Preliminary analysis using 350 pb^{-1} data

On behalf of Wjj 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:

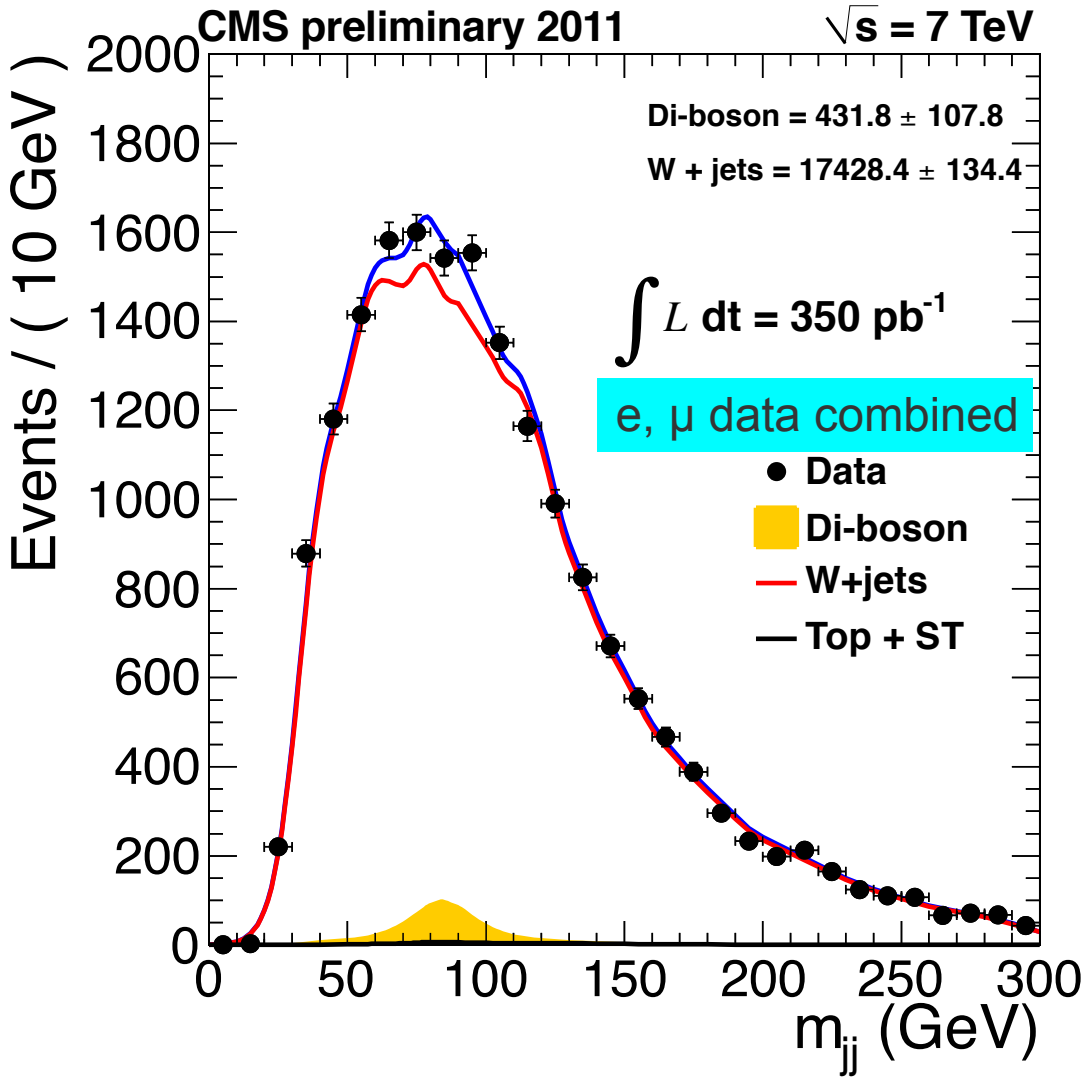
- p_T^{dijet} > 40 GeV
- Δη (j1, j2) < 1.5

Now we are using same cuts as CDF, but cutting tighter on Δη between the two jets to cope with higher background

✦cuts on Δϕ (jet, lepton) turned out to be very inefficient
✦cut on Δϕ (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

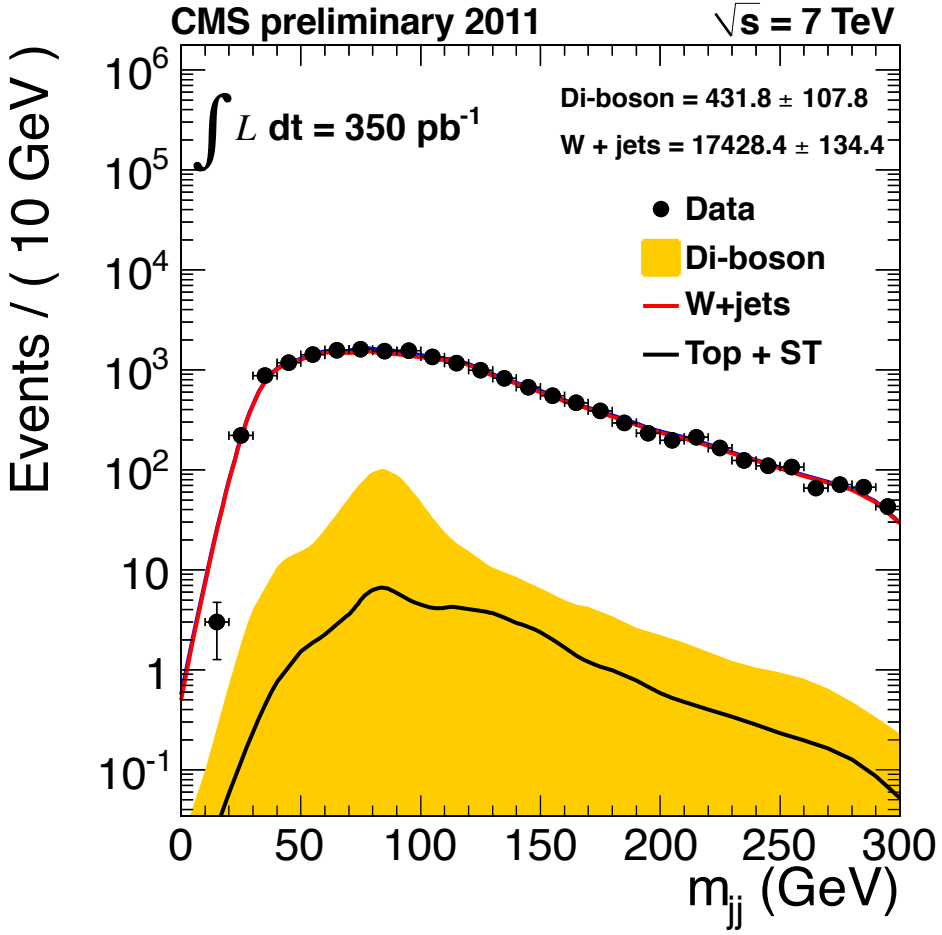
Take the shape from MC. Fit for the normalization. Blue curve shows the fit to data.

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.

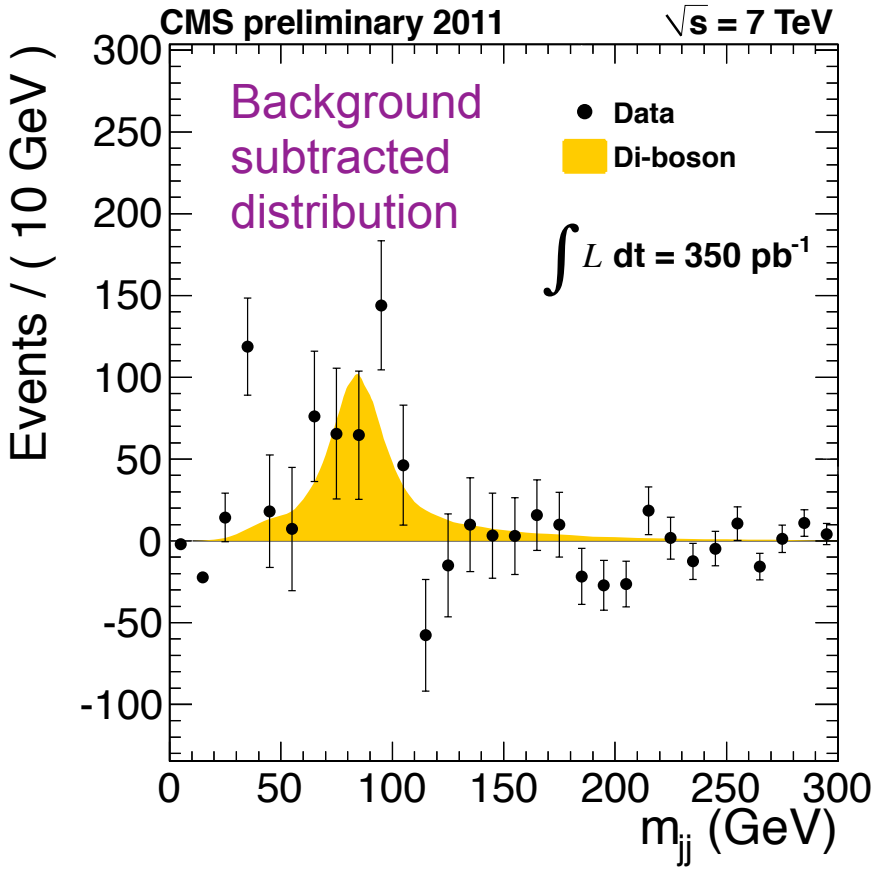
Background subtracted distribution on the next slide

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

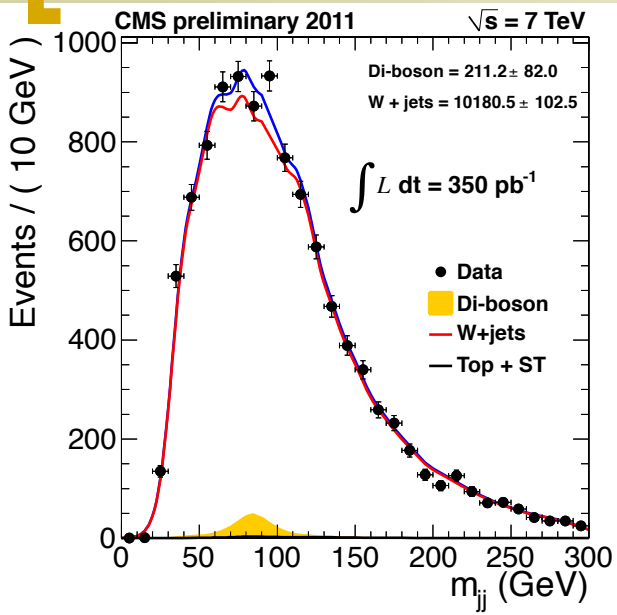
on the log scale

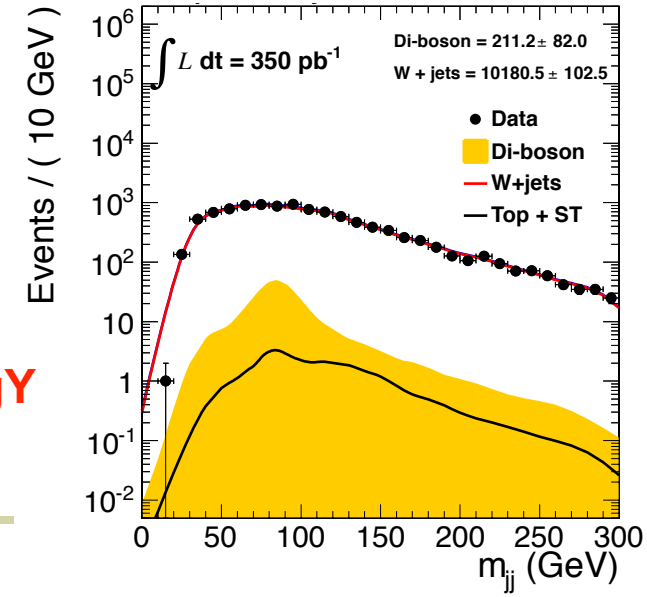


In the W mass window $65 < m_{jj} < 95$ GeV we get:
233 di-boson, **4428** W+jets events

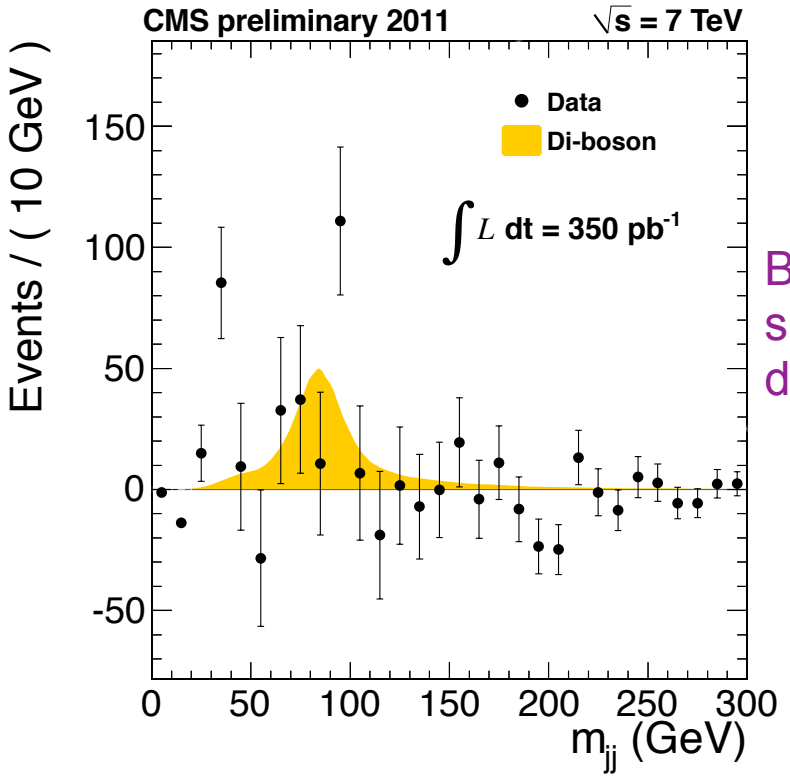


Template fit to m_{jj} in W+2 jet events: μ data





logY

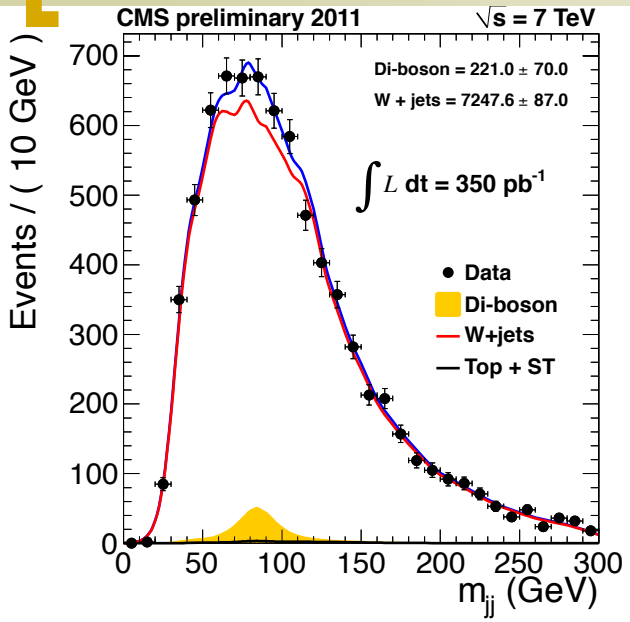


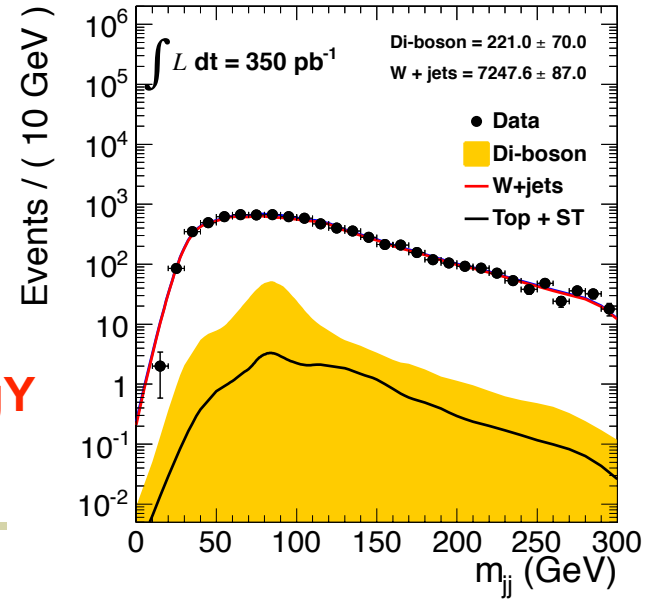
Background subtracted distribution

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

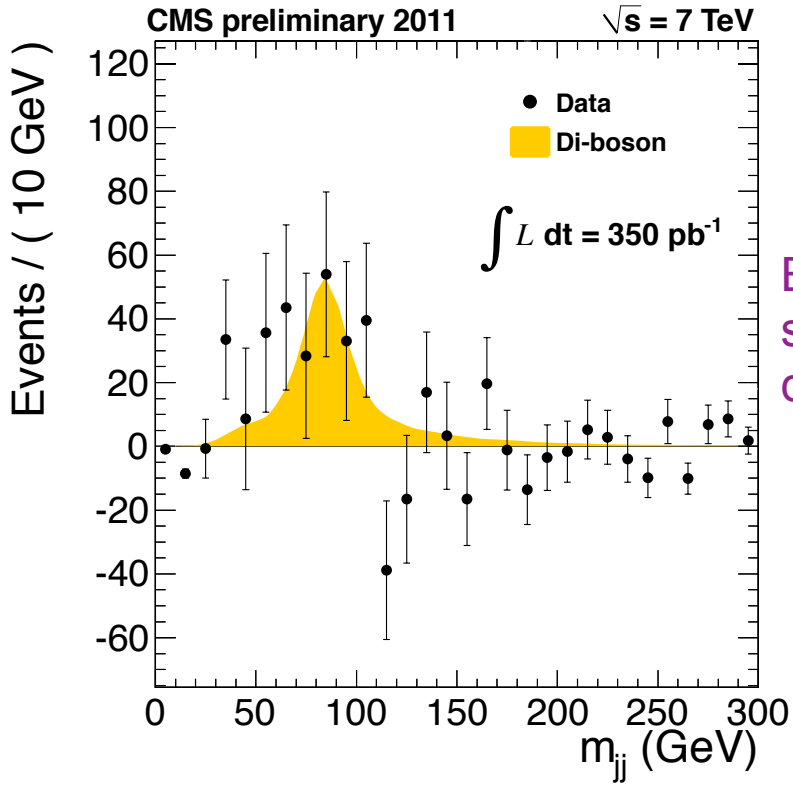
114 di-boson, **2586** W+jets events

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





logY

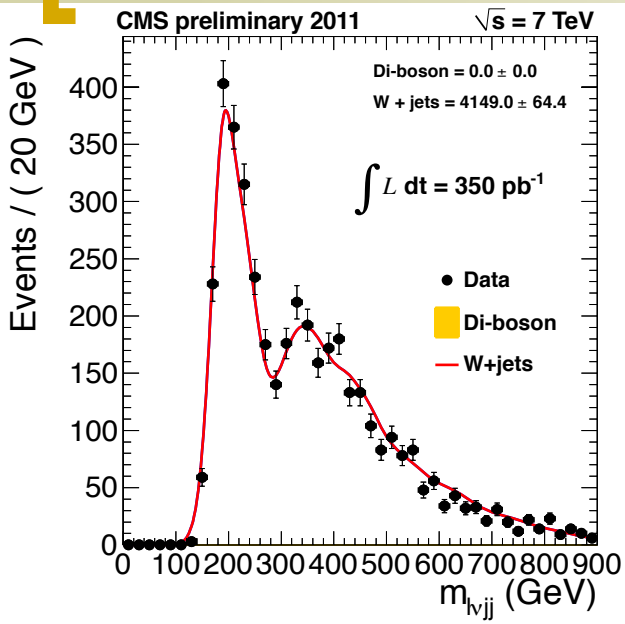


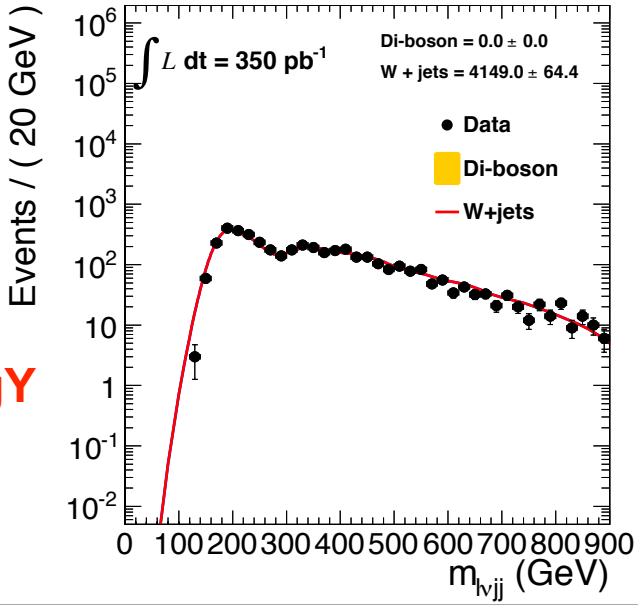
Background subtracted distribution

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

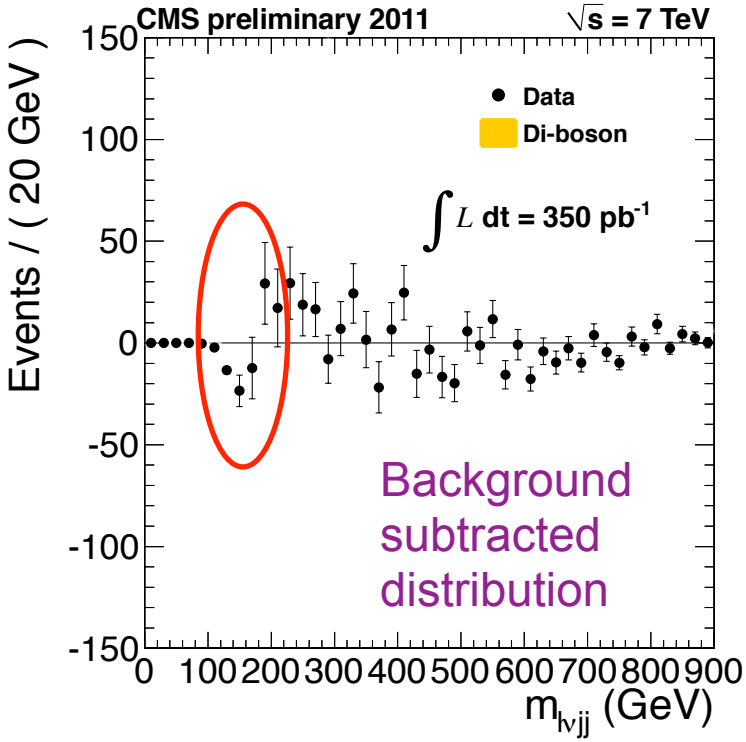




logY

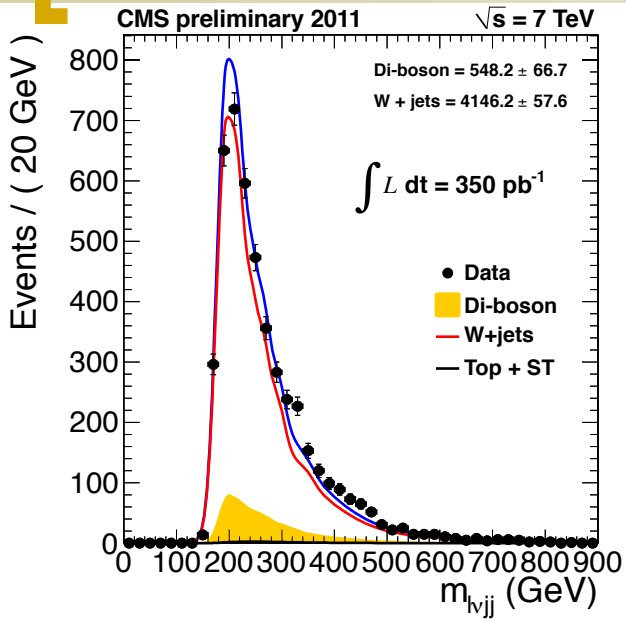
m_{jj} < 60 GeV OR m_{jj} > 200 GeV

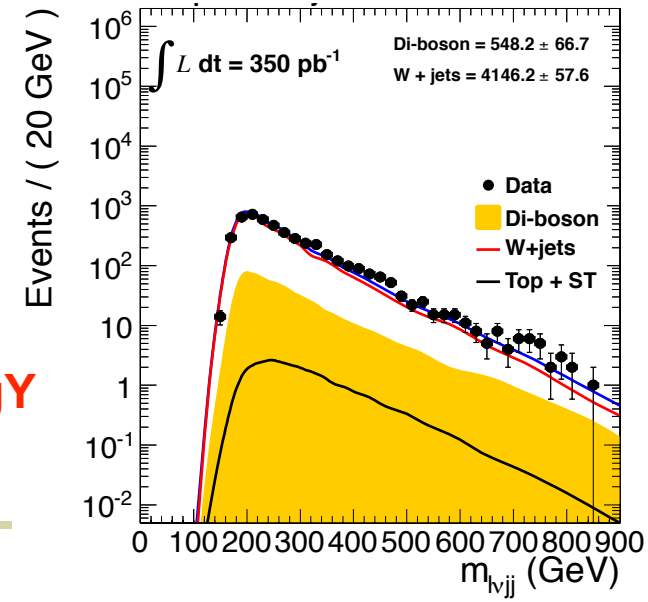
should be able to fit the W+ jets shape



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

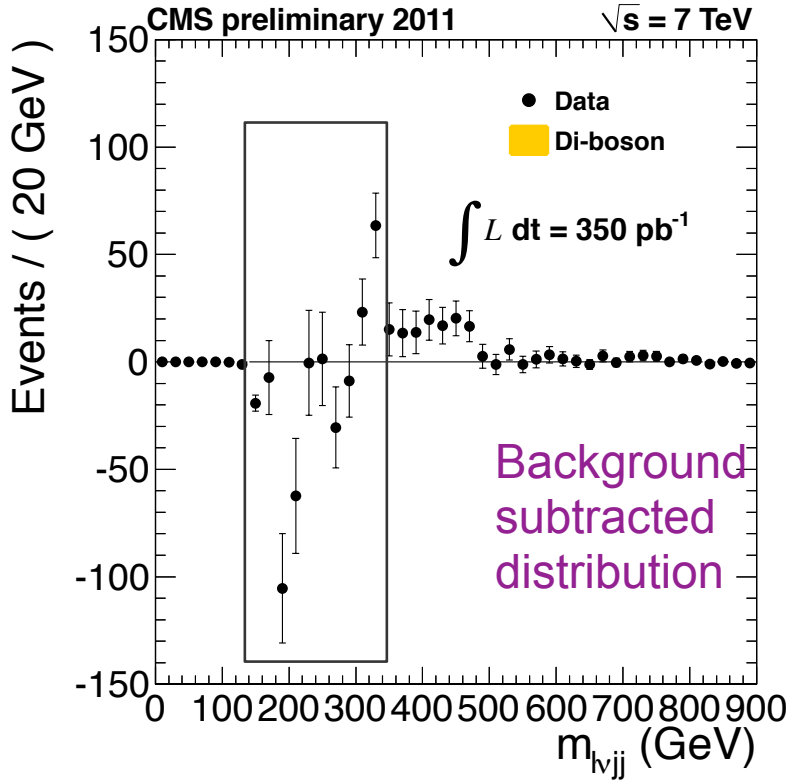
$m_{l\nu jj}$ distribution: $65 < m_{jj} < 95$ GeV





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. Investigating if this dip is caused by inadequacy of the template or resolution effect etc.

Status on the MC generation of new physics

1. Steve Mrenna gave us data card for generating techni-color model in Pythia:

techni-ρ → W + techni-π → W(lν) + jj
techni-ρ has mass 450 GeV, techni-π mass in 150 GeV, W decays leptonic, techni-π decays to two gluon jets

Jake Anderson managed to produce 10k events at FNAL. Analyzing it.

2. Steve Mrenna also gave us data card for generating Higgs (150 GeV) with anomalous decay:

WH/ ZH → W(lν) or Z(ll) + jj
W/Z decays leptonic, H decays to two light flavor jets

3. Adam Martin and Matt Buckley gave us .lhe file (for MadGraph) of their Z' model

Currently working to generate models 2, 3

Next steps for minimal analysis

1. Keep improving the data modeling, fit behavior, and systematics from fit procedure. Generate pseudo experiments for these.

2. Try various MCs to derive m_{jj} template and determine upper bound on uncertainty due to this shape.

3. Include systematics in the likelihood
   - JES/JER are easy to include
   - For uncertainty in template due to NLO effect need NLO MC

[MCFM is a negative weight MC with no interface to CMSSW. Can try to get GEN level shape and smear it for detector response with help from MadGraph sample. Similar story with Blackhats − have started talking to them to figure out details.]

   - Similarly, need MC with Q^2 up/down variation

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.

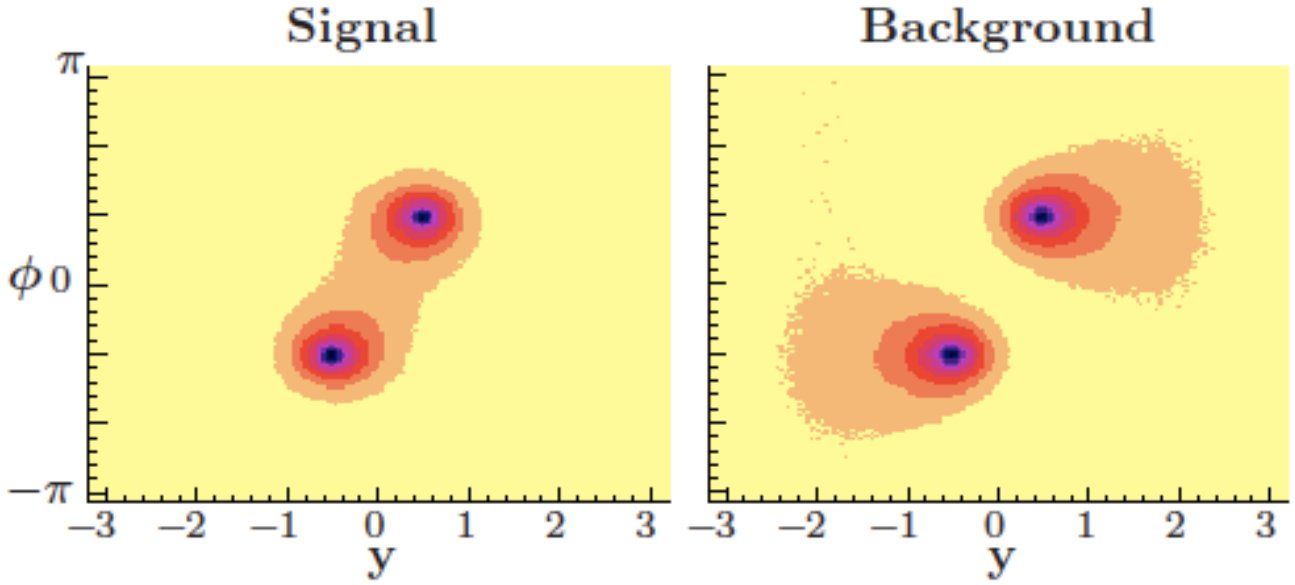
Documentation of RooKeysPdf: http://root.cern.ch/root/htmldoc/RooKeysPdf.html

CMS Higgs combination group also uses this class for templates

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

Some of last week's ideas weren'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).