Z boson production in association with heavy quark jets at D0

by

Joseph Anthony Zennamo III

October 28, 2013

A thesis submitted to the
Faculty of the Graduate School
of the University at Buffalo, State University of New York
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
Department of Physics

Dedication

While this thesis represents my blood, sweat, and tears it was in no small part the effort of those who have supported me. I would like to give special thanks my advisor, Avto Kharchilava, for giving me the opportunity and support necessary to complete this work. I would also like to thank my partner in crime Ashish Kumar, without the countless hours of conversations and painstaking efforts many of the works contained in this thesis would not have converged. Both of you have made this thesis possible and I will be forever grateful for the sacrifices you have made along the way.

Along those lines I would like to thank the QCD group conveners at D0, Dmitry Bandurin and Leo Bellantoni. There leadership and helpful input throughout the publication of these works was invaluable. Additionally, I have to thank the former b jet identification group conveners Bjoern Penning and Seb Greder. Your mentoring and constant backing has helped me to achieve what I have.

Throughout my time as a graduate student I have had many friends whose friendship and support has kept me sane. Specifically I would like to thank Azadeh Moradinezhad and Chris Redino for their friendly competition and many fruitful and enlightening conversations, you both really helped foster my passions which still burn to this day. Alex Radovic and Carrie McGivern for making Fermilab the home that it is. Andrew Godshalk for riding this wave with me. Kenneth James Smith for helping me hit the ground at a full sprint and really pushing me to be what I have become. Thank you all for your support, your friendship, and for all the fun!

To my family, what can I say? You have taught me what hard work, conviction, friendship, and team work are all about. You have put up with my complaining, my late night texts, and my "falling off the earth" for months on end and all with love and support. Thank you all!

Finally, and foremost, Louise Suter. You have been my rock, you have been my ground, you have been everything I have needed and more. A simple thank you will never suffice, but thank you nonetheless. I love you and this thesis is dedicated to you.

Contents

List of Figures

List of Tables

Abstract

The dominant background in searches for a Higgs boson decaying into b-quarks at the Tevatron is production of a Z boson in association with either b- or c-quark initiated jets (b or c jets). This thesis describes the first measurements of the ratio of differential cross sections $\sigma(Z + b\text{ jet})/\sigma(Z + \text{jet})$, and the first measurements of the ratio of cross sections $\sigma(Z + c\text{ jet})/\sigma(Z + \text{jet})$ and $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$. These measurements are performed using the full D0 Run II data set corresponding to an integrated luminosity of 9.7 fb^{-1}. The ratio of differential cross sections $\sigma(Z + b\text{ jet})/\sigma(Z + \text{jet})$ have been measured as a function of jet and Z boson p_T, jet η, and $\Delta\varphi(Z, \text{jet})$. The $Z + c$ jet ratios of differential cross sections are measured as a function of jet and Z boson p_T.

1 Introduction

When performing searches for new phenomena at hadron colliders, an accurate theoretical description of quantum chromodynamic (QCD) processes is important. These processes mimic the characteristic signature of the signal events being searched for at the Tevatron and Large Hadron Collider (LHC). These include searches for and studies of the Standard Model (SM) Higgs boson. On July 4th, 2012, the ATLAS and CMS collaborations announced the discovery of a Higgs boson at $M_H \approx 125$ GeV [1, 2]. The dominant decay mode of the SM Higgs is $H \rightarrow b\bar{b}$, as shown in Figure 1.1.

At the Tevatron, for $M_H = 125$ GeV, one of the most sensitive channels to the production of a Higgs boson is in association with a Z boson [4], where the Higgs boson decays into b-quarks. This means that the study of the Z boson production in association with heavy flavor (HF) quarks ($Z+$HF jets), where heavy flavor refers to either bottom (b) or charm (c) quarks, is essential for our search efforts [3]. This is because $Z+$ HF jet production forms an irreducible background for the $ZH \rightarrow \ell\ell b\bar{b}$ search, as their final states have an identical signature of a Z boson and at least one b or c quark jet. The size of this background (after final selections) in the D0 $ZH \rightarrow \ell\ell b\bar{b}$ search [5] is shown in Figure 1.2.

This thesis aims to improve our understanding of $Z+$ HF jet production by performing a variety of measurements of the ratio of cross sections, both integrated and differential. Measuring the ratio of cross sections allows for the canceling of many uncertainties allowing for a precise comparison to theory. Three ratios of cross sections $\sigma(Z+\text{b jet})/\sigma(Z+\text{jet})$, $\sigma(Z+\text{c jet})/\sigma(Z+\text{jet})$, and $\sigma(Z+\text{c jet})/\sigma(Z+\text{b jet})$ will be presented. The differential mea-

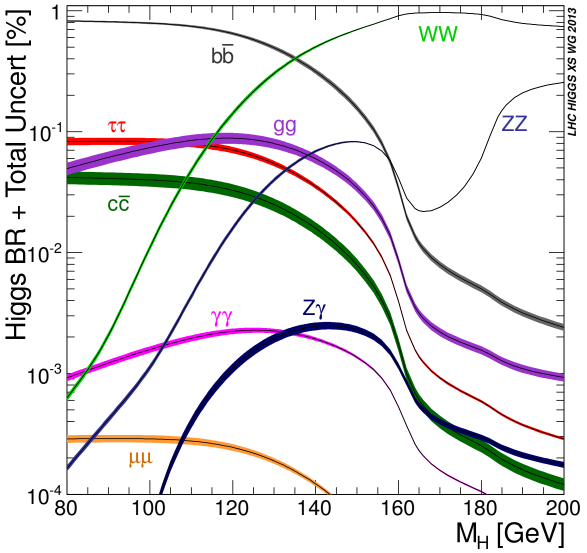


Figure 1.1: The predicted branching fractions for the SM Higgs boson as a function of its mass. Using the knowledge that the mass is $M_H = 125$ GeV we see that its dominant decay mode is $H \rightarrow b\bar{b}$ [3].

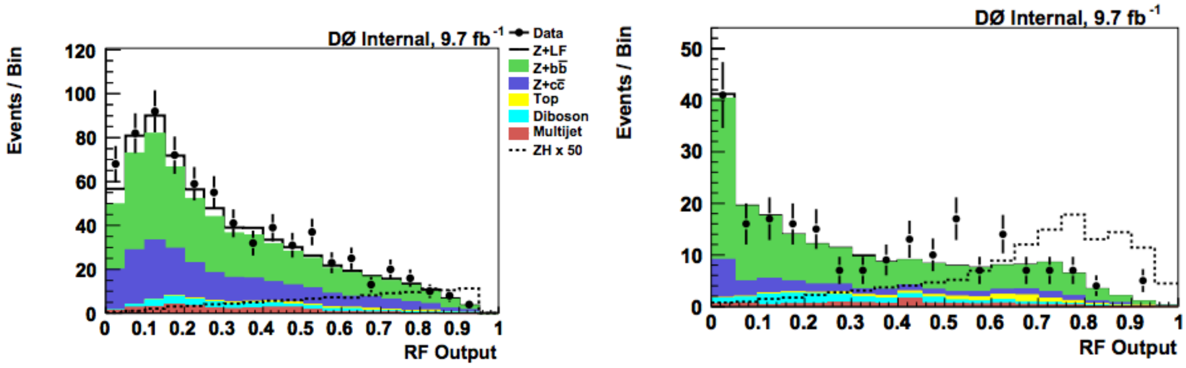


Figure 1.2: The final discriminant output for the $ZH \rightarrow \ell\ell bb$ searches with the requirement of one (left) or two (right) identified b jets. The $Z + b\bar{b}$ (green) and $Z + c\bar{c}$ (dark blue) contributions form large and irreducible backgrounds [5].

surement of $\sigma(Z + \text{b jet})/\sigma(Z + \text{jet})$ has never been performed before and no measurements of $\sigma(Z + \text{c jet})$ have been performed before this measurement.

1.1 Predictions from perturbative QCD

Studies of Z boson production in association with HF jets originating from b or c quarks also provide an important test of perturbative QCD (pQCD) calculations [6]. The leading order (LO) QCD production diagrams of a Z boson with a heavy flavor quark(s) are shown in Figure 1.3. There are two main methods for calculating these processes at next-to-leading order (NLO) accuracy, the four flavor [7] and five flavor [6] schemes. These use different assumptions to calculate the cross section of Z boson plus HF production and are described in the following section.

1.1.1 Z plus HF jet predictions

The fixed (or four) flavor scheme [7] is so named because it considers diagrams which contain only four possible quark-antiquark pairs in the initial state, $q\bar{q} \rightarrow ZQ\bar{Q}$, along with higher order gluon initiated diagrams, $gg \rightarrow Zb\bar{b}$. With this assumption, all events are required to have two real final state b jets. Calculations performed in this scheme have taken the finite b-quark mass ($m_b = 4.62$ GeV) into account during calculations [7]. The uncertainty on the calculations from this scheme is 45% at LO, but reduces to 20% at NLO, dominated by the uncertainty on the renormalization and factorization scale [7].

The variable (or five) flavor scheme [6] loosens the restriction on the initial state quarks and considers all the types of diagrams shown in Figure 1.3, Compton scattering and annihilation. This scheme allows for HF quarks in the initial state coming directly from the proton parton distribution function (PDF). This means diagrams of the type $Qg \rightarrow ZQ$ are

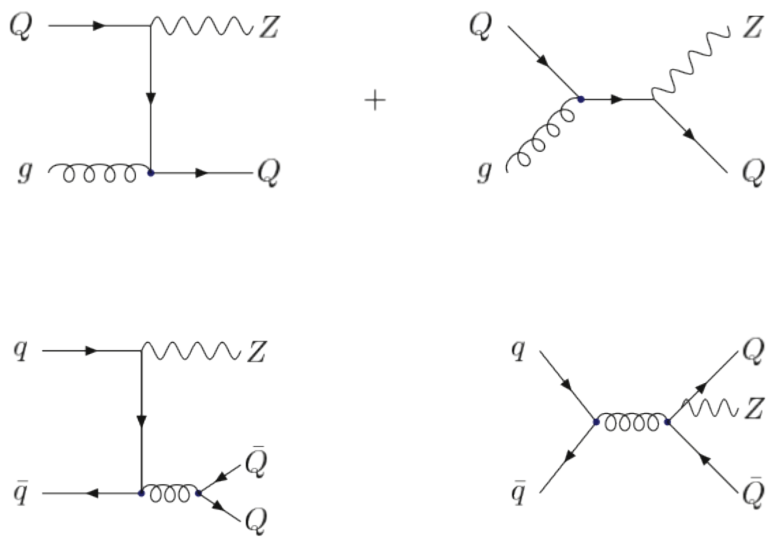


Figure 1.3: Leading order diagrams contributing to $Qg \to ZQ$, or Compton scattering (top row), and $q\bar{q} \to ZQ\bar{Q}$, or annihilation processes (bottom row), where $Q = b, c$ [6].

allowed at LO. Calculations performed using this formalism have assumed that the mass of the b-quarks is negligible ($m_b = 0$), except where the mass is necessary to render the calculations finite [6]. The calculations from these two schemes are compatible within the existing theoretical uncertainties [8].

Due to the low cross section of the $Z+b\bar{b}$ production at the Tevatron (a few picobarns) [6], measurements usually focus on the the production of $Z+ \geq 1\,b$ jet. We compare the measurements presented in this thesis with predictions from the five flavor scheme. The five flavor scheme is implemented into a program known as Monte Carlo for FeMtobarn processes (MCFM) [9]. Using MCFM we can generate a set of NLO predictions for our needs and the selection criteria which we require. These predictions will be used in Chapter 7 and Chapter 8 for comparison to our measurements.

1.2 Previous Tevatron measurements of Z plus b jet production

Due to the fact that $Z + b$ jet production does form a dominant background to Higgs boson searches in $H \to b\bar{b}$ final states it has been studied extensively. None of these analyses have measured the kinematic dependence of the ratio $\sigma(Z + b \text{ jet})/\sigma(Z + \text{jet})$. Furthermore, no analysis to date has measured the rate of $Z + c$ jet production.

The first measurement of $Z + b$ jet production cross section was measured at the D0 experiments, located at the Fermilab Tevatron, in 2004 using 180 pb^{-1} of integrated luminosity [10]. This measurement used a matrix method with inputs from theory to constrain the c jet fraction. They found $\sigma(Z + b \text{ jet})/\sigma(Z + \text{jet}) = 0.023 \pm 0.005$ and was in good agreement with NLO. This was followed by a measurement by the CDF collaboration in 2006 using an integrated luminosity of 330 pb^{-1} [11].

Instead of using a matrix method they constructed a procedure independent of inputs from NLO theory. They extracted the relative contributions of the flavor content from the data using a template fit of a discriminating variable. They proceeded to measure the absolute $Z+b$ jet cross section along with the ratio of cross sections $\sigma(Z+b \text{ jet})/\sigma(Z+\text{jet}) = 0.0235 \pm 0.009$. This measurement is in agreement with the earlier D0 result. The next measurement by CDF, in 2009 [12], used the same methods as laid out in the first CDF paper. CDF repeated the analysis with an integrated luminosity of 2 fb^{-1} and measured $\sigma(Z + b \text{ jet})/\sigma(Z + \text{jet}) = 0.0208 \pm 0.0047$, which again agreed with past measurements and NLO predictions.

Finally, in 2010 D0 [13] adopted the flavor extraction method used by CDF but built a more discriminating variable to help extract the overall flavor fractions from the data. Using this new variable, which will be utilized in this thesis (see Chapter 6), D0 was able to measure $\sigma(Z + b \text{ jet})/\sigma(Z + \text{jet}) = 0.0193 \pm 0.0027$ which stood as the most precise measurement of the integrated ratio and agreed with past measurements and the NLO predictions.

Table 1.1: The number of gluons which split into heavy flavored quarks, $\bar{n}_{g\to Q\bar{Q}}$, as measured in LEP and SLD. These measurements look for gluon radiation off $Z \to q\bar{q}$ decays and look at the flavor of the resulting $g \to Q\bar{Q}$ pair [23].

	$\bar{n}_{g\to c\bar{c}}$ (%)	$\bar{n}_{g\to b\bar{b}}$ (%)
ALEPH [15, 16]	$3.26 \pm 0.23 \pm 0.42$	$0.277 \pm 0.042 \pm 0.057$
DELPHI [17]		$0.21 \pm 0.11 \pm 0.09$
L3 [18]	$2.45 \pm 0.29 \pm 0.53$	
OPAL [19]	$3.20 \pm 0.21 \pm 0.38$	
SLD [20]		$0.307 \pm 0.071 \pm 0.066$
Theory [24]		
$\Lambda^{(5)}_{\bar{M}S} = 150$ MeV	$1.35^{+0.48}_{-0.30}$	0.20 ± 0.02
$\Lambda^{(5)}_{\bar{M}S} = 300$ MeV	$1.85^{+0.69}_{-0.44}$	0.26 ± 0.03

1.3 Previous measurements with c jets

While there have been no measurements of Z boson production in association with c jets, there have been related measurements have provided us with hints of what we might find. First, measurements from LEP [14, 15, 16, 17, 18, 19] and SLD [14, 20] experiments have yielded disagreements with the predictions for gluons splitting into c-quark pairs [21, 22], $g \to c\bar{c}$. This has been documented in Table 1.1 and points to an underestimation by almost a factor of two. Another place a discrepancy has been seen is in measurements of $\gamma + c$ jet production. Both D0 and CDF have measured the cross section of $\gamma + c$ jet events [25, 26] and found that predictions underestimated the total cross sections. In these analyses the contributions from gluons splitting into c jets was enhanced in order for the predictions accurately model the results. This enhancement was derived by varying the $g \to c\bar{c}$ rate and minimizing a χ^2 fit of these predictions to the data [25].

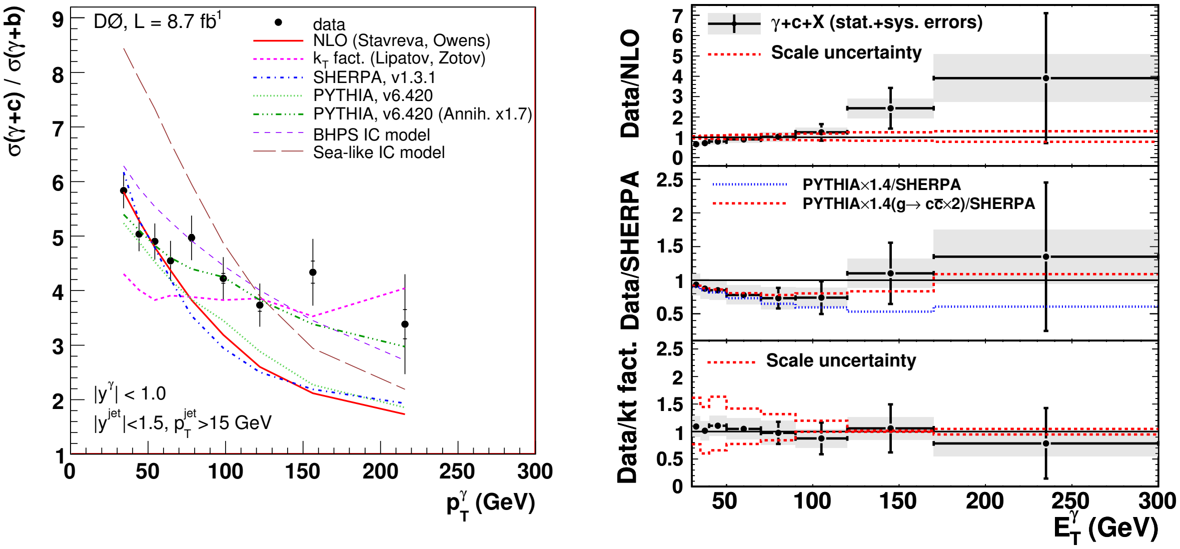


Figure 1.4: Measurements of $\gamma + c$ jet cross sections by D0 [25] (left) and CDF [26] (right). D0 measured the ratio of differential cross section $\sigma(\gamma + c \text{ jet})/\sigma(\gamma + b \text{ jet})$ as a function of photon p_T. While, CDF measured the absolute cross section $\sigma(\gamma + c \text{ jet})$ and has displayed this measurement normalized to various predictions as a function of the transverse energy of the photon. To correctly model the measurements the $g \to c\bar{c}$ contributions had to be enhanced.

2 Experimental setup

To measure the relevant cross sections and test the NLO pQCD calculations we will be using the D0 detector located at the Fermi National Accelerator Laboratory (Fermilab). This chapter will discuss the accelerator, the D0 detector, and its subsystems.

2.1 Fermilab accelerator facility

The Fermilab accelerator complex [27], shown in Figure 2.1, is used to accelerate protons and antiprotons to a center of mass energy of $\sqrt{s} = 1.96$ TeV. Creating particles at such high energies requires a multistep procedure that begins with the Cockcroft-Walton accelerator.

The Cockcroft-Walton takes hydrogen ions (H^{-}) and accelerates them to an energy of 750 keV [29]. These hydrogen ions are then passed into a linear accelerator, or Linac, which uses radio frequency (RF) cavities to accelerate the hydrogen ions to 400 MeV [29]. The Linac can be tuned for a variety of purposes, from continuing on to other parts of the accelerator complex, or to the Neutron Therapy Facility for treating cancer, or even to the muon cooling facility where research into muon acceleration is ongoing. Once the hydrogen ions have been delivered to the Linac they are passed through a carbon foil which removes the electrons orbiting the nuclei of the atoms [27], thus leaving protons which are then injected into the first synchrotron of the acceleration chain, the Booster.

The Booster uses an array of 19 RF cavities spaced around a 75 m circular accelerator to bring the protons from their initial 400 MeV to an energy of 8 GeV [30]. These protons are

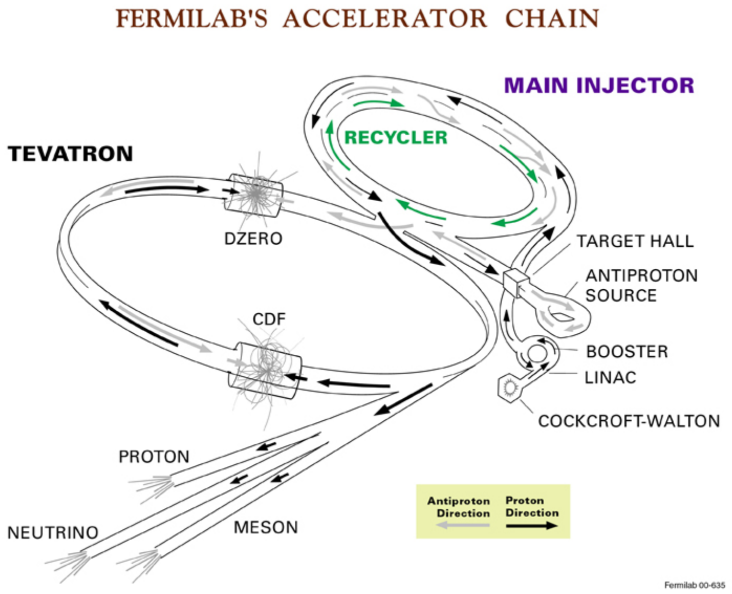


Figure 2.1: The Fermilab accelerator complex which is used to accelerate hydrogen ions from rest to create a proton-antiproton collision at a center of mass energy of $\sqrt{s} = 1.97$ TeV [28].

then passed into the Main Injector which is used for two purposes, accelerating protons for injection into the Tevatron and creating energetic protons for use in the Antiproton Source.

To create antiprotons the Main Injector accelerates two bunches of protons to 120 GeV and merges them into a single bunch [31]. This new bunch is then passed into the Antiproton Source where they collide with a nickel target. The resulting spray of particles is sorted using an array of magnets which select 8 GeV antiprotons [31]. These antiprotons are then passed into the Debuncher, a rounded triangular synchrotron, which uses stochastic cooling [31] to control the momentum spread of the beam. While no actual acceleration takes place in the Debuncher it prepares the beam for acceleration and then passes it back to the Main Injector.

At this point the Main Injector takes the 8 GeV protons and antiprotons and accelerates them to their injection energy of 150 GeV [32]. From this the protons and antiprotons enter the final stage of acceleration, the Tevatron [33]. The Tevatron is a circular ring with a circumference of 4 miles which uses superconducting niobium-titanium magnets, cooled to roughly 4 K, to accelerate the bunches to a collision energy of 980 GeV. Once this energy is reached collisions occur at two collision points along the Tevatron ring, where the CDF and D0 detectors are located [27]. Each collision cycle, or store, occurs for roughly 24 hours.

2.2 D0 detector

The proton-antiproton collisions which occur at D0 are surrounded by a general purpose detector, the D0 detector [34]. The analyses discussed in this thesis makes use of all of the subsystems contained within the D0 detector. Figure 2.2 shows the detector with its major subsystems highlighted. A person is located to the right to give a sense of scale of the detector. The pseudorapidity (described below) coverage for the subsystems is marked on the figure. Descriptions of the important subsystems are documented below.

D0 uses a standard right-handed coordinate system. The nominal collision point is at the center of the detector with coordinate $(0, 0, 0)$. The direction of the proton beam is the $+z$ axis. The $+x$ axis is horizontal, pointing away from the center of the Tevatron ring. The $+y$ axis points vertically upwards. The polar angle, θ, is defined such that $\theta = 0$ is the $+z$ direction. The rapidity is defined as $y = -\ln[(E + p_Z)/(E - p_Z)]$, where E is the energy and p_Z is the momentum component along the proton beam direction. Pseudorapidity is defined as $\eta = -\ln(\tan\frac{\theta}{2})$. φ is defined as the azimuthal angle in the plane transverse to the proton beam direction. Due to momentum conservation is known that the vectorial sum of all particle momenta transverse to the beam direction should equal zero. For this reason we will measure the momentum of all particles transverse to the beam direction, p_T.

2.2.1 Tracking system

The tracking system at D0 is of extreme importance to the analyses presented in this thesis. The algorithms used to identify HF jets rely on the information about the tracks which comprise the jets and on the vertices which are displaced from the initial hard interaction point. A schematic of the tracking system can be found in Figure 2.3 which shows the

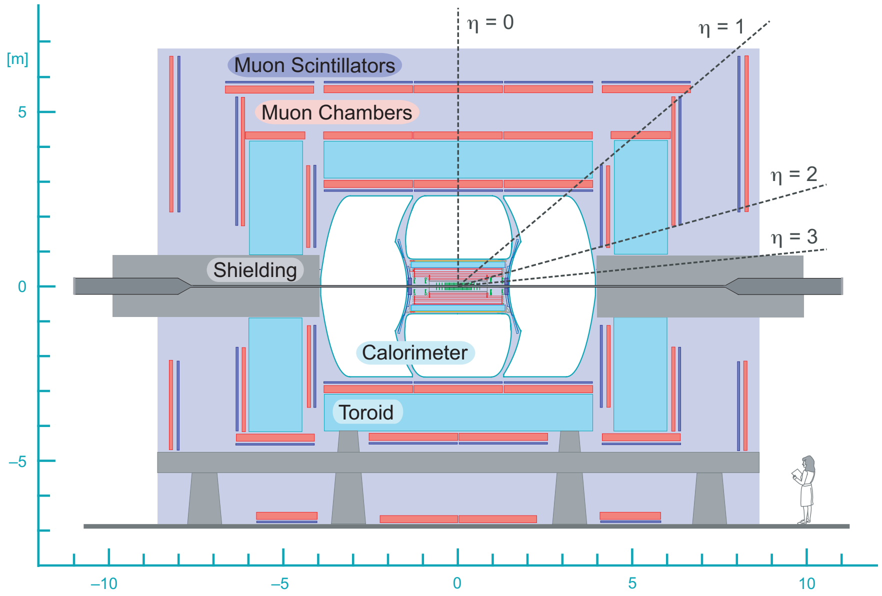


Figure 2.2: A diagrammatical representation of the D0 detector. The axes are marked in meters [35].

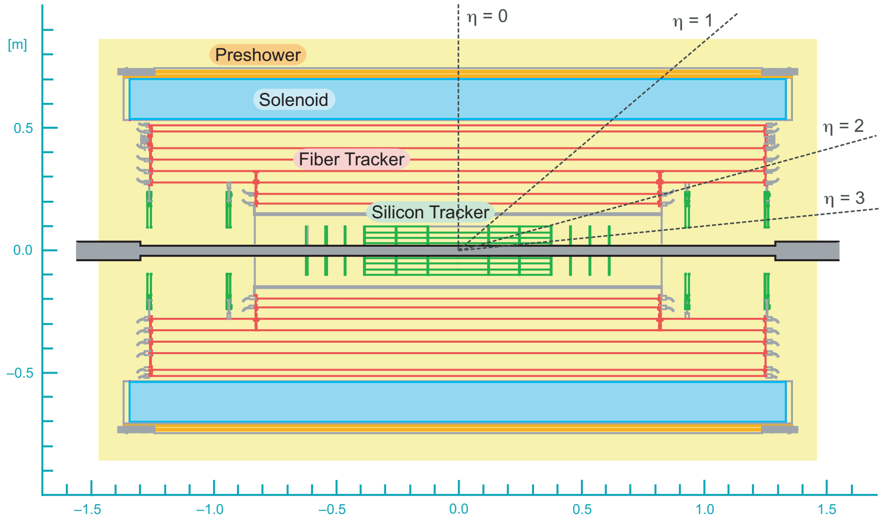


Figure 2.3: A schematic of the D0 tracking volume which is composed of two subsystems, the SMT and CFT. The solenoid is also shown surrounding the tracking volume [35].

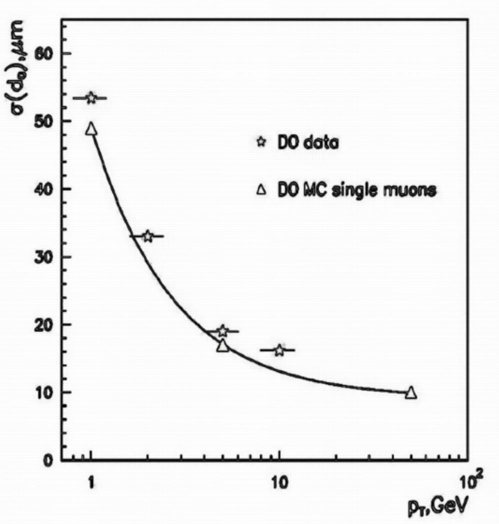


Figure 2.4: The single particle IP resolution as a function of a particles transverse momentum [38].

innermost silicon tracking system and the scintillating fiber tracker surrounding it. Both are immersed in a 1.9 T magnetic field produced by a superconducting solenoid that provides a uniform magnetic field throughout the tracking volume.

The Silicon Microstrip Tracker [36], or SMT, is the closest detector to the primary interaction vertex, PV. The inner-most part of this detector sits just 1.61 cm from the beam [37]. Due to its location this detector is very important for the selection of the PV. The SMT can determine the PV location with an accuracy of 25 μm in the beam direction and 15 μm in the direction tangent to the beam. It also allows us to determine the position of the tracks, measured from charged particles traversing the detector medium, relative to the PV. The shortest distance between the track and the PV is known as the impact parameter, IP. Using the SMT this displacement can be determined to within 54 μm and 16 μm for single particle tracks with a transverse momentum of 1 GeV and 10 GeV, respectively [38], as can be seen in Figure 2.4.

The outer portion of the tracking volume is the Central Fiber Tracker [34], or CFT, which is composed of radiation-hard scintillating fiber arrayed in eight concentric cylinders. The CFT has readout times on par with the bunch crossing time. This fast readout time of

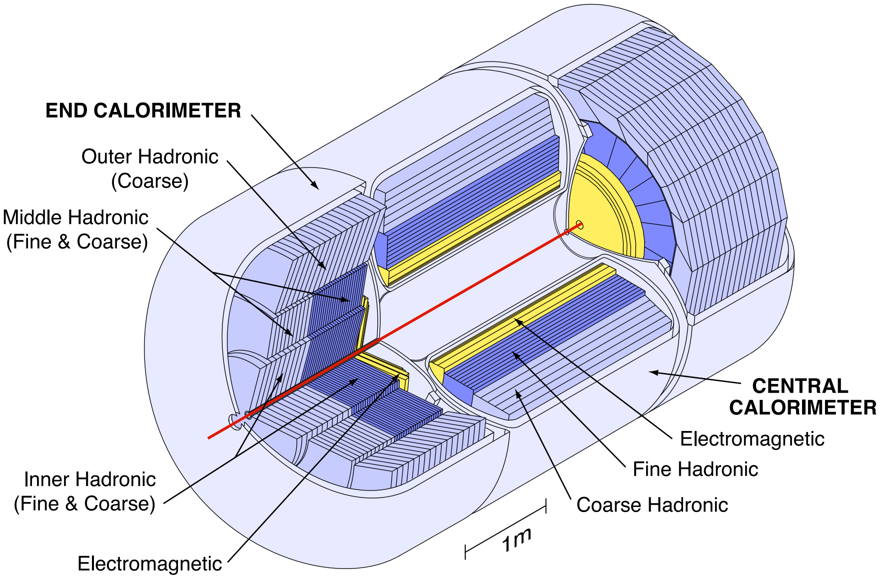


Figure 2.5: A diagram of the D0 calorimeter [34].

material makes it an excellent detector for reading out tracking information.

2.2.2 Calorimeter

D0 employs a sampling calorimeter to determine the energy of the particles and groups of particles which traverse this medium [39]. In Figure 2.5 we see the calorimeter which encases the tracking system and sits just outside the solenoid. The sampling calorimeter technology utilizes materials of different densities to determine the energy of the traversing particles. At D0 this is done with alternating layers of active liquid argon, copper readout pads, and passive pads. Particles create dense showers when they impinge upon the bulk of the passive pads. The showers will then ionize the liquid argon as they traverse the volume. The free electrons are read out by active copper pads which are kept at a positive potential.

The calorimeter is separated into three main parts: a central calorimeter (CC), which covers the region $|\eta| < 1.1$ and two end cap calorimeters (EC) which cover the region of $1.5 < |\eta| < 4$. Each section of these calorimeters is separated into the electromagnetic (EM) calorimeter, which uses passive plates of depleted uranium, and the fine and course hadronic calorimeters. The fine and coarse hadronic calorimeters use uranium-niobium alloy

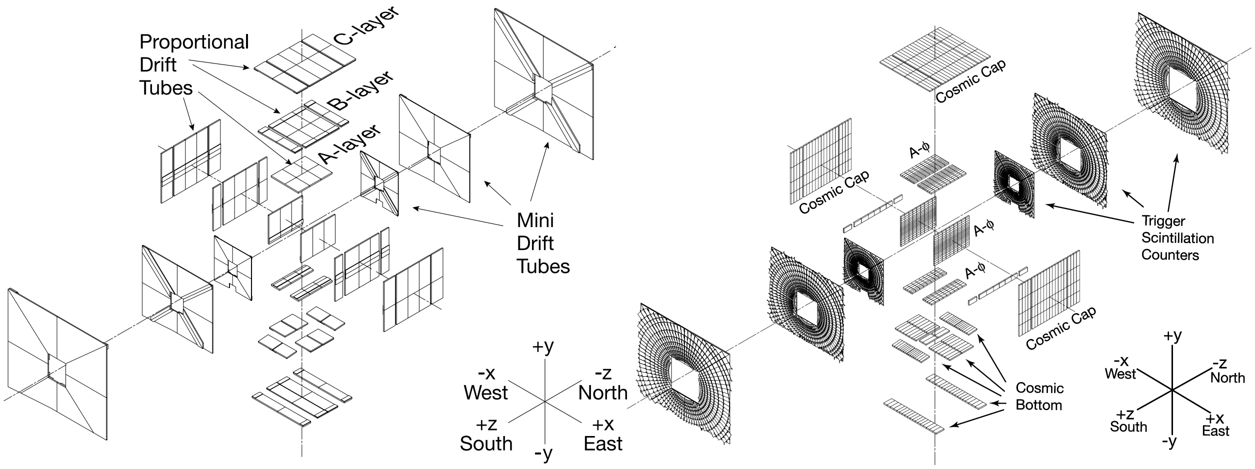


Figure 2.6: An exploded view of the D0 muon system with the PDTs, on the left, and the scintillation counters, on the right [34].

and stainless steel as the passive medium, respectively. Separate cryostats are needed to keep the liquid argon condensed. This leads to a gap in the coverage where the cooling system enters the detector. This is known as the inter-cryostat region, or ICR, and covers the range of $1.1 < |\eta| < 1.5$. To help deal with the limited coverage in this region 16 scintillating tiles are added to match the granularity of the EC.

2.2.3 Muon system

The outermost D0 subsystem is used to account for one of the two particles which do not get absorbed in the calorimeter, the muon [40]. The other particle which escapes detection is the neutrino. The muon system sits around a second toroidal 1.8 T magnet which is used to help determine the momentum of outgoing muons. Both proportional drift tubes (PDTs) and scintillation counters, as shown in Figure 2.6, are used to detect muons. PDTs consist of a charged wire suspended in a volume of gas, 84% argon, 8% methane, and 8% CF_4 [40]. As the muon passes through the volume it ionizes the gas, and the free electrons are collected by the wire. This has a fast readout and is used to trigger events. This system is broken into three layers: A, inside the toroidal magnet, and B and C, which sit outside. The scintillation counters are synchronized with the Tevatron's clock to allow any muon which is detected out of time with a bunch crossing to be rejected, as originating from a cosmic ray.

2.2.4 Triggering

The Tevatron produces $p\bar{p}$ collisions at a rate of 7 MHz which is beyond our ability to record [41]. The fastest that data can be written to magnetic tape is 50 Hz. While many of the events which occur during collisions are important to record, many are more of less interest and need not be recorded. To deal with these two realities a triggering system has been constructed which, in various steps, uses topology of events to reduce the number of events we record to a more manageable 50 Hz.

At D0 a three step triggering system is used. The first level (L1) is based on hardware readouts. The hardware readouts come from the "fast" detector subsystems including the CFT, the calorimeters, and the muon scintillation counters. The decision about whether to reject an event is made in less than 4.2 μs. This reduces the overall rate of 7 MHz to around 50 kHz while enriching the overall data sample in events of interest.

The next layer of the triggering system is the L2, this uses a combination of hardware and software information to select events [41]. The software applies low level reconstruction to determine which objects were produced in the initial hard interaction. The L2 selects an event in 100 μs and reduces the rate further to roughly 1 kHz.

The final layer of the trigger before the events are recorded to tape is the L3 which uses a simplified/fast reconstruction of the whole event to determine if the event is one of interest [41]. This happens in about 50 ms and reduces the final output to a manageable 50 Hz. The selected events are recorded to tape and can be used to complete the following analyses.

3 Data and event selection

The data for these analyses were collected from the Fermilab Tevatron $p\bar{p}$ collisions at $\sqrt{s} =$ 1.96 TeV using the D0 detector, between March 2001 and September 2011 during Run II of the Tevatron and corresponds to an integrated luminosity of 9.7 fb^{-1}. Events are selected by requiring either two muons or two electrons [42]. Muons (electrons) must have $p_T > 10$ GeV (12 GeV). The efficiency for selecting these events is 80% and 100%, respectively [43, 44]. The lower efficiency for selecting muon events originates from limited muon system coverage at the bottom of the detector. Events must also pass data quality requirements. These remove runs and luminosity blocks which are *flagged* by the SMT, CFT, calorimeter, and muon system groups as bad [45]. A run is considered a "bad run" when a specific subsystem is not functioning or has an issue which decreased our confidence in the recorded data. In addition, simulations can be flagged as a "bad run" due to the *zero bias overlays* [46], described in Section 4.3.1.

3.1 Z boson plus jet event selection

These analyses select events which contain a Z boson candidate reconstructed from dilepton decays, $Z \to \mu\mu$ or $Z \to ee$, and at least one hadronic jet. Different methods are used to select events for the different decay channels, while the jet selection is uniform for all events.

3.1.1 Primary vertex reconstruction

To begin selecting $Z+$jet events we must first determine where the PV is located [47]. The reconstruction and identification of the PV at D0 consists of the following steps: (i) selection of tracks with $p_T^{trk} > 0.5$ GeV and which are less than 2 cm apart in the z direction when extrapolated to the PV; (ii) vertex fitting using a Kalman filter algorithm to obtain a list of candidate vertices; (iii) a second vertex fitting iteration using an adaptive algorithm to reduce the effect of outlier tracks; and (iv) selection of the vertex with the lowest probability of originating from a soft underlying event [47]. To select an event, its PV is required to have at least three associated tracks and be reconstructed with a z position within 60 cm of the center of the detector ($z = 0$) along the direction of the beam.

3.1.2 Muon selection

The $Z \to \mu\mu$ event candidates are required to contain at least two muons which have the following requirements applied, increasing the chance that they originated from a Z boson:

- Both muons must have $p_T > 15$ GeV and $|\eta| < 2.0$.
- Both muons must pass the `Loose` muon identification requirements, as defined in Ref. [43].
- The muon object must have a central track pointing towards it.
- A distance of closest approach, *dca*, to the PV of less than 0.04 cm for tracks with SMT hits and less than 0.2 cm for tracks without any SMT hits.
- $\Delta z\,(PV, \mu) < 1$ cm, where $\Delta z\,(\mathrm{PV}, \mu)$ is the distance between the PV and the muon track along the z-axis.

The `Loose` muon criteria requires the muon leaves hits in both the tracking and muon systems as defined in Ref. [43].

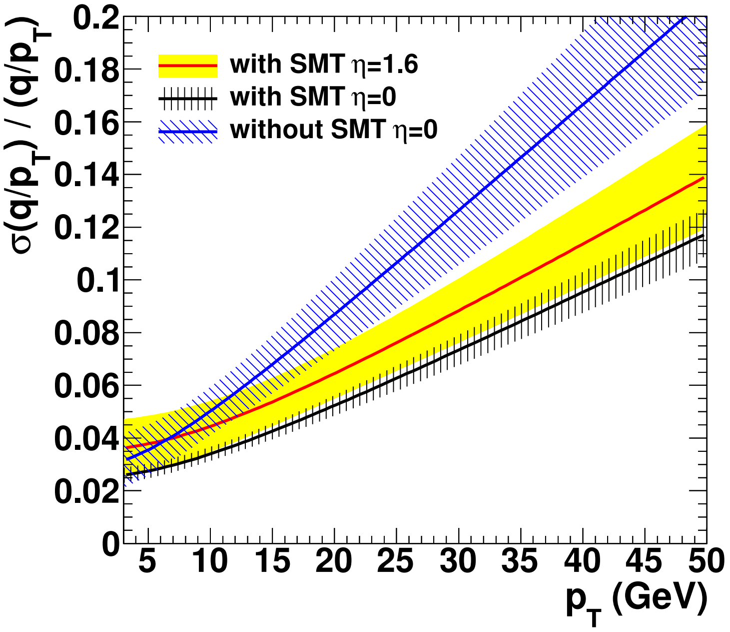


Figure 3.1: The single muon p_T resolution as a function of p_T for muons that either have hits in the SMT or not [43].

If the track associated with the muon has no hits in the SMT the muon p_T is re-measured using the PV as an additional hit for the track. This increases the length of the track and gives additional information about the curvature of the track for measuring the momentum. This requirement leads to an overall single muon momentum resolution that depends on whether we have hits in the SMT or not, as shown in Figure 3.1.

Reconstructing Z boson decaying to muons

A Z boson candidate is required in each event, reconstructed from a pair of selected muons. The muons used to reconstruct a Z boson have additional requirements applied:

- The dimuon invariant mass must be within a window of $70 < M_{\mu\mu} < 110$ GeV.
- A "cosmic veto" is used to suppress muons originating from cosmic rays, $[\pi - \Delta(\phi_{\mu 1}, \phi_{\mu 2})] + |\pi - (\theta_{\mu 1} + \theta_{\mu 2})| > 0.05$.
- The muons are required to be of opposite electric charge.

- *Product scaled isolation* [48], $I(\mu_1)I(\mu_2) < 0.03$,

where the "product scaled isolation" variable is defined as the product of $I(\mu_i)$ of the two muons which compose the Z boson candidate, where

$$I(\mu_i) = \frac{E_T^{cal}(0.1 - 0.4) + p_T^{trk}(0.5)}{p_T^i} \tag{3.1}$$

and $E_T^{cal}(0.1 - 0.4)$ is the transverse energy of the calorimeter inside a hollow cone of $0.1 < \Delta R < 0.4$ around the muon, where $\Delta R = \sqrt{\Delta\varphi^2 + \Delta\eta^2}$; $p_T^{trk}(0.5)$ is the vector sum of track transverse momenta for all tracks inside a cone $\Delta R < 0.5$ around the muon; and p_T^i is the transverse momentum of the i^{th} muon.

3.1.3 Electron selection

The dielectron channel selects events where one electron must be reconstructed in the CC region of the calorimeter, while the second electron can be located either in the CC or the EC, as defined in Section 2.2.2. All electrons which are selected are required to have $p_T > 15$ GeV. To reconstruct a Z boson candidate it is necessary to restrict the dilepton mass to be within a Z boson mass window of $70 < M_{ee} < 110$ GeV.

The following criteria, as defined in Ref. [44], maintains the highest selection efficiency for electrons while suppressing the large multijet background (discussed in Section 4.2). For electrons reconstructed in the CC region we require they pass certain identification and isolation criteria:

- An electromagnetic (EM) cluster must be found in the calorimeter and isolated from other activity by $f_{iso} < 0.09$.

- Energy fraction in EM calorimeter $f_{EM} > 0.9$.

- Tracks associated with the electron must be isolated by requiring the scalar sum of all track p_T in an annulus of $0.04 < \Delta R < 0.4$ surrounding the electron candidate, `IsoHC4`, must be less than 4 GeV.

- `NNout7` > 0.4
- `TrkMatchChi2` > 0.0 or `EMHits_e_f` > 0.6

For electrons located in the EC region a separate set of criteria is used:

- Isolated EM cluster $f_{iso} < 0.05$
- Energy fraction in calorimeter $f_{EM} > 0.97$
- `HMx8` < 10, where `HMx8` is a matrix composed of eight shower shape variables. This takes advantage of the transverse and longitudinal shape of the calorimeter shower to differentiate electrons from hadrons [39].
- `IsoHC4` < 200 GeV
- `NNout4` > 0.2
- `Sigphi` < 100

The `NNout<N>` are outputs of neural networks with `<N>` input variables. These variables can include the energy deposited in the calorimeter, number of cells in the first layer of the calorimeter where energy is deposited, and the track isolation variables [44]. `TrkMatchChi2` is the χ^2 between the track and the electromagnetic shower in the calorimeter. `EMHits_e_f` is a variable which quantifies the number of CFT and SMT hits along the track leading to the EM object's calorimeter shower. Finally, `Sigphi` is the width of the shower in the third layer of the EM calorimeter (EM3), which is the layer with the highest granularity [44]. This variable is defined by

$$\text{Sigphi} = \Sigma \log\left(\frac{E^i_{cell}}{E_{EM3}}\right) \times \left(-\sin\varphi_{EM}\cdot x^i_{cell} + \cos\varphi_{EM}\cdot y^i_{cell}\right)^2,$$

where E^i_{cell}, x^i_{cell}, and y^i_{cell} are the energy and distance along the x and y directions of the i^{th} calorimeter cell, and E_{EM3} and φ_{EM} are the total energy and the azimuthal angle of the energy shower in the EM3.

3.1.4 Jet reconstruction

Jets are reconstructed using the D0 Run II cone algorithm [49] with a cone size of $\Delta\text{R} = 0.5$. This algorithm is also used to reconstruct jets for the prediction of the ratios of cross sections which we compare the results to, discussed in Section 7.6 and Section 8.7. Figure 3.2 shows the resolution of the jets reconstructed in the data as a function of y and p_T. At least one jet is required in each event with the following requirements:

- The highest p_T jet must have $p_T > 20$ GeV.
  - If used, the second highest p_T jet must also have $p_T > 20$ GeV.
- All selected jets are required to be in $|\eta| < 2.5$.
- Jets are required to originate from the PV to suppress jets originating from secondary interactions, known as "pile up."
- All jets must have at least two tracks which are matched to the PV; these are known as "vertex confirmed" jets [51].

Jet energies are corrected in both data and simulation using the jet energy scale (JES) [52]. The goal of these corrections is to relate the measured final state energy of the jet to its true energy by accounting for calorimeter effects (radioactive decay of the uranium, lack of instrumentation in the ICR region, etc.), pile up effects, the varying responses of the different particles which compose the jets, and effects related to the jet reconstruction. These corrections are determined by balancing the energy deposited on either side of back-to-back $\gamma+\text{jet}$ events (events with $\Delta\varphi(\gamma, \text{jet}) > 2.5$). This balancing is parameterized in jet p_T and η and leads to a correction to data and simulations. An example of this correction which

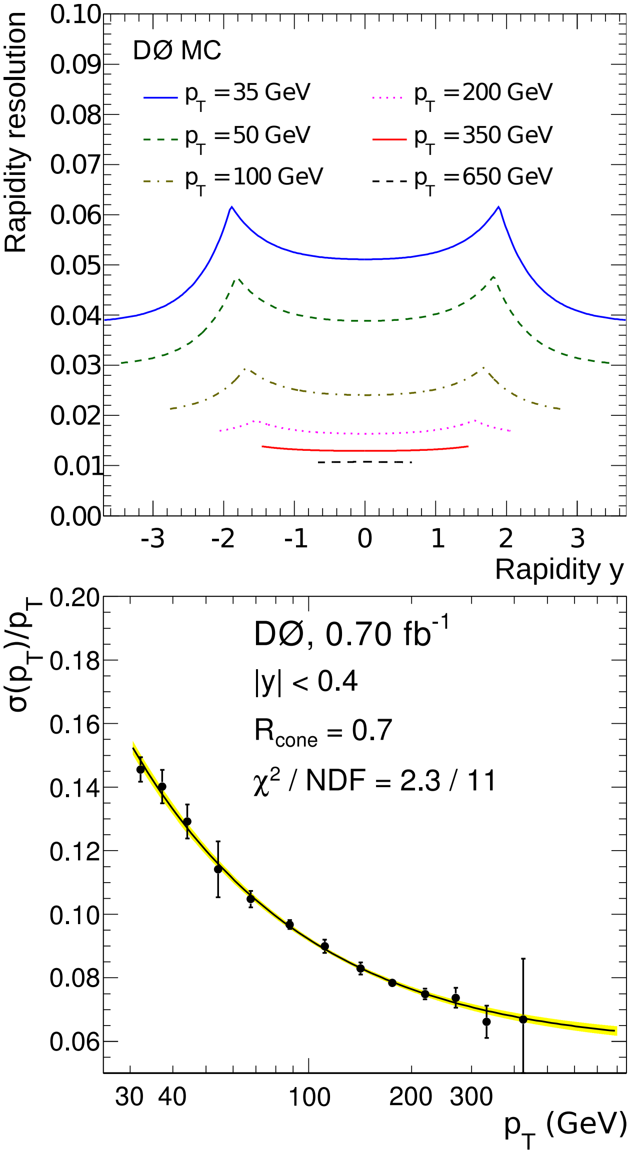


Figure 3.2: The jet rapidity y (top) and p_T (bottom) resolution for the Run II cone algorithm in the D0 detector [50].

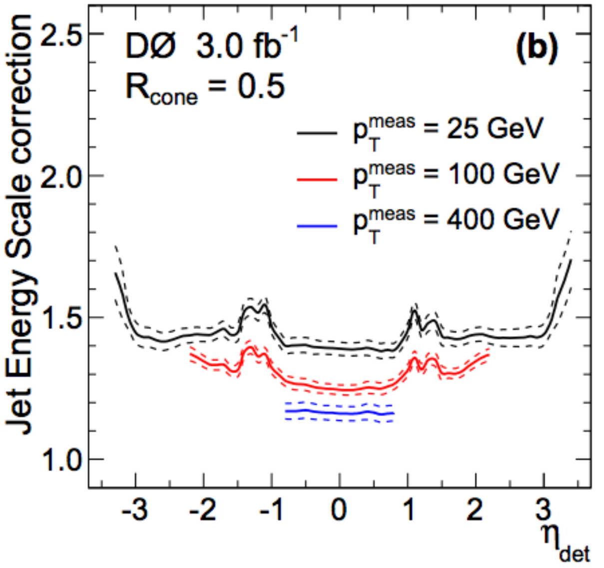


Figure 3.3: The jet energy scale correction as measured in γ + jet events for data (left) and simulation (right) for three different choices of jet p_T [52].

was derived for a 3 fb^{-1} subset of the data is shown in Figure 3.3. An additional flavor-dependent correction can be applied to jets in simulation to account for the difference in the single particle responses for the initial state parton flavor. This correction is derived for the differences between b, gluon, and light-quark jets, as shown in Figure 3.4.

There is an additional correction derived for the jets which are matched ($\Delta R < 0.5$) to a muon. This corrects for energy lost due to the muon and neutrino originating from the decay of hadrons inside the shower [53]. All jets are required to be isolated from the high

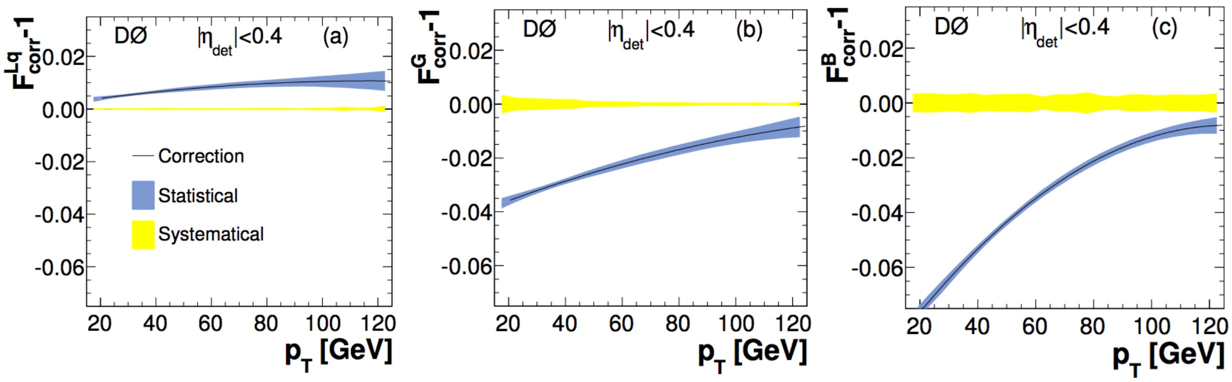


Figure 3.4: The flavor dependent jet energy scale correction for simulations as derived for jets originating from light-partons (left), gluons (center), and b-quarks (right) [52].

p_T electrons and muons which pass the Z boson criteria by $\Delta R > 0.5$.

To suppress the contributions from top quark pair production ($t\bar{t}$) events with a significant imbalance in the measured transverse momentum, $\not{E}_T > 60$ GeV, due to undetected neutrinos from the W boson decay ($t \to Wb \to \ell\nu b$) are rejected.

3.1.5 Z boson plus jet data sample

After the application of these criteria a total of 644,278 (605,633) $Z+$jet candidate events are selected in the $Z \to \mu\mu$ ($Z \to ee$) channel. These are composed of $Z+$light jet, $Z+c$ jet, and $Z+b$ jet, but there are additional backgrounds which mimic the signature of a $Z+$jet event. Various other processes can produce two isolated high p_T leptons with at least one jet. Chapter 4 details the methods for generating simulated events, utilizing orthogonal data samples to model instrumental effects, and then correcting any deficiencies in these models.

4 Event modeling

4.1 Simulated events

Processes such as diboson (WW, WZ, ZZ) production can contribute to the background when their reconstructed final state includes two leptons. Inclusive diboson production is simulated with the PYTHIA [54] MC event generator. The Z + light flavored jet events, Z + HF jet events, and $t\bar{t}$ events are modeled by ALPGEN [55], which generates hard subprocesses including higher order QCD tree-level matrix elements, interfaced with PYTHIA for parton showering and hadronization. Due to this procedure it is possible that there can be an overlap between partons which are generated by the matrix element and those created during the showering. To avoid this double counting the MLM matching scheme is used [56]. The CTEQ6L1 [57] PDFs are used for event modeling. The cross sections of the simulated samples are scaled to the corresponding higher-order theoretical calculations using the MCFM program [9] with the CTEQ6.1M PDF set [57].

4.1.1 Flavor assignment in simulated events

To aid in simulating events it is important that we know the flavor of the simulated jets [58]. To determine the 'true' flavor of a simulated jet we attempt to match jets to a HF hadron at the particle level, by requiring that $\Delta R(\text{jet}, \text{hadron}) < 0.5$. If the jet contains a B hadron it is flagged as a b jet. If no B hadron is contained within the jet, but a C hadron can be matched then it is defined as a c jet. This sequence guards against cases where a b-quark fragments to a c-quark. The remaining jets, which do not contain B or C hadrons, are defined as light jets.

4.1.2 Z boson plus jet cross sections

The normalization of the inclusive Z boson cross sections in simulations are determined by taking the ratio of the ALPGEN cross section to Next-to-Next-to-Leading-Order (NNLO) cross section calculations [59]. This normalization is applied to all events which contain a Z boson, including all $Z+$jet events, and the correction is $k' = 1.30$. This factor has an uncertainty of 6% coming from both ALPGEN and the NNLO calculations [59].

A second correction is applied to the cross sections of the Z + HF jet events. These cross sections are corrected using the NLO calculations from MCFM [9]. To correct the cross sections of the Z + HF jets we take the ratio of the MCFM NLO predictions to the LO predictions, denoted as $HF_{b\bar{b}(c\bar{c})}$. The ALPGEN $Z+$HF cross sections are then scaled by this additional factor for a total scaling of $k' \times HF_{b\bar{b}} = 1.30 \times 1.52 = 1.96$ and $k' \times HF_{c\bar{c}} = 1.30 \times 1.67 = 2.15$. These scale factors are not used when performing the ratio of cross section measurements in Chapter 7 and Chapter 8, but for illustrating the data to simulation (dis-)agreement.

4.1.3 Cross sections for background processes

The diboson and $t\bar{t}$ production cross sections are scaled to NNLO, evaluated using MCFM. The resulting factors are: $k'(ZZ) = 1.030$, $k'(WZ) = 1.062$, $k'(WW) = 1.005$, and $k'(t\bar{t}) = 1.434$, with uncertainties on k'(diboson) of 7% and $k'(t\bar{t})$ of 10% taken from MCFM. The uncertainty on the $t\bar{t}$ cross section is close to the current D0 experimental uncertainty [59, 60] and is the largest cross section reweighting uncertainty.

4.2 Multijet background

The Tevatron's $p\overline{p}$ collisions produce an enormous number of multijet events and some of these jets can be mistakenly identified as leptons. This instrumental multijet background is not well-modeled by simulation, and is determined from data by using a selection of events from an orthogonal data set.

The criteria used to select these multijet events are chosen such that the kinematic properties of the events in the multijet control sample are close to those of the events in the analysis sample, while still providing sufficient statistics. In the dimuon channel, a multijet event must contain a false Z boson candidate which passes all other event selection requirements but fails the product scaled isolation requirement, as described in Section 3.1.2. For the dielectron channel, a multijet event must have at least two fake electrons with `IsoHC4` < 4.0 in CC or `IsoHC4` < 100.0 in EC, `TrkMatchChi2` > 0.0 in CC, `EMHits_e_f` > 0.6 in CC, `NNout7` < 0.1 in CC or `NNout4` < 0.4 in EC (as described in Section 3.1.3).

For each lepton channel, the overall size of each multijet sample must be scaled to match the number of events which pass the $Z+$jet requirements. This scaling is determined by normalizing the simulations to the observed data.

4.2.1 Background normalization

The normalizations of the diboson, $t\bar{t}$, and multijet backgrounds are adjusted by scale factors, determined by a fit, in order to match the number of events in data. To obtain these normalizations, a wide window of the dilepton invariant mass distribution is used, $40 < M_{\ell\ell} < 200$ GeV, and the data is separated into 0, 1, and ≥ 2 jet multiplicity bins. For each jet multiplicity bin and lepton channel, we minimize the χ^2 by varying α (the QCD scale factor) and β (the MC scale factor):

$$\chi^2 = \sum_{i=0}^{\#\text{ of jets}} \frac{(\alpha S_i^{\text{QCD}} + \beta S_i^{\text{MC}} - \sigma_i^2)^2}{\sigma_i^2} \tag{4.1}$$

Table 4.1: Results of the multijet and diboson and $t\bar{t}$ background normalization in the muon and electron samples.

Bin	Muon		Electron (CC-CC)		Electron(CC-EC)	
Run IIa						
	α	β	α	β	α	β
0 jet	4.71 ± 0.4	0.939 ± 0.004	0.445 ± 0.01	1.06 ± 0.006	0.383 ± 0.007	1.02 ± 0.006
1 jet	0.311 ± 0.05	0.895 ± 0.008	0.326 ± 0.02	0.963 ± 0.01	0.264 ± 0.01	0.931 ± 0.01
≥2 jet	0.017 ± 0.006	0.885 ± 0.01	0.209 ± 0.05	0.994 ± 0.03	0.219 ± 0.02	1.009 ± 0.04
Run IIb1						
	α	β	α	β	α	β
0 jet	1.988 ± 0.3	0.929 ± 0.004	0.208 ± 0.005	0.984 ± 0.005	0.181 ± 0.003	0.949 ± 0.005
1 jet	0.222 ± 0.05	0.874 ± 0.009	0.163 ± 0.01	0.889 ± 0.01	0.144 ± 0.005	0.905 ± 0.02
≥2 jet	0.006 ± 0.005	0.915 ± 0.02	0.172 ± 0.03	0.930 ± 0.03	0.124 ± 0.01	1.003 ± 0.04
Run IIb2						
	α	β	α	β	α	β
0 jet	1.168 ± 0.1	0.975 ± 0.003	0.109 ± 0.002	1.010 ± 0.004	0.1034 ± 0.0009	0.975 ± 0.004
1 jet	0.157 ± 0.03	0.918 ± 0.006	0.124 ± 0.005	0.892 ± 0.008	0.103 ± 0.002	0.912 ± 0.01
≥2 jet	0.01 ± 0.003	1.013 ± 0.01	0.135 ± 0.01	0.939 ± 0.02	0.106 ± 0.005	0.975 ± 0.03
Run IIb34						
	α	β	α	β	α	β
0 jet	1.344 ± 0.1	0.937 ± 0.002	0.133 ± 0.001	1.003 ± 0.003	0.119 ± 0.009	0.996 ± 0.003
1 jet	0.178 ± 0.02	0.859 ± 0.005	0.122 ± 0.003	0.910 ± 0.007	0.106 ± 0.002	0.905 ± 0.009
≥2 jet	0.012 ± 0.003	0.969 ± 0.009	0.105 ± 0.008	0.989 ± 0.02	0.094 ± 0.004	0.973 ± 0.02

where S_i^{QCD} is the initial normalization of the multijet sample, S_i^{MC} is the initial normalization of the simulated physics processes, σ_i^2 is defined as the $M_{\ell\ell}$ distribution in data, and finally i refers to the bins of jet multiplicity which contain i number of jets. Table 4.1 shows the results of minimizing Equation 4.1 for each channel and jet multiplicity bin. The overall contribution of the multijet background is very small in the dimuon channel (see Section 3.1.2) which leads to a large relative uncertainty to the α scale factor.

4.3 Corrections to simulated events

The MC samples are corrected to account for various detector and physics effects which are not adequately modeled by the simulation. These corrections are described in this section.

4.3.1 Luminosity profile

In order to provide a realistic simulation of the detector response to beam conditions, data events are collected at random beam crossings and used to define the baseline detector response for simulated events. This is known as a *zero-bias overlay* [46]. The instantaneous luminosity for the zero-bias overlay does not match the luminosity profile of the data sample. Instead, we use independent measurements of the luminosity profiles to reweight the simulations. Figure 4.1 shows the simulations after the application of the zero-bias overlay and the corrections to the instantaneous luminosity in the simulations.

4.3.2 Primary vertex

When the two beams collide the luminous region broadens due to the emittance growth. In the simulations a gaussian distribution with a fixed width is used for the PV distribution. To provide a more realistic model, the PV z position is corrected in simulation to correspond to the measured distributions of PV z locations [61]. The PV z distribution in data and the corrected simulation is shown in Figure 4.1.

4.3.3 Lepton identification efficiency

When selecting electrons there is a difference in the identification efficiency in data and simulations. To correct for these differences a two step process is utilized. First, we correct the electrons which appear in the CC and EC as a function of φ and η, respectively. Then, a second parameterization is applied where the electrons are corrected as a function of instantaneous luminosity, φ, and η. The overall correction factor associated with these

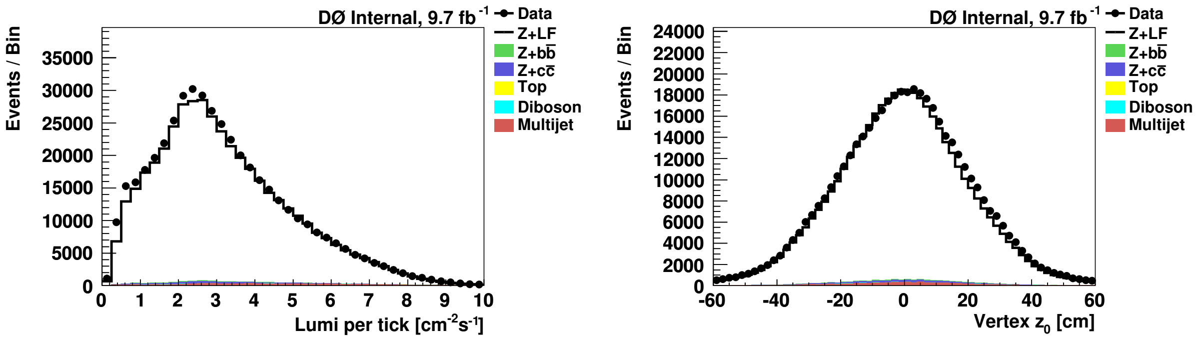


Figure 4.1: The instantaneous luminosity distributions of data and simulated samples for the $Z \rightarrow \mu\mu$ inclusive sample ($\geq$ 0 jet), shown on the left. The simulation has been corrected so that the generated samples mirrors the instantaneous luminosity of the data. The right plot shows the PV z distribution for data and the simulated samples, after corrections have been applied.

reweightings is 3% for CC electrons and 7% for EC electrons [44]. The efficiency for identifying muons is also different in data and simulation and must be corrected. There is a correction applied for the muon identification efficiency, and for the efficiency of reconstructing a matched track. The final correction for the muon identification efficiency is 0.5%, while the loose track correction is 3.6% [62].

4.3.4 Lepton pseudorapidity

The simulation of Z + jet events leads to a 15% underestimate of the number of leptons located at $|\eta| = 1.5$ [48]. This correction is applied as a function of η of the highest p_T lepton [48]. The η distribution of the highest and second highest p_T muons and electrons are shown in Figure 4.2 and Figure 4.3, respectively.

4.3.5 Lepton energy

The lepton energy resolution in the simulated events is better than what is measured in data. To corect this we apply a smearing to the lepton energy. To determine the electron resolution, a Crystal Ball function [63] is fitted to a sample of $Z \rightarrow ee$ data events. This resolution is then used to smear simulated events as a function of p_T, η, and φ. The typical

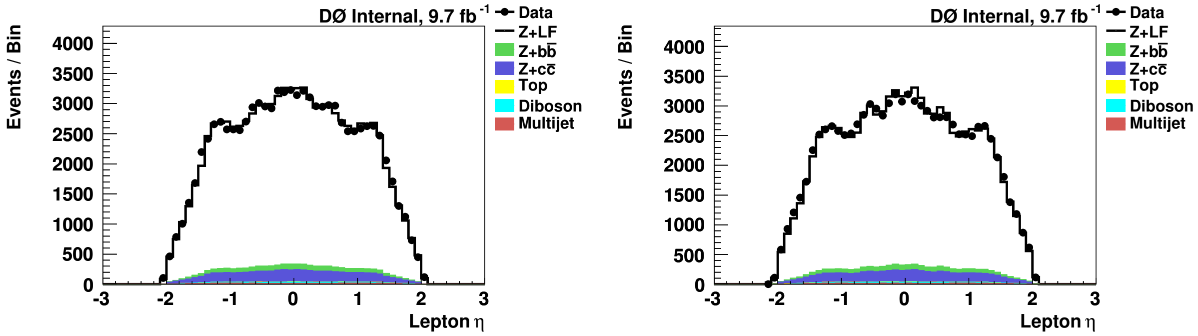


Figure 4.2: Highest p_T (left) and second highest p_T (right) lepton η distribution in the muon channel, after the lepton η correction has been applied for events with ≥ 1 jet.

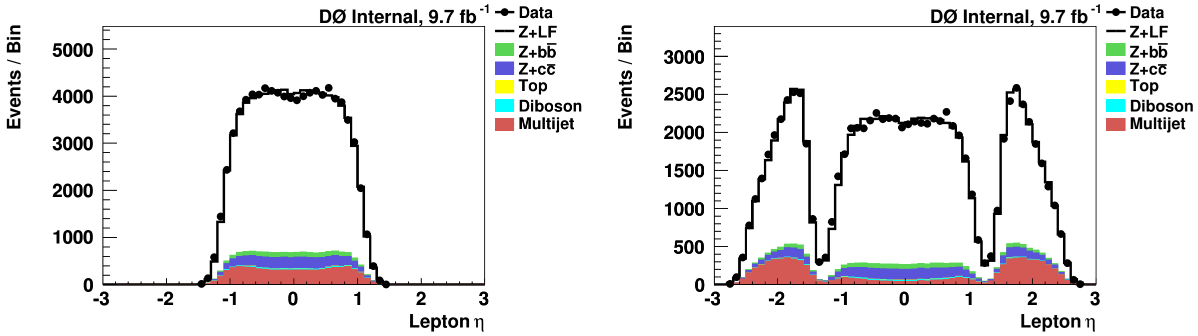


Figure 4.3: Highest p_T (left) and second highest p_T (right) lepton η distribution in the electron channel, after the lepton η correction has been applied for events with ≥ 1 jet.

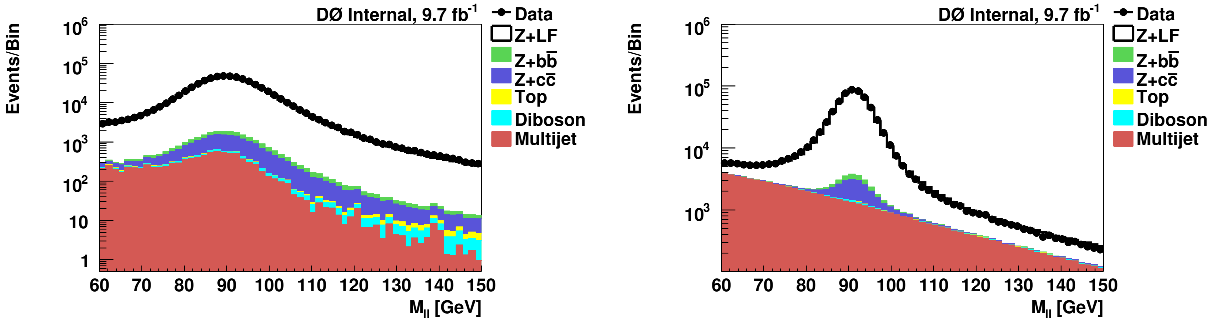


Figure 4.4: Reconstructed Z boson mass peak in muon (left) and electron (right) channels for the inclusive Z sample (≥ 0 jet), displayed with a logarithmic scale.

smearing for the simulated electron momentum is roughly 10% [44]. The muon momentum resolution is measured in samples of $Z \to \mu\mu$ and $J/\psi \to \mu\mu$, and the smearing is applied to randomly selected muons as a function of electric charge, p_T, η, and the length of the track. Applying a smearing of roughly 30% to the simulated muon momenta is necessary at $p_T \backsimeq 40$ GeV [43]. Figure 4.4 shows the reconstructed dilepton mass of the Z boson in both channels after the smearing has been applied.

4.3.6 Z boson transverse momentum

Past measurements at D0 have found that the Z boson p_T, p_T^Z, is not well modeled by either PYTHIA or ALPGEN [64], leading to an underestimate of ~30% at $p_T^Z > 10$ GeV. The discrepancy between data and simulation is corrected by reweighting the p_T^Z distribution. The correction is derived from the p_T distribution at the generator level and the observed spectrum in an unfolded sample of inclusive $Z \to ee$ data. The parametrized correction for the p_T^Z is used for both channels. This p_T^Z dependent correction is roughly 30% over the full p_T^Z range [64]. The correction is applied independent of jet multiplicity. The reweighting function, which is derived as a function of Z boson p_T, or q_T, is shown in Figure 4.5.

After this correction, the data and MC agree in the 0, 1, and ≥ 2 jet multiplicity bins. The p_T of the Z boson candidates are shown in Figure 4.12 and Figure 4.23, after all corrections.

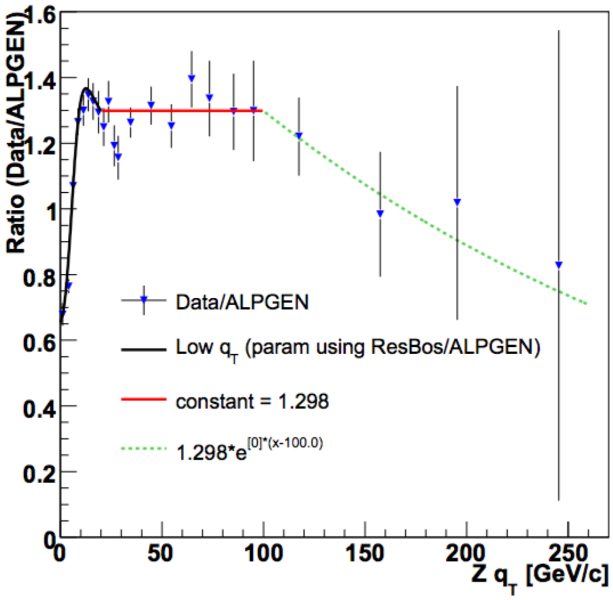


Figure 4.5: The ratio of unfolded $Z \to ee$ data and ALPGEN predictions which is used to reweight the simulated Z boson p_T, or Z q_T, spectrum. This reweighting is applied to events with Z boson $p_T < 100$ GeV [64].

4.3.7 Jet shifting, smearing, and removal

To account for differences in the jet energy resolution as measured in data and modeled by MC we apply a correction known as Jet Shifting, Smearing, and Removal (JSSR) [52]. This correction, as its name implies, uses what is observed in $Z\,(\to ee) + \text{jet}$ data to modify the simulated behavior of jets. It is required that the reconstructed Z boson and the jet be back-to-back, $\Delta\phi\,(Z, \text{jet}) > 2.5$. The correction is mapped as a function of ΔS, which can be defined two ways:

$$\Delta S = \frac{p_T^{\text{jet}} - p_T^{Z}}{p_T^{Z}} \tag{4.2}$$

or more generally by the function

$$f(\Delta S) = A \exp\left[-\frac{(\Delta S - \langle \Delta S \rangle)^2}{2\sigma_{\Delta S}^2}\right]\left[1 + Erf\left(\frac{\Delta S - T}{\sqrt{2}\sigma_T}\right)\right]. \tag{4.3}$$

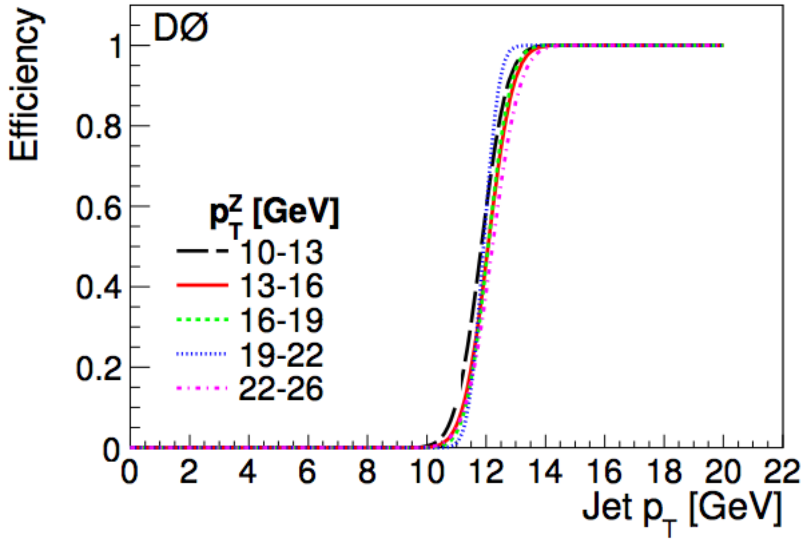


Figure 4.6: The turn on curves as a function of jet p_T. The saturation point represents T and the spread of these saturation points as a function of Z boson p_T corresponds to σ_T as in Equation 4.3 [52].

In this equation the free parameters are: $\langle \Delta S \rangle$ corresponding to the average shifting of the jet momentum relative to the Z boson, $\sigma_{\Delta S}$ represents the smearing which is applied to the jet p_T , T and σ_T are used to model the removal of jets at low p_T, and A is a normalization factor. These correction factors are derived in a flavor-dependent fashion such that we can correct jets originating from gluons, light partons, and b-quarks separately [52].

Figure 4.6 shows how T and σ_T are defined. The saturation location and width of the turn-on curves marks the location of the jet reconstruction threshold [52]. From this figure we see that $T = 12$ GeV and $\sigma_T = 1$ GeV. We take the threshold to be roughly $3\sigma_T$ above T and use a cutoff of 15 GeV. To apply the overall JSSR correction we start by smearing the jet p_T by a random number generated from a Gaussian with $\sigma^2_{corr} = \sigma^2_{DATA} - \sigma^2_{MC}$, found in Figure 4.7. Next we shift the overall jet p_T by $\langle \Delta S_{DATA} \rangle - \langle \Delta S_{MC} \rangle$, shown in Figure 4.7. If the jet falls below the 15 GeV cutoff it is removed.

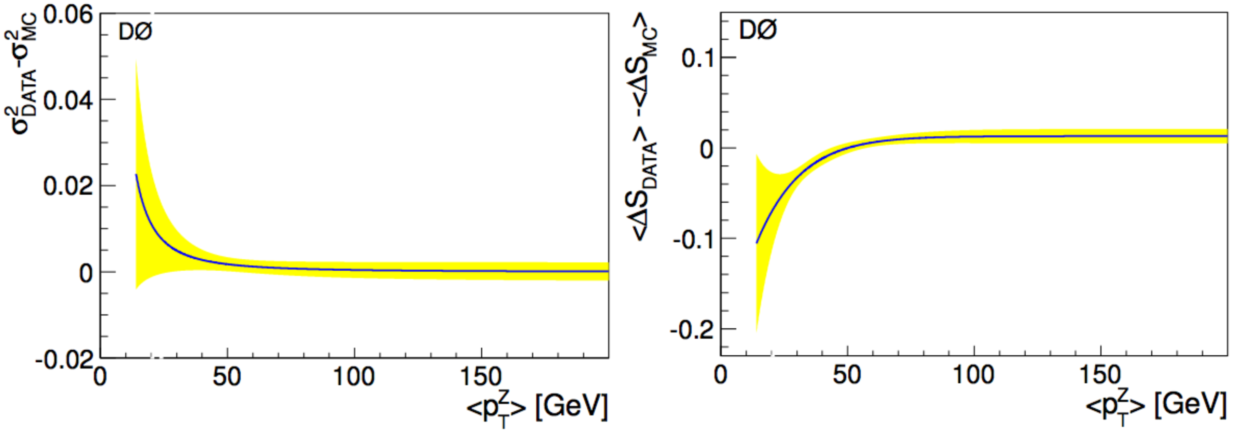


Figure 4.7: Residual resolution correction (left) and jet energy shifting (right) as a function of Z boson p_T for jets with $|\eta| < 0.8$ used to correct the overall jet energy [52]. The yellow band corresponds to the statistical uncertainty on the measurement.

4.3.8 Jet pseudorapidity

The ALPGEN simulation of the distribution of jets in the detector leads to an excess of jets being produced in the ICR region [48]. It has been found that SHERPA [65] correctly models the η distribution of jets [48]. The reconstructed angular distribution is reweighted to match the distribution in the SHERPA MC event generator to account for this issue.

4.4 Z boson plus jet modeling

After applying the above corrections we find that our simulations accurately describe the data. In Table 4.2 the number of events in data, various background components, and the expected number of Z+jet events can be found for both the dimuon and dielectron channels. The uncertainties on the yields found in the table are due to the simulated sample statistics, and are not correlated to the uncertainty on the simulated cross sections. Figure 4.8 - Figure 4.27 show the various kinematic distributions for the selected data events, along with the corrected MC simulations. These simulation are not only used to determine the size of the background contamination in the final samples and to estimate the detector acceptance.

Table 4.2: The number of selected data and simulated events. The uncertainties on the simulated samples are estimated from the MC event statistics.

	$Z(\to \mu\mu) + \text{jet}$ Sample	$Z(\to ee) + \text{jet}$ Sample
Data	94083	86039
$Z +$ light jet	84820 ± 82	70487 ± 59
$Z + b$ jet	2496.8 ± 4.1	2095.9 ± 3.4
$Z + c$ jet	5600.0 ± 8.1	4871.3 ± 6.6
Multijet	592 ± 10	7706 ± 33
ZZ	220.65 ± 0.85	189.52 ± 0.73
WZ	245.7 ± 1.3	222.6 ± 1.1
WW	38.60 ± 0.88	30.17 ± 0.74
$t\bar{t}$	116.87 ± 0.89	97.26 ± 0.55
Sum of Predictions	94130 ± 84	85700 ± 68

4.4.1 Dimuon channel

Figure 4.8 - Figure 4.17 show the kinematic distributions for the selected $Z(\to \mu\mu) + \text{jet}$ sample. Along with the data, the simulated backgrounds and $Z +$jet events are also shown. The simulated samples shown in the figures are the $Z +$ light jet, $Z + b$ jet, $Z + c$ jet, $t\bar{t}$, diboson, and multijet contributions. These events were selected with the criteria laid out in Chapter 3 and the simulations detailed in Chapter 4.

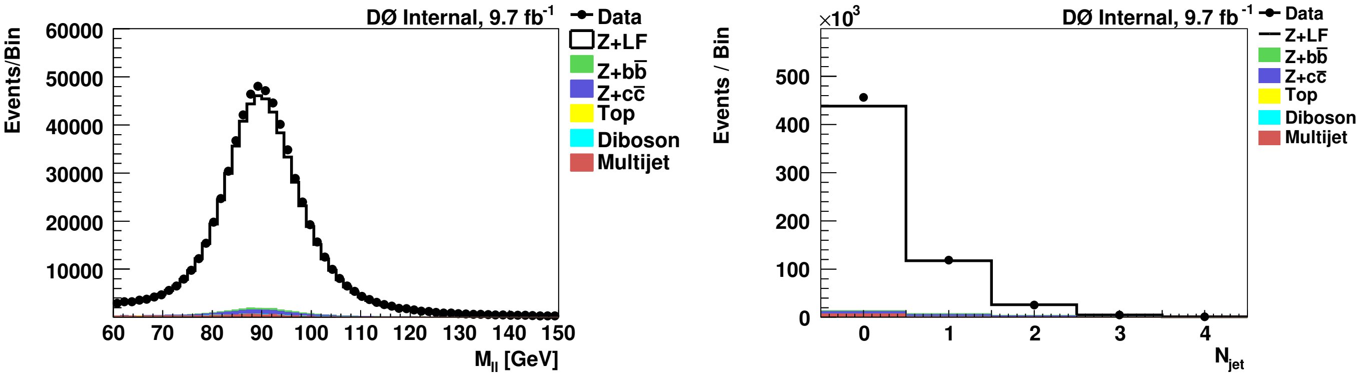


Figure 4.8: The dilepton invariant mass (left) and jet multiplicity (right) in the inclusive Z sample (≥ 0 jet), where the Z boson decays to muons and after the selection described in Chapter 3.

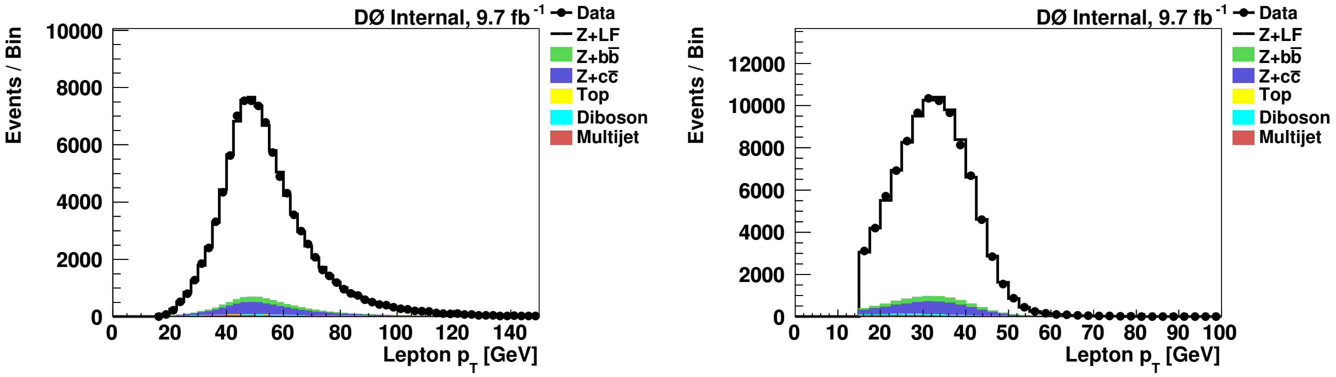


Figure 4.9: The highest (left) and second highest (right) p_T lepton p_T spectrum in $Z+\geq 1$ jet sample, where the Z boson decays to muons and after the selection described in Chapter 3.

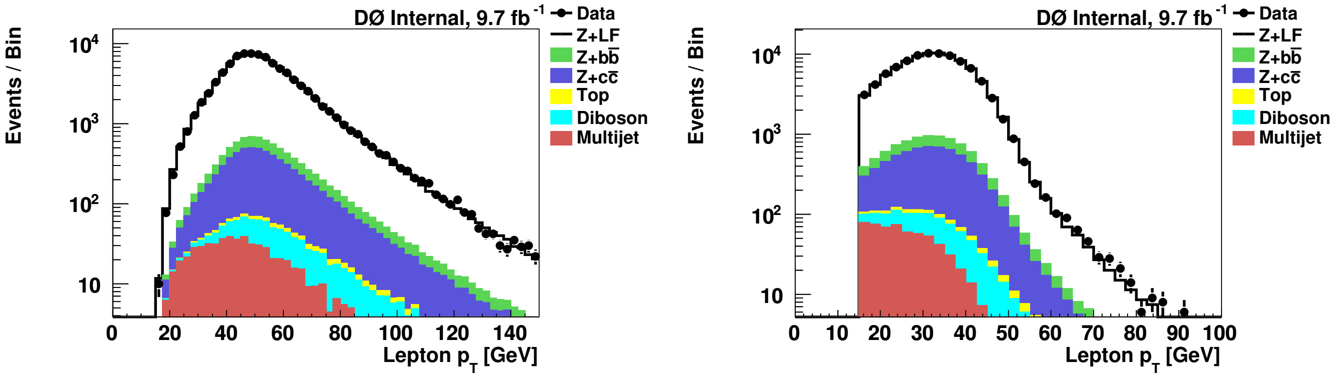


Figure 4.10: The highest (left) and second highest (right) p_T lepton p_T spectrum in $Z+\geq 1$ jet sample, in logarithmic scale, where the Z boson decays to muons and after the selection described in Chapter 3.

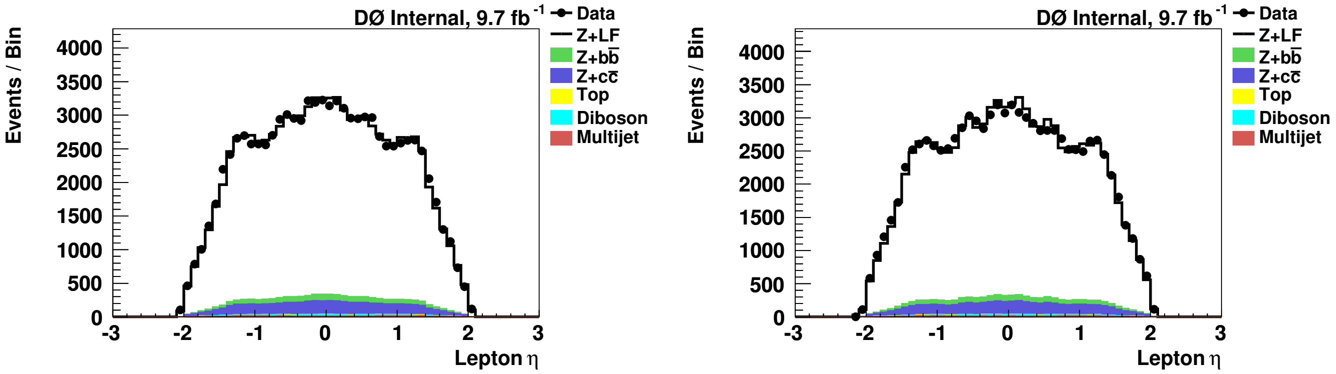


Figure 4.11: The highest (left) and second highest (right) p_T lepton η spectrum in $Z+\geq 1$ jet sample, where the Z boson decays to muons and after the selection described in Chapter 3.

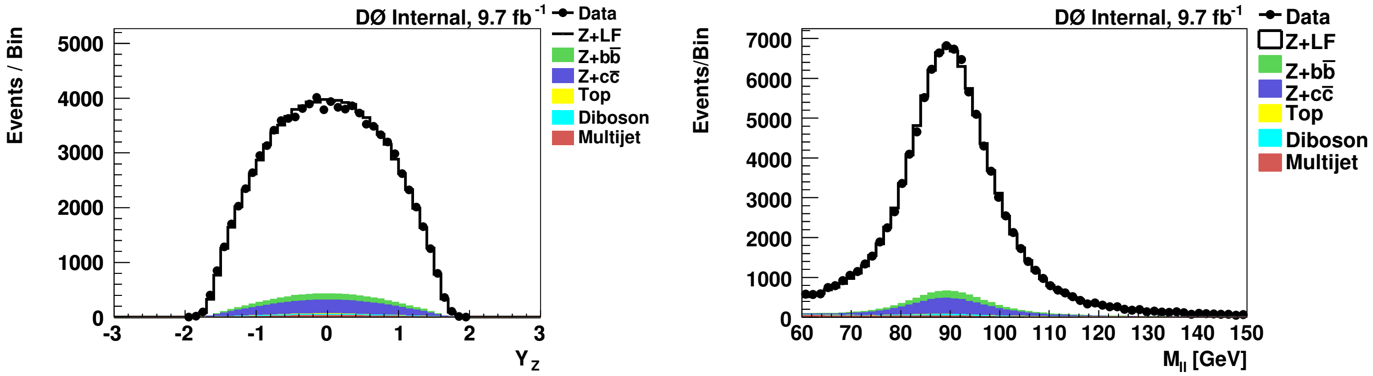


Figure 4.12: The rapidity (left) and mass of the Z boson candidate (right) in the $Z+\geq 1$ jet sample, where the Z boson decays to muons and after the selection described in Chapter 3.

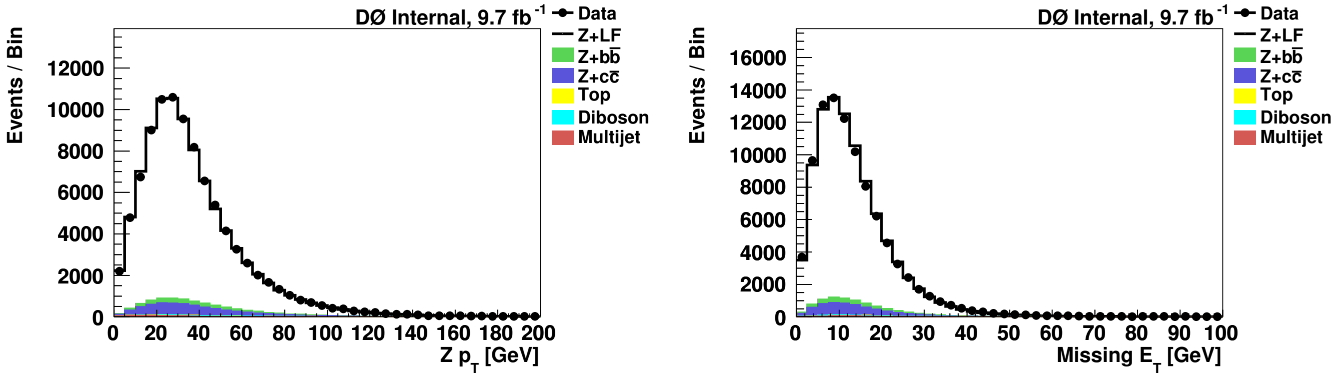


Figure 4.13: The p_T of the Z boson candidate (left) and $\not\!\!E_T$ (right) in the $Z+\geq 1$ jet sample, where the Z boson decays to muons and after the selection described in Chapter 3.

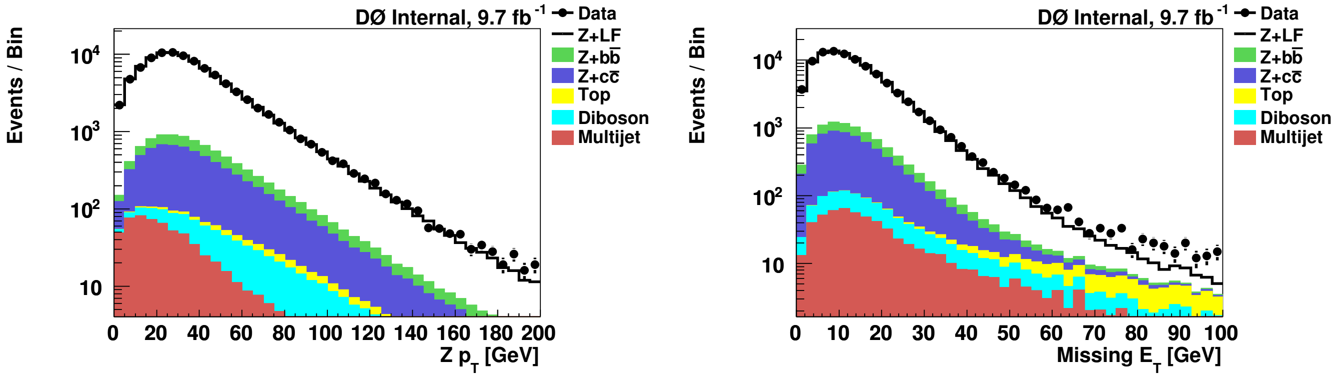


Figure 4.14: The p_T of the Z boson candidate (left) and $\not\!\!E_T$ (right) in the $Z+\geq 1$ jet sample, in logarithmic scale, where the Z boson decays to muons and after the selection described in Chapter 3. The mismodeling observed in the region $\not\!\!E_T > 70$ GeV is due to the mismodeling of the muon momentum resolution.

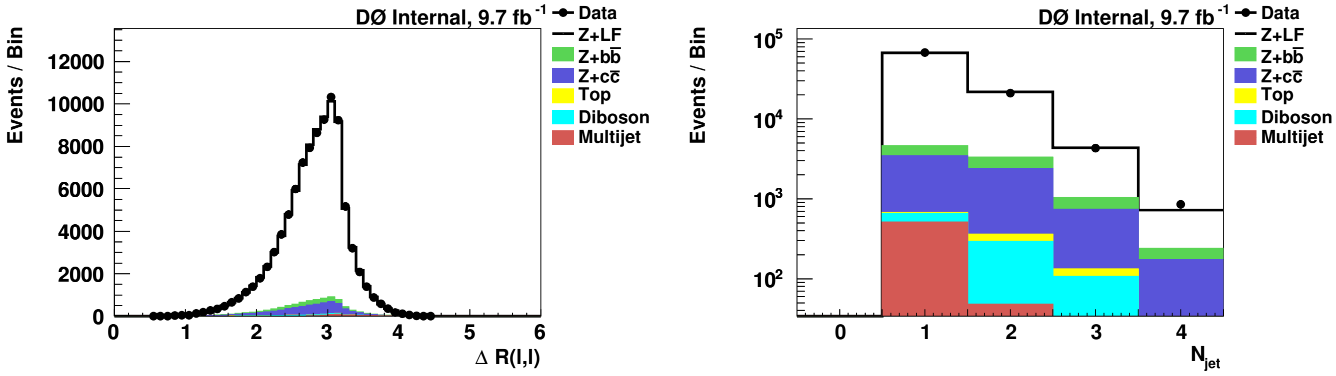


Figure 4.15: The $\Delta R(\mu\mu)$ for the Z boson candidate (left) and jet multiplicity (right) in the $Z+\geq 1$ jet sample, where the Z boson decays to muons and after the selection described in Chapter 3.

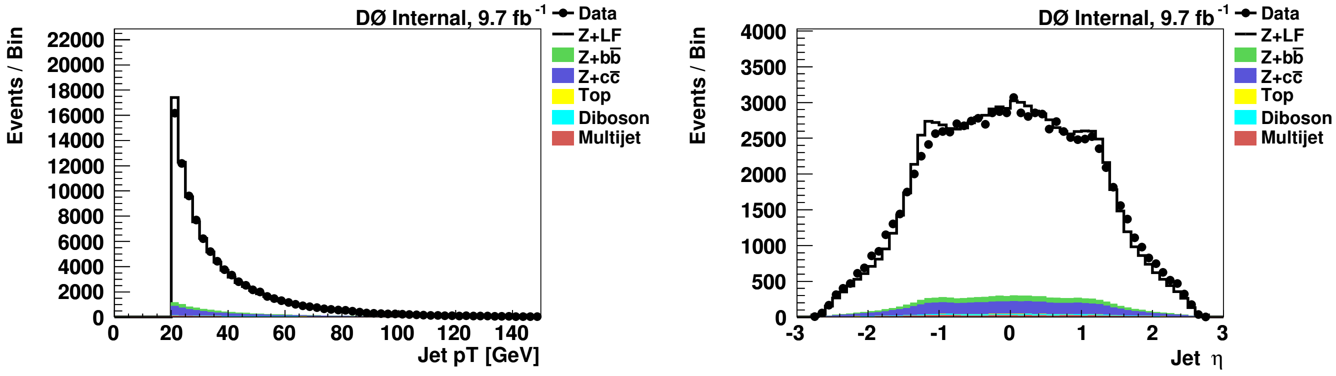


Figure 4.16: The p_T (left) and η (right) distribution of the highest p_T jet for data and background in the $Z+\geq 1$ jet sample, where the Z boson decays to muons and after the selection described in Chapter 3.

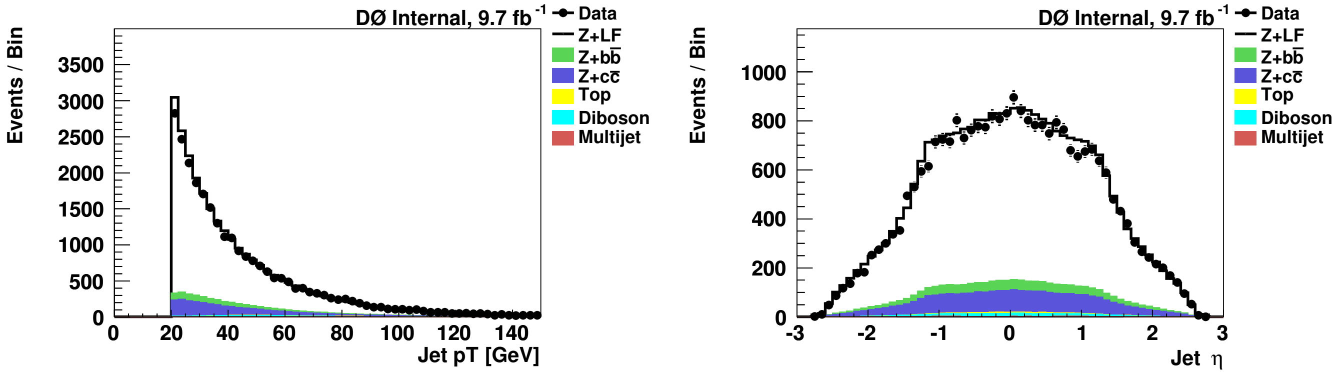


Figure 4.17: The p_T (left) and η (right) distribution of the second highest p_T jet for data and background in the $Z+\geq 1$ jet sample, where the Z boson decays to muons and after the selection described in Chapter 3.

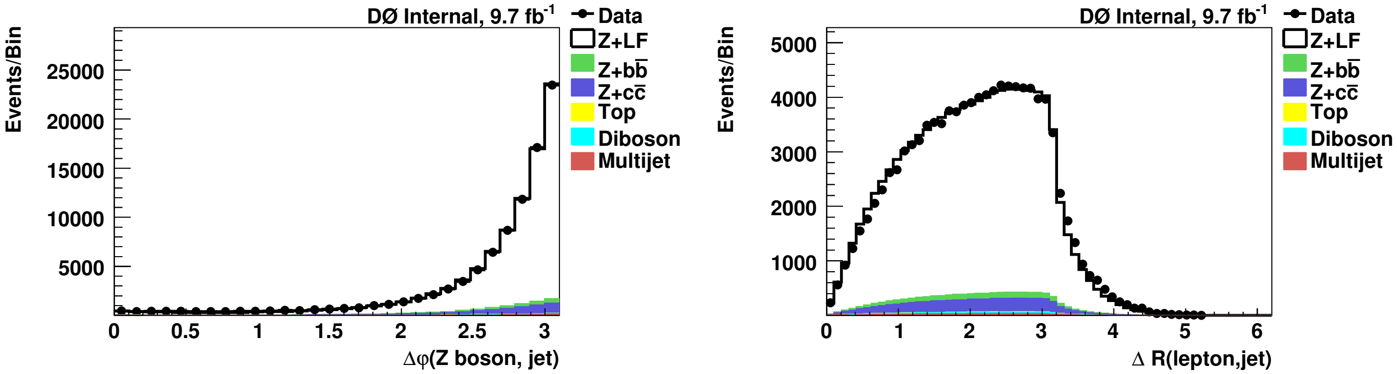


Figure 4.18: The $\Delta\varphi$ between the Z boson and the highest p_T jet (left) and the ΔR between the highest p_T lepton and jet (right) distribution for data and background in the $Z+\geq 1$ jet sample, where the Z boson decays to muons and after the selection described in Chapter 3.

In Figure 4.14 we see some mismodeling of the data by the simulations for the $\not\!\!E_T$ distribution in the region of $\not\!\!E_T > 70$ GeV. This is due to the mismodeling of the muon momentum resolution and its effect on the measurement of the $\not\!\!E_T$. To reduce the large contribution from top quark pair production in this region a requirement of $\not\!\!E_T < 60$ GeV is applied.

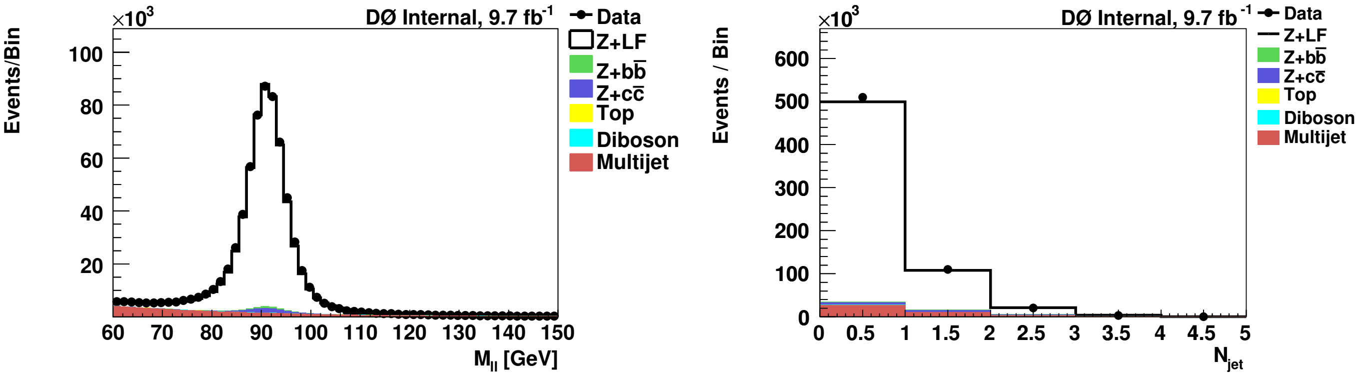


Figure 4.19: The dilepton invariant mass (left) and jet multiplicity (right) in inclusive Z sample (≥ 0 jet), where the Z boson decays to electrons and after the selection described in Chapter 3.

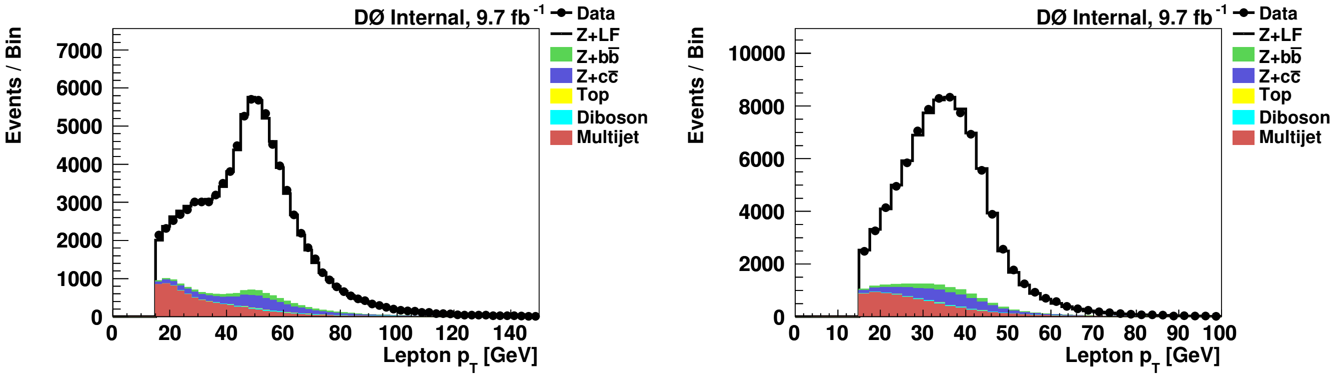


Figure 4.20: The highest (left) and second highest (right) p_T lepton p_T spectrum in $Z+\geq 1$ jet sample, where the Z boson decays to electrons and after the selection described in Chapter 3.

4.4.2 Dielectron channel

Figure 4.19 - Figure 4.29 shows the kinematic distributions for the selected $Z(\rightarrow ee) +$ jet sample. Along with the data events the simulated background and signal events are also shown. These events were selected with the criteria laid out in Chapter 3 and the simulations are detailed in Chapter 4.

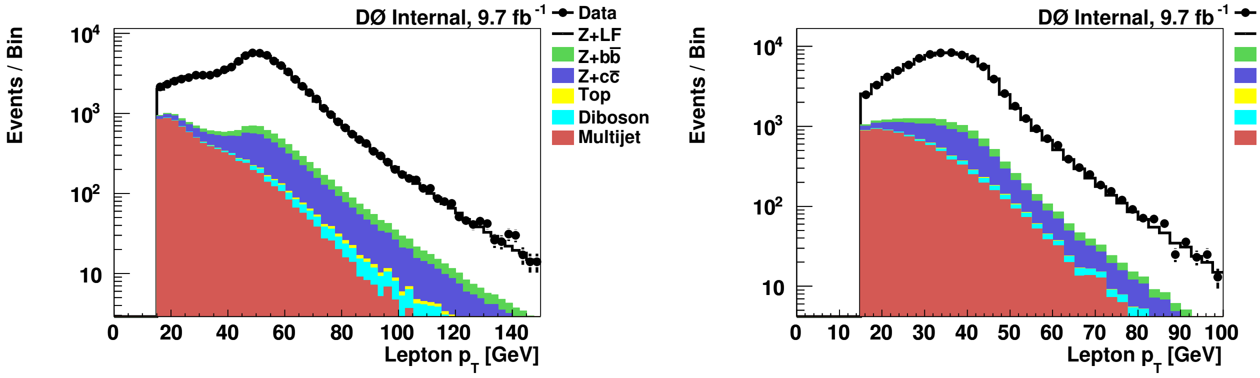


Figure 4.21: The highest (left) and second highest (right) p_T lepton p_T spectrum in $Z+\geq 1$ jet sample, in logarithmic scale, where the Z boson decays to electrons and after the selection described in Chapter 3.

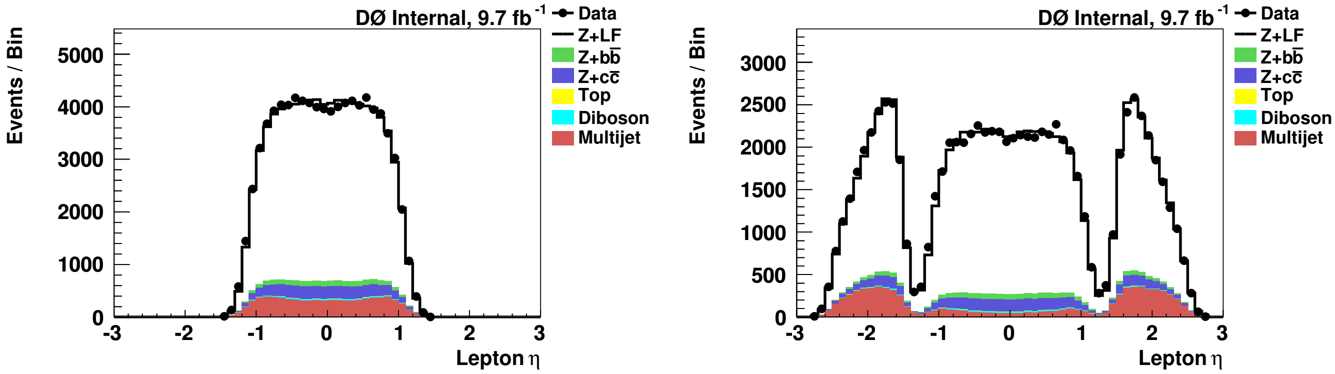


Figure 4.22: The highest (left) and second highest (right) p_T lepton η spectrum in $Z+\geq 1$ jet sample, where the Z boson decays to electrons and after the selection described in Chapter 3.

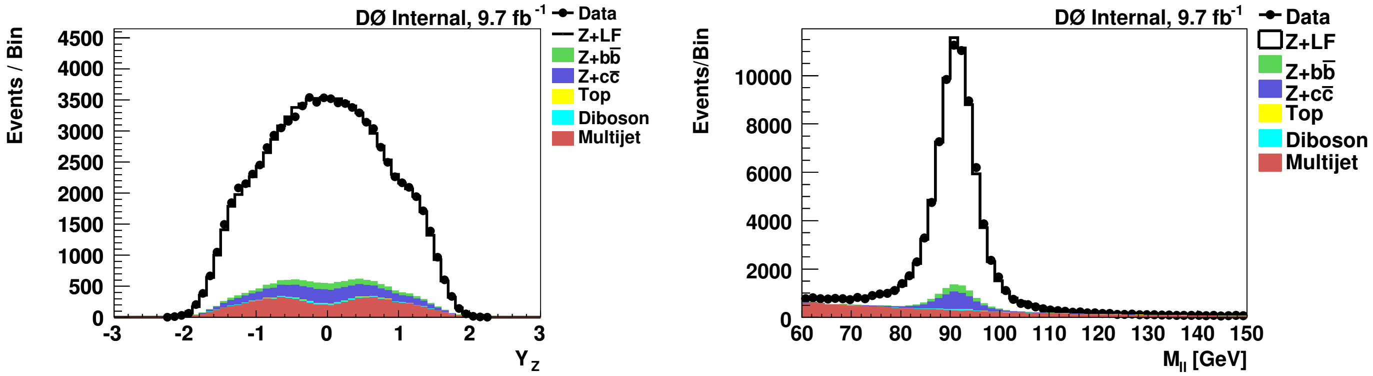


Figure 4.23: The rapidity (left) and mass of the Z boson candidate (right) in the $Z+\geq 1$ jet sample, where the Z boson decays to electrons and after the selection described in Chapter 3.

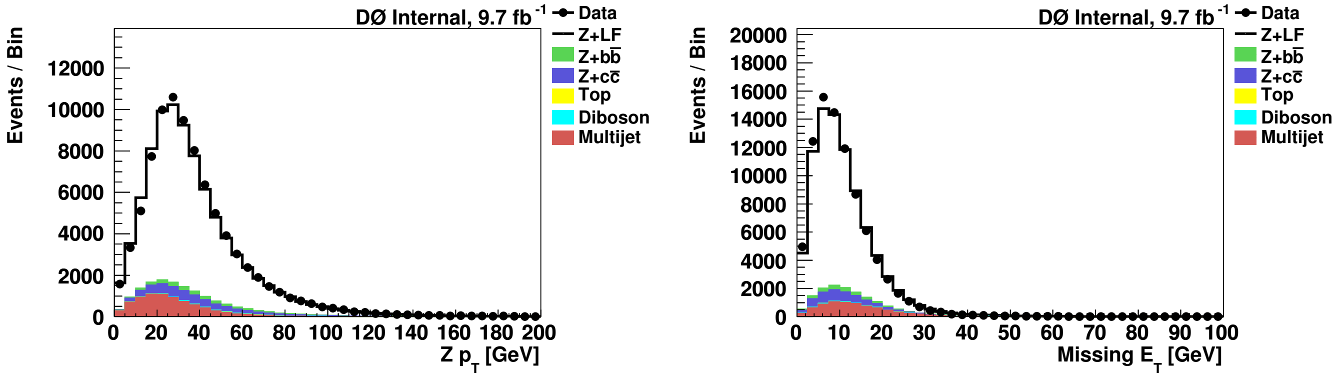


Figure 4.24: The p_T of the Z boson candidate (left) and $\not\!\!E_T$ (right) in the $Z+\geq 1$ jet sample, where the Z boson decays to electrons and after the selection described in Chapter 3.

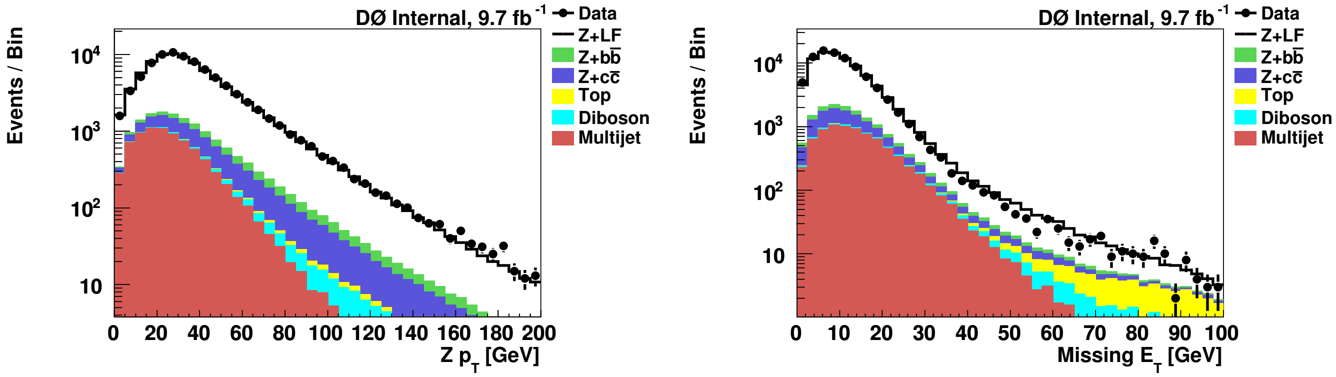


Figure 4.25: The p_T of the Z boson candidate (left) and $\not\!\!E_T$ (right) in the $Z+\geq 1$ jet sample, in logarithmic scale, where the Z boson decays to electrons and after the selection described in Chapter 3.

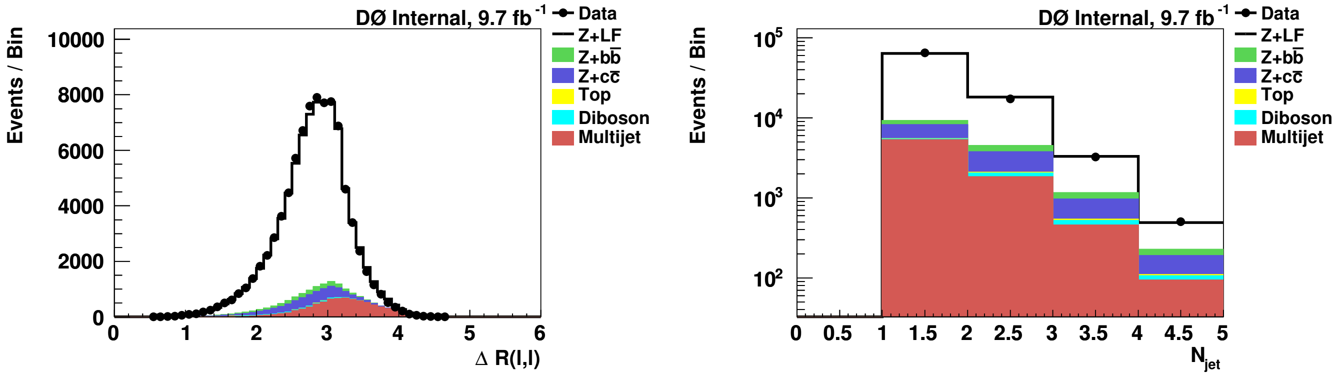


Figure 4.26: The $\Delta R(ee)$ for the Z boson candidate (left) and jet multiplicity (right) in the $Z+\geq 1$ jet sample, where the Z boson decays to electrons and after the selection described in Chapter 3.

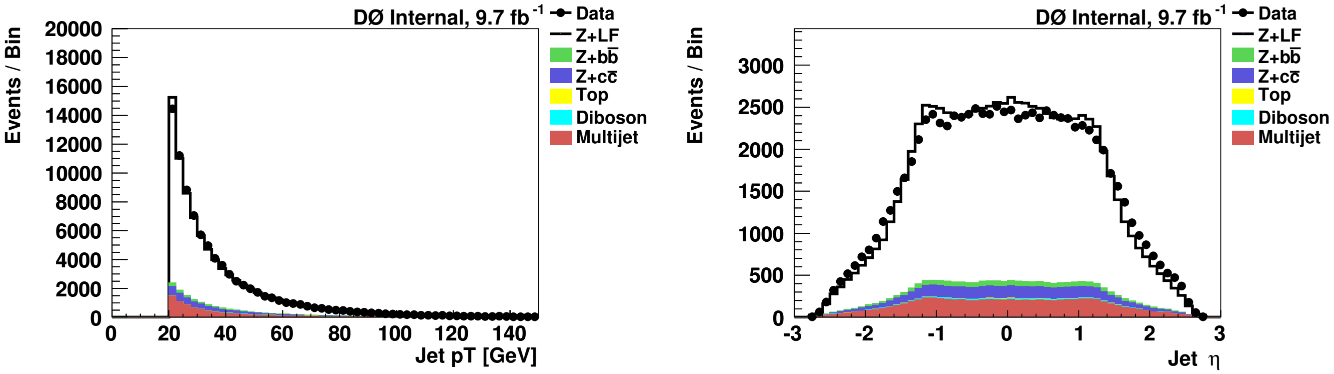


Figure 4.27: The p_T (left) and η (right) distribution of the highest p_T jet for data and background in the $Z+\geq 1$ jet sample, where the Z boson decays to electrons and after the selection described in Chapter 3.

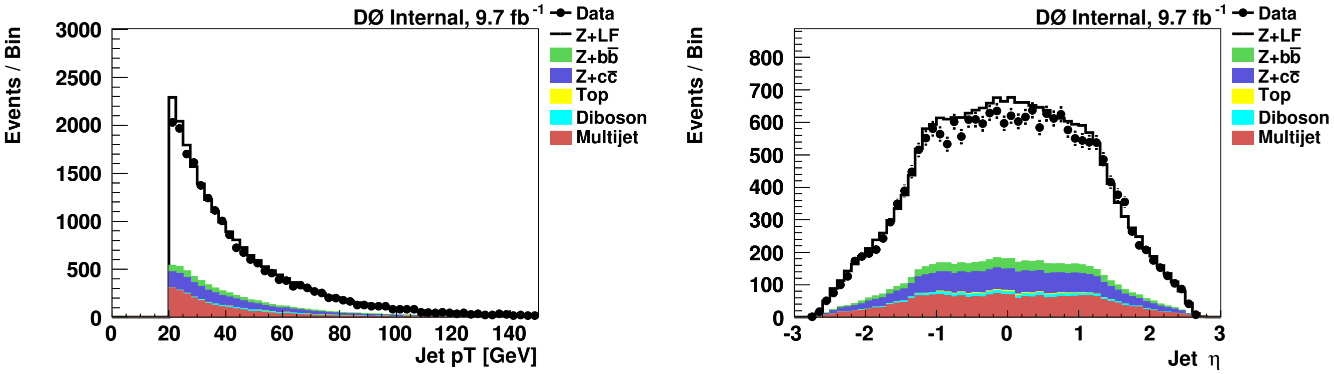


Figure 4.28: The p_T (left) and η (right) distribution of the second highest p_T jet for data and background in the $Z+\geq 1$ jet sample, where the Z boson decays to electrons and after the selection described in Chapter 3.

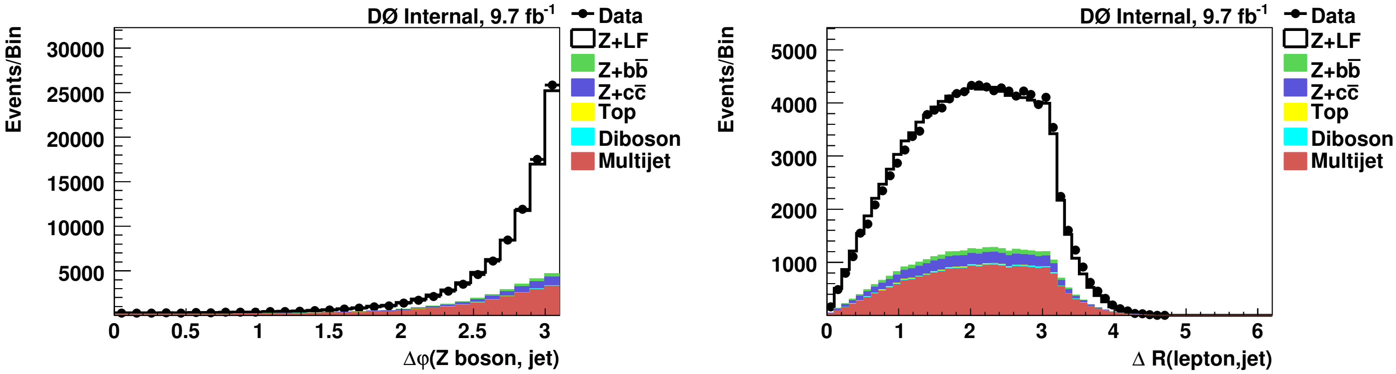


Figure 4.29: The $\Delta\varphi$ between the Z boson and the highest p_T jet (left) and the ΔR between the highest p_T lepton and jet (right) distribution for data and background in the $Z+\geq 1$ jet sample, where the Z boson decays to electrons and after the selection described in Chapter 3.

5 Heavy flavor jet identification

We seek to determine the HF content of the sample of $Z+\text{jet}$ events. To do this we will use an algorithm that utilizes the unique properties of tracks and secondary vertices of jets that originate from HF quarks. Applying a lower threshold on the output of this algorithm will suppress the contributions from light jets while enriching our sample in HF jets. An example of the decay of a HF hadron inside a HF jet can be seen in Figure 5.1, where a secondary vertex provides us information to differentiate it from its light counterparts.

We must also measure the efficiency and misidentification rate (or the rate at which light jets are selected by the algorithm) in data. These measurements are used to correct the simulations which do not accurately model the responses of these algorithms. Differential cross section measurements rely on an accurate modeling of these efficiencies. Additionally, understanding the light jet background which survives this algorithm is necessary for measuring $Z+c$ jet cross sections, described further in Chapter 8.

5.1 Algorithm preselection

Before a candidate jet can be used in the algorithm it must pass certain criteria. First, we must be sure that the jet contains enough information such that we can use it in our algorithm, to do this we match the jet to a set of tracks, discussed in Section 5.1.1. Secondly, we want to be sure that these tracks are not contaminated by signatures which can mimic the HF jets we are hoping to isolate. This contamination can come from displaced vertices

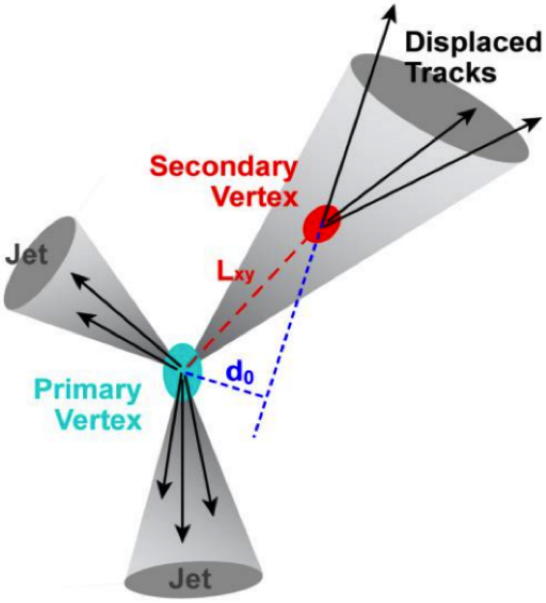


Figure 5.1: An example of a jet with tracks from a secondary vertex recoiling off two jets originating from the PV. A secondary vertex is a signature of a HF jet. The displacement of the vertex, L_{xy}, and the IP, denoted d_0 in this figure, allow us to discriminate this jet from their light counterparts.

originating not from HF hadrons but from other processes. Tracks associated with these contaminants are isolated and removed before the jets are used in the algorithms, shown in Section 5.1.2 and Section 5.1.3.

5.1.1 Taggability

The b jet identification algorithms are based on tracking and vertex information. Therefore we require that each jet reconstructed in the calorimeter can be associated with tracks in the tracking system. We implement this 'taggability' [58] requirement separately from the actual HF identification algorithm, allowing for our measurements of the performance of the algorithm to be independent of possible variations of the tracking system's efficiency with time. All identification efficiencies and light jet misidentification rates are measured relative to *taggable* jets. The taggability requirement is parametrized in terms of jet p_T, η, and z. The z parameterization is necessary due to the extended length of the luminous region

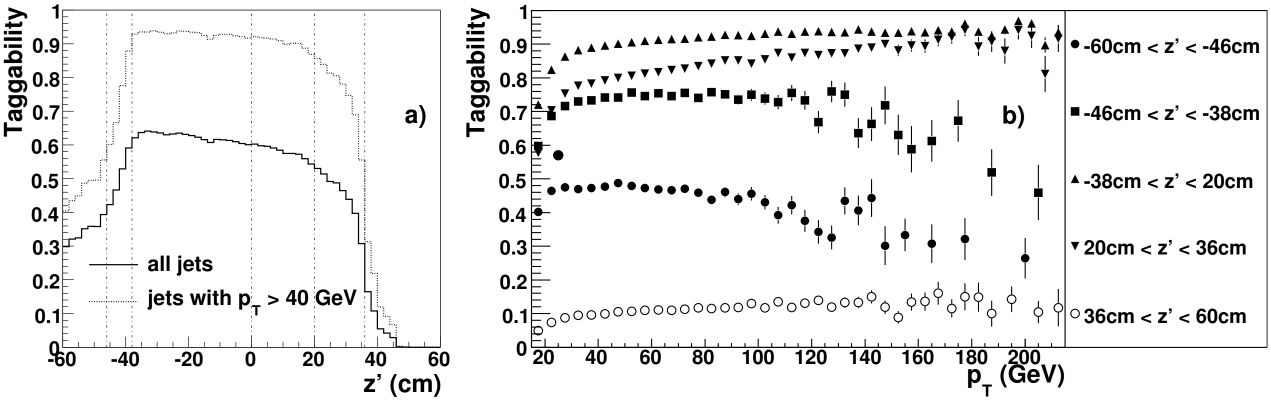


Figure 5.2: (a) The efficiency of the taggability requirement as a function of $z' \equiv |z|\text{sign}(\eta \times z)$. The vertical lines designate the regions of z' that the taggability requirement is parameterized. (b) The taggability efficiency as a function of jet p_T in the various bins of z' [58].

in the D0 detector. For the HF identification studies in this chapter we select a range of $|z| < 25$ cm. For a jet reconstructed in the calorimeter to be considered *taggable* it must be matched to at least two tracks within a cone of $\Delta R < 0.5$. The efficiency for selecting a taggable jet as a function of $|z|$ and jet p_T is shown in Figure 5.2.

5.1.2 Neutral strange hadron rejection

Neutral hadrons containing strange quarks, V^0, have long decay lifetimes and therefore can mimic signatures of b-quarks. In particular, K_S and Λ hadrons have lifetimes of 90 and 263 ps, respectively. To suppress this background we reject secondary vertices with two oppositely charged tracks that satisfy the following criteria:

- For each track, the significance of the distance of closest approach to the selected PV in the transverse plane, $\mathcal{S}_d = dca/\sigma_{dca}$, must be $|\mathcal{S}_d| > 3$.
- For each track, the z coordinate of the point of closest approach in the transverse plane is displaced from the PV by less than 1 cm. This requirement suppresses misreconstructed tracks.
- The V^0 candidate must have a dca to the PV of less than 200 μm. This guarantees

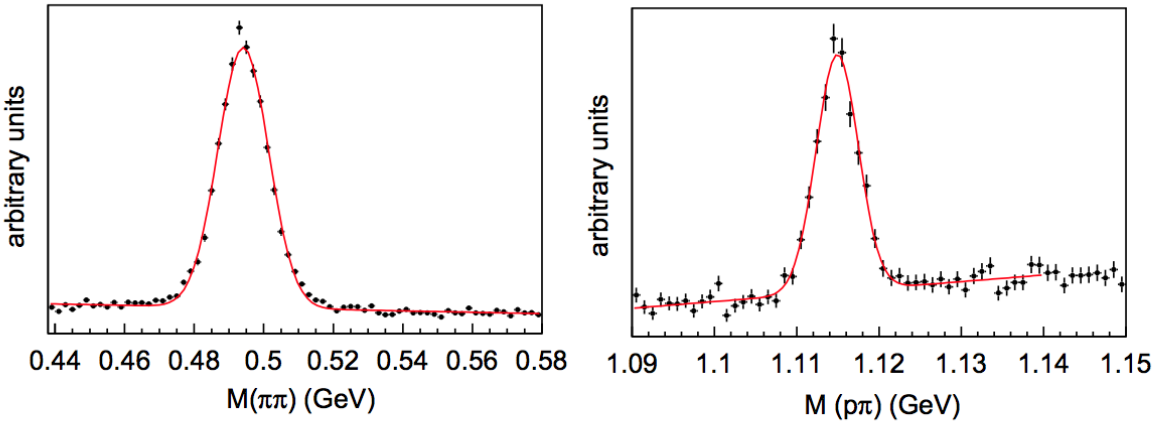


Figure 5.3: V^0 mass peaks, K_s (left) and Λ (right), reconstructed from pairs of tracks identified by the 'V^0 remover' [58].

that only V^0s from the PV are rejected, not those which may have originated from B hadron decays.

- A selected V^0 candidate is rejected if its mass is compatible with the mass of K_S or Λ, $472 < m_{K_S} < 516$ MeV and $1108 < m_\Lambda < 1122$ MeV.

5.1.3 Photon conversion rejection

To identify photon conversion in the plane transverse to the beam line, an opening angle of nearly zero between the electron and positron is utilized. The tracks from the electron and positron are required to be less than 30 μm apart at the point where their trajectories are parallel to each other. In addition, they must have an opposite electric charge, and their invariant mass must be less than 25 MeV.

5.2 b jet identification algorithms

There are currently three low-level tools at D0 to identify whether a jet is produced from a b-quark. The inputs to these tools can be combined using multivariate techniques to further improve their individual discriminating power.

Counting Signed Impact Parameter (CSIP) - CSIP [58] determines the number of tracks identified in a jet based on their S_d with respect to the PV. Events must have at least three tracks with an $S_d > 2$ or two tracks with an $S_d > 3$ to be selected.

Jet Lifetime Impact Parameter (JLIP) - The JLIP algorithm provides, essentially, the confidence level that all the tracks in a jet originate from the PV [58]. First, the probability of the track originating from the PV is computed. These probabilities are then combined for all the tracks matched to a jet to determine the probability that it originated from the PV. Light jets will have a flat JLIP distribution ranging from 0 to 1. While, HF jets will peak at low values. By setting an upper bound on this distribution we can create a sample enriched in HF jets.

Secondary Vertex Tagger (SVT) - SVT [58] uses tracks to reconstruct secondary vertices, those which are significantly displaced from the PV. A jet is selected if a secondary vertex is located within $\Delta R < 0.5$ of the jet. By varying the restrictions on the input variables we can create a set of five SVT algorithms (SVT1-5) that can be used to provide a complimentary set of secondary vertices in the jets. The track selections for the different configurations are listed in Table 5.1. These are tuned by varying the requirements on the track p_T, χ^2 per degree of freedom for the secondary vertex, the distance of closest approach of the secondary vertex to the primary vertex in the plane transverse to the beam direction (dca_{xy}), and S_d.

Table 5.1: Track selection requirements for the five SVT algorithm configurations: SVT1-5.

Track selections	SVT1	SVT2	SVT3	SVT4	SVT5
p_T [GeV]	0.5	0.5	0.5	1.	1.
χ^2	15.	15.	10.	10.	3.
dca_{xy} [cm]	–	1.5	3.	3.	3.5
S_d	–	–	5.	5.	7.

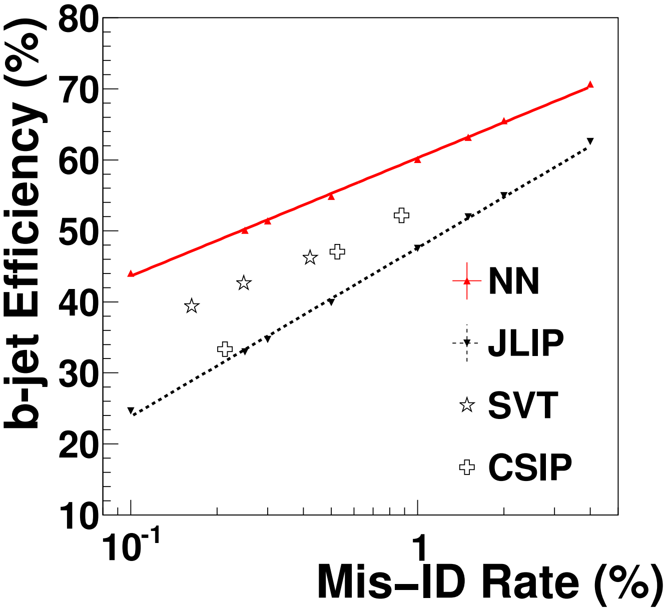


Figure 5.4: The MC performance profile (defined in Section 5.3.3) comparison of the previous identification algorithms at D0. We can see improvement is achieved by combining the low level algorithms using a neural network, the *D0 NN algorithm* [58].

Selected input variables from these tools have been combined using a neural network to form the D0 *NN algorithm* [58]. It has shown significant performance improvements when compared to the individual tools, shown in Figure 5.4. In the following we will discuss how further improvements have been achieved using an extended set of input variables and making use of decision trees and a neural network. The new algorithm which results from these improvements is known as the MVA$_{bl}$.

5.3 MVA$_{bl}$ algorithm

To build our MVA$_{bl}$ algorithm we start by creating a two samples: 10^6 signal events (MC sample of di-b jet events) and 10^6 background (MC sample of QCD light jet events). We then take a large number of variables (discussed in Section 5.3.1) and train six random forests (RF) [66] using the ROOT TMVA [67] framework. One RF is trained using the IP-properties and one for each set of SVT variables extracted from the five different SVT configurations.

These six random forests are then combined using a neural network implementation, the TMultilayerPercepton (MLP), from the ROOT [68] framework. This neural network utilizes the non-linear correlations between these inputs to produce the MVA_{bl} output. This process leads to improved discrimination over the D0 NN algorithm by using an order of magnitude more variables when training. The six intermediate RFs are used to combine a total of 155 variables, where each is used to combine a subset of these. Using the complimentary information supplied by these RFs the final trained algorithm can highly suppress the background events.

5.3.1 Input variables

Each of the six RFs are trained by a set of variables which differentiate b and light jets. A large number of input variables are used for each RF. The most discriminating are discussed below.

Impact parameter variables

When training the track IP RF we combine the outputs of the JLIP and CSIP algorithms with eight additional variables. These variables are: (i) the output of the JLIP algorithm; (ii) the output of the CSIP algorithm; (iii) the reduced JLIP [58], which is computed by removing the track with the lowest probability of originating from the PV and then recalculating the JLIP; (iv) the combined probability associated with the tracks with the highest and second highest probability of coming from the PV; (v) the largest distance in ΔR between any two tracks within a jet, or the total ΔR width of the tracks in the jet, $\max[\Delta R(\text{tracks})]$; (vi) the sum of the ΔR distances between each track matched to the jet and the center of the calorimeter jet, $\Sigma_{trk}\Delta R(trk, jet)$; (vii) the p_T-weighted ΔR width of the tracks relative to the calorimeter jet, which is defined as

$$\Omega = \frac{\Sigma_{trk}\, p_T^{trk} \times \Delta R(trk, jet)}{\Sigma_{trk} p_T^{trk}}; \tag{5.1}$$

(viii) the total transverse momentum of all the selected tracks; and (ix) the total number of tracks matched to the jet. The resulting RF output distribution is displayed in Figure 5.5(a).

Secondary vertex variables

The SVT algorithm preselects tracks according to their kinematic properties and reconstruction qualities. As a consequence, starting from a common set of tracks, the various SVT configurations lead to different secondary vertices with different properties, providing a complimentary set of variables to extract for each jet. The subtle differences in reconstruction properties can vary with jet momentum and pseudorapidity, allowing for improved discrimination between b, c, and light jets.

Along with the variables associated with the displacement of the secondary vertex, such as the decay length and the pseudo-lifetime ($c\tau$), additional variables sensitive to the fragmentation process and jet-shape variables are used for discrimination. Jet shape variables mainly describe the track structure of the jet, such as its average and maximal size in the $\eta - \varphi$ plane. The fragmentation of B hadrons tends to be at higher p_T than its lighter counterparts. This can be probed by studying the charged and neutral activity in the plane transverse to the direction of the tracks originating from the secondary vertex.

In total, each of the SVT RFs use twenty-nine input variables. These include: (i) the p_T fraction carried by the secondary vertex, $p_T^{\mathrm{SVT}}/p_T^{jet}$; (ii) the track multiplicity originating from the secondary vertex; (iii) the signed decay length significance in the plane transverse to the beam direction; (iv) the JLIP probability of the tracks matched to the secondary vertex; (v) the χ^2/n.d.f. of the tracks matched to the secondary vertex; (vi) the number of secondary vertices which can be reconstructed from the tracks matched to the jet; (vii) the signed IP of the track with the highest momentum measured transverse to the direction of flight of the secondary vertex; (viii) the number of tracks matched to the jets; (ix) $c\tau$ computed in the plane transverse to the beam; (x) decay length measured in the plane transverse to the beam direction; (xi) decay length in the beam direction; (xii) the p_T of the highest p_T track in the jet normalized to the p_T^{SVT}, p_T^1/p_T^{SVT}; (xiii) the p_T of the second highest p_T track normalized to the secondary vertex p_T, p_T^2/p_T^{SVT}; (xiv) the *dca* of the secondary vertex to the PV in the plane transverse to the beam; (xv) the *dca* of the secondary vertex to the PV in the beam direction; (xvi) the mass of the secondary vertex ($\mathrm{M_{SV}}$), calculated by summing the invariant masses of all track four-momentum vectors assuming that all tracks originate from pions; (xvii) the p_T of the track which has the highest momentum measured relative to the direction of flight of the secondary vertex; (xviii) the momentum of the secondary vertex in the plane transverse to the calorimeter jet direction; (xix) the p_T of the highest p_T track normalized to the total jet p_T, p_T^1/p_T^{jet}; (xx) the p_T of the second highest p_T track normalized to the total jet p_T, p_T^2/p_T^{jet}; (xxi) the opening angle of the secondary vertex projected into the plane transverse to the beam direction; (xxii) the opening angle of the secondary vertex projected in the beam direction; (xxiii) the Ω as measured for tracks matched to the secondary vertex; (xxiv) $\max[\Delta R(\text{tracks})]$ of the tracks matched to the secondary vertex; (xxv) p_T of the highest p_T track matched to the secondary vertex; (xxvi) p_T of the second highest p_T track matched to the secondary vertex; (xxvii) the charge (q) of the jet, measured as $\Sigma_{trk} p_T^{trk} q^{trk}/p_T^{jet}$; (xxviii) the signed decay length significance in the beam direction; and (xxix) the ΔR size of the tracks matched to the secondary vertex. The five SVT RFs which are trained using these variables can be seen in Figure 5.5(b-f).

5.3.2 Optimized algorithm parameters

The outputs of the six RFs, shown in Figure 5.5, are combined with an MLP neural network into a single variable. The training parameters for the six separate RFs and the final MLP are optimized to minimize the misidentification rate for a fixed b jet identification efficiency using

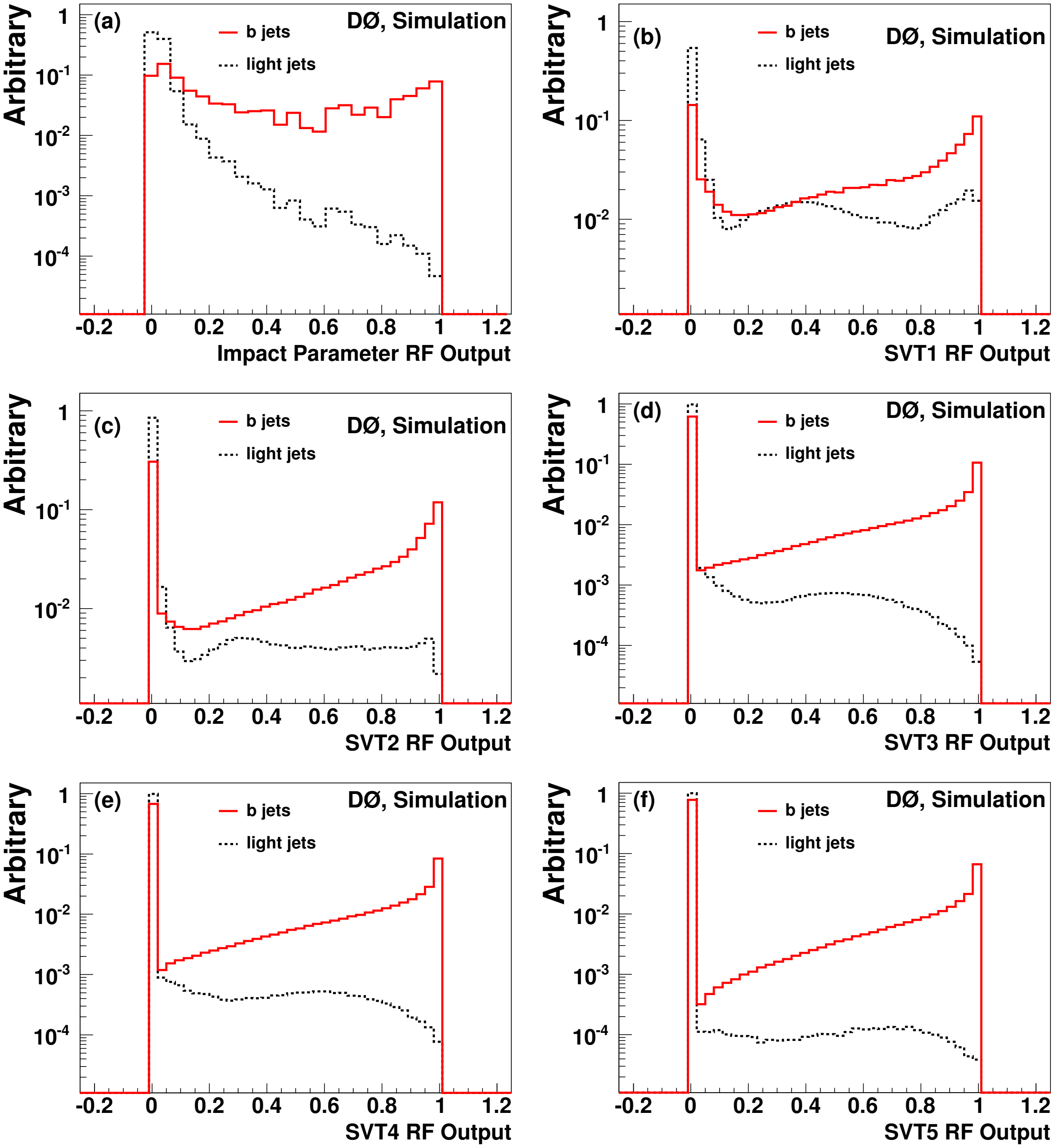


Figure 5.5: Outputs of the 6 trained RFs: (a) IP variables and (b-f) five configurations of the SVT for the b and light jet simulated samples. All distributions are normalized to unity.

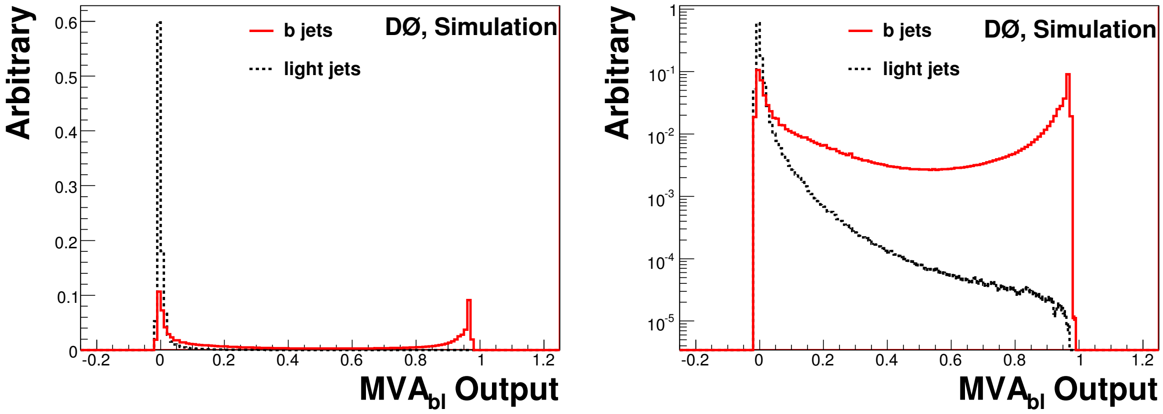


Figure 5.6: The MVA$_{bl}$ output for light flavored jets (dashed line) and b jets (solid line) in simulated events, linear scale on the right, logarithmic scale on the left. Both distributions are normalized to unity.

10^6 simulated multijet events. The parameters selected to optimize the RF were the number of trees in the forest (5) and the number of variables considered at each random split (all). The parameters used for building the final neural network discriminant were the number of nodes (7 input, 1 hidden, and 1 output), and the number of training iterations (50).

5.3.3 Algorithm performance in simulation

The output of the MVA$_{bl}$ algorithm can be seen in Figure 5.6. When compared to the previous D0 NN algorithm, the MVA$_{bl}$ algorithm gives a significant improvement in discriminating power. A measure of this discriminating power is given by the performance profile, or the identification efficiency of a b jet as a function of the misidentification rate. The comparison of the D0 NN and MVA$_{bl}$ algorithms' performance can be seen in Figure 5.7. At low values of misidentification rates the MVA$_{bl}$ performs much better than the D0 NN algorithm, while at high values they yield similar b identification efficiencies. The D0 NN algorithm was trained using only nine variables combined with a neural network. The enhanced performance of the MVA$_{bl}$ algorithm can be attributed to the large increase in the number of variables used in the training and the two step procedure which allowed for a combination of complimentary information.

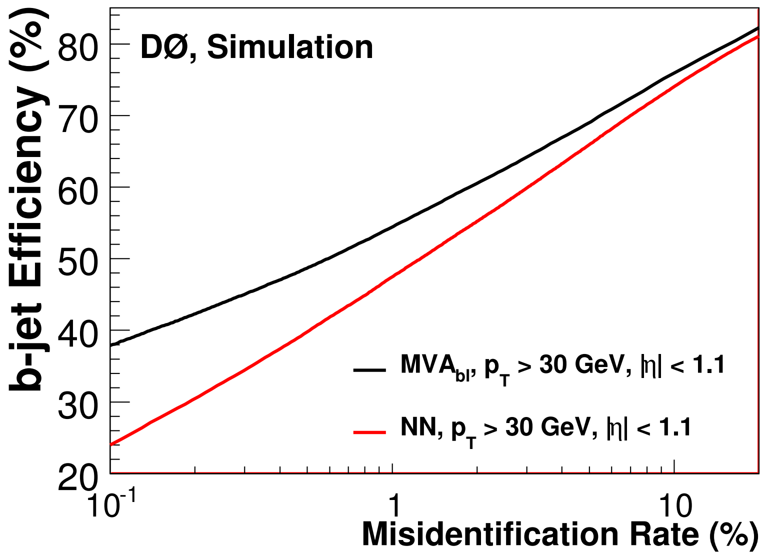


Figure 5.7: The MC performance profile of the MVA$_{bl}$ (in black) and D0 NN algorithm [58] (in red) for jets with $|\eta| < 1.1$ and $p_T > 30$ GeV.

We can create a set of benchmark points, known as operating points (OPs), for which we precisely measure the efficiency and misidentification rates. For the MVA$_{bl}$ algorithm there are twelve OPs assigned the following names:

L6, MVA$_{bl} > 0.02$; **L5**, MVA$_{bl} > 0.025$; **L4**, MVA$_{bl} > 0.035$; **L3**, MVA$_{bl} > 0.042$;

L2, MVA$_{bl} > 0.05$; **Loose**, MVA$_{bl} > 0.075$; **oldLoose**, MVA$_{bl} > 0.1$;

Medium, MVA$_{bl} > 0.15$; **Tight**, MVA$_{bl} > 0.225$; **VeryTight**, MVA$_{bl} > 0.3$;

UltraTight, MVA$_{bl} > 0.4$; **MegaTight**, MVA$_{bl} > 0.5$.

These OPs are displayed in Figure 5.8 where the identification efficiency for b jets and the misidentification efficiency for light jets are shown as a function of the MVA$_{bl}$ output for the simulations.

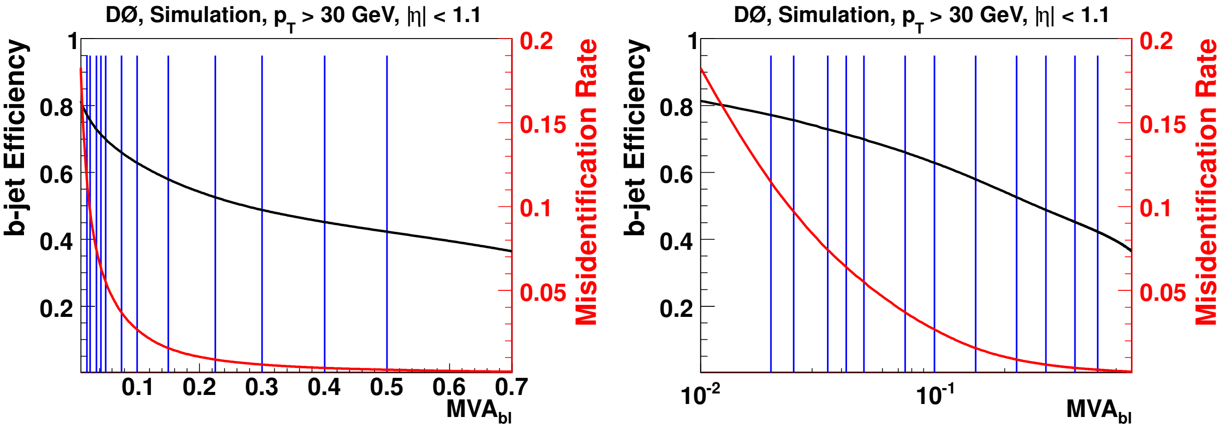


Figure 5.8: The efficiency for selecting a b jet, in black, and the misidentification rate, in red, as a function of the MVA$_{bl}$ requirement. The blue vertical lines correspond to the selected OPs as described in Section 5.3.3.

5.4 Efficiency estimation

Once the algorithm has been defined and its performance is quantified in simulations, we want to measure this performance in data and correct the simulations for any modeling deficiencies. This is a two-step process where we must first determine the efficiencies of this algorithm in both data and MC and then use the ratio of these efficiencies to correct our simulation.

5.4.1 System8 method

The *System8* (S8) formalism allows for the b jet identification efficiencies to be measured directly from data [58]. A system of eight equations with eight unknowns is constructed where the solutions to these nonlinear equations includes the efficiency for selecting b jets.

To determine the efficiency of identifying a b jet we want to construct a HF-enriched data sample. To do this we require that the events contain two back-to-back jets, $\Delta\phi(\text{jet }1, \text{jet }2) > 2.5$, where one must have a $p_T > 15$ GeV and $|\eta| < 2.5$ and be matched to a muon inside a cone of radius $\Delta R = 0.5$. These jets are referred to as muonic jets. The muons inside these jets have the additional requirement that $p_T^{\mu} > 4$ GeV. This sample of events, which is now enriched in HF jets, is contaminated with light jets due to the muonic decays of $\pi^{\pm}$ and $K^{\pm}$.

Since the S8 method can only accommodate a single background sample, we combine the c and light jet background contributions into a single sample referred to as cl jets.

Three additional identification requirements, or “tags”, are individually applied to these jets to further enrich this sample. The first possible *tag* is that the selected muonic jet is required to fulfill a standard MVA$_{bl}$ OP. The second possible *tag* that can be applied is to set a requirement on p_T^{μ} relative to the direction obtained by adding the muon and jet momenta, also known as the p_T^{rel}. Requiring that $p_T^{rel} > 0.5$ GeV allows us to select additional b jets since the high b-quark mass will lead to large muon p_T^{rel}. The final *tag* is a requirement that the jet which is recoiling against the selected jet has JLIP < 0.005. This “away-side tag” allows us to select a sample of pair-produced back-to-back b jets. Using the JLIP to *tag* this *away* jet leads to an enrichment in the overall HF content while not applying any requirement on the selected jet. The following coefficients are introduced into the S8 formulation to account for possible correlations between these *tags*:

- β: Correlations between the *away* tag and MVA$_{bl}$ requirements for b jets.

- α: Correlations between the *away* tag and MVA$_{bl}$ requirements for non-b jets.

- κ_b: Correlations between the p_T^{rel} and MVA$_{bl}$ requirements for b jets.

- κ_{cl}: Correlations between the p_T^{rel} and MVA$_{bl}$ requirements for non-b jets.

The above *tags* are denoted as m, for the MVA$_{bl}$ requirement; p, for the p_T^{rel} requirement; and, a, for the *away* tag. These will be applied independently or concurrently and will appear as superscripts in the following system of equations. Once these requirements are applied to our sample we can build the S8 equations:

$$\begin{array}{lcll}
f_b & + & f_{cl} & = & 1 \\
f_b\varepsilon_b^m & + & f_{cl}\varepsilon_{cl}^m & = & V^m \\
f_b\varepsilon_b^p & + & f_{cl}\varepsilon_{cl}^p & = & V^p \\
f_b\varepsilon_b^a & + & f_{cl}\varepsilon_{cl}^a & = & V^a \\
f_b\kappa_b\varepsilon_b^m\varepsilon_b^p & + & f_{cl}\kappa_{cl}\varepsilon_{cl}^m\varepsilon_{cl}^p & = & V^{m,p} \\
f_b\varepsilon_b^p\varepsilon_b^a & + & f_{cl}\varepsilon_{cl}^p\varepsilon_{cl}^a & = & V^{p,a} \\
f_b\beta\varepsilon_b^a\varepsilon_b^m & + & f_{cl}\alpha\varepsilon_{cl}^a\varepsilon_{cl}^m & = & V^{b,m} \\
f_b\kappa_b\beta\varepsilon_b^m\varepsilon_b^p\varepsilon_b^a & + & f_{cl}\kappa_{cl}\alpha\varepsilon_{cl}^m\varepsilon_{cl}^p\varepsilon_{cl}^a & = & V^{m,p,a}
\end{array} \tag{5.2}$$

where the subscripts b and cl refer either to b or cl jets, V refers to the fraction of the total number of selected jets in the sample which passed a given *tag*, f_X denotes the fraction of events of a given flavor X in the initial *un-tagged* sample, and ε_X^Y refers to the efficiency of a jet of flavor X passing *tag* Y. V is determined from data and α, β, κ_b, and κ_{cl} are determined from simulations [58]. This leaves eight remaining unknowns which can be extracted, including the variable of interest, ε_b^m which is the efficiency of selecting a b jet with the MVA$_{bl}$ requirement. These equations give two possible solutions for ε_b^Y that can be resolved by requiring an additional constraint: $\varepsilon_b^Y > \varepsilon_{cl}^Y$.

The b jet identification efficiency obtained with the S8 method is valid for jets from a semi-leptonic decay of a B hadron. To obtain the efficiency for inclusive b jet decays which is not biased by the requirement of such a decay, a correction factor is determined by using two samples of simulated b jets: where the B hadrons decay either inclusively or semi-leptonically.

The final efficiency is then defined as

$$\varepsilon_b^{data} = \frac{\varepsilon_{b\to\mu X}^{data} \times \varepsilon_b^{MC}}{\varepsilon_{b\to\mu X}^{MC}} = SF \times \varepsilon_b^{MC} \tag{5.3}$$

where $SF = \varepsilon_{b\to\mu X}^{data}/\varepsilon_{b\to\mu X}^{MC}$ is the data-to-simulation efficiency correction factor, and $\varepsilon_{b\to\mu X}^{data}$ is the efficiency as measured by the S8 equations. The identification efficiency for c jets is not measured directly from the data. It is assumed that the data-to-simulation scale factor is identical for b and c jets. The c jet identification efficiency is then derived from the simulation as

$$\varepsilon_c^{data} = SF \times \varepsilon_c^{MC}. \tag{5.4}$$

5.4.2 MVA$_{\mathbf{bl}}$ efficiency

Using this methodology we are able to extract the data efficiencies for a set of specific MVA$_{bl}$ requirements. We have selected two OPs, oldLoose and MegaTight, that will be used later for enriching our $Z+$jet sample with HF jets. These choices are discussed further in Chapter 6.

In Figure 5.9 the efficiency for identifying a muonic b jet, $\varepsilon_{b\to\mu X}$, is shown for data and MC. We also see the ratio of these two efficiencies which form the SFs that will be applied to ε_b^{MC}. Figure 5.10 and Figure 5.11 show the MC and data corrected efficiencies for b and c jets in dijet events, respectively. These efficiency curves are corrected with the parameterized correction factors based on Figure 5.9. Finally, in Figure 5.12, we show the total systematic uncertainty for the S8 method, as discussed in Ref. [58], parameterized as a function of jet p_T.

5.5 Misidentification rate determination

Understanding how the misidentification rates behave in data is allows us to suppress the large uncertainties of the $Z+c$ jet cross section measurement (described in Chapter 8). Pre-

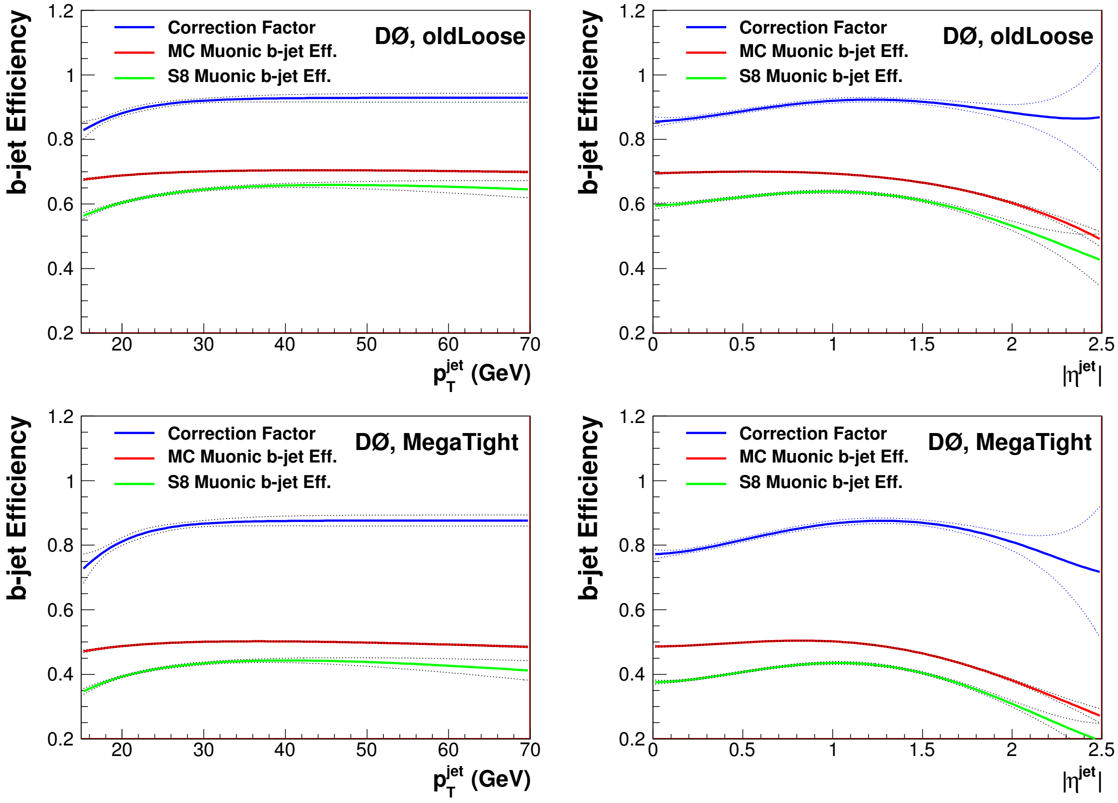


Figure 5.9: The efficiency for selecting a muonic b jet in MC, shown in red, and data, in green, using the S8 method. The correction factor, SF, which will be used to model the algorithm's efficiency is shown in blue. Two OPs are shown, the oldLoose, on the top row, and MegaTight, on the bottom row. The efficiencies are parameterized as a function of p_T, for central jets, on the left, and as a function of η on the right. The band which surrounds the lines corresponds to $\pm 1\sigma$ total uncertainties.

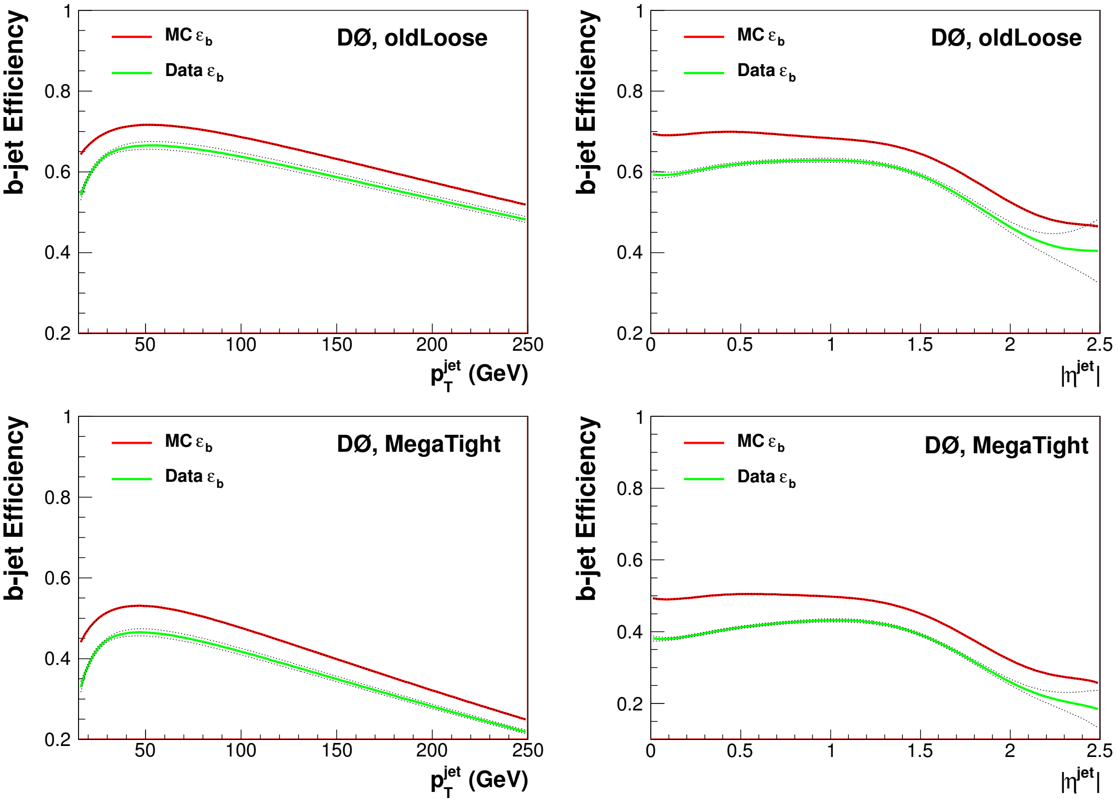


Figure 5.10: The MC b jet identification efficiency, as measured in dijet events, shown in red, along with the data b jet identification efficiency, in green. Two OPs are shown, the oldLoose, the top row, and MegaTight, the bottom row. The efficiencies are parameterized over p_T, for central jets, on the left, and over η on the right.

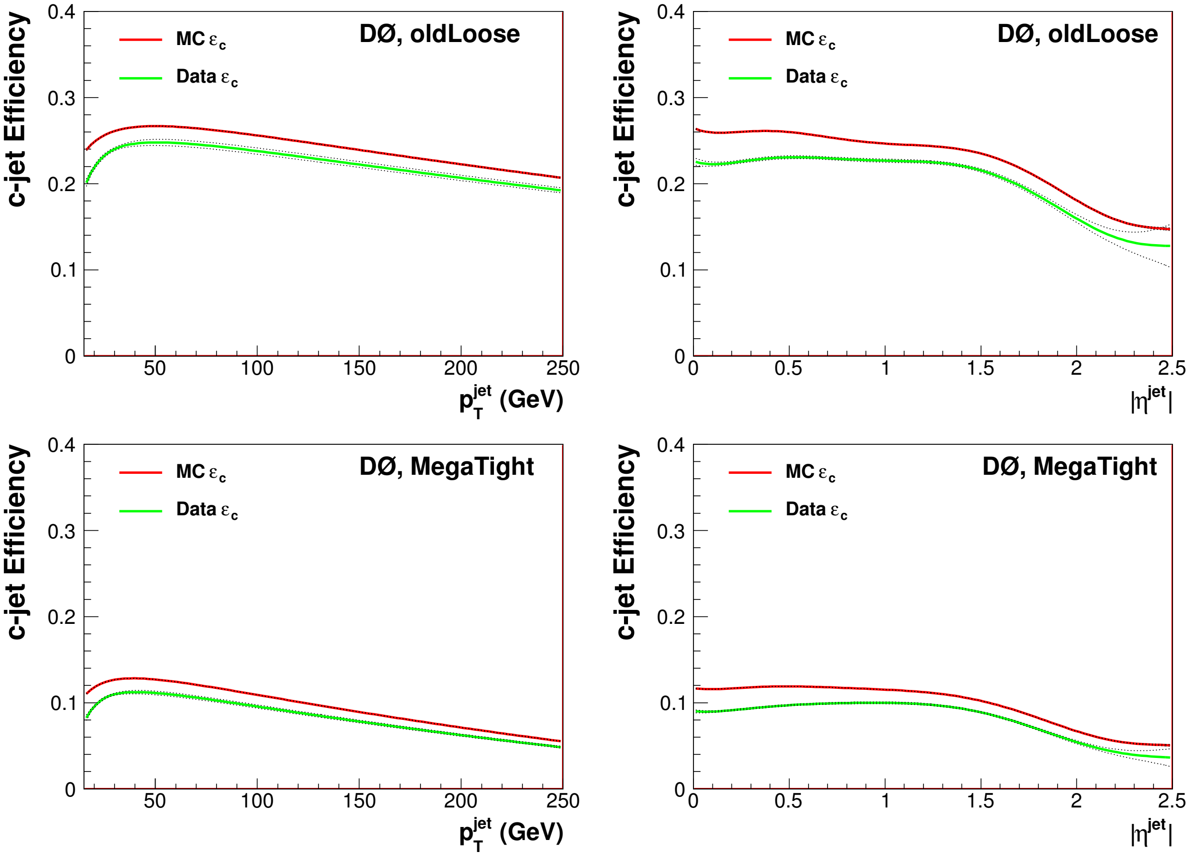


Figure 5.11: The MC c jet identification efficiency, as measured in dijet events, shown in red, along with the data c jet identification efficiency, in green. Two OPs are shown, the oldLoose, the top row, and MegaTight, the bottom row. The efficiencies are parameterized over p_T, for central jets, on the left, and over η on the right.

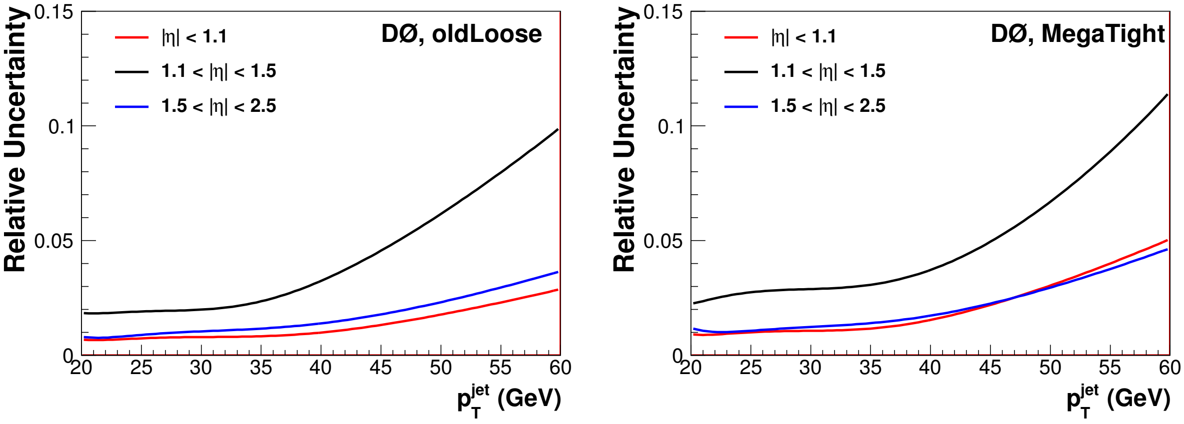


Figure 5.12: The total uncertainty as a function of p_T as measured for two choices of OPs: oldLoose, on the left, and MegaTight, on the right. Three different ranges of jet η were selected $|\eta| < 1.1$, in red, $1.1 < |\eta| < 1.5$, in black, and $1.5 < |\eta| < 2.5$, in blue.

vious methods to determine the misidentification rate relied heavily on simulation [58, 69, 70, 71]. The method described in Ref. [58] for estimating the misidentification rate uses "negatively tagged" (NT) jets, or those with negative IP, with many inputs from simulation. Here we present the *SystemN* (SN) method which was developed to extract misidentification efficiencies directly from data. This method involves several nearly-uncorrelated identification criteria applied to the same data sample. Combining these criteria allows for the definition of a system of equations which can be solved to extract the efficiency of each criterion.

5.5.1 SystemN method

The SN method relies on solving a system of equations based on a binned MVA_{bl} output distribution. The unknown variable of primary interest is the light jet misidentification rate. Additional unknowns are the b, c, and light jet composition. We rely on the b and c jet efficiencies provided by the S8 method to constrain these equations.

The MVA_{bl} output is divided into bins determined by the OP boundaries. If we have n OPs, then there will be n+1 bins, with each bin containing all the jets between the two consecutive OP boundries. The measured b and c jet efficiency distributions from the S8 method are used to predict the rate for selecting b and c jets in each bin. An equation relating the number of jets of each flavor, along with their identification efficiencies, to the total number of retained jets in each bin is formed:

$$N = \varepsilon_l n_l + \varepsilon_c n_c + \varepsilon_b n_b \tag{5.5}$$

where N is the number of selected jets in that bin, ε_X is the efficiency to identify a jet of flavor X, and n_X is the number of jets of flavor X in the total sample. For this method a dijet sample is used without any restrictions on the jets, known as the *all jet* sample. As an example the full set of equations for a selection of five arbitrary OPs are given below (a total of twelve OPs are selected for the final analysis):

$$\begin{array}{lclclcl}
\varepsilon_l^{OP5} n_l & + & \varepsilon_c^{OP5} n_c & + & \varepsilon_b^{OP5} n_b & = & N_{OP5} \\
\varepsilon_l^{OP4-5} n_l & + & \varepsilon_c^{OP4-5} n_c & + & \varepsilon_b^{OP4-5} n_b & = & N_{OP4-5} \\
\varepsilon_l^{OP3-4} n_l & + & \varepsilon_c^{OP3-4} n_c & + & \varepsilon_b^{OP3-4} n_b & = & N_{OP3-4} \\
\varepsilon_l^{OP2-3} n_l & + & \varepsilon_c^{OP2-3} n_c & + & \varepsilon_b^{OP3-4} n_b & = & N_{OP2-3} \\
\varepsilon_l^{OP1-2} n_l & + & \varepsilon_c^{OP1-2} n_c & + & \varepsilon_b^{OP1-2} n_b & = & N_{OP1-2} \\
\varepsilon_l^{aOP1} n_l & + & (1-\varepsilon_c^{OP1}) n_c & + & (1-\varepsilon_b^{OP1}) n_b & = & N - N_{OP1}
\end{array} \tag{5.6}$$

The aOP1 point, or anti-OP1, is all the jets which fall below the OP1 requirement. It is calculated from the other misidentification rates which are all allowed to float in a fit:

$$\varepsilon_l^{aOP1} = 1 - \sum_{x=OP}^{n_{OP}} \varepsilon_l^x \tag{5.7}$$

where n_{OP} corresponds to the number of OPs and x to the MVA$_{bl}$ intervals OP1-OP2, OP2-OP3, OP3-OP4, OP4-OP5 and >OP5. To be solvable these equations require additional constraints. These include the number of jets of a given flavor in the *all jet* sample, n_X, which are extracted from the data by relying on the M$_{\text{SV}}$and performing a template fit, described in Section 5.5.2.

5.5.2 Sample composition

A measurement of the overall flavor composition of the data sample is obtained by fitting M$_{\text{SV}}$ templates for b, c, and light jets to a data distribution. These fits will determine the number of b and c jets in a subsample of the *all jet* sample where the jets pass an SVT and MVA$_{bl}$ requirement, $n_b^{M_{SV}}$ and $n_c^{M_{SV}}$. The sample composition of the *all jet* sample is calculated by extrapolating from the HF sample using b and c jet selection efficiency distributions. These are measured using the S8 procedure for jets passing the MVA$_{bl}$ and SVT requirements. The sample composition measurement is parameterized over jet p_T and η by separating the data into bins of these variables.

A combination of data and MC is used to estimate the M$_{\text{SV}}$ template shapes for the different jet flavors. For b and c jets a data-to-MC correction factor is estimated by comparing the M$_{\text{SV}}$ distributions from a highly HF jet enriched data sample to the MC templates on a bin-by-bin basis and correcting the shape of the MC M$_{\text{SV}}$ templates by any difference. For light jets, M$_{\text{SV}}$ template shapes are estimated using data enriched in light jet, specifically a sample of jets which are designated as NT [58].

In both cases, neither the MC or data templates are expected to perfectly reproduce the M$_{\text{SV}}$ distributions. To account for this, an average of the MC and data-driven M$_{\text{SV}}$ predictions are used to estimate the nominal sample composition and the full difference between the data and MC predictions.

Corrections to the heavy flavor templates

To obtain an estimate of the HF jet M$_{\text{SV}}$ shape from data, a highly HF-enriched dijet sample is constructed by requiring:

- Two taggable jets with a separation of $\Delta\phi(\text{jet 1, jet 2}) > 2.5$
- The selected jet must pass both an MVA$_{bl}$ and SVT requirement
- The recoiling jet must be matched to a muon with $p_T^{\mu} > 8$ GeV and pass an SVT requirement with M$_{\text{SV}} > 1.8$ GeV

A combined M$_{\text{SV}}$ template is constructed by mixing simulated b, c, and light jets using the predicted sample composition. The correction factor is used to alter the HF M$_{\text{SV}}$ template shapes in the separate p_T bins. An example of the corrected mass template is shown in Figure 5.13.

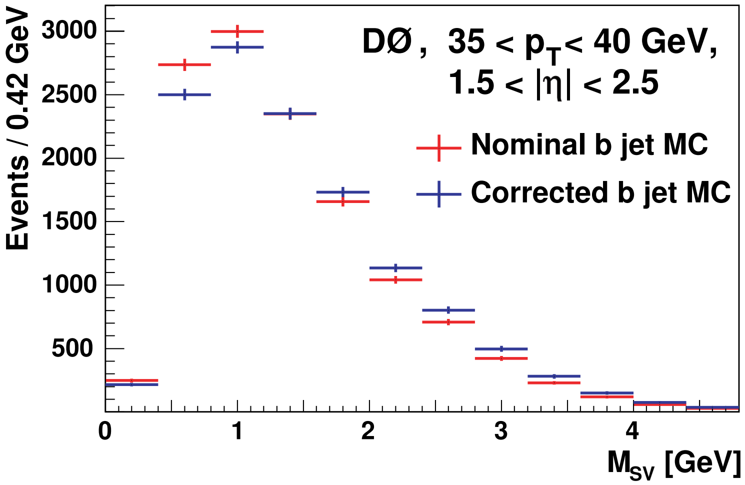


Figure 5.13: Comparison of MC (red) and corrected (blue) b jet $\mathrm{M_{SV}}$ template shapes for jets with $1.5 < |\eta| < 2.5$ and $35 < p_T < 45$ GeV.

Data-driven light jet templates

Due to the difficulty of obtaining a pure sample of light jets in data, the uncertainty due to the shape of the template is estimated by using a sample of NT jets from data. To build this sample it is required that the jets have negative IP and pass an SVT selection. The distribution will be contaminated by the presence of HF jets and as such will not be a perfect representation of the light jet $\mathrm{M_{SV}}$ shape in data. The NT template shapes are measured from data in each p_T and η interval. Figure 5.14 shows a comparison between the NT and MC light jet template shapes. The difference in the shapes are taken as a systematic uncertainty.

Sample composition measurement

These templates are then used to fit the data $\mathrm{M_{SV}}$ distribution using a log likelihood fitter in bins of jet p_T and η. An example of an individual fit of the data $\mathrm{M_{SV}}$ distribution with the data-corrected template shapes can be found in Figure 5.15. This fit results in a measurement of the fraction of each flavored jet type in that bin. The fits in each of the p_T and η regions are subsequently extrapolated back to the full *all jet* sample using the b and c jet efficiency distributions measured for the MVA_{bl} and SVT algorithms. The number of events of flavor

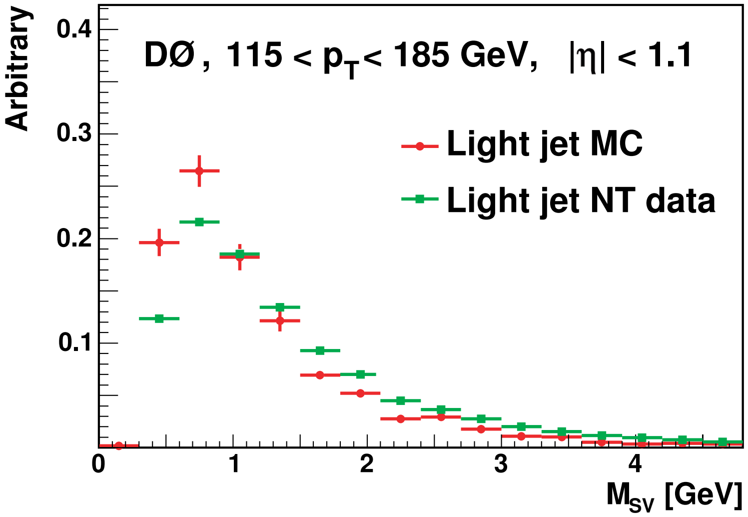


Figure 5.14: Comparison of the MC light jet and NT $\mathrm{M_{SV}}$ mass templates for jets with $|\eta| < 1.1$ and $115 < p_T < 185$ GeV.

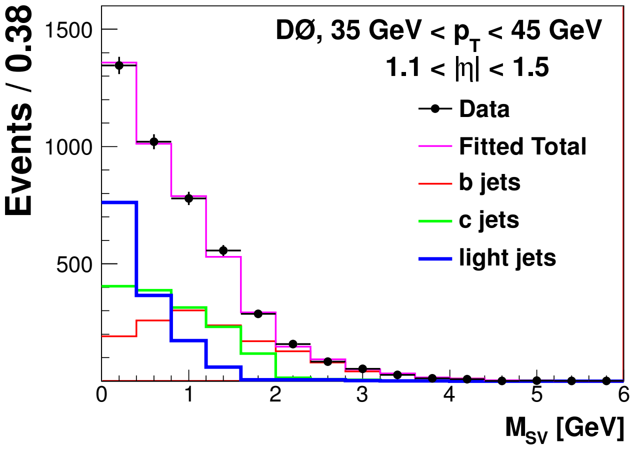


Figure 5.15: An example of the sample composition fit using the $\mathrm{M_{SV}}$ for jets which pass MVA_{bl} and SVT requirements and has a $35 < p_T < 45$ GeV and $1.1 < |\eta| < 1.5$. The b (red), c (green), and light jets (blue) are fit to the data (crosses) resulting in the total fitted contribution (purple).

X in the *all jet* sample is calculated using the following formula:

$$n_{b/c} = N \times f_{b/c} = N \times \frac{f^{Tag}_{b/c}}{\varepsilon^{Tag}_{b/c}} \tag{5.8}$$

where $f^{Tag}_{b/c}$ is the fraction of b or c jets extracted from the HF-enriched sample, f_X is the fraction of jets of flavor X in the *all jet* sample, $\varepsilon^{Tag}_{b/c}$ is the MVA$_{bl}$ and SVT efficiency, and N is the total number of events in that bin. The efficiency is calculated for the average p_T and η of the jets in the region. While $f^{Tag}_{b/c}$ can be corrected to the *all jet* sample, the light jet fraction cannot be. The corresponding light jet fraction in the *all jet* sample is then determined from $f_l = 1 - f_b - f_c$.

The parameterization of the *all jet* sample composition is important so that we can obtain a smoothly varying misidentification rate as a function of p_T and to minimize the effect of statistically limited bins at high p_T. However, the choice of parameterization is not straight-forward. The optimal parameterizations were determined by considering the χ^2 probability of various functional forms, including linear and logarithmic polynomials.

5.5.3 Solutions of the SystemN equations

Instead of solving Equation 5.6 analytically, we form a likelihood function. In this likelihood function we take the SN equations and compare them to what is predicted from simulations. We allow the extracted flavor fractions, f_X, to float within their uncertainties during this fit. To help constrain this likelihood function a second set of SN equations are built using a new data sample, the full procedure is repeated, and it is added to the likelihood fit. This second sample is a sub-set of the *all jet* sample which has an additional requirement on the recoiling “away jet” that it be matched to a muon. This sample will be referred to as the *all jet away* sample.

The resulting likelihood function is then formed by summing over each of the operating bins for both samples,

$$LLH = -2 \sum_{S=All\ \ jet, Away}^{N_S} \sum_{x=OP}^{N_{OP}} (N_x^S \ln(N_x^{MC}) - N_x^{MC}) \tag{5.9}$$

where N_x^S is the number of data events in sample S in the MVA$_{bl}$ interval x, N_x^{MC} is the predicted number of events in interval x, and x represents the intervals: anti-OP1, OP1-OP2, OP2-OP3, OP3-OP4, OP4-OP5, and >OP5. A normalization factor, LLH_{Norm}, is used to ensure that the likelihood values remain well defined:

$$LLH_{Norm} = -2 \sum_{S=All\ \ jet, Away}^{N_S} \sum_{x=OP}^{N_{OP}} (N_x^S \ln(N_x^S) - N_x^S) \tag{5.10}$$

this is subtracted from the likelihood function.

We use the b and c jet fractions measured in Section 5.5.2 to help stabilize the fit through a term which is added to the likelihood function,

$$d^T E^{-1} d \tag{5.11}$$

where E corresponds to the 2×2 covariant matrix of uncertainties resulting from the extraction of the b and c jet content from the M$_{\text{SV}}$ fit and d is a vector

$$d = \begin{pmatrix} n_b - n_b^{M_{SV}} \\ \\ n_c - n_c^{M_{SV}} \end{pmatrix}. \tag{5.12}$$

The result of this likelihood function is the extraction of the data driven light jet efficiency parameterized over jet p_T and η in various MVA$_{bl}$ intervals. These final misidentification rates are shown in Figure 5.16.

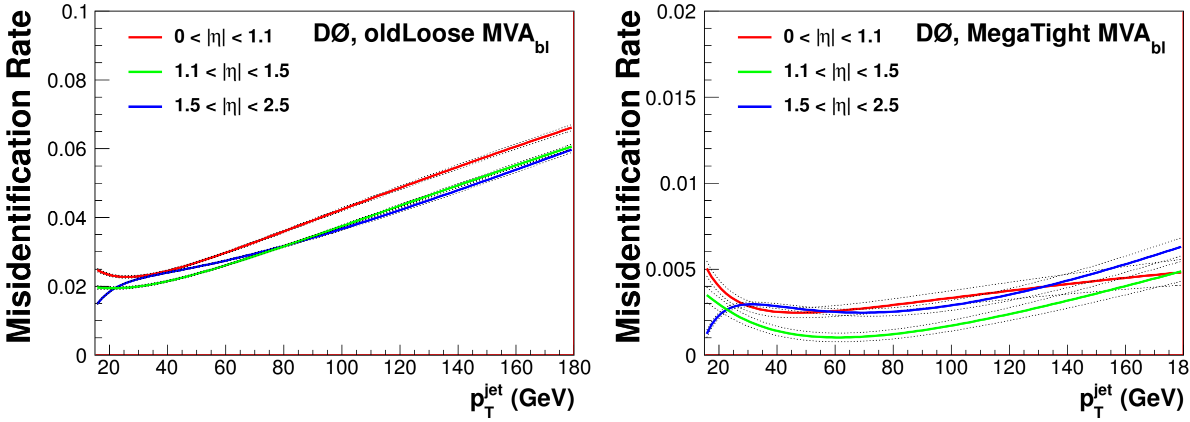


Figure 5.16: The resulting SN data-driven misidentification rates for the MVA$_{bl}$ algorithm. Two OPs are shown, oldLoose, on the left, and MegaTight, on the right. These are further parameterized over jet p_T and for three different jet η intervals: $0 < |\eta| < 1.1$, in red; $1.1 < |\eta| < 1.5$, in green; $1.5 < |\eta| < 2.5$, in blue.

5.5.4 SystemN systematic uncertainties

The three dominant systematic uncertainties that must be accounted for during this procedure are:

- The shape of the b and c jet M$_{\text{SV}}$ templates.
- The shape of the light jet M$_{\text{SV}}$ template.
- The uncertainty on the b and c jet efficiencies from the S8 method.

Heavy flavor templates

The effect of any imperfections in the modeling of the b and c jet M$_{\text{SV}}$ templates is estimated by carrying out the sample composition measurement using a separate set of HF M$_{\text{SV}}$ templates which are not corrected to data in each of the p_T and η intervals. The full difference between the MC and data corrected sample composition predictions are used to assign an uncertainty. Since the MC sample composition is used to determine the shape of the HF template correction factor, we vary this composition and use this varied composition to re-derive the correction factor. The largest deviation from the nominal shape is used to assign an additional uncertainty.

Light flavor templates

The uncertainty due to the shape of the light jet $\mathrm{M_{SV}}$ templates is estimated by remeasuring the sample composition fit using both, the NT and MC light jet template shapes, with the difference in the sample composition used to assign an uncertainty.

b and c jet efficiency

When extrapolating the flavor fractions measured in the HF enriched sample to the *all jet* sample, the efficiencies from the S8 method are used. To account for the uncertainties inherent in this procedure, it is repeated after the efficiencies are varied by $\pm 1\sigma$. This variation will only affect the extrapolation procedure.

The parameterization of the systematic uncertainties is evaluated by carrying out "closure tests." We compare the percentage difference between the number of selected jets and the predicted number of jets in various bins in p_T and η regions. Then the final uncertainty is ascertained from the root mean square of the resulting distributions.

The total uncertainty on the data driven misidentification rates attained using the SN method, given by the statistical and systematic uncertainties combined in quadrature, is shown in Figure 5.17 for the MVA_{bl} algorithm.

5.5.5 Comparison to previous method

A comparison between the misidentification rates measured using the SN method and those estimated by the NT method is shown in Figure 5.18, and they both provide comparable uncertainties. For the looser OPs the new method predicts efficiency distributions 20% higher than the old method. For the tighter OPs the difference is closer to 35%. These two predictions agree within uncertainties across the full range of jet p_T, but the prediction from the NT method is systematically lower.

The source of this disagreement comes from the usage of the simulation in the NT method. With the removal of the V^0s, the main source of the misidentification rate comes from

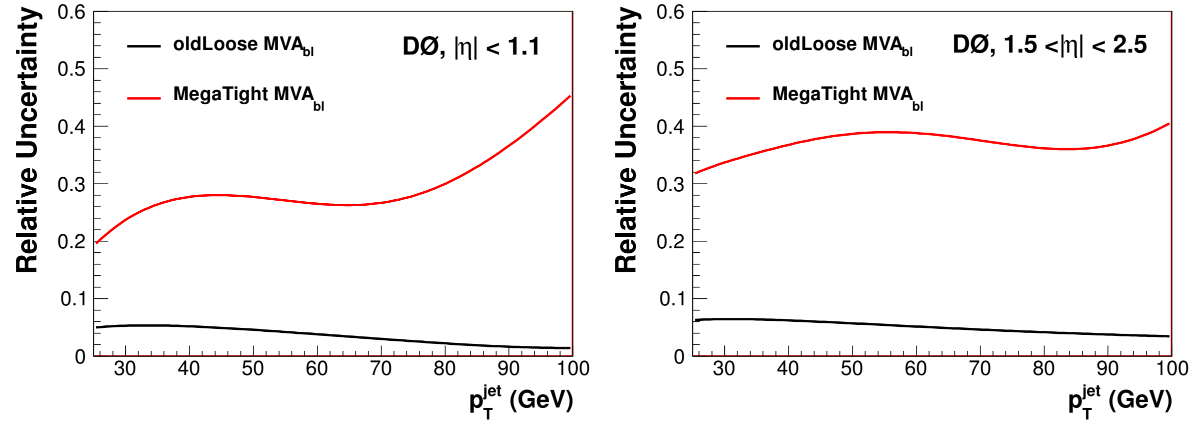


Figure 5.17: The total relative uncertainty of the SN method parameterized in terms of jet p_T and for two different η regions: $|\eta| < 1.1$, on the left; and $1.5 < |\eta| < 2.5$ on the right. Two OPs are shown, the oldLoose, in black, and MegaTight, in red.

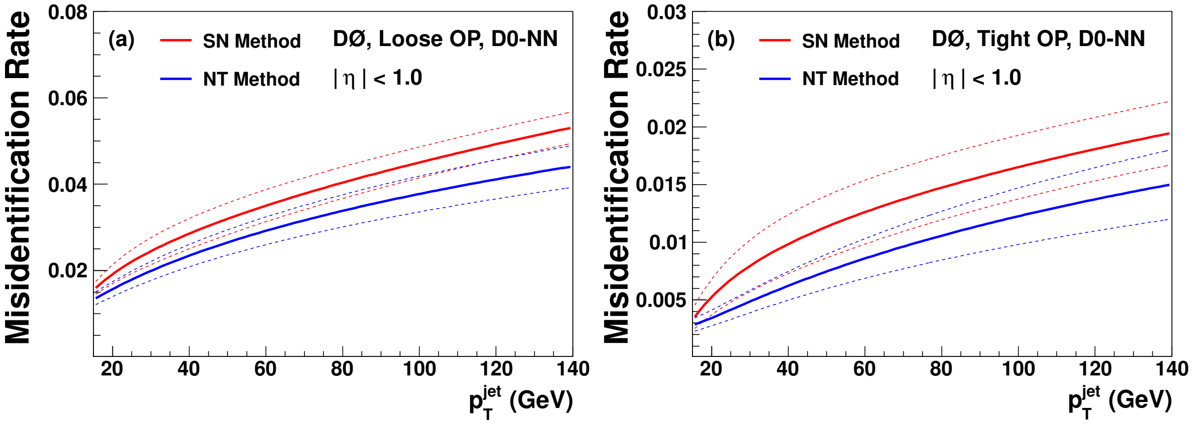


Figure 5.18: A comparison between the misidentification rates of the D0 NN algorithm derived for two OP choices, Loose, on the left, and Tight, on the right, for the new SN method and the old method described in Ref. [58].

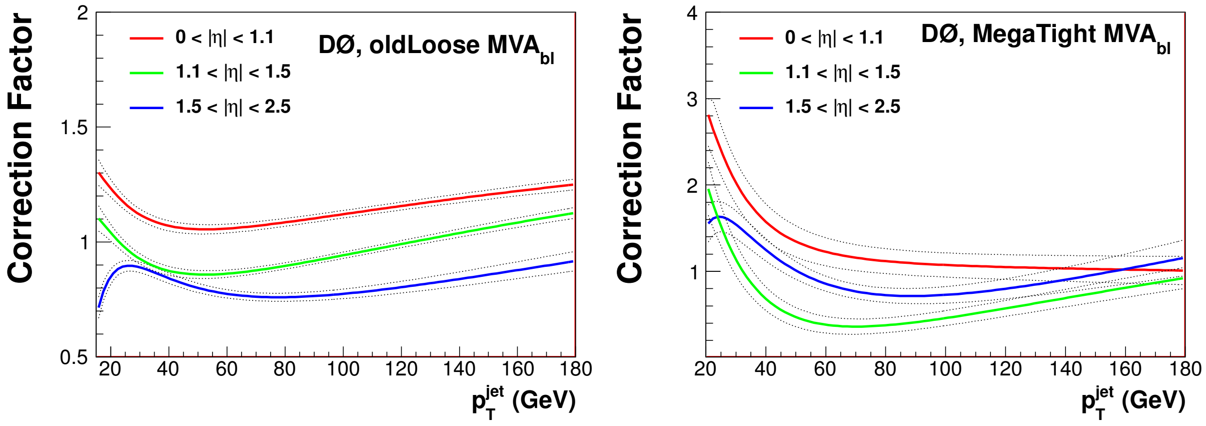


Figure 5.19: The correction factors for the light jet MC which are derived by taking the ratio of the data and MC misidentification rates. Two OPs are shown, oldLoose, on the left, and MegaTight, on the right. These are further parameterized over jet p_T and for three different jet η intervals: $0 < |\eta| < 1.1$, in red; $1.1 < |\eta| < 1.5$, in green; $1.5 < |\eta| < 2.5$, in blue.

the detector resolution and track mis-reconstruction. Simulations inaccurately reproduce these effects. Thereforem the misidentification rate, as determined by the NT method, is underestimated.

5.5.6 MVA$_{\mathbf{bl}}$ misidentification rates

The final result of the SN method is that we can extract directly from data the misidentification rates for light jets. In Figure 5.16 we can see the misidentification rates derived from the data. These are parameterized in terms of p_T for the CC, EC, and ICR calorimeter η regions. This data-driven measurement of the misidentification rate can be combined with the efficiency in simulation and used to derive a correction factor, as shown in Figure 5.19. These correction factors are then applied to the simulation and are used to model the contamination, caused by the light jets, which pass the MVA$_{bl}$ requirements.

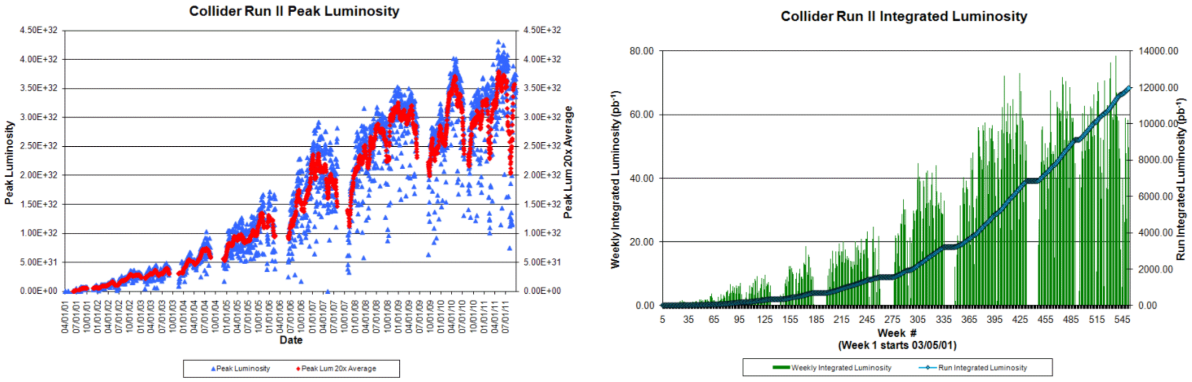


Figure 5.20: The peak instantaneous luminosity as a function of time (left), and the total recorded luminosity for Run II (right) [72].

5.6 Instantaneous luminosity dependence

Over the course of Run II, the Tevatron was able to significantly improve its performance, delivering beams with continually higher instantaneous luminosity. Figure 5.20 shows the peak instantaneous luminosity across the whole Run II data taking period. While this led to the recording of a large amount of data to analyze, it also provided an environment with more “pile up”, or multiple interactions per bunch crossing. Figure 5.20 shows the total recorded data, or integrated luminosity, as a function of time. Understanding how this pile up affects the b identification algorithms is necessary and must be handled correctly.

Studies in the past with the D0 NN algorithm have shown dependence on the instantaneous luminosity, $\mathcal{L}$ [73]. Figure 5.21 shows the results of these studies, where we see that the heavy flavor correction factor has a mild, $1-2\%$ dependence on the instantaneous luminosity. We also can see the dependence on the light jet correction factor which is more substantial, between $5-10\%$. The larger dependence on the misidentification rate is two fold. First, there are additional tracks in the final state due to additional pile up in high luminosity events. Second, the template shapes will also depend strongly on the instantaneous luminosity, which will affect the flavor extraction. These effects are taken into account by parameterizing the corresponding corrections as a function of time. We create a separate set of correction factors from the data for four different periods, known as run epochs. These

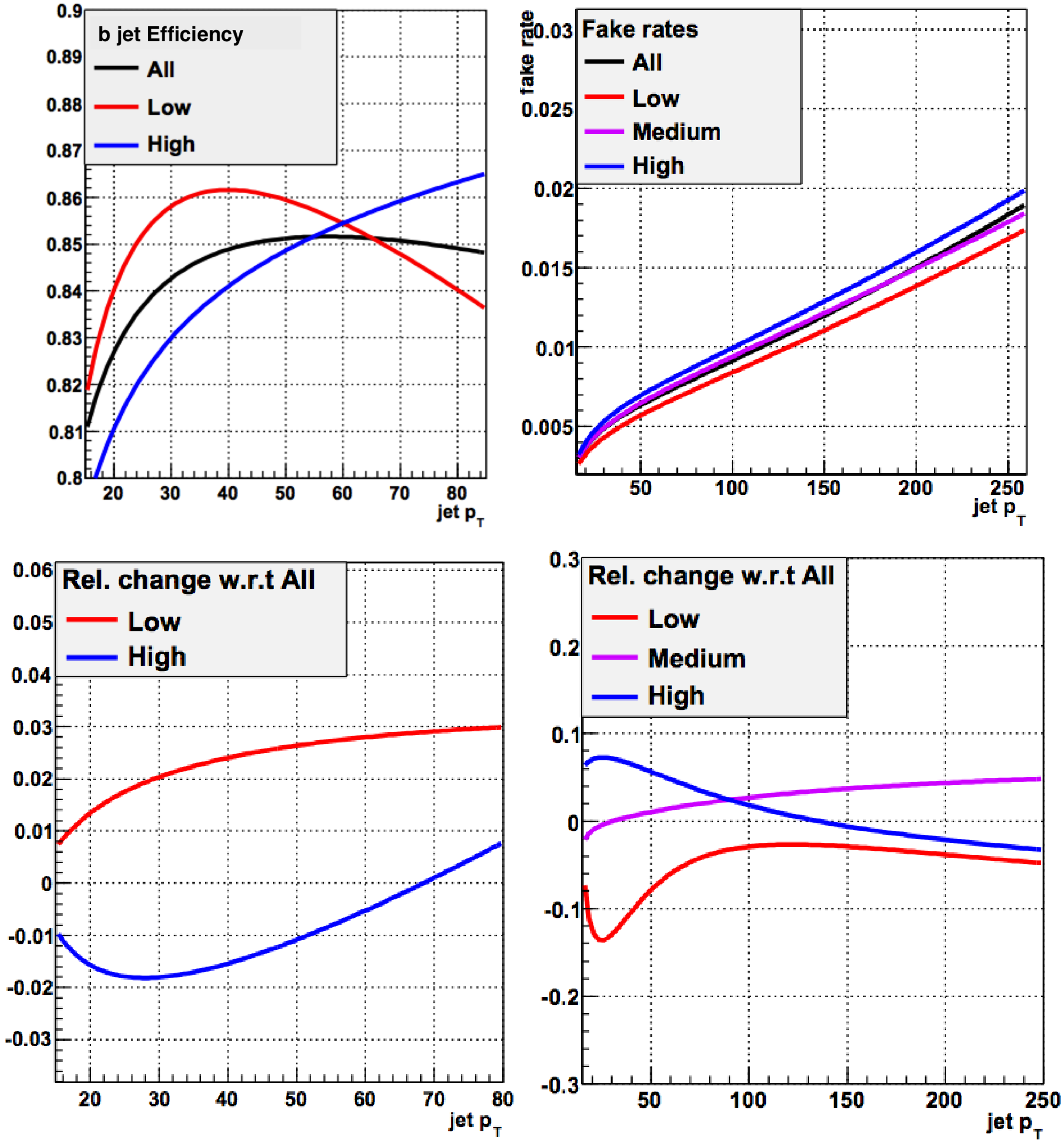


Figure 5.21: The effect of instantaneous luminosity on the heavy flavor scale factor (left) and the light jet scale factor (right) as a function of jet p_T [73]. The "low", "medium", and "high" regions are defined as $\mathcal{L} < 0.6 \times 10^{32}$ cm^{-2}s^{-1}, 0.6×10^{32} cm^{-2}s^{-1} $< \mathcal{L} < 1 \times 10^{32}$ cm^{-2}s^{-1}, and $\mathcal{L} > 1 \times 10^{32}$ cm^{-2}s^{-1}, respectively. For the heavy flavor scale factors the "low" and "medium" curves behaved similarly and were combined to increase the statistical precision. "Rel. change w.r.t All" refers to the ratio of the scale factor determined for a luminosity range to the nominal scale factor.

epochs are used to characterize the changing beam conditions, and changes to the detector configurations across Run II.

6 Analysis methods

To begin measuring the ratio of cross sections we will start with the equation:

$$\frac{\sigma(Z + \mathrm{HF\ jet})}{\sigma(Z + \mathrm{jet})} = \frac{N_{Z+HF\,jet}\mathcal{A}^{jet}}{N_{Z+jet}\varepsilon^{ID}_{HF\,jet}\mathcal{A}^{HF\,jet}} \tag{6.1}$$

where N_{Z+jet} is the number of $Z+$jet events in the total sample, $N_{Z+HF\,jet}$ is the number of $Z+$HF jet events in a sample which has been enriched by a requirement a HF identification algorithm. This selection will have an efficiency of $\varepsilon^{ID}_{HF\,jet}$. Finally, $\mathcal{A}^{jet}$ and $\mathcal{A}^{HF\,jet}$ are the detector acceptance for inclusive and HF jets, respectively. We are more interested in measuring this ratio for a specific flavor of jet therefore we will measure for $Z + b$ jets or $Z + c$ jets instead of $Z +$ HF jets. For example, if we modify Equation 6.1 for $Z + b$ jets the equation will depend on $N_{Z+b\,jet}$, $\varepsilon^{ID}_{b\,jet}$, and $\mathcal{A}^{b\,jet}$. Due to the contamination from other flavors ($Z+c$ jet and $Z+$light jet) it is impossible to count directly the number of $Z+b$ jet events. Therefore we will separate $N_{Z+b\,jet}$ into two components: the total number of events selected by our MVA$_{bl}$ algorithm selection (N^{OP}), and the fraction of those events which are of flavor X (f_X). Such that $N_{Z+b\,jet} = N^{OP}f_b$, where the method for measuring f_X is discussed in Section 6.1. The measurement of the efficiencies which compose the final heavy flavor selection also enters into Equation 6.1 and are discussed in Section 6.2. The estimation of the detector acceptance for the various flavors of jet is discussed in Section 6.3, which also covers the unfolding of the ratio of cross sections. Finally, in Section 6.4 we discuss how we determine the location of the bin center for the differential measurements.

6.1 Measuring the flavor fractions

To begin we must first construct a sample which is enriched in HF events by applying a criteria on the MVA_{bl} requirement. This data sample will still have contamination from light jets and the various other backgrounds. If we create a variable which can discriminate b, c, and light jets we can use a maximum likelihood fitter to determine the fractional contribution of each flavor to the total sample.

To do this we construct a variable from two other variables: $-\ln(\mathrm{JLIP})$ and M_{SV}, which are shown in Figure 6.1. Traditionally, M_{SV} has been used in the template fitting procedures to determine the flavor content of these events (see Section 5.5.2 and Refs. [10, 11, 12, 25, 26, 74]) but we have found that we can exploit additional information about these jets to retain more events and increase the discriminating power. In Figure 6.1(b) we see that a large spike exists at $\mathrm{M}_{\mathrm{SV}} = 0$, this is due to events where a secondary vertex cannot be resolved inside the jet. When fitting with the M_{SV} it is necessary to remove these events before template fitting, these events constitute a large part of the sample and highly suppress our overall statistics. If we include the complimentary information from the JLIP we can benefit from an overall improvement in the discriminating power of the combined discriminant, D_{MJL} [75].

To build this discriminant we combine the variables $-\ln(\mathrm{JLIP})$ and M_{SV} in the following format:

$$D_{MJL} = \frac{M_{SV}/5 - \ln(\mathrm{JLIP})/20}{2}. \tag{6.2}$$

The distribution of D_{MJL} is shown in Figure 6.2, the combined discriminating power of this new variable is greater than its individual components. These templates are dependent on the MVA_{bl} selection that is used and various kinematic variables. These templates are also derived as a luminosity-weighted average with all the derived corrections, applied as described previously in Section 4.3 and Chapter 5. Additionally, to account for any kinematic dependence, they are generated in each differential bin that we study.

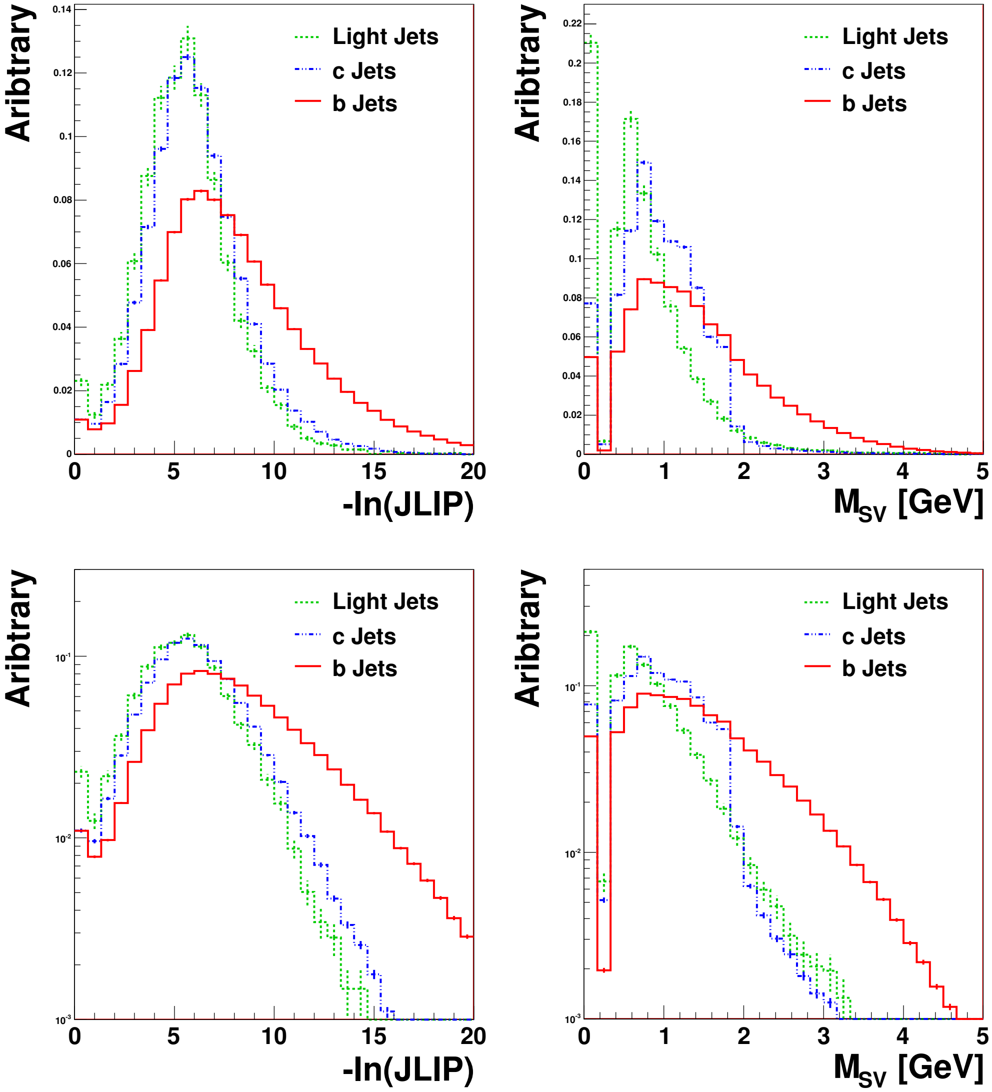


Figure 6.1: The $-\ln(JLIP)$, on the left, and the $\mathrm{M_{SV}}$, on the right, distributions for b, c, and light jets after requiring that $\mathrm{MVA}_{bl} > 0.5$. Using logarithmic scale (bottom) the strong discriminating power for b jets can be observed.

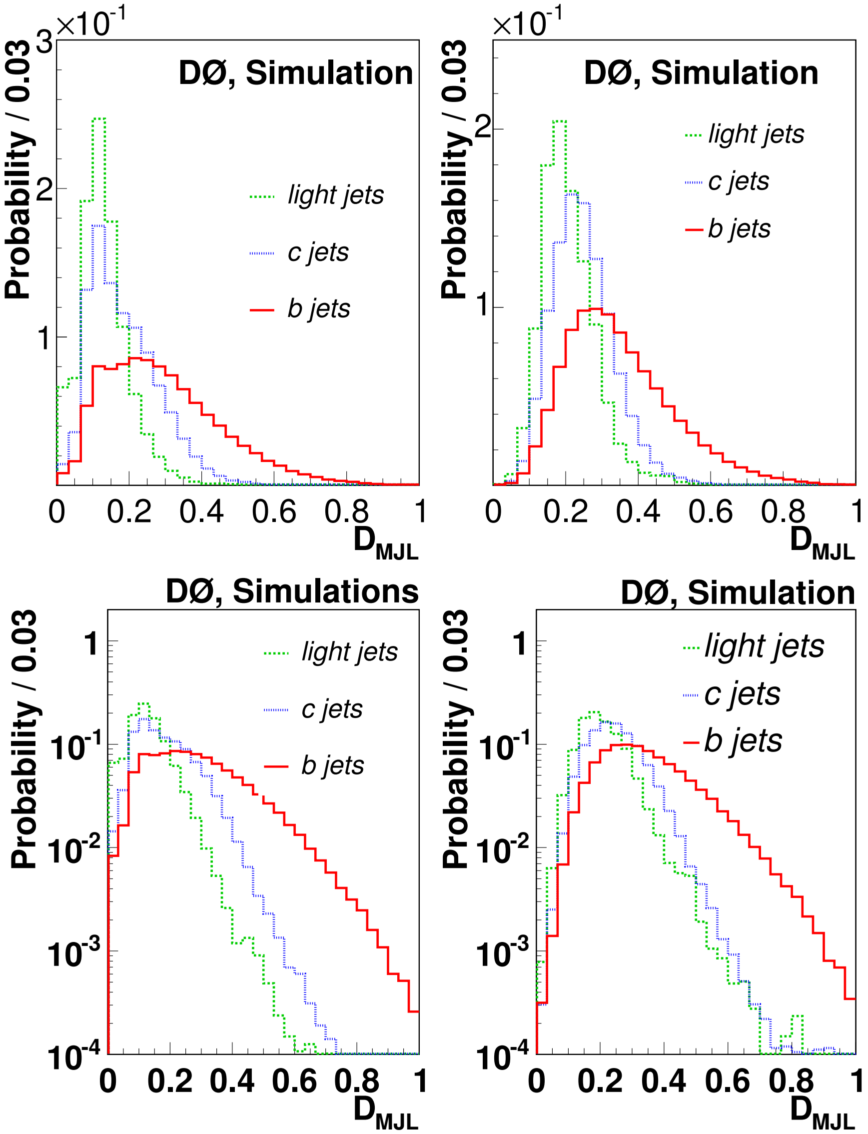


Figure 6.2: D_{MJL} templates for b, c, and light jets after the application the requirments $\mathrm{MVA}_{bl} > 0.1$, on the left, and $\mathrm{MVA}_{bl} > 0.5$, on the right. The bottom row shows these templates on a logarithmic scale which shows the strong discriminating power of the b jet templates.

We remove the residual background (from diboson, $t\bar{t}$, and multijet events) in the final HF enriched sample before fitting. D_{MJL} distributions are constructed for each of the backgrounds and are then subtracted from the data D_{MJL} distribution, weighted by the applicable cross sections discussed in Section 4.1.3. For the $Z + c$ jet analysis the $Z +$ light jet contamination is treated in the same fashion as the other background contributions, which is discussed further in Chapter 8.

6.2 Efficiency measurements

When measuring the ratio of cross sections, the HF identification requirements, both taggability and MVA$_{bl}$, are applied only to the numerator. For each measurement the efficiencies are parameterized in the variable that we are examining. For example, a differential jet η distribution will have the efficiencies measured as a function of jet η. Due to the fact that there are multiple MC releases at D0 to help model detector aging and changing beam conditions, we applied the corrections derived in Chapter 5 for each release and then weight each simulation by their corresponding integrated luminosity. These are then combined to form a single efficiency measurement that is used to correct the cross section.

6.3 Acceptance corrections

To correct our cross section measurements for the detector acceptance we utilize detailed MC simulations. The same kinematic selections as we require in the analysis, listed in Table 6.1, are applied. The acceptance is then defined as:

$$\mathcal{A}_X = \frac{N_X^{Reco}}{N_X^{Part}} \tag{6.3}$$

where N_X^{Reco} and N_X^{Part} are the number of events that are selected by reconstruction and MC "truth" level kinematic and geometric selections, respectively, for events which contain jets

Table 6.1: Criteria applied, both at the reconstruction and particle level, to determine the detector acceptance.

Acceptance Requirements	
N_X^{Reco}	N_X^{Part}
$p_T^{Reco\,Jet} > 20\,\text{GeV}$	$p_T^{Particle\,Jet} > 20\,\text{GeV}$
$\lvert\eta_{Reco\,Jet}\rvert < 2.5$	$\lvert\eta_{Particle\,Jet}\rvert < 2.5$
Reconstruction Level Binning	MC Truth Level Binning
$70 < m_{ll} < 110$ GeV	$70 < m_{ll}^{MC} < 110$ GeV
p_T(leptons) > 15 GeV	p_T(MC leptons) > 15 GeV
$\lvert\eta$(leptons)$\rvert < 2$	$\lvert\eta$(MC leptons)$\rvert < 2$
Number of jets of flavor $X > 0$	

of flavor X.

The ratios of acceptances are comprised of $\mathcal{A}_b$, for b jets, $\mathcal{A}_c$, for c jets, and $\mathcal{A}_{incl}$, for inclusive jets. Since we will be measuring the ratios of the acceptances the contributions from the lepton and Z boson acceptances are highly suppressed, hence the lepton isolation and object identification requirement can be neglected when computing the acceptance.

The acceptance is estimated for the entire detector and it is determined to be

$$\mathcal{A}_b = 0.94 \pm 0.02,\ \mathcal{A}_c = 0.95 \pm 0.02,\ \text{and}\ \mathcal{A}_{incl} = 1.09 \pm 0.02.$$

These values are used to correct the measured cross sections. In addition, we estimate the acceptance in the full range of differential bins that we will be studying. In Figure 6.3 we see the breakdown for jet p_T, Z boson p_T, jet η, and $\Delta\varphi\,(Z, \text{jet})$ where the selected binning is described in Chapter 7.

These corrections also act as a bin-by-bin unfolding procedure. To check whether this unfolding was sufficient and a more sophisticated method was not necessary we studied the migration matrices. A migration matrix is a 2D representation with the reconstruction level information projected on one axis and the MC (or particle) level information projected on the other. Since we measure the ratio of cross sections this bin-by-bin unfolding will correctly

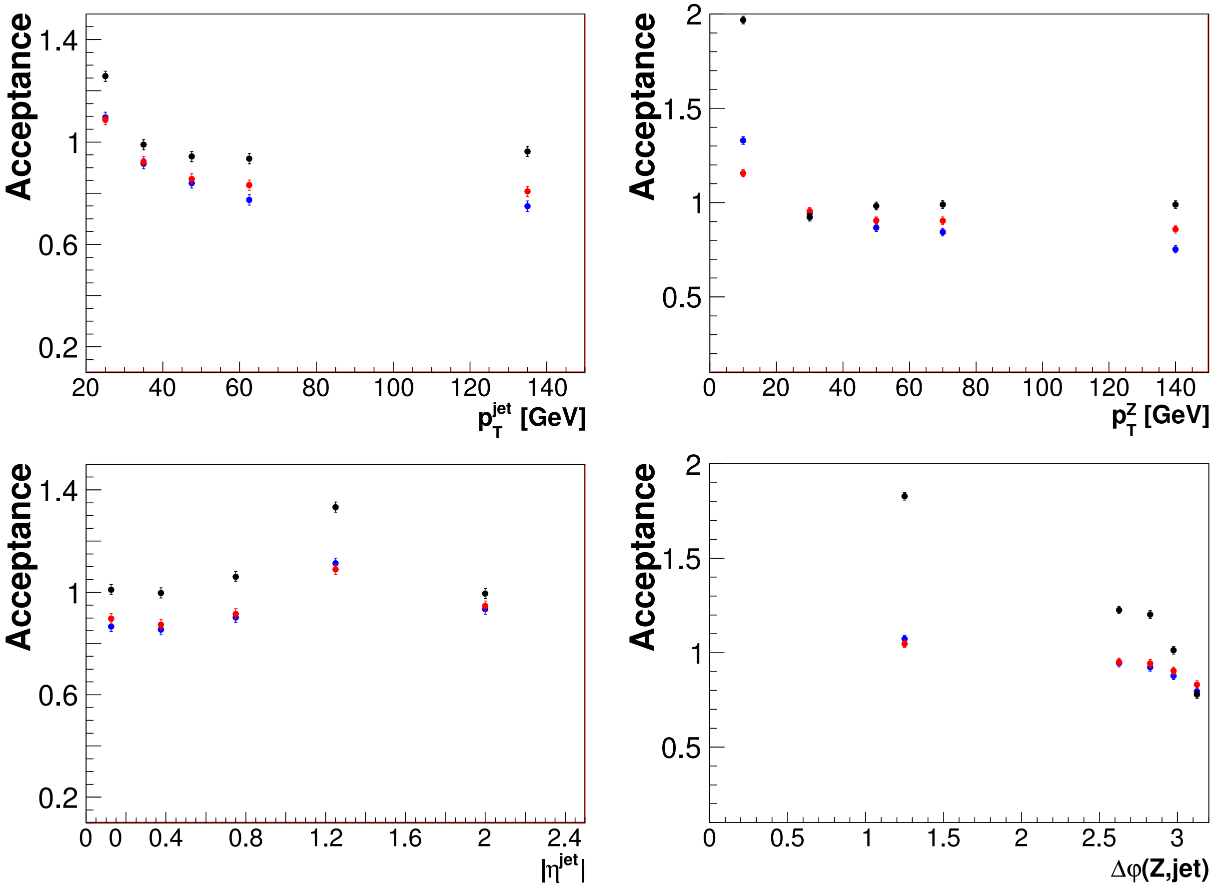


Figure 6.3: Acceptance in each bin of jet p_T (upper left), Z boson p_T (upper right), jet η (lower left), and $\Delta\varphi\,(Z, \text{jet})$ (lower right), for inclusive jets (black), b jets (red), and c jets (blue). The binning was selected for the $Z + b$ jet analysis and is described in Chapter 7.

handle the effects of migration if there is a similar off diagonal migration in the numerator and denominator. The migration matrices for jet p_T (shown in Figure 6.4), Z boson p_T (shown in Figure 6.5), jet η (shown in Figure 6.6), and $\Delta\varphi(Z,\text{jet})$ (shown in Figure 6.7) show similar migration off the diagonal for all flavors. This means that a bin-by-bin unfolding correction will be sufficient to account for any migration effects.

6.4 Differential cross section bin centering

The position of the data point within the bin interval, or bin center, is very important for comparing to theoretical predictions. Instead of using the geometric or weighted mean bin center we instead use the Lafferty-Wyatt method for locating the center of wide bins [76]. This is done by looking at the distribution of the cross section over a given bin and using a function to fit this distribution. Mathematically, we have a cross section whose distribution is fit by a function $g(x)$, to determine the location of the bin center, x_{lw}, in the range of $\Delta x = x_2 - x_1$ we solve for x_{lw}

$$g(x_{lw}) = \frac{1}{\Delta x}\int_{x_1}^{x_2} g(x)\mathrm{d}x. \tag{6.4}$$

Throughout this thesis all bin centers will be found with this method and will appear in the tables in parentheses next to the bin range.

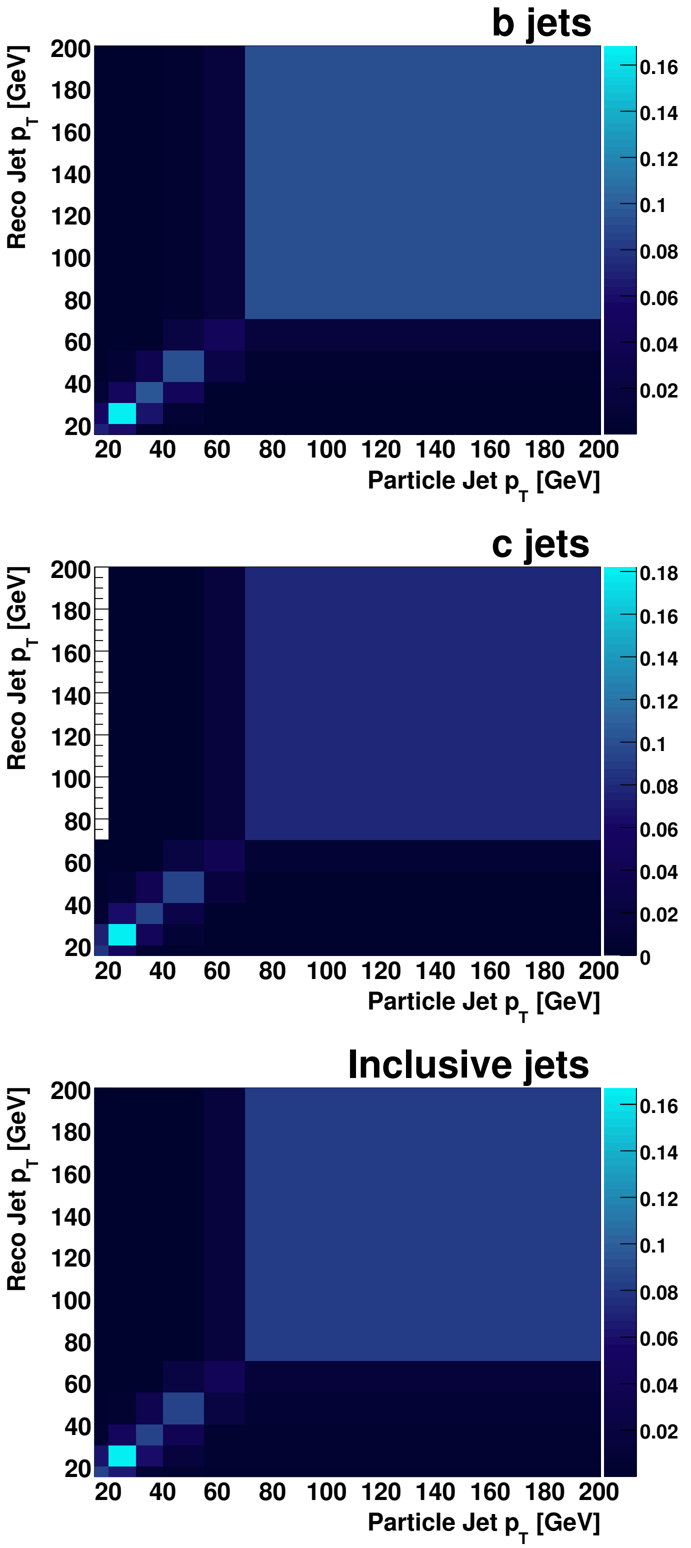


Figure 6.4: Migration matrices as a function of jet p_T for b jets (top), c jets (center), and inclusive jets (bottom). All migration matrices display similar off diagonal bin migrations.

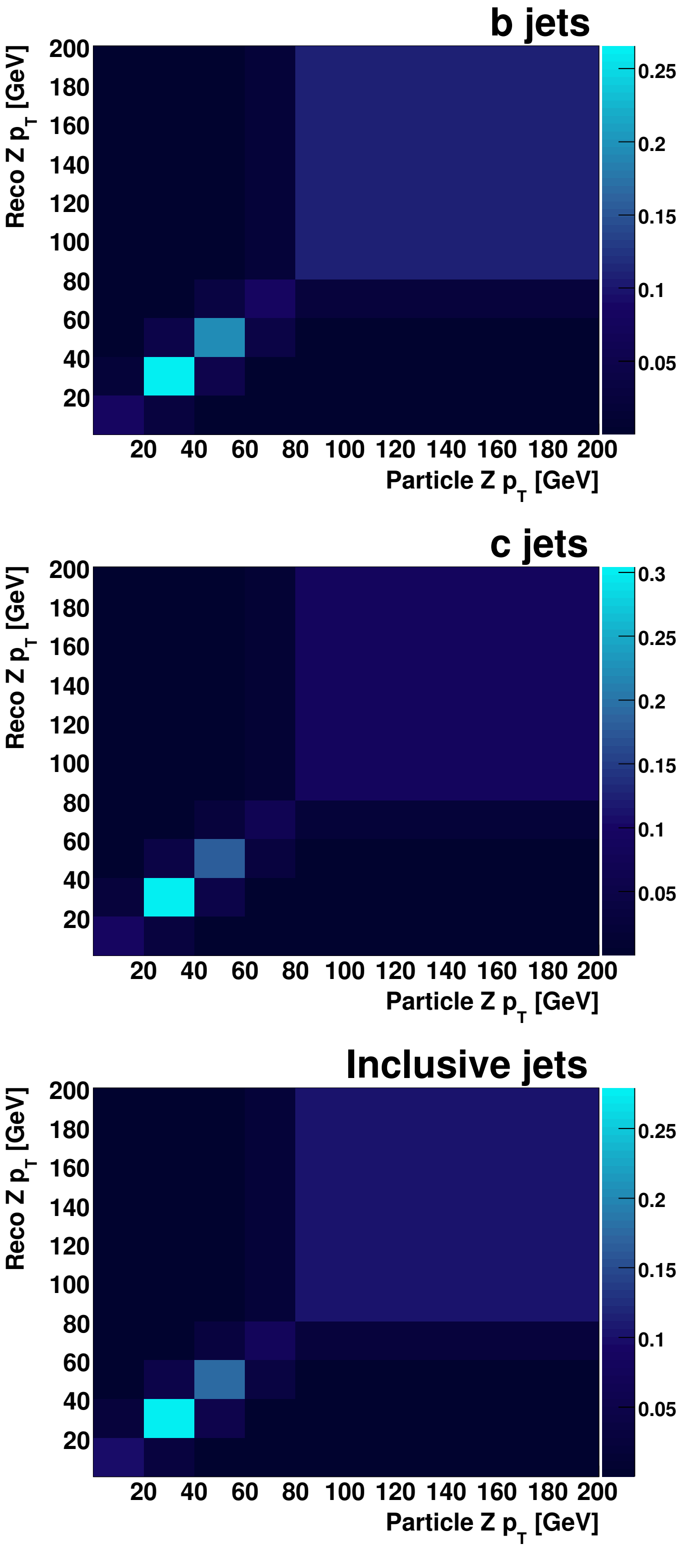


Figure 6.5: Migration matrices as a function of Z boson p_T for b jets (top), c jets (center), and inclusive jets (bottom). All migration matrices display similar off diagonal bin migrations.

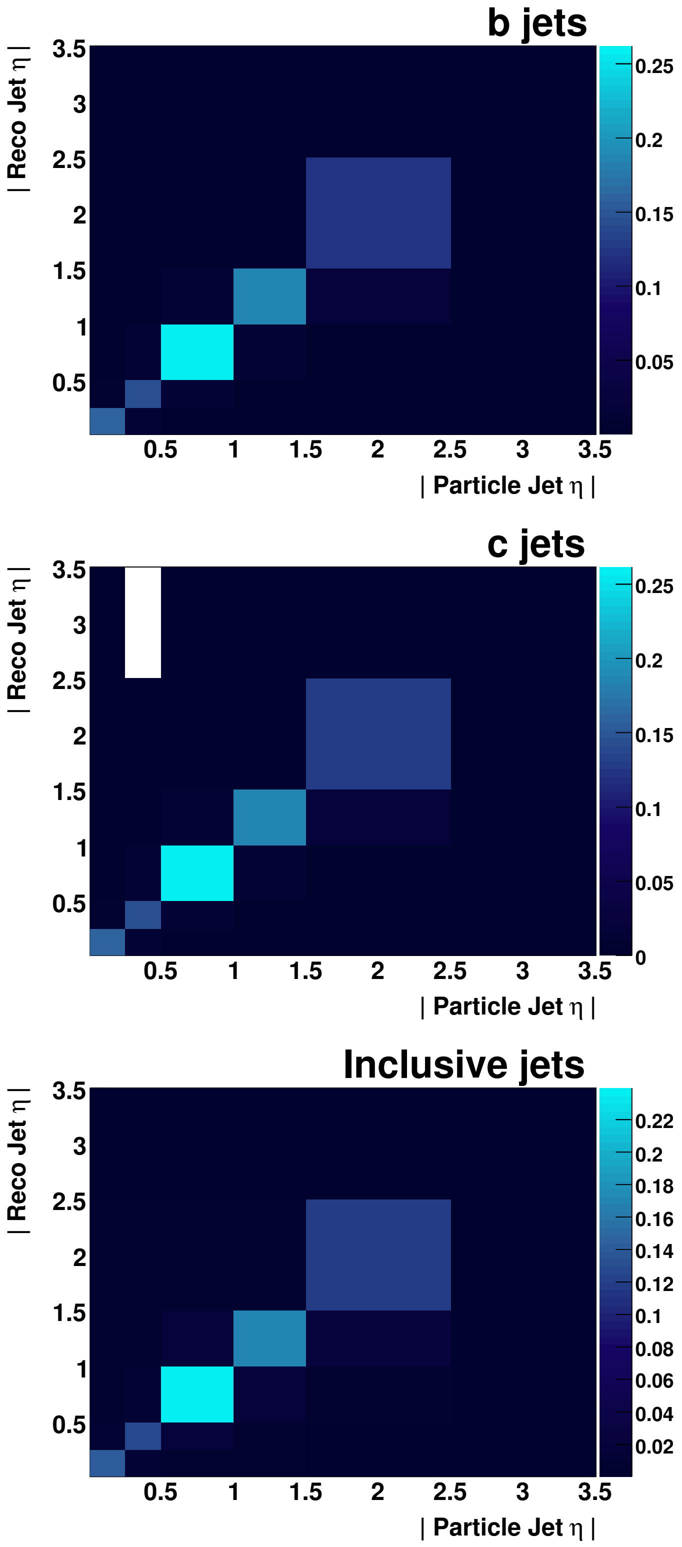


Figure 6.6: Migration matrices as a function of jet η for b jets (top), c jets (center), and inclusive jets (bottom). All migration matrices display similar off diagonal bin migrations.

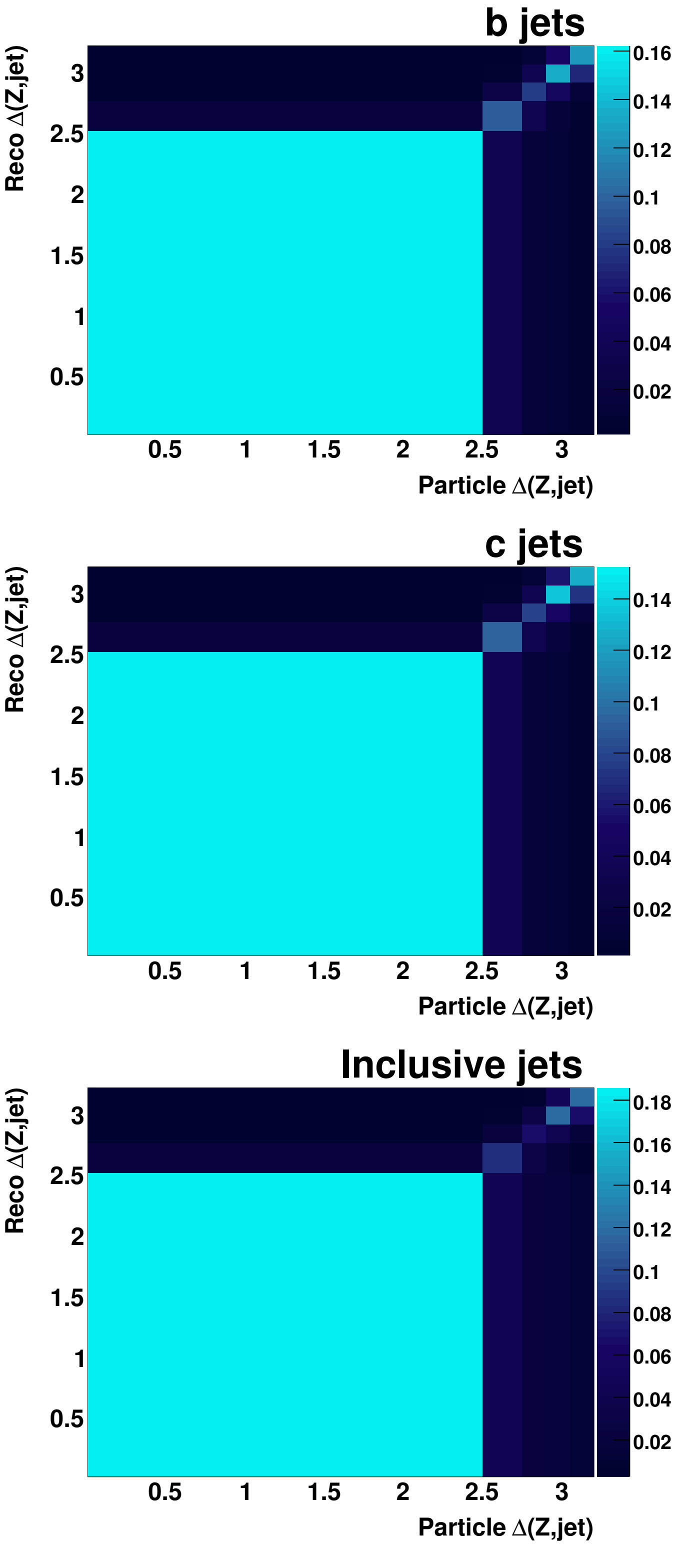


Figure 6.7: Migration matrices as a function of $\Delta\varphi(Z,\text{jet})$ for b jets (top), c jets (center), and inclusive jets (bottom). All migration matrices display similar off diagonal bin migrations.

7 Z boson plus b jet cross sections

Once we have created two samples, the full set of Z+jet events and the $Z + \mathrm{HF}$ jet enriched sample, and have simulations which describe these samples, we can move forward with the cross section measurements. In both analyses we will be measuring the ratio of cross sections. In addition, the differences due to the acceptance for the two lepton channels will cancel, which allows for the two channels to be combined, reducing the overall statistical uncertainty of our measurements.

Starting with the 176,498 events in the $Z +$ jet sample we want to extract the number of $Z + b$ jet events relative to this total number of $Z +$ jet events. Since we are only concerned with the b jet content for this analysis we can afford to have a larger contamination due to light jets. If we refer back to Figure 6.2 we see that it is the b jet tail, the region $D_{\mathrm{MJL}} > 0.3$, that provides the large discriminating power for differentiating b jets from light jets. For this reason we will use the oldLoose OP, which provides a high signal efficiency. After applying this OP we retain a total of 8,142 events in the $Z + \mathrm{HF}^{OL}$ sample, where OL refers to the oldLoose OP.

7.1 Background estimation

Before we use the D_{MJL} templates to fit the data we must first determine the level of background events that have contaminated the sample. Using the simulated event samples and applying the correction factors derived in Chapter 5 we can estimate the contributions from

Table 7.1: Estimated background contributions that have contaminated the $Z+\mathrm{HF}^{OL}$ data sample for the $\mu\mu$ and ee combined sample.

Sample	Number of events
Data	8142
Multijet	345
ZZ	77
WZ	44
WW	2
$t\bar{t}$	73

the various processes which have passed our selection criteria. These are listed in Table 7.1 and are subtracted from the data D_{MJL} distribution.

7.2 Flavor fraction measurement

After we have removed the background contaminations listed in Table 7.1 we are left with a total of 7,497 events in the background-subtracted $Z + \mathrm{HF}^{OL}$ data D_{MJL} distribution. This final D_{MJL} distribution is composed of a mixture of Z + light, c, and b jets, and we can fit this distribution using the templates the we have derived in Figure 6.2(a). Using a maximum likelihood fitter, ROOT's `TFractionFitter` [77], these templates are then simultaneously floated. The final normalization relative to the total sample will yield the fractional contribution for each flavor. This procedure was validated by using a MC closure test, discussed in Section 7.8.1, where we extracted from the MC the predicted flavor fractions and no systematic bias was found. An additional test was performed where we injected a random fraction of b jets, f_b, into a MC sample and then extracted it using `TFractionFitter`. We found that the fraction was extracted within the uncertainty on the fraction, the results of this test can be seen in Figure 7.1.

In Table 7.2 the flavor composition of the electron and muon channels is shown separately. Before we combine the two samples we want to verify that the b jet content, f_b, in the two

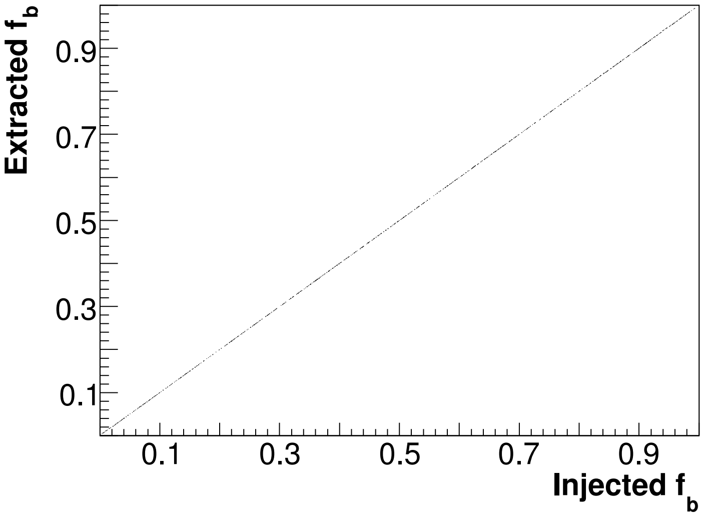


Figure 7.1: The extracted fraction of b jet events as measured as a function of the injected fraction of events. This demonstrates that no bias exists in the fitting procedure.

Table 7.2: Table of the extracted jet flavor fractions from the Z + HF^{OL} sample in the dimuon, dielectron, and combined channel.

Region	$\|\eta\| < 2.5$	$MVA_{bl} > 0.1$	$p_T^{jet} > 20$ GeV
Jet Flavor	Fraction $\mu\mu$	Fraction ee	Fraction $\mu\mu + ee$
Events	3921	3576	7497
$Z + b$ jet, f_b	0.215 ± 0.016	0.198 ± 0.019	0.208 ± 0.012
$Z + c$ jet, f_c	0.227 ± 0.034	0.343 ± 0.042	0.278 ± 0.025
Z + light jet, f_l	0.558 ± 0.026	0.459 ± 0.031	0.514 ± 0.019

samples is consistent. From Table 7.2 we can see that f_b as measured in the separate channels is consistent within uncertainties, allowing for the combination of the samples to increase the overall precision of the flavor extraction. Due to the similarity in the shapes of the light and c jet template the $Z + c$ and Z + light jet fractions have large relative uncertainties and their fractions are less trustworthy. Chapter 8 describes how to reliably extract the c jet fraction.

Since the background-subtracted data distribution is expected to contain only Z + light, $Z+c$, and $Z+b$ jet contributions we can weight the Z+jet templates by their extracted flavor fractions and see how accurately they reproduce the data. We take the templates and set their normalization to what was extracted in Table 7.2 and sum the relative contributions. This along with the background-subtracted data D_{MJL} distribution from the $Z + \mathrm{HF}^{OL}$ sample is shown in Figure 7.2. The uncertainties on the data points come solely from the data statistics from the background-subtracted data sample, additional systematic uncertainties associated with the template shapes and finite template statistics are discussed in Section 7.5.

7.3 Measurement of the integrated ratio of cross sections

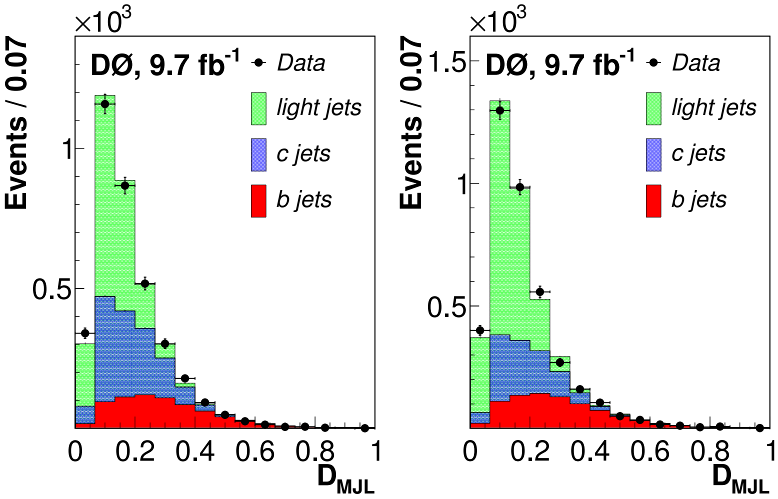
To study the $Z + b$ jet production we will be measuring the ratio of this cross section, $\sigma(Z + b\text{ jet})$, relative to that of the $Z +$ inclusive jet cross section, $\sigma(Z + \text{jet})$. The ratio of the cross sections is calculated using:

$$\frac{\sigma\left(Z+b\text{ jet}\right)}{\sigma\left(Z+\text{jet}\right)} = \frac{N^{OL}\, f_b}{N_{Z+jet}\varepsilon^b_{btag}\varepsilon^b_{tagg}} \times \frac{\mathcal{A}_{incl}}{\mathcal{A}_b} \tag{7.1}$$

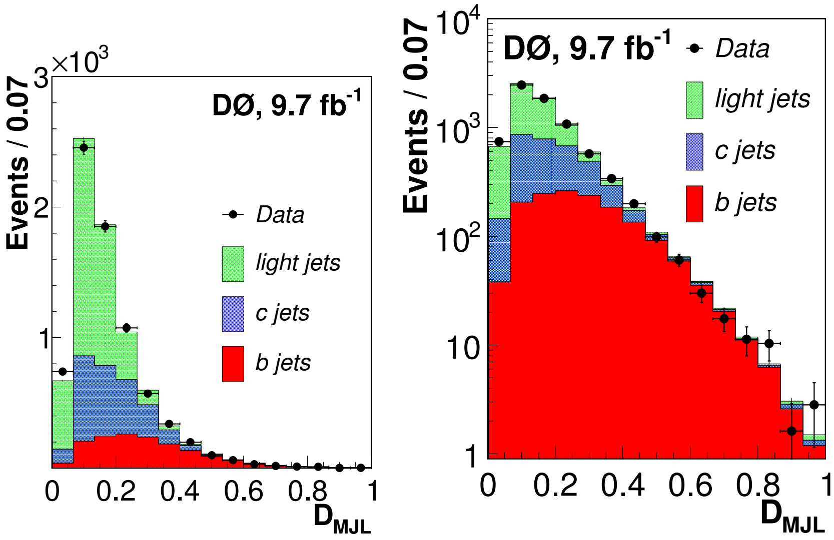
where N^{OL} is the number of events in the background-subtracted $Z + \mathrm{HF}^{OL}$ sample, N_{Z+jet} is the number of events in the background-subtracted $Z + \text{jet}$ sample (shown in Table 4.2), f_b is the extracted b jet flavor fraction, ε^b_{btag} refers to the efficiency of identifying a b jet with the MVA$_{bl}$ algorithm, ε^b_{tagg} is the efficiency of a b jet being taggable, and finally $\mathcal{A}_{incl}/\mathcal{A}_b$ is the ratio of acceptances, as discussed in Section 6.3. The variables in Equation 7.1 are measured and are reported in Table 7.3 and yield:

Table 7.3: Table of efficiencies and event totals used in the measurement of the integrated ratio of cross sections.

Total efficiency and event yields	
ϵ^b_{btag}	0.585 ± 0.009
ϵ^b_{tagg}	0.902 ± 0.013
N_{Z+jet}	167389 ± 409
N^{OL}	7497 ± 87



(a) Electron channel **(b)** Muon channel



(c) Combined electron and muon channel, linear on the left, log on the right.

Figure 7.2: Distribution of observed events for the D_{MJL} discriminant with $\mathrm{MVA}_{bl} > 0.1$. The distributions for the b, c, and light jet templates are shown normalized to their fitted fractions. The uncertainties on the data points are the statistical uncertainties.

$$\frac{\sigma\,(Z + b\text{ jet})}{\sigma\,(Z + \text{jet})} = 0.0196 \pm 0.0012\,(\text{stat})$$

which is in agreement with the previous D0 measurement of $\sigma(Z{+}b\text{ jet})/\sigma(Z{+}\text{jet}) = 0.0193 \pm 0.0022\,(\text{stat}) \pm 0.0015\,(\text{syst})$, measured using 4.2 fb^{-1} [13], and with previous measurements from the Tevatron [10, 11, 12].

7.4 Measurement of the ratios of differential cross sections

This procedure can be repeated in bins of a variety of kinematic variables. For this analysis we have studied the jet p_T, Z boson p_T, jet η, and $\Delta\varphi\,(Z, \text{jet})$, subdivided into five bins each. The bins were selected so that a reliable fit can be achieved in each bin, but also to accurately probe specific features in the kinematic distributions. Every variable, template, and event total which enters into Equation 7.1 was measure again inside every bin, such that each bin is, essentially, independent of its neighboring bins. The acceptance correction acts as a bin-by-bin unfolding procedure, discussed in Section 6.3.

7.4.1 Jet transverse momentum

We start by taking our data and divide it into five bins, the width of which are chosen to allow for us to be able to obtain a stable D_{MJL} fit in each bin. To attain this stability over the desired p_T range the following bins were selected: $20-30$, $30-40$, $40-55$, $55-70$, and $70-200$ GeV. Table 7.4 shows the event totals, measured b jet flavor fraction, and p_T-weighted b jet identification and taggability efficiencies. Figure 7.3 to Figure 7.7 shows the background-subtracted data D_{MJL} distributions being fit with the templates for each bin of jet p_T. All systematic uncertainties quoted in Table 7.4 to Table 7.7 are described in Section 7.5.

The final results of these fits and a comparison of $\sigma(Z{+}b\text{ jet})/\sigma(Z{+}\text{jet})$ to the predictions from NLO calculations and various MC event generators can be found in Section 7.7.

Table 7.4: Table of bin-by-bin information for the measurement of the ratio of differential cross sections as a function of jet p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Jet p_T [GeV]	N^{OL}	N_{Z+jet}	f_b [%]	ϵ^b_{btag}[%]	ϵ^b_{tagg}[%]	$\frac{\sigma(Z+b\text{ jet})}{\sigma(Z+\text{jet})}$ [%]
$20-30$ (25)	2920	80586	20.1 ± 0.20	53.9 ± 0.5	86.6 ± 0.7	$1.72 \pm 0.14 \pm 0.19$
$30-40$ (35)	1669	35508	23.6 ± 0.26	60.5 ± 0.8	93 ± 1	$2.10 \pm 0.20 \pm 0.16$
$40-55$ (47)	1358	25526	22.5 ± 0.26	63 ± 1	94 ± 2	$2.19 \pm 0.22 \pm 0.13$
$55-70$ (62)	616	11910	25.0 ± 0.43	65 ± 2	94 ± 3	$2.36 \pm 0.35 \pm 0.26$
$70-200$ (102)	915	13693	17.6 ± 0.36	65 ± 3	94 ± 5	$2.26 \pm 0.42 \pm 0.22$

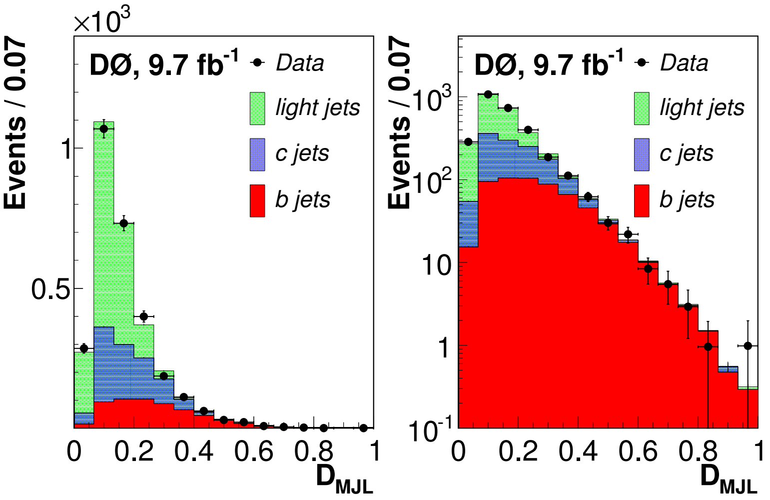


Figure 7.3: MC D_{MJL} templates normalized to their extracted flavor fraction and the background-subtracted data distribution for $20 <$ jet $p_T < 30$ GeV.

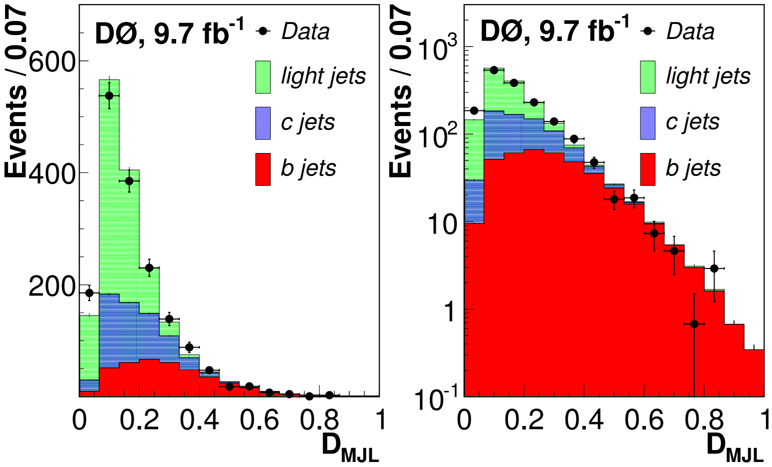


Figure 7.4: MC D_{MJL} templates normalized to their extracted flavor fraction and the background-subtracted data distribution for $30 <$ jet $p_T < 40$ GeV.

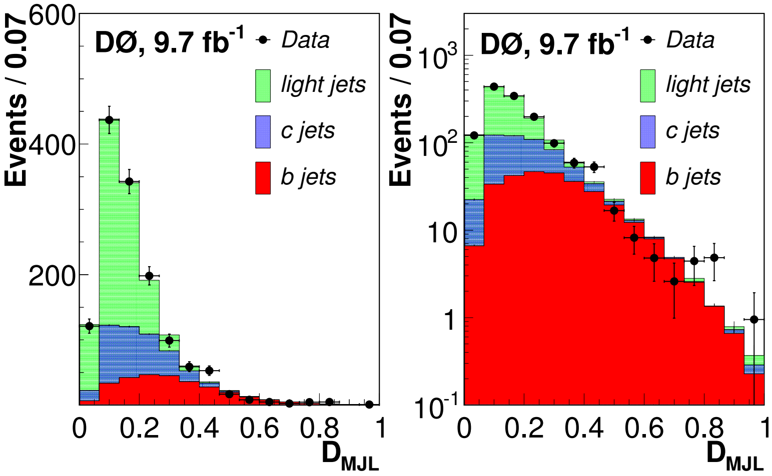


Figure 7.5: MC D_{MJL} templates normalized to their extracted flavor fraction and the background-subtracted data distribution for $40 <$ jet $p_T < 55$ GeV.

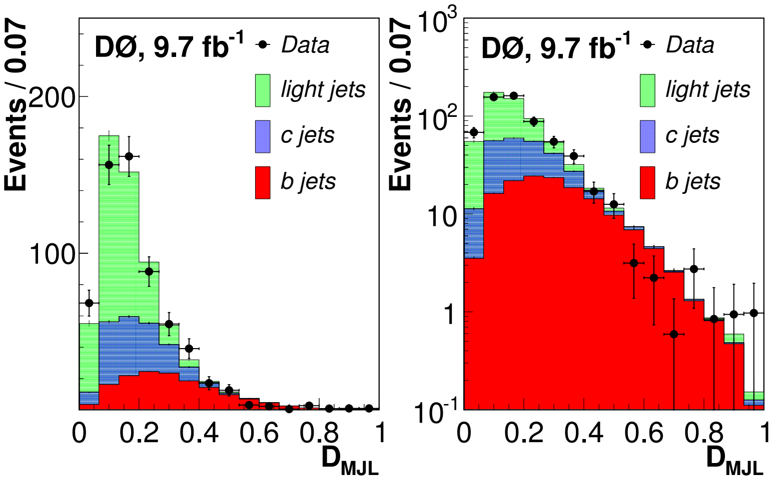


Figure 7.6: MC D_{MJL} templates normalized to their extracted flavor fraction and the background-subtracted data distribution for $55 <$ jet $p_T < 70$ GeV.

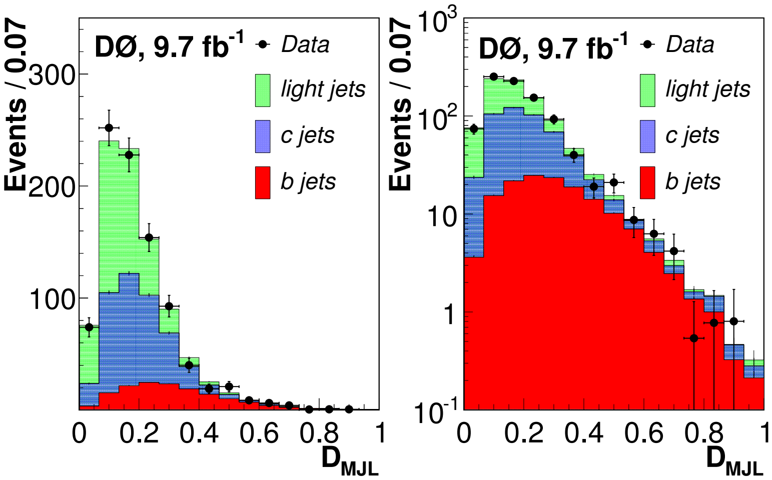


Figure 7.7: MC D_{MJL} templates normalized to their extracted flavor fraction and the background-subtracted data distribution for $70 <$ jet $p_T < 200$ GeV.

Table 7.5: Table of bin-by-bin information for the measurement of the ratio of differential cross sections as a function of Z boson p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Zp_T [GeV]	N^{OL}	N_{Z+jet}	f_b [%]	ϵ^b_{btag}[%]	ϵ^b_{tagg}[%]	$\frac{\sigma(Z+b\text{ jet})}{\sigma(Z+\text{jet})}$ [%]
$0-20$ (12)	1066	35190	29.0 ± 0.37	54.4 ± 0.8	85 ± 1	$2.68 \pm 0.28 \pm 0.37$
$20-40$ (32)	2818	72254	17.7 ± 0.19	57.0 ± 0.8	89 ± 1	$1.19 \pm 0.10 \pm 0.10$
$40-60$ (50)	1925	35984	22.2 ± 0.24	60.9 ± 0.6	92 ± 1	$2.12 \pm 0.19 \pm 0.13$
$60-80$ (68)	887	13900	19.3 ± 0.34	62 ± 1	93 ± 2	$2.18 \pm 0.31 \pm 0.13$
$80-200$ (100)	789	9919	20.2 ± 0.37	62 ± 3	94 ± 5	$3.04 \pm 0.50 \pm 0.19$

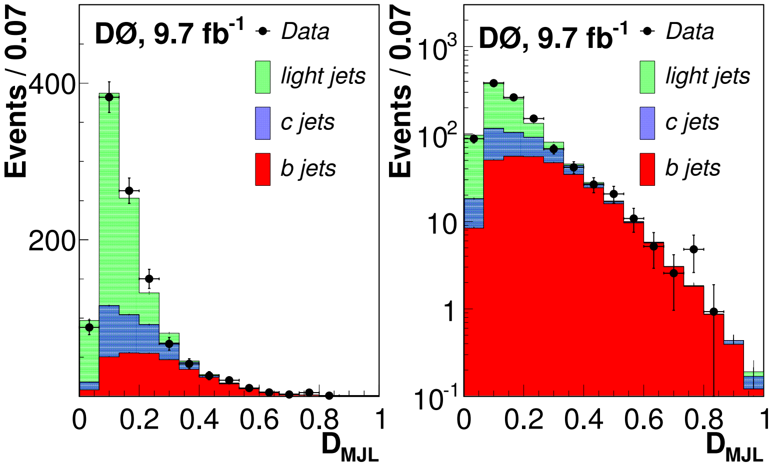


Figure 7.8: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $0 < Z\ p_T < 20$ GeV.

7.4.2 Z boson transverse momentum

Following a similar prescription discussed in Section 7.4.1 we can start by dividing our sample into bins of Z boson p_T: $0-20$, $20-40$, $40-60$, $60-80$, and $80-200$ GeV, where the inputs for Equation 7.1 can be found in Table 7.5. Figure 7.8 to Figure 7.12 shows the template fits for the five bins of Z boson p_T that have been selected. The final $\sigma(Z + b\text{ jet})/\sigma(Z + \text{jet})$ distribution as a function of Z boson p_T is shown in Section 7.7.

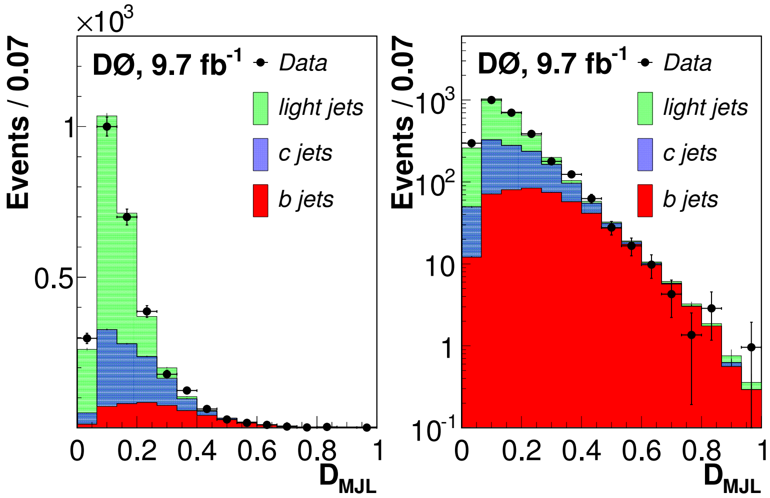


Figure 7.9: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $20 < Z\ p_T < 40$ GeV.

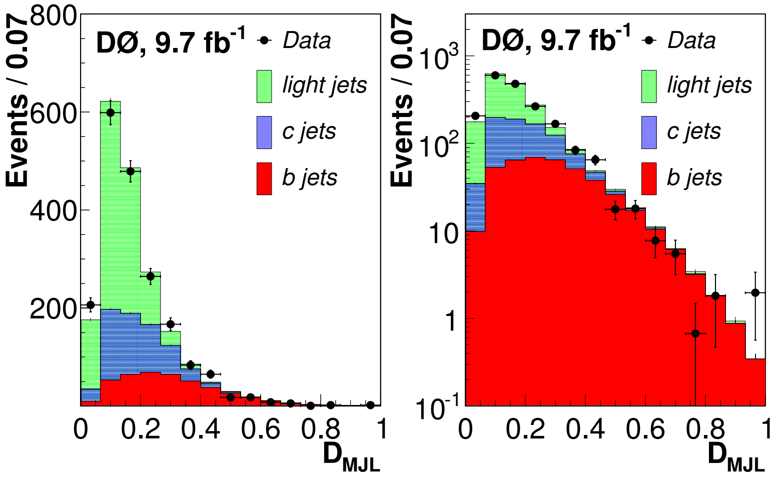


Figure 7.10: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $40 < Z\ p_T < 60$ GeV.

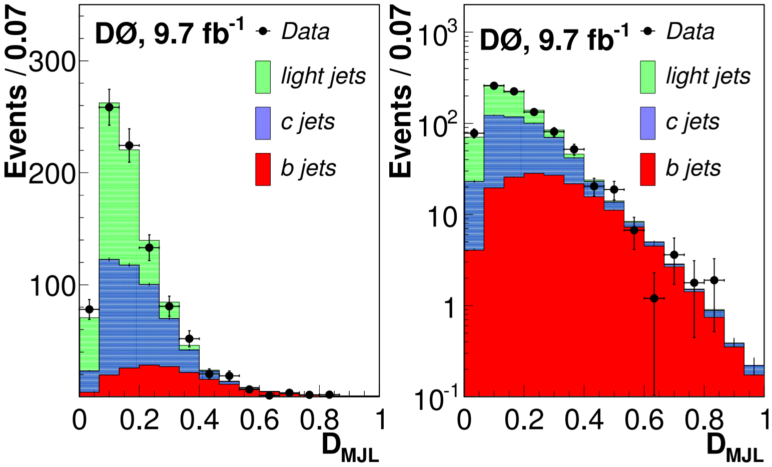


Figure 7.11: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $60 < Z\ p_T < 80$ GeV.

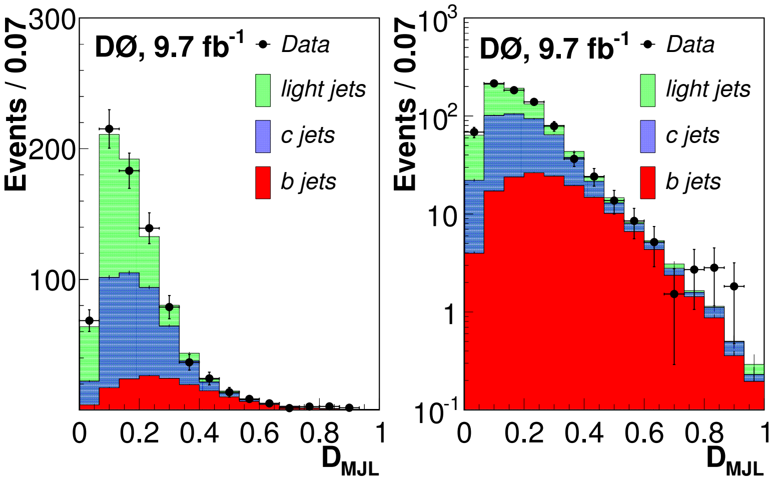


Figure 7.12: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $80 < Z\ p_T < 200$ GeV.

Table 7.6: Table of bin-by-bin information for the measurement of the ratio of differential cross sections as a function of jet η. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Jet η	N^{OL}	N_{Z+jet}	f_b [%]	ϵ^{b}_{btag}[%]	ϵ^{b}_{tagg}[%]	$\frac{\sigma(Z+b\text{ jet})}{\sigma(Z+\text{jet})}$ [%]
$0-0.25$ (0.13)	1203	24993	15.3 ± 0.27	61 ± 1	90 ± 1	$1.39 \pm 0.18 \pm 0.10$
$0.25-0.5$ (0.38)	1207	24426	18.5 ± 0.22	62 ± 1	91 ± 2	$1.72 \pm 0.17 \pm 0.11$
$0.5-1.0$ (0.75)	2217	46619	24.1 ± 0.22	61 ± 1	93 ± 2	$2.13 \pm 0.17 \pm 0.17$
$1.0-1.5$ (1.25)	1695	39813	24.0 ± 0.27	59 ± 1	93 ± 2	$2.02 \pm 0.20 \pm 0.22$
$1.5-2.5$ (2.00)	1174	31535	18.8 ± 0.41	49 ± 1	85 ± 3	$1.61 \pm 0.30 \pm 0.23$

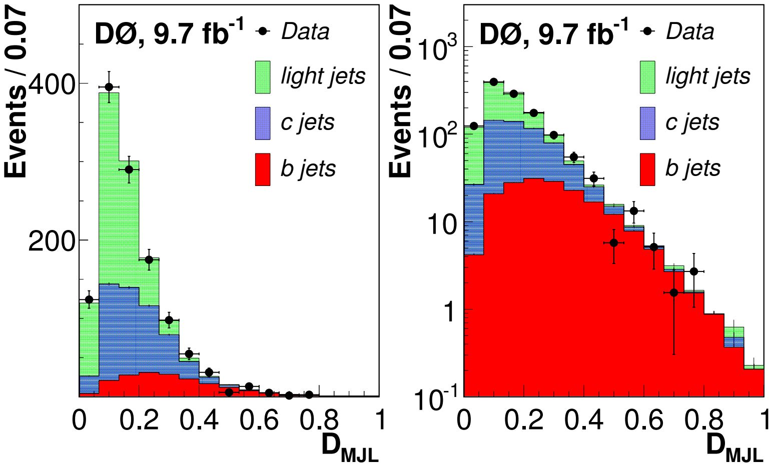


Figure 7.13: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $0.0 < \eta < 0.25$.

7.4.3 Jet pseudorapidity

Again following what has been done in the previous sections we divide our sample into bins of jet η based on the remaining event totals: $0.0-0.25$, $0.25-0.5$, $0.5-1.0$, $1.0-1.5$, and $1.5-2.5$, where the inputs for Equation 7.1 can be found in Table 7.6. Figure 7.13 to Figure 7.17 shows the template fits for the five bins of jet η that have been selected. The final $\sigma(Z+b\text{ jet})/\sigma(Z+\text{jet})$ distribution as a function of jet η is shown in Section 7.7.

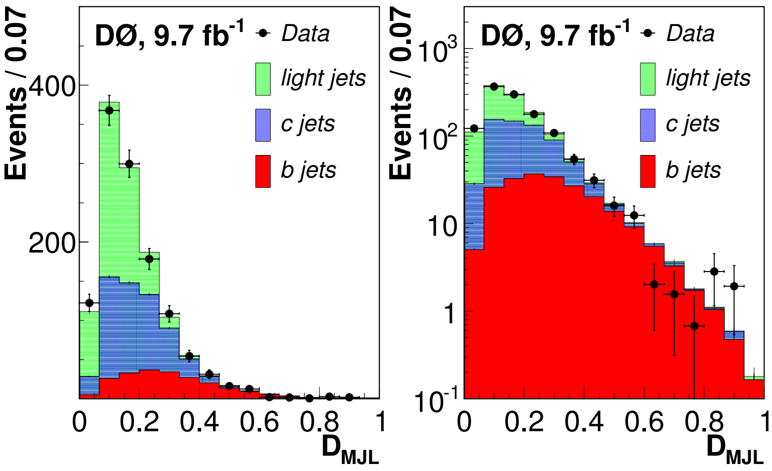


Figure 7.14: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $0.25 < \eta < 0.5$.

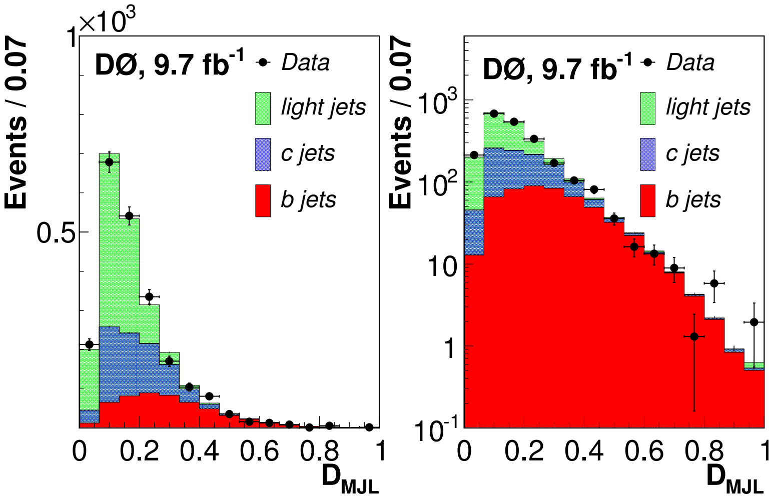


Figure 7.15: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $0.5 < \eta < 1.0$.

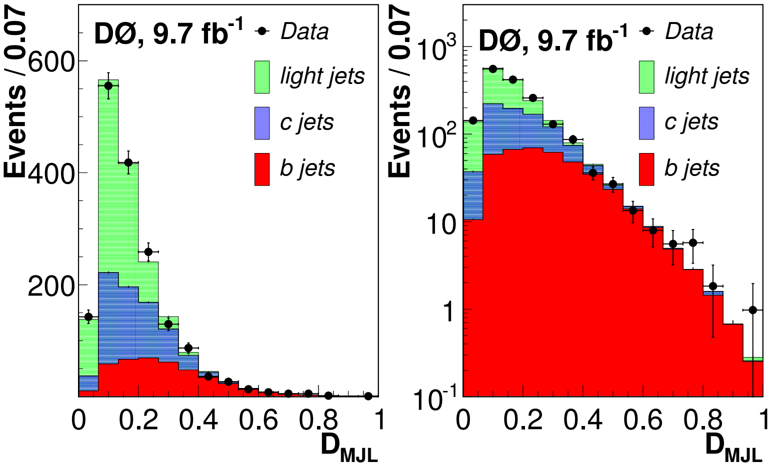


Figure 7.16: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $1.0 < \eta < 1.5$.

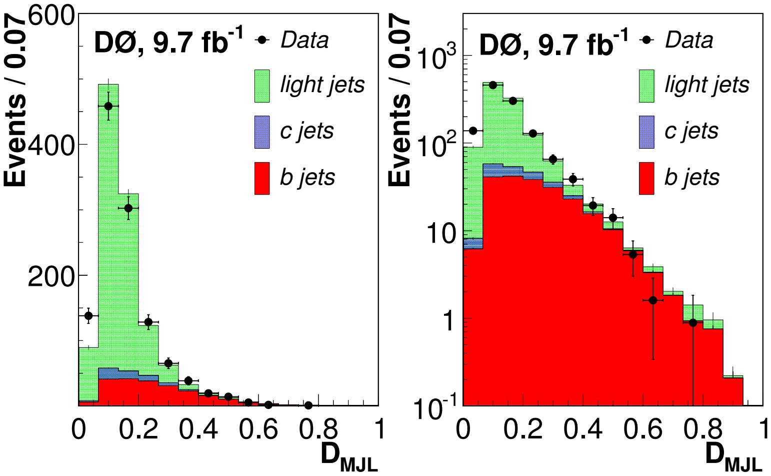


Figure 7.17: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $1.5 < \eta < 2.5$.

Table 7.7: Table of bin-by-bin information for the measurement of the ratio of differential cross sections as a function of $\Delta\varphi\,(Z,\mathrm{jet})$. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

$\Delta\varphi\,(Z,\mathrm{jet})$	N^{OL}	$N_{Z+\mathrm{jet}}$	f_b [%]	ϵ^b_{btag}[%]	ϵ^b_{tagg}[%]	$\frac{\sigma(Z+b\ \mathrm{jet})}{\sigma(Z+\mathrm{jet})}$ [%]
$0-2.5$ (1.62)	1612	29552	23.4 ± 0.27	57 ± 2	90 ± 3	$3.39 \pm 0.37 \pm 0.30$
$2.5-2.75$ (2.63)	957	23224	21.7 ± 0.35	59 ± 1	90 ± 2	$2.00 \pm 0.27 \pm 0.19$
$2.75-2.9$ (2.83)	1155	26463	21.8 ± 0.31	60 ± 1	91 ± 2	$2.10 \pm 0.25 \pm 0.17$
$2.9-3.05$ (2.98)	1937	46546	18.9 ± 0.23	60.9 ± 0.9	92 ± 1	$1.52 \pm 0.15 \pm 0.11$
$3.05-3.2$ (3.13)	1834	41601	18.4 ± 0.24	62.0 ± 0.7	92.0 ± 0.9	$1.29 \pm 0.14 \pm 0.08$

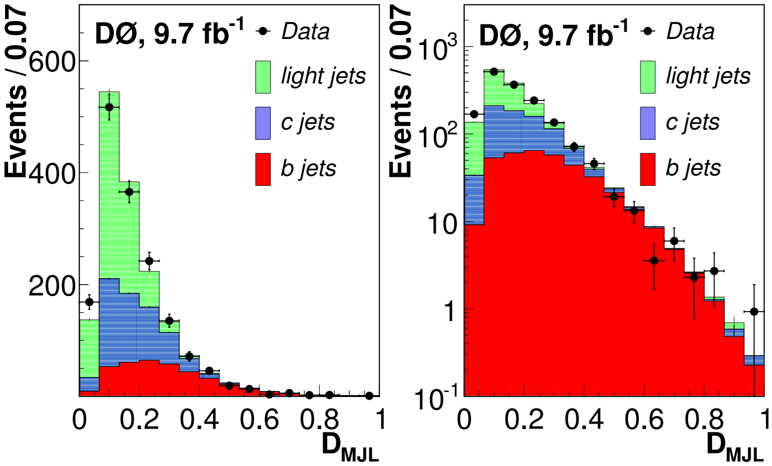


Figure 7.18: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $0 < \Delta\varphi\,(Z,\mathrm{jet}) < 2.5$.

7.4.4 Azimuthal distance between Z boson and jet

Finally, we divide our samples into bins of $\Delta\varphi\,(Z,\mathrm{jet})$: $0-2.5$, $2.5-2.75$, $2.75-2.9$, $2.9-3.05$, and $3.05-3.2$, where the input for Equation 7.1 can be found in Table 7.7. Figure 7.18 to Figure 7.22 shows the template fits for the five bins of $\Delta\varphi\,(Z,\mathrm{jet})$ that have been selected. The final $\sigma(Z+b\ \mathrm{jet})/\sigma(Z+\mathrm{jet})$ distribution as a function of $\Delta\varphi\,(Z,\mathrm{jet})$ is shown in Section 7.7.

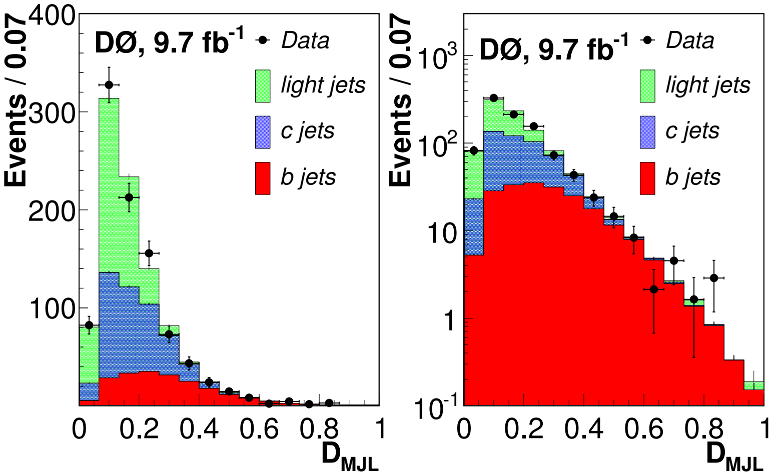


Figure 7.19: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $2.5 < \Delta\varphi\,(Z, \mathrm{jet}) < 2.75$.

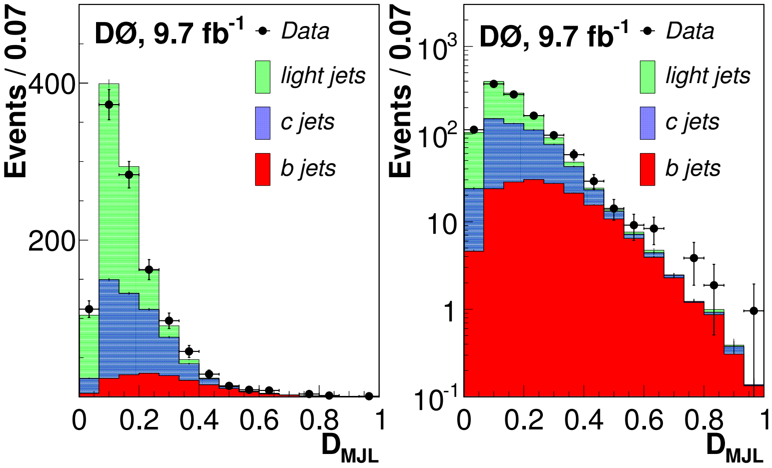


Figure 7.20: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $2.75 < \Delta\varphi\,(Z, \mathrm{jet}) < 2.9$.

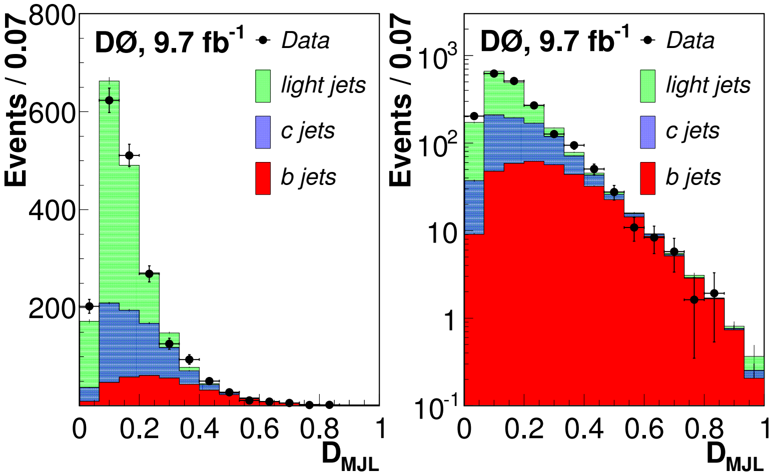


Figure 7.21: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $2.9 < \Delta\varphi\,(Z,\mathrm{jet}) < 3.05$.

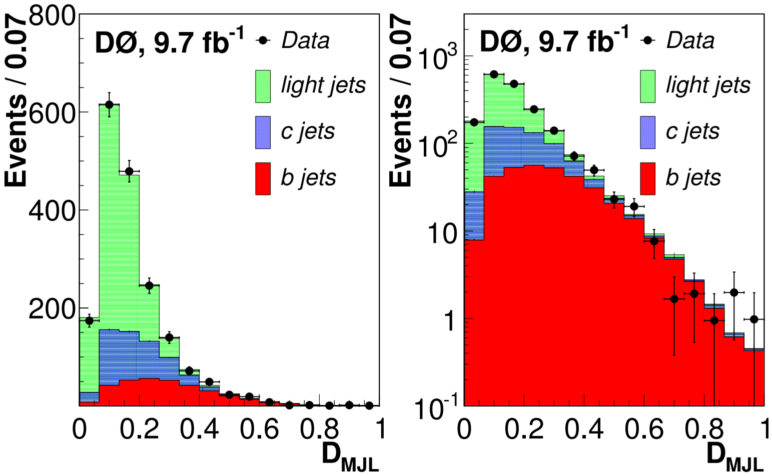


Figure 7.22: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $3.05 < \Delta\varphi\,(Z,\mathrm{jet}) < 3.2$.

7.5 Systematic uncertainties

Measurements of the ratio of cross sections do not benefit from a complete cancelation of the systematic uncertainties. In this section we will discuss and quantify the major systematic uncertainties for the measurement of the $\sigma(Z + b \text{ jet})/\sigma(Z + \text{jet})$. The contributions of the individual systematic uncertainties along with their quadratic sum for each differential bin can be found in Table 7.8 to Table 7.12. Each item in the list below is independently varied and the resulting change to the final result is taken as a systematic uncertainty:

- Heavy flavored jet identification:

  In order to estimate the effect of the uncertainties derived in Section 5.4 we vary the correction factor by $\pm 1\sigma$ and the whole analysis chain is repeated.

- Jet energy resolution (JER):

  The standard resolution of the jet energy is varied $\pm 1\sigma$ and the full analysis chain is repeated.

- Jet energy scale (JES):

  The JES correction is varied within its uncertainties, as determined in Ref. [52], and the full analysis chain is repeated.

- Detector acceptance model dependence:

  To verify that the results do not depend on the choice of MC we compare the results from the reweighted ALPGEN $Z + \text{jet}$ MC to a sample of PYTHIA $Z + \text{jet}$ MC events. The difference between the acceptance values is taken as the uncertainty.

- Subtracted background contributions:

  The size of the background contributions which are subtracted from the data are estimated from MC generators and corrected to higher order calculations. The systematic uncertainty is estimated by conservatively varying the total number of background events which we subtract before fitting up and down by 10%, which is the uncertainty on the $t\bar{t}$ cross sections [78] and corresponds to the largest cross section uncertainty.

Template Shape Uncertainties:

- Data driven template shape:

  To estimate the dependence of the result on the light jet template shape we compare the shapes found in data and MC. To do this we replace the $Z +$ light jet MC D_{MJL} templates with new templates built using a light jet enriched dijet data sample. The sample is created by "anti-tagging" the jet which recoiled away from the jet which we select. This is done by requiring that the *away* jet did not contain a muon, and that it failed an L6 MVA_{bl} requirement. This allows us to the apply the MVA_{bl} selection to the selected jet and maintain a large light jet content. The D_{MJL} distribution from this sample was used in place of the MC light jet templates and the fit was repeated. The relative difference between the b jet fraction was taken as a systematic uncertainty. Since there is no Z boson in the event we cannot parameterize this effect, instead this is taken as a flat systematic across all the differential measurements.

- b and c jet template shape:

  The b and c jet template shapes were checked against a heavy flavor enriched data sample. This data sample was constructed from a back-to-back dijet sample where the "away" jet had to be matched to a muon with a $p_T^{rel} > 0.6$ GeV and pass a MegaTight MVA_{bl} requirement. The selected jet was then required to pass the OldLoose OP requirement. This data sample contained contamination form c jets, so the shape of the resulting data distribution was compared to a combination of MC b and c jet templates. The difference in their shape is found to be negligible and the resulting data driven templates resulted in no significant change to the extracted flavor fractions.

- Template fitting:

  The statistics of the data D_{MJL} distributions are artificially inflated to be as large as possible while still achieving a convergent fit. This preserves the original shape, but minimizes the corresponding statistical uncertainty from the data in the fit. The remaining uncertainty on the flavor fraction is taken as the uncertainty associated with the fitting procedure and the finite MC template statistics. This uncertainty also takes into account the similarity of the template shapes, since the templates are not perfectly discriminating. If the more similar the template shapes are the larger this uncertainty becomes.

- *b*-quark fragmentation model:

  By default, the events have been reweighted from the default PYTHIA *b*-quark fragmentation to a Bowler scheme that has been tuned to LEP data [79]. To evaluate the uncertainty associated with this choice of fragmentation, the events are further reweighted to account for the difference between the SLD and LEP [79]. The analysis is repeated with the new fragmentation scheme and the difference between the two schemes is taken as a systematic uncertainty.

- *c* jet fragmentation:

  It is possible that the *c* jet templates could depend on the relative ratio of various D mesons in the showers. It was observed that in simulations that templates built from D^+ and D^0 decays have a slightly different shape. To test what effect this has on the final result we vary the ratio of D^+/D^0 in our simulations and repeat the analysis. This was found to have an effect of $< 0.4\%$ on the measured fractions and is considered negligible when compared to the other systematic uncertainties.

7.6 Predictions

The measurements are compared to predictions from MCFM NLO [6] calculations and two MC event generators, SHERPA and ALPGEN. The NLO predictions are based on MCFM [9] with the MSTW2008 PDFs [80], with the renormalization, μ_R, and factorization, μ_F, scales

Table 7.8: Systematic uncertainties for the integrated $\sigma\left(Z + b \text{ jet}\right) / \sigma\left(Z + \text{jet}\right)$ measurement.

Systematic Uncertainty		Percentage
JES		4.7%
b jet identification		1.5%
JER		4.7%
Detector acceptance		0.02%
Background estimation		0.4%
Template Shape	Template fitting	0.4%
	b-quark fragmentation	1.0%
	HF merging	0.6%
	Light jet template shape	4.3%
Total		7.1%

Table 7.9: Systematic uncertainties in bins of jet p_T for the measurement of the ratio $\sigma\left(Z + b \text{ jet}\right) / \sigma\left(Z + \text{jet}\right)$.

Systematic Uncertainty		Jet p_T [GeV]				
		$20 - 30$	$30 - 40$	$40 - 55$	$55 - 70$	$70 - 200$
JES		7.2%	3.9%	1.9%	7.1%	5.0%
b jet identification		1.4%	1.4%	1.6%	1.7%	1.7%
JER		4.7%	3.9%	1.9%	7.1%	5.0%
Detector acceptance		3.3%	3.5%	0.4%	3.2%	5.9%
Background estimation		2.1%	0.4%	1.2%	2.9%	2.8%
Template Shape	Template fitting	0.7%	0.9%	0.9%	1.6%	1.2%
	b-quark fragmentation	3.7%	1.4%	1.0%	1.5%	1.4%
	HF merging	1.5%	0.8%	0.7%	0.4%	0.3%
	Light jet template shape	4.3%				
Total		11.0%	7.5%	5.7%	11.0%	9.8%

Table 7.10: Systematic uncertainties in bins of Z boson p_T for the measurement of the ratio $\sigma\left(Z+b \text{ jet}\right)/\sigma\left(Z+\text{jet}\right)$.

Systematic Uncertainty		Z p_T [GeV]				
		0 – 20	20 – 40	40 – 60	60 – 80	80 – 200
JES		8.5%	3.3%	2.4%	0.7%	0.9%
b jet identification		1.4%	1.4%	1.5%	1.5%	1.6%
JER		8.5%	3.3%	2.4%	0.7%	0.9%
Detector acceptance		1.4%	4.8%	0.2%	1.8%	3.2%
Background estimation		0.3%	0.5%	0.4%	0.9%	0.8%
Template Shape	Template fitting	1.4%	0.7%	0.8%	1.3%	1.3%
	b-quark fragmentation	3.5%	2.0%	2.2%	2.5%	1.9%
	HF merging	1.6%	0.9%	1.0%	0.6%	0.7%
	Light jet template shape	4.3%				
Total		13.9%	8.0%	5.9%	5.9%	6.3%

Table 7.11: Systematic uncertainties in bins of jet η for the measurement of the ratio $\sigma\left(Z+b \text{ jet}\right)/\sigma\left(Z+\text{jet}\right)$.

Systematic Uncertainty		Jet η				
		0 – 0.25	0.25 – 0.5	0.5 – 1.0	1.0 – 1.5	1.5 – 2.5
JES		4.4%	2.6%	4.5%	6.1%	7.3%
b jet identification		1.3%	1.3%	1.4%	1.5%	2.5%
JER		4.4%	2.6%	4.5%	6.1%	7.3%
Detector acceptance		1.5%	2.2%	2.7%	5.1%	8.6%
Background estimation		1.1%	0.9%	0.3%	0.4%	0.5%
Template Shape	Template fitting	1.4%	0.9%	0.8%	1.0%	1.4%
	b-quark fragmentation	1.9%	1.8%	2.4%	3.3%	3.0%
	HF merging	0.8%	0.7%	0.9%	1.3%	1.0%
	Light jet template shape	4.3%				
Total		7.1%	6.1%	7.8%	10.9%	14.4%

Table 7.12: Systematic uncertainties in bins of $\Delta\varphi\,(Z, \text{jet})$ for the measurement of the ratio $\sigma\,(Z + b\ \text{jet})\,/\sigma\,(Z + \text{jet})$.

	Systematic Uncertainty	$\Delta\varphi\,(Z, \text{jet})$				
		$0 - 2.5$	$2.5 - 2.75$	$2.75 - 2.9$	$2.9 - 3.05$	$3.05 - 3.2$
	JES	4.5%	6.3%	4.9%	4.1%	3.2%
	b jet identification	1.5%	1.5%	1.5%	1.5%	1.5%
	JER	4.5%	6.3%	4.9%	4.1%	3.2%
	Detector acceptance	3.2%	2.5%	2.2%	1.9%	1.8%
	Background estimation	1.2%	0.8%	0.0%	0.5%	0.0%
Template Shape	Template fitting	0.9%	1.2%	1.1%	0.8%	0.8%
	b-quark fragmentation	3.1%	2.3%	1.8%	1.7%	1.6%
	HF merging	1.0%	0.8%	0.7%	0.9%	0.9%
	Light jet template shape	4.3%				
	Total	8.9%	9.6%	7.9%	7.1%	6.2%

set at $\mu_R^2 = \mu_F^2 = M_Z^2 + p_{T,\text{total}}^2$. Here, M_Z is the Z boson mass and $p_{T,\text{total}}$ is the scalar sum of the transverse momentum for all the jets, with $p_T^{\text{jet}} > 20$ GeV and $|\eta| < 2.5$. Additionally, MCFM offers a choice of jet algorithm used to define jets in the calculations [9]. We have chosen to use the Run II midpoint cone algorithm with the same radius ($\Delta R = 0.5$) which is used to define jets in the analyses (discussed in Section 3.1.4) to reduce any effects due to this choice.

Corrections are applied to account for non-perturbative effects, these are determined by taking the ratio of the parton level and particle level information from ALPGEN+PYTHIA simulations. These correspond to the hardonization and showering corrections along with differences due to initial and final state radiation, and the effects of multiple parton interactions. For the integrated result this is a 5% effect which increases the NLO MCFM predictions. Table 7.13 shows the non-perturbative corrections as a function of the variables which we have measured. The uncertainty on the theoretical predictions are evaluated by simultaneously changing the μ_R and μ_F scales up and down by a factor of two.

ALPGEN generates multi-parton final states using tree-level matrix elements (ME). When

Table 7.13: Non-perturbative corrections as estimated from ALPGEN+PYTHIA simulations and applied to the NLO predictions MCFM.

Non-perturbative Corrections to NLO Predictions	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5
Jet p_T	2%	5%	2%	1%	3%
Z Boson p_T	10%	2%	1%	1%	1%
Jet η	3%	3%	2%	2%	2%
$\Delta\varphi(Z, \text{jet})$	2%	2%	8%	7%	14%

interfaced with PYTHIA, it employs the MLM scheme [56] to match ME partons produced from showering in PYTHIA. This results in an overall improvement over leading-logarithmic accuracy. SHERPA uses the CKKW matching scheme [81] between the leading-order ME partons and the parton-shower jets. This effectively allows for a consistent combination of the ME and the parton shower.

7.7 Results

The ratio $\sigma\left(Z + b \text{ jet}\right)/\sigma\left(Z + \text{jet}\right)$ has been measured in $Z \to \ell\ell\ (\ell = e, \mu)$ final states using 9.7 fb^{-1} of Run II D0 data. We determined the ratio to be $0.0196 \pm 0.0012\,(\text{stat}) \pm 0.0013\,(\text{syst})$ with jets of $p_T > 20$ GeV in the pseudorapidity region of $|\eta| \leq 2.5$, which is in agreement with NLO calculations [82]. This measurement supersedes the earlier result obtained by D0 of $0.0193 \pm 0.0022 \pm 0.0015$, based on 4.2 fb^{-1} of Run II data [13]. We have also measured the differential ratio of cross sections as a function of jet and Z boson p_T, jet η, and $\Delta\varphi\left(Z, \text{jet}\right)$, for direct comparison with NLO theory and two MC event generators, ALPGEN and SHERPA. These are shown in Figure 7.23 as compared to these various predictions.

For the ratio of cross sections as a function of jet p_T we see that all three sets of predictions reproduce the dependence of the ratio of cross sections. The ratio of cross sections' dependence on Z boson p_T is a bit more complicated. Looking at the the full distribution, MCFM does the best job of reproducing the scale of the distribution and has features similar

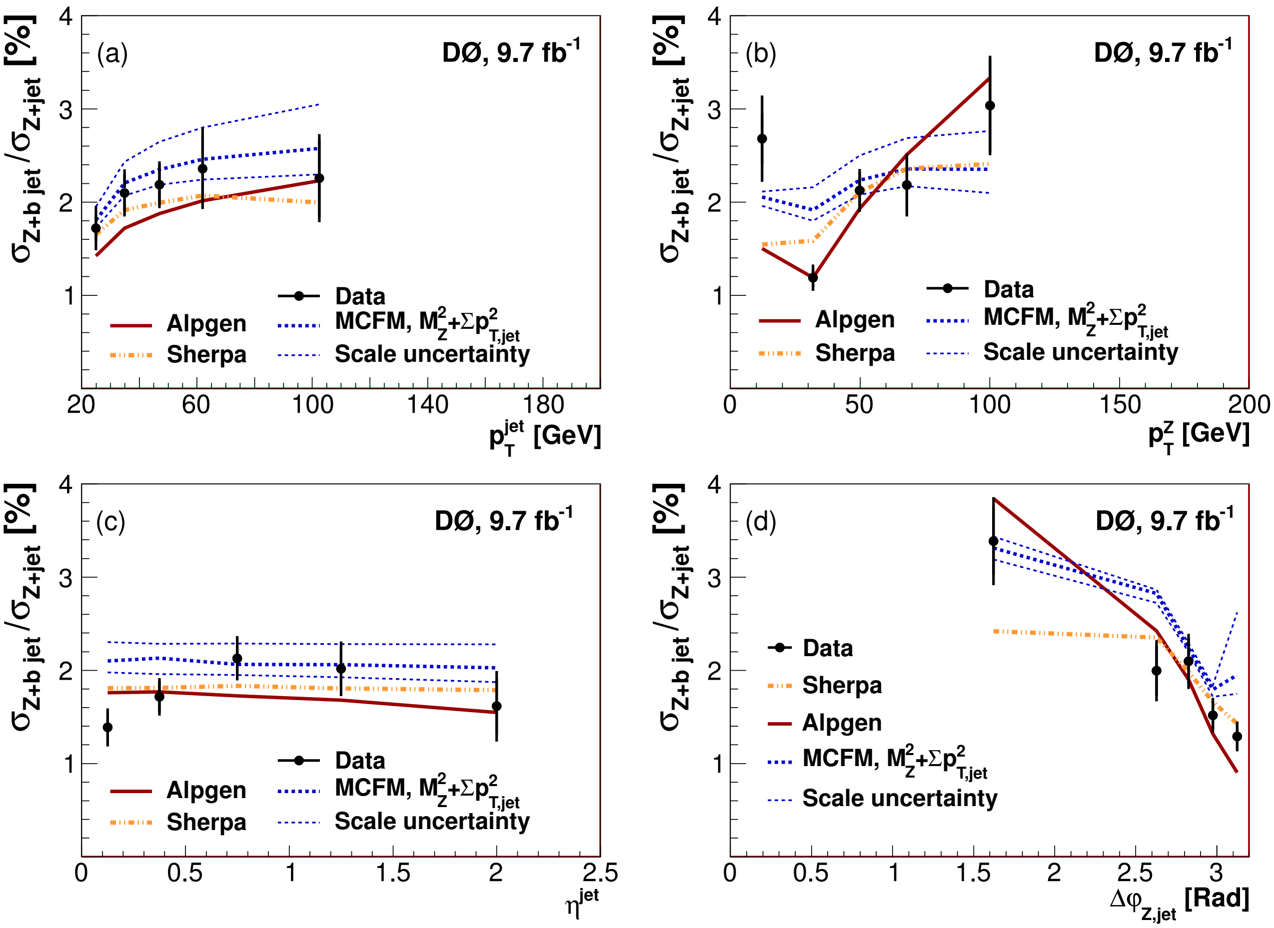


Figure 7.23: Ratios of the differential cross sections (a) jet p_T, (b) Z boson p_T, (c) jet η, and (d) $\Delta\varphi\,(Z, \mathrm{jet})$. The uncertainties on the data include the statistical and systematic uncertainties added in quadrature. The band represents the scale uncertainty for the MCFM calculations [82]. Locations of the bin centers were determined using the prescription described in Section 6.4.

to what are observed in data. If we exclude the first bin of the distribution the predictions from ALPGEN most accurately describe the measured dependence. None of the predictions accurately describe the shape of the jet η distribution, since all three sets predict that the $Z+b$ jet and $Z+$jet cross sections will have identical jet η dependence. Instead we observe a rise in the ratio at central values of η. Finally, we see that all the predictions describing the majority of the $\Delta\varphi\,(Z,\text{jet})$ distribution, except for at high values of $\Delta\varphi\,(Z,\text{jet})$. In this region the jet is recoiling off the Z boson back-to-back, and the theory might suffer from inaccurate modeling of soft radiation, which may cause the NLO calculations to break down.

7.8 Cross checks

Due to the discrepancies that we have observed between our measurements and the NLO predictions, especially with jet η, many cross checks are performed. Below is a sampling of the cross checks which vet the analysis as a function of jet η.

7.8.1 MC closure test

As a basic sanity check a MC closure test was been performed of the full analysis chain. We used ALPGEN to create a “mock data” sample and reproduced the full analysis chain, including detector acceptance, efficiencies, and the template fitting procedure. As an example Figure 7.24 shows the result of this test for the ratio of cross sections dependence on η. No systematic bias is observed in any part of this analysis chain.

7.8.2 Dependence on MVA_{bl} OP choice

There is the possibility that the choice of OP used to enrich our $Z+$jet sample has introduced a systematic bias into our analysis. To check this we create a ratio out of the ratios of cross sections for two different choices of MVA_{bl} requirements. In the numerator we vary the MVA_{bl} selection and in the denominator we fix the MVA_{bl} requirement to what was used in

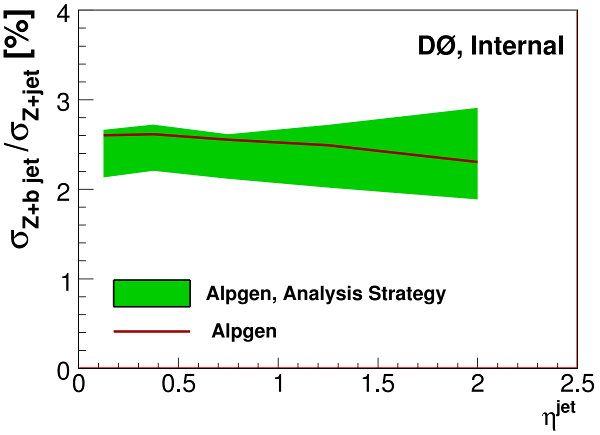


Figure 7.24: The green band represents the output of the analysis chain with all the uncertainties, including the template fitting procedure, while the red line corresponds to the ALPGEN predictions.

the analysis. Any large deviations from unity would imply a strong dependence on the choice of operating point. Figure 7.25 shows the behavior of this ratio when we vary the MVA$_{bl}$ requirement by a factor of roughly two and demonstrates that there is little dependence on the choice of MVA$_{bl}$ selection.

7.8.3 Location of primary interaction vertex

It is possible that the location of the PV may introduce a bias to the analysis. To test this, we first look at the distribution of the PV as a function of jet η (Figure 7.26a) we see no shape dependence. Additionally, the analysis was repeated as a function of jet η by requiring that the PV be reconstructed within 35 cm of the center of the detector in the beam direction, as opposed to the nominal 60 cm requirement. The resulting ratio of cross sections as a function of jet η did not fluctuate outside of their respective uncertainties as seen in Figure 7.26b. Both of these tests demonstrate that the PV location is not introducing any bias.

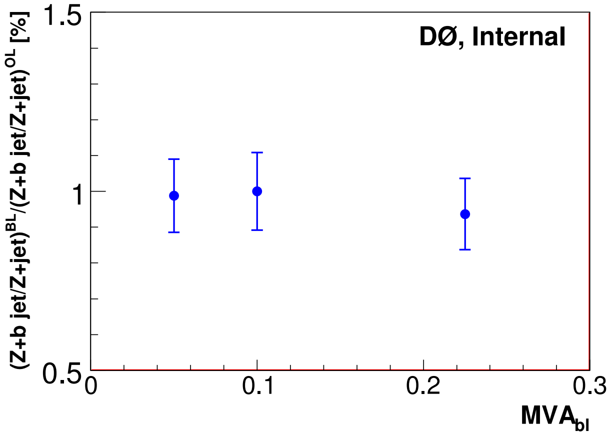


Figure 7.25: To check the dependence on choice of MVA$_{bl}$ OP we vary the MVA$_{bl}$ requirement up and down by roughly a factor of two and study the double ratio of the new value to the nominal value. Our results are consistent with one.

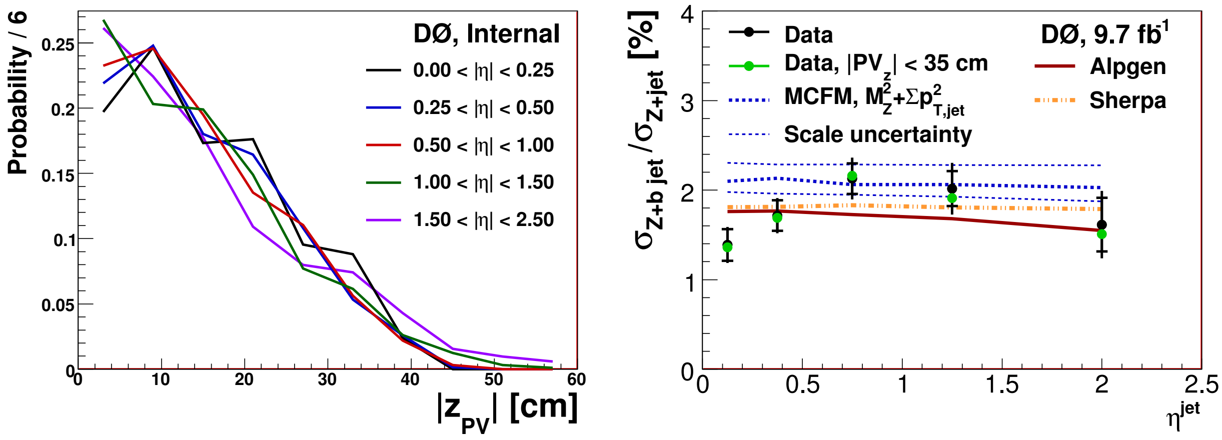


Figure 7.26: The normalized number of events as a function of the location of the PV relative to the center of the detector for the different analysis jet η bins, on the left. The final ratio of cross section measurements as a function of jet η where the green points are what is measured with the addition of a 35 cm requirement on the location of the PV, on the right.

8 Z boson plus c jet cross sections

To measure the relative fraction of c jets in the $Z + \mathrm{HF}$ sample in a reliable way we need to suppress the contribution from the light jets. Looking at Figure 6.2 we see that the light jet template shape is very similar to that of the c jet. Additionally, as it was commented in Section 7.2, the inclusion of both templates in the fit reduces our sensitivity to both f_l and f_c. To deal with this issue we choose the tightest MVA_{bl} requirement of $\mathrm{MVA}_{bl} > 0.5$, the MegaTight OP, to create a sample of $Z + \mathrm{HF}$ jets, $Z + \mathrm{HF}^{MT}$, which has a highly suppressed light jet contribution. We then use the correction factors derived from the data driven SN method, discussed in Section 5.5, to correctly estimate the small residual light jet contribution which survives this stringent requirement. This contribution, estimated from the data-corrected simulations, is then subtracted from the data along with the other backgrounds. This allows for the data to be fit with only b and c jet templates.

After the application of the MegaTight OP we have a total of 2,665 events in the $Z{+}\mathrm{HF}^{MT}$ sample. From here the estimated background contributions are subtracted and a resulting 2,125 $Z + \mathrm{HF}$ jet events are found, containing only $Z + b$ and $Z + c$ jet events. Finally, the ratios of cross sections are measured and compared to a variety of theoretical and MC event generator predictions.

Table 8.1: Estimated background contributions that have contaminated the Z+HF^{MT} data sample for the combined ($\mu\mu + ee$) sample.

Sample	Number of events
Data	2665
Multijet	107
ZZ	46
WZ	16
WW	1
$t\bar{t}$	56
Z + light jets	341

8.1 Background subtraction

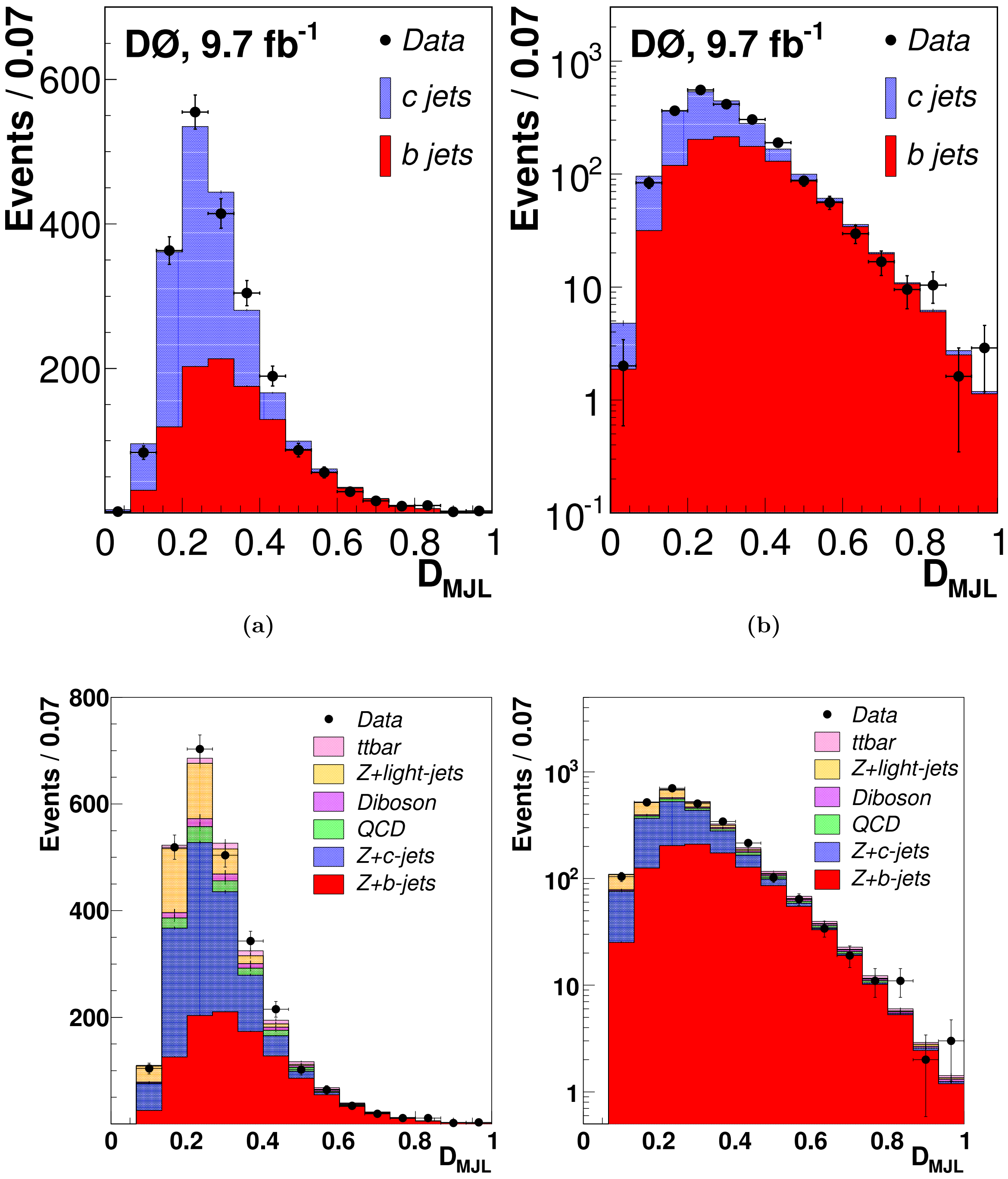
Similar to Section 7.1, we estimate the various non-Z + jet backgrounds which contribute to this sample, but now we also estimate the contribution from the Z + light jet contamination. These background contributions, as estimated for the Z + HF^{MT} sample, are listed in Table 8.1 and are subtracted from the data D_{MJL} distribution. This yields a background-subtracted data D_{MJL} distribution containing 2,125 events which will be fit with the MC Z + HF jet templates.

8.2 Flavor fraction measurement

In Table 8.2 we can see the flavor composition of the electron and muon channels as well as their combination. In Figure 8.1(a,b) we can see the background-subtracted data D_{MJL} distribution fitted with the flavor fraction weighted templates. Additionally, once we have

Table 8.2: Table of extracted jet flavor fractions from the Z + HF^{MT} data sample.

Jet Flavor	Fraction (ee)	Fraction ($\mu\mu$)	Fraction ($\mu\mu + ee$)
Events	1056	1066	2125
Z+b jet, f_b	0.491 ± 0.041	0.530 ± 0.039	0.514 ± 0.028
Z+c jet, f_c	0.509 ± 0.041	0.470 ± 0.039	0.486 ± 0.028



(a) **(b)**

(c) **(d)**

Figure 8.1: Distribution of observed events for D_{MJL} discriminant with $\mathrm{MVA}_{bl} > 0.5$. The distributions for the b and c jet templates are shown normalized to their fitted fraction after background subtraction on a (a) linear and (b) logarithmic scale. We can also add back the contributions due to the backgrounds and view them on a (c) linear and (d) logarithmic scale.

measured the b and c jet flavor fractions, we can add the backgrounds back to the data and display them with the $Z + b$ and $Z + c$ jet signal, to compare the scale of the contributions. Using the flavor fractions from Table 8.2 to scale the templates and the yields from Table 8.1 we can estimate the contribution from each background. These results are shown in Figure 8.1(c,d).

8.3 Integrated ratios of cross sections

The ratio of the cross sections of $\sigma\left(Z + c\ \text{jet}\right)/\sigma\left(Z +\ \text{jet}\right)$ can be calculated using the following form:

$$\frac{\sigma\left(Z + c\ \text{jet}\right)}{\sigma\left(Z + \text{jet}\right)} = \frac{N^{MT}\, f_c}{N_{Z+jet}\varepsilon^{c}_{btag}\varepsilon^{c}_{tagg}} \times \frac{\mathcal{A}_{incl}}{\mathcal{A}_c} \tag{8.1}$$

where N^{MT} is the number of events in the background-subtracted $Z + \mathrm{HF}^{MT}$ sample, N_{Z+jet} is the number of selected events in the background-subtracted $Z + \text{jet}$ sample, f_c is the extracted c jet flavor fraction, ε^{c}_{btag} refers to the efficiency of identifying a c jet with the MVA$_{bl}$ algorithm, ε^{c}_{tagg} is the efficiency identifying a c jet as taggable, and finally $^{\mathcal{A}_{incl}}/_{\mathcal{A}_c}$ is the ratio of acceptances, as discussed in Section 6.3. These variables are listed in Table 8.3 and, when combined, yield the first ever probe of the $Z + c$ jet production,

$$\frac{\sigma\left(Z + c\ \text{jet}\right)}{\sigma\left(Z + \text{jet}\right)} = 0.0829 \pm 0.0052\,(\text{stat})\,.$$

Table 8.3: Table of efficiencies and event totals used in the integrated $\sigma\left(Z + c\ \text{jet}\right)/\sigma\left(Z +\ \text{jet}\right)$ measurement.

	Totals
ϵ^{c}_{btag}	0.0901 ± 0.0001
ϵ^{c}_{tagg}	0.891 ± 0.01
N_{Z+jet}	167389 ± 409
N^{MT}	2125 ± 46

Table 8.4: Table of efficiencies used in the integrated $\sigma\left(Z+c\text{ jet}\right)/\sigma\left(Z+b\text{ jet}\right)$ measurement.

	Total efficiency
ϵ^{b}_{btag}	0.400 ± 0.006
ϵ^{b}_{tagg}	0.902 ± 0.01
ϵ^{c}_{btag}	0.0901 ± 0.0001
ϵ^{c}_{tagg}	0.891 ± 0.01

This can be compared with our measurements using the flavor fractions measured in Chapter 7. The c jet flavor fraction suffers from larger uncertainties due to the similarity with the light jet template, but can be used to cross check the current analysis strategy. Using this and the efficiencies measured in Section 5.4 we find that the three parameter fit yields a consistent result of

$$\frac{\sigma\left(Z+c\text{ jet}\right)}{\sigma\left(Z+\text{jet}\right)} = 0.076 \pm 0.007\,(\text{stat})\,.$$

The ratio $\sigma\left(Z+c\text{ jet}\right)/\sigma\left(Z+b\text{ jet}\right)$ can be interpreted as the double ratio $\sigma\left(Z+c\text{ jet}\right)/\sigma\left(Z+\text{jet}\right)$ over $\sigma\left(Z+b\text{ jet}\right)/\sigma\left(Z+\text{jet}\right)$. Performing this measurement allows us to cancel an extended set of variables and uncertainties. To measure the ratio of the cross sections of $Z+c$ jet to $Z+b$ jet, we calculate using the following equation:

$$\frac{\sigma\left(Z+c\text{ jet}\right)}{\sigma\left(Z+b\text{ jet}\right)} = \frac{f_c\epsilon^{b}_{btag}\epsilon^{b}_{tagg}}{f_b\epsilon^{c}_{btag}\epsilon^{c}_{tagg}} \times \frac{\mathcal{A}_b}{\mathcal{A}_c} \tag{8.2}$$

where f_b and f_c are the b and c jet flavor fractions listed in Table 8.2. The various efficiencies can be found in Table 8.4. $\mathcal{A}_b/\mathcal{A}_c$ is the ratio of acceptances covered in Section 6.3.

Since the extracted flavor fractions will be correlated we need to correctly account for this in our uncertainty propagation. Using the global correlation of the two parameter fit,

ρ_{bc}, as provided by the `TFractionFitter`, we use the following formula:

$$\sigma_F = F\sqrt{\left(\frac{\sigma_b}{f_b}\right)^2 + \left(\frac{\sigma_c}{f_c}\right)^2 - 2\rho_{bc}\frac{\sigma_c\sigma_b}{f_c f_b}} \tag{8.3}$$

where σ_c and σ_b are the uncertainties (statistical in this case) on the extracted c and b jet fractions, respectively, and F is the ratio of ${}^{f_c}/_{f_b}$. For the integrated ratio of cross sections measurement the two template fit has a correlation of $\rho_{bc} = 0.706$. Using all of this information, we measure the ratio of cross sections as

$$\frac{\sigma\left(Z + c\text{ jet}\right)}{\sigma\left(Z + b\text{ jet}\right)} = 4.00 \pm 0.21\,(\text{stat})\,.$$

8.4 Jet transverse momentum differential measurement

The same method used in computing the integrated ratios of cross sections is repeated in bins of jet p_T. Due to the reduced statistics of the $Z + \text{HF}^{MT}$ sample we only use four bins for the determination of the dependence of the ratios of cross sections. To begin, we need to estimate the background contributions, both non-$Z +$ jet and $Z +$ light jet, which will be subtracted on a bin-by-bin basis. Table 8.5 shows the estimated background contributions which will be subtracted from the data before the template fitting. Table 8.5 also lists the percent contribution for each background, with respect to the total number of events in that bin. It can be seen that the $Z +$ light jet background, even with the MegaTight requirement, has the largest contribution.

Table 8.6 and Table 8.7 shows the bin-by-bin results for the measurements of the ratios of cross sections as a function of jet p_T. In Table 8.6 all the variables necessary to measure $\sigma(Z + c\text{ jet})/\sigma(Z + \text{jet})$, along with the relative statistical and systematic uncertainties are provided. While Table 8.7 contains the inputs for $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$ along with the correlation between the two flavor fractions and the relative statistical and systematic uncertainties of the ratios of cross sections measurements. Figure 8.2 to Figure 8.5 show the

Table 8.5: Estimated background contributions that have contaminated the combined ($\mu\mu$+ ee) $Z + \text{HF}^{MT}$ data sample in bins of jet p_T.

Background	Jet p_T [GeV]							
	$20 - 30$		$30 - 40$		$40 - 60$		$60 - 200$	
Data	1303		472		384		473	
Multijet	62	4.7%	17	3.7%	16	4.2%	11	2.4%
ZZ	11	0.9%	9	2.0%	12	3.0%	13	2.8%
WZ	5	0.4%	3	0.7%	4	1.0%	4	0.8%
WW	0.5	0.03%	0.3	0.06%	0.2	0.06%	0.2	0.03%
$t\bar{t}$	6	0.5%	7	1.5%	12	3.2%	30	6.4%
Z+light jets	206	15.8%	49	10.3%	32	8.4%	39	8.3%

Table 8.6: Table of bin-by-bin information for $\sigma(Z + c\text{ jet})/\sigma(Z + \text{jet})$ ratio of differential cross sections for jet p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Jet p_T [GeV]	N^{MT}	N_{Z+jet}	f_c [%]	ϵ^c_{btag}[%]	ϵ^c_{tagg}[%]	$\frac{\sigma(Z+c\text{ jet})}{\sigma(Z+\text{jet})}$	Stat. [%]	Syst. [%]
$20 - 30$ (24.6)	741	80586	46.0 ± 5.7	7.88 ± 0.09	84.7 ± 0.7	0.068	12	16
$30 - 40$ (34.3)	525	35508	48.4 ± 5.8	9.49 ± 0.2	92.1 ± 1.2	0.084	11	12
$40 - 60$ (47.3)	474	25526	48.2 ± 5.6	10.3 ± 0.2	93.4 ± 1.7	0.099	11	9.1
$60 - 200$ (78.0)	380	25603	50.3 ± 6.4	10.5 ± 0.5	93.8 ± 3.9	0.085	13	11

Table 8.7: Table of bin-by-bin information for $\sigma(Z+c\text{ jet})/\sigma(Z+b\text{ jet})$ ratio of differential cross sections for jet p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

	Jet p_T [GeV]			
	$20-30$ (24.6)	$30-40$ (34.3)	$40-60$ (47.3)	$60-200$ (78.0)
f_b [%]	54.0 ± 4.7	51.8 ± 5.5	51.9 ± 5.6	49.8 ± 6.4
f_c [%]	46.0 ± 5.7	48.4 ± 5.8	48.2 ± 5.6	50.3 ± 6.4
ρ_{bc}	0.743	0.769	0.654	0.725
ϵ^b_{btag} [%]	35.0 ± 0.3	42.2 ± 0.6	46.2 ± 0.9	46 ± 2
ϵ^b_{tagg} [%]	86.6 ± 0.7	93 ± 1	94 ± 2	94 ± 4
ϵ^c_{btag} [%]	7.88 ± 0.09	9.49 ± 0.2	10.3 ± 0.2	10.5 ± 0.5
ϵ^c_{tagg} [%]	84.7 ± 0.7	92.1 ± 1.2	93.4 ± 1.7	93.8 ± 3.9
$\frac{\sigma(Z+c\text{ jet})}{\sigma(Z+b\text{ jet})}$	3.64	3.97	3.98	4.30
Stat. [%]	8.5	8.3	10	13
Syst. [%]	21	14	13	14

background-subtracted $Z + \text{HF}^{MT}$ data D_{MJL} distribution fitted with the b and c jet MC templates.

8.5 Z boson transverse momentum differential measurement

The same procedure used in Section 8.4 is applied here to extract the dependence of the ratios of cross sections on Z boson p_T. Table 8.8 lists the background contributions as a function of Z boson p_T, where we see the second bin, $20 < p_T^Z < 40$ GeV, has the largest contribution due to light jets. This bin is dominated by one jet events, in contrast to the first bin which is dominated by two jet events.

Table 8.9 and Table 8.10 show the results from the ratios of cross sections. We see that the measured ratios of cross sections $\sigma(Z+c\text{ jet})/\sigma(Z+\text{jet})$ behaves differently to $\sigma(Z+b\text{ jet})/\sigma(Z+\text{jet})$ at low Z boson p_T. $\sigma(Z+b\text{ jet})/\sigma(Z+\text{jet})$ was enhanced in the first

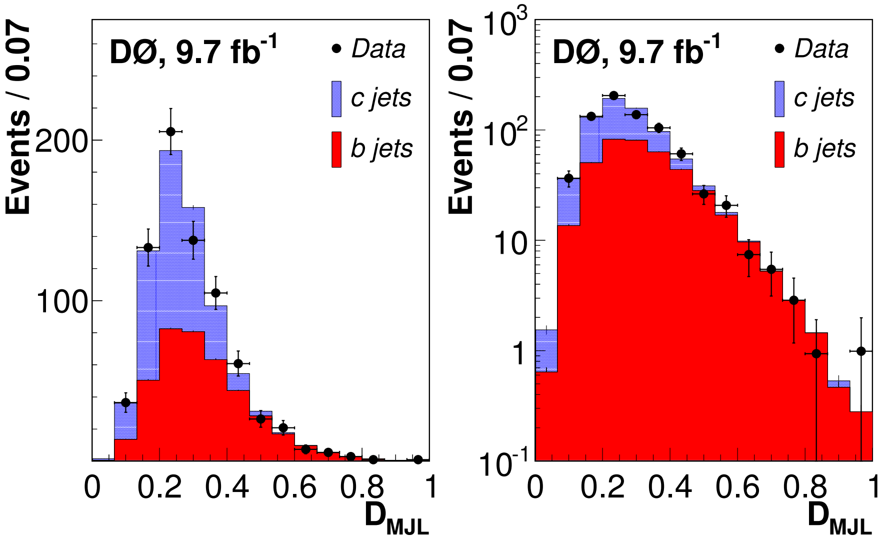


Figure 8.2: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $20 <$ jet $p_T < 30$ GeV.

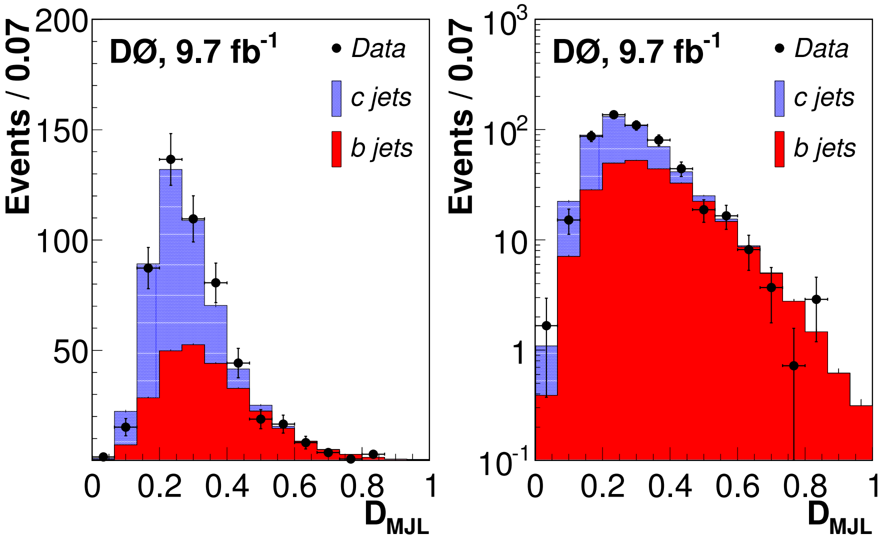


Figure 8.3: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $30 <$ jet $p_T < 40$ GeV.

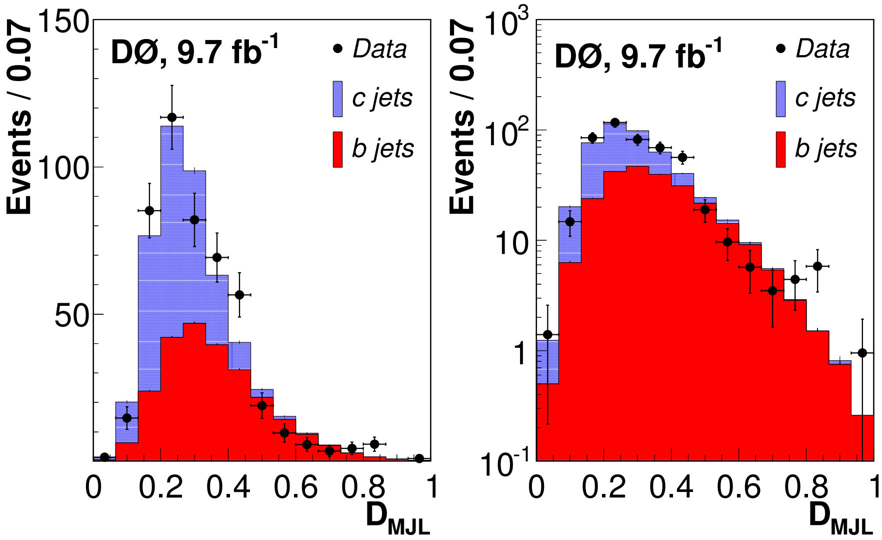


Figure 8.4: MC $D_{\rm MJL}$ templates normalized to their extracted flavor fraction and the data distribution for $40 <$ jet $p_T < 60$ GeV.

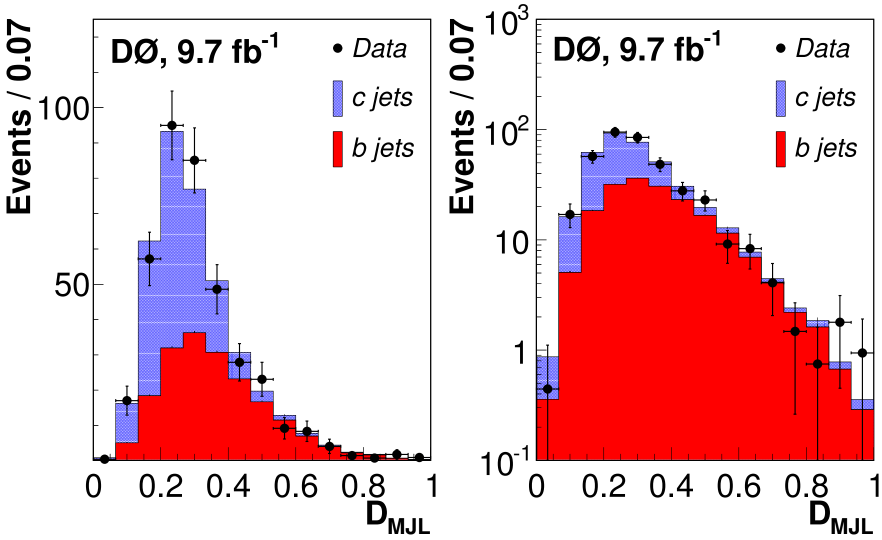


Figure 8.5: MC $D_{\rm MJL}$ templates normalized to their extracted flavor fraction and the data distribution for $60 <$ jet $p_T < 200$ GeV.

Table 8.8: Estimated background contributions that have contaminated the combined ($\mu\mu+ee$) $Z+\mathrm{HF}^{MT}$ data sample in bins of Z boson p_T.

Background	Z boson p_T [GeV]							
	$0-20$		$20-40$		$40-60$		$60-200$	
Data	383		967		692		590	
Multijet	39	10.2%	44	11.6%	18	4.8%	4	1.1%
ZZ	6	1.5%	12	3.0%	11	2.9%	17	4.4%
WZ	2	0.4%	4	1.1%	4	1.0%	6	1.5%
WW	0.2	0.04%	0.5	0.1%	0.2	0.04%	0.4	0.1%
$t\bar{t}$	6	1.4%	12	3.3%	14	3.6%	24	6.3%
Z+light jets	57	14.9%	141	36.8%	72	18.8%	56	14.6%

Table 8.9: Table of bin-by-bin information for $\sigma(Z+c\text{ jet})/\sigma(Z+\text{jet})$ ratio of differential cross sections for Z boson p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Z p_T [GeV]	N^{MT}	N_{Z+jet}	f_c [%]	ϵ^c_{btag}[%]	ϵ^c_{tagg}[%]	$\frac{\sigma(Z+c\text{ jet})}{\sigma(Z+\text{jet})}$	Stat. [%]	Syst. [%]
$0-20$ (10.2)	285	35190	24.2 ± 8.1	8.0 ± 0.2	83 ± 1	0.041	29	22
$20-40$ (29.5)	763	72254	59.1 ± 4.7	9.0 ± 0.1	88.9 ± 0.9	0.073	8.2	12
$40-60$ (49.0)	588	35984	53.3 ± 5.1	9.7 ± 0.2	92 ± 1	0.104	10	11
$60-200$ (92.7)	487	23820	42.7 ± 5.8	9.7 ± 0.4	93 ± 3	0.108	13	8.3

Table 8.10: Table of bin-by-bin information for $\sigma(Z+c \text{ jet})/\sigma(Z+b \text{ jet})$ ratio of differential cross sections for Z boson p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

$\frac{Z+c \text{ jet}}{Z+b \text{ jet}}$	Z p_T [GeV]			
	0 – 20 (10.2)	20 – 40 (29.5)	40 – 60 (49.0)	60 – 200 (92.7)
f_b [%]	75.8 ± 9.0	41.0 ± 4.5	46.7 ± 5.0	57.4 ± 6.0
f_c [%]	24.2 ± 8.1	59.1 ± 4.7	53.3 ± 5.1	42.7 ± 5.8
ρ_{bc}	0.767	0.679	0.660	0.696
ϵ^b_{btag} [%]	34.1 ± 0.8	39.1 ± 0.8	43 ± 1	45 ± 2
ϵ^b_{tagg} [%]	84 ± 2	90 ± 2	93 ± 3	94 ± 4
ϵ^c_{btag} [%]	8.0 ± 0.2	9.0 ± 0.1	9.7 ± 0.2	9.7 ± 0.4
ϵ^c_{tagg} [%]	83 ± 1	88.9 ± 0.9	92 ± 1	93 ± 3
$\frac{\sigma(Z+c \text{ jet})}{\sigma(Z+b \text{ jet})}$	1.15	6.10	5.06	3.41
Stat. [%]	26	8.2	10	13
Syst. [%]	32	20	15	13

bin of Z boson p_T, and we see that $\sigma(Z + c \text{ jet})/\sigma(Z + \text{jet})$ is, instead, highly suppressed. Since we interpret $\sigma(Z+c \text{ jet})/\sigma(Z+b \text{ jet})$ as the double ratio of these two ratios, it follows that we see a very strong suppression of the first bin relative to the second bin. This first bin has a total relative uncertainty of 41%, the largest of any bin.

Figure 8.6 to Figure 8.9 show the background-subtracted data, fitted with the MC b and c jet templates. We see that the first bin suffers from low statistics but is still accurately described by the two template fit.

8.6 Systematic uncertainties

Many of the systematic uncertainties that were discussed in Section 7.5 are applicable for $\sigma(Z + c \text{ jet})/\sigma(Z + \text{jet})$ and $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet})$. Only the systematic uncertainties which are additional, removed due to cancelation, or evaluated differently are discussed further here. Table 8.11 to Table 8.13 and Table 8.14 to Table 8.16 document the full list of

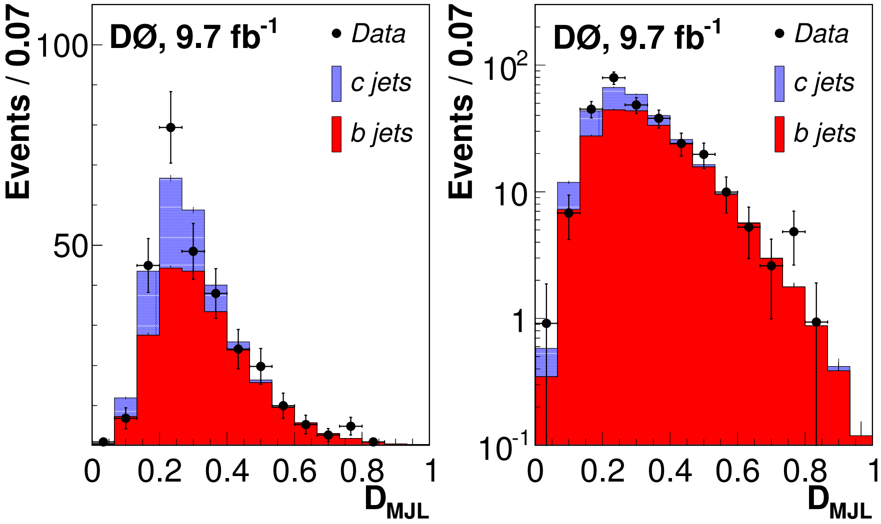


Figure 8.6: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $0 < \mathrm{Z}\ p_T < 20$ GeV.

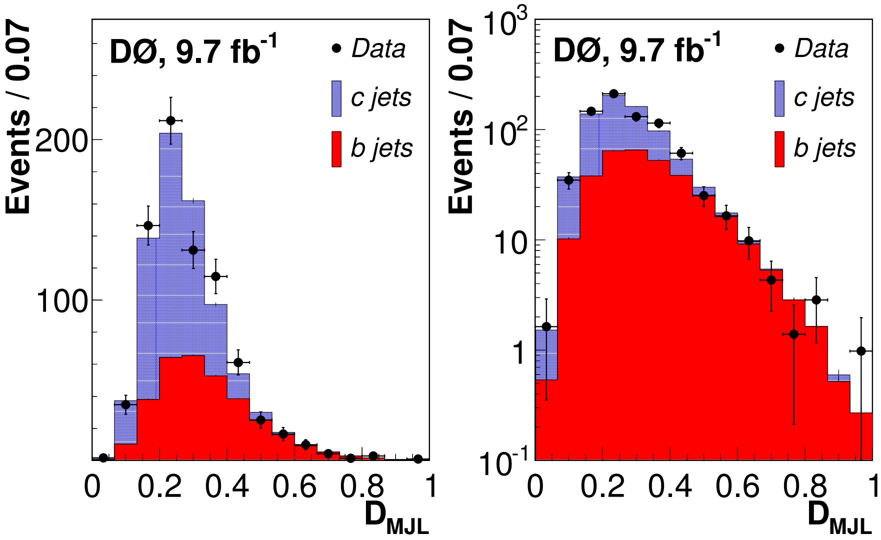


Figure 8.7: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $20 < \mathrm{Z}\ p_T < 40$ GeV.

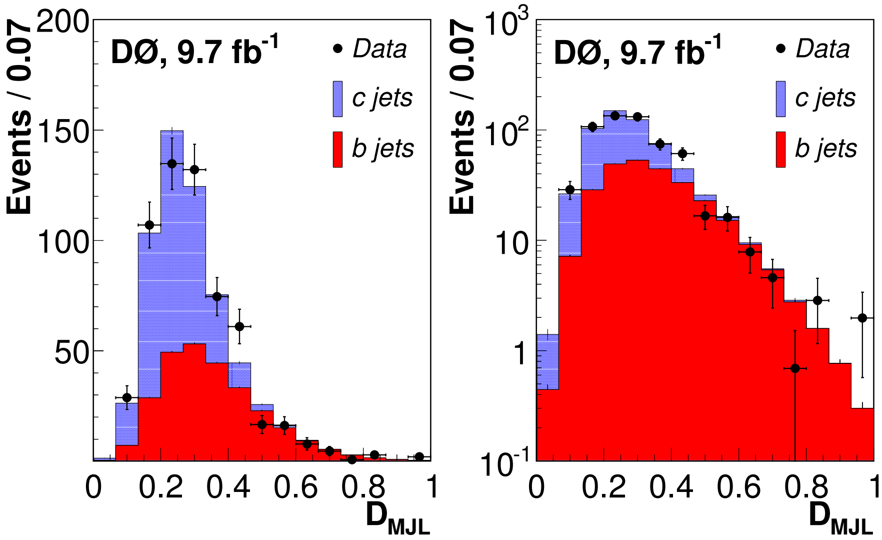


Figure 8.8: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $40 < \text{Z } p_T < 60$ GeV.

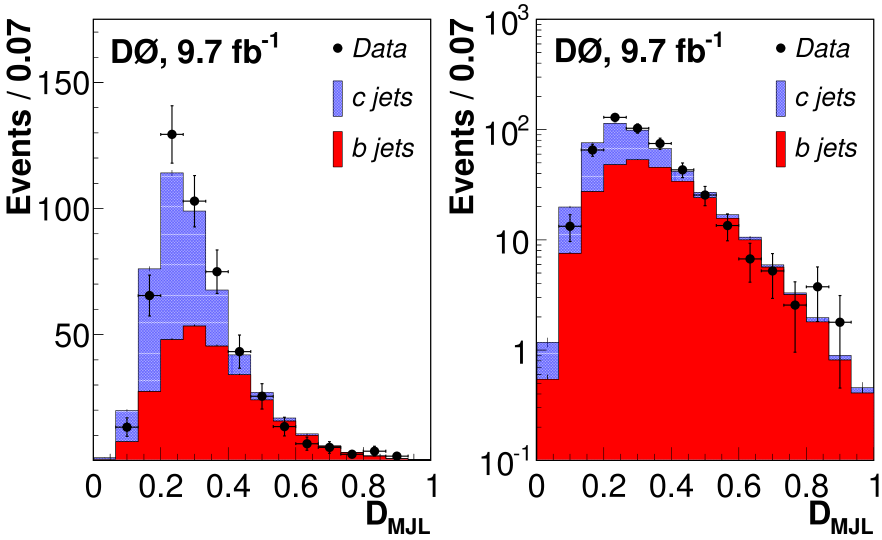


Figure 8.9: MC D_{MJL} templates normalized to their extracted flavor fraction and the data distribution for $60 < \text{Z } p_T < 200$ GeV.

Table 8.11: Systematic uncertainties for the integrated $\sigma(Z+c\text{ jet})/\sigma(Z+\text{jet})$. “T. Shape” refers to the uncertainties which affect the shape of the MC templates used during the fitting procedure.

Systematic Uncertainty		Percentage [%]
JES		1.7
JER		3.0
c jet identification		1.9
c jet detector acceptance		< 0.1
Background estimation		0.5
Light jet estimation		8.1
T. Shape	Light jet template shape	4.8
	HF merging	2.6
	Template fitting	0.9
Total		10.6

applicable systematic uncertainties and the size of their contributions to the total for the $\sigma(Z+c\text{ jet})/\sigma(Z+\text{jet})$ and $\sigma(Z+c\text{ jet})/\sigma(Z+b\text{ jet})$ measurements, respectively.

8.6.1 Z + c jet / Z + jet

- Light jet estimation:

  To vet the modeled Z+light jet contribution remaining in our final selection sample we compare the MC estimation to a 3-parameter `TFractionFitter` output. To do this we select a looser MVA$_{bl}$ OP point choice, VeryTight, where we can get reliably convergent fits. In Figure 8.10 the extracted data light jet fraction is compared to that estimated by the data-corrected simulations. To help constrain the uncertainties on the MC estimation we fit it with a best-fit-line and then measure the χ^2 between the data and this fit. If we modify our χ^2 equation with a penalty term, δ, and then vary it such that we achieve a 70% confidence level ($\chi^2 = 4.88$ for 3 degrees of freedom):

$$\chi^2 = \sum_{i=bins} \frac{(f_i^{MC} - d_i)^2}{(\sigma^2_{f^{MC},i} + \sigma^2_{d,i} + \delta^2)} \rightarrow (6.1 - 4.88) = 1.22 \tag{8.4}$$

Table 8.12: Systematic uncertainties for $\sigma(Z{+}c\text{ jet})/\sigma(Z{+}\text{jet})$ in bins of jet p_T. “T. Shape” refers to the uncertainties which affect the shape of the MC templates used during the fitting procedure.

	Jet p_T	Percentage [%]			
	Systematic Uncertainty	20 – 30 GeV	30 – 40 GeV	40 – 60 GeV	60 – 200 GeV
	JES	3.5	2.0	1.2	0.8
	JER	4.5	2.9	1.9	2.2
	c jet identification	1.6	1.7	2.0	2.5
	c jet detector acceptance	0.3	1.0	1.9	3.3
	Background estimation	0.4	2.5	1.4	3.6
	Light jet estimation	13.4	8.8	5.9	7.5
T. Shape	Light jet template shape	4.8			
	HF merging	2.9	2.3	1.8	0.2
	Template fitting	1.7	1.9	1.9	2.3
	Total	15.8	11.5	8.8	11.0

Table 8.13: Systematic uncertainties for $\sigma(Z + c\text{ jet})/\sigma(Z + \text{jet})$ in bins of Z boson p_T. “T. Shape” refers to the uncertainties which affect the shape of the MC templates used during the fitting procedure.

	Z p_T	Percentage [%]			
	Systematic Uncertainty	0 – 20 GeV	20 – 40 GeV	40 – 60 GeV	60 – 200 GeV
	JES	3.2	1.3	0.3	0.5
	JER	1.7	1.3	0.3	0.7
	c jet identification	1.7	1.7	1.9	2.1
	c jet detector acceptance	10.4	0.7	5.9	0.8
	Background estimation	2.5	0.3	1.0	< 0.1
	Light jet estimation	17.0	10.7	6.1	5.6
T. Shape	Light jet template shape	4.8			
	HF merging	2.6	2.5	3.3	1.0
	Template fitting	2.5	1.7	1.7	2.0
	Total	21.4	12.4	10.6	8.0

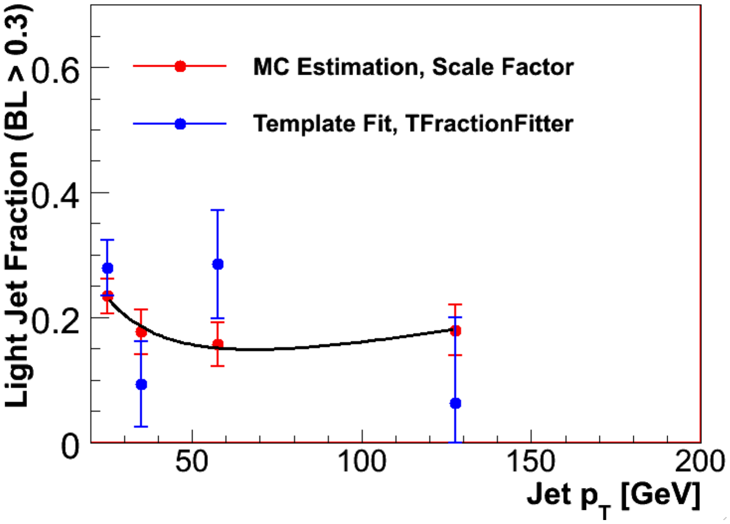


Figure 8.10: A comparison of the MC estimates of the light jet fraction and the extracted light jet fraction after a requirement $\text{MVA}_{bl} > 0.3$. The curve represents the best fit curve of the MC.

this results in a $\delta \approx 20\%$. We then vary the amount of light jets which we subtract by this 20% and take the difference in the final ratio as a systematic uncertainty. Additional methods of determining this systematic were also evaluated, by applying the penalty terms in different ways. All yielded consistent results.

Template Shape Uncertainties:

- Collapsed heavy flavored jets:

  When two HF jets are mis-reconstructed as a single jet they can affect the template shapes and the tagging efficiency. This is a result of two b or c quarks being matched within a cone of radius $\Delta R = 0.5$ to a jet which passes our selection criteria. To estimate this effect we inflate the contribution of these jets in our simulations by a factor of two and repeat the analysis.

8.6.2 Z + c jet / Z + b jet

Since the HF tagging efficiencies in the numerator and denominator use identical scale factors the uncertainty for this identification efficiency will cancel. All the systematics are the same

Table 8.14: Systematic uncertainties for the integrated $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet})$ measurement.

	Systematic Uncertainty	Percentage [%]
	JES	3.0
	JER	2.0
	c jet detector acceptance	< 0.1
	Background estimation	1.0
	Light jet estimation	8.7
Template Shape	Light jet template shape	9.7
	b jet merging	3.7
	c jet merging	2.6
	Template fitting	1.4
	Total	14.4

as the previous section except for:

Template Shape Uncertainties:

- Template fitting:

  This is performed as described in Section 7.5 but the uncertainty associated with the finite template statistics will be diminished slightly by taking into account the correlation between the *c* and *b* jet templates. Using Equation 8.3 and the correlations reported in Table 8.7 and Table 8.10 the correct correlations are utilized.

8.7 Theory and MC comparisons

Using MCFM, as described in Section 7.6, we can produce predictions for the ratio of integrated cross sections for our kinematic selections and using the Run II cone algorithm. As in Section 7.6 we again use the MSTW2008 PDF set when generating these predictions and the dynamic scale choice of $\mu_R^2 = \mu_F^2 = M_Z^2 + p_{T,\text{total}}^2$. This yields a prediction of the NLO

Table 8.15: Systematic uncertainties for $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet})$ in bins of jet p_{T}.

	Jet p_T	Percentage [%]			
	Systematic Uncertainty	$20 - 30$ GeV	$30 - 40$ GeV	$40 - 60$ GeV	$60 - 200$ GeV
	JES	4.0	3.8	2.7	1.9
	JER	2.3	3.1	2.2	3.5
	c jet detector acceptance	0.3	1.0	1.9	3.3
	Background estimation	0.5	1.5	1.3	2.3
	Light jet estimation	16.2	8.3	5.5	5.3
Template Shape	Light jet template shape	9.7			
	b jet merging	6.3	2.2	1.8	2.7
	c jet merging	2.8	2.5	1.8	2.2
	Template fitting	2.1	2.4	2.9	3.4
	Total	20.7	14.4	12.7	13.4

Table 8.16: Systematic uncertainties for $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet})$ in bins of Z boson p_{T}.

	Z p_T	Percentage [%]			
	Systematic Uncertainty	$0 - 20$ GeV	$20 - 40$ GeV	$40 - 60$ GeV	$60 - 200$ GeV
	JES	9.6	2.6	1.3	1.1
	JER	8.8	1.3	0.4	0.4
	c jet detector acceptance	10.4	0.7	5.9	0.8
	Background estimation	0.9	1.2	0.8	1.5
	Light jet estimation	24.2	16.3	7.3	6.2
Template Shape	Light jet template shape	9.7			
	b jet merging	7.8	2.8	4.0	4.9
	c jet merging	2.6	2.4	3.2	0.4
	Template fitting	2.5	4.4	3.2	2.5
	Total	32.1	20.1	15.0	12.9

rate of $Z + c$ jet production of

$$\left[\frac{\sigma(Z + c\text{ jet})}{\sigma(Z + \text{jet})}\right]^{MCFM}_{MSTW2008} = \left(3.68^{+0.63}_{-0.39}\right)\%$$

$$\left[\frac{\sigma(Z + c\text{ jet})}{\sigma(Z + b\text{ jet})}\right]^{MCFM}_{MSTW2008} = 1.64$$

where the uncertainty, which cancels for $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$, is estimated by increasing the normalization and factorization scales up and down by a factor of two. These predictions underestimate the contribution of the $Z + c$ jet production by a factor of 2.5 when compared to the measured ratios of cross sections. Further, if we assume Gaussian uncertainties, the significance of this disagreement is 4σ and 6σ, for the $\sigma(Z + c\text{ jet})/\sigma(Z + \text{jet})$ and $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$ measurements, respectively. We can also compare the feasibility of increasing the contributions of the sea-like intrinsic charm in the initial state by using a specially constructed PDF set, CTEQ6.6c [57]. This PDF set enhances the contribution of the sea-like intrinsic charm by 3.5% at low energies. This is highly disfavored by fits to data but is the maximum increase in the intrinsic charm recommended by CTEQ [57]. Using this new PDF set we obtain predictions of

$$\left[\frac{\sigma(Z + c\text{ jet})}{\sigma(Z + \text{jet})}\right]^{MCFM}_{CTEQ6.6c} = \left(4.25^{+0.48}_{-0.29}\right)\%$$

$$\left[\frac{\sigma(Z + c\text{ jet})}{\sigma(Z + b\text{ jet})}\right]^{MCFM}_{CTEQ6.6c} = 2.23$$

which are still significantly below the ratios that have been measured.

In contrast with the $Z + b$ jet analysis we will add PYTHIA to the set of MC event generators we use to produce predictions. PYTHIA includes only $2 \to 2$ MEs with $gQ \to ZQ$ and $q\bar{q} \to Zg$ scatterings followed by $g \to Q\bar{Q}$ splitting. The Perugia0 tune [83] and the CTEQ6L1 PDF [57] set are used for the PYTHIA predictions.

An additional correction is derived to account for previous measurements at LEP [14, 15, 16, 17, 18, 19] and SLD [14, 20], where it was found that there is tension between the rate at which gluons split into $c\bar{c}$ pairs as predicted [21, 22] and that which is measured. If we look at one of the main processes which we measure, $q\bar{q} \rightarrow Zg$, we find that this is directly analogous to gluon splitting. We can enhance this diagram in our predictions to give some insight into whether this is related to the discrepancy we are measuring. We enhance the contribution to the PYTHIA annihilation diagram by a factor of 1.7, inline with discrepancy observed in the $\gamma + c$ jet analyses at the Tevatron [25, 26].

8.8 Results

The ratios $\sigma(Z{+}c \text{ jet})/\sigma(Z{+}\text{jet})$ and $\sigma(Z{+}c \text{ jet})/\sigma(Z{+}b \text{ jet})$ are measured in the combined $Z{\rightarrow}\ell\ell\,(\ell = e, \mu)$ final state using 9.7 fb^{-1} of Run II D0 data [84]. We find the results to be

$$\frac{\sigma(Z + c)}{\sigma(Z + jet)} = (8.29 \pm 0.52(\text{stat}) \pm 0.89(\text{syst}))\,\%$$

$$\frac{\sigma(Z + c)}{\sigma(Z + b)} = 4.00 \pm 0.21(\text{stat}) \pm 0.58(\text{syst}).$$

These measurements used jets of $p_T > 20$ GeV in the pseudorapidity region of $|\eta| \leq 2.5$. We have also measured the differential ratio as a function of jet and Z boson p_T for direct comparison to theory. Table 8.17 to Table 8.20 show the results of the measurement of the ratio of differential cross sections.

Table 8.17: Final results for the $\sigma(Z + c \text{ jet})/\sigma(Z + \text{jet})$ ratio of differential cross sections in bins of jet p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Jet p_T [GeV]	N^{MT}	$\frac{\sigma(Z+c\text{ jet})}{\sigma(Z+\text{jet})}$(%)	Relative Stat. Uncert.	Relative Syst. Uncert.
20 – 30 (24.6)	741	6.8	12%	16%
30 – 40 (34.3)	525	8.4	11%	12%
40 – 60 (47.3)	474	9.9	11%	9.1%
60 – 200 (78.0)	380	8.5	13%	11%

Table 8.18: Final results for the $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet})$ ratio of differential cross sections in bins of jet p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Jet p_T [GeV]	$\frac{\sigma(Z+c \text{ jet})}{\sigma(Z+b \text{ jet})}$	Relative Stat. Uncert.	Relative Syst. Uncert.
20 – 30 (24.6)	3.64	8.5%	21%
30 – 40 (34.3)	3.97	8.3%	14%
40 – 60 (47.3)	3.98	10%	13%
60 – 200 (78.0)	4.30	13%	14%

Table 8.19: Final results for the $\sigma(Z + c \text{ jet})/\sigma(Z + \text{jet})$ ratio of differential cross sections in bins of Z boson p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Z boson p_T [GeV]	N^{MT}	$\frac{\sigma(Z+c \text{ jet})}{\sigma(Z+\text{jet})}$(%)	Relative Stat. Uncert.	Relative Syst. Uncert.
0 – 20 (10.2)	285	4.1	29%	22%
20 – 40 (29.5)	763	7.3	8.2%	12%
40 – 60 (49.0)	588	10.4	10%	11%
60 – 200 (92.7)	487	10.8	13%	8.3%

Table 8.20: Final results for the $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet})$ ratio of differential cross sections in bins of Z boson p_T. For each bin width the bin center is also given in parenthesis (described in Section 6.4).

Z boson p_T [GeV]	$\frac{\sigma(Z+c \text{ jet})}{\sigma(Z+b \text{ jet})}$	Relative Stat. Uncert.	Relative Syst. Uncert.
0 – 20 (10.2)	1.15	26%	32%
20 – 40 (29.5)	6.10	8.2%	20%
40 – 60 (49.0)	5.06	10%	15%
60 – 200 (92.7)	3.41	13%	13%

Figure 8.11 compares these differential cross sections to the predictions. On average, the NLO predictions significantly underestimate the data, by a factor of 2.5, as they did for the integrated results. For the MC event generators, PYTHIA predictions describe the measurements best. The enhancement of the default rate of $g \to c\bar{c}$ in PYTHIA leads to a significant improvement in these predictions.

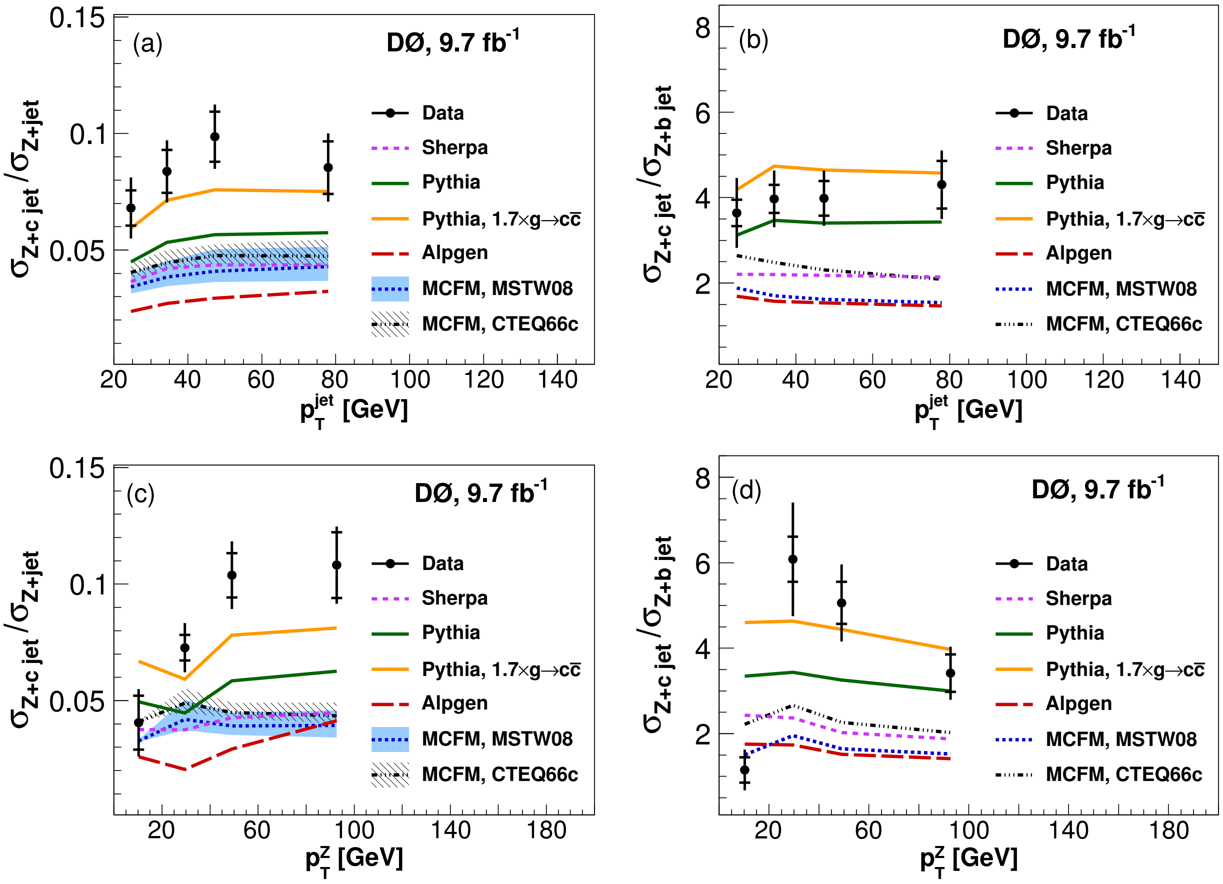


Figure 8.11: The ratios of differential cross sections $\sigma(Z+c \text{ jet})/\sigma(Z+\text{jet})$, on the left, and $\sigma(Z+c \text{ jet})/\sigma(Z+b \text{ jet})$, on the right, as a function of jet p_T, on top, and Z boson p_T, on the bottom. The inner bars represent the statistical uncertainty of the measurements and the full bar denotes the total, quadratic sum, of all the uncertainties [84].

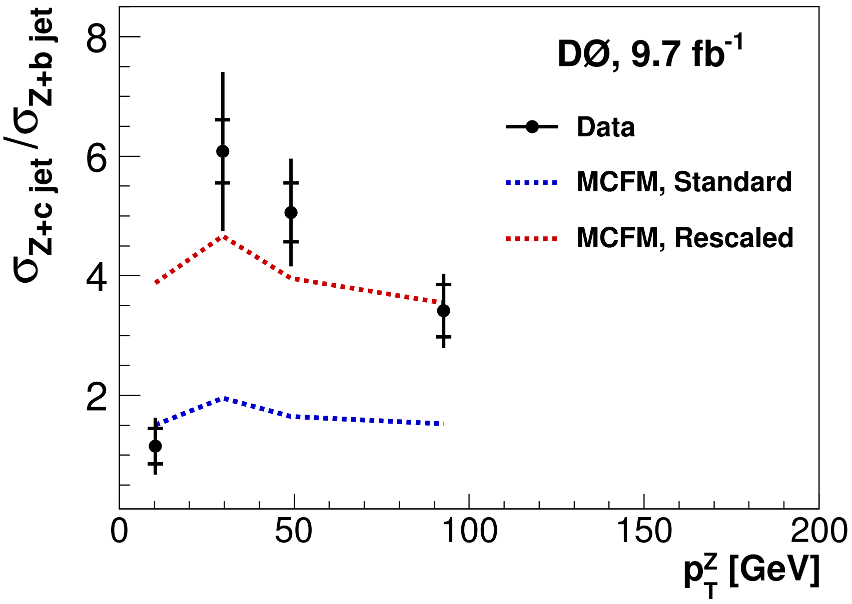


Figure 8.12: The ratios of differential cross sections $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet})$ as a function of Z boson p_T. The predictions from MCFM have been rescaled so that the integrated ratio of cross sections matches the measurement. The inner bars represent the statistical uncertainty of the measurements and the full bar denotes the total, quadratic sum, of all the uncertainties [84].

The largest discrepancy between data and predictions, in particular for the shape of the differential distributions, is for $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet})$ as a function of Z boson p_T (Figure 8.11(d)). To estimate the level of disagreement (in terms of the shape) we rescaled the standard MCFM NLO prediction to match the integrated ratios of cross sections as measured in the data. To do this we renormalize the predicted value of the integrated ratio of cross sections in MCFM to be $\sigma(Z + c \text{ jet})/\sigma(Z + b \text{ jet}) = 4.00$. We then can compare the measured ratio of differential cross sections to the new "rescaled" MCFM predictions, shown in Figure 8.12.

To determine the level of agreement between the measurement and the rescaled MCFM predictions we will measure the p-value for each bin. To we create four Gaussian distributions with means set at the value of the rescaled MCFM predictions and a width set as the total uncertainties on the measurements (found in Table 8.20). These Gaussians are filled with 10^6 "pseudo-experiments", we then count the number of experiments which attain or exceed

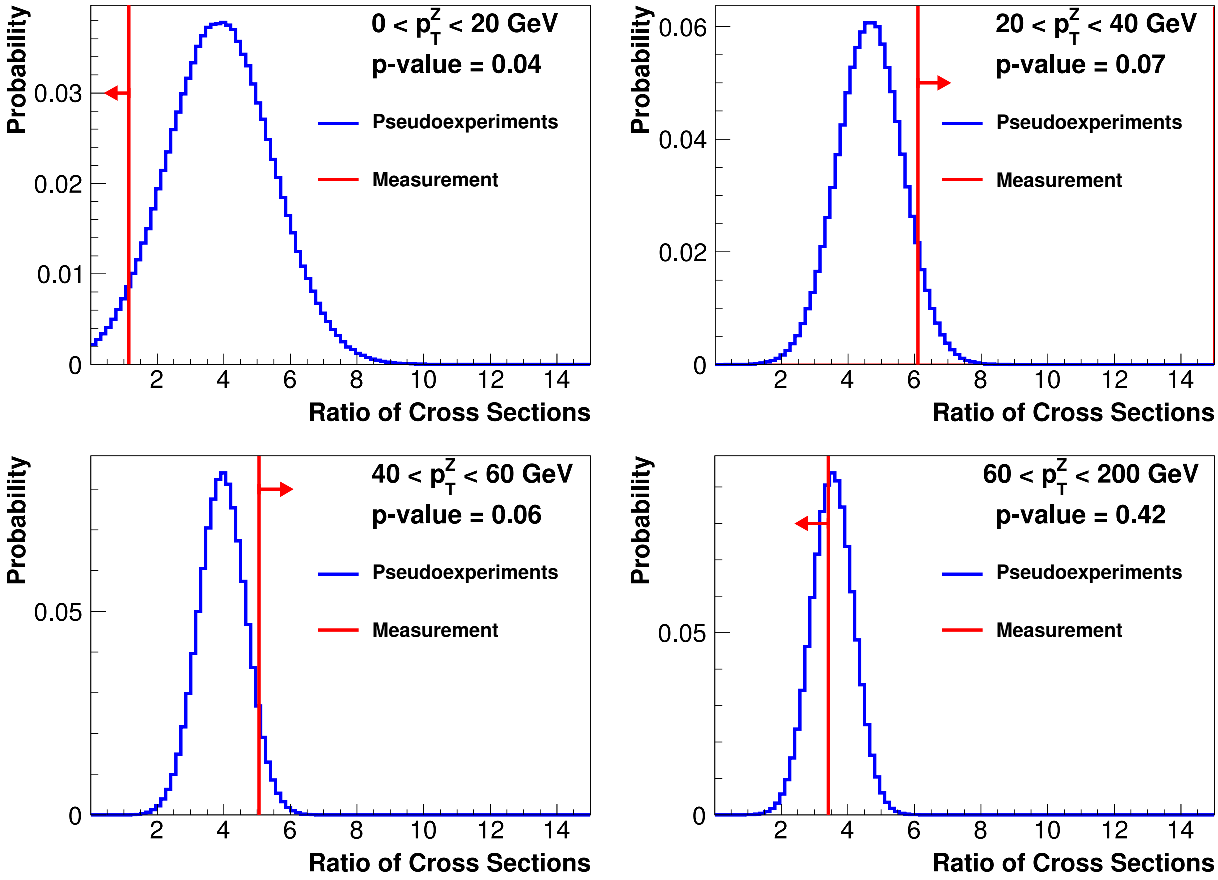


Figure 8.13: The Gaussian distributions used to determine the level of agreement between the measurement of $\sigma(Z{+}c\text{ jet})/\sigma(Z{+}b\text{ jet})$ as a function of Z boson p_T versus the rescaled MCFM predictions shown in Figure 8.12. The blue histograms represents the Gaussian distributions whose means are set to the rescaled MCFM predictions and the widths are set to the total uncertainties on the measurement. The red vertical line represents the value of the measurement.

the value of the measurement. These Gaussians and the p-values of that bin can be found in Figure 8.13. To combine the p-values from each bin we convert each of them into a χ^2 and sum them. This total χ^2 is then converted to a p-value for four degrees of freedom and it is found that the p-value for all four bins of Z boson p_T to simultaneously fluctuate to the observed measurement is 2%.

8.9 Cross checks

Due to the significant discrepancies between our measurements and the predictions from the NLO calculation we have performed many cross check to verify our results. Listed here are a sampling of the major cross checks performed. None of the cross checks performed show any systematic effect or pointed to a bias in the analysis.

8.9.1 Choice of MVA$_{\mathbf{bl}}$ OP

To verify that our choice of MVA$_{bl}$ selection does not affect our measurement we look at the ratio of

$$R = \frac{Nf_c}{\varepsilon^c_{btag}}$$

which is a reduced form of the cross section measurement which retains all the variables which depend on the OP choice. We can study the ratio of R for various choices of MVA$_{bl}$ OP and the R for a MegaTight requirement that was used in the $Z + c$ jet analysis. Figure 8.14 shows the ratio for multiple choices of the MVA$_{bl}$ requirements, including points where there are no correction factors generated. In this region we use the correction factor generated for the MegaTight OP. The uncertainty on this ratio is dominated by the fact that we are subtracting the $Z+$light jet contribution, which becomes very large at low values of MVA$_{bl}$ and is not actively modeled in the extrapolated regime. We see little variation in this ratio with respect to the choice of OP.

8.9.2 Ideal neutral strange hadron remover

The two main sources of the light jet misidentification rate for the MVA$_{bl}$ algorithm come from detector resolution effects and tracks coming from V^0, specifically K^0_s and Λ, decays. While the detector resolution effects are hard to quantify and measure, the effects of V^0

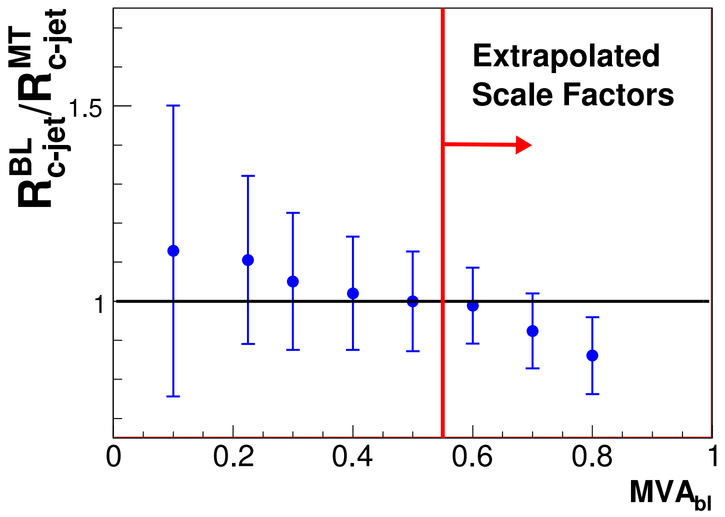


Figure 8.14: The ratio of Rs for a given MVA$_{bl}$ choice normalized to the R for the MegaTight operating point. The uncertainties are a combination of the statistical and light jet uncertainties. The ratio is consistent with one for MVA$_{bl}$ < 0.7. This shows that our analysis is not sensitive to choice of MVA$_{bl}$ requirement.

decays may be easier. These particles travel a measurable distance inside the detector before decaying, creating a displaced secondary vertex which mimics the signature of a HF jet.

These decays are already taken into account in the b jet identification framework, as described in Section 5.1.2. The "V^0 remover" looks for pairs of tracks inside jets whose invariant mass falls inside the mass windows of a V^0 (see Figure 5.3). When a pair of tracks is isolated in this mass range they are not passed into the algorithm. This lowers the overall misidentification rate due to these secondary vertices.

We can probe the effectiveness of this remover by creating our own "perfect" V^0 remover. This isolates MC truth level V^0s in simulation and selects their decay vertex. This method allows us to select all tracks which are coming from the decay of a V^0, including the single track semileptonic decays. When these tracks are located they are removed and not used in the algorithm.

We can use this to compare the misidentification rate for the standard V^0 remover against the new "perfect" V^0 remover. The comparison of the misidentification rates is shown in

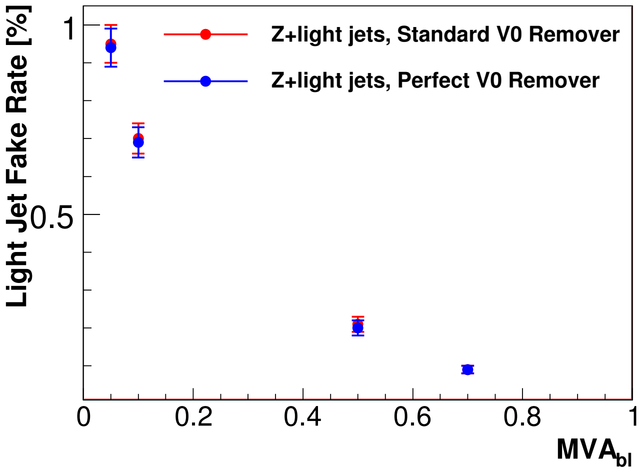


Figure 8.15: The light jet misidentification rate using the standard V^0 remover and our "perfect" V^0 remover, measured as a function of MVA$_{bl}$ requirement in data.

Figure 8.15 where we see the two methods are consistent. It can be concluded from this figure that the main source of the light jet misidentification rate is not coming from the presence of V^0 tracks but instead due to the detector resolution effects. Additionally, in Figure 8.16 we can see these efficiencies measured as a function of jet p_T and η. The tracks from these vertices are correctly taken into account by the standard V^0 remover and the scale factors are reliable in this regard.

8.9.3 Systems of equations

In the 180 pb^{-1} D0 $Z + b$ jet measurement [10] the solution of a system of three equations, instead of fitting templates, was used to extract the ratio of cross sections. To do this, information about the event yield after b jet identification, the efficiencies, and one assumption, that the ratio of $Z + b$ to $Z + c$ jet events can be taken from NLO pQCD predictions, was used and the ratio of cross sections was measured. If we replace this assumption with a constraint derived from the $Z + b$ jet analysis (see Chapter 7), we can repeat this exercise as

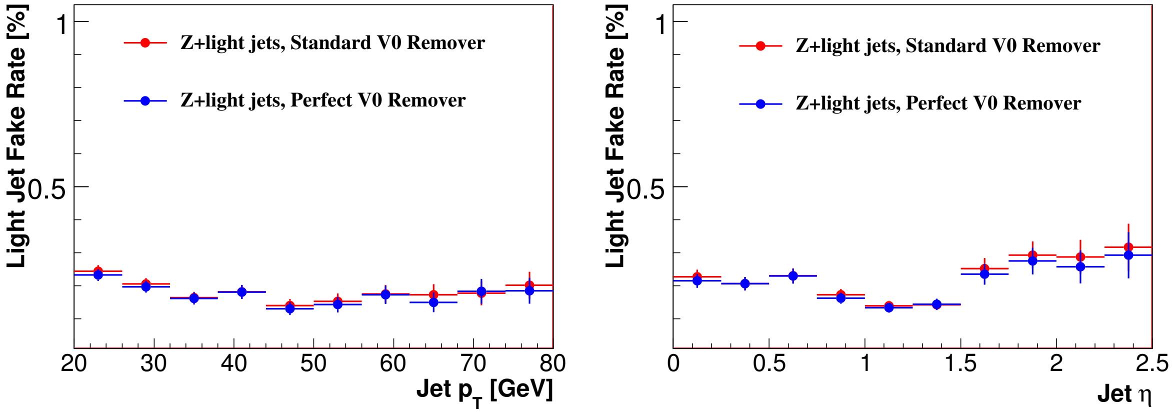


Figure 8.16: The light jet misidentification rate using the standard V^0 remover and our "perfect" V^0 remover for the $\text{MVA}_{bl} > 0.5$, measured as a function of jet p_T and jet η in data.

Table 8.21: Table of efficiencies and event totals used in ratio calculation.

N^{Z+jet}_{tagg}	N^{Z+jet}_{btag}	ϵ^b_{tagg}	ϵ^c_{tagg}	ϵ^l_{tagg}	ϵ^b_{btag}	ϵ^c_{btag}	ϵ^l_{btag}
$148,405$	$2,468$	89%	88%	88%	38%	8.5%	0.228%

a cross check. First we build three equations:

$$\frac{N^{Z+b}}{N^{Z+b}+N^{Z+c}+N^{Z+l}} = 1.96\%$$

$$N^{Z+jet}_{tagg} = \epsilon^b_{tagg}N^{Z+b} + \epsilon^c_{tagg}N^{Z+c} + \epsilon^l_{tagg}N^{Z+l}$$

$$N^{Z+jet}_{btag} = \epsilon^b_{btag}\epsilon^b_{tagg}N^{Z+b} + \epsilon^c_{btag}\epsilon^c_{tagg}N^{Z+c} + \epsilon^l_{btag}\epsilon^l_{tagg}N^{Z+l}$$

where N^{Z+X} is the number of $Z+X$ jet events where X is the flavor of that jet, N^{Z+jet}_{tagg} is the number of background-subtracted Z+jet events in data after the taggability requirement is applied, and N^{Z+jet}_{btag} is the number of background-subtracted Z+jet events in data after a MVA_{bl} requirement is applied. These values, along with the determined efficiencies, can be found in Table 8.21.

Using these values to constrain our equations we can then solve for the ratio of $Z+c$ jet events, without relying on the extracted flavor fractions or using the template fitting method.

This results in event yields

$$N^{Z+l} = 151,521,\ N^{Z+c} = 13,676.5, \text{ and } N^{Z+b} = 3,405.79,$$

which translate into ratios of

$$\frac{N^{Z+c}}{N^{Z+b} + N^{Z+c} + N^{Z+l}} = 8.11\%$$

and

$$\frac{N^{Z+c}}{N^{Z+b}} = 4.01.$$

These are in agreement with the nominal ratio given in Section 8.8, absolving the flavor extraction method of causing this discrepancy.

8.9.4 Dependence on Z boson transverse momentum

Due to the large variation of $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$ in the first two bins of Z boson p_T, we want to verify that we are observing an actual feature of the ratio and not a relic of either detector resolution or some other effect. To check this dependence we measure the ratio of cross sections separately in the two lepton channels. The resolution for measuring the momentum of the dielectron system is roughly 3 GeV at 40 GeV [44], while the resolution for the dimuon system is roughly 7 GeV at 40 GeV [43]. Looking at Figure 8.17 we see that both channels provide consistent results with the combined result.

A separate cross check of the behavior of $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$ as a function of Z boson p_T is whether there is a smooth transition between the first and second bin. A discontinuity could imply a bias in the analysis approach. The range of $0 < p_T^Z < 50$ GeV was divided into 8 bins, and the analysis was performed in each of these bins. The results, seen in Figure 8.18, show a smooth variation from a minimum value at $0 < p_T^Z < 10$ GeV

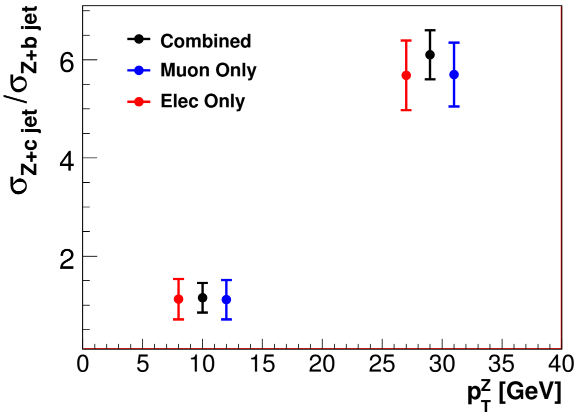


Figure 8.17: The first two bins of $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$ as a function of Z p_T in the separate dilepton channels. The result in the electron, muon, and their combination agree within uncertainties.

to a maximum at $25 < p_T^Z < 30$ GeV. This cross check does not included any bin-to-bin migration effects which become non-negligible with fine binning.

8.9.5 High jet transverse momentum

Referring back to Figure 8.11 we see a sharp transition for $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$ as a function of Z boson p_T which occurs at $p_T = 20$ GeV. To determine whether this is a bias brought on by our requirement that jets must have $p_T > 20$ GeV we can apply a stricter p_T requirement. Due to the limited statistics of the $Z + \text{HF}^{MT}$ sample we will instead examine the effects on $\sigma(Z + b\text{ jet})/\sigma(Z + \text{jet})$. This benefits from much larger statistics, and since we can interpret $\sigma(Z + c\text{ jet})/\sigma(Z + b\text{ jet})$ as the ratio of $\sigma(Z + c\text{ jet})/\sigma(Z + \text{jet})$ and $\sigma(Z + b\text{ jet})/\sigma(Z + \text{jet})$ it provides us with valuable information about the overall behavior.

We apply a p_T requirement on the selected jets of $p_T > 45$ GeV and repeat the measurement of $\sigma(Z + b\text{ jet})/\sigma(Z + \text{jet})$ as a function of Z boson p_T. We expect an overall increase in the ratio due to the fact that we measured the jet p_T dependence of the ratio and know that it increases. Looking at Figure 8.19 we see the first two bins of $\sigma(Z + b\text{ jet})/\sigma(Z + \text{jet})$ as a function of Z boson p_T for the nominal jet p_T requirement and the harder requirement. We see that while the harder criteria yields a slightly higher integrated cross section the

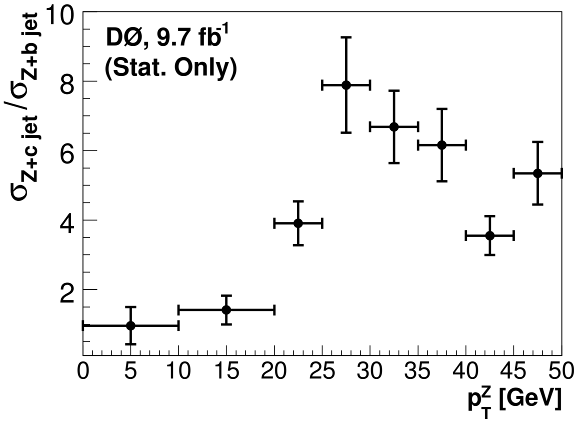


Figure 8.18: $\sigma(Z + c\ \text{jet})/\sigma(Z + b\ \text{jet})$ as a function of Z boson p_T in the range of $0 < p_T^Z < 50$ GeV. Bin-to-bin migrations have been neglected and the uncertainties shown are statistical only. A smooth transition is observed from the first bin until the peak of the distribution at $25 < p_T^Z < 30$ GeV.

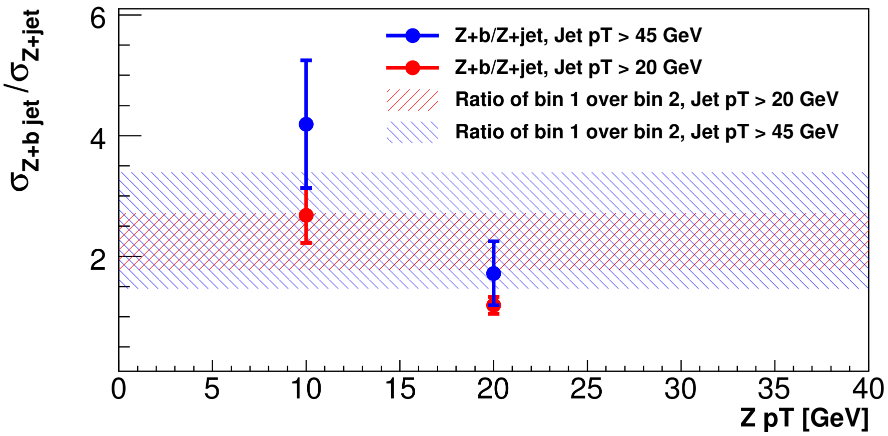


Figure 8.19: The first two bins of $\sigma(Z + b\ \text{jet})/\sigma(Z + \text{jet})$ as a function of Z boson p_T for two different jet p_T requirements. The two bands represent the ratio of the first bin to the second, where the width of the band is the uncertainty on the ratio.

shape of the distribution does not change, further when we compare the ratio of the first and second bin for the two cases we see that they agree within uncertainties.

9 Conclusions and outlook

Using new methods, including a novel method for measuring the MVA_{bl} misidentification rate from data, we have made the first measurements of the ratio of differential cross sections $\sigma(Z+b\text{ jet})/\sigma(Z+\text{jet})$. Additionally we have presented the first measurements of $Z+c$ jet production by studying the integrated and differential ratios of cross sections $\sigma(Z+c\text{ jet})/\sigma(Z+\text{jet})$ and $\sigma(Z+c\text{ jet})/\sigma(Z+b\text{ jet})$. The integrated ratio of cross sections $\sigma(Z+b\text{ jet})/\sigma(Z+\text{jet})$ was measured with the lowest uncertainties to date, improving the precision by 53% over previous measurements. These precision measurements have been compared to NLO pQCD calculations and have shown areas of significant disagreement.

Using the full D0 Run II data set corresponding to an integrated luminosity of 9.7 fb^{-1}, with jets with $p_T > 20$ GeV and $|\eta| < 2.5$, for the combined $Z\rightarrow\ell\ell\,(\ell = e, \mu)$ final state we have measured

$$\frac{\sigma\,(Z+b\text{ jet})}{\sigma\,(Z+\text{jet})} = (1.96 \pm 0.12\,(\text{stat}) \pm 0.13\,(\text{syst}))\,\%,$$

$$\frac{\sigma(Z+c\text{ jet})}{\sigma(Z+\text{jet})} = (8.29 \pm 0.52(\text{stat}) \pm 0.89(\text{syst}))\,\%,$$

and

$$\frac{\sigma(Z+c\text{ jet})}{\sigma(Z+b\text{ jet})} = 4.00 \pm 0.21(\text{stat}) \pm 0.58(\text{syst}).$$

The integrated ratios of cross sections have been compared to NLO pQCD predictions and, whereas $\sigma(Z + b\ \text{jet})/\sigma(Z + \text{jet})$ agreed within uncertainties [82], the $Z+c$ jet measurements significantly disagree [84]. The disagreement observed in the $Z + c$ jet measurements is of the same order as what was observed by related measurements at the Tevatron [25, 26], LEP [14, 15, 16, 17, 18, 19, 20, 24], and SLD [14, 20].

Additionally, by utilizing the unique characteristics of the fragmenting b-quark we have created a new algorithm which allows for the identification of b jets with high precision. The new MVA_{bl} algorithm shows significant improvement over previous algorithms derived at D0. For a fixed misidentification rate of 1% the MVA_{bl} algorithm sees an increase in the efficiency of selecting a b jet of 15% per jet. A new method for extracting the misidentification rate directly from data has also been presented. The SystemN method's data-driven misidentification rates show that the previous method systematically underestimated these rates. This difference is due to the inability of the simulation to accurately model resolution and mis-reconstruction effects.

These measurements can be used to help model the $Z+\text{HF}$ jet backgrounds for Higgs boson searches. At the Tevatron, MCFM is used to predict the cross sections of these processes, as described in Section 4.1.2, and the discrepancies that we have found point to deficiencies with the current models. The high level of precision that we have measured these cross sections can also aid in reducing the uncertainties associated with the modeling of these backgrounds, hence increasing the sensitivity of future Higgs boson searches. An important follow up would be for the ATLAS and CMS collaborations to perform their own measurements in their higher energy regime. The high energies of LHC leads to an increase in the contributions from Compton scattering processes and will reduce the effect of gluon splitting. Hence these measurements will sample a different area of phase space.

References

[1] S. Chatrchyan *et al.*, CMS Collaboration, *Combined results of searches for the standard model Higgs boson in pp collisions at* $\sqrt{s} = 7$ *TeV*, Phys. Lett. **B710**, 26 (2012).

[2] G. Aad *et al.*, ATLAS Collaboration, *Combined search for the Standard Model Higgs boson in pp collisions at* $\sqrt{s} = 7$ *TeV with the ATLAS detector*, Phys. Rev. **D86**, 032003 (2012).

[3] S. Heinemeyer *et al.*, LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties*, 1307.1347, (2013).

[4] T. Aaltonen *et al.*, CDF Collaboration and D0 Collaboration, *Higgs boson studies at the Tevatron*, Phys. Rev. D **88**, 052014 (2013).

[5] S.B. Beri, *et al.*, *Search for* $ZH \to \ell\ell b\bar{b}$ *in 9.7* fb^{-1} *Using a Random Forest in Dimuon and Dielectron Events*, (2012), `www-d0.fnal.gov/Run2Physics/higgs/d0_private/docs/Notes/ICHEP2012/zh_llbb_analysis_v5.00.pdf`.

[6] J. M. Campbell, R. K. Ellis, F. Maltoni, and S. Willenbrock, *Associated production of a Z Boson and a single heavy quark jet*, Phys. Rev. **D69**, 074021 (2004).

[7] F. Febres Cordero, L. Reina, and D. Wackeroth, *NLO QCD corrections to* $Zb\bar{b}$ *production with massive bottom quarks at the Fermilab Tevatron*, Phys. Rev. **D78**, 074014 (2008).

[8] S. Dawson, C. Jackson, L. Reina, and D. Wackeroth, *Higgs boson production with one bottom quark jet at hadron colliders*, Phys.Rev.Lett. **94**, 031802 (2005).

[9] J. M. Campbell and R. Ellis, *MCFM for the Tevatron and the LHC*, Nucl. Phys. B (Proc. Suppl.) **205-206**, 10 (2010).

[10] V. Abazov *et al.*, D0 Collaboration, *A measurement of the ratio of inclusive cross sections* $\sigma(pp \to Z + b - \text{jet})/\sigma(pp \to Z + jet)$ *at* $\sqrt{s} = 1.96$ *TeV*, Phys. Rev. Lett. **94**, 161801 (2005).

[11] A. Abulencia *et al.*, CDF Collaboration, *Measurement of the b jet cross-section in events with a Z boson in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, Phys. Rev. **D74**, 032008 (2006).

[12] T. Aaltonen *et al.*, CDF collaboration, *Measurement of Cross Sections for b Jet Production in Events with a Z Boson in p anti-p Collisions at* $\sqrt{s} = 1.96$ *TeV*, Phys. Rev. **D79**, 052008 (2009).

[13] V. Abazov *et al.*, D0 Collaboration, *A measurement of the ratio of inclusive cross sections* $\sigma(p\bar{p} \to Z + b\,\text{jet})/\sigma(p\bar{p} \to Z + \text{jet})$ *at* $\sqrt{s} = 1.96$ *TeV*, Phys. Rev. **D83**, 031105 (2011).

[14] A. Giammanco, *Gluon splitting into heavy flavours at the Z peak*, Proceeding for DIS 2004 , 783 (2004).

[15] A. Heister *et al.*, ALEPH Collaboration, *A measurement of the gluon splitting rate into c anti-c pairs in hadronic Z decays*, Phys. Lett. **B561**, 213 (2003).

[16] R. Barate *et al.*, ALEPH Collaboration, *A Measurement of the gluon splitting rate into b anti-b pairs in hadronic Z decays*, Phys. Lett. **B434**, 437 (1998).

[17] P. Abreu *et al.*, DELPHI Collaboration, *Measurement of the multiplicity of gluons splitting to bottom quark pairs in hadronic* Z^0 *decays*, Phys. Lett. **B405**, 202 (1997).

[18] M. Acciarri *et al.*, L3 Collaboration, *Measurement of the probability of gluon splitting into charmed quarks in hadronic Z decays*, Phys. Lett. **B476**, 243 (2000).

[19] C. Adloff *et al.*, H1 Collaboration, *Measurement of neutral and charged current cross-sections in electron - proton collisions at high* Q^2, Eur. Phys. J. **C19**, 269 (2001).

[20] K. Abe *et al.*, SLD Collaboration, *Measurement of the probability for gluon splitting into b anti-b in* Z^0 *decays*, hep-ex/9908028, (1999).

[21] M. Seymour, *Heavy quark pair multiplicity in e+ e- events*, Nucl.Phys. **B436**, 163 (1995).

[22] D. Miller and M. H. Seymour, *Secondary heavy quark pair production in e+ e- annihilation*, Phys.Lett. **B435**, 213 (1998).

[23] C. Amsler *et al.*, Particle Data Group, *Review of Particle Physics*, Phys. Lett. **B667**, 1 (2008).

[24] M. Seymour, *Heavy quark pair multiplicity in* e^+e^- *events*, Nucl. Phys. **B436**, 163 (1995).

[25] V. Abazov *et al.*, D0 Collaboration, *Measurement of the differential photon + c-jet cross section and the ratio of differential photon+ c and photon+ b cross sections in proton-antiproton collisions at* $\sqrt{s} = 1.96$ *TeV*, Phys. Lett. **B719**, 354 (2013).

[26] T. Aaltonen *et al.*, CDF Collaboration, *Measurement of the cross section for direct-photon production in association with a heavy quark in* $p\bar{p}$ *collisions at* $\sqrt{s}$ *= 1.96 TeV*, Phys. Rev. Lett. **111**, 042003 (2013).

[27] FNAL Accelerator Division - Operations Department, *Accelerator Concept: Rookie Book*, 2013.

[28] Science at Fermilab webpage, Accelerators, Fermilab's Tevatron.

[29] FNAL Accelerator Division - Operations Department, *Linac Rookie Book*, 2012.

[30] FNAL Accelerator Division - Operations Department, *BOOSTER*, 2009.

[31] FNAL Accelerator Division - Operations Department, *Antiproton Source Rookie Book*, 2011.

[32] FNAL Accelerator Division - Operations Department, *Main Injector Rookie Book*, 2003.

[33] FNAL Accelerator Division - Operations Department, *Tevatron Rookie Book*, 2006.

[34] V. Abazov *et al.*, D0 Collaboration, *The Upgraded D0 detector*, Nucl. Instrum. Methods **A565**, 463 (2006).

[35] D0 Detector Drawings, webpage, `http://www-d0.fnal.gov/Run2Physics/WWW/drawings.htm`.

[36] S. Ahmed *et al.*, D0 Collaboration, *The D0 Silicon Microstrip Tracker*, Nucl. Instrum. Methods **A634**, 8 (2011).

[37] R. Angstadt *et al.*, D0 Collaboration, *The Layer 0 Inner Silicon Detector of the D0 Experiment*, Nucl. Instrum. Methods **A622**, 298 (2010).

[38] S. Ahmed *et al.*, D0 Collaboration, *The D0 Silicon Microstrip Tracker*, Nucl. Instrum. Methods **A634**, 8 (2011).

[39] S. Abachi *et al.*, D0 Collaboration, *Beam tests of the D0 uranium liquid argon end calorimeters*, Nucl. Instrum. Methods **A324**, 53 (1993).

[40] V. Abazov *et al.*, *The Muon system of the Run II D0 detector*, Nucl. Instrum. Methods **A552**, 372 (2005).

[41] M. Abolins *et al.*, *The Run IIb trigger upgrade for the D0 experiment*, IEEE Trans. Nucl. Sci. **51**, 340 (2004).

[42] P. Calfayan, *ORing single muon triggers in p17 data*, D0 Note 5329 (2007).

[43] V. Abazov *et al.*, D0 Collaboration, *Muon reconstruction and identification with the Run II D0 detector*, 1307.5202, (2013).

[44] V. Abazov *et al.*, D0 Collaboration, *Electron and Photon Identification in the D0 Experiment*, to be submitted to Nucl. Instrum. Methods A (2013).

[45] M. B. Su-Jung Park, *Efficiency of the Data Quality Calorimeter Flags*, D0 Note 5324 (2007).

[46] Y. Fisyak and J. Womersley, *D0gstar, D0 GEANT Simulation of the Total Apparatus Response*, D0 Note 3191 (2002).

[47] A. Schwartzman and C. Tully, *Primary Vertex Reconstruction by Means of Adaptive Vertex Fitting*, D0 Note 4918 (2005).

[48] S.B. Beri *et al.*, *Search for* $ZH \rightarrow \ell\ell b\bar{b}$ *in 8.6* fb^{-1} *of Data Using a Random Forest in Dimuon and Dielectron Events*, D0 Note 6158 (2011).

[49] G. Blazey *et al.*, *Run II jet physics*, hep-ex/0005012, (2000).

[50] V. M. Abazov *et al.*, D0 Collaboration, *Measurement of the inclusive jet cross section in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ *TeV*, Phys. Rev. **D85**, 052006 (2012).

[51] D. Lincoln *et al.*, *Taggability Studies in p20*, D0 Note 5773 (2009).

[52] V. Abazov *et al.*, D0 Collaboration, *Jet Energy Scale Determination at D0 Run 2*, to be submitted to Nucl. Instrum. Methods A (2013).

[53] S. Caughron, *Semi-muonic b-jet Corrections*, JES Meeting 20/12/2011.

[54] T. Sjostrand, S. Mrenna, and P. Z. Skands, *PYTHIA 6.4 Physics and Manual*, JHEP **0605**, 026 (2006).

[55] M. L. Mangano, M. Moretti, F. Piccinini, R. Pittau, and A. D. Polosa, *ALPGEN, a generator for hard multiparton processes in hadronic collisions*, JHEP **0307**, 001 (2003).

[56] F. Caravaglios, M. L. Mangano, M. Moretti, and R. Pittau, *A New approach to multijet calculations in hadron collisions*, Nucl. Phys. **B539**, 215 (1999).

[57] J. Pumplin *et al.*, *New generation of parton distributions with uncertainties from global QCD analysis*, JHEP **0207**, 012 (2002).

[58] V. Abazov *et al.*, D0 Collaboration, *b-Jet Identification in the D0 Experiment*, Nucl. Instrum. Methods **A620**, 490 (2010).

[59] J. M. Campbell and R. K. Ellis, *An Update on vector boson pair production at hadron colliders*, Phys. Rev. **D60**, 113006 (1999).

[60] U. Langenfeld, S. Moch, and P. Uwer, *Measuring the running top-quark mass*, Phys. Rev. **D80**, 054009 (2009).

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

[62] O. Brandt *et al.*, *Muon Identification Certification for the Summer 2009 Extended Dataset*, D0 Note 6025 (2010).

[63] J. Gaiser, *Charmonium Spectroscopy From Radiative Decays of the J/ψ and ψ'*, PhD thesis, 1982, SLAC-0255, UMI-83-14449-MC, SLAC-R-0255, SLAC-R-255.

[64] M. Shamim and T. Bolton, *Generator Level Reweighting of Z boson p_T*, D0 Note 5565 (2008).

[65] T. Gleisberg *et al.*, *Event generation with SHERPA 1.1*, JHEP **0902**, 007 (2009).

[66] L. Breiman, *Random Forests*, Machine Learning **45**, 5 (2001).

[67] A. Hocker *et al.*, *TMVA - Toolkit for Multivariate Data Analysis*, PoS **ACAT**, 040 (2007).

[68] R. Brun and F. Rademakers, *ROOT: An object oriented data analysis framework*, Nucl. Instrum. Meth. **A389**, 81 (1997).

[69] J. Freeman *et al.*, *Introduction to HOBIT, a b-Jet Identification Tagger at the CDF Experiment Optimized for Light Higgs Boson Searches*, Nucl. Instrum. Methods **A697**, 64 (2013).

[70] S. Chatrchyan *et al.*, CMS Collaboration, *Identification of b-quark jets with the CMS experiment*, JINST **8**, P04013 (2013).

[71] G. Aad *et al.*, ATLAS Collaboration, *Performance of the ATLAS b-tagging algorithms*, (2009), ATL-PHYS-PUB-2009-018, ATL-COM-PHYS-2009-206.

[72] D0 Luminosity Group, `http://www-d0.fnal.gov/runcoor/RUN/run2_lumi.html`.

[73] T. Gadfort, S. Greder, T. Scanlon, *Performance of the D0 MVA b-tagging Tool on Lepton-Photon 2009 Run IIb Data*, (2009), `http://www-d0.fnal.gov/Run2Physics/d0_private/eb/Run2EB_033/p20mva/p20-mva-v01.pdf`.

[74] V. M. Abazov *et al.*, *Measurement of the photon+b-jet production differential cross section in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96$ TeV, Phys. Lett. **B 714**, 32 (2012).

[75] K. J. Smith, *Measurement of the Ratio of Inclusive Cross Sections* $\sigma(p\bar{p} \rightarrow Z + b - \text{jet})/\sigma(p\bar{p} \rightarrow Z + \text{jet})$ *in the Dilepton Final States*, PhD thesis, University at Buffalo, 2010, FERMILAB-THESIS-2010-68.

[76] G. Lafferty and T. Wyatt, *Where to stick your data points: The treatment of measurements within wide bins*, Nucl. Instrum. Methods **A355**, 541 (1995).

[77] R. J. Barlow and C. Beeston, *Fitting using finite Monte Carlo samples*, Computer Phys. Comm. **77**, 219 (1993).

[78] U. Langenfeld, S. Moch, and P. Uwer, *Measuring the running top-quark mass*, Phys. Rev. **D80**, 054009 (2009).

[79] V. Abazov *et al.*, D0 Collaboration, *Precise measurement of the top-quark mass from lepton+jets events at D0*, Phys. Rev. **D84**, 032004 (2011).

[80] A. Martin, W. Stirling, R. Thorne, and G. Watt, *Parton distributions for the LHC*, Eur. Phys. J. **C63**, 189 (2009).

[81] S. Catani, F. Krauss, R. Kuhn, and B. Webber, *QCD matrix elements + parton showers*, JHEP **0111**, 063 (2001).

[82] V. Abazov *et al.*, D0 Collaboration, *Measurement of the ratio of differential cross sections* $\sigma(p\bar{p} \to Z + b\ jet)/\sigma(p\bar{p} \to Z + jet)$ *in* $p\bar{p}$ *collisions at* $\sqrt{(}s) = 1.96\ TeV$, Phys. Rev. **D87**, 092010 (2013).

[83] P. Z. Skands, *Tuning Monte Carlo Generators: The Perugia Tunes*, Phys. Rev. **D82**, 074018 (2010).

[84] V. Abazov *et al.*, D0 Collaboration, *Measurement of associated production of Z bosons with charm quark jets in* $p\bar{p}$ *collisions at* $\sqrt{s} = 1.96\ TeV$, 1308.4384, (2013).