

# A Search for $ZH \rightarrow l^+l^-b\bar{b}$ in $2.4 \text{ fb}^{-1}$ using a Neural Network discriminant

Richard Hughes, Ben Kilminster, Brandon Parks, Brian Winer

*Ohio State University*

Rob Harr, Shalhout Shalhout

*Wayne State University*

Bo Jayatilaka, Ashutosh Kotwal, Ravi Shekhar

*Duke University*

Daniel Whiteson

*UC Irvine*

## Abstract

We search for the Higgs Boson in  $2.4 \text{ fb}^{-1}$  using the process  $ZH \rightarrow l^+l^-b\bar{b}$  in both  $ee$  and  $\mu\mu$  channels. Since the previous  $1 \text{ fb}^{-1}$  analysis we have developed looser lepton identification by making use of a secondary  $Z$  selection trigger path. We correct the two candidate Higgs jets with an Artificial Neural Network which assigns  $\cancel{E}_T$  to the jets according to their  $\cancel{E}_T$  projections and relative  $\phi$ . To maintain signal efficiency and improve signal discrimination, we employ an additional neural network trained to discriminate event kinematics of  $ZH$  compared to the main  $Z + jets$  background and the kinematically different  $t\bar{t}$  background. We calculate a 95% confidence level upper limit of 11.8 times the Standard Model prediction for a Higgs Boson mass of  $115 \text{ GeV}/c^2$  and observe a limit of  $11.6 \times SM$ .

## 1 Introduction

The search for the Higgs boson in the standard model expected  $ZH \rightarrow l^+l^-b\bar{b}$  process, where  $l$  can be an electron or a muon, has a small cross-section compared with  $ZH \rightarrow \nu\bar{\nu}b\bar{b}$  and  $WH \rightarrow l^+\nu b\bar{b}$ . However, it is the best constrained of the three processes since the final state particles are all measured, and there is both a  $Z$  mass resonance in the dilepton invariant mass distribution and a  $H$  mass resonance in the dijet invariant mass distribution. Further, any imbalance of calorimeter energy in the transverse direction  $\cancel{E}_T$  can be attributed mainly to downward fluctuations in the jet energies.

In this search we update the  $1 \text{ fb}^{-1}$  analysis described in Ref. [1]. To select  $ZH$  events we first identify a high  $P_T$  electron or muon, then we require a second lepton, of the same flavor, such that the reconstructed invariant mass of the two leptons is

consistent with that of a  $Z$  boson. While we don't explicitly search for taus, we retain in our signal region Higgs events produced in association with  $Z \rightarrow \tau\tau$ . Next we identify events with two or more high  $E_T$  jets. We correct the jet energies by fitting for their most likely energy loss by utilizing the direction and magnitude of the  $\cancel{E}_T$ , as well as its transverse projection onto the jets. This correction improves the resolution for identifying dijet resonances. Once the jets are corrected, the events have the same kinematic selection as our final signal region, and we use them to validate our analysis technique as a high statistics control region. Next we divide events into those with one tight secvtx  $b$ -tag and two loose secvtx  $b$ -tags. The final selected events are binned according to a 2D NN which is trained to simultaneously separate  $ZH$  from  $t\bar{t}$ , and  $ZH$  from  $Z$ +jets. Using the expected number of background events and their uncertainties, and the final output of the data, we are able to place an upper limit on the cross-section of  $ZH$  that could be contributing.

Many features of this analysis are the same as the previous analysis. The main differences are outlined here. The Monte Carlo simulation is updated and now includes an estimate of the effect of extra interactions in the simulation. We now model  $Z$ +jets processes using Alpgen version 2 Monte Carlo, which is expected to more correctly add in the hard and soft radiation effects which produce the jets. We have added an additional trigger called Z\_NOTRACK which selects events with two high  $E_T$  electromagnetic deposits in the central or plug regions. This complements the standard high  $P_T$  electron trigger which requires a track pointing to a central electromagnetic deposit. We make use of the Z\_NOTRACK trigger by defining additional lepton categories with less restrictive selection criteria, therefore increasing  $Z$  acceptance. The dijet- $\cancel{E}_T$  corrections from the previous analysis are updated with new MC and with an improved set of variables. Our final NN discriminant is trained for a Higgs mass of  $120 \text{ GeV}/c^2$  and optimized to produce the best error in separating backgrounds. Kinematic selection of jets and  $b$ -tagging remain the same in this analysis. The basics of the NN design also remain the same although we allow the NNs to choose a new set of optimal input variables. Since the new Alpgen V2 MC better models the kinematics of  $Z$ +jets, we allow some variables to enter into the NN which were not previously used in the  $1 \text{ fb}^{-1}$  analysis due to poor modeling.

## 2 Event Selection

To identify  $Z$  boson candidates we search for high  $P_T$  lepton pairs with combined invariant mass between  $76 \text{ GeV}/c^2$  and  $106 \text{ GeV}/c^2$ . Our muon selection remains unchanged from the  $1 \text{ fb}^{-1}$  analysis. We select one tight muon from the high  $P_T$  CMX or CMUP trigger path with  $P_T \geq 18 \text{ GeV}/c$ . The second muon is required to have  $P_T \geq 10 \text{ GeV}/c$  and is not required to leave hits in any muon chambers.

For electrons we accept events from the high  $E_T$  Central Electron Trigger path in which we reconstruct one central electron with  $E_T \geq 18 \text{ GeV}$  and an associated track with  $P_T \geq 9 \text{ GeV}/c$ . The second electron is required to have  $E_T \geq 10 \text{ GeV}$  if central,

Event Selection
2 or more Cone 0.4 jets L5 $E_T > 15$ GeV, $ \eta  < 2$ 1 or more jets with L5 $E_T > 25$ GeV
Additional $b$ -tagging selection for signal region
2 or more loose SECVTX tags If not found 1 or more tight SECVTX tags (with $b$ -tagged jet $E_t \geq 25$ )

Table 1: Summary of event selection.

and  $E_T \geq 18$  GeV if in the plug. We also accept events where the second electron is reconstructed from a high  $P_T$  track pointing to one of the gaps between calorimeter wedges. For Z\_NOTRACK triggered events we require two electron candidates with  $E_T \geq 18$  GeV.

After identifying a  $Z$  candidates, we apply additional selection, as in Table 1. The agreement of our preselection is shown in Table 2. We divide our event sample into four categories according to  $\frac{S}{\sqrt{B}}$ . The final tag categories are *Double Tag High*, *Single Tag High*, *Double Tag Low* and *Single Tag Low*.

<b>Event Totals</b>			
Source	Z	Z + 1 jet	preTag
<b>ZH</b>	3.5	3.5	2.8
<b>tt</b>	54.7	54.3	45.4
<b>WW</b>	82.1	26.0	5.5
<b>WZ</b>	189.9	144.7	84.9
<b>ZZ</b>	206.3	135.5	81.7
<b>Z <math>\rightarrow \tau\tau</math></b>	158.7	28.5	4.3
<b>Z+jets (bb)</b>	1540.6	836.9	243.9
<b>Z+jets (cc)</b>	3363.1	1643.3	474.1
<b>mistags / Z + jets</b>	399711.0	60352.1	7140.8
<b>fakes</b>	8253.0	2697.2	591.6
<b>Total Background</b>	413563.0	65921.8	8674.7
<b>Data</b>	420900	70228	9035

Table 2: Preselection event totals.

L5 Jet 1 $E_T$
L5 Jet 2 $E_T$
Jet 1 $\eta$
Jet 2 $\eta$
$\Delta\phi(\text{jet1}, \text{jet2})$
$\Delta\phi(\cancel{E}_T, \text{jet1})$
$\Delta\phi(\cancel{E}_T, \text{jet2})$
Jet 1 Projection onto $\cancel{E}_T$
Jet 2 Projection onto $\cancel{E}_T$
$\cancel{E}_T$ magnitude

Table 3: Variables used to correct jet energies to parton level.

### 3 Jet Energy Resolution Corrections

We correct jet energies with the  $\cancel{E}_T$  direction and magnitude, and projections onto the jet directions as in the procedure in Ref. [2].

Corrections were redone to use the new MC simulation. In order to sample a wide range of jet  $E_T$ s in order to prevent overcorrecting low  $E_T$  jets, we train over Higgs masses (50 - 200 GeV, in 10 GeV steps). We configure the NN slightly differently than previously. We use relative  $\Delta\phi$  between jets and  $\cancel{E}_T$  direction rather than absolute  $\phi$ . We also use a signed  $\cancel{E}_T$  projection to correct jets, rather than an unsigned projection. This has the advantage of not applying the same correction to jets both pointed in the opposite direction and the same direction as the  $\cancel{E}_T$  direction which should reduce mis-reconstruction of non- $Z$  backgrounds like  $t\bar{t}$ .

The variables used to train the dijet mass correction NN are shown in Table 3. The NN corrections provide major improvement with respect to the Level 5 corrections in terms of the jet energy resolution, as well as a better estimate of the parton  $P_T$ . This leads to a truer and narrower reconstructed dijet mass.

### 4 Background modeling of data

The background consists of  $Z$  + heavy flavor jets ( $b$  or  $c$  jets),  $Z$  + incorrectly tagged light flavor jets,  $t\bar{t}$ ,  $ZZ$ ,  $ZW$ , and events with fake leptons.

To model the  $Z$  + jets,  $t\bar{t}$ ,  $ZZ$  and  $ZW$  backgrounds we use Alpgen+Pythia MC. We model  $Z \rightarrow \mu\mu$  fake events using like sign di-muon events and employ the data based fake method described in Ref. [3] to estimate the  $Z \rightarrow ee$  fake background.

The  $Z$  + tagged light flavor jets (*misTag*) background is modeled using the mistag matrix method described in Ref. [4].

NN Inputs
jet 1 $E_t$
$\Delta R(Z, jj)$
jet 2 $E_t$
Z $P_t$
$M_{jj}$
Missing $E_t$
MET projected onto jet 1
Number of Tight Jets
HT1
$\Delta R(jet1, jet2)$
MET projected onto jet 2
sphericity
Mass Z+jj

Table 4: 2D NN inputs.

## 5 Neural Network Signal-Background discriminant

We develop a 2D NN using the same techniques as the  $1 \text{ fb}^{-1}$  analysis. We consider 29 input variables, after dijet energy corrections are applied, and iteratively add in the best variables until adding additional variables does not improve the training error.

Our final NN configuration uses the inputs listed in Table 4. Projections of the NN output are shown in Figures 1 and 2 for preTag data and MC.

## 6 Results

After applying  $b$ -tagging our final event totals are shown in tables 5 and 6. Since the event yields in data agree with the events expected by our background model, we do not find a significant excess in the counting experiment. The projected NN output distributions for the signal regions are shown in Figures 3 through 10. Since the cross-section for  $ZH$  is small, we do not expect a visible excess with this dataset, and so we quantify the maximum allowed  $ZH$  contamination in the data. We use the mclimit machinery for this, and do a binned fit of the 2D NN distribution, including systematics. Table 6, Table 8, Figure 11 and Figure 12 show our expected limits assuming no Higgs production, and our observed limits for  $ZH$  production in the data.

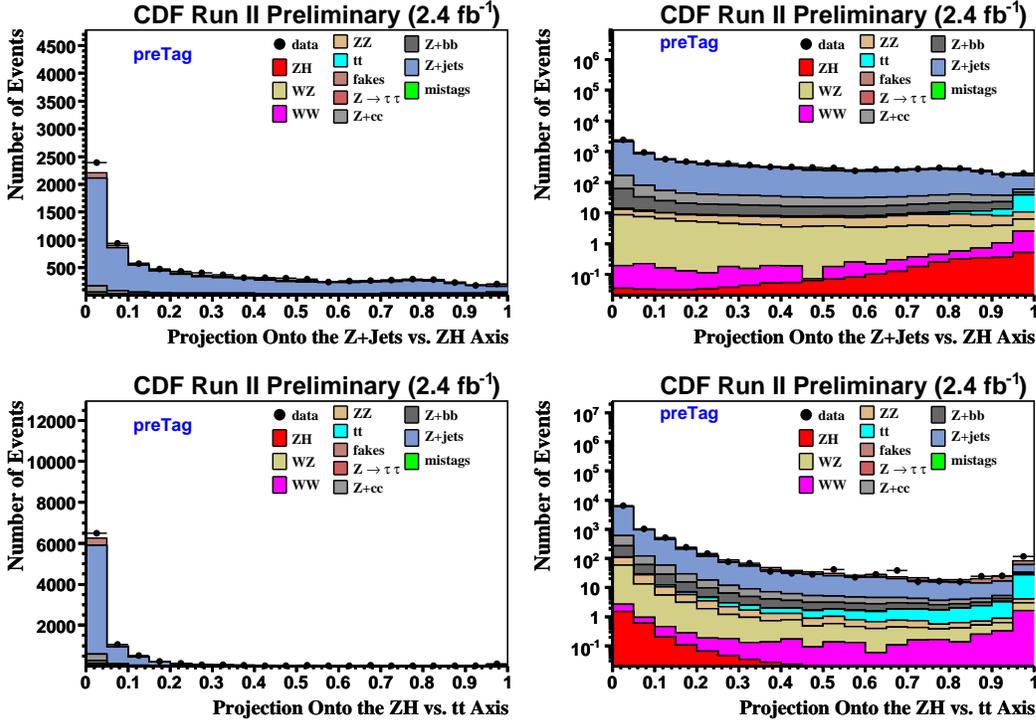


Figure 1: Pre-Tag NN output projections.

Event Totals		
Source	Single Tag (high)	Double Tag (high)
ZH	$0.86 \pm 0.07$	$0.45 \pm 0.04$
tt	$11.36 \pm 2.39$	$6.14 \pm 1.29$
WW	$0.11 \pm 0.02$	$0.01 \pm 0$
WZ	$2.37 \pm 0.35$	$0.13 \pm 0.02$
ZZ	$6.31 \pm 0.95$	$2.23 \pm 0.31$
$Z \rightarrow \tau\tau$	$0.03 \pm 0.01$	$0 \pm 0$
Z+jets (bb)	$53.41 \pm 22.43$	$13.41 \pm 5.5$
Z+jets (cc)	$25.67 \pm 10.78$	$1.92 \pm 0.79$
mistags / Z + jets	$84.21 \pm 36.21$	$2.11 \pm 0.93$
fakes	$13.5 \pm 6.75$	$0.82 \pm 0.41$
<b>Total Background</b>	$197.84 \pm 44.53$	$27.23 \pm 5.8$
<b>Data</b>	205	24

Table 5: Event totals for single and double tag (high).

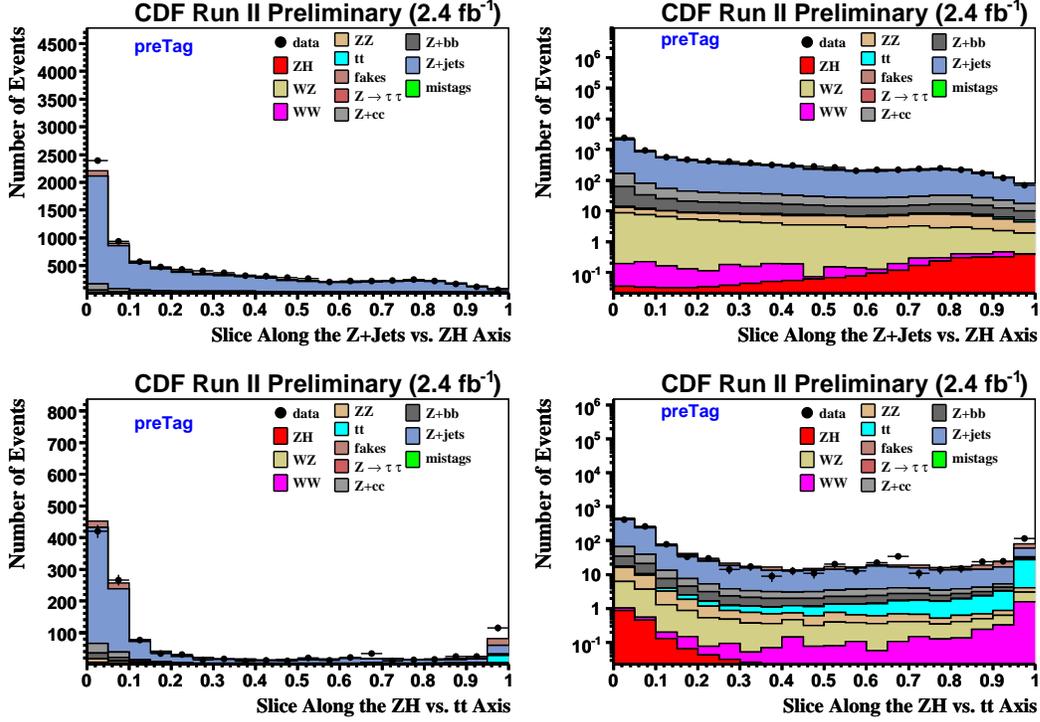


Figure 2: Pre-Tag NN output projections with  $y \leq 0.25$  in the Z+Jets vs. ZH projections and  $x \geq 0.75$  in the ZH vs.  $t\bar{t}$  projections.

Event Totals		
Source	Single Tag (low)	Double Tag (low)
ZH	$0.14 \pm 0.01$	$0.07 \pm 0.01$
$t\bar{t}$	$5.49 \pm 1.15$	$2.51 \pm 0.53$
WW	$0.06 \pm 0.01$	$0 \pm 0$
WZ	$0.63 \pm 0.09$	$0.02 \pm 0$
ZZ	$1.19 \pm 0.18$	$0.41 \pm 0.06$
$Z \rightarrow \tau\tau$	$0.05 \pm 0.02$	$0 \pm 0$
Z+jets (bb)	$11.15 \pm 4.68$	$2.8 \pm 1.15$
Z+jets (cc)	$5.49 \pm 2.31$	$0.37 \pm 0.15$
mistags / Z + jets	$29.34 \pm 12.61$	$0.62 \pm 0.27$
fakes	$12.36 \pm 6.18$	$0.54 \pm 0.27$
<b>Total Background</b>	$65.89 \pm 15.03$	$7.33 \pm 1.33$
<b>Data</b>	74	15

Table 6: Event totals for single and double tag (low).

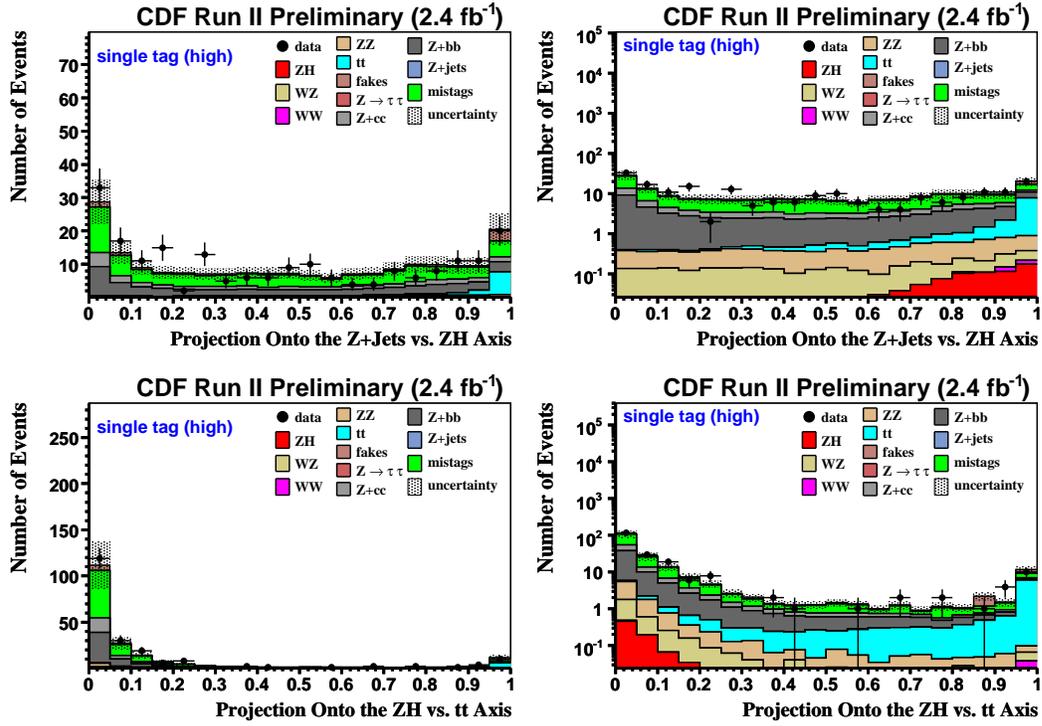


Figure 3: Single Tag (high) NN output projections.

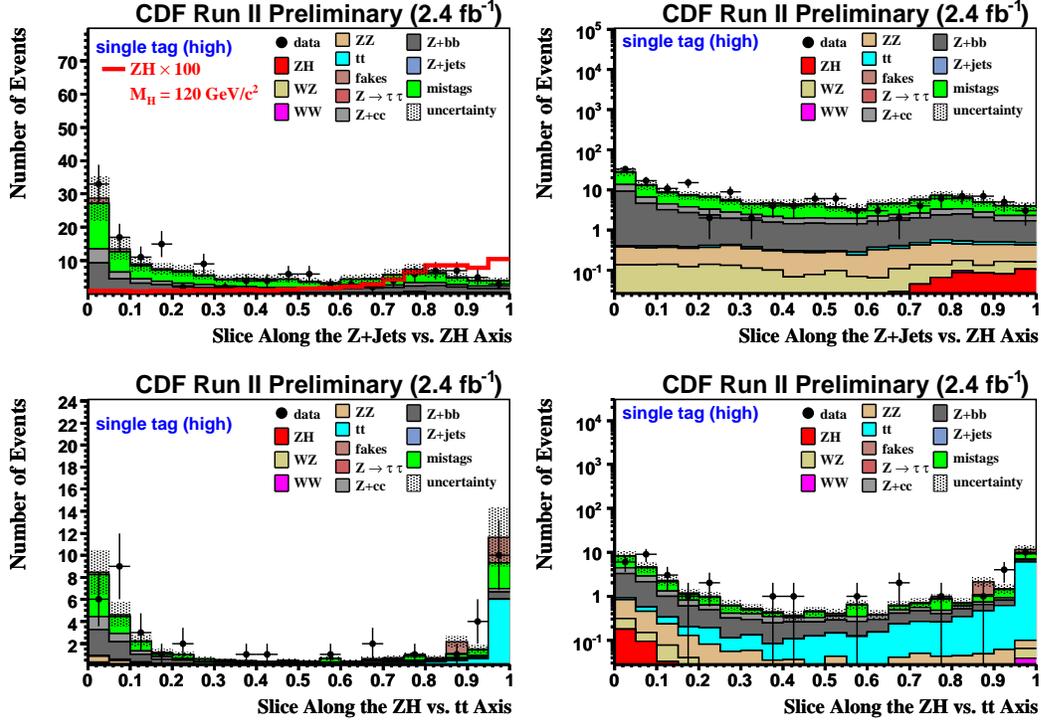


Figure 4: Single Tag (high) NN output projections with  $y \leq 0.1$  in the Z+Jets vs. ZH projections and  $x \geq 0.9$  in the ZH vs.  $t\bar{t}$  projections.

Expected Limits						
$M_H$	Single Tag ( <i>low</i> )	Double Tag ( <i>low</i> )	Single Tag ( <i>high</i> )	Double Tag ( <i>high</i> )	Combined ( <i>high only</i> )	Combined ( <i>low + high</i> )
100	45.63	38.5	18.82	13.16	10.6	9.59
105	39.19	35.1	16.32	11.34	8.99	8.15
110	51.9	42.74	20.53	15.34	11.99	10.7
115	58.42	51.33	23.57	16.49	12.74	11.79
120	79.09	69.89	27.24	19.87	15.99	14.51
125	86.44	80.19	32.34	24.15	18.39	16.67
130	108.56	91	42.15	30.5	23.98	21.58
135	155.81	129.79	58.04	40.87	32.51	29.38
140	209.5	174.59	82.19	58.02	46.06	42.04
145	332.82	277.76	126.26	89.53	70.84	64.36
150	519.5	463.51	213.67	144.6	116.32	104.83

Table 7: Expected limits by each tag category and in combination.

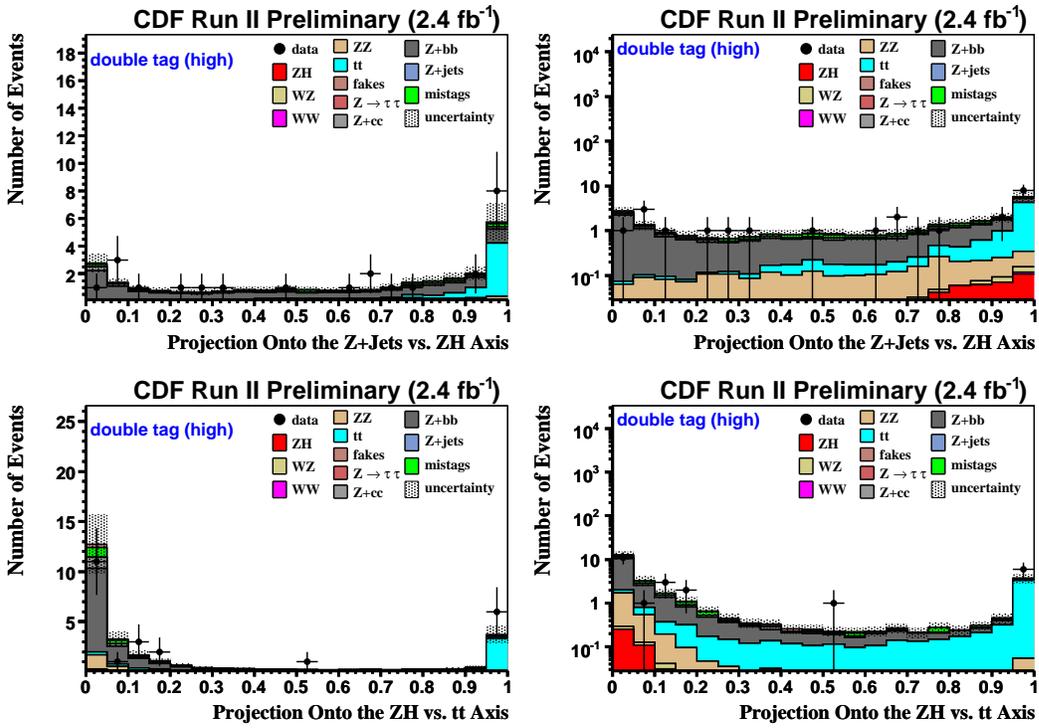


Figure 5: Double Tag (high) NN output projections.

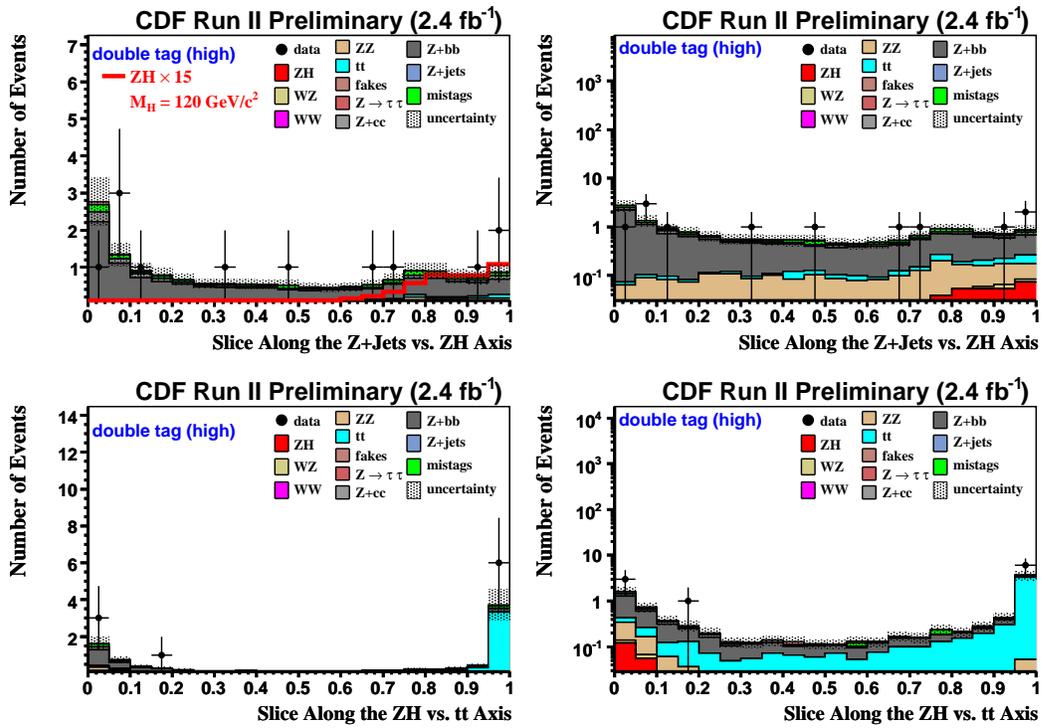


Figure 6: Double Tag (high) NN output projections with  $y \leq 0.1$  in the Z+Jets vs. ZH projections and  $x \geq 0.9$  in the ZH vs.  $t\bar{t}$  projections.

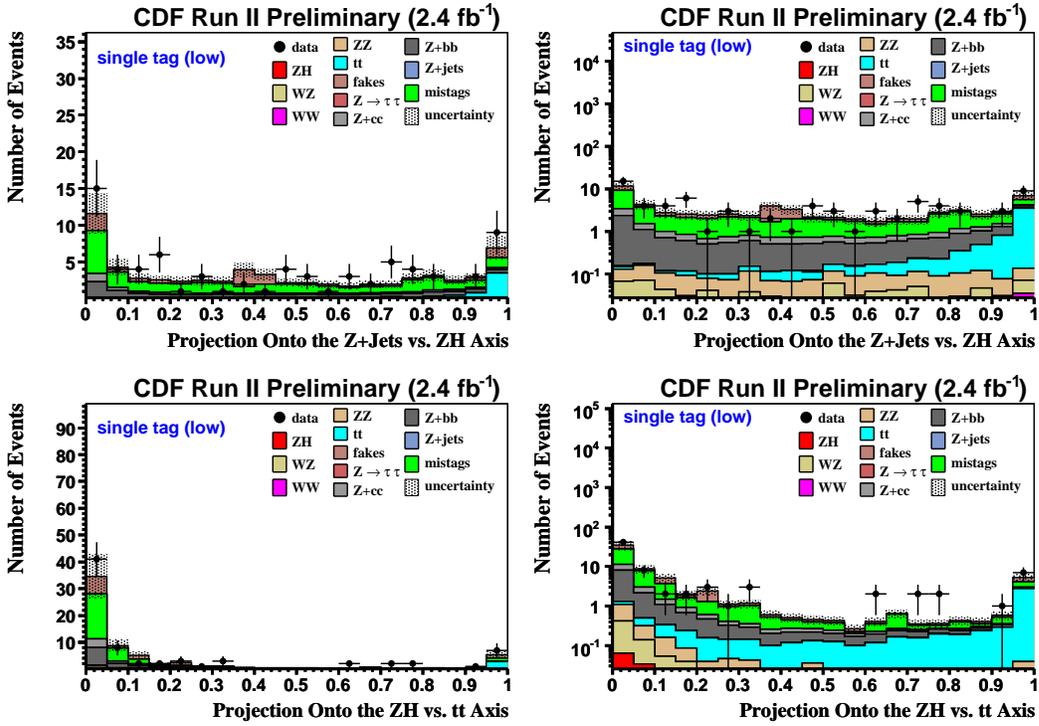


Figure 7: Single Tag (low) NN output projections.

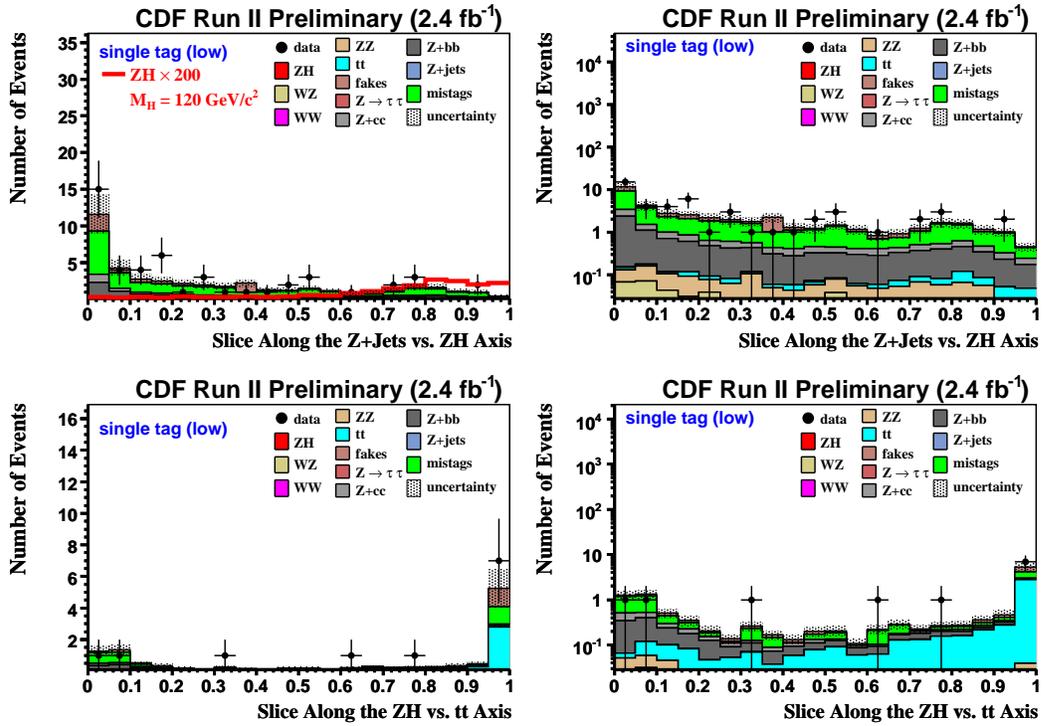


Figure 8: Single Tag (low) NN output projections with  $y \leq 0.1$  in the Z+Jets vs. ZH projections and  $x \geq 0.9$  in the ZH vs.  $t\bar{t}$  projections.

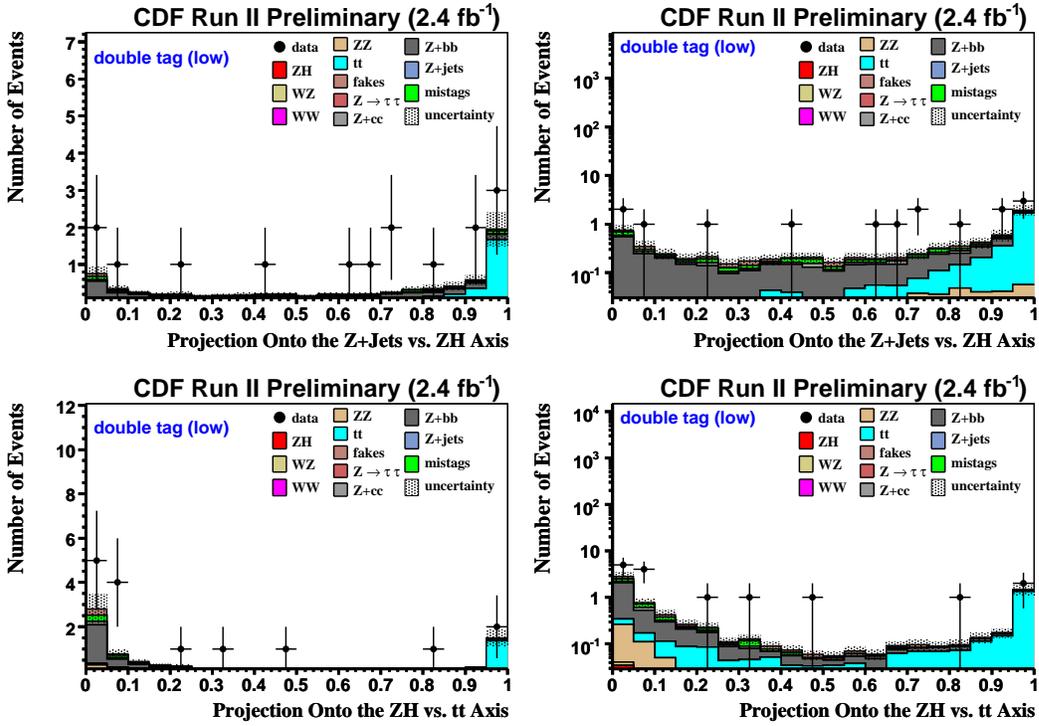


Figure 9: Double Tag (low) NN output projections.

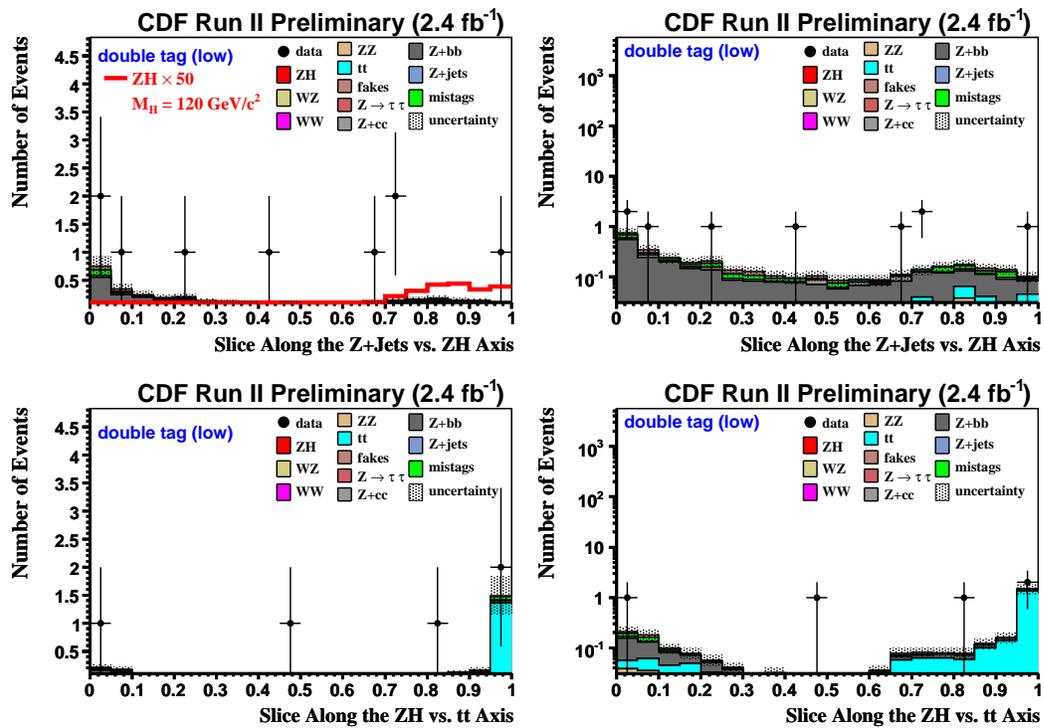


Figure 10: Double Tag (low) NN output projections with  $y \leq 0.1$  in the Z+Jets vs. ZH projections and  $x \geq 0.9$  in the ZH vs.  $t\bar{t}$  projections.

<b>Expected and Observed Limits</b>		
$M_H$	Expected $\pm \frac{1\sigma}{1\sigma}$	Observed
100	9.59 $\pm \frac{4.4}{2.83}$	8.79
105	8.15 $\pm \frac{3.57}{2.43}$	8.02
110	10.7 $\pm \frac{5.01}{3.14}$	10.62
115	11.79 $\pm \frac{4.84}{3.55}$	11.57
120	14.51 $\pm \frac{6.68}{4.35}$	13.36
125	16.67 $\pm \frac{7.71}{4.77}$	17.16
130	21.58 $\pm \frac{10.13}{6.34}$	21.23
135	29.38 $\pm \frac{14.23}{8.97}$	27.11
140	42.04 $\pm \frac{18.34}{12.44}$	39.73
145	64.36 $\pm \frac{27.29}{18.52}$	65.01
150	104.82 $\pm \frac{46.3}{32.49}$	104.83

Table 8: Expected and observed limits.

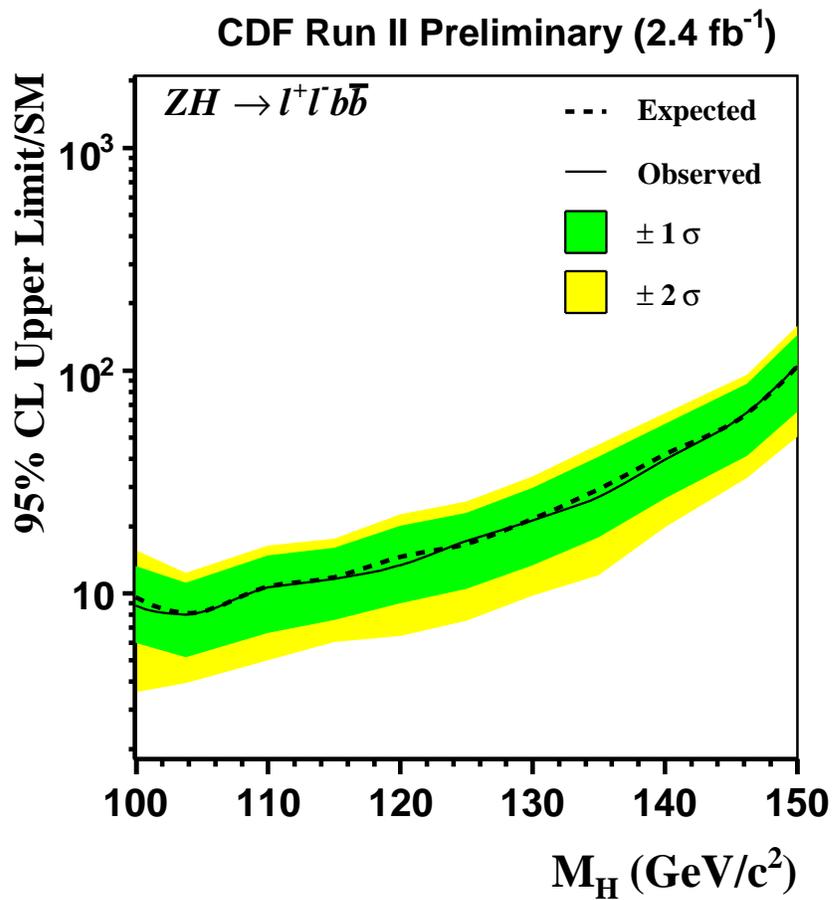


Figure 11: Combined expected and observed limits.

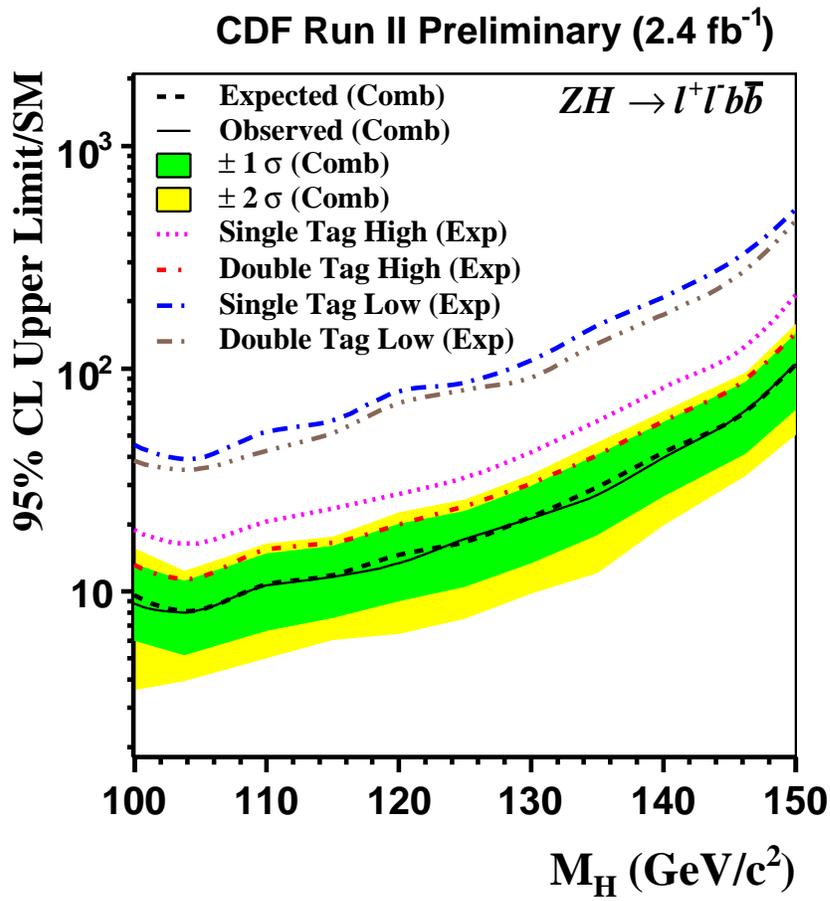


Figure 12: Expected limits for each of the individual channels and in combination.

## 7 Conclusions

We have evaluated new limits with an updated dataset, 2.4 times larger than in the previous analysis. We have made use of new lepton categories to increase  $ZH$  acceptance. In order to gain full sensitivity from our new lepton categories we have added two additional tag categories which reduce the expected limit by  $\sim 10\%$  when compared to the high  $\frac{S}{\sqrt{B}}$  channels alone.

We calculate 95% confidence level upper limits from 9.6 to 104.8 times the Standard Model prediction for Higgs Boson masses between  $100 \text{ GeV}/c^2$  and  $150 \text{ GeV}/c^2$ . For  $m_H = 115 \text{ GeV}/c^2$  the expected 95% confidence level upper limit is 11.8 times the Standard Model prediction with an observed limit of 11.6.

## References

- [1] T. Aaltonen [CDF Collaboration], “Search for the Higgs boson produced with  $Z \rightarrow \ell^+\ell^-$  in  $p\bar{p}$  collisions at  $\sqrt{s} = 1.96$  TeV” arXiv:0807.4493 [hep-ex].
- [2] J. Z. Efron, Ph.D. thesis, Ohio State U. (2007), FERMILAB-THESIS-2007-20
- [3] A. Abulencia *et al.* [CDF Collaboration], Phys. Rev. D **74**, 032008 (2006) [arXiv:hep-ex/0605099].
- [4] R. D. Harrington, “Measurement of the top quark mass in lepton+jets events with secondary vertex tagging” FERMILAB-THESIS-2007-25;