Study of M_{jj} Mass Spectrum and the CDF anomaly in $W(\to l\nu)$+jj data at CMS

Kalanand Mishra

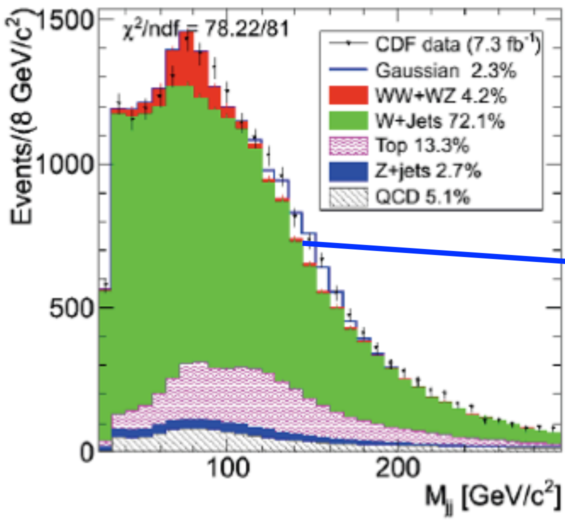
Fermilab

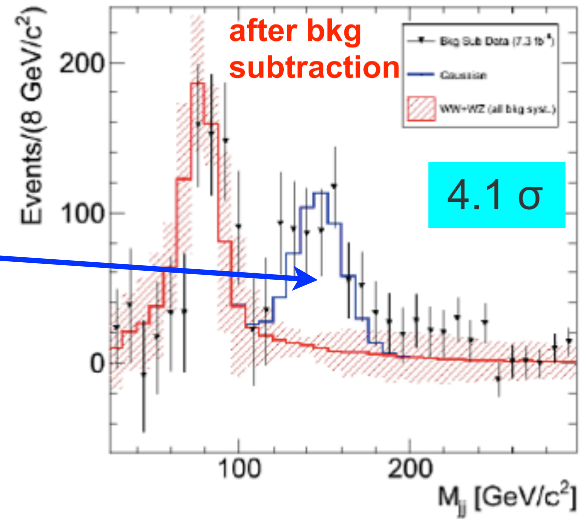
On behalf of CMS Electroweak Mjj subgroup

Disclaimer: CMS results/plots presented in these slides are very preliminary and are NOT approved for presentation outside the CMS.

LPC Physics Forum, September 22, 2011

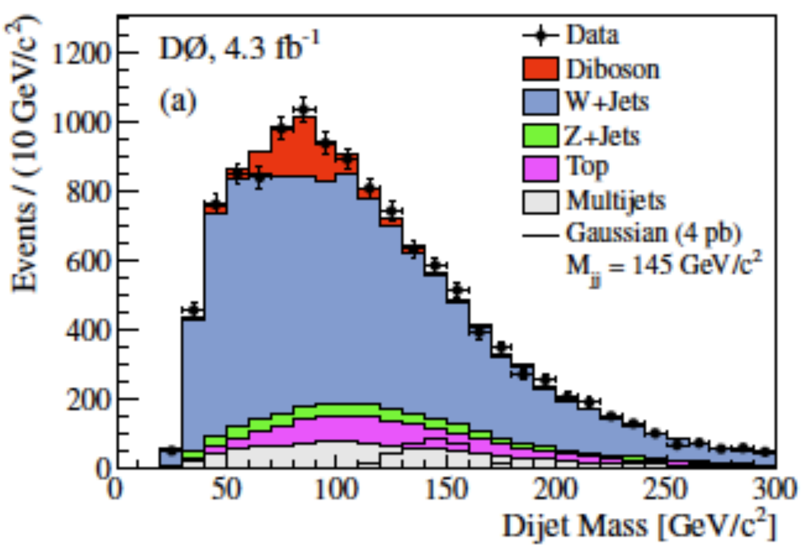
Bump or no bump !!!

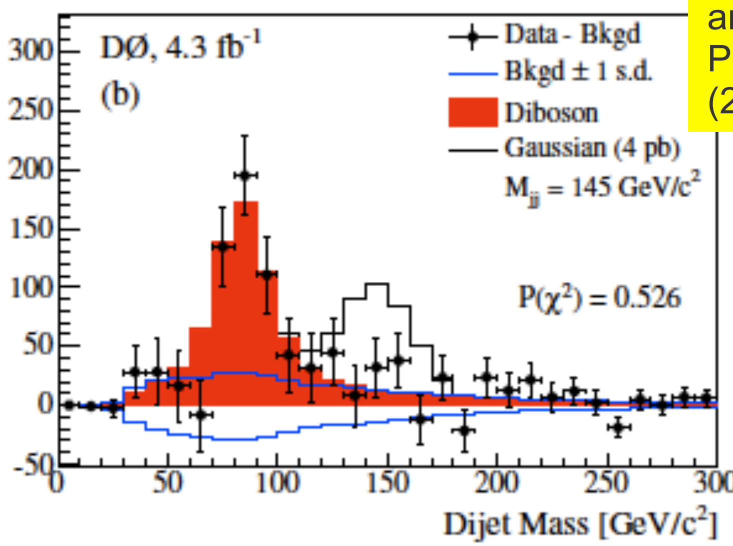




arXiv: 1101.6079, PRL 106:171801 (2011)

- ✦CDF W+jj data doesn't have the featureless falloff of dijet mass spectrum
  - -An excess, peaked at 145 GeV, width = 15 GeV (≈ jet resolution)
- ✦prod. cross section 4 pb

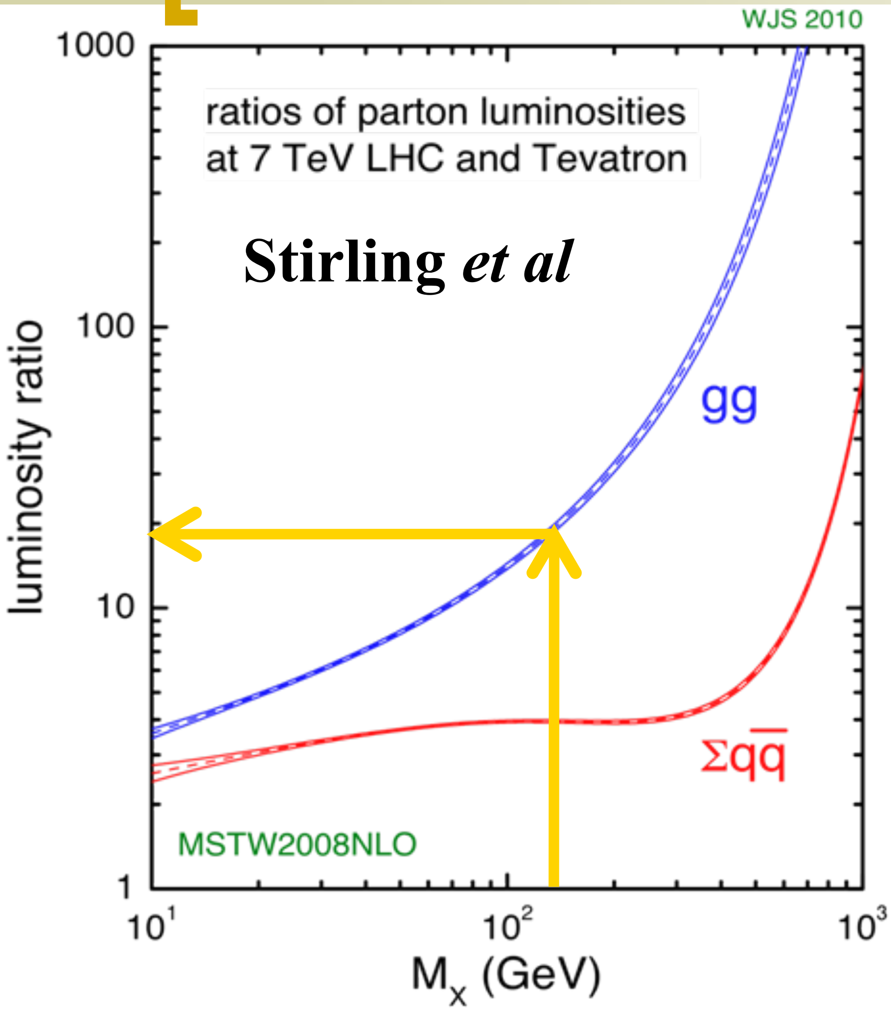




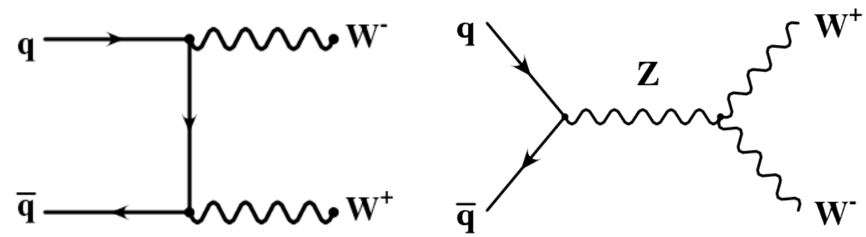
arXiv: 1106.1921, PRL 107:011804 (2011)

- ✦ DØ excludes 4 pb excess at 145 GeV @99.9999% CL

LHC vs Tevatron: can CMS probe it ?



$q\bar{q}$ → WW, WZ cross section at 7 TeV is **~ 3.5 times** that at 2 TeV



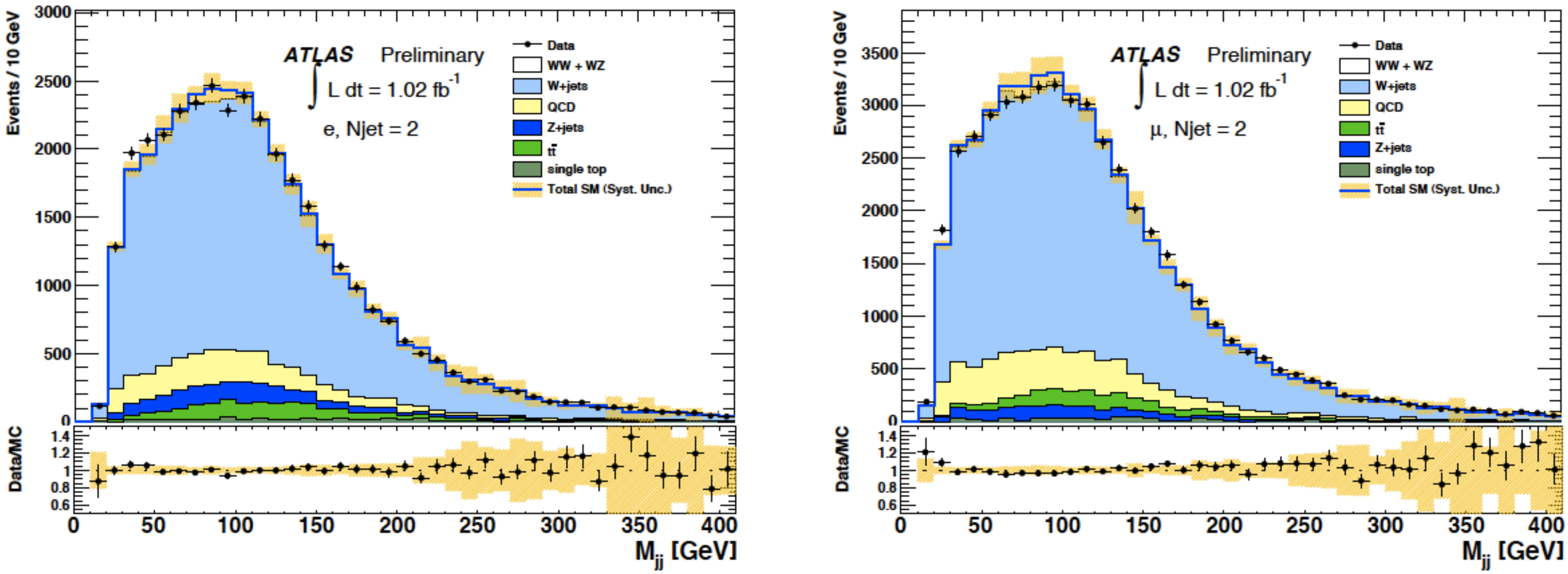
Major backgrounds are W/Z+jets, single top & $t\bar{t}$, QCD multi-jet etc. which rise **by factor 20** due to rise in qg and gg cross sections

⇒ Small signal, worse S/N

Yes, but the S/B is much worse, and stronger cuts need to be applied in order to extract the signal

Life is hard if looking for qqbar signal at LHC

ATLAS result shown in EPS



- Get swamped by W+jets
- **See no diboson peak**, nothing other than W+jets: S/B →0
- Large syst uncertainty, do not even show the bkg-subtracted plot
- Instead plot data/MC distribution which obscures any discrepancy

Clearly, much worse than Tevatron experiments

ELM recommendation to improve S/B @LHC

Testing CDF's Dijet Excess and Technicolor at the LHC

Estia Eichten[1*], Kenneth Lane[2†] and Adam Martin[1‡]
[1]Theoretical Physics Group, Fermi National Accelerator Laboratory
P.O. Box 500, Batavia, Illinois 60510
[2]Department of Physics, Boston University
590 Commonwealth Avenue, Boston, Massachusetts 02215

July 20, 2011

- **Cuts to reduce the main background without removing potential new physics signals**

arXiv:1107.4075

Main recommendations

- Lead jet p_T > 40 GeV, second jet p_T > 30 GeV
- Dijet p_T > 45 GeV
- $\Delta\eta$ (j1, j2) < 1.2
- Plus, some model-dependent cuts for TC which we can ignore

We adopt the ELM cuts and other improvements.

CMS analysis: Data and MC samples

LPC-based institutions have played important role in this analysis

MC:Process in CMSSW_4_2_X

MC ID	name	σ LO(NLO) [pb]
1000030	/WW_TuneZ2_7TeV_pythia6_tauola/Summer11-PU_S4_START42	27.8 (42.9)
1000031	/WZ_TuneZ2_7TeV_pythia6_tauola/Summer11-PU_S4_START42	10.4 (18.3)
4000041	/WJetsToLNu_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42	24640 (31539)
MC:2924	/WJetsToLNu_TuneZ2_matchingup_7TeV-madgraph-tauola/Summer11-PU_S4_START42	
MC:2924	/WJetsToLNu_TuneZ2_matchingdown_7TeV-madgraph-tauola/Summer11-PU_S4_START42	
MC:2924	/WJetsToLNu_TuneZ2_scaleup_7TeV-madgraph-tauola/Summer11-PU_S4_START42	
MC:2924	/WJetsToLNu_TuneZ2_scaledown_7TeV-madgraph-tauola/Summer11-PU_S4_START42	
MC:2924	/DYJetsToLL_TuneZ2_M-50_7TeV-madgraph-tauola/Summer11-PU_S4_START42	2469 (3048)
4000040	/TTJets_TuneZ2_7TeV-madgraph-tauola/Summer11-PU_S4_START42	121 (157.5)
	/Tbar_TuneZ2_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42	(1.44)
	/Tbar_TuneZ2_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42	(22.65)
	/Tbar_TuneZ2_tW-channel-DS_7TeV-powheg-tauola/Summer11-PU_S4_START42	(7.87)
	/T_TuneZ2_s-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42	(3.19)
	/T_TuneZ2_t-channel_7TeV-powheg-tauola/Summer11-PU_S4_START42	(41.92)
	/T_TuneZ2_tW-channel-DS_7TeV-powheg-tauola/Summer11-PU_S4_START42	(7.87)
1000041	/QCD_Pt-30to80_EMEnriched_TuneZ2_7TeV-pythia/Summer11-PU_S4_START42	3866200
1000043	/QCD_Pt-80to170_EMEnriched_TuneZ2_7TeV-pythia6/Summer11-PU_S4_START42	139500
1000040	/QCD_Pt-30to80_BCtoE_TuneZ2_7TeV-pythia6/Summer11-PU_S4_START42	136804
1000042	/QCD_Pt-80to170_BCtoE_TuneZ2_7TeV-pythia/Summer11-PU_S4_START42	9360
1000039	/QCD_Pt-20_MuEnrichedPt-15_TuneZ2_7TeV-pythia6/Summer11-PU_S4_START42	136804

Data

/EG/Run2010A-Apr21ReReco-v1/AOD
/Electron/Run2010B-Apr21ReReco-v1/AOD
/SingleElectron/Run2011A-May10ReReco-v1/AOD
/SingleElectron/Run2011A-PromptReco-v4/AOD
/SingleElectron/Run2011A-05Aug2011-v1/AOD
/SingleElectron/Run2011A-PromptReco-v6/AOD
/Mu/Run2010A-Apr21ReReco-v1/AOD
/Mu/Run2010B-Apr21ReReco-v1/AOD
/SingleMu/Run2011A-May10ReReco-v1/AOD
/SingleMu/Run2011A-PromptReco-v4/AOD
/SingleMu/Run2011A-05Aug2011-v1/AOD
/SingleMu/Run2011A-PromptReco-v6/AOD

Data certification

Cert_136033-149442_7TeV_Apr21ReReco_Collisions10_JSON.txt
Cert_160404-163869_7TeV_May10ReReco_Collisions11_JSON_v3.txt
Cert_170249-172619_7TeV_ReReco5Aug_Collisions11_JSON_v2.txt
Cert_160404-173692_7TeV_PromptReco_Collisions11_JSON.txt

Analyzed 2 fb^{-1} data

Acceptance thresholds and trigger

✦ $W \to l\nu$ reconstruction
- Muon: $p_T > 25$ GeV, $|\eta| < 2.1$, reconstructed as both global & tracker muon
- Electron: $E_T > 30$ GeV, $|\eta| < 2.5$ excluding $1.44 < |\eta| < 1.57$, ECAL seeded gsf electrons

✦ Require exactly two or three PF jets in the event
- corrected $p_T > 30$ GeV and $|\eta| < 2.4$
- $|\Delta R(\text{jet, lepton})| > 0.3$

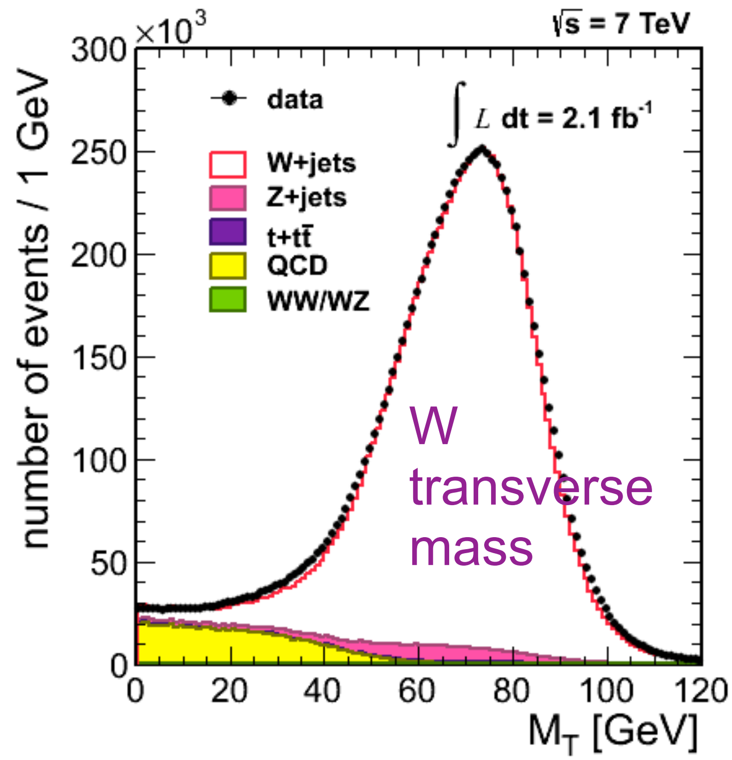
Triggers

For electron use: SingleEle || W_inclusive.
For muons: IsoMu_24 || IsoMu_17 || SingleMu_X

✦ For 2010 data (36 pb^{-1}) use single lepton triggers with $p_T > 17$ GeV (or lower)
✦ For 2011 data before June TS (~200 pb^{-1}) use single IsoMu_17 and Ele_27
✦ For 2011 data after June TS also use "inclusive" W trigger for electron: keeps electron $E_T > 25$ GeV, pf MHT > 25 GeV, W $m_T > 40$ GeV. Iso_mu_24 for muons.

Lepton selection – muon

✦Quality Requirements
- ≥10 tracker hits, ≥1 pixel hits
- ≥1 good muon chamber hit
- Both inside-out & outside-in reconstruction
- Track matching with ≥2 segments in the muon stations
- $\chi2$/ndf < 10 global fit
- Cosmic veto: impact parameter $|d_{xy}|<$ 0.02 cm (w.r.t. the beam spot)



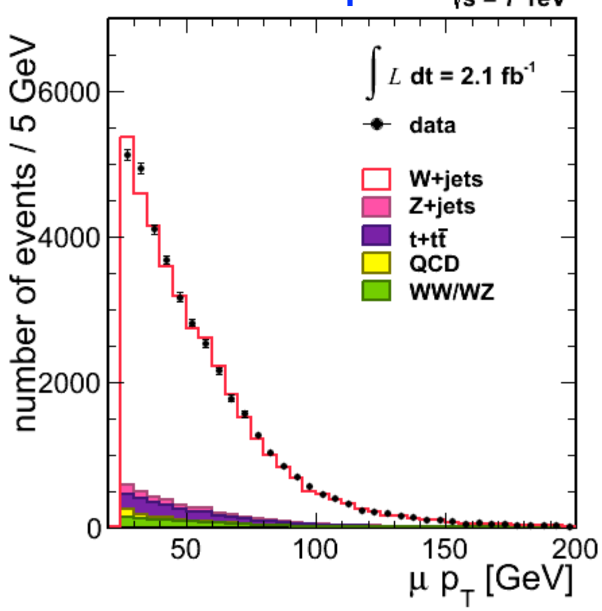
✦Isolation
- Combined relative isolation (R=0.3)

$$I^{\rm rel}_{\rm comb} = \left\{ \sum (p_T(tracks) + E_T(em) + E_T(had)) - \pi r^2 . \rho_{Iso} \right\} / p_T(\mu) < 0.1$$

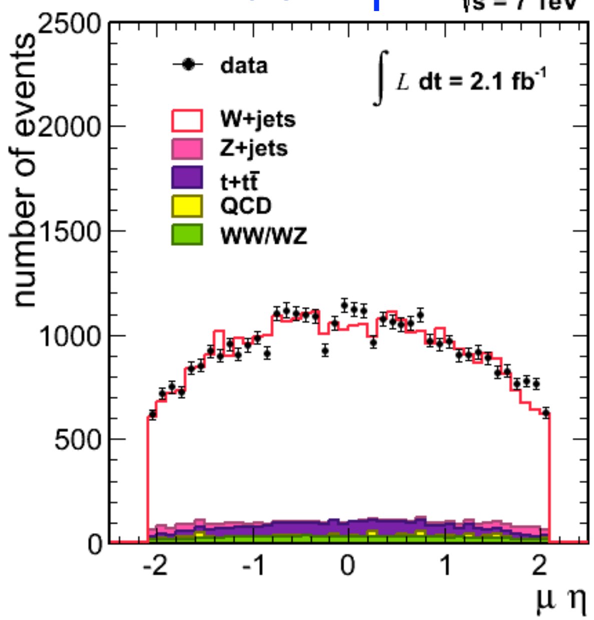
where ρ_{Iso} = PU density in $|\eta| < 2.5$, and r = radius of isolation cone = 0.3

Some control plots

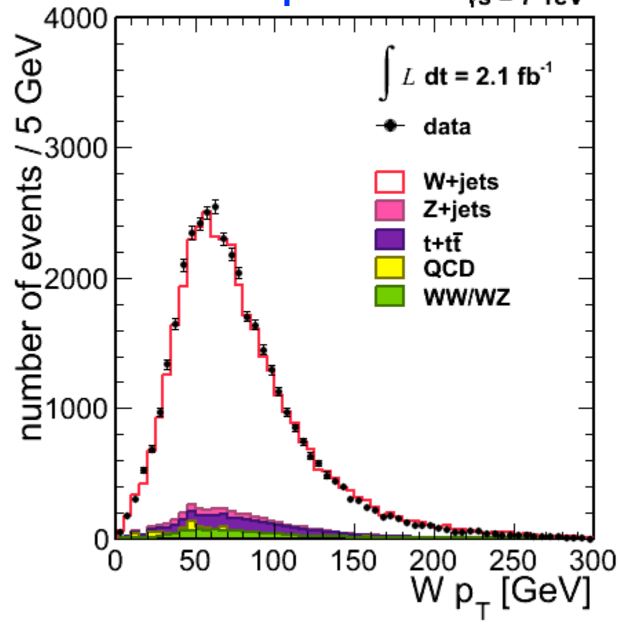
muon pT



muon η



W pT



❖ **Control Plots don't show any significant discrepancies between Data and MC**

Lepton selection – electron

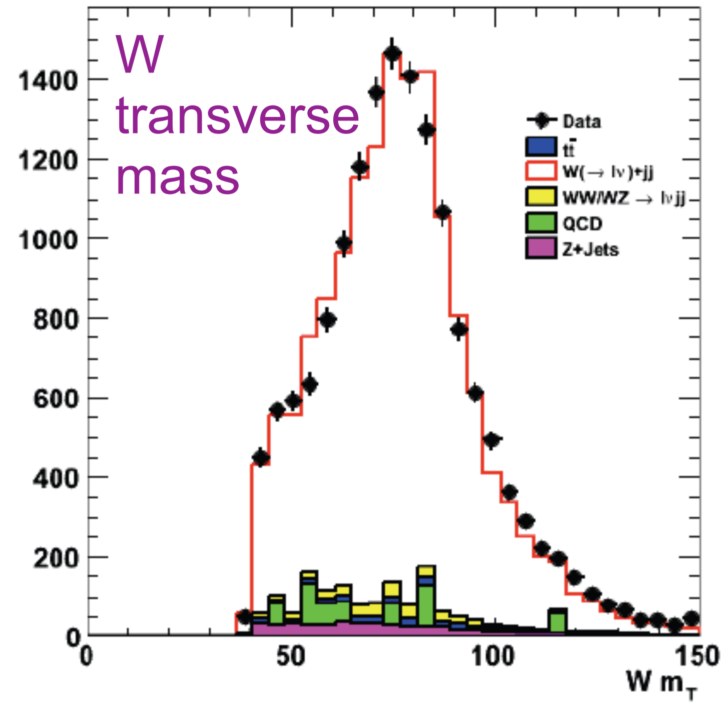
Use VBTF working point 70 with a minor change in isolation cut

http://twiki.cern.ch/twiki/bin/view/CMS/SimpleCutBasedEleID

Conversion Rejection	
missing hits $\leq$	0
Dist	0.02
$\Delta cot\theta$	0.02

Combined Isolation	0.05	after PU subtraction

Electron ID	EB	EE
$\sigma_{i\eta i\eta}$	0.01	0.03
$\Delta\phi$	0.03	0.02
$\Delta\eta$	0.004	0.005
HoE	0.025	0.025



Additional control plots are shown in the following presentation. They all look ok.

https://indico.cern.ch/getFile.py/access?contribId=4&sessionId=1&resId=0&materialId=slides&confId=154503

Jets & MET

Jet

- **At least two PFJets in each event**
- **Corrected p_T >30 GeV and $|\eta|$ <2.4**
- **$|\Delta R(l,j)| > 0.3$**
- **Default (JetMET POG recommended) charge hadron subtraction (PF2PAT/ PfNoPU), FastJet PU subtraction, L2L3 corrections and jet Id.**
- **Loose Jet Id criteria:**
  - fraction of energy due to neutral hadrons < 0.99;
  - fraction of energy due to neutral EM deposits < 0.99;
  - number of constituents > 1;
  - number of charged hadrons candidates > 0;
  - fraction of energy due to charged hadrons candidates > 0;
  - fraction of energy due to charged EM deposits < 0.99.

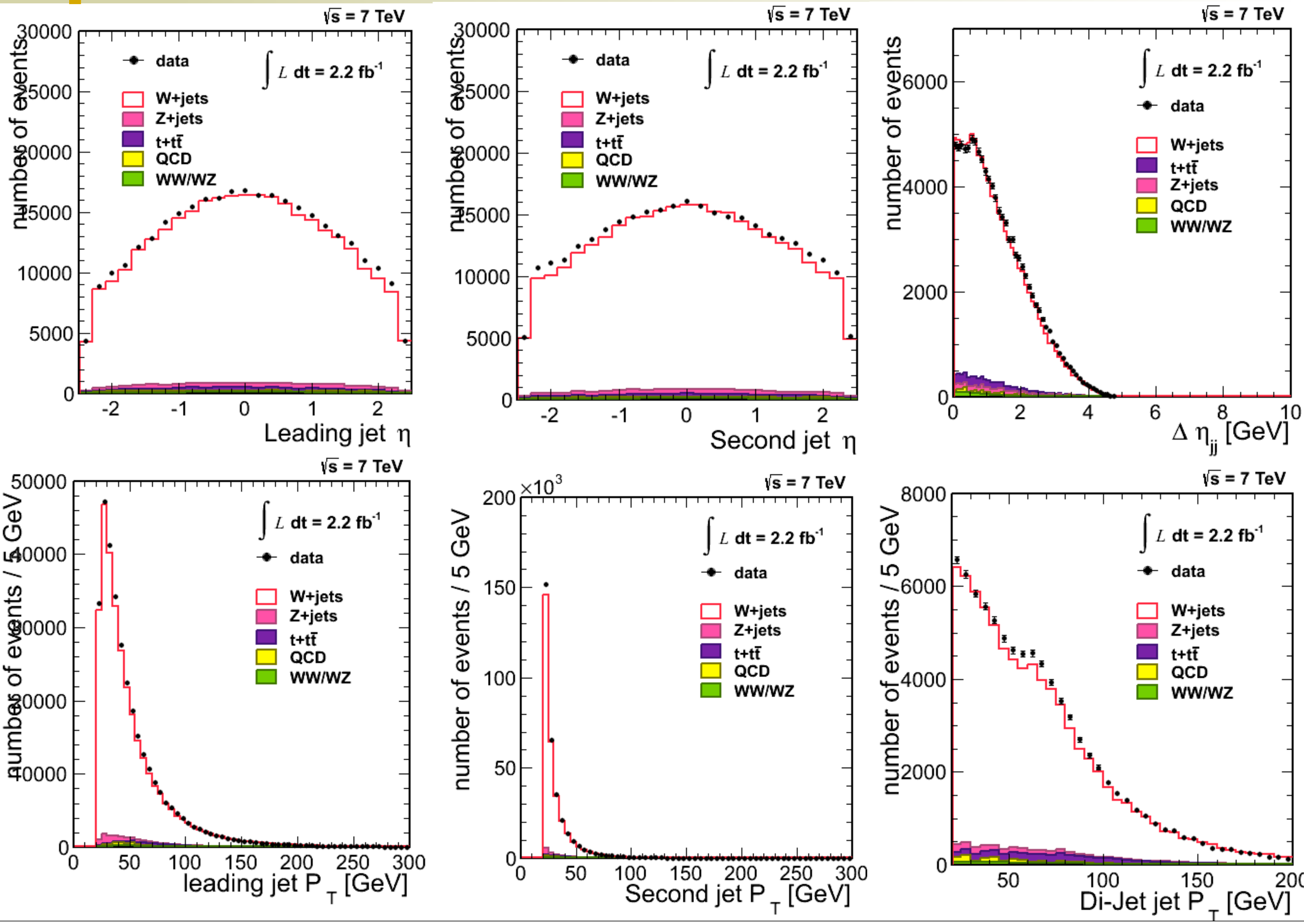
These identification criteria remove fakes due to calorimeter noise etc. Efficiency of passing these criteria for real hard jets is ~99.95%.

- **$p_T(jj) > 40$GeV and $|\Delta\eta(jj)|<1.5$**
- **PF MET > 30 GeV**

MET

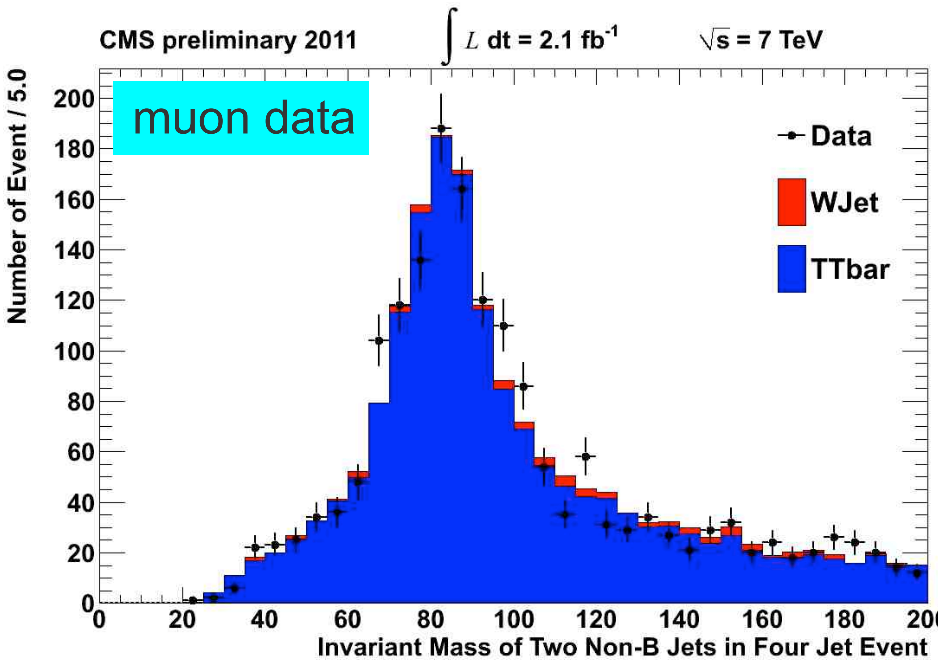
Use default particle flow MET. Require MET > 30 GeV.

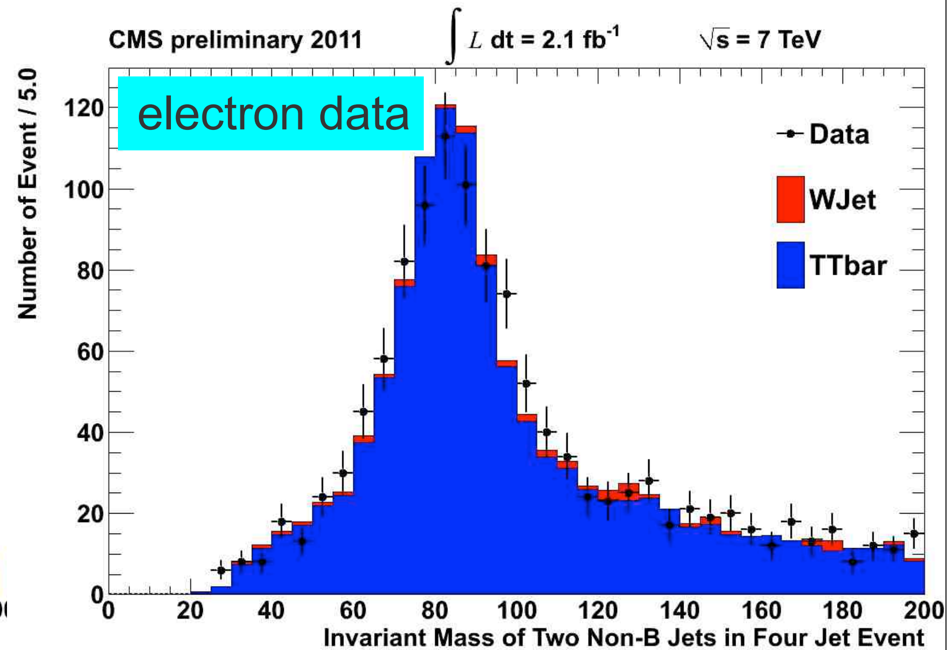
More control plots



Can we reconstruct hadronic W in CMS ? Yes

In top events reconstruct W peak “out-of-box” with good resolution





Require 4 jets above p_T 30 GeV (2 b-tags), and a leptonic W (muon: p_T>25 GeV or electron: E_T>30 GeV, MET>25 GeV). Then plot m_{jj} of the two non-b jets.

Peak is at the right place, and resolution is well modeled in MC.

Topological cuts

Start with simple topological cuts to suppress W+jets background

Leptonic W

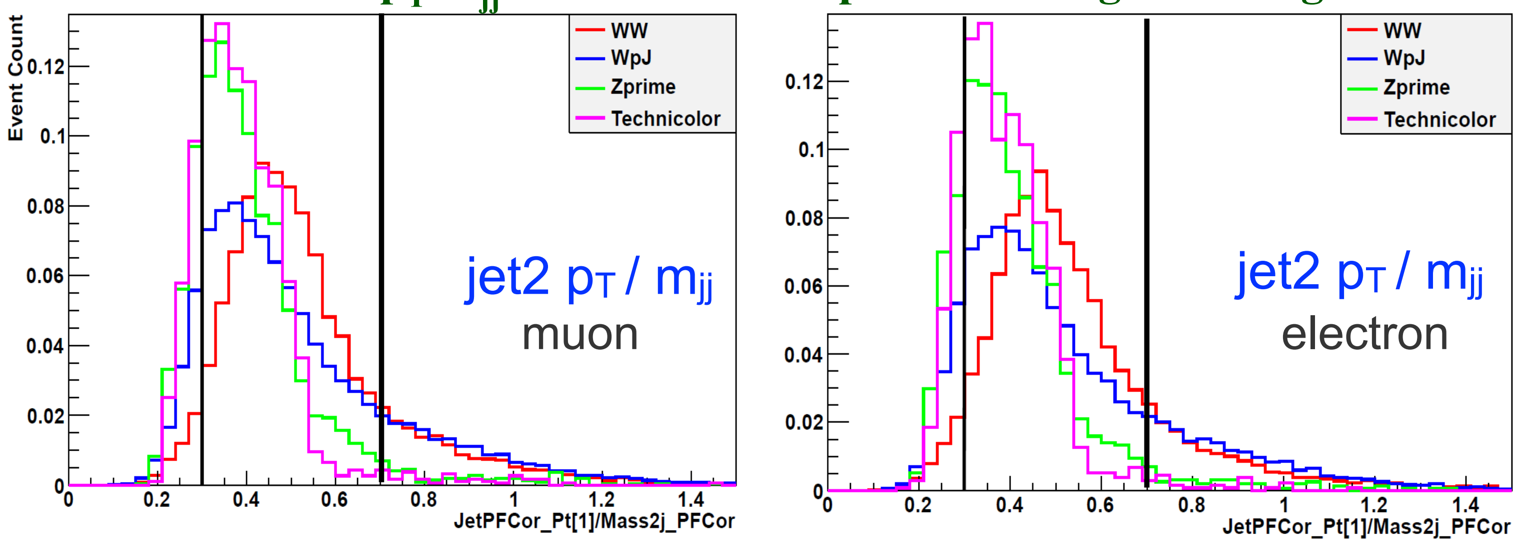
- pf MET > 30 GeV
- W transverse mass > 40 GeV

After this cut essentially pure W events are left : diboson, W+jets, and top pair + single top.

Dijet system

- jet1 p_T > 40 GeV, jet2 p_T > 30 GeV, Dijet p_T > 45 GeV
- $\Delta\eta$ (j1, j2) < 1.2

$0.3 < Jet2p_T/m_{jj} < 0.7$: Further improves the Signal/Background ratio



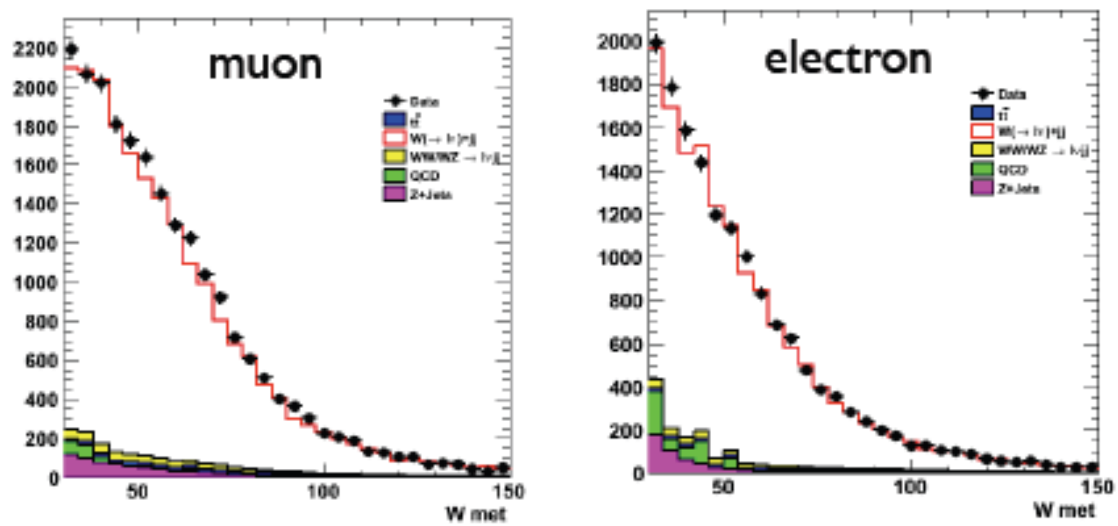
Template fit to extract various contributions

Perform an unbinned maximum likelihood fit of the m_{jj} distribution

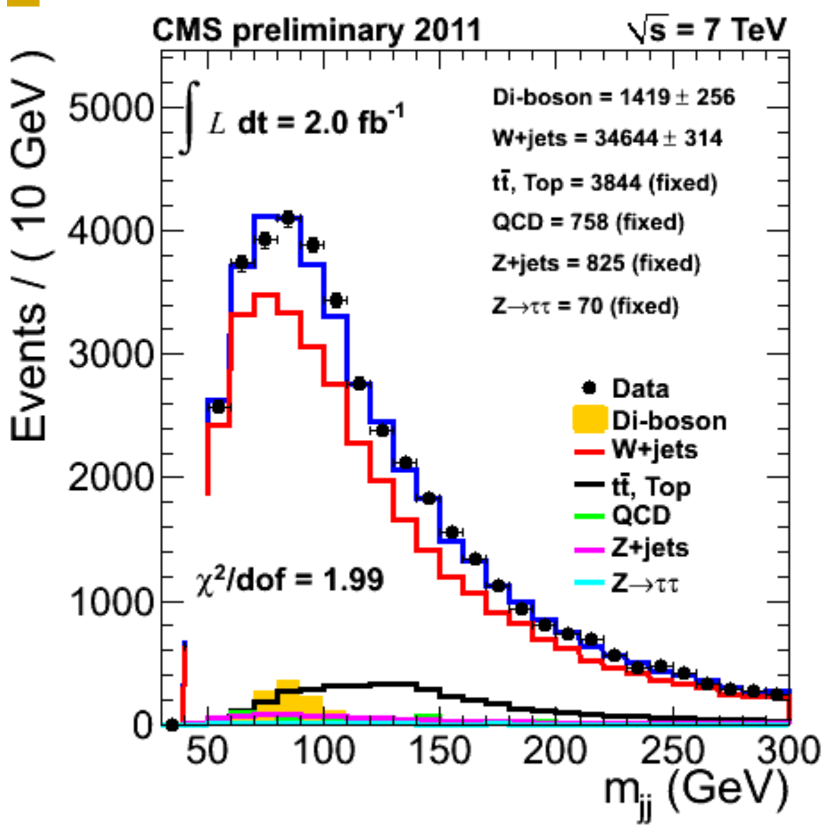
- in the range 40−300 GeV
- take all shapes except for QCD from MC; for QCD invert iso/id cuts in data
- fit for the absolute number of diboson and W+jets events
- fix top pair, single top, Z+jets, Z→ττ contributions to the NLO cross section
- float the jet energy scale within uncertainty

Then plot background-subtracted (i.e., data − bkg from fit) distribution to visually inspect the quality of the fit.

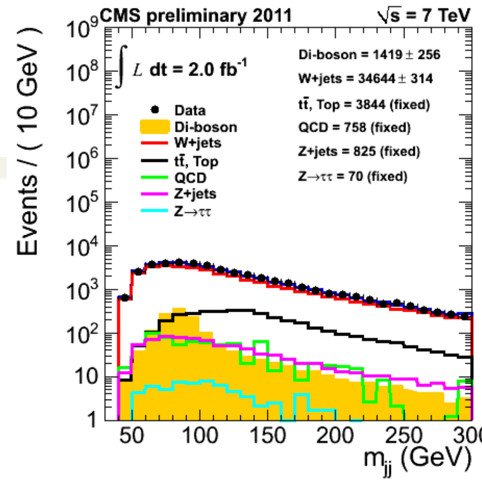
MET distributions in data



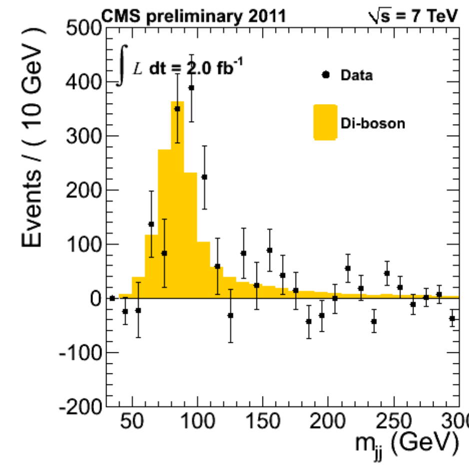
Fit results with improved selection



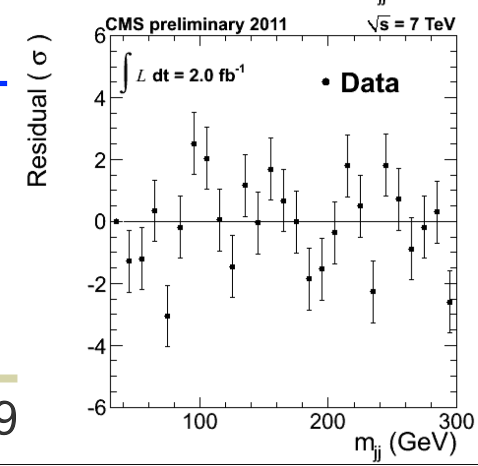
logY



bkg-subtracted



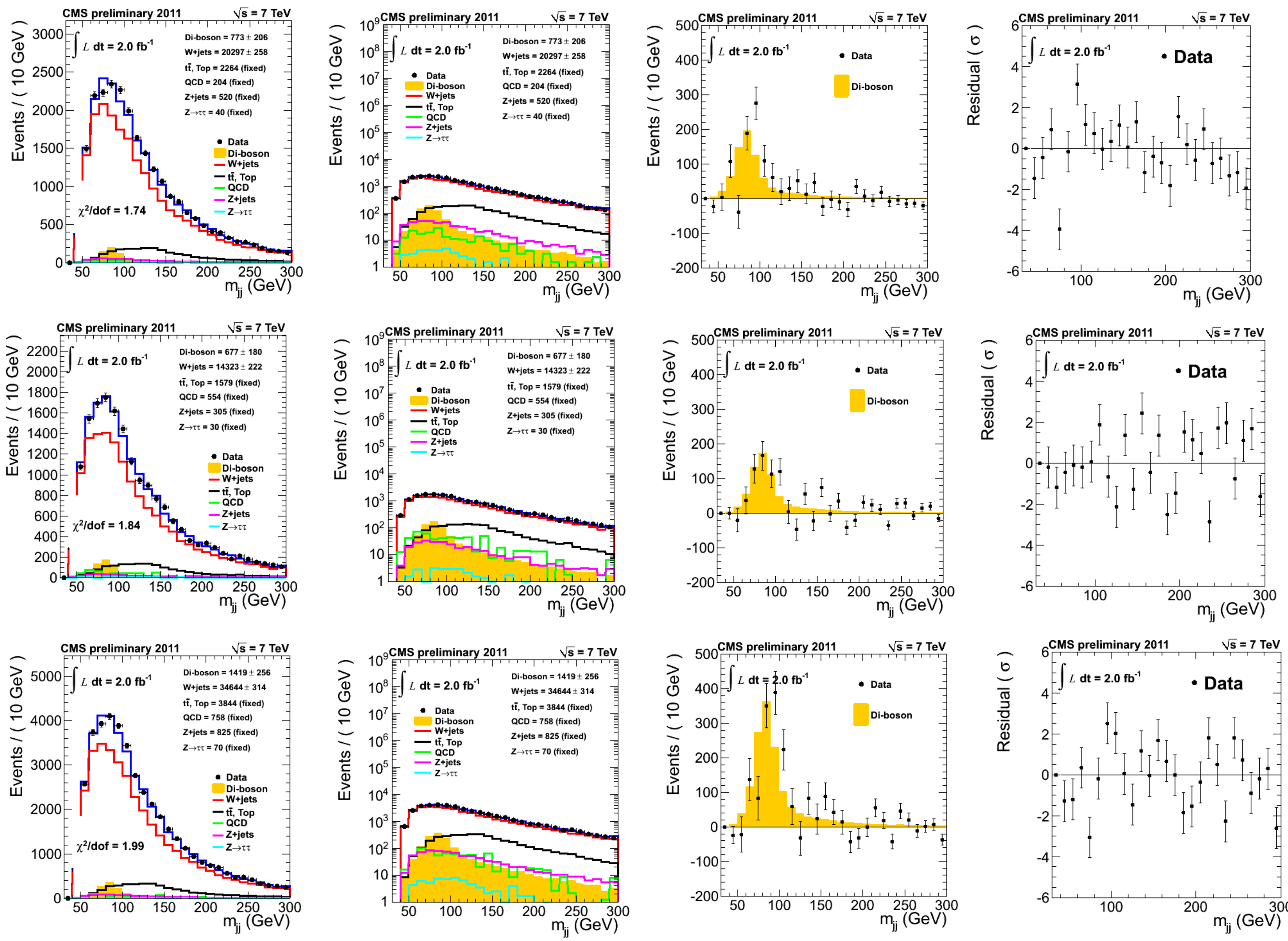
$\frac{\text{Data} - \text{Fit}}{\text{Error}}$



- ✦Systematics are still being finalized. Also, will have some improvement in W+jets template
- ✦Diboson peak established, no anomalous excess

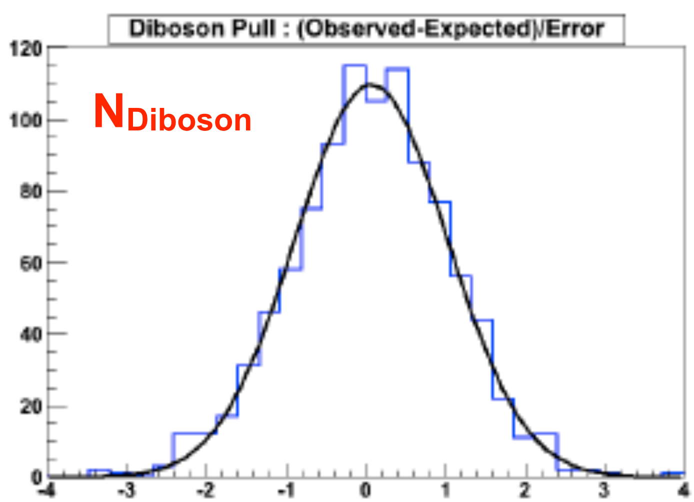
Separately for each lepton channel

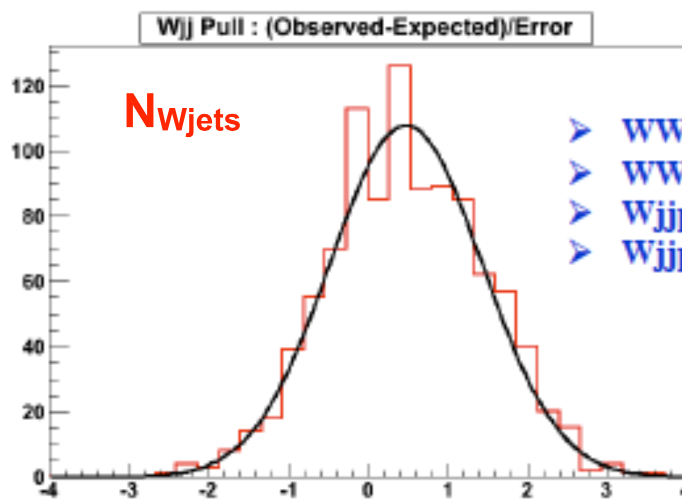
- **muons**
  Diboson=773±206
  Wjj=20297±258
- **electrons**
  Diboson=677±180
  Wjj=14323±222
- **total**
  Diboson=1419±256
  Wjj=34644±314



Validation of fit results

To make sure that the template fit is unbiased and to check the coverage of statistical uncertainty reported by the fit, we generate 1000 pseudo experiments (PE) using the shape that best describes the data. Then we fit each of these PE samples using our nominal shape and plot pull distribution for each parameter.





- WWpull_mean=0.07+/-0.03
- WWpull_sigma=0.96+/-0.02
- Wjjpull_mean=0.48+/-0.03
- Wjjpull_sigma=0.96+/-0.02

• Fitter returns consistent results: negligible bias
• Statistical uncertainties reported by the fit are correct

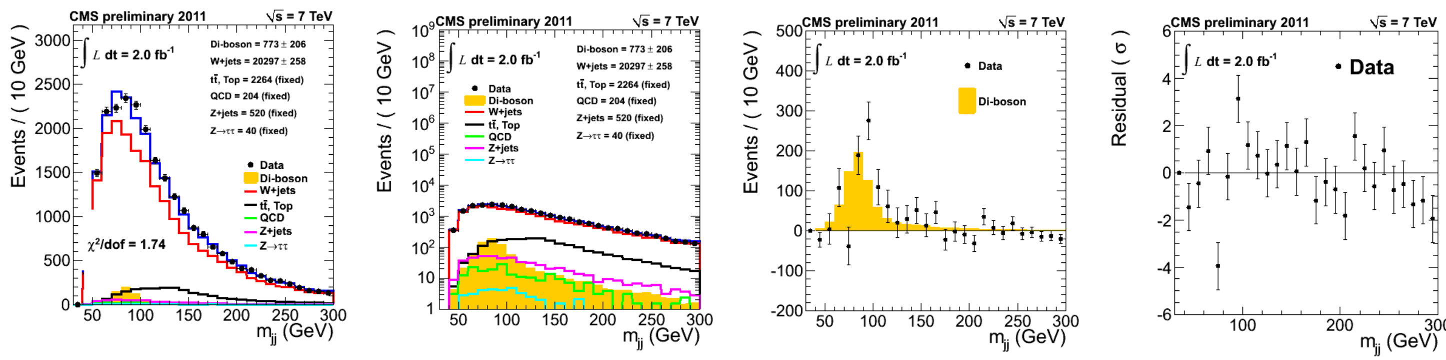
More validation: Pileup dependence

➢ **Examine the nPV≤5 and nPV>5 events separately**

❖ **nPV≤5**

Diboson=773±206

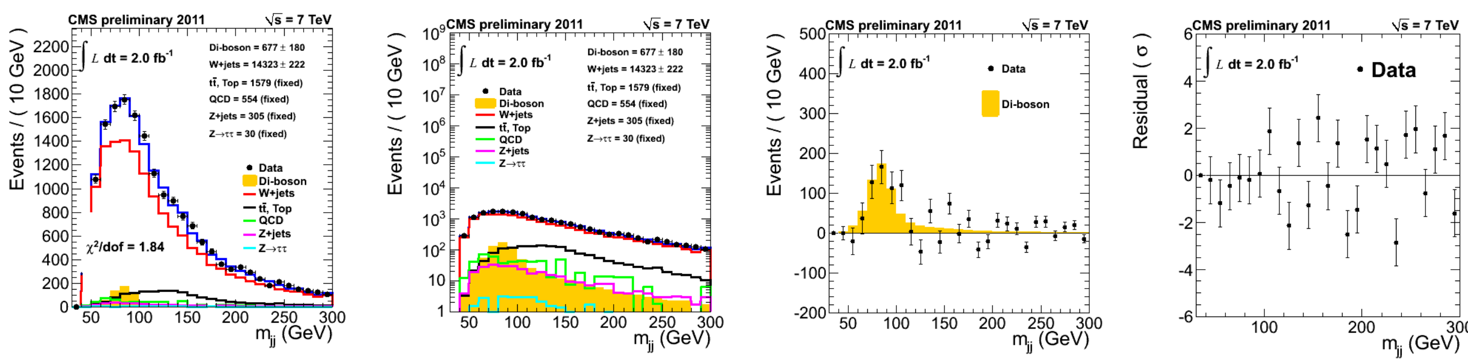
Wjj=20297±258



❖ **nPV>5**

Diboson=677±180

Wjj=14323±222



✓ **Results are consistent with the default fit**

Does the di-boson yield match expectation ?

Fitted Diboson Yield =1419±256(stat.)

Expected Yield = σ*Lumi*ε*A : Diboson=777±78(theory, NLO)

❖ **Compare to the (scaled) Diboson Yield in the 2*l*2ν channel = 1000**

The systematic uncertainty in diboson is large (see next slides) − driven by the systematic uncertainty in W+jets shape. The difference between the observed diboson yield and the SM expectation is covered by systematic uncertainty. Also, note that this analysis is not optimized to enhance diboson signal precision.

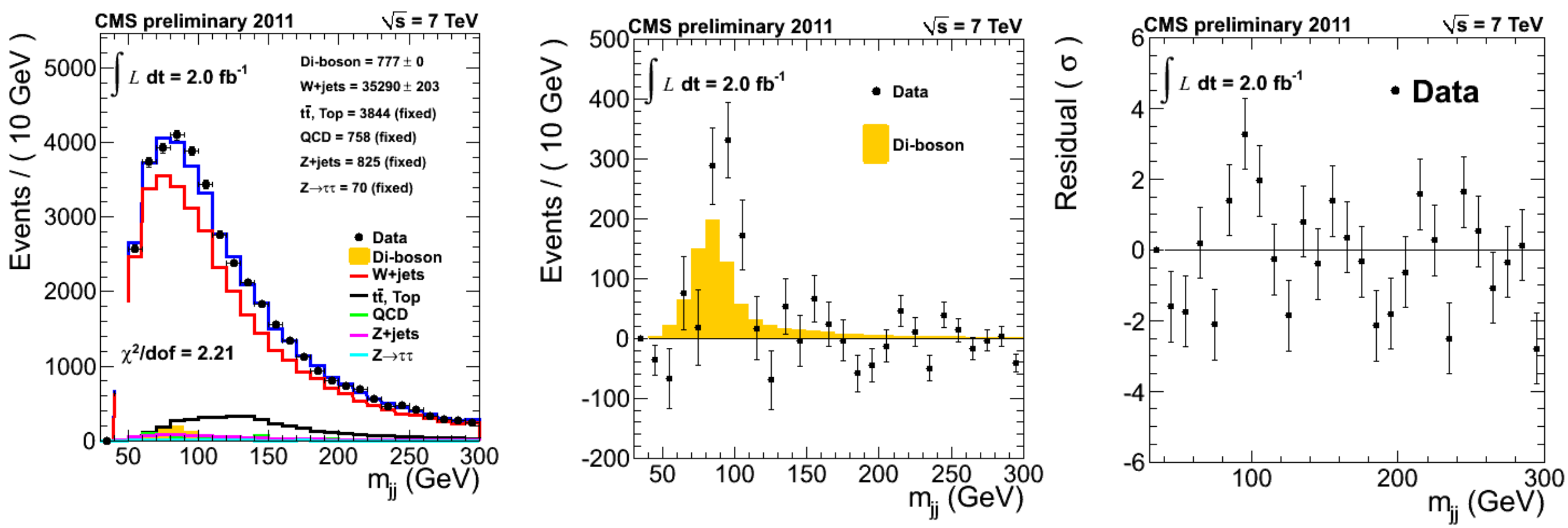
❖ **Several other consistency checks have been performed**

- **Fit with the Diboson Yield fixed to its theory value (777)**
- **Gradually remove the Diboson Signal by taking j1p_T,j2p_T>40,50,60 GeV**
- **Enhance the Diboson Yield by placing a χ^2 cut from the Kinematic Fit**
- **Use the template from the Zjj data, rather than Wjj MC**

See our presentation in the Electroweak meeting on September 20 for details:

https://indico.cern.ch/getFile.py/access?contribId=1&resId=0&materialId=slides&confId=147673

What if we fix the di-boson yield to NLO value?

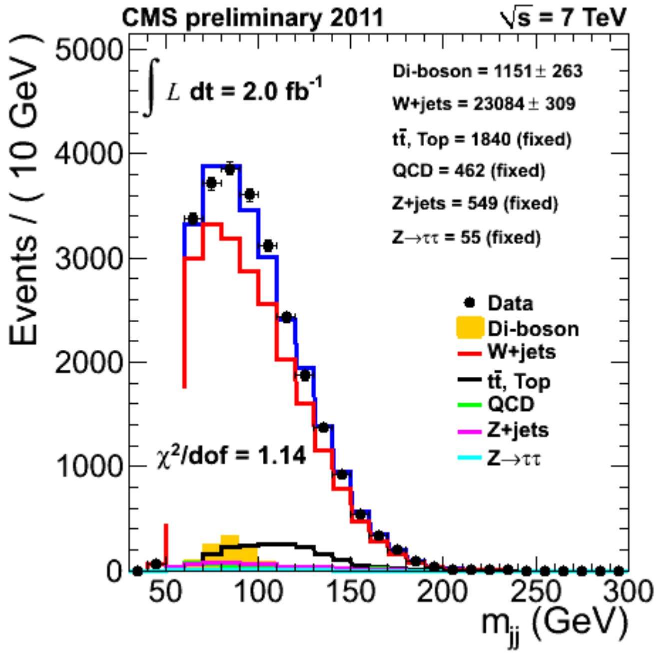


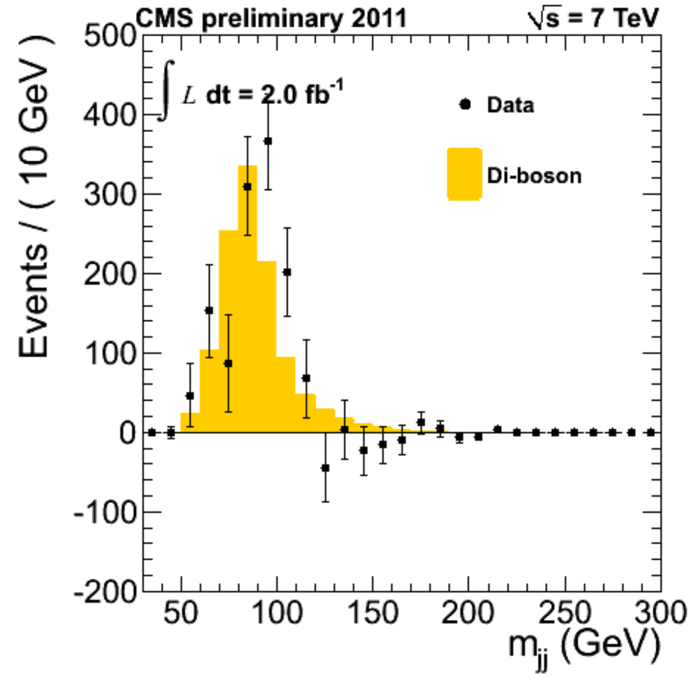
- ❖ **Wjj Yield: 35290 ± 203**
- ✓ **Some excess signal near M_{jj}=80 GeV**

What if we go to "diboson enhanced" selection

Additional cut:

- Kinematic Fit χ2/ NDF<8.5 (not used in generic Mjj analysis)

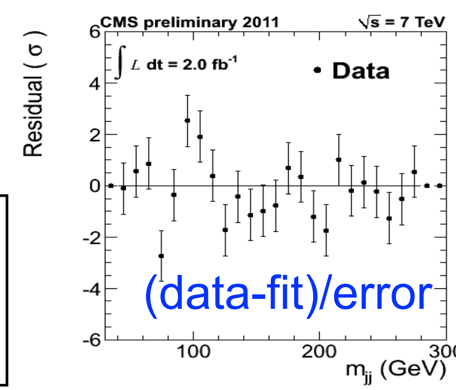




❖ χ^2/Ndof<8.5 nDiboson=1151 ± 263

scale by efficiency of the above cut →

❖ nDiboson fixed: 1376 Evts

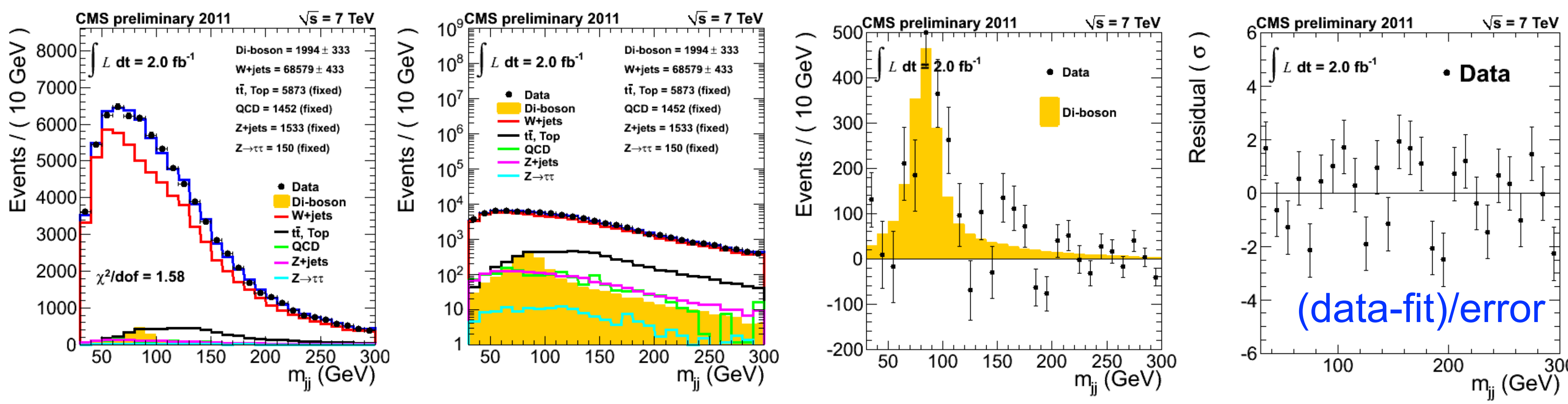


- The fit quality improves with additional cuts
- Diboson peak becomes more well-defined and pronounced
- The yield is consistent with the default fit

What happens if we go to "CDF-like" selection

Cut	CDF	CMS
$\|\Delta\eta_{jj}\| <$	2.5	1.5
$W_{mT} >$	30	40
$MET >$	25	30
$\mu_{pT} >$	20	25
$e_{pT} >$	25	30

- Suppress the Wjj background
- Trigger constraints (ele)



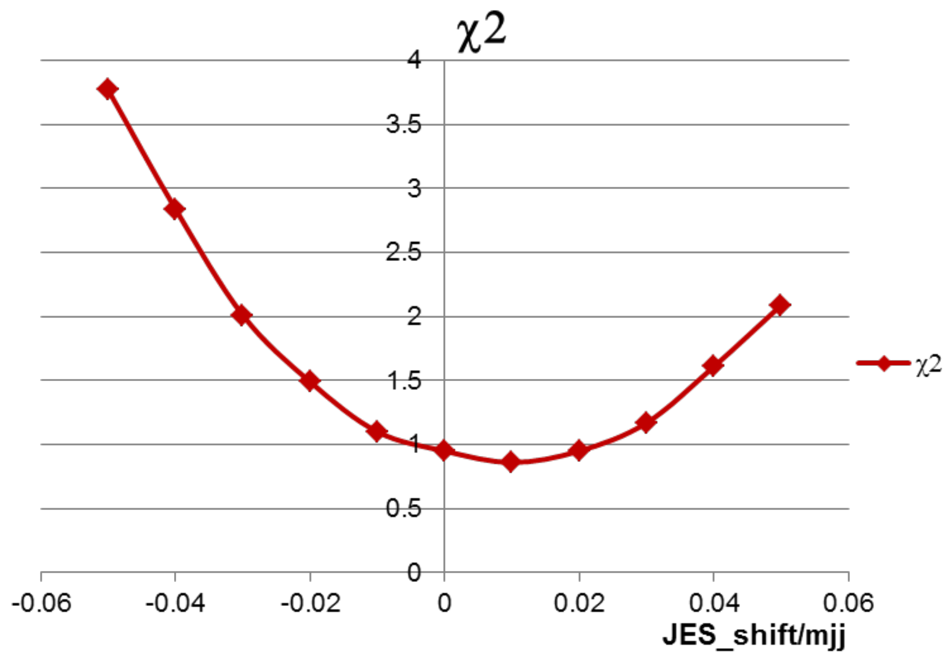
With "CDF-like" loose cut it becomes hard to model the data well, and to have all systematics understood. Although, essential features remain the same.

Yields:
Diboson=1994±333
Wjj=68579±433

Scan of jet energy scale

- The default fit allows the JES to float and returns a value of 0.003 m_{JJ}
- We perform a manual scan by fixing JES and repeating the fit

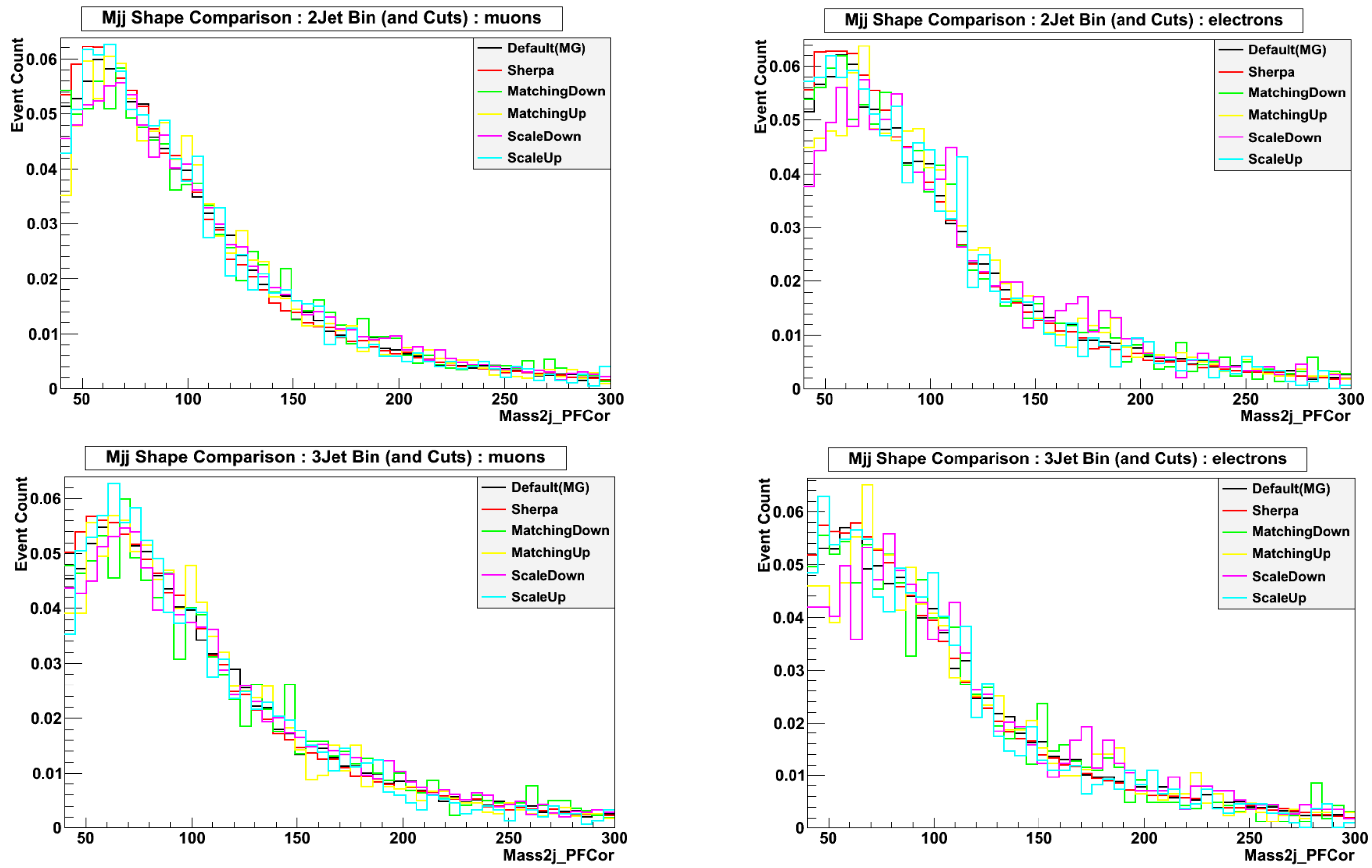
JES_shift/mjj	WW	χ^2
-0.05	1012	3.77
-0.04	960	2.84
-0.03	901	2.01
-0.02	874	1.49
-0.01	877	1.1
0	793	0.95
0.01	676	0.86
0.02	594	0.95
0.03	496	1.17
0.04	354	1.61
0.05	261	2.09



The fit is stable and has a χ^2 minimum near 0.

Shape systematics: W+jets shape uncertainty

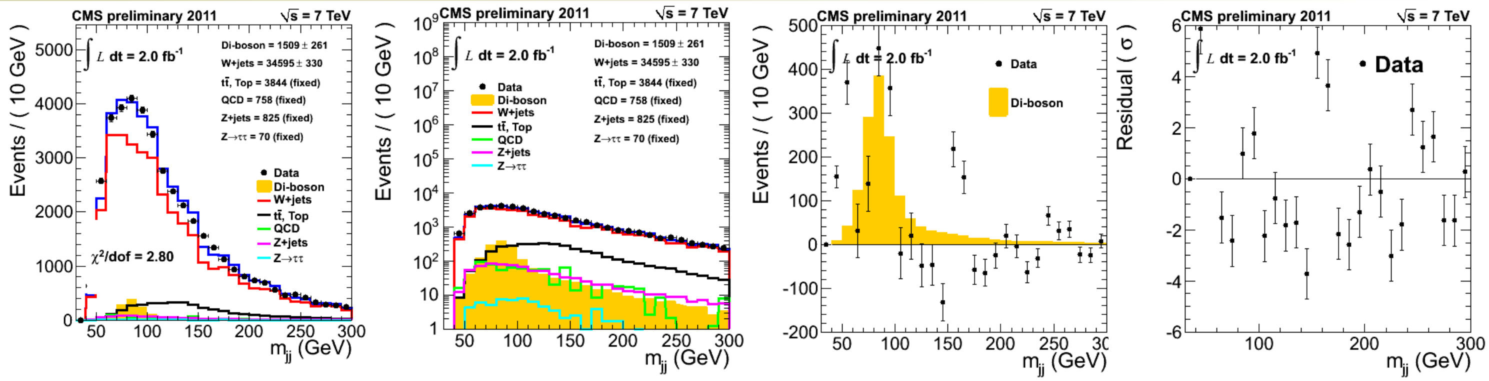
Run over the Matching and Factorization Scale Up/Down MC and compare



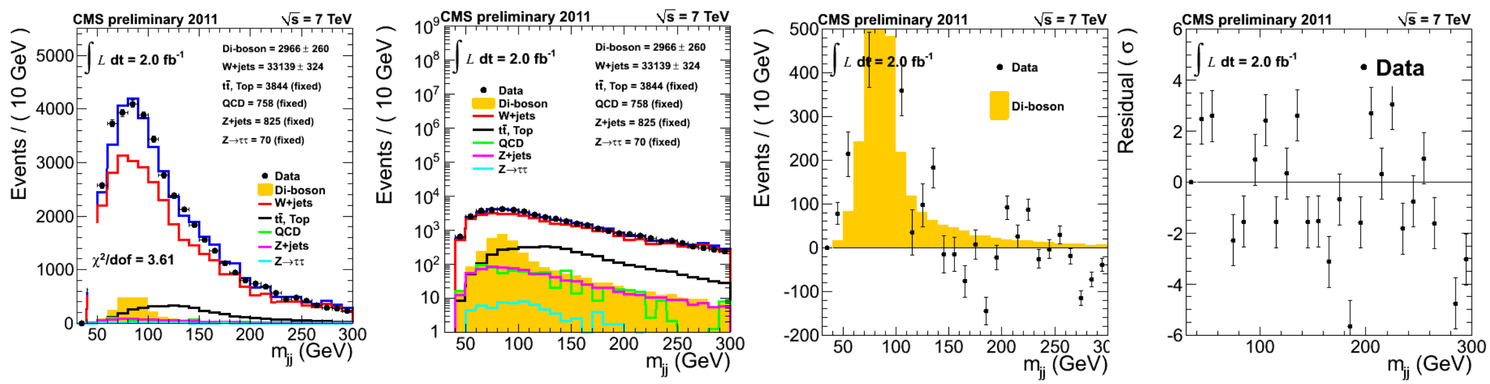
- ❖ Similar Shapes with fluctuations due small size of Monte Carlo samples

Variation of Q^2 scale and ME–PS matching

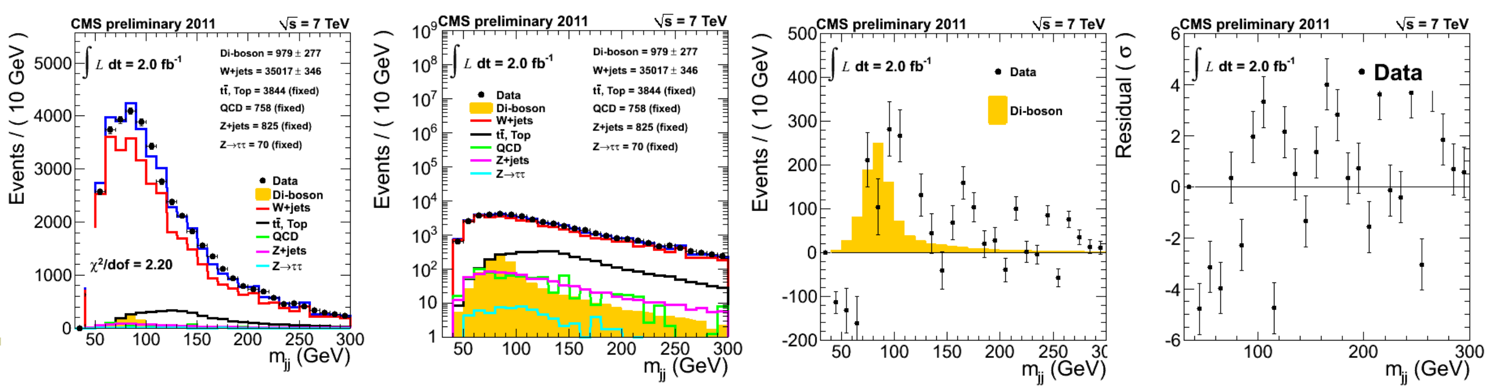
❖ **ME-PS matching @20GeV**
nDiboson=1509±261
nWjj=34595±330



❖ **ME-PS matching @5GeV**
nDiboson=2966±260
nWjj=33139±324



❖ **Q^2 Up**
Diboson=979±277
Wjj=35017±346



Combining Systematic Errors

❖ Three Categories

- Matching: Repeat the fit using Up and Down Samples and take the larger of the two variations to obtain the (symmetric) error.
- Factorization Scale: Likewise, repeat the fit using Up and Down Samples as the Wjj template and take the larger of the two variations to obtain the error.
- Yield Dependent: Remaining uncertainties proportional to the Diboson Yield, added in quadrature.
- ~~Wjj Template Systematics~~: Included (and overcovered) by the above categories.

❖ Assume zero correlation

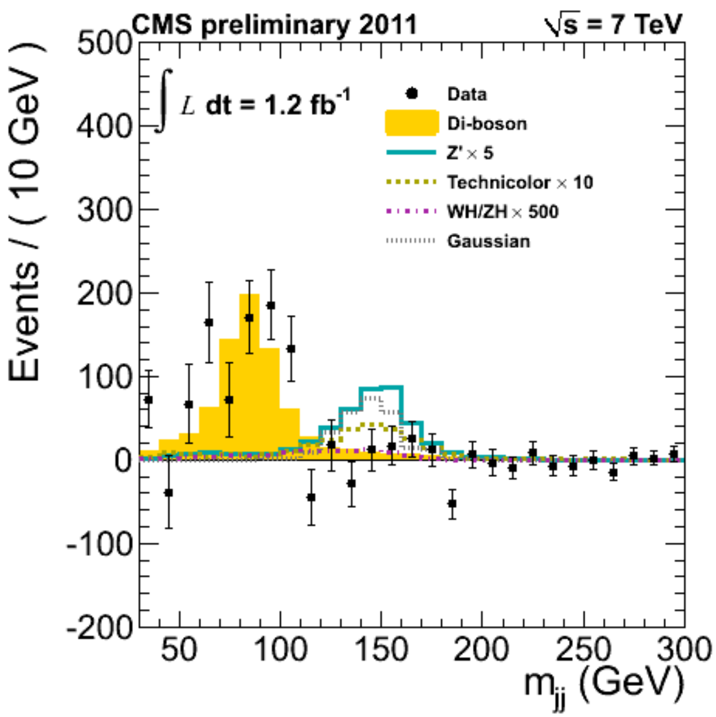
❖ Treat the differences in fit results as coming purely from systematic fluctuations

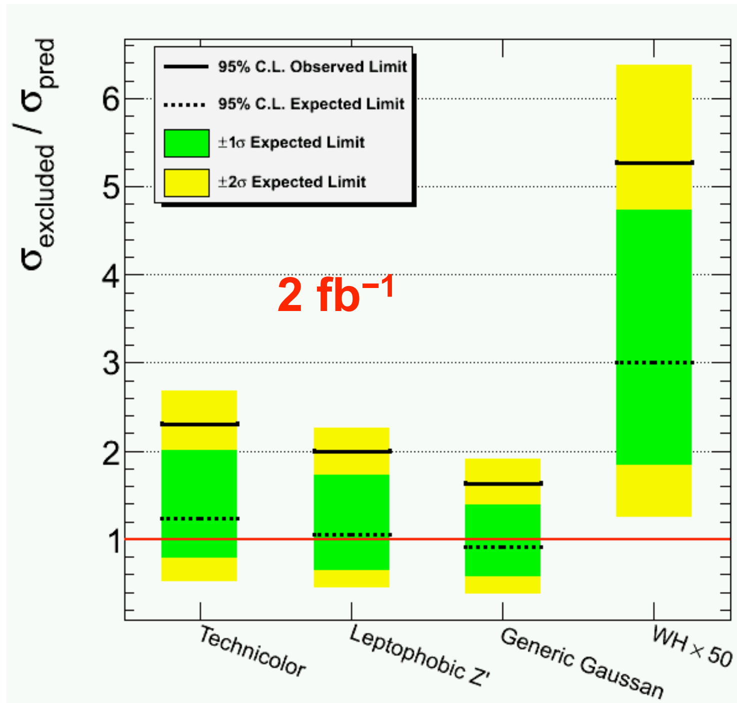
❖ $\sigma^2_{Tot} = \sigma^2_{Matching} + \sigma^2_{Scale} + \sigma^2_{Remaining}$

❖ Compute the uncertainty bin by bin as well as overall

Examine several New Physics Models, set limits

Fit/Process	(WW+WZ)/Wjj	Z'/Wjj	Technicolor/Wjj	H(150GeV)/Wjj
CDF-Like - μ	$1.8x10^{-2}$	$4.6x10^{-3}$	$4.0x10^{-3}$	$6.1x10^{-5}$
Improved - μ	$2.1x10^{-2}$	$5.7x10^{-3}$	$4.8x10^{-3}$	$7.5x10^{-5}$
CDF-Like - el	$1.7x10^{-2}$	$4.6x10^{-3}$	$3.6x10^{-3}$	$5.8x10^{-5}$
Improved - el	$2.1x10^{-2}$	$6.0x10^{-3}$	$5.0x10^{-3}$	$7.8x10^{-5}$





Set limits on Technicolor, Leptophobic Z', Higgs and Generic Gaussian (mass = 150 GeV) models

Summary

- Performed study of dijet mass spectrum in lvjj events @ 2 fb^{-1} data
  - start with basic CDF-like loose selection
  - try various improvements − including those suggested by ELM − to suppress W+jets and to make qq processes stick out
  - settle on selection criteria more appropriate for LHC conditions
- Validated and cross-checked the results
  - extracted the di-boson and Wjj yields: 1419±256, 34644±314 evts
  - examined discrepancy btw predicted & observed di-boson yields
  - analyzed the JES and dominant systematics
- No anomalous peak observed so far
  - placed limit on “CDF bump” effect
  - finalizing the systematics & analysis of di-boson yield discrepancy
  - aiming for publication this year: AN-2011-266, PAS: EWK-11-017

BACKUP SLIDES

What if we vary one cut at a time

Gradually change the cuts from CDF-like to the ones optimized for extracting the diboson yield

- **Stage 0: CDF-like cuts**
- **Stage 1: Kinematic Fit χ^2/NDF<10.0 (not used in generic Mjj analysis)**
- **Stage 2: -0.6<cosJacksonAngle<0.8 (not used in generic Mjj analysis)**
- **Stage 3: Jet2p_T/m_{jj}>0.3**

Since in step 0 we are completely swamped by background, it is hard to say if we do not model the W+jets right or it is statistical fluctuation in the number of W+jets. As the S/B improves, so does our ability to observe qq processes (diboson or otherwise).

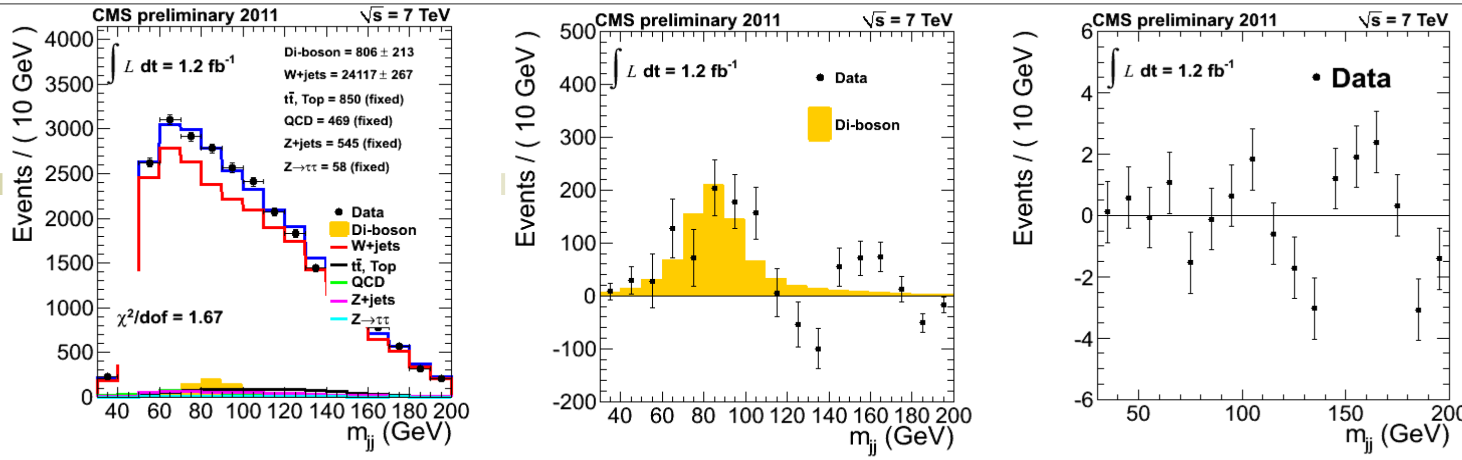
Comparison on the next slide →

Stage 1:

N_{WW}=806±213

N_{Wjj}=24117±267

S/B=0.033

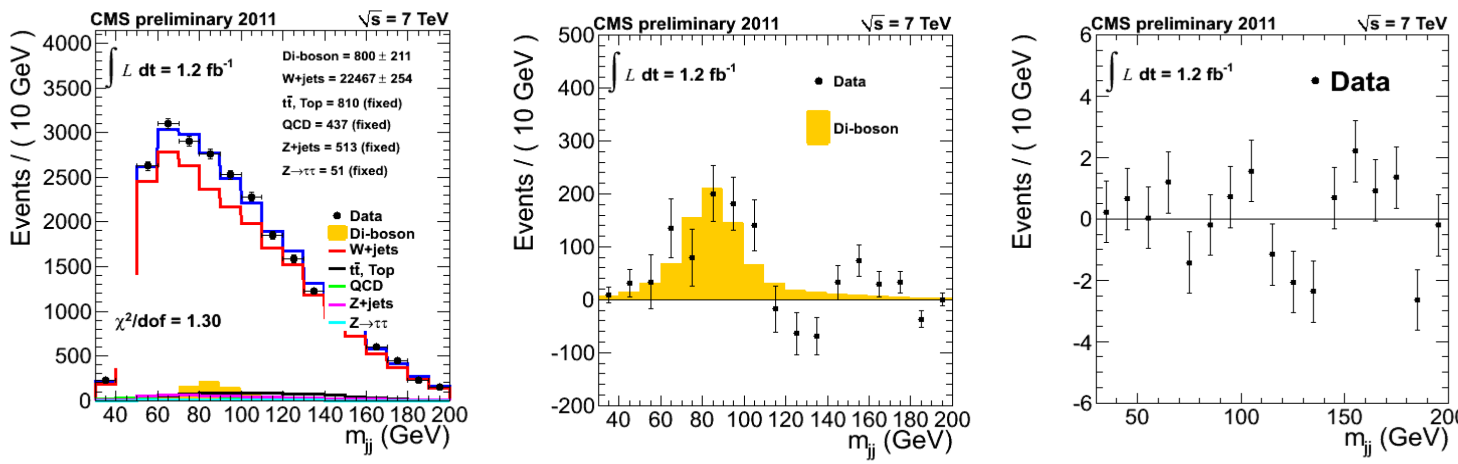


Stage 2:

N_{WW}=800±211

N_{Wjj}=22468±255

S/B=0.036

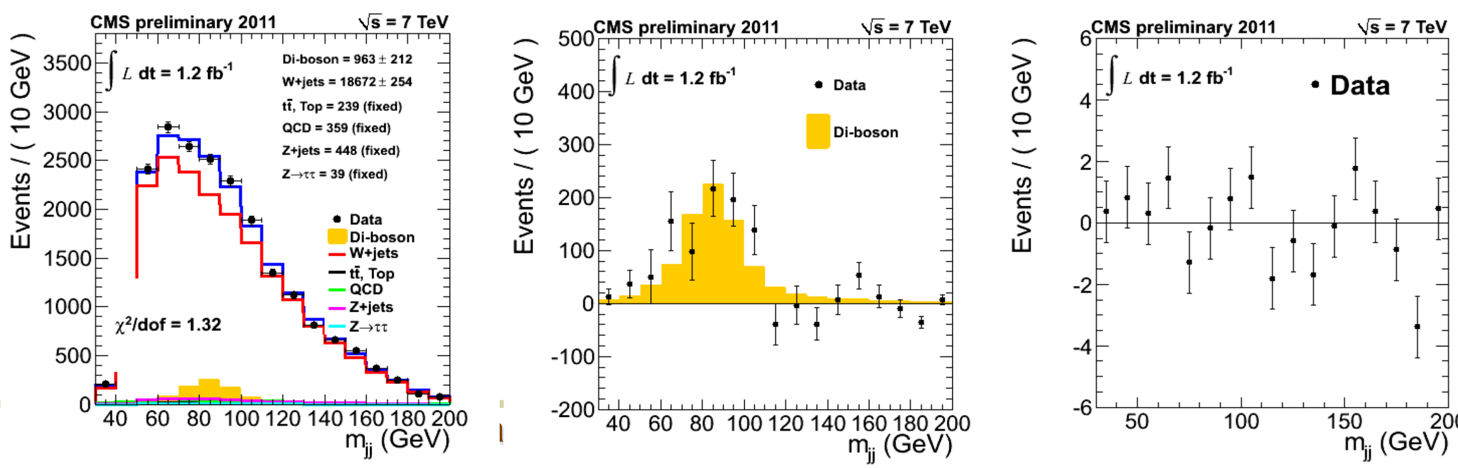


Stage 3:

N_{WW}=847±228

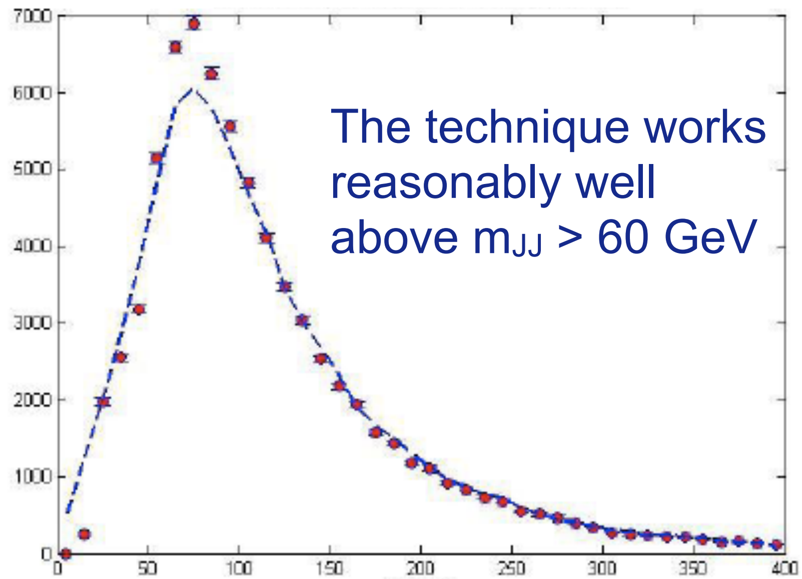
N_{Wjj}=20557±272

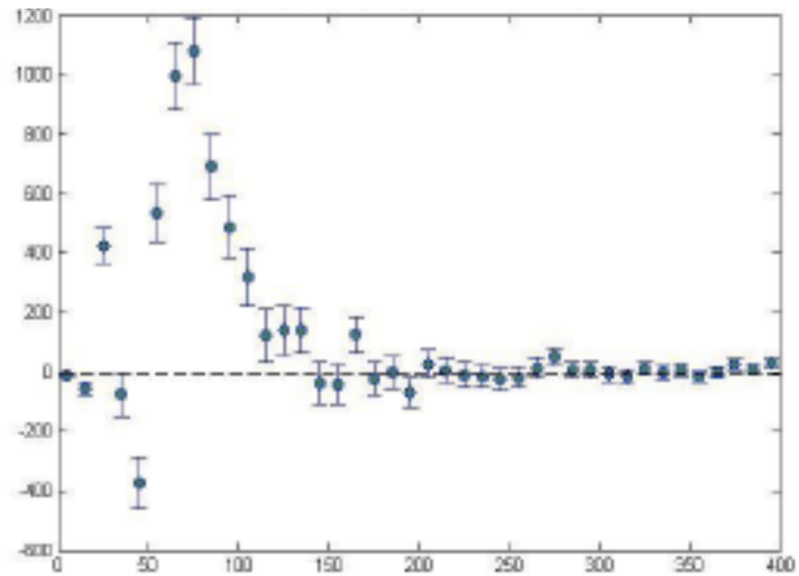
S/B=0.041



Have we tried data-driven shape for W+jets

Yes: For cross check we take shape using “mixed events”. We make random combination of jets by taking one jet from some other event. This gives large ststistics: N*(N-1). This cannot produce a bump. The only challenge is to manage kinematic correlations.





https://indico.cern.ch/materialDisplay.py?contribId=33&materialId=slides&confId=147015

Work ongoing to improve the technique

How confident we are about JES ?

Reasonably confident:

m_{JJ} in top data

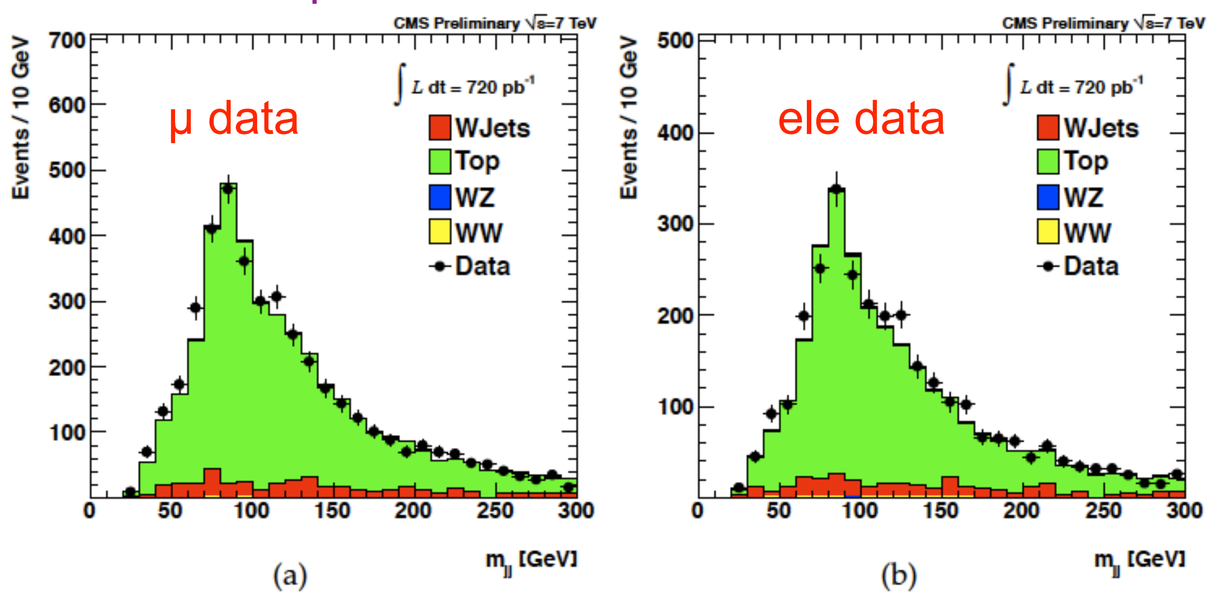
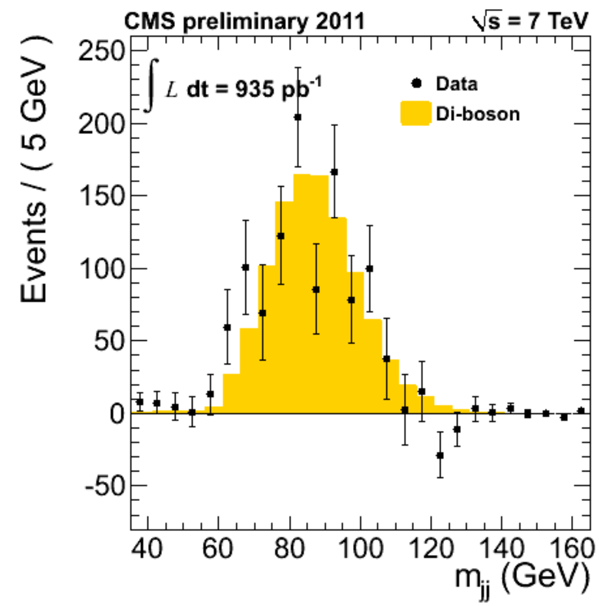


Figure 14: The invariant mass distribution of the two non-b-tagged jets in events containing a high energy lepton, large $\not{E}_T$, and four jets of which two are b-tagged. This sample is dominated by $t\bar{t}$ events. The distribution shown here for both lepton channels (a: muon, b: electron) essentially establishes how well we can reconstruct a hadronically decaying vector boson in CMS detector. The resolution of the reconstructed vector boson is dominated by the jet energy resolution.

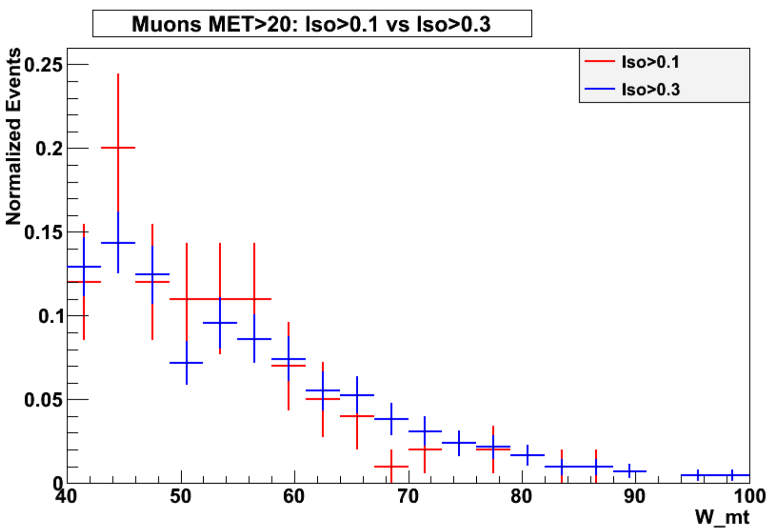


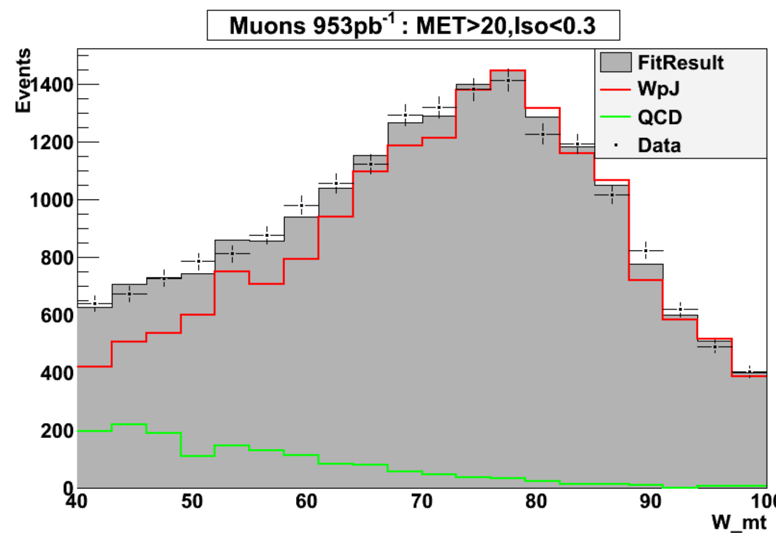
m_{JJ} is diboson data after tight selection

All indications suggest that JES in data and MC agree very well.
Jet resolution is not much worse in data.

QCD: shape and normalization from data

- Fit W_{mT} distributions for QCD and Wjj
- To obtain a sufficient amount of QCD events relax to:
  - MET>20.0
  - Iso<0.3
  - WP70 for e^-





- **The shape remains statistically consistent.**
- **Accounting for acceptances, fix QCD fraction relative to Wjj (0.008 μ, 0.04 e^-).**