Log plots, Fakes and DY

Richard St. Denis University of Glasgow/CDF

19 November 2007

Outline

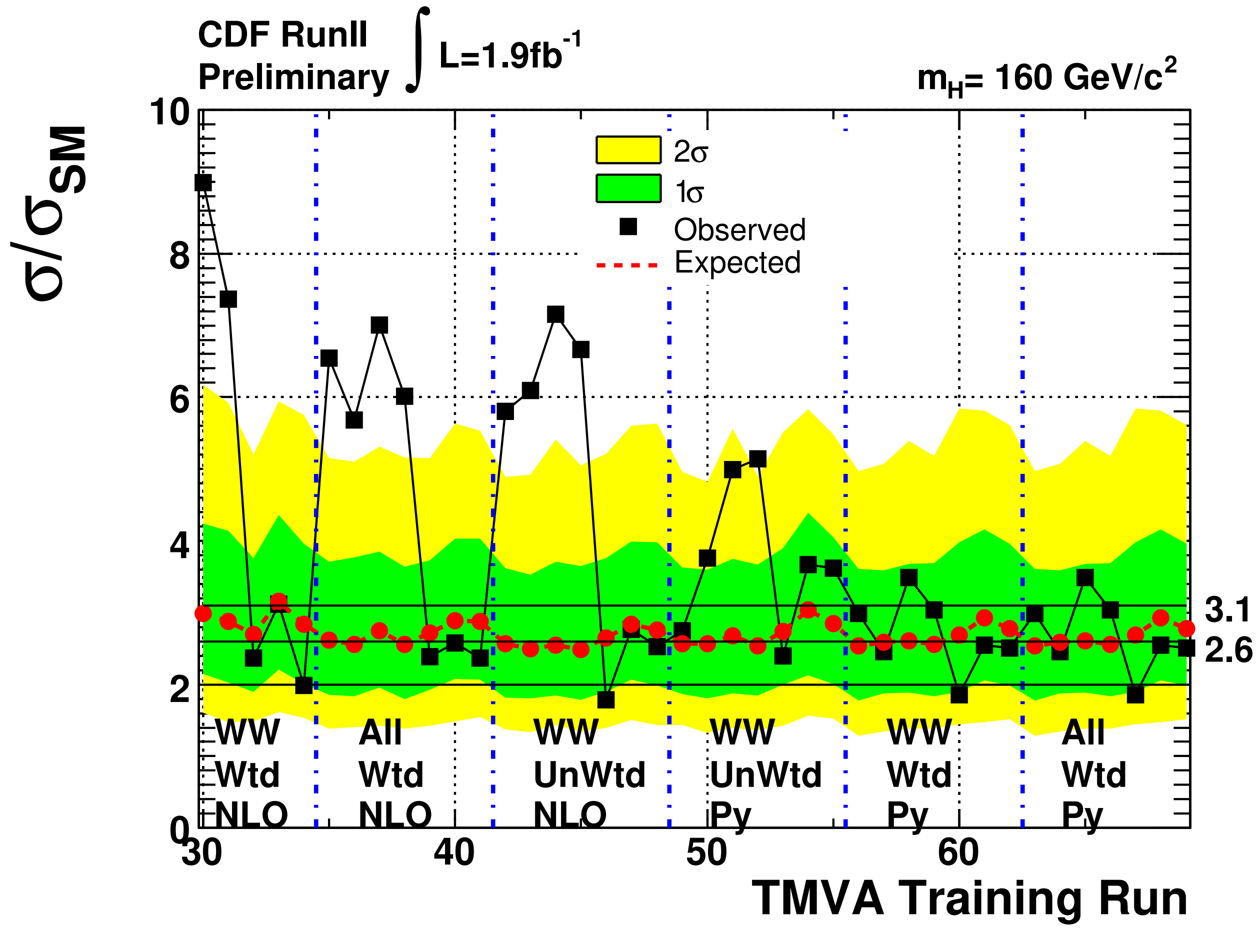
- Summary of Limits and Variations
- Request to see LRHWW on log
- Examination of Fake/DY in MetSig, Mll and NN distn.

Trainings: Conditions

In each group

- Remove Redundant SumJetEt (2 layer)
- Remove Redundant SumJetEt (1 layer)
- Remove SumEtJet and LRHWW (2 layer)
- Remove SumEtJet and all LR (2 layer)
- Remove SumEtJet and MetSig (2 layer)
- Remove SumEtJet, MetSig and LRWW (2 layer)
- Remove SumEtJet, MetSig and All LR (2 layer)

Trainings

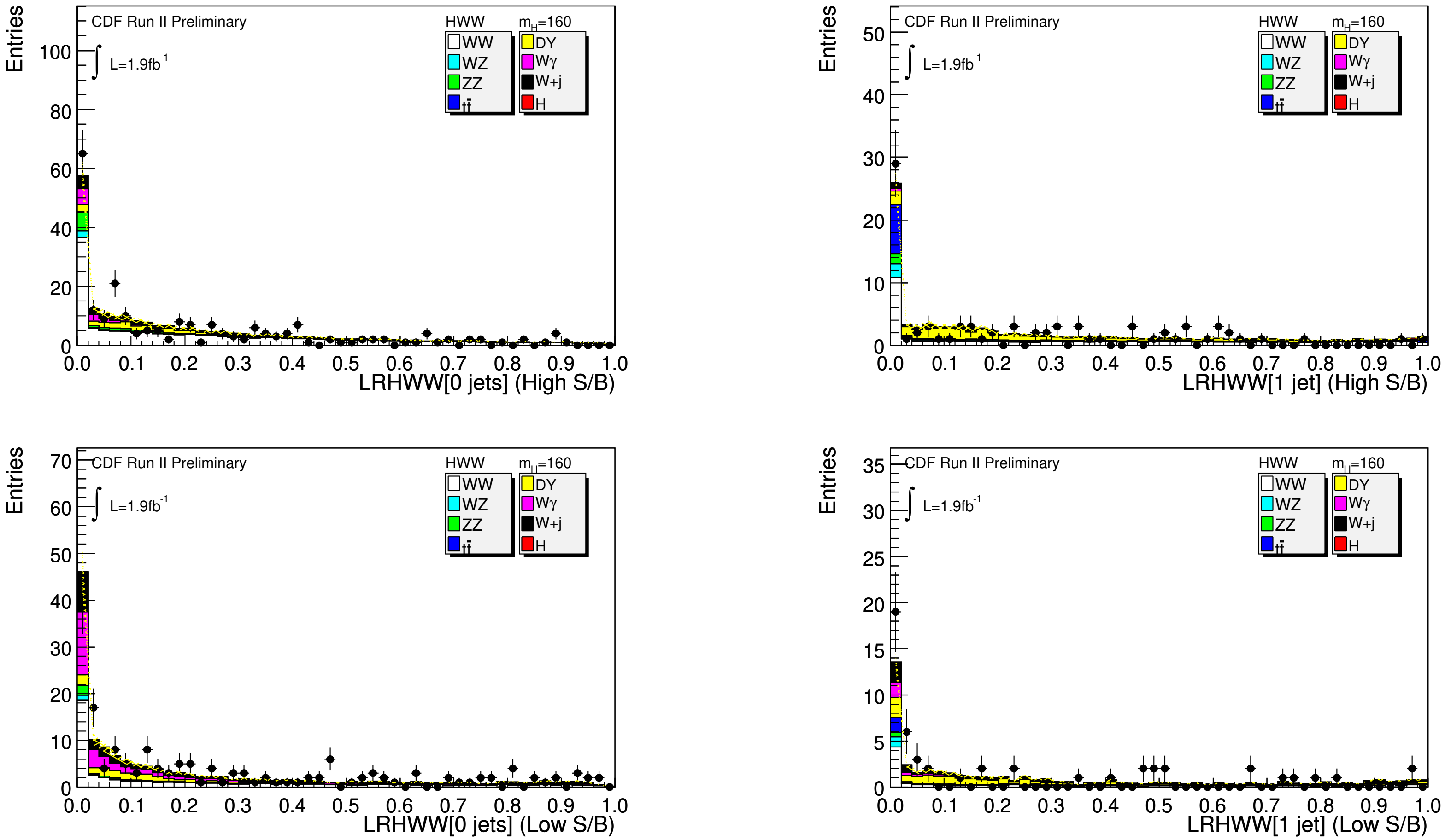

CDF RunII Preliminary
∫L=1.9fb-1
mH= 160 GeV/c2
σ/σSM
2σ
1σ
Observed
Expected
WW Wtd NLO
All Wtd NLO
WW UnWtd NLO
WW UnWtd Py
WW Wtd Py
All Wtd Py
3.1
2.6
TMVA Training Run

Trainings: Observations

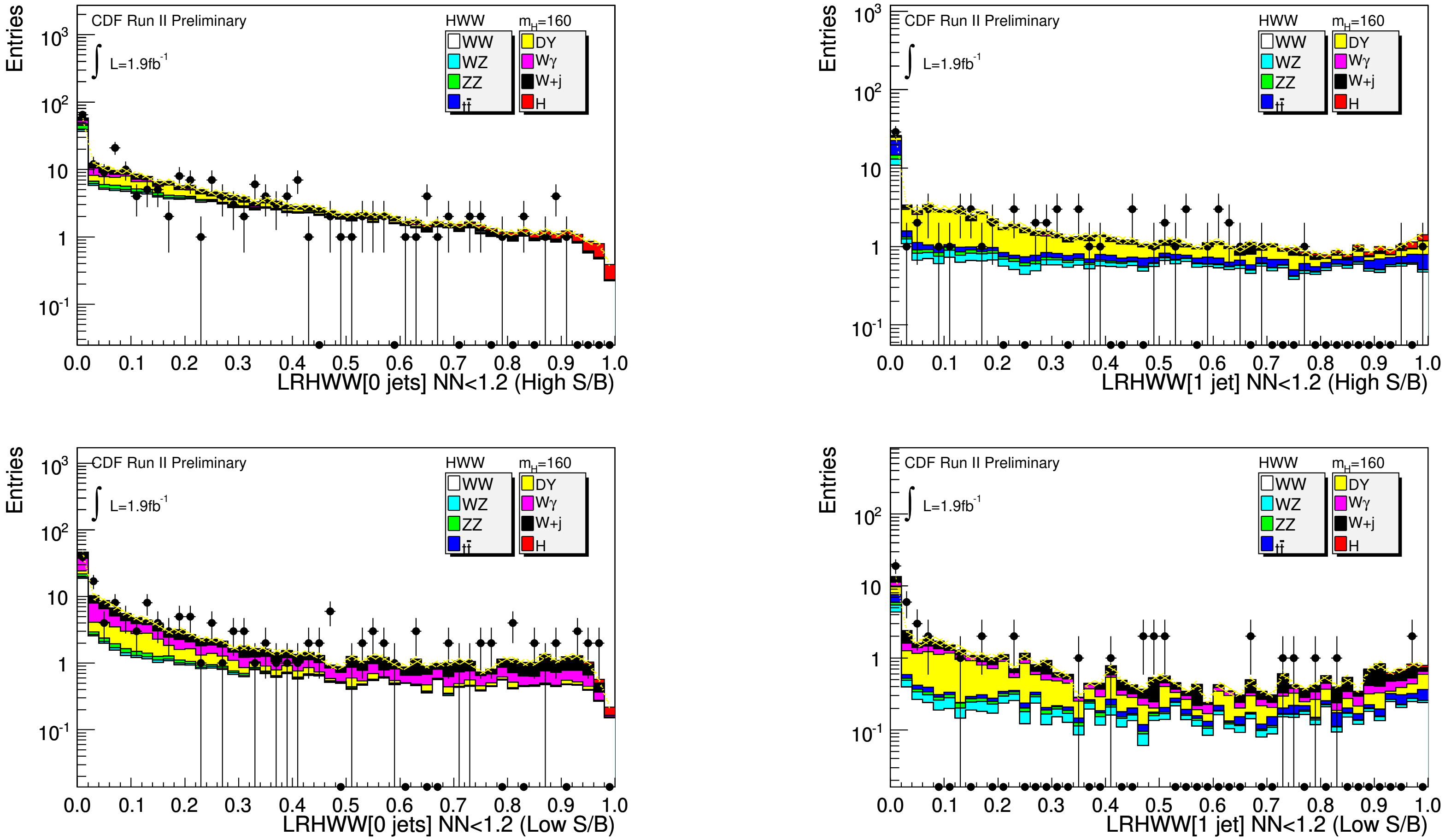
- All *weighted* trainings show deviation in Background training and test NN distribution.
- MetSig has a huge impact on observed with little impact on expected
- Exeception is Pythia: WW and unweighted, all variables (SumJetEt removed) shows same result as when MetSig is removed, although removing LR(HWW) creates MetSig sensitivity
- Pythia shows MetSig Sensitivity to some degree but with varied interaction in LR(HWW).

A higher order modelling problem does seem to be an ingredient. Turn to LRHWW by jet.

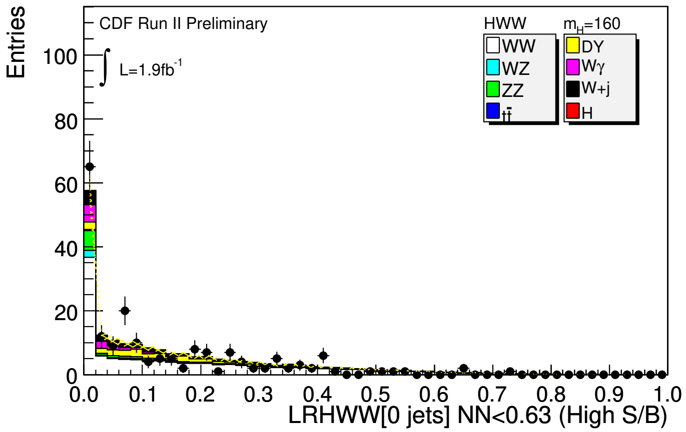
LRHWW:No NN Cut

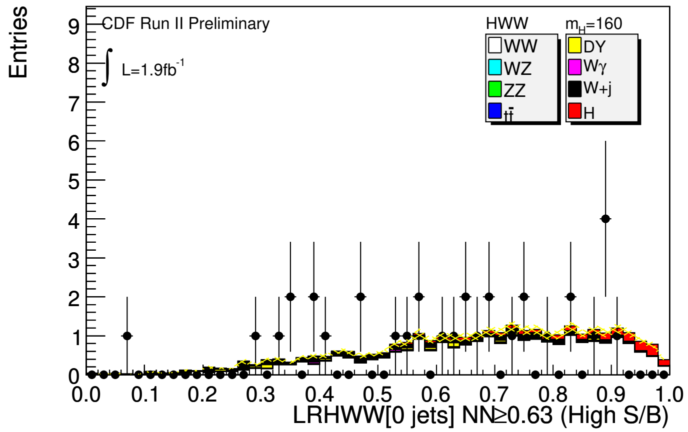


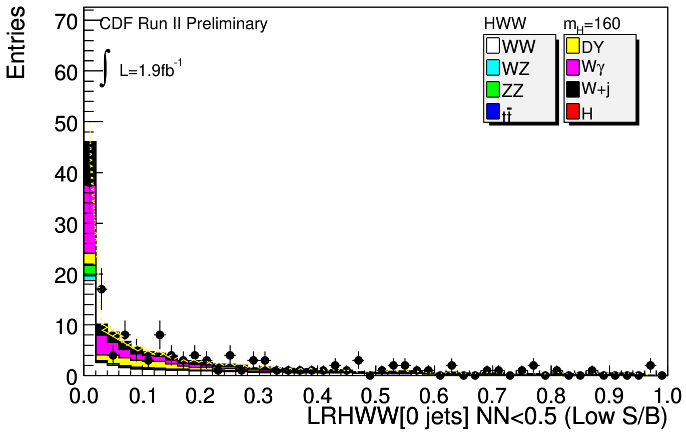
LRHWW: No NN cut

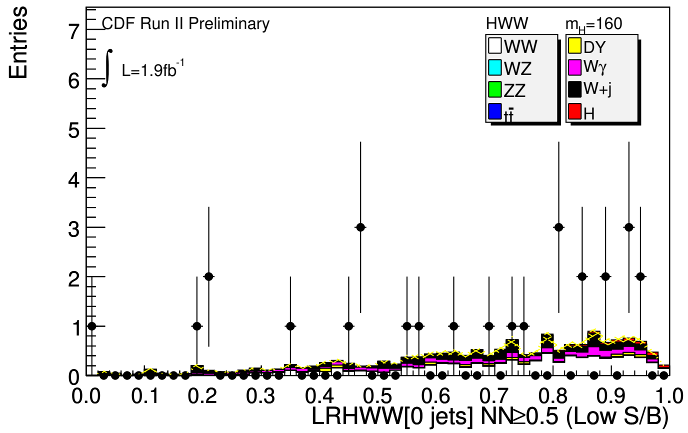


LRHWW: 0 Jets

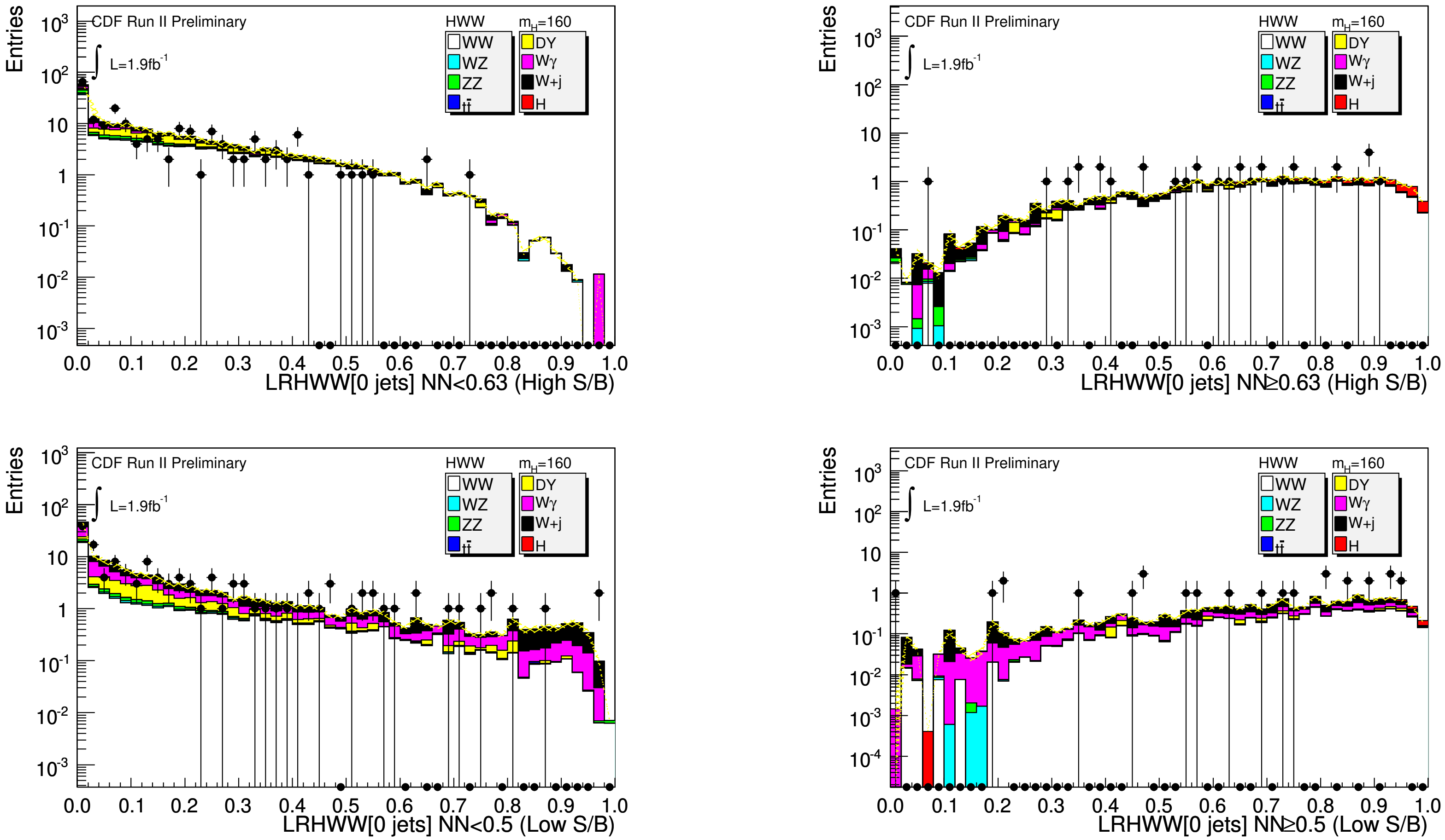

CDF Run II Preliminary
$\int L = 1.9 fb^{-1}$
HWW m_H=160
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
Entries
LRHWW[0 jets] NN<0.63 (High S/B)


CDF Run II Preliminary
$\int L = 1.9 fb^{-1}$
HWW m_H=160
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
Entries
LRHWW[0 jets] NN≥0.63 (High S/B)

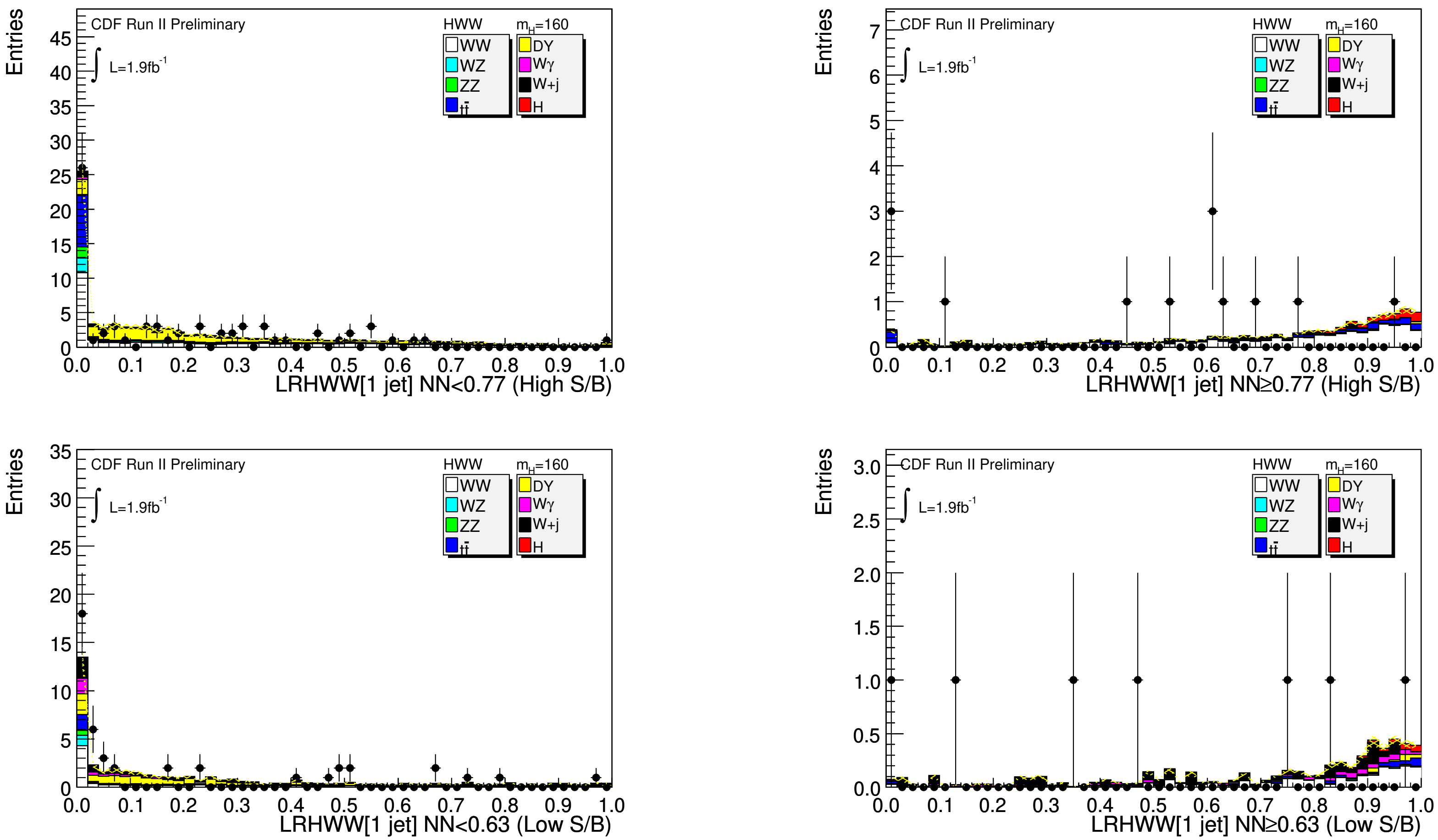

CDF Run II Preliminary
$\int L = 1.9 fb^{-1}$
HWW m_H=160
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
Entries
LRHWW[0 jets] NN<0.5 (Low S/B)


CDF Run II Preliminary
$\int L = 1.9 fb^{-1}$
HWW m_H=160
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
Entries
LRHWW[0 jets] NN≥0.5 (Low S/B)

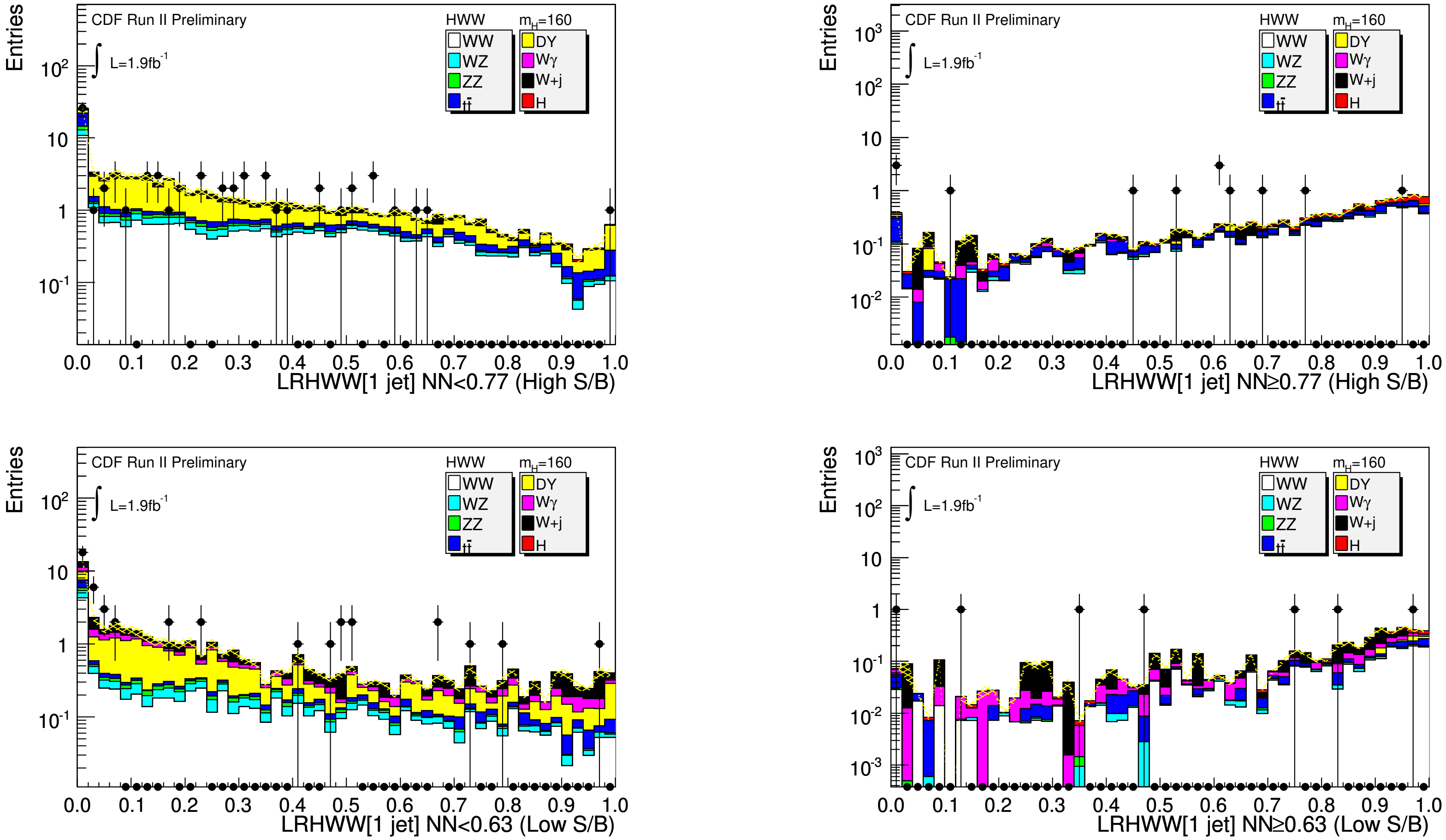
LRHWW: 0 jets


CDF Run II Preliminary
L=1.9fb-1
Entries
HWW
mH=160
WW
WZ
ZZ
tt̄
DY
Wγ
W+j
H
LRHWW[0 jets] NN<0.63 (High S/B)
LRHWW[0 jets] NN≥0.63 (High S/B)
LRHWW[0 jets] NN<0.5 (Low S/B)
LRHWW[0 jets] NN≥0.5 (Low S/B)

LRHWW:1 Jet


CDF Run II Preliminary
L=1.9fb-1
HWW
mH=160
WW
WZ
ZZ
tt̄
DY
Wγ
W+j
H
Entries
LRHWW[1 jet] NN<0.77 (High S/B)
LRHWW[1 jet] NN≥0.77 (High S/B)
LRHWW[1 jet] NN<0.63 (Low S/B)
LRHWW[1 jet] NN≥0.63 (Low S/B)

LRHWW: 1 jet


CDF Run II Preliminary
$\int L = 1.9 fb^{-1}$
HWW $m_H = 160$
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
Entries
LRHWW[1 jet] NN<0.77 (High S/B)
LRHWW[1 jet] NN≥0.77 (High S/B)
LRHWW[1 jet] NN<0.63 (Low S/B)
LRHWW[1 jet] NN≥0.63 (Low S/B)

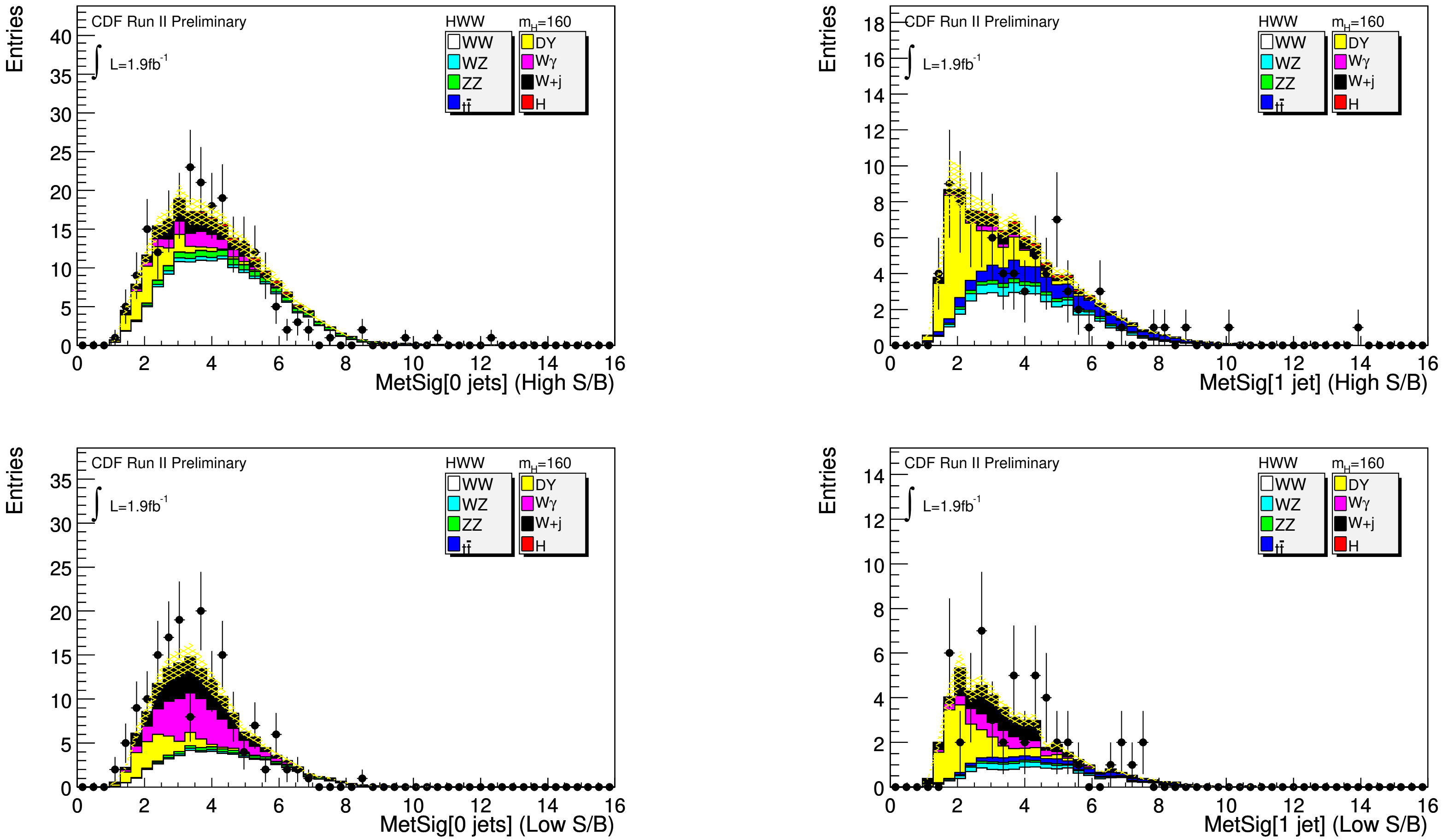
	High S/B 0 Jet			
Discriminant	KS Prob	χ^2/DOF	χ^2 Prob	Comment
LRHWW	0.05	62/50	0.11	
NN<0.63	0.02	49/48	0.43	
NN>0.63	0.41	56/50	0.26	
	Low S/B 0 Jet			
Discriminant	KS Prob	χ^2/DOF	χ^2 Prob	Comment
LRHWW	0.04	65/50	0.08	χ^2 grow after 0.5
NN<0.5	0.18	57/50	0.23	
NN>0.5	0.39	73/50	0.02	
	High S/B 1 Jet			
Discriminant	KS Prob	χ^2/DOF	χ^2 Prob	Comment
LRHWW	0.12	44/50	0.73	
NN<0.77	0.43	39/50	0.88	
NN>0.77	0.01	56/50	0.26	
	Low S/B 1 Jet			
Discriminant	KS Prob	χ^2/DOF	χ^2 Prob	Comment
LRHWW	0.04	59/50	0.18	
NN<0.63	0.02	54/50	0.33	
NN>0.63	0.14	49/50	0.54	

LRHWW 1jet: Observations

- No NN score cut
  - Low S/B, 0 jet, lowest bin ok, data higher in others.
  - Low S/B, 1 jet, $\times 2$ discrepancy in lowest bin, overall data above MC. Fake and Wγ dominate.
  - 1 jet dominated by DY, low s/b shows fakes and Wγ in higher scores.
- 0 Jet, NN cuts
  - High NN Score show data above prediction: more so in Low S/B than High S/B. Background Dominantly WW.
  - Low NN Score in reasonable agreement.
- 1 Jet, NN cuts
  - High NNScore: more events than expected in both high and low s/b

- Low NN score: first bin ok in High S/B, not easy to say in Low S/B; other bins high for low S/B, possibly ok for Low S/B.

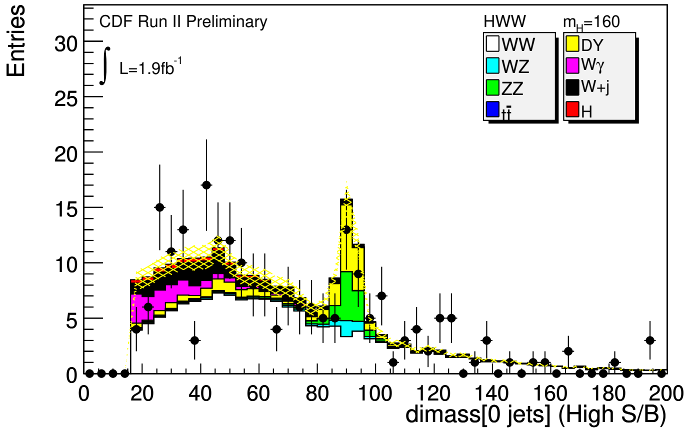
MetSig

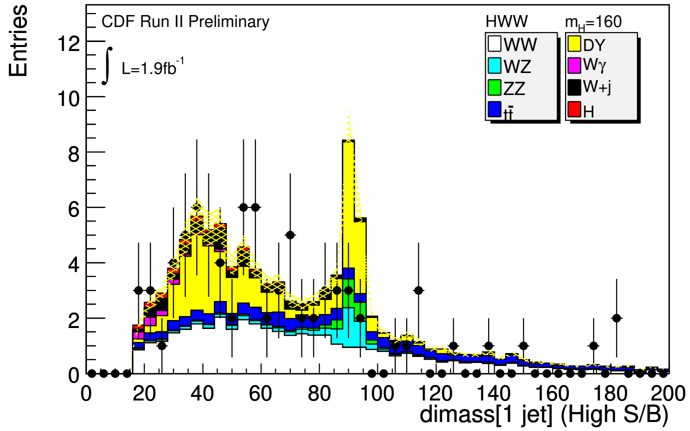

CDF Run II Preliminary
$\int$ L=1.9fb^{-1}
HWW
m_H=160
WW
WZ
ZZ
$t\bar{t}$
DY
Wγ
W+j
H
Entries
MetSig[0 jets] (High S/B)
MetSig[1 jet] (High S/B)
MetSig[0 jets] (Low S/B)
MetSig[1 jet] (Low S/B)

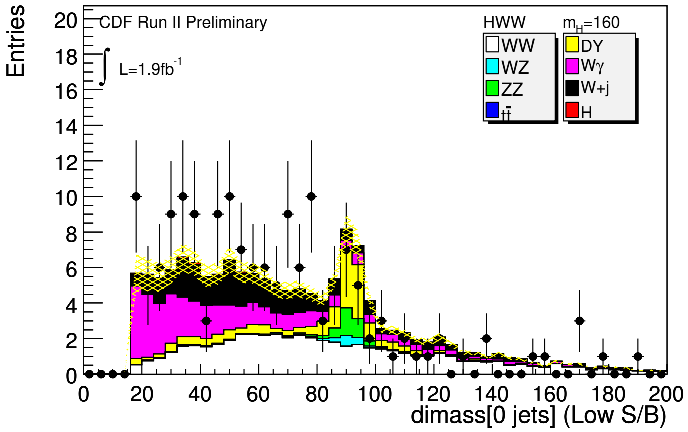
Observations Concerning MetSig

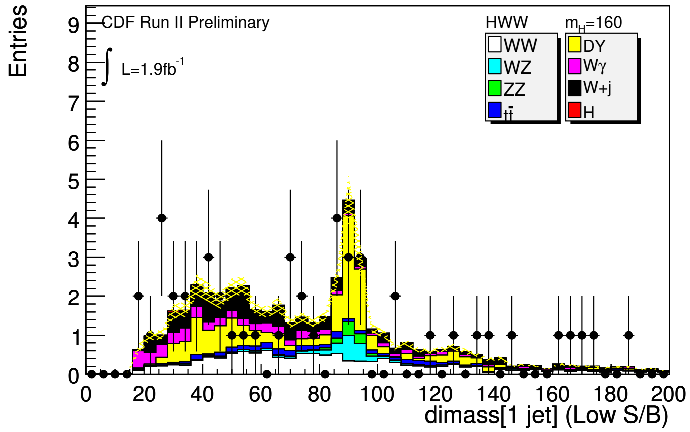
- 0 jet data slightly above prediction in High S/B
- 0 jet data quite above prediction in Low S/B
- Some Fake/Wγ in High S/B, significantly more in Low S/B and shape consistent with the disagreement: a scaling of Fake/Wγ would help with agreement
- High S/B 1 jet has better agreement. The DY contribution good where dominant, but prediction for 3 points around MetSig of 4 is not very good. Relatively little Fake/Wγ.
- Low S/B 1 jet has worse agreement and again more Fake/Wγ expected.

Dilepton Mass







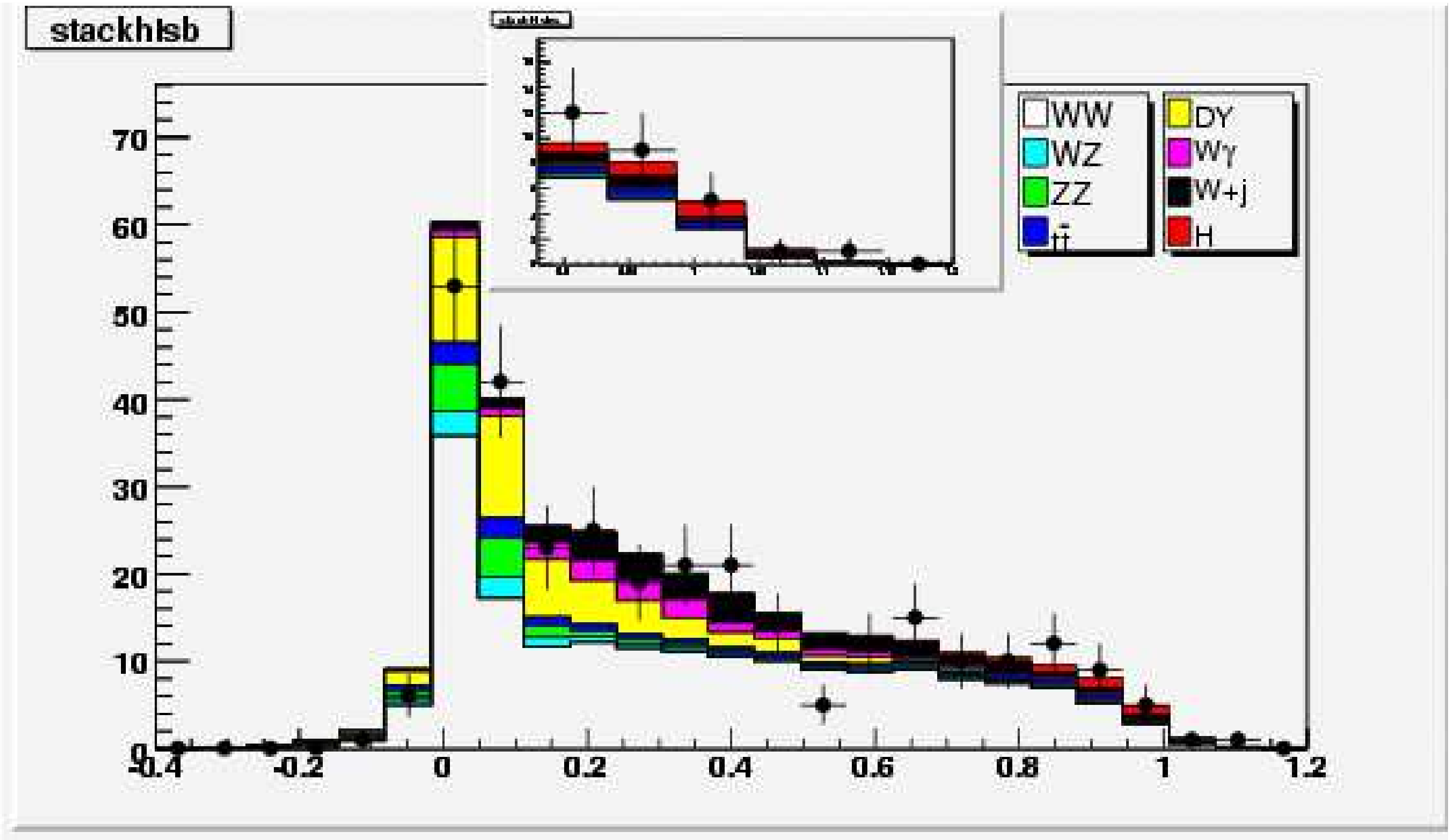


Observations Concerning Dilepton Mass

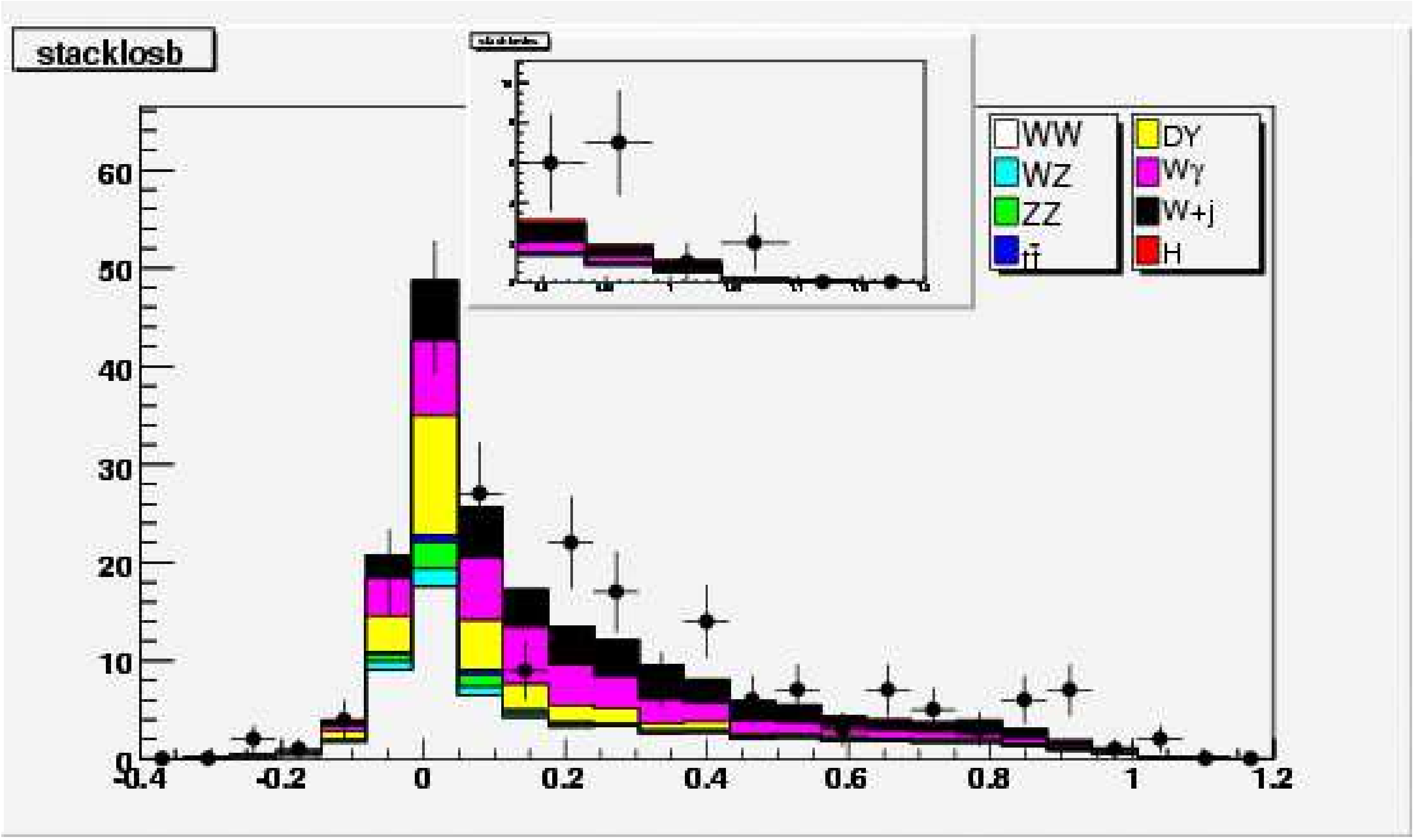
- 0 jet data slightly above prediction in High S/B in the region where more Fake/Wγ is expected.
- 0 jet data quite above prediction in Low S/B in the region where more Fake/Wγ is expected.
- Some Fake/Wγ in High S/B, significantly more in Low S/B and shape consistent with the disagreement: a scaling of Fake/Wγ would help with agreement
- High S/B 1 jet has better agreement execpt at Z resonanace. Relatively little Fake/Wγ.
- Low S/B 1 jet has worse agreement where low Fake/Wγ expected but also there are more events at high mass than expected.

Deficiency in the Fake/Wγ, look at NN distributions.

NN Score Distn, High S/B



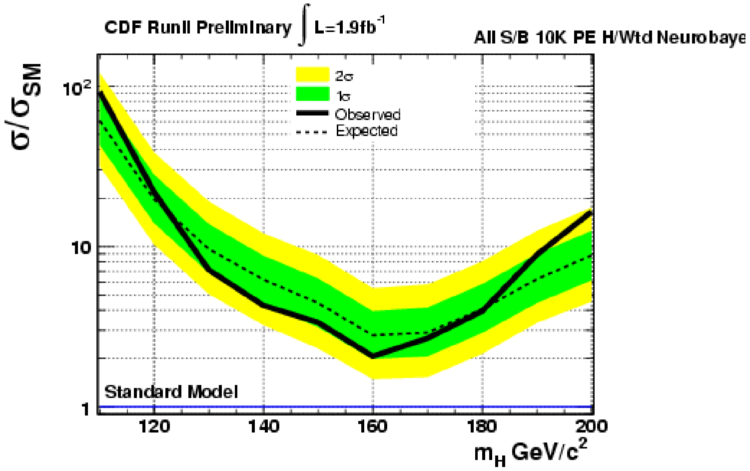
NN Score Distn, Low S/B

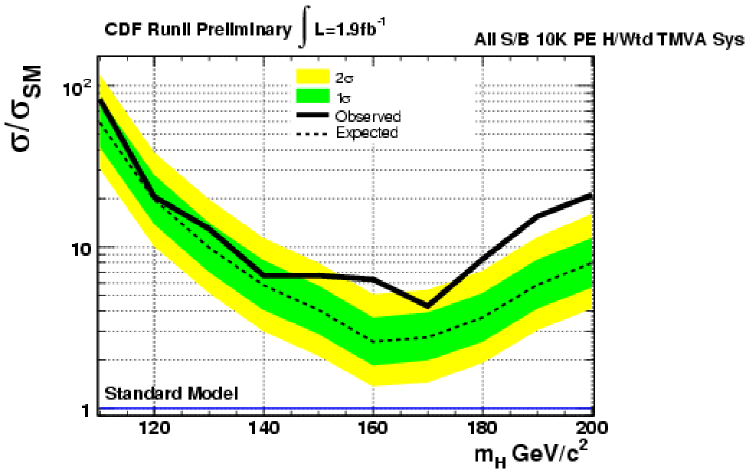


Observations Concerning NN Score

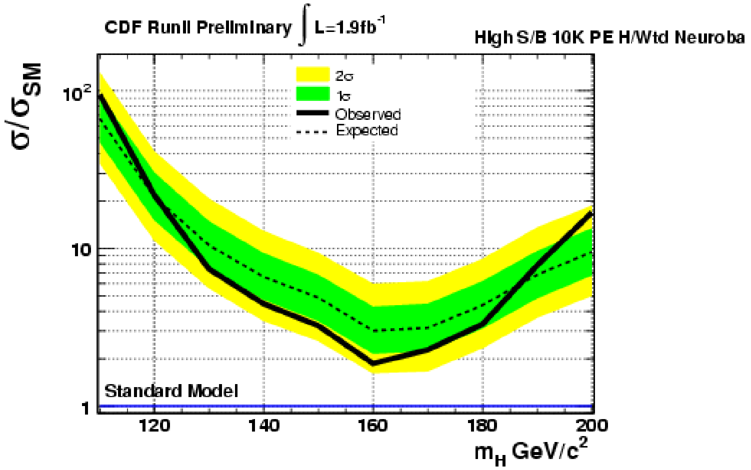
- No $W\gamma$ but there are fakes in High S/B
- Lots $W\gamma$ and Fakes in Low S/B, and points all higher than prediction.

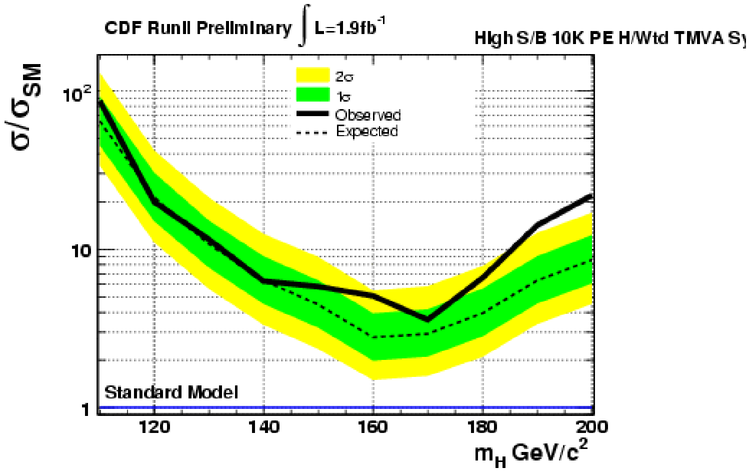
Neurobayes and TMVA: All S/B



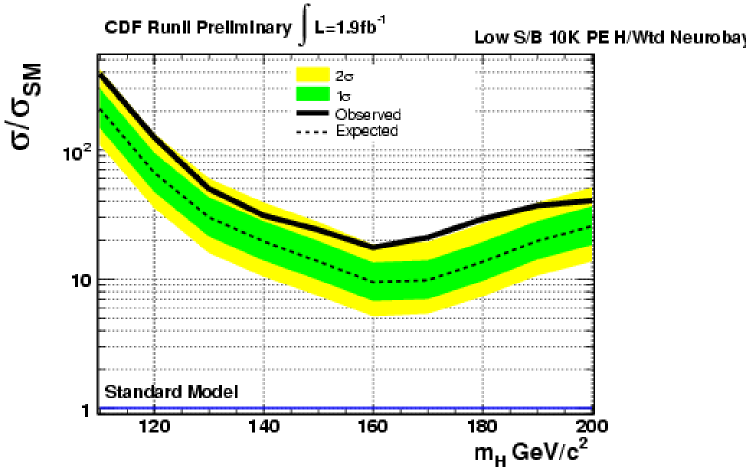


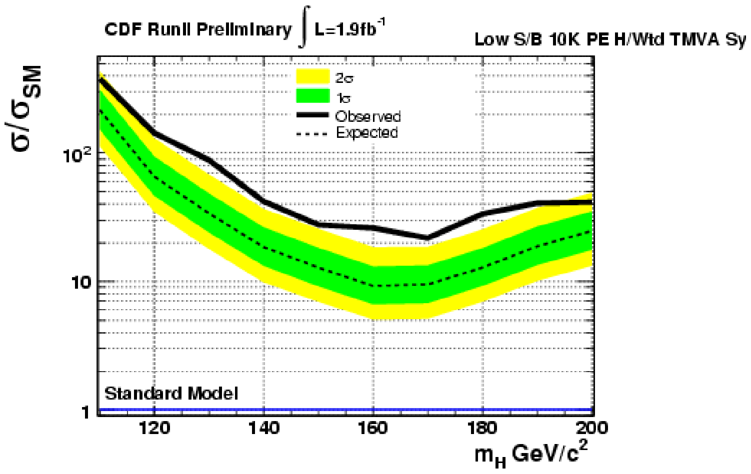
Neurobayes and TMVA: High S/B





Neurobayes and TMVA: Low S/B





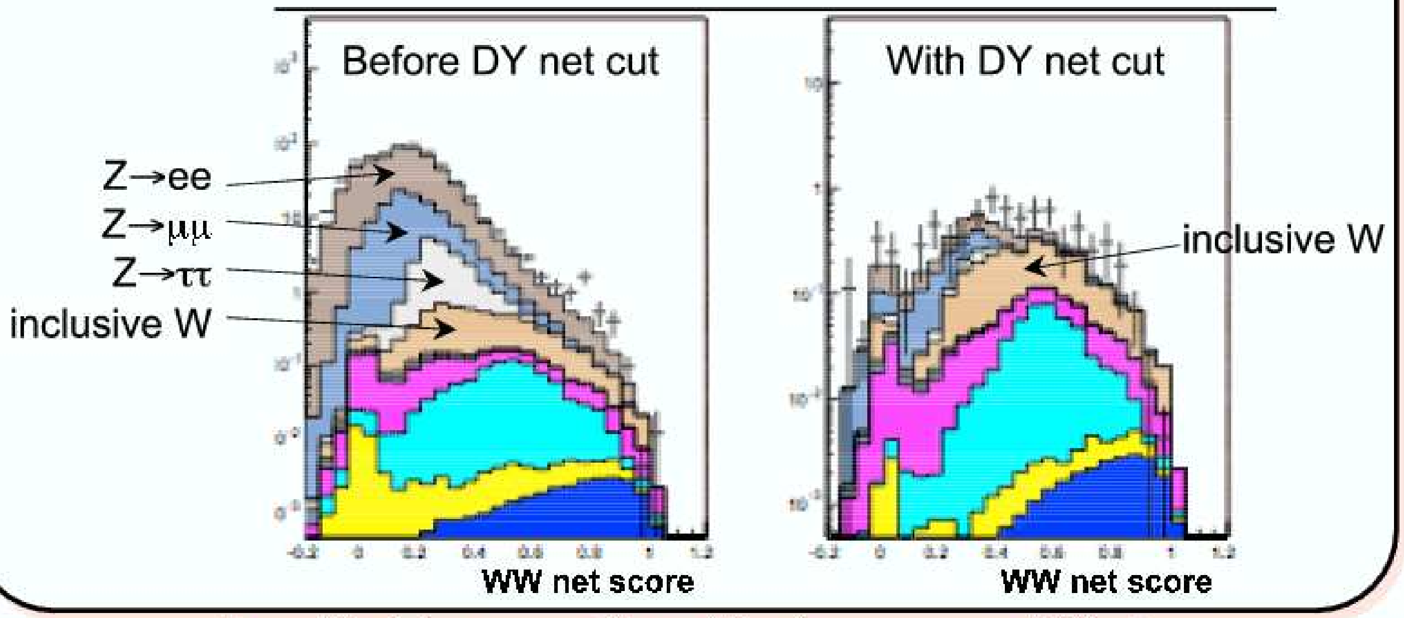
NN lepton + fakeable object

fake = W+jets ???

Fake estimate: ♦ select lepton+"fakeable object" events in data ;
♦ apply jet-based fake rates from data

But what are the lepton +"fakeable object" events? → look in MC

(here, jet-based fake rates from data applied to both data and MC)

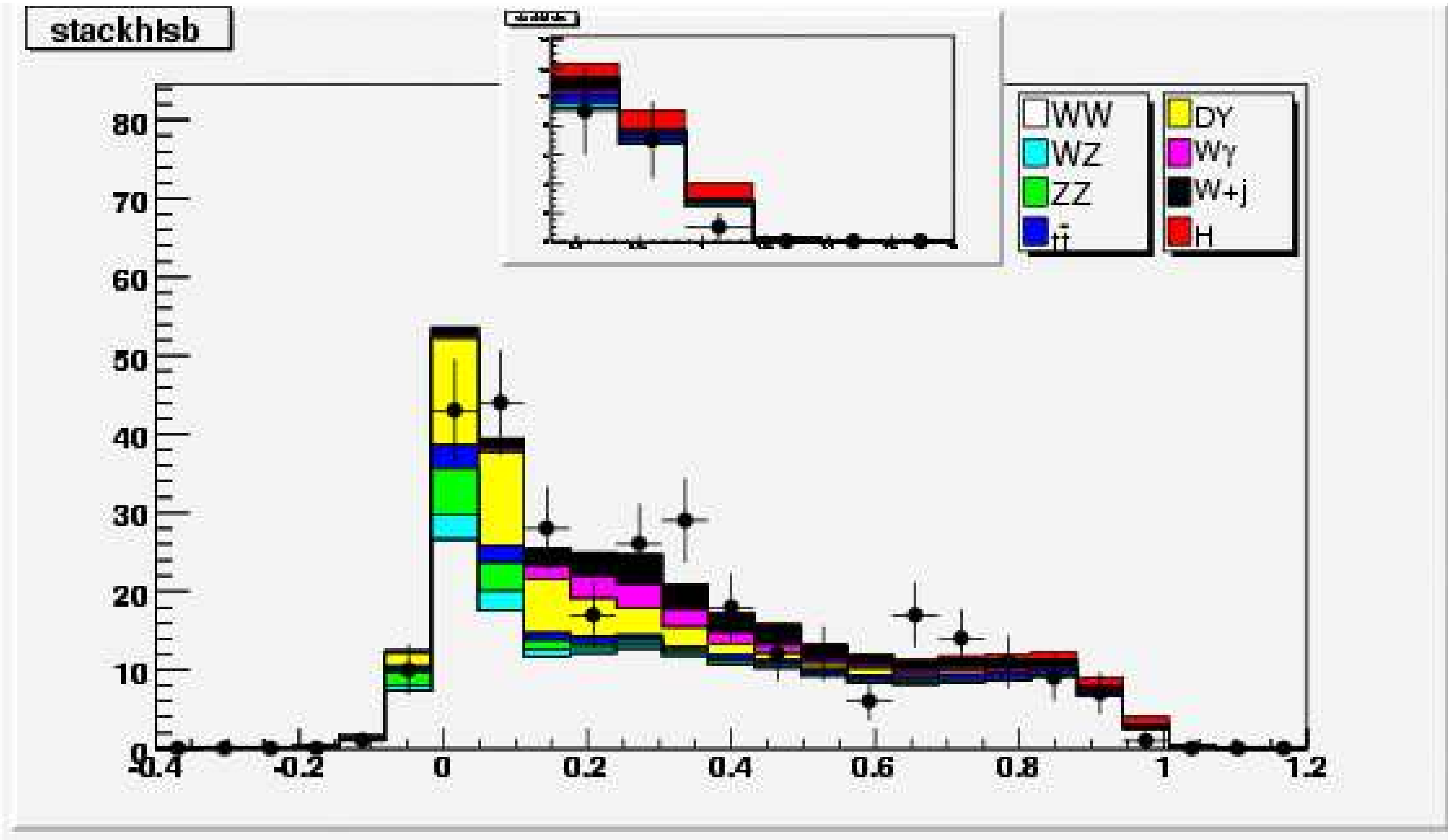


Comment on Fakes

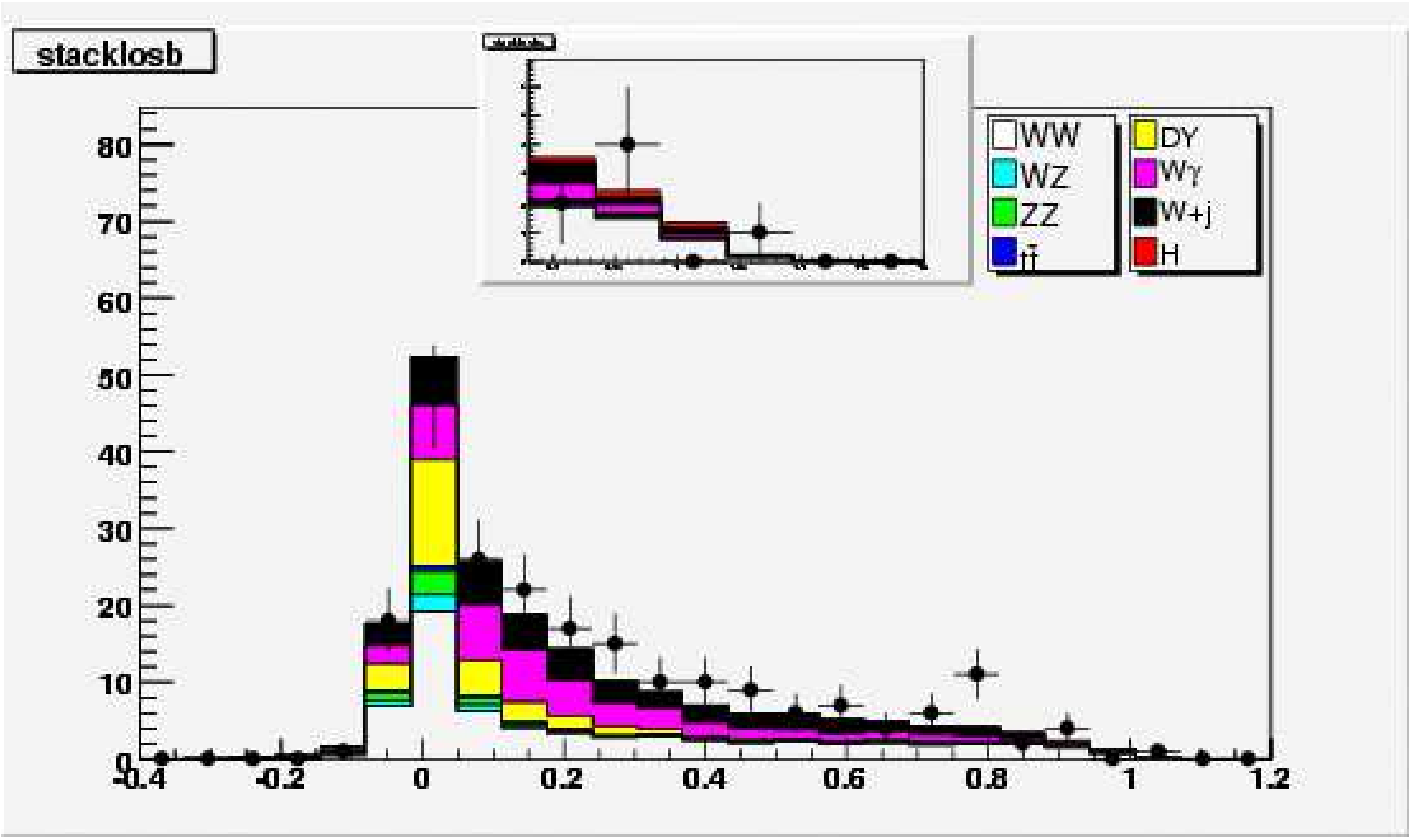
- We select fake denominators in dilepton events
- Most are “dirty leptons” and associated with a good lepton, coming in the dominant channel for dileptons: Drell Yan
- The DY NN removes these nicely and we see that the remaining W+jet and fake is very small and in rough agreement. We took a large systematic.
- The MetSig selection and removal of fakes was not scritinized in the same manner.
- This is difficult without a W+jet MC: We request this sample.
- The DY recognition of fakes may not be as powerful as doing this in two steps.

- We will evaluate the NN scores in a DY/HWW NN and see how MetSig matches this.
- Why does removing MetSig help? Why does Pythia have different sensitivity?

NN Score Distn, High S/B, MetSig removed



NN Score Distn, Low S/B, MetSig removed



Comment on MetSig Removal

- Change in High S/B shape and clearly less events at high NN score.
- Closer match to Low S/B and less Low S/B in tail. More signal.

Summary

- Significant impact of MetSig.
- Interplay between MetSig and LR(HWW).
- Pythia and MC@NLO differences point to the 1 jet modelling.
- 1 jet modelling for overall LRHWW and low S/B is not good
- LRHWW for High NN score shows data above prediction in all bins. This should be examined for other NNs. 1 jet stats low, but more than expected.
- Turning to MetSig: Fakes seem underestimated and the impact in the two neural nets could be different.
- Looking at Fake estimates and differences in High and Low S/B fitting.

Backup

Dilepton Mass:Ratio, Hi S/B