Search for the Standard Model, High-Mass Higgs Boson in the WH→WWW→lν.jj.jj Channel

Zachary Hynes

Massachusetts Institute of Technology

Fermi National Accelerator Laboratory
SIST Program 2009
Supervisor: Dr. Ryuji Yamada

August 5, 2009

Zachary Hynes
Massachusetts Institute of Technology

Abstract:

During the summer of 2009, work began on the preliminary stages of an analysis of associated production of the high-mass Higgs boson. Our analysis group investigated the high mass Higgs channel WH→WWW→lν.jj.jj, a channel that has not yet been studied by other analysis groups. This paper will explain ongoing efforts to process data, Monte Carlo (MC) background and signal samples, model data effectively with MC background processes utilizing analysis and plotting scripts, and develop multivariate classifiers for the purpose of distinguishing between signal and background events.

Table of Contents:

I. Introduction

Search for the Higgs Boson:

The existence of the Higgs boson, originally postulated by Peter Higgs and others in 1964[i], has been the target of many research efforts in recent years at CERN and Fermi National Accelerator Lab (Fermilab). The mass of the Standard Model (SM) Higgs is commonly believed to lie within the range of 114.4 GeV and 200 GeV. Recently, CDF and D0 have been able to exclude the mass range of 160 to 170 GeV at a 95% confidence level[ii]. Within the realm of Standard Model (SM) Higgs searches, there exist two principal categories; high mass Higgs searches within the range of 150 to 200 GeV, and low mass Higgs searches that encompass a mass range from 115 GeV to 150 GeV. The channel through which the Higgs decays varies depending on its mass. At lower masses, the Higgs is expected to decay predominantly to a bottom quark/anti-quark pair while at higher masses the Higgs is expected to decay primarily to a pair of W bosons, a class of particles responsible for the mediation of the electroweak force. The relationship between the Higgs mass and the means by which it will likely decay is illustrated in Figure 1.1.

The SM Higgs can be produced directly through gluon fusion or through associated production methods, where it is produced alongside a W or Z boson. The channel with which our analysis is concerned is associated production of the Higgs alongside a W boson. As can be understood from Figure 1.2, the cross-section of this process is around one order of magnitude smaller for the high mass Higgs in comparison with direct Higgs production.

Figure 1.2 indicates that one should expect more Higgs events from channels involving direct production for a given dataset than from associated production channels. However, the number and cross-section of the background physics processes in these direct production

channels will exceed the number and cross-section of the background physics processes for the WH→WWW→lv.jj.jj production channel.

D0 Detector:

The D0 detector is the engineering marvel at the heart of decades of brilliant work in the field of high-energy physics. Comprised of an inner silicon detector, three liquid argon uranium calorimeters, and a muon system[iii], the detector produces the data that is essential to high-energy physics analyses. Triggers at the detector conduct preliminary event selection, paring down what would otherwise be an unmanageably large dataset by recording in full only those events of interest to high-energy physicists.

Over the course of the past summer, I have gained exposure to many of the Higgs searches that are ongoing at the D0 collaboration in a variety of channels. A considerable number of the tools and processes used in these Higgs searches are uniform across many channels; the existence of such common tools is critical to expediting and improving the quality of these Higgs searches throughout the D0 collaboration.

This paper will explain ongoing efforts to process data, MC background, and MC signal samples, model data effectively with MC background using analysis and plotting scripts, and develop multivariate classifiers for the purpose of distinguishing between signal and background events.

II. Materials and Methods:

C++:

C++ is an object-oriented programming language that is utilized in ROOT and other analysis packages used widely throughout High Energy Physics (HEP). This summer, I had the opportunity to learn C++ from Dr. Walter Brown through the Accelerated C++ Short Course offered at Fermilab. This class gave me crucial insight into a programming language with which I frequently interacted through daily use of packages like ROOT or CAFe, scripts developed by others, and my own code. I benefited tremendously from the opportunity to learn a new programming language in this structured setting.

ROOT:[iv]

High-energy physicists frequently interface with data and MC samples through the user-friendly medium provided by the ROOT software package. TTrees and TMBTrees, their D0-specific relatives provide the overarching structure in the ROOT files with which we have worked in our analysis; the TBranch, TClonesArray, and TLeaf classes, among others, also help to organize ROOT files like the ones with which we worked in our analysis, providing easy access to information relating to HEP events. Typical TTrees for high-energy physics analyses contain information relating to the kinematic properties of leptons and jets, b-tagging (attributing a jet to a b quark), missing transverse energy (MET), and identification information like the event number and run number associated with each event. Lastly, as high-energy physics analyses use event variable distributions regularly, the ROOT package provides simple ways to making, analyzing, and writing histograms.

Common Analysis Format (CAF):

The D0 collaboration employs Common Analysis Format (CAF) to process data stored in TMBTrees. Configuration files specify the processors to be run over the dataset and the parameters that will be used by those processors. These processors are coded in C++ and will typically loop over jets, leptons, and events; common functions carried out by CAF processors include applying cuts, flavor-tagging of jets, producing new TTrees, and writing those TTrees to file. Many analyses employ groups of common processors, depending on the type of lepton(s) in the final state (electron, muon, or some combination of the two) and whether the sample in question is data from the detector or Monte Carlo simulations.

In principle, the functions fulfilled by analysis processors could be carried out interactively in ROOT; however, the potent combination of CAF and the powerful D0 Central Analysis Backend (CAB) computing system is required to tackle the enormous task of processing millions of events in a reasonable amount of time.

Processing Data and MC Backgrounds with wh_cafe

We have conducted data processing efforts using the wh_cafe package, version whtree_v01-01-00, in a D0RunII p21.13.00 release. We have also made successful forays in data processing with hww_cafe v00-02-br and its BtagNtuple processor, but as of the time of this writing, we are planning to use wh_cafe in our analysis. Our effort to process data and MC with wh_cafe also utilizes the packages featured in the D0 working area set-up script featured in Figure 2.2. With the wh_cafe package, we utilized the WHTree processor. We were instructed to modify some components of our wh_cafe package, including WHEvent.hpp (class definition file), WHTree.cpp (processor source file), and WHTree.hpp (processor header file), and recompile the package after running into some issues while using the unaltered whtree_v01-01-

00 version of the package[v]. These partial updates allowed us to successfully process data and MC, but as will be discussed in the "Preliminary Results" section, it appears likely that we will need to update our wh_cafe package in its entirety in order to complete the post-processing stage successfully.

We processed our data and MC background datasets utilizing '.config' files modified from template files provided to us by Yuji Enari. The analysis chain utilized to process the VJets_MUinclusive_Summer09_RunIIb_v2 dataset can be seen in Figure 2.1; the last processor in the chain, the WHTree processor, creates the output file that will be used in later analysis steps. Among other important specifications provided in this .config file, we require at least two jets that have passed certain selection criteria and at least one lepton meeting "loose" lepton criteria in order for the given event to be written to the wh_cafe output file (wh_tree.root). In later stages of our analysis, we will further reduce our samples by requiring four jets and a tight lepton, but it was suggested by members of another analysis group that we may want to refrain from requiring a minimum of four jets at the data processing stage, so we are adhering to this recommendation. One aspect of the configuration file that we will change when we reprocess our data and MC samples is a cut on MET; right now, there is no minimum MET specified in this .config file (the -1 value to which the MET minimum is set signifies "no cut"). While not all leptons that are written to the wh_tree.root output file meet tight selection critieria, the WHEvent class included in the wh_cafe provides an "IsTightLepton()" method by which one can determine whether a given lepton passes tight criteria.

After running the .config file featured in Figure 2.1 over more than 111 million events across the 15,000 files that are specified by the VJets_MUinclusive_Summer09_RunIIb_v2 SAM dataset definition, the resulting output file contains approximately 800,00 events.

Post-Processing Data and MC Backgrounds

After running the data and MC background samples through CAF, the next step in our analysis looks to model the data event variable distributions using weighted sums of the MC and QCD background event variable distributions. As we expect a small number of Higgs events, between five and ten, to be found in approximately 4 fb^{-1} of data for Higgs masses between 150 and 200 GeV[vi], we would expect the MC and QCD backgrounds to model the data effectively regardless of whether the Higgs boson exists given the large number of data events. If the background processes do not model the data well at this stage, this could suggest that the weightings of the backgrounds have not been executed properly or that essential background processes were not represented at all in the data modeling process.

The WHTree processor that we used to process data can access and store in the output tree a number of important event weights. However, unlike more recent analysis scripts, the analysis script WHTreeAnalyze.h from wh_cafe whtree_v01-01-00 applies only the "global" event weight and an event weight based on the cross-section of the physics process modeled by the given event when producing the event variable distribution plots. We could endeavor to modify the analysis scripts from the old version of the package on our own, but this option should be considered a last resort given the level of expertise of those creating and updating these packages. This issue will be explained in further detail in the "Preliminary Results" section.

Backgrounds for the WH→WWW→lν.jj.jj Channel

Our analysis is in the midst of processing a wide variety of backgrounds for the WH→ WWW→lν.jj.jj channel for the associated production and subsequent decay of the Higgs boson

into a final state of four jets, one lepton, and a neutrino. Practically speaking, we deal with two different kinds of backgrounds; backgrounds reflective of other processes that ultimately lead to the same final state of four jets, one lepton, and one neutrino, as well as a background referred to as multijet background, or QCD. QCD background arises due to the fact that the detector can incorrectly label a jet as a lepton. When using CAF to process data and MC background, user-specified criteria are applied to categorize potential leptons as "tight" and/or "loose" (a lepton that is tight, for most tight and loose definition pairs, will also meet loose criteria). Packages used in conjunction with our base CAF package, vjets_cafe, provide some commonly used lepton quality definitions. If one makes the criteria for consideration as a tight lepton too stringent, potential signal events get filtered out of the data. Conversely, lenient "tight" criteria will dilute the sample with QCD background; jets originating from light quarks will be frequently mistaken for leptons. While some QCD is inevitable, it is essential that we model the QCD background correctly in our analysis. The analysis scripts that we are using in association with the wh_cafe package include a method for modeling QCD background based on the actual data.

Multivariate Analysis:

Multivariate analysis is the principal means used by high-energy physicists to distinguish between the various physics processes that may take place when interactions occur in the detector. Firstly, however, let's take a step back and consider the notion of Monte Carlo simulations. Physicists can specify parameters for the Higgs boson, the most essential of these being the Higgs mass, and then use powerful computing systems to simulate how the Higgs would manifest itself in the D0 detector in terms of the kinematic information of the final state

particles that would be recorded by the detector. Unfortunately, these simulations show no evidence that the Higgs leaves a simple, unique "signature". No basic event variable distribution for the signal MC sample shows such drastic deviations from the same event variable distribution for the background MC such that an event of unknown classification, like a data event, could be classified solely on the basis of the value that this event associates with the specified event variable. It is at this critical juncture that supervised learning techniques like neural networks play a role. Neural networks take a fraction of the entire signal and background MC samples, a "training" sample, and develop a complex classification scheme based on user-specified event variables using the events from the signal and background training samples as guides. The classification capacity of these networks is then tested on another portion of the signal and background MC.

An understanding of the physics of the targeted signal process will often allow one to calculate more sophisticated event variables that lead to better discrimination between signal and background processes. In the results section, we will examine distributions of the leptonic W boson transverse mass for data and MC background samples; this is a quantity that is derived from the kinematic variables of the lepton and the MET (associated with the neutrino). In the WH→WWW→lν.jj.jj channel, we can also reconstruct the mass of the other two W bosons and even make final cuts based on the reconstructed mass values. We should expect, for some permutation of jet-jet pairings, the dijet invariant mass of each of the two pairs to fall somewhere within a range centered on the W boson mass for signal MC events; this would not necessarily be the expectation for w+jets or other MC backgrounds.

Toolkit for Multivariate Data Analysis (TMVA)[vii]*:*

Using the TMVA (v. 3.9.5) package in ROOT (v. 5.22.00a), one can choose from a variety of multivariate analysis techniques. For this paper, we trained and tested an example neural network. Neural networks are comprised of several layers of nodes, the first and last being the two most intuitive layers. The first layer represents the input variables coming from the TTree in the ROOT file provided by the user; the last layer ideally represents is the output layer. Ideally, the value in this last output node can easily be associated with signal-like or background-like events. For example, in many neural networks in high-energy physics, signal events cluster near higher output values (close to one) and background events tend to result in output values closer to zero. The architecture of the neural network to be developed by TMVA is left to the discretion of the user. The user must select meaningful input variables for which there are significant differences in the distributions of the signal and background MC samples. He or she must also experiment with the number of layers of nodes in between the input and output layers as well as the number of nodes in each of these layers to achieve an optimal configuration.

After using C++ macros in conjunction with TMVA-specific classes to specify network architecture, train the network, and test the network, we utilize the TMVAGui interface (shown in Figure 2.3) to evaluate the performance of the network. We can examine event variable distributions for signal and background processes for the training sample of signal and background MC events after cuts have been applied, look at the correlations between variables in the analysis, and perhaps most importantly, see how our classifier (the network) fared when evaluating the "test" sample of signal and background events. Ideally, the multivariate analysis (MVA) output distributions for the two classifications, signal and background, will form sharp, distinct peaks with significant separation between the two peaks; if the signal and background MVA output distributions contrast in this way, it is, in a manner of speaking, an indication that

the neural network can classify events of unknown type with confidence if the expected numbers of signal and background events are relatively close. Unfortunately, this final condition concerning equal numbers of expected signal and background events will probably not be true for our case even after all cuts have been applied, but we still hope to obtain these two sharp, distinct MVA output distribution peaks.

After the network has been trained and tested, we apply the classifier to data events (of unknown classification); lastly, we compare the MVA output distribution for the data to the MVA output distribution for the MC background processes. Evidence for the Higgs boson would arise if the distribution for the data deviated significantly from the MVA output distribution for the background MC toward the MVA output distribution for the signal MC.

III. Preliminary Results

Members of our analysis group continue to develop familiarity with the complex web of tools used in these high-energy physics analyses. In the coming weeks, we will be reviewing our methods for processing data and MC with CAF, post-processing with the wh_cafe and hww_cafe packages, and carrying out multivariate analyses. We will likely reprocess some or all of our data and MC samples. When submitting data processing jobs to D0's Central Analysis Backend (CAB) computing system, some jobs unexpectedly failed while others succeeded. We know the mechanism to recover jobs and will ensure that all data and MC files are processed in our final analysis, but we have opted not to implement recovery projects while we continue to refine our approach. We are still testing out new techniques and new packages; all results are preliminary and are subject to future corrections.

Processing and Post-processing Data and MC:

Recently, we have processed data and MC backgrounds using the WHTree processor from wh_cafe and have proceeded to check for agreement between the event variable distributions by utilizing various post-processing tools included in the wh_cafe package. We are currently concerned that the analysis suite from the version of the wh_cafe package with which we were working, labeled whtree_v01-01-00, does not apply the necessary event reweightings based on the reweightings that we have seen applied by others. As evidence of our concerns, I present in Figure 3.6 a comparison of the code utilized to calculate event weights in WHTreeAnalyze.h, one of the major analysis tools in wh_cafe whtree_v01-01-00, and WH_Analysis.h from the wh_cafe package obtained using the addpkg wh_cafe –h command in the context of a D0RunII p21.13.00 release sometime in late July to early August. WH_Analysis.h seems to be analogous to WHTreeAnalyze.h from the version whtree_v01-01-00 in some respects. As one can see in Figure 3.6, functions utilized by the WH_Analysis.h incorporate many more event weights in the more recent version of the package as compared with WHTreeAnalyze.h from version whtree_v01-01-00. Further investigation of this point will be necessary, but at the time of this writing it appears as though fully upgrading the version of wh_cafe that we are using in our analysis would be an appropriate step.

Figure 3.7 shows the event variable output distributions of data and MC background samples for the transverse mass of the leptonic W boson for data and MC background processes after cuts have been applied to require, among other conditions, exactly four jets and one lepton, the key components of the final state of our Higgs decay mode. Plots showing event variable distributions for data and MC backgrounds were obtained by running the rundatavsmc.csh shell script, which subsequently relies upon the tree_datavsmc.cc macro and the WHTreeAnalyze

class as defined in WHTreeAnalyze.h[viii], over many of the significant backgrounds for our channel. This plot is promising in some respects; there appears to be reasonable agreement between the weighted sum of the backgrounds and the data. However, the w+jets scale factor that is applied to help the MC backgrounds model the data more effectively due to uncertainties in the cross-section of w+jets processes is reported by the output of the analysis to be approximately 7.08. Other channels, such as the H→WW→lν.jj (muon) channel being investigated by Shannon Zelitch, Marc Buehler, and Bob Hirosky from the University of Virginia recently reported in a D0 Note submitted for review a w+jets scale factor of 1.035 or 1.05 depending on the type of W p_T reweighting that is applied in the data processing stage[ix]. Given what we believe to be the shortcomings of the analysis script, WHTreeAnalyze.h, from the whtree_v01-01-00 version of wh_cafe, it is not surprising that our w+jets scale factor does not align with values received by others. It should be noted that our list of background processes will likely expand as we head toward a final analysis. We have learned that even MC background samples that wouldn't appear to contain events with four jet and one lepton final states may in fact contain such final states, so there may be a sample that was not incorporated into this analysis that contains significant background in our channel. We still have questions about values that should be used for different parameters for the .config files and may need to add, subtract, or reposition processors in our analysis chain.

As has been discussed, we will be implementing our analysis with a new version of wh_cafe and its associated scripts shortly. We expect to achieve greater agreement between data and MC backgrounds after implementing the new version of wh_cafe.

Multivariate Analysis:

Recently, we have processed Monte Carlo background and signal processes and run TMVA analyses using these results. I developed a macro that takes CAF-processed data and MC as input, applies cuts, and writes a TTree such that access to event variables is not dependent on methods associated with the WHEvent and other classes associated with the wh_cafe package. I originally believed that this intermediate step was necessary if we were to be able to carry out multivariate analysis with certain analysis programs, but it now seems that this will probably not be necessary in the future. It is important, however, because it also applies cuts on a number of variables, including the number of jets, which must be equal to four, the number of leptons, which must be equal to one, and the tight/loose nature of the lepton (it must meet tight lepton criteria). Figures 3.2-3.5 come from TMVA analysis of (unweighted) signal MC for a Higgs mass of 180 GeV in the muon channel and (unweighted) background MC in the w+5lp (inclusive) channel. We do not weight the signal or background MC used as input for our multivariate analysis as we have not yet determined a comprehensive list of weights that need to be applied for each sample. Figure 3.2 shows the cuts that were applied to the signal and MC backgrounds; Figures 3.3 and 3.4 show the event variable distributions for the two processes, and Figure 3.5 shows the MVA (neural network) output for the processes.

As one can see in Figure 3.5, the separation between the signal and background peaks has not reached a level such that the network, as presently constituted, would make for a strong classifier of data events; if an event of unknown classification were to be run through this network, few MVA output values would indicate with much certainty whether the event should be classified as background or signal; the overlap between the two output distributions for the test sample is too significant. Note that the MVA output distributions are normalized so that there are an equal number of signal and background events.

IV. Challenges

As someone who has begun to program within the past six months and did not have familiarity with any of the tools common to high energy physics analyses at D0 prior to arriving at Fermilab in late May, I had to devote a significant portion of my time to learning the system, and how each of the tools available fits into the larger analysis framework. Initially tasked with completing the multivariate analysis portion of the project, my responsibilities changed over the course of the summer. I have made decisions on such matters as the packages to be used in analysis and frequently communicate with other physicists in the D0 collaboration that conduct similar Higgs analyses. While it required significant effort to develop an understanding of the entire high-energy physics analysis framework, I believe that my experience has been much more fulfilling and impactful as a result.

Perhaps one of the most essential lessons that I learned over the course of this experience was one in the organization of one's materials for an analysis effort of this size. In the excitement of testing out new programs and making progress, it can be easy to lose sight of the importance of organizing one's results. Of course, in the initial stages when one is developing familiarity with and testing out new tools, it seems unwise to devote too much time to organization. However, major jobs like data processing require a sound framework in which files relating to job submission, data processing specifications (.config files), and output can be stored. Especially as we look into utilizing new versions of packages, it will be important that I improve my documentation of the packages being used. On the D0 Wiki, the standard packages to be used in analysis are listed with version numbers; we need to make sure we can identify which version of a package is being used at all times. In accessing the new version of wh_cafe, I used the

addpkg wh_cafe –h command to obtain the "head" version of the package within the past several weeks, but parts of the package could have changed even in that short time span. Depending on the part of the package that was changed, my references could be made obsolete as a result. Even those files to which Yuji Enari directed us from one of his areas on clued0 may have been altered or modified since he originally sent the instructions to use his files. Maintaining strong organizational structures and implementing the good practices that we have learned "the hard way" will be essential to the integrity of any ultimate final result and achieving high levels of efficiency along the way.

V. Future Plans:

I will continue work on this analysis through August, with the hope that over the course of the next four weeks, our project will continue to gain momentum and that the analysis stages discussed here will be finalized by that time. I immensely enjoyed my first taste of high-energy physics and appreciated the zeal with which the Fermilab community pursues a greater understanding of fundamental scientific questions. This experience has kindled a desire to continue on to a career in high-energy physics in some capacity, and I look forward to future opportunities to contribute to this field and the work that is ongoing at Fermilab.

VI. Acknowledgments

For the duration of the Summer Internships in Science and Technology program, I have had the great pleasure and privilege of working with Dr. Ryuji Yamada and fellow student intern Vesselin Velev on this project. Our ongoing analysis has also received great support from the D0 collaboration, and in particular, Dr. Joseph Haley, Dr. Shannon Zelitch, Dr. Lidia Zizkovic, and

Dr. Yuji Enari. We look forward to continuing our partnerships with these individuals as we make progress on our analysis.

VII. Figures:

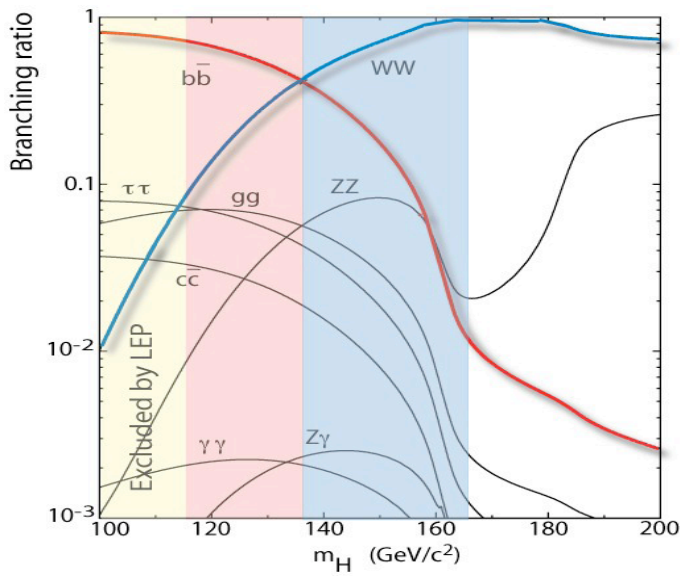

Figure 1.1: Branching Ratios for Higgs Boson as a Function of Mass[x]

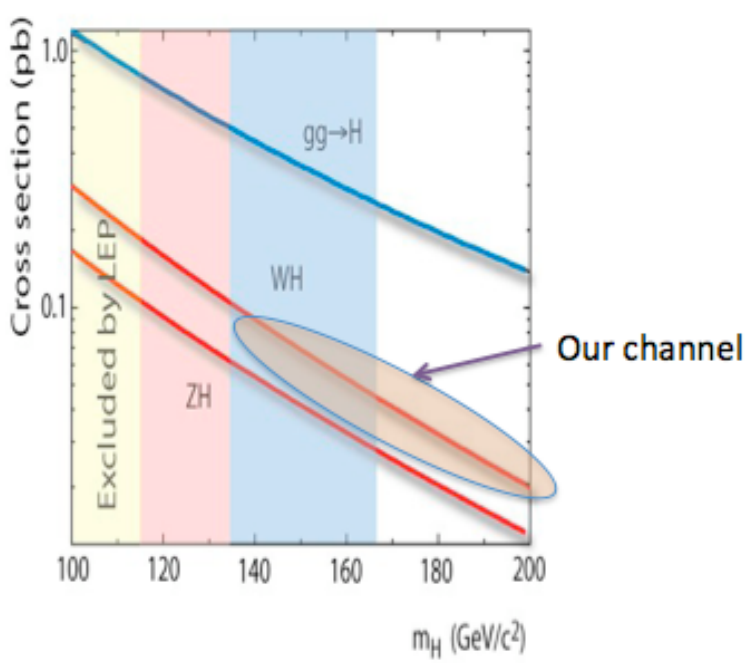

Figure 1.2: Cross Sections for Various Higgs Production Methods[xi]

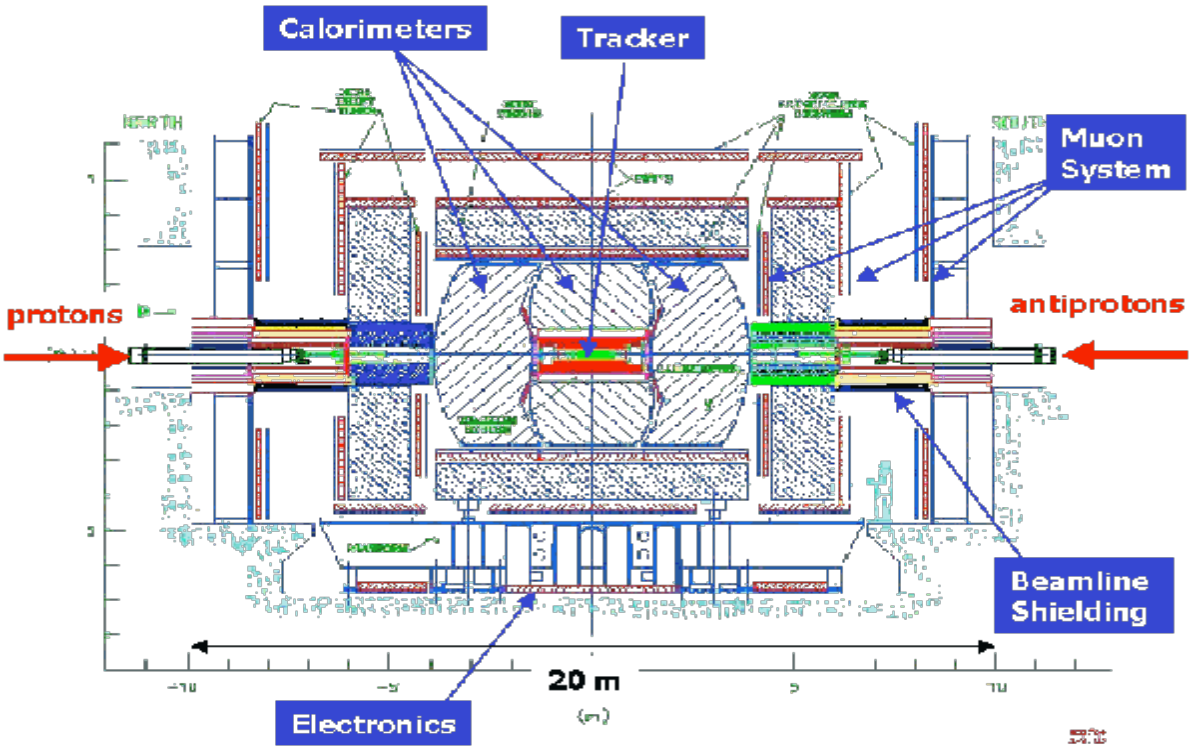

Figure 1.3: Drawing of the D0 Detector[xii]

```
# The data files are not btagged yet
VJets.InputJet_Branch:          %{VJets.JetAlgo}

cafe.Run:                                   Group(MuJets_Data_Base)

# Possible to add to the BEGIN, CUTS and USER groups here:

USER.Run:          Group(VJetsMETSignificance)
+USER.Run:      JetSelector(jet_sele)
+USER.Run:      CafeReadEventProc(event)
+USER.Run:      BTagProcessor(tagger_nn_ol)
+USER.Run:        Group(MySaveOutput)
+USER.Run:        Group(MySaveWeights)
+USER.Run:      WHTree(whtree)

jet_sele.nJets:                    2
jet_sele.nJetsMax:                 %{VJets.GoodJet_Nmax}
jet_sele.Ntracks:                  -1

VJets.GoodJet_Nmax:     -1
VJets.GoodJet_detEtaCut: 2.5
VJets.GoodJet_etaCut: 10
VJets.GoodJet_PtCut: 15.0

whtree.Analysis_Version: 001
whtree.LooseLepton_FromBranch:  %{VJets.LooseMuon_Branch}
whtree.TightLepton_FromBranch:  %{VJets.TightMuon_Branch}
whtree.Lepton:                  MU
whtree.Jets_FromBranch:         GoodJCCB_t3
whtree.VertexConf_Jets_From:    VertexConfirmedJets
whtree.BJets_FromBranch:        GoodJCCB_t3
whtree.MonteCarlo:
```

Figure 2.1: Snapshot of MuJets_Data_Summer09.config[xiii]

```
setup caf_tools -t
setup D0RunII p21.13.00 -O SRT_QUAL=maxopt
newrel p21.13.00 work
cd work
d0setwa

#using csh
setenv PATH ${SRT_PRIVATE_CONTEXT}/shbin/${SRT_SUBDIR}/:${PATH}
setenv LINK_SHARED y

setup d0cvs
cvs co -r p21-13-00 D0reldb/inventory
autoroot.py tmb_tree cafe

addpkg vjets_cafe v03-04-00

addpkg tmb_tree p21-br-61
addpkg jetcorr p21-br-12
addpkg cafe p21-br-28
addpkg cafe_sam p21-br-06
addpkg caf_util p21-br-112
addpkg tau_tmb p21-br-01 # Tau ID
addpkg tmb2ttau v00-00-02
addpkg caf_mc_util p21-br-133
addpkg emid_cuts p21-br-22
addpkg met_util p21-br-01
addpkg eff_utils p21-br-24
addpkg caf_eff_utils p21-br-14
addpkg caf_trigger p21-br-73
addpkg lumi_profiles v2009-07-16

addpkg beamposition v2009-06-24
addpkg emid_eff v7-preliminary-37
addpkg muid_eff v04-04-00
addpkg jetid_eff v03-01-03
addpkg tauid_eff v00-01-02

addpkg dq_defs v2009-04-18
addpkg caf_dq p21-br-03
addpkg dq_util p21-br-05

addpkg btags_cert v08-00-01
addpkg btags_cert_caf v00-09-00
addpkg d0root_analysis v00-09-89
addpkg d0root_sltnn v00-00-04
addpkg d0root_jlip v00-02-02
addpkg d0root_nnbtag v00-01-02
addpkg d0root_mva_btagger v00-00-06
addpkg d0root_tmbtree v00-10-31
(END)
```

Figure 2.2: Packages Used in Data Processing Analysis in Conjunction with wh_cafe whtree_v01-01-00 (Set-up Script for D0RunII p21.13.00 Working Area)

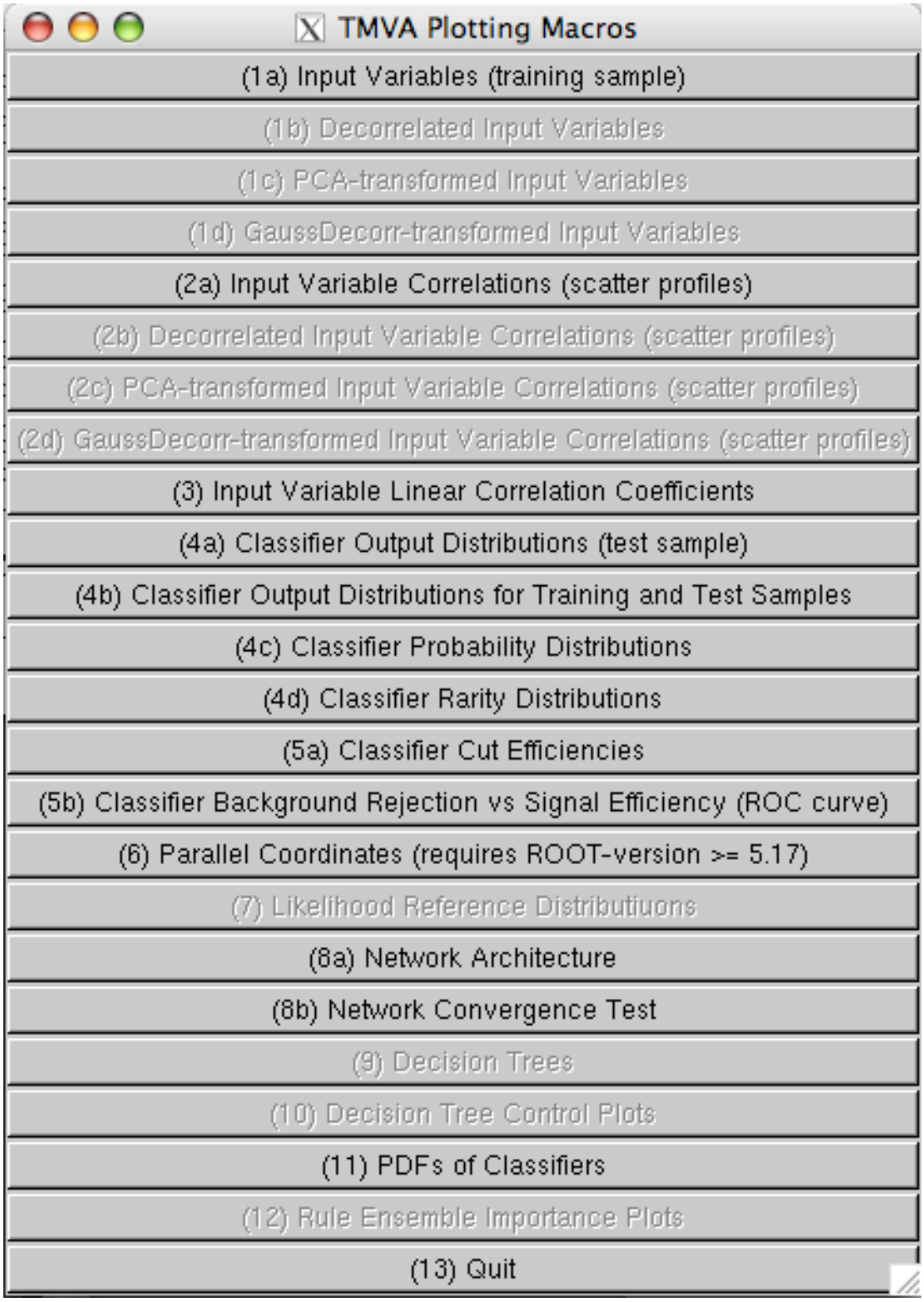

Figure 2.3: TMVA GUI Interface

W+Jets:
CSG_alpgenpythia_w+0lp_lnu+0lp_excl_p211100_v3
CSG_alpgenpythia_w+1lp_lnu+1lp_excl_p211100_v3
CSG_alpgenpythia_w+2lp_lnu+2lp_excl_p211100_v3
CSG_alpgenpythia_w+3lp_lnu+3lp_excl_p211100_v3
CSG_alpgenpythia_w+4lp_lnu+4lp_excl_p211100_v3
CSG_alpgenpythia_w+5lp_lnu+5lp_incl_p211100_v3
CSG_alpgenpythia_w+2b+1lp_lnu+2b+1lp_excl_p211100_v3
CSG_alpgenpythia_w+2b+2lp_lnu+2b+21p_excl_p211100_v3
CSG_alpgenpythia_w+2b+3lp_lnu+2b+3lp_incl_p211100_v3
CSG_alpgenpythia_w+2c+1lp_lnu+2c+1lp_excl_p211100_v3
CSG_alpgenpythia_w+2c+2lp_lnu+2c+21p_excl_p211100_v3
CSG_alpgenpythia_w+2c+3lp_lnu+2c+3lp_incl_p211100_v3

Z+Jets:
CSG_alpgenpythia_gamz+0lp_mumu+0lp_excl_15_75_p211100_v3
CSG_alpgenpythia_gamz+1lp_mumu+1lp_excl_15_75_p211100_v3
CSG_alpgenpythia_gamz+2lp_mumu+2lp_excl_15_75_p211100_v3
CSG_alpgenpythia_gamz+3lp_mumu+3lp_incl_15_75_p211100_v3
CSG_alpgenpythia_gamz+2b+1lp_mumu+2b+1lp_excl_15_75_p211100_v3
CSG_alpgenpythia_gamz+2b+2lp_mumu+2b+2lp_incl_15_75_p211100_v3

Diboson:
CSG_pythia_w+w_incl_p211100_v3
CSG_pythia_w+z_incl_p211100_v3
CSG_pythia_z+z_incl_p211100_v3

Top/anti-Top:
CSG_alpgenpythia_t+t+0lp_lnu+2b+2lpc_excl_m172_p211100_v3
CSG_alpgenpythia_t+t+1lp_lnu+2b+3lpc_excl_m172_p211100_v3
CSG_alpgenpythia_t+t+2lp_lnu+2b+4lpc_incl_m172_p211100_v3

Data:
VJets_MUinclusive_Summer09_RunIIb_v2

Figure 3.1: Data and MC Backgrounds Processed Using Package Versions in My Working Area and Incorporated into Post-processing Analysis

```
TCut c_Jet1_pT="Jet1_pT>40";
TCut c_Jet2_pT="Jet2_pT>30";
TCut c_Jet3_pT="Jet3_pT>25";
TCut c_Jet4_pT="Jet4_pT>20";
TCut
c_Jets_pT=c_Jet1_pT&&c_Jet2_pT&&c_Jet3_pT&&c_Jet4_pT;
//Other topological variable cuts
TCut c_Aplanarity="Aplanarity>0.05";
TCut c_Lepton_pT="Lepton_pT>20";
TCut c_MET="MET>20";

TCut
c_preselection=c_Jets_pT&&c_Aplanarity&&c_Lepton_pT&&c_
MET;
```

Figure 3.2: Cuts Applied to Signal and Background MC in TMVA Analysis Script

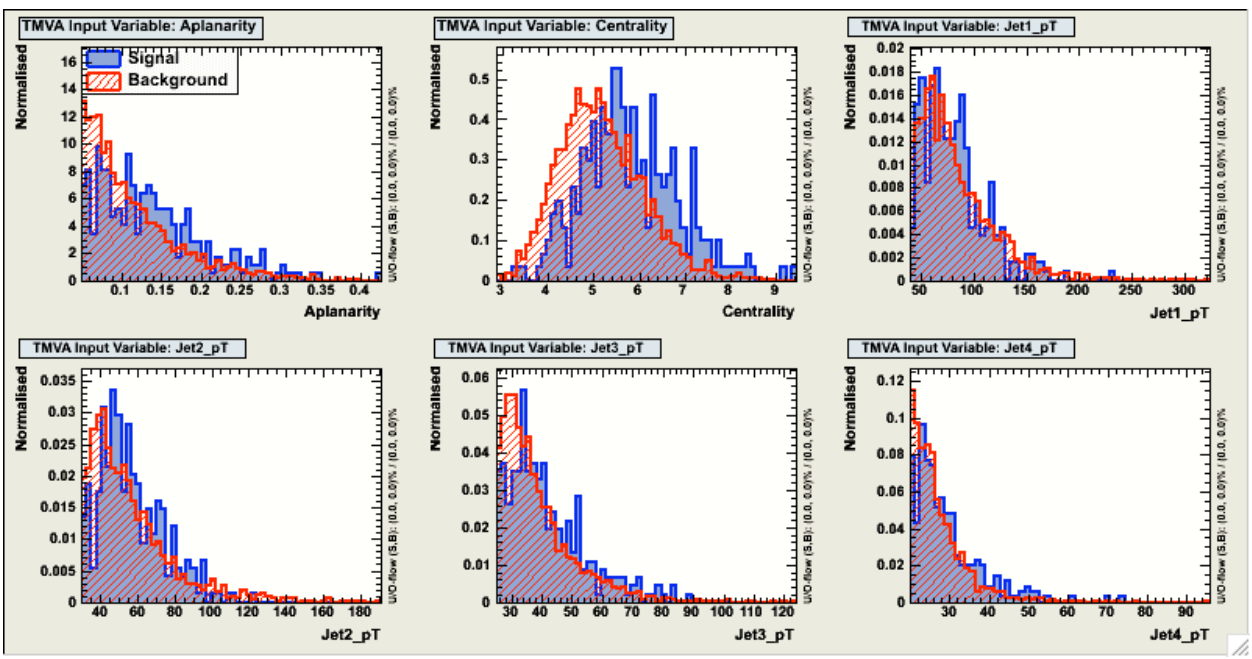

Figure 3.3: Event Variable Distribution

Signal MC: Muon Channel, Higgs Mass of 180 GeV

Background MC: w+5lp → lν + 5lp

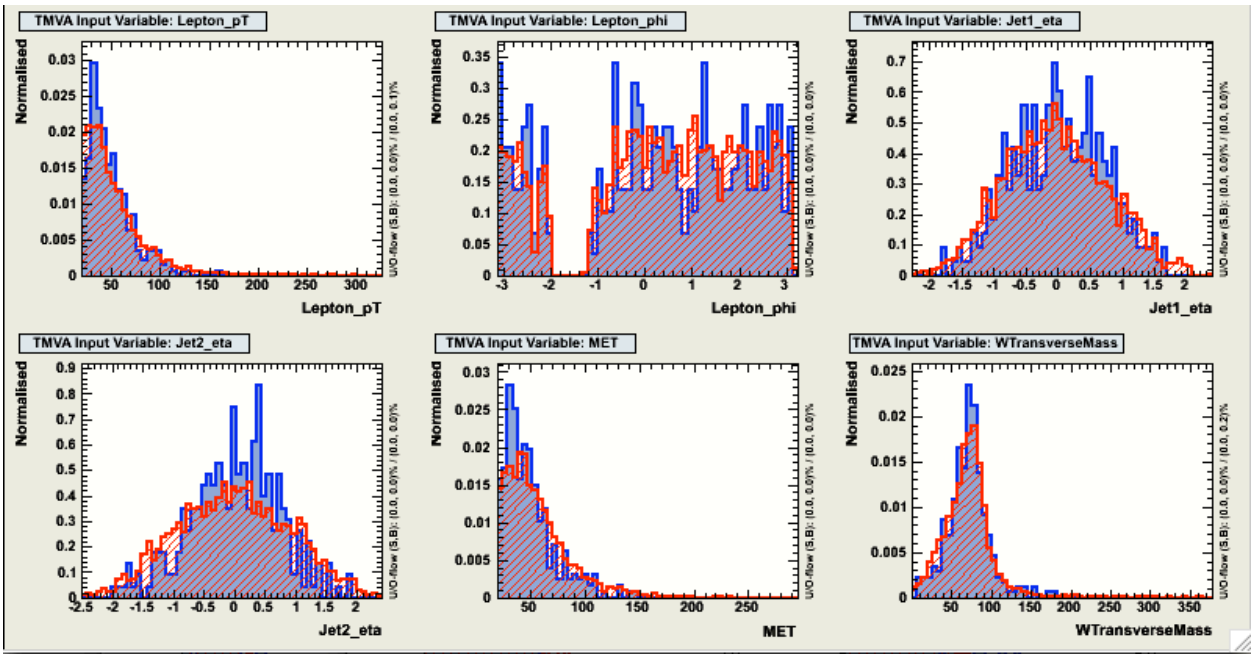

Figure 3.4 Event Variable Distribution

Signal MC: Muon Channel, Higgs Mass of 180 GeV

Background MC: w+5lp → lν + 5lp

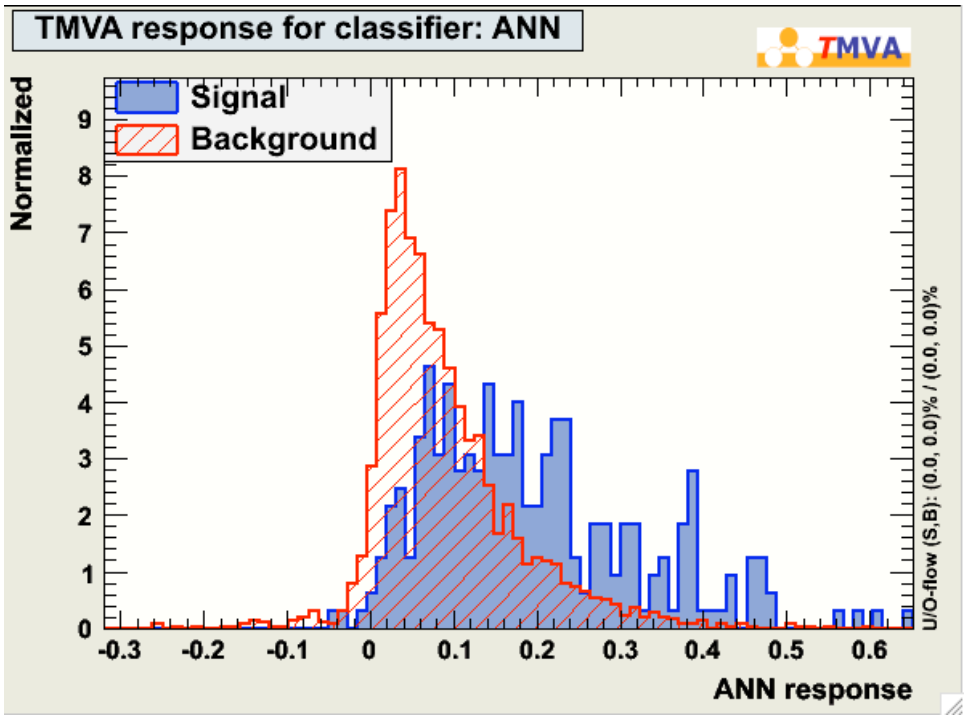

Figure 3.5: Neural Network Output Distribution

Signal MC: Muon Channel, Higgs Mass of 180 GeV

Background MC: w+5lp → lν + 5lp

wh_cafe ("addpkg –h wh_cafe" between late July and early August 2009):

wh_cafe/datavsmc/WH_Analysis.h

```
LooseWeight =
GetWeight()*dEta_weight*dPhi_weight*Lumi_weight*VCJ_weight*Eta1_weight*Eta2_weight*Eta3_weight*Mu_deteta_
weight*dR_weight*xsec_weight;
   TightWeight =
GetWeight()*dEta_weight*dPhi_weight*Lumi_weight*VCJ_weight*Eta1_weight*Eta2_weight*Eta3_weight*Mu_deteta_
weight*dR_weight*xsec_weight;
```

wh_cafe/datavsmc/WH_BaseAnalysis.h

```
float    GetWeight() const {
   float mc_xsec   = map_xsec [fDataName.Data()];
   float mc_nevent = map_event[fDataName.Data()];
   //return mc_nevent != 0. ? fEvt->weight*mc_xsec/mc_nevent*fLumi : 0.;
   return mc_nevent != 0. ? fEvt->weight : 0.;
```

wh_cafe/src/WHTree.cpp

```
_whevt->weight        = stat->eventWeight();
```

wh_cafe whtree_v01-01-00:

wh_cafe/datavsmc/WHTreeAnalyze.h

```
lo_weight = fEvt->weight*mc_xsec/mc_nevent*Lumi;
ti_weight = fEvt->weight*mc_xsec/mc_nevent*Lumi;
```

wh_cafe/src/WHTree.cpp

```
_whevt->weight        = stat->eventWeight();
```

wh_cafe (my working area-/rooms/JFK/projects/zachary_hynes/21_13_vjets_v03-04-00): version whtree_v01-01-00 with WHEvent.hpp, WHTree.cpp, and WHTree.hpp files come from /work/pepe-clued0/enari/WH/p2113_3.0.7_v0/ through clued0):

wh_cafe/wh_cafe/WHEvent.hpp

```
_whevt->weight  =stat->eventWeight()
```

Figure 3.6: Comparison of Event Weights Applied in wh_cafe whtree_v01-01-00 and New Version of wh_cafe Obtained with addpkg –h wh_cafe

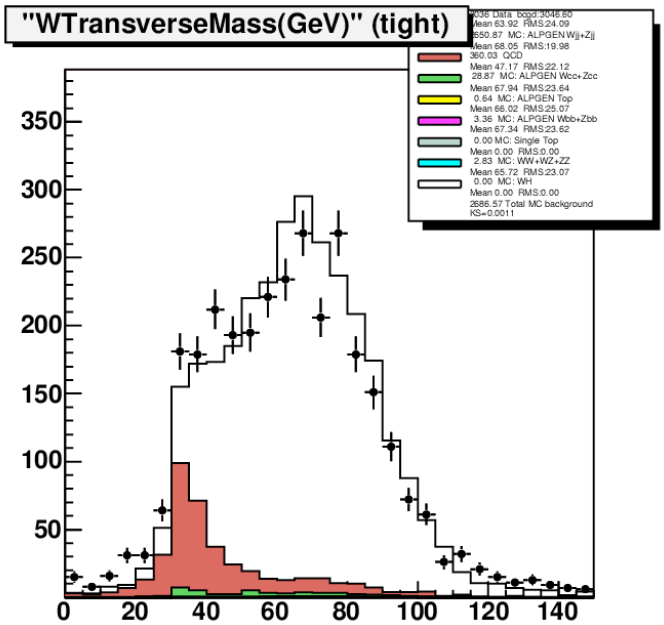

Figure 3.7: Leptonic W Boson Transverse Mass Output Distribution for Data and MC Backgrounds

---

[i] James, Randerson. "Father of the 'God Particle'" *The Guardian*. 30 June 2008. Web. <http://www.guardian.co.uk/science/2008/jun/30/higgs.boson.cern>.

[ii] Tevatron New Phenomena, Higgs Working Group, D0 Collaboration, and CDF Collaboration. "Combined CDF and DZero Upper Limits on Standard Model Higgs-Boson Production with up to 4.2 fb-1 of Data." *ArXiv*. 24 Mar. 2009. Web. 2 Aug. 2009. <http://arxiv.org/pdf/0903.4001v1>.

[iii] "D0 Fact Sheet." *The DZero Experiment*. Fermi National Accelerator Laboratory, Nov. 2008. Web. 3 Aug. 2009. <www-d0.fnal.gov>.

[iv] We have utilized two versions of ROOT throughout this analysis; ROOT version 4.04/02b is the version that is set up in our usual set-up scripts, but we have to use a more recent version in order to use TMVA v. 3.9.5. To obtain this version, we use the "setup root v5_22_00a –q GCC_3_4_3:dzero:rh7" after setting up our working area (see Figure 2.2 for my initial set-up script)

[v] Files "wh_cafe/wh_cafe/WHEvent.hpp", "wh_cafe/wh_cafe/WHTree.hpp," and "wh_cafe/src/WHEvent.hpp" all available at the following address on clued0: work/pepe-clued0/enari/WH/p2113_3.0.7_v0/

[vi] Yamada, Ryuji. PowerPoint Presentation "WH→W(WW) →lv.(jj.jj), or →jj.(jj.jj)". *WH_WWW1.ppt*. 18 Sep. 2008. Obtained from R. Yamada. 4 Aug. 2009.

[vii] Please refer to the TMVA User's Guide for more information on this package: Hoecker, A., et. al. "TMVA - Toolkit for Multivariate Data Analysis." *ArXiv*. 11 June 2007. Web. 4 Aug. 2009. <http://arxiv.org/pdf/physics/0703039v4>.

[viii] rundatavsmc.csh, tree_datavsmc.cc, and WHTreeAnalyze.h can all be read from the wh_cafe/datavsmc/ directory (this is the relative address from our D0RunII p21.13.00 release with the wh_cafe package)

[ix] Buehler, Marc, et. al. "Search for standard-model Higgs production in the lepton + jets final state for H->WW*→lvjj decay in 3.66 fb^{-1} – 4.97 fb^{-1} of Run II Data". Version 2.0.1. 30 Jul. 2009. D0 Note 5851. 4 Aug. 2009.

[x] Yamada, Ryuji. PowerPoint Presentation. "WH→W(WW) →lν.(jj.jj), or →jj.(jj.jj)". *WH_WWW1.ppt*. 18 Sep. 2008. Obtained from R. Yamada. 4 Aug. 2009.

[xi] Yamada, Ryuji. PowerPoint Presentation. "WH→W(WW) →lν.(jj.jj), or →jj.(jj.jj)". *WH_WWW1.ppt*. 18 Sep. 2008. Obtained from R. Yamada. 4 Aug. 2009.

[xii] Side View of the D0 Detector. Digital image. *Fermi National Accelerator Laboratory*. 23 May 2001. Web. 6 Aug. 2009. <http://www.fnal.gov/pub/now/live_events/images/D0sideview.gif>.