First Evidence for WW and WZ Diboson Production with Semi-Leptonic Decays at a Hadron Collider

Joseph Glenn Biddle Haley

A DISSERTATION PRESENTED TO THE

FACULTY OF PRINCETON UNIVERSITY

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

RECOMMENDED FOR ACCEPTANCE

BY THE DEPARTMENT OF PHYSICS

Advisor: Chris Tully

June 2009

UMI Number: 3356716

INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy
submitted. Broken or indistinct print, colored or poor quality illustrations and
photographs, print bleed-through, substandard margins, and improper
alignment can adversely affect reproduction.
In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if unauthorized
copyright material had to be removed, a note will indicate the deletion.

UMI®

To Joseph Brennan Haley 1947 - 2007,

the best dad a boy could have.

Abstract

Presented is a measurement of the simultaneous production of a $W^\pm$ boson in association with a second weak boson ($W^\pm$ or Z^0) in $p\bar{p}$ collisions at $\sqrt{s} = 1.96$ TeV. Events are consider with one electron or one muon, missing transverse energy, and at least two hadronic jets. The data were collected by the D0 detector in Run IIa of the Tevatron accelerator and correspond to 1.07 fb^{-1} of integrated luminosity for each of the two channels ($WW/WZ \to e\nu q\bar{q}$ and $WW/WZ \to \mu\nu q\bar{q}$). The cross section for $WW + WZ$ production is measured to be 20.2 $\pm$2.5(stat) $\pm$3.6(sys) $\pm$1.2(lum) pb with a Gaussian significance of 4.4 standard deviations above the background-only scenario. This measurement is consistent with the Standard Model prediction and represents the first direct evidence for WW and WZ production with semi-leptonic decays at a hadron collider.

The work described in this dissertation
has been published in the following journals:

Contents

5   Event Selection   63

6   Monte Carlo Corrections   67

7   Multivariate Discrimination   86

Acknowledgments

Thank you mom and dad for your everlasting love, support, and guidance. To my wife, thank you for your love and patience. I owe a great debt of gratitude to Wade Fisher for his invaluable guidance as I navigated through my graduate research and in particular his collaboration on this analysis. I also thank Jadranka Sekaric for her contributions to the work presented in this thesis. To my advisor, Chris Tully, I thank you both for your direction and your freedom throughout my graduate career. I thank all of the wonderful professors, as an undergraduate at the University of Washington and as a graduate student at Princeton University, who contributed to my wealth of physics knowledge. Finally, much thanks is owed to the D0 Collaboration, the Fermi National Accelerator Laboratory, and all supporting entities for supplying the resources that made this research possible.

List of Tables

List of Figures

List of Abbreviations

E_T "Transverse" Energy $\equiv E/\cosh(\eta) \equiv E\sin(\theta)$

$\vec{\not{E}}_T$ Imbalance in Transverse Momentum ("Missing E_T")

CC   Central Calorimeter

CFT   Central Fiber Tracker

CH   Course Hadronic

CPS   Central Preshower detector

EC   End Calorimeter

EM   Electromagnetic

FAMUS   Forward Angle Muon System

Fermilab   Fermi National Accelerator Laboratory

FH   Fine Hadronic

FPS   Forward Preshower detector

FV   Final Variable used for the measurement

ICR   Inner-Cryostat Region

JES   Jet Energy Scale

LO   Leading Order

MC   Monte Carlo

MDT   Mini Drift Tube

MPF   Missing E_T Projection Fraction

MPI   Multiple Parton Interaction

NLL   Negative Log-Likelihood

NLO   Next-to-Leading Order

NNLO   Next-to-Next-to-Leading Order

PDT   Proportional Drift Tube

SM   Standard Model of particle physics

SMT   Silicon Microstrip Tracker

TGC   Trilinear Gauge-boson Couplings

WAMUS Wide Angle Muon System

Chapter 1

Introduction

1.1 The Standard Model

The theoretical framework used to describe the high energy particle physics is the so-called Standard Model of particle physics (SM). The SM combines quantum mechanics and special relativity into a theory describing the known fundamental particles and, with the exception of gravity, the interactions between particles. The fundamental particles in the SM consist of quarks and leptons that make up the matter, gauge bosons that mediate the forces, and the as-yet-undiscovered Higgs boson that is believed to exist in order to explain how some particles have mass. These particles are considered fundamental because they do not contain any known substructure (*i.e.*, they are not made up of any finer pieces). They are the basic building blocks for the known universe.

1.1.1 Particles

At the most basic level, matter is composed of leptons and quarks bound together by the force-carrying gauge bosons. As shown in figure 1.1, there are a total of six leptons and 18 quarks (counting color charge), plus corresponding anti-particles. An interesting property is that nature appears to have repeated itself three times by making three generations of

1

quarks and leptons with identical properties across each generation except for an increase in mass. For example, the electron (e), muon (μ), and tau (τ) are identical in every way except mass. Most matter found in nature is only composed of e^-, u quarks, and d quarks from the first generation because the more massive particles from the second and third generation quickly decay into lighter particles.
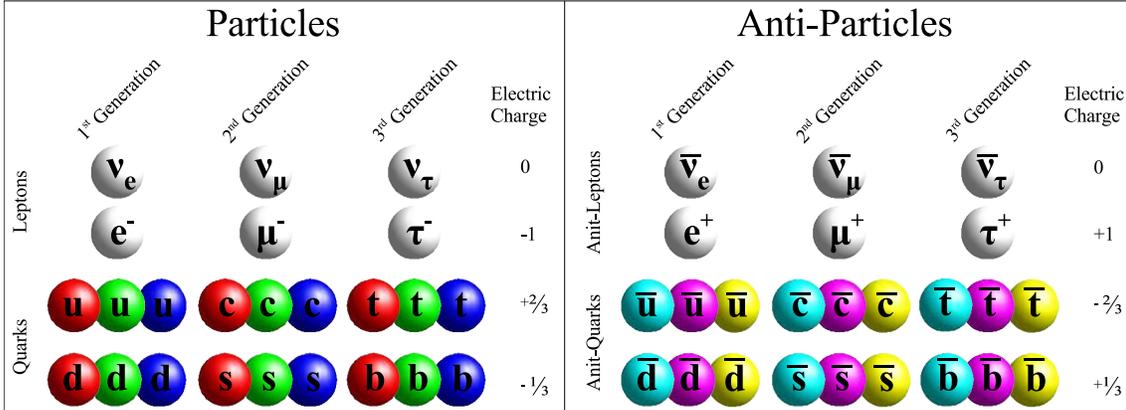
Figure 1.1: Quarks and leptons, the fundamental matter particles.

The primary difference between leptons and quarks is that quarks have color-charge and therefore interact via the strong force while leptons do not. For this reason, single quarks are not found in nature because the strong force binds quarks together until the total color-charge is neutral ("white" or "black"). Therefore, quarks are only found in groups of three quarks with one quark of each color or in quark-antiquark pairs with the quark of a given color and the antiquark of the corresponding anti-color. Three quark groups are called Baryons, such as the proton (uud) and neutron (udd), and quark-antiquark pairs are call Mesons, such as the π^+ ($u\bar{d}$) and K^+ ($d\bar{s}$).

The remaining particles in the SM are the photon (γ), W boson (W), Z boson (Z), Higgs boson (H), and eight gluons (g). The photon interacts with all particles that have electric charge and causes opposite charges to attract and like charges to repel. This results in the familiar electromagnetic force. In a similar manner, the W and Z bosons are responsible for the weak force and the eight gluons mediate the strong force.

The role of the Higgs boson is a little more subtle than the other fundamental bosons.

The interaction of the Higgs boson with a particle results in that particle having mass. Massless particles such as photons and gluons do not directly interact with the Higgs boson, while massive particles such as the electron or W boson do interact with the Higgs boson. The Higgs boson is the only particle in the SM that has not been directly observed by experiment. Experimental confirmation of the existence of the Higgs boson, or alternatively of new physics beyond the SM, are the drive for many research efforts in the field of high energy physics.

1.2   WW and WZ Production

This thesis is dedicated to measuring the production of events containing a W boson that decays leptonically ($W \rightarrow \ell\nu$, $\ell = e$ or μ) in association with a second weak boson (W or Z) that decays hadronically ($W/Z \rightarrow q\bar{q}$). The leading order (LO) Feynman diagrams for these processes are shown in figure 1.2. Next-to-leading order (NLO) calculations for the production of WW and WZ at $\sqrt{s} = 1.96$ TeV predict the cross sections $\sigma(p\bar{p} \rightarrow WW) = 12.4 \pm 0.8$ pb and $\sigma(p\bar{p} \rightarrow WZ) = 3.7 \pm 0.3$ pb [1].
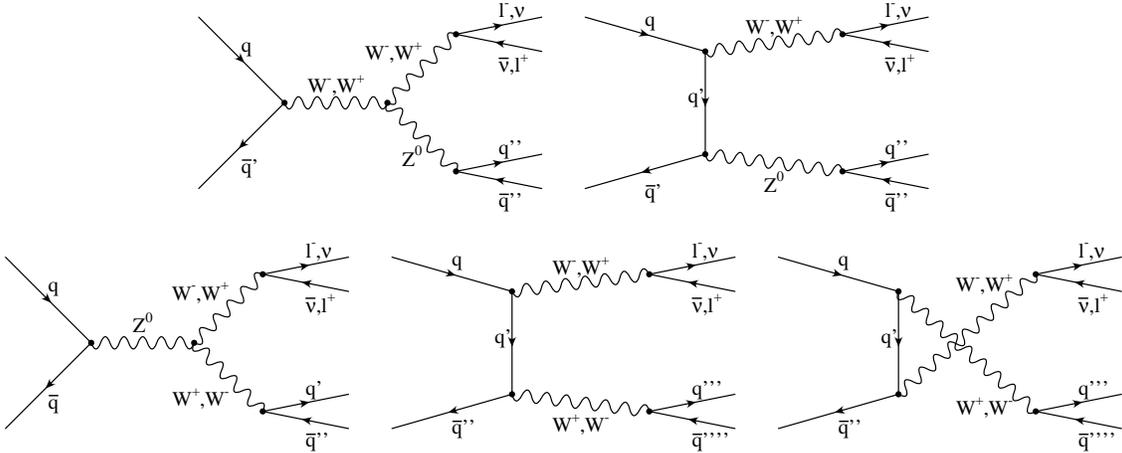
Figure 1.2: Leading order Feynman diagrams for the processes studied in this analysis.

The simultaneous production of two weak gauge bosons is a physical process involved in a large range of measurements at the Tevatron. A primary motivation for studying diboson

physics comes from the fact that their production and interactions provide a test of the Electroweak sector of the SM. Diboson production can be studied by measuring production cross sections and/or their interactions via trilinear gauge-boson couplings (TGC) [2, 3, 4]. The deviation of a TGC or production cross section from the values predicted by the SM would be an indication of new physics beyond the SM and could give some clues about the mechanism responsible for Electroweak Symmetry Breaking.
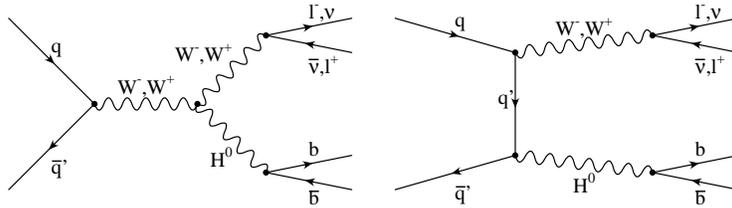
Figure 1.3: Leading order Feynman diagrams for $W^{\pm}H^0 \rightarrow \ell\nu b\bar{b}$.

Furthermore, the SM Higgs boson search landscape is marked by several aspects of diboson production. One of the most promising channels for discovering a low mass Higgs boson $(m_H \lesssim 130 \text{ GeV})$[1] is in associated production with a W boson that decays leptonically $(W^{\pm}H^0 \rightarrow \ell\nu b\bar{b})$. As shown in figure 1.3, the Feynman diagrams for $W^{\pm}H^0 \rightarrow \ell\nu b\bar{b}$ are notably similar to those for $W^+W^-/W^{\pm}Z^0 \rightarrow \ell\nu q\bar{q}$. In both cases, the final state particles are a lepton and neutrino from the decay of a W boson and a quark-antiquark pair from the decay of either the Higgs boson or a weak gauge boson (W or Z). One consequence of this similarity is that $W^+W^-/W^{\pm}Z^0 \rightarrow \ell\nu q\bar{q}$ is an important background for these Higgs searches and making this direct measurement of diboson production supplies an *in situ* measurement of the size of this background. A more important consequence, however, is that this thesis represents a benchmark for these Higgs searches with similar final states. The Higgs boson searches and this analysis share the same challenge of separating a small signal from a large background. Of course, this analysis has the advantage of a signal cross section that is many times larger. Additionally, the ability to discover the Higgs boson (and to mea-

---

[1]This thesis uses the convention $c = 1$. Unless otherwise stated, units of GeV are used for energy, momentum, and mass.

sure $W^+W^-/W^\pm Z^0 \to \ell\nu q\bar{q}$) depends largely on the capacity to reconstruct the resonance in the dijet invariant mass. A measurement of $W^+W^-/W^\pm Z^0 \to \ell\nu q\bar{q}$ provides a scenario to determine the actual resolution of a dijet resonance.

Weak diboson production is also a significant background for high mass SM Higgs boson searches ($m_H \gtrsim 140$ GeV), in which the searches focus on $H \to W^+W^-$ decays. As in the low mass Higgs scenario, both the magnitude and kinematics of diboson production impact the power of the search.

A measurement of $W^+W^-/W^\pm Z^0 \to \ell\nu q\bar{q}$ production provides a "standard candle" with which to calibrate and optimize many of the techniques used in SM Higgs searches. The event selection shares most of the trigger, Monte Carlo (MC) simulation, and normalization methods of the $H \to b\bar{b}$ search. The multivariate event classification schemes that are increasingly common in Higgs boson searches can also be vetted using a known signal. Finally, the statistical search techniques used for the entire SM Higgs mass range can be tested on this known signal, providing opportunities for optimization.

Prior to this analysis, W^+W^- and $W^\pm Z^0$ production had been measured using only the purely leptonic final states. The most recent D0 analysis of semi-leptonic decays of $W^+W^-/W^\pm Z^0$ is from Run I with $\approx$80 pb^{-1} of data [5], in which no statistically significant signal was observed. Thus the reconstruction of W^+W^- and $W^\pm Z^0$ events in their semi-leptonic decay modes ($W^+W^-/W^\pm Z^0 \to \ell\nu q\bar{q}$) represents a challenge regarding the separation of signal from the dominant backgrounds, a challenge shared by Higgs boson searches.

Chapter 2

Apparatus

The data analyzed in this thesis were collected by the D0 detector, which records collisions of protons with antiprotons produced by the Tevatron accelerator located at Fermi National Accelerator Laboratory (Fermilab) in Batavia, IL [6, 7, 8, 9, 10].

2.1 Accelerator Chain

The Tevatron accelerator brings protons and antiprotons into collision, but that is only one link in a long chain responsible for creating, steering, cooling, and accelerating the protons and antiprotons. An illustration of this chain of accelerator components is shown in figure 2.1. The following section retraces the path of a proton that starts in hydrogen gas and ends in a collision with an antiproton. The subsequent section describes the process for producing antiprotons and bringing them into collision with the protons.

2.1.1 Path of a Proton

Each proton starts out in the form of hydrogen gas (H_2) that is converted into negatively charged hydrogen ions (H^-) by a magnetron [11]. A diagram of the magnetron is shown in figure 2.2. The magnetron is composed of an oval-shaped cylindrical cathode oriented
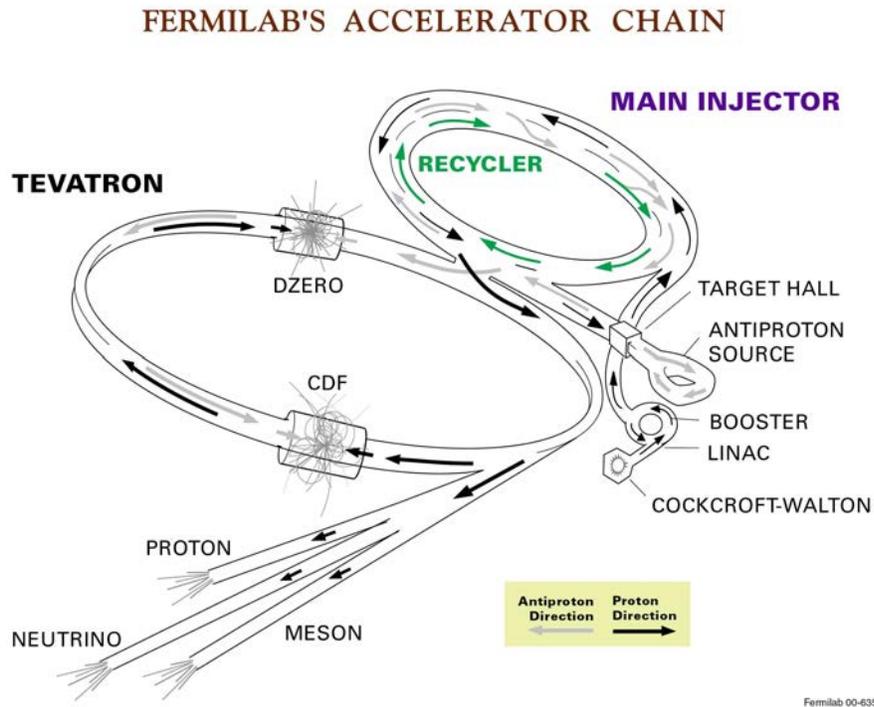
Figure 2.1: Schematic of Fermilab accelerator chain.
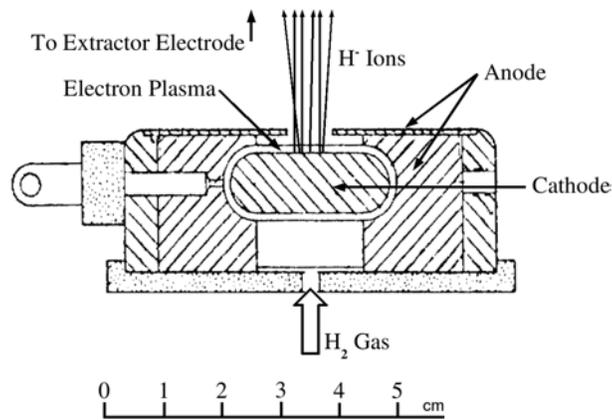
Figure 2.2: Diagram of magnetron used to generate H$^-$ ions.

parallel to a magnetic field and surrounded by an anode. Hydrogen gas is pumped into the gap and a potential difference of a few hundred volts is pulsed with a duration of 60 μs repeated at a rate of 15 Hz. During each pulse, electrons are stripped and spiral tightly in the magnetic field producing a plasma in the gap between the anode and cathode. The protons bend much less and hit the cathode, sometimes picking up two electrons, becoming H^- ions, and then accelerate toward the anode. An aperture in the anode allows some of the H^- ions to escape and accelerate to the extractor electrode.
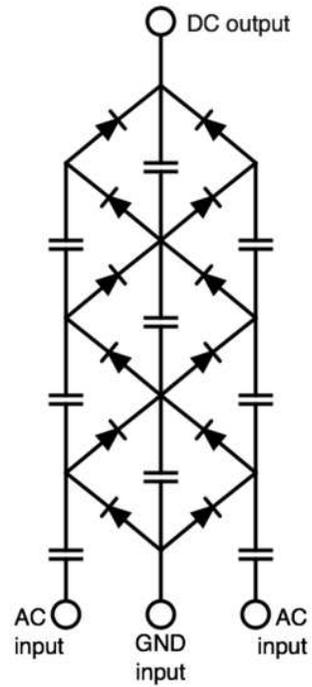
<table>
<tr><td>(a)</td><td>(b)</td></tr>
</table>

Figure 2.3: (a) Image of Cockroft-Walton pre-accelerator at Fermilab. (b) Circuit diagram of a Cockroft-Walton voltage multiplier.

After the magnetron is a Cockroft-Walton pre-accelerator (figure 2.3). The legs of this device form a voltage ladder that charges the dome on top to -750 kV. A resistive tunnel is grounded at one end and has the other end connected to the -750 kV dome, setting up a large static electric field inside. The H^- ions enter the dome and accelerate down the tunnel to an energy of 750 keV.

Next the H$^-$ ions are accelerated by the Linac and sent to the Booster. The Linac [12] is a linear accelerator that accelerates the H$^-$ ions is two stages. The first, older stage is 79 m long and uses five 5 MW amplifiers to boost the H$^-$ ions to 116 MeV. The second, newer stage (upgraded in 1993) is 67 m long and uses seven 12 MW amplifiers to boost the H$^-$ ions to 400 MeV.

The Booster [13] is the first of many synchrotrons in Fermilab's accelerator chain. The Booster uses non-superconducting electromagnets to bend protons around a 75 m radius ring. As the H$^-$ ions enter the Booster the electrons are stripped off of by a carbon foil leaving the bare protons to circulate in the ring. Once the Booster has accumulated enough protons ($\sim 3 \times 10^{12}$) the Linac stops delivering beam and the Booster begins accelerating. As with all synchrotrons, the strength of each magnet is synchronized with the energy of the protons in order to maintain the same orbit as the protons are accelerated. The Booster is able to accelerate the protons up to 8 GeV, limited by the dynamic range of the magnets for the given orbit.

Once the protons reach 8 GeV they are transfered to the Main Injector [14], which is also a synchrotron, but with a radius of approximately 528 m. The Main Injector accelerates the protons from 8 GeV to 150 GeV, then injects them into the Tevatron in a series of 36 bunches. The 36 bunches are divided into three superbunches that are separated by 2.64 μs with each superbunch containing 12 bunches that are separated by 396 ns. The Main Injector is also involved in making and accelerating antiprotons, which is discussed in the next section.

The Tevatron is a synchrotron approximately 1 km in radius. Unlike the other synchrotrons, the Tevatron uses superconducting electromagnets to bend and focus the circulating beams. The Tevatron accelerates the protons to their final energy of 980 GeV before initiating collisions with the antiprotons.

2.1.2 Path of an Antiproton

When creating antiprotons, protons in the Main Injector are accelerated to 120 GeV instead of 150 GeV. The 120 GeV protons are sent to the antiproton source where they collide with a nickel target resulting in a shower of hadrons, including some antiprotons. A lithium lens carrying 0.5 MA of current creates an azimuthal magnetic field that focuses the negatively charged particles. Next a bending magnet selects antiprotons that have an energy of approximately 8 GeV and sends them to the Debuncher.

The Debuncher is another synchrotron, though somewhat triangular in shape, with a circumference of about 505 m. The antiprotons arriving at the Debuncher come in bunches (due to the bunch structure of the protons from the Main Injector) and have a spread of energies. The higher(lower) energy particles follow a longer(shorter) path around the Debuncher and, therefore, complete one revolution later(earlier) than 8 GeV antiprotons. Each revolution an RF cavity supplies a small kick to reduce the spread in energy, thus decreasing the emittance and "cooling" the beam. After about 100 ms the Debuncher produces a continuous beams of antiprotons with a small spread in energy around 8 GeV.

These antiprotons are moved to the Accumulator synchrotron, which sits just inside the Debuncher with a circumference of 474 m. Here the antiprotons are stored and further cooled. After many transfers from the Debuncher the Accumulator fills up to the point where adding more antiprotons to the Accumulator becomes increasingly inefficient. Once full, the Accumulator empties the antiprotons into the Recycler and begins accumulating again.

The Recycler is an antiproton storage ring that sits directly below the Main Injector. Groups of 8 GeV antiprotons from the Accumulator are stored in the Recycler until they are needed for a new store[1]. The Recycler also cools the beam even further using a process called electron cooling [15]; in which a low emittance (*i.e.*, cool) electron beam is circulated with

---

[1]A store is a round of high-energy proton-antiproton collisions starting when the protons and antiprotons are injected into the Tevatron and usually lasting around 24 hours, at which time the beams are dumped in preparation for a new store.

the antiproton beam, thus cooling the antiproton beam as it comes into thermal equilibrium with the electron beam.

When it comes time for a new store (round of collisions), the antiprotons are transfered from the Recycler to the Main Injector where they are accelerated from 8 GeV to 150 GeV. Like the protons, the antiprotons are loaded into the Tevatron as 36 bunches (three super-bunches with 12 bunches per superbunch) and then accelerated to 980 GeV. The protons and antiprotons share the same tunnel with the protons circulating clockwise and the antiprotons counterclockwise. The orbits of the proton and antiproton beams helix around each other except for the two locations where the beams are brought into collision, the center of the CDF detector and the center of the D0 detector.

2.2   D0 Detector

The D0 detector [16, 17, 18] is a multipurpose detector primarily designed for measuring high momentum final state particles resulting from decays of high mass particles such as the W and Z bosons, the top quark, or the Higgs boson. The basic purpose of the detector is to identify (*e.g.*, as an electron, muon, etc.) every final state particle and determine it's four-momentum and point of production[2]. To accomplish this task, the detector is composed of layers of subdetectors as shown in figure 2.4. Starting from the collision point and moving out, the components are:

- The central tracking system with silicon and fiber tracking layers and a 2 T magnet. The central tracking system is used to determine the location of the collision point, measure the paths and momenta of charged particles, and help with particle identification.

- The preshower detectors, which are used for particle identification and to aid in charged

---

[2]Of course, neutrinos (and possibly some non-SM particles) will escape the detector without interacting, but that just makes it all the more important to determine the four-momenta of the detectable particles so that the four-momenta for neutrinos can be accurately inferred from conservation of momentum.

particle tracking.

- The calorimeter system with electromagnetic (EM) and hadronic layers. The calorimeter measures the position and energies of electromagnetic showers from electrons and photons and hadronic showers resulting from quarks produced in the hard interaction.

- The muon system composed of scintillators, drift tubes, and toroidal magnets used to identify and measure the momentum and position of muons.
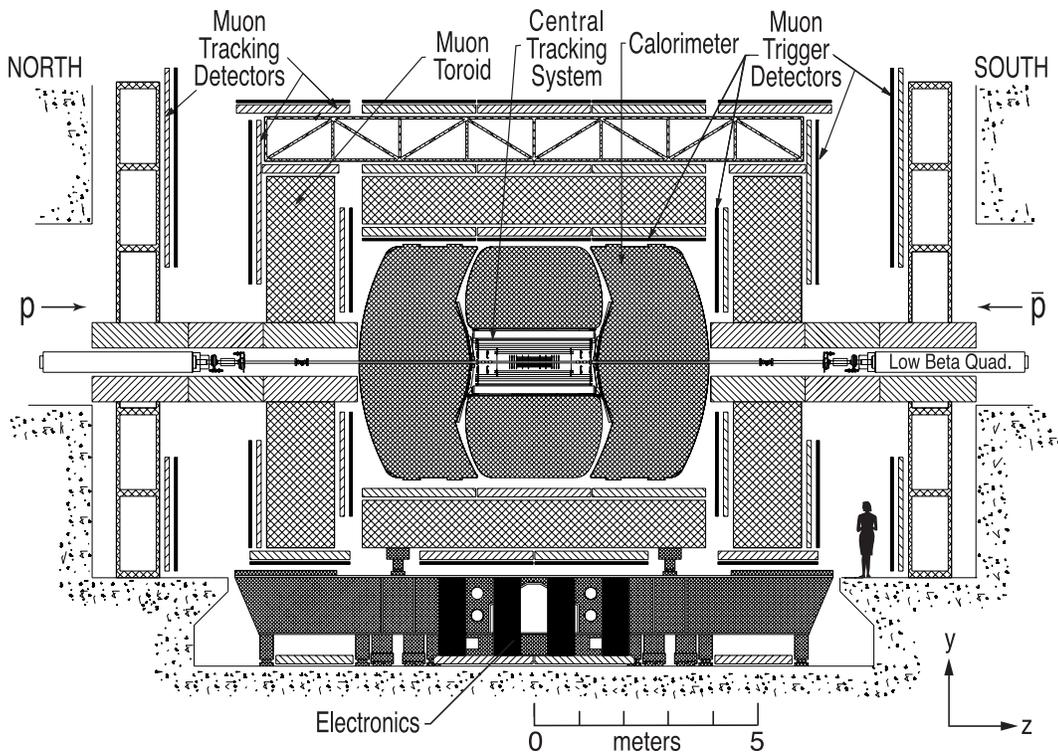
Figure 2.4: Diagram of D0 detector.

2.2.1   D0 Coordinate System

Before going any further, the coordinate system used by the D0 detector and throughout the rest of this thesis must be defined. The detector coordinate system is positioned with the origin at the center of the D0 detector and the positive direction of z-axis pointing in the direction of the proton beam. The positive x-axis points out from the center of the Tevatron

ring and the positive y-axis points up (towards the sky). In cylindrical coordinates, the transverse distance is given by $r = \sqrt{x^2 + y^2}$, the azimuthal angle measure from the x-axis is denoted ϕ, and the polar angle measure from the z-axis is denoted θ. It is usually more convenient, however, to use the pseudorapidity defined as

$$\eta = -\ln\left(\tan\frac{\theta}{2}\right) \tag{2.1}$$

instead of the polar angle. In the relativistic limit ($E \gg m$), η is a good approximation of the rapidity of the particle

$$y = \frac{1}{2}\ln\left(\frac{E + p_z}{E - p_z}\right), \tag{2.2}$$

which is a Lorentz invariant quantity. Furthermore particles tend to be produced more or less uniformly in rapidity, which is why the D0 detector is segmented in equal units of pseudorapidity and why η is a more convenient coordinate than θ.

Proton-antiprotons collisions do not usually coincide exactly with the center of the detector coordinate system. The displacement is small in the radial direction ($\sim 40~\mu$m), but can be substantial in the z direction ($\sim$28 cm). Additionally, the magnetic fields in the detector will curve the paths of charged particles. For these reasons, it is sometimes important to distinguish between "physics" and "detector" pseudorapidity and azimuth. Physics pseudorapidity (azimuth), η_{phys} (ϕ_{phys}), is the pseudorapidity (azimuth) of the reconstructed particle's four-momentum at the point of creation. Detector pseudorapidity (azimuth), η_{det} (ϕ_{det}), is measured with respect to the detector origin and indicates the location of the particle in the detector. In general, detector pseudorapidity and azimuth are used when referring to a location in the detector and physics pseudorapidity and azimuth are used when referring to the properties of a particle; thus the subscripts "phys" and "det" will be omitted except in circumstances that require clarification.

2.2.2 Central Tracking System

The central tracking system is comprised of the Silicon Microstrip Tracker (SMT) surrounded by the Central Fiber Tracker (CFT), which in turn is surrounded by a superconducting solenoid magnet. A digram of one quadrant of the central tracking system is shown in figure 2.6.
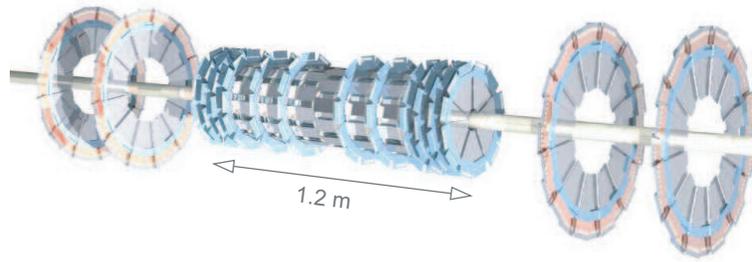
Figure 2.5: Computer generated image of the SMT.

The SMT [19, 20] is composed of silicon microstrip detectors that contain 300 μm thick silicon wafers segmented into parallel strips. The microstrip detectors are arranged in 12 F-disks, four H-disks, and six barrel segments as shown in figure 2.5. Each barrel segment is 12 cm long and contains four concentric cylindrical layers of microstrip detectors, two single sided layers and two double sided layers. The double sided layers have one axial side (microstrips oriented parallel to the beam line) giving azimuthal position of tracks, and one stereo side making an angle of 2° or 90° with respect to the beam line and providing a measurement of track pseudorapidity. The single sided barrel layers are axial. The F-disks are made from 12 wedges of double sided microstrip detectors with each side offset by an angle of ±15° from radial. The H-disks are made from two layers of 12 single sided microstrip detectors with each layer offset by ±7.5° from radial.

The CFT [21] is composed of eight concentric cylinders ranging in radius from 20 cm to 52 cm. As shown in figure 2.6, each cylinder has two double-layers of scintillating fibers, an axial double-layer on the inside and a stereo double-layer on the outside. Fibers in the stereo double-layers make an angle of ±3° (alternating +3° and -3° each cylinder) relative to the

14
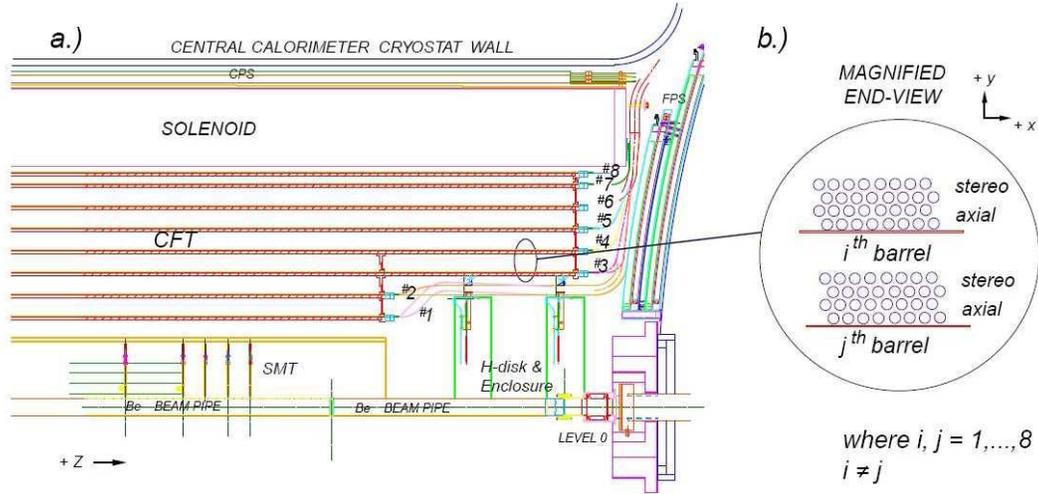
Figure 2.6: Diagrams illustrating the layers of scintillating fibers in the CFT.

axial layers. Each double-layer is composed of 835 μm diameter scintillating fibers arranged in adjacent, parallel layers with the second layer overlapping the gaps of the first layer.

The superconducting solenoid magnet [22] is 2.73 m in length with an outer radius of 71.0 cm and made from niobium-titanium wire that is cooled to less than 4.7 K using liquid helium. The solenoid produces a 2 T magnet field parallel to the z-axis that causes charged particles to curve as they travel radially out from the interaction point. The radius of curvature in meters is give by

$$R = \frac{p_T}{0.3qB};\tag{2.3}$$

where p_T is the transverse momentum in units of GeV, q is the charge of the particle in units of electron change (e), and B is the magnetic field in Tesla; thus providing a measurement of the momenta of charged particles passing through the SMT and CFT.

Besides providing a measurement of momentum, the central tracking system is important for particle identification. Charged particles such as electrons and muons will leave isolated tracks. Neutral particles such as photons and neutrinos will leave no tracks. Particles produced with color charge such as quarks and gluons will produce a shower or "jet" of particles that result in a group of many tracks.

2.2.3 Preshower Detectors
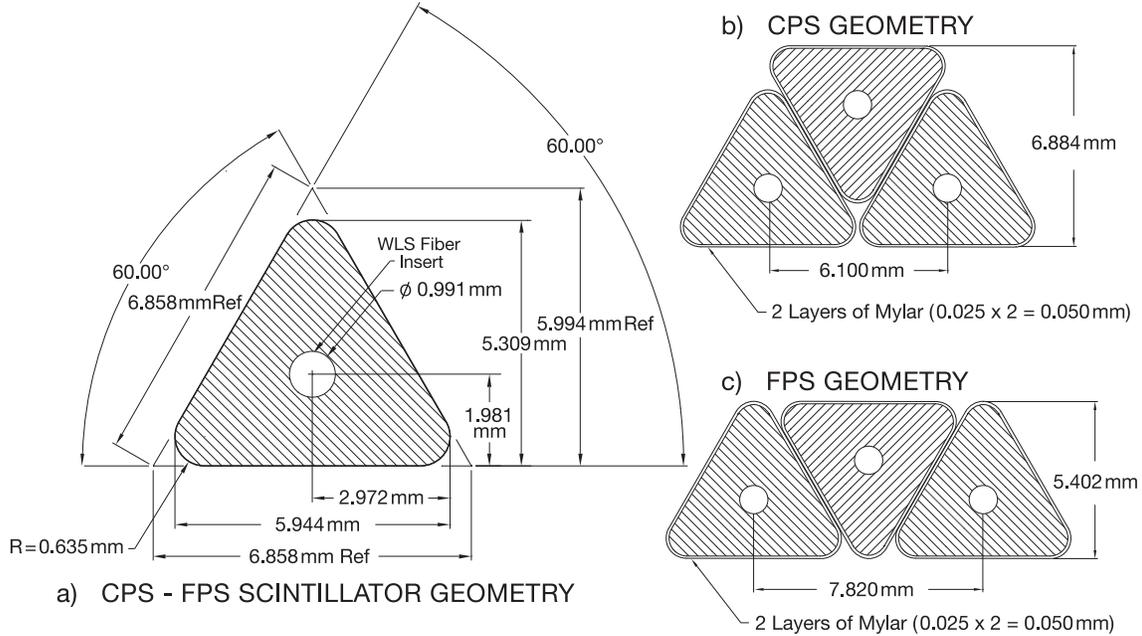
Figure 2.7: (a) Cross section diagram of a scintillating fiber used in the preshower detectors. (b) Diagram showing how fibers are stacked in each layer of the CPS. (c) Diagram showing how fibers are stacked in each layer of the FPS.

Wrapping around the solenoid is the Central Preshower detector (CPS) and at each end of the central tracking system is a Forward Preshower detector. The preshower detectors use scintillating fibers with a triangular cross section arranged into layers as shown in figure 2.7.

The CPS has a layer of lead followed by three layers of scintillating fibers, one axial layer then two stereo layers with stereo angles of 24º then -24º. The layer of lead along with the solenoid results in a thickness of approximately two radiation length at η=0 and increasing to around four radiation lengths near the end of the CPS at $|\eta|$=1.3. Electrons and photons will begin showering before reaching the scintillators. Thus the preshower detectors are used to help distinguish EM particles from hadronic particles such as pions.

Each FPS (one at each end) is divided into 16 wedges. Each wedge has two layers of scintillating fibers, then a layer of lead two radiation lengths thick, then two more layers of scintillating fibers as shown in figure 2.8. EM particles will not shower until they reach the layer of lead, therefore, the first two layers of the FPS are used for precision position

16
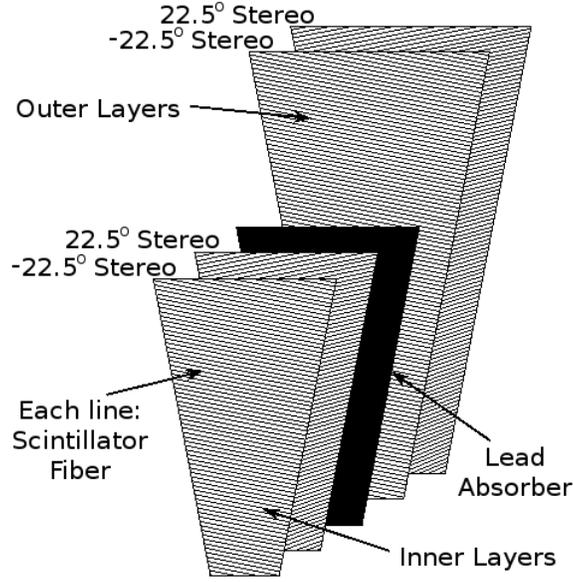
Figure 2.8: Illustration of one wedged from the FPS showing the orientation of the scintillator fibers.

measurements to assist the central tracking system, while the two layers outside the lead are used to help with electron identification.

2.2.4 Calorimeter

Surrounding the central tracking system and preshower detectors is the D0 calorimeter [16]. The calorimeter measures the energies of electrons, photons, and jets (usually the result of high energy quarks and gluons from a hard interaction). As illustrated in figure 2.9, the calorimeter is divided into three separate pieces, a Central Calorimeter (CC) and two End Calorimeters (ECs), each encased in a cryostat cooled to 90 K. The CC covers out to $|\eta| <1.1$ and the two ECs extend the coverage to around $|\eta| <4$.

The calorimeter is made up of many readout layers allowing energy deposits to be sampled at various depths, thus providing additional information for particle identification. The inner four readout layers of the CC and ECs are the EM layers designed to measure electron and photon energies. The next three layers in the CC are fine hadronic (FH) layers and the outer most layer is the course hadronic (CH) layer. The hadronic layers in the ECs (figure 2.9) are
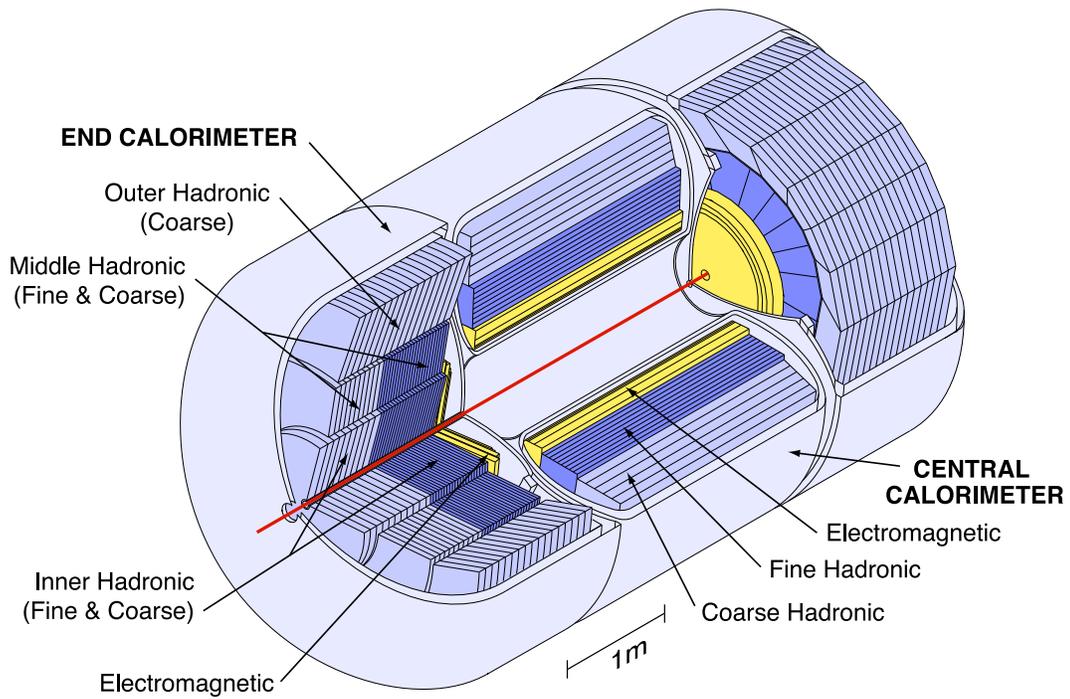
Figure 2.9: Illustration of the D0 calorimeter.
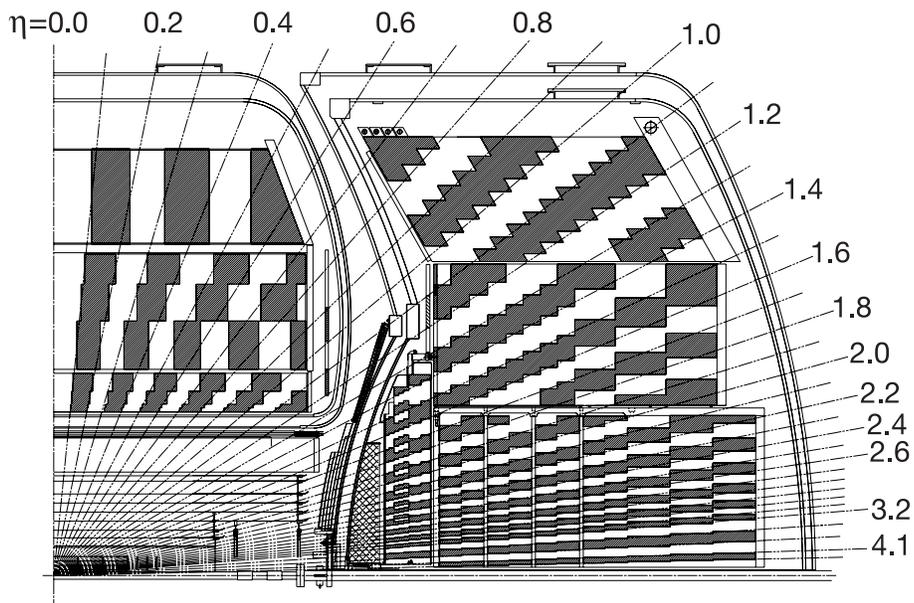
Figure 2.10: Diagram showing calorimeter towers in η.

separated into inner fine hadronic (IFH), inner course hadronic (ICH), middle fine hadronic (MFH), middle course hadronic (MCH), and outer hadronic (OH). Each readout layer is segmented into cells of approximately 0.1×0.1 in $\eta \times \phi$ space. The two exceptions are the third EM layer, which has 0.05×0.05 segmentation for more accurate measurements of EM showers, and the ECs with $|\eta| > 3.2$, which has reduced granularity due to space constraints. The cells in each layer are stacked to form 0.1×0.1 towers in $\eta \times \phi$ as shown in figure 2.10.
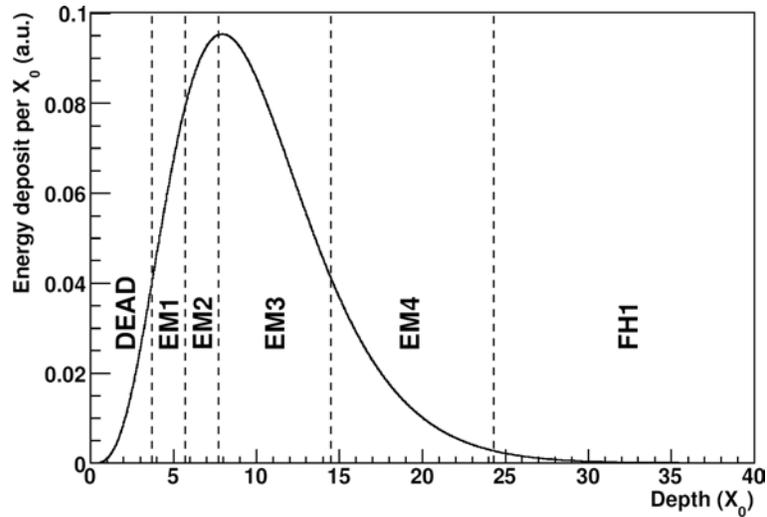
Figure 2.11: Expected energy profile as a function of depth for the shower of a 45 GeV electron. Also shown are the corresponding layers in the calorimeter.

An electron or photon moving through the calorimeter will produce an EM shower through repeated bremsstrahlung and electron-positron pair production. When the energies of the electrons in the shower fall below a critical value ionization becomes dominant over bremsstrahlung and the showering stops. The extent of an EM shower in a particular material is characterized by the radiation length, X_0, which is the average distance an electron or photon will travel before losing 63% $(1 - e^{-1})$ of it's energy. The EM layers have a total thickness of approximately $20.5X_0$ near $\eta = 0$. As shown in figure 2.11, most of the energy from an electron will be captured by EM layers with just the tail end making it into the first hadronic layer.

Jets are composed mostly of pions and kaons that interact via the strong nuclear force. As the hadrons travel through the calorimeter they scatter off nuclei producing more hadrons,

mostly pions. The neutral pions will quickly decay to photons resulting in some EM showering while the charged pions will continue to scatter off nuclei and produce more hadrons resulting in a hadronic shower. The extent of a hadronic shower is defined by the nuclear interaction length, λ_I, analogous to the radiation length for EM showers. However, for a given material λ_I is usually much longer than X_0. In particular, the total thickness of the EM layers is only $0.76\lambda_I$ in the CC and $0.97\lambda_I$ in the ECs. Therefore, it takes much more material to contain a hadronic shower. A summary of the properties of the various calorimeter layers is given in tables 2.1 and 2.2.

Table 2.1: Properties of CC layers.

	EM	FH	CH
Number of readout layers	4	3	1
Signal boards per readout layer	2, 2, 7, 10	20, 16, 14	9
Absorber material	Uranium	Uranium-Niobium	Copper
Radiation lengths (X_0)	20.5	96.0	32.9
Nuclear interaction lengths (λ_I)	0.76	3.2	3.2

Table 2.2: Properties of EC layers.

	EM	IFH	ICH	MFH	MCH	OH
Number of readout layers	4	4	1	4	1	3
Signal boards per readout layer	2, 2, 6, 8	16	14	15	12	8
Absorber material	Uranium	Uranium-Niobium	Stainless Steel	Uranium-Niobium	Stainless Steel	Stainless Steel
Total radiation lengths (X_0)	20.5	121.8	32.8	115.5	37.9	65.1
Total nuclear interaction lengths (λ_I)	0.97	4.9	3.6	4.0	4.1	7.0

Each readout layer is composed of alternating layers of grounded absorbing plates (uranium in most of the calorimeter) and signal boards at a potential of 2 kV. In the space between the absorbing plates and signal boards is liquid argon, as illustrated in figure 2.12. Charged particles from an EM or hadronic shower will ionize the liquid argon resulting in a
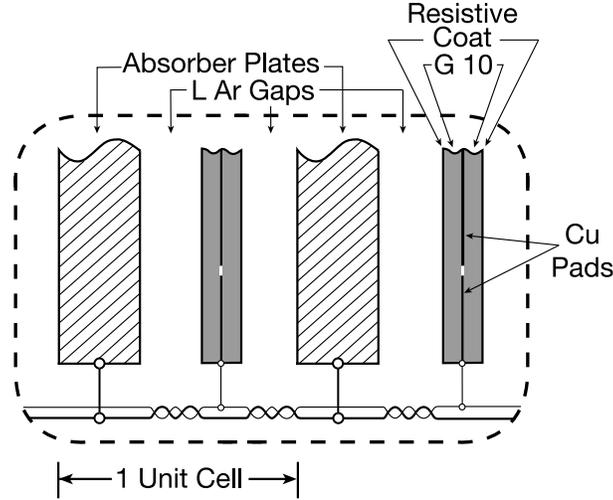
Figure 2.12: Schematic of the internal structure of the calorimeter.

current between the absorber plates and signal boards. The number of signal boards in each readout layer is given in tables 2.1 and 2.2.

In total, the calorimeter has around 50,000 readout cells. With so many cells there is a significant chance for a false signal from electronics noise and/or uranium decay in the calorimeter. Therefore, before object reconstruction (chapter 3) the so called "T42" zero-suppression algorithm [23] is used to reduce the effects of noise by removing cells that do not measure an energy significantly higher than the noise level. Specifically, cells are removed unless they have an energy at least four standard deviations above the noise ($E^{\text{cell}} \geq 4\sigma^{\text{noise}}$) or have $E^{\text{cell}} \geq 2.5\sigma^{\text{noise}}$ and are adjacent to a cell with $E^{\text{cell}} \geq 4\sigma^{\text{noise}}$.

2.2.5 Muon System

Electrons are stopped by the calorimeter, but the 200 times more massive muons pass through leaving only a fraction of their energy behind. Muons with energies between a few hundred MeV and a few hundred GeV are minimum ionizing particles, losing energy at a rate of around 0.25 GeV per nuclear interaction length traversed. Other than neutrinos, which are not detected at all, muons are the only SM particles that live long enough to travel through the detector, but are not stopped by the calorimeter. Therefore, the muon system

is the outermost section of the D0 detector designed to detect the muons as they exit the calorimeter.

The muon system has a rectangular geometry, like a cube, with three layers of detectors as shown in figure 2.4. The "A" layer is closest to the calorimeter, followed by the iron toroid magnet, then the "B" and "C" layers, respectively. The sides of the muon system at each end (perpendicular to the beam line) form the forward angle muon system (FAMUS) and the other four sides form the wide angle muon system (WAMUS). The WAMUS covers the region $|\eta| \lesssim 1$ and the FAMUS covers the region $1 \lesssim |\eta| \lesssim 2$ (the η boundaries are not exact due to the rectangular geometry).

As shown in figure 2.13a, all three layers of the WAMUS contain proportional drift tubes (PDTs). The A layer has four decks of PDTs and the B and C layers each have three decks (figure 2.14). The tubes are 10 cm wide and 5.5 cm thick with copper cathodes on the upper and lower walls and a gold plated tungsten anode wire running down the middle. The tubes are filled with a gas mixture of 84% Argon, 8% CF_4, and 8% methane. The cathode walls are held at 2.3 kV and the anode wires at 4.7 kV.

A muon passing through the PDTs leaves a trail of ions in the gas. Freed electrons are accelerated in the electric field, ionizing more of the gas as they drift toward the anode wire. This results in an avalanche of charge that reaches the anode wire after a time proportional to the distance between the muon's track and the anode. The PDTs have a maximum response time of 500 ns.

The WAMUS PDTs are oriented perpendicular to the beam line providing a measurement of the z position of tracks with a precision of 1 mm. The muon toroid produces a 1.9 T magnetic field between the A and B layers that runs parallel to the PDTs, thus deflecting muons either forward or backward (figure 2.15). As with the central tracking system, the momentum of the muon can be deduced from the angle it is deflected by the magnetic field. The momentum resolution for the central tracking system is much better than for the muon system; therefore, the momentum measured by the muon system is usually only used to help
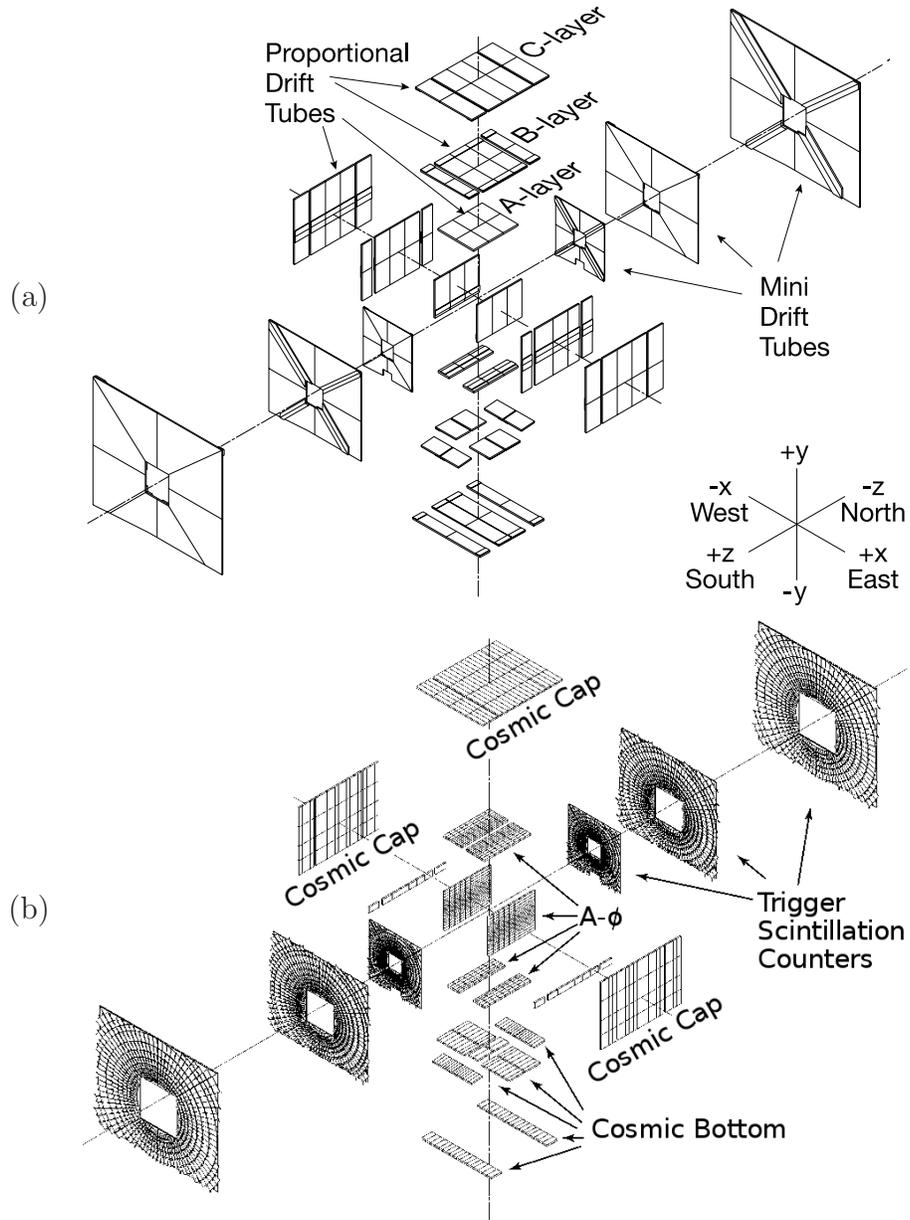
Figure 2.13: Diagram showing (a) the drift tubes and (b) the scintillators for the entire muon system.
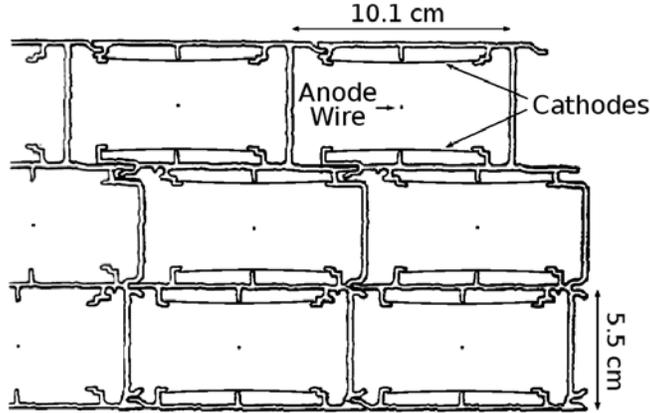
Figure 2.14: Cross section of PDTs used in WAMUS.

find a matching track in the central tracking system then the momentum from the central track is used.
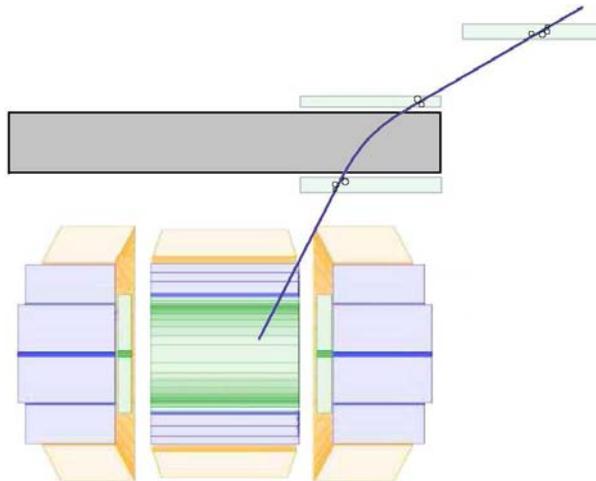
Figure 2.15: Simulation showing a muon deflected by toroid magnet.

As shown in figure 2.13b, the A and C layers and part of the B layer also contain scintillators. The "cosmic cap" consists of scintillators in the C layer on the top and sides of the WAMUS. The scintillators in the bottom B and C layers from the "cosmic bottom." The "A-ϕ" scintillators are in the A layer of the WAMUS. The A-ϕ scintillators are segmented to provide a measurement of ϕ position; however, with a response time of 1.6 ns the primary role of the scintillators is to provide precise timing information required for triggering, matching drift tube hits with the correct event, and rejecting background from cosmic ray muons.
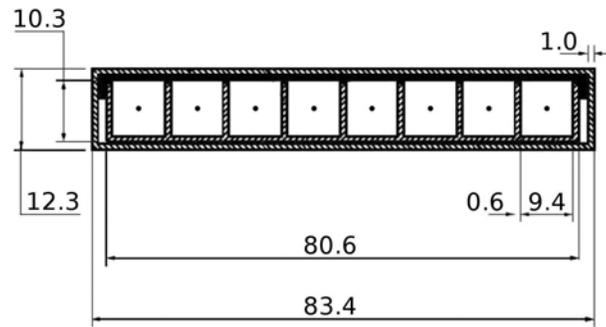
Figure 2.16: Cross section of a deck of MDTs used in FAMUS (dimensions in millimeters).
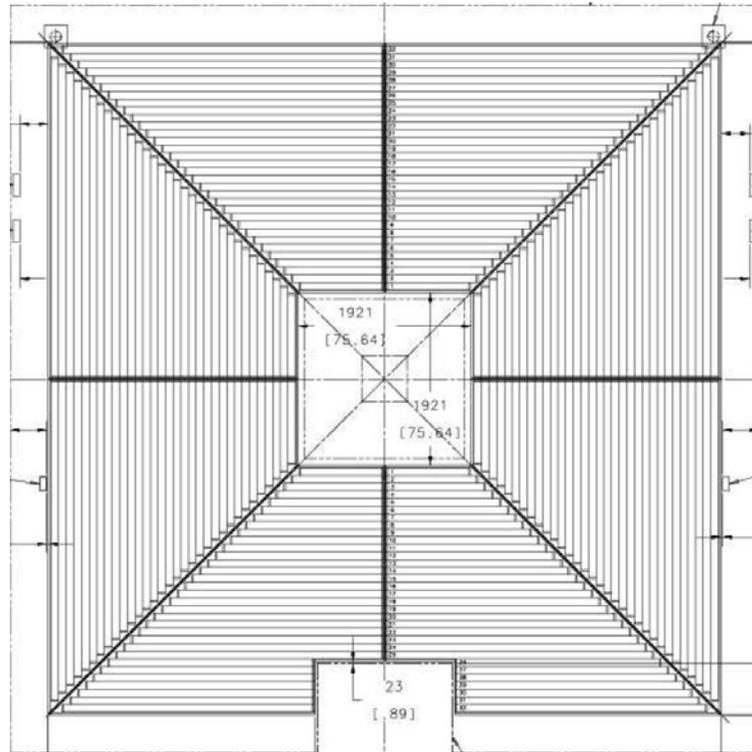
Figure 2.17: Illustration showing the orientation of MDTs in each layer of the FAMUS.
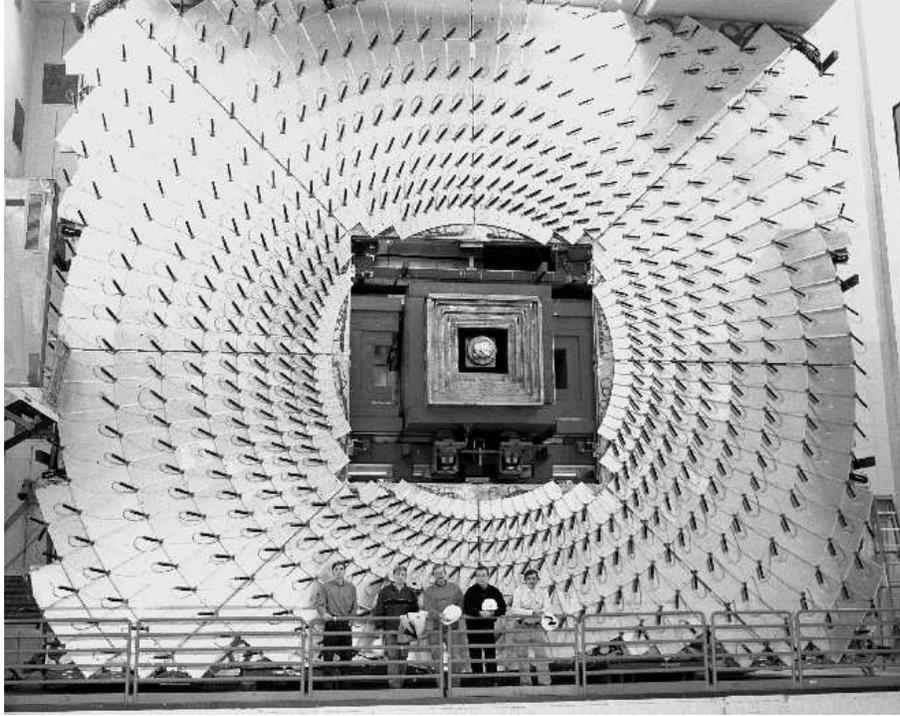
Figure 2.18: Photograph of the scintillating tiles in the C layer of the FAMUS.

The FAMUS uses mini drift tubes (MDTs) with a cross section of 1 cm by 1 cm (figure 2.16). The MDTs are oriented perpendicular to the beam line as shown in figure 2.17 with four decks in the A layer and three in each of the B and C layers. The MDTs are filled with a mixture of 90% CF_4 and 10% methane and operate analogous to the PDTs in the WAMUS, though with a much shorter response time of approximately 100 ns. Each layer of the FAMUS also has a plane of scintillators that are used for triggering and vetoing cosmic ray muons. The scintillating tiles are in concentric circular arrays forming a matrix of large pixels as seen in figure 2.18.

2.2.6 Luminosity Monitor

The luminosity monitor is used to determine the instantaneous luminosity being delivered to the D0 detector by measuring the rate of proton-antiproton collisions. After a proton-antiproton collision, the partons (quarks and gluons) in the colliding proton and antiproton that were not involved in the hard interaction (the remnants) will create hadrons that con-

tinue in the $\pm z$ directions with only a small transverse momentum. To detect the remnants of a collision, the luminosity monitor uses two circular hodoscopes at each end of the tracking system ($z = \pm 144$ cm) and spanning the range $2.7 < |\eta| < 4.4$. Each hodoscope is composed of 24 wedges of scintillating tiles. The scintillators have a timing resolution of around 0.2 ns, which allows the z position of the collision to be determined to within 6 cm. Collisions with $|z| < 100$ cm are used for determining the luminosity.

2.3 Trigger

With 36 bunches of protons and antiprotons traveling around the Tevatron every 21 μs, the average rate of bunch crossings at the center of the D0 detector is 1.7MHz. A three-stage trigger system [24, 25] is used by D0 to reduce this rate to less than 100 events per second that are stored for later analysis. The cross section for inelastic proton-antiproton scattering at the Tevatron is about 50 mb, resulting in an average of a few inelastic collisions per bunch crossing. Of the millions of inelastic collisions happening in the D0 detector only a handful will contain physics that people are interested in studying. For example, only a few W bosons are produced per second and only a few top-quark pairs per hour. If every event was stored for later analysis the experiment would have to store around two TeraBytes of data each day, which is unnecessary. Therefore, sophisticated triggers capable of rejecting nearly 99.999% of the interactions in only a few milliseconds are used to pick out the rare events of interest.

The trigger system works by exploiting the fact that the events of interest have distinctive signatures. For example, a high p_T lepton may indicate the presence of a W or Z boson, a high p_T lepton plus two b quark jets would signal a top quark pair or Higgs boson, and leptons plus large missing energy is a signature for supersymmetry. The job of the trigger is to reject all events that do not display characteristics consistent with a signature of any interesting physics. Therefore, it is the trigger that ultimately determines the physics processes that

can be studied and the ones that are left unexplored.

The trigger system used by D0 is divided into three levels of increasing complexity. An event must pass each of the three levels before it can be stored for later analysis. The Level 1 trigger is designed to find patterns of energy deposition indicating the passage of high energy particles and must be very fast (4.2 μs) to keep up with the rate of interactions. It therefore uses a condensed subset of the full detector readout and is implemented entirely in hardware and firmware. Each detector subsystem (calorimeter, muon, etc.) checks if the event passes preprogrammed trigger conditions (*e.g.*, the calorimeter trigger tests for energy deposits above pre-programmed thresholds), then the results from each detector subsystem are combined to make the final Level 1 decision. After a Level 1 trigger is passed there is 10 μs of dead time required for readout resulting in the maximum accept rate for Level 1 to be set at 5 kHz. After an event passes Level 1 it is sent to Level 2 where it is subjected to more refined tests that may take up to 100 μs. As well as firmware, Level 2 uses microprocessors to take advantage of more precise detector information and spatial correlations to form basic "objects" such as tracks, EM clusters, and jet clusters. Each detector subsystem has a dedicated microprocessor that reduces the data for that subsystem then sends it on to a global processor to make the final Level 2 decision. The rate at which Level 2 can pass events to Level 3 is limited to 1 kHz by the Level 2 processing time. When an event passes the Level 2 trigger the entire detector is read out and sent to the Level 3 CPU farm made up of over 100 Linux computers. The Level 3 CPUs work together to process each event in under 25 ms. During that time the entire event is reconstructed and sophisticated algorithms, close to those used for offline analyses, are applied to make the final decision. The output rate for Level 3 is maintained around 50 Hz, which was chosen to keep the offline event reconstruction from piling up.

Chapter 3

Event Reconstruction

Before analyzing the data, the physical objects (photons, muons, hadronic jets, etc.) in each event must be reconstructed from the raw data (digital and analog signals from the tracking systems and calorimeter). The following sections describe the algorithms and criteria for reconstructing and identifying the physical objects required for an analysis.

3.1   Tracks

Tracks, though not physical objects themselves, are important for measuring momentum and in object identification. For example, electrons and photons have the same experimental signature in the calorimeter, but are differentiated by the presence or lack of a track, respectively. As charged particles traverse the layers of the central tracking system they leave hits that are clustered then analyzed with pattern recognition software to find tracks.

3.1.1   Track Reconstruction

A charged particle traveling through the SMT and CFT will likely interact with multiple channels (strips in the SMT, fibers in the CFT) in each layer, resulting in a clusters of hits (figure 3.1). For each channel, the analog signal is calibrated and converted to digital
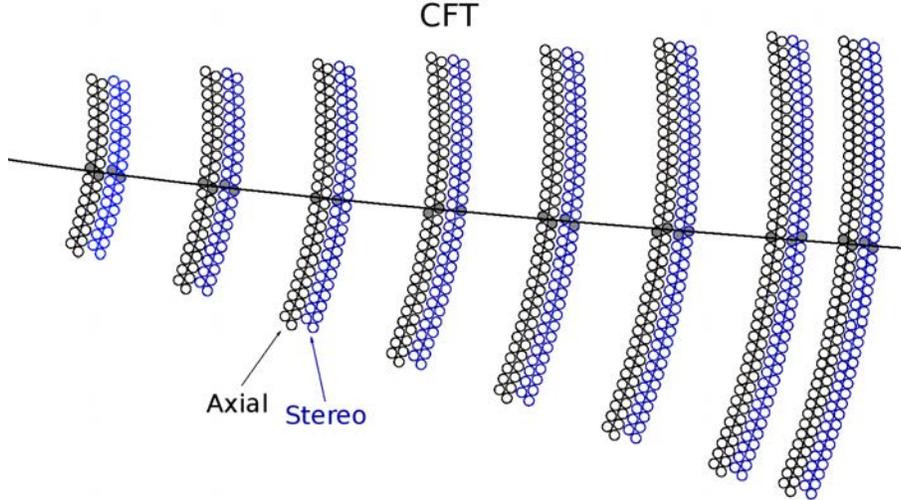
Figure 3.1: Illustration showing the clusters of hits left in each layer of the CFT from a charged particle.

counts. Adjacent channels above a minimum number of digital counts are then clustered and the centers of the clusters are used as the hit locations for the track finding software. The resolution for hits in the SMT (CFT) is approximately 10 μm (100 μm) in the azimuthal direction and 35 μm (2 cm) in the z direction.

It is important to find all of the tracks in each event, which is why D0 uses more than one track finding method. The first technique begins at the innermost layers of the tracking system, then moves incrementally outward. The algorithm starts with a hit in an SMT barrel or F-disk, plus a second hit further out that has a difference in azimuthal angle from the first hit of $\Delta\phi < 0.08$. The algorithms then searches for a third hit consistent with a track that has a radius of curvature $\rho > 30$ cm (*i.e.*, $p_T > 180$ MeV), a distance of closest approach to the beam line $d_0 < 2.5$ cm, and a χ^2 fit value $\chi^2 < 16$. All such candidates are used to construct tracks through the rest of the tracking system using a Kalman filter [26]. Using a detailed mapping of detector material and the magnetic field, the Kalman filter extrapolates each track to the next layer of the tracking system. If a new hit is found within a small window around the extrapolation, then that hit is included in the track. The algorithm allows for missed layers, but stops after three consecutive missed layers. Additionally, tracks cannot share more than 2/3 of their hits with other tracks and the hits must fit the tracks

with a $\chi^2 < 16$.

Tracks are also found by looking at the transverse projection of all the hits in the SMT and applying a Hough transformation [27] to convert the hit positions in $x \times y$ space to lines in $\rho \times \phi_0$ space; where ρ is the curvature of a circle intersecting the origin and the hit, ϕ_0 is the angle of the tangent to that circle at the origin. Hits from the same track intersect at a single point in the $\rho \times \phi_0$ plane, namely the ρ and ϕ_0 of the track. In practice, the procedure is performed at discrete values of ρ and ϕ_0, thus populating a two-dimensional histogram. In this histogramming method, the track candidates show up as peaks in the ρ versus ϕ_0 histogram. A two-dimensional Kalman filter is used to more accurately calculate the track parameters and remove tracks with large χ^2. The z component of each track candidate is determined using another histogramming algorithm (similar to above), then a three-dimensional Kalman filter extrapolates the tracks through the CFT.

Lastly, the previous method is performed starting with hits in the CFT then extrapolating the tracks backward into the SMT.

3.2 Primary Vertex

A collision vertex is the location of a proton-antiproton collision. For a given instantaneous luminosity, the number of collision vertices in each bunch crossing is a Poisson distributed random variable. The average number is typically around two, but due to the long Poisson tail, some bunch crossings will have many. The locations of collision vertices are distributed with a width of about $\sigma_r = 40$ μm in the radial direction and $\sigma_z = 28$ cm along the z-axis.

In bunch crossings that contain an event of interest (such as the production of a W boson), it is important to determine the location of the vertex belonging to that specific event so that the momentum vectors of the resulting particles are accurately calculated. This is the primary vertex. Other collision vertices in the bunch crossing are referred to as minimum-bias interactions because those interactions were recorded only by virtue of occurring in the

same bunch crossing as the hard interaction of interest. Knowing the location of the primary vertex also helps discriminate against particles that are not part of the event of interest, *e.g.*, particles from cosmic rays or minimum-bias interactions.

3.2.1 Primary Vertex Reconstruction

Reconstructing collision vertices is done is two passes. The first pass uses tracks that have a distance of closest approach within 100 standard deviations of the z-axis. These tracks are fit to find the best location for a common vertex. If the fit yields a total $\chi^2/\text{ndf} > 10$, then the track with the highest contribution to the χ^2 (*i.e.*, with the worst fit) is removed and the remaining tracks are refit. This repeats until the fit yields a $\chi^2/\text{ndf} \leq 10$, at which point the location given by the fit is used as the approximate position of a collision vertex. The whole process is then repeated using the removed tracks to find other collision vertex locations.

The second pass is run separately for each vertex found in the first pass. Only tracks that have a distance of closest approach within 5 standard deviation of the given vertex location are used. In addition, the tracks must have $p_T > 0.5$ GeV and at least two hits in the SMT. These tracks undergo a similar procedure of fitting to the primary vertex and removing tracks until the fit yields a $\chi^2/\text{ndf} \leq 10$.

After finding all collision vertices for a bunch crossing, the task is to determine which are minimum-bias interactions and which is the primary vertex. Tracks from a minimum-bias interaction tend to have much lower p_T than tracks from a hard interaction. Using the track p_T, the probability for each track to have originated from a minimum-bias interaction is calculated. The minimum-bias probabilities for all tracks coming from a given vertex are multiplied to determine the minimum-bias probability for that collision vertex. The collision vertex with the lowest probability of being a minimum-bias interactions is taken as the primary vertex.

3.3 Electrons
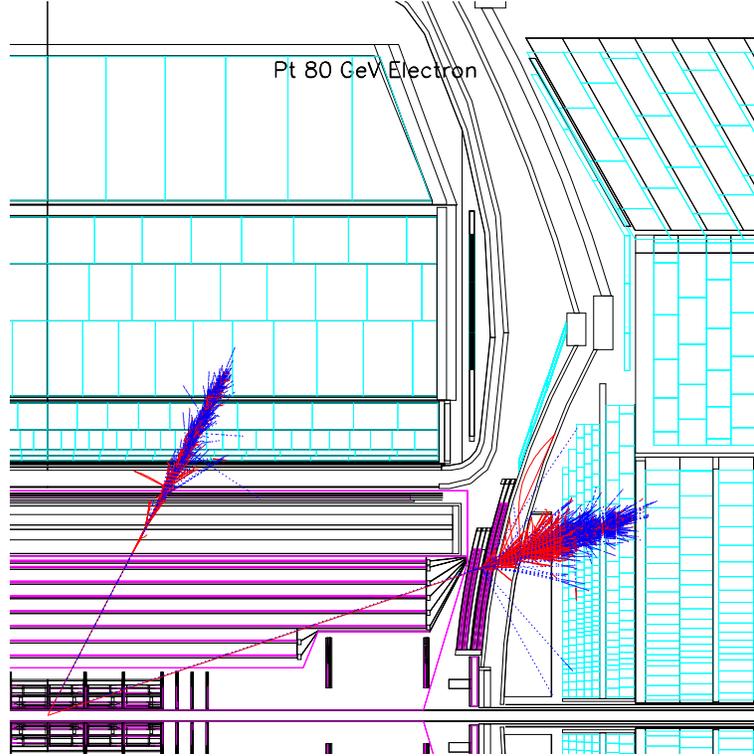
Figure 3.2: Simulation showing the showers from 80 GeV electrons in the CC and EC.

Electrons are characterized by an isolated track in the central tracking system leading to a narrow energy deposit in the calorimeter that starts in the preshower detectors and ends in the first hadronic layer (figure 3.2). To reconstruct electrons[1], a cone algorithm is used to find clusters of energy in the EM layers of the calorimeter then the clusters are matched to tracks and tested for the characteristics of an electron shower. Following a description of the reconstruction are definitions of the electron quality variables, though the actual selection cuts are given in chapter 5.

---

[1]Photons, though not used in this analysis, are basically reconstructed as electrons without a matching track.

3.3.1 Electron Reconstruction

An EM tower is defined as the first five layers of a 0.1×0.1 ($\eta \times \phi$) tower of calorimeter cells (the four EM layers plus the first hadronic layer). All EM towers with a "transverse" energy ($E_T \equiv E/\cosh(\eta) \equiv E\sin(\theta)$) greater than 1.5 GeV go into a list of seed towers. The seed tower with the highest E_T is selected and a cone of radius $R = \sqrt{\Delta\eta^2 + \Delta\phi^2} = 0.4$ is centered on the tower. The energy-weighted center for the cluster of EM towers inside the cone is calculated and the cone is moved to the new center. This repeats until a stable cluster is found. Any seed towers inside the cluster are removed from the list of seeds and the procedure is repeated with the updated list of seeds to find the next cluster.

The clusters have a radius of $R = 0.4$, but the electron shower is actually much narrower. The energy is measured using the EM towers inside a cone of radius $R = 0.2$ (at the center of the cluster). To be considered an electron candidate, there must be a track that extrapolates to within $\Delta\eta = 0.05$ and $\Delta\phi = 0.05$ of the cluster's center. The angle of the track at the primary vertex is used to calculate the four-momentum for the electron.

3.3.2 Electron Energy Scale

The measured electron energy is corrected by selecting $Z \rightarrow ee$ events in data and comparing the dielectron mass resonance with expectation from previous precision measurements [28]. The dielectron mass resonance obtained using uncorrected electron energies is found to be lower than expected, thus the uncorrected electron energies are scaled up to obtain corrected energies that reproduce the expected $Z \rightarrow ee$ resonance. The true energy, E^{true}, is assumed to be related to the measured energy, E^{meas}, by

$$E^{\text{meas}} = \alpha E^{\text{true}} + \beta; \tag{3.1}$$

where α is the energy scale and β is an energy offset. The scale and offset corrections are determined separately for CC and EC electrons using a binned maximum likelihood method

and then the corrected electron energy is calculated by inverting equation 3.1.

3.3.3 Electron Quality Variables

The basic quality variables are EM-fraction and isolation defined as

$$f_{\text{EM}} = \frac{E_{\text{EM}}(0.2)}{E_{\text{total}}(0.2)} \tag{3.2}$$

and

$$f_{\text{iso}} = \frac{E_{\text{total}}(0.4) - E_{\text{EM}}(0.2)}{E_{\text{EM}}(0.2)}, \tag{3.3}$$

respectively. Here $E_{\text{total}}(R)$ is the total energy (all EM and hadronic layers) inside a cone of radius R and $E_{\text{EM}}(R)$ is the energy in only the EM layers inside a cone of radius R. EM-fraction and isolation are remarkably good at rejecting clusters that are not electron showers, however, the purity can be improved even further using more sophisticated quantities, namely those described in the following sections.

Track-Match χ^2 Probability

To determine how well the cluster is matched to a track, a track-match χ^2 is calculated in two ways. The spacial track-match uses the distance between the track and center of the cluster in ϕ and z directions:

$$\chi^2_{\text{spacial}} = \left(\frac{\Delta\phi}{\sigma_{\Delta\phi}}\right)^2 + \left(\frac{\Delta z}{\sigma_{\Delta z}}\right)^2, \tag{3.4}$$

while the E/p track-match also compares the energy, E, measured in the calorimeter to the momentum, p, of the track:

$$\chi^2_{\text{E/p}} = \left(\frac{\Delta\phi}{\sigma_{\Delta\phi}}\right)^2 + \left(\frac{\Delta z}{\sigma_{\Delta z}}\right)^2 + \left(\frac{E/p - 1}{\sigma_{E/p}}\right)^2. \tag{3.5}$$

As usual, σ_X corresponds to one standard deviation for the distribution of variable X. The corresponding χ^2 probability distribution, $P(\chi^2)$, is then used to determine the probability that the track matches the cluster. If there is more than one possible track, the track with the lowest χ^2 is used.

H-matrix χ^2

The H-matrix χ^2 [29] provides a comparison between the shower shape of the cluster and the expected shower shape for an electron. Seven variables are used to parameterize the shower shape:

1. Fraction of energy in the first EM layer,

2. Fraction of energy in the second EM layer,

3. Fraction of energy in the third EM layer,

4. Fraction of energy in the fourth EM layer,

5. Width of shower in the third EM layer,

6. Logarithm of the total energy,

7. z coordinate of the primary vertex.

The expected electron shower shape is determined using a sample of N MC electrons to construct the covariance matrix

$$M_{ij} = \frac{1}{N} \sum_{n=1}^{N} (x_i^n - \bar{x}_i)(x_j^n - \bar{x}_j); \tag{3.6}$$

where x_i^n is the value of variable i in for the nth MC electron. The H-matrix is simply defined as the inverse of the covariance matrix, $\mathrm{H} \equiv \mathrm{M}^{-1}$. The H-matrix can then be used

to construct a χ^2 in the usual way

$$\chi^2_{\text{H-matrix}} = \sum_{i,j=1}^{7} (x_i - \bar{x}_i) H_{ij} (x_j - \bar{x}_j); \tag{3.7}$$

where x_i is the value of variable i for the electron candidate being evaluated. The expected shower shape changes with $|\eta|$, therefore, a different H-matrix is constructed for different $|\eta|$ ranges.

Electron Likelihood

Any electron candidate that is not the result of a real electron is referred to as a "fake" electron. The electron likelihood [30] variable is designed to discriminate between clusters from real electrons (signal) and clusters from fake electrons (background). The likelihood makes use of the following seven variables:

1. EM-fraction (f_{EM}),

2. Spacial track-match χ^2 probability ($P(\chi^2_{\text{spacial}})$),

3. Ratio of calorimeter energy and track momentum (E/p),

4. H-matrix χ^2 ($\chi^2_{\text{H-matrix}}$),

5. Track distance of closest approach to the primary vertex,

6. Total p_T of other tracks inside the cone of radius $R = 0.4$,

7. Number of tracks inside a cone of radius $R = 0.05$.

For each variable (x_i), the probability distributions for signal ($P_{\text{S}}(x_i)$) and background ($P_{\text{B}}(x_i)$) are constructed. The variables are assumed to be uncorrelated, thus the total signal and background probabilities for a set of inputs $\vec{x} = (x_1, ..., x_7)$ are simply

$$P_{\text{S}}(\vec{x}) = \prod_{i=1}^{7} P_{i,\text{S}}(x_i) \quad \text{and} \quad P_{\text{B}}(\vec{x}) = \prod_{i=1}^{7} P_{i,\text{B}}(x_i), \tag{3.8}$$

respectively. The prior probabilities for signal and background are assumed to be unity resulting in the electron likelihood discriminant

$$\mathcal{L}_e(\vec{x}) = \frac{P_{\mathrm{S}}(\vec{x})}{P_{\mathrm{S}}(\vec{x}) + P_{\mathrm{B}}(\vec{x})}. \tag{3.9}$$

3.4 Muons

A muon is identified by a track in the muon system, which can then extrapolated inward to an isolated track in the central tracking system. Below is the procedure for reconstructing a muon followed by descriptions of the variables used for selecting high quality muons for analysis, though the actual selection requirements are given in chapter 5.

3.4.1 Muon Reconstruction

A muon track segment is formed in the A layer by fitting a straight line to hits in two or more decks of the drift tubes. The B and C layers are considered together (called the BC layers) and muon track segments in the BC layers are constructed in the same way. When there is more than one possible track segment for the A or BC layers in a given octant, only the segment with the best fit (lowest χ^2) is used.

After constructing the track segments inside (A layer) and outside (BC layers) the toroid magnet, a fit is performed through the toroid accounting for bending from the magnetic field as well as effects from energy loss and multiple scattering in the material. In principle, this track is a reconstructed muon with a momentum determined by the amount of deflection in the toroid. However, if the muon track can be matched to a track in the central tracking system, then the momentum of the muon can be measured much more accurately using the curvature of the central track. Matching a muon track to a central track is done by extrapolating the muon track inward, modeling the calorimeter as a thin scatterer with the appropriate number of radiation lengths, and fitting to a central track. The track with the

best fit is used and, as stated below, the χ^2 of the fit can be used in muon selection.

3.4.2 Muon Quality Variables

The muon quality variables [31] fall into three categories: muon system quality, central track quality, and isolation variables.

Muon System Quality

Muons can also be reconstructed from a central track extrapolated outward to a hit or track segment in only one layer of the muon system. These were not discussed because this analysis only considers muons with hits in both the A and BC layers. Muons with track segments in both A and BC layers have $|N_{\text{seg}}| = 3$, otherwise $|N_{\text{seg}}| < 3$.

Requirements may also be placed on the number of drift tube hits in the A and BC layers and number of scintillator hits in the A and BC layers. In general, more hits result in a more accurately measured track. In addition, precise timing requirements can be placed on the scintillator hits to help reject muons from cosmic rays.

Central Track Quality

The quantities used to verify the quality of the central track are:

- Distance of closest approach to the beam line for tracks with SMT hits,

- Distance of closest approach to the beam line for tracks without SMT hits,

- χ^2/ndf for the match to the central track.

The distance of closest approach to the beam line is primarily a tool for rejecting muons from cosmic rays, which are unlikely to intersect the beam line. Muon background originating from pions and kaons that decay inside the tracking volume can be significant in the single muon analysis as these muons have high p_T; however, the central tracks will likely be poorly reconstructed resulting in a high value for χ^2/ndf.

Isolation

A muon from a W or Z boson produced in the hard interaction will result in an isolated track originating from the primary vertex. This is in contrast to the non-isolated muon that can result from the decay of a heavy quark inside a jet (*e.g.*, a b quark jet where the b decays muonically ($b \rightarrow u\ W \rightarrow u\ \mu\nu$)). This analysis is concerned with the former, thus muons are required to be isolated from other activity in the event. In addition to requiring reconstructed muons to not be inside a reconstructed jet, the following two quantities are used for ensuring isolation:

- CalorimeterHalo $= E_T^{\text{cal}}(0.4) - E_T^{\text{cal}}(0.1)$,

- TrackHalo $= p_T^{\text{tracks}}(0.5) - p_T(\mu)$;

where $E_T^{\text{cal}}(R)$ is the E_T in all layers of the calorimeter inside a cone (around the muon track) of radius R, $p_T^{\text{tracks}}(R)$ is the total p_T of all tracks inside a cone of radius R, and $p_T(\mu)$ is the p_T of the muon.

3.5   Jets

A jet is a collimated shower of energetic particles coming from the primary vertex. Most jets are the result of the hadronization of a high energy quark or gluon produced in the hard interaction. Jets result in showers of energy in the EM and hadronic layers of calorimeter that are usually accompanied by many tracks in the central tracking system (in principle, one for each charged particle in the jet). This section explains the Run II cone algorithm [32, 33] used to reconstruct jets, followed by the basic jet quality requirements, and finally a description of the jet energy correction [34].

3.5.1 Jet Reconstruction

Jet reconstruction proceeds in three stages. The first stage is constructing preclusters of towers. The "E-scheme" is used to calculate a four-momentum for each 0.1×0.1 ($\eta \times \phi$) tower of calorimeter cells

$$p^{\text{tower}} = \sum_i p_i^{\text{cell}} \; ; \tag{3.10}$$

where p_i^{cell} is a massless four-momentum with an energy of the ith cell in the tower and pointing from the primary vertex to the center of the cell. Towers with $p_T^{\text{tower}} > 0.5$ GeV are then put into a list of seeds for making preclusters. The tower with the highest p_T^{tower} in the seed list is selected and a cone of radius $R = 0.3$ is placed around the tower. Towers within the cone are summed to form a precluster ($p^{\text{precluster}} = \sum p^{\text{tower}}$, towers $\in R < 0.3$). Any seed towers inside the precluster are removed from the list of seeds and the procedure is repeated with the updated list of seed towers to find the next precluster.

The second stage is creating protojets using the preclusters as seeds. Preclusters with $p_T^{\text{precluster}} > 1$ GeV and containing more than one energetic tower are put into a list of seeds for making protojets. The seed precluster with highest $p_T^{\text{precluster}}$ is selected and a cone of radius $\tilde{R} = \sqrt{\Delta y^2 + \Delta \phi^2} = 0.5$ is placed around the precluster; where y is rapidity (equation 2.2). Towers within the cone are summed to form a trial protojet ($p^{\text{protojet}} = \sum p^{\text{tower}}$, towers $\in \tilde{R} < 0.5$) and then the cone is re-centered around the trial protojet. A new trial protojet is made from the updated cone and the process repeats until a stable protojet is found (or if $p_T^{\text{protojet}} < 4$ GeV, then it is discarded). Any seed preclusters within $\tilde{R} = 0.25$ of the stable protojet are removed from the list of seeds and the procedure is repeated with the updated list of seed preclusters to find the next protojet.

The last stage of jet reconstruction is splitting or merging overlapping protojets. A protojet that does not overlap any other protojet becomes a jet. Two overlapping protojets are merged if they share at least than half of the p_T of one of the protojets; otherwise they are split. Merging is done by summing the two protojets to make a new trial protojet that

is iterated as described above until a new stable protojet is found. Splitting is done by assigning each of the shared towers to the closer of the two protojets. Splitting and merging continues until there are no more overlapping protojets and, thus, all the jets are found.

3.5.2 Basic Jet Requirements

The following basic quality requirements are applied to the reconstructed jets to remove "fake" jets that are primarily the result of noisy cells or regions in the calorimeter.

- The fraction of energy coming from cells in the EM layers of the calorimeter must be greater than 5% and less than 95%.

- The fraction of energy coming from cells in the course hadronic layer must by less than 44% for jets in the CC and less then 46% for jets in the EC.

- The jet is rejected if the energy of the most energetic cell in the jet is more than ten times the energy of the second most energetic cell.

- The jet cannot have more then 90% of its energy in a single calorimeter tower.

In addition to the above requirements, each jet is required to be confirmed by the Level 1 trigger (*i.e.*, verify that the jet was "seen" by the Level 1 trigger). The Level 1 trigger uses only a condensed subset of the full calorimeter data consisting of the 100 highest E_T trigger towers; where trigger towers are 0.2×0.2 ($\eta \times \phi$) towers (*i.e.*, groups of four calorimeter towers) excluding the course hadronic layer. The Level 1 confirmation requirement is

$$\frac{E_T^{\text{trigger towers}}}{E_T^{\text{reco}} \cdot (1 - f_{\text{CH}})} > 0.5; \tag{3.11}$$

where $E_T^{\text{trigger towers}}$ is the total E_T of trigger towers inside a cone of radius $R = 0.5$ around the jet, E_T^{reco} is the E_T of the reconstructed jets, and f_{CH} is the fraction of E_T^{reco} in the course hadronic cells.

3.5.3 Jet Energy Scale

The jet energy scale (JES) correction is used to correct a jet's measured energy ($E_{\text{jet}}^{\text{measured}}$) to, on average, more accurately reflect the true energy ($E_{\text{jet}}^{\text{true}}$) of the all the particles in the jet before interacting with the detector. The measured energy for a jet is related to the true energy by

$$E_{\text{jet}}^{\text{true}} = \frac{E_{\text{jet}}^{\text{measured}} - E_O}{\mathcal{R}_{\text{jet}} \cdot \mathcal{S}_{\text{jet}}};$$ (3.12)

where:

- E_O is the offset to the energy resulting from uranium decay, minimum-bias interactions, previous bunch crossings (pile-up), and electronics noise.

- $\mathcal{R}_{\text{jet}}$ is the energy response (the fraction of a particle's energy that is actually measured), which is affected by energy lost before reaching the calorimeter, uninstrumented regions in the calorimeter, differences in EM and hadronic response, and non-linearities in energy response.

- $\mathcal{S}_{\text{jet}}$ is the showering correction due to energy that is lost (gained) from particles inside (outside) the jet cone that have showers that extend outside (inside) the cone boundary.

By measuring the average values for each of the above quantities (as functions of energy and location in the detector), corrected jet energies can be calculated as[2]

$$E_{\text{jet}}^{\text{corr}} = \frac{E_{\text{jet}}^{\text{measured}} - \hat{E}_O}{\hat{\mathcal{R}}_{\text{jet}} \cdot \hat{\mathcal{S}}_{\text{jet}}};$$ (3.13)

where $\hat{X}$ is the measured average value for X.

The average energy offset, $\hat{E}_O$, is measured using zero bias data and minimum bias data. Zero bias data are collected by recording every single bunch crossing (*i.e.*, no trigger

---

[2]Biases in the estimation of $\hat{E}_O$ and $\hat{\mathcal{R}}_{\text{jet}}$ actually result in a slight modification of equation 3.13, but all of the physics motivation is contained in equation 3.13.

requirements). Minimum bias data require only that the event contained an inelastic proton-antiproton collision. The zero bias sample is used to measure the energy offset due to uranium decay, electronics noise, and pile-up, while the energy offset from minimum bias interactions is measured in the minimum bias sample.

The jet energy response, $\hat{\mathcal{R}}_{\mathrm{jet}}$, is measured using the Missing E_T Projection Fraction (MPF) method [35] in data containing a photon and jet (γ+jet) that are back-to-back in the transverse plane. The total transverse momentum in a collision is (very nearly) zero, thus the $\vec{p}_T$ of the photon ($\vec{p}_{T,\gamma}$) will be equal and opposite to the p_T of the of the jet before interacting with the detector ($\vec{p}_{T,\mathrm{hadrons}}$). The photon is required to be in the CC ($|\eta| < 1$) so that its energy is determined very accurately[3]. As illustrated in figure 3.3, the jet energy response is determined by measuring the imbalance in transverse momentum ($\vec{\not{E}}_T$, "missing E_T").

$$\vec{p}_{T,\gamma}^{\;\mathrm{true}} + \vec{p}_{T,\mathrm{jet}}^{\;\mathrm{true}} = \vec{0} \qquad \vec{p}_{T,\gamma}^{\;\mathrm{true}} + \mathcal{R}_{\mathrm{jet}}\vec{p}_{T,\mathrm{jet}}^{\;\mathrm{true}} \approx \vec{p}_{T,\gamma}^{\;\mathrm{meas}} + \vec{p}_{T,\mathrm{jet}}^{\;\mathrm{meas}} = -\vec{\not{E}}_T$$

$$\boxed{\mathcal{R}_{\mathrm{jet}} \approx 1 + \frac{\vec{\not{E}}_T \cdot \vec{p}_{T,\gamma}^{\;\mathrm{meas}}}{(\vec{p}_{T,\gamma}^{\;\mathrm{meas}})^2}}$$
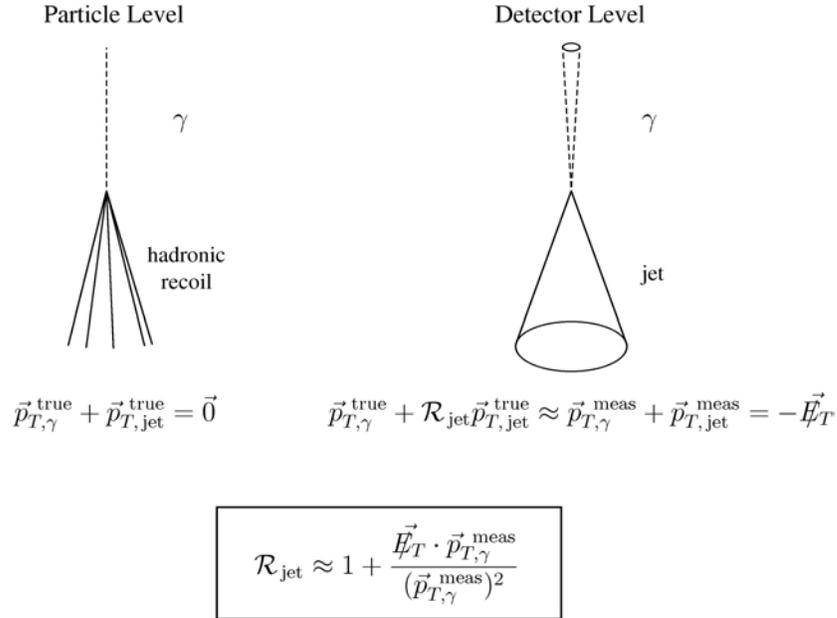
Figure 3.3: Illustration of the MPF method.

The showering correction, $\hat{\mathcal{S}}_{\mathrm{jet}}$, is determined by measuring the shower profile (the energy contained in cones of varying radii, $0.1 \leq \tilde{R} \leq 2$, centered on the jet). MC simulation is used

---

[3]Photons have the same energy response as electrons, which are calibrated using $Z \to ee$ data.

to create templates of the shower profiles for particles inside the jet and particles outside the jet. The templates are fit to the shower profile measured in γ+jet data to determine the ratio of measured energy inside the cone to the true jet energy inside the cone.

3.6 Missing Energy

Momentum conservation requires the total momentum of all particles coming out of a proton-antiproton collision be equal to the total momentum of the colliding proton and antiproton, namely zero. However, usually only a single parton (quark or gluon) from each of the colliding hadrons partake in a hard interaction and the remnants of the hadrons continue down the beam pipe undetected. The fraction of the original hadron's momentum carried away by the remnants is unknown, but the transverse component must be small for the particles to have escaped down the beam pipe. So while the total momentum of the partons in the hard interaction may be unknown, the transverse momentum is approximately zero. Thus, measuring an imbalance in the total transverse momentum or "missing E_T" ($\vec{\not{E}}_T$) indicates the presence of one or more particles that carried away the missing energy without being detected (*e.g.*, a neutrino).

3.6.1 $\vec{\not{E}}_T$ Reconstruction

The first step in reconstructing the missing E_T is calculating the uncorrected missing E_T measured by the calorimeter

$$\vec{\not{E}}_T^{\,\text{uncorr}} = -\sum_i \vec{p}_i^{\,\text{cell}};\tag{3.14}$$

where $\vec{p}_i^{\,\text{cell}}$ is a three-momentum with a magnitude equal to the energy of the ith cell in the EM and fine hadronic layers of the calorimeter, and pointing from the primary vertex to the center of the cell. The course hadronic cells are not included because they can contain a substantial amount of noise.

Muons, which only deposit a couple GeV of energy in the calorimeter, will contribute to

$\vec{E}_T^{\text{uncorr}}$. Additionally, reconstructed electrons, photons, and jets have corrected energies that are more accurate than the simple sum of the calorimeter cells inside those objects. Therefore, the missing E_T is corrected by replacing the uncorrected energy with the corrected energy for each reconstructed electron, muon, photon, and jet in the event. The corrected missing E_T is then given by

$$\vec{E}_T^{\,\text{corr}} = \vec{E}_T^{\,\text{uncorr}} - \sum_{\text{objects}} \left(\vec{p}^{\,\text{object}} - \sum_{i \in \text{object}} \vec{p}_i^{\,\text{cell}} \right) ; \qquad (3.15)$$

where the outer sum is over reconstructed objects, $\vec{p}^{\,\text{object}}$ is the corrected momentum of the object, and i runs over the EM and fine hadronic calorimeter cells in the electron, photon, or jet, or along the interpolated path of the muon through the calorimeter.

Chapter 4

Samples

This chapter describes the dataset analyzed and the samples used to compare the data to the prediction. To be considered for analysis, the data was required to pass trigger and data quality requirements as explained in section 4.1. Section 4.2 describes the signal and background samples containing a high p_T lepton, which were generated by MC simulation. The backgrounds from multijet events in which a jet mimics a lepton in the final state were estimated from data as described in section 4.3.

4.1 Dataset

This analysis uses data collected by the D0 experiment in Run IIa of the Tevatron (2002 - 2006) and reconstructed as described in chapter 3. The full dataset has been skimmed down to two smaller samples: the EMINCLUSIVE sample for the $e\nu q\bar{q}$ channel and MUINCLUSIVE for the $\mu\nu q\bar{q}$ channel. The EMINCLUSIVE dataset contains approximately 355×10^6 events and includes all events with a very loose reconstructed EM object. The MUINCLUSIVE dataset contains approximately 330×10^6 events and includes all events with a very loose reconstructed muon.

4.1.1 Trigger Requirements

Each event in the $e\nu q\bar{q}$ channel analysis is required to have fired at least one of the triggers in a suite of single electron triggers or one of five electron+jets triggers. The suite of single electron triggers is comprised of calorimeter (EM shower) and calorimeter+track (EM shower and a track) triggers requiring an electron with $p_T > 8$ to 30 GeV, depending on the trigger. The electron+jets triggers require at least one electron with $p_T > 15$ GeV and at least two jets with $p_T > 15$ to 30 GeV, depending on the trigger. After final event selection (chapter 5), the selected electron is required to fulfill the criteria of at least one of the triggers responsible for that event passing the trigger selection. This ensures that the selected electron was responsible for firing the trigger and the trigger efficiency can be measure accordingly[1]. The resulting trigger efficiency for the $e\nu q\bar{q}$ channel is $98^{+2}_{-3}\%$.

For the $\mu\nu q\bar{q}$ channel, characteristic trigger efficiencies for muon+jets or single muon triggers are 60% and 70%, respectively. Using both muon+jets and single muon triggers can supply a total trigger efficiency of around 90%. However, accurate modeling for this trigger selection is currently unavailable, which disallows its use in this analysis. Instead, all events in the available dataset are selected that satisfy the kinematic selection requirements (chapter 5), with no specific trigger requirement. Detailed studies of this selection indicate that the efficiency in this kinematic region is very nearly 100% [36, 37, 38]. Further studies specific to this analysis were performed and are outlined in appendix A. The results are consistent with a residual inefficiency of less than 3% with effects on the shapes of differential distributions of less then 5%. Appendix A outlines a means of propagating this shape-dependent uncertainty into the analysis. As a conservative estimated, the overall trigger efficiency is taken to be greater than 95% at a 67% confidence level.

---

[1] In the cases where the event only fired an electron+jets trigger, the selected jets are also require to fulfill the jet criteria for the trigger.

4.1.2   Luminosity and Data Quality

The instantaneous luminosity steadily decreases throughout a store (round of collisions) as antiprotons and protons are removed by collisions and the beams become diffuse. When data is being collected, the average luminosity is measured in blocks of about four minutes and saved to a database. Each luminosity block also records whether any detector subsystems (tracking, calorimeter, muon, luminosity, or trigger/data acquisition systems) had problems during that period that might compromised the quality of the data. Also, the D0 data quality group performs additional checks on the recorded data to identify events of poor quality, for example excess noise in the calorimeter, and composes a list of "bad" luminosity blocks.

Data used in this analysis are required to satisfy quality requirements ensuring that all subsystems were operating adequately and that the corresponding luminosity block was not corrupted or listed as "bad" by the data quality group. The total integrated luminosity is determined by summing the luminosity blocks that meet the data quality criteria[2]. The resulting integrated luminosity is 1.07fb^{-1} for each of the two channels with an uncertainty of 6.1%.

4.2   Simulated Samples

To compare the data with prediction, SM signal and background processes containing a lepton in the final state are simulated by MC event generators. The event-level simulation is processed by a GEANT-based simulation of the D0 detector and overlayed with zero bias events from data to produce a detector-level simulation of raw detector data. The simulated events (or MC events) are then reconstructed using the same software programs as the detector data.

All MC samples were generated at a center of mass energy of $\sqrt{s} = 1.96$ TeV assuming

---

[2]The luminosity blocks must also have been using at least one of the triggers required by the analysis.

a t quark mass of $m_t = 175.0$ GeV and using the CTEQ6L1 [39] PDFs. All MC samples used in this analysis are listed in tables 4.1-4.7, along with the number of events and cross section times branching fraction for each sample.

The diboson (signal) events were generated at LO with the PYTHIA [40] MC generator. To improve the accuracy of the signal modeling, these samples were then corrected to the NLO prediction, as described in section 6.8. Although this analysis is measuring WW and WZ production, ZZ is also included as signal ($\sim$2% contribution) because $ZZ \rightarrow \ell\ell q\bar{q}$ is experimentally identical to $WZ \rightarrow \ell\nu q\bar{q}$ in cases where one of the leptons is not detected. Each diboson process was generated with inclusive decays and all WW, WZ, and ZZ events that pass the selection are considered signal[3].

The fixed order matrix element (FOME) generator ALPGEN [41] was used to generate the $t\bar{t}$, W+jets, and Z+jets events to LO at parton-level. Similarly, the FOME generator COMPHEP [42] was used to produce the single t quark MC samples at parton-level. Each ALPGEN and COMPHEP event is passed to PYTHIA for parton showering and hadronization.

PYTHIA occasionally adds heavy flavor jets during hadronization, which invalidates the exclusive heavy and light flavor jet matching done by ALPGEN [43]. Therefore, all W+jets and Z+jets samples have undergone the process of "PYTHIA heavy flavor skimming" to remove all events with heavy flavor jets added by PYTHIA [44].

4.2.1   Normalization

The diboson and single-top MC samples are normalized to the NLO SM prediction [1, 45]. The $t\bar{t}$ samples are normalized to next-to-next-to-leading order (NNLO) theory prediction [46]. The Z+jets (ee, $\mu\mu$, and $\tau\tau$) samples are normalized by the cross sections determined by the MC generator with an additional NLO k-factor correction of 1.23. The $Z + heavy\ flavor\ partons$ receive an additional heavy flavor k-factor correction of 1.5 for a total correction factor of 1.85 (1.23 $\times$ 1.5). The overall k-factor for W+jets is left as a

---

[3]For example, this includes $WZ \rightarrow \ell\nu\tau\tau$, which can provide a small cross efficiency if both taus decay hadronically and appear as jets.

50

free parameter to be determined by the fit in the final cross section calculation as explained in chapter 9. Because the fit value for the W+jets k-factor is not known until the end of the analysis, the k-factor is estimated by matching the normalization in data. An estimate of 1.53 (1.4×1.09) is used for the W+light flavor partons samples; where 1.4 is a nominal NLO k-factor and the additional factor of 1.09 was determined by a data-to-MC comparison in the two lepton channels. The W+heavy flavor partons receive an additional heavy flavor k-factor correction of 1.5 for a total correction factor of 2.29 ($1.4 \times 1.5 \times 1.09$).

Table 4.1: Diboson samples generated with PYTHIA along with corresponding initial number of events and cross sections from NLO theory predictions [1].

Process	Sample Size	$\sigma_{NLO}[pb]$
$WW \rightarrow$ inclusive	2391k	12.4
$WZ \rightarrow$ inclusive	672k	3.7
$ZZ \rightarrow$ inclusive	671k	1.42

Table 4.2: Single-top samples generated with COMPHEP+PYTHIA along with corresponding initial number of events and LO cross sections times branching fractions from COMPHEP generator.

Process	Sample Size	$\sigma_{LO} \times$BF $[pb]$
$tb \rightarrow e\nu bb$	93k	0.0978
$tb \rightarrow \mu\nu bb$	76k	0.0978
$tb \rightarrow \tau\nu bb$	122k	0.0978
$tqb \rightarrow e\nu bqb$	130k	0.220
$tqb \rightarrow \mu\nu bqb$	138k	0.220
$tqb \rightarrow \tau\nu bqb$	117k	0.220

4.3   Estimation of Multijet Background

In general, the efficiency for multijet events to fake a lepton and pass all selection cuts is minute; however, because the cross section for QCD processes is large, the contribution from multijet events cannot be ignored. Simulating the multijet background is possible; however, it is impractical in practice because it would take an incredibly large sample of generated

Table 4.3: $t\bar{t}$+jets samples generated with ALPGEN+PYTHIA along with corresponding initial number of events and NNLO cross sections times branching fractions.

Process	Sample Size	$\sigma_{NNLO} \times$BF [pb]
$t\bar{t} \to \ell\nu\ell\nu b\bar{b} + 0$lp^a	224k	0.4388
$t\bar{t} \to \ell\nu\ell\nu b\bar{b} + 1$lp	96k	0.1825
$t\bar{t} \to \ell\nu\ell\nu b\bar{b} + \geq 2$lp	148k	0.0899
$t\bar{t} \to \ell\nu b\bar{b}jj + 0$lp	283k	1.8312
$t\bar{t} \to \ell\nu b\bar{b}jj + 1$lp	98k	0.7605
$t\bar{t} \to \ell\nu b\bar{b}jj + \geq 2$lp	93k	0.3743

^{a}lp stands for light parton (gluon or u, d, or s quark).

Table 4.4: $Z \to ee$+jets samples generated with ALPGEN+PYTHIA along with corresponding initial number of events, matched ALPGEN LO cross sections times branching fractions, and NLO k-factors.

Process	(Z/γ Mass)	Sample Size	$\sigma_{LO} \times$BF [pb]	K-factor
$Z \to ee + 0$lp	(15-60)	562k	336.212	
$Z \to ee + 1$lp	(15-60)	427k	39.420	
$Z \to ee + 2$lp	(15-60)	164k	10.297	
$Z \to ee + \geq 3$lp	(15-60)	78k	3.062	
$Z \to ee + 0$lp	(60-130)	1025k	140.291	
$Z \to ee + 1$lp	(60-130)	177k	42.270	
$Z \to ee + 2$lp	(60-130)	83k	10.466	1.23
$Z \to ee + \geq 3$lp	(60-130)	77k	3.416	
$Z \to ee + 0$lp	(130-250)	94k	0.9086	
$Z \to ee + 1$lp	(130-250)	84k	0.3654	
$Z \to ee + 2$lp	(130-250)	87k	0.0986	
$Z \to ee + \geq 3$lp	(130-250)	75k	0.0568	
$Z \to ee + 2b + 0$lp	(60-130)	230k	0.9878	
$Z \to ee + 2b + 1$lp	(60-130)	48k	0.3676	
$Z \to ee + 2b + \geq 2$lp	(60-130)	21k	0.1475	1.23×1.5
$Z \to ee + 2c + 0$lp	(60-130)	47k	3.0500	
$Z \to ee + 2c + 1$lp	(60-130)	43k	1.0725	
$Z \to ee + 2c + \geq 2$lp	(60-130)	22k	0.4237	

Table 4.5: $Z \to \mu\mu$+jets samples generated with ALPGEN+PYTHIA along with corresponding initial number of events, matched ALPGEN LO cross sections times branching fractions, and NLO k-factors.

Process	$(Z/\gamma$ Mass)	Sample Size	$\sigma_{LO} \times$BF $[pb]$	K-factor
$Z \to \mu\mu + $0lp	(15-60)	552k	334.138	
$Z \to \mu\mu + $1lp	(15-60)	423k	38.627	
$Z \to \mu\mu + $2lp	(15-60)	163k	10.146	
$Z \to \mu\mu + \geq$3lp	(15-60)	76k	3.101	
$Z \to \mu\mu + $0lp	(60-130)	985k	139.758	
$Z \to \mu\mu + $1lp	(60-130)	198k	41.858	1.23
$Z \to \mu\mu + $2lp	(60-130)	93k	10.375	
$Z \to \mu\mu + \geq$3lp	(60-130)	86k	3.331	
$Z \to \mu\mu + $0lp	(130-250)	101k	0.9049	
$Z \to \mu\mu + $1lp	(130-250)	91k	0.3641	
$Z \to \mu\mu + $2lp	(130-250)	87k	0.0990	
$Z \to \mu\mu + \geq$3lp	(130-250)	73k	0.0555	
$Z \to \mu\mu + 2b + $0lp	(60-130)	267k	0.9898	
$Z \to \mu\mu + 2b + $1lp	(60-130)	48k	0.3698	
$Z \to \mu\mu + 2b + \geq$2lp	(60-130)	22k	0.1474	1.23×1.5
$Z \to \mu\mu + 2c + $0lp	(60-130)	47k	3.0463	
$Z \to \mu\mu + 2c + $1lp	(60-130)	43k	1.0737	
$Z \to \mu\mu + 2c + \geq$2lp	(60-130)	23k	0.4124	

Table 4.6: $Z \to \tau\tau$+jets samples generated with ALPGEN+PYTHIA along with corresponding initial number of events, matched ALPGEN LO cross sections times branching fractions, and NLO k-factors.

Process	$(Z/\gamma$ Mass)	Sample Size	$\sigma_{LO} \times$BF $[pb]$	K-factor
$Z \to \tau\tau + $0lp	(15-60)	535k	335.434	
$Z \to \tau\tau + $1lp	(15-60)	431k	38.536	
$Z \to \tau\tau + $2lp	(15-60)	167k	10.265	
$Z \to \tau\tau + \geq$3lp	(15-60)	76k	3.062	
$Z \to \tau\tau + $0lp	(60-130)	1557k	139.743	
$Z \to \tau\tau + $1lp	(60-130)	363k	42.011	1.23
$Z \to \tau\tau + $2lp	(60-130)	178k	10.798	
$Z \to \tau\tau + \geq$3lp	(60-130)	156k	3.3498	
$Z \to \tau\tau + $0lp	(130-250)	100k	0.9085	
$Z \to \tau\tau + $1lp	(130-250)	90k	0.3723	
$Z \to \tau\tau + $2lp	(130-250)	80k	0.1022	
$Z \to \tau\tau + \geq$3lp	(130-250)	71k	0.0551	
$Z \to \tau\tau + 2b + $0lp	(60-130)	93k	0.9893	
$Z \to \tau\tau + 2b + $1lp	(60-130)	182k	0.3696	
$Z \to \tau\tau + 2b + \geq$2lp	(60-130)	87k	0.1476	
$Z \to \tau\tau + 2c + $0lp	(60-130)	39k	3.0529	1.23×1.5
$Z \to \tau\tau + 2c + $1lp	(60-130)	43k	1.0764	
$Z \to \tau\tau + 2c + \geq$2lp	(60-130)	21k	0.4203	

Table 4.7: $W \to \ell\nu$+jets samples generated with ALPGEN+PYTHIA along with corresponding initial number of events, matched ALPGEN LO cross sections times branching fractions, and k-factors as explained in section 4.2.1.

Process	Sample Size	$\sigma_{LO} \times$BF $[pb]$	K-factor
$W \to \ell\nu + $0lp	2257k	4574.361	
$W \to \ell\nu + $1lp	2926k	1273.940	
$W \to \ell\nu + $2lp	1564k	298.564	
$W \to \ell\nu + $3lp	789k	70.561	1.4×1.09
$W \to \ell\nu + $4lp	779k	15.831	
$W \to \ell\nu + \geq$5lp	87k	5.7602	
$W \to \ell\nu + 2b + $0lp	739k	19.182	
$W \to \ell\nu + 2b + $1lp	261k	7.9391	
$W \to \ell\nu + 2b + $2lp	171k	2.6365	
$W \to \ell\nu + 2b + \geq$3lp	164k	1.0688	$1.4 \times 1.5 \times 1.09$
$W \to \ell\nu + 2c + $0lp	482k	71.145	
$W \to \ell\nu + 2c + $1lp	336k	29.854	
$W \to \ell\nu + 2c + \geq$2lp	372k	13.740	

events to supply sufficient statistics after applying the selection requirements. Therefore, the multijet background for both channels is estimated using orthogonal datasets as described in the follow sections.

4.3.1 Electron Multijet Background

To estimate the amount of multijet background present in the $e\nu q\bar{q}$ channel, a form of the so-called matrix method is used to calculate weights for a *loose-not-tight* dataset and the MC samples. This method allows one to obtain both the shape and normalization for the multijet background, plus account for signal (real electron) contamination in the *loose-not-tight* sample.

First, a *loose* dataset is defined as the data obtained by applying all of the $e\nu q\bar{q}$ channel selection requirements described in chapter 5, except for the electron likelihood cut. The subset of events in the *loose* dataset that pass the electron likelihood cut constitute the *tight* dataset that is selected for the analysis. The *loose-not-tight* dataset is then defined as the events in the *loose* dataset that are not in the *tight* dataset.

The *loose-not-tight* dataset (of N_{lnt} events) is composed of some events with a real electron ($N_{lnt}^{\text{electron}}$) plus multijet events where a jet has mimicked an electron ($N_{lnt}^{\text{multijet}}$):

$$N_{lnt} = N_{lnt}^{\text{electron}} + N_{lnt}^{\text{multijet}}. \tag{4.1}$$

Similarly for the *loose* and *tight* selections:

$$N_{loose} = N_{loose}^{\text{electron}} + N_{loose}^{\text{multijet}} \tag{4.2}$$

$$N_{tight} = N_{tight}^{\text{electron}} + N_{tight}^{\text{multijet}}. \tag{4.3}$$

If it is know that ϵ_{real} is the efficiency for real electrons satisfying the *loose* requirements to also pass the *tight* requirements and ϵ_{fake} is the efficiency for jets that faked electrons and

passed the *loose* requirement to also pass the *tight* requirement, then it follows that

$$N_{tight}^{\text{electron}} = \epsilon_{\text{real}} \cdot N_{loose}^{\text{electron}} \tag{4.4}$$

$$N_{tight}^{\text{multijet}} = \epsilon_{\text{fake}} \cdot N_{loose}^{\text{multijet}}. \tag{4.5}$$

The *loose-not-tight* sample is the *loose* sample minus the events in the *tight* sample, thus

$$N_{lnt} = N_{loose} - N_{tight} = (1 - \epsilon_{\text{fake}}) \cdot N_{loose}^{\text{multijet}} + (1 - \epsilon_{\text{real}}) \cdot N_{loose}^{\text{electron}}. \tag{4.6}$$

Solving equation 4.6 for $N_{loose}^{\text{multijet}}$ yields

$$N_{loose}^{\text{multijet}} = \left(\frac{1}{1 - \epsilon_{\text{fake}}} \right) \cdot N_{lnt} - \left(\frac{1 - \epsilon_{\text{real}}}{1 - \epsilon_{\text{fake}}} \right) \cdot N_{loose}^{\text{electron}} \tag{4.7}$$

$$= \left(\frac{1}{1 - \epsilon_{\text{fake}}} \right) \cdot N_{lnt} - \left(\frac{1 - \epsilon_{\text{real}}}{1 - \epsilon_{\text{fake}}} \right) \cdot \frac{N_{tight}^{\text{electron}}}{\epsilon_{real}}. \tag{4.8}$$

Finally, the number of multijet events in the *tight* dataset is given by

$$N_{tight}^{\text{multijet}} = \epsilon_{\text{fake}} \cdot N_{loose}^{\text{multijet}} \tag{4.9}$$

$$= \left(\frac{\epsilon_{\text{fake}}}{1 - \epsilon_{\text{fake}}} \right) \cdot N_{lnt} - \left(\frac{\epsilon_{\text{fake}}}{1 - \epsilon_{\text{fake}}} \right) \cdot \left(\frac{1 - \epsilon_{\text{real}}}{\epsilon_{\text{real}}} \right) \cdot N_{tight}^{\text{electron}}. \tag{4.10}$$

There are two terms in equation 4.10, one containing N_{lnt} and a second containing $N_{tight}^{\text{electron}}$. The N_{lnt} term represents the shape of the multijet estimate coming from using the *loose-not-tight* dataset with a factor that sets the normalization. The $N_{tight}^{\text{electron}}$ term represents the real electron contamination in the *loose-not-tight* dataset that must be subtracted to get an accurate multijet estimation. Of course, $N_{tight}^{\text{electron}}$ is given by the tight MC samples. Thus, equation 4.10 gives the shape[4] and normalization of multijet background in terms of the

---

[4] $N_{tight}^{\text{multijet}}$ was presented as the total number of multijet events in the *tight* dataset, but this procedure can also be applied to individual bins in a histogram. In that case, equation 4.10 gives the amount of multijet events in each bin, thus defining the shape.

loose-not-tight dataset and the *tight* MC samples. Now going back to the full *tight* dataset, which is actually what needs to be model, one finds

$$N_{tight} = N_{tight}^{\mathrm{multijet}} + N_{tight}^{\mathrm{electron}} \tag{4.11}$$

$$= \left(\frac{\epsilon_{\mathrm{fake}}}{1 - \epsilon_{\mathrm{fake}}}\right) \cdot N_{lnt} + \left[1 - \left(\frac{\epsilon_{\mathrm{fake}}}{1 - \epsilon_{\mathrm{fake}}}\right) \cdot \left(\frac{1 - \epsilon_{\mathrm{real}}}{\epsilon_{\mathrm{real}}}\right)\right] \cdot N_{tight}^{\mathrm{electron}}. \tag{4.12}$$

Therefore, the *tight* data are modeled by the *loose-not-tight* dataset weighted by $\left(\frac{\epsilon_{\mathrm{fake}}}{1-\epsilon_{\mathrm{fake}}}\right)$ plus the *tight* MC samples weighted by $\left[1 - \left(\frac{\epsilon_{\mathrm{fake}}}{1-\epsilon_{\mathrm{fake}}}\right) \cdot \left(\frac{1-\epsilon_{real}}{\epsilon_{real}}\right)\right]$. In this way, the multijet background is modeled using the *loose-not-tight* dataset with the real electron contamination in that sample accounted for by scaling down the MC by the appropriate amount (about 1.5-2.5% depending on the sample).

The efficiency ϵ_{real} is calculated by taking the ratio of the efficiencies for electrons in data to pass the *tight* and *loose* requirements, which were measured by the D0 electron/photon identification group using the tag-and-probe method [47] with $Z \to ee$ data. The result is $\epsilon_{real} = 0.87 \pm 0.01$.

The efficiency ϵ_{fake} is estimated from data by exploiting the fact that multijet background is the dominant contribution for low $\not{E}_T$ (multijet events do not contain $W \to e\nu$). For the measurement of ϵ_{fake}, the transverse W mass cut is temporarily removed and the $\not{E}_T$ cut is changed to $\not{E}_T < 10$ GeV. This inverted-$\not{E}_T$ selection is applied to the *tight* and *loose* datasets and MC samples. The *loose* (*tight*) MC samples are then used to subtract any real electron contamination present in the *loose* (*tight*) datasets. Then a plot is made showing the fraction of events in the real-electron-subtracted *loose* sample that are also in the real-electron-subtracted *tight* sample, as a function of electron p_T. That distribution is fit to a line+exponential function $(\epsilon_{\mathrm{fake}}(p_T) = p_0 + p_1 \cdot p_T + p_2 \cdot \exp(-p_3 \cdot p_T))$ to obtain the fake rate as a function of the p_T of the (fake) electron, as shown in figure 4.1. For low p_T the fake rate increases as the jet and electron reconstructions get worse, while at high p_T the increase is believed to be the result of γ+jets events in which the photon converts to a collinear e^+e^-

pair that is identified as a single electron.
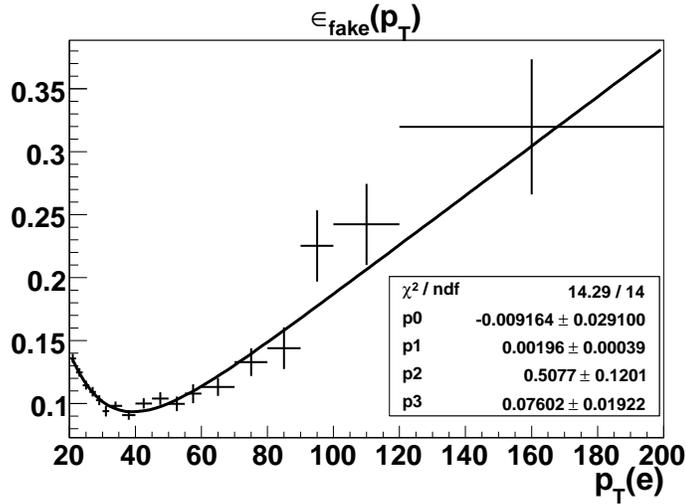
Figure 4.1: Electron fake rate, ϵ_{fake}, for multijet events as a function of electron p_T.

The efficiencies ϵ_{real} and $\epsilon_{\text{fake}}(p_T)$ are used with Eq 4.12 to calculate an event-by-event weight for the *loose-not-tight* dataset and *tight* MC samples. The uncertainty on the determination of $\epsilon_{\text{fake}}(p_T)$ is assigned as a shape-dependent systematic uncertainty on the multijet estimation. The change in shape is determined by calculating new event weights for the *loose-not-tight* dataset and *tight* MC samples after shifting $\epsilon_{\text{fake}}(p_T)$ by $\pm 1\sigma$ from the nominal fit value.

4.3.2 Muon Multijet Background

A multijet event may have a jet reconstructed as a muon either because the jet was not completely stopped by the hadronic calorimeter (punch-through) or the jet actually contains a real muon from the decay of a pion, kaon, or possibly a heavy quark. In either case, the reconstructed muon is usually very close to the jet so the most efficient way to minimize this background is by requiring muons to be isolated. Thus, the multijet contribution is estimated using a multijet-enriched sample of data obtained by changing the muon isolation cuts (chapter 5) to anti-isolation cuts. The *anti-isolated* dataset is defined by selecting muon events with the selection requirements described in chapter 5, except for the CalorimeterHalo

and TrackHalo criteria, which are reversed to be CalorimeterHalo > 2.5 GeV and TrackHalo > 2.5 GeV.

The *anti-isolated* dataset will contain some real muon events, therefore, the same anti-isolation selection is applied to the MC samples to determine the real muon contamination. This real muon contribution is accounted for by applying a weight to each of the *anti-isolated* data as follows. Six distributions are selected that are sensitive to the shape of the multijet background: transverse W mass, dijet mass, muon p_T, $\displaystyle{\not}E_T$, $\Delta\phi$ between the muon and $\vec{\displaystyle{\not}E}_T$, and second jet p_T (figure 4.2). For each of the distributions, a bin-by-bin weight is calculated to scale the *anti-isolated* data to the shape of the multijet background for that distribution. For a given event in the *anti-isolated* dataset, the six weights are averaged to get a single weight for scaling that event. The resulting sample of weighted *anti-isolated* data should reliably reproduce the shape of the multijet background. The estimated multijet shape is shown by a solid red line for the distributions in figure 4.2. As expected, the shape is similar to the difference between data and MC (green line) for each distribution. By averaging the weights from many distributions, the resulting multijet estimation is less sensitive to statistical effects from any single distribution.

The kinematics of the multijet background are not expected to be influenced much by the muon isolation criteria, so the multijet events in the *anti-isolated* dataset should provide a good estimate for the multijet events in the data selected for the final analysis (the *isolated* dataset). A study evaluating the effects of isolation criteria on the multijet estimation is presented in appendix B and a shape-dependent systematic uncertainty is assigned to the multijet estimation as explained in the appendix.

The number of multijet events in the *isolated* dataset (final selection) will generally not be the same as the number in the *anti-isolated* dataset. To estimate the number of multijet events in the final selection a fit is performed using the transverse W mass distribution. The parameter uncertainty in the fit is taken from the diagonalized Hessian error matrix available within the fitting program, MINUIT [48]. A binned fit to the transverse W mass
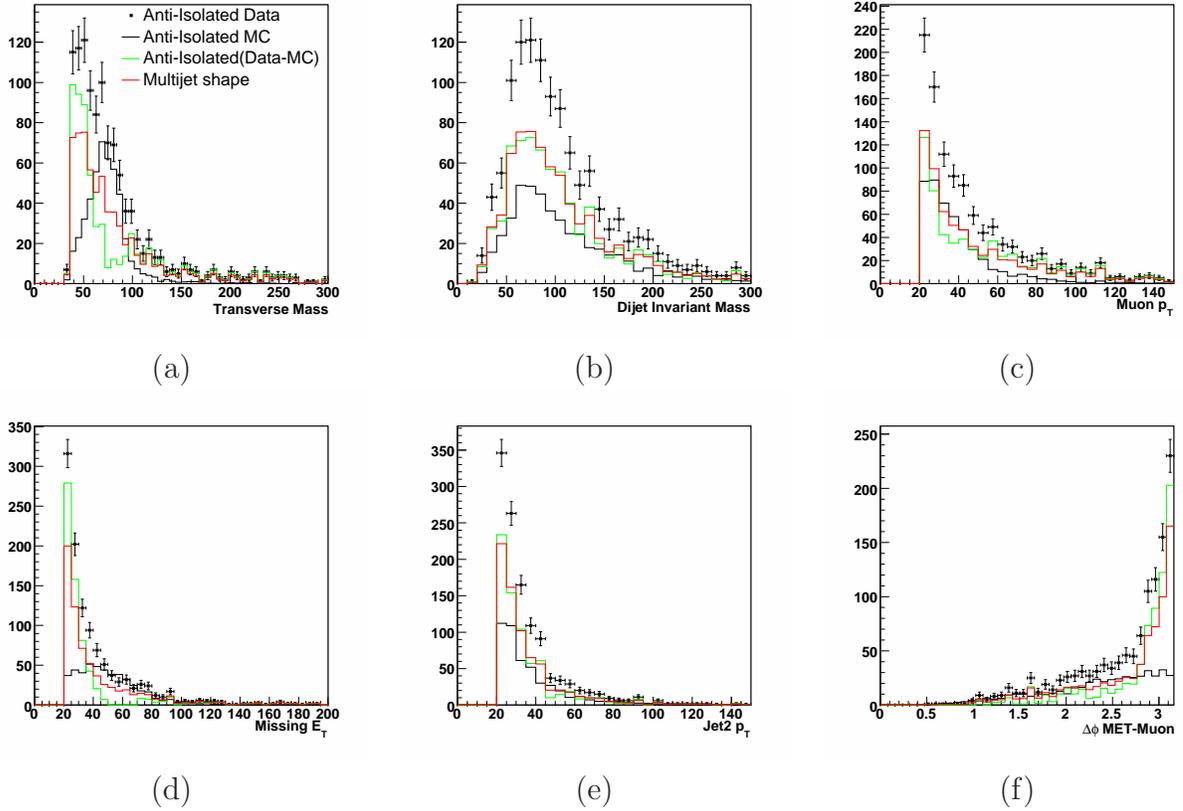
Figure 4.2: Distributions in the $\mu\nu q\bar{q}$ channel after all selection criteria of (a) transverse W mass, (b) dijet mass, (c) muon p_T, (d) $\not{E}_T$, (e) second jet p_T, and (f) $\Delta\phi$ between the muon and $\vec{\not{E}}_T$ for *anti-isolated* data (black squares), *anti-isolated* MC (black solid line), MC-subtracted *anti-isolated* data (green solid line), and multijet shape (red solid line) obtained using the average weight.

distribution is performed with both equidistant and non-equidistant bin sizes over the range $35 \text{ GeV} \geq m_T^W \geq 300 \text{ GeV}$. The fit minimizes the Poisson negative log-likelihood function:

$$-2\ln\mathcal{L} = -2\ln\left[\prod_{i=0}^{n_{\text{bins}}} \frac{\left(N_i^{\text{iso,MC}} + \alpha N_i^{\text{anti,multijet}}\right)^{N_i^{\text{iso,data}}} e^{-\left(N_i^{\text{iso,MC}}+\alpha N_i^{\text{anti,multijet}}\right)}}{\left(N_i^{\text{iso,data}}\right)!}\right] \quad (4.13)$$

$$= -2\sum_{i=1}^{n_{\text{bins}}}\left[N_i^{\text{iso,data}}\ln\left(N_i^{\text{iso,MC}} + \alpha N_i^{\text{anti,multijet}}\right) - \left(N_i^{\text{iso,MC}} + \alpha N_i^{\text{anti,multijet}}\right)\right] + \text{const.};$$

where, for the ith bin, $N_i^{\text{iso,data}}$ is the number of *isolated* data, $N_i^{\text{iso,MC}}$ is the number of *isolated* MC events (set to the number of events predicted by the SM for each process), $N_i^{\text{anti,multijet}}$ is the number of *anti-isolated* multijet events, and α is the fit parameter. The estimated number of multijet events in the *isolated* dataset is then $N^{\text{iso,multijet}} = \alpha \sum_{i=1}^{n_{\text{bins}}} N_i^{\text{anti,multijet}}$. The fit is performed about 30 times, each time using a different binning to estimate the uncertainty related to the choice of binning. The resulting distribution of outcomes for $N^{\text{iso,multijet}}$ (one entry for each fit) is shown in Fig 4.3. This distribution is fit to a Gaussian to determine the mean, 295, which is the estimate for the number multijet events in the range $35 \text{ GeV} \geq m_T^W \geq 300 \text{ GeV}$ for the *isolated* dataset. Approximately 11% of the MC events fall above the transverse mass range used for the fit, therefore, the total multijet contribution for the $\mu\nu q\bar{q}$ channel is estimated to be 327 events.

The width of the Gaussian in figure 4.3a (11 events) is taken as the systematic uncertainty on the multijet normalization coming from the binning; however, this in not the main source of uncertainty on the multijet normalization. Each of the approximately 30 fits (equation 4.13) has an uncertainty on how well the normalization is determined. In each case, the fit uncertainty on $N^{\text{iso,multijet}}$ ($\alpha N^{\text{anti,multijet}}$) was around 91 events. There is also a potential bias from including the signal in the fit with the SM cross section (since this is what is trying to be measured). To assess a possible signal bias, the normalization is also measured with the signal scaled by zero and two times the SM prediction; however, the effect is found to be negligible compared to the fit uncertainty. The total uncertainty assigned to

(a)   (b)
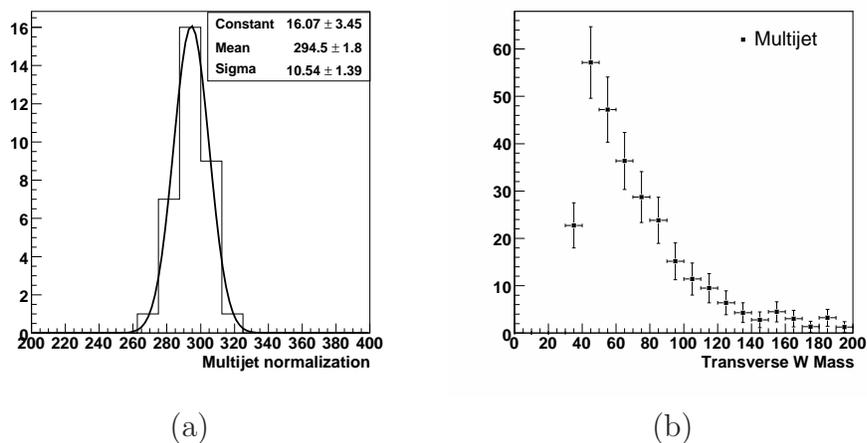
Figure 4.3: (a) The estimated number of multijet events in the $\mu\nu q\bar{q}$ channel obtained from fits of the transverse W mass in the range 35 GeV $\geq m_T^W \geq 300$ GeV. (b) The transverse W mass distribution for the normalized estimate of the multijet background in the $\mu\nu q\bar{q}$ channel.

the normalization of the multijet background in the $\mu\nu q\bar{q}$ channel is 30%.

Chapter 5

Event Selection

This analysis is interested in events with a final state consistent with $WW/WZ \rightarrow \ell\nu q\bar{q}$; where $\ell = e$ or μ. Therefore, events are selected with exactly one high p_T lepton, large missing transverse energy (indicative of a neutrino), and at least two high p_T jets. Following are the details for the selection of each object as well as the global selection requirements.

5.1   Lepton Selection

Leptonic decays of W bosons are selected by requiring each event to contain an energetic reconstructed lepton satisfying strict quality criteria. The energy requirement for the lepton in both channels is $p_T \geq 20$ GeV. Also, all leptons must be separated from any jets by a radius $R \geq 0.5$[1]. The electrons for the $e\nu q\bar{q}$ channel must be reconstructed in the central calorimeter ($|\eta_{\mathrm{det}}| \leq 1.1$) with a matching track in the central tracking system that has a transverse momentum $p_T^{track} \geq 5$ GeV. The quality criteria (described in section 3.3.3) for the selected electron are:

- EM-fraction: $f_{\mathrm{EM}} \geq 0.9$,

- Isolation: $f_{\mathrm{iso}} \leq 0.15$,

---

[1]Muons inside jets are considered part of the jets and removed from the list of muons. An electron inside a jet is considered as an electron and the jet is removed (section 5.2).

- Shower shape: $\chi^2_{H-matrix} \leq 50$,

- Electron Likelihood: $\mathcal{L}_e \geq 0.85$.

Electron candidates passing these criteria are considered *tight* electrons. It is also convenient (*e.g.*, for the multijet estimation) to define *loose* electrons as electron candidates meeting all of these criteria except of the electron likelihood cut.

The muon selected for each event in the $\mu\nu q\bar{q}$ channel must be have a track in both the muon system and central tracking system with $|\eta_{\text{det}}| < 2$. The quality criteria (described in section 3.4.2) for the muon are:

- Muon system quality:

  - $|N_{\text{seg}}| = 3$,

  - ≥ 1 scintillator hit in both the A and BC layers,

  - ≥ 2 drift tube hits in the A and BC layers,

  - scintillator hit times must be within 10 ns of the beam crossing.

- Central track quality:

  - $\chi^2/\text{ndf} \leq 4$ for the match to the central track,

  - distance of closest approach to the beam line ≤ 0.02 cm for tracks with SMT hits,

  - distance of closest approach to the beam line ≤ 0.2 cm for tracks without SMT hits.

- Isolation:

  - TrackHalo ≤ 2.5 GeV

  - CalorimeterHalo ≤ 2.5 GeV

In addition to requiring a lepton meeting the above requirement, events that have more than one lepton are removed to reduce the background from $t\bar{t}$ and Z events with two

leptons in the final state. The $e\nu q\bar{q}$ channel vetoes events with a second electron having $p_T \geq 15$ GeV and $|\eta_{\text{det}}| \leq 2.5$ or with an isolated muon having $p_T \geq 15$ GeV and $|\eta_{\text{det}} \leq 2$. The $\mu\nu q\bar{q}$ channel vetoes events with a second muon with $p_T \geq 15$ GeV and matched to a central track without an isolation requirement or events with an electron with $p_T \geq 20$ GeV and $|\eta_{\text{det}}| \leq 2.5$. As mentioned in section 4.2, some events with two leptons, namely ZZ events, may pass the veto because one of the leptons escapes detection and mimics a neutrino (missing energy). For example, this can happen when an electron goes into the inner-cryostat region (ICR) between the CC and EC or when a muon goes into the bottom hole of the muon system. In this case, ZZ appears identical to WZ, thus ZZ is included as a small ($\sim 2\%$) cross efficiency to the $WW + WZ$ signal.

By design, the muon veto in the $e\nu q\bar{q}$ channel will remove any event that could pass the $\mu\nu q\bar{q}$ channel selection. Similarly, the electron veto in the $\mu\nu q\bar{q}$ channel will remove any event that could pass the $e\nu q\bar{q}$ selection. In this way, the channels are constructed to be orthogonal.

5.2   Jet Selection

All jets are reconstructed and required to satisfy the basic quality requirements as described in section 3.5. Muons that fall inside a jet ($\Delta R(\mu, jet) \leq 0.5$) are considered part of that jet and included in the calculation of the jet's energy. Electrons will often be reconstructed as jets, so to avoid counting an electron as a jet, jets are removed if they overlap a reconstructed electron meeting the *loose* quality criteria and with $p_T \geq 15$ GeV. After jet energy scale corrections, additional smearing, shifting, and efficiency corrections are applied to jets in the MC simulation as described in chapter 6.

The requirements on the jet transverse momenta were determined by performing a limited pre-selection optimization by maximizing signal significance using the MC samples (via limit calculation). The leading jet p_T requirement is nearly 100% efficient for signal events and

the p_T requirement for the second jet is chosen to control the dominant background from W+jets. After all corrections have been applied, the events must pass the following selection:

- All jets must have $|\eta_{\mathrm{det}}| \leq 2.5$,

- At least two jets with $p_T \geq 20$ GeV,

- At least one jet with $p_T \geq 30$GeV.

Finally, so-called "ICR cleaning" is performed to remove events that contain selected jets located close to bad/noisy locations in the ICR.

5.3 $\not{E}_T$ and Global Selection Requirements

Neutrinos created in a collision will escape the detector without depositing any energy, which results in an imbalance in the total transverse momentum of the events (*i.e.*, $\vec{\not{E}}_T$). To ensure a reliable measurement, the $\vec{\not{E}}_T$ is corrected using the reconstructed electrons, muons (not inside a jet), and corrected jets for each event. To select events with high energy neutrinos from the decay of a W boson, the missing energy is required to be $\not{E}_T \geq 20$ GeV.

Assuming the event has a single W boson that decays to a lepton and neutrino, as is the case for the signal process, one can reconstruct the "transverse W mass" defined as $m_T^W = \sqrt{(E_T^W)^2 - (\vec{p}_T^W)^2}$; where E_T^W is the transverse energy of the reconstructed W (lepton $+ \vec{\not{E}}_T$) and $\vec{p}_T^W$ is the transverse momentum of the reconstructed W. Multijet events, which can sometimes fake leptons and $\not{E}_T$, tend to have low values for transverse W mass (m_T^W). To remove most of the multijet background, selected events must satisfy $m_T^W \geq 35$ GeV.

Lastly, the z position of the reconstructed primary vertex is required to be within 60 cm of the center of the detector and the absolute distance in z between the lepton-matched track and the primary vertex is required to be less than 1 cm.

Chapter 6

Monte Carlo Corrections

In an ideal world, MC simulations of SM processes and detector response would result in a perfect reproduction of what is seen in data. However, the complexity of nature ensures that, in reality, this is basically never the case. In order to more accurately model the data, additional studies are performed to determine the differences between simulation and data and then the simulation is modified to match the performance of the data.

6.1 Trigger Efficiency

In order to simulate the effect of the trigger selection in the MC samples, a trigger weight is applied to each MC event. The trigger weight for a given MC event represents the probability that the same event in the data would have passed the trigger requirements.

For the $e\nu q\bar{q}$ channel (single electron and electron+jets triggers), this probability is determined by the efficiency for the electron in the selected event to pass at least one of the available single electron triggers or for the electron and jets to pass the available electron+jets trigger. Only one of the electron+jets triggers and only a subset of the single electron triggers are active for any given trigger list. Therefore, the trigger efficiencies are assessed separately for each configuration. The trigger weight applied to each MC event is then calculated as the weighted average trigger efficiency, weighted by the amount of data collected (integrated

luminosity) for each configuration. The electron and jet efficiencies are parameterized in p_T and η_{det} and were measured by the D0 trigger group in $Z \rightarrow ee$ and μ+jets events [49, 50], respectively.

In the $\mu\nu q\bar{q}$ channel the trigger requirement is removed and it is assume that the trigger probability is $100^{+0}_{-5}\%$. Thus, the trigger efficiency correction is not applied. As mentioned in section 4.1.1, a 5% shape-dependent systematic uncertainty is applied to account for possible kinematic difference between data and MC introduced by not requiring any specific trigger.

6.2 Jet Smearing, Shifting, and Removal

Jets (and for that matter, all objects) are reconstructed using the same algorithms in the MC samples as in the data. Modeling of hadronization, detector response, and detector readout is far from trivial, so it is not surprising that jets in the simulated samples do not exactly match what is seen in data. After reconstruction and applying the JES correction (derived for data) as explained in section 3.5, jets in simulated samples have slightly higher energies (data-MC JES), better energy resolution, and more efficient reconstruction/identification (jet ID) than jets in data.

The relative differences between data and MC are measured by the D0 jet identification group using γ+jet events [51]. For various ranges (*i.e.*, bins) of photon p_T (p_T^γ), histograms are filled with the fractional difference in p_T between the photon and the jet

$$\Delta S = \frac{p_T^{\text{jet}} - p_T^\gamma}{p_T^\gamma}.$$

(6.1)

For high p_T^γ bins (and therefore high p_T^{jet}), the distribution of ΔS is Gaussian with a mean around zero. For low p_T ($\lesssim 40$ GeV) the distribution would be Gaussian except that the low tail is truncated due to inefficiencies in reconstructing/identifying low p_T jets. The truncation is described by an error function, therefore, as shown in figure 6.1 for the range 18 GeV $< p_T^\gamma < 23$ GeV, the ΔS distributions are fit to a Gaussian times an error function.

The difference between data and MC for the means of the Gaussian component of the ΔS distributions gives the relative data-MC JES, the difference in widths of the Gaussian component gives the resolution difference, and the difference in the error function component gives the relative data-MC jet ID efficiency. The MC samples are then corrected by shifting and smearing the jet energies according to the relative data-MC JES and jet resolution and randomly removing jets at a rate proportional to the ratio of the jet ID efficiency in data versus simulation. Systematic uncertainties for the shifting, smearing, and removal are dictated by the uncertainties on the fit parameters.
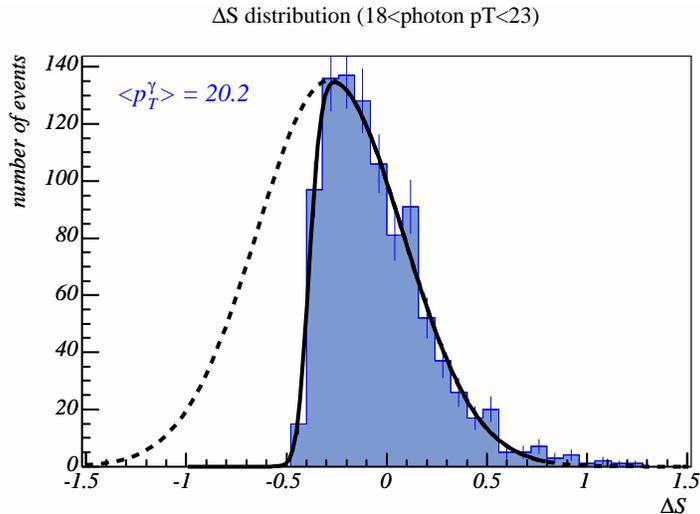
Figure 6.1: Distribution of ΔS for CC jets in data and 18 GeV $< p_T^\gamma <$ 23 GeV, fit to a Gaussian times an error function.

6.3 Lepton Smearing and Efficiencies

Like jets, MC simulation produces leptons with slightly better energy resolutions and higher reconstruction/identification efficiencies (lepton ID) than in data. The D0 electron/photon and muon identification groups use samples of $Z \rightarrow ee$ and $Z \rightarrow \mu\mu$ events to determine how much to smear the lepton energies in MC to match the width of the Z mass peak in data. The the tag-and-probe method [47] is used to measure the lepton ID efficiencies in data and simulation. Instead of randomly removing leptons at a rate proportional to the

ratio of data and MC efficiencies (as done with jets), the MC events are simply weighted by that factor (as done with trigger efficiencies). The primary advantage of weighting versus removal is maintaining the statistics (*i.e.*, number of events) in the MC samples. The systematic uncertainty from the smearing is negligible, but the lepton ID corrections are assigned an uncertainty arising mainly from the limited statistics in deriving the data and MC efficiencies.

6.4   Forward Muon Tracks

When muon tracks are mis-reconstructed, the resulting over-estimation of the muon p_T can generate false $\not{E}_T$. As shown in figure 6.2, there was an observed excess of events containing a single muon with $|\eta_{\mathrm{det}}| > 1.6$ that was back-to-back in ϕ with the $\vec{\not{E}}_T$. These events were confirmed to arise from poorly reconstructed tracks rather than trigger biases, MC modeling, or multijets background. To address this problem, the q/p_T significance of the track,

$$\Sigma_{q/p_T} = \frac{q/p_T}{\sigma_{q/p_T}},$$
(6.2)

was used to assess the quality of the track fit; where q is the charge of the muon (*i.e.*, ± 1) and σ_{q/p_T} is the uncertainty on the q/p_T measurement.
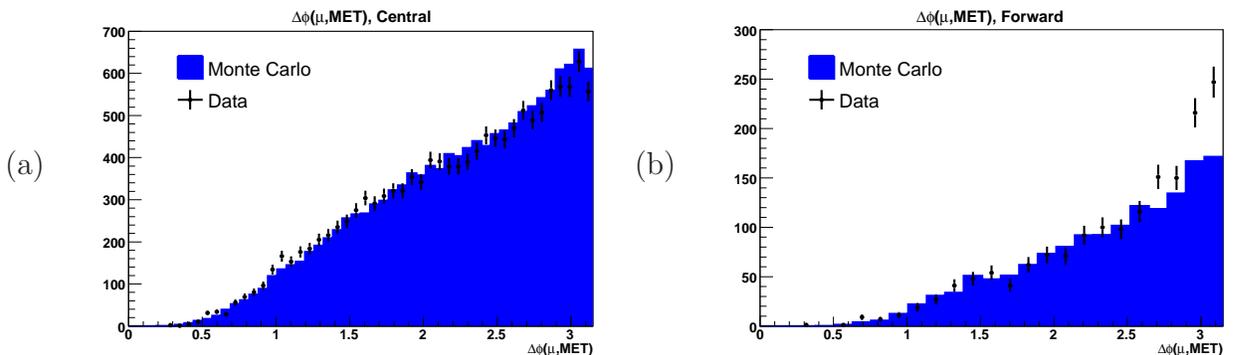
Figure 6.2: $\Delta\phi$ between muon and $\vec{\not{E}}_T$ for (a) central muons and (b) forward muons.

The distribution of σ_{q/p_T} is not properly modeled in the simulation because the muon

p_T smearing is not propagated to σ_{q/p_T}. Thus, an empirical, multiplicative correction of $\sqrt{2}$ is applied to σ_{q/p_T} in MC in order to achieve agreement with data. Figure 6.3 shows the distributions of Σ_{q/p_T} for central ($|\eta_{\mathrm{det}}| \leq 1.6$) and forward ($|\eta_{\mathrm{det}}| > 1.6$) muons, before and after the correction

$$\sigma_{q/p_T}^{\mathrm{MC}} \longrightarrow \sqrt{2} \cdot \sigma_{q/p_T}^{\mathrm{MC}}. \tag{6.3}$$
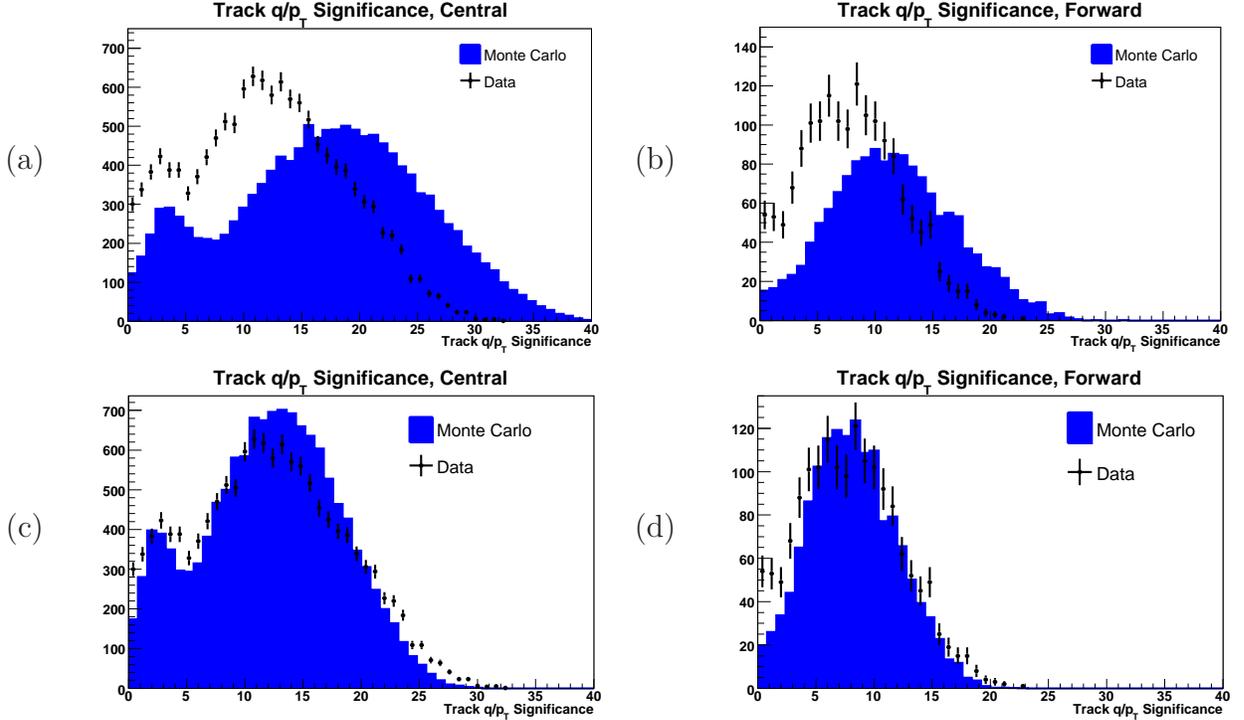
Figure 6.3: Track significance (Σ_{q/p_T}) for (a,c) central muons and (b,d) forward muons before and after the correction to $\sigma_{q/p_T}^{\mathrm{MC}}$ (equation 6.3).

Following the correction to $\sigma_{q/p_T}^{\mathrm{MC}}$, the significance spectrum for forward muons is separated into events where the muon and $\vec{\not{E}}_T$ are not back-to-back ($\Delta\phi(\mu, \vec{\not{E}}_T) \leq 2.5$) and events where the muon and $\vec{\not{E}}_T$ are back-to-back ($\Delta\phi(\mu, \vec{\not{E}}_T) > 2.5$). As seen in figure 6.4, there is an excess of data with low track significance for $\Delta\phi(\mu, \vec{\not{E}}_T) > 2.5$. The χ^2 agreement between data and MC for the $\Delta\phi(\mu, \vec{\not{E}}_T)$ distribution was used as the figure of merit (figure 6.5) for selecting a cut on Σ_{q/p_T} to remove events with poorly measured forward muons. The requirement was chosen to be $\Sigma_{q/p_T} > 5$ and the resulting distribution of $\Delta\phi(\mu, \vec{\not{E}}_T)$ is also shown in figure 6.5. No cut is placed on the track significance for central muons.
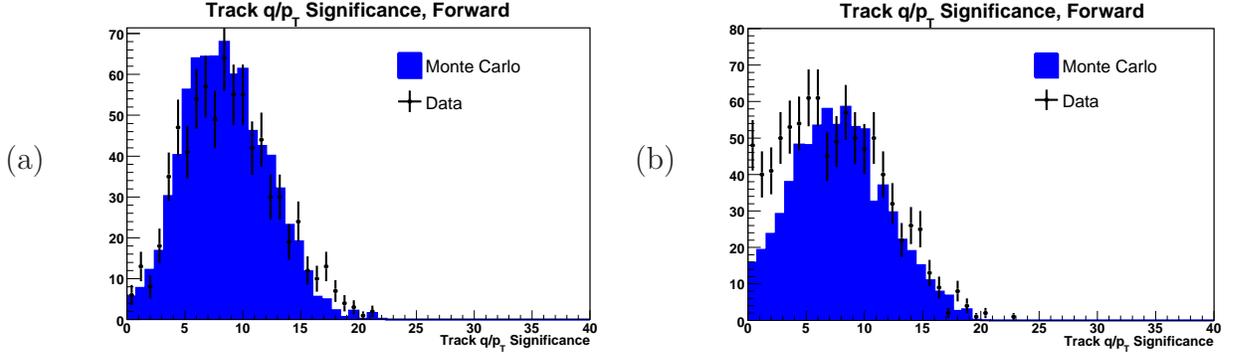
Figure 6.4: Track significance for forward muons with (a) $\Delta\phi(\mu, \not{E}_T) \leq 2.5$ and (b) $\Delta\phi(\mu, \not{E}_T) > 2.5$.
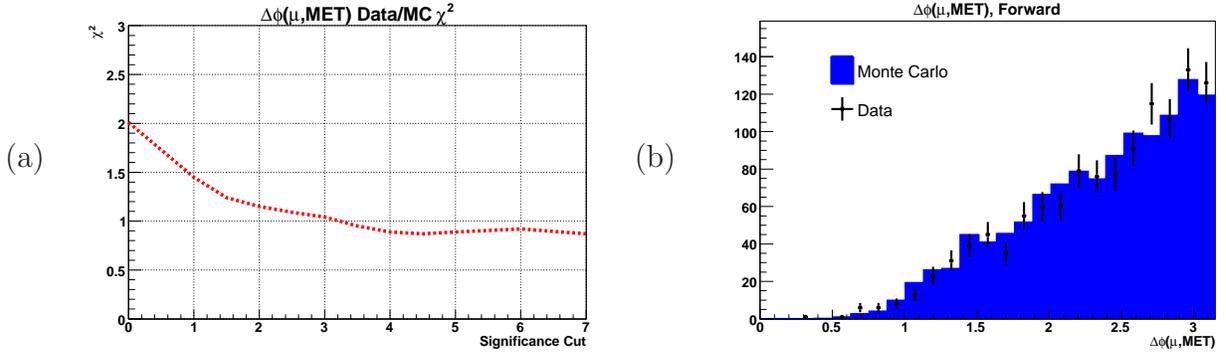
Figure 6.5: (a) The χ^2 between data and MC for the $\Delta\phi(\mu, \not{E}_T)$ distribution versus a cut on Σ_{q/p_T}. (b) The resulting $\Delta\phi(\mu, \not{E}_T)$ distribution after a cut of $\Sigma_{q/p_T} > 5$.

To check for a possible efficiency difference between data and MC for the $\Sigma_{q/p_T} > 5$ requirement on forward muons, a study was performed using the tag-and-probe method [47] in $Z \to \mu\mu$ events. The tag muon had stringent requirements on the central track (high curvature significance, hit requirements, trigger requirements) to ensure a well measured p_T. The probe muon was required to be back-to-back ($\Delta\phi > 2.8$) with the tag muon so that two muons would have equal and opposite $\vec{p}_T$. The efficiency of the $\Sigma_{q/p_T} > 5$ cut is then measured on the probe muons (with the same selection requirements as muons in this analysis) in the $Z \to \mu\mu$ data and MC samples as a function of p_T. The resulting efficiencies and the data/MC ratio are shown in figure 6.6. The data/MC efficiency ratio, which is approximately flat, is fit to find the value of 0.98. Events in this analysis with forward muons are then scaled by this factor.
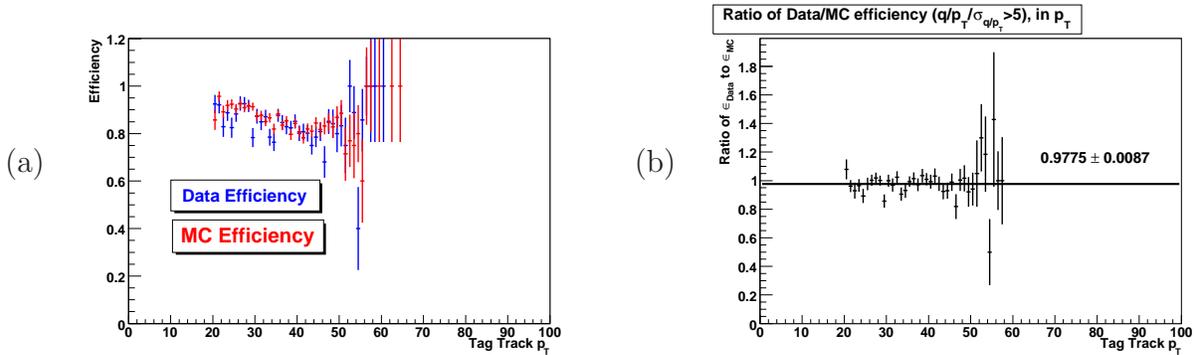
Figure 6.6: The (a) data and MC efficiencies and (b) data/MC efficiency ratio for the track curvature significance cut ($\Sigma_{q/p_T} > 5$) as a function of p_T.

6.5   Multiple Parton Interaction Removal

ALPGEN uses PYTHIA to simulate additional jets that come from multiple parton interactions (MPI) in the proton-antiproton collision. These MPI jets are generated back-to-back in ϕ with equal p_T and are uncorrelated with the hard scatter process. A bug in the production of the MC used in this analysis resulted in PYTHIA adding too many MPI jets to the ALPGEN MC samples. This overabundance of MPI jets has the largest effect on the $W+$0lp (and

Z+0lp) sample because the hard scatter process has zero jets, but an event may still pass the jet selection if it has high-p_T MPI jets. The result of too many MPI jets is clearly shown in the distribution of $\Delta\phi$ between the two leading jets shown in figure 6.7. This bug was fixed in newer versions of ALPGEN and this spike in the $\Delta\phi$ distribution from back-to-back MPI jets is not present [52]. A best χ^2 fit to data suggests that the number of events with a selected MPI jet should be reduced by approximately 80% [53]. Therefore, the MPI jet content is corrected in the ALPGEN samples used in this analysis by removing the MPI jets in 80% of events. The effect of MPI removal is small on the samples that have hard jets from the primary interaction. However, one can see from figure 6.7 that the effect on W+0lp is substantial and results in much better agreement with data.
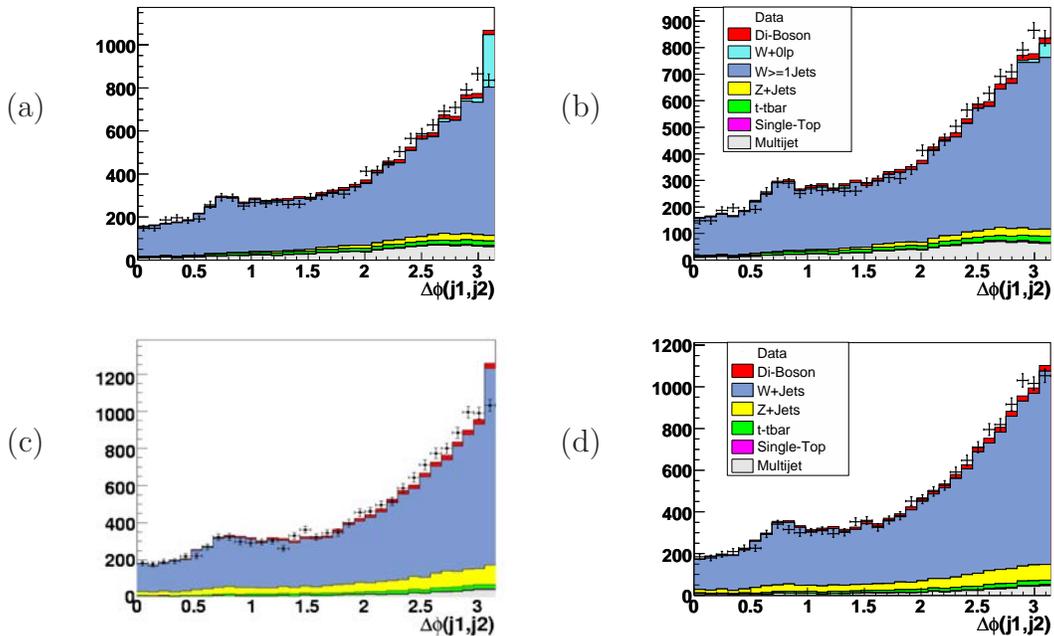
Figure 6.7: $\Delta\phi$ between the two leading jets (a,c) before and (b,d) after removing PYTHIA added MPI jets from 80% of ALPGEN events for the $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channel. It is clear that the contribution from W+0lp (shown separately in light blue for the $e\nu q\bar{q}$ channel) is far too large before MPI jet removal.

6.6 Luminosity and Vertex z Re-Weighting

As stated in section 4.2, the MC samples are overlayed with zero bias events with the intent of making the simulated events more like the detector data. However, the zero bias events were not collected with the same luminosity spectrum as the data, which could result in small difference between the data and MC. Therefore, each of the MC samples has been re-weighted based on the instantaneous luminosity of the overlayed zero bias events in order to have the same luminosity profile as the dataset.

Also, the z position of the primary vertex in MC events is Gaussian distributed, while in data the distribution is slightly non-Gaussian. Thus, the MC samples are re-weighted to make the z position of the primary vertex match the actual vertex distribution measured in the dataset [54].

6.7 Alpgen V+jets Modeling

By far, the most dominant background for this analysis is from W+jets events and the (distant) second most dominant background is from Z+jets. Therefore, it is crucial to model the background from V+jets ($V = W$ or Z) events as accurately as possible. The following sections describe the additional corrections applied to the ALPGEN V+jets MC samples to achieve agreement with data.

6.7.1 Jet η Re-Weighting

From studies performed by the ALPGEN authors [55], it is apparent that ALPGEN may not correctly reproduce the η distribution for jets in V+jets events. When comparing MC to data (figures 6.8 and 6.9) it is clear that the jet η distributions for the data are not described well by the MC. This disagreement could potentially bias the results, therefore, the V+jets

ALPGEN events are re-weighted to better reflect the distributions[1] found in the data.

Every event has two jets, therefore a re-weighting factor was calculated separately for the leading jet ("jet1") and the second-leading jet ("jet2"). For each jet, four histograms of the jet's η were made; one histogram each for the data, multijet estimation, V+jets MC, and all non-V+jets MC. The multijet and non-V+jets MC distributions were then subtracted from the data to generate distributions corresponding to V+jets events in the data. The distributions for V+jets in the data and V+jets MC were both normalized to unity (to remove any difference in overall normalization) and divided. The resulting η-dependent ratio was fit to the function $w(\eta) = p_0 + p_1 \times \eta^2$ to determine the re-weighting required for each of the jets. The jet η region corresponding to the calorimeter ICR was removed from the fits because this region may contain a higher rate of fake or noise jets that are not properly described by MC. The resulting re-weighting parameters are listed in table 6.1 and demonstrated in figure 6.10. The parameters have very good agreement between the electron and $\mu\nu q\bar{q}$ channels, therefore, the fits from each channel were combined to reduce the statistical uncertainty on the parameters. The combined parameters were then used for re-weighting the V+jets MC samples. The η distributions for the $e\nu q\bar{q}$ channel before and after re-weighting are shown in figure 6.8 and the corresponding plots for the $\mu\nu q\bar{q}$ channel are shown in figure 6.9. The χ^2 between data and MC for the jet η distributions before and after the re-weighting procedure are also given in table 6.1. More distributions before and after jet η re-weighting can be found in appendix C.

It was verified that there is very little correlation between the re-weighting of each jet by re-deriving the fit parameters for w_{jet2} after applying the leading jet re-weighting w_{jet1}, and vice-versa. There may be a small bias due to the presence of signal in the non-V+jets MC that is subtracted from the data. To estimate this bias, the derivation of w_{jet1} and w_{jet2} is repeated assuming a signal cross section of zero and also two times the nominal cross section. Combining (in quadrature) the uncertainty of the fit and the uncertainty arising from signal

---

[1]Lepton η is also not described very well; however, it was checked that re-weighting lepton η has no effect ($\ll 1\%$) on distributions used for the measurement.
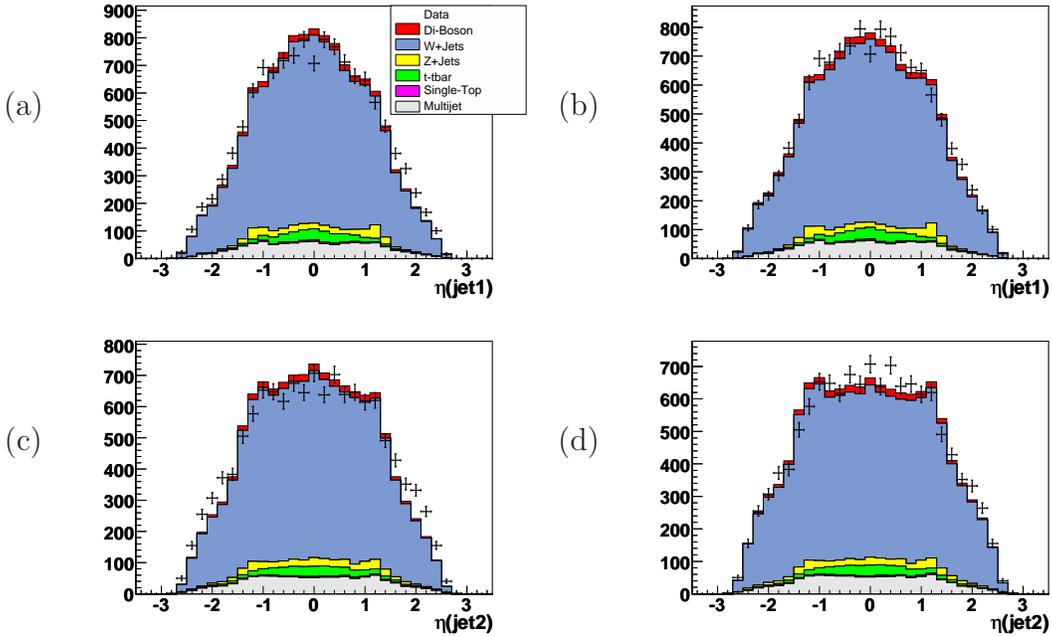
Figure 6.8: Comparison of (a,b) the leading jet η and (c,d) the second jet η distributions in the $e\nu q\bar{q}$ channel before and after performing jet η re-weighting of the V+jets ALPGEN samples.
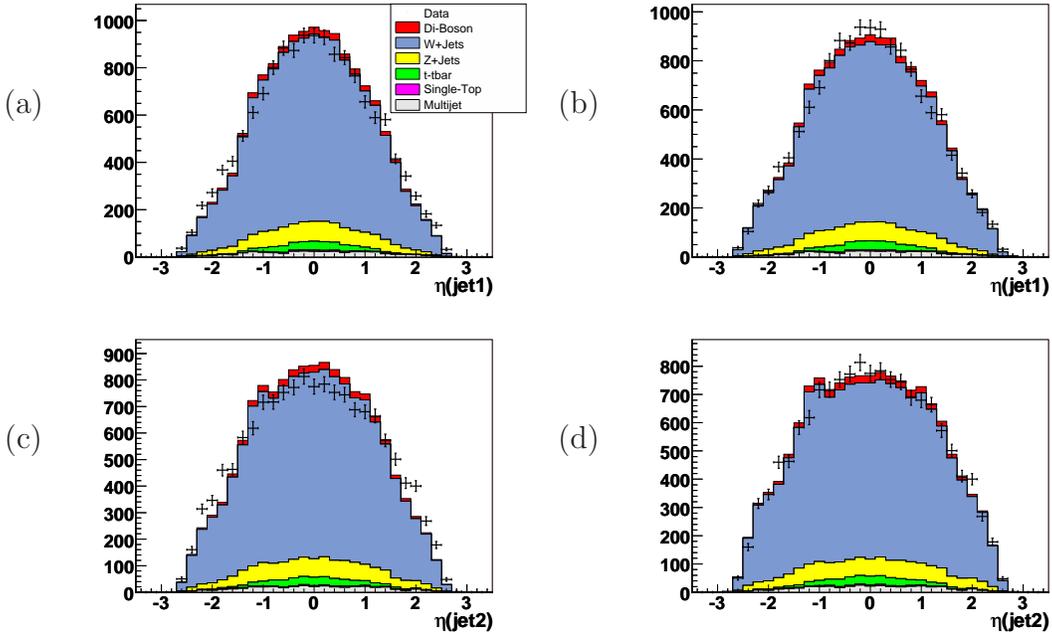
Figure 6.9: Comparison of (a,b) the leading jet η and (c,d) the second jet η distributions in the $\mu\nu q\bar{q}$ channel before and after performing jet η re-weighting of the V+jets ALPGEN samples.

(a) (b) (c)

(d) (e) (f)

Figure 6.10: The fits of the jet η re-weighting function ($w(\eta) = p_0 + p_1 \times \eta^2$, red line) to the ratio of V+jets in data and V+jets MC (black points) for (a,b) the leading jet and (d,e) the second jet in the $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channels. Also, plots of the combined re-weighting function for (c) the leading jet and (f) the second jet showing the full uncertainty on each function as well as the uncertainty due only to varying the signal between zero and two times the nominal cross section.

Table 6.1: Nominal values and uncertainties for the parameters from the individual $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channel fits and the combined values used for ALPGEN V+jets jet η re-weighting.

	p_0	p_1	σ_0	σ_1
Leading jet η, $e\nu q\bar{q}$	0.9217	0.0658	± 0.0158	∓ 0.0098
Second jet η, $e\nu q\bar{q}$	0.8804	0.0870	± 0.0167	∓ 0.0096
Leading jet η, $\mu\nu q\bar{q}$	0.9251	0.0635	± 0.0138	∓ 0.0087
Second jet η, $\mu\nu q\bar{q}$	0.8818	0.0862	± 0.0149	∓ 0.0087
Leading jet η, combined	0.9236	0.0645	± 0.0105	∓ 0.0066
Second jet η, combined	0.8812	0.0865	± 0.0116	∓ 0.0069

Table 6.2: The χ^2/ndf between data and MC distributions of jet η before and after re-weighting.

	Pre-Weight χ^2/ndf	Post-Weight χ^2/ndf
Leading jet η, $e\nu q\bar{q}$	3.20	1.30
Second jet η, $e\nu q\bar{q}$	5.01	1.46
Leading jet η, $\mu\nu q\bar{q}$	3.77	1.83
Second jet η, $\mu\nu q\bar{q}$	6.01	1.59

contamination gives the total systematic uncertainty assigned to each of the fit parameters (table 6.1).

6.7.2 Jet ΔR Re-Weighting

In addition to the η of the individual jets, ΔR between the jets is not described well by ALPGEN (figure 6.11 and table 6.4). Jet η re-weighting supplies a small improvement to the ΔR distribution, but falls short of correcting the discrepancy between data and MC. To correct for the residual difference, a ΔR re-weighting is performed in the same manner as jet η re-weighting. Namely, the ratio of the ΔR distribution for V+jets in data and V+jets MC is fit to a function. In this case the ratio is described well by a linear function, $w(\Delta R) = p_0 + p_1 \times \Delta R$. An additional benefit of using a linear re-weighting (versus a quadratic for example) is that the linear re-weighting is influenced very little by presence or absence of signal, which has a very nonlinear dependence. The resulting fits are shown in figure 6.12 with the fit parameters listed in table 6.3. More distributions before and after ΔR re-weighting can be found in appendix C.

We determine and apply the systematic uncertainty in the same way as for jet η re-weighting, combining the statistical uncertainty from the fit with the uncertainty from varying the signal contribution between zero to two times the nominal rate. The resulting systematic uncertainty on the parameters p_0 and p_1 can be found in table 6.3.

Table 6.3: Nominal values and systematic uncertainties for the parameters from the individual $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channel fits and the combined values used in ALPGEN V+jets ΔR re-weighting.

	p_0	p_1	σ_0	σ_1
ΔR, $e\nu q\bar{q}$	0.8052	0.0761	± 0.0379	∓ 0.0149
ΔR, $\mu\nu q\bar{q}$	0.7925	0.0808	± 0.0340	∓ 0.0133
ΔR, combined	0.7980	0.0788	± 0.0279	∓ 0.0109
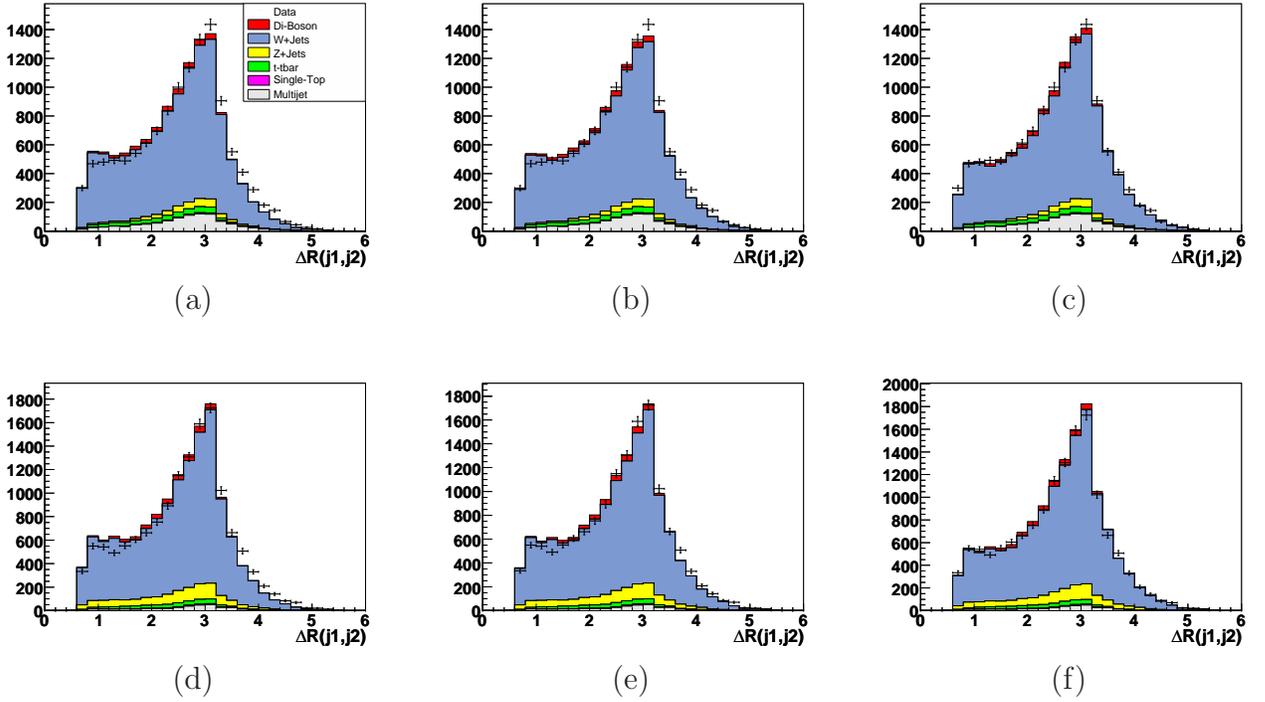
Figure 6.11: Comparison of the ΔR in (a,b,c) the $e\nu q\bar{q}$ channel and (d,e,f) the $\mu\nu q\bar{q}$ channel before re-weighting, after jet η re-weighting, and then after jet η and ΔR re-weighting.
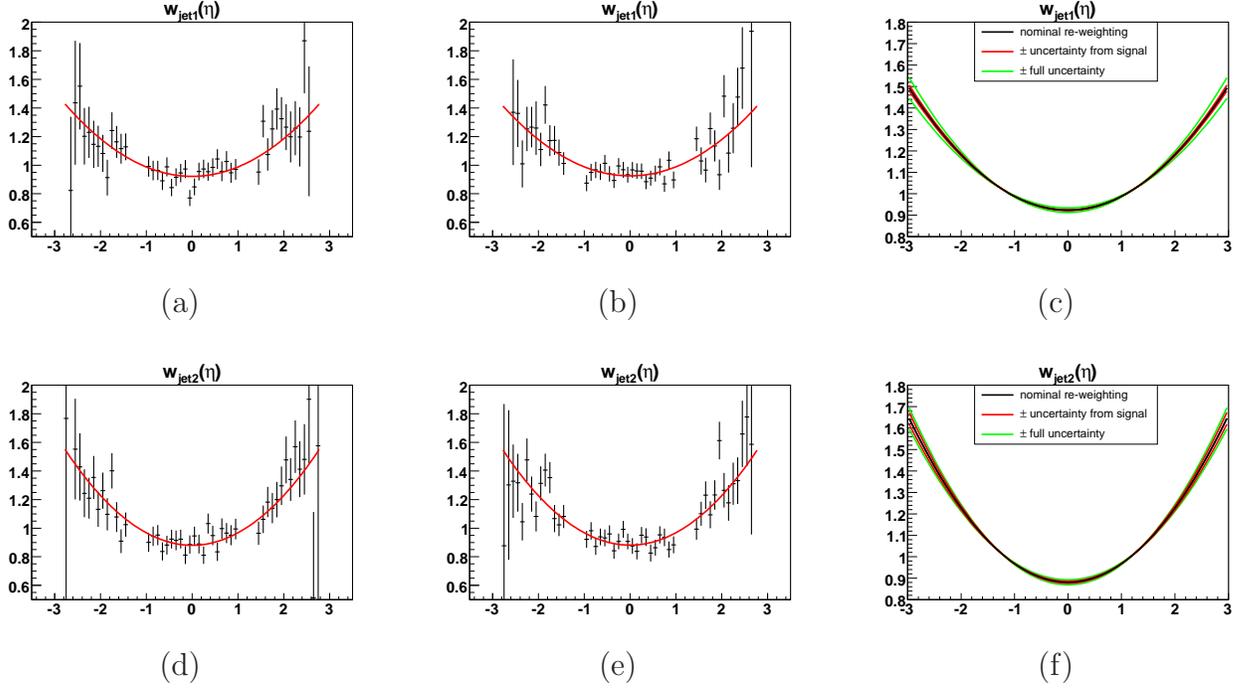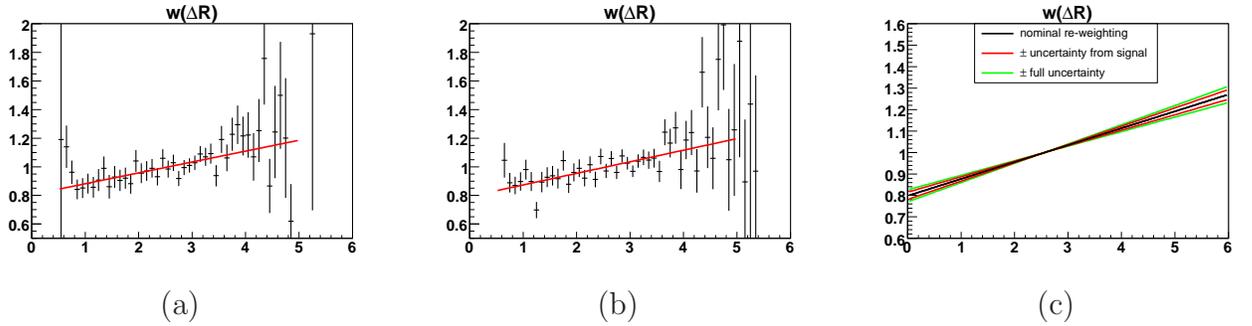

Figure 6.12: The fits of the ΔR re-weighting function ($w(\Delta R) = p_0 + p_1 \times \Delta R$, red line) to the ratio of V+jets in data and V+jets MC (black points) for (a) the $e\nu q\bar{q}$ channel and (b) the $\mu\nu q\bar{q}$ channel. (c) A plot of the combined ΔR re-weighting function showing the full uncertainty as well as the uncertainty due only to varying the signal between zero and two time the nominal cross section.

Table 6.4: The χ^2/ndf between data and MC distributions of ΔR before and after jet η and ΔR re-weightings.

	No Re-Weighting χ^2/ndf	Only Jet η Re-Weighting χ^2/ndf	η and ΔR Re-Weighting χ^2/ndf
ΔR, $e\nu q\bar{q}$	4.91	2.31	0.84
ΔR, $\mu\nu q\bar{q}$	5.95	2.78	1.10

6.7.3 Alpgen Matching Parameters and Renormalization Scale

A study of the adjustable parameters within the ALPGEN generator found that the settings used for generating the V+jets samples were suboptimal. Complete details of this study are provided in appendix D. It was found that ALPGEN provides the best description of data when the parton-jet matching p_T threshold is set to 13.2 GeV; however, the the V+jets samples in this analysis were generated with a parton-jet matching p_T threshold of 8 GeV. The V+jets MC samples used in this analysis were, therefore, corrected in order to achieve a better description of the data. The correction was implemented as a re-weighting using the ratio of the dijet mass distributions for events generated with a parton-jet matching p_T threshold of 13.2 GeV versus 8 GeV. In addition to the correction, appendix D describes the systematic uncertainty assigned to the ALPGEN V+jets MC due to the parton-jet matching parameters and the renormalization/factorization scale.

6.8 Diboson NLO Correction

The diboson signal events are generated using a LO generator, but potential changes to the event kinematics at NLO should be considered. Diboson production at NLO is available in the MC@NLO event generator [56]. Unfortunately this package was not integrated into the full D0 simulation chain at the time of this analysis. Thus, the diboson events simulated with PYTHIA are re-weighted at generator-level to match the kinematics of the W and Z bosons in the diboson samples produced by MC@NLO. Version 3.3 of MC@NLO was used to generate WW and WZ events at NLO using the CTEQ6M PDF set. The events are generated with full final-state radiation and hadronization, but it is the generator-level information for the individual weak bosons prior to decay that is used for re-weighting.

The goal is to re-weight the PYTHIA samples so that the kinematic observables sensitive to the production model will match the NLO predictions. The observables that were analyzed include the p_T, energy, η, and rapidity of the individual weak bosons, as well as the vector

summed p_T, scalar summed p_T, summed energy, η, and rapidity of the diboson system, the separation between the bosons ($\Delta R(V1, V2)$), and the difference in p_T of the two bosons ($\Delta p_T(V1, V2)$). A comparison of diboson p_T, boson p_T, and $\Delta R(V1, V2)$ distributions for PYTHIA and MC@NLO before re-weighting is shown in figure 6.13 and more distributions can be found in appendix E. It is observed empirically that there was no single kinematic variable that could be used to correct the entire event, achieving satisfactory agreement for individual weak bosons and the diboson system. However, agreement is achieved after re-weighting by both the p_T of the diboson system and the p_T of the leading boson. As these two variables contain a degree of correlation, the re-weight is performed with a two-dimensional parameterization instead of two successive one-dimensional re-weightings. A binned re-weighting matrix (figure 6.14) was created by taking the ratio the two-dimensional histograms for MC@NLO and PYTHIA events. Distributions of diboson p_T, boson p_T, and $\Delta R(V1, V2)$ after re-weighting the PYTHIA events are shown in figure 6.15 and more distributions before and after re-weighting can be found in appendix E.

This PYTHIA to MC@NLO re-weighting procedure was observed to generate an increase in overall acceptance of 3.2%. To model the uncertainty on this re-weighting procedure, 50% of the difference between PYTHIA and MC@NLO is incorporated as a symmetric, shape-dependent systematic uncertainty on the signal model.
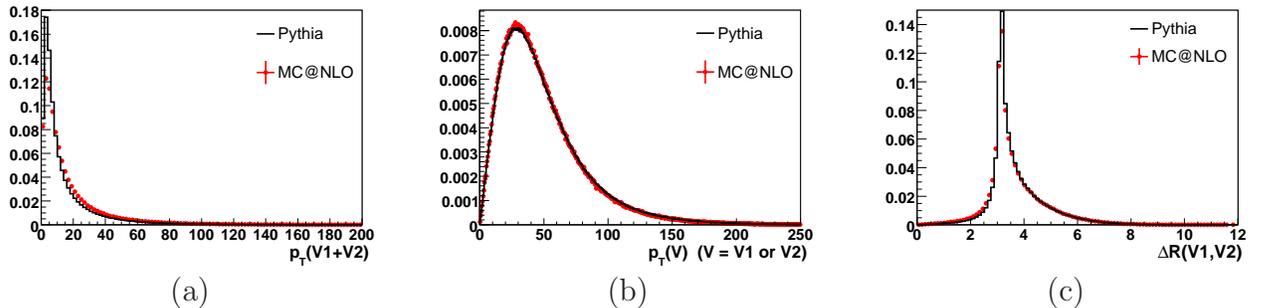
Figure 6.13: Distributions of (a) diboson p_T, (b) boson p_T, and (c) ΔR between bosons comparing PYTHIA and MC@NLO generators (before re-weighting PYTHIA events).
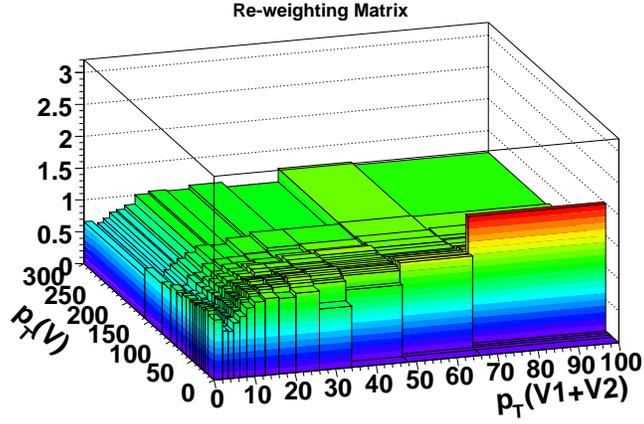
Figure 6.14: The re-weighting matrix used to re-weight PYTHIA to MC@NLO.
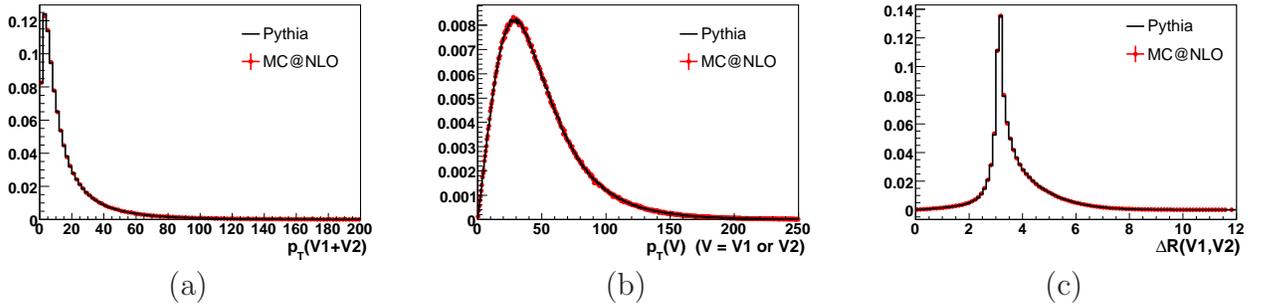
Figure 6.15: Distributions of (a) diboson p_T, (b) boson p_T, and (c) ΔR between bosons comparing PYTHIA and MC@NLO generators after re-weighting PYTHIA events.

6.9   Corrected Distributions

The resulting distributions comparing the data to the prediction from multijet estimates and corrected MC samples and are shown in figures 6.16 and 6.17.
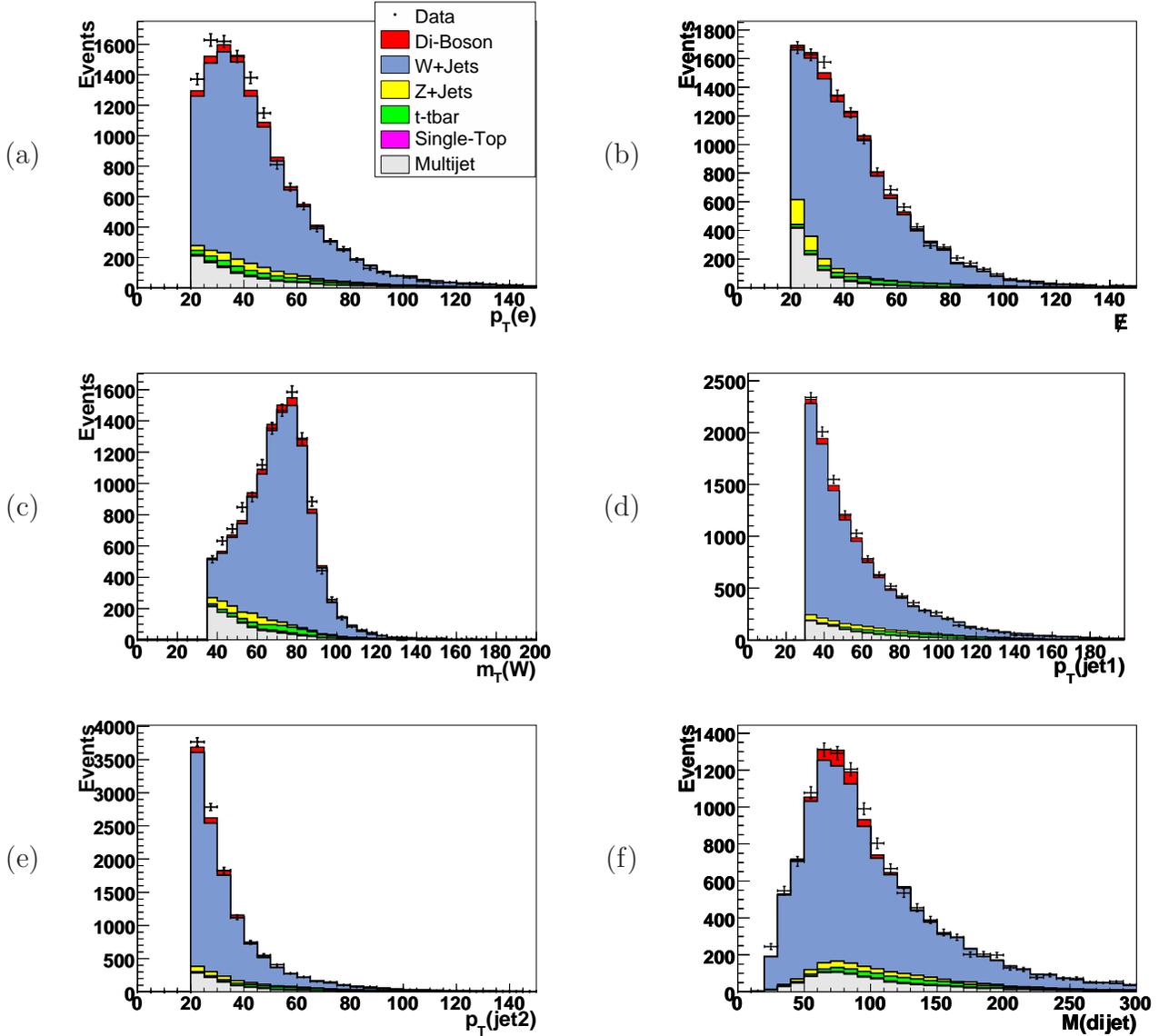
Figure 6.16: Comparison of MC to data in the $e\nu q\bar{q}$ channel for: (a) p_T of the electron, (b) $\not{E}_T$, (c) the transverse mass of the W reconstructed from the electron and $\vec{\not{E}}_T$, (d) p_T of the leading jet, (e) p_T of the second jet, and (f) the mass of the dijet system.
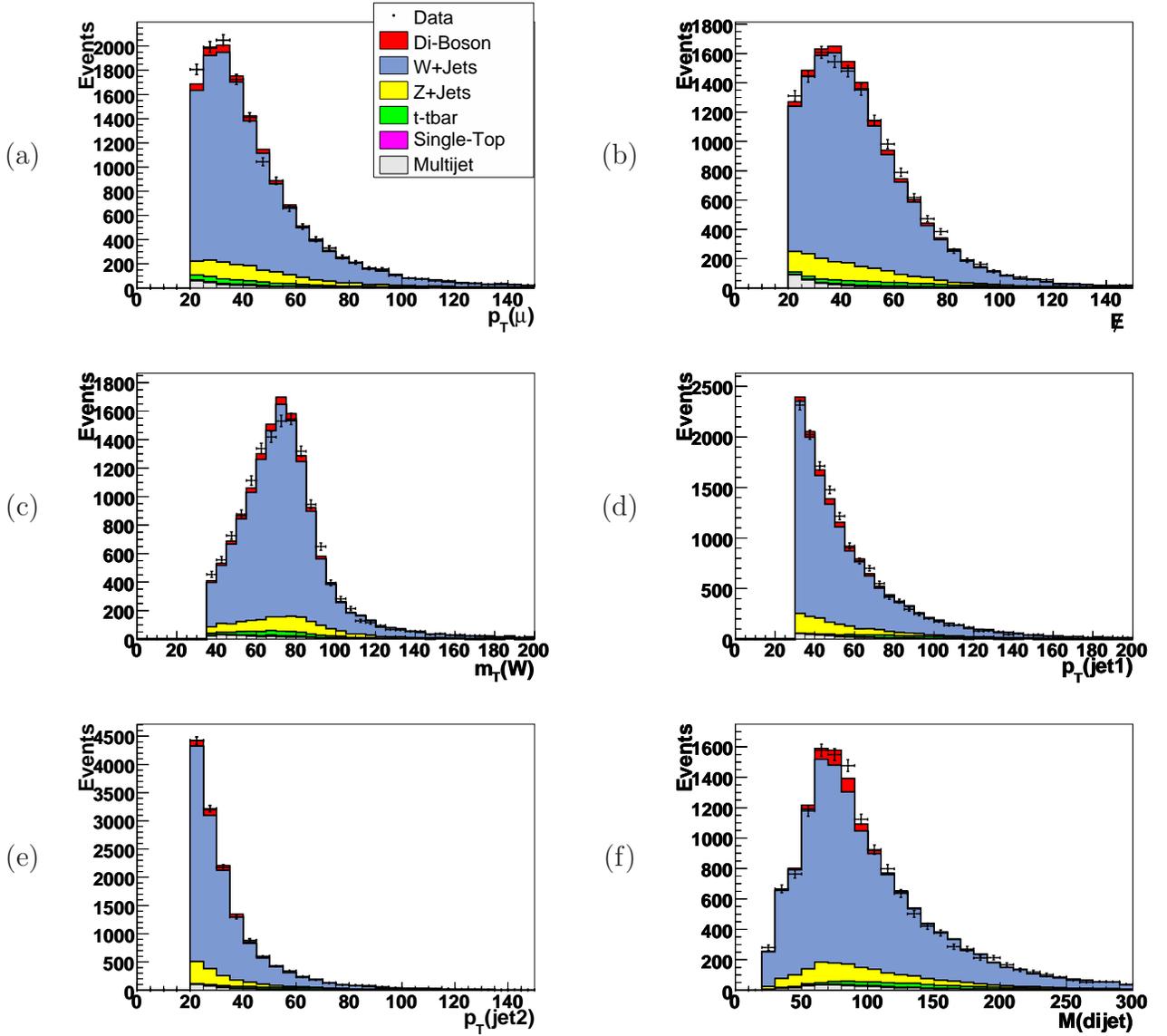
Figure 6.17: Comparison of MC to data in the $\mu\nu q\bar{q}$ channel for: (a) p_T of the electron, (b) $\not{E}_T$, (c) the transverse mass of the W reconstructed from the electron and $\vec{\not{E}}_T$, (d) p_T of the leading jet, (e) p_T of the second jet, and (f) the mass of the dijet system.

Chapter 7

Multivariate Discrimination

In order to increase the sensitivity of an analysis, it is advantageous to combine the information from several variables using a multivariate classifier. There are many multivariate classifiers to choose from, each with varying degrees of robustness and predictive power. This analysis makes use of the random forest classifier from the StatPatternRecognition package [57]. The use of boosted decision trees or artificial neural networks was also investigated, but it was found that, for this application, the random forest classifier was the most powerful. Additionally, random forest classifiers have the advantage of being very robust against over-training, variable correlations, and random noise in the training samples [58].

This chapter provides a description of the random forest classifier implemented in this analysis. Section 7.1 gives a description of the algorithm used for constructing a random forest and section 7.2 describes the procedures and tools used to create and optimize the specific instances of random forests used in the analysis.

7.1   Random Forest Algorithm

In recent years there have been many studies demonstrating the use of ensembles of classifiers to improve classification accuracy [57, 58, 59, 60, 61, 62]. The results of such studies form the basis for random forests introduced by Breiman [59]. In general, a random forest is any

classifier made up of multiple tree-structured classifiers that are each trained (or "grown") using independent, identically distributed training sets. For discrete classification, each tree casts a unit vote (*e.g.*, zero for background or one for signal) and the output of the random forest is the most popular vote (the mode). For numerical/continuous output the random forest takes the mean of the outputs from the trees.

It should be noted that the definition of a random forest is very general, but this discussion is concerned only with the random forest algorithm used for this analysis. Namely, the random forest implemented in version 07-02-00 of StatPatternRecognition [57]. Therefore, unless otherwise stated, the term random forest refers to this implementation.

There are three main components in the construction of the random forest classifier:

1. decision trees - the individual tree classifiers;

2. random subspace sampling - the method for creating multiple trees;

3. bagging - the method for generating the independent, identically distributed, training sets used to train each tree.

Explained below are the details of each of these components; followed by a discussion of combining these components to make the resulting random forest classifier.

7.1.1   Decision Trees

The basic building blocks of a random forest are decision tree classifiers (also know as classi-fication trees). A very nice description of decision trees can be found in the recent single-top analysis [63]. Given a set of N training events $\mathcal{L} = \{(y, \vec{x})_i, i = 1, ..., N\}$, where the $\vec{x}$ are vectors of input variables and the y designate the class of each event ($y = 1$ for signal, $y = 0$ for background), a decision tree is trained ("grown") by recursively dividing the feature space into rectangular regions as follows. The root node containing the full feature space is divided into two children nodes by considering all possible splits in all input dimensions and choosing the split that results in the best separation between signal and background.

Splitting continues on the children nodes until either the separation of signal and background cannot be improved (*e.g.*, the node already contains only background events) or a split would produce children nodes with less then a predetermined minimum number of events [57]. This is shown schematically in figure 7.1.
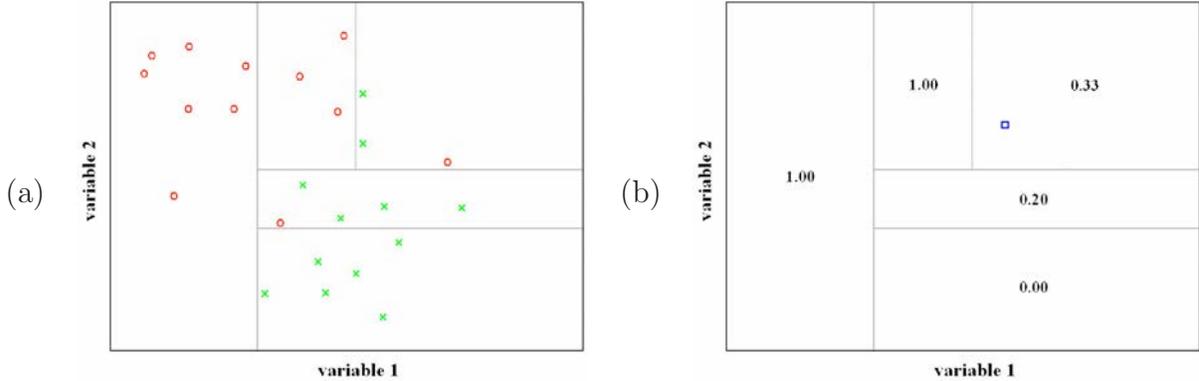
Figure 7.1: Schematic example of a decision tree with two input variables. (a) A decision tree trained using signal (red circles) and background (green crosses) events divides the feature space in to nodes. In this example there must be at least three events per terminal node. (b) The output value assigned to each terminal node. The test event (blue square) has an output of 0.33.

The minimal number of training events per terminal node is used to prevent over-training. With no minimum, splitting would continue until every node contained either only signal or only background events. This would result in 100% accurate classification of the training set, but would not generalize well to unseen events, resulting in a poor classifier. The optimal value for this parameter can vary greatly depending on the specific application and must be tuned to find the optimal balance between over-training and under-training. Determination of this parameter and also the choice for which criteria to use when splitting a node is discussed more in Section 7.2. StatPatternRecognition supplies the user with many options and the user must decide which criteria works best for the particular analysis.

Once a tree is fully grown, each terminal node has a signal purity $p_s = \frac{w_s}{w_s + w_b}$ determined by the weighted number of signal (w_s) and background (w_b) training events in that node. Using this information, the tree can be used to estimate the likelihood that a new/unseen event $\vec{x'}$ (of unknown class y') is a signal event. The new event is filtered through the tree

until it reaches a terminal node and the output value for that event is given by the signal purity of the terminal node. Even though the signal purity can be any number $p_s \in [0, 1]$, strictly speaking the output is discrete because there are a finite number of terminal nodes. However, for all practical purposes the distribution is approximately continuous for trees used in this analysis, which each have around one thousand terminal nodes.

Decision trees have many attractive qualities. Advantages over artificial neural networks include much faster training, the ability to use both discrete and continuous inputs, the ability to handle a large number of input variables, and an intrinsic simplicity that makes decision trees easy to interpret. On the other hand, for many applications a single decision tree will likely be less powerful than more complex classifiers. This is where the methods of random subspace sampling, bagging, and boosting come into play. In each case, classification accuracy is improved by creating multiple decision trees and averaging the outputs. The drawbacks are a reduction in interpretability and somewhat slower training that comes with increased complexity, but these aggregate classifiers retain the other benefits of decision trees and come with benefits of their own. The random forest used in this analysis uses both random subspace sampling and bagging as described in the next two sections.

7.1.2 Random Subspace Sampling

The random subspace method was first proposed by Ho in 1995 [60, 61]. In this method, multiple decision trees are created by training each tree in a randomly selected subspace of the full feature space. Starting with the training set from before, $\mathcal{L} = \{(y, \vec{x})_i, i = 1, ..., N\}$, the input vectors, $\vec{x}$, in the full n-dimensional feature space are projected into a randomly selected m-dimensional subspace $(m \leq n)$ to create the vectors $\vec{x}^{(S_1)}$. A decision tree is then created using the training set $\mathcal{L}^{(S_1)} = \{(y, \vec{x}^{(S_1)})_i, i = 1, ..., N\}$. This is repeated K times and K decision trees are created from the training sets $\{\mathcal{L}^{(S_k)}; k = 1, ..., K\}$. In other works, each of the K decision trees uses only m input variables selected randomly from the n possible input variables. The output for the full forest is then the mean of the outputs from each

tree.

As Ho explains [60], this method can improve the classification accuracy of unseen events because each tree will divide the feature space differently and therefore generalize its classification in a different way. This is shown schematically for a simple two dimensional case in figure 7.2. Note that this example is a gross over simplification because there are only two possible subspaces (besides the full space) to "randomly" select for training new decision trees. This method is most useful when the number of input variables, n, is large and there are many subspaces ($2^n - 1$) to randomly choose from.
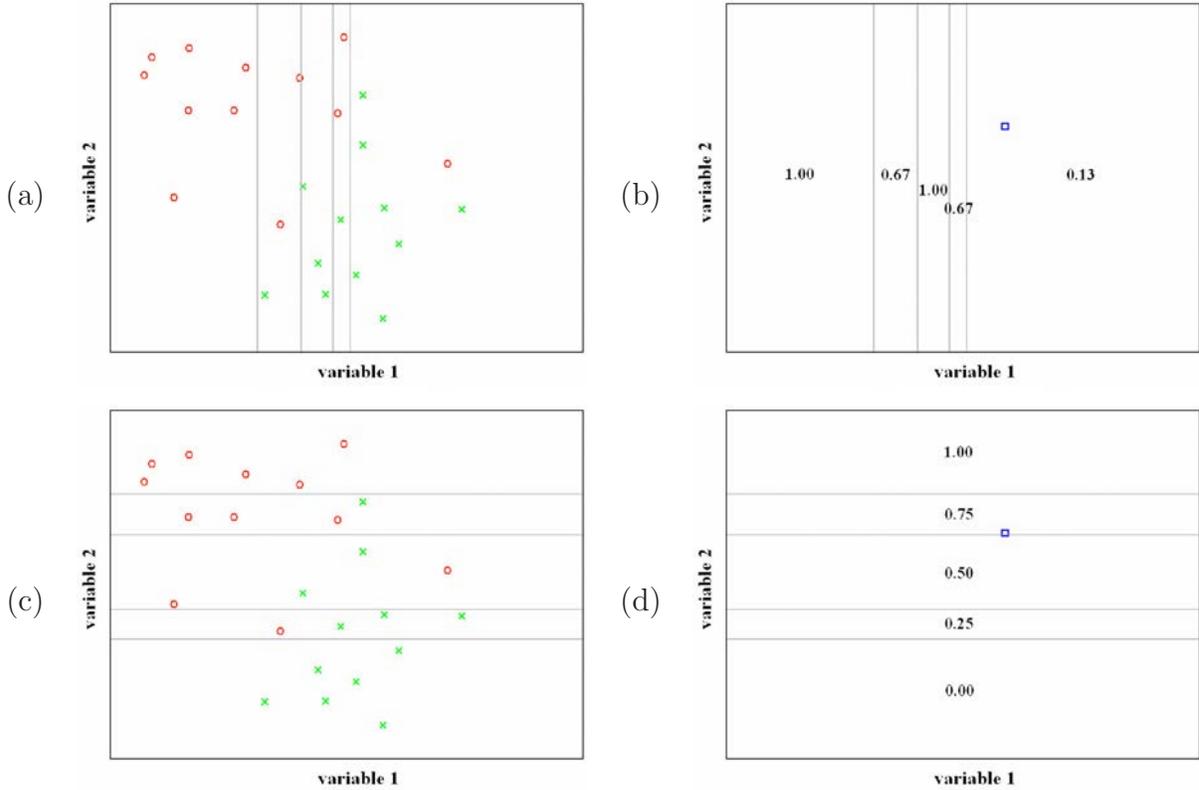
Figure 7.2: Schematic example of how the random subspace method works. (a) A decision tree trained using only "variable 1" divides only that subspace. (c) A decision tree trained using only "variable 2" divides only that subspace. The test event (blue box) has an output of 0.44 calculated by averaging the output for (b) the first tree (0.13) and (d) the second tree (0.75).

Using the random subspace method, Ho shows that dramatic improvement in classification accuracy is achieved compared to single tree classification. It is also shown that

accuracy comparable to, and in some cases better than, the methods of boosting and bagging are achieved [61, 64].

7.1.3 Bagging

Bagging (short for bootstrap aggregating) is a procedure introduced by Breiman [62]. Like the random subspace method, bagging is designed to reduce over-training and increase classification accuracy by averaging the outputs of many decision trees. In this method, multiple copies of a classifier are created each using the full feature space, but trained using different bootstrapped replicas of the training set. A bootstrap replica, $\mathcal{L}^{(B_1)}$, is created by sampling N events *with replacement* from the original training set $\mathcal{L} = \{(y, \vec{x})_i; i = 1, ..., N\}$. By sampling with replacement, each event from $\mathcal{L}$ may appear multiple times or not at all in $\mathcal{L}^{(B_1)}$ (on average the bootstrap replica will contain $(1 - 1/e) \approx 0.632$ of the original events). Using this procedure, a set of K bootstrap replica training sets $\{\mathcal{L}^{(B_k)}; k = 1, ..., K\}$ are created, each approximating the distribution underlying the original training set $\mathcal{L}$.

Studies of bagging have shown that this procedure can lead to substantial improvements in classification accuracy for systems that are unstable to small changes in the training set (*e.g.*, due to random noise). In Breiman's original study using bagging with decision trees, he found reductions of 20% to 47% in misclassification when analyzing a variety of real and simulated data sets [62]. To help understand this, first consider systems that are *not* sensitive to small changes in the training set. These systems will not benefit from bagging because, even though each tree is trained on a slightly different training set, all the trees will have the same behavior. In this case, one can be confident that a single decision tree properly describes the features and nothing is gained by averaging multiple trees. On the contrary, a single decision tree for an unstable system will likely perform poorly on unseen data. One could improve stability and generality by simply increasing the minimum number of events per terminal node, but this comes at the expense of accuracy. Applying the method of bagging to an unstable system will result in multiple trees that are each affected slightly

differently by fluctuations in the training data. For example, some trees may be affected by certain noisy events while other trees are not. Averaging the output of the different trees washes out these fluctuations resulting in a more stable classifier.

7.1.4 Random Forest

The random forest used in this analysis combines the methods of random subspace sampling and bagging. Each tree in the forest is created by randomly selecting a subset of input variables (as with the random subspace method), then trained on a bootstrap replica of the training data (as with the bagging method). The result is many trees that are each trained with slightly different input variables and/or training events. Each tree will therefore generalize differently on unseen events and, even though individual trees may be over-trained, the net effect of averaging all the trees is an accurate and stable classifier.

The number of input variables[1] (m) to randomly select for each tree and the total number of trees to create (K) are parameters that must be set by the user. Of course, before any random forest can be made the user must choose what input variables to use. All of these details are discussed in the following section.

In summary, the random forest classifier works by instantiating and training many decision trees, each using different subsets of the full training set and different subsets of the total number of input variables. The result is a "forest" of many random decision trees. The output from the random forest is then constructed by averaging (with equal weights) the output from each decision tree. In other words, given an event $\vec{x}$, each decision tree in the forest evaluates the inputs and supplies an output $\phi_k(\vec{x}) \in [0, 1]$. The output of the random forest is the mean of all the tree outputs, $\Phi(\vec{x}) = \frac{1}{K} \sum_{k=0}^{K} \phi_k$.

---

[1]Actually, the user specifies the *maximum* number of input variables per tree because the selection of input variables for a particular tree is done with replacement (*i.e.*, the same input variable can be selected more than once). By virtue of how decision trees work, multiple inputs of the same variable are ignored. Therefore, cases where any input variable is selected more than once will result in decision trees trained on fewer then the specified maximum number of input variables.

7.2 Diboson Random Forest

For the analysis presented in this thesis, a random forest classifier was constructed for each channel ($e\nu q\bar{q}$ and $\mu\nu q\bar{q}$) to separate diboson signal events from the various background processes. The details of these random forests are discussed in the following sections.

7.2.1 Training

The training sets consist of signal and background MC events passing the selection requirements for each channel. The random forest output distributions for events in the training sets will almost certainly be biased to have somewhat better signal and background separation than unseen events. Therefore, some of the MC samples must be reserved for performing the measurement. The $W+$0lp and $W+$1lp samples have relatively low statistics (compared to their cross sections), thus these events were not used for training in order to preserve sufficient statistics for the comparison to data. Also, samples with less than 100 events, regardless of cross section, were not used for training. All other MC samples were divided in half (randomly) with one half of each sample used for training and the other half reserved for the measurement. All event weights (most importantly cross section) were used when training. Additionally, signal events are scaled up by a constant factor to make the total weighted signal integral equal to the total weighted background integral. This produces a more symmetric random forest output and improves the separation power.

7.2.2 Input Variables

Input variables were selected that demonstrate differences between the signal distribution and at least one of the background distributions. It should be noted, however, that performance of a decision tree (and by extension random forest) is not degraded by the addition of a well modeled variable even if the variable is uninformative. If an input variable is added that does not separate signal from background, then the tree will simply never make a cut on

that variable.

The χ^2 probability between the data and MC distributions is used to ensure that the input variables are well modeled by the MC events. Before calculating the χ^2, all events in the signal peak of the dijet mass (55 GeV$< m_{jj} < 100$ GeV) are removed in order to effectively eliminate an influence from the signal events. Distributions using events from the dijet mass sidebands must then have a χ^2 probability greater than 5% to be used as an input variable. Resulting input variables are:

1. $\not{E}_T =$ imbalance in transverse momentum (missing E_T);

2. $p_T(\text{jet2}) = p_T$ of second-leading jet;

3. $\Delta\phi(\ell, \vec{\not{E}}_T) =$ azimuthal separation of the charged lepton and $\vec{\not{E}}_T$;

4. $M_T(W) =$ transverse mass of W reconstructed from charged lepton and $\vec{\not{E}}_T$;

5. $p_T(W) = p_T$ of reconstructed W from charged lepton and $\not{E}_T$;

6. $M(\text{dijet}) =$ invariant mass of dijet system;

7. $p_T^{\text{Rel}}(\text{dijet}, \text{jet1})_{\text{WF}} =$ magnitude of $\vec{p}_T(\text{jet1}) \perp$ to dijet system $= \frac{|\vec{p}_T(\text{jet1}+\text{jet2}) \times \vec{p}_T(\text{jet1})|}{|\vec{p}_T(\text{jet1}+\text{jet2})|}$, calculated in the rest frame of the reconstructed leptonic W boson;

8. $p_T^{\text{Rel}}(\text{dijet}, \text{jet2}) =$ magnitude of $\vec{p}_T(\text{jet2}) \perp$ to dijet system $= \frac{|\vec{p}_T(\text{jet1}+\text{jet2}) \times \vec{p}_T(\text{jet2})|}{|\vec{p}_T(\text{jet1}+\text{jet2})|}$, calculated in the lab frame;

9. $\text{ktMin}_{\text{WF}} = \Delta R(\text{jet1}, \text{jet2}) \cdot \frac{E_T(\text{jet2})}{E_T(\ell)+\not{E}_T}$ calculated in the rest frame of the reconstructed leptonic W boson;

10. centrality $=$ sum of scalar p_T divided by sum of energy for the lepton plus jets in the event;

11. $\cos(\angle(\text{dijet}, \text{jet1})) =$ cosine of the angle between dijet system and the leading jet in the lab frame;

12. $\cos(\angle(\mathrm{dijet}, \mathrm{jet2})) = $ cosine of the angle between dijet system and the second jet in the lab frame;

13. $\cos(\angle(W, \mathrm{jet1}))_{\mathrm{DF}} = $ cosine of the angle between the reconstructed leptonic W boson and the leading jet in the rest frame of the dijet system.

Histograms with the expected SM distributions (from the MC) and the observed data distributions for each input variable are shown in figures 7.3 and 7.4 for the $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channels, respectively. Figures 7.5 and 7.6 show comparisons between the shapes of the signal and background distributions for each input variable. Plots of the input variables can also be found in appendix F showing the systematic uncertainty bands on each distribution and comparisons of the signal to the background-subtracted data after the final measurement.

7.2.3   Optimization of Random Forest Parameters

Besides the input variables, there are a few additional parameters that must be supplied by the user and optimized for the application at hand. As described in section 7.1, these parameters are:

- the minimal number of events per terminal node in each decision tree,

- the split criteria for the decision tree nodes,

- the maximum number of variables used by each tree,

- the number of trees in the forest.

The primary method used in determining these parameters was a 3-fold cross-validation test. In general, k-fold cross-validation involves dividing the training data into k equal pieces. The first piece is reserved for validation and the remaining $k - 1$ pieces for training, then the second piece is reserved for validation and the remaining $k - 1$ pieces for training another random forest, etc. until all k pieces have been used for validation. The results of each

Figure 7.3: Comparison of data to MC for the input variables used by the random forest for the $e\nu q\bar{q}$ channel.
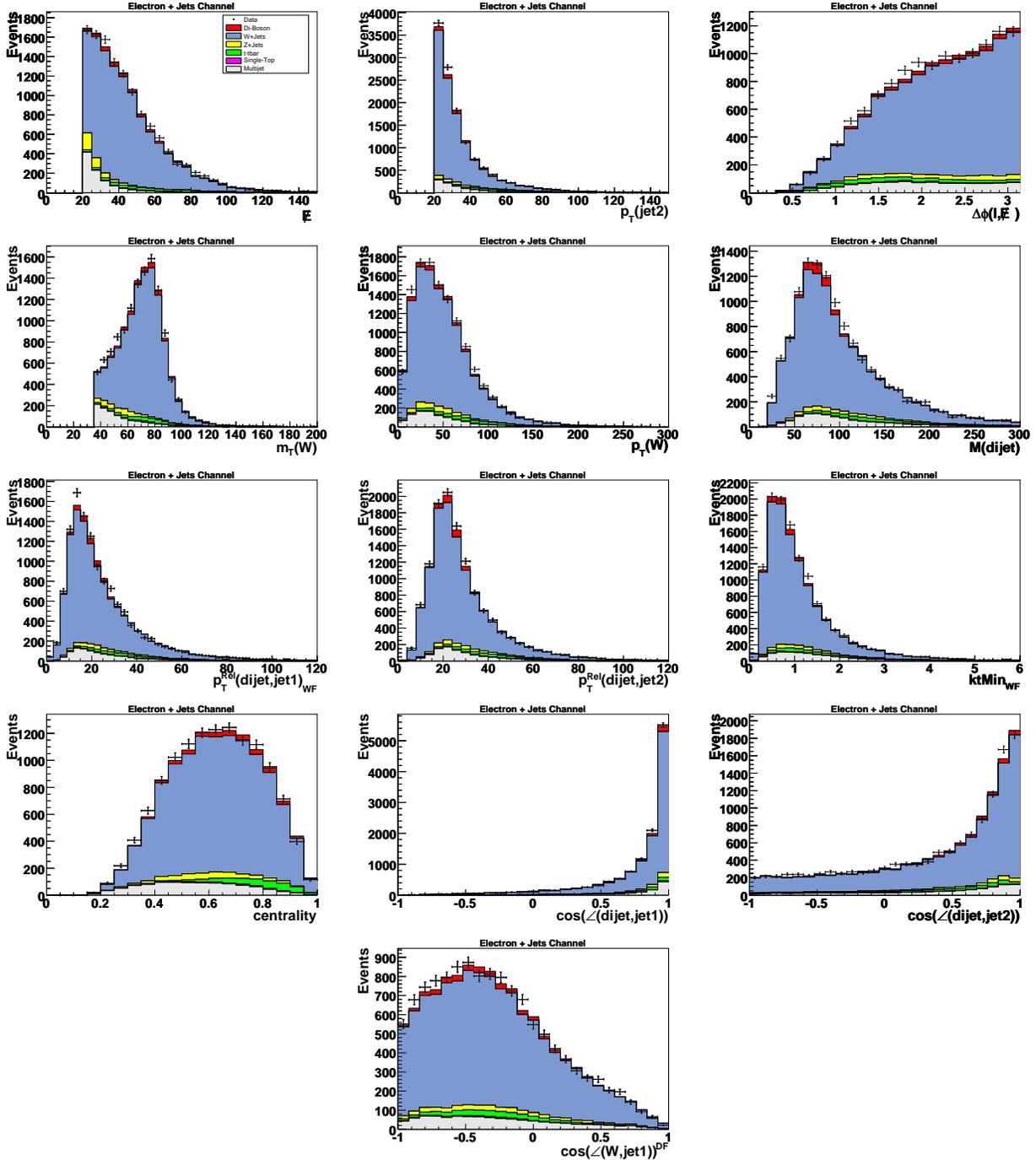
Figure 7.4: Comparison of data to MC for the input variables used by the random forest for the $\mu\nu q\bar{q}$ channel.

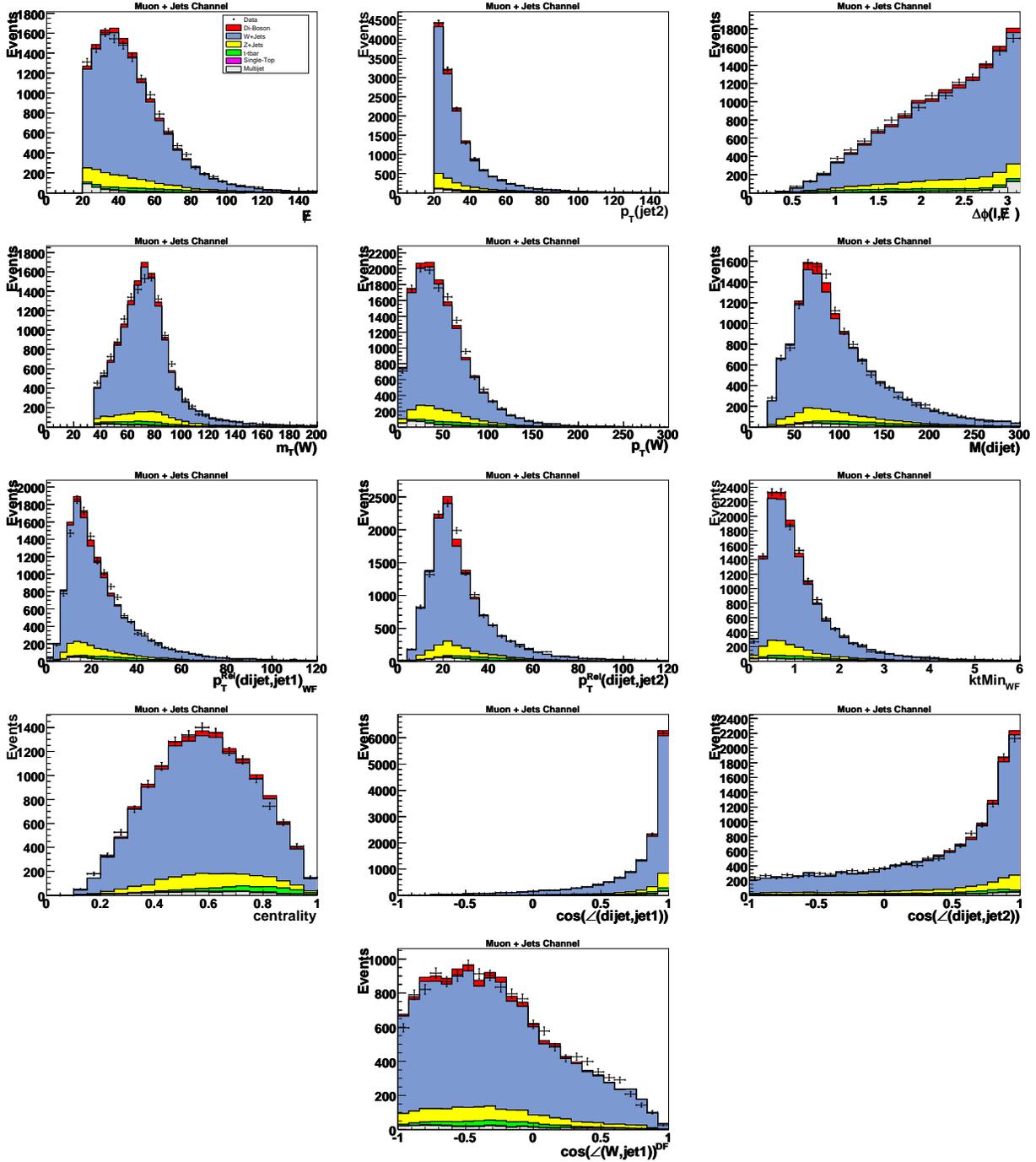Figure 7.5: Shape comparison for different physics processes (each integral normalized to unity) for the input variables used by the random forest for the $e\nu q\bar{q}$ channel.
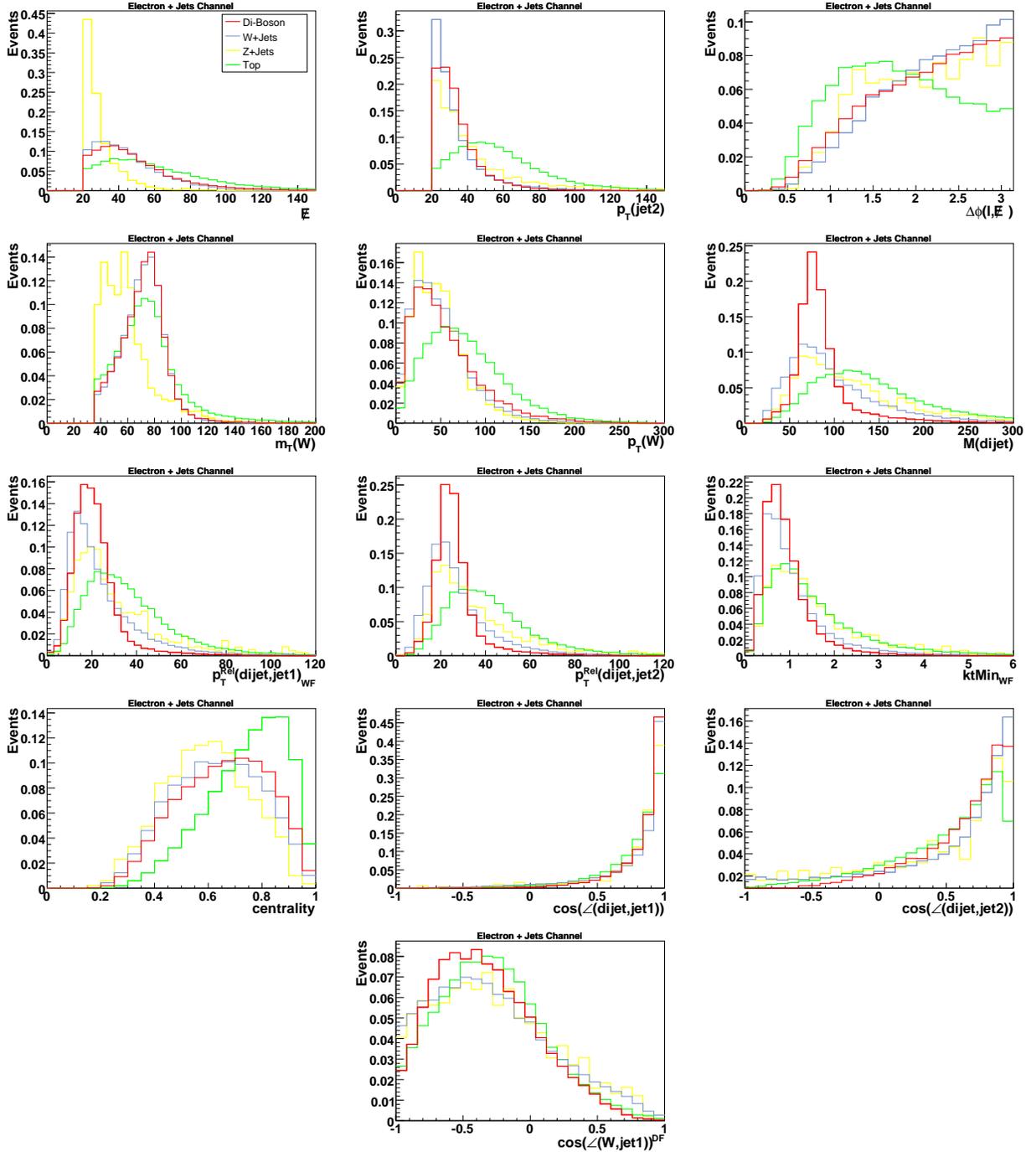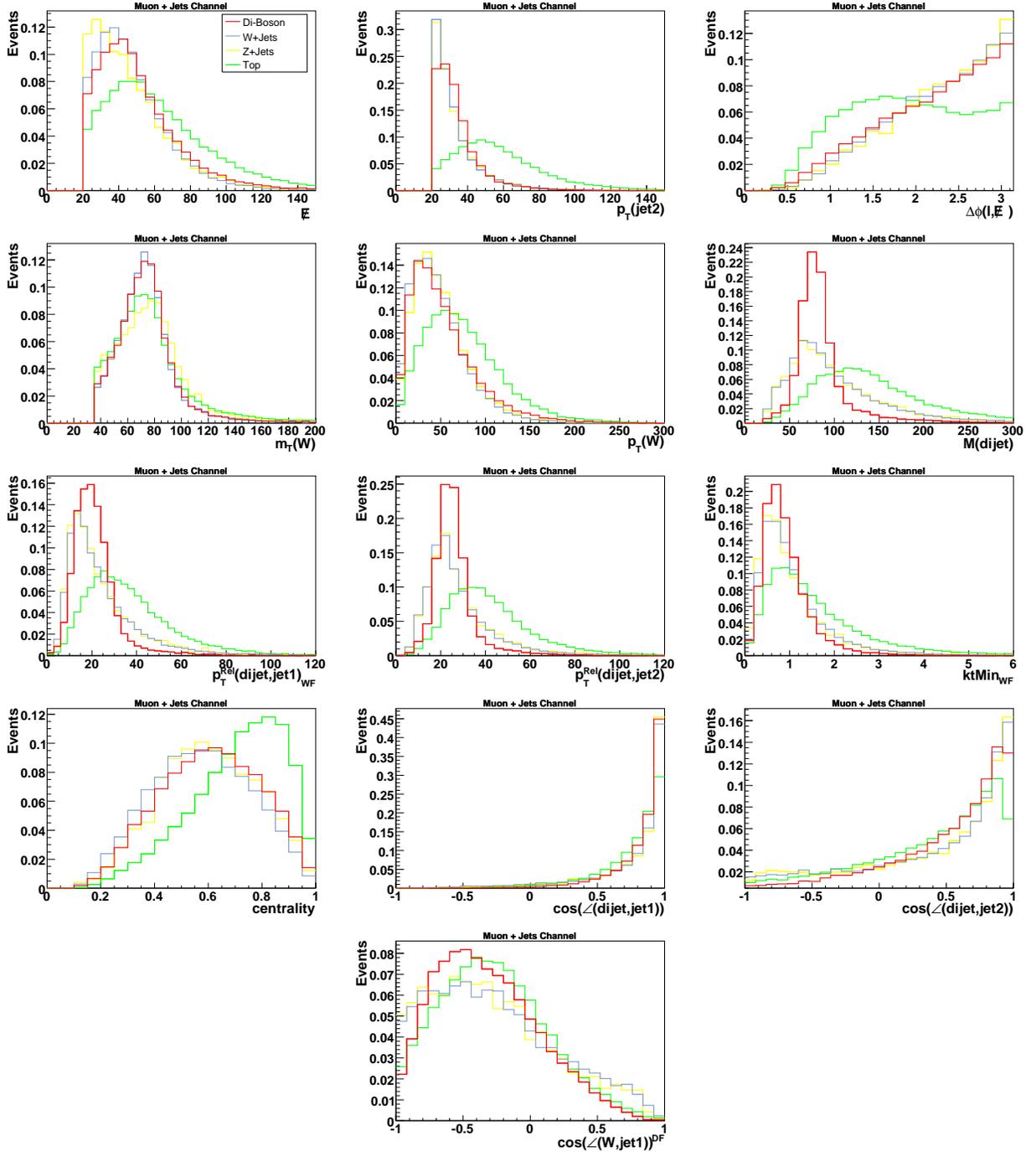
Figure 7.6: Shape comparison for different physics processes (each integral normalized to unity) for the input variables used by the random forest for the $\mu\nu q\bar{q}$ channel.

validation are then averaged. The validation figure-of-merit used is the average quadratic loss

$$\Lambda = \frac{1}{\sum_{i=0}^{N} w_i} \sum_{i=0}^{N} w_i \left(y_i - \Phi(\vec{x}_i)\right)^2 ; \qquad (7.1)$$

where y_i is the desired output for the ith training event (*i.e.*, zero for background and one for signal), w_i is the event weight, and $\Phi(\vec{x}_i)$ is the classifier output for that event. This is a measure of how far the classifier output, $\Phi(\vec{x}_i)$, is from the desired output, y_i. Smaller loss means better classification.

Determining the best value for each of the parameters was an iterative process because the optimal value for any given parameter may depend on the values used for the other parameters. Starting from rough estimates for each of the parameters, the parameters were cycled through one-by-one, optimizing and updating each parameter, then moving to the next. The cycle continued until the parameter settings stabilized.

All split criteria available in StatPatternRecognition were investigated and it was found that the Gini index, cross-entropy, signal purity, and correctly classified fraction all gave similar performance. The final decision was to use the Gini index (also Gini coefficient or Gini ratio) as it is one of the most commonly used criteria. The performance tends to be best when the maximum number of variables for each tree is in the neighborhood of three-quarters of the total number of inputs, but is fairly insensitive to the exact value and was ultimately set at eleven. The minimum number of events per terminal node should be relatively small for a random forest (or any bagged classifier) and the optimal value was found to be around 100.

To determine the number of decision trees to use in the forest, instead of using cross-validation, the training set was simply divided into two pieces with 60% used for training and the remaining 40% for validation. The change in the event loss was monitored as the number of decision trees increased. The loss should decrease and reach a stable value as more trees are added. Using more trees than necessary increases training and evaluation time without improving the accuracy of the classifier. It was found that 50 trees resulted in

a fairly fast classifier with close to the lowest attainable event loss.

7.2.4   Final Discriminant

Once the input variables were chosen and the parameter settings were optimized, a random forest classifier was constructed for each channel using the full training sets. The data and all of the MC events not used for training were then evaluated by the respective channel's random forest, producing the distributions to be used for the cross section measurement. The random forest output distributions are shown in figures 7.7 for both the $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channels. These figures show the comparison between data and the nominally predicted rates for signal and all backgrounds.
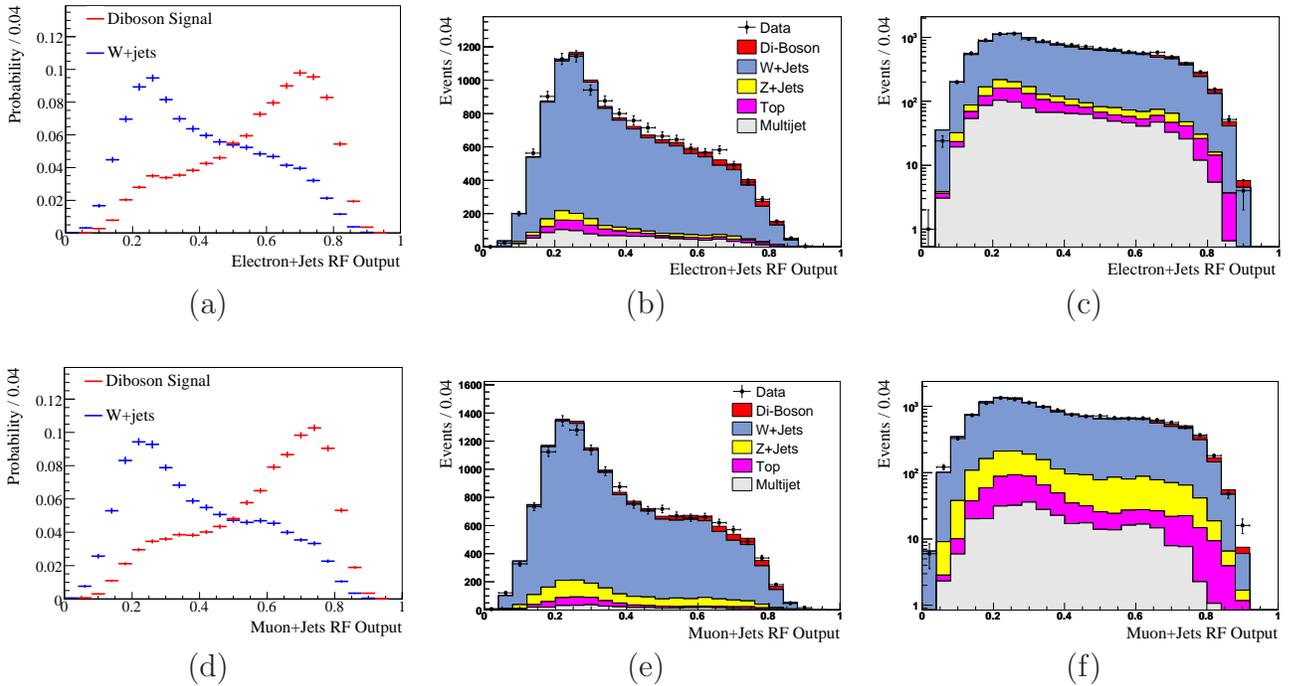

Figure 7.7: (a) Comparison of the random forest output in the $e\nu q\bar{q}$ channel for signal and W+jets background. (b,c) Comparison of data to the SM prediction for the random forest output in the $e\nu q\bar{q}$ channel in linear and log scale. (d) Comparison of the random forest output in the $\mu\nu q\bar{q}$ channel for signal and W+jets background. (e,f) Comparison of data to the SM prediction for the random forest output in the $\mu\nu q\bar{q}$ channel in linear and log scale.

Following the construction of a common final discriminant for the $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channels, a comparison of the events selected in data and MC samples can be made. The expected

numbers of events include all reported normalizations. To further compare the relative significance of the signal rate, the expected and observed events are compared after a simple cut of 0.5 on the output of the random forest classifier. This cut is performed for illustration only and has little bearing on the final measurement. Tables 7.1 and 7.2 outline the expected and observed number of events for the $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channel, respectively.

Table 7.1: The expected and observed number of events in the $e\nu q\bar{q}$ channel before and after requiring the random forest classifier output to be above 0.50. The event yields are accompanied by the Poisson uncertainty from limited statistics (*i.e.*, without the systematic uncertainties). Also given is the simple statistical significance of the signal before considering systematic uncertainties.

Source	Pre-Cut	Post-Cut
Diboson Signal	360.5 ± 2.3	246.5 ± 1.9
W+jets	10226 ± 76	3331 ± 42
Z+jets	408 ± 13	104.6 ± 6.5
Top	463.3 ± 2.2	113.6 ± 1.1
Multijet	825 ± 11	225.0 ± 5.8
Total Predicted	12283 ± 78	4021 ± 43
Data	12473 ± 112	4111 ± 64
$S/\sqrt{S+B}$	3.25	3.85

Table 7.2: The expected and observed number of events in the $\mu\nu q\bar{q}$ channel before and after requiring the random forest classifier output to be above 0.50. The event yields are accompanied by the Poisson uncertainty from limited statistics (*i.e.*, without the systematic uncertainties). Also given is the simple statistical significance of the signal before considering systematic uncertainties.

Source	Pre-Cut	Post-Cut
Diboson	427.3 ± 2.7	290.6 ± 2.2
W+jets	12012 ± 88	3690 ± 48
Z+jets	1239 ± 20	382 ± 10
Top	437.0 ± 2.2	104.5 ± 1.0
Multijet	327.0 ± 9.6	86.7 ± 5.0
Total Predicted	14442 ± 91	4553 ± 49
Data	14392 ± 120	4633 ± 68
$S/\sqrt{S+B}$	3.56	4.31

Chapter 8

Systematic Uncertainties

This analysis considers the effects from systematic uncertainties that change the shape of kinematic distributions as well as systematics uncertainties that only change the overall normalization of the contributing physics processes. To distinguish systematic uncertainties with a kinematic dependence from those without, the former are referred to as shape uncertainties and the latter as normalization uncertainties. Although shape uncertainties may also impact efficiencies and normalization, any uncertainty that changes the expected multivariate classifier distribution for any of the contributing processes is considered a shape uncertainty.

The largest normalization uncertainties are the cross section factors used to normalize the background processes. These are related to the accuracy of the theoretical cross section calculation and the uncertainty on the normalization of the multijet estimations. Also, the component of the W+jets background having heavy flavor (b and c quarks) jets has an additional degree of theoretical uncertainty on top of the overall W+jets cross section uncertainty, which must be taken into account because the kinematics for these events can be slightly different than W events with light flavor (u, d, and s quark) jets[1]. Normalization uncertainties are also associated with the efficiencies for reconstructing and triggering

---

[1]The contribution from Z+jets is more than an order of magnitude smaller than W+jets so the additional normalization uncertainty for Z events with heavy flavor jets is negligible.

on leptons and the uncertainty on the additional energy smearing required for jets in the MC samples, each of which was explained in chapter 6. Each normalization uncertainty is considered to arise from a Gaussian parent distribution.

The shape uncertainties are those that, when propagated through the analysis selection, impact the shape of the distribution (*e.g.*, the random forest output or dijet mass) used to perform the measurement (chapter 9). The shape uncertainty on the signal and background simulations due to PDF uncertainties is described in appendix G. All other shape uncertainties considered in this analysis were measured as described in chapter 6. The dependence of the random forest output distribution (or any other distribution) on these uncertainties is determined by varying each parameter by its associated uncertainty ($\pm 1\sigma$) and re-evaluating the shape of the distribution. The magnitude of the resulting shape dependence is considered to arise from a Gaussian parent distribution.

Each systematic uncertainty is assumed to be 100% correlated amongst backgrounds and signals. All sources of systematic uncertainty are assumed to be mutually independent, and no inter-correlation is propagated.

A list of the systematic uncertainties used in this analysis can be found in table 8.1. The dependence of each shape uncertainty on the random forest output distribution for each affected physics process can be found in appendix H. As explained in the next chapter, when measuring the diboson signal cross section the W+jets normalization is a free parameter and the uncertainty on the W+jets cross section is not used. However, the size of the uncertainty must be specified for generating the pseudo-data used to estimate the significance of the signal excess.

Given in table 8.2 is the contribution of each systematic uncertainty to the total systematic uncertainty of 3.6 pb on the measured cross section. As explained in chapter 9, the measured cross section is determined by the best fit of the random forest output distributions to the data, in which the size of the signal contribution is floated. The measured cross section is then given by $\sigma^{\mathrm{meas}}(WV) = S_{\mathrm{fit}} \cdot \sigma^{\mathrm{th}}(WV)$; where $\sigma^{\mathrm{th}}(WV)$ is the theoretical cross

section and S_{fit} is the scale factor for the signal determined by the fit. The systematic uncertainty of 3.6 pb on the measured cross section is determined by the systematic uncertainty on the fit parameter multiplied by $\sigma^{\text{th}}(WV)$ and does not include the uncertainty on the integrated luminosity (6.1%) of the dataset, which is assigned as an addition uncertainty on the measurement.

Table 8.1: Systematic uncertainties in percent for MC simulations and multijet estimates. The nature of the uncertainty, *i.e.*, whether it refers to a shape (S) or normalization (N) uncertainty, is also provided. The values given for shape uncertainties correspond to the approximate amplitude of the fluctuations in the random forest output distribution, the full shape dependence can be found in appendix H.

Source of systematic uncertainty	Diboson	W+jets	Z+jets	Top	Multijet	Nature
Trigger efficiency, $e\nu q\bar{q}$	+2/-3	+2/-3	+2/-3	+2/-3		N
Trigger efficiency, $\mu\nu q\bar{q}$	+0/-5	+0/-5	+0/-5	+0/-5		S
Lepton identification	±4	±4	±4	±4		N
Jet identification	±1	±1	±1	± <1		S
Jet energy scale	±4	±9	±9	±4		S
Jet energy resolution	±3	±4	±4	±4		N
Cross section		±20[a]	±6	±10		N
Heavy flavor jet contribution		±20				N
Diboson NLO correction	±10					S
PDF set	±1	±1	±1	±1		S
ALPGEN jet η and ΔR corrections		±1	±1			S
Renormalization and factorization scale		±3	±3			S
ALPGEN parton-jet matching parameters		±4	±4			S
Multijet normalization, $e\nu q\bar{q}$					±20	N
Multijet normalization, $\mu\nu q\bar{q}$					±30	N
Multijet shape, $e\nu q\bar{q}$					±6	S
Multijet shape, $\mu\nu q\bar{q}$					±10	S

[a]The cross section uncertainty on W+jets is not used when measuring the diboson signal cross section (the W+jets normalization is a free parameter); however, it is necessary for generating pseudo-data used in the significance estimation.

Table 8.2: The contribution of each systematic uncertainty to the total systematic uncertainty of 3.6 pb on the measured cross section, which does not include the additional uncertainty of 6.1% on the integrated luminosity of the dataset.

Source of systematic uncertainty	$\Delta\sigma$ [pb]
Trigger efficiency, $e\nu q\bar{q}$	< 0.1
Trigger efficiency, $\mu\nu q\bar{q}$	< 0.1
Lepton identification	< 0.1
Jet identification	0.3
Jet energy scale	1.9
Jet energy resolution	< 0.1
Cross section	1.1
Heavy flavor jet contribution	< 0.1
Diboson NLO correction	< 0.1
PDF set	0.2
ALPGEN jet η and ΔR corrections	< 0.1
Renormalization and factorization scale	0.9
ALPGEN parton-jet matching parameters	2.4
Multijet normalization, $e\nu q\bar{q}$	0.9
Multijet normalization, $\mu\nu q\bar{q}$	0.5
Multijet shape, $e\nu q\bar{q}$	< 0.1
Multijet shape, $\mu\nu q\bar{q}$	< 0.1

Chapter 9

Measurement

The $WW + WZ$ cross section is measured by performing a fit to the data using the signal and background distributions for a chosen "final variable" (FV), as described in the following section. The most precise measurement will come from a FV with the best separation of signal from background, which was the motivation for constructing a multivariate classifier (chapter 7). The cross section measurement is also performed using the dijet mass as the FV to supply a cross check and to illustrate the improvement in precision achieved by using the random forest classifier.

Following the measurement of the cross section, the significance of the signal excess is evaluated. A Frequentist approach is used to count the fraction of times one would expect the background-only scenario to fluctuate such that a measurement of the $WW + WZ$ cross section would produce a result as high as the observed (or expected) cross section.

9.1 $WW + WZ$ Cross Section

A signal template of the FV distribution (*e.g.*, the random forest output or dijet mass) is obtained for each channel by populating a histogram using the simulated signal events. Similarly, the simulated background samples and multijet estimation provide FV templates for the contributing background processes. The signal and background templates are then

fit to the FV distribution given by the data in each channel in order to determine the size of any excess (over the background-only scenario) that is consistent with the kinematics of W^+W^- and $W^\pm Z^0$ production. The fit procedure is performed as follows:

1. The signal and background templates are initialized to the SM predictions with all nuisance parameters (systematic uncertainties) set to their central values. The total number of predicted events in the ith bin for the nominal templates is denoted p_i.

2. The N_{sys} nuisance parameters are then allowed to deviate from their central values, thus changing the predicted number of events in each bin to

$$p_i' = p_i \prod_k^{N_{\mathrm{sys}}} \left(1 + R_k \sigma_{k,i}\right); \tag{9.1}$$

where R_k is the number of standard deviations by which kth nuisance parameter was deviated from the central values and $\sigma_{k,i}$ is size of one standard deviation for the kth nuisance parameter in the ith bin (*i.e.*, the size of the kth systematic uncertainty in the ith bin).

3. The best fit is determined by minimizing the negative log-likelihood (NLL) function [65]

$$\mathrm{NLL} = -2\ln\mathcal{L} = 2\sum_i^{N_{\mathrm{bins}}} \left[p_i' - d_i - d_i \ln\left(\frac{p_i'}{d_i}\right)\right] + \sum_k^{N_{\mathrm{sys}}} (R_k)^2; \tag{9.2}$$

where d_i is the observed number of data in the ith bin and N_{bins} is the total number of bins.

The first sum in equation 9.2 (over N_{bins}) is the Poisson χ^2 between the observed data and the templates that have been varied within the systematic uncertainties. The second sum in equation 9.2 (over N_{sys}) supplies a penalty to the NLL for the number of standard deviations by which each systematic uncertainty has been deviated.

The normalization of the signal is included in the fit as an unconstrained parameter (*i.e.*, floated), allowing the fit to freely determine the diboson signal cross section that best

matches the data. The uncertainty on the W+jets cross section (or k-factor) is also floated, simultaneously fitting the W+jets cross sections. Floating the W+jets cross section necessarily reduces the sensitivity of the measurement, but it basically amounts to determining the normalization of the dominant background using on the sideband (low signal) regions of the FV distribution (*e.g.*, random forest output $\lesssim 0.5$).

The results from the measurements using the random forest output are given in first three rows of table 9.1. The measurements performed separately for each channel indicate a signal rate slightly larger than expected, but consistent with prediction and between each channel. The most precise measurement is obtained by fitting both channels simultaneously. The combined measurement[1] using the random forest output distribution yielded a diboson cross section of $\sigma(WW + WZ) = 20.2 \pm 4.5$ pb. The individual contributions to the systematic uncertainty on the measurement are given in table 8.2, with the dominant contributions coming from the modeling of the W+jets background and the JES. The random forest output distributions resulting from the combined fit are shown in figures 9.1 and 9.2. Figure 9.3 shows a plot of the dijet mass distribution after this combined fit. The dijet mass resonance of the signal is clearly visible in the background-subtracted data and has a peak within one half of a standard deviation of the extracted signal.

Table 9.1: The measured cross section for $W^+W^- + W^{\pm}Z^0$ production determined by the best fit to the data for the specified FV and channel.

Final Variable	Channel	Measured Signal Cross Section [pb]
Random Forest	$e\nu q\bar{q}$	18.0 $\pm$3.7(stat) $\pm$5.2(sys) $\pm$1.1(lum)
Random Forest	$\mu\nu q\bar{q}$	22.8 $\pm$3.3(stat) $\pm$4.9(sys) $\pm$1.4(lum)
Random Forest	Combined	20.2 $\pm$2.5(stat) $\pm$3.6(sys) $\pm$1.2(lum)
Dijet Mass	Combined	18.5 $\pm$2.8(stat) $\pm$4.9(sys) $\pm$1.1(lum)

As a cross check, the measurement was also performed using the dijet mass distribution as the FV. The signal cross section measured from the combined fit using the dijet mass distributions is given in the last row of table 9.1. In all cases, the measurements indicate a

---

[1]The combined measurement fits the FV distribution from both channels simultaneously by summing over bins in both channels when calculating the NLL (equation 9.2).

preference for a signal cross section slightly larger than the input values of $\sigma(WW) = 12.4$ pb and $\sigma(WZ) = 3.7$ pb, but the behavior is consistent between the channels and when using the dijet mass or the random forest output. The combined fit using the dijet mass distribution yields a cross section for $WW + WZ$ production of 18.5 ± 5.7 pb, which, as expected, is considerably less precise then 20.2 ± 4.5 pb obtained from the combined fit using the random forest output.
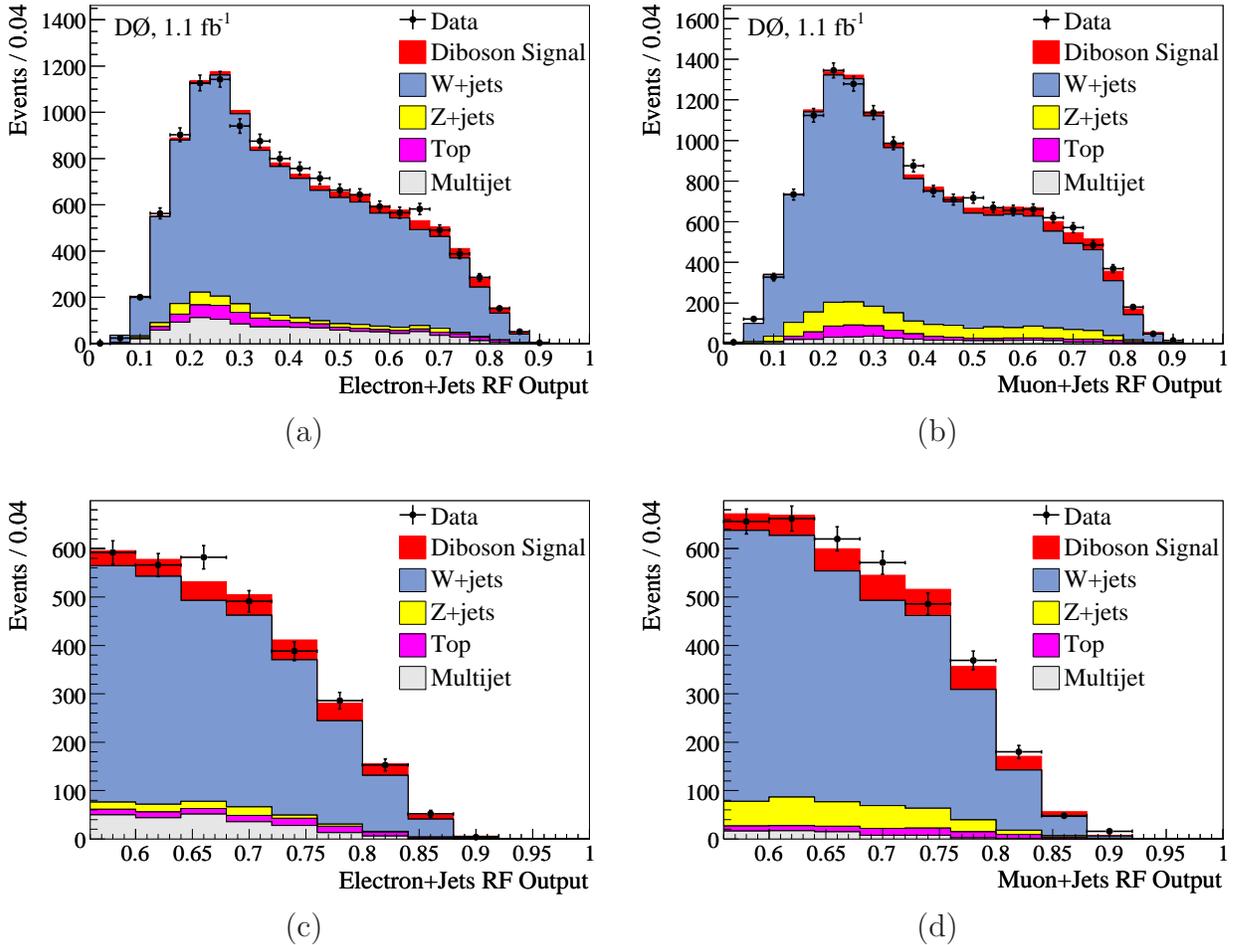

Figure 9.1: A comparison of data to the signal and background estimations for full distribution and the signal region of the random forest output in (a,c) the $e\nu q\bar{q}$ channel and (b,d) the $\mu\nu q\bar{q}$ channel after the combined fit of the random forest output.

In each case the best fit for the W+jets k-factor was found to be approximately 1.53. The correlation of the NLL response in the two-dimensional plane of the signal cross section and W+jets k-factor is shown in figures 9.4 and 9.5. For figure 9.4, all other fit parameters
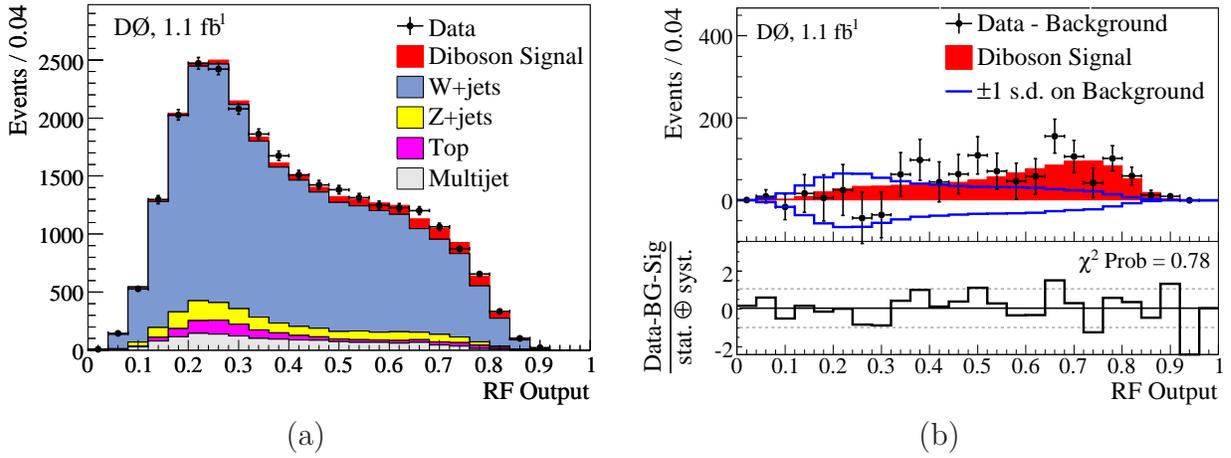
Figure 9.2: (a) A comparison of data to the signal and background estimations for the random forest output in both channels together after the combined fit of the random forest output. (b) A comparison of the extracted signal to the background-subtracted data and the ±1 standard deviation (s.d.) bands of the systematic uncertainty on the background. Also given is the residual distance between the data and measurement in units of total uncertainty.
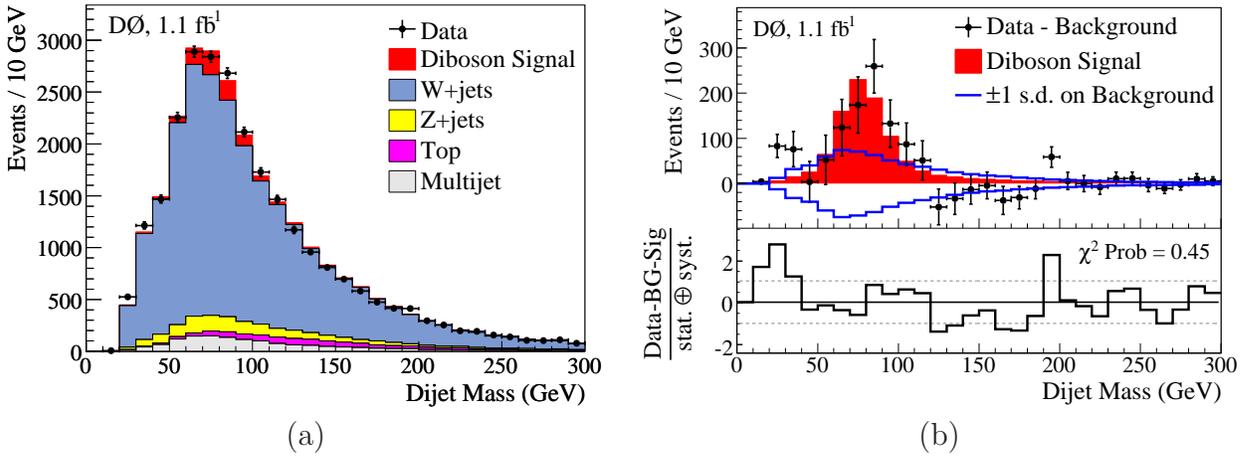
Figure 9.3: (a) A comparison of data to the signal and background estimations for the dijet mass distribution in both channels together after the combined fit of the random forest output. (b) A comparison of the extracted signal to the background-subtracted data and the ±1 standard deviation (s.d.) bands of the systematic uncertainty on the background. Also given is the residual distance between the data and measurement in units of total uncertainty.

were fixed at the best-fit values while the two test parameters are scanned. For figure 9.5, the NLL function was re-minimized for each set of test parameters.

(a)   (b)   (c)
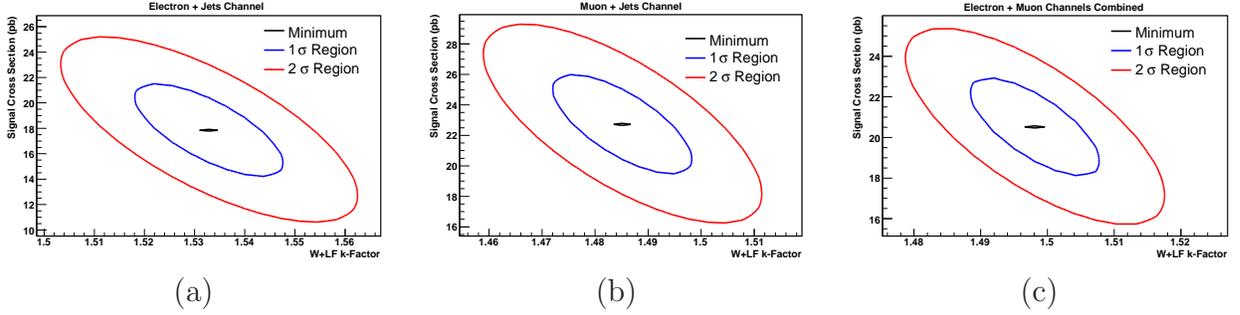
Figure 9.4: The two-dimensional correlation of the signal cross section and the W+jets k-factor for the (a) $e\nu q\bar{q}$ channel, (b) $\mu\nu q\bar{q}$ channel, and (c) combined fit using the random forest output; here the one (two) σ region is defined by the points where the NLL increases by 1.0 (4.0) as the two parameters are deviated from the best fit values and all other fit parameters are fixed at the values determined from the best fit.

(a)   (b)   (c)
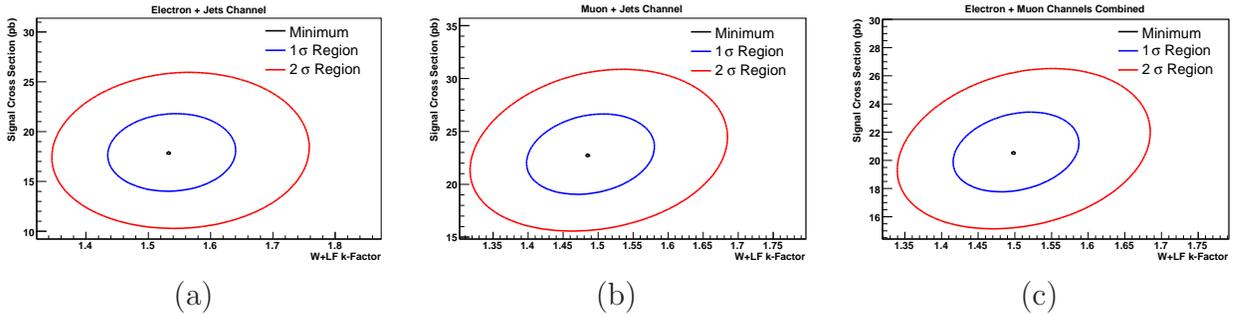
Figure 9.5: The two-dimensional correlation of the signal cross section and the W+jets k-factor for the (a) $e\nu q\bar{q}$ channel, (b) $\mu\nu q\bar{q}$ channel, and (c) combined fit using the random forest output; here the one (two) σ region is defined by the points where the NLL increases by 1.0 (4.0) as the two parameters are deviated from the best fit values and all other fit parameters are adjusted to minimized the NLL for the given signal cross section and W+jets k-factor.

9.2   Significance

Arguably just as important as the measurement itself is the significance of the measurement, *i.e.*, the likelihood that there actually is no signal and the measured signal is just a signal-like fluctuation in the background. This analysis takes the Frequentist approach, asking

the question: If there truly is no signal and this experiment is repeated many times (many equivalent datasets are collected and analyzed, each subject to the same uncertainties), what fraction of the experiments will report a signal cross section as large as this analysis observed or expected?

To answer this question, the data distributions from such experiments are simulated by generating sets of "pseudo-data" for the background-only hypothesis. It is assumed that for a real experiment the data in each bin of the FV distribution is a stochastic variable from a Poisson distribution with a mean given by the expected number of events for that bin. Thus, a pseudo-data histogram is generated by setting the value of each bin to the result of a random Poisson trial seeded by the predicted mean number of events in the bin. For the background-only hypothesis, the predicted mean number of events in each bin is given by summing the contribution from each background template. In order to include the uncertainties on the nuisance parameters, each nuisance parameter is shifted by a random amount (based on its uncertainty) and the templates are adjusted accordingly before throwing the Poisson trials for each bin.

Repeating this process many ($\gtrsim 10^6$) times produces an array of background-only pseudo-data distributions with a range of systematic uncertainties and statistical uncertainties consistent with repeated experiments on the background-only scenario. Each background-only pseudo-data distribution is then fit in the same manner as the data (section 9.1), with the signal cross section and W+jets k-factor taken as free parameters in the fit. The resulting signal cross section measurements from all of the background-only pseudo-data distributions are put into a histogram. The expected significance is determined by the fraction of outcomes with a signal cross section equal to or greater than the predicted cross section and the observed significance is determined by the fraction at or above the measured cross section. The results for the $e\nu q\bar{q}$ channel, $\mu\nu q\bar{q}$ channel, and combined measurement significance are shown in figures 9.6, 9.7, and 9.8, respectively, and summarized in table 9.2. Also shown in figures 9.6, 9.7, and 9.8 is a comparison of the observed cross section to the pseudo-

experiments drawn from the signal-plus-background hypothesis, which determine the degree of compatibility with the SM prediction.

Table 9.2: The p-values for the expected and observed significance obtained by comparison with pseudo-data distributions drawn from a background-only model. In parentheses is the corresponding Gaussian significance in number of standard deviations.

Channel	Expected p-value	Observed p-value
$e\nu q\bar{q}$	6.8×10^{-3} (2.5σ)	3.2×10^{-3} (2.7σ)
$\mu\nu q\bar{q}$	1.8×10^{-3} (2.9σ)	5.2×10^{-5} (3.9σ)
Combined	1.5×10^{-4} (3.6σ)	5.4×10^{-6} (4.4σ)

(a)   (b)   (c)
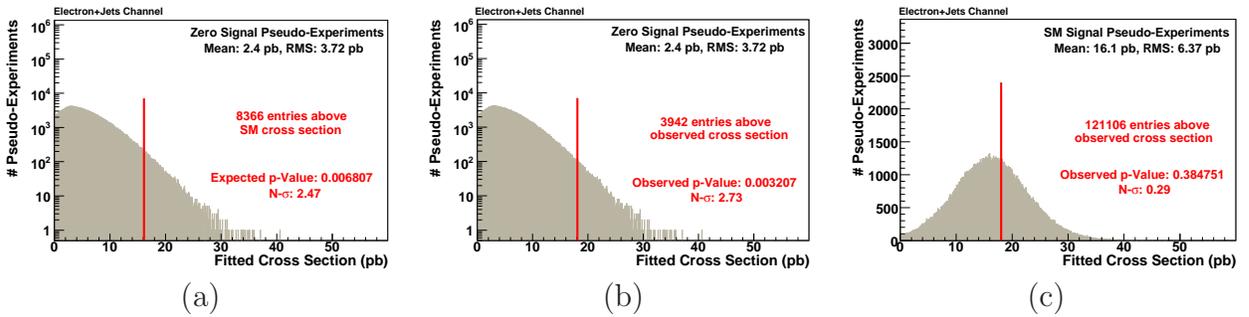
Figure 9.6: (a,b) Distribution of fitted signal cross sections in pseudo-data drawn from the background-only hypothesis compared to the expected and observed signal cross sections for the $e\nu q\bar{q}$ channel. (c) Distribution of fitted signal cross sections in pseudo-data drawn from the SM signal-plus-background hypothesis compared to the observed signal cross sections for the $e\nu q\bar{q}$ channel.
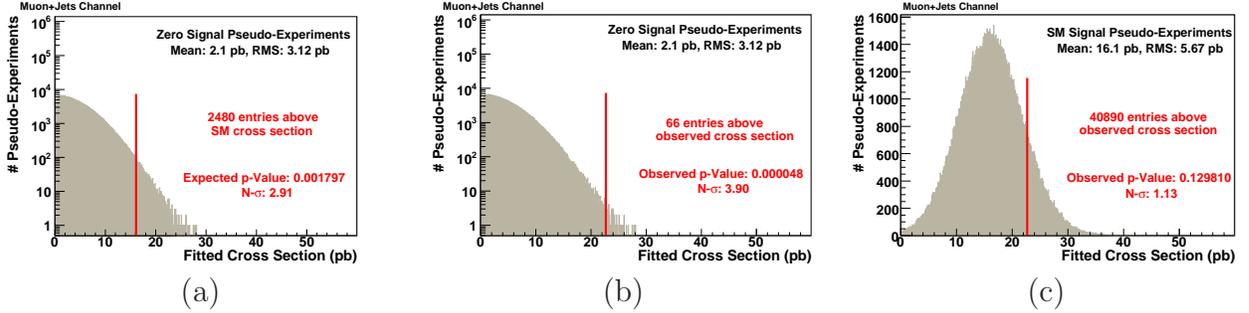
Figure 9.7: (a,b) Distribution of fitted signal cross sections in pseudo-data drawn from the background-only hypothesis compared to the expected and observed signal cross sections for the $\mu\nu q\bar{q}$ channel. (c) Distribution of fitted signal cross sections in pseudo-data drawn from the SM signal-plus-background hypothesis compared to the observed signal cross sections for the $\mu\nu q\bar{q}$ channel.
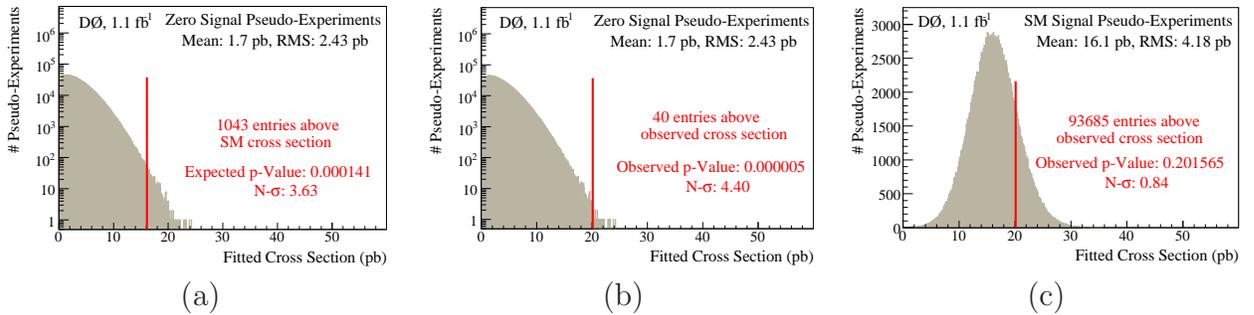
Figure 9.8: (a,b) Distribution of fitted signal cross sections in pseudo-data drawn from the background-only hypothesis compared to the expected and observed signal cross sections for the combined measurement. (c) Distribution of fitted signal cross sections in pseudo-data drawn from the SM signal-plus-background hypothesis compared to the observed signal cross sections for the combined measurement.

116

Chapter 10

Conclusions

This thesis presented a description of the detection, selection, and analysis of data consistent with the production of $WW/WZ \to \ell\nu q\bar{q}$ events in proton-antiproton collisions recorded by the D0 experiment at Fermilab. The observed data were compared to simulated signal and background samples and yielded a measurement of $\sigma(p\bar{p} \to W^+W^-) + \sigma(p\bar{p} \to W^\pm Z^0) = 20.2 \pm 4.5$ pb at $\sqrt{s} = 1.96$ TeV. The observed significance of this measurement is 4.4 σ. This measurement is compatible at the level of 0.8 σ with the theoretical prediction of $\sigma(W^+W^-) + \sigma(W^\pm Z^0) = 16.1 \pm 0.9$ pb [1].

This analysis represents the first evidence for $W^+W^-/W^\pm Z^0$ production with lepton plus jets final states at a hadron collider. Measurements of WW and WZ production at the Tevatron in the fully leptonic final states have yielded results of $\sigma(W^+W^-) = 13.8 \pm 4.6$ pb in roughly 240 pb^{-1}, $\sigma(W^+W^-) = 12.3 \pm 2.0$ pb in roughly 1.0 fb^{-1}, and $\sigma(W^\pm Z^0) = 2.7^{+1.7}_{-1.3}$ pb in roughly 1.0 fb^{-1} by the D0 experiment [66]. Results from the CDF experiment found $\sigma(W^+W^-) = 13.6 \pm 3.1$ pb in roughly 845 pb^{-1} and $\sigma(W^\pm Z^0) = 4.3^{+1.4}_{-1.1}$ pb in roughly 1.9 fb^{-1} [67]. A summary plot of electroweak cross section measurements for Run II of the Tevatron is shown in figure 10.

The techniques used in this analysis, in particular the use of a multivariate discriminant, are becoming ubiquitous in searches for the SM Higgs boson (and beyond the SM Higgs
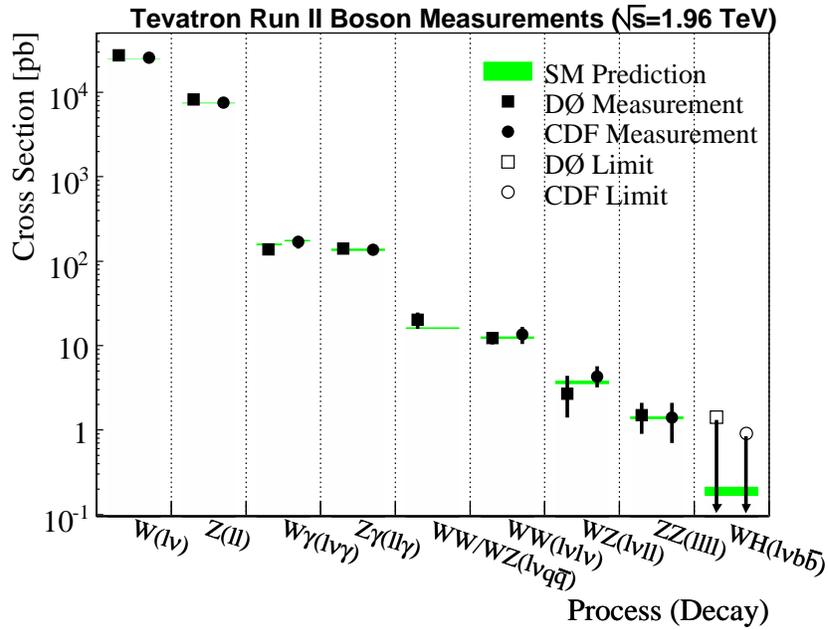
Figure 10.1: Summary of electroweak cross section measurements and limits on $WH \rightarrow \ell\nu b\bar{b}$ for a Higgs mass of 115 GeV from the D0 and CDF collaborations.

bosons). As physicists push ever closer to discovering the existence (or absence) of the Higgs boson, it becomes increasingly important to verify the validity of the analysis techniques used in these Higgs boson searches. The similar final states and the common experimental challenges associated with finding a small signal in an enormous background make this analysis the ideal proving ground; and the results of this analysis provide an experimental validation of these advanced analysis techniques.

Appendix A

Analysis of Inclusive Muon Trigger Events

This appendix contains the results of a study to determine the possible shape-dependence biases in the kinematic variables for the $\mu\nu q\bar{q}$ channel as a result of using an inclusive trigger selection. For the dataset, comparisons were made between using the following three trigger selections: single muon trigger suite, single muon and muon+jet trigger suite, and inclusive triggers (*i.e.*, not trigger requirement). For the simulated MC samples, comparisons were only made between using the single muons trigger suite efficiencies and using no trigger efficiencies (inclusive triggers) because a proper set of trigger efficiencies for the single muon and muon+jet trigger suite was not available at the time of this analysis. Figures A.1, A.2, and A.3 show comparisons between the different trigger selections for various kinematic distributions in data. Figures A.4 and A.5 show comparisons between the trigger selections for various kinematic distributions in the signal and W+jets MC samples, respectively. Lastly, figures A.6-A.9 have plots showing the effect of the various triggers on the random forest output.

The number of selected data passing each set of the triggers are 9070, 13436, and 14831 for the single muon trigger suite, single muon and muon+jet trigger suite, and inclusive triggers,

respectively. Of the 1395 (9.4%) events not selected by the single muon and muon+jet trigger suite, 284 (1.9%) are selected by only one trigger. Events that fired more than one trigger can be expected to be selected with a high efficiency, while those which fire only one trigger would likely be subject to trigger inefficiency. The triggers that fired for these 284 events are outlined in table A.1, indicating that the majority come from muon-based triggers.

The residual inefficiency is therefore estimated to be on the order of 2% with any residual shape difference for kinematic distributions likely bounded by the shapes given by a fully efficient trigger and the shape of the single muon and muon+jet trigger suite. The residual shape uncertainty is then obtained via the relative difference in shape between the distributions obtained using the single muon and muon+jet trigger suite and an inclusive trigger selection. This relative difference is applied as a shape uncertainty in the cross section and significance calculations, allowing the data to resolve the true value. As a conservative treatment, the uncertainty is parameterized as a linear interpolation between the two fixed boundaries with a flat probability distribution (as opposed to the usual assumption of a Gaussian prior).

Table A.1: The list triggers that fired for the 284 events that fire only one trigger and were not selected by the single muon and muon+jet trigger suite. The triggers are grouped according to primary selection criteria and show the number of events that fired for each trigger (and group).

Muon+X Triggers	Fired	EM+X Triggers	Fired	Jet Triggers	Fired	Tau Triggers	Fired
2MU_A_L2M0	1	2CEM6	1	4CJT5	2	TAU1_ITU20_MET15	4
DMU1_1L1MM2	1	CEM10_2CJT5	1	4JT10	3	TAU1_NN10_MET10	1
DMU1_2LM2_LM5	1	E17_ITK10_M25	1	3JT15_L2L0_PV	1	TAU1_NN20_MET15	1
DMU1_JT12_TLM3	1	E1_SH15_2JHA100	1	3CJT5_JT20_L2M0	4	TAU2_2NN10_15INN	1
DMU4_2TAM_IMJ_VX	1	EM_HI_2EM5_TR	6	JT1_ACO_MHT_BDV	1	TAU2_2NN15_20ITK	3
MM1_JT15_HA_TK10	1	EM_HI_METF0	2	JT1_ACO_MHT_HT	14	TU3_2N03_2T10I10	2
MM1_JT25	2	EM_HI_TR	4	JT1_ACO_MHT_LM0	1		
MM2_LEL12	1	EM_MX_F0	1	JT2_3JT15L_IP_VX	1		
MT10_JT20_L2M0	4			JT3_3JT10L_LM3_V	10		
MT3_L2M0_MM3	1			JT3_3JT12L_MM3_V	4		
MUH1_TK10H	3			JT4_HT_LM3_2LM0	1		
MUH2_ILM12_ITK12	4			JT6_ACO_MHT_LM0	2		
MUH4_ILM12_ITK12	1			JT_125TT	1		
MUI_TAU_L2M0	1			JT_45TT	3		
MUJ1_ILM10T10NN1	3			JT_45TT_GAPN	1		
MUJ1_ILM10T5NN3	1			JT_65TT	1		
MUJ1_JT15HA_LM6V	2			MET20_TIS10_CJT7	12		
MUJ2_JT15HA_LM8V	3			MHT30_3CJT5	14		
MUJ2_JT20_LM10	15						
MUJ2_JT20_TK10	15						
MUJ2_JT25_ILM3	3						
MUJ2_JTHATK_LMVB	1						
MU_2TRK3_L2M0	10						
MU_A_EM10	1						
MU_A_L2M3_TRK10	3						
MU_EM_L2M5	8						
MU_JT10_L2M0_IP	2						
MU_JT15_L3M0	1						
MU_JT15_MET10	9						
MU_JT20_MET10	30						
MU_TAU10_2T_L2M0	1						
MU_TAU10_L2M0	6						
MU_W_L2M0_2TRK3	1						
MWLT10_2EM3_SHT7	1						
MWLXT10_L0_L3L15	1						
MWTXT10_TK10	18						
mu1ptxatxx_CEM5	4						
mu1ptxatxx_CJT3	1						
mu1ptxatxx_CJT5	1						
mu1ptxbtxx_fz	1						
mu1ptxctxx_CEM5	1						
MUEM2_LEL12_MM5	3						
MUEM2_LEL15_MM5	1						
MUEM3_LEL15_TRK5	1						
MUEM3_SH7_TRK3	1						
MUEM5_LEL15	2						
MEB1_TM0_VX	5						
TOTAL	179		17		76		12

Figure A.1: Comparison between selected **data** using inclusive triggers (black) and using the single muon trigger suite (red) for various kinematic distributions (labeled on the x-axis). The plots on the left show each distribution normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the $\pm 5\%$ region and vertical red lines indicating the highest signal density region.
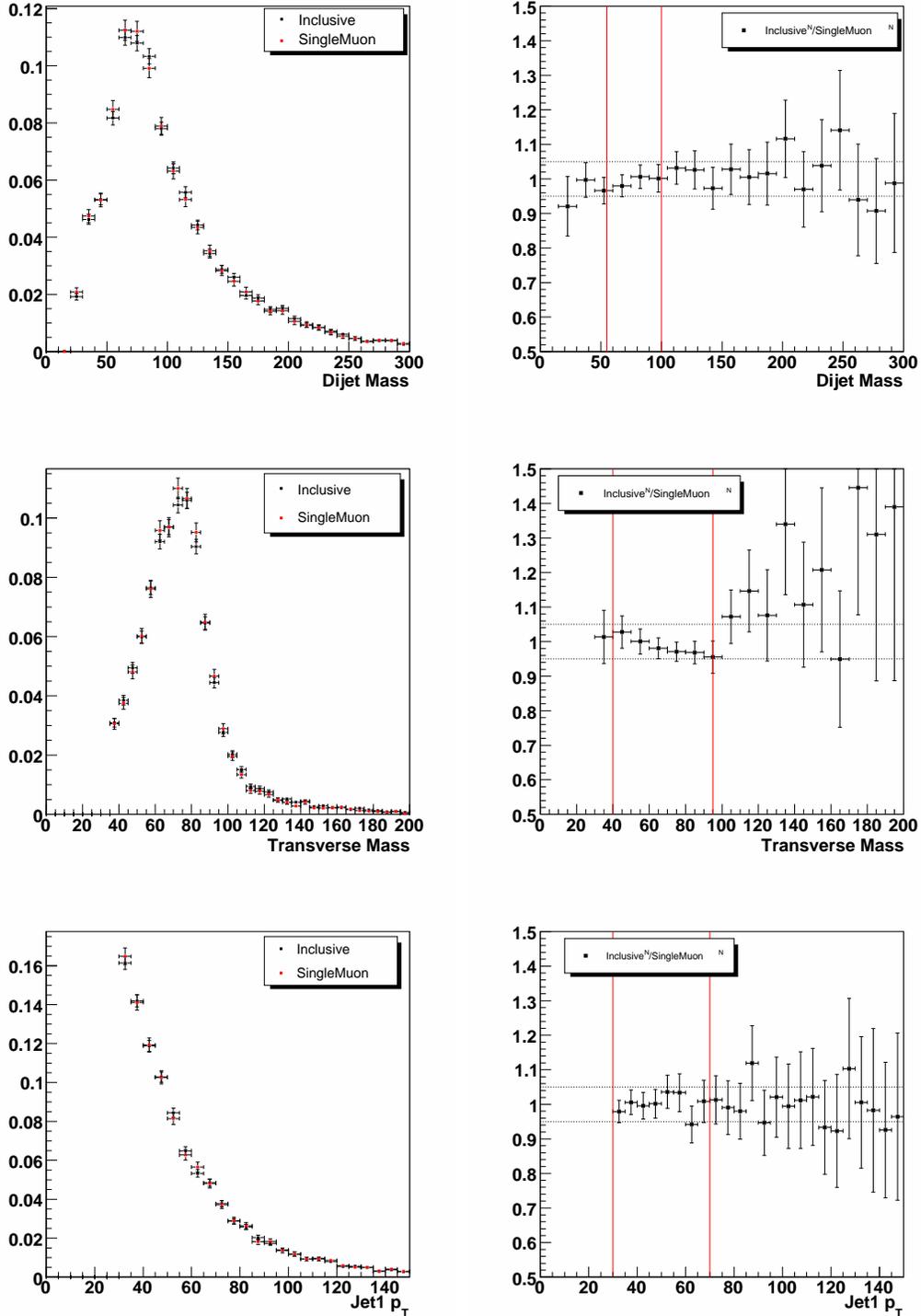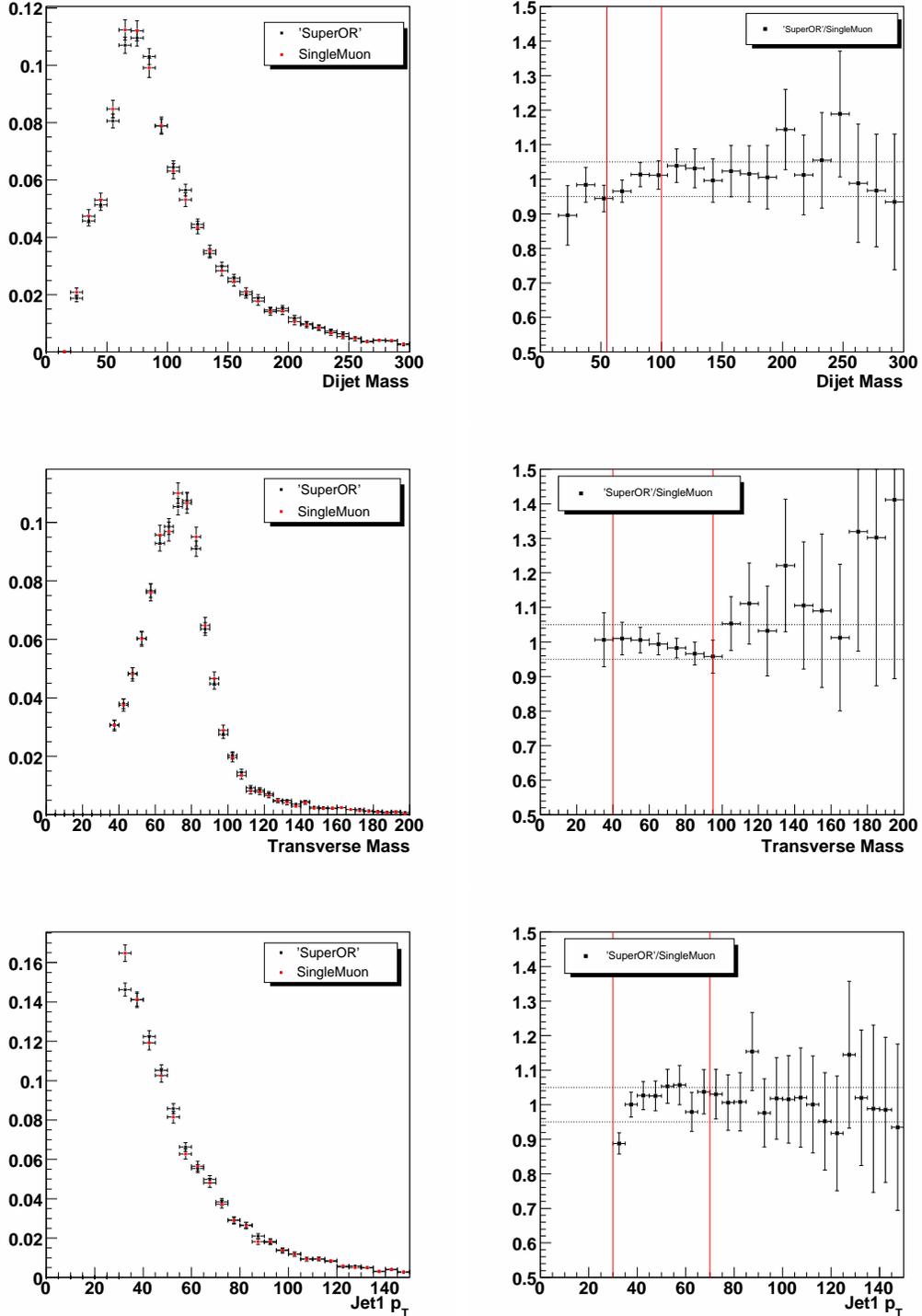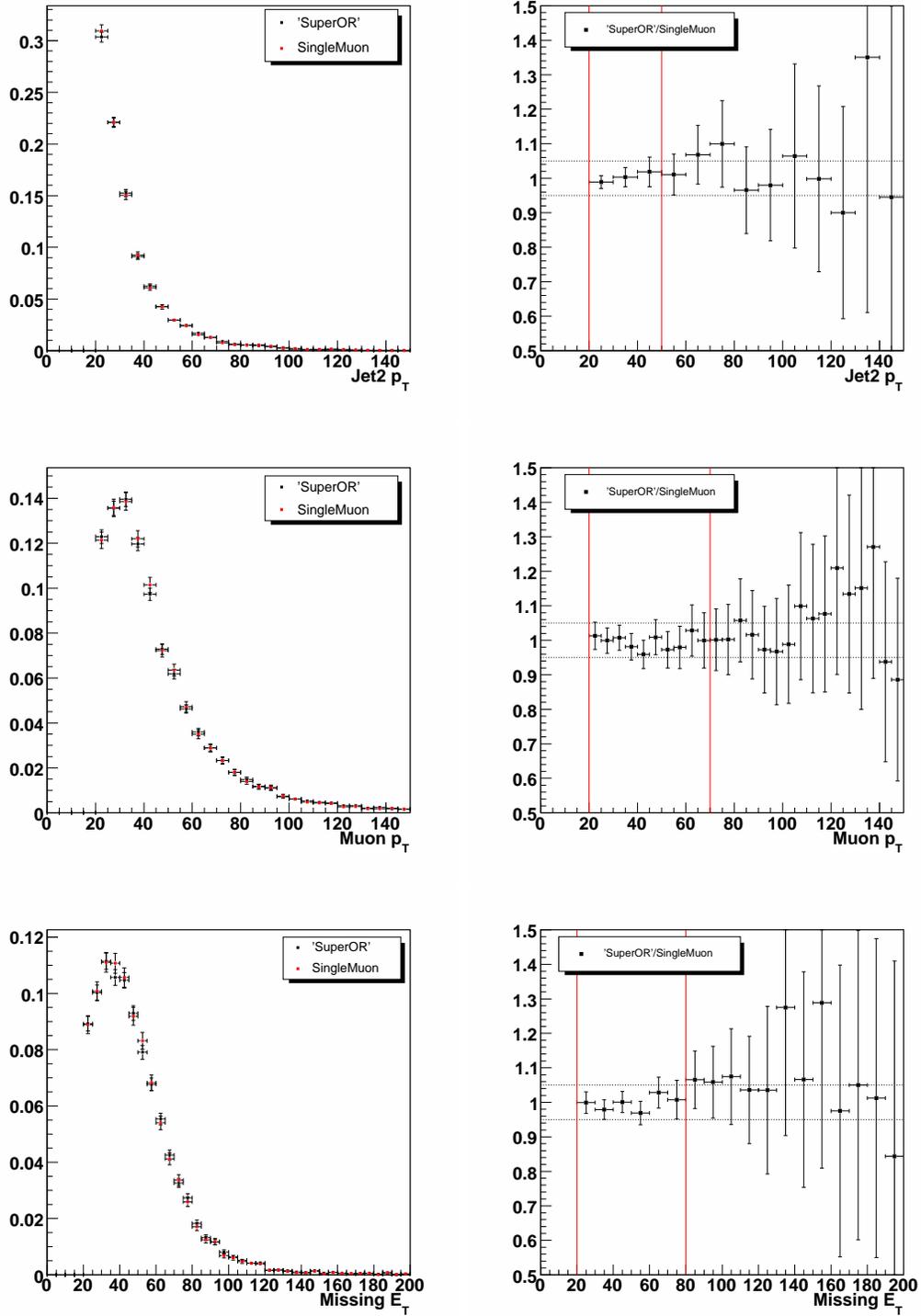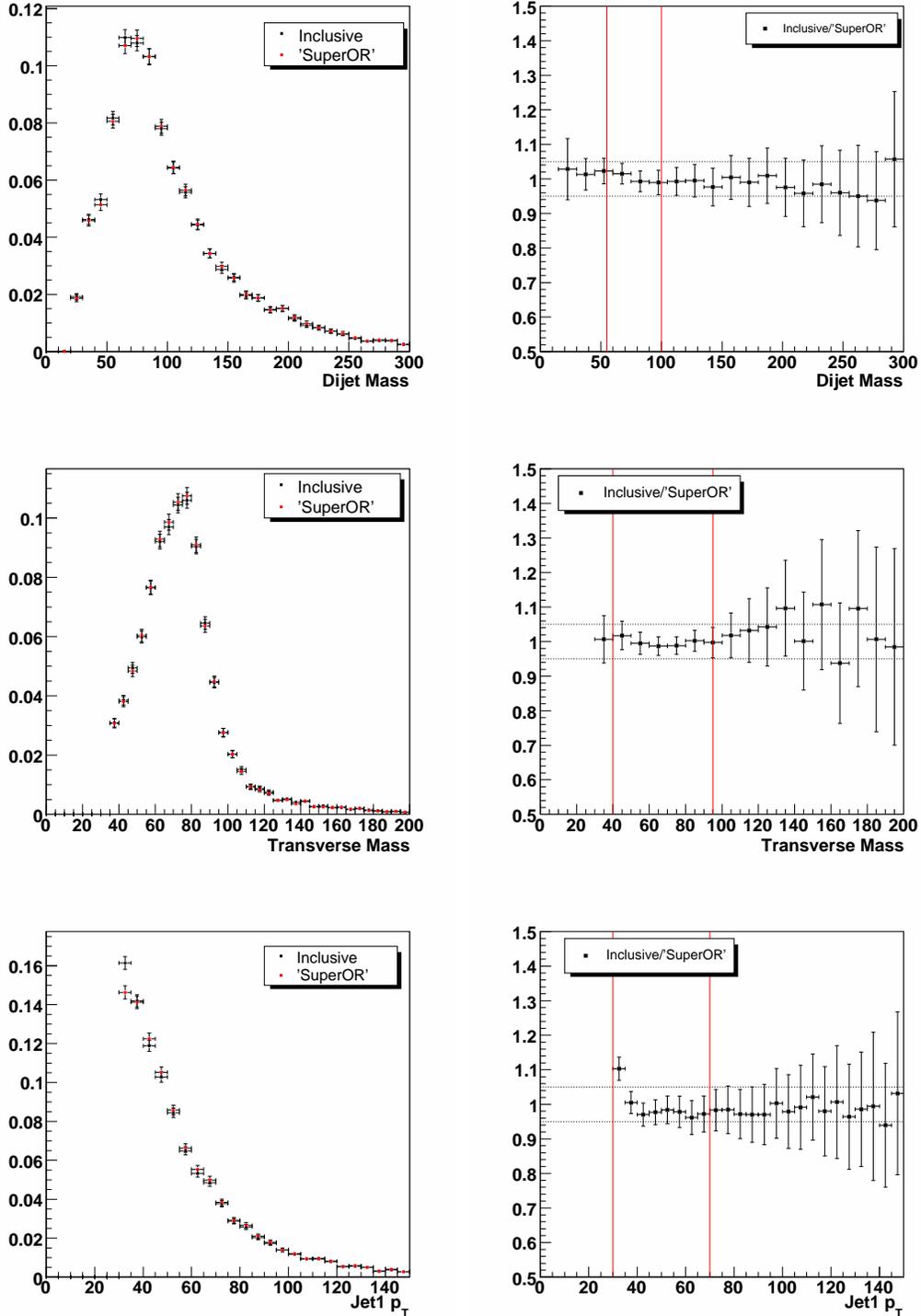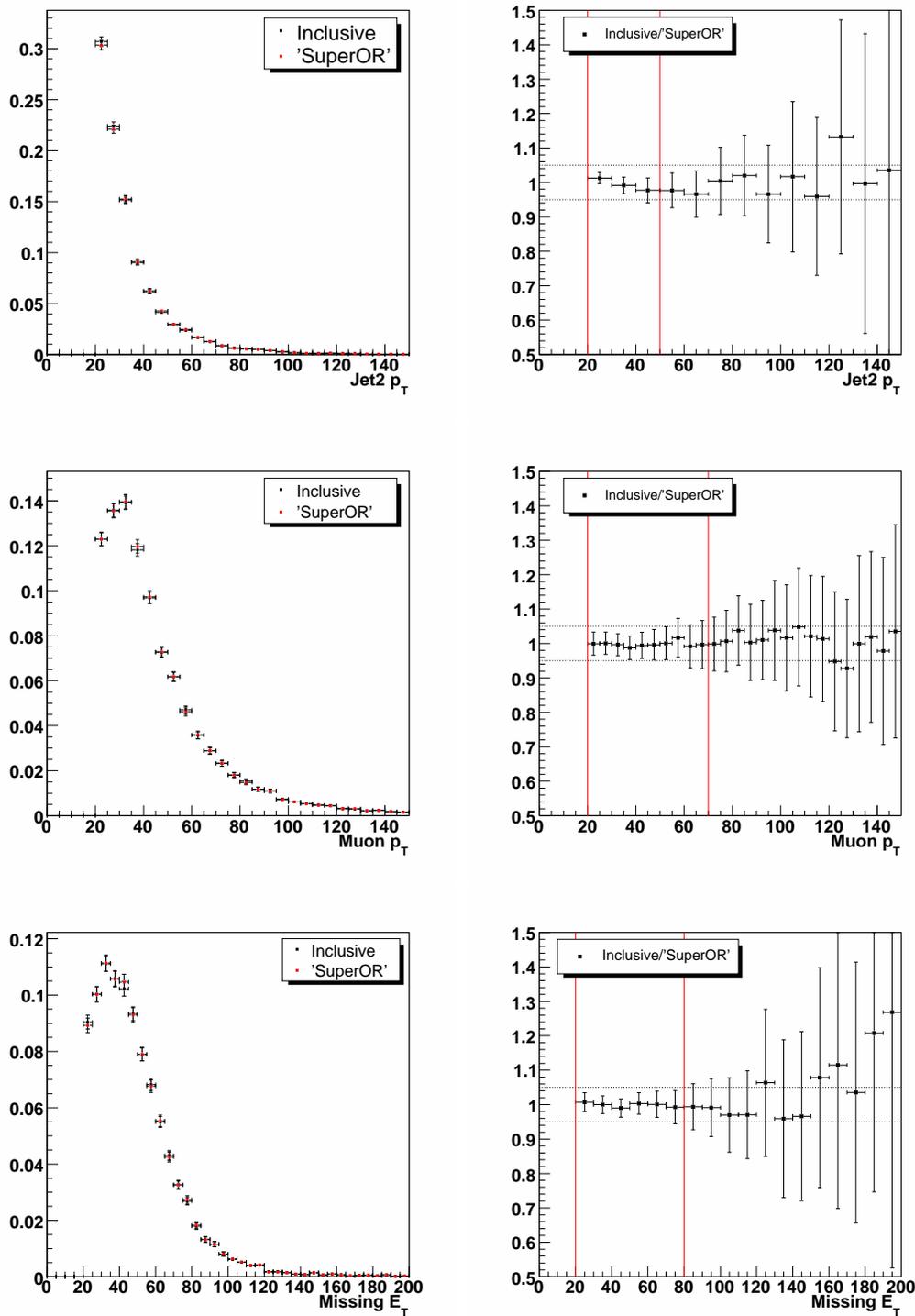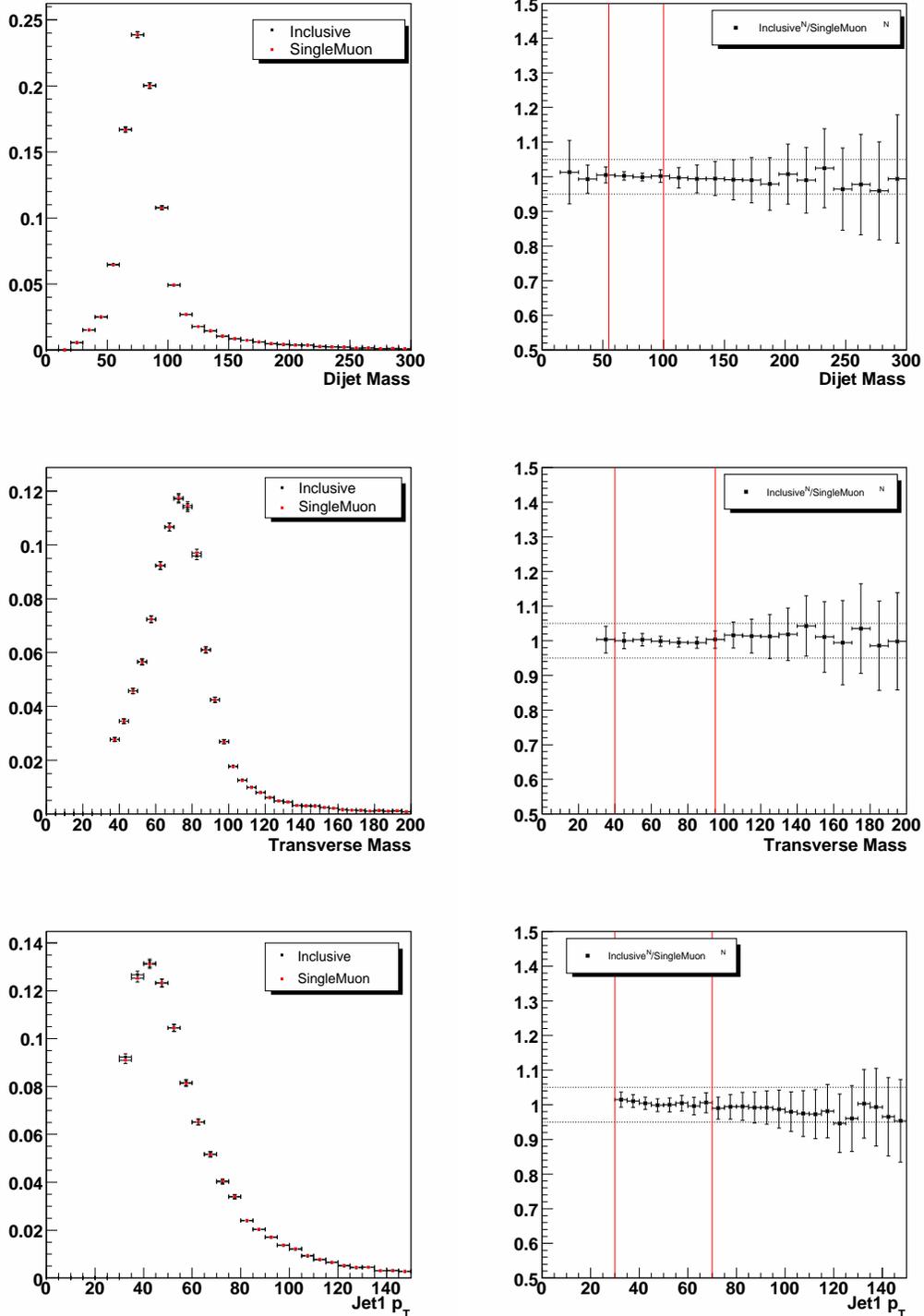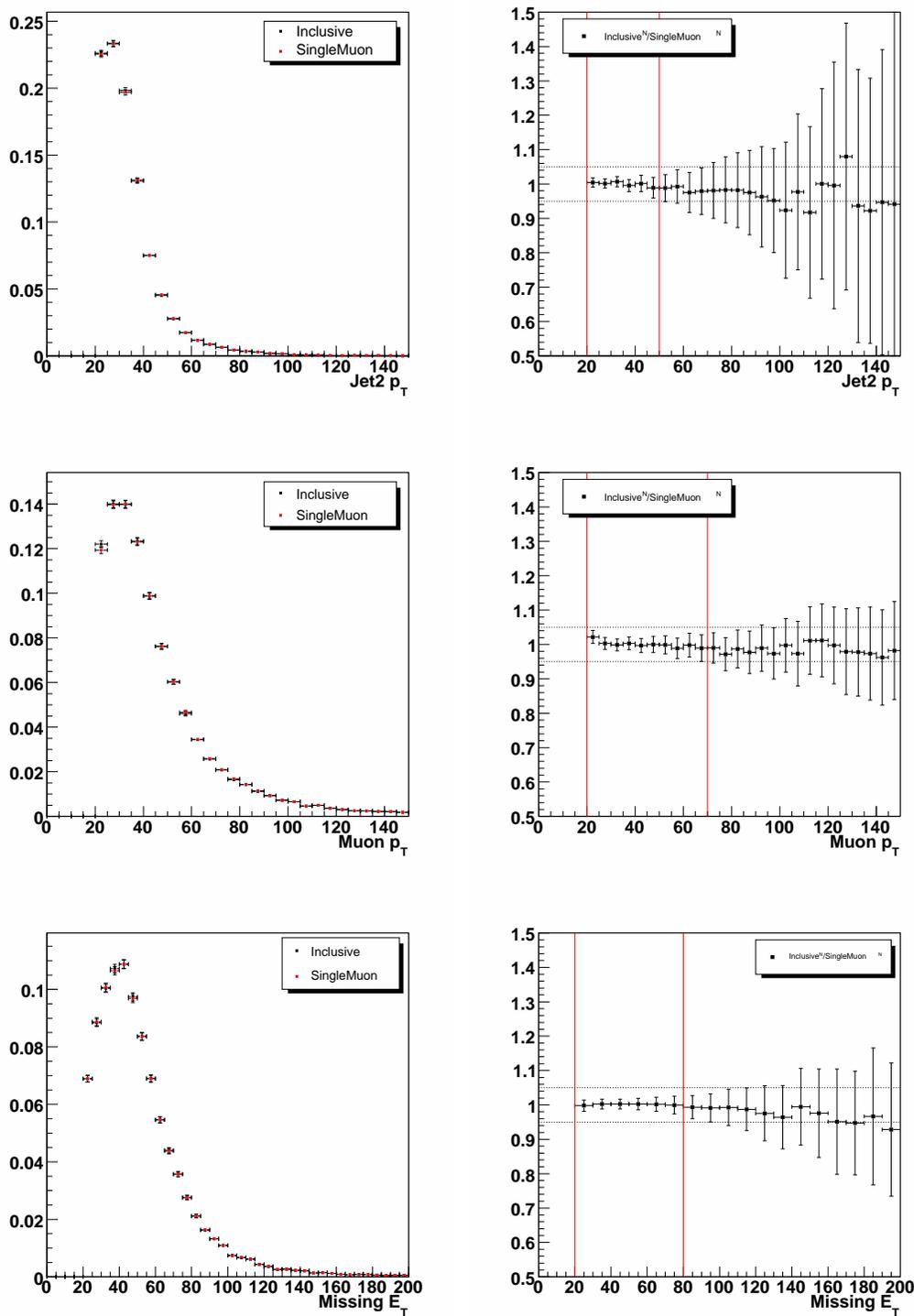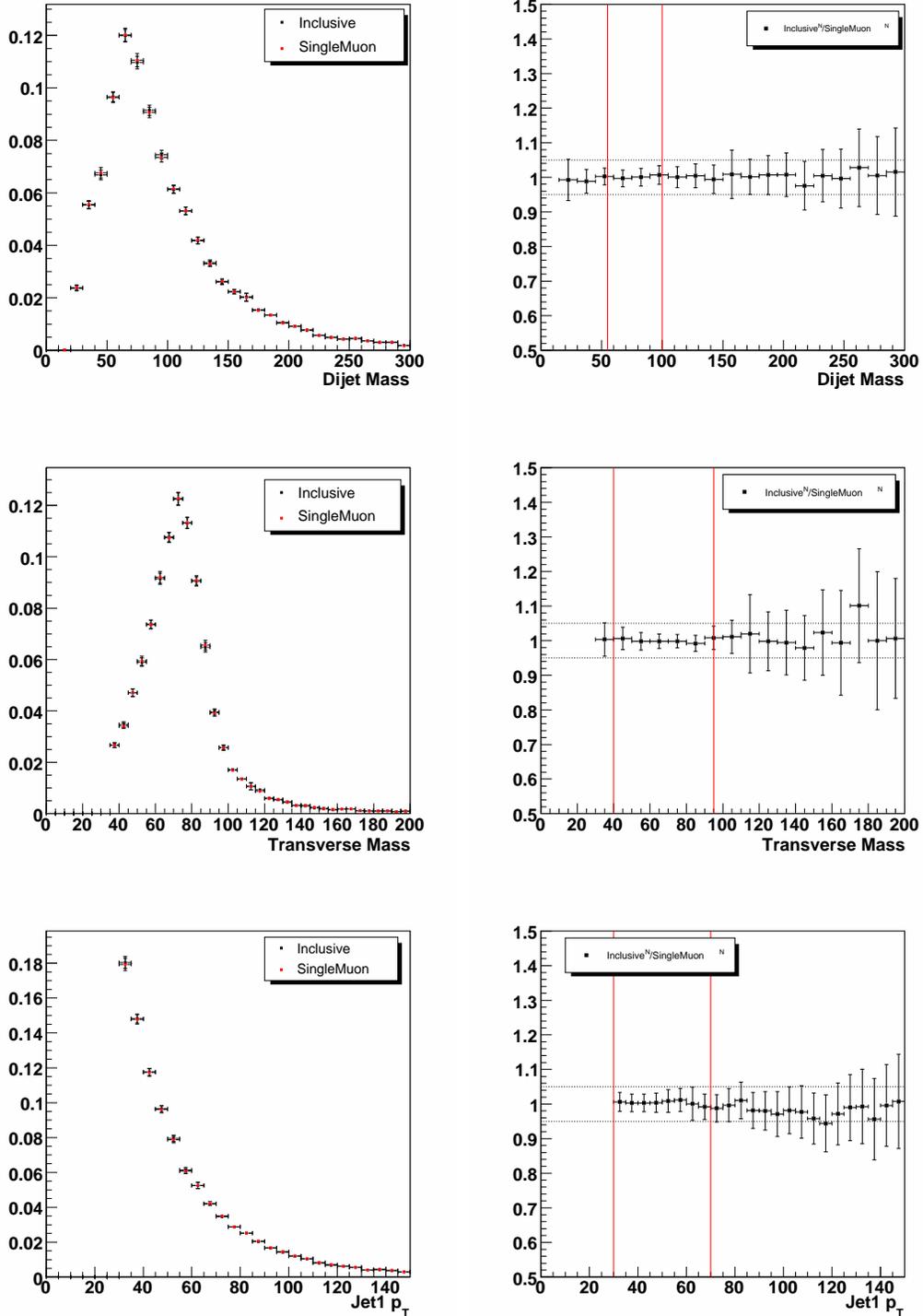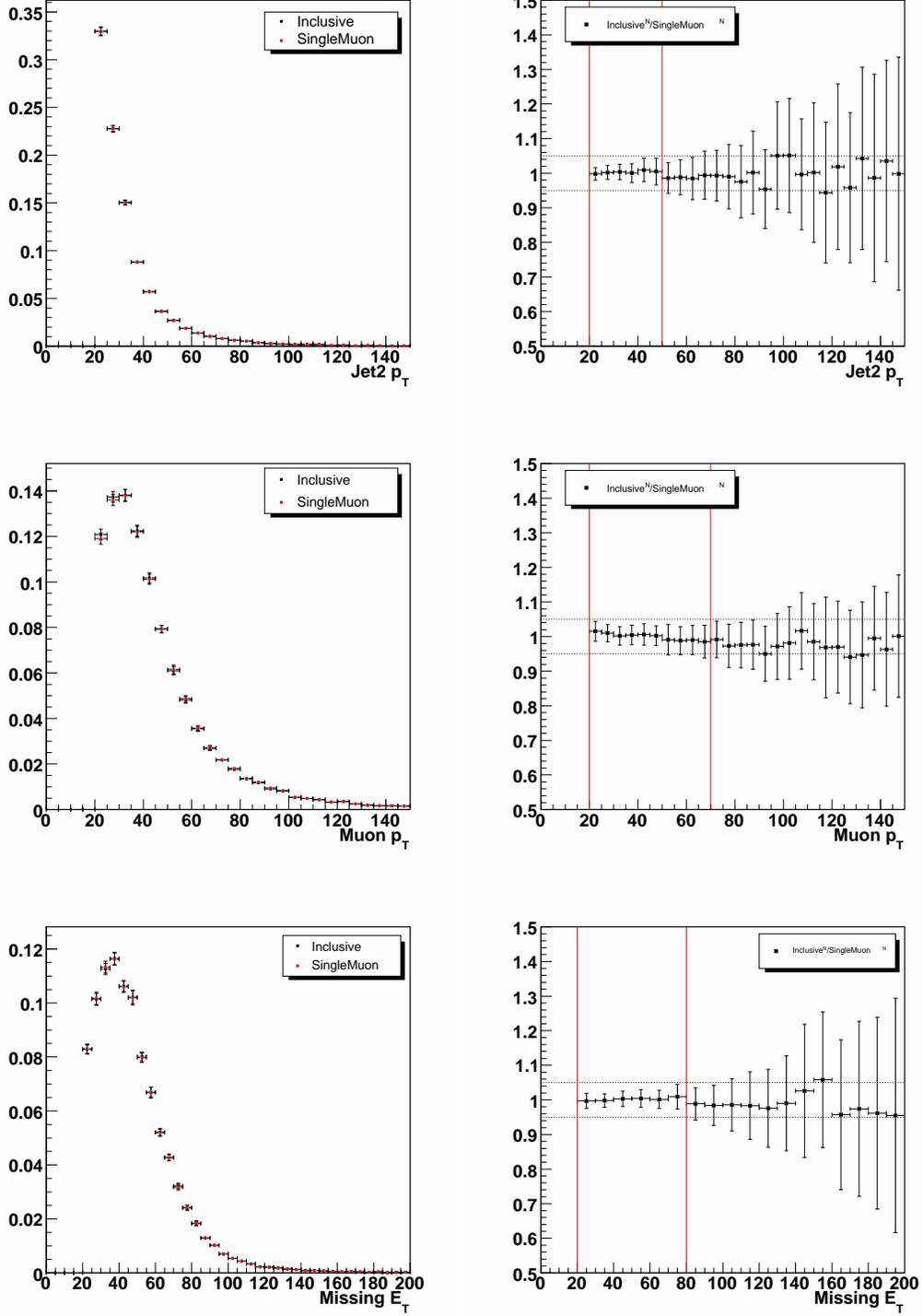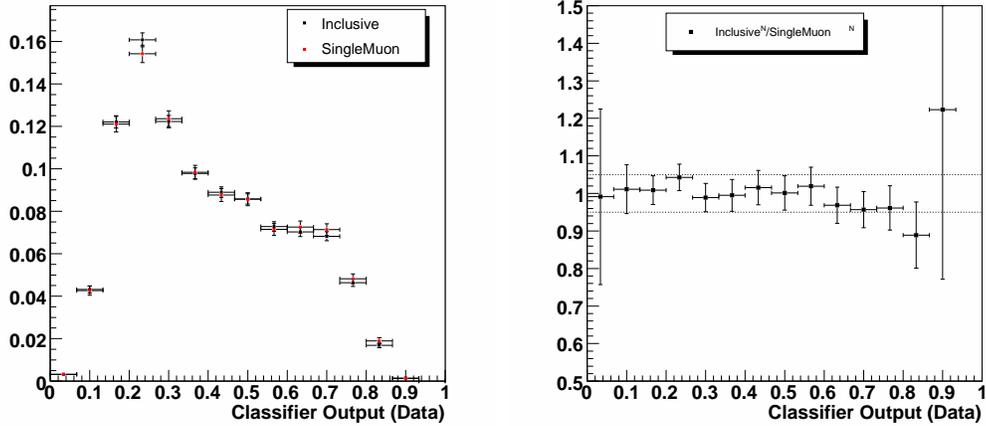
Figure A.1: *continued*
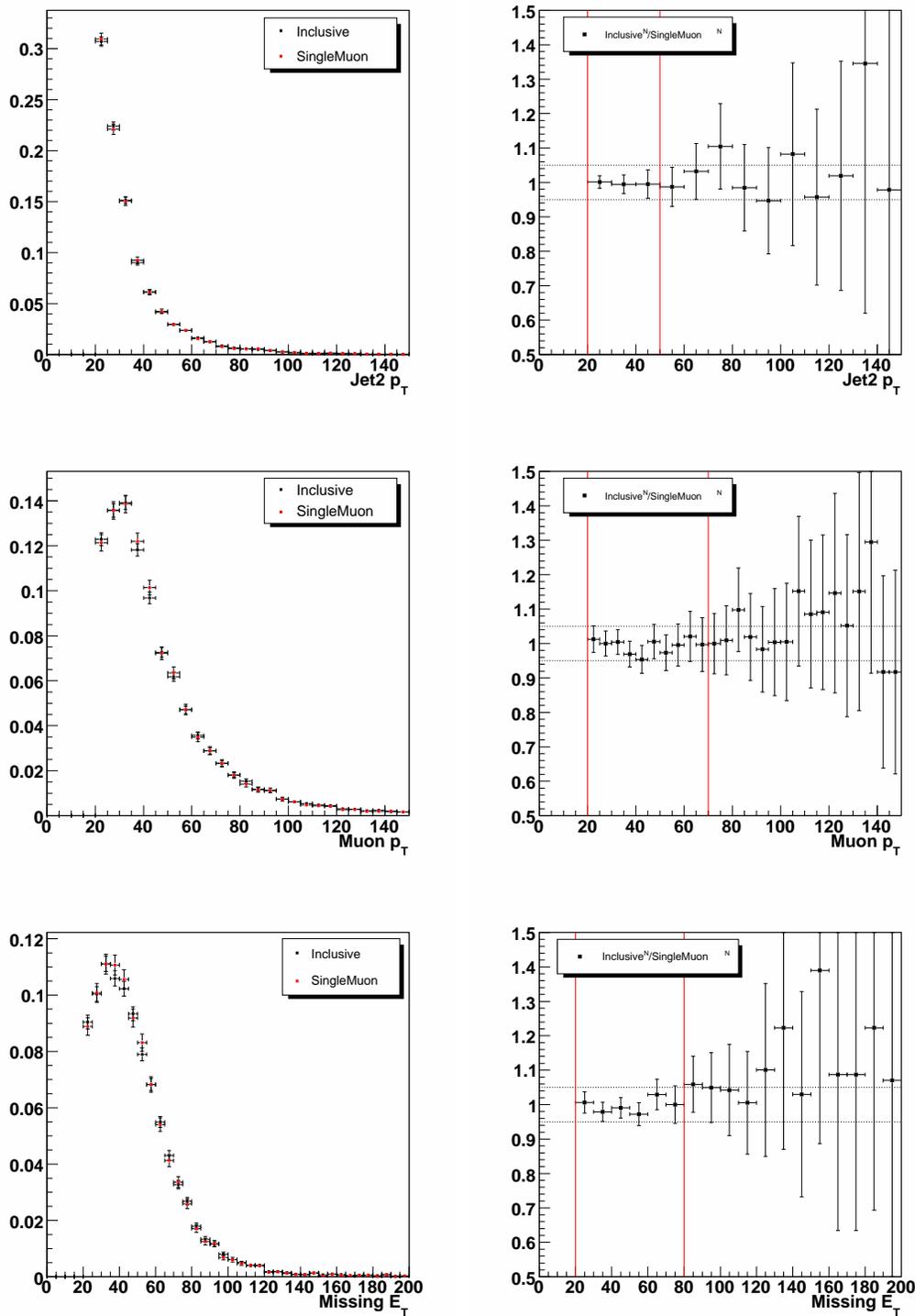
Figure A.2: Comparison between selected **data** using the single muon and muon+jet trigger suite (black) and using the single muon trigger suite (red) for various kinematic distributions (labeled on the x-axis). The plots on the left show each distribution normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the $\pm 5\%$ region and vertical red lines indicating the highest signal density region.

Figure A.2: *continued*

Figure A.3: Comparison between selected **data** using inclusive triggers (black) and using the single muon and muon+jet trigger suite (red) for various kinematic distributions (labeled on the x-axis). The plots on the left show each distribution normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the $\pm 5\%$ region and vertical red lines indicating the highest signal density region.

Figure A.3: *continued*

Figure A.4: Comparison between selected $WW + WZ$ MC events using inclusive triggers (black) and using the single muon trigger suite (red) for various kinematic distributions (labeled on the x-axis). The plots on the left show each distribution normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the $\pm 5\%$ region and vertical red lines indicating the highest signal density region.

Figure A.4: *continued*

Figure A.5: Comparison between selected W+jets MC events using inclusive triggers (black) and using the single muon trigger suite (red) for various kinematic distributions (labeled on the x-axis). The plots on the left show each distribution normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the $\pm 5\%$ region and vertical red lines indicating the highest signal density region.

Figure A.5: *continued*

Figure A.6: Comparison of the random forest output for selected **data** using inclusive triggers (black) and using the single muon trigger suite (red). The plots on the left show the distributions normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the ±5% region.

Figure A.7: Comparison of the random forest output for selected **data** using the single muon and muon+jet trigger suite (black) and using the single muon trigger suite (red). The plots on the left show the distributions normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the ±5% region.
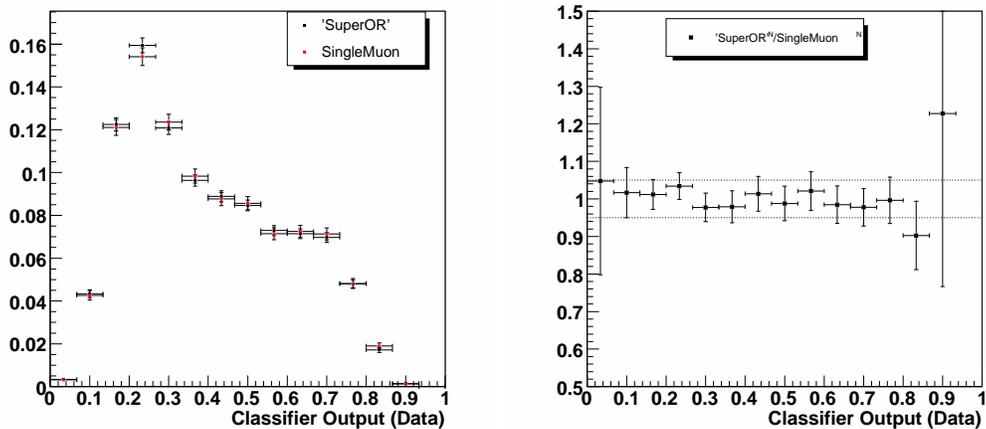
Figure A.8: Comparison of the random forest output for selected **data** using inclusive triggers (black) and using the single muon and muon+jet trigger suite (red). The plots on the left show the distributions normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the $\pm 5\%$ region.
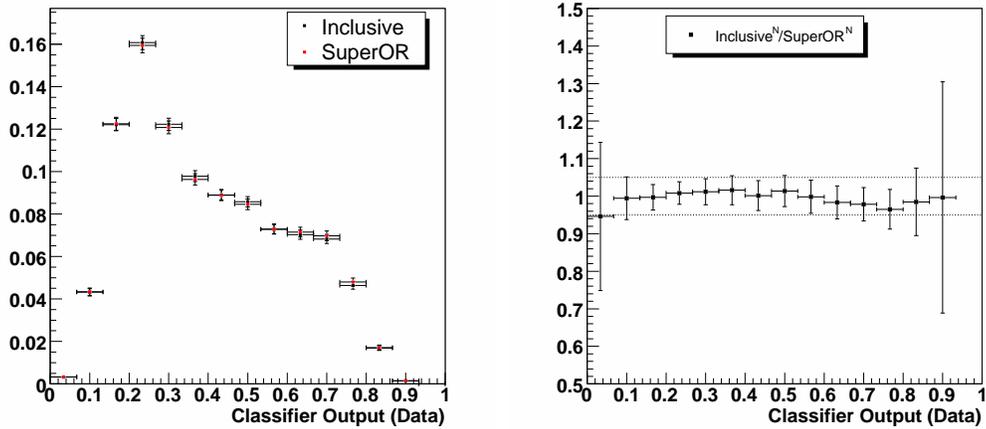
Figure A.9: Comparison of the random forest output for $WW + WZ$ and W+jets MC samples using inclusive triggers (black) and using the single muon trigger suite (red). The plots on the left show the distributions normalized to unit area and on the right is the ratio of the two distributions with dashed horizontal lines illustrating the $\pm 5\%$ region.
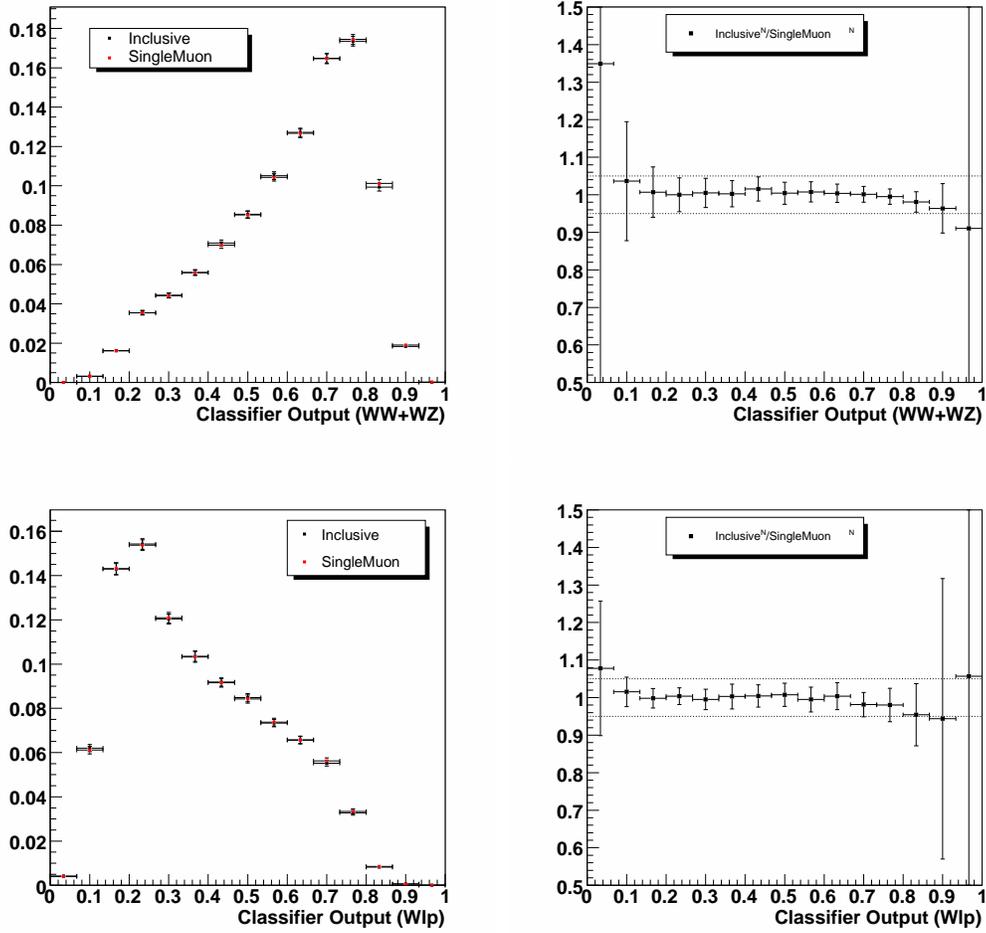
Appendix B

Study of Muon Isolation in Multijet Estimation

As explained in section 4.3.2, the $\mu\nu q\bar{q}$ multijet estimation is constructed from so-called *anti-isolated* data having the requirements CalorimeterHalo > 2.5 GeV and TrackHalo > 2.5 GeV. This appendix presents the results of a study to determine the possible shape-dependent influence of these anti-isolation requirements on the $\mu\nu q\bar{q}$ channel multijet estimation.

The influence of the CalorimeterHalo and TrackHalo (anti-)isolation cuts was studied on six kinematic variables: muon p_T, $\not{E}_T$, dijet mass, transverse W mass, second jet p_T, and $\Delta\phi$ between the muon and $\vec{\not{E}}_T$. First, the *anti-isolated* dataset was divided into two roughly equal sub-regions of TrackHalo (a "less" anti-isolated region 2.5 GeV < TrackHalo < 6 GeV and a "more" anti-isolated region 6 GeV < TrackHalo) and the shapes were compared the to the nominal multijet estimation as shown in figure B.1. Next, the shapes were compared for two different CalorimeterHalo sub-regions of the *anti-isolated* dataset (2.5 GeV < Calorimeter < 4.5 GeV and 4.5 GeV < CalorimeterHalo) as shown in figure B.2. The last two sub-regions studied (figure B.3) were a "more" anti-isolated sample defined by events in the *anti-isolated* dataset with either TrackHalo > 8.0 GeV or CalorimeterHalo > 7.0 GeV and the accompanying "less" anti-isolated sample containing the orthogonal set of events for the

anti-isolated dataset.

Within the limited statistics for the sub-regions studied, there appears to be no significant bias in the shape of the anti-isolated events for the six kinematic variables used in the determination of the $\mu\nu q\bar{q}$ channel multijet background. It is concluded that the bias in the shape due to the different isolation criteria is of order of 15% or less and the shape differences for the samples in figure B.3 are taken into account as the shape systematics assigned to the multijet estimation in the $\mu\nu q\bar{q}$ channel.

Figure B.1: Deviations relative to the nominal $\mu\nu q\bar{q}$ channel multijet shape for data that are "more" (green) and "less" (red) anti-isolated in TrackHalo. The dashed lines illustrate $\pm\,10\%$ deviations.
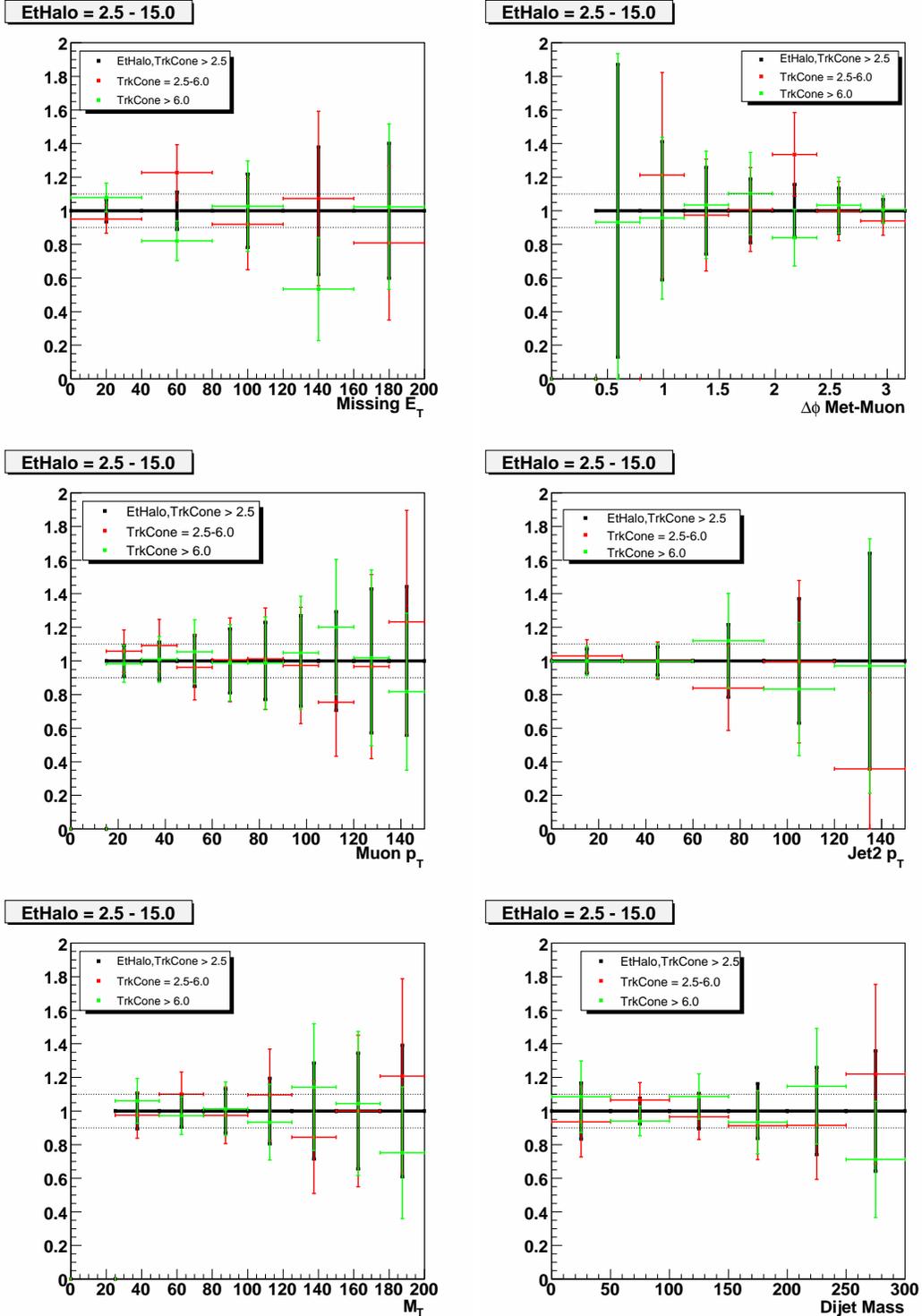
Figure B.2: Deviations relative to the nominal $\mu\nu q\bar{q}$ channel multijet shape for data that are "more" (green) and "less" (red) anti-isolated in CalorimeterHalo. The dashed lines illustrate $\pm 10\%$ deviations.
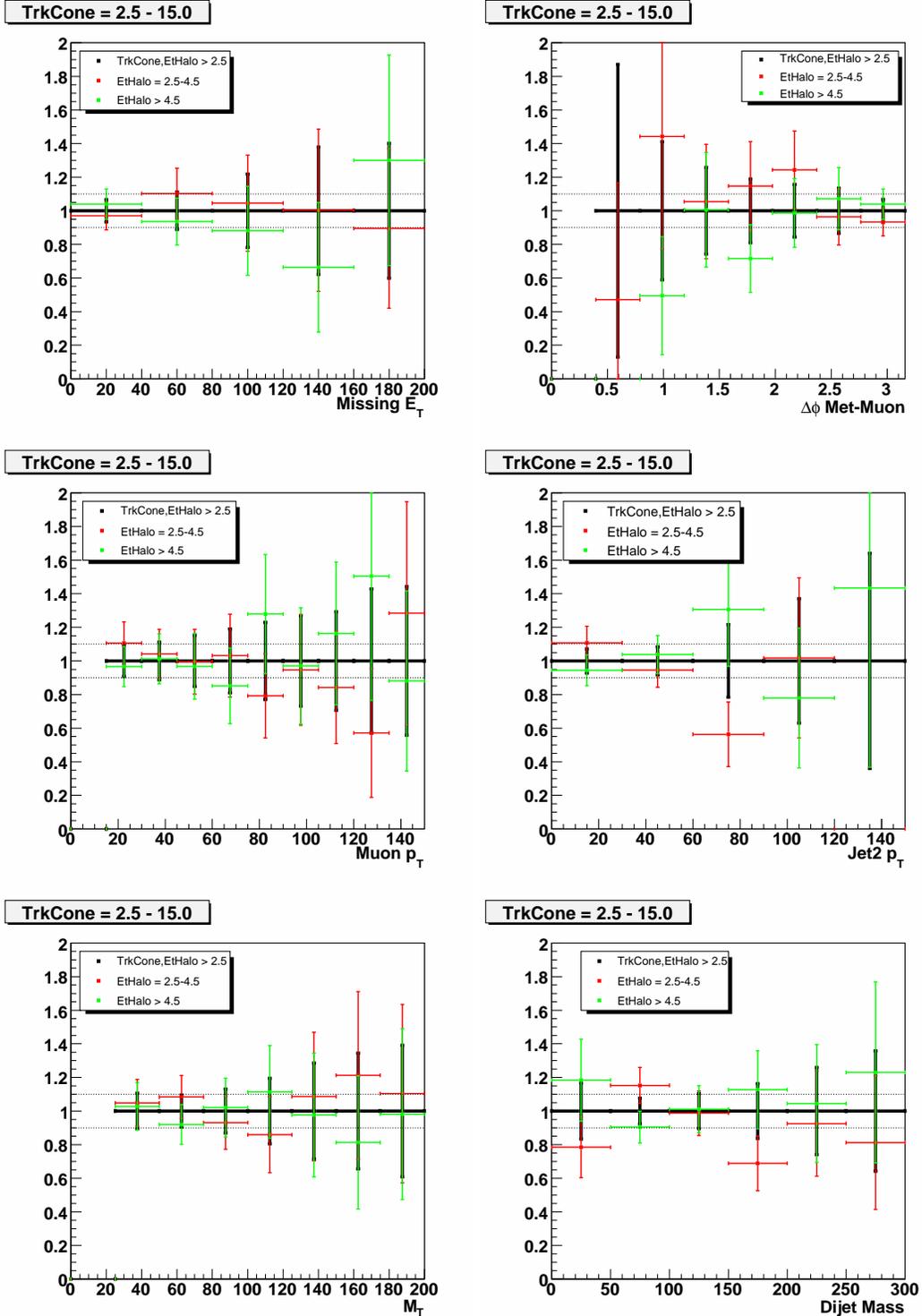
Figure B.3: Deviations relative to the nominal $\mu\nu q\bar{q}$ channel multijet shape for data that are "more" anti-isolated in either TrackHalo or CalorimeterHalo (green) and the orthogonal "less" anti-isolated data (red). The dashed lines illustrate $\pm\,10\%$ deviations.
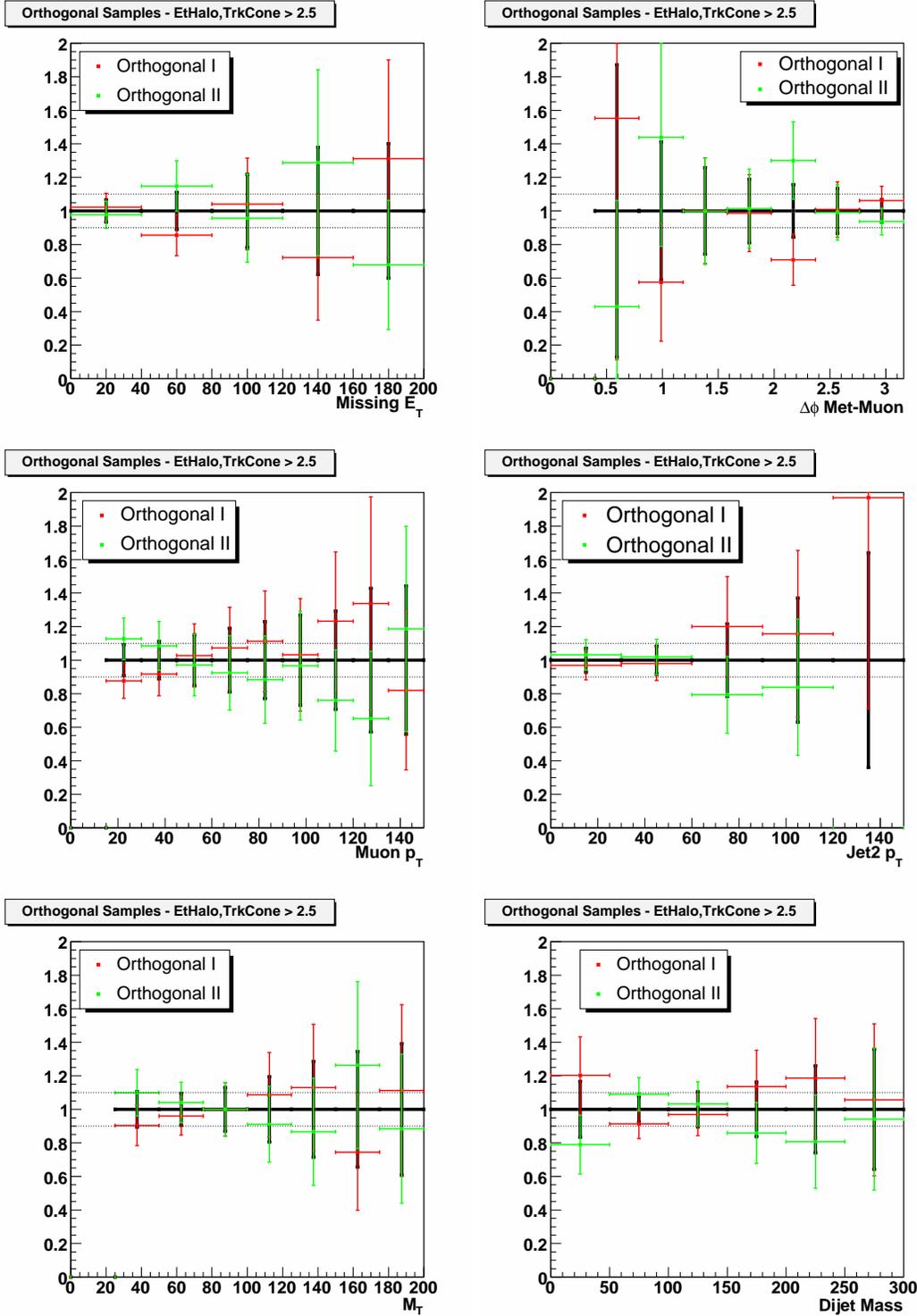
Appendix C

Effects of Jet η and ΔR Re-weighting

Here the effects of jet η and ΔR re-weighting (Section 6.7) on other kinematic distributions are shown. Distributions before and after each re-weighting step are shown in figures C.1 and C.2 for the $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ channel, respectively. Tables C.1 and C.2 list the χ^2/ndf agreement between the data and predictions for the distributions in figures C.1 and C.2, respectively.

Table C.1: The χ^2/ndf between data and prediction for distributions shown in figure C.1.

	No Re-Weighting χ^2/ndf	Only Jet η Re-Weighting χ^2/ndf	η and ΔR Re-Weighting χ^2/ndf
M(dijet)	2.31	1.74	1.30
centrality	8.45	1.29	0.79
$p_T(W)$	1.99	1.56	1.15
p_T(dijet)	2.83	2.17	1.25

Table C.2: The χ^2/ndf between data and prediction for distributions shown in figure C.2.

	No Re-Weighting χ^2/ndf	Only Jet η Re-Weighting χ^2/ndf	η and ΔR Re-Weighting χ^2/ndf
M(dijet)	3.57	2.57	1.23
centrality	8.34	1.95	1.34
$p_T(W)$	0.72	0.62	0.86
p_T(dijet)	1.82	1.41	0.86
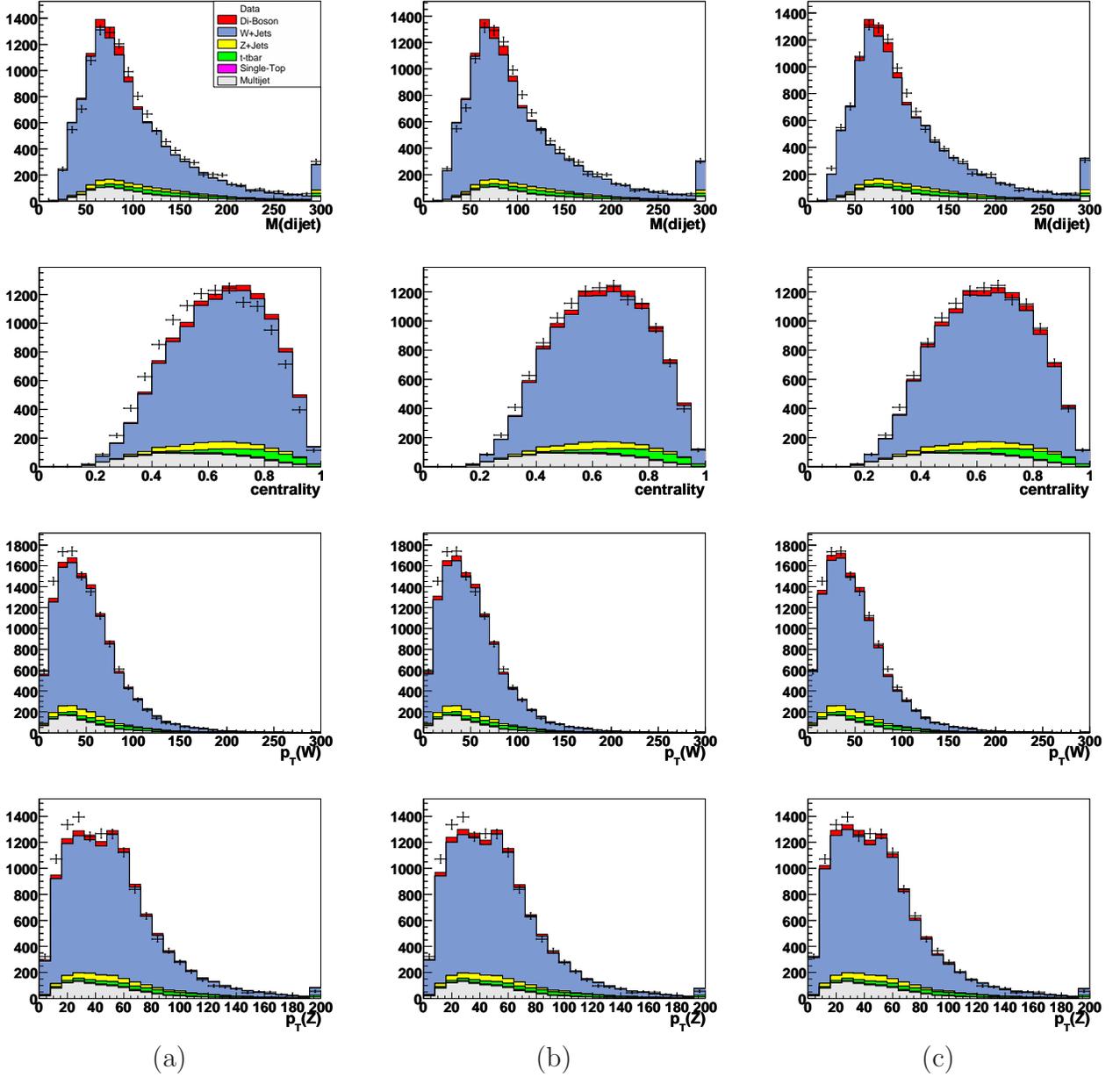
Figure C.1: Comparison of data to prediction for kinematic distributions in the $e\nu q\bar{q}$ channel (a) before re-weighting, (b) after applying leading and second jet η re-weighting, then (c) after jet η and ΔR re-weighting.
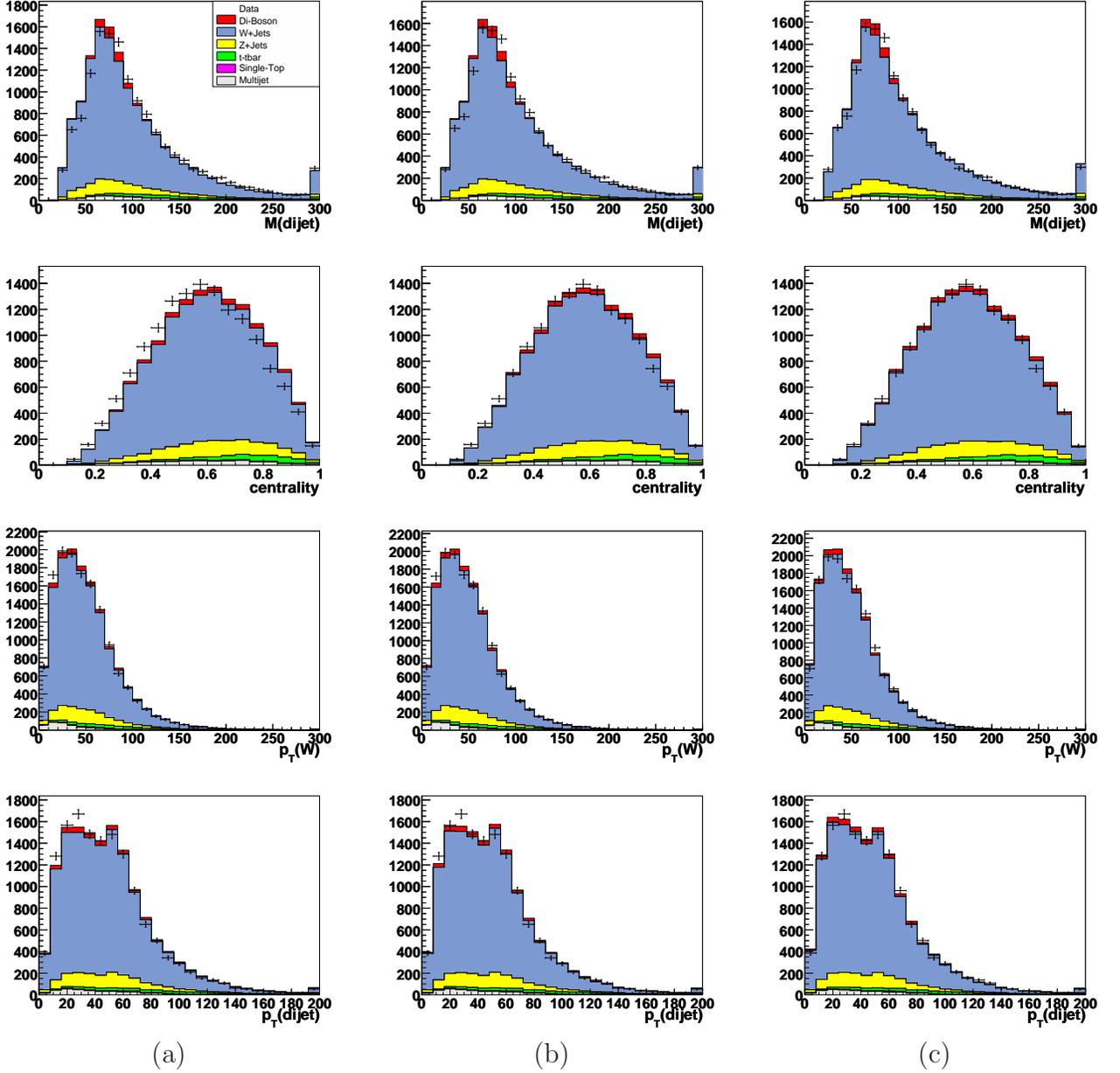
Figure C.2: Comparison of data to prediction for kinematic distributions in the $\mu\nu q\bar{q}$ channel (a) before re-weighting, (b) after applying leading and second jet η re-weighting, then (c) after jet η and ΔR re-weighting.

Appendix D

Alpgen Modeling

Comparisons between data and MC showed large differences in jet η and ΔR distributions that were corrected by re-weighting the W+jets and Z+jets MC events to data as described in section 6.7. However, there may still be residual differences between data and MC arising from the modeling of the dominant background from W+jets events. In particular, the tuning of the ALPGEN generator has yet to be demonstrated to be correct in a sample of events such as the one selected for this analysis. This appendix describes the investigation of the following parameters that were found to impact the modeling of the ALPGEN W+jets MC used in this analysis:

- Renormalization (and factorization) factor: ALPGEN uses a dynamic renormalization and factorization scale of $q^2 = (M_W)^2 + \sum(jetp_T)^2$, which can then be scaled by a constant factor (D0 default is 1.0). ALPGEN does not allow for the renormalization and factorization scales to be varied independently, so the factorization scale is varied simultaneously.

- $k_\perp$-factor(ktfac): ALPGEN determines the scale of α_S at each vertex using a $k_\perp$ measure, which can be scaled by a constant factor "ktfac" (D0 default is 1.0).

- ALPGEN parton-jet matching p_T threshold: The minimum p_T for jet clusters that are used for the MLM matching procedure [43] (D0 default is 8 GeV). Note: the parton-jet

> matching p_T threshold cannot be lower than the generator parton p_T threshold, which
> is 8 GeV.

- ALPGEN parton-jet matching radius: The radius of the jet cone ($\eta \times \phi$ space) used
  when creating clusters for the MLM matching procedure (D0 default is 0.4). Note:
  the parton-jet matching radius cannot be smaller than the generator ΔR cut between
  partons, which is 0.4.

The default values used by D0 for the renormalization/factorization factor, $k_\perp$-factor, and
parton-jet matching radius are the same as those suggested by the ALPGEN authors. How-
ever, for the generator cuts used, the default value for the parton-jet matching p_T recom-
mended by the ALPGEN authors is 13 GeV [55], while the default used by D0 is 8 GeV.

These parameters were studied using D0's alpgen_prod_x executable to generate ALP-
GEN+PYTHIA events (ALPGEN v2.11_d0 and PYTHIA v6_409) using the same generator-level
requirements as was used for the full simulation samples. Events were generated for $W+1$,
2, and 3 jets exclusively and $W+ \geq 4$ jets inclusively. The full detector simulation and recon-
struction was not performed due to the extraordinary about of time and resources required
for such a task. Instead, the events were processed with the SISCONE clustering algorithm to
construct parton-level jets from the final-state MC particles. Such jets will not be subject
to detector specific effects, however, they should provide a reasonably good approximation
of reconstructed jets, with the same being true for the leptons. All events were then passed
through the selection cuts used in this analysis. The separate jet multiplicities were com-
bined according to the relative cross sections determined by ALPGEN and normalized to
$1 fb^{-1}$ per sample.

One sample was generated with the D0 default settings (the settings used to generate the
fully simulated and reconstructed $W+$jets samples used in the analysis) and then many other
samples were generated with alternative settings. Each ALPGEN parameter was sampled over
a range of values to determine how the kinematic distributions change as the parameters are
varied. The renormalization/factorization factor was sampled at 0.5 and 2.0; $k_\perp$-factor was

sampled at 0.5 and 2.0; the parton-jet matching p_T threshold was sampled at 12, 16, and 20 GeV; and the parton-jet matching radius was sampled at 0.6 and 0.8. In order to obtain sufficient statistics, approximately 10-20 million events were generated for each sampled parameter value. Additionally, for a given parameter, *e.g.*, parton-jet matching p_T, the distributions were interpolated between the sampled values to estimate the distributions for intermediate values of that parameter.

To determine the parameter values that best describe the data, the $\Delta\chi^2$ between data and MC was measured as each parameter was varied. The following Poisson χ^2 function was used:

$$\chi^2 = -2\ln\left(\prod_i^{N_{\text{bins}}} \frac{\mathcal{L}(m_i|d_i)}{\mathcal{L}(d_i|d_i)}\right) = 2\sum_i^{N_{\text{bins}}} m_i - d_i - d_i \ln\left(\frac{m_i}{d_i}\right); \tag{D.1}$$

where $\mathcal{L}(\lambda|k) = \text{P}(k|\lambda)$ is the discrete Poisson probability for k events with a mean value of λ events, d_i is the number of data in the ith bin, m_i is the number of predicted events in the ith bin, and N_{bins} is the number of bins. The data was compared to the fully simulated and reconstructed MC after it was re-weighted according to the variations determined from the ALPGEN+PYTHIA samples generated for these studies. In other words, for a given distribution, the difference between the D0 default settings (the nominal shape) and some variation (the shifted shape) was determined from the aforementioned ALPGEN+PYTHIA samples and then that difference was applied to the fully simulated and reconstructed W+jets samples for comparisons to data. Because a correction was already applied to achieve agreement between MC and data for the ΔR between jets, any ΔR differences between the nominal and shifted ALPGEN+PYTHIA samples was removed (via a re-weighting of the shifted events) before evaluating the shape changes for other distributions. Jet η differences between samples were found to be negligible and were ignored. It should be noted, however, that these ΔR differences (figure D.1) were considerably smaller than the ΔR correction already applied.

The change in the $\Delta\chi^2$ by varying each parameter was first measured using the dijet mass distribution. These plots are shown in the left column of figure D.2. The data prefers a parton-jet matching radius near the default value of 0.4. The preferred values for the
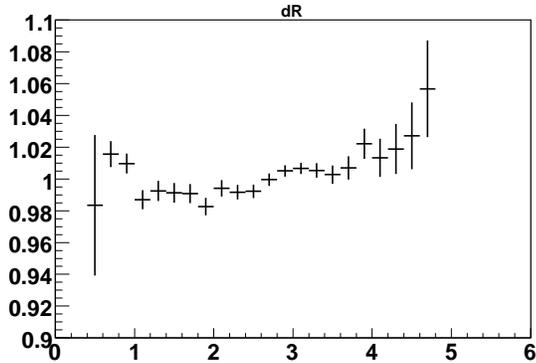
Figure D.1: Fractional difference for the distribution of ΔR between jets for a parton-jet matching p_T of 13.2 versus nominal. This difference was re-weighted away before determining the shape differences for other distributions.

renormalization/factorization factor and $k_\perp$-factor were near or somewhat below 1.0. The parton-jet matching p_T threshold appears to match data best for a value in the range 12-16 GeV.

In order to remove the possible bias from a non-SM rate of signal events, all events were removed that had a dijet mass in the window around the signal peak (55 GeV < m_{jj} <110 GeV). With these events removed, the p_T distributions of the leading jet and of the recoiling $W \rightarrow \ell\nu$ were used, comparing the spectra found in data to those obtained by varying the ALPGEN parameters. The middle and right columns of figure D.2 show the change in the χ^2 for these distributions as each parameter was varied. Not surprisingly, these variables exhibited a similar behavior to what was seen for the dijet mass. An exception was the parton-jet matching radius variation for the W boson p_T, in which the muon channel seems to prefer a value away from the nominal. However, the more interesting trend was the preference for a parton-jet matching p_T threshold larger than the D0 default of 8 GeV.

Next was a simultaneous $\Delta\chi^2$ measurement using both the p_T distribution of the leading jet and of the recoiling W boson. These tests were performed as a simple sum of the χ^2 values from the individual tests and are shown in figure D.3. This combination has the effect of reducing random statistical effects, strengthening a common minimum, averaging out incoherent behaviors, and broadening the resolution of any near minima. The resulting
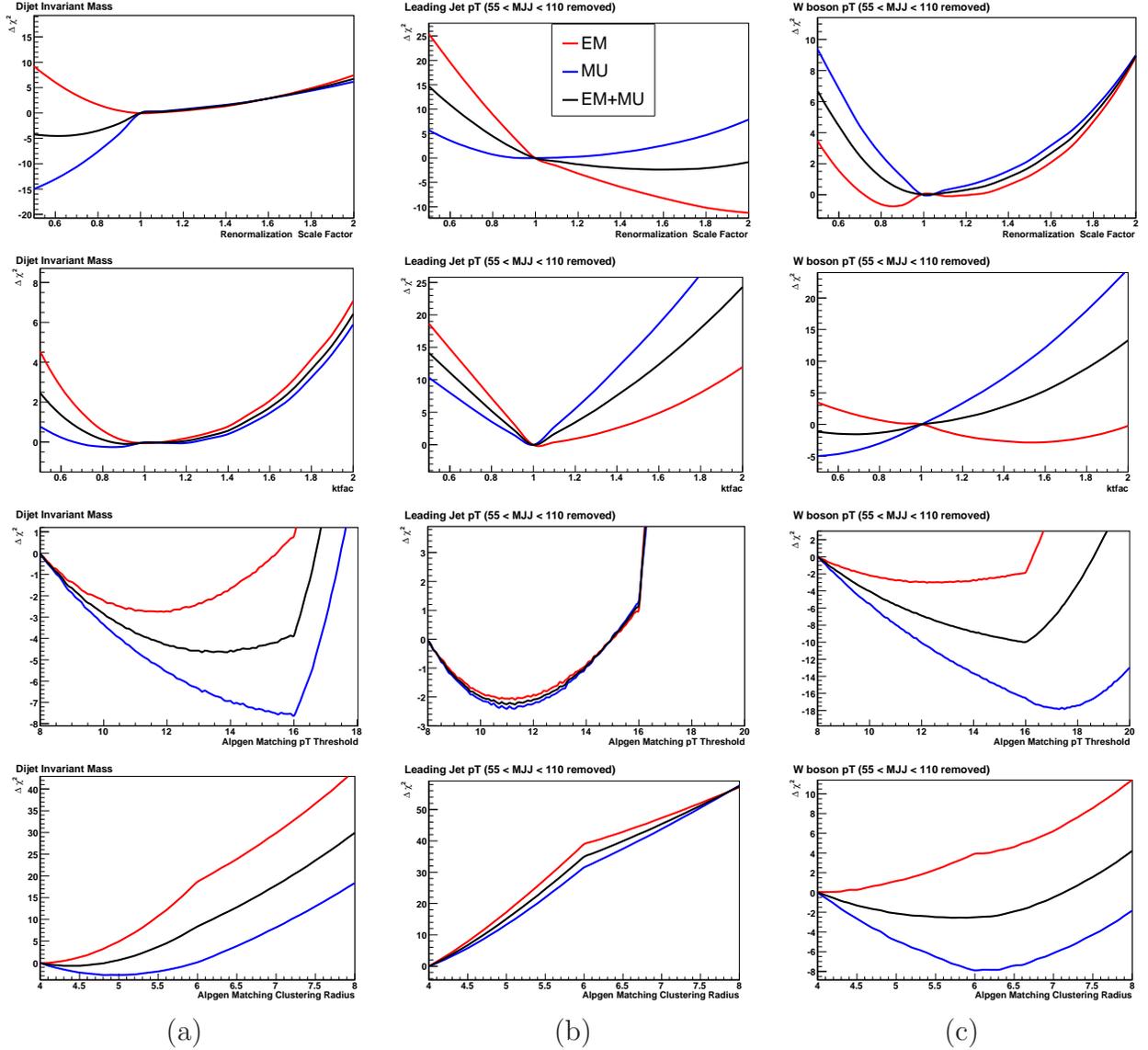
Figure D.2: Change in χ^2 between data and MC as each ALPGEN parameter was varied when comparing (a) the dijet mass, (b) the leading jet p_T, or (c) the W boson p_T distributions. Comparisons using the leading jet p_T and leptonic W boson p_T have only events outside of the dijet mass region 55 GeV $< m_{jj} <$ 110 GeV.
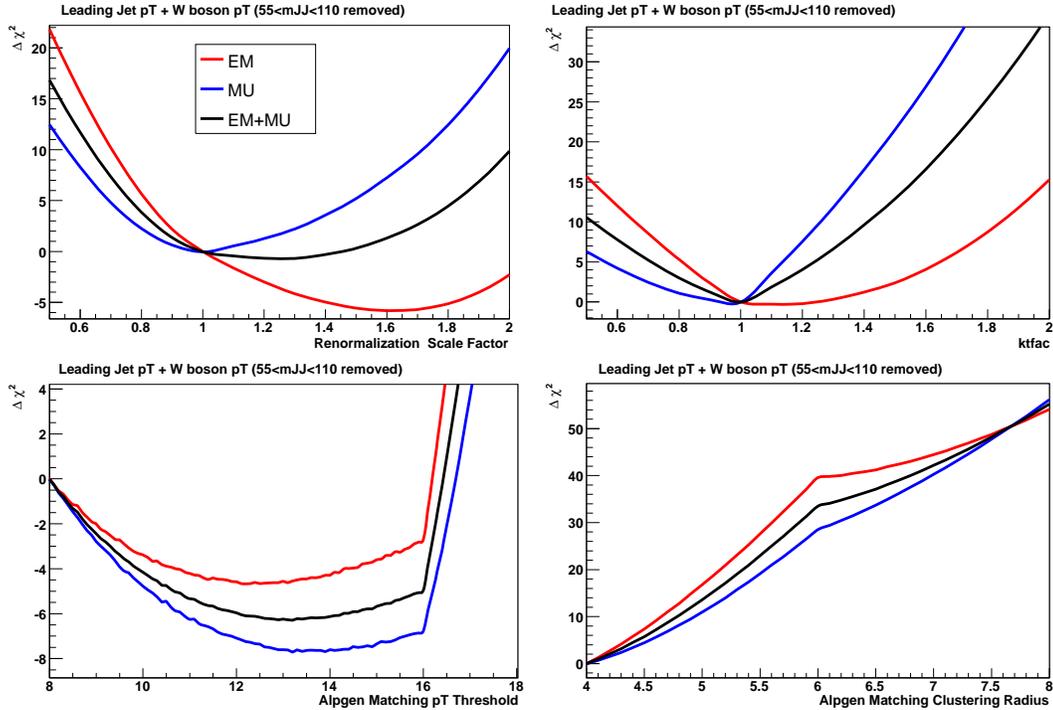
Figure D.3: Change in χ^2 between data and MC as each ALPGEN parameter was varied and simultaneously comparing the leading jet p_T and leptonic W boson p_T distributions for evens outside of the dijet mass region 55 GeV $< m_{jj} <$ 110 GeV.

$\Delta\chi^2$ distributions indicate that the only variable with a minimum that was significantly different from the D0 default was the parton-jet matching p_T threshold. As all three variables (dijet mass, leading jet p_T, and W boson p_T) exhibited a similar behavior for the parton-jet matching p_T threshold, the fully simulated and reconstructed ALPGEN W+jets MC was re-weighted to the parton-jet matching p_T threshold value found to give the minimum χ^2 in the simultaneous $\Delta\chi^2$ test (figure D.3). This value was found to be 13.2 GeV, which agrees well with the minimum value from using the dijet mass (13.5 GeV) and also appears to agree well with the value suggested to be used by the ALPGEN authors (13 GeV).

Re-weighting of the ALPGEN W+jets MC samples to a parton-jet matching p_T threshold of 13.2 GeV was done using the dijet mass shape change shown in figure D.4. Also shown in figure D.4 is the resulting shape change propagated to the random forest classifier. Compared to the dijet mass, the random forest classifier was less sensitive to the change in ALPGEN parton-jet matching p_T threshold. This is not too surprising, the effect on the dijet mass

(a) (b)
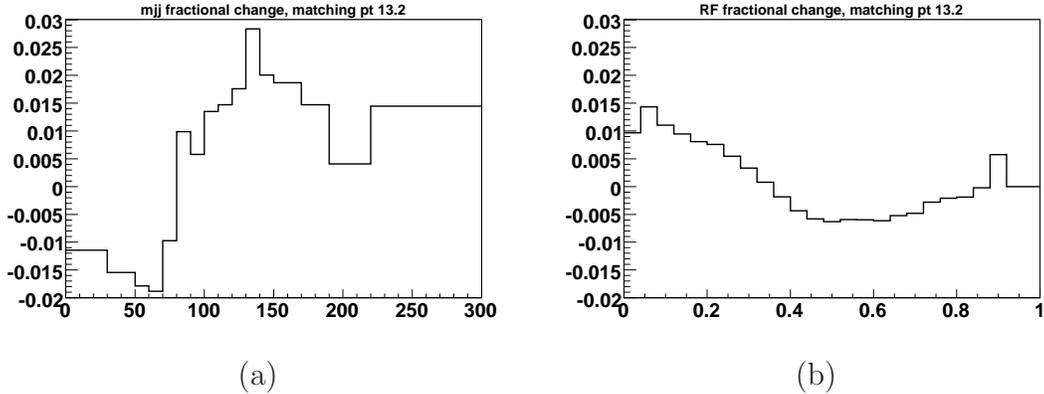
Figure D.4: Fractional shape change between using the D0 default parton-jet matching p_T threshold of 8 GeV and the preferred value of 13.2 GeV for the distributions of (a) dijet mass and (b) random forest output.

was to increase the number of events with very low dijet mass and decrease the number with high dijet mass; however, both high and low dijet mass will result in a low random forest output values and the effects somewhat cancel.

After re-weighting the ALPGEN W+jets MC to the preferred value for parton-jet matching p_T (giving the new "nominal" W+jets model), the $\Delta\chi^2$ between data and MC was then used to determine the uncertainties associated with each parameter; these plots are shown in figure D.5. The uncertainty on each parameter was determined by the points where the $\Delta\chi^2$ increased by 2.0 from the minimum value (2.0 instead of 1.0 because the χ^2 is actually the sum of two χ^2 from distributions with the same data). For example, the parton-jet matching p_T range of $[10.0, 16.2]$ covers the potential outcomes observed in the tests at the level of 1-σ. The 1-σ uncertainty ranges found for the other parameters were: $[0.9, 1.8]$ for the renormalization/factorization factor, $[0.9, 1.2]$ for $k_\perp$-factor, and $[0.35, 0.45]$ for parton-jet matching radius (only the upward uncertainty for the parton-jet matching radius can be determined, thus the downward uncertainty was assumed to be equal and opposite to the upward uncertainty). The resulting shape uncertainties on the random forest output are shown in figure D.6 and were assigned as systematic uncertainties for renormalization/factorization scale and ALPGEN parton-jet matching parameters.
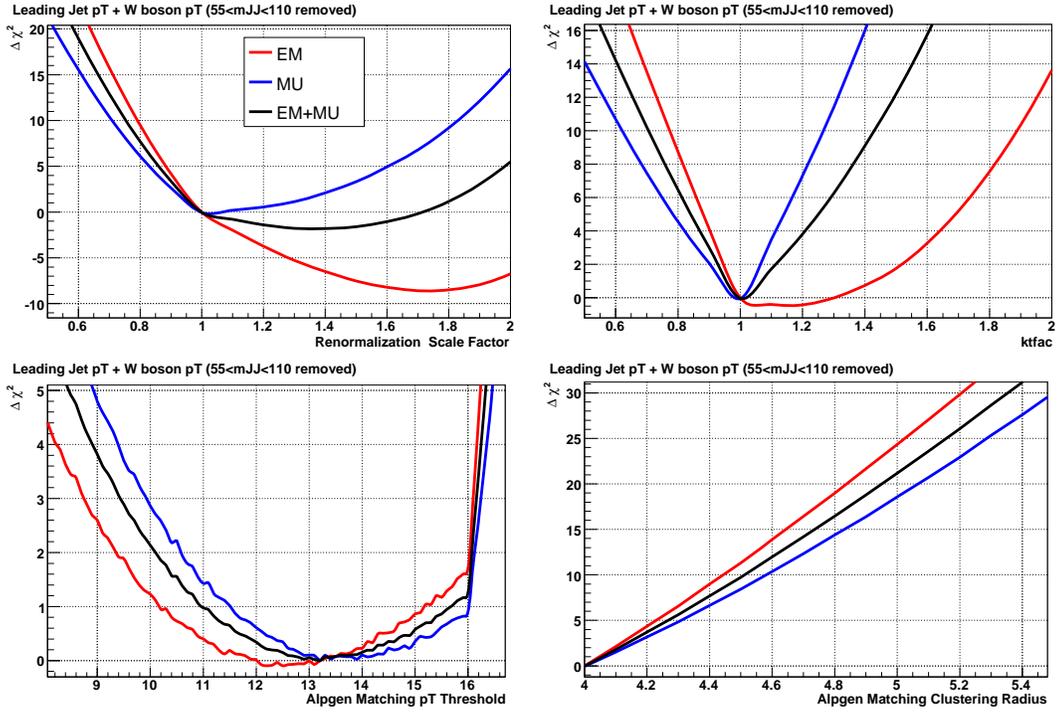
Figure D.5: Change in χ^2 between data and MC as each ALPGEN parameter was varied, but after the W+jets samples have been re-weighted to a "nominal" parton-jet matching p_T threshold of 13.2 GeV. The χ^2 was calculated using the simultaneous comparison of the leading jet p_T and W boson p_T distributions with events outside of the dijet mass region 55 GeV $< m_{jj} <$ 110 GeV.
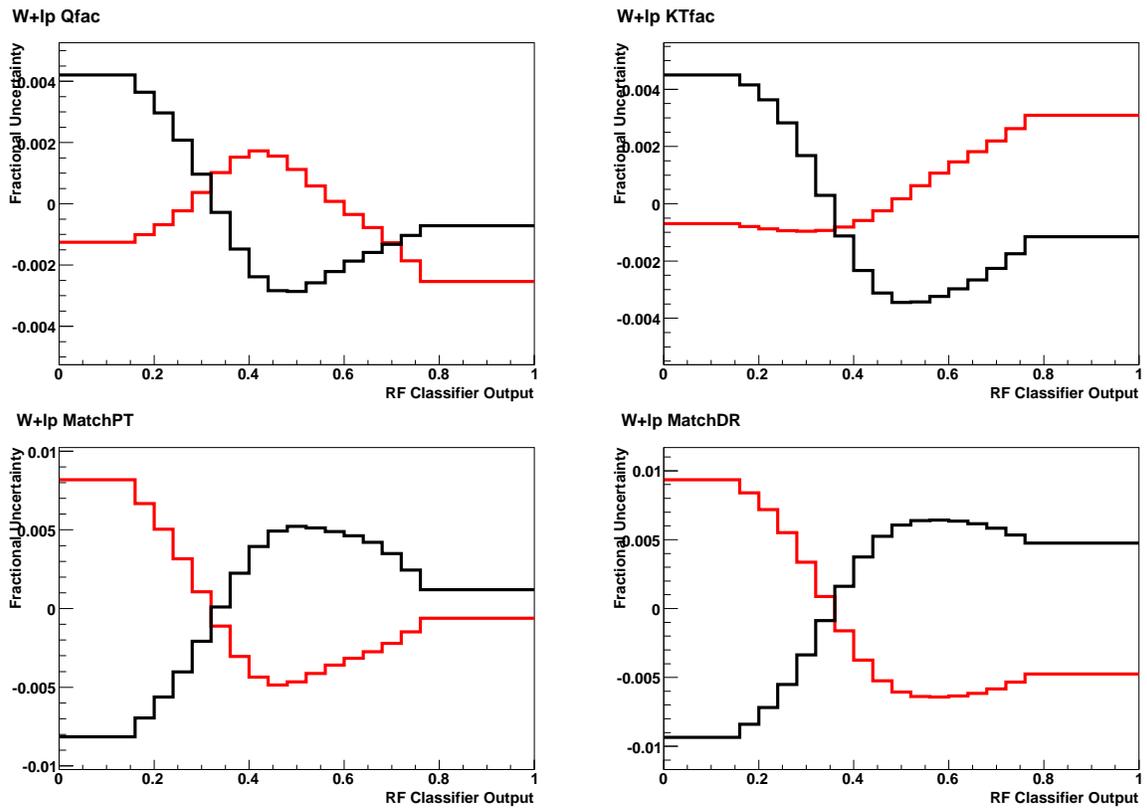
Figure D.6: Uncertainties due to ALPGEN parameters as a function of random forest output.

Appendix E

Additional Plots of Signal NLO Correction

This appendix contains additional plots comparing the kinematic distributions for diboson events generated with the PYTHIA and MC@NLO MC generators before and after applying the re-weighting described in Section 6.8.

Figure E.1: Plots of the following distributions before and after NLO re-weighting: (a,b) p_T of the diboson system (linear and log scale), (c,d) p_T of each boson (linear and log scale), (e,f) Δp_T between the bosons (linear and log scale), (g) η of the diboson system, (h) η of each boson, (i,j) ΔR between the bosons (linear and log scale).
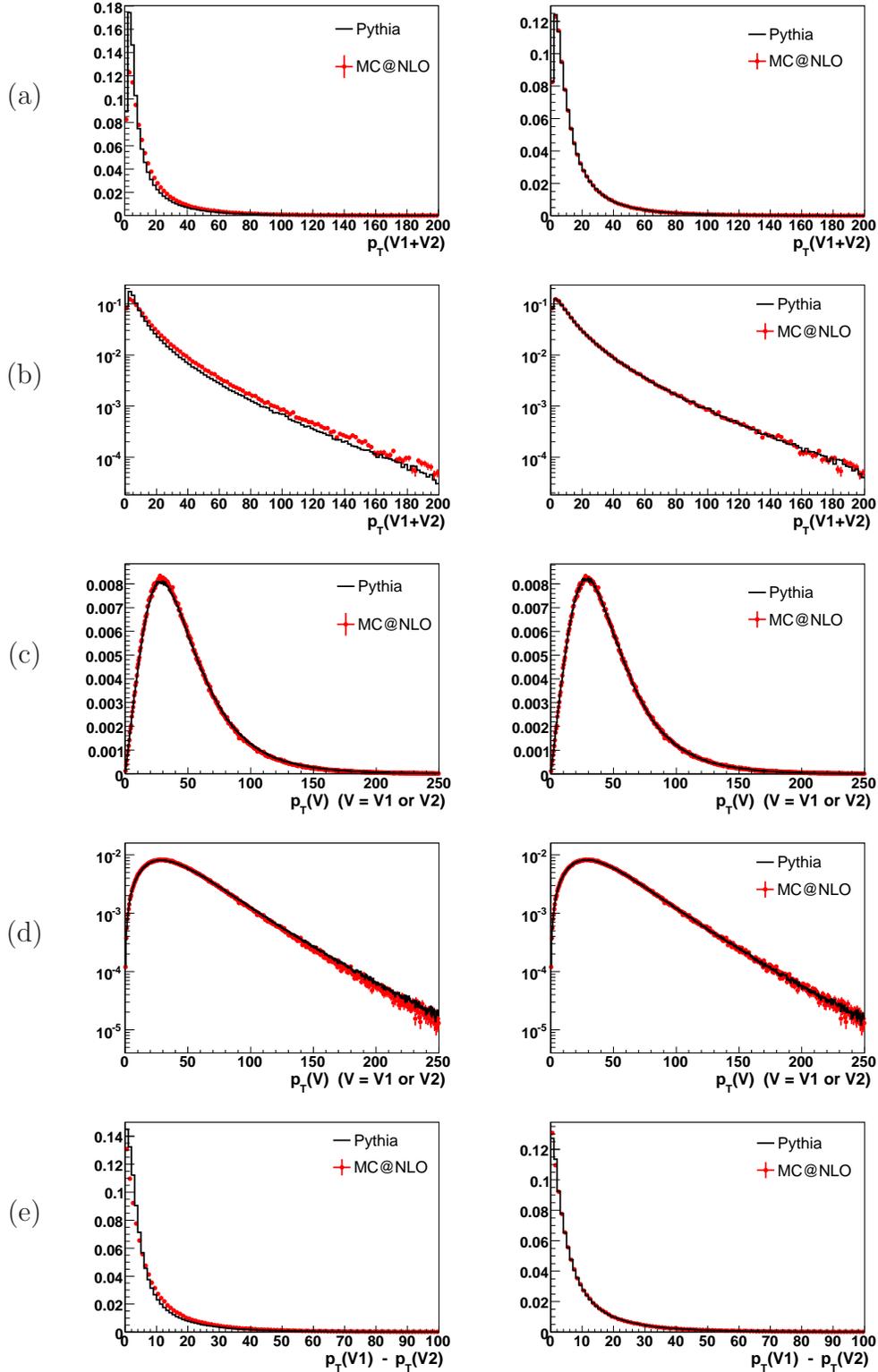
Figure E.1: *continued*

(f)

(g)

(h)

(i)

(j)
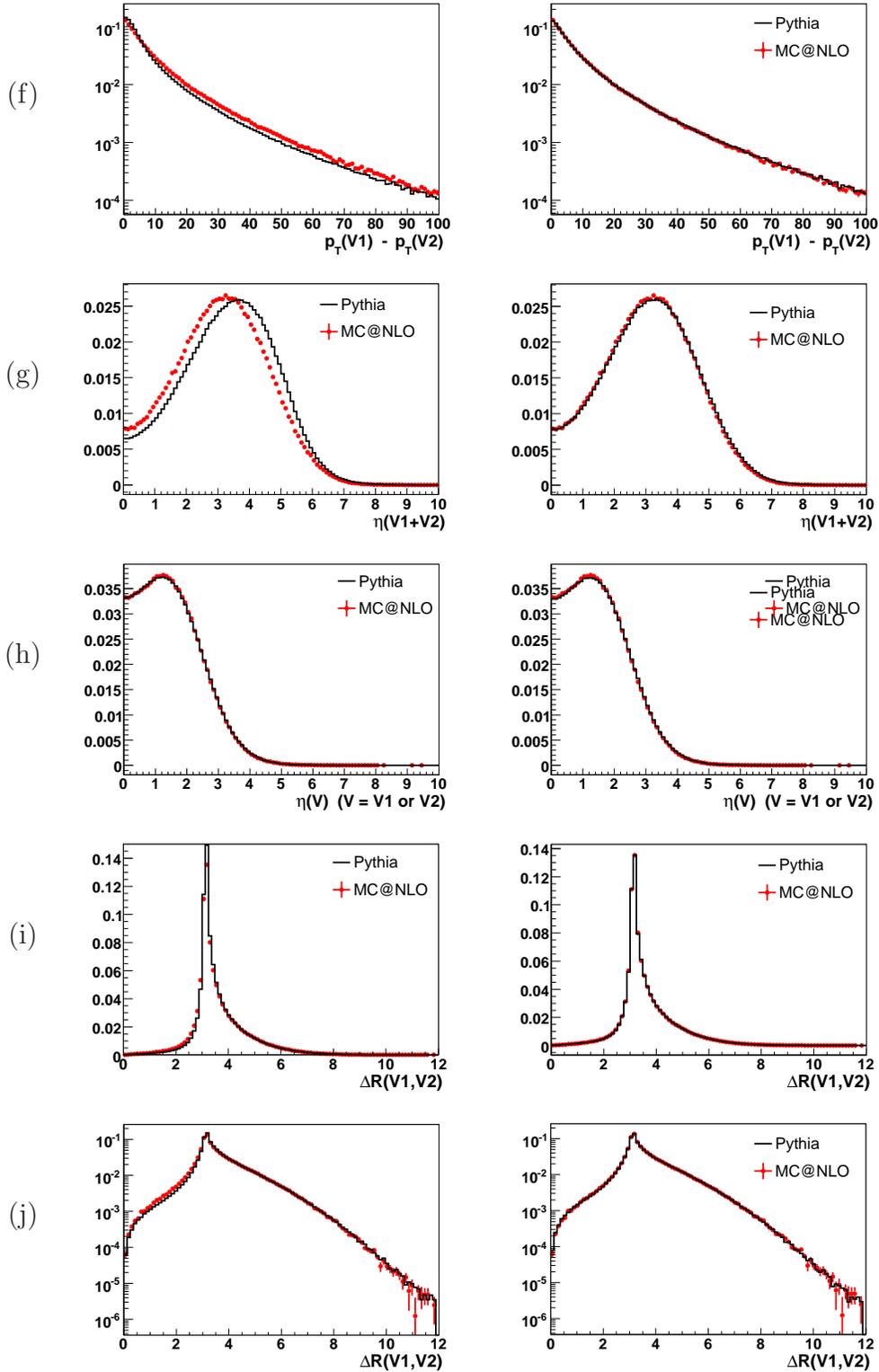
Appendix F

Random Forest Classifier Inputs

This appendix contains additional plots of the input variables for the random forest classifiers. Figures F.1 and F.3 show each input variable after the combined fit to data using the random forest output distributions (chapter 9). Figures F.2 and F.4 show the analogous plots comparing the background-subtracted data to the signal.
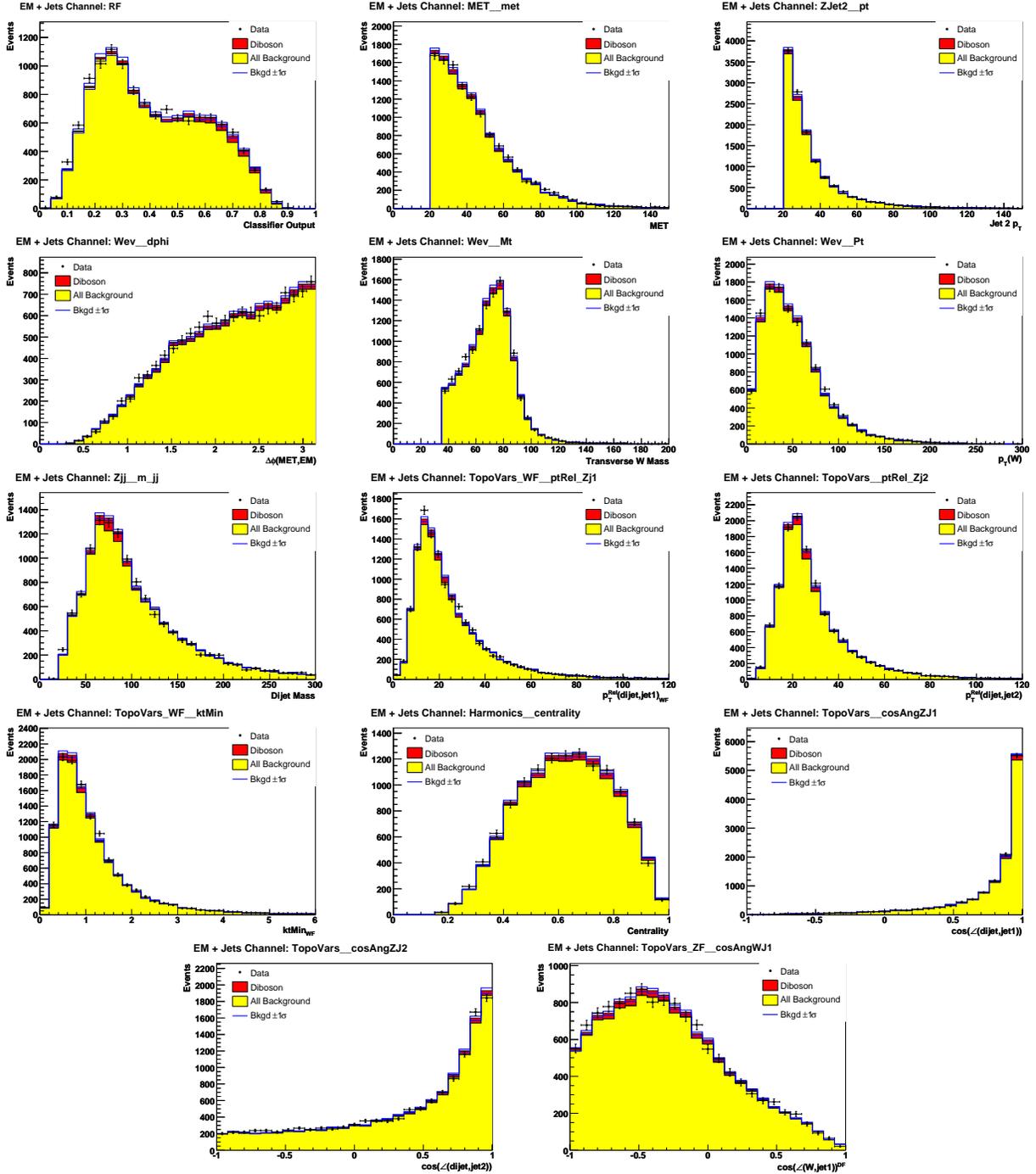
Figure F.1: Post-fit comparison of data to signal plus background in the $e\nu q\bar{q}$ channel for the random forest input variables. The signal and backgrounds are adjusted to the best fit values of each nuisance parameter as determined by the cross section measurement. The blue lines show ± 1 standard deviation of the systematic uncertainty.
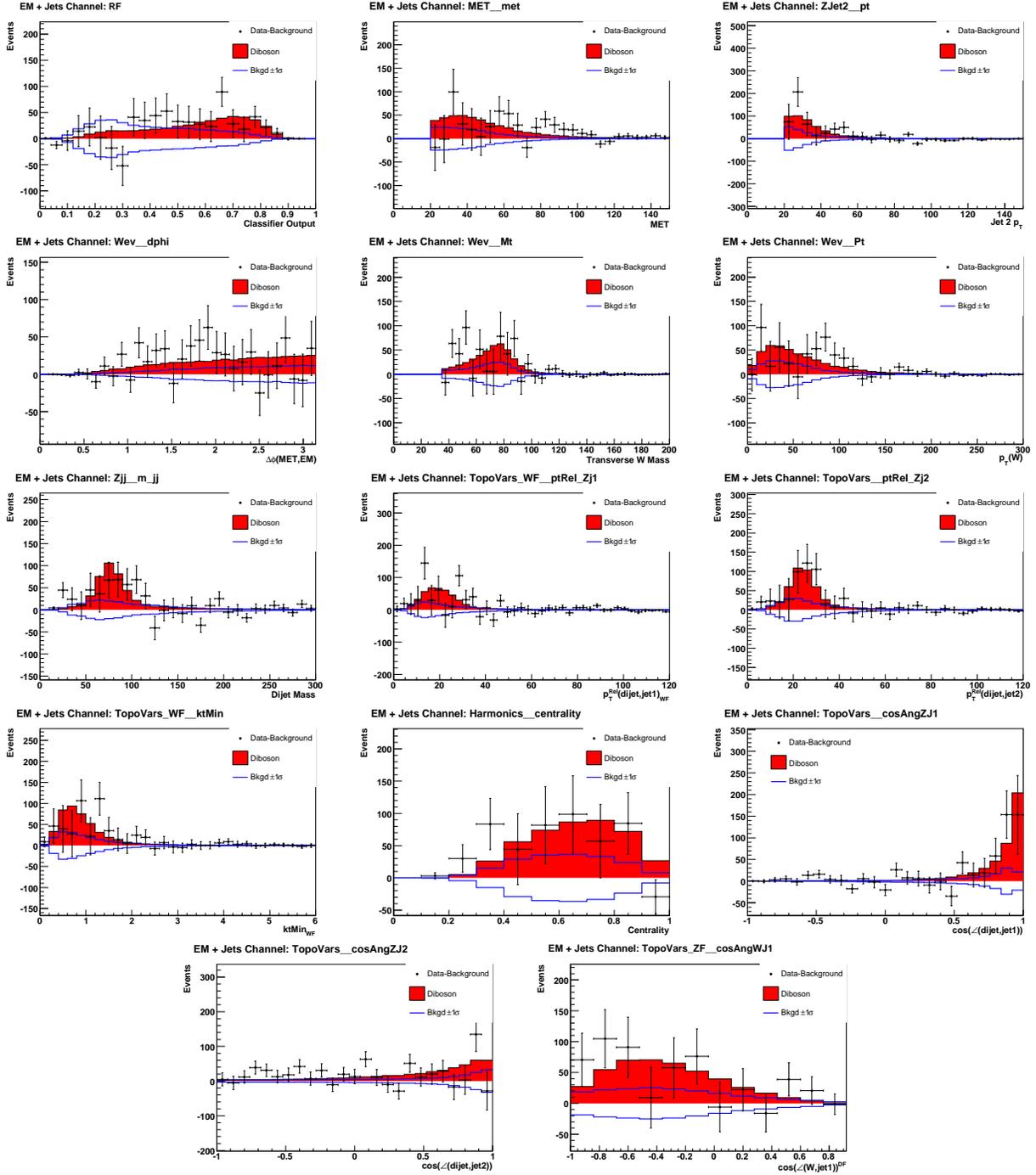
Figure F.2: Post-fit comparison of background-subtracted data to signal in the $e\nu q\bar{q}$ channel for the the random forest input variables. The signal and subtracted backgrounds are adjusted to the best fit values of each nuisance parameter as determined by the cross section measurement. The blue lines show ± 1 standard deviation of the systematic uncertainty.
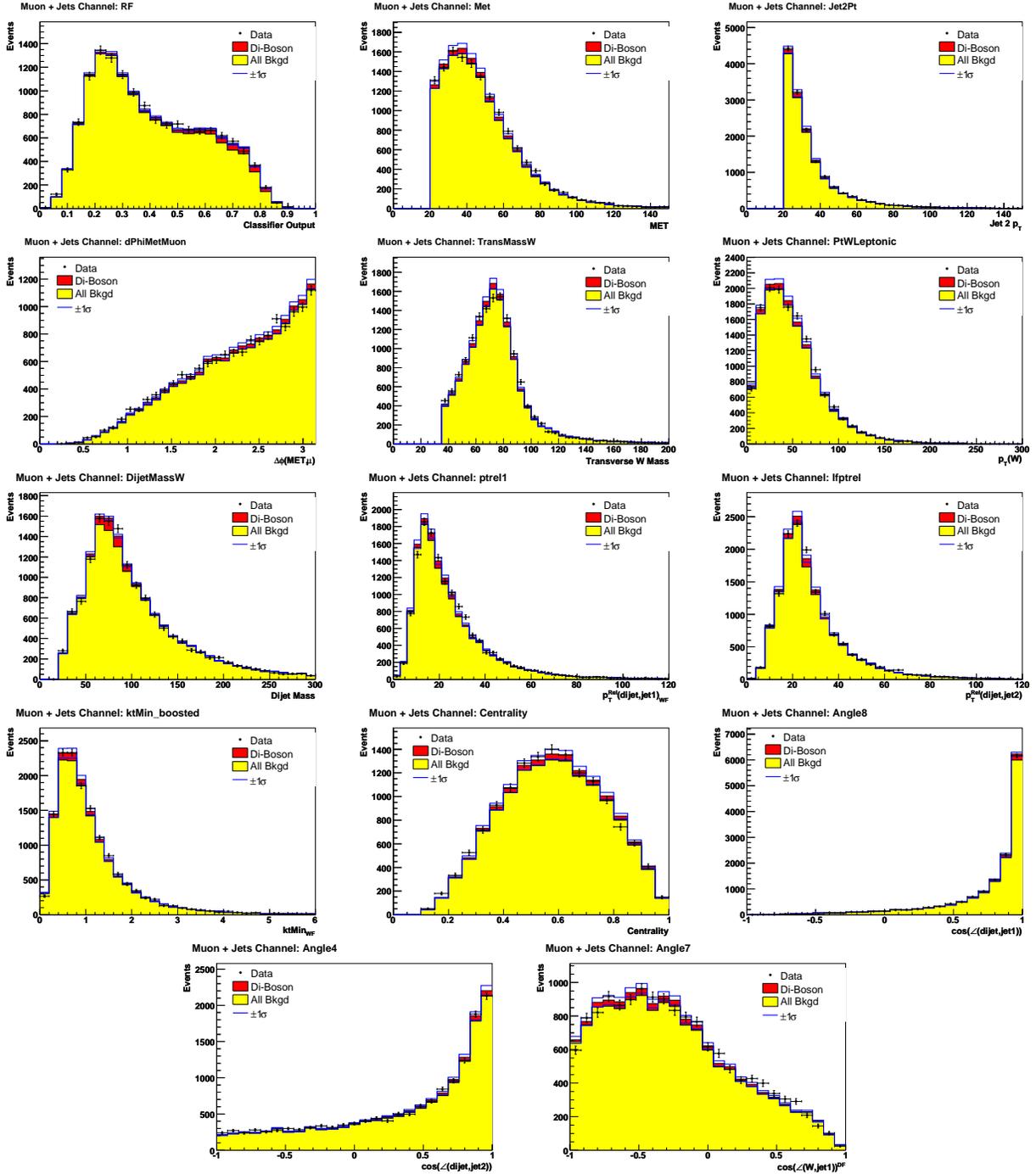
Figure F.3: Post-fit comparison of data to signal plus background in the $\mu\nu q\bar{q}$ channel for the random forest input variables. The signal and backgrounds are adjusted to the best fit values of each nuisance parameter as determined by the cross section measurement. The blue lines show ± 1 standard deviation of the systematic uncertainty.

Figure F.4: Post-fit comparison of background-subtracted data to signal in the $\mu\nu q\bar{q}$ channel for the the random forest input variables. The signal and subtracted backgrounds are adjusted to the best fit values of each nuisance parameter as determined by the cross section measurement. The blue lines show ± 1 standard deviation of the systematic uncertainty.

Appendix G

PDF Uncertainties

The 40 error sets (20 plus and 20 minus variations) for the PDF set CTEQ6.1M (also called CTEQ6.1) were used to evaluate the shape uncertainty for each sample due to the PDF set. The MC samples used in this analysis were generated with PDF set CTEQ6L1 (also called CTEQ6ll), thus, the relative shift between each of the 40 CTEQ6.1M error sets and the nominal CTEQ6.1M set was applied as a weight to each MC event. The random forest output distributions for each sample with and without applying the weights were then compared to determine the shape dependencies.

Figures G.1 and G.2 show each of the PDF fractional uncertainties (20 uncertainties, each with a plus and minus shift) as a function of the random forest output in the electron channel for the W+jets background. (The PDF uncertainty is most important for the W+jets sample as it is the dominant background, however, is was determined for all simulated samples.) Figure G.3 illustrates the total magnitude of all 20 systematic uncertainties obtained via quadrature sum. The largest uncertainty comes from PDF variation #15 with an amplitude of $\simeq 1\%$ and it can be seen that many of the other uncertainties have a shape similar to that of #15. For this reason, the shape from #15 is use with an amplitude scaled to match the total magnitude of the quadrature sum of all 20 uncertainties. This procedure partially overestimates the uncertainty size by double-counting the six PDF uncertainties for which

both the positive and negative fluctuations return nearly identical shapes.
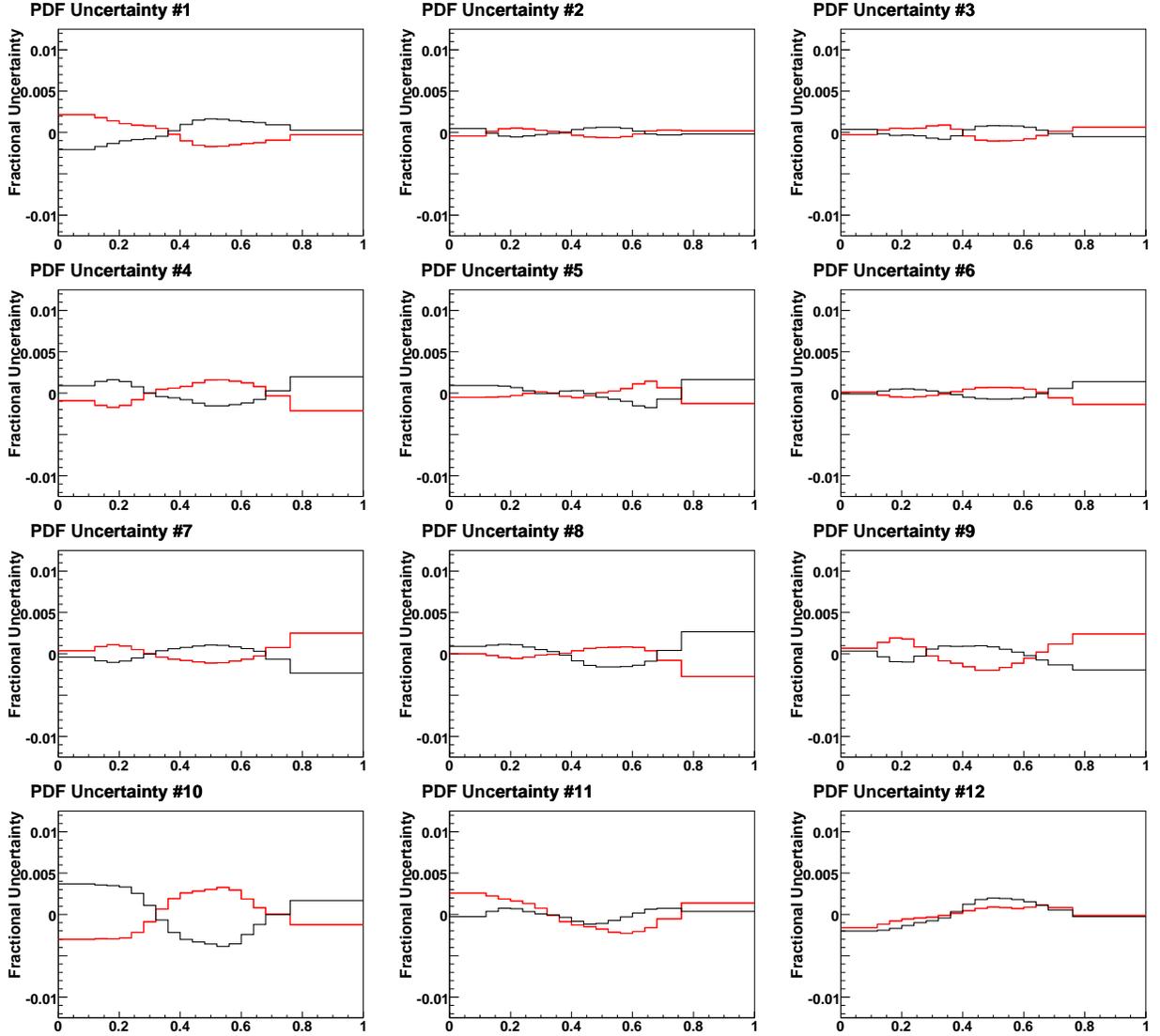
Figure G.1: PDF uncertainties #1 - #12 as a function of the random forest output in the $e\nu q\bar{q}$ channel for the W+jets background.
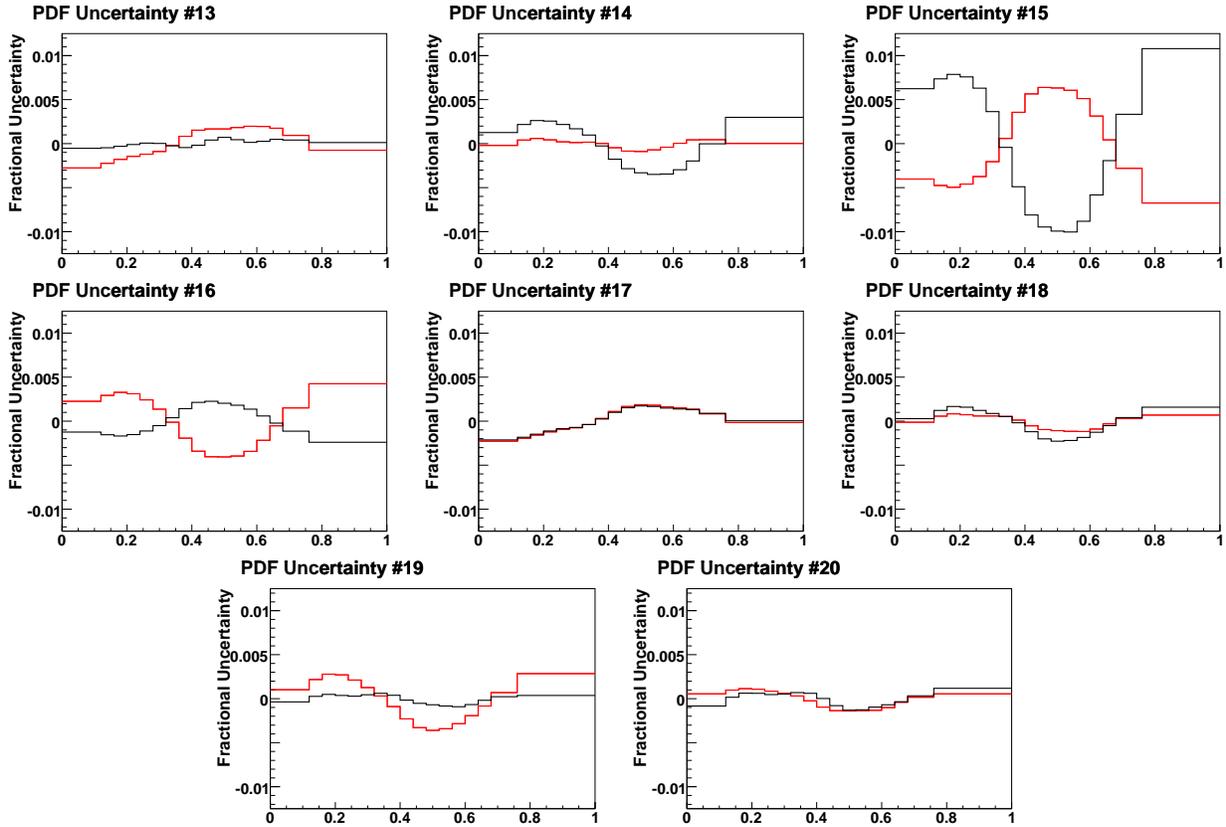
Figure G.2: PDF uncertainties #13 - #20 as a function of the random forest output in the $e\nu q\bar{q}$ channel for the W+jets background.
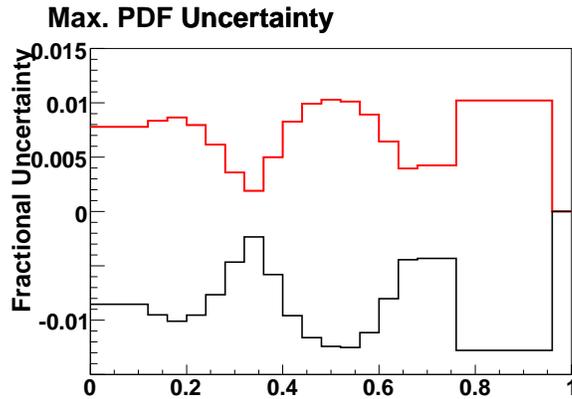
Figure G.3: Maximum PDF uncertainty for the W+jets background obtained by quadrature sum.

162

Appendix H

Shape Uncertainties

This appendix shows each of the shape uncertainties mentioned in chapter 8 (table 8.1) as a function of the random forest output. Shape uncertainties were also measured as a function of the dijet mass in order to perform the measurement using the dijet mass, as well as for each of the random forest input variables in order to determine the $\pm 1\sigma$ bands for the plots in appendix F. However, the actual shapes of these systematics are shown only as a function of random forest output due to space considerations and the importance of the random forest output in the final measurement.

The $\pm 1\sigma$ (standard deviation) shape distortions for each shape uncertainty can be determined for any given distribution by changing that nuisance parameter by $\pm 1\sigma$ from the nominal value and measuring the fractional change in the given distribution. The simulated samples do not have infinite statistics, so the distributions take the form of histograms. Bins with low statistics are subject to statistical fluctuations that could distort the true shape of the uncertainty. In order to mitigate this, bins with low statistics are merged with neighboring bins so that all bins have at least half the number of entries as the highest bin.

When multiple nuisance parameters are varied in the cross section measurement (chapter 9), the signal and each background will be affected differently. Therefore, each shape uncertainty must be measured separately for each sample. Also, because W events with

jets from heavy flavor partons (denoted W+hf) are affected by an additional normalization uncertainty not affecting W events with light parton jets (denoted W+lp), each shape uncertainty is measured separately for the W+lp and W+hf.

H.1 Trigger Efficiency for $\mu\nu q\bar{q}$ Channel

As described in appendix A, a shape uncertainty is assigned to the trigger efficiency in the $\mu\nu q\bar{q}$ channel based on the difference between data collected with the single muon trigger suite versus with the inclusive triggers. Because it is measured from the data, the same shape uncertainty is applied to every sample.
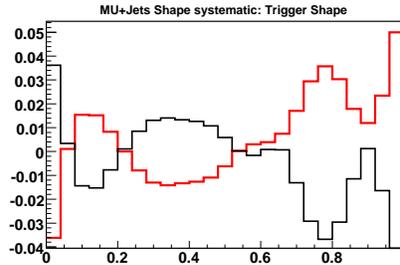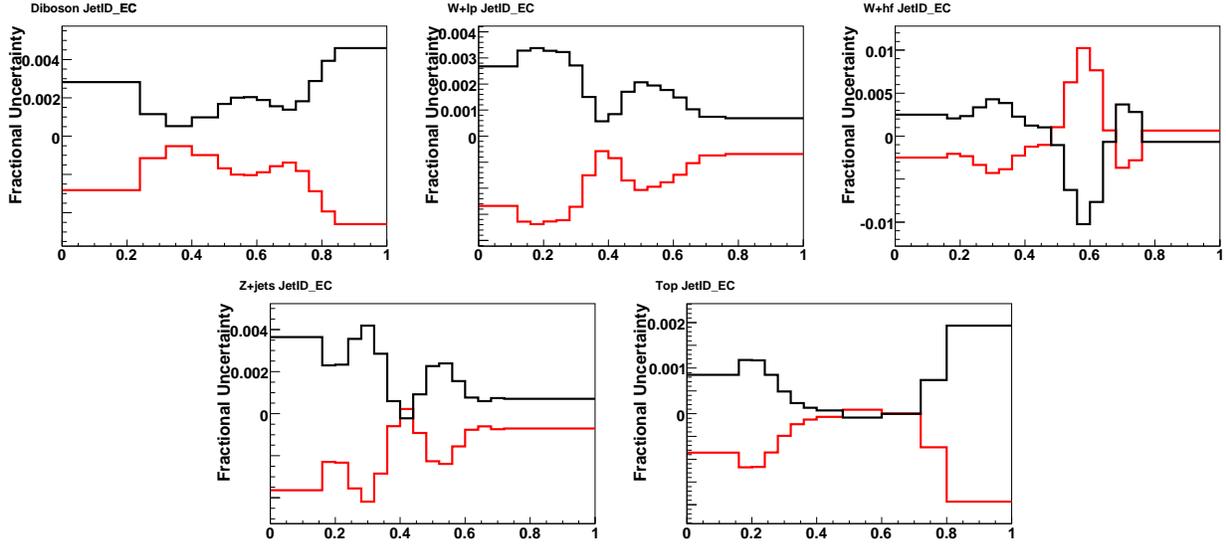

Figure H.1: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution assigned to all samples as a result of shifting the muon trigger shape by $+1\sigma$ (black line) or -1σ (red line).

H.2   Jet Identification

The primary source for inefficiencies in jet ID are different for jets in the CC, ICR, and EC. Therefore, the jet ID uncertainty is evaluated separately for each region. Also, while it is possible to simulated a decreased jet ID efficiency in the MC samples by removing more jets than otherwise would be removed, the reverse is not true. Thus, the change in shape due to increasing the jet ID efficiency is assumed to be equal and opposite to change in shape from decreasing the jet ID efficiency.
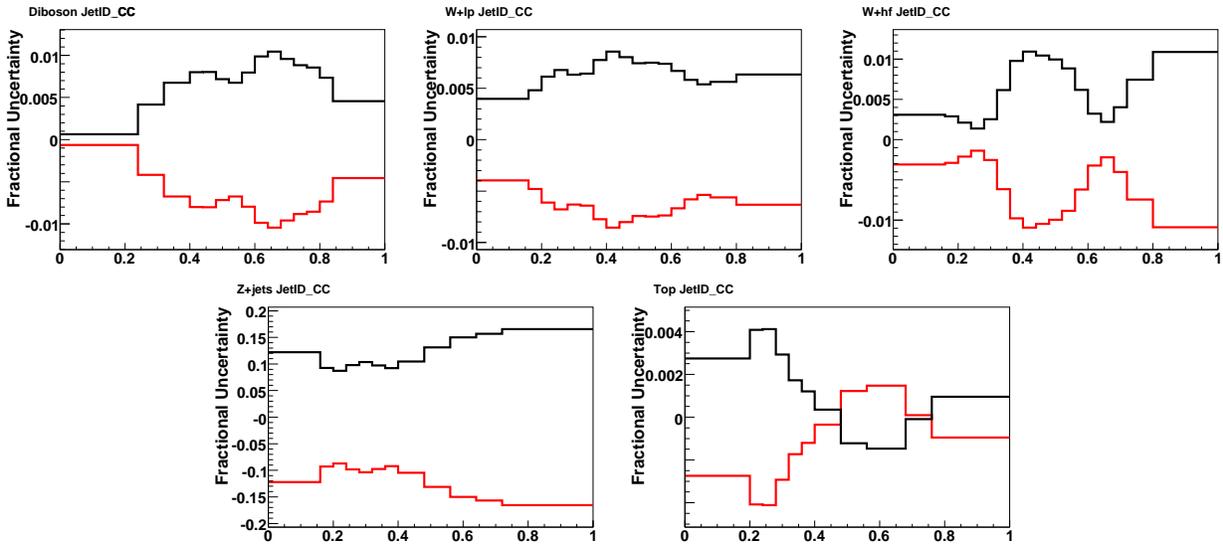
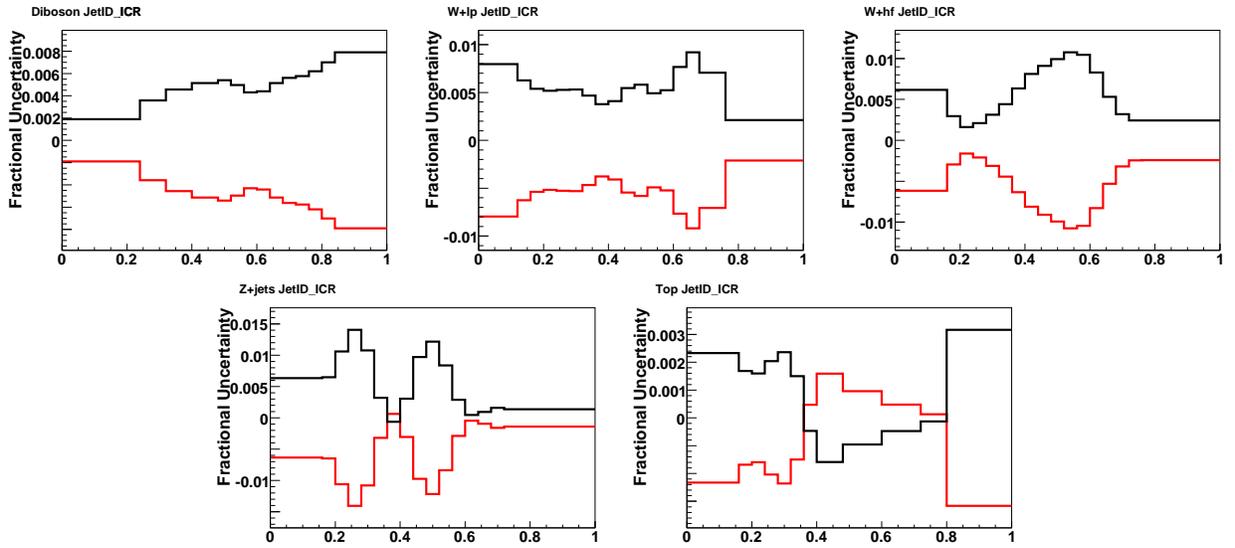Figure H.2: Fractional change in $e\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the CC jet ID efficiency by -1σ (red line) and the equal and opposite change assumed for shifting the jet ID efficiency by $+1\sigma$ (black line).
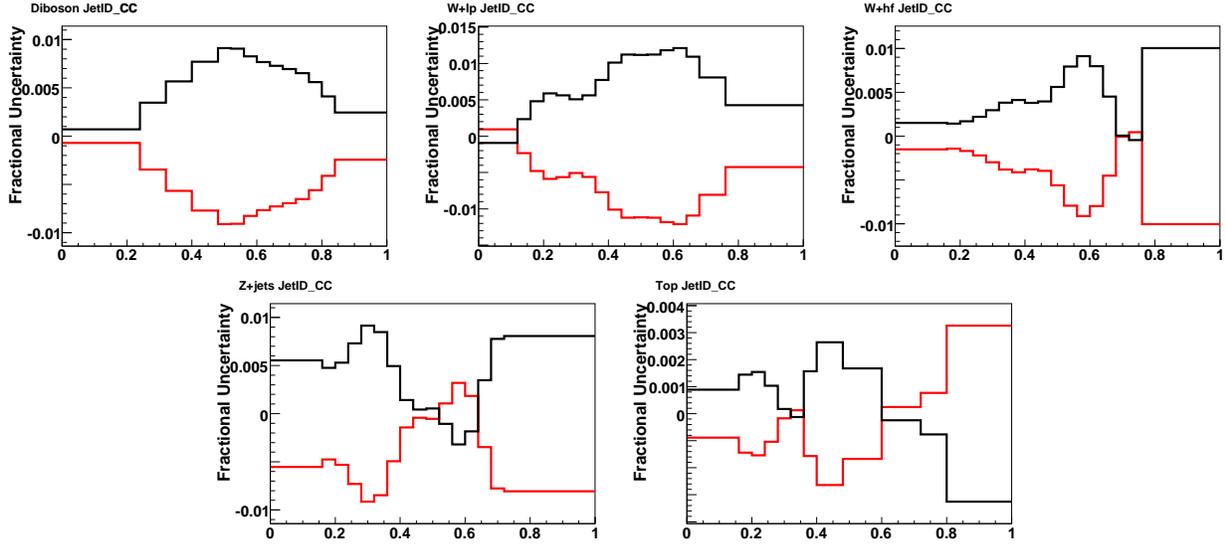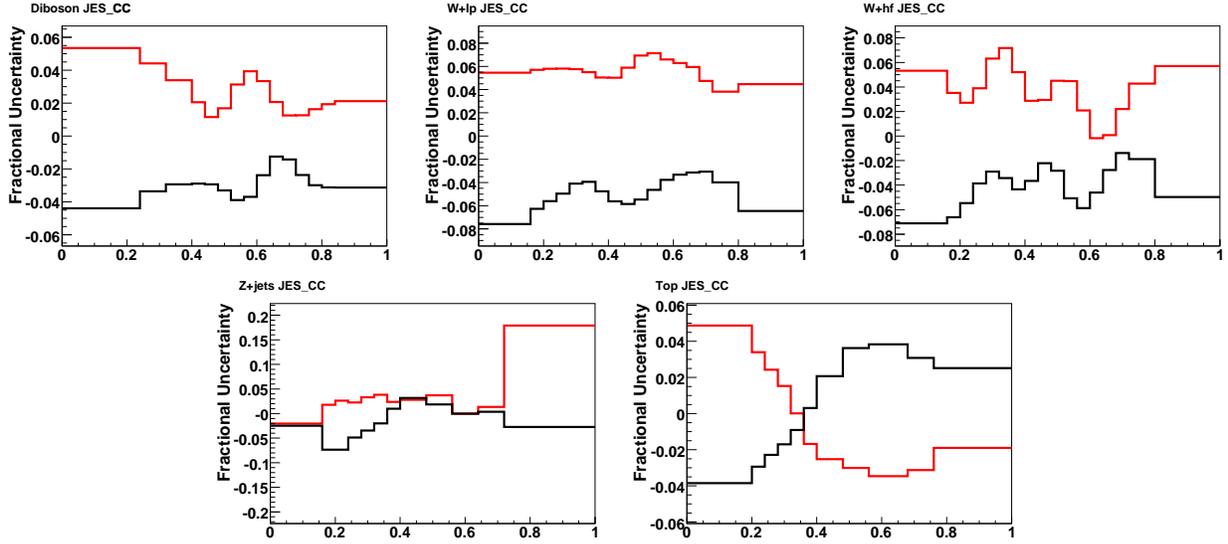
Figure H.3: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the CC jet ID efficiency by -1σ (red line) and the equal and opposite change assumed for shifting the jet ID efficiency by $+1\sigma$ (black line).
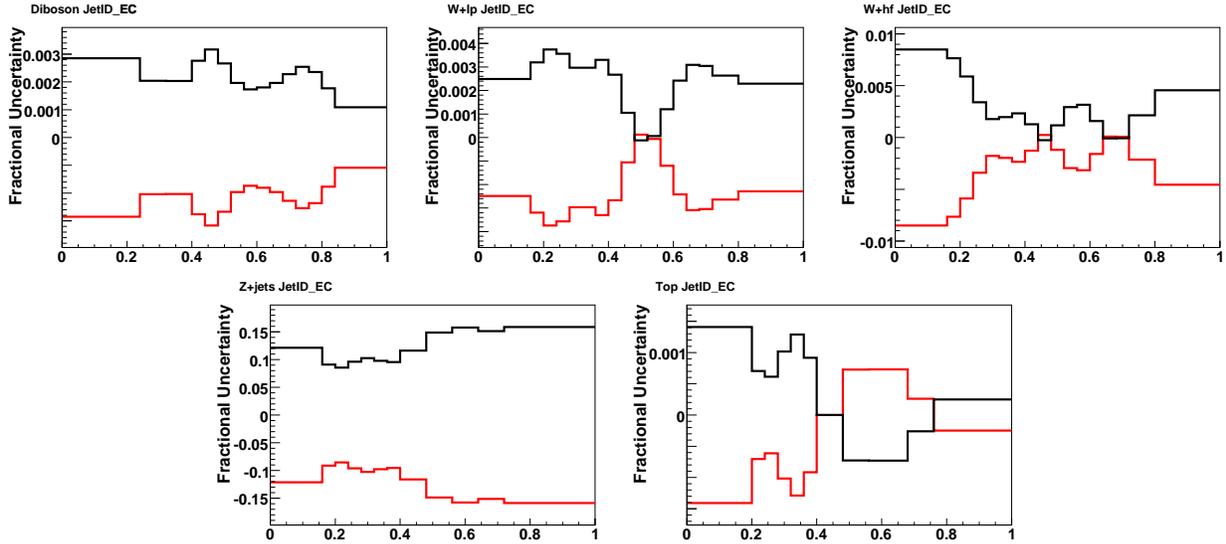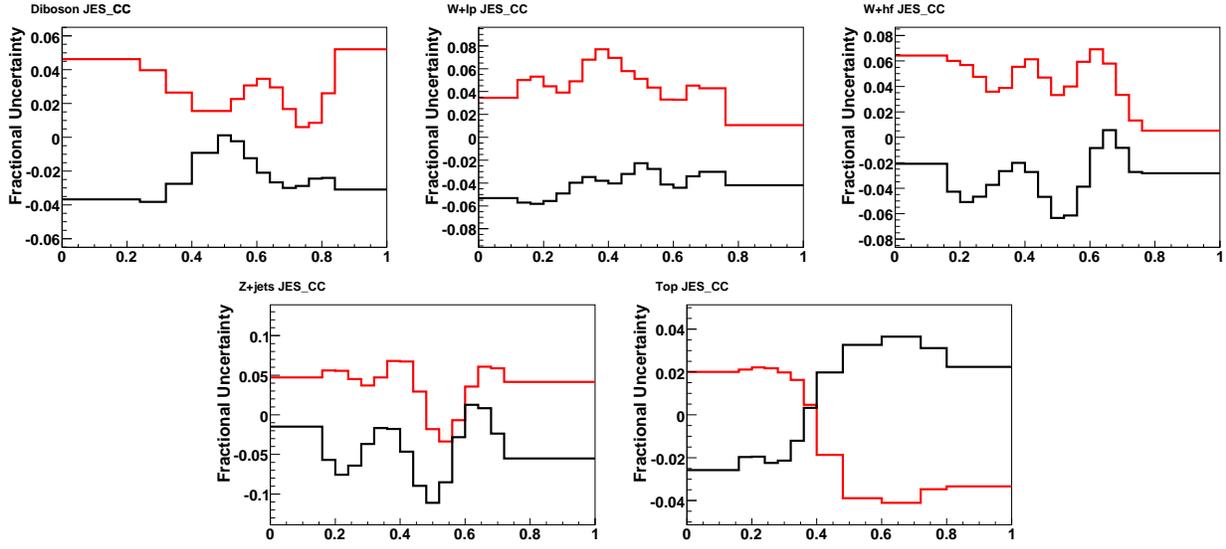

Figure H.4: Fractional change in $e\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the EC jet ID efficiency by -1σ (red line) and the equal and opposite change assumed for shifting the jet ID efficiency by $+1\sigma$ (black line).

Figure H.5: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the EC jet ID efficiency by -1σ (red line) and the equal and opposite change assumed for shifting the jet ID efficiency by $+1\sigma$ (black line).
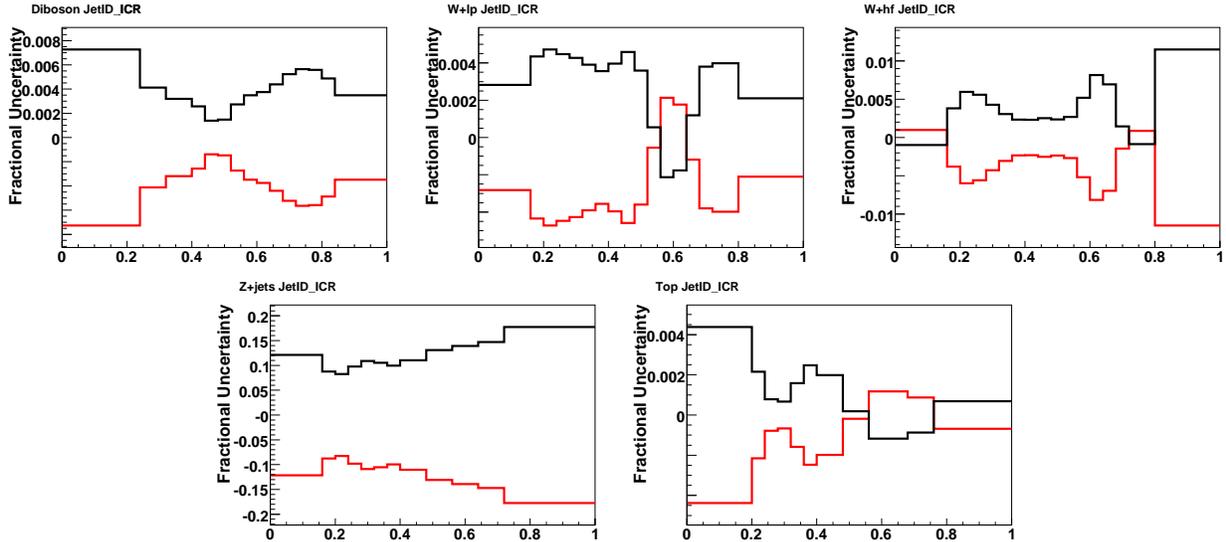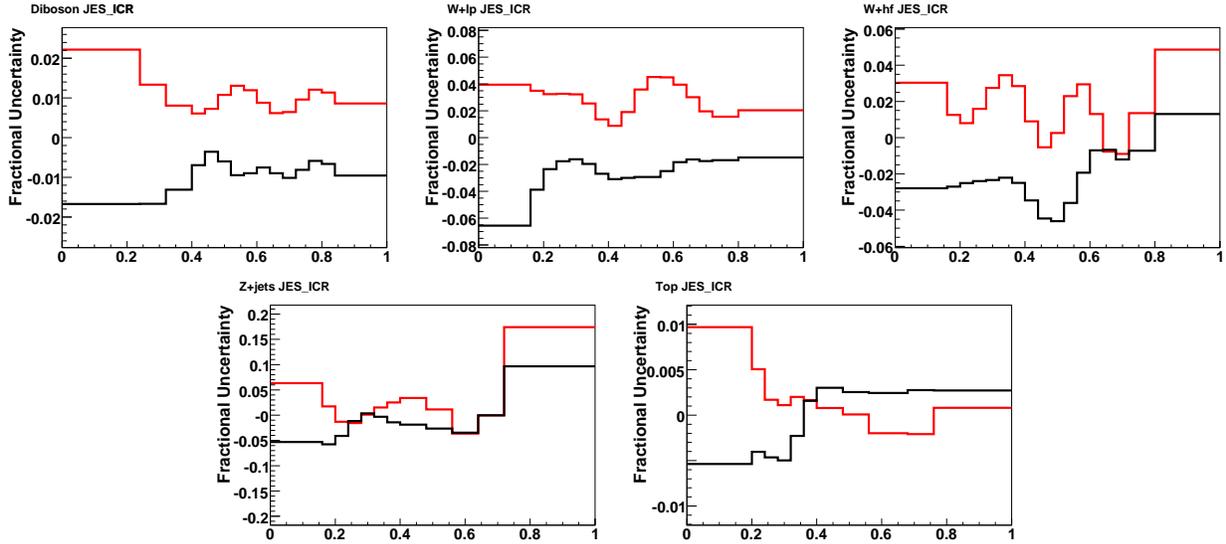

Figure H.6: Fractional change in $e\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the ICR jet ID efficiency by -1σ (red line) and the equal and opposite change assumed for shifting the jet ID efficiency by $+1\sigma$ (black line).
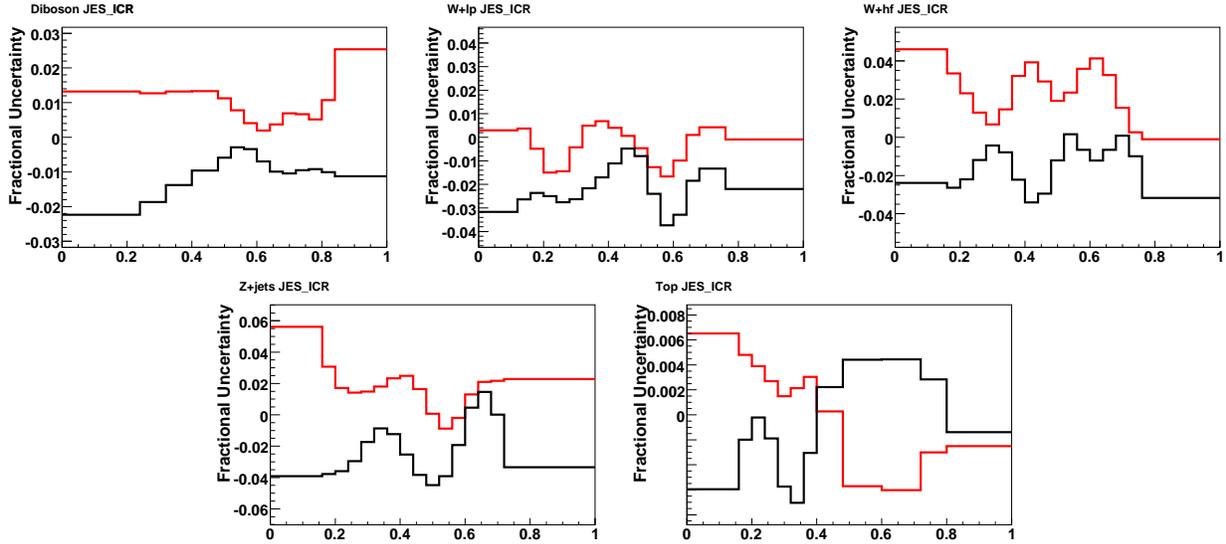
Figure H.7: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the ICR jet ID efficiency by -1σ (red line) and the equal and opposite change assumed for shifting the jet ID efficiency by $+1\sigma$ (black line).

H.3   Jet Energy Scale

Like the jet ID efficiency, the JES uncertainty has been assessed separately for the three different regions of the calorimeter: CC, ICR, and EC.
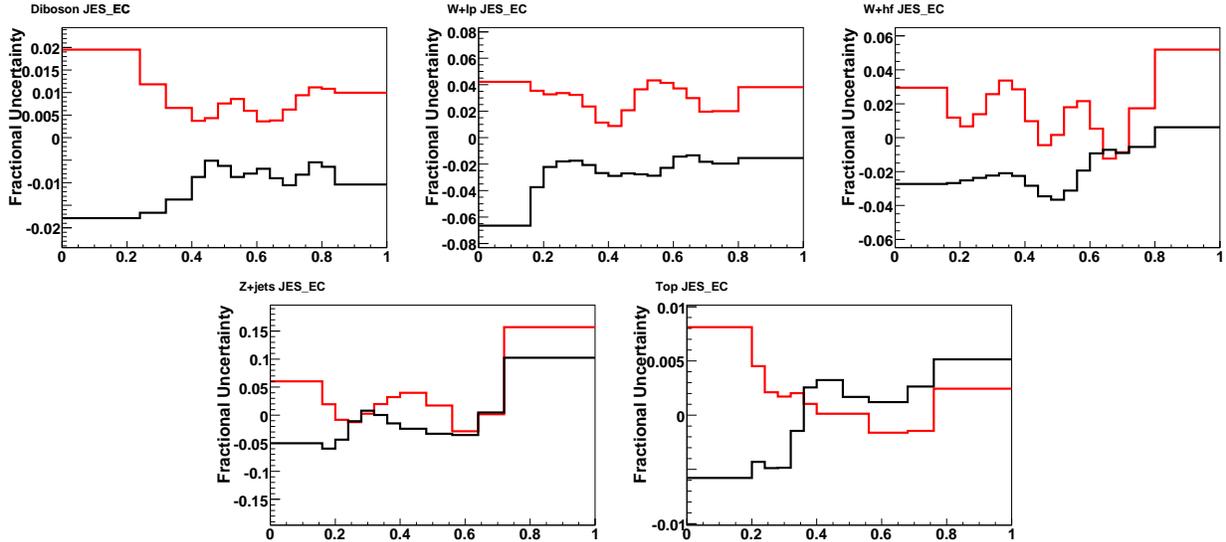

Figure H.8: Fractional change in $e\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the JES in the CC by $+1\sigma$ (black line) or by -1σ (red line).
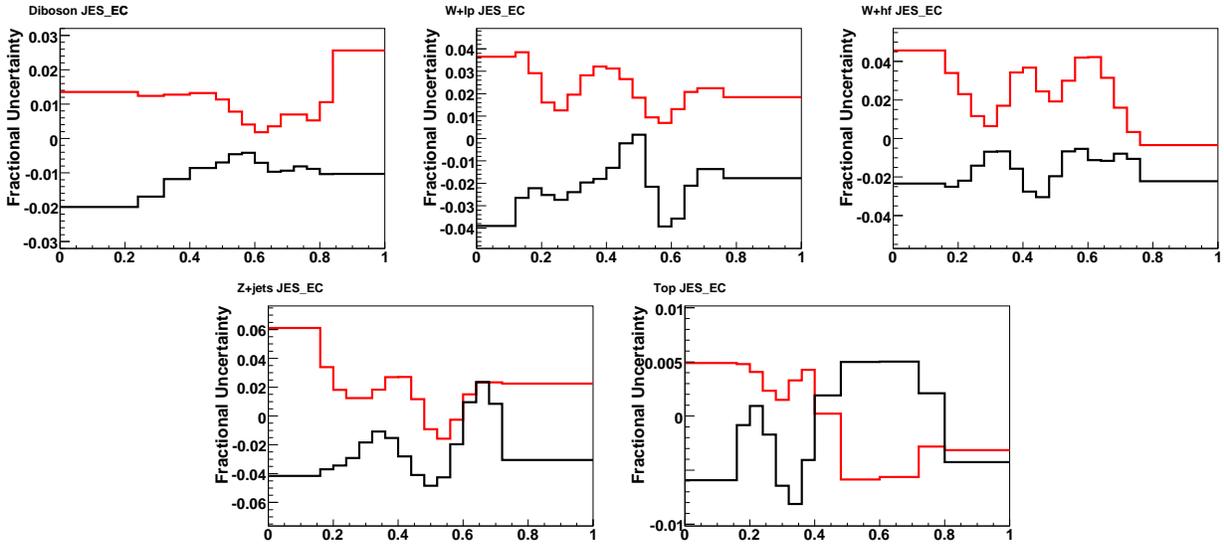
Figure H.9: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the JES in the CC by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.10: Fractional change in $e\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the JES in the ICR by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.11: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the JES in the ICR by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.12: Fractional change in $e\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the JES in the EC by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.13: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the JES in the EC by $+1\sigma$ (black line) or by -1σ (red line).

H.4   Diboson NLO Correction

As explained in section 6.8, the NLO correction for the diboson sample is assigned an uncertainty of half the difference between the LO prediction and the NLO corrected distribution.
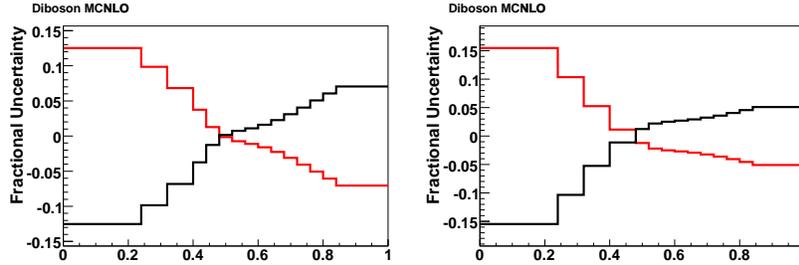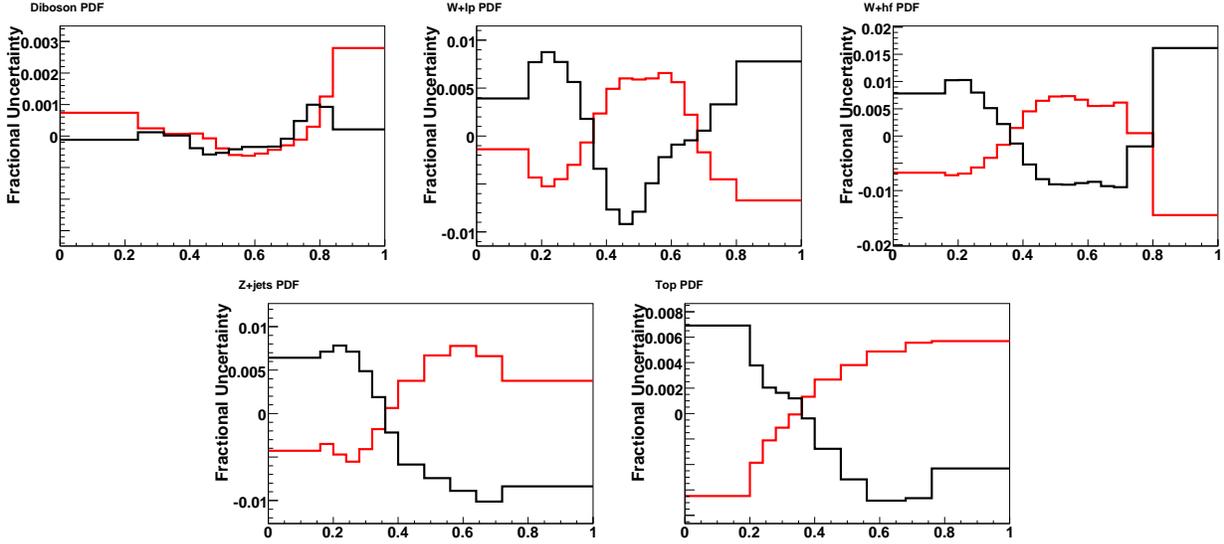

Figure H.14: Fractional change in $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ random forest output distributions for the signal sample as a result of shifting the MC@NLO correction by $+1\sigma$ (black line) or by -1σ (red line).

H.5 PDF Set

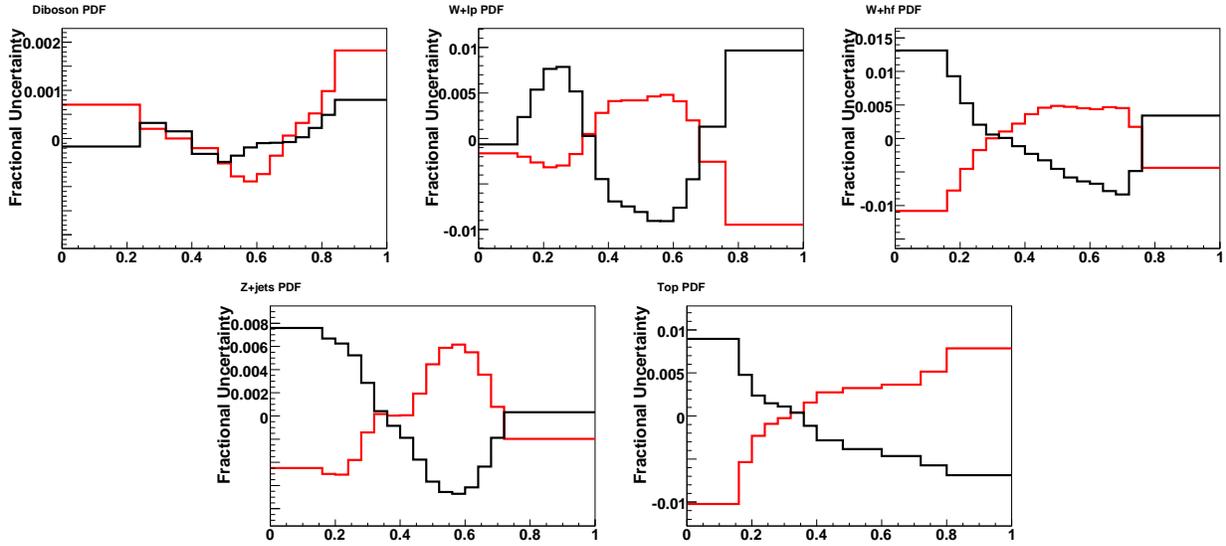Derivation of the shape uncertainty due to uncertainties on the PDF set is explained in appendix G.
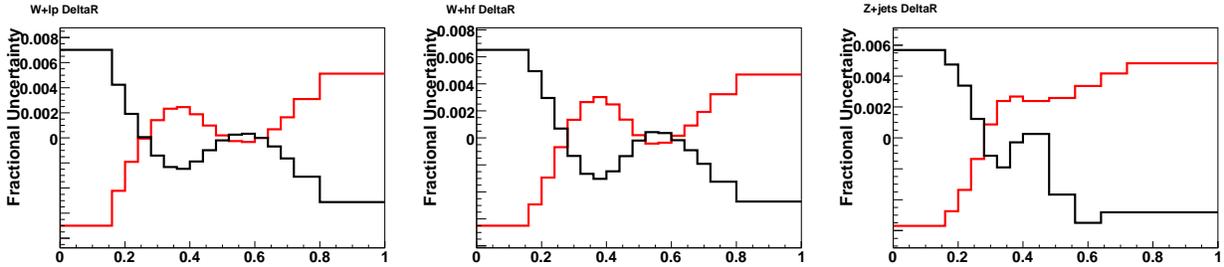

Figure H.15: Fractional change in $e\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the PDF by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.16: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for each sample as a result of shifting the PDF by $+1\sigma$ (black line) or by -1σ (red line).

H.6   Alpgen Jet η and ΔR Corrections

The uncertainty assigned to the jet η and ΔR corrections are determined as explained in section 6.7.

Figure H.17: Fractional change in $e\nu q\bar{q}$ random forest output distribution for the W+lp, W+hf, and Z+jets samples as a result of shifting the ΔR re-weighting by $+1\sigma$ (black line) or by -1σ (red line).
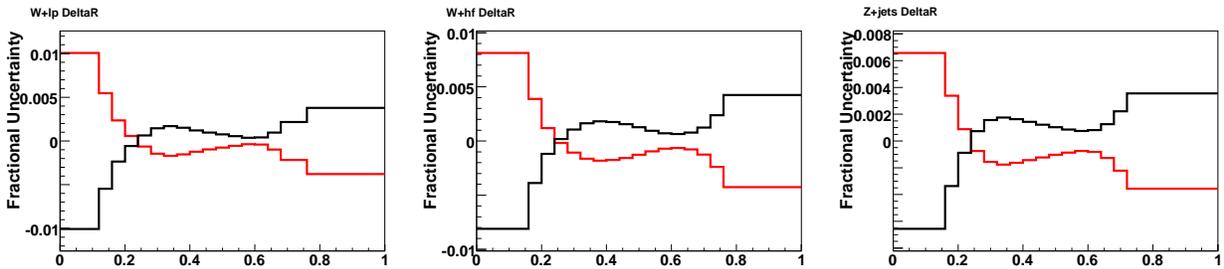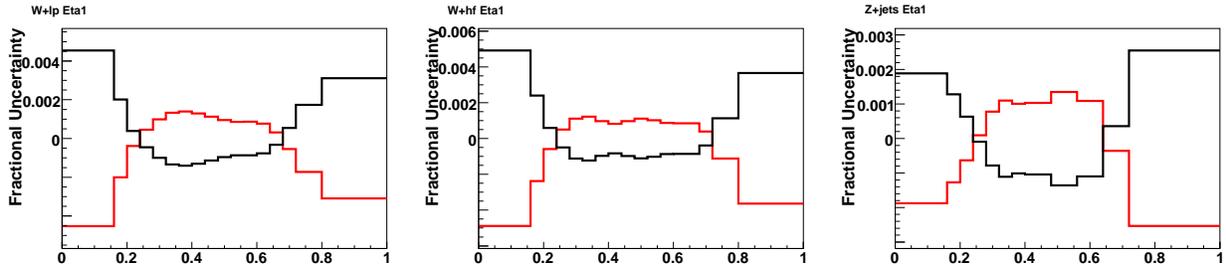

Figure H.18: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for the W+lp, W+hf, and Z+jets samples as a result of shifting the ΔR re-weighting by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.19: Fractional change in $e\nu q\bar{q}$ random forest output distribution for the W+lp, W+hf, and Z+jets samples as a result of shifting the leading jet η re-weighting by $+1\sigma$ (black line) or by -1σ (red line).
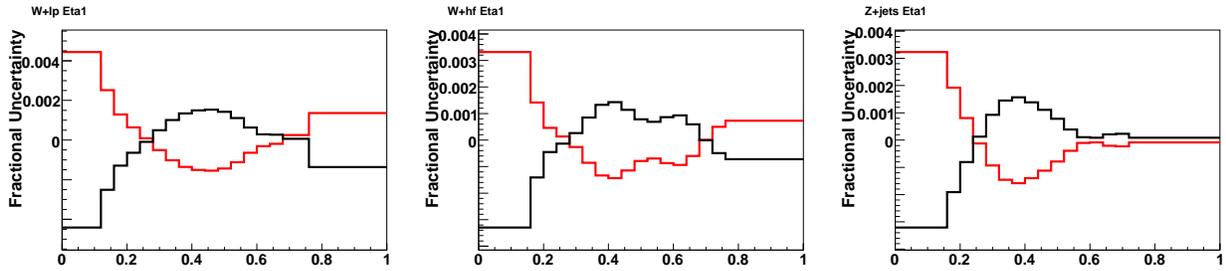

Figure H.20: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for the W+lp, W+hf, and Z+jets samples as a result of shifting the leading jet η re-weighting by $+1\sigma$ (black line) or by -1σ (red line).
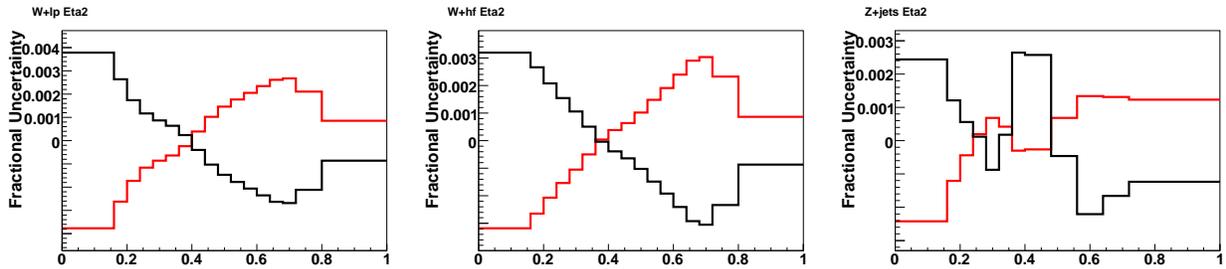

Figure H.21: Fractional change in $e\nu q\bar{q}$ random forest output distribution for the W+lp, W+hf, and Z+jets samples as a result of shifting the second jet η re-weighting by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.22: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for the W+lp, W+hf, and Z+jets samples as a result of shifting the second jet η re-weighting by $+1\sigma$ (black line) or by -1σ (red line).
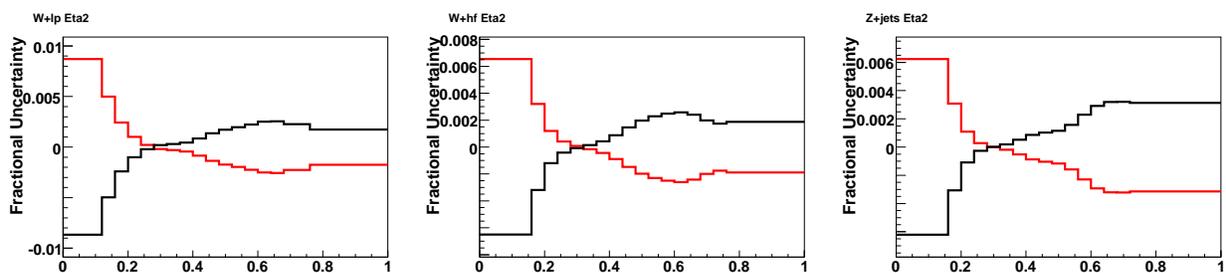
H.7 Renormalization and Factorization Scale

The renormalization and factorization scale for the W+jets sample are assigned an uncertainty as explained in appendix D.
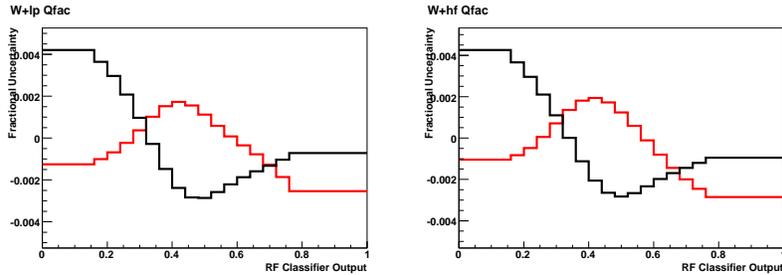

Figure H.23: Fractional change in $e\nu q\bar{q}$ random forest output distribution for the W+lp and W+hf samples as a result of shifting the renormalization and factorization scale by $+1\sigma$ (black line) or by -1σ (red line).
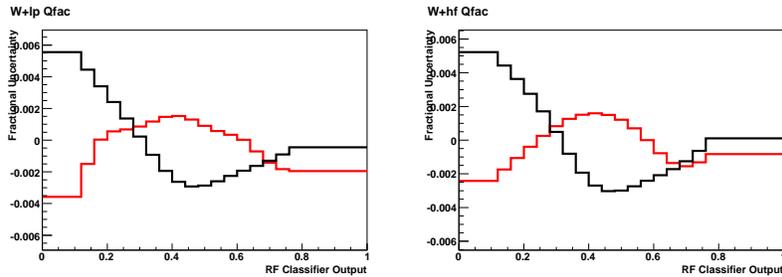

Figure H.24: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for the W+lp and W+hf samples as a result of shifting the renormalization and factorization scale by $+1\sigma$ (black line) or by -1σ (red line).

H.8   Alpgen Parton-Jet Matching Parameters

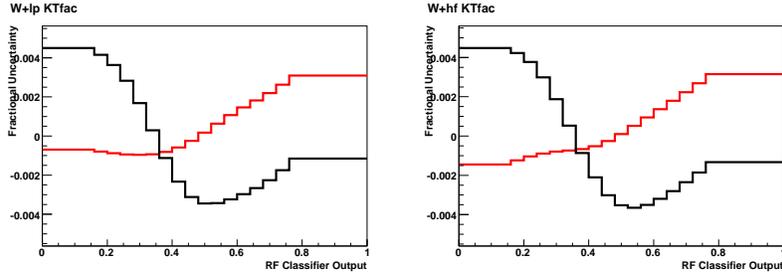The uncertainties are assigned to the ALPGEN parton-jet matching parameters are explained in appendix D.

Figure H.25: Fractional change in $e\nu q\bar{q}$ random forest output distribution for the W+lp and W+hf samples as a result of shifting $k_\perp$-factor by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.26: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for the W+lp and W+hf samples as a result of shifting $k_\perp$-factor by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.27: Fractional change in $e\nu q\bar{q}$ random forest output distribution for the W+lp and W+hf samples as a result of shifting the parton-jet matching p_T threshold by $+1\sigma$ (black line) or by -1σ (red line).
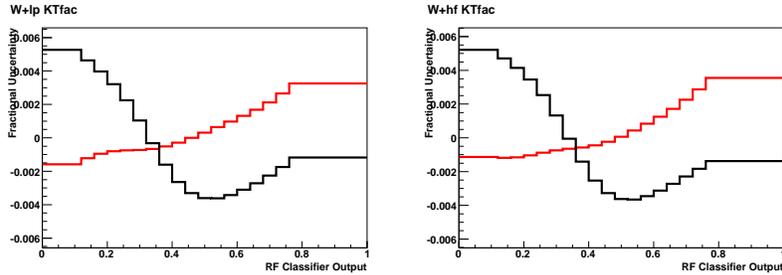

Figure H.28: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for the W+lp and W+hf samples as a result of shifting the parton-jet matching p_T threshold by $+1\sigma$ (black line) or by -1σ (red line).

Figure H.29: Fractional change in $e\nu q\bar{q}$ random forest output distribution for the W+lp and W+hf samples as a result of shifting the parton-jet matching radius by $+1\sigma$ (black line) or by -1σ (red line).
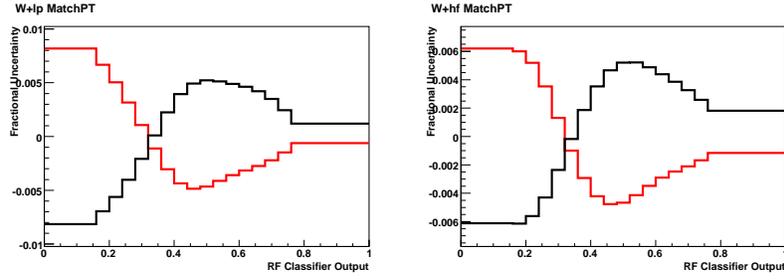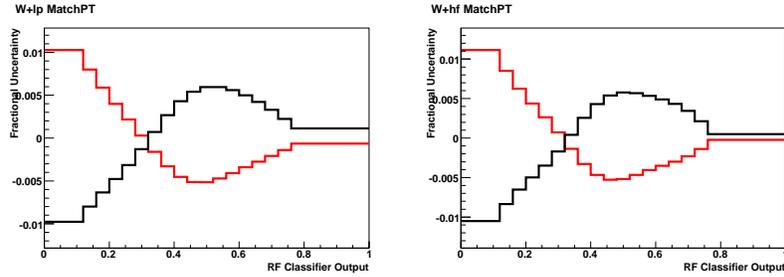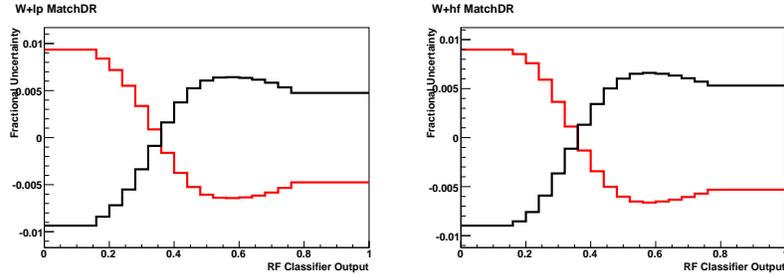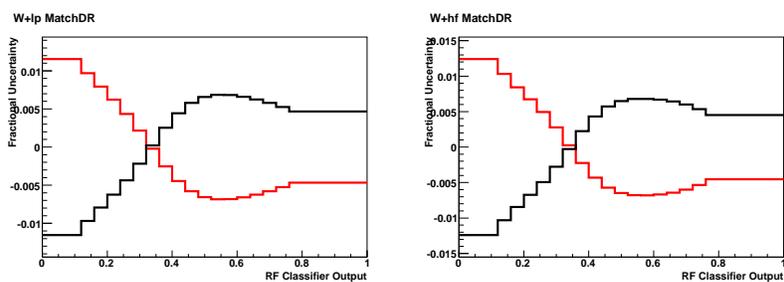
Figure H.30: Fractional change in $\mu\nu q\bar{q}$ random forest output distribution for the W+lp and W+hf samples as a result of shifting the parton-jet matching radius by $+1\sigma$ (black line) or by -1σ (red line).

H.9   Multijet Shape

As stated in section 4.3.1, the uncertainty on the multijet estimation for the electron channel is determined by varying ϵ_{fake} by $\pm1\sigma$ and reevaluating the event weights for the multijet estimation. The uncertainty on the multijet estimation for the muon channel is derived as described in appendix B.
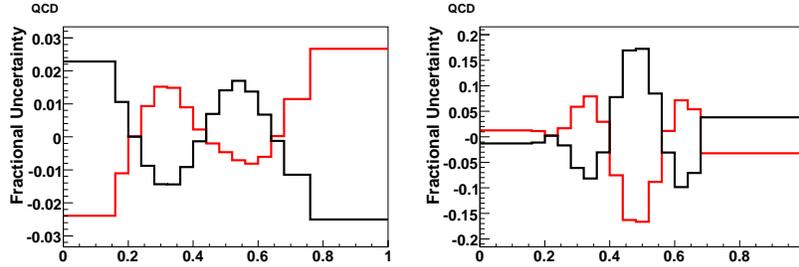

Figure H.31: Fractional change in $e\nu q\bar{q}$ and $\mu\nu q\bar{q}$ random forest output distribution for the multijet estimation as a result of shifting each by $+1\sigma$ (black line) or by -1σ (red line).

Bibliography

[1] J. M. Campbell and R. K. Ellis, "Update on vector boson pair production at hadron colliders", Phys. Rev. D **60**, 113006 (1999).

[2] F. Abe *et al.* (CDF Collaboration), "Evidence for W^+W^- Production in $p\bar{p}$ Collisions at $\sqrt{s} = 1.8$ TeV", Phys. Rev. Lett. **78**, 4536 (1997).

[3] B. Abbott *et al.* (D0 Collaboration), "Studies of WW and WZ production and limits on anomalous WWγ and WWZ couplings", Phys. Rev. D **60**, 072002 (1999).

[4] K. Hagiwara, J. Woodside, and D. Zeppenfeld, "Measuring the WWZ coupling at the Fermilab Tevatron", Phys. Rev. D **41**, 2113 (1990).

[5] A. Sanchez-Hernandez, "Search for Anomalous $WW/WZ \rightarrow e\nu jj$ Production at D0", PhD. Thesis (1997).

[6] L. M. Lederman, "The Tevatron", Scientific American, **264(3)**, 48 (1991).

[7] H. T. Edwards, "The Tevatron Energy Doubler: A Superconducting Accelerator", Ann. Rev. Nucl. Part. Sci., **35**, 605 (1985).

[8] TeVI Group, "Design Report Tevatron I projects", Fermilab Technical Report FERMILAB-DESIGN-1984-01 (1984).

[9] F. T. Cole *et al.*, "A Report on the Design of the Fermi National Accelerator Laboratory Superconducting Accelerator", Beams Document 1888-v1 (1979).

[10] J. Thompson, "Introduction to Colliding Beams at Fermilab", Fermilab Technical Report FermiLab TM-1909 (1994).

[11] C. W. Schmidt and C. D. Curtis, "A 50 mA Negative Hydrogen-Ion Source", IEEE Transactions on Nuclear Science, NS-26, 4120 (1979).

[12] C. D. Curtis *et al.*, "Linac H-Beam Operation and Uses at FermiLab", IEEE Transactions on Nuclear Science, NS-26, 3760 (1979).

[13] E. L. Hubbard *et al.*, "Booster synchrotron", Fermilab Technical Report Fermilab TM-405 (1973).

[14] C. S. Mishra, "The Fermilab Main Injector", Fermilab Technical Report Fermilab Conf-92/372 (1992).

[15] S. Nagaitsev *et al.*, "Antiproton Cooling in the Fermilab Recycler Ring", Fermilab Technical Report Fermilab-Conf-05-550-AD (2005).

[16] S. Abachi *et al.* (D0 Collaboration), "The D0 Detector", Nucl. Inst. Methods **A338**, 185 (1994).

[17] S. Abachi *et al.* (D0 Collaboration), "The D0 Upgrade: The Detector and its Physics", Fermilab Technical Report Fermilab-Pub-96/357-E (1996).

[18] V. M. Abazov *et al.* (D0 Collaboration), "The Upgraded D0 Detector", Nucl. Inst. Methods **A565**, 463 (2006).

[19] M. Roco, "The Silicon Microstrip Tracker for the D0 Upgrade", Internal D0 Note 3553 (1998).

[20] D0 Upgrade Collaboration, "D0 Silicon Tracker Technical Design Report" (1994).

[21] F. Borcherding *et al.* (D0 Collaboration), "Technical Design Report for the Upgrade L1/L2 Tracking Trigger - TDR for the CTT", Internal D0 Note 3551 (1998).

[22] J. Brzezniak (D0 Collaboration), "Conceptual Design of a 2 Tesla Superconducting Solenoid for the Fermilab D0 Detector Upgrade", Fermilab Technical Report FERMILAB-TM-1886 (1994).

[23] G. Bernardi *et al.*, "Improvements from the T42 Algorithm on Calorimeter Objects Reconstruction", Internal D0 Note 4335 (2004).

[24] M. Abolins *et al.*, "The Run IIb Trigger Upgrade for the D0 Experiment" IEEE Trans. Nucl. Sci. **51**, 340344 (2004).

[25] V. M. Abazov *et al.* (D0 Collaboration), "Run IIb Upgrade Technical Design Report", Fermilab Pub-02-327-E (2002).

[26] R. Fruhwirth, "Application of Kalman Filtering to Track and Vertex Fitting", Nucl. Inst. Methods **A262**, 444 (1987).

[27] P. V. C. Hough, "Machine Analysis of Bubble Chamber Pictures", Proc. Int. Conf. High Energy Accelerators and Instrumentation (1959).

[28] J. Zhu, "Determination of Electron Energy Scale and Energy Resolution using P14 $Z^0 \rightarrow e^+e^-$ data", Internal D0 Note 4323 (2003).

[29] M. S. Anzelc and J. Hays, "Comparison of em_cert H-Matrix Cuts and Efficiencies in Data and Monte Carlo", Internal D0 Note 5075 (2006).

[30] C. Schwanenberger, "Electron Likelihood Efficiency in p17", Internal D0 Note 5114 (2006).

[31] P. Calfayan *et al.*, "Muon Identification Certification for p17 data", Internal D0 Note 5157 (2007).

[32] G. C. Blazey *et al.*, "Run II Jet Physics", arXiv:hep-ex/0005012 (2000).

[33] E. Busato and B. Andrieu, "Jet Algorithms in the D0 Run II Software: Description and User's Guide", Internal D0 Note 4457 (2004).

[34] D0 JES Group, "Jet Energy Scale Determination at D0 Run II (final p17 version)", Internal D0 Note 5382 (2008).

[35] D. Gillberg *et al.*, "Measuring Jet Response Using the Missing E_T Projection Fraction Method in γ + jet Events", Internal D0 Note 4571 (2004).

[36] G. Bernardi *et al.*, "Improved Limits on WH Production using Neural Network Selection with 1 fb^{-1} of Run IIa Data", Internal D0 Note unnumbered (2008).

[37] G. Bernardi *et al.*, "WH-cut based. Updated analysis without trigger requirements", D0 Collaboration General Higgs Meeting, (2007).

[38] J. Lellouch *et al.*, "WH$\rightarrow \mu\nu$bb, distributions with new triggers", D0 Collaboration General Higgs Meeting, (2007).

[39] J. Pumplin, *et al.*, JHEP **0207**, 012 (2002).

[40] T. Sjöstrand, *et al.*, "High-Energy-Physics Event Generation with PYTHIA 6.1", Computer Phys. Commun. **135** 238 (2001) [arXiv:hep-ph/0010017].

[41] M. L. Mangano *et al.*, "ALPGEN, a generator for hard multiparton processes in hadronic collisions", JHEP **0307**, 001 (2003) [arXiv:hep-ph/0206293].

[42] A. Pukhov *et al.*, "CompHEP: A package for evaluation of Feynman diagrams and integration over multi-particle phase space. User's manual for version 33", (1999) [arXiv:hep-ph/9908288].

[43] S. Höche *et al.*, "Matching Parton Showers and Matrix Elements", (2006) [arXiv:hep-ph/0602031v1].

[44] D. Gillberg, "Heavy Flavour Removal and Determination of Weighting Factors for ALP-GEN W+jet Monte Carlo", Internal D0 Note 5129 (2006).

[45] Z. Sullivan, "Understanding single-top-quark production and jets at hadron colliders", Phys. Rev. D **70**, 114012 (2004) [arXiv:hep-ph/0408049].

[46] Kidonakis *et al.*, "Top quark production at the Tevatron at NNLO", Eur. Phys. J. **C33**, s466-s468 (2004).

[47] Tag-and-probe is a method used to measure efficiencies for an object (the probe) in data by requiring another object (the tag) to satisfy the trigger/selection criteria such that the efficiency of the probe object is (for the most part) unbiased. The canonical example is $Z \to \mu\mu$ events where the event is selected by requiring one high quality muon that fired the trigger (the tag) and a second muon with no quality requirements (the probe) that have a dimuon mass near the Z mass peak.

[48] F. James, "MINUIT Function Minimization and Error Analysis, Reference Manual", http://wwwasdoc.web.cern.ch/wwwasdoc/minuit/minmain.html.

[49] M. Owen and M. Hohlfeld, "Trigger Efficiencies for the OR of Single Electron Triggers in p17 Data", Internal D0 Note 5409 (2007).

[50] M. Voutilainen, "Jet trigger efficiencies for Run IIa", Internal D0 Note 5549 (2008).

[51] N. Makovec and J.-F. Grivaz, "Shifting, Smearing and Removing Simulated Jets", Internal D0 Note 4914 (2005).

[52] M. Sanders, "Multi-Parton Interaction Studies in p17.09 W+jets", D0 Collaboration Joint MC Generators and MC Verification Meeting (2007). http://www-d0.hef.kun.nl//fullAgenda.php?ida=a072093&fid=109

[53] H. Nilsen, "Alpgen multiple parton reweighting", D0 Collaboration Joint MC Generators and MC Simulation Meeting (2007).
http://www-d0.hef.kun.nl//fullAgenda.php?ida=a071637&fid=47

[54] H. Schellman, "The longitudinal shape of the luminous region at D0", Internal D0 Note 5142 (2006).

[55] J. Alwall *et al.*, "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions", arxiv:0706.2569v1 [hep-ph] (2007).

[56] S. Frixione and B. R. Webber, "Matching NLO QCD computations and parton shower simulations", JHEP **0206** 029 (2002) [hep-ph/0204244]; S. Frixione and B. R. Webber, "The MC@NLO 3.3 Event Generator", Cavendish-HEP-06/28, GEF-TH-19/2006 (2006) [arXiv:hep-ph/0612272v1]

[57] I. Narsky, "StatPatternRecognition: A C++ Package for Statistical Analysis of High Energy Physics Data", arXiv:physics/0507143v1 (2005).
http://sourceforge.net/projects/statpatrec

[58] T. G. Dietterich, "An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization", Machine Learning **40** (2) 139 (2000).

[59] L. Breiman, "Random Forests", Machine Learning **45** (2001).
http://www.stat.berkeley.edu/~breiman/RandomForests/

[60] T. K. Ho, "Random Decision Forests", Proceedings of the 3rd International Conference on Document Analysis and Recognition (1995).
http://citeseer.ist.psu.edu/ho95random.html

[61] T. K. Ho, "The Random Subspace Method for Constructing Decision Forests", IEEE Transactions on Pattern Analysis and Machine Intelligence **20** (8) (1998).

[62] L. Breiman, "Bagging Predictors", Machine Learning **24** (1996).

[63] E. Aguilo, *et al.*, "Using Boosted Decision Trees to Search for Single Top Quarks in 1 fb^{-1} of Data", Internal D0 Note 5286 (2007).

[64] T. K. Ho, "A Data Complexity Analysis of Comparative Advantages of Decision Forest Constructors", Pattern Analysis and Applications **5** (2002).

[65] W. Fisher, "Systematics and Limit Calculations", FERMILAB-TM-2386-E (2006).

[66] D0 Collaboration: V. M. Abazov *et al.*, Phys. Rev. Lett. **94**, 151801 (2005); Phys. Rev. D **76**, 111104(R) (2007); arXiv:0808.0703 [hep-ex] (2008).

[67] CDF Collaboration: D. Acosta *et al.*, Phys. Rev. Lett. **94**, 211801 (2005); A. Abulencia *et al.*, Phys. Rev. Lett. **98**, 161801 (2007); T. Aaltonen *et al.*, Phys. Rev. Lett. **100**, 201801 (2008).