Analysis status using 600 pb^{-1} data

On behalf of Wjj working group

(June 21, 2011)

The "minimal plan" for summer conferences

[1] Make m_{jj} invariant mass distribution from lepton+jets+MET events for both inclusive (NJet ≥ 2) and exclusive (NJet = 2) jet multiplicity and ~ same selection criteria as used in CDF/DØ analysis. ☑ Done

[2] Take all shapes from Monte Carlo and perform a fit to data extract normalization for W+jets and di-boson. Fix all other components.

☑ Under control, waiting for more W+jets MC

[3] Plot the distribution: Data − [all components except di-boson] in the range 0−300 GeV. We should clearly see the di-boson peak in right place and right magnitude. ☑ Done, under control

[4] Overlay expected contribution from new physics models: technicolor, Z', and WH with H(150)→qq. If time permits set limit on these models.

☐ Good progress, see Jake's talk

[5] Go through similar steps with dilepton+jets events.

☐ Progress, see Phil's talk

Event selection

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 > 40 GeV

Kinematic cuts to suppress W+jets:

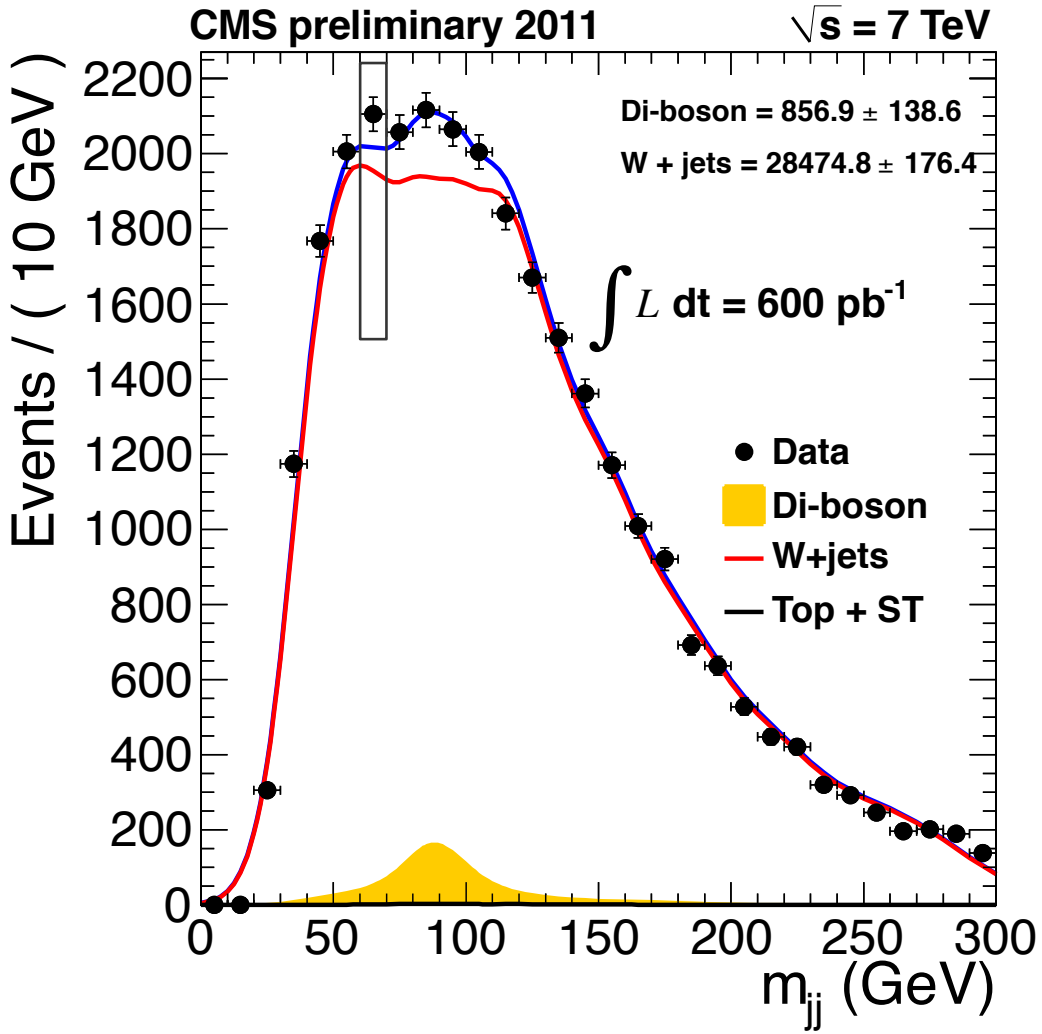
- p_T^{dijet} > 40 GeV
- $\Delta\eta$ (j1, j2) < 2.5
- $\Delta\phi$ (j1, MET) > 0.4

Now we are using same cuts as CDF / D∅

Analyzed ~ 600 pb^{-1} of data so far.

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

e, μ data combined



MC predicts ~ 700 di-bosons.
σ = 67 pb, BR = 0.22 x 0.7
Acceptance ~ 0.45 x 0.45
Efficiency ~ 0.7 x 0.8, Lumi = 600

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

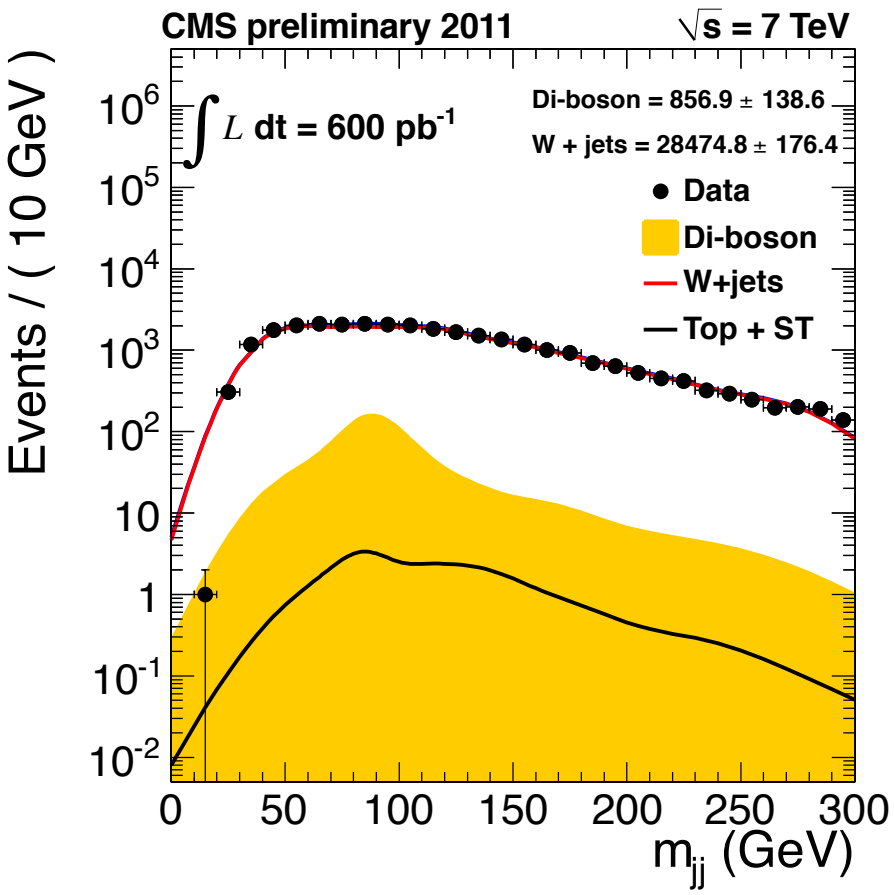
Wiggle near 60 GeV is not captured by W+jets MC template because sampling is not fine enough.

We'll get ~ 2 fb^{-1} W+jets MC today. This will help improve template.

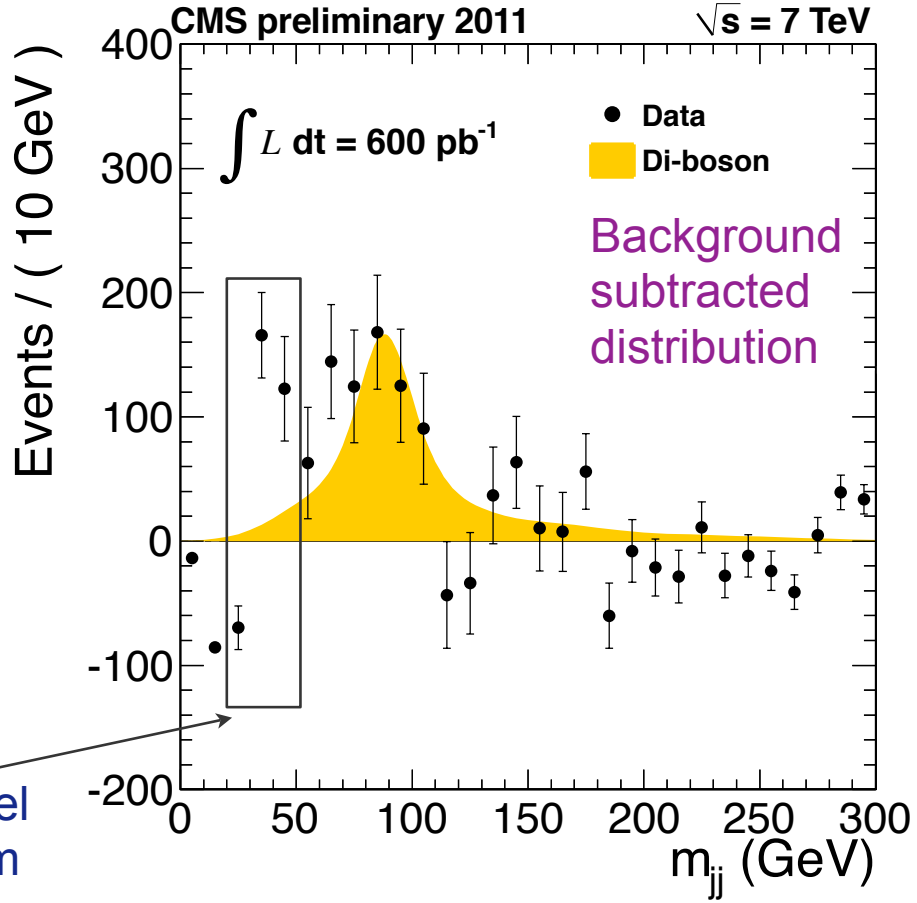
Background subtracted distribution on the next slide

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:
376 di-boson, **5802** W+jets events



Also see Jake's talk re ideas on how to model the sharply rising rough edge in the spectrum

Simulation of new physics models

1. Adam Martin + Steve Mrenna gave us data card for generating techni-color model in Pythia: ☐ 10k events. Adam is fine-tuning masses.

techni-ρ → W + techni-π → W(lν) + jj
π_T mass = 150 GeV, W decays leptonic, π_T decays to two jets

2. Steve Mrenna also gave us data card for generating Higgs (150 GeV) with anomalous decay: ☑ 50k events. This model was also used by CDF/DØ to compute acceptance for NP.

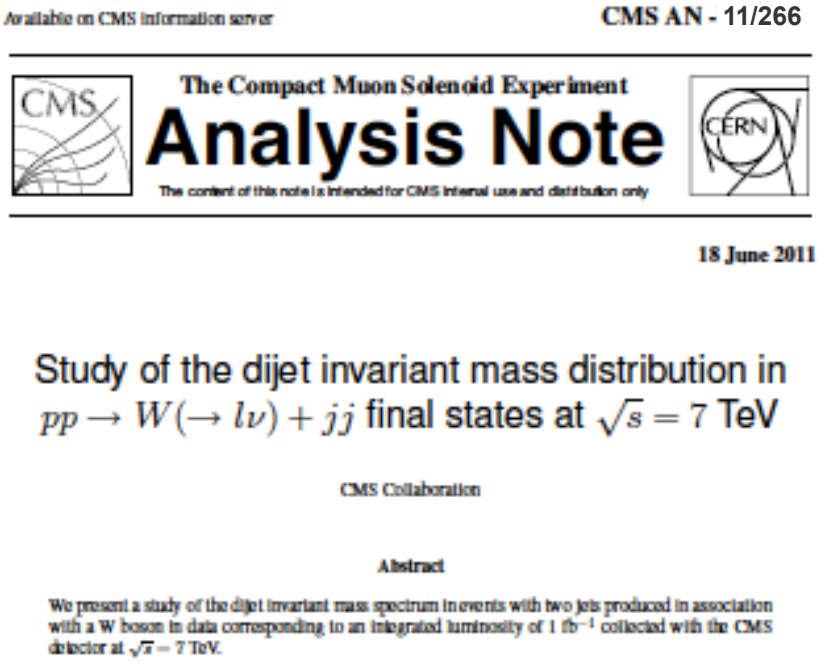
WH/ ZH → W(lν) or Z(ll) + jj
W/Z decays leptonic, H decays to two light flavor jets

3. Adam Martin and Matt Buckley gave us .lhe file (for MadGraph) of their leptophobic Z' (150)→qq model ☑ Done, generated 50k events

Production cross section of these processes at LHC is ~ 8 pb and acceptance x efficiency ~ 5%, so expect ~400 events from 1 fb^{-1}. We have sensitivity to detect excess of this magnitude or set real tight limits. See talk by Jake Anderson.

Documentation

We have started documenting the details of the analysis
CMS Analysis Note 2011/266



Available on CMS information server

CMS AN - 11/266

The Compact Muon Solenoid Experiment

Analysis Note

The content of this note is intended for CMS internal use and distribution only

18 June 2011

Study of the dijet invariant mass distribution in $pp \to W(\to l\nu) + jj$ final states at $\sqrt{s} = 7$ TeV

CMS Collaboration

Abstract

We present a study of the dijet invariant mass spectrum in events with two jets produced in association with a W boson in data corresponding to an integrated luminosity of 1 fb^{-1} collected with the CMS detector at $\sqrt{s} = 7$ TeV.

Ambitiously aiming to have first draft ready by next week. Afterward will focus on a PAS / Conf Note.

Missing steps in minimal analysis: systematics

[1.] Systematics from fit procedure. Need to study using pseudo expts.

[2.] Systematics from data modeling. Try various MCs to derive m_{jj} template and determine upper bound on uncertainty due to this shape. Also investigate uncertainty from NLO effect using MCFM/Blackhats.

[Unfortunately, MCFM is a negative weight MC with no interface to CMSSW. Similar story with Blackhats. Are in contact with developers.]

[3.] Include systematics in the likelihood

- JES/JER can be directly included as nuisance parameters in LH
- For uncertainty in template due to NLO effect need some NLO MC
- Factorization/ normalization scale: use Q^2 up/down variation
- Propagate uncertainty in the top pair and single top production cross section and NJet survival rate for the former

A careful estimation of all systematics will take some time. Meanwhile we will start with the simpler ones and will work through the rest.

Be prepared to answer some questions

I have heard the following sort of questions being asked to CDF/D∅ speakers:

➡Can you show me the analysis results using inclusive ≥2 jets ? Because a third of the signal has more than 2 jets.

➡Do you understand the uncertainty in estimation of the amount of top background which survives the Njet=2 requirement ?

➡How much the single top shape and normalization affect the overall background level between 130−170 GeV ?

➡Why do you trust your AlpGen/ Sherpa / MadGraph to get the W+jets shape right ? The shape between 100−200 GeV is quite sensitive to even small change in quark/gluon ratio which is hard to get right. I don't trust that your systematics accounts for all the uncertainties.

➡Is your jet energy scale right ? What fraction of your jets are quark initiated and what fraction light flavor ? How different is JEC for them ?

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