Search for the Higgs Boson in the WH → lνbb and WH → WWW → lνjjjj Channels

Stephanie Hamilton

Michigan State University

The DØ Collaboration, Fermi National Accelerator Laboratory

August 9, 2011

Abstract

Throughout the summer of 2011, work continued on the analysis of the associated production of the Higgs boson. Our particular analysis was the WH → lν*bb* decay channel. This channel has the highest branching ratio of the low-mass Higgs decay channels and thus is one of the most sensitive channels to analyze, resulting in a solid framework and a good foundation. Work was also done on the WH → WWW → lνjjjj decay channel. This channel is unique within DØ because there are only a few people working on it, all whom are summer students. This paper explains ongoing efforts to process data and Monte Carlo (MC) samples, model data correctly, and utilize the output of multivariate training to effectively distinguish between signal and background events and perform a search for the Higgs Boson. It also discusses new variables that were added to the search and how they impacted the overall analysis.

Table of Contents

1 Introduction

The Higgs boson is the only particle predicted by the Standard Model (see Figure 1.1) that has not yet been observed. The Standard Model describes all fundamental particles and their interactions with each other, describes the mediation of forces using gauge bosons, and predicts particles that have not yet been seen. The discovery of the Higgs would explain spontaneous electroweak symmetry breaking and the masses of most particles[1]. As a result, the SM Higgs has become the focus of many searches at Fermi National Accelerator Laboratory (Fermilab) and the European Center for Nuclear Research (CERN). This paper discusses the ongoing analysis at DØ in the WH → lν*bb* and WH → WWW → lνjjjj channels.

Higgs production at the Tevatron is dominated by two major production mechanisms. The first is gluon fusion (Figure 1.2), in which two gluons collide with sufficient energy to form a Higgs. The second is by associated production (Figure 1.3), in which a W or Z boson is produced with the Higgs. The cross section for gluon fusion is 10 times larger than that for WH production, and the cross section for ZH production is smaller still (Figure 1.4). The Higgs can then follow one of approximately 30 decay paths. The likelihood of the Higgs decaying by a certain channel is known as the branching ratio and varies with the mass of the Higgs. Because the mass is not known, it must be determined by experimental means. A low mass Higgs (115 GeV to 135 GeV) prefers to decay into a $b\bar{b}$ pair, i.e. a bottom and an antibottom quark. A high mass Higgs (135 GeV to 200 GeV[2]) would prefer to decay into two W bosons. The branching ratios are depicted in Figure 1.5 and an updated limit plot from July 27, 2011 constraining the mass of the Higgs from the Tevatron can be seen in Figure 1.6. The WH → lν*bb* channel continues to be the favored decay channel for a low mass Higgs because requiring the decay of the W into a lepton and a neutrino greatly reduces $b\bar{b}$ background rates. Figure 1.7 depicts the Feynman diagram for a WH → lν*bb* decay.

Another decay channel of the SM Higgs is the WH → WWW → lνjjjj channel, as shown in Figure 1.5. H→WW is the preferred decay mode of a high-mass Higgs; gg→H→WW is the most sensitive channel at high mass, but WH→WWW→lνjjjj is a complimentary channel that can enhance combined sensitivity. Requiring four jets in the end state greatly reduces several types of background that include only one or two jets. Figure 1.8 details the decay of WH → WWW → lνjjjj.

The WH channel has a very low signal to background ratio and thus requires very high integrated luminosities and good cuts or variables to discriminate signal from background. This paper discusses the continued analyses of the WH → lν*bb* and the WH → WWW → lνjjjj decay channels at $\sqrt{s}$ = 1.96 TeV at the Tevatron. It examines new variables added and their effects on the search. It also discusses some of the analysis done on the WH → WWW → lνjjjj decay channel, as I was placed on that project towards the end of the summer to assist my coworkers Alex Abbinnante, Tony Podkowa, and Youssef Mobarak.

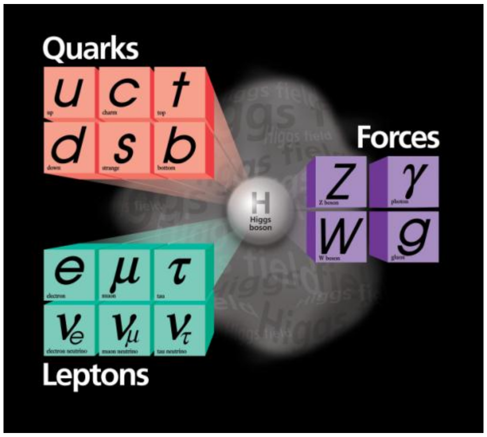


Figure 1.1. The Standard Model

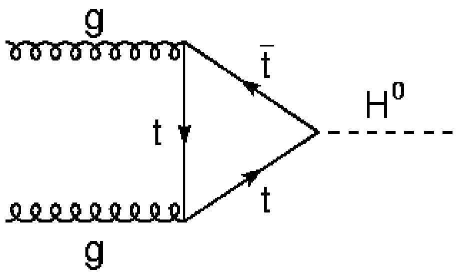


Figure 1.2. Gluon Fusion

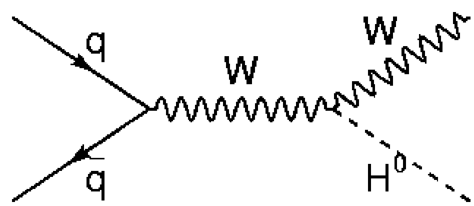


Figure 1.3. Associated Production

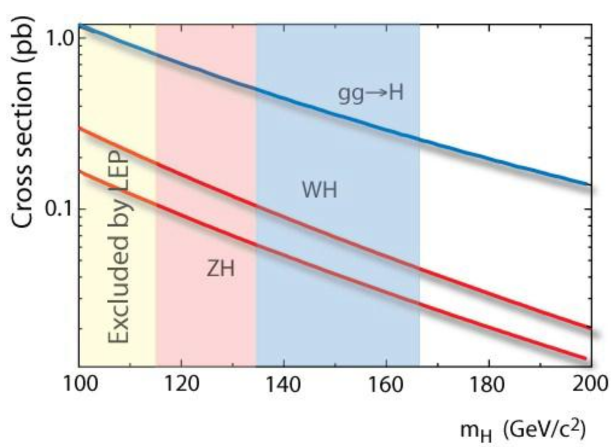


Figure 1.4. Production Cross Sections of the SM Higgs Boson

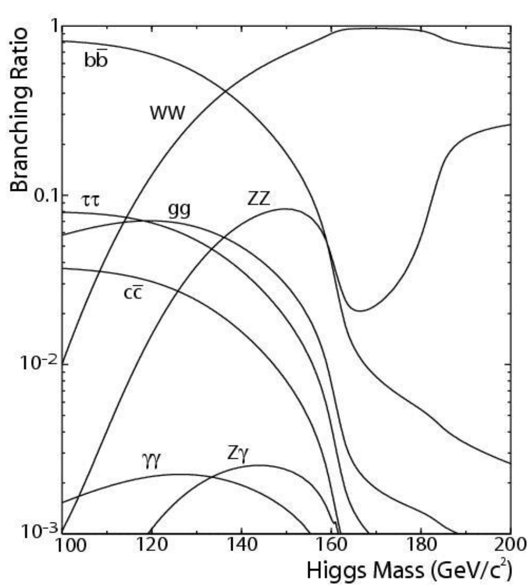


Figure 1.5. Branching Ratios of the SM Higgs boson

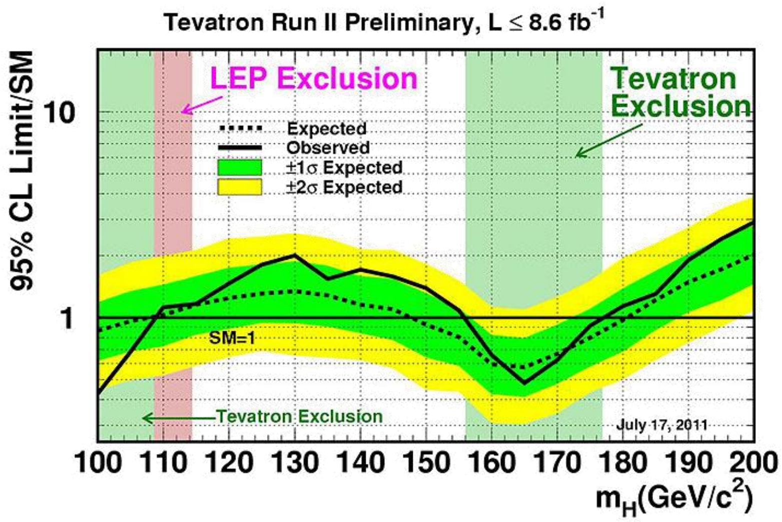


Figure 1.6. Limits on the Mass of the Higgs at the Tevatron

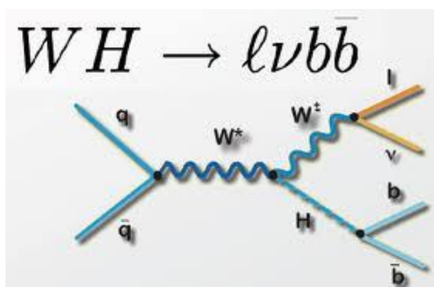


Figure 1.7. Associated production of the Higgs boson: WH → lν*bb*

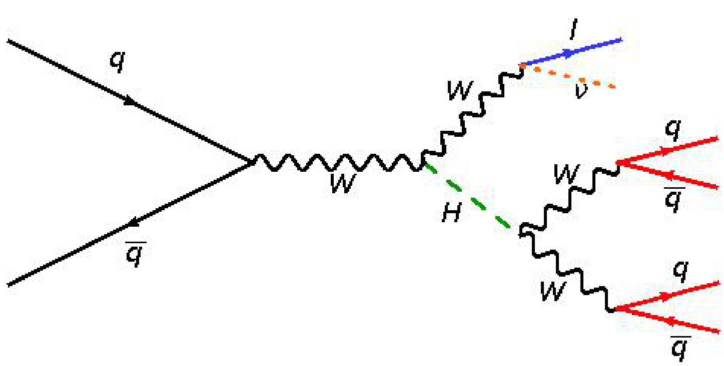


Figure 1.8. Associated Production of the Higgs boson: WH → WWW → lνjjjj

2 Methods and Materials

2.1 The DØ Detector

The DØ detector consists of three major parts designed to measure each aspect of a collision. The innermost component is the tracking system, composed of the Silicon Microstrip Tracker (SMT) and the Central Fiber Tracker (CFT). The SMT consists of barrels and wedges of silicon sensors placed in both the central and forward regions to cover low and high pseudorapidity, η[3]. Outside the SMT, the CFT is composed of scintillating fibers that produce light as a charged particle strikes them and simultaneously shifts the wavelength of light to an optimal wavelength for electronic readout. Surrounding the tracking system is the calorimeter, consisting of uranium and liquid argon. The uranium is responsible for stopping particles by causing them to turn into a spray of lower-energy particles, while the liquid argon ionizes, allowing us to detect the showers and measure their energy. The outermost layer is the muon system, designed to tag muons that leave a track in the central tracker, but little energy deposit in the calorimeter.

Systems of computers utilize the information encoded in the signals sent by the subdetectors to recreate events. Particles are identified according to the tracks and energy deposits they leave in the detector. Electrons will leave a track in the tracker along with a large, concentrated energy deposit in the calorimeter. Photons, because they are neutral particles, do not leave a track in the tracker but leave a concentrated energy deposit in the calorimeter like electrons. Quarks and gluons will hadronize into jets of many different particles and will leave a track through the tracker and a wide energy distribution in the calorimeter. Muons tend to travel straight through the detector, depositing little energy anywhere and leaving a long, nearly straight track through each component. A depiction of how each particle discussed appears in the detector can be seen in Figure 2.2

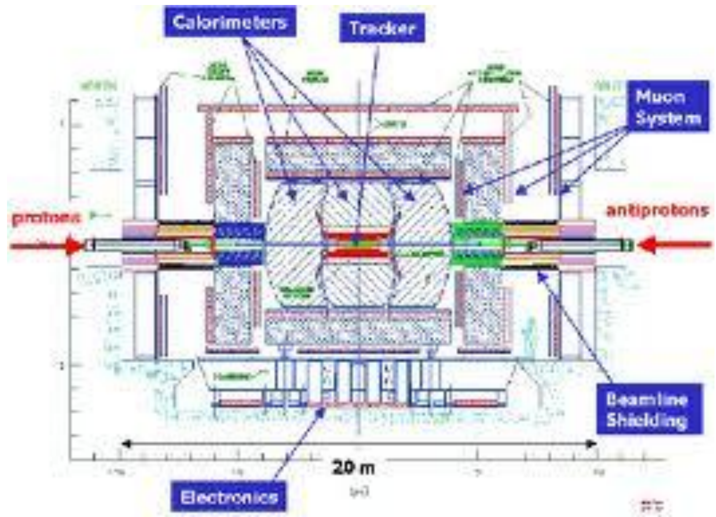


Figure 2.1. The DØ Detector

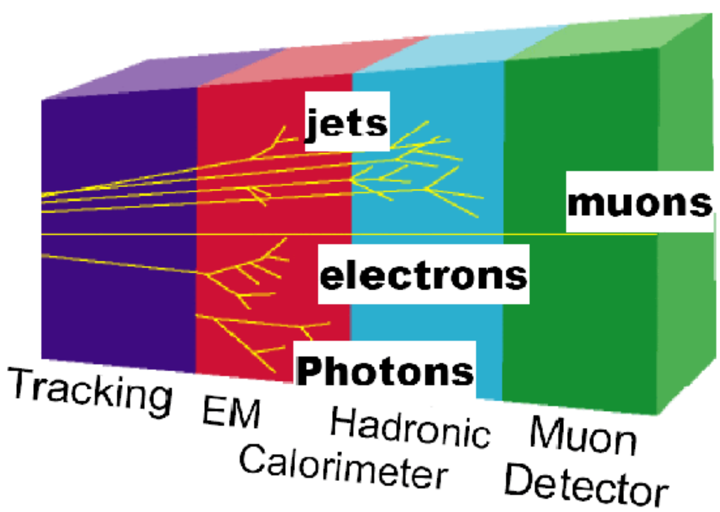


Figure 2.2. Depiction of Particles in the DØ Detector

2.2 C++

C++ is an intermediate-level, object-orientated programming language widely used throughout High Energy Physics today. It is used to implement a variety of other programs and packages, including ROOT. It is used as a basis for Common Analysis Format (CAF) packages before they are submitted to the DØ Central Analysis Backbend (CAB) to analyze samples of data and Monte Carlo simulations.

2.2 ROOT

ROOT is the primary way in which high energy physicists interact with data and MC. It is used for making and analyzing of histograms, along with making four-vector computations and utilizing statistical tools for data analysis. ROOT includes a CINT interpreter, interactively allowing the user to input C and C++ commands. ROOT consists of many classes, including TLeaf, TBranch, TTree, and the DØ-specific TMBTree. A typical TTree for a high physics analysis contains information about kinematic properties of all particles relevant to the event, *b*-tagging, missing transverse energy, and detailed particle or jet identification information.

2.3 *b*-tagging

b-tagging is the process by which jets originating from bottom quarks are identified. *b*-tagging can greatly reduce the W + jets or $t\bar{t}$ backgrounds. The signature of a *b*-jet is a secondary vertex a few millimeters from the primary vertex, indicating a lifetime of a couple picoseconds. The secondary vertex must also have an impact parameter that is slightly displaced from the primary vertex. The *b*-tagging efficiency is defined as the amount of tagged *b*-jets over the number of true *b*-jets. This value is determined by the data and used in Monte Carlo simulations.

2.5 Common Analysis Format Environment and wh_cafe

The Common Analysis Format Environment (CAFÉ) is the framework utilized by the DØ collaboration to process the data collected by the detector and create much smaller and more manageable files. This takes place within the DØ Central Analysis Backbend (CAB) servers. Utilizing the thousands of CAB servers dramatically reduces the amount of time this process takes, although it can still take days or weeks to complete. CAB splits up the jobs required of this process and allows them to run in parallel simultaneously.

wh_cafe is the framework specific to the WH analysis. It is composed of compiled packages of processors and configuration files. The configuration files specify the type of analysis (MVA or otherwise) to be done or what should be done within a particular analysis, while the processors include groups of packages written in C++ code that contain functions that specify and extract different information within the analysis. This information can be pertaining to flavor tagging of quarks, specifying primary/secondary/etc. jets, or various aspects of possible color flow between quarks, among many other things.

2.6 Multivariate Techniques

Multivariate techniques are used at DØ to turn many different low sensitivity variables into one strongly sensitive variable. They then assess its discriminating power in rejecting background events and keeping signal events. Various multivariate techniques are used, including neural networks, the matrix element method, and decision trees. Neural networks (Figure 2.3) take an input of event variables and give them a weight. Each input is then summed and modified with its specific weight. The activity of the neuron (the summation) is then fed into an activation function and if it is above a threshold value, an output is given. This method is very opaque however. The matrix method uses differential cross sections of signal and major backgrounds to estimate the likelihood of the signal. This method is very computer intensive and thus other methods may be preferred. Decision trees (Figure 2.4) are the way in which the data is analyzed to gain sensitivity to a Higgs signal at DØ. They work by placing the most significant cut on the sample, dividing the full sample into subsamples that either pass or fail this cut. It then continues to place cuts on each subset, creating finer sets of subsamples until a designated ending point, usually reached when statistics run out. Each final "leaf" then has a specific signal to background ratio which is used to classify test events. The "boosting" process repeats this training hundreds of times, applying higher weights to misclassified events each time after the first.

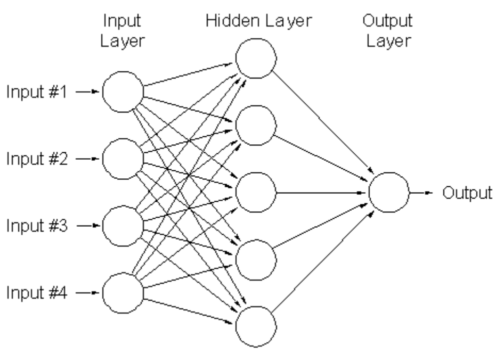


Figure 2.3. Neural Network

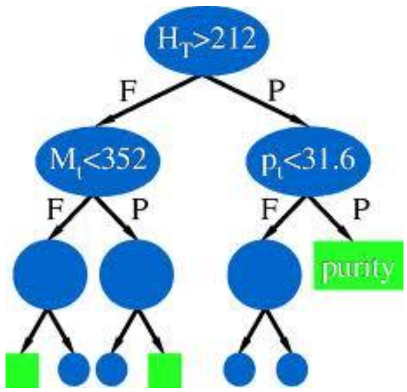


Figure 2.4. Decision Tree

3 Results

3.1 Adding to the Framework

The data and MC for the WH → lν*bb* group were processed using version v05-04-14 of the wh_cafe package. Several new variables were added to the search, including the significance of the curvature of the lepton track, various color flow variables, and fake track killer information. These particular variables utilized and analyzed information that had not been looked at before and it was hoped that they would have a significant impact on the WH → lν*bb* analysis. Because it was late in the analysis schedule and the WH group was working towards finalizing a result, these variables were not incorporated into their analysis during the summer of 2011, but several are planned to be incorporated for the next result.

In order to add the various variables, modifications were made to parts of the wh_cafe package, including WHEvent.cpp, WH_ResultsManager.h, WH_ResultsManager.C, and WH_PlotManager.C. A complete list of the variables added to the WH → lν*bb* framework and used as training inputs for multivariate analysis can be found in Table 3.1. Example plots of some of these variables can be seen in Figures 3.1 to 3.3. The color flow variables were added first, and after processing these variables by using the WH framework, we examined the "MVA Sensitivity" output contained in the log file. This particular output is coded as a cout statement within WH_PlotManager.C and provides a fast way to find the most effective variables that are good candidates for an MVA training. We found that the j12_cf_sum variable was the best candidate of the color flow variables, superior to the original color flow information contained in the framework. The results of the MVA training of all variables are discussed in the next section.

In addition to being added to the WH → lν*bb* analysis, these variables were also added to the WH → WWW → lνjjjj analysis. The group working on this analysis had developed their own version of code, and was able to extend the original WH framework. This extension took some time to get working. After approximately four weeks, the code was finally functional and the group was starting to get outputs and plots of variables. I joined this group after the code was finally functional, and I worked to add in the variables to their framework that I had added to the WH → lv*bb* framework. At time of writing, the version of code for the WWW analysis was v05-05-07WCorr.

Table 3.1. Variables Added to the Analysis

Variable	**Description**
j12_cf_sum, j12_cf_diff, j12_cf_asym	Addition, difference, and ratio of the angles between the apparent color flow and the line between the leading jets
lep_curvsig	Significance of the curvature of the track of the lepton in the event
lep_trk_ftk	Multivariate with the purpose of discriminating real lepton tracks from fake tracks
lep_trk_nsmt, lep_trk_hassmt	Determines the number/whether or not an event had a hit on the SMT tracker
j12_cf_etaangle	Determines the angle between the line between the two leading jets in the event and the η axis
j12_cf_beam_dphimin	Compares the color flow angles (j12_cf_12 or j12_cf_21 in the framework) to the "j12_cf_etaangle" angle listed above

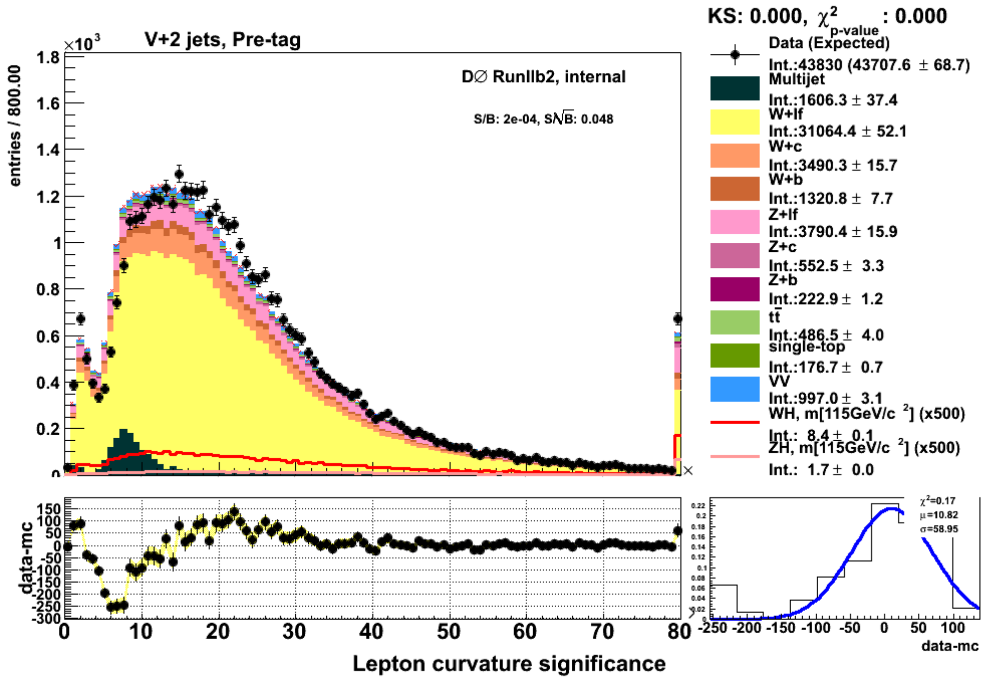


Figure 3.1. Example Plot of Lepton Curvature Significance

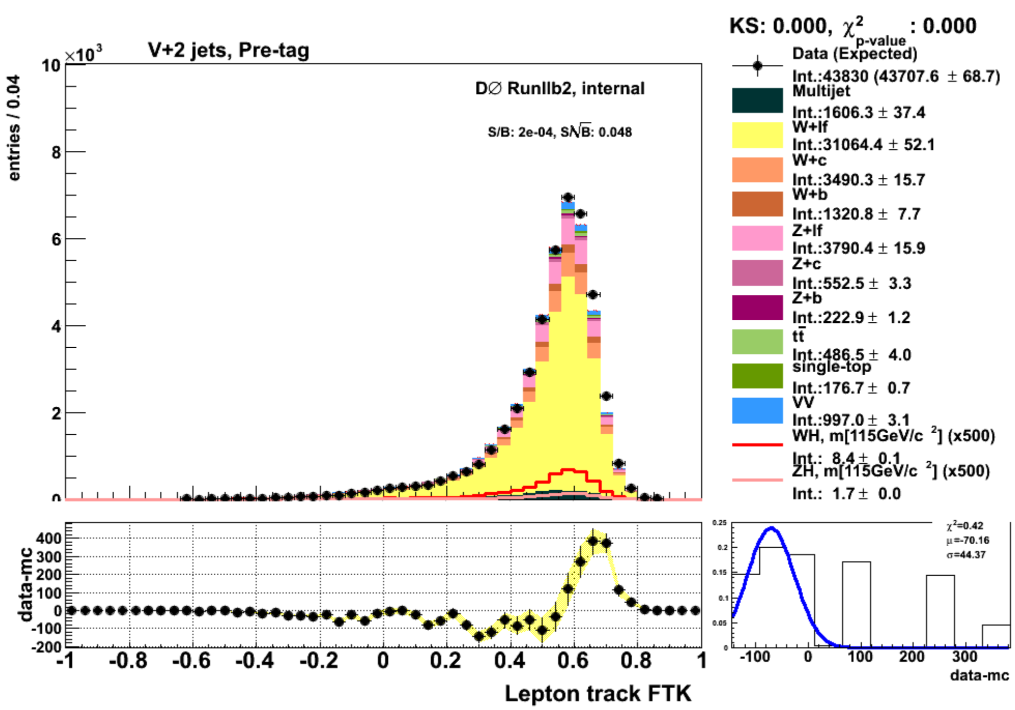


Figure 3.2. Example Plot of Lepton Track Fake Track Killer

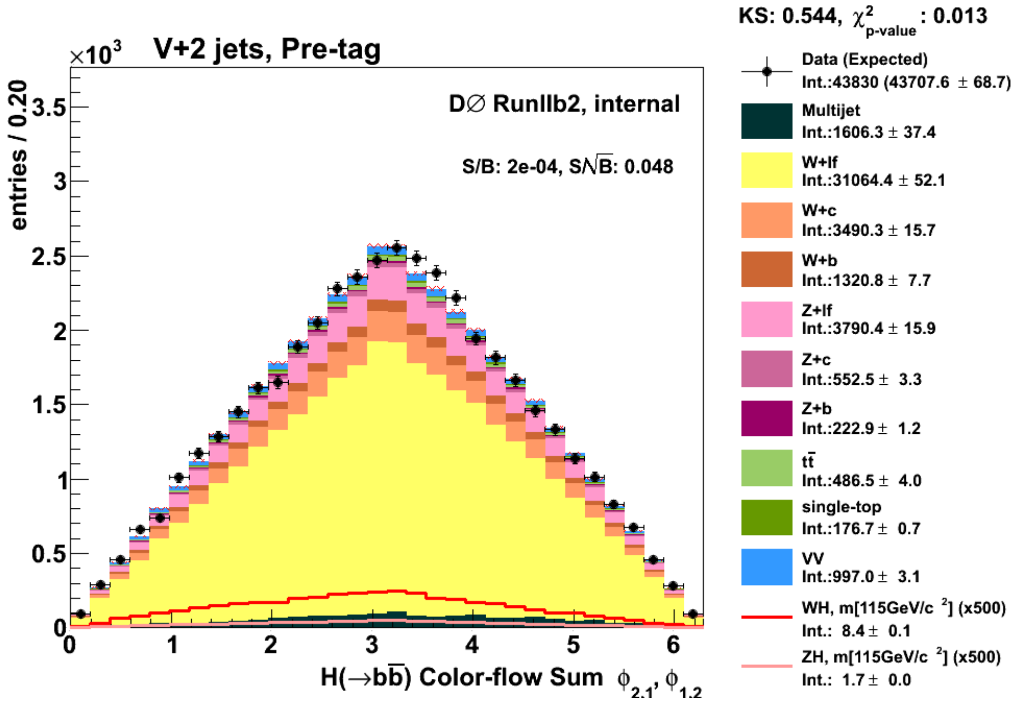


Figure 3.3. Example Plot of the Sum of the Color flow Angles

3.2 Multivariate Analysis

After processing the data and MC, a multivariate analysis (MVA) was trained on these variables. An "MVA QCD" was trained using boosted decision trees to discriminate against multijet background, in order to place a cut on the output of the MVA and reduce the multijet background. An example TMVA[4] BDTG output is depicted in Figure 3.1. Multiple versions of the MVA QCD were trained in the WH → lν*bb* channel. The purpose of this process was to determine the significance of each variable at this stage in the analysis. This is determined by examining the Signal Acceptance v. Background Rejection curve, one of the outputs of TMVA. The majority of the variables listed above in Table 3.1, including the color flow variables and the fake track killer variable, had no observable effect on this curve. However, the significance of curvature variable had a noticeable effect on the curve; in particular, it allowed more background rejection and kept more signal. The effect was slight, but even a slight effect is important for improving the sensitivity to a Higgs signal.

We had hoped that the color flow variables (including j12_cf_sum) and the fake track killer variable would have some impact in rejecting multijet background, but this was not the case. These variables did not have a noticeable effect on the Signal Acceptance v. Background Rejection curve, but it was speculated that much of the multijet background theoretically rejected by these variables had already been eliminated using other variables that were incorporated into the framework prior to the summer.

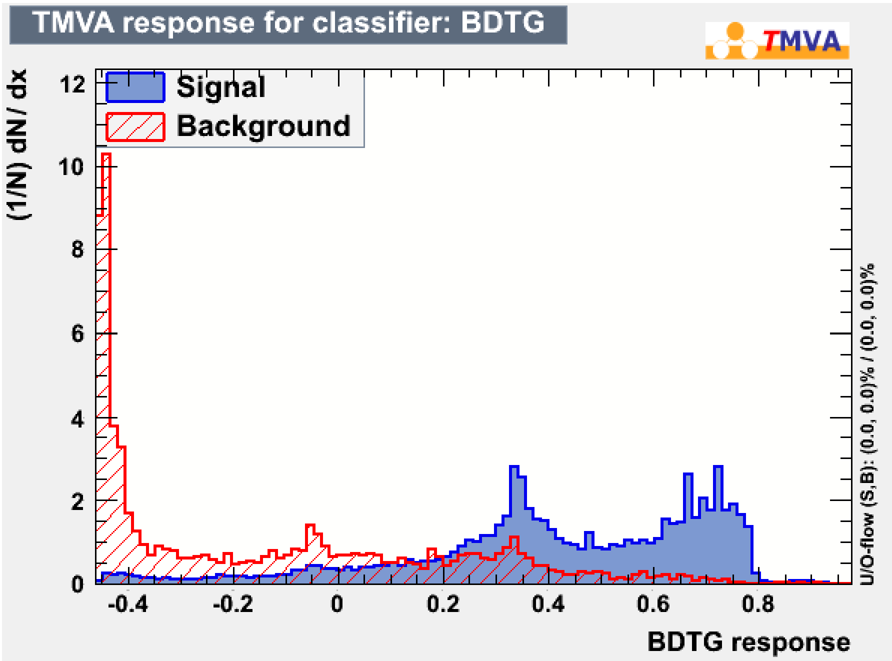
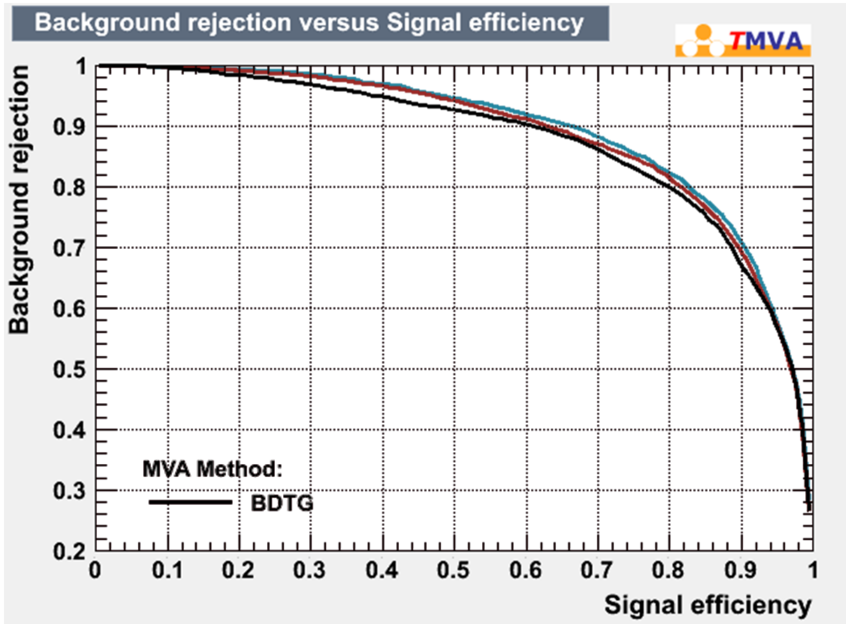


Figure 3.4. Example BDTG Output Distribution

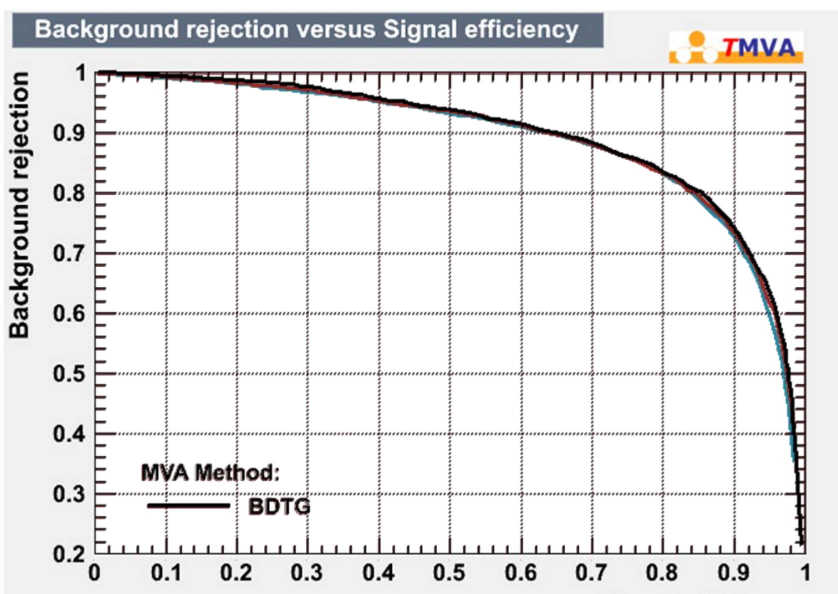
At the suggestion of Dr. Michael Cooke, a correction was added to adjust the lepton significance of curvature and, indirectly, the fake track killer information. This correction adjusted the Monte Carlo modeling of both variables according to the number of SMT hits there were on the detector in the event. There was a significant improvement in the modeling of both variables, and a second MVA QCD was trained for each variable, taking the correction into account. The Signal Acceptance v. Background Rejection curves had improved further after this correction. The figure comparing these three curves can be seen in Figure 3.5.

An “MVA Final” was then trained using a specific list of variables. This test would show how influential these variables were for WH versus all backgrounds, and the outputs of this training are used as inputs when the limit-setting procedure is implemented. The majority of the variables trained were those that the WH group had analyzed in great detail and determined to be significant variables. The color flow variables that had not significantly impacted the multijet background were added to this list. These variables were speculated to reject more W + jets background, and it was important to train a final MVA to determine their significance in rejecting all types of background. Figure 3.6 depicts the result of the final MVA training, and although the curve had moved slightly after adding these variables, it will be uncertain how significant they were at rejecting background until new limits can be produced that incorporate these variables.



Version 1: Original
Version 2: lep_curvsig added, without correction
Version 4: lep_curvsig added, with correction

Figure 3.5. The Effect of lep_curvsig on Rejecting Multijet Background



Version 1: Original – no color flow
Version 2: Original color flow angles added
Version 3: Variations on color flow angles added

Figure 3.6. Effect of Color Flow on Rejecting all Background

A similar process was implemented in the WH → WWW → lνjjjj channel. The results were much the same as the WH → lν*bb* channel, but because the WH → WWW → lνjjjj analysis was not as fine-tuned as the WH → lν*bb* group, and because they were still developing framework and experimenting with different variables, the results of MVA QCD training had a much bigger impact on multijet background rejection. There was a significant improvement in the Signal v. Background Rejection curve from variables that had not had any effect in the WH → lν*bb* MVA QCD. Because this group took much of their work from the WH → lν*bb* group, many of these initial variables were chosen for their sensitivity in searches for a low-mass Higgs. As a result, many of the variables had a much greater effect in the low-mass range compared to the high-mass range in the WWW search. At time of writing, there is ongoing work to develop new variables and modify the current ones for specific high-mass sensitivity.

4 Conclusion

Throughout the summer of 2011, much work was done on the WH → lν*bb* and WH → WWW → lνjjjj analyses. Many variables were added to each framework to try to help distinguish between background and signal events. The most notable of these variables was the significance of curvature of the lepton track, which caused a noticeable improvement in the Signal Acceptance v. Background Rejection curve for the MVA QCD in the WH → lν*bb* channel. Because the analysis is so far along and has been so fine-tuned, a small difference in the curve is significant. Within the WH → WWW → lνjjjj channel, immense progress was made, bringing the analysis framework from nonfunctional to functional. All of the variables listed in Table 3.1 had a significant impact on multijet background rejection, though that channel had not been as highly optimized as WH → lν*bb*. There remains much work to be done on both channels; in particular, the WH → WWW → lνjjjj requires that new variables specific to the WWW analysis be made, along with much adjusting of the modeling of background by Monte Carlo. Systematics must still be incorporated into this channel's summer result. Within the WH → lν*bb* channel, further optimization of multivariate analysis, along with the creation of additional variables, can allow the channel to gain sensitivity. The variables added over the course of the summer had significant effects on both analyses, and it is hoped that in the future they will help to improve outputs and final limits and improve the overall search for the SM Higgs.

5 Footnotes

[1] The Higgs mechanism works by coupling right- and left-handed particles. Because there is no observed right-handed neutrino, the Higgs does not give mass to neutrinos.

[2] Combined results from CDF and DØ have recently excluded the 158-175 GeV mass region at 95% CL.

[3] Pseudorapidity $\eta = -\ln(\tan(\theta/2))$

[4] TMVA, or Toolkit for Multivariate Analysis, is the way in which a multivariate analysis is performed. It consists of object-oriented C++ for each area of the multivariate analysis, and provides evaluation algorithms for training and testing the data and MC, along with analyzing the performance of each variable submitted.

6 Acknowledgments

I would like to thank Dr. Michael Cooke for his continued patience and teaching throughout the summer. I also had the pleasure of working with fellow interns Alex Abbinante, Tony Podkowa, and Youssef Mobarak and was able to gain insight on various aspects of the analysis from them. I would like to acknowledge my mentors Jamieson Olsen and Elliott McCrory, and I would like to thank Fermi National Accelerator Laboratory and the Fermilab SIST Committee for allowing me to have this educational, life-changing experience.

7 References

Z. Hynes, “Search for the Standard Model, High-Mass Higgs Boson in the WH → WWW → lvjjjj Channel.” Fermilab SIST Intern Papers (2009).

T. Junk, “Prospects for a Low-Mass Higgs Boson.” FERMILAB-CONF-07-636-E (2008).

A. Podkowa, “Continuation of the Search for the Standard Model Higgs Boson in the WH → WWW → lvjjjj Channel.” Fermilab SIST Intern Papers (2010).

V.M. Abazov et al. (D0 Collaboration), Nucl. Instrum. Methods, Phys. Res. A 565, 463 (2006). [arXiv:physics/0507191v1]

F. Englert and R. Brout, Phys. Rev. Lett. 13, 321 (1964).

P. Higgs, Phys. Rev. Lett. 13, 508 (1964).

G. Guralnik, C.R. Hagen, T.W.B. Kibble, Phys. Rev. Lett. 13, 585 (1964).

V.M. Abazov et al. "Search for WH Associated Production in 5.3 fb^{-1} of pp Collisions at the Fermilab Tevatron." (D0 Collaboration), Phys. Lett. B 698, 6 (2011) [arXiv:1012.0874].

V.M Abazov et al. “Search for Associated W and Higgs boson Production in pp Collisions at $\sqrt{s}$ = 1.96 TeV.” Phys. Rev. Lett. 102, 051803 (2009).

"Combined CDF and D0 Upper Limits on Standard Model Higgs Boson Production with up to 8.6 fb-1 of Data," The Tevatron New Phenomena and Higgs Working Group for the CDF and D0 Collaborations, [arXiv:1107.5518].

"The Tevatron's Massive Search for the Higgs," Seminar by Michael Cooke at Columbia University, 11/17/2010.