Preliminary analysis using 350 pb^{-1} data

On behalf of H→WW (lvjj) working group

(June 14, 2011)

Data & selection used in this presentation

Acceptance

- Tight lepton selection from top PAG
- Exactly two jets with **p_T > 30 GeV** (using PF2PAT cleaning)
- pf MET > 25 GeV
- W transverse mass > 50 GeV

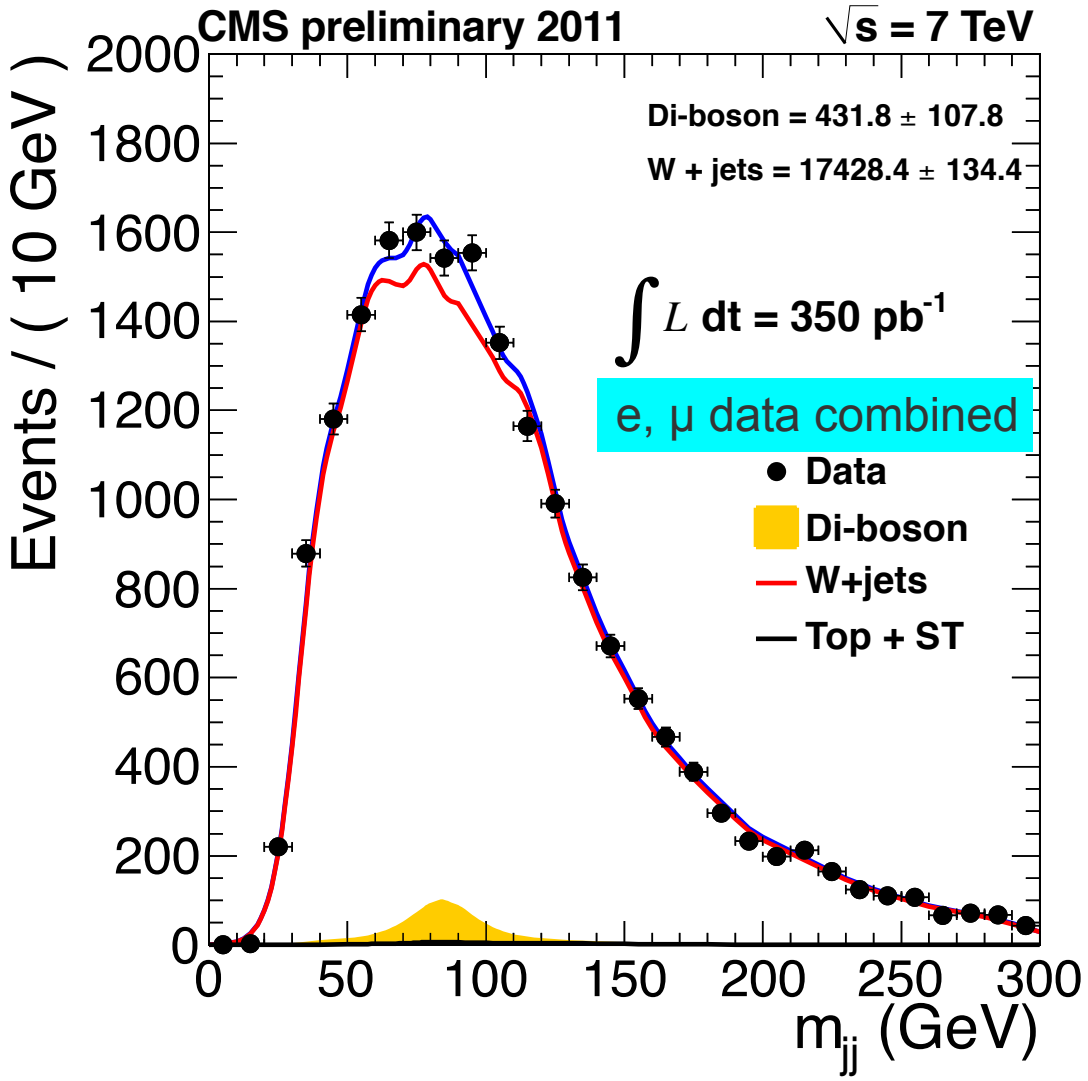
Kinematic cuts to suppress W+jets:

- p_T^{dijet} > 40 GeV
- $\Delta\eta$ (j1, j2) < 1.5

✦cuts on $\Delta\phi$ (jet, lepton) turned out to be very inefficient
✦cut on $\Delta\phi$ (W, dijet) is also inefficient, and has data/MC disagreement

Processed ~ 350 pb^{-1} of data so far (340 pb^{-1} for electron, 360 pb^{-1} for muon).
Still use 4.1.X MC.

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



MC predicts ~ 350 di-bosons.
σ = 61 pb, BR = 0.22 x 0.7
Acceptance ~ 0.45 x 0.45
Efficiency ~ 0.7 x 0.7, Lumi = 350

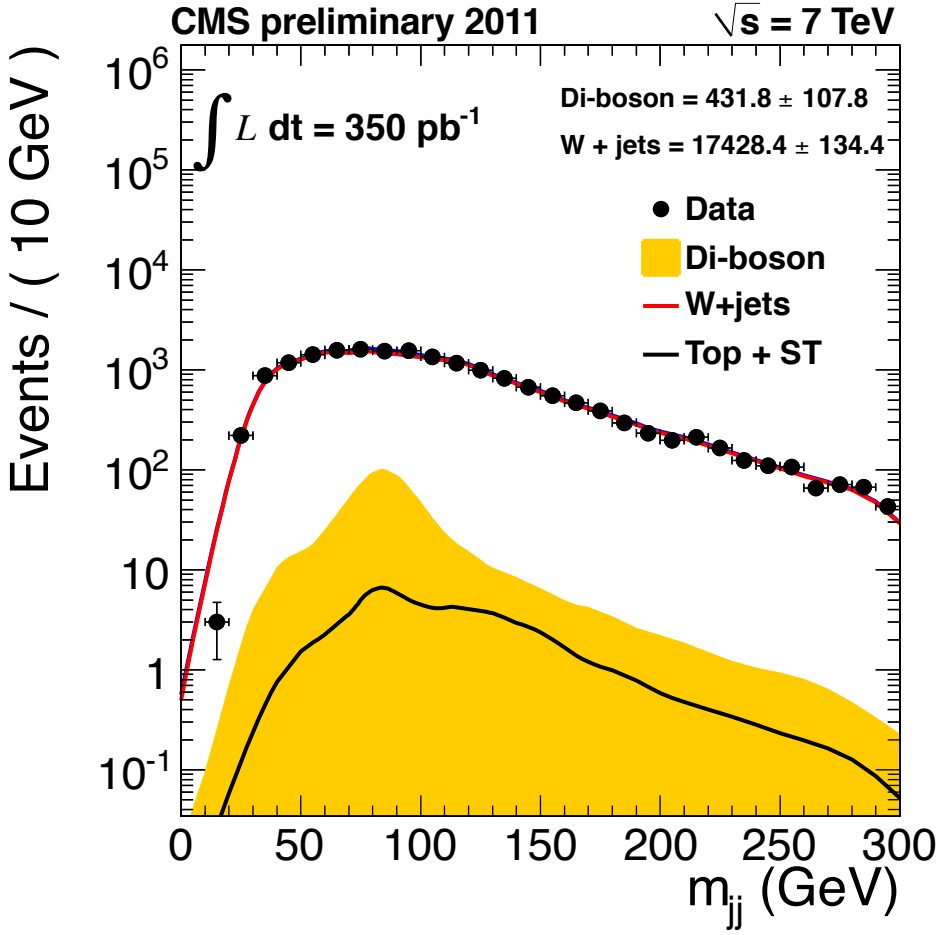
Take the shape from MC. Fit for the normalization. Blue curve shows the fit to data.

This fit is still a work in progress. Will get ~ 3 fb^{-1} W +jets MC by next week. This will help improve the template.

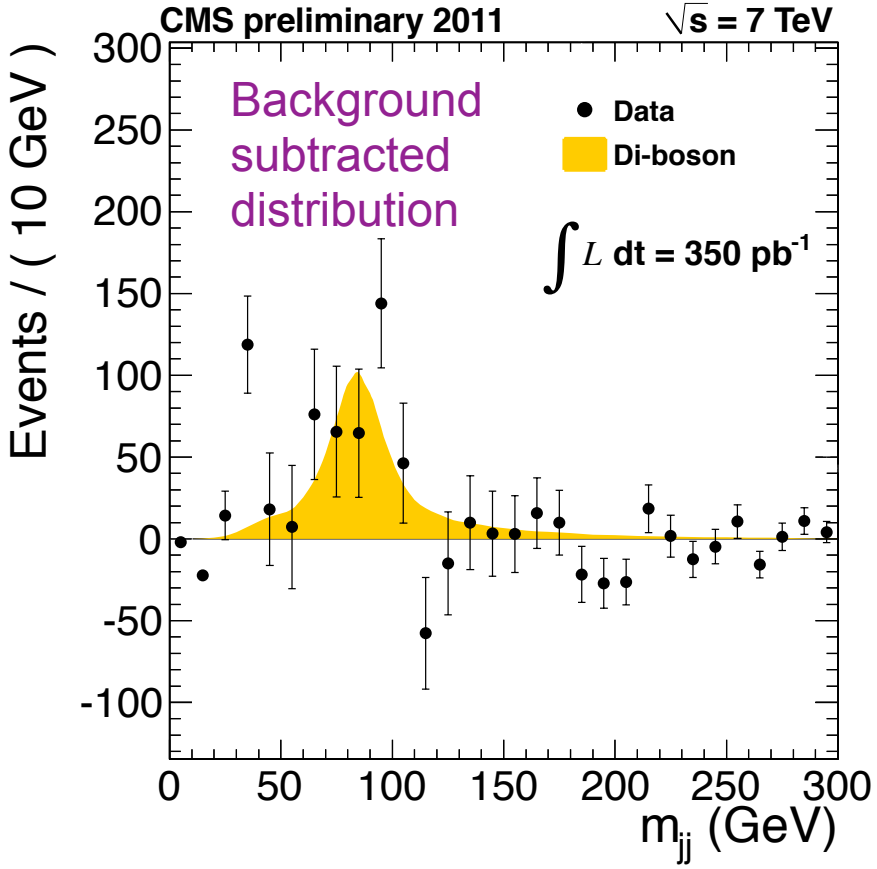
Background subtracted distribution on the next slide

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

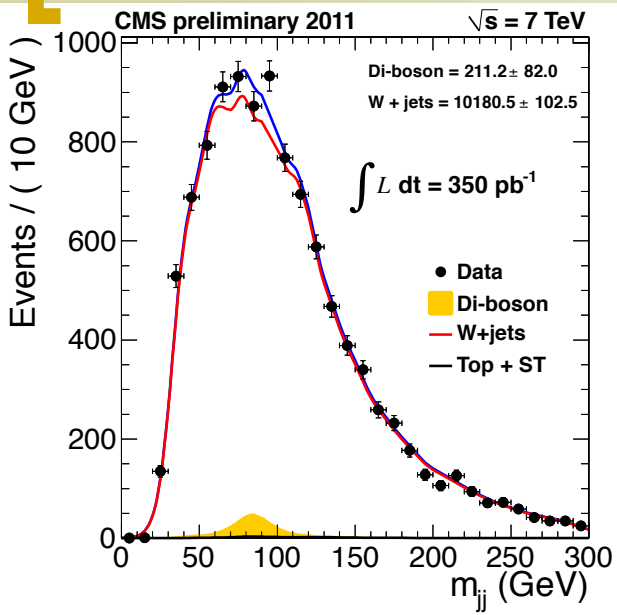
on the log scale

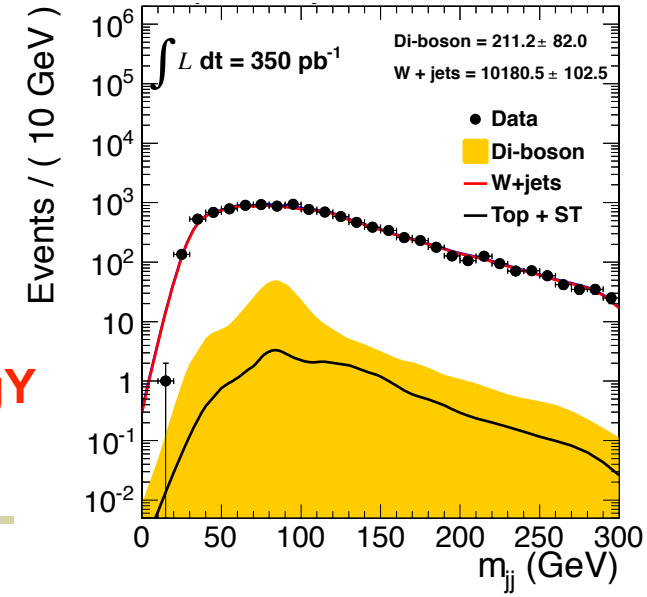


In the W mass window $65 < m_{jj} < 95$ GeV we get:
233 di-boson, **4428** W+jets events

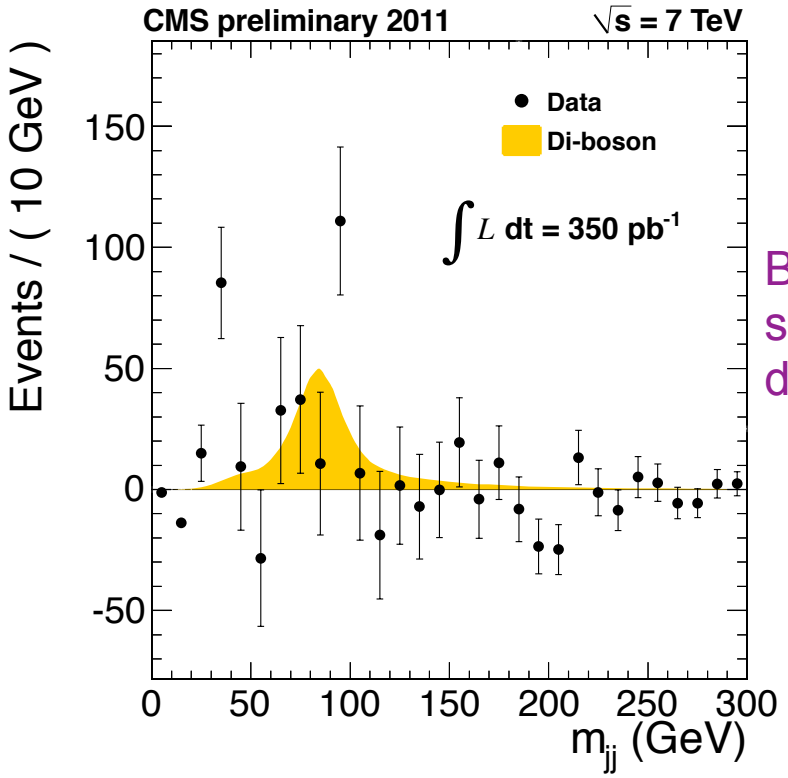


Template fit to m_{jj} in W+2 jet events: μ data





logY

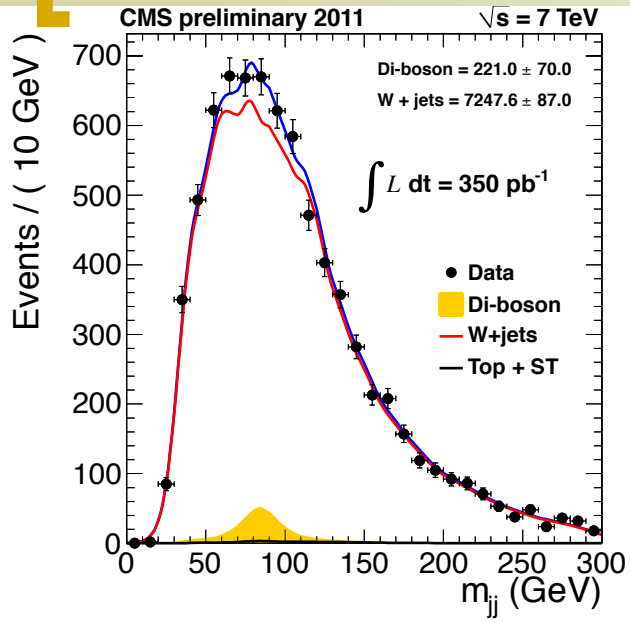


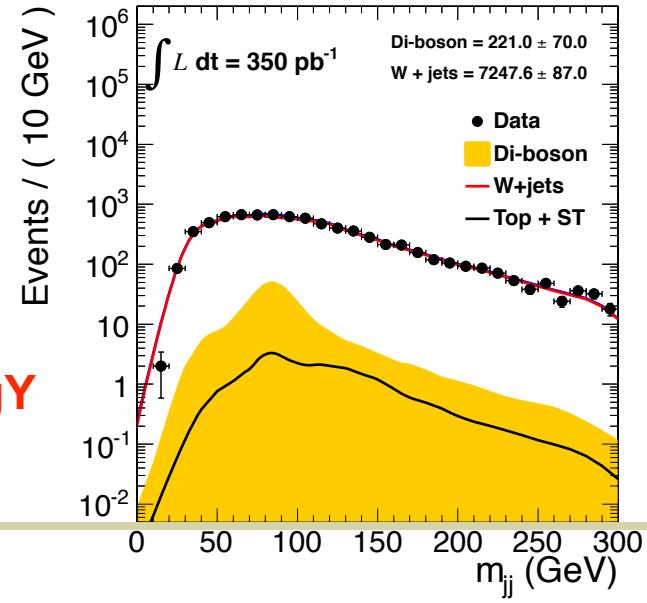
Background subtracted distribution

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

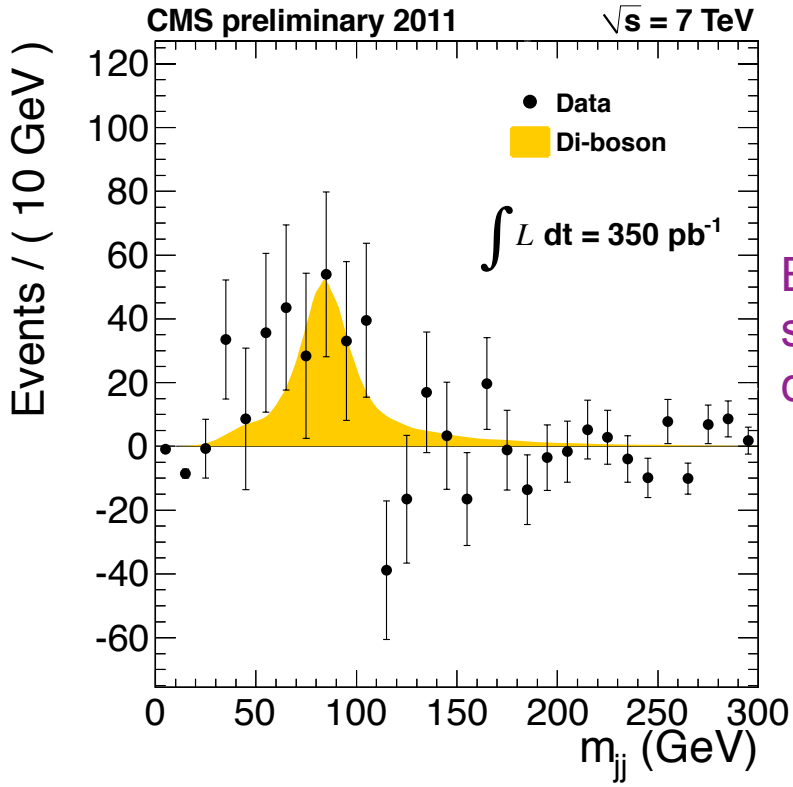
114 di-boson, **2586** W+jets events

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





logY

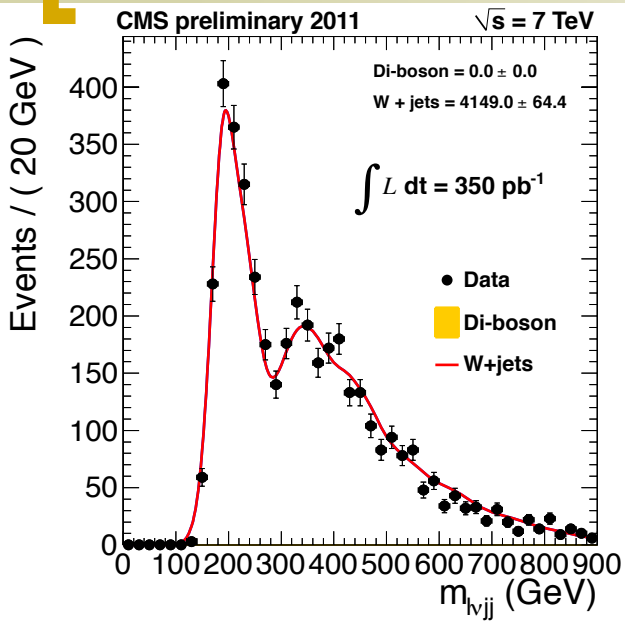


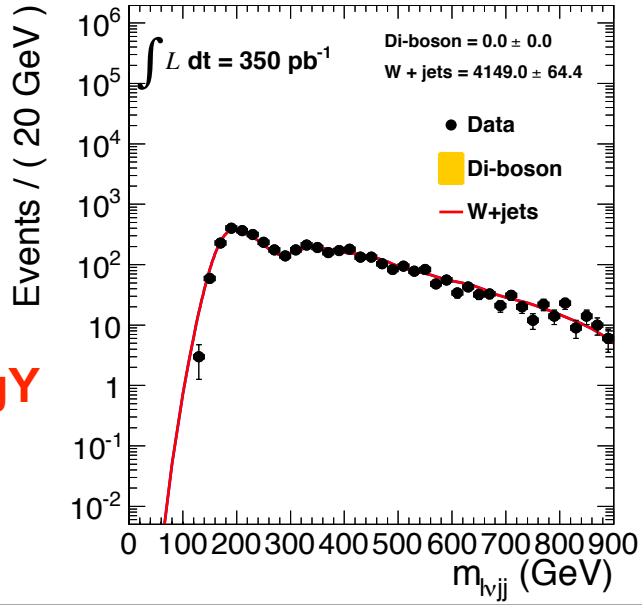
Background subtracted distribution

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

119 di-boson, **1841** W+jets events

$m_{l\nu jj}$ fit: start with sidebands

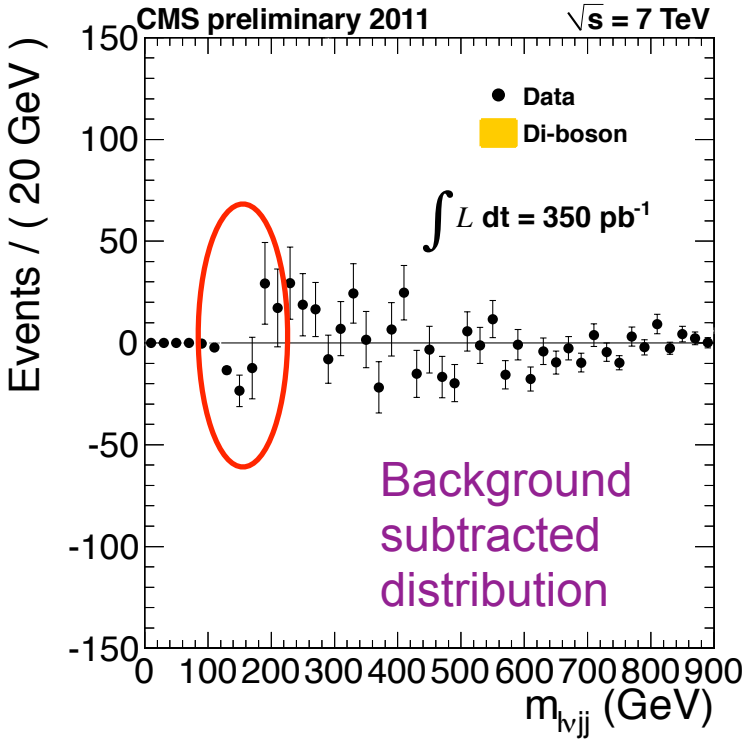




logY

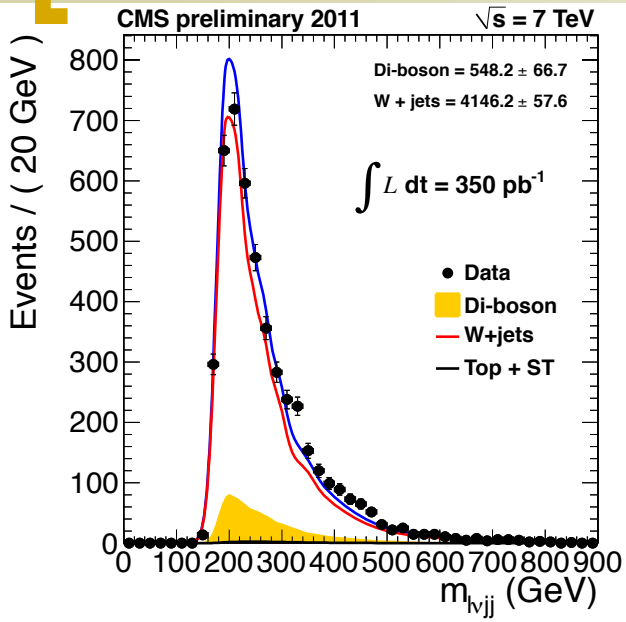
m_{jj} < 60 GeV OR m_{jj} > 200 GeV

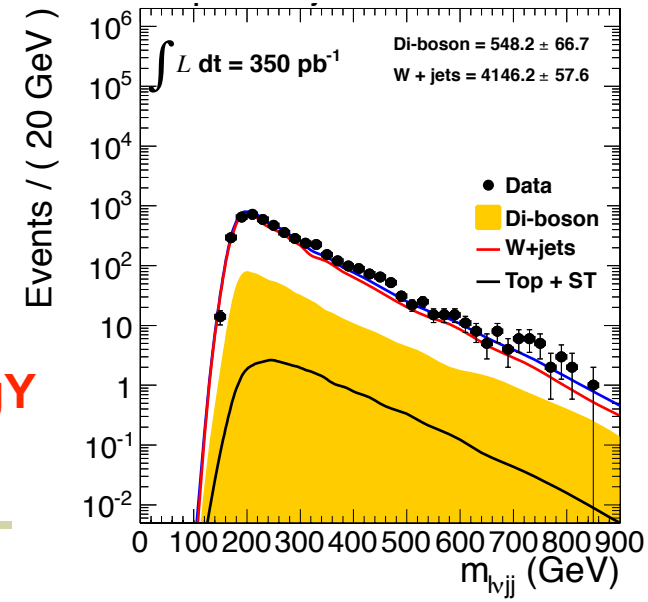
should be able to fit the W+ jets shape



We are doing ok except near the rough edge.
We can renormalize the residual if needed.

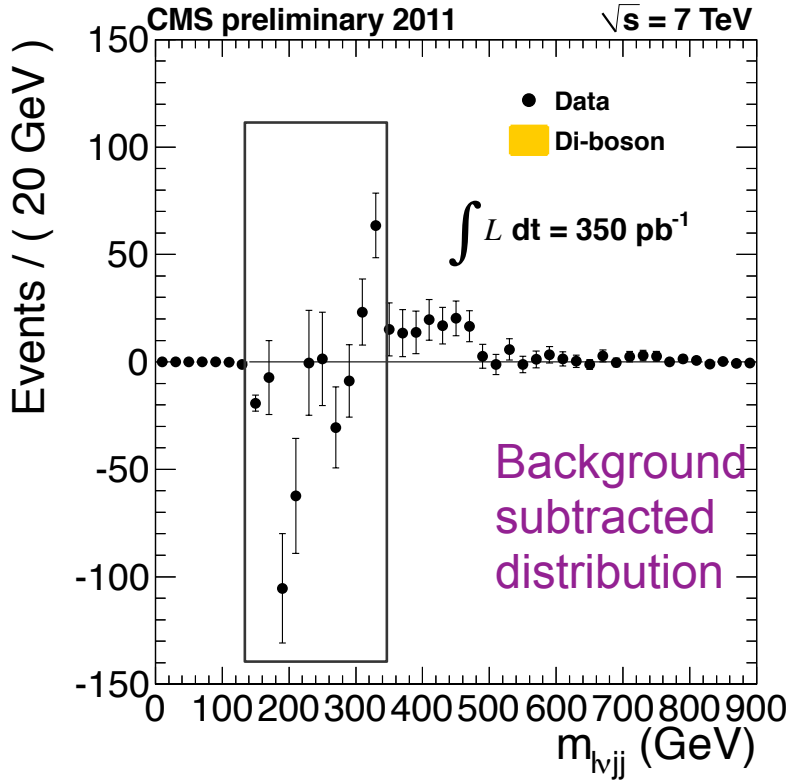
$m_{l\nu jj}$ distribution: $65 < m_{jj} < 95$ GeV





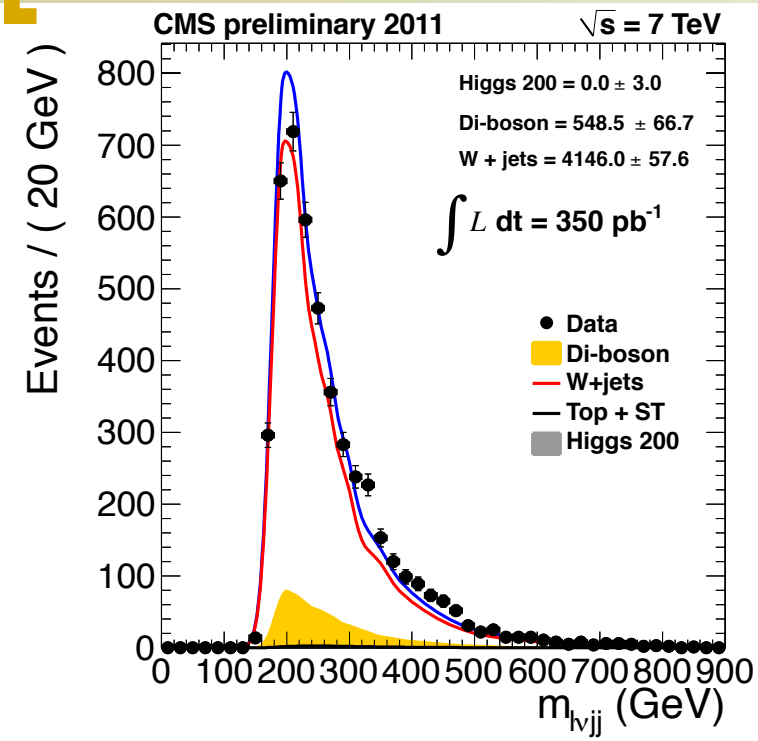
logY

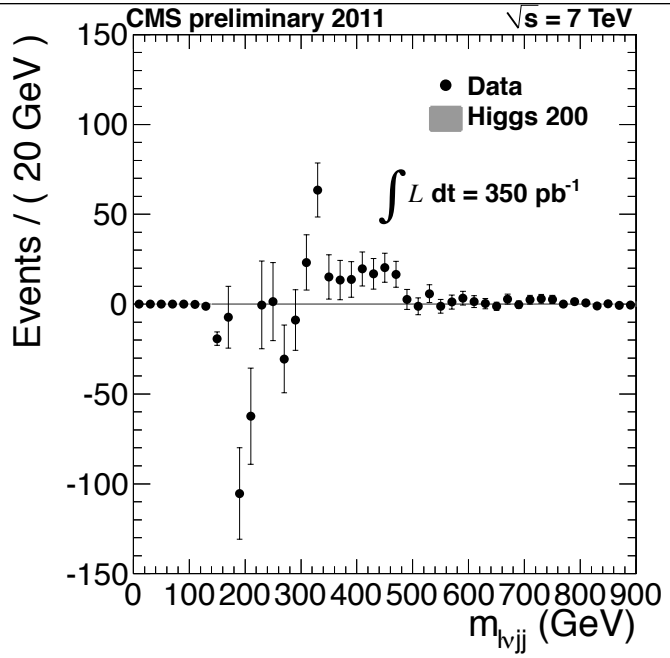
Take the shape directly from MC. Additionally, constrain the W+jets normalization from m_{jj} fit.

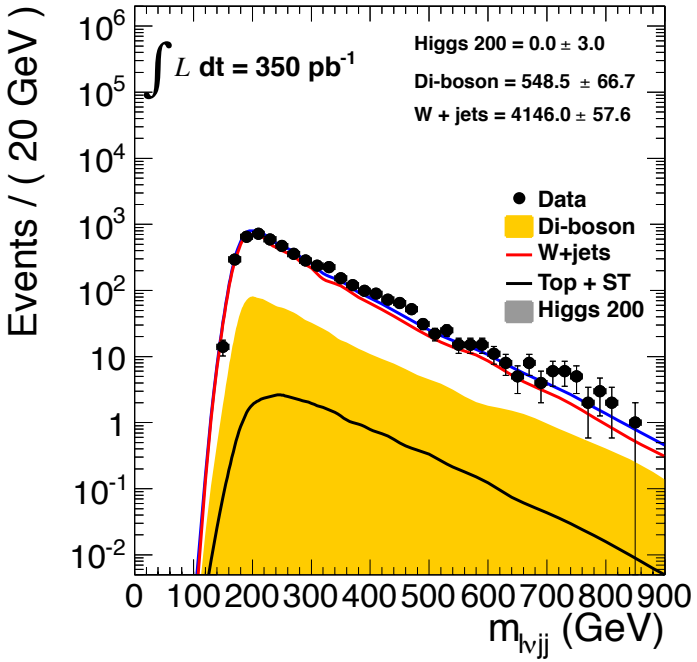


See the same dip from rough edge near 200 GeV. You will see this dip in all the subsequent plots. From next slide start including Higgs in the fit.

$m_{l\nu jj}$ fit: m_H = 200 GeV



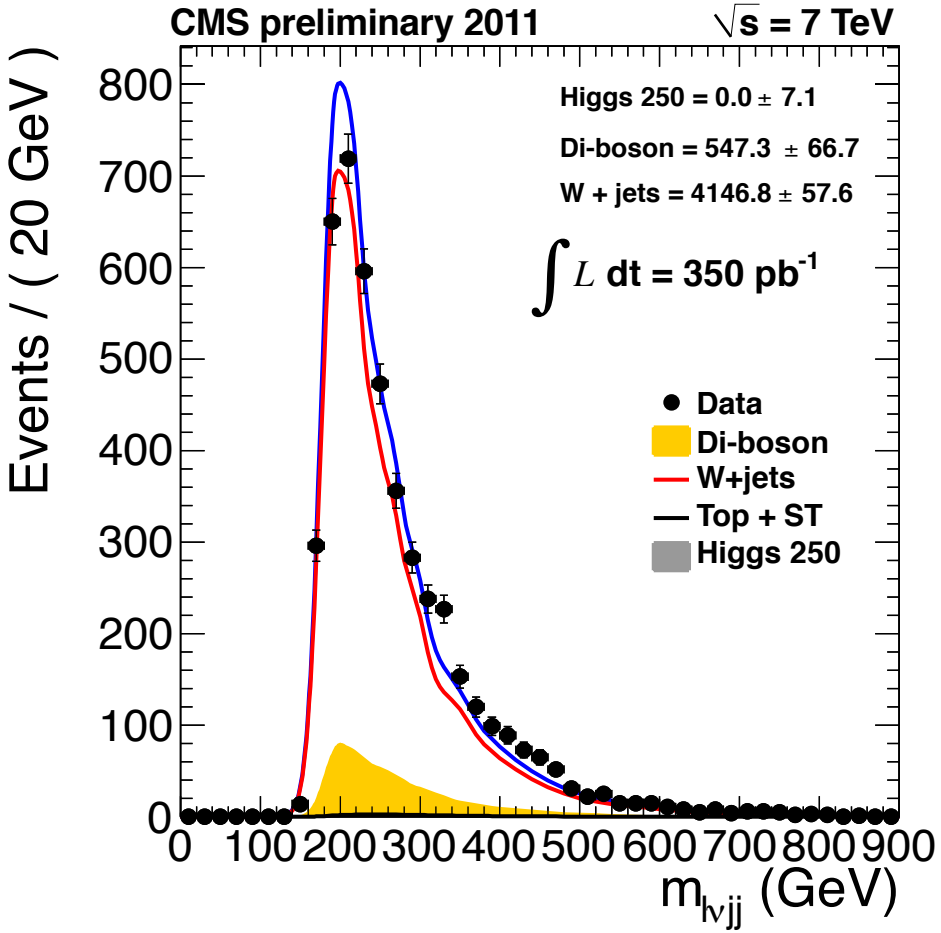


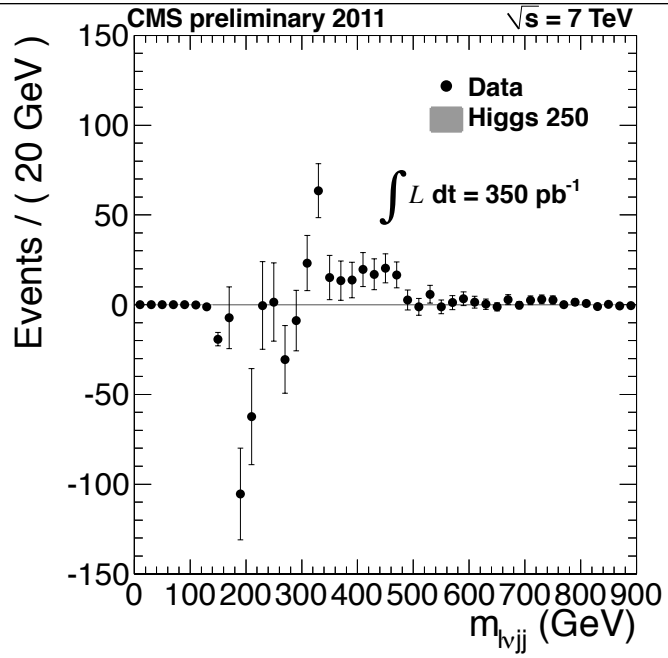


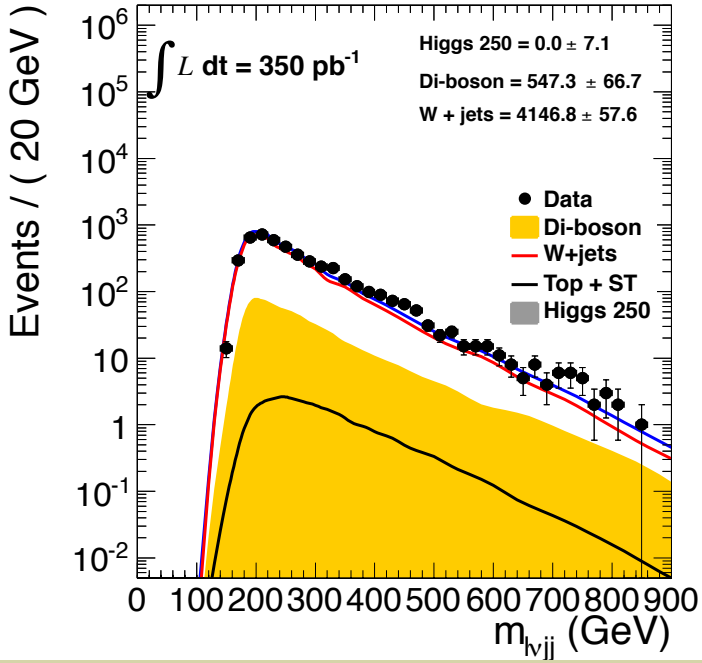
Methodology

1.) Take shapes directly from MC
2.) Constrain W+jets normalization to that obtained from m_{jj} fit to data
3.) Constrain di-boson + Higgs normalization to that obtained from m_{jj} fit

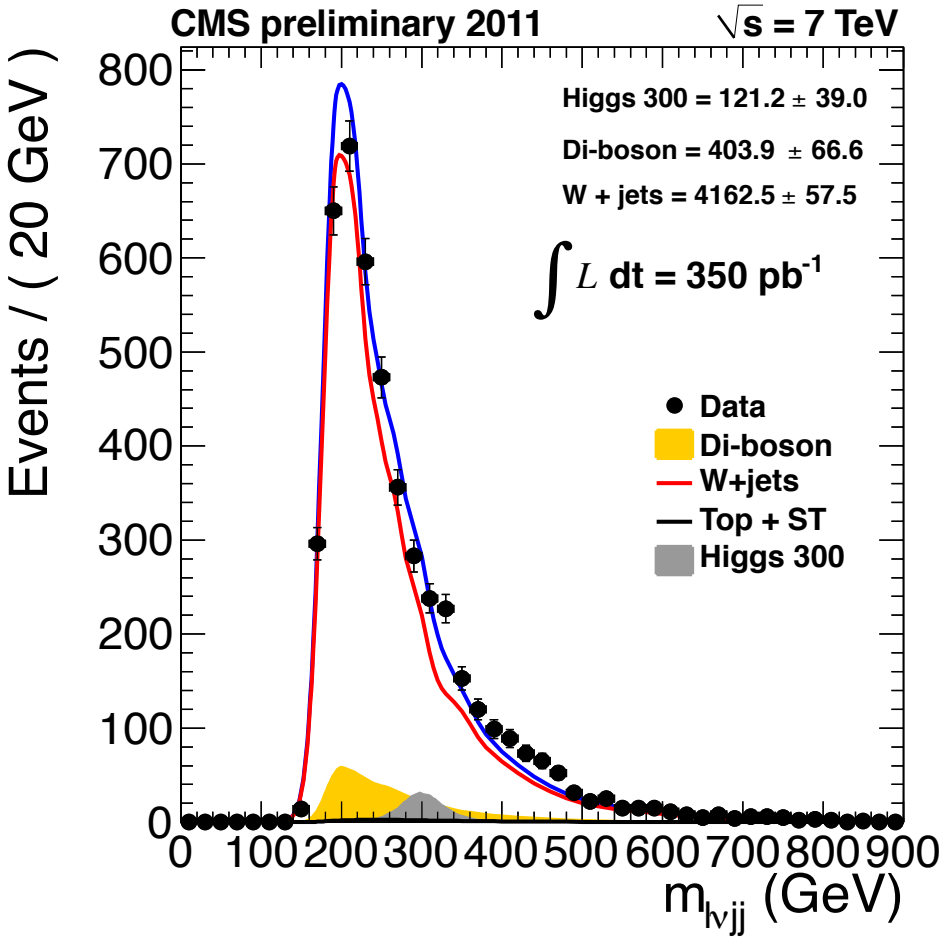
$m_{l\nu jj}$ fit: m_H = 250 GeV

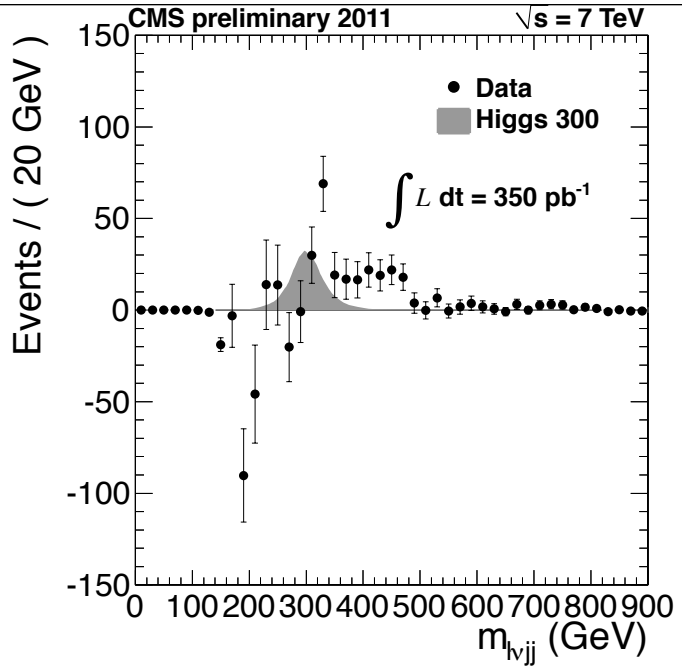


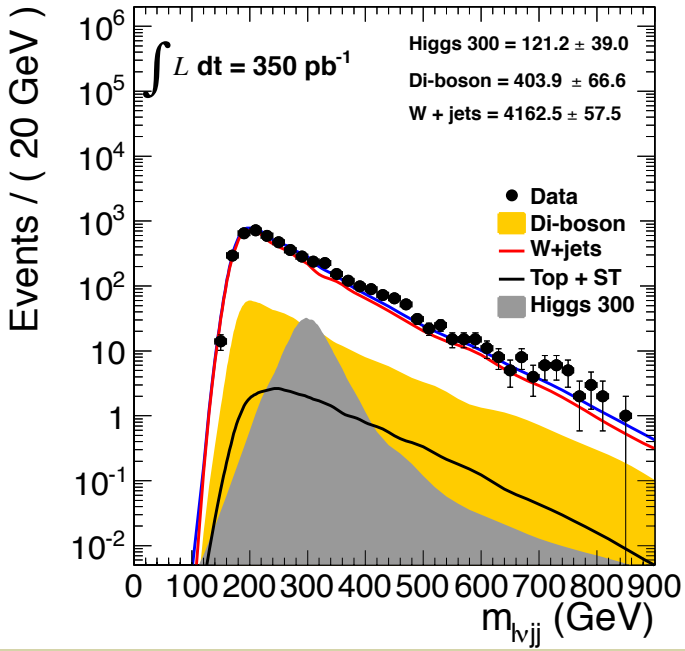




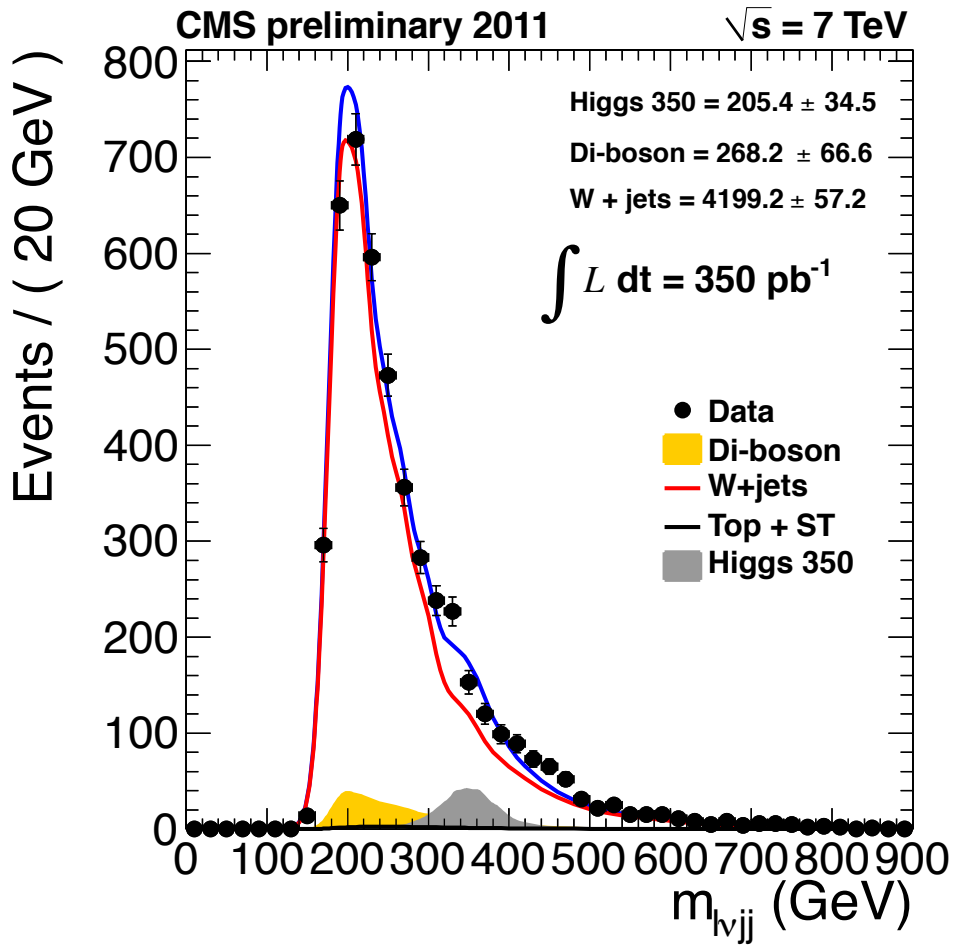
$m_{l\nu jj}$ fit: m_H = 300 GeV

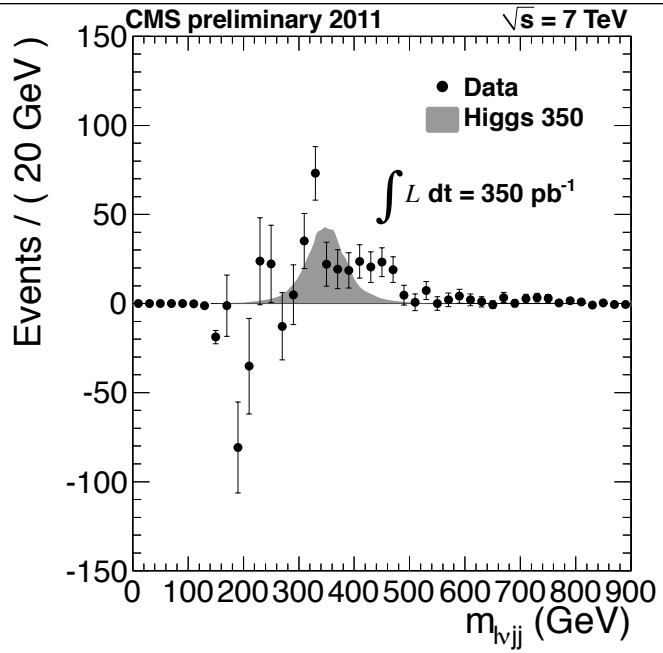


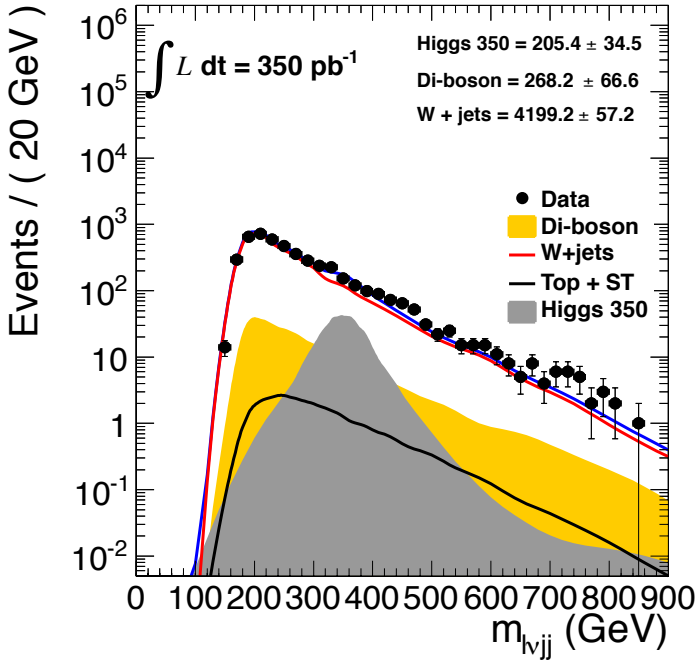




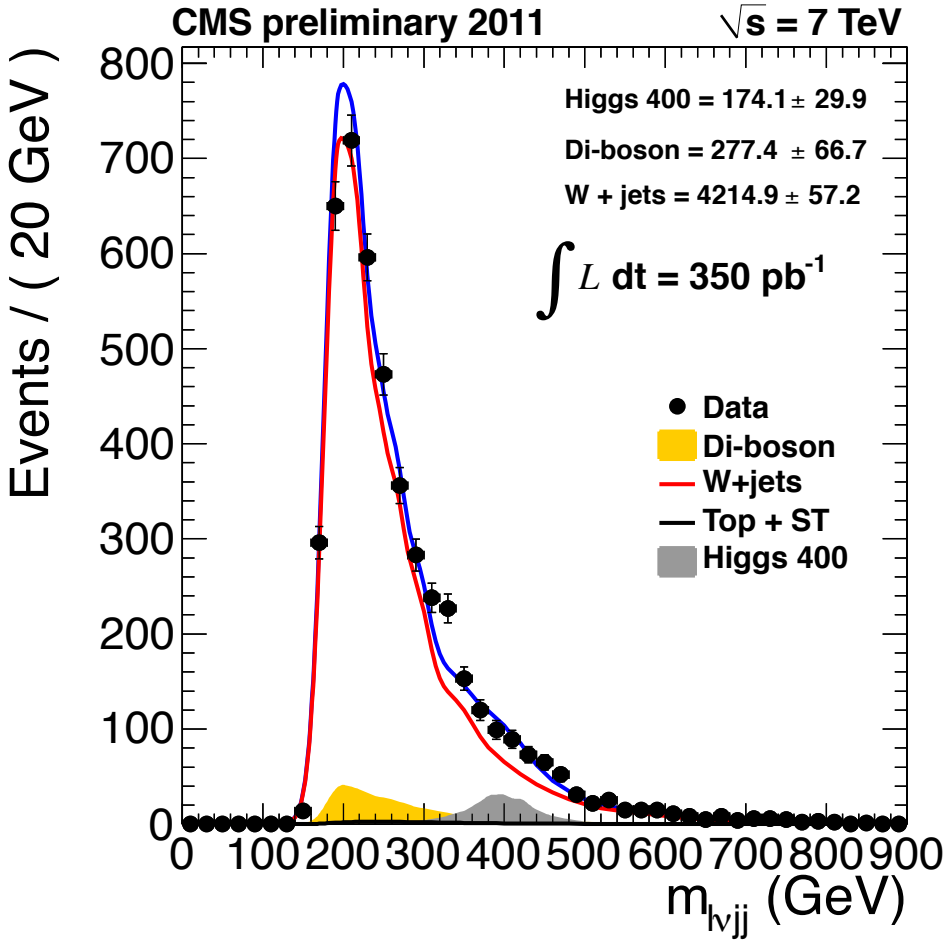
$m_{l\nu jj}$ fit: m_H = 350 GeV

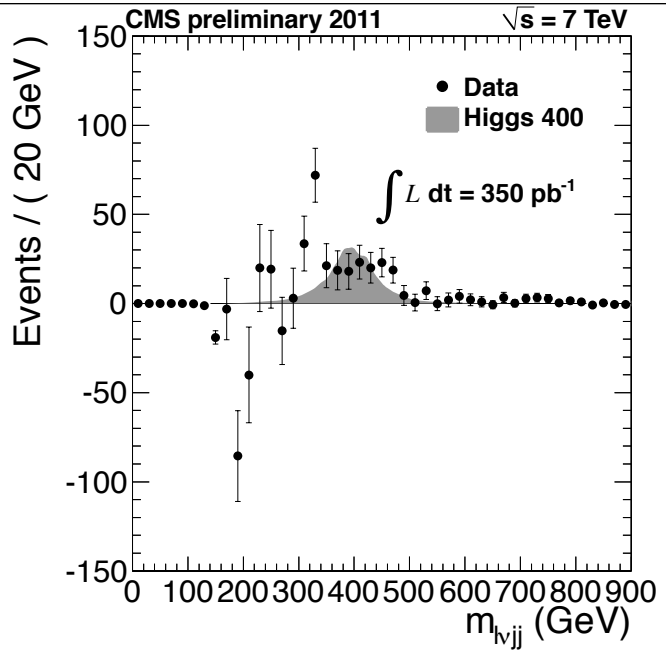


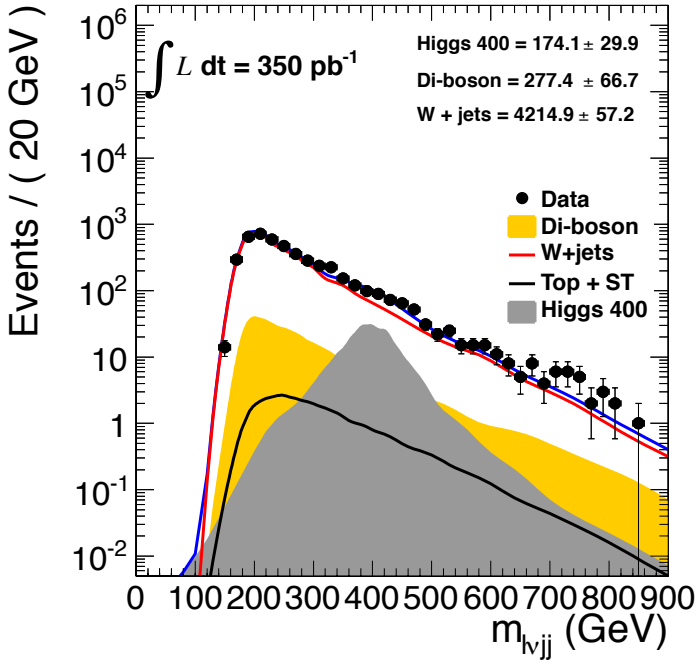




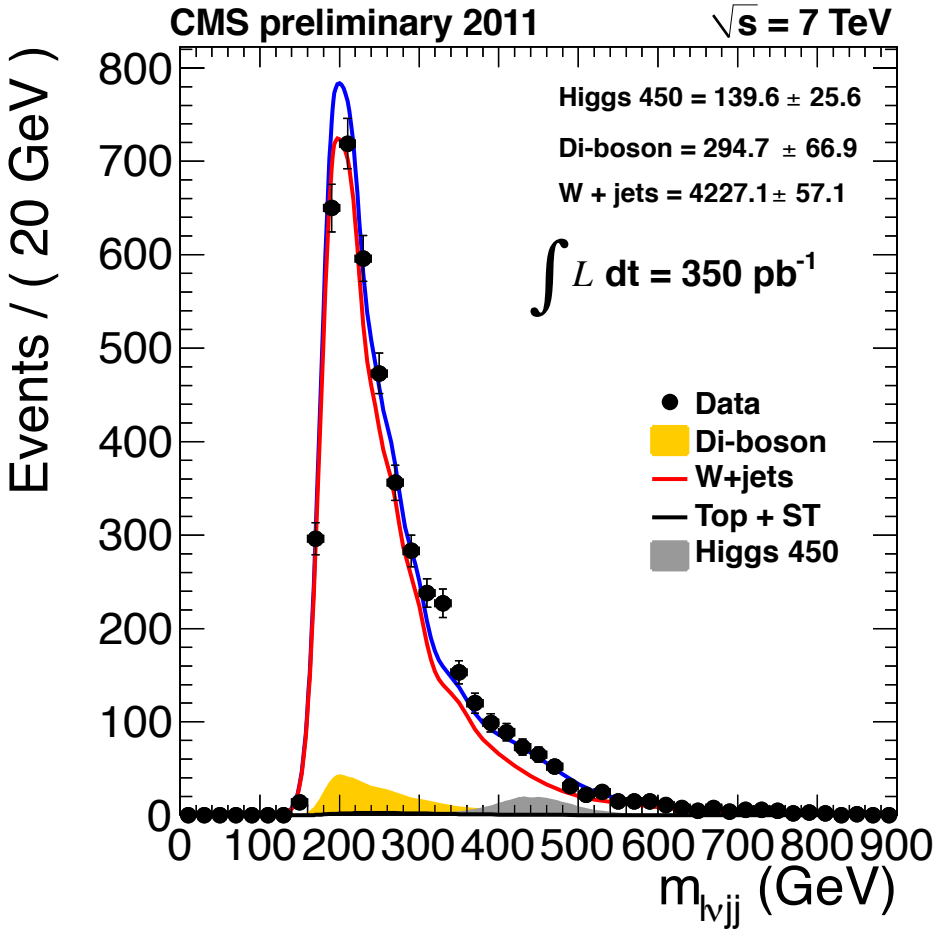
$m_{l\nu jj}$ fit: m_H = 400 GeV

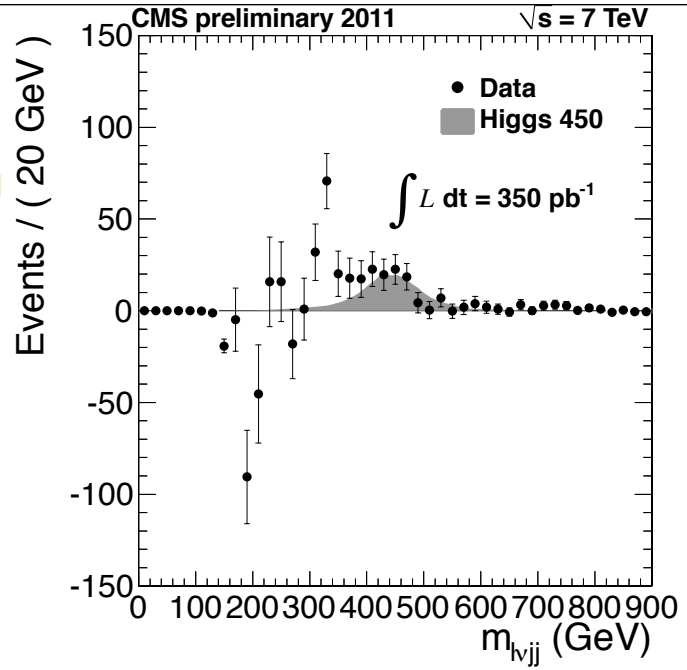


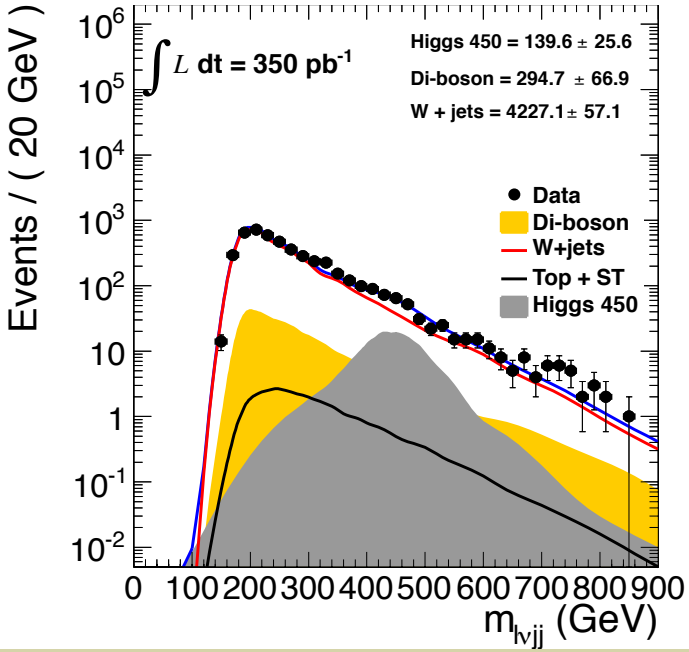




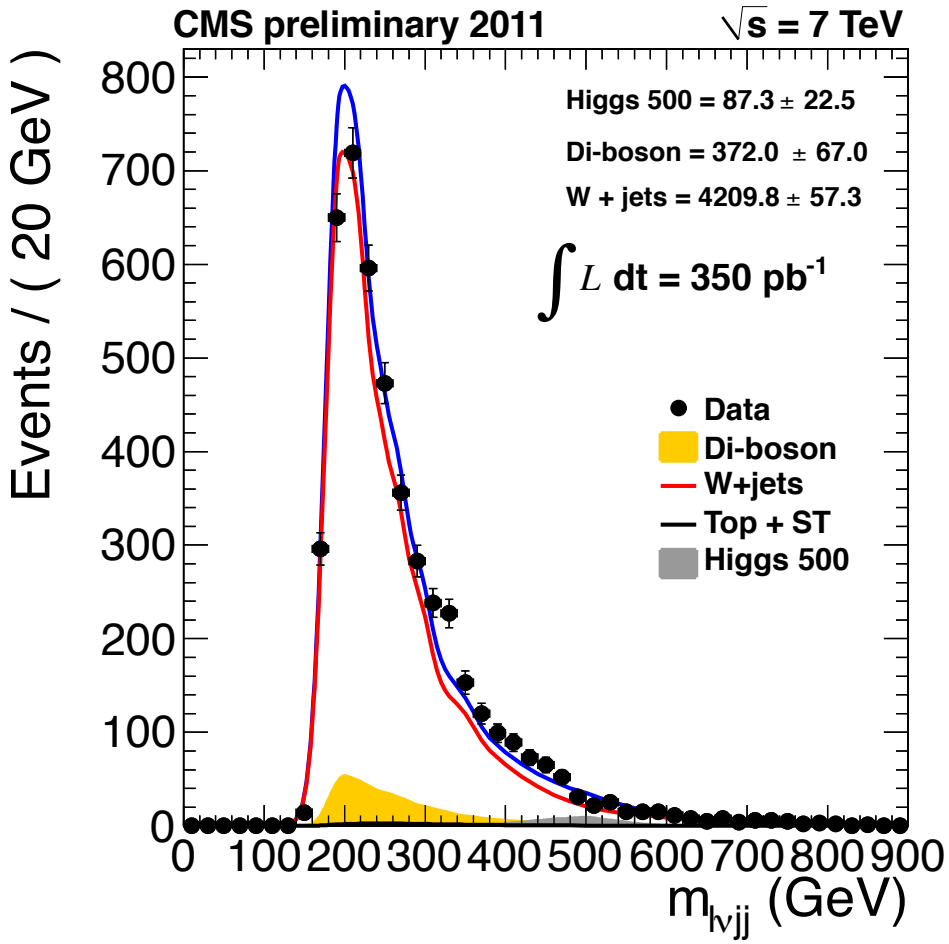
$m_{l\nu jj}$ fit: m_H = 450 GeV

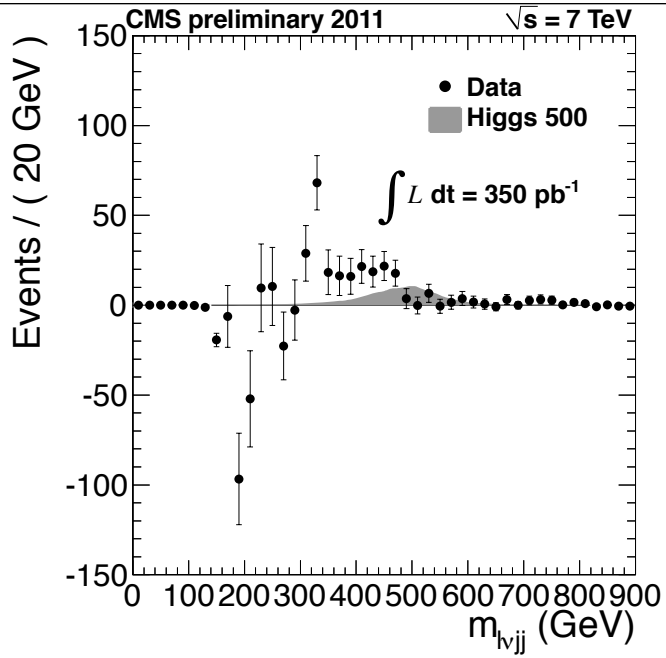


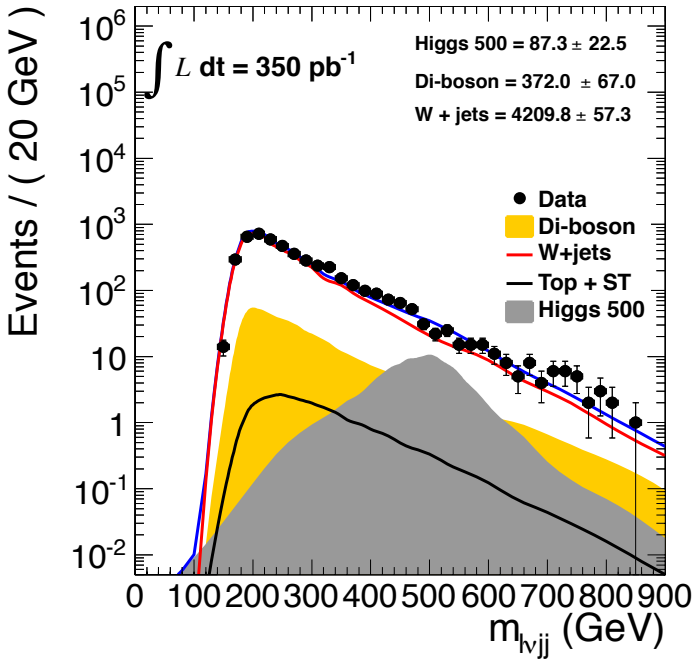




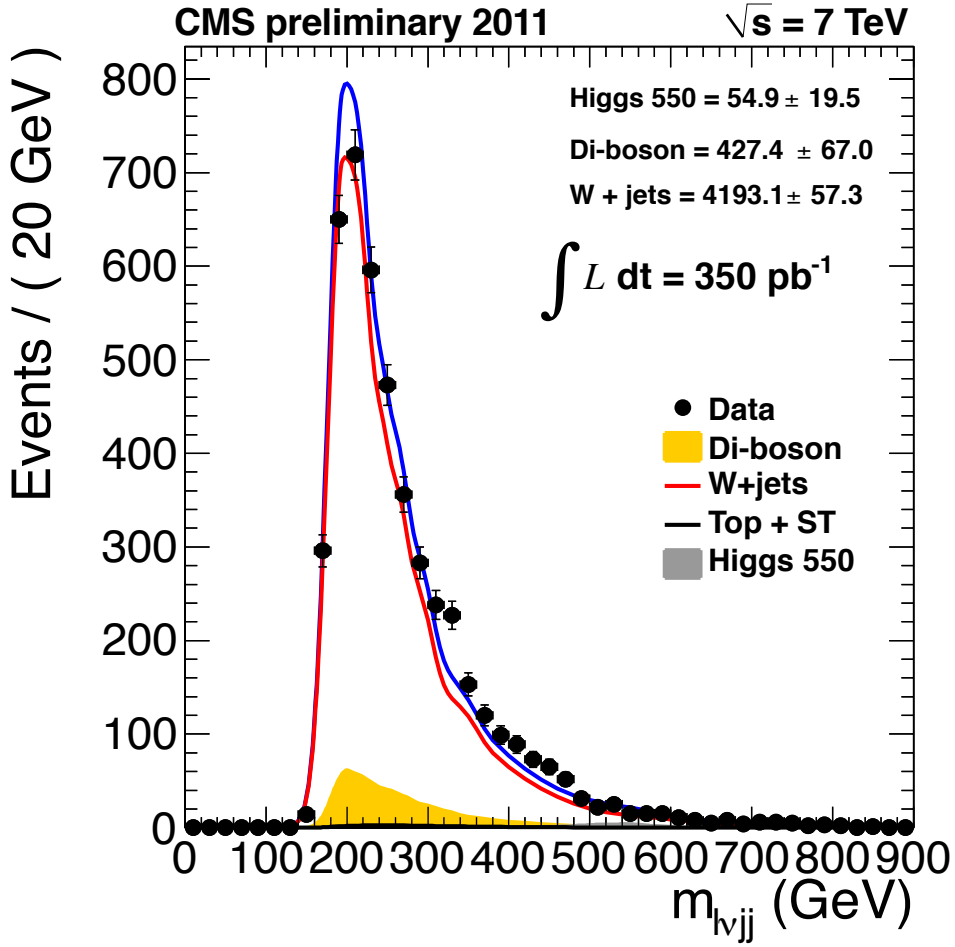
$m_{l\nu jj}$ fit: m_H = 500 GeV

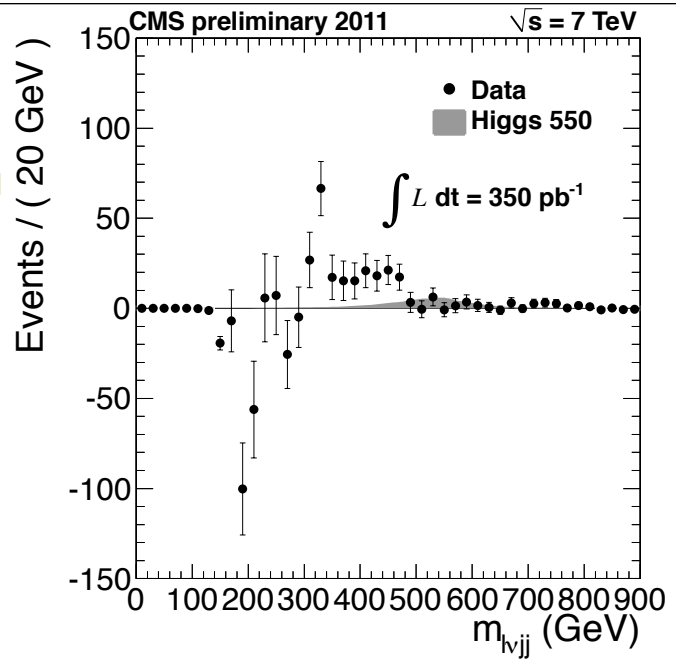


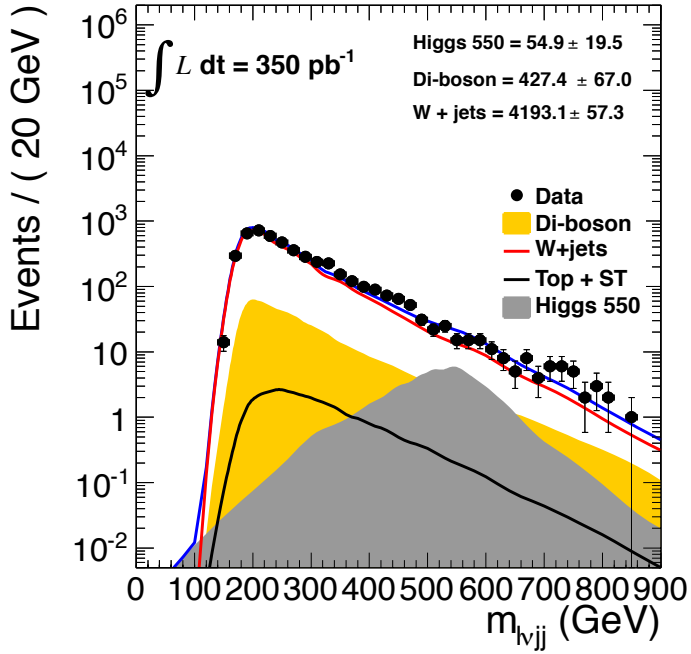




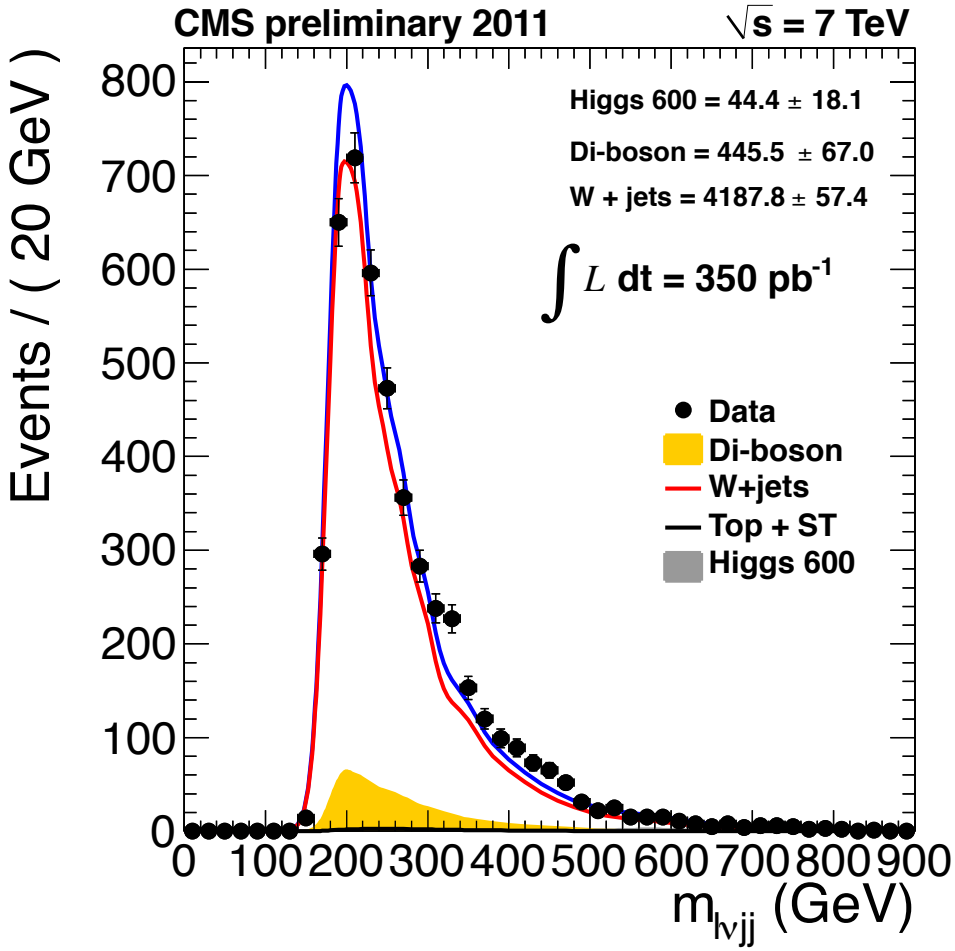
$m_{l\nu jj}$ fit: m_H = 550 GeV

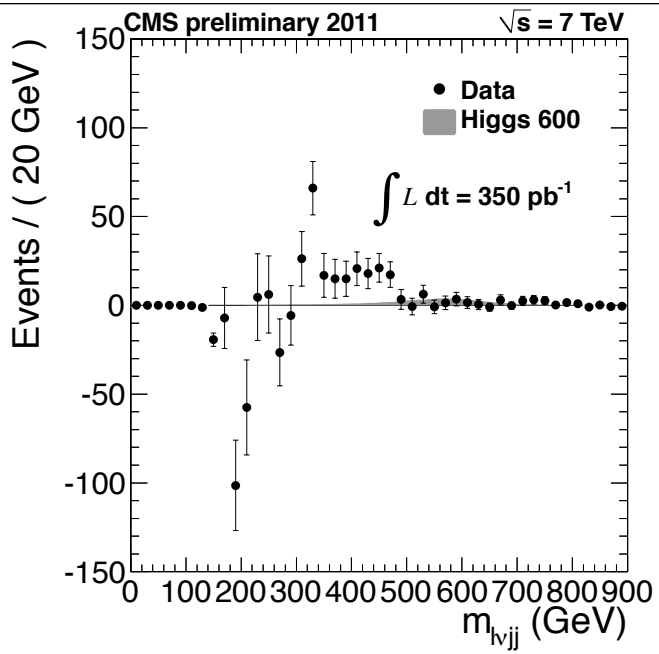


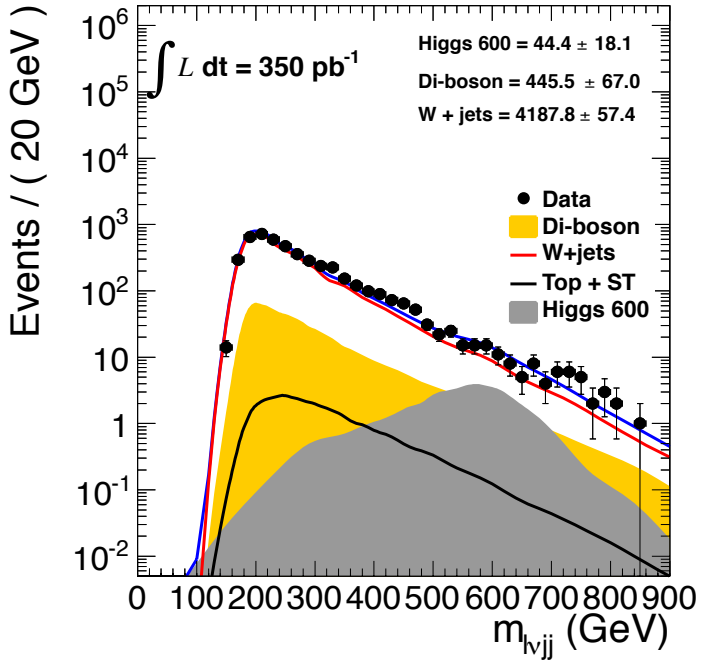




$m_{l\nu jj}$ fit: m_H = 600 GeV







Next steps

1. Compute acceptance (with systematics) for each mass point.

2. Compute lepton efficiency, and efficiency for other kinematic cuts. Include the QCD multi-jet in background estimation.

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

At the same time keep improving the data modeling, fit behavior, and systematics from fit procedure. Generate pseudo experiments for these.

I computed limits for few mass points using CMS Higgs Combination machinery. Haven't produced the band yet - needs lots of toy MC / CPU time. Will have soon.

backup slides

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.

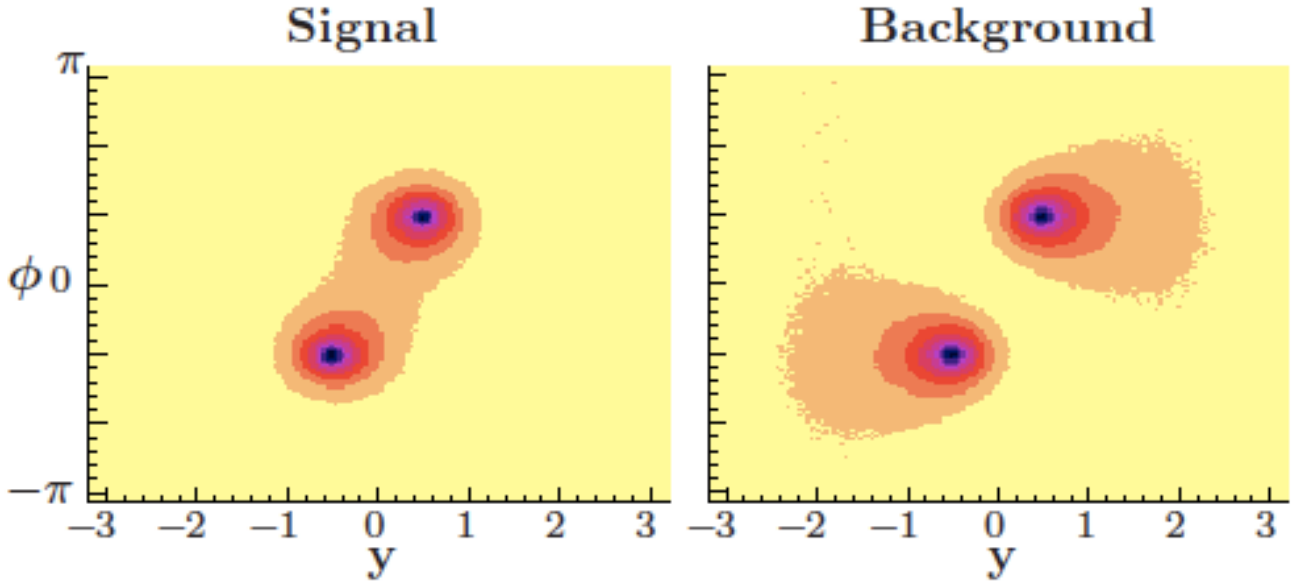
Documentation of RooKeysPdf: http://root.cern.ch/root/htmldoc/RooKeysPdf.html

CMS Higgs combination group also uses this class for templates

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

Some of last week's ideas aren't very helpful

- Use information about color correlation between between the two jets



arXiv:1001.5027

color pull:

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

But gives only marginal improvement, if any.

Include plot showing pull from WW and W+jets here

Similarly, kinematic fit should give some small improvement since we are already using the leptonic W mass constraint and most of the hadronic W mass constraint (by requiring the window 65--95 GeV).