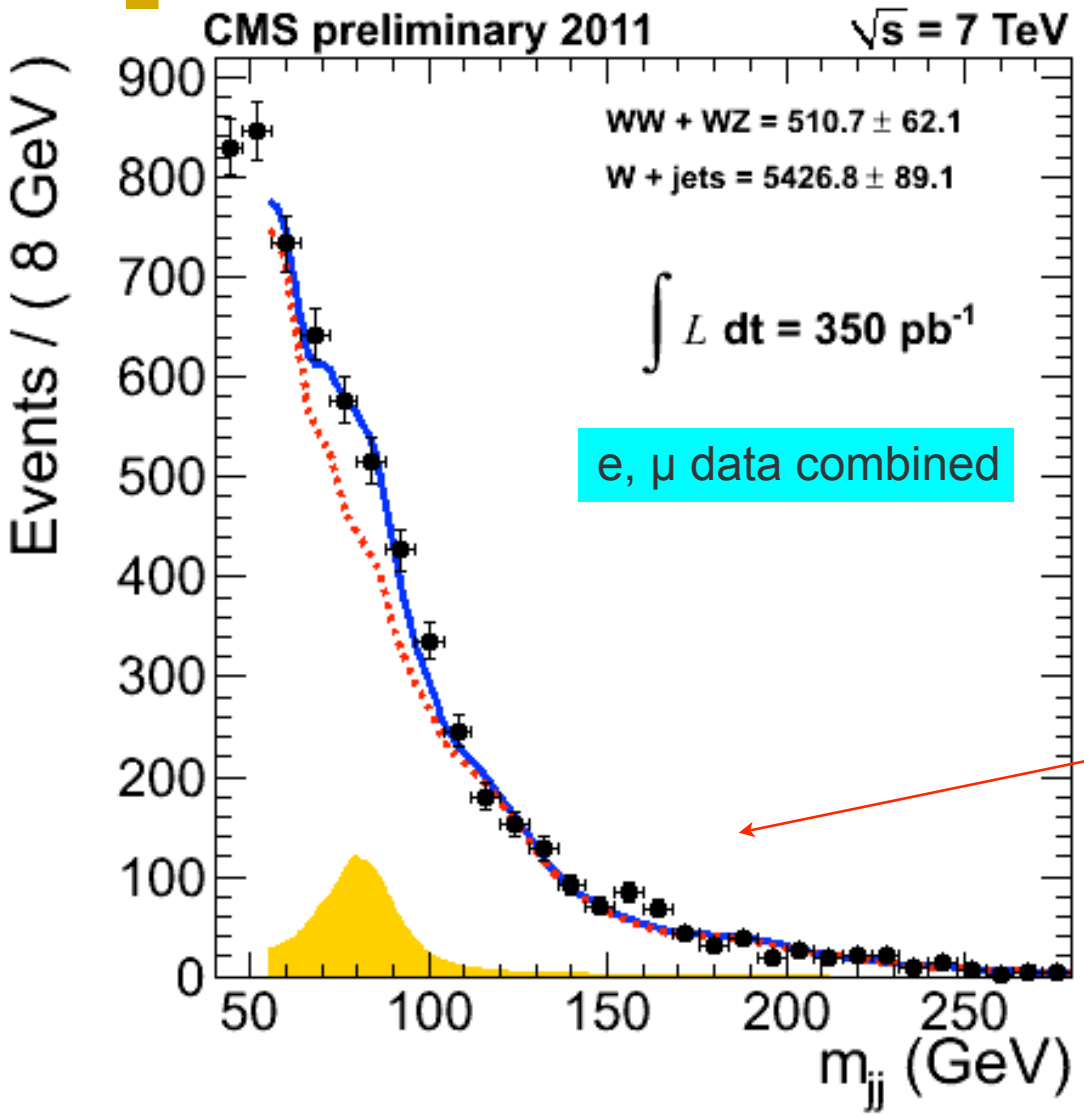
Instead of using fixed bin histograms to derive templates, I use a ROOT functionality called 'RooKeysPdf'. This class is useful if one has to deal with histograms with poor statistics and the trade-offs between having too large bins and having spikes in the plots. It's a class that behaves like a histogram, but internally saves the un-binned events and finally produces a smooth histogram.

Documentation of RooKeysPdf: http://root.cern.ch/root/htmldoc/RooKeysPdf.html

CMS Higgs combination group also uses this class for templates

see for example: HiggsAnalysis/CombinedLimit/interface/TH1Keys.h

Template fit to m_{jj} in W+2 jet events



MC predicts ~ 400 WW + WZ events

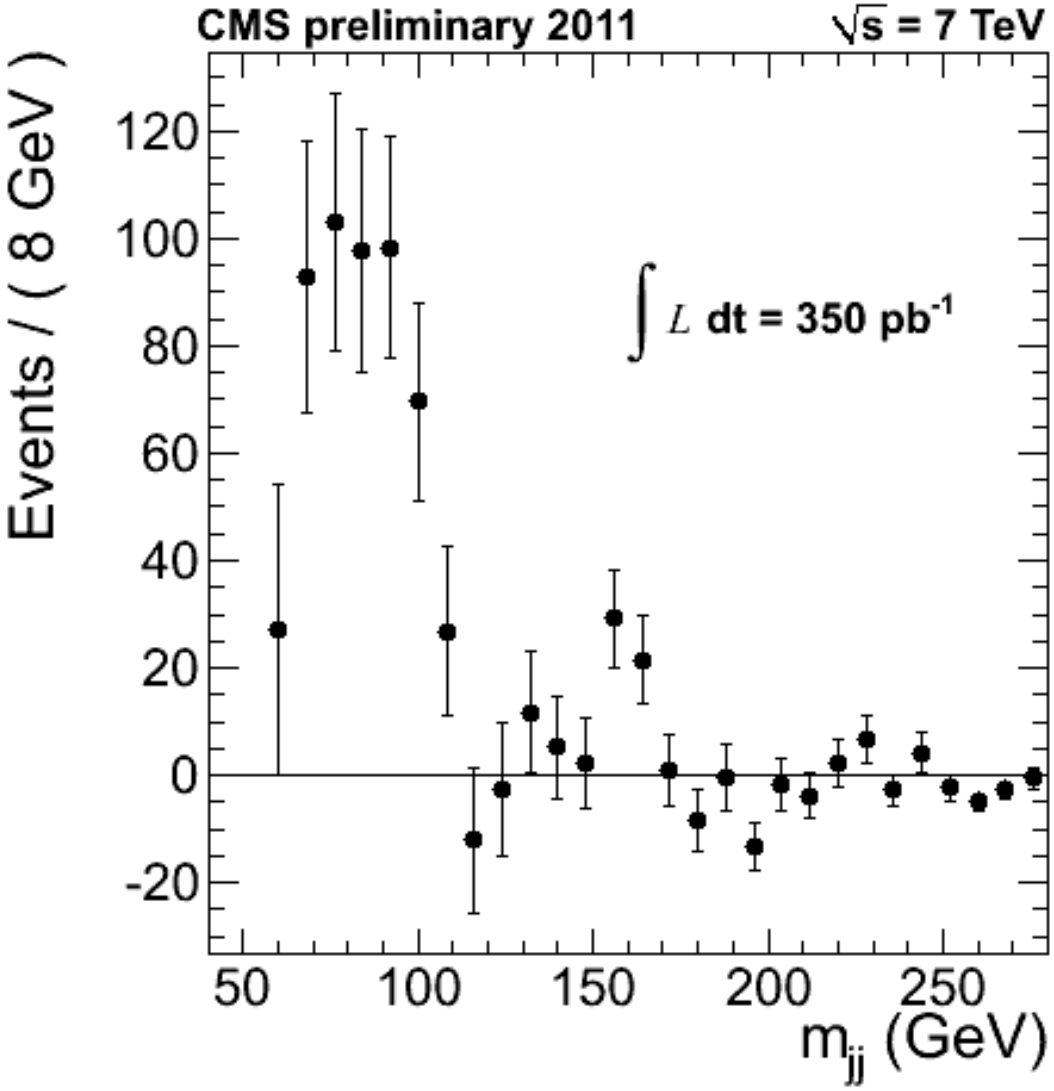
Take the shape from MC. Just fit for the normalization. Blue curve shows the fit to data, dashed red curve is W+jets component, shaded area in orange is di-boson component.

The fit is completely unbinned. Only the plots have binning.

Haven't included single top events in the fit which peak around 150 GeV. Working on it.

Background subtracted distribution on the next slide

m_{jj} in W+2 jet events after bkg subtraction

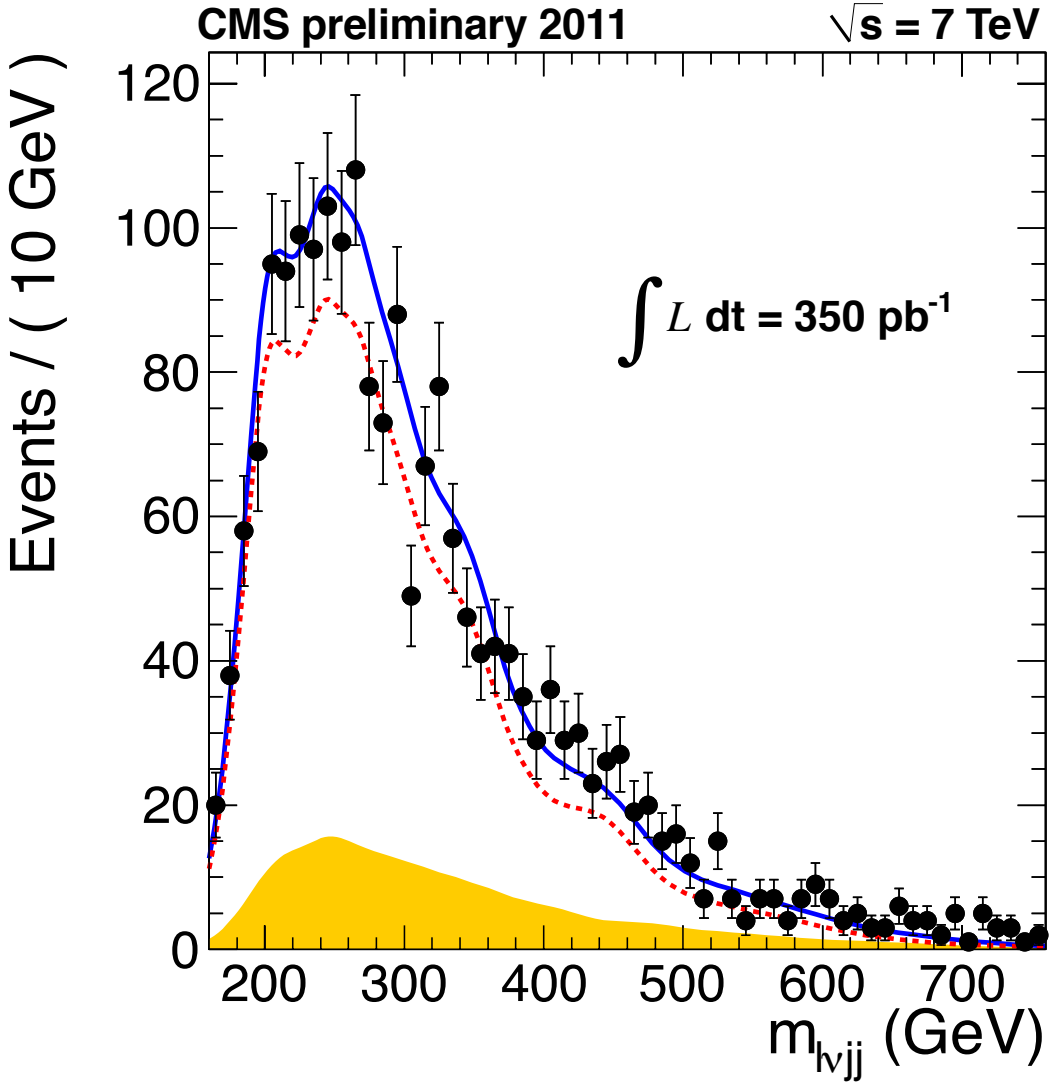


In the W mass window 65 < m_{jj} < 95 GeV we get:

338 di-boson events
1702 W+jets events

We will fix the normalization in the m_{lvjj} fit to above yields. On the next slide.

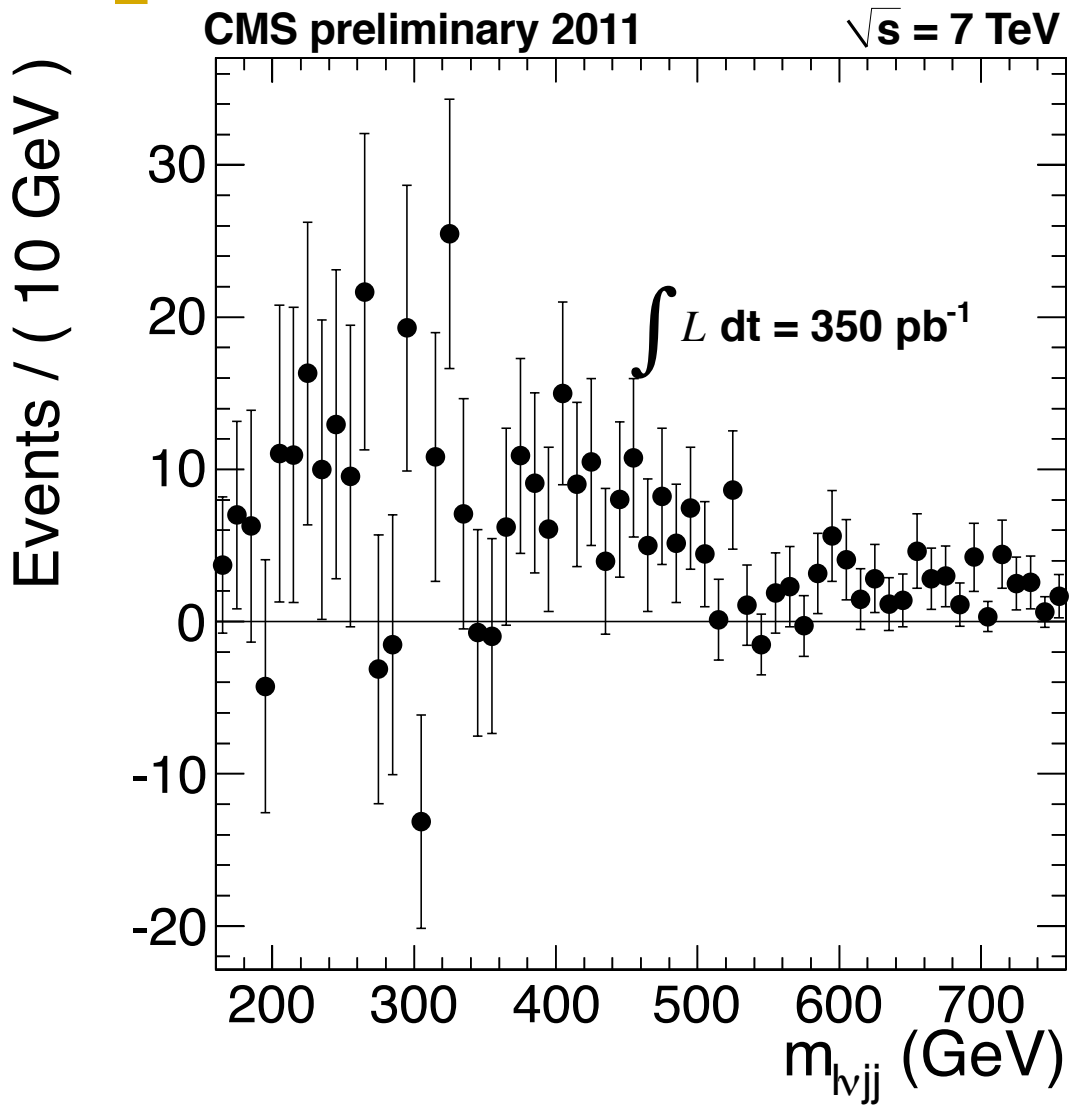
$m_{l\nu jj}$ distribution: $65 < m_{jj} < 95$ GeV



Take the shape directly from MC without any correction. Additionally, fix the normalization from fit to m_{jj} distribution in data.

Background subtracted distribution on the next slide

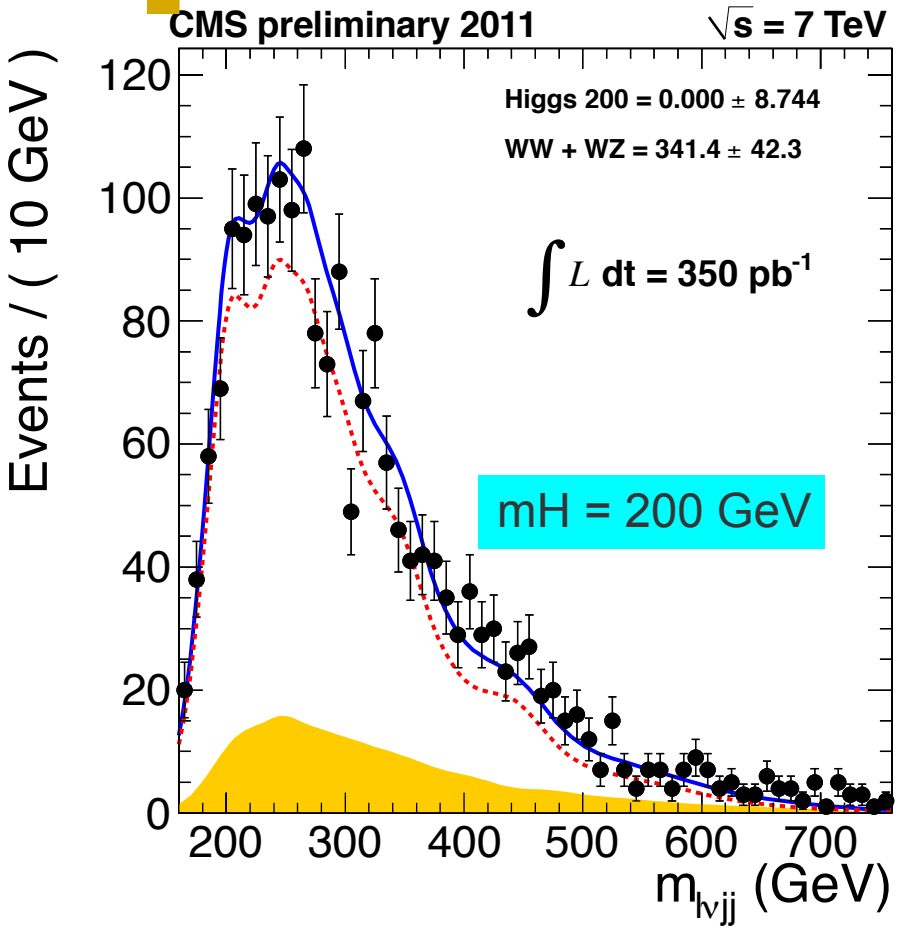
$m_{l\nu jj}$ distribution after bkg subtraction



Some local enhancements between 220−280 GeV and 400−500 GeV.

Now we will try to include various Higgs masses in the template fit.

Template fit of m_{lvjj} after including Higgs



Methodology

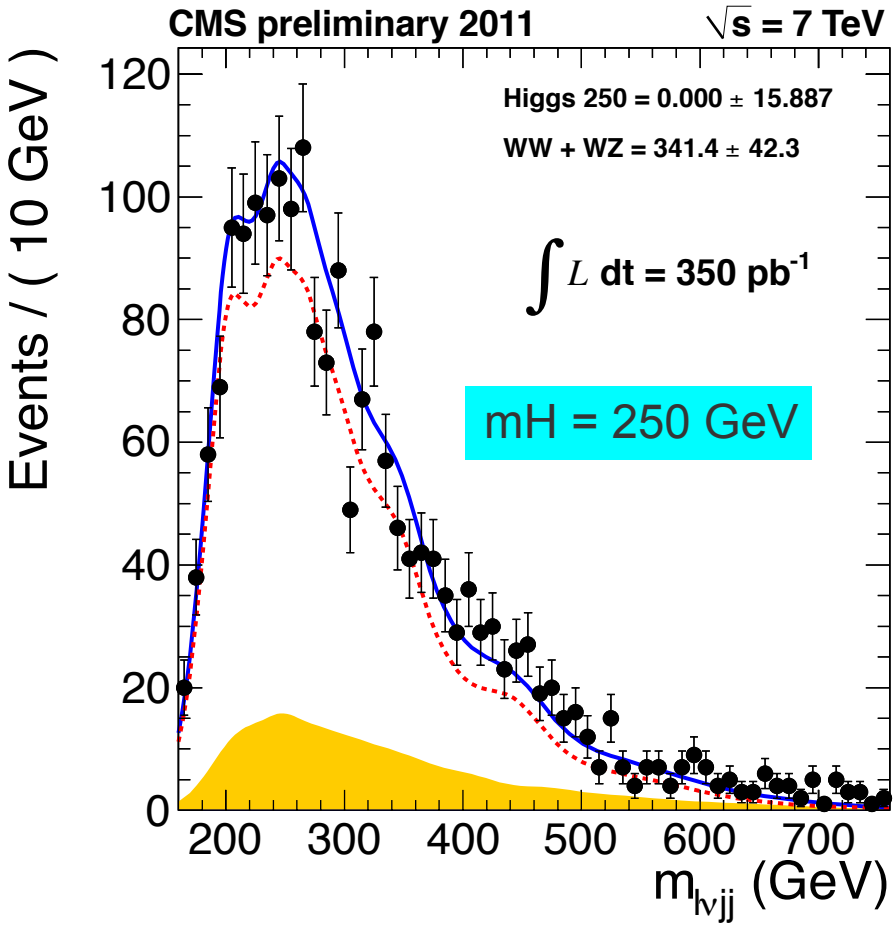
In the following slides I

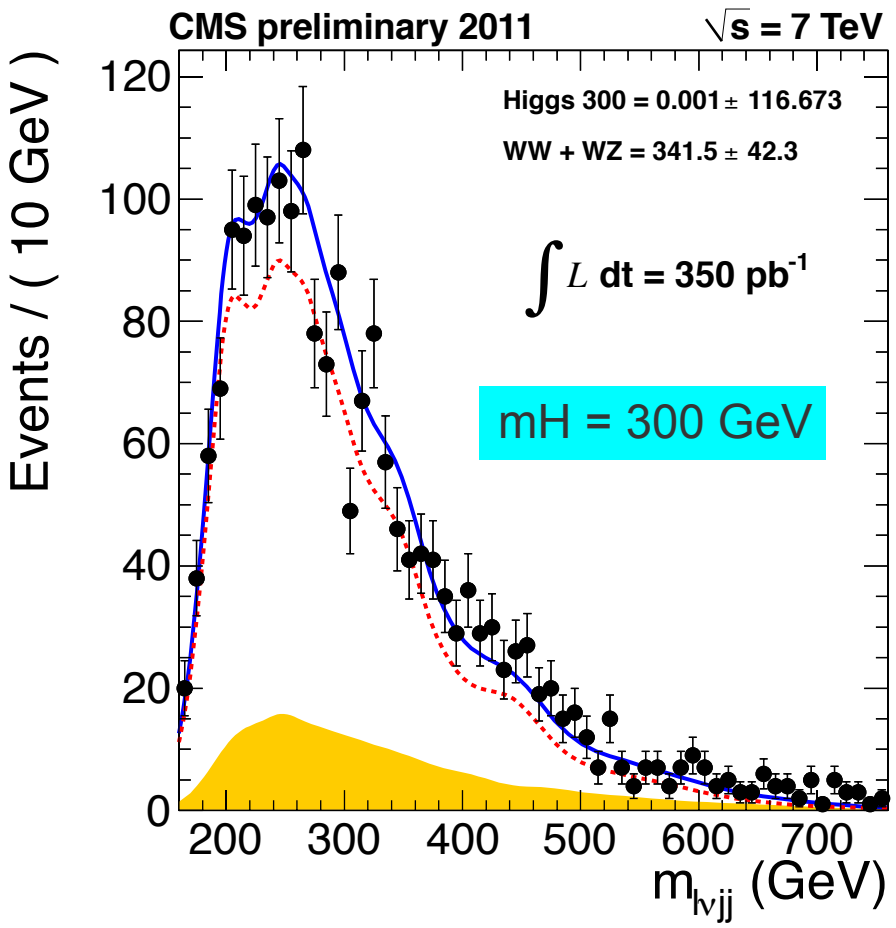
1.) Take W+jets, di-boson, and Higgs shapes directly from MC
2.) Fix the W+jets normalization to that obtained from m_{jj} fit to data
3.) Float the di-boson and Higgs yields

We will convert to Higgs yields into cross section (or cross section / SM predicted cross section) and thus set a limit on it.

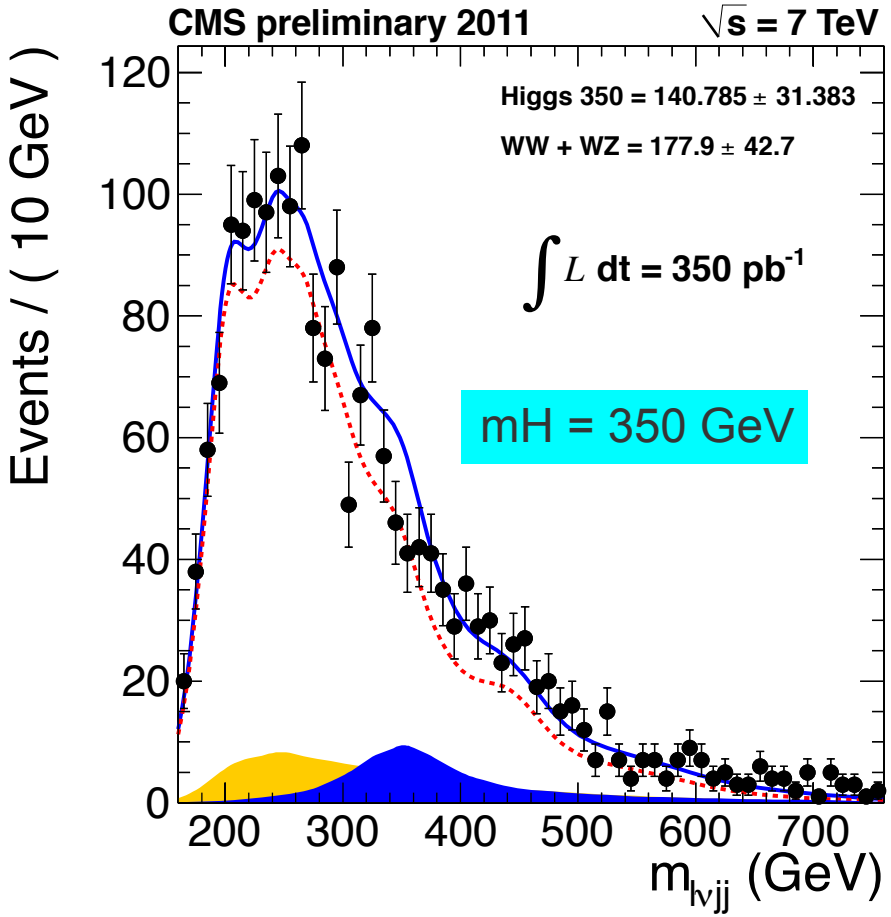
I should also fix the **sum** of di-boson and Higgs yields to the total hadronic W +Z yield from m_{jj} fit. I will do this in the next iteration of the fits. Meanwhile I do two extremes: fix the di-boson yield to the ones from m_{jj} fit or float it completely.

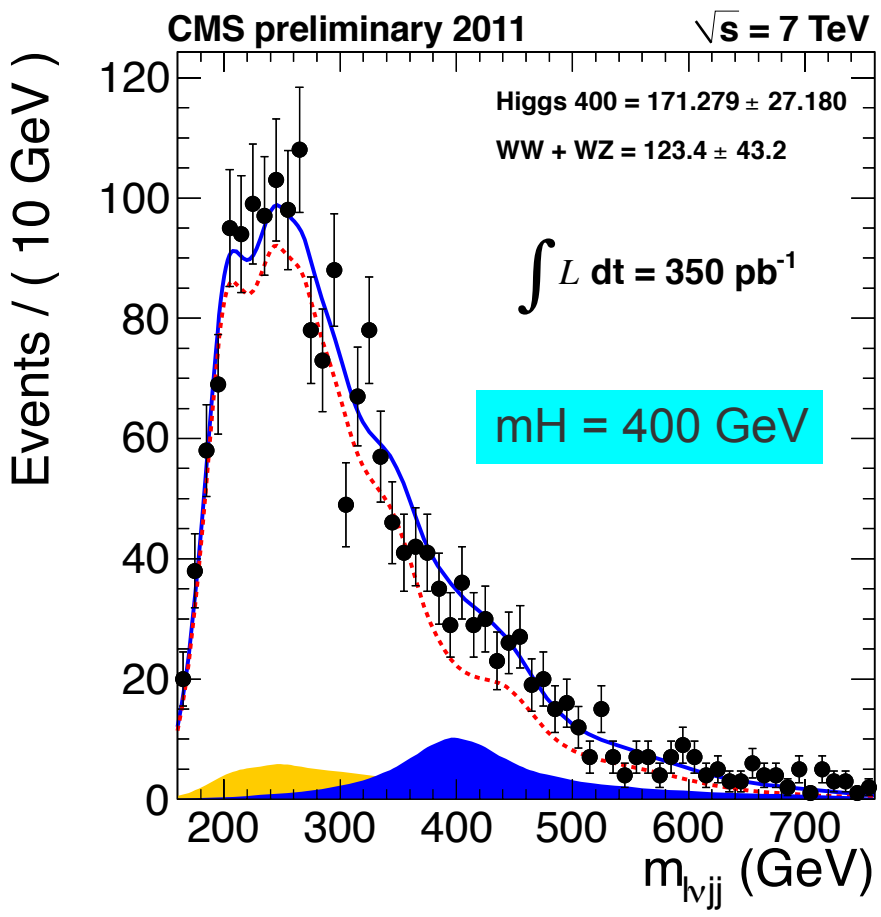
Template fit after including Higgs (1)





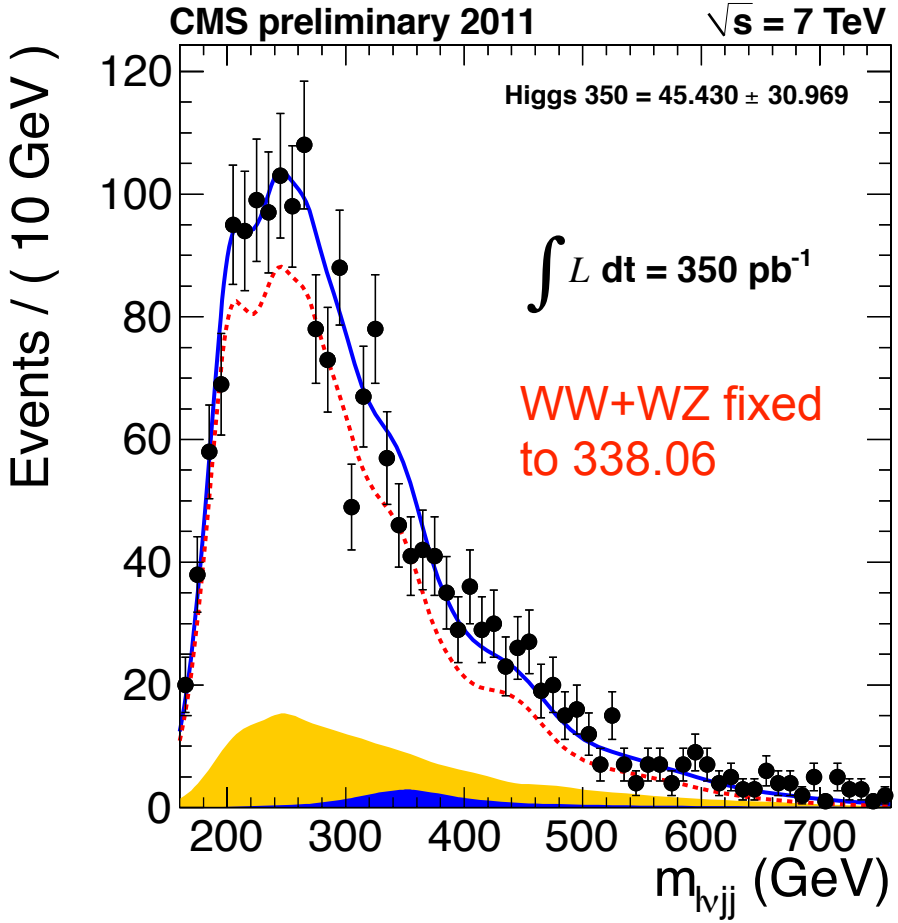
Template fit after including Higgs (2)

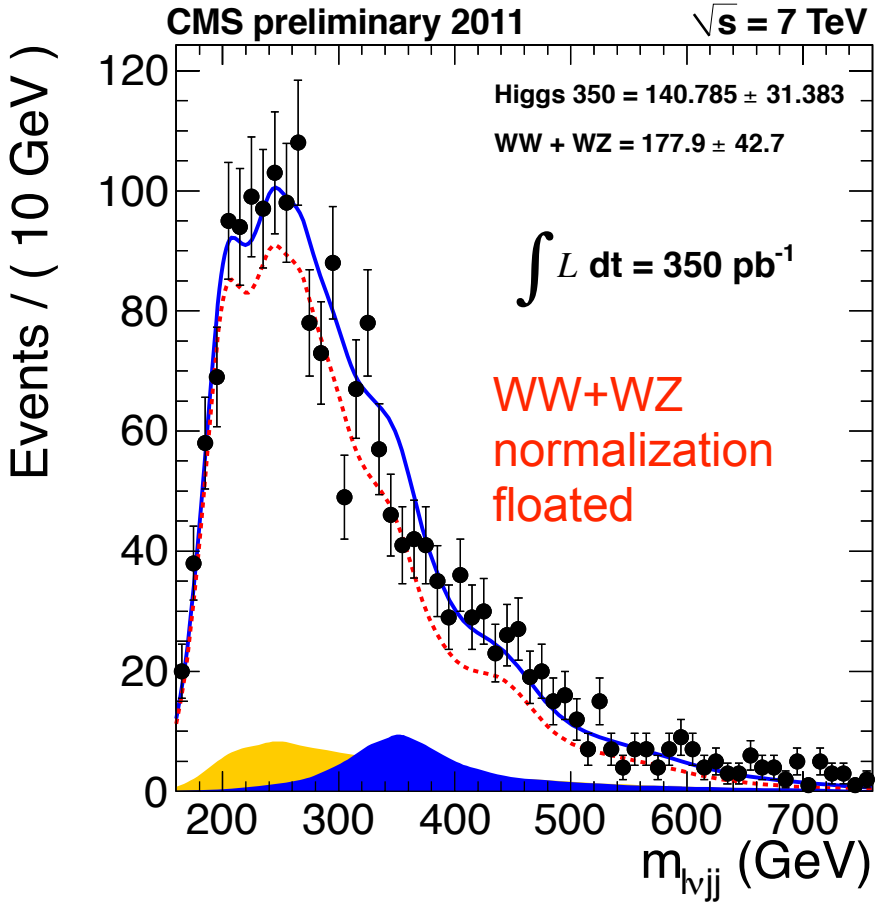




mH = 350 fit: effect of constraining WW yield

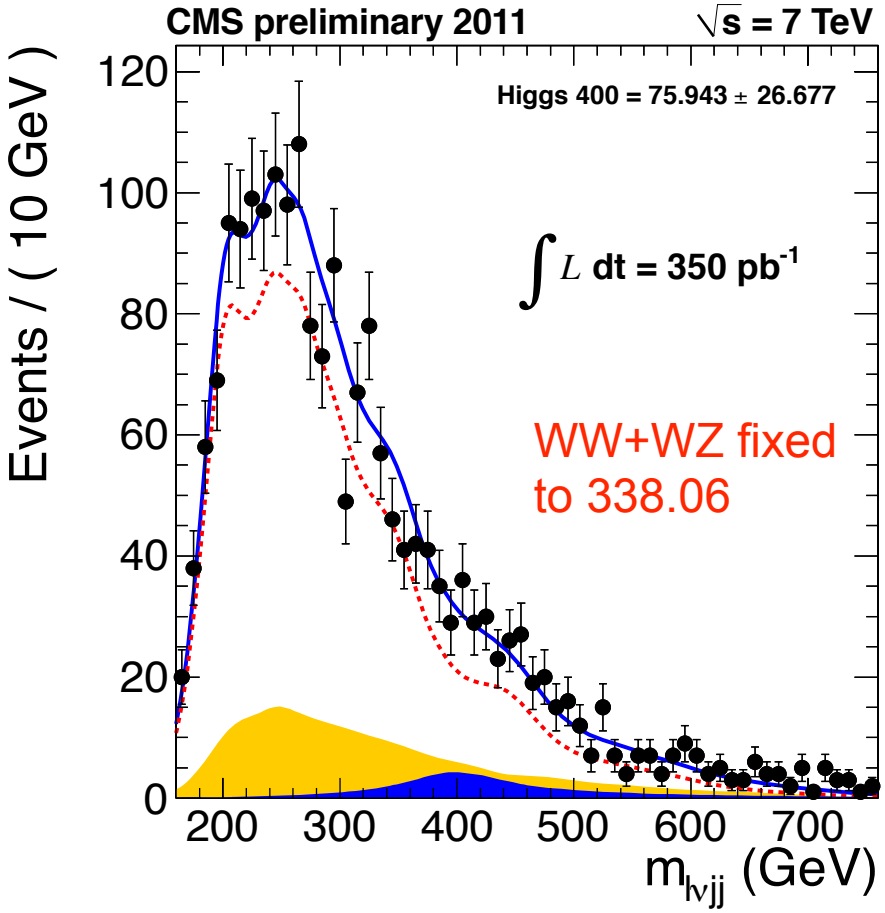
mH = 350 GeV

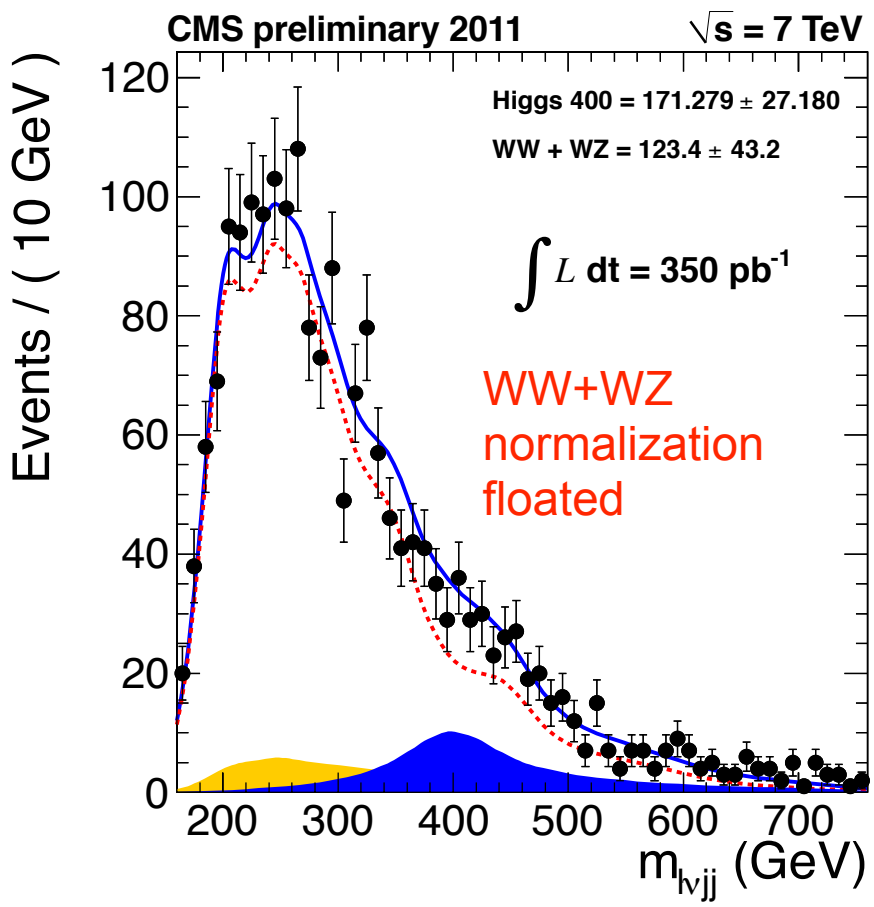




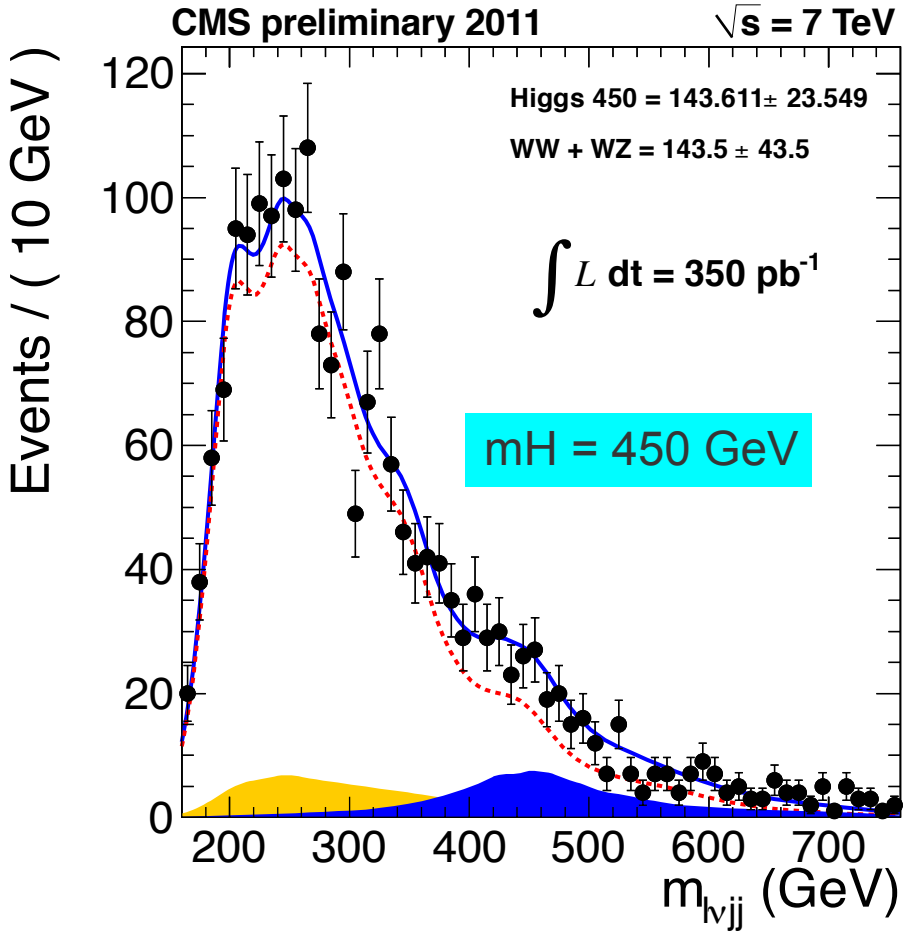
mH = 400 fit: effect of constraining WW yield

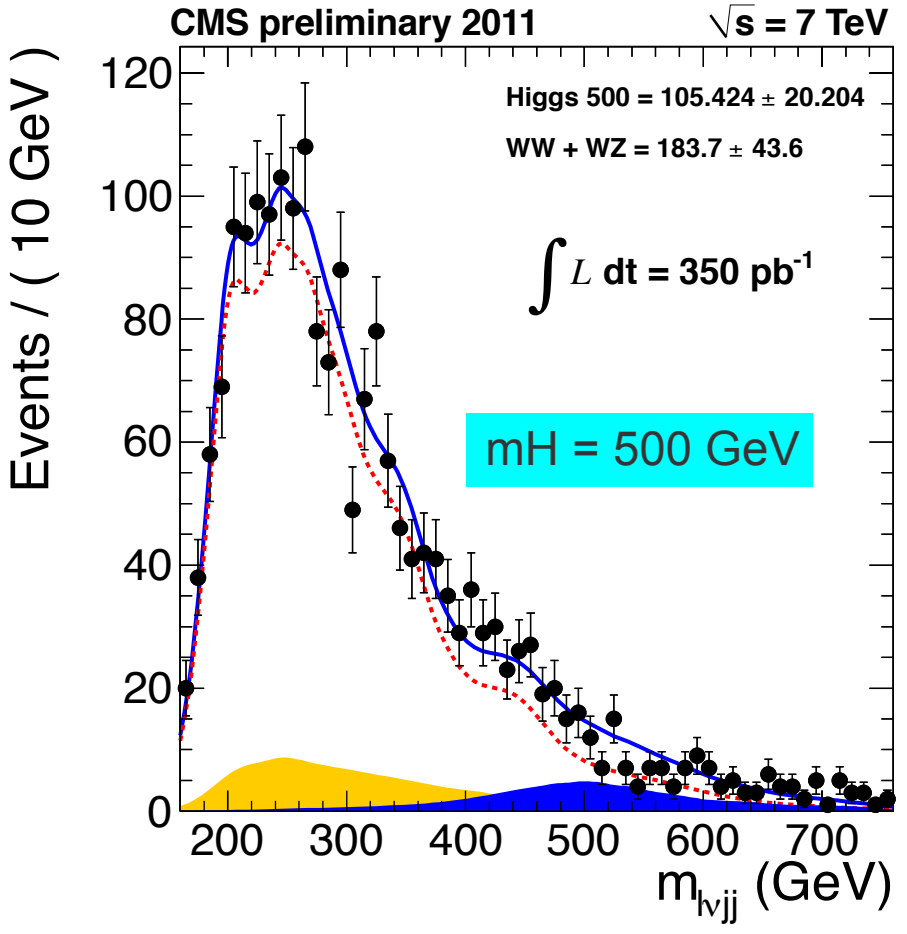
mH = 400 GeV





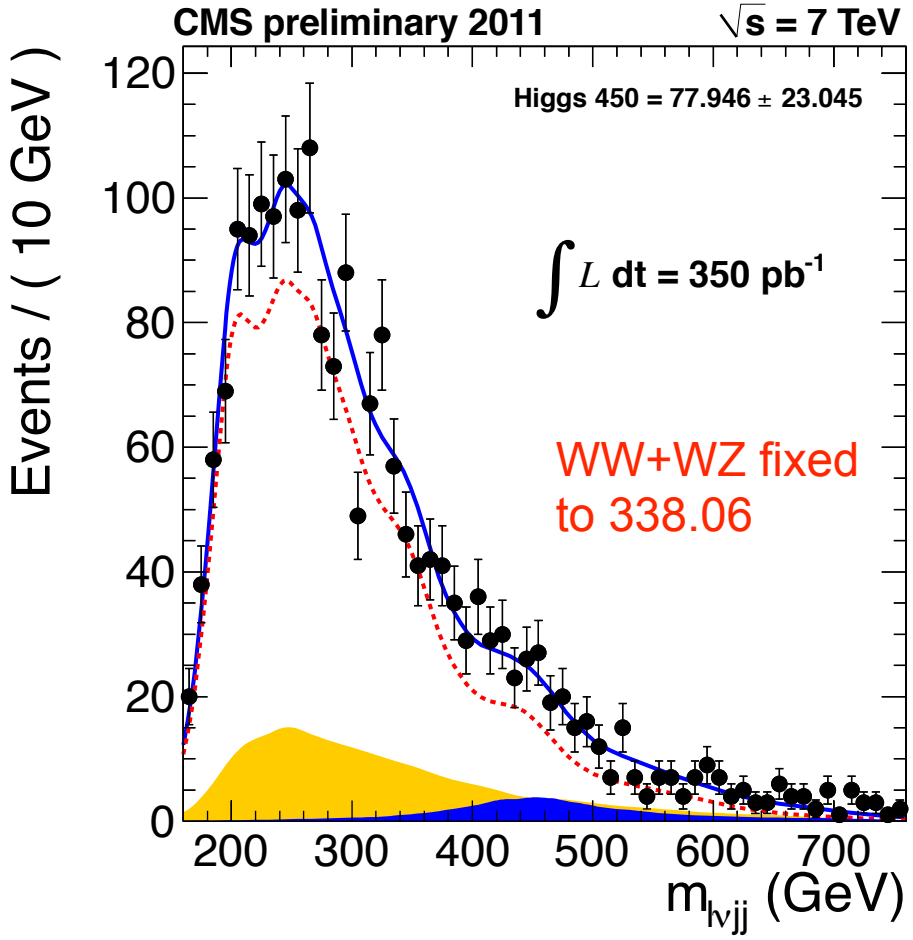
Template fit after including Higgs (3)

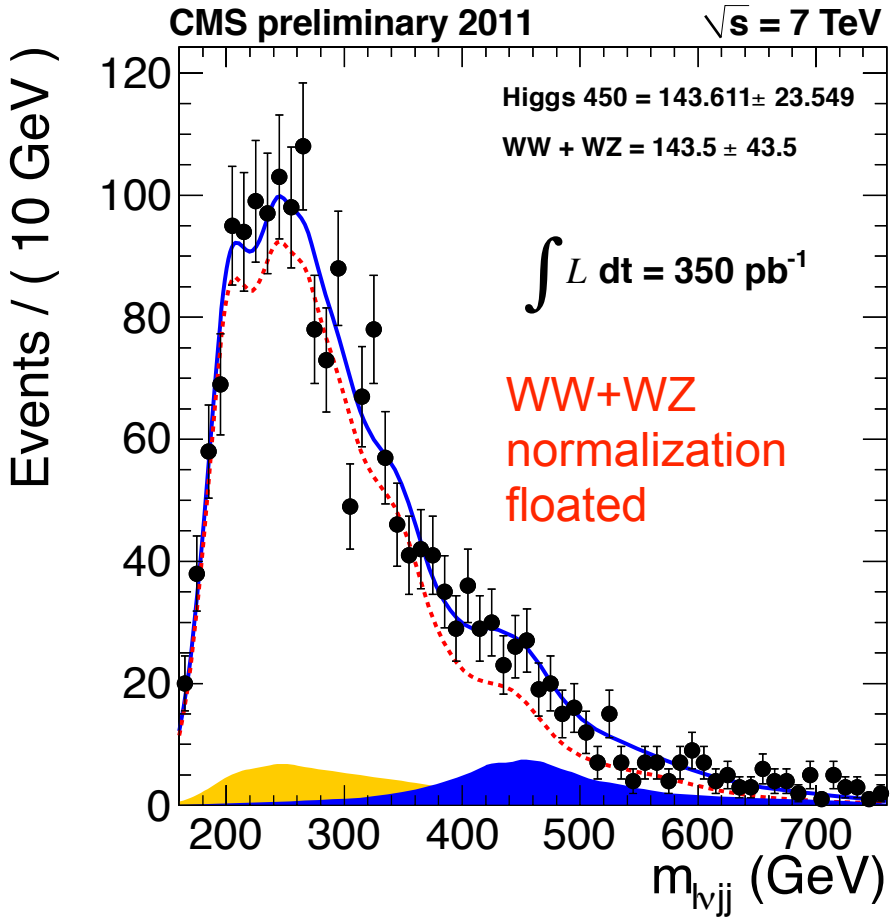




mH = 450 fit: effect of constraining WW yield

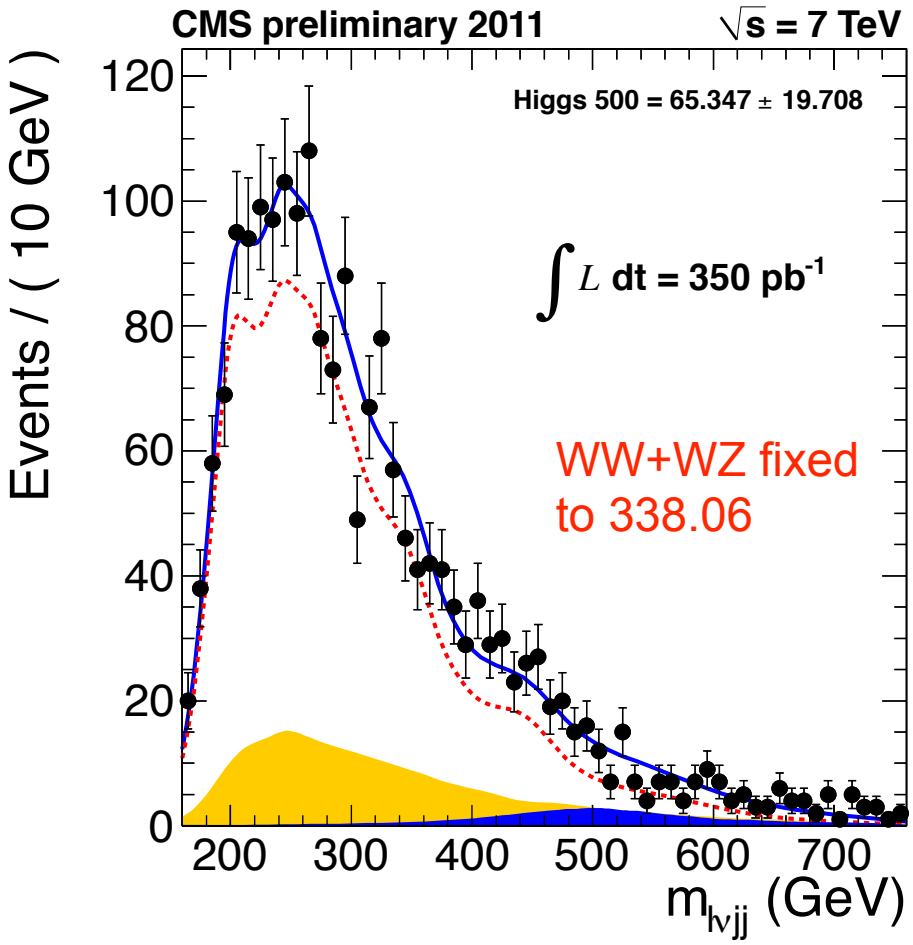
mH = 450 GeV

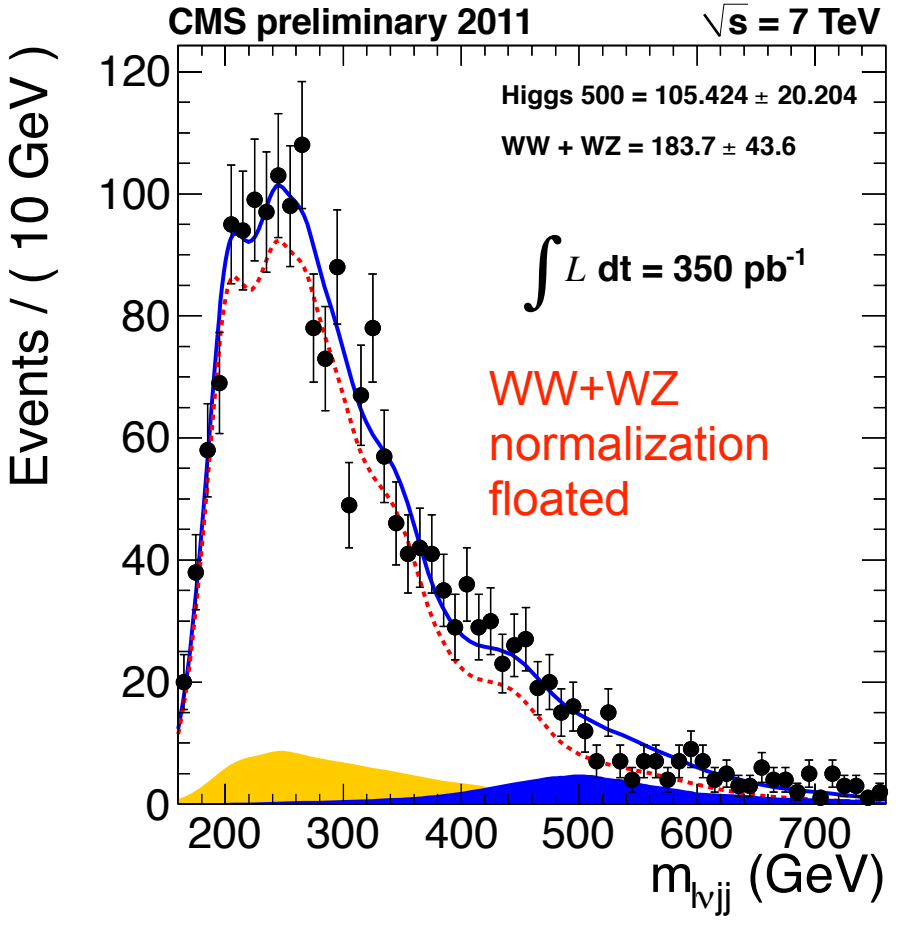




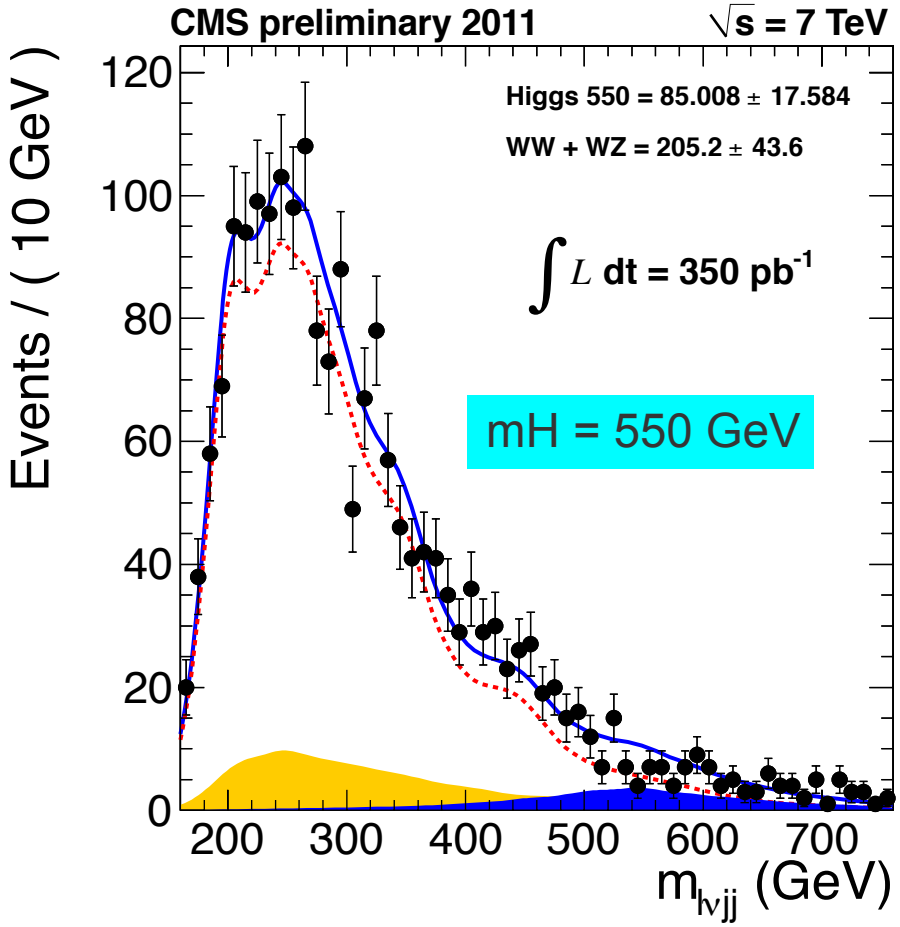
mH = 500 fit: effect of constraining WW yield

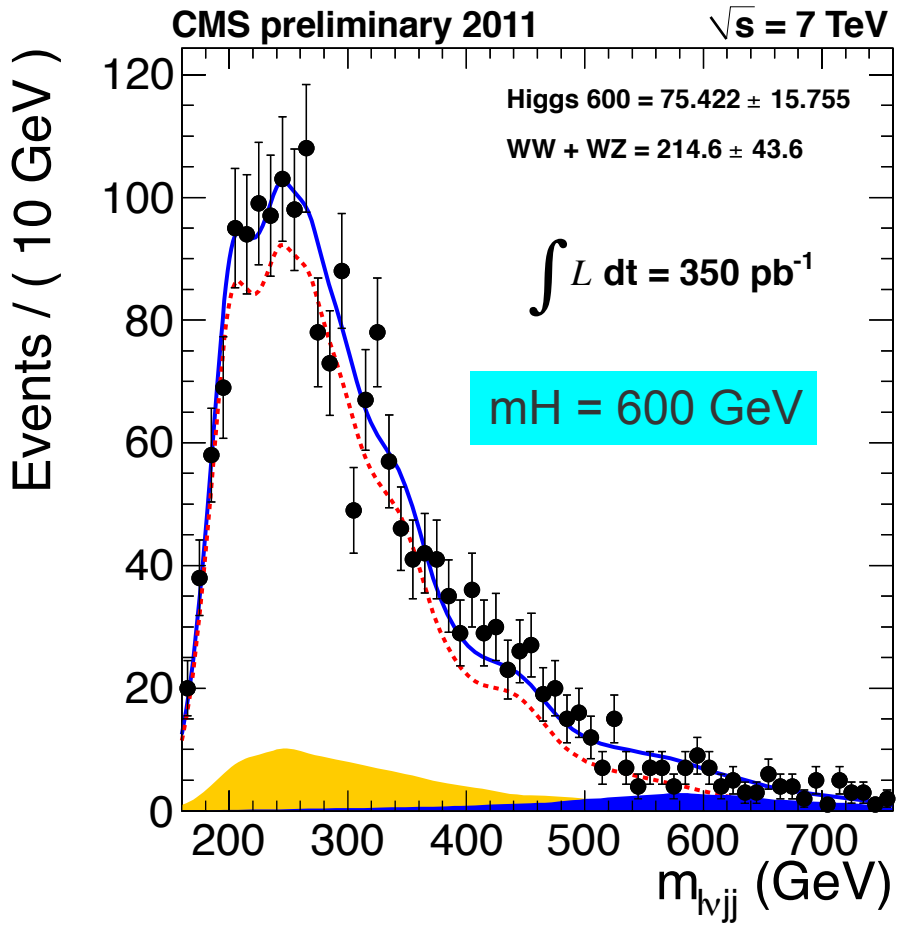
mH = 500 GeV





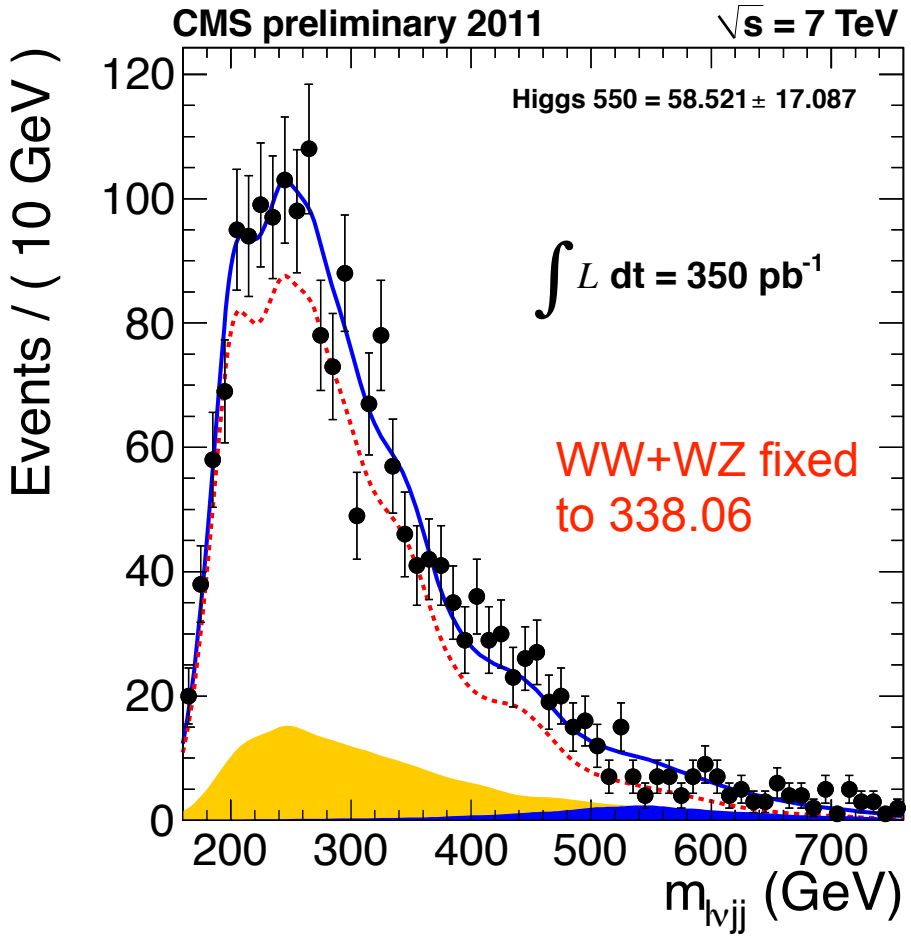
Template fit after including Higgs (4)

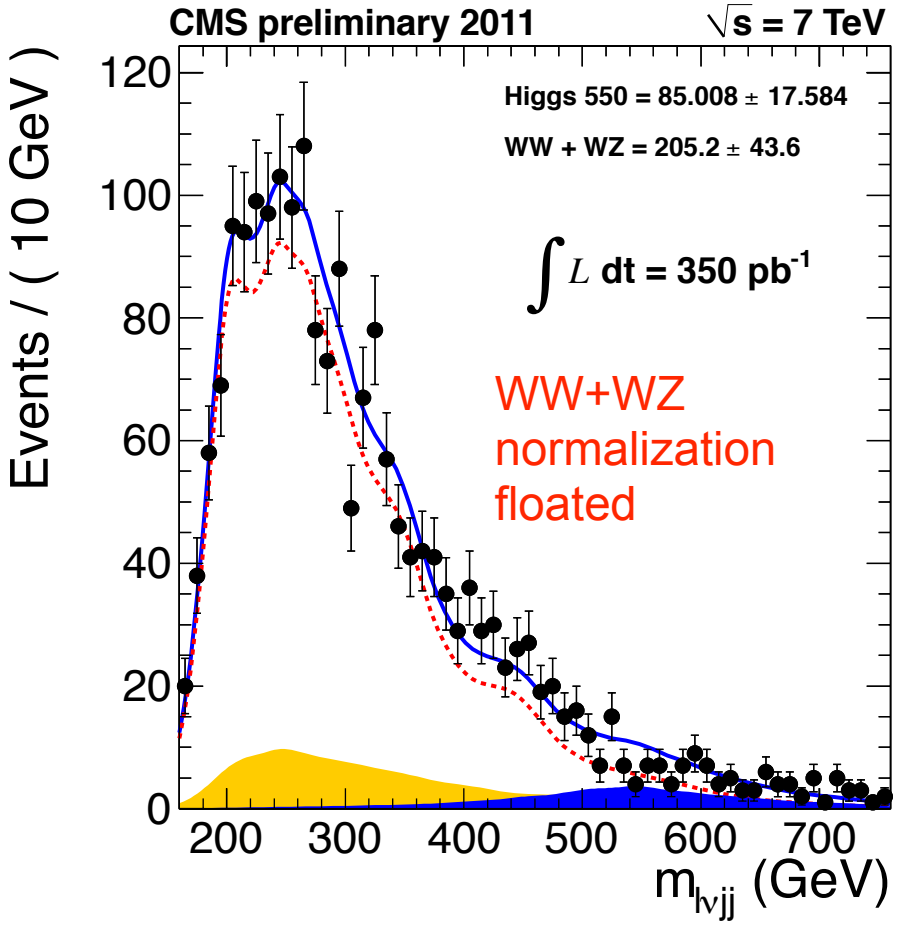




mH = 550 fit: effect of constraining WW yield

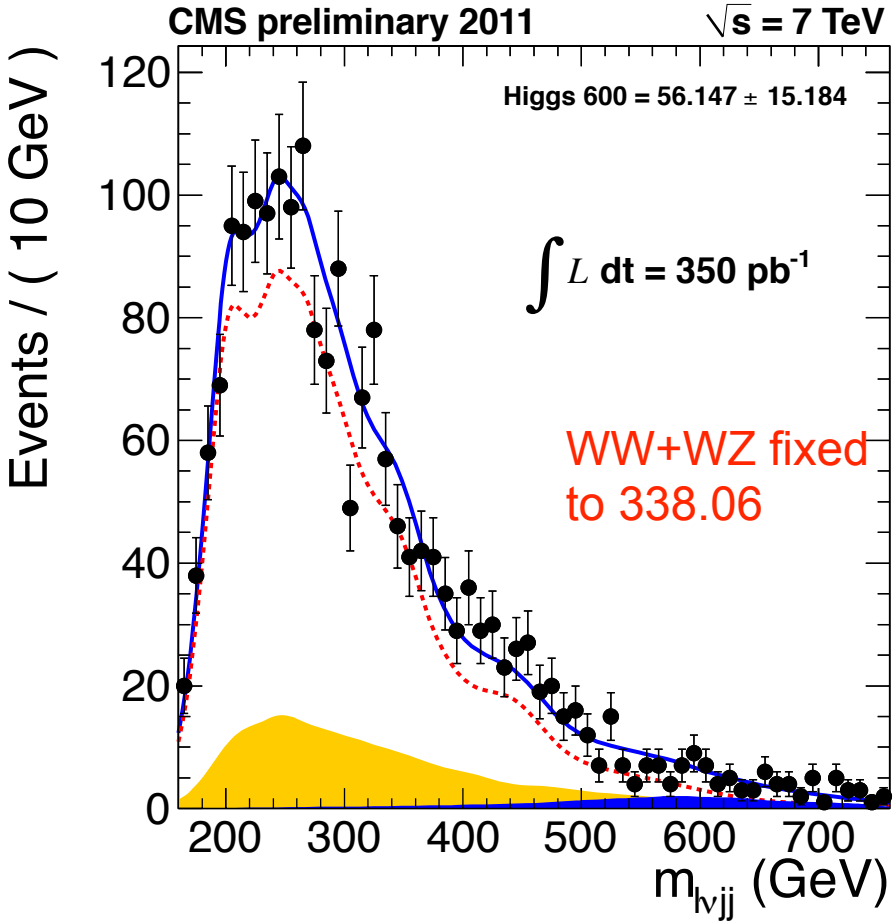
mH = 550 GeV

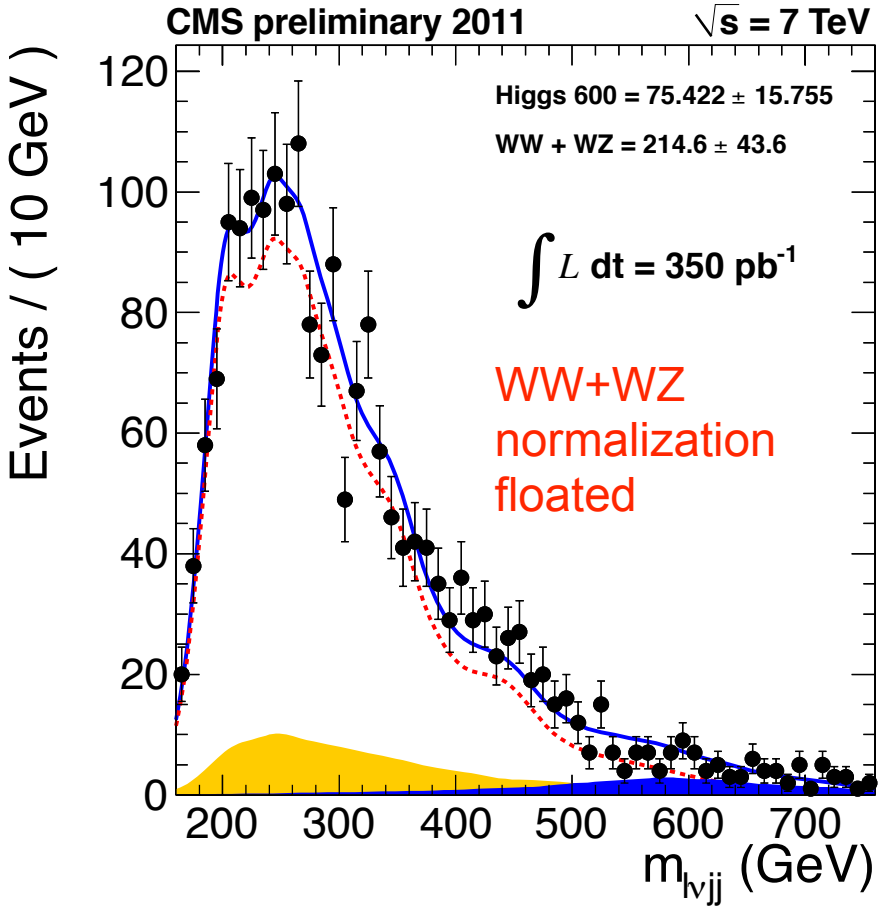




mH = 600 fit: effect of constraining WW yield

mH = 600 GeV





A big concern

Our step 1 is optimizing cuts to improve S/B on Wjj vs WW.
However, WW contains both EWK WW and H or other new physics.
Dan pointed out particular pitfall of optimizing EWK WW against W+jets.
Such optimization also cut into low mass Higgs signal.

We need to make sure that our cuts do not reduce either H(180) or H(250).
Dan fears that our cuts are removing low mass WW and therefore making the “signal” at ~ (200,250) GeV harder to extract cleanly.

This is important action item before we go to next steps

Next steps

1. Constrain the sum of di-boson and Higgs normalization in m_{lvjj} distribution to the hadronic W+Z yield obtained from m_{jj} fit

2. Do kinematic fit: this will constrain the two W masses to nominal W boson mass and will improve resolution in m_{lvjj}

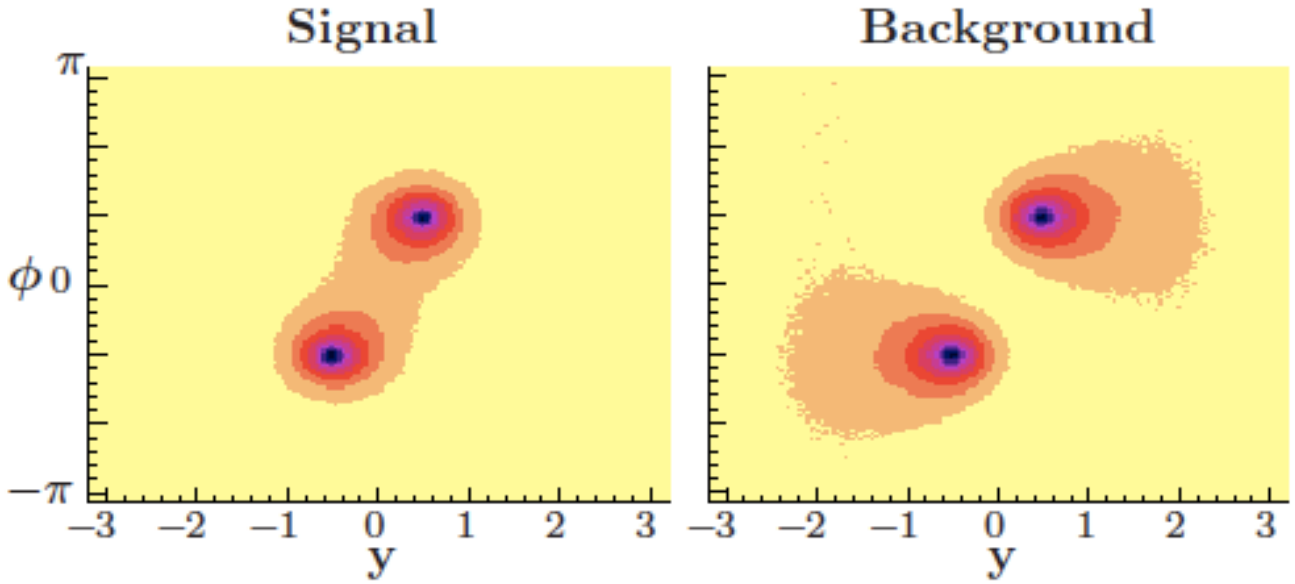
3. Convert the Higgs yield into cross section (divided by SM predicted cross section) and make the standard limit plot as a function of mH

4. Include systematics in the likelihood
   - JES/JER are easy to include
   - For uncertainty in template due to NLO effect need NLO MC
   - Similarly, need MC with Q^2 up/down variation
   - Include single top, QCD multi-jet, top etc. contributions

Some future improvements

- Use information about color correlation between between the two jets
- Use angular information after kinematic fit
- Optimize for separately for various Higgs mass: < 250, 250−400, >400

Should give us another 20−30% more discrimination.



arXiv:1001.5027

color pull:

$$\vec{t} = \sum_{i \in \text{jet}} \frac{p_T^i \, |r_i|}{p_T^{\text{jet}}} \vec{r}_i \, .$$

backup slides