



Search for $ZH(\rightarrow l^+l^-b\bar{b})$ in $p\bar{p}$ collisions at $\sqrt{s} = 1960$ GeV

The DØ Collaboration
URL <http://www-d0.fnal.gov>
(Dated: August 14, 2007)

We search for the standard model Higgs boson produced in association with a Z boson in approximately 1.1 fb^{-1} of $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV using the DØ detector at the Fermilab Tevatron collider. Events where the Z boson decays to e^+e^- or $\mu^+\mu^-$ and the Higgs boson decays to $b\bar{b}$ are considered. To improve sensitivity, the data is divided into single and double b -tagged channels, and neural networks are trained to separate signal from background. Good agreement between the data and expected backgrounds are observed, and upper limits on the ZH production cross section are set for Higgs boson masses between 105 and 145 GeV.

Preliminary Results for Summer 2007 Conferences

I. INTRODUCTION

One of the most sensitive search channels at the Tevatron for a standard model Higgs boson with a mass below approximately 140 GeV is the associated production of a Higgs boson with a Z boson, where the Higgs decays to $b\bar{b}$. We present in this note a search for ZH production in the $l^+l^-b\bar{b}$ final states, where $l = e, \mu$. The Z boson is reconstructed from a pair of high p_T leptons with an invariant mass constraint. Events are required to have either at least two loosely b -tagged jets or one tightly b -tagged jet. The dominant backgrounds result from the associated production of a Z boson with jets, among which the $Zb\bar{b}$ production is an irreducible background. The other main backgrounds are $t\bar{t}$, WZ , ZZ , and multijet production from QCD processes. Instead of simply searching for a $H \rightarrow b\bar{b}$ resonance in the dijet mass distribution, a multivariate neural-net (NN) algorithm is trained on simulated signal against simulated backgrounds. An excess in data from a Higgs signal is searched for in the high NN output region.

II. THE DØ DETECTOR

The DØ detector has a central-tracking system, consisting of a silicon microstrip tracker (SMT) and a central fiber tracker (CFT), both located within a 2 T superconducting solenoidal magnet [1], with tracking and vertexing at pseudorapidities $|\eta| < 3$ and $|\eta| < 2.5$, respectively. A liquid-argon and uranium calorimeter has a central section (CC) covering $|\eta|$ up to ≈ 1.1 , and two end calorimeters (EC) that extend coverage to $|\eta| \approx 4.2$. An outer muon system, at $|\eta| < 2$, consists of a layer of tracking detectors and scintillation trigger counters in front of 1.8 T toroids, followed by two similar layers after the toroids.

III. ANALYSIS

These analyses are based on data taken between April 2002 and March 2006. No explicit trigger requirements were made, in order to retain the highest possible efficiency for signal. The integrated luminosities were found to be approximately 1.1fb^{-1} for both the dielectron and dimuon channels, after the requirement of good data quality.

The selection of hadronic jets is common to both dilepton final states. A jet is reconstructed using the RunII cone algorithm with $\Delta R = 0.5$ [2]. The jet must have $p_T > 15$ GeV after the jet energy scale correction and a pseudorapidity of $|\eta| < 2.5$. In the simulation, jets are corrected to account for the difference in reconstruction efficiency and energy resolution from the data. All corrections to the jets and leptons are propagated in the computation of the missing transverse energy, \cancel{E}_T , and scalar E_T .

Events for the dielectron channel are required to have at least two electron candidates, with $p_T > 15$ GeV, $|\eta_{detector}| < 1.1$ or $1.5 < |\eta_{detector}| < 2.5$, and satisfying electron shower shape criteria and a match with a central track. In the simulation, the EM identification efficiency is corrected to the data one as a function of η for each electron. A good Z candidate is required, reconstructed from two electrons. The reconstructed Z mass, m_Z must satisfy $70 < m_Z < 110$ GeV. At least two jets are required in each event with each jet being away from both good electrons by $\Delta R > 0.5$.

Events for the dimuon channel are required to have at least two loose-quality muons, each matched with a central track, and $p_T > 10$ GeV. The muon p_T in data is corrected using the beam spot position for each run, if the muon track has no SMT hits. A good Z candidate is required with $70 < m_Z < 110$ GeV. The two muons which form the Z are together required to have *product scaled isolation* < 0.01 . The product scaled isolation is the product of the scaled isolation variables for each of the two muons from the Z : $productscaledisolation = scaledisolation(muon1) \times scaledisolation(muon2)$ where $scaledisolation = (p_T^{track} + p_T^{calorimeter})/p_T^{muon}$. The p_T^{track} is the sum of all tracks in a cone of $\Delta R < 0.5$ around the muon track, except the muon track. The $p_T^{calorimeter}$ is the sum of all calorimeter energy in a hollow cone around the muon from $0.1 < \Delta R < 0.4$.

For both analyses, at least two jets must be loosely (2L), or one jet tightly (1T), b -tagged using a neural network b -tagging algorithm [3]. Overlapping 2L events are removed from the 1T channel in both data and simulations. The tag-rate-function (TRF) was measured in data and parameterized, then used to predict the probability that a *taggable* jet of a given flavor would be tagged in the simulation. A taggable jet is defined as a jet which matches a track-jet cluster within $\Delta R < 0.5$ containing two tracks with at least one SMT hit and $p_T > 1.0$ (0.5) GeV for the first (other) tracks. The loose NN b -tagging requirement was set to accept a 6% light-jet fake rate, providing a 72% b -tagging efficiency in the central region ($|\eta| < 1.5$), for $p_T > 15$ GeV, for taggable jets. The tight NN b -tagging efficiency was 45%, for a 0.5% light-jet fake rate in the same kinematic region.

TABLE I: The number of data, background, and expected SM signal events (and systematic uncertainty) after 0, 1 tight (but not ≥ 2 loose), and ≥ 2 loose b -tagged jets are required, in the dielectron and dimuon channels.

Sample	dielectron channel			dimuon channel		
	Before b -tagging	1T	2L	Before b -tagging	1T	2L
Data	2985	73	24	4669	87	53
Total Bgnd.	2961 \pm 533	57.4 \pm 21	27.9 \pm 7.8	4671 \pm 841	101.4 \pm 38	45.7 \pm 12.8
ZH 115 GeV	0.76 \pm 0.034	0.23 \pm 0.012	0.23 \pm 0.019	0.69 \pm 0.031	0.31 \pm 0.17	0.30 \pm 0.025
QCD	89.4 \pm 44.7	1.73 \pm 0.87	0.74 \pm 0.37	47.3 \pm 23.7	0.25 \pm 0.13	0.88 \pm 0.44
$Z + (udscg)$	2451 \pm 490 (udsg)	7.17 \pm 1.4 (udsg)	5.12 \pm 1.0 (udsg)	4400 \pm 880	46.0 \pm 9.2	19.2 \pm 3.8
	252 \pm 50 (c)	14.1 \pm 2.8 (c)	5.53 \pm 1.1 (c)			
$Z + 2b$	120 \pm 26	29.6 \pm 9.5	12.0 \pm 3.8	168 \pm 37	47.6 \pm 15.2	18.5 \pm 5.9
$t\bar{t}$	9.6 \pm 0.86	2.8 \pm 0.28	3.1 \pm 0.37	10.6 \pm 0.96	4.8 \pm 0.48	5.3 \pm 0.64
WZ	20.3 \pm 1.8	0.60 \pm 0.07	0.17 \pm 0.02	24.2 \pm 2.2	0.82 \pm 0.09	0.24 \pm 0.03
ZZ	18.0 \pm 1.4	1.40 \pm 1.3	1.17 \pm 0.13	21.4 \pm 1.7	1.95 \pm 0.18	1.48 \pm 0.16

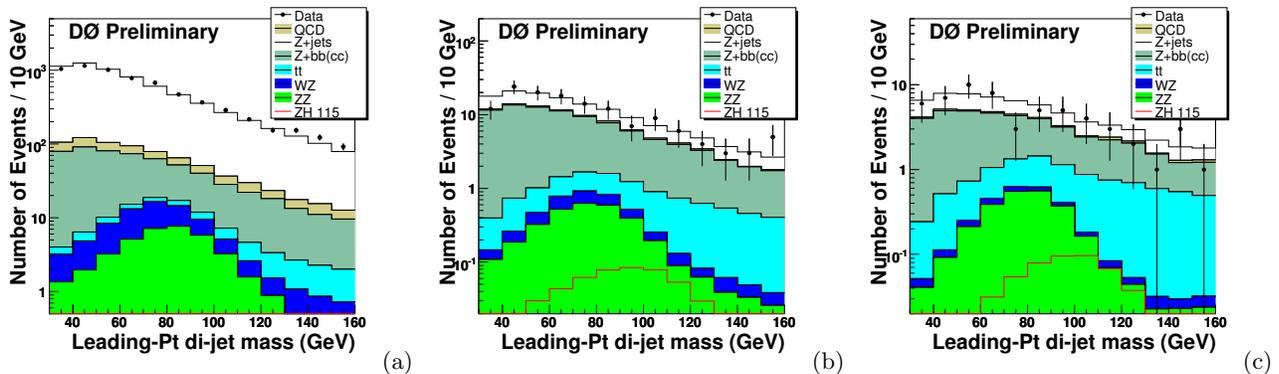


FIG. 1: The leading- p_T di-jet invariant mass distributions for non- b -tagged (a), 1T (b), and 2L (c) channels, for a Higgs mass of 115 GeV.

IV. SIMULATION

Using the CTEQ6L1 [4] leading-order parton distribution functions, the following physics processes are simulated to estimate the signal acceptance and the number of background events: $Z(\rightarrow l^+l^-)H(\rightarrow b\bar{b})$ by PYTHIA [6], $Z(\rightarrow l^+l^-)jj$ including $Z(\rightarrow l^+l^-)cc$ by ALPGEN [7], $Z(\rightarrow l^+l^-)bb$ by ALPGEN, $t\bar{t} \rightarrow l^+\nu b l^-\bar{\nu}b$ and $t\bar{t} \rightarrow bbjjl\nu$ by ALPGEN, inclusive ZZ and WZ by PYTHIA. The samples generated by ALPGEN are interfaced with PYTHIA for parton showering and hadronization. All the samples were run through the full GEANT based detector simulation, digitization, and reconstruction. The signal cross sections, as well as those for $t\bar{t}$, WZ , and ZZ are taken from MCFM [5], which is NLO. For the ALPGEN samples, a matching procedure (MLM) was used so as to not double count the radiation of additional jets between ALPGEN and PYTHIA. In addition, NLO k-factors of 1.23 and 1.35 were applied to the Z +light and $Z+b\bar{b}$ samples, respectively, to account for the NLO cross sections of these process (as calculated with MCFM) as compared to the LO cross sections from ALPGEN. All simulated samples are scaled, prior to b -tagging to agree with the number of observed data event under the Z mass peak. Also, since the Z p_T distribution is poorly modeled in Z + light jet simulation, the Z + light jet simulated samples are re-weighted to the Z p_T observed in data before b -tagging. The shape of the QCD background, where jets create the reconstructed leptons, is obtained from data where the QCD contribution is enhanced by inverting the lepton quality requirements. The QCD contribution is then normalized by fitting the dilepton invariant mass by a Breit-Wigner function convoluted with a Gaussian for the Z peak, and an exponential shape for QCD and Drell-Yan processes. The fraction of Drell-Yan events is determined from the simulation.

The total amount of data, various backgrounds, and expected signal are shown in Table I after 0, 1, and at least 2 b -tagged jets are required, in the dielectron and dimuon channels. Figure 1 shows the invariant mass of the two leading- p_T jets, in the non- b -tagged, 1T, and 2L channels, for the combined dielectron and dimuon samples.

V. NEURAL NETWORK

To improve the separation of signal from backgrounds, an artificial NN is used. The NN is trained using approximately half of the background and signal events, the rest being used to test the network performance and derive signal cross section limits. The background events come from MC simulation for all backgrounds, except for QCD, which comes from data with either non-isolated muons or reversed electron shower shape criteria. Separate NNs were trained for the 1T and 2L channels. About 10,000 signal events and 100,000 background events were available for training in each channel, with $\pm 3\sigma$ around the signal mass peak. Background events were weighted such that the total contribution of each sample made up that expected from it in either the 1T or 2L channel.

The parameters of the NN were optimized to give the best signal significance, defined as S/\sqrt{B} , by counting the number of signal (S) and background (B) events which pass the optimal NN output cut. Sensitive variables were hypothesized based on physical reasoning of the signal vs. background processes, and then added, in order of their ability to improve the NN significance, until the NN significance failed to improve further. The final set of NN input variables are (in order of separation power):

1. M_{bb} : The invariant mass of a randomly chosen b -tagged jet pair. For the 1T channel, this variable was set to the invariant mass of the b -tagged jet and the highest p_T non- b -tagged jet.
2. $p_T(1)$: The p_T of the leading p_T jet.
3. $p_T(2)$: The p_T of the second-leading p_T jet.
4. $\Delta R(l, l)$: ΔR between the two leptons of the Z candidate.
5. $|\Delta\eta|$: The absolute value of the difference in η between the two highest p_T jets.
6. $|\Delta\phi|$: The absolute value of the difference in ϕ between the two highest p_T jets.
7. $\Delta R(Z, jet1)$: The ΔR between the Z candidate and the highest p_T jet.
8. $|\eta_Z|$: The $|\eta|$ of the Z candidate.
9. \cancel{E}_T : The missing E_T of the event (useful against $t\bar{t}$).
10. SE_T : The scalar E_T of the event.

Each NN was trained for 150 epochs, using six hidden neurons (in a single layer), and one output neuron. A separate NN was trained against each candidate signal Higgs mass. A NN was also trained against $ZZ(\rightarrow l^+l^-b\bar{b})$ as a candidate signal, ignoring it as a background. The NN outputs for $M_H = 115$ GeV in the non- b -tagged, 1T, and 2L channels, for the combined dielectron and dimuon samples, are shown in Figure 5. Additional plots of some NN input variables for $M_H = 115$ GeV, non- b -tagged, 1T, and 2L channels, are shown in Figures 2–4.

VI. RESULTS

The systematic uncertainties from jet energy scale (JES) and b -tagging are estimated by varying JES and b -tagging TRFs by ± 1 standard deviation. JES uncertainties are between 1–7% for background samples and about 1% for signals. b -tagging uncertainties are about 7% for all samples. Cross section uncertainties are included for all background processes: about 10% for $t\bar{t}$, WZ , and ZZ , 20% for Z +jets, and 30% for Z + $2b$. The uncertainties of lepton efficiency or scale factor for dilepton events is 4%. The total background uncertainty is 28%, and the signal efficiency uncertainty is 8%.

No significant excess is observed, hence ZH cross section limits are calculated, with dielectron and dimuon samples combined, using a modified frequentist approach [8]. The full NN output distributions for data, backgrounds, and signal are used as input. Limits are set including all systematic uncertainties including their correlations. In addition, the distortions of the NN output shape due to uncertainties of the JES and M_{bb} distribution shape are considered during the limit setting. The expected distributions for background are evaluated by minimizing a profile likelihood function, referencing the shape and rate of the observed distributions in the sideband regions. Table II lists the expected and observed ZH cross section limits for each Higgs mass, derived from the combination of dielectron and dimuon samples. Figure 6 shows the expected and observed ZH cross section limits, compared to the standard model expectation. The branching-ratio of $Z \rightarrow e^+e^-$ or $\mu^+\mu^-$ is taken to be 0.03366 [9].

In addition, the $ZZ(\rightarrow l^+l^-b\bar{b})$ process is searched for as a signal, and limits are set on its cross section of only ~ 5 times larger than predicted in the SM. Given the strong similarities between the $ZZ(\rightarrow l^+l^-b\bar{b})$ process and SM Higgs

signal in this channel, the observation of $ZZ(\rightarrow l^+l^-b\bar{b})$ would be an important validation of the analysis, paving the way for observation of the Higgs boson.

We thank the staffs at Fermilab and collaborating institutions, and acknowledge support from the DOE and NSF (USA), CEA and CNRS/IN2P3 (France), FASI, Rosatom and RFBR (Russia), CAPES, CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil), DAE and DST (India), Colciencias (Colombia), CONACyT (Mexico), KRF (Korea), CONICET and UBACyT (Argentina), FOM (The Netherlands), PPARC (United Kingdom), MSMT (Czech Republic), CRC Program, CFI, NSERC and WestGrid Project (Canada), BMBF and DFG (Germany), SFI (Ireland), A.P. Sloan Foundation, Research Corporation, Texas Advanced Research Program, Alexander von Humboldt Foundation, and the Marie Curie Fellowships.

-
- [1] DØ Collaboration, V. Abazov *et al.*, “The Upgraded DØ Detector,” Nucl. Instr. and Methods A565, 463 (2006).
 - [2] G. C. Blazey *et al.*, in *Proceedings of the Workshop: “QCD and Weak Boson Physics in Run II,”* edited by U. Baur, R. K. Ellis, and D. Zeppenfeld, (Fermilab, Batavia, IL, 2000) p. 47.
 - [3] T. Scanlon, ‘B-tagging and the Search for Neutral Supersymmetric Higgs Bosons at Dzero’, FERMILAB-THESIS-2006-43.
 - [4] J. Pumplin *et al.*, J. High Energy Phys. 07 (2002) 012.
 - [5] J. Campbell and K. Ellis, MCFM, *Monte Carlo for FeMtobarn processes*, <http://mcfm.fnal.gov/> .
 - [6] T. Sjöstrand *et al.*, Comp. Phys. Comm. **135**, 238 (2001).
 - [7] M.L. Mangano, M. Moretti, F. Piccinini, R. Pittau, A. Polosa, JHEP 0307:001,2003, hep-ph/0206293.
 - [8] A. L. Read, Workshop on Confidence Limits, CERN-OPEN-2000-205 (2000); T. Junk, Nucl. Instrum. Methods Phys. Res., Sect. A **434**, 435 (1999).
 - [9] W.-M.Yao *et al.* (Particle Data Group), J. Phys. G 33, 1 (2006).

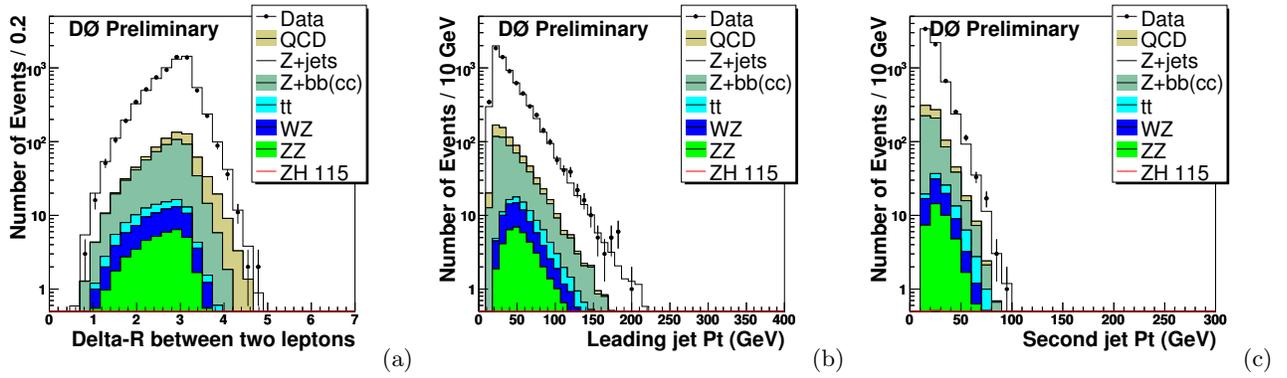


FIG. 2: A selection of important NN input variable distributions: the ΔR between the two leptons (a), the leading jet p_T (b), and the second-leading jet p_T (c), in the non- b -tagged channel, for a Higgs mass of 115 GeV.

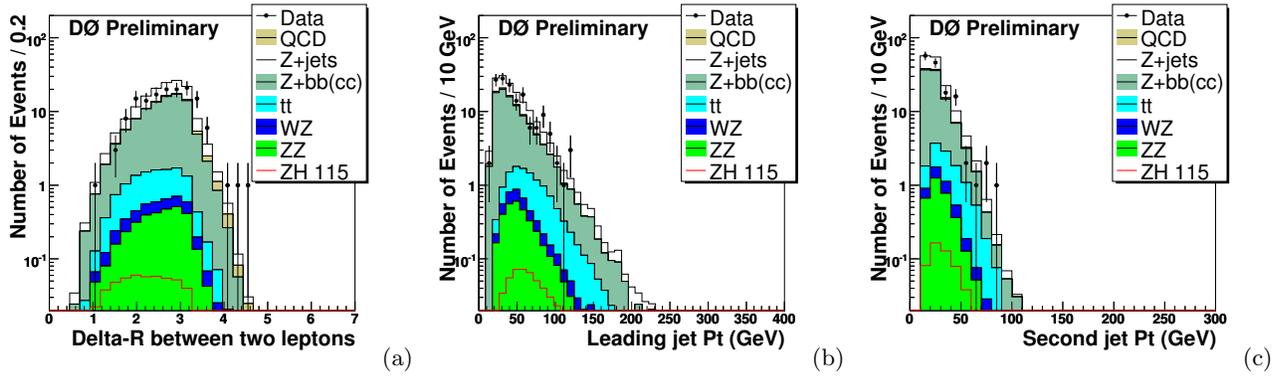


FIG. 3: A selection of important NN input variable distributions: the ΔR between the two leptons (a), the leading jet p_T (b), and the second-leading jet p_T (c), in the 1T channel, for a Higgs mass of 115 GeV.

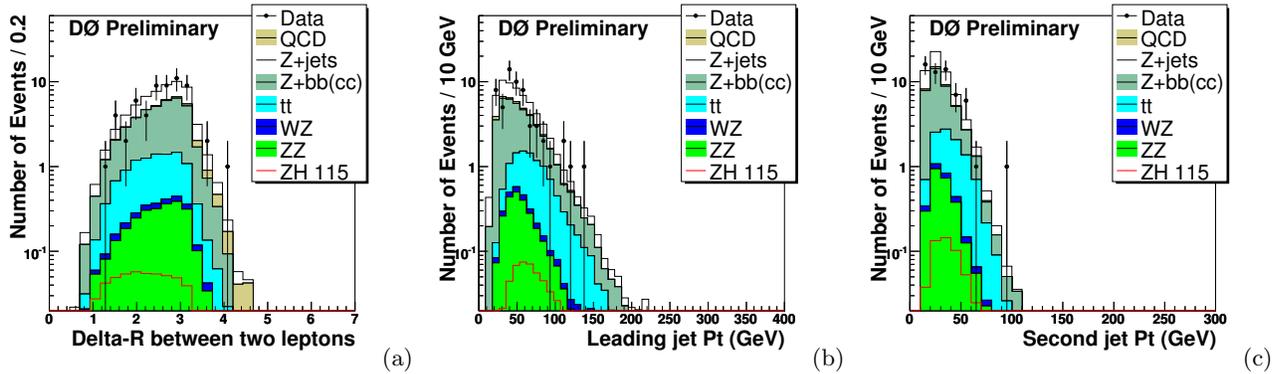


FIG. 4: A selection of important NN input variable distributions: the ΔR between the two leptons (a), the leading jet p_T (b), and the second-leading jet p_T (c), in the 2L channel, for a Higgs mass of 115 GeV.

TABLE II: The expected and observed ZH cross section limits for each Higgs mass and ZZ (and ratios to the SM cross section), derived from the combination of dielectron and dimuon samples.

m_H (GeV)	ZZ	105	115	125	135	145
Observed limit (pb)	7.38 (5.2)	1.28 (11.2)	1.39 (17.8)	1.49 (30.4)	1.38 (51.4)	1.26 (103.5)
Expected limit (pb)	9.51 (6.7)	1.69 (14.9)	1.60 (20.4)	1.34 (27.3)	1.15 (42.8)	1.07 (88.0)

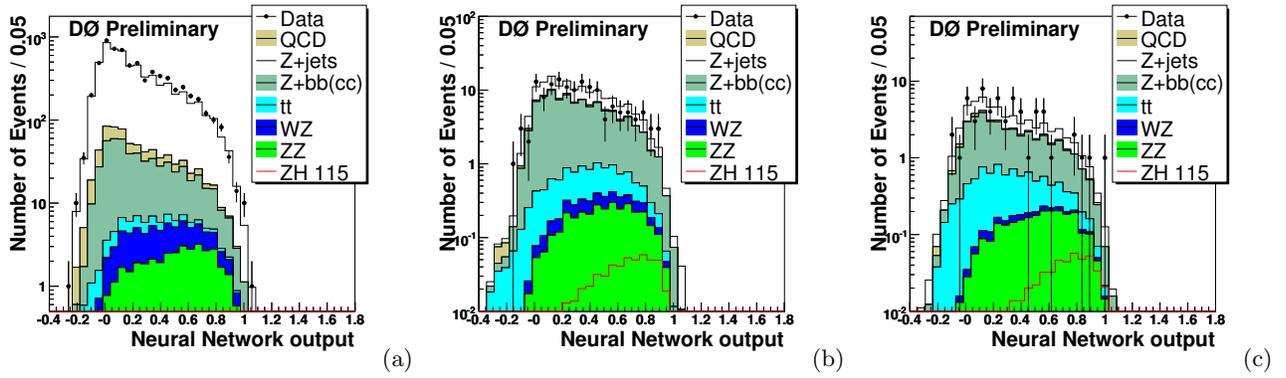


FIG. 5: The NN output distributions for non- b -tagged (a), 1T (b), and 2L (c) channels, for a Higgs mass of 115 GeV.

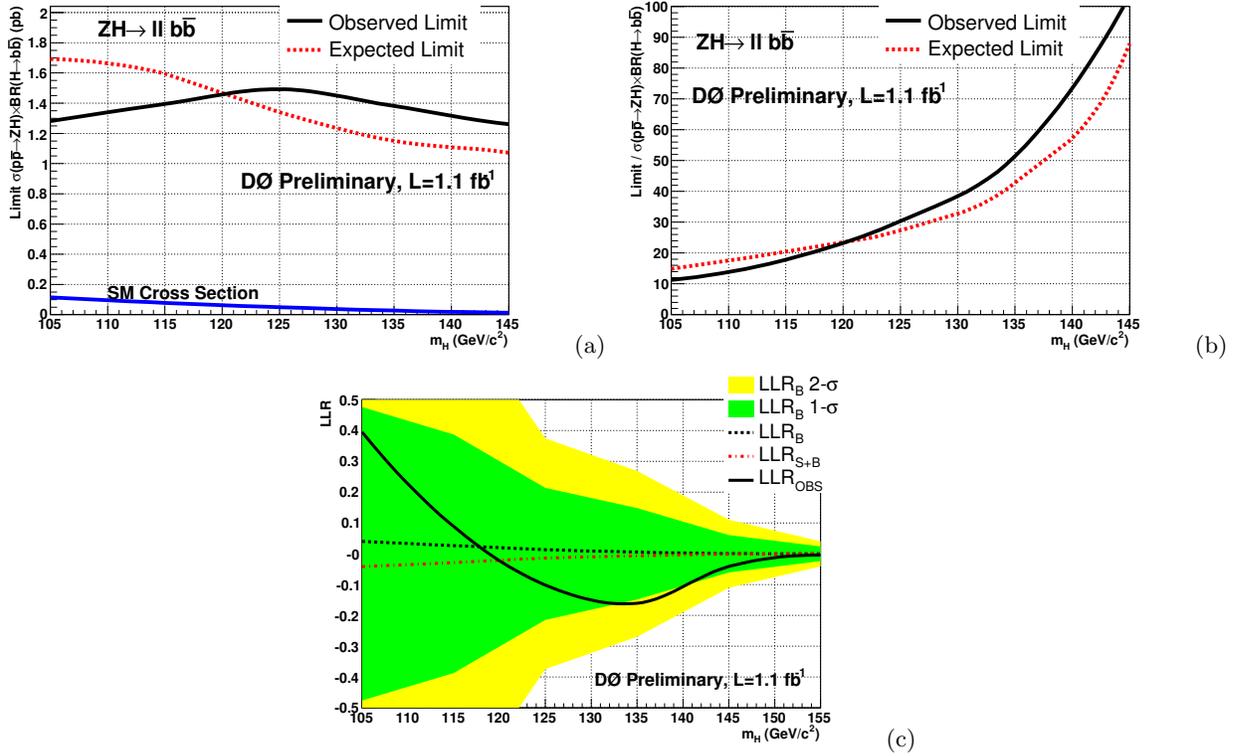


FIG. 6: The 95% confidence level upper limits on the ZH cross sections (a), the ratios to the SM cross section (b), and log-likelihood ratios (c).