



Search for the SM Higgs boson in the $\tau^+\tau^-q\bar{q}$ final state

The DØ Collaboration
URL: <http://www-d0.fnal.gov>

(Dated: August 14, 2009)

We search for the standard model Higgs boson using 3.9 fb^{-1} of Run IIb data from the DØ experiment at the Fermilab $p\bar{p}$ collider. We examine five processes in which the final state $\tau^+\tau^- + 2 \text{ jets}$ can contribute. Three involve associated production of a Higgs boson and a vector boson: $ZH \rightarrow \tau^+\tau^-b\bar{b}$, $HZ \rightarrow \tau^+\tau^-q\bar{q}$ and $HW \rightarrow \tau^+\tau^-q\bar{q}$. The vector boson fusion and gluon gluon fusion processes with $H \rightarrow \tau\tau$ can also contribute to this final state. We require that one τ decays via $\tau \rightarrow \mu\nu_\tau\bar{\nu}_\mu$ and the other τ into hadrons and ν_τ . We construct a boosted decision tree whose output characterizes the difference between signal and background, and set a 95% C.L. limit on the the measured cross sections which is 30.3 times larger than that expected in the standard model, for $M_H = 115 \text{ GeV}$. When combined with a previous DØ analysis using 1.0 fb^{-1} of data for the same signal processes we obtain this ratio to be 27.0, again for $M_H = 115 \text{ GeV}$.

I. INTRODUCTION

At the Tevatron, the standard model (SM) Higgs boson has been sought in the mass range 100 – 200 GeV using different strategies for $M_H < 135$ GeV and $M_H > 135$ GeV. In the low mass region, the dominant and subdominant Higgs boson decays are $H \rightarrow b\bar{b}$ ($\sim 90\%$) and $H \rightarrow \tau^+\tau^-$ ($\sim 10\%$), but because of the large multijet background, the dominant gluon gluon fusion production channel with $H \rightarrow b\bar{b}$ is not accessible in this region, and analyses have instead relied primarily on associated VH production ($V = W$ or Z), with W/Z boson decay modes involving e , μ or ν . To date, the low mass searches have focussed primarily on the $H \rightarrow b\bar{b}$ decay mode. In the high mass region, the increase in the branching fraction of $H \rightarrow WW$ provides sensitivity to gluon gluon fusion production.

In this note, we extend the low mass Higgs search by selecting $\tau^+\tau^-$ decays of either the Z or H , using 3.9 fb $^{-1}$ of data from Run IIb collected in the DØ detector from June 2006 to April 2009. We require that one tau decays through $\tau \rightarrow \mu\nu_\tau\bar{\nu}_\mu$ and the other decays to hadrons and a neutrino. This analysis extends a similar Run IIa analysis [1] and complements the recent result from associated WH production that used $W \rightarrow \tau\nu$, $H \rightarrow b\bar{b}$ [1] decays. Here, we select the final state $\tau^+\tau^-q\bar{q}$ in which the jets arising from the quarks are not b -tagged. This final state is populated by the following production and decay processes (the first three are referred to as VH processes):

$$p\bar{p} \rightarrow ZH, \quad Z \rightarrow \tau^+\tau^-, \quad H \rightarrow q\bar{q} \quad (\text{denoted ZH}) \quad (1)$$

$$p\bar{p} \rightarrow HZ, \quad H \rightarrow \tau^+\tau^-, \quad Z \rightarrow q\bar{q} \quad (\text{denoted HZ}) \quad (2)$$

$$p\bar{p} \rightarrow HW, \quad H \rightarrow \tau^+\tau^-, \quad W \rightarrow q\bar{q}' \quad (\text{denoted HW}) \quad (3)$$

The $\tau\tau$ + two jets final state is also sensitive to the vector boson fusion process

$$q\bar{q} \rightarrow Hq\bar{q} \quad H \rightarrow \tau^+\tau^- \quad (\text{denoted VBF}) \quad (4)$$

and to the gluon-gluon fusion process

$$gg \rightarrow H, \quad H \rightarrow \tau^+\tau^- + 2 \text{ additional jets} \quad (\text{denoted GGF}) \quad (5)$$

Reaction (1) is dominated by $H \rightarrow b\bar{b}$ decay, but the remaining reactions typically do not produce b jets. Thus, in this analysis we do not require b tagging of the jets since the τ selection provides good rejection of most background processes, and we treat the searches for all five processes in a single combined analysis.

The cross section multiplied by appropriate V , H and τ branching ratios for the VH and VBF processes are given in Table I. GGF production with Higgs decay to $\tau\tau$ is typically a factor of 50 times larger than that of the combined VH processes, but the requirement of two extra jets brings the yield for this reaction down to levels comparable to the other reactions.

Process	$m_H = 120$ GeV	$m_H = 140$ GeV
HZ	0.54	0.16
ZH	1.05	0.32
WH	1.80	0.53
VBF	1.33	0.51

TABLE I: Cross section multiplied by branching fractions in fb for $m_H = 120$ and 140 GeV. The W , Z , H and τ branching fractions are all included.

II. DATA AND MONTE CARLO EVENT SAMPLES

A. DØ Detector

The DØ detector [2, 3] contains tracking, calorimeter and muon subdetector systems. Silicon microstrip tracking detectors (SMT) near the interaction point cover pseudorapidity $|\eta| < 3$ and provide tracking and vertexing information. The central fiber tracker surrounds the SMT, providing coverage to about $|\eta| = 2$. A 1.9 T solenoid surrounds these tracking detectors. Three uranium-liquid argon calorimeters measure particle energies. The central calorimeter

(CC) covers $|\eta| < 1$, and two end calorimeters (EC) extend coverage to about $|\eta| = 4$. Intercryostat detectors provide added sampling in the region $1.1 < |\eta| < 1.4$ where the CC and EC cryostat walls degrade the energy resolution of the calorimeter. Muons are measured in stations that use scintillation counters and several layers of tracking chambers over the range $|\eta| < 2$. One such station is located just outside the calorimeters, with two more outside of 1.8 T iron toroidal magnets. Scintillators surrounding the exiting beams are used to determine the instantaneous luminosity. A three level trigger system selects events for data logging at about 100 Hz.

B. Trigger

The data for this analysis was collected in Run IIb, using the OR of single muon and muon + jet triggers, for which the integrated luminosity is 3.9 fb^{-1} after removing events of poor data quality. The trigger efficiency is parametrized as a function of muon azimuthal angle ϕ and pseudorapidity η , and used as a weight for Monte Carlo (MC) simulated events. We find the trigger efficiency to be independent of muon transverse momentum (p_T) for selections detailed below. The luminosity is computed using a high transverse momentum jet trigger which remained unrescaled throughout this run.

C. Monte Carlo samples

The MC samples for the Higgs boson signal are generated with PYTHIA [4] using the CTEQ6L1 [5] leading-order (LO) parton distribution functions (PDF). The signal cross sections are normalized to the next-to-leading order (NLO) calculation using Ref. [6]. SM background samples for $t\bar{t}$ and V +jets production are generated using ALPGEN [7] with parton showering and hadronization provided by PYTHIA. The gauge boson pair (diboson) production is generated with PYTHIA. Higgs and tau decays are simulated by HDECAY [8] and TAUOLA [9] respectively. For the combined ALPGEN/PYTHIA simulations, jets added in the fragmentation stage are removed so as to create properly exclusive or inclusive MC samples. The V +jets samples are skimmed to remove events with additional heavy flavor jets introduced by PYTHIA. Normalizations of the SM backgrounds to NLO cross sections are obtained using k -factors ranging from 1.3 to 1.7 for $t\bar{t}$, V +light jets and V +heavy flavor jets, as determined by the MCFM program [10] and data. The NLO cross sections for dibosons are also taken from MCFM.

SM backgrounds are processed with the standard DØ GEANT3 [11] detector simulation, digitization and event reconstruction programs. Events from minimum bias data runs are added to the MC events, after reweighting to account for the distribution of the observed luminosity. We apply reweighting of MC events to provide better agreement between data and MC in the distributions of the primary vertex along the beam axis and of the p_T of Z bosons.

D. Object and event preselection

The muon from the decay of one tau is found by associating hits in the muon detectors and a central track with $|\eta| < 2.0$ and $p_T > 15 \text{ GeV}$ that intersects the primary event vertex within 1.5 cm in the coordinate along the beam axis (z_{pv}). Two isolation requirements are imposed on muon candidates: (a) the sum of calorimeter transverse energies within an annular ring $0.1 \leq \mathcal{R} \leq 0.4$ must be less than 2.5 GeV, and (b) the sum of track transverse momenta, excluding the muon track, within a cone $\mathcal{R} \leq 0.5$ must be less than 2.5 GeV, where $\mathcal{R} = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2}$ is the distance in η - ϕ space, oriented relative to the muon track. Muon candidates in the MC are reweighted as a function of z_{pv} and pseudorapidity with respect to the detector center to improve the data-MC agreement.

The tau decaying to hadrons is identified with the procedure given in [12]. This employs (i) calorimeter clusters found using a simple cone algorithm using $\mathcal{R} = 0.3$; (ii) energy in an annular cone $0.3 \leq \mathcal{R} \leq 0.5$; (iii) electromagnetic (EM) subclusters, and (iv) up to three tracks with $p_T > 1.5 \text{ GeV}$ within $\mathcal{R} < 0.5$ of the tau direction (defined by the visible decay products) with invariant mass consistent with that expected for tau decays. Neural networks are used to identify each of three tau types, motivated by the characteristics of the tau decay modes (1) $\tau^\pm \rightarrow \pi^\pm \nu$, (2) $\tau^\pm \rightarrow \pi^\pm \pi^0 \nu$ and (3) $\tau^\pm \rightarrow \pi^\pm \pi^\pm \pi^\mp (\pi^0) \nu$. The transverse momentum of the tau, p_T^τ , is computed from the calorimetric energy within the τ cone, improved by the measured track momenta. We require $p_T^\tau > 15 \text{ GeV}$ for Types 1 and 2 and $> 20 \text{ GeV}$ for Type 3. Tau candidates are required to have $|\eta| < 2$ measured with respect to the center of the detector. We require the sum of the transverse momenta of the τ -associated tracks, p_T^{trk} , to exceed 7 (15) GeV for tau type 2 (3). In addition, for Type 3 taus, we require at least one track with $p_T > 7 \text{ GeV}$. A Type 3 tau must have at least two reconstructed tracks; if only two are present we require that they be of the same charge. A Type 1 tau candidate is required to have $p_T^{\text{trk}}/E_T^\tau > 0.7$ to reduce misidentification of type 2 taus within the region between

calorimeter cryostats and $p_T^{\text{trk}}/E_T^\tau < 1.5$ to eliminate contamination from cosmic ray muons. The direction of the tau must extrapolate to the primary vertex along the beam direction to within 1.5 cm. To be considered as a tau candidate for a signal process, the neural network output variable, NN_i , for tau Type i is required to exceed 0.9 for Types 1 and 2 and 0.95 for Type 3.

Jets are reconstructed using a cone algorithm [13] with a cone size of $\mathcal{R} = 0.5$. The jet energy is corrected for detector effects to the particle level and corrected for missing neutrino energy in the case of an indication of semi-muonic decays within the jet. Jet objects in the MC are corrected for differences between data and MC in identification efficiency and energy resolution. Jets are required to have $|\eta| < 3.4$ and $p_T > 20$ GeV. The missing transverse energy is calculated as $\cancel{E}_T = \sqrt{(\sum_j E_{x,j})^2 + (\sum_j E_{y,j})^2}$, where $E_{x,j}$ ($E_{y,j}$) is the energy in the j -th calorimeter cell multiplied by $\cos \phi_j$ ($\sin \phi_j$) with ϕ_j the azimuthal angle of the j -th cell, with standard energy corrections applied to all observed objects, including any observed muons. The observed \cancel{E}_T is ascribed to the neutrinos from the two tau decays in proportion to the transverse and longitudinal momenta of their visible decay products, and is used to calculate the $\tau\tau$ invariant mass and $\mu\nu$ or $\tau\nu$ transverse masses.

With the above object definitions, we select signal events with the requirements:

- Only one isolated muon
- At least one hadronically decaying tau candidate separated from the muon candidate by $\Delta\mathcal{R} > 0.5$. If more than one tau candidate exists, we select the one with the highest p_T^τ
- Oppositely charged muon and tau
- At least two good jets, separated from both μ and τ by $\Delta\mathcal{R} > 0.5$
- No good electron with $p_T > 12$ GeV.

E. Estimation of background from multijet events

Additional event selections, orthogonal to the main signal preselection requirements above, use data to provide an estimate of the multijet (MJ) background. The primary MJ sample is selected as above, except that the muon is required to fail either the calorimeter or track isolation requirement, and the tau must have a NN output in the range $0.3 \leq \text{NN} \leq 0.8$. For the purpose of estimating the systematic uncertainty of the contribution from the MJ background, two alternate MJ samples are constructed in which either the μ anti-isolation or reduced τ NN output cuts (but not both) are applied.

The primary MJ sample contains dominantly QCD multijet processes. We estimate that the multijet processes comprise about 95% of this sample, and that the remainder are contributed by other SM processes such as $t\bar{t}$, V +jets or diboson production. We use this MJ sample to train and characterize the final event selection based on a multivariate technique discussed below.

The MJ background is determined from two data samples: the signal selection described in the previous section and the MJ selection. The one derived from the signal sample after subtraction of MC-generated SM processes is named the SG sample and that derived from the MJ sample after subtracting the SM MC contributions is termed the BG sample. We split both SG and BG samples into same-sign (SS) and opposite sign (OS) subsamples, based on the charges of the muon and tau candidate. The resulting numbers of events are termed N_C^Q , where $C = (\text{SG}, \text{BG})$ and $Q = (\text{OS}, \text{SS})$. The number of MJ background events in the signal sample is then calculated as $N_{\text{SG}}^{\text{OS}} = f \times N_{\text{SG}}^{\text{SS}}$, where the MJ normalization factor is measured in the BG sample using $f = N_{\text{BG}}^{\text{OS}}/N_{\text{BG}}^{\text{SS}}$. The scale factor f is determined separately for each tau type, and is observed to be close to unity and independent of p_T^μ , p_T^τ , jet p_T , dijet p_T and jet η . We use the primary MJ sample defined above as the default description of the MJ background for the multivariate analysis, and use the two alternate MJ samples to assess the systematic uncertainty on this component of background.

F. Preselection sample

The number of data events, expected contributions from the background processes and the expected number of signal events at $M_H = 115$ GeV after preselection are given in Table II.

Data	Σ Bknd	$t\bar{t}$	W +jets	Z +jets	DB	MJ
433	439.9	66.7	81.5	222.7	10.2	80.7

Process	ZH	HZ	HW	VBF	GGF
Event yield	0.11	0.23	0.72	0.12	0.15

TABLE II: Event yields at the preselection level for data and various backgrounds, and for the signal processes with $m_H = 115$ GeV. ‘DB’ refers to diboson production processes, ‘MJ’ to multijets, ‘hp’ to heavy partons and ‘lp’ to light partons.

III. MULTIVARIATE ANALYSIS

After preselection, the expected backgrounds exceed the expected signal by a large factor, so additional discrimination is needed. No single discriminating variable is found that reduces the backgrounds sufficiently, and therefore we employ a boosted decision tree (BDT) [14] multivariate analysis that exploits the correlations among such variables. We have verified that signal/background discrimination and the expected Higgs cross section limits using BDTs are superior to those obtained with neural network multivariate classifiers in a study using the Run IIa data. Since the five signal processes and the four main classes of background ($t\bar{t}$, W +jets, Z +jets and multijets) have different kinematic characteristics, we use separate BDTs to characterize and discriminate each of the four background categories from each of the first four signal processes. After the preselection cuts requiring two jets, the MC samples for GGF are too small to allow independent BDT training. Nevertheless, when calculating the limits on the Higgs boson cross section, we add the GGF signal contributions to those from VBF since these reactions both contain non-resonant di-jets. We denote these as BDT(S,B), where S corresponds to the now four reactions for signal in Eqs. 1 – 4 and B refers to $t\bar{t}$, Wj , Zj and MJ. We further build the BDT(S,B) for low postulated M_H (105, 115 or 125 GeV) and high M_H (135, 145 GeV), yielding 32 such BDTs in all.

We have examined many kinematic variables that can discriminate at least one of the signals from one of the backgrounds. From these, we select those variables with good discrimination and good agreement between data and expected backgrounds. All BDTs use the same set of 17 input variables, given in Table III. We have checked that adding variables with little discriminating power for a particular BDT(S,B) does not cause deterioration in the signal to background separation.

Name	Description
p_T^μ	p_T of the μ candidate
p_T^{j1}	p_T of the highest p_T jet
\cancel{E}_T	missing transverse energy
$M_{\tau\tau}$	invariant mass of the two tau candidates
M_{jj}	invariant mass of the two highest E_T jets
ΔR_{jj}	distance in η - ϕ space between the two highest E_T jets
M_T^μ	transverse mass calculated from the μ and \cancel{E}_T
M_T^τ	transverse mass calculated from the τ and \cancel{E}_T
H_T	scalar sum of p_T of all jets with $ \eta < 2.5$
S_T	scalar sum of the p_T of the μ , τ , two leading jets and \cancel{E}_T
V_T	magnitude of the vector sum of the objects included in S_T
C	centrality: the ratio of H_T to the scalar sum of the energies of all jets
$A(\cancel{E}_T, \cancel{H}_T)$	asymmetry between \cancel{E}_T and \cancel{H}_T , $(\cancel{E}_T - \cancel{H}_T)/(\cancel{E}_T + \cancel{H}_T)$
$\Delta\phi(\cancel{E}_T, T_T)$	azimuthal difference between \cancel{E}_T and the missing transverse momentum of all tracks
$\Delta\phi(\mu, \tau)$	azimuthal difference between μ and τ
$\Delta\phi(\cancel{E}_T, jets)$	minimum azimuthal difference between \cancel{E}_T and any jet
\mathcal{A}	aplanarity [15] constructed from the four-vectors of the μ , τ , and all jets

TABLE III: Input variables for the BDTs.

Figure 1 shows a comparison of data to MC for some of the BDT input variables. The agreement for variables employed in the BDTs is generally good, but there is some discrepancy at small values of leading jet p_T . We have checked that this effect is not caused by the muon + jets triggers.

Figure 2 shows representative BDT output plots for the $HW \rightarrow \tau^+\tau^-q\bar{q}'$ signal at low M_H trained against the $t\bar{t}$, W +jets and MJ backgrounds. The BDT output distributions show good agreement between data and the expected backgrounds, and tend to show substantial concentrations of events from the background used in the training at negative BDT output, and a tendency for the events from the signal process used to be at positive BDT output. Similar behavior and good agreement between data and MC is seen for the other three signals, although a BDT trained using one signal or background does not necessarily discriminate well against a different signal or background.

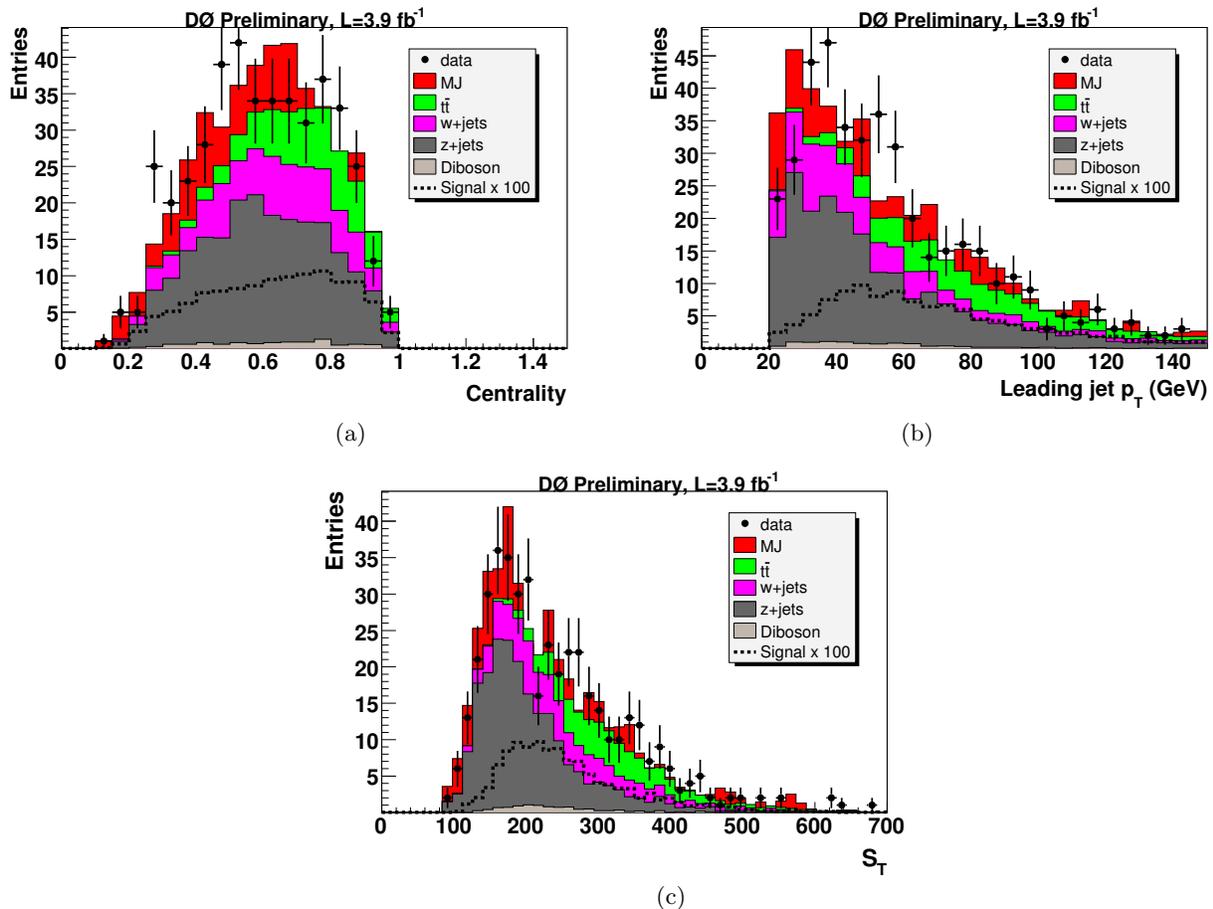


FIG. 1: Representative distributions after preselection showing comparisons of data and MC for representative variables used in BDT training: (a) centrality; (b) p_T^{j1} ; and (c) S_T . The sum of all signal processes is shown as the dashed line in (b), (c) and (d) for $m_H=115$ GeV scaled up by 100.

This suggests combining these individual BDTs for each background by taking the maximum value of the BDTs over the four signal processes:

$$\text{BDT}^B = \max[\text{BDT}(\text{ZH}, B), \text{BDT}(\text{HZ}, B), \text{BDT}(\text{HW}, B), \text{BDT}(\text{VBF}, B)] \quad (6)$$

for $B = t\bar{t}$, W + jets and MJ. We then make a final selection of events by requiring the BDT^B value to exceed -0.2 , -0.2 and 0.0 for $B = t\bar{t}$, W + jets and MJ respectively. These values were optimized by maximizing the expected significance of the signal. They reduce the backgrounds considerably, with relatively little loss of signal. The distributions of BDT^B for the $t\bar{t}$, W + jets and MJ backgrounds before these final selections are shown in Fig. 3. The final event yields for data and backgrounds are shown in Table IV.

	Data	Σ bknd	$t\bar{t}$	W +jets	Z +jets	DB	MJ	Signal
low M_H	37	32.2	0.66	0.116	27.7	1.09	2.32	0.59
high M_H	13	13.02	0.27	0.98	11.13	0.32	0.83	0.19

TABLE IV: Event yields after the selections on BDT^B for $B = t\bar{t}$, W + jets and MJ, both for high and low mass Higgs bosons. Here ‘hp’ refers to heavy partons (b , c), ‘lp’ to light partons, ‘DB’ to dibosons and ‘MJ’ to multijet. The signal yields are for the sum of reactions (1) – (5) for $M_H = 115$ GeV (low M_H) and $M_H = 135$ GeV (high M_H).

The BDTs trained to discriminate the four signals against the Z + jets background have good agreement between data and the MC prediction, but show poorer separation of signal and background than for the other backgrounds. The Z + jets background has a resonant tau pair and is therefore more signal-like than the other backgrounds. Figure

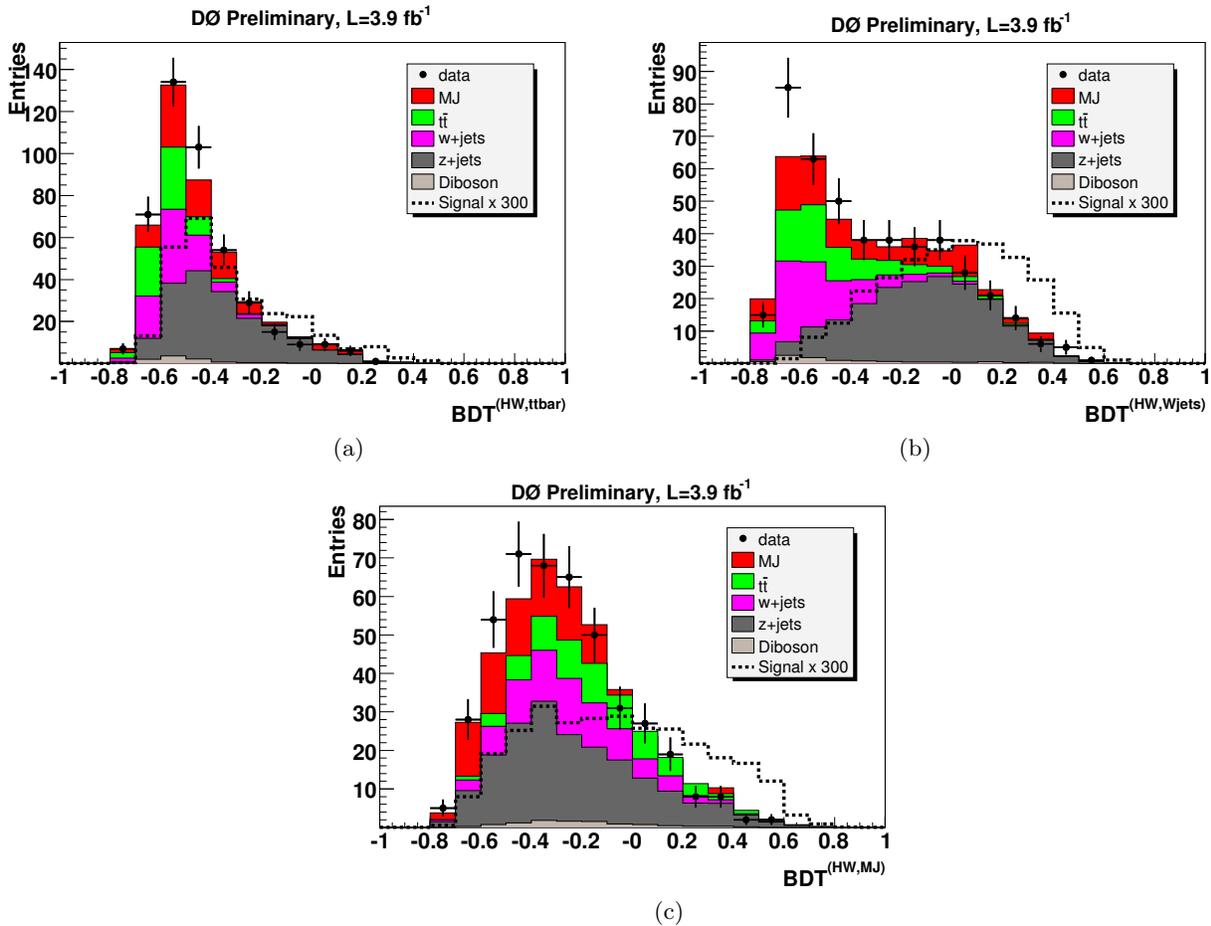


FIG. 2: The output distributions of the BDTs for the low Higgs boson mass $HW \rightarrow \tau^+\tau^-q\bar{q}'$ signal trained against (a) $t\bar{t}$, (b) $W + \text{jets}$ and (c) MJ backgrounds. The signal curves are shown for $m_H=115$ GeV, scaled up by 300.

4 shows the low mass Higgs boson signal BDT output distributions for each of the signals against the $Z + \text{jets}$ background, before imposing the selections on the BDT^B values. As indicated above, the GGF signal is included here with the VBF signals. After the final BDT^B selections, we form the final variable $\text{BDT}^{Z\text{jets}}$,

$$\text{BDT}^{Z\text{jets}} = \sum_{S=1}^5 w_S \text{BDT}(S, Z\text{jets}) / \sum_S w_S \quad (7)$$

for setting limits, where $S = (ZH, HZ, HW, \text{VBF}, \text{GGF})$ and w_S is the product of the cross section, the branching fraction and the selection efficiency of signal process S . Figure 5 shows the distribution in $\text{BDT}^{Z\text{jets}}$ for both low and high mass Higgs bosons. These $\text{BDT}^{Z\text{jets}}$ distributions are taken as the final test statistic for setting limits on the cross section for the production of Higgs bosons.

IV. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties are assigned in two categories. Those that modify the overall normalization of signal or backgrounds but do not modify the shape of the $\text{BDT}^{Z\text{jets}}$ test statistic distributions are termed “flat” uncertainties. For those sources whose variation within their systematic uncertainties modify the shape of $\text{BDT}^{Z\text{jets}}$ distribution, we provide these relative shapes to the limit setting algorithm (“shape dependent” uncertainties).

The following uncertainties are taken to be flat, with the values quoted:

- Integrated luminosity (6.1%) [16]
- Muon identification, track match and isolation efficiencies (5%)

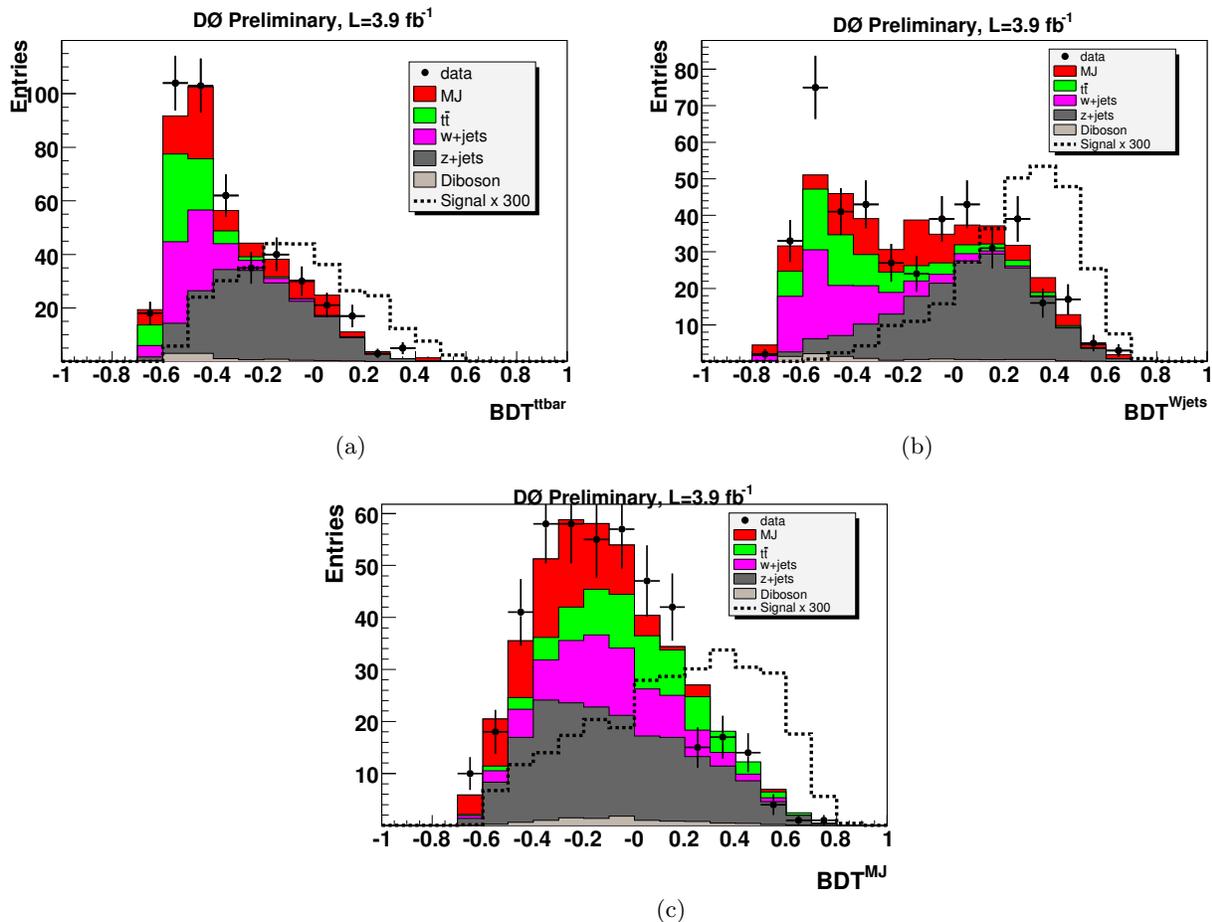


FIG. 3: The maximum BDT output, BDT^B , over all low mass Higgs boson signals for the backgrounds (a) $t\bar{t}$, (b) $W + \text{jets}$ and (c) MJ backgrounds. The sum of all signals is shown for $m_H=115$ GeV, scaled up by 300.

- Trigger efficiency (5%)
- Tau neural network efficiency (8.9%), taken as the weighted average over the three tau types [12]
- Jet identification and reconstruction efficiency (3%) is determined by changing the identification and reconstruction corrections by ± 1 standard deviation (sd) from their nominal values
- Jet energy scale correction (7.5%) is determined by changing the jet energy scale corrections by ± 1 sd from nominal values
- Jet energy resolution correction (5%) is determined by changing the jet energy resolution corrections by ± 1 sd from nominal values
- The modelling uncertainty of jets at low jet p_T (10%) is determined by reweighting the multijet background as a function of jet p_T
- Uncertainties on the theoretical cross sections for the SM processes are taken to be 6% for SM Higgs boson production, 10% for $t\bar{t}$ and single top production, 6% for W/Z +light quark jet production, 20% for W/Z +heavy quark jet production and 7% for diboson production.

Two sources of systematic uncertainty in particular cause changes in the shape of the $BDT^{Z_{\text{jets}}}$ distribution. The multijet background enters into the construction of the BDTs, and can in principle alter their shapes. As discussed in Section II.E, two alternates to the standard MJ background samples were constructed. To evaluate the shape dependence arising from this source, the alternative samples were used to construct the final $BDT^{Z_{\text{jets}}}$ distributions. These are compared with those using the nominal MJ selection to obtain an overall scale uncertainty of 17% and a shape

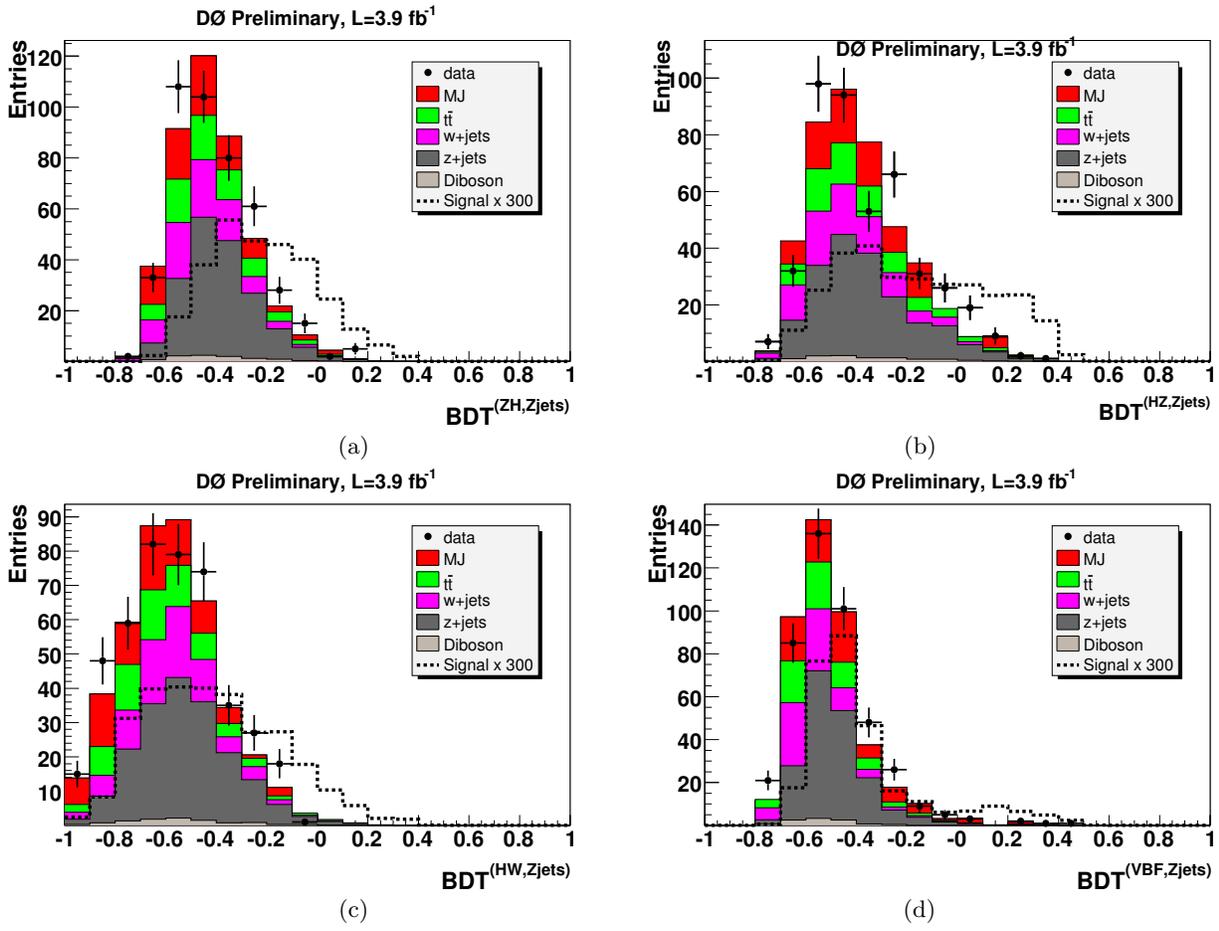


FIG. 4: The output distributions of the BDTs trained against $Z + \text{jets}$ background for the low Higgs boson mass signals before the cut on BDT^B , for (a) $ZH \rightarrow \tau^+\tau^-b\bar{b}$, (b) $HZ \rightarrow \tau^+\tau^-q\bar{q}$, (c) $HW \rightarrow \tau^+\tau^-q\bar{q}'$ and (d) VBF. The signal curves are shown for $m_H=115$ GeV, multiplied by 300.

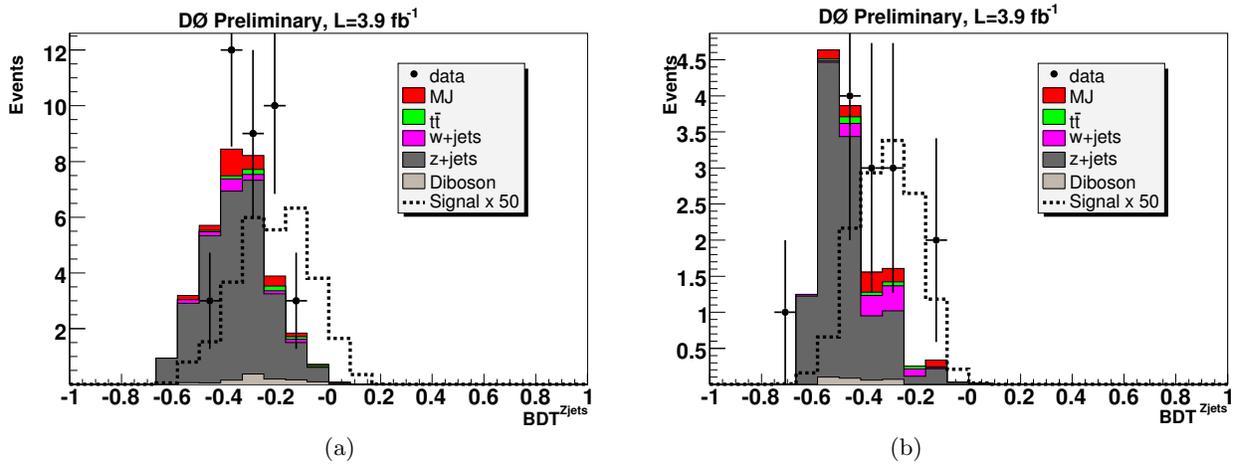


FIG. 5: The weighted average output distributions of the BDTs trained of the $Z + \text{jets}$ background against signals, BDT^{Zjets} , for (a) low mass Higgs and (b) high mass Higgs hypotheses. The sum of all signals are shown as the dashed line for (a) $m_H=115$ GeV and (b) $m_H=145$ GeV, multiplied by 50.

dependent uncertainty taken as the absolute difference in the final $\text{BDT}^{Z_{\text{jets}}}$ between the nominal MJ background and the two alternatives. The shape dependent uncertainty from the choice of PDF is obtained from the variations among the CTEQ6 pdf eigenvector sets [5].

V. LIMIT CALCULATION

The upper limits on the Higgs boson production cross section are calculated using the modified frequentist method [18] and implemented in the Collie [19] framework. The test statistic is a binned negative Poisson log-likelihood ratio (LLR) for the $\text{BDT}^{Z_{\text{jets}}}$ variable (Eq. 7) with a bin size of 0.083. The confidence levels CL_{s+b} (CL_b) give the probability that the signal + background (background) LLR value from a set of 50,000 simulated pseudo-experiments is less likely than that observed, at the quoted confidence level. The hypothesized signal cross sections are scaled up from their SM values until the value of $\text{CL}_s = \text{CL}_{s+b}/\text{CL}_b$ reaches 0.05 which defines the limit cross sections at 95% C.L. In the calculation, all contributions to systematic uncertainty are fitted, subject to the constraints given by their estimated values, to yield the best fit (‘profiling’). Correlations of systematic uncertainties among signal and/or background processes are accounted for in the minimization.

The expected and observed upper limits for the current data are shown in Table V and in Fig. 6(a). At $M_H = 115$ GeV, we set a limit on SM Higgs boson production at a factor of 30.3 times larger than that predicted in the SM for the five signal processes of Eqs. 1 – 5. The expected LLRs for signal plus background and background only (with ± 1 sd and ± 2 sd bands for LLR_b), and the observed LLR are shown in Fig. 6(b).

M_H (GeV)	Cross section ratio expected	Cross section ratio observed
105	15.1	25.4
115	18.8	30.3
125	30.5	49.8
135	36.4	73.9
145	71.2	174.9

TABLE V: Expected and observed 95% upper limits on the cross section ratios relative to those predicted in the SM, from this analysis.

The Run IIa analysis [1] obtained limits on SM Higgs boson production for the same set of signal processes using a neural network technique using 1.0 fb^{-1} of data. Combining the results from that analysis with the current results for Run IIb shown in Table V, taking into account the correlations among the uncertainties, we obtain the combined limits shown in Table VI and in Fig. 7 for a total accumulated luminosity of 4.9 fb^{-1} .

M_H (GeV)	Cross section ratio expected	Cross section ratio observed
105	13.4	21.9
115	15.9	27.0
125	24.4	41.9
135	32.2	37.6
145	61.4	86.0

TABLE VI: Expected and observed 95% upper limits on the cross section ratios relative to those predicted in the SM for the combination of the Run IIb data set with that from Run IIa [1].

VI. CONCLUSION

We have set limits on SM Higgs boson production for masses between 105 and 145 GeV from a search using $\tau\tau$ jet jet final states. For a Higgs boson mass of 115 GeV, the ratio of the limit on the cross section at 95% C.L. to that expected in the SM is 30.3 for the Run IIb analysis alone, and in combination with the Run IIa data this ratio is 27.0.

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); CNPq, FAPERJ, FAPESP and FUNDUNESP (Brazil); DAE and DST (India); Colciencias (Colombia); CONACyT (Mexico); KRF and KOSEF

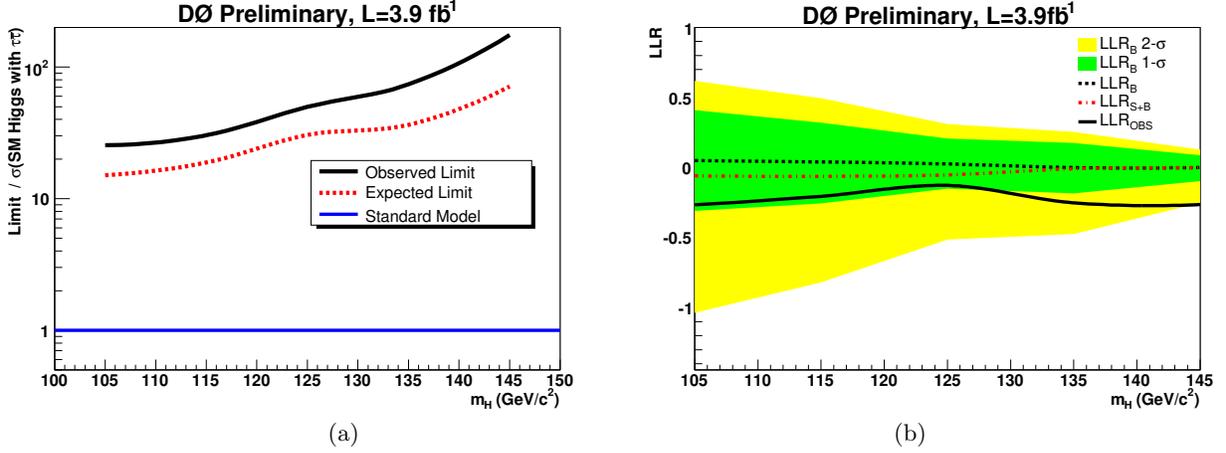


FIG. 6: For the 3.9 fb^{-1} Run IIb analysis presented here, (a) expected and observed limits on the cross section of five SM Higgs signal production processes of Eqs. 1 – 5, relative to that predicted in the SM at the 95% C.L.; (b) LLR for the expected background only (black dashed line) and expected signal plus background (red dash-dotted line), with ± 1 sd (green) and ± 2 sd (yellow) bands around the LLR_b expectation, and the observed LLR (solid line).

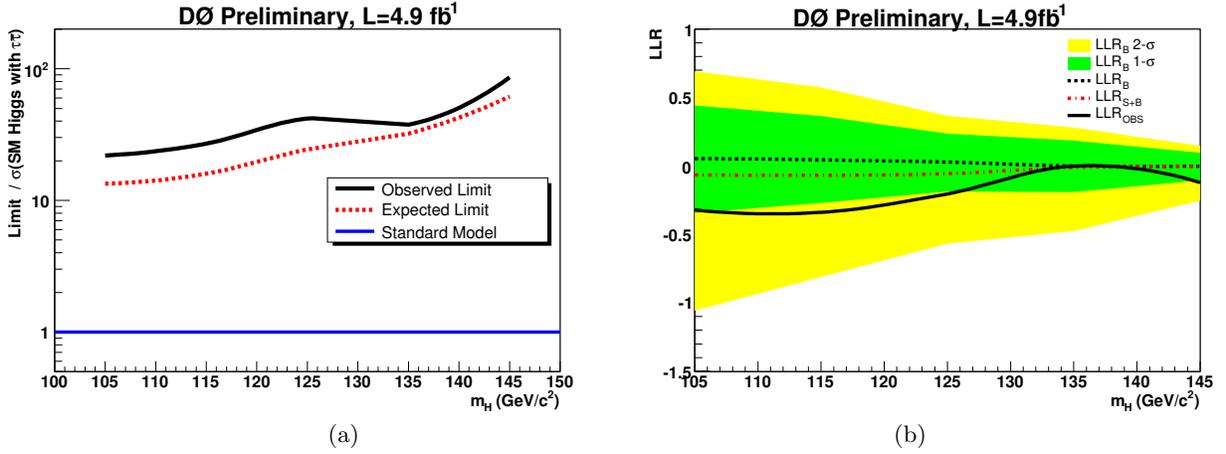


FIG. 7: For the 4.9 fb^{-1} combination of the Run IIa and Run IIb analyses, (a) expected and observed limits on the cross section of five SM Higgs signal production processes of Eqs. 1 – 5, relative to that predicted in the SM at the 95% C.L.; (b) LLR for the expected background only (black dashed line) and expected signal plus background (red dash-dotted line), with ± 1 sd (green) and ± 2 sd (yellow) bands around the LLR_b expectation, and the observed LLR (solid line).

(Korea); CONICET and UBACyT (Argentina); FOM (The Netherlands); STFC (United Kingdom); MSMT and GACR (Czech Republic); CRC Program, CFI, NSERC and WestGrid Project (Canada); BMBF and DFG (Germany); SFI (Ireland); The Swedish Research Council (Sweden); CAS and CNSF (China); and the Alexander von Humboldt Foundation (Germany).

-
- [1] V.M. Abazov *et al.*, Phys. Rev. Lett **102**, 251801 (2009).
[2] V.M. Abazov *et al.*, (D0 Collaboration), Nucl. Instrum. Methods Phys. Res A **565**, 463 (2006).
[3] V.M. Abazov *et al.*, (D0 Collaboration), Nucl. Instrum. Methods Phys. Res A **552**, 372 (2005).
[4] T. Sjöstrand *et al.*, Computer Phys. Commun. **135**, 238 (2001); T. Sjöstrand *et al.*, arXiv:(hep-ph) 0108264 (2001).

- [5] J. Pumplin *et al.*, J. High Energy Phys. **07**, 12 (2002).
- [6] T. Hahn *et al.*, arXiv: (hep-ph) 0607308v2 (2006).
- [7] M. Mangano *et al.*, J. High Energy Physics **0307**, 001 (2003).
- [8] A. Djouadi, J. Kalinowski, M. Spira, arXiv: (hep-ph) 9704448 (1997).
- [9] S. Jadach *et al.*, Computer Phys. Commun. **76**, 361 (1993).
- [10] J. Campbell and K. Ellis, <http://mcfm.fnal.gov/>.
- [11] R. Brun and F. Carminati, CERN Program Library Long Writeup W5013, 1993 (unpublished), version V3.21.
- [12] V.M. Abazov *et al.* (D0 Collaboration), Phys. Rev. D **71**, 072004 (2005).
- [13] G. 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-Pub-00/297 (2000).
- [14] L. Breiman *et al.*, *Classification and Regression Trees* (Wadsworth, Stamford CT, 1984); D. Bowser-Chao and D.L. Dzialo, Phys. Rev D **47**, 1900 (1993).
- [15] V.D. Barger and R.J.N. Phillips, *Collider Physics* (Addison- Wesley, Reading MA, 1987), p.281.
- [16] T. Andeen *et al.* FERMILAB-TM-2365 (2007).
- [17] V.M. Abazov *et al.* (DØ Collaboration), Phys. Rev. Lett. **101**, 062001 (2008).
- [18] A. Read, J. Phys. G: Nucl. Part. Phys. **28**, 2693 (2002); T. Junk, Nucl. Instrum. Methods A **434**, 435 (1999).
- [19] W. Fisher, FERMILAB-TM-2386-E (2006).